DISCONTINUOUS GALERKIN FINITE ELEMENT METHOD FOR THE NONLINEAR HYPERBOLIC PROBLEMS WITH ENTROPY-BASED ARTIFICIAL VISCOSITY STABILIZATION

A Dissertation

by

VALENTIN NIKOLAEVICH ZINGAN

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

Doctor of Philosophy

May 2012

Major Subject: Nuclear Engineering

DISCONTINUOUS GALERKIN FINITE ELEMENT METHOD FOR THE NONLINEAR HYPERBOLIC PROBLEMS WITH ENTROPY-BASED ARTIFICIAL VISCOSITY STABILIZATION

A Dissertation

by

VALENTIN NIKOLAEVICH ZINGAN

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

Doctor of Philosophy

Approved by:

Co-Chairs of Committee,	Jim E. Morel
	Jean-Luc Guermond
Committee Members,	Marvin L. Adams
	Jean C. Ragusa
	Pavel V. Tsvetkov
Head of Department,	Yassin A. Hassan

May 2012

Major Subject: Nuclear Engineering

ABSTRACT

Discontinuous Galerkin Finite Element Method for the Nonlinear Hyperbolic Problems with Entropy-Based Artificial Viscosity Stabilization. (May 2012)
Valentin Nikolaevich Zingan, B.S.N.E., Moscow State Engineering Physics Institute (MEPhI); M.S.N.E., Moscow State Engineering Physics Institute (MEPhI)

Co-Chairs of Committee: Dr. Jim E. Morel
Dr. Jean-Luc Guermond

This work develops a discontinuous Galerkin finite element discretization of nonlinear hyperbolic conservation equations with efficient and robust high order stabilization built on an entropy-based artificial viscosity approximation.

The solutions of equations are represented by elementwise polynomials of an arbitrary degree $p > 0$ which are continuous within each element but discontinuous on the boundaries. The discretization of equations in time is done by means of high order explicit Runge-Kutta methods identified with respective Butcher tableaux.

To stabilize a numerical solution in the vicinity of shock waves and simultaneously preserve the smooth parts from smearing, we add some reasonable amount of artificial viscosity in accordance with the physical principle of entropy production in the interior of shock waves. The viscosity coefficient is proportional to the local size of the residual of an entropy equation and is bounded from above by the first-order artificial viscosity defined by a local wave speed. Since the residual of an entropy equation is supposed to be vanishingly small in smooth regions (of the order of the Local Truncation Error) and arbitrarily large in shocks, the entropy viscosity is almost zero everywhere except the shocks, where it reaches the first-order upper bound.

One- and two-dimensional benchmark test cases are presented for nonlinear hyperbolic scalar conservation laws and the system of compressible Euler equations. These tests demonstrate the satisfactory stability properties of the method and optimal convergence rates as well. All numerical solutions to the test problems agree well with the reference solutions found in the literature.

We conclude that the new method developed in the present work is a valuable alternative to currently existing techniques of viscous stabilization.

DEDICATION

To my father Nikolay and my mother Svetlana

ACKNOWLEDGMENTS

This research project would not have been possible without the support of many people. First of all, I wish to express my gratitude to my advisor, Prof. Jim E. Morel who was abundantly helpful and offered invaluable assistance, support and guidance. I would like to sincerely thank my co-advisor Prof. Jean-Luc Guermond for giving me valuable ideas and getting me familiar with the world of finite elements. Prof. Bojan Popov is gratefully acknowledged for his collaboration and useful suggestions.

I would like to express my special thanks to the deal.II community, especially Prof. Wolfgang Bangerth. I always received the valuable replies to all my requests.

I am forever indebted to my father Nikolay Zingan and my mother Svetlana Zingan for their love, understanding, endless patience and encouragement when it was most required. Finally, yet importantly, I would like to express my heartfelt thanks to my beloved wife Alena and my little son Daniel for their love, support and encouragement during the entire period of this work.

NOMENCLATURE

BC's	boundary conditions
BDF	backward differentiation formula
CD	centered difference
CFD	computational fluid dynamics
CFL	Courant-Friedrichs-Lewy
CG	continuous Galerkin
CN	Crank-Nicholson
DG	discontinuous Galerkin
DoF	degree of freedom
ENO	essentially non-oscillatory
EoS	equation of state
FE	finite element
FEM	finite element method
FVM	finite volume method
IBVP	initial boundary value problem
IVP	initial value problem
JLG	Jean-Luc Guermond
KPP	Kurganov-Petrov-Popov
LES	large eddy simulation
LF	Lax-Friedrichs
LTE	local truncation error
ODE	ordinary differential equation
PDE	partial differential equation
RK	Runge-Kutta
TV	total variation
TVD	total variation diminishing

WENO weighted essentially non-oscillatory

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

1. INTRODUCTION

1.1 Motivation

Inviscid compressible shock-dominated flows in multi-dimensional gas dynamics problems are modeled by nonlinear hyperbolic conservation laws whose solutions exhibit a wide range of localized structures, such as shock waves, contact discontinuities, and rarefaction waves. The most complex unsteady flows include multiple interactions of these structures with each other as well as with different kinds of bodies around which such flows exist. Some flows related to compressible turbulence simulations, aeroacoustics, or detonation processes exhibit both shocks and complicated smooth solution structures.

It is not an ordinary task to satisfactorily describe all possible characteristics of these phenomena in terms of numerical solutions of the governing equations. The accurate numerical resolution of the features of intricate flows necessitates the use of higher-order numerical methods. Unfortunately, higher-order numerical methods are known to produce spurious oscillations (Gibbs phenomenon) in the regions of strong discontinuities, which can also affect the smooth parts of the solution. To suppress these undesired effects, the most efficient methods rely at some point on limiters whose development and implementation in two and three space dimensions on unstructured meshes with an arbitrary polynomial degree is a highly nontrivial task. The theoretical understanding of the stability and convergence of these nonlinear methods is currently limited to uniform grids and scalar equations in one space dimension [1–5]. A true two-dimensional non-oscillatory reconstruction which could be applied to arbitrary unstructured meshes (without any additional post processing) seems to be available only in the piecewise linear case [6], and extensions to the higher degree polynomial reconstructions do not seem to be evident.

This dissertation follows the style of Journal of Computational Physics.

In this thesis, we introduce the combination of a *Discontinuous Galerkin Finite Element Method* with an *Entropy-Based Artificial Viscosity Approximation* [7] for the numerical solution of hyperbolic conservation laws. We avoid the use of limiters and non-oscillatory reconstructions by adding a degenerate nonlinear dissipation to the numerical discretization of the equation or system at hand. The new approach is shown to be robust and to perform well by testing on a wide variety of multi-dimensional benchmark problems including different types of hyperbolic scalar conservation laws and the system of compressible Euler equations.

1.2 Background

1.2.1 Governing equations

The Euler equations are the set of *Partial Differential Equations* (PDE) governing inviscid flows. The equations correspond to the Navier-Stokes equations with zero viscosity and zero thermal conduction terms and describe the conservation of mass, momentum, and total energy.

Although inviscid flows do not truly exist in nature, there are many practical flows where the influence of transport phenomena, which give rise to both viscosity and heat conduction, is small, and we can model the flow as being inviscid. Theoretically, inviscid flow is approached in the limit as the Reynolds number $Re \to \infty$. However, for practical problems, many flows with high but finite Re can be assumed to be inviscid. For such flows, the influence of diffusion, friction, and thermal conduction is limited to a very thin region adjacent to a body surface, a so-called boundary layer, and the remainder of the flow outside this region is essentially inviscid. For flows over slender bodies, such as airfoils, inviscid theory adequately predicts the pressure distribution and lift on the body and gives a valid representation of the streamlines and flow field away from the body. However, because friction (shear

stress) is a major source of aerodynamic drag, inviscid theories by themselves cannot adequately predict the total drag.

Another example related to the significant dissipation of momentum and energy is the interior of strong shock waves, which is a very thin dissipative layer across which the properties of flow change drastically. Because of large gradients in velocity and temperature, the mechanisms of friction and thermal conduction are very strong in this layer. These dissipative irreversible mechanisms always increase the entropy of the fluid. However, the thickness of such a layer is of the order of the mean free path of the molecules, which is approximately 10^{-5} cm. Therefore, for the majority of practical problems, the detailed knowledge about the distribution of properties of fluid within the interior of shock waves is insignificant. Instead, one can replace the transient layer with a jump of all fields of flow, such as particles velocity, density, pressure, etc.

The Euler equations can be applied to compressible as well as to incompressible flow - using either an appropriate *Equation of State* (EoS) or assuming a constant density and a velocity field with zero divergence, respectively. Although all flows, to a greater or lesser extent, are compressible, it is always safe to assume the density to be constant if the Mach number $M < 0.3$. On the other hand, high-speed flows (M near or above 1) must be treated as compressible. For such flows, density can vary over a wide range of magnitudes. Moreover, the compressibility of flow gives rise to shock waves formation.

1.2.2 Higher-order numerical methods

Unfortunately, it is essentially impossible to solve the fundamental equations governing fluid flow analytically, even with simplifying assumptions. Hence, one must employ numerical methods to solve fluid flow problems, which generally involve complex geometries and shapes. Simulation of flow around cars, airplanes, rockets, birds, and supersonic jets are actual examples of such problems from real life.

A *Finite Volume Method* (FVM) is the industry standard approach to *Computational Fluid Dynamics* (CFD) for inviscid compressible shock-dominated flows [8]. However, higher-order methods are not routinely used in the FVM community, despite the significant growth in computational resources. Instead, second-order accurate methods are the most prevalent. Higher-order spatial accuracy for the FVM is typically achieved by non-oscillatory polynomial reconstructions of cell or nodal averages. This creates expanded numerical stencils, which in turn complicate boundary condition discretizations and adversely impact iterative algorithms.

To meet the modern-day requirements of accuracy, stability, and applicability to real-world problems, we use a *Discontinuous Galerkin Finite Element Method* (DG FEM or DG method) [9]. The DG FEM is often referred to as a hybrid method since it combines features of both a *Continuous Galerkin Finite Element Method* (CG FEM or CG method) and the FVM. Within each element, the solution is approximated by a polynomial of degree $p \geq 0$ (as in CG FEM), while the continuity conditions applied to the solution are relaxed at the boundaries of elements (as in FVM), which necessitates the construction of interelement numerical fluxes. Theoretically, solutions may be obtained to an arbitrarily high order of accuracy.

While the DG FEM was proposed in the early 1970's by Reed and Hill for the neutron transport equation [10], it was not used for CFD simulations until the early 1990's when it was first used to solve the Euler equations by Cockburn and Shu [11]. The solution of the Navier-Stokes equations with the DG method was first accomplished by Bassi and Rebay in 1997 [12]. As the method gained more attention in the CFD research community, further advantages have come fairly rapidly. Researchers are now using the DG method to perform simulations of a wide variety of flow regimes. The method has been adapted to study compressible and incompressible, steady and unsteady, as well as laminar and turbulent flows.

In addition to being arbitrarily high order accurate, the DG method also permits the formulation of very compact numerical schemes. The compactness of the schemes

stems from the fact that the solution within each element is kept independent of the solutions in other elements, with interelement communication occurring only at the boundaries of the adjacent elements. This characteristic, along with other properties, makes this method extremely flexible in the sense of easily handling a wide variety of element types and mesh topologies, and also allows a number of adaptive techniques like $h-$ or $p-$ refinement or both [13–15].

1.2.3 Shock capturing

As mentioned above, shock waves are mathematically treated as discontinuities, but it was quickly recognized that this representation of the phenomenon caused problems for any higher-order numerical solution, because higher-order numerical methods are known to produce spurious unphysical oscillations (Gibbs phenomenon) in the regions of shocks. Discontinuities exist in the solutions of many hyperbolic conservation laws. Shock and contact waves can manifest themselves in the solutions of scalar equations, such as Burgers' equation, or systems of equations, such as the Euler equations. Numerical schemes designed to solve these PDEs must be able to capture any discontinuity that might arise in the solution.

There have been a number of higher-order numerical schemes developed in the CFD community to capture different kinds of discontinuities. The best known and successful classes of such schemes are *Total Variation Diminishing* (TVD) and *Essentially Non-Oscillatory* (ENO). Both classes of methods were primarily designed in the FVM framework, and therefore, utilize higher degree polynomial reconstructions as well as limiters to identify the so-called "trouble cells", namely the cells which might need the limiting procedure.

TVD :

Consider a one-dimensional hyperbolic scalar conservation law:

$$\partial_t u + \partial_x f(u) = 0. \tag{1.1}$$

The *Total Variation* (TV) of the solution u is defined to be:

$$TV(u) = \int \left| \frac{\partial u}{\partial x} \right| dx. \tag{1.2}$$

The TV of a numerical approximation u_h is the discrete analogue of (1.2):

$$TV(u_h) = \sum_i \left| u_{h,i+1} - u_{h,i} \right|. \tag{1.3}$$

A time dependent numerical method $u_h^n \mapsto u_h^{n+1}$ is said to be TVD if:

$$TV(u_h^{n+1}) \leq TV(u_h^n). \tag{1.4}$$

Bounded TV in a scheme implies that no new local extrema are created - the values of local minima do not decrease, and the values of local maxima do not increase. In flow regions where the limiters are active, that is, in the "trouble cells", TVD methods reduce the degree of approximating polynomial to the piecewise constant representation, leading to the solution which is TVD in the mean values of each cell. This idea was first incorporated into the DG method by Cockburn and Shu. In their work, they developed a DG scheme with Runge-Kutta time stepping and a slope limiter based on the minmod operator [16–20].

ENO :

The main disadvantage of the TVD methods is the first-order spatial accuracy in the regions of discontinuities. To overcome that, schemes of the ENO class employ multiple candidate stencils to choose the least oscillatory one, which then is used to reconstruct a higher-order polynomial representation from a set of local cell averages [21–23].

An improvement over the traditional ENO methods is the class of methods called *Weighted Essentially Non-Oscillatory* (WENO) [24–26]. The basic idea is the following: instead of using only one of the candidate stencils to form the reconstruction, one uses a weighted combination of all of them. The choice of weights is based on the local smoothness of the solution.

A few researchers have also applied the ENO/WENO methodology of shock capturing schemes to the DG formulation [27, 28]. The general approach is to replace the DG approximating polynomials in the "trouble cells" with the ENO/WENO reconstructed polynomials, which maintain the original cell averages, have the same order of accuracy as before, but are less oscillatory. This approach looks very promising in combining the advantages of the ENO/WENO schemes and the DG method.

1.2.4 Artificial viscosities

The use of artificial viscosities to solve hyperbolic conservation laws has been pioneered by Neumann and Richtmyer [29], and has been popularized later by Smagorinsky [30] for *Large Eddy Simulation* (LES) purposes and by Ladyzhenskaya [31,32] for theoretical purposes in the analysis of the Navier-Stokes equations. The early versions of artificial viscosities were overly dissipative. The interest for this technique has faded over the years, especially in the DG FEM literature, where upwinding and limiters have been shown to be efficient and to yield high order accuracy [9, 33–35].

Despite their indisputable success story, limiters have some disadvantages. For instance, limiters may not be consistent in the steady-state limit, and thus may

sometimes lead to difficulties when trying to use time marching schemes to reach steady-state solutions [36, §3.5] and [37]. Furthermore, up to a few exceptions [34, §3.4], slope limiting is essentially a one-dimensional concept that does not generalize easily to unstructured meshes in two and three space dimensions. The theoretical understanding of the stability and convergence properties of limiters is currently restricted to uniform grids and scalar equations in one space dimension [1–5]. A true two-dimensional non-oscillatory reconstruction which could be applied to arbitrary unstructured meshes (without any additional post processing) seems to be available only in the piecewise linear case [6], and extensions to the higher degree polynomial reconstructions do not seem to be evident. For the reasons stated above, the interest for artificial viscosities has recently been revived in the DG literature [36–40] and in the CG literature as well [7, 41].

In the present work, we use the nonlinear artificial viscosity, which is proportional to the local size of the residual of an entropy equation. For this reason, this nonlinear artificial viscosity is called the entropy viscosity. The entropy equation itself holds only in smooth flow regions, and becomes an inequality in shocks. Therefore, the basic idea is that the entropy viscosity is large in strong shocks and small (of the order of the truncation error) in the smooth parts of the flow. This simple idea is mesh and approximation independent and can be applied to equations or physical systems of equations that are supplemented with an auxiliary entropy inequality.

The use of a residual to construct an artificial viscosity is not a new idea. For instance, several works rely on the residuals of conservation equations which then are used to construct the so-called PDE-based artificial viscosity [39, 42]. However, using the residuals of PDEs may be less robust than using the entropy residual. This argument is based on the observation that consistency requires the residual of the PDE to vanish in the distribution sense as the mesh size goes to zero, whereas the very nature of entropy implies that the entropy residual converges to a Dirac measure supported in shocks.

Another immediate consequence of the choice of the entropy viscosity is that the viscosity coefficient is proportional to the residual of an entropy equation, which is known to be large in shocks and to be zero in contact discontinuities. As a result, this strategy makes an automatic distinction between shocks and contact discontinuities, and this subtle distinction cannot be made by the other artificial viscosity methods.

1.3 Thesis overview

The thesis is organized as follows: First we give a brief background on hyperbolic scalar conservation laws and the system of compressible Euler equations in Chapter II, where we will also mention some elementary aspects of equilibrium thermodynamics. Then in Chapter III we discuss the main idea of stabilization of numerical solutions to hyperbolic equations. In Chapter IV we give a general description of the entropy-based artificial viscosity approximation in the framework of DG methods for solving hyperbolic scalar conservation laws, and present computational results. We discuss the extension of the method to compressible gas dynamics equations and present computational results in Chapter V. We end the thesis with a chapter relating to our conclusions.

2. CONSERVATION LAWS

2.1 Hyperbolic scalar conservation laws

In this section we consider scalar conservation laws of hyperbolic type. Some difficulties encountered with systems of hyperbolic equations are discussed here.

2.1.1 Integral and differential forms

To understand how conservation laws arise from physical principles, we will begin by deriving the equation for conservation of mass in a d-dimensional gas dynamics problem. Let $\delta\Omega \in \mathbb{R}^d$ be an arbitrarily chosen open connected domain with the boundary $\delta\Gamma$. Let $t \in \mathbb{R}_+$ be an arbitrary time. If $u(\mathbf{x}, t)$ and $\mathbf{v}(\mathbf{x}, t)$ are the density and velocity of the gas, respectively, then the vector function $\mathbf{f}(\mathbf{x}, t) = u(\mathbf{x}, t)\mathbf{v}(\mathbf{x}, t)$ is the mass flux. The total mass of gas within $\delta\Omega$ at time t is defined as the integral of the density:

$$m(t) = \int_{\delta\Omega} u(\mathbf{x}, t) d\mathbf{x}. \tag{2.1}$$

We assume that mass is neither created nor destroyed, then the mass in $\delta\Omega$ can change only because of the gas flow across the boundary $\delta\Gamma$:

$$m(t + dt) = m(t) + \overbrace{\left[- \int_{\delta\Gamma_+} \mathbf{f}(\mathbf{x}, t) \cdot d\boldsymbol{\sigma} dt \right]}^{inflow} - \overbrace{\left[\int_{\delta\Gamma_-} \mathbf{f}(\mathbf{x}, t) \cdot d\boldsymbol{\sigma} dt \right]}^{outflow}, \tag{2.2}$$

where $m(t + dt)$ is the total mass of gas within $\delta\Omega$ at time $t + dt$, $\delta\Gamma_+$ and $\delta\Gamma_-$ are the inflow and outflow portions of the boundary $\delta\Gamma$, respectively, and $d\boldsymbol{\sigma} = d\sigma \mathbf{n}(\mathbf{x})$ is a vector area. The equation (2.2) can be modified to obtain the integral form of the mass conservation law:

$$\frac{d}{dt} \int_{\delta\Omega} u(\mathbf{x}, t) d\mathbf{x} + \oint_{\delta\Gamma} \mathbf{f}(\mathbf{x}, t) \cdot d\boldsymbol{\sigma} = 0. \tag{2.3}$$

Another integral form of the same conservation law is obtained by integrating (2.3) in time from any t_1 to any $t_2 > t_1$:

$$\int_{\delta\Omega} u(\mathbf{x}, t_2) d\mathbf{x} - \int_{\delta\Omega} u(\mathbf{x}, t_1) d\mathbf{x} + \int_{t_1}^{t_2} \oint_{\delta\Gamma} \mathbf{f}(\mathbf{x}, t) \cdot d\boldsymbol{\sigma} dt = 0. \tag{2.4}$$

To derive the differential form of mass conservation law, we must assume the differentiability of functions $u(\mathbf{x}, t)$ and $\mathbf{v}(\mathbf{x}, t)$. Then using:

$$\int_{\delta\Omega} u(\mathbf{x}, t_2) d\mathbf{x} - \int_{\delta\Omega} u(\mathbf{x}, t_1) d\mathbf{x} = \int_{t_1}^{t_2} \int_{\delta\Omega} \partial_t u(\mathbf{x}, t) d\mathbf{x} dt \tag{2.5}$$

and

$$\int_{t_1}^{t_2} \oint_{\delta\Gamma} \mathbf{f}(\mathbf{x}, t) \cdot d\boldsymbol{\sigma} dt = \int_{t_1}^{t_2} \int_{\delta\Omega} \nabla \cdot \mathbf{f}(\mathbf{x}, t) d\mathbf{x} dt \tag{2.6}$$

in (2.4) gives that for all $\delta\Omega \subset \mathbb{R}^d$ and for all $(t_1 < t_2) \in \mathbb{R}_+$:

$$\int_{t_1}^{t_2} \int_{\delta\Omega} \left[\partial_t u(\mathbf{x}, t) + \nabla \cdot \mathbf{f}(\mathbf{x}, t)\right] d\mathbf{x} dt = 0. \tag{2.7}$$

Since the domain $\delta\Omega \subset \mathbb{R}^d$ as well as times $(t_1 < t_2) \in \mathbb{R}_+$ are arbitrarily chosen, we conclude that in fact the integrand in (2.7) must be identically zero for all $\mathbf{x} \in \mathbb{R}^d$ and for all $t \in \mathbb{R}_+$:

$$\partial_t u(\mathbf{x}, t) + \nabla \cdot \mathbf{f}(\mathbf{x}, t) = 0 \tag{2.8}$$

or

$$\partial_t u(\mathbf{x}, t) + \nabla \cdot \left(u(\mathbf{x}, t)\mathbf{v}(\mathbf{x}, t)\right) = 0, \tag{2.9}$$

which is the desired differential form of the mass conservation law. Equation (2.8) or (2.9) must be typically solved in conjunction with the equations for conservation of momentum and total energy. These equations will be discussed later in Section 2.2 of the current chapter.

The derived conservation law can be solved in isolation only if the velocity $\mathbf{v}(\mathbf{x}, t)$ is known *a priori* or is known as a function of density $u(\mathbf{x}, t)$. If it is, then the mass

flux $\mathbf{f}(\mathbf{x},t)$ is a function of density $u(\mathbf{x},t)$ alone, say $\mathbf{f}(\mathbf{x},t)=\mathbf{f}(u(\mathbf{x},t))$, and the law of conservation of mass becomes a hyperbolic scalar conservation law whose integral and differential forms are given by:

$$\frac{d}{dt}\int_{\delta\Omega} u d\mathbf{x} + \oint_{\delta\Gamma} \mathbf{f}(u)\cdot d\boldsymbol{\sigma} = 0, \tag{2.10}$$

$$\int_{\delta\Omega} u(\mathbf{x},t_2)d\mathbf{x} - \int_{\delta\Omega} u(\mathbf{x},t_1)d\mathbf{x} + \int_{t_1}^{t_2}\oint_{\delta\Gamma} \mathbf{f}(u)\cdot d\boldsymbol{\sigma} dt = 0, \tag{2.11}$$

$$\partial_t u + \nabla\cdot\mathbf{f}(u) = 0. \tag{2.12}$$

To solve any of the equations (2.10)-(2.12), one must define appropriate initial and boundary conditions.

Although such decoupling does not happen in gas dynamics, it can occur in other applications where the same conservation law holds. For instance, it is possible for modeling flows of chemicals in water, when the velocity of flowing water is known from other sources. Another example is a traffic flow along a highway, where the velocity of vehicles depends only on their local density on the road [43,44]. Moreover, the hyperbolic scalar conservation law of the forms (2.10)-(2.12) serves as a solid basis for studying systems of hyperbolic equations whose solutions exhibit discontinuities.

At the end of this section we formulate two problems which involve the differential form (2.12) and hold in the classical sense only if the solution u is a differentiable function. The *Initial Value Problem* (IVP) for the d-dimensional hyperbolic scalar conservation law is defined as follows:

$$\begin{cases} \partial_t u + \nabla\cdot\mathbf{f}(u) = 0, & (\mathbf{x},t)\in\mathbb{R}^d\times\mathbb{R}_+, \\ u(\mathbf{x},0) = u_0(\mathbf{x}), & \mathbf{x}\in\mathbb{R}^d. \end{cases} \tag{2.13}$$

where $u = u(\mathbf{x}, t) : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}$ is the solution, $\mathbf{f}(u) : \mathbb{R} \to \mathbb{R}^d$ is a known differentiable function of the solution u called the inviscid flux, and $u_0(\mathbf{x})$ is the initial data which is also known. The *Initial Boundary Value Problem* (IBVP) is defined as follows:

$$\begin{cases} \partial_t u + \nabla \cdot \mathbf{f}(u) = 0, & (\mathbf{x}, t) \in \Omega \times \mathbb{R}_+, \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}), & \mathbf{x} \in \Omega. \end{cases} \tag{2.14}$$

where $\Omega \subset \mathbb{R}^d$ is an open connected domain with the boundary Γ. The appropriate choice of boundary conditions depends on the specific problem at hand and will be discussed later in Chapter IV.

2.1.2 Diffusion

Consider the problem of propagating some small amount of a chemical or pollutant in water, when the velocity of flowing water is *a priori* known. If this velocity is identically zero everywhere throughout the domain of interest, then the equations (2.10)-(2.12) yield that the density of the chemical remains unchanged as time evolves. This result does not reflect the physical reality observed in practice. We expect molecular diffusion to occur as well as advection. Since the chemical is still conserved, we can incorporate this effect into the derivation of the corresponded conservation law:

$$\begin{aligned} m(t+dt) = m(t) + & \overbrace{\left[-\int_{\delta\Gamma_+} \mathbf{f}(\mathbf{x}, t) \cdot d\boldsymbol{\sigma} dt \right]}^{inflow} - \overbrace{\left[\int_{\delta\Gamma_-} \mathbf{f}(\mathbf{x}, t) \cdot d\boldsymbol{\sigma} dt \right]}^{outflow} + \\ & \overbrace{\left[-\oint_{\delta\Gamma} \mathbf{J}^-(\mathbf{x}, t) \cdot d\boldsymbol{\sigma} dt \right]}^{inflow} - \overbrace{\left[\oint_{\delta\Gamma} \mathbf{J}^+(\mathbf{x}, t) \cdot d\boldsymbol{\sigma} dt \right]}^{outflow}, \end{aligned} \tag{2.15}$$

where $\mathbf{J}^-(\mathbf{x},t)$ and $\mathbf{J}^+(\mathbf{x},t)$ are the partial molecular currents and all notations now relate to the properties of the chemical. The latter equation can be modified as:

$$\frac{d}{dt}\int_{\delta\Omega} u(\mathbf{x},t)d\mathbf{x} + \oint_{\delta\Gamma} \mathbf{f}(\mathbf{x},t)\cdot d\boldsymbol{\sigma} = -\oint_{\delta\Gamma} \mathbf{J}(\mathbf{x},t)\cdot d\boldsymbol{\sigma}, \tag{2.16}$$

where $\mathbf{J}(\mathbf{x},t) = \mathbf{J}^+(\mathbf{x},t)+\mathbf{J}^-(\mathbf{x},t)$ is the molecular net current. Further modifications of this equation along with the Fick's law:

$$\mathbf{J}(\mathbf{x},t) = -\epsilon(\mathbf{x})\nabla u(\mathbf{x},t), \tag{2.17}$$

where for all $\mathbf{x}\in\mathbb{R}^d : \epsilon(\mathbf{x}) > 0$ is the diffusion coefficient, will eventually lead to the following PDE:

$$\partial_t u + \nabla\cdot\mathbf{f}(u) = \nabla\cdot(\epsilon\nabla u). \tag{2.18}$$

Equation (2.18) is the differential form of a scalar conservation law of parabolic type. The major difference between equations (2.12) and (2.18) is that (2.18) always has smooth unique solutions, even if the initial data is discontinuous. We can view (2.12) as an approximation to (2.18) valid for very small ϵ, which however, play an important role in the proper interpretation of possible discontinuities in the solution of the equation (2.12).

2.1.3 Shock formation

The nonlinear hyperbolic conservation laws are known to produce discontinuities in the solution, even if the initial data is smooth. To understand the mechanism of shock wave formation, we consider a one-dimensional version of the IVP (2.13):

$$\begin{cases} \partial_t u + \partial_x f(u) = 0, & (x,t)\in\mathbb{R}\times\mathbb{R}_+, \\ u(x,0) = u_0(x), & x\in\mathbb{R}. \end{cases} \tag{2.19}$$

If $f(u)$ is a differentiable function of the solution u, then we can rewrite the first equation of the IVP (2.19) in the following form:

$$\partial_t u + f'(u)\partial_x u = 0. \tag{2.20}$$

We seek a curve $x = x(t)$ in the (x,t)-plane, such that $u(x(t),t) = const$. The latter requirement implies:

$$\frac{du(x(t),t)}{dt} = \frac{\partial u(x(t),t)}{\partial t} + \frac{\partial u(x(t),t)}{\partial x(t)}\frac{dx(t)}{dt} = 0. \tag{2.21}$$

Comparing equations (2.20) and (2.21) yields that equation (2.21) holds only if:

$$\frac{dx(t)}{dt} = f'(u(x(t),t)). \tag{2.22}$$

Taking into account that $x(0) = \xi$ and that $u(x(t),t) = u(x(0),0) = u_0(\xi) = const$ by definition, we obtain the following *Ordinary Differential Equation* (ODE):

$$\frac{dx(t)}{dt} = f'(u_0(\xi)), \quad x(0) = \xi \in \mathbb{R}, \quad t \in \mathbb{R}_+. \tag{2.23}$$

The solution of the ODE (2.23) is defined by:

$$x = \xi + f'(u_0(\xi))t \tag{2.24}$$

and represents a characteristic of the PDE under discussion. Note, that $\frac{x-\xi}{t} = f'(u_0(\xi)) = f'(u(x(t),t))$ is the speed of the initial data propagation, which never changes along the characteristic. In our case, this characteristic is a straight line in the (x,t)-plane along which the solution u of the IVP (2.19) is a constant. For any given pair $(x,t) \in \mathbb{R} \times \mathbb{R}_+$ we can solve the algebraic equation (2.24) for ξ and then:

$$u(x,t) = u_0(\xi). \tag{2.25}$$

If $f(u)$ is a linear function of u, such as in the case of linear transport equation, for which $f(u) = cu$, then $f'(u) = const$ for all u and all straight lines from the family of characteristics have the same slope and never cross each other. In this particular case, $u(x,t) = u_0(x - ct)$ - all points of the initial profile $u_0(x)$ are traveling in the space with the same speed c as shown in Figure 2.1.

Otherwise, if $f(u)$ is a nonlinear function of u, such as for Burgers' equation, for which $f(u) = \dfrac{u^2}{2}$, then $f'(u) \neq const$ for all u and some characteristics of the PDE with different slopes may cross over, originating the formation of infinite gradients of the function u (shock waves) and eventually yielding a solution analogous to a breaking wave, see Figure 2.2.

Before the characteristics of a nonlinear hyperbolic PDE first cross, we can still use (2.24) along with (2.25) to find the classical solution u of the IVP (2.19).

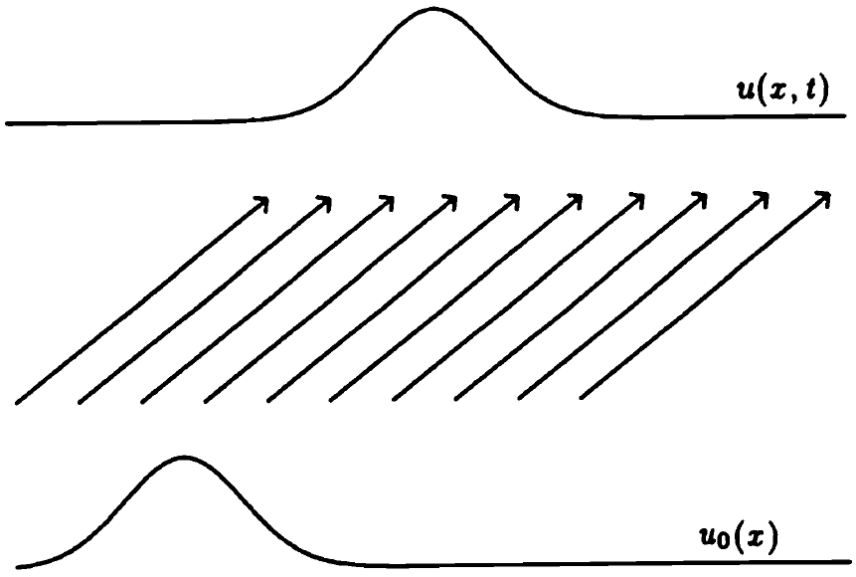

Fig. 2.1. Characteristics and solution for linear transport equation with $c > 0$.

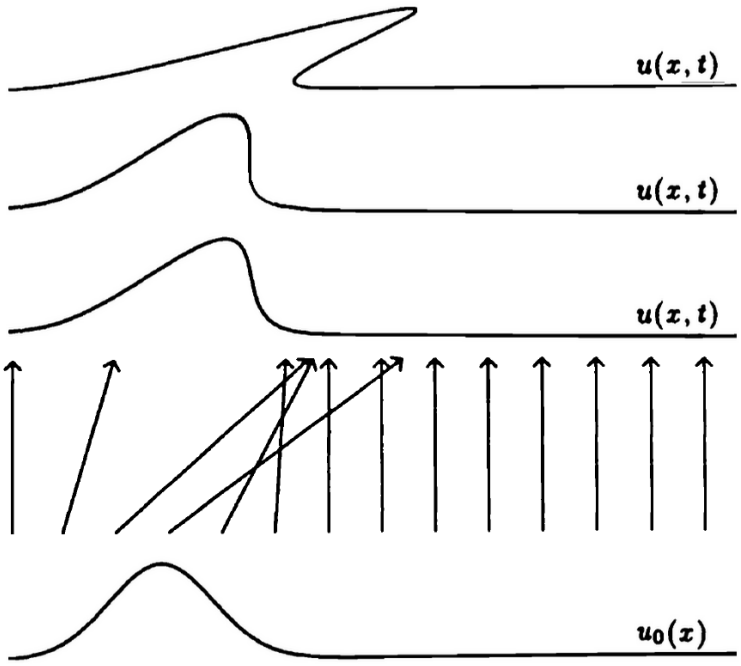

Fig. 2.2. Characteristics and solution for Burgers' equation.

Let $T_{shock} \in \mathbb{R}_+$ be the time when characteristics first cross. Let us consider two different characteristics whose origins are at vanishingly small distance $\Delta\xi$ from each other:

$$x_1 = \xi + f'(u_0(\xi))t, \tag{2.26}$$

$$x_2 = \xi + \Delta\xi + f'(u_0(\xi + \Delta\xi))t. \tag{2.27}$$

If the two cross over at time $t = T \in \mathbb{R}_+$, then $x_1 = x_2$:

$$\xi + f'(u_0(\xi))T = \xi + \Delta\xi + f'(u_0(\xi + \Delta\xi))T, \tag{2.28}$$

or

$$\frac{f'(u_0(\xi + \Delta\xi)) - f'(u_0(\xi))}{\Delta\xi} T = -1, \tag{2.29}$$

or

$$\frac{f'(u_0(\xi+\Delta\xi))-f'(u_0(\xi))}{u_0(\xi+\Delta\xi)-u_0(\xi)}\frac{u_0(\xi+\Delta\xi)-u_0(\xi)}{\Delta\xi}T=-1. \tag{2.30}$$

As $\Delta\xi \to 0$:

$$f''(u_0(\xi))u_0'(\xi)T=-1 \tag{2.31}$$

and

$$T=\frac{-1}{f''(u_0(\xi))u_0'(\xi)}. \tag{2.32}$$

Finally:

$$T_{shock}=\frac{-1}{\min\limits_{\xi\in\mathbb{R}}\left[f''(u_0(\xi))u_0'(\xi)\right]}. \tag{2.33}$$

To exclude the trivial cases, we assume that there exist u such that $f''(u)\neq 0$, and that $\forall\xi_1\in\mathbb{R},\exists\xi_2\in\mathbb{R}$ such that $u_0(\xi_1)\neq u_0(\xi_2)$. In other words, we do not consider the cases when $f(u)$ is a linear function of u or when the initial data $u_0(\xi)$ is a constant function or both. Otherwise $T_{shock}=\infty$, which means that a shock wave will never form. If $f(u)$ is a convex function of u, such that $f''(u)>0$ for all u, then to form a shock wave, there must be points ξ where the slope of the initial data $u_0(\xi)$ is negative. And vice versa - if $f(u)$ is a concave function of u, such that $f''(u)<0$ for all u, then the initial data $u_0(\xi)$ must have points ξ where its slope is positive. By minimization of the denominator in (2.32) we obtain the origin ξ_{shock} of the characteristic which first crosses, and then the position x_{shock} of this intersection:

$$x_{shock}=\xi_{shock}+f'(u_0(\xi_{shock}))T_{shock}. \tag{2.34}$$

The bottom line of this section is that one can use (2.24) along with (2.25) to determine the classical solution u of the IVP (2.19) only for time $t<T_{shock}$.

If $T_{shock} = \infty$, the classical solution is defined everywhere in the (x, t)-plane. If $T_{shock} = 0$, the classical solution does not exist at all.

For time $t \geq T_{shock}$ the algebraic equation (2.24) may not have a unique solution. It necessitates the multi-valued solutions to the IVP (2.19). These sorts of solutions probably make sense in some contexts. However, in most practical situations it is not physically reasonable. For example, the density of a gas cannot be multi-valued at any spatial location. Therefore, we need to somehow define a nonclassical or generalized solution of the equation or problem that represents physical reality.

2.1.4 Vanishing viscosity solution

One way to introduce a generalized solution of the hyperbolic scalar conservation law (2.12) is to replace it with the equation of parabolic type (2.18) where $\epsilon > 0$ is a small constant parameter (diffusion or viscosity coefficient):

$$\partial_t u^\epsilon + \nabla \cdot \mathbf{f}(u^\epsilon) = \epsilon \Delta u^\epsilon. \tag{2.35}$$

This equation has a smooth unique solution u^ϵ for all time $t > 0$, for any set of initial data, provided that $\epsilon > 0$. If ϵ is extremely small, then we might expect the solution u^ϵ of the parabolic equation (2.35) to be very close to the solution u of the hyperbolic equation (2.12). Away from shock waves, Δu^ϵ is bounded and the new term $\epsilon \Delta u^\epsilon$ is negligible. If a shock wave starts to form, the gradients of the solution u^ϵ of (2.35) begin to blow up and the $\epsilon \Delta u^\epsilon$ term becomes important. Figures 2.3 and 2.4 show an example where the same piecewise linear initial data forms the solutions u and u^ϵ to the one-dimensional inviscid and viscous Burgers' equations, respectively.

Definition 2.1.1. *A generalized solution u of the hyperbolic scalar conservation law* (2.12) *is called admissible in vanishing viscosity sense or just a vanishing viscosity solution if there is a sequence of smooth unique solutions u^ϵ of the parabolic equation* (2.35) *which converges to u as $\epsilon \to 0$.*

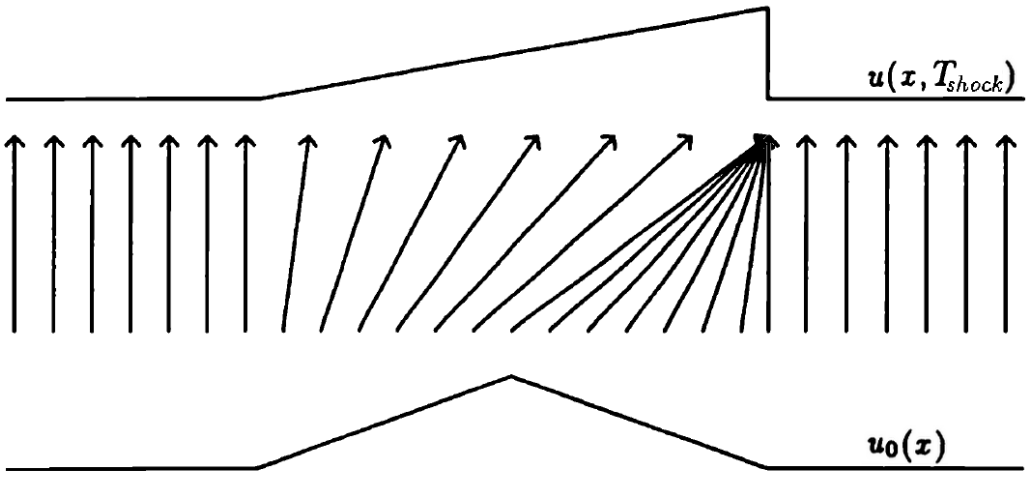

Fig. 2.3. Shock formation in inviscid Burgers' equation.

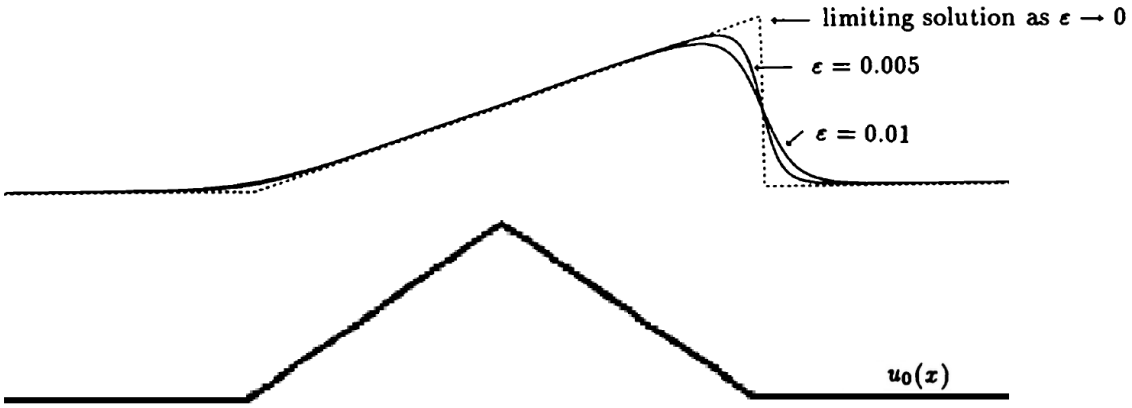

Fig. 2.4. Solution to viscous Burgers' equation for two different values of ϵ.

2.1.5 Weak solution

Another way to define a generalized solution of the inviscid equation (2.12) is to go back to the more fundamental integral form (2.11), which does not require the differentiability of the integrands, and say that u is a generalized solution of the hyperbolic scalar conservation law (2.12) if the associated integral form (2.11) is satisfied for all $\delta\Omega \subset \mathbb{R}^d$ and for all $(t_1 < t_2) \in \mathbb{R}_+$.

There is another approach that results in a different integral formulation that is often more convenient to work with. The basic idea is to take the PDE, multiply by a smooth test function, integrate one or more times over some domain, and then use integration by parts to move derivatives off the function u and onto the smooth test function. The result is an integral equation involving fewer derivatives on u, and hence requiring less smoothness.

In our case, we use test functions $\phi \in C_0^1(\mathbb{R}^d \times \mathbb{R}_+)$. Here C_0^1 is the space of functions that are continuously differentiable with compact support. The latter requirement means that ϕ is identically zero outside of some bounded domain. In particular, the test functions ϕ vanish at infinity, i.e. $\phi\,|_{\mathbf{x}\in\partial\mathbb{R}^d} = 0$ as well as $\phi\,|_{t=+\infty} = 0$.

If we multiply the equation (2.12) by a test function ϕ and then integrate over space and time, we obtain:

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}^d} \left[\phi\partial_t u + \phi\nabla\cdot\mathbf{f}(u)\right] d\mathbf{x}dt = 0. \tag{2.36}$$

Integration by parts yields:

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}^d} \left[u\partial_t\phi + \mathbf{f}(u)\cdot\nabla\phi\right] d\mathbf{x}dt + \int_{\mathbb{R}^d} \phi(\mathbf{x},0)u_0(\mathbf{x})d\mathbf{x} = 0. \tag{2.37}$$

Note that nearly all the boundary terms that normally arise through integration by parts drop out. The remaining boundary term brings in the initial condition of the PDE.

Definition 2.1.2. *A generalized solution u of the hyperbolic scalar conservation law* (2.12) *is called a weak solution if the weak integral formulation* (2.37) *is satisfied for all test functions $\phi \in C_0^1(\mathbb{R}^d \times \mathbb{R}_+)$.*

We should expect a direct connection between the fundamental integral form (2.11) and the weak integral formulation (2.37). This can be achieved by considering special test functions ϕ with the property that:

$$\phi = \begin{cases} 1 & \text{if } (\mathbf{x}, t) \in \delta\Omega \times [t_1, t_2], \\ 0 & \text{if } (\mathbf{x}, t) \notin \delta\Omega \bigcup \delta\Delta \times [t_1 - \Delta, t_2 + \Delta]. \end{cases} \tag{2.38}$$

and with ϕ smooth in the intermediate strip of width Δ. It is not difficult to demonstrate that the weak integral formulation (2.37) becomes the integral form (2.11) as $\Delta \to 0$. Therefore, any weak solution of the hyperbolic scalar conservation law (2.12) also satisfies its original integral form.

2.1.6 One-dimensional Riemann problem as an example of nonuniqueness of a weak solution

Unfortunately, weak solutions as well as solutions to the fundamental integral form (2.11) are often not unique. We will demonstrate this fact by considering a one-dimensional Riemann problem for the hyperbolic scalar conservation law (2.12) with a convex inviscid flux $f(u)$:

$$\begin{cases} \partial_t u + \partial_x f(u) = 0, & (x, t) \in \mathbb{R} \times \mathbb{R}_+, \\ u(x, 0) = u_0(x), & x \in \mathbb{R}. \end{cases} \tag{2.39}$$

where $u_0(x)$ is piecewise constant initial data with a single discontinuity:

$$u_0(x) = \begin{cases} u_L & \text{if } x < 0, \\ u_R & \text{if } x > 0. \end{cases} \tag{2.40}$$

Obviously, there are no classical solutions to the IVP (2.39). If $u_L > u_R$, then at any given time t, one will obtain a multi-valued "solution" for any spatial location $f'(u_R)t < x < f'(u_L)t$. If $u_L < u_R$, then at any given time t, the "solution" is not defined at all for all $f'(u_L)t < x < f'(u_R)t$. The form of a nonclassical or generalized solution to the IVP (2.39) depends on the relation between u_L and u_R.

Case I : $\mathbf{u_L > u_R}$:

In this case, the generalized solution is represented by a travelling shock wave as shown in Figure 2.5 and has the following form:

$$u = \begin{cases} u_L & \text{if } x < st, \\ u_R & \text{if } x > st. \end{cases} \tag{2.41}$$

where s is the shock speed. Generally, s is a function of time t, but in the case of the Riemann problem, we will show that s only depends on the left and right states.

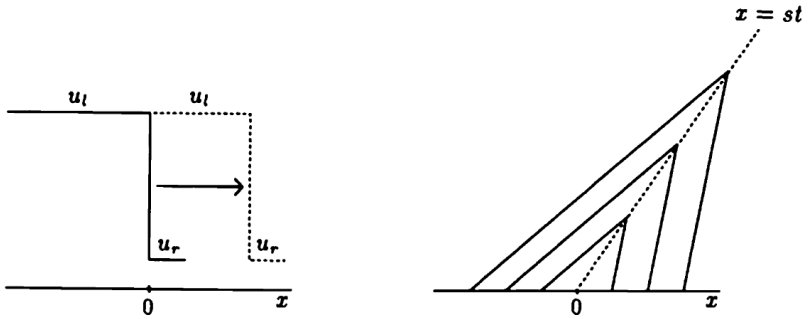

Fig. 2.5. Shock wave.

To find the shock speed s, we consider a one-dimensional version of the integral form (2.11):

$$\begin{aligned}&\forall(x_1 < x_2) \in \mathbb{R} \quad \text{and} \quad \forall(t_1 < t_2) \in \mathbb{R}_+ : \\ &\int_{x_1}^{x_2} u(x,t_2)dx - \int_{x_1}^{x_2} u(x,t_1)dx + \int_{t_1}^{t_2} f(u(x_2,t))dt - \\ &\int_{t_1}^{t_2} f(u(x_1,t))dt = 0,\end{aligned} \tag{2.42}$$

where x_1 is the shock location at time t_1 and x_2 is the shock location at time t_2, see Figure 2.6. In accordance with this figure, equation (2.42) can be modified as follows:

$$u_L\Delta x - u_R\Delta x + f(u_R)\Delta t - f(u_L)\Delta t = 0 \tag{2.43}$$

or

$$s = \frac{\Delta x}{\Delta t} = \frac{f(u_L) - f(u_R)}{u_L - u_R} = \frac{[\![f]\!]}{[\![u]\!]}. \tag{2.44}$$

The relation (2.44) between the shock speed s and the states u_L and u_R is called the Rankine-Hugoniot jump condition. The generalized solution (2.41) along with the jump condition (2.44) is unique. It is both the vanishing viscosity and weak solution, and obviously satisfies the fundamental integral form (2.42).

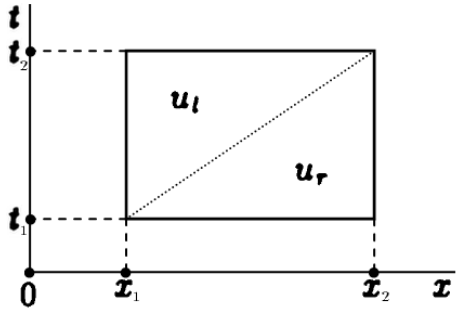

Fig. 2.6. Shock path inside a rectangular region in (x,t)-plane.

Case II : $\mathbf{u_L < u_R}$:

Such relation between two pieces of the initial data necessitates the existence of infinitely many weak solutions, which also satisfy the integral equation (2.42). One of these is again (2.41) together with (2.44) in which the discontinuity propagates with speed s as shown in Figure 2.7. Note that characteristics now go out of the shock. However, this weak solution is not stable to perturbations - if the initial data is smeared out slightly, or if a small amount of viscosity is added to the equation, the weak solution changes completely. Another example of physically incorrect weak solutions is any sort of piecewise linear connection between the left and right states, which allows a finite number of discontinuities. The physically relevant generalized solution is represented in Figure 2.8 by a weak solution in the form of the rarefaction wave:

$$u = \begin{cases} u_L & \text{if } x < f'(u_L)t, \\ v(x,t) & \text{if } f'(u_L)t \le x \le f'(u_R)t, \\ u_R & \text{if } x > f'(u_R)t. \end{cases} \tag{2.45}$$

where $v(x,t)$ is the solution to $f'(v(x,t)) = \dfrac{x}{t}$. This weak solution is stable to perturbations of both the initial data and the equation itself and is in fact the vanishing viscosity solution.

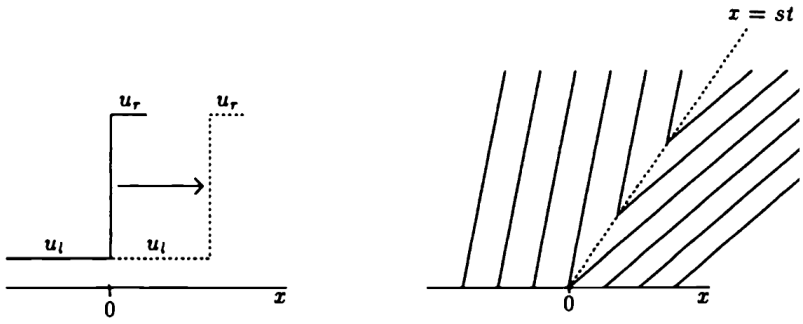

Fig. 2.7. Spurious or expansion shock.

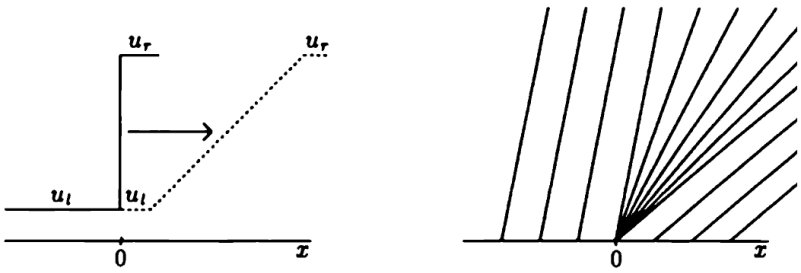

Fig. 2.8. Rarefaction wave.

2.1.7 Entropy conditions

We see that there might be many weak solutions to the hyperbolic scalar conservation law (2.12). The existence of these spurious solutions is not merely a mathematical curiosity. Under some circumstances nonphysical solutions of this type are all too easily computed numerically, in spite of the fact that numerical methods typically contain some numerical viscosity.

In order to effectively use the hyperbolic equations we should impose some additional condition along with the differential equation in order to insure that the problem has a unique weak solution that is physically correct. Often the condition we want to impose is simply that the weak solution must be the vanishing viscosity solution to the proper viscous equation. However, this condition is hard to work with directly in the context of hyperbolic equations. Instead, other conditions have been developed and can be applied directly to weak solutions of hyperbolic equations to check if they are physically admissible. Such conditions are called entropy conditions. This name comes from gas dynamics, where the second law of thermodynamics demands that the entropy of a system must be nondecreasing with time. Across a physically admissible shock the entropy of the gas increases. Across a spurious or expansion shock, however, the entropy of the gas would decrease, which is not allowed. The entropy of a perfect gas can be computed as a simple function of

the pressure and density, and the behavior of this function can be used to test a weak solution for admissibility. For hyperbolic scalar equations it is possible to define a function $\eta(u)$ called an entropy function, which has similar diagnostic properties.

Similar to gas dynamics, we find that if the solution u of the hyperbolic scalar conservation law (2.12) is smooth everywhere in $\mathbb{R}^d \times \mathbb{R}_+$, then the function $\eta(u)$ satisfies an additional differential equation:

$$\partial_t \eta(u) + \nabla \cdot \boldsymbol{\psi}(u) = 0 \tag{2.46}$$

for some entropy flux $\boldsymbol{\psi}(u)$. Moreover, if $\eta(u)$ and $\boldsymbol{\psi}(u)$ are differentiable functions of the solution u, then we can rewrite (2.46) as:

$$\eta'(u)\partial_t u + \boldsymbol{\psi}'(u) \cdot \nabla u = 0. \tag{2.47}$$

The hyperbolic equation (2.12) can be written as:

$$\partial_t u + \mathbf{f}'(u) \cdot \nabla u = 0. \tag{2.48}$$

Then multiplication of (2.48) by $\eta'(u)$ and comparison with (2.47) yields:

$$\boldsymbol{\psi}'(u) = \eta'(u)\mathbf{f}'(u). \tag{2.49}$$

The equation (2.49) admits many solutions $\eta(u)$ and $\boldsymbol{\psi}(u)$. For reasons that will be seen below, we assume that the entropy function is convex, i.e. $\eta''(u) > 0$ for all u.

If the solution u contains a discontinuity, then the manipulations performed above are not valid everywhere in $\mathbb{R}^d \times \mathbb{R}_+$ space. Since we are interested in how the entropy function $\eta(u)$ behaves for the vanishing viscosity solution, we look at the related

viscous or parabolic equation (2.35) with solution u^ϵ, and then let the viscosity go to zero. Multiplication of (2.35) by $\eta'(u^\epsilon)$ yields:

$$\eta'(u^\epsilon)\partial_t u^\epsilon + \eta'(u^\epsilon)\nabla\cdot\mathbf{f}(u^\epsilon) = \eta'(u^\epsilon)\epsilon\Delta u^\epsilon \tag{2.50}$$

or

$$\partial_t\eta(u^\epsilon) + \nabla\cdot\boldsymbol{\psi}(u^\epsilon) = \epsilon\nabla\cdot(\eta'(u^\epsilon)\nabla u^\epsilon) - \epsilon\eta''(u^\epsilon)(\nabla u^\epsilon)^2. \tag{2.51}$$

Integrating the last equation over an arbitrarily chosen open connected domain $\delta\Omega \in \mathbb{R}^d$ with the boundary $\delta\Gamma$ as well as in time from any t_1 to any $t_2 > t_1$ gives:

$$\begin{aligned}\int_{t_1}^{t_2}\int_{\delta\Omega}\left[\partial_t\eta(u^\epsilon) + \nabla\cdot\boldsymbol{\psi}(u^\epsilon)\right]d\mathbf{x}dt = \quad & \epsilon\int_{t_1}^{t_2}\int_{\delta\Omega}\nabla\cdot(\eta'(u^\epsilon)\nabla u^\epsilon)d\mathbf{x}dt - \\ & \epsilon\int_{t_1}^{t_2}\int_{\delta\Omega}\eta''(u^\epsilon)(\nabla u^\epsilon)^2 d\mathbf{x}dt\end{aligned} \tag{2.52}$$

or

$$\begin{aligned}\int_{t_1}^{t_2}\int_{\delta\Omega}\left[\partial_t\eta(u^\epsilon) + \nabla\cdot\boldsymbol{\psi}(u^\epsilon)\right]d\mathbf{x}dt = \quad & \epsilon\int_{t_1}^{t_2}\oint_{\delta\Gamma}\eta'(u^\epsilon)\nabla u^\epsilon\cdot d\boldsymbol{\sigma}dt - \\ & \epsilon\int_{t_1}^{t_2}\int_{\delta\Omega}\eta''(u^\epsilon)(\nabla u^\epsilon)^2 d\mathbf{x}dt.\end{aligned} \tag{2.53}$$

As $\epsilon \to 0$, the first term on the right hand side of the last equation vanishes. This is clearly true if u is at least continuous at the boundary $\delta\Gamma$ for all $t \in [t_1, t_2]$. If u is discontinuous at any spatial point $\mathbf{x} \in \delta\Gamma$ and at any time $t \in [t_1, t_2]$, then the first term vanishes anyway because the double integral is bounded. As $\epsilon \to 0$, the second term vanishes only if u is at least continuous throughout the domain $\delta\Omega$ for all $t \in [t_1, t_2]$. If there exist $(\mathbf{x}, t) \in \delta\Omega \times [t_1, t_2]$ such that $u(\mathbf{x}, t)$ is discontinuous, then

the second term does not vanish but, since $\eta(u)$ is convex, approaches a non-negative limit. Therefore, as $\epsilon \to 0$:

$$\int_{t_1}^{t_2} \int_{\delta\Omega} \left[\partial_t \eta(u) + \nabla \cdot \boldsymbol{\psi}(u)\right] d\mathbf{x}dt \leq 0. \tag{2.54}$$

Since the domain $\delta\Omega \subset \mathbb{R}^d$ as well as times $(t_1 < t_2) \in \mathbb{R}_+$ are arbitrarily chosen, we conclude that, at least formally, or more precisely - in vanishing viscosity sense, that:

$$\partial_t \eta(u) + \nabla \cdot \boldsymbol{\psi}(u) \leq 0 \tag{2.55}$$

for all $\mathbf{x} \in \mathbb{R}^d$ and for all $t \in \mathbb{R}_+$. We will later use the inequality (2.55) as a key ingredient for development of the numerical schemes for solving hyperbolic scalar conservation equations in Chapter IV. The weak integral formulation of the inequality (2.55) is the following: for all non-negative test functions $\phi \in C_0^1(\mathbb{R}^d \times \mathbb{R}_+)$ we find that:

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \left[\eta(u)\partial_t \phi + \boldsymbol{\psi}(u) \cdot \nabla \phi\right] d\mathbf{x}dt + \int_{\mathbb{R}^d} \phi(\mathbf{x}, 0)\eta(u_0(\mathbf{x}))d\mathbf{x} \geq 0. \tag{2.56}$$

Definition 2.1.3. *A weak solution u of the hyperbolic scalar conservation law* (2.12) *is called an entropy admissible solution or just an entropy solution if the inequality* (2.56) *is satisfied for all convex entropy functions $\eta(u)$ and corresponding entropy fluxes $\boldsymbol{\psi}(u)$.*

Example :

Consider a one-dimensional Riemann problem (2.39)-(2.40) for Burgers' equation for which $f(u) = \frac{u^2}{2}$ and $u_L > u_R$. The physically relevant generalized solution for this problem is a travelling shock wave with speed $s = \frac{u_L + u_R}{2}$. If we choose $\eta(u) = \frac{u^2}{2}$ (note that $\eta''(u) = 1 > 0$), then $\psi(u) = \frac{u^3}{3}$. If for a given $(x,t) \in \mathbb{R} \times \mathbb{R}_+$ the function u is smooth, then:

$$\begin{aligned} \partial_t \eta(u) + \partial_x \psi(u) = \quad & \partial_t \left(\frac{u^2}{2} \right) + \partial_x \left(\frac{u^3}{3} \right) = \\ & u \left(\partial_t u + \partial_x \left(\frac{u^2}{2} \right) \right) = 0. \end{aligned} \tag{2.57}$$

Otherwise, if u has a discontinuity at that point, then the differential operations are not valid anymore, and instead, we will integrate $\partial_t \eta(u) + \partial_x \psi(u)$ over an infinitesmal rectangle $[x_1, x_2] \times [t_1, t_2]$ which includes this space-time point:

$$\begin{aligned} & \int_{t_1}^{t_2} \int_{x_1}^{x_2} \left[\partial_t \eta(u) + \partial_x \psi(u) \right] dx dt = \int_{x_1}^{x_2} \eta(u(x,t_2)) dx - \int_{x_1}^{x_2} \eta(u(x,t_1)) dx + \\ & \int_{t_1}^{t_2} \psi(u(x_2,t)) dt - \int_{t_1}^{t_2} \psi(u(x_1,t)) dt = \\ & \eta(u_L)\Delta x - \eta(u_R)\Delta x + \psi(u_R)\Delta t - \psi(u_L)\Delta t = \\ & \Delta t \left(s(\eta(u_L) - \eta(u_R)) + \psi(u_R) - \psi(u_L) \right) = \frac{\Delta t}{12} (u_R - u_L)^3 < 0. \end{aligned} \tag{2.58}$$

Hence, the solution is in fact the entropy solution. Besides that, the relation (2.58) clearly indicates that the propagation of a rarefaction wave in the form of a shock wave is not allowed.

2.2 Foundations of fluid dynamics

2.2.1 Compressible Navier-Stokes equations

A general description of compressible flows is given by the system of compressible Navier-Stokes equations, which play a fundamental role in understanding and simulating compressible gases and fluids and are widely used in science and industry.

The compressible Navier-Stokes equations are derived under the assumption of the continuum hypothesis and express conservation of mass, momentum, and total energy for a fluid enclosed in an arbitrarily chosen open connected domain $\delta\Omega \in \mathbb{R}^d$ with the boundary $\delta\Gamma$ over a finite time interval.

The conservation of mass states that the time rate of change of mass in $\delta\Omega$ must be equal to the net mass flowing into $\delta\Omega$ through its boundary $\delta\Gamma$. The conservation of momentum comes directly from Newton's second law, which states that the time rate of change of momentum in $\delta\Omega$ must be equal to the sum of all forces applied to the fluid in $\delta\Omega$. And finally the conservation of total energy states that total energy of an isolated system remains constant.

We seek the density ρ, momentum $\mathbf{m} = \rho\mathbf{u}$, where $\mathbf{u}$ is the velocity, and total energy $E = \dfrac{\mathbf{m}\cdot\mathbf{m}}{2\rho} + \rho e$, where e is the internal energy per unit mass, as functions of $(\mathbf{x}, t) \in \mathbb{R}^d \times \mathbb{R}_+$. The compressible Navier-Stokes equations for $\mathbf{U} = (\rho, \mathbf{m}, E)^{\mathrm{T}}$ read:

$$\begin{cases} \partial_t \rho + \nabla\cdot\mathbf{m} = 0, \\ \partial_t \mathbf{m} + \nabla\cdot\left(\dfrac{1}{\rho}\mathbf{m}\otimes\mathbf{m} + p\mathbb{I}\right) = \nabla\cdot\left(2\mu\nabla_s\mathbf{u}\right) + \nabla\cdot\left(\lambda\nabla\cdot\mathbf{u}\mathbb{I}\right) + \mathbf{g}, \\ \partial_t E + \nabla\cdot\left(\dfrac{\mathbf{m}}{\rho}(E+p)\right) = \nabla\cdot\left(2\mu\nabla_s\mathbf{u}\mathbf{u}\right) + \nabla\cdot\left(\lambda\nabla\cdot\mathbf{u}\mathbf{u}\right) + \nabla\cdot\left(\kappa\nabla T\right), \\ \mathbf{U}(\mathbf{x}, 0) = \mathbf{U}_0(\mathbf{x}). \end{cases} \tag{2.59}$$

where p is the pressure, T is the temperature, μ is the dynamic viscosity coefficient, λ is the bulk viscosity coefficient, κ is the thermal conduction coefficient, $\mathbf{g}$ is a given

external volume force, and $\nabla_s \mathbf{u} = \frac{1}{2}(\nabla \mathbf{u} + \nabla^{\mathrm{T}} \mathbf{u})$ is the strain rate tensor. The initial data $\mathbf{U}_0(\mathbf{x})$ is supposed to be known. If the equations have to be solved on some bounded domain, then we also need to include the appropriate types of boundary conditions, such as inflow or outflow boundaries, slip or no-slip boundaries, and others. These boundary conditions will be discussed later in Chapter V.

2.2.2 Viscosity and thermal conduction

The basic physical phenomena of viscosity and thermal conduction in a fluid are due to the transport of momentum and energy via random molecular motion. Each molecule in a fluid has momentum and energy, which it carries with it when it moves from one location to another in space before colliding with another molecule. The transport of molecular momentum gives rise to the macroscopic effect we call viscosity, and the transport of molecular energy gives rise to the macroscopic effect we call thermal conduction. This is why viscosity and thermal conduction are labeled as transport phenomena. A study of these transport phenomena at the molecular level is part of kinetic theory, which is beyond the scope of this work. Instead, we simply state the macroscopic results of such molecular motion.

In general, both the dynamic and bulk viscosity coefficients depend on the thermodynamic state and properties of the fluid. For most normal situations, μ is a constant or is a function of temperature only. A conventional relation for the temperature variation of μ for air is given by Sutherland's law $\mu \sim \sqrt{T}$ and the pressure dependence is neglected.

To this day, the correct expression for the bulk viscosity coefficient λ is still somewhat controversial. In 1845, George Stokes hypothesized that $\lambda = -\frac{2}{3}\mu$. The kinetic theory found that λ must be nonzero if the kinetic energy of fluid molecules can be transferred to internal degrees of freedom. The value of λ depends on the characteristic time of the energy transfer and vanishes for monatomic fluids.

The thermal conduction coefficient κ also depends primarily on the temperature, because the results of elementary kinetic theory show that $\kappa \sim \mu$. All these simple relationships are only approximate and do not hold at high temperatures. For any detailed viscous flow calculation, one should consult the published literature for more precise values of these coefficients.

2.2.3 Compressible Euler equations

For inviscid flows, both the dynamic and bulk viscosity coefficients as well as the thermal conduction coefficient are zero, which corresponds to the compressible Euler equations:

$$\begin{cases} \partial_t \rho + \nabla \cdot \mathbf{m} = 0, \\ \partial_t \mathbf{m} + \nabla \cdot \left(\dfrac{1}{\rho} \mathbf{m} \otimes \mathbf{m} + p \mathbb{I} \right) = \mathbf{g}, \\ \partial_t E + \nabla \cdot \left(\dfrac{\mathbf{m}}{\rho} (E + p) \right) = 0, \\ \mathbf{U}(\mathbf{x}, 0) = \mathbf{U}_0(\mathbf{x}). \end{cases} \tag{2.60}$$

Such flows often exhibit a complex behavior of the solution, and may form different types of localized structures including shock waves, contact discontinuities, and rarefaction waves. Capturing these structures numerically is a highly nontrivial task.

The number of equations in the systems (2.59) and (2.60) is $d+2$. However, the number of unknown functions including the pressure is $d+3$. It means that we need to close the systems with an appropriate equation, which is called the *Equation of State* (EoS) and links the pressure with other unknowns. The immense broadening of theoretical understanding, experimental and measurement capabilities, and specialized technical applications have contributed to the large diversity of equations of state that are presently available to describe compressible fluids. In the present

work, we utilize the simplest possible model of a compressible fluid called a perfect gas.

2.2.4 Equilibrium thermodynamics of a perfect gas

A hypothetical gas composed of a set of randomly-moving point particles that exert no intermolecular forces is defined as a perfect gas. If at a given time $t \in \mathbb{R}_+$ the thermodynamic parameters of such a perfect gas do not vary over the entire volume occupied by the gas, then the gas is said to be in a state of thermodynamic equilibrium at time t. Equilibrium thermodynamics of perfect gases studies a wide variety of processes by which gases pass through a sequence of states of thermodynamic equilibrium. Hereinafter we assume the model of a perfect gas whose behavior is subject to the equilibrium thermodynamics approach. Though this approach is not perfect, it simplifies the analysis and study of the thermodynamic systems by introducing lumped parameters, and often agrees well with experimental data.

We first introduce the Equation of State which determines the connection between different thermodynamic parameters of a gas:

$$p = k\frac{NT}{V}, \tag{2.61}$$

where p is the pressure, N is the total number of particles, T is the temperature, V is the volume, and $k = 1.381 \cdot 10^{-23} JK^{-1}$ is the Boltzmann constant. Since the number of particles per unit volume is defined as $n = \dfrac{N}{V}$, the EoS (2.61) can be written in the form:

$$p = knT. \tag{2.62}$$

For a mix of perfect gases in a state of thermodynamic equilibrium the EoS (2.62) is modified as follows:

$$p = \sum_i p_i = \sum_i kn_i T, \tag{2.63}$$

where $p_i = kn_iT$ is the fractional pressure of the i-th gas. Given that $N = \frac{m}{M}N_A$, where m is the total mass, M is the molar mass, and $N_A = 6.022 \cdot 10^{23} mol^{-1}$ is Avogadro's number, that the specific gas constant is given by $R = \frac{kN_A}{M}$, and that the density is given by $\rho = \frac{m}{V}$, we define another useful form of the EoS for a perfect gas:

$$p = R\rho T. \tag{2.64}$$

Equation (2.64) is then used to construct the closure to the systems (2.59) and (2.60). To do that, we first recall that the elementary kinetic theory of perfect monatomic gases gives the following relation for the pressure:

$$p = \frac{2}{3}n\bar{\epsilon}, \tag{2.65}$$

where $\bar{\epsilon}$ is the average kinetic energy per particle. Comparing the equations (2.62) and (2.65) yields:

$$\bar{\epsilon} = \frac{3}{2}kT. \tag{2.66}$$

Then we use some physical intuition to recognize that since the equation (2.66) is derived for monatomic gases for which the number of translational degrees of freedom per particle in three space dimensions is exactly 3, then $\frac{1}{2}kT$ is associated with each of these translational degrees of freedom. The logical extension to polyatomic gases and the equipartition theorem yield:

$$\bar{\epsilon} = \frac{i}{2}kT, \tag{2.67}$$

where $i = i_{\mathbf{translational}} + i_{\mathbf{rotational}} + 2i_{\mathbf{vibrational}}$ is the total number of degrees of freedom per polyatomic particle. Since there are no intermolecular forces in a perfect gas, then the total internal energy of such a gas is defined as:

$$U = \frac{m}{M}N_A\bar{\epsilon} = m\frac{i}{2}RT. \tag{2.68}$$

And the internal energy per unit mass:

$$e = \frac{U}{m} = \frac{i}{2}RT. \tag{2.69}$$

Combining the equations (2.64) and (2.69) yields:

$$p = \frac{2}{i}\rho e. \tag{2.70}$$

To define the microscopic property i in terms of macroscopic characteristics, we will utilize the first law of equilibrium thermodynamics:

$$\delta Q = dU + pdV \tag{2.71}$$

along with the definitions of heat capacity at constant volume:

$$C_V = \left(\frac{\delta Q}{dT}\right)_V = \left(\frac{dU + pdV}{dT}\right)_V = \frac{i}{2}mR \tag{2.72}$$

and heat capacity at constant pressure:

$$C_p = \left(\frac{\delta Q}{dT}\right)_p = \left(\frac{dU + pdV}{dT}\right)_p = \frac{i+2}{2}mR. \tag{2.73}$$

Then introducing the adiabatic gas constant γ as:

$$\gamma = \frac{C_p}{C_V} = \frac{i+2}{i}, \tag{2.74}$$

we will find that:

$$i = \frac{2}{\gamma - 1}, \tag{2.75}$$

and finally one more form of the EoS for a perfect gas reads:

$$p = (\gamma - 1)\rho e. \tag{2.76}$$

The final form of the closure is obtained by expressing equation (2.76) in terms of conservative variables relating to the systems of equations (2.59) and (2.60):

$$p = (\gamma - 1)\left(E - \frac{\mathbf{m}\cdot\mathbf{m}}{2\rho}\right). \tag{2.77}$$

We study the compressible Euler equations, i.e. the equations with zero viscosity and thermal conduction terms. It means that any fluid particle participating in the global flow modeled by these equations is adiabatically isolated ($\delta Q = 0$) from its surroundings. Therefore, the fluid particles of a perfect gas whose motion is studied can only undergo equilibrium compression or expansion with no heat transfer between them. Then, according to the first law of equilibrium thermodynamics:

$$\delta Q = dU + pdV = 0. \tag{2.78}$$

Using the equations (2.64) and (2.68) in (2.78) yields:

$$d\left(m\frac{i}{2}RT\right) + pd\left(mR\frac{T}{p}\right) = 0 \tag{2.79}$$

or

$$\frac{i+2}{2}\frac{dT}{T} = \frac{dp}{p}. \tag{2.80}$$

Taking into account equation (2.75) and that $\frac{da}{a} = d\ln a$ we have the following relation:

$$d\ln T^{\frac{\gamma}{\gamma-1}} = d\ln p \tag{2.81}$$

or

$$\int d\ln T^{\frac{\gamma}{\gamma-1}} = \int d\ln p + C, \tag{2.82}$$

where C is an arbitrary constant. Finally:

$$pT^{\frac{\gamma}{1-\gamma}} = C. \tag{2.83}$$

It is not difficult to see, that:

$$pV^{\gamma} = C \quad \text{or} \quad \frac{p}{\rho^{\gamma}} = C \tag{2.84}$$

as well as

$$TV^{\gamma-1} = C \quad \text{or} \quad \frac{T}{\rho^{\gamma-1}} = C. \tag{2.85}$$

Equations (2.83)-(2.85) reflect the relationship between thermodynamic parameters of a perfect gas undergoing adiabatic compression or expansion.

2.2.5 Entropy

If we take the first law of equilibrium thermodynamics and divide it by the equilibrium temperature T of the system, we obtain:

$$\frac{\delta Q}{T} = \frac{dU}{T} + \frac{pdV}{T}. \tag{2.86}$$

Using the equations (2.64) and (2.68) in (2.86) yields:

$$\frac{\delta Q}{T} = \frac{1}{T} d\left(m\frac{i}{2}RT\right) + \frac{p}{T} d\left(mR\frac{T}{p}\right). \tag{2.87}$$

From equation (2.64) we can also define $T = \dfrac{p}{R\rho}$, which modifies (2.87) as:

$$\frac{\delta Q}{T} = m\frac{i}{2}R\frac{\rho}{p} d\left(\frac{p}{\rho}\right) + mR\rho d\left(\frac{1}{\rho}\right). \tag{2.88}$$

Further modifications of the equation (2.88) along with $i = \dfrac{2}{\gamma - 1}$ lead to the following differential form:

$$\frac{\delta Q}{T} = d\left(\frac{mR}{\gamma - 1} \ln \frac{p}{\rho^{\gamma}}\right). \tag{2.89}$$

The right hand side of this equation is a full differential. Hence, the left hand side of the same equation must be a full differential of some function S, which is called the entropy:

$$\frac{\delta Q}{T} = dS \tag{2.90}$$

with

$$S = \frac{mR}{\gamma - 1} \ln \frac{p}{\rho^{\gamma}}. \tag{2.91}$$

If we consider a fluid particle of total mass m within a perfect gas, for which the above derivations hold, then it is immediately seen from the equations (2.84) and (2.91) that the entropy of the fluid particle undergoing adiabatic changes of its volume remains constant and therefore $\dfrac{dS}{dt} = 0$. This result can be easily explained (at least qualitatively) in the framework of statistical physics, where the entropy of an equilibrium system is proportional to the logarithm of the number of possible microscopic configurations of the individual molecules of the system (microstates) which could give rise to the observed macroscopic state (macrostate) of the system. The total number of microstates $\hat{\Gamma}$ can be represented for example as:

$$\hat{\Gamma} = \hat{\Gamma}_{\mathbb{R}^d} \hat{\Gamma}_{\mathbb{M}^d}, \tag{2.92}$$

where $\hat{\Gamma}_{\mathbb{R}^d}$ is the number of possible locations of molecules in the Euclidean space $\mathbb{R}^d$ and $\hat{\Gamma}_{\mathbb{M}^d}$ is the number of possible locations of molecules in the space of momentum $\mathbb{M}^d$. For adiabatic compression, for instance, the volume of the system decreases and $\hat{\Gamma}_{\mathbb{R}^d}$ goes down. At the same time, the temperature of the system increases and $\hat{\Gamma}_{\mathbb{M}^d}$ goes up. As a result, these two trends cancel out and $\hat{\Gamma} = const$.

As we have mentioned above, the full time rate of change of the entropy S of an isolated fluid particle is zero, i.e. $\dfrac{dS}{dt} = 0$. On the other hand:

$$\frac{dS}{dt} = \frac{\partial S}{\partial t} + \sum_{i=1}^{n} \frac{\partial S}{\partial x_i}\frac{\partial x_i}{\partial t} = \frac{\partial S}{\partial t} + \mathbf{u}\cdot\nabla S. \tag{2.93}$$

Therefore:

$$\frac{\partial S}{\partial t} + \mathbf{u}\cdot\nabla S = 0. \tag{2.94}$$

Since we do not travel with the fluid particles in the framework of Eulerian gas dynamics, we reexpress equation (2.94) in accordance with the continuum fluid assumption. To achieve that, we first note that $S = sV$, where s is the entropy of a fluid particle per unit volume. Then:

$$\frac{\partial (sV)}{\partial t} + \mathbf{u}\cdot\nabla(sV) = 0 \tag{2.95}$$

or

$$\frac{\partial s}{\partial t} + \mathbf{u}\cdot\nabla s + s\frac{1}{V}\left(\frac{\partial V}{\partial t} + \mathbf{u}\cdot\nabla V\right) = 0. \tag{2.96}$$

Recognizing that:

$$\frac{1}{V}\left(\frac{\partial V}{\partial t} + \mathbf{u}\cdot\nabla V\right) = \frac{1}{V}\frac{dV}{dt} \tag{2.97}$$

and further that:

$$\frac{1}{V}\frac{dV}{dt} = \nabla\cdot\mathbf{u}, \tag{2.98}$$

we obtain the following equation:

$$\partial_t s + \nabla\cdot(\mathbf{u}s) = 0. \tag{2.99}$$

It has been shown, see for example [45], that the entropy S of a fluid particle can only increase or jump to a higher value when the particle passes a shock:

$$\frac{\partial S}{\partial t} + \mathbf{u} \cdot \nabla S > 0. \tag{2.100}$$

Hence, in this case:

$$\partial_t s + \nabla \cdot (\mathbf{u} s) > 0. \tag{2.101}$$

Eventually, we introduce the following inequality:

$$\partial_t s + \nabla \cdot (\mathbf{u} s) \geq 0, \tag{2.102}$$

where

$$s = \frac{\rho R}{\gamma - 1} \ln \frac{p}{\rho^\gamma}. \tag{2.103}$$

As it will be seen in Chapter V, that inequality plays the crucial role in further discussion of numerical schemes which are based on the entropy production in the interior of shock waves.

3. THE IDEA OF NUMERICAL STABILIZATION AND ARTIFICIAL VISCOSITIES

In this chapter we will introduce the so-called first-order artificial viscosity via regularization of hyperbolic equations. Then we derive the criteria of stability and monotonicity for certain primitive numerical schemes with subsequent extension to more general cases. At the end of this chapter we will also give the preliminary recipe for the entropy-based artificial viscosity construction but discuss it in further detail in Chapter IV.

3.1 Finite difference approximation

Finite difference methods use the nodal values of a numerical solution u_h to approximate both time and space derivatives of an equation. This technique can be represented by means of numerical stencils. By $u_h\left(x_i, t^n\right) \equiv u_{h,i}^n$ we denote a discrete function u_h at a spatial point x_i at time t^n. In our notation $\forall n: \quad \tau = t^{n+1} - t^n$ and $\forall i: \quad h = x_{i+1} - x_i$.

3.1.1 Linear transport equation

Consider the following Initial Value Problem (IVP) for linear transport equation in one space dimension:

$$\begin{cases} \partial_t u + c\partial_x u = 0, \quad (x,t) \in \mathbb{R} \times \mathbb{R}_+, \\ u(x,0) = u_0(x), \quad x \in \mathbb{R}. \end{cases} \tag{3.1}$$

A solution to this problem is represented by:

$$u(x,t) = u_0(x - ct), \tag{3.2}$$

and reflects the simple fact of propagation of the initial profile $u_0(x)$ along $\mathbb{R}$ with a constant speed c, as for example in Figure 3.1. If the initial data $u_0(x)$ is smooth, then (3.2) is a classical solution. Otherwise (3.2) is a nonclassical or generalized solution which is both a vanishing viscosity and an entropy admissible weak solution.

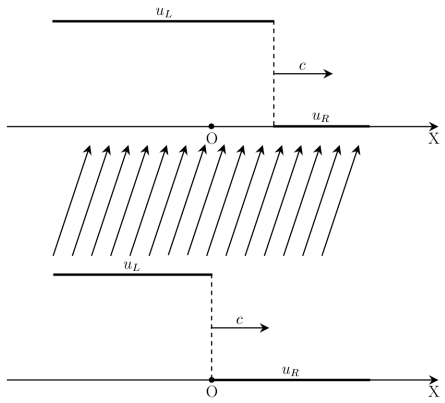

Fig. 3.1. Solution to linear transport equation with $c > 0$.

Let us instead of dealing with the original or unperturbed linear transport equation in the IVP (3.1) consider a regularized or perturbed equation of the following form:

$$\partial_t u + c\partial_x u = \epsilon \partial_{xx} u, \tag{3.3}$$

and show the tight connection between a viscosity coefficient ϵ and the most well-known finite difference schemes for numerically solving the original or unperturbed linear transport equation.

We first need to discretize the regularized equation (3.3) in space and time. For this purpose we employ the following centered difference scheme:

$$\frac{u_{h,i}^{n+1} - u_{h,i}^{n}}{\tau} + c\frac{u_{h,i+1}^{n} - u_{h,i-1}^{n}}{2h} = \epsilon \frac{u_{h,i-1}^{n} - 2u_{h,i}^{n} + u_{h,i+1}^{n}}{h^2} \tag{3.4}$$

or

$$u_{h,i}^{n+1} = \frac{\tau}{h}\left(\frac{c}{2} + \frac{\epsilon}{h}\right) u_{h,i-1}^{n} + \left(1 - \frac{2\epsilon\tau}{h^2}\right) u_{h,i}^{n} + \frac{\tau}{h}\left(-\frac{c}{2} + \frac{\epsilon}{h}\right) u_{h,i+1}^{n}. \qquad (3.5)$$

The scheme does not depend on the direction of flow and is identified by the numerical stencil shown in Figure 3.2.

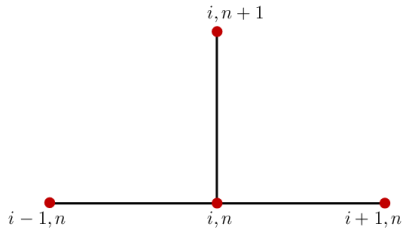

Fig. 3.2. CD numerical stencil.

Formula (3.5) provides an explicit algorithm for evolving the numerical solution u_h in time and has resulted from the first- and second-order approximations to the time and space derivatives of the perturbed equation (3.3), respectively.

At the next stage of the current discussion we perform a von Neumann stability analysis to determine within which range of variation of the viscosity coefficient ϵ the centered difference scheme (3.5) is stable. The numerical solution u_h can be expanded in the orthonormal basis $\left\{e^{Imx}\right\}_{m=-\infty}^{+\infty}$ as follows:

$$u_h = \sum_{m=-\infty}^{+\infty} C_m(t^n) e^{Imx_i}, \qquad (3.6)$$

where $C_m(t^n)$ is an m-th complex amplitude at time t^n and $I = \sqrt{-1}$. Substituting the expansion (3.6) into the formula (3.5) yields:

$$\forall m: \quad C_m^{n+1} = \overbrace{\left(\left(1 - \frac{2\epsilon\tau}{h^2}\right) + \frac{\tau}{h}\left(\frac{c}{2} + \frac{\epsilon}{h}\right) e^{-Imh} + \frac{\tau}{h}\left(-\frac{c}{2} + \frac{\epsilon}{h}\right) e^{Imh}\right)}^{G} C_m^n. \qquad (3.7)$$

Using the Euler formula:

$$e^{Ix} = \cos x + I \sin x \tag{3.8}$$

simplifies (3.7) as:

$$\forall m: \quad C_m^{n+1} = \overbrace{\left(1 - \frac{2\epsilon\tau}{h^2}\left(1 - \cos mh\right) - I\frac{c\tau}{h}\sin mh\right)}^{G} C_m^n. \tag{3.9}$$

For stability one requires $|G|^2 \leq 1$. In our case, this requirement is equivalent to:

$$\forall m: \quad \epsilon^2 - \frac{h^2}{\tau(1-\cos mh)}\epsilon + \frac{c^2h^2}{4}\frac{1+\cos mh}{1-\cos mh} \leq 0. \tag{3.10}$$

It is not difficult to see that the roots of the corresponded quadratic equation are:

$$\epsilon_1 = \frac{h^2}{4\tau \sin^2 \frac{mh}{2}}\left(1 - \sqrt{d}\right) \quad \text{and} \quad \epsilon_2 = \frac{h^2}{4\tau \sin^2 \frac{mh}{2}}\left(1 + \sqrt{d}\right), \tag{3.11}$$

where

$$d = 1 - \left(\frac{c\tau}{h}\right)^2 \sin^2 mh. \tag{3.12}$$

Since the viscosity coefficient ϵ is assumed to be a real number, we require that $d \geq 0$. This requirement yields the following condition:

$$\left|\frac{c\tau}{h}\right| \leq 1, \tag{3.13}$$

where $a = \left|\frac{c\tau}{h}\right|$ is called the CFL number. Conditions of this type are named for Courant, Friedrichs, and Lewy, and are referred to as CFL conditions. Using the roots ϵ_1 and ϵ_2 of the corresponded quadratic equation, we can rewrite the inequality (3.10) in the following form:

$$\forall m: \quad \left(\epsilon - \frac{h^2}{4\tau}\overbrace{\frac{1-\sqrt{d}}{\sin^2\frac{mh}{2}}}^{f_1(mh)}\right)\left(\epsilon - \frac{h^2}{4\tau}\overbrace{\frac{1+\sqrt{d}}{\sin^2\frac{mh}{2}}}^{f_2(mh)}\right) \leq 0. \tag{3.14}$$

To solve the above inequality, we first need to analyze the behavior of the functions $f_1(mh)$ and $f_2(mh)$. The function $f_1(mh) = \dfrac{1-\sqrt{d}}{\sin^2\frac{mh}{2}}$ is bounded on the interval $[0, 2\pi]$ from below and above:

$$\forall mh \in [0, 2\pi]: \quad f_1(mh) \in [0, 2a^2]. \tag{3.15}$$

The function $f_2(mh) = \dfrac{1+\sqrt{d}}{\sin^2\frac{mh}{2}}$ is bounded on the interval $[0, 2\pi]$ only from below:

$$\forall mh \in [0, 2\pi]: \quad f_2(mh) \in [2, +\infty]. \tag{3.16}$$

We also note that:

$$\forall mh \in [0, 2\pi]: \quad f_1(mh) \leq f_2(mh). \tag{3.17}$$

These functions are plotted for $a = 0.9$ in Figure 3.3.

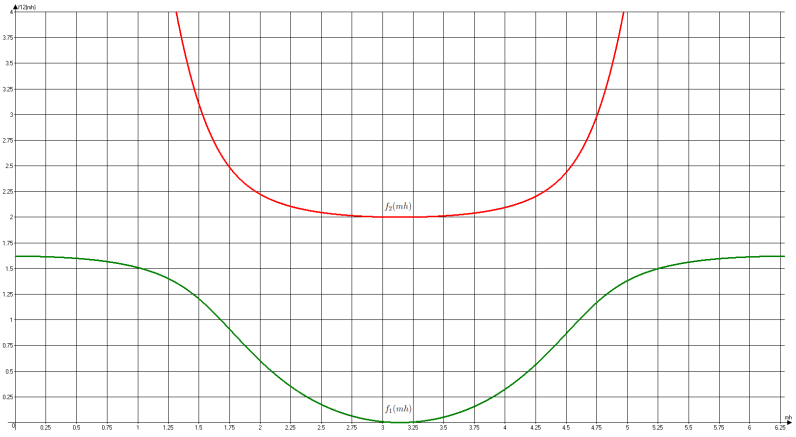

Fig. 3.3. Functions $f_1(mh)$ and $f_2(mh)$ with $a = 0.9$.

The inequality (3.14) is equivalent to the following two separate systems of inequalities:

$$\forall m: \quad \begin{cases} \epsilon - \dfrac{h^2}{4\tau}\dfrac{1-\sqrt{d}}{\sin^2\frac{mh}{2}} \le 0, \\ \epsilon - \dfrac{h^2}{4\tau}\dfrac{1+\sqrt{d}}{\sin^2\frac{mh}{2}} \ge 0. \end{cases} \qquad \text{or} \quad \forall m: \quad \begin{cases} \epsilon - \dfrac{h^2}{4\tau}\dfrac{1-\sqrt{d}}{\sin^2\frac{mh}{2}} \ge 0, \\ \epsilon - \dfrac{h^2}{4\tau}\dfrac{1+\sqrt{d}}{\sin^2\frac{mh}{2}} \le 0. \end{cases} \tag{3.18}$$

It turns out that the left system of inequalities does not have a solution. It obviously follows from the condition (3.17). To find the solution of the right system of inequalities, we can use the diagram depicted in Figure 3.4:

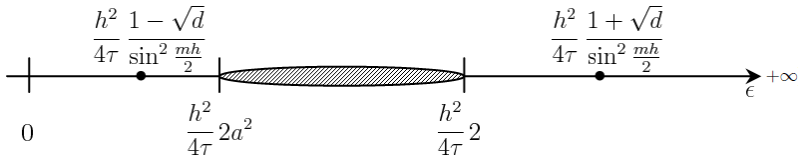

Fig. 3.4. Diagram.

from which it is easily seen that the centered difference scheme (3.5) for the regularized linear transport equation (3.3) is conditionally stable with the following stability conditions:

$$a \le 1 \quad \text{and} \quad \frac{h^2}{2\tau}a^2 \le \epsilon \le \frac{h^2}{2\tau}. \tag{3.19}$$

It is more convenient to determine the viscosity coefficient ϵ as follows:

$$\epsilon = C_{\max} h \left| f'(u) \right| = C_{\max} h \left| c \right|, \tag{3.20}$$

where the range of variation of $C_{\max}$ can be obtained from the stability conditions (3.19). Indeed, the requirements:

$$C_{\max}^{\min} h \left| c \right| = \frac{h^2}{2\tau}a^2 \quad \text{and} \quad C_{\max}^{\max} h \left| c \right| = \frac{h^2}{2\tau} \tag{3.21}$$

yield the new definition of the stability conditions (3.19):

$$a \leq 1 \quad \text{and} \quad \epsilon = C_{\max} h \left| c \right| \quad \text{with} \quad C_{\max} \in \left[\frac{a}{2}, \frac{1}{2a} \right]. \tag{3.22}$$

Remark 3.1.1. *We first observe that the viscosity coefficient ϵ vanishes as $\mathcal{O}(h)$ when a grid size h goes to zero. Therefore, it is sometimes referred to as the first-order artificial viscosity.*

Remark 3.1.2. *A von Neumann stability analysis itself does not use the information about the smoothness of the initial profile. Therefore, the above derivations are equally valid for both smooth and discontinuous sets of the initial data.*

Remark 3.1.3. *We also note that the centered difference scheme for the original or unperturbed linear transport equation can be obtained by setting $\epsilon = 0$ in the related formula* (3.5) *and is unconditionally unstable since $0 < \frac{a}{2} h \left| c \right|$.*

Let us now check if the centered difference scheme (3.5) for the regularized linear transport equation (3.3) under the stability conditions (3.22) is monotone. To do that, we first note that a new computed value $u_{h,i}^{n+1}$ defined by the formula (3.5) is in fact a linear combination of the old computed values $u_{h,i-1}^{n}$, $u_{h,i}^{n}$, and $u_{h,i+1}^{n}$. Then we introduce the definition of a convex linear combination.

Definition 3.1.1. *Given a finite number of points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ in a real vector space $\mathbb{R}^d$, a convex linear combination of these points is a point of the form:*

$$A_1 \mathbf{x}_1 + A_2 \mathbf{x}_2 + ... + A_n \mathbf{x}_n,$$

where the real numbers A_i satisfy $A_i \geq 0$ and $A_1 + A_2 + ... + A_n = 1$.

As a particular example, every convex linear combination of n real numbers $x_1, x_2, ..., x_n \in \mathbb{R}$ lies between $\min\limits_{i=1,n} x_i$ and $\max\limits_{i=1,n} x_i$. It is now clear that if the formula (3.5) represents

a convex linear combination of $u^n_{h,i-1}$, $u^n_{h,i}$, and $u^n_{h,i+1}$, then the new computed value $u^{n+1}_{h,i}$ lies between the minimum and maximum of the values computed at the previous time level t^n and the centered difference scheme (3.5) is monotone. It is not difficult to verify that the sum of all coefficients of the linear combination (3.5) is exactly 1 and the requirement $A_i \geq 0$ yields:

$$\frac{1}{2}h\,|c| \leq \epsilon \leq \frac{h^2}{2\tau} \tag{3.23}$$

or

$$\epsilon = C_{\max} h\,|c| \quad \text{with} \quad C_{\max} \in \left[\frac{1}{2}, \frac{1}{2a}\right]. \tag{3.24}$$

Compare the monotonicity condition (3.24) with the stability conditions (3.22) and note that $\left[\frac{1}{2}, \frac{1}{2a}\right] \subset \left[\frac{a}{2}, \frac{1}{2a}\right]$ and that $\frac{a}{2} \leq \frac{1}{2} \leq \frac{1}{2a}$.

Let us now summarize. The centered difference scheme (3.5) for the regularized linear transport equation (3.3) is stable if:

$$a \leq 1 \quad \text{and} \quad \epsilon = C_{\max} h\,|c| \quad \text{with} \quad C_{\max} \in \left[\frac{a}{2}, \frac{1}{2a}\right]. \tag{3.25}$$

The centered difference scheme (3.5) for the regularized linear transport equation (3.3) is both stable and monotone if:

$$a \leq 1 \quad \text{and} \quad \epsilon = C_{\max} h\,|c| \quad \text{with} \quad C_{\max} \in \left[\frac{1}{2}, \frac{1}{2a}\right]. \tag{3.26}$$

We now show three classical finite difference algorithms for numerical solving the original or unperturbed linear transport equation. These numerical algorithms are obtained from the centered difference scheme (3.5) by the following values of the

first-order artificial viscosity: $\epsilon = \frac{a}{2}h\,|c|$ - Lax-Wendroff scheme, $\epsilon = \frac{1}{2}h\,|c|$ - upwind scheme, and $\epsilon = \frac{1}{2a}h\,|c|$ - Lax−Friedrichs scheme.

Lax − Wendroff scheme :

The Lax-Wendroff scheme can be obtained from the centered difference formula (3.5) by $\epsilon = \frac{a}{2}h\,|c|$ and is defined as follows:

$$u_{h,i}^{n+1} = u_{h,i}^{n} - \frac{1}{2}a\ \mathrm{sgn}(c)\left(u_{h,i+1}^{n} - u_{h,i-1}^{n}\right) + \frac{1}{2}a^2\left(u_{h,i-1}^{n} - 2u_{h,i}^{n} + u_{h,i+1}^{n}\right), \qquad (3.27)$$

where $\mathrm{sgn}(c) = \frac{c}{|c|}$. As it is seen from the previous analysis - the scheme is stable if $a \leq 1$ but is not monotone. Interestingly, the Lax-Wendroff scheme was originally derived for the unperturbed equation as a predictor-corrector method. In this method we first compute the values $u_{h,i-\frac{1}{2}}^{n+\frac{1}{2}}$ and $u_{h,i+\frac{1}{2}}^{n+\frac{1}{2}}$ at the mid-points:

$$u_{h,i-\frac{1}{2}}^{n+\frac{1}{2}} = \overbrace{\frac{1}{2}\left(u_{h,i-1}^{n} + u_{h,i}^{n}\right)}^{u_{h,i-\frac{1}{2}}^{n}} - \frac{1}{2}a\ \mathrm{sgn}(c)\left(u_{h,i}^{n} - u_{h,i-1}^{n}\right), \qquad (3.28)$$

$$u_{h,i+\frac{1}{2}}^{n+\frac{1}{2}} = \overbrace{\frac{1}{2}\left(u_{h,i}^{n} + u_{h,i+1}^{n}\right)}^{u_{h,i+\frac{1}{2}}^{n}} - \frac{1}{2}a\ \mathrm{sgn}(c)\left(u_{h,i+1}^{n} - u_{h,i}^{n}\right), \qquad (3.29)$$

and after that we write for the desired value $u_{h,i}^{n+1}$:

$$u_{h,i}^{n+1} = u_{h,i}^{n} - a\ \mathrm{sgn}(c)\left(u_{h,i+\frac{1}{2}}^{n+\frac{1}{2}} - u_{h,i-\frac{1}{2}}^{n+\frac{1}{2}}\right). \qquad (3.30)$$

Manipulating the above results will directly yield (3.27).

Upwind scheme :

The upwind scheme can be obtained from the centered difference formula (3.5) by setting $\epsilon = \frac{1}{2}h\,|c|$ and reads as:

$$u_{h,i}^{n+1} = u_{h,i}^{n} - \frac{1}{2}a\,(\mathrm{sgn}(c)+1)\left(u_{h,i}^{n} - u_{h,i-1}^{n}\right) - \frac{1}{2}a\,(\mathrm{sgn}(c)-1)\left(u_{h,i+1}^{n} - u_{h,i}^{n}\right). \quad (3.31)$$

The scheme is both stable and monotone if $a \leq 1$. Note that formula (3.31) automatically chooses which part of the CD numerical stencil must be employed in accordance with the direction of flow, see Figure 3.5.

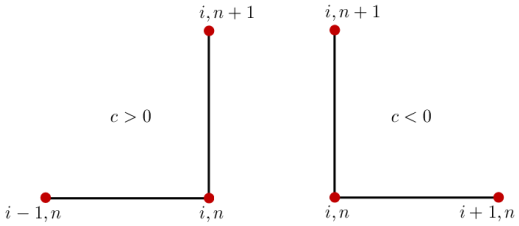

Fig. 3.5. Upwind numerical stencils.

Lax − Friedrichs scheme :

The most diffusive scheme of this class of methods is known as the Lax−Friedrichs scheme and corresponds to the centered difference formula (3.5) with $\epsilon = \frac{1}{2a}h\,|c|$:

$$u_{h,i}^{n+1} = \frac{1}{2}\left(u_{h,i-1}^{n} + u_{h,i+1}^{n}\right) - \frac{1}{2}a\ \mathrm{sgn}(c)\left(u_{h,i+1}^{n} - u_{h,i-1}^{n}\right). \quad (3.32)$$

The scheme is both stable and monotone if $a \leq 1$. The Lax−Friedrichs method can be considered as a successful attempt to retrieve the pure centered difference scheme for the unperturbed equation (see Remark 3.1.3.) by replacing $u_{h,i}^{n}$ term with the average $\frac{1}{2}\left(u_{h,i-1}^{n} + u_{h,i+1}^{n}\right)$. The method is identified by the numerical stencil depicted in Figure 3.6.

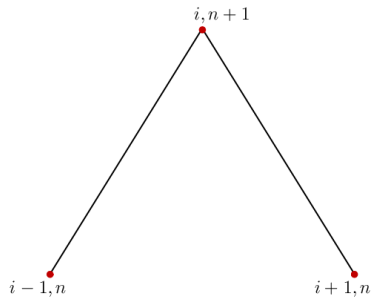

Fig. 3.6. Lax−Friedrichs numerical stencil.

Examples :

The examples of numerical solutions to linear transport equation, in the context of the current discussion, are shown in Figures 3.7 and 3.8. It is seen from Figure 3.7 that using the Lax-Wendroff artificial viscosity is more preferable for the transport of smooth or regular initial conditions. Whereas the upwind artificial viscosity performs better with discontinuities, see Figure 3.8. It can thus be understood that the distribution of various amount of artificial viscosity between different parts of the same numerical solution is probably the best strategy. Indeed, we add less viscosity where the solution is smooth, thereby preserving the information from smearing, and we add more viscosity to discontinuities to provide a monotone transition between "before" and "after" state variables.

Remark 3.1.4. *The Lax-Wendroff numerical solution in Figure 3.8 contains the considerable overshoot that, however, does not much grow with time.*

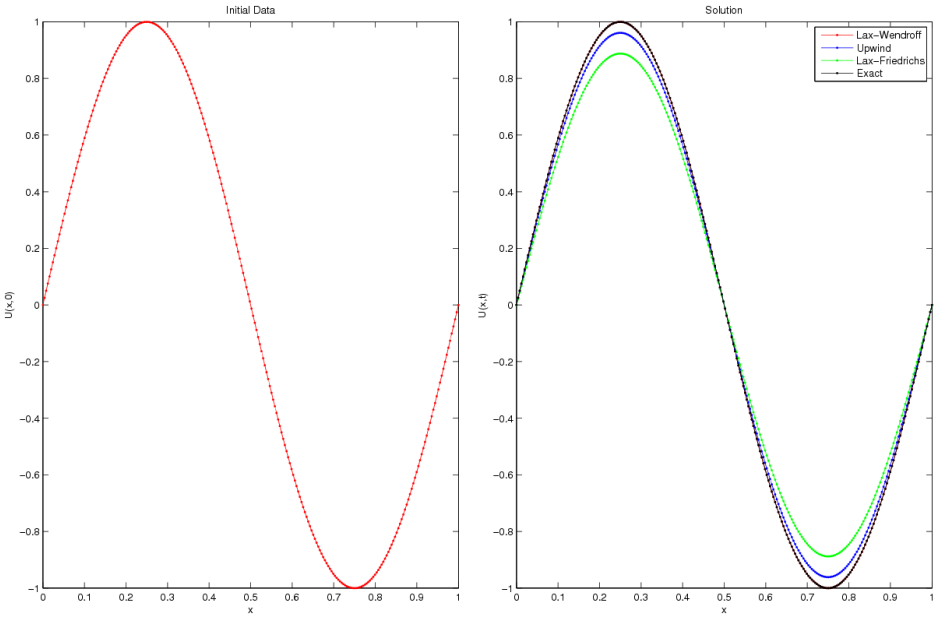

Fig. 3.7. Linear transport of smooth data.

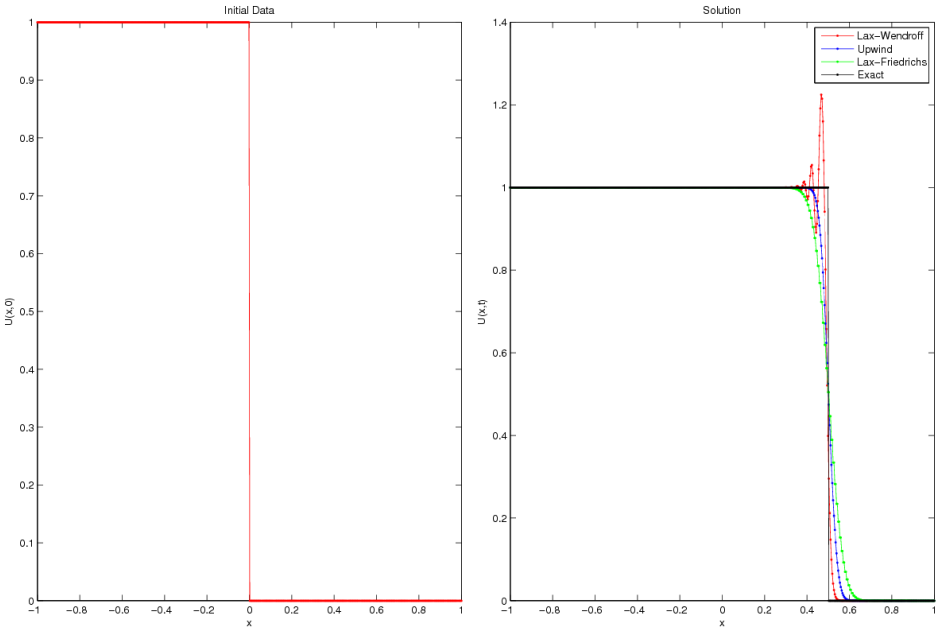

Fig. 3.8. Linear transport of discontinuous data.

3.1.2 Extension to nonlinear scalar equations

The extension of the first-order artificial viscosity numerical stabilization technique to the nonlinear scalar equations is straightforward. Consider a nonlinear scalar equation in one space dimension:

$$\partial_t u + \partial_x f(u) = 0. \tag{3.33}$$

On the one hand, the centered difference spatial discretization of the original or unperturbed equation generates unconditionally unstable numerical solutions, as for instance, for inviscid nonlinear Burgers' equation in Figure 3.9.

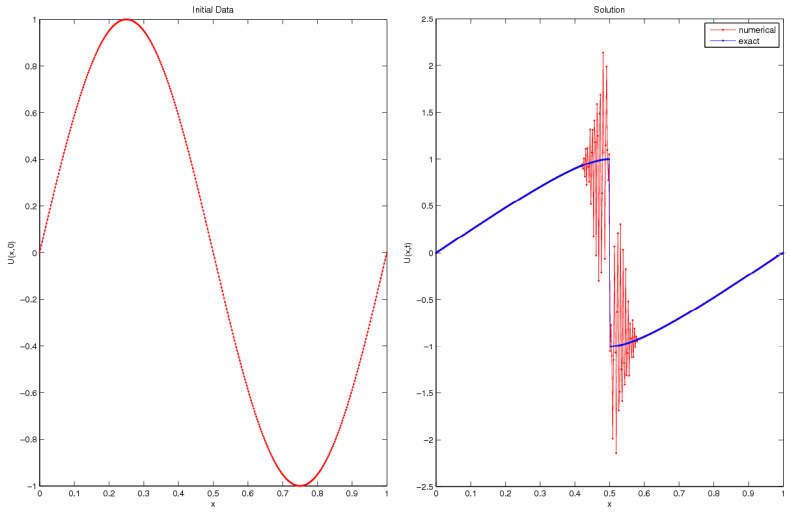

Fig. 3.9. CD-based numerical solution to inviscid nonlinear Burgers' equation with no stabilization.

On the other hand, the most natural upwind scheme requires information on the direction of flow at a spatial point x_i at time t^n, which itself is difficult to determine in the case of nonlinear equations. To overcome these difficulties, we first regularize the original equation:

$$\partial_t u + \partial_x f(u) = \partial_x \left(\epsilon(u) \partial_x u \right), \tag{3.34}$$

and then apply the simplest centered difference scheme:

$$\frac{u_{h,i}^{n+1}-u_{h,i}^{n}}{\tau}+\frac{f_{h,i+1}^{n}-f_{h,i-1}^{n}}{2h}= \tag{3.35}$$

$$\frac{1}{2}\left(\frac{\epsilon_{h,i-1}^{n}+\epsilon_{h,i}^{n}}{h^2}u_{h,i-1}^{n}-\frac{\epsilon_{h,i-1}^{n}+2\epsilon_{h,i}^{n}+\epsilon_{h,i+1}^{n}}{h^2}u_{h,i}^{n}+\frac{\epsilon_{h,i}^{n}+\epsilon_{h,i+1}^{n}}{h^2}u_{h,i+1}^{n}\right)$$

or

$$u_{h,i}^{n+1}=u_{h,i}^{n}-\frac{\tau}{2h}\left(f_{h,i+1}^{n}-f_{h,i-1}^{n}\right)+ \tag{3.36}$$

$$\frac{\tau}{2}\left(\frac{\epsilon_{h,i-1}^{n}+\epsilon_{h,i}^{n}}{h^2}u_{h,i-1}^{n}-\frac{\epsilon_{h,i-1}^{n}+2\epsilon_{h,i}^{n}+\epsilon_{h,i+1}^{n}}{h^2}u_{h,i}^{n}+\frac{\epsilon_{h,i}^{n}+\epsilon_{h,i+1}^{n}}{h^2}u_{h,i+1}^{n}\right)$$

along with the first-order artificial viscosity:

$$\epsilon_{h,i}^{n}\equiv\epsilon\left(u_{h,i}^{n}\right)=C_{\max}h\left|f'\left(u_{h,i}^{n}\right)\right|. \tag{3.37}$$

Note that the numerical approximation to the derivative $\partial_x\left(\epsilon(u)\partial_x u\right)$ is symmetrized for better performance. The example of a numerical solution to inviscid nonlinear Burgers' equation is shown in Figure 3.10. Note that in this case, $C_{\max}=\frac{1}{2}$ does not correspond to the upwind difference scheme for the unperturbed equation and that a new stability analysis is required to determine the acceptable range of $C_{\max}$. From this perspective $C_{\max}$ can be considered as a free or tunable parameter of the numerical scheme which we define empirically.

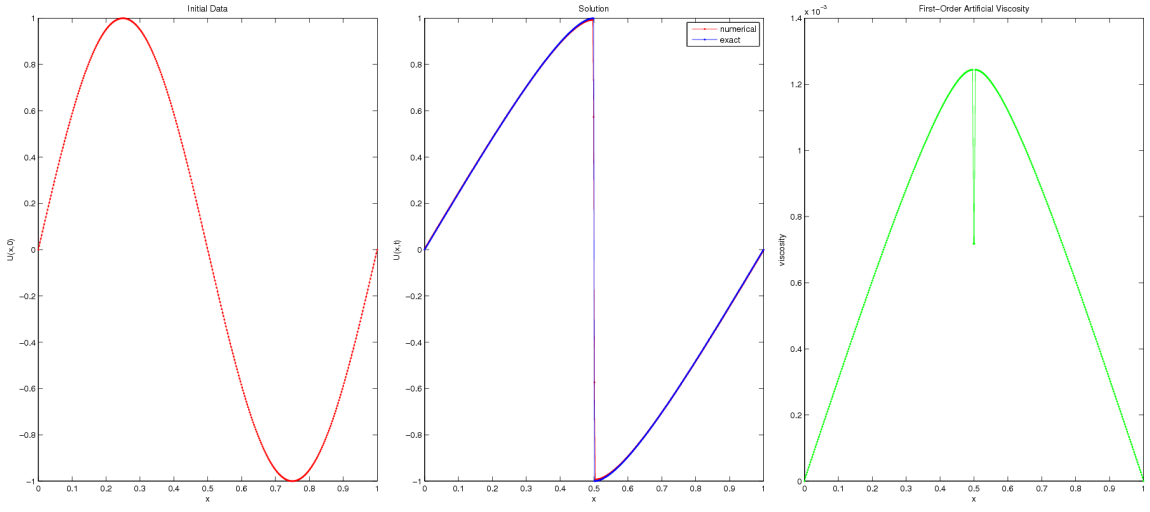

Fig. 3.10. CD-based numerical solution to inviscid nonlinear Burgers' equation with stabilization $\left(C_{\max} = \frac{1}{2}\right)$.

3.1.3 Extension to multi-dimensional grids

As usual we discretize the regularized PDE:

$$\partial_t u + \nabla \cdot \mathbf{f}(u) = \nabla \cdot \left(\epsilon(u) \nabla u\right) \tag{3.38}$$

by means of the centered difference scheme with the first-order artificial viscosity defined as follows:

$$\epsilon^n_{h,ijk} \equiv \epsilon\left(u^n_{h,ijk}\right) = C_{\max} h \left|\mathbf{f}'\left(u^n_{h,ijk}\right)\right|. \tag{3.39}$$

3.1.4 Extension to nonuniform grids

Instead of h in formula (3.39) we should take the maximum local grid size associated with the node ijk, i.e. $h = \max\limits_{l=1,2^d} h_{ijk,l}$.

3.2 DG(0) approximation and numerical fluxes

In contrast to the finite difference approximation, the DG(0) or basic finite volume approach deals with the cell averaged values $\bar{u}^n_{h,i}$ of a numerical solution u_h at time t^n and relies on the intercell numerical fluxes. In our notation $\forall n: \quad \tau = t^{n+1} - t^n$ and $\forall i: \quad h = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}} = x_{i+1} - x_i$, see Figure 3.11.

To demonstrate the stabilization capability of the first-order artificial viscosity in the framework of DG(0) approximation, we first represent the regularized equation in the following form:

$$\partial_t u + \partial_x f(u) = -\partial_x q, \tag{3.40}$$

where

$$q = -\epsilon(u)\partial_x u. \tag{3.41}$$

Then we construct a grid:

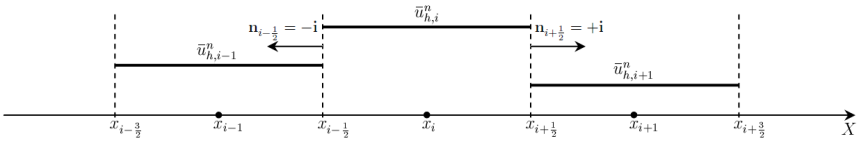

Fig. 3.11. DG(0) grid.

and integrate the regularized equation (3.40) over space and time:

$$\int_{t^n}^{t^{n+1}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left[\partial_t u + \partial_x f(u)\right] dxdt = -\int_{t^n}^{t^{n+1}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \partial_x q dxdt, \tag{3.42}$$

or

$$\frac{1}{\tau}\overbrace{\left[\frac{1}{h}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u\left(x,t^{n+1}\right)dx\right]}^{\bar{u}_{h,i}^{n+1}} - \frac{1}{\tau}\overbrace{\left[\frac{1}{h}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u\left(x,t^{n}\right)dx\right]}^{\bar{u}_{h,i}^{n}} +$$

$$\frac{1}{h}\overbrace{\left[\frac{1}{\tau}\int_{t^n}^{t^{n+1}} f\left(u\left(x_{i+\frac{1}{2}},t\right)\right)dt\right]}^{f_{h,i+\frac{1}{2}}^{n}} - \frac{1}{h}\overbrace{\left[\frac{1}{\tau}\int_{t^n}^{t^{n+1}} f\left(u\left(x_{i-\frac{1}{2}},t\right)\right)dt\right]}^{f_{h,i-\frac{1}{2}}^{n}} =$$

$$-\frac{1}{h}\overbrace{\left[\frac{1}{\tau}\int_{t^n}^{t^{n+1}} q\left(x_{i+\frac{1}{2}},t\right)dt\right]}^{q_{h,i+\frac{1}{2}}^{n}} + \frac{1}{h}\overbrace{\left[\frac{1}{\tau}\int_{t^n}^{t^{n+1}} q\left(x_{i-\frac{1}{2}},t\right)dt\right]}^{q_{h,i-\frac{1}{2}}^{n}}, \tag{3.43}$$

or

$$\frac{\bar{u}_{h,i}^{n+1}-\bar{u}_{h,i}^{n}}{\tau} + \frac{f_{h,i+\frac{1}{2}}^{n} - f_{h,i-\frac{1}{2}}^{n}}{h} = -\frac{q_{h,i+\frac{1}{2}}^{n} - q_{h,i-\frac{1}{2}}^{n}}{h}, \tag{3.44}$$

or

$$\bar{u}_{h,i}^{n+1} = \bar{u}_{h,i}^{n} - \frac{\tau}{h}\left(f_{h,i+\frac{1}{2}}^{n} - f_{h,i-\frac{1}{2}}^{n}\right) - \frac{\tau}{h}\left(q_{h,i+\frac{1}{2}}^{n} - q_{h,i-\frac{1}{2}}^{n}\right). \tag{3.45}$$

Here $\bar{u}_{h,i}^{n}$ and $\bar{u}_{h,i}^{n+1}$ are the averaged values of the numerical solution u_h in cell i at time t^n and t^{n+1}, respectively. We use a centered approximation to define the inviscid numerical fluxes $f_{h,i-\frac{1}{2}}^{n}$ and $f_{h,i+\frac{1}{2}}^{n}$ as:

$$f_{h,i-\frac{1}{2}}^{n} \approx \frac{1}{2}\left(f_{h,i-1}^{n} + f_{h,i}^{n}\right) \tag{3.46}$$

and

$$f_{h,i+\frac{1}{2}}^{n} \approx \frac{1}{2}\left(f_{h,i}^{n} + f_{h,i+1}^{n}\right). \tag{3.47}$$

To find an appropriate approximation to the viscous numerical fluxes $q^n_{h,i-\frac{1}{2}}$ and $q^n_{h,i+\frac{1}{2}}$, let us integrate the discrete version of the equation (3.41):

$$q^n_h = -\epsilon^n_h \partial_x u^n_h, \tag{3.48}$$

where $\epsilon^n_h = \left\{\bar{\epsilon}^n_{h,i}\right\}^{\text{cells}}_{i=1}$ and $u^n_h = \left\{\bar{u}^n_{h,i}\right\}^{\text{cells}}_{i=1}$ as follows:

$$\overbrace{\frac{1}{h}\int_{x_{i-1}}^{x_i} q^n_h dx}^{q^n_{h,i-\frac{1}{2}}} = -\frac{1}{h}\int_{x_{i-1}}^{x_i} \epsilon^n_h \partial_x u^n_h dx \tag{3.49}$$

and

$$\overbrace{\frac{1}{h}\int_{x_i}^{x_{i+1}} q^n_h dx}^{q^n_{h,i+\frac{1}{2}}} = -\frac{1}{h}\int_{x_i}^{x_{i+1}} \epsilon^n_h \partial_x u^n_h dx. \tag{3.50}$$

Taking into account the discontinuous nature of the piecewise constant numerical solution u_h at any arbitrary time t^n, we can write these equations as:

$$q^n_{h,i-\frac{1}{2}} = -\frac{1}{h}\int_{x_{i-1}}^{x_i} \epsilon^n_h \left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i-1}\right) \delta\left(x - x_{i-\frac{1}{2}}\right) dx \tag{3.51}$$

and

$$q^n_{h,i+\frac{1}{2}} = -\frac{1}{h}\int_{x_i}^{x_{i+1}} \epsilon^n_h \left(\bar{u}^n_{h,i+1} - \bar{u}^n_{h,i}\right) \delta\left(x - x_{i+\frac{1}{2}}\right) dx, \tag{3.52}$$

which is identical to:

$$q^n_{h,i-\frac{1}{2}} = -\frac{1}{h}\epsilon^n_{h,i-\frac{1}{2}} \left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i-1}\right) \tag{3.53}$$

and

$$q^n_{h,i+\frac{1}{2}} = -\frac{1}{h}\epsilon^n_{h,i+\frac{1}{2}} \left(\bar{u}^n_{h,i+1} - \bar{u}^n_{h,i}\right). \tag{3.54}$$

Using:

$$\epsilon^n_{h,i-\frac{1}{2}} \approx \frac{1}{2}\left(\bar{\epsilon}^n_{h,i-1} + \bar{\epsilon}^n_{h,i}\right) \tag{3.55}$$

and

$$\epsilon^n_{h,i+\frac{1}{2}} \approx \frac{1}{2}\left(\bar{\epsilon}^n_{h,i} + \bar{\epsilon}^n_{h,i+1}\right) \tag{3.56}$$

eventually yields:

$$q^n_{h,i-\frac{1}{2}} \approx -\frac{1}{h}\left(\frac{\bar{\epsilon}^n_{h,i-1} + \bar{\epsilon}^n_{h,i}}{2}\right)\left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i-1}\right) \tag{3.57}$$

and

$$q^n_{h,i+\frac{1}{2}} \approx -\frac{1}{h}\left(\frac{\bar{\epsilon}^n_{h,i} + \bar{\epsilon}^n_{h,i+1}}{2}\right)\left(\bar{u}^n_{h,i+1} - \bar{u}^n_{h,i}\right). \tag{3.58}$$

To unify the notation and summarize for the time being, let us slightly change the above formulations. We first write the equations (3.44) or (3.45) in the following form:

$$\frac{\bar{u}^{n+1}_{h,i} - \bar{u}^n_{h,i}}{\tau} + \frac{\hat{F}^n_{h,i-\frac{1}{2}} + \hat{F}^n_{h,i+\frac{1}{2}}}{h} = -\frac{\hat{Q}^n_{h,i-\frac{1}{2}} + \hat{Q}^n_{h,i+\frac{1}{2}}}{h} \tag{3.59}$$

or

$$\bar{u}^{n+1}_{h,i} = \bar{u}^n_{h,i} - \frac{\tau}{h}\left(\hat{F}^n_{h,i-\frac{1}{2}} + \hat{F}^n_{h,i+\frac{1}{2}}\right) - \frac{\tau}{h}\left(\hat{Q}^n_{h,i-\frac{1}{2}} + \hat{Q}^n_{h,i+\frac{1}{2}}\right), \tag{3.60}$$

then define the inviscid numerical fluxes $\hat{F}^n_{h,i-\frac{1}{2}}$ and $\hat{F}^n_{h,i+\frac{1}{2}}$ as:

$$\hat{F}^n_{h,i-\frac{1}{2}} \equiv \mathbf{f}^n_{h,i-\frac{1}{2}} \cdot \mathbf{n}_{i-\frac{1}{2}} \approx \frac{1}{2}\left(\mathbf{f}^n_{h,i-1} + \mathbf{f}^n_{h,i}\right) \cdot \mathbf{n}_{i-\frac{1}{2}} \tag{3.61}$$

and

$$\hat{F}^n_{h,i+\frac{1}{2}} \equiv \mathbf{f}^n_{h,i+\frac{1}{2}} \cdot \mathbf{n}_{i+\frac{1}{2}} \approx \frac{1}{2}\left(\mathbf{f}^n_{h,i} + \mathbf{f}^n_{h,i+1}\right) \cdot \mathbf{n}_{i+\frac{1}{2}}, \tag{3.62}$$

and viscous numerical fluxes $\hat{Q}^n_{h,i-\frac{1}{2}}$ and $\hat{Q}^n_{h,i+\frac{1}{2}}$ as:

$$\begin{aligned}\hat{Q}^n_{h,i-\frac{1}{2}} \equiv \mathbf{q}^n_{h,i-\frac{1}{2}} \cdot \mathbf{n}_{i-\frac{1}{2}} \approx \quad & \frac{1}{2}\left(\mathbf{q}^n_{h,i-1} + \mathbf{q}^n_{h,i}\right) \cdot \mathbf{n}_{i-\frac{1}{2}} + \\ & \frac{1}{h}\left(\frac{\bar{\epsilon}^n_{h,i-1} + \bar{\epsilon}^n_{h,i}}{2}\right)\left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i-1}\right)\end{aligned} \tag{3.63}$$

and

$$\hat{Q}^n_{h,i+\frac{1}{2}} \equiv \mathbf{q}^n_{h,i+\frac{1}{2}} \cdot \mathbf{n}_{i+\frac{1}{2}} \approx \quad \frac{1}{2}\left(\mathbf{q}^n_{h,i} + \mathbf{q}^n_{h,i+1}\right) \cdot \mathbf{n}_{i+\frac{1}{2}} + \frac{1}{h}\left(\frac{\bar{\epsilon}^n_{h,i} + \bar{\epsilon}^n_{h,i+1}}{2}\right)\left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i+1}\right), \tag{3.64}$$

where $\mathbf{n}_{i-\frac{1}{2}} = -\mathbf{i}$ and $\mathbf{n}_{i+\frac{1}{2}} = +\mathbf{i}$, see Figure 3.11. Note that in the framework of the DG(0) approximation, the first term in both equations (3.63) and (3.64) is identically zero. Substituting both inviscid (3.61)-(3.62) and viscous (3.63)-(3.64) numerical fluxes into the formulation (3.59) or (3.60) will exactly reproduce the centered difference discretization (3.35) or (3.36) - with the only exception that instead of nodal values $u^n_{h,i}$ one will have the cell averaged values $\bar{u}^n_{h,i}$ of the numerical solution u_h at time t^n. As usual:

$$\bar{\epsilon}^n_{h,i} \equiv \epsilon\left(\bar{u}^n_{h,i}\right) = C_{\max} h \left|\mathbf{f}'\left(\bar{u}^n_{h,i}\right)\right|. \tag{3.65}$$

Remark 3.2.1. *Note that the viscous numerical fluxes* $\hat{Q}^n_{h,i-\frac{1}{2}}$ *and* $\hat{Q}^n_{h,i+\frac{1}{2}}$ *defined by the equations* (3.63) *and* (3.64) *can be also represented as:*

$$\hat{Q}^n_{h,i-\frac{1}{2}} \equiv \mathbf{q}^n_{h,i-\frac{1}{2}} \cdot \mathbf{n}_{i-\frac{1}{2}} \approx \quad \frac{1}{2}\left(\mathbf{q}^n_{h,i-1} + \mathbf{q}^n_{h,i}\right) \cdot \mathbf{n}_{i-\frac{1}{2}} + \max\left(\frac{\bar{\epsilon}^n_{h,i-1}}{h}, \frac{\bar{\epsilon}^n_{h,i}}{h}\right)\left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i-1}\right) \tag{3.66}$$

and

$$\hat{Q}^n_{h,i+\frac{1}{2}} \equiv \mathbf{q}^n_{h,i+\frac{1}{2}} \cdot \mathbf{n}_{i+\frac{1}{2}} \approx \quad \frac{1}{2}\left(\mathbf{q}^n_{h,i} + \mathbf{q}^n_{h,i+1}\right) \cdot \mathbf{n}_{i+\frac{1}{2}} + \max\left(\frac{\bar{\epsilon}^n_{h,i}}{h}, \frac{\bar{\epsilon}^n_{h,i+1}}{h}\right)\left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i+1}\right). \tag{3.67}$$

Another useful observation that we would like to mention here is that the terms including the jump of the numerical solution u_h can be moved from the viscous numerical flux definitions (3.63) and (3.64) to the inviscid numerical flux definitions (3.61) and (3.62). In this case, we redefine the inviscid numerical fluxes $\hat{F}^n_{h,i-\frac{1}{2}}$ and $\hat{F}^n_{h,i+\frac{1}{2}}$ defined by the equations (3.61) and (3.62) as:

$$\begin{aligned}\hat{F}^n_{h,i-\frac{1}{2}} \equiv \mathbf{f}^n_{h,i-\frac{1}{2}} \cdot \mathbf{n}_{i-\frac{1}{2}} \approx \quad & \frac{1}{2}\left(\mathbf{f}^n_{h,i-1} + \mathbf{f}^n_{h,i}\right) \cdot \mathbf{n}_{i-\frac{1}{2}} + \\ & \frac{1}{h}\left(\frac{\bar{\epsilon}^n_{h,i-1} + \bar{\epsilon}^n_{h,i}}{2}\right)\left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i-1}\right) \end{aligned} \tag{3.68}$$

and

$$\begin{aligned}\hat{F}^n_{h,i+\frac{1}{2}} \equiv \mathbf{f}^n_{h,i+\frac{1}{2}} \cdot \mathbf{n}_{i+\frac{1}{2}} \approx \quad & \frac{1}{2}\left(\mathbf{f}^n_{h,i} + \mathbf{f}^n_{h,i+1}\right) \cdot \mathbf{n}_{i+\frac{1}{2}} + \\ & \frac{1}{h}\left(\frac{\bar{\epsilon}^n_{h,i} + \bar{\epsilon}^n_{h,i+1}}{2}\right)\left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i+1}\right). \end{aligned} \tag{3.69}$$

Using formula (3.65) in equations (3.68) and (3.69) yields:

$$\begin{aligned}\hat{F}^n_{h,i-\frac{1}{2}} \equiv \mathbf{f}^n_{h,i-\frac{1}{2}} \cdot \mathbf{n}_{i-\frac{1}{2}} \approx \quad & \frac{1}{2}\left(\mathbf{f}^n_{h,i-1} + \mathbf{f}^n_{h,i}\right) \cdot \mathbf{n}_{i-\frac{1}{2}} + \\ & C_{\max}\left(\frac{\left|\mathbf{f}'\left(\bar{u}^n_{h,i-1}\right)\right| + \left|\mathbf{f}'\left(\bar{u}^n_{h,i}\right)\right|}{2}\right)\left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i-1}\right) \end{aligned} \tag{3.70}$$

and

$$\begin{aligned}\hat{F}^n_{h,i+\frac{1}{2}} \equiv \mathbf{f}^n_{h,i+\frac{1}{2}} \cdot \mathbf{n}_{i+\frac{1}{2}} \approx \quad & \frac{1}{2}\left(\mathbf{f}^n_{h,i} + \mathbf{f}^n_{h,i+1}\right) \cdot \mathbf{n}_{i+\frac{1}{2}} + \\ & C_{\max}\left(\frac{\left|\mathbf{f}'\left(\bar{u}^n_{h,i}\right)\right| + \left|\mathbf{f}'\left(\bar{u}^n_{h,i+1}\right)\right|}{2}\right)\left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i+1}\right). \end{aligned} \tag{3.71}$$

The viscous numerical fluxes $\hat{Q}^n_{h,i-\frac{1}{2}}$ and $\hat{Q}^n_{h,i+\frac{1}{2}}$ defined by the equations (3.63) and (3.64) now become:

$$\hat{Q}^n_{h,i-\frac{1}{2}} \equiv \mathbf{q}^n_{h,i-\frac{1}{2}} \cdot \mathbf{n}_{i-\frac{1}{2}} \approx \frac{1}{2}\left(\mathbf{q}^n_{h,i-1} + \mathbf{q}^n_{h,i}\right) \cdot \mathbf{n}_{i-\frac{1}{2}} = 0 \longleftarrow \text{for DG(0)} \tag{3.72}$$

and

$$\hat{Q}^n_{h,i+\frac{1}{2}} \equiv \mathbf{q}^n_{h,i+\frac{1}{2}} \cdot \mathbf{n}_{i+\frac{1}{2}} \approx \frac{1}{2}\left(\mathbf{q}^n_{h,i} + \mathbf{q}^n_{h,i+1}\right) \cdot \mathbf{n}_{i+\frac{1}{2}} = 0 \longleftarrow \text{for DG(0)} \tag{3.73}$$

Remark 3.2.2. *Note that the inviscid numerical fluxes $\hat{F}^n_{h,i-\frac{1}{2}}$ and $\hat{F}^n_{h,i+\frac{1}{2}}$ defined by the equations* (3.70) *and* (3.71) *can be also represented as:*

$$\begin{aligned}\hat{F}^n_{h,i-\frac{1}{2}} \equiv \mathbf{f}^n_{h,i-\frac{1}{2}} \cdot \mathbf{n}_{i-\frac{1}{2}} \approx \quad & \frac{1}{2}\left(\mathbf{f}^n_{h,i-1} + \mathbf{f}^n_{h,i}\right) \cdot \mathbf{n}_{i-\frac{1}{2}} + \\ & C_{max} \max\left(\left|\mathbf{f}'\left(\bar{u}^n_{h,i-1}\right)\right|, \left|\mathbf{f}'\left(\bar{u}^n_{h,i}\right)\right|\right)\left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i-1}\right)\end{aligned} \tag{3.74}$$

and

$$\begin{aligned}\hat{F}^n_{h,i+\frac{1}{2}} \equiv \mathbf{f}^n_{h,i+\frac{1}{2}} \cdot \mathbf{n}_{i+\frac{1}{2}} \approx \quad & \frac{1}{2}\left(\mathbf{f}^n_{h,i} + \mathbf{f}^n_{h,i+1}\right) \cdot \mathbf{n}_{i+\frac{1}{2}} + \\ & C_{max} \max\left(\left|\mathbf{f}'\left(\bar{u}^n_{h,i}\right)\right|, \left|\mathbf{f}'\left(\bar{u}^n_{h,i+1}\right)\right|\right)\left(\bar{u}^n_{h,i} - \bar{u}^n_{h,i+1}\right).\end{aligned} \tag{3.75}$$

Remark 3.2.3. *The inviscid numerical fluxes $\hat{F}^n_{h,i-\frac{1}{2}}$ and $\hat{F}^n_{h,i+\frac{1}{2}}$ defined by the equations* (3.74) *and* (3.75) *are in fact the Lax-Friedrichs numerical fluxes.*

Remark 3.2.4. *Note that for the DG(0) approximation - the jump $[\![\bar{u}^n_h]\!] \sim \mathcal{O}(h)$, which implies the extension to general DG(p) methods as $[\![u^n_h]\!] \sim \mathcal{O}(h^{p+1})$.*

3.3 Entropy-based artificial viscosity

The concept of the entropy-based artificial viscosity implies the use of the residual of an entropy equation to impose the appropriate dissipation to different (smooth and discontinuous) parts of a numerical solution. To maintain the stability of the numerical solution, the entropy viscosity is bounded by the first-order artificial viscosity from above. Since the residual of an entropy equation is supposed to be vanishingly small - of the order of the *Local Truncation Error* (LTE) in smooth regions and arbitrarily large in shocks, the entropy viscosity is almost zero everywhere except in shocks, where it reaches the first-order upper bound.

It certainly does not make much sense to apply the above strategy to the numerical schemes which have been previously considered in this chapter. These schemes were actually discussed to develop some intuition of what will happen in Chapter IV. For instance, the attempt to replace the first-order viscosity by the entropy viscosity in centered difference formulas (3.5) or (3.36) will lead to loss of the stability of a numerical solution because of the restrictions imposed on the viscosity coefficient ϵ. The same is true in the case of the DG(0) approximation unless we stabilize both inviscid (3.70)-(3.71) and viscous (3.63)-(3.64) numerical fluxes, which, however, adds too much viscosity, especially to shocks.

In the present work we employ higher-order DG methods along with higher-order time integration schemes. In principle it accepts the use of edge stabilization terms in both inviscid and viscous numerical flux definitions without significant overdissipation, see Remark 3.2.4.

The entropy viscosity method can be formally constructed in a few steps:

- Given an entropy pair $(\eta(u), \boldsymbol{\psi}(u))$, define the entropy residual:

$$D_h = \partial_t \eta(u_h) + \nabla \cdot \boldsymbol{\psi}(u_h), \quad (\mathbf{x}, t) \in \Omega \times \mathbb{R}_+. \tag{3.76}$$

- Use this residual to define a viscosity, say ϵ_E:

$$\epsilon_E = C_E \mathcal{F}_+(D_h) \frac{h^2}{\|\eta(u_h) - \bar{\eta}_h\|_{L^\infty(\Omega)}}, \tag{3.77}$$

  where C_E is a tunable constant, $\mathcal{F}_+$ is a positive functional that remains to be specified, $h \equiv h(\mathbf{x})$ is the local mesh size at $\mathbf{x} \in \Omega$, and $\bar{\eta}_h = \dfrac{1}{\|\Omega\|} \displaystyle\int_\Omega \eta(u_h) d\mathbf{x}$ is the space-averaged value of the entropy. Since D_h is expected to oscillate, and this is especially true in shocks where D_h approximates a Dirac measure, the simplest functional that one can use to avoid negative values of D_h is $\mathcal{F}_+(D_h) = |D_h|$. Introducing the scaling coefficient h^2 together with the normalizing term $\|\eta(u_h) - \bar{\eta}_h\|_{L^\infty(\Omega)}$ gives to ϵ_E the dimension of a viscosity.

- Introduce an upper bound to the entropy viscosity:

$$\epsilon_{\max} = C_{\max} h \max_{\Delta_{\mathbf{x}}} |\mathbf{f}'(u_h)|, \tag{3.78}$$

  where $\Delta_{\mathbf{x}}$ is some neighborhood of $\mathbf{x}$.

- Define the entropy viscosity:

$$\epsilon = \min(\epsilon_E, \epsilon_{\max}). \tag{3.79}$$

In order to further describe the entropy viscosity method, we have to specify the numerical discretization. Implementation details depend on the numerical approximation. Implementation of the method for DG finite elements is described hereafter.

The above method is simple to implement, but simplicity has a price and the price paid here is the introduction of two tunable constants C_E and $C_{\max}$. In practice these two constants are tuned by testing the method on a coarse grid. For any given problem, the tuning is done quickly once and for all on a coarse mesh.

Remark 3.3.1. *The idea of using the entropy to design numerical methods for nonlinear conservation equations is not new. For instance, it is shown in [46, 47] that the entropy production can be used as a posteriori error indicator and therefore is useful for adaptive strategies.*

Remark 3.3.2. *Using a residual to construct a viscosity is not a new idea. For example, the residual of the conservation equation itself can be employed, see [39,42]. Although the residual of the conservation equation is a good error indicator, it is far less robust than the entropy residual. The reason is that consistency requires the equation residual vanishes as a grid size h goes to zero, whereas the entropy residual converges to a Dirac measure supported in shocks.*

4. DISCONTINUOUS GALERKIN FINITE ELEMENT METHOD AND ENTROPY-BASED ARTIFICIAL VISCOSITY APPROXIMATION FOR SOLVING HYPERBOLIC SCALAR CONSERVATION LAWS

4.1 Introduction

This chapter provides a detailed description of a discontinuous Galerkin spatial discretization technique with an embedded entropy-based artificial viscosity approximation to obtain stable and accurate numerical solutions to different types of hyperbolic scalar conservation equations. We will also discuss the procedure of explicit integration in time which we apply to spatially semi-discrete forms. The chapter ends up with considering one- and two-dimensional numerical tests which clearly demonstrate the satisfactory stability properties of the method and optimal convergence rates as well. The extension to compressible gas dynamics equations will be discussed later in Chapter V.

4.2 Preface

The hyperbolic scalar conservation equations themselves are frequently used to model a number of physically important phenomena, as for example inviscid nonlinear Burgers' equation - probably the simplest model equation which has been widely used to study shock waves propagation, acoustic transmission, and traffic flow. The reader is referred to Fletcher [48] for some of the phenomena that can be exactly or approximately modeled by this equation. Beyond this fact these relatively simple equations can also inherit some particular nonlinear features of the general systems of equations - and therefore can be considered as a good experimental platform for development, implementation, and testing of the numerical schemes for systems of hyperbolic conservation laws.

To make the last statement more clear, let us consider the system of compressible Euler equations in one space dimension:

$$\begin{cases} \partial_t \rho + \partial_x (\rho u) = 0, \\ \partial_t (\rho u) + \partial_x (\rho u^2 + p) = 0, \\ \partial_t E + \partial_x (u (E + p)) = 0. \end{cases} \tag{4.1}$$

Let us now make the first assumption that the temperature of the flow is constant, i.e. $T = T_0$. Therefore, the Equation of State becomes $p = R\rho T_0$ and we do not need the energy conservation equation. In this case, we will be left with the mass and momentum conservation equations only:

$$\begin{cases} \partial_t \rho + \partial_x (\rho u) = 0, \\ \partial_t (\rho u) + \partial_x (\rho u^2 + p) = 0, \\ p = R\rho T_0. \end{cases} \tag{4.2}$$

Expanding the momentum conservation equation, we can write:

$$u\partial_t \rho + \rho \partial_t u + u\partial_x (\rho u) + \rho u \partial_x u + \partial_x p = 0. \tag{4.3}$$

Using the mass conservation equation, we can simplify the above equation as:

$$\partial_t u + u\partial_x u + \frac{1}{\rho}\partial_x p = 0. \tag{4.4}$$

To isolate the convection process from the acoustic signal propagation, let us now make the second assumption that the pressure is constant throughout the entire domain and so the pressure gradient term disappears, yielding:

$$\partial_t u + u\partial_x u = 0. \tag{4.5}$$

The same equation in conservative form reads:

$$\partial_t u + \partial_x \left(\frac{u^2}{2} \right) = 0, \tag{4.6}$$

which is an inviscid nonlinear Burgers' equation in one space dimension. The equation (4.6) captures the non-linearity of the convection terms of the original system (4.1), and is often used to test the numerical schemes before proceeding with the system of compressible Euler equations. Note that in multi-dimensional case the equation (4.6) becomes:

$$\partial_t u + \nabla \cdot \left(\frac{u^2}{2} \mathbf{e} \right) = 0, \tag{4.7}$$

where the components e_i of the vector $\mathbf{e}$ are defined as $\{e_i\}_{i=1}^{d} = 1$.

Another potentially useful equation which we employ to test the linear propagation of both smooth and discontinuous initial conditions is the linear transport equation, which in one space dimension is defined as follows:

$$\partial_t u + c \partial_x u = 0 \tag{4.8}$$

or

$$\partial_t u + \partial_x \left(cu \right) = 0, \tag{4.9}$$

where c is some constant speed of propagation. Note that in multi-dimensional case the equation (4.9) becomes:

$$\partial_t u + \nabla \cdot \left(\boldsymbol{\beta} u \right) = 0, \tag{4.10}$$

where $\boldsymbol{\beta} = \boldsymbol{\beta} \left(\mathbf{x} \right)$ is a smooth vector field.

Both Burgers' and linear transport equations may be solved analytically, which makes them a valuable tool to measure the convergence properties of the numerical methods.

4.3 Model problem

We consider the DG method for spatial discretization of the following model problem:

$$\begin{cases} \partial_t u + \nabla \cdot \mathbf{f}(u) = 0, & (\mathbf{x}, t) \in \Omega \times (0, T], \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}), & \mathbf{x} \in \Omega. \end{cases} \tag{4.11}$$

where $u = u(\mathbf{x}, t) : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}$ is the solution, $\mathbf{f}(u) : \mathbb{R} \to \mathbb{R}^d$ is a known differentiable function of the solution u called the inviscid flux, $u_0(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ is the initial data which is also known, $\Omega \subset \mathbb{R}^d$ is an open connected domain with the boundary Γ, and $(0, T] \in \mathbb{R}_+$ is a time interval.

This Initial Boundary Value Problem for the multi-dimensional hyperbolic scalar conservation law (2.12) has been previously introduced in Chapter II, see equation (2.14), and is supposed to be equipped with the appropriate boundary conditions on Γ. For the sake of simplicity we assume that an appropriate boundary function g is prescribed on the boundary Γ. That can be done either by using the periodic boundary conditions or by assuming that the initial data $u_0(\mathbf{x})$ is compactly supported and we are interested in the solution u before the domain of dependence of $u_0(\mathbf{x})$ reaches the boundary of the domain Ω.

We regularize the IBVP (4.11) as follows:

$$\begin{cases} \partial_t u + \nabla \cdot \mathbf{f}(u) = -\nabla \cdot \mathbf{q}, & (\mathbf{x}, t) \in \Omega \times (0, T], \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}), & \mathbf{x} \in \Omega. \end{cases} \tag{4.12}$$

where $\mathbf{q} \equiv \mathbf{q}(\mu, \nabla u) = -\mu \nabla u$ is the viscous flux and $\mu \equiv \mu(u)$ is the artificial viscosity whose construction will be later based on the residual of the following entropy inequality:

$$\partial_t \eta(u) + \nabla \cdot \boldsymbol{\psi}(u) \leq 0, \tag{4.13}$$

where $(\eta(u), \boldsymbol{\psi}(u))$ is a given entropy pair.

4.4 Discretization in space

For practical computations we first need to discretize the regularized model problem (4.12) in space. For this purpose, we consider the DG finite element discretization. We begin by introducing a mesh.

4.4.1 Mesh and some notation

We discretize the domain of interest Ω into the set $\mathbb{T}_h$ of disjoint elements K such that $\bigcup_{K \in \mathbb{T}_h} \bar{K} = \bar{\Omega}$. Here, $\bar{K} = K \cup \partial K$ where ∂K is the boundary of K, $\bar{\Omega} = \Omega \cup \Gamma$, and h denotes the piecewise constant mesh function defined by $h\,|_K \equiv h_K = \mathrm{diam}(K)$ for all $K \in \mathbb{T}_h$. An element K is assumed to be either a polygon (we use quadrilaterals) in two space dimensions or a polyhedron (we use hexahedrons) in three space dimensions. We assume the shape regularity of the mesh $\mathbb{T}_h$, meaning that if ρ_K is the diameter of the largest ball inscribed in K, then the ratio $\max_{K \in \mathbb{T}_h} \dfrac{h_K}{\rho_K}$ is finite, i.e. the elements are not too flat. For all $K \in \mathbb{T}_h$ the collection of elements in $\mathbb{T}_h$ adjacent to K is denoted by Δ_K. We assume that the mesh $\mathbb{T}_h$ is locally quasi-uniform in the sense that the quantity $\max_{K \in \mathbb{T}_h} \dfrac{h_K}{\min_{K' \in \Delta_K} h_{K'}}$ is finite, i.e. all the elements adjacent to K have a diameter of the order h_K.

Let us suppose that each $K \in \mathbb{T}_h$ is an image of a fixed reference element $\hat{K}$, that is, $K = \sigma_K(\hat{K})$ for all $K \in \mathbb{T}_h$. Here, we only consider the case when $\hat{K}$ is the open unit hypercube in $\mathbb{R}^d$. The mapping σ_K of the reference element $\hat{K}$ to the real element $K \in \mathbb{T}_h$ is assumed to be bijective and smooth. For the elements K which

are located in the interior of the domain Ω, such that the intersection $\partial K \cap \Gamma = \emptyset$, the mapping σ_K is represented by a d-linear function. For the elements on the boundary Γ, such that the intersection $\partial K \cap \Gamma \neq \emptyset$, the mapping relied on the polynomials of higher degree might be required.

On the reference element $\hat{K}$ we define the space $\hat{Q}_p$ of polynomials of degree $p \geq 0$ as follows:

$$\hat{Q}_p = \text{span}\left\{ \prod_{i=1}^{d} \hat{x}_i^{\omega_i} : 0 \leq \omega_i \leq p \right\}. \tag{4.14}$$

Any polynomial function $\hat{q} \equiv \hat{q}(\hat{\mathbf{x}}) \in \hat{Q}_p$ of the degree $p \geq 0$ can be represented on the reference element $\hat{K}$ through the values $\hat{q}_i \equiv \hat{q}(\hat{\mathbf{x}}_i)$ at the support points $\hat{\mathbf{x}}_i \in \hat{K}$ and the respective basis functions $\hat{\Phi}_i(\hat{\mathbf{x}})$ associated with each of these support points:

$$\hat{q} = \sum_{i=1}^{N_s} \hat{q}_i \hat{\Phi}_i(\hat{\mathbf{x}}), \tag{4.15}$$

where $N_s = (p+1)^d$ is the number of support points. The concrete appearance of $\left\{\hat{\Phi}_i(\hat{\mathbf{x}})\right\}_{i=1}^{N_s}$ depends on the number of dimensions of the reference element $\hat{K}$ as well as the polynomial degree p.

Now, we introduce the finite element space $\mathcal{V}_h^p$ consisting of discontinuous scalar-valued polynomial functions v_h of degree $p \geq 0$:

$$\mathcal{V}_h^p = \left\{ v_h \in L^2(\Omega) : v_{h,K} \circ \sigma_K \in \hat{Q}_p \right\}. \tag{4.16}$$

For each element $K \in \mathbb{T}_h$ the following representation holds:

$$v_h = \sum_{i=1}^{N_s} v_{h,i} \Phi_i(\mathbf{x}), \tag{4.17}$$

where $v_{h,i} \equiv v_h(\mathbf{x}_i)$ are the values of v_h at the support points $\mathbf{x}_i \in K$ and $\Phi_i(\mathbf{x})$ are the associated basis functions which are compactly supported on K.

Note that:

$$v_h = \sum_{i=1}^{N_s} v_{h,i}\Phi_i(\mathbf{x}) = \sum_{i=1}^{N_s} v_{h,i}\Phi_i(\sigma_K(\hat{\mathbf{x}})) = \sum_{i=1}^{N_s} v_{h,i}\hat{\Phi}_i(\hat{\mathbf{x}}). \tag{4.18}$$

In particular, a d-linear mapping of the reference element $\hat{K}$ to the real element $K \in \mathbb{T}_h$ can be defined as follows:

$$\mathbf{x} = \sum_{v=1}^{N_v} \mathbf{x}_v \hat{\Phi}_v(\hat{\mathbf{x}}), \tag{4.19}$$

where v is the index of the mapped vertex, $N_v = 2^d$ is the number of vertices on the d-dimensional hypercube, and $\mathbf{x}_v$ is the coordinate vector of the vertex v on the real element $K \in \mathbb{T}_h$.

For each $K \in \mathbb{T}_h$ and each $\mathbf{x} \in \partial K$ we denote by $v^+_{h,\partial K}$ the interior trace of v_h on ∂K and by $v^-_{h,\partial K}$ the exterior trace of v_h on ∂K. Actually, the inner trace $v^+_{h,\partial K}$ of v_h on ∂K is taken from within the element K. And the outer trace $v^-_{h,\partial K}$ of v_h on ∂K is defined from within the neighboring element K'. If $\mathbf{x} \in \partial K \cap \Gamma$, then $v^-_{h,\partial K} = g$ with g be an appropriate boundary function.

Figure 4.1 shows some examples of meshes which we used in our computations, and biquadratic (Q_2) quadrilateral elements with the support points are depicted in Figure 4.2.

4.4.2 Standard DG formulation

To devise the discontinuous Galerkin method, we first derive a weak formulation of the regularized model problem (4.12). We multiply the first equation of the regularized IBVP (4.12) by an arbitrary smooth weight function v and then integrate the result over an arbitrary element $K \in \mathbb{T}_h$:

$$\int_K v\partial_t u d\mathbf{x} + \int_K v\nabla \cdot \mathbf{f}(u) d\mathbf{x} + \int_K v\nabla \cdot \mathbf{q} d\mathbf{x} = 0. \tag{4.20}$$

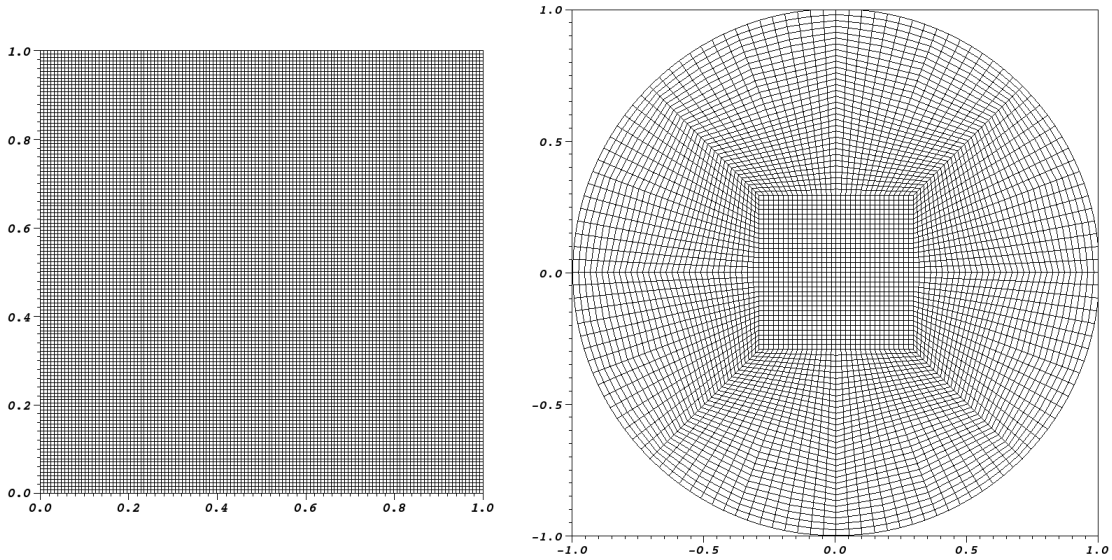

Fig. 4.1. Square and circular meshes.

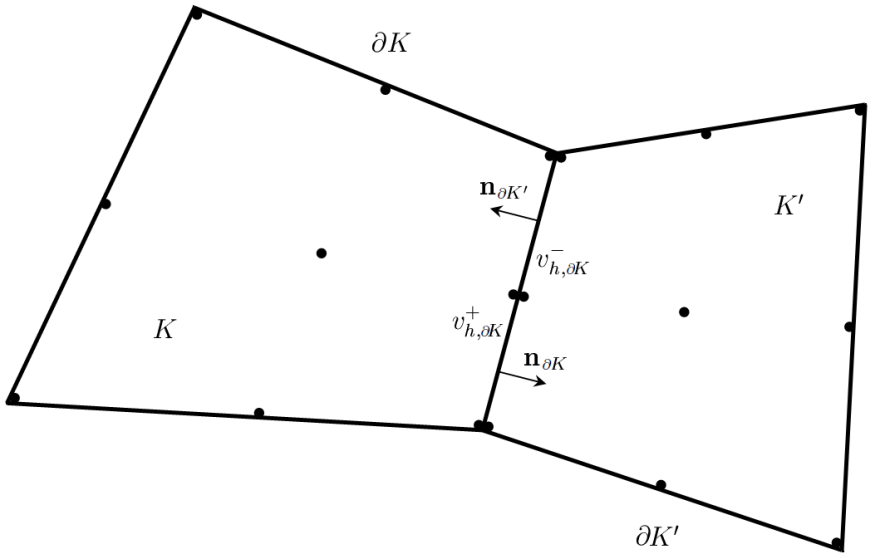

Fig. 4.2. Q_2 quadrilateral elements.

The integration by parts yields:

$$\int_K v\partial_t u d\mathbf{x} + \int_K \nabla\cdot\left(v\mathbf{f}(u)\right) d\mathbf{x} - \int_K \nabla v\cdot\mathbf{f}(u) d\mathbf{x} + \int_K \nabla\cdot\left(v\mathbf{q}\right) d\mathbf{x} - \int_K \nabla v\cdot\mathbf{q} d\mathbf{x} = 0. \tag{4.21}$$

The application of Gauss-Ostrogradski formula gives:

$$\int_K v\partial_t u d\mathbf{x} + \oint_{\partial K} v\mathbf{f}(u)\cdot\mathbf{n}_{\partial K} d\sigma - \int_K \nabla v\cdot\mathbf{f}(u) d\mathbf{x} + \oint_{\partial K} v\mathbf{q}\cdot\mathbf{n}_{\partial K} d\sigma - \int_K \nabla v\cdot\mathbf{q} d\mathbf{x} = 0, \tag{4.22}$$

where $\mathbf{n}_{\partial K}$ is an outward unit normal vector to the boundary ∂K. To obtain a discrete form of the weak formulation (4.22), we replace the smooth weight function v with its analogue $v_h \in \mathcal{V}_h^p$ and the analytical solution u with a numerical solution $u_h \in \mathcal{V}_h^p$:

$$\int_K v_h\partial_t u_h d\mathbf{x} + \oint_{\partial K} v_h\mathbf{f}(u_h)\cdot\mathbf{n}_{\partial K} d\sigma - \int_K \nabla v_h\cdot\mathbf{f}(u_h) d\mathbf{x} + \oint_{\partial K} v_h\mathbf{q}_h\cdot\mathbf{n}_{\partial K} d\sigma - \int_K \nabla v_h\cdot\mathbf{q}_h d\mathbf{x} = 0, \tag{4.23}$$

where $\mathbf{q}_h \equiv \mathbf{q}\left(\mu_h, \nabla u_h\right) = -\mu_h\nabla u_h$ and $\mu_h = \mu(u_h)$. Since the numerical solution u_h is allowed to be discontinuous across the boundaries of all elements, the dot products $\mathbf{f}(u_h)\cdot\mathbf{n}_{\partial K}$ and $\mathbf{q}_h\cdot\mathbf{n}_{\partial K}$ are not uniquely defined on ∂K. Hence, we uniquely define them by replacing the physical flux functions $\mathbf{f}(u_h)\cdot\mathbf{n}_{\partial K}$ and $\mathbf{q}_h\cdot\mathbf{n}_{\partial K}$ by the numerical flux functions $\hat{F}_{h,\partial K}$ and $\hat{Q}_{h,\partial K}$, respectively.

Eventually, we formulate a very compact numerical scheme of an arbitrarily high order of accuracy for which the numerical solution $u_h \in \mathcal{V}_h^p$ within each element $K \in \mathbb{T}_h$ is kept independent of the numerical solutions in other elements from $\mathbb{T}_h$, with interelement communication occurring only via numerical flux functions at the boundaries of the adjacent elements. This scheme is called the discontinuous Galerkin

finite element method of degree p, or DG(p) method in short, and can be formulated as follows:

$$\forall K \in \mathbb{T}_h \text{ find } u_h \in \mathcal{V}_h^p \text{ such that } \forall v_h \in \mathcal{V}_h^p :$$

$$\int_K v_h \partial_t u_h d\mathbf{x} + \oint_{\partial K} v_h \hat{F}_{h,\partial K} d\sigma - \int_K \nabla v_h \cdot \mathbf{f}(u_h) d\mathbf{x} + \oint_{\partial K} v_h \hat{Q}_{h,\partial K} d\sigma - \int_K \nabla v_h \cdot \mathbf{q}_h d\mathbf{x} = 0. \tag{4.24}$$

The inviscid numerical flux $\hat{F}_{h,\partial K}$ basically depends on the inner trace $u^+_{h,\partial K}$ of u_h on ∂K and the outer trace $u^-_{h,\partial K}$ of u_h on ∂K as well as the outward unit normal vector $\mathbf{n}_{\partial K}$ to the boundary ∂K:

$$\hat{F}_{h,\partial K} = \hat{F}_{h,\partial K}\left(u^+_{h,\partial K}, u^-_{h,\partial K}, \mathbf{n}_{\partial K}\right). \tag{4.25}$$

The viscous numerical flux $\hat{Q}_{h,\partial K}$, in addition to that, also depends on $\mu^+_{h,\partial K}\nabla u^+_{h,\partial K}$ and $\mu^-_{h,\partial K}\nabla u^-_{h,\partial K}$:

$$\hat{Q}_{h,\partial K} = \hat{Q}_{h,\partial K}\left(u^+_{h,\partial K}, u^-_{h,\partial K}, \mu^+_{h,\partial K}\nabla u^+_{h,\partial K}, \mu^-_{h,\partial K}\nabla u^-_{h,\partial K}, \mathbf{n}_{\partial K}\right). \tag{4.26}$$

4.4.3 Numerical fluxes

Inviscid fluxes :

The inviscid numerical flux $\hat{F}_{h,\partial K}(\,\cdot\,,\,\cdot\,,\,\cdot\,)$ must be both consistent with the respective physical flux $\mathbf{f}(\,\cdot\,)\cdot\mathbf{n}_{\partial K}$ and conservative. It means that:

- $\forall K \in \mathbb{T}_h$ the inviscid numerical flux $\hat{F}_{h,\partial K}(\,\cdot\,,\,\cdot\,,\,\cdot\,)$ is consistent with the physical flux $\mathbf{f}(\,\cdot\,)\cdot\mathbf{n}_{\partial K}$, meaning that:

$$\hat{F}_{h,\partial K}(v, v, \mathbf{n}_{\partial K}) = \mathbf{f}(v)\cdot\mathbf{n}_{\partial K}. \tag{4.27}$$

- $\forall K \in \mathbb{T}_h$ the inviscid numerical flux $\hat{F}_{h,\partial K}(\,\cdot\,,\,\cdot\,,\,\cdot\,)$ is conservative, i.e. given any two adjacent elements $K \in \mathbb{T}_h$ and $K' \in \Delta_K \subset \mathbb{T}_h$, at each point $\mathbf{x} \in \partial K \cap \partial K'$, taking into account that $\mathbf{n}_{\partial K'} = -\mathbf{n}_{\partial K}$, we have that:

$$\hat{F}_{h,\partial K}(u, v, \mathbf{n}_{\partial K}) = -\hat{F}_{h,\partial K}(v, u, -\mathbf{n}_{\partial K}). \tag{4.28}$$

There exist several types of inviscid numerical fluxes satisfying these two conditions. Among them - the Godunov, Osher, Lax-Friedrichs, Roe numerical fluxes [49] and others. As an example, here we consider two similar inviscid numerical fluxes, namely the Lax-Friedrichs numerical flux and JLG numerical flux [50].

The Lax-Friedrichs numerical flux $\hat{F}^{\mathrm{LF}}_{h,\partial K}$ has been previously introduced in the framework of DG(0) approximation, see Chapter III and equations (3.74) and (3.75), and is now defined as follows:

$$\hat{F}^{\mathrm{LF}}_{h,\partial K} = \frac{1}{2}\left(\mathbf{f}(u^+_{h,\partial K}) + \mathbf{f}(u^-_{h,\partial K})\right) \cdot \mathbf{n}_{\partial K} + \omega\alpha_{\partial K}\left(u^+_{h,\partial K} - u^-_{h,\partial K}\right) \tag{4.29}$$

or

$$\hat{F}^{\mathrm{LF}}_{h,\partial K} = \left\{\mathbf{f}(u_h)\right\}_{\partial K} \cdot \mathbf{n}_{\partial K} + \omega\alpha_{\partial K}\left[\!\left[u_h\right]\!\right]_{\partial K}. \tag{4.30}$$

The JLG numerical flux $\hat{F}^{\mathrm{JLG}}_{h,\partial K}$ can be written in the following form:

$$\hat{F}^{\mathrm{JLG}}_{h,\partial K} = \mathbf{f}\left(\frac{u^+_{h,\partial K} + u^-_{h,\partial K}}{2}\right) \cdot \mathbf{n}_{\partial K} + \omega\alpha_{\partial K}\left(u^+_{h,\partial K} - u^-_{h,\partial K}\right) \tag{4.31}$$

or

$$\hat{F}^{\mathrm{JLG}}_{h,\partial K} = \mathbf{f}\left(\left\{u_h\right\}_{\partial K}\right) \cdot \mathbf{n}_{\partial K} + \omega\alpha_{\partial K}\left[\!\left[u_h\right]\!\right]_{\partial K}. \tag{4.32}$$

Here, we use $\left\{\,\cdot\,\right\}_{\partial K}$ to denote the average on ∂K which serves for consistency purposes and $[\![\,\cdot\,]\!]_{\partial K}$ to denote the jump on ∂K which is added for stabilization of both inviscid numerical fluxes. The parameter ω is typically 0 or $\frac{1}{2}$ and the value

of $\alpha_{\partial K}$ is defined as the maximum propagation speed associated with the whole boundary ∂K of the element K:

$$\alpha_{\partial K} = \| \ |\mathbf{f}'(u_h) \cdot \mathbf{n}_{\partial K}| \ \|_{L^\infty(K \cup \Delta_K)}, \tag{4.33}$$

where $\| \cdot \|_{L^\infty(K \cup \Delta_K)}$ means the maximum on the element K and all elements adjacent to K, see Figure 4.3.

Remark 4.4.1. *The definition* (4.33) *of* $\alpha_{\partial K}$ *can be slightly changed to associate only one value of* $\alpha_{\partial K}$ *with each* j*-th part of the boundary* ∂K*, see Figure 4.4:*

$$\alpha^j_{\partial K} = \| \ |\mathbf{f}'(u_h) \cdot \mathbf{n}_{\partial K}| \ \|_{L^\infty(K \cup K_j)}. \tag{4.34}$$

The both definitions are similar and we did not observe any significant difference in the results.

Remark 4.4.2. *We note that the dependence of both Lax-Friedrichs and JLG inviscid numerical fluxes upon* $\alpha_{\partial K}$ *makes them a function of not only the immediate internal* $u^+_{h,\partial K}$ *and external* $u^-_{h,\partial K}$ *states of the numerical solution* u_h*, but implies the local dependence within some neighborhood of the element* K*. However, for the sake of simplicity we hold the form* $\hat{F}_{h,\partial K}\left(u^+_{h,\partial K}, u^-_{h,\partial K}, \mathbf{n}_{\partial K}\right)$*.*

Remark 4.4.3. *Note that in the particular case of linear transport equation,* $\mathbf{f}(u) = \boldsymbol{\beta} u$*, using* $\omega = \dfrac{1}{2}$ *and* (4.33) *in either* (4.29)-(4.30) *or* (4.31)-(4.32) *is equivalent to using the upwind numerical flux, since:*

$$\begin{aligned} \hat{F}_{h,\partial K} = \quad & \frac{1}{2}\left(\boldsymbol{\beta} u^+_{h,\partial K} + \boldsymbol{\beta} u^-_{h,\partial K}\right) \cdot \mathbf{n}_{\partial K} + \frac{1}{2}|\boldsymbol{\beta} \cdot \mathbf{n}_{\partial K}|\left(u^+_{h,\partial K} - u^-_{h,\partial K}\right) = \\ & \begin{cases} u^+_{h,\partial K} \boldsymbol{\beta} \cdot \mathbf{n}_{\partial K} & \text{if} \quad \boldsymbol{\beta} \cdot \mathbf{n}_{\partial K} \geq 0, \\ u^-_{h,\partial K} \boldsymbol{\beta} \cdot \mathbf{n}_{\partial K} & \text{if} \quad \boldsymbol{\beta} \cdot \mathbf{n}_{\partial K} \leq 0. \end{cases} \end{aligned} \tag{4.35}$$

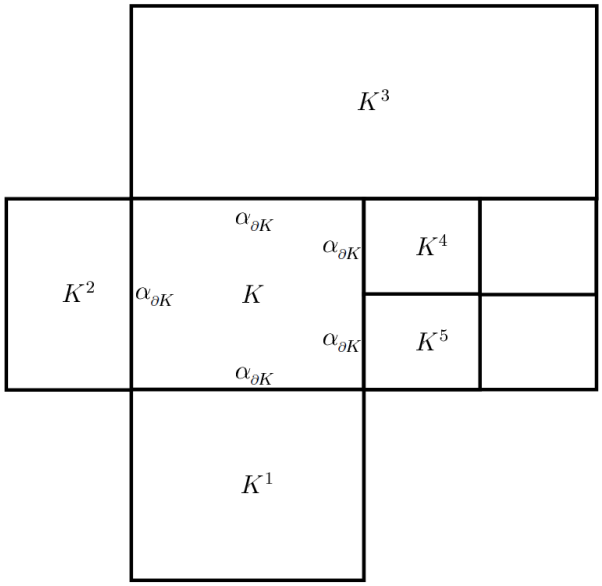

Fig. 4.3. Stencil for $\alpha_{\partial K}$.

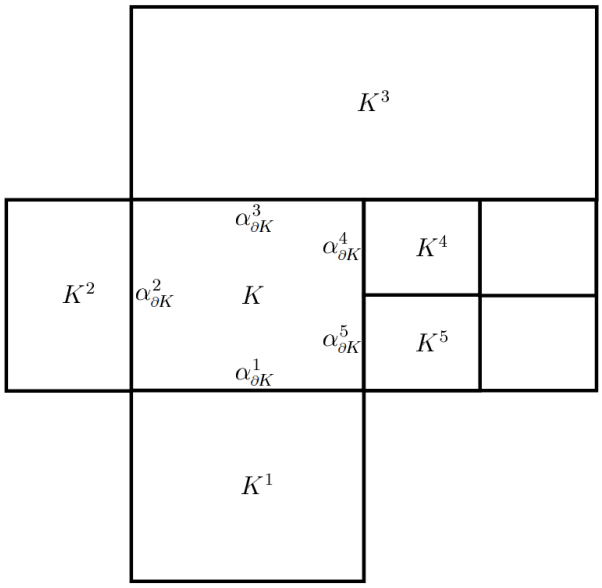

Fig. 4.4. Stencil for $\alpha^j_{\partial K}$.

Viscous fluxes :

In a spirit of Chapter III, see equations (3.66) and (3.67), we define the viscous numerical flux $\hat{Q}_{h,\partial K}$ as follows:

$$\begin{aligned}\hat{Q}_{h,\partial K} = \quad & \frac{1}{2}\left(\mathbf{q}\left(\mu^+_{h,\partial K}, \nabla u^+_{h,\partial K}\right) + \mathbf{q}\left(\mu^-_{h,\partial K}, \nabla u^-_{h,\partial K}\right)\right) \cdot \mathbf{n}_{\partial K} + \\ & \delta\beta_{\partial K}\left(u^+_{h,\partial K} - u^-_{h,\partial K}\right)\end{aligned} \tag{4.36}$$

or

$$\hat{Q}_{h,\partial K} = \left\{\mathbf{q}(\mu_h, \nabla u_h)\right\}_{\partial K} \cdot \mathbf{n}_{\partial K} + \delta\beta_{\partial K}\left[\!\left[u_h\right]\!\right]_{\partial K}. \tag{4.37}$$

The parameter δ is typically 0 or 1 and the value of $\beta_{\partial K}$ is defined as the maximum relative artificial viscosity associated with the whole boundary ∂K of the element K:

$$\beta_{\partial K} = \left\| \frac{\mu_h}{h} \right\|_{L^\infty(K\cup\Delta_K)}. \tag{4.38}$$

Remark 4.4.4. *Similar to the formulation* (4.34) *for* $\alpha^j_{\partial K}$*, the definition* (4.38) *can be also slightly changed as:*

$$\beta^j_{\partial K} = \left\| \frac{\mu_h}{h} \right\|_{L^\infty(K\cup K_j)}. \tag{4.39}$$

Note that the viscous numerical flux $\hat{Q}_{h,\partial K}$ defined by the equations (4.36) or (4.37) is both consistent with the respective physical flux $\mathbf{q}_h \cdot \mathbf{n}_{\partial K}$ and conservative.

Remark 4.4.5. *We note that the discontinuous Galerkin finite element methods are similar to finite volume schemes, especially in the use of numerical fluxes. In fact, the basic finite volume schemes exactly correspond to the DG(0) method, i.e. to the discontinuous Galerkin method using a piecewise constant representation of the numerical solution* u_h*. Consequently, the DG(p) methods with* $p > 0$ *can be treated as the natural generalization of finite volume schemes to methods of higher order of accuracy.*

4.5 Discretization in time

In this section we will describe the time integration procedure which we apply to the discontinuous Galerkin finite element discretization (4.24) of the regularized model problem (4.12). For the time discretization of (4.24) we first introduce some notation. By $0 < t^0 < t^1 < t^2 < ... < t^N = T$ we define a partition of the time interval $(0, T]$ with time stepping $\tau_n = t^{n+1} - t^n$ where $n = 0, ..., N-1$. By $u_h^n \in \mathcal{V}_h^p$ we denote a discrete function $u_h \in \mathcal{V}_h^p$ at time t^n.

If $\zeta_K \in \mathbb{N}_+$ is a local numbering of degrees of freedom within an element $K \in \mathbb{T}_h$, then a weight function $v_h \in \mathcal{V}_h^p$ can be defined in the following form:

$$v_h = \sum_{i \in \zeta_K} v_{h,i} \Phi_i(\mathbf{x}). \tag{4.40}$$

Since $\left\{v_{h,i}\right\}_{i \in \zeta_K} \in \mathbb{R}$ are arbitrarily chosen, we obtain the following system of PDEs:

$$\forall K \in \mathbb{T}_h \text{ find } u_h \in \mathcal{V}_h^p \text{ such that } \forall i \in \zeta_K :$$

$$\begin{aligned} &\int_K \Phi_i(\mathbf{x}) \partial_t u_h d\mathbf{x} + \oint_{\partial K} \Phi_i(\mathbf{x}) \hat{F}_{h,\partial K} d\sigma - \int_K \nabla \Phi_i(\mathbf{x}) \cdot \mathbf{f}(u_h) d\mathbf{x} + \\ &\oint_{\partial K} \Phi_i(\mathbf{x}) \hat{Q}_{h,\partial K} d\sigma - \int_K \nabla \Phi_i(\mathbf{x}) \cdot \mathbf{q}_h d\mathbf{x} = 0. \end{aligned} \tag{4.41}$$

In fact, the basis functions $\left\{\Phi_i(\mathbf{x})\right\}_{i \in \zeta_K} \in \mathcal{V}_h^p$ form a complete set of functions in the sense that any polynomial form $a_h \in \mathcal{V}_h^p$ can be represented as a linear combination of these functions. Thereby, for each element $K \in \mathbb{T}_h$ the equations (4.41) are simply equivalent to the requirement of the regularized PDE residual minimization by making the projection of this residual onto all basis functions be equal to 0. Taking into account that for each $K \in \mathbb{T}_h$:

$$u_h = \sum_{j \in \zeta_K} u_{h,j}(t) \Phi_j(\mathbf{x}), \tag{4.42}$$

we formulate the following system of *Ordinary Differential Equations* (ODE) with respect to the nodal unknown values $\left\{\left\{u_{h,j}(t)\right\}_{j\in\zeta_K}\right\}_{K\in\mathbb{T}_h}$:

$$\forall K \in \mathbb{T}_h \text{ find } \left\{u_{h,j}(t)\right\}_{j\in\zeta_K} \text{ such that } \forall i \in \zeta_K :$$

$$\begin{aligned}
&\sum_{j\in\zeta_K} \int_K \Phi_i(\mathbf{x})\Phi_j(\mathbf{x})d\mathbf{x} \cdot \frac{d}{dt}u_{h,j}(t) = \\
&\int_K \nabla\Phi_i(\mathbf{x}) \cdot \mathbf{f}(u_h)d\mathbf{x} + \int_K \nabla\Phi_i(\mathbf{x}) \cdot \mathbf{q}_h d\mathbf{x} - \\
&\oint_{\partial K} \Phi_i(\mathbf{x})\hat{F}_{h,\partial K}d\sigma - \oint_{\partial K} \Phi_i(\mathbf{x})\hat{Q}_{h,\partial K}d\sigma.
\end{aligned} \tag{4.43}$$

The system (4.43) can be written globally on the whole triangulation $\mathbb{T}_h$ in the matrix-vector form as follows:

$$\mathbb{M}\frac{d}{dt}\mathbf{u}_h = \mathbf{R}_h, \tag{4.44}$$

where $\mathbb{M}$ is the mass matrix of the system, $\mathbf{u}_h$ is the time-dependent vector of all nodal unknowns, and $\mathbf{R}_h$ is the vector of all right hand sides of the equations (4.43). It is clear that $\mathbf{R}_h$ is a functional of the numerical solution, its derivatives, and the artificial viscosity as well, and does not explicitly depend on time. However, for a brief discussion on explicit Runge-Kutta methods we will use one of the following two forms: $\mathbf{R}_h\left(t, \mathbf{u}_h\right)$ or $\mathbf{R}_h\left(t, \mathbf{u}_h, \boldsymbol{\mu}_h\right)$.

Since for each $K \in \mathbb{T}_h$ the basis functions $\left\{\Phi_i(\mathbf{x})\right\}_{i\in\zeta_K} \in \mathcal{V}_h^p$ are compactly supported on K, the mass matrix $\mathbb{M}$ of the system of ODEs is a block-diagonal matrix consisting of blocks $\mathbb{M}_K$ corresponding to respective $K \in \mathbb{T}_h$. The entries of such blocks are defined by:

$$M_{K,ij} = \int_K \Phi_i(\mathbf{x})\Phi_j(\mathbf{x})d\mathbf{x}. \tag{4.45}$$

Each of these blocks can be inverted by using, for example, the Gauss–Jordan elimination [51]. If the mesh $\mathbb{T}_h$ is globally uniform, then all the blocks $\mathbb{M}_K$ have exactly

the same content. The structures of the mass matrix $\mathbb{M}$ and the inverse mass matrix $\mathbb{M}^{-1}$ are schematically shown below and hold for any fully explicit time integration scheme:

$$\mathbb{M} = \begin{bmatrix} \mathbb{M}_1 & 0 & \dots & 0 \\ 0 & \mathbb{M}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbb{M}_{N_K} \end{bmatrix} \quad \mathbb{M}^{-1} = \begin{bmatrix} \mathbb{M}_1^{-1} & 0 & \dots & 0 \\ 0 & \mathbb{M}_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbb{M}_{N_K}^{-1} \end{bmatrix} \tag{4.46}$$

where N_K is the number of elements forming the triangulation $\mathbb{T}_h$.

To discretize the system of ODEs (4.44) in time, we employ explicit Runge-Kutta algorithms. Generally, the explicit Runge-Kutta methods take the following form:

$$\mathbf{u}_h^{n+1} = \mathbf{u}_h^n + \tau_n \sum_{m=1}^{S} b_m \mathbf{k}_m, \tag{4.47}$$

where:

$$\begin{aligned} &\mathbf{k}_1 = \mathbb{M}^{-1}\mathbf{R}_h\,(t^n, \mathbf{u}_h^n\,), \\ &\mathbf{k}_2 = \mathbb{M}^{-1}\mathbf{R}_h\,(t^n + c_2\tau_n, \mathbf{u}_h^n + a_{21}\mathbf{k}_1\,), \\ &\mathbf{k}_3 = \mathbb{M}^{-1}\mathbf{R}_h\,(t^n + c_3\tau_n, \mathbf{u}_h^n + a_{31}\mathbf{k}_1 + a_{32}\mathbf{k}_2\,), \\ &... \\ &\mathbf{k}_S = \mathbb{M}^{-1}\mathbf{R}_h\,(t^n + c_S\tau_n, \mathbf{u}_h^n + a_{S1}\mathbf{k}_1 + a_{S2}\mathbf{k}_2 + a_{S3}\mathbf{k}_3 + ... + a_{S,S-1}\mathbf{k}_{S-1}\,). \end{aligned} \tag{4.48}$$

The methods of this class are each identified by the respective Butcher tableau (after John C. Butcher), which puts the coefficients of the method in a table as follows, see Table 4.1 below.

Table 4.1
General Butcher tableau.

0						
c_2	a_{21}					
c_3	a_{31}	a_{32}				
$.$	$.$	$.$	$.$			
c_S	a_{S1}	a_{S2}	a_{S3}	$...$	$a_{S,S-1}$	
	b_1	b_2	b_3	$...$	b_{S-1}	b_S

The estimation of a time step τ_n is based on the CFL condition. If the elements composing $\mathbb{T}_h$ are not too flat, then the following estimate is correct:

$$\tau_n = \min_{K \in \mathbb{T}_h} \overbrace{\left(\text{CFL} \, \frac{h_K}{|\mathbf{f}'(u_h^n)|_{L^\infty(K)}} \right)}^{\tau_{n,K}}. \tag{4.49}$$

For instance, the Butcher tableaux corresponding to the explicit Runge-Kutta methods RK3 and RK4 are presented as shown in Table 4.2.

Table 4.2
Explicit RK3 (left) and RK4 (right) Butcher tableaux.

0			
$\frac{1}{2}$	$\frac{1}{2}$		
1	-1	2	
	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

0				
$\frac{1}{2}$	$\frac{1}{2}$			
$\frac{1}{2}$	0	$\frac{1}{2}$		
1	0	0	1	
	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$

We have not noticed significant differences in behavior of the entropy viscosity methods when using either explicit RK3 or RK4, besides the fact that RK4 is more accurate than RK3 on problems with smooth solutions.

4.6 Inclusion of the entropy-based artificial viscosity

The right hand side of the system of ODEs in the form of equations (4.43) or (4.44) includes the viscous fluxes $\mathbf{q}_h$ and $\hat{Q}_{h,\partial K}$ with embedded and still unknown artificial viscosity μ_h. The construction of μ_h can be directly performed within the Runge-Kutta methods and shown schematically as follows:

- at the outset
  - Use $\mathbf{u}_h^{n-1}$ and $\mathbf{u}_h^n$ to find $\boldsymbol{\mu}_h^n$
- stage No. 1
  - $\mathbf{u}_h^1 = \mathbf{u}_h^n$
  - $\boldsymbol{\mu}_h^1 = \boldsymbol{\mu}_h^n$
  - $\mathbf{k}_1 = \mathbb{M}^{-1}\mathbf{R}_h\left(t^n, \mathbf{u}_h^1, \boldsymbol{\mu}_h^1\right)$
- stage No. 2
  - $\mathbf{u}_h^2 = \mathbf{u}_h^1 + a_{21}\mathbf{k}_1$
  - Use $\mathbf{u}_h^1$ and $\mathbf{u}_h^2$ to find $\boldsymbol{\mu}_h^2$
  - $\mathbf{k}_2 = \mathbb{M}^{-1}\mathbf{R}_h\left(t^n + c_2\tau_n, \mathbf{u}_h^2, \boldsymbol{\mu}_h^2\right)$
- stage No. 3
  - $\mathbf{u}_h^3 = \mathbf{u}_h^1 + a_{31}\mathbf{k}_1 + a_{32}\mathbf{k}_2$
  - Use $\mathbf{u}_h^1$ and $\mathbf{u}_h^3$ to find $\boldsymbol{\mu}_h^3$
  - $\mathbf{k}_3 = \mathbb{M}^{-1}\mathbf{R}_h\left(t^n + c_3\tau_n, \mathbf{u}_h^3, \boldsymbol{\mu}_h^3\right)$
- . . .
- stage No. S
  - $\mathbf{u}_h^S = \mathbf{u}_h^1 + a_{S1}\mathbf{k}_1 + a_{S2}\mathbf{k}_2 + a_{S3}\mathbf{k}_3 + ... + a_{S,S-1}\mathbf{k}_{S-1}$

  - Use $\mathbf{u}_h^1$ and $\mathbf{u}_h^S$ to find $\boldsymbol{\mu}_h^S$
  - $\mathbf{k}_S = \mathbb{M}^{-1}\mathbf{R}_h\left(t^n + c_S\tau_n, \mathbf{u}_h^S, \boldsymbol{\mu}_h^S\right)$

- at the end
  - $\mathbf{u}_h^{n+1} = \mathbf{u}_h^n + \tau_n \sum_{m=1}^{S} b_m \mathbf{k}_m$

In the present work, the entropy viscosity is defined as a piecewise constant function over the domain Ω, meaning that only one value of the entropy viscosity is assigned to each element $K \in \mathbb{T}_h$ at time $t^k \in (0, T]$, where the upper index k might mean either a time level n or a Runge-Kutta stage s between two time levels n and $n+1$.

Let us consider two different sets of a numerical solution $\mathbf{u}_h$, namely $\mathbf{u}_h^a$ and $\mathbf{u}_h^{b>a}$ at time t^a and $t^{b>a}$, respectively. For example, $a = n-1$ and $b = n$ or $a = 1$ and $b = s$ with $s = 2, ..., S$, see the scheme at the beginning of this section.

Then, based on the data $\mathbf{u}_h^a$ and $\mathbf{u}_h^b$, the construction of the entropy viscosity field $\boldsymbol{\mu}_h^b$ can be done in a few steps. So, for each element $K \in \mathbb{T}_h$ we do the following:

- Given an entropy pair $(\eta(u), \boldsymbol{\psi}(u))$, compute the entropy residual $D_{h,K,q}^b$ at all quadrature points $\mathbf{x}_q \in K$ at time t^b:

$$D_{h,K,q}^b = \partial_t \eta_{h,K,q}^b + \nabla \cdot \boldsymbol{\psi}_{h,K,q}^b. \tag{4.50}$$

  The equation (4.50) can be discretized in time by means of, for example, the second-order Crank-Nicholson scheme:

$$D_{h,K,q}^b = \frac{\eta_{h,K,q}^b - \eta_{h,K,q}^a}{t^b - t^a} + \frac{1}{2}\left(\nabla \cdot \boldsymbol{\psi}_{h,K,q}^a + \nabla \cdot \boldsymbol{\psi}_{h,K,q}^b\right). \tag{4.51}$$

- Compute the jump of the entropy flux $J_{h,\partial K,g}^b$ at all quadrature points $\mathbf{x}_g \in \partial K$ at time t^b:

$$J_{h,\partial K,g}^b = \left(\boldsymbol{\psi}_{h,\partial K,g}^{b,+} - \boldsymbol{\psi}_{h,\partial K,g}^{b,-}\right) \cdot \mathbf{n}_{\partial K,g}. \tag{4.52}$$

- Define the maximum, say $R^b_{h,K}$ associated with element K at time t^b:

$$R^b_{h,K} = \max_{q,g} \left(\, h_K \left| D^b_{h,K,q} \right| \, , \, \left| J^b_{h,\partial K,g} \right| \, \right). \tag{4.53}$$

- Use this maximum to define a viscosity, say $\mu^b_{h,K,E}$ associated with element K at time t^b:

$$\mu^b_{h,K,E} = C_E \, \frac{h_K}{\max\limits_{q \in K \in \mathbb{T}_h} \left| \eta^b_{h,K,q} - \bar{\eta}^b_h \right|} \, R^b_{h,K}, \tag{4.54}$$

  where C_E is a tunable constant, h_K is a diameter of element K, $\max\limits_{q \in K \in \mathbb{T}_h} |\cdot|$ is taken over all quadrature points that can be possibly found in the entire triangulation $\mathbb{T}_h$, and $\bar{\eta}^b_h = \dfrac{1}{\|\Omega\|} \displaystyle\int_\Omega \eta \left(u_h \left(\mathbf{x}, t^b \right) \right) d\mathbf{x}$ is the space-averaged value of the entropy computed at time t^b.

- Introduce an entropy viscosity upper bound $\mu^b_{h,K,\max}$ associated with element K at time t^b:

$$\mu^b_{h,K,\max} = C_{\max} h_K \max_{q \in K} \left| \mathbf{f}' \left(u^b_{h,K,q} \right) \right|. \tag{4.55}$$

- Define the entropy viscosity $\mu^b_{h,K}$ associated with element K at time t^b:

$$\mu^b_{h,K} = \min \left(\, \mu^b_{h,K,E} \, , \, \mu^b_{h,K,\max} \, \right). \tag{4.56}$$

- Use $\left\{ \mu^b_{h,K} \right\}_{K \in \mathbb{T}_h}$ along with other data to march in time.

Remark 4.6.1. *Note that $\mu^b_{h,K,E} \sim \mathcal{O}(h^4)$ is the higher-order artificial viscosity that we add to the regions where the solution is smooth, and $\mu^b_{h,K,max} \sim \mathcal{O}(h)$ is the first-order artificial viscosity which fits the shocks.*

Remark 4.6.2. *Instead of using Crank-Nicholson formula* (4.51)*, other explicit methods, for instance, BDF-like multistep methods might be employed.*

4.7 Numerical tests

We illustrate the capability of the above technique by solving various types of hyperbolic scalar conservation equations in one and two space dimensions. The two-dimensional problems are solved by using the deal.II finite element library.

Deal.II is a C++ program library targeted at the computational solution of partial differential equations using adaptive finite elements. The main aim of deal.II is to enable rapid development of modern finite element codes, using among other aspects adaptive meshes and a wide array of tools classes often used in finite element programs. For more details about the deal.II finite element library, see, for example, [52, 53].

4.7.1 One-dimensional tests

Linear transport equation with smooth data :
We consider the linear transport problem:

$$\begin{cases} \partial_t u + \partial_x u = 0 \ , \quad (x,t) \in \Omega \times (0,T], \\ u(x,0) = u_0(x) = \sin(2\pi x) \ , \quad x \in \Omega. \end{cases} \tag{4.57}$$

with periodic boundary conditions. The entropy pair $(\eta, \psi) = \left(\frac{u^2}{2}, \frac{u^2}{2}\right)$ and parameters of the computation are gathered in Table 4.3.

Table 4.3
1D linear transport equation with smooth data. Parameters.

Ω	$(0,1)$	ω	0.5
T	1.0	δ	0 or 1
RK	4	C_E	1.0
CFL	0.5	$C_{\max}$	$0.5/p$

The numerical solution is computed up to time $T = 1$ on various uniform meshes for polynomial degrees 1, 2, and 3. The errors measured in the L^1- and L^2-norms at time $T = 1$ are reported in Tables 4.4, 4.5, and 4.6. The observed convergence rates are 2, 3, and 4 for both norms, respectively. The method gives optimal convergence orders in one space dimension.

The graphs of the P_3 numerical solution with 100 elements and related entropy viscosity at time $T = 1$ are shown in Figure 4.5. Note that since the solution is completely smooth, then $\mu_E \sim C_E\mathcal{O}(h^4) \sim 10^{-10}$ everywhere, whereas $\mu_{\max} \sim C_{\max}\mathcal{O}(h) \sim 10^{-4}$. Therefore, the entropy viscosity field is exclusively composed of μ_E which is very close to 0. The maxima of the entropy viscosity graph correspond to the largest slopes on the graph of the numerical solution, and minima - to the zero slopes.

Table 4.4
1D linear transport equation with smooth data. Convergence test for P_1.

cells	dofs	h	L^1-error	R_1	L^2-error	R_2
5	10	2e-01	6.355e-01	-	6.951e-01	-
10	20	1e-01	5.325e-01	0.26	5.917e-01	0.23
20	40	5e-02	2.276e-01	1.23	2.573e-01	1.20
40	80	2.5e-02	3.725e-02	2.61	4.276e-02	2.59
80	160	1.25e-02	5.198e-03	2.84	6.023e-03	2.83
160	320	6.25e-03	5.155e-04	3.33	6.097e-04	3.30
320	640	3.125e-03	6.864e-05	2.91	8.338e-05	2.87
640	1280	1.5625e-03	9.741e-06	2.82	1.262e-05	2.73
1280	2560	7.8125e-04	1.631e-06	2.58	2.248e-06	2.49
2560	5120	3.90625e-04	3.435e-07	2.25	4.712e-07	2.25
5120	10240	1.953125e-04	8.068e-08	2.09	1.090e-07	2.11
10240	20480	9.765625e-05	1.977e-08	2.03	2.710e-08	2.01

Table 4.5
1D linear transport equation with smooth data. Convergence test for P_2.

cells	dofs	h	L^1-error	R_1	L^2-error	R_2
5	15	2e-01	5.570e-01	-	6.159e-01	-
10	30	1e-01	2.854e-01	0.97	3.217e-01	0.94
20	60	5e-02	1.771e-02	4.01	2.026e-02	3.99
40	120	2.5e-02	8.131e-04	4.45	8.920e-04	4.51
80	240	1.25e-02	7.221e-05	3.49	8.117e-05	3.46
160	480	6.25e-03	7.442e-06	3.28	8.367e-06	3.28
320	960	3.125e-03	8.357e-07	3.16	9.433e-07	3.15
640	1920	1.5625e-03	9.859e-08	3.08	1.117e-07	3.08
1280	3840	7.8125e-04	1.196e-08	3.04	1.359e-08	3.04
2560	7680	3.90625e-04	1.483e-09	3.01	1.687e-09	3.01

Table 4.6
1D linear transport equation with smooth data. Convergence test for P_3.

cells	dofs	h	L^1-error	R_1	L^2-error	R_2
5	20	2e-01	4.610e-01	-	5.121e-01	-
10	40	1e-01	2.405e-01	0.94	2.678e-01	0.94
20	80	5e-02	3.817e-02	2.66	4.319e-02	2.63
40	160	2.5e-02	5.457e-04	6.13	6.347e-04	6.09
80	320	1.25e-02	2.850e-05	4.26	3.247e-05	4.29
160	640	6.25e-03	1.496e-06	4.25	2.026e-06	4.00
320	1280	3.125e-03	9.021e-08	4.05	1.279e-07	3.98
640	2560	1.5625e-03	5.554e-09	4.02	8.050e-09	3.99
1280	5120	7.8125e-04	3.502e-10	3.99	5.036e-10	4.00

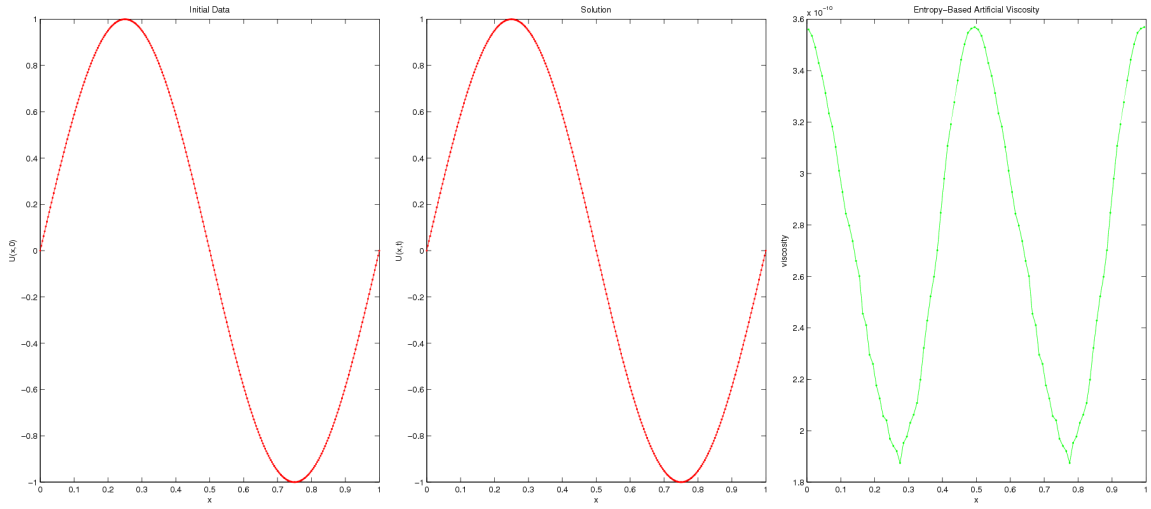

Fig. 4.5. 1D linear transport equation with smooth data. Initial data, P_3 solution, and entropy viscosity at time $T = 1$.

Linear transport equation with discontinuous data :

We now evaluate the performance of the method with nonsmooth data. We consider again the linear transport problem (4.57) with periodic boundary conditions, but this time the initial data is defined as follows:

$$u_0(x) = \begin{cases} 1 & \text{if} \quad 0.25 \leq x \leq 0.75, \\ 0 & \text{otherwise.} \end{cases} \tag{4.58}$$

The entropy pair $(\eta, \psi) = \left(\frac{u^2}{2}, \frac{u^2}{2}\right)$ and parameters of the computation are given in Table 4.7.

Table 4.7
1D linear transport equation with discontinuous data. Parameters.

Ω	$(0,1)$	ω	0.5
T	1.0	δ	0 or 1
RK	4	C_E	1.0
CFL	0.5	$C_{\max}$	$0.5/p$

The numerical solution is computed up to time $T = 1$ on various uniform meshes for polynomial degrees 1,...,5. The L^1- and L^2-norms of the error are computed and along with the respective convergence rates are reported in Tables 4.8, 4.9, 4.10, 4.11, and 4.12. As the polynomial degree p increases, the convergence rates in the L^1- and L^2-norms approach 1 and 0.5 and are compatible with the theoretical estimates $\frac{p+\frac{1}{2}}{p+1}$ and $\frac{1}{2}\frac{p+\frac{1}{2}}{p+1}$, respectively. The higher the polynomial degree the more accurate the method. This test shows that higher-order methods perform better than lower-order methods even on nonsmooth solutions. This statement is reinforced by looking at Figure 4.6, where we show the L^1-norm of the error at time $T = 1$ as a function of the total number of degrees of freedom for P_1,...,P_5 approximations. For any given number of degrees of freedom the error is a decreasing function of the polynomial

degree p. Note that the P_3, P_4, and P_5 errors almost coincide because the accuracy is limited to fourth-order by our using RK4.

Table 4.8

1D linear transport equation with discontinuous data. Convergence test for P_1.

cells	dofs	h	L^1-error	R_1	L^2-error	R_2
5	10	2e-01	4.709e-01	-	4.807e-01	-
10	20	1e-01	3.539e-01	0.41	4.009e-01	0.26
20	40	5e-02	1.860e-01	0.93	2.447e-01	0.71
40	80	2.5e-02	1.022e-01	0.87	1.757e-01	0.48
80	160	1.25e-02	6.142e-02	0.73	1.365e-01	0.36
160	320	6.25e-03	3.632e-02	0.76	1.048e-01	0.38
320	640	3.125e-03	2.142e-02	0.76	8.029e-02	0.38
640	1280	1.5625e-03	1.259e-02	0.77	6.132e-02	0.39
1280	2560	7.8125e-04	7.427e-03	0.76	4.711e-02	0.38

Table 4.9

1D linear transport equation with discontinuous data. Convergence test for P_2.

cells	dofs	h	L^1-error	R_1	L^2-error	R_2
5	15	2e-01	3.985e-01	-	4.550e-01	-
10	30	1e-01	2.041e-01	0.97	2.574e-01	0.82
20	60	5e-02	1.005e-01	1.02	1.759e-01	0.55
40	120	2.5e-02	5.693e-02	0.82	1.319e-01	0.42
80	240	1.25e-02	3.233e-02	0.82	9.911e-02	0.41
160	480	6.25e-03	1.846e-02	0.81	7.440e-02	0.41
320	960	3.125e-03	1.047e-02	0.82	5.557e-02	0.42
640	1920	1.5625e-03	5.916e-03	0.82	4.152e-02	0.42
1280	3840	7.8125e-04	3.351e-03	0.82	3.113e-02	0.42

Table 4.10
1D linear transport equation with discontinuous data. Convergence test for P_3.

cells	dofs	h	L^1-error	R_1	L^2-error	R_2
5	20	2e-01	2.570e-01	-	2.812e-01	-
10	40	1e-01	1.194e-01	1.11	1.717e-01	0.71
20	80	5e-02	6.217e-02	0.94	1.156e-01	0.57
40	160	2.5e-02	3.535e-02	0.82	8.407e-02	0.46
80	320	1.25e-02	1.941e-02	0.87	6.138e-02	0.45
160	640	6.25e-03	1.066e-02	0.87	4.515e-02	0.45
320	1280	3.125e-03	5.867e-03	0.86	3.338e-02	0.44
640	2560	1.5625e-03	3.239e-03	0.87	2.453e-02	0.45
1280	5120	7.8125e-04	1.770e-03	0.87	1.801e-02	0.45

Table 4.11
1D linear transport equation with discontinuous data. Convergence test for P_4.

cells	dofs	h	L^1-error	R_1	L^2-error	R_2
5	25	2e-01	2.297e-01	-	2.667e-01	-
10	50	1e-01	1.068e-01	1.11	1.550e-01	0.78
20	100	5e-02	5.683e-02	0.91	1.065e-01	0.54
40	200	2.5e-02	3.211e-02	0.82	7.702e-02	0.47
80	400	1.25e-02	1.706e-02	0.91	5.672e-02	0.45
160	800	6.25e-03	9.095e-03	0.91	4.092e-02	0.47
320	1600	3.125e-03	4.827e-03	0.91	2.951e-02	0.47
640	3200	1.5625e-03	2.580e-03	0.90	2.131e-02	0.47
1280	5400	7.8125e-04	1.373e-03	0.91	1.548e-02	0.46

Table 4.12
1D linear transport equation with discontinuous data. Convergence test for P_5.

cells	dofs	h	L^1-error	R_1	L^2-error	R_2
5	30	2e-01	2.001e-01	-	2.518e-01	-
10	60	1e-01	9.117e-02	1.13	1.430e-01	0.82
20	120	5e-02	4.734e-02	0.95	1.002e-01	0.51
40	240	2.5e-02	2.465e-02	0.94	7.213e-02	0.47
80	480	1.25e-02	1.290e-02	0.93	5.258e-02	0.46
160	960	6.25e-03	6.747e-03	0.94	3.766e-02	0.48
320	1920	3.125e-03	3.514e-03	0.94	2.719e-02	0.47
640	3840	1.5625e-03	1.831e-03	0.94	1.954e-02	0.48
1280	7680	7.8125e-04	9.559e-04	0.94	1.404e-02	0.48

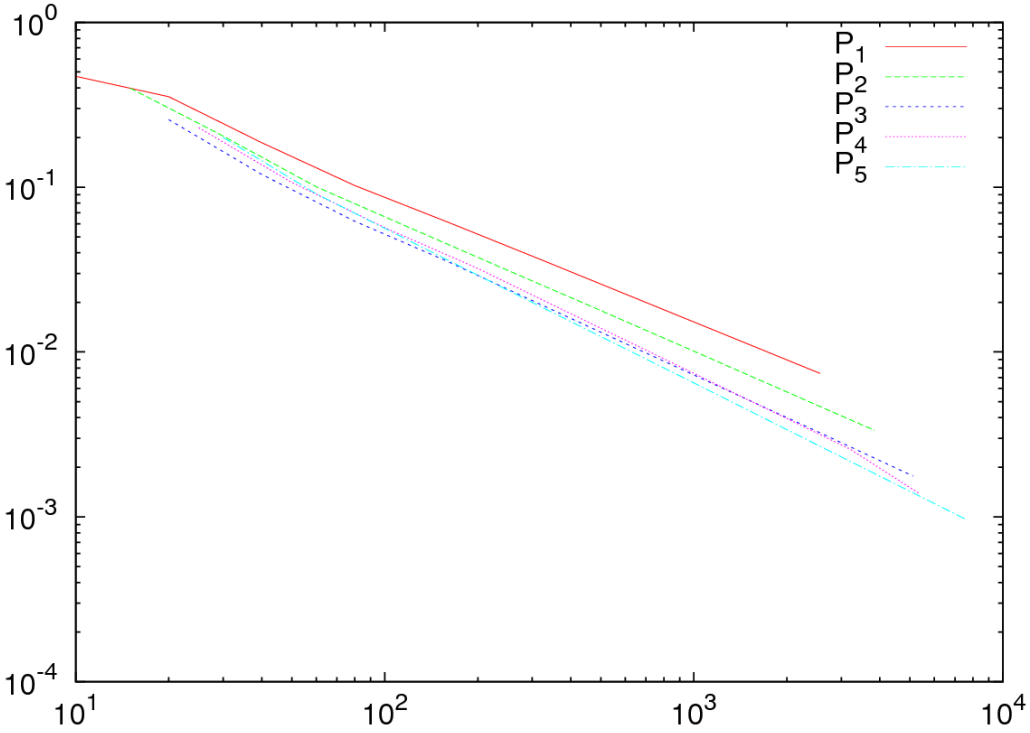

Fig. 4.6. 1D linear transport equation with discontinuous data. L^1-norm of the error at time $T = 1$ vs No. of dofs for P_1,...,P_5 polynomials.

The graphs of P_5 numerical solution with 100 elements and associated entropy viscosity at time $T = 1$ are shown in Figure 4.7. Note that the entropy viscosity field concentrates on discontinuities.

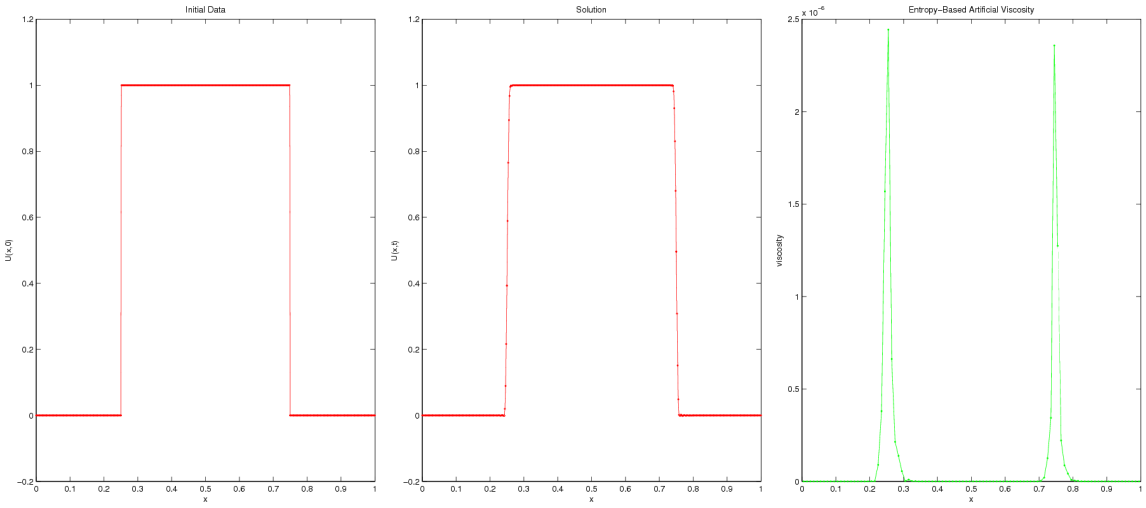

Fig. 4.7. 1D linear transport equation with discontinuous data. Initial data, P_5 solution, and entropy viscosity at time $T = 1$.

Burgers′ equation :

Consider the following IBVP:

$$\begin{cases} \partial_t u + \partial_x \left(\dfrac{u^2}{2} \right) = 0 \ , \quad (x,t) \in \Omega \times (0,T], \\ u(x,0) = u_0(x) = \sin(2\pi x) \ , \quad x \in \Omega. \end{cases} \tag{4.59}$$

with periodic boundary conditions. The method of characteristics can be involved to solve this problem analytically, see Chapter II. We use the entropy pair $(\eta, \psi) = \left(\dfrac{u^2}{2}, \dfrac{u^3}{3} \right)$, and computational parameters are defined in Table 4.13.

Table 4.13
1D Burgers' equation. Parameters.

Ω	$(0,1)$	ω	0.5
T	0.25	δ	0 or 1
RK	4	C_E	1.0
CFL	0.5	$C_{\max}$	$0.5/p$

The numerical solution is computed up to time $T = 0.25$ when the stationary shock is fully developed. We compute the L^1- and L^2-norms of the error and the corresponding convergence rates at time $T = 0.25$ on various uniform meshes for the P_1, P_2, and P_3 approximations. For more details, the reader is referred to the Tables 4.14, 4.15, and 4.16, respectively. We observe that the convergence rates in the L^1- and L^2-norms are 1 and 0.5, respectively and independently of the polynomial degree of the approximation.

Table 4.14
1D Burgers' equation. Convergence test for P_1.

cells	dofs	h	L^1-error	R_1	L^2-error	R_2
5	10	2e-01	2.617e-01	-	3.722e-01	-
10	20	1e-01	1.483e-01	0.82	2.811e-01	0.41
20	40	5e-02	7.247e-02	1.03	1.910e-01	0.56
40	80	2.5e-02	3.339e-02	1.12	1.313e-01	0.54
80	160	1.25e-02	1.603e-02	1.06	9.160e-02	0.52
160	320	6.25e-03	7.944e-03	1.01	6.460e-02	0.50
320	640	3.125e-03	3.959e-03	1.01	4.565e-02	0.50

Table 4.15
1D Burgers' equation. Convergence test for P_2.

cells	dofs	h	L^1-error	R_1	L^2-error	R_2
5	15	2e-01	1.242e-01	-	2.623e-01	-
10	30	1e-01	5.058e-02	1.30	1.567e-01	0.74
20	60	5e-02	2.409e-02	1.07	1.056e-01	0.57
40	120	2.5e-02	1.191e-02	1.02	7.405e-02	0.51
80	240	1.25e-02	5.922e-03	1.01	5.219e-02	0.51
160	480	6.25e-03	2.957e-03	1.00	3.688e-02	0.50
320	960	3.125e-03	1.478e-03	1.00	2.608e-02	0.50

Table 4.16
1D Burgers' equation. Convergence test for P_3.

cells	dofs	h	L^1-error	R_1	L^2-error	R_2
5	20	2e-01	6.621e-02	-	1.698e-01	-
10	40	1e-01	3.201e-02	1.05	1.040e-01	0.71
20	80	5e-02	1.609e-02	0.99	7.327e-02	0.51
40	160	2.5e-02	7.988e-03	1.01	5.182e-02	0.50
80	320	1.25e-02	3.983e-03	1.00	3.664e-02	0.50
160	640	6.25e-03	1.991e-03	1.00	2.591e-02	0.50
320	1280	3.125e-03	9.953e-04	1.00	1.832e-02	0.50

The numerical solution based on the P_3 approximation and artificial stabilizing viscosity at time $T = 0.25$ are shown in Figure 4.8. The number of elements used here is 320. The entropy viscosity field is highly concentrated on the shock.

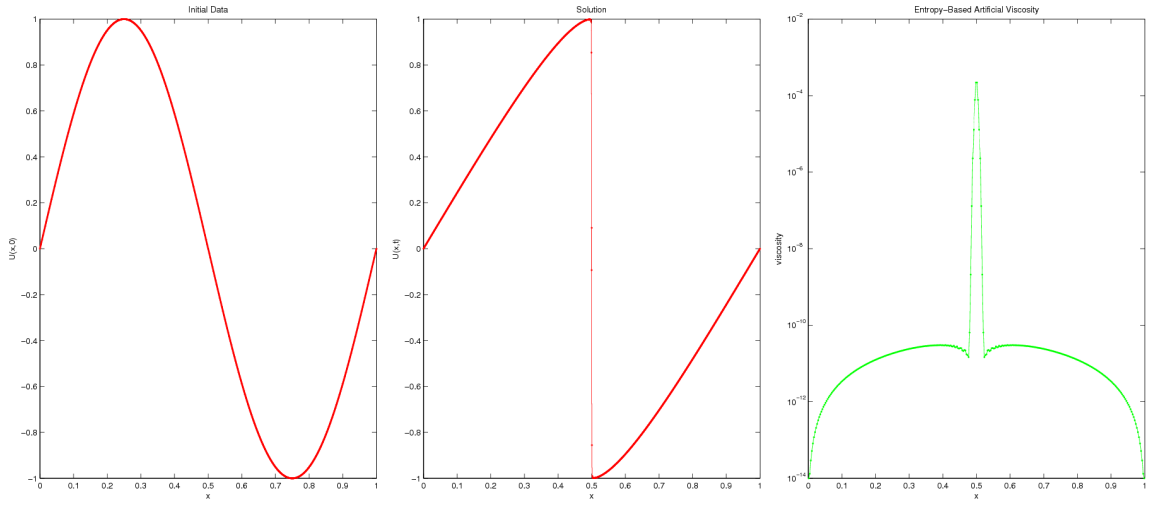

Fig. 4.8. 1D Burgers' equation. Initial data, P_3 solution, and logarithm of entropy viscosity at time $T = 0.25$.

Nonconvex problem :

To demonstrate the capability of the method to solve nonlinear scalar conservation equations with nonconvex fluxes, we consider the following Riemann problem:

$$\begin{cases} \partial_t u + \partial_x f(u) = 0 \ , \quad (x,t) \in \Omega \times (0,T], \\ u(x,0) = u_0(x) \ , \quad x \in \Omega. \end{cases} \tag{4.60}$$

where

$$f(u) = \begin{cases} \frac{1}{4}u(1-u) \quad \text{if} \quad u \leq 0.5, \\ \frac{1}{2}u(u-1) + \frac{3}{16} \quad \text{if} \quad u > 0.5. \end{cases} \tag{4.61}$$

with the initial data defined as follows:

$$u_0(x) = \begin{cases} 0 & \text{if} \quad x \leq 0.35, \\ 1 & \text{if} \quad x > 0.35. \end{cases} \tag{4.62}$$

Note that both $f(u)$ and $f'(u)$ are continuous functions whereas $f''(u) = -\frac{1}{2} < 0$ if $u \leq 0.5$ and $f''(u) = 1 > 0$ if $u > 0.5$. The BC's are $u(0,t) = 0$ and $u(1,t) = 1$.

The problem (4.60)-(4.62) and many similar problems are known to produce numerical solutions converging to physically incorrect or entropy violating solutions. The reason is that in the case of nonconvex inviscid fluxes, one entropy pair is not enough to guarantee that an entropy inequality holds for all possible entropy pairs.

For this computation we define an entropy function as $\eta(u) = \left(u - \frac{1}{2}\right)^2$, we employ P_3 approximation with 320 elements, and use parameters shown in Table 4.17.

Table 4.17
1D nonconvex problem. Parameters.

Ω	$(0,1)$	ω	0.5
T	1.0	δ	0 or 1
RK	4	C_E	1.0
CFL	0.5	$C_{\max}$	$0.25/p$

The physically relevant solution to the problem above at time $T = 1$ consists of a shock wave, located at $x = \frac{5\sqrt{6} - 3}{20}$, followed by a rarefaction wave. The numerical solution generated by the entropy viscosity method is shown in Figure 4.9 and exactly captures the behavior of the correct solution at time $T = 1$.

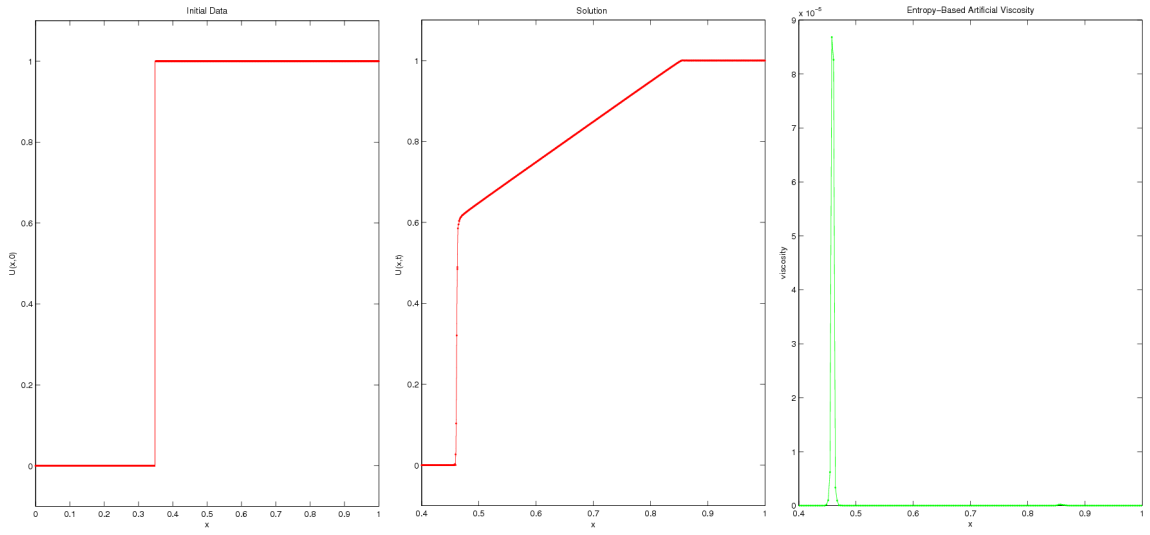

Fig. 4.9. 1D nonconvex problem. Initial data, P_3 solution, and entropy viscosity at time $T = 1$.

4.7.2 Two-dimensional tests

Linear transport equation with smooth data :

The linear transport problem in two space dimensions reads:

$$\begin{cases} \partial_t u + \nabla \cdot (\boldsymbol{\beta} u) = 0 \,, \quad (\mathbf{x}, t) \in \Omega \times (0, T], \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}) \,, \quad \mathbf{x} \in \Omega. \end{cases} \tag{4.63}$$

where $\boldsymbol{\beta} = 2\pi(-y, x)$ is a smooth vector field and $\Omega = \left\{\mathbf{x} \in \mathbb{R}^2 : |\mathbf{x}| < 1\right\}$ is a circular domain of interest. Note that $\nabla \cdot \boldsymbol{\beta} = 0$ and the boundary Γ of the domain Ω is a characteristics boundary, meaning that $\boldsymbol{\beta} \cdot \mathbf{n}|_\Gamma = 0$. The initial data is smooth and defined as follows:

$$u_0(\mathbf{x}) = \frac{1}{2} \left(1 - \tanh\left(\frac{|\mathbf{x} - \mathbf{r}_0|^2}{a_0^2} - 1\right)\right), \tag{4.64}$$

where $a_0 = 0.3$ and $\mathbf{r}_0 = (0.4, 0)$. The example of a numerical solution at time $T = 1$ using the Q_2 approximation with 81920 elements or equivalently $81920 \cdot 9 = 737280$ dofs, is shown in Figure 4.10. The corresponding entropy viscosity field with $\mu \sim \mu_E \sim C_E \mathcal{O}(h^4) \sim 10^{-8} - 10^{-9}$ is represented in Figure 4.11.

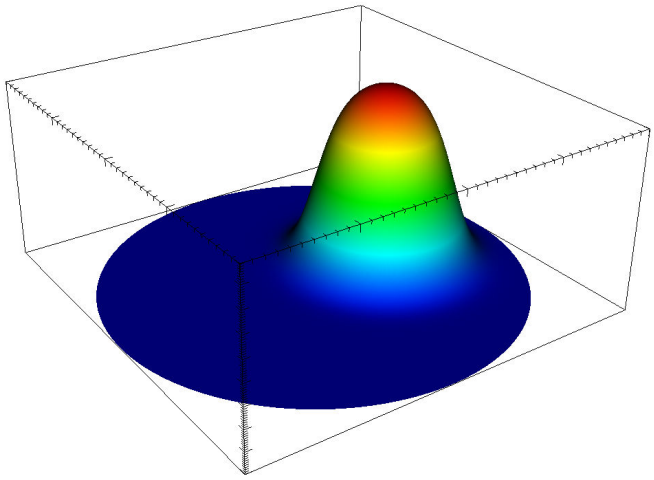

Fig. 4.10. 2D linear transport equation with smooth data. Q_2 solution at time $T = 1$.

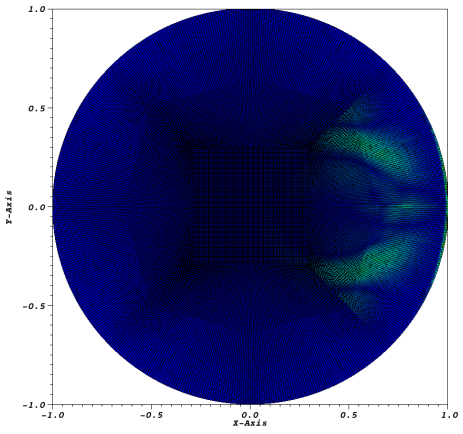

Fig. 4.11. 2D linear transport equation with smooth data. Entropy viscosity at time $T = 1$.

Two convergence tests are done by using Q_2 discontinuous Galerkin finite elements on various grids. The boundary approximation utilizes the piecewise graphs of cubic polynomials to approximate the true boundary Γ of the domain Ω. In the first series of tests we use the upwind inviscid numerical fluxes with $\omega = \frac{1}{2}$, and in the second test we use the centered fluxes with $\omega = 0$. In both cases we take $\delta = 1$ in the definition of the viscous numerical fluxes. The other computational parameters are set up in Table 4.18.

Table 4.18
2D linear transport equation with smooth data. Parameters.

Ω	unit circle in $\mathbb{R}^2$	CFL	0.25
T	1.0	C_E	0.5
RK	4	$C_{\max}$	$0.1/p$

The goal of these tests is to evaluate the importance of the parameter ω in the definition of the inviscid numerical fluxes when a smooth data is linearly transported. The errors measured in the L^1- and L^2-norms at time $T = 1$ together with the respective convergence orders are reported in Tables 4.19 and 4.20.

Table 4.19
2D linear transport equation with smooth data. Convergence test for Q_2 approximation with upwind fluxes.

dofs	h	L^1-error	R_1	L^2-error	R_2
180	0.2929	2.685e-01	-	2.482e-01	-
720	0.1768	6.570e-02	2.03	7.464e-02	1.73
2880	0.0937	8.774e-03	2.90	1.162e-02	2.68
11520	0.0496	7.655e-04	3.52	1.137e-03	3.35
46080	0.0252	5.721e-05	3.74	1.132e-04	3.33
184320	0.0128	5.699e-06	3.33	1.685e-05	2.81
737280	0.0064	7.101e-07	3.01	2.261e-06	2.90

Table 4.20
2D linear transport equation with smooth data. Convergence test for Q_2 approximation with centered fluxes.

dofs	h	L^1-error	R_1	L^2-error	R_2
180	0.2929	2.633e-01	-	2.415e-01	-
720	0.1768	9.589e-02	1.46	9.972e-02	1.28
2880	0.0937	1.895e-02	2.34	2.299e-02	2.12
11520	0.0496	2.036e-03	3.22	2.469e-03	3.22
46080	0.0252	1.895e-04	3.43	1.961e-04	3.65
184320	0.0128	1.826e-05	3.38	1.604e-05	3.61

It is remarkable that the convergence rate is 3 irrespective of the definition of the flux. Nonetheless, the upwind fluxes result in better accuracy by about a factor of 5.

Burgers′ equation :

One more benchmark problem which we consider here is a Riemann problem for two-dimensional inviscid Burgers' equation:

$$\begin{cases} \partial_t u + \nabla \cdot \left(\dfrac{u^2}{2} \mathbf{e} \right) = 0 \ , \quad (\mathbf{x}, t) \in \Omega \times (0, T], \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}) \ , \quad \mathbf{x} \in \Omega. \end{cases} \tag{4.65}$$

where the components e_i of the vector $\mathbf{e}$ are defined as $\{e_i\}_{i=1}^2 = 1$ and $\Omega = \left\{ \mathbf{x} \in \mathbb{R}^2 : 0 < x < 1 \ \text{ and } \ 0 < y < 1 \right\}$ is a unit square in $\mathbb{R}^2$. The above equation is subject to the following initial condition:

$$u_0(\mathbf{x}) = \begin{cases} +0.5 & \text{if } \ x < 0.5 \text{ and } y < 0.5, \\ -0.2 & \text{if } \ x < 0.5 \text{ and } y > 0.5, \\ +0.8 & \text{if } \ x > 0.5 \text{ and } y < 0.5, \\ -1.0 & \text{if } \ x > 0.5 \text{ and } y > 0.5. \end{cases} \tag{4.66}$$

The entropy pair that we choose for this problem is $(\eta, \boldsymbol{\psi}) = \left(\frac{u^2}{2}, \frac{u^3}{3}\mathbf{e}\right)$ and the parameters for this computation are given in Table 4.21.

Table 4.21
2D Burgers' equation. Parameters.

Ω	unit square in $\mathbb{R}^2$	ω	0.5
T	0.5	δ	0 or 1
RK	4	C_E	1.0
CFL	0.25	$C_{\max}$	$0.25/p$

The solution is computed at time $T = 0.5$ to facilitate comparisons with [6, 54]. The exact solution to this problem for $t > 0$ is given in [55]. We compute the L^1- and L^2-norms of the error and the corresponding convergence rates at time $T = 0.5$ on various uniform meshes composed of squares. The results for Q_1, Q_2, and Q_3 approximations are shown in Tables 4.22, 4.23, and 4.24, respectively. Similar to the one-dimensional case, we observe that the convergence rates in the L^1- and L^2-norms are 1 and 0.5, respectively and independently of the polynomial degree of the approximation.

Table 4.22
2D Burgers' equation. Convergence test for Q_1.

cycle	cells	dofs	h	L^1-error		L^2-error	
1	4	16	0.5000	5.845e-01	-	6.527e-01	-
2	16	64	0.2500	2.788e-01	1.07	3.782e-01	0.79
3	64	256	0.1250	1.127e-01	1.31	2.328e-01	0.70
4	256	1024	0.0625	6.415e-02	0.81	1.762e-01	0.40
5	1024	4096	0.0312	3.174e-02	1.02	1.248e-01	0.50
6	4096	16384	0.0156	1.697e-02	0.90	9.238e-02	0.43
7	16384	65536	0.0078	8.506e-03	1.00	6.678e-02	0.47

Table 4.23
2D Burgers' equation. Convergence test for Q_2.

cycle	cells	dofs	h	L^1-error		L^2-error	
1	4	36	0.5000	4.843e-01	-	5.596e-01	-
2	16	144	0.2500	1.500e-01	1.69	2.601e-01	1.11
3	64	576	0.1250	7.441e-02	1.01	1.911e-01	0.44
4	256	2304	0.0625	3.702e-02	1.01	1.370e-01	0.48
5	1024	9216	0.0312	1.948e-02	0.93	1.010e-01	0.44
6	4096	36864	0.0156	9.718e-03	1.00	7.157e-02	0.50
7	16384	147456	0.0078	5.118e-03	0.93	5.308e-02	0.43

Table 4.24
2D Burgers' equation. Convergence test for Q_3.

cycle	cells	dofs	h	L^1-error		L^2-error	
1	4	64	0.5000	4.510e-01	-	5.126e-01	-
2	16	256	0.2500	1.237e-01	1.87	2.332e-01	1.14
3	64	1024	0.1250	5.628e-02	1.14	1.638e-01	0.51
4	256	4096	0.0625	2.945e-02	0.93	1.223e-01	0.42
5	1024	16384	0.0312	1.445e-02	1.03	8.728e-02	0.49
6	4096	65536	0.0156	7.618e-03	0.92	6.395e-02	0.45
7	16384	262144	0.0078	3.731e-03	1.03	4.527e-02	0.50

The graph of the Q_3 numerical solution at time $T = 0.5$ is shown in the left panel in Figure 4.12 and the entropy viscosity field is shown in the right panel. The entropy viscosity focuses on the shock as anticipated. The boundary conditions for this test simulate a transparent boundary and imposed on the boundary Γ weakly via the DG numerical fluxes.

KPP rotating wave :

The last benchmark problem which we consider in the context of the current chapter is a so-called KPP (Kurganov-Petrov-Popov) rotating wave problem:

$$\begin{cases} \partial_t u + \nabla \cdot (\sin u, \cos u)^{\mathrm{T}} = 0 \ , \quad (\mathbf{x}, t) \in \Omega \times (0, T], \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}) \ , \quad \mathbf{x} \in \Omega. \end{cases} \tag{4.67}$$

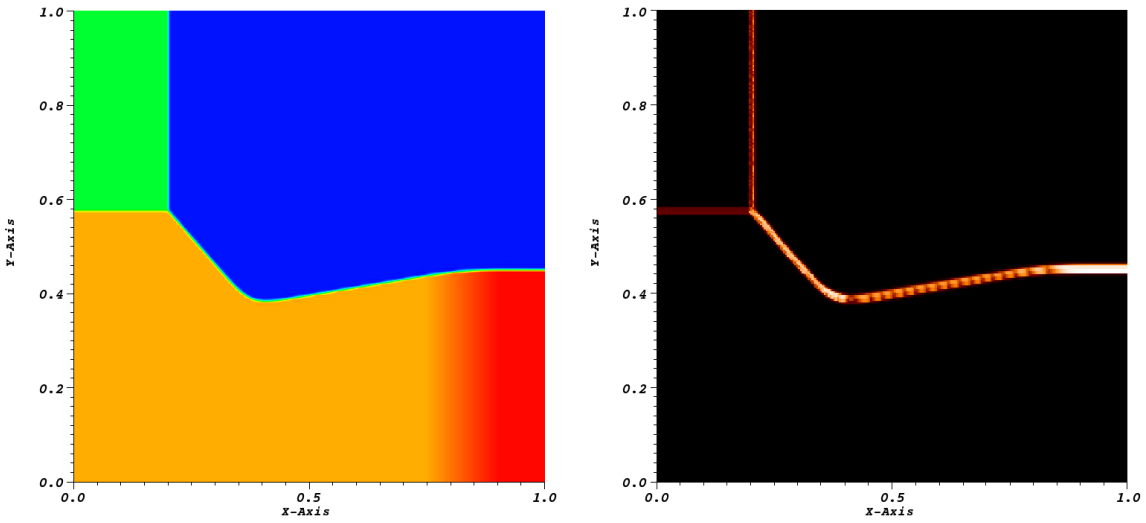

Fig. 4.12. 2D Burgers' equation. Q_3 solution and entropy viscosity at time $T = 0.5$.

where the discontinuous initial data is defined as follows:

$$u_0(\mathbf{x}) = \begin{cases} 3.5\pi & \text{if } \; x^2 + y^2 < 1, \\ .25\pi & \text{otherwise.} \end{cases} \tag{4.68}$$

The local velocity is $\mathbf{f}'(u) = (\cos u, -\sin u)^{\mathrm{T}}$. This test was proposed in [56] and is challenging to many higher-order numerical schemes because the solution has a two-dimensional composite wave structure. For example centered-upwind schemes based on WENO5, Minmod 2, and SuperBee reconstructions fail in this test case, see [56] for more details.

The entropy pair for this problem is $(\eta, \boldsymbol{\psi}) = \left(\dfrac{u^2}{2}, (u \sin u + \cos u, u \cos u - \sin u)^{\mathrm{T}}\right)$ and parameters of the computation are given in Table 4.25.

The initial data is shown in Figure 4.13. The Q_2 entropy viscosity solution built on $4^8 = 65536$ elements and associated entropy viscosity field are depicted in Figures

Table 4.25
2D KPP rotating wave. Parameters.

Ω	$(-2,2)\times(-2.5,1.5)$	ω	0.5
T	1	δ	0 or 1
RK	4	C_E	5.0
CFL	0.2	$C_{\max}$	$0.5/p$

4.14 and 4.15, respectively. The solution shows the expected rotating composite wave structure. As expected, dissipation is only added in the shock region.

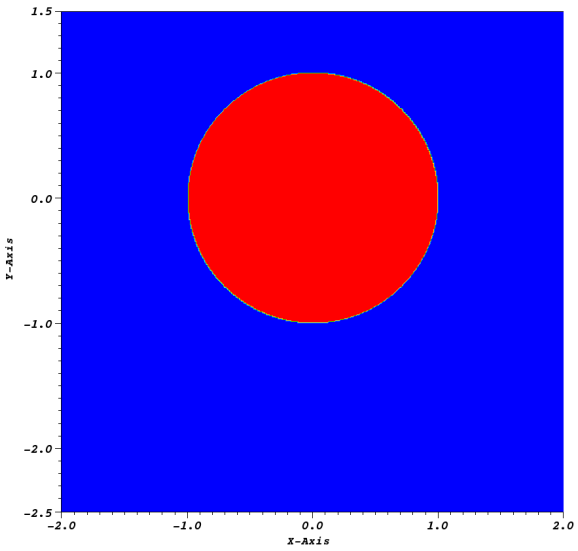

Fig. 4.13. 2D KPP rotating wave. Initial data.

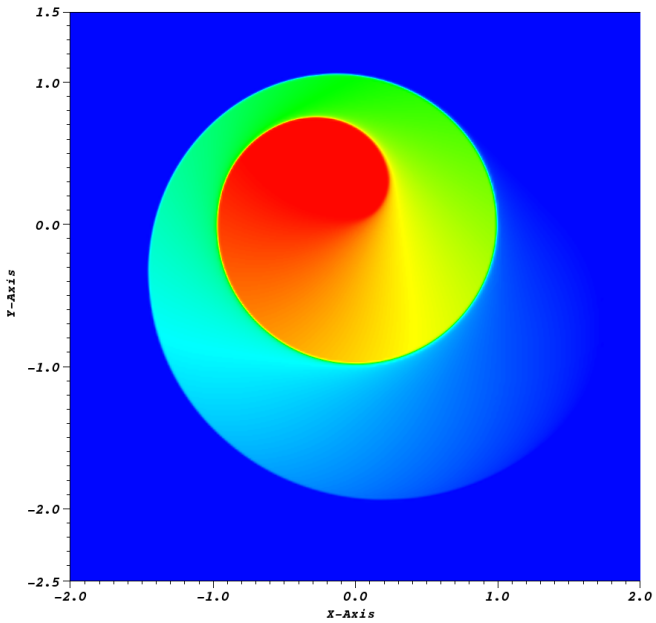

Fig. 4.14. 2D KPP rotating wave. Q_2 solution at time $T = 1$.

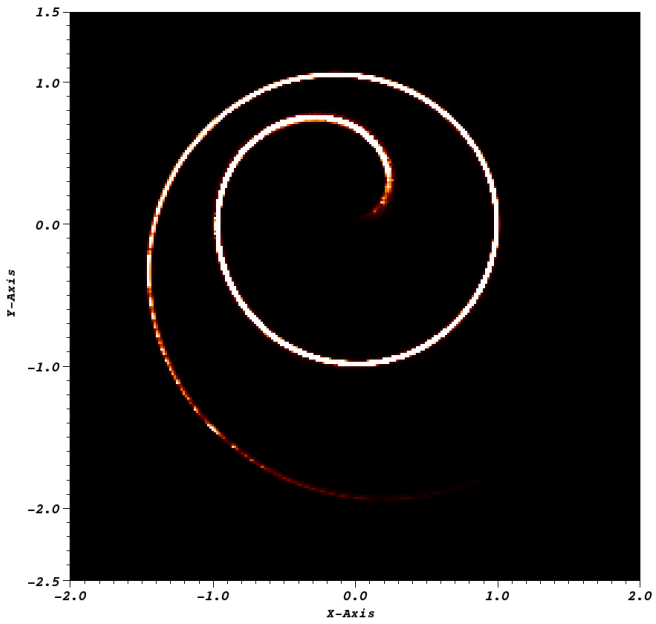

Fig. 4.15. 2D KPP rotating wave. Entropy viscosity trace at time $T = 1$.

4.8 Conclusion

The method which we have explored in this chapter performs very well on all test problems specially designed for assessment of efficiency and robustness of numerical algorithms for hyperbolic scalar conservation equations. As we know, many of these equations accept analytical solutions (in closed form or not), which allows researchers to perform informative convergence tests. The convergence orders measured in these tests in both the L^1- and L^2-norms of the error apparently approach the optimal thresholds predicted theoretically. The excellent results obtained for scalar equations justify the extension of the entropy viscosity method to gas dynamics problems.

5. EXTENSION OF THE METHOD TO COMPRESSIBLE GAS DYNAMICS EQUATIONS

5.1 Introduction

In the last chapter of this dissertation we discuss the application of the DG spatial discretization technique in the framework of the entropy viscosity method to the system of compressible Euler equations. This chapter provides the natural extension of the method described in Chapter IV to the more general case of a system of nonlinear hyperbolic conservation laws supplemented with an auxiliary entropy inequality. At the end of the chapter we consider several benchmark test cases in one and two space dimensions which agree well with the reference solutions found in the literature.

5.2 Regularization of Euler equations

We write the original system of compressible Euler equations in the following conservative form:

$$\begin{cases} \partial_t \mathbf{c} + \nabla \cdot [\mathbf{f}(\mathbf{c})] = 0, & (\mathbf{x}, t) \in \Omega \times (0, T], \\ \mathbf{c}(\mathbf{x}, 0) = \mathbf{c}_0(\mathbf{x}), & \mathbf{x} \in \Omega. \end{cases} \tag{5.1}$$

where $\mathbf{c} = \begin{pmatrix} \rho \\ \mathbf{m} \\ E \end{pmatrix}$ is a vector of conserved variables, $\mathbf{f}(\mathbf{c}) = \begin{pmatrix} \mathbf{m} \\ \frac{1}{\rho}\mathbf{m} \otimes \mathbf{m} + p\mathbb{I} \\ \frac{\mathbf{m}}{\rho}(E + p) \end{pmatrix}$ is a vector of inviscid fluxes, $\Omega \subset \mathbb{R}^d$ is an open connected domain with the boundary Γ, $(0, T] \in \mathbb{R}_+$ is a time interval, and $\mathbf{c}_0(\mathbf{x})$ is the initial data which is known. Here, $\mathbf{m} \otimes \mathbf{m}$ and $\mathbb{I}$ are symmetric tensors of rank 2 whose entries are $(\mathbf{m} \otimes \mathbf{m})_{ij} = m_i m_j$ and $\mathbb{I}_{ij} = \delta_{ij}$, respectively, and indices i and j represent coordinate axes and vary

from 1 to d. Since pressure p is supposed to be another unknown, we use the Equation of State for a perfect gas to express the pressure p through the density of a gas ρ, momentum $\mathbf{m}$, and total energy E as follows:

$$p = (\gamma - 1)\left(E - \frac{\mathbf{m}\cdot\mathbf{m}}{2\rho}\right). \tag{5.2}$$

To solve the Euler equations, one must equip the system (5.1) with the appropriate boundary conditions on Γ. In the framework of DG methods the boundary conditions of different types are commonly imposed implicitly via respective numerical fluxes. For instance, we use implicit inflow, outflow, or transparent boundary conditions, whereas the slip boundary condition $\mathbf{u}\cdot\mathbf{n}|_\Gamma = 0$ is applied to a solution vector $\mathbf{c}$ explicitly within each Runge-Kutta iteration.

An additional entropy inequality introduced in Chapter II reads:

$$\partial_t s + \nabla\cdot(\mathbf{u}s) \geq 0, \tag{5.3}$$

where $\mathbf{u} = \dfrac{\mathbf{m}}{\rho}$ is a particles velocity vector field and s is an entropy function which we previously defined as:

$$s = \frac{\rho R}{\gamma - 1}\ln\frac{p}{\rho^\gamma}. \tag{5.4}$$

Conforming to the technique discussed in Chapter IV, we regularize the original system of equations (5.1) and write it as follows:

$$\begin{cases} \partial_t \mathbf{c} + \nabla\cdot[\mathbf{f}(\mathbf{c})] = -\nabla\cdot[\mathbf{q}], & (\mathbf{x},t)\in\Omega\times(0,T], \\ \mathbf{c}(\mathbf{x},0) = \mathbf{c}_0(\mathbf{x}), & \mathbf{x}\in\Omega. \end{cases} \tag{5.5}$$

where $\mathbf{q}$ is a vector of artificial viscous fluxes taken from the right hand side of the system of Navier-Stokes equations and defined in the following manner:

$$\mathbf{q} = \begin{pmatrix} -\nu\nabla\rho \\ -\mu\nabla_s\mathbf{u} \\ -\mu\nabla_s\mathbf{u}\mathbf{u} - \kappa\nabla T \end{pmatrix}. \tag{5.6}$$

Here, $\nabla_s\mathbf{u}$ is a symmetric tensor of rank 2 with the entries $(\nabla_s\mathbf{u})_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$, ν is a diffusion viscosity, μ is a dynamic viscosity, and κ is a thermal viscosity. All these viscosities are artificial and will be further defined, based on the entropy inequality (5.3). The new variable - temperature T is obtained from the same Equation of State for a perfect gas:

$$T = \frac{\gamma - 1}{R}\left(\frac{E}{\rho} - \frac{\mathbf{m}\cdot\mathbf{m}}{2\rho^2}\right). \tag{5.7}$$

The regularized system of PDEs (5.5) is closed by the relations (5.2) and (5.7). We first need to discretize it in space.

5.3 Discretization in space

The spatial discretization of equations (5.5) goes along the same line as we discussed in Chapter IV relating to hyperbolic scalar conservation laws. The only difference is that instead of one unknown per node, we now have $d+2$ unknown values of a numerical solution, namely: density ρ, all components of momentum $\{m_i\}_{i=1}^d$, and total energy E. We use the same mesh notation and functional spaces as previously defined, and write each individual PDE of the system (5.5) in the following form:

$$\partial_t c + \nabla\cdot\mathbf{f}_c = -\nabla\cdot\mathbf{q}_c\,, \quad (\mathbf{x},t) \in \Omega\times(0,T], \tag{5.8}$$

where c means one of the following: density ρ, or a component of momentum m_i, or total energy E. If $c = \rho$, then $\mathbf{f}_\rho = \mathbf{m}$ and $\mathbf{q}_\rho = -\nu\nabla\rho$. If $c = m_i$, then $\mathbf{f}_{m_i}$ can be obtained as follows:

$$\mathbf{f}_{m_1} = \begin{pmatrix} \frac{1}{\rho}m_1^2 + p \\ \frac{1}{\rho}m_2 m_1 \\ \frac{1}{\rho}m_3 m_1 \end{pmatrix}, \quad \mathbf{f}_{m_2} = \begin{pmatrix} \frac{1}{\rho}m_1 m_2 \\ \frac{1}{\rho}m_2^2 + p \\ \frac{1}{\rho}m_3 m_2 \end{pmatrix}, \quad \mathbf{f}_{m_3} = \begin{pmatrix} \frac{1}{\rho}m_1 m_3 \\ \frac{1}{\rho}m_2 m_3 \\ \frac{1}{\rho}m_3^2 + p \end{pmatrix}, \tag{5.9}$$

and respective viscous fluxes $\mathbf{q}_{m_i}$ read:

$$\mathbf{q}_{m_1} = -\frac{\mu}{2}\begin{pmatrix} 2\frac{\partial u_1}{\partial x_1} \\ \frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2} \\ \frac{\partial u_3}{\partial x_1} + \frac{\partial u_1}{\partial x_3} \end{pmatrix}, \quad \mathbf{q}_{m_2} = -\frac{\mu}{2}\begin{pmatrix} \frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \\ 2\frac{\partial u_2}{\partial x_2} \\ \frac{\partial u_3}{\partial x_2} + \frac{\partial u_2}{\partial x_3} \end{pmatrix},$$

$$\mathbf{q}_{m_3} = -\frac{\mu}{2}\begin{pmatrix} \frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} \\ \frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2} \\ 2\frac{\partial u_3}{\partial x_3} \end{pmatrix}. \tag{5.10}$$

If $c = E$, then $\mathbf{f}_E = \dfrac{\mathbf{m}}{\rho}(E + p)$ and respective viscous flux $\mathbf{q}_E$ is defined in the following form:

$$\mathbf{q}_E = \mathbf{q}_{E_\mu} + \mathbf{q}_{E_T} \quad \text{where} \quad \mathbf{q}_{E_\mu} = -\mu\nabla_s\mathbf{u}\mathbf{u} \quad \text{and} \quad \mathbf{q}_{E_T} = -\kappa\nabla T. \tag{5.11}$$

Eventually, the discontinuous Galerkin finite element method of degree $p > 0$ can be formulated for each individual equation of the system (5.5) as follows:

$$\forall K \in \mathbb{T}_h \text{ find } c_h \in \mathcal{V}_h^p \text{ such that } \forall v_{c,h} \in \mathcal{V}_h^p :$$

$$\int_K v_{c,h}\partial_t c_h d\mathbf{x} + \oint_{\partial K} v_{c,h}\hat{F}_{c,h,\partial K} d\sigma - \int_K \nabla v_{c,h} \cdot \mathbf{f}_{c,h} d\mathbf{x} + \oint_{\partial K} v_{c,h}\hat{Q}_{c,h,\partial K} d\sigma - \int_K \nabla v_{c,h} \cdot \mathbf{q}_{c,h} d\mathbf{x} = 0. \tag{5.12}$$

We next say a few words about numerical fluxes $\hat{F}_{c,h,\partial K}$ and $\hat{Q}_{c,h,\partial K}$ which we use in the DG formulation above.

5.3.1 Inviscid numerical fluxes

The definitions of the Lax-Friedrichs inviscid numerical fluxes are given below:

Density :

$$\hat{F}_{\rho,h,\partial K} = \frac{1}{2}\left(\mathbf{f}^{+}_{\rho,h,\partial K} + \mathbf{f}^{-}_{\rho,h,\partial K}\right)\cdot\mathbf{n}_{\partial K} + \omega\alpha_{\partial K}\left(\rho^{+}_{h,\partial K} - \rho^{-}_{h,\partial K}\right). \tag{5.13}$$

A component of momentum :

$$\hat{F}_{m_i,h,\partial K} = \frac{1}{2}\left(\mathbf{f}^{+}_{m_i,h,\partial K} + \mathbf{f}^{-}_{m_i,h,\partial K}\right)\cdot\mathbf{n}_{\partial K} + \omega\alpha_{\partial K}\left(m^{+}_{i,h,\partial K} - m^{-}_{i,h,\partial K}\right). \tag{5.14}$$

Total energy :

$$\hat{F}_{E,h,\partial K} = \frac{1}{2}\left(\mathbf{f}^{+}_{E,h,\partial K} + \mathbf{f}^{-}_{E,h,\partial K}\right)\cdot\mathbf{n}_{\partial K} + \omega\alpha_{\partial K}\left(E^{+}_{h,\partial K} - E^{-}_{h,\partial K}\right). \tag{5.15}$$

The parameter ω is typically $\frac{1}{2}$ and the value of $\alpha_{\partial K}$ was defined in Chapter IV as the maximum propagation speed associated with the whole boundary ∂K (or with any of its parts):

$$\alpha_{\partial K} = \| \, |\mathbf{u}_h\cdot\mathbf{n}_{\partial K}| + \sqrt{\gamma R T_h} \, \|_{L^\infty(K\cup\Delta_K)}, \tag{5.16}$$

where $\mathbf{u}_h \cdot \mathbf{n}_{\partial K}$ is a projection of speed of particles $\mathbf{u}_h$ onto the direction defined by $\mathbf{n}_{\partial K}$ and $\sqrt{\gamma R T_h}$ is a speed of sound.

5.3.2 Viscous numerical fluxes

We define the viscous numerical fluxes as follows:

Density :

$$\hat{Q}_{\rho,h,\partial K} = \frac{1}{2}\left(\mathbf{q}^{+}_{\rho,h,\partial K} + \mathbf{q}^{-}_{\rho,h,\partial K}\right) \cdot \mathbf{n}_{\partial K} + \delta\beta_{\rho,\partial K}\left(\rho^{+}_{h,\partial K} - \rho^{-}_{h,\partial K}\right), \tag{5.17}$$

where

$$\beta_{\rho,\partial K} = \left\| \frac{\nu_h}{h} \right\|_{L^\infty(K\cup\Delta_K)}. \tag{5.18}$$

A component of momentum :

$$\hat{Q}_{m_i,h,\partial K} = \frac{1}{2}\left(\mathbf{q}^{+}_{m_i,h,\partial K} + \mathbf{q}^{-}_{m_i,h,\partial K}\right) \cdot \mathbf{n}_{\partial K} + \delta\beta_{m_i,\partial K}\left(u^{+}_{i,h,\partial K} - u^{-}_{i,h,\partial K}\right), \tag{5.19}$$

where

$$\beta_{m_i,\partial K} = \left\| \frac{\mu_h}{h} \right\|_{L^\infty(K\cup\Delta_K)}. \tag{5.20}$$

Total energy. **Viscosity dependence** :

$$\hat{Q}_{E_\mu,h,\partial K} = \frac{1}{2}\left(\mathbf{q}^{+}_{E_\mu,h,\partial K} + \mathbf{q}^{-}_{E_\mu,h,\partial K}\right)\cdot\mathbf{n}_{\partial K} + \delta\beta_{E_\mu,\partial K}\left(\frac{E^{+}_{h,\partial K}}{\rho^{+}_{h,\partial K}} - \frac{E^{-}_{h,\partial K}}{\rho^{-}_{h,\partial K}}\right), \tag{5.21}$$

where

$$\beta_{E_\mu,\partial K} = \left\| \frac{\mu_h}{h} \right\|_{L^\infty(K\cup\Delta_K)}. \tag{5.22}$$

Total energy. **Temperature dependence** :

$$\hat{Q}_{E_T,h,\partial K} = \frac{1}{2}\left(\mathbf{q}^{+}_{E_T,h,\partial K} + \mathbf{q}^{-}_{E_T,h,\partial K}\right)\cdot\mathbf{n}_{\partial K} + \delta\beta_{E_T,\partial K}\left(T^{+}_{h,\partial K} - T^{-}_{h,\partial K}\right), \tag{5.23}$$

where

$$\beta_{E_T,\partial K} = \left\| \frac{\kappa_h}{h} \right\|_{L^\infty(K\cup\Delta_K)}. \tag{5.24}$$

All of the numerical fluxes listed above are both consistent with the respective physical fluxes and conservative. Parameter δ can be chosen to be either 0 or 1.

5.4 Discretization in time

The time discretization procedure which we apply to spatially semi-discrete forms (5.12) is quite similar to that employed in Chapter IV. Substitution of the following expressions:

$$v_{c,h} = \sum_{i\in\zeta_{c,K}} v_{c,h,i}\Phi_i(\mathbf{x}) \tag{5.25}$$

and

$$c_h = \sum_{j \in \zeta_{c,K}} c_{h,j}(t)\Phi_j(\mathbf{x}) \tag{5.26}$$

into (5.12), where $\zeta_{c,K} \in \mathbb{N}_+$ is a local numbering of degrees of freedom within an element $K \in \mathbb{T}_h$ assigned to a component c_h, yields the following system of ODEs:

$$\forall K \in \mathbb{T}_h \text{ find } \left\{c_{h,j}(t)\right\}_{j \in \zeta_{c,K}} \text{ such that } \forall i \in \zeta_{c,K} :$$

$$\begin{aligned}
&\sum_{j \in \zeta_{c,K}} \int_K \Phi_i(\mathbf{x})\Phi_j(\mathbf{x})d\mathbf{x} \cdot \frac{d}{dt}c_{h,j}(t) = \\
&\int_K \nabla\Phi_i(\mathbf{x}) \cdot \mathbf{f}_{c,h}d\mathbf{x} + \int_K \nabla\Phi_i(\mathbf{x}) \cdot \mathbf{q}_{c,h}d\mathbf{x} - \\
&\oint_{\partial K} \Phi_i(\mathbf{x})\hat{F}_{c,h,\partial K}d\sigma - \oint_{\partial K} \Phi_i(\mathbf{x})\hat{Q}_{c,h,\partial K}d\sigma.
\end{aligned} \tag{5.27}$$

The matrix-vector form of the system is defined as follows:

$$\mathbb{M}\frac{d}{dt}\mathbf{c}_h = \mathbf{R}_h. \tag{5.28}$$

To solve the system (5.28), we use the explicit Runge-Kutta methods as described in Chapter IV with a time step τ_n defined by the following CFL condition:

$$\tau_n = \min_{K \in \mathbb{T}_h} \overbrace{\left(\text{CFL } \frac{h_K}{\| \, |\mathbf{u}_h^n| + \sqrt{\gamma R T_h^n} \, \|_{L^\infty(K)}}\right)}^{\tau_{n,K}}, \tag{5.29}$$

where $\mathbf{u}_h^n$ and $\sqrt{\gamma R T_h^n}$ are the speed of particles and speed of sound at time t^n, respectively. Since we now have $d+2$ components of a numerical solution per DoF, it means that we pass through each element $K \in \mathbb{T}_h$ exactly $d+2$ times while assembling the mass matrix $\mathbb{M}$ and the right hand side $\mathbf{R}_h$. Therefore, each particular block $\mathbb{M}_K$ of the block-diagonal mass matrix $\mathbb{M}$ is itself a block-diagonal matrix consisting of $d+2$ sub-blocks. Inverting of such sub-blocks and blocks as well as the whole mass matrix $\mathbb{M}$ is straightforward.

5.5 Artificial viscosities

Construction of the entropy viscosity fields $\boldsymbol{\nu}_h$, $\boldsymbol{\mu}_h$, and $\boldsymbol{\kappa}_h$ is essentially based on the entropy inequality (5.3). Given numerical solutions $\mathbf{c}_h^a$ at time t^a and $\mathbf{c}_h^{b>a}$ at time $t^{b>a}$, the construction of artificial viscosities proceeds as follows:

- Compute the entropy residual $D_{h,K,q}^b$ at all quadrature points $\mathbf{x}_q \in K$ at time t^b:

$$D_{h,K,q}^b = \partial_t s_{h,K,q}^b + \nabla \cdot \left(\mathbf{u}_{h,K,q}^b s_{h,K,q}^b \right). \tag{5.30}$$

  For practical computations we replace the equation (5.30) with a discrete analogue:

$$D_{h,K,q}^b = \frac{s_{h,K,q}^b - s_{h,K,q}^a}{t^b - t^a} + \frac{1}{2} \left(\nabla \cdot \left(\mathbf{u}_{h,K,q}^a s_{h,K,q}^a \right) + \nabla \cdot \left(\mathbf{u}_{h,K,q}^b s_{h,K,q}^b \right) \right). \tag{5.31}$$

- Compute the jump of the entropy flux $J_{h,\partial K,g}^b$ at all quadrature points $\mathbf{x}_g \in \partial K$ at time t^b:

$$J_{h,\partial K,g}^b = \left(\mathbf{u}_{h,\partial K,g}^{b,+} s_{h,\partial K,g}^{b,+} - \mathbf{u}_{h,\partial K,g}^{b,-} s_{h,\partial K,g}^{b,-} \right) \cdot \mathbf{n}_{\partial K,g}. \tag{5.32}$$

- Define the maximum, say $R_{h,K}^b$ associated with element K at time t^b:

$$R_{h,K}^b = \max_{q,g} \left(\, h_K \left| D_{h,K,q}^b \right| \, , \, \left| J_{h,\partial K,g}^b \right| \, \right). \tag{5.33}$$

- Use this maximum to define a dynamic viscosity, say $\mu_{h,K,E}^b$ associated with element K at time t^b:

$$\mu_{h,K,E}^b = C_E \, \frac{h_K \max\limits_{q \in K} \rho_{h,K,q}^b}{\max\limits_{q \in K \in \mathbb{T}_h} \left| s_{h,K,q}^b - \bar{s}_h^b \right|} \, R_{h,K}^b, \tag{5.34}$$

where C_E is a tunable constant, h_K is a diameter of element K, $\max\limits_{q\in K\in\mathbb{T}_h}|\cdot|$ is taken over all quadrature points that can be possibly found in the entire triangulation $\mathbb{T}_h$, and $\bar{s}_h^b = \dfrac{1}{\|\Omega\|}\displaystyle\int_\Omega s_h\left(\mathbf{x},t^b\right)d\mathbf{x}$ is the space-averaged value of the entropy computed at time t^b.

- Introduce an entropy dynamic viscosity upper bound $\mu^b_{h,K,\max}$ associated with element K at time t^b:

$$\mu^b_{h,K,\max} = C_{\max}h_K \max_{q\in K}\rho^b_{h,K,q}\ \max_{q\in K}\left(\left|\mathbf{u}^b_{h,K,q}\right| + \sqrt{\gamma R T^b_{h,K,q}}\right). \tag{5.35}$$

- Define the entropy dynamic viscosity $\mu^b_{h,K}$ associated with element K at time t^b:

$$\mu^b_{h,K} = \min\left(\ \mu^b_{h,K,E}\ ,\ \mu^b_{h,K,\max}\ \right). \tag{5.36}$$

- Define the entropy diffusion viscosity $\nu^b_{h,K}$ and entropy thermal viscosity $\kappa^b_{h,K}$ associated with element K at time t^b as follows:

$$\nu^b_{h,K} = \frac{\mathcal{P}_\rho}{\max\limits_{q\in K}\rho^b_{h,K,q}}\mu^b_{h,K} \quad , \quad \kappa^b_{h,K} = \frac{\gamma\mathcal{P}_T}{\gamma-1}\mu^b_{h,K}, \tag{5.37}$$

  where $\mathcal{P}_\rho = \mathcal{O}(1)$ and $\mathcal{P}_T = \mathcal{O}(1)$ are two tunable Prandtl numbers for density ρ and temperature T, respectively. Our experience is that using $\mathcal{P}_\rho \in [0,\frac{1}{4}]$ and $\mathcal{P}_T \in [0,\frac{1}{4}]$ works well in most cases.

- Use the entropy viscosity fields $\left\{\nu^b_{h,K}\right\}_{K\in\mathbb{T}_h}$, $\left\{\mu^b_{h,K}\right\}_{K\in\mathbb{T}_h}$, and $\left\{\kappa^b_{h,K}\right\}_{K\in\mathbb{T}_h}$ together with solution data to reach the next time level.

Formulas (5.37) are derived using the definition of the Prandtl number. Originally the Prandtl number was introduced as the ratio of momentum diffusivity to thermal diffusivity:

$$\mathcal{P}r = \frac{\text{viscous diffusion rate}}{\text{thermal diffusion rate}} = \frac{\text{kinematic viscosity}}{\text{thermal diffusivity}} = \frac{\mu/\rho}{\kappa/\left(\rho c_p\right)} = \frac{c_p\mu}{\kappa}. \tag{5.38}$$

Since for a perfect gas:

$$c_p = \frac{C_p}{m} = \frac{i+2}{2}R = \frac{\gamma}{\gamma-1}R, \tag{5.39}$$

then:

$$\kappa = \frac{\gamma}{\gamma-1}\,\frac{R}{\mathcal{P}r}\,\mu. \tag{5.40}$$

Designating $\dfrac{R}{\mathcal{P}r}$ as $\mathcal{P}_T$ yields the second formula in (5.37). Introducing $\mathcal{P}_\rho$ as:

$$\mathcal{P}_\rho = \frac{\text{viscous diffusion rate}}{\text{mass diffusion rate}} = \frac{\text{kinematic viscosity}}{\text{mass diffusivity}} = \frac{\mu/\rho}{\nu}, \tag{5.41}$$

we obtain the first formula in (5.37).

5.6 Numerical tests

We demonstrate the performance of the method by solving several benchmark test problems in one and two space dimensions. The two-dimensional problems are solved using the deal.II finite element library.

5.6.1 One-dimensional tests

We now consider three classical one-dimensional Riemann problems taken from [49]. These tests are also used to illustrate some typical wave patterns resulting from the solution of the Riemann problems. Test 1 is a modified version of Sod's problem [57]. This is a very mild test and its solution consists of a left sonic rarefaction wave, a right travelling contact discontinuity, and a right travelling shock wave. This test is useful for assessing the ability of numerical methods to capture the correct entropy admissible solution. Test 2 is a very severe test problem which is designed to assess the robustness and accuracy of numerical methods. This test is actually the left half of the blast wave problem of Woodward and Colella [58]. The solution contains

a left rarefaction wave, a contact surface, and a strong shock wave of shock Mach number 198. Test 3 is also a very challenging problem associated with the collision of two strong shocks and its solution consists of three strong discontinuities travelling to the right: a very slow travelling left shock wave, a contact discontinuity, and a right shock wave. Table 5.1 shows the left ($x < x_0$) and right ($x > x_0$) constant states for all three tests in terms of primitive variables. The computational domain is $\Omega = (0, 1)$ and the final time of the computation is denoted by T. We take $\gamma = 1.4$ for the adiabatic gas constant in all these tests. The computations are done using

Table 5.1
Data for one-dimensional Riemann problems.

Test	T	x_0	ρ_l	u_l	p_l	ρ_r	u_r	p_r
1	0.200	0.300	1.000	0.750	1.000	0.125	0.000	0.100
2	0.012	0.500	1.000	0.000	1000	1.000	0.000	0.010
3	0.035	0.400	5.99924	19.5975	460.894	5.99242	-6.19633	46.095

P_3 discontinuous finite elements on a uniform mesh composed of 200 cells and the time stepping is performed with explicit RK4. The computational parameters are reported in Table 5.2. The results for each test case are shown in Figures 5.1, 5.2,

Table 5.2
Entropy viscosity parameters for one-dimensional Riemann problems.

Test	CFL	ω	δ	C_E	$C_{\max}$	$\mathcal{P}_\rho$	$\mathcal{P}_T$
1	0.33	0.5	1.0	1.0	$0.30/p$	0.15	0.15
2	0.33	0.5	1.0	1.0	$0.40/p$	0.15	0.15
3	0.33	0.5	1.0	1.0	$0.45/p$	0.15	0.15

and 5.3. We observe that the results are non-oscillatory even though we used cubic finite elements. We have verified that the computed numerical solutions are in fact the entropy admissible solutions.

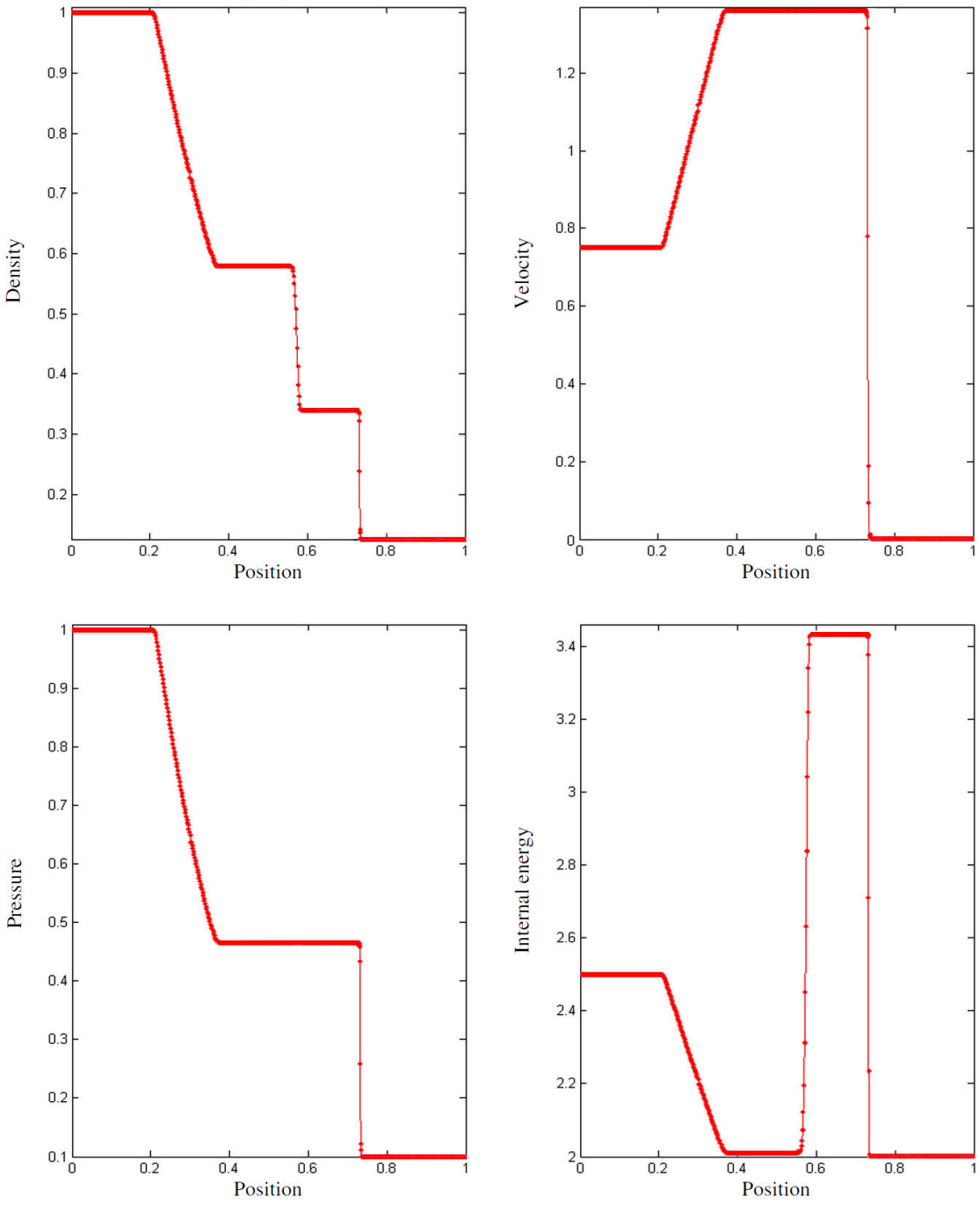

Fig. 5.1. 1D Riemann problem for Euler equations. Test 1. P_3 solution at time $T = 0.2$.

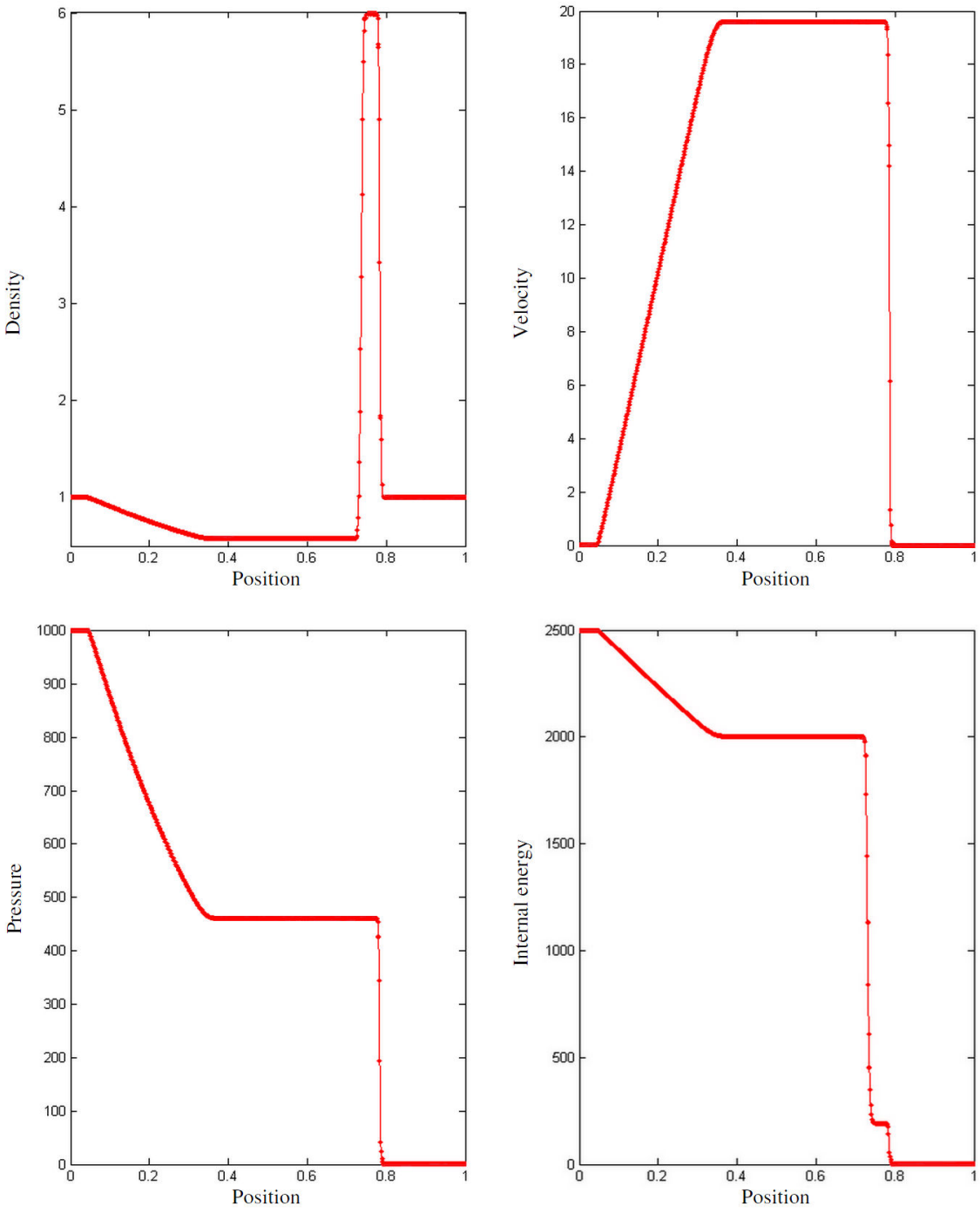

Fig. 5.2. 1D Riemann problem for Euler equations. Test 2. P_3 solution at time $T = 0.012$.

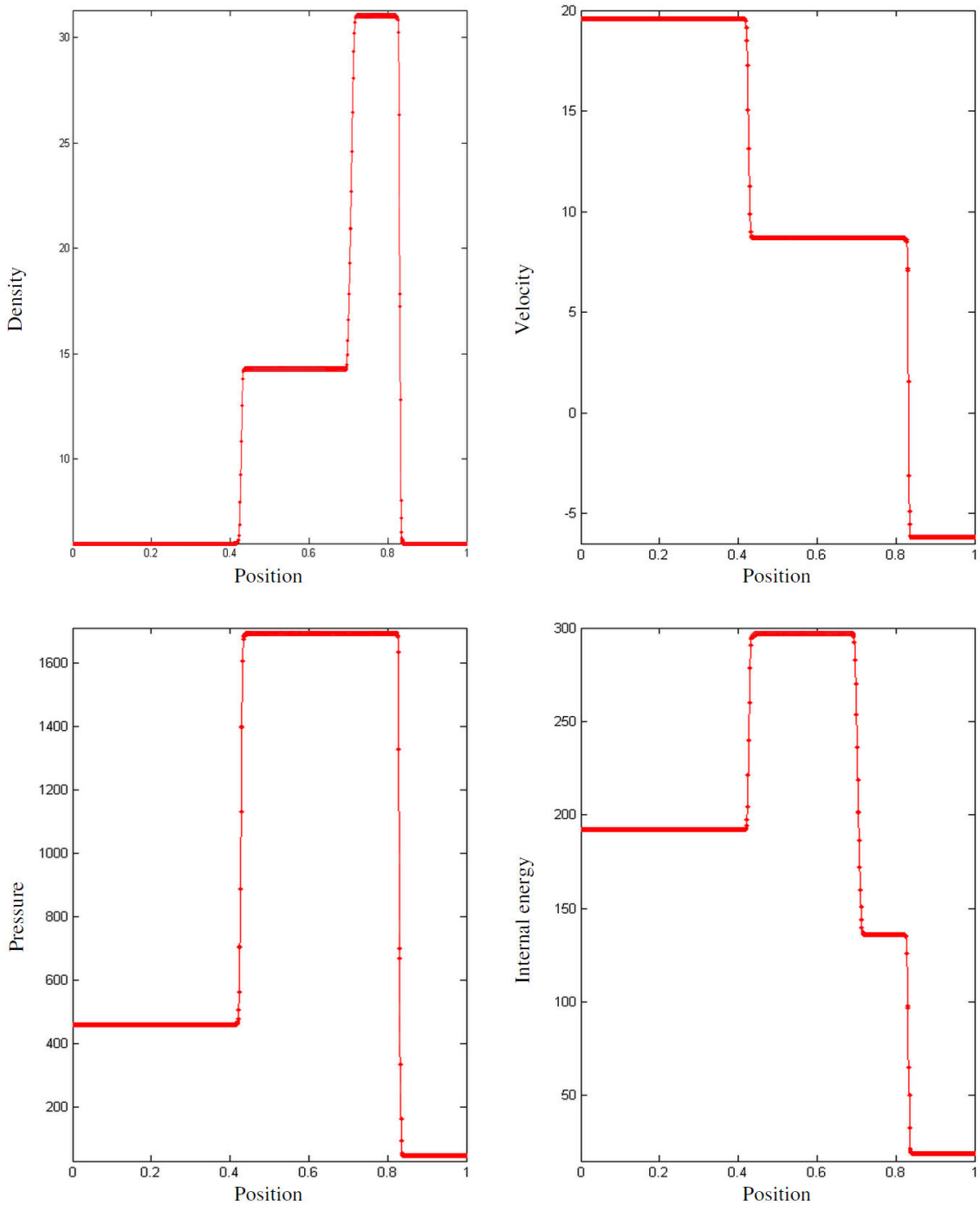

Fig. 5.3. 1D Riemann problem for Euler equations. Test 3. P_3 solution at time $T = 0.035$.

Figures 5.4, 5.5, and 5.6 show the graphs of density and the related entropy-based artificial dynamic viscosity field. The dynamic viscosity is mainly focused on shocks as expected. We also observe small peaks of viscosity in the regions of contact discontinuities.

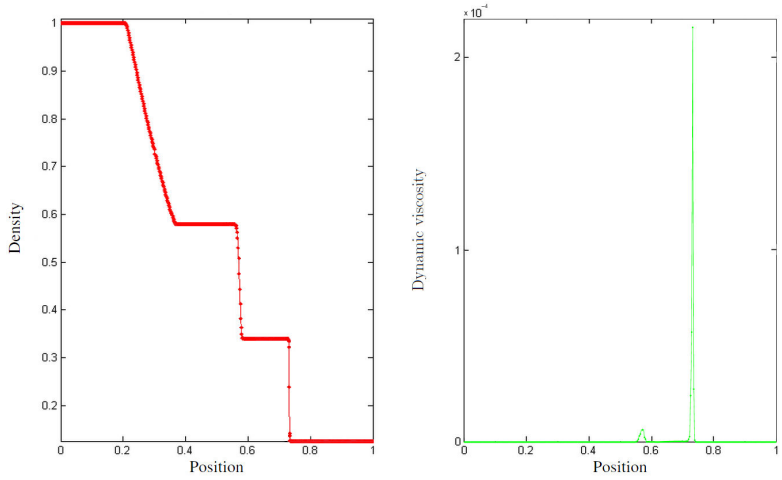

Fig. 5.4. 1D Riemann problem for Euler equations. Test 1. Density and dynamic viscosity at time $T = 0.2$.

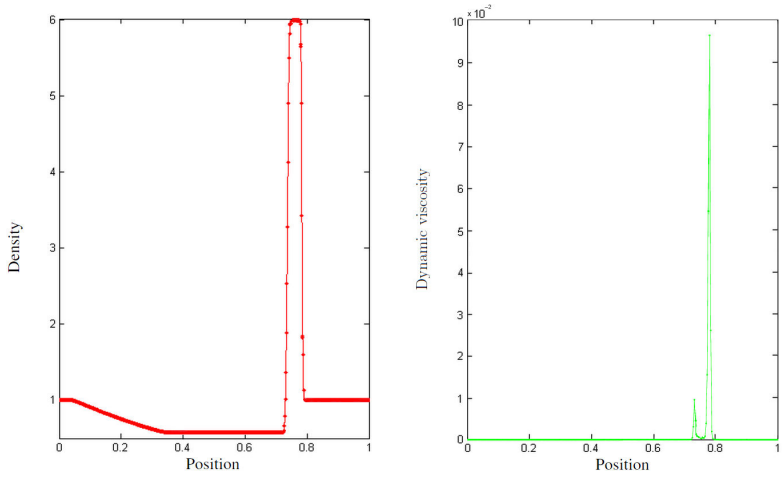

Fig. 5.5. 1D Riemann problem for Euler equations. Test 2. Density and dynamic viscosity at time $T = 0.012$.

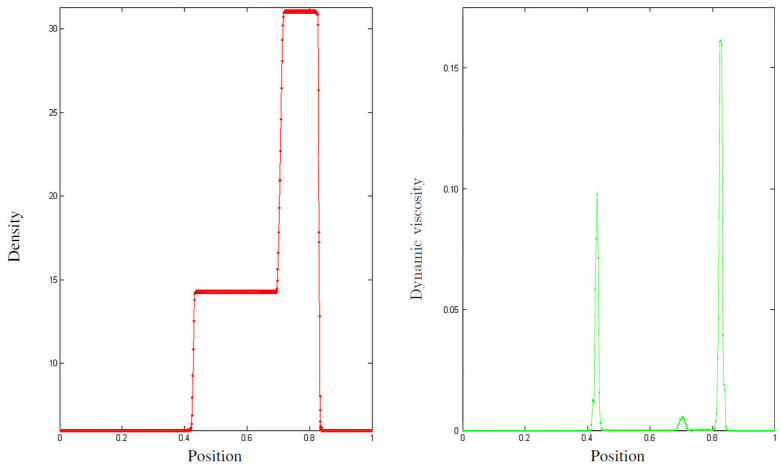

Fig. 5.6. 1D Riemann problem for Euler equations. Test 3. Density and dynamic viscosity at time $T = 0.035$.

5.6.2 Two-dimensional tests

The method is now tested in two space dimensions on three classical benchmark problems.

Riemann problem :

We consider the Riemann problem No. 12, see [59]. It is a two-dimensional problem developing both fine structures of the solution and complex localized structures including shock waves and contact discontinuities as well. The computational domain is $\Omega = (0,1)^2$ and the final time of the computation is $T = 0.2$. The adiabatic gas constant is $\gamma = 1.4$ and the initial data is defined as follows:

$$\begin{aligned} &\rho = 4/5, \quad \mathbf{u} = (0,0), && p = 1, \quad x \in (0.0, 0.5); y \in (0.0, 0.5), \\ &\rho = 1, \quad \mathbf{u} = (3/\sqrt{17}, 0), && p = 1, \quad x \in (0.0, 0.5); y \in (0.5, 1.0), \end{aligned}$$

$$\begin{aligned} &\rho = 1, && \mathbf{u} = (0, 3/\sqrt{17}), && p = 1, && x \in (0.5, 1.0); y \in (0.0, 0.5), \\ &\rho = 17/32, && \mathbf{u} = (0, 0), && p = 2/5, && x \in (0.5, 1.0); y \in (0.5, 1.0). \end{aligned} \tag{5.42}$$

The control parameters of the entropy viscosity method are $\omega = 0.5$, $\delta = 1.0$, $C_E = 1.0$, $C_{\max} = 0.5/p$, $\mathcal{P}_\rho = 0.25$, and $\mathcal{P}_T = 0.25$. The time stepping is implemented using RK4 with CFL $= 0.25$. The computations are done with Q_1, Q_2, and Q_3 discontinuous finite elements on a uniform grid composed of $4^7 = 16384$ quadrangular cells. The total number of scalar degrees of freedom for the Q_1, Q_2, and Q_3 approximations are 65536, 147456, and 262144, respectively, i.e. $16384 \cdot (p+1)^2$. The transparent boundary conditions for this test problem are imposed implicitly on Γ using the DG numerical fluxes.

We show in Figure 5.7 the density fields at time $T = 0.2$ for the Q_1 and Q_2 approximations. The Q_3 density along with respective dynamic viscosity field at time $T = 0.2$ are shown in Figure 5.8. The results compare well with those from [59, 60]. The shock waves and the fine structures that develop behind them are very well described. The method behaves well as the polynomial degree p of the approximation increases. The dynamic viscosity field is sharply concentrated on the shocks. Figure 5.8 reveals some weak instability of the contact surfaces, also known as the Kelvin–Helmholtz instability.

Mach 3 supersonic flow over a forward – facing step :

We now focus our attention on the flow past a forward-facing step in a wind tunnel at Mach 3. This benchmark test has been proposed in [61] and has been popularized by Woodward and Colella's extensive study [58]. The geometry of the computational domain Ω is shown in Figure 5.9 and the initial data along with inflow boundary conditions are specified in terms of primitive variables as follows:

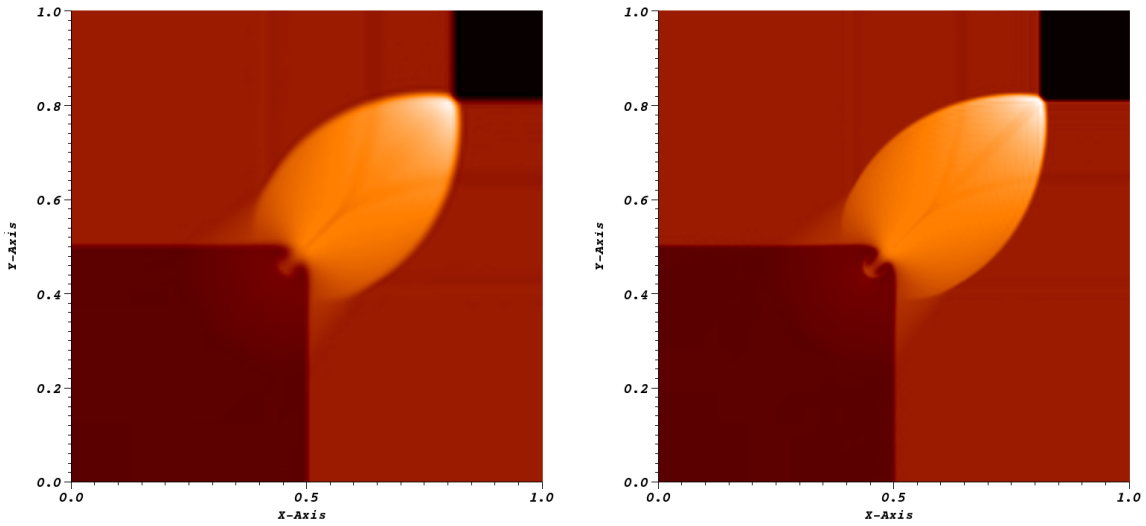

Fig. 5.7. Riemann problem No. 12 for Euler equations. Q_1 (left) and Q_2 (right) density fields at time $T = 0.2$.

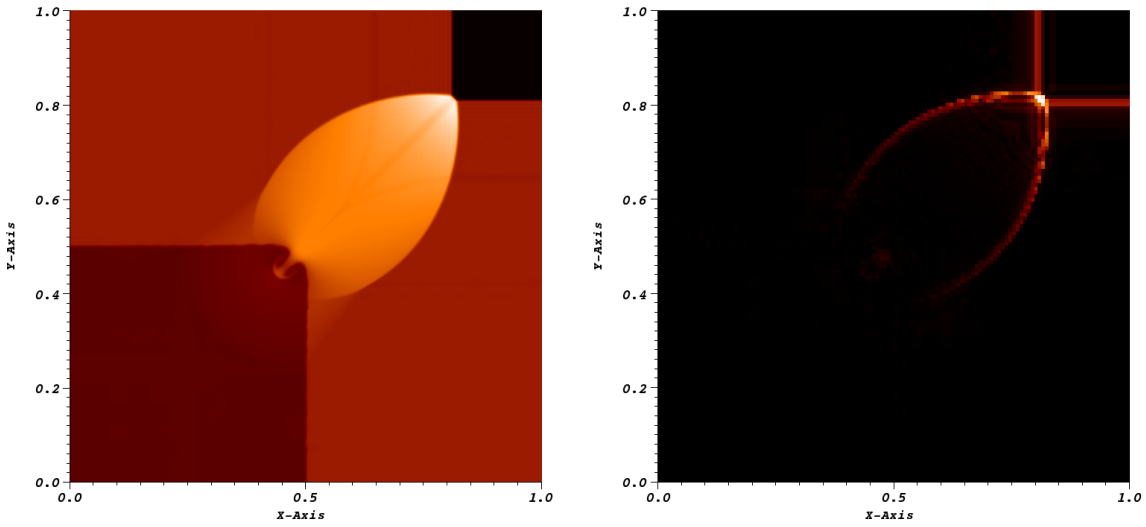

Fig. 5.8. Riemann problem No. 12 for Euler equations. Q_3 density (left) and dynamic viscosity (right) at time $T = 0.2$.

$$\left.\begin{array}{l}(\rho,\mathbf{u},p)^{\mathrm{T}}(x,y,0)\\(\rho,\mathbf{u},p)^{\mathrm{T}}(0,y,t)\end{array}\right\}=(1.4,(3.0,0.0),1.0)^{\mathrm{T}} \tag{5.43}$$

The outflow boundary at $x = 3$ is free and slip boundary conditions $\mathbf{u}\cdot\mathbf{n} = 0$ are specified on the solid walls of the tunnel, $\mathbf{n}$ being an outward unit normal vector to the boundary Γ of domain Ω.

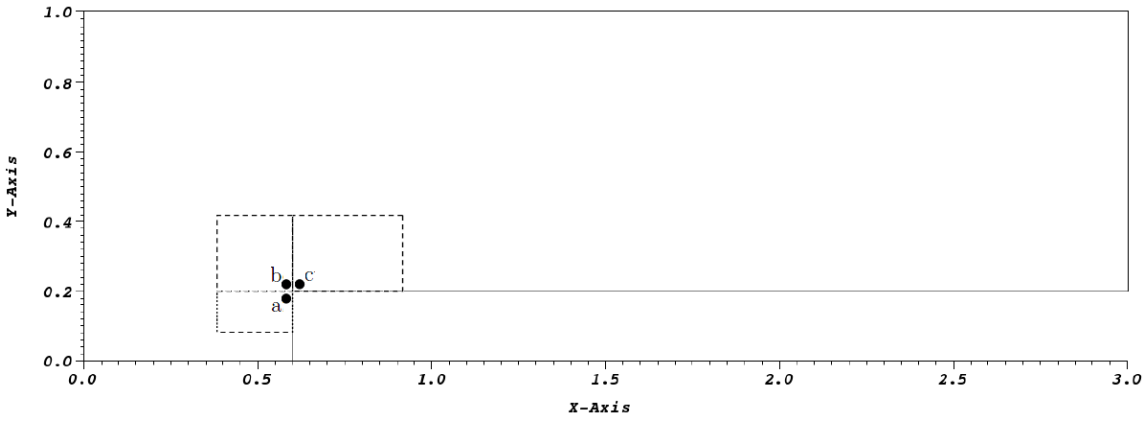

Fig. 5.9. Mach 3 problem for Euler equations. Geometry of domain.

The corner of the step is a singular point in the flow. Using the DG approximation implies three degrees of freedom to be assigned to that particular point, see Figure 5.9. DoF "b" is supposed to be a free stream point with no restrictions applied from the boundaries, whereas DoFs "a" and "c" logically belong to the vertical and horizontal walls of the wind tunnel, respectively. We have observed that better results are obtained by assigning the numerical solution computed at DoF "b" to DoFs "a" and "c".

We solve this problem by using Q_1 discontinuous finite elements and an adaptive mesh refinement strategy available in the deal.II library.

The local Kelly error indicator for the density [62] is employed in this computation:

$$\eta_K^2 = h_K \oint_{\partial K} [\![\nabla \rho_h \cdot \mathbf{n}]\!]^2 \, d\sigma, \tag{5.44}$$

and we limit the number of refinement levels to 7. This error indicator tries to approximate the error per cell by integration of the jump of the gradient of the solution along the faces of each cell. It gives good indications for mesh refinement, but the error estimate is not to be trusted. For higher order trial spaces the integrals computed here tend to zero faster than the error itself, thus ruling out the values as actual error estimators.

The refining and coarsening are done so that 30% of the cells with the largest error indicators are refined, whereas 10% of the cells with the smallest error indicators are coarsened. Mesh adaption is done at every time step. The control parameters of the entropy viscosity method are $\omega = 0.5$, $\delta = 1.0$, $C_E = 1.0$, $C_{\max} = 0.5/p$, $\mathcal{P}_\rho = 0.05$, and $\mathcal{P}_T = 0.1$. The time stepping is performed using RK4 with CFL $= 0.25$.

The graphs of density and associated dynamic viscosity fields at time $T = 4$ are shown in Figures 5.10 and 5.11, respectively. The mesh obtained at time $T = 4$ is shown in Figure 5.12. This solution agrees, at least in the "visual norm", with other reference solutions that can be found in the literature. The contact discontinuity emerging from the three-shocks interaction point is presented in the simulation and is captured satisfactorily, considering the small number of elements (25500) used in this computation. We observe a spurious Mach stem at the bottom wall of the tunnel. This artifact is due to the entropy layer produced at the bottom wall by the corner singularity. The phenomenon is well explained in [34, p.217]. It can be removed either by applying special boundary conditions for the cells close to the corner to avoid the effect from the singularity or by aggressive refining the mesh at the corner. Another approach could be to round off the nonphysically sharp corner and allow adaptive mesh refinement to follow the smooth structure of the rounded corner.

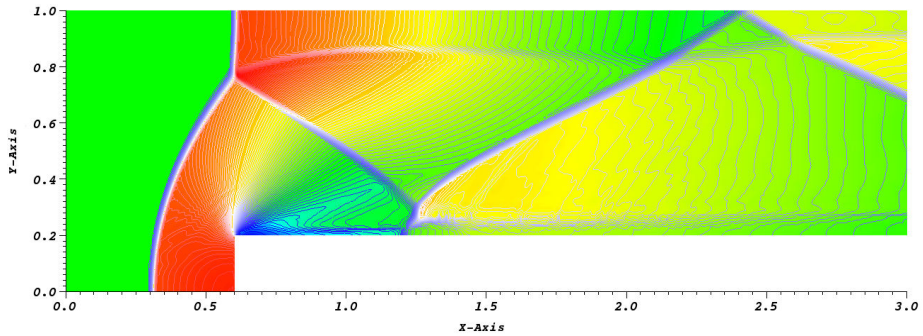

Fig. 5.10. Mach 3 problem for Euler equations. Q_1 density at time $T = 4$.

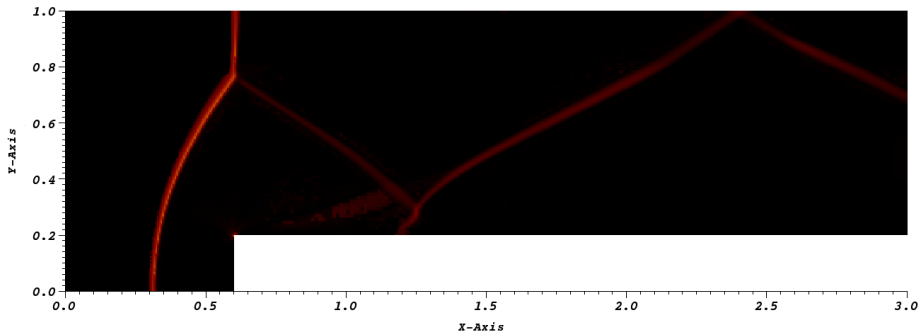

Fig. 5.11. Mach 3 problem for Euler equations. Dynamic viscosity field at time $T = 4$.

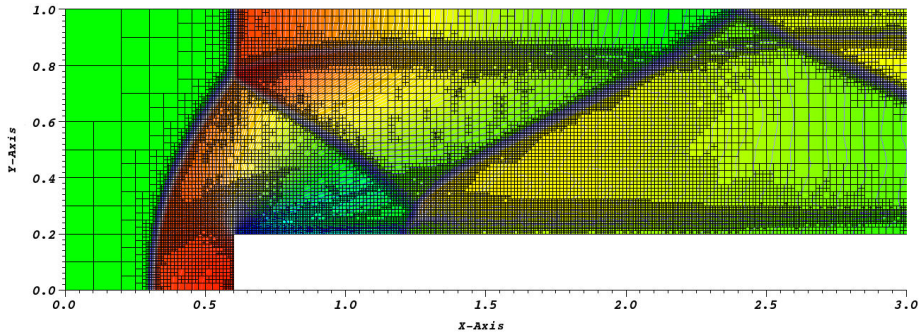

Fig. 5.12. Mach 3 problem for Euler equations. Adaptive mesh at time $T = 4$.

Circular explosion on a square mesh :

We finish the series of numerical examples by considering a circular explosion in two space dimensions. This numerical test has been proposed in [49] and centres around the production of an unstable contact layer. The computational domain is $\Omega = (0,1)^2$ and the final time of the computation is $T = 0.25$. The adiabatic gas constant is $\gamma = 1.4$ and the initial data is shown in Figure 5.13.

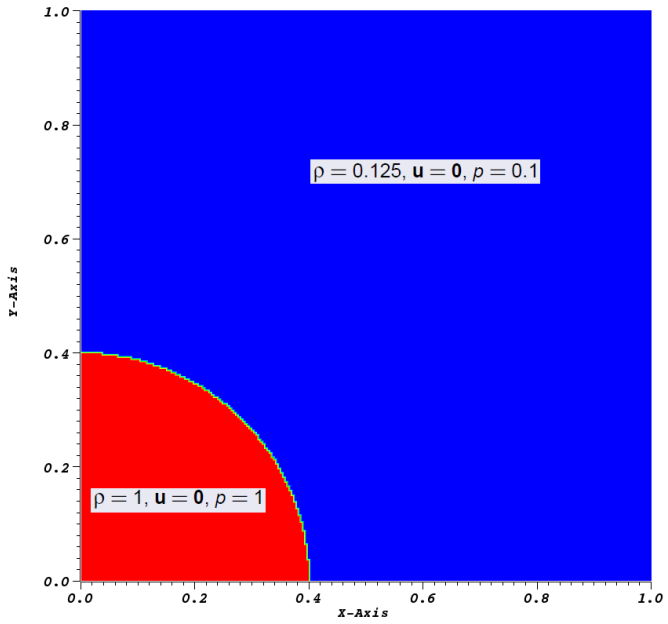

Fig. 5.13. Circular explosion. Initial data.

The boundaries at $x = 0$ and $y = 0$ are symmetric and the boundaries at $x = 1$ and $y = 1$ are transparent. We use the following control parameters: $\omega = 0.5$, $\delta = 1.0$, $C_E = 1.0$, $C_{\max} = 0.5/p$, $\mathcal{P}_\rho = 0.1$, and $\mathcal{P}_T = 0.1$. The time stepping is implemented using RK4 with CFL $= 0.20$. The computation involves Q_1 discontinuous finite elements on an adaptive grid with approximately 17000 cells at the end. We show numerical density and entropy dynamic viscosity fields in Figures 5.14 and 5.15, respectively. Figure 5.16 shows three-dimensional image of density.

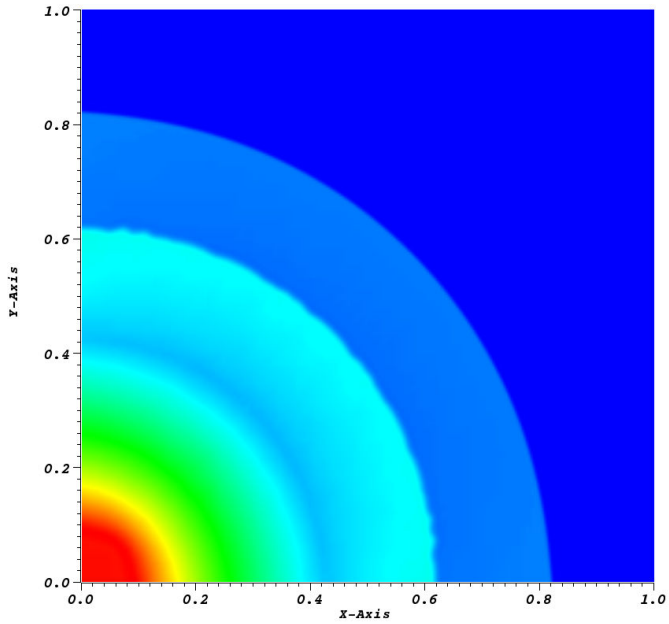

Fig. 5.14. Circular explosion. Q_1 density at time $T = 0.25$.

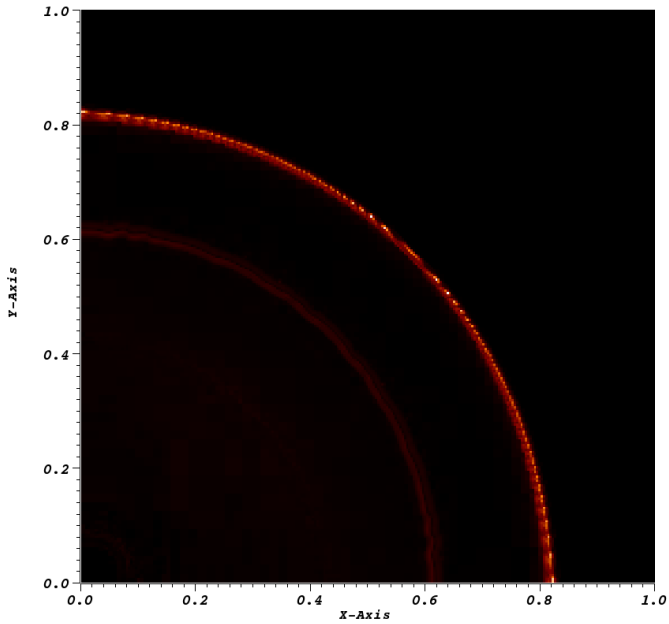

Fig. 5.15. Circular explosion. Dynamic viscosity field at time $T = 0.25$.

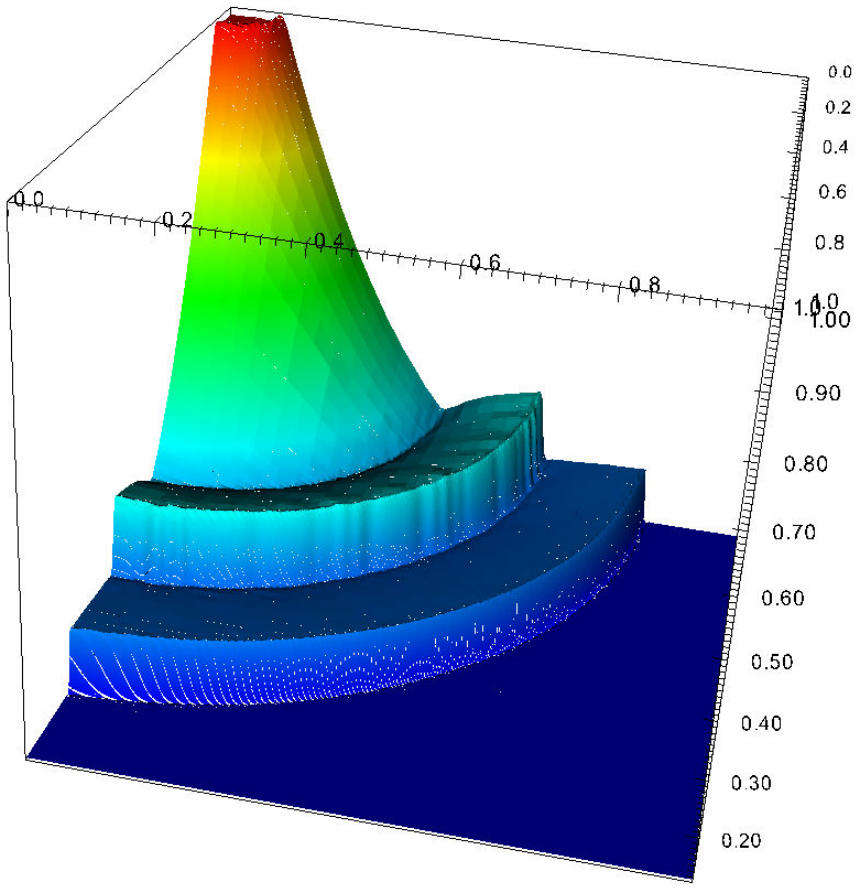

Fig. 5.16. Circular explosion. 3D image of Q_1 density at time $T = 0.25$.

6. CONCLUSIONS

We have developed and implemented a cutting-edge numerical technique which combines higher-order discontinuous Galerkin finite element methods with an efficient and robust high order shock-capturing and stabilization procedure based upon the entropy-based artificial viscosity approximation. The new method can be successfully applied to a broad class of time-dependent nonlinear hyperbolic problems supplemented with an auxiliary entropy inequality. The arbitrarily order accuracy of DG spatial discretizations together with reliable and precise time integration schemes enable resolution of the small-scale features of most complex unsteady flows. Due to the "intelligent" distribution of artificial viscosity between smooth and discontinuous flow regions, the new technique keeps the fine structures of smooth flows essentially untouched and simultaneously provides sharp and stable modulations of the numerical solution across discontinuities.

The implementation of the method is relatively simple on unstructured grids with arbitrary polynomial DG approximations. Flux or slope limiting procedures in the so-called "trouble cells" are avoided in the presence of stabilizing artificial viscosity. This characteristic, along with other favorable properties, makes this method extremely flexible in the sense of easily handling a wide variety of element types and mesh topologies as well as a number of adaptive techniques.

The method we have developed extends the entropy viscosity approach in a more systematic way into the discontinuous Galerkin finite element framework. The interelement jumps of the entropy flux are added to the definition of artificial viscosity, which makes the entropy viscosity field more concentrated across discontinuities. Our construction of DG inviscid numerical fluxes differs slightly from that associated with standard versions of the DG method, while our construction of DG viscous numerical fluxes differs significantly. The JLG numerical fluxes (after Jean-Luc Guermond) [50] have been introduced.

Our technique has been shown to perform well on many benchmark test problems. While performing calculations not presented here, we have observed that the DG version of the entropy viscosity method is at least as efficient as CG and spectral element versions on the same class of problems. Since DG methods are routinely used for radiative transfer calculations in the high energy density physics regime [63], the use of DG finite elements provides a natural setting for coupling the hydrodynamics and radiation transport equations.

The overall performance of our method leads us to conclude that it is a very valuable alternative to existing techniques for solving the Euler equations and other systems of nonlinear hyperbolic conservation laws.

REFERENCES

[1] P. G. LeFloch, J.-G. Liu, Generalized monotone schemes, discrete paths of extrema, and discrete entropy conditions, Math. Comp. 68 (227) (1999) 1025–1055.

[2] P.-L. Lions, P. E. Souganidis, Convergence of MUSCL and filtered schemes for scalar conservation laws and Hamilton-Jacobi equations, Numer. Math. 69 (4) (1995) 441–470.

[3] B. Popov, O. Trifonov, One-sided stability and convergence of the Nessyahu-Tadmor scheme, Numer. Math. 104 (4) (2006) 539–559.

[4] H. Yang, On wavewise entropy inequalities for high-resolution schemes. I. The semidiscrete case, Math. Comp. 65 (213) (1996) 45–67.

[5] H. Yang, On wavewise entropy inequality for high resolution schemes. II. Fully discrete MUSCL schemes with exact evolution in small time, SIAM J. Numer. Anal. 36 (1) (1999) 1–31.

[6] I. Christov, B. Popov, New non-oscillatory central schemes on unstructured triangulations for hyperbolic systems of conservation laws, J. Comput. Phys. 227 (11) (2008) 5736–5757.

[7] J.-L. Guermond, R. Pasquetti, Entropy-based nonlinear viscosity for fourier approximations of conservation laws, C. R. Math. Acad. Sci. Paris 346 (2008) 801–806.

[8] H. Versteeg, W. Malalasekera, An Introduction to Computational Fluid Dynamics: The Finite Volume Method, 2nd Edition, Prentice Hall, Upper Saddle River, NJ, 2007.

[9] B. Cockburn, G. E. Karniadakis, C.-W. Shu, Discontinuous Galerkin Methods: Theory, Computation, and Applications, 1st Edition, Springer, New York, NY, 2000.

[10] W. Reed, T. Hill, Triangular mesh methods for the neutron transport equation, Tech. Rep. LA-UR-73-479, Los Alamos Scientific Laboratory, Los Alamos, NM, (1973).

[11] B. Cockburn, C. Shu, The local discontinuous Galerkin method for time-dependent convection-diffusion systems, SIAM J. Numer. Anal. 35 (1998) 2440–2463.

[12] F. Bassi, S. Rebay, A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations, J. Comput. Phys. 131 (2) (1997) 267–279.

[13] C. Baumann, J. Oden, A discontinuous hp finite element method for the Euler and Navier-Stokes equations, Int. J. Numer. Methods Engrg. 31 (1999) 79–95.

[14] C. Baumann, J. Oden, A discontinuous hp finite element method for convection-diffusion problems, Comput. Methods Appl. Mech. Engrg. 175 (3-4) (1999) 311–341.

[15] J. Oden, I. Babuška, C. Baumann, A discontinuous hp finite element method for diffusion problems, J. Comput. Phys. 146 (2) (1998) 491–519.

[16] B. Cockburn, C.-W. Shu, Tvb runge-kutta local projection discontinuous galerkin finite element method for conservation laws ii: general framework, Math. Comp. 52 (1989) 411–435.

[17] B. Cockburn, S.-Y. Lin, C.-W. Shu, Tvb runge-kutta local projection discontinuous galerkin finite element method for conservation laws iii: one dimensional systems, J. Comput. Phys. 84 (1989) 90–113.

[18] B. Cockburn, S. Hou, C.-W. Shu, The runge-kutta local projection discontinuous galerkin finite element method for conservation laws iv: the multidimensional case, Math. Comp. 54 (1990) 545–581.

[19] B. Cockburn, C.-W. Shu, The runge-kutta discontinuous galerkin method for conservation laws v: multidimensional systems, J. Comput. Phys. 141 (1998) 199–224.

[20] B. Cockburn, C.-W. Shu, The runge-kutta local projection p1-discontinuous galerkin finite element method for scalar conservation laws, ESAIM: Mathematical Modelling and Numerical Analysis 25 (1991) 337–361.

[21] A. Harten, B. Engquist, S. Osher, S. R. Chakravarthy, Uniformly high-order accurate essentially nonoscillatory schemes. III, J. Comput. Phys. 71 (2) (1987) 231–303.

[22] S. Osher, C.-W. Shu, High-order essentially nonoscillatory schemes for Hamilton-Jacobi equations, SIAM J. Numer. Anal. 28 (4) (1991) 907–922.

[23] C. Shu, S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes, ii, J. Comput. Phys. 83 (1989) 32–78.

[24] G.-S. Jiang, C.-W. Shu, Efficient implementation of weighted eno schemes, J. Comput. Phys. 126 (1996) 202–228.

[25] C. Hu, C.-W. Shu, Weighted essentially non-oscillatory schemes on triangular meshes, J. Comput. Phys. 150 (1999) 97–127.

[26] S. Zhang, S. Jiang, C.-W. Shu, Development of nonlinear weighted compact schemes with increasingly higher order accuracy, J. Comput. Phys. 227 (2008) 7294–7321.

[27] T. Zhou, Y. Li, C.-W. Shu, Numerical comparison of weno finite volume and runge-kutta discontinuous galerkin methods, SIAM J. Sci. Comput 16 (2001) 145–171.

[28] J. Qiu, C.-W. Shu, Hermite weno schemes and their application as limiters for runge-kutta discontinuous galerkin method: one dimensional case, J. Comput. Phys. 193 (2003) 115–135.

[29] J. von Neumann, R. D. Richtmyer, A method for the numerical calculation of hydrodynamic shocks, J. Appl. Phys. 21 (1950) 232–237.

[30] J. Smagorinsky, General circulation experiments with the primitive equations, part i: the basic experiment, Monthly Wea. Rev. 91 (1963) 99–152.

[31] O. Ladyženskaja, New equations for the description of motion of viscous incompressible fluds and solvability in the large of boundary value problems for them, in: O. Ladyženskaja (Ed.), Proc. of the Stheklov Institute of Mathematics, number 102 (1967), Boundary value problems of mathematical physics. V, American Mathematical Society (AMS), Providence, RI, 1970.

[32] O. Ladyženskaja, Modification of the Navier–Stokes equations for large velocity gradients, in: O. Ladyženskaja (Ed.), Seminars in Mathematics V.A. Stheklov Mathematical Institute, Volume 7, Boundary value problems of mathematical physics and related aspects of function theory, Part II, Consultant Bureau, New York, NY, 1970.

[33] C.-W. Shu, S. Osher, Efficient implementation of essentially nonoscillatory shock-capturing schemes, J. Comput. Phys. 77 (2) (1988) 439–471.

[34] B. Cockburn, C.-W. Shu, The Runge-Kutta discontinuous Galerkin method for conservation laws. V. Multidimensional systems, J. Comput. Phys. 141 (2) (1998) 199–224. doi:10.1006/jcph.1998.5892.
URL http://dx.doi.org/10.1006/jcph.1998.5892

[35] B. Cockburn, C. Johnson, C. Shu, E. Tadmor, Advanced Numerical Approximation of Nonlinear Hyperbolic Equations, Vol. 1697 of Lecture Notes in Mathematics, Springer, New York, NY, 1998.

[36] J. J. W. van der Vegt, H. van der Ven, Space-time discontinuous galerkin finite element method with dynamic grid motion for inviscid compressible flows: I. general formulation, Journal of Computational Physics 182 (2) (2002) 546 – 585. doi:DOI: 10.1006/jcph.2002.7185.
URL http://www.sciencedirect.com/science/article/pii/S0021999102971858

[37] R. Hartmann, Adaptive discontinuous Galerkin methods with shock-capturing for the compressible Navier-Stokes equations, Internat. J. Numer. Methods Fluids 51 (9-10) (2006) 1131–1156. doi:10.1002/fld.1134.
URL http://dx.doi.org/10.1002/fld.1134

[38] F. Bassi, S. Rebay, G. Mariotti, S. Pedinotti, M. Savini, A high-order accurate discontinuous finite element method for inviscid and viscous turbomachinery flows, in: R. Decuypere, G. Dibelius (Eds.), Proceedings of the 2nd European Conference on Turbomachinery Fluid Dynamics and Thermodynamics, 1997, pp. 99–109.

[39] G. Barter, D. Darmofal, Shock capturing with pde-based artificial viscosity for dgfem: Part i, formulation, J. Comput. Phys. 229 (2010) 1810–1827.

[40] P. Persson, J. Peraire, Sub-cell shock capturing for discontinuous galerkin methods, Vol. AIAA-2006-112 of 44th AIAA Aerospace Sciences Meeting, AIAA, 2006.

[41] E. Burman, On nonlinear artificial viscosity, discrete maximum principle and hyperbolic conservation laws, BIT 47 (4) (2007) 715–733. doi:10.1007/s10543-007-0147-7.
URL http://dx.doi.org/10.1007/s10543-007-0147-7

[42] G. Barter, D. Darmofal, Shock capturing with higher-order, pde-based artificial viscosity, Miami, FL, 2007.

[43] P. Nelson, On deterministic developments of traffic stream models, Transportation Research 29 (1995) 297–302.

[44] P. Nelson, A kinetic model of vehicular traffic and its associated bimodal equilibrium solutions, Transport Theory and Statistical Physics 24 (1995) 383–409.

[45] J. Anderson, Fundamentals of Aerodynamics, 5th Edition, McGraw-Hill Science/Engineering/Math, New York, NY, 2010.

[46] J. G. Andrews, K. W. Morton, A posteriori error estimation based on discrepancies in an entropy variable, Int. J. Comput. Fluid Dyn. 10 (3) (1998) 183–198. doi:10.1080/10618569808961684.
URL http://dx.doi.org/10.1080/10618569808961684

[47] G. Puppo, Numerical entropy production for central schemes, SIAM Journal on Scientific Computing 25 (4) (2003) 1382–1415. doi:10.1137/S1064827502386712.
URL http://link.aip.org/link/?SCE/25/1382/1

[48] C. Fletcher, Burgers' equation: A model for all reasons, Numerical solutions of Partial Differential Equations, Amsterdam: North-Holland Publ. Co., Amsterdam, 1982.

[49] E. F. Toro, Riemann solvers and numerical methods for fluid dynamics, 2nd Edition, Springer-Verlag, Berlin, 1999.

[50] V. Zingan, J.-L. Guermond, B. Popov, J. Morel, Implementation of the entropy viscosity method with the discontinuous galerkin method, submitted to Computer Methods in Applied Mechanics and Engineering.

[51] E. Luther, Gauss-Jordan Elimination, Acu Publishing, San Diego, CA, 2011.

[52] W. Bangerth, R. Hartmann, G. Kanschat, deal.II – a general purpose object oriented finite element library, ACM Trans. Math. Softw. 33 (4) (2007) 24/1–24/27.

[53] W. Bangerth, G. Kanschat, deal.II Differential Equations Analysis Library, Technical Reference.
URL http://www.dealii.org

[54] G.-S. Jiang, E. Tadmor, Nonoscillatory central schemes for multidimensional hyperbolic conservation laws, SIAM J. Sci. Comput. 19 (6) (1998) 1892–1917.

[55] E. Tadmor, Convergence of spectral methods for nonlinear conservation laws, SIAM J. Numer. Anal. 26 (1) (1989) 30–44.

[56] A. Kurganov, G. Petrova, B. Popov, Adaptive semi-discrete central-upwind schemes for nonconvex hyperbolic conservation laws, SIAM J. Sci. Comput. 29 (6) (2007) 2381–2401.

[57] G. Sod, A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws, J. Comput. Phys. 27 (1978) 1–31.

[58] P. Woodward, P. Colella, The numerical simulation of two-dimensional fluid flow with strong shocks, J. Comput. Phys. 54 (1) (1984) 115–173.

[59] R. Liska, B. Wendroff, Comparison of several difference schemes on 1D and 2D test problems for the Euler equations, SIAM J. Sci. Comput. 25 (3) (2003) 995–1017 (electronic). doi:10.1137/S1064827502402120.
URL http://dx.doi.org/10.1137/S1064827502402120

[60] A. Kurganov, E. Tadmor, Solution of two-dimensional riemann problems for gas dynamics without riemann problem solvers, Numerical Methods for Partial Differential Equations 18 (2002) 584–608.

[61] A. F. Emery, An evaluation of several differencing methods for inviscid fluid flow problems, J. Computational Phys. 2 (1968) 306–331.

[62] D. Kelly, The self-equilibration of residuals and complementary a posteriori error estimates in the finite element method, International Journal for Numerical Methods in Engineering 20 (1984) 1491–1506.

[63] J. E. Morel, E. W. Larsen, Nuclear Computational Science: A Century in Review, Chapter 1: Advances in Discrete-Ordinates Methods, 1st Edition, Springer, New York, NY, 2010.