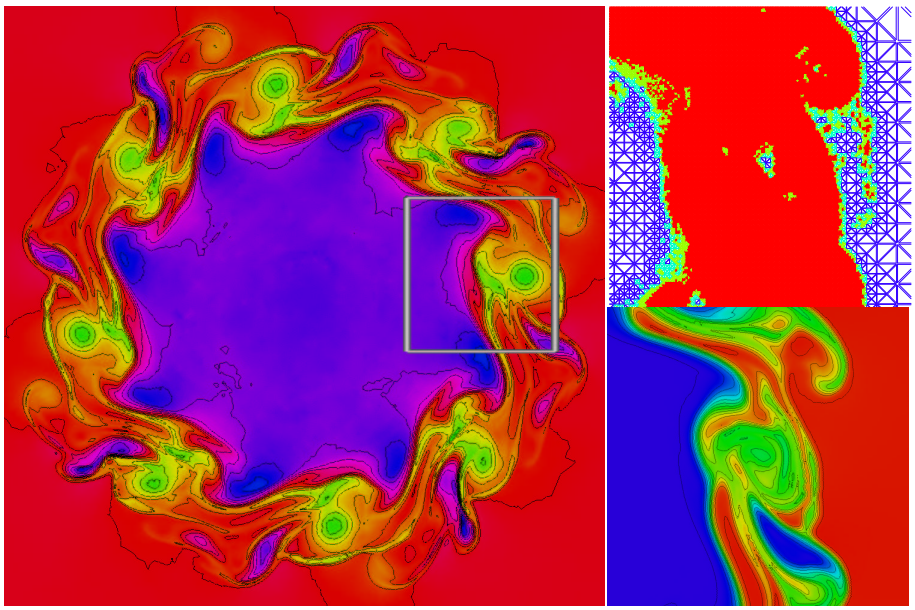
Albert-Ludwigs-Universität Freiburg i. Br.
Fakultät für Mathematik und Physik

Solving the System of Radiation Magnetohydrodynamics

for solar physical simulations in 3d



Andreas Dedner

Dissertation zur Erlangung des Doktorgrades der Fakultät für Mathematik und Physik der Albert-Ludwigs-Universität Freiburg im Breisgau
Betreuer: Prof. Dr. Dietmar Kröner

Abteilung für Angewandte Mathematik
Freiburg im Breisgau, April 2003

Dekan	: Prof. Dr. Rolf Schneider
Referenten	: Prof. Dr. Dietmar Kröner
	: Prof. Dr. Gerald Warnecke, Universität Magdeburg
Datum der Promotion	: 22. September 2003

Picture on title page:
simulation of a circular slip stream. In the purely hydrodynamic setting the interface is unstable with respect to perturbations (Kelvin–Helmholtz instability). By means of a sufficiently strong magnetic field that is tangential to the interface, this instability can be suppressed. For the simulation shown here, the magnetic field is not yet strong enough so that the development of the Kelvin–Helmholtz instabilities can be clearly seen. The large picture shows the density of the fluid. The two smaller pictures show the locally adapted grid (top) and the third component of the magnetic field (bottom) in a small section of the domain in the vicinity of the interface.

Abstract

In this study we present a finite–volume scheme for solving the equations of radiation magnetohydrodynamics in two and three space dimensions. Among other applications this system is used to model the plasma in the solar convection zone and in the solar photosphere. It is a non–linear system of balance laws derived from the Euler equations of gas dynamics and the Maxwell equations; the energy transport through radiation is also included in the model. The starting point of our presentation is a standard explicit first and second order finite–volume scheme on both structured and unstructured grids. We first study the convergence of a finite–volume scheme applied to a scalar model problem for the full system of radiation magnetohydrodynamics. We then present modifications of the base scheme. These make it possible to approximate the system of magnetohydrodynamics with an arbitrary equation of state; they reduce errors due to a violation of the divergence constraint on the magnetic field, and they lead to an improved accuracy in the approximation of solution near an equilibrium state. These modifications significantly increase the robustness of the scheme and are essential for an accurate simulation of processes in the solar atmosphere. For simulations in the solar photosphere, we have to take the radiation intensity into account. A scheme for solving the radiation transport equation is a further focus of this study. We present both analytical results and numerical tests, comparing our scheme with some standard schemes found in the literature. We conclude our presentation with a study of the parallelization strategy for distributed memory computers that we use in our 3d code.

Introduction

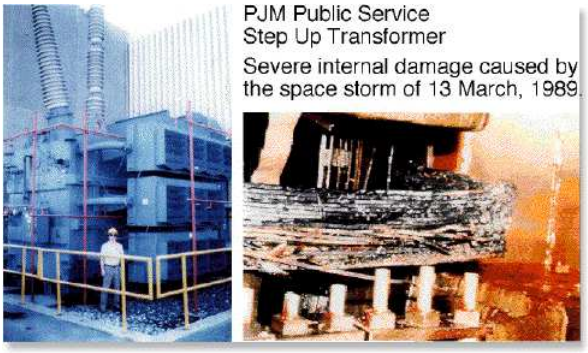
Numerical simulations have become an important tool for studying many different physical and technical problems. Ranging from the formation of galaxies to weather forecasts to the design for parts of complex machinery, the applications are numerous. On the one hand, numerical simulations serve as a tool for the verification of physical theories deduced from observation; on the other hand, they play an important role in reducing development cost in manufacturing. Although the range of applications is extremely broad, the methods used for solving problems numerically have many features in common. This is due to the fact that the physical models used have similar properties. For example, fluid flow in the atmosphere of stars or in car engines can be modeled by very similar systems of equations and can be simulated using very similar numerical methods.

In this study we investigate numerical schemes that can be used to simulate the evolution of a compressible fluid. The governing system of partial differential equations is based on the Euler equations of gas dynamics. Over the last centuries, this system has been the focus of both analytical and numerical studies. A large number of different schemes have been developed and tested using this system. One very successful approach turned out to be the finite–volume framework, and many different schemes based on this approach have been presented. The same methods have also been applied to different extensions of the basic system of Euler equations, including, for example, reactive flow and magnetohydrodynamics. The latter will be the main focus of our study.

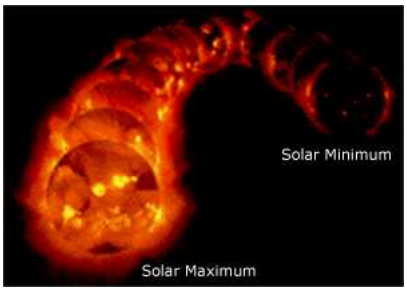
Solar physical applications

The material presented in the following is part of a project financed by the Deutsche Forschungsgemeinschaft (DFG) aimed at deriving and analytically justifying numerical methods for studying fluid flow in the solar atmosphere. The development of many of the methods is a direct consequence of the interaction between members of our group here in Freiburg (Dietmar Kröner, Christian Rohde, Matthias Wesenberg, and myself) and solar physicists (Manfred Schüssler and Peter Vollmöller from the Max–Planck Institute for Aeronomie in Kattlenburg–Lindau), whose ideas greatly influenced our work. Many of the problems discussed here occur only if the methods are applied not to academic test cases, but to realistic settings. Therefore, the discussions with the solar physicists and their help in developing and testing the numerical methods influenced the direction in which our work progressed.

Although a variation in solar activity has a strong impact on life here on earth, a thorough understanding of the physical processes behind these phenomena is still the subject of research all over the world:



http://image.gsfc.nasa.gov/poetry/storm0/black1.html



http://sec.noaa.gov/SWN/index.html

Storms are usually responsible for the losses of electricity we endure, but did you know that "storms" as far away as the sun are capable of knocking out large areas of electric service? Amazingly, the sun is capable of not only disrupting electrical power, but also short wave radio, television and telegraph signals, navigational equipment (GPS and LORAN), defense (military) early warning radar systems, the climate, and can even knockout our communication satellites in space.

http://image.gsfc.nasa.gov/poetry/storm0/black1.html

Early records of sunspots indicate that the Sun went through a period of inactivity in the late 17th century. Very few sunspots were seen on the Sun from about 1645 to 1715. [...] This period of solar inactivity also corresponds to a climatic period called the "Little Ice Age" when rivers that are normally ice-free froze and snow fields remained year-round at lower altitudes. There is evidence that the Sun has had similar periods of inactivity in the more distant past. The connection between solar activity and terrestrial climate is an area of on-going research.

http://sec.noaa.gov/SWN/index.html

Since the possibilities for direct observation of physical processes below the solar surface are limited, numerical simulations play an important role in obtaining a clearer understanding of solar phenomena. A further example of a solar phenomena not yet fully understood is the eleven year cycle in which the number of sun spots on the solar surface increase and decrease. One difficulty is that the filaments at the boundary of the sun spots are made up of *magnetic fluxtubes* that are formed about $2 \cdot 10^5$ kilometers below the solar surface in the lower convection zone of the sun. In this region direct observation is hardly possible so that the formation and evolution of the fluxtubes has to be studied by means of numerical simulations. Although the presentation here is far more general and such solar phenomena are not the immediate focus, the application of our method to problems in solar physics has been a constant motivation.

Mathematical model

The mathematical model consists of a system of balance laws combining the equations of magnetohydrodynamics (MHD) and the radiation transport (RT) equation. The MHD equations are a non–linear system of eight conservation laws; the energy transport through radiation leads to an additional source term that is non–local in space.
The MHD equations describe the evolution of an electrically conductive plasma in the presence of magnetic fields and combine the Euler equations of gas dynamics and the

Maxwell equations. The latter also introduce a constraint equation on the divergence of the magnetic field. In the solar atmosphere the force of gravity plays an important role and is included in our model via source terms. To perform the simulations, we have to prescribe suitable initial conditions for the fluid. These often consist of a perturbation of a stratified and static background atmosphere. One intrinsic problem of simulations with this type of initial data is the size of the computational domain. The setting allows for no physical boundaries, and the construction of suitable artificial boundary conditions that can be used in numerical simulations is no easy undertaking.
The main difficulty in approximating the radiation field is the high dimensionality of the problem and the propagation speed of the radiation, which is several orders of magnitude above the speed of the fluid. In our model we deal with the second problem by assuming an instantaneous radiation equilibrium. We have thus removed the different time scales, but we introduce a non–local dependency into our problem, which we have to cope with in our numerical scheme. The high dimensionality of the radiation intensity — it depends on space, time, propagation angle, and frequency — forces us to construct a very efficient solver to compute the radiation field.

Numerical scheme

We use a first and second order finite–volume scheme on locally adapted structured and unstructured grids. We have implemented this method in one, two, and three space dimensions, using both Cartesian and triangular grids in 2d and hexahedral and tetrahedral grids in 3d. To increase the efficiency of the scheme, we make use of parallelization strategies including distributed memory parallelization with dynamic load balancing. Most of the methods presented in this study are, however, not restricted to use with a finite–volume scheme and have been constructed to be easily added to any existing method for solving the system of magnetohydrodynamics. We have already pointed out that the application to solar physical problems serves as a motivation for the development and test of the scheme; the presentation, however, is kept at a far more general level. For example, the correction method used to compute solutions near an equilibrium state can be used for many types of atmospheric flow, or even for totally different applications where the problem of balancing source terms and flux gradients plays a crucial role.
An important consideration for the development of our methods is their simple implementation within the framework of an existing numerical scheme, which we modify as little as possible. The complexity of our applications also requires an efficient solution algorithm. Consequently, none of the modifications should lead to an increase in the computational cost. Furthermore, we try to reduce the number of free parameters as much as possible; if available, we use an analytically motivated choice for the parameters, otherwise we try to find suitable values by means of numerical tests.

Analytical justification

There are very few analytical results for complex non–linear coupled systems of the type studied here. Even for the MHD system without radiation very little is known

concerning the existence and uniqueness of solutions for general initial data. Consequently, convergence analysis for numerical schemes is not yet available. One approach often used in the analysis of complex systems is to reduce the complexity (often down to a scalar balance law), taking care to retain the important characteristic features of the original system. A multitude of analytical results are available for scalar balance laws, ranging from existence and uniqueness results to the convergence of numerical schemes in higher space dimensions. We employ this approach to justify the use of a finite–volume scheme to solve the coupled system of radiation magnetohydrodynamics. We carefully derive a scalar balance law that includes a non–local operator. This source term, which models the radiation transport, is the novel feature of our model problem. We first study the influence of this non–local term on the solution of the model problem; then we prove the convergence of a finite–volume scheme including an explicit approximation of the non–local operator.

Numerical tests

The mathematical model consists of two parts, one describing the evolution of the fluid and the other the radiation field. The construction of our numerical schemes is based on this splitting, which leads to a MHD and a RT "module". These modules are discussed and tested separately since the coupling of the radiation and the fluid flow occurs only on a source term level. We thereby assume that the performance of the full scheme can be measured by the performance of both contributing modules. This indirect test of our algorithm is necessary since we are not aware of any simple test cases for the full coupled system. A rigorous test of the full algorithm is very difficult and requires a detailed understanding of the underlying physical processes in the solar atmosphere; this is beyond the scope of this presentation.
The main focus of our study is a comparison of the efficiency of different numerical schemes. We compare often used approaches from the literature with newly developed schemes. We measure the efficiency of a scheme by studying the error to runtime ratio. Since the complexity of our problems (especially of our simulations in 3d) leads to a high demand on computational cost, the runtime efficiency of the numerical scheme has to be the essential aspect of our study.

Hardware and software used

The numerical scheme is implemented in C++. We used many different computer systems for our numerical tests, including single processor Linux PC, a shared memory SGI computer system (Origin with 46 processors), the IBM RS/6000 SP computer at the Rechenzentrum in Karlsruhe, and the IBM Regatta at the Rechenzentrum in Freiburg. Both GnuPlot and the graphics library GraPE were used for the visualization of the data. Detailed references to all software packages used are given later.

Outline of the thesis

In the first chapter we derive the relevant system of equations for our solar physical applications. The physical derivation of the full system is not discussed in detail, only the relevant notation is introduced. The main part of our study is divided into three parts, each of which is preceded by an overview chapter and concluded with a summary. In the first part we outline a very general numerical scheme for solving the system of radiation magnetohydrodynamics. We justify the numerical scheme through the analysis of a simplified setting. In the second and third parts, we extend this basic numerical scheme, treating the parts for the fluid and the radiation separately. We now describe the three parts in more detail.

In the first part (Chapters 2–5) we present a standard **finite–volume scheme** for solving the system of radiation magnetohydrodynamics. At this stage we describe only the standard building block as can be found in the literature. The scheme described in Chapter 3 does not yet include all aspects of our mathematical model and in its basic form it is not suitable for use in challenging applications. It serves rather as the skeleton for the extensions described in the second and third parts. Before we study the necessary modifications of the scheme, we justify the general approach with an analytical study of a simplified model problem. The fact that the central non–standard aspect of the system is the non–local effect of the radiation source term dictates the choice of material presented in Chapter 4. The model problem consists of a scalar balance law with a right hand side including a **non–local integral operator**. We first study the properties of special solutions to the model scalar balance law and then present a general **convergence proof** for finite–volume schemes in 2d.

In the second part (Chapters 6–10) we present modifications of that part of the numerical scheme in which the evolution of the fluid variables is computed. In the overview Chapter 6 we present a number of challenges that our numerical scheme must meet and also describe approaches found in the literature, approaches we then use as comparison schemes for our own solution technique. The comparison methods are chosen in accordance with the guidelines we set up for our own modification as discussed above (simple extension of existing scheme and no additional computational cost). Chapters 7–9 are devoted to the description of the methods and numerical tests for three central challenges: we first study a relaxation approach that allows us to extend a solver for a perfect gas to approximate the MHD equations with a **general equation of state**; then we present a general framework in which the **divergence constraint** on the magnetic field is coupled with the evolution equation for the magnetic field. Based on this approach we derive a number of different correction mechanisms for reducing errors in the divergence of the magnetic field. Finally we study a modification of the base scheme that facilitates the accurate approximation of solutions near an **equilibrium state**.

The third part of our investigation (Chapters 11–14) is devoted to the presentation and study of numerical schemes for solving the **radiation transport equation**. Again we start with an overview in which we discuss the central aspects of this part of the numerical scheme and present a standard solver found in the literature that we use as a reference method. In Chapter 12 we derive a numerical scheme for approximating the radiation intensity for a fixed propagation direction. This is the central building

block used for the approximation of the radiation source term that enters into the balance law for the total energy. We present a **convergence proof** for our method and, after presenting numerical tests for fixed propagation directions, we conclude our investigation of the radiation transport module in Chapter 13 by studying the approximation of the **radiation source term** itself.
In the last Chapter we then present some results using our **3d MHD code**, including a simulation for a problem from solar physics.
The enclosed CD ROM contains a pdf version of this thesis and the sources of our 2d and 3d MHD code. Furthermore, we have included the web pages of our project. The CD ROM also contains additional material including movies and posters that were produced during the project. The file (*MHD.html*) in the root directory of the CD ROM gives specific details of the layout.

Acknowledgments

I would like to express my thanks to all my colleagues here at the IAM in Freiburg — especially to Matthias Wesenberg and Christian Rohde for the discussions of the methods, of the results, and of life, the universe, and everything. There are in fact many reasons for thanking Matthias Wesenberg with whom I have been working together since the beginning of our studies. All my colleagues helped me by supplying such important raw material as coffee and tea; special thanks, however, should go to the different people who, over the last few years, invested a huge amount of their time to keeping our computer system in working order.
Let me continue by saying thank you to my supervisor Dietmar Kröner for suggesting and planning this project and for the many fruitful discussions and suggestions. I would also like to thank him for introducing me to several other scientists who also influenced the work presented here, but of whom I can mention only a few here. On the one hand, these include our project partners from the Max–Planck Institute of Aeronomie, Manfred Schüssler and Peter Vollmöller. Many of the problems studied here were brought to our attention only through their continuing desire to apply our methods to problems from solar physics; the development of the new radiation transport solver was one of the results of this cooperation. On the other hand, I want to mention Ivan Sofronov from the Keldysh Institute of Applied Mathematics RAS in Moscow, with whom we developed transparent boundary conditions for our atmospheric flow problems, and Claus–Dieter Munz from the Institut für Aerodynamik und Gasdynamik, Universität Stuttgart, who suggested extending his divergence cleaning technique derived for the Maxwell equations to the MHD system. The analytical results for the finite–volume scheme were proven in cooperation with Christian Rohde, large parts of the numerical scheme were implemented together with Matthias Wesenberg, and the grid concept goes back to Bernhard Schupp, whom I would also like to thank for saving me from the tedious task of having to design the hierarchical grid concept.
My thanks for proofreading this thesis go to Christian Rohde and to my mother. To my wife Sabine Voigt go my very special thanks for reasons of which she is well aware.
This study was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the priority research program *Analysis and Numerics for Conservation Laws* (ANumE).

Contents

Chapter 1

Mathematical Model

In this chapter the mathematical model is derived that is used to describe the physical processes in the solar convection zone and photosphere. Three sets of equations have to be combined: the first set describes the evolution of the mass, the momentum, and the energy density of the fluid; the second set the evolution of the magnetic field. These two systems are the Euler equations of gas dynamics and the Maxwell equations. By taking into account the interaction between the magnetic field and the fluid these two systems are combined to yield the system of magnetohydrodynamics (MHD). This system is described in Section 1.1.
In the solar photosphere the energy transport by radiation plays an important role for the energy balance of the fluid. This makes it necessary to add the equation governing the radiation intensity. The absorption, emission, and transport of the radiation intensity is strongly influenced by the temperature and the density of the underlying fluid. The radiation transport (RT) equation is discussed in Section 1.2. The absorption and emission of radiation energy lead to a source term in the balance law for the energy density. The coupled system (RMHD) of the MHD system with the RT equation is introduced in Section 1.3. A detailed derivation of the MHD system can be found in [Cab70] and the interaction of radiation and hydrodynamics is discussed in [MM84].

1.1 The Equations of Magnetohydrodynamics (MHD)

Combining the Maxwell equations and the Euler equations of gas dynamics in the case where heat conduction, relativistic, viscous, and resistive effects can be neglected leads to the system of ideal magnetohydrodynamics (MHD). This system describes the motion of an electrically conducting fluid in the presence of a magnetic field in a domain $\Omega \times [0,T] \subset \mathbb{R}^3 \times \mathbb{R}^+$. It consists of eight balance laws together with a set of algebraic relations and a constraint equation on the magnetic field. The gravitational force on the fluid is taken into account through source terms.

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0 \qquad \text{(conservation of mass)}, \tag{1.1a}$$

$$\partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u}\mathbf{u}^T + \mathcal{P}) = \mathbf{q}_{\rho \mathbf{u}} \qquad \text{(conservation of momentum)}, \tag{1.1b}$$

$$\partial_t \mathbf{B} + \nabla \cdot (\mathbf{u}\mathbf{B}^T - \mathbf{B}\mathbf{u}^T) = 0 \qquad \text{(induction equation)}, \tag{1.1c}$$

$$\partial_t (\rho e) + \nabla \cdot (\rho e \mathbf{u} + \mathcal{P}\mathbf{u}) = q_{\rho e} \qquad \text{(conservation of energy)}, \tag{1.1d}$$

$$\nabla \cdot \mathbf{B} = 0 \qquad \text{(divergence constraint)}. \tag{1.1e}$$

The conserved quantities are the density $\rho > 0$, the momentum $\rho\mathbf{u} \in \mathbb{R}^3$, the magnetic field $\mathbf{B} \in \mathbb{R}^3$, and the total energy density $\rho e > 0$. The MHD system is augmented by some algebraic relations. The internal energy $\varepsilon > 0$ is defined by the relation between the total energy, the kinetic, and the magnetic energies

$$e = \varepsilon + \frac{1}{2}|\mathbf{u}|^2 + \frac{1}{8\pi\rho}|\mathbf{B}|^2 \qquad \text{(equation for the total energy)}. \tag{1.1f}$$

The pressure tensor combines the influence of the hydrodynamic and the magnetic pressure

$$\mathcal{P} = \left(p + \frac{1}{8\pi}|\mathbf{B}|^2\right)\mathcal{I} - \frac{1}{4\pi}\mathbf{B}\mathbf{B}^T \qquad \text{(equation for the pressure tensor)} \tag{1.1g}$$

where $\mathcal{I}$ denotes the unit tensor.
An *equation of state* (EOS) is used to close the system. This defines the (hydrodynamic) pressure $p > 0$ as a function of the internal energy and the density. We also define the temperature $\theta > 0$ in the same way

$$p = p(\rho, \varepsilon) \qquad \text{(EOS for the pressure)}, \tag{1.2a}$$
$$\theta = \theta(\rho, \varepsilon) \qquad \text{(EOS for the temperature)}. \tag{1.2b}$$

The speed of sound $c > 0$ in the fluid is also defined through the EOS:

$$c^2(\rho, \varepsilon) = -\tau^2\left(\partial_\tau p - p\partial_\varepsilon p\right) \quad \text{(speed of sound)}. \tag{1.2c}$$

As in the definition for the speed of sound in (1.2c) the density ρ or the specific volume $\tau = 1/\rho$ is used in the EOS adding the variables only when necessary.
The force of gravity leads to the following expressions for the source terms in the momentum and energy equations ((1.1b) and (1.1d)):

$$\mathbf{q}_{\rho\mathbf{u}} = \rho\mathbf{g}\ , \tag{1.3a}$$
$$q_{\rho e} = \rho\mathbf{g} \cdot \mathbf{u} \tag{1.3b}$$

where the vector valued function $\mathbf{g} = \mathbf{g}(\mathbf{x})$ is defined on Ω.
By virtue of (1.2a) and (1.1f) the MHD system can be written as a system of balance laws in the conserved variables

$$\mathbf{U} = (\rho, \rho\mathbf{u}, \mathbf{B}, \rho e)^T \in \mathcal{U} \tag{1.4}$$

where the state space $\mathcal{U} \subset \mathbb{R}^8$ is given by

$$\mathcal{U} := \big\{\mathbf{U} \in \mathbb{R}^+ \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^+ \mid p(\mathbf{U}_1, \varepsilon) > 0, \theta(\mathbf{U}_1, \varepsilon) > 0, c^2(\mathbf{U}_1, \varepsilon) > 0 \\ \text{with } \varepsilon = \frac{\mathbf{U}_8}{\mathbf{U}_1} - \frac{\mathbf{U}_2^2 + \mathbf{U}_3^2 + \mathbf{U}_4^2}{2\mathbf{U}_1^2} - \frac{\mathbf{U}_5^2 + \mathbf{U}_6^2 + \mathbf{U}_7^2}{8\pi\mathbf{U}_1} > 0\big\}\ . \tag{1.5}$$

ρ	:	density,	τ	=	$1/\rho$: specific volume,
$\mathbf{u}$	=	(u_x, u_y, u_z): velocity,	$\mathbf{B}$	=	(B_x, B_y, B_z): magnetic field,
e	:	total energy,	ε	:	internal energy,
p	:	(hydrodynamic) pressure,	θ	:	temperature,
c	:	speed of sound,	$\mathbf{g}$	:	gravitational force,
$\mathcal{P}$	:	pressure tensor,			
$\mathbf{U}$	=	$(\rho, \rho\mathbf{u}, \mathbf{B}, \rho e)^T$: conserved variables,			
$\mathbf{V}$	=	$(\rho, \mathbf{u}, \mathbf{B}, p)^T$: primitive variables.			

Table 1.1: Physical quantities

In the following we either write $\mathbf{U}_1$ to denote the first component of a space vector or also simply ρ. In the same sense $\mathbf{U}_2$ and $\rho\mathbf{u}_x$ are taken to be equivalent. In many cases it is convenient to study the MHD equations in the set of *primitive variables* given by

$$\mathbf{V} = (\rho, \mathbf{u}, \mathbf{B}, p)^T \ . \tag{1.6}$$

A summary of the notation is given in Table 1.1.
At time $t = 0$ the conserved variables are prescribed:

$$\mathbf{U}(\mathbf{x}, 0) = \mathbf{U}_0(\mathbf{x}) = (\rho_0(\mathbf{x}), (\rho\mathbf{u})_0(\mathbf{x}), \mathbf{B}_0(\mathbf{x}), (\rho e)_0(\mathbf{x}))^T \in \mathcal{U} \quad (\mathbf{x} \in \Omega).$$

In accordance with (1.1e) the initial conditions have to satisfy $\nabla \cdot \mathbf{B}_0 = 0$. Note that the flux in the induction equation (1.1c) can be rewritten as a curl of a vector field $\nabla \cdot (\mathbf{u}\mathbf{B}^T - \mathbf{B}\mathbf{u}^T) = \nabla \times (\mathbf{u} \times \mathbf{B})$. Therefore taking the divergence of equation (1.1c) yields $\partial_t \nabla \cdot \mathbf{B} = 0$. This shows that (1.1e) is a condition on the initial data of the magnetic field and not an additional elliptic constraint.
Introducing

$$\mathbf{q} = \mathbf{q}(\mathbf{U}) = (0, \mathbf{q}_{\rho\mathbf{u}}, 0, 0, 0, q_{\rho e})^T \tag{1.7}$$

the MHD system (1.1) can be rewritten in the compact form

$$\begin{aligned} \partial_t \mathbf{U} + \nabla \cdot \mathbf{F}(\mathbf{U}) &= \mathbf{q}(\mathbf{U}) \ , \\ \nabla \cdot \mathbf{B} &= 0 \ , \\ \mathbf{U}(\cdot, 0) &= \mathbf{U}_0 \end{aligned} \tag{1.8}$$

using the three dimensional flux vector $\mathbf{F} = (F_1, F_2, F_3)$. To close the system suitable boundary conditions for $\mathbf{U}(\mathbf{x}, t)$ for $(\mathbf{x}, t) \in \partial\Omega \times [0, T]$ have to be added.

1.1 Definition
We call a system of the form (1.8) hyperbolic *if for any unit vector* $\mathbf{n} \in \mathbb{R}^3$ *the Jacobian of the flux* $\mathbf{F}(\mathbf{U}) \cdot \mathbf{n}$ *has only real eigenvalues and a full set of right eigenvectors. It is* strictly hyperbolic *if all eigenvalues are distinct.*

1.2 Theorem
System (1.8) *is* hyperbolic *in the unknowns*

$$\mathbf{U}(\mathbf{x},t) = (\rho(\mathbf{x},t), (\rho\mathbf{u})(\mathbf{x},t), \mathbf{B}(\mathbf{x},t), (\rho e)(\mathbf{x},t)) \quad (\mathbf{x} \in \Omega \subset \mathbb{R}^3, t > 0) \tag{1.9}$$

if $\mathbf{U}(\mathbf{x},t) \in \mathcal{U}$ *for all* $(\mathbf{x},t) \in \Omega \times \mathbb{R}^+$.

A rigorous proof can be found for example in [Wes02b]. Here we restrict ourselves to a brief discussion. Since the MHD system (1.1) is invariant under rotation, it suffices to study the MHD equations in one space dimension. Consequently, we study the flux Jacobian of F_1. Note that the evolution equation for B_x and the divergence constraint, (1.1c) and (1.1e), respectively, lead to $B_x \equiv$ const. Therefore we have one zero eigenvalue that we denote with λ_{div}. In the literature it is common to view the MHD equations in 1d as a seven by seven system, treating B_x as a constant parameter that enters into the flux function F_1 and thus also influences the eigensystem. The eigenvalues of the flux Jacobian are in this case u_x, $u_x \pm c_s$, $u_x \pm c_a$, $u_x \pm c_f$ with speeds $c_s \leq c_a \leq c_f$ defined by

$$c_s = \sqrt{\frac{1}{2}\left(c^2 + b^2 - \sqrt{(c^2+b^2)^2 - 4c^2b_1^2}\right)}\,, \tag{1.10a}$$

$$c_a = |b_1|\,, \tag{1.10b}$$

$$c_f = \sqrt{\frac{1}{2}\left(c^2 + b^2 + \sqrt{(c^2+b^2)^2 - 4c^2b_1^2}\right)} \tag{1.10c}$$

using the abbreviations

$$b_1^2 = \frac{B_x^2}{\rho}\,, \quad b^2 = \frac{|\mathbf{B}|^2}{\rho}\,. \tag{1.11}$$

Given that the conserved quantities lie in the state space $\mathcal{U}$ so that $c^2 > 0$ and $\rho > 0$, all wave speeds are real numbers. We denote the eigenvalues in increasing order with λ_i for $i = 1, \ldots, 7$. Furthermore the flux Jacobian has seven linearly independent right eigenvectors (see [ZC92]). This shows that the Jacobian is diagonalizable and therefore the MHD system is hyperbolic. It is not strictly hyperbolic since, depending on the magnetic field $\mathbf{B}$, up to five eigenvalues can be identical (e.g. $\mathbf{B} = 0$).
In many problems — for example, in the lower convection zone — the fluid can be assumed to be a perfect gas. In this case the EOS (1.2a) is given by

$$p = (\gamma - 1)\rho\varepsilon \tag{1.12}$$

with a constant $\gamma > 1$. The temperature is given by

$$\theta = \frac{1}{R}\tau p$$

with a constant $R > 0$. This simple form of the EOS cannot be applied in the solar photosphere, since here the plasma is partially ionized. In this case a far more complicated law has to be applied. In fact, in this situation the pressure can only be computed by

solving additional ordinary differential equations. A simplified model is given by the *Saha Equations* [VK65, Chapter V,Section 4] and [Sch99b].
In the following we always make the assumption that our state vector $\mathbf{U}$ belongs to the state space $\mathcal{U}$ so that the MHD system is hyperbolic. This is summarized in the following assumption.

1.3 Assumption
We assume that the quantities p, θ, and c^2 given by (1.2) *are positive for all pairs of values ρ, ε appearing during a simulation. Furthermore for fixed ρ we assume that the pressure law satisfies $\partial_\varepsilon p > 0$. Consequently we can define the function*

$$\varepsilon = \varepsilon(\rho, p) \tag{1.13}$$

as the inverse of $p(\rho, \cdot)$ for fixed ρ, i.e.

$$p(\rho, \varepsilon(\rho, \bar{p})) = \bar{p} \ . \tag{1.14}$$

1.4 Remark: *Our definition of the temperature θ using* (1.2b) *is used for simplicity. Physically the pressure p is given as a function of the density ρ and the specific entropy s, i.e. $p = p(\rho, s)$. The temperature is then defined as the derivative of the pressure function with respect to the entropy: $\theta = \partial_s p(\rho, s)$. This derivative is always greater than zero. The internal energy ε is also defined as a function of ρ and s which is strictly increasing with respect to s, i.e. $\partial_s \varepsilon(\rho, s) > 0$. Therefore one can define the entropy s as a function of ρ and the internal energy ε. This then leads to $p = p(\rho, \varepsilon)$ and $\theta = \theta(\rho, \varepsilon)$ as used in* (1.2a) *and* (1.2b)*, respectively. Since it is not our intention to give a precise physical derivation of the MHD system, we assume for simplicity that the temperature and the pressure are defined by the EOS as functions of ρ and ε as summarized in* (1.2).
The assumption on the monotonicity of the pressure law with respect to ε is satisfied for most fluids. Due to this assumption it is always possible to map primitive variables onto their conserved counterparts since for given ρ and p the internal energy ε can be defined by the inverse of $p(\rho, \cdot)$ (cf. (1.13)*). By means of equation* (1.1f) *it is then possible to compute the total energy e. Note that it is always possible to define a mapping from conserved variables to primitive variables. For more details on the derivation and the hydrodynamic properties of the EOS see for example [VK65, MP89].*

1.2 The Radiation Transport Equation (RT)

The transport of electromagnetic radiation and its interaction with a fluid is generally described by the time dependent radiation transport (RT) equation

$$\frac{1}{c_{\text{light}}} \partial_t I_\nu(\mathbf{x}, t, \boldsymbol{\mu}) + \boldsymbol{\mu} \cdot \nabla I_\nu(\mathbf{x}, t, \boldsymbol{\mu}) + \chi_\nu(\mathbf{x}, t) I_\nu(\mathbf{x}, t, \boldsymbol{\mu}) = \chi_\nu(\mathbf{x}, t) B_\nu(\mathbf{x}, t) \ , \tag{1.15}$$

which is a linear Boltzmann type equation. The intensity I_ν represents the amount of energy transported (with the speed of light c_{light}) by radiation of frequency ν across an area $d^2\boldsymbol{\mu}$ in the direction $\boldsymbol{\mu}$. I_ν is a function of time $t > 0$, location $\mathbf{x} \in \Omega$, propagation direction $\boldsymbol{\mu} \in S^2$, and frequency $\nu \in \mathbb{R}^+$. $B_\nu = \epsilon_\nu / \kappa_\nu$ and $\chi_\nu = \rho \kappa_\nu$ are functions

depending on the temperature θ and the mass density ρ of the underlying fluid. χ_ν is the inverse mean free photon path, $\kappa_\nu(\rho,\theta)$ is the absorption coefficient, and $\epsilon_\nu(\rho,\theta)$ is the emission coefficient.
The source function B_ν depends on the temperature θ and not on the density ρ of the underlying fluid. It is described by Planck's law for black body radiation

$$B_\nu(\theta) = 2\pi h c_{\text{light}}^2 \frac{\nu^5}{\exp(\frac{h c_{\text{light}} \nu}{k\theta}) - 1}$$

where k and h are the Boltzmann and Planck constants, respectively. The absorption of radiation is described by the function χ_ν, which depends on the fluid density ρ and temperature θ. In applications this can be a very complicated function, which is sometimes only defined in tables.
Under the assumption that the radiation relaxation time is small compared with all other time scales of the physical system considered, the RT equation can be reduced to its stationary form. For simplicity we also neglect the dependency on the frequency, thereby using frequency averaged data χ, B:

$$\boldsymbol{\mu} \cdot \nabla I(\mathbf{x}, t, \boldsymbol{\mu}) + \chi(\mathbf{x}, t) I(\mathbf{x}, t, \boldsymbol{\mu}) = \chi(\mathbf{x}, t) B(\mathbf{x}, t). \tag{1.16a}$$

This equation is well posed if the intensity is given on the *inflow boundary:*

$$I(\mathbf{x}, t, \boldsymbol{\mu}) = g(\mathbf{x}, t, \boldsymbol{\mu}) \quad \mathbf{x} \in \partial\Omega_-. \tag{1.16b}$$

The inflow boundary belonging to a direction $\boldsymbol{\mu}$ is given by

$$\partial\Omega_- = \partial\Omega_-(\boldsymbol{\mu}) := \{\mathbf{x} \in \partial\Omega : \mathbf{n}(\mathbf{x}) \cdot \boldsymbol{\mu} < 0\} \tag{1.17}$$

where $\mathbf{n}(\mathbf{x})$ denotes the unit outer normal to $\partial\Omega$ at $\mathbf{x} \in \partial\Omega$. Since $\theta > 0$ we compute

$$B(\theta) := \int_0^\infty B_\nu(\theta) d\nu = \sigma\theta^4 \tag{1.18}$$

where $\sigma = \frac{2\pi^5 k^4}{15 h^3 c_{\text{light}}^2}$ is the Stefan–Boltzmann constant.

1.3 The Coupled System (RMHD)

For the dynamic interaction of radiation and matter in the solar photosphere, the system of radiation magnetohydrodynamics (RMHD) has to be studied. The coupling of the RT equation (1.16a) and the MHD equations (1.1) leads to a source term Q_{rad} in the balance law for the total energy (1.1d). Therefore the set of equations (1.1) has to be augmented as follows:

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0 \qquad \text{(conservation of mass)}, \tag{1.19a}$$

$$\partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u}\mathbf{u}^T + \mathcal{P}) = \mathbf{q}_{\rho u} \qquad \text{(conservation of momentum)}, \tag{1.19b}$$

$$\partial_t \mathbf{B} + \nabla \cdot (\mathbf{u}\mathbf{B}^T - \mathbf{B}\mathbf{u}^T) = 0 \qquad \text{(induction equation)}, \tag{1.19c}$$

$$\partial_t(\rho e) + \nabla \cdot (\rho e \mathbf{u} + \mathcal{P}\mathbf{u}) = q_{\rho e} + Q_{\text{rad}} \qquad \text{(conservation of energy)}, \tag{1.19d}$$

$$\nabla \cdot \mathbf{B} = 0 \qquad \text{(divergence constraint)}, \tag{1.19e}$$

$$\boldsymbol{\mu} \cdot \nabla I + \chi I = \chi B \qquad \text{(radiation transport)}. \tag{1.19f}$$

The radiation source term takes on the following form:

$$Q_{\text{rad}}(\mathbf{x}, t) = \int_{S^2} \chi(\mathbf{x}, t)\,(I(\mathbf{x}, t, \boldsymbol{\mu}) - B(\mathbf{x}, t))\,d\boldsymbol{\mu} \qquad \text{(radiation source term)}. \tag{1.19g}$$

Both χ and B are given functions of the density ρ and the temperature θ:

$$B(\mathbf{x}, t) = B(\theta(\mathbf{x}, t)) = \sigma\theta(\mathbf{x}, t)^4, \tag{1.19h}$$

$$\chi(\mathbf{x}, t) = \rho(\mathbf{x}, t)\kappa(\rho(\mathbf{x}, t), \theta(\mathbf{x}, t)). \tag{1.19i}$$

Depending on the context we use both $B = B(\mathbf{x}, t)$, $\chi = \chi(\mathbf{x}, t)$ and $B = B(\theta)$, $\chi = \chi(\rho, \theta)$. The system is closed using the algebraic relations (1.1f), (1.1g) and the relations (1.2) defined through the EOS. Again initial data $\mathbf{U}_0$ and suitable boundary conditions have to be added. A derivation of the full system of radiation magnetohydrodynamics can be found in [MM84].
Defining

$$\mathbf{Q}_{\text{rad}} := (0, 0, 0, 0, 0, 0, 0, Q_{\text{rad}})^T\ , \tag{1.20}$$

the coupled system (1.19) can be rewritten in the compact form:

$$\begin{aligned} \partial_t \mathbf{U} + \nabla \cdot \mathbf{F}(\mathbf{U}) &= \mathbf{q}(\mathbf{U}) + \mathbf{Q}_{\text{rad}}\ , \\ \nabla \cdot \mathbf{B} &= 0\ , \\ \boldsymbol{\mu} \cdot \nabla I &= \chi(B - I)\ , \\ \mathbf{U}(\cdot, 0) &= \mathbf{U}_0\ . \end{aligned} \tag{1.21}$$

1.5 Remark: *The coupling between the radiation and the fluid occurs only on a source term level so that the hyperbolicity of the fluid part of the system is still guaranteed as long as the sound speed c defined by* (1.2c) *is positive. Taking the RT equation in its time–independent form, however, leads to a non–local effect and the finite speed of propagation is lost. Thus the finite domain of dependence so characteristic of hyperbolic systems is not maintained. The influence of the non–local nature of the radiation source term Q_{rad} is discussed in Chapter 4.*
Note that although the radiation transport equation is taken in its time–independent form, the radiation intensity and therefore Q_{rad} depend on time because the time dependent temperature θ and density ρ of the fluid influences the radiation field. On the other hand no initial data for the intensity has to be prescribed since it is uniquely defined by $\mathbf{U}_0$ and the boundary conditions.

2. Overview

Base Scheme

In recent years upwind finite–volume schemes have become very popular in numerical gas dynamics. This is due to the fact that by using this approach one can obtain both sharp shock profiles without generating spurious oscillations and second or higher order accuracy in smooth parts of the flow. Among the first to successfully apply Godunov–type schemes to the MHD equations were Brio and Wu [BW88]; further examples include [DW94, ZMC94, PRM$^+$95, BKP96, CG97, BD99, DPRV99, DRW99, PRL$^+$99, Wes02b]. Many other approaches developed for systems of balance laws have also been applied to the system of magnetohydrodynamics. These include finite–difference schemes; finite–element schemes such as the Discontinuous Galerkin method of Cockburn [CKS00, DKR$^+$03]; central schemes following the ideas of Nessyahu and Tatmoor [NT90] can be easily applied to the real gas MHD system. Schemes specially derived for multidimensional systems include, e.g., the evolution Galerkin method [LMMW00] and the method of transport, which has been applied to the perfect gas MHD system [FNT01]. Furthermore, many methods have been presented that are specially designed for the approximation of the MHD equations in higher space dimensions on structured grids, an overview can be found in [Tót00].

In addition to their good numerical properties, which have been shown through many examples, the analytical properties of finite–volume schemes also justify their use for approximating systems of balance laws such as the system of radiation magnetohydrodynamics (1.19). For a scalar balance law the convergence of finite–volume schemes even in higher space dimensions has been shown, requiring only very few restrictions on the non–linearities and under realistic assumptions on the regularity of the solution (e.g. [EGH00]). This is discussed in Chapter 4, where we extend the convergence result for finite–volume schemes to include a class of scalar model problems derived from our coupled system (1.19) of radiation magnetohydrodynamics. Also for special systems convergence of general finite–volume schemes in higher space dimension has been shown, e.g. for weakly coupled systems in [Roh98] and for linear systems of Friedrich's type in [JR02]. For special schemes, convergence proofs are available even for more general non–linear systems in one space dimension (e.g. [Gli65, BJ00]).

2.1 General Concept

We describe the concept of the finite–volume approach for the full coupled system (1.21) — the purely magnetohydrodynamic setting is included by setting the radiation source term $\mathbf{Q}_{\text{rad}}$ to zero. In the finite–volume approach the conserved variables $\mathbf{U}$ are approximated by average values on elements of a grid $\mathcal{T} = \{T_i : i \in \mathcal{I}\}$. For simplicity we assume that $\Omega = \bigcup_{i\in\mathcal{I}} T_i$. The derivation starts from the integral form of (1.21)

$$\int_{T_i} \partial_t \mathbf{U} + \int_{\partial T_i} \mathbf{F}(\mathbf{U}) \cdot \mathbf{n} = \int_{T_i} \left(\mathbf{q}(\mathbf{U}) + \mathbf{Q}_{\text{rad}}\right) . \tag{2.1}$$

For $i \in \mathcal{I}$ we define average values

$$\mathbf{U}_i(t) = \frac{1}{|T_i|} \int_{T_i} \mathbf{U}(\mathbf{x}, t), \quad \mathbf{Q}_{\text{rad}i}(t) = \frac{1}{|T_i|} \int_{T_i} \mathbf{Q}_{\text{rad}}(\mathbf{x}, t)$$

and a grid function

$$\mathbf{U}_h(\mathbf{x}, t) = \mathbf{U}_i(t) \quad \text{for} \quad \mathbf{x} \in T_i \tag{2.2}$$

which we use to approximate equation (2.1):

$$\frac{d}{dt}\mathbf{U}_i(t) = -\frac{1}{|T_i|} \int_{\partial T_i} \mathbf{F}(\mathbf{U}_h(\cdot, t)) \cdot \mathbf{n} + \frac{1}{|T_i|} \int_{T_i} \mathbf{q}(\mathbf{U}_h(\cdot, t)) + \mathbf{Q}_{\text{rad}i}(t) . \tag{2.3}$$

To arrive at a fully explicit scheme suitable for simulations on a computer we introduce a few more approximation steps in Chapter 3:

- The computation of the radiation source term Q_{rad} requires the approximation of the integral in (1.19g); this step is detailed in Section 3.1.
- We use standard Runge–Kutta methods for approximating the time derivative and quadrature rules to approximate the integrals in (2.3); this is described in Section 3.2 and Section 3.3.
- Since $\mathbf{U}_h(\cdot, t)$ is discontinuous over the cell boundaries, the flux $\mathbf{F}(\mathbf{U}_h(\cdot, t)) \cdot \mathbf{n}$ is not well defined and has to be approximated by a *numerical flux function*. One possible choice is described in Section 3.4.

As a consequence of these steps the values $\mathbf{U}_i(t)$ and $\mathbf{Q}_{\text{rad}i}(t)$ will only be approximations of the exact average values.
In Chapter 3 we focus only on a simple base scheme as found in the literature. It does not include, for example, the treatment of arbitrary equations of state or the treatment of the divergence constraint. These and further extensions of the base scheme, which are indispensable for real life applications, are studied in the Chapters 7, 8, and 9. Furthermore we only briefly describe the approximation of the radiation source term, but not of the radiation intensity itself. This is dealt with in Chapters 12 and 13.

2.2 Grid Structure

In this section we fix some notation used throughout this study (for triangular grids these are summarized in Figure 2.1). With T_i we denote the elements of a grid where i is in some index set $\mathcal{I}$. The set of all elements is given by $\mathcal{T} = \{T_i : i \in \mathcal{I}\}$. We assume that each element is uniquely defined by a finite set of vertices $\mathbf{p}_k \in \mathcal{P}(i)$. The set of vertices of the grid is denoted by $\mathcal{P} = \{\mathbf{p}_k : k \in \mathcal{I}_P\}$ for an index set $\mathcal{I}_P$. Each element T_i has a finite set of neighbors T_j with $j \in \mathcal{N}_1(i)$ where $T_j \cap T_i$ has codimension one. The intersection of two neighboring grid elements defines a face $S_{ij} = T_i \cap T_j$ $(j \in \mathcal{N}_1(i))$. We assume that the computational domain Ω is the disjoint union of the elements T_i. The boundary segments of the grid we denote with S_j for $j \in \mathcal{I}_B$; for each $j \in \mathcal{I}_B$ there exists an unique index $b_j \in \mathcal{I}$ with $S_j \cap T_{b_j} = S_j$. All index sets are held to be disjoint. We define $\mathcal{N}_2(i) = \{j \in \mathcal{I}_B : b_j = i\}$ and corresponding faces $S_{ij} = S_j \cap T_i$ for $j \in \mathcal{N}_2(i)$. The set of all neighbors of an element T_i is given by $\mathcal{N}(i) = \mathcal{N}_1(i) \cup \mathcal{N}_2(i)$. The index set of all faces is given by $\mathcal{I}_S = \{(i,j) : j \in \mathcal{N}_1(i), i < j\} \cup \{(i,j) : j \in \mathcal{N}_2(i)\}$ and the set of all faces by $\mathcal{S} = \{S_{ij} : (i,j) \in \mathcal{I}_S\}$. For a face S_{ij} we denote with $\mathbf{n}_{ij}(\mathbf{x})$ the unit normal pointing outwards from the element T_i at $\mathbf{x} \in S_{ij}$. Furthermore we define the set of vertices on the face S_{ij} with $\mathcal{P}(i,j) = \{k \in \mathcal{P} : p_k \in S_{ij}\}$. We denote the minimum height over the face S_{ij} in T_i by h_{ij}, i.e. $h_{ij} = \max\{h > 0 : \mathbf{x} - h\mathbf{n}_{ij}(\mathbf{x}) \in T_i, \mathbf{x} \in S_{ij}\}$. We distinguish the elements of a family of grids $\{\mathcal{T}_h\}_{h>0}$ by the parameter $h = \min_{i \in \mathcal{I}} h_i$ with $h_i = \max_{j \in \mathcal{N}(i)} h_{ij}$. Functions defined piecewise on each grid $\mathcal{T}_h$ of a family $\{\mathcal{T}_h\}_{h>0}$ are denoted with a subscript h. All index sets of $\mathcal{T}_h$ also depend on h; in most cases this is not included in the notation.

For simplicity we assume in the following that the grid consists of triangles in two space dimensions and tetrahedrons in three space dimensions – although we have also implemented the methods on Cartesian and hexahedral grids [DKSW01b, DRSW02]. For triangular grids the faces S_{ij} are straight line segments and hypersurfaces for tetrahedral grids; the normal $\mathbf{n}_{ij}$ is constant. The height can be easily computed as $h_{ij} = 2\frac{|T_i|}{|S_{ij}|}$ for a triangular grid and $h_{ij} = 3\frac{|T_i|}{|S_{ij}|}$ for a tetrahedral grid.

To implement our finite–volume scheme we have to evaluate integrals on the elements T_i of the grid and on the faces S_{ij}. For these integrals we use quadrature rules that approximate the integrals by weighted sums. We also present these quadrature rules only in the case of triangular and tetrahedral grids. We have to define the center of gravity of the element T_i and of the face S_{ij}: $\boldsymbol{\omega}_i = \frac{1}{|\mathcal{P}(i)|}\sum_{k \in \mathcal{P}(i)} \mathbf{p}_k$ and $\mathbf{z}_{ij} = \frac{1}{|\mathcal{P}(i,j)|}\sum_{k \in \mathcal{P}(i,j)} \mathbf{p}_k$, respectively.

In our implementation the grid is organized in a hierarchy, starting with a coarse macro grid at the top level (Figure 2.2(a)). In each refinement step the new elements are organized in a tree structure below the node which corresponds to the refined element (Figure 2.2); the grid $\mathcal{T}$ then consists only of the leafs of this hierarchy. This storage technique has the advantage that local refinement and especially local coarsening of elements can be performed very fast without influencing global structures. This has many advantages for parallel computations. For more details see also [DRW99, Sch99a, DRSW02]).

To minimize the computational cost we adapt the grid to the structure of the solution. In regions where the solution is smooth, large grid elements can be used; whereas in

global index sets

$\mathcal{I} = \{0, 1, 2, \dots\}/\mathcal{I}_B = \{3, \dots\}$: index set of all cells/boundaries

$\mathcal{I}_S = \{(0,1), (0,2), (0,3), \dots\}$: index set of surfaces

$\mathcal{I}_P = \{0, 1, 2, 3, 4, \dots\}$: index set of vertices

quantities relative to T_0

$\boldsymbol{\omega}_0$: center of gravity

$\mathcal{N}_1(0) = \{1, 2\}/\mathcal{N}_2(0) = \{3\}$: index set of cell/boundary neighbors

$\mathcal{N}(0) = \{1, 2, 3\}$: index set of all neighbors

$\mathcal{P}(0) = \{0, 1, 3\}$: index set of vertices

quantities relative to S_{02}

$\mathbf{z}_{02}, \mathbf{n}_{02}, h_{02}$: midpoint,unit normal, and height

$\mathcal{P}(0, 2) = \{1, 3\}$: index set of vertices

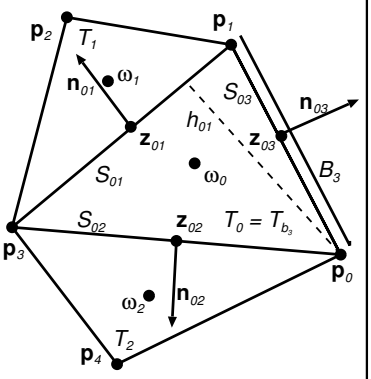
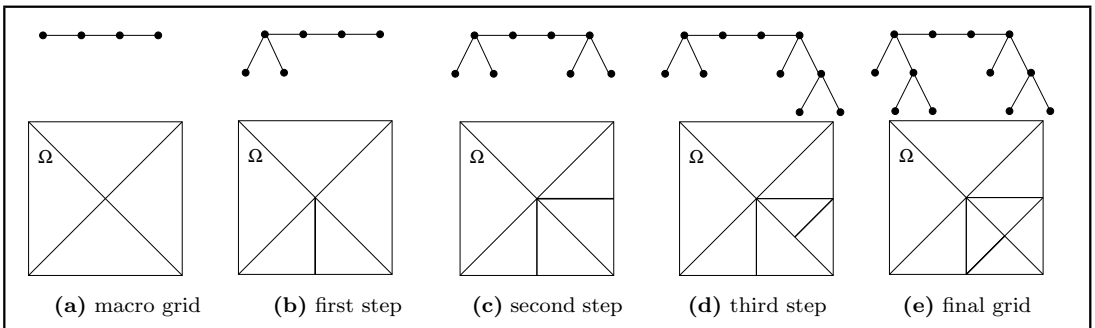


Figure 2.1: Unstructured triangular grid around a cell T_0 and surface S_{02}



(a) macro grid **(b)** first step **(c)** second step **(d)** third step **(e)** final grid

Figure 2.2: Hierarchical storage of a triangular grid during refinement.

regions with shocks, small elements are necessary to obtain accurate results. Since the position of these smooth and discontinuous structures in the computational domain move in time, we cannot use a grid that is refined a priori. We have to keep track of the different regions and adapt the grid dynamically during the simulations. This leads to a series of grids $(\mathcal{T}^n)_{n\in\mathbb{N}}$ where $\mathcal{T}^n$ is the grid on which the approximation is defined for $t \in [t^n, t^{n+1})$. The index sets corresponding to $\mathcal{T}^n$ naturally also depend on n and are also denoted with a superscript n when necessary. In 2d we perform an iteration procedure in each adaption step, which leads to a conform triangulation, i.e. a grid with no hanging nodes; for example, the grid in Figure 2.2(d) would be refined one step further to produce the grid in Figure 2.2(e). Since this iteration can be computationally expensive, our 3d code allows for a difference of one level between two elements T_i, T_j adjacent to a face $S_{ij} \in \mathcal{S}^n$. Since in most cases the macro grid is too coarse for a good approximation of the initial data, we first refine the grid until all h_i are below some given constant h_{start}. Then we use our local refinement strategy a fixed number of times, always projecting the initial data $\mathbf{U}_0$ onto the new grid. This strategy leads to a initial grid $\mathcal{T}^0$ on which we start our simulation. Since we use a time–stepping scheme that operates on a fixed grid to advance the solution from one time–level to the next, we neglect the index n for the grid in the following.

Chapter 3

Finite–Volume Schemes

In the following sections we detail the necessary approximation steps which enable us to use the finite–volume approach (2.3) to compute an approximation of the solution to the coupled system of balance laws (1.21).

3.1 Approximation of the Radiation Source

To approximate the average radiation source term

$$\mathbf{Q}_{\mathrm{rad}_i}(t) \approx \frac{1}{|T_i|} \int_{T_i} \mathbf{Q}_{\mathrm{rad}}(\cdot, t)$$

for $t \geq 0$ on an element $T_i \in \mathcal{T}$ the integral over the propagation directions $\boldsymbol{\mu}$ in (1.19g) is approximated by a quadrature rule. For a fixed set $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_M$ in S^2 with $M \geq 1$ we denote with $I_m = I_m(\mathbf{x}, t)$ the intensity in direction $\boldsymbol{\mu}_m$ for $m = 1, \ldots, M$ defined by the temperature θ and the density ρ at time t. With the *discrete ordinate method* (DOM) we obtain a semi–discrete approximation of the radiation source term Q_{rad}

$$\begin{aligned}\frac{1}{|T_i|} \int_{T_i} Q_{\mathrm{rad}}(\mathbf{x}, t) d\mathbf{x} &= \frac{1}{|T_i|} \int_{T_i} \int_{S^2} \chi(\mathbf{x}, t) \left(I(\mathbf{x}, t, \boldsymbol{\mu}), -B(\mathbf{x}, t)\right) d\boldsymbol{\mu} d\mathbf{x} \\ &\approx \frac{1}{|T_i|} \int_{T_i} \sum_{m=1}^{M} \omega_m \chi(\mathbf{x}, t) \left(I_m(\mathbf{x}, t) - B(\mathbf{x}, t)\right) \ .\end{aligned}$$

The constants $\omega_1, \ldots, \omega_M$ are the weights of the quadrature rule. Using a further quadrature rule for the integral over T_i we obtain the approximation $Q_{\mathrm{rad}_i}(t)$ of the average radiation intensity.
The choice of the directions $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_M$ and weights $\omega_1, \ldots, \omega_M$ defining a quadrature rule on S^2 is a non–trivial problem and can strongly influence the quality of the approximation. We will not study this aspect of the scheme . We use the quadrature rule with $M = 24$ given in Table 3.1 in all our calculations involving radiation. (Note that in 2d due to symmetry we have to compute only one half of the intensities values since $I_m = I_n$ for all $m, n \in \{1, \ldots, M\}$ with $(\boldsymbol{\mu}_{m,1}, \boldsymbol{\mu}_{m,2}, \boldsymbol{\mu}_{m,3}) = (\boldsymbol{\mu}_{n,1}, \boldsymbol{\mu}_{n,2}, -\boldsymbol{\mu}_{n,3})$.)
In [Car63] some general guidelines for the construction of suitable quadrature rules

$$\begin{aligned}
\boldsymbol{\mu}_{1,4,7,10,13,16,19,22} &= (\pm 0.88191710, \pm 0.33333333, \pm 0.33333333),\\
\boldsymbol{\mu}_{2,5,8,11,14,17,20,23} &= (\pm 0.33333333, \pm 0.33333333, \pm 0.88191710),\\
\boldsymbol{\mu}_{3,6,9,12,15,18,21,24} &= (\pm 0.33333333, \pm 0.88191710, \pm 0.33333333).
\end{aligned}$$

Table 3.1: Quadrature rule for integral over S^2 with $M = 24$ points. The weights are equal to $\omega_m = \frac{4\pi}{24}$ for $m = 1, \ldots, 24$.

were derived focusing on some special physical aspects of the problem. In [BVS99] the different quadrature rules were compared for a model problem from solar physics. We use the quadrature favored by the authors.
Now the source term is given by $\mathbf{Q}_{\mathrm{rad}i}(t) = (0, 0, 0, 0, 0, 0, 0, Q_{\mathrm{rad}i}(t))^T$, and its computation is reduced to solving the radiation transport equation (1.19f) for a set of fixed directions $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_M$:

$$\begin{aligned}
\boldsymbol{\mu}_m \cdot \nabla I_m + \chi I_m &= \chi B \qquad \text{in } \Omega\ ,\\
I_m &= g \qquad \text{on } \partial\Omega_-(\boldsymbol{\mu}_m)
\end{aligned} \tag{3.1}$$

in Ω for $1 \le m \le M$. The data B and χ are thereby given piecewise smooth functions defined on Ω. The boundary data g is defined on the inflow part of the boundary (cf. (1.17)).
We use an explicit method to construct an approximation to (3.1) which computes I_m element-wise. The details are described in Chapter 12. In the following we just assume that we have a method for approximating I_m on a given element T_i for fixed data B, χ and with the intensity I_m given on the inflow boundary $\partial T_{i-}(\boldsymbol{\mu}_m)$ of T. Using this method as a building block, we can construct an approximation on all T_i for $i \in \mathcal{I}$ if we can find a permutation π_m on the index set $\mathcal{I}$ with the following property:

$$\text{For all } (i,j) \in \mathcal{I}_S : S_{ij} \subset \partial T_{i-}(\boldsymbol{\mu}_m) \Rightarrow \pi_m(j) \le \pi_m(i)\ . \tag{3.2}$$

Such a sequence allows us to use the solution scheme for one element to construct the discrete solution on the whole grid. We start the approximation on the element $T_{\pi_m(1)}$ since, due to property (3.2), the inflow boundary of $T_{\pi_m(1)}$ must lie on $\partial\Omega_-(\boldsymbol{\mu}_m)$, where the intensity is given by the boundary data g. Assuming that we have constructed a solution on $T_{\pi_m(1)}, \ldots, T_{\pi_m(j-1)}$, we can compute the intensity on $T_{\pi_m(j)}$ since the inflow boundary of T_j is either part of $\partial\Omega_-(\boldsymbol{\mu}_m)$ or it is part of the boundary of the triangles $T_{\pi_m(1)}, \ldots, T_{\pi_m(j-1)}$, where we can use the intensity computed so far to define the inflow intensity.
With this iterative approach the two major parts for solving (3.1) consist of the solution algorithm for a given element T_i of the grid and of the construction of the permutation π_m for each propagation direction $\boldsymbol{\mu}_m$ for $m = 1, \ldots, M$. It has been shown in [LR74] that at least for triangular grids in 2d it is always possible to construct a suitable permutation. In 3d this has not been shown, and the authors of [WMMD01] claim to have found tetrahedral grids for which they could not construct such a permutation; but they could not construct a simple counter example. In Chapter 12 we review the ideas sketched in this section in detail and focus on the two important building blocks.

3.2 Time Evolution

To approximate the time derivative in (2.3) we use either the forward Euler scheme for a first order approximation or Heun's method for a second order scheme [HW91, Krö97]. The forward Euler method is an explicit one step method: we denote by

$$\mathbf{U}_i^n = (\rho_i^n, (\rho\mathbf{u})_i^n, \mathbf{B}_i^n, (\rho e_i)^n)^T$$

the approximation to the conserved variables at a time level t^n for $n \in \mathbb{N}$. For derived values we use a similar notation, for example, p_i^n denotes the pressure defined by the value $\mathbf{U}_i^n$. We set $t^0 = 0$ and define $\mathbf{U}_i^0$ as the average value of the initial data on T_i:

$$\mathbf{U}_i^0 = \frac{1}{|T_i|} \int_{T_i} \mathbf{U}_0(\cdot) \ . \tag{3.3}$$

The approximation of $\mathbf{U}$ on $\Omega \times [0,T]$ is then given by $\mathbf{U}_h(\mathbf{x},0) = \mathbf{U}_i^0$ for $\mathbf{x} \in T_i$ and

$$\mathbf{U}_h(\mathbf{x},t) = \mathbf{U}_i^{n+1} \quad (\mathbf{x} \in T_i, t^n < t \le t^{n+1}) \ . \tag{3.4}$$

The values $\mathbf{U}_i^{n+1}$ at the time level $t^{n+1} = t^n + \Delta t^n$ are computed from the values $\mathbf{U}_i^n$ via

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \frac{\Delta t^n}{|T_i|} \int_{\partial T_i} \mathbf{F}(\mathbf{U}_h(\cdot,t^n)) \cdot \mathbf{n} + \frac{\Delta t^n}{|T_i|} \int_{T_i} \mathbf{q}(\mathbf{U}_h(\cdot,t^n)) + \Delta t^n \mathbf{Q}_{\mathrm{rad}\,i}(t^n) \tag{3.5}$$

(cf. 2.3). This method is formally first order in space and time.
For a second order approximation $\mathbf{U}_h$ we use a reconstruction that defines a linear function on the element T_i taking the values on all neighboring elements into account. Given a set of average values $\{\mathbf{V}_j\}_{j\in\mathcal{J}}$ we denote the linear reconstruction on an element T_i ($i \in \mathcal{J}$) by mean of an operator

$$\mathcal{L}_i[\mathbf{V}](\mathbf{x}) = \mathcal{L}_i\big[\mathbf{V}_i, (\mathbf{V}_j)_{j\in\mathcal{N}(i)}\big](\mathbf{x}) \ . \tag{3.6}$$

We define the approximation $\mathbf{U}_h$ in $\Omega \times [0,T]$ by

$$\begin{array}{rcccll} \mathbf{U}_h(\mathbf{x},0) & = & \mathcal{L}_i^0(\mathbf{x}) & = & \mathcal{L}_i[\mathbf{U}^0](\mathbf{x}) & (\mathbf{x} \in T_i) \ , \\ \mathbf{U}_h(\mathbf{x},t) & = & \mathcal{L}_i^{n+\frac12}(\mathbf{x}) & = & \mathcal{L}_i[\mathbf{U}^{n+\frac12}](\mathbf{x}) & (\mathbf{x} \in T_i, t^n < t \le t^{n+\frac12}) \ , \\ \mathbf{U}_h(\mathbf{x},t) & = & \mathcal{L}_i^{n+1}(\mathbf{x}) & = & \mathcal{L}_i[\mathbf{U}^{n+1}](\mathbf{x}) & (\mathbf{x} \in T_i, t^{n+\frac12} < t \le t^{n+1}) \ . \end{array} \tag{3.7}$$

The average values $\mathbf{U}^0$ are defined by (3.3) and $\mathbf{U}^{n+\frac12}, \mathbf{U}^{n+1}$ are constructed by means of a two step Runge–Kutta method:

$$\begin{aligned} \mathbf{U}_i^{n+\frac12} &= \mathbf{U}_i^n - \frac{\Delta t^n}{|T_i|} \int_{\partial T_i} \mathbf{F}(\mathbf{U}_h(\cdot,t^n)) \cdot \mathbf{n} + \\ &\quad \frac{\Delta t^n}{|T_i|} \int_{T_i} \mathbf{q}(\mathbf{U}_h(\cdot,t^n)) + \Delta t^n \mathbf{Q}_{\mathrm{rad}\,i}(t^n) \ , \\ \mathbf{U}_i^{n+1} &= \frac12 \Big\{ \mathbf{U}_i^n + \mathbf{U}_i^{n+\frac12} - \frac{\Delta t^n}{|T_i|} \int_{\partial T_i} \mathbf{F}(\mathbf{U}_h(\cdot,t^{n+\frac12})) \cdot \mathbf{n} + \\ &\quad \frac{\Delta t^n}{|T_i|} \int_{T_i} \mathbf{q}(\mathbf{U}_h(\cdot,t^{n+\frac12})) + \Delta t^n \mathbf{Q}_{\mathrm{rad}\,i}(t^{n+\frac12}) \Big\} \ . \end{aligned} \tag{3.8}$$

On Cartesian and special triangular grids this leads to a formally second order scheme in space and time. In our numerical experiments we choose the DEOMOD method for constructing the linear reconstruction $\mathcal{L}$. This method is described in [DRW02a].
In the following we need the restriction $\mathbf{U}_i(\cdot,t)$ of the approximate function $\mathbf{U}_h(\cdot,t)$ on a cell $T_i \in \mathcal{T}$: in the first order scheme we have $\mathbf{U}_i(\cdot,t) = \mathbf{U}_i^{n+1}$ for $t^n < t \leq t^{n+1}$, and in the second order scheme the restriction is the linear function $\mathbf{U}_i(\cdot,t) = \mathcal{L}_i^{n+\frac{1}{2}}(\mathbf{x})$ for $t^n < t \leq t^{n+\frac{1}{2}}$ and $\mathbf{U}_i(\cdot,t) = \mathcal{L}_i^{n+1}(\mathbf{x})$ for $t^{n+\frac{1}{2}} < t \leq t^{n+1}$. The density components of the vector $\mathbf{U}_i(\cdot,t)$ is denoted by $\rho_i(\cdot,t)$ and analogous expressions are used for the other components.
A final ingredient of the time discretization is the choice of the time step Δt^n, which has to satisfy the usual CFL stability condition (cf. [Krö97]). The time step restriction is computed from the fastest wave speeds of the approximate solution at time t^n in the midpoints of the faces S_{ij} and in the direction $\mathbf{n}_{ij}$. We define a local time step for $(i,j) \in \mathcal{I}_S$

$$\Delta t_{ij}^n = \frac{h_{ij}}{\max\{\lambda_{\max}(\mathbf{U}_i(\mathbf{z}_{ij},t^n),\mathbf{n}_{ij}), \lambda_{\max}(\mathbf{U}_j(\mathbf{z}_{ij},t^n),\mathbf{n}_{ij})\}} \ . \tag{3.9}$$

The values in the denominator are the maximum of the absolute values of the wave speeds in the direction of the normal $\mathbf{n}_{ij}$ associated with the values on the two neighboring cells T_i and T_j, respectively. Following the discussion in Section 1.1 this value is

$$\lambda_{\max}(\mathbf{U},\mathbf{n}) = |\mathbf{u}\cdot\mathbf{n}| + c_f(\mathbf{U},\mathbf{n}) \tag{3.10}$$

where c_f is given by (cf. (1.10c)):

$$c_f(\mathbf{U},\mathbf{n}) = \sqrt{\frac{1}{2}\left(c(\mathbf{U})^2 + \frac{|\mathbf{B}|^2}{\rho} + \sqrt{\left(c(\mathbf{U})^2 + \frac{|\mathbf{B}|^2}{\rho}\right)^2 - 4c(\mathbf{U})^2\frac{(\mathbf{B}\cdot\mathbf{n})^2}{\rho}}\right)} \ .$$

With $c(\mathbf{U})$ we denote the speed of sound associated with the conserved values $\mathbf{U}$ and the EOS as defined by (1.2c). The global time step is

$$\Delta t^n = c_{\text{cfl}} \min_{(i,j)\in\mathcal{I}_S} \Delta t_{ij}^n \ . \tag{3.11}$$

The constant c_{cfl} is hereby smaller than 1. In the second order scheme the time step has to be the same for both steps of the Runge–Kutta method, thus only the approximation at time t^n enters into the computation of Δt^n. Note also that we do not take into account the stability restriction introduced by the source terms in (1.19); in none of our computations did this lead to any problems.
Next we turn to the approximation of the spatial integrals in our scheme. We use quadrature rules which naturally depend on the space dimension and the type of element. As already noted we restrict our presentation to the case of triangular and tetrahedral elements. On the one hand, we have an element integral due to the source term in the system (1.21). For this integral we use a quadrature rule that is exact for quadratic functions with quadrature points in the midpoints $\mathbf{z}_{ij}$ of the cell faces:

$$\frac{1}{|T_i|}\int_{T_i} \mathbf{q}(\mathbf{U}_h(\cdot,t)) \approx \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \mathbf{q}(\mathbf{U}_i(\mathbf{z}_{ij},t)) \ . \tag{3.12a}$$

The initial data is also approximated using this quadrature rule

$$\mathbf{U}_i^0 \approx \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \mathbf{U}_0(\mathbf{z}_{ij}) \ . \tag{3.12b}$$

The boundary integral, on the other hand, is first decomposed into a sum over the faces S_{ij}, and then the integral over a given face is approximated by the value at the center $\mathbf{z}_{ij}$:

$$\begin{aligned} \int\limits_{\partial T_i} \mathbf{F}(\mathbf{U}_h(\cdot,t)) \cdot \mathbf{n} &= \sum_{j \in \mathcal{N}(i)} \int\limits_{S_{ij}} \mathbf{F}(\mathbf{U}_h(\cdot,t)) \cdot \mathbf{n}_{ij} \\ &\approx \sum_{j \in \mathcal{N}(i)} |S_{ij}| \mathbf{F}(\mathbf{U}_h(\mathbf{z}_{ij},t)) \cdot \mathbf{n}_{ij} \ . \end{aligned} \tag{3.12c}$$

In the case of the element integral we use the restriction of $\mathbf{U}_h$ on the cell T_i. For the boundary integral this is not sufficient since in this case the flux over the cell interface has to be approximated; the flux is influenced by the values on both sides of the face S_{ij}, i.e. $\mathbf{U}_i(\mathbf{z}_{ij},t)$ and $\mathbf{U}_j(\mathbf{z}_{ij},t)$, respectively. Therefore we introduce a family of numerical flux function $\mathbf{g}_{ij}$ for $(i,j) \in \mathcal{I}_S$. The function $\mathbf{g}_{ij}$ is an approximation of the flux over the face S_{ij} in the direction of the normal $\mathbf{n}_{ij}$ and depends upon the approximate solution on either side of the interface, i.e.

$$\mathbf{g}_{ij} = \mathbf{g}_{ij}(\mathbf{U}_i(\mathbf{z}_{ij},t), \mathbf{U}_j(\mathbf{z}_{ij},t)) \approx |S_{ij}| \mathbf{F}(\mathbf{U}_h(\mathbf{z}_{ij},t)) \cdot \mathbf{n}_{ij} \ . \tag{3.13}$$

As pointed out this approximation is necessary since $\mathbf{U}_h$ can be discontinuous over cell interfaces. In the case where $\mathbf{U}_h$ is continuous it would be desirable for the numerical flux to be identical to the analytical flux. This leads to the following requirement on the numerical flux function:

$$\mathbf{g}_{ij}(\mathbf{U},\mathbf{U}) = |S_{ij}| \mathbf{F}(\mathbf{U}) \cdot \mathbf{n}_{ij} \ . \tag{3.14}$$

We also require that the flux from the cell T_i into its neighbor T_j over the face S_{ij} should be the same as the inverse of the flux from T_j into T_i, i.e.

$$\mathbf{g}_{ij}(\mathbf{U},\mathbf{V}) = -\mathbf{g}_{ji}(\mathbf{V},\mathbf{U}) \tag{3.15}$$

for all $\mathbf{U},\mathbf{V} \in \mathcal{U}$. These two properties play an important role in the convergence analysis of the finite–volume scheme (cf. Definition 4.5 in the following chapter). Property (3.15) allows us to write the sum over the faces of a cell in (3.12c) as a global sum over all faces so that in numerical schemes the flux only has to be computed once per face:

$$\begin{aligned} &\sum_{j \in \mathcal{N}(i)} |S_{ij}| \mathbf{g}_{ij}(\mathbf{U}_i(\mathbf{z}_{ij},t), \mathbf{U}_j(\mathbf{z}_{ij},t)) \\ &= \sum_{\substack{(k,l) \in \mathcal{I}_S \\ k=i}} \mathbf{g}_{kl}(\mathbf{U}_k(\mathbf{z}_{kl},t), \mathbf{U}_l(\mathbf{z}_{kl},t)) + \sum_{\substack{(k,l) \in \mathcal{I}_S \\ l=i}} \mathbf{g}_{lk}(\mathbf{U}_l(\mathbf{z}_{lk},t), \mathbf{U}_k(\mathbf{z}_{lk},t)) \end{aligned}$$

$$= \sum_{\substack{(k,l)\in\mathcal{I}_S \\ k=i}} \mathbf{g}_{kl}(\mathbf{U}_k(\mathbf{z}_{kl},t),\mathbf{U}_l(\mathbf{z}_{kl},t)) - \sum_{\substack{(k,l)\in\mathcal{I}_S \\ l=i}} \mathbf{g}_{kl}(\mathbf{U}_k(\mathbf{z}_{kl},t),\mathbf{U}_l(\mathbf{z}_{kl},t)) \ .$$

For $(i,j)\in\mathcal{I}_S$ with $j\in\mathcal{I}_B$ the data $\mathbf{U}_j(\mathbf{z}_{ij},t)$ has to be computed using the boundary conditions given by the setting of the simulation. This is described in the next section. Now taking into account (3.5), (3.12), (3.13), and the abbreviation $\mathbf{g}_{kl}^n = \mathbf{g}_{kl}(\mathbf{U}_k^n,\mathbf{U}_l^n)$, for $(k,l)\in\mathcal{I}_S$ the first order finite–volume scheme is given by

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \frac{\Delta t^n}{|T_i|}\Big(\sum_{\substack{(k,l)\in\mathcal{I}_S \\ k=i}} \mathbf{g}_{kl}^n - \sum_{\substack{(k,l)\in\mathcal{I}_S \\ l=i}} \mathbf{g}_{kl}^n\Big) + \Delta t^n \mathbf{q}(\mathbf{U}_i^n) + \Delta t^n \mathbf{Q}_{\mathrm{rad}_i}(t^n) \ . \tag{3.16}$$

To write down the second order finite–volume scheme we define

$$\mathbf{g}_{kl}^n = \mathbf{g}_{kl}(\mathcal{L}_k^n(\mathbf{z}_{kl}),\mathcal{L}_l^n(\mathbf{z}_{kl})),\ \mathbf{g}_{kl}^{n+\frac12} = \mathbf{g}_{kl}(\mathcal{L}_k^{n+\frac12}(\mathbf{z}_{kl}),\mathcal{L}_l^{n+\frac12}(\mathbf{z}_{kl}))$$

for $(k,l)\in\mathcal{I}_S$. Then our two step method reads

$$\begin{aligned}
\mathbf{U}_i^{n+\frac12} &= \mathbf{U}_i^n - \frac{\Delta t^n}{|T_i|}\Big(\sum_{\substack{(k,l)\in\mathcal{I}_S \\ k=i}} \mathbf{g}_{kl}^n - \sum_{\substack{(k,l)\in\mathcal{I}_S \\ l=i}} \mathbf{g}_{kl}^n\Big) + \\
&\quad \Delta t^n \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \mathbf{q}(\mathcal{L}_i^n(\mathbf{z}_{ij})) + \Delta t^n \mathbf{Q}_{\mathrm{rad}_i}(t^n) \ , \\
\mathbf{U}_i^{n+1} &= \frac12\Big\{\mathbf{U}_i^n + \mathbf{U}_i^{n+\frac12} - \frac{\Delta t^n}{|T_i|}\Big(\sum_{\substack{(k,l)\in\mathcal{I}_S \\ k=i}} \mathbf{g}_{kl}^{n+\frac12} - \sum_{\substack{(k,l)\in\mathcal{I}_S \\ l=i}} \mathbf{g}_{kl}^{n+\frac12}\Big) + \\
&\quad \Delta t^n \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \mathbf{q}(\mathcal{L}_i^{n+\frac12}(\mathbf{z}_{ij})) + \Delta t^n \mathbf{Q}_{\mathrm{rad}_i}(t^{n+\frac12})\Big\} \ .
\end{aligned} \tag{3.17}$$

3.1 Remark: *The radiation source term $\mathbf{Q}_{\mathrm{rad}_i}(t^n)$ is defined by the values $\mathbf{U}_j^n$ for $j\in\mathcal{I}$ through the temperature and the density that enter into equation* (1.19f) *for the radiation intensity. Note that the flux only has to be computed once on each face S_{ij}. For the definition of the linear reconstruction we refere to [DRW02a]. The DEOMOD method described there is a modification of the method described in [DEO92]. The steps of the algorithm are given in the summary Chapter 5.*

3.3 Boundary Conditions

To compute the flux on the faces S_{ij} for $j\in\mathcal{I}_B$ we have to define values $\mathbf{U}(\mathbf{z}_{ij},t)$ approximating the boundary conditions on the boundary elements S_j. These values may depend on some given data and on the solution in the interior of the domain (for example on the data defined on the neighboring element T_{b_j}). In many cases the definition of suitable boundary conditions is a non–trivial task since for hyperbolic problems the number of quantities that can be prescribed depends on the structure of the solution in the vicinity of the boundary. In the following we define a set of very simple boundary conditions that we use for our simulations. For $(i,j)\in\mathcal{I}_S$ and $j\in\mathcal{I}_B$ we define the following boundary conditions:

inflow: The boundary data is given by some function $\mathrm{BND}_j(\mathbf{x}, t)$ and does not depend on the solution in the interior of the domain:

$$\mathbf{U}_j(\mathbf{z}_{ij}, t) = \mathrm{BND}_j(\mathbf{z}_{ij}, t).$$

These boundary conditions can also be seen as *Dirichlet* boundary conditions.

outflow: The flow in the inside is prolonged into the exterior of the domain:

$$\mathbf{U}_j(\mathbf{z}_{ij}, t) = \mathbf{U}_i(\mathbf{z}_{ij}, t).$$

These boundary conditions are an approximation of Neumann boundary conditions.

slip: This is a solid wall condition. The normal components of the vector valued quantities vanish; the tangential components and the scalar values are continuous at the boundary:

$$\mathbf{U}_j(\mathbf{z}_{ij}, t) = \begin{pmatrix} \rho_i(\mathbf{z}_{ij}, t) \\ \rho_i(\mathbf{z}_{ij}, t)\big(\mathbf{u}_i(\mathbf{z}_{ij}, t) - (\mathbf{u}_i(\mathbf{z}_{ij}, t) \cdot \mathbf{n}_{ij})\mathbf{n}_{ij}\big) \\ \mathbf{B}_i(\mathbf{z}_{ij}, t) - (\mathbf{B}_i(\mathbf{z}_{ij}, t) \cdot \mathbf{n}_{ij})\mathbf{n}_{ij} \\ \rho_i(\mathbf{z}_{ij}, t) e_i(\mathbf{z}_{ij}, t) \end{pmatrix}.$$

reflecting: This boundary condition represents a symmetry axis in the solution and allows us to compute the solution on only one half of the domain. The scalar quantities are identical on both sides of the symmetry axis as are the tangential components of the vector valued quantities. The normal components of the vector valued quantities have an opposite sign:

$$\mathbf{U}_j(\mathbf{z}_{ij}, t) = \begin{pmatrix} \rho_i(\mathbf{z}_{ij}, t) \\ \rho_i(\mathbf{z}_{ij}, t)\big(\mathbf{u}_i(\mathbf{z}_{ij}, t) - 2(\mathbf{u}_i(\mathbf{z}_{ij}, t) \cdot \mathbf{n}_{ij})\mathbf{n}_{ij}\big) \\ \mathbf{B}_i(\mathbf{z}_{ij}, t) - 2(\mathbf{B}_i(\mathbf{z}_{ij}, t) \cdot \mathbf{n}_{ij})\mathbf{n}_{ij} \\ \rho_i(\mathbf{z}_{ij}, t) e_i(\mathbf{z}_{ij}, t) \end{pmatrix}.$$

periodic: The solution is periodic in the direction normal to the boundary, i.e. $\mathbf{U}(\mathbf{x}, t) = \mathbf{U}(\mathbf{x} + \alpha\mathbf{n}_{ij}, t)$ with some fixed $\alpha \in \mathbb{R}$. Therefore

$$\mathbf{U}_j(\mathbf{z}_{ij}, t) = \mathbf{U}_h(\mathbf{z}_{ij} + \alpha\mathbf{n}_{ij}, t).$$

With the exception of the slip boundary conditions all these boundary conditions are non–physical in the sense that the boundary is used to reduce the size of the computational domain. Inflow and outflow boundary conditions are only exact if all the characteristic waves of the flow at the boundary are moving into the domain or out of the domain, respectively. Reflecting and periodic boundary conditions can be used if some special property of the solution is known a priori. In many cases these boundary conditions are used to approximate the unknown physical boundary conditions. The problem of constructing suitable boundary conditions for our solar physical simulations is briefly discussed in Section 6.1.4, where we summarize results from [DKSW01b].

3.4 Rotated Riemann Solvers

For the computation of the flux across $S_{ij} \in \mathcal{S}$ we use a numerical flux function $\mathbf{G} = \mathbf{G}(\mathbf{U}, \mathbf{V})$ for the one dimensional MHD system, i.e., $\mathbf{G}$ is an approximation of the flux vector F_1 in (1.21). For simulations in higher space dimensions the cells T_i and T_j are rotated so that $\mathbf{n}_{ij}$ is transformed into the unit vector in x–direction. Now the numerical flux function is computed and the resulting vector is rotated back into the original frame. These rotations act as orthogonal mappings $\mathcal{R}(\mathbf{n}_{ij})$, $\mathcal{R}^{-1}(\mathbf{n}_{ij})$ on the data. Thus the numerical flux function for $\mathbf{U}, \mathbf{V} \in \mathcal{U}$, and $(i,j) \in \mathcal{I}_S$ is given by

$$\mathbf{g}_{ij}(\mathbf{U}, \mathbf{V}) = |S_{ij}| \mathcal{R}^{-1}(\mathbf{n}_{ij}) \mathbf{G}(\mathcal{R}(\mathbf{n}_{ij})\mathbf{U}, \mathcal{R}(\mathbf{n}_{ij})\mathbf{V}) \ . \tag{3.18}$$

In three space dimensions the rotation matrix $\mathcal{R}$ has the form

$$\mathcal{R}(\mathbf{n}) = \begin{cases} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{n}_1 & \mathbf{n}_2 & \mathbf{n}_3 & 0 & 0 & 0 & 0 \\ 0 & -\frac{\mathbf{n}_2}{d(\mathbf{n})} & \frac{\mathbf{n}_1}{d(\mathbf{n})} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{\mathbf{n}_1\mathbf{n}_3}{d(\mathbf{n})} & -\frac{\mathbf{n}_2\mathbf{n}_3}{d(\mathbf{n})} & d(\mathbf{n}) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{n}_1 & \mathbf{n}_2 & \mathbf{n}_3 & 0 \\ 0 & 0 & 0 & 0 & -\frac{\mathbf{n}_2}{d(\mathbf{n})} & \frac{\mathbf{n}_1}{d(\mathbf{n})} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{\mathbf{n}_1\mathbf{n}_3}{d(\mathbf{n})} & -\frac{\mathbf{n}_2\mathbf{n}_3}{d(\mathbf{n})} & d(\mathbf{n}) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, & \text{if } |\mathbf{n}_3| < 1 \ , \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, & \text{if } |\mathbf{n}_3| = 1 \ , \end{cases}$$

where $d(\mathbf{n}) = \sqrt{1 - \mathbf{n}_3^2}$.

3.2 Remark: *It is common in the literature to remove B_x from the vector of conserved variables when studying the MHD equations in 1d. This is possible since, due to the evolution equation for B_x and the divergence constraint, B_x must be a constant. Since we use the one dimensional flux only to construct a flux function for simulations in higher space dimensions, we include a component for B_x in the flux function $\mathbf{G}$, which is set to zero. Thus we have $\mathbf{G} = (\mathbf{G}_i)_{i=1}^8$ with $\mathbf{G}_5 = 0$.*

We are quite free in our choice of the one dimensional flux function $\mathbf{G}$ as long as the corresponding 2d flux function $\mathbf{g}_{ij}$ satisfies (3.14) and (3.15). In our numerical tests we use the MHD–HLLEM scheme developed in [Wes02b], where it was found to be the best choice compared to a range of other flux functions found in the literature.

3.5 Grid Adaptation

Local refinement and coarsening require the definition of an indicator that tells us which elements should be refined and which can be coarsened. This indicator is defined on each element T_i at a given time step t^n and is based on the structure of the approximate solution in the neighborhood of T_i. Every few time steps two indicators ref_i^n and crs_i^n for $i \in \mathcal{I}^n$ are computed, using the available data $\mathbf{U}_i^n$ on the element T_i together with the data $\mathbf{U}_j^n$ for $j \in \mathcal{N}^n(i)$ from the neighboring cells. Then elements for which ref_i^n is larger than some threshold value $\mathrm{ref}_{\mathrm{limit}}$ are refined, and the remaining elements are coarsened if crs_i^n is smaller than some threshold value $\mathrm{crs}_{\mathrm{limit}}$. To reduce the computational cost in our 2d code, the grid is modified only every five steps. To ensure that moving fronts are still resolved accurately, we also refine two layers of neighboring elements together with the element that is marked for refinement. Since we do not enforce conformity of the grid in our 3d code, grid adaptation is less expensive and is therefore performed every time step without adding a layer of elements around a refinement zone.
The indicators we use in our simulations are based only on heuristic arguments. In regions where the solution varies very little, the size of the elements can be large; strong variation in the solution requires small elements. Therefore we use the size of the jumps in the discrete solution over cell boundaries as indicators. Since especially in atmospheric simulation the values of the components of the solution vary over several orders of magnitude, some local normalization of the jumps is required. We define for each surface S_{ij} for $(i,j) \in \mathcal{I}_S^n$ the value

$$\mathrm{jmp}_{ij}^n = \max \left\{ \frac{|\rho_i^n(\mathbf{z}_{ij}) - \rho_j^n(\mathbf{z}_{ij})|}{\frac{1}{2}(\rho_i^n(\mathbf{z}_{ij}) + \rho_j^n(\mathbf{z}_{ij}))}, \frac{|p_i^n(\mathbf{z}_{ij}) - p_j^n(\mathbf{z}_{ij})|}{\frac{1}{2}(p_i^n(\mathbf{z}_{ij}) + p_j^n(\mathbf{z}_{ij}))}, \right.$$
$$\left. \sum_{k=1}^{3} \frac{|(\mathbf{u}_i^n)_k(\mathbf{z}_{ij}) - (\mathbf{u}_j^n)_k(\mathbf{z}_{ij})|}{\bar{u}_0}, \sum_{k=1}^{3} \frac{|(\mathbf{B}_i^n)_k(\mathbf{z}_{ij}) - (\mathbf{B}_j^n)_k(\mathbf{z}_{ij})|}{\bar{B}_0} \right\} . \quad (3.19)$$

We use relative jumps between neighboring elements since the different physical quantities can be of very different magnitudes. Since $\mathbf{u}$ and $\mathbf{B}$ can be zero we use some fixed constants $\bar{u}_0 > 0$ and $\bar{B}_0 > 0$ that have to be chosen a priori depending on the simulation. Due to the atmosphere the density can vary strongly in magnitude over the whole domain; therefore we use the mean value of the two densities, which is always greater than zero, to compute the relative indicator. For the same reason the jump in the pressure is also taken relative to the mean of the pressure values on the elements T_i and T_j. The choice of primitive variables is arbitrary, and in some cases taking the jump in the conserved variables might be advantages. Note that the indicator is symmetric with respect to T_i and T_j so that $\mathrm{jmp}_{ij} = \mathrm{jmp}_{ji}$. As indicator on T_i we take the maximum indicator from the surfaces belonging to T_i. Like the flux the local indicators are values defined on the faces since they are symmetric with respect to the elements on both sides of the face. On each element T_i we define

$$\mathrm{ref}_i^n = \max_{\substack{(k,l) \in \mathcal{I}_S^n \\ k=i \text{ or } l=i}} \mathrm{jmp}_{kl}^n , \quad \mathrm{crs}_i^n = \mathrm{ref}_i^n . \quad (3.20)$$

A slightly different approach and a numerical study is published in [DRW02a].

To ensure that elements do not become too large a maximal diameter $h_{\max}$ is prescribed. Since our choice for $\mathrm{ref}_i^n, \mathrm{crs}_i^n$ is not based on a rigorous a posteriori analysis, we have to prescribe a constant $h_{\min}$ to restrict the element size from below. On the one hand, the number of elements has to be as small as possible to minimize computational cost; on the other hand, the element size has to be small enough in regions where the solution shows a strong variation. The minimal element size also has to be chosen in such a way that the time step Δt^n, which is strongly influenced by the minimum element size (cf. (3.9)), is not too small; at the same time we are interested in sharp shock profiles, which require a high grid resolution.

3.3 Remark: *Our choice for the indicator* ref_i^n *is based only on the magnetohydrodynamic quantities and neglects the radiation intensities* I*. Since* I *is directly coupled to the fluid structure, this works reasonably well. In Chapter 12 we discuss the question of adaptation in the context of the RT equation, but in our simulations we have so far not included the radiation in the adaptation process.*
We already remarked upon the fact that our indicator is based on a heuristic approach and that we are not aware of a rigorous a posteriori analysis for complex system like the MHD equations. This approach was successfully used for similar complex systems of balance laws, for example, for reactive flows in [Geß01]. A further justification for this approach is given by rigorous a posteriori results for scalar equations [KO00].

3.6 Parallelization

To improve the performance of numerical schemes parallelization of the code is an essential tool. The use of more than one processor can considerably reduce the runtime for a given problem, and it can even be the only possible way to perform the simulation if, for example, the memory requirements are too large for a single processor computer. For an efficient parallelization strategy the architecture of the parallel machine has to be taken into account. In recent years parallel computers have become widely available, especially with shared and distributed memory architectures. Therefore we give a short overview of the general parallelization concepts for these two architectures; more details can be found in Chapter 15, where we describe some central aspects of the implementation of our 3d code in detail.

3.6.1 Shared Memory Architecture

For our 2d simulations we use a multithread model making strong use of the assumption that all threads have (almost) equal access to a global memory space where we can store all the data for the simulation. Therefore this strategy is well suited for use with a shared memory architecture and cannot be used on a distributed memory machine. The task of computing the solution is evenly distributed among all the threads. For the computation of the radiation source term Q_{rad} the propagation directions $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_M$ used to approximate the integral over S^2 (cf. Section 3.1) are partitioned into equally sized subsets and each thread computes the radiation intensities from one subset. To compute the intensity I_m in a fixed direction $\boldsymbol{\mu}_m$ at a given time level t^n and on an element T, we require the approximated hydrodynamic quantities ρ, θ on T given by

the finite–volume scheme and the intensity I_m from all the neighboring triangles in downwind direction (cf. Section 3.1). Therefore, we need the data on the whole grid to compute I_m at time t^n. Due to the assumption that each thread has access to all the data, this requirement presents no problem, and since the approximations of the intensity for two different directions $\boldsymbol{\mu}_i, \boldsymbol{\mu}_j$ $(i \neq j)$ are independent of each other, the parallelization of the computation of Q_{rad} can be performed very efficiently on a shared memory machine.

The evolution of the conserved quantities $\mathbf{U}_i^n$ on an element T_i from time level t^n to the new time level t^{n+1} only requires the conserved quantities $\mathbf{U}_j^n$ on the neighboring elements T_j for $j \in \mathcal{N}(i)$ at the time level t^n (cf. Section 3.2). Consequently, this part of the algorithm can also be parallelized very efficiently: the index set $\mathcal{I}$ is partitioned into subsets of equal size and each processor computes the update of the conservative quantities $\mathbf{U}$ for all the elements of one subset.

Only the reorganization of the grid during the refinement and coarsening process is not straightforward, and we have refrained from a parallelization of this part of the scheme. The overall efficiency of this approach was demonstrated in [Ded98].

3.6.2 Distributed Memory Architecture

The advantage of a shared memory architecture lies in the fact that all data are easily available. At the same time this leads to some restrictions on the size of the problem that can be computed with this strategy. The scalability of this approach is limited since access to the memory can become a bottleneck, and the number of processors that can efficiently be used to compute a given problem is consequently restricted. Since, furthermore, large shared memory computers are very expensive whereas large distributed memory machines are comparatively cheap, we employ a different approach for our 3d calculations. Here we distribute the macro grid over a number of processors, each of which has access to a local memory and which can communicate efficiently with each other to exchange data. We distribute the elements of the macro grid, and each processor only has access to the data on one part of the grid. This processor also carries out the time evolution for this data. The evolution of the hydrodynamic quantities can still be performed quite efficiently since only the data on the neighboring elements has to be available. If this data is stored on another processor it has to be exchanged; this is done using the standard message passing library MPI (cf. [Sch99a, DRSW02]).

The computation of the radiation source term is far more complicated on distributed memory computers. Up until now we have not extended our solvers for the RT equation in 2d to our 3d code. The main difficulty is that the radiation source term at a time level t^n depends on the hydrodynamic quantities on the whole grid. Consequently a high amount of communication is required to compute Q_{rad}. This can severely reduce the efficiency of the solver. One way to overcome this problem is to use similar iteration strategy like the one used to take care of periodic boundary conditions (cf. Section 12.4 and the discussion in the Outlook on page 271). This approach for computing the radiation source term seems promising since the hydrodynamic quantities barely change from one time step to the next so that the radiation intensity from the previous time level is a good starting point for computing the radiation field for the next time level.

3.7 Experimental Verification of the Scheme

Since we use an explicit finite–volume scheme for the solution of the coupled system (1.19), the main task lies in the construction of a numerical flux function $\mathbf{g}_{ij}$ for the real gas MHD equations (1.1) and the construction of a scheme for solving the radiation transport equation (1.16) for fixed data χ, B, and a fixed set of propagation angles $\{\boldsymbol{\mu}_m\}_{m=1}^M$. We assume that the overall performance of the scheme can be measured by the performance of the MHD and RT schemes, respectively. Consequently we focus on the derivation, analysis, and the numerical study of these separate modules. The independent study of the two main parts of the solution scheme is also necessary because to our knowledge there are no test cases with a non–trivial exact solution for the full real gas MHD system with radiation. The full scheme can only be verified in demanding physical tests. Therefore, the main aspects of this study are the derivation and verification of higher order schemes for the real gas MHD equations (1.1) and for the RT equation (1.16).

3.7.1 Experimental Order of Convergence (EOC)

For model problems the convergence properties of the finite–volume scheme can be studied with analytical tools (cf. Chapter 4), but for complex systems like (1.19) these issues can only be investigated in numerical tests. One method that allows us to study the convergence rate of a scheme is the computation of the *experimental order of convergence* (EOC). This approach is based on the assumption that the error e_h of a scheme measured in some suitable norm can be expressed in powers of the grid parameter h:

$$e_h := \|u - u_h\|_\Omega = C_1 h^\alpha + C_2 h^{\alpha+1} + C_3 h^{\alpha+2} + \dots \ . \tag{3.21}$$

Here u_h is the approximation on a grid $\mathcal{T}_h$ to the exact solution u of a given problem. The order of the scheme is given by the lowest power in h, i.e. by the constant α. With the error on two grids with parameters h_1 and h_2 we can approximate this value α in the following way:

$$\begin{aligned}
\ln\left(\frac{e_{h_1}}{e_{h_2}}\right) &= \ln\left(\frac{h_1^\alpha(C_1 + C_2 h_1 + C_3 h_1^2 + \dots)}{h_2^\alpha(C_1 + C_2 h^2 + C_3 h_2^2 + \dots)}\right) \\
&= \alpha \ln\left(\frac{h_1}{h_2}\right) + \ln(C_1 + C_2 h_1 + C_3 h_1^2 + \dots) - \ln(C_1 + C_2 h_2^2 + C_3 h_2^2 + \dots) \\
&\approx \alpha \ln\left(\frac{h_1}{h_2}\right) .
\end{aligned}$$

This approximation allows us to compute the relevant parameter α.

3.4 Definition (Experimental Order of Convergence (EOC))
Given a sequence of grids with parameters $h, \beta h, \beta^2 h, \dots$ and corresponding approximations $u_h, u_{\beta h}, u_{\beta^2 h}, \dots$ we define the EOC for a given norm $\|\cdot\|_\Omega$ by

$$\mathrm{EOC}_\beta := \frac{\ln\left(e_{\beta^i h}\right) - \ln\left(e_{\beta^{i+1} h}\right)}{\ln(\beta)} \ .$$

Here e_h denotes the error between the exact solution u and the approximation u_h

$$e_h := \|u - u_h\|_\Omega \ .$$

The main difficulty is that to compute the EOC of a given scheme suitable problems with exact solution u have to be available.

3.7.2 Efficiency of a Numerical Scheme

Another issue that is perhaps of even greater importance than the order of the scheme is its efficiency. We measure the efficiency in terms of the runtime required to reach a given error. The EOC gives no information concerning the actual error or required runtime of a given scheme. For a high grid resolution a higher order scheme will always be superior to a low resolution scheme — but this is not necessarily true on grids that can be used in applications. Furthermore, two schemes with identical order can perform very differently for a fixed grid resolution: on the one hand, this is due to two possibly different constants C_1 in (3.21); on the other hand, the computational costs of two schemes on a fixed grid can differ, so that the time required to reach a fixed error can vary considerably between two different schemes even if they are of identical order. This can result in a higher order scheme being less efficient on coarse grids than a lower order scheme. Consequently, the study of the error to runtime ratio is essential for a comparison of different schemes. As for the computation of the experimental order of convergence, however, we have to know the exact solution to a given problem to compute that error to runtime ratio. Therefore, the construction of solutions to a given system of PDEs is an important task for the experimental verification of numerical schemes; this issue is discussed in Section 3.7.5 below.

3.7.3 Efficiency of the Parallel Algorithm

The efficiency of a parallelization strategy is strongly influenced by the percentage of the computational load that can be distrubuted between the processors. Denote with $\Theta(1)$ the computational cost of the algorithm on one processor. If the algorithm can be split into P parts that run independently of each other, then the computational cost of the algorithm on P processors is $\Theta(P) = \Theta(1)/P$. This is optimal, but in many cases there are points in the algorithm where the processors have to be synchronized and where data have to be exchanged. Therefore we have to expect that $\Theta(P) > \Theta(1)/P$. For example, consider a case where $\Theta(1) = \Theta_{\text{ser}} + \Theta_{\text{par}}$ with Θ_{par} denoting the computational cost of that part of the algorithm that can be parallelized without any synchronization and with Θ_{ser} denoting the computational cost of that part of the algorithm that is executed by one processor alone. Then we compute that $\Theta(P) = \Theta_{\text{ser}} + \Theta_{\text{par}}/P$. For P small this function behaves like $\Theta(1)/P$, but for P large it is dominated by Θ_{ser}. Therefore the efficiency of the parallelization is reduced by increasing P. In many cases Θ_{par} increases with the size of the problem (e.g. the number of elements in the grid), whereas Θ_{ser} is often more or less independent of the problem size. Therefore the efficiency of the scheme depends, for example, on the number of elements in the grid. In the following we give a definition that allows us to measure the efficiency of a parallel algorithm for a given simulation.

3.5 Definition
The speedup $S(K;L)$ *of a parallel code running on* K *processors relative to the same code running on* L *processors is given by the ratio*

$$S(K;L) := \Theta(L)/\Theta(K) \ .$$

Here $\Theta(.)$ *denotes the total runtime of the complete parallel algorithm. The* optimal speedup $S_{\text{opt}}(K;L)$ *is given by* K/L*, which corresponds to a situation without parallel overhead.*
The efficency $E(K;L)$ *of a parallel code is given by* $S(K;L)/S_{\text{opt}}(K;L)$*, i.e.*

$$E(K;L) := S(K;L)(L/K) \tag{3.22}$$

3.7.4 Evaluation of Results

In the case where an exact solution $\mathbf{U}$ is available, a plot of the error versus the grid resolution h or versus the computational time can be used to compare different schemes with each other quantitatively. In many cases this gives a good impression of the quality of the methods. In addition it can be necessary to take a closer look at the approximate solutions to determine, for example, in which regions a scheme leads to a smearing of discontinuities and thus to a high approximation error. Or a scheme can produce a small error, but still show oscillatory behavior, which results in an unstable scheme. In the case where no exact solution is available, this type of qualitative study of a numerical scheme is often the only possible way to determine the quality of the approximation.
In one space dimension the evaluation of the results is straight forward in most cases. The approximation $\mathbf{U}_h$ can be simply plotted versus the space variable x. In higher space dimensions this is not so simple. For some problems we can utilize some symmetry in the solution $\mathbf{U}$ or at least in some component v of the solution vector. For example, v can only depend on one space variable or might be rotationally symmetric so that we can plot $v(|\mathbf{x}|)$ versus the radius $r = |\mathbf{x}|$. In this case the approximation can be visualized in a similar way as in 1d. We call the type of representation *scatterplot* in which we plot the approximation in the barycenter $\boldsymbol{\omega}_T$ of each element $T \in \mathcal{T}$ versus, for example, $|\boldsymbol{\omega}_T|$ or $\boldsymbol{\omega}_{T,z}$ (in the case of a radially symmetric solution or a solution that only depends on z, respectively). For these 1d representations of the solution we use the *GNUPlot* package [WK]. For an impression of the approximation in two or three space dimensions, we use the *GraPE programming library* [UFUB]. With this graphics environment we can plot isosurfaces of 3d solutions and isolines for 2d solutions. A representation of the piecewise constant or linear approximation on a triangular or tetrahedral grid is also possible. Note that we always use the same colorbar for a series of plots that we want to compare directly with each other.

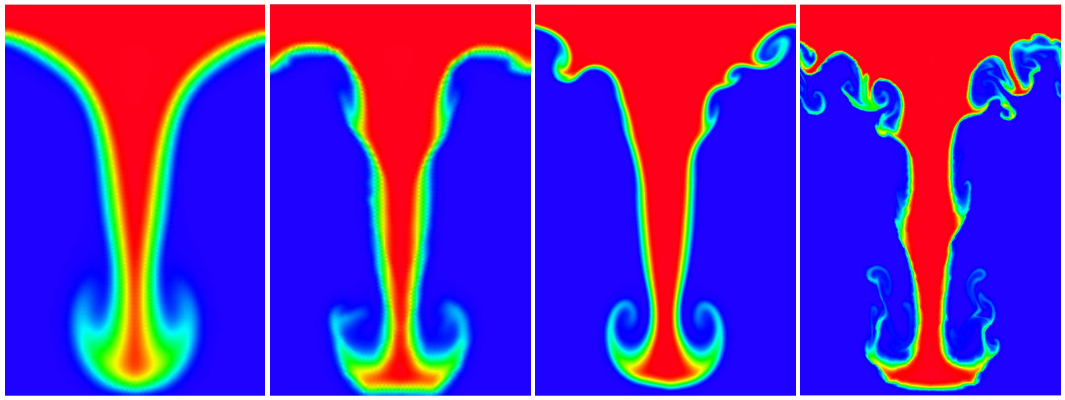
3.7.5 Constructing Solutions

Problems with exact solutions are often either rather trivial and, therefore, of very little interest or are very hard to construct. In one space dimension the most common test case for conservation laws like the MHD system (1.1) is the *Riemann problem.*

For many situations one can construct solutions for this problem or at least compute approximate solutions up to any given accuracy. The solution to the Riemann problem is, furthermore, the most important building block of the finite–volume scheme. In higher space dimensions the Riemann problem is also often used for test calculations (for example [LL98]).
A further possibility to determine the approximation quality of a scheme is to use a finely resolved approximation as reference solution in place of the exact solution. This is easily done, but one should keep in mind that the numerical scheme that is to be tested is also used to compute the reference solution, and, therefore, we have to treat this kind of test with caution. A last resort is to choose a function $\mathbf{U}(\mathbf{x}, t)$ and to define suitable source terms $\mathbf{q}$ so that $\mathbf{U}$ is a solution to the system of PDEs augmented by the artificial source terms $\mathbf{q}(\mathbf{U}, \mathbf{x}, t)$. This method gives us the possibility of prescribing special features of the solution $\mathbf{U}$, for example, its smoothness, so that we can study the behavior of the scheme under special circumstances. On the other hand, the source term $\mathbf{q}$ is totally artificial and has no physical relevance, so that it is not always clear if the insights obtained in this way can be extended to relevant physical settings. Nevertheless, this is an important tool for verifying numerical schemes.

3.7.6 Instabilities

The last problem we want to discuss in this section is the problem of instabilities in the solution. Even in the case of apparently simple initial data, small scale structures can develop that grow in time and are not reduced by grid refinement. Higher order schemes or local adaptivity can, in fact, lead to an additional amplification of these small scale structures. In our applications two types of instabilities play an important role. The first type of instability is the so called *Rayleigh–Taylor* instability, which occurs when a heavy fluid is superimposed on a light fluid with respect to the gravitational force vector $\mathbf{g}$. This instability leads to so–called fingering, where the heavy fluid "falls" into the light fluid. Linear stability analysis can be found, for example, in [Cha81, Chapter 10], where it is shown that a magnetic field normal to the gravitational force leads to a stabilization of the interface. At a certain critical magnetic field strength the setting becomes stable. In Figure 3.1 we show a series of simulations for a hydrodynamic Rayleigh–Taylor instability. As our results demonstrate, the solution develops an increasing amount of small scale structures if a higher order scheme or locally adapted grids are used. To verify if, nevertheless, the higher resolution schemes lead to better results an independent quantity has to be found that can be measured a posteriori. An important factor is the so–called *growth rate* of the instability. In [DKRW01b] we have performed numerical tests and measured the growth rate following the ideas published in [JNS95].
A second type of instability arises at the sides of the interface, where the fluid is moving downwards. This is the so–called *Kelvin–Helmholtz* instability. This occurs at an interface between two fluids moving at different speeds. This type of interface is also studied in [Cha81, Chapter 11]. Again it can be shown that a magnetic field tangential to the interface has a stabilizing effect. The simulation shown on the title page demonstrates the effects of this type of instability: the interface should be circular, but due to the moving plasma in the interior of the circle the interface is unstable.



(a) no adaptation, first order scheme

(b) no adaptation, second order scheme

(c) with adaptation, first order scheme

(d) with adaptation, second order scheme

Figure 3.1: Solution to the Rayleigh–Taylor instability using different schemes and grid structures. Although the initial conditions are always the same, the approximations at $t = 2$ differ strongly at least qualitatively. In this case it is important to find a relevant quantity that can be used to measure the quality of the solution. Results are taken from [DKRW01b].

We study this problem in more detail in the following chapters; a film showing the development of the Kelvin–Helmholtz instability from the title page can be found in the enclosed CD.

Chapter 4

Analytical Results

For complex systems of balance laws like the RMHD equations (1.19) very few analytical results are available. Even the question of the existence and the uniqueness of solutions for simple initial data in one space dimension has not yet been fully answered. Consequently, the convergence of numerical schemes — such as the finite–volume scheme presented in Chapter 3 — has not been shown. The analysis of scalar balance laws is much more developed. Even in higher space dimensions, questions like the existence and uniqueness of solutions are solved and convergence results for finite–volume schemes are available. In this chapter we summarize some of these results. The scope of our presentation is motivated by our applications; general concepts can be found, for example, in [Krö97, War99, Daf00]. We especially focus on results for the Cauchy problem for balance laws with non–local operators $\widehat{\mathcal{T}}$ of the general form

$$\begin{aligned} \partial_t u(\mathbf{x},t) + \nabla \cdot \mathbf{f}(u(\mathbf{x},t),\mathbf{x},t) = \widehat{\mathcal{T}}[u(\cdot,t)](\mathbf{x},t) \ ,&\mathbf{x} \in \mathbb{R}^d,\ t \in (0,T), \\ u(\mathbf{x},0) = u_0(\mathbf{x}) \ , \qquad\qquad &\mathbf{x} \in \mathbb{R}^d. \end{aligned}$$

These serve as a model problem for our coupled system of MHD and instantaneous radiation transport (1.19). The scalar quantity u is a lumped quantity of the hydrodynamic variables defining the radiation field and $\widehat{\mathcal{T}}$ represents the radiation source term Q_{rad}. A detailed analytical understanding of simplified model problems and of the numerical schemes applied to this setting is one important step in the justification of the use of these methods in the case of complex systems. A second step is the numerical study of these schemes applied to complex system with simplified data. This step is the focus of the following chapters.

After fixing some notation we describe the model problem and its connection to the system (1.19) of radiation magnetohydrodynamics in more detail (Section 4.1). The novel feature of this model is the non–local operator $\widehat{\mathcal{T}}$. The case of a local source term, i.e., $\widehat{\mathcal{T}}[u(\cdot,t)](\mathbf{x},t) = q(u(\mathbf{x},t),\mathbf{x},t)$, which is included in our model, has been thoroughly studied and the existence and uniqueness of a solution in a suitable function space has been proven. This result will be given in Section 4.2. In Section 4.3 we study the influence of the non–local operator. We concentrate on the one–dimensional setting using analytical results when available and high resolution numerical results in the other cases. The main aspect is the study of the regularization effects of the non–local term, which we compare with the approximation of the solution due to a

viscous regularization of the homogeneous model problem. The analytical results for the non–linear case are taken from [KN99b, KN99a]. In the last two sections of this chapter we study the convergence of finite–volume schemes in two space–dimensions. The homogeneous case and the case with local source terms have been thoroughly studied in the literature, for example in [Vil94, EGGH98, Roh98, CHC01]. First we summarize a convergence result for the case of local source terms taken from [CHC01]. We extend this result in Section 4.5 to include non–local operators. For a special class of operators that include the operator defining the radiation field we present a fully discrete approximation for which our convergence theorem is applicable. These results were published in [DR02b, DR02a].
In the following we use the standard notation for function spaces and corresponding norms. In particular we use $L^p(S)$, $p \in [1,\infty]$ to denote the usual Lebesgue spaces on $S \subset \mathbb{R}^d$ and with $C^k(S)$, $k \in \mathbb{N}_0$, the space of k-times continuously differentiable functions on S. We also need the space BV of functions of bounded variation:

$$\mathrm{BV}(S) = \left\{ w \in L^1(S) \,\middle|\, |w|_{BV(S)} := \sup_{\phi \in C_0^\infty(S),\, \|\phi\|_\infty \le 1} \left\{ \int_S w(\mathbf{x}) \mathrm{div}(\phi(\mathbf{x}))\, d\mathbf{x} \right\} < \infty \right\} .$$

Furthermore we make the subsequent convention for the corresponding norms of functions $u \in L^p(\mathbb{R}^d \times [0,T])$ and $w \in L^p(\mathbb{R}^d)$, $p \in [1,\infty) \cup \{\infty\}, T > 0$:

$$\|u\|_p = \|u\|_{L^p(\mathbb{R}^d \times [0,T])}, \quad \|w\|_p = \|w\|_{L^p(\mathbb{R}^d)} .$$

The expressions $\|\cdot\|_{C^1}$ and $|\cdot|_{BV}$ have to be understood in the same manner. Norms of spaces of functions acting on other subsets of $\mathbb{R}^d \times [0,T]$ or $\mathbb{R}^d$ will be explicitly given. Furthermore we define the space $W(0,T)$ by

$$W(0,T) = L^\infty(0,T; L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)). \tag{4.1}$$

With C we denote a generic constant that can take different values in the set of positive real numbers $\mathbb{R}^+$. In the following we restrict ourselves to the case of $d = 1$ or $d = 2$ although some of the results presented in this chapter have been shown for arbitrary $d \in \mathbb{N}$.

4.1 A Model Problem

To derive the model problem we study the solution of the radiation transport equation (1.16a) in more detail. The radiation intensity I satisfies

$$\boldsymbol{\mu} \cdot \nabla I(\mathbf{x}, t, \boldsymbol{\mu}) = \chi(B(\theta(\mathbf{x}, t)) - I(\mathbf{x}, t, \boldsymbol{\mu})) , \tag{4.2}$$

and we assume in the rest of this chapter that the absorption coefficient χ is a positive constant. As in Section 1.2 we set $B(\theta) = \sigma\theta^4$ with $\sigma > 0$ constant. Note that every unit vector $\boldsymbol{\mu} \in S^2$ can be written in the form

$$\boldsymbol{\mu} = \begin{pmatrix} \sin(\vartheta)\cos(\varphi) \\ \sin(\vartheta)\sin(\varphi) \\ \cos(\vartheta) \end{pmatrix} \tag{4.3}$$

with $\vartheta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$. For our model problem we couple the radiation source term to a scalar conservation law for the temperature $\theta > 0$. Thus we seek the positive solution to the scalar balance law

$$\partial_t \theta(\mathbf{x}, t) + \nabla \cdot \mathbf{f}(\theta(\mathbf{x}, t)) = Q_{\text{rad}}(\mathbf{x}, t)$$

where $\mathbf{f}$ is some general non–linear flux function. The radiation source term Q_{rad} is defined by integrating the intensity over the unit sphere (cf. (1.19g)). Using the representation (4.3) we can express the integral over $\boldsymbol{\mu}$ as an integral over ϑ and φ

$$Q_{\text{rad}}(\mathbf{x}, t) = \chi \int_0^{\pi} \int_0^{2\pi} I(\mathbf{x}, t, \varphi, \vartheta) \sin(\vartheta) d\varphi d\vartheta - 4\pi\chi B(\theta(\mathbf{x}, t)) \ . \tag{4.4}$$

Since in the situation studied here the intensity depends only on the temperature θ, we can think of the radiation source term as a mapping of θ to Q_{rad} by some operator $\widehat{\mathcal{T}}$. In the following we derive some explicit expressions for this operator first in 2d and then in 1d. Note that this operator is non–linear due to the non–linearity B in the radiation transport equation. It also has to be a non–local mapping in space since the radiation field at a fixed point is influenced by the temperature of the fluid everywhere. In the following derivation, we do not always explicitly include the dependency of the radiation intensity on φ and ϑ.

4.1.1 The Radiation Operator in Two Space Dimensions

If we assume $\partial_z I \equiv 0$ we arrive at the equation

$$\bar{\boldsymbol{\mu}} \cdot \nabla I(\mathbf{x}, t) = \frac{\chi}{\sin(\vartheta)} (B(\theta(\mathbf{x}, t)) - I(\mathbf{x}, t))$$

for $\vartheta \in (0, \pi)$. We use the abbreviation $\bar{\boldsymbol{\mu}} := (\cos(\varphi), \sin(\varphi))^T$. For $\vartheta = 0$ or $\vartheta = \pi$ the solution to (4.2) is given by $I = B$. To arrive at a closed form for the operator defining Q_{rad} we derive the formal solution for the intensity I. The radiation transport equation can be rewritten as

$$\frac{d}{ds} I(\mathbf{x} + s\bar{\boldsymbol{\mu}}, t) = \frac{\chi}{\sin(\vartheta)} (B(\theta(\mathbf{x} + s\bar{\boldsymbol{\mu}}, t)) - I(\mathbf{x} + s\bar{\boldsymbol{\mu}}, t)) \ . \tag{4.5}$$

Therefore the intensity at a point $\mathbf{x} \in \mathbb{R}^2$ under the assumption that $\theta(\cdot, t)$ is bounded is given by

$$I(\mathbf{x}, t) = \frac{\chi}{\sin(\vartheta)} \int_0^{\infty} B(\theta(\mathbf{x} - \sigma\bar{\boldsymbol{\mu}}, t)) e^{-\frac{\chi}{\sin(\vartheta)}\sigma} d\sigma \ . \tag{4.6}$$

Thus we arrive at the following expression for the radiation source term using (4.4)

$$Q_{\text{rad}}(\mathbf{x}, t) = \int_0^{2\pi} \int_0^{\pi} \chi^2 \int_0^{\infty} B(\theta(\mathbf{x} - \sigma\bar{\boldsymbol{\mu}}, t)) e^{-\frac{\chi}{\sin(\vartheta)}\sigma} d\sigma d\vartheta d\varphi - 4\pi\chi B(\theta(\mathbf{x}, t))$$

$$= \int_0^{2\pi} \int_0^{\infty} \chi^2 B(\theta(\mathbf{x} - \sigma\bar{\boldsymbol{\mu}}, t)) \int_0^{\pi} e^{-\frac{\chi}{\sin(\vartheta)}\sigma} d\vartheta d\sigma d\varphi - 4\pi\chi B(\theta(\mathbf{x}, t)) \ .$$

Switching from the polar coordinates (σ, φ) to Cartesian coordinates leads to

$$\begin{aligned} Q_{\text{rad}}(\mathbf{x}, t) &= \chi^2 \int_{\mathbb{R}^2} B(\theta(\mathbf{x} - \mathbf{y}, t)) \frac{\int_0^{\pi} \exp\big(-\frac{\chi}{\sin(\vartheta)}|\mathbf{y}|\big) d\vartheta}{|\mathbf{y}|} d\mathbf{y} - 4\pi\chi B(\theta(\mathbf{x}, t)) \\ &= \chi^2 \int_{\mathbb{R}^2} B(\theta(\mathbf{y}, t)) \frac{\int_0^{\pi} \exp\big(-\frac{\chi}{\sin(\vartheta)}|\mathbf{x} - \mathbf{y}|\big) d\vartheta}{|\mathbf{x} - \mathbf{y}|} d\mathbf{y} - 4\pi\chi B(\theta(\mathbf{x}, t)) \ . \end{aligned}$$

This leads to the following expression for the mapping from θ to Q_{rad}:

$$\widehat{\mathcal{T}}[w(\cdot)](\mathbf{x}) := \chi^2 \int_{\mathbb{R}^2} B(w(\mathbf{y})) k(\mathbf{x}, \mathbf{y}) d\mathbf{y} - \chi 4\pi B(w(\mathbf{x})) \ , \tag{4.7}$$

for a function $w : \mathbb{R}^2 \to \mathbb{R}$. The kernel function $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is given by

$$\begin{aligned} k(\mathbf{x}, \mathbf{y}) &= \begin{cases} \tilde{k}(\mathbf{x}, \mathbf{y}) |\mathbf{x} - \mathbf{y}|^{-1} & \mathbf{x} \neq \mathbf{y} \ , \\ 0 & \mathbf{x} = \mathbf{y} \ , \end{cases} \\ \tilde{k}(\mathbf{x}, \mathbf{y}) &= \int_0^{\pi} \exp\Big(-\frac{\chi|\mathbf{x} - \mathbf{y}|}{\sin(\vartheta)}\Big) d\vartheta \quad (\mathbf{x}, \mathbf{y} \in \mathbb{R}^2) \ . \end{aligned} \tag{4.8}$$

Since $\chi^2 \int_{\mathbb{R}^2} k(\mathbf{x}, \mathbf{y}) d\mathbf{y} = 4\pi\chi$ we can rewrite (4.7) in the form

$$\widehat{\mathcal{T}}[w(\cdot)](\mathbf{x}) = \chi^2 \int_{\mathbb{R}^2} \big(B(w(\mathbf{y})) - B(w(\mathbf{x}))\big) k(\mathbf{x}, \mathbf{y}) d\mathbf{y} \ . \tag{4.9}$$

4.1.2 The Radiation Operator in One Space Dimension

A similar analysis to the one presented above can be performed for the one dimensional case. It is convenient to assume that $\partial_x I \equiv 0$ and $\partial_y I \equiv 0$, and we use x instead of z in the following. Then the radiation transport equation is

$$\mu I'(x, t) + \chi I(x, t) = \chi B(\theta(x, t)) \tag{4.10}$$

for $x \in \mathbb{R}$ and $\mu = \cos(\vartheta) \in [-1, 1]$. In the integral (4.4) we can substitute μ for $\cos(\theta)$

$$Q_{\text{rad}} = 2\pi\chi \int_{-1}^{1} I d\mu - 4\pi\chi B(\theta) \ .$$

Note that I does not depend on φ so that the integral over φ is not present in the representation for Q_{rad}. Now we can derive an expression for the radiation operator by

again replacing I with the formal solution of the ODE (4.10), which depends upon the sign of μ:

$$I(x,t) = \begin{cases} \frac{\chi}{\mu} \int_x^\infty B(\theta(y,t)) e^{-\frac{\chi}{\mu}(x-y)} dy & \text{for } \mu < 0 \ , \\ \frac{\chi}{\mu} \int_{-\infty}^x B(\theta(y,t)) e^{-\frac{\chi}{\mu}(x-y)} dy & \text{for } \mu > 0 \ . \end{cases} \tag{4.11}$$

At a point $(x,t) \in \mathbb{R} \times \mathbb{R}^+$ the radiation source term is given by

$$\begin{aligned} Q_{\text{rad}} &= 2\pi\chi\Big(-\int_{-1}^0 \int_x^\infty \frac{\chi}{\mu} B(\theta(y,t)) e^{-\frac{\chi}{\mu}(x-y)} dy d\mu + \int_0^1 \int_{-\infty}^x \frac{\chi}{\mu} B(\theta(y,t)) e^{-\frac{\chi}{\mu}(x-y)} dy d\mu \Big) \\ &\quad - 4\pi\chi B(\theta(x,t)) \\ &= 2\pi\chi\Big(\int_0^1 \int_x^\infty \frac{\chi}{\mu} B(\theta(y,t)) e^{\frac{\chi}{\mu}(x-y)} dy d\mu + \int_0^1 \int_{-\infty}^x \frac{\chi}{\mu} B(\theta(y,t)) e^{-\frac{\chi}{\mu}(x-y)} dy d\mu \Big) \\ &\quad - 4\pi\chi B(\theta(x,t)) \\ &= 2\pi\chi \int_0^1 \Big(\int_x^\infty \frac{\chi}{\mu} B(\theta(y,t)) e^{-\frac{\chi}{\mu}|x-y|} dy d\mu + \int_{-\infty}^x \frac{\chi}{\mu} B(\theta(y,t)) e^{-\frac{\chi}{\mu}|x-y|} dy d\mu \Big) \\ &\quad - 4\pi\chi B(\theta(x,t)) \ . \end{aligned}$$

Now we can combine the two integrals over y to one integral over $\mathbb{R}$. For simplicity of notation we also drop the factor 2π from both summands. We arrive at $Q_{\text{rad}} = \widehat{\mathcal{T}}[\theta(\cdot,t)]$ using the following integral operator:

$$\widehat{\mathcal{T}}[w](x) = \chi^2 \int_{-\infty}^\infty B(w(y)) \int_0^1 \frac{e^{-\frac{\chi}{\mu}|x-y|}}{\mu} d\mu dy - 2\chi B(w(x)) \tag{4.12}$$

For further details see [Ded98].

4.1.3 A General Model Problem

With the model problem for the system of radiation hydrodynamics in mind, we study the following more general problem for $d = 1$ or $d = 2$:

4.1 Definition (Model Problem for the RMHD system)
We seek a scalar function $u = u(\mathbf{x},t) : \mathbb{R}^d \times [0,T) \to \mathbb{R}$ with $T > 0$ fixed that satisfies

$$\partial_t u(\mathbf{x},t) + \nabla \cdot \mathbf{f}(u(\mathbf{x},t),\mathbf{x},t) = \widehat{\mathcal{T}}[u(\cdot,t)](\mathbf{x},t) \ , \quad \mathbf{x} \in \mathbb{R}^d, \, t \in (0,T), \tag{4.13a}$$

$$u(\mathbf{x},0) = u_0(\mathbf{x}) \ , \qquad \mathbf{x} \in \mathbb{R}^d. \tag{4.13b}$$

Here $f : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^d$ denotes a general flux function, and u_0 is a scalar function on $\mathbb{R}^d$. For functions $B : \mathbb{R} \to \mathbb{R}$, $q : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}$, and a kernel function

$k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, the balance term $\widehat{\mathcal{T}}$ takes the form

$$\begin{aligned} \widehat{\mathcal{T}}[w](\mathbf{x},t) &= \mathcal{T}[w](\mathbf{x}) + q(w(\mathbf{x}),\mathbf{x},t) \ , \\ \mathcal{T}[w](\mathbf{x}) &= \int_{\mathbb{R}^d} k(\mathbf{x},\mathbf{y}) B(w(\mathbf{y}))\, d\mathbf{y} \ , \end{aligned} \tag{4.13c}$$

for $\mathbf{x} \in \mathbb{R}^d, t \in [0,T)$ and functions $w : \mathbb{R}^d \to \mathbb{R}$.

4.2 Remark: *As we already saw that problem (4.13) is a model for the coupled system (1.19) in the same sense that a scalar conservation law is taken as a model for the hydrodynamics equations. The non–local operator $\widehat{\mathcal{T}}$ is derived from the radiation source term $Q_{\rm rad}$ in the energy balance equation. The unknown scalar function u replaces the temperature θ of the fluid, which defines the radiation field.*
In [DR02b] we prove a convergence theorem for a finite–volume scheme for the model problem (4.13) where the flux vector is indpendent of $\mathbf{x}$ and t. In Chapter 9 we need some analytical results for the model problem with local source term ($\mathcal{T}[w] = 0$) but with a flux function $\mathbf{f}$ and a local source q that explicitly depend on $\mathbf{x}$ and t. In the following we, therefore, distinguish between the two cases $\mathcal{T}[w] = 0$ and $\mathcal{T}[w] \neq 0$.

4.3 Assumption (Continuous Data)

(i) The flux function $\mathbf{f} = \mathbf{f}(s,\mathbf{x},t)$ satisfies

- $\mathbf{f} \in C^1(\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^+)$,
- $\partial_s \mathbf{f}$ *is locally Lipschitz,*
- *for every compact $K \subset \mathbb{R}$ there exists $V_K < \infty$ such that $|\partial_s f| < V_K$ in $K \times \mathbb{R}^d \times \mathbb{R}^+$,*
- $\mathrm{div}_{\mathbf{x}} f$ *is locally Lipschitz and there exist constants D^0, D^1 with $\sup_{\mathbb{R}^d \times \mathbb{R}^+} |\mathrm{div}_{\mathbf{x}} f(0,\mathbf{x},t)| < D^0$ and $\sup_{\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^+} |\partial_s \mathrm{div}_{\mathbf{x}} f(s,\mathbf{x},t)| < D^1$.*

(ii) For the balance term we distinguish between two cases: If the term is local, i.e., $\mathcal{T} \equiv 0$ then we assume that the source term $q = q(s,\mathbf{x},t)$ satisfies

- $q \in C^1(\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^+)$,
- *q is locally Lipschitz and there exist constants Q^0, Q^1 with $\sup_{\mathbb{R}^d \times \mathbb{R}^+} |q(0,\mathbf{x},t)| < Q^0$ and $\sup_{\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^+} |\partial_s q| < Q^1$.*

If $\mathcal{T} \not\equiv 0$ then we assume that q, f do not depend explicitly on $\mathbf{x}$ and t, i.e., $q = q(s), \mathbf{f} = \mathbf{f}(s)$ and for $\widehat{\mathcal{T}} = \widehat{\mathcal{T}}[w]$ given by (4.13c) we assume that there exists a continuous function $C_{\widehat{\mathcal{T}}} : \mathbb{R}^+ \to \mathbb{R}^+$ such that for all $w, \tilde{w} \in L^\infty(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ with $\|w\|_\infty, \|\tilde{w}\|_\infty \leq M$, $M > 0$

- $\widehat{\mathcal{T}} : L^\infty(\mathbb{R}^d) \cap L^1(\mathbb{R}^d) \to L^\infty(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$,
- $\|\widehat{\mathcal{T}}[w]\|_\infty < C_{\widehat{\mathcal{T}}}(M)$ *and* $\|\widehat{\mathcal{T}}[w]\|_1 < C_{\widehat{\mathcal{T}}}(M) \|w\|_1$,
- $\int_{\mathbb{R}^d} \left| \widehat{\mathcal{T}}[w](\mathbf{x}) - \widehat{\mathcal{T}}[\tilde{w}](\mathbf{x}) \right| d\mathbf{x} < C_{\widehat{\mathcal{T}}}(M) \int_{\mathbb{R}^d} \left| w(\mathbf{x}) - \tilde{w}(\mathbf{x}) \right| d\mathbf{x}$.

(iii) We assume that

- $u_0 \in L^\infty(\mathbb{R}^d) \cap BV(\mathbb{R}^d)$ *and for some* $R_0 > 0$*:* $\operatorname{supp}(u_0) \subset B_{R_0}(0)$.

For the discretization of (4.13) we use a general finite–volume scheme as presented in Section 3.2 on a family of grids $\{\mathcal{T}_h\}$ for $h \in (0, h_0]$. For each $h \in (0, h_0]$, let V_h denote the set of functions on $\mathbb{R}^d$ which are piecewise constant on the cell volumes T_i with $i \in \mathcal{I}_h$. The constant value of a grid function $w_h \in V_h$ on T_i is denoted with w_i. We make the following assumptions on the grid and the time–step:

4.4 Assumption (Grid and time step)
Let $\{\mathcal{T}_h\}$ *be a family of unstructured grids on* $\mathbb{R}^d$ *and* $\Delta t > 0$*. In contrast to the definition of the grid parameter* h *used in the case of locally adapted grids (cf. Section 2.2), we use* $h = \max_{i \in \mathcal{I}_h} h_i$ *in this Chapter. We assume that there exist constants* $c_G, c_1 > 0$ *such that for all* $j \in \mathcal{I}_h$ *and* $l \in \mathcal{N}(j)$

$$\frac{h}{\Delta t} \le c_1, \quad c_G h^d \le |T_j|, \quad c_G |S_{jl}| \le h. \tag{4.14}$$

From these estimates it follows that $|T_j| \le C h^d$ *for all* $j \in \mathcal{I}_h$.

The flux $\mathbf{f}$ is discretized by a family of monotone numerical flux functions $\{g_{ij}\}_{(i,j) \in \mathcal{I}_S}$:

4.5 Definition (Monotone Numerical Flux Functions)
For $i \in \mathcal{I}$ *and* $j \in \mathcal{N}(i)$ *let* $g_{ij} = g_{ij}^n(u, v)$ *for* $(u, v) \in \mathbb{R}^2$ *be a numerical flux function which satisfies*

(i) $g_{ij}^n(u, v) = -g_{ji}^n(v, u)$,

(ii) $g_{ij}^n(u, u) = \frac{1}{\Delta t} \int_{t^n}^{t^{m+1}} \int_{S_{ij}} \mathbf{f}(u, \mathbf{x}, t) \cdot \mathbf{n}_{ij}(\mathbf{x}) \, d\mathbf{x} dt$,

(iii) $g_{ij}^n(u, v)$ *is non–decreasing with respect to* v,

(iv) $g_{ij}^n(u, v)$ *is locally Lipschitz continuous on* $\mathbb{R}^d$*, i.e., for each* $M > 0$ *there exists a constant* $L_g(M) > 0$ *so that for all* u_1, u_2, v_1, v_2 *with* $|u_1|, |u_2|, |v_1|, |v_2| < M$

$$|g_{ij}^n(u_1, v_1) - g_{ij}^n(u_2, v_2)| \le |S_{ij}| L_g(M) \big(|u_1 - u_2| + |v_1 - v_2| \big) \ .$$

4.6 Remark: *Since the continuous flux depends explicitly on time* t*, the numerical flux function has to depend on the time–step* t^n*; this is expressed by the additional superscript* n*. In the case where* $\mathbf{f}$ *does not depend explicitly on* $\mathbf{x}$ *and* t*, conditions 4.5(i) and 4.5(ii) are equivalent to conditions* (3.14) *and* (3.15) *given in Section 3.2. In this case many numerical flux functions that satisfy Assumption 4.5 can be found in the literature, e.g. [Krö97, Example 3.3.19].*

For the discretization of the balance term we assume the existence of a discrete operator $\widehat{\mathcal{T}}_h : V_h \to V_h$. The precise assumptions on this operator are given in Sections 4.4 and 4.5, where the convergence of the finite–volume scheme is discussed. One possible choice for $(\mathbf{x}, t) \in T_i \times (t^n, t^{n+1}]$ is

$$\widehat{\mathcal{T}}_h[w_h](\mathbf{x}, t) = \frac{1}{\Delta t |T_i|} \int_{t^n}^{t^{n+1}} \int_{T_i} \widehat{\mathcal{T}}[w_h](\mathbf{x}, t) d\mathbf{x} dt \ . \tag{4.15}$$

4.7 Definition (Finite–Volume Scheme)
Consider the problem (4.13) *in* $\mathbb{R}^d \times [0,T)$ *for* $T > 0$. *Let a family of unstructured grids* $\{\mathcal{T}_h\}$, $h \in (0, h_0]$, *and* $\Delta t > 0$ *be given that satisfy Assumption 4.4. Define* $N_T \in \mathbb{N}$ *as the smallest number with* $\Delta t N_T > T$. *For* $n = 1, \ldots, N_T - 1$, $i \in \mathcal{I}_h$, *and* $j \in \mathcal{N}(i)$ *let* $\{g_{ij}^n\}$ *be a family of monotone numerical flux functions. Assume that* $\widehat{\mathcal{T}}_h$ *is an admissible discrete operator for* $\widehat{\mathcal{T}}$ *on* $\mathcal{T}_h$.
For $n = 1, \ldots, N_T - 1$ *and* $j \in \mathcal{I}_h$, *we define iteratively*

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{|T_i|} \sum_{j \in \mathcal{N}(i)} g_{ij}^n(u_i^n, u_j^n) + \Delta t \widehat{\mathcal{T}}_h[u_h(\cdot, t^n)](\boldsymbol{\omega}_i, t^n) \ , \tag{4.16a}$$

where u_i^0 *is given by*

$$u_i^0 = \frac{1}{|T_i|} \int_{T_i} u_0(\mathbf{x}) \, d\mathbf{x} \ . \tag{4.16b}$$

The approximate solution $u_h : \mathbb{R}^d \times [0, N_T \Delta t] \to \mathbb{R}$ *is defined by* $u_h(\mathbf{x}, 0) = u_i^0$ *for* $\mathbf{x} \in T_i$, *and by*

$$u_h(\mathbf{x}, t) = u_i^{n+1} \text{ for } (\mathbf{x}, t) \in T_i \times (t^n, t^{n+1}] \tag{4.17}$$

for $n \in \{0, \ldots, N_T - 1\}$ *and* $i \in \mathcal{I}_h$.

4.8 Remark: *Definition 4.7 corresponds to the first order finite–volume scheme introduced in Section 3.2. The existence of a discrete operator* $\widehat{\mathcal{T}}_h$ *for* $\widehat{\mathcal{T}}$ *in the case of the radiation source term in 2d (cf.* (4.9)*) is discussed in Section 4.5, where all the assumptions on the data and discretization for this setting are studied.*

4.2 Existence and Uniqueness of Solutions

For non–linear scalar conservation law with zero right hand side, it is well known that one cannot expect a smooth solution for large times even for smooth initial data. We demonstrate the problem briefly for the one dimensional setting using Burgers equation with $f(u) = \frac{1}{2}u^2$ and $\widehat{\mathcal{T}} \equiv 0$ in (4.13). It has been shown that for $u_0 \in H^1(\mathbb{R})$ with $u_0'(x) \geq 0$ for all $x \in \mathbb{R}$ there exists a classical solution for all time. In the case where $u_0'(x) < 0$ for some $x \in \mathbb{R}$ there exists a pair $(x_0, t_0) \in \mathbb{R} \times \mathbb{R}^+$ with $\partial_x u(x,t) \to -\infty$ for $(x,t) \to (x_0, t_0)$, i.e., the derivative of the solution blows up in finite time and no global classical solution exists [Krö97, Lemma 2.1.2]). Therefore the notion of classical solutions is not suitable in the context of conservation laws. A regularization of the solution can be achieved by introducing a second order term into the equation. For example the parabolic equation

$$\partial_t u_\varepsilon(x,t) + \partial_x f(u_\varepsilon(x,t)) = \varepsilon \partial_{xx}^2 u_\varepsilon(x,t) \tag{4.18}$$

with $\varepsilon > 0$ admits a globally smooth solution in $\mathbb{R} \times \mathbb{R}^+$. The question arises if a global integral operator of the type that we derived for the radiation source term can also lead to the same kind of regularization. This was studied for a similar problem in [KNN98, KN99b, KN99a].

4.9 Theorem (Regularity of Solutions)
Consider the model problem (4.13) *with* $d = 1$, *flux function* $f(u) = \frac{1}{2}u^2$, *and balance term* $\widehat{\mathcal{T}}[w](x) = \frac{1}{2}\int_{-\infty}^{\infty}\big(w(y) - w(x)\big)\exp(-|x-y|)dy$. *Assume that the initial data* u_0 *is smooth and bounded and define the constants* $\delta_0 := \sup_x u_0(x) - \inf_x u_0(x)$, $k_0 := \min\{\frac{1}{2}\delta_0, \sup_x u_0'(x)\}$, $\omega_* := \frac{-1-\sqrt{1+4k_0}}{2} \leq -1$, *and* $\omega_{**} := \frac{-1-\sqrt{1-2\delta_0}}{2} > \omega_*$.

(i): If there exists a $x_0 \in \mathbb{R}$ *with* $u_0'(x_0) < \omega_*$ *then problem* (4.13) *does not admit a global classical solution since* $\partial_x u(x,t)$ *blows up in finite time. If we set* $R := \frac{\omega_2 - u_0'(x_0)}{\omega_1 - u_0'(x_0)}$ *with* $\omega_1 := \omega_*$ *and* $\omega_2 := \frac{-1+\sqrt{1+4k_0}}{2} \geq 0$ *then* $\partial_x u(x,t) \to -\infty$ *before* t *reaches* $t_0 := \log(R)$. *Note that* $R > 1$ *and therefore* $t_0 \in (0, \infty)$.

(ii): If $\delta_0 \leq \frac{1}{2}$ *and* $u_0'(x) > \omega_{**}$ *for all* $x \in \mathbb{R}$ *then the problem* (4.13) *admits a global smooth solution with*

$$\inf_x u_0(x) \leq u(x,t) \leq \sup_x u_0(x) \ .$$

Proof:
The proof of this Theorem relies on a maximum principle for classical solutions of (4.13) and their derivatives. The blow up is shown by studying the solution along characteristics. Details can be found in [KN99a]. □

4.10 Remark: *The operator used in the previous theorem is a slightly simplified version of our radiation transport operator* (4.12). *This result together with numerical experiments indicate that balance terms of the form studied here only lead to a moderate regularization that is by no means as strong as the regularization due to a second order term. Therefore one has to assume that solutions* u *to* (4.13) *are discontinuous in general.*
In contrast to the situation for the homogenous Burgers equation the threshold value ω_* *from Theorem 4.9 is not a constant, but depends on the values of the initial data on the whole real axis (through the value* k_0*). This mirrors the non–local influence of the operator* $\widehat{\mathcal{T}}$.

Since weak solutions are in general not unique, the notion of entropy solution turns out to be a suitable choice for hyperbolic balance laws. We use the *Kruzhkov* definition of an entropy solution:

4.11 Definition (Entropy Solution)
For a function $v \in W(0,T)$, $\kappa \in \mathbb{R}$, *and* $\phi \in C_0^\infty(\mathbb{R}^d \times [0,T))$ *we introduce the form*

$$\begin{aligned} E(v,\kappa,\phi) = &\int_0^T\int_{\mathbb{R}^d} |v(\mathbf{x},t) - \kappa|\partial_t\phi(\mathbf{x},t)d\mathbf{x}dt + \\ &\int_0^T\int_{\mathbb{R}^d} \big(\mathbf{f}(v(\mathbf{x},t)\top\kappa,\mathbf{x},t) - \mathbf{f}(v(\mathbf{x},t)\perp\kappa,\mathbf{x},t)\big)\cdot\nabla\phi(\mathbf{x},t)d\mathbf{x}dt - \\ &\int_0^T\int_{\mathbb{R}^d} \operatorname{sgn}(u(\mathbf{x},t)-\kappa)\Big(\operatorname{div}_{\mathbf{x}}\mathbf{f}(v(\mathbf{x},t),\mathbf{x},t) - \widehat{\mathcal{T}}[v(\cdot,t)](\mathbf{x},t)\Big)\phi(\mathbf{x},t)d\mathbf{x}dt \ . \end{aligned}$$

Here we put $a \top b = \max\{a, b\}$ *and* $a \perp b = \min\{a, b\}$ *for* $a, b \in \mathbb{R}$.
A function $u \in W(0,T)$ *is called an* entropy solution *of* (4.13) *if for all* $\kappa \in \mathbb{R}$ *and* $\phi \in C_0^\infty(\mathbb{R}^2 \times [0,T))$ *with* $\phi > 0$ *the inequality*

$$E(u, \kappa, \phi) \geq -\int_{\mathbb{R}^2} |u_0(\mathbf{x}) - \kappa| \phi(\mathbf{x}, 0)\, d\mathbf{x} \tag{4.19}$$

holds.

The existence and uniqueness of an entropy solution for the model problem (4.13) in the case where the non–local term $\mathcal{T}$ vanishes can be shown by studying the solution of the viscous equation:

4.12 Theorem (Existence/Uniqueness of an Entropy Solution)
Let Assumption 4.3 be satisfied for $\mathcal{T} \equiv 0$. *If the data* $\mathbf{f}, q,$ *and* u_0 *are in* C^2 *then there exists a unique classical solution* u_ε *to the viscous problem*

$$\partial_t u_\varepsilon(\mathbf{x}, t) + \nabla \cdot \mathbf{f}(u_\varepsilon(\mathbf{x}, t), \mathbf{x}, t) = \widehat{\mathcal{T}}[u_\varepsilon(\cdot, t)](\mathbf{x}, t) + \varepsilon \triangle u_\varepsilon(\mathbf{x}, t)\ , \tag{4.20a}$$

$$u_\varepsilon(\mathbf{x}, 0) = u_0(\mathbf{x})\ , \tag{4.20b}$$

for $\varepsilon > 0$ *and* $\mathbf{x} \in \mathbb{R}^2$, $t \in (0,T)$. *The sequence* $\{u_\varepsilon\}_{\varepsilon > 0}$ *converges in* L^1 *to a function* $u \in W(0,T)$ *which is the unique entropy solution to the problem* (4.13).

Proof:
The proof of this theorem can be found e.g. [Daf00, Theorem 6.2.1 and Corollary 6.2.1]) The idea is to show the existence of a classical solution to (4.20) and to prove boundedness of the solutions u_ε independent of ε. This leads to the convergence of u_ε to a function u for which it is possible to show that u is a entropy solution. A contraction principle shows that there exists at most one entropy solution. □

To our knowledge the existence of an entropy solution in the case where $\mathcal{T} \not\equiv 0$ has not yet been shown in the literature. In the following we therefore assume the existence of an entropy solution in the case with a non–local operator, as well.

4.13 Assumption (Existence/Uniqueness of an Entropy Solution)
Let Assumptions 4.3 be satisfied. Then there exists an unique entropy solution $u \in W(0,T)$ *to the initial value problem* (4.13).

We conclude this section with the only existence result known to the author for the coupled system. The result is concerned with the question of the existence of classical solution for small time. It is not derived for the coupled system (1.19) but for the case where the time dependent radiation transport equation (1.15) is coupled to the MHD equations in the manner described in Section 1.3

4.14 Theorem (Short Time Existence)
We consider the Cauchy problem for the coupled system (1.19) *in three space dimensions with the instationary radiation transport equation* (1.15) *instead of* (1.19f). *Let the initial conditions* $\mathbf{U}_0 \in H^3(\mathbb{R}^3)$ *be contained in some compact subset of the state space* $\mathcal{U}$ *(especially* $\rho_0 > \bar{\rho} > 0$*) with* $\nabla \cdot \mathbf{B}_0 = 0$, *then there exists a* $T_* > 0$, *such that the Cauchy problem has a* classical solution *in* $[0, T_*)$.

Proof:
For the proof a standard iteration technique is used (cf. [RZ01]). □

4.3 Partial Regularization

In this section we study some effects of the non–local radiation source term Q_{rad} on the PDE level. We first investigate special solutions to our model problem (4.13) using analytical results when available and high resolution one–dimensional simulations in the other cases. We demonstrate to which extend radiation has a regularizing effect on the solution u; we study the effects of the radiation source term in relation to the regularizing effects of a viscous approximation. As stated in Theorem 4.12 the solution to the viscous equations are smooth even for discontinuous initial data. The radiation, on the other hand, only leads to smooth solutions for small jumps in the initial data at least for $t \to \infty$, whereas for large jumps the solution stays discontinuous for all time; discontinuities can even develop for smooth data depending on the size of the initial data and on the size of its derivative (cf. Theorem 4.9). Therefore we only have a *partial regularization* through radiation.
In this section we study solutions to the one dimensional model problem (4.13) with the balance term given by (4.12) in the special case where the operator is linear, i.e., $B(u) = u$:

$$\partial_t u(x,t) + \partial_x f(u(x,t)) = \widehat{\mathcal{T}}[u(\cdot,t)](x) \tag{4.21a}$$

with initial condition

$$u(x,0) = u_0(x) \tag{4.21b}$$

defined on the whole real axis. We use any one of the following equivalent formulations of (4.12):

$$\begin{aligned}\widehat{\mathcal{T}}[w](x) &= \chi \int_0^1 \left(\frac{\chi}{\mu} \int_{-\infty}^{\infty} w(x-y) e^{-\frac{\chi}{\mu}|y|} dy - 2w(x) \right) d\mu \\ &= \chi \int_0^1 \left(\frac{\chi}{\mu} \int_{-\infty}^{\infty} (w(y) - w(x)) e^{-\frac{\chi}{\mu}|x-y|} dy \right) d\mu.\end{aligned}$$

In the following two sections we construct special solutions to (4.21) and study the effects of using a quadrature for the computation of the integral with respect to μ (cf. Section 3.1). We compare the effects of the radiation operator $\widehat{\mathcal{T}}$ on the scalar quantity u with the solution to the viscous regularization

$$\partial_t v(x,t) + \partial_x f(v(x,t)) = \varepsilon \partial_{xx} v(x,t) \ , \tag{4.22a}$$

$$v(x,0) = u_0(x) \ . \tag{4.22b}$$

4.3.1 Linear Advection

We begin our study by setting the flux function to $f(u) = au$ for $a \in \mathbb{R}$ constant. By a simple transformation we can assume without loss of generality that $a = 0$ both in (4.21a) and (4.22a): let $u(x,t)$ be a solution to (4.21a) (or (4.22a)) with $f \equiv 0$ then $w(x,t) = u(x - at, t)$ is a solution to (4.21a) (or (4.22a)) with $f(u) = au$, respectively.

In the case $f \equiv 0$ the viscous equation (4.22) reduces to the linear heat equation for which the formal solution can be easily obtained. For linear equations, plane wave solutions are especially simple to construct.

4.15 Theorem (Linear Model Problem)
Let the initial data be of the form $u_0(x) = \sum_{l=1}^{L} e^{i\lambda_l x}$ *with* $L \in \mathbb{N}$ *and* $\lambda_1, \ldots, \lambda_L \in \mathbb{R}$. *Define for* $\lambda \in \mathbb{R}$

$$D_{\text{visc}}(\lambda) := \lambda^2 \quad \textit{and} \quad D(\lambda) := 2\frac{\chi\lambda - \chi^2 \arctan\left(\frac{\lambda}{\chi}\right)}{\lambda} \ .$$

Then the solution to problem (4.22) *with initial data* u_0 *and* $f \equiv 0$ *is given by*

$$v(x,t) = \sum_{l=1}^{L} e^{-t\varepsilon D_{\text{visc}}(\lambda_l)} e^{i\lambda_l x}$$

and the solution to problem (4.21) *with initial data* u_0 *and* $f \equiv 0$ *is given by*

$$u(x,t) = \sum_{l=1}^{L} e^{-tD(\lambda_l)} e^{i\lambda_l x} \ .$$

Proof:
Since both equation (4.21) and equation (4.22) are linear it is sufficient to study initial data of the form $u_0(x) = e^{i\lambda x}$. Since the result for (4.22) is standard and the proof is very similar to the proof for our model problem, we restrict ourselves to constructing the solution for (4.21). To solve (4.21) with $f \equiv 0$ and $u_0 = e^{i\lambda x}$ we assume that the solution has the form $u(x,t) = w(t)e^{i\lambda x}$. Substituting this expression into equation (4.21a) we find:

$$w'(t)e^{i\lambda x} = \chi \int_0^1 \left(\frac{\chi}{\mu} \int_{-\infty}^{\infty} w(t) e^{i\lambda(x-y)} e^{-\frac{\chi}{\mu}|y|} dy - 2w(t)e^{i\lambda x} \right) d\mu.$$

Dividing by $e^{i\lambda x}$ leads to a linear ordinary differential equation for $w(t)$:

$$\begin{aligned} w'(t) &= w(t)\chi \int_0^1 \left(\frac{\chi}{\mu} \int_{-\infty}^{\infty} e^{-i\lambda y - \frac{\chi}{\mu}|y|} dy - 2 \right) d\mu = w(t)\chi \int_0^1 \left(\frac{\chi}{\mu} \frac{2\frac{\chi}{\mu}}{\frac{\chi^2}{\mu^2} + \lambda^2} - 2 \right) d\mu \\ &= w(t)\chi \int_0^1 \frac{-2\lambda^2\mu^2}{\chi^2 + \lambda^2\mu^2} d\mu = w(t)\frac{-2\chi}{\lambda} \int_0^{\lambda} \frac{s^2}{\chi^2 + s^2} ds \ . \end{aligned}$$

Now since $\frac{d}{ds}\frac{s^2}{a^2+s^2} = s - a\arctan(\frac{s}{a})$ we can evaluate the remaining integral

$$w'(t) = w(t)\frac{-2\chi}{\lambda}\left[s - \chi \arctan\left(\frac{s}{\chi}\right) \right]_{y=0}^{\lambda} = w(t)(-2)\frac{\chi\lambda - \chi^2 \arctan\left(\frac{\lambda}{\chi}\right)}{\lambda} \ .$$

We thus arrive at the following expression for the solution u:

$$u(x,t) = e^{-tD(\lambda)}e^{i\lambda x}.$$

Direct computation shows that this is a solution to (4.21). □

4.16 Remark: *We have $D(\lambda) = -D(\lambda)$ and since $s > \arctan(s)$ for $s > 0$ we also have $D(\lambda) > 0$ for $\lambda > 0$ and $D(0) = 0$. Furthermore we have $D(\lambda) \to 2\chi$ for $\lambda \to \infty$. Thus both equation* (4.21) *and equation* (4.22) *lead to a damping of all modes with $\lambda > 0$. In the case of* (4.22)*, however, the rate of the damping is unbounded for $\lambda \to \infty$; this is not the case for equation* (4.21)*.*

In Figure 4.1(a) we plot the rate of the damping $D(\lambda)$ and λ^2, respectively. In Figure 4.2 we show a time sequence of the solutions $u(x,t), v(x,t)$ for the initial condition

$$u_0(x) = e^{i0.5x} + e^{i5x}. \tag{4.23}$$

The different damping rate especially for the fast mode e^{i5x} is clearly visible.
Our approximation of the radiation source term involves a quadrature for the integral over the propagation directions $\boldsymbol{\mu}$ (cf. Section 3.1). The influence of this quadrature for this simple setting is easily studied.

4.17 Theorem
For $M \in \mathbb{N}$ fixed define the operator

$$\widehat{\mathcal{T}}_M[w](x) := \chi\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\chi}{\mu_m}\int_{-\infty}^{\infty} w(x-y)e^{-\frac{\chi}{\mu_m}|y|}dy - 2w(x)\right)$$

with $\mu_m = \frac{m-0.5}{M}$ and define the function

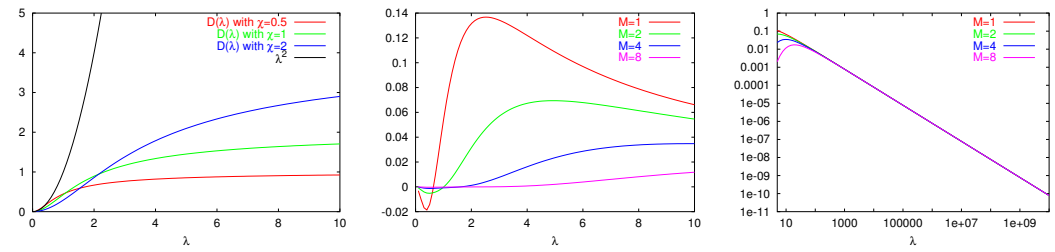
$$D_M(\lambda) := \chi\frac{1}{M}\sum_{m=1}^{M}\left(\frac{-2\mu_m^2\lambda^2}{\chi^2+\mu_m^2\lambda^2}\right)$$

for $\lambda \in \mathbb{R}$. Consider the problem (4.21) *with $f \equiv 0$, initial data u_0 given as in Theorem 4.15, and with the right hand side defined by $\widehat{\mathcal{T}}_M[w]$. Then the solution u is given by*

$$u_M(x,t) = \sum_{l=1}^{L} e^{-tD_M(\lambda_l)}e^{i\lambda_l x}\ .$$

Proof:
The proof is a direct consequence of the proof of Theorem 4.15. First let us recall the ODE for the function w:

$$w'(t) = w(t)\chi\int_0^1\left(\frac{-2\mu^2\lambda^2}{\chi^2+\mu^2\lambda^2}\right)d\mu.$$



(a) damping rates for viscous equation and radiation operator for different values of χ

(b) approximation error of damping rate due to quadrature for small λ and $\chi = \frac{1}{2}$

(c) approximation error of damping rate due to quadrature for large λ and $\chi = \frac{1}{2}$

Figure 4.1: Damping rate for viscous and radiation operator, and approximation error of damping rate due to use of quadrature rule in radiation operator.

In the definition of $\widehat{\mathcal{T}}_M[w]$ we approximate the integral in μ using the simple midpoint rule

$$\int_0^1 f(\mu)d\mu \approx \frac{1}{M}\sum_{j=1}^{M} f(\mu_j) \tag{4.24}$$

This leads to an approximation w_M for w which satisfies

$$w'_M(t) = w_M(t)\chi\frac{1}{M}\sum_{m=1}^{M}\left(\frac{-2\mu_m^2\lambda^2}{\chi^2+\mu_m^2\lambda^2}\right) .$$

Thus we arrive at the damping rate $D_M(\lambda)$ for $M \in \mathbb{N}$ as stated in the theorem. □

The errors due to the approximation of the integral in μ for a few values of M are plotted in Figure 4.1(b) and Figure 4.1(c). The error decreases significantly for large values of λ (note that the y–axis is logarithmic in Figure 4.1(c)). The approximation errors between the solution u of (4.21) for the initial conditions (4.23) and the solutions u_M using the quadrature rule with $M = 4$ and $M = 8$ are shown in Figure 4.3. It is clearly visible that even for small values of M the exact solution is reproduced with a high accuracy.

The results in this section show that the radiation source term leads to a damping of all frequency nodes, and in this sense we can speak of a regularization of the initial data. The rate of the damping especially for the high frequency modes is quite different from the rate produced by dissipation effects. Note that in the homogenous case equation (4.21a) reduces to $\partial_t u = 0$ so that the initial condition is maintained for all time, i.e., no damping occurs. In the next section we present some results that demonstrate this difference in the case of a nonlinear advection term.

4.3.2 Burgers' Equation

We now study the case where the flux function is non–linear. The analytical results presented here are taken from [KN99b, KN99a], where the authors study a special form of (4.21). For the convection nonlinearity they study the case of Burgers' equation, i.e.,

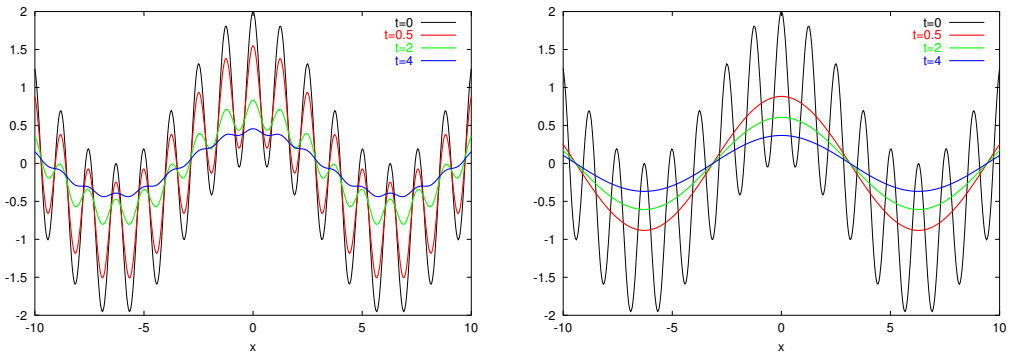


Figure 4.2: Time sequence of solution to linear problem with radiation operator (left) and viscous operator (right).

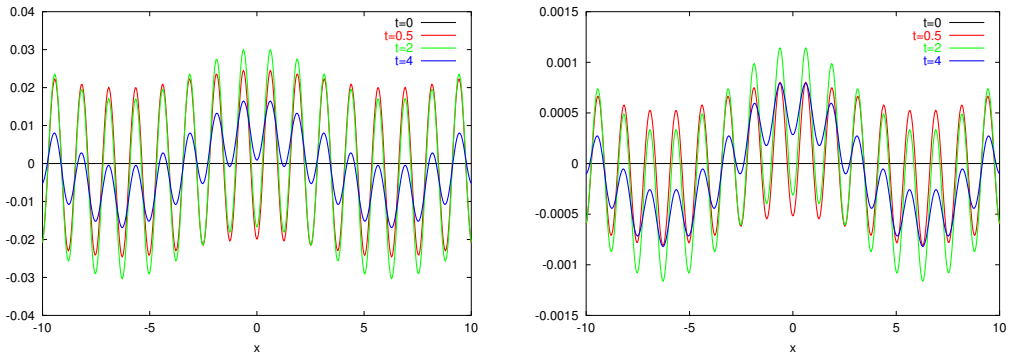


Figure 4.3: Time sequence of deviation between exact solution to linear problem (cf. Figure 4.2) and solution obtained by approximation of radiation operator by quadrature rule with $M = 4$ (left) and $M = 8$ (right).

$f(u) = \frac{1}{2}u^2$, to which they couple an elliptic equation. This leads to the following set of PDEs:

$$\partial_t u(x,t) + \partial_x \left(\frac{1}{2} u^2(x,t) \right) + \partial_x q(x,t) = 0 \ , \tag{4.25a}$$

$$-\partial_{xx} q(x,t) + q(x,t) + \partial_x u(x,t) = 0 \tag{4.25b}$$

with $(x,t) \in \mathbb{R} \times \mathbb{R}^+$. This is a model for a radiating gas in the diffusion limit using a special scaling of the radiation transport equation. We can derive this coupled system as an approximation of (4.21) by including the radiation intensity only for the two propagation directions $\{-\frac{1}{2}, \frac{1}{2}\}$. This is equivalent to using the midpoint role for evaluating the integral in (4.12) over $[0,1]$ in μ. For the derivation we assume that all quantities are smooth. We start with the following system:

$$\partial_t u(x,t) + \partial_x \left(\frac{1}{2} u^2(x,t) \right) = \chi \left(I_+(x,t) - u(x,t) \right) + \chi \left(I_-(x,t) - u(x,t) \right) \ , \tag{4.26a}$$

$$\frac{1}{2}\partial_x I_+(x,t) + \chi I_+(x,t) = \chi u(x,t) \ , \tag{4.26b}$$

$$-\frac{1}{2}\partial_x I_-(x,t) + \chi I_-(x,t) = \chi u(x,t) \ . \tag{4.26c}$$

Taking the derivative with respect to x and combining the equations for I_+, I_- one easily derives the equation

$$\partial_{xx}(I_+ - I_-) = 4\chi^2(I_+ - I_-) + 4\chi\partial_x u \ .$$

Defining $q := \frac{1}{2}(I_+ - I_-)$ we arrive at equation (4.25b) by setting $\chi = \frac{1}{2}$. Inserting the equations for I_+, I_- into the right hand side of the equation for u we find:

$$\partial_t u(x,t) + \partial_x\left(\frac{1}{2}u^2(x,t)\right) = -\frac{1}{2}\partial_x(I_+ - I_-)$$

which reduces to (4.25a) due to our definition of q. Introducing the operator $\mathcal{K}$ as the inverse of the operator $-\frac{d^2}{dx^2} + 1$, we can set $q = -\mathcal{K}\partial_x u$. It is easy to see that $\mathcal{K}[f](x) = \frac{1}{2}\int_{\mathbb{R}} e^{-|x-y|} f(y)dy$. It follows that $\partial_x q = u - \mathcal{K}u$. Therefore we can rewrite the system (4.25) as

$$\partial_t u(x,t) + \partial_x\left(\frac{1}{2}u^2(x,t)\right) + u - \mathcal{K}u = 0 \ , \tag{4.27a}$$

$$q = -\mathcal{K}\partial_x u \ . \tag{4.27b}$$

This is in accordance with our definition of q as $\frac{1}{2}(I_+ - I_-)$ (cf. (4.11) with $\mu = \pm\frac{1}{2}$ and $\chi = \frac{1}{2}$). Combining these equations, and, using integration by parts, we derive the integral representation of q.

In [KN99b] the authors study the existence and asymptotic stability of *admissible* traveling waves for the system (4.25). These are entropy solutions of the form $(u,q)(x,t) = (U,Q)(\xi)$ with $\xi = x - st$ and $\lim_{\xi\to\pm\infty} U(\xi) = u_\pm$. The constant value s is called the speed of the traveling wave, and the constants $u_\pm$ are called the asymptotic states of the wave. As admissibility conditions the authors use the following form of the Kruzhkov entropy condition:

$$\begin{aligned}\int_{\mathbb{R}} -s|U-\kappa|\phi' + \operatorname{sign}(U-\kappa)\frac{1}{2}(U^2-\kappa^2)\phi' \\ + \operatorname{sign}(U-\kappa)(U-\mathcal{K}U)\phi d\xi \geq 0 \ ,\end{aligned} \tag{4.28a}$$

$$\int_{\mathbb{R}} Q\psi d\xi = \int_{\mathbb{R}} \mathcal{K}U\psi d\xi \tag{4.28b}$$

for arbitrary $\phi \in C_0^\infty(\mathbb{R})$ with $\phi \geq 0$, $\kappa \in \mathbb{R}$, and where ψ is a fast decreasing function. We first study the case of no radiation ($q \equiv 0$) and the case of the viscous equation:

$$\partial_t w(x,t) + \partial_x\left(\frac{1}{2}w^2(x,t)\right) = \varepsilon\partial_{xx} w(x,t) \ .$$

For the homogeneous Burgers' equation there exists a unique admissible traveling wave if and only if $u_- > u_+$ (cf. [Krö97, Remark 2.1.21]). In this case we have the following discontinuous solution

$$u(x,t) = \begin{cases} u_- & x < st \ , \\ u_+ & x > st \end{cases}$$

with s defined by the Ranking–Hugoniot relation

$$s = \frac{1}{2}(u_- + u_+) \ .$$

In the case of the viscous equation there exists a smooth continuous wave connecting u_- and u_+ again under the condition $u_- > u_+$ with the same speed as in the case without dissipation [Daf00, Theorem 8.6.1]. The next theorem gives an overview of the structure of traveling waves solutions for (4.25).

4.18 Theorem (Traveling Wave Solutions for Radiation Operator)

(i): Let (U,Q) be an admissible traveling wave, i.e., $u(x,t) := U(x-st)$ and $q(x,t) := Q(x-st)$ are admissible solutions of (4.25) and $\lim_{\xi\to\pm} U(\xi) = u_\pm$. If (U,Q) are piecewise smooth functions with only first kind discontinuities, then the following relations must hold

$$u_- > u_+ \ , \qquad s = \frac{1}{2}(u_+ + u_-) \ .$$

Furthermore we have that $\lim_{\xi\to\pm\infty} Q(\xi) = 0$.

(ii): Assume that $u_- > u_+$ and that $s = \frac{1}{2}(u_+ - u_-)$ then there exists an admissible traveling wave (U,Q) that is unique up to a shift in the class of piecewise smooth functions with only the first kind discontinuities. Two cases can be distinguished:

(a) If $|u_+ - u_-| > \sqrt{2}$ then U is continuous except for one point, while Q is Lipschitz continuous. If we denote with U_l the limit of U from the left at the discontinuity and with U_r the limit from the right (U_l', U_r', Q_l', Q_r' are defined accordingly) the following algebraic relations hold:

$$U_l - Q_l' = U_r - Q_r' \ , \qquad U_l' = -1 \ , \qquad U_r' = -1 \ .$$

(b) If $|u_+ - u_-| \le \frac{2\sqrt{2n}}{n+1}$ for some $n > 0$ then we have

$$U \in C^n(\mathbb{R}) \cap H^{n,\infty}(\mathbb{R}) \quad \textit{and}$$
$$Q \in C^{n+1}(\mathbb{R}) \cap H^{n+1,\infty}(\mathbb{R}) \ .$$

This theorem shows that an admissible traveling wave exists under the same conditions as in the case of the homogeneous Burgers' equation, but that in contrast to the viscous case not every traveling wave is smooth. The regularity greatly depends on the size of $|u_+ - u_-|$, i.e., the size of the jump. The proof of Theorem 4.18 can be found in [KN99b].

We now demonstrate the implications of the theorems given above. Since the computation of stationary waves is much simpler than that of moving waves, we use asymptotic left and right hand states that lead to a vanishing speed of the traveling wave, i.e. $u_- = -u_+ =: u_*$. We use a numerical scheme similar to the base scheme described in Chapter 3: for the evolution of the scalar solution u we use a first order finite–volume scheme with the Enquist–Osher flux [Krö97, Example 2.2.7], and to approximate the radiation operator (4.12) we use a quadrature rule as in (4.24). Finally the ODE for the radiation intensity in (4.26) is solved using the backward Euler method.

4.19 Remark: *The traveling wave for the homogeneous Burgers' equation is u_* for $x < 0$ and $-u_*$ for $x > 0$. The solution of the viscous equation* (4.22a) *is continuous for all $t > 0$. The stationary solution with the correct asymptotic states for* (4.22a) *with $f(u) = \frac{1}{2}u^2$ is given by $v_*(x) = u_* \tanh(-\frac{u_*}{\varepsilon}x)$.*

We study the time evolution of problem (4.21) with the following smooth initial value:

$$u_0(x) = u_* \tanh(-\delta x) \ , \tag{4.29}$$

where $\delta > 0$ is a constant that we use to change the amplitude of the derivative of u_0:

$$u_0'(x) = -\frac{u_*\delta}{\cosh^2(-\delta x)} \ .$$

Therefore $\min_{x\in\mathbb{R}} u_0'(x) = -u_*\delta$. Note that u_0 has the correct asymptotic states for all δ. We use a one point quadrature so that the radiation operator is identical to the one studied in [KN99b, KN99a]. Therefore we can apply Theorem 4.9: the threshold value for which the solution develops a discontinuity is given by $\omega_* = -1$. Therefore, if $-u_*\delta < \omega_*$ the solution must become discontinuous for finite time. Note that for $-u_*\delta > \omega_*$ it is not clear whether the solution will become discontinuous or not. Due to Theorem 4.18 we know that for $u_* < \frac{\sqrt{2}}{2}$ the traveling wave is continuous and that it is discontinuous for $u_* > \frac{\sqrt{2}}{2}$. Whether the discontinuity develops in finite time or only in the limit $t \to \infty$ is not clear. We choose two different values for δ in the following: $\delta_1 := -(\omega_* + 0.4)/u_* = 3$ and $\delta_2 := -(\omega_* - 0.4)/u_* = 7$. Obviously $-\delta_1 u_* > \omega_*$ and $-\delta_2 u_* < \omega_*$.

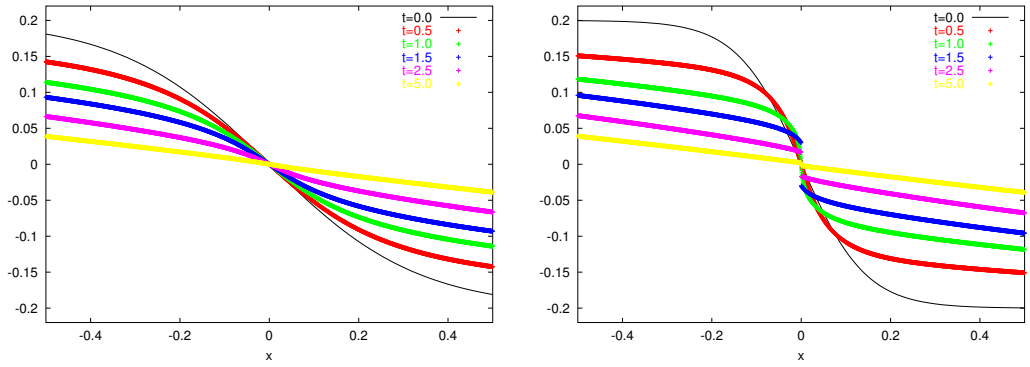
Small Jump ($\mathbf{u_* = 0.2}$): The solution for different values of t are shown in Figure 4.4(a). Note that with δ_2 a discontinuity has developed for $t = 1.5$ (the time given in Theorem 4.9 is $t_0 \approx 1.25276$). From Figure 4.4(a) it is not obvious at which time the discontinuity developed and if the solution is still discontinuous, for example, at time $t = 5$. To give a clearer indication of what is happening, we have also computed the central difference quotient at $x = 0$ for a series of different grid resolutions. If the solution is discontinuous, we expect that this value diverges if the grid size goes to zero. On the other hand for continuous solutions this value should converge. The table for δ_2 in Figure 4.4(b) indicates that the solution is still continuous at $t = 1$ but is discontinuous for all samples with $t > 1$. On the other hand the solution using δ_1 remains continuous for all time. This is in accordance with the second part of Theorem 4.9 since $\delta_0 = 0.4 < \frac{1}{2}$ and $\omega_{**} \approx -0.7236 < -u_*\delta_1$.

Note that since the jump between the left and right hand state is 0.4 and is therefore smaller than $\frac{2\sqrt{2n}}{n+1}$ for $n = 47$, the resulting traveling wave with asymptotic states $u_*, -u_*$ must be at least in $C^{42}(\mathbb{R})$ due to Theorem 4.18. Thus, if the solution approaches steady state for either δ_1 or δ_2 and $t \to \infty$, then this must be a smooth function. In Figure 4.4(c) we have also plotted the stationary solution as far as our algorithm is able to compute it. We computed the steady state for initial data given by (4.29) using δ_1 and δ_2 and also using the discontinuous Riemann initial data given by u_* for $x < 0$ and $-u_*$ for $x > 0$. We obtained the identical stationary solution in all three cases.

Large Jump ($\mathbf{u_* = 1}$): In this case the jump $2u_*$ is well over $\sqrt{2}$ so that the traveling wave is discontinuous according to Theorem 4.18. The same plots as for $u_* = 0.2$ are given in Figure 4.5 this time with $u_* = 1$. Again we show a time sequence for δ chosen in such a way that $\min_{x\in\mathbb{R}} u_0'(x)$ has the same value as in the case of the small jump discussed above. It is clear from Figure 4.5(a) that, although the absolute value of the derivative of the initial data is pointwise small, the solution develops a discontinuity at $x = 0$ for finite time for both δ_1 and δ_2 because the influence of the radiation operator is non–local. This behavior is a clearly different from the homogenous Burgers' equation, where the solution always develops a discontinuity as long as u_0' is negative somewhere, but its development does not depend on the amplitude of the initial data.

Quadrature in μ: We now study the solution to the Riemann problem for (4.21) with left and right hand state u_* and $-u_*$, respectively, using $u_* = 1$. But this time we use a higher order quadrature rule for the approximation of the integral in μ (cf. (4.24)). In the previous analysis we used a simple midpoint rule, i.e. $M = 1$. In this case our model problem can be reformulated as a hyperbolic balance law coupled with a single elliptic equation. We can always reformulate the model problem as a scalar hyperbolic balance law to which M elliptic equations are coupled for any M. The proof of Theorem 4.18, however, strongly relies on the fact that we only have a two by two system in (u, q). For $M > 1$ the proof cannot be repeated in this form. Thus we have to rely on numerical experiments. In Figure 4.6(a) we plot the stationary traveling wave for $M = 1$ and $M = 2$. Already for moderate values of M the solution only changes slightly when M is increased. This is shown in Figure 4.6(b), where the difference between solutions with two different values of M are plotted. One can see that the difference between $M = 8$ and $M = 16$ is already quite small. As in the linear case this shows that at least for the setting studied here a good approximation of the angular integral in the radiation operator Q_{rad} can be achieved with small values of M.

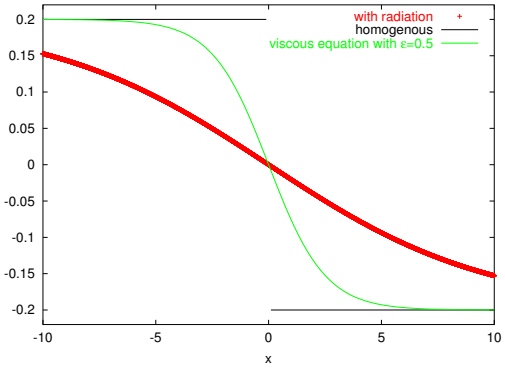
Non–linear operator ($\mathbf{B(u) = u^4}$): We conclude this section with a brief look at the traveling wave solution for a non–linear radiation operator. As is discussed in the next section it is reasonable to use $B(u) = \text{sgn}(u)u^4$ (note that this corresponds to (1.18) for $u \geq 0$ and that $u < 0$ is not physically meaningful since u corresponds to the temperature of the fluid). In Figure 4.7(a) we plot the traveling waves in the linear case and the non–linear case. The qualitative structure of the solutions are similar. The influence of the quadrature rule on the solution is shown in Figure 4.7(b).



(a) time evolution with δ_1 (left) and δ_2 (right)

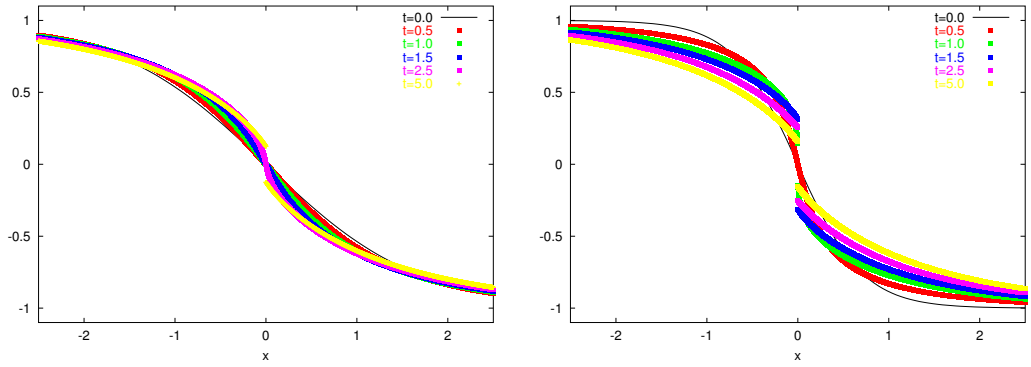
	with δ_1					with δ_2				
h	$t=\frac{1}{2}$	$t=1$	$t=\frac{3}{2}$	$t=\frac{5}{2}$	t=5	$t=\frac{1}{2}$	$t=1$	$t=\frac{3}{2}$	$t=\frac{5}{2}$	t=5
8e-3	-0.70	-0.73	-0.70	-0.52	-0.16	-3.48	-7.02	-7.75	-4.29	-0.56
4e-3	-0.70	-0.73	-0.70	-0.53	-0.16	-3.66	-10.59	-15.06	-8.51	-1.04
2e-3	-0.70	-0.73	-0.70	-0.53	-0.16	-3.71	-14.94	-29.86	-17.00	-1.99
1e-3	-0.70	-0.73	-0.70	-0.53	-0.16	-3.73	-19.24	-59.75	-33.98	-3.90
5e-4	-0.70	-0.73	-0.70	-0.53	-0.16	-3.73	-22.36	-119.71	-67.94	-7.71
2.5e-4	-0.70	-0.73	-0.70	-0.53	-0.16	-3.73	-23.91	-239.72	-135.85	-15.33

(b) finite difference quotient at $x = 0$



(c) stationary state

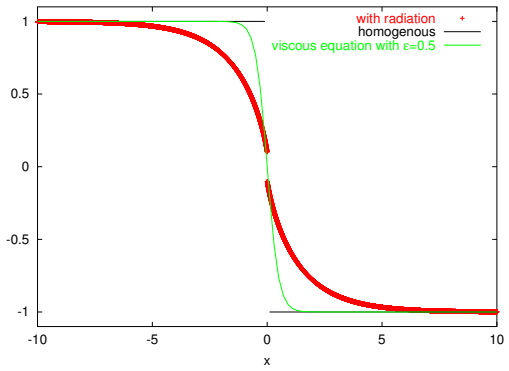
Figure 4.4: Traveling wave solution to Burgers' equation with asymptotic states $u_\pm = \pm 0.2$. The top row shows the time evolution of the solution for the initial conditions (4.29) (left: $\delta = \delta_1$, right: $\delta = \delta_2$). The bottom plot shows the quasi–stationary solution for Burgers' equation with radiation operator, with second order term, and in the homogenous case. The tables show the value of the central difference quotient at $x = 0$ at different times and for different values of h. The difference quotient indicates at which time the solution is discontinuous.



(a) time evolution with δ_1 (left) and δ_2 (right)

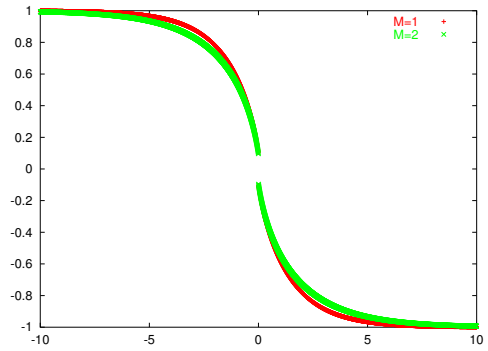
	with δ_1					with δ_2				
h	$t=\frac{1}{2}$	$t=1$	$t=\frac{3}{2}$	$t=\frac{5}{2}$	t=5	$t=\frac{1}{2}$	$t=1$	$t=\frac{3}{2}$	$t=\frac{5}{2}$	t=5
1.6e2	-0.91	-1.33	-2.02	-6.65	-16.74	-4.75	-26.26	-39	-32	-20
8e-3	-0.91	-1.33	-2.03	-8.00	-32.33	-4.81	-46.54	-78	-64	-40
4e-3	-0.91	-1.33	-2.03	-8.85	-63.45	-4.82	-84.26	-156	-127	-79
2e-3	-0.91	-1.33	-2.03	-9.21	-125.67	-4.83	-156.16	-312	-254	-157
1e-3	-0.91	-1.33	-2.03	-9.31	-250.11	-4.83	-297.23	-625	-508	-313
5e-4	-0.91	-1.33	-2.03	-9.34	-498.98	-4.83	-578.48	-1250	-1016	-625

(b) finite difference quotient at $x = 0$

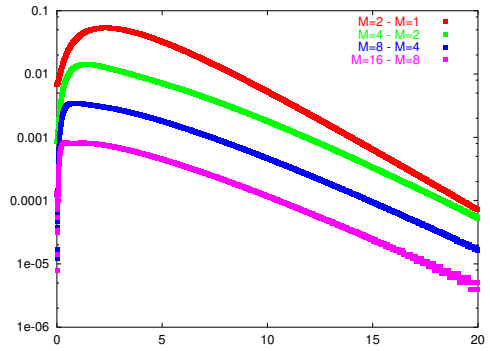


(c) stationary state

Figure 4.5: Traveling wave solution to Burgers' equation with asymptotic states $u_\pm = \pm 1$. The top row shows the time evolution of the solution for the initial conditions (4.29) (left: $\delta = \delta_1$, right: $\delta = \delta_2$). The bottom plot shows the quasi–stationary solution for Burgers' equation with radiation operator, with second order term, and in the homogenous case. The tables show the value of the central difference quotient at $x = 0$ at different times and for different values of h. The difference quotient indicates at which time the solution is discontinuous.

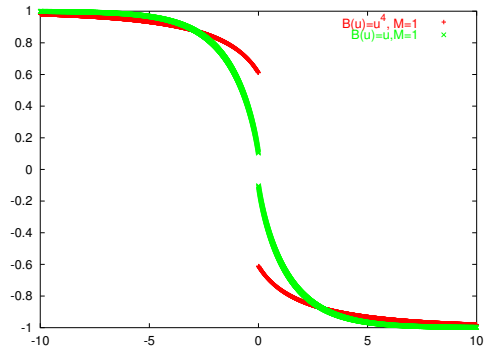


(a) different values for M

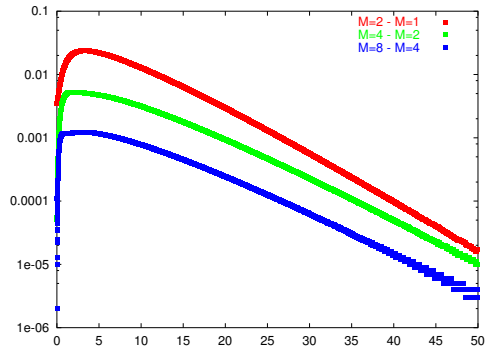


(b) difference between two solution with different value for M

Figure 4.6: Traveling wave solution to Burgers' equation with asymptotic states $u_{\pm} = \pm 1$. Using a quadrature rule, the integral defining the radiation source term is approximated with M integration points. Already with a moderate number of quadrature points the solution shows little variation.



(a) traveling waves for linear and non–linear radiation operator



(b) difference between two solution with different value for M

Figure 4.7: Traveling wave solution to Burgers' equation with asymptotic states $u_{\pm} = \pm 1$. The physically correct non–linear radiation source term with $B(u) = \sigma u^4$ leads to a slightly different traveling wave, but with the same structure as in the linear case. Again only a moderate number of quadrature points are required to obtain a good approximation.

4.4 Convergence Result: Local Source Term

In this section we assume that the balance term $\widehat{\mathcal{T}}$ is a local source term, i.e. $\widehat{\mathcal{T}}[w](\mathbf{x},t) = q(w(\mathbf{x}),\mathbf{x},t)$. This operator is discretized on T_i ($i \in \mathcal{I}_h$) by

$$\widehat{\mathcal{T}}_h[w_h](\boldsymbol{\omega}_i, t^n) = \frac{1}{\Delta t|T_i|} \int_{t^n}^{t^{n+1}} \int_{T_i} q(w_i,\mathbf{x},t) d\mathbf{x}dt \tag{4.30}$$

for $w_h \in V_h$. The following theorem establishes the rate of convergence for a finite–volume scheme under the assumptions given in Section 4.1. The proof is omitted and can be found in [CHC01].

4.20 Theorem (Convergence with Local Source Term)
Let $u \in W(0,T)$ be the entropy solution to (4.13) *and let $\{\mathcal{T}_h\}_h$ be a family of grids and Δt a time step satisfying Assumption 4.4. Consider a family of discrete approximations $\{u_h\}_h$ given by the finite–volume scheme from Definition 4.7 with $\widehat{\mathcal{T}}_h$ as in* (4.30). *Assume that the time step Δt satisfies the CFL condition*

$$\frac{\Delta t}{h} \leq \frac{(1-\xi)c_G^2}{2L_g(\|u_0\|_\infty)}$$

for some $\xi \in (0,1)$. If Assumption 4.3 is satisfied then there exists a constant K depending on $\mathbf{f}, u_0, q, c_G, \xi$ such that

$$\int_0^T \int_{\mathbb{R}^d} |u_h(\mathbf{x},t) - u(\mathbf{x},t)| d\mathbf{x}dt \leq K h^{\frac{1}{4}} \ . \tag{4.31}$$

4.5 Convergence Result: Non–Local Source Term

In this section we study the convergence properties of the finite–volume scheme for our model problem (4.13) in two space dimensions including a non–local balance term. For simplicity we assume in the following that the local part q of the source term and that the flux function f do not depend explicitly on $\mathbf{x}$ and t. Consider $\mathbf{f}, q, u_0$, and $\widehat{\mathcal{T}}$ with $\mathcal{T} \not\equiv 0$ satisfying Assumption 4.3. We only summarize the basic steps of the convergence proof, which is published in [DR02b]. Furthermore we discuss a fully explicit finite–volume scheme with a discretization of the radiation operator that can be directly implemented on a computer. To this discretization we apply our general convergence result at the end of this section. The main difficulties are the weak singularity of the integral kernel (cf. (4.8)) and the small amount of regularity of the entropy solution u of the model problem (4.13) (cf. Definition 4.11) . We have demonstrated in the previous section that we cannot expect a regularization effect of the radiation operator similar to the regularization of a second order term. Therefore we must study the case where the entropy solutions u is in the space $W(0,T)$ defined in (4.1).
In [Ded98] we discussed the convergence of a finite–volume approximation in one space dimension without deriving a computable approximation for the integral operator. The main result from [Ded98] was the existence of a weak solution, which followed from the

convergence result using standard arguments. For the convergence result presented here we have to assume that an entropy solution for our model problem exists (cf. Assumption 4.13). This is typical for convergence results on unstructured grids in higher space dimensions. Nevertheless the results presented here cover a far larger range of problems including, for example, the case of non–linear operators, which was not included in [Ded98].
The main problem in applying the standard theory for finite–volume schemes to this setting lies in the control of the domain of dependence, which is not finite due to the integral operator $\mathcal{T}$ in (4.13a). Therefore, we introduce a truncated problem where $\mathcal{T}$ is only evaluated on some compact set.

4.21 Definition (Truncated Model Problem for RMHD)
For some compact subset Ω of $\mathbb{R}^2$ with the characteristic function χ_Ω, consider the truncated Cauchy problem for $u_\Omega : \mathbb{R}^2 \times [0,T) \to \mathbb{R}$

$$\partial_t u_\Omega(\mathbf{x},t) + \nabla \cdot \mathbf{f}(u_\Omega(\mathbf{x},t)) = \widehat{\mathcal{T}}_\Omega[u_\Omega(\cdot,t)](\mathbf{x}) \ , \ (\mathbf{x},t) \in \mathbb{R}^2 \times (0,T) \ , \tag{4.32a}$$

$$u(\mathbf{x},0) = u_0(\mathbf{x}) \ , \qquad \mathbf{x} \in \mathbb{R}^2 \ . \tag{4.32b}$$

Here the truncated operator $\widehat{\mathcal{T}}_\Omega$ is defined by

$$\widehat{\mathcal{T}}_\Omega[w](\mathbf{x}) = \chi_\Omega(\mathbf{x})\widehat{\mathcal{T}}[w](\mathbf{x}) \tag{4.32c}$$

4.22 Remark: *Note that the operator $\widehat{\mathcal{T}}_\Omega$ satisfies Assumption 4.3 with $C_{\widehat{\mathcal{T}}_\Omega} = C_{\widehat{\mathcal{T}}}$. Since the truncated problem from Definittion 4.21 is only a special case of the original model problem 4.1, the Assumption 4.13 on the existence and uniqueness of entropy solutions also holds true for the truncated problem. Note that any distributional solution of* (4.32) *has* compact *support if u_0 is compactly supported as is the case due to Assumption 4.3.*

We approximate the truncated problem by a finite–volume scheme for which we have to define an admissible discrete operator for $\widehat{\mathcal{T}}_\Omega$ (cf. Definition 4.7).

4.23 Definition (Admissible Discrete Radiation Transport Operator)
For $h \in (0, h_0]$, an operator $\widehat{\mathcal{T}}_{\Omega,h} : V_h \to V_h$ is called an admissible discrete operator for $\widehat{\mathcal{T}}_\Omega$ on the grid $\mathfrak{T}_h$ if there exists a continuous function $\gamma_{app} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that for all $w_h, \tilde{w}_h \in V_h \cap L^\infty(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$ with $\|w_h\|_\infty, \|\tilde{w}_h\|_\infty \leq M$ for $M > 0$, we have

(i) $\|\widehat{\mathcal{T}}_{\Omega,h}[w_h]\|_\infty \leq C_{\widehat{\mathcal{T}}}(M)$,

(ii) $\|\widehat{\mathcal{T}}_{\Omega,h}[w_h]\|_1 \leq C_{\widehat{\mathcal{T}}}(M)\|w_h\|_1$,

(iii) $\operatorname{supp}(\widehat{\mathcal{T}}_{\Omega,h}[w_h]) \subset \{\mathbf{x} \in \mathbb{R}^2 \,|\, \operatorname{dist}(\mathbf{x},\Omega) \leq h\}$,

(iv) $\sum_{j\in\mathcal{I}_h} |T_j| \left| \widehat{\mathcal{T}}_{\Omega,h}[w_h](\boldsymbol{\omega}_j) - \widehat{\mathcal{T}}_{\Omega,h}[\tilde{w}_h](\boldsymbol{\omega}_j) \right| \leq C_{\widehat{\mathcal{T}}}(M) \sum_{j\in\mathcal{I}_h} |T_j| \left| (w_h - \tilde{w}_h)(\boldsymbol{\omega}_j) \right|$,

(v) $\int_{T_j} \left| \widehat{\mathcal{T}}_\Omega[w_h](\mathbf{x}) - \widehat{\mathcal{T}}_{\Omega,h}[w_h](\mathbf{x}) \right| d\mathbf{x} \leq |T_j| \gamma_{app}(h) \quad (j \in \mathcal{I}_h)$.

The continuous function γ_{app} can depend on M (which is omitted in the notation). It satisfies for $\Delta > 0$

$$\gamma_{app}(\Delta) > 0 \ , \quad \gamma_{app}(0) = 0 \ . \tag{4.33}$$

Furthermore our convergence result depends on the *(spatial) modulus of continuity* of $\widehat{\mathcal{T}}[u(\cdot,t)]$ for $t \in [0,T]$, i.e., the operator $\widehat{\mathcal{T}}$ applied to the entropy solution u.

4.24 Definition (Modulus of Continuity)
For a function $w \in L^\infty(\mathbb{R}^2)$, $\Delta > 0$, and a subset S of $\mathbb{R}^2$, the modulus of continuity $\varepsilon(\Delta, S, w)$ is defined by

$$\varepsilon(\Delta, S, w) := \sup_{|\Delta\mathbf{x}| \le \Delta} \left\{ \int_S |w(\mathbf{x}+\Delta\mathbf{x}) - w(\mathbf{x})| \, d\mathbf{x} \right\}. \tag{4.34}$$

For an admissible operator $\widehat{\mathcal{T}}$ and the entropy solution u from Assumption 4.13, we define the function $\gamma_\varepsilon : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ by

$$\gamma_\varepsilon(\Delta) := \operatorname*{ess\,sup}_{t\in[0,T]} \left\{ \varepsilon(\Delta, \mathbb{R}^2, \widehat{\mathcal{T}}[u(\cdot,t)]) \right\}, \tag{4.35}$$

for $\Delta > 0$ and $\lim_{\Delta\to 0} \gamma_\varepsilon(\Delta) = 0$.

4.25 Remark: *Since we have $u(\cdot,t) \in L^\infty(\mathbb{R}^2)$ and therefore by assumption $\widehat{\mathcal{T}}[u(\cdot,t)] \in L^\infty(\mathbb{R}^2)$, classical results for L^∞-functions lead to $\gamma_\varepsilon(\Delta) = o(1)$ (cf. [Kru70]). For operators with $\widehat{\mathcal{T}}[u(\cdot,t)] \in BV(\mathbb{R}^2)$, $t \in [0,T]$, we even have $\gamma_\varepsilon(\Delta) = \mathcal{O}(\Delta)$.*

Now we have gathered all the definitions necessary for formulating our main result.

4.26 Theorem (Convergence with Non–Local Source Term)
Let $u \in W(0,T)$ be the entropy solution to problem (4.13) with data satisfying Assumption 4.3. With $u_\Omega \in W(0,T)$ we denote the entropy solution to problem (4.32) with $|\Omega| \ge 1$. Let $\{u_{\Omega,h}\}_h$ be a family of discrete solutions for the truncated problem (4.32) defined by the finite–volume scheme from Definition 4.7 using an admissible discrete operator as given in Definition 4.23. For the grid $\mathcal{T}$ we assume that Assumption 4.4 holds with some constant c_G. The local Lipschitz constant of the numerical flux is denoted with $L_g(M)$ as in Definition 4.5(iv). Furthermore we suppose that there is a constant $M_T > 0$ (independent of h and Ω) such that for u and the functions $u_{\Omega,h}$

$$\|u\|_\infty, \|u_{\Omega,h}\|_\infty \le M_T \ . \tag{4.36}$$

If the time–step satisfies the CFL condition

$$\frac{\Delta t}{h} \le \frac{(1-\xi)c_G^2}{2L_g(M_T)} \tag{4.37}$$

for some $\xi \in (0,1)$, then there exists a constant $C > 0$ such that the estimate

$$\int_0^T \int_{\mathbb{R}^2} |u(\mathbf{x},t) - u_{\Omega,h}(\mathbf{x},t)| \, d\mathbf{x}dt \le$$

$$C\Big((1+R_T(\Omega)^2)h^{1/4}+\gamma_\varepsilon(h^{1/4})+R_T(\Omega)^2\gamma_{app}(h)+ \operatorname*{ess\,sup}_{t\in[0,T]}\int_{\mathbb{R}^2\setminus\Omega}|\widehat{\mathcal{T}}[u(\cdot,t)](\mathbf{x})|\,d\mathbf{x}dt\Big) \quad (4.38)$$

holds. The constant $C>0$ depends on the quantities $C_{\widehat{\mathcal{T}}}(M_T), c_G, \xi, T, u_0$ but not on h or Ω. The function R_T is defined as the smallest number such that for all $t\in[0,T]$

$$\operatorname{supp}(u_{\Omega,h}(\cdot,t))\cup\operatorname{supp}(u_\Omega(\cdot,t))\subset B_{R_T(\Omega)}(0) \quad (4.39)$$

holds.

Proof:
The proof is based on the splitting of the error into two parts: one associated with the error due to the finite–volume scheme and the other with the error due to the truncation of the non–local operator. Thus we study the errors $\|u-u_\Omega\|_1$ and $\|u_\Omega-u_{\Omega,h}\|_1$ according to

$$\|u-u_{\Omega,h}\|_1\le\|u_\Omega-u_{\Omega,h}\|_1+\|u-u_\Omega\|_1. \quad (4.40)$$

For the first term an analysis similar to the one leading to Theorem 4.20 leads to the first part of the error estimate. The main difficulty is keeping track of how the error not only decreases with decreasing h but how it also increases with the increasing size of Ω. For the second part of the error we show that it can be bounded by $\operatorname{ess\,sup}_{t\in[0,T]}\int_{\mathbb{R}^2\setminus\Omega}|\widehat{\mathcal{T}}[u(\cdot,t)](\mathbf{x})|\,d\mathbf{x}dt$. Details of the proof can be found in [DR02b]. □

Theorem 4.26 gives an a priori error estimate in terms of h and the size of Ω. To obtain *convergence* for our numerical scheme in terms of the discretization parameter h alone, we couple the — so far arbitrary — set Ω to h.

4.27 Corollary
For $h\in(0,h_0]$ let $\Omega=\Omega_h$ be the ball of radius $\min\{h^{-1/16},(\gamma_{app}(h))^{-1/4}\}$ around the origon. Then under the same assumptions as in Theorem 4.26 we have

$$\lim_{h\to0}\|u-u_{\Omega_h,h}\|_1=0.$$

Proof:
From the definition of Ω_h and R_T we conclude that there is a constant $C>0$ that does not depend on h with

$$R_T(\Omega_h)^2(h^{1/4}+\gamma_{app}(h))\le C(h^{1/8}+\gamma_{app}(h)^{1/2})\to0 \text{ for } h\to0. \quad (4.41)$$

Furthermore we have $\lim_{h\to0}|\Omega_h|\to\infty$ and therefore

$$\lim_{h\to0}\operatorname*{ess\,sup}_{t\in[0,T]}\int_{\mathbb{R}^2\setminus\Omega_h}|\widehat{\mathcal{T}}[u(\cdot,t)](\mathbf{x})|=0 \quad (4.42)$$

for the L^1-function $\widehat{\mathcal{T}}[u(\cdot,t)]$. Since $\gamma_\varepsilon(h^{1/4})=o(1)$ (cf. Remark 4.25), the result follows from Theorem 4.26 using (4.41) and (4.42). □

4.28 Remark: *Corollary 4.27 does not establish to any positive rate of convergence. To prove such a rate some additional assumption on the decay of the L^1-function $\widehat{\mathcal{T}}[u(\cdot,t)]$ for $|\mathbf{x}| \to \infty$ is required. Furthermore we have to assume the existence of a constant $\alpha > 0$ such that*

$$\gamma_\varepsilon(h^{1/4}) + \gamma_{app}(h) = \mathcal{O}(h^\alpha).$$

While the first is an a priori assumption on the entropy solution u of the model problem (4.13), the latter is an assumption on the quality of the approximation. Let us consider the most basic choice for discretizing $\widehat{\mathcal{T}}_\Omega$ — also used in Theorem 4.20 —

$$\widehat{\mathcal{T}}_{\Omega,h}[w_h](\mathbf{x}) := \frac{1}{|T_j|}\int_{T_j} \widehat{\mathcal{T}}_\Omega[w_h](\mathbf{y})\,d\mathbf{y} \qquad (\mathbf{x} \in T_j, w_h \in V_h)\ .$$

Provided we have $\widehat{\mathcal{T}}_\Omega : L^1(\mathbb{R}^2) \cap L^\infty(\mathbb{R}^2) \to BV(\mathbb{R}^2)$, standard theory for BV-functions shows $\alpha = 1/4$ (cf. [Kru70]). Integral operators with smooth kernel functions that decay sufficiently fast provide examples of such compact operators (e.g. convolution operators). The question becomes more delicate when we consider weakly singular integral operators, which in general map into a compact subset of $L^1(\mathbb{R}^2)$ but not necessarily into $BV(\mathbb{R}^2)$. Here one has to consider each case separately.

4.5.1 The Model Problem from Radiation Hydrodynamics

We now apply our convergence result from Theorem 4.26 to the model problem from radiation hydrodynamics in two space dimensions derived in Section 4.1.1, i.e., for the balance term given by (4.9) with k given by (4.8) and $B(u) = \sigma u^4$. We study a more general setting that includes this case but does not depend on the special form of B and $\widetilde{k}$. In the following the operator is assumed to be of the form

$$\widehat{\mathcal{T}}[w](\mathbf{x}) = \int_{\mathbb{R}^2} \big(B(w(\mathbf{y})) - B(w(\mathbf{x}))\big) k(\mathbf{x},\mathbf{y}) d\mathbf{y} \tag{4.43}$$

with functions k, B satisfying:

4.29 Assumption (Continuous Operator)
Let the function $\tilde{k} : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined by $\tilde{k}(\mathbf{x},\mathbf{y}) := k(\mathbf{x},\mathbf{y})|\mathbf{x}-\mathbf{y}|$ be smooth. Furthermore we assume that there exists a monotone decreasing function $\beta : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ and a constant $C_k > 0$ such that

$$\int_0^\infty \beta(s)ds \ \leq\ C_k, \tag{4.44a}$$

$$0 \leq \tilde{k}(\mathbf{x},\mathbf{y}) \leq \beta(|\mathbf{x}-\mathbf{y}|) \ \leq\ C_k \frac{1}{|\mathbf{x}-\mathbf{y}|+1}, \tag{4.44b}$$

$$|\nabla_{\mathbf{x}}\tilde{k}(\mathbf{x},\mathbf{y})|, |\nabla_{\mathbf{y}}\tilde{k}(\mathbf{x},\mathbf{y})| \ \leq\ C_k \frac{1}{(|\mathbf{x}-\mathbf{y}|+1)^2}. \tag{4.44c}$$

For the function B we suppose $B \in C^1(\mathbb{R})$ and $B' \geq 0$. Define for $M > 0$:

$$C_B(M) = \max_{w\in[-M,M]} B'(w)\ .$$

How these assumptions fit into the framework of Theorem 4.26 and to what extent our model from radiation hydrodynamics (cf. Section 4.1.1) satisfies these assumptions is clarified in the following Lemmata and in Remark 4.36 below.
As a first step towards the error estimate we verify that the operator $\widehat{\mathcal{T}}$ with the data satisfying Assumption 4.29 belongs to the class of admissible operators as defined in Assumption 4.3.

4.30 Lemma
Consider the operator $\widehat{\mathcal{T}}$ of the form (4.43) such that Assumption 4.29 is satisfied. Then $\widehat{\mathcal{T}}$ is an admissible operator, i.e., there exists in particular a function $C_{\widehat{\mathcal{T}}} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that Assumption 4.3 is satisfied. Furthermore there exists a constant $C > 0$ such that we have for $w \in L^\infty(\mathbb{R}^2) \cap L^1(\mathbb{R}^2) \cap BV(\mathbb{R}^2)$ and $\Delta > 0$

$$\varepsilon(\Delta, \mathbb{R}^2, \widehat{\mathcal{T}}[w]) \leq C\Delta \tag{4.45}$$

(cf. Definition 4.24). The constant C depends only on $B, C_k, \|w\|_\infty, \|w\|_1, |w|_{BV}$.

Proof:
We first prove the estimate (4.45); similar arguments can be used to show that Assumption 4.3 is satisfied. With $\Delta > 0$ and $\mathbf{z} \in \mathbb{R}^2$ fixed with $|\mathbf{z}| < \Delta$ we consider for functions $w \in L^\infty(\mathbb{R}^2) \cap BV(\mathbb{R}^2)$

$$\begin{aligned}
&\int_{\mathbb{R}^2} |\widehat{\mathcal{T}}[w](\mathbf{x}+\mathbf{z}) - \widehat{\mathcal{T}}[w](\mathbf{x})| d\mathbf{x} \\
&= \int_{\mathbb{R}^2} \Big| \int_{\mathbb{R}^2} B(w(\mathbf{y}))k(\mathbf{x}+\mathbf{z},\mathbf{y})d\mathbf{y} - \int_{\mathbb{R}^2} B(w(\mathbf{y}))k(\mathbf{x},\mathbf{y})d\mathbf{y} \\
&\qquad - B(w(\mathbf{x}+\mathbf{z})) \int_{\mathbb{R}^2} k(\mathbf{x}+\mathbf{z},\mathbf{y})d\mathbf{y} + B(w(\mathbf{x})) \int_{\mathbb{R}^2} k(\mathbf{x},\mathbf{y})d\mathbf{y} \Big| d\mathbf{x} \\
&= \int_{\mathbb{R}^2} \Big| \int_{\mathbb{R}^2} B(w(\mathbf{y}+\mathbf{z}))k(\mathbf{x}+\mathbf{z},\mathbf{y}+\mathbf{z})d\mathbf{y} - \int_{\mathbb{R}^2} B(w(\mathbf{y}))k(\mathbf{x},\mathbf{y})d\mathbf{y} \\
&\quad + B(w(\mathbf{x}+\mathbf{z})) \int_{\mathbb{R}^2} k(\mathbf{x}+\mathbf{z},\mathbf{y}+\mathbf{z})d\mathbf{y} - B(w(\mathbf{x})) \int_{\mathbb{R}^2} k(\mathbf{x},\mathbf{y})d\mathbf{y} \Big| d\mathbf{x} \\
&\leq \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \Big|B(w(\mathbf{y}+\mathbf{z}))\Big|\Big|k(\mathbf{x}+\mathbf{z},\mathbf{y}+\mathbf{z}) - k(\mathbf{x},\mathbf{y})\Big| d\mathbf{y}d\mathbf{x} \\
&\quad + \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \Big|B(w(\mathbf{y}+\mathbf{z})) - B(w(\mathbf{y}))\Big| k(\mathbf{x},\mathbf{y})d\mathbf{y}d\mathbf{x} \\
&\quad + \int_{\mathbb{R}^2} \Big|B(w(\mathbf{x}+\mathbf{z}))\Big| \int_{\mathbb{R}^2} \Big|k(\mathbf{x}+\mathbf{z},\mathbf{y}+\mathbf{z}) - k(\mathbf{x},\mathbf{y})\Big| d\mathbf{y}d\mathbf{x} \\
&\quad + \int_{\mathbb{R}^2} \Big|B(w(\mathbf{x}+\mathbf{z})) - B(w(\mathbf{x}))\Big| \int_{\mathbb{R}^2} k(\mathbf{x},\mathbf{y})d\mathbf{y}d\mathbf{x}
\end{aligned}$$

$$=: \quad A_1 + B_1 + A_2 + B_2 \ .$$

Using $k(\mathbf{x},\mathbf{y}) \le \frac{\beta(|\mathbf{x}-\mathbf{y}|)}{|\mathbf{x}-\mathbf{y}|}$ (cf. Assumption 4.29) we can combine B_1 and B_2:

$$\begin{aligned} B_1 + B_2 &\le 2 \int_{\mathbb{R}^2} \Big| B(w(\mathbf{y})) - B(w(\mathbf{y}+\mathbf{z})) \Big| \int_{\mathbb{R}^2} \frac{\beta(|\mathbf{x}-\mathbf{y}|)}{|\mathbf{x}-\mathbf{y}|} d\mathbf{x} d\mathbf{y} \\ &\le 2C_B(\|w\|_\infty) \int_{\mathbb{R}^2} \Big| w(\mathbf{y}+\mathbf{z}) - w(\mathbf{y}) \Big| \int_0^\infty \beta(r) dr d\mathbf{y} \le 2C_B(\|w\|_\infty)\Delta |w|_{BV} C_k \end{aligned}$$

where we have used $w \in BV(\mathbb{R}^2)$, $|\mathbf{z}| \le \Delta$, and (4.44a). To bound A_1 and A_2 we expand the smooth function $\tilde{k}$ around $(\mathbf{x},\mathbf{y})$ and exploit (4.44c) and $|\mathbf{z}| \le \Delta$. Again we can combine both terms:

$$\begin{aligned} B_1 + B_2 &\le 2 \int_{\mathbb{R}^2} \Big| B(w(\mathbf{y})) \Big| \frac{|\tilde{k}(\mathbf{x}+\mathbf{z},\mathbf{y}+\mathbf{z}) - \tilde{k}(\mathbf{x},\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|} d\mathbf{x} d\mathbf{y} \\ &\le 2 \int_{\mathbb{R}^2} C_k \Big| B(w(\mathbf{y})) \Big| \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{\Delta}{(|\mathbf{x}-\mathbf{y}|+1)^2 |\mathbf{x}-\mathbf{y}|} d\mathbf{x} d\mathbf{y} \\ &\le 2C_B(\|w\|_\infty) C_k \int_{\mathbb{R}^2} \Big| w(\mathbf{y}) \Big| \int_0^\infty \frac{\Delta}{|r+1|^2} dr d\mathbf{y} = 2C_B(\|w\|_\infty) C_k \|w\|_1 \Delta \ . \end{aligned}$$

This leads to the desired result $\varepsilon(\Delta, \mathbb{R}^2, w) \le C\Delta$ for some constant $C > 0$.
Straightforward computations show that the estimates on the operator stated in Assumption 4.3*(ii)* are satisfied for any function $w \in L^1(\mathbb{R}^2) \cap L^\infty(\mathbb{R}^2)$ with $\|w\|_\infty \le M$ and $M > 0$ by choosing $C_{\widehat{\mathcal{T}}}(M) = 4\pi C_B(M) C_k$. □

Now we turn to the discrete setting. We extend Assumption 4.4:

4.31 Assumption
Consider a time step $\Delta t > 0$ and, for $h \in (0, h_0]$, grids $\mathcal{T}_h$ such that Assumption 4.4 holds. Furthermore, for $j \in \mathcal{I}_h$, let $\rho_j = \rho_j(h)$ be the radius of the largest circle with midpoint $\boldsymbol{\omega}_j$ that lies in the closure of T_j. We define

$$\rho = \inf\{\rho_j \ : \ j \in \mathcal{I}_h\} \ . \tag{4.46}$$

Note that ρ depends on h. In addition to Assumption 4.4 we assume that there is a constant $\hat{c} > 0$ independent of h such that

$$\frac{h}{\rho} \le \hat{c} \ . \tag{4.47}$$

We define subsets of $\mathcal{I}_h$ for $l \in \mathbb{N}$ and $\mathbf{x} \in \mathbb{R}^2$: $\mathcal{I}_h^l(\mathbf{x}) := \{j \in \mathcal{I}_h \ : \ l\rho \le |\mathbf{x} - \boldsymbol{\omega}_j| < (l+1)\rho\}$. Note that there exists a constant $\tilde{c}$ independent of h, ρ, and $\mathbf{x}$ such that $card(\mathcal{I}_h^l(\mathbf{x})) \le \tilde{c}l$.

The main difficulty in defining a suitable discrete operator $\widehat{\mathcal{T}}_h$ for our finite–volume scheme lies in the singularity of the kernel k for $\mathbf{x} = \mathbf{y}$. For the approximations used in our RMHD simulations (cf. Chapters 11–13) we are not able to verify the admissibility criteria (Definition 4.23) required in our convergence proof. Nevertheless we are able to define a fully discrete operator:

4.32 Definition (Discrete Operator)
For a family of grids $\{\mathcal{T}_h\}_h$ satisfying Assumption 4.31 and for functions $\beta, \tilde{k}$ satisfying Assumption 4.29, we consider the operator $\widehat{\mathcal{T}}_{\Omega,h} : V_h \to V_h$ defined for grid functions $w_h \in V_h$ and for $\mathbf{x} \in T_j$ $(j \in \mathcal{I}_h)$ by

$$\widehat{\mathcal{T}}_{\Omega,h}[w_h](\mathbf{x}) = \chi_\Omega(\mathbf{x}) \sum_{i \in \mathcal{I}_h \setminus \{j\}} |T_i| \big(B(w_h(\boldsymbol{\omega}_i)) - B(w_h(\boldsymbol{\omega}_j))\big) k_h(\boldsymbol{\omega}_j, \boldsymbol{\omega}_i) \ , \tag{4.48}$$

$$k_h(\mathbf{x}, \mathbf{y}) = \frac{\tilde{k}(\mathbf{x}, \mathbf{y})}{|\mathbf{x} - \mathbf{y}| + h} \quad \textit{for} \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^2 \ .$$

Here Ω is an arbitrary compact subset of $\mathbb{R}^2$ with the characteristic function χ_Ω. From the definition (4.46) of ρ and since β is monotone decreasing according to Assumption 4.29 we deduce for $j \in \mathcal{I}_h$ that

$$\sum_{i \in \mathcal{I}_h \setminus \{j\}} |T_i| k_h(\boldsymbol{\omega}_j, \boldsymbol{\omega}_i) \le \sum_{l \in \mathbb{N}} \sum_{i \in \mathcal{I}_h^l(\boldsymbol{\omega}_j)} h^2 \frac{\beta(|\boldsymbol{\omega}_j - \boldsymbol{\omega}_i|)}{|\boldsymbol{\omega}_j - \boldsymbol{\omega}_i| + h} \le C(\hat{c}, C_k) \sum_{l \in \mathbb{N}} h \beta(l\rho).$$

The last sum is bounded due to (4.44a); similarly it follows that $\sum |T_i| k_h(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j)$ is bounded. We introduce the constant $C_k^h = C_k^h(\tilde{c}, \hat{c}, C_k)$ such that

$$\sum_{i \in \mathcal{I}_h \setminus \{j\}} |T_i| k_h(\boldsymbol{\omega}_j, \boldsymbol{\omega}_i), \ \sum_{i \in \mathcal{I}_h \setminus \{j\}} |T_i| k_h(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) \le C_k^h. \tag{4.49}$$

The discrete operator from Definition 4.32 is an admissible discrete operator in the sense of Definition 4.23 as we show in the next Lemma.

4.33 Lemma
The discrete operator $\widehat{\mathcal{T}}_{\Omega,h}$ from (4.48) is an admissible discrete operator in the sense of Definition 4.23. A function γ_{app} such that Definition 4.23(v) is satisfied for each $w_h \in V_h \cap L^\infty(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$ with $\|w_h\|_\infty \le M$ with $M > 0$ is given by

$$\gamma_{app}(h) = Ch|\ln(h)| \quad (h \in (0, h_0]), \tag{4.50}$$

where $C = C(B, M, C_k, \hat{c}, \tilde{c})$ does not depend on h with $h \le h_0 < 1$.

Proof:
For w_h given as in the statement, we use the abbreviation $w_j = w_h(\boldsymbol{\omega}_j)$ for $j \in \mathcal{I}_h$. To verify 4.23*(i)* we observe that by (4.49) we have for all $j \in \mathcal{I}_h$ and $\mathbf{x} \in T_j$

$$\begin{aligned} |\widehat{\mathcal{T}}_{\Omega,h}[w_h](x)| &\le C_B(M) M \Big(\sum_{i \in \mathcal{I}_h \setminus \{j\}} |T_i| k_h(\boldsymbol{\omega}_j, \boldsymbol{\omega}_i) + \sum_{i \in \mathcal{I}_h \setminus \{j\}} |T_i| k_h(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) \Big) \\ &\le C(B, M, C_k^h). \end{aligned}$$

To verify 4.23*(ii)* we again use (4.49) to compute

$$\begin{aligned}\|\widehat{\mathcal{T}}_{\Omega,h}[w_h]\|_1 &\leq C_B(M)\sum_{j\in\mathcal{I}_h}|T_j|\sum_{i\in\mathcal{I}_h\setminus\{j\}}(|w_i|+|w_j|)|T_i|k_h(\boldsymbol{\omega}_j,\boldsymbol{\omega}_i)\\ &\leq C(B,M,C_k^h)\|w_h\|_1.\end{aligned}$$

Property 4.23*(iii)* is clear. The estimate 4.23*(iv)* follows using the same arguments as for 4.23*(ii)*.
For a further grid function $\tilde{w}_h \in V_h \cap L^\infty(\mathbb{R}^2)\cap L^1(\mathbb{R}^2)$, $\|\tilde{w}_h\|_\infty \leq M$ we compute

$$\begin{aligned}\sum_{j\in\mathcal{I}_h}|T_j|\left|\widehat{\mathcal{T}}_{\Omega,h}[w_h](\boldsymbol{\omega}_j)-\widehat{\mathcal{T}}_{\Omega,h}[\tilde{w}_h](\boldsymbol{\omega}_j)\right| &\leq 2C_B(M)\sum_{j\in\mathcal{I}_h}|T_j||w_j-\tilde{w}_j|\sum_{i\in\mathcal{I}_h\setminus\{j\}}|T_i|k_h(\boldsymbol{\omega}_j,\boldsymbol{\omega}_i)\\ &\leq C(B,M,C_k^h)\sum_{j\in\mathcal{I}_h}|T_j||w_j-\tilde{w}_j|.\end{aligned}$$

It remains to prove the estimate from Definition 4.23*(v)*. For some $\mathbf{x}\in T_j$ and $j\in\mathcal{I}_h$ consider

$$\begin{aligned}&\left|\widehat{\mathcal{T}}_{\Omega}[w_h](\mathbf{x})-\widehat{\mathcal{T}}_{\Omega,h}[w_h](\mathbf{x})\right|\\ &\leq \left|\sum_{i\in\mathcal{I}_h\setminus\{j\}}(B(w_i)-B(w_j))\int_{T_i}(k(\mathbf{x},\mathbf{y})-k_h(\boldsymbol{\omega}_j,\boldsymbol{\omega}_i))d\mathbf{y}\right|\\ &\leq C_B(M)M\Bigg(\sum_{i\in\mathcal{I}_h\setminus\{j\}}\int_{T_i}|k(\mathbf{x},\mathbf{y})-k_h(\mathbf{x},\mathbf{y})|d\mathbf{y}+\sum_{i\in\mathcal{I}_h\setminus\{j\}}\int_{T_i}|k_h(\mathbf{x},\mathbf{y})-k_h(x,\omega_i)|d\mathbf{y}\\ &\qquad+\sum_{i\in\mathcal{I}_h\setminus\{j\}}\int_{T_i}|k_h(\mathbf{x},\boldsymbol{\omega}_i)-k_h(\boldsymbol{\omega}_j,\boldsymbol{\omega}_i)|d\mathbf{y}\Bigg)\\ &=: C(B,M)\left(A_1(\mathbf{x})+A_2(\mathbf{x})+A_3(\mathbf{x})\right).\end{aligned}$$

Using (4.44b) and our definition of ρ in (4.46) it follows that

$$\begin{aligned}A_1(\mathbf{x}) &\leq \int_{\mathbb{R}^2\setminus B_\rho(\boldsymbol{\omega}_j)}\frac{h\tilde{k}(\mathbf{x},\mathbf{y})}{|\mathbf{x}-\mathbf{y}|(|\mathbf{x}-\mathbf{y}|+h)}d\mathbf{y}\\ &\leq hC_k\int_{\mathbb{R}^2\setminus B_\rho(\boldsymbol{\omega}_j)}\frac{1}{|\mathbf{x}-\mathbf{y}|(|\mathbf{x}-\mathbf{y}|+h)(|\mathbf{x}-\mathbf{y}|+1)}d\mathbf{y}\leq Ch|\ln(h)|.\end{aligned}$$

To estimate A_2 we use Taylor expansion for the smooth function k_h:

$$A_2(\mathbf{x})\leq\sum_{i\in\mathcal{I}_h\setminus\{j\}}\|\nabla_{\mathbf{y}}k_h(\mathbf{x},\cdot)\|_{L^\infty(T_i)}\int_{T_i}|\mathbf{y}-\boldsymbol{\omega}_i|d\mathbf{y}\leq h^3\sum_{l\in\mathbb{N}}\sum_{i\in I_l^h(\mathbf{x})}\|\nabla_{\mathbf{y}}k_h(\mathbf{x},\cdot)\|_{L^\infty(T_i)}.$$

Since $h<h_0<1$ and due to (4.44a), (4.44b), we can bound $|\nabla_{\mathbf{y}}k_h(\mathbf{x},\mathbf{y})|$ for $\mathbf{y}\in T_i$ by $\frac{C_k}{(|\mathbf{x}-\mathbf{y}|+1)(|\mathbf{x}-\mathbf{y}|+h)^2}$; this implies

$$\|\nabla_{\mathbf{y}}k_h(\mathbf{x},\cdot)\|_{L^\infty(T_i)}\leq\frac{C_k}{(l\rho-h+1)(l\rho)^2}.$$

Note that we have $|\mathbf{x}-\mathbf{y}| \geq |\mathbf{x}-\boldsymbol{\omega}_i| - |\boldsymbol{\omega}_i - \mathbf{y}| \geq l\rho - h$ for all $\mathbf{y} \in T_i$. With (4.47) we deduce by elementary calculus

$$A_2(\mathbf{x}) \leq C(C_k, \hat{c}, \tilde{c})h|\ln(h)|.$$

Similar calculations as for A_2 show that we have for all $i \in \mathcal{I}_h^l(\boldsymbol{\omega}_i)$

$$\|\nabla_{\mathbf{x}} k_h(\cdot, \boldsymbol{\omega}_i)\| L^\infty(T_j) \leq \frac{C_k}{(l\rho - h + 1)(l\rho)^2}.$$

Therefore we arrive at the same bound for A_3 as for A_2. Adding the three estimates for A_1, A_2, A_3 leads us to

$$\int_{T_j} \left| \widehat{\mathcal{T}}[w_h](\mathbf{x}) - \widehat{\mathcal{T}}_h[w_h](\mathbf{x}) \right| \leq C(B, M, C_k, \hat{c}, \tilde{c})|T_j|h|\ln(h)|.$$

This proves estimate 4.23*(v)* and concludes the proof of the lemma. □

4.34 Remark: *With Lemma 4.33 at hand and Definition 4.7 we obtain a fully explicit finite–volume scheme to approximate entropy solutions of* (4.13)*. Note that, although the scheme can be used to approximate (4.13), the computation cost for evaluating the discrete operator $\widehat{\mathcal{T}}_h$ grows quadratically with the number of elements. Approximations that grow only linearly are available (cf. Chapter 12), but any approximation results in the sense of Definition 4.23 known to us require some additional smoothness assumptions that are not fulfilled in the setting studied here (e.g. Theorem 11.1 and Theorem 12.10).*

We can now prove the boundedness of the finite–volume approximation $u_{\Omega,h}$.

4.35 Lemma (Maximum Principle)
Let the time–step Δt satisfy the CFL-like condition

$$\Delta t \leq \frac{c_G^2 h}{K L_g(\|u_0\|_\infty) + c_G^2 h C_B(\|u_0\|_\infty) C_k^h} \ . \tag{4.51}$$

where K denotes the maximum number of neighbors of each element, i.e. $|\mathcal{N}(i) \leq K$ for all $i \in \mathcal{I}_h$; we assume that K is independent of h. Then for $h \in (0, h_0]$ and for compact sets $\Omega \subset \mathbb{R}^2$ the discrete solution $u_{\Omega,h}$ satisfies the maximum principle

$$\operatorname*{ess\,inf}_{\mathbf{x}\in\mathbb{R}^2}\{u_0(\mathbf{x})\} \leq u_{\Omega,h}(\mathbf{x}, t) \leq \operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{R}^2}\{u_0(\mathbf{x})\} \quad (\mathbf{x} \in \mathbb{R}^2, t \in [0, T]) \ . \tag{4.52}$$

Proof:
For all $n \in \{0, \ldots, N_T\}$, $i, j \in \mathcal{I}_h$, and $l \in \mathcal{N}(j)$, we define

$$\Delta G_{jl}^n := \frac{g_{jl}(u_j^n, u_l^n) - g_{jl}(u_j^n, u_j^n)}{u_j^n - u_l^n}, \qquad \Delta B_{ji}^n := \frac{B(u_i^n) - B(u_j^n)}{u_i^n - u_j^n}$$

provided $u_j^n \neq u_{jl}^n$, $u_j^n \neq u_i^n$. Otherwise set $\Delta G_{jl}^n = 0$ and $\Delta B_{ji}^n = 0$. Note that from the consistency of g_{jl} (cf. Definition 4.5*(ii)*) it follows that $\sum_{l\in\mathcal{N}(j)} g_{jl}(u_j^n, u_j^n) = 0$. Rewriting formula (4.16) for $n = 1$ and using the discretization (4.48) we obtain

$$u_j^1 \ = \ u_j^0 - \frac{\Delta t}{|T_j|} \sum_{l\in\mathcal{N}(j)} \Delta G_{jl}^0 (u_j^0 - u_l^0) + \Delta t \sum_{i\in\mathcal{I}_h} \Delta B_{ji}^0 (u_i^0 - u_j^0)|T_i| k_h(\boldsymbol{\omega}_j, \boldsymbol{\omega}_i)$$

$$= u_j^0\left(1-\frac{\Delta t}{|T_j|}\sum_{l\in\mathcal{N}(j)}\Delta G_{jl}^0-\Delta t\sum_{i\in\mathcal{I}_h}\Delta B_{ji}^0|T_i|k_h(\boldsymbol{\omega}_j,\boldsymbol{\omega}_i)\right)$$
$$+\frac{\Delta t}{|T_j|}\sum_{l\in\mathcal{N}(j)}\Delta G_{jl}^0u_l^0+\Delta t\sum_{i\in\mathcal{I}_h}\Delta B_{ji}^0u_i^0|T_i|k_h(\boldsymbol{\omega}_j,\boldsymbol{\omega}_i).$$

Since g_{jl} is monotone non–increasing in the second variable (cf. Definition 4.5), we have $\Delta G_{jl}^0\geq 0$. Since $B'\geq 0$ and using (4.44b), we furthermore have $\Delta B_{ji}^0k_h(\boldsymbol{\omega}_j,\boldsymbol{\omega}_i)\geq 0$. This implies the lower bound in

$$0\leq\frac{1}{|T_j|}\sum_{l\in\mathcal{N}(j)}\Delta G_{jl}^0+\sum_{i\in\mathcal{I}_h}\Delta B_{ji}^0|T_i|k_h(\boldsymbol{\omega}_j,\boldsymbol{\omega}_i)\leq\frac{KL_g(\|u_0\|_\infty)}{c_G^2h}+C_B(\|u_0\|_\infty)C_k^h.$$

The upper bound follows from Definition 4.5, Assumption 4.4, and (4.49). Due to the CFL condition (4.51) and the compact support of $u_{\Omega,h}$ we have

$$\begin{aligned}u_j^1 &\leq \max_{j\in\mathcal{I}_h}\{u_j^0\}\left(1-\frac{\Delta t}{|T_j|}\sum_{l\in\mathcal{N}(j)}\Delta G_{jl}^0-\Delta t\sum_{j\in\mathcal{I}_h}\Delta B_{ji}^0|T_i|k_h(\boldsymbol{\omega}_j,\boldsymbol{\omega}_i)\right)+\\ &\quad \max_{j\in\mathcal{I}_h}\{u_j^0\}\frac{\Delta t}{|T_j|}\sum_{l\in\mathcal{N}(j)}\Delta G_{jl}^0+\max_{j\in\mathcal{I}_h}\{u_j^0\}\Delta t\sum_{j\in\mathcal{I}_h}B_{ji}^0|T_i|k_h(\boldsymbol{\omega}_j,\boldsymbol{\omega}_i)\\ &\leq \max_{j\in\mathcal{I}_h}\{u_j^0\}\leq\operatorname{ess\,sup}_{\mathbf{x}\in\mathbb{R}^2}\{u_0(\mathbf{x})\}.\end{aligned}$$

In the same way we can prove that $u_j^1\geq\operatorname{ess\,inf}_{\mathbf{x}\in\mathbb{R}^2}\{u_0(\mathbf{x})\}$. With a standard induction argument we arrive at the estimate

$$\operatorname*{ess\,inf}_{\mathbf{x}\in\mathbb{R}^2}\{u_0(\mathbf{x})\}\leq u_i^n\leq\operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{R}^2}\{u_0(\mathbf{x})\}$$

for all $n\in\{0,\ldots,N_T\}$ and $i\in\mathcal{I}_h$. Due to Definition 4.7 this proves (4.52). □

4.36 Remark: *Using Lemma 4.35 it is now easy to see that our model problem from radiation hydrodynamics derived in Section 4.1.1 fits into the class of problems discussed here. The integral kernel* $\int_0^\pi\exp\left(-\frac{|\mathbf{x}-\mathbf{y}|}{\sin(\vartheta)}\right)d\vartheta$ *satisfies Assumption 4.29 with* $\beta(s)=e^{-s}$ *since* $\sin(\vartheta)$ *is bounded. It remains to show that the function* $B(u)=\sigma u^4$ *fits into the context studied here. For* $u\geq 0$ *this is obvious. For* $u<0$ *the assumption* $B'(u)>0$ *is not valid. Note that in this case we are not in the physical regime and the formula* (1.18)*, which leads to the definition of* B*, does not hold in this case. Due to Lemma 4.35 the approximate solution* $u_{\Omega,h}$ *always remains within the bounds given by the initial data* u_0*. Since the initial data should lie in the physical regime we have* $u_0>0$*. It follows that the approximation* $u_{\Omega,h}$ *is positive, and the definition of* B *on the negative real axis is not relevant for the approximation of our model problem. We can therefore redefine* $B(u)$ *for* $u<0$ *so that Assumption 4.29 is satisfied without influencing* $u_{\Omega,h}$*. Since we show the convergence of* $u_{\Omega,h}$ *to the entropy solution* u*, this must also remain in the physical regime independent of the definition of* B *on the negative real axis. In this sense our model problem fits into the framework studied here.*

In the concluding theorem we use the lemmata above to obtain a convergent finite–volume algorithm for the approximation of entropy solutions of (4.13). As in Corollary 4.27 we let the domain Ω depend on h, i.e. $\Omega = \Omega_h$. Since in the case of this special model problem much stronger estimates than in the general case of Corollary 4.27 are available, we obtain convergence with only mild (in fact logarithmical) growth of Ω_h for $h \to 0$. If additionally we assume some (natural) decay rate for the entropy solution u, we recover (again up to a logarithmically small factor) the classical convergence rate of 1/4 as obtained in the case without non–local operator (cf. Theorem 4.20).

4.37 Theorem (Convergence with Radiation Source Term)
Let $u \in W(0,T)$ be the entropy solution to problem (4.13) with data satisfying Assumption 4.3. We assume that the balance term $\widehat{\mathcal{T}}$ is of the form (4.43) with data satisfying Assumption 4.29. Consider the discrete operator $\widehat{\mathcal{T}}_{\Omega,h}$ from Definition 4.32 with $\Omega = \Omega_h = \{\mathbf{x} \in \mathbb{R}^2 \,|\, |\mathbf{x}| \leq \sqrt{|\ln(h)|}\}$. Suppose that the CFL-like conditions (4.37) and (4.51) hold. Then the sequence of approximate solutions $\{u_{\Omega_h,h}\}$ given by Definition 4.7 on a family of grids $\{\mathcal{T}_h\}_h$ satisfying Assumption 4.31 converges to u:

$$\lim_{h\to 0} \|u - u_{\Omega_h,h}\|_1 = 0. \tag{4.53}$$

If additionally the operator $\widehat{\mathcal{T}}$ satisfies

$$\widehat{\mathcal{T}}[u(\cdot,t)](\mathbf{x}) = \mathcal{O}(\exp(-|\mathbf{x}|)) \quad (|\mathbf{x}| \to \infty) \tag{4.54}$$

for all $t \in [0,T]$, we obtain for some positive constant C independent of h the error estimate

$$\|u - u_{\Omega_h,h}\|_1 \leq C h^{\frac{1}{4}} |\ln(h)|. \tag{4.55}$$

Proof:
The conditions of Theorem 4.26 are satisfied due to the Lemmata 4.30, 4.33, and 4.35. Note that we can choose $\gamma_\varepsilon(h^{1/4}) = Ch^{1/4}$ due to (4.45). With (4.50) and $\widehat{\mathcal{T}}[u(\cdot,t)] \in L^1(\mathbb{R}^2)$ we get the same convergence statement (4.53) as in Corollary 4.27. Straightforward calculus using (4.54) establishes the estimate (4.55). □

5. Summary

Base Scheme

In Chapter 1 we briefly derived the basic equations that serve as a mathematical model for the physical processes in the solar convection zone and photosphere. The model consists of eight balance laws for the density, the momentum, the magnetic field, and the energy density. This system is coupled via a source term in the energy equation with the stationary radiation transport equation. This equation describes the radiation intensity, which is influenced by the density and the temperature of the fluid. In Chapter 3 we detailed the solution algorithm used for solving this system of equations. We proposed an explicit finite–volume scheme on unstructured, locally adapted grids using a one dimensional Riemann solver for the approximation of the fluxes over the cell faces. To meet all the aspects of the different physical regimes (e.g. including general equation of state) and to get a stable and reliable code (e.g. reducing error in the divergence constraint (1.1e)) the flux has to be modified. This is detailed in Chapters 7–9.
A second and largely independent part of our algorithm is the computation of the radiation field for a given temperature and density distribution. By approximating the integrals defining the radiation source term using the discrete ordinate method (cf. Section 3.1), this problem is reduced to the approximation of the radiation transport equation for a fixed set of directions. In Chapter 12 and Chapter 13 we focus on this part of the scheme. The same idea can also be used to approximate the frequency dependent radiation transport equation.
In Chapter 4 we studied a model problem for the complex coupled system. The major simplification was to reduce the problem to the case of a scalar balance law (4.13). The novel aspect of this balance law is the non–local operator on the right hand side modelling the radiation source term Q_{rad} in our coupled system. A discussion of the characteristic properties of the solution to the model problem served as justification for studying the scalar problem instead of the full system. Due to these simplifications, we were in a position to give some answers to important questions like the existence and uniqueness of entropy solutions for special initial data and were able to study the convergence properties of our finite–volume scheme. We extended some results from the standard theory for scalar balance laws to the case including a non–local operator of the form given by the radiation source term. Large parts of the following chapters will be devoted to a further justification of our scheme by experimental means.
A sketch of the finite–volume scheme presented so far is given in Algorithm 1 on page 65 and Algorithm 2 on page 66. In the algorithm the approximation $\{U_i^n\}_{i\in\mathcal{I}}$ on a given

grid $\mathcal{T}^n$ and time t^n is evolved to the next time level $t^{n+1} = t^n + \Delta t^n$. The new approximation is thereby defined on a new grid $\mathcal{T}^{n+1}$. Since in our code we only store the most recent grid and approximate values, we leave out the index n. Some realistic simulations using this scheme are presented in Chapter 15. For our 2d code we use a parallelization strategy for machines with shared memory architecture. The 3d code uses MPI for execution on distributed memory machines.
We can distinguish two parts of the algorithm. One part requires computations on the faces S_{ij} for all $(i, j) \in \mathcal{I}_S^n$ — for example, the computation of the fluxes, the local time steps, and the indicators for the adaptation process. The second group consists of calculations on the elements T_i for all $i \in \mathcal{I}^n$, most notably the computation of the reconstruction $\mathcal{L}_i$, the source term calculation, and the update step. If we neglect the radiation transport, we see that to compute the new conserved vectors we need the old conserved quantities on each cell and for each face we need the conserved quantities on either side of the face. This is a minimum stencil, which we require for both our first order scheme and our second order scheme. This small stencil allows us to use an efficient and simple parallelization strategy as discussed in Section 3.6. The computation of the radiation field, the evolution of the magnetohydrodynamic quantities, and the organization of the locally adapted grid are a second possible grouping of the steps in the algorithm.

Algorithm 1: Algorithm for computing a sequence of grids and average values $\{\mathcal{T}^n, \{\mathbf{U}_i^n\}_{i\in\mathcal{I}^n}, \{\mathbf{U}_i^{n+\frac{1}{2},t^n}\}_{i\in\mathcal{I}^n}\}_{n=1}^{N_T}$ using an explicit second order time–stepping scheme. For the computation of the solution a grid $\mathcal{T}$ is used together with values $(\mathbf{U}_i, \mathbf{g}_i, \mathrm{ref}_i, \mathrm{crs}_i)$ stored on each element T_i for $i \in \mathcal{I}$. In the parallel case the algorithm is executed by each processor separately, and the grid $\mathcal{T}$ is in this case only a part of the whole computational grid. The highlighted steps indicate parts of the algorithm that lead to communication between the processors. The algorithm presented here does not represent our implementation in every detail. For a more accurate description of the 2d code we refer to [Ded98, DRW99] and for the 3d code to [Sch99a, DRSW02]. The parts of the finite–volume scheme either act on the surfaces of the grid requiring the left and right limits of the approximation or on the elements requiring only the values stored there. Some additional postprocessing steps are required that act on the whole grid. These methods are sketched on the following page.

(a) Finite–Volume Scheme

Initialize
while $t < T$ **do**
  $\widehat{\mathbf{U}}_i \leftarrow \mathbf{U}_i$ $(i \in \mathcal{I})$
  $(\{\mathbf{g}_i\}, \{\mathrm{ref}_i\}, \{\mathrm{crs}_i\}, \Delta t^n) \leftarrow$ **Surface**
  $(\{\mathbf{g}_i\}) \leftarrow$ **Element**
  minimize Δt^n over all partitions
  for all $i \in \mathcal{I}$ **do**
    $\mathbf{U}_i \leftarrow \mathbf{U}_i + \Delta t^n \mathbf{g}_i$
  end for
  $(\{\mathcal{L}_i\}, \{\mathbf{U}_i\}_{i\in\mathcal{I}_B}) \leftarrow$ **Poststep**
  $\mathbf{U}_i^{n+\frac{1}{2}} \leftarrow \mathbf{U}_i$ $(i \in \mathcal{I}^n)$
  $(\{\mathbf{g}_i\}, \{\mathrm{ref}_i\}, \{\mathrm{crs}_i\}, \Delta t^{n+\frac{1}{2}}) \leftarrow$ **Surface**
  $(\{\mathbf{g}_i\}, \{M_i\}) \leftarrow$ **Element**
  for all $i \in \mathcal{I}$ **do**
    $\mathbf{U}_i \leftarrow \frac{1}{2}\big(\widehat{\mathbf{U}}_i + \mathbf{U}_i + \Delta t^n \mathbf{g}_i\big)$
  end for
  $(\{\mathcal{L}_i\}, \{\mathbf{U}_i\}_{i\in\mathcal{I}_B}) \leftarrow$ **Poststep**
  adapt grid $\mathcal{T}$ and load balance
  $(\{\mathcal{L}_i\}, \{\mathbf{U}_i\}_{i\in\mathcal{I}_B}) \leftarrow$ **Poststep**
  $\mathcal{T}^{n+1} \leftarrow \mathcal{T}$, $\mathbf{U}_i^{n+1} \leftarrow \mathbf{U}_i$ $(i \in \mathcal{I}^{n+1})$
  $n \leftarrow n+1, t \leftarrow t + \Delta t^n$
end while

(b) **Initialize**

$\mathcal{T} \leftarrow$ macro grid, refine $\mathcal{T}$ until h_{start} reached
for all $i \in \mathcal{I}$ **do**
  $\mathbf{U}_i \leftarrow \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \mathbf{U}_0(\mathbf{z}_{ij})$
end for
for $r = 1$ to 10 **do**
  adapt grid $\mathcal{T}$ and load balance
  for all $i \in \mathcal{I}$ **do**
    $\mathbf{U}_i \leftarrow \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \mathbf{U}_0(\mathbf{z}_{ij})$
  end for
end for
$t \leftarrow 0, n \leftarrow 0$
$(\{\mathcal{L}_i\}, \{\mathbf{U}_i\}_{i\in\mathcal{I}_B}) \leftarrow$ **Poststep**
$\mathcal{T}^n \leftarrow \mathcal{T}, \mathbf{U}_i^n \leftarrow \mathbf{U}_i$ $(i \in \mathcal{I}^n)$

(c) **Poststep**

for all $i \in \mathcal{I}_B$ **do**
  $\mathbf{U}_i \leftarrow$ [cf. Section 3.3]
end for
exchange data on inner boundary
for all $i \in \mathcal{I}$ **do**
  $\{\mathcal{L}_i\} \leftarrow$ **Reconstruction**
  $\{Q_{\mathrm{rad}\,i}\} \leftarrow$ [cf. Chapter 12–13]
  $\mathbf{g}_i \leftarrow 0, \mathrm{ref}_i \leftarrow 0, \mathrm{crs}_i \leftarrow 0$
end for

Algorithm 2: The parts of the algorithm on the left hand side have to be executed for each face S_{ij} of the grid. They modify values on the elements T_i and T_j and only require data stored there. On the right hand side are the parts of the algorithm called on the elements T_i. These may also require the data on the neighboring elements but only modify the data on T_i.

(a) **Surface**

$\Delta t \leftarrow \infty$
for all $(i,j) \in \mathcal{I}_S$ **do**
  $(\mathbf{g}_{ij}, \Delta t_{ij}, \mathrm{jmp}_{ij}) \leftarrow$
    $\mathbf{Flux}(\mathcal{L}_i(\mathbf{z}_{ij}), \mathcal{L}_j(\mathbf{z}_{ij}), \mathbf{n}_{ij}, h_{ij})$
  $\mathbf{g}_i \leftarrow \mathbf{g}_i + \mathbf{g}_{ij}$
  $\mathrm{ref}_i \leftarrow \max\{\mathrm{ref}_i, \mathrm{jmp}_{ij}\}$
  $\mathrm{crs}_i \leftarrow \max\{\mathrm{crs}_i, \mathrm{jmp}_{ij}\}$
  if $j \in \mathcal{I}$ **then**
    $\mathbf{g}_j \leftarrow \mathbf{g}_j - \mathbf{g}_{ij}$
    $\mathrm{ref}_j \leftarrow \max\{\mathrm{ref}_j, \mathrm{jmp}_{ij}\}$
    $\mathrm{crs}_j \leftarrow \max\{\mathrm{crs}_j, \mathrm{jmp}_{ij}\}$
  end if
  $\Delta t \leftarrow \min\{\Delta t, \Delta t_{ij}\}$
end for

(b) **FLUX($\mathbf{U}_l, \mathbf{U}_r, \mathbf{n}, h$)**

$\bar{\mathbf{U}}_l \leftarrow \mathcal{R}(\mathbf{n}_{ij})\mathbf{U}_l$
$\bar{\mathbf{U}}_r \leftarrow \mathcal{R}(\mathbf{n}_{ij})\mathbf{U}_r$
$\bar{\mathbf{G}} \leftarrow \mathbf{G}(\bar{\mathbf{U}}_l, \bar{\mathbf{U}}_r)$
$\mathbf{g}_{ij} \leftarrow$
  $|S_{ij}|\mathcal{R}^{-1}(\mathbf{n}_{ij})\bar{\mathbf{G}}(\bar{\mathbf{U}}_l, \bar{\mathbf{U}}_r)$ [cf. Chapter 7–9]
$\Delta t_{ij} \leftarrow \frac{h}{\max\{|\bar{u}_{l,x}|+c_f(\bar{\mathbf{U}}_l), |\bar{u}_{r,x}|+c_f(\bar{\mathbf{U}}_r)\}}$
$\mathrm{jmp}_{ij} \leftarrow$ [cf. (3.19)]

(c) **Element**

for all $i \in \mathcal{I}$ **do**
  $\mathbf{g}_i \leftarrow -\mathbf{g}_i + \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \mathbf{q}(\mathcal{L}_i(\mathbf{z}_{ij})) + \mathbf{Q}_{\mathrm{rad}\,i}$
  if $\mathrm{ref}_i > \mathrm{ref}_{\mathrm{limit}}$ and $h_i > h_{min}$ **then**
    mark T_i for refinement
  else if $\mathrm{crs}_i < \mathrm{crs}_{\mathrm{limit}}$ and $h_i < h_{max}$ **then**
    mark T_i for coarsening
  end if
end for

(d) **Reconstruction**

for all $i \in \mathcal{I}$ and $l \in \mathcal{N}(i)$ **do**
  for all components $U_{i,k}$ of $\mathbf{U}_i$ **do**
    for all $l \in \mathcal{N}(i)$ **do**
      compute interpolation $D_{l,k}$ of
        $(\boldsymbol{\omega}_i, U_{i,k}), \big((\boldsymbol{\omega}_l, U_{l,k})\big)_{j\in\mathcal{N}(i)\setminus\{l\}}$
      $g \leftarrow \nabla D_{l,k} \cdot (\boldsymbol{\omega}_l - \boldsymbol{\omega}_i)$
      $d \leftarrow U_{l,k} - U_{i,k}$
      if $dg > 0$ and $|g| > |d|$ **then** $D_{l,k} \leftarrow \frac{d}{g} D_{l,k}$
      if $dg \leq 0$ **then** $D_{l,k} = 0$
    end for
    $\mathcal{L}_{i,k}(\mathbf{x}) \leftarrow U_{i,k} + \nabla D_{l_0,k} \cdot (\mathbf{x} - \boldsymbol{\omega}_i)$
      for $\big|\nabla D_{l_0,k}\big| = \max_{j\in\mathcal{N}(i)} \{\big|\nabla D_{l,k}\big|\}$
  end for
end for

6. Overview

MHD Scheme

The base scheme presented in Chapter 3 is not yet suitable for performing realistic simulations. On the one hand it cannot be used for simulations of a non–perfect gas. Simple tests also reveal stability problems caused, for example, by the fact that the divergence constraint (1.1e) was not taken into account during the construction of the finite–volume scheme. Therefore we present modifications of the base scheme in the following three chapters, which are essential for our simulations in the solar atmosphere. Since the problems and the solution strategies are more generally applicable, we do not focus solely on the astrophysical problems although these always serve as a motivation. All three modifications require only some pre– or postprocessing of the numerical flux function and are therefore easy to add to our base scheme. Large parts of Algorithm 1 on page 65 remain unchanged since all the modifications presented here can be implemented by minor changes in the numerical flux function $\mathbf{g}_{ij}$ (cf. Algorithm 2(b) on page 66).

That the modifications should only require minor modifications of the base scheme was one of the main aspects of their development. Furthermore, it was of great importance that the modifications should require hardly any additional cost in runtime; for example, the scheme must remain stable using the time step Δt defined by the base scheme. We always compare the modified scheme with the base scheme and, if available, also with other standard methods from the literature. The choice of these comparison schemes is always motivated by the two important aspects mentioned above: they have to be simple to add to the base scheme and should require hardly any additional CPU time.

In Chapter 7 we extend the energy relaxation method from [CP98] to the system of real gas MHD. In Chapter 8 we present a method that stabilizes the MHD solver against divergence errors. Finally in Chapter 9 we derive a method that can be used to compute solutions near an equilibrium state.

In Section 6.1 we give an overview of the central problems and discuss some approaches found in the literature. Before we derive the new methods in the following chapters, we describe the test cases that we use to verify the schemes in Section 6.2. Since we are interested in computing approximation errors, we study test cases for which we can construct exact solutions or for which we can at least use finely resolved 1d approximations as reference solutions.

6.1 Numerical Challenges

In the following we discuss a number of problems that we encountered during the development of our MHD code and its application to model problems from solar physics. These challenges serve as a motivation for the extensions of the base scheme described in the following chapters. Many of these challenges arise in a multitude of different applications and are not only relevant in the solar physical context; in many cases they are not even restricted to the MHD setting studied here. Therefore, many approaches for solving these problems can be found in the literature. We consequently also include a brief overview of some of the relevant literature.

6.1.1 Arbitrary Equation of State

As discussed in Chapter 1 the MHD system has to be closed by an equation that describes the relationship between the internal energy ε, the density ρ, and the pressure p. A great deal of different physical regimes can be modeled by means of this equation; for example, in the lower convection zone the plasma can be assumed to be a perfect polytrophic gas. This can be modeled by a very simple pressure law.

6.1 Definition (EOS for a Perfect Gas)

$$p(\rho,\varepsilon) = (\gamma-1)\rho\varepsilon \ , \ \theta(\rho,\varepsilon) = \frac{\varepsilon}{c_v} \tag{6.1}$$

with some constants $\gamma > 1, c_v > 0$.

Over the last few years many numerical flux functions have been derived for the MHD equations with a perfect gas law (cf. [DKRW01a, Wes02b] and the references therein). In many applications it is not feasible to assume a perfect gas law; in the following we will talk of a "*real gas*" in these cases. As already mentioned the plasma in the solar photosphere is assumed to be partially ionized. This leads to a far more complicated equation of state [Sch99b]. Therefore flux functions constructed for a perfect gas law are not applicable in this regime. In Chapter 7 we describe a possibility of extending any perfect gas flux function so that it can be used for solving the MHD equations for an arbitrary real gas.
The method we present in Chapter 7 is only one possibility of solving the MHD equation with an arbitrary EOS. One of the simplest numerical flux functions is the Lax–Friedrichs (LF) flux, which can be used without modification in the real gas setting. The idea is to use a central difference scheme and to add artificial viscosity for stabilization (cf. [Krö97, Example 2.2.6]). In 1d and in its simplest form the LF flux is

$$\mathbf{G}(\mathbf{U},\mathbf{V}) = \frac{1}{2}\big(F_1(\mathbf{U}) - F_1(\mathbf{V})\big) + \frac{\Delta x}{2\Delta t}\big(\mathbf{U}-\mathbf{V}\big) \ . \tag{LF}$$

Here Δx denotes the grid spacing. The simplicity of the scheme arises from the fact that only the analytical flux F_1 is required, whereas many other schemes require the eigensystem of the flux Jacobian. Since the flux evaluation only requires the knowledge of the pressure p as a function of ρ, ε and this is given by the EOS, the scheme can be

used without modification to solve the real gas MHD equations. On the other hand the Lax–Friedrich scheme shows a very poor resolution especially of contact discontinuities. We use this scheme only as a reference scheme since all our tests have shown that the error to runtime ratio is poor in comparison to other schemes [Wes02b, DW01]. Staggered central schemes present a further possibility of solving the real gas MHD equations. These schemes also require only flux evaluations and no additional information [NT90]. The drawback is that these schemes require the implementation of a dual grid, which can be very complicated for unstructured locally adapted grids — especially in 3d. For that reason we have refrained from testing these schemes. A comparison of some further Godunov type schemes can be found in [MYV88].
For the compressible Euler equations two methods of extending numerical fluxes for perfect gases to the real gas case have recently been suggested [CP98, BGH00]. In Chapter 7 we show how the *energy relaxation method* from [CP98] can be applied to the real gas MHD equations. This method allows the use of any existing solver without modification. Only a minor prestep is required. The method presented in [BGH00] exploits the fact that the structure of the eigensystem is independent of the EOS if quantities such as the pressure p and the sound speed c are assumed to be EOS dependent functions. Both methods were extended to the real gas MHD system and their efficiency compared in [DW01]. Both methods lead to very similar results. For the hydrodynamic case a comparison was recently published in [GHN02], where the authors arrived at the same conclusion. In Chapter 7 we will briefly cover the major results and add some further tests.
Since the partial ionization leads to a very complicated equation of state, we test our new method using a simpler setting. Real gases also arise, if — at low temperatures and at high pressures — intermolecular forces have to be taken into account. The van der Waals EOS is an extension of the perfect gas law, which can be used to model these molecular forces.

6.2 Definition (EOS for a van der Waals Gas)
The EOS of van der Waals is defined by

$$p(\rho,\varepsilon) = \rho\frac{R(\varepsilon+a\rho)}{c_v(1-b\rho)} - a\rho^2 \ , \ \theta(\rho,\varepsilon) = \frac{\varepsilon-\varepsilon_0+a\rho}{c_v} \ . \tag{6.2}$$

We use the same choice for the constants a, b, R, c_v, and ε_0 as in [BGH00]

$$a = 1684.54 \ , \ b = 0.001692 \ , \ R = 461.5 \ , \ c_v = 401.88 \text{ and } \varepsilon_0 = 0 \ .$$

6.3 Remark: *Note that for $\rho < \frac{1}{b} \approx 591$ we have $p > 0$ but for $\rho \to \frac{1}{b}$ we have $p \to \infty$. In contrast to the perfect gas EOS the MHD system* (1.1) *augmented by the van der Waals EOS is not hyperbolic for all $\mathbf{U}$ with $\rho > 0$ and $\varepsilon > 0$. The pressure law in* (6.2) *is still linear in ε with $\partial_\varepsilon p(\rho,\cdot) > 0$, we also have $c^2 > 0$ (cf. Assumption 1.3). As in the case of a perfect gas* (6.1)*, where we have $p = \rho R\theta$, we can also write the pressure law of a van der Waals gas as a function of the density and the temperature, which leads to $p(\rho,\theta) = \rho\frac{R\theta}{1-b\rho} - a\rho^2$. For ρ small the pressure is close to $\rho R\theta$, i.e., in this case we recover the perfect gas law.*

Although the van der Waals EOS is quite simple (for example, p is still given explicitly) the structure of the solution already includes the new features typical of a more

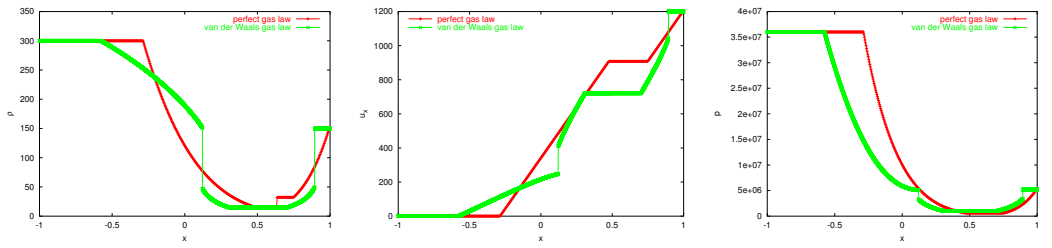


Figure 6.1: Quasi exact solutions to a Riemann problem for the Euler equations with two different equations of state. The solution in the perfect gas case was obtained using Chorin's method [Krö97]; the solution for the van der Waals gas was computed by the exact Riemann solver developed in [MV99]. In the van der Waals case the two acoustic waves are not simple waves anymore but compound waves consisting of a series of rarefactions and discontinuities.

complicated EOS. The example in Figure 6.1 demonstrates the influence of the EOS on the structure of the solution.
A second example which we use in our numerical tests is a pressure law that includes the vibrational motion of the oxygen and nitrogen in the air at very high temperatures. This setting was studied in [MS99].

6.4 Definition (EOS for a two molecule vibrating (tmv) Gas)
The tmv *pressure law is defined by the relations*

$$p(\rho,\varepsilon) = \rho r\theta \ , \ \varepsilon = c_v\theta + \frac{\alpha\Theta_{\text{vib}}}{\exp\left(\frac{\Theta_{\text{vib}}}{\theta}\right) - 1} \ . \tag{6.3}$$

For the constants we choose the values published in [MS99]:

$$r = 287.086 \ , \ c_v = 717.715 \ , \ \Theta_{\text{vib}} = 1000 \ , \ \textit{and}\ \alpha = 287.086 \ .$$

6.5 Remark: *Note that the pressure law has the same form as in the perfect gas case but that now the temperature is only given implicitly. For fixed ρ, ε we have to solve the algebraic relation defining the temperature using, for example, Newton iteration. This temperature is then used to compute the pressure. This is a setting similar to the Saha equations modeling the partial ionization of the solar photosphere (cf. Chapter 1). Thus this is a good EOS for studying the efficiency of a scheme in the case of a computationally expensive pressure law.*

6.1.2 Divergence Constraint

In Chapter 1 we saw that the divergence constraint on the magnetic field (1.1e) only has to be satisfied by the initial data and is then satisfied for all time. Therefore numerical methods are usually based only on the hyperbolic evolution equations (1.1a) — (1.1d). Due to the fact that the discrete divergence of the discrete curl is usually not exactly zero, $\nabla \cdot \mathbf{B}$–errors arise in numerical simulations; the approximation of the initial data can also introduce divergence errors. These error lead to an unphysical behavior of the system: Magnetic field lines may have wrong topologies leading to plasma transport

orthogonal to the magnetic field. This effect is discussed in [BB80, BS99]. These divergence error will often lead to the breakdown of the simulation due, for example, to non–physical states.
Schemes have been developed that imitate the analytical fact that the divergence of a curl equals zero. These schemes are often referred to as "constrained transport methods". This approach is used for the MHD equations in many different versions; some recent methods can be found, for example, in [DW98, RMJF98, BS99, LZ00, Tót00]. The main idea of the "constrained transport" approach is to use a special discretization of the magnetic field equations. This means that the underlying base scheme is only partially used and thus some of its desired properties may be lost. Moreover, these schemes are restricted to structured grids and require large stencils for the spatial discretization, cf. [Tót00, p. 646].
In the finite–volume approach each component of the curl of a vector field is interpreted as the divergence of a flux and integrated using Gauß' theorem. In many implementations a discrete divergence applied to a discrete curl will give zero only in an approximate way. Therefore, to prevent divergence errors from increasing with time, some correction technique has to be added to these schemes. A well–known correction method is the projection of the magnetic field into the space of divergence–free vector fields, also known as "Hodge projection". This method was implemented e.g. by Balsara [Bal98a, Bal98b], who discretized the Laplace operator in Fourier space.
In the finite–volume approach the numerical fluxes between adjacent grid cells are usually calculated by considering the 1d wave propagation in the normal direction to the element faces. In this one–dimensional setting, condition (1.1e) means that there is no jump in the normal component of the magnetic field across the interface. In multidimensional simulations this constraint cannot be generally fulfilled. Hence the one–dimensional wave considerations must allow for a jump in the normal component of the **B**–field. In the method developed by Powell et al. [BB80, Asl93, Pow94] the derivation of one–dimensional fluxes is based on the symmetrizable form of the MHD equations, which was, for example, introduced by Godunov in [God72]. In this form some additional terms that are not in divergence form are added to the MHD equations (1.8):

$$\partial_t \mathbf{U} + \nabla \cdot \mathbf{F}(\mathbf{U}) = \mathbf{q}(\mathbf{U}) + \mathbf{q}_{\mathrm{div}}(\mathbf{U}) \ .$$

The new "source terms" are proportional to $\nabla \cdot \mathbf{B}$

$$\mathbf{q}_{\mathrm{div}}(\mathbf{U}(\mathbf{x},t)) = \big(0, -\mathbf{B}(\mathbf{x},t), -\mathbf{u}(\mathbf{x},t), -\mathbf{u}(\mathbf{x},t)\cdot\mathbf{B}(\mathbf{x},t)\big)^T \nabla \cdot \mathbf{B}(\mathbf{x},t) \ .$$

In the original approach, a Roe–type solver for a modified system is used that admits jumps in the normal component of the magnetic field and advects them with the fluid velocity. Additionally the new terms are evaluated in each timestep. It was later discovered that the robustness of a MHD code can be improved just by adding these so–called "divergence source terms" to an arbitrary solver [TO96]. In the following we use this *source term fix* as a reference method for reducing divergence errors. The advantage of this method over many of the other ones mentioned above is that a given base scheme can be easily extended simply by adding a discretization for the term $\mathbf{q}_{\mathrm{div}}$.

We do this in a finite–volume spirit by discretizing $\int_{T_i} \mathbf{q}_{\text{div}}(\mathbf{U})$ (cf. (2.1)):

$$\begin{aligned}
&\int_{T_i} \mathbf{q}_{\text{div}}(\mathbf{U}(\cdot,t)) \\
&\approx \big(0, -\mathbf{B}_i(\boldsymbol{\omega}_i,t), -\mathbf{u}_i(\boldsymbol{\omega}_i,t), -\mathbf{u}_i(\boldsymbol{\omega}_i,t)\cdot\mathbf{B}_i(\boldsymbol{\omega}_i,t)\big)^T \int_{T_i} \nabla\cdot\mathbf{B}(\cdot,t) \\
&= \big(0, -\mathbf{B}_i(\boldsymbol{\omega}_i,t), -\mathbf{u}_i(\boldsymbol{\omega}_i,t), -\mathbf{u}_i(\boldsymbol{\omega}_i,t)\cdot\mathbf{B}_i(\boldsymbol{\omega}_i,t)\big)^T \sum_{j\in\mathcal{N}(i)} \int_{S_{ij}} \widehat{B}_{ij}(\cdot,t)\ .
\end{aligned}$$

With $\mathbf{B}_i, \mathbf{u}_i$ we denote the approximation on the element T_i. The scalar quantity $\widehat{B}_{ij}$ is a suitable approximation of the magnetic field in normal direction $(\mathbf{B}\cdot\mathbf{n}_{ij})$ on the face S_{ij}. In the case where the approximation is divergence–free, the values of $\mathbf{B}_i\cdot\mathbf{n}_{ij}$ and $\mathbf{B}_j\cdot\mathbf{n}_{ij}$ are the same on both sides of the face S_{ij}. In general this will not be the case and the approximations on the elements T_i and T_j will be discontinuous. Since a suitable value for $\widehat{B}_{ij}$ is thus not available, we use a simple average of the discrete values on both sides of S_{ij}:

$$\begin{aligned}
\int_{T_i} \mathbf{q}_{\text{div}}(\mathbf{U}(\cdot,t)) \approx \big(0, -\mathbf{B}_i(\boldsymbol{\omega}_i,t), -\mathbf{u}_i(\boldsymbol{\omega}_i,t), -\mathbf{u}_i(\boldsymbol{\omega}_i,t)\cdot\mathbf{B}_i(\boldsymbol{\omega}_i,t)\big)^T \\
\sum_{j\in\mathcal{N}(i)} |S_{ij}| \frac{1}{2}\big(\mathbf{B}_i(\mathbf{z}_{ij},t) + \mathbf{B}_j(\mathbf{z}_{ij},t)\big)\cdot\mathbf{n}_{ij}\ .
\end{aligned} \qquad \text{(source term fix)}$$

For the Maxwell equations Munz et al. [MSSV99, MOS$^+$00] introduced a technique to couple the divergence constraint for the electric field to the hyperbolic system. They called this modified system *Generalized Lagrange Multiplier* (GLM) formulation of the Maxwell equations. In [DKK$^+$02] we have derived the *GLM–MHD* equations following the same ideas. The modified divergence constraint can be chosen to be either an elliptic, a parabolic, or a hyperbolic equation. The most promising choice, especially in the finite–volume framework, seems to be a mixed hyperbolic/parabolic approach. In Chapter 8 we give a summary of [DKK$^+$02] and add some new results, including some test calculations also using the elliptic and the parabolic approach. The choice of the free parameters is also studied in more detail, taking into account analytical results for a model problem.

6.1.3 Balancing Source Terms and Flux Gradients

The initial conditions $\mathbf{U}_0$ to many problems, especially in the lower convection zone, are given as a local perturbation of a static, stratified, and purely hydrodynamic background atmosphere i.e.

$$\mathbf{U}_0 = \mathring{\mathbf{U}} + \widetilde{\mathbf{U}}$$

where $\mathring{\mathbf{U}}$ only depends on height z and $\widetilde{\mathbf{U}}$ has compact support. $\mathring{\mathbf{U}}$ is a solution to the stationary MHD equations and therefore has to satisfy

$$\partial_z \mathring{p} = \mathring{\rho} g. \tag{6.4}$$

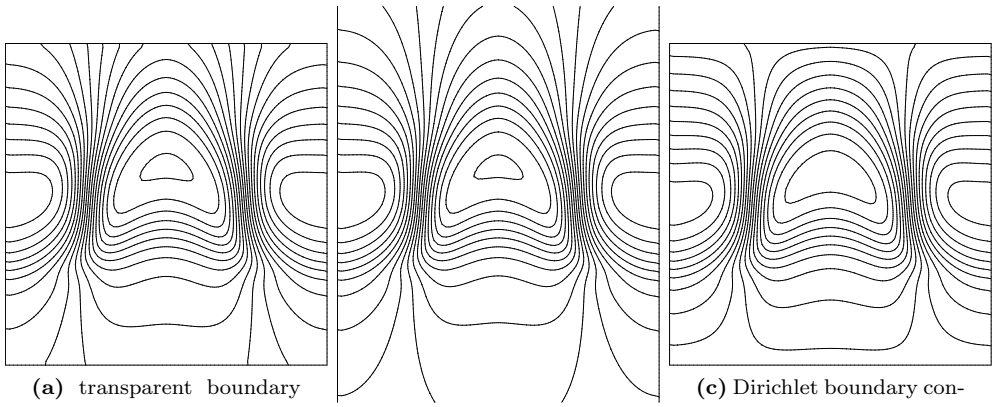
This equation follows from (1.1) if we set $\partial_t = 0, \partial_x = 0, \partial_y = 0$ and assume that the velocity and the magnetic field in the background atmosphere are zero. The force of gravity acts in the z–direction, leading to a gravity source term $\mathbf{g}(\mathbf{x}) = (0, 0, g(z))^T$. If due to approximation errors the identity (6.4) is violated, then the background atmosphere starts to shift, leading to errors that can severely influence the whole simulation. Since on the coarse grid used, for example, in 3d simulations these perturbations can easily be of the same magnitude as the difference between the background atmosphere $\mathring{\mathbf{U}}$ and the solution $\mathbf{U}$ itself, the structure of the solution can be totally lost. The unphysical perturbations not only lead to an unsatisfactory approximation but may also lead to stability and efficiency problems in the numerical scheme; for example, the oscillations cause local grid refinement in regions where the solution is smooth, so that a fine grid is used even in regions where the solution is equal to the background solution; furthermore, the limiter used in the higher order reconstruction reduces the method to first order in these regions since local maxima and minima are detected.
The problem of balancing flux gradients and source terms and of computing solutions close to a static state arises in many different fields, ranging from atmospheric flow, as considered here, to the approximation of the shallow water equations, where the source terms are used to model the ground topology. Many different approaches have been suggested in recent years for example in [LeV98, Gos00, Jin01, BPV03]. In Chapter 9 we suggest a simple modification of our base scheme that guarantees that a given background atmosphere remains static during a simulation and that can be used if the background solution $\mathring{\mathbf{U}}$ is known. In addition to its simple implementation the method leads to very good locally adapted grids — coarse elements are used in regions where the solution does not vary from $\mathring{\mathbf{U}}$. This allows us to use large domains for the simulations without investing a large amount of computational time in regions of the domain where the exact solution is very close to $\mathring{\mathbf{U}}$. Consequently, the problems arising from unphysical boundary conditions that are too close to the interesting structures in the solution can be reduced.

6.1.4 Open Boundaries

For our simulations in the lower solar convection zone the computational domain can only be a small section of the full convection zone. It has to be chosen with two aims in mind. On the one hand, small structures have to be resolved and their evolution has to be tracked over a long time period; on the other hand, the computational domain has to be as small as possible to minimize computational costs. Covering only a small portion of the full domain in a simulation requires the specification of suitable boundary conditions on the artificial vertical and horizontal boundaries to close the MHD system. It would be desirable if all boundaries — the vertical as well as the horizontal — were transparent for outgoing waves in the following sense:

6.6 Definition (Transparent Boundary Conditions)
Let Ω_1 and Ω_2 be two compact subsets of $\mathbb{R}^d$ with $\Omega_1 \subset \Omega_2$. Let $\mathbf{U}_i$ be the solution of the MHD equation on Ω_i using the boundary condition (BC) for $i = 1, 2$. Then the boundary condition (BC) is said to be transparent if $\mathbf{U}_1$ is identical to $\mathbf{U}_2$ in Ω_1.

In most solar physical simulations the important structures move upwards through the atmosphere and, therefore, the top boundary is the most critical one; the influence



(a) transparent boundary conditions

(b) reference solution on ten times larger domain

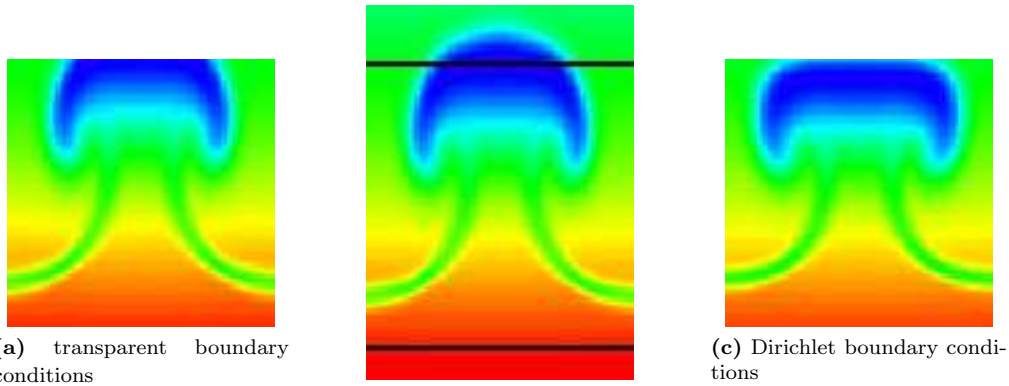
(c) Dirichlet boundary conditions

Figure 6.2: Isolines of the vertical velocity for a 2d atmospheric test case for $t = 25$. The main structure is still far away from the horizontal boundaries; but note that small perturbations have already reached the top and bottom boundaries via acoustic waves. Figure taken from [DKSW01b].

of the vertical boundaries is much smaller. For instance, in [EMI98] both vertical boundaries and the bottom boundary are assumed to be "closed lids". According to [NS90] the bottom boundary should also be transparent. Therefore we focus on both horizontal boundaries, while we assume periodic vertical boundaries in accordance with [NS90, CMIS95a, FZL98]. The influence of the top and bottom boundary on the solution is demonstrated in Figure 6.2 and Figure 6.3.

The conditions on the horizontal boundaries should lead to solutions that are (practically) independent of the height of the computational domain. Waves generated in the interior of the computational domain must be allowed to pass through the top and bottom boundary; i.e. an ideal artificial boundary should be transparent for outgoing perturbations. One method of achieving this is to absorb outgoing waves by introducing additional layers at the boundaries. (For solar physical simulations this method was used in [NS90, EMI98].) As far as we know there is neither an analytical argument nor a detailed numerical study that shows that this approach meets the stated requirements for a transparent boundary in the case of our application. However, the idea of absorbing layers seems to be a promising approach. This has recently been demonstrated for many different problems in the form of "perfectly matched layers", see e.g. [Ber94, AG98, TY98, Pet00].

In [DKSW01a, DKSW01b] we derived boundary conditions that fulfill the requirement stated in Definition 6.6 at least for small perturbations. Our method of formulating non–reflecting boundary conditions belongs to the class of so–called exact boundary conditions, cf. the reviews [Giv91, Tsy98]. It follows the technique presented in [Sof98]. Our method is based on the derivation of an analytically exact boundary condition for the hyperbolic equation describing the evolution of the pressure perturbation. The condition necessarily includes a *non–local* convolution term with respect to time at the artificial boundaries. However, by using a special approximation of the convolution



(a) transparent boundary conditions

(b) reference solution on ten times larger domain

(c) Dirichlet boundary conditions

Figure 6.3: Density for a 2d atmospheric test case for $t = 35$. The main structure is starting to move through the upper boundary. The result using our transparent boundary conditions is still very close to the reference solutions. Figure taken from [DKSW01b].

kernel, this non–local term can be evaluated in a time–stepping manner so that the numerical method stays *local* with respect to time.
A necessary first step in the derivation of our boundary conditions is the linearization of the MHD equations about a stratified background atmosphere $\mathring{\mathbf{U}}$ (cf. Section 6.1.3). We assume that the perturbations at the boundary are sufficiently small and smooth. Furthermore, we study the special case of an exponentially decaying atmosphere that permits a sufficiently far–reaching analytical study. At the same time, the application of our boundary conditions to other models for the background atmosphere inside the computational domain seems to pose no problems. For the linearized system we can prove the transparency of our boundary conditions in the sense of Definition 6.6. In the following Theorem we concentrate on the top boundary. A similar result also holds for the conditions on the lower boundary, but due to the unboundedness of the background solution the formulation of the theorem is slightly more complicated.

6.7 Theorem (Transparent Boundary Conditions)
Let the computation domain be $\Omega = [x_l, x_r] \times [-\infty, z_t]$ *with* $x_l, x_r, z_t \in \mathbb{R}$ *and* $x_l < x_r$. *The initial conditions are* $\mathbf{U}_0(x,z) = \widetilde{\mathbf{U}}_0(x,z) + \mathring{\mathbf{U}}(z)$ *with* $\mathring{p}'(z) = g(z)\mathring{\rho}(z)$, $\mathring{\mathbf{u}} \equiv 0$, *and* $\mathring{\mathbf{B}} \equiv 0$. *The background atmosphere is, furthermore, assumed to satisfy* $\mathring{\rho}(z)^{\gamma-1} = \frac{\gamma-1}{a\gamma}\exp(-2\alpha z)$ *and* $\mathring{p}(z) = \mathring{\rho}(z)^\gamma$ *with some constants* $\gamma > 1$, $\alpha > 0$ *and* $a > 0$. *We assume that the perturbations are compactly supported in* Ω, *i.e., we assume that there exists a* $z_b < z_t$ *so that the support of* $\widetilde{\mathbf{U}}_0$ *is in* $[x_l, x_r] \times [z_b, z_t]$. *Define* $\Omega_\infty := [x_l, x_r] \times \mathbb{R}$ *and set* $\widetilde{\mathbf{U}}(z) = 0$ *for* $z < z_b$ *and* $z > z_t$. *We consider the following two problems:*

(A) Consider the MHD equations linearized around $\mathring{\mathbf{U}}$ *in* Ω_∞ *with initial conditions* $\widetilde{\mathbf{U}}_0$, $|\widetilde{\mathbf{U}}| = 0$ *for* $|z| \to \infty$, *and periodic boundary conditions in* x.

(B) Consider the MHD equations linearized around $\mathring{\mathbf{U}}$ *in* Ω *with initial conditions* $\widetilde{\mathbf{U}}_0$, $|\widetilde{\mathbf{U}}| = 0$ *for* $z \to -\infty$, *and periodic boundary conditions in* x. *At the boundary* $z = z_t$ *we prescribe our transparent boundary conditions.*

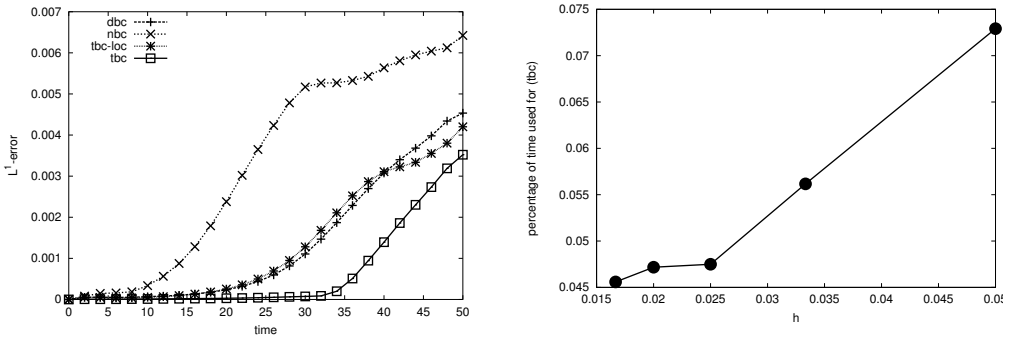


Figure 6.4: Evolution of L^1–error for different boundary conditions. At around $t = 30$ the large perturbations reach the upper boundary. Up to this time the transparent boundary conditions lead to almost no error compared to the reference solution. Time used for a 2D–calculation for different grid sizes in comparison to the time required for the calculation of the boundary values using our boundary condition (taken from [DKSW01b]).

Then the following two statements hold

(i): Any solution to problem (A) is a solution to problem (B).

(ii): Consider a solution $\mathbf{U}$ *to problem (B), which is continuously differentiable up to* $z = z_t$*. Then there exists a solution to problem (A) that coincides with* $\mathbf{U}$ *in* Ω*.*

The proof can be found in [DKSW01b].
In [DKSW01b] we also discuss implementational aspects and compare our boundary conditions with other more direct approaches. Our numerical examples illustrate that the structure of the solution is considerably influenced by the choice of the boundary conditions. Moreover, using our boundary conditions we find that even large perturbations are hardly reflected at the artificial boundaries. The examples indicate that the proposed transparent boundary conditions yield good results (Figure 6.4) and are very cheap with respect to their computational costs. In fact, the costs for the numerical evaluation of the boundary conditions are almost negligible: in a 2d test calculation it took less than 6% of the overall CPU time, see Figure 6.4.

6.2 Constructing Solutions

6.2.1 The Riemann Problem

The most thoroughly studied initial value problem for hyperbolic conservation laws is the one dimensional Riemann problem. The initial conditions for this Cauchy problem are defined by a left and right hand state $\mathbf{U}_l, \mathbf{U}_r$:

$$\mathbf{U}_0(x) = \begin{cases} \mathbf{U}_l & \text{for } x < 0 \ , \\ \mathbf{U}_r & \text{for } x > 0 \ , \end{cases} \quad (x \in \mathbb{R}) \ .$$

In the context of finite–volume schemes this problem is of special interest because numerical flux functions can be constructed from a detailed knowledge of the solution

to the Riemann problem; it can furthermore be used as the main building block for existence results, even for systems of hyperbolic equations in one space dimension (see, for example, [Daf00]). The most important features of the solution $\mathbf{U}$ to the Riemann problem is its self–similarity: $\mathbf{U}(x,t) = \widehat{\mathbf{U}}(x/t)$. If a system of conservation law of dimension m is strictly hyperbolic (cf. Definition 1.1), then, by the Theorem of Lax and Liu, the function $\widehat{\mathbf{U}}$ consists of $m+1$ constant states $\mathbf{U}_0 = \mathbf{U}_l, \mathbf{U}_1, \ldots, \mathbf{U}_m, \mathbf{U}_{m+1} = \mathbf{U}_r$ connected either by a smooth transition (rarefaction wave) or by a discontinuity (shock or contact discontinuity) or by a combination of smooth transitions and discontinuities (compound wave).
For the Euler equation of gas dynamics with a perfect gas law, Chorin's method can be used to construct a solution to the Riemann problem. The method only requires the solution of a simple ODE. Therefore, the Riemann problem for arbitrary left and right hand states — at least if they are sufficiently close together — can be solved up to any given accuracy. The Riemann problem is thus also very well suited to verify numerical schemes. For more complicated equations of state the algorithm for constructing the solution has recently been implemented in [MV99]. For the MHD equations — even in the case of a perfect gas law — the construction of the solution is not yet available for general left and right hand states.
We also study a Riemann problem in two space dimensions that consists of four constant initial states:

$$\mathbf{U}_0(x,y) = \begin{cases} \mathbf{U}_1 & \text{for } x > 0, y > 0 \ , \\ \mathbf{U}_2 & \text{for } x < 0, y > 0 \ , \\ \mathbf{U}_3 & \text{for } x < 0, y < 0 \ , \\ \mathbf{U}_4 & \text{for } x > 0, y < 0 \ , \end{cases} \quad ((x,y) \in \mathbb{R}^2) \ .$$

The solution to this problem is far more complicated than in the 1d case, and to our knowledge the construction of the solution for arbitrary states is not yet possible even for simple systems. Some numerical investigations for the Euler equations can be found, for example, in [LL98]. We use this problem to verify our scheme for two reasons. On the one hand, the solution is intrinsically two dimensional, so that problems arising only in 2d simulations can be studied. On the other hand the finite speed of propagations means that at least at some distance from the origin the solution is given by the solution to the one dimensional Riemann problems between two of the four states. To be more precise, there exist $x_l, x_r, y_b, y_t \in \mathbb{R}$ for any given time t so that for $y > y_t$ the solution is given by the Riemann problem between $\mathbf{U}_2, \mathbf{U}_1$, for $x < x_l$ by the Riemann problem between $\mathbf{U}_3, \mathbf{U}_2$. Similarly, the solution for $y < y_b$ and for $x > x_r$ can be obtained by studying the 1d Riemann problems between $\mathbf{U}_3, \mathbf{U}_4$ and $\mathbf{U}_4, \mathbf{U}_1$, respectively. A sketch of the structure of the solution both to the 1d Riemann problem and the 2d Riemann problem is shown in Figure 6.5.

6.2.2 The Rotation Problem

The following problem is an extension of a purely hydrodynamic problem suggested by Tim Kröger [Krö02]. We added a stabilizing magnetic field and extended the setting to include general equations of state and non–constant density profiles. This problem has

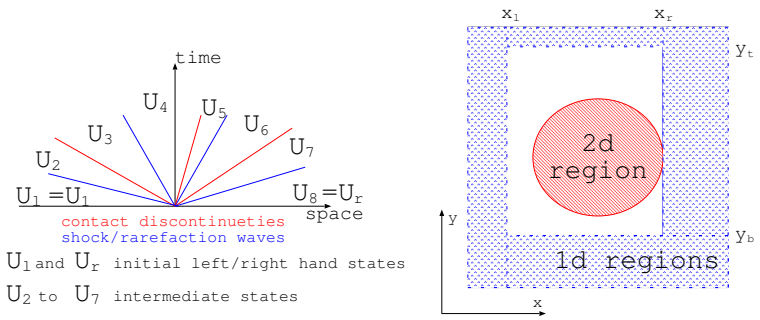


Figure 6.5: Sketch of the structure of the (MHD) Riemann problem in 1d (left) and 2d (right). In the 1d setting we show the evolution of the self–similar solution in the x/t plane. The solution consists of seven waves connecting the two original states. In the 2d setting we show the structure of the solution at a time T. Due to the finite speed of propagation the 2d structure originating from the point where initially all four states meet has not influenced the solution on the whole domain. Away from the origin, the solution is still given by the four 1d Riemann problems defined by the four original states.

a stationary weak solution where outside the ball $B_R(0)$ the fluid is at rest. Inside $B_R(0)$ the motion of the fluid is always tangential to the radial direction: $\mathbf{u} = u_0(-y, x, 0)^T$. Therefore the solution is discontinuous on $\partial B_R(0)$. The pressure and the density are chosen in such a way that a gravitational source term is balanced. The gravitational force always points to the origin: $\mathbf{g} = -g(x, y, 0)^T$ with some constant $g \geq 0$. If $g = 0$ then the density is constant. For $g > 0$ the density is a function of the radius. The gas pressure is then defined in such a way that we obtain a stationary weak solution of the MHD system (1.1) in two space dimensions. A sketch of the structure of the solution is shown in Figure 6.6.

Theorem: *Let functions $\mathring{\rho} = \mathring{\rho}(z)$ and $\mathring{p} = \mathring{p}(z)$ be given which satisfy*

$$\frac{d}{dz}\mathring{p}(z) = \mathring{\rho}(z) \ , \qquad \mathring{\rho}(z) > 0 \ .$$

With the constants $R > 0$, $g \geq 0$, $u_0 > 0$, $B_0 \geq 0$ we define for $(x, y) \in \Omega$ and $r = \sqrt{x^2 + y^2}$

$$\rho(x, y) = \mathring{\rho}(r^2) \ ,$$

$$\mathbf{u}(x, y) = \begin{cases} \big(u_0 y, -u_0 x, 0\big)^T & \text{for } r < R \ , \\ \big(0, 0, 0\big)^T & \text{for } r > R \ , \end{cases}$$

$$\mathbf{B}(x, y) = \begin{cases} \big(B_0 y, -B_0 x, 0\big)^T & \text{for } r < R \ , \\ \big(0, 0, B_0 R\big)^T & \text{for } r > R \ , \end{cases}$$

$$p(x, y) = \begin{cases} p_0 + \frac{1}{2}u_0^2\big(\mathring{p}(r^2) - \mathring{p}(R^2)\big) - \frac{B_0^2}{4\pi}\big(r^2 - R^2\big) - g\mathring{p}(r^2) & \text{for } r < R \ , \\ p_0 - g\mathring{p}(r^2) & \text{for } r > R \ . \end{cases}$$

The constant p_0 is fixed so that $p(x, y) > 0$ for all (x, y) in the computational domain Ω. We define the vector of conservative variables using an arbitrary EOS to define the

internal energy (cf. Assumption 1.3)

$$\mathbf{U} = \left(\rho, \rho\mathbf{u}, \mathbf{B}, \varepsilon(\rho, p) + \frac{1}{2}\rho\mathbf{u}^2 + \frac{\mathbf{B}^2}{8\pi}\right)^T .$$

Then $\mathbf{U}(x, y)$ *is a stationary weak solution to the MHD equations* (1.1) *in* $\mathbb{R}^2$ *with the gravity force vector* $\mathbf{g}$ *given by*

$$\mathbf{g}(x, y) = (-2gx, -2gy, 0) .$$

Proof:
To show that $\mathbf{U}$ is a stationary solution we first have to verify that the flux gradient $\nabla \cdot \mathbf{F}(\mathbf{U})$ is balanced by the gravity source term $(0, \rho\mathbf{g}, \mathbf{0}, \rho\mathbf{u} \cdot \mathbf{g})^T$ in the regions of $\mathbb{R}^2$ where $\mathbf{U}$ is smooth.

$r < R$: It is easy to see that the flux in the equation for the density given by $\nabla \cdot (\rho\mathbf{u})$ vanishes identically. Next we compute the flux for the first component of the moment; the same analysis can be used for the second component by noting that the roles of x and y can be directly exchanged.

$$\begin{aligned}
&\partial_x\left(\rho u_x^2 + p + \frac{\mathbf{B}^2}{8\pi} - \frac{B_x^2}{4\pi}\right) + \partial_y\left(\rho u_x u_y - \frac{B_x B_y}{4\pi}\right) \\
&= 2u_0^2 x y^2 \mathring{\rho}' + u_0^2 x \mathring{p}' - \frac{B_0^2}{2\pi}x - 2gx\mathring{p}' + \frac{B_0^2}{4\pi}x - u_0^2 x(2y^2\mathring{\rho}' + \mathring{\rho}) + \frac{B_0^2}{4\pi}x \\
&= -2gx\mathring{p}' + u_0^2 x(\mathring{p}' - \mathring{\rho}) \\
&= -2gx\mathring{\rho} .
\end{aligned}$$

The last equation follows since we have assumed that $\mathring{p}' = \mathring{\rho}$. We also have $\mathring{\rho} = \rho$ so that the remaining expression is equal to the gravity source term. The arguments for u_z, B_x, B_y, and B_z are quite straightforward so that only the result for the equation for the total energy remains to be shown. With the observation that $\zeta = (\rho e + p + \frac{B^2}{8\pi}$ is a function of r^2 the result easily follows:

$$\begin{aligned}
&\partial_x\Big(u_x\zeta - B_x\mathbf{u}\cdot\mathbf{B}\Big) + \partial_y\Big(u_y\zeta - B_y\mathbf{u}\cdot\mathbf{B}\Big) \\
&= 2u_0 xy\zeta' - 2u_0 B_0 xy - 2u_0 xy\zeta' + 2u_0 B_0 xy \\
&= 0 .
\end{aligned}$$

Since $\mathbf{u} \perp \mathbf{g}$ we also have $\rho\mathbf{u} \cdot \mathbf{g} = 0$. Thus concludes the proof for $r < R$.

$r > R$: Again we can concentrate on the equations for u_x and ρe:

$$\partial_x\left(\rho u_x^2 + p + \frac{\mathbf{B}^2}{8\pi} - \frac{B_x^2}{4\pi}\right) + \partial_y\left(\rho u_x u_y - \frac{B_x B_y}{4\pi}\right) = -2xg\mathring{p}'$$

and

$$\partial_x\Big(u_x\zeta - B_x\mathbf{u}\cdot\mathbf{B}\Big) + \partial_y\Big(u_y\zeta - B_y\mathbf{u}\cdot\mathbf{B}\Big) = 0 .$$

Thus both equations lead to the same result as in the previous case and the same arguments hold.

To conclude our proof we show that the discontinuity at $r = R$ is a contact discontinuity. First we verify that the Ranking–Hugoniot relations are satisfied. Since the source term is continuous in the whole domain, we can study the jump relations as for the homogeneous system. Let (x_0, y_0) be a point of $\partial B_R(0)$. We have to rotate the limits of the solution in (x_0, y_0) in the direction $\mathbf{n} = (x_0, y_0)/R$. In primitive variables we compute

$$\mathbf{V}_I = (\mathring{\rho}(R^2), 0, u_0, 0, 0, B_0, 0, p_0 - g\mathring{p}(R^2))^T \ ,$$
$$\mathbf{V}_O = (\mathring{\rho}(R^2), 0, 0, 0, 0, 0, B_0, p_0 - g\mathring{p}(R^2))^T \ .$$

It is easy to see that these states satisfy the Ranking–Hugoniot jump relation with the speed $s = 0$. By computing the wave speeds for these two states, we find using the notation of (1.10)

$$c_a = 0 \ ,$$
$$c_s^2 = \frac{1}{2}\left(c^2 + \frac{B_0^2}{\mathring{\rho}(R)} - \sqrt{\left(c^2 + \frac{B_0^2}{\mathring{\rho}(R)}\right)^2}\right) = 0 \ ,$$
$$c_f^2 = \frac{1}{2}\left(c^2 + \frac{B_0^2}{\mathring{\rho}(R)} + \sqrt{\left(c^2 + \frac{B_0^2}{\mathring{\rho}(R)}\right)^2}\right) = \left(c^2 + \frac{B_0^2}{\mathring{\rho}(R)}\right)^2 > 0 \ .$$

since $b_1^2 = 0$. These relations hold for both the inner state $\mathbf{V}_I$ and the outer state $\mathbf{V}_O$. Five of the seven eigenvalues (all except the fast waves) are identical to the shock speed $s = 0$. Therefore the discontinuity is a contact discontinuity. Since the magnetic field vector is rotated but has the the same length on both sides, this contact is an Alfven wave; in the hydrodynamic case $B_0 = 0$ it reduces to a simple entropy wave. □

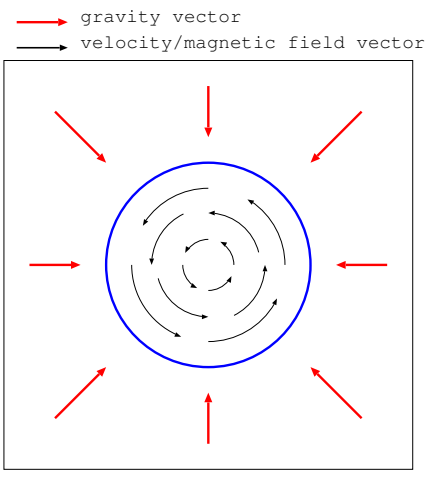


Figure 6.6: This sketch shows the structure of the solution to the rotation problem described in Section 6.2.2. In the interior of the blue circle the velocity **u** and the magnetic field **B** are "twisted" so that they are always orthogonal to the radial direction. Outside the circle the velocity is zero and the magnetic field is constant and points into the plane. The density ρ and the pressure p depend on the radial direction and are balanced with the force of gravity **g**, which always points to the origin.
In the case where the force of gravity is zero, the density ρ is constant and the hydrodynamic pressure p is chosen so that the total pressure $p + \frac{1}{8\pi}|\mathbf{B}|^2$ is constant.

6.2.3 Advection Problem in B_z

As in the previous case the following class of problems are built on top of a background solution $\mathring{\rho}, \mathring{p}$, which is used to balance some source terms. In this setting the z compo-

nent of the magnetic field (B_z) is advected; in all the other components the solution does not depend on time. Compared to the previous setting the additional difficulty is the moving structure in B_z which require dynamic grid refinement and coarsening. A sketch of the structure of the solution is shown in Figure 6.7.

Theorem: *We assume in the following a perfect gas law with $\gamma = 2$ and that u_0, g are given constants. Let a piecewise smooth function $\mathring{B}_z = \mathring{B}_z(x, y)$ be given, then*

$$\mathbf{U}(x, y, t) = \Big(\mathring{\rho}(y), u_0, 0, 0, 0, 0, \mathring{B}_z(x - u_0 t, y), \mathring{p}(y) + \frac{u_0^2}{2}\mathring{\rho}(y)\Big)^T$$

is a stationary solution to the MHD equation in 2d with gravitational source term $\mathbf{g} = (0, g, 0)$ if $\mathring{\rho}$ and $\mathring{p}$ satisfy the following relations

$$\begin{aligned} \mathring{p}(y)' &= g\mathring{\rho} \ , \\ \mathring{\rho}(y) &> 0 \ , \\ \mathring{p}(y) &> \sup_x \frac{B_z^2(x, y)}{8\pi} \ . \end{aligned}$$

Proof:
The two inequalities on $\mathring{\rho}, \mathring{p}$ guarantee that the density and the gas pressure are positive so that $\mathbf{U}$ is in the physical regime. The ODE for $\mathring{p}$ leads to a balancing of the pressure gradient with the gravitational source term. The proof that $\mathbf{U}$ is a solution is straightforward and we only show the necessary computation for B_z. For simplicity we assume that B_z is differentiable

$$\partial_t B_z + \partial_x\big(u_x B_z - B_x u_z\big) + \partial_y\big(u_y B_z - B_y u_z\big) = -u_0 \partial_x \mathring{B}_z + u_0 \partial_x \mathring{B}_z = 0 \ ,$$

The proof that discontinues B_z is quite simple. Assume that $\mathbf{U}_L, \mathbf{U}_R$ are two conservative states that are identical up to the value in B_z. Then the wave speeds differ only in c_f. Both c_a and c_s are zero for $\mathbf{U}_L$ and $\mathbf{U}_R$. The Ranking–Hugoniot relations are satisfied for $s = u_0$. Therefore we have $\lambda_i = s$ for $i = 2, \ldots, 6$ and a discontinuity in B_z is a contact discontinuity. □

Remark: *As we have already pointed out, the exact solution does not depend on time in the conservative variables $\rho, \mathbf{u}, B_x, B_y$, and ρe. The initial condition in B_z given by the function $\mathring{B}_z$ are transport with velocity u_0 normal to the force of gravity. For $B_z = 0$ the gas pressure is given by $\mathring{p}$. In the other regions the gas pressure $p(x, y, t) = \mathring{p}(y) - \frac{B_z^2(x-u_0 t, y)}{8\pi}$ is reduced so that in this case the total pressure is equal to $\mathring{p}$.*

6.3 Test Cases

For some of the settings described in the previous section we can define an arbitrary one dimensional *stratified background atmosphere* given by two functions $\mathring{\rho}(z)$ and $\mathring{p}(z)$ defining the density and the pressure as a function of the height in the atmosphere. The main condition is that $\mathring{p}' = \overline{g}\mathring{\rho}$ for some constant $\overline{g}$. This condition guarantees that the pressure gradient balances the gravitational source term. With our applications in mind we use a *model solar atmosphere*:

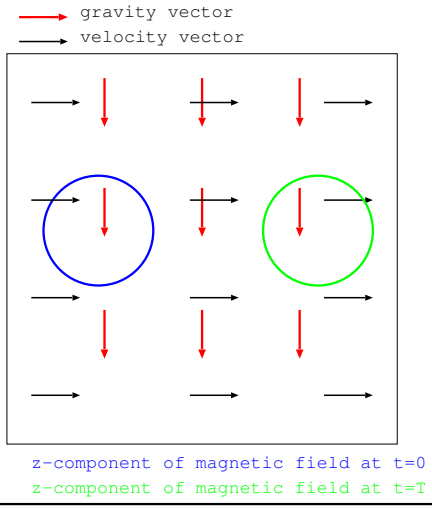


Figure 6.7: This sketch shows the structure of the solution to the advection problem described in Section 6.2.3. The velocity $\mathbf{u} = (u_0, 0, 0)$ is constant and is orthogonal to the force of gravity $\mathbf{g} = (0, g, 0)$. The density ρ and the pressure p depend only on y and are balanced with the force of gravity $\mathbf{g}$. The magnetic field is zero in the x and y component. Only B_z is non–zero (for example, B_z is initially non–zero in the blue circle) and is advected with the fluid (during the simulation the non–zero region moves to the right, as indicated by the green circle).

Problem 6.1(ATM) Stratified Atmosphere
This is a model for the solar atmosphere in the lower convection zone (see, for example, [CMIS95b]). Since in this region effects like the partial ionization of the plasma are negligible, we can use a perfect gas law with $\gamma > 1$. We define the density, the pressure, and the gravitation constant

$$\mathring{\rho}_{sun}(z) = (1 - 0.32(z+1))^{\frac{1}{\gamma-1}} ,$$
$$\mathring{p}_{sun}(z) = (1 - 0.32(z+1))^{\frac{\gamma}{\gamma-1}} ,$$
$$g_{sun} = -0.32\frac{\gamma}{\gamma-1} .$$

The variable z denotes the height in the atmosphere. Together with $\mathbf{u} \equiv 0$ and $\mathbf{B} \equiv 0$ we arrive at a stationary solution to the 1d MHD equations with gravity source term since we have $\frac{d}{dz}\mathring{p}_{sun}(z) = -g_{sun}\mathring{\rho}_{sun}$.

The next three test cases are planar Riemann problems:

Problem 6.2(RPDWT) Dai–Woodward–Tóth 1d Riemann Problem
This is a Riemann problem for the perfect gas MHD equations which was suggested by Dai and Woodward in [DW94] and which is also considered by Tóth [Tót00].

Equation of state: perfect gas with $\gamma = 5.0/3.0$
computational domain: $[-0.5, 0.5]$ and $T = 0.08$
boundaries: Dirichlet data given by the initial data on the boundaries

	ρ	u_x	u_y	u_z	B_x	B_y	B_z	p
$\mathbf{U}_l$:	1	10	0	0	5	5	0	2
$\mathbf{U}_r$:	1	−10	0	0	5	5	0	1

Problem 6.3(RPWAALS1) 1d van der Waals Riemann Problem
This is a Riemann problem for the Euler equations which we use to verify our real gas solver together with a 1d version of our finite–volume code. The initial conditions

were constructed by the authors of [MV99] who also supplied us with the exact solution. The solution shows the maximum number of compound waves, which are the main new feature when switching from a perfect gas law to a more complicated EOS.

Equation of state: van der Waals (cf. (6.2)*)*
computational domain: $[-1,1]$ *and* $T = 0.0009$
boundaries: Dirichlet data given by the initial data on the boundaries

	ρ	u_x	u_y	u_z	B_x	B_y	B_z	p
$\mathbf{U}_l$*:*	333	0	0	0	0	0	0	$3.6e7$
$\mathbf{U}_r$*:*	150	1200	0	0	0	0	0	$5.2e6$

Problem 6.4(RPWALLS2**) 1d van der Waals Riemann Problem**
This Riemann problem was suggested for the real gas Euler equations in [BGH00]. For this problem we again have a quasi exact reference solution that we can use to verify our code.

Equation of state: van der Waals (cf. (6.2)*)*
computational domain: $[-1,1]$ *and* $T = 0.0005$
boundaries: Dirichlet data given by the initial data on the boundaries

	ρ	u_x	u_y	u_z	B_x	B_y	B_z	p
$\mathbf{U}_l$*:*	333	0	0	0	0	0	0	$3.7311358e7$
$\mathbf{U}_r$*:*	111	0	0	0	0	0	0	$2.1770768e7$

So far we have described problems for the Euler equations of gas dynamics since the magnetic field in the initial conditions is zero. In all the following problems the magnetic field is non–zero.

Problem 6.5(RPtmv2d) 2d Riemann Problem
As we have already pointed out, the solution is only available on parts of the domain. Even in these regions we can only use a high resolution 1d approximation to the corresponding Riemann problems as reference solution.

Equation of state: tmv (cf. (6.3)*)*
computational domain: $[-1,1] \times [-1,1]$ *and* $T = 0.0004$
boundaries: Dirichlet data using finely resolved 1d solution of the corresponding 1d Riemann problems

	ρ	u_x	u_y	u_z	B_x	B_y	B_z	p
$\mathbf{U}_1$*:*	200	-700	0	-200	50000	50000	-20000	$1e8$
$\mathbf{U}_2$*:*	200	700	-500	200	50000	50000	10000	$1e8$
$\mathbf{U}_3$*:*	500	150	0	0	50000	50000	-5000	$5e8$
$\mathbf{U}_4$*:*	150	0	-200	200	50000	50000	10000	$6e7$

For the adaptation process we choose $\bar{u} = 1000$ *and* $\bar{B} = 100000$ *in* (3.19).

Problem 6.6(RPTMV2D(I,J)) **1d Riemann Problems based on RPtmv2d**

The initial data for this 1d Riemann problem is given by two of the four states of the 2d Riemann problem 6.5(RPtmv2d). The left hand state is $\mathbf{U}_i$ *and the right hand state is* $\mathbf{U}_j$ *for* $1 \leq i, j \leq 4$*. In our tests we use a numerical solution computed on 15000 grid points as a reference solution.*

We now define two rotation problems. To define these problems uniquely, we have to fix the constants $R > 0$, $g \geq 0$, $u_0 > 0$, $B_0 \geq 0$, and p_0 together with the two functions $\mathring{\rho}, \mathring{p}$.

Problem 6.7(ROTCONST) **Constant Rotation Problem**

We start out with a setting where the gravity source term vanishes and where the density is constant everywhere. We also choose a constant density $\rho_0 > 0$*.*

Equation of state: van der Waals (cf. (6.2)*)*
computational domain: $[-2, 2]^2$ *and* $T = 0.005$
boundaries: periodic boundary conditions

R	g	u_0	p_0	$\mathring{\rho}(z)$	$\mathring{p}(z)$
1	0	1000	$1e08$	100	$100z$

We choose the constant B_0 *as a function of a parameter* β *as follows:*

$$B_0 = \beta \sqrt{2\pi u_0^2 \frac{|\mathring{p}(R^2)|}{R^2}} \ .$$

For $\beta = 1$ *this leads to a constant pressure that is equal to* p_0*. In our simulations we use* $\beta \in \{0, \frac{1}{2}, 1, 2\}$*. For the adaptation process we choose* $\bar{u} = 5000$ *and* $\bar{B} = 10000$ *in* (3.19)*.*

Problem 6.8(ROTATM) **Atmosphere Rotation Problem**

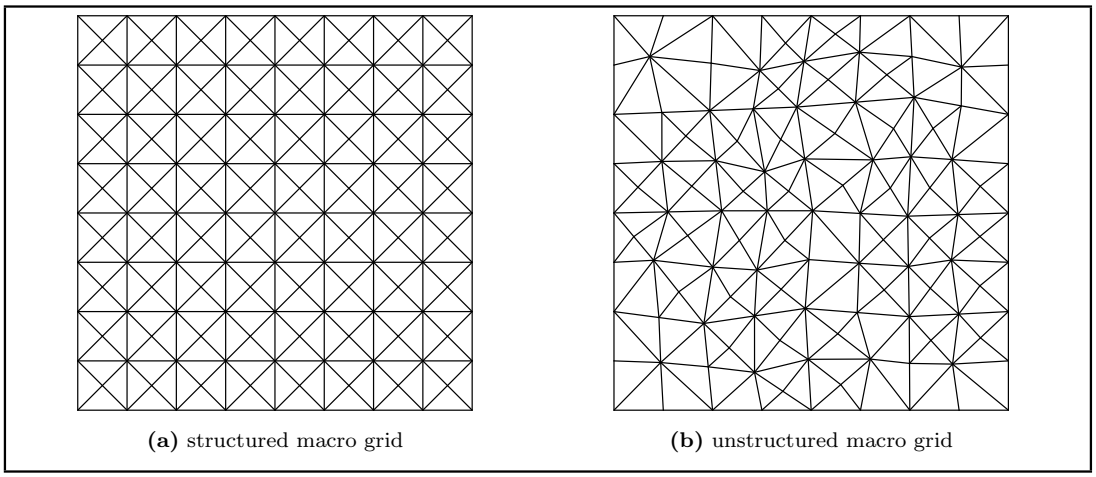
For this setting we use the model solar atmosphere as background solution:

Equation of state: perfect gas with $\gamma = 1.4$
computational domain: $[-2, 2]^2$ *and* $T = 5$
boundaries: Dirichlet data given by the initial data on the boundaries

R	g	u_0	p_0	$\mathring{\rho}(z)$	$\mathring{p}(z)$
1	1	0.1	0.01	$g_{sun}\, \mathring{\rho}_{sun}(z-6)$	$-\mathring{p}_{sun}(z-6)$

Again we choose B_0 *as in the previous case but with a fixed* $\beta = 0.1$*. For the adaptation process we choose* $\bar{u} = 1$ *and* $\bar{B} = 1$ *in* (3.19)*.*

Problem 6.9(AdvAtm) Atmosphere Advection Problem



(a) structured macro grid **(b)** unstructured macro grid

Equation of state: perfect gas with $\gamma = 2$

computational domain: $[-1,1]^2$ *and* $T = 4$

boundaries: Dirichlet data given by the initial data on the top and bottom, periodic on left and right

g	u_0	$\mathring{\rho}(z)$	$\mathring{p}(z)$	$\mathring{B}_z(x,y)$
g_{sun}	0.5	$\mathring{\rho}_{sun}(z)$	$\mathring{p}_{sun}(z)$	$\begin{cases} 4096.0(x^2+y^2)^4 - 128(x^2+y^2)^2 + 1 & x^2+y^2 \leq \frac{1}{8}\ , \\ 0 & \textit{otherwise.} \end{cases}$

For the adaptation process we choose $\bar{u} = 1$ *and* $\bar{B} = 1$ *in* (3.19).

Remark: *For all problems we use the exact solution (at least up to a high accuracy) as boundary conditions; therefore we have no problems with artificial boundary conditions.*

To use the finite–volume scheme presented in Chapter 3 we have to fix the macro grid and some parameters. If not noted otherwise, we use the same set of parameters in all calculations. To define the time step we use $c_{\text{cfl}} = 0.4$ (cf. (3.11)). For the adaptation process we have to choose values for $h_{\min}, \text{ref}_{\text{limit}}, \text{crs}_{\text{limit}}$ (cf. Section 3.5):

$$h_{\min} = 0.005\ , \quad \text{ref}_{\text{limit}} = 0.1\ , \quad \text{crs}_{\text{limit}} = 0.05\ .$$

As discussed in Section 2.2 the computational grid is constructed from a macro triangulation of the computational domain Ω. Since all the test cases presented so far are defined on a rectangular domain $[-a,a]^2$, we can use a scaled version of one fixed macro grid. To reduce effects due to grid alignment like, for example, superconvergence and cancellation of errors we use an unstructured macro grid (cf. Figure 6.3). It is constructed by a random perturbation of a structured grid.
In each of the following chapters we always compare the modification suggested in that chapter with the *base scheme*. With the base scheme we thereby refer to the basic scheme presented in Algorithm 1 on page 65 augmented by the modifications described in the previous chapter. Thus in Chapter 8 we already include the modifications required for a general EOS as described in Chapter 7. In Chapter 9 base scheme refers to our finite–volume scheme together with the mixed GLM method described in Chapter 8. We recall this convention in each of the following chapters.

Chapter 7

General Equation of State: the Energy Relaxation Scheme

In this chapter we discuss a possible modification of our base scheme for a perfect gas, which allows us to approximate the MHD equations with a general equation of state (EOS). Instead of solving the original MHD system (1.1) directly, the idea of the *energy relaxation method* (ER) is to solve the MHD system with a much simpler EOS and to use a relaxation mechanism to obtain an approximation to the original system. This idea was first suggested in [CP98] for the Euler equations of gas dynamics and successfully used, for example, in [MS99, In99]. The idea is based on a splitting of the internal energy into two parts $\varepsilon = \varepsilon_1 + \varepsilon_2$. The first part ε_1 governs a simple pressure law p_1 (for example, a polytropic gas law). The disturbing nonlinearities in the original pressure law p are simply advected with the fluid using ε_2. In each step of the time evolution ((3.16) or (3.17)) of the finite–volume scheme the fluxes are computed using any numerical flux function $\mathbf{g}_{ij}$ for the pressure law p_1 together with an additional relaxation step. This scheme can, therefore, be easily implemented as an add–on to a given finite–volume scheme for a perfect gas.

In the MHD system the relationship between the pressure and the total energy has the same form as in the purely hydrodynamic case. Therefore, the idea of the ER scheme can also be applied to solve the MHD system with an arbitrary equation of state. This is presented in the following. Some preliminary results are published in [DW01], where we first presented the ER scheme for the MHD equations; we compared the ER approach with a direct solver following the ideas from [BGH00].

Given a simple pressure law $p_1 = p_1(\rho, \varepsilon)$, i.e. a perfect gas law, we study the *relaxation system* for $\lambda \in \mathbb{R}^+$

$$
\begin{aligned}
\partial_t \rho^\lambda + \nabla \cdot (\rho^\lambda \mathbf{u}^\lambda) &= 0 \,, && \text{(7.1a)}\\
\partial_t (\rho^\lambda \mathbf{u}^\lambda) + \nabla \cdot (\rho^\lambda \mathbf{u}^\lambda (\mathbf{u}^\lambda)^T + \mathcal{P}_1^\lambda) &= 0 \,, && \text{(7.1b)}\\
\partial_t \mathbf{B}^\lambda + \nabla \cdot (\mathbf{u}^\lambda (\mathbf{B}^\lambda)^T - \mathbf{B}^\lambda (\mathbf{u}^\lambda)^T) &= 0 \,, && \text{(7.1c)}\\
\partial_t (\rho^\lambda e_1^\lambda) + \nabla \cdot (\rho^\lambda e_1^\lambda \mathbf{u}^\lambda + \mathcal{P}_1^\lambda \mathbf{u}^\lambda) &= \lambda \rho^\lambda \big(\varepsilon_2^\lambda - \Phi(\rho^\lambda, \varepsilon_1^\lambda)\big) \,, && \text{(7.1d)}\\
\partial_t (\rho^\lambda \varepsilon_2^\lambda) + \nabla \cdot (\rho^\lambda \varepsilon_2^\lambda \mathbf{u}^\lambda) &= -\lambda \rho^\lambda \big(\varepsilon_2^\lambda - \Phi(\rho^\lambda, \varepsilon_1^\lambda)\big) \,, && \text{(7.1e)}
\end{aligned}
$$

together with the divergence constraint

$$\nabla \cdot \mathbf{B}^\lambda = 0 , \tag{7.1f}$$

the algebraic relations

$$e_1^\lambda = \varepsilon_1^\lambda + \frac{1}{2}|\mathbf{u}^\lambda|^2 + \frac{1}{8\pi\rho^\lambda}|\mathbf{B}^\lambda|^2 , \tag{7.1g}$$

$$\mathcal{P}_1^\lambda = \left(p_1^\lambda + \frac{1}{8\pi}|\mathbf{B}^\lambda|^2\right)\mathcal{I} - \frac{1}{4\pi}\mathbf{B}^\lambda(\mathbf{B}^\lambda)^T , \tag{7.1h}$$

and the pressure law

$$p_1^\lambda = p_1(\rho^\lambda, \varepsilon_1^\lambda) . \tag{7.1i}$$

The energy function $\Phi = \Phi(\rho, \varepsilon)$ is chosen so that in the *equilibrium limit* $\lambda \to \infty$ the original MHD system is recovered. We have not included the gravity source terms from (1.1) in (7.1) since they have no influence on the following discussion.
Using the general energy relaxation framework described above, we can derive an extension of our base scheme to compute solutions for arbitrary equations of state. The ER scheme is best understood as an operator splitting scheme for the relaxation system (7.1) in the equilibrium limit. In the splitting approach the evolution of the conserved quantities is performed in two steps. The first step is the relaxation step. Here we neglect the spatial derivatives in (7.1) and solve the remaining system of ODEs for $\lambda \to \infty$. The solution computed in the first step is then used in the second step as initial condition for the homogeneous system, which does not depend on λ. Since the homogeneous system coincides with the standard MHD system with pressure law p_1, we can reuse a numerical scheme for this system in the second step. The solution to the first step can be derived analytically. This is discussed in Section 7.1. In Section 7.2 we show how these two steps can be combined to yield a scheme for the real gas MHD equations, which fits into the finite–volume framework described in Chapter 3. After we discuss the requirements on the EOS imposed by the ER scheme (in Section 7.3 and Section 7.4) we present numerical tests in Section 7.5 and Section 7.6.

7.1 Analytical Motivation

We first show how the energy $\Phi(\rho, \varepsilon_1)$ has to be chosen to allow us to formally recover the original MHD system in the equilibrium limit.

7.1 Theorem
Let Assumption 1.3 hold and consider a family of classical solutions

$$(\rho^\lambda, \rho^\lambda \mathbf{u}^\lambda, \mathbf{B}^\lambda, \rho^\lambda e_1^\lambda, \rho^\lambda \varepsilon_2^\lambda)_{\lambda>0}$$

of system (7.1) *that is uniformly bounded with respect to* λ. *Assume that the* equilibrium limit

$$\mathbf{U}(\mathbf{x}, t) := \lim_{\lambda\to\infty} (\rho^\lambda, \rho^\lambda \mathbf{u}^\lambda, \mathbf{B}^\lambda, \rho^\lambda e_1^\lambda + \rho^\lambda \varepsilon_2^\lambda)^T(\mathbf{x}, t)$$

exists. Then $\mathbf{U}$ *is a solution of the MHD equations* (1.1) *if we choose*

$$\Phi(\rho, \varepsilon_1) = \varepsilon(\rho, p_1(\rho, \varepsilon_1)) - \varepsilon_1 \ . \tag{7.2}$$

Here $\varepsilon(\rho, \cdot)$ *is the inverse of* $p(\rho, \cdot)$ *as defined in Assumption 1.3.*

Proof:
First note that the total energy density $\rho^\lambda e^\lambda = \rho^\lambda e_1^\lambda + \rho^\lambda \varepsilon_2^\lambda$ is a conserved quantity of the relaxation system since, by adding (7.1d) and (7.1e), we arrive at the conservation law

$$\partial_t(\rho^\lambda e^\lambda) + \nabla \cdot (\rho^\lambda e^\lambda \mathbf{u}^\lambda + \mathcal{P}_1^\lambda \mathbf{u}^\lambda) = 0 \ . \tag{7.3}$$

Due to equation (7.1g) the following algebraic relation holds

$$e^\lambda = \varepsilon^\lambda + \frac{1}{2}|\mathbf{u}^\lambda|^2 + \frac{1}{8\pi\rho^\lambda}|\mathbf{B}^\lambda|^2 \tag{7.4}$$

with

$$\varepsilon^\lambda = \varepsilon_1^\lambda + \varepsilon_2^\lambda \ . \tag{7.5}$$

Since we assume that all quantities are uniformly bounded in λ, we have in the equilibrium limit $\varepsilon_2 = \Phi(\rho, \varepsilon_1)$ and therefore $\varepsilon = \varepsilon_1 + \Phi(\rho, \varepsilon_1)$. This lets us recover the MHD system if the *consistency condition*

$$p(\rho, \varepsilon_1 + \Phi(\rho, \varepsilon_1)) = p_1(\rho, \varepsilon_1) \tag{7.6}$$

is satisfied for all $\rho > 0, \varepsilon_1 > 0$. This condition holds if we choose Φ according to (7.2) since ε is the inverse of $p(\rho, \cdot)$. □

In the general context of relaxation approximations the most important condition on the relaxation system are the so–called *subcharacteristic conditions* as found, for example, in [CLL94, JX95]): for stability reasons the wave speeds of the relaxation system in the equilibrium limit must be greater than the wave speeds of the original system. Consequently we must study the wave speeds of the relaxation system (7.1).

7.2 Theorem
Denote with c_1 *the sound speed defined by the pressure law* p_1 *through* (1.2c)*. The wave structure of the relaxation system* (7.1) *consists of nine waves. Eight of these waves are identical to the waves of the MHD system* (1.1) *with the EOS given by* p_1*, i.e. the wave speed are given by* u_x*,* $u_x \pm c_{1,s}$*,* $u_x \pm c_{1,a}$*,* $u_x \pm c_{1,f}$ *with* $c_{1,s}, c_{1,a}$*, and* $c_{1,f}$ *defined by* (1.10) *with the sound speed* $c = c_1$*. The additional wave carries information in* ε_2^λ *and its speed coincides with the speed of the entropy wave* u_x*.*
If the sound speed c_1 *defined by the pressure law* p_1 *satisfies the* subcharacteristic condition

$$c_1(\rho, \varepsilon_1) > c(\rho, \varepsilon_1 + \Phi(\rho, \varepsilon_1)) \tag{7.7}$$

for all $\rho, \varepsilon > 0$ *then each wave speed of the system* (7.1) *is larger or equal to the corresponding wave speed of the original MHD system, i.e.*

$$c_{1,s} \geq c_s \ , \ c_{1,a} \geq c_a \ , \ c_{1,f} \geq c_f \ .$$

Proof:
Since the flux function of the system (7.1a)–(7.1d) corresponds to the flux function of the MHD system (1.1) for the pressure law p_1 and is independent of ε_2 the eigenvalues are given by (1.10) with the sound speed c_1 defined by the pressure law p_1 (cf. (1.2c)). In addition we have a further wave due to equation (7.1e). This new wave travels with the speed of the entropy wave. We conclude that a modification of the pressure law influences only the speed of sound c. The other terms constituting to the wave speeds of the MHD system given in (1.10) are not influenced by the relaxation framework. It remains to show that the wave speeds c_s, c_a, and c_f given by (1.10) are monotone increasing with respect to the sound speed c since then the subcharacteristic conditions are satisfied if $c_1 > c$. The Alfvén speed c_a is independent of c so that we only have to study c_s and c_f. Consider the functions $f_\pm(c) = c^2+b^2 \pm \sqrt{(c^2+b^2)^2 - 4b_1^2c^2}$ with b and b_1 given by (1.11); note that $b^2 \geq b_1^2$. We have to show that $f'_\pm(c) \geq 0$ for all $c > 0$. A simple calculation shows that this is equivalent to $\sqrt{(c^2+b^2)^2 - 4b_1^2c^2} \geq \left|c^2+b^2-2b_1^2\right|$ which holds for $b^2 \geq b_1^2$. □

In addition to condition (7.7) the authors in [CP98] require the function Φ to be monotone increasing with respect to ε_1 for fixed ρ; like the subcharacteristic condition (7.7) the monotonicity condition on Φ is a condition on the pressure law p_1 due to (7.2). If we choose a perfect gas law for p_1 with some constant γ_1, then both these conditions lead to a lower bound on γ_1.

7.3 Theorem
Consider a constant $\gamma_1 > 1$ which satisfies

$$\gamma_1 > \sup_{\rho,\varepsilon>0} \max\{\gamma(\rho,\varepsilon), \Gamma(\rho,\varepsilon)\} \tag{7.8}$$

with $\gamma(\rho,\varepsilon), \Gamma(\rho,\varepsilon)$ defined by

$$\gamma(\rho,\varepsilon) := \frac{\rho c^2(\rho,\varepsilon)}{p(\rho,\varepsilon)} \ , \tag{7.9}$$

$$\Gamma(\rho,\varepsilon) := 1 + \tau\partial_\varepsilon p(\rho,\varepsilon) \ . \tag{7.10}$$

Assume that $\partial_\varepsilon p > 0$. If we choose $p_1(\rho,\varepsilon_1) = (\gamma_1 - 1)\rho\varepsilon_1$, then condition (7.7) is satisfied and the function Φ defined by (7.2) is monotone increasing in ε_1.

Proof:
First we verify that $c_1 > c$

$$\begin{aligned} c_1^2(\rho,\varepsilon_1) &= \gamma_1 \frac{p_1(\rho,\varepsilon_1)}{\rho} \\ &> \gamma(\rho,\varepsilon)\frac{p_1(\rho,\varepsilon_1)}{\rho} \\ &= \gamma(\rho,\varepsilon)\frac{p(\rho,\varepsilon_1+\Phi(\rho,\varepsilon_1))}{\rho} \\ &= c^2(\rho,\varepsilon_1+\Phi(\rho,\varepsilon_1)) \ . \end{aligned}$$

Here we used the definition of Φ from (7.2). The monotonicity of Φ follows from the relation

$$\partial_{\varepsilon_1}\Phi = \frac{\partial_{\varepsilon_1} p_1}{\partial_\varepsilon p} - 1 = \frac{(\gamma_1 - 1)\rho}{\partial_\varepsilon p} - 1 \geq \frac{(\Gamma - 1)\rho}{\partial_\varepsilon p} - 1 = 0$$

since $\gamma_1 > \Gamma$ and $\partial_\varepsilon p > 0$. □

We conclude our study of the system (7.1) by computing the solution to the system of ODEs which we obtain from (7.1) by neglecting the spatial derivatives. As outlined at the beginning of this chapter, we use the solution to this system in the construction of our numerical scheme.

7.4 Theorem
Let the energy Φ be given by (7.2) and assume that Φ is monotone increasing in ε_1. Consider the following system of ODEs

$$\partial_t \rho^\lambda = 0 \ , \tag{7.11a}$$
$$\partial_t (\rho \mathbf{u})^\lambda = 0 \ , \tag{7.11b}$$
$$\partial_t \mathbf{B}^\lambda = 0 \ , \tag{7.11c}$$
$$\partial_t (\rho^\lambda e_1^\lambda) = \lambda \rho^\lambda \big(\varepsilon_2^\lambda - \Phi(\rho^\lambda, \varepsilon_1^\lambda)\big) \ , \tag{7.11d}$$
$$\partial_t (\rho^\lambda \varepsilon_2^\lambda) = -\lambda \rho^\lambda \big(\varepsilon_2^\lambda - \Phi(\rho^\lambda, \varepsilon_1^\lambda)\big) \ . \tag{7.11e}$$

Let the initial conditions be given by

$$\big(\rho_0, (\rho\mathbf{u})_0, \mathbf{B}_0, (\rho e_1)_0, (\rho\varepsilon_2)_0\big) \ . \tag{7.12}$$

Denote with $(\varepsilon_1)_0$ the internal energy of the initial data defined through the relation (7.1g). Then the solution to (7.11) for $\lambda \to \infty$ is

$$\big(\rho_0, (\rho\mathbf{u})_0, \mathbf{B}_0, \rho_0\varepsilon_1^* + \frac{1}{2}\rho_0|\mathbf{u}_0|^2 + \frac{1}{8\pi}|\mathbf{B}_0|^2, \rho_0\varepsilon_2^*\big) \tag{7.13}$$

The constants ε_1^* and ε_2^* are defined by the algebraic relations

$$\begin{aligned} p(\rho_0, (\varepsilon_1)_0 + (\varepsilon_2)_0) &= p_1(\rho_0, \varepsilon_1^*) \ , \\ \varepsilon_1^* + \varepsilon_2^* &= (\varepsilon_1)_0 + (\varepsilon_2)_0 \ . \end{aligned} \tag{7.14}$$

If $p_1(\rho, \varepsilon_1) = (\gamma_1 - 1)\rho\varepsilon_1$ then ε_1^* and ε_2^* are given by the explicit relations

$$\varepsilon_1^* := \frac{p(\rho_0, \varepsilon_0)}{(\gamma_1 - 1)\rho_0} \ , \tag{7.15a}$$
$$\varepsilon_2^* := \varepsilon_0 - \varepsilon_1^* \tag{7.15b}$$

with $\varepsilon_0 := (\varepsilon_1)_0 + (\varepsilon_2)_0$.

Proof:
Obviously we have $\rho^\lambda = \rho_0$ for all λ and also $(\rho\mathbf{u})^\lambda = (\rho\mathbf{u})_0, \mathbf{B}^\lambda = \mathbf{B}_0$ — this remains

true in the equilibrium limit. If we replace $\rho^\lambda e_1^\lambda$ in (7.11d) using (7.1g) then the last two equations in (7.11) can be rewritten as

$$\begin{aligned}\partial_t \varepsilon_1^\lambda &= \lambda\big(\varepsilon_2^\lambda - \Phi(\rho^\lambda, \varepsilon_1^\lambda)\big) \ , \\ \partial_t \varepsilon_2^\lambda &= -\lambda\big(\varepsilon_2^\lambda - \Phi(\rho^\lambda, \varepsilon_1^\lambda)\big) \ .\end{aligned}$$

Using a rescaling of the time $t = \frac{s}{\lambda}$ we arrive at

$$\begin{aligned}\partial_s \varepsilon_1 &= (\varepsilon_2 - \Phi(\rho_0, \varepsilon_1)) \ , \\ \partial_s \varepsilon_2 &= -(\varepsilon_2 - \Phi(\rho_0, \varepsilon_1)) \ ,\end{aligned}$$

with initial data given by $((\varepsilon_1)_0, (\varepsilon_2)_0)$ where $(\varepsilon_1)_0$ is computed from the initial data (7.12) using (7.1g). The equilibrium limit is now the solution of this two by two system for $s \to \infty$. We denote these limit functions with ε_1^* and ε_2^*. Adding the two equations leads to $\partial_s(\varepsilon_1 + \varepsilon_2) = 0$ and therefore $\varepsilon_1^* + \varepsilon_2^* = (\varepsilon_1)_0 + (\varepsilon_2)_0$. Since we assumed that $\Phi(\rho_0, \cdot)$ is a monotone increasing function, it can be seen that the equilibrium limit exists and is characterized by the algebraic relation $\varepsilon_2^* = \Phi(\rho_0, \varepsilon_1^*)$. Due to (7.2) this relation is equivalent to

$$p(\rho_0, \varepsilon_1^* + \varepsilon_2^*) = p_1(\rho_0, \varepsilon_1^*) \ .$$

Therefore ε_1^* and ε_2^* are defined by the two relations (7.14).
Replacing p_1 in the first equation in (7.14) with $(\gamma_1 - 1)\rho\varepsilon$ we find

$$p(\rho_0, (\varepsilon_1)_0 + (\varepsilon_2)_0) = (\gamma_1 - 1)\rho_0 \varepsilon_1^* \ .$$

Solving this equation for ε_1^* and the second equation in (7.14) for ε_2^* leads to the explicit equations (7.15) for ε_1^* and ε_2^*. This concludes the proof. □

7.5 Remark: *Neither ε_1^* nor ε_1^* directly depend on the initial conditions $(\varepsilon_1)_0, (\varepsilon_2)_0$ but only on the sum $\varepsilon_0 = (\varepsilon_1)_0 + (\varepsilon_2)_0$. This makes it easy to use the splitting approach to construct a modified flux function for solving the MHD system directly. Also note that the relaxation step does not influence the magnetic field so that the divergence constraint is not touched by the relaxation framework.*

7.2 Numerical Scheme

Using the general framework described above we now derive an extension of our base scheme to compute solutions for arbitrary equations of state. Since we want to reuse our numerical flux functions for the perfect gas MHD equations, we assume in the following that the simple pressure law used is a perfect gas law

$$p_1(\rho, \varepsilon_1) = (\gamma_1 - 1)\rho\varepsilon_1 \tag{7.16}$$

for some $\gamma_1 > 1$ satisfying (7.8). We incorporate the splitting approach described above into our finite–volume framework. In the following we denote with $\mathbf{G}^{\text{ideal}}(\cdot, \cdot; \gamma_1) = (G_i^{\text{ideal}}(\cdot, \cdot; \gamma_1))_{1 \le i \le 8}$ an arbitrary flux function for the MHD equations in one space

dimension for the perfect gas law (7.16) (cf. Section 3.4). For a general EOS we obtain in two steps a flux $\mathbf{G}(\mathbf{U}_l, \mathbf{U}_r)$ for two given left and right hand states $\mathbf{U}_{l,r} = (\rho_{l,r}, (\rho\mathbf{u})_{l,r}, \mathbf{B}_{l,r}, (\rho e)_{l,r})$ by following the splitting approach described above. We first construct the mappings $\mathbf{\Psi}_1, \Psi_2$ which define the solution to the relaxation step for the initial data $\mathbf{U} \in \mathcal{U}$ using Theorem 7.4

$$\begin{aligned}\Psi_2(\mathbf{U}) &:= \varepsilon - \frac{p(\rho,\varepsilon)}{\rho(\gamma_1 - 1)} \ , \\ \mathbf{\Psi}_1(\mathbf{U}) &:= \Big(\rho, (\rho\mathbf{u}), \mathbf{B}, \rho e - \rho\Psi_2(\mathbf{U})\Big)\end{aligned}$$

(cf. (7.13) and (7.15)). We can now compute the fluxes for $(\rho, \rho\mathbf{u}, \mathbf{B}, \rho e_1)$ using the numerical flux

$$\mathbf{G}^{\text{ideal}}(\mathbf{\Psi}_1(\mathbf{U}_l), \mathbf{\Psi}_1(\mathbf{U}_r); \gamma_1) \ .$$

The flux $\mathbf{G}^{\varepsilon_2}(\cdot,\cdot)$ for $\rho\varepsilon_2$ can be computed using the flux for ρ following the ideas from [Lar91] derived for the similar situation of multicomponent flow. The mass flux G_1^{ideal} is an approximation of $\rho\mathbf{u}$ on the interface. If it is positive ε_2 should be advected to the left and otherwise to the right. Thus we define

$$\begin{aligned}&G^{\varepsilon_2}(\mathbf{U}_l, \mathbf{U}_r) := \\ &\begin{cases} \mathbf{G}_1^{\text{ideal}}(\mathbf{\Psi}_1(\mathbf{U}_l), \mathbf{\Psi}_1(\mathbf{U}_r); \gamma_1)\Psi_2(\mathbf{U}_l) & \text{for} \quad \mathbf{G}_1^{\text{ideal}}(\mathbf{\Psi}_1(\mathbf{U}_l), \mathbf{\Psi}_1(\mathbf{U}_r); \gamma_1) \geq 0 \ , \\ \mathbf{G}_1^{\text{ideal}}(\mathbf{\Psi}_1(\mathbf{U}_l), \mathbf{\Psi}_1(\mathbf{U}_r); \gamma_1)\Psi_2(\mathbf{U}_r) & \text{for} \quad \mathbf{G}_1^{\text{ideal}}(\mathbf{\Psi}_1(\mathbf{U}_l), \mathbf{\Psi}_1(\mathbf{U}_r); \gamma_1) < 0 \ . \end{cases}\end{aligned} \tag{7.17}$$

Since we want to compute a flux for ρe we use the relation (7.4). Thus we define the flux $\mathbf{G}(\cdot,\cdot)$ by

$$\mathbf{G}(\mathbf{U}_l, \mathbf{U}_r) := \mathbf{G}^{\text{ideal}}(\mathbf{\Psi}_1(\mathbf{U}_l), \mathbf{\Psi}_1(\mathbf{U}_r); \gamma_1) + (0, \mathbf{0}, \mathbf{0}, G^{\varepsilon_2}(\mathbf{U}_l, \mathbf{U}_r)) \ . \tag{7.18}$$

7.6 Remark: *The functions $\mathbf{\Psi}_1, \Psi_2$ are clearly Lipschitz continuous functions if we assume that p is smooth and that the density is bounded away from zero. The flux $\mathbf{G}$ is a combination of the original flux function and $\mathbf{\Psi}_1, \Psi_2$. Thus it is easy to verify that it has the same basic properties as the original flux function with respect to conservation, consistency, and Lipschitz continuity (cf. Definition 4.5). It is important to note that both $\mathbf{\Psi}_1, \Psi_2$ are very simple to compute and require only one evaluation of the pressure law. We do not require the function Φ in our algorithm and thus we do not need the inverse function $\varepsilon(\rho, p)$ either. On the other hand computing γ_1 through (7.19) requires some additional evolution of the pressure law and its derivatives. In numerical simulations it turns out that the condition $\gamma_1 > \gamma$ is the important one. It requires the computation of the speed of sound c. In the following section we discuss the requirements on the EOS in more detail.*
Finally, it is worth noting that the functions $\mathbf{\Psi}_1, \Psi_2$ are not changed by the rotation necessary in our 2d flux computation (cf. Section 3.4); this is easy to verify if one keeps in mind that the rotation does not change the length of the vectors $\mathbf{u}$ and $\mathbf{B}$ and therefore neither the internal energy ε nor the pressure p is effected. Consequently, the order in which the relaxation step and the rotation of the two states are performed has no influence on the result.

Since we are approximating the flux of the relaxation system, the time step has to be chosen in accordance with the eigensystem of (7.1). In Theorem 7.2 we have shown that the eigensystem of (7.1) has the same structure as the perfect gas MHD system with an additional wave for ε_2 that moves with the same speed as the entropy wave (cf. Theorem 7.2). In the definition of the slow and the fast wave (1.10) the sound speed c_1 has to be used instead of the original sound speed c. Since we have chosen γ_1 according to Theorem 7.2, so that c_1 is greater than c, these waves move faster than the original waves in the MHD system. If we choose Δt with local time steps computed using c_1 instead of c in the definition of $\lambda_{\max}$ (cf. (3.10)), this leads to a stable scheme.

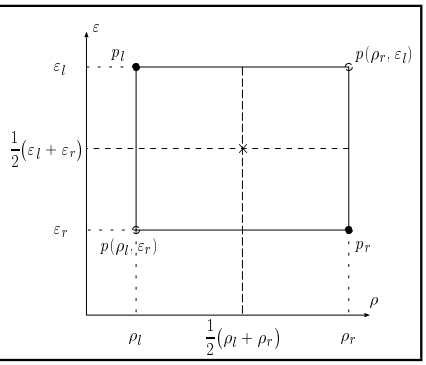
7.7 Remark: *The time steps that we have to use in the ER scheme is smaller than the original time steps so that the requirements on the modified scheme, stated at the beginning of the overview Chapter 6, are not met. On the other hand in the case that the pressure law p is a perfect gas law with a constant $\gamma_0 > 0$ equation* (7.19) *leads to $\gamma_1 = \gamma_0$. Therefore the ER scheme is identical to the base scheme in this case, and since $c_1 = c$ the time step also remains unaltered. Consequently the ER scheme hardly increases the computational cost of the scheme if the pressure law p is a perfect gas law.*

Since larger values for γ_1 lead to a higher amount of numerical viscosity, the choice of this constant can be very crucial. The authors of [CP98] point out that it is enough to satisfy (7.8) *locally*, i.e., we may choose γ_1 separately at every interface depending on $\mathbf{U}_l$ and $\mathbf{U}_r$. In [DW01] we experimented with the following choice:

$$\gamma_1(\mathbf{U}_l, \mathbf{U}_r) := \max\left\{\gamma(\rho_l, \varepsilon_l), \gamma(\rho_r, \varepsilon_r), \Gamma(\rho_l, \varepsilon_l), \Gamma(\rho_r, \varepsilon_r)\right\} \ . \tag{7.19}$$

The same definition was also used in [In99]. If the speed of sound and $\partial_\varepsilon p$ are available and can be cheaply evaluated, γ_1 can be directly computed using (7.19). In case where this is not possible or too expensive we have also tested central differences to approximate the derivatives of p at $\left(\frac{1}{2}(\rho_l + \rho_r), \frac{1}{2}(\varepsilon_l + \varepsilon_r)\right)$ as shown in Figure 7.1. In

Figure 7.1: To compute γ_1 using (7.19) the pressure function and its derivatives are approximated. If $|\rho_l - \rho_r| > 10^{-8}$ and $|\varepsilon_l - \varepsilon_r| > 10^{-8}$, we use the four pressure values indicated in the sketch to compute a finite difference approximation at the point $\left(\frac{1}{2}(\rho_l + \rho_r), \frac{1}{2}(\varepsilon_l + \varepsilon_r)\right)$. If the left and right hand state are too close together, we use $\gamma_1 = 1 + \max\left\{\frac{p(\rho_l,\varepsilon_l)}{\rho_l \varepsilon_l}, \frac{p(\rho_r,\varepsilon_r)}{\rho_r \varepsilon_r}\right\}$.



Algorithm 3 on page 94 we summarize all the necessary steps for the modification of the base scheme.

7.3 Demands on the EOS

The number of calls to the EOS required in each time step of a numerical scheme have a strong influence on its performance. Clearly different methods require a dif-

Algorithm 3: Modification of the base scheme Algorithm 2(b) on page 66 to cope with a general EOS. The flux function $\mathbf{G}^{\text{ideal}}(\cdot,\cdot;\gamma_1)$ denotes an arbitrary flux function for the 1d perfect gas MHD equations with adiabatic exponent γ_1. The new parts are highlighted.
The right algorithm sketches the computation of the pressure function using an adaptive table to store discrete pressure values.

(a) **FLUX**$(\mathbf{U}_l, \mathbf{U}_r, \mathbf{n}, h)$

$\gamma_1 \leftarrow \max \left\{ \begin{array}{ll} \gamma(\rho_l,\varepsilon_l), & \gamma(\rho_r,\varepsilon_r), \\ \Gamma(\rho_l,\varepsilon_l), & \Gamma(\rho_r,\varepsilon_r) \end{array} \right\}$
$(\varepsilon_1)_l \leftarrow \frac{p(\rho_l,\varepsilon_l)}{\rho_l(\gamma_1-1)}$
$(\varepsilon_1)_r \leftarrow \frac{p(\rho_r,\varepsilon_r)}{\rho_l(\gamma_1-1)}$
$(\varepsilon_2)_l \leftarrow \varepsilon_l - (\varepsilon_1)_l$
$(\varepsilon_2)_r \leftarrow \varepsilon_r - (\varepsilon_1)_r$
$\mathbf{W}_l \leftarrow \big(\rho_l, \rho_l\mathbf{u}_l, \mathbf{B}_l,$
$\quad \rho_l(\varepsilon_1)_l + \frac{1}{2}\rho_l|\mathbf{u}_l|^2 + \frac{1}{8\pi}|\mathbf{B}_l|^2\big)^T$
$\mathbf{W}_r \leftarrow \big(\rho_r, \rho_r\mathbf{u}_r, \mathbf{B}_r,$
$\quad \rho_r(\varepsilon_1)_r + \frac{1}{2}\rho_r|\mathbf{u}_r|^2 + \frac{1}{8\pi}|\mathbf{B}_r|^2\big)^T$
$\bar{\mathbf{U}}_l \leftarrow \mathcal{R}(\mathbf{n}_{ij})\mathbf{W}_l$
$\bar{\mathbf{U}}_r \leftarrow \mathcal{R}(\mathbf{n}_{ij})\mathbf{W}_r$
$\bar{\mathbf{G}} \leftarrow \mathbf{G}(\bar{\mathbf{U}}_l, \bar{\mathbf{U}}_r)$
$\mathbf{g}_{ij} \leftarrow |S_{ij}|\mathcal{R}^{-1}(\mathbf{n}_{ij})\bar{\mathbf{G}}^{\text{ideal}}(\bar{\mathbf{U}}_l, \bar{\mathbf{U}}_r; \gamma_1)$
$\Delta t_{ij} \leftarrow \frac{h}{\max\{|\bar{u}_{l,x}|+c_f(\bar{\mathbf{U}}_l), |\bar{u}_{r,x}|+c_f(\bar{\mathbf{U}}_r)\}}$
$\text{jmp}_{ij} \leftarrow$ [cf. (3.19)]
if $(\mathbf{g}_{ij})_1 \geq 0$ **then**
  $\bar{G}^{\varepsilon_2} \leftarrow (\mathbf{g}_{ij})_1(\varepsilon_2)_l$
else
  $(\mathbf{g}_{ij})_1(\varepsilon_2)_r$
end if
$(\mathbf{g}_{ij})_8 \leftarrow (\mathbf{g}_{ij})_8 + \bar{G}^{\varepsilon_2}$

(b) **Adaptive table for computing** $p(\rho,\varepsilon)$

Require: $I > 1$, $J > 1$, $K > 2$, $L > 1$ and an *empty* grid $(\rho_i, \varepsilon_j)_{1\leq i\leq I,\, 1\leq j\leq J}$ on level 0.
compute (i_0, j_0) with $(\rho,\varepsilon) \in [\rho_{i_0}, \rho_{i_0+1}] \times [\varepsilon_{j_0}, \varepsilon_{j_l+1}]$
$l \leftarrow 1$
while the value of $p(\rho,\varepsilon)$ is unknown **do**
  if subgrid $(\rho_{i_0,\ldots,i_{l-1}}, \varepsilon_{j_0,\ldots,j_{l-1}})$ exists **then**
    compute (i_l, j_l) with $(\rho,\varepsilon) \in [\rho_{i_0,\ldots,i_l}, \rho_{i_0,\ldots,i_l+1}] \times [\varepsilon_{j_0,\ldots,j_l}, \varepsilon_{j_0,\ldots,j_l+1}]$
    if subgrid $(\rho_{i_0,\ldots,i_{l-1}}, \varepsilon_{j_0,\ldots,j_{l-1}})$ contains values **then**
      determine $p(\rho,\varepsilon)$ from $\{p(\rho_{i_0,\ldots,i_l+r}, \varepsilon_{j_0,\ldots,j_l+s}) | r,s \in \{0,1\}\}$.
    else
      $l \leftarrow l+1$
    end if
  else
    create *empty* $K \times K$–subgrid at $(\rho_{i_0,\ldots,i_{l-1}}, \varepsilon_{j_0,\ldots,j_{l-1}})$ on level l
    if $l = L$ or error tolerance reached on level $l-1$ **then**
      compute and store $\{p(\rho_{i_0,\ldots,i_{l-1},r}, \varepsilon_{j_0,\ldots,j_{l-1},s}) | r,s \in \{1,\ldots K\}\}$
    end if
  end if
end while

ferent number of calls and also a different amount of additional information such as the derivatives of the pressure law. In this section we compare the ER scheme with the simple Lax–Friedrichs scheme (LF) with respect to their requirements on and the number of calls to the EOS. The LF scheme requires no modification when used with an arbitrary EOS. For the flux computation it requires the evaluation of the analytical flux function and thus the pressure p has to be computed once for $\mathbf{U}_l$ and $\mathbf{U}_r$. (In a first order scheme it is possible to reduce the number of calls even further, but we want to compare the methods in the case of the second order scheme using reconstruction and a two step Runge–Kutta method.) This leads to four calls to the EOS for each face S_{ij} for $(i,j) \in \mathcal{I}_S$. In addition we also have to compute the local time steps Δt_{ij} for $(i,j) \in \mathcal{I}_S$. This requires the computation of the sound speed c twice once for $\mathbf{U}_i(\mathbf{z}_{ij}, t^n)$ and once for $\mathbf{U}_j(\mathbf{z}_{ij}, t^n)$ for every face S_{ij} (recall that we compute Δt only in the first step of our Runge–Kutta method).

In its simplest form, i.e. with a constant γ_1, the ER scheme also requires only two calls to the EOS per face in each step of the time evolution. Since for the computation of the local time steps we use c_1 instead of c, we do not require any further knowledge of the EOS or calls to the EOS. In complex applications the choice of a constant γ_1 in

scheme	requirement for						no. of calls to		
	flux evaluation				time step				
	p	$\partial_\tau p$	$\partial_\varepsilon p$	ε	$\partial_\tau p$	$\partial_\varepsilon p$	p	$\partial_\tau p$	$\partial_\varepsilon p$
Lax–Friedrichs	x	—	—	—	x	x	4	1	1
ER, constant γ_1	x	—	—	—	—	—	4	0	0
ER, γ_1 from (7.19)	x	x	x	—	—	—	4	2	2
ER, FD–scheme for γ_1	x	—	—	—	—	—	8	0	0

Table 7.1: Comparison of second order schemes w.r.t. their requirements and the number of calls to the EOS at each interface. Note that the time step in the energy relaxation ("ER") schemes is computed from the relaxation system.

space and time leads either to very small time steps and thus to a scheme that suffers from high numerical viscosity or the scheme will be unstable. Therefore a local choice of γ_1 on each face and in each time step seems more promising. The most important bound on γ_1 is imposed by the condition $c_1 > c$. This leads to the bound $\gamma_1 > \gamma$ with γ defined by (7.9). To compute γ we again have to compute the sound speed c on each interface for both states in each call to the flux function. In this form the ER scheme requires four calls to the pressure law and four computations of the sound speed and is, therefore, slightly more expensive than the LF scheme. To arrive at the same number of calls as in the case of the LF scheme we could also store γ_1 on each face in the first step of our Runge–Kutta method and reuse the value in the second step. If we, furthermore, impose $\gamma_1 > \Gamma$ we have to compute the derivative of p with respect to ε. We can also use a finite difference approximation for the computation of the sound speed and $\partial_\varepsilon p$; but this requires two additional calls to the pressure (cf. Figure 7.1). Consequently, the cost is comparable to the direct calculation of γ_1; but no additional information of the EOS is required.

In Table 7.1 we give a summary of our comparison of the ER scheme and the Lax–Friedrichs scheme. We should point out that the analysis is not complete. For example, one could also use a finite difference approximation to compute the sound speed in the case of the Lax–Friedrichs scheme. The important thing to note is that the ER scheme requires only slightly more calls to the EOS than the Lax–Friedrichs scheme, and to our knowledge there is no scheme available which requires fewer calls.

7.4 Tabularized Equation of State

The ER scheme described above can be implemented very cheaply in respect to the number of calls to the EOS. In our applications, however, even a few calls to the EOS in each time–step lead to an unacceptable performance of the scheme. Therefore, a further reduction of the number of evaluations of the pressure law is necessary; this can be achieved, for example, by storing values of the pressure (and required derived values) in a table. Values for the pressure can then be interpolated from the table for arbitrary values of ρ, ε. This reduces the cost of the method considerably. In some

applications the use of a table for the pressure values may even be the only possibility — in the case where the pressure is only known from experimental data and cannot be directly computed from any given pair ρ, ε.
The simplest approach is to use bilinear interpolation on a structured Cartesian grid in the ρ/ε plane. Using bilinear interpolation guarantees that if p is positive and monotone increasing with respect to ε then this is also true for the approximate pressure function p_h. A further advantage is the low cost for computing the pressure during the simulation — it is comparable to computing the pressure for an ideal gas law. On the other hand in order to achieve a suitable resolution, a very fine grid may be necessary depending on the complexity of the EOS. The numerical experiments presented at the end of this chapter show that the resolution of the grid for storing the pressure values has to be chosen in accordance with the spatial grid on which the MHD system is solved.
In our application we can compute pressure values on the fly and in this case an adaptive table built during each simulation has proven to be very successful, as we will show at the end of this chapter. By using a hierarchical structure for storing the table we can easily refine it in regions where it is necessary. We do not compute and store pressure values for pairs of ρ, ε that do not occur during a simulation. Consequently we save a considerable amount of memory, avoid problems arising from unphysical combinations of ρ, ε for which the pressure cannot be computed, and we achieve the necessary resolution. Details of the algorithm are published in [DRW02a] and are sketched in Algorithm 3(b) on page 94.

7.5 Numerical Results in 1d

Since the ER method only influences the one dimensional flux function $\mathbf{G}$, we study the different aspects of the ER modification mainly in one space dimension. If not noted otherwise, we use equation (7.19) to define γ_1 and the HLLEM flux function for $\mathbf{G}^{\text{ideal}}$. Results for the Riemann problem 6.3(RPWaals1) are shown in Figure 7.2. As can been seen, the complicated compound wave structure is captured quite well by the relaxation scheme. In the following three subsections we focus on the Riemann Problem 6.4(RPWalls2), for which we have an exact solution.

7.5.1 Linear Reconstruction

We start our study of the energy relaxation scheme with a brief look at the construction of the higher order scheme. The observations discussed here are not restricted to the case of the ER scheme. To construct higher order finite–volume schemes the reconstruction and limiting of the averaged values is the most crucial task. In [DRW02a] we describe the method we use for our 2d simulations. In one space dimension or on Cartesian grids the reconstruction process is far simpler. The method we use here is the standard *minmod* limiter (see, for example, [Krö97, Example 2.5.10]). In both cases a method for scalar quantities is applied to each component of the state vector $\mathbf{U}$. On unstructured grids the main problem is to construct a stable scheme that still leads to higher order in smooth regions. In this study we do not focus on this problem. We only demonstrate in Figure 7.3 the difficulties that are present even in 1d simulations. Especially at the contact discontinuity, where the velocity and the pressure

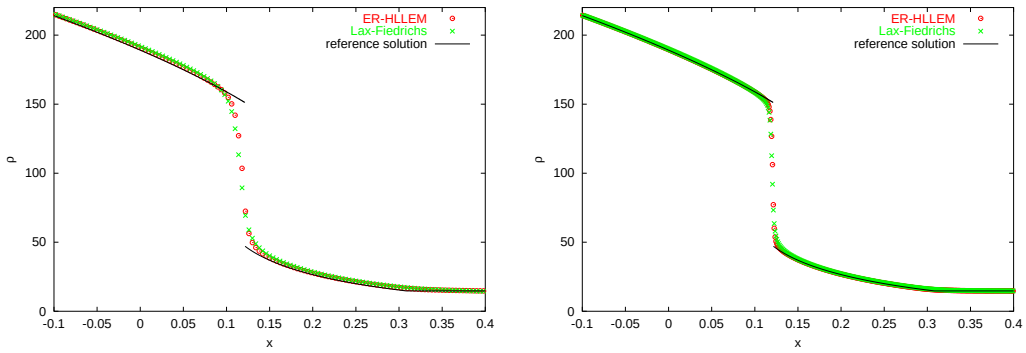
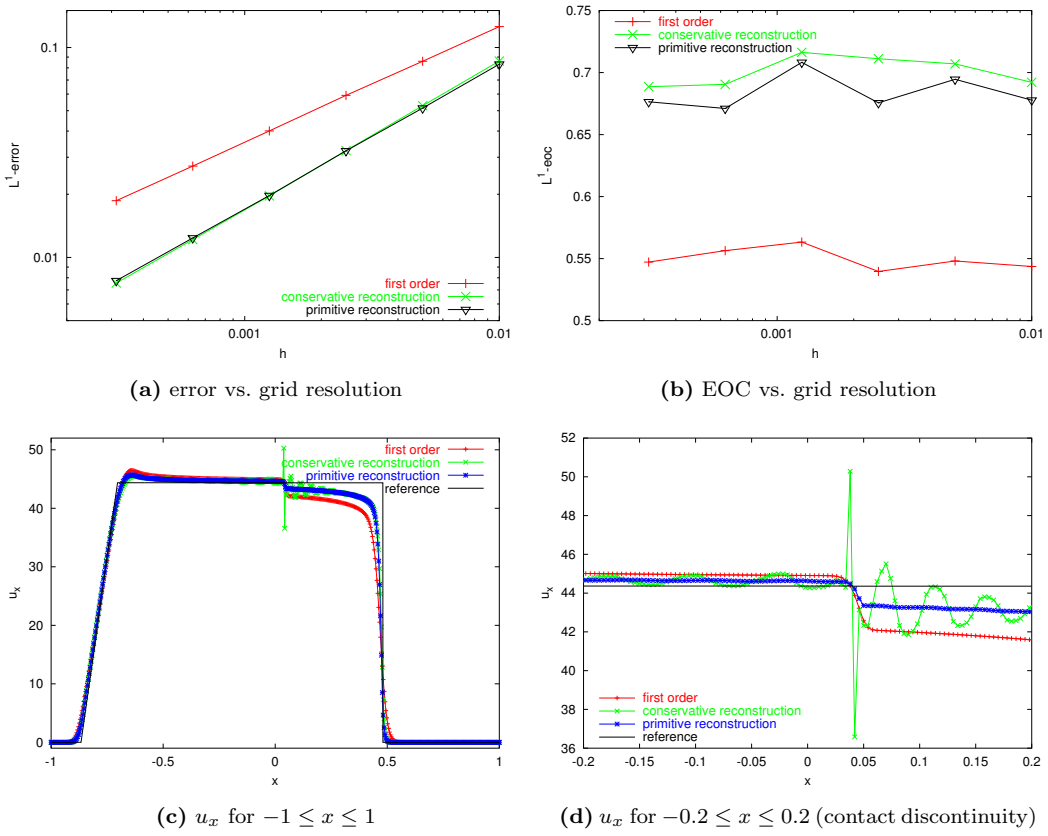


Figure 7.2: Problem: Riemann Problem 6.3(RPWaals1). Method: ER–HLLEM scheme and LF scheme applied to the real gas MHD equations with minmod reconstruction in primitive variables. Left: 500 grid points. Right: 2000 grid points. The reference solution was computed by the authors of [MV99] (cf. Figure 6.1 on page 70). Both numerical schemes lead to a good approximation of all the different waves, even of the compound wave structure, which we show here. A close look reveals that the ER scheme produces slightly less numerical viscosity than the LF scheme.

should be constant, strong oscillations are evident if we use the standard technique of reconstructing the conservative variables $\rho, \rho\mathbf{u}, \mathbf{B}, \rho e$. Neither $\rho\mathbf{u}$ nor ρe are constant over contact discontinuities. Therefore even if we start out with constant $\mathbf{u}$ and p the reconstructed values at the interface will lead to an artificial jump in both components. This problem can be considerably reduced if the reconstruction and limiting technique is not applied to the conservative variables but to the primitive variables $\rho, \mathbf{u}, \mathbf{B}, p$. We thus construct linear functions for each of these quantities on every cell using the same technique for each scalar quantity as for the conservative variables. With these linear functions we define the conservative quantities in each point of the cell. Note that this leads to a reconstruction of the conservative quantities, which is not linear. For example, $\rho\mathbf{u}$ is a quadratic function. To define the energy, we have to use the internal energy function $\varepsilon(\rho, p)$ as defined in Assumption 1.3. This is a drawback of this strategy since this function is not necessarily available or its evaluation leads to additional computational cost. The reconstruction of the quantities $\rho, \rho\mathbf{u}, \mathbf{B}, \rho e$ in this manner is, furthermore, not conservative in the sense that the reconstructed functions can have a different average value than the original constant approximation. This does not seem to lead to any problems since the reconstruction is only used to supply left and right states for the flux computation. The actual update is still performed using the original average values. Thus conservation is guaranteed if the flux function satisfies (3.15). We conclude from our test that using primitive reconstruction often leads to a more stable scheme but at the same time also to a loss of efficiency.

7.5.2 Choosing the Parameter γ_1

The stability of the relaxation scheme is directly influenced by the choice of γ_1. We discussed a number of possibilities for defining γ_1 in Section 7.2. The simplest choice



(a) error vs. grid resolution

(b) EOC vs. grid resolution

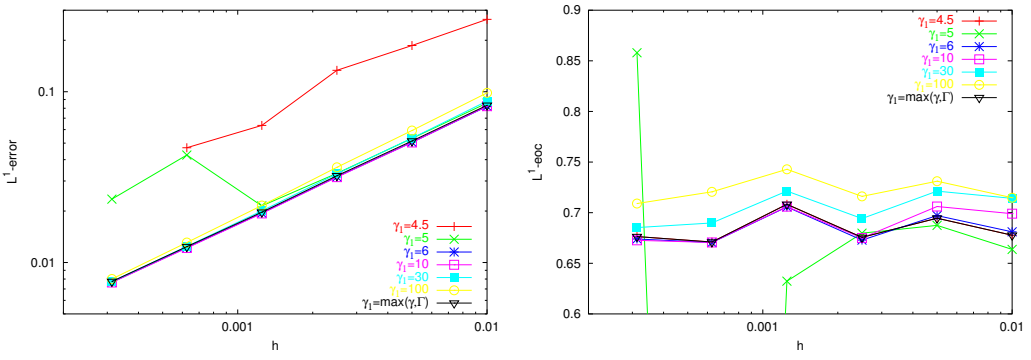
(c) u_x for $-1 \leq x \leq 1$

(d) u_x for $-0.2 \leq x \leq 0.2$ (contact discontinuity)

Figure 7.3: Problem: Riemann Problem 6.4(RPWalls2). Method: ER–HLLEM scheme with minmod reconstruction in primitive and conservative variables. The velocity is constant over the contact discontinuity. In the first order scheme and for both types of reconstruction a small jump is introduced into the approximate solution. The conservative reconstruction leads to additional oscillations.

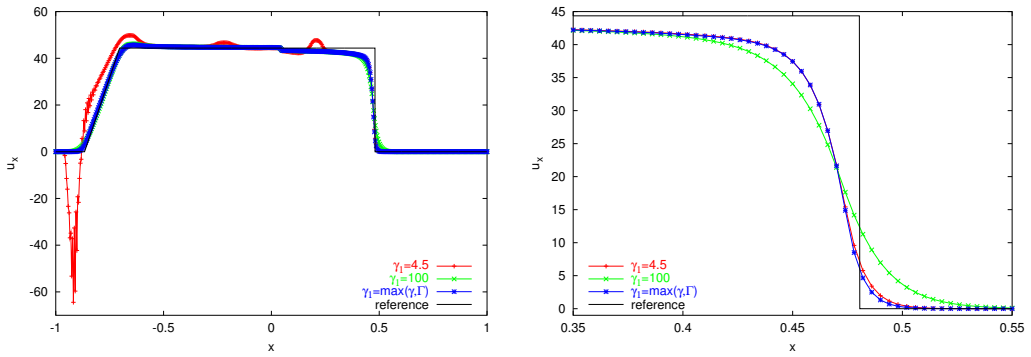
is to use a fixed value chosen a priori. The advantage of this method is that it leads to no additional computational cost as required for computing a local γ_1. The main drawback is that some a priori knowledge of the solution must be available — a γ_1 which is too small leads to an unstable scheme; at the same time a large value leads to additional numerical viscosity, which reduces the efficiency of the scheme. The effect of the choice of γ_1 on the solution of the Riemann Problem 6.4(RPWalls2) is shown in Figure 7.4. As reference we include the results obtained using (7.19) to compute γ_1 locally in every flux calculation. Since the van der Waals EOS is still quite simple, the additional overhead to compute γ_1 locally is negligible. Therefore we do not include a study of the efficiency of the scheme using a local γ_1 compared to choosing γ_1 constant. This might be an issue if the EOS is more complicated. The stability problems that are evident from the results shown in Figure 7.4, however, lead us to favor the local choice in any case.

As already mentioned above, the derivatives of the pressure law might not be available or they might be far too expensive to compute. This situation occurs in our applica-



(a) error vs. grid resolution

(b) EOC vs. grid resolution



(c) u_x for $-1 \leq x \leq 1$

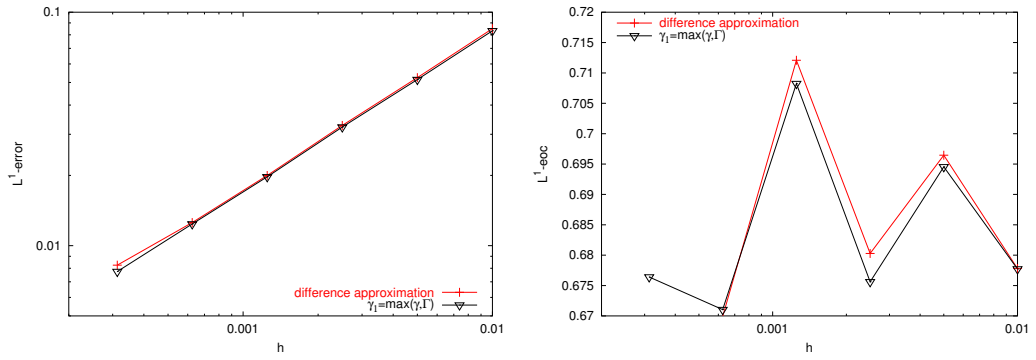
(d) u_x for $0.35 \leq x \leq 0.55$ (shock wave)

Figure 7.4: Problem: Riemann Problem 6.4(RPWalls2). Method: ER–HLLEM scheme with minmod reconstruction in primitive variables. ER method with a constant γ_1 compared to using (7.19). If γ_1 is chosen too small, the method is unstable and can even break down. For large γ_1 the numerical viscosity is increased. The results show that using (7.19) gives near optimal results, but that a choice that is too large only leads to a small reduction of the quality of the solution.

tion in the solar photosphere. A first step towards handling this case is to use finite differences to approximate the derivatives in (7.19). Figure 7.5 shows that this approximation barely reduces the quality of the scheme. Since even four evaluations of the pressure law on each face S_{ij} of the grid can, however, still be unacceptable, we focus on a different approach that severely reduces the computational cost of the scheme.

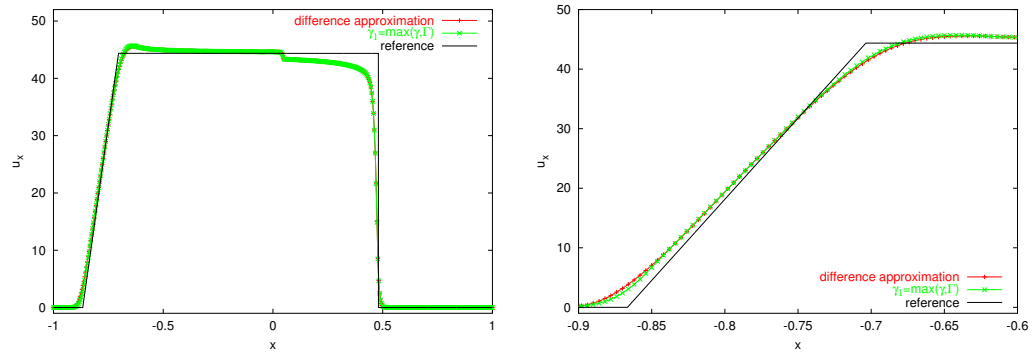
We do not include a thorough study of the finite difference approach since it does not reduce the computational cost enough to be used in our applications. A last resort is the use of a table to store the pressure values. For that reason the study of the influence of tabularized pressure values on the performance of the ER method is important. We interpolate the pressure values given at the nodes of a Cartesian grid using bilinear interpolation. Thus we can compute pressure values for any combination of ρ, ε. This also enables us to compute derivatives cheaply using the derivative of the discrete pressure function p_h. Since we use a Cartesian table, finding the relevant grid cell for a given pair ρ, ε is straight forward and very cheap.

In Figure 7.6 we show results for our Riemann problem using a tabularized version of



(a) error vs. grid resolution

(b) EOC vs. grid resolution



(c) u_x for $-1 \leq x \leq 1$

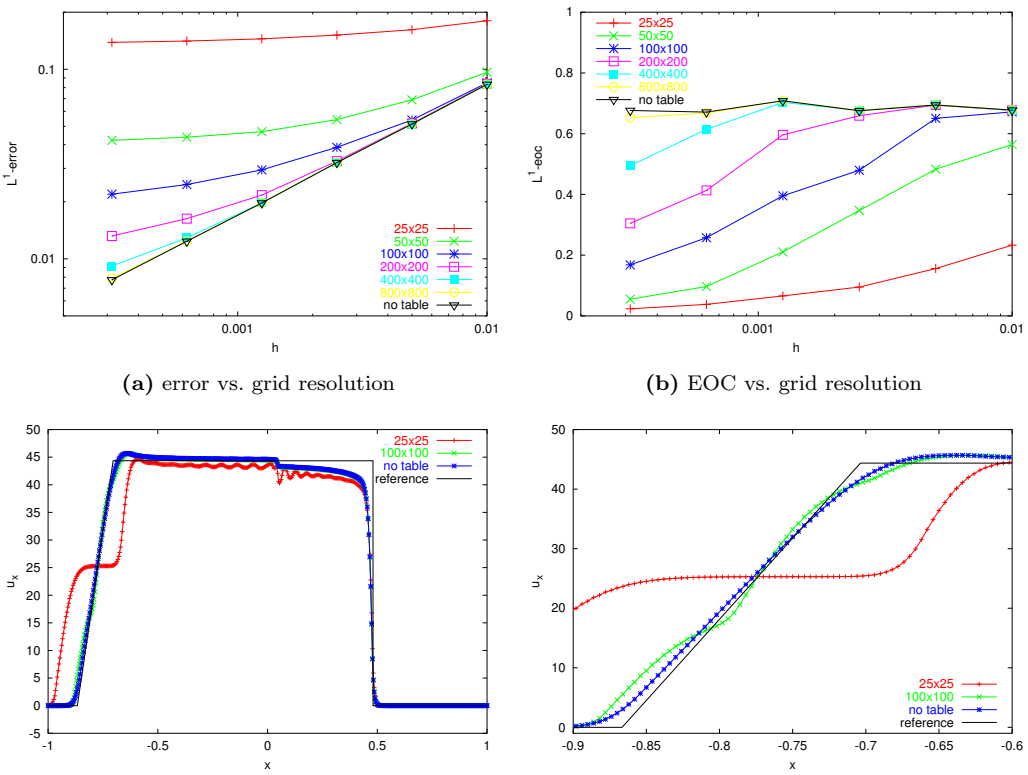
(d) u_x for $-0.9 \leq x \leq -0.6$ (rarefaction wave)

Figure 7.5: Problem: Riemann Problem 6.4(RPWalls2). Method: ER–HLLEM scheme with minmod reconstruction in primitive variables. ER scheme using a finite difference approximation for (7.19) as compared to using (7.19). The finite–difference approximation leads to a stable scheme which is slightly more diffusive than the original scheme but only requires the evaluation of the pressure law and of no derived values.

the van der Waals EOS with different resolutions. The results show that the method works well as long as the resolution of the table is chosen according to the resolution of the spatial grid. The importance of this observation becomes clear in the context of locally adapted grids where the resolution of the spatial grid can be very high and varies strongly in the computational domain. This leads to a high demand on memory to store the pressure values; this greatly reduces the maximum size of a simulation which can still be performed on standard computed systems. Some tests using the Saha equations to define the EOS show that to achieve a relative interpolation error for the discrete pressure function p_h of no more than ten percent requires over one million elements in the pressure table (cf. [DRW02a]).

7.5.3 Efficiency of the ER scheme

After having studied the different aspects of the ER scheme in detail, we conclude our investigations in one space dimension with a study of the efficiency of our method. We



(a) error vs. grid resolution

(b) EOC vs. grid resolution

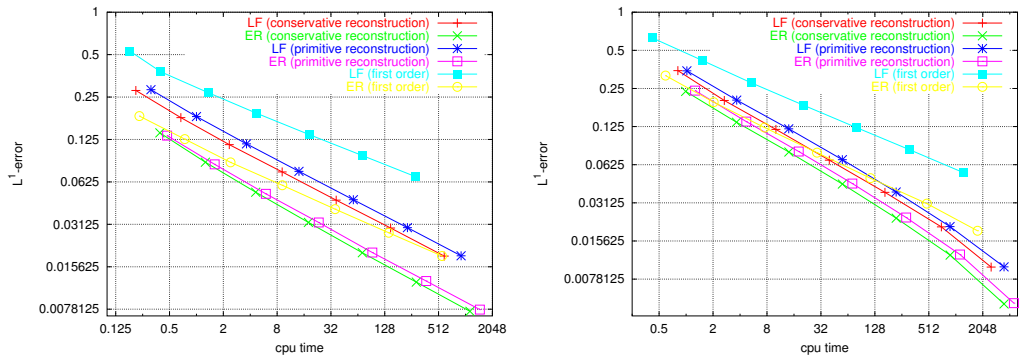
(c) u_x for $-1 \leq x \leq 1$

(d) u_x for $-0.9 \leq x \leq -0.6$ (rarefaction wave)

Figure 7.6: Problem: Riemann Problem 6.4(RPWalls2). Method: ER–HLLEM scheme with minmod reconstruction in primitive variables. ER scheme using a table for the pressure values. Derived values are also computed from the table. The resolution of the table (here in ρ, ε) has to be chosen in accordance with the grid resolution. A table that is too coarse can lead to a totally wrong solution especially in smooth regions (for example, in rarefaction waves).

compare the ER scheme with the standard Lax–Friedrichs (LF) solver (cf. page 68), which we apply directly to the real gas MHD system (1.1). In Figure 7.7 we plot the error to runtime of the first and second order ER scheme and of the LF scheme. For all grid resolutions we see that the second order schemes are more efficient than the corresponding first order schemes and that the reconstruction in conservative variables leads to a smaller error than the reconstruction in primitive variables. From the results in Figure 7.7 it is also very clear that the energy relaxation scheme with the HLLEM flux is much more efficient than the Lax–Friedrichs scheme. Especially the results for the van der Waals Riemann problem (Figure 7.7(a)) demonstrate this very clearly, since in this case even the first order ER scheme is more efficient than the second order LF scheme (at least up to the grid resolution of 6400 points shown here).

Summary of Section 7.5: *In this section we studied a number of different aspects of our MHD solver for the real gas MHD equations. We studied problems in one space dimension but since our finite–volume scheme in higher space dimension greatly relies on the 1d numerical flux function, many of the insights gained here are also applicable*



(a) Riemann problem 6.4(RPWalls2) **(b)** Riemann problem 6.6(RP2D(2,1))

Figure 7.7: Problem: Riemann Problem 6.4(RPWalls2) and 6.6(RP2D(2,1)). Comparison of first and second order schemes using the ER–HLLEM scheme and the LF scheme applied to the real gas MHD system (1.1). The ER scheme is by far more efficient than the LF scheme, although it requires slightly more cpu time on any given grid. On most of the grids used here the first order ER scheme was even more efficient than the second LF scheme.

in the 2d case. Overall we can conclude that the ER scheme is an efficient method for solving the real gas MHD equations. It can be easily adapted to cope with different settings; even very little information concerning the pressure function or approximate pressure values presents no problem. Only the mapping from ρ, ε to p must be implemented in some way.

To use the energy relaxation method a parameter γ_1 has to be chosen. It controls the amount of numerical viscosity in the scheme. We tested a number of different possible choices for defining γ_1. A γ_1 fixed a priori reduces the computational cost of the scheme; at the same time a great deal of a priori knowledge of the structure of the problem is required. If we choose a new γ_1 in every time step and on every face of the grid using (7.19)*, no parameter tuning is required. Computing γ_1 by means of* (7.19) *requires the evaluation of derivatives of the pressure function $p(\rho, \varepsilon)$, which can be expensive or even impossible. In these cases a numerical approximation of the derivatives scarcely reduces the approximation quality of the scheme. Even in the case where the pressure p can be computed for arbitrary ρ and ε, the use of a tabularized pressure function can greatly increase the experimental order of convergence of a scheme. By using a dynamically generated and locally adapted table the computational cost of the scheme can be significantly reduced: the computation of the pressure is reduced to index arithmetic, derivatives can be evaluated cheaply by computing the derivative of the discrete pressure function, and by dint of the local adaptivity the pressure can be computed up to a high accuracy while at the same time memory requirements are kept to a minimum.*

7.6 Numerical Results in 2d

We start our investigation of the 2d scheme with the rotation problem 6.7(RotConst). The solution is stationary and in primitive variables it is independent of the EOS. For

the results shown here we use the van der Waals EOS. The density, the pressure, and the normal velocity are continuous; but the tangential velocity jumps from $u_0 R$ to zero for $r = R$: this type of discontinuity is called a shear discontinuity and is *Kelvin–Helmholtz unstable* as already noted in Section 3.7. In the hydrodynamic case this interface is unstable against perturbations of any wave length. Therefore even very small perturbations of the interface are amplified in time. Since the projection of the initial data onto a grid always leads to a perturbation of the circular interface, we cannot expect grid convergence for this setting. This is confirmed by our numerical experiments in the following. Some numerical studies for the simple setting are published, for example, in [MBR96]. Linear stability analysis of the simple setting of a straight interface shows that a sufficiently strong magnetic field that is tangential to the interface has a stabilizing effect [Cha81]. In our setting we influence the strength of the magnetic field with the free parameter β.

In Figure 7.8 we present some calculations with zero magnetic field ($\beta = 0$) to demonstrate the problems caused by the unstable interface. We show a time series on different refinement levels using the structured macro grid. It can be clearly seen that the approximation is not converging to the stationary solution due to the unstable nature of the Kelvin–Helmholtz interface. This can also be seen if we look at the development of the error during grid refinement in Figure 7.9(a). We emphasize that the problem of the failing convergence of the scheme should be treated with caution. Due to the instability we cannot expect any grid convergence (or at least only at a very high grid resolution). This is not only an academic problem, but it also arises in applications. Thus, in the purely hydrodynamic setting, this is not a good test for studying the performance of a scheme.

We now turn our attention to the case of a stabilizing magnetic field ($\beta > 0$). Again we study the error at different grid resolutions (in Figure 7.9). As macro grid we use the unstructured grid. For β we choose $0.5, 1$, and 2. The results show that on the one hand the approximation error is now clearly reduced through grid refinement so that the scheme seems to converge for all values of $\beta > 0$. But now the problem is that the scheme breaks down due to states $\mathbf{U} \notin \mathcal{U}$, i.e., unphysical values like negative pressure or speed of sound. The larger β (and thus the size of the magnetic field) is, the more problems we have with the breakdown of the simulation. In the case of $\beta = 2$ we could perform the calculations on a grid with 90 elements (the macro grid) and on a grid with 360 elements but not on a grid with 206 or on any finer grid.

It is not trivial to pinpoint the reason for the breakdown of the simulation. It is important to remember that for all choices of β we have the same values at the interface for $\rho, \mathbf{u}$ and p; only the size of the magnetic field $\mathbf{B}$ is influenced by the parameter β. The most probable reason for the stability problems is thus the divergence constraint. Due to the projection of the magnetic field onto the grid the discrete initial data are far from divergence–free and for large β the error in the divergence of $\mathbf{B}$ is very large. To corroborate this assumption we have repeated the simulation using the source term fix described in Section 6.1.2. The results are also shown in Figure 7.9. They show that the convergence rate of the scheme increases with increasing β; this demonstrates the stabilizing effect of the magnetic field on the interface. With this correction mechanism the scheme is stable and performs the simulations on all grids for $\beta = \frac{1}{2}$ and $\beta = 1$. For $\beta = 2$ we can compute the solution on far finer grids than with the base scheme;

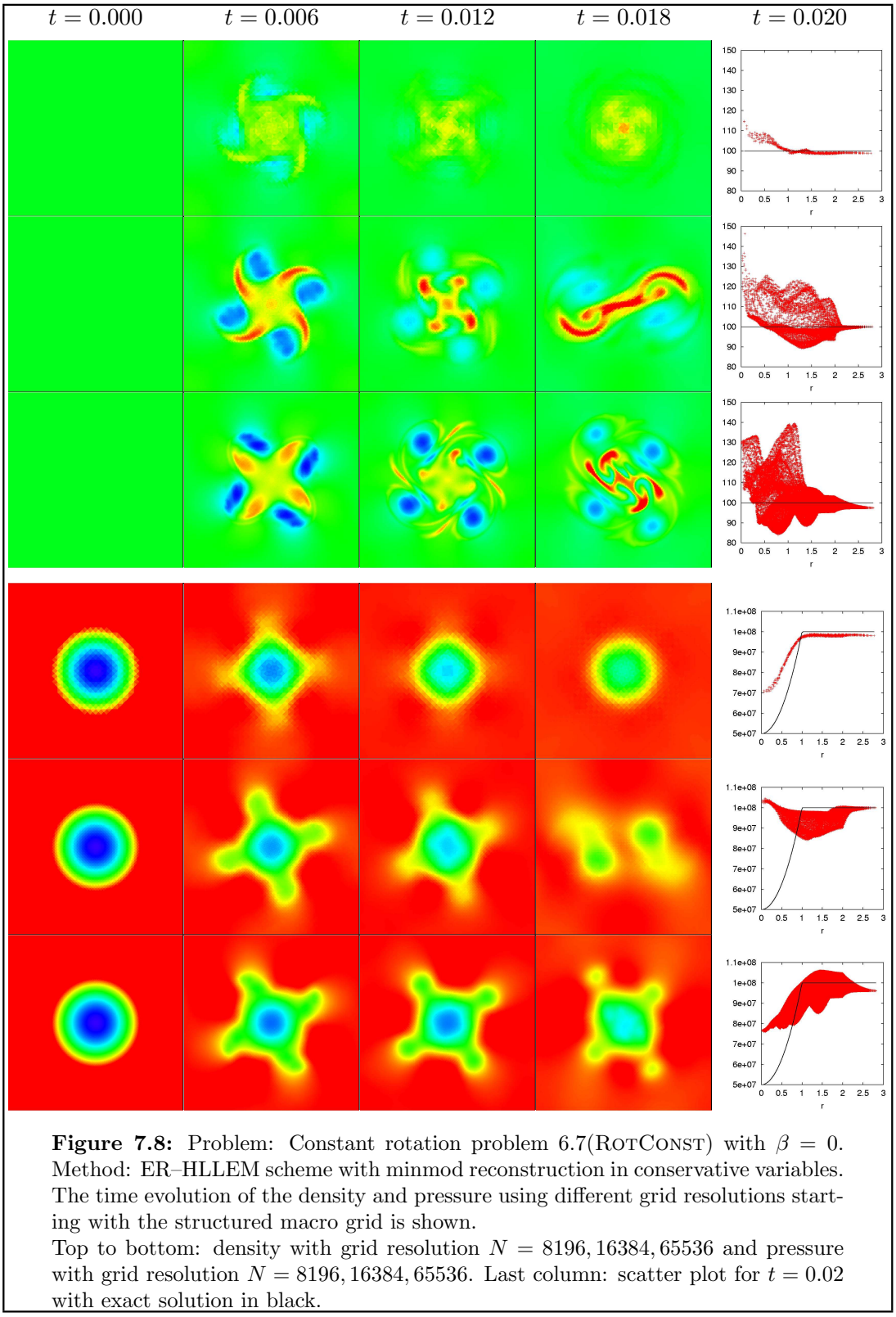
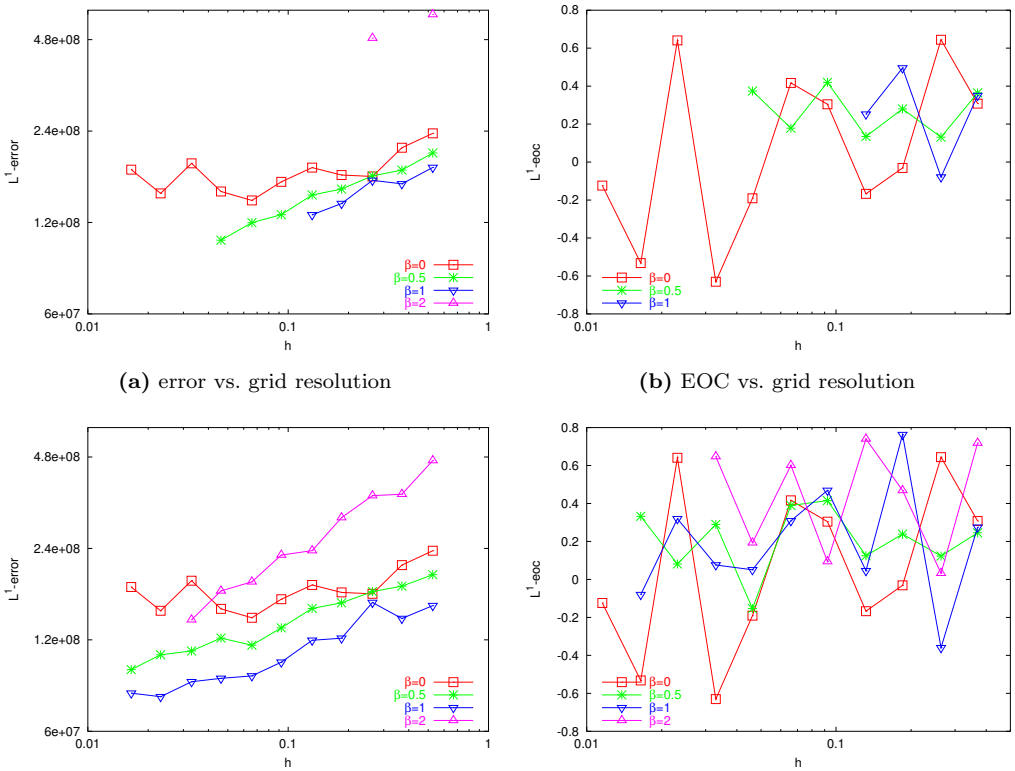


Figure 7.8: Problem: Constant rotation problem 6.7(RotConst) with $\beta = 0$. Method: ER–HLLEM scheme with minmod reconstruction in conservative variables. The time evolution of the density and pressure using different grid resolutions starting with the structured macro grid is shown.
Top to bottom: density with grid resolution $N = 8196, 16384, 65536$ and pressure with grid resolution $N = 8196, 16384, 65536$. Last column: scatter plot for $t = 0.02$ with exact solution in black.



(a) error vs. grid resolution

(b) EOC vs. grid resolution

(c) error vs. grid resolution

(d) EOC vs. grid resolution

Figure 7.9: Problem: Constant rotation problem 6.7(RotConst) with $\beta = 0, 0.5, 1, 2$. Method: ER–HLLEM scheme with minmod reconstruction in conservative variables. Middle row: results using the base scheme. Bottom row: results using the source term fix as described in Section 6.1.2. The source term fix leads to a stabilization of the base scheme so that we can compute approximations with higher resolution. But note that the order of convergence is unsatisfactory; for $\beta = 0$ the scheme does not converge due to the unstable interface.

but with this correction as well the scheme breaks down at high grid resolutions. Due to the obvious stability problems of our scheme resulting from a violation of the divergence constraint, we first derive methods for reducing the divergence errors before we continue our study of the 2d scheme.

Summary of Section 7.6: *Our results from simulations in two space dimensions confirm the results from our tests in one space dimension. The ER scheme leads to an accurate approximation of the MHD system with a complex equation of state. At the same time our results demonstrated that the finite–volume scheme derived so far suffers from stability problems that are caused by the violation of the divergence constraint* (1.1e)*. Using the source term approach described in Section 6.1.2, these problems are reduced; the overall performance of the scheme is, nevertheless, still not satisfactory. Therefore in the following chapter, we derive an extension of our base scheme that greatly increases the stability of the scheme.*

Chapter 8

Divergence Constraint: the GLM–MHD Scheme

In this chapter we study a new method for reducing the problems of divergence errors in MHD simulations. This approach was introduced in [MOS+00] for the Maxwell equations and we extended it to the MHD equations in [DKK+02, DRW02b]. The method is based on an extension of the MHD system that we call the GLM–MHD system. The divergence constraint is coupled with the induction equation by means of an auxiliary function ψ, and the structure of the modified divergence constraint can be varied, leading to a variety of different correction mechanisms.
The GLM–MHD system consists of equations (1.1a), (1.1b), and (1.1d). The divergence constraint (1.1e) is coupled with the induction equation (1.1c) by introducing a new auxiliary function ψ and a linear differential operator $\mathcal{D}$

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0 \qquad \text{(conservation of mass)}, \tag{8.1a}$$
$$\partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u}\mathbf{u}^T + \mathcal{P}) = 0 \qquad \text{(conservation of momentum)}, \tag{8.1b}$$
$$\partial_t \mathbf{B} + \nabla \cdot (\mathbf{u}\mathbf{B}^T - \mathbf{B}\mathbf{u}^T) + \nabla\psi = 0 \qquad \text{(modified induction equation)}, \tag{8.1c}$$
$$\partial_t (\rho e) + \nabla \cdot (\rho e \mathbf{u} + \mathcal{P}\mathbf{u}) = 0 \qquad \text{(conservation of energy)}. \tag{8.1d}$$
$$\mathcal{D}(\psi) + \nabla \cdot \mathbf{B} = 0 \qquad \text{(modified divergence constraint)}. \tag{8.1e}$$

The gravity source terms can be directly added as in (1.1). Our aim is to choose $\mathcal{D}$ and the initial and boundary data for ψ in such a way that a numerical approximation to (8.1) is a good approximation to the original system (1.1). Note that independent of the choice for $\mathcal{D}$ the quantities $\rho, \rho\mathbf{u}, \mathbf{B}, \rho e$ are conservative quantities for the GLM–MHD system if we choose suitable boundary conditions for ψ. Furthermore, if the initial conditions for $\mathbf{B}$ are divergence–free and the initial data ψ_0 for ψ is zero, then ψ remains zero for all time and the original induction equation (1.1c) is recovered. Thus, under suitable initial and boundary conditions on ψ the GLM–MHD system is equivalent to the original MHD system if $\nabla \cdot \mathbf{B}_0 \equiv 0$.
In the following section we study different choices for the linear operator $\mathcal{D}$ according to the suggestions in [DKK+02]. In Section 8.2 we then extend our finite–volume scheme to approximate the GLM–MHD system. The definition of the free parameters is discussed in Section 8.3. The last section is then devoted to some numerical examples

that demonstrate the possibilities of our method. In [DKK+02] we focused on the *hyperbolic* and the *mixed* approach ($\mathcal{D} = \frac{1}{c_h^2}\partial_t$, $\mathcal{D} = \frac{1}{c_h^2}\partial_t + \frac{1}{c_p^2}$, respectively) since these seemed to be the most promising ones. The *elliptic* correction ($\mathcal{D} = 0$) leads to the well known Hodge projection scheme and the *parabolic* approach ($\mathcal{D} = \frac{1}{c_p^2}$) seemed less likely to lead to enough damping of divergence errors. In this chapter we also discuss the implementation of the parabolic approach. A further version that was only briefly discussed in [DKK+02] leads to a Galilean invariant form of the GLM–MHD system. This is also discussed in more detail here. For comparison we include numerical results for the elliptic approach and the source term approach (cf. page 72).

8.1 Analytical Motivation

We first derive some simple equations that follow directly from the GLM–MHD system. By applying ∂_t to (8.1e) and the divergence operator to (8.1c) we obtain for sufficiently smooth solutions:

$$\partial_t(\nabla \cdot \mathbf{B}) + \Delta\psi = 0 \ , \quad \partial_t \mathcal{D}(\psi) + \partial_t(\nabla \cdot \mathbf{B}) = 0 \ .$$

By applying the operator $\mathcal{D}$ to the first equation and the Laplace operator to (8.1e) we arrive at

$$\partial_t(\mathcal{D}(\nabla \cdot \mathbf{B})) + \Delta(\mathcal{D}(\psi)) = 0 \ , \quad \Delta(\mathcal{D}(\psi)) + \Delta(\nabla \cdot \mathbf{B}) = 0 \ .$$

These equations lead us to the following equations for $\nabla \cdot \mathbf{B}$ and ψ:

$$\partial_t \mathcal{D}(\nabla \cdot \mathbf{B}) - \Delta(\nabla \cdot \mathbf{B}) = 0 \ , \tag{8.2a}$$
$$\partial_t \mathcal{D}(\psi) - \Delta\psi = 0 \ , \tag{8.2b}$$

i.e., $\nabla \cdot \mathbf{B}$ and ψ satisfy the same equation for any choice of $\mathcal{D}$. These equations are used in the following to determine the type of the correction at the PDE level for the different choices for $\mathcal{D}$.
In one space dimension the modified equation for the first component of the magnetic field B_x and the equation for ψ are

$$\partial_t B_x + \partial_x \psi = 0 \ ,$$
$$\mathcal{D}(\psi) + \partial_x B_x = 0 \ .$$

These two equations are decoupled from the remaining seven equations, which are identical to the original MHD system in one space dimension (with the parameter B_x). Note that in the case of the GLM–MHD system B_x is no longer a constant but can vary in space and time.
In the following we present a number of different possible choices for the linear operator $\mathcal{D}$ and study the impact on the evolution of divergence errors. We start out by recalling some properties of the original MHD system. The magnetic field $\mathbf{B}$ has to satisfy

$$\partial_t \mathbf{B} + \nabla \times (\mathbf{B} \times \mathbf{u}) = 0 \qquad \text{in} \quad \Omega \times \mathbb{R}^+, \tag{8.3a}$$
$$\nabla \cdot \mathbf{B} = 0 \qquad \text{in} \quad \Omega \times \mathbb{R}^+, \tag{8.3b}$$

$$\mathbf{B}(\cdot,0) = \mathbf{B}_0(\cdot) \quad \text{in} \quad \Omega \tag{8.3c}$$

with $\Omega \subset \mathbb{R}^3$. Here $\mathbf{B}_0$ denotes the initial data for the magnetic field. Since $\nabla\cdot(\nabla\times\cdot) \equiv 0$ we have $\partial_t(\nabla\cdot\mathbf{B}) = 0$. This means that $\nabla\cdot\mathbf{B}(\cdot,0) \equiv 0$ implies $\nabla\cdot\mathbf{B}(\cdot,t) \equiv 0$ for all $t > 0$. Thus, the divergence constraint (8.3b) is a condition for the initial data. However, in numerical simulations (8.3b) can be violated. Therefore, we have to study the evolution of $\mathbf{B}$ in the case where the initial conditions have the form $\mathbf{B}(\cdot,0) \equiv \mathbf{B}_0(\cdot) + \mathbf{b}(\cdot)$ with $\nabla\cdot\mathbf{B}_0 = 0$ but $\nabla\cdot\mathbf{b} \neq 0$. (The perturbation $\mathbf{b}$ is, for example, due to discretization errors.) In this case the evolution equations (8.3a) yield $\nabla\cdot\mathbf{B}(\cdot,t) \equiv \nabla\cdot\mathbf{b}(\cdot)$ for all $t > 0$. Hence (8.3b) is violated for any $t > 0$.
In one space dimension the evolution equation for $B := B_x$ in (8.3a) is decoupled from the other equations. Equation (8.3b) reduces to $\partial_x B = 0$ and the evolution equation for B reads $\partial_t B = 0$. As before, $\partial_x B(\cdot,0) \equiv 0$ leads to $\partial_x B(\cdot,t) \equiv 0$ for all $t > 0$. Therefore B_0, i.e. the divergence–free part of the initial data, is a constant. In order to study the evolution of divergence errors, we therefore have to consider the following equations:

$$\partial_t B = 0 \qquad \text{in} \quad [x_l, x_r] \times \mathbb{R}^+, \tag{8.4a}$$
$$\partial_x B = 0 \qquad \text{in} \quad [x_l, x_r] \times \mathbb{R}^+, \tag{8.4b}$$
$$B(\cdot,0) = B_0 + b(\cdot) \quad \text{in} \quad [x_l, x_r]\ , \tag{8.4c}$$
$$B(x_l,t) = B(x_r,t) = B_0 \qquad \text{in} \quad \mathbb{R}^+. \tag{8.4d}$$

with $b' \neq 0$. These equations reproduce the situation in 3d: (8.4a) and (8.4c) result in $B(\cdot,t) \equiv B_0 + b(\cdot)$ for all $t > 0$, which implies $\partial_x B = b' \neq 0$ for all time. Thus (8.4b) is uniformly violated. In the following we show that the situation is substantially improved if we use the GLM–MHD equations instead of the original system. Let $B = B_x$ and ψ denote the solution obtained from the one–dimensional GLM–MHD system for the initial conditions (8.4c):

$$\partial_t B + \partial_x \psi = 0 \qquad \text{in} \quad [x_l, x_r] \times \mathbb{R}^+, \tag{8.5a}$$
$$\mathcal{D}(\psi) + \partial_x B = 0 \qquad \text{in} \quad [x_l, x_r] \times \mathbb{R}^+, \tag{8.5b}$$
$$B(\cdot,0) = B_0 + b(\cdot) \quad \text{in} \quad [x_l, x_r]\ , \tag{8.5c}$$
$$B(x_l,t) = B(x_r,t) = B_0 \qquad \text{in} \quad \mathbb{R}^+. \tag{8.5d}$$

For simplicity and without loss of generality we set $x_l = 0$ and $x_r = 2\pi$. We assume that b is a smooth function defined on $[0, 2\pi]$ with $b(0) = b(2\pi) = 0$ which can be written as a finite Fourier sum

$$b(x) = \sum_{k=1}^{n} \alpha_k \sin(kx) \tag{8.6}$$

with constants $\alpha_k \in \mathbb{R}$. Parts of the analysis can be extended to a more general setting (cf. [DKK$^+$02]) but for simplicity we restrict ourselves to this case. Note that clearly we have $b' \neq 0$.
In the rest of this section we present different versions of the linear operator $\mathcal{D}$ and study the resulting solutions B, ψ of the model problem given above. We show that if

the boundary and initial conditions for ψ are suitably chosen, we find $|\partial_t B(x,t)| \to 0$, $|\partial_x B(x,t)| \to 0$, and $|B(x,t) - B_0| \to 0$ as $t \to \infty$. Thus for large times, (8.4a) and (8.4b) hold at least approximately.

8.1 Remark: *We decide not to modify the boundary conditions for B since these are given by the original problem. In some cases it might be advantageous to change these conditions as well. In the case where the magnetic field is not divergence–free we are not in the physical regime and the physical boundary conditions have no meaning. We have not experimented with this approach.*

The Elliptic Approach

The simplest choice for the linear operator $\mathcal{D}$ is given by

$$\mathcal{D}(\psi) = 0 \ . \tag{8.7}$$

Since due to (8.2a) $\nabla \cdot \mathbf{B}$ satisfies the Laplace equation, we call the resulting system the *elliptic* GLM–MHD system. We have shown in [DKK$^+$02] that in this case the well–known Hodge projection scheme is rediscovered (see, for example, [BB80, Tót00]). More details on the implementation, boundary conditions, and some numerical results can be found in [Wes02a] and we only include this method as a comparison scheme in our numerical tests.

The Parabolic Approach

If we choose

$$\mathcal{D}(\psi) = \frac{1}{c_p^2}\psi \tag{8.8}$$

with $c_p \in \mathbb{R}^+$ then, due to (8.2a), the divergence of $\mathbf{B}$ satisfies the heat equation. Therefore we call this approach the *parabolic* GLM–MHD method. Since equation (8.1e) leads to $\psi = -c_p^2 \nabla \cdot \mathbf{B}$ we can eliminate the auxiliary function ψ from the GLM–MHD system. This leads to a second order term in equation (8.1c) for $\mathbf{B}$:

$$\partial_t \mathbf{B} + \nabla \cdot \left(\mathbf{u}\mathbf{B}^T - \mathbf{B}\mathbf{u}^T\right) = c_p^2 \nabla\big(\nabla \cdot \mathbf{B}\big) \ .$$

Since ψ can also be eliminated from our model problem (8.5), the question of boundary and initial conditions for ψ does not arise. The only free parameter remaining is the constant c_p. Its influence on the solution to our model problem is easily studied.

8.2 Theorem (Model Problem for Parabolic GLM–MHD System)
Let B and $\psi = -\frac{1}{c_p^2}\partial_x B$ be the solution of our model problem (8.5) with $\mathcal{D}$ given by (8.8). Then $|B(x,\cdot) - B_0|, |\partial_t B(x,\cdot)|$, and $|\partial_x B(x,\cdot)|$ are monotone decreasing and decay exponentially fast to zero for $t \to \infty$.

Proof:
The statements follow directly from the fact that the function

$$B(x,t) = B_0 + \sum_{k=1}^{n} \alpha_k \sin(kx) e^{-c_p^2 k^2 t}$$

together with $\psi(x,t) = -c_p^2 \partial_x B(x,t)$ is a solution to our model problem. □

The choice of the free parameter c_p directly influences the rate of decay, which is equal to $c_p^2 t$. Due to the fast decay of the divergence errors and the simple structure of the correction this approach seems promising. The drawback is the loss of the first order, hyperbolic structure of the MHD system and the necessity of computing second order derivatives; this leads to severe stability restrictions on the time step Δt. Note that we formally recover the elliptic correction for $c_p \to \infty$ — in this case $B(x,t) = B_0$ for all $t > 0$.

The Hyperbolic Approach

If we choose

$$\mathcal{D}(\psi) = \frac{1}{c_h^2} \partial_t \psi \tag{8.9}$$

with $c_h \in \mathbb{R}^+$, then equation (8.2a) is the wave equation; consequently, we call this approach the *hyperbolic* GLM–MHD method. The eigensystem of the resulting GLM–MHD system was studied in [DKK+02], where we showed that the system is hyperbolic. The observation that in one space dimension the equation for B_x and ψ decouple from the other equations allows us to obtain a right eigenvector $\mathbf{r}$ from the eigenvectors of the original MHD system: Let $\mathbf{r}' \in \mathbb{R}^7$ be a right eigenvector for the original MHD system to the eigenvalue λ, then we obtain a right eigenvector for the hyperbolic GLM–MHD system to the same eigenvalue by extending $\mathbf{r}'$ by two zero entries: $\mathbf{r} := (r'_1, \ldots, r'_4, 0, r'_5, \ldots, r'_7, 0)^T$. Moreover, also due to the decoupling, the system has the additional eigenvalues $-c_h$ and c_h. These are distinct from the MHD eigenvalues if c_h is sufficiently large. Hence we find nine linearly independent right eigenvectors and the eigenvalues read in non–decreasing order

$$\begin{aligned} &\lambda_1 = -c_h, \quad \lambda_2 = u_x - c_f \ , \quad \lambda_3 = u_x - c_a \ , \quad \lambda_4 = u_x - c_s \ , \\ &\lambda_5 = u_x \ , \quad \lambda_6 = u_x + c_s \ , \quad \lambda_7 = u_x + c_a \ , \quad \lambda_8 = u_x + c_f \ , \quad \lambda_9 = c_h \ . \end{aligned} \tag{8.10}$$

Therefore the GLM–MHD system with $\mathcal{D}$ given by (8.9) is hyperbolic. Furthermore, the structure of the right eigenvectors shows that only the two additional waves traveling with speeds $\pm c_h$ carry a change in B_x or ψ. In Section 1.1 we have seen that in addition to the seven eigenvalues, which are symmetric to the fluid velocity u_x, we also have an eigenvalue λ_{div} corresponding to the equation for the first magnetic field component; this eigenvalue is always equal to zero. In the source term approach this eigenvalue is shifted to u_x. In our approach we replace this eigenvalue with two new waves, which are symmetric to zero. In the following we call these waves *divergence waves* and denote them with $\lambda_{\text{div}\pm} = \pm c_h$.

We now turn our attention to our model problem. In this case we have to prescribe boundary and initial conditions for ψ. We start our analysis using the simplest choice of homogenous conditions.

8.3 Theorem (Model Problem for Hyperbolic GLM–MHD System)
Consider problem (8.5) *with* $\mathcal{D}$ *as in* (8.9) *and initial and boundary data for* ψ *given by*

$$\psi(0,t) = \psi(2\pi,t) = 0 \qquad \psi(\cdot,0) = 0 \ .$$

Let B be the solution to our model problem (8.5). *Then the limits for $t \to \infty$ of $|\partial_t B|$ and $|\partial_x B|$ do not exist.*

Proof:
Under the initial and boundary conditions given in the theorem the statement follows directly since the solution to our model problem (8.5) is

$$B(x,t) = B_0 + \sum_{k=1}^{n} \alpha_k \sin(kx)\cos(c_h k t) \ ,$$
$$\psi(x,t) = -c_h \sum_{k=1}^{n} \alpha_k \cos(kx)\sin(c_h k t) \ .$$

□

With these boundary conditions on ψ we have no decay of the divergence errors for all choices of c_h. One reason is that we decided to retain the boundary conditions on B, which do not allow divergence errors to be transported out of the computational domain. By studying the corresponding Cauchy problem to (8.5) we can reach a better understanding of the mechanism behind the hyperbolic correction.

8.4 Theorem (Cauchy Problem for Hyperbolic GLM–MHD System)
Consider the problem (8.5a)–(8.5c) *in $\mathbb{R} \times \mathbb{R}^+$ with $\mathcal{D}$ given by* (8.9). *Assume that the initial data b is a smooth function with compact support in $[0, 2\pi]$. Let $\psi(x,0) = 0$ for $x \in \mathbb{R}$. The solution B then satisfies the inequalities*

$$|B(x,t) - B_0| \le \|b\|_\infty \qquad \text{for} \quad x \in [0,2\pi] \quad \text{and} \quad 0 \le t < \frac{1}{2c_h} \ ,$$
$$|B(x,t) - B_0| \le \frac{1}{2}\|b\|_\infty \qquad \text{for} \quad x \in [0,2\pi] \quad \text{and} \quad \frac{1}{2c_h} \le t < \frac{1}{c_h} \ ,$$
$$B(x,t) = B_0 \qquad \text{for} \quad x \in [0,2\pi] \quad \text{and} \quad \frac{1}{c_h} \le t \ .$$

It follows that

$$\partial_t B(x,t) = 0 \quad \text{and} \quad \partial_x B(x,t) = 0 \quad \text{for} \quad x \in [0,2\pi] \quad \text{and} \quad t \ge \frac{1}{c_h} \ .$$

Proof:
Using the standard representation formula for solutions to systems of linear conservation laws, the function B, ψ are given by

$$B(x,t) = B_0 + \frac{1}{2}\big(b(x + c_h t) + b(x - c_h t)\big) \ ,$$
$$\psi(x,t) = -c_h\big(b(x + c_h t) - b(x - c_h t)\big) \ .$$

The statements follow since $b(x) = 0$ for $x \le 0$ and $x \ge 2\pi$. □

In this case our goal of reducing divergence errors in the initial conditions is achieved even in finite time (depending on the choice of c_h). The disturbances in the divergence–free part of the magnetic field are transported out of the computational domain $[0, 2\pi]$. This behavior is strongly linked to the fact that we have neglected the boundary conditions on B, which we have so far assumed to be fixed by the original problem. Note that for $c_h \to 0$ we recover the original MHD equations and with $c_h \to \infty$ the elliptic approach. This is reflected in the statements of Theorem 8.4.

Galilean Invariance

One disadvantage of the hyperbolic GLM–MHD system is that it is not Galilean invariant. This can be seen from the structure of the eigenvalues given in (8.10). The two new *divergence* waves $\lambda_{\text{div}\pm}$ are not symmetric with respect to the fluid velocity u_x as is the case for all the other eigenvalues. It is possible to modify the GLM–MHD system by adding some non–conservative terms to the modified induction equation (8.1c) and the modified divergence constraint (8.1e) so that the resulting system is Galilean invariant:

$$\partial_t \mathbf{B} + \nabla \cdot \left(\mathbf{u}\mathbf{B}^T - \mathbf{B}\mathbf{u}^T\right) + \nabla\psi = -\mathbf{u}\nabla\cdot\mathbf{B} \,, \tag{8.11a}$$

$$\partial_t \psi + c_h^2 \nabla \cdot \mathbf{B} = -\mathbf{u}\cdot\nabla\psi \,. \tag{8.11b}$$

The proof that this system is Galilean invariant is sketched in [DKK+02]. Due to the new terms the two divergence waves are now $\lambda_{\text{div}\pm} = u_x \pm c_h$. The eigenvalues of this system are thus

$$\lambda_1 = u_x - c_h, \quad \lambda_2 = u_x - c_f \,, \quad \lambda_3 = u_x - c_a \,, \quad \lambda_4 = u_x - c_s \,,$$
$$\lambda_5 = u_x \,, \quad \lambda_6 = u_x + c_s \,, \quad \lambda_7 = u_x + c_a \,, \quad \lambda_8 = u_x + c_f \,, \quad \lambda_9 = u_x + c_h \,.$$

It is important to note that in this system the magnetic field $\mathbf{B}$ is no longer conserved if $\nabla \cdot \mathbf{B} \neq 0$. As in the hyperbolic case we study the Cauchy problem for our model problem. For simplicity we assume that the velocity is constant.

8.5 Theorem (Cauchy Problem for Galilean Invariant GLM–MHD System) *Consider the solution B, ψ of the Cauchy problem*

$$\begin{aligned}
\partial_t B + \partial_x \psi &= -u\partial_x B && \text{in} \quad \mathbb{R}\times\mathbb{R}^+,\\
\partial_t \psi + c_h^2 \partial_x B &= -u\partial_x \psi && \text{in} \quad \mathbb{R}\times\mathbb{R}^+,\\
B(\cdot,0) &= B_0 + b(\cdot) && \text{in} \quad \mathbb{R}\,,\\
\psi(\cdot,0) &= 0 && \text{in} \quad \mathbb{R}
\end{aligned}$$

with smooth b having compact support in $[0,2\pi]$ and constants $c_h \in \mathbb{R}^+$ and $u \in \mathbb{R}$. Then B satisfies

$$\begin{aligned}
|B(x,t) - B_0| &\leq \|b\|_\infty && \text{for} \quad x\in[0,2\pi] \quad \text{and} \quad 0 \leq t < \frac{1}{2\bar{\lambda}_{\text{div}}}\,,\\
|B(x,t) - B_0| &\leq \frac{1}{2}\|b\|_\infty && \text{for} \quad x\in[0,2\pi] \quad \text{and} \quad \frac{1}{2\bar{\lambda}_{\text{div}}} \leq t < \frac{1}{\bar{\lambda}_{\text{div}}}\,,\\
B(x,t) &= B_0 && \text{for} \quad x\in[0,2\pi] \quad \text{and} \quad \frac{1}{\bar{\lambda}_{\text{div}}} \leq t
\end{aligned}$$

with $\bar{\lambda}_{\text{div}} := |u| + c_h$. Therefore it follows that

$$\partial_t B(x,t) = 0 \quad \text{and} \quad \partial_x B(x,t) = 0 \quad \text{for} \quad x\in[0,2\pi] \quad \text{and} \quad t \geq \frac{1}{\bar{\lambda}_{\text{div}}}\,.$$

Proof:
As before we use the standard representation formula to obtain B, ψ

$$B(x,t) = B_0 + \frac{1}{2}\big(b(x-(u-c_h)t) + b(x-(u+c_h)t)\big) \ ,$$
$$\psi(x,t) = -c_h\big(b(x-(u-c_h)t) - b(x-(u+c_h)t)\big) \ .$$

The statements follow since b has compact support in $[0, 2\pi]$. □

The Mixed Approach

In the last approach we combine the transport of divergence errors present in the hyperbolic approach with the damping property of the parabolic approach. This *mixed* approach has the advantage of retaining the hyperbolic structure of the MHD system, while simultaneously divergence errors are reduced in time even in the case where the original boundary conditions on the magnetic field are used. Given $c_h, c_p \in \mathbb{R}^+$ we choose

$$\mathcal{D}(\psi) = \frac{1}{c_h^2}\partial_t\psi + \frac{1}{c_p^2}\psi \ . \tag{8.12}$$

With this choice of the linear operator, equation (8.2a) is the *telegraph equation*

$$\partial_{tt}^2(\nabla\cdot\mathbf{B}) + \frac{c_h^2}{c_p^2}\partial_t(\nabla\cdot\mathbf{B}) - c_h^2\triangle(\nabla\cdot\mathbf{B}) = 0 \ .$$

Since this approach seems the most promising one, we extend the study of our model problem to the case of more than one space dimension.

8.6 Theorem (Model Problem for Mixed GLM–MHD System)
There exists a smooth solution B, ψ of the model problem (8.5) *with $\mathcal{D}$ given by* (8.12) *and with homogeneous Dirichlet boundary conditions and zero initial conditions for ψ. The function B satisfies: $|B(x,t) - B_0| \to 0$, $|\partial_t B(x,t)| \to 0$, and $|\partial_x B(x,t)| \to 0$ for $t \to \infty$.*
Furthermore let $\beta := \nabla\cdot\mathbf{B}$ and ψ be the solution of the system

$$\partial_t\beta + \triangle\psi = 0 \qquad \text{in} \quad \Omega\times\mathbb{R}^+, \tag{8.13a}$$
$$\partial_t\psi + \frac{c_h^2}{c_p^2}\psi + c_h^2\beta = 0 \qquad \text{in} \quad \Omega\times\mathbb{R}^+, \tag{8.13b}$$
$$\beta(\cdot,0) = \nabla\cdot\mathbf{B}_0 \quad \text{in} \quad \Omega \ , \tag{8.13c}$$
$$\psi(\cdot,0) = 0 \qquad \text{in} \quad \Omega \ , \tag{8.13d}$$
$$\beta(\mathbf{x},t) = 0 \qquad \text{on} \quad \partial\Omega\times\mathbb{R}^+ \ , \tag{8.13e}$$
$$\psi(\mathbf{x},t) = 0 \qquad \text{on} \quad \partial\Omega\times\mathbb{R}^+ \tag{8.13f}$$

in a domain $\Omega \subset \mathbb{R}^n$ where $\nabla\cdot\mathbf{B}_0$ is a smooth function. Then we have for all $\mathbf{x} \in \Omega$ that $|\beta(\mathbf{x},t)| \to 0$ with an exponential rate of decay which is largest if the constants c_p, c_h satisfy

$$c_p^2 > \frac{c_h}{c_{\text{rel}}} \tag{8.14}$$

Here $c_{\rm rel} := 2\sqrt{\lambda_{min}}$ *where* λ_{min} *is the smallest eigenvalue of the Laplace operator in the domain* Ω *with Dirichlet boundary conditions.*

8.7 Remark: *The system* (8.13) *is derived from the GLM–MHD equations with* $\mathbf{B}_0$ *as initial conditions for* $\mathbf{B}$ *by taking the divergence of* (8.1c) *and using* (8.12)*. In the case of more than one space dimension we can only derive a simple system for* $\nabla\cdot\mathbf{B}$ *and* ψ*; therefore we can only prove the decay of* $\nabla\cdot\mathbf{B}$ *and, in contrast to the situation in one space dimension, we cannot prove any results for the magnetic field itself.*

Proof:
We first derive the general solution of the telegraph equation

$$\partial_{tt}^2 u + \frac{c_h^2}{c_p^2}\partial_t u - c_h^2 \triangle u = 0 \qquad \text{for } \mathbf{x}\in\Omega, t>0\ , \tag{8.15}$$

with smooth initial data and Dirichlet boundary conditions

$$\begin{aligned} u(\mathbf{x},0) &= u_0(\mathbf{x}) && \text{for } \mathbf{x}\in\Omega\ ,\\ \partial_t u(\mathbf{x},0) &= 0 && \text{for } \mathbf{x}\in\Omega\ ,\\ u(\mathbf{x},t) &= 0 && \text{for } \mathbf{x}\in\partial\Omega, t>0\ . \end{aligned}$$

We want to set $u=\beta$; therefore we have $u_0 = \nabla\cdot\mathbf{B}_0$, and equation (8.13a) together with (8.13d) lead to the initial conditions for the time derivative. Assume that we can write $u_0(\mathbf{x}) = \sum_{k=1}^m \alpha_k w_k(\mathbf{x})$ where w_k are the eigenfunctions of the Laplace operator in Ω with homogenous boundary conditions, i.e. $-\triangle w_k = \lambda_k w_k$ and $w_k = 0$ on $\partial\Omega$. We make the ansatz $u(\mathbf{x},t) = \sum_{k=1}^m \alpha_k v_k(t) w_k(\mathbf{x})$ with some functions v_k. Inserting this ansatz into the telegraph equations leads to an ODE for each v_k:

$$v_k'' + \frac{c_h^2}{c_p^2} v_k' + c_h^2\lambda_k v_k = 0.$$

Defining $\overline{v_k}(t) := v_k(t)\exp\left(\frac{c_h^2}{2c_p^2}t\right)$ it follows that

$$v_k'(t) = \overline{v_k}'(t)\exp\left(\frac{c_h^2}{2c_p^2}t\right) - \frac{c_h^2}{2c_p^2}\overline{v_k}(t)\exp\left(\frac{c_h^2}{2c_p^2}t\right)\ ,$$

and

$$v_k''(t) = \overline{v_k}''(t)\exp\left(\frac{c_h^2}{2c_p^2}t\right) - \frac{c_h^2}{2c_p^2}\overline{v_k}'(t)\exp\left(\frac{c_h^2}{2c_p^2}t\right) - \frac{c_h^2}{2c_p^2}v_k'(t)\ .$$

Inserting this into the telegraph equation leads to

$$\begin{aligned} 0 &= v_k''(t) + \frac{c_h^2}{c_p^2}v_k'(t) + c_h^2\lambda_k v_k(t)\\ &= \overline{v_k}''(t)\exp\left(\frac{c_h^2}{2c_p^2}t\right) - \frac{c_h^2}{2c_p^2}\overline{v_k}'(t)\exp\left(\frac{c_h^2}{2c_p^2}t\right) - \frac{c_h^2}{2c_p^2}v_k'(t) + \frac{c_h^2}{c_p^2}v_k' + c_h^2\lambda_k v_k \end{aligned}$$

$$
\begin{aligned}
&= \overline{v_k}''(t)\exp\Big(\frac{c_h^2}{2c_p^2}t\Big) - \frac{c_h^2}{2c_p^2}\overline{v_k}'(t)\exp\Big(\frac{c_h^2}{2c_p^2}t\Big) + \\
&\quad \frac{c_h^2}{2c_p^2}\Big(\overline{v_k}'(t)\exp\Big(\frac{c_h^2}{2c_p^2}t\Big) - \frac{c_h^2}{2c_p^2}\overline{v_k}(t)\exp\Big(\frac{c_h^2}{2c_p^2}t\Big)\Big) + c_h^2\lambda_k\overline{v_k}(t)\exp\Big(\frac{c_h^2}{2c_p^2}t\Big) \\
&= \overline{v_k}''(t)\exp\Big(\frac{c_h^2}{2c_p^2}t\Big) + \overline{v_k}(t)\exp\Big(\frac{c_h^2}{2c_p^2}t\Big)\Big(c_h^2\lambda_k - \frac{c_h^4}{4c_p^4}\Big)
\end{aligned}
$$

Therefore $\overline{v_k}$ satisfies

$$
\overline{v_k}'' + \Big(c_h^2\lambda_k - \frac{c_h^4}{4c_p^4}\Big)\overline{v_k}(t) = 0.
$$

The solution to this second order ODE can be easily computed and depends on the sign of $c_h^2\lambda_k - \frac{c_h^4}{4c_p^4}$. For $|v_k(t)|$ we arrive at the following estimate

$$
|v_k(t)| \leq \begin{cases} \exp\Big(-\frac{c_h^2}{2c_p^2}t\Big), & c_h^2\lambda_k \geq \frac{c_h^4}{4c_p^4} \\ |\cosh(q_k t)|\exp\Big(-\frac{c_h^2}{2c_p^2}t\Big), & c_h^2\lambda_k < \frac{c_h^4}{4c_p^4} \end{cases} \tag{8.16}
$$

with $q_k := \sqrt{\Big|c_h^2\lambda_k - \frac{c_h^4}{4c_p^4}\Big|}$. Note that $v_k(t)$ describes the time evolution of the corresponding node in the initial conditions, so that we require a decay of v_k for all $k \in \{1,\dots,m\}$; v_k decays exponentially fast with the rate $\frac{c_h^2}{2c_p^2}$ in the first case of (8.16). In the second case we also find an exponential decay: if $c_h^2\lambda_k < \frac{c_h^4}{4c_p^4}$, then we have $q < \frac{c_h^2}{2c_p^2}$ and therefore v_k decays like $\exp\Big(q - \frac{c_h^2}{2c_p^2}\Big)$. Thus v_k decays for $t \to \infty$ for all $k \in \{1,\dots,m\}$ but the decay is fastest for those k with $c_h^2\lambda_k \geq \frac{c_h^4}{4c_p^4}$. This fast decay applies to all nodes in the case where c_p and c_h satisfy (8.14).
The existence of a solution to our model problems (8.5) and (8.13) is a direct consequence of the above derivation since B and β both satisfy telegraph equations of the form (8.15). The functions ψ can then be defined using B or β respectively from the PDE describing the time evolution of ψ. The statements concerning the decay property of B and β then follow directly. □

8.8 Remark: *The combination of the hyperbolic with the parabolic approach leads to a set of equations where the initial disturbances in the divergence of* $\mathbf{B}$ *are, on the one hand, transported as in the hyperbolic approach and, at the same time, damped even for Dirichlet boundary conditions. Thus the advantages of both approaches are combined in the mixed GLM–MHD system.*
In the mixed approach ψ *no longer satisfies a conservation law as is the case in the hyperbolic approach. In* (8.1e) *we have an additional linear source term* $-\frac{c_h^2}{c_p^2}\psi$ *on the right hand side. Since* ψ *is not a physical quantity this is not a problem, and the important physical conservative variables* $\rho, \rho\mathbf{u}, \mathbf{B}, \rho e$ *still satisfy conservation laws.*
The modifications of the hyperbolic GLM–MHD system described above for achieving Galilean invariance can also be applied to the mixed GLM–MHD system, which differs

from the hyperbolic system only by the source term in the equation for ψ. But again we lose the conservation property of the magnetic field if we add the divergence "source–terms" to the GLM–MHD system.

8.2 Numerical Scheme

All the different approaches for choosing the linear operator $\mathcal{D}$ studied in the previous section can be easily implemented as an extension of a given *base scheme* for the original MHD system. We demonstrate the technique by means of our first order finite–volume scheme (3.16). We assume in the following that a numerical flux function $\mathbf{g}_{ij}$ is given. Since only the update for the magnetic field and for the auxiliary function ψ are affected by the GLM–MHD method, we restrict ourselves to detailing these modifications. Let therefore $\mathbf{g}_{ij}^{\mathbf{B}}$ denote the three components of the flux vector corresponding to the magnetic field. Then the update of the magnetic field using the base scheme is computed via

$$\mathbf{B}_i^{n+1} = \mathbf{B}_i^n - \frac{\Delta t^n}{|T_i|} \sum_j \mathbf{g}_{ij}^{\mathbf{B}} \ .$$

The Parabolic Approach

In the previous section we already saw that ψ can be eliminated from the GLM–MHD system if $\mathcal{D}$ is defined by (8.8). The evolution equation (8.1c) for $\mathbf{B}$ is now a second order equation

$$\partial_t \mathbf{B} + \nabla \cdot (\mathbf{u}\mathbf{B}^T - \mathbf{B}\mathbf{u}^T) = c_p^2 \nabla(\nabla \cdot \mathbf{B})$$

and we have no equation for ψ. Since the right hand side of this equation can be reformulated in divergence form, $\mathbf{B}$ is still a conservative quantity of the GLM–MHD system. To implement the parabolic approach we have to approximate $\nabla(\nabla \cdot \mathbf{B})$, which we do in a finite–volume spirit using Gauß' Theorem:

$$\int_{T_i} \nabla(\nabla \cdot \mathbf{B}) = \int_{\partial T_i} (\nabla \cdot \mathbf{B})\mathbf{n} = \sum_j \int_{S_{ij}} (\nabla \cdot \mathbf{B})\mathbf{n}_{ij} \ .$$

This leads to the following expression for the update of the magnetic field:

$$\begin{aligned} \mathbf{B}_i^{n+1} &= \mathbf{B}_i^n - \frac{\Delta t^n}{|T_i|} \sum_j \mathbf{g}_{ij}^{\mathbf{B}} + \frac{\Delta t^n}{|T_i|} c_p^2 \sum_j |S_{ij}| \widehat{\psi}_{ij}^n \mathbf{n}_{ij} \\ &= \mathbf{B}_i^n - \frac{\Delta t^n}{|T_i|} \sum_j \left(\mathbf{g}_{ij}^{\mathbf{B}} - c_p^2 |S_{ij}| \widehat{\psi}_{ij}^n \mathbf{n}_{ij} \right) \ , \end{aligned}$$

with

$$\widehat{\psi}_{ij}^n \approx (\nabla \cdot \mathbf{B})_{|_{S_{ij}}} \ .$$

The main difficulty is to define $\widehat{\psi}_{ij}^n$ as a good approximation of $\nabla \cdot \mathbf{B}$ on S_{ij}. We use the approach sketched in Algorithm 4 on page 117, which utilizes a continuous reconstruction of the magnetic field on triangular grid — the method can be directly extended to 3d. In the first step values for the magnetic field are defined on each vertex p of the grid by averaging the data of all the elements surrounding p. These values allow us to construct a linear function $\bar{\mathbf{B}}_i$ on each triangle (or tetrahedra) T_i. Now the divergence of the magnetic field is approximated on a surface S_{ij} by taking the average of the two constant values $\nabla \cdot \bar{\mathbf{B}}_i$ and $\nabla \cdot \bar{\mathbf{B}}_j$.

Algorithm 4: Modification of the base scheme for the parabolic GLM–MHD correction. In the postprocessing step a linear reconstruction is computed, which is used to approximate the divergence of $\mathbf{B}$ during the calculation of the flux.

(a) **Poststep**

for all $i \in \mathcal{I}_B$ **do**
  $\mathbf{U}_i \leftarrow$ [cf. Section 3.3]
end for
exchange data on inner boundary
for all $i \in \mathcal{I}$ **do**
  $\{\mathcal{L}_i\} \leftarrow$ **Reconstruction**
  $\{Q_{\mathrm{rad}\,i}\} \leftarrow$ [cf. Chapter 12–13]
  $\mathbf{g}_i \leftarrow 0, \mathrm{ref}_i \leftarrow 0, \mathrm{crs}_i \leftarrow 0$
end for
for all $k \in \mathcal{I}_P$ **do**
  $\mathbf{B}_k \leftarrow 0, \mathrm{area}_k \leftarrow 0$
end for
for all $i \in \mathcal{I}$ **do**
  for all $k \in \mathcal{P}(i)$ **do**
    $\mathbf{B}_k \leftarrow \mathbf{B}_k + |T_i|\mathbf{B}_i^n$
    $\mathrm{area}_k \leftarrow \mathrm{area}_k + |T_i|$
  end for
end for
for all $i \in \mathcal{I}$ **do**
  Construct $\bar{\mathbf{B}}_i$ as linear function through
    $(\mathbf{p}_k, \frac{\mathbf{B}_k}{\mathrm{area}_k})$ for $k \in \mathcal{P}(i)$
  $\widehat{\psi}_i^n \leftarrow \nabla \cdot \bar{\mathbf{B}}_i$
end for

(b) **FLUX**$(\mathbf{U}_l, \mathbf{U}_r, \mathbf{n}, h)$

$\bar{\mathbf{U}}_l \leftarrow \mathcal{R}(\mathbf{n}_{ij})\mathbf{U}_l$
$\bar{\mathbf{U}}_r \leftarrow \mathcal{R}(\mathbf{n}_{ij})\mathbf{U}_r$
$\bar{\mathbf{G}} \leftarrow \mathbf{G}(\bar{\mathbf{U}}_l, \bar{\mathbf{U}}_r)$
$\mathbf{g}_{ij} \leftarrow |S_{ij}|\mathcal{R}^{-1}(\mathbf{n}_{ij})\bar{\mathbf{G}}(\bar{\mathbf{U}}_l, \bar{\mathbf{U}}_r)$
$\Delta t_{ij} \leftarrow \frac{h}{\max\{|\bar{u}_{l,x}|+c_f(\bar{\mathbf{U}}_l), |\bar{u}_{r,x}|+c_f(\bar{\mathbf{U}}_r)\}}$
$\mathrm{jmp}_{ij} \leftarrow$ [cf. (3.19)]
$\mathbf{g}_{ij} \leftarrow \mathbf{g}_{ij} - (0, \mathbf{0}, c_p^2|S_{ij}|\frac{1}{2}(\widehat{\psi}_i^n + \widehat{\psi}_j^n)\mathbf{n}_{ij}, 0)^T$

The Hyperbolic Approach

The modification of the base scheme in the case of the hyperbolic correction can be written down in a fashion very simular to the parabolic correction. In this case we have to update $\mathbf{B}_i^n$ and ψ_i^n:

$$\mathbf{B}_i^{n+1} = \mathbf{B}_i^n - \frac{\Delta t^n}{|T_i|} \sum_j \left(\mathbf{g}_{ij}^{\mathbf{B}} + |S_{ij}| \widehat{\psi}_{ij}^n \mathbf{n}_{ij} \right) ,$$

$$\psi_i^{n+1} = \psi_i^n - \frac{\Delta t^n}{|T_i|} \sum_j |S_{ij}| c_h^2 \widehat{B}_{ij}^n \, .$$

In this case the tuple $(\widehat{B}_{ij}^n, \widehat{\psi}_{ij}^n)$ are some suitably upwinded values derived from $\mathbf{B}_i^n \cdot \mathbf{n}_{ij}, \mathbf{B}_j^n \cdot \mathbf{n}_{ij}$ and ψ_i^n, ψ_j^n, respectively. Following the idea of rotated one dimensional

Riemann solvers detailed in Section 3.2, we can define $(\widehat{B}^n_{ij}, \widehat{\psi}^n_{ij})$ by studying the solution to the one dimensional hyperbolic GLM–MHD system. We have already remarked upon the fact that in one space dimension the equations governing the evolution of B_x, ψ decouple from the other equations. Since B_x in the one dimensional system corresponds to $\mathbf{B} \cdot \mathbf{n}$ in the rotated frame, we see that $\widehat{B} := \mathbf{B} \cdot \mathbf{n}$ and ψ satisfy a decoupled linear two by two system

$$\partial_t \widehat{B} + \partial_x \psi = 0 \ ,$$
$$\partial_t \psi + c_h^2 \partial_x \widehat{B} = 0 \ .$$

The Riemann problem for this system with left hand state $(\widehat{B}_l, \psi_l) = (\mathbf{B}^n_i \cdot \mathbf{n}_{ij}, \psi_i)$ and right hand state $(\widehat{B}_r, \psi_r) = (\mathbf{B}^n_j \cdot \mathbf{n}_{ij}, \psi_j)$ can be solved explicitly. On the cell interface the solution has the value

$$\begin{pmatrix} \widehat{B}_m \\ \psi_m \end{pmatrix} = \begin{pmatrix} \widehat{B}_l \\ \psi_l \end{pmatrix} + \begin{pmatrix} \frac{1}{2}(\widehat{B}_r - \widehat{B}_l) - \frac{1}{2c_h}(\psi_r - \psi_l) \\ \frac{1}{2}(\psi_r - \psi_l) - \frac{c_h}{2}(\widehat{B}_r - \widehat{B}_l) \end{pmatrix} \ . \tag{8.17}$$

Therefore a suitable choice for the value of $(\mathbf{B} \cdot \mathbf{n}_{ij}, \psi)$ on the interface S_{ij} is given by $(\widehat{B}^n_{ij}, \widehat{\psi}^n_{ij}) := (\widehat{B}_m, \psi_m)$.
So far we have discretized the new term $\nabla\psi$ in the induction equation without modifying the flux of the MHD equations itself. This flux is an approximation of the analytical flux on the surface S_{ij} and is constructed by using a rotated numerical flux $\mathbf{G}$ for the 1d MHD equations. Our aim is not to have to modify $\mathbf{G}$; but the modification of the left and right values $\mathbf{U}_l$ and $\mathbf{U}_r$ for which $\mathbf{G}$ is evaluated presents no problem in our code (cf. Algorithm 2(b) on page 66). In our numerical tests we found that it is advantageous to replace the first magnetic field components $B_{x,l}$ and $B_{x,r}$ with the approximation $\widehat{B}^n_{ij}$ of B_x on the interface. A modification of the magnetic field leads to a modification of the internal energy computed using the algebraic relation (1.1f). Thus, the left and right hand pressure $p(\mathbf{U}_l)$ and $p(\mathbf{U}_r)$ are modified. The best results are obtained if the pressure is maintained, thus modifying the total energy density $(\rho e)_{l,r}$. The resulting implementation of the hyperbolic GLM–MHD scheme is sketched in Algorithm 5 on page 120.

Galilean Invariance

To extend the hyperbolic GLM–MHD scheme to approximate the Galilean invariant GLM–MHD system we have to include the non hyperbolic terms in the finite–volume scheme. This can be achieved by the approximations

$$\frac{1}{|T_i|}\int_{T_i} \mathbf{u}\nabla\cdot\mathbf{B} \approx \frac{1}{|T_i|}\mathbf{u}^n_i \int_{T_i} \nabla\cdot\mathbf{B} = \frac{1}{|T_i|}\mathbf{u}^n_i \sum_j \int_{S_{ij}} \mathbf{B}\cdot\mathbf{n} \approx \frac{1}{|T_i|}\mathbf{u}^n_i \sum_j |S_{ij}|\widehat{B}^n_{ij} \ ,$$

and

$$\frac{1}{|T_i|}\int_{T_i} \mathbf{u}\cdot\nabla\psi \approx \frac{1}{|T_i|}\mathbf{u}^n_i \cdot \int_{T_i} \nabla\psi = \frac{1}{|T_i|}\mathbf{u}^n_i \cdot \sum_j \int_{S_{ij}} \psi\mathbf{n}_{ij} \approx \frac{1}{|T_i|}\sum_j |S_{ij}|\widehat{\psi}^n_{ij}\mathbf{u}^n_i\cdot\mathbf{n}_{ij} \ ,$$

Hereby we use the values $\widehat{B}^n_{ij}$ and $\widehat{\psi}^n_{ij}$ computed for the hyperbolic correction to approximate $\mathbf{B}\cdot\mathbf{n}_{ij}$ and ψ on the interface S_{ij}. The modification of the base scheme is described in Algorithm 5 on page 120.

The Mixed Approach

The mixed approach differs from the hyperbolic correction only by a linear source term in the equation for ψ. This source term can be implemented in a finite–volume framework; but this leads to an unnecessary stability restriction on the size of the factor $\frac{c_h^2}{c_p^2}$. By using an operator splitting approach we can overcome the stability restriction. If we denote the update of ψ from the homogeneous part of equation (8.1e) by ψ_i^{n*}, the actual value of ψ at the next time level t^{n+1} is then defined as the solution to the ODE

$$\psi_i(t) = -\frac{c_h^2}{c_p^2}\psi_i(t)$$

at time Δt^n with initial conditions $\psi_i(0) = \psi_i^{n*}$. Since this is a linear ODE, the solution is easily computed, leading to the following equation for ψ at the new time level:

$$\psi_i^{n+1} = \exp\Big(-\frac{c_h^2}{c_p^2}\Delta t^n\Big)\psi_i^* \ .$$

This corresponds to the damping term in (8.16) in the proof of Theorem 8.6. The corresponding scheme is summarized in Algorithm 5 on page 120.

8.3 Choice of Parameters

In this section we discuss the choice of the free parameters c_p and c_h. Formally these constants influence the damping and the transport of divergence errors as described in Section 8.1. The main idea behind our choice for these values is that we want to achieve as large a reduction of divergence errors as possible, while at the same time we want to maintain the size of the time step Δt^n given by the base scheme. Therefore our choice of c_p and c_h must not introduce any additional stability restrictions that would require the reduction of the time step Δt^n. Thus we choose new values for $c_p = c_p^n$ and $c_h = c_h^n$ in every time step as functions of Δt^n and the grid resolution. In the following Δt^n always denotes the time step given by the base scheme (for example as defined in (3.11)).

Defining c_h

We start off by detailing our choice for the constant c_h in the hyperbolic and the Galilean invariant approach. As we have seen in Section 8.1 the parameter c_h directly enters into the eigensystem of the GLM–MHD equations leading to two additional waves. In the hyperbolic case the new wave speeds are $\lambda_{\text{div}\pm}(\mathbf{U}, \mathbf{n}) = \pm c_h$, and in the Galilean invariant approach we have two new waves with the speeds $\lambda_{\text{div}\pm}(\mathbf{U}, \mathbf{n}) = \mathbf{u}\cdot\mathbf{n} \pm c_h$. We denote with $\bar{\lambda}_{\text{div}}(\mathbf{U}, \mathbf{n})$ the maximum of these new eigenvalues, i.e., for c_h large enough

$$\bar{\lambda}_{\text{div}}(\mathbf{U}, \mathbf{n}) = \begin{cases} c_h & \text{in the hyperbolic case,} \\ |\mathbf{u}\cdot\mathbf{n}| + c_h & \text{in the Galilean invariant case.} \end{cases}$$

Algorithm 5: Modification of the base scheme for the hyperbolic GLM–MHD correction. To achieve Galilean invariance an additional step in **Element** is required to update the source terms. In this case the values for the normal magnetic field and the auxiliary function on the surfaces have to be stored. Both corrections can be enhanced by an additional damping of ψ in **Poststep**. All vectors in state space ($\mathbf{U}_i, \mathbf{g}_{ij}, \bar{\mathbf{U}}_i$ etc.) are now assumed to be vectors in $\mathbb{R}^9$ instead of in $\mathbb{R}^8$ as before. Thus the auxiliary function ψ is directly incorporated into the scheme including, for example, the linear reconstruction process.

(a) **FLUX**$(\mathbf{U}_l, \mathbf{U}_r, \mathbf{n}, h)$

$\bar{\mathbf{U}}_l \leftarrow \mathcal{R}(\mathbf{n}_{ij})(\mathbf{U}_{l,1}, \ldots, \mathbf{U}_{l,8})$
$\bar{\mathbf{U}}_r \leftarrow \mathcal{R}(\mathbf{n}_{ij})(\mathbf{U}_{r,1}, \ldots, \mathbf{U}_{r,8})$
$\widehat{B}_{ij} \leftarrow \bar{B}_{l,x} + \frac{1}{2}(\bar{B}_{r,x} - \bar{B}_{l,x}) - \frac{1}{2c_h}(\psi_r - \psi_l)$
$\widehat{\psi}_{ij} \leftarrow \widehat{\psi}_l + \frac{1}{2}(\psi_r - \psi_l) - \frac{c_h}{2}(\bar{B}_{r,x} - \bar{B}_{l,x})$
$\bar{\rho e}_l \leftarrow \bar{\rho e}_l - \frac{\bar{B}_{l,x}^2}{8\pi} + \frac{\bar{B}_{ij}^2}{8\pi}$
$\bar{\rho e}_r \leftarrow \bar{\rho e}_r - \frac{\bar{B}_{r,x}^2}{8\pi} + \frac{\bar{B}_{ij}^2}{8\pi}$
$\bar{B}_{l,x} \leftarrow \widehat{B}_{ij}$
$\bar{B}_{r,x} \leftarrow \widehat{B}_{ij}$
$\bar{\mathbf{G}} \leftarrow \mathbf{G}(\bar{\mathbf{U}}_l, \bar{\mathbf{U}}_r)$
$\bar{\mathbf{g}}_{ij} \leftarrow |S_{ij}|\mathcal{R}^{-1}(\mathbf{n}_{ij})\bar{\mathbf{G}}(\bar{\mathbf{U}}_l, \bar{\mathbf{U}}_r)$
$\mathbf{g}_{ij} \leftarrow (\bar{\mathbf{g}}_{ij}, 0) + (0, \mathbf{0}, |S_{ij}|\widehat{\psi}_{ij}\mathbf{n}_{ij}, 0, |S_{ij}|c_h^2\widehat{B}_{ij})^T$
$\Delta t_{ij} \leftarrow \frac{h}{\max\{|\bar{u}_{l,x}|+c_f(\bar{\mathbf{U}}_l), |\bar{u}_{r,x}|+c_f(\bar{\mathbf{U}}_r)\}}$
$\mathrm{jmp}_{ij} \leftarrow$ [cf. (3.19)]

(b) **Element**

for all $i \in \mathcal{I}$ **do**
  $\mathbf{g}_i \leftarrow -\mathbf{g}_i + \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \mathbf{q}(\mathcal{L}_i(\mathbf{z}_{ij})) + \mathbf{Q}_{\mathrm{rad}\,i}$
  $\mathbf{g}_i \leftarrow \mathbf{g}_i + (0, \mathbf{0}, \frac{1}{|T_i|}\mathbf{u}_i^n \sum_{j\in\mathcal{N}(i)} |S_{ij}|\widehat{B}_{ij}^n, 0)^T$
  if $\mathrm{ref}_i > \mathrm{ref}_{\mathrm{limit}}$ and $h_i > h_{min}$ **then**
    mark T_i for refinement
  else if $\mathrm{crs}_i < \mathrm{crs}_{\mathrm{limit}}$ and $h_i < h_{max}$ **then**
    mark T_i for coarsening
  end if
end for

(c) **Poststep**

for all $i \in \mathcal{I}_B$ **do**
  $\mathbf{U}_i \leftarrow$ [cf. Section 3.3]
end for
exchange data on inner boundary
for all $i \in \mathcal{I}$ **do**
  $\psi_i \leftarrow \exp\left(-\frac{c_h^2}{c_p^2}\Delta t\right)\psi_i$
  $\{\mathcal{L}_i\} \leftarrow$ **Reconstruction**
  $\{Q_{\mathrm{rad}\,i}\} \leftarrow$ [cf. Chapter 12–13]
  $\mathbf{g}_i \leftarrow 0, \mathrm{ref}_i \leftarrow 0, \mathrm{crs}_i \leftarrow 0$
end for

Since the wave speeds of the eigensystem directly influence the size of the time–step, we have to choose c_h small enough so that the time–step can be maintained. Let us recall the definition of Δt^n given by equations (3.9) and (3.11)

$$\Delta t^n = c_{\mathrm{cfl}} \min_{(i,j)\in\mathcal{I}_S} \frac{h_{ij}}{\max\{\lambda_{\max}(\mathbf{U}_i(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij}), \lambda_{\max}(\mathbf{U}_j(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij})\}} .$$

The value $\lambda_{\max}(\mathbf{U}_i(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij})$ denotes the fastest wave speed in the direction $\mathbf{n}_{ij}$. Since we want to use the same time–step for the modified scheme as well, the following inequality must hold

$$\Delta t^n \leq c_{\mathrm{cfl}} \min_{(i,j)\in\mathcal{I}_S} \frac{h_{ij}}{\bar{\lambda}_{\mathrm{div}}(\mathbf{U}_i(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij})} . \tag{8.18}$$

The Hyperbolic Approach: In the hyperbolic approach this inequality is especially easy to fulfill since the new wave speeds do not depend on the conservative quantities or the unit normal $\mathbf{n}$. Inequality (8.18) is equivalent to

$$\Delta t^n \leq c_{\mathrm{cfl}} \min_{(i,j)\in\mathcal{I}_S} \frac{h_{ij}}{c_h} .$$

This inequality leads to an upper bound for c_h, and since a fast transport of divergence errors with the two additional waves is desirable, we choose c_h^n as large as possible

$$c_h^n := c_{\text{cfl}} \frac{\min_{(i,j)\in\mathcal{I}_S} h_{ij}}{\Delta t^n} \ . \tag{8.19}$$

The Galilean Invariant Approach: Inequality (8.18) is satisfied if

$$\max\{\bar{\lambda}_{\text{div}}(\mathbf{U}_i(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij}), \bar{\lambda}_{\text{div}}(\mathbf{U}_j(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij})\} \leq \max\{\lambda_{\max}(\mathbf{U}_i(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij}), \lambda_{\max}(\mathbf{U}_j(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij})$$

for all $(i,j) \in \mathcal{I}_S$. Using equation (3.10) this leads to the following inequality for all $i \in \mathcal{I}$ and $j \in \mathcal{N}(i)$

$$|\mathbf{u}_i \cdot \mathbf{n}_{ij}| + c_h \leq \min\{|\mathbf{u}_i \cdot \mathbf{n}_{ij}| + c_f(\mathbf{U}_i(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij}), |\mathbf{u}_j \cdot \mathbf{n}_{ij}| + c_f(\mathbf{U}_j(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij})\} \ .$$

Therefore we define

$$c_h^n := \min_{(i,j)\in\mathcal{I}_S} \{c_f(\mathbf{U}_i(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij}), c_f(\mathbf{U}_j(\mathbf{z}_{ij}, t^n), \mathbf{n}_{ij})\} \ . \tag{8.20}$$

Defining c_p

For the mixed approach we use the same choice for c_h as for the hyperbolic or the Galilean invariant approach. In addition we must fix the constant c_p. In the parabolic approach only c_p must be chosen.

The Mixed Approach: We are quite free in our choice of the constant c_p since our method for discretizing the mixed approach is unconditionally stable with respect to c_p. In [DKK+02] we tested many different possibilities for defining c_p as a function of c_h^n and Δt^n. Our tests have led us to believe that the best choice is to fix the ration $c_{\text{rel}} := \frac{c_h}{c_p^2}$ in each time step. This choice was also motivated by some analysis of the solution to the wave equation and the heat equation similar to the ones presented in Section 8.1. The analysis of the telegraph equation given in Section 8.1 leads to the same result (cf. Theorem 8.6). Thus to define c_p we first fix a constant c_{rel} a priori. Then we compute c_h^n using the method described above. The value for $(c_p^n)^2$ is then given by

$$(c_p^n)^2 := \frac{c_h^n}{c_{\text{rel}}} \ . \tag{8.21}$$

From our analysis of the telegraph equation a good choice seems to be given by $c_{\text{rel}} := 2\sqrt{\lambda_{min}}$ — a value that is close to the optimal value found in our numerical experiments. For rectangular domains $\Omega = [-L_x, L_x] \times [-L_y, L_y] \times [-L_z, L_z]$ this value is easily computed:

$$c_{\text{rel}} = \pi \sqrt{\frac{1}{L_x^2} + \frac{1}{L_y^2} + \frac{1}{L_z^2}}. \tag{8.22}$$

8.9 Remark: *Our numerical tests have shown that the choice of c_{rel} does not have a significant impact on the quality of the scheme; also the results for fixed c_{rel} depend very little upon grid resolution and the order of the scheme (cf. [DKK+02]).*

The Parabolic Approach: In the parabolic case the choice of the parameter c_p^n is critical. A large value leads to a high damping of the divergence errors, while at the same time a large value of the factor $(c_p^n)^2$ in front of the second order term leads to a severe restriction of the time step. In scalar convection diffusion equations the factor ε in front of the diffusion term enters into the time step control via $\Delta t \approx \frac{h^2}{\varepsilon}$. Since we want to maintain the time step given by the base scheme (which is proportional to h), we have to choose $(c_p^n)^2 \approx h$. We have tested many different approaches to define c_p as a function of the grid size h. One choice that leads to a stable scheme is to choose c_p in the same way as in the mixed approach with a constant c_{rel} that scales with the grid resolution:

$$(c_p^n)^2 := c_{\text{cfl}} \frac{\min_{(i,j)\in\mathcal{I}_S} h_{ij}}{\Delta t^n} \frac{\min_{(i,j)\in\mathcal{I}_S} h_{ij}}{c_{\text{rel}}} \tag{8.23}$$

This corresponds to our definition of c_p in the mixed approach scaled with the minimum grid resolutions if we replace c_h in (8.21) using (8.19).

8.10 Remark: *Due to the dependence of c_p on h, an increase in grid resolution leads to less damping of divergence errors in the parabolic approach. As our numerical examples demonstrate, an increase in the resolution does not necessarily lead to smaller errors in $\nabla \cdot \mathbf{B}$, so that less damping is not desirable (cf. Figure 8.7 on page 134).*

8.4 Initial and Boundary Conditions

We still have to detail our choice for the initial and boundary conditions for the auxiliary function ψ. Initially we set $\psi \equiv 0$ in all our simulations. In Section 8.1 we used Dirichlet boundary conditions for ψ. Our numerical tests have shown that this choice leads to a reduction of divergence errors — but in some cases these boundary conditions are not optimal. For example, in the case of periodic boundary conditions the periodic structure of the physical variables can be disturbed if we do not also prescribe periodic boundary conditions for ψ. In fact, the best results were obtained when we used the same boundary conditions for ψ as for the other scalar quantities [DKK+02]. For the boundary conditions presented in Section 3.3 we, thus, define for $(i,j) \in \mathcal{I}_S$ and $j \in \mathcal{I}_B$:

inflow: $\psi_j(\mathbf{z}_{ij}, t) = 0.$

outflow: $\psi_j(\mathbf{z}_{ij}, t) = \psi_i(\mathbf{z}_{ij}, t).$

slip: $\psi_j(\mathbf{z}_{ij}, t) = \psi_i(\mathbf{z}_{ij}, t).$

reflecting: $\psi_j(\mathbf{z}_{ij}, t) = \psi_i(\mathbf{z}_{ij}, t).$

periodic: Let $\mathbf{U}(\mathbf{x}, t) = \mathbf{U}(\mathbf{x} + \alpha \mathbf{n}_{ij}, t)$ with some fixed $\alpha \in \mathbb{R}$. Then $\psi_j(\mathbf{z}_{ij}, t) = \psi_h(\mathbf{z}_{ij} + \alpha \mathbf{n}_{ij}, t)$.

In the following we only show results using these boundary conditions.

8.11 Remark: *According to Theorem 8.3 no damping of divergence errors can be expected in the hyperbolic case if wrong boundary conditions are used. The artificial*

waves that should transport the divergence errors out of the domain are reflected at the boundary if no transparent boundary conditions are used. This problem is reduced if, in addition to the transport of divergence errors, these are also damped as in the mixed approach. In our numerical tests we observed that the boundary conditions have very little impact on the performance of the scheme even if only the hyperbolic correction is used. This is presumably due to the damping introduced by numerical viscosity, which is always present in finite–volume schemes. We are not aware of suitable transparent boundary conditions that can be prescribed on the whole boundary and where the physical boundary conditions for the magnetic field is maintained.

8.5 Numerical Results

In this chapter we derived a number of different approaches to coping with problems arising from a violation of the divergence constraint (1.1e) on the magnetic field $\mathbf{B}$. They all have in common that they can easily be added to an existing *base scheme*. We have described two approaches that have often been used in the literature: the *source term fix* and the *Hodge projection scheme*. The latter is included in our GLM–MHD framework in the form of the *elliptic correction* — we use this notation in the following. Using different choices for the linear differential operator $\mathcal{D}$ in the evolution equation (8.1e) for the auxiliary function ψ, we arrive at three further approaches, which we identify as *parabolic*, *hyperbolic*, and *mixed correction*. Furthermore, we can add some additional terms to the hyperbolic approach; these lead to a *Galilean invariant* formulation of the hyperbolic GLM–MHD system. This modification can also be applied to the mixed correction. Together with the base scheme we now have eight different methods for approximating the MHD system.

We reduce the number of schemes by not including results for the pure hyperbolic and Galilean invariant correction. In most numerical tests the quality of the approximation is barely influenced by the additional source term in the mixed correction, so that the results of the hyperbolic approach are in most cases comparable to the results obtained using the mixed correction. The additional cost of using the mixed correction is negligible, and our analysis has shown that the purely hyperbolic approach can lead to problems at the boundaries. Therefore we favor the mixed correction. The remaining six methods (base scheme plus five correction methods) are summarized in Table 8.1.

We concentrate on the comparison of the approximation errors of the schemes on a given grid and not on the runtime efficiency of the schemes. The mixed corrections with or without divergence source terms can be implemented with hardly any additional cost in cpu time so that the efficiency of the method in relation to the base scheme is not an issue. For the parabolic approach the reconstruction of the magnetic field using vertex averages leads to additional computational cost. This is not the simplest approach to compute the divergence of the magnetic field on the faces of the grid. The problem of finding an efficient implementation is even more difficult for the elliptic correction. Solving the Laplace equation in the projection step is a very time consuming part of the algorithm, which severely influences the efficiency of the scheme. Since all our numerical tests show that the mixed correction techniques are superior to both the parabolic and the elliptic approach, we decide to neglect the efficiency issue in the following.

To compare the approximate solutions even in the case where we have no exact solution, we compare the errors in the divergence of the magnetic field $\mathbf{B}$ — since in the exact solution we have $\nabla \cdot \mathbf{B} \equiv 0$. On Cartesian grids $\nabla \cdot \mathbf{B}$ is easily computed for a given approximation using, for example, central differences. On unstructured grids a consistent approximation of the divergence operator is not so straightforward. We do not discuss the problem in detail. A thorough study can be found in [Wes02a]. We use the following expressions to compute $\nabla \cdot \mathbf{B}$ on a given element T_i for $i \in \mathcal{I}$:

$$(\nabla \cdot \mathbf{B})^n_{\mathrm{el},i} := \frac{1}{|T_i|} \sum_{j \in \mathcal{N}(i)} \frac{|S_{ij}|}{2} \big(\mathbf{B}_i(\mathbf{z}_{ij}, t^n) + \mathbf{B}_j(\mathbf{z}_{ij}, t^n)\big) \cdot \mathbf{n}_{ij} \quad \text{(element error)} , \tag{8.24}$$

$$(\nabla \cdot \mathbf{B})^n_{\mathrm{fce},i} := \max_{j \in \mathcal{N}(i)} \Big| \frac{|S_{ij}|}{|T_i|} \big(\mathbf{B}_i(\mathbf{z}_{ij}, t^n) - \mathbf{B}_j(\mathbf{z}_{ij}, t^n)\big) \cdot \mathbf{n}_{ij} \Big| \qquad \text{(jump error)} . \tag{8.25}$$

In the following we use the sum of the L^1–norms and the L^∞–norms of the functions $(\nabla \cdot \mathbf{B})^n_{\mathrm{el}}(\mathbf{x}) := (\nabla \cdot \mathbf{B})^n_{\mathrm{el},i}$ and $(\nabla \cdot \mathbf{B})^n_{\mathrm{fce}}(\mathbf{x}) := (\nabla \cdot \mathbf{B})^n_{\mathrm{fce},i}$ for $\mathbf{x} \in T_i$. Note that since $\mathbf{B}$ is discontinues in many of our test cases, we cannot expect convergence in any norm — especially not in L^∞. Nevertheless the L^∞–norm is important since the stability problems are often caused by local divergence errors, so that a comparison of the magnitude of the L^∞–errors gives some indication of the quality of the scheme.

8.5.1 Influence of the parameter c_{rel}

In the case of the GLM–MHD schemes presented here the parameters that have to be chosen are c_h and c_p: the parameter c_h influences the speed of propagation of the divergence waves and c_p the amount of damping. Since coupling c_h to the time–step — as suggested in Section 8.3 — seems to be the optimal choice, we do not show results using, for example, a fixed c_h. As a consequence of our analysis and of earlier studies we choose c_p according to (8.21) in the mixed approach and according to (8.23) in the parabolic approach. Consequently, we only have to investigate the influence of the remaining parameter c_{rel} in (8.21) and (8.23) on the mixed approach and on the parabolic approach, respectively.

First we study the mixed GLM–MHD scheme. Motivated by the analysis of our model problem in Section 8.3, we suggest defining this value in relation to the computational domain Ω. In the following we denote the scheme with c_{rel} given by (8.22) as *mixed* correction. The results shown in Figure 8.1 demonstrate that the choice of this constant has very little influence on the performance of the scheme. The results for the 1d Riemann problem seem to indicate that a large value for c_{rel} leads to a reduction of the error at least for high grid resolution. The results for the rotation problem, on the other hand, show that this is not true in general; consequently, the optimal choice for c_{rel} is not easy to find. Formula (8.22) seems to lead to almost optimal results for both problems. This choice is used in the mixed approach in all further calculations.

Figure 8.2 shows a series of calculations using the parabolic approach with different values for c_{rel}. The results show that by increasing c_{rel} the error is reduced slightly; but at the same time the scheme becomes unstable. With $c_{\mathrm{rel}} = 0.2$ we have not encountered stability problems in any of our tests, so that we use this value in the following.

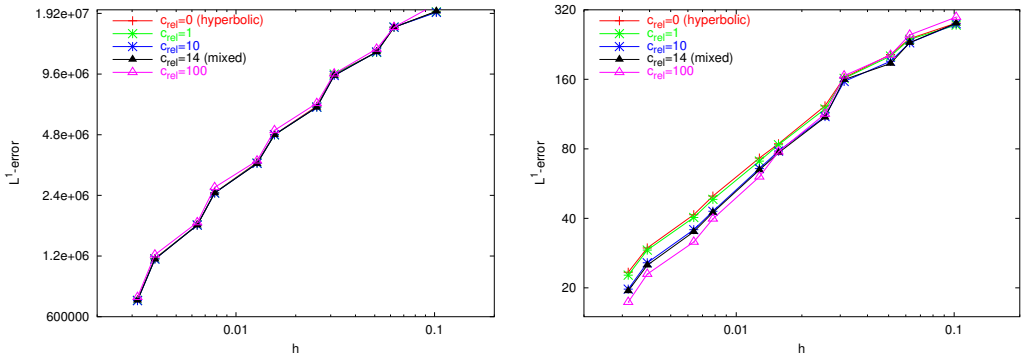
Scheme	operator with $\hat{\mathcal{D}}(\nabla \cdot \mathbf{B}) = 0$	effect	implementation	stencil	conservation
base	∂_t	stationary	fluxes	neighbors	$(\rho, \rho\mathbf{u}, \mathbf{B}, \rho e)$
elliptic	$\triangle$	projection	fluxes Laplace	full domain	$(\rho, \rho\mathbf{u}, \mathbf{B}, \rho e)$
(mixed) Galilean invariant	*no simple expression*	transport damping	fluxes source term	neighbors	$(\rho, \rho\mathbf{u}, \rho e)$
mixed	$\partial_{tt} + \frac{c_h^2}{c_p^2}\partial_t - c_h^2\triangle$	transport damping	fluxes source term	neighbors	$(\rho, \rho\mathbf{u}, \mathbf{B}, \rho e)$
parabolic	$\partial_t - c_p^2\triangle$	damping	fluxes	"vertex neighbors"	$(\rho, \rho\mathbf{u}, \mathbf{B}, \rho e)$
source terms	$\partial_t + \mathrm{div}(\mathbf{u}\cdot)$	transport	source term	neighbors	(ρ)

Table 8.1: Of the eight correction mechanisms presented in this chapter we concentrate on the six listed above. The main characteristics of the different approaches are summarized. Four of the methods are directly based on the GLM–MHD formalism: together with the base scheme these are the elliptic, the parabolic, and the mixed approach. The Galilean invariant approach is an extension of the mixed approach. Additional source terms in the evolution equations for $\mathbf{B}$ and ψ lead to a Galilean invariant system. Due to these source terms the magnetic field is no longer conserved. Note that a simple evolution equation for $\nabla \cdot \mathbf{B}$ cannot be derived for this setting. In the source term fix additional terms that are not in divergence form are added to the equations for the momentum, the magnetic field, and the total energy density. Therefore only the density is conserved. The elliptic approach corresponds to the Hodge projection scheme, where an additional Laplace equation is solved in each time step. In the case of $\nabla \cdot \mathbf{B} = 0$ all schemes are equivalent to the base scheme.

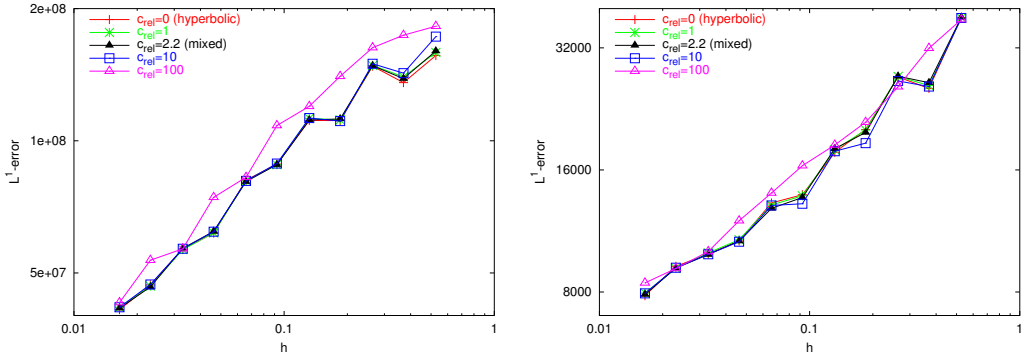
Summary of Section 8.5.1: *Since our goal is to reduce the amount of parameter tuning as much as possible, we present possibilities of choosing the parameters in our GLM–MHD scheme automatically. We choose new values for c_h and c_p in each time step. In the hyperbolic approaches we couple c_h to the time step Δt via* (8.20) *and in the mixed approach c_p to c_h by means of* (8.21)*. With these choices we have no parameter left in the hyperbolic and the Galilean invariant scheme; only in the mixed approach are we left with a parameter c_{rel}. Motivated by our analytical results for the model problem* (8.5) *in Theorem 8.6 and our numerical tests, we conclude that a good choice for this parameter is given by* (8.22)*, where c_{rel} is chosen subject to the computational domain Ω.*
In the parabolic approach we have to choose c_p proportional to the grid size h (or the time step Δt) to obtain a stable scheme. Our choice for the proportionality constant c_{rel} in equation (8.23) *is more or less arbitrary and motivated only by a number of numerical tests where $c_{\text{rel}} = 0.2$ resulted in a stable scheme with a measurable reduction of the divergence errors.*

8.5.2 Rotation Problem

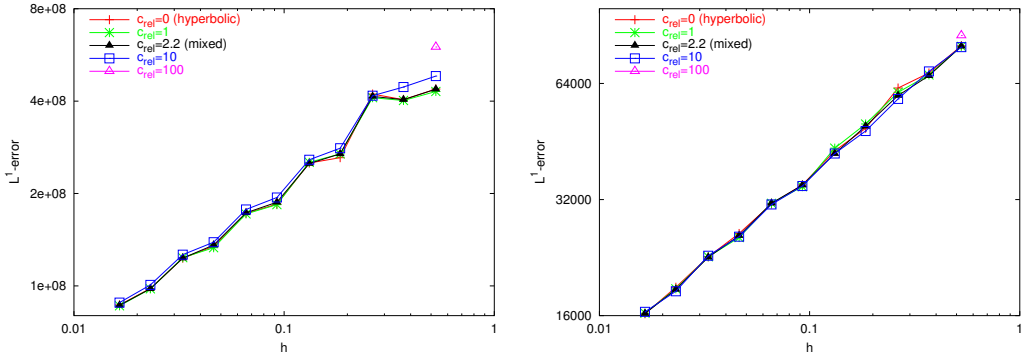
Next we continue the study of Problem 6.7(RotConst) begun in Section 7.6. There, on the one hand, we saw how a magnetic field leads to a stabilization of the interface; on the other hand, however, we also observed that an increase in the strength of the magnetic field leads to problems with the stability of the scheme (cf. Figure 7.9) that become more severe with increasing grid resolution. While we can compute the solution using the base scheme at all refinement levels from zero to ten (starting with the unstructured macro grid) for both $\beta = 0$ and $\beta = \frac{1}{2}$, it crashed for $\beta = 1$ at level five. For $\beta = 2$ we arrive at the final time $T = 5e-3$ only at level zero and level two. At all other levels the simulation terminates well before the final time. At level three, for example, it crashes at time $t = 1.05e-3$, at level four at $t = 1.28e-3$, and at level five at $t = 5.24e-4$. In the previous Chapter we made the conjecture that these problems are caused by unphysical magnetic fields. One indication is that with the inclusion of divergence source terms the performance of the scheme is improved: for $\beta = 1$ all refinement levels presented no problem — although the EOC was very small. For $\beta = 2$ we compute the solution on all levels up to and including level eight before we observe a breakdown of the simulation. In the following we extend this test using the GLM–MHD methods.
We start off by plotting the L^1–error versus the grid resolution for $\beta = 1, 2$. The results are shown in Figure 8.4. (For $\beta = \frac{1}{2}$ we observe no problems, although the unstable interface leads to a very poor convergence rate so that we concentrate on $\beta \geq 1$.) For $\beta = 1$ we see that the source term fix does not converge or at least with only a poor rate. The elliptic approach leads to an improvement compared to the base scheme if we look at the errors; but crashes for the same grid resolution (level 5) as the base scheme. On level 6 it again reaches the final time T, as well as on levels 7,8, and 10. For refinement level 9, on the other hand, it crashes. The parabolic approach computes the solution up to level eight with a noticeable reduction of the error; it crashes on level 9 and computes a solution again on level 10. Next to the source term fix only the mixed and the Galilean invariant correction compute the solution for all



(a) Problem 6.6(RP2D(2,1)): Left: error in full state vector, right: error in B_x



(b) Problem 6.7(RotConst) ($\beta = 1$): Left: error in full state vector, right: error in B_x



(c) Problem 6.7(RotConst) ($\beta = 2$): Left: error in full state vector, right: error in B_x

Figure 8.1: Problem: 1d Riemann problem 6.6(RP2D(2,1)) with tmv EOS and constant rotation problem 6.7(RotConst) with van der Waals EOS. Method: ER–HLLEM scheme with DEOMOD reconstruction in conservative variables. The results using the mixed GLM–MHD scheme with different values for c_{rel} are shown. The 1d Riemann problem is computed on $\Omega = [-\frac{1}{2}, \frac{1}{2}]\times[-\frac{1}{4}, \frac{1}{4}]$; this leads to $c_{\text{rel}} \approx 14$ for the mixed approach. The rotation problem is computed on $[-2, 2]^2$, which leads to $c_{\text{rel}} \approx 2.2$. The results are not conclusive since $c_{\text{rel}} = 100$ leads to good results for the 1d Riemann problem, but the scheme crashes for the same value of c_{rel} in the case of the rotation problem with $\beta = 2$.

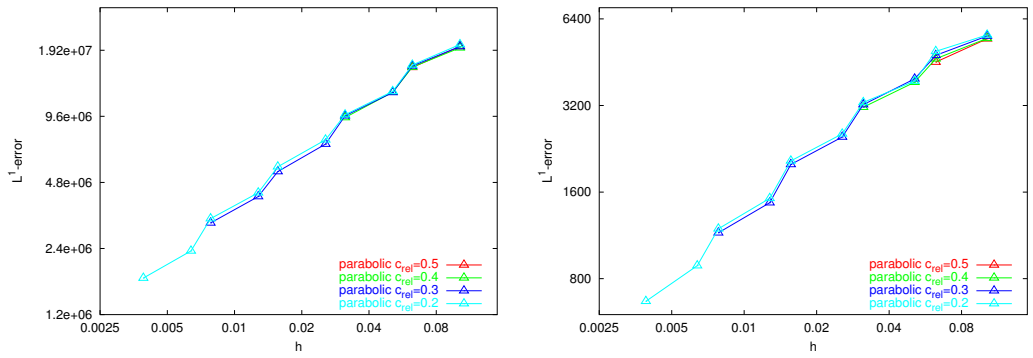
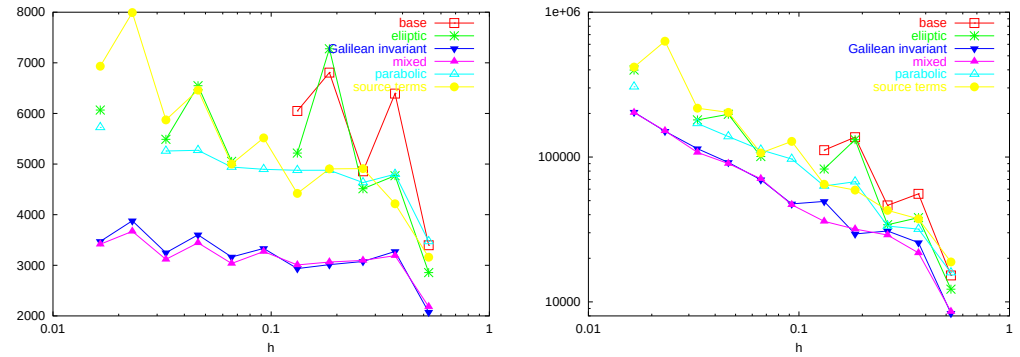


Figure 8.2: Problem: 1d Riemann problem 6.6(RP2D(2,1)) with tmv EOS. Method: ER–HLLEM scheme with DEOMOD reconstruction in conservative variables. The constant c_{rel} only slightly influences the error introduced by the parabolic GLM–MHD scheme. At the same time, the scheme is unstable if c_{rel} is chosen too large. The left figure shows the error for the whole state vector; the right shows the error in the component u_y.

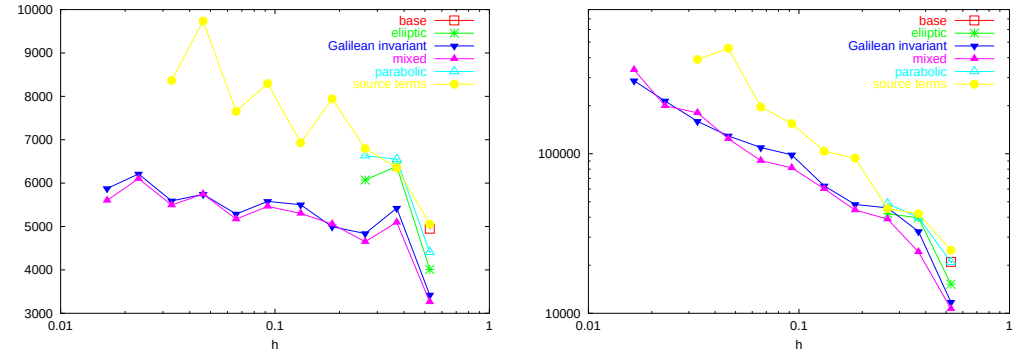
grid resolutions. At level five the overall error is only slightly reduced, but the error in the density is reduced by almost fifty percent. (Note that the error summed over all conservative variables is dominated by the total energy density ρe, which is about $2.7e8$ at the interface $r = R$ due to the high pressure p.) What is even more important is that the EOC is around 0.3–0.4 even at very high grid resolutions. The observations made for $\beta = 1$ are confirmed by the results for $\beta = 2$. Of the six schemes tested only the mixed and the Galilean invariant schemes are able to compute the solution for all grid resolutions. For very large values of the magnetic field even the GLM–MHD correction mechanism is not sufficient as the results for $\beta = 4$ in Figure 8.5 demonstrate. In this case only the source term fix, the Galilean invariant, and the mixed corrections lead to results — but only on very coarse grids.

Let us now take a closer look at the magnetic field. In the following we concentrate on the base scheme, the mixed correction, and the source term fix. All results are computed on refinement level eight. In Figure 8.6 we show a scatter plot of $|(\mathbf{B}_x, \mathbf{B}_y)|$. We plot the numerical solution versus the radius r. The exact solution grows linear for $r < 1$ and is zero for $r > 1$. At the interface $r = 1$ the solution is discontinuous. For $\beta = 1$ and at a very early time in the simulation, the three results shown are still quite similar (cf. Figure 8.6 top row). A closer look shows that the base scheme and the source term fix lead to oscillations at the interface. Without a divergence correction these oscillations grow in time as can be seen in the second row of Figure 8.6. Note the different scaling of the vertical axis. Both the mixed and the source term approaches lead to a smearing of the interface. The source term fix introduces strong oscillations behind the interface. The oscillations present in the base scheme cause the simulation to crash. We also observe a stronger break in the rotational symmetry in the approximation using the source term fix. Even with $\beta = 2$ the rotational symmetry is almost intact when the mixed approach is used.

We conclude our tests for the constant rotation problem with a look at the divergence



(a) $\beta = 1$: divergence errors left in the L^1–norm, right in the L^∞–norm



(b) $\beta = 2$: divergence errors left in the L^1–norm, right in the L^∞–norm

Figure 8.3: Problem: Constant rotation problem 6.7(RotConst) with van der Waals EOS. Method: ER–HLLEM scheme with DEOMOD reconstruction in conservative variables. Divergence errors for the base scheme and the five correction methods versus the grid resolution are plotted using the expressions (8.24) to approximate $\nabla \cdot \mathbf{B}$.

errors measured in the L^1– and in the L^∞–norm for different grid parameters h. The results are shown in Figure 8.3. For both $\beta = 1$ and $\beta = 2$ we see a similar picture, which confirms the observations made so far. Both the mixed and the Galilean invariant approaches lead to the smallest errors on all grids whereas the other methods all lead to comparable results. In the L^1–norm the mixed and the Galilean invariant approach lead to an almost constant error for all grid resolutions whereas the error is clearly increasing in all other schemes. Since the solution is discontinuous and the base scheme converges with a rate smaller than one, we cannot expect convergence of the derivatives of the solution. For the same reason the divergence errors measured in the L^∞–norm increase with decreasing grid resolution. We conclude by noting that the mixed correction and the Galilean invariant approach clearly produce the smallest errors.

Summary of Section 8.5.2: *In the previous chapter we already noted that with increasing magnetic field strength the violation of the divergence constraint* (1.1e) *causes the base scheme presented in Chapter 3 to become unstable and even to crash. All the extensions of the base scheme presented here greatly increase the stability of the scheme, although to different degrees. From our tests we conclude that the mixed and the Galilean*

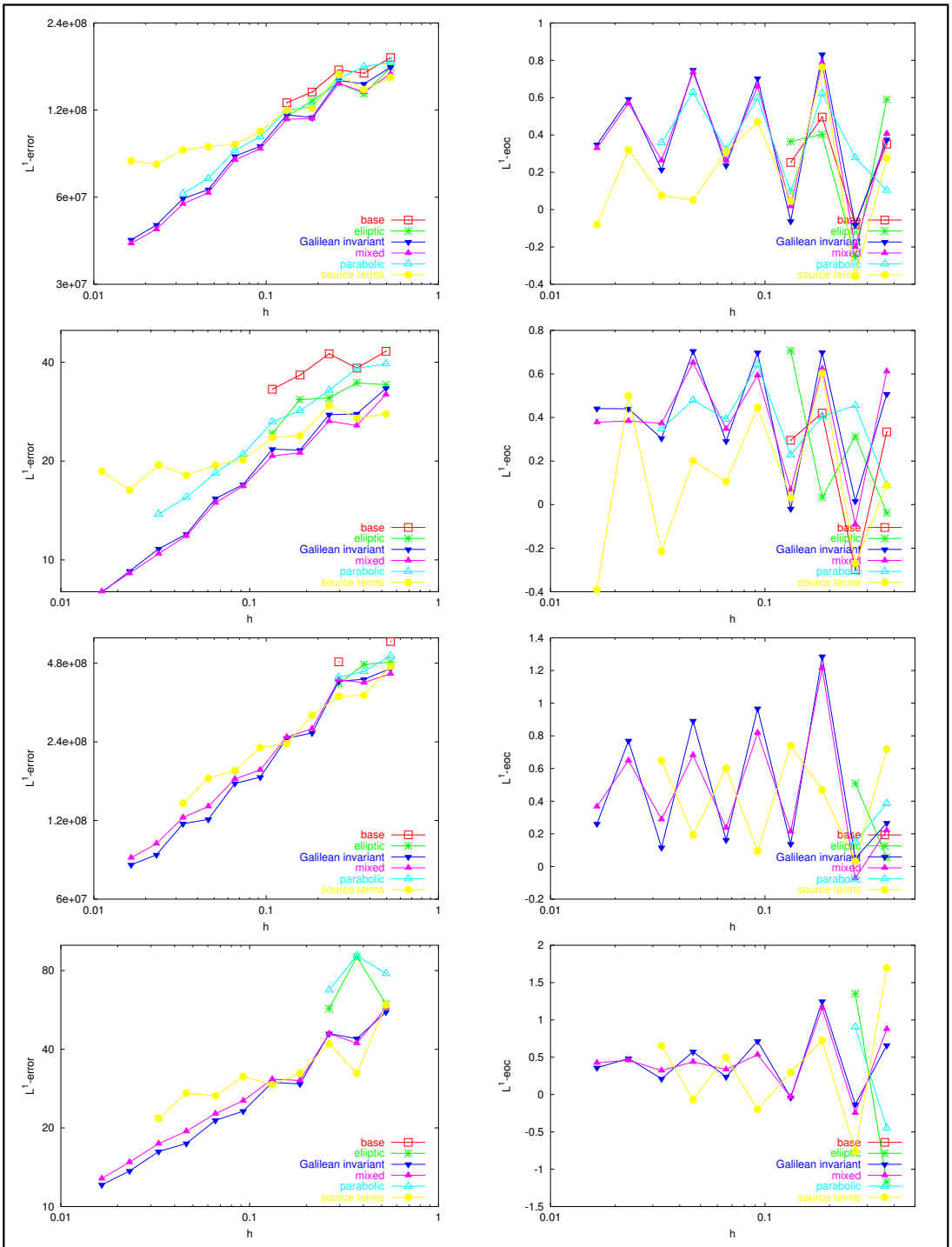


Figure 8.4: Problem: Constant rotation problem 6.7(RotConst) with van der Waals EOS. Method: ER–HLLEM scheme with DEOMOD reconstruction in conservative variables. The L^1–errors EOCs for different correction approaches are plotted for the whole state vector and for the density ρ (top two rows: $\beta = 1$, bottom two rows: $\beta = 2$).

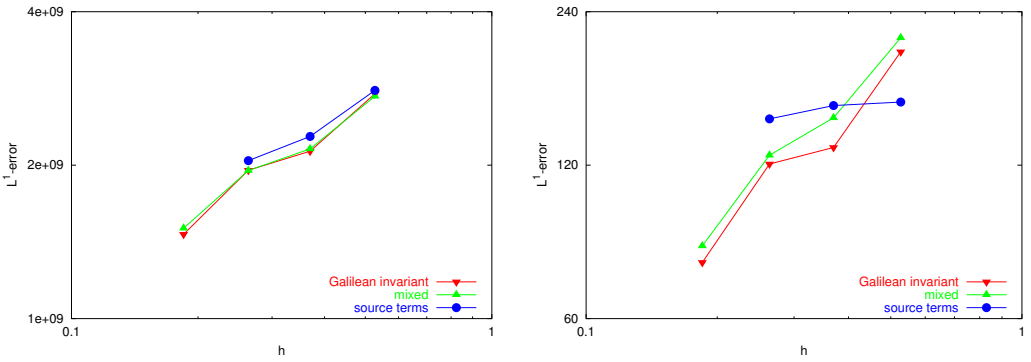


Figure 8.5: Problem: Constant rotation problem 6.7(RotConst) with van der Waals EOS. Method: ER–HLLEM scheme with DEOMOD reconstruction in conservative variables. The L^1–error for the whole state vector (left) and for the density (right) is plotted for very strong magnetic field ($\beta = 4$). Only three of the six tested schemes where able to compute results.

invariant approach lead to the most stable scheme; they also produce the smallest errors for all grid resolutions. A direct comparison of these two correction techniques gives no definite answer as to which scheme should be used. For $\beta = 1$ the errors are smallest when the mixed correction is used, whereas for $\beta = 2$ the Galilean invariant correction leads to slightly better results (cf. Figure 8.4).

8.5.3 2d Riemann Problem

We now turn to a 2d Riemann problem (cf. Problem 6.5(RPtmv2d)). As already discussed in Section 6.2.1 we have no exact solution to this problem; only near the boundaries can we compute a reference solution by using a 1d approximation of the solutions to 1d Riemann problems. Consequently, we know that near the vertical boundaries B_y must be constant and near the horizontal boundaries B_x must be constant. We can use these observations to determine qualitative differences between the correction mechanisms. We show results for the time $t = 0.00032$ since the base scheme crashed shortly after this time; all the other schemes reached the final time $T = 0.0004$ without any problems.

In Figures 8.8 and 8.9 we plot the 2d solutions in the density, the pressure, and the magnetic field components B_x and B_y. A closer look at the isolines shows perturbations in the base scheme solution, which are considerably reduced by all schemes. This is especially easy to see in the magnetic field components in Figure 8.9. In the regions where one of the magnetic field components should be constant due to the divergence constraint, we see disturbances in all the solutions shown, but they are more obvious when the base scheme is used.

We have also plotted the divergence errors and a representation of the grid in Figure 8.10. The plots of the divergence errors indicate that the mixed GLM–MHD method leads to the best results; this can be seen especially in the center of the domain, where the full 2d structure of the solution is developing. The divergence errors are located almost entirely on the discontinuities in the solution. The plots of the different refine-

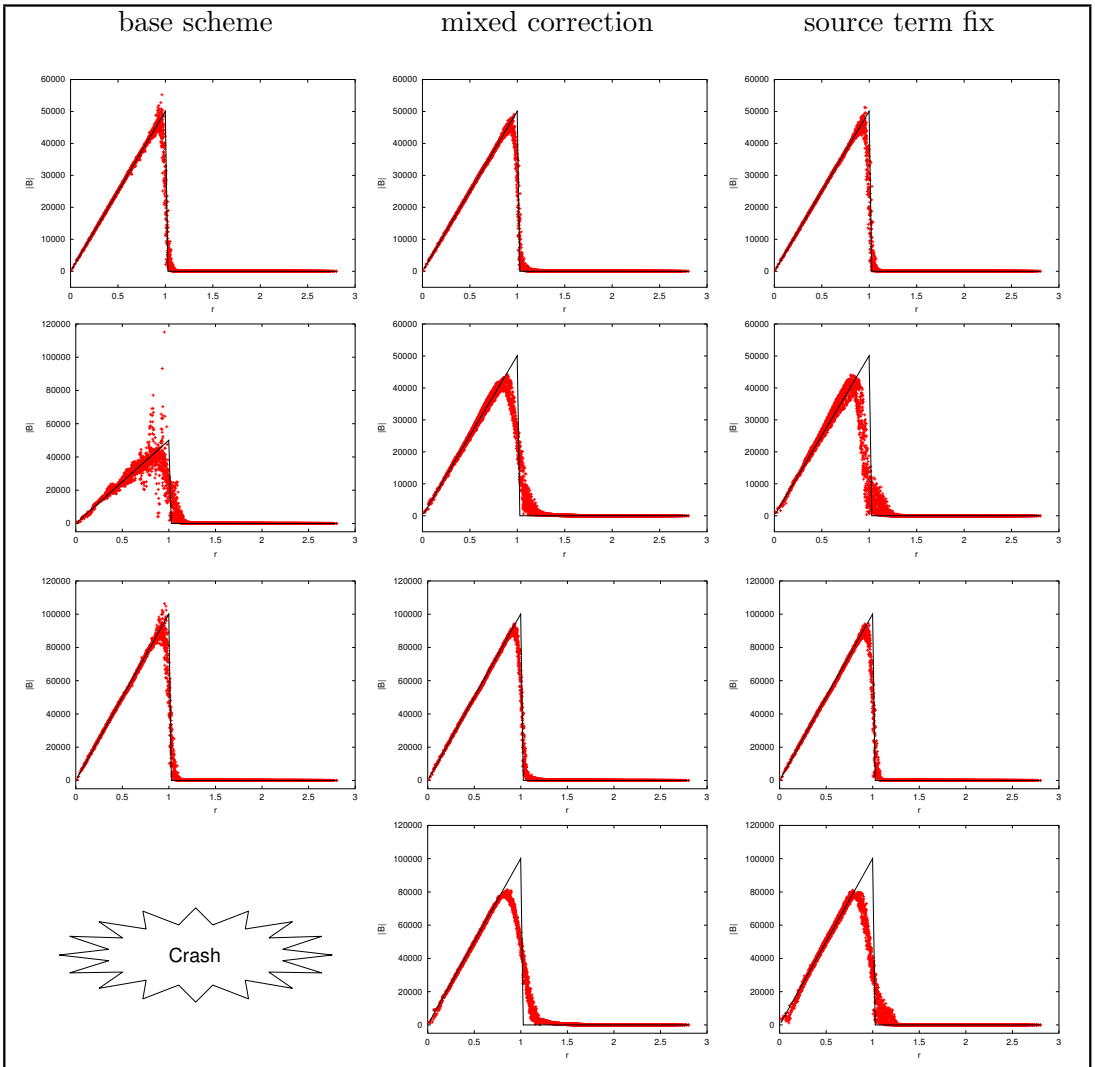


Figure 8.6: Problem: Constant rotation problem 6.7(RotConst) with van der Waals EOS. Method: ER–HLLEM scheme with DEOMOD reconstruction in conservative variables. Scatter plot of the magnetic field strength $|(B_x, B_y)|$ using the base scheme, the mixed correction, and the source term fix (left to right). From top to bottom the value of β is increased: first two rows $\beta = 1$, bottom two rows $\beta = 2$. The results are shown at the beginning of the simulation ($t = 2.5e-4$) and at a later time ($t = 2.5e-3$). After this point in the simulation the base scheme crashed even for $\beta = 1$.

ment levels of the locally adapted grid demonstrate that the correction schemes also improve the efficiency of the base scheme. Due to the spurious oscillations, regions in the domain are refined although the exact solution is constant. This increases the size of the grid and consequently also the computational cost. For the base scheme the grid at time $t = 0.00032$ had about 190000 elements, but only about 140000 elements for the mixed correction and the source term fix, 156000 elements for the elliptic approach, 141000 for the Galilean invariant correction, and about 162000 elements for the parabolic GLM–MHD correction.
It is not easy to find large differences between the different correction methods. To get a more quantitative impression of the approximation quality we, therefore, show scatterplots of the solution near the right boundary together with a reference solution to the Riemann problem 6.6(RP2D(4,1)). In Figures 8.11–8.13 we plot the density ρ, B_x, and B_y, respectively. The spurious oscillations are now clearly visible. Of the correction schemes the parabolic approach again leads to the most problems; the other four corrections are quite similar although the magnitude of the oscillations is smallest in the mixed correction and the source term approach. Note that in contrast to the previous example, the approximation produced by the Galilean invariant approach is not quite as good as the approximation produced by the mixed approach. The results shown so far suggest that the source term fix leads to the best approximation, closely followed by the mixed GLM–MHD correction.

Summary of Section 8.5.3: *Since we have no exact solution to the 2d Riemann problem, we cannot compare the approximation errors of the schemes directly but have to use other means of determining the advantages and disadvantages of the different approaches of handling the problem of divergence errors. All our results clearly demonstrate that all corrections lead to an improvement when compared to the base scheme. The differences between the correction methods are less obvious than in the previous example. So far we have found that the mixed and the Galilean invariant approach are the best methods followed by the source term fix. Here the source term fix seems to be the best approach together with the mixed correction; the Galilean invariant approach leads to stronger oscillations. As before, the parabolic approach is less effective in reducing the errors than the other four correction techniques. So far we have only plotted results at a fixed time. In Figure 8.7 we plot the evolution of the divergence error measured in the L^1– and in the L^∞–norm. Note that initially the divergence error is zero, even after the projection onto the grid since the initial data for both B_x and B_y are constant in the whole domain. This plot shows that the mixed correction leads to the smallest divergence errors, especially in the L^1–norm. In the base scheme the error in L^1–norm increases monotone up to the point where the scheme crashes; after the corrections have been applied to the base scheme the error remains more or less constant in time.*

8.5.4 Conservation Property

We conclude this chapter with a closer look at the conservation property of the different schemes. This issue is only relevant for the source term fix and the Galilean invariant approach since in both cases additional terms that are not in divergence form are added to the MHD equations. These "source terms" are proportional to $\nabla \cdot \mathbf{B}$ so

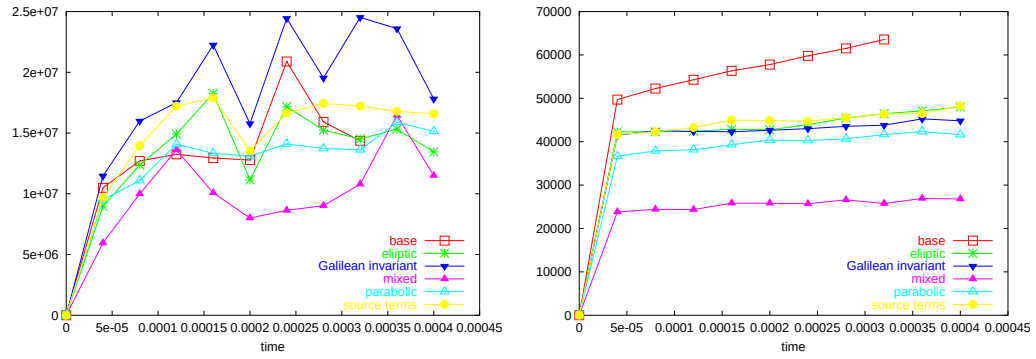
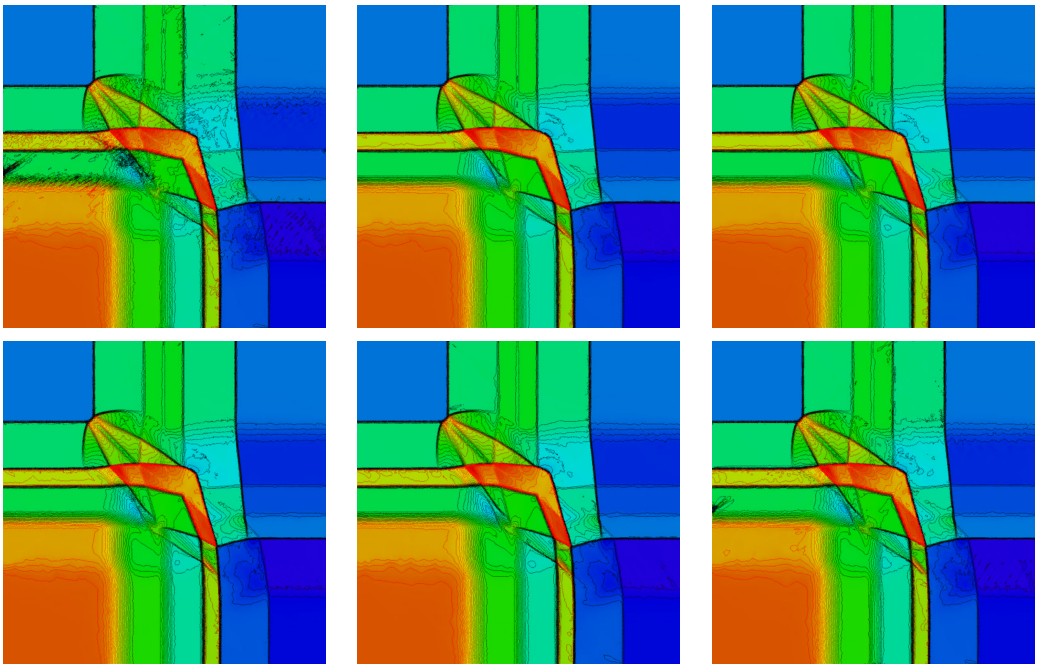


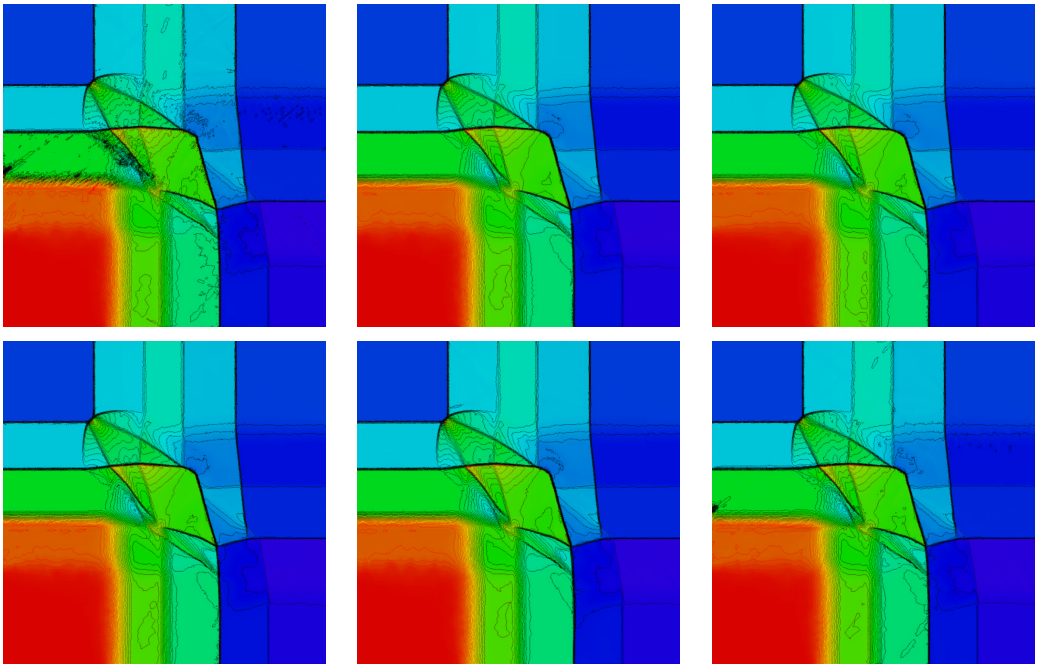
Figure 8.7: Problem 6.5(RPtmv2d) with HLLEM flux function and DEOMOD reconstruction in conservative variables. All computations are performed on the unstructured grid using local grid adaptation. We plot the evolution of the divergence errors using different correction mechanisms in the L^∞–norm (left) and in the L^1–norm (right).

that large divergence errors lead to a significant loss of conservation. In the Galilean invariant approach only the induction equation (8.1c) is modified (the source term in the equation for ψ is not relevant since ψ is not a physical variable). In addition to this source term in the induction equation, the equations for the momentum $\rho\mathbf{u}$ and the total energy density ρe are also modified by the source term fix. Thus only the density satisfies a conservation law. In Figure 8.14 we show the time evolution of the average deviation of the total mass for different components v of the state vector, i.e. we compute $\left|\frac{\sum_{i\in\mathcal{I}}|T_i|(v_i^n - v_i^0)}{\sum_{i\in\mathcal{I}}|T_i|v_i^0}\right|$. Note that this value should be zero since all boundaries are periodic and thus no mass leaves or enters the domain. The results show that the source terms lead to a substantial change in the total mass. Since the divergence errors are smaller in the case of the Galilean invariant approach, the loss of conservation is not quite as large as for the source term fix. Nevertheless, the loss of conservation is still severe and can lead to problem with the convergence of the scheme as our next example indicates.

We study the 1d Riemann problem (Problem 6.2(RPDWT)), which was also used by Tóth [Tót00] to emphasize the problems that can arise if a non–conservative scheme is used. The results are taken from [DKK+02] and are computed using the first order DW method as base scheme. In Figure 8.15 we show 1d–cuts at $y = 0.0424$ of the 2d solutions obtained using the mixed and the source term approach. The 1d reference solution is computed for $h = 0.0002$. In [Tót00] it was already observed that the source term approach leads to a wrong solution if this Riemann problem is solved on a rotated Cartesian grid in 2d. Similarly, we see that the solution obtained using the source term fix on 16384 triangles seems to contain wrong intermediate states; their development can be attributed to the lack of conservation in the source term approach. The L^1–errors and corresponding EOCs shown in Table 8.2 support the observation that wrong intermediate states are computed: for the mixed scheme we have a uniform first order convergence, whereas the convergence rate decreases monotonically for the source term approach. Note that the base scheme fails for all grid resolutions studied.

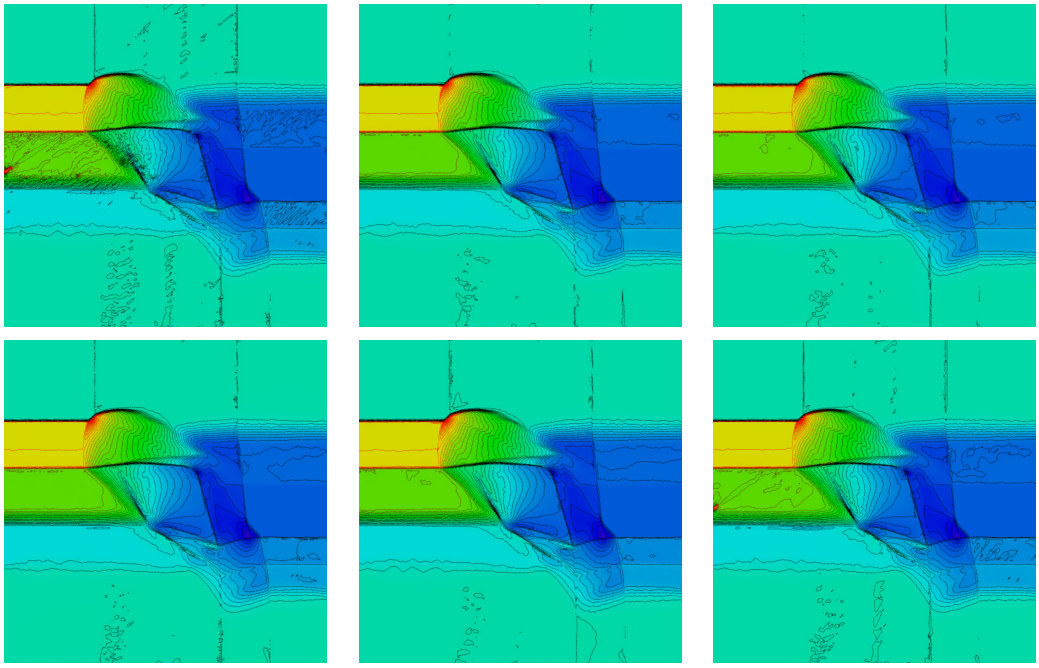


(a) ρ: left to right (top) base,mixed,source term fix, (bottom) elliptic,Galilean invariant, parabolic

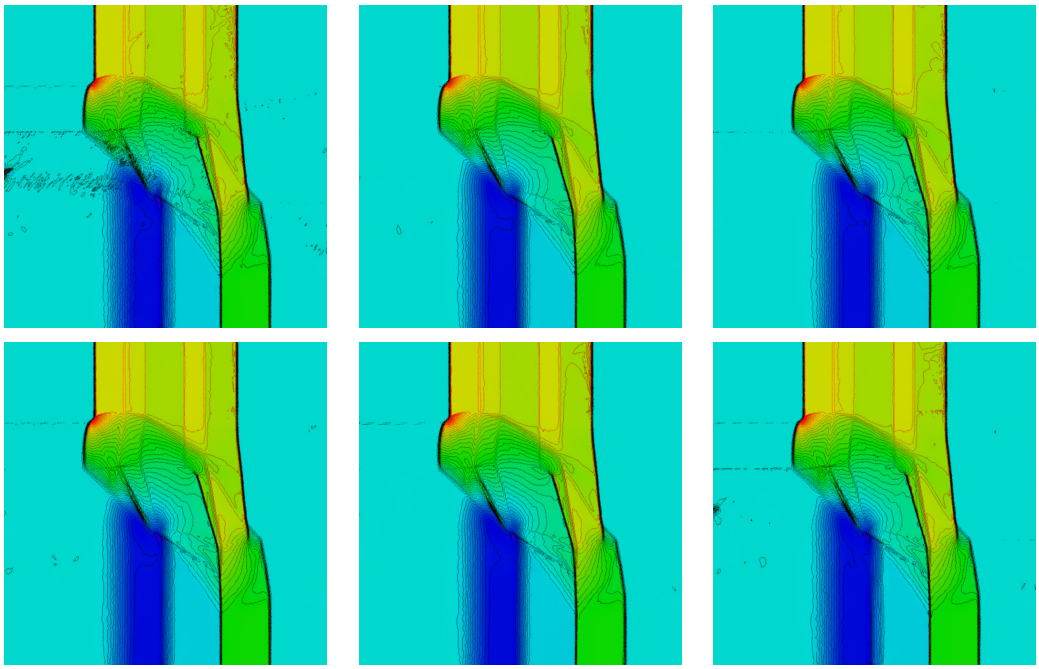


(b) p: left to right (top) base,mixed,source term fix, (bottom) elliptic,Galilean invariant, parabolic

Figure 8.8: Problem 6.5(RPtmv2d) with HLLEM flux function and DEOMOD reconstruction in conservative variables. All computations are performed on the unstructured grid using local grid adaptation. We show a 2d representation of the density (a) and the pressure (b).

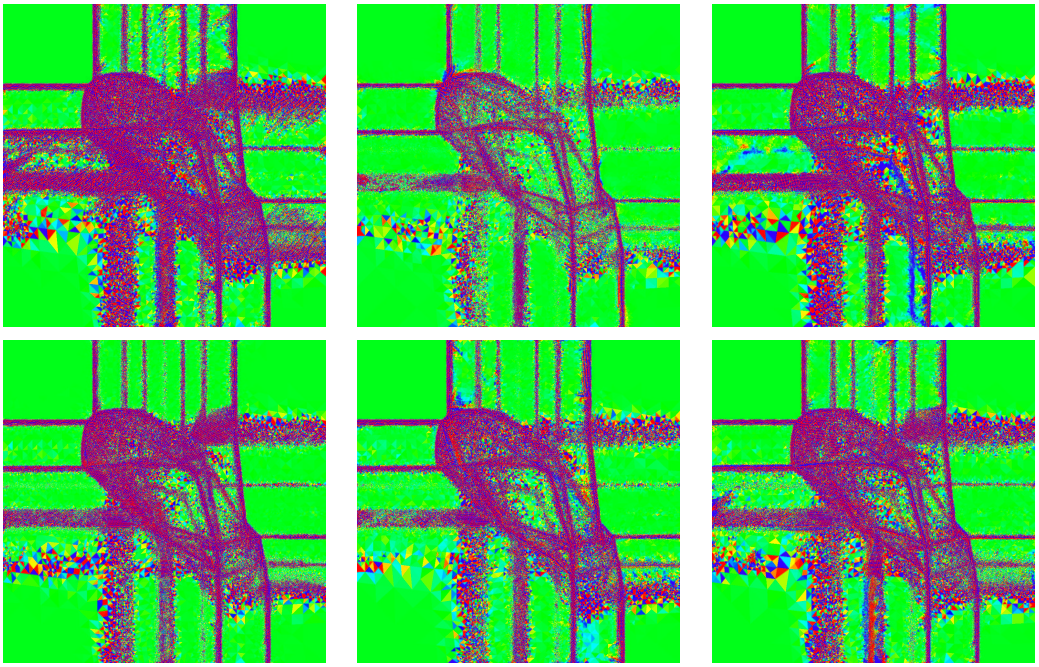


(a) B_x: left to right (top) base,mixed,source term fix, (bottom) elliptic,Galilean invariant, parabolic

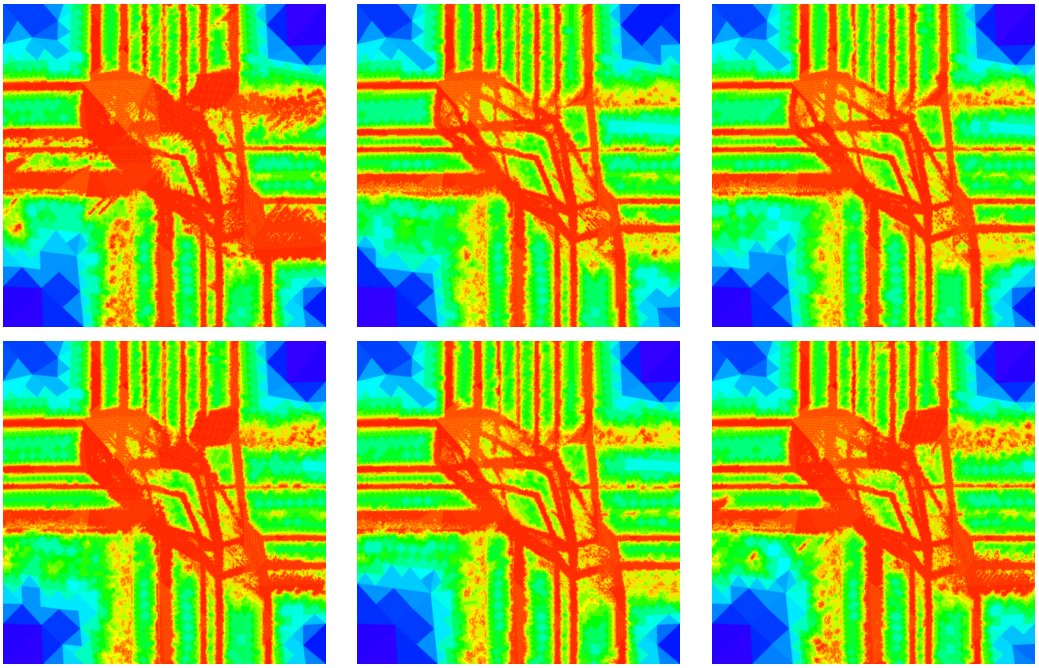


(b) B_y: left to right (top) base,mixed,source term fix, (bottom) elliptic,Galilean invariant, parabolic

Figure 8.9: Problem 6.5(RPtmv2d) with HLLEM flux function and DEOMOD reconstruction in conservative variables. All computations are performed on the unstructured grid using local grid adaptation. We show a 2d representation of the magnetic field components B_x (a) and B_y (b).



(a) $\nabla \cdot \mathbf{B}$: left to right (top) base,mixed,source term fix, (bottom) elliptic,Galilean invariant, parabolic



(b) level: left to right (top) base,mixed,source term fix, (bottom) elliptic,Galilean invariant, parabolic

Figure 8.10: Problem 6.5(RPtmv2d) with HLLEM flux function and DEOMOD reconstruction in conservative variables on the unstructured grid using local grid adaptation. We show a representation of $\nabla \cdot \mathbf{B}$ (a) and of the grid refinement (b).

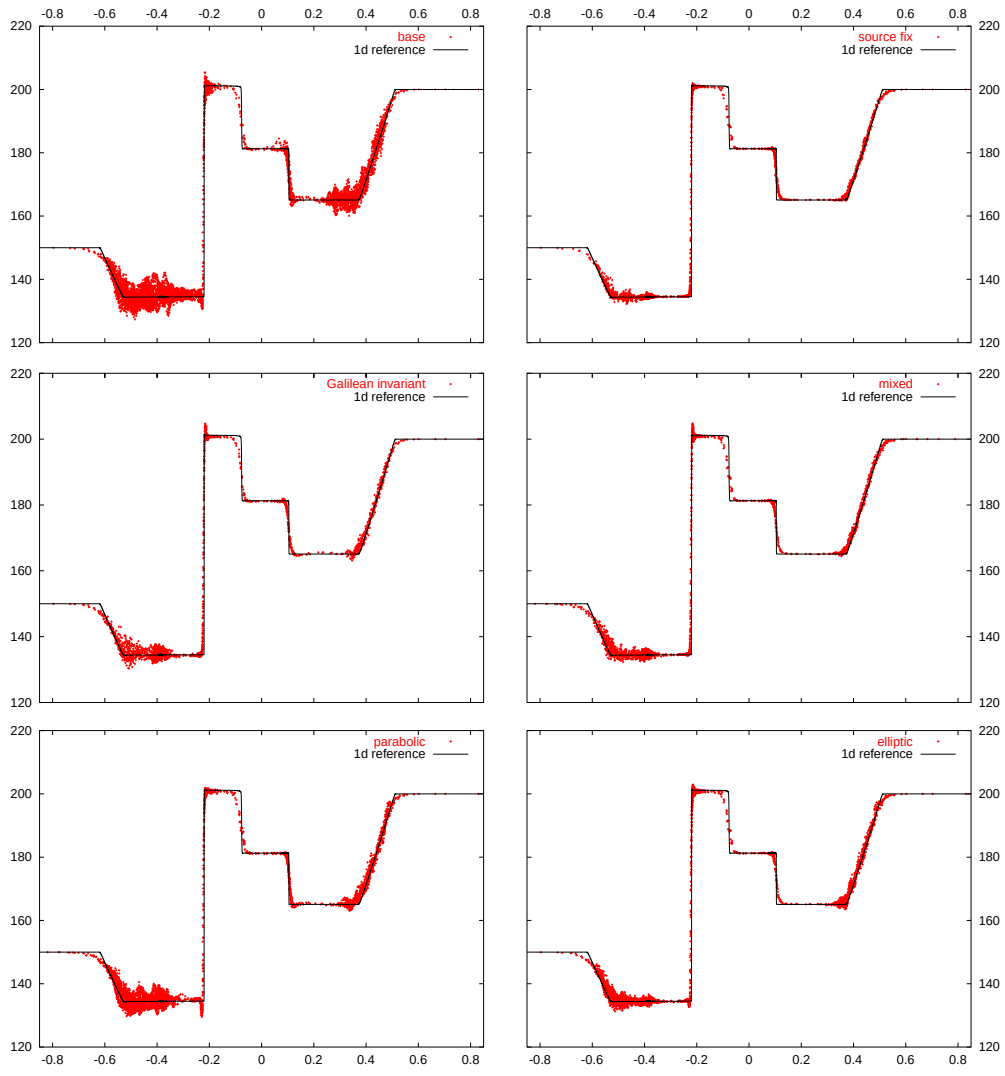


Figure 8.11: Problem 6.5(RPtmv2d) with HLLEM flux function and DEOMOD reconstruction in conservative variables. All computations are performed on the unstructured grid using local grid adaptation. The plots show the density for $x \in [0.87, 0.97]$ together with a 1d reference solution of the Riemann problem between the fourth and the first state. Top to bottom and left to right we plot results for the base scheme, the source term fix, the Galilean invariant, the mixed, the parabolic, and the elliptic GLM–MHD methods.

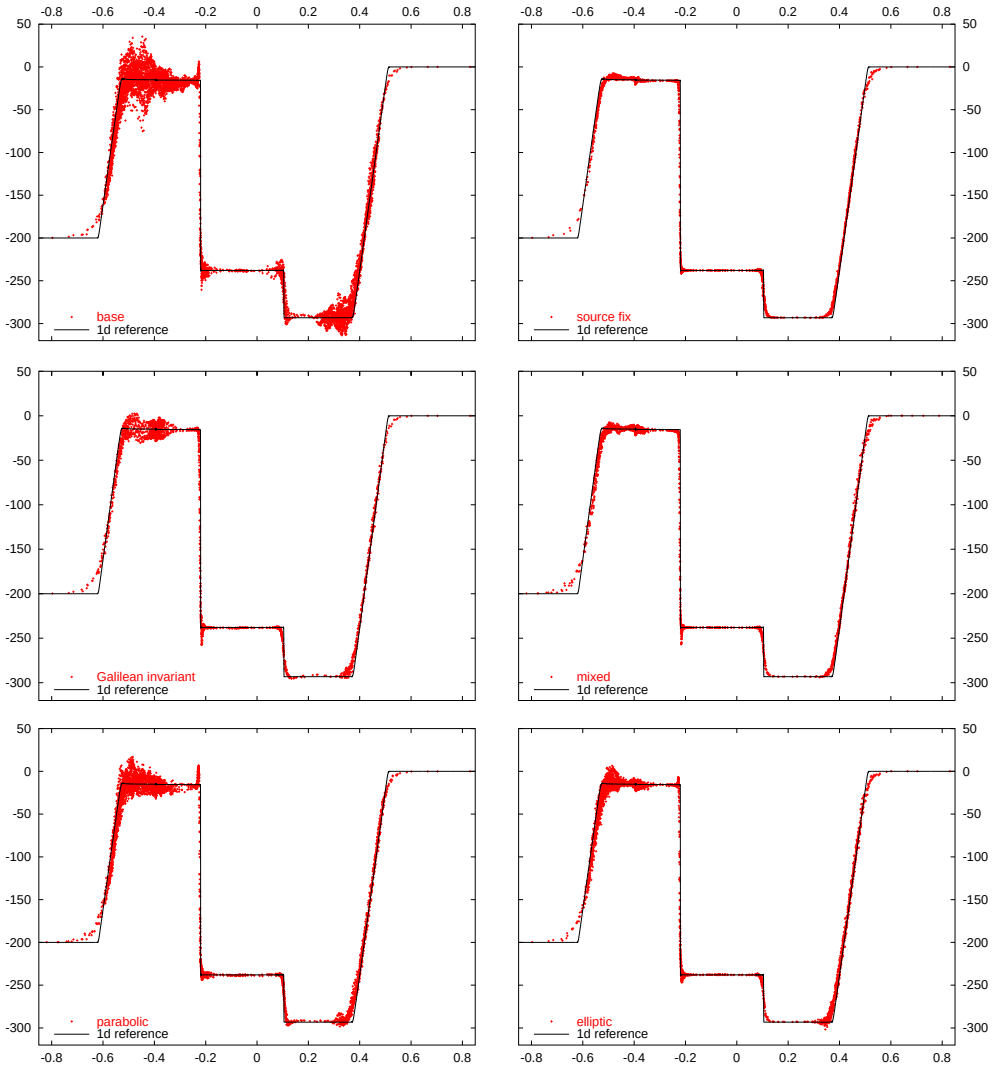


Figure 8.12: Problem 6.5(RPtmv2d) with HLLEM flux function and DEOMOD reconstruction in conservative variables. All computations are performed on the unstructured grid using local grid adaptation. The plots show u_y for $x \in [0.87, 0.97]$ together with a 1d reference solution of the Riemann problem between the fourth and the first state. Top to bottom and left to right we plot results for the base scheme, the source term fix, the Galilean invariant, the mixed, the parabolic, and the elliptic GLM–MHD methods.

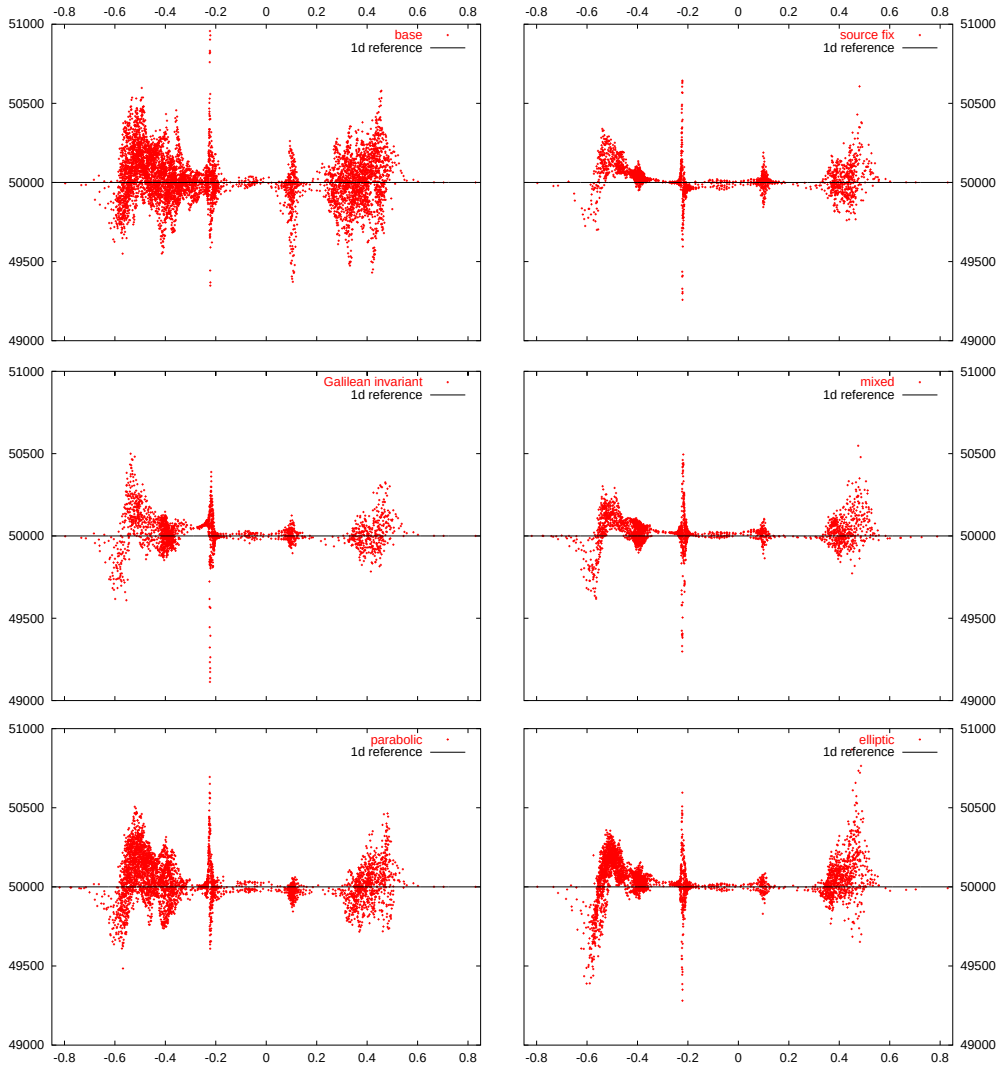
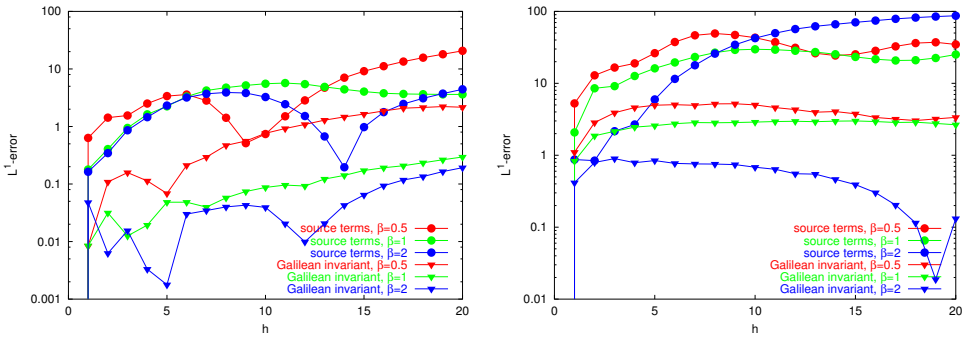
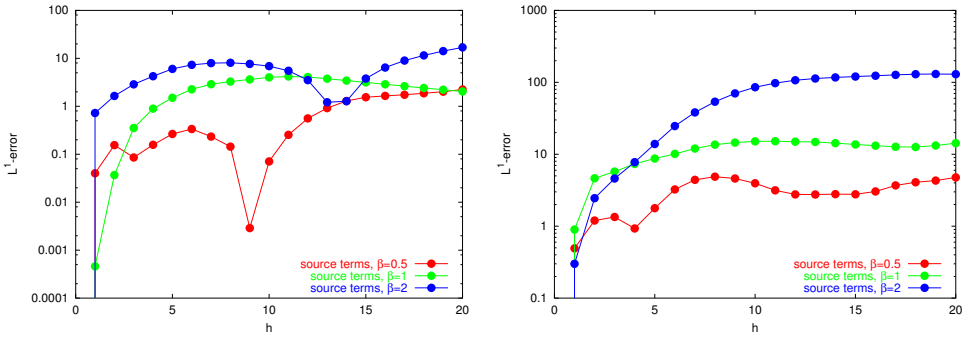


Figure 8.13: Problem 6.5(RPtmv2d) with HLLEM flux function and DEOMOD reconstruction in conservative variables. All computations are performed on the unstructured grid using local grid adaptation. The plots show the B_y for $x \in [0.87, 0.97]$ together with a 1d reference solution of the Riemann problem between the fourth and the first state. Note that due to the divergence constraint, B_y should be constant in this part of the computational domain. Top to bottom and left to right we plot results for the base scheme, the source term fix, the Galilean invariant, the mixed, the parabolic, and the elliptic GLM–MHD methods.



(a) conservation error in B_x,B_y for source term fix and Galilean invariant method



(b) conservation error in ρu_x,ρu_y for source term fix (momentum is conserved in all other methods)

Figure 8.14: Problem: Constant rotation problem 6.7(RotConst) with van der Waals EOS. Method: ER–HLLEM scheme with DEOMOD reconstruction in conservative variables. The relative loss of conservation for different conservative variables is shown. Four of the six schemes tested are conservative and only the source term fix and the Galilean invariant approach lead to errors in the conservation.

Summary of Section 8.5.4: *All the results shown so far lead us to favor the mixed approach either with or without the source terms leading to a Galilean invariant formulation. Without these source terms we only have a non–conservative term in the equation for the auxiliary function ψ, so that the relevant physical quantities still satisfy conservation laws. In the case of the Galilean invariant formulation a term that is not in divergence form is added to the induction equation* (1.1c)*; consequently, the magnetic field* $\mathbf{B}$ *is no longer a conserved quantity. Since these terms are proportional to* $\nabla\cdot\mathbf{B}$*, large errors in* $\nabla\cdot\mathbf{B}$*, which cause the scheme to become unstable, also increase the violation of the conservation property. The results demonstrate that the loss of conservation can be quite severe due to the source terms. Furthermore, we demonstrated that in one example this leads to the computation of a wrong solution. Since the difference in performance between the Galilean invariant scheme and the mixed scheme are very small and considering the problems caused by the loss of conservation, we favor the mixed approach.*

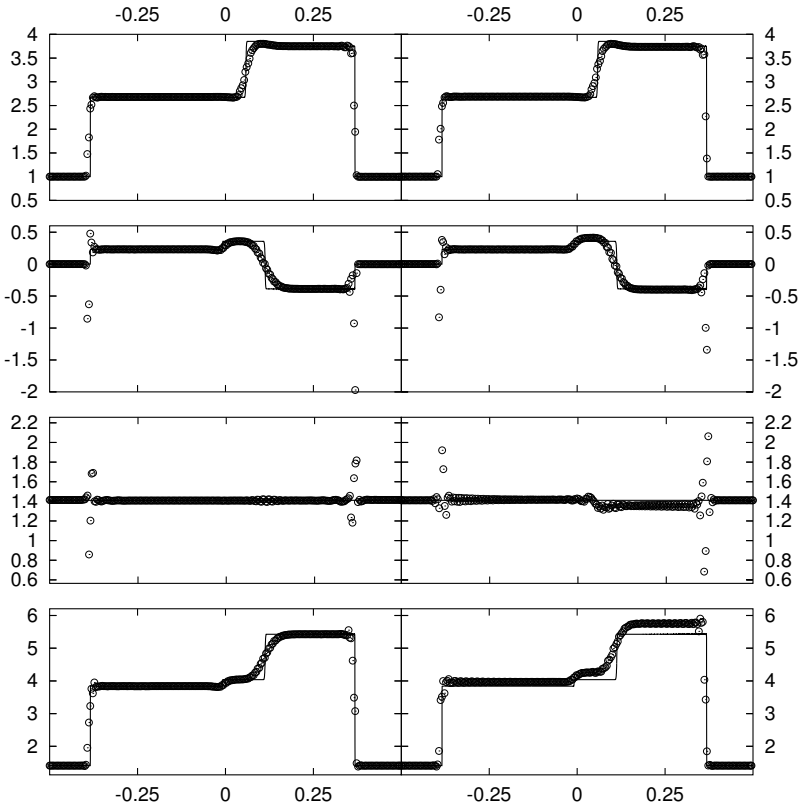


Figure 8.15: Results for 1d Riemann problem 6.2(RPDWT) computed with the first order DW scheme on a 2d grid with 16384 triangles (circles) in comparison with a 1d reference solution ($h = 0.0002$, solid line). Left: mixed GLM ansatz, right: source term approach. From top to bottom: ρ, u_y, B_x, and B_y.

	mixed GLM–MHD		source term	
elements	L^1–error	EOC	L^1–error	EOC
1024	6.45517676		7.03537990	
4096	3.09306537	1.061	3.82213248	0.880
16384	1.68627598	0.875	2.48360779	0.622
65536	0.96600380	0.804	1.84060872	0.432
262144	0.46045425	1.069	no solution	
1048576	0.22792806	1.014	no solution	

Table 8.2: L^1–error and EOC for the 1d Riemann problem 6.2(RPDWT). The errors are computed between a 1d reference solution with $h = 0.0002$ and 1d–cuts of the 2d solutions at $y = 0.0424$.

Chapter 9

Balancing Source Terms: the Bgfix Scheme

In many applications — for example in atmospheric flows — the solution $\mathbf{U}$ to a system of PDEs can be represented as a local perturbation $\widetilde{\mathbf{U}}$ of a known solution $\mathring{\mathbf{U}}$, which we call *background solution* in the following. In the situation of atmospheric flows the background solution is given by balancing the forces of pressure and gravity. Even in the case where the convergence of a numerical scheme has been shown for a grid size $h \to 0$, simulations have to be performed for large values of h. On these grids approximation errors, although small, can still be of the same magnitude or even larger than the perturbations $\widetilde{\mathbf{U}}$ that contribute to the solution. Consequently, the physical characteristics of the problem cannot be captured. This makes it necessary to modify the numerical scheme. The modification we present here is based on a suitable equation for the perturbations themselves. Its main feature is that the background solution is captured without any approximation error. As we will see, this allows for a very efficient solver for this setting.
We derive the scheme for a general system of balance laws (cf. (1.8)):

$$\partial_t \mathbf{U}(\mathbf{x},t) + \nabla \cdot \mathbf{F}(\mathbf{U}(\mathbf{x},t)) = \mathbf{q}(\mathbf{U}(\mathbf{x},t)), \tag{9.1a}$$

$$\mathbf{U}(\cdot,0) = \mathring{\mathbf{U}}(\cdot,0) + \widetilde{\mathbf{U}}_0(\cdot) \tag{9.1b}$$

where $\mathring{\mathbf{U}}$ is a smooth function satisfying

$$\partial_t \mathring{\mathbf{U}}(\mathbf{x},t) + \nabla \cdot \mathbf{F}(\mathring{\mathbf{U}}(\mathbf{x},t)) = \mathbf{q}(\mathring{\mathbf{U}}(\mathbf{x},t)). \tag{9.2}$$

By subtracting (9.2) from (9.1a) we arrive at an equation for the perturbation

$$\widetilde{\mathbf{U}}(\mathbf{x},t) := \mathbf{U}(\mathbf{x},t) - \mathring{\mathbf{U}}(\mathbf{x},t) \tag{9.3}$$

that can be written in the following form

$$\partial_t \widetilde{\mathbf{U}}(\mathbf{x},t) + \nabla \cdot \widetilde{\mathbf{F}}(\widetilde{\mathbf{U}}(\mathbf{x},t),\mathbf{x},t) = \widetilde{\mathbf{q}}(\widetilde{\mathbf{U}}(\mathbf{x},t),\mathbf{x},t), \tag{9.4a}$$

$$\widetilde{\mathbf{U}}(\cdot,0) = \widetilde{\mathbf{U}}_0(\cdot) \tag{9.4b}$$

if we define the flux $\widetilde{\mathbf{F}}$ and the source $\widetilde{\mathbf{q}}$ via

$$\widetilde{\mathbf{F}}(\mathbf{u},\mathbf{x},t) := \mathbf{F}(\mathbf{u}+\mathring{\mathbf{U}}(\mathbf{x},t)) - \mathbf{F}(\mathring{\mathbf{U}}(\mathbf{x},t)), \tag{9.4c}$$

$$\widetilde{\mathbf{q}}(\mathbf{u},\mathbf{x},t) := \mathbf{q}(\mathbf{u}+\mathring{\mathbf{U}}(\mathbf{x},t)) - \mathbf{q}(\mathring{\mathbf{U}}(\mathbf{x},t)). \tag{9.4d}$$

Therefore the function $\widetilde{\mathbf{U}}$ also satisfies a system of hyperbolic balance laws with a flux function and a source term, which depend explicitly on $\mathbf{x}$ and t. If we define $\mathbf{V}(\mathbf{x},t) := \widetilde{\mathbf{U}}(\mathbf{x},t) + \mathring{\mathbf{U}}(\mathbf{x},t)$ where $\widetilde{\mathbf{U}}$ is a weak solution to (9.4), it is easy to see that $\mathbf{V}$ is a weak solution of (9.1). For the scalar case, we show in Section 9.1 that not only does (9.4) have the same weak solutions, but also that if $\widetilde{\mathbf{U}}$ is the entropy solution of (9.4) then $\mathbf{V}$ is the entropy solution of (9.1).

9.1 Remark: *If $\widetilde{\mathbf{U}}_0 \equiv 0$ then $\widetilde{\mathbf{U}} \equiv 0$ is the classical solution to* (9.4) *and $\mathbf{V}$ is equal to the background solution $\mathring{\mathbf{U}}$. Since even first order finite–volume schemes reproduce constant solutions without any approximation error, a scheme based on* (9.4) *can fulfill the requirement that the background solution be reproduced exactly in the case of vanishing initial perturbation $\widetilde{\mathbf{U}}_0$.*

Since the main difficulty in approximating the original problem (9.4a) lies in the errors due to the spatial discretization, we explain the idea of our method in the case of a semi–discrete scheme; the full scheme is presented in Section 9.2 and Section 9.3.

9.2 Definition (Semi–discrete Bgfix scheme)
Consider an approximation $\mathbf{U}_i(\mathbf{x},t)$ to the solution $\mathbf{U}$ of (9.1) *on a grid element T_i satisfying*

$$\partial_t \mathbf{U}_i(\mathbf{x},t) = \mathcal{S}_i\big[(\mathbf{U}_j(\cdot,t))_{j\in\mathcal{J}}\big](\mathbf{x}) \quad \textit{for} \quad \mathbf{x}\in T_i, t>0, \tag{9.5a}$$

$$\mathbf{U}_i(\mathbf{x},0) = \mathcal{P}_i[\mathbf{U}_0(\cdot)](\mathbf{x}). \tag{9.5b}$$

The operator $\mathcal{S}_i$ is assumed to be an approximation of the spatial derivatives and the source term in (9.1a) *and $\mathcal{P}_i$ is some projection operator for the initial data onto the grid element T_i. In the* Bgfix scheme *we define an approximation $\widetilde{\mathbf{U}}_i$ via*

$$\partial_t \widetilde{\mathbf{U}}_i(\mathbf{x},t) = \widetilde{\mathcal{S}}_i\big[(\widetilde{\mathbf{U}}_j(\cdot,t))_{j\in\mathcal{J}}\big](\mathbf{x}) \quad \textit{for} \quad \mathbf{x}\in T_i, t>0\ , \tag{9.6a}$$

$$\widetilde{\mathbf{U}}_i(\mathbf{x},0) = \mathcal{P}_i[\widetilde{\mathbf{U}}_0(\cdot)](\mathbf{x}) \tag{9.6b}$$

with

$$\widetilde{\mathcal{S}}_i\big[(\widetilde{\mathbf{U}}_j(\cdot,t))_{j\in\mathcal{J}}\big] := \mathcal{S}_i\big[(\widetilde{\mathbf{U}}_j(\cdot,t) + \mathring{\mathbf{U}}(\cdot,t)_{|_{T_j}})_j\big] - \mathcal{S}_i\big[(\mathring{\mathbf{U}}(\cdot,t)_{|_{T_j}})_{j\in\mathcal{J}}\big]\ . \tag{9.6c}$$

As approximation of $\mathbf{U}$ on T_i we define

$$\mathbf{V}_i(\mathbf{x},t) := \widetilde{\mathbf{U}}_i(\mathbf{x},t) + \mathring{\mathbf{U}}(\mathbf{x},t)\ . \tag{9.7}$$

9.3 Remark: *If $\widetilde{\mathbf{U}}_0 \equiv 0$ or equivalently $\mathbf{U}_0 \equiv \mathring{\mathbf{U}}(\cdot,0)$ then $\widetilde{\mathbf{U}} \equiv 0$ is a solution to* (9.6) *and therefore $\mathbf{V}_i \equiv \mathring{\mathbf{U}}$. Thus our modified scheme reproduces the background solution without any approximation error, which in general will not be true for the original approximation* (9.5).

In the next section we quantify in which sense our new method is superior to the original scheme. We also answer the question concerning the entropy solutions to scalar versions of the original and of the modified problem (9.1) and (9.4), respectively. For our analysis we use the notation and the results presented in Chapter 4. Then in Section 9.2 and Section 9.3 we derive a fully discrete version of our finite–volume scheme and present numerical results in Section 9.4.

9.1 Analytical Motivation

For the analysis we use the notations and results from Chapter 4. We focus on the scalar versions of (9.1) and (9.4):

$$\partial_t u(\mathbf{x},t) + \nabla \cdot \mathbf{f}(u(\mathbf{x},t)) = q(u(\mathbf{x},t)) \qquad \text{in } \mathbb{R}^d \times (0,T)\ , \tag{9.8a}$$

$$u(\cdot,0) = \mathring{u}(\cdot,0) + \widetilde{u}_0(\cdot) \quad \text{in } \mathbb{R}^d \tag{9.8b}$$

and

$$\partial_t \widetilde{u}(\mathbf{x},t) + \nabla \cdot \widetilde{\mathbf{f}}(\widetilde{u}(\mathbf{x},t),\mathbf{x},t) = \widetilde{q}(\widetilde{u}(\mathbf{x},t),\mathbf{x},t) \quad \text{in } \mathbb{R}^d \times (0,T)\ , \tag{9.9a}$$

$$\widetilde{u}(\cdot,0) = \widetilde{u}_0(\cdot) \qquad \text{in } \mathbb{R}^d \tag{9.9b}$$

where $\mathring{u} \in C^2(\mathbb{R}^d \times \mathbb{R}^+) \cap H^{1,\infty}(\mathbb{R}^d \times \mathbb{R}^+)$ satisfies

$$\partial_t \mathring{u}(\mathbf{x},t) + \nabla \cdot \mathbf{f}(\mathring{u}(\mathbf{x},t)) = q(\mathring{u}(\mathbf{x},t)) \quad \text{in } \mathbb{R}^d \times (0,T)\ . \tag{9.10}$$

For simplicity we assume that $\mathring{u}(\cdot,0)$ and $\widetilde{u}_0(\cdot)$ have compact support although the results presented here can be extended to the general case by using the local L^1–norm instead of using the L^1–norm on the whole space $\mathbb{R}^d$. We assume that flux function and the source term in (9.8) satisfy Assumption 4.3. The flux and the source term for the perturbation are defined by

$$\begin{aligned}\widetilde{\mathbf{f}}(s,\mathbf{x},t) &:= \mathbf{f}(s + \mathring{u}(\mathbf{x},t)) - \mathbf{f}(\mathring{u}(\mathbf{x},t)),\\ \widetilde{q}(s,\mathbf{x},t) &:= q(s + \mathring{u}(\mathbf{x},t)) - q(\mathring{u}(\mathbf{x},t)).\end{aligned}$$

The following result then follows directly:

9.4 Lemma
If the flux function $\mathbf{f}$ *and the source term* q *satisfy Assumption 4.3 then the flux* $\widetilde{\mathbf{f}}$ *and the source term* $\widetilde{q}$ *for the perturbation problem* (9.9) *also satisfy Assumption 4.3.*

We first study the existence of entropy solutions as defined in Definition 4.11.

9.5 Theorem (Entropy solution for Bgfix modification)
Let the data $\mathbf{f}, q,$ *and* u_0 *satisfy Assumption 4.3 then the problems* (9.8) *and* (9.9) *have unique entropy solutions* $u \in W(0,T)$ *and* $\widetilde{u} \in W(0,T)$*, respectively. Furthermore the identity* $\widetilde{u} = u - \mathring{u}$ *holds or, equivalently, the function* $\widetilde{u} + \mathring{u}$ *is the unique entropy solution to* (9.8)*.*

Proof:
As in Theorem 4.12 we can prove the existence of an entropy solution using the vanishing viscosity method. We study the following regularizations: for $\varepsilon > 0$ let u_ε and $\widetilde{u}_\varepsilon$ be the classical solutions of the parabolic equations

$$\partial_t u_\varepsilon(\mathbf{x},t) + \nabla \cdot \mathbf{f}(u_\varepsilon(\mathbf{x},t)) = q(u_\varepsilon(\mathbf{x},t)) + \varepsilon \triangle u_\varepsilon(\mathbf{x},t),$$
$$\partial_t \widetilde{u}_\varepsilon(\mathbf{x},t) + \nabla \cdot \widetilde{\mathbf{f}}(\widetilde{u}_\varepsilon(\mathbf{x},t),\mathbf{x},t) = \widetilde{q}(\widetilde{u}_\varepsilon(\mathbf{x},t),\mathbf{x},t) + \varepsilon \triangle \widetilde{u}_\varepsilon(\mathbf{x},t).$$

with initial data $\widetilde{u}_0 + \mathring{u}$ and $\widetilde{u}_0$, respectively. Under the conditions on $\mathbf{f}$,q, and u_0 we can apply Theorem 4.12. Therefore there exists a function $u \in W(0,T)$ with $u_\varepsilon \to u$ in L^1. Furthermore u is the unique entropy solution of (9.8). As stated in Lemma 9.4 $\widetilde{\mathbf{f}}$ and $\widetilde{q}$ both also fulfill the requirements stated for $\mathbf{f}$ and q in Theorem 4.12. Consequently there also exists a unique entropy solution $\widetilde{u} \in W(0,T)$ of (9.9) with $\widetilde{u}_\varepsilon \to \widetilde{u}$ in L^1.
For $\varepsilon > 0$ define $v_\varepsilon := \widetilde{u}_\varepsilon + \mathring{u}$. Then v_ε is the classical solution of

$$\partial_t v_\varepsilon(\mathbf{x},t) + \nabla \cdot \mathbf{f}(v_\varepsilon(\mathbf{x},t)) = q(v_\varepsilon(\mathbf{x},t)) + \varepsilon \triangle \big(v_\varepsilon(\mathbf{x},t) - \mathring{u}(\mathbf{x},t)\big). \tag{9.11}$$

Since $\mathring{u}$ does not depend on ε, the additional term $\varepsilon \triangle \mathring{u}(\mathbf{x},t)$ vanishes for $\varepsilon \to 0$; as in Theorem 4.12 we conclude that there exists a v with $v_\varepsilon \to v$ and v is the unique entropy solution to (9.8). Therefore, the identity $v \equiv u$ holds, since due to the definition of v_ε we also have $v = \lim_{\varepsilon \to 0} \widetilde{u}_\varepsilon + \mathring{u} = \widetilde{u} + \mathring{u}$. This concludes the proof. □

9.6 Remark: *Only the limits of (v_ε) and (u_ε) are identical; for $\varepsilon > 0$ the functions v_ε and u_ε satisfy different regularizations of the system* (9.8).

We now turn our attention to the convergence properties of the modified scheme.

9.7 Theorem (Convergence result for the Bgfix scheme)
Let the data $\mathbf{f}, q,$ and u_0 satisfy Assumption 4.3. Consider the unique entropy solutions $u, \widetilde{u} \in W(0,T)$ to (9.8) *and* (9.9)*, respectively. Let $\{\mathcal{T}_h\}_h$ be a family of unstructured grids in $\mathbb{R}^d$ satisfying Assumption 4.4 with some $\Delta t > 0$. Consider a family of monotone numerical flux functions $\{g_{ij}\}$ for $\mathbf{f}$ satisfying Definition 4.5 with local Lipschitz constant L_g. For $i \in \mathcal{I}$ and $j \in N(i)$ define the numerical flux*

$$\widetilde{g}_{ij}^n(u,v) := \frac{1}{|S_{ij}|\Delta t} \int_{t^n}^{t^{n+1}} \int_{S_{ij}} \big(g_{ij}(u + \mathring{u}(\mathbf{x},t), v + \mathring{u}(\mathbf{x},t)) - |S_{ij}| \mathbf{f}(\mathring{u}(\mathbf{x},t)) \cdot \mathbf{n}_{ij}\big)\, d\mathbf{x}dt \ . \tag{9.12}$$

Consider the discrete solution $\widetilde{u}_h$ to our first order finite volume scheme (cf. Definition 4.7) with average values defined by

$$\widetilde{u}_i^{n+1} = \widetilde{u}_i^n - \frac{\Delta t}{|T_i|} \sum_{j \in \mathcal{N}(i)} \widetilde{g}_{ij}^n(\widetilde{u}_i^n, \widetilde{u}_j^n) + \frac{1}{|T_i|} \int_{t^n}^{t^{n+1}} \int_{T_i} \widetilde{q}(\widetilde{u}_i^n, \mathbf{x}, t)\, d\mathbf{x}dt \ ,$$

where $\widetilde{u}_i^0$ is given by

$$\widetilde{u}_i^0 = \frac{1}{|T_i|} \int_{T_i} \widetilde{u}_0(\mathbf{x})\, d\mathbf{x} \ .$$

If Δt satisfies

$$\frac{\Delta t}{h} \leq \frac{(1-\xi)c_G^2}{2L_g(\|\widetilde{u}_0 + \mathring{u}\|_\infty)}$$

for some $\xi \in (0,1)$ then $\widetilde{u}_h$ converges in L^1 to $\widetilde{u}$. If we define $v_h = \widetilde{u}_h + \mathring{u}$ then there exists a constant $\widetilde{K}$ depending on $\mathbf{f}, \widetilde{u}_0, \mathring{u}, q, c_G, \xi$ such that

$$\int_0^T \int_{\mathbb{R}^d} |v_h(\mathbf{x},t) - u(\mathbf{x},t)| \, d\mathbf{x}dt \leq \widetilde{K} h^{\frac{1}{4}} \ . \tag{9.13}$$

Proof:
We want to apply Theorem 4.20 to the approximation $\widetilde{u}_h$. From Lemma 9.4 we know that the data in problem (9.9) satisfy the conditions from Theorem 4.20. We still have to verify that $\widetilde{g}_{ij}^n$ defines a family of monotone numerical flux functions for the analytical flux $\widetilde{\mathbf{f}}$. Using the definition (9.12) of $\widetilde{g}$, it is easy to verify, that all the conditions from Definition 4.5 are satisfied by $\widetilde{g}_{ij}$, if they are satisfied by g_{ij}. The Lipschitz constant is given by

$$L_{\widetilde{g}}(M) := L_g\big(M + \|\mathring{u}\|_\infty\big)$$

so that Δt satisfies

$$\frac{\Delta t}{h} \leq \frac{(1-\xi)c_G^2}{2L_{\widetilde{g}}(\|\widetilde{u}_0\|_\infty)}$$

Therefore all conditions from Theorem 4.20 are satisfied. Consequently there exists a constant $\widetilde{K}$ depending on $\widetilde{\mathbf{f}}, \widetilde{u}_0, \widetilde{q}, c_G, \xi$ such that

$$\int_0^T \int_{\mathbb{R}^d} |\widetilde{u}_h(\mathbf{x},t) - \widetilde{u}(\mathbf{x},t)| d\mathbf{x}dt \leq \widetilde{K} h^{\frac{1}{4}} \ .$$

Using our definition of v_h it directly follows that

$$\begin{aligned}\int_0^T \int_{\mathbb{R}^d} |v_h(\mathbf{x},t) - u(\mathbf{x},t)| d\mathbf{x}dt &= \int_0^T \int_{\mathbb{R}^d} |\widetilde{u}_h(\mathbf{x},t) + \mathring{u}(\mathbf{x},t) - u(\mathbf{x},t)| d\mathbf{x}dt \\ &= \int_0^T \int_{\mathbb{R}^d} |\widetilde{u}_h(\mathbf{x},t) - \widetilde{u}(\mathbf{x},t)| d\mathbf{x}dt \leq \widetilde{K} h^{\frac{1}{4}}\end{aligned}$$

since due to Theorem 9.5 we have $u = \widetilde{u} + \mathring{u}$. This concludes the proof. □

9.8 Remark: *Both schemes converge to the unique entropy solution, under the same conditions on the grid, the time–step, and the data — the time–step restriction introduced by the finite–volume scheme is the same for the Bgfix scheme as for the base scheme. For the base scheme we have a convergence rate of $Kh^{\frac{1}{4}}$ and for the Bgfix the*

same order of convergence but in this case with a constant $\widetilde{K}$. The relationship between K and $\widetilde{K}$ indicates the difference between the base scheme and the modified scheme. An improvement in the order of the scheme cannot be expected as long as the smoothness assumptions on $\widetilde{u}$ and u are the same. Thus only through a smaller constant can an improvement be expected. We are not able to quantify the relation between K and $\widetilde{K}$; consequently, the conditions under which the quality of approximation is improved by our modification are not clear from the analysis.

In the remaining part of this section we show how the Bgfix modification qualitatively changes the approximation to the original problem. We first study a semi–discrete first order finite–volume scheme in one space dimension. For simplicity we use the Lax–Friedrichs scheme to discretize the spatial derivative in (9.1a). Using the notation introduced in Definition 9.2, the approximation on each cell $T_i := [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ is then given by $\mathcal{P}_i[U_0(\cdot)](x) := U_0(x_i)$, and

$$\begin{aligned}&\mathcal{S}_i\big[(U_j(t))_j\big](x) :=\\ &\quad -\frac{1}{\Delta x}\big(g(U_i(x_{i+\frac{1}{2}},t),U_{i+1}(x_{i+\frac{1}{2}},t)) - g(U_{i-1}(x_{i-\frac{1}{2}},t),U_i(x_{i-\frac{1}{2}},t))\big) + q(U_i(x_i,t))\end{aligned}$$

with $x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ and $x_i := i\Delta x, x_{i+\frac{1}{2}} := (i+\frac{1}{2})\Delta x$. The numerical flux function is given by (cf. page 68)

$$g(v,w) := \frac{1}{2}\big(f(v)+f(w)\big) - \frac{1}{2}\big(v-w\big)\ .$$

Using the definition of our modified scheme, we arrive at the following discretization:

$$\begin{aligned}\widetilde{\mathcal{S}}_i\big[(\widetilde{U}_j(t))_j\big](x) &= -\frac{1}{\Delta x}\\ &\quad \big(g(\widetilde{U}_i(x_{i+\frac{1}{2}},t)+\mathring{U}(x_{i+\frac{1}{2}},t),\widetilde{U}_{i+1}(x_{i+\frac{1}{2}},t)+\mathring{U}(x_{i+\frac{1}{2}},t)) -\\ &\qquad g(\widetilde{U}_{i-1}(x_{i-\frac{1}{2}},t)+\mathring{U}(x_{i-\frac{1}{2}},t),\widetilde{U}_i(x_{i-\frac{1}{2}},t)+\mathring{U}(x_{i-\frac{1}{2}},t))\big)\\ &\quad + q(\widetilde{U}_i(x_i,t)+\mathring{U}(x_i,t))\\ &\quad + \frac{1}{\Delta x}\big(g(\mathring{U}(x_{i+\frac{1}{2}},t),\mathring{U}(x_{i+\frac{1}{2}},t)) - g(\mathring{U}(x_{i-\frac{1}{2}},t),\mathring{U}(x_{i-\frac{1}{2}},t))\big)\\ &\quad - q(\mathring{U}(x_i,t))\\ &= -\frac{1}{\Delta x}\big(\widetilde{g}(\widetilde{U}_i(x_{i+\frac{1}{2}},t),\widetilde{U}_{i+1}(x_{i+\frac{1}{2}},t),x_{i+\frac{1}{2}},t) -\\ &\qquad \widetilde{g}(\widetilde{U}_{i-1}(x_{i-\frac{1}{2}},t),\widetilde{U}_i(x_{i-\frac{1}{2}},t),x_{i-\frac{1}{2}},t)\big)\\ &\quad + \widetilde{q}(U_i(x_i,t),x_i,t)\ .\end{aligned}$$

The modified flux function $\widetilde{g}$ is defined by

$$\begin{aligned}\widetilde{g}(\widetilde{v},\widetilde{w},x,t) &:= g(\widetilde{v}+\mathring{U}(x,t),\widetilde{w}+\mathring{U}(x,t)) - g(\mathring{U}(x,t),\mathring{U}(x,t))\\ &= \frac{1}{2}\big(f(\widetilde{v}+\mathring{U}(x,t))+f(\widetilde{w}+\mathring{U}(x,t))\big) - \frac{1}{2}\big(\widetilde{u}-\widetilde{w}\big) - f(\mathring{U}(x,t))\\ &= \frac{1}{2}\big(\widetilde{f}(\widetilde{v},x,t)+\widetilde{f}(\widetilde{w},x,t)\big) - \frac{1}{2}\big(\widetilde{u}-\widetilde{w}\big)\ .\end{aligned}$$

9.9 Remark: *The discretization given by $\widetilde{\mathcal{S}}_i$ is equivalent to the Lax–Friedrichs scheme for the modified equation* (9.4a)*. The additional cost of evaluating $\widetilde{\mathcal{S}}_i$ compared to $\mathcal{S}_i$ is equal to the cost of one evaluation of the analytical flux f and one source term evaluation. Thus no additional numerical flux evaluations are required, which are in general far more expensive than the evaluation of the analytical flux f.*

We now study the approximation properties of U_i and $\widetilde{U}_i$ to (9.1) and (9.4), respectively.

9.10 Theorem
The function U_i is a first order approximation of (9.8) *and a second order approximation of*

$$\partial_t W(x,t) + \partial_x f(W(x,t)) = q(W(x,t)) + \frac{1}{2}\Delta x \partial_x^2 W(x,t). \tag{9.14}$$

The function $\widetilde{U}_i$ is a first order approximation of (9.9) *and a second order approximation of*

$$\partial_t \widetilde{W}(x,t) + \partial_x \widetilde{f}(\widetilde{W}(x,t),x,t) = \widetilde{q}(\widetilde{W}(x,t),x,t) + \frac{1}{2}\Delta x \partial_x^2 \widetilde{W}(x,t). \tag{9.15}$$

Furthermore $V_i(x_i,t) := \widetilde{U}_i(x_i,t) + \mathring{U}(x_i,t)$ is a first order approximation of (9.8) *and a second order approximation of*

$$\partial_t W(x,t) + \partial_x f(W(x,t)) = q(W(x,t)) + \frac{1}{2}\Delta x \partial_x^2 \Big(W(x,t) - \mathring{U}(x,t)\Big). \tag{9.16}$$

Proof:
In the following let $t \in \mathbb{R}^+$ be fixed. The first two statements of the Theorem are a direct consequence of the following result: For a smooth function $w = w(x,t)$ we write w_i to denote $w(x_i,t)$. Let $k = k(w,x,t)$ be some smooth flux functions, then the corresponding Lax–Friedrichs flux satisfies

$$\begin{aligned} -\big(g(w_i, w_{i+1}, x_{i+\frac{1}{2}}, t) - g(w_{i-1}, w_i, x_{i-\frac{1}{2}}, t)\big) = \\ -\frac{1}{2}\big(k(w_i, x_{i+\frac{1}{2}}, t) + k(w_{i+1}, x_{i+\frac{1}{2}}, t) - k(w_{i-1}, x_{i-\frac{1}{2}}, t) - k(w_i, x_{i-\frac{1}{2}}, t)\big) \\ +\frac{1}{2}\big(w_{i+1} - 2w_i + w_{i-1}\big) \\ = -\Delta x \partial_x k(w(x_i), x_i, t) + \frac{\Delta x^2}{2}\partial_x^2 w(x_i) + O(\Delta x^3). \end{aligned}$$

The last equation follows by 2d Taylor expansion of k at $(w(x_i), x_i, t)$.

The third statement follows using the approximation result (9.15) for $\widetilde{U}_i$:

$$\begin{aligned}
\partial_t V_i(x_i,t) &= \widetilde{\mathcal{S}}_i\big[(\widetilde{U}_i(\cdot,t))_i\big](x) + \partial_t \mathring{U}(x,t) \\
&= -\partial_x \widetilde{f}(\widetilde{U}_i(x_i,t),x_i,t) + \widetilde{q}(\widetilde{U}_i(x_i,t),x_i,t) \\
&\quad + \frac{1}{2}\Delta x \partial_x^2 \widetilde{U}_i(x_i,t) + O(\Delta x^2) + \partial_t \mathring{U}(x,t) \\
&= -\partial_x f(\widetilde{U}_i(x_i,t) + \mathring{U}(x_i,t)) + \partial_x f(\mathring{U}(x_i,t)) \\
&\quad + q(\widetilde{U}_i(x_i,t) + \mathring{U}(x_i,t)) - q(\mathring{U}(x_i,t)) \\
&\quad + \frac{1}{2}\Delta x \partial_x^2 \widetilde{U}_i(x_i,t) + O(\Delta x^2) + \partial_t \mathring{U}(x_i,t) \\
&= -\partial_x f(V_i(x_i,t)) + q(V_i(x_i,t)) \\
&\quad + \frac{1}{2}\Delta x \partial_x^2 \big(V_i(x_i,t) - \mathring{U}(x_i,t)\big) + O(\Delta x^2).
\end{aligned}$$

This concludes the proof. □

9.11 Remark: *The parabolic equations* (9.14), (9.15), *and* (9.16) *are called* modified equations. *The main idea behind the study of the modified equation for a given scheme is that it gives an idea of the form of the numerical viscosity. The results from Theorem 9.10 correspond to the parabolic regularizations used in the proof of Theorem 9.5 (compare equation* (9.16) *with* (9.11)*).*

We conclude this section with an example demonstrating the effect of the Bgfix method in the special case of a linear ordinary differential equation, i.e., in the simple case where the flux $\mathbf{f}$ vanishes and the source term q is linear.

9.12 Example
Instead of (9.8) *consider the solution u to the linear ODE*

$$\begin{aligned}
u'(t) &= \lambda u(t) \quad \textit{for } t > 0, \\
u(0) &= \widetilde{u}_0 + \mathring{u}
\end{aligned}$$

with $\widetilde{u}_0, \mathring{u} \in \mathbb{R}$. We use the forward Euler scheme to discretize this ODE:

$$\begin{aligned}
u^0 &= \widetilde{u}_0 + \mathring{u}\ , \\
u^{n+1} &= (1 + \Delta t \lambda) u^n
\end{aligned}$$

with $\Delta t > 0$ fixed. Consider the corresponding ODE for the perturbation (cf. (9.9)*)*

$$\begin{aligned}
\widetilde{u}'(t) &= \lambda \widetilde{u}(t) \quad \textit{for } t > 0\ , \\
\widetilde{u}(0) &= \widetilde{u}_0
\end{aligned}$$

which we also discretize using the forward Euler method

$$\begin{aligned}
\widetilde{u}^0 &= \widetilde{u}_0\ , \\
\widetilde{u}^{n+1} &= (1 + \Delta t \lambda) \widetilde{u}^n\ .
\end{aligned}$$

Define $v^n := \widetilde{u}^n + e^{n\Delta t}\mathring{u}$. *Then the following inequalities hold*

$$|u^n - u(n\Delta t)| \leq |(1+\Delta t\lambda)^n - e^{\lambda n\Delta t}|\ |\widetilde{u}_0 + \mathring{u}|\ ,$$
$$|v^n - u(n\Delta t)| \leq |(1+\Delta t\lambda)^n - e^{\lambda n\Delta t}|\ |\widetilde{u}_0|.$$

This can easily be shown using a standard induction argument.

9.13 Remark: *As in Theorem 9.7 we only encounter an improvement in the constant in front of the error term. The rate of convergence and the evolution of the error in time are not influenced by the modification. If* $|\widetilde{u}_0|$ *is small compared to* $|\widetilde{u}_0+\mathring{u}|$ *then* v^n *leads to a better approximation of* $u(n\Delta t)$ *than* u^n. *On the other hand, if, for example,* $\widetilde{u} = -\mathring{u}$ *then the original scheme produces the exact solution (the constant zero), whereas the modified scheme is a bad approximation. Therefore, using our modified scheme, we expect a big improvement in the approximation only if the perturbations* $\widetilde{u}_0$ *are small. In our numerical examples at the end of this chapter we see that even for large perturbations our modified scheme gives results that are at least comparable to the results of the original scheme.*

9.2 Numerical Scheme

We now return to the system case (9.1) and apply our modification to the first order finite volume scheme (3.5):

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \frac{\Delta t^n}{|T_i|}\sum_j \mathbf{g}_{ij}(\mathbf{U}_i^n, \mathbf{U}_j^n) + \Delta t^n \mathbf{q}(\mathbf{U}_i^n),$$

with

$$\mathbf{U}_i^0 = \mathring{\mathbf{U}}(\boldsymbol{\omega}_i, 0) + \widetilde{\mathbf{U}}_0(\boldsymbol{\omega}_i).$$

The approximation to $\mathbf{U}$ is defined on $T_i \times (t^n, t^{n+1}]$ by

$$\mathbf{U}_h(\mathbf{x}, t) = \mathbf{U}_i^n.$$

Following the approach used in Theorem 9.7 we use the numerical flux function $\mathbf{g}_{ij}$ to define a numerical flux for $\widetilde{\mathbf{F}}$:

$$\widetilde{\mathbf{g}}_{ij}^n(\mathbf{V}, \mathbf{W}) := \mathbf{g}_{ij}(\mathbf{V} + \mathring{\mathbf{U}}(\mathbf{z}_{ij}, t^n), \mathbf{W} + \mathring{\mathbf{U}}(\mathbf{z}_{ij}, t^n)) - |S_{ij}|\mathbf{F}(\mathring{\mathbf{U}}(\mathbf{z}_{ij}, t^n)) \cdot \mathbf{n}_{ij}. \quad (9.17)$$

We can now derive a finite–volume scheme to approximate $\widetilde{\mathbf{U}}$:

$$\widetilde{\mathbf{U}}_i^{n+1} = \widetilde{\mathbf{U}}_i^n - \frac{\Delta t^n}{|T_i|}\sum_j \widetilde{\mathbf{g}}_{ij}^n(\widetilde{\mathbf{U}}_i^n, \widetilde{\mathbf{U}}_j^n) + \Delta t^n \widetilde{\mathbf{q}}(\widetilde{\mathbf{U}}_i^n, \boldsymbol{\omega}_i, t^n), \quad (9.18)$$

with

$$\widetilde{\mathbf{U}}_i^0 = \widetilde{\mathbf{U}}_0(\boldsymbol{\omega}_i).$$

Since we are interested in an approximation of $\mathbf{U}$ and not of $\widetilde{\mathbf{U}}$, we define the approximation on $T_i \times (t^n, t^{n+1}]$ by

$$\mathbf{V}_h(\mathbf{x}, t) = \widetilde{\mathbf{U}}_i^n + \mathring{\mathbf{U}}(\mathbf{x}, t).$$

Note that if $\widetilde{\mathbf{U}}_0 \equiv 0$ then $\widetilde{\mathbf{U}}_i^n = 0$ for all i, n and therefore we have $\mathbf{V}_h(\mathbf{x}, t) = \mathring{\mathbf{U}}(\mathbf{x}, t)$ as required.
The scheme can also be rewritten to compute volume data $\mathbf{V}_i^n$ for $\mathbf{U}$ directly:

$$\mathbf{V}_i^{n+1} = \mathbf{V}_i^n - \frac{\Delta t^n}{|T_i|} \sum_j \mathbf{g}_{ij}(\mathbf{V}_i^n - \mathring{\mathbf{U}}(\boldsymbol{\omega}_i, t^n) + \mathring{\mathbf{U}}(\mathbf{z}_{ij}, t^n), \mathbf{V}_j^n - \mathring{\mathbf{U}}(\boldsymbol{\omega}_j, t^n) + \mathring{\mathbf{U}}(\mathbf{z}_{ij}, t^n)) + \Delta t^n \mathbf{q}(\mathbf{V}_i^n) + \Delta t^n \mathrm{Corr}_i^n \quad (9.19)$$

where the correction term to the first order finite–volume scheme is defined as

$$\mathrm{Corr}_i^n := \frac{1}{|T_i|} \sum_j |S_{ij}| \mathbf{F}(\mathring{\mathbf{U}}(\mathbf{z}_{ij}, t^n)) \cdot \mathbf{n}_{ij} - \mathbf{q}(\mathring{\mathbf{U}}(\boldsymbol{\omega}_i, t^n)) \quad (9.20)$$

where we used the consistency of the numerical flux function with the analytical flux, i.e. $\mathbf{g}_{ij}(\mathring{\mathbf{U}}(\mathbf{z}_{ij}, t^n), \mathring{\mathbf{U}}(\mathbf{z}_{ij}, t^n) = |S_{ij}| \mathbf{F}(\mathring{\mathbf{U}}(\mathbf{z}_{ij}, t^n)) \cdot \mathbf{n}_{ij}$. In this form the modification can be easily added to an existing scheme. Note that the approximation on $T_i \times (t^n, t^{n+1}]$ is now given by $\widetilde{\mathbf{U}}_i^n + \mathring{\mathbf{U}}(\boldsymbol{\omega}_i, t^n)$ and therefore an additional approximation error of the background solution is introduced into the approximation of $\mathbf{U}$, which is not present if we use the scheme to compute the perturbations themselves.

9.14 Remark: *The correction term in the first order finite–volume scheme represents the consistency error introduced by the spatial discretization of the scheme applied to the background solution $\mathring{\mathbf{U}}$. In general we do not simply subtract $\mathring{\mathbf{U}}(\boldsymbol{\omega}_{ij}, t^n)$ but have to use the same projection as for the initial data, i.e., we have to use $\mathcal{P}_i[\mathring{\mathbf{U}}(\cdot)](\mathbf{z}_{ij})$ for the flux computation. In the same sense the source contribution to the correction term has to be modified according to the approximation of the source in the base scheme. Since the background solution is assumed to be continuous, the consistency of the numerical flux allows us to use the analytical flux in the correction term. Consequently, we can write the correction term in the form* (9.20)*, which is independent of the numerical flux function used in the base scheme.*
Using the first form of the Bgfix scheme (9.17)*, we have to compute the values of the background solution in the points $(\mathbf{z}_{ij}, t^n)$ for $(i, j) \in \mathcal{I}_S$ and each time–step t^n; furthermore, depending on the approximation of the source term, some additional values of $\mathring{\mathbf{U}}$, for example at the center of gravity of each cell, have to be computed. If the background solution $\mathring{\mathbf{U}}$ does not depend on time t but only on the space variable $\mathbf{x}$ — as is often the case (e.g. atmospheric flow) — we can store these values during the generation of the grid. In the case where it is costly to compute $\mathring{\mathbf{U}}$ this can greatly improve the performance of the scheme. In the second form of the Bgfix scheme we can also store the required values of the background solution such as $\mathcal{P}_i[\mathring{\mathbf{U}}(\cdot)](\mathbf{z}_{ij})$. In the case of a static background atmosphere it suffices to store the density and the pressure so that the additional cost in memory is not too high.*

9.3 Higher Order and Adaptivity

Besides its ability to reproduce the background solution $\mathring{\mathbf{U}}$, the Bgfix scheme has an important additional advantage when used with local grid adaption and higher order reconstruction. Since the exact value of the background solution at a point $\mathbf{x}, t$ is added without approximation errors, a very coarse grid suffices in regions of the domain where the perturbation from the background solution is small. Thus in simulations where the perturbations are local in space a large computational domain can be used without requiring a fine grid — even in regions where the background solution varies strongly a coarse grid is sufficient. Taking a large computational domain has the advantage that the influence of artificial boundary conditions is substantially diminished. In the original scheme we would require a fine grid in the full domain, which leads to an inefficient scheme. In our adaption indicator (3.19) we use the jump of the discrete approximation on the cell interfaces. In the case of our first order base scheme these are given by the difference of the constant values on neighboring cells. In the case of the modified scheme these are equal to the difference of the perturbations. Thus if the perturbations are small the grid will automatically not be refined. Therefore we do not have to modify our adaptation criteria to take advantage of the Bgfix correction. Only the prolongation and restriction of the data on the new grid must now be performed for the perturbations themselves.
The same property of the Bgfix scheme that allows us to use a coarse grid is also advantageous for the linear reconstruction of the data in the higher order finite–volume scheme. Since we can reconstruct the perturbations directly, the nonlinear behavior in the solution due to $\mathring{\mathbf{U}}$ does not reduce the quality of the reconstruction. For example in regions where $\widetilde{\mathbf{U}}$ is linear but $\mathring{\mathbf{U}}$ is nonlinear, the reconstruction of the perturbations $\widetilde{\mathbf{U}}$ leads to a good approximation of the solution, whereas the reconstruction of $\mathbf{U}$ itself leads to poor results.

9.15 Remark: *If the numerical scheme is used to approximate the perturbations directly as shown in the previous section, then the algorithm does not have to be modified to take advantage of the improved features presented above. If, on the other hand,* $\mathbf{V}_i^n$ *is computed from* (9.19) *then the reconstruction should be performed for* $\mathbf{V}_i^n - \mathring{\mathbf{U}}(\boldsymbol{\omega}_i, t^n)$.

9.4 Numerical Results

To study the Bgfix scheme we apply our correction technique to the mixed GLM–MHD method described in the previous chapter; in the following the term *base scheme* has to be understood in this sense. The main application for which we developed and tested the Bgfix method is atmospheric flow. In many ways this is a special case since the gravitational source terms given by (1.3) are linear with respect to the conservative variables $\mathbf{U}$. In this situation the modified source term $\widetilde{\mathbf{q}}$ is equal to the original source term $\mathbf{q}$, as can easily be seen from equation (9.4d). Consequently, the source term does not depend on the background solution $\mathring{\mathbf{U}}$. Another aspect of this type of problem is that all components of the background solution are zero, with the exception of the density ρ and the pressure p. The flux in a given direction $\mathbf{n}$ therefore depends only on the pressure and is given by $(0, p\mathbf{n}, \mathbf{0}, 0)$; since the pressure is a non–linear function

of the conservative variables, the flux is also non–linear. Before we study variations of atmospheric flow in detail, we study a purely academic test case where we define the source term $\mathbf{q}$ in such a way that a given function $\mathring{\mathbf{U}}$ is a background solution to the MHD equations.

9.4.1 Model Problem

We define the function $\mathring{\mathbf{U}}$ and the source term $\mathbf{q}$ via

$$\mathring{\mathbf{U}}(\mathbf{x}) := \begin{pmatrix} \frac{1}{u_0+\varphi(\mathbf{x})} \\ 1 \\ 1 \\ 0 \\ \mathbf{0} \\ \frac{p_0}{\gamma-1} + u_0 + \varphi(\mathbf{x}) \end{pmatrix} \text{ and } \mathbf{q}(\mathbf{U},\mathbf{x}) := \begin{pmatrix} 0 \\ \mathrm{div}\varphi(\mathbf{x}) \\ \mathrm{div}\varphi(\mathbf{x}) \\ 0 \\ \mathbf{0} \\ \frac{\gamma}{\gamma-1} p(\mathbf{U})\mathrm{div}\varphi(\mathbf{x}) + 2\mathbf{u}\cdot\nabla\varphi(\mathbf{x}) \end{pmatrix}.$$

for some given scalar function $\varphi \in C^1(\mathbb{R}^2)$ and positive constants u_0, p_0. The function $\mathring{\mathbf{U}}$ is a solution to the MHD equations for a perfect gas law augmented by the source term $\mathbf{q}$. Note that the pressure is equal to the constant p_0 on the whole domain. In our simulations we chose $\gamma = \frac{5}{3}$ and

$$\varphi(\mathbf{x}) = \begin{cases} (1 - 10(x^2+y^2))^2 & \text{if } x^2+y^2 < 0.1\ , \\ 0 & \text{otherwise}\ ; \end{cases}$$

furthermore $u_0 = 2$ and $p_0 = 1.5$. With this choice $\mathring{\mathbf{U}}$ is once continuously differentiable but has discontinuous second derivatives, and $\mathring{\mathbf{U}}(\mathbf{x})$ lies in the state space $\mathcal{U}$ for all $\mathbf{x} \in \mathbb{R}^2$ since $\varphi \geq 0$ and $u_0, p_0 > 0$. If we choose the initial data $\mathbf{U}_0$ for our simulation equal to $\mathring{\mathbf{U}}$, the Bgfix scheme approximates the solution without an error on any grid. This is therefore not a suitable test for the scheme since this does not represent the situation found in applications. Instead we choose the initial and boundary data in such a way that the solution $\mathbf{U}(\cdot,t)$ is close to $\mathring{\mathbf{U}}$ for sufficiently large t. As initial conditions we take $\rho_0 \equiv \frac{1}{u_0}$, $\mathbf{u}_0 \equiv (\frac{1}{2}u_0, \frac{1}{2}u_0)$, $\mathbf{B}_0 \equiv \mathbf{0}$, and choose the pressure equal to the constant p_0. We perform the simulation on the domain $\Omega = [-\frac{1}{2}, \frac{1}{2}]^2$ with outflow boundary conditions for $x = \frac{1}{2}$ and $y = \frac{1}{2}$ and inflow boundary conditions on the lower and the left boundary. As inflow function we choose $\mathring{\mathbf{U}}$, which is constant on the boundaries of Ω. Note that the initial data is not close to the solution $\mathring{\mathbf{U}}$ in any component so that the initial perturbation $\widetilde{\mathbf{U}}_0$ is quite large.
In Figure 9.1 we plot the evolution of the difference in the L^1–norm between $\mathring{\mathbf{U}}$ and the approximations for different values of h using both the base scheme and the Bgfix method. For $t \geq 4$ the approximation has reached a quasi–stationary state so that the difference $\|\mathbf{U}_h(\cdot,t) - \mathring{\mathbf{U}}(\cdot,t)\|$ is constant. For small t the approximation is almost independent of the scheme used. In this case the solution $\mathbf{U}(\cdot,t)$ is not known, but it is not close to the background solution $\mathring{\mathbf{U}}$, which is used for the Bgfix correction. Consequently, the Bgfix scheme cannot be expected to lead to an improvement compared to the base scheme. It is important to note that the Bgfix scheme apparently does not

lead to any loss of accuracy. In Figure 9.2 we show a sequence of the discrete density and pressure for different values of t together with the plot of the difference between the base scheme and the Bgfix modification.
For large values of t the advantage of our modification becomes apparent. The stationary solution obtained with the Bgfix method is identical up to roundoff errors with the background solution $\mathring{\mathbf{U}}$, whereas the error between $\mathring{\mathbf{U}}$ and the approximation using the base scheme is several orders of magnitude larger and only decreases slowly through grid refinement (cf. Figure 9.1). The difference in magnitude of the error on grid levels five and nine, compared to the error on level seven, is not clear; but, although the error on level seven is two orders of magnitude larger than the error on the other levels, it is nevertheless three orders of magnitude smaller than the errors produced by the base scheme on all grids.
Figure 9.3 shows the error in the density and the pressure at $t = 6$ using the base scheme. Upwind of the region where the function φ is non–zero the approximation shows disturbances that do not decrease in time. These are not present in the Bgfix scheme, where the difference between the approximation and $\mathring{\mathbf{U}}$ is too small to plot. Note that the results for $t = 4$ is almost identical to the results shown here so that a quasi–stationary solution has been reached. Only the small scale perturbations downwind from the center are transported out of the domain, and new perturbations are constantly being generated at the boundary of the regions where φ does not vanish.

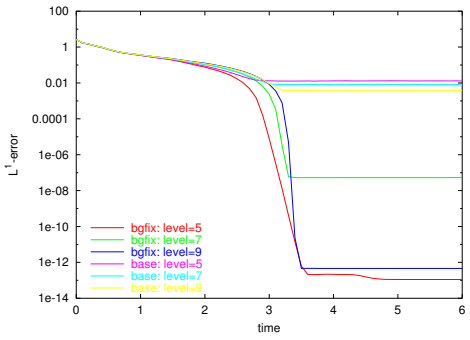


Figure 9.1: Model problem for Bgfix scheme described in Section 9.4.1. The evolution of the difference in the L^1–norm between the background solution and the approximation for different refinement levels of the unstructured macro grid is shown. For $t < 2$ the Bgfix scheme leads to the same approximation as the base scheme. In the stationary state the difference between the Bgfix approximation and $\mathring{\mathbf{U}}$ is in the regime of roundoff errors, whereas the error introduced by the base scheme is still many orders of magnitude larger.

Summary of Section 9.4.1: *Using the construction principle described in Section 3.7.5, we obtaining a test case for the MHD equations, with which we can demonstrate the efficiency of the Bgfix scheme. On the one hand, our results show that the Bgfix modification does not decrease the quality of the base scheme in the case where the computed solution $\mathbf{U}_h$ is far away from $\mathring{\mathbf{U}}$; in the initial phase, where the initial data are transported out of the domain, both schemes lead to the same approximate solution. On the other hand, as the solution approaches steady state, the Bgfix method almost reproduces the exact solution $\mathring{\mathbf{U}}$ whereas the base scheme leads to significant errors downwind of the region where the source does not vanish; these errors are barely reduced by increasing the simulation time or through grid refinement.*

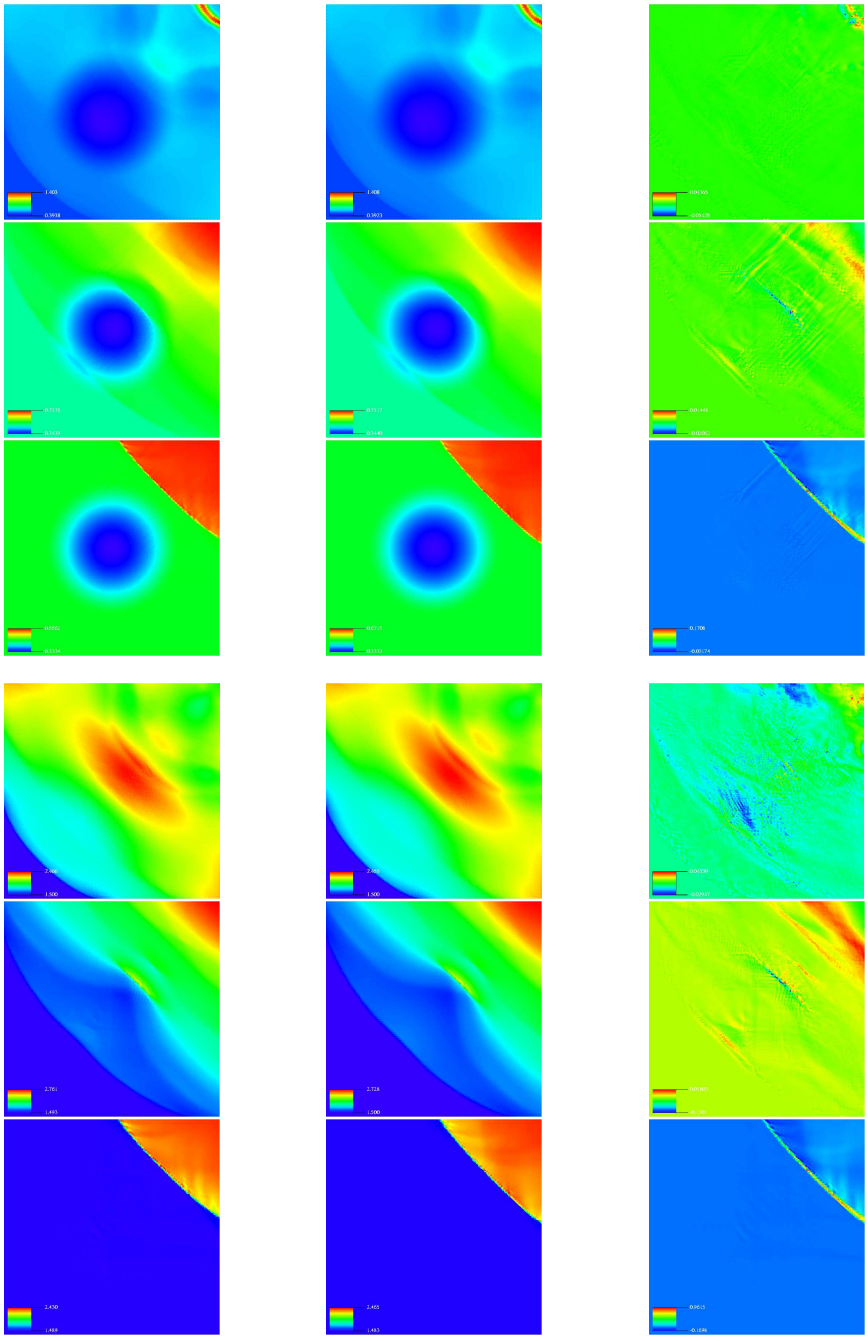


Figure 9.2: Model problem for Bgfix scheme described in Section 9.4.1 computed on the unstructured grid (refinement level nine). The first three rows show the density and the bottom three rows the pressure for $t = \frac{1}{2}, 1, 2$, respectively. The left column shows results using the base scheme, the middle column the approximation using the Bgfix scheme. The right column shows the difference between the two approximations. Note that the minimum and the maximum of the colorbar is adjusted in each plot.

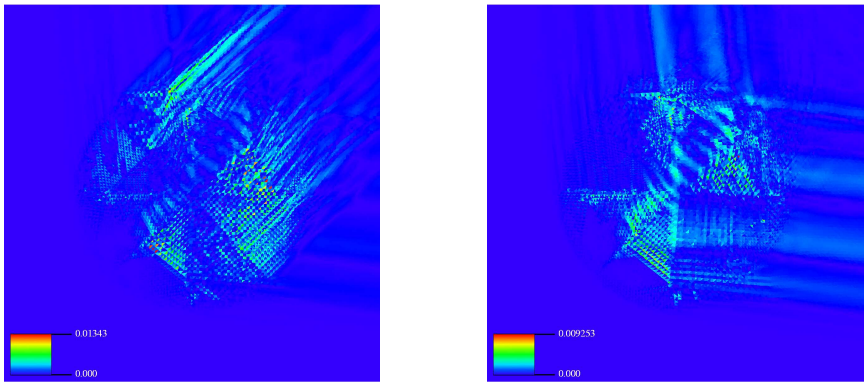


Figure 9.3: Model problem for Bgfix scheme described in Section 9.4.1 computed on the unstructured grid (refinement level nine). The plots show the difference between $\mathring{\mathbf{U}}$ and the approximation using the base scheme in the density (left) and the pressure (right). The results are for $t = 6$. The solution at this point is quasi–stationary. The corresponding results for the Bgfix scheme are not shown since $\mathring{\mathbf{U}}$ is reproduced with an accuracy of 10^{-10}.

9.4.2 Rotation Problem

Next we study the rotation problem but this time embedded in the model solar atmosphere (Problem 6.1(Atm)). As in the constant rotation problem studied in the previous chapter, we have a stationary radial symmetric solution $\mathbf{U} = \mathbf{U}(r)$ with $r = \sqrt{x^2 + y^2}$, but the density is no longer constant and we have a gravity source with the force of gravity pointing to the origin (cf. Problem 6.8(RotAtm) on page 84). For the Bgfix modification we have to define a suitable background solution $\mathring{\mathbf{U}}$. Since we are interested in the computation of a solution superimposed on a static background atmosphere, we only use the density and the pressure profile to define the background solution. The simplest choice is to define the pressure in the background atmosphere as $p_0 + \mathring{p}_{\text{sun}}(r^2 - 6)$. Together with $g_{\text{sun}}\mathring{\rho}_{\text{sun}}(r^2 - 6)$ for the density we then have $\mathring{\mathbf{U}} = \mathbf{U}$ for $r > 1$. Although a perturbation with compact support is the typical case in our applications, we use a different setting that leads to a more challenging test for the Bgfix scheme. We rewrite the pressure of the exact solution as follows

$$p(r) = \begin{cases} \widetilde{p}_0 + \frac{1}{2}u_0^2\mathring{p}_{\text{sun}}(r^2 - 6) - \frac{B_0^2}{4\pi}r^2 + g_{\text{sun}}\mathring{p}_{\text{sun}}(r^2 - 6) & \text{for } r < R\,, \\ \widetilde{p}_0 + \frac{1}{2}u_0^2\mathring{p}_{\text{sun}}(R^2 - 6) - \frac{B_0^2}{4\pi}R^2 + g_{\text{sun}}\mathring{p}_{\text{sun}}(r^2 - 6) & \text{for } r > R \end{cases}$$

with the constant $\widetilde{p}_0 = p_0 - \frac{1}{2}u_0^2\mathring{p}_{\text{sun}}(R^2 - 6) + \frac{B_0^2}{4\pi}R^2$. Now we use $\widetilde{p}_0$ instead of p_0 to define the background solution

$$\mathring{\mathbf{U}}(r) := \left(g_{\text{sun}}\mathring{\rho}_{\text{sun}}\big(r^2 - 6\big), \mathbf{0}, \mathbf{0}, \frac{\widetilde{p}_0 + \mathring{p}_{\text{sun}}\big(r^2 - 6\big)}{\gamma - 1} \right)^T .$$

With this setting we can study the Bgfix scheme in the case where the structure but not the exact values of the background solution are known. Note that the perturbation $\widetilde{\mathbf{U}}(r) = \mathbf{U}(r) - \mathring{\mathbf{U}}(r)$ is zero only in the density but large in all other components.

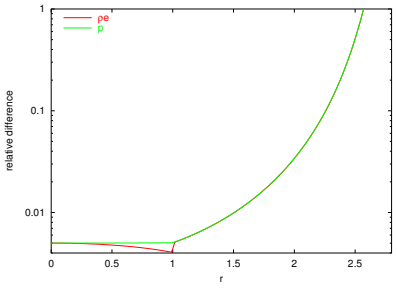


Figure 9.4: The plot shows the relative deviation of the background solution $\mathring{\mathbf{U}}$ used for the Bgfix modification from the stationary solution $\mathbf{U}$ of the atmosphere rotation problem 6.8(RotAtm). Only the density is identical; all other components differ in the whole domain. We show the values of $\frac{|\mathring{p}-p|}{p}$ and $\frac{|\mathring{\rho E}-\rho E|}{\rho E}$ as functions of the radius r.

For example, since the velocity and the magnetic field in the exact solution of Problem 6.8(RotAtm) are non–zero, the total energy density differs significantly between the background solution and the exact solution. Thus, the smallness assumption made in the derivation of the Bgfix scheme is — as in the previous example — not satisfied. The relative deviation for some components of the background solution from the exact solution is plotted in Figure 9.4. Since the energy density for large radius r is very small, the relative deviation is far larger for large r than for r small. In the velocity and the magnetic field components the difference between $\mathbf{U}$ and $\mathring{\mathbf{U}}$ is non–zero, as well.

We start our numerical experiments with a comparison of the error on a series of globally refined grids. Since the Bgfix modification requires hardly any additional cpu time, we do not compare the error to runtime ratio in this case. In Figure 9.5 the error for both the base scheme and the Bgfix scheme are plotted up to about 1.5 million elements. As is to be expected, the Bgfix modification leads to a significant improvement especially at low grid resolutions because here the approximation errors in the base scheme are more severe; at very low grid resolutions the base scheme crashes, whereas the Bgfix modification stabilizes the scheme. To facilitate the comparison of the two schemes we include an approximation of the error curves in the form Ch^α (as used in the definition of the EOC, cf. Definition 3.4). The Bgfix scheme clearly leads to a smaller constant C — this was to be expected from our analysis in Section 9.1. At the same time the convergence rate α seems to be smaller, as well. The value $\alpha \approx 1$ observed for the base scheme seems to be larger than the expected asymptotic convergence rate since the exact solution is discontinuous. The high rate of convergence observed at low grid resolutions is therefore possibly due to the reduction of the large error in the background atmosphere, which is smooth. Note that our experiments for the rotation problem with constant density presented in the previous chapter lead to a convergence rate below 0.5 (cf. Figure 8.4); a higher asymptotic convergence rate for the atmospheric rotation problem does not seem likely. At higher grid resolutions, the convergence rate for the base scheme also seems to deteriorate: the average over all values shown is close to 1, the average over the last six values is only about 0.82, and if we take only the last four values into account the average is merely 0.67. This clearly suggests that we are not yet in the assymptotic regime. In the case of the Bgfix scheme different averages all lead to about 0.5.

In Figure 9.6 the quality of the approximations for the base scheme and the Bgfix scheme are compared on a locally adapted grid. Both in the density and the velocity field the deviation from the exact solution is quite large when the base scheme is di-

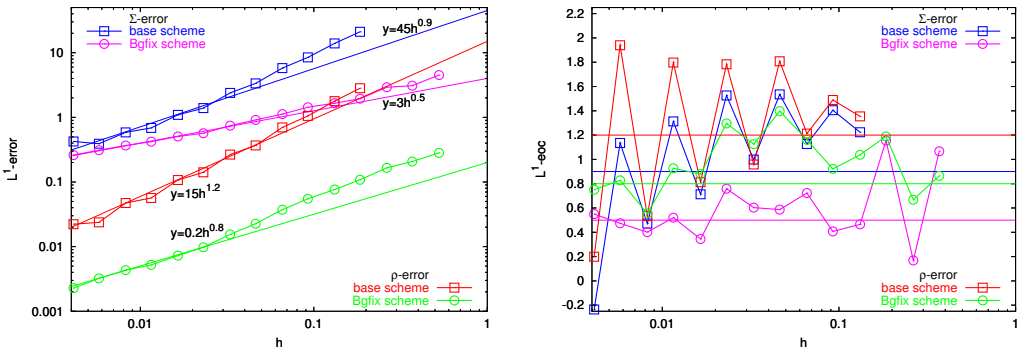


Figure 9.5: Problem: 6.8(RotAtm). Scheme: HLLEM flux with DEOMOD reconstruction in conservative variables and mixed GLM–MHD correction. The results are computed on the unstructured grid. In the left plot the error vs. the grid resolution is shown for the Bgfix scheme and the base scheme. The error in the full conservative vector $\mathbf{U}$ is shown together with the error in the density ρ. For an easy comparison an approximation of the errors in the form Ch^{α} is included. The right plot shows the corresponding experimental order of convergence.

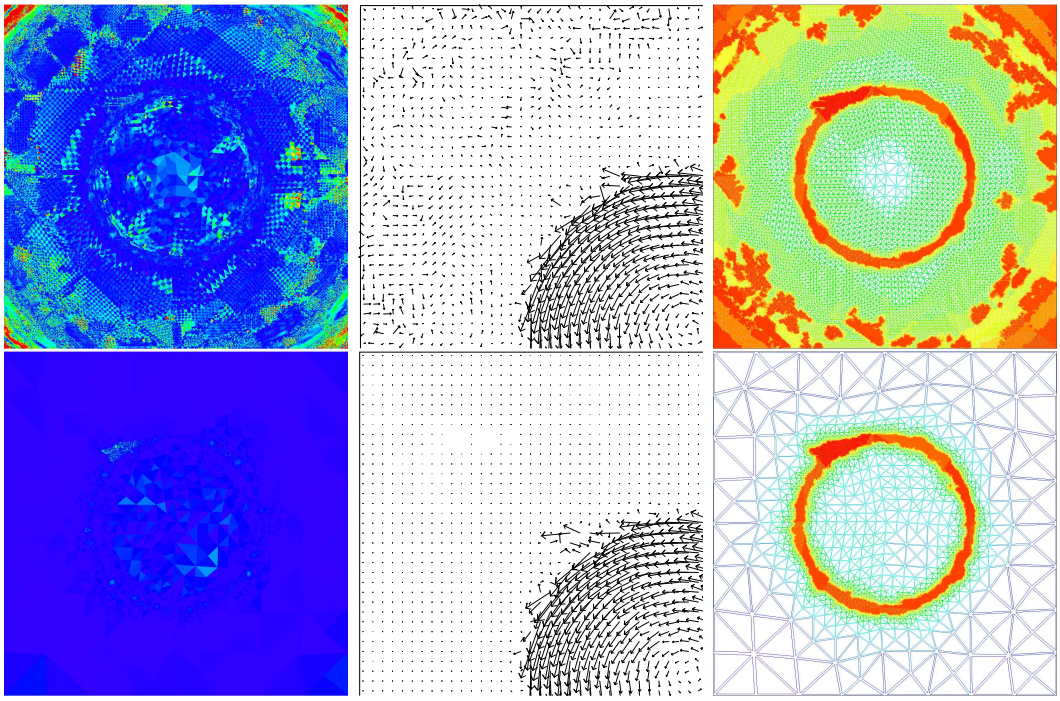
rectly applied to the rotation problem. These approximation errors are greatly reduced by our modification. Outside the ball with radius R the velocity should be zero. This is the case for the Bgfix scheme. Since the balance of the force of gravity and the pressure is not satisfied in the base scheme, errors in the velocity field are clearly visible. The variation in the density and the pressure requires a high grid resolution even in those parts of the domain where the solution is stationary and smooth. In addition the perturbation introduced by the lack of balance between the flux vector and the source term leads to additional grid refinement. Neither effect is present in the Bgfix scheme. A coarse grid can be used especially in the outlying regions, whereas all grid elements are far smaller if no modification is used. The simulation without modification leads to a grid with 253380 elements; by using the Bgfix modification only 53338 elements are required; and although the grid has fewer elements, the accuracy of the approximation is higher in the case where the modification is used. In addition to higher accuracy, the Bgfix method thus leads to a considerable gain in cpu time since only approximately 20 percent of the grid elements are required. To further quantify the difference between the Bgfix method and the base scheme, we also show scatter plots of the velocity and the magnetic field in Figure 9.6. The perturbations in the approximation using the base scheme are clearly visible for large values of r. Here the disturbances in the velocity field are nearly of the same magnitude as the velocity at the interface so that the structure of the exact solution is hardly recognizable.

Summary of Section 9.4.2: *One major disadvantage of the Bgfix scheme is that we have to prescribe a priori a background solution* $\mathring{\mathbf{U}}$ *since this has to be evaluated to compute the approximation* $\mathbf{U}_h$*. It is plausible that the scheme improves the approximation quality of the scheme in the case where the approximation* $\mathbf{U}_h$ *is almost identical to* $\mathring{\mathbf{U}}$ *— in the case where* $\mathbf{U}$ *is identical to* $\mathring{\mathbf{U}}$ *the scheme produces no approximation error. An important aspect of the scheme is its performance in the case where* $\mathring{\mathbf{U}}$ *has the same structure as* $\mathbf{U}$ *but is, for example, shifted by some constant or includes only*

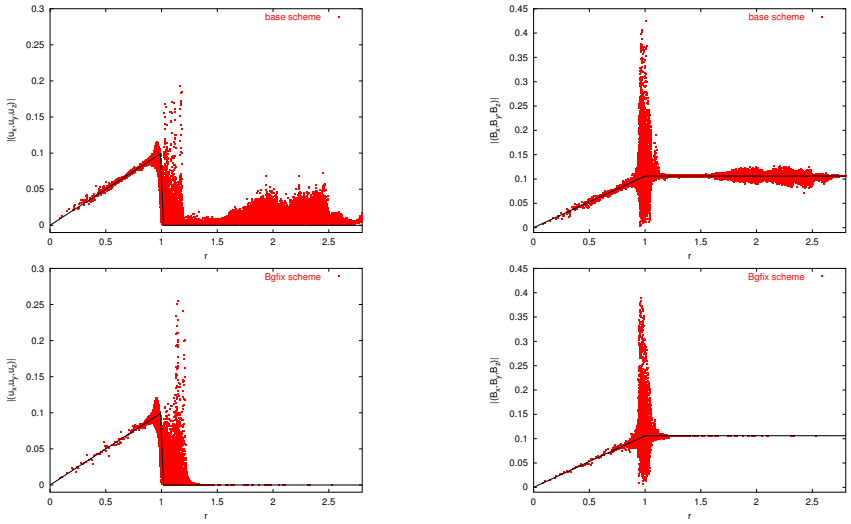
some part of the important structure; we studied this case in this section.
*Our results show that the Bgfix modification leads to a significant reduction of the approximation error. The reduction is largest on coarse grids and diminishes with increasing grid resolution. However, for all grid resolutions on which we tested the Bgfix scheme the reduction in the error was still significant. (Note that the finest grid had 1.5 million elements.) On these grids the base scheme showed a higher order of convergence, higher even than in the case of the far simpler constant rotation problem (Problem 6.7(*RotConst*)). This last observation suggests that we have not yet reached the asymptotic regime. An increase in the grid resolution will probably led to a reduced order of convergence so that the base scheme will eventually converge with the same order as the Bgfix scheme.*
Furthermore we demonstrated that on a locally adapted grid the Bgfix scheme leads to a significant reduction in the computational cost since a far coarser grid is generated, while at the same time the quality of the approximation is increased.

9.4.3 Advection Problem

We continue our study of the Bgfix scheme with the advection problem 6.9(AdvAtm). In contrast to the rotation problem studied above the solution to this problem is not stationary. Thus dynamic local grid adaption and coarsening lead to an improved performance of the scheme, and the advantages of the Bgfix scheme sketched in Section 9.3 become apparent. We use the density and the pressure from the background atmosphere to construct the Bgfix scheme. Therefore the perturbation $\widetilde{\mathbf{U}}$ is zero in the density and non–zero in the momentum and the magnetic field. We also have additional perturbations in the energy density due to the balance of the total pressure in those regions of the domain where $B_z \neq 0$. Note that since $\gamma = 2$ the density in the background atmosphere given by Problem 6.1(Atm) is a linear function. Our choice of the reconstruction guarantees that this linear function is reproduced (cf. [DRW02a]). Therefore the initial error in the density is zero not only in the Bgfix scheme but also in the base scheme. In Figure 9.7 we plot the error versus the grid resolution as in Figure 9.5. In this case we do not plot the global error, rather the error in the density, which is identical to the background atmosphere, and the error in the magnetic field component B_z, which is the advected quantity, is shown. The total error is dominated by the error in B_z. Since $\mathbf{B}$ and also $\mathbf{u}$ are zero in the background solution $\mathring{\mathbf{U}}$, the Bgfix scheme does not lead to an improvement for the approximation of B_z. Although initially the error in the density is zero in both the base and the Bgfix scheme, the errors at the end of the simulation differ by an order of magnitude. In contrast to the results shown in Figure 9.5 we do not observe a reduction in the convergence rate due to the Bgfix modification in Figure 9.7.
We now turn to the case of locally adapted grids. The density and the pressure are identical to the background atmosphere, and the velocity and the magnetic field are constant in those regions of the domain where $B_z = 0$. Consequently when we use the Bgfix modification, the grid has to be refined only in the regions where $B_z \neq 0$. For the base scheme a high grid resolution is required in the whole domain to allow for sufficient accuracy of the approximation of the background atmosphere. In Figure 9.9 we show the grids produced using the base scheme and the Bgfix scheme with the parameters



(a) in the top row we show results using the base scheme and in the bottom row using the Bgfix modification. On the left the relative deviation from the exact solution in the density is shown using identical scaling for both approximations. The middle column shows the velocity vector field for $x < 0, y > 0$, and on the right the locally refined grid is shown. The triangles are colored according to their size with identical scaling in both cases.



(b) scatter plots of the velocity field (left) and the magnetic field (right) together with the exact solution. The bottom row is with the Bgfix correction, the row above without correction.

Figure 9.6: Problem: 6.8(RotAtm). Scheme: HLLEM flux with DEOMOD reconstruction in conservative variables and mixed GLM–MHD correction on the unstructured grid using local grid adaption.

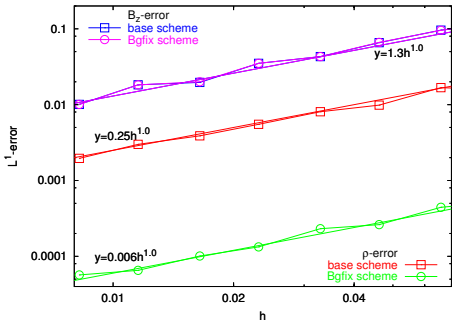
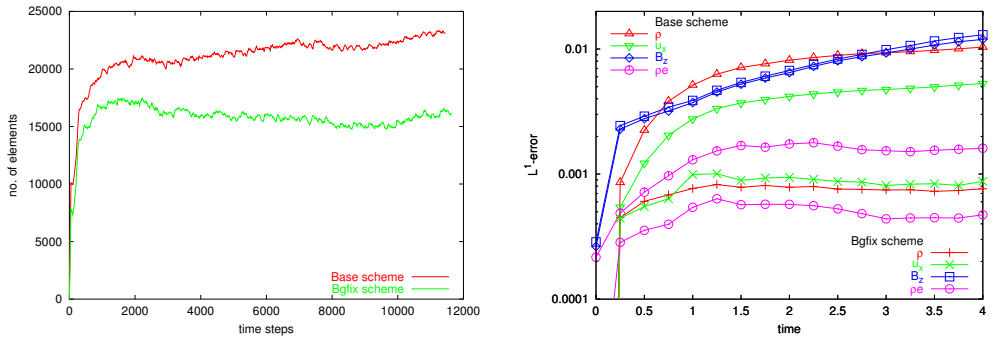


Figure 9.7: Problem: 6.9(AdvAtm). Scheme: DW flux with DEOMOD reconstruction in primitive variables and mixed GLM–MHD correction. The results are computed on the unstructured grid. The error vs. the grid resolution is shown for the Bgfix scheme and the base scheme for a series of uniformly refined grids. We plot the error in ρ and in B_z together with an approximation of the errors in the form Ch^α.

for the grid adaption chosen as specified at the end of Section 6.3. At the beginning of the simulation the grids generated by the adaption procedure have 10099 and 7811 elements for the base scheme and the Bgfix scheme, respectively. The time evolution of the number of elements for both schemes is shown in Figure 9.8(a). The time evolution of the error for different quantities is shown in Figure 9.8(b). The projection of the initial data onto the grid is (after the reconstruction process) identical to the initial data in the density and also in the momentum. Consequently the errors in ρ and u_x are in the range of the roundoff error. Since the background pressure is a quadratic function, the error in the energy is approximately 10^{-4} for the base scheme and only 10^{-6} for the Bgfix scheme. The main approximation error is in B_z where the difference between the base scheme and the Bgfix scheme is negligible. But note that in the base scheme the error in the density is of the same order as the error in B_z and the error in u_x is also similar. Due to the Bgfix modification, the errors in these components are reduced by an order of magnitude.

Since the force of gravity points downwards, an approximation error in the balance between the pressure gradient and the gravity source term leads to an error in the vertical velocity. In Figure 9.10 we plot the values of u_y at $t = 4$ for both the base and the Bgfix scheme. In the region where B_z does not vanish and where therefore the horizontal advection is relevant, both schemes produce perturbations in the vertical velocity that are of a similar magnitude. In the top and the bottom regions of the domain where $B_z = 0$, the base scheme leads to perturbations that are of the same magnitude as the perturbations in the middle of the domain. The Bgfix scheme, on the other hand, shows very few perturbations in the top and bottom parts of the domain.

The results shown so far demonstrate only to what extend the Bgfix scheme leads to an improvement in the efficiency of the scheme, due to the reduction of the error in some components of the state vector and due to the reduction of the grid size. In none of the tests so far have we taken into account the additional computational cost of the Bgfix modification. In Figure 9.11 we, therefore, study the error versus the execution time. Together with the results on a series of globally refined grids we also show results computed on two series of locally refined grids; these were generated dynamically during the simulation using the strategy from Section 3.5 with values for $h_{\min}$ in the interval $[0.002, 0.5]$. The "high resolution" results were obtained using $\text{crs}_{\text{limit}} = 0.005$ and $\text{ref}_{\text{limit}} = 0.025$; the other results were computed with our standard values $\text{crs}_{\text{limit}} = 0.05$



(a) time evolution of grid size **(b)** time evolution of error

Figure 9.8: Problem: 6.9(AdvAtm). Scheme: DW flux with DEOMOD reconstruction in primitive variables and mixed GLM–MHD correction. The results are computed on the unstructured grid.

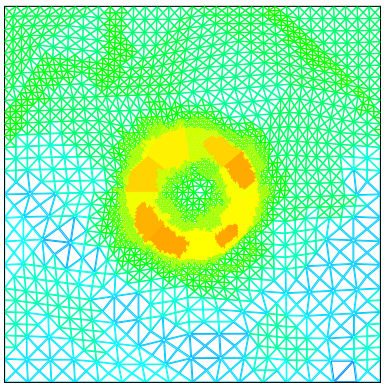
and $\text{ref}_{\text{limit}} = 0.01$. Since the solution is smooth, the grid resolution is not increased for small $h_{\min}$. Consequently, we see no reduction in the error. With the time step Δt computed from the local time steps of the smallest elements, the cpu time increases since a few elements are always refined down to $h_{\min}$. This is the reason why we have included results using different refinement and coarsening limits.

On the globally refined grids the Bgfix scheme leads to an increase in the execution time of approximately eight percent. But due to the reduction in the error the Bgfix scheme is still slightly more efficient. The advantage of the Bgfix method becomes obvious on the locally refined grids. In the computations using our standard values for the coarsening and refinement indicators the reduction in cpu time is always more than 75 percent; for the "high resolution" results the gain is even more than 85 percent.

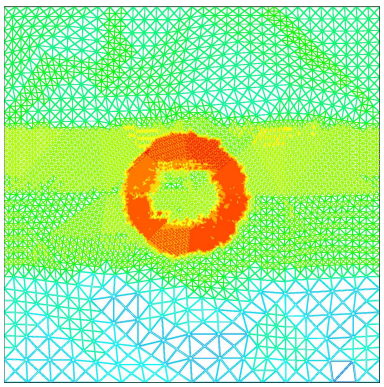
Summary of Section 9.4.3: *The test case studied in this section is close to possible applications because it describes a moving structure in a background atmosphere. The major difference to the solar physical applications is that the structure moves with a constant speed through the atmosphere and remains unchanged; furthermore, the direction of motion is not against the force of gravity* $\mathbf{g}$ *but orthogonal to* $\mathbf{g}$*.*

On the uniformly refined grid the Bgfix method leads to a reduction in the error in the density ρ*, the profile of which is used in the background solution* $\mathring{\mathbf{U}}$*. In the advected magnetic field component* B_z*, which is zero in* $\mathring{\mathbf{U}}$*, the Bgfix scheme leads to the same error as the base scheme. In this case an error in the balance between the force the gravity and the pressure gradient does not seem to influence the structure in* B_z *significantly; possibly this is due to the fact that the error caused by a shift in the atmosphere is small compared to the error caused by the numerical viscosity during the advection process.*

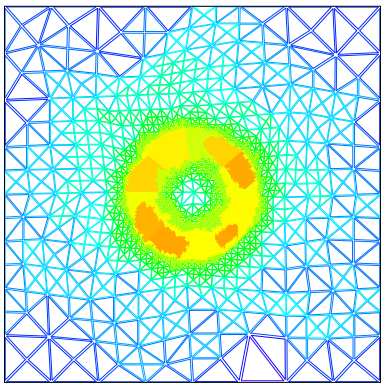
The same calculation on a series of locally refined grids demonstrates the advantages of the Bgfix approach. Since only the region where $B_z \neq 0$ *is refined, the computational cost of the Bgfix method is considerably smaller compared to the base scheme; therefore, a fixed error is reached at a far smaller computational cost. The adaptation strategy works so well together with the Bgfix scheme that the final grid is very similar to the*



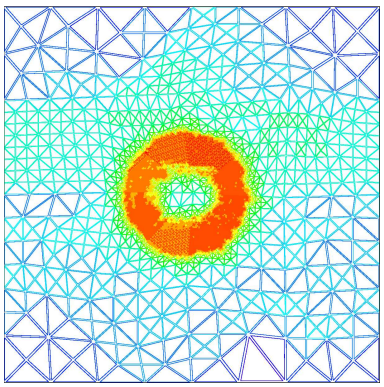
(a) initial grid using base scheme



(b) grid at $t = 4$ using base scheme



(c) initial grid using Bgfix scheme



(d) grid at $t = 4$ using Bgfix scheme

Figure 9.9: Problem: 6.9(AdvAtm). Scheme: DW flux with DEOMOD reconstruction in primitive variables and mixed GLM–MHD correction. The results are computed on the unstructured grid.

grid produced for the initial data. (Note that the solution at the final time is identical to the initial data.) The grid produced by the base scheme at the final time T is refined everywhere where the advected structure in B_z passed through for $t < T$; due to perturbations the grid was not completely coarsened after the structure in B_z had moved through.

9.4.4 Smoothness of Background Solution

As we saw in Section 9.1 we can only expect an improvement in the approximation when using the Bgfix scheme if the initial perturbation of the background solution is small. Our numerical results show that even for large perturbations the Bgfix scheme can lead to a reduction in the approximation error. For the motivation of the Bgfix modification we made the additional assumption that the background solution is smooth (cf. Section 9.1). To construct the Bgfix scheme as shown in Section 9.2 we need far less regularity of the background solution $\mathring{\mathbf{U}}$ than the C^2 regularity assumed for the

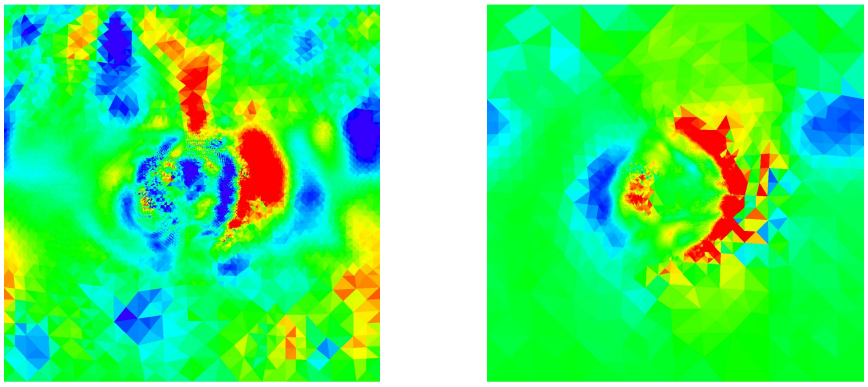


Figure 9.10: Problem: 6.9(AdvAtm). Scheme: DW flux with DEOMOD reconstruction in primitive variables and mixed GLM–MHD correction. The results are computed on the unstructured grid. We plot the perturbation in the vertical velocity u_y. The green color indicates regions where the vertical velocity is close to zero, which is the correct value. Red indicates upward motion and blue downward motion.

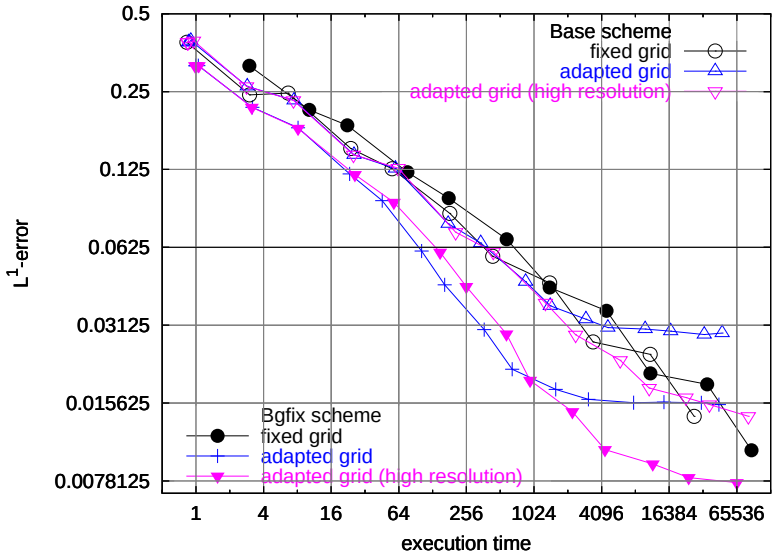


Figure 9.11: Problem: 6.9(AdvAtm). Scheme: DW flux with DEOMOD reconstruction in primitive variables and mixed GLM–MHD correction. The results are computed on the unstructured grid. The plot shows the error vs. the execution time on a series of globally refined grids and on two series of locally refined grids using different values for the refinement parameters $\text{crs}_{\text{limit}}$ and $\text{ref}_{\text{limit}}$.

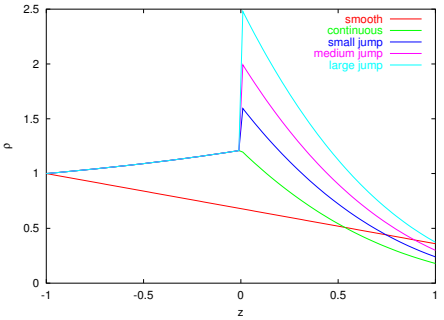
analytical results. On the one hand, we have to compute cell averages so that $\mathring{\mathbf{U}}$ at least has to be in L^1 locally. Furthermore, we have to evaluate the integral of $\mathbf{F}(\mathring{\mathbf{U}}(\cdot,t))$ on the interfaces for fixed t. In practical applications the evaluation of these integrals requires the use of a quadrature rule (cf. (9.17) and (9.18)) so that $\mathring{\mathbf{U}}$ and $\mathbf{F}(\mathring{\mathbf{U}})$ and also $\mathbf{q}(\mathring{\mathbf{U}})$ should be defined pointwise on the whole computational domain. In the case of atmospheric flow this is satisfied if the density $\mathring{\rho}$ and the pressure $\mathring{p}$ in the background atmosphere can be evaluated pointwise since $\mathbf{u}$ and $\mathbf{B}$ are zero in the background atmosphere.
To study the influence of the smoothness of the background solution on the Bgfix scheme we return to the advection problem (Problem 6.9(AdvAtm)) studied in the previous section. In the following we modify the density $\mathring{\rho}$ (and consequently also the pressure $\mathring{p}$) in the background atmosphere using a parameter $\alpha \in \mathbb{R}$:

$$\begin{aligned} \mathring{\rho}(z) &= \begin{cases} (1-0.32(z+1))^{-0.5} & z > 0\ , \\ 8\alpha(1-0.32(z+1))^{3} & z \le 0\ , \end{cases} \\ \mathring{p}(z) &= \begin{cases} (1-0.32(z+1))^{0.5} & z > 0\ , \\ \alpha(1-0.32(z+1))^{4} + \bar{p} & z \le 0\ . \end{cases} \end{aligned} \tag{9.21}$$

The constant $\bar{p}$ is chosen in such a way that the pressure is continuous at $z = 0$; consequently the flux $\mathbf{F}(\mathring{\mathbf{U}})$ is continuous. With a constant gravity source term $\mathbf{g} = (0, -0.16, 0)$ we have $\mathring{p}' = \mathbf{g}\mathring{\rho}$ as required for a static background atmosphere. The density $\mathring{\rho}$ can be discontinuous at $z = 0$ depending on our choice for α. Let $\bar{\alpha}$ denote the value of α for which the density is continuous, i.e.

$$\bar{\alpha} := \frac{1}{8}(1-0.32)^{-3.5} \approx 0.48209\ . \tag{9.22}$$

Using this constant we define in addition to the original setting (Problem 6.9(AdvAtm)) four more settings for our advection problem leading to a continuous density profile and to three density profiles with jumps of varying magnitude. The corresponding choices for α and a plot of the density profiles are shown in Figure 9.12.



smooth: advection problem with model solar atmosphere,
continuous: $\alpha = \bar{\alpha}$,
small jump: $\alpha = \frac{4}{3}\bar{\alpha}$,
medium jump: $\alpha = \frac{5}{3}\bar{\alpha}$,
large jump: $\alpha = 1$.

Figure 9.12: Five different profiles for the background atmosphere used in the advection problem 6.9(AdvAtm). Together with the smooth setting studied in the previous section, four additional density profiles are shown using the definition (9.21) with $\bar{\alpha}$ given by (9.22).

In Figure 9.13 we combine the error to runtime ratios for the smooth density profile (using our standard adaption indicators), for the continuous setting, and for the setting

with the small jump in the density. As in the smooth case studied in the previous section the advantage of the Bgfix scheme is also clearly visible in the case where the background solution $\mathring{\mathbf{U}}$ is less smooth. The gain in the runtime required to reach a fixed error even increases with the decreasing smoothness of the density. To reach an error of 0.06 the Bgfix scheme requires in the case of the smooth background solution only about one–fourth of the time and only about one–eighth in the case of the small discontinuity. Furthermore, the increase in the error on a fixed grid caused by the decrease in the regularity of the solution is less pronounced when the Bgfix modification is used.
We conclude our studies with the results for the three discontinuous settings described in Figure 9.12. The results are presented in Figure 9.14 and confirm the observations made so far. Again the Bgfix scheme leads to a substantial reduction in cpu time compared to the unmodified scheme. Only in the case of the large jump did we observe difficulties with the stability of the scheme. At refinement level 10 the scheme broke down whereas the base scheme reached level 15 without any difficulties. On the other hand, the approximation error of the Bgfix scheme at level 9 is smaller than the approximation error of the base scheme at level 15. The problems with the unstable behavior of the scheme is probably due to the fact that the perturbations are not small. If we include the constant velocity u_x into the background solution (so that $\mathring{\mathbf{U}}$ differs from the exact solution only in those regions where $B_z \neq 0$), then the scheme remains stable and the error is substantially reduced (cf. Figure 9.14).
Table 9.1 quantifies the reduction in both the error and the grid size when the Bgfix modification is used. Independent of the smoothness of the background solution, we observe a reduction in the grid size of about 40 percent, and, at the same time, the error is reduced by about 50 per cent. This demonstrates the high efficiency of the Bgfix modification even in the case of non–smooth background solutions. In Figure 9.15 we show a scatter plot of the density at time $t = 4$ for four of our five settings; furthermore we have included the case where the background solution consists of the density profile with the large jump together with the constant velocity field; results using the first order scheme are also shown. Note that the advantage of the Bgfix scheme is even more apparent if a first order scheme is modified — in the first order scheme the problems with balancing the pressure gradient and the gravity source term is even more severe than in the case where the constant values are reconstructed. The problems with oscillations at the discontinuity is clearly visible in the results using the Bgfix scheme. These oscillations are not present when the base scheme is used, but, on the other hand, the base scheme leads to a stronger smearing of the interface. Away from the discontinuity the Bgfix scheme reproduces the background solution up to a high order of accuracy whereas the base scheme introduces small scale oscillation. By including u_x in the background solution we can suppress the oscillations in the Bgfix method, and the discontinuity is captured almost without smearing. Note that, even if we include u_x, we still have a deviation from the exact solution in those parts of the domain where B_z does not vanish. This region is located around $y = 0$ so that it contains the discontinuity in ρ.

Summary of Section 9.4.4: *For the derivation of the Bgfix method we had to assume a certain amount of regularity of the background solution $\mathring{\mathbf{U}}$. In our applications $\mathring{\mathbf{U}}$ is a model for the quiet solar atmosphere, and we can assume that this is a smooth function. In other application such as shallow water flow with discontinuous bottom*

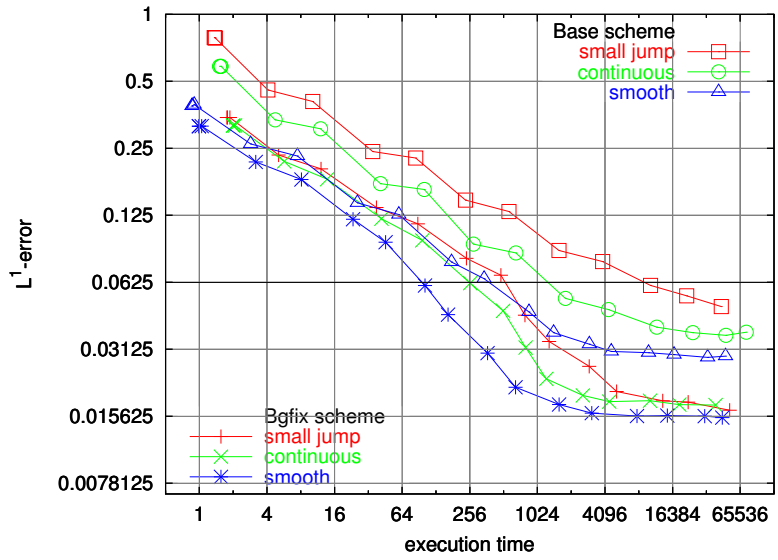


Figure 9.13: Problem: 6.9(AdvAtm) with different density profiles (cf. Figure 9.12). Scheme: DW flux with DEOMOD reconstruction in primitive variables and mixed GLM–MHD correction. The results are computed on the unstructured grid. The plot shows the error vs. the execution time on a series of locally refined grids.

topology (e.g. [BPV03]) the function $\mathring{\mathbf{U}}$ *is discontinuous. In this section we, therefore, tested the Bgfix method for different degrees of regularity for* $\mathring{\mathbf{U}}$.
Our results show that the Bgfix modification leads to an improvement in the base scheme even for large discontinuities in $\mathring{\mathbf{U}}$ *and that the improvement in the error to runtime ratio on locally adapted grids is even more obvious. The oscillatory behavior at the discontinuity was the only disadvantage of the Bgfix method we found; whereas the base scheme leads to a strong smearing of the discontinuity, the Bgfix scheme produces a sharp profile with over and undershoots. This problem can possibly be solved by a different limiting approach in the higher order scheme that takes the smoothness of the background solution into account.*

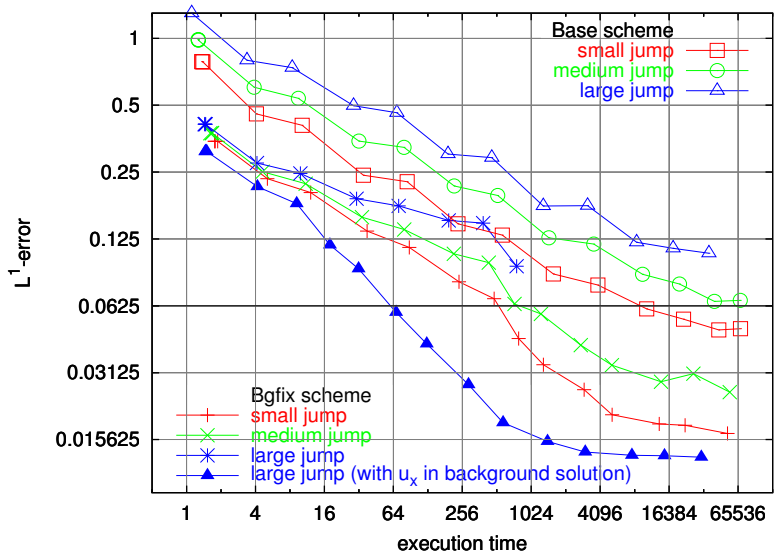
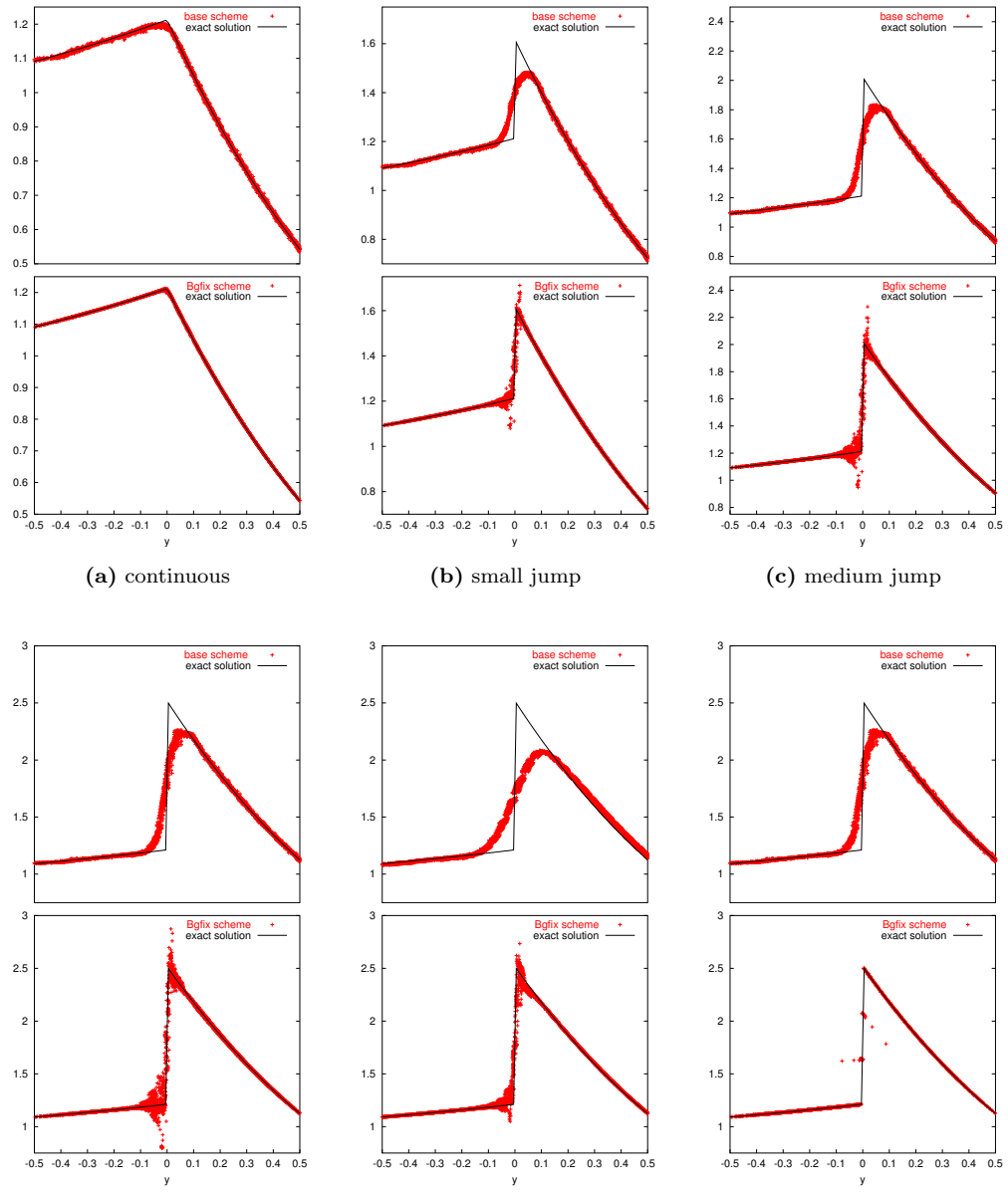


Figure 9.14: Problem: 6.9(AdvAtm) with different discontinuous density profiles (cf. Figure 9.12). Scheme: DW flux with DEOMOD reconstruction in primitive variables and mixed GLM–MHD correction. The results are computed on the unstructured grid. The plot shows the error vs. the execution time on a series of locally refined grids.

setting	base scheme		Bgfix scheme	
	grid size	L^1–error	grid size	L^1–error
		$t = 0$		
continuous	5442	2.383481e-03	3354	1.048754e-03
small jump	6204	9.845641e-03	4244	1.038314e-03
medium jump	6204	1.773566e-02	4296	1.038315e-03
large jump	6204	2.738398e-02	4296	1.038315e-03
		$t = 4$		
continuous	7992	5.293264e-02	4852	3.185965e-02
small jump	8270	8.709184e-02	5266	4.466570e-02
medium jump	8312	1.267274e-01	5242	6.375453e-02
large jump	8368	1.760125e-01	5276	9.421814e-02

Table 9.1: Problem: 6.9(AdvAtm) with different density profiles (cf. Figure 9.12). Scheme: DW flux with DEOMOD reconstruction in primitive variables and mixed GLM–MHD correction. The results are computed on the unstructured grid at level 9. The L^1–error is computed for all components of the state vector.



(a) continuous **(b)** small jump **(c)** medium jump

(d) large jump **(e)** large jump (first order) **(f)** large jump (u_x included in background solution)

Figure 9.15: Problem: 6.9(AdvAtm) with different density profiles (cf. Figure 9.12). Scheme: DW flux with DEOMOD reconstruction in primitive variables and mixed GLM–MHD correction. The results are computed on the unstructured grid on level 9. All plots show the density at $t = 4$ with varying size of the discontinuity. Note the different scaling of the vertical axis.

10. Summary

MHD Scheme

In the previous three chapters we presented extensions of our basic finite–volume scheme from Chapter 3. Each of these three chapters was devoted to one of the numerical challenges discussed in the overview Chapter 6. We have not discussed all the problems sketched there but have concentrated on three important issues; we did not study the derivation of suitable boundary conditions satisfying Definition 6.6 in detail. (A brief summary of our results from [DKSW01b] is given in Section 6.1.4.) We concentrated on modifications of the flux computation in the base scheme that allow us to solve the real gas MHD equations, that reduce stability problems due to a violation of the divergence constraint, and that produce accurate approximations for initial conditions near an equilibrium state.
In Chapter 7 we extended the energy relaxation scheme presented in [CP98] for the Euler equations of hydrodynamics to the real gas MHD equations (1.1). The main idea of this approach is to replace the complex pressure function $p(\rho, \varepsilon)$ with a simpler function $p_1(\rho, \varepsilon_1)$ and to use a relaxation framework to recover the original MHD system with the pressure law p in the equilibrium limit. Since we are interested in deriving an extension of our base scheme for the perfect gas MHD equations, the natural choice for p_1 is a perfect gas pressure law of the form $p_1(\rho, \varepsilon_1) = (\gamma_1 - 1)\rho\varepsilon_1$. This leads to a simple extension of the base scheme, which is easy to implement and which is not restricted to the use with the finite–volume approach. The authors of [CP98] derived a lower bound for the free parameter γ_1. Since a large value for γ_1 increases the amount of numerical viscosity, the optimal choice for γ_1 seems to be given by this lower bound. For our extension to the MHD system we use this lower bound to define γ_1 in each time step and for each flux computation. Our numerical results show that this choice leads to a stable and efficient discretization of the real gas MHD equations. In the following we briefly summarize the main results from the numerical tests presented in Chapter 7.

- The implementation of the ER method can be adapted to the amount of information available for the EOS. If, for example, only the pressure function itself is known then the ER method can nevertheless be used either with a constant γ_1 or by using a finite–difference approximation of equation (7.19) to compute a local γ_1.
- The analytically justified lower bound (7.19) for the parameter γ_1 leads to a stable scheme. We choose γ_1 locally in each time step and on each interface subject to

the left and right hand states for which the flux is computed. In this way we greatly increase the efficiency of the scheme since in regions where a small γ_1 is sufficient for the stability of the scheme no unnecessary numerical viscosity is added.

- The computation cost of the scheme can be considerably reduced by the use of a tabularized equation of state. Even in those cases where the pressure can be computed for arbitrary pairs ρ, ε the use of a table might be advisable, especially if the pressure is very expensive to compute. Furthermore the use of a pressure table reduces the cost of computing the derivatives of the pressure function that are required to compute γ_1 via (7.19). (For any explicit scheme the sound speed has to be computed in some way and this also requires the computation of the derivatives of p.) Our results show that by using a dynamically generated and locally adapted Cartesian table the approximation quality of the scheme is unaffected, whereas the efficiency of the scheme is considerably increased.

- We studied the efficiency of the ER extension of the MHD–HLLEM flux and of the local Lax–Friedrichs scheme applied directly to the real gas MHD equations. Although the ER scheme requires more time on a fixed grid, the high amount of numerical viscosity introduced by the LF approach means that the second order LF scheme can be even less efficient than the first order ER scheme — at least up to a very high grid resolution. The results of our comparison are comparable to the ones published in [Wes02b], where the MHD–HLLEM scheme was first presented and compared with the LF scheme in the case of the perfect gas MHD equations. This indicates that the ER extension of a numerical flux function $\mathbf{g}_{ij}$ does not reduce the efficiency of the flux function. This observation is confirmed by our comparison in [DW01] of the direct extension of perfect gas flux functions with their ER extension; both methods where found to lead to identical results.

Although these observations were derived by studying 1d problems, they also hold true for problems in higher space dimension. This can be attributed to the fact that the 1d flux function is the central building block of our higher dimensional schemes. Therefore all convergence tests using the ER scheme in 2d led to results similar to the corresponding 1d test (cf. Chapter 8).
In Chapter 8 we presented a general way in which the divergence constraint (1.1e) can be coupled to the induction equation (1.1c) using an auxiliary function ψ and a general linear operator $\mathcal{D}$. From this modified system, called the GLM–MHD equations, we derived a number of simple extensions of our base scheme: the *elliptic*, the *parabolic*, the *hyperbolic*, and the *Galilean invariant* approach. In our numerical tests we studied an extension of the hyperbolic, and of the Galilean invariant approach where an additional damping term in the equation for the auxiliary function ψ is added. In the case of the Galilean invariant approach we have not modified the terminology; we termed the hyperbolic correction augmented by this additional damping term the *mixed* correction. Together with the *source term fix* from [Pow94] we thus compared five possible modifications of the base scheme suitable for reducing divergence errors in the magnetic field $\mathbf{B}$. To justify the GLM–MHD system we studied the long time behavior of the solution to a model problem using the different choices for the operator $\mathcal{D}$. With the

exception of the elliptic approach (which corresponds to the well known Hodge projection) the influence of all the GLM–MHD methods on the solution can be controlled by parameters. Since we are interested in deriving a general purpose solver for the MHD equations that does not require a large amount of parameter tuning, we used the insights obtained by the analytical study of the model problem and performed simple numerical tests to determine general rules of how these parameters should be chosen. One important side condition for our choice of the parameters was that the scheme should remain stable with the time step Δt used in the base scheme. By following this guideline we arrived at numerical schemes, which lead to practically no increase in computation cost. Consequently, we could measure the increase in efficiency of our new schemes by comparing the errors for fixed grid resolutions. Let us now summarize the major results of the numerical tests presented in Chapter 8.

- With the exception of the elliptic and the parabolic approach all methods can be easily implemented by a simple extension of the numerical flux function in the base scheme. The parabolic approach requires an additional reconstruction step. This can be done with little extra cost; only in the case of a distributed memory parallelization does this step lead to a considerable amount of extra communication. For the elliptic approach the solution to a Laplace equation has to be computed. This leads to a considerable increase in the computational cost compared to the base scheme and leads to additional problems if distributed memory parallelization is used.

- In the case where the base scheme was able to compute a solution without breaking down, the error in the conservative variables was often comparable to that obtained with the modifications. Thus we conclude that the correction mechanisms do not introduce a significantly higher amount of numerical viscosity. The advantage of the correction methods becomes noticeable at the point where the simulation using the base scheme breaks down, for example, due to negative pressure values; all correction schemes stabilize the base scheme considerably. The differences between the correction schemes can also be seen most clearly if one compares their stability. Here the Galilean invariant and the mixed approach are clearly superior to the other schemes.

- The problem of spurious oscillations caused by a violation of the divergence constraint is more severe for smaller grid sizes. Thus, the stability problems of the schemes are increased by an increase in the grid resolution. Consequently, a divergence fix is especially important for simulations using locally adapted grids.

- In the Galilean invariant approach and for the source term fix expressions that are not in divergence form are added to the original conservation laws. Thus in the Galilean invariant approach the magnetic field $\mathbf{B}$ and in the source term fix $\rho\mathbf{u}$, $\mathbf{B}$, and ρe no longer satisfy conservation laws. This loss of conservation can lead to problems, which considerably reduce the quality of the approximation. We observed a low convergence rate of the source term fix both for the rotation Problem with $\beta = 1$ as well as for a 1d Riemann problem. Especially in the second case it seems possible that the scheme is converging to a function with wrong intermediate states.

All the results presented in Chapter 8 let us favor the mixed GLM–MHD scheme: the additional cost of the scheme is very small; together with the Galilean invariant approach it was clearly the most stable; it does not suffer from loss of conservation as does the Galilean invariant approach.
In Chapter 9 we presented a scheme that allows an efficient approximation of a solution near an equilibrium state. It is designed for problems where an equilibrium solution $\mathring{\mathbf{U}}$ is known that can be separated from the solution $\mathbf{U}$. In our applications $\mathring{\mathbf{U}}$ described the stratified, static background atmosphere in the solar convection zone. The initial conditions differ from $\mathring{\mathbf{U}}$ only in a region that is small compared to the whole computational domain. Our method, termed *Bgfix* scheme, can be applied to any system of balance laws. The method can be viewed as a finite–volume scheme for a modification of the original system of balance laws where the flux functions and the source terms explicitly depend on space and time. Using this interpretation of the scheme we could prove the convergence of the Bgfix scheme for scalar balance laws using the general convergence theorem from Section 4.4. However, our analysis does not indicate under which conditions we can expect a reduction in the approximation error compared to the base scheme. Our numerical tests show:

- The Bgfix scheme improves the approximation quality of the base scheme when the solution $\mathbf{U}$ is close to the background solution $\mathring{\mathbf{U}}$; if the difference between the two is large, the Bgfix scheme does not reduce the quality of the base scheme.

- Even in the case where only the structure of the background solution is known but not the exact values, we found that the Bgfix modification leads to a reduction in the approximation error.

- The Bgfix scheme leads to an improvement even in the case where the background solution does not meet the smoothness assumptions required for the derivation and the analysis of the Bgfix scheme. Even for discontinuous $\mathring{\mathbf{U}}$ we found a significant reduction in the approximation error due to the Bgfix modification.

- The advantage of the Bgfix scheme is most obvious in its interplay with local grid adaptation. On locally adapted grids the Bgfix method shows a far better error to runtime ratio than the base scheme. This is due to the fact that in the regions of the domain where the approximation $\mathbf{U}_h$ is identical or at least close to the background solution $\mathring{\mathbf{U}}$, a very coarse grid is sufficient. In our applications — where the perturbation of the background atmosphere has compact support — we can thus use a very coarse grid in large parts of the domain. This also allows an increase in the size of the computational domain (for example, to reduce effects from unphysical boundary conditions) without a substantial increase in computational cost.

We conclude our study of the finite–volume scheme for the MHD system (1.1) with a test case from solar physics that models the rise of a magnetic fluxtube through a stratified atmosphere. The initial setting can be found in [DRW99]. In Figure 10.1 we plot a time sequence of the simulation using both the mixed GLM–MHD and the Bgfix method. During the rise of the fluxtube its boundaries are subject to Kelvin–Helmholtz and Rayleigh–Taylor type instabilities. To reduce the influence of these

instabilities the magnetic field in the fluxtube is twisted in manner similar to the way it was in our rotation problem (Problem 6.8(RotAtm)). Due to this rotation the magnetic field components B_x and B_y, which in the untwisted setting are zero, now take on non–zero values, so that the divergence constraint cannot be ignored. Due to the surrounding atmosphere the balance of the pressure gradient and the gravitational force also have to be taken into account. Thus this problem is both a test case for the GLM–MHD method as well as for the Bgfix scheme.
The shape and position of the fluxtube can be most easily observed in B_z, which is initially zero outside of the fluxtube and equal to 2 inside. In Figure 10.3 we plot B_z at time $t = 20$. Note that we show only a small section of the full computational domain and that we have adjusted the colorbar to cover the minimum and maximum values of B_z for this time step (cf. Figure 10.1). We show results for the base scheme and for the mixed GLM–MHD scheme, which we hold to be the most efficient of the correction approaches tested. Since the source term fix and the Hodge projection (elliptic approach) are widely used, we have also include these in our test. For this problem we have no exact solution so that we have to compare the results in a qualitative manner. In Figure 10.2 we plot the time evolution of the divergence errors; note that in this case, due to the projection of the magnetic field onto the grid, the discrete divergence of $\mathbf{B}$ is even initially not zero. Especially in the maximum norm we observe a high error at the beginning of the simulation, which is quickly reduced by all the schemes; this is most likely due to a smearing of the interface. The mixed GLM–MHD correction leads to a far smaller value of $\nabla \cdot \mathbf{B}$ in both the L^∞– and the L^1–norm. This low value is maintained over the whole simulation. All the other correction methods lead to a similar error in $\nabla \cdot \mathbf{B}$, which is barely smaller than the error in the base scheme. In the L^1–norm the error increases for all the schemes with the exception of the mixed GLM–MHD scheme; the L^1–error of the mixed GLM–MHD scheme is, furthermore, an order of magnitude smaller than the errors from all the other schemes (note that y-axis in the plot of the L^1–error is scaled logarithmically).
If we now look at the structure of the fluxtube (cf. Figure 10.3) we see that, on the one hand, its position is lower for the mixed GLM–MHD method than for the other schemes. Both the elliptic and the source term fix lead to a slightly higher position compared to the base scheme. Since we do not know the exact position of the fluxtube, it is hard to determine which is the correct position. On the other hand, we also see that especially the source term fix leads to a break in symmetry. Only the mixed approach produces a solution with intact axial symmetry; we already observed the good preservation of symmetry for the rotation problem (cf. Figure 8.6).
We conclude our discussion of the MHD schemes by demonstrating the influence of the Bgfix scheme in the case of the fluxtube simulation. In Figure 10.4 we show results using the mixed GLM–MHD scheme with and without the Bgfix correction mechanism. We see a loss of symmetry when the Bgfix scheme is not used, and we observe that the Bgfix scheme requires more than twice the number of elements.

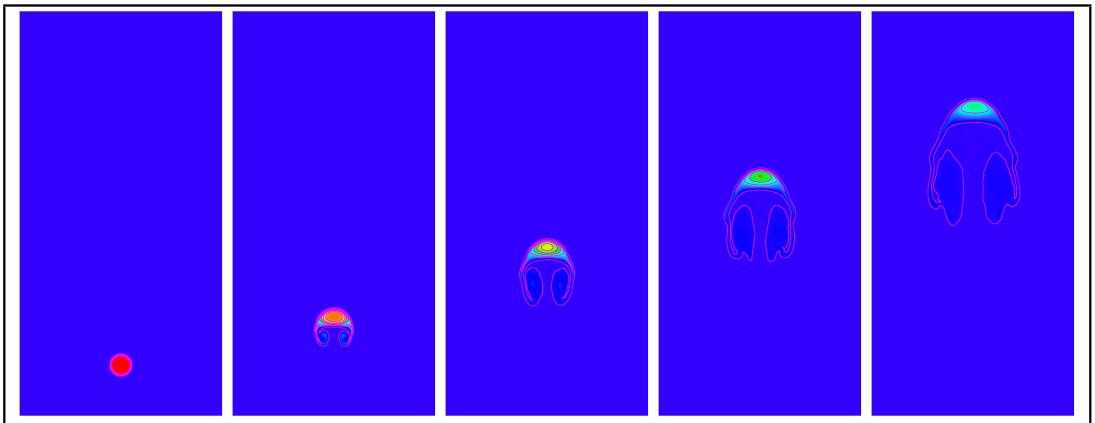
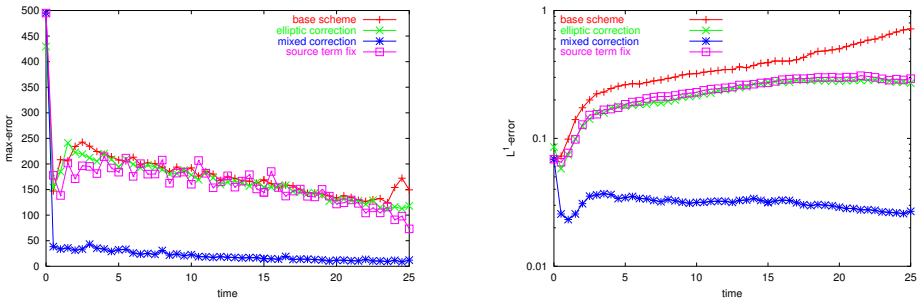
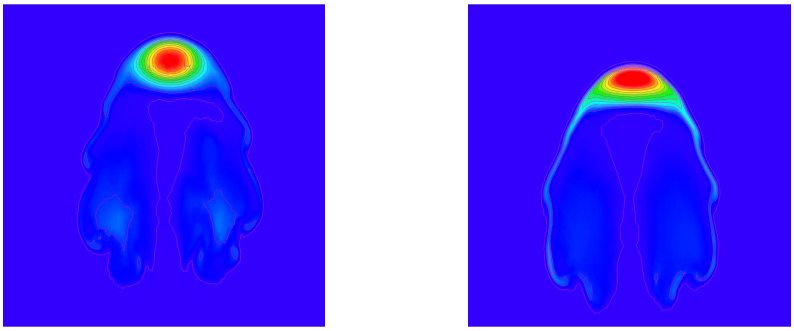


Figure 10.1: Time evolution of a twisted magnetic fluxtube in 2d using the Bgfix scheme with the GLM–MHD correction of the DW flux. From left to right: $t = 0, 5, 10, 15, 20$. The results were computed on the locally adapted unstructured grid using $\bar{u} = 1$ and $\bar{B} = 1$ in (3.19). The remaining constant used in the adaptation process are chosen as described at the end of Section 6.3.



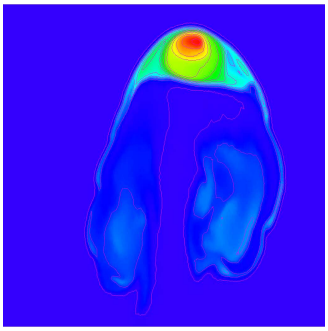
(a) L^∞–norm of $\nabla \cdot \mathbf{B}$ **(b)** L^1–norm of $\nabla \cdot \mathbf{B}$

Figure 10.2: Time evolution of the divergence error measured in the L^∞– and in the L^1–norm for different divergence correction schemes.

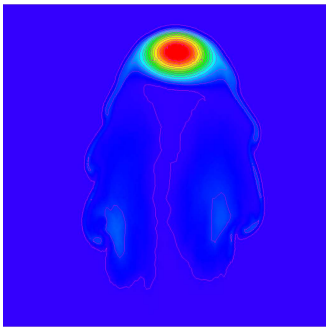


(a) base scheme

(b) mixed GLM–MHD scheme

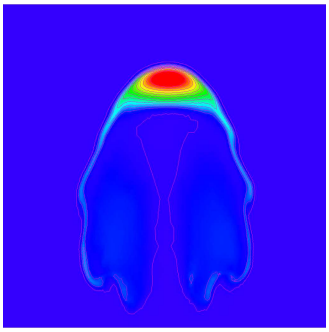


(c) source term fix

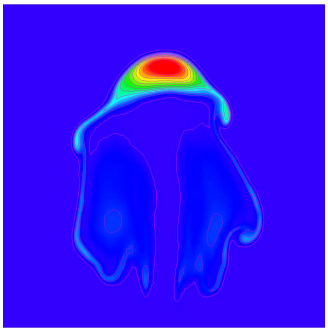


(d) Hodge projection (or elliptic GLM–MHD scheme)

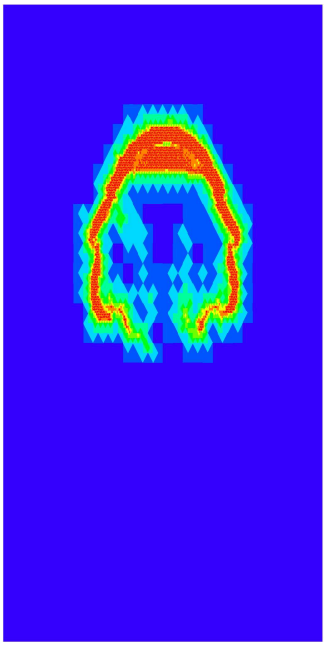
Figure 10.3: We compare the results of the simulation of a magnetic fluxtube in 2d using different correction techniques. In every case we used the Bgfix scheme with the DW numerical flux function on a locally adapted grid.



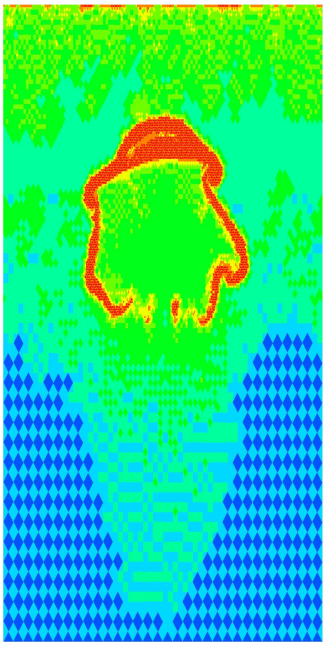
(a) B_z with Bgfix scheme



(b) B_z without Bgfix scheme



(c) refinement levels with Bgfix scheme



(d) refinement levels without Bgfix scheme

Figure 10.4: In this plot we show results using the mixed GLM–MHD scheme with (left) and without (right) the Bgfix correction. Together with the B_z component (top) we also show a color representation of the refinement levels (bottom). Red indicates a very high grid resolution and blue a very coarse grid. The grid produced with the Bgfix correction consists of about 12000 elements, without the correction a grid with 32000 elements is generated.

11. Overview

Radiation Transport Scheme

As discussed in Section 3.1 the computation of the radiation source term Q_{rad} in the energy equation (1.1d)

$$\partial_t(\rho e) + \nabla \cdot (\rho e \mathbf{u} + \mathcal{P}\mathbf{u}) = Q_{\text{rad}} \,, \tag{11.1a}$$

$$Q_{\text{rad}} = \int_{S^2} \chi\,(I - B)\, d\boldsymbol{\mu} \tag{11.1b}$$

can be reduced to solving the radiation transport (RT) equation

$$\boldsymbol{\mu} \cdot \nabla I + \chi I = \chi B \tag{11.1c}$$

for a fixed and finite set of directions $\boldsymbol{\mu} \in S^2$. The absorption coefficient χ and source term B are given functions of the fluid's density ρ and temperature θ. On the inflow boundary of the computational domain Ω the intensity is prescribed

$$I = g \quad \text{on } \partial\Omega_- := \{\mathbf{x} \in \partial\Omega_- : \boldsymbol{\mu} \cdot \mathbf{n}(\mathbf{x}) < 0\} \,. \tag{11.1d}$$

In the following two chapters we study the necessary steps for discretizing the radiation source term Q_{rad}. Our major interest lies in the construction of a scheme that can be easily added to our finite–volume code. Therefore we concentrate on first and second order methods for approximating the radiation source term. Our main emphasis will be the derivation and study of a new class of schemes that allow an efficient discretizing of the radiation transport equation (11.1c) (cf. Chapter 12). In Chapter 13 we then study the discretization of the radiation source term Q_{rad}.

11.1 Numerical Challenges

11.1.1 Non–Local Effects

As discussed in Section 1.2 we assume an instantaneous radiation equilibrium. This is a reasonable assumption since the fluid velocity is very small compared with the propagation velocity of the radiation. We can thus use an explicit discretization of the full system (1.19). The drawback is that, due to the instantaneous radiation equilibrium, the local domain of dependence is lost. This is a major difference between the MHD

system with radiation and without. The influence of this global dependency on the solution to a model problem was discussed in Chapter 4. For a numerical scheme this dependency leads to demands on the underlying grid structure and the approximation, which differ from the purely hydrodynamic case. In Section 3.6 we already sketched some of the consequences for the parallelization strategy. Further aspects are discussed in the following chapters.

11.1.2 Computational Cost

By using the discrete ordinate method (cf. Section 3.1), we reduced the problem of computing the radiation source term Q_{rad} to solving the radiation transport equation (11.1c) for a fixed set of directions $\{\boldsymbol{\mu}_m\}_{m=1}^{M}$ with some $M \geq 1$. Since the radiation intensity depends upon the density and the temperature of the fluid, this procedure is repeated every time step. Even a simple quadrature of the unit sphere S^2 leads to $M = 12$ in 2d simulations and to $M = 24$ in 3d simulations (cf. Table 3.1). Consequently, the computational cost for the approximation of Q_{rad} is far greater than for the flux calculation and, therefore, a very fast algorithm is required. This problem becomes even more severe if the dependence of the radiation intensity on the frequency ν is to be taken into account (cf. Section 1.2).
The necessity for a fast solution algorithm and, furthermore, the difficult physical regime have led us to pose the following demands on the RT solver in [DV02]:

- The RT solver should be easily implemented for 2d and 3d calculations, independent of the underlying grid structure and without extensive recoding. It should require only a small stencil for the computation of the intensity I to minimize communication in a parallel environment.
- Modern (M)HD–solvers are at least second order accurate, and the method for solving (11.1c) should be of the same order as the (M)HD solver.
- The RT solver has to be able to handle a stiff as well as a non–stiff source term $\chi(B - I)$ in (11.1c), because in the solar photosphere χ varies between 10^{-3} and 10^4.
- The RT solver must cope with large gradients in the temperature θ and the density ρ. In some regions of the solar photosphere, χ is proportional to θ^{10} and the source function B varies with θ^4 everywhere. Therefore even small changes in θ lead to strong variations in χ and hence also in the solution I of (11.1c).

In Chapter 12 we compare different methods for solving the RT equation (11.1c) with these demands in mind. On the one hand, we tested the *Discontinuous–Galerkin* finite element scheme introduced in [LR74]. The intensity is approximated on each element T in a space of polynomials $P_k(T)$ using a variational formulation of the radiation transport problem (11.1c). To ensure stability, no continuity of the approximation is enforced over cell boundaries. This leads to the following linear system of equations for the approximate intensity $I_T \in P_k(T)$ on an element T for a given direction $\boldsymbol{\mu}$:

$$\int_T (\boldsymbol{\mu} \cdot \nabla I_T + \chi I_T)\varphi_i - \int_{\partial T_-} I_T \varphi_i \boldsymbol{\mu} \cdot \mathbf{n} = \int_T \chi B \varphi_i - \int_{\partial T_-} I_g \varphi_i \boldsymbol{\mu} \cdot \mathbf{n} \qquad \text{(DG)}$$

where I_g is the intensity on the inflow boundary of the element T. The test functions φ_i $(i = 1, \ldots, r)$ are a basis of the space $P_k(T)$ of polynomials of order k on T. The convergence of the DG method was established in [JP86] (cf. also [Ric88, Pet91]). A study of some other finite element approaches can be found, for example, in [Füh93, Tur93, Kan96].

11.1 Theorem (Convergence of DG Method)
Consider a locally quasi–uniform family of triangulations $\{\mathcal{T}_h\}_h$ *of a convex and polygonal domain* $\Omega \subset \mathbb{R}^2$ *and let* $k \geq 0$ *be fixed. Denote for* $T \in \mathcal{T}_h$ *the space of polynomials of degree* k *on* T *with* $P_k(T)$ *and define* $V_k := \{v \in L^2(\Omega) | v_{|T} \in P_k(T) \forall T \in \mathcal{T}_h\}$. *Furthermore for* $v \in V_k$ *define* $\|v\|^2_{L^2(\Omega),\boldsymbol{\mu}} := \sum_{T\in\mathcal{T}_h} \|\boldsymbol{\mu} \cdot \nabla v\|^2_{L^2(T)}$.
Let $I \in V := \{u \in L^2(\Omega) | \boldsymbol{\mu} \cdot \nabla u \in L^2(\Omega)\}$ *be the solution to the RT equation* (11.1c) *with data satisfying:* $\chi S \in L^2(\Omega)$, $g \in L^2(\partial\Omega_-)$, *and* $\chi \in L^\infty(\Omega)$. *Let* $I_h \in V_k$ *be the discrete approximation using the kth order DG method. If* $I \in H^{k+1,2}$ *then the following error estimates hold*

$$\|I - I_h\|_{L^2(\Omega)} \leq Ch^{k+\frac{1}{2}} |I|_{H^{k+1,2}} \ ,$$
$$\|I - I_h\|_{L^2(\Omega),\boldsymbol{\mu}} \leq Ch^k |I|_{H^{k+1,2}}$$

for h *sufficiently small and with a constant* $C > 0$ *independent of* h.

Proof:
The proof can be found in [JP86]. □

We also studied two approaches based on characteristics that were developed for Cartesian grids: the *long–characteristics* method suggested by [MAM78] and the *short–characteristics* method proposed by [KA88]; the latter was recently adapted to triangular grids [BVS99]. Although these methods are often used, we are not aware of a convergence proof or even of a numerical investigation of the order of convergence. We study both issues in the following chapter.
Both the finite element and the characteristic based approaches have been used in many applications concerning neutron and radiation transport problems, e.g. [Tur93, AP94, Kan96, FK97, SGDKS98, Ada99, HSS00]. However, to our knowledge [DV02] is the first comparison of these methods in respect to their error to runtime ratios.
In view of the demands sketched above, both methods suffer either from computational inefficiency, low accuracy, or are not suitable in conjunction with new computational strategies, such as parallel processing and locally adapted unstructured grids. Higher order versions of these methods also tend to lead to over– and undershoots in the solution, which can even result in unphysical negative intensity values. Therefore it is necessary to study possible improvements of well–known methods and to develop new methods for the solution of the RT equation (11.1c).
In [DV02] we present a new higher order method that combines the finite element and the short–characteristics approaches. The advantage of our approach is that the order of the method can be easily changed without major recoding. Furthermore we are able to reduce problems caused by spurious oscillations through a simple modification of the algorithm. The scheme works locally with data delivered by a (M)HD–solver on structured or unstructured grids in two and three space dimensions. We compare the

first and second order versions of our method with other well–known methods in regard to the experimental order of convergence (EOC) and the error to runtime ratio. We also study some simple adaptation strategies. In Chapter 12 we also present some more recent developments: we study a simple extension that leads to a *conservative scheme*, and we present some analytical results including a convergence proof for our scheme on unstructured grids in 2d.

11.1.3 Approximation of Data

The functions χ and B defining the radiation intensity I through (11.1c) depend on the fluid temperature θ and density ρ. In the case of frequency integrated data, the function B is proportional to θ^4 in the whole domain (cf. Section 1.2); the function χ is defined only by tabularized values and is proportional to θ^{10} in some parts of the solar photosphere. Due to the partial ionization of the plasma in this region of the sun, the temperature is defined only implicitly by the conservative quantities $\mathbf{U}$. In our time dependent numerical simulations, these quantities are approximated on some grid by constant values. Due to the non–linear dependence of the data in the RT equation on the conservative variables, the approximation of the hydrodynamic quantities leads to difficulties in calculating the radiation source term Q_{rad}. Therefore a suitable reconstruction of the data is essential for a meaningful approximation of Q_{rad}. We focus on these aspects of the approximation in Chapter 13.

11.2 Test Cases

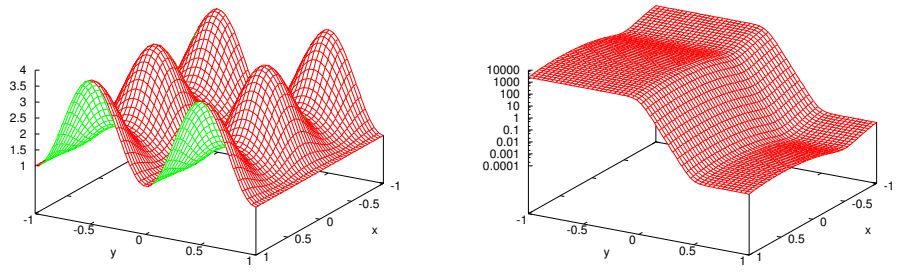
In the following we describe the test cases that we use to study our RT scheme. The radiation intensity I in a fixed direction $\boldsymbol{\mu}$ is defined by the data functions χ, B in (11.1c). In our finite–volume scheme we must compute the radiation intensity for fixed ρ, θ. Consequently, we can assume in our tests that the data functions B, χ are given functions of the space variable $\mathbf{x} \in \Omega$. Furthermore the intensity g on the inflow boundary $\partial\Omega_-$ has to be prescribed. The solution I to the RT equation (11.1c) for given data χ, B, g can always be described in a point $\mathbf{x} \in \Omega$ by integrating (11.1c) along the characteristic connecting $\mathbf{x}$ with a point $\mathbf{q}$ on the inflow boundary $\partial\Omega_-$:

$$I(\mathbf{x}, \boldsymbol{\mu}) = \underbrace{g(\mathbf{q})\ e^{-\Delta\tau(0,s,\boldsymbol{\mu})}}_{\textit{attenuated incident intensity}} + \underbrace{\int_0^s \chi(\mathbf{q}+\sigma\boldsymbol{\mu}) B(\mathbf{q}+\sigma\boldsymbol{\mu})\ e^{-\Delta\tau(\sigma,s,\boldsymbol{\mu})}\ d\sigma}_{\textit{emitted intensity}} \tag{11.2a}$$

with (for $a, b \in \mathbb{R}$)

$$\Delta\tau(a, b, \boldsymbol{\mu}) := \int_a^b \chi(\mathbf{q}+\sigma\boldsymbol{\mu})\ \mathrm{d}\sigma \tag{11.2b}$$

and $\mathbf{q} + s\boldsymbol{\mu} = \mathbf{x}$. Only in very special cases is it possible to derive a closed form for I. Therefore we mainly make use of the technique in which a function I is chosen and then the data is computed in such a way that I is a solution of the RT equation (11.1c)



(a) intensity function I **(b)** absorption function χ

Figure 11.1: The data for Problem 11.1(SMOOTH) is defined by using the function χ and the function I to define the source term B using (11.3). Note that on the right the χ axis is scaled logarithmically.

(cf. Section 3.7.5). In our first two test cases we choose I independent of $\boldsymbol{\mu}$ and also a function $\chi = \chi(\mathbf{x})$. Note that in contrast to the physically relevant situation, this approach leads to a source term B that depends on the direction $\boldsymbol{\mu}$. Furthermore since I is independent of $\boldsymbol{\mu}$, it is easy to verify that $Q_{\text{rad}} \equiv 0$. The source function B is defined as

$$B(\mathbf{x}; \boldsymbol{\mu}) = \frac{\boldsymbol{\mu} \cdot I(\mathbf{x})}{\chi(\mathbf{x})} + I(\mathbf{x}) \tag{11.3}$$

and the boundary data as $g(\mathbf{x}) = I(\mathbf{x})$.

Problem 11.1(SMOOTH) Smooth Solution
For a positive constant α we choose

$$\begin{aligned} I(x, y, \boldsymbol{\mu}) &:= (\cos(2\pi x) + 2)\sin(\pi y)^2 + 1 \ , \\ \chi(x, y) &:= (1000 \tanh(-\alpha y) + 1000.001)(\sin(\pi x) + 1.25) \ . \end{aligned}$$

Then we solve the RT equation (11.1c) *with absorption coefficient χ and with B given by* (11.3). *Note that χ is chosen so that in one region ($y < 0$) of the domain the absorption is very large, whereas in the second region ($y > 0$) the absorption is very small. The transition between these two regions is controlled by the constant α with α large corresponding to a sharp transition. For all choices of α the data and the solution are in $C^\infty(\mathbb{R}^2)$. In Figure 11.1 we plot the functions I and χ for $\alpha = 15$.*

Problem 11.2(H3) H^3 Solution
For the following problem we choose $I \in H^{3,\infty}(\mathbb{R}^2)$:

$$\begin{aligned} I(x, y, \boldsymbol{\mu}) &:= \begin{cases} -r^6 + 3r^4 - 3r^2 + 2 & r < 1 \ , \\ 1 & \textit{otherwise,} \end{cases} \\ \chi(x, y) &:= \alpha \end{aligned}$$

with $r = \frac{16}{9}(x^2 + y^2)$ and $\alpha \in \mathbb{R}^+$. We solve the RT equation (11.1c) *with absorption coefficient χ and with B given by* (11.3). *The solution I is plotted in Figure 11.2.*

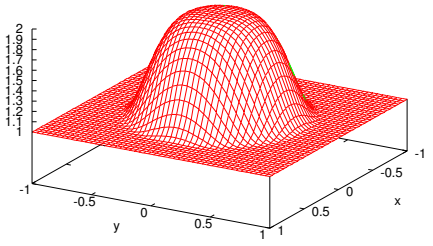


Figure 11.2: The data for Problem 11.2(H3) is defined by using the function I and a constant absorption coefficient χ to define the source term B using (11.3).

For the next two problems the exact solution I can be computed using (11.2). In both cases the absorption χ is equal to a constant χ_0 in the whole domain; thus (11.2b) reduces to $\Delta\tau(a, b, \boldsymbol{\mu}) = \chi_0(b - a)$. In the first case B is a characteristic function, yet the solution I is still continuous. In the second case B is zero, so that only the first part in (11.2a) is relevant. The boundary data g is a characteristic function, which leads to a discontinuous intensity I. Both problems have been studied in the literature [KA88, Füh93, Ded98].

Problem 11.3(STAR) Star Problem
The data for the RT equation (11.1c) *is given by*

$$\begin{aligned} B(x,y) &:= \begin{cases} 11 & x^2 + y^2 < 0.09 \ , \\ 1 & \textit{otherwise} \ , \end{cases} \\ \chi(x,y) &:= 2 \ , \quad g(x,y) := 1 \ . \end{aligned}$$

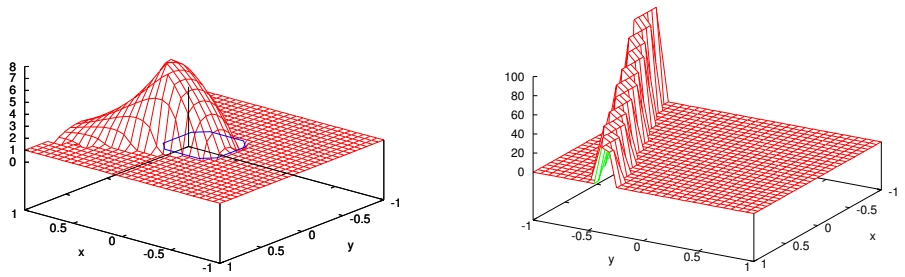
Due to a localized source, the radiation intensity is transported into the surrounding domain (cf. Figure 11.3(a)).

Problem 11.4(SEARCHLIGHT) Searchlight Problem
The data for the RT equation (11.1c) *is given by*

$$\begin{aligned} B(x,y) &:= 0 \ , \quad \chi(x,y) := 0.5 \ , \\ g(x,y) &:= \begin{cases} 100 & x \in [-0.95, -0.55] \ , \\ 0 & \textit{otherwise} \ . \end{cases} \end{aligned}$$

Due to a localized source on the boundary, the radiation intensity is transported into the domain (cf. Figure 11.3(b)).

The following two problems serve as physically realistic test cases for the RT solvers. The setting for these problems is taken from [BVS99] and was also studied in [DV02]. Both our test cases for the solar photosphere. The first problem simply describes a stratified, quiet atmosphere, and the second describes a magnetic fluxsheet embedded in this atmospheric model. In the following chapters we will mostly use y to denote the height in the atmosphere, although in some cases z is used instead in accordance with the usual physical conventions.

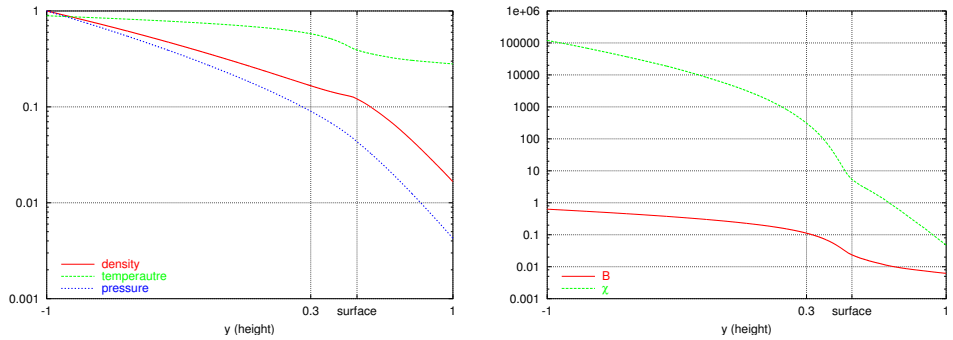


(a) Problem 11.3(STAR) **(b)** Problem 11.4(SEARCHLIGHT)

Figure 11.3: Solutions for Problem 11.3(STAR) and Problem 11.4(SEARCHLIGHT) for a propagation direction $\boldsymbol{\mu} = \boldsymbol{\mu}_1$ (cf. Table 3.1). In Problem 11.3(STAR) the radiation is generated by a localized source term (indicated by the circle in the left plot). In Problem 11.4(SEARCHLIGHT) radiation is generated by a local increase in the intensity on the boundary. In both cases the intensity is then transported through the domain and simultaneously absorbed ($\chi \equiv \chi_0 > 0$).

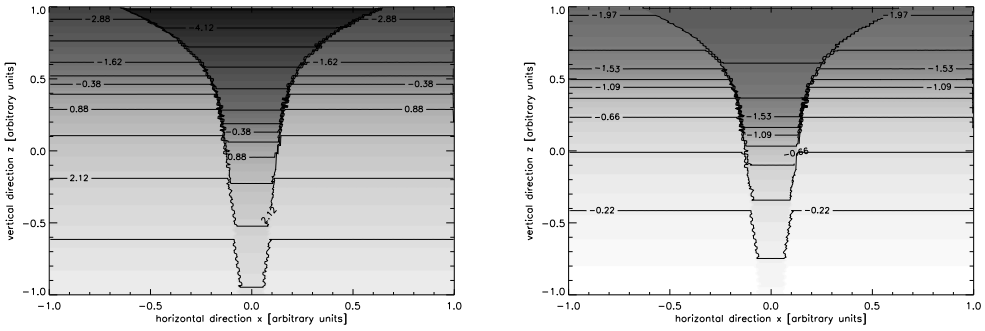
Problem 11.5(SOLARATM) Model Solar Atmosphere
This setting is a model for the solar photosphere. The radiation source term B is given by $B = \sigma\theta^4$ where θ is the temperature in the atmosphere (cf. Section 1.2). The absorption coefficient χ is the averaged Rosseland absorption coefficient (cf. [Kur96]) and is only defined through a table depending on the temperature θ and the density ρ. The hydrodynamic quantities ρ, θ depend only on the height in the atmosphere (cf. Figure 11.4). This setting was studied in [BVS99].

Problem 11.6(FLUXSHEET) Fluxsheet
This setting describes an embedded schematic magnetic fluxsheet in the solar atmosphere given by Problem 11.5(SOLARATM). At a given height z the hydrodynamic quantities in the interior of the fluxsheet are also defined by the model solar atmosphere, but from a height that lies slightly above z. Consequently the plasma in the interior of the fluxsheet is slightly cooler and less dense. At the boundary of the fluxsheet the hydrodynamic quantities show a sharp transition, which leads to a sharp transition in the source term B and the absorption coefficient χ as shown in Figure 11.5.



(a) hydrodynamic quantities **(b)** data defining radiation intensity

Figure 11.4: The data for Problem 11.5(SolarAtm) is defined by a model for a stratified atmosphere. The visible solar surface is in the region where the absorption coefficient drops by several orders of magnitude. Here radiation is emitted and transported outwards through the atmosphere.



(a) isoline representation of $\log_{10}\chi$ **(b)** isoline representation of $\log_{10} B$

Figure 11.5: Data functions χ, B for solar fluxsheet model (Problem 11.6(Fluxsheet)). Both functions grow with increasing geometrical depth z and show a sharply marked transition between the interior and the exterior of the fluxsheet. The atmosphere in the interior is shifted downwards compared to the surrounding atmosphere that is given by Problem 11.5(SolarAtm).

Chapter 12

The Extended Short–Characteristics Method

In this chapter we study a new second order method for solving the radiation transport equation (11.1c) that is well suited to be used with our finite–volume scheme in 2d as well as in 3d. It is constructed according to the demands given in Section 11.1.2. The presentation closely follows [DV02] where the *extended short–characteristics* method was first suggested. We add to the discussion in [DV02] by developing a modification of the scheme that increases its order and results in a scheme that is compatible with the integral form of the RT equation (11.1c). We also study analytical properties of the ESC–method. In the next chapter we then study the computation of the integral source term Q_{rad}.
In this chapter we assume that the data defining the radiation intensity I through (11.1c) is fixed; in particular we assume that $\boldsymbol{\mu} \in S^2$ is fixed and that χ, B are functions defined on the whole computational domain $\Omega \subset \mathbb{R}^2$. Section 12.1 describes the general framework of the solution strategy irrespective of the underlying grid structure. A detailed study of the implementation of first and second order versions of the solver in 2d on triangular grids is discussed in the following Sections 12.2–12.4. In Section 12.5 we present an extension of the ESC–method that leads to a conservative scheme. Before we present a detailed numerical study in Section 12.7, we prove the convergence of the first order ESC scheme on unstructured grids in Section 12.6. We conclude our study with the magnetic fluxsheet problem: in Section 12.8 we compare the performance of the different solution method, and in Section 12.9 we discuss adaptation strategies such as local grid adaptation and the local variation of the order of the scheme.

12.1 General Framework

In this section we develop a general framework for a class of solution schemes for the RT equation (11.1c). We extend the technique of short–characteristics described in [KA88] by embedding the method in a finite element framework. This permits a very general formulation of the scheme that can be used to construct methods of higher order on different types of grids. For the description of our scheme we study the RT equation on a bounded, open, and connected subset ω of $\mathbb{R}^n$ for $n \geq 2$; this can be either the full

computational domain Ω or a subset of Ω, such as a single grid element. In ω we now seek an approximation I_ω of the solution I to the radiation transport equation (11.1c) for a fixed vector $\boldsymbol{\mu} \in \mathbb{R}^n$ assuming that the radiation intensity on the inflow boundary

$$\partial\omega_- = \partial\omega_-^{\boldsymbol{\mu}} := \{\mathbf{x} \in \partial\omega_- : \boldsymbol{\mu} \cdot \mathbf{n}(\mathbf{x}) < 0\}$$

is known (here $\mathbf{n}(\mathbf{x})$ is an outer normal to $\partial\omega$, which we assume exists for almost all $\mathbf{x} \in \partial\omega$). We denote the inflow intensity with I_g and assume that $I_g \in C^0(\partial\omega_-)$. For the other data functions we assume $\chi, B \in C^0(\overline{\omega})$. Our aim is to find an approximation I_ω in a given function space $P(\omega)$ with finite dimension r. For example, if ω is an element of a given triangulation, then $P(\omega)$ could be a space of polynomials on that triangle and I_g is the approximate intensity function computed on the neighboring triangles in downwind direction (cf. Section 3.1).

For a set of basis functions $\{\varphi_i\}_{1\le i\le r}$ of $P(\omega)$ and points $\mathbf{p}_i \in \overline{\omega}$ $(1 \le i \le r)$ we assume

$$\varphi_i(\mathbf{p}_j) = \delta_{ij} \quad \text{with} \quad 1 \le i, j \le r \qquad (12.1)$$

and that there exists a $m \ge 0$ with

$$\begin{aligned} \mathbf{p}_i &\in \partial\omega_- \qquad \text{for} \quad 1 \le i \le m\ , \\ \mathbf{p}_i &\in \omega \setminus \partial\omega_- \qquad \text{for} \quad m < i \le r\ . \end{aligned}$$

This setting is sketched in Figure 12.1. We represent I_ω in the basis $\{\varphi_i\}_{1\le i\le r}$

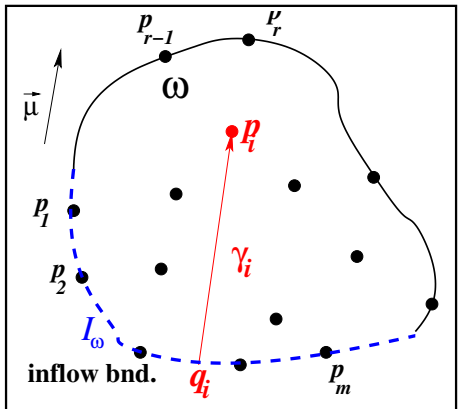


Figure 12.1: General setting of the ESC–method

$$I_\omega(\mathbf{x}) = \sum_{j=1}^{r} I_j \varphi_j(\mathbf{x}) \qquad (12.2)$$

with unknown coefficients $I_1, \ldots, I_r \in \mathbb{R}$. Using (12.1) it follows that

$$I_j = I_\omega(\mathbf{p}_j) \quad (1 \le j \le r).$$

We utilize this to calculate the coefficients I_j in such a way that $I_\omega(\mathbf{p}_j)$ is an approximation of $I(\mathbf{p}_j)$.

The intensity is given at the points $\mathbf{p}_j$ $(j = 1, \ldots, m)$ since these lie on the inflow boundary of ω; therefore we choose

$$I_j = I_g(\mathbf{p}_j) \quad \text{for} \quad 1 \le j \le m. \qquad (12.3)$$

To find the other coefficients we use the method of short–characteristics. Since the characteristic of (11.1c) through $\mathbf{p}_j$ is a straight line and since ω is bounded, the characteristic must intersect the inflow boundary $\partial\omega_-$. We denote the first intersection with $\mathbf{q}_j$ and the length of the characteristic between $\mathbf{q}_j$ and $\mathbf{p}_j$ with $s_j > 0$, i.e.

$$\mathbf{q}_j = \mathbf{p}_j - s_j\boldsymbol{\mu}\ , \quad \gamma_j := \{\mathbf{x} : \mathbf{x} = \mathbf{q}_j + s\boldsymbol{\mu}, 0 \le s \le s_j\} \subset \omega$$

(cf. Figure 12.1). Along the characteristic γ_j the function $v_j(s) := I(\mathbf{q}_j + s\boldsymbol{\mu})$ satisfies the following initial value problem:

$$\begin{aligned} v_j'(s) + \chi(\mathbf{q}_j + s\boldsymbol{\mu}) v_j(s) &= \chi(\mathbf{q}_j + s\boldsymbol{\mu}) B(\mathbf{q}_j + s\boldsymbol{\mu}) \quad \text{for} \quad 0 \le s \le s_j, \\ v_j(0) &= I_g(\mathbf{q}_j)\ . \end{aligned} \tag{12.4}$$

With a suitable approximation $\overline{v_j}$ of v_j, we now choose the remaining coefficients:

$$I_j := \overline{v_j}(s_j) \quad \text{for} \quad m < j \le r. \tag{12.5}$$

In the following we call (12.4) the *short–characteristic problem*. In the ESC–method the approximate solution I_ω of the radiative transport equation (11.1c) on ω is given through (12.2) together with (12.3) and (12.5).

12.2 Implementation on Unstructured Grids

For simplicity we describe the details of the ESC–method only in the case of two space dimensions focusing on triangular grid; the case of a structured grid or the extension to higher space dimensions is easily derived. We construct an approximate solution I_h to the RT equation

$$\begin{aligned} \boldsymbol{\mu} \cdot \nabla I + \chi I &= \chi B \qquad \text{in} \quad \Omega, \\ I &= g \qquad \text{on} \quad \partial\Omega_-, \end{aligned} \tag{12.6}$$

with a given propagation direction $\boldsymbol{\mu}$. We assume that we have a given grid $\mathcal{T}_h$ on the bounded domain $\Omega \subset \mathbb{R}^2$. In our applications the data are given by a finite–volume scheme, so that we assume in the following that χ, B are functions that are continuous on each element of the grid $\mathcal{T}_h$. The inflow intensity $g \in L^\infty(\partial\Omega_-)$ is assumed to be continuous on each boundary segment of $\mathcal{T}_h$.
One way to find an approximate solution to (12.6) with the ESC–method is to set $\omega := \Omega$ in Section 12.1 and to let $\mathbf{p}_j$ be the nodes of the grid $\mathcal{T}_h$. The initial value problem (12.4) is then solved for each characteristic by connecting the nodes of the grid with the corresponding starting point $\mathbf{q}_j$ on the inflow boundary as sketched in Figure 12.2. The resulting scheme is called the *method of long–characteristics* and was first proposed in [MAM78] for Cartesian grids. The same idea can also be used on triangular grids, where it results in a piecewise linear and continuous approximation. Using standard results from interpolation theory and assuming that the short–characteristic problem (12.4) is solved with second order accuracy, it is easy to verify that the method of long–characteristics leads to a second order accurate approximation, if $I \in H^2(\Omega)$ holds. The computational cost of the method is, however, very high: let N denote the number of elements in the grid, then the number of elements that are intersected by each characteristic is in $O(\sqrt{N})$, and we have $O(N)$ characteristics. Thus the complexity is about $O(N^{\frac{3}{2}})$. This method is especially inefficient in combination with parallelization strategies such as domain decomposition (cf. Section 3.6).
Another way to use the ESC–method is to apply the scheme on each grid element $T \in \mathcal{T}_h$ separately (i.e. $\omega = T$). This requires a decomposition of the algorithm into three

consecutive steps. First we have to construct a processing sequence of the elements as discussed in Section 3.1. The construction of the discrete intensity approximation on an element T then requires choosing a suitable function space with basis functions φ_i and points $\mathbf{p}_j$ satisfying (12.1). Finally we have to specify a method for approximating the short–characteristic problem (12.4). In the following we focus on the implementation for triangular grids.

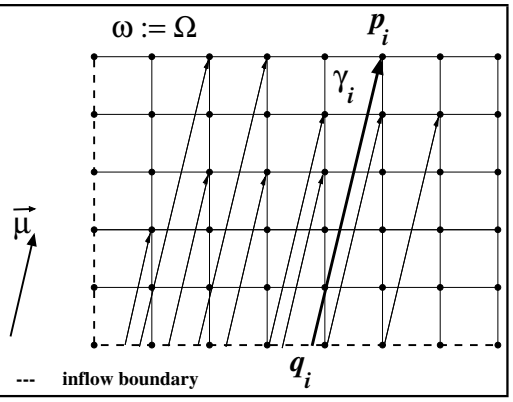


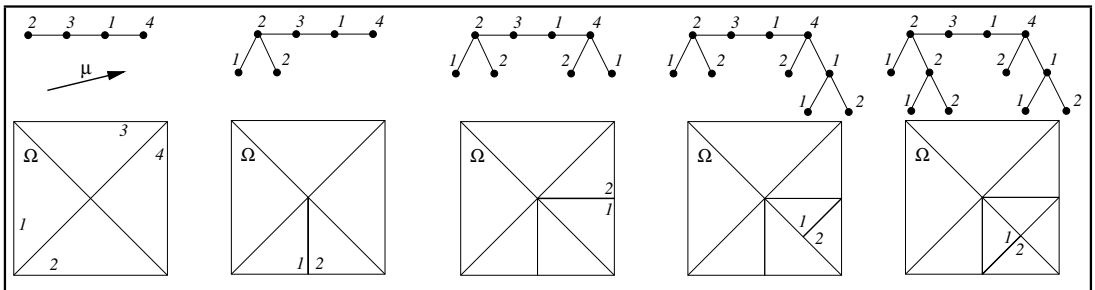
Figure 12.2: Long–characteristics method ($\omega = \Omega$): the radiation transport equation is solved along each characteristic γ_i that connects a vertex $\mathbf{p}_i$ in a Cartesian grid with the inflow boundary at $\mathbf{q}_i$.

12.2.1 Step 1: ordering the triangles

The solution technique described in Section 12.1 with $\omega = T$ (T is an element of a given triangulation $\mathcal{T}_h$ of Ω) can be used directly to compute the intensity I_T, if the intensity on the inflow boundary ∂T_- is known. If this inflow boundary coincides with the outflow boundary of some other triangle $\hat{T}$, the intensity $I_{\hat{T}}$ on $\hat{T}$ has to be calculated prior to I_T. Therefore a processing sequence of the elements $T_1, \ldots, T_N$ of $\mathcal{T}_h$ has to be found that satisfies:

$$\text{For two triangles } T_i, T_j \text{ with } T_i \cap T_j \subset \partial T_{j-}, \; i \leq j \text{ holds.} \tag{12.7}$$

Such a sequence allows us to use the technique described in Section 12.1: the construction of the discrete solution to (12.6) is started on T_1 since, due to property (12.7), the inflow boundary of T_1 must lie on $\partial\Omega_-$, where the intensity is given. Assuming that we have constructed a solution on $T_1, \ldots, T_{j-1}$, we can compute I_{T_j} using our method, since the inflow boundary of T_j is either part of $\partial\Omega_-$ or it is part of the boundary of the triangles $T_1, \ldots, T_{j-1}$, due to (12.7). We already discussed this idea for solving the RT equation in Section 3.1. We now show how to construct such a sequence in the case of a locally adapted grid stored in a hierarchy (cf. Section 2.2).
We first sort the macro grid using a simple algorithm: we start with one element T of the macro grid and check whether all the neighboring triangles in the downstream direction are included in the sequence. If this is the case, T is taken as the next element in the sequence, and the procedure is repeated with some other triangle. If not, the procedure is repeated with one of the downstream neighbors, until an element is found that can be included in the sequence. The ordering of the macro grid has to be constructed *only once*, at the beginning of the calculation. Since the number of elements in the macro grid is usually small compared to the total number of elements,



(a) macro grid **(b)** 1st step **(c)** 2nd step **(d)** 3rd step **(e)** final grid

Figure 12.3: Example of a refinement sequence of a triangulation shown together with the hierarchical storage and the *local* numbering.

the cost of this step is negligible. In each refinement step we sort the new elements (most often two or four) *locally*, i.e., only with respect to each other; the information stored for the other elements remains untouched. This can be done in a short time irrespective of the total size of the grid and leads to a local numbering of the nodes in the tree as shown in Figure 12.3.
For the solution of the radiation transport problem we traverse down the subtrees in the order given by the *local* numbering. This is repeated on each level until all the elements of the tree have been visited. An example is shown in Figure 12.4.

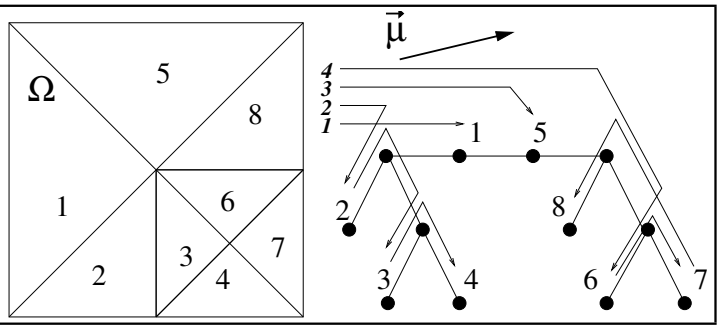


Figure 12.4: Construction of a sequence satisfying (12.7) using the *local* numbering from Figure 12.3(e).

12.2.2 Step 2: solution on a single element

We now consider the approximation on a given element T of a triangulation $\mathcal{T}_h$, i.e. $\omega = T$ in Section 12.1. To use our method, the points $\mathbf{p}_i \in T$ $(i = 1, \ldots, r)$ and an ansatz space $P(T)$ with a set of basis functions $\{\varphi_i\}_{1 \le i \le r}$ satisfying $\varphi_i(\mathbf{p}_j) = \delta_{ij}$ $(1 \le i, j \le r)$ have to be selected. We examine two possibilities — which we call ESC1 and ESC2, respectively — for choosing $\mathbf{p}_i$ and φ_i using the space

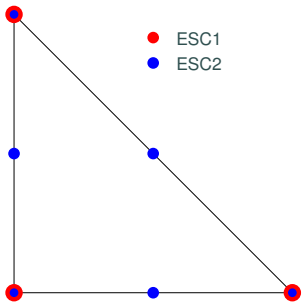
$$P_k(T) := \{p : p \text{ is a polynomial on } T \text{ of degree less than or equal to } k\} \qquad (12.8)$$

for $k \in \mathbb{N}$. We use the space of linear polynomial for the ESC1–method and the space of quadratic polynomials for the ESC2–method. The details are summarized in Figure 12.5.

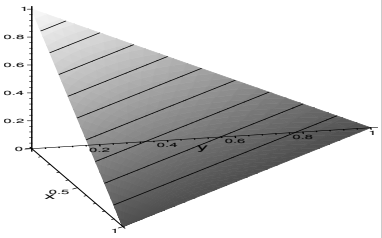
- **ESC1**: For the first order method linear ansatz functions are chosen: $P(T) := P_1(T)$ with $\dim P(T) = 3$. Three points $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3 \in T$ must be defined, at which the coefficients in the representation (12.2) are computed. The natural choice for these points are the three nodes of the triangle T. This choice results in basis functions shown in (b).

- **ESC2**: For the desired higher order scheme, the choice is the six dimensional function space $P(T) := P_2(T)$. For the points $\mathbf{p}_j$ we choose the nodes and the midpoints of the edges. This leads to two types of basis functions, shown in (c) and (d).

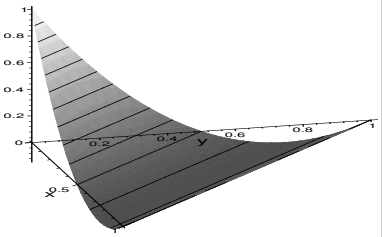
Figure 12.5: The top figure shows the location of the points $\mathbf{p}_j$ used for the ESC1– and the ESC2–method studied here. The points from the ESC1–method are also used in the ESC2–method — the vertices are also included in the ESC2–method. The other figures show graphs of typical basis functions.



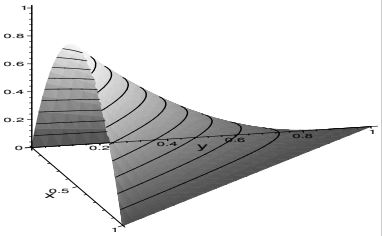
(a) location of $\mathbf{p}_j$ in reference triangle



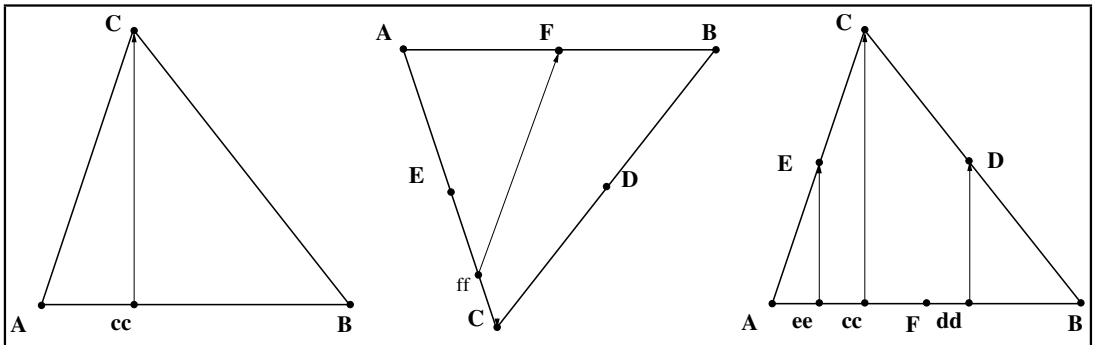
(b) nodal basis for ESC1



(c) nodal basis for ESC2



(d) mid–point basis for ESC2



(e) one inflow edge **(f)** two inflow edges **(g)** one inflow edge

Figure 12.6: To solve the RT equation on a single element using the ESC–method, it is necessary to calculate the intensity for different points, depending on the desired order.
(a): in the ESC1–method the intensity at point C has to be computed on triangles with only one inflow edge along the characteristic starting at point cc.
(b) and (c): in the ESC2–method the intensities have to be computed at point F (two inflow edges) or E, C, D (one inflow edge) using the corresponding inflow intensities at point ff or at points ee, cc, and dd, respectively.

In the ESC1–method we use a linear approximation and the coefficients (12.5) must be determined at the vertices of each triangle. There are two possibilities: either all three vertices lie on the inflow boundary ∂T_- ($m = r = 3$ cf. (12.3)), in which case the linear function I_T is already uniquely determined; or we have $m = 2$, and the intensity I_C at point C (cf. Figure 12.6(a)) must be computed. The intensity I_{cc} at the starting point cc of the characteristic, which serves as the inflow intensity, is given by the solution on the neighboring triangle.
In the ESC2–method a quadratic behavior of the ansatz functions is assumed. Again, depending on the number of inflow boundaries, two cases have to be distinguished. In the first case (two inflow boundaries, $m = 5$ Figure 12.6(b)), the short–characteristic problem has to be solved for the mid point F of the outflow boundary, with the starting value given by the intensity at the point ff. In the second case ($m = 3$, Figure 12.6(c)) the coefficients at the points E, C and D have to be computed. In this case the inflow intensities at the points ee, cc and dd, (I_{ee}, I_{cc}, I_{dd}, respectively) are known.

12.2.3 Step 3: solution of the short–characteristic problem

We now discuss the solution of the short–characteristic problem (12.4), which was one of the main aspects of our numerical tests in [DV02], where we compared four different solvers for the ODE (12.4). On the one hand we tested the two methods used in [KA88, BVS99] termed KA1 and KA2. Furthermore we used fully implicit Runge–Kutta solvers and simple diagonally implicit Runge–Kutta methods (e.g. [HW91]). We note that the ODE is linear and that therefore only a linear system of equations has to be solved when using an implicit Runge–Kutta method. We have also tested further methods, including several quadrature methods for the integrals in (11.2a) and (11.2b), as well as a number of explicit and implicit Runge–Kutta methods. Most methods

$$\begin{array}{c|ccc} c_1 & a_{11} & \dots & a_{1r} \\ \dots & \dots & \dots & \dots \\ c_r & a_{r1} & \dots & a_{rr} \\ \hline & b_1 & \dots & b_r \end{array}$$

(a) table for coefficients of general r step Runge–Kutta method

$$\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}$$

(b) one step method

$$\begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & \frac{-1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array}$$

(c) two step method

$$\begin{array}{c|ccc} \frac{4-\sqrt{6}}{10} & \frac{88-7\sqrt{6}}{360} & \frac{296-169\sqrt{6}}{1800} & \frac{-2+3\sqrt{6}}{225} \\ \frac{4+\sqrt{6}}{10} & \frac{296+169\sqrt{6}}{1800} & \frac{88+7\sqrt{6}}{360} & \frac{-2-3\sqrt{6}}{225} \\ 1 & \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \\ \hline & \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \end{array}$$

(d) three step method

Table 12.1: Coefficients for the Radau IIa ODE solver of order $1, 3$, and 5.

showed difficulties with the strong variations in χ occurring in our applications. The best results were obtained using the r–step Radau IIa implicit Runge–Kutta method, which is of the order $2r - 1$. In Table 12.1 the coefficients for the Radau IIa methods for $r = 1, 2, 3$ are summarized (cf. [HW91]). Using the notation from Table 12.1(a) the approximation for the ODE (12.4) is given by

$$\overline{v_j} = I_g(\mathbf{q}_j) + s_j \sum_{i=1}^{r} b_i \chi(\mathbf{q}_j + c_i s_j)\big(B(\mathbf{q}_j + c_i s_j) - u_i\big) \ ,$$

where for $i = 1, \dots, r$ the values u_i are given by the linear system of equations

$$u_i = I_g(\mathbf{q}_j) + s_j \sum_{k=1}^{r} a_{ik} \chi(\mathbf{q}_j + c_k s_j)\big(B(\mathbf{q}_j + c_k s_j) - u_k\big) \ .$$

12.3 Suppressing Spurious Oscillations

In higher order schemes for hyperbolic equations the computed solutions can show spurious oscillatory behavior. In the ESC2–method this results from the fact that the quadratic interpolation of three intensity values I_A, I_F and I_B given at the corresponding points on an inflow edge (e.g. Figure 12.6(c)) might generate a new extremum as sketched in Figure 12.7. As shown in this example, a characteristic originating between the points A and F (e.g. ee and cc in Figure 12.6(c)) carries a negative intensity, although the intensities at A, F and B are all non–negative. This results in a negative value for the coefficients I_{ee} or I_{cc}, leading to an unphysical solution on the whole triangle.

By applying a suitable operator $\mathcal{G}$ to the discrete intensity I_h on the inflow edge S of an element T to determine the initial value for the short–characteristic problem (12.4),

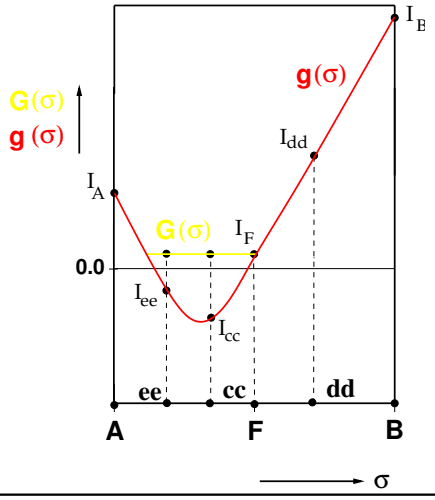


Figure 12.7: Quadratic interpolation can generate new extrema, leading to spurious oscillations, which are suppressed by replacing the inflow intensity g by a modified function G satisfying:
$\min\{I_A, I_F, I_B\} \leq G(\sigma) \leq \max\{I_A, I_F, I_B\}$.

we can eliminate these difficulties. The operator $\mathcal{G}$ must satisfy

$$\min\{I_h(\mathbf{p}_j^T)\} \leq \mathcal{G}(I_h)(\mathbf{x}) \leq \max\{I_h(\mathbf{p}_j^T)\}$$

for all $\mathbf{x} \in S$; the minimum and the maximum are taken over all $\mathbf{p}_j^T \in S$. In our applications we use the following operator on $S \subset \partial T_-$:

$$\mathcal{G}(u)(\mathbf{x}) := u(\mathbf{x}) + \left(\max_{\mathbf{p}_j^T \in S} u(\mathbf{p}_j^T) - u(\mathbf{x}) \right)^- + \left(\min_{\mathbf{p}_j^T \in S} u(\mathbf{p}_j^T) - u(\mathbf{x}) \right)^+ . \qquad (12.9)$$

This operator guarantees that any extremum on the inflow boundary of a triangle corresponds to one of the calculated values as sketched in Figure 12.7. It can be applied in any version of our ESC–method in order to reduce oscillations, as long as at least two $\mathbf{p}_j^T$ lie on S.

12.4 Periodic Boundary Conditions

The algorithm described so far can be used on any domain Ω for which the inflow intensity g is known everywhere on $\partial\Omega_-$. In many applications the intensity is known only on parts of the inflow boundary, and periodic boundary conditions are prescribed on the remaining inflow boundary. In the following we describe an iteration process for solving this type of problem. For simplicity we assume that $\Omega = [x_1, x_2] \times [y_1, y_2]$ for $x_1 < x_2$ and $y_1 < y_2$. The intensity is known on the lower boundary for upward directions and on the top boundary for downward directions; at the vertical boundaries periodic boundary conditions are used. In this situation the sorting algorithm from Section 12.2.1 cannot be used directly since, in general, there is no sequence satisfying (12.7) that takes into account the fact that the inflow intensity at the vertical boundaries is unknown. We solve this problem by iteration, using the intensities at the vertical boundaries that were calculated in one step of the iteration as inflow intensity for the next step. We stop the iteration when the change in the calculated intensities does not exceed a given threshold:

$$|I_{new} - I_{old}| \leq \varepsilon \max\{|I_{old}|, \delta\} \ .$$

Here $\varepsilon > 0$ and $\delta > 0$ are given constants.

12.1 Remark: *We have no proof that this iteration converges in all situations. In our calculations with* $\varepsilon = \frac{1}{10}$ *and* $\delta = 0$ *the algorithm terminates after* $n(\boldsymbol{\mu})$ *steps, with* $n(\boldsymbol{\mu}) = \left[\frac{y_2-y_1}{x_2-x_1}\frac{\boldsymbol{\mu}_x}{\boldsymbol{\mu}_y}\right] + 1$. *The algorithm cannot be used for* $\mu_y = 0$.

12.5 The Conservative ESC–method

In the study of schemes for hyperbolic equations, a lack of conformity with the integral form of the equation (e.g. (2.1)) has been shown to lead to problems — especially in producing correct shock speeds (e.g. [LeV90]). Although a non–conservative scheme can converge to the right solution, a conservative scheme may be desirable. In many applications this is not so relevant, and often schemes are used without taking the integral form of the equations into account. For the approximation of the radiation transport equation (11.1c), the question of satisfying the integral form of (11.1c),

$$\int\limits_{\partial\omega_-} I_g\boldsymbol{\mu}\cdot n + \int\limits_{\partial\omega_+} I\boldsymbol{\mu}\cdot n + \int\limits_{\omega} \chi(I-B) = 0\ . \tag{12.10}$$

on some control volume $\omega \subset \Omega$, is often not so relevant. Other aspects such as the positivity of the discrete intensity I is of greater importance. In the case of neutron transport (which is modeled by the same equation) a *conservative* scheme is, however, often preferable. Since the importance of satisfying (12.10) depends on the application, we present in the following a modification of the ESC–method that leads to a conservative scheme. We first demonstrate that the ESC–method described so far is not conservative:

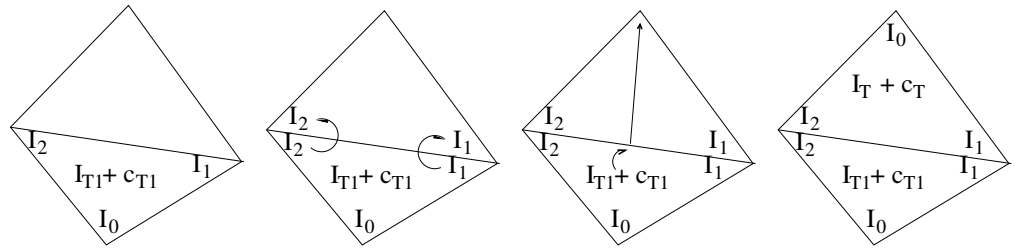
12.2 Example
For $B \equiv 0, \chi \equiv 1, \boldsymbol{\mu} = (0,1), I_g \equiv g_0$ *we compute for* $\omega = \widehat{T}$ *where* $\widehat{T}$ *is the unit simplex:* $I(x,y) = g_0 e^{-y}$. *Taking, for example, the ESC1–method with* $I_h(x,y) = I_0(1-x-y) + I_1 x + I_2 y$ *and assuming that the short–characteristic problem is solved exactly, we compute* $I_0 = g_0$, $I_1 = g_0$, *and* $I_2 = g_0 e^{-1}$ *and therefore* $I_h(x,y) = g_0(1-(1-e^{-1})y)$. *Plugging this expression into* (12.10) *we find*

$$\int\limits_{\partial\omega_-} g_0\boldsymbol{\mu}\cdot n + \int\limits_{\partial\omega_+} I_h\boldsymbol{\mu}\cdot n + \int\limits_{\omega} I_h = g_0\left(\frac{2}{3}e^{-1} - \frac{1}{6}\right) \approx 0.079 g_0.$$

Therefore (12.10) *is not satisfied if* $g_0 \neq 0$.

We can easily modify the ESC–method to make it satisfy the integral form (12.10) on each triangle $T \in \mathfrak{T}_h$. We achieve this by adding a constant c_T to the intensity on each triangle, so that (12.10) is satisfied. As in the DG–scheme this leads to a discrete solution with discontinuities between elements.
Assume now that I_T has been computed using the ESC–method. We then compute c_T — using I_h for the approximation of the intensity already computed on the inflow boundary ∂T_- — by plugging $I_T + c_T$ into (12.10)

$$\int\limits_{\partial T_-} I_h\boldsymbol{\mu}\cdot n + \int\limits_{\partial T_+} (I_T + c_T)\boldsymbol{\mu}\cdot n + \int\limits_{T} \chi(I_T + c_T - B) = 0 \tag{12.11}$$



(a) intensity computed on T_1

(b) reuse coefficients on inflow boundary of T

(c) solve SC problem using $I_{T_1} + c_T$ as inflow intensity

(d) compute conservation fix c_T

Figure 12.8: strategy for computing $\widehat{I}_T = I_T + c_T$ using the CESC1–method.

and solving for c_T. This leads to the following equation for c_T:

$$c_T = -\frac{\int_T \chi(I_T - B) + \int_{\partial T_+} I_T \boldsymbol{\mu}\cdot\mathbf{n} + \int_{\partial T_-} I_h \boldsymbol{\mu}\cdot\mathbf{n}}{\int_T \chi + \int_{\partial T_+} \boldsymbol{\mu}\cdot\mathbf{n}} =: -\frac{Q^J_{\mathrm{rad}T,\boldsymbol{\mu}} + Q^F_{\mathrm{rad}T,\boldsymbol{\mu}}}{\alpha_{T,\boldsymbol{\mu}} + \beta_{T,\boldsymbol{\mu}}} \ . \quad (12.12)$$

In a numerical algorithm the integrals defining c_T have to be replaced by quadrature rules; the values $Q^J_{\mathrm{rad}T,\boldsymbol{\mu}}, Q^F_{\mathrm{rad}T,\boldsymbol{\mu}}, \alpha_{T,\boldsymbol{\mu}}, \beta_{T,\boldsymbol{\mu}}$ can be used for computing the radiation source term and are therefore already defined here. The discrete intensity I_h on the triangle T is now given by

$$I_h(\mathbf{x}) = \widehat{I}_T(\mathbf{x}) := I_T(\mathbf{x}) + c_T \quad \text{for } \mathbf{x} \in T \ . \quad (12.13)$$

For the initial values of the short–characteristic problems on the neighboring triangles, we use the intensity defined by $\widehat{I}_T$. For the coefficients I_j (for $\mathbf{p}_j \in \partial T_-$) we still take the values calculated by the short–characteristic solver; we consequently have to store only one additional value (c_T) on each element. The construction of the approximation in the linear case is sketched in Figure 12.8. We call this modification of the ESC–method the *conservative extended short–characteristics* method and use the abbreviation CESC.
In addition to the CESC1–method and CESC2–method we can also define a CESC–scheme with constant polynomials (CESC0–method). In this case I_T in (12.13) is set to zero and the approximation on T is given by $\widehat{I}_T \equiv c_T$. The defining equation (12.11) for c_T is equivalent to the discretization of the RT equation using the DG–method with zero order polynomials (DG0):

$$\int_T \boldsymbol{\mu}\cdot\nabla I_T + \int_T \chi I_T - \int_{\partial T_-} I_T \boldsymbol{\mu}\cdot\mathbf{n} = \int_T \chi B - \int_{\partial T_-} I_h \boldsymbol{\mu}\cdot\mathbf{n}.$$

Shifting the discrete intensity I_T by a constant value leads to new maxima or minima on the element boundary. As discussed in Section 12.3, this can lead to spurious oscillations, if intensity values that are larger (or smaller) than those calculated by the short–characteristic solver are used to compute coefficients for the intensity approximation. Following the technique presented in Section 12.3, we can considerably reduce these oscillations by modifying the inflow intensity via (12.9); an example is sketched

in Figure 12.9. Note that as in Section 12.3 we take the minimum and maximum of the values on the edge computed by the short–characteristic solver and not the values shifted by c_T.

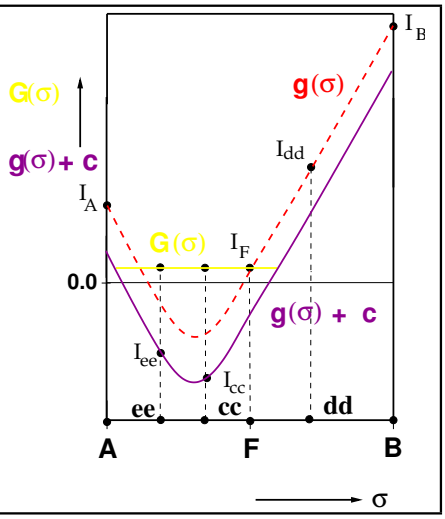


Figure 12.9: The constant shift of the intensity approximation on an element can generate new extrema even in the case of linear ansatz functions. This can lead to spurious oscillations, which can be suppressed by the technique used in the ESC2–method (cf. Section 12.3). For example, in the CESC2–method the inflow intensity g is replaced by a modified function G satisfying: $\min\{I_A, I_F, I_B\} \leq G(\sigma) \leq \max\{I_A, I_F, I_B\}$ where I_A, I_F, I_B are the intensity values computed with the short–characteristics method without the shift.

12.6 Convergence Result

In this section we study some stability issues and prove the convergence of the ESC1–method in the L^∞–norm; furthermore we discuss the convergence of $\boldsymbol{\mu} \cdot \nabla I_h$, which is an important quantity since it can be used to compute the radiation source term Q_{rad} (cf. Chapter 13). For the higher order ESC–methods our convergence proof requires severe restrictions on the operator $\mathcal{G}$ used to reduce oscillations (cf. Section 12.3). We can present only a simple setting where these assumptions are satisfied, but at the same time the ESC–method is reduced to first order accuracy.
In our applications the absorption coefficient χ and the source functions B are both positive, and this is also true for the inflow intensity g. Therefore we assume in the following that $B > 0, \chi > 0$, and $g > 0$ and that B, χ are smooth on each element of a given grid $\mathcal{T}$. Let $I \in L^\infty(\Omega)$ with $\boldsymbol{\mu} \cdot \nabla I \in L^\infty(\Omega)$ be the solution to the radiation transport equation

$$\begin{aligned} \boldsymbol{\mu} \cdot \nabla I + \chi I &= \chi B \quad \text{in} \quad \Omega\,, \\ I &= g \quad \text{on} \quad \partial\Omega_- \end{aligned} \tag{12.14}$$

on some bounded domain $\Omega \subset \mathbb{R}^2$. Denote with I_h the numerical approximations to (12.14) on a family of grids $\{\mathcal{T}_h\}_h$ using the ESC–method with the oscillation reduction technique described in Section 12.3.
In this section we make some assumptions concerning the different components of the ESC–method and the underlying grid. The grid has to satisfy the same assumptions used for our convergence proof of the finite–volume scheme in Chapter 4 (cf. Assumption 4.4). The second set of assumptions restricts the choice of the ansatz space and the set of points $\mathbf{p}_j^T$ for which the short–characteristic problem is solved. They are

posed with standard Lagrange finite elements in mind. For the ODE solver used to approximate the short–characteristic problem (12.4), we assume a simple error estimate that is easily fulfilled, if the data χ, B are sufficiently smooth on each triangle. The most severe assumption is placed on the operator $\mathcal{G}$ used to limit the inflow intensity for the short–characteristic problem (cf. Section 12.3). To apply the following results to the higher order ESC–methods, this operator must be a linear approximation operator that does not generate new minima and maxima. We discuss these assumptions in more detail in Remark 12.4.

12.3 Assumption

(i) The family $\{\mathcal{T}_h\}_h$ of grids on Ω satisfies

$$c_G h^2 \leq |T_i| \ , \quad c_G |S_{ij}| \leq h \ , \tag{12.15}$$

for all $i \in \mathcal{I}^h$ and $j \in \mathcal{N}(i)$ with some constant $c_G > 0$ independent of h. The grid parameter h is again defined as in Definition 4.4: $h = \max_{i \in \mathcal{I}_h} h_i$ and therefore $|T_i| \leq Ch^2$.

(ii) Denote with $\widehat{T}$ the reference triangle with the nodes $\widehat{\mathbf{p}}_1 = (0,0)$, $\widehat{\mathbf{p}}_2 = (1,0)$, and $\widehat{\mathbf{p}}_3 = (0,1)$. We assume that there exist points $\mathbf{p}_j \in \widehat{T}$ and a set of basis functions $\{\varphi_j\}_{j=1}^r$ of some function space $V_r(\widehat{T})$ satisfying $\varphi_j(\mathbf{p}_i) = \delta_{ij}$ for $1 \leq i, j \leq r$. Furthermore, we assume that the vertices of $\widehat{T}$ are included in the set $\{\mathbf{p}_j\}$, i.e. $\mathbf{p}_1 = \widehat{\mathbf{p}}_1$, $\mathbf{p}_2 = \widehat{\mathbf{p}}_2$, and $\mathbf{p}_3 = \widehat{\mathbf{p}}_3$. For $T \in \mathcal{T}_h$ let $\mathbf{F}_T : \widehat{T} \to T$ be the linear mapping from the reference triangle $\widehat{T}$ onto T. Then for $j \in \{1, \ldots, r\}$ the points $\mathbf{p}_j^T$ and basis functions φ_j^T used to construct the ESC approximation are given by

$$\mathbf{p}_j^T = \mathbf{F}_T(\mathbf{p}_j) \ , \quad \varphi_j^T(\mathbf{x}) = \varphi_j(\mathbf{F}^{-1}(\mathbf{x})) \ . \tag{12.16}$$

Note that $\{\mathbf{p}_j^T\}$ and $\{\varphi_j^T\}$ satisfy (12.1).

Consider the operator Π^T that maps some function space $V(T) \subset C^0(T)$ onto $V_r(T) = span(\varphi_1^T, \ldots, \varphi_r^T)$ by means of

$$\Pi^T(u)(\mathbf{x}) = \sum_{j=1}^{r} u(\mathbf{p}_j^T)\varphi_j^T(\mathbf{x}) \ . \tag{12.17}$$

We assume that Π^T satisfies the following interpolation estimate:

$$\|\Pi^T(u) - u\|_{L^\infty(T)} \leq C_{\text{inter}} h^\gamma \tag{12.18}$$

for all $u \in V(T)$ with some $\gamma > 0$.

(iii) Consider $\mathbf{p}_j^T \notin \partial T_-$ for $T \in \mathcal{T}_h$ and $j \in \{1, \ldots, r\}$. We assume that the ODE solver used to approximate the short–characteristic problem (12.4) satisfies the error estimate

$$|\overline{v_j^T}(s_j^T) - v_j^T(s_j^T)| \leq C_{\text{ODE}}(s_j^T)^\alpha \tag{12.19}$$

for some $\alpha > 1$. As in Section 12.1 the positive value s_j^T denotes the length of the characteristic connecting $\mathbf{p}_j^T$ with the inflow boundary of T, v_j^T is the exact solution of (12.4), and $\overline{v_j^T}$ is the approximation of v_j^T at s_j^T.

(iv) Our final assumption is concerned with the operator $\mathcal{G}$ used in Section 12.3 to reduce the oscillations in higher order schemes. This operator has to be defined on each face $S \in \mathcal{S}_h$ of the grid and is therefore denoted with $\mathcal{G}_S$. We make the following assumptions:

- *$\mathcal{G}_S$ is a linear operator on $C^0(S)$.*
- *$\|\mathcal{G}_S(u) - u\|_{L^\infty(S)} \leq C_{\mathcal{G}}|S|^\beta$ for some $\beta > 1$.*
- *For all $\mathbf{x} \in S$ the following estimates hold*

$$\begin{aligned} \mathcal{G}_S(u)(\mathbf{x}) &\leq \max\{u(\mathbf{p}_j^T) : T \in \mathcal{T}_h, 1 \leq j \leq r, \mathbf{p}_j^T \in S\} \ , \\ \mathcal{G}_S(u)(\mathbf{x}) &\geq \min\{u(\mathbf{p}_j^T) : T \in \mathcal{T}_h, 1 \leq j \leq r, \mathbf{p}_j^T \in S\} \ . \end{aligned} \tag{12.20}$$

Note that the vertices of the grid are included in the set of points $\mathbf{p}_j^T$ so that we are taking the maximum and minimum at least over the values of u at the vertices of the grid lying on S.

12.4 Remark: *The assumptions concerning the grid, the basis functions, and the points $\mathbf{p}_j^T$ are standard and can be found in many books on finite element theory, e.g. [Cia78, Theorem 3.1.6]. By choosing $V_r(\widehat{T})$ to be the space of polynomials of degree k and by taking the Lagrange points for $\mathbf{p}_j^T$, the assumptions concerning the interpolation operator Π_T are satisfied with $V(T) = H^{k+1,\infty}(T)$ and $\gamma = k + 1$. This is the setting that we have in mind in the following and that corresponds to our choice in the ESC1– and the ESC2–method (cf. Section 12.2.2).*
If the data is sufficiently smooth, so that the solution I to the radiation transport problem is smooth along each characteristic, then the assumption concerning the ODE method is easily satisfied by choosing a suitable solver with a local truncation error of α. For example, the one step Radau IIa method (better known as the backward Euler scheme) used in the ESC1–method satisfies (12.19) *with $\alpha = 2$, and the two step Radau IIa method used in the ESC2–method satisfies* (12.19) *with $\alpha = 3$.*
The assumptions made on the limiting operator $\mathcal{G}_S$ are the most difficult to satisfy. To prove convergence, a piecewise linear interpolation at the points $\mathbf{p}_j^T$ on S is sufficient to satisfy all assumptions with $\beta = 2$. In the case of the ESC1–method this operator does not change the approximation I_h since I_h is linear on S. For a higher order ESC–method this choice would reduce the method to first order. The operator presented in Section 12.3, which we use in the ESC2–method, is clearly not linear. We are not aware of a suitable approximation operator that satisfies Assumption 12.3(iv) *for $\beta > 2$.*

We first summarize some simple consequences of our assumptions.

12.5 Lemma
Let Assumptions 12.3 be satisfied. Consider the short–characteristic problem for a point $\mathbf{p}_j^T \notin \partial T_-$ ($T \in \mathcal{T}_h, j \in \{1, \ldots, r\}$). Then the length s_j^T of the characteristic satisfies

$$s_j^T \geq \widetilde{c}_G h \ , \quad s_j^T \leq \frac{1}{\widetilde{c}_G} h \tag{12.21}$$

for some constant $\widetilde{c}_G$. *Let* $C_\Phi := \left\| \sum_{j=1}^r |\varphi_j| \right\|_{L^\infty(\widehat{T})}$ *then for all* $T \in \mathfrak{T}_h$ *we have*

$$\left\| \sum_{j=1}^r |\varphi_j^T| \right\|_{L^\infty(T)} \leq C_\Phi \ . \tag{12.22}$$

Proof:
Both estimates are a simple consequence of Assumption 12.3*(ii)*. The second follows directly due to the definition of φ_j^T. For the first we also use Assumption 12.3*(i)*. □

We begin our analysis by studying some stability properties of the ESC–method. For a good approximation of (12.14), it is desirable that special features of the structure of the solution I to (12.14) be recovered by the numerical approximation. We focus on two such properties in the following:

(i): $I(\mathbf{x}) > 0$ for all $\mathbf{x} \in \Omega$ (the intensity is always in the physical regime),

(ii): If $B \equiv 0$ than $I(\mathbf{x}) \leq \max_{\mathbf{y}\in\partial\Omega} g(\mathbf{y})$ for all $\mathbf{x} \in \Omega$ (maximum principle).

By studying the solution I of (12.14) along the characteristic through $\mathbf{x}$, it is easy to see that it satisfies both conditions (cf. (11.2)). The next theorem shows that both properties are also satisfied by the intensity values computed using the ESC–method, if similar conditions are satisfied by the ODE solver used for the solution of the short–characteristic problem (12.4). Furthermore we show that I_h is uniformly bounded.

12.6 Theorem (Stability of the ESC–method)
Let Assumptions 12.3 be satisfied. Consider the RT equation (12.14) *with positive data* χ, B, *and* g. *Let* I_h *be the discrete solution to the RT equation using the ESC–method and including the oscillation suppressing technique with the operator* $\mathcal{G}_S$. *Denote with* I_j^T *(*$T \in \mathfrak{T}_h, j = 1, \ldots, r$*) the intensity approximations computed by the ODE solver for the short–characteristic problem on the element* T. *We assume that this ODE solver satisfies the following stability estimates for the approximation* $\overline{v}$ *to the ODE* (12.4) *with initial condition* $v(0)$*:*

$$\overline{v} > 0 \quad \textit{if} \quad v(0) > 0 \ , \qquad \overline{v} < v(0) \quad \textit{if} \quad B \equiv 0 \quad \textit{on } T \ . \tag{12.23}$$

Then the intensity values I_j^T *are also in the physical regime, i.e.* $I_j^T > 0$ *for all* $j \in \{1, \ldots, r\}$ *and* $T \in \mathfrak{T}_h$.
If $B \equiv 0$ *then* I_j^T *satisfies the maximum principle*

$$I_j^T \leq \max_{\mathbf{y}\in\partial\Omega} g(\mathbf{y}) \tag{12.24}$$

for all $j \in \{1, \ldots, r\}$ *and* $T \in \mathfrak{T}_h$.
If there exists a constant $\chi_0 > 0$ *with* $\chi > \chi_0$ *in* Ω, *then the approximation* I_h *is uniformly bounded in* h, *i.e., there exists a constant* $C > 0$ *independent of* h *so that*

$$\|I_h\|_{L^\infty(\Omega)} \leq C \ . \tag{12.25}$$

Proof:
We use an induction argument based on the construction sequence for the values I_j^T (cf. Section 12.2.1). We first prove that $I_j^T > 0$: Let $T_1, \dots, T_N$ be a suitable construction sequence satisfying (12.7). The values $I_j^{T_1}$ are positive since the initial data for the short–characteristic problem is given by the inflow intensity g, which is positive. Now assume that $I_j^{T_i} > 0$ for all $i \leq n$ and $1 \leq j \leq r$. On the element T_{n+1} the intensity is computed using the short–characteristic solver with initial conditions $v(0)$ given either by the inflow intensity g or by the approximation I_h on one of the inflow edges S of T_{n+1}. In the first case we have $v(0) > 0$. In the second case we also have $v(0) > 0$ since $v(0) = I_h(\mathbf{q}_j^T)$ lies between all the intensity values $I_j^{T_i}$ that belong to points on S due to (12.20):

$$v(0) \geq \min\{I_j^{T_i} : 1 \leq i \leq n, 1 \leq j \leq r, \mathbf{p}_j^{T_i} \in S\} \ .$$

We made the assumption that intensity values have been computed for both nodes of the edge S (cf. Assumption 12.3*(ii)*) so that $v(0) > 0$ follows due the induction hypothesis. Using (12.23) this proves that $I_j^{T_{n+1}} > 0$ for all $j = 1, \dots, r$. With the same argument we prove the maximum principle (12.24).
To prove the boundedness of I_h, we use Lemma 12.5 for the estimate

$$\|I_h\|_{L^\infty(\Omega)} = \max_{T \in \mathfrak{T}_h} \Big\| \sum_{j=1}^{r} I_j^T \varphi_j^T \Big\|_{L^\infty(T)} \leq C_\Phi \max_{j=1,\dots,r, T \in \mathfrak{T}_h} |I_j^T| = C_\Phi |I_{j_0}^{T_0}|$$

with some $j_0 \in \{1, \dots, r\}$ and $T_0 \in \mathfrak{T}_h$. If $\mathbf{p}_{j_0}^{T_0} \in \partial\Omega_-$ then $I_{j_0}^{T_0} = g(\mathbf{p}_{j_0}^{T_0})$, and we continue with

$$\|I_h\|_{L^\infty(\Omega)} \leq C_\Phi \|g\|_{L^\infty(\partial\Omega_-)} \ . \tag{12.26}$$

This concludes the proof. In the case where $\mathbf{p}_{j_0}^{T_0} \notin \partial\Omega_-$ the value $I_{j_0}^{T_0}$ is computed by the short–characteristic solver for some point $\mathbf{p}_{j_1}^{T_1} = \mathbf{p}_{j_0}^{T_0}$ on some element $T_1 \in \mathfrak{T}_h$ with $j_1 \in \{1, \dots, r\}$, i.e. $I_{j_0}^{T_0} = \overline{v}_{j_1}^{T_1}$. To simplify the notation we set $j = j_1$ and $T = T_1$ in the following.
Let $S \in \mathcal{S}_h$ be the inflow edge of T that intersects the characteristic through $\mathbf{p}_j^T$ and let $\mathbf{q}_j^T \in S$ be the starting point for the short–characteristic problem. Using Assumption 12.3*(iii)* and the representation formula for the solutions v_j^T of the short–characteristic problem (12.4) (cf. (11.2)), it follows that

$$\begin{aligned} |I_j^T| \leq & \big(|\overline{v}_j^T - v_j^T| + |v_j^T|\big) \leq C_{\mathrm{ODE}}(s_j^T)^\alpha + \\ & \Big| \mathcal{G}_S(I_h)(\mathbf{q}_j^T) e^{-\Delta\tau(0, s_j^T, \boldsymbol{\mu})} + \int_0^{s_j^T} \chi(\mathbf{q}_j^T + s\boldsymbol{\mu}) B(\mathbf{q}_j^T + s\boldsymbol{\mu})\ e^{-\Delta\tau(s, s_j^T, \boldsymbol{\mu})} ds \Big| \ . \end{aligned} \tag{12.27}$$

As a direct consequence of (12.20) we find

$$|\mathcal{G}_S(I_h)(\mathbf{q}_j^T)| \leq \max\{|I_h(\mathbf{p}_k^{T'})| : T' \in \mathfrak{T}_h, 1 \leq k \leq r, \mathbf{p}_k^{T'} \in S\} \leq |I_j^T| \tag{12.28}$$

using $I_j^T = I_{j_0}^{T_0} = \max_{k=1,\dots,r,T'\in\mathcal{T}_h} |I_k^{T'}|$. As shown in Lemma 12.5 the length s_j^T of the characteristic is bounded from above by $\frac{h}{\widetilde{c}_G}$. Since $\Delta\tau(s, s_j^T, \boldsymbol{\mu}) \geq 0$ for all $s \in [0, s_j^T]$ (cf. (11.2b)), the integral in (12.27) is bounded from above by $\|\chi B\|_{L^\infty(\Omega)} \frac{h}{\widetilde{c}_G}$. Since $\chi > \chi_0$, it follows from (11.2b) using (12.21) that

$$\Delta\tau(0, s_j^T, \boldsymbol{\mu}) = \int_0^{s_j^T} \chi(\mathbf{q}_j^T + \sigma\boldsymbol{\mu}) d\sigma \geq \chi_0 s_j^T \geq \chi_0 \widetilde{c}_G h \ .$$

Thus it follows from (12.27) and (12.28) that

$$|I_j^T| \leq C_{\mathrm{ODE}}(s_j^T)^\alpha + |I_j^T| e^{-\chi_0 \widetilde{c}_G h} + \|\chi B\|_{L^\infty(\Omega)} \frac{1}{\widetilde{c}_G} h$$

and therefore

$$|I_j^T| \leq \Big(C_{\mathrm{ODE}}(s_j^T)^\alpha + \|\chi B\|_{L^\infty(\Omega)} \frac{1}{\widetilde{c}_G} h \Big) \frac{1}{1 - e^{-\chi_0 \widetilde{c}_G h}} \ . \tag{12.29}$$

Since $\frac{h}{1-e^{-\chi_0 \widetilde{c}_G h}}$ is bounded and $\alpha \geq 1$, the left hand side of (12.29) is uniformly bounded in h. Consequently

$$\|I_h\|_{L^\infty(\Omega)} \leq C_\Phi |I_{j_0}^{T_0}| \leq C_\Phi \Big(C_{\mathrm{ODE}}(s_j^T)^{\alpha-1} + \|\chi B\|_{L^\infty(\Omega)} \frac{1}{\widetilde{c}_G} \Big) C \ .$$

Together with (12.26) this concludes the proof. □

12.7 Remark: *With the assumptions made in Theorem 12.6 we can prove the maximum principle and the positivity only for the coefficients I_j^T. The result does not directly follow for I_h since our oscillation fix is used only to bound the inflow intensity values, but not to modify the intensity approximation itself. For example, we can only prove the following maximum principle*

$$I_h(\mathbf{x}) \leq C_\Phi \|g\|_{L^\infty(\partial\Omega_-)}$$

for $B \equiv 0$. In the case of the linear ansatz space used in the ESC1–method we have $C_\Phi = 1$, but in the case of the ESC2–method $C_\Phi > 1$.
We did not use the linearity of the operator $\mathcal{G}_S$ in the proof of Theorem 12.6; we only required that the estimate (12.20) *holds. Consequently, the results from Theorem 12.6 also hold for the higher order ESC–methods with the oscillation reduction technique as presented in Section 12.3.*
Neither the positivity nor the maximum principle are satisfied by the discontinuous Galerkin method, as we will demonstrate in our numerical tests.

The conditions (12.23) on the ODE solver can be satisfied in most cases by (adaptively) subdividing each characteristic and thus reducing the size of each step. Since in our applications the efficiency of the scheme is essential, this approach is not an option. In the following we summarize some standard results for the Radau IIa method that give some indication as to why this ODE solver is a good choice for our applications.

12.8 Theorem
Let $B \equiv 0$ and $\chi \equiv \chi_0$ for some constant $\chi_0 > 0$ so that we have the following short–characteristic problem (cf. (12.4)*)*

$$v'(s) = -\chi_0 v(s), \quad v(0) = v_0 \ .$$

We denote with $s > 0$ the (fixed) length of the characteristic. Let v_1 denote the approximation of $v(s)$ using the Radau IIa Runge–Kutta solver. It follows that

$$v_1 \leq v_0 \ , \tag{12.30a}$$

$$v_1 \to 0 \quad \textit{for} \quad \chi_0 \to \infty \ . \tag{12.30b}$$

Proof:
The first estimate is a consequence of the *A–stability* of the Radau IIa method; the second estimate holds since the Radau IIa method is *L–stable* (cf. [SW95, Section 6.2]). □

12.9 Remark: *The boundedness result* (12.30a) *for v_1 is also unconditionally satisfied by the Gauß Runge–Kutta method, but it is not satisfied by any explicit scheme for s arbitrary. The estimate* (12.30b) *is important since in the case where v_1 is only bounded, we would need small values of s in regions where χ is large. This estimate is, for example, not satisfied by the Gauß method; in the case of the one–step Gauß method — better known as Crank–Nicholson method — we find $u_1 \to -u_0$ for $\chi \to \infty$ and consequently the solution to the short–characteristic problem can become negative in regions of large χ. We observed this in our tests: although the one–step Gauß method is second order accurate and requires less computational cost than the two–step Radau IIa method, a study of the error to runtime ratio shows that the Gauß method is inefficient for large values of χ. Even the two–step Gauß method leads to a significantly higher error than the two–step Radau IIa method despite the fact that the Gauß method is fourth order accurate, whereas the Radau IIa method is only third order accurate.*

We now study the convergence of the ESC–method under Assumption 12.3.

12.10 Theorem (Convergence of the ESC–method)
Let Assumption 12.3 be satisfied. Denote with I_h the approximation of the solution I to the radiation transport equation (12.14) *using the ESC–method on the triangulation $\mathfrak{T}_h$. Let $V(\Omega)$ be a function space consisting of functions u satisfying: $u_{|T} \in V(T)$ for all $T \in \mathfrak{T}_h$ (where the function spaces $V(T)$ are those given in Assumption 12.3(ii)). If $I \in V(\Omega)$ and if there exists a constant $\chi_0 > 0$ with $\chi \geq \chi_0$ on Ω, then for h small enough the following estimate holds*

$$\|I_h - I\|_{L^\infty(\Omega)} \leq C h^{\min\{\alpha-1, \beta-1, \gamma\}} \tag{12.31}$$

with a constant $C > 0$ not depending on h. The values α, β, and γ are given in Assumption 12.3.

Proof:
Let $T_0 \in \mathfrak{T}_h$ be an element with $\|I_h - I\|_{L^\infty(\Omega)} = \|I_h - I\|_{L^\infty(T_0)}$. Since $I_{|T} \in V(T)$ we can use Assumption 12.3*(ii)* and Lemma 12.5 to estimate

$$\|I_h - I\|_{L^\infty(\Omega)} \leq \|I_h - \Pi_{T_0}(I)\|_{L^\infty(T_0)} + \|\Pi_{T_0}(I) - I\|_{L^\infty(T_0)}$$

$$\leq \max_{\mathbf{x}\in T_0} \sum_{j=1}^{r} |I_j^{T_0} - I(\mathbf{p}_j^{T_0})||\varphi_j^{T_0}(\mathbf{x})| + C_{\text{inter}} h^\gamma$$
$$\leq C_\Phi |I_{j_0}^{T_0} - I(\mathbf{p}_{j_0}^{T_0})| + C_{\text{inter}} h^\gamma \ . \tag{12.32}$$

If $\mathbf{p}_{j_0}^{T_0}$ lies on the inflow boundary $\partial\Omega_-$, we are finished since then $I(\mathbf{p}_{j_0}^{T_0}) = g(\mathbf{p}_{j_0}^{T_0})$ and $I_{j_0}^{T_0} = g(\mathbf{p}_{j_0}^{T_0})$. Therefore let $\mathbf{p}_{j_0}^{T_0} \notin \partial\Omega_-$. By construction the value $I_{j_0}^{T_0}$ is then the approximation to a short–characteristic problem in a point $\mathbf{p}_j^T = \mathbf{p}_{j_0}^{T_0}$. Thus $I_{j_0}^{T_0} = \overline{v_j^T}(s_j^T)$ where $\overline{v_j^T}$ is an approximation to the solution of the ODE (12.4)

$$\frac{d}{ds} v_j^T(s) + \chi(\mathbf{q}_j^T + s\boldsymbol{\mu}) v_j^T(s) = \chi(\mathbf{q}_j^T + s\boldsymbol{\mu}) B(\mathbf{q}_j^T + s\boldsymbol{\mu}) \quad \text{for} \quad 0 < s < s_j^T,$$
$$v_j^T(0) = \mathcal{G}_S(I_h)(\mathbf{q}_j^T)$$

with $\mathbf{q}_j^T \in S$ for $S \in \mathcal{S}_h$ and $S \subset \partial T_-$. The exact solution v_j^T is given by

$$v_j^T(s) = \mathcal{G}_S(I_h)(\mathbf{q}_j^T) e^{-\Delta\tau(0, s_j^T, \boldsymbol{\mu})} + \int_0^{s_j^T} \chi(\mathbf{q}_j^T + s\boldsymbol{\mu}) B(\mathbf{q}_j^T + s\boldsymbol{\mu}) \ e^{-\Delta\tau(s, s_j^T, \boldsymbol{\mu})} ds$$

with $\Delta\tau(a, b, \boldsymbol{\mu}) = \int_a^b \chi(\mathbf{q}_j^T + \sigma\boldsymbol{\mu}) d\sigma$ (cf. (11.2)). Note that I solves the same ODE with initial conditions $I(\mathbf{q}_j^T)$ and therefore

$$I(\mathbf{p}) = I(\mathbf{q}_j^T) e^{-\Delta\tau(0, s_j^T, \boldsymbol{\mu})} + \int_0^{s_j^T} \chi(\mathbf{q}_j^T + s\boldsymbol{\mu}) B(\mathbf{q}_j^T + s\boldsymbol{\mu}) \ e^{-\Delta\tau(s, s_j^T, \boldsymbol{\mu})} ds \ .$$

Using our assumption (12.19) on the ODE solver, it follows that

$$\begin{aligned} |I_{j_0}^{T_0} - I(\mathbf{p}_{j_0}^{T_0})| &\leq |\overline{v_j^T}(s_j^T) - v_j^T(s_j^T)| + |v_j^T(s_j^T) - I(\mathbf{p}_j^T)| \\ &\leq C_{\text{ODE}}(s_j^T)^\alpha + \left|\mathcal{G}_S(I_h)(\mathbf{q}_j^T) e^{-\Delta\tau(0, s_j^T, \boldsymbol{\mu})} - I(\mathbf{q}_j^T) e^{-\Delta\tau(0, s_j^T, \boldsymbol{\mu})}\right| \\ &\leq C_{\text{ODE}}(s_j^T)^\alpha + |\mathcal{G}_S(I_h)(\mathbf{q}_j^T) - I(\mathbf{q}_j^T)| e^{-\chi_0 s_j^T} \ . \end{aligned}$$

The estimate $e^{-\Delta\tau(0, s_j^T, \boldsymbol{\mu})} < e^{-\chi_0 s_j^T}$ follows from $\chi \geq \chi_0$. Next we use the Assumption 12.3*(iv)* on the linear operator $\mathcal{G}_S$

$$\begin{aligned} |I_{j_0}^{T_0} - I(\mathbf{p}_{j_0}^{T_0})| &\leq C_{\text{ODE}}(s_j^T)^\alpha + \Big(|\mathcal{G}_S(I)(\mathbf{q}_j^T) - I(\mathbf{q}_j^T)| + |\mathcal{G}_S(I_h - I)(\mathbf{q}_j^T)|\Big) e^{-\chi_0 s_j^T} \\ &\leq C_{\text{ODE}}(s_j^T)^\alpha + C_{\mathcal{G}} |S|^\beta e^{-\chi_0 s_j^T} + \\ &\quad \max\{|I_h(\mathbf{p}_k^{T'}) - I(\mathbf{p}_k^{T'})| : T' \in \mathcal{T}_h, 1 \leq k \leq r, \mathbf{p}_k^{T'} \in S\} e^{-\chi_0 s_j^T} \end{aligned}$$

Due to our choice of T_0, j_0 the last term is less than or equal to $|I_{j_0}^{T_0} - I(\mathbf{p}_{j_0}^{T_0})|$. As in the final part of the proof of Theorem 12.6 (cf. (12.29)) we conclude that

$$|I_{j_0}^{T_0} - I(\mathbf{p}_{j_0}^{T_0})| \leq \Big(C_{\text{ODE}} \frac{1}{\widetilde{c}_G^\alpha} h^\alpha + C_{\mathcal{G}} \frac{1}{c_G^\beta} h^\beta\Big) \frac{1}{1 - e^{-\chi_0 \widetilde{c}_G h}}$$

$$\leq C_{\mathrm{ODE}} \frac{1}{\widetilde{c}_G^{\alpha}} h^{\alpha-1} + C_{\mathcal{G}} \frac{1}{c_G^{\beta}} h^{\beta-1} \tag{12.33}$$

where we made use of our assumptions on the grid and of Lemma 12.5. The combination of (12.32) and (12.33) concludes the proof. □

12.11 Remark: *In general we cannot expect $\beta > \gamma$ because the polynomial interpolation of the points $\mathbf{p}_j^T$ on S leads to $\beta = \gamma$ since γ measures the interpolation quality on an element T of the polynomial space with the same degree. Since for sufficiently smooth data, we can choose an ODE solver with $\alpha \geq \beta$, and we conclude that the convergence rate of the ESC–scheme is equal to $\beta - 1$. Consequently, an optimal choice for the operator $\mathcal{G}_S$ ($\beta = \gamma$) leads to a convergence rate of $\gamma - 1$; this is an order smaller than the error of the interpolation operator Π_T, but corresponds to the experimental order of convergence found in our numerical tests (cf. Section 12.7.2).*

As already pointed out, Assumption 12.3*(iv)* is satisfied if we choose $\mathcal{G}_S$ to be the piecewise linear interpolation in the points $\mathbf{p}_j^T \in S$. In this case we have $\beta = 2$ so that Theorem 12.10 only leads to a convergence rate of one. If we use the ESC1–method, we have $\mathcal{G}_S(I_h) = I_h$ so that this operator does not lead to a modification of the ESC1–method. Since the one–step Radau IIa method satisfies Assumption 12.3*(iii)*, we conclude from Theorem 12.10 that the ESC1–method as used in our numerical tests converges at least with the order of one. In the following we show that this convergence result is optimal.

12.12 Corollary
Let the assumptions of Theorem 12.10 be satisfied and consider the approximation I_h using the ESC1–method. Then there exists a constant C with

$$\|I_h - I\|_{L^\infty(\Omega)} \leq Ch \ . \tag{12.34}$$

The estimate is optimal in h, i.e., there exist functions χ, B, a vector $\boldsymbol{\mu} \in S^1$, and a family of triangulations $\{T_h\}$ satisfying all the assumptions from Theorem 12.10 so that the ESC1 approximation I_h satisfies

$$\|I_h - I\|_{L^\infty(\Omega)} \geq \bar{C}h \tag{12.35}$$

with some constant $\bar{C} > 0$.

Proof:
The upper bound follows directly from Theorem 12.10 since $\alpha = \beta = 2$ and we can choose $\mathcal{G}_S(u) = u$. To prove the lower bound we consider the setting sketched in Figure 12.10. The propagation angle is $\tan^{-1}(2)$, i.e. $\boldsymbol{\mu} = \frac{1}{\sqrt{5}}(1,2)$; with this choice for $\boldsymbol{\mu}$ only the lower boundary is an inflow boundary. We choose $\chi \equiv 1$ and $B \equiv 0$. Denoting with $g(x)$ the inflow function on the lower boundary, the function I solving the radiation transport equation (11.1c) is easily computed:

$$I(x,y) = g\big(x - \frac{1}{2}y\big) \exp\big(-\frac{\sqrt{5}}{2}y\big) \ .$$

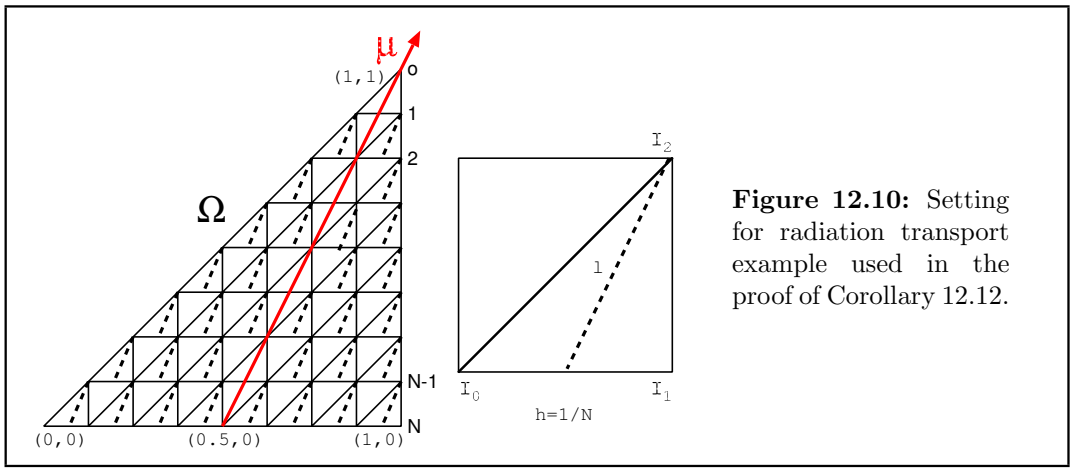


Figure 12.10: Setting for radiation transport example used in the proof of Corollary 12.12.

For the discretization we use a grid as show in Figure 12.10, where the macro grid consists of a single triangle. For the refinement we use a standard quartering strategy. We denote the number of refinement steps with N. It is easy to see that $h = \frac{1}{N}$. The radiation intensity at the nodes of the grid can be computed by starting with the nodes one row N (the lower boundary) and moving up on row at a time. We denote the intensities on row k with I_{kl} with $l = 0, \ldots, k$ and $k = 0, \ldots, N$. The setting for the short–characteristic problem for computing I_{lk} is also sketched in Figure 12.10. The length of the characteristic is $l = \frac{\sqrt{5}}{2N}$, and it intersects the lower edge of the triangle in the middle; consequently the inflow intensity used as the starting value for the short–characteristic problem is $\frac{1}{2}\big(I_{(k+1)(l+1)} + I_{(k+1)l}\big)$. By induction we compute that the approximation at the point $(1,1)$ is given by

$$I_h(1,1) = I_{00} = \Big(\frac{1}{2}e^{-l}\Big)^n \sum_{k=0}^{n} \binom{n}{k} I_{nk}$$

for $1 \leq n \leq N$. At the points on the lower boundary the intensity is given by the inflow function so that $I_{Nk} = g\big(\frac{k}{N}\big)$. Thus we arrive at the following formula for the intensity at the point $(1,1)$:

$$I_h(1,1) = \Big(\frac{1}{2}e^{-l}\Big)^N \sum_{k=0}^{N} \binom{N}{k} g\Big(\frac{k}{N}\Big).$$

If we choose $g(x) = x^2$ then the above expression can be simplified to

$$I_h(1,1) = e^{-Nl}\frac{N+1}{4N} = e^{-\frac{\sqrt{5}}{2}}\frac{N+1}{4N} \ .$$

We now arrive at the following estimate

$$\|I_h - I\|_{L^\infty(\Omega)} \geq |I_h(1,1) - I(1,1)| = \Big|e^{-\frac{\sqrt{5}}{2}}\frac{N+1}{N} - \frac{1}{4}e^{-\frac{\sqrt{5}}{2}}\Big| = e^{-\frac{\sqrt{5}}{2}}\frac{1}{4N} = e^{-\frac{\sqrt{5}}{2}}\frac{h}{4} \ .$$

This concludes the proof. □

Comparing our convergence result for the ESC1–method with the corresponding result for the DG–method with linear ansatz functions (termed DG1, cf. Theorem 11.1), we

find that we lose half a power in h. Compared with the approximation properties of the continuous linear function space we even lose a full power. Furthermore the convergence result for the DG scheme also controls the error of the derivative in the direction $\boldsymbol{\mu}$. On the other hand, the approximation using the ESC1–method is constructed with far fewer degrees of freedom (the DG1–method requires three on each element, the ESC1–method only one per node); this can explain the reduction in the convergence rate. The lack of control of the error in the derivatives of I is a more severe problem. The blow–up observed in the standard Galerkin approach is not a problem in the ESC context since we prove convergence in L^∞. Nevertheless a control of the derivatives is desirable. In the numerical results presented in the following, we show that the derivatives converge with the same order as the approximation itself. This is a surprising result since in most cases the derivatives converge with a lower rate (e.g. Theorem 11.1). Unfortunately we cannot prove this result. The main problem is that we have no control over the derivative of the solution on those elements where no short–characteristic problem is solved, i.e. on those elements with two inflow boundaries. On the other elements a simple argument shows that for the ESC1–method the convergence rate is, in fact, of the order h:

12.13 Corollary
Let the assumptions of Theorem 12.10 be satisfied. Consider an element T with only one inflow boundary and assume that χ, B are continuous on T. Then the following estimate holds

$$\|\boldsymbol{\mu} \cdot \nabla (I_T - I)\|_{L^\infty(T)} \leq Ch \tag{12.36}$$

where I_T denotes the approximation of the intensity on T using the ESC1–method.

Proof:
Let $\mathbf{p}$ be the node of the triangle T opposite the inflow boundary, let $\mathbf{q}$ be the intersection of the characteristic through $\mathbf{p}$ with the inflow boundary, and let s denote the length of the characteristic. Since I_T is linear, $\boldsymbol{\mu} \cdot \nabla I_T$ is constant and therefore

$$\boldsymbol{\mu} \cdot \nabla I_T = \frac{I_T(\mathbf{p}) - I_T(\mathbf{q})}{s} \ .$$

Since we use a backward Euler scheme to compute $I_T(\mathbf{p})$, it follows that

$$\boldsymbol{\mu} \cdot \nabla I_T = \chi(\mathbf{p})(B(\mathbf{p}) - I_T(\mathbf{p})) \ .$$

(This equation is also true up to an $O(s)$ for other ODE solvers. Since $s = O(h)$ the result of the Theorem does not depend on the ODE solver used for the short–characteristic problem.)
Since $I \in H^{2,\infty}(T)$, it follows by the Sobolev embedding theorem that $I \in C^0(T)$. Using the RT equation (11.1c), we conclude $\boldsymbol{\mu} \cdot \nabla I \in C^0(T) \cap H^{1,\infty}(T)$ and

$$\boldsymbol{\mu} \cdot \nabla I(\mathbf{x}) = \boldsymbol{\mu} \cdot \nabla I(\mathbf{p}) + O(h) \ .$$

Again using the RT equation it follows that

$$\boldsymbol{\mu} \cdot \nabla I(\mathbf{x}) = \chi(\mathbf{p})(B(\mathbf{p}) - I(\mathbf{p})) + O(h) \ .$$

Since by Lemma 12.5 $s = O(h)$, we conclude

$$\begin{aligned}\|\boldsymbol{\mu} \cdot \nabla(I_T - I)\|_{L^\infty(T)} &= \sup_{\mathbf{x}\in T} |\boldsymbol{\mu} \cdot \nabla I_T(\mathbf{x}) - \boldsymbol{\mu} \cdot \nabla I(\mathbf{x})| \\ &= |\chi(\mathbf{p})(B(\mathbf{p}) - I_T(\mathbf{p})) - \chi(\mathbf{p})(B(\mathbf{p}) - I(\mathbf{p}))| + O(h) \\ &= |I_T(\mathbf{p}) - I(\mathbf{p})| + O(h) \ .\end{aligned}$$

This proves the desired estimate using Corollary 12.12. □

12.7 Numerical Results

In the numerical tests presented in the following we use the abbreviation ESC1 and ESC2 to denote the linear and quadratic ESC–methods with the Radau IIa ODE solver; — these methods were termed ESC1–IRK and ESC2–IRK in [DV02]. For the ESC1–method we use the one step Radau IIa ODE solver, which is identical to the backwards Euler scheme and satisfies the assumption from Corollary 12.12 (cf. (12.19)). For the ESC2–method we use the two step Radau IIa solver. For the CESC extension we use the same ODE solver as for the corresponding ESC–method and a simple quadrature for the computation of c_T (cf. (12.12)) based on the values in the midpoints of the edges of the triangle T. We add *–fix* to the abbreviation of the schemes to denote the use of the oscillation fix discussed in Section 12.3. With the exception of the results presented in Section 12.7.1, we use the oscillation fix in most of our numerical tests so that we do not alway add the abbreviation.

Our main interest lies in the quality of the approximation of the average radiation source term Q_{rad} (cf (1.19g)). In Chapter 3 we discussed an approximation in two steps, in which first the integral over the unit sphere is approximated by a quadrature rule involving a fixed set of M directions $\{\boldsymbol{\mu}_m\}_{m=1}^M$; in the second step the intensity I_m in each direction $\boldsymbol{\mu}_m$ is approximated for $m = 1, \ldots, M$. We will not study the influence of this quadrature rule, i.e., the influence of the parameter M on the solution. An investigation of this parameter can be found in [BVS99], where the quadrature given in Table 3.1 was found to give satisfactory results. In 3d this quadrature requires the approximation of the RT equation (11.1c) for $M = 24$ different values of $\boldsymbol{\mu}$; in $2d$ we still have $M = 12$.

In [DV02] we studied the ESC–scheme for some fixed directions $\boldsymbol{\mu}$, focusing on the EOC and the error to runtime ratio of the ESC1– and the ESC2–methods using different ODE solvers for the short–characteristic problem. We use some of the same test problems for the following investigations. As computational domain we use $\Omega := [-1, 1]^2$ and as macro grid we use the unstructured triangulation shown on page 85. We use inflow boundary conditions on the lower boundary for direction pointing upwards and on the top boundary for downwards pointing $\boldsymbol{\mu}$. On the vertical boundaries we prescribe periodic boundary conditions for all problems, using the iteration technique described in Section 12.4. Note that due to the periodic boundary conditions the beam in Problem 11.4(SEARCHLIGHT) moves through the domain not only once, as sketched in Figure 11.3(b), but until it reaches the top or the bottom boundary. The solution for Problem 11.3(STAR) must also be modified accordingly. Since grid alignment with the propagation direction $\boldsymbol{\mu}$ of the intensity can strongly influence the approximation

error of the numerical schemes, we compute the maximum of the errors between the exact solutions I_m and the approximations $I_{m,h}$ for all directions $\boldsymbol{\mu}_m$ from Table 3.1:

$$\begin{aligned} \mathrm{err}_h^1 &:= \max\{\|I_m - I_{m,h}\| : m = 1, \ldots, M\} , \\ \mathrm{err}_h^2 &:= \max\{\|\boldsymbol{\mu}_m \cdot \nabla(I_m - I_{m,h})\| : m = 1, \ldots, M\} . \end{aligned} \tag{12.37}$$

In [DV02] we studied the solution in the L^2–norm. Here we choose the L^1–norm in accordance with our analysis and the numerical tests for the MHD solver. The results, however, hardly depend on the norm used [DV02].
We start our investigation in Section 12.7.1 by quantifying the influence of the oscillation fix described in Section 12.3. In Section 12.7.2 we study the experimental order of convergence (EOC) of the ESC– and the CESC–methods. A very important consideration is the error to runtime ratio of the different schemes; this is studied in Section 12.7.3 for the test cases with smooth solutions and in Section 12.7.4 with non–smooth solutions. To allow a better classification of the schemes, we include the first and second order discontinuous Galerkin (DG) schemes introduced in [LR74] as standard reference methods. Since the coupling of the RT solver with a method for solving the MHD equations is the focus of our study, we concentrate on the first and second order versions of the (C)ESC method and of the DG scheme. We conclude this chapter with a model problem from solar physics: in Section 12.8 we compare the different schemes, and in Section 12.9 we discuss some issues concerning the local grid adaptation and local adaptation of the order of the scheme. In the next chapter we then shift our attention to the approximation of the radiation source term Q_{rad}.

12.7.1 Suppressing Oscillations

As already pointed out in Section 12.3, higher order schemes for hyperbolic equations tend to lead to oscillations, especially in the vicinity of discontinuities in the solution. For this reason some limiting process is required to produce stable schemes. In the following we test the method outlined in 12.3 for the ESC– and the CESC–methods. As test problem we use Problem 11.4(Searchlight). In Figure 12.11 the intensity at the top boundary of the domain is shown using $\boldsymbol{\mu} = \boldsymbol{\mu}_1$ (cf. Table 3.1). (Due to the periodic boundary condition the beam always reaches the top boundary. The intensity in the beam is reduced by $e^{-1/\boldsymbol{\mu}_y}$ due to the constant absorption coefficient $\chi = 0.5$.) The exact solution consists of a sharp beam with an intensity value of around 4.98 in the interior; outside the intensity is zero. At the boundaries of the beam the solution is discontinuous. At these boundaries the higher order schemes lead to over- and undershoots and, as a result, to negative intensity values. The CESC1–, the CESC2–, and the ESC2–methods lead to a good resolution of the discontinuity, especially when compared to the results of the ESC1– and the CESC0–methods, which do not produce any oscillations, but lead to a strong smearing of the beam. The aim of a correction mechanism must be to reduce the over– and undershoots, while at the same time maintaining the high resolution of the discontinuity. The corresponding results are shown in Figure 12.12, where we have reduced the range of the x–axis to the region of the beam. As can be clearly seen, the fix reduces the amplitude of the oscillations considerably while barely reducing the resolution of the discontinuity.

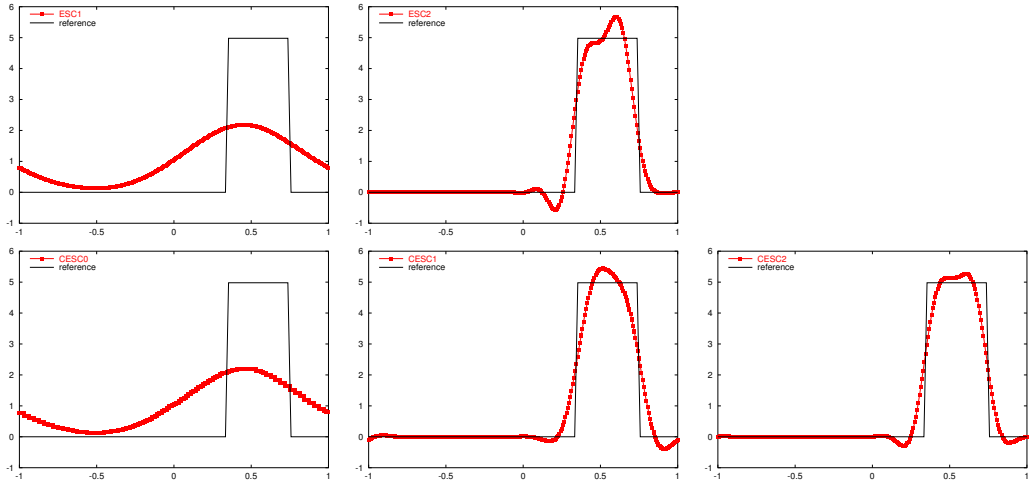


Figure 12.11: Problem: 11.4(SEARCHLIGHT). Top row: ESC–methods, bottom row: CESC–methods. The higher order ESC– and CESC–methods without oscillation fix produce oscillations near discontinuities. In the case of the CESC1–, the ESC2–, and the CESC2–methods over– and undershoots at the discontinuity are visible. The ESC1– and the CESC0–method do not lead to over– and undershoots but to a strong smearing of the discontinuity.

Summary of Section 12.7.1: *The oscillation fix leads to the desired reduction in the size of the oscillations without severely reducing the resolution of the discontinuity. Since in [DV02] we found that even for smooth data the quality of the approximation is barely reduced by the fix, we restrict our attention in the following to the schemes including the fix without always mentioning this in the discussion of the results.*

12.7.2 Convergence Rate of the (C)ESC–Methods

We now study the experimental order of convergence (EOC) (cf. Definition 3.4) of the (C)ESC–methods. To compute the error we use both expressions from (12.37) together with the L^1–norm. We use Problem 11.1(SMOOTH) since both the data and the solution are in C^∞. This is not the typical setting in applications, but it allows us to determine the maximum possible rate of convergence. In Figure 12.13 and Figure 12.14 we plot the error versus the grid size h and the corresponding EOC for Problem 11.1(SMOOTH) with the parameter $\alpha = 5$ and $\alpha = 15$, respectively. Note that the approximation using the CESC0–method is constant on each element and consequently the derivative is zero; therefore we do not include the CESC0–method in the plots of the $\boldsymbol{\mu} \cdot \nabla$–error. In the err^1 norm the conservation fix from Section 12.5 leads to an increase in the convergence rate by about one; the ESC1–method converges with the order one, whereas the CESC1–method converges with an order of about two. The ESC2–method also converges with an order of two, whereas the CESC2–method with an order of almost three. If we study the results for the err^2 norm we see that the order of convergence for the ESC–methods is identical to the rate of convergence observed in the err^1 norm. As already pointed out in Section 12.6, this is surprising since we would normally expect

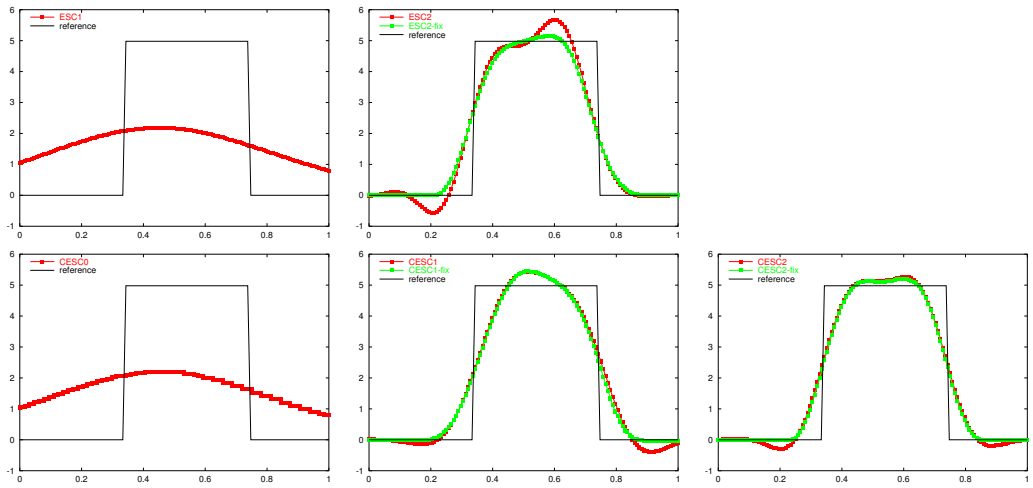
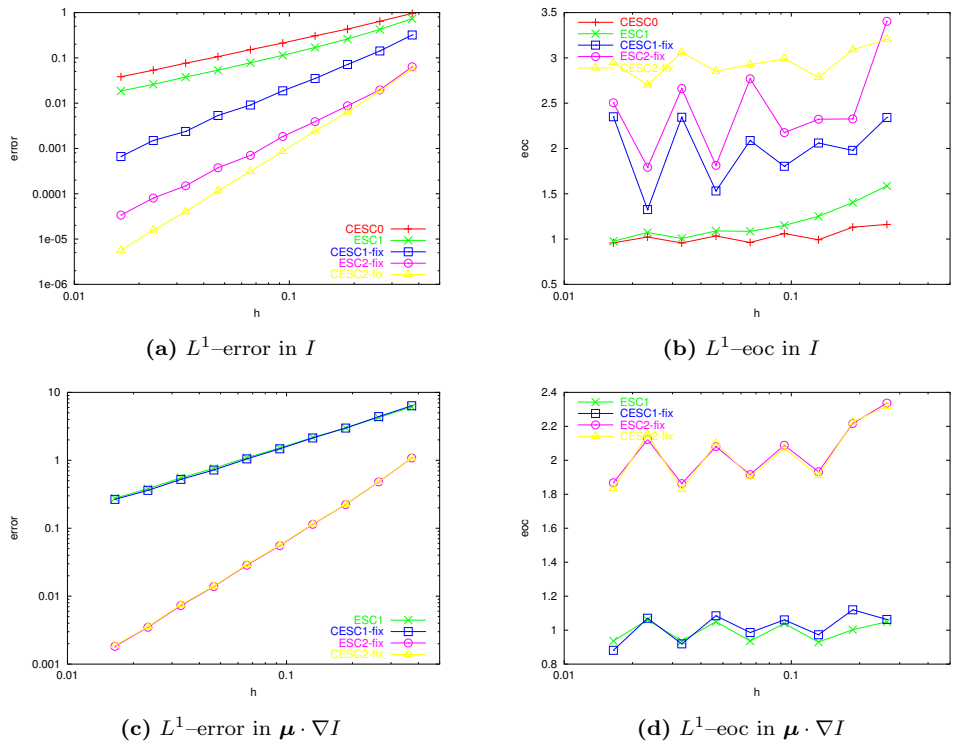


Figure 12.12: Problem: 11.4(SEARCHLIGHT). Top row: ESC–methods, bottom row: CESC–methods. The oscillation fix as outlined in Section 12.3 leads to a considerable reduction in the amplitude of the oscillations; at the same time the resolution of the discontinuities is barely reduced by the fix.

a decrease in the order of convergence, as observed for the CESC–methods. The errors measured in the derivatives of I are, however, about an order of magnitude larger than the errors measured in the intensity itself. Consequently, the increased convergence rate of the CESC–method observed in the err^1 norm leads to an insignificant reduction of the error if we study the sum of err^1 and err^2. Since the conservation fix leads to a higher computational cost, our results indicate that we can only expect an increase in the efficiency due the conservation fix if we study the error in the err^1 norm.

Next we compare schemes with identical order, i.e., the CESC0– with the ESC1–scheme and the CESC1– with the ESC2–scheme. The results in Figure 12.13 and Figure 12.14 show that on a fixed grid the ESC–method is superior to the CESC–method of the same order. For example, in the err^1 norm the ESC2–method produces an error, which is about an order of magnitude smaller than the error produced by the CESC1–method. Since the computational cost of these two methods is not easily compared, this result gives no indication as to which scheme is more efficient. This is studied in the following section.

Summary of Section 12.7.2: *We use Problem 11.2(*H3*) with $\alpha = 5$ to confirm the observations made so far. Note that the solution to this problem is only in $H^{3,\infty}$. In Figure 12.15 we again plot the errors and EOCs for the ESC– and CESC–methods. The convergence rates of the schemes are the same both for this problem and for the previous one; the difference in the errors is also similar. If we measure the error in the intensity I we find that the CESC0– and the ESC1–methods are first order accurate; this is a confirmation of our analytical results from Corollary 12.12. The CESC1– and the ESC2–methods are both second order accurate and the CESC2–method is of third order. If we measure the error in $\boldsymbol{\mu} \cdot \nabla I$, we find that the ESC–methods converge with the same order as in the intensities (cf. Corollary 12.13), the CESC–methods show a reduced rate of convergence.*



(a) L^1–error in I **(b)** L^1–eoc in I

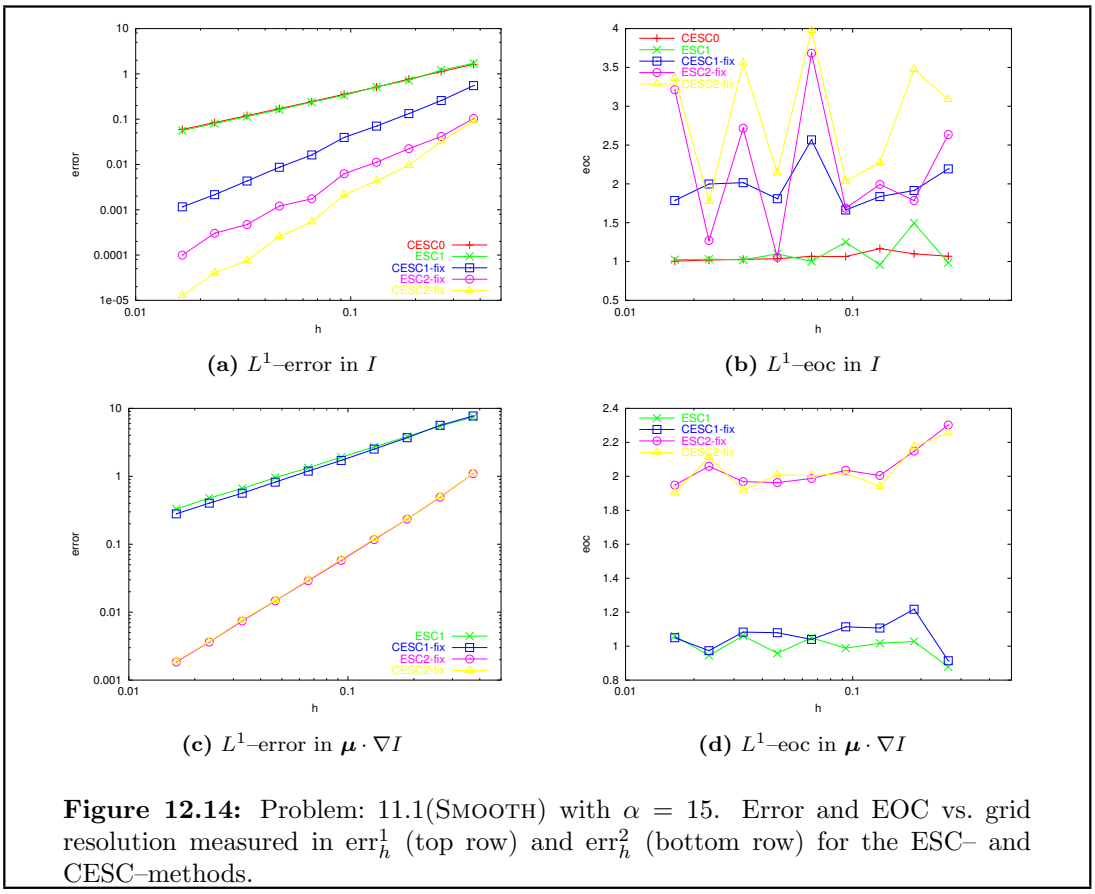
(c) L^1–error in $\boldsymbol{\mu} \cdot \nabla I$ **(d)** L^1–eoc in $\boldsymbol{\mu} \cdot \nabla I$

Figure 12.13: Problem: 11.1(Smooth) with $\alpha = 5$. Error and EOC vs. grid resolution measured in err_h^1 (top row) and err_h^2 (bottom row) for the ESC– and CESC–methods.

12.7.3 Efficiency of the (C)ESC–Methods for Smooth Data

As in the development of the MHD solver described in the previous chapters, our focus must lie on the efficiency of the numerical schemes. Therefore we revisit the problems studied so far, but this time we compare not the approximation errors of the schemes on a fixed grid, but rather the error as a function of the runtime as discussed in Section 3.7.2. We again use the two expressions from (12.37) to compute the error for a given approximation. Since we are interested in first and second order schemes, the CESC2–method is somewhat outside of the scope of our presentation. Nevertheless we include this method here since it is as simple to implement as the ESC2–method itself. We include in our comparison the first and second order discontinuous Galerkin methods (DG0 denotes the method with constant ansatz functions and DG1 the scheme with linear ansatz functions). We already noted in Section 12.5 that the DG0–method is identical to the CESC0–method.

In Figure 12.16 we plot the error versus the runtime for Problem 11.1(Smooth) with the parameter $\alpha = 5$. First we observe that the higher order schemes are more efficient than the first order schemes even on the coarse grids, i.e., the additional cost involved in computing the higher order approximations is not too great compared to the reduction in the error — at least for smooth solutions. For the conservative schemes the order of convergence differs in the err^1 norm and the err^2 norm, but it is the same for the



(a) L^1–error in I **(b)** L^1–eoc in I

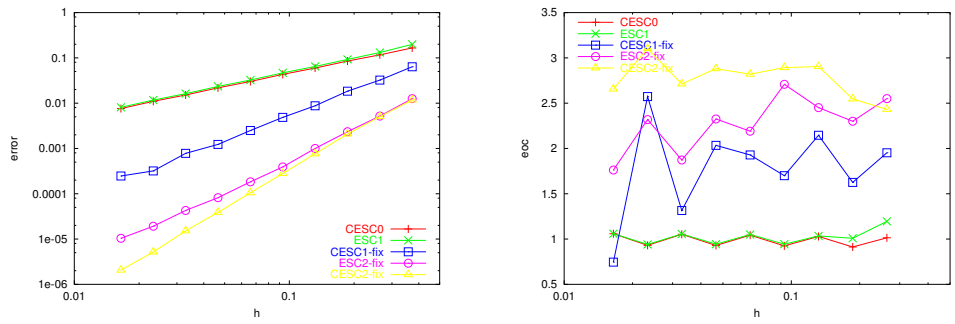
(c) L^1–error in $\boldsymbol{\mu} \cdot \nabla I$ **(d)** L^1–eoc in $\boldsymbol{\mu} \cdot \nabla I$

Figure 12.14: Problem: 11.1(SMOOTH) with $\alpha = 15$. Error and EOC vs. grid resolution measured in err_h^1 (top row) and err_h^2 (bottom row) for the ESC– and CESC–methods.

non–conservative schemes. Consequently, we have to study these two cases separately. We start with Figure 12.16(a), where we plot the err^1 error. Starting with the two first order schemes, we see that the ESC1–method reaches a fixed error in about twenty percent of the time required by the DG0–method. In the case of the ESC2– and the DG1–method, we observe an even greater reduction in the runtime. The second order CESC–method, on the other hand, leads to almost the same result as the DG1–method; both second order conservative schemes are thus comparable. (On a fixed grid the DG1–method leads to a smaller error but requires more runtime.) Finally we see that the third order CESC2–method is more efficient than the second order ESC2–method only for sufficiently high grid resolutions.

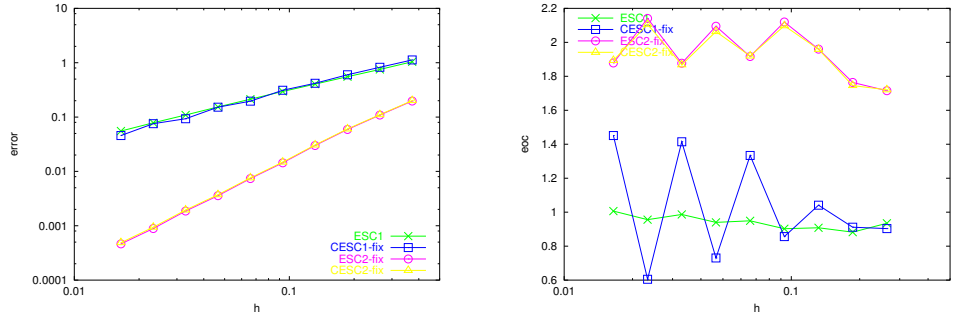
We now turn to Figure 12.16(b). Again the DG1–method and the CESC1–method lead to very similar results. Since both are, however, only first order schemes (in the err^2 norm), they are far less efficient than the ESC2–method, which is still of second order. In fact, they are even less efficient than the ESC1–method. The CESC2–method is also second order and slightly less efficient than the ESC2–method. Note that again we have not included the DG0–method (corresponding to the CESC0–method) since $\nabla I_h \equiv 0$.

The results for Problem 11.1(SMOOTH) with $\alpha = 15$ confirm our observations, although the differences between the schemes is less pronounced than in the previous case. The runtime required by the DG–method is about a factor of four larger than the runtime



(a) L^1–error in I

(b) L^1–eoc in I



(c) L^1–error in $\boldsymbol{\mu} \cdot \nabla I$

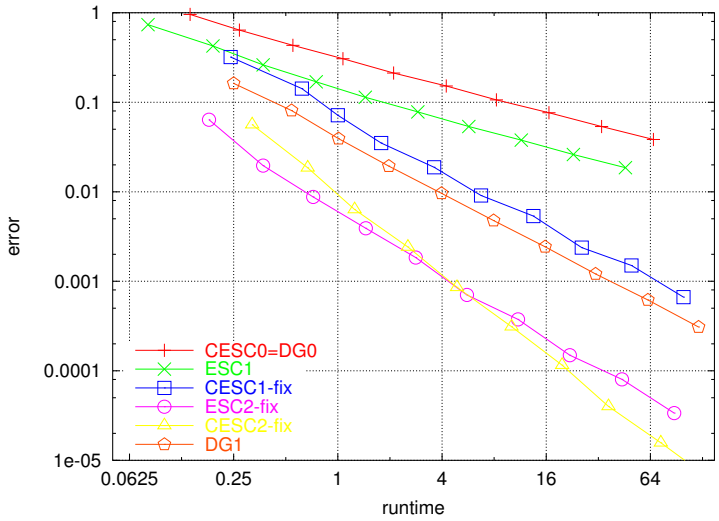
(d) L^1–eoc in $\boldsymbol{\mu} \cdot \nabla I$

Figure 12.15: Problem: 11.2(H3) with $\alpha = 5$. Error and EOC vs. grid resolution measured in err_h^1 (top row) and err_h^2 (bottom row) for the ESC– and CESC–methods.

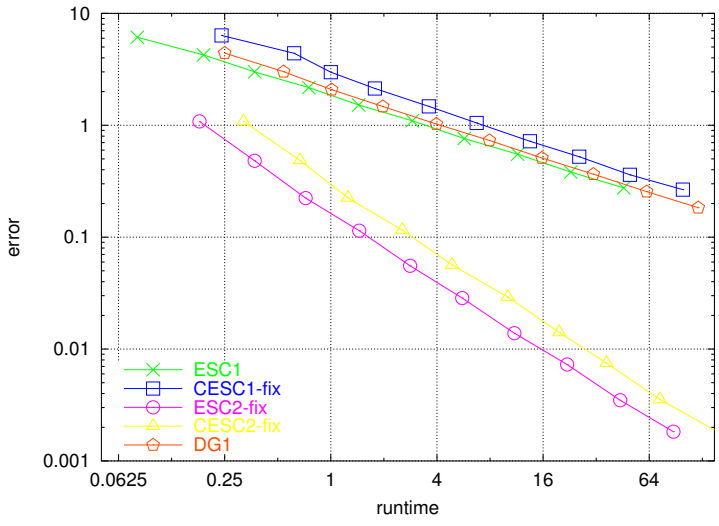
for the ESC–method of the same order. We also see that in this case the DG1–method is clearly more efficient than the CESC1–method.

Summary of Section 12.7.3: *We conclude our study of problems with smooth data with results for Problem 11.2(*H3*) with $\alpha = 5$ (cf. Figure 12.18). The reduction in the smoothness of the solution has very little influence on the results. If we study the error in the intensities, all the observations made above are still true, only the ESC1–method is now slightly less efficient than the DG0–method (cf. Figure 12.18(a)). However if we include the approximation of the derivatives, then the DG0–method does not converge since $\boldsymbol{\mu} \cdot \nabla I_h \equiv 0$, whereas the ESC1–method is still of first order and almost as efficient as the DG1–method. Since the ESC2–method is second order accurate (also in the* err^2 *norm), it is far more efficient than all the other schemes tested.*

If we measure the error in I, then the ESC2–method turns out to be the most efficient (with the exception of the third order CESC2–method). The gain in runtime of the ESC–method compared to the DG–method of the same order is above 75 percent. The conservative schemes of the same order lead to comparable results with the DG–methods faring slightly better. Since the details of the implementation can be varied in many respects, it is possible that a more efficient implementation of the DG–scheme would lead to a more obvious advantage of the DG1–method over the CESC1–method; but the factor of almost ten with respect to the ESC2–method is difficult to obtain by merely modifying the implementation of the scheme.

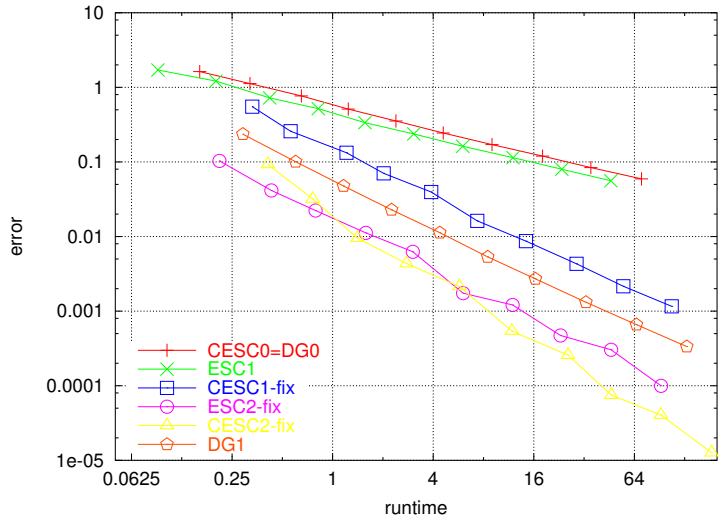


(a) L^1–error in I

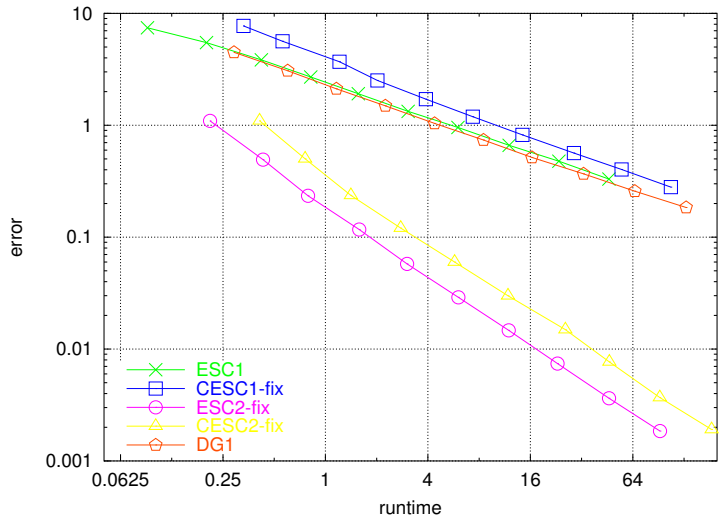


(b) L^1–error in $\boldsymbol{\mu} \cdot \nabla I$

Figure 12.16: Problem: 11.1(SMOOTH) with $\alpha = 5$. Error vs. runtime measured in err_h^1 (top) and err_h^2 (bottom) for the ESC–, CESC–, and DG–methods.

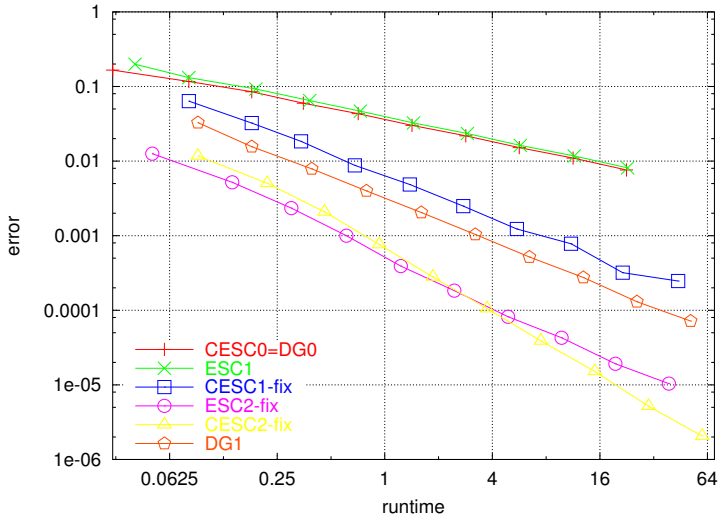


(a) L^1–error in I

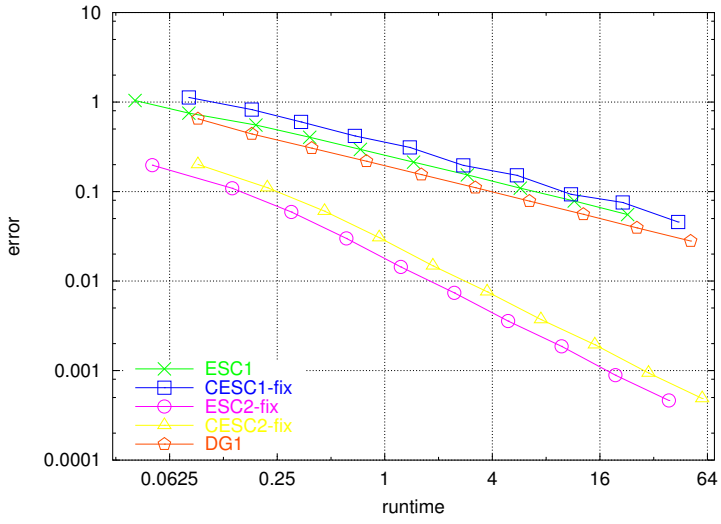


(b) L^1–error in $\boldsymbol{\mu} \cdot \nabla I$

Figure 12.17: Problem: 11.1(SMOOTH) with $\alpha = 15$. Error vs. runtime measured in err_h^1 (top) and err_h^2 (bottom) for the ESC–, CESC–, and DG–methods.

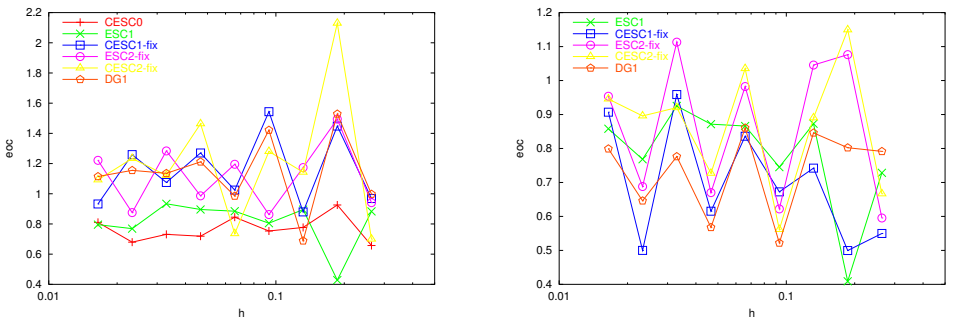


(a) L^1–error in I



(b) L^1–error in $\boldsymbol{\mu} \cdot \nabla I$

Figure 12.18: Problem: 11.2(H3) with $\alpha = 5$. Error vs. runtime measured in err_h^1 (top) and err_h^2 (bottom) for the ESC–, CESC–, and DG–methods.



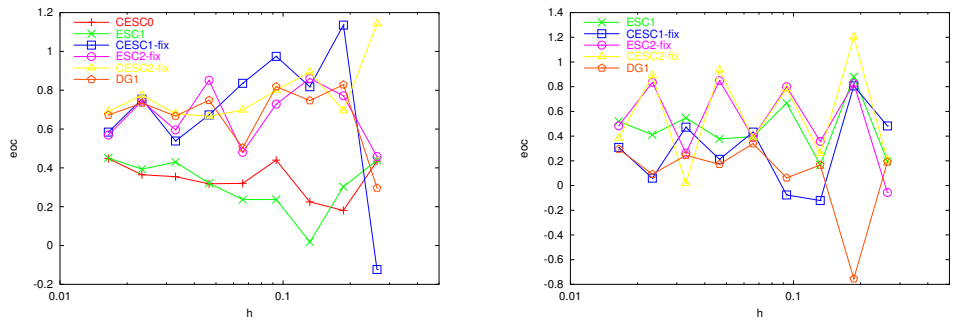
(a) L^1–eoc in I **(b)** L^1–eoc in $\boldsymbol{\mu} \cdot \nabla I$

Figure 12.19: Problem: 11.3(STAR). EOC vs. grid resolution measured in err_h^1 (left) and err_h^2 (right) for the ESC–, CESC–, and DG–methods.

12.7.4 Efficiency of the (C)ESC–Methods for Discontinuous Data

We continue our discussion of the efficiency of the different RT solvers by studying the case of discontinuous data. Note that we use the oscillation fix in all the higher order (C)ESC–methods. We investigate two different settings with discontinuous source term B (Problem 11.3(STAR)) and discontinuous boundary data g (Problem 11.4(SEARCHLIGHT)). Before we study the error to runtime ratio, we plot the EOCs for all six schemes in Figure 12.19 and Figure 12.20. Although the convergence rates are now far from the rates observed for the smooth problems (cf. Section 12.7.2), we can still clearly distinguish the first order schemes (DG0, ESC1) and the higher order schemes (DG1, CESC1, ESC2, CESC2). For Problem 11.3(STAR) we have an EOC of about 1.1 for all higher order methods and an EOC of about 0.8 for the first order methods in the err^1 norm. The EOC is not easy to determine in the err^2 norm (note that $\boldsymbol{\mu} \cdot \nabla I$ is discontinuous); it is about 0.7 for all schemes. For Problem 11.4(SEARCHLIGHT) we find the same general picture, but with an overall lower EOC. Furthermore, as in the case of the smooth solutions studied previously, we see that the DG1–method has an EOC similar to the second order ESC–method in the err^1 norm but an EOC like a first order method in the err^2.

We now study the efficiency of the schemes in the non–smooth settings. The corresponding results are shown in Figures 12.21 and 12.22 for Problem 11.3(STAR) and Problem 11.4(SEARCHLIGHT), respectively. In the err^1 norm the difference in the convergence rate between the two first order schemes and the second order schemes leads to a significant difference in the efficiency of the schemes. On the finest grid the computational cost of the first order schemes is more than sixteen times greater than the cost of the higher order schemes when a fixed error is prescribed. The difference between the two first order schemes and between all the higher order schemes is less significant than in the previous examples. For Problem 11.3(STAR) the DG1–method is the most efficient scheme, whereas for Problem 11.4(SEARCHLIGHT) it is the least efficient higher order scheme. The ESC2–method is very close to being the most efficient scheme in both cases. The situation is altered if we study the error term err^2. The lower convergence rate of the DG1–method now leads to a loss of efficiency. It



(a) L^1–eoc in I **(b)** L^1–eoc in $\boldsymbol{\mu} \cdot \nabla I$

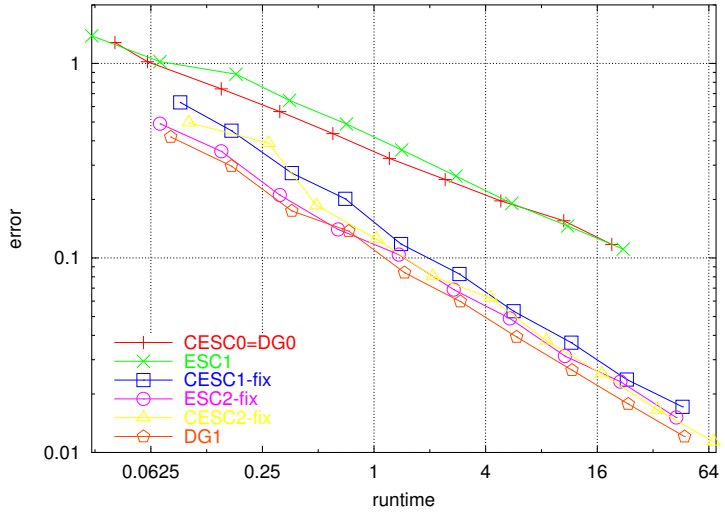
Figure 12.20: Problem: 11.4(SEARCHLIGHT). EOC vs. grid resolution measured in err_h^1 (left) and err_h^2 (right) for the ESC–, CESC–, and DG–methods.

is still far more efficient than the ESC1–method, but loses considerably compared to the ESC2–method, which is the most efficient method for both problems. Especially in the case of Problem 11.4(SEARCHLIGHT), we see that the DG1–method converges more slowly than the ESC–methods. The lower convergence rate is probably due to oscillations in the vicinity of the boundary of the beam. In Figure 12.23 we show the intensity at the top boundary for the ESC2–fix–method and the DG1–method — the corresponding plots for the other schemes can be found in Figure 12.12. Finally we plot a 3d representation of the intensity using the ESC1–, the ESC2–, and the DG1–method in Figure 12.24. The high amount of dissipation in the ESC1–method is clearly visible. Due to the oscillation fix the result using the second order ESC–scheme shows a sharp resolution of the beam without oscillations, whereas some oscillations are visible in the DG1 solution.

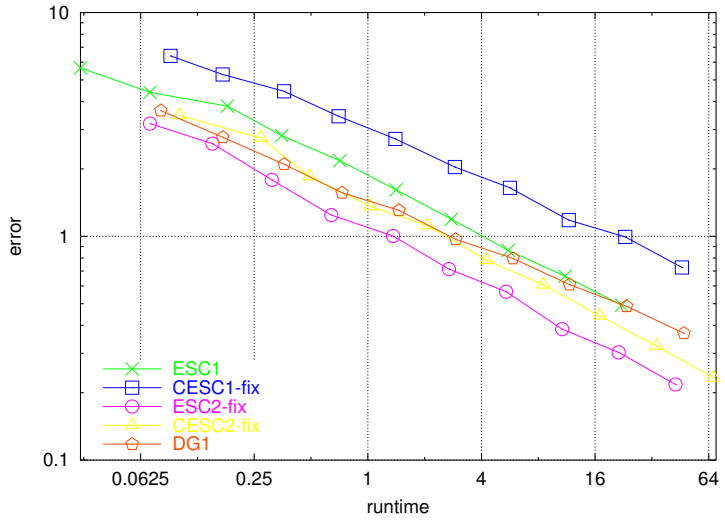
Summary of Section 12.7.4: *First it is important to note that all schemes also converge in the case where the solution is not smooth. The rate of convergence is around* 1 *in the case when* I *is still continuous and is around* 0.5 *when* I *is only piecewise continuous. Although the EOCs of the different schemes are not so far apart as in the case of smooth solutions, the error to runtime ratio of the schemes is, nevertheless, very different. Overall we conclude that in this case, as well, the ESC2–method is (almost) always the most efficient scheme tested; this is especially true if we study the efficiency in the sum* $\mathrm{err}^1 + \mathrm{err}^2$.

12.8 Solar Physical Application: A Magnetic Fluxsheet

Next we study the model problem from solar physics (Problem 11.6(FLUXSHEET)) using an inclination of 80°, i.e. $\boldsymbol{\mu} = (0.17364818, 0.98480775)$. For $0 < z < 0.5$ the jump in χ at the boundary of the fluxsheet leads to a sudden increase in the photon mean free path $l = 1/\chi$ so that we have $l > h$ (where h is the grid spacing). This leads to a significant intensity increase in the vicinity of the sheet boundaries (cf. Figure 12.25). In regions with $l \gg h$ ($z > 0.5$), where there is no significant absorption and emission, the intensity originating from layers with strong χ–discontinuities is transported outwards. This

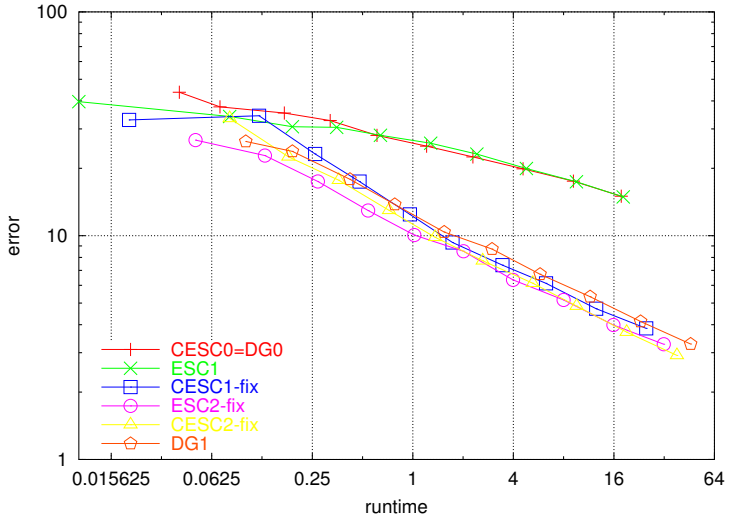


(a) L^1–error in I

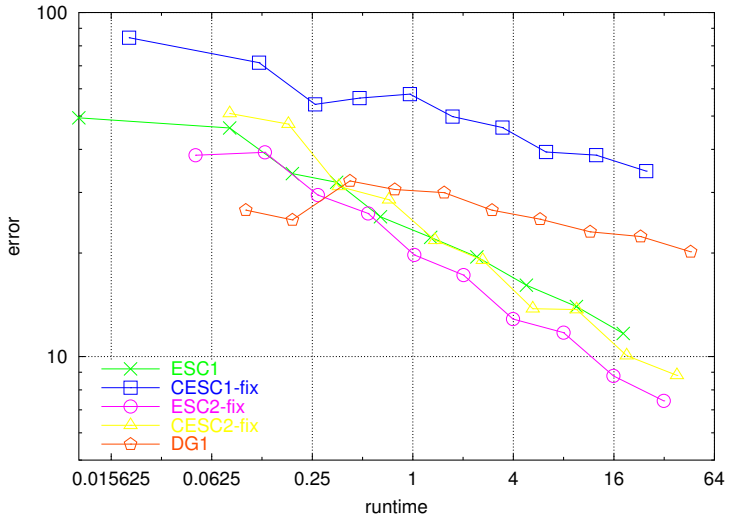


(b) L^1–error in $\boldsymbol{\mu} \cdot \nabla I$

Figure 12.21: Problem: 11.3(STAR). Error vs. runtime measured in err_h^1 (top) and err_h^2 (bottom) for the ESC–, CESC–, and DG–methods.



(a) L^1–error in I



(b) L^1–error in $\boldsymbol{\mu} \cdot \nabla I$

Figure 12.22: Problem: 11.4(SEARCHLIGHT). Error vs. runtime measured in err_h^1 (top) and err_h^2 (bottom) for the ESC–, CESC–, and DG–methods.

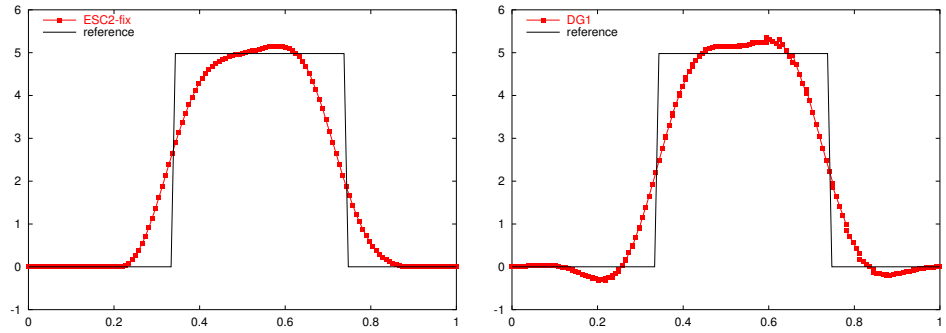
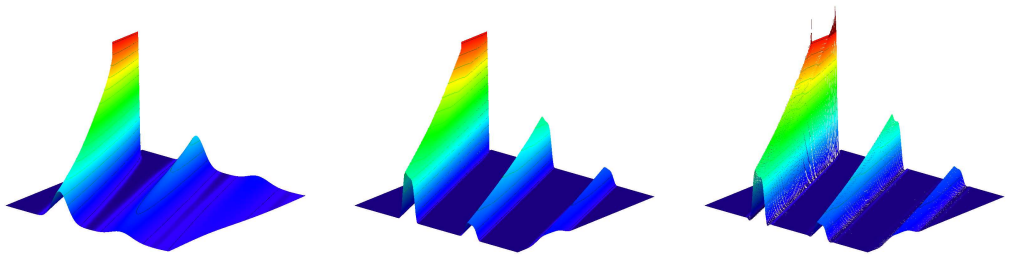


Figure 12.23: Problem: 11.4(SEARCHLIGHT). The intensity on the top boundary of the domain is plotted using the ESC2–method including the oscillation fix (left) and the DG1–method (right). Both schemes lead to overshoots inside the beam. Due to the oscillation fix (cf. Figure 12.11) the ESC2–method shows no oscillations at the base of the beam, whereas the DG1–method leads to small oscillations.



(a) ESC1–method **(b)** ESC2–method **(c)** DG1–method

Figure 12.24: Problem: 11.4(SEARCHLIGHT). Plot of the radiation intensity for the fixed direction $\boldsymbol{\mu}_1$. Dark blue represents zero intensity, which should be assumed outside of the beam. Clearly visible is the high crosswind diffusion in the ESC1–scheme. Both the ESC2– and the DG1–scheme produce a sharp transition between the interior and the exterior of the beam. By using the oscillation fix, over– and undershoots, which are visible in the DG1 solution, are suppressed in the ESC2–scheme.

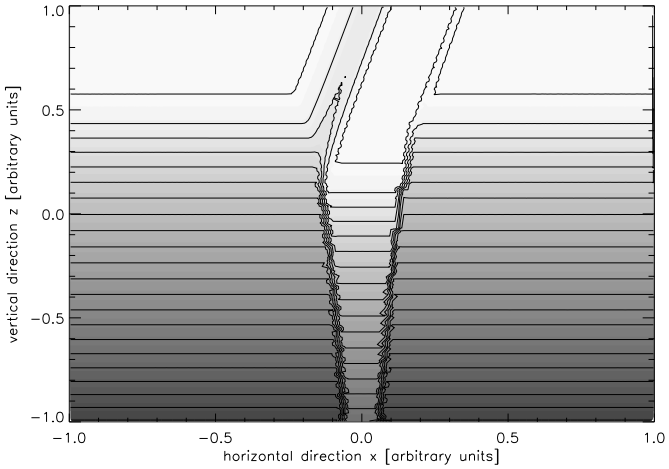


Figure 12.25: The intensity for $\boldsymbol{\mu} = (0.17364818, 0.98480775)$ is shown as a combination of a contour and a gray–scale representation. The intensity increases with increasing geometrical depth z. There are two intensity maxima detectable at the top boundary, originating from atmospheric regions around $(0 \leq z \leq 0.5)$ caused by the sudden decrease in χ in this region (cf. Figure 11.5).

intensity is detectable at the upper domain boundary because the intensity emitted in warmer layers of the atmosphere is transported without being significantly attenuated. Figure 12.26 shows the intensities at the upper domain boundary for different grid resolutions. We see that the approximations converge towards the reference solution (computed with ESC2–method for 10^6 grid cells). The results obtained with the ESC2–method and the DG1–method are similar and are very close to the reference solution on only 33770 grid cells. The ESC1–method clearly shows a poorer resolution. In Figure 12.27 we show the intensity at a height of $z = 0$. In this region the DG1–method leads to oscillations in the vicinity of sheet boundaries. These oscillations are not detectable at the top boundary, but obviously affect the radiation source term in the lower regions.

Summary of Section 12.8: *Again we find that the ESC2–method leads to a good approximation of the radiation intensity without producing oscillations. The DG1–method leads to a similar approximation at the top boundary, but produces oscillations in layers below the solar surface. A more detailed study can be found in [DV02].*

12.9 Local Adaptivity

In the previous sections we saw that a higher order scheme is essential for an adequate resolution of the important structures in the solution. On the other hand, a second order scheme requires much more computational effort than a first order method. In the case of the ESC–scheme, the complexity can be measured by the number of characteristics for which the intensity has to be computed. For the ESC1–method this number is approximately equal to $\frac{1}{2}N$, where N is the number of grid elements. In the case of the

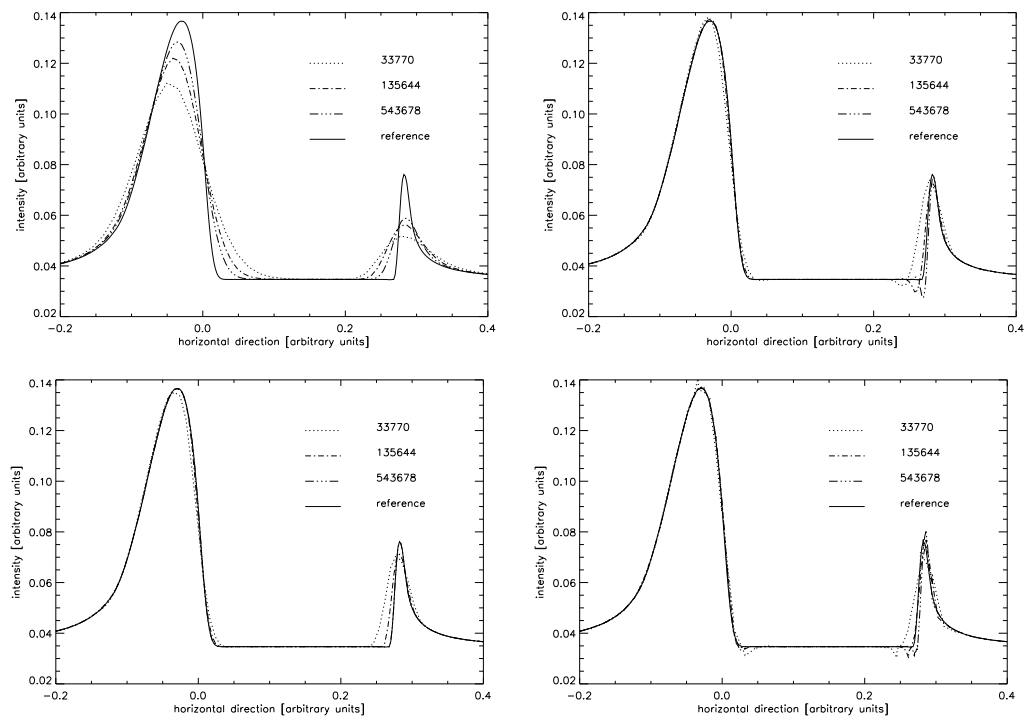


Figure 12.26: The computed intensities at the top domain boundary. Upper left and upper right: ESC1, ESC2. Lower left and Lower right: ESC2–fix, DG1. Results for various numbers of grid cells (33770, 42478, 135644, 543678) are shown. The reference solution is obtained with ESC2–fix and 10^6 grid cells.

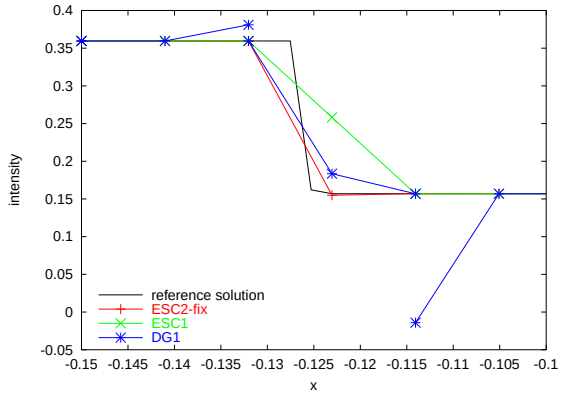


Figure 12.27: Intensity at height $z = 0$ computed on a grid with 33770 elements using the ESC1, ESC2–fix, and the DG1 method; the reference solution is computed using the ESC2–fix method on a grid with 10^6 elements. Only the region around the left boundary of the fluxsheet is shown. The DG1–method produces oscillations and the ESC1–method a high amount of cross-wind diffusion.

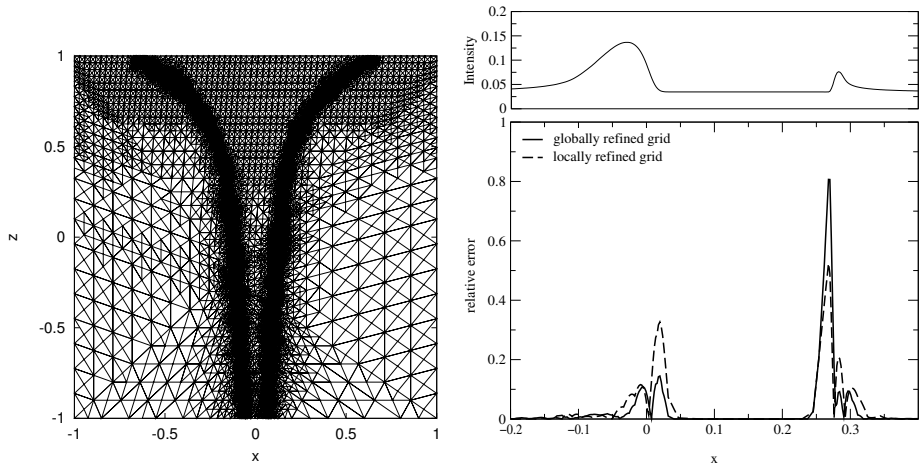


Figure 12.28: Test of grid adaption based on relative jumps in the density for the fluxsheet problem. Left: Grid. Right top: Intensity on upper boundary. Right bottom: Relative error on the top boundary between a reference solution calculated with 543678 elements and the solution on the locally adapted grid. As a comparison, the relative error computed for a solution on a regular grid with approximately the same number of elements is also plotted. Although both computations require the same amount of cpu time, the solution on the locally adapted grid has a larger error than the solution on the globally refined grid.

ESC2–method, we can approximate the complexity by $\frac{1}{2}N + \frac{3}{2}N = 2N$ and therefore have a four fold increase in the computational cost (cf. Section 12.2.2). In the following we sketch two possibilities of reducing the computational cost:

- **h–Adaptivity:** The number of elements is reduced in those parts of the domain where a high resolution of the grid is not required. In our MHD code local adaptivity is achieved by evaluating relative jumps of the hydrodynamic quantities over element edges (cf. Section 3.5). In the case of the solar magnetic fluxsheet considered in the previous section, this will not lead to an adequate adaption for the radiation transport problem (cf. Figure 12.28), since the upper regions are not sufficiently refined to allow for a good resolution of the intensity peak at the top boundary (cf. Figure 12.26). By including the intensity gradients or a similar indicator based on the values of the intensities, this problem could be avoided.

- **p–Adaptivity:** In the derivation of all the RT schemes discussed here we can choose a different function space $P(T)$ on each triangle T. This space can be chosen according to the complexity of the problem on the element T. During the calculation of the approximation on T we have to determine which order of the scheme should be used. This can be done without a recalculation of the approximation on the other elements.

  In a simple two step approach a solution $I_T^1 \in P^1(T)$ is computed and some error indicator $\mathcal{E}(I_T^1)$ is evaluated. If this indicator is large, I_T^1 is discarded and a new solution $I_T^2 \in P^2(T)$ is computed. In the case of the combination DG0/DG1

this approach is very expensive; the values computed for I_T^1 cannot be used in the computation of I_T^2. If ESC1/ESC2 is used, the coefficients calculated for I_T^1 can be reused for the representation of the solution I_T^2, since they approximate the intensity at the vertices, which are also used in the ESC2 approximation (cf. Figure 12.5).

12.14 Remark: *Both adaptation techniques described above estimate only the error produced locally on the element T. The transported error (i.e. the error produced upwind from T) cannot be estimated in this way. To quantify this part of the error, a rigorous a–posteriori error analysis for the ESC–scheme would be necessary. Such error estimators have been constructed for finite element schemes applied to the RT equation (e.g. [FK97, Sül98, HSS00]). For the ESC–method no suitable estimate is yet available.*

In the following we focus on an implementation of a p–adaptive algorithm using a combination of the ESC1– and the ESC2–method. The question is how to choose the local error indicator $\mathcal{E}_T$. Our aim is to construct a p–adaptive scheme using the lower order ESC1–method as often as possible, while at the same time maintaining the second order convergence rate of the ESC2–method, i.e., the approximation I_h should satisfy

$$\|I - I_h\|_{L^2(\Omega)} \leq Ch^2.$$

In the following we motivate the residuum based indicator used here: since the error generated by an insufficient approximation of the transport term $\boldsymbol{\mu} \cdot \nabla$ grows for diminishing χ, we switch from the radiation transport equation (11.1c) to the equivalent formulation

$$\begin{aligned} \tfrac{1}{\chi}\boldsymbol{\mu} \cdot \nabla I + I - B &= 0 \quad \text{in} \quad \Omega, \\ I &= g \quad \text{on} \quad \partial\Omega_-. \end{aligned}$$

The indicator that is presented here is based on the assumption that the error in the L^2–norm can be controlled by the L^2–norm of the residual:

12.15 Assumption
There exists $q \in \mathbb{R}$ with

$$\|I - I_h\|_{L^2(\Omega)} \leq Ch^q \|R(I_h)\|_{L^2(\Omega)} + M_h$$

where the residual is given by

$$R(I_h) := \frac{1}{\chi}\boldsymbol{\mu} \cdot \nabla I_h + I_h - B.$$

M_h describes the approximation error of the data and is not taken into account in the following, i.e. we set $M_h = 0$.

12.16 Remark: *The above assumption does not hold in this generality. For some finite element approximations of equation (11.1c) such a bound with $q = 1$ has been shown using the H^{-1}–norm for the error [Sül98]. Our tests indicate that $q = 1$ is a good choice in this case, as well.*

Taking our assumption into account, we have to control the norm of the local residual $\|R_T(I_T)\|_{L^2(\Omega)} := \|\frac{1}{\chi}\boldsymbol{\mu}\cdot\nabla I_T + I_T - B\|_{L^2(\Omega)}$ since we have

$$\begin{aligned}\|I - I_h\|^2_{L^2(\Omega)} &\leq C^2h^{2q}\|R(I_h)\|^2_{L^2(\Omega)} = C^2h^{2q}\int\limits_{\Omega}\left(\frac{1}{\chi}\boldsymbol{\mu}\cdot\nabla I_h + I_h - B\right)^2 \\ &= C^2h^{2q}\sum_{T\in\mathfrak{T}_h}\|R_T(I_T)\|^2_{L^2(T)} \leq C^2h^{2q}N\max_{T\in\mathfrak{T}_h}\|R_T(I_T)\|^2_{L^2(T)}.\end{aligned}$$

The number of elements N in the triangulation is of the same order as the area $|T|$ of the elements in grid $\mathfrak{T}$. Therefore, we see that if the solution I_T on each triangle satisfies $\|R_T(I_T)\|_{L^2(T)} \leq Ch^{2-(q-1)}$, our method is second order accurate. We therefore choose

$$\mathcal{E}_T(I_T) := h^{q-1}\frac{\|R_T(I_T)\|_{L^2(T)}}{|T|}\ . \tag{12.38}$$

We have tested this indicator using $q = 0, \frac{1}{2}, 1$ and achieved the best results with $q = 1$. We approximate the integral in (12.38) by a midpoint rule:

$$\mathcal{E}_T(I_T) := \left|\frac{1}{\chi(\boldsymbol{\omega}_T)}\boldsymbol{\mu}\cdot\nabla I_T(\boldsymbol{\omega}_T) + I_T(\boldsymbol{\omega}_T) - B(\boldsymbol{\omega}_T)\right|, \tag{12.39}$$

where $\boldsymbol{\omega}_T$ is the barycenter of the triangle T. To reduce as much as possible the computational cost of the indicator, we approximate the derivative $\boldsymbol{\mu}\cdot\nabla I_T(\boldsymbol{\omega}_T)$ in (12.39). Using the notation introduced in Section 12.2 (cf. Figure 12.6), an approximation is given by:

$$\boldsymbol{\mu}\cdot\nabla I_T(\boldsymbol{\omega}_T) \approx \begin{cases} \frac{\frac{1}{2}(I_a+I_b)-I_c}{\sqrt{|T|}} & \text{for the two–inflow case (Figure 12.6(middle)),} \\ \frac{I_c-I_f}{\sqrt{|T|}} & \text{for the one–inflow case (Figure 12.6(left)).}\end{cases} \tag{12.40}$$

In Figure 12.29 we compare the results of the ESC2–method with the results achieved by the combination of the ESC1/ESC2–method. We use the indicator (12.39) together with (12.40) and switch to the ESC2–method on those triangles where $\mathcal{E}_T(I_T) > 1$. In Table 12.2 we compare the p–adaptive method using different values for the tolerance. In the case shown in Figure 12.29 we gain up to 30% in computational time, and the difference in L^2 between this solution and the solution obtained with the ESC2–method on the same grid is only 10^{-4}. Therefore we conclude that the ESC1/ESC2 p–adaptive method leads to a considerable decrease in computational time, and the deviation from the ESC2 solution is orders of magnitude smaller than the difference between the ESC2–method and the ESC1–method.

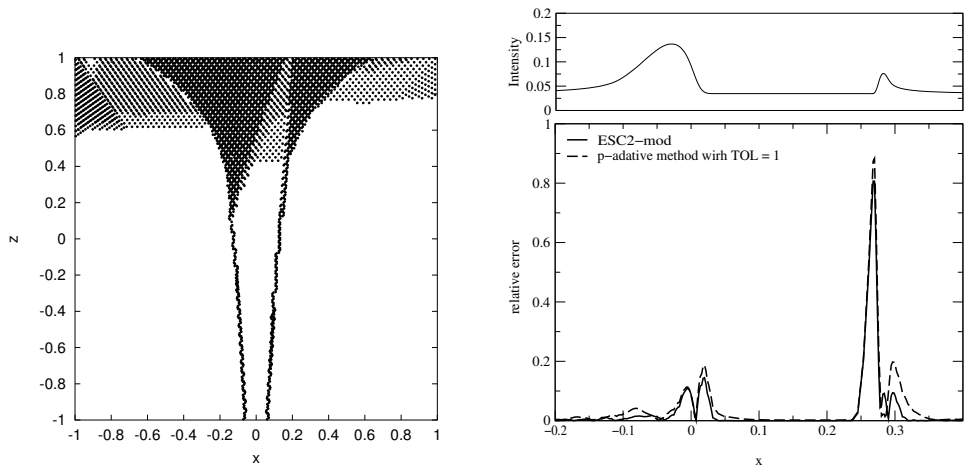


Figure 12.29: Left: Indicator (dark=ESC2,light=ESC1). Right top: Intensity on upper boundary. Right bottom: Relative error between a reference solution computed with the ESC2–method on 543678 elements and the solutions using the p–adaptive method (TOL=1) and the ESC2–method, both computed on 33770 elements.

Fluxsheet: L^2–error						
	elements = 33770			**elements = 543678**		
TOL	%ESC2*	%time**	error***	%ESC2*	%time**	error***
Adaptive schemes						
$\infty^{\S}$	0	67.3	1.43e-02	0	68.8	7.04e-03
4	8	37.9	3.57e-03	10	36.8	5.55e-04
2	8	38.8	1.44e-03	16	33.7	4.50e-04
1§§	**11**	**37.0**	**1.40e-03**	**24**	**29.0**	**2.90e-04**
0.5	17	32.6	1.26e-03	27	27.5	1.63e-04
0.25	25	27.8	7.98e-04	29	24.8	4.51e-05
0.125	28	27.3	6.84e-04	31	25.0	3.54e-05

* Percentage of the total elements on which the higher order method was used.
** Percentage of time gained with respect to the reference method using the same grid.
*** Deviation compared to the solution obtained with the reference method (ESC2–fix) on the same grid meassured with the L^2–norm.
§ only ESC1–scheme
§§ Tolerance used in Figure 12.29.

Table 12.2: The p–adaptive method vs. the other methods with respect to computational time and accuracy measured in L^2.

Chapter 13

Approximating the Radiation Source Term

To include the energy transport through radiation in our finite–volume scheme, the average radiation source term Q_{rad} has to be approximated (cf. Section 3.1). Using the average radiation intensity

$$J(\mathbf{x},t) = \frac{1}{4\pi} \int\limits_{S^2} I(\mathbf{x},t,\boldsymbol{\mu}) d\boldsymbol{\mu}$$

we have to find an approximation of

$$\overline{Q^J_{\text{rad}}}(t) := \frac{1}{|T|} \int\limits_T Q_{\text{rad}}(\mathbf{x},t) d\mathbf{x} = \frac{1}{|T|} \int\limits_T 4\pi\chi(\mathbf{x},t) \left(J(\mathbf{x},t) - B(\mathbf{x},t)\right) d\mathbf{x} \tag{13.1}$$

for each element T of a given grid $\mathcal{T}$. By means of a quadrature rule for the integral over the unit sphere defining J, we reduce the computation of the radiation source term to a summation involving the radiation intensities $I_m(\mathbf{x},t) := I(\mathbf{x},t,\boldsymbol{\mu}_m)$ for a fixed set of directions $\{\boldsymbol{\mu}_m\}_{m=1}^M$. In the previous chapter we studied different schemes for computing an approximation $I_{m,h}$ of the solution I_m to the radiation transport equation (11.1c) in a fixed direction $\boldsymbol{\mu}_m$. These approximations lead to the following approximation for $\overline{Q^J_{\text{rad}}}$ on an element $T \in \mathcal{T}$:

$$\overline{Q^J_{\text{rad}h}}(t) := \frac{1}{|T|} \int\limits_T \chi(\mathbf{x},t) \sum_{m=1}^M \omega_m \left(I_{m,h}(\mathbf{x},t) - B(\mathbf{x},t)\right) d\mathbf{x} \ . \tag{13.2}$$

The constants ω_m are the weights of the quadrature rule satisfying $\sum_{m=1}^M \omega_m = 4\pi$.
A second possible approximation can be derived by using the radiation flux

$$F(\mathbf{x},t) = \int\limits_{S^2} \boldsymbol{\mu} \cdot \nabla I(\mathbf{x},t,\boldsymbol{\mu}) d\boldsymbol{\mu}$$

since we have (using the RT equation (11.1c) and Gauß' theorem)

$$\overline{Q^J_{\text{rad}}} = -\frac{1}{|T|} \int\limits_T F(\mathbf{x},t) d\mathbf{x} = -\frac{1}{|T|} \int\limits_{\partial T} \int\limits_{S^2} I(\mathbf{x},t,\boldsymbol{\mu}) \boldsymbol{\mu} \cdot \mathbf{n} =: \overline{Q^F_{\text{rad}}} \ .$$

The approximation steps outlined above lead to the following expression

$$\overline{Q^F_{\mathrm{rad}h}}(t) := \frac{1}{|T|}\int\limits_{\partial T}\sum_{m=1}^{M}\omega_m I_{m,h}(\mathbf{x},t)\boldsymbol{\mu}_m\cdot\mathbf{n}\ . \tag{13.3}$$

Depending on the scheme used for computing $I_{m,h}$, the approximations (13.2) and (13.3) might not lead to the same result for a fixed grid $\mathcal{T}$.
Since the approximate intensity functions $I_{m,h}$ can be discontinuous over element boundaries, we have to clarify how we compute the boundary integrals in (13.3) defining $\overline{Q^F_{\mathrm{rad}h}}$ on an element $T = T_i$. Let S_{ij} be a face of the grid and denote with $I_{m,i}$ the discrete intensity defined on the element T_i and with $I_{m,j}$ the intensity defined on the neighboring element T_j. We define the value of the integral $\int_{S_{ij}} I_{m,h}$ required to compute (13.3) by using the intensity in downwind direction, i.e.

$$\int_{S_{ij}} I_{m,h} = \begin{cases}\int_{S_{ij}} I_{m,j} & \text{if } \mathbf{n}_{ij}\cdot\boldsymbol{\mu} < 0\ ,\\ \int_{S_{ij}} I_{m,i} & \text{otherwise}\ .\end{cases}$$

In the following all boundary integrals involving the discrete intensity function are to be understood in this sense. For the element integrals in (13.2) we use only values defined on the element itself.
During the derivation of the conservation fix for the ESC–method in Section 12.5 we encountered the term $Q^J_{\mathrm{rad}T,\boldsymbol{\mu}}, Q^F_{\mathrm{rad}T,\boldsymbol{\mu}}$:

$$Q^J_{\mathrm{rad}T,\boldsymbol{\mu}} = \int_T \chi_T(I_{m,h} - B_T)\ , \quad Q^F_{\mathrm{rad}T,\boldsymbol{\mu}} = \int_{\partial T_+} I_{m,h}\boldsymbol{\mu}\cdot\mathbf{n} + \int_{\partial T_-} I_{m,h}\boldsymbol{\mu}\cdot\mathbf{n}$$

(cf. (12.12)). These expressions can be used for approximating the radiation source terms since, by summation with respect to m, we arrive at the expressions (13.2) and (13.3), respectively.

13.1 Remark: *The two terms $Q^J_{\mathrm{rad}T,\boldsymbol{\mu}}$ and $Q^F_{\mathrm{rad}T,\boldsymbol{\mu}}$ differ only in the case of a non–conservative scheme. In the case of the DG–methods and the CESC–methods both expressions are identical and can differ only after the integrals have been approximated by quadrature rules. However, this difference is negligible so that we only have to study the difference between the approximations $\overline{Q^J_{\mathrm{rad}h}}$ and $\overline{Q^F_{\mathrm{rad}h}}$ in the case of the non–conservative ESC–methods.*

The data χ, B defining the radiation field are functions of the hydrodynamic temperature θ and the density ρ. In our finite–volume scheme we require the radiation source term Q_{rad} for the update of the total energy density only for fixed grid functions θ, ρ, so that the time dependency of these functions is not relevant for studying the approximation of Q_{rad}. Consequently, we drop the time variable t in the following and assume, as in the previous chapter, that θ, ρ (and therefore χ, B) are given functions of the space variable $\mathbf{x}$.
In Section 12.7 we compared different schemes for approximating the radiation intensity I by studying the errors in I (err^1_h) and in $\boldsymbol{\mu}\cdot\nabla I$ (err^1_h) (cf. (12.37)). In this chapter

we measure the quality of the approximation to $\overline{Q_{\text{rad}}}$ using the L^1–norm:

$$\text{err}_h^J := \sum_{T\in\mathcal{T}} |T| \left| \overline{Q_{\text{rad}}^J} - \overline{Q_{\text{rad}\,h}^J} \right| , \quad \text{err}_h^F := \sum_{T\in\mathcal{T}} |T| \left| \overline{Q_{\text{rad}}^F} - \overline{Q_{\text{rad}\,h}^F} \right| . \tag{13.4}$$

Using the semi–discrete approximations

$$\overline{Q_{\text{rad}\,M}^J} := \frac{1}{|T|} \int_T \chi(\mathbf{x}) \sum_{m=1}^M \omega_m \left(I_m(\mathbf{x}) - B(\mathbf{x}) \right) d\mathbf{x} ,$$

$$\overline{Q_{\text{rad}\,M}^J} := -\frac{1}{|T|} \int_{\partial T} \sum_{m=1}^M \omega_m I_m(\mathbf{x}) \boldsymbol{\mu}_m \cdot \mathbf{n}$$

to $\overline{Q_{\text{rad}}^J}$ and $\overline{Q_{\text{rad}}^F}$, respectively, we can decompose the expressions in (13.4) into two parts

$$\text{err}_h^J \le \sum_{T\in\mathcal{T}} |T| \left| \overline{Q_{\text{rad}}^J} - \overline{Q_{\text{rad}\,M}^J} \right| + \sum_{T\in\mathcal{T}} |T| \left| \overline{Q_{\text{rad}\,M}^J} - \overline{Q_{\text{rad}\,h}^J} \right| = E_1^J + E_2^J .$$

In the same way we can derive error terms E_1^F and E_2^F. The error E_1^J (E_1^F) measures the quality of the quadrature rule for the unit sphere S^2. The second term E_2^J (E_2^F) measures the quality of the spatial approximation. The error terms err_h^1 and err_h^2 used in the previous chapter can be interpreted as upper bounds for E_2^J and E_2^F, respectively:

$$\begin{aligned}
E_2^J &= \sum_{T\in\mathcal{T}} |T| \left| \frac{1}{|T|} \int_T \chi(\mathbf{x}) \sum_{m=1}^M \omega_m \left(I_m(\mathbf{x}) - B(\mathbf{x}) \right) - \omega_m \left(I_{m,h}(\mathbf{x}) - B(\mathbf{x}) \right) d\mathbf{x} \right| \\
&\le \|\chi\|_{L^\infty(\Omega)} \sum_{m=1}^M \omega_m \|I_m - I_{m,h}\|_{L^1(T)} \le 4\pi \|\chi\|_{L^\infty(\Omega)} \, \text{err}_h^1 , \\
E_2^F &= \sum_{T\in\mathcal{T}} |T| \left| \frac{1}{|T|} \int_{\partial T} \sum_{m=1}^M \omega_m \big(I_m(\mathbf{x}) - I_{m,h}(\mathbf{x}) \big) \boldsymbol{\mu}_m \cdot \mathbf{n} \right| \\
&= \sum_{T\in\mathcal{T}} \left| \int_T \sum_{m=1}^M \omega_m \boldsymbol{\mu}_m \cdot \nabla \big(I_m(\mathbf{x}) - I_{m,h}(\mathbf{x}) \big) d\mathbf{x} \right| \\
&\le \sum_{T\in\mathcal{T}} \sum_{m=1}^M \omega_m \|\boldsymbol{\mu}_m \cdot \nabla (I_m - I_{m,h})\|_{L^1(T)} \le 4\pi \, \text{err}_h^2 .
\end{aligned}$$

Therefore the results from the last section also give some indication of the quality of the different numerical schemes for approximating Q_{rad}. The estimates err_h^1 and err_h^2 are, however, only very crude. For example, if $\|\chi\|_{L^\infty(\Omega)}$ is very large then err_h^1 can be a bad estimate for the actual error err_h^J. Therefore the separate study of the approximation quality for the radiation source term Q_{rad} is important.
In Section 13.1 we begin our study by investigating the differences between the approximations of Q_{rad} by (13.2) and (13.3). We derive an indicator based on χ that

allows us to distinguish between those regions where (13.2) should be used and those regions where (13.3) is the better approximation of $Q_{\rm rad}$. In Section 13.2 we study the efficiency of the numerical schemes described in the previous chapter. Note that in all our numerical tests we use the oscillation fix described in Section 12.3.

Finally we concern ourselves with the approximation of the data χ, B defining the radiation intensity I through (11.1c). In all the numerical tests presented so far, we assumed that both χ and B are defined pointwise on the whole domain. In our applications both χ and B depend on the hydrodynamic quantities, which are given element-wise by the finite–volume scheme. The quality of this approximation is essential for a stable approximation of the radiation source; we demonstrate this in Section 13.3.

13.1 The Average Intensity versus the Radiation Flux

First we study the approximation to $Q_{\rm rad}$ for Problem 11.1(SMOOTH). Since the exact solution I is independent of $\boldsymbol{\mu}$, we easily compute that $Q_{\rm rad} \equiv 0$. In Figures 13.1 and 13.2 we plot the approximation to $\overline{Q^J_{\rm rad}}$ and $\overline{Q^F_{\rm rad}}$ using the ESC1– and the ESC2–method, respectively. We can clearly distinguish two regions. For $y > 0$ the approximation of $Q^J_{\rm rad}$ is very close to zero, whereas the error in $Q^F_{\rm rad}$ is quite large. The situation is reversed for $y < 0$, where the error in $Q^J_{\rm rad}$ is very large. Between these two regions the absorption coefficient χ varies considerably in magnitude: for $y < 0$ the absorption is dominant (χ large) and for $y > 0$ the transport of radiation is dominant (χ small). A closer look at the RT equation (11.1c) can help to explain the influence of χ on the radiation source term. When χ is large the radiation intensity I is close to B. To approximate $Q^J_{\rm rad}$ in (13.2) the difference of I and B is taken and then multiplied by χ. Thus small errors in $I - B$ are amplified and this leads to the observed oscillations. On the other hand when χ is small then $\boldsymbol{\mu} \cdot \nabla I$ is also small. This causes oscillations when $Q^F_{\rm rad}$ is used. These observations were already made in [BVS99] for the standard short–characteristics method. With the following example we quantify the different quality of the approximation to $Q_{\rm rad}$ by using either the average radiation intensity J or the radiation flux F.

13.2 Example

We study the solution to a model problem that corresponds to the radiation transport equation (11.1c) *in 1d with only two directions* $\boldsymbol{\mu}$ *and constant data (cf. Section 4.1.2):*

$$u'(s) = \chi(B - u(s)) \qquad \textit{for} \quad s \in [0, h] \ , \tag{13.5a}$$

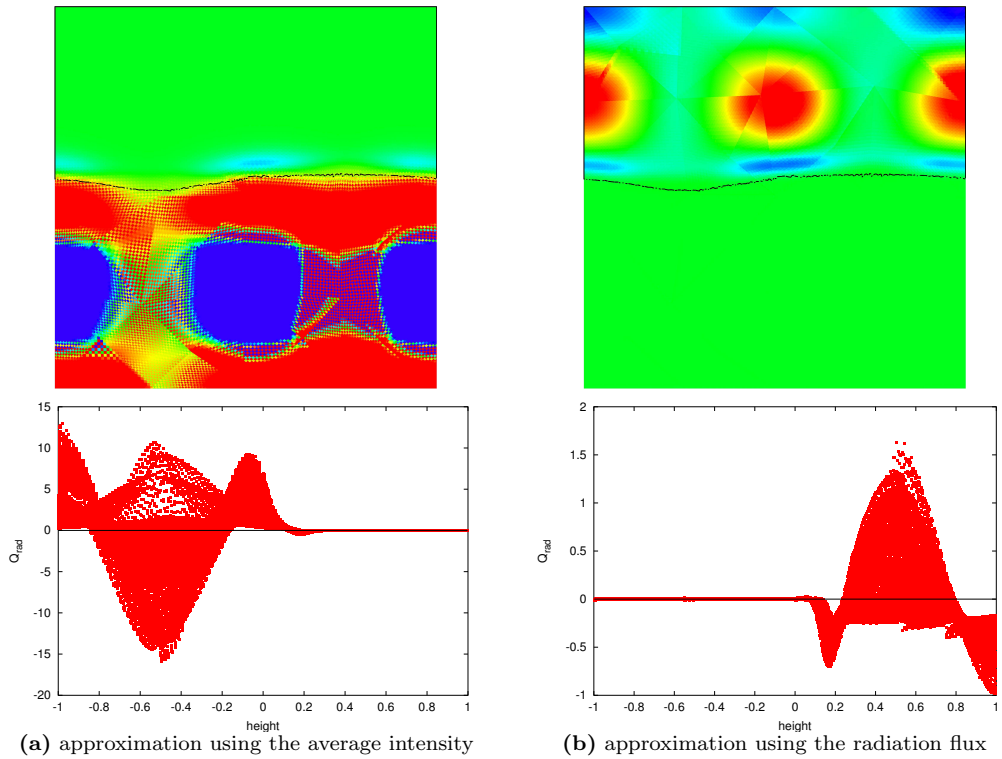
$$-v'(s) = \chi(B - v(s)) \qquad \textit{for} \quad s \in [0, h] \ , \tag{13.5b}$$

$$u(0) = u_0 \ , \tag{13.5c}$$

$$v(h) = v_1 \ , \tag{13.5d}$$

$$Q^J_{\rm rad}(s) = \chi\big(u(s) + v(s) - 2B\big) \qquad \textit{for} \quad s \in [0, h] \ . \tag{13.5e}$$

We assume that $\chi, B, u_0,$ *and* v_0 *are given constants.*



(a) approximation using the average intensity **(b)** approximation using the radiation flux

Figure 13.1: Problem 11.1(SMOOTH) with $\alpha = 15$. Isoline plots and scatter plots (along the y–axis) of the radiation source term using the ESC1–method. The black line represents the isoline where $\chi = \frac{1}{h}$. On the left Q_{rad} is computed using the average radiation intensity; on the right the radiation flux is used. Note that the exact solution is $Q_{\text{rad}} \equiv 0$ and that the value zero corresponds to the green color. The black line corresponds to $\chi h_T \equiv 1$.

We are interested in approximating the average radiation source term

$$\overline{Q^J_{\text{rad}}} := \frac{1}{h}\int_0^h \chi\big(u(s)+v(s)-2B\big)ds \ . \tag{13.6}$$

Using the ODEs (13.5a) *and* (13.5b) *defining* u *and* v*, respectively, we obtain*

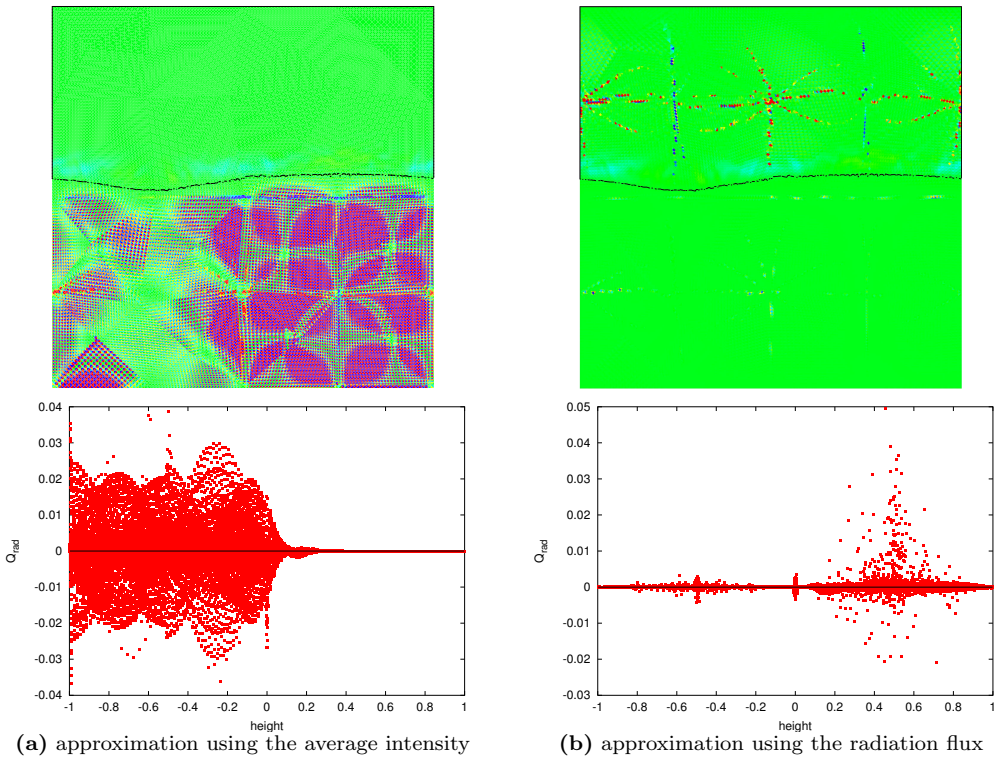
$$\overline{Q^J_{\text{rad}}} = -\frac{1}{h}\int_0^h \big(u'(s)-v'(s)\big)ds = -\frac{1}{h}\big(u(h)-u(0)-v(h)+v(0)\big) \ .$$

According to the notation introduced at the beginning of this chapter, we define the abbreviation

$$\overline{Q^F_{\text{rad}}} := -\frac{1}{h}\big(u(h)-u(0)-v(h)+v(0)\big) \ . \tag{13.7}$$

It is straightforward to compute the solutions to the initial value problems in (13.5)

$$u(s) = (u_0 - B)e^{-\chi s} + B \ , \quad v(s) = (v_1 - B)e^{-\chi(1-s)} + B \ .$$



(a) approximation using the average intensity

(b) approximation using the radiation flux

Figure 13.2: Problem 11.1(SMOOTH) with $\alpha = 15$. The ESC2–method leads to problems similar to the ESC1–method (cf. Figure 13.1) although the size of the oscillations is reduced by two orders of magnitude.

We thus compute

$$\overline{Q^F_{\rm rad}} = \overline{Q^J_{\rm rad}} = (u_0 + v_1 - 2B)\frac{1 - e^{-\chi h}}{h} \ . \tag{13.8}$$

We now approximate u, v by linear functions on $[0, h]$. As in the ESC1–method, we approximate $u(h)$ and $v(0)$ by values u_1 and v_0, respectively, which we use to define linear functions on $[0, h]$:

$$u_h(s) = \frac{h - s}{h}u_0 + \frac{s}{h}u_1 \ , \quad v_h(s) = \frac{h - s}{h}v_0 + \frac{s}{h}v_1 \ .$$

Now our approximation to $\overline{Q^J_{\rm rad}}$ is given by

$$\overline{Q^J_{{\rm rad}h}} = \frac{1}{h}\int_0^h \chi\big(u_h(s) + v_h(s) - 2B\big)ds \ .$$

Since both u_h and v_h are linear and B is constant we compute

$$\overline{Q^J_{{\rm rad}h}} = \chi\Big(u_h\big(\frac{h}{2}\big) + v_h\big(\frac{h}{2}\big) - 2B\Big) = \chi\big(\frac{1}{2}(u_0 + u_1) + \frac{1}{2}(v_0 + v_1) - 2B\big) \ . \tag{13.9}$$

As an approximation of $\overline{Q^F_{\mathrm{rad}}}$ *we obtain*

$$\overline{Q^F_{\mathrm{rad}h}} = -\frac{1}{h}\big(u_1 - u_0 - v_1 + v_0\big) \ . \tag{13.10}$$

To define u_1, v_0 *we use the backwards Euler scheme (which corresponds to the one step Radau IIa method used in the ESC1–scheme). This leads to the following equations*

$$u_1 = \frac{u_0 + h\chi B}{1 + h\chi} \ , \quad v_0 = \frac{v_1 + h\chi B}{1 + h\chi} \ .$$

Plugging these two values into (13.9) *and* (13.10) *leads to the following approximations:*

$$\overline{Q^J_{\mathrm{rad}h}} = (u_0 + v_1 - 2B)\frac{\chi + \frac{1}{2}h\chi^2}{1 + h\chi} \ , \quad \overline{Q^F_{\mathrm{rad}h}} = (u_0 + v_1 - 2B)\frac{\chi}{1 + h\chi} \ .$$

We first study the limits of these values for $\chi \to 0$ *and* $\chi \to \infty$*. For* $\chi \to 0$ *we compute that both* $\overline{Q^J_{\mathrm{rad}h}}$ *and* $\overline{Q^F_{\mathrm{rad}h}}$ *tend to zero. On the other hand, for* $\chi \to \infty$ *we have* $\overline{Q^J_{\mathrm{rad}h}} \to \infty$*, but* $\overline{Q^F_{\mathrm{rad}h}} \to \frac{1}{h}(u_0 + v_1 - 2B)$*. Thus the approximations differ greatly for* χ *large. Due to* (13.8) *we compute* $\overline{Q^J_{\mathrm{rad}}} \to 0$ *for* $\chi \to 0$ *and* $\overline{Q^J_{\mathrm{rad}}} \to \frac{1}{h}(u_0 + v_1 - 2B)$ *for* $\chi \to \infty$*. Therefore* $\overline{Q^F_{\mathrm{rad}h}}$ *must be a better approximation for the radiation source term for* χ *large. This simple analysis gives no indication of the quality of the two approximations for* χ *small. In Figure 13.3 we plot both approximations and the errors* $|\overline{Q^J_{\mathrm{rad}h}} - \overline{Q^J_{\mathrm{rad}}}|$ *and* $|\overline{Q^F_{\mathrm{rad}h}} - \overline{Q^J_{\mathrm{rad}}}|$ *as functions of* χ*. In the plots we have taken* $h = 1$ *and* $(u_0 + v_1 - 2B) = 1$*; the general structure of the approximations is not influenced by these values. We see that* $\overline{Q^J_{\mathrm{rad}h}}$ *is always above* $\overline{Q^J_{\mathrm{rad}}}$ *and that* $\overline{Q^F_{\mathrm{rad}h}}$ *is always below* $\overline{Q^J_{\mathrm{rad}}}$*. Furthermore the approximation using* $\overline{Q^J_{\mathrm{rad}h}}$ *is closer to the correct value for* $h\chi < 1$ *whereas only* $\overline{Q^F_{\mathrm{rad}h}}$ *is close to the correct value for* $h\chi > 1$*; the errors in the two approximations are equal for* $h\chi \approx 1.1$*. Note that for* $h \to 0$ *both approximations converge to the correct value* $\lim_{h\to 0} \overline{Q^J_{\mathrm{rad}}} = \chi(u_0 + v_1 - 2B)$*.*

To compute Q_{rad} the authors of [BVS99] propose using $\overline{Q^F_{\mathrm{rad}h}}$ on those elements T where χ is greater than some threshold value χ_0 and to use $\overline{Q^J_{\mathrm{rad}h}}$ otherwise. They suggest a threshold value that is either a constant or depends on the element size h_T. Our example indicates that the threshold should not be chosen independent of the element size, but rather that $h_T\chi$ is the relevant parameter. By means of Problem 11.2(H3) with varying constants α for the absorption coefficient χ, we verify this assumption. In Figure 13.4 we plot the error of the approximations $\overline{Q^J_{\mathrm{rad}h}}$ and $\overline{Q^F_{\mathrm{rad}h}}$ in the L^1–norm using both the ESC1– and the ESC2–methods. Again we observe a variation in the quality of the approximations depending on the size of the absorption coefficient $\chi \equiv \alpha$. For α small the approximation to Q_{rad} using the average radiation intensity ($\overline{Q^J_{\mathrm{rad}h}}$) is clearly more accurate than the approximation using the radiation flux ($\overline{Q^F_{\mathrm{rad}h}}$). For α large the opposite is the case. This is in accordance with the observations made at the beginning of the section. For $\alpha = 5$ we observe that the two error curves for the ESC1–method intersect so that Q^F_{rad} is the correct choice for h large, whereas Q^J_{rad} is to be favored for h small. For the ESC2–method we observe the same behavior for $\alpha = 50$. As we already surmised, together with χ the grid size h also plays an important role in

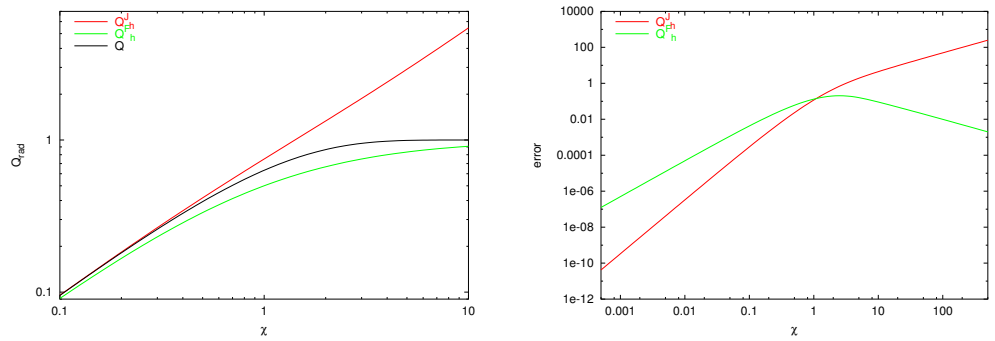


Figure 13.3: Approximation results for the model problem from Example 13.2 using the average radiation intensity Q^J_{rad} and the radiation flux Q^F_{rad}. We set $h = 1$ and $(u_0 + v_1 - 2B) = 1$ and plot the results as functions of χ. The approximation $\overline{Q^J_{\text{rad}h}}$ is always above the correct value, and the approximation $\overline{Q^F_{\text{rad}h}}$ always below it. For χ small $\overline{Q^J_{\text{rad}h}}$ is the better approximation, but for χ large only $\overline{Q^F_{\text{rad}h}}$ is close to the correct average radiation source term.

defining the threshold value, so that a reasonable indicator seems to be $h\chi$, which is a dimensionless quantity. The indicator $h\chi$ allows us to take into account the intersection of the error curves since it is small for h small, and we thus choose Q^J_{rad} in accordance with the observed behavior of the errors. Consequently we define

$$\overline{Q_{\text{rad}h}} := \begin{cases} \frac{1}{|T|} \sum_{m=1}^{M} Q^J_{\text{rad}T,\boldsymbol{\mu}_m} & h_T\chi(\boldsymbol{\omega}_T) < C \ , \\ -\frac{1}{|T|} \sum_{m=1}^{M} Q^F_{\text{rad}T,\boldsymbol{\mu}_m} & \text{otherwise} \ . \end{cases} \tag{13.11}$$

The results presented in Figure 13.4 show that choosing a different constant C for the ESC1– and the ESC2–methods would increase the efficiency of both schemes (since the intersection of the lines occur for different values of α). Since the generalization of the results for the simple problem studied here is not straightforward, we decide to choose $C = 1$ in all our calculations, independent of the problem and the scheme used. Thus (13.11) with $C = 1$ is used for approximating Q_{rad} in the rest of this chapter. In Figure 13.4 we have also plotted the error in $\overline{Q_{\text{rad}h}}$. In Figure 13.1 and Figure 13.2 the black line in the isoline plots shows the isolevel where $h_T\chi(\boldsymbol{\omega}_T) = 1$. We clearly see that for both problems (13.11) leads to an evident improvement in the approximation compared to using $\overline{Q^J_{\text{rad}h}}$ or $\overline{Q^F_{\text{rad}h}}$ alone.

In Figure 13.5 we plot the L^1–error versus the runtime for Problem 11.1(SMOOTH). We plot the error for the approximations $\overline{Q^J_{\text{rad}h}}$, $\overline{Q^F_{\text{rad}h}}$, and also for $\overline{Q_{\text{rad}h}}$ as defined in (13.11) using $C = 1$. The advantage of our mixed definition of Q_{rad} can be clearly seen. Since χ is large in parts of the domain and small in others, the error produced by using (13.11) is far below the errors produced by using Q^J_{rad} or Q^F_{rad} on the whole domain. In the case of the ESC1–method, the approximation using (13.11) on the coarsest grid is just as good as the approximations of Q^J_{rad} and Q^F_{rad} on the finest grid level. The quality of the approximation using the ESC2–method is substantially improved by (13.11), as well; on the finest grid level the use of our indicator leads to

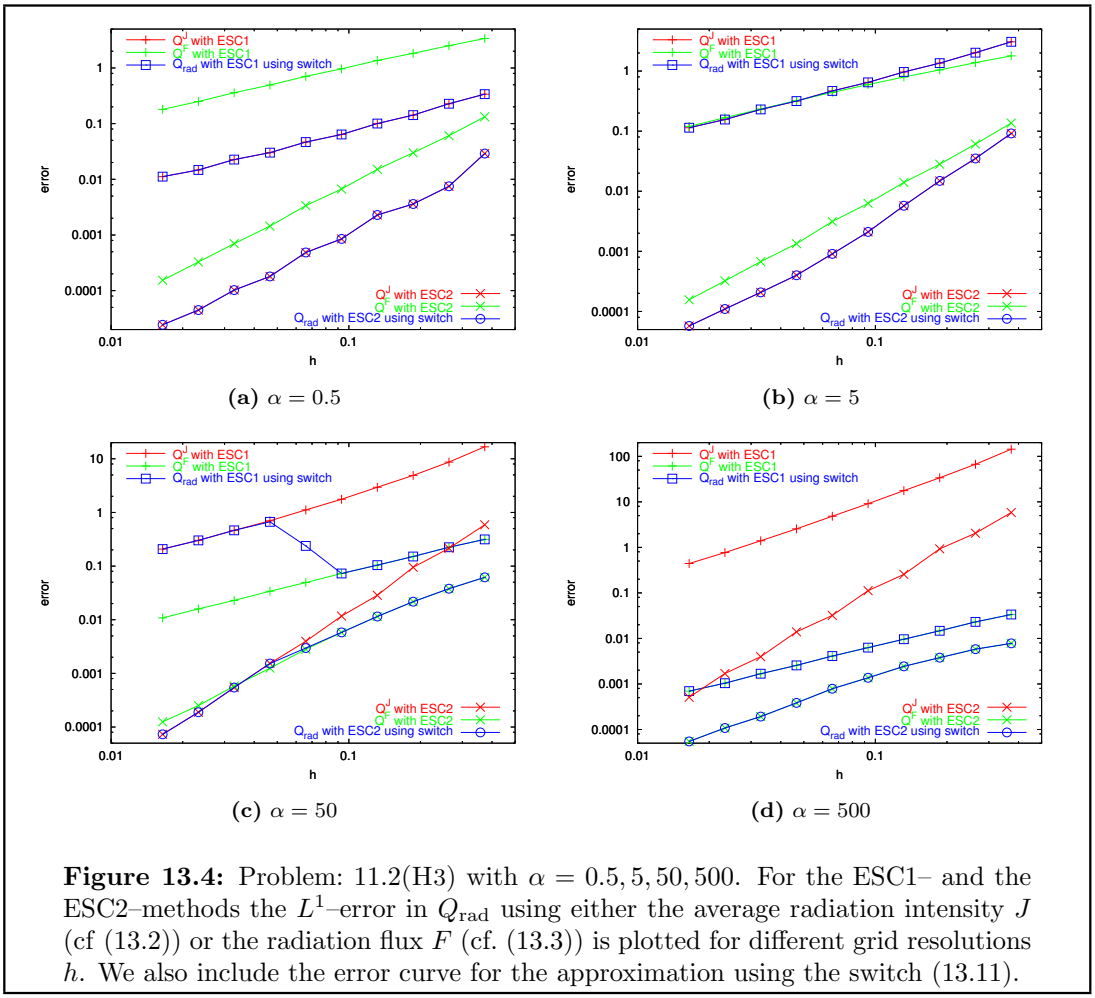


Figure 13.4: Problem: 11.2(H3) with $\alpha = 0.5, 5, 50, 500$. For the ESC1– and the ESC2–methods the L^1–error in Q_{rad} using either the average radiation intensity J (cf (13.2)) or the radiation flux F (cf. (13.3)) is plotted for different grid resolutions h. We also include the error curve for the approximation using the switch (13.11).

a gain in runtime of more than a factor of 50 compared to using only Q^J_{rad} and to a gain of a factor of four compared to using Q^F_{rad}. The curves for $\overline{Q^J_{\text{rad}h}}$ show a higher convergence rate than the curves for $\overline{Q^F_{\text{rad}h}}$, which have about the same slope as the corresponding curves for $\overline{Q_{\text{rad}h}}$. Consequently, these curves will intersect for h small enough. Our indicator takes this behavior into account since for small h we use Q^J_{rad} to compute Q_{rad}. Furthermore we can clearly see that the higher order ESC2–scheme is far more efficient than the ESC1–method: the error for the ESC2 approximation on the coarsest grid is of the same size as the error on the finest grid using the ESC1–method; this leads to a gain in runtime by a factor of more than 200.

Before we compare the performance of all the different schemes in the next section, we verify that our indicator (13.11) for computing Q_{rad} leads to reasonable results for the more realistic setting of the model solar atmosphere (Problem 11.5(SolarAtm)). In many ways this problem influenced our choice for Problem 11.1(Smooth): Q_{rad} is close to zero in large parts of the domain, on the one hand, because χ is very large (below the visible surface of the sun) and, on the other hand, because χ is very small (above the visible surface). In the vicinity of the solar surface ($y_0 \approx 0.5$) energy is transformed

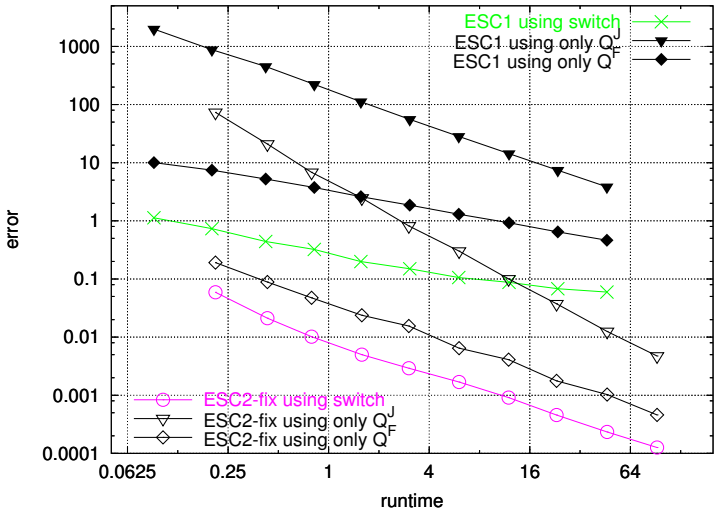


Figure 13.5: Problem: 11.1(SMOOTH) with $\alpha = 15$. The error vs. the runtime is plotted for the approximation of the radiation source term Q_{rad}. For both methods we plot the error in $\overline{Q^J_{\text{rad}_h}}$ and $\overline{Q^F_{\text{rad}_h}}$ together with the error in $\overline{Q_{\text{rad}_h}}$, which we compute using (13.11) with $C = 1$.

into radiation, which is transported outwards. This leads to a cooling (Q_{rad} is negative) of the plasma. Directly above the solar surface the plasma is heated ($Q_{\text{rad}} > 0$).
Since the data χ, B depend only on height (z), we can use a one dimensional model to compute a reference solution for the radiation source term (cf. Section 4.1.2). In one space–dimension the angular integral is reduced to an integral over the interval $[-1, 1]$, so that we can compute Q_{rad} very accurately. We use a set of M points equally distributed in the interval $[-1, 1]$ for the propagation directions and a set of N spatial points to approximate the radiation intensity I. For the approximation of the ODE defining I for a fixed direction we use the three step Radau IIa method described in Table 12.1. In Figure 13.6 we plot the result of the one dimensional calculation. We have portioned the computational domain into three parts corresponding to the three different physical regimes found in our model; the structure of the solution sketched above is clearly visible. In addition we see some disturbance in the solution, which is reduced by a higher resolution. These oscillations are localized around one grid point, so that the convergence of the scheme in an integral norm can still be expected. We use this high resolution approximation to measure the quality of our two dimensional schemes.
In Figure 13.7 and Figure 13.8 we plot $\overline{Q^J_{\text{rad}_h}}$ and $\overline{Q^F_{\text{rad}_h}}$ for the ESC1– and the ESC2–method, respectively. In the isoline plots we have also included the isolevel of the indicator at which we switch from Q^J_{rad} (above the black line) to Q^F_{rad} (below the black line). Note that the grid is irregular so that h_T is not a function of height. Consequently, the line defined by $h_T \chi(\boldsymbol{\omega}_T) = 1$ is not a straight line. We see that (13.11) leads to satisfactory results — although for this example good results are already obtained by using Q^F_{rad} everywhere in the domain. However, Q^J_{rad} leads to very poor results in lower parts of the domain where χ is very large. In many other respects the results point

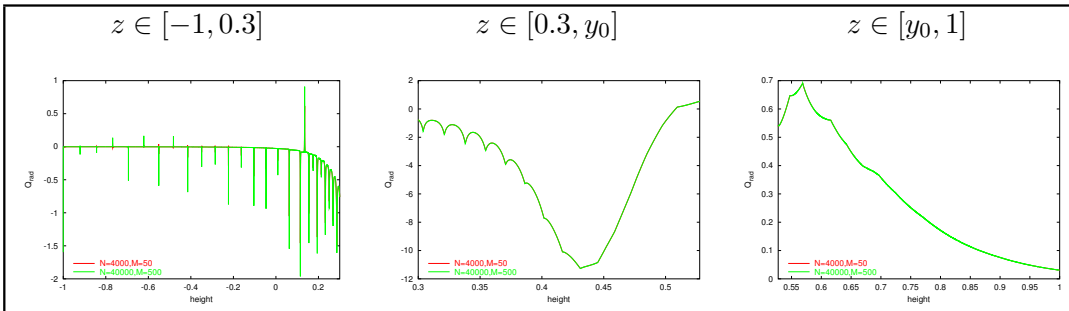


Figure 13.6: Problem 11.5(SolarAtm). Reference solution for Q_{rad} computed using the 1d form of the RT equation on two different grids. We plot the result on three non–overlapping subintervals with different scaling of the y–axis in each case. Due to problems with the data, the scheme leads to some small–scale oscillations. Note that these are located only on one grid point so that the error in the L^1–norm is still reduced by grid refinement.

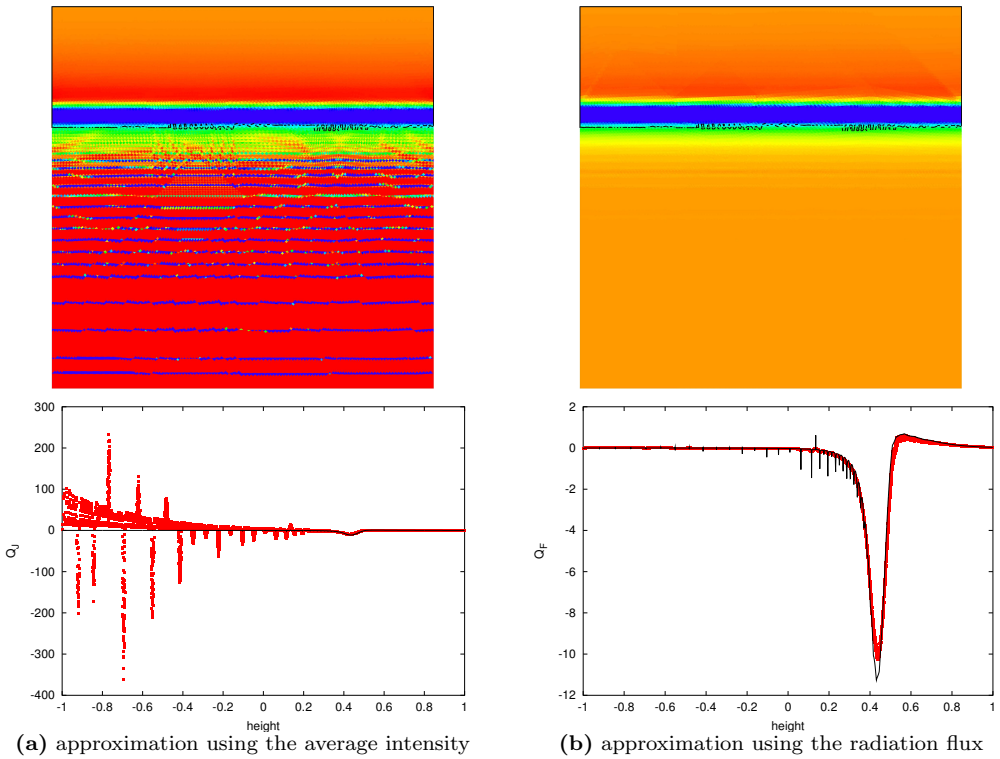
in the same direction as the results for the test case 11.1(Smooth). The size of the oscillations in Q_{rad}^F are substantially reduced and the minimum in the source term is far more accurately resolved by using the second order ESC2–scheme. Furthermore we can see that the 1d structure of the solution is far better captured by the ESC2–scheme (for example, in the region around $x = 0.6$, where the heating is greatest).

Summary of Section 13.1: *As Example 13.2 demonstrates, the quality of the approximation to the radiation source term Q_{rad} depends greatly on whether the average radiation intensity J or the radiation flux F is used. If the absorption coefficient χ is large, then the approximation by means of J leads to very bad results. For χ small, on the other hand, J leads to a more accurate approximation. We observed the same behavior in our numerical tests for both the ESC1– and the ESC2–method, as well. (For the conservative DG– and CESC–schemes both approximations are identical.) Furthermore we observed that the difference between the two approximations also depends on the grid size h. Consequently we derived a simple indicator that switches between both approximations depending on the size of $h_T\chi(\boldsymbol{\omega}_T)$. We also tested the indicator in the case of our model solar atmosphere (Problem 11.5(SolarAtm)), where it seems to lead to reasonable results, as well. Consequently the approximation* (13.11) *with $C = 1$ is used in the following sections.*

In Figure 13.5 we began our study of the efficiency of the ESC–schemes. Our focus lay on the different possibilities of approximating Q_{rad}. As expected, the approximation using (13.11) *was clearly superior to the other approaches. Furthermore we saw that even on the coarsest grid the second order scheme was far more efficient than the first order scheme. The study of the efficiency of the schemes is continued in the following section.*

13.2 Efficiency of the Numerical Schemes

As in the previous chapter, we are mainly interested in quantifying the efficiency of the different schemes for approximating the radiation source term Q_{rad} defined by (11.1b).



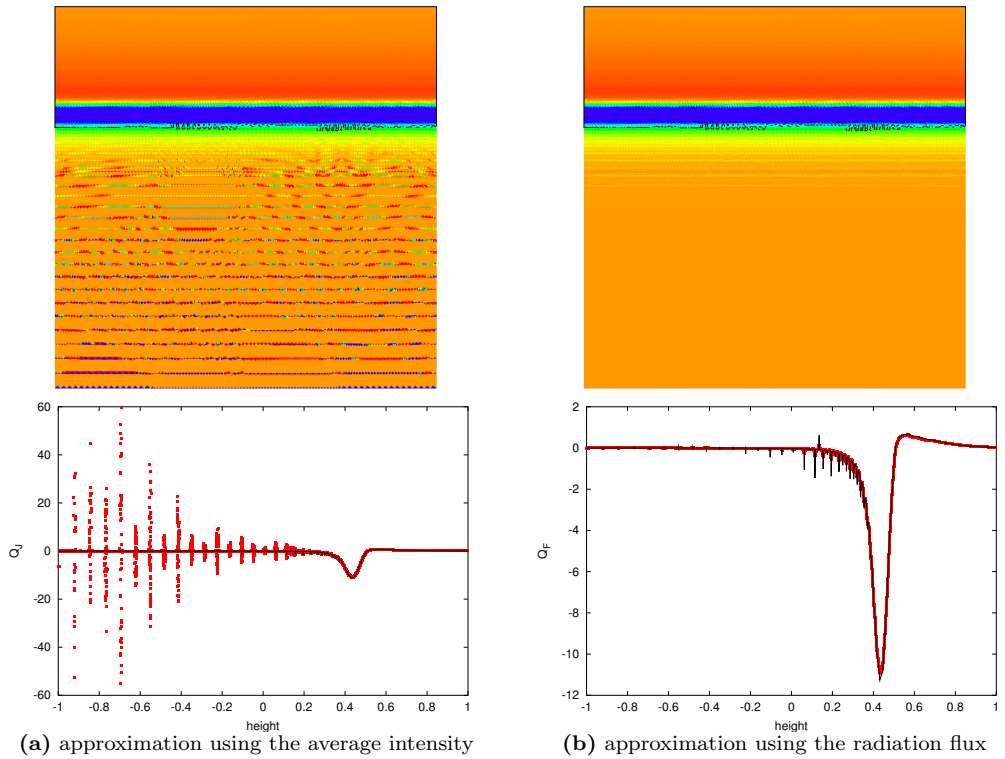
(a) approximation using the average intensity

(b) approximation using the radiation flux

Figure 13.7: Problem 11.5(SolarAtm) using the ESC1–method. Oscillations in the lower part of the atmosphere, where χ is large, lead to large errors when Q_{rad} is computed using the average radiation intensity J. This problem is not present if the radiation flux F is used. The black lines in the scatter plot shows the 1d reference solution.

In the case of the ESC–schemes we approximate Q_{rad} using (13.11) with $C = 1$; in the case of the conservative schemes the difference between using the approximations $\overline{Q^J_{\text{rad}h}}$ and $\overline{Q^F_{\text{rad}h}}$ is negligible. In Figure 13.9 we plot the error to runtime ratio for Problem 11.1(Smooth). The advantage of using the ESC2–method together with the indicator for defining Q_{rad} is very clear. Due to strong oscillations in those parts of the domain where χ is large (cf. Figure 13.10), the DG1–method is far less efficient than the ESC2-method (even less efficient than the ESC1–method for the grid resolutions shown here). In fact, since there is no possibility of switching between Q^J_{rad} and Q^F_{rad}, all the conservative schemes show a reduced error to runtime ratio compared with the results presented in the previous chapter, where in Figure 12.17 we measured the error in the intensity for the same problem. Note that the error to runtime ratio for the ESC–methods is very similar to the error to runtime ratio observed in Section 12.7.3.

In Figure 13.11 we plot Q_{rad} for the model atmosphere (Problem 11.5(SolarAtm)) as computed by our two dimensional schemes. Especially the first order schemes have difficulty resolving the different parts of the solution. In the lower part of the computational domain the DG0–method produces oscillations that are an order of magnitude larger than the global maximum and minimum of Q_{rad}. If we use Q^J_{rad} in the whole domain, the results obtained using the ESC1–method resemble the ones obtained using



(a) approximation using the average intensity **(b)** approximation using the radiation flux

Figure 13.8: Problem 11.5(SolarAtm) using the ESC2–method. As in Figure 13.7 we observe problems with oscillations in the lower part of the atmosphere. Although these are reduced by the higher order scheme, using the radiation flux to compute Q_{rad} leads to a significantly improved approximation.

the DG0–method (cf. Figure 13.7). Since the DG0–method is conservative, we have no alternative way to compute Q_{rad} so that we cannot improve the results.

The ESC1–method clearly shows its first order accuracy: the maximum and the minimum values in Q_{rad} are not resolved. In the upper regions of the atmosphere all the higher order schemes are comparable. In the lower regions, however, the conservative methods have problems with resolving the correct solution. For example, a closer look at the results for the DG1–method shows that the value zero is not obtained, but rather that two values (one slightly above zero the other slightly below zero) are assumed.

Summary of Section 13.2: *The higher efficiency of the ESC–schemes — observed in the previous chapter — is even more evident if we study the approximation of the radiation source term Q_{rad}. This is mainly due to the possibility of switching between the approximations $\overline{Q^J_{\text{rad}h}}$ and $\overline{Q^F_{\text{rad}h}}$ in the ESC–methods. Especially the approximation using the average radiation intensity J leads to very bad results for χ large.*

The possibility of distinguishing between regions where χ is large and regions where χ is small also leads to a considerable improvement in the approximation for the model solar atmosphere. The oscillations in the lower regions of the atmosphere can be severely reduced by using the approximation (13.11)*. (cf. Figure 13.7, Figure 13.8, and Figure 13.11). Here the first order ESC–scheme reproduces the solution (at least roughly),*

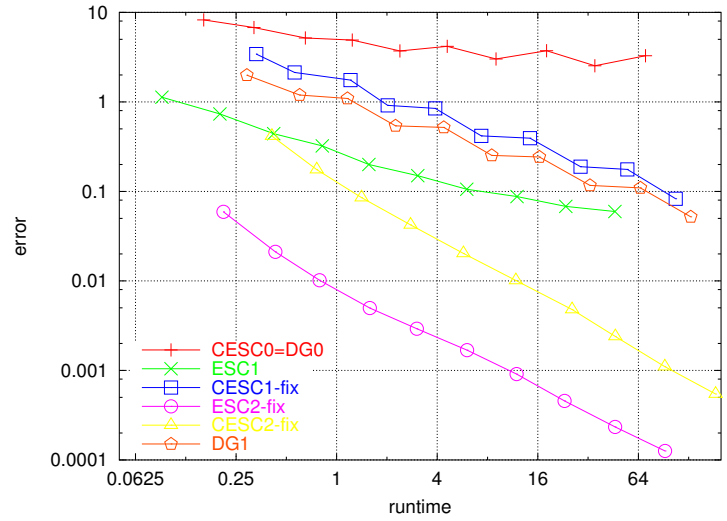


Figure 13.9: Problem: 11.1(SMOOTH) with $\alpha = 15$. The error vs. the runtime is plotted for the approximation of Q_{rad}. For the ESC–methods we use (13.11) with $C = 1$ to compute Q_{rad}. In the case of the conservative schemes both Q^J_{rad} and Q^F_{rad} lead to the same results.

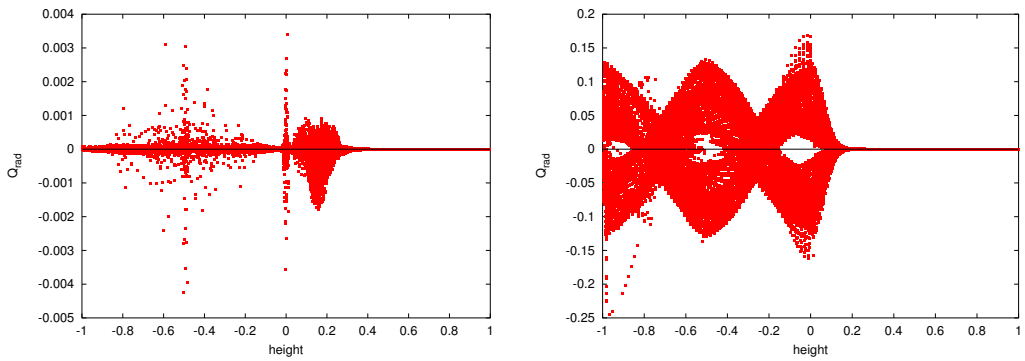


Figure 13.10: Problem 11.1(SMOOTH) with $\alpha = 15$. We compute Q_{rad} with the ESC2–method using (13.11) with $C = 1$ (left). The DG1–scheme (right) leads to oscillations in the lower part of the domain that are similar to those observed for Q^J_{rad} using the ESC2–scheme (cf. Figure 13.2). Since average values $\overline{Q^J_{\text{rad}}}$ and $\overline{Q^F_{\text{rad}}}$ are identical in the case of the DG1–scheme, the oscillations cannot be reduced by the technique presented in Section 13.1 and which we use for the ESC2–scheme.

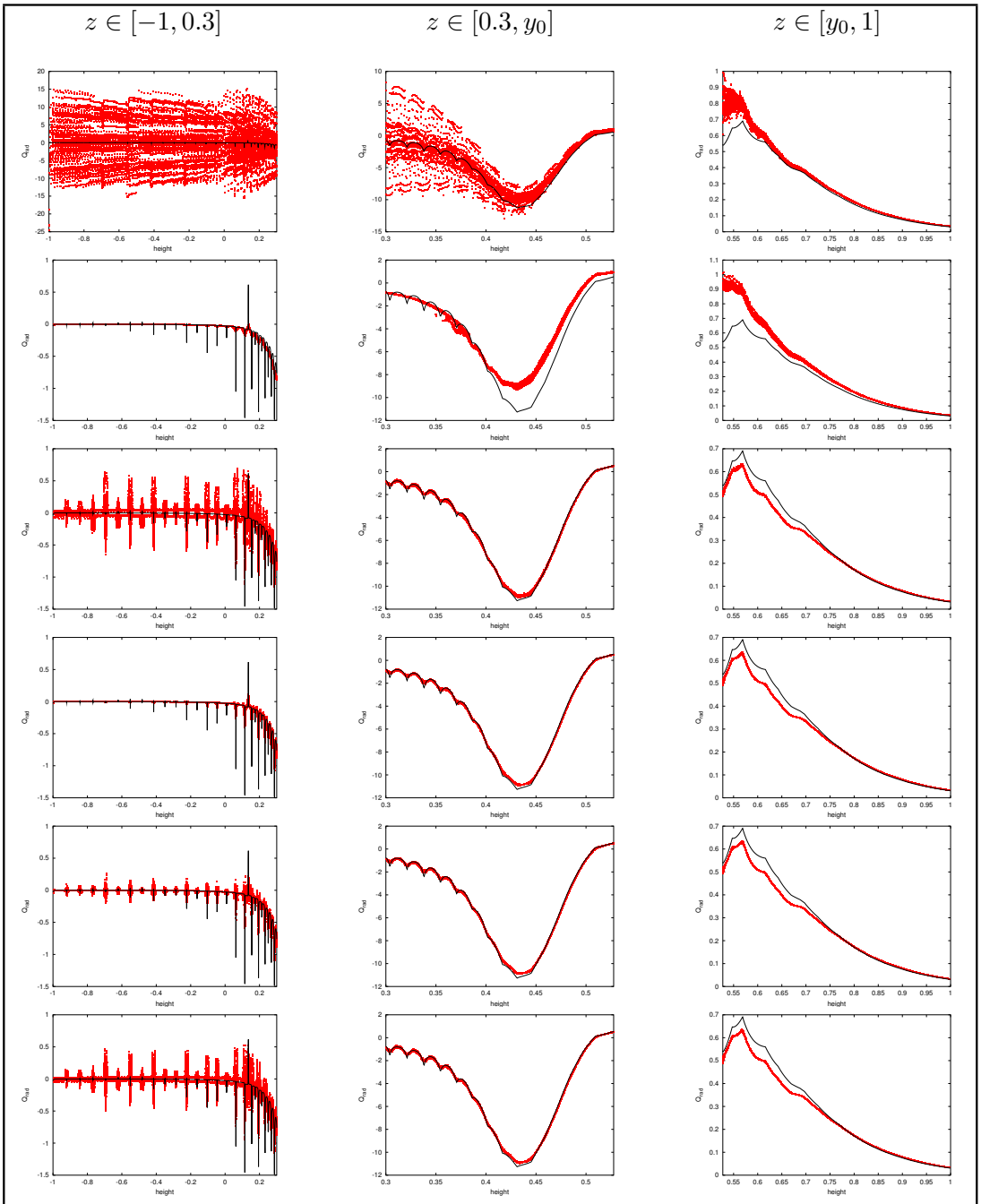


Figure 13.11: Problem 11.5(SolarAtm). The approximation of the radiation source term Q_{rad} is plotted for the schemes DG0=CESC0,ESC1,CESC1,ESC2,CESC2, and DG1 (top to bottom). In the case of the ESC1– and the ESC2–method Q_{rad} is computed using (13.11) with $C = 1$. This results in a considerable reduction in the size of the oscillations. For the conservative schemes no similar technique can be applied, so that oscillations in the lower part of the atmosphere are still present. All the higher order schemes lead to a good resolution of the cooling and heating zone above the solar surface.

but the first order DG–scheme (and CESC–scheme) leads to an unacceptable amount of oscillations. The oscillations in the second order schemes are far less severe than in the first order schemes, and among the second order conservative methods the DG–scheme clearly leads to the better approximation. Compared to the ESC2–method, however, the oscillations are still quite a bit larger. If we also take into account the fact that the DG1–method required approximately 25 *percent more computation time on a fixed grid, we conclude that the ESC2–method is the most suitable for our applications.*

13.3 Approximation of the Data

The last step towards a realistic setting for our radiation transport problem lies in the realistic approximation of the data defining the radiation intensity. Both χ and B are functions of the hydrodynamic quantities ρ and θ (cf. Section 1.2). After coupling the radiation transport solver to the finite–volume scheme, these quantities are defined on each element and may be discontinuous on element edges. On each edge we may thus have two different values for ρ and θ and thus for χ and B. For the vertices of the grid the situation is even worse since an arbitrary number of elements can intersect with a given vertex. This number depends on the macro triangulation, and in typical situations about eight elements can have one vertex in common. Due to the highly non–linear dependency of B and χ on the temperature θ, even small jumps between neighboring elements can lead to very different values for the data in the radiation transport equation. For computing the element integrals on T for the DG–methods and for the CESC–methods, we use only the functions defined on this element; for the solution of the short–characteristic problems (12.4) we use only the values defined on T. For example, in the case of the ESC1–method using the one step Radau IIa ODE solver this leads to the following situation: for a fixed direction the approximation of the radiation intensity in a given vertex $\mathbf{p}$ is computed using the values $\rho_T(\mathbf{p})$ and $\theta_T(\mathbf{p})$ given by the approximation on an element T. For a second direction, values $\rho_{T'}(\mathbf{p}), \theta_{T'}(\mathbf{p})$ are used, which can differ greatly from $\rho_T(\mathbf{p}), \theta_T(\mathbf{p})$. (Note that T and T' might not even have a common surface.) Similar arguments show that all the schemes studied here have this problem in common. In the following we study the consequences for the approximation of Q_{rad} using again the model solar atmosphere Problem 11.5(SolarAtm) as test case.

The results presented in this section were computed by first averaging the density ρ and the temperature θ onto the elements of a grid and then using these average values to compute a suitable reconstruction. We compare three different approaches: we use the constant data given by the finite–volume scheme, we use the reconstructed data given by the second order finite–volume scheme, and we test the continuous reconstruction already used in Section 8.2 (cf. Algorithm 4(a) on page 117). For the third approach we have to average the data on all elements surrounding a vertex $\mathbf{p}$ to define the hydrodynamic quantities on this vertex; these values are used to define a piecewise linear and continuous reconstruction of the hydrodynamic quantities. For the piecewise linear reconstruction we use the same method as for our MHD solver (cf. [DRW02a]). For the results shown here the reconstruction is performed in conservative variables; the results using reconstruction in primitive variables are very similar since for simplicity

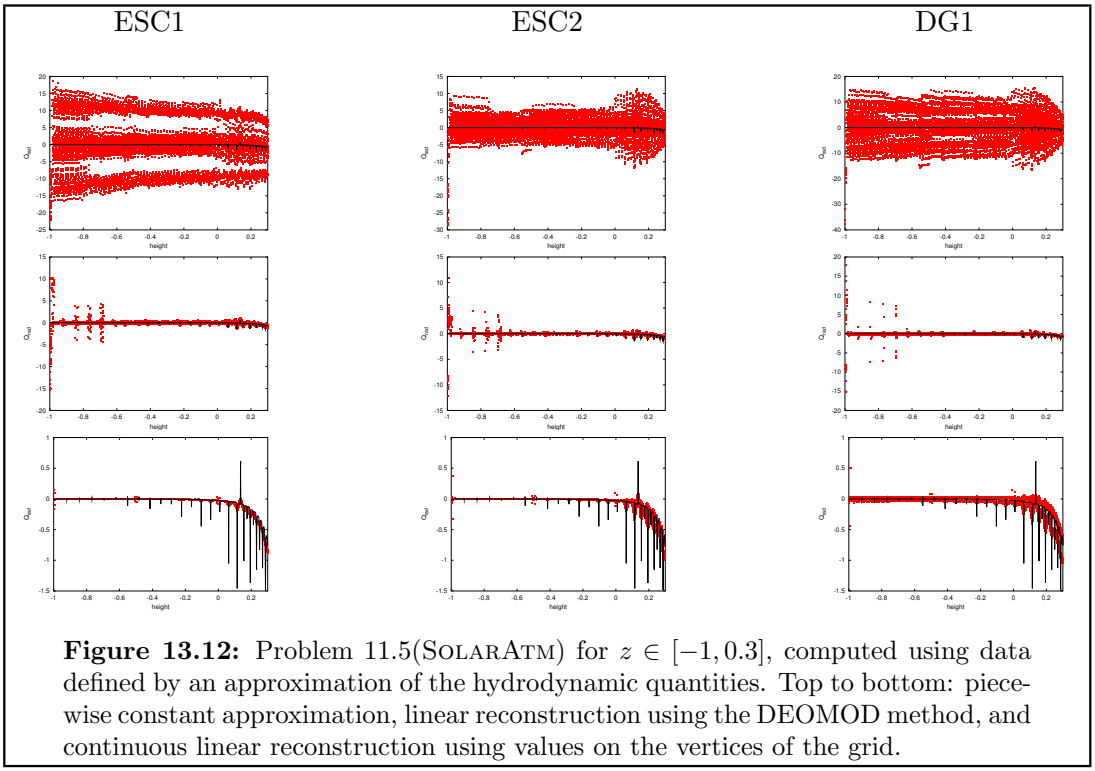


Figure 13.12: Problem 11.5(SolarAtm) for $z \in [-1, 0.3]$, computed using data defined by an approximation of the hydrodynamic quantities. Top to bottom: piecewise constant approximation, linear reconstruction using the DEOMOD method, and continuous linear reconstruction using values on the vertices of the grid.

we assumed a perfect gas law for our tests. If we were to use the more realistic EOS for a partially ionized plasma, a difference between conservative and primitive reconstruction might well be observed.

We only show results for the ESC1– and the ESC2–methods since these are the first and second order schemes that we favor. For a better comparison, we also include results for the DG1–scheme. As before, we have partitioned the computational domain into three parts: the results computed in the lower part of the atmosphere are shown in Figure 13.12, the region just below the solar surface is shown in Figure 13.13, and the upper part of the domain is shown in Figure 13.14. We clearly see that discontinuities in the data lead to problems with oscillations, especially in the upper parts of the domain. Constant data are not sufficiently smooth to capture the solution in the region below the solar surface (cf. Figure 13.12 and Figure 13.13); this is true for all the schemes tested. In the upper part of the domain even constant data lead to results that are comparable to those obtained with continuously given data (cf. Figure 13.11). In the region around the solar surface, we see little difference between the results using the finite–volume reconstruction and the vertex reconstruction (cf. Figure 13.13). The difference between these two techniques is, however, clearly noticeable in the lower regions of the atmosphere (cf. Figure 13.12).

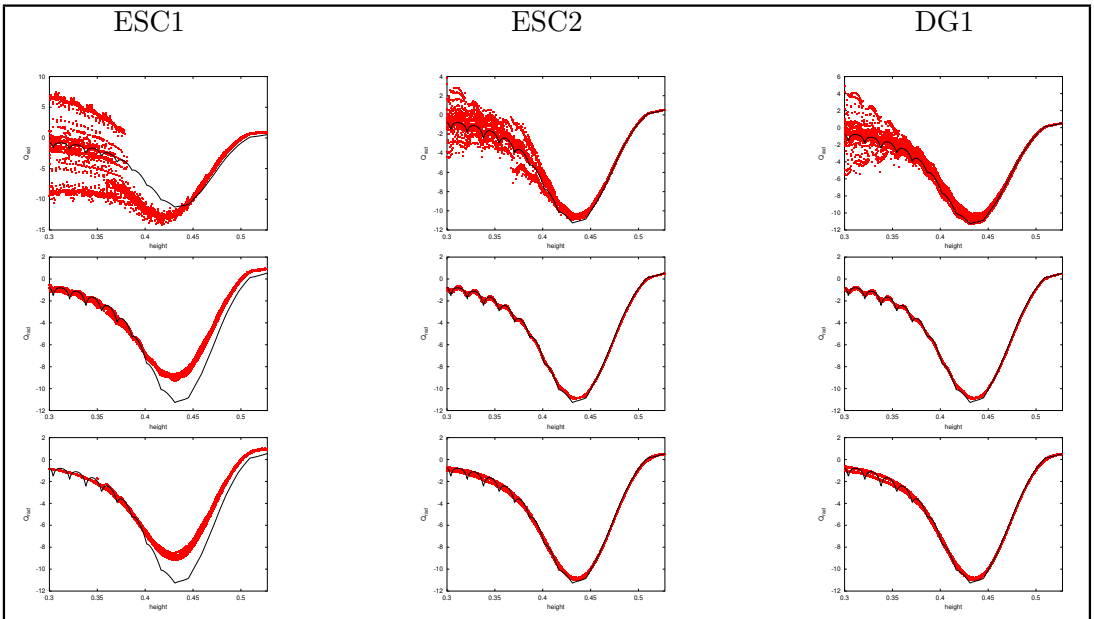


Figure 13.13: Problem 11.5(SolarAtm) for $z \in [0.3, y_0]$, computed using data defined by an approximation of the hydrodynamic quantities. Top to bottom: piecewise constant approximation, linear reconstruction using the DEOMOD method, and continuous linear reconstruction using values on the vertices of the grid.

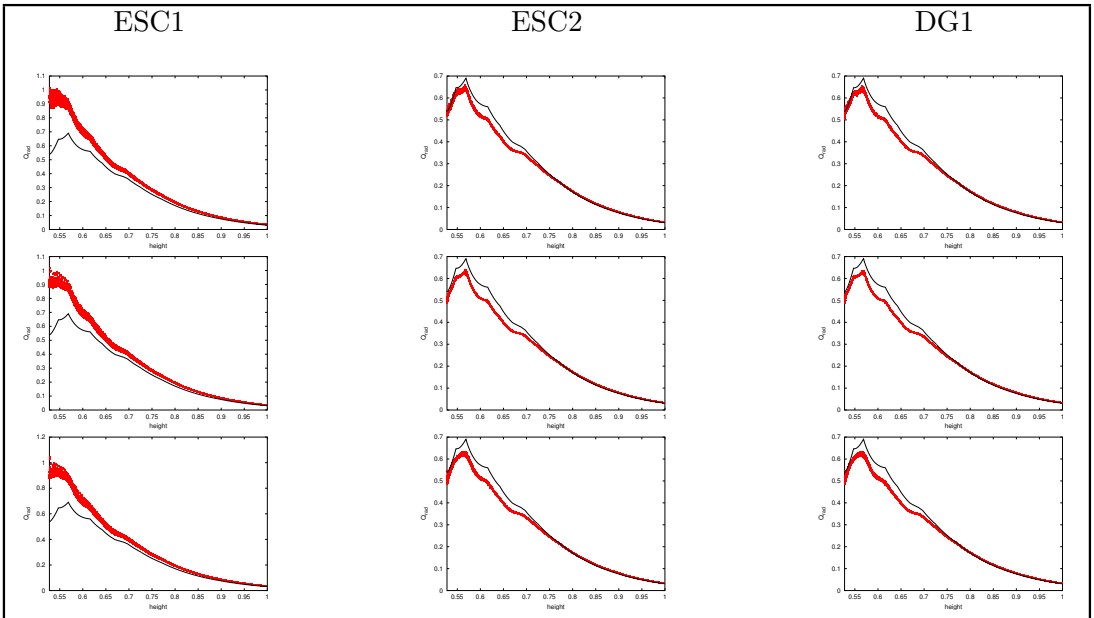


Figure 13.14: Problem 11.5(SolarAtm) for $z \in [y_0, 1]$, computed using data defined by an approximation of the hydrodynamic quantities. Top to bottom: piecewise constant approximation, linear reconstruction using the DEOMOD method, and continuous linear reconstruction using values on the vertices of the grid.

14. Summary

Radiation Transport Scheme

In the previous two chapters we studied a number of different ways of approximating the radiation transport equation (11.1c). We derived the first order ESC1–method and the second order ESC2–method based on the short–characteristics approach. Through a simple modification of the scheme, we also derived a conservative extension of the ESC approach: the first order CESC0–method, the second order CESC1–method, and the third order CESC2–method. This extension of the ESC–scheme was facilitated by the general framework that we derived for the ESC approach in Section 12.1. The general setting allows for a wide range of further modifications of the scheme that we have not discussed so far. For example, our choice of the points $\mathbf{p}_i$, in which the short–characteristic problems are solved, is not the only possible choice; the midpoints $\mathbf{z}_{ij}$ of the edges of the triangle T might lead to an improved convergence rate since the continuity of the approximation is then given only in the midpoints of the edges and not globally. Our technique for suppressing oscillations can, however, not be applied to this setting since it requires at least two approximate intensity values per edge. A further class of schemes is given by not choosing the coefficients I_j^T in (12.2) according to $I_j^T \approx I(\mathbf{p}_j^T)$, but rather by requiring that the approximation I_T satisfies the RT equation (11.1c) in all the points $\mathbf{p}_j^T$. Further flexibility comes from the decoupling of the short–characteristic problems. Since we have to cope with stiff source terms in our applications, we use an implicit Runge–Kutta solver for this step in the scheme. In those regions where χ is small either an explicit solver or also the simple second order Crank–Nicholson method might be sufficient and would increase the efficiency of the scheme. This is another example of the ease with which the ESC–methods can be dynamically adapted to the given problem; we already demonstrated this possibility in the case of the adaptation of the order of the scheme in Section 12.9. We compared the ESC– and the CESC–schemes with the first and second order discontinuous Galerkin methods (termed DG0 and DG1, respectively). These method are derived from a totally different approach based on the variational formulation of the RT equation (11.1c).

In Section 11.1.2 we posed five demands that an approximation method should fulfill in order to allow for an efficient numerical scheme for solving the coupled system (1.19) of radiation magnetohydrodynamics. These demands served as a starting point for our numerical investigations; our solar physical applications and the high computational cost involved in computing the radiation field of a non–stationary plasma were the

primary factors in their formulation. Consequently, the efficiency of the scheme was the main focus of our study. The difficult physical regimes in the solar photosphere also place high demands on the solution method. Especially the large differences in the absorption coefficient χ between the regions below the solar surface where $\chi \approx 10^5$ and the regions above the solar surface where $\chi \approx 0.1$ have to be taken into account. Furthermore the data defining the radiation field depend highly non–linear on the hydrodynamic quantities so that the interaction between the approximation due to the finite–volume scheme and the numerical scheme for the RT equation have to be studied in detail.
The discussion in the previous two chapters has lead us to the following conclusions. We summarize first the central results for the approximation of the radiation transport equation (11.1c) presented in Chapter 12:

- The (C)ESC and the DG approaches differ little with respect to the complexity of the implementation. In the (C)ESC method the main part of the algorithm is the solution of the RT equation along one characteristic, and this part can be used for any type of grid in 2d and in 3d. Only the evaluation of the basis functions and that part of the algorithm where the intersection of a given characteristic with the inflow boundary ∂T_- of a given element T is computed depends on the type of element.

- Higher order methods are essential to accurately approximate regions with large gradients in the intensity. First order schemes lead to a high amount of crosswind diffusion so that, irrespective of the smoothness of the data, first order schemes are far less efficient then second order schemes.

- Higher order methods lead to oscillations in regions with high gradients in the intensity. In the case of the ESC– and CESC–methods these can be removed efficiently through a simple limiting process, which barely reduces the quality of the approximation. For the DG–method we are not aware of any way to reduce these oscillations.

- If the compatibility of the numerical scheme with the integral form of the RT equation is not relevant in the physical application, then the ESC–methods are the most efficient methods tested here. Especially if we study the approximation of both I and $\boldsymbol{\mu}\cdot\nabla I$, we find that the ESC1–method is the most efficient first order method and that the ESC2–method is the most efficient second order method. If the conservation property of the scheme is essential, then the ESC–method cannot be used. In this case we have to compare the CESC– and the DG–methods. Here we found very little difference between the schemes with the same polynomial ansatz space.

- The main disadvantage of the ESC– and the CESC–methods is the incomplete analytical justification of the approach. At least in the case where the solution I to the RT equation is smooth, the convergence analysis of the DG–scheme is well–established. For the ESC– and the CESC–methods we can only prove the convergence of the first order schemes ESC1 and CESC0. For the higher order ESC–methods, the restriction on the limiting operator $\mathcal{G}$ is so severe that we can

show only first order convergence using a special choice for $\mathcal{G}$. Furthermore we do not have complete control of the error in $\boldsymbol{\mu} \cdot \nabla I$. Although we cannot expect the high amount of regularity required for the convergence proof of the DG scheme in our applications, the lack of analytical results for the higher order ESC– and CESC–methods is a point in favor of the finite element approach.

In many ways the observations summarized so far also apply to the approximation of the radiation source term Q_{rad}. Based on our numerical tests in Chapter 13 we add the following observations:

- The approximations of the radiation source term Q_{rad} (cf. (1.19g)) show large oscillations in regions with large χ. These oscillations are much stronger when a first order scheme is used instead of a higher order scheme. In the case of the non–conservative ESC–methods, the use of the radiation flux F to compute Q_{rad} reduces these oscillations considerably. Since for conservative schemes the approximation using F is identical to the approximation using the average radiation intensity J, this approach cannot be used in this case. With the help of a simple indicator (13.11), the ESC–method leads to a very efficient approximation of Q_{rad}. With this indicator the second order ESC2–method was in some cases even more efficient than the CESC2–method, which is third order accurate.

- For an accurate approximation of Q_{rad}, a reconstruction of the piecewise constant approximation given by the finite–volume scheme is unavoidable. In the case of a higher order scheme like ESC2 or DG1, the piecewise linear but discontinuous reconstruction used in the second order finite–volume scheme already leads to good results. Only in those regions where χ is very large does the discontinuity of the data over element edges lead to oscillation in the approximation of Q_{rad}. A piecewise linear and continuous (but expensive to compute) reconstruction also leads to a good approximation in those regions where χ is large.

All our numerical tests demonstrate that the ESC approach leads to a more efficient scheme than the other approaches tested here. This is due to the possibility of suppressing oscillations in the vicinity of large gradients, and the possibility of switching between two approximations for Q_{rad} also increases the efficiency of the scheme considerably. Both conservative approaches show oscillatory behavior in the approximation of Q_{rad} when χ is large. These oscillations reduce the quality of the scheme to the extent that even the first order ESC–method can be more efficient than the second order CESC– and DG–methods. Therefore, if it is not essential that the approximation satisfies the integral form of the RT equation, we conclude that our ESC approach leads to the most efficient and robust solution schemes for approximating the radiation intensity I and the radiation source term Q_{rad} in our applications.

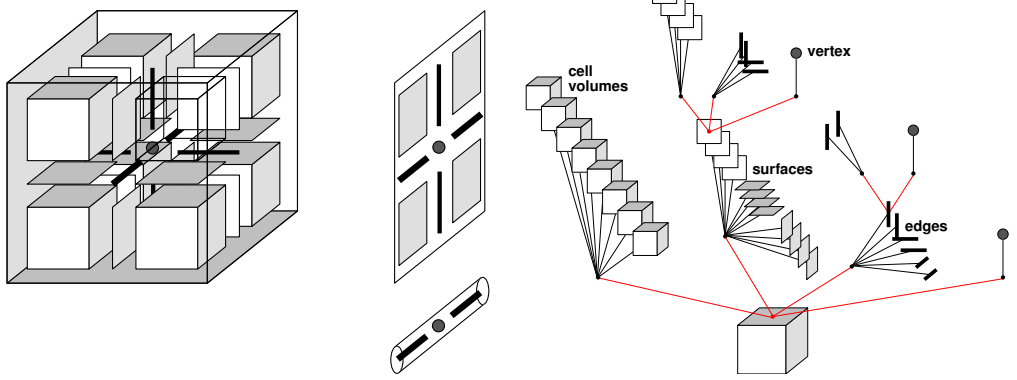
Chapter 15

Applications in 3D

In Section 2.2 and Section 3.6 we described the basic concepts of the grid structure and the parallelization strategy of our 2d and 3d MHD code. Since numerical tests in 3d are very time consuming and the evaluation of the results is very difficult, the thorough testing of the numerical schemes in 2d is an important step towards performing realistic 3d simulations. We have, therefore, so far only presented numerical results using our 2d code. In this chapter we focus on 3d simulations. From the numerical schemes presented in the previous chapters, the GLM–MHD solver (cf. Chapter 8) and the Bgfix scheme (cf. Chapter 9) have so far been included in our 3d code. For the local grid adaptation we use a strategy simular to that described in Section 3.5.
In Section 15.1 we focus on two aspects of the design of our 3d code, namely on the organization of the grid and on the load balancing process. In Section 15.2 we investigate the efficiency of the adaptation strategy by studying the error to runtime ratio of the scheme on a series of locally adapted grids and a series of globally refined grids. We show results both for tetrahedral and hexahedral meshes. As in our tests of the 2d scheme, we use planar Riemann problems (cf. Section 6.2.1) for which we either know the solution or can with high accuracy compute a reference solution. In Section 15.3 we present results for a solar physical application using a setup supplied by M. Rempel [Rem01]. Our main focus is not the physical interpretation of the results (which lies outside the scope of this investigation), but the study of the efficiency of our numerical scheme, the adaptation, and the load–balancing strategy in the case of a realistic 3d simulation. Most of the results shown here are taken from [DRSW02]; preliminary results are published in [DKRW03].

15.1 The Structure of our 3d MHD Code

In the following we focus on two aspects of the organization of the hierarchical grid in our distributed memory code; a full description of the grid structure and the underlying concepts of the algorithms can be found in [Sch99a]. For the following discussion consider a grid $\mathcal{T}$ with index sets $\mathcal{I}$ and $\mathcal{I}_S$ for the elements and faces, respectively. The partitioning of the grid is performed only on the macro grid level, which we denote with $\mathcal{T}^0$ (with corresponding index sets $\mathcal{I}^0$ and $\mathcal{I}^0_S$). For a simulation on P processors consider a given partitioning $(\mathcal{T}^0_p)_{p=1}^P$ of $\mathcal{T}^0$ with corresponding index subsets $\mathcal{I}^0_p$ of $\mathcal{I}^0$. Since the



(a) refinement of an element of hexahedral type and refinement of all objects on its boundary

(b) elements (faces and edges) administrate all objects that are created in their interior during refinement

Figure 15.1: Local administration of all geometrical objects facilitates an efficient refinement and coarsening of the grid. (Taken from [Sch99a].)

computational domain $\Omega = \cup_{i \in \mathcal{I}^0} T_i$, the partitioning of $\mathcal{I}^0$ also induces a partitioning of Ω:

$$\Omega_p := \bigcup_{i \in \mathcal{I}_p^0} T_i \quad \text{for } 1 \leq p \leq P \ .$$

Surfaces S_{ij} with $i \in \mathcal{I}_p^0$ and $j \in \mathcal{I}_q^0$ and $p \neq q$ are called *inner boundaries* in the following. With $l(i)$ for $i \in \mathcal{I}$ we denote the level of the element T in the hierarchy, i.e., the length of the path from the macro grid level (where all elements have zero level). Note that in contrast to our 2d code, we do not enforce conformity of the grid so that hanging nodes are permitted. We only require that for two neighboring elements T_i and T_j the difference in level is not greater than one: $|l(i) - l(j)| \leq 1$.

15.1.1 Grid Storage and Adaptation

In our 3d code we store all geometrical objects of the grid in tree–like structures using the following abstract classes: element (tetrahedron, hexahedron), surface and boundary surface (triangle, quadrangle), edge, and vertex. In parentheses we write the derived classes that are implemented in our code. For a given object it is always possible to identify all objects that belong to its boundary. We store each object, not in global lists or tree–like structures, but rather *locally* according to its geometrical location. If, for example, an element T is split during the refinement process, then not only are the new elements $T_1, \ldots, T_s$ administrated by the element T, but also all other new geometrical objects — surfaces, edges, and vertices — lying in the interior of T. Similarly, new object generated during the refinement of a surface $S \subset \partial T$ are administrated by the surface S itself. With this strategy, it is always possible to identify all objects that lie in the interior of a given object. In Figure 15.1 we show the situation after the refinement of a hexahedron.

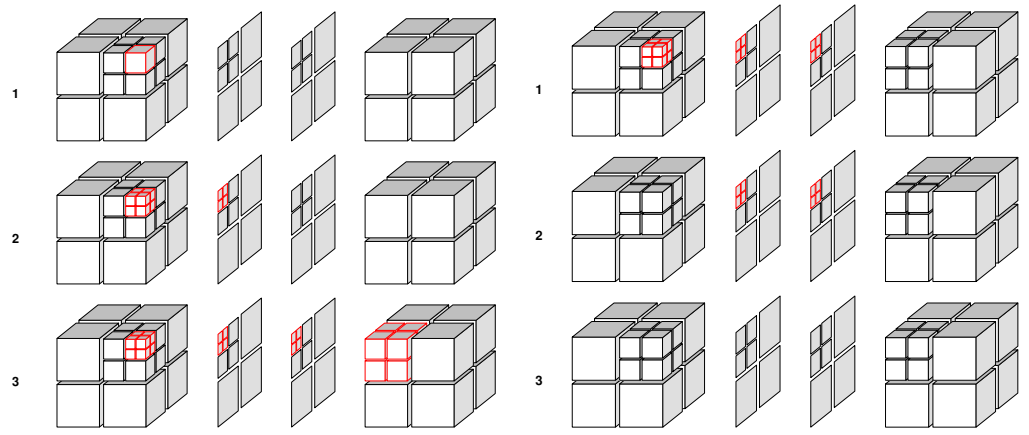
The storage of objects in *local* tree–like structures facilitates an efficient implementation of the coarsening process: we can simply remove all objects that are stored locally in an element that is to be coarsened, and we obtain a new grid without having to pay special attention to the objects of lower dimension. In Remark 15.2 at the end of this section we discuss some further advantages of this storage strategy.
Some parts of our numerical scheme, such as the flux computation, act on the surfaces S_{ij} of the grid. Since the numerical flux depends on the values of the approximate solution on both elements T_i and T_j adjacent to S_{ij}, it must be possible to access both T_i and T_j from the surface S_{ij}. Due to the construction of our grid and the definition of the faces S_{ij}, this presents no problem. However, note that the surfaces are the only objects of dimension $d-1$ that can be uniquely associated with a fixed number of objects of dimension d for any $d \in \{1,2,3\}$ (cf. Section 2.2); for instance, the number of associated elements or surfaces of an edge is not known a priori. Therefore, it is not possible to access all elements or faces that are adjacent to a given edge.

15.1 Remark: *It turns out that object-oriented concepts greatly facilitate the administration of (hierarchical) meshes. In our case, object orientation is, also, useful from a second point of view. In the finite–volume scheme, numerical integration is the basic numerical tool. Quadrature rules, however, depend on the type of geometrical object considered. Object orientation makes a simple treatment of all objects possible, as long as they are equipped with suitable integration methods. The same is true in the case of the ESC–method. Here one of the main tasks is the computation of the intersection of a characteristic (through a given point* $\mathbf{p} \in T$*) with the inflow boundary of the element* T*. Together with the evaluation of the basis functions, this is the only part of the ESC–method that depends on the geometry of the element* T*. Again object orientation can be a considerable help in simplifying the implementation of the algorithm.*

A key feature of the parallel algorithm is the handling of the refinement and coarsening process for elements adjacent to an inner boundary of a partition. An inner boundary element is artificial and corresponds to a single surface in the serial algorithm. Consequently, the grid should be conform across any inner boundary, and the inner boundaries should be refined and coarsened following the rules used for the surfaces in the serial algorithm. To illustrate the adaptation process, we add two figures that display the necessary steps for a typical situation of refinement and coarsening in a hexahedral mesh. Figure 15.2(a) shows the refinement of a hexahedron (marked in red) on one side of an inner boundary. The refinement of this element leads to the refinement of an element in the other partition with the result that, after the refinement is completed, the grid is conform over the inner surface. The coarsening process is illustrated in Figure 15.2(b): the red hexahedron in the left partition is to be coarsened and, as in the serial case, both boundary surfaces have to be coarsened, as well.

15.1.2 Load Balancing

To ensure an efficient algorithm, the parallelization strategy discussed in Section 3.6.2 relies on a distribution of the grid that ensures that each processor requires about the same amount of time to evolve the approximation from one time level to the next. Due to local grid adaptation, a balanced distribution may become unbalanced as parts of



(a) refinement **(b)** coarsening

Figure 15.2: Refinement and coarsening process for partitions with an inner boundary in common. The elements and boundary surfaces of the two partitions are displayed. The hexahedron (marked in red) in the first line has to be refined or coarsened. (Taken from [Sch99a].)

the domain are refined and others are coarsened. This requires a repartitioning of the grid.
To compute a suitable partitioning of a given macro grid $\mathcal{T}^0$, we estimate the computational cost associated with a partition $\mathcal{T}_p^0$ by the number of

(i): elements $T \in \mathcal{T}$ with $T \subset \Omega_p$ and

(ii): surfaces S_{ij} with $(i,j) \in \mathcal{I}_S$, $T_i, T_j \subset \Omega_p$.

Since the number of surfaces scale with the number of elements, the computational effort for one partition is given by $\sum_{i \in \mathcal{I}_p^0} \alpha(i)$, where $\alpha(i)$ is a measure of the computational cost for one element T_i for $i \in \mathcal{I}_p^0$:

$$\alpha(i) = |\{k \in \mathcal{I} \,:\, T_k \subset T_i\}|. \tag{15.1}$$

We also estimate the cost of the numerical scheme associated with a surface S_{ij} of the macro grid $\mathcal{T}^0$ by the number of

(i): surfaces of the computation grid $\mathcal{T}$ that lie in S_{ij} and are, therefore, not in any element of the macro grid.

Consequently, this part of the computational cost is proportional to

$$\beta(i,j) = |\{(k,l) \in \mathcal{I}_S \,:\, S_{kl} \subset S_{ij}\}| \tag{15.2}$$

for $(i,j) \in \mathcal{I}_S^0$. For each surface S_{kl} that is a subset of an inner boundary S_{ij} there has to be an exchange of data and the numerical flux $\mathbf{g}_{kl}$ is computed twice (cf. [DRSW02]). This leads to additional computational cost that is proportional to $\beta(i,j)$ and that we attempt to minimize by choosing an appropriate partitioning of the grid. Our analysis

so far has included neither the cost of the synchronization steps required prior to any global operation (e.g. the computation of the next time step Δt), nor the cost incurred by the necessary exchange of data between processors. However, no redistribution of the grid, can make up for the loss in efficiency caused by this parallel overhead, unless we use some specific knowledge of the underlying hardware.
Based on the considerations raised so far, we compute a partitioning of a macro grid $\mathcal{T}^0$ using standard methods developed for graph partitioning. For a macro grid $\mathcal{T}^0$ we introduce the weighted graph $\mathcal{G} = (\mathcal{I}^0, \mathcal{I}^0_S, \alpha, \beta)$. The functions $\alpha : \mathcal{I}^0 \to \mathbb{N}$ and $\beta : \mathcal{I}^0_S \to \mathbb{N}$ are the weight functions for the nodes ($\mathcal{I}^0$) and the edges ($\mathcal{I}^0_S$) of the graph, respectively. Define the set $\mathcal{A}$ of partition functions $\pi : \mathcal{I}^0 \to \{1, \ldots, P\}$ that satisfy

$$\sum_{i \in \mathcal{I}^0,\, \pi(i)=p} \alpha(i) - \frac{1}{P} \sum_{i \in \mathcal{I}^0} \alpha(i) \leq \max_{i \in \mathcal{I}^0} \alpha(i)$$

for all $p \in \{1, \ldots, P\}$. We define the functional $F : \mathcal{A} \to [0, \infty)$ that measures the computational cost caused by the inner surfaces of the partitioning based on $\pi \in \mathcal{A}$:

$$F[\pi] \;=\; \sum_{(i,j) \in \mathcal{I}^0_S,\, \pi(i) \neq \pi(j)} \beta(i,j) \;.$$

Now consider the following discrete optimization problem: find a $\bar{\pi} \in \mathcal{A}$ with

$$F[\bar{\pi}] = \min_{\pi \in \mathcal{A}} F[\pi]. \tag{15.3}$$

A solution to (15.3) leads to a partitioning of $\mathcal{T}^0$ where the computational cost due to the inner boundary surfaces is minimized under the condition that all processors have approximately the same load. It is not clear whether a solution to problem (15.3) exists. We approximate a solution $\bar{\pi}$ using standard codes for graph partitioning ([KK, PD]). In our algorithm load balancing is not performed in each time step. Instead we fix a number $\tau > 1$ and balance the load if and only if there is a $\widetilde{p} \in \{1, \ldots, P\}$ with

$$\sum_{i \in \mathcal{I}^0,\, \pi(i)=\widetilde{p}} \alpha(i) \geq \frac{\tau}{P} \sum_{i \in \mathcal{I}^0} \alpha(i) \;. \tag{15.4}$$

15.2 Remark:

(i): For $p \in \{1, \ldots, P\}$ let $\Theta(p)$ denote the total runtime for one time step of the algorithm on the partition Ω_p and let $\bar{\Theta} = \sum \Theta(p)/P$ be the corresponding mean value. If we neglect the cost $\beta(i,j)$ for $(i,j) \in \mathcal{I}^0_S$, the criterion (15.4) *is equivalent to*

$$\Theta(\widetilde{p}) \geq \tau \bar{\Theta}. \tag{15.5}$$

(ii): The functions α and β defined in (15.1) *and* (15.2)*, respectively, are only an estimate of the computational effort. However, as we show in the following numerical examples, these simple estimates turn out to be sufficient in our case.*

(iii): Due to the local administration of all geometrical objects of the grid, the redistribution of the grid can be performed very efficiently. As we show in our numerical examples, the removal of parts of the grid and the rebuilding of the grid on a different processor is not too time consuming.

Ryu–Jones Riemann Problem ($\gamma = 5.0/3.0$; $x_{\max} = 2.25$, $T = 0.8$)								
	ρ	u_x	u_y	u_z	B_x	B_y	B_z	p
$x < 0$	1.08	1.20	0.01	0.5	2.0	3.6	2.0	0.95
$x \geq 0$	0.98911301	−0.01312230	0.02693733	0.01003856	2.0	4.02442111	2.00259931	0.97158833
Dai–Woodward–Tóth Riemann Problem (cf. Problem 6.2(RPDWT)) ($\gamma = 5.0/3.0$; $x_{\max} = 1.25$, $T = 0.16$)								
	ρ	u_x	u_y	u_z	B_x	B_y	B_z	p
$x < 0$	1.0	10.0	0.0	0.0	5.0	5.0	0.0	2.0
$x \geq 0$	1.0	−10.0	0.0	0.0 5.0	5.0	0.0	1.0	

Table 15.1: Initial data for the one-dimensional Riemann problems. Within the setup of the simulation we rotate (u_y, u_z) and (B_y, B_z) by 17°.

15.2 Planar Riemann Problems in 3d

To test the efficiency of the adaptation procedure, we consider problems for the MHD system (1.1) with computable exact solutions. Within the framework of nonlinear conservation laws the natural candidates for such problems are Riemann-type problems, which we introduced in Section 6.2.1. With respect to the space variable x, the solution is a planar wave pattern consisting of elementary waves (shocks, contacts, rarefactions) separated by constant states. For the tests we treat two different choices for the left and right hand states $\mathbf{U}_{l,r}$:

(a) Ryu-Jones Riemann problem (cf. [RJ95, Wes02b]),

(b) Dai-Woodward-Tóth Riemann problem (cf. [DKK+02] and references therein).

All computations are performed on the domain $\Omega = (-1.25, x_{\max}) \times (-0.25, 0.25)^2$ up to a final time T. The corresponding states $\mathbf{U}_{l,r}$, $x_{\max}$, and T are displayed in Table 15.1. We use Dirichlet conditions on the boundaries at $x = -1.25$ and $x = x_{\max}$ and periodic boundaries elsewhere. In order to minimize the influence of mesh orientation on the quality of the approximate solution, we rotate the transversal components of $\mathbf{u}_0$ and $\mathbf{B}_0$ by 17°. The macro grid is rotated by $\tan^{-1}(1/3)^\circ$ around the z-axis; the boundary conditions are chosen according to the rotation.
At a time level n the grid is refined and coarsened using the mesh indicators $\mathrm{ref}_i^n, \mathrm{crs}_i^n$ (cf. (3.20)) which are given by

$$\mathrm{ref}_i^n = \max_{\substack{(k,l)\in\mathcal{J}_S^n \\ k=i \text{ or } l=i}} \mathrm{jmp}_{kl}^n \ , \quad \mathrm{crs}_i^n = \mathrm{ref}_i^n$$

with

$$\mathrm{jmp}_{ij}^n := 2 \max \left\{ \frac{|\rho_i^n - \rho_j^n|}{\rho_i^n + \rho_j^n}, \frac{|B_{x,i}^n - B_{x,j}^n| + |B_{y,i}^n - B_{y,j}^n| + |B_{z,i}^n - B_{z,j}^n|}{|\mathbf{B}_l + \mathbf{B}_r|} \right\} .$$

The vectors $\mathbf{B}_{l,r}$ are the magnetic field components of the initial states $\mathbf{U}_{l,r}$. An element T_i ($i \in \mathcal{J}$) is refined if $\mathrm{ref}_i^n > \underline{c}^n$ and the level $l(i)$ of the element in the hierarchy is less than some prescribed value $l_{\max}$. (Here the level of all the elements in the macro grid is zero.) An element $T_i \in \mathcal{T}$ can be coarsened if $\mathrm{crs}_i^n < \overline{c}^n$. The threshold values $\underline{c}^n > \overline{c}^n > 0$ are defined as

$$\underline{c}^n := \frac{\overline{c}^n}{2} := c_{\lim} \max_{i\in\mathcal{J}} \{\mathrm{ref}_i^n\} \ . \tag{15.6}$$

We use $c_{\lim} := 0.05$ for the Ryu-Jones problem and $c_{\lim} := 0.01$ for the Dai-Woodward-Tóth problem. As CFL number we take 0.2. All the following computations have been performed on 8 (POWER3-II@375MHz) processors of an IBM RS/6000 SP. To quantify the efficiency of the grid adaptation, we compare the results from a series of calculations on grids that are globally refined down to some level $l_{\max}$ with the results from a series on locally adapted grids (again with varying $l_{\max}$). We plot the L^1-error $e_{l_{\max}}$ for $l_{\max} \in \{0, \ldots, 5\}$ versus the cpu-time. For each series of measured L^1-errors $e_{l_{\max}}$ we add the graph of the curve $\Gamma = \{(x(s), y(s)) \,|\, s \in [0, 5)\}$, where for $l = 0, \ldots, 4$ the coordinates x, y are given by

$$\begin{aligned} x(s) &= ab^s \ , \\ y(s) &= (1 - (s-l))e_l + (s-l)e_{l+1} \ , \end{aligned} \qquad (l \le s < l+1) \ . \tag{15.7}$$

The numbers $a, b \in \mathbb{R}$ are chosen such that Γ fits the graph of the error.

15.2.1 Ryu-Jones Riemann Problem

In Figure 15.3 we display the L^1-error $e_{l_{\max}}$ versus the cpu-time for a tetrahedral and a hexahedral mesh. For both mesh types we show the results for a locally-adapted mesh (with the finest mesh level equal to $l_{\max}$) and a globally-refined mesh (with a mesh level equal to $l_{\max}$ for all elements $T \in \mathcal{T}$). The efficiency of the adaptive method is apparent for $l_{\max}$ not too small. Note that both axes in the graphs in Figure 15.3 are scaled logarithmically. We stress that the size of the memory of 8 processors is insufficient for computing the result on the globally refined tetrahedral mesh with $l_{\max} = 5$. Due to the significantly smaller number of elements the calculation on the locally adapted grid causes no problems. With increasing refinement level the difference between the two curves $e_{l_{\max}}$ and Γ decreases, indicating that we have reached the asymptotic regime of vanishing mesh diameters.
The graphs in Figures 15.4(a) and 15.4(b) show components of the approximate solutions as scatterplot on a section of the whole domain: for each element $T \in \mathcal{T}$ with barycenter $\boldsymbol{\omega}_T = (x_T, y_T, z_T)^T$ we plot the approximate solution at x_T, if $z_T \in [-0.05, 0.05]$. With both meshes we capture all elementary waves. Note that in hexahedral meshes there are many barycenters with identical x-coordinates. Thus in Figure 15.4 the resolution of the pictures corresponding to the tetrahedral and hexahedral meshes only appears to be different.
A detailed analysis of the relation between the mesh refinement levels and the position of elementary waves in the solution is displayed in Figure 15.4(c). In the case of the Ryu-Jones Riemann problem we have fast/slow waves (shock waves) and entropy/Alfvén waves (contacts). The mesh indicator detects all of these waves, and the highest mesh level is used only close to the discontinuities where the computational error is large.

15.2.2 Dai-Woodward-Tóth Riemann Problem

As our results displayed in Figure 15.5 and Figure 15.6 show, the Dai-Woodward-Tóth Riemann problem is very challenging due to big variations in the amplitude of

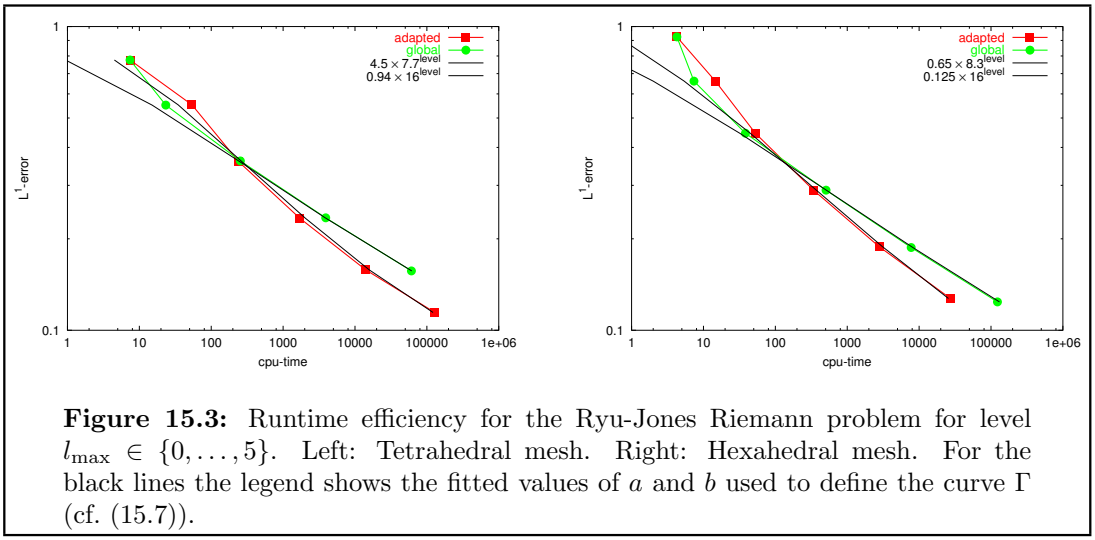


Figure 15.3: Runtime efficiency for the Ryu-Jones Riemann problem for level $l_{\max} \in \{0, \ldots, 5\}$. Left: Tetrahedral mesh. Right: Hexahedral mesh. For the black lines the legend shows the fitted values of a and b used to define the curve Γ (cf. (15.7)).

the elementary waves in the solution. Except for the fact that we display the B_y-component of the approximate solution instead of u_y, these figures correspond to those for the Ryu-Jones Riemann problem

Both mesh types – tetrahedral and hexahedral – lead to a considerable amount of oscillations close to discontinuities. However, similar oscillations also occur in two-dimensional computations (cf. Section 8.5.4). In addition we note that the mesh indicator has difficulty detecting the small-amplitude slow wave with negative speed (Figure 15.6(c)). This difficulty can be overcome by choosing a different mesh indicator (see [DRW02a]).

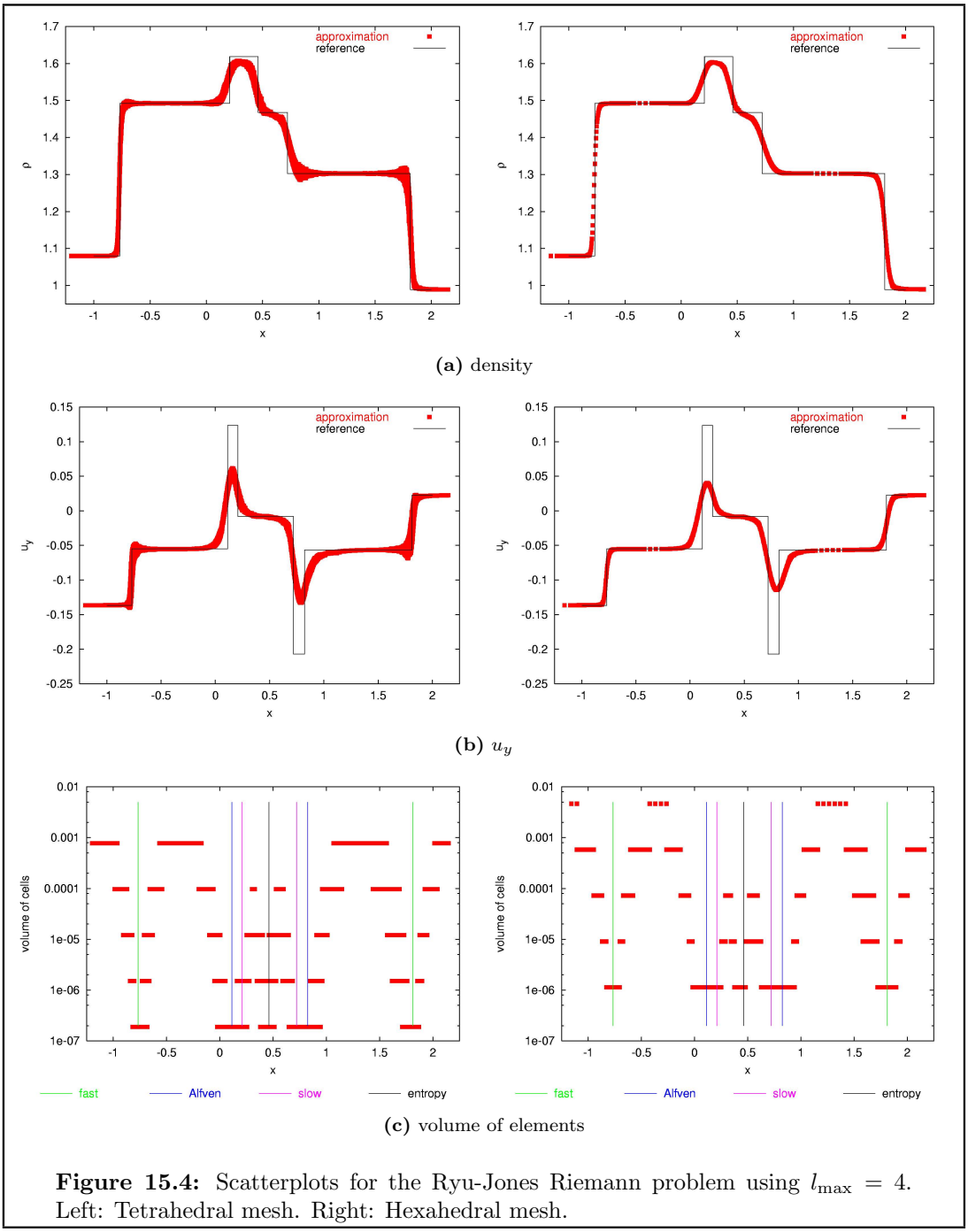
15.3 Magnetic Fluxtube in 3d

The setup for the simulation of an exploding magnetic fluxtube was supplied by M. Rempel [Rem01]. It is a model problem for the lower convection zone, and we can assume that the plasma is governed by a perfect gas law. We thus consider the system of MHD equations (1.1) together with the perfect gas law $p(\mathbf{U}) = (\gamma - 1)\rho\varepsilon$ for some $\gamma > 1$. The gravitational source term is given by $\mathbf{g}(x, y, z) = (0, 0, -g(z))$ where z denotes the height in the atmosphere. The initial values for the fluxtube are computed assuming hydrostatic equilibrium with the background atmosphere along the central field line. The entropy within the tube is higher than in the surrounding atmosphere, and the radius of the fluxtube is determined by the conservation of magnetic flux. We use the following definitions for prescribing the initial conditions $\mathbf{U}_0$:

- **background atmosphere $\mathring{\mathbf{U}}$:**
  Pressure and density are given as solutions to

$$\mathring{p}'(z) = g(z)\mathring{\rho}(z), \quad \mathring{\rho}(z) := \mathring{p}(z)^{\frac{1}{\gamma}}$$

  for $\gamma := 5/3$ and the boundary condition $\mathring{p}(0) = 1$. The height-dependent gravi-



(a) density

(b) u_y

(c) volume of elements

Figure 15.4: Scatterplots for the Ryu-Jones Riemann problem using $l_{\max} = 4$. Left: Tetrahedral mesh. Right: Hexahedral mesh.

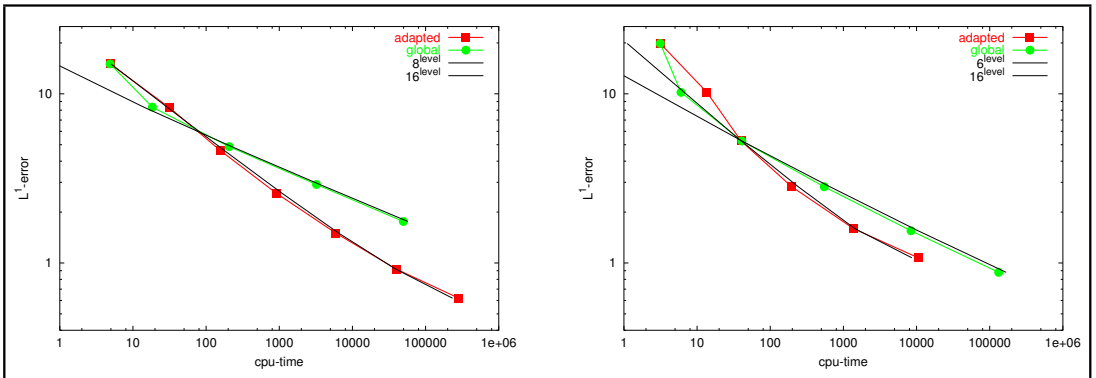


Figure 15.5: Runtime efficiency for the Dai-Woodward-Tóth Riemann problem. Left: Tetrahedral mesh. Right: Hexahedral mesh. For the black lines the legend shows the fitted values of a and b used to define the curve Γ (cf. (15.7)).

tation is given by

$$g(z) := 0.25\Big[1 + \tanh(20z)\Big]\Big[1 - \tanh\big(20(z-2)\big)\Big].$$

Furthermore $\mathring{\mathbf{B}} \equiv \mathring{\mathbf{u}} \equiv 0$.

- **magnetic fluxtube:**
  - *central field line:*

$$f(y) := -0.25 + 0.4\,\exp\Big(-\big(2(y-3)\big)^2\Big)$$

  - *magnetic pressure:*

$$p_m^z{}'(z) = -\frac{\mathring{\rho}(z)g(z)}{\gamma}\left(\frac{p_m^z(z)}{\mathring{p}(z)} - 1\right)$$

    with boundary condition $p_m^z(z_0) = 0.1\,\mathring{p}(z)$ for $z_0 := f(0)$. We use p_m as a function of y by dint of

$$p_m(y) := p_m^z\big(f(y)\big).$$

  - *radius of tube:*

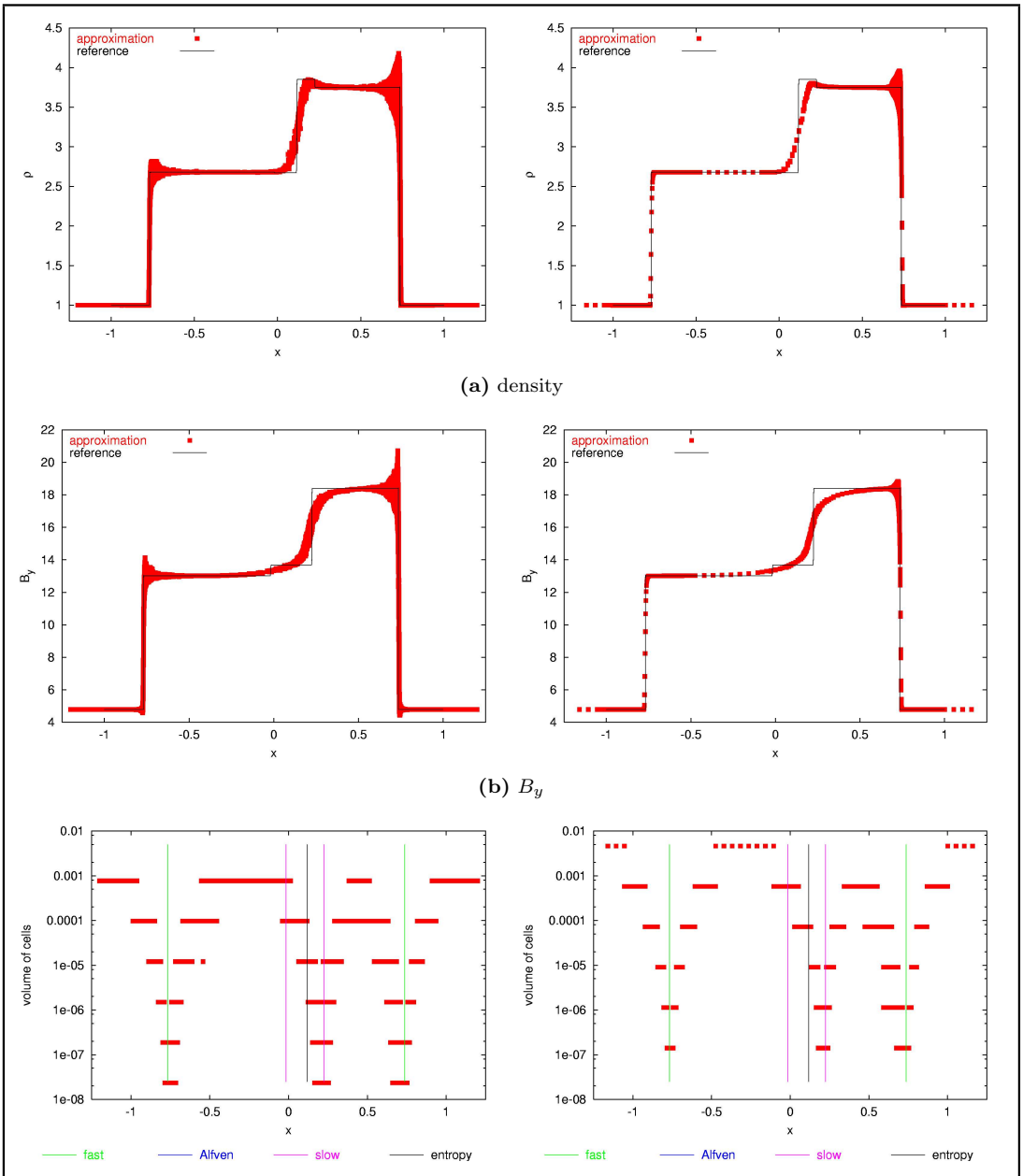
$$R(y) := 0.15\left(\frac{p_m(0)}{p_m(y)}\right)^{0.25}$$

- **initial conditions $\mathbf{U}_0$:**
  For

$$T(x,y,z) := \exp\left(-\left(\frac{(z-f(y))^2 + x^2}{R(y)^2}\right)^4\right)$$

  we define

$$B_x(x,y,z) := \frac{f(y)-z}{R(y)}\,\sqrt{8\pi p_m(y)}\,T(x,y,z),$$



(a) density

(b) B_y

(c) volume of elements

Figure 15.6: Scatterplot for the Dai-Woodward-Tóth Riemann problem using $l_{\max} = 5$. Left: Tetrahedral mesh. Right: Hexahedral mesh.

$$
\begin{aligned}
B_y(x,y,z) &:= \cos\Big(\arctan\big(f'(y)\big)\Big)\sqrt{8\pi p_m(y)}\,T(x,y,z),\\
B_z(x,y,z) &:= \left[\sin\Big(\arctan\big(f'(y)\big)\Big) + \frac{x}{R(y)}\right]\sqrt{8\pi p_m(y)}\,T(x,y,z),\\
\rho(x,y,z) &:= \mathring{\rho}(z)\left(1 - \frac{|\mathbf{B}(x,y,z)|^2}{8\pi\mathring{p}(z)}\right)^{\frac{1}{\gamma}} \exp\left(-\frac{T(x,y,z)}{\gamma}\right),\\
u_x(x,y,z) &:= 0, \quad u_y(x,y,z) := 0, \quad u_z(x,y,z) := 0,\\
\big(\rho e\big)(x,y,z) &:= \frac{1}{\gamma-1}\left(\mathring{p}(z) + \frac{(\gamma-2)\,|\mathbf{B}(x,y,z)|^2}{8\pi}\right).
\end{aligned}
$$

- **computational domain Ω:** $(-1.5, 1.5) \times (-3.0, 9.0) \times (-0.6, 2.4)$.
- **boundary conditions:**
  Periodic boundary in y-direction. Across the remaining boundaries it is assumed that the *normal* components of the moments and *all* the components of the magnetic field maintain their absolute value but change their sign; the remaining components are simply copied to ghost cells.

For the simulation of this problem we use the GLM–MHD method to cope with divergence errors (cf. Chapter 8) and the Bgfix method to stabilize the stratified background atmosphere (cf. Chapter 9). As base scheme we use the HLLEM–MHD Riemann solver (cf. [Wes02b]). The interior of the fluxtube can be characterized by its higher entropy

$$s(\mathbf{U}) := \ln(p(\mathbf{U})/\rho^\gamma)\ .$$

Therefore the boundary of the fluxtube is best determined by using the jump indicator

$$\mathrm{jmp}_{ij}^n := |s(\mathbf{U}_i^n(\mathbf{z}_{ij}) - s(\mathbf{U}_j^n(\mathbf{z}_{ij}))| \tag{15.8}$$

to perform the local grid adaptation. For the threshold values we use (15.6) with $c_{\mathrm{lim}} = 0.015$. Again all computations described below were performed on an IBM RS/6000 SP architecture using a CFL number equal to 0.2.
Initially, the *total pressure* within the tube (consisting of gas pressure p plus magnetic pressure $|\mathbf{B}|^2/(8\pi)$) is in equilibrium with the gas pressure of the background atmosphere, which is itself equal to the total pressure because the atmosphere contains no magnetic field. As the tube rises the gas pressure in the fluxtube becomes equal to the gas pressure in the surrounding atmosphere since the background pressure is monotone decreasing. Therefore at a certain height, the equilibrium of the total pressures can no longer be maintained. This leads to an "explosion" of the central part of the tube. A time sequence of our numerical results is shown in Figure 15.7. The film on the enclosed CD shows the full dynamics of the exploding fluxtube, the mesh adaption, and the partitioning. The physical interpretation of the results is beyond the scope of our investigation. For details concerning the physical background and interpretation we refer to [Rem01].
We use this problem as a test case for our 3d code, beginning with a demonstration of the effectiveness of both the GLM–MHD schemes and the Bgfix method in a realistic 3d setting. Then we study the effects of the load balancing strategy described in

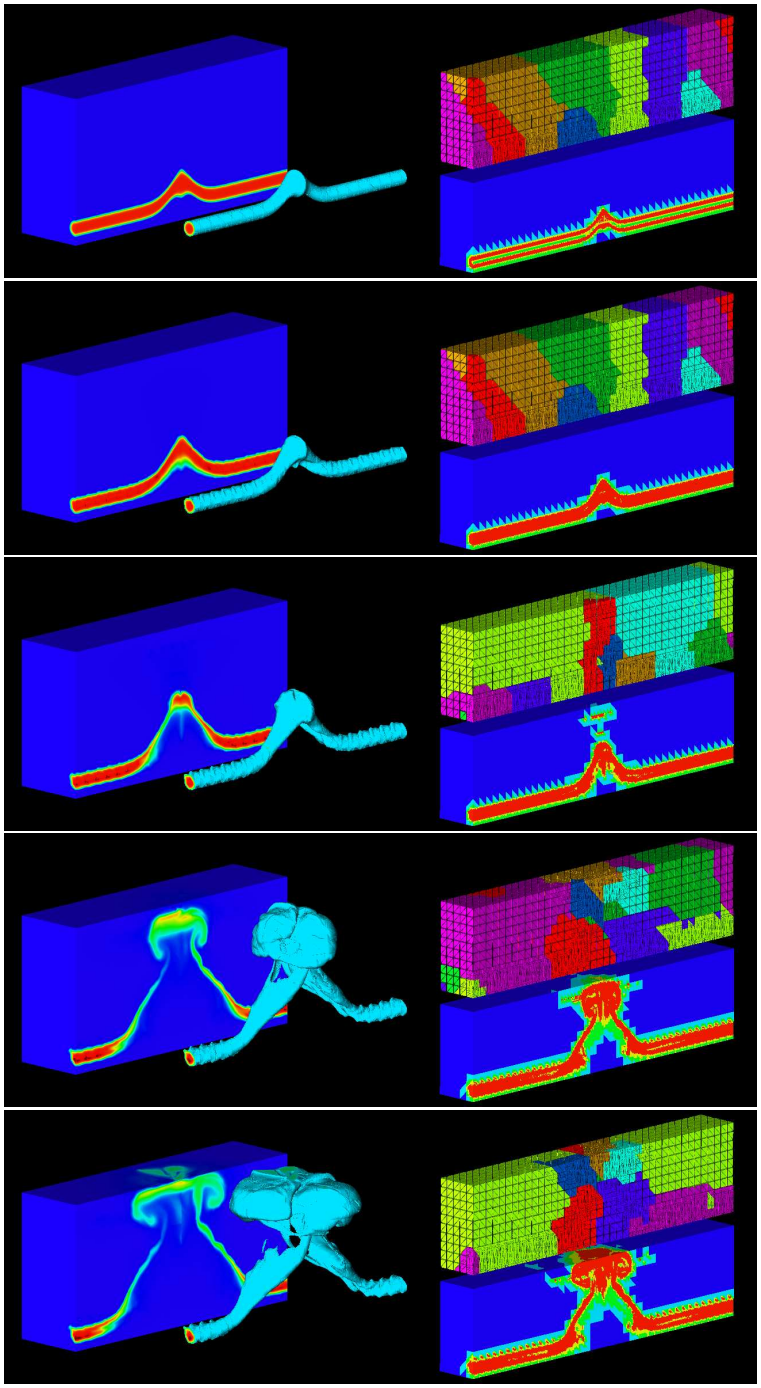
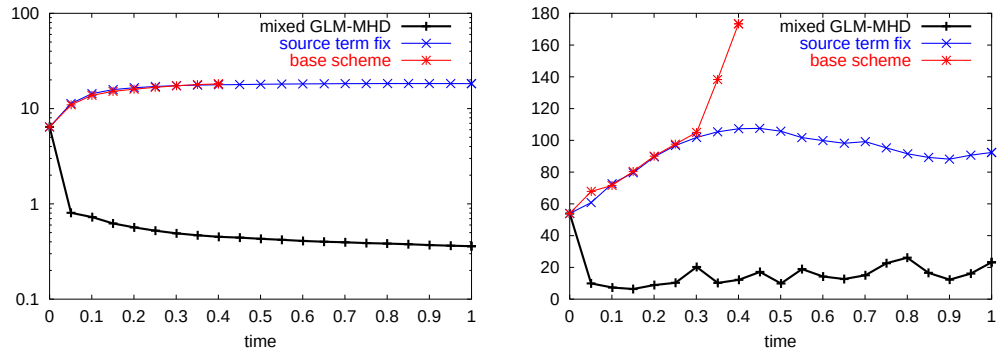


Figure 15.7: Results for the exploding magnetic fluxtube at times $t = 0.0$, $t = 3.0$, $t = 4.5$, $t = 6.0$, and $t = 7.5$ (top to bottom). Each picture contains two visualizations of the entropy in the central part of the domain (left) and, in a cross section of the full domain, the grid partitioning together with the fineness of the grid (right).



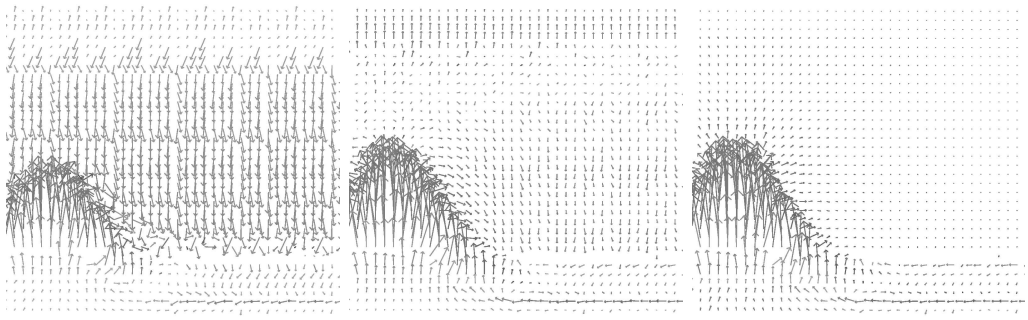
(a) L^1–error **(b)** L^∞-error

Figure 15.8: Evolution of divergence error for the exploding magnetic fluxtube. We show results for the base scheme, the mixed GLM–MHD scheme, and the source term fix.

Section 15.1.2. A further important issue is the efficiency of the whole parallel code. At the end of this section, we measure the speedup and the efficiency of the code using the technique described in Section 3.7.3.

15.3.1 The GLM–MHD and the Bgfix Schemes

We compute the 3d fluxtube using both the Bgfix scheme (cf. Chapter 9) and the mixed GLM–MHD scheme (cf. Chapter 8). Together with the mixed GLM–MHD scheme we also test the source term fix, which can likewise be added to be base scheme without difficulty. In Figure 15.8 we plot the evolution of the divergence errors at the beginning of the simulation. Without correction the base scheme breaks down quite early in the simulation ($t \approx 0.44$) due to negative pressure values. As in our 2d tests we can clearly see that the GLM–MHD method leads to a significant reduction in the magnitude of the divergence errors. The source term fix leads to a stabilization of the scheme, but — at least in L^1 — barely reduces the error compared to the base scheme.
Now let us turn our attention to the Bgfix scheme. The adaptation indicator (15.8) based on the entropy s does not lead to good result when used without the Bgfix scheme. If we use the same choice for the adaptation as in the results shown so far but without the Bgfix modification, the size of the grid quickly increases to over eight million elements (instead of about one million). We, therefore, modified the indicator for the adaptation using all the conservative quantities. In Figure 15.9 we show a comparison of the velocity vector field using the numerical scheme with and without the Bgfix correction. We show results in which we chose the adaptation parameters in such a way that the unmodified base scheme and the Bgfix modification lead to approximately the same grid with about 1.4 million elements (Figure 15.9(a) and Figure 15.9(c), respectively). With this choice for the error indicator and parameters the grid in the region away from the fluxtube remains on the macro grid level. In combination with the Bgfix modification this presents no problem, and the velocity vector field remains close to zero away from the fluxtube. In the plot using the unmodified base scheme it can be clearly seen that the background atmosphere is moving downwards. In Figure 15.9(b)



(a) without Bgfix modification (coarse grid, 1.4 million elements)

(b) without Bgfix modification (fine grid, 2.8 million elements)

(c) with Bgfix modification (coarse grid, 1.4 million elements)

Figure 15.9: Velocity vector field for the exploding magnetic fluxtube problem at time $t = 2$. The results show the importance of using a scheme that takes into account the balance of the background atmosphere. Note that the vectors are identically scaled in all plots.

we show a result using the base scheme, but this time all elements have been refined down to level two before the start of the computation and are not coarsened beyond that level during the simulation. This leads to a grid with approximately twice the number of elements (2.8 million). A shift in the atmosphere is still clearly visible, although it is not so extreme as in the simulation using the coarse grid. The fluxtube looks similar to the Bgfix scheme approximation (but note the lower number of elements).

15.3.2 Efficiency of the Load Balancing Strategy

Figure 15.10 shows the effect of the load balancing routine on the mesh partitioning for two different situations. Within each column, load balancing is performed exactly once (between the pictures in the middle and on the bottom, see also the corresponding periods in Figure 15.11). In Figure 15.10(a) we observe that load balancing leads to an almost complete redistribution of the mesh, while in the situation shown in Figure 15.10(b) the algorithm produces an equal distribution without significant alterations in the mesh topology. To analyze the performance of the parallel algorithm, in particular of the load balancing, we display in Figure 15.11 the graphs of the complete temporal evolution of a number of quantities. In the upper figure we see that the total number of elements increases during the simulation. This reflects the fact that the fluxtube expands and that the (approximate) solution develops more and more structures that are detected by the mesh indicator. The increase in the total runtime is proportional to the number of elements. This is the desired effect of the load balancing strategy. The peaks in the total runtime occur each time load balancing is executed, i.e. when condition (15.4) is satisfied. In the lower part of Figure 15.11 we show the detailed evolution of the runtime and the number of elements close to a time where load balancing is performed. Directly before the load is balanced, the maximum runtime increases faster than the number of elements, while the minimum runtime stays more or less constant. After load balancing both curves move in line with the number of

K	$S_{\text{opt}}(K;4)/\tau$	$S_{\text{opt}}(K;4)$	$S(K;4)$	$E(K;4)$
		$t \in [0,6]$		
8	1.67	2	1.88	0.94
16	3.33	4	3.51	0.88
32	6.67	8	6.08	0.75
		$t \in [3,6]$		
8	1.67	2	1.92	0.96
16	3.33	4	3.56	0.89
32	6.67	8	6.25	0.78

Table 15.2: Speedup and efficiency for the parallel code performed on K processors. $S_{\text{opt}}(K;4)$ denotes the theoretically possible optimal speedup.

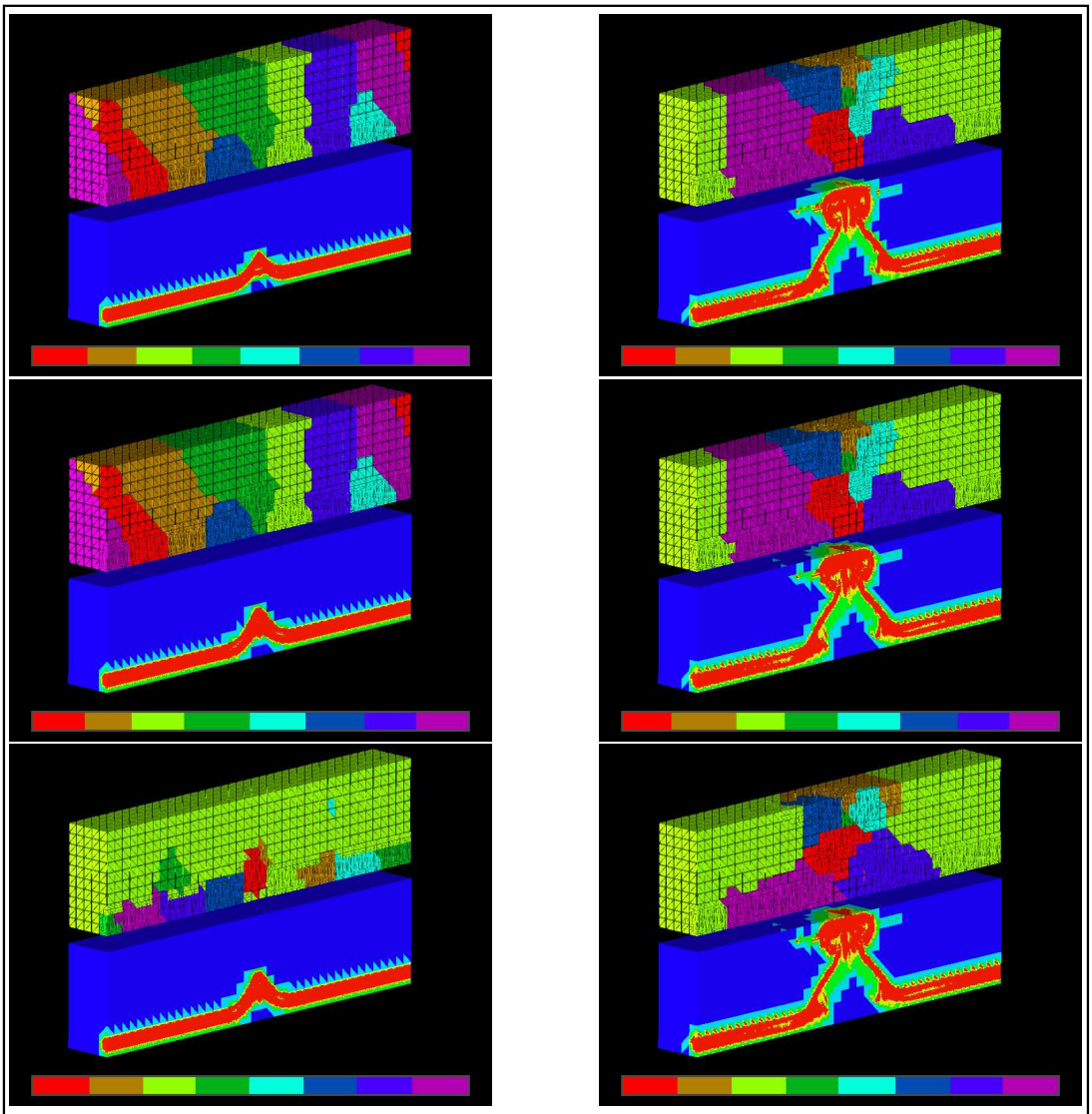
elements and the average runtime.

15.3.3 Efficiency of the Parallel Implementation

We perform computations on 4, 8, 16, and 32 processors with $\tau = 1.2$ (cf. (15.4)). In Table 15.2 we show the speedup $S(K;L)$ and the efficiency $E(K;L)$, fixing $L = 4$ (cf. Definition 3.5). We also include the optimal speedup $S_{\text{opt}}(K;L)$ and a second value that, under simplified assumptions, leads to a lower bound for the expected speedup. As discussed in Section 15.1.2, we repartition the grid only if condition (15.4) holds, i.e., using the parameter τ we allow the partitions to become unbalanced up to a certain extent. The effect of this strategy can be observed in Figure 15.11, where we see that the minimum and maximum runtime are allowed to diverge up to a certain degree before the grid is repartitioned. In Remark 15.2 we already noted that condition (15.5) can be seen as a simplified version of condition (15.4). First assume that we have no load balancing, but that condition (15.5) holds with equality for *all* time steps. Then the speedup $S(K;L)$ is bounded from above by $S_{\text{opt}}(K;L)/\tau < S_{\text{opt}}(K;L)$. If load balancing is performed using condition (15.5), a lower (theoretical) bound for the speedup is given by

$$S_{\text{opt}}(K;L;\tau) = S_{\text{opt}}(K;L)/\tau \ . \tag{15.9}$$

In Table 15.2 we observe a better speedup for the time interval $t \in [3,6]$ since the (time) averaged number of elements is much higher than for the interval $t \in [0,6]$. In the optimal case the speedup $S(K;L)$ should be the interval $[S_{\text{opt}}(K;L)/\tau, S_{\text{opt}}(K;L)]$ as discussed above. In our results this is not always the case; this is not due to the failure of the load balancing strategy, but rather to the fact that the costs of synchronization cannot be ignored as we did in the derivation of the theoretical bounds.



(a) $t = 0.5$, $t = 1.85$. and $t = 1.9$

(b) $t = 6.2$, $t = 6.6$, and $t = 6.65$

Figure 15.10: Effect of the load balancing algorithm during the simulation of the exploding magnetic fluxtube. Each picture displays the generated mesh partitioning, the fineness of the grid, and the *relative* size of the partitions. On the left and the right we show two time sequences around points where the grid is repartitioned.
(a): In the first line the fourth partition (color: dark green) lies right in the center of the domain and in this region the grid is refined due to the rise of the fluxtube. This leads to an disproportional growth of the partition size, which is compensated by the load balancing strategy. In this case load balancing leads to a complete alteration of the grid partitioning.
(b): In this case the grid size increases in partitions number two and five (colors: brown and light blue). This time the redistributions remain quite local.

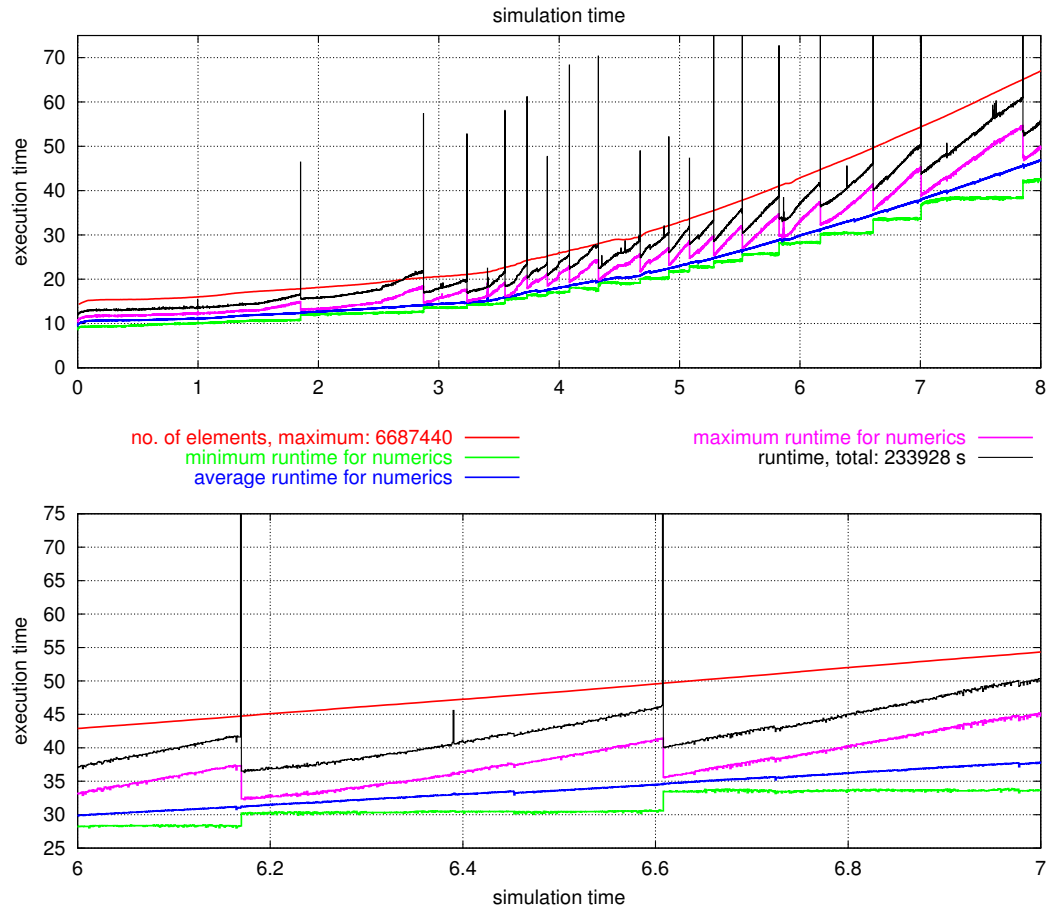


Figure 15.11: Effect of the total number of elements and of the load balancing on the runtime for the whole simulation time (upper picture), together with an excerpt for $t \in [6, 7]$ (lower picture).

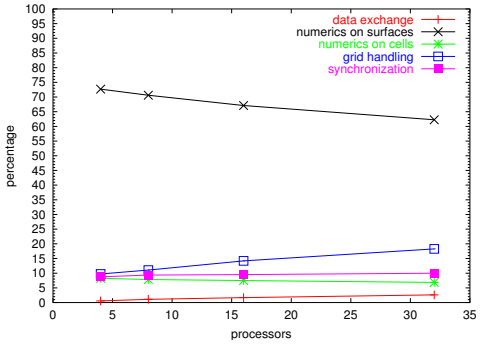
Conclusions and Outlook

We presented an explicit finite–volume scheme for solving the coupled system (1.19) of radiation magnetohydrodynamics consisting of the system of magnetohydrodynamics and the radiation transport equation. We implemented the scheme in one, two, and three space dimensions, in a first and second order version on structured and unstructured grids. The starting point of our discussion was a standard finite–volume scheme for solving the equation of ideal magnetohydrodynamics, based on the very efficient MHD–HLLEM approximate Riemann solver for a perfect gas pressure law. We first modified this method to include more general pressure laws: in Chapter 7 we presented a relaxation mechanism (**ER scheme**) that requires only a minor modifications of the base scheme since it uses the flux function for the perfect gas law. The remaining modifications were aimed at improving the robustness and accuracy of the scheme. In Chapter 8 we introduced different methods (**GLM–MHD schemes**) that take the divergence constraint on the magnetic field into account; the method presented in Chapter 9 improves the approximation of solutions near an equilibrium state (**Bgfix scheme**). These methods require only small modifications of the flux computation and were, therefore, easy to add to our base scheme. In fact, none of the modification presented in this study are restricted to the finite–volume framework; they can be used in other schemes as well. In Chapter 12 and Chapter 13 we studied methods (**(C)ESC schemes**) for computing the radiation source term (1.19g). We found that the ESC framework leads to very efficient schemes that can be easily coupled to a scheme for solving the equations of magnetohydrodynamics.

To study the quality of our methods, we had to rely primarily on numerical tests since analytical results for complex systems of balance laws are scarce. For an analytical justification of the schemes, we studied simplified settings: for a scalar model problem derived from the coupled system (1.19), we were able to establish the **convergence** of a first order finite–volume approximation in L^1; we studied the **convergence** of the Bgfix scheme applied to a scalar balance law; furthermore, we have proven the **convergence** in L^∞ of our first order ESC approximation to the radiation transport equation. Since we already presented a complete summary of both our analytical and numerical results in Chapter 5, Chapter 10, and Chapter 14, we do not have to go into detail here.

In the design of every part of our scheme, we followed two central guidelines. One concern was their efficiency measured by the error to runtime ratio. The computational cost involved in simulating the evolution of a plasma — including the energy transport by radiation — is very high; therefore, the efficiency of the scheme is a central aspect. Since our scheme falls into two parts, one for evolving the fluid quantities and the

other for computing the radiation field, we studied these two parts separately. We took care not to introduce any new stability restriction into the base scheme and not to increase its computational cost either. For the solution of the MHD equations, the flux computation is by far the most expensive part (cf. the following Figure). Since all the extensions of the MHD part of our scheme require a modification of the numerical flux, a significant increase in the cost of the flux computation would significantly reduce the efficiency of the whole scheme. Therefore, it was one of our main concerns that our modifications lead to very little additional cost.
The second guideline was that we wanted to develop a stable and robust scheme for a wide range of applications. Therefore, we attempted to make good and analytically justified choices for all the free parameters. In the case of some parameters that are necessary for the local grid adaptation and for the CFL number, however, we still have to make arbitrary choices. With a CFL number that depends only on the space dimension (0.4 in 2d, 0.2 in 3d) the scheme remained stable in all our tests. The parameters used for the grid adaptation depend to a certain extent on the setting of the simulation and have to be determined by numerical tests. Here further analysis is necessary to increase the efficiency of the scheme.



Percentage of runtime used for different parts of the 3d algorithm during the simulation of the fluxtube problem using a different number of processors. All our methods for increasing the performance of the finite–volume scheme lead to a modification of the numerical flux function, which is that part of the scheme that acts on the surfaces of the grid and requires by far the most computational effort. Therefore, an efficient flux calculation is the most essential tool for increasing the efficiency of the whole scheme. The amount of parallel overhead (synchronization/data exchange) increases with the number of processors since we did not change the size of the problem. However, the parallel overhead takes up only a small percentage of the overall cost.

Outlook

In the course of our cooperation with the group around Manfred Schüssler — part of the DFG priority research program ANumE — Peter Vollmöller used many of the methods presented here to simulate 2d magneto–convection in the solar photosphere as part of his PhD thesis [Vol02]; together with Matthias Rempel (also a former PhD student of Manfred Schüssler's [Rem01]), we used the methods for the simulation of a 3d magnetic fluxtube in the lower convection zone (cf. Chapter 15). We have not yet performed 3d simulations of the solar photosphere. For this purpose we still have to incoperate a method for approximating the radiation source term into our 3d code. The methods presented here can be easily used in the 3d case, but further research is required to derive

an efficient implementation: if we include the radiation transport, the computation of the radiation source term will be by far the most expensive part of our algorithm. The efficiency of the parallelization strategy using domain decomposition is uncertain, as well, because the computation of the radiation intensity for a fixed direction $\boldsymbol{\mu}$ requires traversing the whole grid, and the radiation transport module may well result in an unacceptably high amount of additional parallel overhead. To overcome this obstacle and to derive an efficient method, we intend to use an iteration strategy similar to the one used to cope with the periodic boundary conditions (cf. Chapter 12). This is a promising approach since the hydrodynamic quantities undergo only slight changes from one time step to the next, and, therefore, the change in the radiation intensity is similarly small. Consequently, the intensity at one time step t^n might be a sufficiently good starting point for computing the intensity at the next time step t^{n+1}. We sketch the structure of the algorithm resulting from these innovations in the following:

Let the discrete fluid state $\mathbf{U}^n$ *on each partition be given, together with the corresponding intensities* I_m^n *for* $m = 1, \ldots, M$*. Then on each partition*

(i): compute $\mathbf{U}^{n+1}$ from $\mathbf{U}^n$ and I_m^n,

(ii): compute I_m^{n*} using $\mathbf{U}^{n+1}$ in the interior of the domain and I_m^n as boundary data on the inner boundaries,

(iii): exchange $\mathbf{U}^{n+1}$ and I_m^{n*} at the inner boundaries,

(iv): compute I_m^{n+1} using $\mathbf{U}^{n+1}$ in the interior of the domain and I_m^{n*} as boundary data on the inner boundaries.

With this algorithm we have only one synchronization/data exchange step — thus no increase in the parallel overhead — yet we solve the radiation transport equation twice. In the second step of the algorithm, the hydrodynamic quantities in the interior of the domain have changed while the intensity on the inflow boundary remains unaltered; for step *(iv)*, we use the same hydrodynamic quantities but new inflow intensity values. We assume that we can compute an accurate approximation of the radiation field with this strategy; the study of the 3d radiation transport solver is part of our forthcoming work.

Figures

Tables

Algorithms

Schemes

Test Cases

Assumptions

Definitions

Theorems

Publications

[Ded98] A. Dedner, *Numerik und Analysis für ein gekoppeltes System der Eulergleichungen mit der Strahlungstransportgleichung in zwei Raumdimensionen*, Diplomarbeit, Albert–Ludwigs–Universität, Mathematische Fakultät, Freiburg, Juli 1998.

[DKK+02] A. Dedner, F. Kemm, D. Kröner, C.-D. Munz, T. Schnitzer, and M. Wesenberg, *Hyperbolic divergence cleaning for the MHD equations*, J. Comput. Phys. **175** (2002), no. 2, 645–673, doi:10.1006/jcph.2001.6961.

[DKR+03] A. Dedner, D. Kröner, C. Rohde, T. Schnitzer, and M. Wesenberg, *Comparison of finite volume and discontinuous Galerkin methods of higher order for systems of conservation laws in multiple space dimensions*, Geometric Analysis and Nonlinear Partial Differential Equations (S. Hildebrandt and H. Karcher, eds.), Springer, Berlin, 2003, pp. 573–589.

[DKRW01a] A. Dedner, D. Kröner, C. Rohde, and M. Wesenberg, *Godunov–type schemes for the MHD equations*, Godunov Methods: Theory and Applications (E.F. Toro, ed.), Kluwer Academic/Plenum Publishers, November 2001, pp. 209–216.

[DKRW01b] ———, *MHD instabilities arising in solar physics: A numerical approach*, Hyperbolic Problems: Theory, Numerics, Applications (Basel) (H. Freistühler and G. Warnecke, eds.), International Series of Numerical Mathematics, vol. 140, Birkhäuser, 2001, Eighth International Conference in Magdeburg, February/March 2000, pp. 277–286.

[DKRW02] ———, *Efficient divergence cleaning in three–dimensional MHD simulations*, High Performance Computing in Science and Engineering '02 (E. Krause and W. Jäger, eds.), Springer, Berlin, 2003, pp. 323–334.

[DKSW01a] A. Dedner, D. Kröner, I.L. Sofronov, and M. Wesenberg, *Absorbing boundary conditions for astrophysical MHD simulations*, Godunov Methods: Theory and Applications (E.F. Toro, ed.), Kluwer Academic/Plenum Publishers, November 2001, pp. 217–224.

[DKSW01b] ———, *Transparent boundary conditions for MHD simulations in stratified atmospheres*, J. Comput. Phys. **171** (2001), no. 2, 448–478.

[DR02a] A. Dedner and C. Rohde, *FV–schemes for a scalar model problem of radiation magnetohydrodynamics*, Finite Volumes for Complex Applications III: Problems and Perspectives (Paris) (R. Herbin and D. Kroöner, eds.), Hermès Science Publications, 2002, pp. 179–186.

[DR02b] ———, *Numerical approximation of entropy solutions for hyperbolic integro–differential equations*, Preprint 15, Albert–Ludwigs–Universität, Mathematische Fakultät, Freiburg, April 2002, submitted to Numer. Math.

[DRSW02] A. Dedner, C. Rohde, B. Schupp, and M. Wesenberg, *A parallel, load–balanced MHD code on locally adapted, unstructured grids in 3d*, Preprint 30, Albert–Ludwigs–Universität, Mathematische Fakultät, Freiburg, 2002, submitted to Comput. Visual. Sci.

[DRW99] A. Dedner, C. Rohde, and M. Wesenberg, *A MHD–simulation in solar physics*, Finite Volumes for Complex Applications II: Problems and Perspectives (Paris) (R. Vilsmeier, F. Benkhaldoun, and D. Hänel, eds.), Hermès Science Publications, 1999, pp. 491–498.

[DRW02a] ———, *Efficient higher–order finite volume scheme for (real gas) magnetohydrodynamics*, to appear in proceedings of the Ninth International Conference on Hyperbolic Problems: Theory, Numerics, Applications, 2002.

[DRW02b] ———, *A new approach to divergence cleaning in magnetohydrodynamic simulations*, to appear in proceedings of the Ninth International Conference on Hyperbolic Problems: Theory, Numerics, Applications, 2002.

[DRW02c] ———, *A note on analysis and numerics for radiative MHD in 3d*, submitted to the GAMM 2002 in Augsburg, 2002.

[DV02] A. Dedner and P. Vollmöller, *An adaptive higher order method for solving the radiation transport equation on unstructured grids*, J. Comput. Phys. **178** (2002), 263–289.

[DW01] A. Dedner and M. Wesenberg, *Numerical methods for the real gas MHD equations*, Hyperbolic Problems: Theory, Numerics, Applications (Basel) (H. Freistühler and G. Warnecke, eds.), International Series of Numerical Mathematics, vol. 140, Birkhäuser, 2001, Eighth International Conference in Magdeburg, February/March 2000, pp. 287–296.

Bibliography

[Ada99] M.L. Adams, *Short characteristic solution to neutral and photon transport*, Massachussets Institute of Technology Preprint **PSFC/RR-99-7** (1999), 1.

[AG98] S. Abarbanel and D. Gottlieb, *On the construction and analysis of absorbing layers in CEM*, Appl. Numer. Math. **27** (1998), no. 4, 331–340.

[AP94] L. Auer and F. Paletou, *Two-dimensional radiative transfer with partial frequency redistribution*, Astron. Astrophys. **284** (1994), 657.

[Asl93] N. Aslan, *Computational investigations of ideal magnetohydrodynamic plasmas with discontinuities*, Ph.D. thesis, University of Michigan, 1993.

[Bal98a] D.S. Balsara, *Linearized formulation of the Riemann problem for adiabatic and isothermal magnetohydrodynamics*, Astrophys. J. Suppl. **116** (1998), 119–131.

[Bal98b] ———, *Total variation diminishing scheme for adiabatic and isothermal magnetohydrodynamics*, Astrophys. J. Suppl. **116** (1998), 133–153.

[BB80] J.U. Brackbill and D.C. Barnes, *Note: The effect of nonzero $\nabla \cdot \mathbf{B}$ on the numerical solution of the magnetohydrodynamic equations*, J. Comput. Phys. **35** (1980), 426–430.

[BD99] F. Bezard and B. Despres, *An entropic solver for ideal Lagrangian magnetohydrodynamics*, J. Comput. Phys. **154** (1999), no. 1, 65–89.

[Ber94] J.-P. Berenger, *A perfectly matched layer for the absorption of electromagnetic waves*, J. Comput. Phys. **114** (1994), no. 2, 185–200.

[BGH00] T. Buffard, T. Gallouët, and J. M. Hérard, *A sequel to a rough Godunov scheme: Application to real gases*, Comput. Fluids (2000), no. 29, 813–847.

[BJ00] A. Bressan and H.K. Jenssen, *On the convergence of Godunov scheme for nonlinear hyperbolic systems*, Chin. Ann. Math., Ser. B **21** (2000), no. 3, 269–284.

[BKP96] A.A. Barmin, A.G. Kulikovskiy, and N.V. Pogorelov, *Shock–capturing approach and nonevolutionary solutions in magnetohydrodynamics*, J. Comput. Phys. **126** (1996), no. 1, 77–90.

[BPV03] Ramaz Botchorishvili, Benoit Perthame, and Alexis Vasseur, *Equilibrium schemes for scalar conservation laws with stiff sources.*, Math. Comput. **72** (2003), no. 241, 131–157.

[BS99] D.S. Balsara and D.S. Spicer, *A staggered mesh algorithm using high order Godunov fluxes to ensure solenoidal magnetic fields in magnetohydrodynamic simulations*, J. Comput. Phys. **149** (1999), no. 2, 270–292.

[BVS99] J.H.M.J. Bruls, P. Vollmöller, and M. Schüssler, *Computing radiative heating on unstructured spatial grids*, Astron. Astrophys. **348** (1999), 233.

[BW88] M. Brio and C.C. Wu, *An upwind differencing scheme for the equations of ideal magnetohydrodynamics*, J. Comput. Phys. **75** (1988), no. 2, 400–422.

[Cab70] H. Cabannes, *Theoretical magnetofluiddynamics*, Applied Mathematics and Mechanics, vol. 13, Academic Press, New York, 1970.

[Car63] B.G. Carlson, *The numerical theory of neutron transport*, vol. 1, Alder, Berni and Ferbach, Sidney, 1963.

[CG97] P. Cargo and G. Gallice, *Roe matrices for ideal MHD and systematic construction of Roe matrices for systems of conservation laws*, J. Comput. Phys. **136** (1997), 446–466.

[Cha81] S. Chandrasekhar, *Hydrodynamic and hydromagnetic stability*, Dover, New York, 1981.

[CHC01] C. Chainais-Hillairet and S. Champier, *Finite volume schemes for non-homogeneous scalar conservation laws: Error estimate*, Numer. Math. **88** (2001), no. 4, 607–639.

[Cia78] Philippe G. Ciarlet, *The finite element method for elliptic problems*, vol. 4, Studies in Mathematics and its Applications, Amsterdam - New York - Oxford, 1978.

[CKS00] B. Cockburn, E. Karniadakis, and C.-W. Shu, *The development of discontinuous Galerkin methods*, Lecture Notes in Computational Science and Engineering, vol. 11, Springer, Berlin, 2000.

[CLL94] G.-Q. Chen, C.D. Levermore, and T.-P. Liu, *Hyperbolic conservation laws with stiff relaxation terms and entropy*, Commmun. Pure Appl. Math **47** (1994), no. 1, 787–830.

[CMIS95a] P. Caligari, F. Moreno-Insertis, and M. Schüssler, *Emerging flux tubes in the solar convection zone. I: Asymmetry, tilt, and emergence latitude*, Astrophys. J. **441** (1995), no. 2, 886–902.

[CMIS95b] ———, *Emerging flux tubes in the solar convection zone. I: Asymmetry, tilt, and emergence latitude*, Astrophys. J. **441** (1995), no. 2, 886–902.

[CP98] F. Coquel and B. Perthame, *Relaxation of energy and approximate Riemann solvers for general pressure laws in fluid dynamics*, SIAM J. Numer. Anal. **35** (1998), no. 6, 2223–2249.

[Daf00] C. Dafermos, *Hyperbolic conservation laws in continuum physics*, first ed., Grundlehren der Mathematischen Wissenschaften, vol. 325, Springer, Heidelberg, Berlin, 2000.

[Ded98] A. Dedner, *Numerik und Analysis für ein gekoppeltes System der Eulergleichungen mit der Strahlungstransportgleichung in zwei Raumdimensionen*, Diplomarbeit, Albert–Ludwigs–Universität, Mathematische Fakultät, Freiburg, Juli 1998.

[DEO92] L.J. Durlofsky, B. Engquist, and S. Osher, *Triangle based adaptive stencils for the solution of hyperbolic conservation laws*, J. Comput. Phys. **98** (1992), no. 1, 64–73.

[DKK+02] A. Dedner, F. Kemm, D. Kröner, C.-D. Munz, T. Schnitzer, and M. Wesenberg, *Hyperbolic divergence cleaning for the MHD equations*, J. Comput. Phys. **175** (2002), no. 2, 645–673, doi:10.1006/jcph.2001.6961.

[DKR+03] A. Dedner, D. Kröner, C. Rohde, T. Schnitzer, and M. Wesenberg, *Comparison of finite volume and discontinuous Galerkin methods of higher order for systems of conservation laws in multiple space dimensions*, Geometric Analysis and Nonlinear Partial Differential Equations (S. Hildebrandt and H. Karcher, eds.), Springer, Berlin, 2003, pp. 573–589.

[DKRW01a] A. Dedner, D. Kröner, C. Rohde, and M. Wesenberg, *Godunov–type schemes for the MHD equations*, Godunov Methods: Theory and Applications (E.F. Toro, ed.), Kluwer Academic/Plenum Publishers, November 2001, pp. 209–216.

[DKRW01b] ———, *MHD instabilities arising in solar physics: A numerical approach*, Hyperbolic Problems: Theory, Numerics, Applications (Basel) (H. Freistühler and G. Warnecke, eds.), International Series of Numerical Mathematics, vol. 140, Birkhäuser, 2001, Eighth International Conference in Magdeburg, February/March 2000, pp. 277–286.

[DKRW03] ———, *Efficient divergence cleaning in three–dimensional MHD simulations*, High Performance Computing in Science and Engineering '02 (E. Krause and W. Jäger, eds.), Springer, Berlin, 2003, pp. 323–334.

[DKSW01a] A. Dedner, D. Kröner, I.L. Sofronov, and M. Wesenberg, *Absorbing boundary conditions for astrophysical MHD simulations*, Godunov Methods: Theory and Applications (E.F. Toro, ed.), Kluwer Academic/Plenum Publishers, November 2001, pp. 217–224.

[DKSW01b] ———, *Transparent boundary conditions for MHD simulations in stratified atmospheres*, J. Comput. Phys. **171** (2001), no. 2, 448–478.

[DPRV99] P. Degond, P.-F. Peyrard, G. Russo, and P. Villedieux, *Polynomial upwind schemes for hyperbolic systems*, C. R. Acad. Sci., Paris, Ser. I, Math. **328** (1999), no. 6, 479–483.

[DR02a] A. Dedner and C. Rohde, *FV–schemes for a scalar model problem of radiation magnetohydrodynamics*, Finite Volumes for Complex Applications III: Problems and Perspectives (Paris) (R. Herbin and D. Kroöner, eds.), Hermès Science Publications, 2002, pp. 179–186.

[DR02b] _____, *Numerical approximation of entropy solutions for hyperbolic integro–differential equations*, Preprint 15, Albert–Ludwigs–Universität, Mathematische Fakultät, Freiburg, April 2002, submitted to Numer. Math.

[DRSW02] A. Dedner, C. Rohde, B. Schupp, and M. Wesenberg, *A parallel, load–balanced MHD code on locally adapted, unstructured grids in 3d*, Preprint 30, Albert–Ludwigs–Universität, Mathematische Fakultät, Freiburg, 2002, submitted to Comput. Visual. Sci.

[DRW99] A. Dedner, C. Rohde, and M. Wesenberg, *A MHD–simulation in solar physics*, Finite Volumes for Complex Applications II: Problems and Perspectives (Paris) (R. Vilsmeier, F. Benkhaldoun, and D. Hänel, eds.), Hermès Science Publications, 1999, pp. 491–498.

[DRW02a] _____, *Efficient higher–order finite volume scheme for (real gas) magnetohydrodynamics*, to appear in proceedings of the Ninth International Conference on Hyperbolic Problems: Theory, Numerics, Applications, 2002.

[DRW02b] _____, *A new approach to divergence cleaning in magnetohydrodynamic simulations*, to appear in proceedings of the Ninth International Conference on Hyperbolic Problems: Theory, Numerics, Applications, 2002.

[DRW02c] _____, *A note on analysis and numerics for radiative MHD in 3d*, submitted to the GAMM 2002 in Augsburg, 2002.

[DV02] A. Dedner and P. Vollmöller, *An adaptive higher order method for solving the radiation transport equation on unstructured grids*, J. Comput. Phys. **178** (2002), 263–289.

[DW94] W. Dai and P.R. Woodward, *Extension of the piecewise parabolic method to multidimensional ideal magnetohydrodynamics*, J. Comput. Phys. **115** (1994), no. 2, 485–514.

[DW98] _____, *A simple finite difference scheme for multidimensional magnetohydrodynamical equations*, J. Comput. Phys. **142** (1998), no. 2, 331–369.

[DW01] A. Dedner and M. Wesenberg, *Numerical methods for the real gas MHD equations*, Hyperbolic Problems: Theory, Numerics, Applications (Basel) (H. Freistühler and G. Warnecke, eds.), International Series of Numerical Mathematics, vol. 140, Birkhäuser, 2001, Eighth International Conference in Magdeburg, February/March 2000, pp. 287–296.

[EGGH98] R. Eymard, T. Gallouët, M. Ghilani, and R. Herbin, *Error estimates for the approximate solutions of a nonlinear hyperbolic equation given by some finite volume schemes*, IMA J. Numer. Anal. **18** (1998), 563–594.

[EGH00] R. Eymard, T. Gallouẽt, and R. Herbin, *Finite volume methods.*, Solution of equations in $\mathbb{R}^n$ (Part 3). Techniques of scientific computing (Part 3). (P. G. et al. Ciarlet, ed.), Handbook of numerical analysis, vol. 7, Elsevier, Amsterdam, 2000.

[EMI98] T. Emonet and F. Moreno-Insertis, *The physics of twisted magnetic tubes rising in a stratified medium: Two–dimensional results*, Astrophys. J. **492** (1998), 804–821.

[FK97] C. Führer and G. Kanschat, *A posteriori error control in radiative transfer*, Computing **58** (1997), no. 4, 317–334.

[FNT01] M. Fey, S. Noelle, and C. Törne, *The MoT-ICE: a new multi-dimensional wave-propagation-algorithm based on Fey's method of transport. with application to the Euler- and MHD-equations*, Hyperbolic Problems: Theory, Numerics, Applications (Basel) (H. Freistühler and G. Warnecke, eds.), International Series of Numerical Mathematics, vol. 140, Birkhäuser, 2001, Eighth International Conference in Magdeburg, February/March 2000, pp. 373–389.

[Füh93] C. Führer, *A comparative study of finite element solvers for hyperbolic problems with applications to radiative transfer*, Preprint 09/93, Universität Heidelberg, 1993.

[FZL98] Y. Fan, E.G. Zweibel, and S.R. Lantz, *Two–dimensional simulations of buoyantly rising, interacting magnetic flux tubes*, Astrophys. J. **493** (1998), 480–493.

[Geß01] Thomas Geßner, *Dynamic mesh adaption for supersonic combustion waves modeled with detailed reaction mechanisms*, Ph.D. thesis, Univ. Freiburg im Br., Mathematische Fakultät, 2001.

[GHN02] T. Gallouët, J. M. Hérard, and Seguin N., *Some recent finite volume schemes to compute Euler equations using real gas eos*, Int. J. Numer. Methods Fluids **39** (2002), no. 12, 1073–1138.

[Giv91] D. Givoli, *Non–reflecting boundary conditions*, J. Comput. Phys. **94** (1991), no. 1, 1–29.

[Gli65] J. Glimm, *Solutions in the large for nonlinear hyperbolic systems of equations*, Commun. Pure Appl. Math. **18** (1965), 697–715.

[God72] S.K. Godunov, *The symmetric form of magnetohydrodynamics equation*, Num. Meth. Mech. Cont. Media **1** (1972), 26–34.

[Gos00] L. Gosse, *A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms*, Comput. Math. Appl. **39** (2000), no. 9-10, 135–159.

[HSS00] Paul Houston, Christoph Schwab, and Endre Süli, *Stabilized hp–finite element methods for first–order hyperbolic problems*, SIAM J. Numer. Anal. **37** (2000), no. 5, 1618–1643.

[HW91] E. Hairer and G. Wanner, *Solving ordinary differential equations II*, Springer series in computational mathematics 14, 1991.

[In99] A. In, *Numerical evaluation of an energy relaxation method for inviscid real fluids*, SIAM J. Sci. Comput. **21** (1999), no. 1, 340–365.

[Jin01] Shi Jin, *A steady-state capturing method for hyperbolic systems with geometrical source terms.*, M2AN, Math. Model. Numer. Anal. **35** (2001), no. 4, 631–645.

[JNS95] B.-I. Jun, M.L. Norman, and J.M. Stone, *A numerical study of Rayleigh–Taylor instability in magnetic fluids*, Astrophys. J. **453** (1995), 332–349.

[JP86] C. Johnson and J. Pitkäranta, *Analysis of the discontinuous Galerkin method for linear hyperbolic equations*, Math.Comp. **46** (1986), 1–26.

[JR02] V. Jovanovic and C. Rohde, *Finite-volume schemes for Friedrichs systems in multiple space dimensions: A-priori and a-posteriori error estimates*, Preprint 36, Albert–Ludwigs–Universität, Mathematische Institut, Freiburg, 2002.

[JX95] S. Jin and Z. Xin, *The relaxation scheme for systems of conservation laws in arbitrary space dimensions*, Commmun. Pure Appl. Math **48** (1995), no. 3, 235–276.

[KA88] P.B. Kunasz and L. Auer, *Short characteristic integration of radiative transfer problems: Formal solution in two-dimensional slabs*, J. Quant. Spectrosc. Radiat. Transfer **39** (1988), 67.

[Kan96] G. Kanschat, *Parallel and adaptive Galerkin methods for radiative transfer problems*, Preprint 96-29, Universität Heidelberg, Mai 1996.

[KK] G. Karypis and V. Kumar, *Metis, a software package for partitioning unstructured graphs, unstructured meshes, and computing fill-reducing orderings of sparse matrices, version 3.0.3*, University of Minnesota (1997), http://www-users.cs.umn.edu/~karypis/metis/index.html.

[KN99a] S. Kawashima and S. Nishibata, *Cauchy problem for a model system of the radiating gas: weak solutions with a jump and classical solutions*, Math. Models Methods Appl. Sci. **9** (1999), no. 1, 69–91.

[KN99b] ———, *Shock waves for a model system of the radiating gas*, SIAM J. Math. Anal. **30** (1999), no. 1, 95–117.

[KNN98] S. Kawashima, Y. Nikkuni, and S. Nishibata, *The initial value problem for hyperbolic-elliptic coupled systems and applications to radiation hydrodynamics.*, Freistühler, H., Analysis of systems of conservation laws., Chapman & Hall/CRC, 1998.

[KO00] D. Kröner and M. Ohlberger, *A posteriori error estimates for upwind finite volume schemes for nonlinear conservation laws in multi dimensions.*, Math. Comput. **69** (2000), no. 229, 25–39.

[Krö97] D. Kröner, *Numerical schemes for conservation laws*, first ed., Wiley–Teubner Series Advances in Numerical Mathematics, B.G. Teubner Verlagsgesellschaft mbH, Stuttgart, 1997.

[Krö02] T. Kröger, RWTH Aachen, private communication., July 2002.

[Kru70] S.N. Kruzhkov, *First order quasilinear equations in several independent variables*, Math. USSR, Sb. **10** (1970), 217–243.

[Kur96] R.L. Kurucz, *Status of the ATLAS 12 Opacity Sampling Program and New Programs for Rosseland and Distribution Function Opacity*, M.A.S.S.; Model Atmospheres and Spectrum Synthesis (S.J. Adelman, F. Kupka, and W.W. Weiss, eds.), ASP Conference Series, 1996, p. 160.

[Lar91] B. Larrouturou, *How to preserve the mass fraction positivity when computing compressible multicomponent flows*, J. Comput. Phys. **95** (1991), no. 1, 59–84.

[LeV90] R.J. LeVeque, *Numerical methods for conservation laws*, first ed., Lectures in Mathematics, Birkhäuser, Basel; Boston; Berlin, 1990.

[LeV98] R. LeVeque, *Balancing source terms and flux gradients in high-resolution Godunov methods: The quasi-steady wave-propagation algorithm*, J. Comput. Phys. **146** (1998), no. 1, 346–365.

[LL98] P.D. Lax and X.-D. Liu, *Solution to the tow–dimensional Riemann problems of gas dynamics by positive schemes*, SIAM J. Sci. Comput. **19** (1998), no. 2, 319–340.

[LMMW00] M. Lukácová-Medvid'ová, K.W. Morton, and Gerald Warnecke, *Evolution Galerkin methods for hyperbolic systems in two space dimensions.*, Math. Comput. **69** (2000), no. 232, 1355–1384.

[LR74] P. Lesaint and P.A. Raviart, *On a finite element method for solving the neutron transport equation*, Mathematical Aspects of Finite Elements in Partial Differential Equations (New York) (C. de Boor, ed.), Mathematics Research Center University of Wisconsin-Madison, Academic Press, April 1974, pp. 89–123.

[LZ00] P. Londrillo and L. Del Zanna, *High–order upwind schemes for multidimensional magnetohydrodynamics*, Astrophys. J. **530** (2000), no. 1, 508–524.

[MAM78] D. Mihalas, L. Auer, and B. Mihalas, *Two-dimensional radiative transfer. i. planar geometry*, Astrophys. J. **220** (1978), 1001.

[MBR96] A. Malagoli, G. Bodo, and R. Rosner, *On the nonlinear evolution of magnetohydrodynamic Kelvin–Helmholtz instabilities*, Astrophys. J. **456** (1996), 708–716.

[MM84] D. Mihalas and B.W. Mihalas, *Foundations of radiation hydrodynamics*, Oxford University Press, New York, Oxford, 1984.

[MOS+00] C.-D. Munz, P. Omnes, R. Schneider, E. Sonnendrücker, and U. Voss, *Divergence correction techniques for Maxwell solvers based on a hyperbolic model*, J. Comput. Phys. **161** (2000), no. 2, 484–511, doi:10.1006/jcph.2000.6507.

[MP89] R. Menikoff and B. Plohr, *The Riemann problem for fluid flow of real materials*, Rev. Mod. Physics **61** (1989), no. 1, 75–120.

[MS99] P. Montarnal and C.-W. Shu, *Real gas computation using an energy relaxation method and high-order WENO schemes*, J. Comput. Phys. **148** (1999), no. 1, 59–80.

[MSSV99] C.-D. Munz, R. Schneider, E. Sonnendrücker, and U. Voss, *Maxwell's equations when the charge conservation is not satisfied*, C. R. Acad. Sci. Paris Ser. I **328** (1999), 431–436.

[MV99] S. Müller and A. Voß, *A Riemann solver for the Euler equations with non–convex equation of state*, Bericht 168, IGPM, RWTH Aachen, 1999.

[MYV88] J.-L. Montagné, H.C. Yee, and M Vinokur, *Comparative study of high-resolution shock-capturing schemes for a real gas.*, Proceedings of the 7th GAMM-Conference on numerical methods in fluid mechanics (Michel Deville, ed.), Notes Numer. Fluid Mech., vol. 20, Triedr. Vieweg & Sohn, Braunschweig/Wiesbaden, 1988, pp. 219–228.

[NS90] A. Nordlund and R.F. Stein, *3–D simulations of solar and stellar convection and magnetoconvection*, Comput. Phys. Commun. **59** (1990), no. 1, 119–125.

[NT90] H. Nessyahu and E. Tadmor, *Non–oscillatory central differencing for hyperbolic conservation laws*, J. Comput. Phys. **87** (1990), no. 2, 408–463.

[PD] R. Preis and R. Diekmann, *The PARTY partitioning-library user guide - version 1.1*, Technical Report, University of Paderborn, (1996), http://www.uni-paderborn.de/fachbereich/AG/monien/RESEARCH/PART/party.html.

[Pet91] T. Peterson, *Note on the convergence of the discontinuous Galerkin method for a scalar hyperbolic equation*, SIAM J.Numer.Anal. **28** (1991), 133–140.

[Pet00] P.G. Petropoulos, *Reflectionless sponge layers as absorbing boundary conditions for the numerical solution of Maxwell equations in rectangular, cylindrical, and spherical coordinates*, SIAM J. Appl. Math. **60** (2000), no. 3, 1037–1058.

[Pow94] K.G. Powell, *An approximate Riemann solver for magnetohydrodynamics (that works in more than one dimension)*, ICASE–Report 94–24 (NASA CR–194902), NASA Langley Research Center, Hampton, VA 23681–0001, 8. April 1994.

[PRL+99] K.G. Powell, P.L. Roe, T.J. Linde, T.I. Gombosi, and D.L. De Zeeuw, *A solution–adaptive upwind scheme for ideal magnetohydrodynamics*, J. Comput. Phys. **154** (1999), no. 2, 284–309, Art. No. jcph.1999.6299.

[PRM+95] K.G. Powell, P.L. Roe, R.S. Myong, T. Gombosi, and D. De Zeeuw, *An upwind scheme for magnetohydrodynamics*, Numerical methods for fluid dynamics (Oxford) (K.W. Morton et al., ed.), vol. V, Clarendon Press, 1995, Proceedings of the conference, Oxford, UK, April 1995, pp. 163–180.

[Rem01] M. Rempel, *Struktur und Ursprung starker Magnetfelder am Boden der solaren Konvektionszone*, Ph.D. thesis, Georg–August–Universität, Göttingen, June 2001, http://webdoc.sub.gwdg.de//diss/2001/rempel/index.html.

[Ric88] G. Richter, *An optimal-order error estimate for the discontiuous Galerkin method*, Math.Comput. **50** (1988), 75–88.

[RJ95] D. Ryu and T.W. Jones, *Numerical magnetohydrodynamics in astrophysics: algorithms and tests for one-dimensional flow*, Astrophys. J. **442** (1995), 228–258.

[RMJF98] D. Ryu, F. Miniati, T.W. Jones, and A. Frank, *A divergence–free upwind code for multidimensional magnetohydrodynamic flows*, Astrophys. J. **509** (1998), no. 1, 244–255.

[Roh98] C. Rohde, *Upwind finite volume schemes for weakly coupled hyperbolic systems of conservation laws in 2d*, Numer. Math. **81** (1998), no. 1, 85–124.

[RZ01] C. Rohde and W. Zajaczkowski, *On the Cauchy problem for the equations of ideal compressible MHD fluids with radiation*, Preprint 18, Albert–Ludwigs–Universität, Mathematische Fakultät, Freiburg, 2001, accepted for publication in Appl. Math.

[Sch99a] B. Schupp, *Entwicklung eines effizienten Verfahrens zur Simulation kompressibler Strömungen in 3D auf Parallelrechnern*, Ph.D. thesis, Albert–Ludwigs–Universität, Mathematische Fakultät, Freiburg, Dezember 1999, http://www.freidok.uni–freiburg.de/volltexte/68.

[Sch99b] M. Schüssler, *Zur Bestimmung der thermodynamischen Größen eines partiell ionisierten Gases*, Internal memorandum, Max–Planck–Institut für Aeronomie, Katlenburg–Lindau, November 1999.

[SGDKS98] O. Steiner, U. Grossmann-Doerth, M. Knölker, and M. Schüssler, *Simulation of the interaction of convective flow with magnetic elements in the solar atmosphere*, Astrophys. J. **495** (1998), 468.

[Sof98] I.L. Sofronov, *Non–reflecting inflow and outflow in a wind tunnel for transonic time–accurate simulation*, J. Math. Anal. Appl. **221** (1998), no. 1, 92–115.

[Sül98] E. Süli, *A posteriori error analysis and adaptivity for finite element approximations of hyperbolic problems*, An Introduction to recent developments in theory and numerics for conservation laws (D. Kröner, M. Ohlberger, and C. Rohde, eds.), Springer, Lecture Notes in Computational Science and Engineering, 1998.

[SW95] K. Strehmel and R. Weiner, *Numerik gewöhnlicher Differentialgleichungen*, Teubner Stuttgart, 1995.

[TO96] G. Tóth and D. Odstrčil, *Comparison of some flux corrected transport and total variation diminishing numerical schemes for hydrodynamic and magnetohydrodynamic problems*, J. Comput. Phys. **128** (1996), no. 1, 82–100.

[Tót00] Gábor Tóth, *The $\nabla \cdot B = 0$ constraint in shock–capturing magnetohydrodynamics codes*, J. Comput. Phys. **161** (2000), 605–652, doi:10.1006/jcph.2000.6519.

[Tsy98] S.V. Tsynkov, *Numerical solution of problems on unbounded domains. A review*, Appl. Numer. Math. **27** (1998), no. 4, 465–532.

[Tur93] S. Turek, *An efficient solution technique for the radiative transfer equation*, Impact in Comp. in Sci. and Eng. **5** (1993), 201–214.

[TY98] E. Turkel and A. Yefet, *Absorbing PML boundary layers for wave–like equations*, Appl. Numer. Math. **27** (1998), no. 4, 533–557.

[UFUB] IAM Universität Freiburg and IAM SFB 256 Universität Bonn, *GRAPE — GRAphics Programming Environment*, http://www.mathematik.uni-freiburg.de/IAM/Research/grape/GENERAL/index.html.

[Vil94] J.-P. Vila, *Convergence and error estimate in finite volume schemes for general multidimensional conservation laws.*, Math. Model. Numer. Anal. **28** (1994), 267–285.

[VK65] W. Vincenti and C. Kruger, *Introduction to physical gas dynamics*, Willey, New York, 1965.

[Vol02] P. Vollmöller, *Untersuchung der Wechselwirkung von Magnetfeldkonzentrationen und konvektiven Strömungen mit dem Strahlungsfeld in der Photosphäre der Sonne*, Ph.D. thesis, Max–Planck–Institut für Aeronomie / Universität Göttingen, Göttingen, 2002.

[War99] G. Warnecke, *Analytische Methoden in der Theorie der Erhaltungsgleichungen. (Analytic methods in the theory of conservation laws)*, Teubner-Texte zur Mathematik, Stuttgart, 1999.

[Wes02a] M. Wesenberg, *Efficient finite volume schemes for MHD simulations in solar physics*, Ph.D. thesis, Albert–Ludwigs–Universität, Mathematische Fakultät, Freiburg, 2002.

[Wes02b] ———, *Efficient MHD Riemann solvers for simulations on unstructured triangular grids*, J. Numer. Math. **10** (2002), no. 1, 37–71.

[WK] T. Williams and C. Kelley, *Guplot: An interactive plotting program version 3.7*, organized by: David Denholm (3 December 1998), http://www.ucc.ie/gnuplot/gnuplot.html.

[WMMD01] T. A. Wareing, J. M. McGhee, J. E. Morel, and Pautz S. D., *Discontinuous finite element Sn methods on 3-d unstructured grids*, Sci. Eng. **138 No.3** (2001), 256–268.

[ZC92] A.L. Zachary and P. Colella, *Note: A higher–order Godunov method for the equations of ideal magnetohydrodynamics*, J. Comput. Phys. **99** (1992), 341–347.

[ZMC94] A.L. Zachary, A. Malagoli, and P. Colella, *A higher–order Godunov method for multidimensional ideal magnetohydrodynamics*, SIAM J. Sci. Comput. **15** (1994), no. 2, 263–284.