Ph.D. Thesis in the Doctoral Program
Computer Science and Computational Mathematics

Tesi di Dottorato di Ricerca in
Informatica e Matematica del Calcolo
Ciclo XXXIII

Università degli Studi dell'Insubria
Dipartimento di Scienza e Alta Tecnologia

Numerical methods for all-speed flows for the Euler equations including well-balancing of source terms

Author:
Andrea Christine THOMANN
Marie Skłodowska-Curie fellow from
Istituto Nazionale di Alta Matematica

Supervisor:
Prof. Gabriella PUPPO
La Sapienza Università di Roma

Academic year 2019/2020

This work has been funded under the

INdAM Doctoral Programme in Mathematics and/or Applications Cofunded by Marie Sklodowska-Curie Actions.

Acronym: INdAM-DP-COFUND-2015
Grant number: 713485

Abstract

This thesis regards the numerical simulation of inviscid compressible ideal gases which are described by the Euler equations. We propose a novel implicit explicit (IMEX) relaxation scheme to simulate flows from compressible as well as near incompressible regimes. The Mach number, which gives the ratio between the gas and sound speed, plays an important role in the design of the scheme, as it has great influence on the flow behaviour and physical properties of solutions of the Euler equations. Our focus is on an accurate resolution of the Mach number independent wave associated with the gas velocity, also called material wave. Apart from strongly supersonic regimes, the material wave is rather slow compared to the other two acoustic waves which tend to infinity for small Mach numbers, thus also called the fast waves. A special feature of our scheme is that it can account for the influence of a gravitational field on the fluid flow and is applicable also in small Froude number regimes which are characterized by a strong gravitational potential.

To address these problems, we use a relaxation model which is designed to separate the slow Mach number independent waves from the fast Mach number dependent acoustic waves. The relaxation model is then numerically solved with an IMEX approach in which the slow waves are treated explicitly while the fast waves are integrated implicitly. The approach proposed in this work results in a single scalar PDE for the implicit fast waves, while on the explicit part a Godunov-type scheme based on an approximate Riemann solver is applied. The time step of the IMEX scheme is constrained only by the eigenvalues of the explicitly treated system of equations and is independent of the Mach number allowing for large time steps especially in low Mach number regimes.

Despite using an upwind scheme in the main explicit part, where the physical variables are updated, the scheme has a Mach number independent diffusion. This is a necessary property to obtain accurate solutions also for small Mach numbers. In addition, the scheme is provably asymptotic preserving (AP), which means that the obtained numerical solutions in the near incompressible flow regime have the correct physical limit behaviour. Contributing to the AP property is the use of centred differences in the implicit part which preserve the well-preparedness of the pressure.

In presence of a gravitational source term, the scheme is well-balanced for arbitrary a priori known hydrostatic equilibria independently of the considered Mach and Froude regime. As the scheme provably preserves the positivity of density and internal energy throughout the simulation, the scheme is well suited for physical applications. To increase the accuracy of the presented scheme a natural extension to second order is provided which keeps all properties of the first order scheme. Applications in multiple space dimensions are realized by applying dimensional splitting in the explicit step of the IMEX scheme while the implicit scalar PDE is extended to include multiple space derivatives.

The theoretical properties of the given schemes are numerically validated by various test cases performed on Cartesian grids in multiple space dimensions.

Keywords All-speed schemes, IMEX discretization, relaxation, Euler equations, gravity, finite volumes, asymptotic preserving, well-balanced

Riassunto

Questa tesi riguarda la simulazione numerica dei gas ideali comprimibili non viscosi che sono descritti dalle equazioni di Eulero. Proponiamo un nuovo schema di rilassamento implicito esplicito (IMEX) per simulare flussi in regimi comprimibili e quasi incomprimibili. Il numero di Mach, che dà il rapporto tra la velocità del gas e quella del suono, gioca un ruolo importante nel design dello schema, in quanto ha una grande influenza sul comportamento del flusso e sulle proprietà fisiche delle soluzioni delle equazioni di Eulero. Il nostro obiettivo è una risoluzione accurata dell'onda associata alla velocità del gas, detta anche onda materiale che è indipendente dal numero di Mach. A parte regimi fortemente supersonici, l'onda materiale è piuttosto lenta rispetto alle altre due onde acustiche che tendono all'infinito nel regime a basso numero di Mach, perciò chiamate anche onde veloci. Una caratteristica speciale di questo schema risiede nel fatto che può tener conto dell'influenza di un campo gravitazionale sul flusso del fluido ed è applicabile anche per regimi con numero di Froude piccolo, che sono caratterizzati da un forte potenziale gravitazionale.

Per affrontare questi problemi, utilizziamo un modello di rilassamento che è stato progettato appositamente in questo lavoro per separare le onde lente indipendenti dal numero di Mach dalle onde acustiche veloci dipendenti dal numero di Mach. Il modello di rilassamento viene poi risolto numericamente con un approccio IMEX in cui le onde lente sono trattate esplicitamente mentre le onde veloci sono integrate implicitamente. L'approccio proposto in questo lavoro risulta in un'unica PDE scalare per le onde veloci implicite, mentre alla parte esplicita viene applicato uno schema di tipo Godunov basato su un solutore di Riemann approssimato. Il passo temporale dello schema IMEX è vincolato solo dagli autovalori del sistema trattato esplicitamente ed è indipendente dal numero di Mach, consentendo grandi passi temporali soprattutto nei regimi a basso numero di Mach.

Nonostante l'utilizzo di uno schema *upwind* nella parte principale esplicita, dove le variabili fisiche vengono aggiornate, lo schema presenta una diffusione indipendente dal numero di Mach che è necessaria per ottenere soluzioni accurate anche per bassi valori del numero di Mach. Inoltre, si può dimostrare che lo schema preserva il limite asintotico, il che significa che le soluzioni numeriche ottenute nel regime di flusso quasi incomprimibile presentano un comportamento limite dal punto di vista fisico corretto. A ciò, contribuisce l'uso delle differenze centrate nella parte implicita che conserva la dipendenza correta rispetto al numero di Mach.

In presenza di un termine di sorgente gravitazionale lo schema è ben bilanciato per equilibri idrostatici arbitrari noti a priori e questa proprietà è indipendente dal regime di Mach e Froude considerati. Poiché si verifica che lo schema preserva la positività della densità e dell'energia interna durante tutta la simulazione, esso è adatto alle applicazioni fisiche. Per aumentare l'accuratezza dello schema presentato è prevista una naturale estensione al secondo ordine che mantiene tutte le proprietà dello schema del primo ordine. Si possono realizzare applicazioni a più dimensioni spaziali applicando il solutore di Riemann dimensione per dimensione nella fase esplicita dello schema IMEX mentre la PDE scalare implicita è estesa per includere multipli derivati spaziali.

Le proprietà teoriche degli schemi presentati sono validate numericamente con vari casi test eseguiti su griglie cartesiane in varie dimensioni spaziali.

Parole chiave schemi a tutte le velocità, discretizzazione IMEX, rilassamento, equazioni di Eulero, gravità, volumi finiti, conservazione asintotica, mantenimento dell'equilibrio

Kurzzusammenfassung

Diese Arbeit befasst sich mit der numerischen Simulation von reibungsfreien kompressiblen idealen Gasen, die durch die Euler-Gleichungen beschrieben werden. Wir schlagen ein neuartiges implizites explizites (IMEX) Verfahren vor, welches auf einem Relaxationsansatz basiert, um kompressible und nahezu inkompressible Strömungen zu simulieren. Die Machzahl, welche das Verhältnis zwischen der Gas- und Schallgeschwindigkeit angibt, spielt eine wichtige Rolle in der Entwicklung des Verfahrens, da sie großen Einfluss auf das Strömungsverhalten und die physikalischen Eigenschaften der Lösungen der Euler-Gleichungen hat. Unser Schwerpunkt liegt auf einer genauen Auflösung der von der Machzahl unabhängigen Welle, die durch die Gasgeschwindigkeit beschrieben und als Materialwelle bezeichnet wird. Abgesehen von stark supersonischen Bereichen, ist die Materialwelle ziemlich langsam verglichen mit den beiden anderen akustischen Wellen, die für kleine Machzahlen gegen unendlich streben und deswegen auch als schnelle Wellen bezeichnet werden. Eine Besonderheit unseres Verfahrens ist, dass es den Einfluss eines Gravitationsfeldes auf die Flüssigkeitsströmung berücksichtigen kann und auch auf Strömungen anwendbar ist, die durch kleine Froudezahlen, welche auf ein starkes Gravitationspotential zurückgehen, gekennzeichnet sind.

Um diese Probleme zu lösen, verwenden wir ein Relaxationsmodell, das so konzipiert ist, dass die langsamen machzahlunabhängigen Wellen von den schnellen machzahlabhängigen akustischen Wellen getrennt werden können. Das Relaxationsmodell wird dann numerisch mit einem IMEX-Ansatz gelöst, bei welchem die langsamen Wellen explizit und die schnellen Wellen implizit integriert werden. Der in dieser Arbeit vorgeschlagene Ansatz führt zu einer einzigen skalaren PDE für die impliziten schnellen Wellen, während auf den expliziten Teil ein auf einem approximativen Riemann-Löser basierendes Godunov-Verfahren angewendet wird. Der Zeitschritt des IMEX-Verfahrens wird nur durch die Eigenwerte des explizit diskretisierten Gleichungssystems begrenzt und ist unabhängig von der Machzahl, so dass insbesondere in Anwendungen mit niedriger Machzahl große Zeitschritte möglich sind.

Trotz der Verwendung einer *upwind*-Diskretisierung im expliziten Hauptteil des Verfahrens, in dem die physikalischen Variablen aktualisiert werden, hat das Verfahren eine von der Machzahl unabhängige Diffusion, die notwendig ist, um auch für kleine Machzahlen akkurate Lösungen zu erhalten. Darüber hinaus ist das Verfahren nachweislich asymptotisch erhaltend (AP), was bedeutet, dass numerische Lösungen in nahezu inkompressiblen Strömungsbereichen das korrekte physikalische Grenzverhalten aufweisen. Zu dieser Eigenschaft trägt auch die Verwendung von zentralen Differenzen im impliziten Teil bei, welche die korrekte Abhängigkeit des Durcks hinsichtlich der Machzahl bewahren.

Ist ein Gravitationsquellterm gegeben, erhält das Verfahren beliebige a priori bekannte hydrostatische Gleichgewichte, unabhängig von den betrachteten Mach- und Froudezahlen. Da das Verfahren nachweislich die Positivität von Dichte und interner Energie während der gesamten Simulation bewahrt, eignet sich das Schema gut für aus der Physik stammenden Anwendungen. Um die Genauigkeit des vorgestellten Verfahrens zu erhöhen, wird

eine natürliche Erweiterung auf zweite Ordnung hergeleitet, die alle Eigenschaften des Verfahrens erster Ordnung erhält. Applikationen in mehreren Raumdimensionen können durch die Anwendung des Riemann-Lösers entlang jeder Raumrichtung in der expliziten Phase des IMEX-Verfahrens realisiert werden, während die skalare implizite PDE auf mehrere räumliche Ableitungen erweitert wird.

Die theoretischen Eigenschaften der vorgestellten Verfahren werden durch verschiedene numerische Tests, die auf kartesischen Gittern in mehreren Raumdimensionen durchgeführt werden, validiert.

Schlüsselwörter Verfahren geeignet für alle Strömungsgeschwindikeiten bezüglich der Machzahl, IMEX Diskretisierung, Relaxation, Euler Gleichungen, Schwerkraft, finite Volumen, asymptotisch erhaltend, gleichgewichtserhaltend

Acknowledgements

Many people were involved until the finishing of this thesis and I want to thank them for accompanying me on the way.

First, I would like to deeply thank my supervisor Gabriella Puppo for her support and trust in me, from applying for the PhD scholarship until the end of the PhD time, especially during the last year that was not easy for all of us. She taught, and still teaches, me a lot and guided me not only through academic challenges, but also helped me to adapt to the new environment and encouraged me to experience as many facets of Italian culture and lifestyle as possible.

I also want to say thank you to Matteo Semplice and Sebastiano Boscarino with whom we often discussed about this work and who had always good advice. A thank-you goes to Gero Schnücke for suggestions and useful comments during the writing of this thesis. I would also like to thank my reviewers Angelo Iollo and Manuel Torrilhon for accepting to read and review this thesis, as well as for their valuable comments and insights.

Second, I would like to warmly thank Christian Klingenberg who initially brought me in touch with mathematical fluid dynamics and encouraged me to do the PhD in Italy. Also I am very grateful towards Markus Zenk who won me over with his enthusiasm to study this field and introduced me to the beautiful world of fluid dynamics. Both were a great support during my PhD time and beyond and I thank them for all their time, suggestions and contributions.

Further, I would like to thank the INdAM Doctoral Programme for the financial support, which made it possible for me to fully concentrate on my studies and gave me the opportunity to travel to winter/summer schools and conferences to not only get in touch and connect with fellow researchers, but also to present and discuss my work with them. Especially I want to thank Mauro Petrucci from INdAM, who was always available for questions and helped to smoothly navigate through any bureaucratic problem, as well as Marco Donatelli on the part of the PhD school of Insubria University as coordinator of the PhD program.

I also would like to thank my colleagues and friends in Como who made my time there unforgettable, especially Isa, Mele, Paola, Rafael and Riccardo. A big thank you also to all my friends that are scattered all over the world and despite the distance were always there for me, be it in difficult or happy times, especially Felix, Jayesh, Pablo, Rafael, Thanasis and Victor.

Last but not least I would like to thank my parents who are always there for me and without them this thesis would not have happened.

Contents

Introduction

Context and scope of the study

Atmospheric flows of compressible invicid gases, as arising e.g. in meteorology and astrophysics, are often modelled by the compressible Euler equations. Thereby large scale differences in the flow regime driven by thermodynamical processes are observed. The main quantity through which the flow is characterised is the Mach number. It can be used to measure the compressibility of the fluid flow since it is defined as the ratio of the gas velocity in relation to the present sound speed. Compressible flows are described by a Mach number around one, whereas flows with higher Mach number are called supersonic. Especially low Mach number flow, with Mach numbers reaching close to zero, is challenging since it exhibits properties that are characteristic in incompressible flows. Amongst them are a divergence free velocity field and constant density.

Including a gravitational field in the considered model is important to obtain a proper description of atmospheric events. Thereby hydrostatic equilibrium states can arise which are characterized by a balance of the gravity against a pressure-gradient force. Many physical flow phenomena are in principle perturbations around such a stationary state. Adding gravity introduces another scale into the considered model given by the Froude number which states the ratio of the flow inertia to the external gravitational field. High Froude numbers correspond to a negligible gravitational influence leading back to the homogeneous Euler equations widely studied in mathematical fluid dynamics.

Here, in contrast, we want to take into account both low Mach and low Froude fluid flows and aim at constructing stable and computationally efficient numerical methods applicable in all flow regimes, also called all-speed schemes. These kind of schemes are necessary as one scheme suffices to accurately approximate compressible as well as near incompressible flows. To guarantee the correct incompressible limit behaviour with respect to a vanishing Mach number, schemes have to be asymptotic preserving (AP), which means that in the low Mach number limit a consistent dicretization of the incompressible Euler equations must be achieved. Closely connected to the AP property is the requirement of a Mach number independent bound for the artificial diffusion of the numerical scheme. This bound is essential in the low Mach number regime as it has been observed that standard schemes are extremely diffusive for small Mach numbers leading to unreliable numerical solutions.

A property that becomes crucial when working with a gravitational source terms in this context is the capturing of hydrostatic equilibria on machine precision, also called well-balancing. As mentioned above, one is interested in the simulation of small perturbations around equilibria. To resolve those small perturbations, it is necessary that the underlying

equilibrium is captured exactly, even on coarse grids and thus the well-balanced property is required. Besides, for a numerical scheme to be of interest in physical applications, quantities as density and internal energy have to be positive throughout the simulation. This important property is also referred to as positivity preserving or positivity property. Summarizing, the purpose of this thesis is the construction of an all-speed scheme with the above mentioned properties.

Numerical method

The Euler equations exhibit two acoustic waves whose speeds tend to infinity as the Mach number goes to zero. We are not much interested in the propagation of those fast waves transporting the sound but rather in the slow material wave which carries most of the energy. Therefore the focus of the numerical scheme is on an accurate resolution of this slow waves.

Standard explicit methods are not well suited for this task mainly for two reasons. Due to stability reasons, the time step is oriented to the fast acoustic waves decreasing the time step almost to zero as the Mach number is reduced leading to huge CPU times. In addition all waves are resolved including the acoustic waves which are not our prior interest.

Implicit methods on the other hand do not require a time step restriction and the time step can be chosen to obtain the desired resolution of the slow waves. A big drawback with respect to the implicit treatment of the Euler equations is that without any further simplifications or treatments, one has to solve a non-linear implicit system. Therefore a lot of computational effort has to be put into solving this non-linear system implicitly, often with iterative methods whose convergence is not always guaranteed.

Trying to combine the best of both approaches, we decided for an implicit-explicit (IMEX) discretization of the equations. With a clever splitting of the Euler flux, a Mach number independent time step restriction can be achieved resulting only from the explicitly treated flux terms. Thereby the decreasing time step in the simulation of near incompressible flow can be avoided and the focus of the simulation lies on the resolution of the slow waves.

Part of this clever splitting is to treat the non-linear advective terms explicitly, as the crucial point in choosing the implicitly treated terms is to avoid solving a non-linear implicit system. To achieve this, we make use of relaxation techniques which are applied on the continuous equations obtaining a viscous approximation of the Euler equations and a linearisation of the wave structure. This allows us to use complex Riemann solvers which provide the basis for the positivity and well-balanced property of the all-speed scheme. The implicit part can be reduced to solving only one linear scalar PDE reducing the computational effort compared to a standard fully implicit scheme and moreover avoiding the possible non-convergence of non-linear solvers. Although the discretized linear system is ill conditioned for small Mach numbers yielding slowly convergent linear iterative solvers, we can overcome this problem by using a simple standard preconditioner leading to accurate solutions after few iterations. Due to the structure of the relaxation source term, it suffices to only update the physical variables in the explicit main part of the scheme and the computational effort does not increase due to the presence of relaxation variables

which can be seen as mere auxiliary variables.

The thesis provides a three step guide on how an all-speed scheme with the above explained properties is obtained. Starting from the compressible Euler equations with a gravitational source term, we pass on to an all-speed scheme for the homogeneous Euler equations, arriving finally at a well-balanced all-speed scheme for the Euler equations with gravity.

Organization of the thesis

In detail, the thesis is organized as follows.

The *first* chapter is concerned with the basics of the Euler equations on a continuous level. Beginning with the definition of solutions to general conservation laws, to which the homogeneous Euler equations belong, their physical properties as well as their wave structure are addressed. Fundamental to the work in this thesis are the solution of Riemann problems and relaxation techniques which are described subsequently. The chapter is concluded by the extension of the Euler equations to include gravitational source terms.

The *second* chapter is devoted to the explicit finite volume discretization which provides the basis of the schemes presented in the remainder of the thesis. The focus lies on Godunov-type schemes and the construction of approximate Riemann solvers. To give an idea how approximate Riemann solvers can be obtained using a relaxation model, an example based on a Suliciu relaxation model is given. This is followed by the extension of Godunov type schemes to multiple dimensions in case of the Euler equations followed by an extension to second order accuracy in space and different time integration methods. Chapter 2 is concluded by briefly discussing an extension of Godunov-type methods to source terms and defining discrete stationary solutions and well-balanced schemes in the context of Riemann solvers.

In the *third* chapter, an explicit Godunov-type scheme for the compressible Euler equations with a source term is constructed. Important features are the well-balanced property for arbitrary given hydrostatic equilibria and the positivity property. The central task is thereby to extend the scheme to second order preserving the properties of the first order scheme which are numerically verified. The main focus of the test cases is on the performance of the scheme in the resolution of small perturbations of hydrostatic equilibria and its behaviour in case of strong rarefactions. The content of this chapter is already published in the *Int. Journal for Numerical Methods in Fluids* [78].

The goal of the *forth* chapter is the development of an all-speed scheme for the homogeneous Euler equations. The task therein was the construction of a stable asymptotic preserving IMEX scheme, which exhibits a time step and numerical diffusion that are both independent of the Mach number. It is based on a relaxation model which allows for one linear implicit scalar equation only and the use of a Riemann solver in the explicit part to which the positivity property of the resulting scheme is closely connected. An extension to second order accuracy is provided which possesses the properties of the first order scheme verified also numerically. The applicability of the scheme on low Mach number flows and its performance in compressible as well as near incompressible flows are the main focus of the numerical tests. The content of this chapter is already published in *Communications in Computational Physics* [79].

The all-speed IMEX scheme from Chapter 4 is augmented in the *fifth* chapter to be able to numerically treat gravitational source terms. One task thereby is the derivation of the limit equations and the definition of well-prepared data in the presence of source terms. An essential part of the work is to combine the concepts of well-balancing arbitrary hydrostatic equilibria from Chapter 3 with the IMEX formalism derived in Chapter 4. The resulting all-speed IMEX scheme is asymptotic and positivity preserving as well as well-balanced. The extension to second order is based on the time integration scheme introduced in Chapter 4 with the space reconstruction techniques discussed in Chapter 3 leading to the preservation of all the aforementioned properties. Numerical examples verify the low Mach, low Froude properties of the numerical scheme as well as its ability to preserve hydrostatic equilibria on machine precision. It is concluded with a rising bubble test case, a meteorological phenomenon, which requires at least a second order scheme, as well as well-balancing and a good performance of the scheme in the low Mach, low Froude regime. The content of this chapter is already published in the *Journal of computational Physics* [77].

The thesis is completed by a chapter of *Conclusions* wherein the results of this work are summarized and perspectives for future research are given. Especially interesting are the extension of the all-speed IMEX schemes to higher order accuracy and their application on two phase flows.

Chapter 1

The Euler equations

This chapter is dedicated to an introduction to the Euler equations and a review of its properties that will be important in the remainder of this work. The given review is non-exhaustive and is mainly based on the following standard text-books for hyperbolic problems [52, 37, 80, 11].

We first give a brief repetition of the notion of solutions in the context of hyperbolic conservation laws, of which the homogeneous Euler equations are an example. Then we turn to the thermodynamic background of the Euler equations followed by a brief analysis of the wave structure. This lays the ground for the study of Riemann problems which also play an important role in the numerical methods that will be applied throughout this work. Since in general the solution of Riemann problems for non-linear hyperbolic problems are complex and difficult to obtain, we consider relaxation methods which provide a viscous approximation of the original equations. Riemann problems for the relaxation model are in general easier to solve as they have a linearised wave structure. Subsequently we give the non-dimensional formulation of the Euler equations and address their asymptotic behaviour with respect to a vanishing Mach number which lay the ground for the construction of the all-speed schemes described later on in this work. Th chapter is completed by considering the Euler equations in the presence of a gravitational field modelling atmospheric flows.

1.1 Solutions of conservation laws

Physical models that describe atmospheric flows are often based on conservation laws. Let d denote the number of space dimension and $w : \mathbb{R}^d \times \mathbb{R} \to \Omega \subset \mathbb{R}^k$ be a vector of k conserved quantities mapping to the set of admissible states Ω. Regarding the homogeneous Euler equations, w consists of the density, momentum and energy. The time variation of w on any volume $V \subset \mathbb{R}^d$ is equal to the losses through the boundary ∂V described by a flux function $f : \Omega \to \mathbb{R}^k$ with $f = (f_1, \dots, f_d)^T$. Formally this process can be written as

$$\partial_t \int_V w(\mathbf{x}, t) d\mathbf{x} + \int_{\partial V} f(w(\mathbf{x}, t)) \cdot \mathbf{n} \, dV = 0, \tag{1.1}$$

where $\mathbf{n}$ denotes the outward pointing normal to the boundary ∂V, $\mathbf{x} = (x_1, \dots, x_d)$ the spatial coordinate and ∂_t the partial derivative with respect to the time t. System (1.1) is

referred to as the *integral formulation* of conservation laws. Assuming w and $f(w)$ to be at least continuously differentiable, we can apply the Gauss divergence theorem on (1.1) to obtain

$$\int_V \partial_t w(\mathbf{x},t) + \nabla \cdot f(w(\mathbf{x},t))\, d\mathbf{x} = 0. \tag{1.2}$$

Since relation (1.2) has to be true for all control volumes V, we can write the *differential form* as

$$\partial_t w(\mathbf{x},t) + \sum_{i=1}^{d} \partial_{x_i} f_i(w(\mathbf{x},t)) = 0 \tag{1.3}$$

with $(\mathbf{x},t) \in \mathbb{R}^d \times \mathbb{R}^+$. We will consider (1.3) equipped with an initial condition

$$w^0(\mathbf{x}) = w(\mathbf{x},0), \quad \mathbf{x} \in \mathbb{R}^d, \tag{1.4}$$

which is referred to as *Cauchy* or *initial value problem* (IVP). Throughout the manuscript we will consider different aspects of the Euler equations which are an example of a *hyperbolic* system of PDEs.

Definition 1.1. *A conservation law (1.3) is called hyperbolic if the flux Jacobians $\nabla_w f_i(w)$ are diagonalizable for any $w \in \Omega$ and $i = 1,\dots,d$. This means it has k real eigenvalues $\lambda_\kappa(w)$ and k linearly independent corresponding eigenvectors $\mathbf{r}_\kappa(w)$, $\kappa = 1,\dots,k$. If all eigenvalues λ_κ are distinct, system (1.3) is called strictly hyperbolic.*

A solution of the differential form (1.3) has to be continuously differentiable. Such a solution is called *classical* or *strong solution*. In practice classical solutions may exists only for a short time interval until discontinuities are developed. This may occur even for smooth initial data. Since discontinuous functions are not solutions of (1.3) in the classical sense, we introduce the concept of weak solutions in the distributional sense.

Definition 1.2. *A function $w \in (L^\infty_{loc}(\mathbb{R}^d \times \mathbb{R}^+))^k$ is called a weak solution of the Cauchy problem (1.3) with an initial condition $w^0 \in (L^\infty_{loc}(\mathbb{R}^d))^k$, if it satisfies*

$$\int_0^\infty \int_{\mathbb{R}^d} \left(\langle w, \partial_t \phi \rangle + \sum_{i=1}^{d} \langle f_i(w), \partial_{x_i} \phi \rangle \right) d\mathbf{x}\, dt + \int_{\mathbb{R}^d} \left\langle w^0(\mathbf{x}), \phi(\mathbf{x},0) \right\rangle d\mathbf{x} = 0 \tag{1.5}$$

for all test functions $\phi \in (C^\infty_0(\mathbb{R}^d \times \mathbb{R}^+))^k$, where $\langle \cdot,\cdot \rangle$ denotes the euclidean product in $\mathbb{R}^k$.

The weak formulation (1.5) is obtained by taking the scalar product of (1.3) and the test function ϕ which is then integrated on $\mathbb{R}^+ \times \mathbb{R}^d$. Thereby all derivatives are passed to the test function ϕ. By construction, every classical solution is a weak solution. Unfortunately not every weak solution is admissible. The following condition helps to identify which discontinuities are admissible. Let Γ be a surface in the $(\mathbf{x},t)$ space across which w is discontinuous and outside of which w is continuously differentiable. Let $\tilde{\mathbf{n}} = (n_t, n_1, \dots, n_d)$ be the normal to Γ and w_+, w_- the limits of w on each side of Γ given by

$$w_\pm(\mathbf{x},t) = \lim_{\varepsilon \to 0} w((\mathbf{x},t) \pm \varepsilon \mathbf{n}).$$

Then the following jump condition, called the *Rankine-Hugoniot* condition,

$$(w_+ - w_-)n_t + \sum_{i=1}^{d}(f_i(w_+) - f_i(w_i))n_i = 0 \tag{1.6}$$

must hold along the surface of discontinuity Γ. If $\tilde{\mathbf{n}} \neq \mathbf{0}$, we can set $\tilde{\mathbf{n}} = (-S, \mathbf{e})$, where $S \in \mathbb{R}$ and $\mathbf{e}$ is a unit vector of $\mathbb{R}^d$. Then (1.6) can be written as

$$S(w_+ - w_-) = \sum_{i=1}^{d}(f_i(w_+) - f_i(w_i))e_i. \tag{1.7}$$

Thereby $\mathbf{e}$ and S may be interpreted as the direction and the speed of the discontinuity described by Γ. Summarizing, to obtain admissible discontinuous solutions, one may reformulate the problem in terms of the differential form (1.3) for regions of smooth solutions while the Rankine-Hugoniot condition must hold across discontinuities.

To illustrate the non-uniqueness of weak solutions, we consider briefly the Burgers' equation which is a non-linear scalar conservation law. The following example can be found in [52].

Example 1.3. *The one dimensional Burgers' equation is given by*

$$\partial_t u + \partial_x \left(\frac{u^2}{2}\right) = 0. \tag{1.8}$$

We consider a discontinuous initial datum consisting of two constant states u_L, u_R given by

$$u^0(x) = \begin{cases} u_L, & \text{if } x \leq 0, \\ u_R, & \text{if } x > 0. \end{cases} \tag{1.9}$$

Following the Rankine-Hugoniot conditions, the speed of propagation of the discontinuity is given with $u_- = u_L, u_+ = u_R$ by $S = (u_L + u_R)/2$. Then the function

$$u^1(x,t) = \begin{cases} u_L, & \text{if } x \leq S\ t, \\ u_R, & \text{if } x > S\ t, \end{cases} \tag{1.10}$$

is a weak solution of the Cauchy problem (1.8),(1.9) for $u_L < u_R$ and $u_L > u_R$. The family of functions

$$u^m(x,t) = \begin{cases} u_L & \text{if } x \leq S_m t, \\ u_m & \text{if } S_m\ t < x \leq u_m\ t, \\ x/t & \text{if } u_m\ t \leq x \leq u_R\ t, \\ u_R & \text{if } x > u_R\ t, \end{cases} \tag{1.11}$$

are weak solutions of the same Cauchy problem for $u_L < u_R$ for any $u_m \in [u_L, u_R]$ and $S_m = (u_L + u_m)/2$. This shows that for $u_L < u_R$ infinitely many weak solutions can be constructed.

This example shows that using only the Rankine-Hugoniot conditions does not lead to physical meaningful or unique solutions. To select the physically relevant solution, we can use in addition *entropy conditions*. Assuming a convex domain Ω, then a strictly convex function $\eta : \Omega \to \mathbb{R}$ is called an *entropy function* for system (1.3) if there exist d functions $\Psi_i : \Omega \to \mathbb{R}$, called *entropy fluxes*, such that

$$\langle \nabla_w \eta(w), \nabla_w f_i(w) \rangle = \nabla_w \Psi_i(w), \quad i = 1, \dots, d. \tag{1.12}$$

For a classical solution w of (1.3) an additional conservation law can be obtained by taking the scalar product of (1.3) with $\nabla_w \eta$ and which is given by

$$\partial_t \eta(w) + \sum_{i=1}^{d} \partial_{x_i} \Psi_i(w) = 0. \tag{1.13}$$

From equation (1.13) immediately follows that the entropy is conserved in smooth regions whereas this is not in general true for a weak solution and in particular not true for discontinuous solutions. To determine the behaviour of the entropy across discontinuities, the evolution of the entropy function for the related viscous problem in the vanishing viscosity limit is studied and a weak formulation of the entropy condition can be obtained.

Definition 1.4. *A weak solution w of (1.3) is called entropy solution, if it satisfies*

$$\int_0^\infty \int_{\mathbb{R}^d} \left(\eta(w)\ \partial_t \phi + \sum_{i=1}^{d} \Psi_i(w)\ \partial_{x_i} \phi \right) d\mathbf{x}\ dt + \int_{\mathbb{R}^d} \eta(w^0(\mathbf{x}))\ \phi(\mathbf{x}, 0) d\mathbf{x} \geq 0 \tag{1.14}$$

for all non-negative test functions $\phi \in C_0^\infty(\mathbb{R}^d \times \mathbb{R}^+), \phi(\mathbf{x}, t) \geq 0$ and for every convex entropy function.

Summarizing, we can write the entropy inequality in the distributional sense as

$$\partial_t \eta(w) + \sum_{i=1}^{d} \partial_{x_i} \Psi_i(w) \leq 0 \tag{1.15}$$

and we obtain the formal criterion that the entropy has to decrease across discontinuities.

This shifts the problem to finding entropy functions and if possible all entropy functions associated with a non-linear system of conservation laws. However in practical examples derived from physical principles it is possible to find an entropy function which has a physical meaning as we will see in case of the Euler equations in the following section.

1.2 Properties of the Euler equations

The homogeneous Euler equations are derived from the physical principles of conservation of mass and energy together with Newton's second law of motion which states that a change in momentum is directly proportional to the applied force. Since the Euler equations are used to describe inviscid gas flows, the change in momentum is only driven by the

applied pressure, as viscous effects and external forces are absent. Putting these physical considerations together gives the system of the Euler equations

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{u}) &= 0, \\ \partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u} + p \mathbb{I}) &= 0, \\ \partial_t E + \nabla \cdot (\mathbf{u}(E+p)) &= 0, \end{aligned} \tag{1.16}$$

where $\rho(\mathbf{x}, t) > 0$ denotes the density, $\rho\mathbf{u}(\mathbf{x}, t)$ the momentum field, $\mathbb{I} \in \mathbb{R}^{d \times d}$ the identity matrix and $E(\mathbf{x}, t) > 0$ the total energy. The latter is given by the sum of kinetic and internal energy as

$$E = \rho \left(\frac{1}{2} |\mathbf{u}|^2 + e \right). \tag{1.17}$$

In addition to the conserved variables $w = (\rho, \rho\mathbf{u}, E)$ arise the velocity field $\mathbf{u}(\mathbf{x}, t)$ and the specific internal energy $e(\mathbf{x}, t) > 0$. The set of physical admissible states contain the states w with positive density and internal energy denoted by

$$\Omega_{phy} = \{ w \in \Omega; \rho > 0, e > 0 \}. \tag{1.18}$$

The system is closed by a pressure law $p(\mathbf{x}, t) > 0$ which is given by an equation of state (EOS). It connects the pressure arising in the momentum and energy equations to the internal energy and the specific volume $\tau = 1/\rho$ via the first law of thermodynamics. Here we assume an ideal gas law for which the pressure can computed by

$$p(\tau, e) = (\gamma - 1) \frac{e}{\tau} = (\gamma - 1) \rho e. \tag{1.19}$$

The adiabatic exponent γ can be estimated by $\gamma = 5/3$ for a mono-atomic gas and by $\gamma = 1.4$ for a two-atomic gas. To complete the review of the underlying physical properties of the Euler equations, we shortly turn to the notion of entropy.

1.2.1 Entropy

The physical entropy can be interpreted as the amount of disorder present in the system. The second law of thermodynamics states that the total entropy in an isolated system is non-decreasing and is constant for reversible processes. The thermodynamical specific entropy $s(\tau, e) > 0$ is defined by the following relations

$$\frac{\partial s}{\partial \tau}(\tau, e) = \frac{p(\tau, e)}{T(\tau, e)} > 0, \quad \frac{\partial s}{\partial e}(\tau, e) = \frac{1}{T(\tau, e)} > 0, \tag{1.20}$$

where $T(\tau, e)$ is the temperature. For an ideal gas, we have

$$s(\tau, e) = s_0 + c_v \ln \left(p(\tau, e) \tau^\gamma \right), \tag{1.21}$$

where c_v is the specific heat at constant volume and s_0 constant. A strictly convex (mathematical) entropy function η is given by $\eta(\tau, e) = -s(\tau, e)/\tau$ associated with the entropy fluxes $\Psi_i = -\rho u_i s$, $i = 1, \ldots, d$. We can write the entropy inequality for the Euler equations (1.16) in the sense of distributions as

$$\partial_t \eta + \nabla \cdot (\eta \mathbf{u}) \leq 0. \tag{1.22}$$

We would like to remark here, that the thermodynamical entropy increases with time, according to the second law of thermadynamics, whereas the (mathematical) convex entropy is a decreasing function.

1.2.2 Eigenstructure and Hyperbolicity

This section is devoted to the hyperbolicity of the Euler equations and an analysis of its eigenstructure. The algebraic properties that are established in this section prepare the ground for the study of Riemann problems in the subsequent section.

For simplicity, we consider the one dimensional (1D) case where the vectors $\mathbf{u} = u$ and $\mathbf{x} = x$ consist only of one component. The Euler equations in 1D are given by

$$\begin{aligned} \partial_t \rho + \partial_x(\rho u) &= 0, \\ \partial_t(\rho u) + \partial_x(\rho u^2 + p) &= 0, \\ \partial_t E + \partial_x(u(E+p)) &= 0. \end{aligned} \tag{1.23}$$

We can write equations 1.23 in quasi-linear form

$$\partial_t w + \mathcal{A}(w)\partial_x w = 0, \tag{1.24}$$

where $\mathcal{A}(w) = \partial_w f(w)$ denotes the flux Jacobian with

$$f(w) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E+p) \end{pmatrix}. \tag{1.25}$$

The quasi-linear formulation is invariant under a smooth variable transformation $v = \varphi(w)$, as well as all concepts derived from it in the remainder of this section. Therefore, we can use primitive variables $v = (\rho, u, e)$ instead of the conserved variables $w = (\rho, \rho u, E)$ to determine the hyperbolicity of the Euler equations (1.23). Using the primitive variables v we can write with $\mathcal{B}(v) = \nabla_w \varphi(w) \mathcal{A}(w) \nabla_w \varphi(w)^{-1}$ the following non-conservative system

$$\partial_t v + \mathcal{B}(v)\partial_x v = 0 \tag{1.26}$$

with

$$\mathcal{B}(v) = \begin{pmatrix} u & \rho & 0 \\ \frac{\partial_\rho p}{\rho} & u & \frac{\partial_e p}{\rho} \\ 0 & \frac{p}{\rho} & u_1 \end{pmatrix}. \tag{1.27}$$

The eigenvalues of $\mathcal{B}(v)$ are given by

$$\lambda_1 = u - c, \quad \lambda_2 = u, \quad \lambda_3 = u + c, \tag{1.28}$$

where c denotes the speed of sound which is defined as

$$c = \sqrt{\tau(p\ \partial_e p - \partial_\tau p)} = \sqrt{\gamma \tau p} = \sqrt{\gamma\ \frac{p}{\rho}}\ > 0$$

for an ideal gas (1.19). For $w \in \Omega_{phy}$, that is positive density and internal energy, the sound speed c is positive. Therefore all eigenvalues are real and can be given in the order $\lambda_1 < \lambda_2 < \lambda_3$. The eigenvalues represent the wave speeds with which information is propagated. Since $\lambda_{1,3}$ contain the sound speed, they are also referred to as *acoustic waves*. The associated eigenvectors are given by

$$\mathbf{r}_1 = \begin{pmatrix} 1 \\ -c/\rho \\ p/\rho^2 \end{pmatrix}, \quad \mathbf{r}_2 = \begin{pmatrix} 1 \\ 0 \\ -e/\rho \end{pmatrix}, \quad \mathbf{r}_3 = \begin{pmatrix} 1 \\ c/\rho \\ p/\rho^2 \end{pmatrix} \tag{1.29}$$

and are linearly independent which establishes that the one dimensional Euler equations (1.23) are strictly hyperbolic.

Connected to each eigenvalue λ_i is a *characteristic field*, called λ_i-field, whose properties are determined by the associated eigenvector(s). A λ_i-field is called *linear degenerate* if for all $w \in \Omega$ and all corresponding eigenvectors holds

$$\langle \nabla_w \lambda_i, \mathbf{r}_i \rangle = 0 \tag{1.30}$$

and *genuinely non-linear* if $\langle \nabla_w \lambda_i, \mathbf{r}_i \rangle \neq 0$ for all $w \in \Omega$. Straightforward calculations show that the characteristic fields associated to the acoustic waves $\lambda_{1,3} = u \pm c$ are genuinely non-linear and the field associated to $\lambda_2 = u$ is linear degenerate.

Another useful concept connected to the eigenstructure are *Riemann invariants*. With their help can be specified which quantities remain invariant across a wave λ_i. For a λ_i-field, an associated Riemann invariant $I_i(w)$ fulfils the following PDE

$$\langle \nabla_w I_i(w), \mathbf{r}_i \rangle = 0. \tag{1.31}$$

Using the eigenvectors (1.29) associated to the eigenvalues λ_i of matrix $\mathcal{B}(v)$, we find the following Riemann invariants

$$\lambda_1 = u - c: \quad I_1^1 = u + \frac{2c}{\gamma - 1}, \quad I_1^2 = e\rho^{-(\gamma-1)}, \tag{1.32}$$

$$\lambda_2 = u: \quad I_2^1 = u, \quad I_2^2 = p, \tag{1.33}$$

$$\lambda_3 = u + c: \quad I_3^2 = u - \frac{2c}{\gamma - 1}, \quad I_3^2 = e\rho^{-(\gamma-1)}. \tag{1.34}$$

Riemann invariants play an important role in finding a solution to a Riemann problem as detailed in the next section.

1.2.3 Riemann Problem

A Riemann problem is a special type of Cauchy problem, where the initial data consists of two constant states w_L and w_R. Riemann problems arise in a physical context in the so called shock-tube problem. Here, we consider one gas consisting of two different initial configurations defined by w_L and w_R which are separated by a membrane. Removing the membrane triggers the interaction of the two gas configurations. The study of Riemann problems is also important regarding the construction of numerical schemes as they naturally arise from a finite volume discretization which is addressed in Chapter 2. For the 1D Euler equations (1.23) the Riemann problem is given by

$$\begin{cases} \partial_t w + \partial_x f(w) = 0, \\ w^0(x) = \begin{cases} w_L & \text{if } x < 0, \\ w_R & \text{if } x > 0, \end{cases} \end{cases} \tag{1.35}$$

where the initial states are $w_L = (\rho_L, (\rho u)_L, E_L)$ and $w_R = (\rho_R, (\rho u)_R, E_R)$. The solution to the Riemann problem (1.35) is a similarity solution, that means it is a function of x/t,

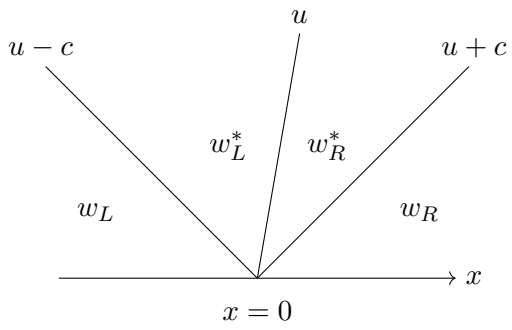


Figure 1.1: Solution structure (1.36) of the Riemann Problem (1.35) of the 1D Euler equations (1.23).

alone, where $t > 0$, which means it is self similar. The solution consists of four constant states separated by the three waves $\lambda_1 < \lambda_2 < \lambda_4$ and can be written as

$$w(x,t) = \tilde{w}(x/t) = \begin{cases} w_L & \text{if } \frac{x}{t} < \lambda_1, \\ w_L^* & \text{if } \lambda_1 < \frac{x}{t} < \lambda_2, \\ w_R^* & \text{if } \lambda_2 < \frac{x}{t} < \lambda_3, \\ w_R & \text{if } \lambda_3 < \frac{x}{t}. \end{cases} \tag{1.36}$$

The states w_L^*, w_R^* are called *intermediate states*. The structure of solution (1.36) is depicted in Figure 1.1.

In principle, we distinguish three different types of waves also called elementary waves. In a *shock wave*, the left and right states w_L, w_R are connected through a single jump discontinuity in a genuinely non-linear λ_i-field. The discontinuity moves with a speed S_i which is given by the Rankine Hugoniot condition (1.6). Entropy violating shocks are ruled out via the Lax entropy condition, that is

$$\lambda_i(w_L) > S_i > \lambda_i(w_R). \tag{1.37}$$

The Lax entropy condition can be deduced for a convex flux from the entropy inequality (1.15). For details, we refer to [25]. This means the family of curves $\frac{dx(t)}{dt} = \lambda_i$ for a curve $x(t)$, also called *characteristic lines* or *characteristics*, run into the discontinuity.

In a *contact wave*, the left and right states w_L, w_R are connected through a discontinuity in a linearly degenerate λ_i-field. As in the case of a shock wave, the Rankine-Hugoniot condition determines the shock speed S_i, but due to (1.30) the characteristic lines run parallel to the discontinuity and we have $\lambda_i(w_L) = S_i = \lambda_i(w_R)$.

In a *rarefaction wave*, the left and right states w_L, w_R are connected by a smooth transition in a genuinely non-linear λ_i-field. In this case, the characteristic lines diverge from each other such that $\lambda_i(w_L) < \lambda_i(w_R)$

We have already seen that the 1D Euler equations (1.23) exhibit three waves of which the middle wave $\lambda_2 = u$ is linearly degenerate and the acoustic waves are genuinely non-linear. Accordingly, the middle wave, connecting the states w_L^* and w_R^*, is a contact wave with the Riemann invariants u and p. This means velocity and pressure are invariant across this wave, whereas there is a jump in the density. This also applies for all density dependent quantities as specific internal energy and sound speed.

The acoustic waves λ_1 and λ_3 can correspond to a rarefaction or a shock wave. They connect the states w_L, w_L^* and w_R^*, w_R respectively. According to the Riemann invariants $I_{1,3}$ in (1.32) and (1.34) all three quantities ρ, u and p change across the acoustic waves. The exact solution of the Riemann problem is difficult to obtain due to the non-linear nature of the Euler equations and we refer to [80, 52] for a detailed study of the solution strategy. The obtained exact solution is complex and cannot be given in a closed form.

Therefore we turn to an approximation of the Euler equations using relaxation techniques. The advantage of considering relaxation models is that they have a linearised wave structure leading to less complex Riemann solutions.

1.3 Relaxation approaches for the Euler equations

A relaxation model for a system of conservation laws (1.3) is a system of PDEs of the form

$$\partial_t W + \nabla \cdot F(W) = \frac{1}{\varepsilon} R(W), \tag{1.38}$$

with a state vector $W : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^q$ and a flux function $F : \mathbb{R}^q \to \mathbb{R}^q$ with $q > k$. The relaxation model (1.38) is a perturbation of the original system (1.3). Therefore, we assume it to be hyperbolic. The relaxation source term $R : \mathbb{R}^q \to \mathbb{R}^q$ drives the solution W to its relaxation equilibrium $\mathcal{M}(w) \in \mathbb{R}^q$ as the relaxation time ε tends to zero. In relaxation equilibrium it has to hold $W = \mathcal{M}(w)$ and $R(\mathcal{M}(w)) = 0$, where $\mathcal{M}$ denotes the equilibrium manifold. Formally we can connect the relaxation system to the original system (1.3) via a linear operator $Q : \mathbb{R}^q \to \mathbb{R}^k$ for which in equilibrium holds $Q\mathcal{M}(w) = w$ and $QF(\mathcal{M}(w)) = f(w)$.

The relaxation model (1.38) consists of more equations than the original system of conservation laws, but is in general easier to solve due to a linearised wave structure. The thereby obtained solution W of the relaxation model provides an approximate solution $w = QW$ of the original system of conservation laws (1.3).

In the following we will review two relaxation strategies. The first one is mainly connected to Jin & Xin [43] and we will use that approach to derive a relaxation system fitted for the Euler equations (1.23) as described by Suliciu in [74, 75].

1.3.1 Jin-Xin Relaxation

To provide a basis for the Suliciu relaxation, which will be the relaxation method used in the remainder of this work, it suffices to consider a scalar conservation law in one dimension

$$\partial_t u + \partial_x f(u) = 0 \tag{1.39}$$

with $u(x, 0) = u^0(x)$ and the eigenvalue is given by $\lambda = \partial_u f(u)$. The associated relaxation system is given by

$$\begin{aligned} \partial_t u + \partial_x v &= 0, \\ \partial_t v + a^2 \partial_x u &= \frac{1}{\varepsilon}\big(f(u) - v\big), \end{aligned} \tag{1.40}$$

with the initial condition for the relaxation variable v given by $v(x, 0) = f\left(u^0(x)\right)$. The relaxation time is again denoted by $\varepsilon > 0$, where the constant $a > 0$ is a relaxation

parameter. The eigenvalues of the relaxation system (1.40) are given by $\lambda^{\pm} = \pm a$ and the flux function is diagonalisable in the characteristic variables $\upsilon \pm au$. The relaxation system is therefore hyperbolic. We will show in the following that the original system can be obtained from the vanishing viscosity limit $\varepsilon \to 0$ of the relaxation system. Therefore we perform a Chapman Enskog analysis and look at the first order approximation in ε of the relaxation variable $\upsilon = \upsilon_0 + \varepsilon\upsilon_1 + \mathcal{O}(\varepsilon^2)$ where $\upsilon_0 = f(u)$. We can rewrite the second equation in (1.40) for υ and obtain

$$\begin{aligned} \upsilon &= f(u) - \varepsilon(\partial_t \upsilon + a^2 \partial_x u) \\ &= f(u) - \varepsilon \left(a^2 - (\partial_u f(u))^2\right) \partial_x u + \mathcal{O}(\varepsilon^2) \end{aligned} \tag{1.41}$$

Inserting the expansion (1.41) into the first equation of (1.40) yields

$$\partial_t u + \partial_x f(u) = \varepsilon \partial_x \left(\left(a^2 - (\partial_u f(u))^2\right) \partial_x u \right). \tag{1.42}$$

Provided that

$$a^2 \geq (\partial_u f(u))^2, \tag{1.43}$$

equation (1.42) is a parabolic equation with a positive dissipation. For $\varepsilon \to 0$ we obtain from equations (1.41) and (1.42) in the limit the original equations

$$\upsilon = f(u), \quad \partial_t u + \partial_x f(u) = 0. \tag{1.44}$$

Therefore we can set as equilibrium $\mathcal{M}(u) = (u, f(u))^T$.

Condition (1.43) guarantees a stable diffusive approximation of the original equations and also provides an interlacing of the eigenvalues

$$-a \leq \partial_u f(u) \leq a \tag{1.45}$$

which was first referred to as the *subcharacteristic condition* for the relaxation parameter a by Liu in [57]. A dissipative entropy condition was formulated for general non-linear relaxation systems by Chen, Levermore and Liu [19] and therein, as well as a little later in [61], was also established that (1.41) and (1.42) indeed govern the asymptotic behaviour of the relaxation system if time goes to infinity or ε goes to zero.

1.3.2 Suliciu Relaxation

To directly apply the relaxation (1.40) on (1.23), we write the one dimensional Euler equations in Lagrangian coordinates. Unlike the Eulerian framework, in which the Euler equations (1.23) are given, the Lagrangian framework tracks a particular fluid particle. The position of the particle ξ at time t is given by $\bar{x}(\xi, t)$ for which holds

$$\frac{\partial \bar{x}}{\partial \xi} = \frac{1}{\rho} = \tau. \tag{1.46}$$

According to the first Newtonian law the velocity of the particle is given by

$$\frac{\partial \bar{x}}{\partial t}(\xi, t) = u(\bar{x}(\xi, t), t). \tag{1.47}$$

Let $\bar{\varphi}(\xi,t) = \varphi(\bar{x}(\xi,t),t) \in L^\infty(\mathbb{R},\mathbb{R}^+)$ denote an Eulerian variable φ in Lagrangian coordinates. Then we can write the transformation

$$\begin{aligned}\frac{\partial\bar{\varphi}}{\partial\xi}(\xi,t) &= \tau\frac{\partial\varphi}{\partial\bar{x}}(\xi,t),\\ \frac{\partial\bar{\varphi}}{\partial t}(\xi,t) &= \rho u\frac{\partial\bar{\varphi}}{\partial\xi}(\xi,t) + \frac{\partial\varphi}{\partial t}(\xi,t),\end{aligned} \tag{1.48}$$

with which we obtain

$$\partial_x\varphi = \rho\ \partial_\xi\bar{\varphi}, \quad \partial_t\varphi = \partial_t\bar{\varphi} - \rho u\ \partial_\xi\bar{\varphi}. \tag{1.49}$$

Using this transformation on the Euler equations (1.23), we can write in Lagrangian coordinates

$$\begin{aligned}\partial_t\tau - \partial_\xi u &= 0,\\ \partial_t u + \partial_\xi p &= 0,\\ \partial_t\left(e + \frac{u^2}{2}\right) + \partial_\xi(pu) &= 0.\end{aligned} \tag{1.50}$$

For the sake of simplicity, we have dropped the bar in the notation of the Lagrangian variables. The non-linear character of (1.50) is given by the non-linear pressure law p. Therefore the pressure is relaxed by applying the Jin Xin relaxation (1.40) on the second equation of (1.50) and we obtain

$$\begin{aligned}\partial_t u + \partial_\xi\pi &= 0,\\ \partial_t\pi + a^2\partial_\xi u &= \frac{1}{\varepsilon}(p-\pi).\end{aligned} \tag{1.51}$$

Transforming back in Eulerian variables, we can write a relaxation system for (1.23) in conserved variables as

$$\begin{aligned}\partial_t\rho + \partial_x(\rho u) &= 0,\\ \partial_t\rho u + \partial_x(\rho u^2+\pi) &= 0,\\ \partial_t E + \partial_x(u(E+\pi)) &= 0,\\ \partial_t\rho\pi + \partial_x(\rho u\pi + a^2 u) &= \frac{\rho}{\varepsilon}(p-\pi).\end{aligned} \tag{1.52}$$

Comparing the relaxation equation for $\rho\pi$ with the actual pressure evolution

$$\partial_t\rho p + \partial_x(\rho u p) + \rho^2 c^2\partial_x u = 0, \tag{1.53}$$

which is derived from the internal energy equation, we see that the relaxation parameter is an approximation of ρc. Indeed we can show that under the subcharacteristic condition

$$a \geq \rho c \tag{1.54}$$

the relaxation system (1.52) is a stable diffusive approximation of the original equations (1.23). Following the Chapman Enskog analyis of the Jin-Xin relaxation, we find using (1.53)

$$\pi = p - \varepsilon\rho\left(\frac{a^2}{\rho^2} - c^2\right)\partial_x u + \mathcal{O}(\varepsilon^2). \tag{1.55}$$

Inserting this in the momentum and energy equation of (1.52), we find

$$\partial_t \rho u + \partial_x(\rho u^2 + p) = \varepsilon\ \partial_x \left(\rho \left(\frac{a^2}{\rho^2} - c^2 \right) \partial_x u \right), \tag{1.56}$$

$$\partial_t E + \partial_x(u(E+p))) = \varepsilon\ \partial_x \left(\rho \left(\frac{a^2}{\rho^2} - c^2 \right) \partial_x \left(\frac{u^2}{2} \right) \right). \tag{1.57}$$

Due to the subcharacteristic condition, the diffusion coefficient is positive and in the limit $\varepsilon \to 0$ the original equations are recovered. The relaxation equilibrium is given by

$$\mathcal{M}(w) = (\rho, \rho u, E, \rho p). \tag{1.58}$$

To obtain further information about the relaxation model, we perform an analysis of the eigenstructure. For the relaxation model, we define primitive variables $\mathcal{V} = (\rho, u, e, \pi)$ and rewrite the relaxation system (1.52) in non-conservative form

$$\partial_t \mathcal{V} + \mathcal{B}(\mathcal{V}) \partial_x \mathcal{V} = 0. \tag{1.59}$$

The matrix $\mathcal{B}(\mathcal{V})$ is given by

$$\mathcal{B}(\mathcal{V}) = \begin{pmatrix} u & \rho & 0 & 0 \\ 0 & u & 0 & \frac{1}{\rho} \\ 0 & \frac{\pi}{\rho} & u & 0 \\ 0 & \frac{a^2}{\rho} & 0 & u \end{pmatrix}. \tag{1.60}$$

The eigenvalues are given by

$$\lambda^- = u - \frac{a}{\rho} < \lambda^u = u < \lambda^+ = u + \frac{a}{\rho}, \tag{1.61}$$

where λ^u has multiplicity two. Due to the subcharacteristic condition, we have the interlacing of the eigenvalues given by

$$u - \frac{a}{\rho} < u - c < u < u + c < u + \frac{a}{\rho}. \tag{1.62}$$

which is depicted in Figure 1.2. The associated eigenvectors are linearly independent and are given by

$$\mathbf{r}_1^u = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{r}_2^u = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{r}^\pm = \begin{pmatrix} \frac{\rho^2}{a^2} \\ \pm\frac{1}{a} \\ \frac{\pi}{a^2} \\ 1 \end{pmatrix}. \tag{1.63}$$

Therefore, the relaxation model is hyperbolic. Furthermore we can compute that all characteristic fields are linearly degenerate which makes the computation of the exact solution of the associated Riemann problem to the relaxation model rather uncomplicated. The exact solution of the Riemann problem is given in Section 2.1.2 where it is used to construct an approximate Riemann solver for the Euler equations (1.23).

Until now we have considered the Euler equations in dimensional variables which are used to construct numerical methods for compressible flow. The basis for the development of numerical methods for all-speed flow form the non-dimensional Euler equations whose derivation and properties are addressed in the following section.

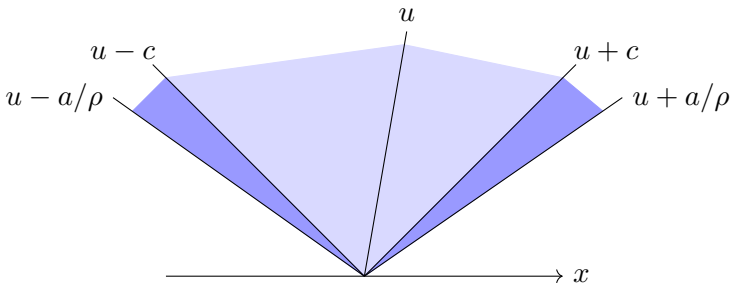


Figure 1.2: Interlacing of the eigenvalues of the Euler equations (1.23) and the Suliciu relaxation model (1.52).

1.4 The incompressible limit of the Euler equations

To keep the derivations as general as possible, we turn back to the Euler equations (1.16) in d space dimensions. The flow regime of the Euler equations can be characterized by the so called *Mach number* which gives the ration between the gas velocity $|\mathbf{u}|$ and the sound speed c. Even though the Mach number does not directly appear in the dimensional form of the Euler equations (1.16), it plays a crucial role for the properties of the gas under consideration as well as for the design of stable numerical schemes. To visualize the influence of the Mach number in the equations, we rewrite (1.16) in its so called scaled or *non-dimensional* form. Therefore we write each variable as a non-dimensional quantity, denoted by $(\tilde{\cdot})$, multiplied by a reference value, denoted by the subscript $(\cdot)_r$. The reference value contains the units in the SI standard formulation which means we give the reference time t_r in seconds $[s]$, the reference space x_r in meters $[m]$, the reference velocities u_r, c_r in $[m/s]$ and the reference density ρ_r in $[kg/m^3]$. The remaining reference pressure p_r and internal energy e_r can be computed by $p_r = \rho_r c_r^2$ and $e_r = c_r^2$. In this spirit, the dimensional variables can be written as

$$\begin{aligned} x &= \tilde{x}\ x_r, \quad t = \tilde{t}\ t_r, \quad \mathbf{u} = \tilde{\mathbf{u}}\ \frac{x_r}{t_r}, \\ \rho &= \tilde{\rho}\ \rho_r, \quad p = \tilde{p}\ \rho_r c_r^2, \quad e = \tilde{e}\ c_r^2. \end{aligned} \tag{1.64}$$

Relations (1.64) are inserted into the Euler equations (1.16) and after some algebraic manipulations the non-dimensional Euler equations are obtained. Thereby most of the reference values cancel out and the only non-dimensional quantity left is the reference Mach number defined as

$$M = \frac{u_r}{c_r}. \tag{1.65}$$

Omitting the tilde in the notation of the non-dimensional variables, the scaled equations read

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{u}) &= 0, \\ \partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u} + \frac{p}{M^2} \mathbb{I}) &= 0, \\ \partial_t E + \nabla \cdot (\mathbf{u}(E+p)) &= 0. \end{aligned} \tag{1.66}$$

where the non-dimensional total energy is given by

$$E = \rho\left(\frac{1}{2}M^2|\mathbf{u}|^2 + e\right). \tag{1.67}$$

We see from equations (1.66) that the Mach number has an influence on the pressure gradient in the momentum equation. The Mach number also arises in the acoustic waves. The wave speeds for (1.66) are given by

$$\lambda^- = u_n - \frac{c}{M}, \quad \lambda^u = u_n, \quad \lambda^+ = u_n + \frac{c}{M}, \quad u_n = \langle \mathbf{u}, \mathbf{n}\rangle. \tag{1.68}$$

Especially for small Mach numbers, the pressure term becomes large and the acoustic wave speeds tend to infinity. Possible consequences that arise in the construction of a stable numerical scheme for small Mach numbers regarding the time integration are discussed in Section 2.3.

Small or vanishing Mach numbers can arise when the reference sound speed tends to infinity which is typical for incompressible materials. The investigation of the low Mach limit started with the seminal work by Klainerman & Majda [46, 47] for the isentropic Euler equations. The non-isentropic case is studied for example in [26, 70, 40]. Therein it was formally shown that in the low Mach limit $M \to 0$ the compressible Euler equations (1.66) tend under suitable boundary conditions to the incompressible Euler equations

$$\begin{aligned} &\rho = \text{const}, \\ &\partial_t \mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} + \nabla P = 0, \\ &\nabla\cdot\mathbf{u} = 0. \end{aligned} \tag{1.69}$$

We will give a short review on how the incompressible equations (1.69) are formally obtained from (1.66) following Dellacherie [26]. First the non-dimensional equations are reformulated in primitive variables $\rho, \mathbf{u}, p$ obtaining

$$\begin{aligned} \partial_t\rho + \nabla\cdot(\rho\mathbf{u}) &= 0, \\ \rho(\partial_t\mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u}) + \frac{\nabla p}{M^2} &= 0, \\ \partial_t p + \mathbf{u}\cdot\nabla p + \rho c^2\nabla\cdot\mathbf{u} &= 0. \end{aligned} \tag{1.70}$$

Suppose an asymptotic expansion of the variables $\rho, \mathbf{u}, p$ with respect to the Mach number as

$$\begin{aligned} \rho(\mathbf{x},t) &= \rho_0(\mathbf{x},t) + \mathcal{O}(M), \\ p(\mathbf{x},t) &= p_0(\mathbf{x},t) + Mp_1(\mathbf{x},t) + M^2p_2(\mathbf{x},t) + \mathcal{O}(M^3), \\ \mathbf{u}(\mathbf{x},t) &= \mathbf{u}_0(\mathbf{x},t) + \mathcal{O}(M). \end{aligned} \tag{1.71}$$

Then we insert (1.71) into (1.70) and separate the $\mathcal{O}(M^{-2})$ and $\mathcal{O}(M^{-1})$ terms. Thereby we find $\nabla p_0 = 0$ and $\nabla p_1 = 0$ which means that the pressure is constant in space up to a perturbation in M^2. Therefore we can write for the pressure expansion $p(\mathbf{x},t) = p_0(t) + M^2p_2(\mathbf{x},t) + \mathcal{O}(M^3)$ and we find for the $\mathcal{O}(M^0)$ terms

$$\begin{aligned} &\partial_t\rho_0 + \nabla\cdot(\rho_0\mathbf{u}_0) = 0, \\ &\rho_0\left(\partial_t\mathbf{u}_0 + \mathbf{u}_0\cdot\nabla\mathbf{u}_0\right) + \nabla p_2 = 0, \\ &\nabla\cdot\mathbf{u}_0 = \frac{\partial_t p_0(t)}{\rho_0 c_0^2}, \end{aligned} \tag{1.72}$$

where $c_0 = c(p_0, \rho_0)$. For a bounded domain V with slipping or periodic boundary conditions, we find via the Gauss theorem that $\int_V \nabla \cdot \mathbf{u}\, d\mathbf{x} = 0$. Therefore when solving the Euler equations on such a domain V, we find $\partial_t p_0(t) = 0$ since $\rho_0 c_0^2 > 0$. This means the pressure is given by

$$p(\mathbf{x}, t) = p_0 + \mathcal{O}(M^2), \text{ where } p_0 = \text{ const.} \tag{1.73}$$

Then it follows directly from (1.72) that $\nabla \cdot \mathbf{u}_0 = 0$. At this point we can define the so called set of *well-prepared data* in the context of the Euler equations as

$$\Omega_{wp} = \{w \in \Omega_{phy}; \nabla\rho_0 = 0, \nabla \cdot \mathbf{u}_0 = 0, \nabla p_0 = 0, \nabla p_1 = 0\}. \tag{1.74}$$

Thereby a solution w of the compressible Euler equations which is well-prepared, is close to a solution of the incompressible Euler equations (1.69) for small Mach numbers.

The goal of this work is to develop all-speed schemes, which are numerical methods that are stable and produce an accurate solution for all Mach regimes. This includes taking into account the physical properties of the flow. This means that for small Mach numbers, a discretization of the non-dimensional Euler equations (1.66) has to be a consistent discretization of the incompressible Euler equations (1.69) such that the diagram depicted in Fig. 1.3 is satisfied. A scheme that has this property is called *asymptotic preserving* (AP). In terms of invariant domains we can formulate the AP property such that if the initial condition $w^0(\mathbf{x})$ is well-prepared, i.e. lies in Ω_{wp}, then the numerical approximation of $w(\mathbf{x}, t)$ lies in Ω_{wp} for all later times $t > 0$.

A great part of the description of the all-speed scheme for the Euler equations (1.66) presented in Chapter 4 is dedicated to the proof and the numerical testing of the AP property. The same applies for the all-speed scheme for the Euler equations with a gravitational potential presented in Chapter 5. To lay the basis for the latter case, we give a short review on the compressible Euler equations with a gravitational source term in the following section.

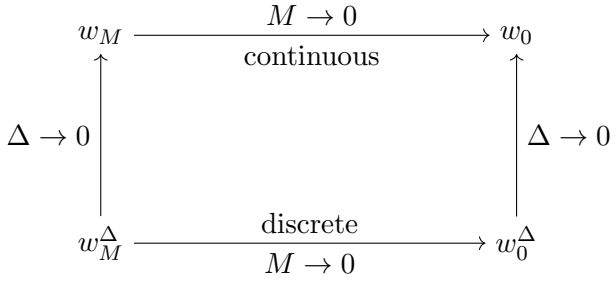


Figure 1.3: Diagram of the asymptotic limit: w_M denotes the solution of the compressible equations, w_M^Δ the discrete solution of compressible equations, w_0 solution to the limit equations and w_0^Δ the discrete solution of the limit equations.

1.5 The Euler equations with a source term

For atmospheric flows, arising in meteorology or in astrophysical applications, the influence of a *gravitational potential* is very important. Due to the presence of a gravitational

field, the gas is accelerated and the conservation of momentum does not hold any more. Therefore the Euler equations (1.16) are augmented by source terms and the full equations read

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{u}) &= 0, \\ \partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u} + p \mathbb{I}) &= -\rho \nabla \Phi, \\ \partial_t E + \nabla \cdot (\mathbf{u}(E+p)) &= -\mathbf{u} \cdot \rho \nabla \Phi, \end{aligned} \tag{1.75}$$

where the gravitational potential $\Phi : \mathbb{R}^d \to \mathbb{R}$ is assumed to be a smooth function. For simplicity we consider here the special case of a given time-independent potential. In the case of self-gravitation, which we will not consider here, the gravitational potential Φ is described by a Poisson equation

$$\Delta \Phi = 4\pi g \rho, \tag{1.76}$$

where g denotes a gravitational constant. The total energy of the system (1.75) is given by

$$E^\Phi = E + \rho\Phi = \rho \left(\frac{1}{2} |\mathbf{u}|^2 + e + \Phi \right) \tag{1.77}$$

containing also the potential energy $\rho\Phi$. Rewriting the last equation in (1.75) in terms of the total energy E^Φ with a time-independent potential Φ, gives the conservation of the total energy

$$\partial_t E^\Phi + \nabla \cdot \big(\mathbf{u}(E^\Phi + p) \big) = 0, \tag{1.78}$$

as considered for example in [10]. Since the momentum is not conserved, system (1.75) is called a system of *balance laws*.

Equations (1.75) can be written in compact notation in the variables (w, Φ) as a Cauchy problem

$$\begin{aligned} &\partial_t w + \nabla \cdot f(w) = -s(w) \nabla \Phi, \\ &\partial_t \Phi = 0, \\ &w^0(\mathbf{x}) = w(\mathbf{x}, 0), \end{aligned} \tag{1.79}$$

where the flux function and the source term are given by

$$f(w) = \begin{pmatrix} \rho \mathbf{u} \\ \rho \mathbf{u} \otimes \mathbf{u} + p\, \mathbb{I} \\ \mathbf{u}(E+p) \end{pmatrix}, \quad s(w) = \begin{pmatrix} 0 \\ \rho \\ \rho \cdot \mathbf{u} \end{pmatrix}. \tag{1.80}$$

Analogue to the homogeneous case, we can define the concept of weak solutions for the Cauchy problem (1.79).

Definition 1.5. *A function $w \in (L^\infty_{loc}(\mathbb{R}^d \times \mathbb{R}^+))^k$ is a weak solution of (1.79) with an initial condition $w^0 \in (L^\infty_{loc}(\mathbb{R}^d))^k$, if it satisfies*

$$\begin{aligned} \int_0^\infty \int_{\mathbb{R}}^d \left(\langle w, \partial_t \phi \rangle + \sum_{i=1}^d \langle f_i(w), \partial_{x_i} \phi \rangle + \langle s(w) \nabla \Phi, \phi \rangle \right) d\mathbf{x} dt \\ + \int_{\mathbb{R}^d} \big\langle w^0(\mathbf{x}), \phi(\mathbf{x}, 0) \big\rangle \, d\mathbf{x} = 0 \end{aligned} \tag{1.81}$$

for all test functions $\phi \in (C_0^\infty(\mathbb{R}^d \times \mathbb{R}^+))^k$.

To rule out inadmissible weak solutions, we can define Rankine-Hugoniot conditions. Since the potential is assumed to be smooth and therefore bounded, the Rankine-Hugoniot conditions established for the homogeneous case (1.6) still hold. A notion of entropy can also be established for (1.79) and we refer for details for example to [11, 28, 65] and references therein.

To analyse the wave structure of (1.75), let us consider for simplitiy the one dimensional case of the equations (1.75) with $\mathbf{x} = x$. We can rewrite the inhomogeneous equations (1.75) in primitive variables $\mathcal{V} = (\rho, u, e, \Phi)$ in non-conservative form

$$\partial_t \mathcal{V} + \mathcal{B}(\mathcal{V})\partial_x \mathcal{V} = 0 \tag{1.82}$$

with

$$\mathcal{B}(\mathcal{V}) = \begin{pmatrix} u & \rho & 0 & 0 \\ \frac{\partial_\rho p}{\rho} & u & \frac{\partial_e p}{\rho} & 1 \\ 0 & \frac{p}{\rho} & u & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.83}$$

The matrix $\mathcal{B}(\mathcal{V})$ has the eigenvalues

$$\lambda_1 = u - c, \quad \lambda_2 = u, \quad \lambda_3 = u + c, \quad \lambda_4 = 0, \tag{1.84}$$

where the zero wave speed corresponds to the gravitational potential. The associated eigenvalues are given by

$$\mathbf{r}_1 = \begin{pmatrix} 1 \\ -c/\rho \\ p/\rho^2 \\ 0 \end{pmatrix}, \quad \mathbf{r}_2 = \begin{pmatrix} 1 \\ 0 \\ -e/\rho \\ 0 \end{pmatrix}, \quad \mathbf{r}_3 = \begin{pmatrix} 1 \\ c/\rho \\ p/\rho^2 \\ 0 \end{pmatrix} \mathbf{r}_4 = \begin{pmatrix} 1 \\ -u/\rho \\ p/\rho^2 \\ (u^2 - c^2)/\rho \end{pmatrix}. \tag{1.85}$$

The eigenvectors are linearly degenerate as long as $u \neq 0$. However in a resonant point, that is $u = c$, the eigenvalue $\lambda_4 = 0$ has multiplicity two and the eigenvector $\mathbf{r}_4$ coincides with $\mathbf{r}_1$ which means $\mathcal{B}(\mathcal{V})$ is not diagonalizable. Apart from that resonant point, the system is hyperbolic and we immediately see that the characteristic field associated to the zero eigenvalue is linear degenerate. Therefore it exhibits a contact wave which is well-defined. Considering a Riemann problem with piecewise constant initial data, the potential only jumps across the $\lambda = 0$ curve.

Especially interesting are stationary solutions of (1.79) which are independent of time and fulfil the following balance between flux and source term

$$\nabla \cdot f(w) = s(w)\nabla\Phi. \tag{1.86}$$

An important class of stationary solutions of the Euler equations with gravity are given by *hydrostatic equilibria* which are stationary states at rest, that is $\mathbf{u} = 0$. Inserting this constraint into (1.86) with the flux given by system (1.75), we find the hydrostatic equilibrium equation

$$\nabla p = -\rho\nabla\Phi. \tag{1.87}$$

System (1.87) is under-determined since the pressure p is a dependent variable of density and internal energy and therefore (1.87) equipped with an initial condition $w^0(\mathbf{x})$ might

have infinitely many solutions. If however additional information on the physical framework is supplied, the uniqueness of solutions can be recovered. Here we will assume one class of EOS, that is frequently used to model atmospheric or astrophysical flows. Let p fulfil the following relation

$$p = \chi \rho^{\Gamma} \tag{1.88}$$

with the constant $\chi > 0$ and $\Gamma \in (0, \infty)$. An *isothermal atmosphere* is given by $\chi = R\,T$, where R is the specific gas constant and T a fixed temperature. The adiabatic exponent is given by $\Gamma = 1$ and the isothermal EOS reads

$$p = RT\rho. \tag{1.89}$$

Inserting this relation in the hydrostatic equation (1.87) yields for a constant C

$$\rho(\mathbf{x}) = \exp\left(\frac{C - \Phi(\mathbf{x})}{RT}\right), \quad p(\mathbf{x}) = RT\rho(\mathbf{x}). \tag{1.90}$$

For $\Gamma \in (0,1) \cup (1,\infty)$ one obtains a *polytropic atmosphere*, where the equilibrium solution reads for a constant C

$$\rho(\mathbf{x}) = \left(\frac{\Gamma - 1}{\chi^{\Gamma}}\left(C - \Phi(\mathbf{x})\right)\right)^{1/(\Gamma-1)}, \quad p(\mathbf{x}) = \chi\rho(\mathbf{x})^{\Gamma}. \tag{1.91}$$

A special case of the polytropic atmosphere is the *isentropic atmosphere* which is obtained for $\Gamma = \gamma$. This choice leads to the fact, that the energy equation in (1.16) and (1.75) is always fulfilled and the mass and momentum equations are sufficient to describe the gas.

In nature, often fluctuations around such equilibrium solutions occur. To accurately compute their evolution, one has to make sure that hydrostatic equilibrium solutions of (1.87) are numerically well captured and the numerical method does not contribute spurious fluctuations itself. Therefore it is a goal of this work to develop schemes for the Euler equations with a gravitational source term (1.75) that are able to preserve a discrete version of the hydrostatic equilibrium on machine precision.

Chapter 2

Finite volume methods

This chapter is dedicated to the numerical methods used to approximate the evolution of compressible gases given by Euler equations which were described in the previous chapter. A good approximation of the continuous equations has to preserve their properties also on the discrete level. Since integral formulation of the considered homogeneous equations have the conservation property, the finite volume (FV) framework provides a fitting discretization framework which naturally leads to methods in conservation form.

The numerical methods based on the FV framework discussed in this chapter are taken from the text-books by Bouchut [11], Leveque [52] and Toro [80]. First we give a brief derivation of the general form of conservative schemes in one space dimension in the FV framework continued by the Godunov scheme based on exact Riemann solvers and Godunov-type schemes based on approximate Riemann solvers. A strategy to extend the one dimensional Godunov schemes to multiple space dimensions is given by considering split dimensional Riemann problems. Subsequent we give a brief review on the MUSCL approach to obtain a scheme that is second order accurate in space based on a first order FV scheme introduced in the previous section. Concerning the numerical treatment of the time derivatives, we give examples of second order explicit time integration schemes and motivate the use of an implicit-explicit (IMEX) approach when considering the Euler equations for small Mach numbers. The chapter is completed by addressing Godunov type schemes for hyperbolic problems with source terms as arising in the case of the Euler equations with a gravitational potential.

2.1 First order finite volume schemes for conservation laws

We consider the following one dimensional (1D) initial value problem for a system of hyperbolic conservation laws

$$\begin{cases} \partial_t W(x,t) + \partial_x F(W(x,t)) = 0, \\ W^0(x) = W(x,0), \end{cases} \tag{2.1}$$

where $W : \mathbb{R} \times \mathbb{R}^+ \mapsto \Omega \subset \mathbb{R}^k$ is the vector of conserved variables, $F : \Omega \to \mathbb{R}^k$ the flux function. The set Ω denotes the set of admissible states.

The first step towards a numerical scheme is the discretization of the computational domain $I \subset \mathbb{R}$ which is given by an interval in 1D. Throughout this work we will consider

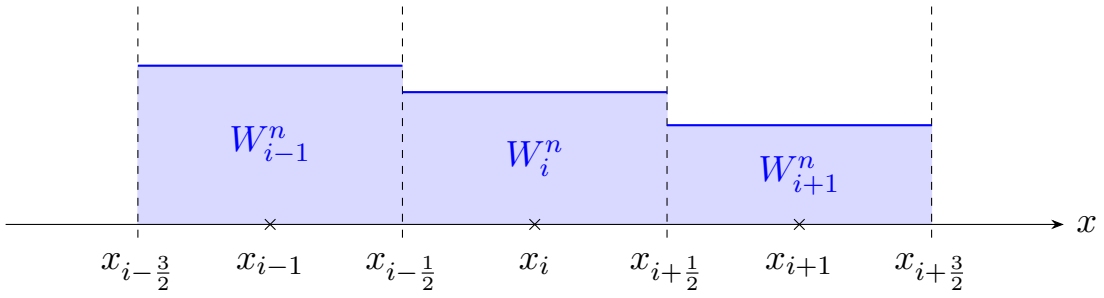


Figure 2.1: Finite volume discretization.

a Cartesian mesh with a uniform grid size Δx. In this spirit, the interval I is subdivided into N grid cells $C_i = (x_{i-1/2}, x_{i+1/2})$ of length Δx with the cell center $x_i = i\Delta x$ for $i = 1, \dots, N$. The time line $\mathbb{R}^+$ is discretized as $t^{n+1} = t^n + \Delta t$ where the time increment Δt is variable and can underlie a stability restriction.

As the name finite volume indicates, the solution of (2.1) on the cell C_i is approximated by the average of the solution at time t^n over C_i as follows

$$W_i^n \approx \frac{1}{\Delta x} \int_{C_i} W(x, t^n) dx. \tag{2.2}$$

This results in a piecewise constant approximation at time t^n as illustrated in Fig. 2.1. To obtain the update W_i^{n+1} at the next time level t^{n+1} from known cell averages W_i^n at the previous time step, let us look at the average of the IVP (2.1). Assuming W sufficiently smooth on $(t^n, t^{n+1}]$, we integrate (2.1) on $C_i \times (t^n, t^{n+1}]$ which yields

$$\begin{aligned} \int_{C_i} W\left(x, t^{n+1}\right) dx - \int_{C_i} W\left(x, t^n\right) dx \\ + \int_{t^n}^{t^{n+1}} F\left(W(x_{i+1/2}, t)\right) dt - \int_{t^n}^{t^{n+1}} F\left(W(x_{i-1/2}, t)\right) dt = 0. \end{aligned} \tag{2.3}$$

Rearranging the terms and using the definition of cell averages (2.2), leads to

$$W_i^{n+1} = W_i^n - \frac{1}{\Delta x} \left(\int_{t^n}^{t^{n+1}} F(W(x_{i+1/2}, t)) dt - \int_{t^n}^{t^{n+1}} F(W(x_{i-1/2}, t)) dt \right). \tag{2.4}$$

According to (2.4), the cell average on cell C_i at the new time is updated by determining the fluxes through the cell interfaces $x_{i-1/2}$, $x_{i+1/2}$. In practise the exact integrals of the flux function are in general difficult to obtain. Therefore, we define the numerical flux as an approximation of the flux average along $x = x_{i\pm1/2}$ on $(t^n, t^{n+1}]$ given by

$$\mathcal{F}_{i+1/2}^{\Delta} \approx \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} F(W(x_{i+1/2}, t)) dt.$$

Inserting this definition into (2.4), the state W^{n+1} can be obtained by the recursive relation

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x} \left(\mathcal{F}_{i+1/2}^{\Delta} - \mathcal{F}_{i-1/2}^{\Delta} \right), \quad \text{where} \quad W_i^0 \approx \frac{1}{\Delta x} \int_{C_i} W^0(x) dx. \tag{2.5}$$

An explicit scheme can be obtained, when the numerical flux $\mathcal{F}^{\Delta}_{i+1/2}$ is given in dependence of data at the previous time step only. For a first order approximation it suffices to base the numerical flux at the interface $x_{i+1/2}$ on the direct neighbours W^n_i, W^n_{i+1} and we can write $\mathcal{F}^{\Delta}_{i+1/2}$ as a function of these values given by $\mathcal{F}^{\Delta}_{i+1/2} = \mathcal{F}(W^n_i, W^n_{i+1})$. The explicit scheme is then given by

$$W^{n+1}_i = W^n_i - \frac{\Delta t}{\Delta x}\left(\mathcal{F}(W^n_i, W^n_{i+1}) - \mathcal{F}(W^n_{i-1}, W^n_i)\right). \tag{2.6}$$

A method, that can be written in the form of (2.6) is said to be *conservative* as it is a discrete analogue of the integral formulation (1.1) of the IVP. Indeed, summing (2.6) over all cells C_i for $i = 1, \dots, N$, we obtain the Riemann sum approximation

$$\Delta x \sum_{i=1}^{N} W^{n+1}_i = \Delta x \sum_{i=1}^{N} W^n_i - \Delta t \left(\mathcal{F}^{\Delta}_{N+1/2} - \mathcal{F}^{\Delta}_{1/2}\right). \tag{2.7}$$

Due to the telescope sum, the flux terms cancel out and only the fluxes over the boundaries remain. To obtain a scheme that is conservative on the whole computational domain $I = [x_{1/2}, \dots, x_{N+1/2}]$, the choice of boundary condition is important. To simplify the notation, we introduce ghost cells to extend the computational domain as depicted in Figure 2.2. In the fully discrete scheme (2.6), the fluxes at the boundaris depend on the ghost cells W_0 at the left and on W_{N+1} at the right boundary. Those are set according to the chosen boundary condition. Throughout the manuscript, the following boundary conditions are used

- *Periodic* boundary conditions: Due to the periodicity, we have $W_0 = W_N$ and $W_{N+1} = W_1$ which means that in the update (2.6), the fluxes at the boundaries coincide and the simulated quantities are conserved on the computational domain.
- *Dirichlet/Exact* boundary conditions: The value at the boundary is given exactly. Especially when knowing the exact solution $W(x,t)$ of (2.1), for example when computing convergence rates, the values of the ghost cells are given by

$$W^n_0 \approx \frac{1}{\Delta x}\int_{C_0} W(x, t^n)dx, \quad W^n_{N+1} \approx \frac{1}{\Delta x}\int_{C_{N+1}} W(x, t^n)dx. \tag{2.8}$$

- *Neumann* boundary conditions: They describe a zero flux over the boundary and are given by $\partial_x W(x_{1/2}, t) = 0$ and $\partial_x W(x_{N+1/2}, t) = 0$ in one dimension. The simplest way of approximation is to set $W_0 = W_1$ and $W_{N+1} = W_N$. Another option is to approximate the derivative by higher order extrapolation methods based on interior domain values.

Before defining a specific numerical flux function $\mathcal{F}$, we have to give criteria what forms a good flux function. Above all, the numerical solution should converge to the true solution of (2.1) as the grid is refined, that is $\Delta = (\Delta x, \Delta t) \to 0$. A necessary condition for convergence is *consistency*. A scheme is consistent with the original conservation law (2.1) if the numerical flux function reduces to the true flux F in case of constant flow, i.e. $\mathcal{F}(W, W) = F(W)$.

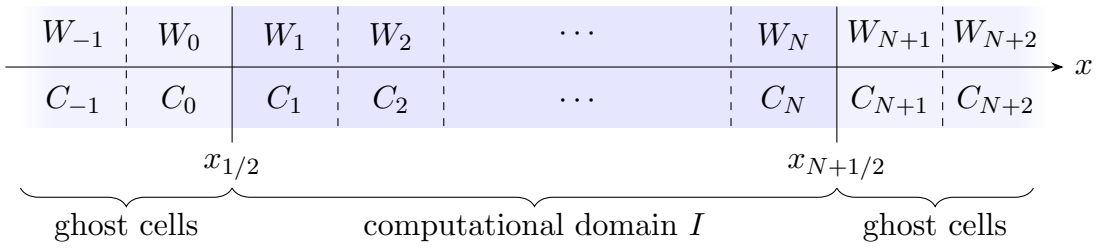


Figure 2.2: Discretization of the computational domain with ghost cells.

A second necessary condition is that information must be propagated at the correct physical speeds determined by the eigenvalues of the flux Jacobian $\partial_W F(W)$. This condition can be reformulated as a restriction on the time step Δt also known as the Courant-Friedrichs-Lewy (CFL) condition

$$\frac{\Delta x}{\Delta t} \leq \lambda_{\max}, \tag{2.9}$$

where

$$\lambda_{\max} = \max_{i,\kappa} |\lambda_\kappa(W_i^n)|, \quad \kappa = 1, \ldots, k, \quad i = 1, \ldots, N, \tag{2.10}$$

is the maximum of the absolute value of all eigenvalues. It ensures, that the information in the numerical scheme is not propagated faster than the correct physical speed.

Unfortunately these conditions are only necessary but not sufficient to ensure convergence. However, the following result of Lax and Wendroff reassures that when a convergent method is found that the resulting numerical solution is an approximation of a weak solution for (2.1).

Theorem 2.1 (Lax-Wendroff, taken from [37])**.** *Let W_{Δ_l} be given by a scheme (2.5) which is in conservation form and consistent with (2.1). Assume there exists a sequence $\Delta x_l \to 0$ such that the ratio $\Delta x_l / \Delta t_l$ is constant. Suppose the following holds:*

- $\|W_{\Delta_l}\|_{L^\infty(\mathbb{R}\times\mathbb{R}^+)} \leq K, \quad K$ *const.*
- *The sequence $W_{\Delta l}$ converges in $L^1_{loc}(\mathbb{R} \times \mathbb{R}^+)$ to a function W.*

Then W is a weak solution of (2.1).

To numerically verify convergence and the rate of convergence of the schemes throughout this work, we compute the L^1 norm of the difference between numerical and exact solution at a final time $t^m = T$ on a sequence of refined grids

$$\mathrm{err}_{1,T}^l = \Delta x_l \sum_{i=1}^{N_l} |W_i^m - W(x, T)| \,. \tag{2.11}$$

The convergence rates between two consecutive grids are then computed as

$$\mathrm{rate}_{1,T} = \frac{\log\left(\mathrm{err}_{1,T}^{l-1} / \mathrm{err}_{1,T}^{l}\right)}{\log\left(\Delta x_{l-1} / \Delta x_l\right)}. \tag{2.12}$$

The Lax-Wendroff Theorem does not ensure that the obtained weak solution is an admissible weak solution. To rule out entropy violating numerical solutions, it is required to be consistent with the entropy condition. We refer to e.g. [11, 25, 37] for details on entropy stability for explicit FV schemes of type (2.5).

2.1.1 Godunov's scheme

Looking at the piecewise constant solution given by the cell averages W_i for $i = 1, \dots, N$, depicted in Figure 2.1, we see that at each interface arises a local Riemann problem

$$\begin{cases} \partial_t W + \partial_x F(W) = 0 \\ W^0(x) = \begin{cases} W_i^n & \text{if } x < x_{i+1/2} \\ W_{i+1}^n & \text{if } x > x_{i+1/2}. \end{cases} \end{cases} \tag{2.13}$$

The basis of the (exact) Godunov scheme is the assumption that a solution of the Riemann problem (2.13) exists and that it can be given analytically in an explicit or implicit expression. In general, the scheme can be divided in three parts, also known as the Reconstruct-Evolve-Average (REA) algorithm. In the *reconstruction* step, a piecewise polynomial function $\widetilde{W}(x, t^n)$ is defined from cell averages at time t^n. For a first order scheme Godunov proposed to use the cell average value on the grid cell, that is $\widetilde{W}(x, t^n) = W_i^n$ for $x \in (x_{i-1/2}, x_{i+1/2})$. In the *evolution* step, the local Riemann Problems (2.13) are solved exactly. The solution $\widetilde{W}(x, t^{n+1})$ at the next time step $t^{n+1} = t^n + \Delta t$ is piecewise defined by the local Riemann solutions $\mathcal{W}_{i+1/2}\left(\frac{x - x_{i+1/2}}{t - t^n}; W_i^n, W_{i+1}^n\right)$ at $t = t^{n+1}$ on $x \in (x_i, x_{i+1})$ for $i = 1, \dots, N$. To ensure that the Riemann solutions do not interact, it has to be ensured that the waves from adjacent Riemann problems travel maximal of distance $\Delta x/2$. This can be achieved by adjusting the time increment Δt as depicted in Figure 2.3 for the local Riemann problems at the interfaces $x_{i-1/2}, x_{i+1/2}$. The CFL restriction for Godunov's scheme is thus given by

$$\Delta t \leq \frac{1}{2} \frac{\Delta x}{\lambda_{\max}}, \tag{2.14}$$

where $\lambda_{\max}$ is the largest absolute wave speed on the whole computational domain at time t^n as defined in (2.10). To maintain the cell average structure at the new time level, the solution $\widetilde{W}(x, t^{n+1})$ is *averaged* in the last step of the RSA algorithm over each grid cell as

$$W_i^{n+1} = \frac{1}{\Delta x} \int_{C_i} \widetilde{W}(x, t^{n+1}) dx. \tag{2.15}$$

The procedure of Godunov's scheme is depicted in Figure 2.4.

As we have seen previously, it is desirable to have a scheme in conservation form when approximating conservation laws. To see that Godunov's scheme fulfils this property, we analyse the solution $\widetilde{W}(x, t^{n+1})$ on cell C_i. Let $\lambda_{i+1/2}^-$ be the fastest wave speed from the Riemann problem at interface $x_{i+1/2}$ travelling into the cell C_i from the right and $\lambda_{i-1/2}^+$ the fastest wave from the Riemann problem at $x_{i-1/2}$ travelling into the cell C_i from the left, as illustrated in Figure 2.4. We can divide the cell C_i into three distinct intervals on

which holds

$$\widetilde{W}(x,t^{n+1}) = \begin{cases} \mathcal{W}_{i-1/2}(\frac{x-x_{i-1/2}}{t-t^n}), & x \in [x_{i-1/2}, x_{i-1/2} + \lambda^+_{i-1/2}\Delta t], \\ W_i^n, & x \in [x_{i-1/2} + \lambda^+_{i-1/2}\Delta t, x_{i+1/2} - \lambda^-_{i+1/2}\Delta t], \\ \mathcal{W}_{i+1/2}(\frac{x-x_{i+1/2}}{t-t^n}), & x \in [x_{i+1/2} - \lambda^-_{i+1/2}\Delta t, x_{i+1/2}]. \end{cases} \tag{2.16}$$

Since $\widetilde{W}(x,t^{n+1})$ is an exact solution of the local Riemann problem (2.13) on $(t^n, t^{n+1}]$, integrating the equations (2.1) on these three distinct intervals given in (2.16) yields

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x}\left(F(\mathcal{W}_{i+1/2}(0)) - F(\mathcal{W}_{i-1/2}(0))\right). \tag{2.17}$$

While rearranging (2.17), we have used that the Riemann solutions $\mathcal{W}_{i-1/2}\left(\frac{0}{t-t^n}\right)$ and $\mathcal{W}_{i+1/2}\left(\frac{0}{t-t^n}\right)$ are constant along $t \in (t^n, t^{n+1}]$.

The update (2.17) in conservation form is also known as the second formulation of Godunov's method which is easy to apply in practice. In addition, since the Riemann solutions are only evaluated at the interfaces, the quite restrictive CFL condition (2.14) can be relaxed to allowing the incoming waves to travel up to the whole distance Δx.

2.1.2 Godunov type scheme

In practice, calculating the exact Riemann solution can be very complicated and computationally inefficient, as in the case of the Euler equations where the pressure is given by an implicit relation. Therefore, in *Godunov type* schemes, the exact Riemann solution is replaced by an approximation. Like Godunov's scheme, schemes of Godunov type have to be consistent with the original conservation law. Therefore, an *approximate Riemann solver* $W_{\mathcal{R}}(x/t; W_i, W_{i+1})$ has to satisfy

$$W_{\mathcal{R}}\left(x/t; W, W\right) = W, \text{ for all } W \in \Omega. \tag{2.18}$$

In addition, it has to fulfil the *integral consistency condition*

$$\int_{x_i}^{x_{i+1}} W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}; W_i, W_{i+1}\right) dx = \int_{x_i}^{x_{i+1}} \widetilde{W}\left(\frac{x - x_{i+1/2}}{\Delta t}; W_i, W_{i+1}\right) dx, \tag{2.19}$$

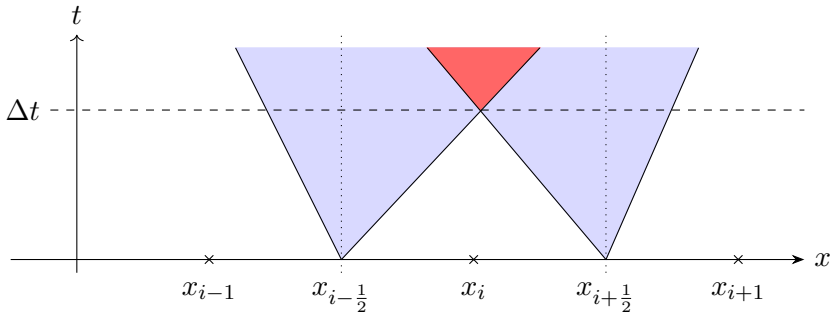


Figure 2.3: Illustration of the CFL condition for the first version of Godunov's method.

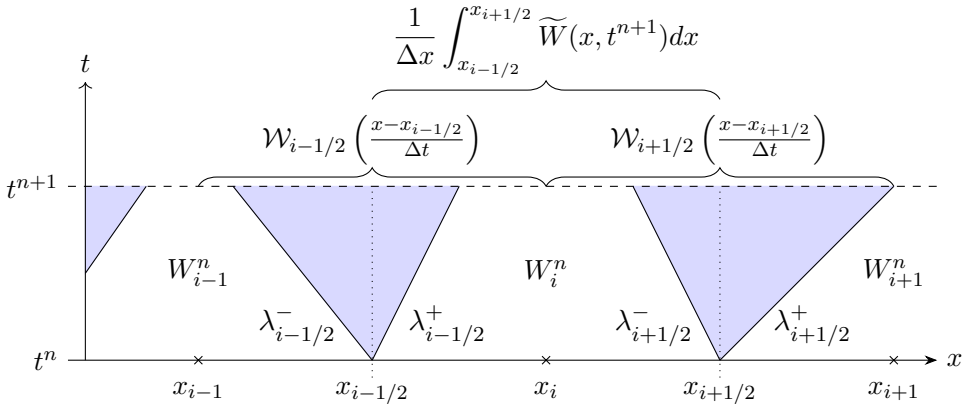


Figure 2.4: Illustration of the (exact) Godunov scheme.

where $\widetilde{W}(x/t; W_i, W_{i+1})$ denotes the exact Riemann solution of (2.13) on $(t^n, t^{n+1}]$. Analogously to Godunov's scheme, the approximate solution $\widetilde{W}_\Delta(x, t^{n+1})$ at the new time level t^{n+1} is piecewise given by the local approximate Riemann solutions $W_\mathcal{R}$ stemming from the Riemann problems at the interfaces $x_{i+1/2},\ i = 1, \dots, N$. The cell averages on C_i at t^{n+1} are then updated by

$$\begin{aligned} W_i^{n+1} &= \frac{1}{\Delta x}\int_{C_i} \widetilde{W}_\Delta(x, t^{n+1})dx \\ &= \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_i} W_\mathcal{R}\left(\frac{x - x_{i-1/2}}{\Delta t}, W_{i-1}, W_i\right) dx \\ &\qquad + \frac{1}{\Delta x}\int_{x_i}^{x_{i+1/2}} W_\mathcal{R}\left(\frac{x - x_{i+1/2}}{\Delta t}, W_i, W_{i+1}\right) dx. \end{aligned} \tag{2.20}$$

To be able to apply the Lax-Wendroff theorem in case of convergence, we will conclude the introduction to Godunov-type methods with the derivation of its conservation form. We reformulate the integral consistency condition (2.19) by integrating the Riemann problem (2.13) on $[x_i, x_{i+1}] \times (t^n, t^{n+1}]$. For an exact solution $\widetilde{W}$ follows

$$\begin{aligned} &\int_{x_i}^{x_{i+1}} \widetilde{W}\left(\frac{x - x_{i-1/2}}{\Delta t}; W_i, W_{i+1}\right) dx = \\ &\qquad \int_{x_i}^{x_{i+1}} \widetilde{W}\left(0; W_i, W_{i+1}\right) dx - \int_{t^n}^{t^{n+1}} F\left(\widetilde{W}\left(\frac{\Delta x}{2(t - t^n)}; W_i, W_{i+1}\right)\right) dt \\ &\qquad\qquad + \int_{t^n}^{t^{n+1}} F\left(\widetilde{W}\left(-\frac{\Delta x}{2(t - t^n)}; W_i, W_{i+1}\right)\right) dt. \end{aligned} \tag{2.21}$$

From the initial condition of the Riemann Problem we obtain immediately

$$\int_{x_i}^{x_{i+1}} \widetilde{W}\left(0; W_i, W_{i+1}\right) dx = \frac{\Delta x}{2}\left(W_i^n + W_{i+1}^n\right). \tag{2.22}$$

Under the CFL condition of Godunov's scheme (2.14), that ensures

$$\widetilde{W}\left(-\frac{\Delta x}{2(t - t^n)}; W_i, W_{i+1}\right) = W_i^n \text{ and } \widetilde{W}\left(\frac{\Delta x}{2(t - t^n)}; W_i, W_{i+1}\right) = W_{i+1}^n \tag{2.23}$$

for all $t \in (t^n, t^{n+1}]$, we can determine the flux integrals in (2.21). Using the integral consistency condition (2.19) we obtain for the approximative Riemann solver

$$\frac{1}{\Delta x}\int_{x_i}^{x_{i+1}} W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}; W_i, W_{i+1}\right) dx = \frac{1}{2}\left(W_i^n + W_{i+1}^n\right) - \frac{\Delta t}{\Delta x}\left(F\left(W_{i+1}^n\right) - F\left(W_i^n\right)\right). \quad (2.24)$$

The update of the cell average (2.20) can then be expressed by

$$\begin{aligned} W_i^{n+1} = \; & \frac{1}{2}\left(W_i^n + W_{i+1}^n\right) - \frac{\Delta t}{\Delta x}\left(F\left(W_{i+1}^n\right) - F\left(W_i^n\right)\right) \\ & + \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_i} W_{\mathcal{R}}\left(\frac{x - x_{i-1/2}}{\Delta t}; W_{i-1}, W_i\right) dx \\ & - \frac{1}{\Delta x}\int_{x_{i+1/2}}^{x_{i+1}} W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}; W_i, W_{i+1}\right) dx. \end{aligned} \quad (2.25)$$

Defining the numerical flux as

$$\mathcal{F}\left(W_i^n, W_{i+1}^n\right) = F\left(W_{i+1}^n\right) - \frac{\Delta x}{2\Delta t}\left(W_{i+1}^n - \frac{2}{\Delta x}\int_{x_{i+1/2}}^{x_{i+1}} W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}; W_i, W_{i+1}\right) dx\right), \quad (2.26)$$

we find the conservation form of the Godunov type scheme as

$$W_i^{n+1} = W_i^n - \frac{\Delta x}{\Delta t}\left(\mathcal{F}(W_i^n, W_{i+1}^n) - \mathcal{F}(W_{i-1}^n, W_i^n)\right). \quad (2.27)$$

To arrive at the numerical flux (2.26), we have reformulated the right local Riemann solution in (2.25) in terms of the integral consistency (2.24). Therefore $\mathcal{F}$ is also called the right numerical flux. An equivalent numerical flux formulation can be obtained if the left local Riemann solution is reformulated instead. This yields the following left numerical flux

$$\overline{\mathcal{F}}\left(W_i^n, W_{i+1}^n\right) = F\left(W_i^n\right) + \frac{\Delta x}{2\Delta t}\left(W_i^n - \frac{2}{\Delta x}\int_{x_i}^{x_{i+1/2}} W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}; W_i, W_{i+1}\right) dx\right). \quad (2.28)$$

Both flux definitions are consistent and as soon as the integral consistency condition is satisfied by the approximate Riemann solver $W_{\mathcal{R}}$, the fluxes fulfil the so called *conservativity identity*

$$\mathcal{F}(W_i, W_{i+1}) = \overline{\mathcal{F}}(W_i, W_{i+1}). \quad (2.29)$$

Finding approximative Riemann solvers is still an active field of research. Well known is for example the HLL scheme which was developed by Harten, Lax and van Leer [41]. It is based on an approximate Riemann solver with a simplified wave structure that consists only of the slowest and fastest wave speed arising in the Riemann fan. This means an

expression for only one intermediate state has to be found. Consequently, when looking at the Euler equations, the acoustic waves have a reasonable resolution, whereas the contact wave in between is quite diffusive. To correct this, a modification, the HLLC scheme was proposed, introducing back the contact wave into the approximate Riemann solver. This generates two unknown intermediate states and is thus closer to the true wave structure. The intermediate states can be determined by using additional knowledge of Riemann invariants across the middle wave allowing for a better description of the contact wave. See [81] for a short review of HLLC schemes for different conservation and balance laws.

2.1.3 Suliciu relaxation Godunov type scheme for the Euler equations

To conclude the revision of Godunov type schems, we give the derivation of a Godunov type scheme for the Euler equations using the Suliciu relaxation model (1.52). The approximate Riemann solvers that are presented in the manuscript lateron are derived following the steps given here. We consider the homogeneous part of the relaxation model (1.52) to construct the Riemann solver which corresponds to "$\varepsilon = \infty$" given by

$$\begin{aligned} \partial_t \rho + \partial_x(\rho u) &= 0, \\ \partial_t(\rho u) + \partial_x(\rho u^2 + \pi) &= 0, \\ \partial_t E + \partial_x(u(E+\pi)) &= 0, \\ \partial_t(\rho\pi) + \partial_x(\rho u \pi + a^2 u) &= 0. \end{aligned} \tag{2.30}$$

The relaxation source term will be treated in a subsequent step

$$\partial_t(\rho\pi) = \frac{\rho}{\varepsilon}(p - \pi). \tag{2.31}$$

Equation (2.31) is treated by projection which means, we consider the relaxation source term in the limit $\varepsilon = 0$. Due to its stiffness for small $\varepsilon > 0$ it is treated implicitly by a backward Euler step

$$\rho^{n+1}\pi^{n+1} = \rho^n\pi^n + \frac{\Delta t}{\varepsilon}\left(\rho^{n+1}p^{n+1} - \rho^{n+1}\pi^{n+1}\right). \tag{2.32}$$

Rearranging the terms we have the following update for π^{n+1}

$$\rho^{n+1}\pi^{n+1} = \frac{\varepsilon}{\varepsilon + \Delta t}\rho^n\pi^n + \frac{\Delta t}{\varepsilon + \Delta t}\rho^{n+1}p^{n+1}. \tag{2.33}$$

Taking the limit $\varepsilon \to 0$ we find immediately $\pi^{n+1} = p^{n+1}$ since $\rho^{n+1} > 0$, where p^{n+1} is calculated from the total energy E^{n+1} at t^{n+1} via the considered equation of state.

Concerning the equations (2.30), we have seen in Section 1.3.2 that the wave structure has a fixed ordering which is given by

$$\lambda^- = u - \frac{a}{\rho} < \lambda^u = u < \lambda^+ = u + \frac{a}{\rho}. \tag{2.34}$$

The fields associated to the eigenvalues are linearly degenerate and we can calculate the exact solution $W_{\mathcal{RS}}$ to the Riemann problem of the homogeneous part of the Suliciu relaxation model (2.30). To illustrate what happens at the interfaces, we will consider a

general Riemann problem for the equations (2.30) consisting of two initial constant states W_L and W_R at $x = 0$. The Riemann solution consists of four constant states, separated by the three waves $\lambda^-, \lambda^u, \lambda^+$ leading to two unknown intermediate states W_L^*, W_R^*. It is given by

$$W_{\mathcal{RS}}\left(\frac{x}{t}; W_L, W_R\right) = \begin{cases} W_L & \text{for } \frac{x}{t} < \lambda^-, \\ W_L^* & \text{for } \lambda^- < \frac{x}{t} < \lambda^u, \\ W_R^* & \text{for } \lambda^u < \frac{x}{t} < \lambda^+, \\ W_R^* & \text{for } \lambda^+ < \frac{x}{t}, \end{cases} \tag{2.35}$$

for $t > 0$. Due to the linear degeneracy of all characteristic fields, we find algebraic relations to compute the intermediate states W_L^*, W_R^* by determining the Riemann invariants across each wave. According to the eigenvectors $\mathbf{r}_{1,2}^u$ in (1.63), the Riemann invariants associated with λ^u are given by

$$I_1^u = u, \quad I_2^u = \pi, \tag{2.36}$$

which means that velocity and pressure do not change across the middle wave which gives

$$u_L^* = u_R^* = u^* \text{ and } \pi_L^* = \pi_R^* = \pi^*. \tag{2.37}$$

For the acoustic waves $\lambda^\pm$ we have the following invariants

$$I_1^\pm = u \pm \frac{a}{\rho}, \quad I_2^\pm = \pi \mp au, \quad I_3^\pm = e - \frac{\pi^2}{2a^2}. \tag{2.38}$$

Combining the relations (2.37) with (2.38), we can form an algebraic system from which the intermediate states can be computed. It is given as follows

$$u_L - \frac{a}{\rho_L} = u^* - \frac{a}{\rho_L^*}, \qquad u_R + \frac{a}{\rho_R} = u^* + \frac{a}{\rho_R^*}, \tag{2.39}$$

$$\pi_L + au_L = \pi^* + au^*, \qquad \pi_R - au_R = \pi^* - au^*, \tag{2.40}$$

$$e_L - \frac{\pi_L^2}{2a^2} = e_L^* - \frac{\pi^{*2}}{2a^2}, \qquad e_R - \frac{\pi_R^2}{2a^2} = e_R^* - \frac{\pi^{*2}}{2a^2}. \tag{2.41}$$

Straightforward computations yield

$$u^* = \frac{1}{2}(u_L + u_R) - \frac{1}{2a}(\pi_R - \pi_L), \tag{2.42}$$

$$\pi^* = \frac{1}{2}(\pi_L + \pi_R) - \frac{a}{2}(u_R - u_L), \tag{2.43}$$

$$\frac{1}{\rho_L^*} = \frac{1}{\rho_L} - \frac{1}{a}(u_L - u^*), \qquad \frac{1}{\rho_R^*} = \frac{1}{\rho_R} - \frac{1}{a}(u^* - u_R), \tag{2.44}$$

$$e_L^* = e_L - \frac{1}{2a^2}(\pi_L^2 - \pi^{*2}), \qquad e_R^* = e_R - \frac{1}{2a^2}(\pi_R^2 - \pi^{*2}). \tag{2.45}$$

To fully determine the intermediate states, we have to set the initial value for the relaxed pressure π. We start with initial data in relaxation equilibrium and set $\pi_L = p_L$ and $\pi_R = p_R$ calculated from the total energy E via the considered equation of state.

Since we have constructed an exact Riemann solution to the relaxation system, the numerical flux function at the interface $x_{i+1/2}$ is given by $F(W_{\mathcal{RS}}(0; W_i, W_{i+1}))$. In detail it is given by

$$\mathcal{F}(W_i, W_{i+1}) = F(W_{\mathcal{RS}}(0; W_i, W_{i+1})) = \begin{cases} F(W_i) & \text{if } 0 < \lambda^-, \\ F(W_i^*) & \text{if } \lambda^- < 0 < \lambda^u, \\ F(W_{i+1}^*) & \text{if } \lambda^u < 0 < \lambda^+, \\ F(W_{i+1}) & \text{if } \lambda^+ < 0. \end{cases} \tag{2.46}$$

To fully determine the fluxes, we have to specify how to compute the relaxation parameter a in the Riemann solver. The aim is to choose the parameter a as small as possible to reduce the diffusiveness of the Riemann solution and to choose it as large as necessary to guarantee stability by fulfilling the subcharacteristic condition $a \geq c\rho$ given in (1.54). One possibility would be to determine the parameter a globally on the whole computational domain according to the subcharacteristic condition (1.54). This can be done by setting

$$a \geq \max_{i \in \{1,\dots,N\}} \left(\sqrt{\gamma \rho_i^n p_i^n} \right). \tag{2.47}$$

However, looking at the local Riemann solution using the global estimate (2.47) can result in a larger relaxation parameter than necessary, since the Riemann problem depends only on the left and right states W_L, W_R. Bouchut in [11, Prop. 2.18] gives a formula on how to set the relaxation parameter a locally such that the subcharacteristic condition is satisfied. In our context for an ideal gas law with $\rho_L, \rho_R > 0$, we can directly compute the local relaxation parameter $a^* = \max(a_L, a_R)$ following Prop. 2.18, where a_L, a_R are defined by

$$\left.\begin{aligned} a_L &= \sqrt{\gamma \rho_L p_L} + 2\rho_L \max\left(\tfrac{p_R - p_L}{\sqrt{\gamma \rho_R p_R}} + u_L - u_R, 0 \right) \\ a_R &= \sqrt{\gamma \rho_R p_R} + 2\rho_R \max\left(\tfrac{p_L - p_R}{a_L} + u_L - u_R, 0 \right) \end{aligned}\right\} \quad \text{if } p_R - p_L \geq 0,$$

$$\left.\begin{aligned} a_R &= \sqrt{\gamma \rho_R p_R} + 2\rho_R \max\left(\tfrac{p_L - p_R}{\sqrt{\gamma \rho_L p_L}} + u_L - u_R, 0 \right) \\ a_L &= \sqrt{\gamma \rho_L p_L} + 2\rho_L \max\left(\tfrac{p_R - p_L}{a_R} + u_L - u_R, 0 \right) \end{aligned}\right\} \quad \text{if } p_R - p_L < 0. \tag{2.48}$$

Now that the intermediate states are well-defined, the first order scheme for the relaxation model is given by

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x} \left(\mathcal{F}(W_i, W_{i+1}) - \mathcal{F}(W_{i-1}, W_i) \right) \tag{2.49}$$

with the numerical flux $\mathcal{F}$ defined in (2.46) under the CFL condition according to (2.14)

$$\frac{\Delta t}{\Delta x} \max\left(\left| u_i + \frac{a_{i-1/2}^*}{\rho_i} \right|, \left| u_i - \frac{a_{i+1/2}^*}{\rho_i} \right| \right) \leq \frac{1}{2}. \tag{2.50}$$

Now, we have constructed a Godunov scheme for the relaxation system (1.52), but actually we are interested in solving the original equations (1.16). The connection can be established in the notation introduced in Section 1.3 as follows. Let the relaxation system be denoted by

$$\partial_t W + \partial_x F(W) = R(W) \tag{2.51}$$

and the original equations by

$$\partial_t w + \partial_x f(w) = 0. \tag{2.52}$$

Let $W_{\mathcal{RS}}$ be an approximate or exact Riemann solver for the relaxation system (1.52), then

$$W_{\mathcal{R}}(x/t; w_L, w_R) = QW_{\mathcal{RS}}(x/t; \mathcal{M}(w_L), \mathcal{M}(w_R)) \tag{2.53}$$

is an approximate Riemann solver for the original equations (1.16), where $\mathcal{M}(w)$ denotes the state vector of the relaxation system in equilibrium

$$\mathcal{M}(w) = (\rho, \rho u, E, \rho p)^T \tag{2.54}$$

and the linear operator Q is given by

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{2.55}$$

The consistency of $W_{\mathcal{R}}$ follows directly from the consistency of $W_{\mathcal{RS}}$ as

$$W_{\mathcal{R}}(x/t; w, w) = QW_{\mathcal{RS}}(x/t; \mathcal{M}(w), \mathcal{M}(w)) = Q\mathcal{M}(w) = w. \tag{2.56}$$

Since $W_{\mathcal{RS}}$ fulfils the integral consistency condition and Q is a linear operator, we have from (2.24) that

$$\begin{aligned} \int_{x_i}^{x_{i+1}} W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}; w_i, w_{i+1}\right) dx = \\ \int_{x_i}^{x_{i+1}} QW_{\mathcal{RS}}\left(\frac{x - x_{i+1/2}}{\Delta t}; \mathcal{M}(w_i), \mathcal{M}(w_{i+1})\right) dx = \\ \frac{1}{2}\Big(Q\mathcal{M}\left(w_i^n\right) + Q\mathcal{M}\left(w_{i+1}^n\right)\Big) - \frac{\Delta t}{\Delta x}\Big(QF\left(\mathcal{M}\big(w_{i+1}^n\big)\right) - QF\big(\mathcal{M}\left(w_i^n\right)\big)\Big). \end{aligned} \tag{2.57}$$

Following the steps in Section 2.1.1, we can write the Godunov type scheme for the original equations as

$$w_i^{n+1} = w_i^n - \frac{\Delta x}{\Delta t}\Big(\mathfrak{f}(w_i^n, w_{i+1}^n) - \mathfrak{f}(w_{i-1}^n, w_i^n)\Big) \tag{2.58}$$

with

$$\begin{aligned} \mathfrak{f}(w_i, w_{i+1}) = \; & QF(\mathcal{M}(w_{i+1})) - \frac{\Delta x}{2\Delta t} Q\mathcal{M}(w_{i+1}) \\ & - \frac{1}{\Delta x} Q \int_{x_{i+1/2}}^{x_{i+1}} W_{\mathcal{RS}}\left(\frac{x - x_{i+1/2}}{\Delta t}; \mathcal{M}(w_i), \mathcal{M}(w_{i+1})\right) dx \\ & = Q\mathcal{F}(\mathcal{M}(w_i), \mathcal{M}(w_{i+1})), \end{aligned}$$

where $\mathcal{F}$ is the numerical flux of the Godunov type scheme (2.46) developed for the relaxation system.

2.1.4 The dimensional split Riemann Problem

So far, we have focused on Godunov's and Godunov type schemes to solve one dimensional conservation laws. To numerically solve the two or three dimensional Euler equations, requires basically only the knowledge about the one dimensional *split multi-dimensional Riemann problem*. The two dimensional Suliciu relaxation model with the velocity field $\mathbf{u} = (u_1, u_2)^T$ and $\mathbf{x} = (x_1, x_2)^T$ is given by

$$\begin{aligned} \partial_t \rho + \partial_{x_1}(\rho u_1) + \partial_{x_2}(\rho u_2) &= 0, \\ \partial_t(\rho u_1) + \partial_{x_1}(\rho u_1^2 + \pi) + \partial_{x_2}(\rho u_1 u_2) &= 0, \\ \partial_t(\rho u_2) + \partial_{x_1}(\rho u_1 u_2) + \partial_{x_2}(\rho u_2^2 + \pi) &= 0, \\ \partial_t E + \partial_{x_1}(u_1(E+\pi)) + \partial_{x_2}(u_2(E+\pi)) &= 0, \\ \partial_t(\rho\pi) + \partial_{x_1}(\rho u_1 \pi + a^2 u_1) + \partial_{x_2}(\rho u_2 \pi + a^2 u_2) &= \frac{\rho}{\varepsilon}(p - \pi), \end{aligned} \tag{2.59}$$

which can be written in compact notation as

$$\partial_t W + \partial_{x_1} F_1(W) + \partial_{x_2} F_2(W) = 0 \tag{2.60}$$

with the state vector W and the flux functions F_1 and F_2 given by

$$W = \begin{pmatrix} \rho \\ \rho u_1 \\ \rho u_2 \\ E \\ \rho\pi \end{pmatrix}, \quad F_1(W) = \begin{pmatrix} \rho u_1 \\ \rho u_1^2 + \pi \\ \rho u_1 u_2 \\ u_1(E+\pi) \\ \rho u_1 \pi + a^2 u_1 \end{pmatrix}, \quad F_2(W) = \begin{pmatrix} \rho u_2 \\ \rho u_2^2 + \pi \\ \rho u_2 u_1 \\ u_2(E+\pi) \\ \rho u_2 \pi + a^2 u_2 \end{pmatrix}. \tag{2.61}$$

The dimensional split Riemann problem in x_1-direction is then given by

$$\begin{aligned} \partial_t W + \partial_{x_1} F_1(W) &= 0 \\ W^0(\mathbf{x}) &= \begin{cases} W_L & \text{if } x_1 < 0, \\ W_R & \text{if } x_1 > 0. \end{cases} \end{aligned} \tag{2.62}$$

The initial data consists of a jump along the x_1-direction while it is constant along the x_2-direction. The wave structure compared to the genuinely one dimensional problem (1.52) is almost identical, whereas the middle contact wave u in the 1D case is now associated to u_1 and has multiplicity three instead of two. All characteristic fields are still linear degenerate and the structure of the Riemann solution is the same as in the truly one dimensional case (2.35). Since the multiplicity of the acoustic waves remains one, we obtain an additional Riemann invariant for the outer waves given by $I_4^{\pm} = u_2^*$. This means that the additional velocity component varies only over the middle wave leading to

$$u_{2L}^* = u_{2L} \quad \text{and} \quad u_{2R}^* = u_{2R}. \tag{2.63}$$

The solution of the intermediate states ρ, u_1, e and π is then given by the one-dimensional Riemann Problem (2.42) with $u = u_1$ and for u_2 by (2.63).

To numerically solve the two dimensional problem (2.60), we first extend the finite volume framework to two dimensions. On a two dimensional Cartesian grid, the x_1- and x_2-

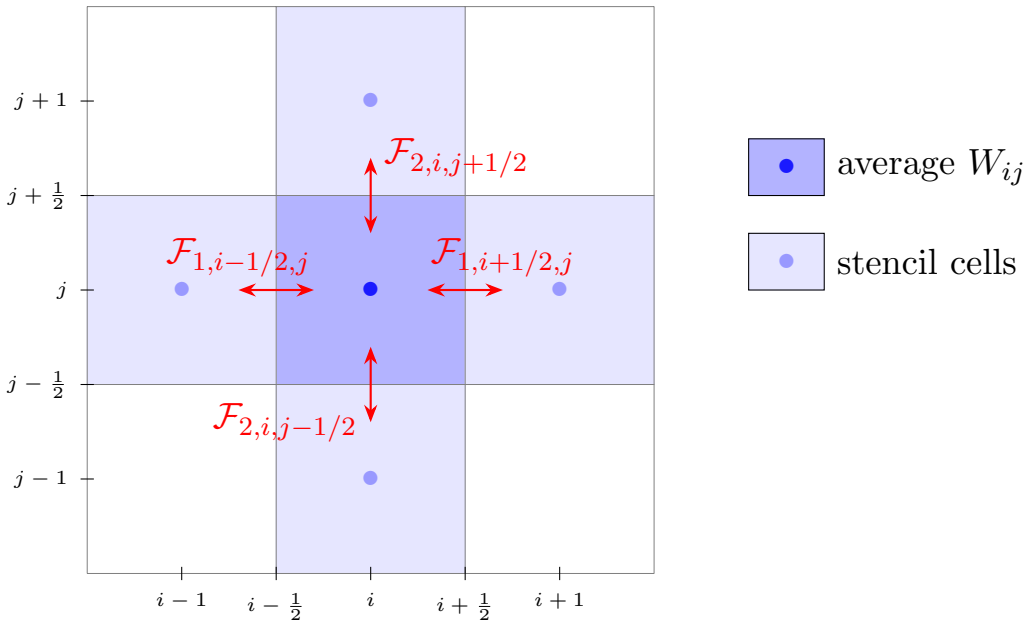


Figure 2.5: Update of the cell averages on a Cartesian grid in two dimensions. The fluxes along the interfaces are obtained by solving dimensional split Riemann problems in x_1-and x_2-direction.

directions are aligned with the coordinate directions, given by the vectors $\mathbf{e}_1 = (1,0)^T$ and $\mathbf{e}_2 = (0,1)^T$. The computational domain is subdivided equidistant along the coordinate directions. Let $C_{ij} = [x_{1,i-1/2}, x_{1,i+1/2}] \times [x_{2,j-1/2}, x_{2,j+1/2}]$ denote a rectangle grid cell of area $\Delta x_1 \Delta x_2$, then the cell average W_{ij}^n at time t^n on cell C_{ij} is given by

$$W_{ij}^n = \frac{1}{\Delta x_1 \Delta x_2} \int_{C_{ij}} W(\mathbf{x}, t^n) d\mathbf{x}. \tag{2.64}$$

An explicit finite volume scheme to solve (2.60), is given by

$$\begin{aligned} W_{ij}^{n+1} = W_{ij}^n &+ \frac{\Delta t}{\Delta x}\Big(\mathcal{F}_1(W_{ij}^n, W_{i+1,j}^n) - \mathcal{F}_1(W_{i-1,j}^n, W_{ij}^n)\Big) \\ &+ \frac{\Delta t}{\Delta y}\Big(\mathcal{F}_2(W_{ij}^n, W_{i,j+1}^n) - \mathcal{F}_2(W_{i,j-1}^n, W_{ij}^n)\Big). \end{aligned} \tag{2.65}$$

The cell averages are updated in a single step involving the flux contributions along all cell interfaces as depicted in Figure 2.5. The numerical fluxes $\mathcal{F}_1$ and $\mathcal{F}_2$ are obtained by solving the dimensional split Riemann problem (2.62) along x_1- and x_2-direction respectively. Each dimensional split Riemann problem consists of a truly one dimensional problem (1.52) with $u = u_i = \langle \mathbf{e}_i, \mathbf{u} \rangle$ and passively advecting the remaining velocity component $u_j = \langle \mathbf{e}_j, \mathbf{u} \rangle, j \neq i$ with u_i for $i, j = 1, 2$.

The time step restriction for scheme (2.65) is given by the more restrictive CFL condition that arises when solving the two dimensional split Riemann problems. A three dimensional problem can be solved analogously on a Cartesian grid by adding a third component in x_3-direction to the numerical scheme (2.65).

Since multi-dimensional problems can be reduced to one dimension via dimensional splitting, we will consider only the 1D case in the remainder of this chapter.

2.2 Second order reconstructions

Godunov's and the Godunov type schemes, developed in the previous sections, are first order accurate. In space, this is mainly connected to the fact, that in the reconstruction step of the REA algorithm, the piecewise polynomial function $\widetilde{W}(x,t^n) = \widetilde{W}^n(x)$ coincides with the constant cell average W_i^n on cell C_i which is a conservative first order reconstruction. To ensure the conservation property of the scheme, the reconstructed piecewise polynomial function $\widetilde{W}^n(x)$ has to be conservative, that is

$$\frac{1}{\Delta x}\int_{C_i}\widetilde{W}^n(x)dx = W_i^n. \tag{2.66}$$

To find a good second order reconstruction, we follow the MUSCL (Monotonic Upstream centred scheme for Conservation laws) approach for simplicity in one dimension. The MUSCL approach can be straightforwardly extended to multiple dimensions using the notion of split dimensional Riemann problems and applying the reconstruction along each direction separately. The stencil is the same as for the first order scheme and is depicted in Figure 2.5.

We construct a linear function with slope σ_i on cell C_i based on the cell averages W_i^n as follows

$$\widetilde{W}^n(x) = W_i^n + \sigma_i(x - x_i) \text{ for } x \in (x_{i-1/2}, x_{i+1/2}) \tag{2.67}$$

which has the conservation property independent of the slopes σ_i. We immediately see that $\widetilde{W}^n(x_i) = W_i^n$. The values for the one sided limits $x \to x_{i-1/2}^+$ and $x \to x_{i+1/2}^-$ are called the inner *interface values* and are given by

$$W_{i-1/2}^+ = W_i^n - \sigma_i\frac{\Delta x}{2} \text{ and } W_{i+1/2}^- = W_i^n + \sigma_i\frac{\Delta x}{2}. \tag{2.68}$$

Since the slopes σ_i are a first order approximation of $\partial_x W(x,t^n)$ of the true solution, we find with a Taylor expansion

$$W_{i-1/2}^+ = W(x_{i-1/2},t^n) + \mathcal{O}(\Delta x^2) \text{ and } W_{i+1/2}^- = W(x_{i+1/2},t^n) + \mathcal{O}(\Delta x^2). \tag{2.69}$$

This means that the interface values obtained by the reconstruction (2.67) are second order accurate. The Riemann problem that has to be solved is now based on the left and right inner interface values

$$\begin{cases} \partial_t W + \partial_x F(W) = 0 \\ W^0(x) = \begin{cases} W_{i+1/2}^- & \text{if } x < x_{i+1/2} \\ W_{i+1/2}^+ & \text{if } x > x_{i+1/2}. \end{cases} \end{cases} \tag{2.70}$$

A Godunov type scheme for a numerical flux $\mathcal{F}$, based on an approximate Riemann solution for (2.70), is then given by

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x}\left(\mathcal{F}\left(W_{i+1/2}^-, W_{i+1/2}^+\right) - \mathcal{F}\left(W_{i-1/2}^-, W_{i-1/2}^+\right)\right). \tag{2.71}$$

The scheme is second order accurate in space and and only first first order accurate in time, see [11]. To completely determine the linear function $\widetilde{W}^n(x)$ we need to find the slopes σ_i. Therefore define the left and right slopes on cell C_i as

$$\sigma_L = \frac{1}{\Delta x}(W_i - W_{i-1}), \quad \sigma_R = \frac{1}{\Delta x}(W_{i+1} - W_i). \tag{2.72}$$

A natural choice would be centred slopes due to symmetry

$$\sigma_i = \frac{\sigma_L + \sigma_R}{2} = \frac{W_{i+1}^n - W_{i-1}^n}{2\Delta x} \tag{2.73}$$

or upwind slopes

$$\sigma_i = \sigma_R = \frac{W_{i+1}^n - W_i^n}{\Delta x}, \quad \sigma_i = \sigma_L = \frac{W_i^n - W_{i-1}^n}{\Delta x} \tag{2.74}$$

which yield a centred three point stencil. Unfortunately, both choices are known to induce spurious oscillations especially when calculating shock solutions due to over or undershoots produced by the reconstruction. Therefore, so called *slope limiters* can be used which have the task to prevent the appearance of those over and undershoots in the numerical solution. Consequently spurious oscillations can be reduced or completely avoided. The slope limiter that is used throughout the manuscript is the *minmod limiter* which is defined as

$$\sigma_i = \text{minmod}(\sigma_L, \sigma_R) = \begin{cases} \min(\sigma_L, \sigma_R) & \text{if } \sigma_L, \sigma_R \geq 0, \\ \max(\sigma_L, \sigma_R) & \text{if } \sigma_L, \sigma_R \leq 0, \\ 0 & \text{else.} \end{cases} \tag{2.75}$$

The minmod function is only one possibility to achieve an explicit oscillation free second order scheme. For other limiting strategies and a theoretical and numerical study comparing different limiters, including the minmod function (2.75) see [76].

Reconstructions using centred slopes (2.73) and minmod slopes (2.75) are compared in Figure 2.6. We see from the definition (2.75) that using minmod slopes preserves the minimum and maximum of cell average values over the computational domain by setting the slope σ_i to zero whenever the sign of the left and right slopes σ_L, σ_R on cell C_i differ. This leads to a reduction of the order near maxima or minima. In contrast, the interface values reconstructed with centred slopes under and overshoot the minimum and maximum of the numerical solution. This becomes problematic when an upper or lower bound of certain variables has be preserved by the scheme.

In context of the Euler equations, this concerns the positivity of density ρ and internal energy e. To ensure $\rho, e > 0$ throughout the simulation, we can profit from the property of the minmod limiter to keep the interface values of both quantities positive. We remark, that the set of physical states (1.18) for the Euler equations (1.16) given by

$$\Omega_{phy}(w) = \left\{ w \in \Omega \subset \mathbb{R}^{d+2}; \rho > 0, e > 0 \right\} \tag{2.76}$$

is a *convex invariant domain*. This means if $W^0(x) \in \Omega_{phy}$ then it follows $W(x,t) \in \Omega_{phy}$ for all $x \in I \subset \mathbb{R},\ t > 0$. A numerical scheme preserves the convex invariant domain Ω_{phy} if from $W_i^n \in \Omega_{phy}$ follows $W_i^{n+1} \in \Omega_{phy}$ for all $i = 1, \ldots, N$ under some CFL condition. For the second order scheme in space (2.71) holds the following result taken from [11].

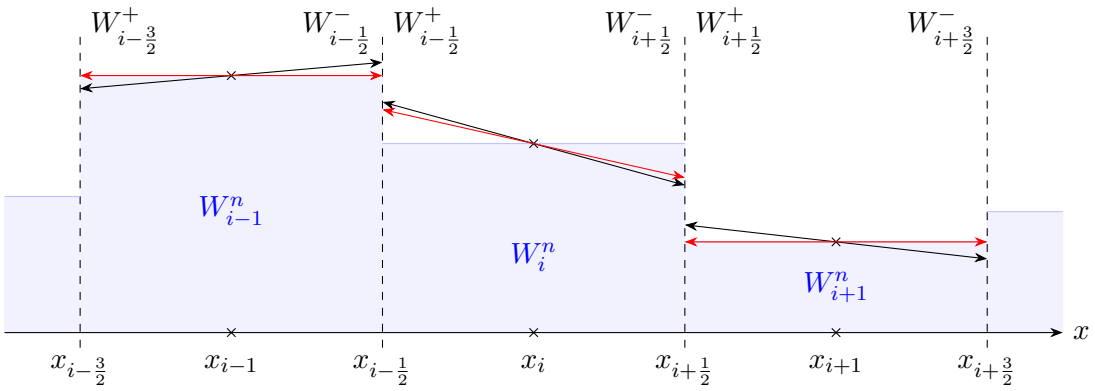


Figure 2.6: Second order reconstruction of interface values, red: minmod slopes, black: centred slopes. The reconstruction with the centred slopes overshoots the maximum value W_{i-1} and undershoots the minimum W_{i+1} whereas the maximum and minimum values are preserved by the reconstruction using a minmod limiter.

Proposition 2.2. *If under a CFL condition the first order scheme* (2.27) *preserves an invariant domain, and if the reconstruction* (2.67) *also preserves this invariant domain, then under the half original CFL condition, the second-order scheme* (2.71) *also preserves this invariant domain.*

Therefore it will be a goal to develop approximate Riemann solvers that preserve the convex invariant domain Ω_{phy}, in the sense that if for two states holds $W_L, W_R \in \Omega_{phy}$, then the approximate Riemann solver ensures

$$W_{\mathcal{R}}(x/t; W_L, W_R) \in \Omega_{phy} \text{ for any } x/t. \tag{2.77}$$

From this property follows directly that the first order scheme (2.27) preserves the invariant domain Ω_{phy} as due to the integral consistency of the approximate Riemann solver, we can write the update

$$W_i^{n+1} = \frac{1}{\Delta x}\left(\int_{x_{i-1/2}}^{x_i} W_{\mathcal{R}}\left(\frac{x - x_{i-1/2}}{\Delta t}, W_{i-1}, W_i\right) dx + \int_{x_i}^{x_{i+1/2}} W_{\mathcal{R}}\left(\frac{x - x_{i-1/2}}{\Delta t}, W_i, W_{i+1}\right) dx\right) \tag{2.78}$$

as a convex combination of states in Ω_{phy} since the CFL condition prevents the interaction of the Riemann solutions.

2.3 Time integration methods

So far, we have focused on the space accuracy of the numerical scheme. To extend the scheme (2.71) to second order accuracy in time, we can use time integration concepts from the theory of ordinary differential equations (ODE). Therefore we integrate the conservation law (2.1) on cell C_i and write it as a system of ODEs in time

$$\partial_t W_i(t) = -\varphi_i(W(t)) \tag{2.79}$$

with $W_i(t) \approx \frac{1}{\Delta x}\int_{C_i} W(x,t)dx$ and $\varphi_i(W(t)) \approx \frac{1}{\Delta x}\int_{\partial C_i} F(W(x,t))dx$. Depending on the properties of the right hand side function φ_i, which contains the flux function, explicit, implicit or implicit-explicit (IMEX) numerical time integrators can be applied.

A popular choice to solve (2.79) with a non-stiff right hand side are explicit Runge Kutta (RK) methods since they are available for high orders and easily applicable. For a second order scheme, the Heun method is widely used. A feature of this method is that it preserves invariant domains without any further reduction of the CFL condition associated to the second order scheme in space (2.71) provided that the Riemann solver fulfils property (2.77). It belongs to the class of predictor corrector methods. To evolve data from t^n to $t^n + \Delta t$ the solution is predicted using a forward Euler scheme (2.6) and then corrected using the trapezoidal rule. In total it is given by

$$\begin{aligned} W_i^{(1)} &= W_i^n - \Delta t \varphi_i\left(W^n\right), \\ W_i^{(2)} &= W_i^{(1)} - \Delta t \varphi_i\left(W^{(1)}\right), \\ W_i^{n+1} &= \frac{1}{2}\left(W_i^n + W_i^{(2)}\right). \end{aligned} \tag{2.80}$$

The time step Δt is determined by the CFL condition of the first Euler step when computing $W_i^{(1)}$. When computing the second stage $W^{(2)}$, the CFL condition is not necessarily fulfilled. A modification of the Heun method which allows different time increments for each Euler step was proposed in [6]. It is also second order accurate and given by

$$\begin{aligned} W_i^{(1)} &= W_i^n - \Delta t_1 \varphi_i\left(W^n\right), \\ W_i^{(2)} &= W_i^{(1)} - \Delta t_2 \varphi_i\left(W^{(1)}\right), \\ W_i^{n+1} &= \left(1 - \frac{2\Delta t_1 \Delta t_2}{(\Delta t_1 + \Delta t_2)^2}\right) W_i^n + \frac{2\Delta t_1 \Delta t_2}{(\Delta t_1 + \Delta t_2)^2} W_i^{(2)}. \end{aligned} \tag{2.81}$$

The time steps $\Delta t_1, \Delta t_2$ are chosen in accordance with the CFL condition of the first and second step respectively. For fixed $\Delta t_1 = \Delta t_2 = \Delta t$, we recover Heun's method (2.80). The final time increment is given by

$$\Delta t = \frac{2\Delta t_1 \Delta t_2}{\Delta t_1 + \Delta t_2}. \tag{2.82}$$

Since the final update of (2.81) is a convex combination of the states $W_i^n \in \Omega_{phy}$ and $W_i^{(2)}$, which is obtained by subsequent first order updates that lie in Ω_{phy}, also this scheme preserves invariant domains.

As we have seen, when deriving Godunov's scheme, the CFL condition for an explicit scheme depends on the fastest absolute value of the wave speeds arising in the Riemann fan. Explicit schemes become very costly when due to stability requirements the time step is strongly restricted. We illustrate this problem of stiffness at the example of the Euler equations in the low Mach regime. As we have seen in Section 1.4, the fastest wave speed is given by $u + \frac{c}{M}$ and therefore the time restriction for a first order scheme is given by

$$\frac{\Delta t}{\Delta x} \leq \tilde{c}\, \frac{M}{\max_i |M u_i^n + c_i^n|} \tag{2.83}$$

where $\tilde{c}$ is a positive constant. Especially in the low Mach regime the CFL condition (2.83) is problematic since $\Delta t \to 0$ in the limit $M \to 0$. An option to circumvent the restrictive CFL condition (2.83) is the use of implicit schemes. They have the advantage that the time step can be chosen unconditionally large. Keeping in mind the non-linear character of the Euler equations, using an implicit scheme means solving a non-linear implicit system each time step. This requires iterative methods like the Newton algorithm which is very costly since each step the Jacobian of the residual and large matrix vector products have to be calculated. In addition, to ensure a good quality of the numerical solution, the time step cannot be chosen arbitrarily large but has to be oriented at the focus of the numerical simulation. In case of the low Mach regime of the Euler equations, we are not interested in the fast acoustic waves, that is the spreading of the sound, but rather in the slower dynamics that evolve with the wave speed u. Therefore a CFL restriction

$$\frac{\Delta t}{\Delta x} \leq \frac{\tilde{c}}{\max_i |u_i^n|} \tag{2.84}$$

would be sufficient to follow the slow wave. This can be achieved by using an IMEX approach where $\varphi_i(t)$ in the ODE (2.79) is split in two parts φ_i^{ex} and φ_i^{im} which will be integrated explicitly or implicitly in time respectively:

$$\partial_t W_i(t) = -\varphi_i^{ex}(W(t)) - \varphi_i^{im}(W(t)). \tag{2.85}$$

A first order scheme consisting of a forward and a backward Euler step is given by

$$W_i^{n+1} = W_i^n - \Delta t \varphi_i^{ex}(W^n) - \Delta t \varphi_i^{im}(W^{n+1}). \tag{2.86}$$

The CFL condition associated to the IMEX scheme (2.86) is based on the fastest wave speeds associated to φ_i^{ex}. To obtain a CFL condition independent of the Mach number in case of the Euler equations, the terms associated to the fast acoustic waves have to be treated implicitly. The main challenge consists in identifying those terms and avoiding to solve a non-linear implicit system at the same time.

The all-speed IMEX schemes given in Chapters 4 and 5 are developed under the aspect of being computationally efficient with respect to allowing large time steps in the low Mach number regime. Non-linear implicit systems can be avoided due to the linearisation of the characteristic fields by using a Suliciu relaxation approach. An extension of the IMEX scheme to second order will be discussed in the Chapters 4 and 5 since it is not based on a standard IMEX Runge Kutta method.

2.4 Source terms

We consider a one dimensional initial value problem for a balance law of the following form

$$\begin{aligned} &\partial_t W(x,t) + \partial_x F(W(x,t)) = S(W(x,t))\partial_x Z(x), \\ &\partial_t Z(x) = 0, \\ &W^0(x) = W(x,0), \end{aligned} \tag{2.87}$$

where $Z : \mathbb{R} \to \mathbb{R}$ is a given smooth scalar time independent function and $S(W) : \Omega \to \mathbb{R}^k$. We remark, that the extra equation for the time derivative of Z is not important for the formulation of the continuous equations, but is essential when analysing the solution to a Riemann problem for (2.87). The IVP problem for the one dimensional Euler equations with a gravitational source term (1.75) can be written in the form of (2.87). The balance law (2.87) is not conservative in the presence of a non-zero source term and thus a numerical scheme approximating solutions of (2.87) is not in conservation form.

2.4.1 Godunov type scheme with a source term

To construct a Godunov type scheme for (2.87), we define the associated Riemann problem at the interface $x_{i+1/2}$ by

$$\partial_t W + \partial_x F(W) = S(W)\partial_x Z, \tag{2.88}$$

$$W^0(x) = \begin{cases} W_i & \text{if } x < x_{i+1/2}, \\ W_{i+1} & \text{if } x > x_{i+1/2}, \end{cases}. \tag{2.89}$$

The function Z is approximated in the finite volume framework as piecewise constant

$$Z_i = \frac{1}{\Delta x}\int_{C_i} Z(x)dx \quad \text{for} \quad x \in (x_{i-1/2}, x_{i+1/2}), \quad \text{for all } i = 1, \dots, N. \tag{2.90}$$

An exact solution of the inhomogeneous Riemann problem (2.88) is in general very challenging. Therefore we turn directly to finding an approximative Riemann solver $W_{\mathcal{R}}$. Due to the presence of the source term, $W_{\mathcal{R}}$ also depends on the values of Z. To emphasize this, we write

$$W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{t - t^n}; W_i, Z_i, W_{i+1}, Z_{i+1}\right). \tag{2.91}$$

For completeness, we state again the definition of an approximate Riemann solver accounting for its dependence on Z. First, it has to verify

$$W_{\mathcal{R}}(x/t; W, Z, W, Z) = W \text{ for all } W \in \Omega \text{ and } x/t \in \mathbb{R}. \tag{2.92}$$

It has to fulfil the integral consistency condition with respect to an exact solution $\widetilde{W}$ of the local Riemann problem (2.88)

$$\int_{x_i}^{x_{i+1}} W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}; W_i, Z_i, W_{i+1}, Z_{i+1}\right) dx = \int_{x_i}^{x_{i+1}} \widetilde{W}\left(\frac{x - x_{i+1/2}}{\Delta t}; W_i, Z_i, W_{i+1}, Z_{i+1}\right) dx. \tag{2.93}$$

Analogously to the integral consistency for the conservation law in Section 2.1.2, we can reformulate it by integrating (2.88) on $[x_i, x_{i+1}] \times (t^n, t^{n+1}]$ to obtain

$$\begin{aligned} \frac{1}{\Delta x}\int_{x_i}^{x_{i+1}} W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}; W_i, Z_i, W_{i+1}, Z_{i+1}\right) dx = \\ \frac{1}{2}(W_i^n + W_{i+1}^n) - \frac{\Delta t}{\Delta x}(F(W_{i+1}^n) - F(W_i^n)) \\ + \frac{1}{\Delta x}\int_{x_i}^{x_{i+1}}\int_{t^n}^{t^{n+1}} S\left(\widetilde{W}\left(\frac{x - x_{i+1/2}}{t - t^n}; W_i, Z_i, W_{i+1}, Z_{i+1}\right)\right)\partial_x Z(x) dt dx. \end{aligned} \tag{2.94}$$

In general it is not possible to calculate the exact integral expression of the source term. Therefore, we define the first order numerical approximation

$$\mathcal{S}(W_i, W_{i+1})\frac{Z_{i+1}-Z_i}{\Delta x} \approx \\ \frac{1}{\Delta x \Delta t}\int_{x_i}^{x_{i+1}}\int_{t^n}^{t^{n+1}} S\left(\widetilde{W}\left(\frac{x-x_{i+1/2}}{t-t^n}; W_i, Z_i, W_{i+1}, Z_{i+1}\right)\right)\partial_x Z(x) dt dx, \quad (2.95)$$

where $\mathcal{S}$ denotes the numerical source which has to fulfil the consistency condition

$$\mathcal{S}(W, W) = S(W) \text{ for all } W \in \Omega. \quad (2.96)$$

Turning back to the REA algorithm, we consider a piecewise constant initial data

$$\widetilde{W}_\Delta(x, t^n) = W_i^n \text{ for } x \in C_i. \quad (2.97)$$

Then the solution at time t^{n+1} is piecewise given by approximate Riemann solutions

$$\widetilde{W}_\Delta(x, t^{n+1}) = W_{\mathcal{R}}\left(\frac{x-x_{i+1/2}}{\Delta t}; W_i, Z_i, W_{i+1}, Z_{i+1}\right) \text{ for } x \in (x_i, x_{i+1}), \quad (2.98)$$

under the CFL condition

$$\frac{\Delta t}{\Delta x}\max_{i,\kappa}|\lambda_\kappa(W_i^n, Z_i, W_{i+1}^n, Z_{i+1})| \le \frac{1}{2}, \quad i = 1, \dots, N, \quad \kappa = 1, \dots, k. \quad (2.99)$$

The update at time t^{n+1} on cell C_i is then defined as

$$\begin{aligned} W_i^{n+1} &= \frac{1}{\Delta x}\int_{C_i}\widetilde{W}_\Delta(x, t^{n+1})dx \\ &= \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_i} W_{\mathcal{R}}\left(\frac{x-x_{i-1/2}}{\Delta t}, W_{i-1}, Z_{i-1}, W_i, Z_i\right)dx \\ &\qquad + \frac{1}{\Delta x}\int_{x_i}^{x_{i+1/2}} W_{\mathcal{R}}\left(\frac{x-x_{i+1/2}}{\Delta t}, W_i, Z_i, W_{i+1}, Z_{i+1}\right)dx. \end{aligned} \quad (2.100)$$

Defining the right numerical flux as

$$\mathcal{F}(W_i, Z_i, W_{i+1}, Z_{i+1}) = \\ F(W_{i+1}) - \frac{\Delta x}{2\Delta t}\left(W_{i+1} - \frac{2}{\Delta x}\int_{x_{i+1/2}}^{x_{i+1}} W_{\mathcal{R}}\left(\frac{x-x_{i+1/2}}{\Delta t}; W_i, Z_i, W_{i+1}, Z_{i+1}\right)dx\right), \quad (2.101)$$

we can formulate a numerical scheme consistent with the balance law (2.87) as

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x}\left(\mathcal{F}(W_i, Z_i, W_{i+1}, Z_{i+1}) - \mathcal{F}(W_{i-1}, Z_{i-1}, W_i, Z_i)\right) \\ + \Delta t \mathcal{S}(W_i, W_{i+1})\frac{Z_{i+1}-Z_i}{\Delta x}. \quad (2.102)$$

Defining analogously the left numerical flux as

$$\overline{\mathcal{F}}(W_i, Z_i, W_{i+1}, Z_{i+1}) = \\ F(W_i) + \frac{\Delta x}{2\Delta t}\left(W_i - \frac{2}{\Delta x}\int_{x_i}^{x_{i+1/2}} W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}; W_i, Z_i, W_{i+1}, Z_{i+1}\right) dx\right). \quad (2.103)$$

we can define the numerical scheme as

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x}\Big(\overline{\mathcal{F}}(W_i, Z_i, W_{i+1}, Z_{i+1}) - \overline{\mathcal{F}}(W_{i-1}, Z_{i-1}, W_i, Z_i)\Big) \\ + \Delta t \mathcal{S}(W_{i-1}, W_i)\frac{Z_i - Z_{i-1}}{\Delta x}. \quad (2.104)$$

Both numerical fluxes, $\mathcal{F}, \overline{\mathcal{F}}$ are consistent because of the consistency property of the approximate Riemann solver. The schemes (2.102) and (2.104) are equivalent since the approximate Riemann solver fulfils the integral consistency condition (2.93) and are a consistent discretization of the IVP (2.87). Due to the dependence of the Riemann problem on the source term, the numerical fluxes don't fulfil the conservativity identity (2.29) which is not surprising since the underlying equations are not conservative.

2.4.2 Discrete stationary solutions and well-balanced schemes

Stationary states of the balance law (2.87) are given by $\partial_x F(W) = S(W)\partial_x Z$. For the numerical scheme (2.102) we can define *discrete stationary states*. Let two states $(W_i, Z_i), (W_{i+1}, Z_{i+1})$ fulfil the discrete equilibrium

$$\mathcal{F}(W_i, Z_i, W_{i+1}, Z_{i+1}) - \mathcal{F}(W_{i-1}, Z_{i-1}, W_i, Z_i) = \mathcal{S}(W_i, W_{i+1})(Z_{i+1} - Z_i) \quad (2.105)$$

for all $i = 1, \dots N$, then the states W_i define a *discrete piecewise constant stationary solution*.

A numerical scheme, that is able to fulfil (2.105) exactly for a discrete stationary solution is called *well-balanced*. This means the scheme fulfils

$$W_i^{n+1} = W_i^n \text{ for all } i = 1, \dots, N. \quad (2.106)$$

In general, numerical schemes are not well-balanced. For a Godunov type scheme, the well-balanced property is closely connected to the approximative Riemann solver. Let $W_{\mathcal{R}}(x/t; W_i, Z_i, W_{i+1}, Z_{i+1})$ be an approximate Riemann solver based on the two states (W_i, Z_i), and (W_{i+1}, Z_{i+1}) fulfilling the discrete equilibrium (2.105). Then it is called well-balanced or *at rest* if it satisfies

$$W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{t - t^n}; W_i, Z_i, W_{i+1}, Z_{i+1}\right) = \begin{cases} W_i & \text{if} \quad \dfrac{x - x_{i+1/2}}{t - t^n} < 0, \\ W_{i+1} & \text{if} \quad \dfrac{x - x_{i+1/2}}{t - t^n} > 0. \end{cases} \quad (2.107)$$

With this definition follows that a Godunov type scheme is well-balanced if the approximate Riemann solver is at rest. This follows directly from the update (2.100) where due

to the well-balanced property of the approximate Riemann solver holds

$$\begin{aligned} W_{\mathcal{R}}\left(\frac{x - x_{i-1/2}}{\Delta t}, W_{i-1}, Z_{i-1}, W_i, Z_i\right) &= W_i \text{ for all } x \in (x_{i-1/2}, x_i) \\ W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}, W_i, Z_i, W_{i+1}, Z_{i+1}\right) &= W_i \text{ for all } x \in (x_i, x_{i+1/2}). \end{aligned} \tag{2.108}$$

From this we can conclude $W_i^{n+1} = W_i^n$ and therefore the Godunov type scheme is well-balanced.

The objective of Chapter 3 and 5 is to define a well-balanced approximate Riemann solver for compressible and all-speed flow respectively. Therein also the extension to second order is addressed. The Godunov type scheme with source term can be straightforwardly extended to multiple dimensions following the approach of split dimensional Riemann problems from Section 2.1.4.

Chapter 3

A well-balanced scheme for the Euler equations with gravity

The compressible Euler equations with a gravitational source term (1.75) are important in many applications, be it in atmospheric modelling or in astrophysical stellar evolution. It is typical for those applications that solutions are not far from a hydrostatic equilibrium state. Therefore it is necessary to have a numerical method that captures those hydrostatic equilibria on machine precision in order to resolve the evolution of small perturbations near the equilibrium even on a coarse grid. This can be achieved with well-balanced methods.

In literature a variety of approaches to develop those methods can be found. A wide range of schemes are based on finite volume methods [44, 45], central schemes [82] or discontinuous Galerkin approaches [55, 56, 18], see also references therein. Challenging is also the treatment of the source term. Since in Riemann solver based schemes, as the Godunov type scheme, the intermediate states depend on the source term, it can be included in the flux formulation, as done in [28, 20]. More often the source term is discretised separately from the numerical flux as done eg. in [18, 84, 44, 45].

Common to a lot of well-balanced schemes is that they are algebraically consistent with a certain class of equilibria, for example hydrostatic equilibria with constant entropy [44] or isothermal and polytropic equilibria [28, 17, 18]. An interesting approach is found in Käppeli [45], where a discrete approximation of the hydrostatic equilibrium is used to achieve a second order accuracy to general equilibria. Depending on the distribution of the hydrostatic equilibria this second order approximation might still give reasonable results with an error close to machine precision, whereas in general the numerical error remains of second order.

In contrast, here we present a method where a given arbitrary hydrostatic equilibrium is well-balanced by reformulating the relevant hydrostatic background. This hydrostatic background is not restricted to be of a certain class of hydrostatic equilibria. We give a general recipe on how to use this reformulation in terms of a reference equilibrium state in the scheme to obtain a generality in application. However we would like to stress that this reference state can be chosen by the user to fit the practical purpose.

Furthermore we show that by choosing a reconstruction procedure in so called equilibrium variables leads to a second order scheme that is well-balanced as well as preserves the invariant domain Ω_{phy} with respect to the positivity of density and internal energy.

The results presented in this chapter are already published in [78]. The chapter is organized as follows. Section 3.1 is devoted to the parametrisation of the source term with respect to an arbitrary reference equilibrium. We will utilize the time-independent nature of stationary solutions to rewrite the derivative of the potential in terms of time-independent functions that describe the steady-state solution. In Section 3.2, we describe the relaxation model that is used to derive an approximate Riemann solver. Subsequent we give the description of the numerical scheme which is of Godunov type as discussed in Section 2.4. This includes an extension to second order with special focus on a positivity preserving linear reconstruction of the interface values.

For the resulting scheme, the main properties, which are the positivity and the well-balanced property of the scheme, are proven in Section 3.5. It is followed by a section with numerical results to validate the main properties given in Section 3.5. Therein are included well-balanced test cases for isothermal, polytropic and a general stationary equilibrium. To show the accuracy of the second order scheme, we consider two analytical solutions of the Euler equations with gravity (1.75). The first one is a two dimensional solution from [18] and we also give a novel analytic solution based on an isothermal equation of state which we test in three dimensions. Additionally, the convergence of the source term is shown by balancing a general non-isothermal stationary state against an isothermal equilibrium. To demonstrate that the scheme is indeed applicable to simulate fluctuations around an equilibrium state, we calculate a Rayleigh-Taylor instability taken from [53] in two dimensions. The numerical section is concluded by a modification of the Einfeldt strong rarefaction test [32] adopted to the presence of a gravitational field.

3.1 The reformulation of the gravitational field

As we have seen in Section (1.5), hydrostatic steady states for the Euler equations with gravity (1.75) are solutions of the hydrostatic equilibrium described by the equations

$$\begin{aligned} \mathbf{u} &= 0, \\ \nabla p &= -\rho \, \nabla \Phi. \end{aligned} \tag{3.1}$$

Since the system is under-determined depending on the chosen pressure law, the solutions to (3.1) might have quite a different structure and physical behaviour. This can be already seen from the isothermal (1.90) and polytropic (1.91) examples given in Section (1.5). In practise one might not want to change the numerical scheme in dependence of the EOS under consideration. In addition, for most applications the underlying hydrostatic equilibrium is known in advance and can be computed from the EOS describing the physical regime. Therefore we seek for a reformulation of the gravitational potential in terms of the hydrostatic equilibrium relation (3.1) that can be easily incorporated in the system of equations (1.75). Let $\bar{\rho}$ and $\bar{p}$ be solutions of the hydrostatic equilibrium (3.1). Since they are stationary, they are time-independent. Following the approach used in [36], we choose two time-independent functions α and β such that they coincide with the given hydrostatic solution described by $\bar{\rho}$ and $\bar{p}$ as

$$\bar{\rho}(\mathbf{x}) = \alpha(\mathbf{x}) \quad \text{and} \quad \bar{p}(\mathbf{x}) = \beta(\mathbf{x}). \tag{3.2}$$

Since the density and the pressure are strictly positive, we also require $\alpha, \beta > 0$. By construction α and β fulfil the hydrostatic equation and we can write

$$\nabla\beta = -\alpha\nabla\Phi. \tag{3.3}$$

This relation describes a connection between the reference equilibrium defined by α, β and the gravitational potential Φ. Rewriting (3.3), leads to the following expression of the gravitational potential

$$\nabla\Phi(\mathbf{x}) = -\frac{\nabla\beta(\mathbf{x})}{\alpha(\mathbf{x})}. \tag{3.4}$$

Now we can replace the potential gradient in the Euler equations (1.75) and in the hydrostatic equation (3.1). We will consider the following formulation of the Euler equations with gravity

$$\begin{aligned} \partial_t\rho + \nabla\cdot(\rho\mathbf{u}) &= 0, \\ \partial_t(\rho\mathbf{u}) + \nabla\cdot(\rho\mathbf{u}\otimes\mathbf{u} + p\mathbb{I}) &= \frac{\rho}{\alpha}\nabla\beta, \\ \partial_t E + \nabla\cdot(\mathbf{u}(E+p)) &= \mathbf{u}\cdot\frac{\rho}{\alpha}\nabla\beta, \end{aligned} \tag{3.5}$$

which exhibits the following formulation of the hydrostatic equation

$$\nabla p = \frac{\rho}{\alpha}\nabla\beta. \tag{3.6}$$

We would like to remark, that both formulations of the Euler equations with gravity given by (1.75) and (3.5) are equivalent, as long as α, β are smooth functions. The approach (3.2) to rewrite the potential strongly relies on the fact that the considered gravitational field is assumed to be time-independent during the evolution of gas given by the Euler equations (3.5).

Relation (3.4) suggests that $\nabla\Phi$ is a function of α and β. However in practise it is found that rather α and β are depending on Φ as it is initially known. Using the definition of α, β in (3.2) we can give the reference equilibrium states for the isentropic and polytropic equilibria given in (1.90) and (1.91) respectively as

$$\alpha(\mathbf{x}) = \exp\left(\frac{C-\Phi(\mathbf{x})}{RT}\right), \quad \beta(\mathbf{x}) = RT\alpha(\mathbf{x}), \tag{3.7}$$

and

$$\alpha(\mathbf{x}) = \left(\frac{\Gamma-1}{\chi^\Gamma}\left(C-\Phi(\mathbf{x})\right)\right)^{1/(\Gamma-1)}, \quad \beta(\mathbf{x}) = \chi\alpha^\Gamma. \tag{3.8}$$

As we have seen in Section 1.5, associated to the time-independent potential Φ is a zero eigenvalue. Due to the rewritten potential gradient (3.4) this zero eigenvalue is now associated to β. This can be seen when analysing the eigenstructure of the quasi-linear formulation in the primitive variables given by $\mathcal{V} = (\rho, \mathbf{u}, e, \beta)$. This zero eigenvalue is problematic when deriving a Riemann solver, since the waves do not have a fixed order. This problem can be resolved, when modifying the standard Suliciu relaxation model given in Section 1.3.2. How it is done exactly is subject of the next section.

3.2 The Suliciu relaxation model

The overall aim of this chapter is to derive a Godunov type scheme based on an approximate Riemann solver. The Suliciu relaxation model that is used to construct the approximate Riemann solver for the Euler equations (3.5) with the modified source term is based on an approach from [28] that was applied on the original system (1.75). Therein was suggested to approximate the potential Φ by a new variable Z which is transported with the gas velocity $\mathbf{u}$. This yields an additional relaxation equation, apart from the Suliciu relaxation of the pressure which can be derived analogously as in the homogeneous case (1.52). Here Z approximates β instead of Φ, since the zero eigenvalue is associated to β due to the rewritten source term. The full relaxation model then reads

$$\begin{aligned}
\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) &= 0, \\
\partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u} + \pi \mathbb{I}) &= \frac{\rho}{\alpha} \nabla Z, \\
\partial_t E + \nabla \cdot (\mathbf{u}(E + \pi)) &= \mathbf{u} \cdot \frac{\rho}{\alpha} \nabla Z, \\
\partial_t (\rho \pi) + \nabla \cdot (\rho \pi \mathbf{u} + a^2 \mathbf{u}) &= \frac{\rho}{\epsilon}(p - \pi), \\
\partial_t (\rho Z) + \nabla \cdot (\rho Z \mathbf{u}) &= \frac{\rho}{\epsilon}(\beta - Z).
\end{aligned} \tag{3.9}$$

The constant $a > 0$ denotes the relaxation parameter which has to fulfil a subcharacteristic condition that is specified later on in Lemma 3.1. To construct a scheme for multiple space dimensions on a Cartesian grid, we consider dimensional split Riemann problems. Therefore, we analyse the relaxation model in two dimensions with $\mathbf{u} = (u_1, u_2)$ considering the x_1-direction. The equations read

$$\begin{aligned}
\partial_t \rho + \partial_{x_1} (\rho u_1) &= 0, \\
\partial_t (\rho u_1) + \partial_{x_1} (\rho u_1^2 + \pi) &= \frac{\rho}{\alpha} \partial_{x_1} Z, \\
\partial_t (\rho u_2) + \partial_{x_1} (\rho u_1 u_2) &= 0 \\
\partial_t E + \partial_{x_1} (u_1 (E + \pi)) &= u_1 \frac{\rho}{\alpha} \partial_{x_1} Z \\
\partial_t (\rho \pi) + \partial_{x_1} (\rho \pi u_1 + a^2 \pi) &= \frac{\rho}{\varepsilon}(p - \pi) \\
\partial_t (\rho Z) + \partial_{x_1} (\rho Z u_1) &= \frac{\rho}{\varepsilon}(\beta - Z).
\end{aligned} \tag{3.10}$$

The following lemma sums up some important properties regarding the structure and stability of the relaxation model (3.10). The proof can be obtained by adapting the steps explained in Sections 1.2 and 1.3.2. Alternatively, the proof can be established analogously to [28].

Lemma 3.1. *The relaxation system (3.10) is hyperbolic and is a stable diffusive approximation of (3.5) under a subcharacteristic condition $a \geq \rho c$. It has the following ordered eigenvalues*

$$\lambda^- = u_1 - \frac{a}{\rho} < \lambda^u = u_1 < \lambda^+ = u_1 + \frac{a}{\rho} \tag{3.11}$$

where λ^u has multiplicity 4. In addition, the characteristic fields associated to the eigenvalues $\lambda^\pm, \lambda^u$ are linear-degenerate.

To shorten notation, we will refer to the relaxation model (3.10) by

$$\partial_t W + \partial_{x_1} F_1(W) = S_1(W) + \frac{1}{\varepsilon} R(W) \tag{3.12}$$

where

$$W = \begin{pmatrix} \rho \\ \rho u_1 \\ \rho u_2 \\ E \\ \rho\pi \\ \rho Z \end{pmatrix}, \; F_1(W) = \begin{pmatrix} \rho u_1 \\ \rho u_1^2 + \pi \\ \rho u_1 u_2 \\ u_1(E+\pi) \\ \rho\pi u_1 + a^2\pi \\ \rho Z u_1 \end{pmatrix}, \; S_1(W) = \begin{pmatrix} 0 \\ \frac{\rho}{\alpha}\partial_x Z \\ 0 \\ u_1 \frac{\rho}{\alpha}\partial_x Z \\ 0 \\ 0 \end{pmatrix}, \; R(W) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \rho(p-\pi) \\ \rho(\beta - Z) \end{pmatrix}. \tag{3.13}$$

The corresponding original equations to the dimensional split relaxation model (3.10) is referred to as

$$\partial_t w + \partial_{x_1} f_1(w) = s_1(w) \tag{3.14}$$

where $w = (\rho, \rho u_1, \rho u_2, E)^T$ and

$$f_1(w) = \begin{pmatrix} \rho u_1 \\ \rho u_1^2 + \pi \\ \rho u_1 u_2 \\ u_1(E+\pi) \end{pmatrix}, \quad s_1(w) = \begin{pmatrix} 0 \\ \frac{\rho}{\alpha}\partial_x \beta \\ 0 \\ u_1 \frac{\rho}{\alpha}\partial_x \beta \end{pmatrix}. \tag{3.15}$$

The relaxation equilibrium of the relaxation model is given by

$$W^{\text{eq}} = \mathcal{M}(w) = (\rho, \rho u_1, \rho u_2, E, \rho p, \rho\beta)^T. \tag{3.16}$$

The linear operator which connects the relaxation model to the original system is given by

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \tag{3.17}$$

We can easily see that $Q\mathcal{M}(w) = w$ and $QF_1(\mathcal{M}(w)) = f_1(w)$ as well as $QS_1(\mathcal{M}(w)) = s_1(w)$.

3.3 The approximate Riemann solver

Since we are interested in the $\varepsilon \to 0$ limit of the relaxation model, the relaxation source term $\frac{1}{\varepsilon}R(W)$ is stiff. Therefore, we split the operators as follows

$$\partial_t W + \partial_{x_1} F_1(W) = S_1(W), \tag{3.18}$$

$$\partial_t W = \frac{1}{\varepsilon} R(W). \tag{3.19}$$

System (3.18) is then used to define the numerical flux at the interfaces via solving the arising Riemann problems. Afterwards in a separate step, system (3.19) is used to ensure

that the initial conditions at each stage for (3.18) are at relaxation equilibrium W^{eq} [11]. Due to the stiffness, the relaxation source term is treated implicitly for a time step Δt and then the limit $\varepsilon \to 0$ is taken. Following the steps in (2.32) and (2.33) we find directly the update at the new time step $t^n + \Delta t$ as

$$\pi^{n+1} = p^{n+1}, \quad Z^{n+1} = \beta. \tag{3.20}$$

As established in Lemma 3.1, all characteristic fields are linear degenerate and this enables us to solve the Riemann problem of (3.18) without much effort. The basis for this is laid by the Riemann invariants associated to the linear-degenerate fields given in the following lemma.

Lemma 3.2. *The Riemann invariants with respect to $\lambda^{\pm}$ are*

$$I_1^{\pm} = u_1 \pm \frac{a}{\rho}, \quad I_2^{\pm} = \pi \mp a u_1, \quad I_3^{\pm} = e - \frac{\pi^2}{2a^2}, \quad I_4^{\pm} = Z, \quad I_5^{\pm} = u_2 \tag{3.21}$$

and with respect to λ^u

$$I_1^u = u_1, \quad I_2^u = \pi - \frac{\rho}{\alpha} Z. \tag{3.22}$$

Proof. The proof consists in calculating the eigenvectors to the respective eigenvalues and verifying condition (1.30). The straightforward computations are for example given in [28, 52] and can be easily adapted to the equations considered here. □

From the Riemann invariants (3.21) and (3.22), we obtain 12 relations for in total 12 unknown intermediate values, as the solution consists of four constant states separated by the eigenvalues given in Lemma 3.1 and which is depicted in Figure 3.1. The fact, that the eigenvalues are ordered, simplifies the construction of the Riemann solver, as only one wave configuration has to be considered for the determination of the intermediate states. For the problematic of constructing a Riemann solver for the Euler equations with gravity with a changing order of eigenvalues see [85], where 6 different wave configurations have to be considered.

To simplify the solution and to obtain a free parameter that will later ensure the well-balanced property, we keep $\frac{\rho}{\alpha} =: \hat{\kappa}(\rho, \alpha)$ fixed in the Riemann problem. This means the second Riemann invariant for $\lambda = u_1$ is approximated by $I_2^u = \pi - \hat{\kappa} Z$, $\hat{\kappa} > 0$. The discretization of $\hat{\kappa}$, appearing in the intermediate states of the Riemann solution and denoted by κ, will be later defined such that the Riemann solver is at rest when the solution is in hydrostatic equilibrium. This strategy has already been applied successfully in [28, 29]. Now with the help of Lemma 3.2 we can construct a solution to a Riemann problem in dependence of κ for system (3.18).

Lemma 3.3. *Consider an initial value problem for system (3.18) with initial data*

$$W^0(\mathbf{x}) = \begin{cases} W_L & x_1 < 0, \\ W_R & x_1 > 0, \end{cases} \tag{3.23}$$

which is constant along the x_2-direction. Then the solution consists of four constant states separated by contact discontinuities and has the following structure for $t > 0$

$$W_{\mathcal{RS}}\left(\frac{x_1}{t}; W_L, W_R\right) = \begin{cases} W_L & \frac{x_1}{t} < \lambda^- \\ W_L^* & \lambda^- < \frac{x_1}{t} < \lambda^u \\ W_R^* & \lambda^u < \frac{x_1}{t} < \lambda^+ \\ W_R & \lambda^+ < \frac{x_1}{t} \end{cases}. \tag{3.24}$$

Furthermore a solution for the intermediate states $W^*_{L/R}$ is given by

$$u_1^* = \frac{1}{2}(u_{1,L} + u_{1R}) - \frac{1}{2a}(\pi_R - \pi_L - \kappa(Z_R - Z_L)), \tag{3.25}$$

$$\pi_L^* = \pi_L + a(u_{1,L} - u_1^*), \qquad \pi_R^* = \pi_R + a(u_1^* - u_{1,R}) \tag{3.26}$$

$$\frac{1}{\rho_L^*} = \frac{1}{\rho_L} - \frac{1}{a}(u_{1,L} - u_1^*), \qquad \frac{1}{\rho_R^*} = \frac{1}{\rho_R} - \frac{1}{a}(u_1^* - u_{1,R}) \tag{3.27}$$

$$e_L^* = e_L - \frac{1}{2a^2}(\pi_L^2 - \pi_L^{*2}), \qquad e_R^* = e_R - \frac{1}{2a^2}(\pi_R^2 - \pi_R^{*2}), \tag{3.28}$$

$$Z_L^* = Z_L, \qquad Z_R^* = Z_R, \tag{3.29}$$

$$u_{2,L}^* = u_{2,L}, \qquad u_{2,R}^* = u_{2,R}, \tag{3.30}$$

where $\kappa = \kappa(\rho_L, \alpha_L, \rho_R, \alpha_R)$.

Proof. The solution structure (3.24) directly follows from the linear degeneracy of the eigenvalues stated in Lemma 3.2. To derive the solution for the states $W^*_{L/R}$ one uses the Riemann invariants (3.21) and (3.22) and solves the resulting system. □

The relaxation parameter a can be determined locally analogously to the homogeneous case defined in equation (2.48) which was taken from [11]. The value a in the Riemann problem is then obtained by taking the maximum $a = \max(a_L, a_R)$ where the left and right relaxation parameters are obtained from the initial condition of the Riemann problem W_L, W_R. Since the solution of the Riemann problem depends on the source term, so do

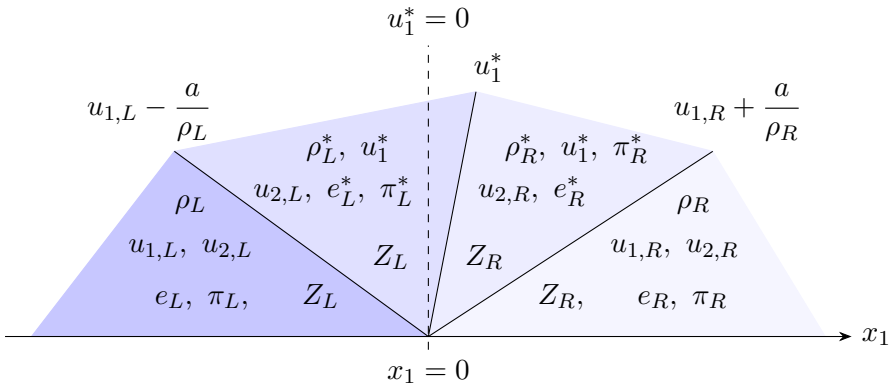


Figure 3.1: Wave structure of the Riemann problem of the Relaxation model (3.18).

a_L and a_R. This is expressed by replacing the pressure terms $p_R - p_L$ in (2.48) by the discrete hydrostatic equation $p_R - p_L - \kappa(\beta_R - \beta_L)$.

To fully determine the approximate Riemann solution, it remains to give an explicit form of the function $\kappa(\rho_L, \alpha_L, \rho_R, \alpha_R)$. We seek κ such that we achieve the well-balanced property. Indeed from (3.25) we see, that controlling the intermediate velocity u_1^* is the key to achieve this property. The choice for κ is discussed in the next lemma.

Lemma 3.4. *Let the initial condition W_L, W_R in (3.23) be given at relaxation equilibrium (3.16). Furthermore assume that W_L, W_R are a discrete stationary solution, that is they fulfil the relations*

$$u_{1,L} = 0, \quad u_{1,R} = 0, \quad \frac{\rho_{L,R}}{\alpha_{L,R}} = 1, \quad \frac{p_{L,R}}{\beta_{L,R}} = 1. \tag{3.31}$$

If the function κ is discretized such that

$$\kappa(\rho_L, \alpha_L, \rho_R, \alpha_R) = 1, \tag{3.32}$$

for a discrete stationary solution W_L, W_R, then the intermediate states in (3.24) satisfy

$$W_L^* = W_L, \quad W_R^* = W_R,$$

that is the approximate Riemann solver is at rest.

Proof. Let W_L, W_R be given satisfying (3.31). If $u_1^* = 0$, we immediately have from (3.25)-(3.29) together with $u_{1,L,R} = 0$ that $W_L^* = W_L$ and $W_R^* = W_R$. It remains to show, that $u_1^* = 0$. Combining (3.33) with (3.31) we have that

$$\pi_R - \pi_L - \kappa(Z_R - Z_L) = p_R - p_L - (\beta_R - \beta_L) = 0$$

since the initial condition of the Riemann problem is in relaxation equilibrium. Therefore we find $u_1^* = 0$ and the proof is completed. □

In the following we will use a symmetric second order discretization for κ given by

$$\kappa(\rho_L, \rho_R, \alpha_L, \alpha_R) = \frac{1}{2}\left(\frac{\rho_R}{\alpha_R} + \frac{\rho_L}{\alpha_L}\right). \tag{3.33}$$

We can immediately verify, that this approximation fulfils the requirement (3.32) for an piecewise constant stationary solution W_L, W_R.

We conclude the section by stating the positivity property of the Riemann solver defined in Lemma 3.3. Therefore we define the analogue of Ω_{phy} for the relaxation system (3.9) which is given by

$$\Omega_{phy}(W) = \left\{ W \in \mathbb{R}^{d+4}; \rho > 0, e > 0 \right\}. \tag{3.34}$$

The connection to $\Omega_{phy}(w)$ defined for the original equations (1.75) is easily computed by $\Omega_{phy}(QW^{eq}) = \Omega_{phy}(Q\mathcal{M}(w)) = \Omega_{phy}(w)$.

Lemma 3.5. *Given initial data $W_L, W_R \in \Omega_{phy}(W^{eq})$, then there is $W_{\mathcal{RS}}(\frac{x}{t}; W_L, W_R) \in \Omega_{phy}(W^{eq})$ for the relaxation parameter a sufficiently large. This means the approximate Riemann solver is robust with respect to the positivity of density ρ and internal energy e.*

Proof. Using the intermediate states for the density (3.27) and the velocity (3.25), we can write

$$\frac{1}{\rho_L^*} = \frac{1}{\rho_L} - \frac{1}{2a}(u_{1,R} - u_{1,L}) + \frac{1}{2a^2}\left(p_R - p_L - \kappa(\beta_R - \beta_L)\right). \tag{3.35}$$

Since $\rho_L > 0$, all terms that could become negative can be controlled by choosing the relaxation parameter large enough to ensure the positivity of ρ_L^*. The positivity of ρ_R^* can be shown analogously. For the intermediate state for the internal energy we have

$$\begin{aligned} e_L^* = e_L &+ \frac{1}{2a^2}\Big(- \pi_L^2 + \frac{1}{2}\left(p_L + p_R - \kappa(\beta_R - \beta_L)\right)^2 \Big) \\ &+ \frac{1}{2a}\Big(p_L + p_R - \kappa(\beta_R - \beta_L)\Big)(u_L - u_R) \\ &+ \frac{1}{8}(u_L - u_R)^2. \end{aligned}$$

This formula contains only positive terms or terms which can be controlled by a chosen sufficiently large. For e_R^* there can be found a similar formula. □

Especially this result means that $QW_{\mathcal{RS}}(\frac{x}{t}; W_L, W_R) \in \Omega_{phy}(w)$. We remark that the positivity of the intermediate densities $\rho_{L,R}^*$ guarantees the ordering of the approximate wave speeds

$$u_{1,L} - \frac{a}{\rho_L} < u_1^* < u_{1,R} + \frac{a}{\rho_R}. \tag{3.36}$$

We conclude the discussion of an approximate Riemann solver for (3.18) by a remark on the entropy stability.

Remark 3.6. *Let $\partial_t \eta + \partial_x(\eta u_1) \leq 0$ be an entropy inequality for the Euler equations with gravity. Then the approximate Riemann solver is consistent with the entropy inequality. This follows directly from Theorem 8 given in [28], since the proof given there is independent of the treatment of source term and thus can be directly applied here.*

3.4 Numerical scheme

In this section, we describe the numerical scheme associated with the approximate Riemann solver developed in the previous section. We consider a Cartesian mesh with a uniform grid size Δx as defined in Chapter 2. The numerical scheme consists of a Godunov type scheme and a subsequent projection step (3.20). The fundamentals of the Godunov type scheme with a source term were discussed in Section 2.4. We use the properties stated therein to construct the first order scheme. Additionally an extension to second order is given with the focus on a linear reconstruction that preserves the well-balanced and the positivity property.

3.4.1 First order scheme

To define the Godunov-type scheme, we consider the local Riemann problem for (3.18) at the interface $x = x_{1,i+1/2}$ depending on the left and right cell averages W_i^n, W_{i+1}^n. The cell

averages are defined in (2.2). The local dimensional split Riemann problems are given by

$$\begin{aligned} &\partial_t W + \partial_{x_1} F_1(W) = S_1(W), \\ &W^0(\mathbf{x}) = \begin{cases} \mathcal{M}(w_i^n) = W_i^{eq,n} & \text{if} \quad x_1 < x_{1,i+1/2} \\ \mathcal{M}(w_{i+1}^n) = W_{i+1}^{eq,n} & \text{if} \quad x_1 > x_{1,i+1/2} \end{cases}. \end{aligned} \tag{3.37}$$

Note that in comparison to the IVP for the Godunov type scheme with a source term discussed in Section 2.4, the variable Z is now contained in the state variables of the relaxation system W as defined in (3.13) and the initial condition for Z is defined by $W^0(\mathbf{x})$. The initial condition is given in relaxation equilibrium (3.16).

To give the numerical scheme as compact as possible, we write the numerical flux and source term together in one formulation. Looking at the Riemann solution, we see that the update of u_1^* in (3.25) is dependent on the source term $\kappa_{i+1/2}^n(Z_{i+1} - Z_i)$ and therefore also influences the position of the middle wave u_1^* in the wave structure of the Riemann problem depicted in Figure 3.1. Depending on u_1^*, the source term has an influence on the fluxes at the interface. Let $\lambda_L^- = u_{1,L} + a/\rho_L$ and $\lambda_R^+ = u_{1,R} + a/\rho_R$. Then, the update for the physical variables $w = (\rho, \rho u_1, \rho u_2, E)$ is given by

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x} \left(\mathfrak{f}_1^-(w_i^n, w_{i+1}^n) - \mathfrak{f}_1^+(w_{i-1}^n, w_i^n) \right), \tag{3.38}$$

where the numerical fluxes are defined by

$$\mathfrak{f}_1^\pm(w_i, w_{i+1}) = QF_1^\pm(\mathcal{M}(w_i), \mathcal{M}(w_{i+1})), \tag{3.39}$$

where

$$\begin{aligned} &\Big(F_1^-(W_L, W_R), F_1^+(W_L, W_R)\Big) = \\ &\begin{cases} \Big(F_1(W_L), & F_1(W_L) + \bar{S}_1(W_L, W_R)\Big) & \text{if} \quad \lambda_L^- > 0 \\ \Big(F_1(W_L^*), & F_1(W_L^*) + \bar{S}_1(W_L, W_R)\Big) & \text{if} \quad u_1^* > 0 > \lambda_L^- \\ \Big(F_1(W_L^*), & F_1(W_R^*)\Big) & \text{if} \quad u_1^* = 0 \\ \Big(F_1(W_R^*) - \bar{S}_1(W_L, W_R), & F_1(W_R^*)\Big) & \text{if} \quad \lambda_R^+ > 0 > u_1^* \\ \Big(F_1(W_R) - \bar{S}_1(W_L, W_R), & F_1(W_R)\Big) & \text{if} \quad \lambda_R^+ < 0 \end{cases}, \end{aligned} \tag{3.40}$$

where

$$\begin{aligned} \bar{S}_1(W_L, W_R) &= \mathcal{S}_1(W_L, W_R)(Z_R - Z_L) \\ &= \left(0, \frac{1}{2}\left(\frac{\rho_L}{\alpha_L} + \frac{\rho_R}{\alpha_R}\right), 0, u_1^* \frac{1}{2}\left(\frac{\rho_L}{\alpha_L} + \frac{\rho_R}{\alpha_R}\right), 0, 0\right)^T (Z_R - Z_L). \end{aligned}$$

From the flux definition (3.40), we directly see that in general $\mathfrak{f}_1^-(w_i, w_{i+1}) \neq \mathfrak{f}_1^+(w_i, w_{i+1})$, since we include the source term into the flux definition and the Euler equations with a gravitational potential are not conservative.

To see that the numerical scheme (3.38) is consistent with the balance law (3.5), we first show that the approximate Riemann solver

$$W_{\mathcal{R}}\left(\frac{x}{t}; w_L, \beta_L, w_R, \beta_R\right) = QW_{\mathcal{RS}}\left(\frac{x}{t}; \mathcal{M}(w_L), \mathcal{M}(w_R)\right)$$

for the original equations (3.5) and the source term discretization

$$\bar{\mathfrak{s}}_1(w_L, \beta_L, w_R, \beta_R) = Q\bar{S}_1(\mathcal{M}(w_L), \mathcal{M}(w_R)$$

are consistent. Using the definition of the intermediate states and the structure of the Riemann solver for the relaxation model given in Lemma 3.3, we immediately find for all $w \in \Omega$ that

$$W_{\mathcal{R}}\left(\frac{x}{t}; w, \beta, w, \beta\right) = QW_{\mathcal{RS}}\left(\frac{x}{t}; \mathcal{M}(w), \mathcal{M}(w)\right) = w \tag{3.41}$$

and for the source term discretization

$$Q\mathcal{S}_1(\mathcal{M}(w), \mathcal{M}(w)) = s_1(w). \tag{3.42}$$

Following the steps in [27], we integrate the original system (3.14) on $C_i \times (t^n, t^{n+1}]$ and use the solution of the Riemann problem $W_{\mathcal{R}}$ to find

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\Big(f_1\left(W_{\mathcal{R}}(0; w_i^n, \beta_i, w_{i+1}^n, \beta_{i+1})\right) - f_1\left(W_{\mathcal{R}}(0; w_{i-1}^n, \beta_{i-1}, w_i^n, \beta_i)\right)\Big) + \frac{1}{\Delta x}\int_{C_i}\int_{t^n}^{t^{n+1}} s_1\left(\widetilde{W}_\Delta(x,t)\right) dtdx. \tag{3.43}$$

We recall that $\widetilde{W}_\Delta(x,t)$ is a function piecewise given by approximate Riemann solutions and defined in (2.98). The source term integral is split into two parts given by

$$\frac{1}{\Delta x}\int_{C_i}\int_{t^n}^{t^{n+1}} s_1\left(\widetilde{W}_\Delta(x,t)\right) dtdx = \Delta t\Big(\bar{\mathfrak{s}}_1^+(w_{i-1}^n, \beta_{i-1}, w_i^n, \beta_i) + \bar{\mathfrak{s}}_1^-(w_i^n, \beta_i, w_{i+1}^n, \beta_{i+1})\Big) \tag{3.44}$$

with

$$\bar{\mathfrak{s}}_1^+(w_L, \beta_L, w_R, \beta_R) = \frac{1}{\Delta x \Delta t}\int_{x<0}\int_{t^n}^{t^{n+1}} s_1\left(W_{\mathcal{R}}\left(\frac{x}{t}; w_L, \beta_L, w_R, \beta_R\right)\right) dtdx \tag{3.45}$$

$$\bar{\mathfrak{s}}_1^-(w_L, \beta_L, w_R, \beta_R) = \frac{1}{\Delta x \Delta t}\int_{x>0}\int_{t^n}^{t^{n+1}} s_1\left(W_{\mathcal{R}}\left(\frac{x}{t}; w_L, \beta_L, w_R, \beta_R\right)\right) dtdx. \tag{3.46}$$

Depending on the wave position u_1^* we can determine the value of the sorce terms $\bar{\mathfrak{s}}^\pm$. We begin with the rectangle $[x_{1,i}, x_{1,i+1/2}] \times (t^n, t^{n+1}]$ and study the value of the source term at the left of the interface $x = x_{i+1/2}$. If $u_{1,i+1/2}^* > 0$, then $Z_L = Z_R = Z_i$ is constant, see also Figure 3.1, and we find due to the consistency of the source discretization

$$\bar{\mathfrak{s}}_1^-(w_i, \beta_i, w_{i+1}, \beta_{i+1}) = 0. \tag{3.47}$$

If $u^*_{1,i-1/2} < 0$, then Z jumps across the u^*_1 wave and

$$\bar{\mathfrak{s}}_1^-(w_i, \beta_i, w_{i+1}, \beta_{i+1}) = \frac{1}{\Delta x}\mathfrak{s}_1(w_i, \beta_i, w_{i+1}, \beta_{i+1}). \tag{3.48}$$

Using the sign function we can write

$$\bar{\mathfrak{s}}_1^-(w_i, \beta_i, w_{i+1}, \beta_{i+1}) = \frac{-\text{sign}(u^*_{1,i+1/2}) + 1}{2\Delta x}\mathfrak{s}_1(w_i, \beta_i, w_{i+1}, \beta_{i+1}). \tag{3.49}$$

Considering the rectangle $[x_{1,i-1/2}, x_{1,i}] \times (t^n, t^{n+1}]$, we find analogously

$$\bar{\mathfrak{s}}_1^+(w_{i-1}, \beta_{i-1}, w_i, \beta_i) = \frac{\text{sign}(u^*_{1,i-1/2}) + 1}{2\Delta x}\mathfrak{s}_1(w_{i-1}, \beta_{i-1}, w_i, \beta_i). \tag{3.50}$$

Using the source term definitions (3.50),(3.49), we directly obtain the formulation of scheme (3.38) with an upwinded source as

$$\begin{aligned} w_i^{n+1} = w_i^n + \frac{\Delta t}{\Delta x}\left(f_1(W_{\mathcal{R}}(0; w_i, \beta_i, w_{i+1}, \beta_{i+1})) - f_1(W_{\mathcal{R}}(0; w_{i-1}, \beta_{i-1}, w_i, \beta_i))\right) \\ + \Delta t \left(\frac{-\text{sign}(u^*_{1,i+1/2}) + 1}{2\Delta x}\mathfrak{s}_1(w_i, \beta_i, w_{i+1}, \beta_{i+1}) \right. \\ \left. + \frac{\text{sign}(u^*_{1,i-1/2}) + 1}{2\Delta x}\mathfrak{s}_1(w_{i-1}, \beta_{i-1}, w_i, \beta_i) \right). \end{aligned} \tag{3.51}$$

Alternatively, one can derive a scheme that uses a centred symmetric discretization of the source term as done in [27, 28, 29].

3.4.2 A second order extension

In the following, we want to extend the first order scheme to second order accuracy. We seek this extension such that the second order scheme has the well-balanced and positivity property. As already discussed in Section (2.2) for the homogeneous case, we solve a Riemann problem for (3.18), where its initial condition is based on the left and right interface values

$$W^0(\mathbf{x}) = \begin{cases} \mathcal{M}\left(w^{n,-}_{i+1/2}\right) & \text{if } x_1 < x_{1,i+1/2} \\ \mathcal{M}\left(w^{n,+}_{i+1/2}\right) & \text{if } x_1 > x_{1,i+1/2} \end{cases}, \tag{3.52}$$

where with $\Delta w = \sigma_i \frac{\Delta x}{2}$ the interface values are defined as

$$w^{n,+}_{i-1/2} = w_i^n - \Delta w \text{ and } w^{n,-}_{i+1/2} = w_i^n + \Delta w. \tag{3.53}$$

The second order scheme in space is then given by

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(\mathfrak{f}_1^-(w^{n,-}_{i+1/2}, w^{n,+}_{i+1/2}) - \mathfrak{f}_1^+(w^{n,-}_{i-1/2}, w^{n,+}_{i-1/2})\right) \tag{3.54}$$

with the definition of the numerical flux given by (3.39) and (3.40). In the following we will give a strategy how to obtain the slopes Δw in (3.53) such that the Riemann solver defined in (3.24) with initial data (3.52) is at rest and preserves the invariant domain Ω_{phy}.

A well-balanced hydrostatic reconstruction

To obtain a second order reconstruction of the interface values, we follow the MUSCL approach described in Section 2.2. Therein a reconstruction based on the conserved variables W_i was described. However, to ensure that the Riemann solver is at rest for an initial condition given by (3.52) computed from a piecewise constant stationary solution, we base the reconstruction on so called *equilibrium variables*. Therefore we adopt an approach known as the *surface gradient method* in the context of the shallow-water equations [89, 3] for treating hydrostatic equilibria. In this context, we define the equilibrium variables by $v = (\rho, u_1, u_2, q)$ which consist of the density, the two velocity components and a modified pressure q. It is defined by the following transformation of the pressure

$$\begin{aligned} q_{i-1} &= \pi_{i-1} + S_{i-1/2}, \\ q_i &= \pi_i \\ q_{i+1} &= \pi_{i+1} - S_{i+1/2}, \end{aligned} \tag{3.55}$$

where

$$S_{i+1/2} = \frac{1}{2}\left(\frac{\rho_i}{\alpha_i} + \frac{\rho_{i+1}}{\alpha_{i+1}}\right)(Z_{i+1} - Z_i) \tag{3.56}$$

is the second order approximation of the source term $\kappa \partial_x Z$ with the discretization of κ defined in (3.33). Note that $S_{i+1/2}$ appears also in the intermediate state u_1^* in the Riemann solution (3.25). We can rewrite the intermediate state u_1^* for two initial states W_i, W_{i+1} in terms of $S_{i+1/2}$ as follows

$$u^*_{1,i+1/2} = \frac{1}{2}(u_{1,i} + u_{1,i+1}) - \frac{1}{2a}(\pi_{i+1} - \pi_i - S_{i+1/2}). \tag{3.57}$$

The slopes σ_i for the equilibrium variables are computed by using the minmod limiter (2.75). The interface values in equilibrium variables are then given by

$$v^{n,+}_{i-1/2} = v^n_i - \sigma_i \frac{\Delta x}{2}, \quad v^{n,-}_{i+1/2} = v^n_i + \sigma_i \frac{\Delta x}{2}. \tag{3.58}$$

The interface values for the relaxation pressure π are obtained by setting

$$\pi^{n,-}_{i+1/2} = q^{n,-}_{i+1/2} \quad \text{and} \quad \pi^{n,+}_{i-1/2} = q^{n,+}_{i-1/2}. \tag{3.59}$$

The interface values for the momentum and energy are calculated from $v^{n,\mp}_{i\pm1/2}$ and the considered EOS. This reconstruction ensures that the Riemann solver (3.24) is at rest for an initial discrete stationary solution and is subject to the following lemma.

Lemma 3.7. *Let the cell averages $w^n_{i-1}, w^n_i, w^n_{i+1}$ form are a discrete stationary solution, that is they fulfil the relations (3.31). Further, let the initial condition (3.23) of the Riemann problem be composed of the interface values $w^{n,-}_{i+1/2}, w^{n,+}_{i+1/2}$ obtained under the reconstruction in equilibrium variables $\rho, \mathbf{u}, q$. Then the approximate Riemann solver $W_{\mathcal{RS}}$ defined in (3.24) with the intermediate velocity defined by (3.57), where $S_{i+1/2}$ is defined by (3.56), is at rest.*

Proof. Since the initial condition $W_{i-1}^n = \mathcal{M}(w_{i-1}^n), W_i^n = \mathcal{M}(w_i^n), W_{i+1}^n = \mathcal{M}(w_{i+1}^n)$ is given in hydrostatic equilibrium, the velocity components are zero and we have from the definition of the minmod limiter (2.75) that

$$u_{1,i+1/2}^- = u_{1,i+1/2}^+ = u_{1,i}^n = 0 \quad \text{and} \quad u_{2,i+1/2}^- = u_{2,i+1/2}^+ = u_{2,i}^n = 0. \tag{3.60}$$

We show that $u_1^* = 0$ from which immediately follows that $W_i^* = W_i^n$ and $W_{i+1}^* = W_{i+1}^n$. From the definition of u_1^* given in (3.57) remains

$$u_1^* = -\frac{1}{2a}\left(\pi_{i+1/2}^{n,+} - \pi_{i-1/2}^{n,-} - S_{i+1/2}\right). \tag{3.61}$$

Since W_i^n, W_{i+1}^n are in relaxation equilibrium and are a discrete hydrostatic equilibrium solution, we have $S_{i+1/2} = \beta_{i+1} - \beta_i$. Furthermore, we have due to the reconstruction in equilibrium variables (3.55) for the left and right slopes

$$q_i - q_{i-1} = \pi_i - \pi_{i-1} - S_{i-1/2} = \beta_i - \beta_{i-1} - S_{i-1/2} = 0 \tag{3.62}$$

and analogue $q_{i+1} - q_i = 0$. From the definition of the minmod limiter we obtain $\sigma = 0$ and therefore

$$\pi_{i+1/2}^{n,-} = \pi_i^n = p_i^n \quad \text{and} \quad \pi_{i+1/2}^{n,+} = \pi_{i+1}^n = p_{i+1}^n. \tag{3.63}$$

Therefore we can further simplify u_1^* and obtain

$$u_1^* = -\frac{1}{2a}(\pi_{i+1}^n - \pi_i^n - \beta_{i+1} + \beta_i) = 0. \tag{3.64}$$

This completes the proof. □

A positivity preserving reconstruction

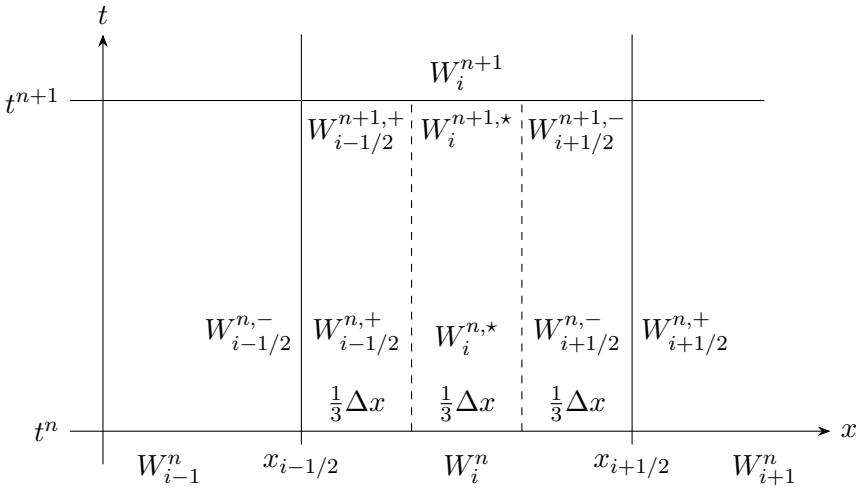


Figure 3.2: Positivity reconstruction.

As we have seen in Section 2.2, the positivity of the interface values for density and internal energy followed directly from the minmod reconstruction in conserved variables.

However here, we do a reconstruction in equilibrium variables which is in principle a reconstruction in primitive variables ρ, u_1, u_2, p. To ensure that the reconstruction also provides positive density and energy interface values, we derive a limiting procedure for the slopes σ obtained from the minmod function in the hydrostatic reconstruction. To do so, we follow the approach of Berthon [6] depicted in Figure 3.2. The update w_i^{n+1} obtained with the MUSCL scheme (3.54) can be understood as the average on the cell C_i composed of three values $w_{i-1/2}^{n+1,+}, w_i^{n+1,\star}, w_{i+1/2}^{n+1,-}$ that were evolved with the first order scheme respectively. The third state $w_i^{n,*}$ is uniquely defined by the conservation property

$$w_i^n = \frac{1}{3}\left(w_{i-1/2}^{n,+} + w_i^{n,*} + w_{i+1/2}^{n,-}\right). \tag{3.65}$$

In other words, the cell C_i is now divided into three regions with the volume $\frac{\Delta x}{3}$ filled with three constant states. Now, a necessary condition for the second order scheme in space in order to preserve the invariant domain Ω_{phy} is according to [6] given by

$$w_{i-1/2}^{n,+}, w_i^{n,\star}, w_{i+1/2}^{n,-} \in \Omega_{phy} \tag{3.66}$$

under one third of the CFL condition of the first order scheme (3.38). This necessary condition can be rewritten as to require in relaxation equilibrium

$$\rho_{i-1/2}^{n,+},\ \rho_i^{n,\star},\ \rho_{i+1/2}^{n,-} > 0 \quad \text{and} \quad p_{i-1/2}^{n,+},\ p_i^{n,\star},\ p_{i+1/2}^{n,-} > 0. \tag{3.67}$$

Since we have employed a minmod function to obtain the slope for the density $\Delta\rho$, the positivity of the interface values is guaranteed. For any conserved state w, we immediately obtain from the conservation property (3.65) that for the middle state holds

$$w_i^\star = w_i^n. \tag{3.68}$$

Therefore we obtain $\rho_i^\star = \rho_i^n$ and the necessary condition for the density is already fulfilled by the hydrostatic reconstruction. The interface values of the pressure however are obtained by calculating the slopes with the transformed pressure q. Therefore in general we do not obtain the positivity from the minmod limiter. Let the slope for the equilibrium variable q be denoted by Δq. Then the positivity of $p_{i\pm1/2}^{n,\mp}$ is guaranteed as soon as

$$\left|\frac{\Delta q}{p_i}\right| < 1. \tag{3.69}$$

This can be achieved by limiting Δq and the slope for the pressure can be defined by

$$\Delta p = p_i^n \max\left(-1, \min\left(1, \frac{\Delta q}{p_i^n}\right)\right). \tag{3.70}$$

It is left to verify whether also $p_i^\star > 0$. From the conservation property, we have $E_i^\star = E_i^n > 0$. From this we obtain with straightforward calculations the following relation

$$p_i^\star = p_i^n - (\gamma-1)(\rho_i^n|\Delta\mathbf{u}|^2) - 2(\gamma-1)(\Delta\rho\langle\Delta\mathbf{u}, \mathbf{u}_i^n\rangle) \tag{3.71}$$

which is a function of $\Delta\mathbf{u} = (\Delta u_1, \Delta u_2)^T$. Therefore, the positivity of $p_i^\star$ depends on the value of $\Delta\mathbf{u}$ which leads to a limiting of the velocity slopes $\Delta\mathbf{u}$. Therefore we do the following Ansatz

$$\Delta\mathbf{u} = \omega\Delta\bar{\mathbf{u}}, \tag{3.72}$$

with $0 < \omega \leq 1$ that preserves the sign of $\Delta\mathbf{u}$. For $\Delta\mathbf{u} = 0$ we immediately obtain $p_i^\star = p_i^n > 0$. For $\omega > 0$ and $\Delta\bar{\mathbf{u}} \neq 0$, we use (3.72) in (3.71) and consider the following quadratic function in ω

$$\Pi(\omega) = p_i^n - 2(\gamma-1)\Delta\rho\langle\Delta\bar{\mathbf{u}}, \mathbf{u}_i^n\rangle\omega - (\gamma-1)\rho_i^n|\Delta\bar{\mathbf{u}}|^2\omega^2.$$

The roots $\omega_{1,2}$ of $\Pi(\omega)$ are real, distinct and given by

$$\omega_{1,2} = -\frac{\Delta\rho\langle\Delta\bar{\mathbf{u}}, \mathbf{u}_i^n\rangle}{\rho_i^n|\Delta\bar{\mathbf{u}}|^2} \mp \frac{1}{\rho_i^n|\Delta\bar{\mathbf{u}}|^2}\sqrt{\Delta\rho^2\langle\Delta\bar{\mathbf{u}}, \mathbf{u}_i^n\rangle^2 + |\Delta\bar{\mathbf{u}}|^2\frac{\rho_i^n p_i^n}{\gamma-1}}$$

A straightforward computation shows that

$$\omega_1 < 0 < \omega_2.$$

Since $\Pi(\omega)$ is concave, $p_i^\star > 0$ for $\omega \in [0, \omega_2)$. Let $\Delta\bar{\mathbf{u}}$ be the velocity slopes obtained by the minmod function in the hydrostatic reconstruction, then the limited slopes are given by

$$\Delta\mathbf{u} = \min(1, \omega_2)\Delta\bar{\mathbf{u}}. \tag{3.73}$$

Using this limiting procedure does not make the resulting scheme more complex, but allows for the preservation of the positivity of density and energy also for the second order scheme. Summing up, we obtain the following result.

Lemma 3.8. *Let the initial data be composed by cell averages $W_{i-1}^n = \mathcal{M}(w_{i-1}^n)$, $W_i^n = \mathcal{M}(w_i^n)$, $W_{i+1}^n = \mathcal{M}(w_{i+1}^n)$ contained in $\Omega_{phy}(\mathcal{M}(w))$. Further, let the initial condition (3.23) of the Riemann problem be composed by the interface values $W_{i+1/2}^{n,-}, W_{i+1/2}^{n,+}$ obtained by the limited hydrostatic reconstruction (3.70),(3.73). Then the approximate Riemann solver $W_{\mathcal{RS}}$ defined in (3.24) preserves the invariant domain Ω_{phy} for a sufficiently large relaxation parameter a.*

Proof. Due to the limiting of the slopes, the initial data $W_L = W_{i+1/2}^{n,-}$ and $W_R = W_{i+1/2}^{n,+}$ of the Riemann solver $W_{\mathcal{RS}}$ are in Ω_{phy}. The statement follows then directly from Lemma 3.5. □

Second order in time

To obtain a fully second order scheme in space and time, we combine the second order scheme in space given by (3.54) with the second order time integration scheme (2.81), proposed in [6], that allows for different time increments $\Delta t_1, \Delta t_2$ for each computational stage respectively.

The second order update for the physical variables is given by

$$\begin{aligned}
w_i^{(1)} &= w_i^n - \frac{\Delta t_1}{\Delta x}\left(\mathfrak{f}_1^-\left(w_{i+1/2}^{n,-}, w_{i+1/2}^{n,+}\right) - \mathfrak{f}_1^+\left(w_{i-1/2}^{n,-}, w_{i-1/2}^{n,+}\right)\right),\\
w_i^{(2)} &= w_i^{(1)} - \frac{\Delta t_2}{\Delta x}\left(\mathfrak{f}_1^-\left(w_{i+1/2}^{(1),-}, w_{i+1/2}^{(1),+}\right) - \mathfrak{f}_1^+\left(w_{i-1/2}^{(1),-}, w_{i-1/2}^{(1),+}\right)\right),\\
w_i^{n+1} &= \left(1 - \frac{\Delta t}{\Delta t_1 + \Delta t_2}\right) w_i^n + \frac{\Delta t}{\Delta t_1 + \Delta t_2} w_i^{(2)}
\end{aligned} \tag{3.74}$$

with the total time step

$$\Delta t = \frac{2\Delta t_1 \Delta t_2}{\Delta t_1 + \Delta t_2}. \tag{3.75}$$

This time integration method has the advantage that the CFL criterion can be met for every single stage independently. In the following section, the CFL conditions for the first and second order scheme are given. Due to requiring that the first and second order schemes (3.38) and (3.54) preserve the invariant domain Ω_{phy}, the respective CFL conditions are more restrictive than one would expect from a second order finite volume scheme. However, the CFL bounds are necessary for the theoretical study of the properties of the schemes and could be relaxed in practice.

3.5 Properties of the numerical schemes

In this section we state the central properties in the two-dimensional setting of the first and second order scheme derived above. As motivated in Section 2.1.4, a scheme for multiple space dimensions can be obtained from considering the dimensional split Riemann problems in each direction. The fully two dimensional first order scheme for a uniform grid size Δx is then given by

$$\begin{aligned} w_{ij}^{n+1} = w_{ij}^n &- \frac{\Delta t}{\Delta x}\left(\mathfrak{f}_1^-(w_{i,j}^n, w_{i+1,j}^n) - \mathfrak{f}_1^+(w_{i-1,j}^n, w_{i,j}^n)\right) \\ &- \frac{\Delta t}{\Delta x}\left(\mathfrak{f}_2^-(w_{i,j}^n, w_{i,j+1}^n) - \mathfrak{f}_2^+(w_{i,j-1}^n, w_{i,j}^n)\right), \end{aligned} \tag{3.76}$$

where $\mathfrak{f}_1$ denotes the flux in x_1-direction and $\mathfrak{f}_2$ the analogue numerical flux in x_2-direction, as discussed in Section 2.1.4 about dimensional split Riemann problems. The fully second order scheme can be obtained analogously by adding to (3.74) the flux terms in x_2-direction and performing the linear construction in each space direction separately. We recall that C_{ij} denotes the two dimensional rectangular grid cell of area Δx^2 for a uniform space discretization $\Delta x_1 = \Delta x_2$ and the cell averages W_{ij}^n are computed according to (2.64). We begin with the well-balanced property.

Theorem 3.9. *Let the initial data w_{ij}^n be a discrete stationary solution, that is fulfilling (3.31) for all cells C_{ij} on the computational domain. Then the first order scheme (3.76) is well-balanced.*

Proof. Let $W_{i,j}^n = \mathcal{M}(w_{i,j}^n), W_{i+1,j}^n = \mathcal{M}(w_{i+1,j}^n)$ be the initial data for the Riemann problem (3.23) in x_1-direction and $W_{i,j}^n = \mathcal{M}(w_{i,j}^n), W_{i,j+1}^n = \mathcal{M}(w_{i,j+1}^n)$ for the Riemann problem in x_2-direction. As they form by assumption a discrete stationary solution, we know from Lemma 3.4 that the approximate Riemann solvers are at rest in each space direction. Therefore in the definition of the numerical flux (3.40) we have $F_1^-(W_{ij}^n, W_{i+1,j}^n) = F_1(W_{ij}^n)$ and $F_1^+(W_{i-1,j}^n, W_{ij}^n) = F_1(W_{ij}^n)$. The analogue relations we find for the fluxes $F_2^\pm$ in x_2-direction. Therefore we get in the update (3.76) with $QF_1(W_{ij}^n) = f_1(w_{ij}^n)$ and $QF_2(W_{ij}^n) = f_2(w_{ij}^n)$

$$w_{ij}^{n+1} = w_{ij}^n - \frac{\Delta t}{\Delta x}\Big(f_1(w_{ij}^n) - f_1(w_{ij}^n) + f_2(w_{ij}^n) - f_2(w_{ij}^n)\Big) = w_{ij}^n$$

and thus the first order scheme is well-balanced. □

The analogue result can be stated for the second order scheme.

Theorem 3.10. *Let the initial data w_{ij}^n be a discrete stationary solution, that is fulfilling (3.31) for all cells C_{ij} on the computational domain. Further let the interface values $w_{i\pm1/2,j}^{n,\mp}$, $w_{i,j\pm1/2}^{n,\pm}$ be obtained by the hydrostatic reconstruction. Then the second order scheme (3.76) in two space dimensions is well-balanced.*

Proof. Since the requirements of Lemma 3.7 are fulfilled, it follows that the Riemann solver at each interface is at rest. Since the second order time integration method is composed of first order steps, we can follow the lines of the proof of Theorem 3.9 and obtain

$$w_{ij}^{(1)} = w_{ij}^n, \quad w_{ij}^{(2)} = w_{ij}^n$$

and therefore $w_{ij}^{n+1} = w_{ij}^n$. Thus the second order scheme is well-balanced. □

Now we turn to the preservation of the invariant domain Ω_{phy}. For the first order scheme we can proof the following result.

Theorem 3.11. *Let the initial data be given by $w_{ij}^n \in \Omega_{phy}$ on all cells C_{ij} on the computational domain. Then under the CFL condition*

$$\frac{\Delta t}{\Delta x} \max\left(\left| u_{1,ij}^n \pm \frac{a_{i\mp1/2,j}}{\rho_{ij}^n} \right|, \left| u_{2,ij}^n \pm \frac{a_{i,j\mp1/2}}{\rho_{ij}^n} \right| \right) \leq \frac{1}{4} \tag{3.77}$$

and a sufficiently large local relaxation parameter $a_{i\pm1/2,j}$, $a_{i,j\pm1/2}$, the first order scheme (3.76) preserves the invariant domain Ω_{phy}, that is $w_{ij}^{n+1} \in \Omega_{phy}$.

Proof. We begin the proof with the analysis of one spatial dimension. Due to the integral consistency of the approximate Riemann solver, the update (3.38) can be rewritten as

$$w_i^{n+1} = \frac{1}{\Delta x_1} \left(\int_{x_{i-1/2}}^{x_i} W_{\mathcal{R}}\left(\frac{x - x_{i-1/2}}{\Delta t}, w_{i-1}^n, w_i^n \right) dx + \int_{x_i}^{x_{i+1/2}} W_{\mathcal{R}}\left(\frac{x - x_{i-1/2}}{\Delta t}, w_i^n, w_{i+1}^n \right) dx \right), \tag{3.78}$$

see also [11]. Due to the CFL condition, we know that the Riemann solutions are not interacting and therefore we have $W_{\mathcal{R}}(\frac{x}{t}; w_{i-1}^n, w_i^n), W_{\mathcal{R}}(\frac{x}{t}, w_i^n, w_{i+1}^n) \in \Omega_{phy}$ from Lemma 3.5. By convexity of Ω_{phy} and the relation (3.78) we have $w_i^{n+1} \in \Omega_{phy}$. Now we extend the analysis to two space dimensions. For this we can rewrite the update formula (3.76) as

$$\begin{aligned} w_i^{n+1} &= w_i^n - \frac{\Delta t}{\Delta x} \sum_{l=1}^{2} \left(\mathfrak{f}_l^-(w_i, w_{i+1}) - \mathfrak{f}_l^+(w_{i-1}, w_i) \right) \\ &= \frac{1}{2} \sum_{l=1}^{2} \underbrace{\left(w_i^n - \frac{2\Delta t}{\Delta x} \left(\mathfrak{f}_l^-(w_i, w_{i+1}) - \mathfrak{f}_l^+(w_{i-1}, w_i) \right) \right)}_{\in\, \Omega_{phy} \quad \text{due to (3.78)}}. \end{aligned} \tag{3.79}$$

Therefore by the convexity of (3.79), we have $w_i^{n+1} \in \Omega_{phy}$. □

Now we prove the preservation of the invariant domain for the second order method. Since we decided to reconstruct in equilibrium variables and followed the strategy in [6] to define a slope limiting, we will use results of Berthon given in [6] to ensure the positivity. We give the following Theorem as a summary of the work from [6] within our context.

Theorem 3.12. *Let the initial data be given by $w_{ij} \in \Omega_{phy}$ on all cells C_{ij} on the computational domain. Then under the minimum Δt computed according to the directional CFL conditions*

$$\frac{\Delta t}{\Delta x} \max\left(\left| u^{n,+}_{1,i-1/2,j} - \frac{a_{i-1/2,j}}{\rho^{n,+}_{i-1/2,j}} \right|, \left| u^{n,-}_{1,i+1/2,j} + \frac{a_{i+1/2,j}}{\rho^{n,-}_{i+1/2,j}} \right|, \left| u^{n}_{1,ij} \pm \frac{a}{\rho^{n}_{ij}} \right| \right) < \frac{1}{4} \cdot \frac{1}{3} \tag{3.80}$$

$$\frac{\Delta t}{\Delta x} \max\left(\left| u^{n,+}_{2,i,j-1/2} - \frac{a_{i,j-1/2}}{\rho^{n,+}_{i,j-1/2}} \right|, \left| u^{n,-}_{2,i,j+1/2} + \frac{a_{i,j+1/2}}{\rho^{n,-}_{i,j+1/2}} \right|, \left| u^{n}_{2,ij} \pm \frac{a}{\rho^{n}_{ij}} \right| \right) < \frac{1}{4} \cdot \frac{1}{3} \tag{3.81}$$

and a sufficiently large local relaxation parameter $a_{i\pm1/2,j}, a_{i,j\pm1/2}$, the two dimensional second order scheme preserves the invariant domain Ω_{phy}, that is $w^{n+1}_{ij} \in \Omega_{phy}$.

Proof. The proof is a straightforward application of the Theorem 2.5 given in [6] since by Theorem 3.11 the first order scheme is preserves the invariant domain. Due to the limiting procedure, the interface values $w^{n,-}_{i-1/2,j}, w^{n,+}_{i+1/2,j}$ are contained in Ω_{phy}. Due to the CFL condition the Riemann solutions do not interact and by Lemma 3.8 the Riemann solutions lie in Ω_{phy}. The second order time integrator (2.81) used in the second order scheme is therefore a convex combination of positivity preserving states and we obtain $w^{n+1}_{ij} \in \Omega_{phy}$. □

The results given here can be straightforwardly extended to three dimensional problems.

3.6 Numerical results

To illustrate the properties presented in Section 3.5, we present several numerical experiments. We assume an ideal gas law with the equation of state given by

$$p = RT\rho = (\gamma - 1)\rho e$$

where we set $\gamma = 5/3$ if not otherwise specified.

We would like to stress that in all test cases the slope limiter given by (3.70),(3.73) was not necessary to ensure the positivity. All computations are carried out on double precision and errors are given in the L^1-norm. We start with well-balanced tests for different atmospheres and potentials.

3.6.1 Well-balanced tests

For the well-balanced tests, we consider different stationary solutions in one, two and three space dimensions using the second order scheme. The computations are performed on a uniform grid on the domain $[0,1]^d$, where d denotes the number of dimensions, up to a final time $T_f = 1.0$.

As a first example, we consider an isothermal hydrostatic atmosphere (1.90) in three space dimensions with a quadratic potential $\Phi(\mathbf{x}) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2)$ and choosing for the constant $C = 0$. The gas constant and temperature is set as $R = 1, T = 1$. The reference equilibrium given in terms of α and β is set according to (3.7).

The second example as is a polytropic atmosphere (1.91) in two dimensions for which we set $C = 0$ and $\chi = 1$. In this case we choose a linear gravitational field given by $\Phi(\mathbf{x}) = x_1 + x_2$. The functions α and β are given by (3.8).

The last example we consider is a non-polytropic and non-isothermal stationary state with the potential $\Phi(\mathbf{x}) = -\sum_{j=1}^{d} \sin(2\pi x_j)$. This corresponds to a non-constant gravitational acceleration over the computational domain. A steady state solution using this potential is given by

$$\begin{aligned} \rho(\mathbf{x}) &= c_\rho - 2\ \Phi(\mathbf{x}), \\ \mathbf{u} &= 0, \\ p(\mathbf{x}) &= c_p - c_\rho\ \Phi(\mathbf{x}) - \frac{1}{2}\sum_{j=1}^{d} \cos(4\pi x_j) + \sum_{i,j=1, j>i}^{d} \sin(2\pi x_i)\sin(2\pi x_j). \end{aligned} \tag{3.82}$$

The parameters c_ρ and c_p are chosen such that the initial condition for the density and pressure are positive. Since the hydrostatic equilibrium solution (3.82) is periodic on the domain $[0, 1]^d$ we use periodic boundary conditions. The reference equilibrium states α, β are given by definition by $\alpha = \rho$ and $\beta = p$. For this general stationary state calculations are performed in one, two and three space dimensions with $c_\rho = 3.0, 5.0, 7.0$ and $c_p = 3.0, 8.0, 14.5$ respectively. In all tests Neumann boundary conditions are used, since they fulfil the hydrostatic equilibrium. Alternatively also exact boundary conditions can be applied. The numerical results in terms of the L^1 error with respect to the exact hydrostatic solution are given in Tables 3.1 for the isothermal test, in Table 3.2 for the polytropic test and in Table 3.3 for the stationary solution (3.82). As can be seen, the errors are of order of machine precision independent of the number of space dimensions and space resolution. Even for coarse grids the well-balancing is ensured and it will not be necessary to use an especially fine grid to capture fluctuations around an equilibrium as we will see in Section 3.6.3. But before we will have a look at the accuracy of the schemes.

N	ρ	ρu	ρv	ρw	E
50	5.996E-017	1.450E-016	1.450E-016	1.438E-016	7.404E-017
150	4.930E-017	2.121E-016	2.121E-016	2.128E-016	6.426E-017

Table 3.1: L^1 error with respect to initial values in density, momentum and energy for isothermal equilibrium at $T_f = 1.0$.

N	ρ	ρu	ρv	E
100	4.796E-017	1.188E-016	1.188E-016	8.684E-017
500	1.036E-016	5.525E-016	5.525E-016	2.209E-016

Table 3.2: L^1 error with respect to initial values in density, momentum and energy for a polytropic equilibrium at $T_f = 1.0$.

N	ρ	ρu_1	ρu_2	ρu_3	E
200	2.187E-016	3.742E-015			1.263E-015
1000	2.220E-018	5.149E-016			4.529E-017
100	9.294E-017	2.285E-015	2.285E-015		1.214E-015
500	5.571E-016	6.459E-015	6.459E-015		3.257E-015
50	2.137E-016	3.092E-015	3.098E-015	3.032E-015	2.174E-015
150	4.930E-015	1.292E-014	1.299E-014	1.298E-014	5.000E-014

Table 3.3: L^1 error with respect to initial values in density, momentum and energy for a general stationary state at $T_f = 1.0$ for the dimensions $d = 1, 2, 3$.

3.6.2 Accuracy

To numerically verify the order of convergence, we consider the accuracy of the whole scheme, as well as the order of accuracy of the source term discretization.

Source term

To verify the accuracy of the source term, we model the case, where an equilibrium solution is balanced with a not corresponding reference hydrostatic equilibrium solution α, β. In practise this can happen if the reference equilibrium is too complicated to compute or the nature of the chosen pressure law changes during the simulation. However the scheme will still converge with second order to the correct equilibrium solution. To demonstrate this, we consider in one dimension the equilibrium solution (3.82) where we choose an isothermal reference equilibrium (3.7). To see the accuracy of the discretization of the hydrostatic equilibrium (3.1), we choose the first order scheme.

For the given discretization $S_{i+1/2}$ from (3.56) one finds with a straightforward Taylor expansion in one dimension that

$$\partial_x p(x_{i+1/2}) - \frac{\rho}{\alpha}(x_{i+1/2})\partial_x \beta(x_{i+1/2}) = \frac{1}{\Delta x}(p_{i+1} - p_i - S_{i+1/2}) + \mathcal{O}(\Delta x^2).$$

This shows that the hydrostatic equilibrium is approximated by second order for a hydrostatic equilibrium solution that does not coincide with the chosen α and β. From Table 3.4, we see that indeed the first order scheme converges to the hydrostatic equilibrium solution (3.82) with order two.

N	ρ		ρu		E	
50	9.083E-06	—	5.171E-05	—	1.841E-05	—
100	1.673E-06	2.441	1.303E-05	1.989	3.433E-06	2.423
200	3.428E-07	2.287	3.268E-06	1.995	7.114E-07	2.271
400	7.632E-08	2.167	8.182E-07	1.998	1.595E-07	2.157
800	1.790E-08	2.092	2.047E-07	1.999	3.757E-08	2.086
1600	4.328E-09	2.048	5.120E-08	1.999	9.106E-09	2.045

Table 3.4: L^1 error and convergence rates at $T_f = 1.0$ balancing (3.82) with an isothermal equilibrium.

Exact solutions

Now we turn to numerically verify the accuracy of the second order scheme. To demonstrate that the second order extension of the first order scheme has the expected accuracy, we compare the numerical solution to an exact solution of the Euler equations with gravity. We will consider two different solutions. The first exact solution of (1.75) for a quadratic gravitational potential is given by

$$\begin{aligned} \rho(\mathbf{x},t) &= \exp\left(\frac{1}{RT}\left(\frac{1}{2}\sum_{i=1}^{d} u_i^2 - \sum_{i=1}^{d}\frac{\zeta_i}{\eta_i}\cos(\eta_i t)\partial_{x_i}\Phi(\mathbf{x}) - \Phi(\mathbf{x})\right)\right), \\ u_i(t) &= \zeta_i \sin(\eta_i t), \\ p(\mathbf{x},t) &= RT\rho(\mathbf{x},t), \\ \Phi(\mathbf{x}) &= \frac{1}{2}\sum_{i=1}^{d}\eta_i^2 x_i^2, \end{aligned} \tag{3.83}$$

where the constants $\zeta_i > 0$ denote the amplitude of the velocities u_i and $\eta_i > 0$ are scaling constants. The velocity field is time dependent and for $\mathbf{u} = 0$ we obtain the isothermal hydrostatic equilibrium solution (1.90). Thus, we have set α and β according to the isothermal equilibrium solution (3.7). The calculations are performed for the three dimensional problem on the domain $[0,1]^3$ starting with 25 cells using exact boundary conditions (2.8). To be far enough from the equilibrium solution to be able to observe stable convergence rates, we have set high velocity amplitudes $\zeta_{1,2,3} = 20$ and the scaling constant is set to $\eta_{1,2,3} = 1$. The time frame is chosen such that the velocities are monotone increasing in time, where the density and pressure is exponentially decaying on the computational domain. The L^1 error and convergence rates are given in Table 3.6. It can be seen that the convergence rates are around two for the velocity field. The rates for density and energy are slightly below but approaching two. This can be explained by the use of the minmod limiter function as the density and pressure are almost constant in some parts of the computational domain and the calculated slopes are almost zero.

A second exact solution in two space dimensions for a linear potential taken from [18]

is given by

$$\begin{aligned}
\rho(\mathbf{x},t) &= 1 + 0.2\sin(\pi(x_1 + x_2 - t(u_{1_0} + u_{2_0}))),\\
u_1(t) &= u_{1_0},\\
u_2(t) &= u_{2_0},\\
p(\mathbf{x},t) &= p_0 + t(u_{1_0} + u_{2_0}) - (x_1 + x_2) + \frac{0.2}{\pi}\cos(\pi(x_1 + x_2 - t(u_{1_0} + u_{2_0}))).
\end{aligned} \tag{3.84}$$

Here the velocity field is constant throughout the simulation. For a zero velocity field, the exact solution (3.84) fulfils the hydrostatic equation (3.1) which is used to define α and β. They are given by

$$\begin{aligned}
\alpha(\mathbf{x}) &= 1 + 0.2\sin(\pi(x_1 + x_2)),\\
\beta(\mathbf{x}) &= p_0 - (x_1 + x_2) + \frac{0.2}{\pi}\cos(\pi(x_1 + x_2)).
\end{aligned} \tag{3.85}$$

For the parameters, we choose $u_{1_0} = 20, u_{2_0} = 20$ to be sufficiently far from the equilibrium solution and $p_0 = 4.5$ which ensures the positivity of the pressure. The two dimensional computational domain is given by $[0,1]^2$ and the computations are performed with exact boundary conditions up to a final time $T_f = 0.01$. The L^1 error and convergence rates are given in Table 3.5. It can be seen that the convergence rates throughout all variables are around 1.9 but approaching two. Since the density and pressure have a maximum on the computational domain, the slopes calculated with the minmod limiter function reduces to zero at the maximum and therefore the numerical convergence rates are reduced.

3.6.3 Perturbations of equilibrium solutions

As already mentioned before, the main motivation of using well-balanced schemes is that they are able to capture the hydrostatic equilibrium at machine precision. Therefore they allow an accurate description of small fluctuations on the equilibrium state even for coarse meshes.

Evolution of a small pressure perturbation

With the following numerical test, taken from [53], the evolution of a small perturbation in one dimension which is added to an initial isothermal hydrostatic solution is investigated. Therefore the equilibrium pressure is perturbed by a Gaussian centred at $x = 0.5$ with an amplitude η. The initial data is given by

$$\rho(x) = \exp(-\Phi(x)), \tag{3.86}$$

$$p(x) = \exp(-\Phi(x)) + \eta\exp(-100(x - 0.5)^2). \tag{3.87}$$

For the simulation we use a linear potential $\Phi(x) = x$ and the functions α and β are chosen according to the isothermal atmosphere (3.7). In the first numerical experiment, we choose quite a large perturbation with the amplitude of $\eta = 10^{-3}$. The final time $T = 0.2$ is chosen such that the resulting waves are still inside the computational domain $[0,1]$ and at the boundary the hydrostatic equilibrium is fulfilled. Therefore Neumann boundary conditions are sufficient. To illustrate the behaviour of the perturbation, the

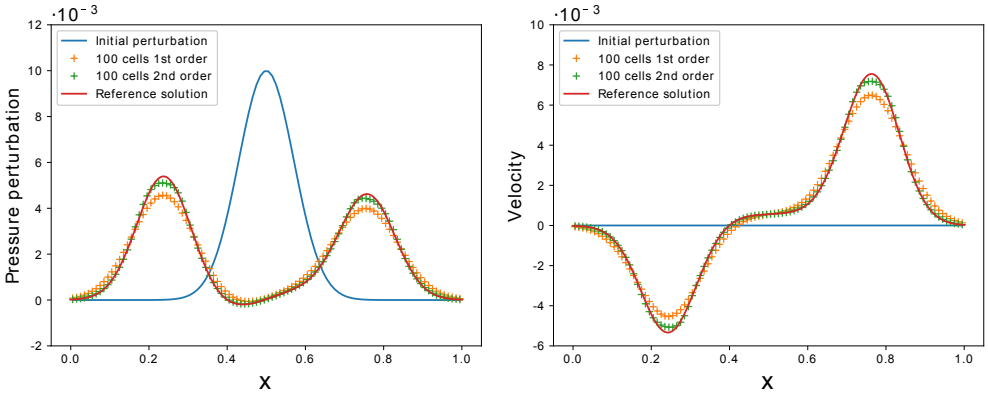


Figure 3.3: Evolution of the perturbation in pressure (left) and velocity (right) at time $T_f = 0.2$ for an initial Gaussian pressure perturbation (3.86) with an amplitude $\eta = 10^{-3}$.

difference between the pressure and initial data $p(x) - p_0(x)$ and the velocity perturbation are plotted.

The numerical solutions plotted in Figure 3.3 are computed by the first and second order scheme with 100 cells respectively. In addition the initial pressure perturbation as well as a reference solution also computed with the second order scheme using 32000 cells are given. Overall can be said that the first and second order well-balanced scheme capture the evolution of the perturbation accurately. This can especially be seen at the boundaries, where the solution is still in equilibrium. Due to the higher accuracy of the second order scheme, the resulting waves travelling towards the left and right boundary are well captured. As expected, the first order scheme is more diffusive but still gives a good description.

To see how the well-balanced scheme performs compared to an unbalanced one, we redo the same test case and choose the reference equilibrium polytropic instead of isothermal. This results in a not well-balanced scheme. As the before chosen amplitude of $\eta = 10^{-3}$ is quite large, we set now $\eta = 10^{-8}$. This perturbation is now of the order of the error that arises when the background atmosphere is calculated with the not well-balanced scheme. In Figure 3.4, the initial perturbation, the solution for the well-balanced scheme with 500 cells and the solution for the unbalanced scheme with 250 and 500 cells are displayed. As can be seen, the peaks in the pressure perturbation are also captured from the unbalanced scheme but the behaviour near the boundary, where the solution is in a stationary state, is off as well as the behaviour of the velocity. As the number of cells increases, the solution seems to converge to the solution of the well-balanced scheme, but a quite fine mesh is needed to have a good resolution of the travelling waves. This underlines the importance of using well-balanced schemes to correctly resolve perturbations upon an equilibrium state especially on coarse grids.

Rayleigh-Taylor instability

Rayleigh-Taylor instabilities arise at the interface between two gases with different densities where the lighter gas is accelerated towards the heavier gas for example by gravity.

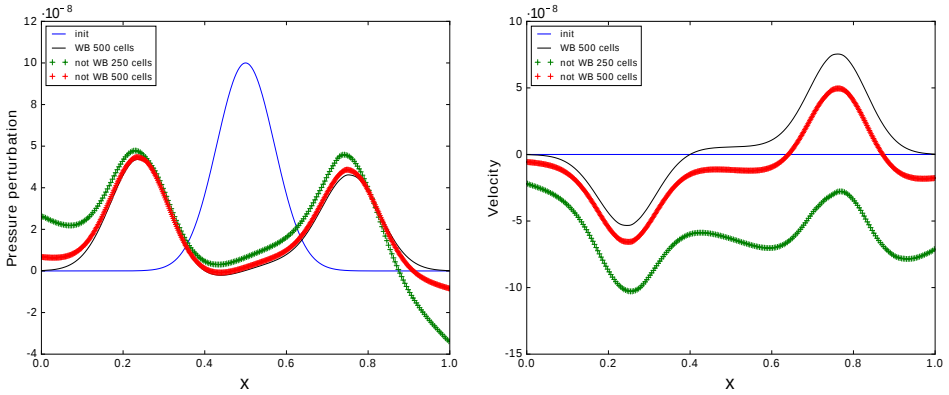


Figure 3.4: Evolution of the perturbation in pressure (left) and in velocity (right) comparing a well-balanced scheme to a not well-balanced scheme for an initial Gaussian pressure perturbation (3.86) with an amplitude $\eta = 10^{-8}$.

In this test case, taken from [53], in a radial set up from [17], this two layers of gas are modelled by a gas in isothermal equilibrium at the core, where a perturbation in density builds the second layer. The gravitational field is modelled in radial coordinates by $\Phi(r) = r$ acting away from the core. The considered computational domain is given by $[-1, 1]^2$ and the two initial density layers are given by

$$p = \begin{cases} \exp(-r) & r \leq r_0 \\ \exp(-\frac{r}{\mu} + r_0 \frac{1-\mu}{\mu}) & r > r_0 \end{cases}, \quad \rho = \begin{cases} \exp(-r) & r \leq r_i(\theta) \\ \frac{1}{\mu} \exp(-\frac{r}{\mu} + r_0 \frac{1-\mu}{\mu}) & r > r_i(\theta) \end{cases},$$

where

$$\mu = \frac{\exp(-r_0)}{\exp(-r_0) + \Delta\rho}$$

and the interface curve is given by

$$r_i(\theta) = r_0(1 + \nu \cos(k\theta)).$$

This results in a jump in density by an amount of $\Delta\rho$ at the interface defined by $r = r_i(\theta)$ whereas the pressure is continuous. Following [17], we take $\Delta\rho = 0.1$, $\eta = 0.02$, $k = 20$. For the computation, we use the second order well-balanced scheme on a mesh consisting of 240×240 cells and the results for the density are given in Figure 3.5. The picture on the top left displays the initial condition. For $r < r_0(1-\eta)$ and $r > r_0(1+\eta)$, it is in stable equilibrium but due to the discontinuous density, a Rayleigh-Taylor instability develops at the interface of the two density layers. The other pictures in Figure 3.5 show the numerical solution at times $t = 2.9, 3.8$ and $t = 5.0$ from the top right to the bottom right. It can be seen that the characteristic mushroom like instabilities develop around the discontinuous interface. These can be seen due to the application of a well-balanced scheme. Using a non well-balanced scheme instead would lead to spurious disturbances developing away from the density interface. Since a Cartesian mesh and dimensional splitting is used to calculate a numerical solution in a radial set-up, the instabilities evolve along the coordinate axes and thus the solution is not completely radially symmetric.

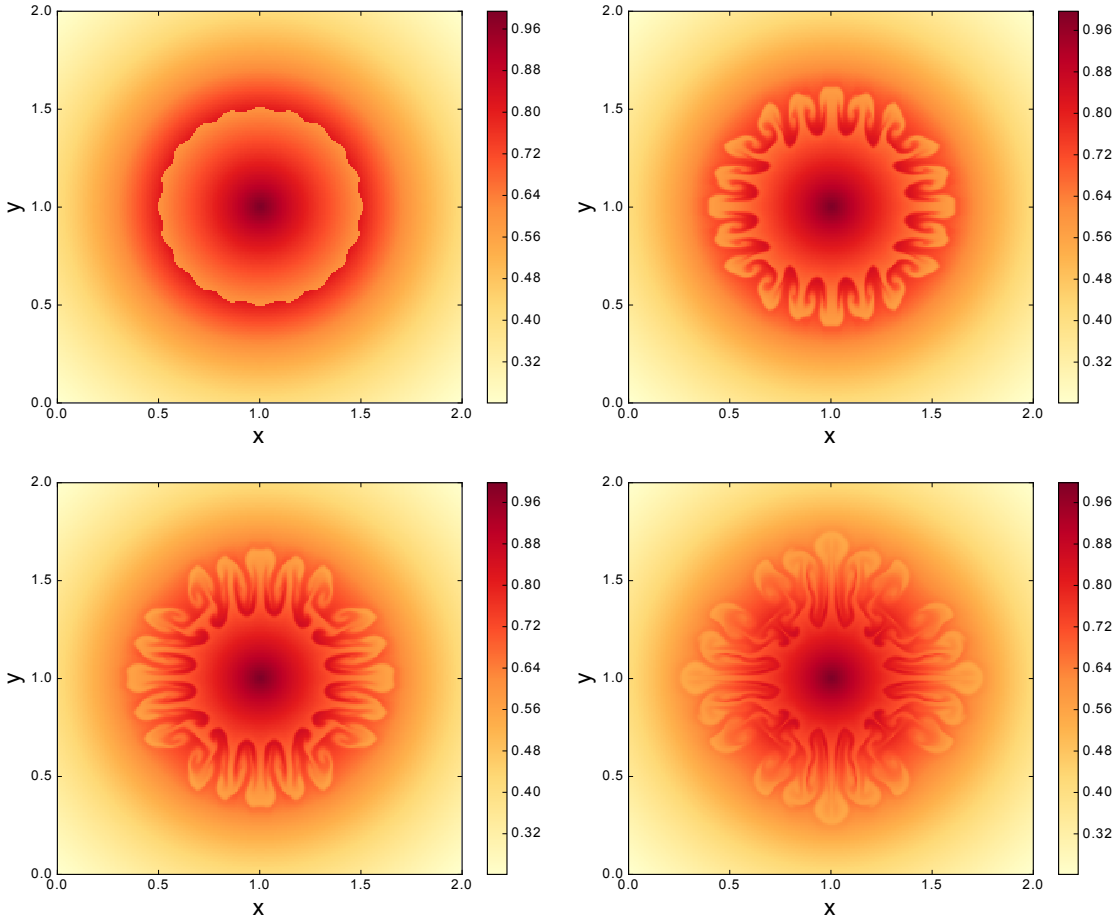


Figure 3.5: Rayleigh-Taylor instability in density in radial gravitational field at times $t = 0$ (top left), $t = 2.9$ (top right), $t = 3.8$ (bottom left) and $t = 5.0$ (bottom right).

3.6.4 Strong rarefraction test

We conclude this section on numerical validation of the properties of the scheme by turning to the positivity preserving property. A classical test-case in this context is a strong rarefaction test. The rarefaction test performed here, follows the two dimensional 1-2-0-3 rarefaction test from [32] which is developed for the homogeneous Euler equations with a zero source term. Originally the density and energy are chosen constant and are given by $\rho = 1$ and $E = 3$. The x_1-component of the velocity jumps from -2 to 2 whereas the x_2 component of the velocity is zero, that is

$$u_1 = \begin{cases} -2 & \text{for } x_1 < 0.5, \\ +2 & \text{for } x_1 \geq 0.5, \end{cases} \qquad u_2 = 0$$

The jump in u_1 launches two strong rarefaction waves in opposite directions along the x_1 coordinate. To include the presence of the gravitational field, we modify this set-up and define ρ and p to be in isothermal equilibrium with a quadratic potential

$$\Phi(\mathbf{x}) = \frac{1}{2}\big(x_1 - 0.5\big)^2 + \big(x_2 - 0.5\big)^2\big)$$

centred around $\mathbf{x} = (0.5, 0.5)$. To have the initial maximal energy to be $E_{\max} = 3$, we set $R\,T = \gamma - 1$ and $\gamma = 1.4$.

The calculations were performed with the first and the second order scheme on 100 cells in each direction on the domain $[0, 1]^2$ up to $T_f = 0.1$. The results given in Figure 3.6 are projected on the x_1 axis and show the density, the velocity component u_1 and the energy. As can be seen, the density and energy come close to zero at the center of the cell as the rarefaction waves travel to the left and right, but remain positive throughout the computations.

3.7 Conclusion

In this chapter, we constructed a method for well-balancing arbitrary given hydrostatic equilibria of the compressible Euler equations with gravity. A key element in achieving the well-balanced and the positivity preserving property of the developed Godunov type scheme was the use of a relaxation procedure to be able to compute an approximate Riemann solution to the Riemann problems arising at the interfaces due to the finite volume discretization. To this first order scheme, a quite natural extension to multiple dimensions as well as to second order was given that inherited the well-balanced and positivity property of the first order scheme.

The overall aim of this work is to extend the scheme, that provably works well in the compressible regime, to all speeds, especially to near incompressible regimes. As a first step in this direction we focus on the construction of an all-speed scheme for the homogeneous Euler equations in the following chapter before we turn to the construction of an all-speed scheme for the Euler equations with a gravitational potential in Chapter 5.

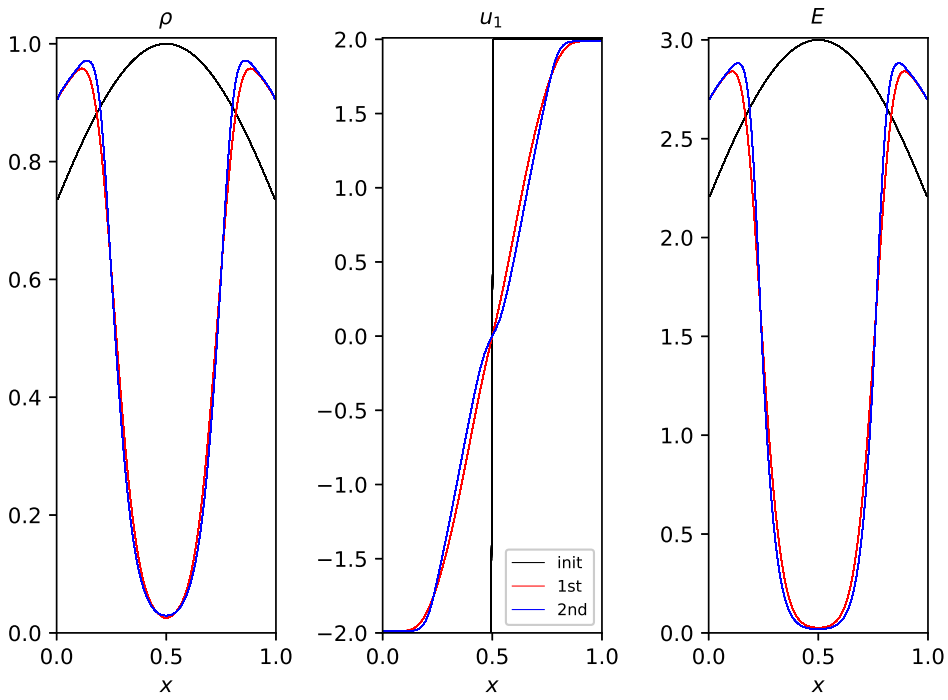


Figure 3.6: Density, velocity u_1 and energy for the strong rarefaction test based on an isothermal atmosphere at $T_f = 0.1$.

N	ρ		ρu_1		ρu_2		E	
50	3.223E-04	—	6.438E-03	—	6.438E-03	—	1.290E-01	—
100	8.604E-05	1.905	1.716E-03	1.907	1.716E-03	1.907	3.444E-02	1.905
150	3.983E-05	1.899	7.936E-04	1.901	7.936E-04	1.901	1.594E-02	1.899
200	2.298E-05	1.910	4.573E-04	1.915	4.573E-04	1.915	9.206E-03	1.909
250	1.497E-05	1.919	2.975E-04	1.926	2.975E-04	1.926	6.000E-03	1.919
300	1.058E-05	1.904	2.099E-04	1.912	2.099E-04	1.912	4.244E-03	1.898
350	7.885E-06	1.909	1.562E-04	1.919	1.562E-04	1.919	3.166E-03	1.901
400	6.095E-06	1.928	1.205E-04	1.937	1.205E-04	1.937	2.453E-03	1.912

Table 3.5: L^1 error and convergence rates with respect to the exact solution (3.84) at $T_f = 0.01$.

N	ρ		ρu_1		ρu_2		ρu_3		E	
25	5.209E-06	—	2.031E-06	—	2.031E-06	—	2.031E-06	—	1.215E-05	—
50	1.739E-06	1.582	5.346E-07	1.926	5.346E-07	1.926	5.346E-07	1.926	4.205E-06	1.531
75	8.507E-07	1.763	2.379E-07	1.996	2.379E-07	1.996	2.379E-07	1.996	2.077E-06	1.739
100	5.013E-07	1.838	1.332E-07	2.014	1.332E-07	2.014	1.332E-07	2.014	1.228E-06	1.824
125	3.295E-07	1.879	8.478E-08	2.026	8.478E-08	2.026	8.478E-08	2.026	8.093E-07	1.871
150	2.330E-07	1.900	5.871E-08	2.015	5.871E-08	2.015	5.871E-08	2.015	5.730E-07	1.893
175	1.732E-07	1.923	4.291E-08	2.033	4.291E-08	2.033	4.291E-08	2.033	4.262E-07	1.919
200	1.339E-07	1.929	3.280E-08	2.015	3.280E-08	2.015	3.280E-08	2.015	3.297E-07	1.925

Table 3.6: L^1 error and convergence rates with respect to the exact solution (3.83) at $T_f = 0.01$.

Chapter 4

An all-speed scheme for the Euler equations

In this chapter, we consider the non-dimensional Euler equations in d-space dimensions which were derived from the homogeneous Euler equations (1.16) in Section (1.4) by a scaling argument. They are given by the following set of equations

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{u}) &= 0, \\ \partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u} + \frac{p}{M^2} \mathbb{I}) &= 0, \\ \partial_t E + \nabla \cdot (\mathbf{u}(E+p)) &= 0. \end{aligned} \tag{4.1}$$

where the total energy is given by

$$E = \rho \left(e + M^2 \frac{1}{2} |\mathbf{u}|^2 \right) \tag{4.2}$$

and M is the reference Mach number as defined in (1.65) which gives the ratio between the velocity of the gas and the sound speed. Depending on the magnitude of the Mach number the characteristic nature of the flow changes. This renders the numerical simulation of these flows very challenging but also yields a very interesting research subject with a wide range of applications, for example in astrophysical stellar evolution or multi-material flows [60, 2]. For large Mach numbers, the flow is governed by compressible effects, whereas in the low Mach limit the solution of the Euler equations (4.1) are close to solutions of the incompressible Euler equations (1.69). This behaviour was studied for example in [47, 26, 70]. We refer to [34] for a study of the full Euler equations.

Standard schemes designed for compressible flows like the Roe scheme [66] or standard Godunov type schemes fail due to excessive diffusion when applied in the low Mach regime. A lot of work is dedicated to cure this defect, see for instance [40, 83, 26, 60]. To ensure accurate solutions in the low Mach regime is the development of asymptotic preserving (AP) schemes which are consistent with the limit behaviour of the Euler equation as M tends to zero, see for example [23, 31, 9] and references therein.

Due to the hyperbolic nature of (4.1) the time step for an explicit scheme is restricted by a stability CFL condition that depends on the inverse of the fastest wave speed. In the case of (4.1), the acoustic wave speeds tend to infinity as M tends to zero which

leads to very small time steps to guarantee the stability of an explicit scheme. As a side effect all waves will be resolved by the explicit scheme although the fast acoustic waves are often not of interest in application. Implicit methods on the other hand allow larger time steps but introduce diffusion on the slow wave which leads to a loss of resolution. In addition at each iteration a non-linear often ill conditioned algebraic system has to be solved. Implicit-explicit (IMEX) methods try to overcome those disadvantages by treating the stiff parts implicitly and thus allow for a Mach number independent time step. Many of those schemes are based on a splitting of the pressure in the spirit of Klein [48] in a slow compressible and fast acoustic component, since the stiffness of the system is closely related with the pressure, see for example [23, 62, 48]. As the IMEX scheme presented in [62] is only weakly asymptotic preserving as it suffers of a lack of stability for small Mach numbers, the authors recently presented an AP IMEX scheme applicable in the low Mach regime of the full Euler equations [87].

A way to avoid solving non-linear implicit systems is using a relaxation approach. The idea behind using relaxation is to modify the flux function in the equations which makes it easier to solve them implicitly as done in [1] through Jin Xin relaxation [43] or in [7] by applying a Suliciu relaxation. The linear degenerate structure of models using a Suliciu relaxation approach, as described in Section 1.3.2, allows to use accurate Godunov-type finite volume methods due to the knowledge about the exact Riemann solution. This approach was used in [12] to construct an explicit AP relaxation scheme for the barotropic Euler equations. The scheme was very recently extended in [13] to a first order AP IMEX relaxation scheme for the full Euler equations. We want to stress that the Suliciu relaxation model and the IMEX relaxation scheme described in [13] differ in important aspects from those presented in this chapter and were developed independently.

The relaxation model we use here is based on a pressure splitting and results from relaxing the slow pressure component in a Suliciu relaxation manner whereas the fast pressure part is relaxed by coupling it with an additional fast velocity obtained by relaxing the momentum equation. The flux is then split such that only two terms concerning relaxation variables are treated implicitly which leads to solving only one linear scalar equation implicitly. We do not split the flux terms of the original Euler equations but treat the whole flux in the explicit step. This leads to a conservative explicit part for which we use a Godunov-type scheme. The use of a Riemann solver allows us to prove with little effort the positivity preserving property of the scheme which is important in physical applications. In addition the way the flux is split leads to an asymptotic preserving scheme.

To have a relevant scheme for applications, an extension to second or higher order is necessary. Higher order schemes in time can be achieved by using IMEX Runge Kutta (RK) methods as for example in [9, 30]. As standard in finite volume schemes, higher order accuracy in space is achieved by reconstructing the cell interface values using WENO schemes, as for example in [72, 24]. Here, we use a MUSCL approach as described in Section 2.2 to achieve a second order extension which maintains the positivity of density and internal energy from the first order scheme.

The results presented in this chapter are already published in [79]. The chapter is organized as follows. In the next section we describe the relaxation model for the full Euler equations that is used to derive the IMEX scheme. Section 4.2 is dedicated to the structure of the first order time semi-discrete scheme for which in the subsequent section

the asymptotic preserving property is proven. It is followed by the derivation of the fully discrete scheme which is given in Section 4.4. Thereby mainly the construction of the Riemann solver for the Godunov type scheme used in the explicit part is addressed. In addition the main properties of the resulting IMEX scheme are proven. Especially, the diffusion introduced by the Riemann solver is studied which is provably independent of the Mach number. In Section 4.5 we give a second order extension for the first order scheme based on the MUSCL approach. It is followed by a section of numerical results to validate the main properties given in Sections 4.4 and 4.5.

4.1 Suliciu Relaxation model

To simplify the non-linear structure of the Euler equations (4.1) we make use of the Suliciu relaxation approach that is described for example in [74, 11, 22] and references therein. Compared to the relaxation studied by Jin and Xin in [43], the original system of equations can be found almost unchanged as a part of the relaxation model. Whereas in the Jin Xin relaxation approach a linearisation of the flux is achieved by relaxing every component of the flux function, in the Suliciu type relaxation only a few terms of the flux are relaxed in a fashion that is tailored to the problem. This leads to a reduced diffusion introduced by the relaxation process compared to the Jin Xin relaxation.

Following the usual Suliciu relaxation procedure, as described in Section 1.3.2, the key element is the relaxation of the whole pressure by introducing a new variable π, where its evolution is given by

$$\partial_t(\rho\pi) + \nabla \cdot (\rho\pi\mathbf{u} + a^2\mathbf{u}) = \frac{\rho}{\varepsilon}(p - \pi). \tag{4.3}$$

Here ε denotes the relaxation time and $a > 0$ is the constant relaxation parameter. To guarantee a stable diffusive approximation of the original Euler equations the relaxation parameter must meet the subcharacteristic condition $a > \rho c$. We want to profit from the properties of the Suliciu relaxation also for the non-dimensional Euler equations. In the following, we describe the relaxation model that was first introduced in [7]. Following [50], in a first step the pressure is decomposed into a slow dynamics and a fast acoustic component given by

$$\frac{p}{M^2} = p + \frac{1 - M^2}{M^2} p.$$

Approximating the slow and the fast pressure in the momentum equation in (4.1) by the variables π and ψ respectively, the momentum equation becomes

$$\partial_t(\rho\mathbf{u}) + \nabla \cdot \left(\rho\mathbf{u} \otimes \mathbf{u} + \pi + \frac{1 - M^2}{M^2}\psi \right) = 0$$

The evolution of the new variables π and ψ are then developed in the spirit of a Suliciu relaxation approach. The evolution for π is given by the Suliciu relaxation equation (4.3). However, applying the standard Suliciu relaxation method also on the pressure ψ, would only lead to non relevant diffusion terms and not to a low Mach scheme as shown in [7]. To overcome this, the authors of [7] introduced a new velocity variable $\hat{\mathbf{u}} \in \mathbb{R}^d$ which is relaxed to the gas velocity $\mathbf{u}$ and couples with the pressure ψ. The form of the evolution equations for $\hat{\mathbf{u}}$ and ψ is chosen, such that

- the resulting model is in conservation form
- the resulting model has ordered eigenvalues, which results in a clear wave structure
- the resulting model is a stable diffusive approximation of the non-dimensional Euler equations (4.1)
- the resulting numerical scheme has a Mach number independent diffusion.

Considering the above points leads to the following relaxation model

$$\begin{aligned}
\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) &= 0, \\
\partial_t (\rho \mathbf{u}) + \nabla \cdot \left(\rho \mathbf{u} \otimes \mathbf{u} + \pi \mathbb{I} + \frac{1 - M^2}{M^2} \psi \mathbb{I} \right) &= 0, \\
\partial_t E + \nabla \cdot \left(\mathbf{u}(E + M^2 \pi + (1 - M^2)\psi) \right) &= 0, \\
\partial_t (\rho \pi) + \nabla \cdot (\rho \pi \mathbf{u} + a^2 \mathbf{u}) &= \frac{\rho}{\varepsilon}(p - \pi), \\
\partial_t (\rho \hat{\mathbf{u}}) + \nabla \cdot \left(\rho \mathbf{u} \otimes \hat{\mathbf{u}} + \frac{1}{M^2} \psi \mathbb{I} \right) &= \frac{\rho}{\varepsilon}(\mathbf{u} - \hat{\mathbf{u}}), \\
\partial_t (\rho \psi) + \nabla \cdot (\rho \psi \mathbf{u} + a^2 \hat{\mathbf{u}}) &= \frac{\rho}{\varepsilon}(p - \psi).
\end{aligned} \tag{4.4}$$

The relaxation model (4.4) differs from the one given in [7] in the following points:

1. In the relaxation equation for $\hat{\mathbf{u}}$, we use $1/M^2$ instead of $1/M^4$ as proposed by the authors in [7]. This is due to the upwind discretization used in [7] which requires $1/M^4$ in order to have a Mach number independent diffusion of the numerical scheme. Here instead we use centred differences in the implicit part to ensure the Mach number independent diffusion of the numerical scheme and therefore $1/M^2$ suffices.

2. We have simplified the model in the sense that we do not distinguish between the given reference Mach number M and the local Mach number M_{loc}. This is not a restriction in practice, because the choice of M is given by the application as illustrated in the numerical results. Especially for the Mach number dependent shock test case in Section 7.1.2 we directly compare our results to a scheme that uses the local Mach number M_{loc}.

The following lemma sums up some properties of system (4.4). The proof including the Chapman-Enskog stability analysis can be done by adapting the steps explained in Sections 1.2.2 and 1.3. Alternatively, the proof can be established analogously to [7].

Lemma 4.1. *The relaxation system (4.4) is hyperbolic and is a stable diffusive approximation of (4.1) under the Mach number independent subcharacteristic condition $a > \rho c$. Considering the x_i-direction, the relaxation system (4.4) has the following linearly degenerate eigenvalues*

$$\lambda^u = u_i, \ \lambda^{\pm} = u_i \pm \frac{a}{\rho}, \ \lambda^{\pm}_M = u_i \pm \frac{a}{M\rho},$$

where λ^u has multiplicity $2d$, where d is the number of considered space dimensions. For $M < 1$, the eigenvalues have the ordering

$$\lambda^-_M < \lambda^- < \lambda^u < \lambda^+ < \lambda^+_M.$$

In the case of $M = 1$ the waves given by $\lambda^{\pm}$ and $\lambda_M^{\pm}$ collapse to the waves $\lambda^{\pm}$ which have then multiplicity 2 respectively. Lemma 4.1 also states that the eigenvalues depend on the relaxation parameter a and therefore the diffusion of the resulting scheme depends on a. In general, a larger a leads to a more diffusive numerical solution whereas if a is chosen too small it can give rise to instabilities.

To shorten notations we will refer to the original system (4.1) by

$$\partial_t w + \nabla \cdot f(w) = 0, \tag{4.5}$$

where $w = (\rho, \rho\mathbf{u}, E)^T$ denotes the physical variables and the flux function is given by

$$f(w) = \begin{pmatrix} \rho\mathbf{u} \\ \rho\mathbf{u} \otimes \mathbf{u} + \frac{p}{M^2}\mathbb{I} \\ \mathbf{u}(E+p) \end{pmatrix}.$$

The relaxation model (4.4) is given by

$$\partial_t W + \nabla \cdot \widetilde{F}(W) = \frac{1}{\varepsilon} R(W), \tag{4.6}$$

where

$$W = \begin{pmatrix} \rho \\ \rho\mathbf{u} \\ E \\ \rho\pi \\ \rho\hat{\mathbf{u}} \\ \rho\psi \end{pmatrix}, \quad \widetilde{F}(W) = \begin{pmatrix} \rho\mathbf{u} \\ \rho\mathbf{u} \otimes \mathbf{u} + \pi\mathbb{I} + \frac{1-M^2}{M^2}\psi\mathbb{I} \\ \mathbf{u}(E + M^2\pi + (1-M^2)\psi) \\ \rho\pi\mathbf{u} + a^2\mathbf{u} \\ \rho\mathbf{u} \otimes \hat{\mathbf{u}} + \frac{1}{M^2}\psi\mathbb{I} \\ \rho\psi\mathbf{u} + a^2\hat{\mathbf{u}} \end{pmatrix}, \quad R(W) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \rho(p-\pi) \\ \rho(\mathbf{u}-\hat{\mathbf{u}}) \\ \rho(p-\psi) \end{pmatrix}. \tag{4.7}$$

Since ε is the relaxation time, it is indicating how fast the perturbed system (4.4) is reaching its equilibrium (4.1). The relaxation equilibrium state is defined as

$$W^{\text{eq}} = \mathcal{M}(w) = (\rho, \rho\mathbf{u}, E, \rho p(\rho, e), \rho\mathbf{u}, \rho p(\rho, e))^T. \tag{4.8}$$

The connection between (4.5) and (4.6) can be established through the matrix $Q \in \mathbb{R}^{(2+d)\times 2(2+d)}$ defined by

$$Q = \begin{pmatrix} \mathbb{I}_{2+d} & 0_{2+d} \end{pmatrix} \tag{4.9}$$

where d denotes the number of space dimensions. For all equilibrium states $\mathcal{M}(w)$, we have $R(\mathcal{M}(w)) = 0$ and the physical variables are then recovered by $w = Q\mathcal{M}(w)$ and the fluxes are connected by $f(w) = Q\widetilde{F}(\mathcal{M}(w))$.

4.2 Time semi-discrete scheme

As we have seen in Lemma 4.1, the largest absolute eigenvalue $|\lambda_M^{\pm}|$ of the relaxation model (4.4) tends to infinity as M goes to 0. Using a time explicit scheme results in a very restrictive CFL condition (2.83) that tends to 0 as $M \to 0$. By using an IMEX approach as done for example in [64, 9], we can avoid the Mach number dependence of the time step.

We rewrite the relaxation system (4.4)/(4.6) in the following form:

$$\partial_t W + \nabla \cdot F(W) + \frac{1}{M^2}\nabla \cdot G(W) = \frac{1}{\varepsilon}R(W). \tag{4.10}$$

In (4.10), we have split the flux $\widetilde{F}$ in (4.6) into a flux function F which will contain the explicit terms and a flux function G which will contain the terms treated implicitly. For efficiency, we want to have as many explicit terms as possible as long as the eigenvalues of F are independent of the Mach number. To avoid especially inverting a large non-linear system, we treat the non-linear advection terms explicitly. This results in the following flux functions

$$F(W) = \begin{pmatrix} \rho\mathbf{u} \\ \rho\mathbf{u}\otimes\mathbf{u} + \pi\mathbb{I} + \frac{1-M^2}{M^2}\psi\mathbb{I} \\ \mathbf{u}\left(E + M^2\pi + (1-M^2)\psi\right) \\ \rho\pi\mathbf{u} + a^2\mathbf{u} \\ \rho\mathbf{u}\otimes\hat{\mathbf{u}} \\ \rho\psi\mathbf{u} \end{pmatrix} \text{ and } G(W) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \psi\mathbb{I} \\ a^2M^2\ \hat{\mathbf{u}} \end{pmatrix}. \tag{4.11}$$

We see, that F contains in equilibrium $\mathcal{M}(w)$ the flux f of the Euler equations (4.1) whereas G only acts on the relaxation variables $\hat{\mathbf{u}}, \psi$. To obtain a time semi-discrete scheme we perform the implicit and explicit steps in the following order

$$\text{Implicit: } \partial_t W + \frac{1}{M^2}\nabla \cdot G(W) = 0, \tag{4.12}$$

$$\text{Explicit: } \quad \partial_t W + \nabla \cdot F(W) = 0, \tag{4.13}$$

$$\text{Projection: } \quad \partial_t W = \frac{1}{\varepsilon}R(W). \tag{4.14}$$

The relaxation source term in (4.14) is solved by projecting the variables onto the equilibrium manifold corresponding to $\varepsilon = 0$ and thereby reaching the relaxation equilibrium state $\mathcal{M}(w)$ defined in (4.8). As the relaxation source term in (4.14) is stiff for small ε, it is treated implicitly for a time step Δt and then the limit $\varepsilon \to 0$ is taken. Following the steps (2.32) and 2.33, we find

$$\pi^{n+1} = p^{(2)}, \quad \psi = p^{(2)}, \quad \hat{\mathbf{u}}^{n+1} = \mathbf{u}^{(2)}, \tag{4.15}$$

where the superscript (2) denotes the solution after the second step (4.13). The projection step can be then summarized as $W^{n+1} = \mathcal{M}(w^{(2)}) = W^{(2),\text{eq}}$. The formal time semi-discrete scheme for the relaxation model (4.4) is then given by

$$W^{(1)} - W^{n,\text{eq}} + \frac{\Delta t}{M^2}\nabla \cdot G\left(W^{(1)}\right) = 0, \tag{4.16}$$

$$W^{(2)} - W^{(1)} + \Delta t \nabla \cdot F\left(W^{(1)}\right) = 0, \tag{4.17}$$

$$W^{n+1} = W^{(2),\text{eq}}, \tag{4.18}$$

where we consider the data at time t^n to be at relaxation equilibrium $W^{n,\text{eq}}$. First we solve the implicit equation (4.38) to gain $W^{(1)}$, followed by the explicit step where we calculate $W^{(2)}$. The final projection step ensures that the initial condition at the subsequent time step is in relaxation equilibrium W^{eq}. The procedure given by (4.16) - (4.18) results in a first order scheme in time.

4.3 Asymptotic properties

We consider as continuous limit equations the incompressible Euler equations given by

$$\begin{aligned} \rho &= \text{const.} \\ \partial_t \mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} + \nabla P &= 0 \\ \nabla\cdot\mathbf{u} &= 0 \end{aligned} \tag{4.19}$$

with a dynamical pressure described by P. As we have seen in Section 1.4, the limit equations can be formally derived using a Mach number expansion of the variables $\rho, \mathbf{u}, p$ given by

$$\begin{aligned} \rho(\mathbf{x},t) &= \rho_0(\mathbf{x},t) + \mathcal{O}(M), \\ \mathbf{u}(\mathbf{x},t) &= \mathbf{u}_0(\mathbf{x},t) + \mathcal{O}(M), \\ p(\mathbf{x},t) &= p_0(\mathbf{x},t) + Mp_1(\mathbf{x},t) + M^2 p_2(\mathbf{x},t) + \mathcal{O}(M^3), \end{aligned} \tag{4.20}$$

and considering slipping or periodic boundary conditions, see for example [40, 26]. This formal limit derivation leads to the following set of well-prepared data

$$\Omega_{wp} = \{w \in \Omega_{phy}, \nabla\rho_0 = 0, \nabla p_0 = 0, \nabla\cdot\mathbf{u}_0 = 0\} \tag{4.21}$$

which can be written in terms of the Mach number expansion as

$$\rho = \rho_0 + \mathcal{O}(M), \qquad \rho_0 = \text{const}, \tag{4.22}$$

$$\mathbf{u} = \mathbf{u}_0 + \mathcal{O}(M), \qquad \nabla\cdot\mathbf{u}_0 = 0, \tag{4.23}$$

$$p = p_0 + \mathcal{O}(M^2), \qquad p_0 = \text{const}. \tag{4.24}$$

For simplicity we show the AP property for the time semi-discrete scheme. The same steps can be followed with the fully discretized scheme that will be given in Section 4.4.

To show the AP property, we will exploit some properties of the fast pressure $\psi^{(1)}$ obtained in the implicit step (4.25).

4.3.1 Asymptotic behaviour of $\psi^{(1)}$

Due to the sparse structure of G defined in (4.11), the implicit part reduces to solving only two coupled equations in the relaxation variables $\hat{\mathbf{u}}, \psi$ given by

$$\begin{aligned} \partial_t(\rho\hat{\mathbf{u}}) + \frac{1}{M^2}\nabla\psi &= 0, \\ \partial_t(\rho\psi) + a^2\nabla\cdot\hat{\mathbf{u}} &= 0, \end{aligned} \tag{4.25}$$

with the eigenvalues $\tilde{\lambda}_M^{\pm} = \pm\frac{a}{\rho M}$. As done in [23], we rewrite the coupled system (4.25) into one linear equation for ψ starting from the time-semi-discrete scheme

$$\frac{\rho^{(1)} - \rho^n}{\Delta t} = 0, \tag{4.26}$$

$$\frac{(\rho\hat{\mathbf{u}})^{(1)} - (\rho\hat{\mathbf{u}})^n}{\Delta t} + \frac{1}{M^2}\nabla\psi^{(1)} = 0, \tag{4.27}$$

$$\frac{(\rho\psi)^{(1)} - (\rho\psi)^n}{\Delta t} + a^2\nabla\cdot\hat{\mathbf{u}}^{(1)} = 0. \tag{4.28}$$

To emphasize that (4.25) also depends on the density, we have included the density update (4.26) into the time semi-discrete system. From equation (4.26) we see that $\rho^{(1)} = \rho^n$. To simplify notation we define $\tau^n = 1/\rho^n$. Inserting (4.27) into (4.28) we can reduce the implicit system to only one equation with an elliptic operator for ψ given as

$$\psi^{(1)} - \frac{\Delta t^2 a^2}{M^2}\tau^n \nabla \cdot (\tau^n \nabla \psi^{(1)}) = \psi^n - \Delta t a^2 \tau^n \nabla \cdot \mathbf{u}^n. \tag{4.29}$$

On the right hand side of (4.29) we have already made use of the fact that $\hat{\mathbf{u}}^n = \mathbf{u}^n$ since we start from equilibrium data. We will see that the correct scaling of $\psi^{(1)}$ with respect to the Mach number is important not only for showing the AP property of the scheme, but also for the positivity of density and internal energy as well as for a Mach number independent diffusion of the fully discretized scheme.

To prevent $\mathcal{O}(M)$ pressure perturbations at the boundaries which would destroy the well-prepared nature of the pressure, we require boundary conditions on ψ which preserve the scaling of the pressure in time. For a computational domain D, we set

$$\left.\begin{aligned} \psi_0^{(1)} &= p_0^n \\ \psi_1^{(1)} &= 0 \end{aligned}\right\} \text{ on } \partial D. \tag{4.30}$$

The following lemma states that the fast pressure $\psi^{(1)}$ after the implicit step is still well-prepared using the boundary condition (4.30).

Lemma 4.2 (Scaling of $\psi^{(1)}$)**.** *Let $w^n \in \Omega_{wp}$ be equipped with the boundary conditions (4.30). Then the Mach number expansion of $\psi^{(1)}$ after the first stage satisfies*

$$\psi^{(1)} = p_0^n + M^2 \psi_2^{(1)} + \mathcal{O}(M^3),$$

where p_0^n is constant.

Proof. Let us assume that the expansion of $\psi^{(1)}$ is given by

$$\psi^{(1)} = \psi_0^{(1)} + M\psi_1^{(1)} + M^2\psi_2^{(1)} + \mathcal{O}(M^3). \tag{4.31}$$

Since $w^n \in \Omega_{wp}$, the data at t^n is well-prepared as defined in (4.22),(4.23) and (4.24). We insert the therein given Mach number expansion of w^n and the Mach number expansion of $\psi^{(1)}$ given by (4.31) into the implicit update for ψ (4.29). Separating the $\mathcal{O}(M^{-2})$ terms, we find

$$\begin{cases} \Delta\psi_0^{(1)} &= 0 \text{ in } D \\ \psi_0^{(1)} &= p_0^n \text{ on } \partial D \end{cases}$$

with the boundary condition for $\psi_0^{(1)}$ given in (4.30). This leads to $\psi_0^{(1)} = p_0^n$ on $\overline{D}$. Separating the $\mathcal{O}(M^{-1})$ terms and using that $\psi_0^{(1)} = p_0^n = \text{const}$, we find

$$\begin{cases} \Delta\psi_1^{(1)} &= 0 \text{ in } D \\ \psi_1^{(1)} &= 0 \text{ on } \partial D \end{cases}$$

which leads to $\psi_1^{(1)} = 0$ on $\overline{D}$. As a last step, we collect the $\mathcal{O}(M^0)$ terms and use that $\psi_0^{(1)} = p_0^n$ as well as $\psi_1^{(1)} = 0$ on $\overline{D}$. It is not necessary to impose special boundary conditions for $\psi_2^{(1)}$. Thus we find

$$\Delta \psi_2^{(1)} = 0 \text{ in } D, \tag{4.32}$$

from which follows the stated Mach number expansion for $\psi^{(1)}$ and the proof is completed. □

4.3.2 Asymptotic preserving property

We show that the time discretization of (4.4) in the $M \to 0$ limit coincides with a time discretization of the incompressible Euler equations (4.19). We consider well-prepared data $w^n \in \Omega_{wp}$ and the Mach number expansion of $\psi^{(1)}$ from Lemma 4.2. For the total energy defined in (4.2), we find the following Mach number expansion

$$E = \rho_0 e_0 + M(\rho_1 e_0 + \rho_0 e_1) + M^2 \left(\frac{1}{2} |\mathbf{u}_0|^2 + \rho_2 e_0 + \rho_1 e_1 + \rho_0 e_2 \right) + \mathcal{O}(M^3).$$

Inserting the Mach number expansions of $w^n, \psi^{(1)}$ and E^n into the density, momentum and energy equation of (4.17) and considering the $\mathcal{O}(M^0)$ order terms we have

$$\begin{aligned} \frac{\rho_0^{n+1} - \rho_0^n}{\Delta t} + \nabla \cdot \rho_0^n \mathbf{u}_0^n &= 0, \\ \frac{\rho_0^{n+1}\mathbf{u}_0^{n+1} - \rho_0^n \mathbf{u}_0^n}{\Delta t} + \nabla \cdot (\rho_0^n \mathbf{u}_0^n \otimes \mathbf{u}_0^n) + \nabla \psi_2^{(1)} &= 0, \\ \frac{\rho_0^{n+1} e_0^{n+1} - \rho_0^n e_0^n}{\Delta t} + \nabla \cdot \mathbf{u}_0^n \left(\rho_0^n e_0^n + \psi_0^{(1)} \right) &= 0. \end{aligned} \tag{4.33}$$

Let us assume that the pressure at time t^{n+1} has the following Mach number expansion $p^{n+1} = p_0^{n+1} + M p_1^{n+1} + \mathcal{O}(M^2)$. Note that we have from Lemma 4.2 that $\psi_0^{(1)} = p_0$ and thus $\nabla \psi_0^{(1)} = \nabla p_0 = 0$. Since $w^n \in \Omega_{wp}$ we have $\nabla \rho_0^n = 0$ and $\nabla \cdot \mathbf{u}_0^n = 0$. Further we use $p_0 = (\gamma - 1)\rho_0 e_0$. Equipped with that we can simplify (4.33) to

$$\frac{\rho_0^{n+1} - \rho_0^n}{\Delta t} = 0, \tag{4.34}$$

$$\frac{\mathbf{u}_0^{n+1} - \mathbf{u}_0^n}{\Delta t} + \mathbf{u}_0^n \cdot \nabla \mathbf{u}_0^n + \frac{\nabla \psi_2^{(1)}}{\rho_0^n} = 0, \tag{4.35}$$

$$\frac{p_0^{n+1} - p_0^n}{\Delta t} = 0. \tag{4.36}$$

Especially from equations (4.34) and (4.36) we see that ρ_0^{n+1} and p_0^{n+1} are constants. Looking at the $\mathcal{O}(M^1)$ terms we have from the energy equation

$$p_1^{n+1} + \Delta t \nabla \cdot \mathbf{u}_1^n = 0.$$

This means the density and pressure at time t^{n+1} are well-prepared up to $\mathcal{O}(\Delta t)$ perturbation as in (4.22), (4.24). To be consistent with a time discretization of the incompressible

Euler equations (4.19) the divergence of $\mathbf{u}_0$ at time t^{n+1} defined by $\nabla \cdot \mathbf{u}_0^{n+1}$ has to be at least of order Δt. To show this, we apply the divergence on the velocity update (4.35) which gives

$$\nabla \cdot \mathbf{u}_0^{n+1} = \nabla \cdot \mathbf{u}_0^n + \Delta t \nabla \cdot \left(\mathbf{u}_0^n \cdot \nabla \mathbf{u}_0^n + \frac{\Delta \psi_2^{(1)}}{\rho_0^n} \right). \tag{4.37}$$

In the proof of Lemma 4.2, we have shown that $\Delta \psi_2^{(1)} = 0$ on ∂D, see (4.32). Using (4.32) together with $\nabla \cdot \mathbf{u}_0^n = 0$, we can simplify (4.37) to

$$\nabla \cdot \mathbf{u}_0^{n+1} = \Delta t \nabla \cdot (\mathbf{u}_0^n \cdot \nabla \mathbf{u}_0^n) = \mathcal{O}(\Delta t).$$

In summary, we have shown the following theorem.

Theorem 4.3 (AP property)**.** *Let $w^n \in \Omega_{wp}$. Then under the boundary conditions (4.30) the scheme (4.16), (4.17), (4.18) is asymptotic preserving when M tends to 0, in the sense that if $w^n \in \Omega_{wp}$ then it is $w^{n+1} \in \Omega_{wp}$ up to $\mathcal{O}(\Delta t)$ and in the limit $M \to 0$ the time-semi-discrete scheme is a consistent discretization of the incompressible Euler equations (4.19).*

4.4 Derivation of the fully discrete scheme

For simplicity, we develop the fully discretized scheme in one space dimension, but it can be straightforwardly extended to d dimensions. In the implicit update (4.29), the space derivatives read

$$\nabla \cdot (\tau \nabla \psi) = \partial_{x_1}(\tau \partial_{x_1} \psi) + \cdots + \partial_{x_d}(\tau \partial_{x_d} \psi) \text{ and } \nabla \cdot \mathbf{u} = \partial_{x_1} u_1 + \cdots + \partial_{x_d} u_d$$

for $\mathbf{u} = (u_1, \ldots, u_d)$ and in the explicit part we can obtain a multi-dimensional scheme by considering dimensional split Riemann problems as discussed in Section 2.1.4.

In the following we use a Cartesian grid on a computational domain D devided in N cells $C_i = (x_{i-1/2}, x_{i+1/2})$ of step size Δx. We use a standard finite volume setting, where we define at time t^n the piecewise constant functions

$$w(x, t^n) = w_i^n, \text{ for } x \in C_i.$$

Using this notation, we apply centred differences on the implicit update (4.29) and obtain

$$\psi_i^{(1)} - \frac{\Delta t^2}{\Delta x^2} \frac{a^2}{M^2} \tau_i^n \left(\tau_{i-1/2}^n \psi_{i-1}^{(1)} - \left(\tau_{i-1/2}^n + \tau_{i+1/2}^n \right) \psi_i^{(1)} + \tau_{i+1/2}^n \psi_{i+1}^{(1)} \right) = \psi_i^n - \frac{\Delta t}{2\Delta x} a^2 \tau_i^n \left(u_{i+1}^n - u_{i-1}^n \right), \tag{4.38}$$

where $\tau_{i+1/2} = (\tau_{i+1} + \tau_i)/2$. Due to the centred differences, the update is second order accurate in space.

For the explicit part, we will use a Godunov type finite volume scheme following [41] which we will describe in the following section.

4.4.1 Godunov type finite volume scheme

In the explicit step we consider the following equations as defined in (4.11), (4.17)

$$\begin{aligned}
\partial_t \rho + \partial_x \rho u &= 0,\\
\partial_t(\rho u) + \partial_x\left(\rho u^2 + \pi + \frac{1-M^2}{M^2}\psi\right) &= 0,\\
\partial_t E + \partial_x\Big(\big(E + M^2\pi + (1-M^2)\psi\big)u\Big) &= 0,\\
\partial_t(\rho\pi) + \partial_x\left(\rho\pi u + a^2 u\right) &= 0,\\
\partial_t(\rho\hat{u}) + \partial_x(\rho\hat{u}u) &= 0,\\
\partial_t(\rho\psi) + \partial_x(\rho\psi u) &= 0,
\end{aligned} \tag{4.39}$$

which have the following properties:

Lemma 4.4. *System (4.39) admits the ordered linear degenerate eigenvalues*

$$\lambda^- = u - \frac{a}{\rho} < \lambda^u = u < \lambda^+ = u + \frac{a}{\rho},$$

where the eigenvalue λ^u has multiplicity 4. The relaxation parameter a, as well as all eigenvalues, are independent of the Mach number .

The Riemann invariants with respect to λ^u are

$$I_1^u = u,\ I_2^u = M^2\pi + (1-M^2)\psi \tag{4.40}$$

and with respect to $\lambda^\pm$

$$\begin{aligned}
&I_1^\pm = u \pm \frac{a}{\rho},\ I_2^\pm = \pi \mp au,\\
&I_3^\pm = e - \frac{M^2}{2a^2}\pi^2 - \frac{1-M^2}{a^2}\pi\psi,\\
&I_4^\pm = \hat{u}, \quad I_5^\pm = \psi.
\end{aligned} \tag{4.41}$$

Proof. We rewrite the equations (4.39) using primitive variables $\mathcal{V} = (\rho, u, e, \pi, \hat{u}, \psi)$ in quasi-linear form

$$\partial_t \mathcal{V} + \mathcal{B}(\mathcal{V})\partial_x \mathcal{V} = 0, \tag{4.42}$$

where the matrix $\mathcal{B}(\mathcal{V})$ is given by

$$\mathcal{B}(\mathcal{V}) = \begin{pmatrix}
u & \rho & 0 & 0 & 0 & 0\\
0 & u & 0 & \frac{1}{\rho} & 0 & \frac{1-M^2}{M^2}\\
0 & \frac{M^2\pi + (1-M^2)\psi}{\rho} & u & 0 & 0 & 0\\
0 & \frac{a^2}{\rho} & 0 & u & 0 & 0\\
0 & 0 & 0 & 0 & u & 0\\
0 & 0 & 0 & 0 & 0 & u
\end{pmatrix}.$$

It is easy to check that $\lambda^u, \lambda^\pm$ are eigenvalues of $\mathcal{B}(\mathcal{V})$. Associated to these eigenvalues, we find for $M > 0$ the respective linearly independent eigenvectors

$$r_1^u = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 - \frac{1}{M^2} \\ 0 \\ 1 \end{pmatrix}, \; r_2^u = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \; r_3^u = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; r_4^u = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; r^\pm = \begin{pmatrix} \frac{\rho^2}{a^2} \\ \pm\frac{1}{a} \\ \frac{M^2\pi + (1-M^2)\psi}{a^2} \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Therefore, system (4.39) is hyperbolic. Furthermore, using equation (1.30), we can compute that all characteristic fields are linear degenerate. It is straightforward to check with the help of equation (1.31), that (4.40) and (4.41) are Riemann invariants. Since Riemann invariants are invariant under change of variables, the Riemann invariants of (4.42) are the same as for the equations in conservation form (4.39). □

We remark that in the case $M = 1$, the first four equations of system (4.39) coincide with the Suliciu relaxation model for the homogeneous compressible Euler equations considered in Section 1.3.2. Especially the eigenstructure and Riemann invariants of the compressible model (1.52) are recovered in for $M = 1$.

We will follow the theory of Harten, Lax and van Leer [41] for deriving a Riemann solver $W_{\mathcal{RS}}\left(\frac{x}{t}; W_L^{(1)}, W_R^{(1)}\right)$ based on the states $W^{(1)}$ after the implicit step. Due to the linear-degeneracy of the characteristic fields from Lemma 4.4, we can solve the Riemann problem exactly. The Riemann solution, as displayed in Figure 4.1, consists of four constant states separated by contact discontinuities and is given by

$$W_{\mathcal{RS}}\left(\frac{x}{t}; W_L^{(1)}, W_R^{(1)}\right) = \begin{cases} W_L^{(1)} & \frac{x}{t} < \lambda^-, \\ W_L^* & \lambda^- < \frac{x}{t} < \lambda^u, \\ W_R^* & \lambda^u < \frac{x}{t} < \lambda^+, \\ W_R^{(1)} & \lambda^+ < \frac{x}{t}. \end{cases} \tag{4.43}$$

To compute the intermediate states $W_{L,R}^*$, we use the Riemann invariants as given in Lemma 4.4. Note that since the eigenvalues $\lambda^\pm$ have multiplicity 1, we get the expected 5 Riemann invariants. This does not hold in general for eigenvalues with higher multiplicity, see eg. [11]. Nevertheless, the invariants (4.40) and (4.41) give enough relations to determine the solution to a Riemann problem for (4.39) as shown in the following lemma.

Lemma 4.5. *Consider an initial value problem for system (4.39) with initial data given by*

$$W^0(x) = \begin{cases} W_L^{(1)} & x < 0, \\ W_R^{(1)} & x > 0. \end{cases} \tag{4.44}$$

Then the solution consists of four constant states separated by contact discontinuities and has the structure given in (4.43). Furthermore, dropping the superscript (1), an exact solution for the intermediate states $W^*_{L/R}$ is given by

$$
\begin{aligned}
u^* = u^*_L = u^*_R &= \frac{1}{2}(u_L + u_R) + \frac{1}{2a}\left((\pi_L - \pi_R) + \frac{1-M^2}{M^2}(\psi_L - \psi_R)\right),\\
\pi^*_L &= \frac{1}{2}(\pi_L + \pi_R) + \frac{a}{2}(u_L - u_R) - \frac{1-M^2}{M^2}\frac{1}{2}(\psi_L - \psi_R),\\
\pi^*_R &= \frac{1}{2}(\pi_L + \pi_R) + \frac{a}{2}(u_L - u_R) + \frac{1-M^2}{M^2}\frac{1}{2}(\psi_L - \psi_R),\\
\frac{1}{\rho^*_L} &= \frac{1}{\rho_L} + \frac{1}{a}(u^* - u_L),\\
\frac{1}{\rho^*_R} &= \frac{1}{\rho_R} + \frac{1}{a}(u_R - u^*),\\
e^*_L &= e_L - \frac{1}{2a^2}\left(\pi_L^2 - {\pi^*_L}^2 + (1-M^2)(\pi_L - \pi^*_L)\psi_L\right),\\
e^*_R &= e_R - \frac{1}{2a^2}\left(\pi_R^2 - {\pi^*_R}^2 + (1-M^2)(\pi_R - \pi^*_R)\psi_R\right),\\
\psi^*_{L,R} &= \psi_{L,R},\\
\hat{u}^*_{L,R} &= \hat{u}_{L,R}.
\end{aligned}
\tag{4.45}
$$

Proof. The solution structure follows directly from the linear degeneracy of the eigenvalues given in Lemma 4.4 and the ordering of the eigenvalues. To derive the solution for the intermediate states $W^*_{L,R}$ one uses the Riemann invariants given in (4.40) and (4.41) and solves the resulting system of equations. □

From the intermediate states (4.45) and from the equations (4.39) we see, that the relaxation variables $\psi, \hat{u}$ are only passively advected in the explicit step. To define a Godunov scheme for (4.39), we consider the local Riemann problem at the interface $x = x_{i+1/2}$ depending on the left and right cell averages $W^{(1)}_i, W^{(1)}_{i+1}$ after the implicit step. It

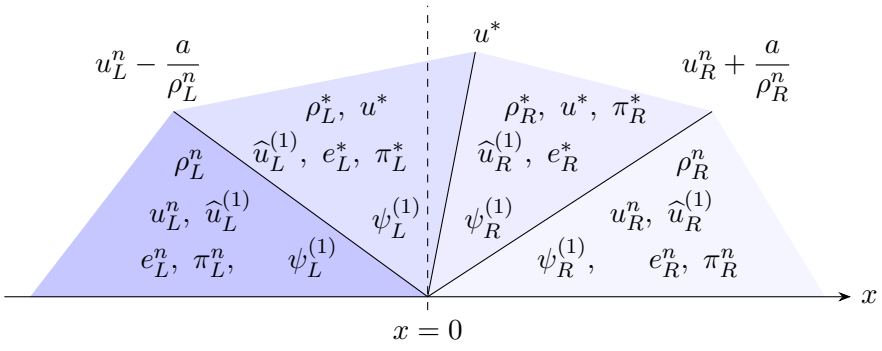


Figure 4.1: Structure of the Riemann solution of the explicit part of the relaxation model (4.4).

is given by

$$\begin{aligned} &\partial_t W + \partial_x F(W) = 0, \\ &W^0(x_i) = \begin{cases} W_i^{(1)} & \text{if} \quad x < x_{i+1/2} \\ W_{i+1}^{(1)} & \text{if} \quad x > x_{i+1/2} \end{cases}. \end{aligned} \tag{4.46}$$

Given the solution of the Riemann problem (4.43), the numerical fluxes are defined as follows

$$\mathcal{F}\left(W_i^{(1)}, W_{i+1}^{(1)}\right) = \begin{cases} F\left(W_i^{(1)}\right) & \lambda_L^- > 0 \\ F\left(W_i^{*,(1)}\right) & u^* > 0 > \lambda_L^- \\ F\left(W_{i+1}^{*,(1)}\right) & \lambda_R^+ > 0 > u^* \\ F\left(W_{i+1}^{(1)}\right) & \lambda_R^+ < 0 \end{cases}$$

where $\lambda_L^- = u_L - a/\rho_L$ and $\lambda_R^+ = u_R + a/\rho_R$. To avoid interactions between the approximate Riemann solvers at the interfaces $x_{i+1/2}$, we have a CFL restriction on the time step of

$$\frac{\Delta t}{\Delta x} \max_i \left| u_i^n \pm \frac{a}{\rho_i^n} \right| \leq \frac{1}{2} \tag{4.47}$$

which is independent of the Mach number. This leads to the following update of the explicit part

$$W_i^{(2)} = W_i^{(1)} - \frac{\Delta t}{\Delta x}\left(\mathcal{F}\left(W_i^{(1)}, W_{i+1}^{(1)}\right) - \mathcal{F}\left(W_{i-1}^{(1)}, W_i^{(1)}\right)\right). \tag{4.48}$$

The complete numerical scheme is then given by (4.38),(4.48) and (4.18). It also includes the updates of the relaxation variables in the explicit step which are redundant, since they will get updated in the subsequent projection step. Defining the numerical fluxes by

$$\mathfrak{f}\left(w_i^n, \psi_i^{(1)}, w_{i+1}^n, \psi_{i+1}^{(1)}\right) = QF\left(W_{RS}(0; W_i^{(1)}, W_{i+1}^{(1)})\right), \tag{4.49}$$

where $W_i^{(1)}, W_{i+1}^{(1)}$ are obtained from the implicit update (4.38) with initial data $W_i = \mathcal{M}(w_i)$ for $i = 1, \ldots, N$, the update of the physical variables can be written as

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(\mathfrak{f}\left(w_i^n, \psi_i^{(1)}, w_{i+1}^n, \psi_{i+1}^{(1)}\right) - \mathfrak{f}\left(w_{i-1}^n, \psi_{i-1}^{(1)}, w_i^n, \psi_i^{(1)}\right)\right). \tag{4.50}$$

We want to remark that writing the update for the physical variables in the form (4.50), it is not necessary to compute the relaxation variable $\hat{\mathbf{u}}$ at all in the numerical scheme since it does not appear in the implicit or explicit step. This reduces the computational costs of the IMEX scheme.

As we have remarked earlier, in the limit case $M = 1$, the relaxation model (4.4) reduces to a Suliciu relaxation model for the compressible Euler equations since all ψ-terms are being cancelled and the eigenvalues $\lambda^{\pm}$ and $\lambda_M^{\pm}$ collapse. Analogously, in the scheme the implicit step becomes redundant, since in the Riemann solution (4.45) the

ψ-terms are cancelled, too. Thus the scheme reduces to an explicit Godunov-type scheme based on an approximate Riemann solver for the compressible Euler equations discussed in Section (2.1.2).

For $M > 1$ the order of the eigenvalues changes to $\lambda^- < \lambda_M^- < \lambda^u < \lambda_M^+ < \lambda^+$. The scheme is still stable because the CFL condition (4.47) is determined by the largest eigenvalue in the explicit step which are now the largest ones. Since the order of the eigenvalues of the explicit part has not changed, the Riemann solver is still valid in the case of $M > 1$.

We continue now with some properties of the first order scheme.

4.4.2 Positivity of density and internal energy

The property of the scheme to preserve the domain Ω_{phy} is linked to the used Riemann solver. In our case it is essential that the density and internal energy of the Riemann solution (4.45) are positive. This is shown in the following lemma.

Lemma 4.6. *Let the initial data $W_{L,R}^{(1)}$ of the Riemann problem (4.44) be composed of $w_{L,R}^{(1)} \in \Omega_{phy} \cap \Omega_{wp}$ and $\psi^{(1)}$ satisfying the boundary conditions (4.30). Then there is a relaxation parameter a large enough but independent of M such that*

$$QW_{\mathcal{RS}}\left(\frac{x}{t}; W_L^{(1)}, W_R^{(1)}\right) \in \Omega_{phy}.$$

Proof. Since the proof only concerns data after the implicit step, we will drop the superscript (1). We have to prove, that $\rho_{L,R}^* > 0$ and $e_{L,R}^* > 0$. Using the intermediate states for the density and the velocity in the Riemann solution (4.45), we can write

$$\frac{1}{\rho_L^*} = \frac{1}{\rho_L} - \frac{1}{2a}(u_R - u_L) + \frac{1}{2a^2}\left(\pi_R - \pi_L - \frac{1-M^2}{M^2}(\psi_R - \psi_L)\right). \tag{4.51}$$

For the internal energy e_L^*, we insert the definition of π_L^* from (4.45) into e_L^* to obtain

$$\begin{aligned} e_L^* = e_L &+ \frac{1}{8}(u_L - u_R)^2 + \\ &\frac{1}{2a^2}\left(-\pi_L^2 + \frac{1}{4}\left(\pi_L + \pi_R + \frac{1-M^2}{M^2}(\psi_R - \psi_L)\right)^2\right. \\ &\qquad \left. + \frac{1}{2}\psi_L(1-M^2)\left(\pi_R - \pi_L + \frac{1-M^2}{M^2}(\psi_R - \psi_L)\right)\right) \\ &+ \frac{1}{4a}(u_R - u_L)\left(\pi_L + \pi_R + \frac{1-M^2}{M^2}(\psi_R - \psi_L) + (1-M^2)\psi_L\right). \end{aligned}$$

Both the update of the intermediate density and internal energy depend on the left and right states $W_{L,R}$ only. From Lemma 4.2, we know $\psi_{L,R} = p_0 + \mathcal{O}(M^2)$. The difference $\psi_R - \psi_L = \mathcal{O}(M^2)$ cancels with the $1/M^2$. All possibly negative terms in e_L^* and ρ_L^* can then be controlled by the relaxation parameter $a > 0$ independent of the Mach number. The same argument holds for ρ_R^* and e_R^*. □

We remark that the positivity of the intermediate densities $\rho^*_{L,R}$ guarantees the ordering of the approximate wave speeds

$$u_L - \frac{a}{\rho_L} < u^* < u_R + \frac{a}{\rho_R}. \tag{4.52}$$

The positivity property for the first order scheme is given in the next result.

Theorem 4.7 (Positivity property 1)**.** *Let the initial state be given as*

$$w_i^n \in \Omega = \Omega_{phy} \cap \Omega_{wp}$$

satisfying the boundary condition described in (4.30)*. Then under the Mach number independent CFL condition*

$$\frac{\Delta t}{\Delta x} \max_i \left| u_i \pm \frac{a}{\rho_i^n} \right| \leq \frac{1}{2},$$

the numerical scheme defined by (4.38),(4.50) *preserves the positivity of density and internal energy, that is*

$$w_i^{n+1} \in \Omega_{phy}$$

for a relaxation parameter a sufficiently large, but independent of M*.*

Proof. Due to the construction of the numerical scheme, the update of the physical variables is only done through the explicit step (4.50). Therefore we can adopt the proof of Theorem 3.11 in Chapter 3 using the positivity of the Riemann solver from Lemma 4.6. The key element is, that we can write the update for w_i^{n+1} as a convex combination of Riemann solvers which satisfy $QW_{\mathcal{RS}} \in \Omega_{phy}$ according to Lemma 4.6:

$$w_i^{n+1} = \frac{1}{\Delta x} \left(\int_{x_{i-\frac{1}{2}}}^{x_i} QW_{\mathcal{RS}} \left(\frac{x - x_{i-1/2}}{\Delta t}, W_{i-1}^{(1)}, W_i^{(1)} \right) dx \right.$$
$$\left. + \int_{x_i}^{x_{i+\frac{1}{2}}} QW_{\mathcal{RS}} \left(\frac{x - x_{i+1/2}}{\Delta t}, W_i^{(1)}, W_{i+1}^{(1)} \right) dx \right) \tag{4.53}$$

□

This result can be extended to multiple space dimensions by considering dimensional split Riemann problems and defining the update for w_i^{n+1} as a sum of the contributions of the resulting fluxes at the interface. Since we are using dimensional splitting, the update in d dimensions can be written as a sum of updates as given in (4.53) in each dimension and due to convexity we have $w_i^{n+1} \in \Omega_{phy}$ as was done in the proof of Theorem 3.11.

4.4.3 Mach number independent diffusion

Although we are using a Godunov type upwind scheme in the explicit part, our scheme does not suffer from an excessive numerical diffusion as M tends to 0. As we will show in the following this is due to the well-prepared implicitly treated fast pressure $\psi^{(1)}$. In order to do so, we investigate the numerical diffusion vector $\mathcal{D} \in \mathbb{R}^{2+d}$ defined by

$$\mathcal{D} = \frac{f(w_i^n) + f(w_{i+1}^n)}{2} - QF\left(W_{\mathcal{RS}}\left(W_i^{(1)}, W_{i+1}^{(1)} \right) \right), \tag{4.54}$$

where the Matrix Q is defined in (4.9). Given well-prepared initial data $w_i^n \in \Omega_{wp}$, we have the following Mach number expansion for the physical variables as given in (4.22),(4.23) and (4.24)

$$\begin{aligned} \rho_i &= \rho_0 + \mathcal{O}(M), & \rho_{i+1} &= \rho_0 + \mathcal{O}(M) \\ u_i &= u_{0,i} + \mathcal{O}(M), & u_{i+1} &= u_{0,i+1} + \mathcal{O}(M) \\ e_i &= e_0 + \mathcal{O}(M), & e_{i+1} &= e_0 + \mathcal{O}(M) \\ \pi_i &= p_0 + \mathcal{O}(M^2), & \pi_{i+1} &= p_0 + \mathcal{O}(M^2) \end{aligned} \tag{4.55}$$

From Lemma 4.2, we have for ψ after the implicit step

$$\psi_i = p_0 + \mathcal{O}(M^2), \qquad \psi_{i+1} = p_0 + \mathcal{O}(M^2). \tag{4.56}$$

The Mach number expansion of the states $W_i^{(1)}, W_{i+1}^{(1)}$ used in the Riemann solver $W_{\mathcal{RS}}$ is composed of the expansions (4.55) and (4.56). Inserting them into the formulas of the intermediate states (4.45) of the Riemann solution (4.43), we have the following scaling of W_i^*, W_{i+1}^* with respect to the Mach number

$$\begin{aligned} u_{i+1/2}^* &= (u_{0,i} + u_{0,i+1})/2 + \mathcal{O}(1) & & \\ &= u_{0,i+1/2} + \mathcal{O}(1) & & \\ \tau_i^* &= \tau_0 + \mathcal{O}(1), & \tau_{i+1}^* &= \tau_0 + \mathcal{O}(1) \\ e_i^* &= e_0 + \mathcal{O}(1), & e_{i+1}^* &= e_0 + \mathcal{O}(1) \\ \pi_i^* &= p_0 + \mathcal{O}(1), & \pi_{i+1}^* &= p_0 + \mathcal{O}(1), \\ \psi_i^* &= p_0 + \mathcal{O}(M^2), & \psi_{i+1}^* &= p_0 + \mathcal{O}(M^2). \end{aligned} \tag{4.57}$$

From (4.57) it is evident that the lowest order of M in the intermediate states is $\mathcal{O}(M^0)$. Inserting (4.57) in the interface flux gives for $\lambda_L^- < 0 < \lambda_R^+$:

$$QF\left(W_{\mathcal{RS}}\left(W_i^{(1)}, W_{i+1}^{(1)}\right)\right) = \begin{pmatrix} \rho_0 u_{0,i+1/2} + \mathcal{O}(1) \\ \rho_0 u_{0,i+1/2}^2 + \frac{p_0}{M^2} + \mathcal{O}(1) \\ u_{0,i+1/2}(E_0 + p_0) + \mathcal{O}(1) \end{pmatrix}, \tag{4.58}$$

for $\lambda_L^- > 0$:

$$QF\left(W_{\mathcal{RS}}\left(W_i^{(1)}, W_{i+1}^{(1)}\right)\right) = \begin{pmatrix} \rho_0 u_{0,i} + \mathcal{O}(1) \\ \rho_0 u_{0,i}^2 + \frac{p_0}{M^2} + \mathcal{O}(1) \\ u_{0,i}(E_0 + p_0) + \mathcal{O}(1) \end{pmatrix}, \tag{4.59}$$

and for $\lambda_R^+ < 0$:

$$QF\left(W_{\mathcal{RS}}\left(W_i^{(1)}, W_{i+1}^{(1)}\right)\right) = \begin{pmatrix} \rho_0 u_{0,i+1} + \mathcal{O}(1) \\ \rho_0 u_{0,i+1}^2 + \frac{p_0}{M^2} + \mathcal{O}(1) \\ u_{0,i+1}(E_0 + p_0) + \mathcal{O}(1) \end{pmatrix}. \tag{4.60}$$

Therefore, using respectively (4.58), (4.59) and (4.60) in (4.54), the diffusion vector with respect to the Mach number is given by

$$\mathcal{D} = \begin{pmatrix} \mathcal{O}(1) \\ \mathcal{O}(1) \\ \mathcal{O}(1) \end{pmatrix}.$$

This shows that the diffusion introduced by the Riemann solver does not suffer from a $\mathcal{O}(M^{-1})$ dependent diffusion in the momentum equation.

4.5 Second order extension

In this section we extend the first order scheme given by (4.50) and (4.38) to second order accuracy. We seek a natural extension of the first order scheme that preserves the positivity property and the low Mach properties. We begin with a second order IMEX time integration method based on the SSP-RK2 (2.80) and its variable time step version (2.81) based on results from [6].

4.5.1 A second order time integration scheme

Instead of solving one explicit update in one stage of the SSP-RK2 method (2.80), here, one stage consists of solving an implicit equation and a system of explicit equations with a subsequent projection to relaxation equilibrium. The time integration scheme is a convex combination of first order steps. The second order time semi-discrete scheme is given by

$$\begin{aligned} W^{(1)} =& W^{n,\mathrm{eq}} + \frac{\Delta t}{M^2}\nabla\cdot G\left(W^{(1)}\right), \\ W^{(2)} =& W^{(1)} + \Delta t\nabla\cdot F\left(W^{(1)}\right), \\ \overline{W}^{(1)} =& W^{(2),\mathrm{eq}} + \frac{\Delta t}{M^2}\nabla\cdot G\left(\overline{W}^{(1)}\right), \\ \overline{W}^{(2)} =& \overline{W}^{(1)} + \Delta t\nabla\cdot F\left(\overline{W}^{(1)}\right), \\ W^{n+1} =& \frac{1}{2}\overline{W}^{(2),\mathrm{eq}} + \frac{1}{2}W^{n,\mathrm{eq}}. \end{aligned} \tag{4.61}$$

The relaxation equilibrium states $W^{n,\mathrm{eq}}, W^{(2),\mathrm{eq}}, \overline{W}^{(2),\mathrm{eq}}$ are defined as in (4.8). As in the first order scheme, we can rewrite the integration scheme (4.61) in terms of the update for the physical variables only. Writing the first order time update (4.50) in time-semi discrete form

$$w^{n+1} = w^n - \Delta t\nabla\cdot QF(W^{(1)}),$$

where Q is defined in (4.9), we can directly give the second order time semi-discrete scheme for the physical variables $w = (\rho, \rho\mathbf{u}, E)$. Based on this we can write (4.61) in a more compact form as

$$\begin{aligned} \overline{w} =& w^n - \Delta t\nabla\cdot QF\left(W^{(1)}\right) \\ \overline{\overline{w}} =& \overline{w} - \Delta t\nabla\cdot QF\left(\overline{W}^{(1)}\right) \\ w^{n+1} =& \frac{1}{2}w^n + \frac{1}{2}\overline{\overline{w}}. \end{aligned} \tag{4.62}$$

where $W^{(1)}$ and $\overline{W}^{(1)}$ are solutions of the implicit step (4.38) where the initial condition is given by w^n and $\overline{w}$ respectively.

The time integration scheme (4.62) can be extended to variable step sizes $\Delta t_1, \Delta t_2$ for each stage respectively as given in [6] for an explicit scheme, see also Section 2.3 equation (2.81). This has the advantage that the CFL criterion can be met at each stage

independently. It is given by

$$\begin{aligned} \overline{w} =& w^n - \Delta t_1 \nabla \cdot QF\left(W^{(1)}\right) \\ \overline{\overline{w}} =& \overline{w} - \Delta t_2 \nabla \cdot QF\left(\overline{W}^{(1)}\right) \\ w^{n+1} =& \left(1 - \frac{\Delta t}{\Delta t_1 + \Delta t_2}\right) w^n + \frac{\Delta t}{\Delta t_1 + \Delta t_2} \overline{\overline{w}}. \end{aligned} \tag{4.63}$$

with an overall time step

$$\Delta t = \frac{2\Delta t_1 \Delta t_2}{\Delta t_1 + \Delta t_2}. \tag{4.64}$$

In Section 4.6 we will use the time integration method (4.63) in the numerical simulations and numerically verify that it is second order accurate.

4.5.2 A second order space reconstruction

As it is standard in the finite volume setting, we apply a reconstruction to get a higher accuracy for the values at the interfaces $x_{i+1/2}$. To get second order accuracy, we follow the MUSCL approach discussed in Section 2.2 and consider piecewise linear functions in the conserved variables $w = (\rho, \rho\mathbf{u}, E)$ at time level t^n and in addition reconstruct also $\psi^{(1)}$. As we are working on a Cartesian grid, we reconstruct along each dimension separately. In one dimension the linear function in (w_i, x_i) for $x \in (x_{i-1/2}, x_{i+1/2})$ is given by

$$\widetilde{w}^n(x) = w_i^n + \sigma_i (x - x_i),$$

where $\sigma_i = (\sigma_i^\rho, \sigma_i^{\rho u}, \sigma_i^E)$ and for the fast pressure ψ, we define

$$\widetilde{\psi}^{(1)}(x) = \psi_i^{(1)} + \sigma_i^\psi (x - x_i).$$

The slopes σ_i are computed from the neighbouring cells using a limiter function. To have a second order extension that preserves the positivity properties of the first order scheme, see Theorem 4.7, we choose the minmod limiter defined in (2.75). The slopes are then computed as

$$\sigma_i = \operatorname{minmod}\left(\frac{w_i^n - w_{i-1}^n}{\Delta x}, \frac{w_{i+1}^n - w_i^n}{\Delta x}\right), \quad \sigma_i^\psi = \operatorname{minmod}\left(\frac{\psi_i^{(1)} - \psi_{i-1}^{(1)}}{\Delta x}, \frac{\psi_{i+1}^{(1)} - \psi_i^{(1)}}{\Delta x}\right).$$

The interface values are then defined as in (2.68) by

$$w_{i\mp1/2}^{\pm} = w_i^n \mp \sigma_i \frac{\Delta x}{2}, \quad \psi_{i\mp1/2}^{(1),\pm} = \psi_i^{(1)} \mp \sigma_i^\psi \frac{\Delta x}{2}. \tag{4.65}$$

To define the Godunov type scheme for the explicit step, the initial values of the local Riemann problems are given by the values at the interface $x_{i+1/2}$ as

$$w^0(x) = \begin{cases} w_{i+1/2}^- & \text{if} \quad x < x_{i+1/2} \\ w_{i+1/2}^+ & \text{if} \quad x > x_{i+1/2} \end{cases}, \quad \psi^0(x) = \begin{cases} \psi_{i+1/2}^{(1),-} & \text{if} \quad x < x_{i+1/2} \\ \psi_{i+1/2}^{(1),+} & \text{if} \quad x > x_{i+1/2} \end{cases}. \tag{4.66}$$

The full second order scheme based on the time integration method (4.63) is given by

$$\begin{aligned}
\overline{w}_i =& w^n - \frac{\Delta t_1}{\Delta x}\left(\mathfrak{f}\left(w^-_{i+1/2}, \psi^{(1),-}_{i+1/2}, w^+_{i+1/2}, \psi^{(1),+}_{i+1/2}\right) - \mathfrak{f}\left(w^-_{i-1/2}, \psi^{(1),-}_{i-1/2}, w^+_{i-1/2}, \psi^{(1),+}_{i-1/2}\right)\right)\\
\overline{\overline{w}}_i =& \overline{w}_i - \frac{\Delta t_2}{\Delta x}\left(\mathfrak{f}\left(\overline{w}^-_{i+1/2}, \overline{\psi}^{(1),+}_{i+1/2}, \overline{w}^+_{i+1/2}, \overline{\psi}^{(1),+}_{i+1/2}\right) - \mathfrak{f}\left(\overline{w}^-_{i-1/2}, \overline{\psi}^{(1),-}_{i-1/2}, \overline{w}^+_{i-1/2}, \overline{\psi}^{(1),+}_{i-1/2}\right)\right)\\
w_i^{n+1} =& \left(1 - \frac{\Delta t}{\Delta t_1 + \Delta t_2}\right) w_i^n + \frac{\Delta t}{\Delta t_1 + \Delta t_2}\overline{\overline{w}}_i.
\end{aligned} \tag{4.67}$$

with the definition of the flux function $\mathfrak{f}$ from (4.49) based on the Riemann solver (4.43) with the intermediate states (4.45) and the time step Δt defined by (4.64).

Since we are using the minmod limiter on the conservative variables to determine the slopes, we immediately get that for the interface values given by (2.68) holds

$$w^+_{i-1/2}, w^-_{i+1/2} \in \Omega_{phy} \cap \Omega_{wp}$$

based on cell averages with $w_i^n \in \Omega_{phy} \cap \Omega_{wp}$. In addition, due to the linear reconstruction, the expansion of the interface values $\psi^{(1),+}_{i-1/2}$ and $\psi^{(1),-}_{i+1/2}$ with respect to the Mach number is preserved. This means by Lemma 4.6 that the Riemann solver with initial data (4.66) still ensures the positivity of density and internal energy. By Theorem 4.7 the first order scheme has the positivity property and therefore, the second order scheme (4.67) is a convex combination of states in Ω_{phy}. From this we have $w_i^{n+1} \in \Omega_{phy}$. Thus, we have proven the following result for the second order scheme:

Theorem 4.8 (Positivity property 2)**.** *Let the initial state be given as*

$$w_i^n \in \Omega_{phy} \cap \Omega_{wp}$$

with the boundary condition described in (4.30)*. Then under the Mach number independent CFL condition*

$$\frac{\Delta t}{\Delta x}\max_i\left(\left|u^{n,+}_{i-1/2} \pm \frac{a}{\rho^{n,+}_{i-1/2}}\right|, \left|u^{n,-}_{i+1/2} \pm \frac{a}{\rho^{n,-}_{i+1/2}}\right|\right) \leq \frac{1}{4},$$

the numerical scheme defined by (4.67),(4.38) *preserves the positivity of density and internal energy, that is*

$$w_i^{n+1} \in \Omega_{wp}$$

for a relaxation parameter sufficiently large, but independent of M*.*

From the fact, that the second order scheme is a convex combination of first order steps, follows also the AP property of the second order scheme, as well as the Mach number independent diffusion. The latter one is ensured since ψ is still well-prepared after the linear reconstruction of the interface values.

The second order scheme, as well as the first order scheme, can be extended to multiple space dimensions considering dimensional split Riemann problems and updating the variables by considering the contribution of all the fluxes at the cell interfaces as discussed in Section 2.1.4.

Test case	dim	M	a	c_a
SOD shock tube test	1D	1.0	1.77	1.5
Mach number dependent shock test	1D	$6.2 \cdot 10^{-3}$	1.87	2.5
Gresho vortex test	2D	$10^{-1}, 10^{-2}, 10^{-3}$	3.08	2.5
Accuracy test	2D	$10^{-1}, 10^{-2}, 10^{-3}$	6.61	5.0

Table 4.1: Relaxation parameter a for the initial data for each test case.

4.6 Numerical results

In the following section, we numerically validate the theoretical properties of the proposed scheme. In all considered test cases, we assume an ideal gas law with the equation of state given by

$$p = (\gamma - 1)\rho e.$$

For solving the implicit non-symmetric linear system given by (4.38), we use the GMRES algorithm combined with a preconditioner based on an incomplete LU decomposition [4]. For obtaining a global estimate for the relaxation parameter a, that is needed to determine Δt and is also used in the implicit part (4.38), we choose the maximum over all locally computed values of a, given in (2.47), multiplied by a constant c_a independent of M to guarantee the subcharacteristic condition and the stability in accordance with Lemma 4.1. The Riemann solver allows for a smaller relaxation parameter which is calculated locally on each cell following the procedure of Bouchut [11] given in (2.48) fulfilling the subcharacteristic condition locally. This leads to a reduced numerical diffusion in the Riemann solver that is introduced by the relaxation procedure.

In Table 4.1, we give c_a and the resulting relaxation parameter a computed from the initial data. Since a depends on ρ and p it can vary during the simulation, but it is of roughly the same order independently of the Mach number in the respective test case.

4.6.1 Shock test cases

To verify that our proposed IMEX schemes capture discontinuities accurately, we perform a SOD shock tube test [73] in the regime $M = 1$ and a Mach number dependent Riemann problem with $M \approx 6.2 \cdot 10^{-3}$ taken from [1].

SOD shock tube test

The computational domain for the SOD shock tube test [73] is $[0, 1]$ and the initial data is given using $\gamma = 1.4$ by

$$\begin{aligned} \rho_L &= 1 \ \tfrac{kg}{m^3}, & \rho_R &= 0.125 \ \tfrac{kg}{m^3}, \\ u_L &= 0 \ \tfrac{m}{s}, & u_R &= 0 \ \tfrac{m}{s}, \\ p_L &= 1 \ \tfrac{kg}{ms^2}, & p_R &= 0.1 \ \tfrac{kg}{ms^2}, \end{aligned}$$

where we place the initial discontinuity at $x = 0.5$. Since the regime is compressible, we set $M = 1$ and the initial data is given in dimensional form. This test also demonstrates that the collapse of the eigenvalues $\lambda^{\pm}$ and $\lambda_M^{\pm}$ in the case of $M = 1$ is not problematic. We see in Figure 4.2 that the first as well as the second order scheme captures the shock positions correctly. As expected, the second order scheme is more accurate than the first order scheme.

Mach number dependent Riemann problem

The setting of this test case is in the low Mach regime, where the contact wave travels with a Mach number of $M \approx 6.2 \cdot 10^{-3}$. The initial data is taken from [1] and is given by

$$\begin{aligned} \rho_L &= 1 \ \tfrac{kg}{m^3}, & \rho_R &= 1 \ \tfrac{kg}{m^3}, \\ u_L &= 0 \ \tfrac{m}{s}, & u_R &= 0.008 \ \tfrac{m}{s}, \\ p_L &= 0.4 \ \tfrac{kg}{ms^2}, & p_R &= 0.399 \ \tfrac{kg}{ms^2}. \end{aligned} \tag{4.68}$$

The discontinuity is placed at $x_0 = 0.5$ on the domain $[0, 1]$ with $\gamma = 1.4$ and the final time is given by $T = 0.25s$. To transform the initial data (4.68) into non-dimensional quantities, we define for a variable ϕ the relation $\phi = \phi_r \, \hat{\phi}$, where ϕ denotes the dimensional variable, ϕ_r the reference value which contains the units and $\hat{\phi}$ the non-dimensional quantity. For the reference values we have the following relations

$$u_r = \frac{x_r}{t_r}, \ p_r = \rho_r c_r^2, \ c_r = \frac{u_r}{M}.$$

We choose the scaling in space to be $x_r = 1m$, the scaling in density to be $\rho_r = 1\frac{kg}{m^3}$ and the scaling in velocity to be $u_r = M\frac{m}{s}$. This yields the following scaling of the sound speed $c_r = 1\frac{m}{s}$, the time $t_r = Ms$ and the pressure $p_r = 1\frac{kg}{ms^2}$. Then the non-dimensional initial data is given by

$$\begin{aligned} \rho_L &= 1, & \rho_R &= 1, \\ u_L &= 0, & u_R &= 0.008/M, \\ p_L &= 0.4, & p_R &= 0.399. \end{aligned} \tag{4.69}$$

In the simulation, we choose $M = 6.2 \cdot 10^{-3}$ which is the Mach number on the contact wave.

In Figure 4.3a we show the influence of the space and time step on the density profile computed with the first order scheme (IMEX1) and the second order scheme (IMEX2). The contact wave is always reproduced sharply whereas the acoustic waves are smoothed since the time step is chosen according to the CFL restriction (4.47) associated with the contact wave. Choosing a finer mesh, and thus having a time step closer to the one oriented at the acoustic waves, allows us to also resolve the acoustic waves as depicted in Figure 4.3b. Moreover can be observed that the numerical approximation converges towards the exact solution. In Figure 4.4 our results computed with IMEX1 and IMEX2 are plotted against the implicit Jin Xin relaxation scheme presented in [1] and an explicit upwind

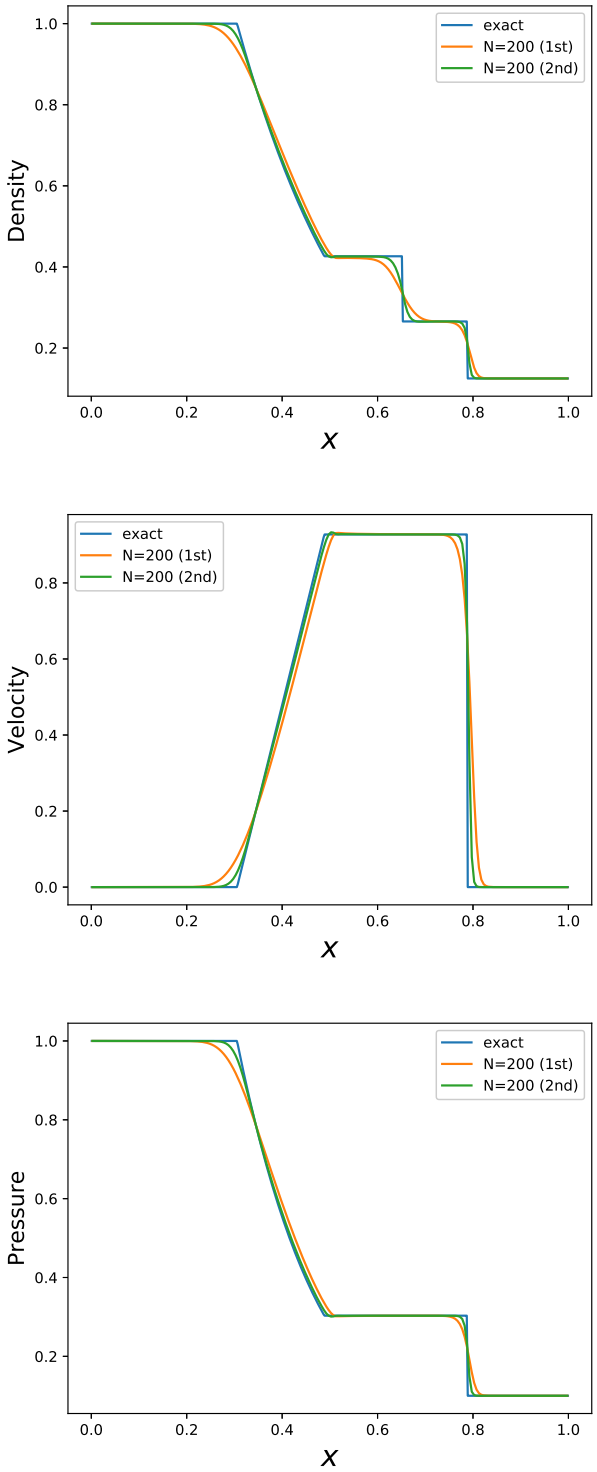
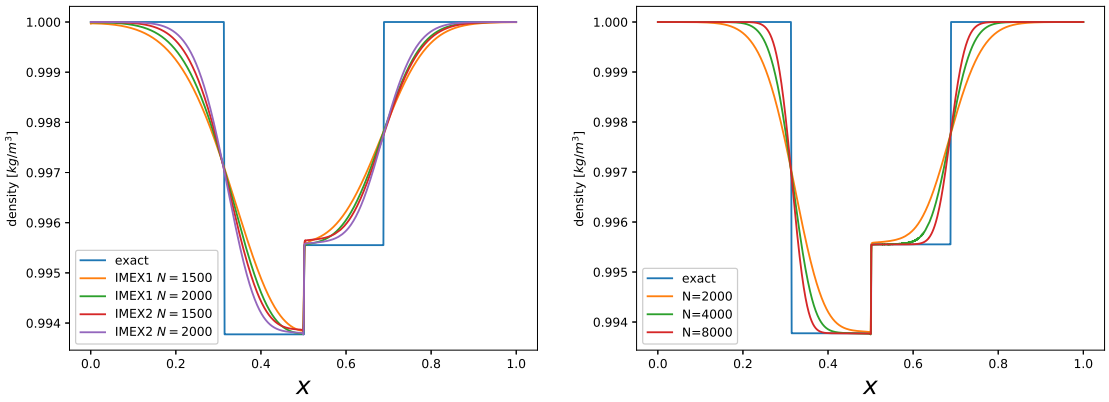


Figure 4.2: SOD test case for $T = 0.1644, \gamma = 1.4$.



(a) Time steps for IMEX1: $\Delta t = 7.8 \cdot 10^{-3}s$ and $\Delta t = 5.8 \cdot 10^{-3}s$, time steps for IMEX2: $\Delta t = 3.9 \cdot 10^{-3}s$ and $\Delta t = 2.9 \cdot 10^{-3}s$.

(b) Increased resolution of the acoustic waves due to mesh refinement (IMEX2).

Figure 4.3: Density profile for the Mach number dependent test case with different number of grid points and time steps.

second order Suliciu relaxation scheme. Our results obtained with the IMEX schemes are in good agreement with the results of the implicit scheme. Since the time step for the explicit upwind scheme depends on the Mach number dependent acoustic waves, all waves are resolved. This is rather costly since it results in a very small time step. The implicit scheme is unconditionally stable and the time step can be chosen with respect to the desired accuracy of the numerical solution.

4.6.2 Gresho Vortex test

In order to demonstrate the low Mach properties, we calculate the solution to the Gresho Vortex test as given in [60]. The Gresho vortex is a stationary solution of the compressible Euler equations and its initial data is well-prepared.

The velocity field $\mathbf{u}$ is divergence free and defined by the angular velocity in $\frac{m}{s}$, given by

$$u_\theta = \begin{cases} 5r & \text{for } 0 \leq r < 0.2 \\ 2 - 5r & \text{for } 0.2 \leq r < 0.4 \\ 0 & \text{for } 0.4 \leq r \end{cases},$$

where $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$ with $x_0 = 0.5, y_0 = 0.5$ on a computational domain of $[0, 1]$ and $\gamma = 5/3$. The pressure distribution in $\frac{kg}{ms^2}$ is given by

$$p = \begin{cases} p_0 + 12.5r^2 & \text{for } 0 \leq r < 0.2 \\ p_0 + 12.5r^2 + 4(1 - 5r - \log(0.2) + \log(r)) & \text{for } 0.2 \leq r < 0.4 \\ p_0 - 2 + 4\log(2) & \text{for } 0.4 \leq r \end{cases}$$

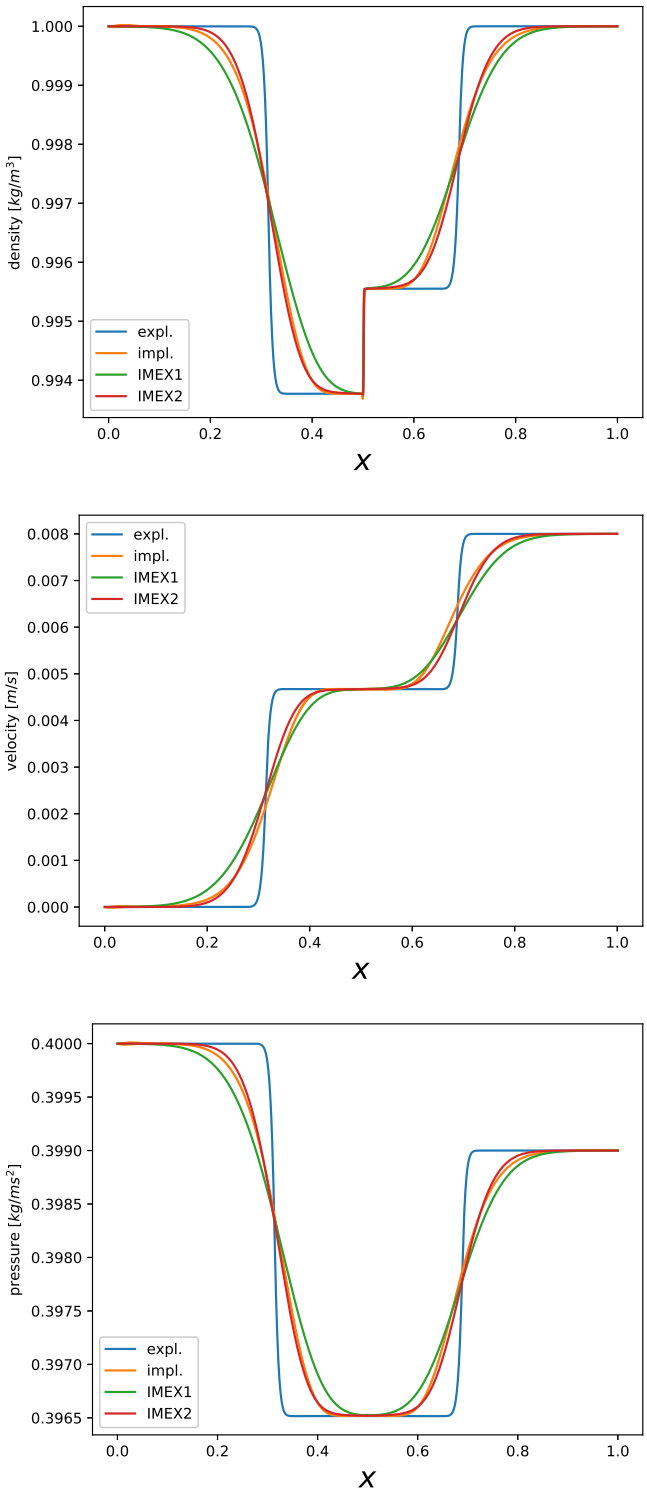


Figure 4.4: Mach dependent shock test case: Time steps are given by $\Delta t = 1.1 \cdot 10^{-7} s$ (expl.), $\Delta t = 2.2 \cdot 10^{-2} s$ (impl.), $\Delta t = 5.0 \cdot 10^{-3} s$ (IMEX1), $\Delta t = 2.5 \cdot 10^{-3} s$ (IMEX2).

with $p_0 = \frac{\rho_0 u_{\Phi,\max}^2}{\gamma M^2} \frac{kg}{ms^2}$, where $\rho_0 = 1$. The initial density is given by $\rho = 1\frac{kg}{m^3}$ and we transform the initial condition in non-dimensional quantities by using $x_r = 1m$, $\rho_r = 1\frac{kg}{m^3}$, $u_r = 2 \cdot 0.2\ \pi\frac{m}{s}$, $p_r = \frac{\rho_0 u_0^2}{\gamma M^2}\frac{kg}{ms^2}$ and $t_r = 1\frac{m}{u_r}$. The computational domain is given by $[0,1] \times [0,1]$. We use a 40×40 grid with periodic boundary conditions. The results for a full turn of the vortex using the IMEX2 scheme together with the initial distribution of the Mach number are given in Figure 4.5. We see that even for a low resolution, the solution at $T = 1$ shows only little dissipation throughout all tested Mach numbers. To further check the quality of the numerical simulation, we monitor the loss of kinetic energy. Since the Gresho vortex is a stationary solution of the incompressible Euler equations, the kinetic energy should be preserved. In Figure 4.6 we show the ratio between the initial kinetic energy $E_{kin,0}$ and the kinetic energy after each time step $E_{kin,t}$ for the Mach numbers $M = 10^{-2}, 10^{-3}$. The graphs for the different Mach numbers are indistinguishable which shows that the loss of kinetic energy does not depend on the chosen Mach number but depends on the chosen space discretization and time-step.

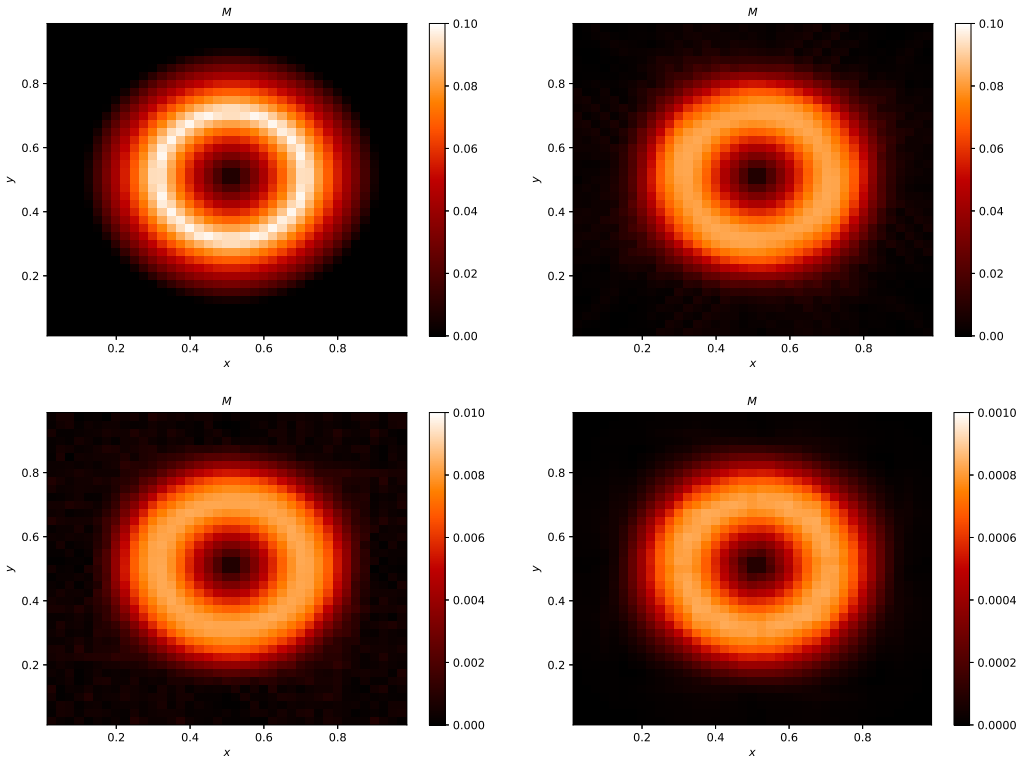


Figure 4.5: Mach number distribution for different maximal reference Mach numbers. Top left: Initial state for $M = 10^{-1}$. Top right: $M = 10^{-1}$, bottom left: $M = 10^{-2}$, bottom right: $M = 10^{-3}$ at $t = 1$.

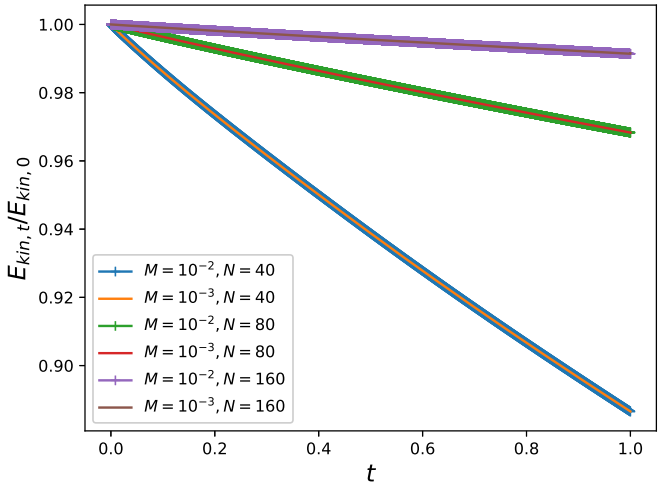


Figure 4.6: Loss of kinetic energy for different grids and reference Mach numbers after one full turn of the vortex (non-dimensional).

4.6.3 Accuracy

Now we turn to the accuracy of the second order scheme given in Section 4.5. Unfortunately the Gresho vortex is not a smooth enough solution to test second order accuracy, since its velocity profile is only continuous but not continuously differentiable. Therefore we propose a C^1 velocity profile with which we then calculate a pressure profile to gain a stationary vortex. A continuously differentiable angular velocity in $\frac{m}{s}$ with $u_{\Phi,\max} = 1$ and $u_\theta(0) = 0$ and $u_\theta(0.4) = 0$ is given by

$$u_\theta = \begin{cases} 75r^2 - 250r^3, & \text{for } 0 \leq r < 0.2 \\ -4 + 60r - 225r^2 + 250r^3, & \text{for } 0.2 \leq r < 0.4 \\ 0 & \text{for } 0.4 \geq r \end{cases}$$

with the radius $r = \sqrt{(x-0.5)^2 + (y-0.5)^2}$. The profile can be easily modified to be a C^k function, where $k \in \mathbb{N}$ denotes the degree of continuous differentiability. Under the condition that the centrifugal forces are balanced, that is $\partial_r p = \frac{u_\theta(r)^2}{r}$, we can calculate the pressure in $\frac{kg}{ms^2}$ as

$$p = \begin{cases} p_0 + 1406.25r^4 - 7500r^5 + (10416 + \frac{2}{3})r^6 & \text{for } 0 \leq r < 0.2 \\ p_0 + p_2(r) & \text{for } 0.2 \leq r < 0.4 \\ p_0 + p_2(0.4) & \text{for } 0.4 \geq r \end{cases}$$

where

$$\begin{aligned} p_2(r) =& 65.8843399322788 - 480r + 2700r^2 - (9666 + \frac{2}{3})r^3 \\ &+ 20156.25r^4 - 22500r^5 + (10416 + \frac{2}{3})r^6 + 16\ln(r). \end{aligned}$$

M	N	ρ		u_1		u_2		p	
10^{-1}	20	$1.810{\cdot}10^{-3}$	—	$1.729{\cdot}10^{-2}$	—	$1.729{\cdot}10^{-2}$	—	$1.921{\cdot}10^{-3}$	—
	40	$3.705{\cdot}10^{-4}$	2.288	$5.070{\cdot}10^{-3}$	1.770	$5.070{\cdot}10^{-3}$	1.770	$3.956{\cdot}10^{-4}$	2.279
	60	$1.246{\cdot}10^{-4}$	2.688	$2.403{\cdot}10^{-3}$	1.842	$2.403{\cdot}10^{-3}$	1.842	$1.343{\cdot}10^{-3}$	2.665
	80	$5.510{\cdot}10^{-5}$	2.835	$1.396{\cdot}10^{-3}$	1.887	$1.396{\cdot}10^{-3}$	1.887	$5.922{\cdot}10^{-5}$	2.845
10^{-2}	20	$1.812{\cdot}10^{-3}$	—	$1.731{\cdot}10^{-2}$	—	$1.731{\cdot}10^{-2}$	—	$1.912{\cdot}10^{-3}$	—
	40	$3.582{\cdot}10^{-4}$	2.339	$5.057{\cdot}10^{-3}$	1.775	$5.057{\cdot}10^{-3}$	1.775	$3.781{\cdot}10^{-4}$	2.337
	60	$1.162{\cdot}10^{-4}$	2.777	$2.402{\cdot}10^{-3}$	1.837	$2.402{\cdot}10^{-3}$	1.837	$1.226{\cdot}10^{-4}$	2.777
	80	$4.881{\cdot}10^{-5}$	3.014	$1.386{\cdot}10^{-3}$	1.910	$1.386{\cdot}10^{-3}$	1.910	$5.151{\cdot}10^{-5}$	3.014
10^{-3}	20	$1.811{\cdot}10^{-3}$	—	$1.731{\cdot}10^{-2}$	—	$1.731{\cdot}10^{-2}$	—	$1.912{\cdot}10^{-3}$	—
	40	$3.580{\cdot}10^{-4}$	2.339	$5.057{\cdot}10^{-3}$	1.775	$5.057{\cdot}10^{-3}$	1.775	$3.778{\cdot}10^{-4}$	2.339
	60	$1.162{\cdot}10^{-4}$	2.775	$2.402{\cdot}10^{-3}$	1.836	$2.402{\cdot}10^{-3}$	1.836	$1.227{\cdot}10^{-4}$	2.775
	80	$4.875{\cdot}10^{-5}$	3.019	$1.382{\cdot}10^{-3}$	1.920	$1.382{\cdot}10^{-3}$	1.920	$5.146{\cdot}10^{-5}$	3.019

Table 4.2: L^1-error and convergence rates for the solution of the smooth Gresho vortex test at $T = 0.05$ (non-dimensional).

As in the Gresho vortex test case, the background pressure is scaled with the Mach number as

$$p_0 = \frac{\rho_0 u_{\Phi,\max}^2}{\gamma M^2} \frac{kg}{ms^2}.$$

To transform the dimensional data into non-dimensional quantities, we use the same reference values as in the Gresho vortex test case.

To show the accuracy of the IMEX2 scheme, we compute the solution of the smooth Gresho vortex test on the domain $[0,1]^2$ with periodic boundary conditions. In Table 4.2 the L^1-error between the numerical solution at $T = 0.05$ and the initial configuration in non-dimensional quantities as well as the convergence rates are displayed for $M = 10^{-1}, 10^{-2}, 10^{-3}$. It can be seen that we reach the expected accuracy independently of the chosen Mach number although we do not recover a full second order convergence in u_1 and u_2. This is due to the usage of the minmod limiter in the reconstruction step which is truncating the slopes when reaching a minimum or maximum in the solution to guarantee the positivity of ρ and e. However, using an unlimited linear reconstruction instead of the minmod limiter to calculate the slopes will lead to the full second order in all variables.

4.7 Conclusion

We have proposed an all-speed IMEX scheme for the full Euler equations of gas dynamics which is based on a Suliciu relaxation model. The proposed IMEX scheme is an improvement of explicit schemes since the time step is restricted by the material and not the acoustic wave speeds. It is also an improvement of implicit schemes since the implicit part consists only of one scalar linear equation and can be solved very efficiently. The scheme has the correct numerical viscosity for all Mach numbers as verified by the Gresho

vortex test case, can capture the correct shock positions as shown by the SOD shock tube test case as well as by the Mach number dependent Riemann problem. In addition it is positivity preserving and shows the expected second order convergence rates.

Starting from this encouraging results of the homogeneous model, we extend the schemes in the following chapter to be able to treat flows in the presence of a gravitational source term. The aim is to design well-balanced scheme, that inherits all the properties of the IMEX scheme presented in this chapter combined with the well-balancing techniques discussed in Chapter 3.

Chapter 5

An all-speed scheme for the Euler equations with gravity

The aim of this chapter is the construction of an all-speed scheme for the Euler equations of gas dynamics with a given gravitational source term in multiple space dimensions. Applications of this model can be found for example in astrophysics and meteorology. A broad overview is given in the review of Klein [49] where it is demonstrated that atmospheric flows can have large scale differences. To reflect those scales in the equations, we use the non-dimensional formulation of the equations which are characterized by the reference Mach and Froude number denoted by M and Fr respectively.

In the homogeneous case, the behaviour of the fluid changes in dependence of the Mach number only. It ranges from compressible flow for large Mach numbers to the incompressible limit equations for M going to zero. The derivation of the limit equations can be found eg. in [47, 26, 70] and references therein. To accurately approximate all-speed flows, asymptotic preserving (AP) schemes are well suited since they are consistent with the limit behaviour as M tends to zero. The development of those schemes is an active field of research and we refer to the review of Jin [42] for an introduction. An important role in the achievement of the AP property is played by the splitting of the pressure following the studies of Klein [48, 50] as used in the schemes [62, 23, 79, 7]. In [79, 7] the pressure splitting is combined with a Suliciu relaxation approach [74] which allows for an easy construction of Riemann Solvers. An example for a Jin-Xin relaxation approach [43] can be found in [1].

Since for explicit schemes the time step is restricted by the inverse of the largest wave speed, which scales with $1/M$, explicit schemes are not practical for low Mach applications. Therefore implicit schemes, see for example [7, 1], or implicit-explicit (IMEX) schemes, see for example [23, 9, 30], are used to have a Mach number independent time step.

The presence of a source term makes it interesting to look at steady states. For a zero velocity field, we find hydrostatic equilibria, which are characterized by the balance of the pressure gradient with the weight of the fluid. Most atmospheric-flow phenomena may be understood as perturbations of such a balanced background state. The scope of well-balanced schemes is to maintain the background atmosphere at machine precision to be able to resolve those small perturbations accurately. Since the shape of the equilibrium state depends on the underlying pressure law, there are schemes focused on well-balancing

a specific class of equilibria, for example isothermal and polytropic atmospheres [28] or equilibria with constant entropy [44]. The latter was extended in [45] to the preservation of hydrostatic equilibria with arbitrary entropy stratification using a second order reconstruction of the discrete equilibrium equation. A different approach can be found in [35], where the well-balanced property is achieved by using path-conservative finite volumes schemes. Higher order well-balanced schemes can be realized by using a high order hydrostatic reconstruction, as done in [86, 51, 39]. Since our aim is to exactly well-balance arbitrary hydrostatic equilibria, we follow the approaches from [86, 36], also used in [51, 78], and rewrite the gravitational potential in terms of a reference equilibrium state. Note that the above mentioned well-balancing techniques were developed for the compressible regime. Instead the scheme presented here is designed to be well-balanced also in the low Mach, low Froude regime. It is a natural continuation of the second order AP IMEX scheme developed for the homogeneous Euler equations [79], presented in the previous chapter, where the focus was mainly on the construction of an AP scheme with a Mach number independent time-step as well as the preservation of the positivity of density and internal energy. Here, we complete the scheme adding a gravitational source term treated in the framework presented in Chapter 3 and we use that approach to show the well-balanced and positivity property of the new scheme. One of the main difficulties is to combine the well-balanced property with the IMEX framework as the approach in Chapter 3 was developed for a purely explicit scheme.

To our knowledge, this is the first case in which the construction of a well balanced scheme for general equilibria is addressed which, at the same time, preserves asymptotic properties in the low Mach regime under a gravitational field for the full Euler equations. We show the AP property of the scheme by proving that it preserves the divergence free constraint in the zero Mach number limit when starting from well prepared initial data. The limit equations are given by the incompressible Euler equations in a gravitational field. Similar results were found in [8] for the isentropic case with potential temperature. We refer to [33, 34] for theoretical studies on the isothermal and isentropic case with a one component linear gravitational field and to [5] for a low Mach scheme that allows for a gravitational source term, but lacks the well-balanced property.

The results presented in this chapter are already published in [77]. The chapter is organized as follows. In Section 5.1, we introduce the equations, the notion of hydrostatic equilibria and the limit equations. Then we give the derivation of the Suliciu type relaxation model in Section 5.2. The time semi-discrete scheme with the flux splitting together with the Mach number expansion of the fast pressure and the asymptotic preserving property are discussed in Section 5.3. Subsequent, we give the derivation of the fully discrete scheme which includes a Godunov type finite volume scheme based on an approximative Riemann solver in the explicit part. We show that the scheme is well-balanced and that it preserves the positivity of density and internal energy. The section ends with the extension to second order. All properties are numerically validated in Section 5.5. In particular, we give an example of low Mach flow, starting from well prepared initial data, and we study a low Mach stationary vortex in a gravitational field, with a test we derived from the classical Gresho vortex test case from [60]. We conclude the numerical tests with a simulation of a rising hot air bubble which arises in meteorology.

5.1 The Euler equations with a gravitational source term

The Euler equations with a gravitational source term in d dimensions are given by

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{u}) &= 0, \\ \partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \nabla p &= -\rho \nabla \Phi, \\ \partial_t E + \nabla \cdot (\mathbf{u}(E+p)) &= -\rho \mathbf{u} \cdot \nabla \Phi \end{aligned} \tag{5.1}$$

where the total energy E is composed of the kinetic and internal energy given by

$$E = \rho \left(\frac{1}{2} |\mathbf{u}|^2 + e \right)$$

and Φ denotes a smooth given gravitational potential. To highlight the impact of slow and fast scales in the equations, we formulate (5.1) in its non-dimensional form by rewriting the variables as a scalar reference value that contains the units, indicated by the subscript $(\cdot)_r$, multiplied by a non-dimensional quantity indicated by $\widetilde{(\cdot)}$, as defined in (1.64). Choosing the reference length x_r, time t_r, density ρ_r, sound speed c_r and gravitational acceleration Φ_r, we can compute the missing reference values as

$$u_r = \frac{x_r}{t_r}, \quad p_r = \rho_r c_r^2, \quad e_r = c_r^2. \tag{5.2}$$

Inserting the decomposition (1.64),(5.2) in the dimensional equations (5.1) and using the relations (5.2), we arrive at the non-dimensional Euler equations with a gravitational source term:

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{u}) &= 0 \\ \partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \frac{1}{M^2} \nabla p &= -\frac{1}{Fr^2} \rho \nabla \Phi \\ \partial_t E + \nabla \cdot (\mathbf{u}(E+p)) &= -\frac{M^2}{Fr^2} \rho \mathbf{u} \cdot \nabla \Phi. \end{aligned} \tag{5.3}$$

For simplicity, we have dropped the tilde and, if not otherwise mentioned, we will use the non-dimensional variables throughout this chapter. The total energy of system (5.3) is given by

$$E = \rho \left(\frac{1}{2} M^2 |\mathbf{u}|^2 + e \right).$$

Equations (5.3) depend on two non-dimensional quantities, the Mach number M and the Froude number Fr. The reference Mach number, that is used here, is defined as the ratio between the reference velocity of the gas and the reference sound speed

$$M = \frac{u_r}{c_r}$$

while the reference Froude number is defined as the ratio between the reference velocity of the gas and the velocity introduced by the reference gravitational acceleration

$$Fr = \frac{u_r}{\sqrt{\Phi_r}}.$$

5.1.1 Hydrostatic equilibria

Also for the non-dimensional equations, we can define hydrostatic equilibria which now depend on the Mach and Froude number. The stationary states at rest of the non-dimensional equations (5.3) satisfy for $\mathbf{u} = 0$

$$\frac{1}{M^2}\nabla p = -\frac{1}{Fr^2}\rho\nabla\Phi. \tag{5.4}$$

As in the dimensional case discussed in Section 1.5, solutions to (5.4) are not unique and, depending on the relation between the pressure and the density, they can have completely different behaviour. To demonstrate this, let us for a moment consider the following class of pressure laws

$$p = \chi\rho^{\Gamma} \tag{5.5}$$

with constants $\chi > 0$, $\Gamma \in (0, \infty)$. For the class of equation of states (5.5), we obtain for $\Gamma = 1$ (isothermal) with a constant $C \in \mathbb{R}$ and $\chi = RT$

$$\rho(\mathbf{x}) = \exp\left(\frac{C - \frac{M^2}{Fr^2}\Phi(\mathbf{x})}{RT}\right), \quad p(\mathbf{x}) = RT\rho(\mathbf{x}) \tag{5.6}$$

and for $\Gamma \in (0,1) \cup (1,\infty)$ (polytropic) with a constant $C \in \mathbb{R}$

$$\rho(\mathbf{x}) = \left(\frac{\Gamma - 1}{\chi\Gamma}\left(C - \frac{M^2}{Fr^2}\Phi(\mathbf{x})\right)\right)^{\frac{1}{\Gamma-1}}, \quad p(\mathbf{x}) = \chi\rho(\mathbf{x})^{\Gamma}. \tag{5.7}$$

Comparing the isothermal and polytropic equilibria (5.6) and (5.7) to the solutions in dimensional form given by (1.90) and (1.91) respectively, we see that in front of the gravitational potential Φ, appears now the ratio between Mach and Froude number. Especially if M and Fr are of the same order, the equilibria are independent of these parameters.

To obtain a scheme, that can well-balance arbitrary known equilibria, we rewrite the potential Φ in terms of a reference equilibrium, expressed by α, β as introduced in Section 3.1. Since arbitrary solutions $\bar{\rho}$ and $\bar{p}$ of the hydrostatic equilibrium (5.4) are stationary, we follow [36] and define two time-independent positive functions

$$\alpha(\mathbf{x}) = \bar{\rho}(\mathbf{x}) \quad \text{and} \quad \beta(\mathbf{x}) = \bar{p}(\mathbf{x})$$

representing the equilibrium density and pressure respectively. Since α, β satisfy (5.4), we can find a new relation for $\nabla\Phi$ due to the following equivalent description

$$\frac{1}{M^2}\nabla\beta = -\frac{1}{Fr^2}\alpha\nabla\Phi \quad \Leftrightarrow \quad \nabla\Phi = -\frac{Fr^2}{M^2}\frac{\nabla\beta}{\alpha}. \tag{5.8}$$

With this definition of the gravitational potential, we can rewrite (5.3) into

$$\begin{aligned} \partial_t\rho + \nabla\cdot(\rho u) &= 0, \\ \partial_t(\rho\mathbf{u}) + \nabla\cdot(\rho\mathbf{u}\otimes\mathbf{u}) + \frac{1}{M^2}\nabla p &= \frac{1}{M^2}\frac{\rho}{\alpha}\nabla\beta, \\ \partial_t E + \nabla\cdot(\mathbf{u}(E+p)) &= \frac{\rho}{\alpha}\mathbf{u}\cdot\nabla\beta. \end{aligned} \tag{5.9}$$

We emphasize, that the reference equilibrium has to be known a priori. In general, this is not a restriction, as in many applications the interest lies in preserving a particular equilibrium state, from which the functions α and β can be calculated in advance. In the case of an isothermal or polytropic equilibrium, they can be set according to (5.6) or (5.7) respectively. Since they represent the given stationary potential $\nabla\Phi$, they are also considered as given stationary functions and it is not necessary to update them in time. Note, that the equations (5.9) are only depending on the Mach number, but the dependence on the Froude number is implicitly given in the definition of β in (5.8).

5.1.2 The low Mach limit

Most interesting for the asymptotic analysis from a well-balancing point of view is the case $M = Fr$. In this case one can find a balance between the source and pressure terms in the momentum and energy equation which can also be recovered in the limit $M, Fr \to 0$. This choice was also considered in [8, 33, 34]. To analyse multi-scale effects and the formal asymptotic behaviour of (5.3), we express the variables in form of a Mach number expansion and compare the orders of terms in M. The expansions are given by

$$\begin{aligned} \rho &= \rho_0 + M\rho_1 + M^2\rho_2 + \mathcal{O}(M^3), \quad \mathbf{u} = \mathbf{u}_0 + M\mathbf{u}_1 + M^2\mathbf{u}_2 + \mathcal{O}(M^3), \\ e &= e_0 + Me_1 + M^2e_2 + \mathcal{O}(M^3), \quad p = p_0 + Mp_1 + M^2p_2 + \mathcal{O}(M^3). \end{aligned} \tag{5.10}$$

Inserting the expansions (5.10) into the Euler equations (5.9) and collecting the terms of order $\mathcal{O}(M^{-2})$, we have

$$\nabla p_0 = -\rho_0 \nabla\Phi. \tag{5.11}$$

For the $\mathcal{O}(M^{-1})$ terms, we find

$$\nabla p_1 = -\rho_1 \nabla\Phi. \tag{5.12}$$

This means that the couples p_0, ρ_0 and p_1, ρ_1 fulfil the hydrostatic equilibrium and thus are time-independent. Using this fact, the $\mathcal{O}(M^0)$ terms can be reduced to

$$\begin{aligned} \nabla \cdot (\rho_0 \mathbf{u}_0) &= 0, \\ \partial_t \mathbf{u}_0 + \mathbf{u}_0 \cdot \nabla \mathbf{u}_0 + \frac{\nabla p_2}{\rho_0} &= -\frac{\rho_2 \nabla\Phi}{\rho_0}, \\ \nabla \cdot \mathbf{u}_0 &= \frac{\mathbf{u}_0 \cdot \nabla\Phi}{c_0^2}, \end{aligned} \tag{5.13}$$

where we have used $c_0^2 = \gamma \frac{p_0}{\rho_0}$. An important class of solutions of (5.13) are solutions with a velocity field orthogonal to the gravitational direction. This directly implies a divergence free velocity field $\nabla \cdot \mathbf{u}_0 = 0$ in (5.13). Associated to this case for a given potential Φ is the following set of well-prepared data

$$\begin{aligned} \Omega_{wp} = \Big\{ w \in \mathbb{R}^{d+2};\ \nabla p_0 = -\rho_0 \nabla\Phi,\ \nabla p_1 = -\rho_1 \nabla\Phi, \\ \nabla \cdot (\rho_0 \mathbf{u}_0) = 0, \nabla \cdot \mathbf{u}_0 = 0,\ \mathbf{u}_0 \cdot \nabla\Phi = 0 \Big\}. \end{aligned} \tag{5.14}$$

The well-prepared data for given α, β regarding the modified equations (5.9) are defined analogously by

$$\Omega_{wp}^{\alpha\beta} = \Big\{ w \in \mathbb{R}^{d+2}; \nabla p_0 = \rho_0 \frac{\nabla \beta}{\alpha},\ \nabla p_1 = \rho_1 \frac{\nabla \beta}{\alpha}, \\ \nabla \cdot (\rho_0 \mathbf{u}_0) = 0, \nabla \cdot \mathbf{u}_0 = 0,\ \mathbf{u}_0 \cdot \frac{\nabla \beta}{\alpha} = 0 \Big\}. \quad (5.15)$$

We can conclude from this short analysis that well-prepared pressure and density fulfil the hydrostatic equilibrium up to a perturbation of M^2. We want to remark that a more general set of well-prepared data can be obtained by only requiring the balance between $\nabla \cdot \mathbf{u}_0$ and $\mathbf{u} \cdot \nabla \Phi / c_0^2$ in (5.13).

5.2 Suliciu Relaxation model

This section is devoted to the relaxation model which is the bases for the IMEX scheme. Using a Suliciu relaxation approach as done for example in [74, 11, 22] is one way of simplifying the non-linear structure of the Euler equations (5.3) in such a way that the characteristic fields of the relaxation model are linearly degenerate. This structure provides a natural Riemann solver. The derivation of the relaxation model follows the arguments given in [23, 79, 7]. As already done in the homogeneous case described in Section 4.1, following Klein [48], we apply in the momentum and energy equation a splitting of the pressure p into a slow and a fast component

$$\frac{p}{M^2} = p + \frac{1 - M^2}{M^2} p.$$

The aim is to relax both the slow and the fast pressure in a Suliciu relaxation manner. The pressure in relaxation equilibrium is then characterized by

$$p = M^2 \pi + (1 - M^2)\psi,$$

where π is the approximation of the slow and ψ of the fast part. To obtain the evolution of π, we can directly apply the Suliciu relaxation technique which leads to the addition of the following equation in conservation form

$$\partial_t(\rho\pi) + \nabla \cdot (\rho\pi \mathbf{u}) + a^2 \nabla \cdot \mathbf{u} = \frac{\rho}{\varepsilon}(p - \pi),$$

where $a > 0$ denotes the relaxation parameter which has to fulfil a subcharacteristic condition specified in Lemma 5.1. As discussed in [7], applying this Suliciu relaxation technique also on the fast pressure does not lead to scheme that is accurate for small Mach numbers. Instead a relaxation equation for the velocity $\hat{\mathbf{u}}$ coupled with the pressure ψ is added. We apply the same strategy as in the homogeneous case described in Section 4.1 or in [7, 88]. Here in addition, the influence of the source term in the momentum equation has to be accounted for. As a consequence, the source term will also appear in the relaxation equation for $\hat{\mathbf{u}}$. The full relaxation model is developed under the following objectives:

- It has a ordered eigenvalues that lead to a clear wave structure easing the construction of a Riemann solver.
- It is a stable diffusive approximation of the non-dimensional Euler equations with gravitational source term (5.9).
- The resulting numerical scheme has Mach number independent diffusion.

The achievement of the first objective depends also on the treatment of the source term, since it is associated to β with a zero eigenvalue. Following [28], we avoid this zero eigenvalue by relaxing also β. It is approximated by a new variable Z that is transported with $\mathbf{u}$ as

$$\partial_t Z + \mathbf{u} \cdot \nabla Z = \frac{1}{\varepsilon}(\beta - Z).$$

This associates the source term with the eigenvalue $\mathbf{u}$. All these considerations lead to the following relaxation model

$$\begin{aligned}
\partial_t \rho + \nabla \cdot (\rho u) &= 0, \\
\partial_t(\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \nabla \pi + \frac{1 - M^2}{M^2} \nabla \psi &= \frac{1}{M^2} \frac{\rho}{\alpha} \nabla Z, \\
\partial_t E + \nabla \cdot (\mathbf{u}(E + M^2 \pi + (1 - M^2)\psi)) &= \frac{\rho}{\alpha} \mathbf{u} \cdot \nabla Z, \\
\partial_t(\rho \pi) + \nabla \cdot (\rho \mathbf{u} \pi + a^2 \mathbf{u}) &= \frac{\rho}{\varepsilon}(p - \pi), \\
\partial_t(\rho \hat{\mathbf{u}}) + \nabla \cdot (\rho \mathbf{u} \otimes \hat{\mathbf{u}}) + \frac{1}{M^2} \nabla \psi &= \frac{1}{M^2} \frac{\rho}{\alpha} \nabla Z + \frac{\rho}{\varepsilon}(\mathbf{u} - \hat{\mathbf{u}}), \\
\partial_t(\rho \psi) + \nabla \cdot (\rho \mathbf{u} \psi + a^2 \hat{\mathbf{u}}) &= \frac{\rho}{\varepsilon}(p - \psi), \\
\partial_t(\rho Z) + \nabla (\rho \mathbf{u} Z) &= \frac{\rho}{\varepsilon}(\beta - Z).
\end{aligned} \tag{5.16}$$

The following lemma summarizes important properties of the relaxation model (5.16) regarding the structure and stability. The proof including the Chapman-Enskog stability analysis can be done by adapting the steps explained in Sections 1.2.2 and 1.3. Alternatively, the proof can be established analogously to [88].

Lemma 5.1. *The relaxation system (5.16) is hyperbolic and a stable diffusive approximation of (5.9) under the Mach number independent subcharacteristic condition for the relaxation parameter $a > \rho c$. Considering the x_i-direction, it has the following linearly degenerate eigenvalues*

$$\lambda^u = u_i, \ \lambda^{\pm} = u_i \pm \frac{a}{\rho}, \ \lambda_M^{\pm} = u_i \pm \frac{a}{M\rho}$$

where λ^u has multiplicity $2d + 1$, where d is the number of considered space dimensions.

Note, that in the case of $M = 1$, the waves associated with $\lambda_M^{\pm}$ and $\lambda^{\pm}$ collapse to $\lambda^{\pm}$ with multiplicity 2 respectively. Then the core of the relaxation model (5.16) reduces to the hyperbolic relaxation model for the Euler equations with gravity (3.9). For $M < 1$, the eigenvalues have the following order

$$\lambda_M^- < \lambda^- < \lambda^u < \lambda^+ < \lambda_M^+.$$

To shorten notation, we will refer to the original system (5.9) as

$$w_t + \nabla \cdot f(w) = s(w). \tag{5.17}$$

where $w = (\rho, \rho\mathbf{u}, E)^T$ denotes the vector of physical variables, while the flux function $f(w)$ and the source term $s(w)$ are given by

$$f(w) = \begin{pmatrix} \rho\mathbf{u} \\ \rho\mathbf{u} \otimes \mathbf{u} + \dfrac{1}{M^2} p\, \mathbb{I} \\ \mathbf{u}(E+p) \end{pmatrix} \quad \text{and} \quad s(w) = \begin{pmatrix} 0 \\ \dfrac{1}{M^2}\dfrac{\rho}{\alpha}\nabla\beta \\ \dfrac{\rho}{\alpha}\mathbf{u}\cdot\nabla\beta \end{pmatrix}.$$

The relaxation model (5.16) is given by

$$W_t + \nabla \widetilde{F}(W) = S(W) + \frac{1}{\varepsilon} R(W), \tag{5.18}$$

where $W = (\rho, \rho\mathbf{u}, E, \rho\pi, \rho\hat{\mathbf{u}}, \rho\psi, \rho Z)^T$ denotes the state vector and $\widetilde{F}$ the flux function as defined in (5.16). The gravitational source term $S(W)$ and the relaxation source term $R(W)$ are given by

$$S(W) = \begin{pmatrix} 0 \\ \dfrac{1}{M^2}\dfrac{\rho}{\alpha}\nabla Z \\ \dfrac{\rho}{\alpha}\mathbf{u}\cdot\nabla Z \\ 0 \\ \dfrac{1}{M^2}\dfrac{\rho}{\alpha}\nabla Z \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad R(W) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \rho(p-\pi) \\ \rho(\mathbf{u}-\hat{\mathbf{u}}) \\ \rho(p-\psi) \\ \rho(\beta - Z) \end{pmatrix}.$$

The relaxation time ε indicates how fast the perturbed system (5.18) is reaching its equilibrium (5.17). The relaxation equilibrium state is given by

$$W^{\text{eq}} = \mathcal{M}(w) = (\rho, \rho\mathbf{u}, E, \rho p(\rho, e), \rho\mathbf{u}, \rho p(\rho, e), \rho\beta)^T. \tag{5.19}$$

Following [19], we can connect (5.18) to (5.17) through the matrix $Q \in \mathbb{R}^{(2+d)\times(2(2+d)+1)}$ defined as

$$Q = \begin{pmatrix} \mathbb{I}_{2+d} & 0 \end{pmatrix}, \qquad 0 \in \mathbb{R}^{(2+d)\times(2+d+1)}$$

where d denotes the number of space dimensions. For all equilibrium states $\mathcal{M}(w)$ the relation $R(\mathcal{M}(w)) = 0$ is satisfied and the physical variables $w = (\rho, \rho\mathbf{u}, E)^T$ are recovered by $w = Q\mathcal{M}(w)$ and the flux function by $f(w) = Q(\widetilde{F}(\mathcal{M}(w)))$.

5.3 Time semi-discrete scheme

To avoid the very restrictive CFL condition that would arise when using an explicit scheme, we will construct an IMEX scheme such that the CFL condition is independent of the Mach number. Therefore, we split the flux function $\widetilde{F}(W)$ and source term $S(W)$ in (5.16) in the following way:

$$\partial_t W + \nabla \cdot F(W) + \frac{1}{M^2}\nabla \cdot G(W) = S_E(W) + \frac{1}{M^2}S_I(W) + \frac{1}{\varepsilon}R(W). \tag{5.20}$$

Therein, $F(W)$ and $S_E(W)$ will be treated explicitly while $G(W)$ and $S_I(W)$ contain the implicit terms. The functions $F(W)$, $S_E(W)$, $G(W)$ and $S_I(W)$ are chosen carfully to avoid the need to invert a huge non-linear system while the waves associated to the explicitly treated flux function $F(W)$ should be independent of the Mach number. We propose the following definition of flux functions and source terms:

$$\begin{aligned}
F(W) &= \begin{pmatrix} \rho\mathbf{u} \\ \rho\mathbf{u}\otimes\mathbf{u} + \pi\mathbb{I} + \dfrac{1-M^2}{M^2}\psi\mathbb{I} \\ \left(E + M^2\pi + (1-M^2)\psi\right)\mathbf{u} \\ \rho\pi\mathbf{u} + a^2\mathbf{u} \\ \rho\mathbf{u}\otimes\hat{\mathbf{u}} \\ \rho\psi\mathbf{u} \\ \rho Z\mathbf{u} \end{pmatrix}, \quad S_E(W) = \begin{pmatrix} 0 \\ \dfrac{1}{M^2}\dfrac{\rho}{\alpha}\nabla Z \\ \dfrac{\rho}{\alpha}\mathbf{u}\cdot\nabla Z \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \\
G(W) &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \psi\mathbb{I} \\ a^2M^2\hat{\mathbf{u}} \\ 0 \end{pmatrix}, \qquad S_I(W) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \dfrac{\rho}{\alpha}\nabla Z \\ 0 \\ 0 \end{pmatrix}.
\end{aligned} \tag{5.21}$$

Since the implicit operators G and S_I are sparse and do not act on the physical variables, we perform the implicit before the explicit step which is then followed by the integration of the relaxation source term. This procedure is summarized in the following sequence of operations:

$$\text{Implicit: } \partial_t W + \frac{1}{M^2}\nabla \cdot G(W) = \frac{1}{M^2}S_I(W), \tag{5.22}$$

$$\text{Explicit: } \partial_t W + \nabla \cdot F(W) = S_E(W), \tag{5.23}$$

$$\text{Projection: } \partial_t W = \frac{1}{\varepsilon}R(W). \tag{5.24}$$

The projection step (5.24) ensures that the updated variables at the new time level W^{n+1} are in relaxation equilibrium (5.19) corresponding to $\varepsilon = 0$. Following the same line of arguments as in the previous chapters, we can set the update of the relaxation variables directly as

$$\pi^{n+1} = p^{(2)}, \quad \hat{\mathbf{u}}^{n+1} = \mathbf{u}^{(2)}, \quad \psi^{n+1} = p^{(2)}, \quad Z^{n+1} = \beta, \tag{5.25}$$

where the superscript (2) denotes the solution after the second step (5.23). Considering the whole state vector W, we can summarize the projection step by $W^{n+1} = W^{(2),\text{eq}}$. The formal time semi-discrete scheme is then given by

$$W^{(1)} - W^n + \frac{\Delta t}{M^2}\nabla \cdot G\left(W^{(1)}\right) = \frac{\Delta t}{M^2} S_I\left(W^{(1)}\right), \tag{5.26}$$

$$W^{(2)} - W^{(1)} + \Delta t\ \nabla \cdot F\left(W^{(1)}\right) = \Delta t\ S_E\left(W^{(1)}\right), \tag{5.27}$$

$$W^{n+1} = W^{(2),\text{eq}}. \tag{5.28}$$

To construct a well-balanced scheme, we have to preserve the hydrostatic equilibria of the time semi-discrete scheme (5.26)-(5.28) which are described by

$$\text{Implicit} \begin{cases} \hat{\mathbf{u}}^{(1)} & = 0, \\ \dfrac{1}{M^2}\nabla\psi^{(1)} & = \dfrac{1}{M^2}\dfrac{\rho^{(1)}}{\alpha}\nabla Z^{(1)}, \end{cases} \tag{5.29}$$

$$\text{Explicit} \begin{cases} \mathbf{u}^{(1)} & = 0, \\ \nabla\pi^{(1)} + \dfrac{1-M^2}{M^2}\nabla\psi^{(1)} & = \dfrac{1}{M^2}\dfrac{\rho^{(1)}}{\alpha}\nabla Z^{(1)}. \end{cases} \tag{5.30}$$

From (5.29) and (5.30) we see that if the implicit step is well-balanced, then the hydrostatic equation for the explicit step reduces to solving

$$\begin{cases} \mathbf{u}^{(1)} & = 0, \\ \nabla\pi^{(1)} & = \dfrac{\rho^{(1)}}{\alpha}\nabla Z^{(1)}, \end{cases}$$

which is independent of the Mach number. The aim is to ensure that the explicit as well as the implicit step are well-balanced to obtain an overall well-balanced IMEX scheme.

5.3.1 Mach number expansion of $\psi^{(1)}$

Due to the sparse structure of the flux function G and the source term S_I in (5.21), the implicit part reduces to solving only two coupled equations in the relaxation variables $\hat{\mathbf{u}}, \psi$ given by

$$\begin{aligned} &\partial_t(\rho\hat{\mathbf{u}}) + \frac{1}{M^2}\nabla\psi = \frac{1}{M^2}\kappa\nabla Z, \\ &\partial_t(\rho\psi) + a^2\nabla\cdot\hat{\mathbf{u}} = 0, \end{aligned} \tag{5.31}$$

with the eigenvalues $\tilde{\lambda}_M^{\pm} = \pm\frac{a}{M\rho}$, where we defined $\kappa = \rho/\alpha$. As done in [23, 79, 8], we rewrite the coupled system (5.31) into a single equation with an elliptic operator for ψ starting from the time-semi-discrete scheme

$$\frac{\rho^{(1)} - \rho^n}{\Delta t} = 0, \tag{5.32}$$

$$\frac{(\rho\hat{\mathbf{u}})^{(1)} - (\rho\hat{\mathbf{u}})^n}{\Delta t} + \frac{1}{M^2}\nabla\psi^{(1)} - \frac{1}{M^2}\kappa^n\nabla Z^n = 0, \tag{5.33}$$

$$\frac{(\rho\psi)^{(1)} - (\rho\psi)^n}{\Delta t} + a^2\nabla\cdot\hat{\mathbf{u}}^{(1)} = 0. \tag{5.34}$$

Note, that Z is a relaxation variable approximating the equilibrium pressure β and is not updated in the implicit step and therefore appears at time level t^n. From the density equation (5.32) it follows that $\rho^{(1)} = \rho^n$. Together we have $\kappa^{(1)} = \frac{\rho^n}{\alpha} = \kappa^n$. Inserting (5.33) into (5.34) we have

$$\begin{aligned}\psi^{(1)} - \Delta t^2 a^2 \tau^n \nabla\cdot(\tau^n\frac{1}{M^2}\nabla\psi^{(1)}) = \psi^n - \Delta t^2 a^2\tau^n\nabla\cdot(\tau^n\frac{\kappa^n}{M^2}\nabla\beta)\\ - \Delta t a^2\tau^n\nabla\cdot\mathbf{u}^n,\end{aligned} \tag{5.35}$$

where we have used $\tau = 1/\rho$. Since the data at time t^n is in relaxation equilibrium, we can set $\hat{\mathbf{u}}^n = \mathbf{u}^n$ and $Z^n = \beta$ on the right hand side of (5.35). Note that, the update (5.35) is linear in ψ.

Now we analyse the implicit update of $\psi^{(1)}$ with respect to the Mach number. We assume that the initial data is well-prepared, that is $w^n \in \Omega_{wp}^{\alpha\beta}$ as defined in (5.15). To preserve the scaling of the pressure, we define the following boundary conditions for ψ on a computational domain D

$$\left.\begin{aligned}\nabla\psi_0^{(1)} &= \nabla p_0^n\\ \nabla\psi_1^{(1)} &= \nabla p_1^n\end{aligned}\right\} \text{ on } \partial D. \tag{5.36}$$

Inserting the Mach number expansion according to $\Omega_{wp}^{\alpha\beta}$ for well-prepared data into (5.35) and separating the $\mathcal{O}(M^{-2})$ terms we find

$$\begin{cases}\nabla\cdot\left(\tau_0^n\nabla\psi_0^{(1)}\right) = \nabla\cdot(\tau_0^n\nabla p_0^n) & \text{in } D\\ \nabla\psi_0^{(1)} = \nabla p_0^n & \text{on } \partial D\end{cases}. \tag{5.37}$$

This boundary value problem has the unique solution $\nabla\psi_0^{(1)} = \nabla p_0^n$ on the whole domain $\overline{D}$. Substituting the Mach number expansions of ψ and τ and collecting the $\mathcal{O}(M^{-1})$ terms leads to

$$\begin{aligned}&\tau_1^n\nabla\cdot\left(\tau_0^n\nabla\psi_0^{(1)}\right) + \tau_0^n\nabla\cdot\left(\tau_1^n\nabla\psi_0^{(1)} + \tau_0^n\nabla\psi_1^{(1)}\right)\\ &\qquad = \tau_1^n\nabla\cdot\left(\tau_0^n\frac{\rho_0^n}{\alpha}\nabla\beta\right) + \tau_0^n\nabla\cdot\left(\tau_1^n\frac{\rho_0^n}{\alpha}\nabla\beta + \tau_0^n\frac{\rho_1^n}{\alpha}\nabla\beta\right).\end{aligned} \tag{5.38}$$

Due to the well-prepared data, we have the relation $\rho_1 \frac{\nabla\beta}{\alpha} = \nabla p_1$ from (5.15). Then we can simplify the equation (5.38) using $\nabla\psi_0^{(1)} = \nabla p_0^n$ to

$$\begin{cases} \nabla \cdot \left(\tau_0^n \nabla \psi_1^{(1)}\right) = \nabla \cdot (\tau_0^n \nabla p_1) & \text{in } D \\ \nabla \psi_1^{(1)} = \nabla p_1^n & \text{on } \partial D \end{cases} \tag{5.39}$$

which has the unique solution $\nabla\psi_1^{(1)} = \nabla p_1^n$ on the whole domain $\overline{D}$. As a last step we look at the $\mathcal{O}(M^0)$ terms and find using the results from (5.37) and (5.39) that

$$\nabla \cdot \left(\tau_0^n \nabla \psi_2^{(1)}\right) = \nabla \cdot \left(\tau_0^n \frac{\rho_2^n}{\alpha} \nabla\beta\right) \quad \text{in } D.$$

This means the first two terms in the expansion of $\psi^{(1)}$ fulfil the hydrostatic equilibrium (5.11), (5.12). This proves that the pressure $\psi^{(1)}$ has the correct asymptotic behaviour.

5.3.2 Asymptotic preserving property

Having established the Mach number expansion of $\psi^{(1)}$, we can show now that the time semi-discrete scheme (5.26) - (5.28) for $M \to 0$ coincides with a time-discretization of the limit equations (5.13) and that the scheme preserves the set of well-prepared data $\Omega_{wp}^{\alpha\beta}$. We start by inserting the Mach number expansions given in (5.10) into (5.27). Then we find for the zero order terms in the density, momentum and energy equation

$$\begin{aligned} \rho_0^{n+1} - \rho_0^n + \Delta t \ \nabla \cdot \rho_0^n \mathbf{u}_0^n &= 0, \\ \rho_0^{n+1}\mathbf{u}_0^{n+1} - \rho_0^n \mathbf{u}_0^n + \Delta t \ \left(\rho_0^n \mathbf{u}_0^n \otimes \mathbf{u}_0^n + \nabla \psi_2^{(1)}\right) &= \Delta t \frac{\rho_2^n}{\alpha} \nabla\beta, \\ \rho_0^{n+1} e_0^{n+1} - \rho_0^n e_0^n + \Delta t \ \left(\nabla \cdot \mathbf{u}_0^n \left(\rho_0^n e_0^n + \psi_0^{(1)}\right)\right) &= \Delta t \ \frac{\rho_0}{\alpha} \mathbf{u}_0^n \cdot \nabla\beta. \end{aligned}$$

We can simplify the equations by using $\nabla\psi_0^{(1)} = \nabla p_0^n$ and well-prepared data $w^n \in \Omega_{wp}^{\alpha\beta}$:

$$\begin{aligned} \rho_0^{n+1} - \rho_0^n &= 0, \\ \mathbf{u}_0^{n+1} - \mathbf{u}_0^n + \Delta t \ \left(\mathbf{u}_0^n \cdot \nabla \mathbf{u}_0^n + \frac{\nabla \psi_2^{(1)}}{\rho_0^n}\right) &= \Delta t \frac{\rho_2^n}{\rho_0^n \ \alpha} \nabla\beta, \\ p_0^{n+1} - p_0^n &= 0. \end{aligned}$$

From the first and the last equation we see that ρ_0 and p_0 do not change in time and looking at the $\mathcal{O}(M^1)$ terms in the energy equation we have $p_1^{n+1} = p_1^n + \mathcal{O}(\Delta t)$. This means the pressure and density at t^{n+1} are still well-prepared up to perturbations of Δt. Next, we analyse the divergence free property of $\mathbf{u}_0^{n+1}$ and $\rho_0^{n+1}\mathbf{u}_0^{n+1}$. This is done by applying the divergence operator on the momentum equation and simplifying using (5.3.1). We obtain

$$\begin{aligned} \nabla \cdot \mathbf{u}_0^{n+1} &= \Delta t \ \nabla \cdot (-\mathbf{u}_0^n \cdot \nabla \mathbf{u}_0^n) = \mathcal{O}(\Delta t), \\ \nabla \cdot \left(\rho_0^{n+1}\mathbf{u}_0^{n+1}\right) &= \Delta t \ \nabla \cdot \left(-\rho_0^n \mathbf{u}_0^n \cdot \nabla \mathbf{u}_0^n - \nabla \psi_2^{(1)} + \frac{\rho_2^n}{\alpha} \nabla\beta\right) = \mathcal{O}(\Delta t). \end{aligned}$$

For showing the orthogonality condition for $\mathbf{u}_0^{n+1}$ we multiply the momentum equation by $\frac{\nabla\beta}{\alpha}$ and obtain

$$\mathbf{u}_0^{n+1}\cdot\frac{\nabla\beta}{\alpha^n}=\Delta t\,\left(-\mathbf{u}_0^n\cdot\nabla\mathbf{u}_0^n-\frac{\nabla\psi_2^{(1)}}{\rho_0^n}+\frac{\rho_2^n}{\rho_0^n\,\alpha}\nabla\beta\right)\cdot\frac{\nabla\beta}{\alpha}=\mathcal{O}(\Delta t).$$

Therefore all three conditions are satisfied up to a perturbation in Δt. An analogue estimate for the homogeneous case can be found in the method proposed in [23]. This analysis yields the following result about the asymptotic preserving property.

Theorem 5.2 (AP property)**.** *For well-prepared initial data $w^n\in\Omega_{wp}^{\alpha\beta}$ and under the boundary conditions (5.36) the time semi-discrete scheme (5.26)- (5.28) is asymptotic preserving when $M\to 0$ in the sense that if $w^n\in\Omega_{wp}^{\alpha\beta}$ then also $w^{n+1}\in\Omega_{wp}^{\alpha\beta}$ up to $\mathcal{O}(\Delta t)$ and in the limit $M\to 0$ the time semi-discrete scheme is a consistent time discretization of the limit equations (5.13) within $\mathcal{O}(\Delta t)$ terms.*

We remark that the analysis still holds if instead of $\Omega_{wp}^{\alpha\beta}$ the original well-prepared set Ω_{wp} is used.

5.4 Derivation of the fully discrete scheme

The derivation of the fully discrete scheme is done in one spatial direction for simplicity, but it can be extended straightforwardly to multiple dimensions considering dimensional split Riemann problems in the explicit part and discretizing the expressions

$$\begin{aligned}&\nabla\cdot(\tau\nabla(\cdot))=\partial_{x_1}(\tau\partial_{x_1}(\cdot))+\cdots+\partial_{x_d}(\tau\partial_{x_d}(\cdot))\text{ and}\\&\nabla\cdot\mathbf{u}=\partial_{x_1}u_1+\cdots+\partial_{x_d}u_d\end{aligned}\tag{5.40}$$

with $\mathbf{u}=(u_1,\dots,u_d)$ component-wise in the implicit step. We use a uniform Cartesian grid on a computational domain D divided in N cells $C_i=(x_{i-1/2},x_{i+1/2})$ of step size Δx. We use a standard finite volume setting, where we define at time t^n the piecewise constant functions $w(x,t^n)=w_i^n,\ \text{for } x\in C_i$.

5.4.1 Well-balanced property of the implicit part

Applying central differences in (5.35) we obtain

$$\begin{aligned}\psi_i^{(1)}-\frac{\Delta t^2}{\Delta x^2}\frac{a^2}{M^2}\tau_i^n\left(\tau_{i+1/2}^n(\psi_{i+1}^{(1)}-\psi_i^{(1)})-\tau_{i-1/2}^n(\psi_i^{(1)}-\psi_{i-1}^{n+1})\right)=\\ \psi_i^n-\frac{\Delta t^2}{\Delta x^2}\frac{a^2}{M^2}\tau_i^n\left(\tau_{i+1/2}^n\kappa_{i+1/2}^n(\beta_{i+1}-\beta_i)-\tau_{i-1/2}^n\kappa_{i-1/2}^n(\beta_i-\beta_{i-1})\right)\\ -\frac{\Delta t}{2\Delta x}a^2\left(u_{i+1}^n-u_{i-1}^n\right),\end{aligned}\tag{5.41}$$

where $\tau_{i+1/2}=\frac{1}{2}(\tau_{i+1}+\tau_i)$. To fully determine the implicit update, we have to define $\kappa_{i+1/2}$. Its discretization is important to obtain a well-balanced implicit step, as given in the following result.

Lemma 5.3 (Well-balancedness of the implicit part)**.** *Let the initial condition w_i^n be well-balanced, that is*

$$u_i = 0, \quad \frac{\rho_i^n}{\alpha_i} = 1, \quad \frac{p_i^n}{\beta_i} = 1. \tag{5.42}$$

If the function κ is discretized such that in the hydrostatic equilibrium holds

$$\kappa_{i+1/2} = 1, \tag{5.43}$$

then it is $\psi_i^{(1)} = \psi_i^n$ for all cells $i = 1, \dots N$, that means (5.26) is well-balanced in the sense that $W^{(1)}$ fulfils (5.29).

Proof. From the condition (5.42) we have $\kappa_{i+1/2} = 1$. At time level t^n we know that $\psi^n = p^n$. Therefore we can write

$$\psi_{i+1}^n - \psi_i^n = \beta_{i+1} - \beta_i = \kappa_{i+1/2}^n(\beta_{i+1} - \beta_i). \tag{5.44}$$

Using $u = 0$ and inserting (5.44) into (5.41), we have

$$\begin{aligned}\psi_i^{(1)} - \frac{\Delta t^2}{\Delta x^2}\frac{a^2}{M^2}\tau_i^n \left(\tau_{i+1/2}^n(\psi_{i+1}^{(1)} - \psi_i^{(1)}) - \tau_{i-1/2}^n(\psi_i^{(1)} - \psi_{i-1}^{n+1})\right) = \\ \psi_i^n - \frac{\Delta t^2}{\Delta x^2}\frac{a^2}{M^2}\tau_i^n \left(\tau_{i+1/2}^n(\psi_{i+1}^n - \psi_i^n) - \tau_{i-1/2}^n(\psi_i^n - \psi_{i-1}^n)\right).\end{aligned} \tag{5.45}$$

Define the tridiagonal coefficient matrix A by

$$A = \operatorname{diag}(-\mu\tau_i^n\tau_{i-1/2}^n, 1 + \mu\tau_i^n(\tau_{i+1/2}^n + \tau_{i-1/2}^n), -\mu\tau_i^n\tau_{i+1/2}^n),$$

where $\mu = \frac{\Delta t^2}{\Delta x^2}\frac{a^2}{M^2}$. Then we can write (5.45) as

$$A\psi^{(1)} = A\psi^n \Leftrightarrow A(\psi^{(1)} - \psi^n) = 0. \tag{5.46}$$

Since the matrix A is strict diagonal dominant it is invertible. Then we have from (5.46) that $\psi_i^{(1)} = \psi_i^n$ for all $i = 1, \dots, N$. The proof can be extended to d dimensions using (5.40) for the space discretization. In d dimensions the coefficient matrix A is an invertible strict diagonal dominant banded Matrix with $2d+1$ diagonals. Therefore the results holds also in d dimensions. □

In the following we will use a second order accurate discretization of $\kappa_{i+1/2}$ that fulfils (5.43) and is given by

$$\kappa_{i+1/2} = \frac{1}{2}\left(\frac{\rho_{i+1}}{\alpha_{i+1}} + \frac{\rho_i}{\alpha_i}\right). \tag{5.47}$$

5.4.2 Godunov type finite volume scheme

Now, we consider the explicit step (5.27) using the explicit operators F and S_E defined in (5.21). In one dimension, we can write

$$\begin{aligned}
\partial_t \rho + \partial_x \rho u &= 0 \\
\partial_t(\rho u) + \partial_x(\rho u^2 + \pi + \frac{1-M^2}{M^2}\psi) &= \frac{1}{M^2}\kappa \partial_x Z \\
\partial_t E + \partial_x((E + M^2\pi + (1-M^2)\psi)u) &= u\kappa\partial_x Z \\
\partial_t(\rho\pi) + \partial_x((\rho\pi + a^2)u) &= 0 \\
\partial_t(\rho\hat{u}) + \partial_x(\rho u\hat{u}) &= 0 \\
\partial_t(\rho\psi) + \partial_x(\rho\psi u) &= 0 \\
\partial_t(\rho Z) + \partial_x(\rho Z u) &= 0.
\end{aligned} \tag{5.48}$$

The derivation of the Godunove type finite volume scheme follows closely the steps given eg. in [79, 7, 28, 78, 80]. The omitted proofs to the results given in this section can be done analogously following those references or the steps for similar results described in previous chapters. To construct a Riemann solver for (5.48), we follow [28] and include the source term in the flux formulation. To calculate the Riemann invariants given in Lemma 5.4, we rewrite (5.48) in non-conservative form using the primitive variables $(\rho, u, e, \pi, \hat{u}, \psi, Z)$. Since Riemann invariants are invariant under change of variables, they are the same as for the equations in conservation form.

Lemma 5.4. *System (5.48) admits the linear degenerate eigenvalues $\lambda^{\pm} = u \pm \frac{a}{\rho}$ and $\lambda^u = u$, where the eigenvalue λ^u has multiplicity 5. The relaxation parameter a as well as the eigenvalues are independent of the Mach number M. The Riemann invariants with respect to λ^u are*

$$I_1^u = u, \quad I_2^u = M^2\pi + (1-M^2)\psi - \kappa Z$$

and with respect to $\lambda^{\pm}$

$$I_1^{\pm} = u \pm \frac{a}{\rho}, \quad I_2^{\pm} = \pi + \frac{a^2}{\rho}, \quad I_3^{\pm} = e - \frac{M^2}{2a^2}\pi^2 - \frac{1-M^2}{a^2}\pi\psi,$$

$$I_4^{\pm} = \hat{u}, \quad I_5^{\pm} = \psi, \quad I_6^{\pm} = Z.$$

We will follow the theory of Harten, Lax and van Leer [41] for deriving an approximate Riemann solver $W_{\mathcal{RS}}\left(\frac{x}{t}; W_L^{(1)}, W_R^{(1)}\right)$ based on the states $W^{(1)}$ after the implicit step keeping κ fixed while computing the Riemann solution. Due to the linear-degeneracy from Lemma 5.4, the structure of the approximate Riemann solver is given as follows

$$W_{\mathcal{RS}}\left(\frac{x}{t}; W_L^{(1)}, W_R^{(1)}\right) = \begin{cases} W_L^{(1)} & \frac{x}{t} < \lambda^-, \\ W_L^* & \lambda^- < \frac{x}{t} < \lambda^u, \\ W_R^* & \lambda^u < \frac{x}{t} < \lambda^+, \\ W_R^{(1)} & \lambda^+ < \frac{x}{t}. \end{cases} \tag{5.49}$$

To compute the intermediate states $W_{L,R}^*$, we use the Riemann invariants as given in Lemma 5.4.

Lemma 5.5. *Consider an initial value problem with initial data*

$$W^0(x) = \begin{cases} W_L^{(1)} & x < 0, \\ W_R^{(1)} & x > 0. \end{cases}$$

Then, the solution consists of four constant states separated by contact discontinuities with the structure given in (5.49). Dropping the superscript (1), a solution for the intermediate states W_L^*, W_R^* in dependence of κ, with $u^* = u_L^* = u_R^*$, is given by

$$\begin{aligned}
u^* &= \frac{1}{2}(u_L + u_R) - \frac{1}{2a}\left(\pi_R - \pi_L + \frac{1-M^2}{M^2}(\psi_R - \psi_L) - \frac{\kappa}{M^2}(Z_R - Z_L)\right), \\
\pi_L^* &= \frac{1}{2}(\pi_L + \pi_R) - \frac{a}{2}(u_R - u_L) + \frac{1-M^2}{2M^2}(\psi_R - \psi_L) - \frac{\kappa}{2M^2}(Z_R - Z_L), \\
\pi_R^* &= \frac{1}{2}(\pi_L + \pi_R) - \frac{a}{2}(u_R - u_L) - \frac{1-M^2}{2M^2}(\psi_R - \psi_L) + \frac{\kappa}{2M^2}(Z_R - Z_L), \\
\frac{1}{\rho_L^*} &= \frac{1}{\rho_L} + \frac{1}{a^2}(\pi_L - \pi_L^*), \\
\frac{1}{\rho_R^*} &= \frac{1}{\rho_R} + \frac{1}{a^2}(\pi_R - \pi_R^*), \\
e_L^* &= e_L - \frac{1}{2a^2}(\pi_L^2 - (\pi_L^*)^2 + (1-M^2)(\pi_L - \pi_L^*)\psi_L), \\
e_R^* &= e_R - \frac{1}{2a^2}(\pi_R^2 - (\pi_R^*)^2 + (1-M^2)(\pi_R - \pi_R^*)\psi_R), \\
\psi_{L,R}^* &= \psi_{L,R}, \\
\hat{u}_{L,R}^* &= \hat{u}_{L,R}, \\
Z_{L,R}^* &= Z_{L,R}.
\end{aligned} \tag{5.50}$$

Having established the structure of the Riemann solver, we can show that it is preserving hydrostatic equilibria.

Lemma 5.6 (Well-balancedness of Riemann Solver)**.** *Let the initial condition w_L^n, w_R^n be given in hydrostatic equilibrium (5.42). Let the function κ be defined as in (5.43). Then the intermediate states (5.50) satisfy*

$$W_L^{(1)*} = W_L^{(1)}, \; W_R^{(1)*} = W_R^{(1)}$$

that is, the approximate Riemann solver as defined in Lemma 5.5 is at rest.

Proof. From Lemma 5.3, we know that $\psi^{(1)} = p^n$ and satisfies

$$\psi_L^{(1)} - \psi_R^{(1)} = \kappa(Z_L^{(1)} - Z_R^{(1)}). \tag{5.51}$$

We also know that $\pi_{L,R}^{(1)} = \pi_{L,R}^n = p_{L,R}^n$ and since w^n is fulfilling (5.42) and with (5.43) we have $\pi_L^{(1)} - \pi_R^{(1)} = \kappa(Z_L^{(1)} - Z_R^{(1)})$. Then we have

$$\begin{aligned}
&\pi_R^{(1)} - \pi_L^{(1)} + \frac{1-M^2}{M^2}(\psi_R^{(1)} - \psi_L^{(1)}) - \frac{\kappa}{M^2}(Z_R^{(1)} - Z_L^{(1)}) = \\
&\pi_R^{(1)} - \pi_L^{(1)} - \kappa(Z_R^{(1)} - Z_L^{(1)}) = 0.
\end{aligned}$$

Since $u^{(1)}_{L,R} = u^n_{L,R} = 0$, we find $u^{(1)*} = 0$. With $u^{(1)*} = 0$ and (5.51) and the fact that $\psi^n = p^n = \pi^n = \pi^{(1)}$, we can write

$$\begin{aligned}\pi^*_L &= \frac{1}{2}(\pi^{(1)}_L + \pi^{(1)}_R) + \frac{1-M^2}{2M^2}(\psi^{(1)}_R - \psi^{(1)}_L) - \frac{\kappa}{2M^2}(Z^{(1)}_R - Z^{(1)}_L)\\ &= \frac{1}{2}(\pi^{(1)}_L + \pi^{(1)}_R) - \frac{1}{2}(\pi^{(1)}_R - \pi^{(1)}_L)\\ &= \pi_L.\end{aligned}$$

Analogously follows $\pi^*_R = \pi_R$. Then it follows directly from the intermediate states (5.50) that $\rho^*_L = \rho_L$, $\rho^*_R = \rho_R$ and $e^*_L = e_L$, $e^*_R = e_R$. □

Another important property is that the density and pressure remain positive during the simulation. This is equivalent to preserving the following invariant domain

$$\Omega_{phy} = \{w \in \Omega, \rho > 0, e > 0\}.$$

We show that the Riemann solver preserves Ω_{phy}.

Lemma 5.7 (Positivity preserving property of Riemann Solver)**.** *Suppose the initial data $W^{(1)}_{L,R}$ is composed of $w^{(1)}_{L,R} \in \Omega_{phy} \cap \Omega^{\alpha,\beta}_{wp}$ and $\psi^{(1)}$ satisfies the boundary conditions (5.36). Then the solution of the Riemann problem given by $QW_{\mathcal{RS}}(\frac{x}{t}; W^{(1)}_L, W^{(1)}_R)$ is contained in Ω_{phy} for a relaxation parameter a sufficiently large but independent of M.*

Proof. The proof for the intermediate states for the density is done analoguosly to the proof of Lemma 3.5 or can be taken from [79, 78]. After the implicit step we have $u^{(1)} = u^n$, $\pi^{(1)} = \pi^n$ and $Z^{(1)} = Z^n$. We use the following notation $\Delta(\cdot) = (\cdot)_R - (\cdot)_L$. For the internal energy, the intermediate state $\pi^{(1)*}_L$ is inserted into e^*_L and we have

$$\begin{aligned}e^{(1)*}_L = \; & e^n_L + \frac{1}{8}\Delta u^2 + \frac{1}{4a}\Delta u^n\left(\Delta\pi^n + 2\pi^n_L - \Delta\psi^{(1)} + \frac{1}{M^2}H^{(1)} + (1-M^2)\psi^{(1)}_L\right)\\ & + \frac{1}{2a^2}\left(-\;(\pi^n_L)^2 + \frac{1}{4}\left(\pi^n_L + \pi^n_R - \Delta\psi^{(1)} + \frac{1}{M^2}H^{(1)}\right)^2\right.\\ & \left. + \frac{1}{2}\,\psi^{(1)}_L\left(1-M^2\right)\left(\Delta\pi^n - \Delta\psi^{(1)} + \frac{1}{M^2}H^{(1)}\right)\right) \qquad (5.52)\end{aligned}$$

where we have defined $H^{(1)} = (\psi^{(1)}_R - \psi^{(1)}_L) - \kappa(Z^n_R - Z^n_L)$. We know from the Mach number analysis in Section 4.3 that $\psi^{(1)}$ preserves the hydrostatic equilibrium up to a perturbation of M^2, thus $H^{(1)} = \mathcal{O}(M^2)$. Therefore we find a relaxation parameter $a > \rho c$ independent of M that can control the potentially negative terms in (5.52) and we have $e^{(1)*}_L > 0$. Analogously we obtain $e^{(1),*}_R > 0$. □

Analogously to the homogeneous case given in Chapter 4, we consider the local Riemann problem at the interface $x_{i+1/2}$ consisting of data $W^{(1)}_i, W^{(1)}_{i+1}$ after the implicit step.

To define a first order scheme, we consider the Riemann problem given by

$$\begin{cases} \partial_t W + \partial_x F(W) = S_E(W), \\ W^0(x) = \begin{cases} W_i^{(1)} & \text{if } x < x_{i+1/2}, \\ W_{i+1}^{(1)} & \text{if } x > x_{i+1/2}. \end{cases} \end{cases} \tag{5.53}$$

The problematic of solving the IVP (5.53) is very close to solving the Riemann problem for the Euler equations with gravity in the compressible regime described in Section 3.4.1. Therefore, the derivation of the update for the explicit step follows the steps and argumentation given therein. First we note that the intermediate stat u^* in the Riemann solution (5.50) is depending on the source term $\kappa(Z_{i+1} - Z_i)$. Therefore, the contribution of the source term to the fluxes over the interface depends on the positioning of the middle wave u^*. Due to this connection, we choose to define the fluxes at the interface including this source contribution which leads to a compact formulation of the explicit update. It is given by

$$W_i^{(2)} = W_i^{(1)} + \frac{\Delta t}{\Delta x}\left(\mathcal{F}^-\left(W_i^{(1)}, W_{i+1}^{(1)}\right) - \mathcal{F}^+\left(W_{i-1}^{(1)}, W_i^{(1)}\right)\right), \tag{5.54}$$

where the left and right numerical fluxes are defined by

$$\begin{aligned} &(\mathcal{F}^-(W_L, W_R), \mathcal{F}^+(W_L, W_R)) = \\ &\qquad \begin{cases} (F(W_L), & F(W_L) + \bar{S}(W_L, W_R) & \text{if} \quad \lambda_L^- > 0 \\ (F(W_L^*), & F(W_L^*) + \bar{S}(W_L, W_R) & \text{if} \quad u_1^* > 0 > \lambda_L^- \\ (F(W_L^*), & F(W_R^*)) & \text{if} \quad u_1^* = 0 \\ (F(W_R^*) - \bar{S}(W_L, W_R), & F(W_R^*)) & \text{if} \quad \lambda_R^+ > 0 > u_1^* \\ (F(W_R) - \bar{S}(W_L, W_R), & F(W_R)) & \text{if} \quad \lambda_R^+ < 0 \end{cases}, \end{aligned} \tag{5.55}$$

where $\lambda_L^- = u_L + a/\rho_L$ and $\lambda_R^+ = u_R + a/\rho_R$ and

$$\begin{aligned} \bar{S}(W_L, W_R) &= \mathcal{S}(W_L, W_R)(Z_R - Z_L) \\ &= \left(0, \frac{1}{2}\left(\frac{\rho_L}{\alpha_L} + \frac{\rho_R}{\alpha_R}\right), u^* \frac{1}{2}\left(\frac{\rho_L}{\alpha_L} + \frac{\rho_R}{\alpha_R}\right), 0, 0, 0, 0\right)^T (Z_R - Z_L). \end{aligned}$$

Since the source term is included into the flux definition, the left and right fluxes in general do not coincide and it is $\mathcal{F}_{i+1/2}^- \neq \mathcal{F}_{i+1/2}^+$. To avoid interactions between the approximate Riemann solutions originating from the interfaces $x_{i+1/2}$, we have a CFL restriction on the time step of

$$\frac{\Delta t}{\Delta x} \max_i \left| u_i^n \pm \frac{a}{\rho_i^n} \right| \leq \frac{1}{2} \tag{5.56}$$

which is independent of the Mach number. The explicit update (5.54) is consistent with the equations in the explicit step (5.23). To see this, the update (5.54) can be reformulated with an upwinded source term depending on the sign of u^* analogue to the numerical scheme (3.51) for the compressible Euler equations with gravity.

The complete first order IMEX scheme is given by (5.41), (5.54) and (5.25). Looking at the update of the relaxation variables, we see that the explicit update is immediately overwritten in the subsequent projection step (5.25). To avoid the unnecessary update of

$d+3$ relaxation variables in the explicit step (5.54), we can directly give the update only for the physical variables w in dependence of the pressure $\psi^{(1)}$. Defining the numerical fluxes

$$\mathfrak{f}^-(w_L,\psi_L,w_R,\psi_R) = Q\mathcal{F}^-(W_L,W_R), \quad \mathfrak{f}^+(w_L,\psi_L,w_R,\psi_R) = Q\mathcal{F}^+(W_L,W_R), \tag{5.57}$$

the update for the physical variables can be written as

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(\mathfrak{f}^-\left(w_i^n,\psi_i^{(1)},w_{i+1}^n,\psi_{i+1}^{(1)}\right) - \mathfrak{f}^+\left(w_{i-1}^n,\psi_{i-1}^{(1)},w_i^n,\psi_i^{(1)}\right)\right). \tag{5.58}$$

We want to remark that even though through the relaxation procedure the number of equations in the model was more than doubled, we effectively only solve an explicit scheme for the physical variables plus one linear scalar implicit equation.

The IMEX scheme for which we show the well-balanced and positivity property, is now based on the three parts beginning with obtaining $\psi^{(1)}$ from the implicit equation (5.41), then updating the physical variables (5.58) and finally projecting the relaxation variables onto equilibrium (5.25).

Theorem 5.8 (Well-balanced property 1)**.** *Let w_i^n for all cells $C_i, i \in \{1, N\}$ be given in hydrostatic equilibrium (5.42) and κ be defined as in (5.43). Then the first order scheme given by the steps (5.41),(5.58),(5.25) is well-balanced.*

Proof. Since w^n fulfils the hydrostatic equilibrium, we know from Lemma 5.3 that $W_i^{(1)} = W_i^n$ fulfils the hydrostatic equilibrium. From Lemma 5.6 we have that the approximate Riemann solver at the cell interfaces is at rest. With the definition of the fluxes (5.55), we have

$$\mathcal{F}^+\left(W_{i-1}^{(1)},W_i^{(1)}\right) = F\left(W_i^{(1)}\right), \qquad \mathcal{F}^-\left(W_i^{(1)},W_{i+1}^{(1)}\right) = F\left(W_i^{(1)}\right).$$

Using the formulation (5.58) for the update of the variables w, we have

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x} Q\left(\mathcal{F}^-(W_i^{(1)},W_{i+1}^{(1)}) - \mathcal{F}^+(W_{i-1}^{(1)},W_i^{(1)})\right) = w_i^n.$$

This shows the well-balanced property in one dimension. Since we apply dimensional splitting in the multi-dimensional set-up, the proof can be easily extended by giving the update (5.4.2) as a sum of the flux differences along each dimension. □

Theorem 5.9 (Positivity preserving 1)**.** *Let the initial state be given as*

$$w_i^n \in \Omega = \Omega_{phy} \cap \Omega_{wp}^{\alpha\beta}$$

Then under the Mach number independent CFL condition

$$\frac{\Delta t}{\Delta x}\max_i \left| u_i^n \pm \frac{a}{\rho_i^n}\right| \leq \frac{1}{2},$$

and the boundary conditions (5.36) the numerical scheme defined by (5.41),(5.58) preserves the positivity of density and internal energy, that is $w_i^{n+1} \in \Omega_{phy}$ for a sufficiently large relaxation parameter a independent of M.

An important property for any low Mach scheme is the behaviour of the diffusion. Due to the fact that $\psi^{(1)}$ is still well-prepared after the implicit step, the diffusion of the scheme is of order $\mathcal{O}(M^0)$. The computations are performed analogously to the homogeneous case in Section 4.4.3.

5.4.3 Second order extension

In this section, we give a strategy to extend the first order scheme to second order accuracy such that the well-balanced and the positivity preserving property are maintained. Thereby we combine the hydrostatic and positivity preserving reconstructions from Section 3.4.2 with the second order IMEX time integration scheme described in Section 4.5.

A second order reconstruction in space

The second order extension in space is realized by a modification of the hydrostatic reconstruction combined with the limiting procedure which were both described in Section 3.4.2. We reconstruct on each cell the equilibrium variables $v = (\rho, \mathbf{u}, q, \xi)$ which consist of the density, the velocity field and the modified pressures q and ξ. They are defined by the following transformations

$$\begin{cases} q_{i-1} &= \pi^{(1)}_{i-1} + S_{i-1/2} \\ q_i &= \pi^{(1)}_i \\ q_{i+1} &= \pi^{(1)}_{i+1} - S_{i+1/2} \end{cases}, \qquad \begin{cases} \xi_{i-1} &= \psi^{(1)}_{i-1} + S_{i-1/2} \\ \xi_i &= \psi^{(1)}_i \\ \xi_{i+1} &= \psi^{(1)}_{i+1} - S_{i+1/2} \end{cases}, \tag{5.59}$$

where

$$S_{i+1/2} = \frac{1}{2}\left(\frac{\rho^n_i}{\alpha_i} + \frac{\rho^n_{i+1}}{\alpha_{i+1}}\right)(\beta_{i+1} - \beta_i)$$

denotes the second order discretization of the source term $\kappa \partial_x Z$ with the discretization of κ defined in (5.47). Note that we can write the intermediate velocity u^* at the interface in terms of $S_{i+1/2}$ as follows

$$u^{(1),*}_{i+1/2} = \frac{1}{2}(u^n_i + u^n_{i+1}) - \frac{1}{2a}\left(\pi^n_{i+1} - \pi^n_i - S_{i-1/2} + \frac{1-M^2}{M^2}\left(\psi^{(1)}_{i+1} - \psi^{(1)}_i - S_{i+1/2}\right)\right).$$

We turn back to the computation of the slopes σ_i for the equilibrium variables v which are obtained by applying the minmod limiter (2.75) on each component of v. The interface values in equilibrium variables are then given by

$$v^+_{i-1/2} = v^{(1)}_i - \Delta v, \quad v^-_{i+1/2} = v^{(1)}_i + \Delta v, \tag{5.60}$$

where $\Delta v = \sigma_i \Delta x/2$ denotes the total slope. The interface values in conserved variables are calculated from the equilibrium variables. To obtain the pressure variables, we have the simple relation of

$$\pi^-_{i+1/2} = q^-_{i+1/2}, \quad \pi^+_{i-1/2} = q^+_{i-1/2}, \quad \psi^-_{i+1/2} = \xi^-_{i+1/2}, \quad \psi^+_{i-1/2} = \xi^+_{i-1/2}. \tag{5.61}$$

It can be shown analogously to the proof of Lemma 3.7 that by using the hydrostatic reconstruction to obtain the interface values $W^{\pm}_{i+1/2}$, the Riemann solver (5.49) is at rest, when considering the following initial values

$$W^0(x) = \begin{cases} W^{(1),-}_{i+1/2} & \text{if } x < x_{i+1/2}, \\ W^{(1),+}_{i+1/2} & \text{if } x > x_{i+1/2}. \end{cases} \tag{5.62}$$

To meet the positivity criterion we apply a limiting procedure on the slopes Δv to guarantee $w^-_{i+1/2}, w^+_{i-1/2} \in \Omega_{phy}$. We use the limiter from Section 3.4.2 which we summarize in the following. We can directly apply the limiter design for the compressible regime, since only $\pi^n_i = p^n_i$, but not $\psi^{(1)}$, is important for the positivity for the internal energy at time level t^n. The slope limiter applied on the slopes after the hydrostatic reconstruction $\Delta\bar{v}$ is given by

$$\begin{aligned} \Delta\rho &= \Delta\bar{\rho}, \\ \Delta\mathbf{u} &= \min(1,\omega_2)\Delta\bar{\mathbf{u}}, \\ \Delta\pi &= \pi_i \max\left(-1, \min\left(1, \frac{\Delta\bar{q}}{\pi_i}\right)\right), \end{aligned} \tag{5.63}$$

where

$$\omega_2 = -\frac{\Delta\rho\langle\Delta\bar{\mathbf{u}}, \mathbf{u}_i\rangle}{\rho_i|\Delta\bar{\mathbf{u}}|^2} + \frac{1}{\rho_i|\Delta\bar{\mathbf{u}}|^2}\sqrt{\Delta\rho^2\langle\Delta\bar{\mathbf{u}}, \mathbf{u}_i\rangle^2 + |\Delta\bar{\mathbf{u}}|^2\frac{\rho_i\pi_i}{\gamma-1}}. \tag{5.64}$$

A second order time integration method

To obtain a fully second order scheme in space and time, we combine the second order space reconstruction with the second order time integration scheme (4.62) used for the homogeneous case in Section 4.5. The fully discretized second order scheme is given by

$$\begin{aligned} \overline{w}_i &= w^n - \frac{\Delta t_1}{\Delta x}\left(\mathfrak{f}^-\left(w^-_{i+1/2}, \psi^{(1),-}_{i+1/2}, w^+_{i+1/2}, \psi^{(1),+}_{i+1/2}\right)\right. \\ &\qquad \left. - \mathfrak{f}^+\left(w^-_{i-1/2}, \psi^{(1),-}_{i-1/2}, w^+_{i-1/2}, \psi^{(1),+}_{i-1/2}\right)\right) \\ \overline{\overline{w}}_i &= \overline{w}_i - \frac{\Delta t_2}{\Delta x}\left(\mathfrak{f}^-\left(\overline{w}^-_{i+1/2}, \overline{\psi}^{(1),+}_{i+1/2}, \overline{w}^+_{i+1/2}, \overline{\psi}^{(1),+}_{i+1/2}\right)\right. \\ &\qquad \left. - \mathfrak{f}^+\left(\overline{w}^-_{i-1/2}, \overline{\psi}^{(1),-}_{i-1/2}, \overline{w}^+_{i-1/2}, \overline{\psi}^{(1),+}_{i-1/2}\right)\right) \\ w^{n+1}_i &= \left(1 - \frac{\Delta t}{\Delta t_1 + \Delta t_2}\right) w^n_i + \frac{\Delta t}{\Delta t_1 + \Delta t_2}\overline{\overline{w}}_i. \end{aligned} \tag{5.65}$$

The pressures $\psi^{(1)}$ and $\overline{\psi}^{(1)}$ are obtained by solving the linear implicit system (5.41) with initial data given by $\mathcal{M}(w^n)$ and $\mathcal{M}(\overline{w})$ respectively. The step sizes $\Delta t_{1,2}$ are obtained according to the CFL condition of the respective stages in (5.65). The time integration method (5.65) consists of a convex combination of first order temporal integrators and therefore the second order time integration preserves the AP property shown in Section 5.3.2.

We summarize the well-balanced and positivity preserving property of the second order scheme (5.65) in the following theorems. The proofs are analogous to the ones shown in the Sections 3.5 for the explicit scheme for the compressible Euler equations with gravity and in Section and 4.5 for the all-speed scheme for the homogeneous Euler equations respectively.

Theorem 5.10 (Well-balanced property 2)**.** *Let the initial condition w^n be given in hydrostatic equilibrium (5.42) and κ be defined as in (5.47). Then, using the hydrostatic reconstruction (5.59), the second order scheme given by (5.65) is well-balanced.*

Theorem 5.11 (Positivity property 2)**.** *Let the initial state be given as $w_i^n \in \Omega_{phy} \cap \Omega_{wp}^{\alpha\beta}$ satisfying the boundary conditions (5.36) and the limiting procedure given in (5.63) is used. Then for a sufficiently large relaxation parameter a, under the Mach number independent CFL condition*

$$\frac{\Delta t}{\Delta x}\max_i\left(\left|u_{i-1/2}^{n,+}-\frac{a}{\rho_{i-1/2}^{n,+}}\right|,\left|u_i^n\pm\frac{a}{\rho_i^n}\right|,\left|u_{i+1/2}^{n,-}-\frac{a}{\rho_{i+1/2}^{n,-}}\right|\right)<\frac{1}{2\cdot 3},$$

where d denotes the dimension, the second order scheme (5.65) preserves the domain Ω_{phy}.

Both results can be extended straightforwardly to multiple space dimensions.

5.5 Numerical results

In this section, we give numerical test cases to validate the theoretical properties of the first and second order scheme. For all test cases we assume an ideal gas law $p = (\gamma-1)\rho e$. The implicit non-symmetric linear system given by (5.41) is solved with the GMRES algorithm combined with a preconditioner based on an incomplete LU decomposition [4]. To choose the relaxation parameter a, we follow the procedure given in [11] to obtain a local estimate for a. We calculate a global estimate by taking the maximum of the local values of a and multiply by a constant c_a independent of M to ensure the stability property given in Lemma 5.1. Even though the proof for the AP property was restricted to the case $M = Fr$, the scheme can be applied in regimes with different Mach and Froude numbers. Especially the well-balanced and accuracy test cases were performed with different Mach and Froude numbers.

5.5.1 Well-balanced test case

To numerically verify the well-balanced property of the scheme, we compute an isothermal equilibrium with a linear potential in two dimensions as given in (5.6) where $\mathbf{u} = (u_1, u_2) = 0$, $\chi = 1$ and $\gamma = 1.4$. The equilibrium solutions α and β are set according to the isothermal equilibrium (5.6). In Table 5.1 we give the L^1 error with respect to the initial configuration at the final time $T_f = 1$ for different Mach and Froude numbers on the domain $D = [0,1]^2$ with 100 cells in each direction. In Table 5.1 the results for the first order scheme are given. It can be seen that the error compared to the initial isothermal equilibrium is of machine precision. This means that the solution is still in equilibrium at a later time T_f. This verifies the well-balanced property given in Theorem 5.4.2. In Table (5.2), we have repeated the simulation with the second order scheme. The results clearly show, that also the second order scheme is well-balanced.

5.5.2 Accuracy

To numerically validate the second order accuracy of the proposed scheme, we compare the numerical solution obtained with the second order scheme to an exact solution of the Euler equations with gravity as given in [17]. In physical variables, it is given in 2

M	Fr	ρ	ρu_1	ρu_2	E
10^{-1}	10^{-1}	2.459E-017	3.605E-016	3.605E-016	2.419E-017
10^{-2}	10^{-2}	5.606E-017	9.999E-017	9.999E-017	5.507E-017
10^{-3}	10^{-3}	2.506E-017	9.811E-016	9.811E-016	2.457E-017
10^{-4}	10^{-4}	2.539E-017	5.304E-017	5.304E-017	2.495E-017
10^{-2}	10^{-1}	6.111E-017	5.6222E-016	5.6222E-016	1.517E-016
10^{-4}	10^{-2}	2.386E-017	5.679E-017	5.679E-017	5.879E-017

Table 5.1: L^1-error with respect to the isothermal equilibrium for the first oder IMEX scheme at $T = 1$ (non-dimensional).

M	Fr	ρ	ρu_1	ρu_2	E
10^{-1}	10^{-1}	1.332E-015	1.479E-015	1.479E-015	6.641E-015
10^{-2}	10^{-2}	1.116E-015	1.315E-015	1.315E-015	5.761E-015
10^{-3}	10^{-3}	1.043E-015	1.324E-015	1.324E-015	5.531E-015
10^{-4}	10^{-4}	5.828E-016	5.848E-016	5.848E-016	2.585E-015
10^{-2}	10^{-1}	3.330E-016	4.047E-016	4.047E-016	1.885E-015
10^{-4}	10^{-2}	7.950E-016	6.265E-016	6.265E-016	3.632E-015

Table 5.2: L^1-error with respect to the isothermal equilibrium for the second order IMEX scheme at $T = 1$ (non-dimensional).

dimensions with $\mathbf{x} = (x_1, x_2)$ and $\mathbf{u} = (u_1, u_2)$ as

$$\begin{aligned} \rho(\mathbf{x}, t) &= 1 + 0.2 \sin\left(\pi(x_1 + x_2 - t(u_{1_0} + u_{2_0}))\right) \frac{kg}{m^3} \\ u_1(\mathbf{x}, t) &= u_{1_0} \frac{m}{s} \\ u_2(\mathbf{x}, t) &= u_{2_0} \frac{m}{s} \\ p(\mathbf{x}, t) &= p_0 + t(u_{1_0} + u_{2_0}) - (x_1 + x_2) \\ &\quad + \frac{0.2}{\pi} \cos\left(\pi(x_1 + x_2 - t(u_{1_0} + u_{2_0}))\right) \frac{kg}{ms^2}. \end{aligned} \tag{5.66}$$

For the parameters we set $u_{1_0} = 20, u_{2_0} = 20$ and $p_0 = 4.5$. The gravitational potential is linear and given as $\Phi(\mathbf{x}) = x_1 + x_2$. For $\mathbf{u} = 0$, the solution (5.66) is in hydrostatic equilibrium and we set α and β as the density and pressure of the stationary state respectively. We want to remark that this equilibrium is neither isothermal nor polytropic. The computational domain is $D = [0, 1]^2$ and the final time $T = 0.01s$.

To transform the initial data (5.66) into non-dimensional quantities, we define the following reference values

$$x_r = 1m,\ u_r = 1\frac{m}{s},\ \rho_r = 1\frac{kg}{m^3},\ p_r = \frac{1}{M^2}\frac{kg}{ms^2},\ \Phi_r = \frac{1}{Fr^2}\frac{m^2}{s^2}.$$

We use different values for M and Fr to show that our scheme is second order accurate independently of the chosen regime. In the computations we use exact boundary conditions and $\gamma = 5/3$. As can be seen from Table 5.3 the L^1 error and the convergence rates are of the same magnitude for all displayed Mach numbers and we achieve the expected second order accuracy. In addition, to illustrate that the accuracy is independent of the Mach number, we have plotted the L^1- error in Figure 5.1. Due to the limiting procedure that we apply on the slopes in the reconstruction step to ensure the positivity property, we are not recovering a full second order convergence. Using unlimited slopes in the reconstruction step however will lead to the full second order.

5.5.3 Strong Rarefaction Test

Linearized Riemann solvers can fail producing negative pressures or densities in the intermediate states $W^*_{L,R}$ for very strong rarefactions. In order to demonstrate the positivity preserving property of our schemes proven in Theorems 5.9 and 5.11, we follow the 1-2-0-3 strong rarefaction test proposed in [32] designed for the homogeneous equations. We modified the set-up by launching two rarefaction waves in x-direction on top of an isothermal atmosphere. The initial states for ρ and p follow the isothermal equilibrium (5.6) with a quadratic potential $\Phi(\mathbf{x}) = \frac{1}{2}\left((x_1 - 0.5)^2 + (x_2 - 0.5)^2\right)$ centred at $\mathbf{x} = (0.5, 0.5)$ with $\gamma = 1.4$ and $\chi = \gamma - 1$. As reference equilibrium expressed by α and β, we set the isothermal equilibrium given by (5.6). As in [32], we choose a compressible regime, i.e. $M = 1$ and $Fr = 1$. As initial velocity $\mathbf{u} = (u_1, u_2)$, we set

$$u_1 = \begin{cases} -2 & \text{for } x_1 < 0.5, \\ 2 & \text{for } x_1 \geq 0.5, \end{cases} \qquad u_2 = 0.$$

M, Fr	N	$\rho \left[\frac{kg}{m^3}\right]$		$\rho u_1 \left[\frac{kg}{m^2 s}\right]$		$\rho u_2 \left[\frac{kg}{m^2 s}\right]$		$E \left[\frac{kg}{ms^2}\right]$	
$10^{-1}, 10^{-1}$	25	1.139E-03	—	2.278E-02	—	2.278E-02	—	4.562E-01	—
	50	3.142E-04	1.858	6.276E-03	1.859	6.276E-03	1.859	1.257E-01	1.859
	100	8.427E-05	1.898	1.680E-03	1.901	1.680E-03	1.901	3.366E-02	1.901
	200	2.232E-05	1.916	4.438E-04	1.920	4.438E-04	1.920	8.894E-03	1.920
$10^{-2}, 10^{-2}$	25	1.140E-03	—	2.280E-02	—	2.280E-02	—	4.567E-01	—
	50	3.144E-04	1.859	6.280E-03	1.860	6.280E-03	1.860	1.258E-01	1.859
	100	8.430E-05	1.899	1.680E-03	1.901	1.680E-03	1.901	3.367E-02	1.901
	200	2.233E-05	1.916	4.441E-04	1.919	4.441E-04	1.919	8.901E-03	1.919
$10^{-3}, 10^{-3}$	25	1.141E-003	—	2.281E-02	—	2.281E-02	—	4.569E-01	—
	50	3.144E-04	1.859	6.280E-03	1.861	6.280E-03	1.861	1.258E-01	1.860
	100	8.431E-05	1.898	1.680E-03	1.901	1.680E-03	1.901	3.368E-02	1.901
	200	2.233E-05	1.916	4.441E-04	1.919	4.441E-04	1.919	8.901E-03	1.919
$10^{-4}, 10^{-4}$	25	1.141E-03	—	2.280E-02	—	2.280E-02	—	4.582E-01	—
	50	3.143E-04	1.860	6.277E-03	1.860	6.277E-03	1.860	1.257E-01	1.864
	100	8.430E-05	1.898	1.680E-03	1.901	1.680E-03	1.901	3.367E-02	1.901
	200	2.233E-05	1.916	4.441E-04	1.919	4.441E-04	1.919	8.900E-03	1.919
$10^{-4}, 10^{-1}$	25	1.141E-003	—	2.280E-02	—	2.280E-02	—	4.581E-01	—
	50	3.143E-04	1.860	6.277E-03	1.860	6.277E-03	1.860	1.257E-01	1.864
	100	8.430E-05	1.898	1.680E-03	1.901	1.680E-03	1.901	3.367E-02	1.901
	200	2.233E-05	1.916	4.441E-04	1.919	4.441E-04	1.919	8.900E-03	1.919
$10^{-1}, 10^{-4}$	25	1.139E-03	—	2.278E-02	—	2.278E-02	—	4.562E-01	—
	50	3.142E-04	1.858	6.276E-03	1.859	6.276E-03	1.859	1.257E-01	1.859
	100	8.427E-05	1.898	1.680E-03	1.901	1.680E-03	1.901	3.366E-02	1.901
	200	2.232E-05	1.916	4.438E-04	1.920	4.438E-04	1.920	8.894E-03	1.920

Table 5.3: L^1-error and convergence rates for different Mach and Froude numbers.

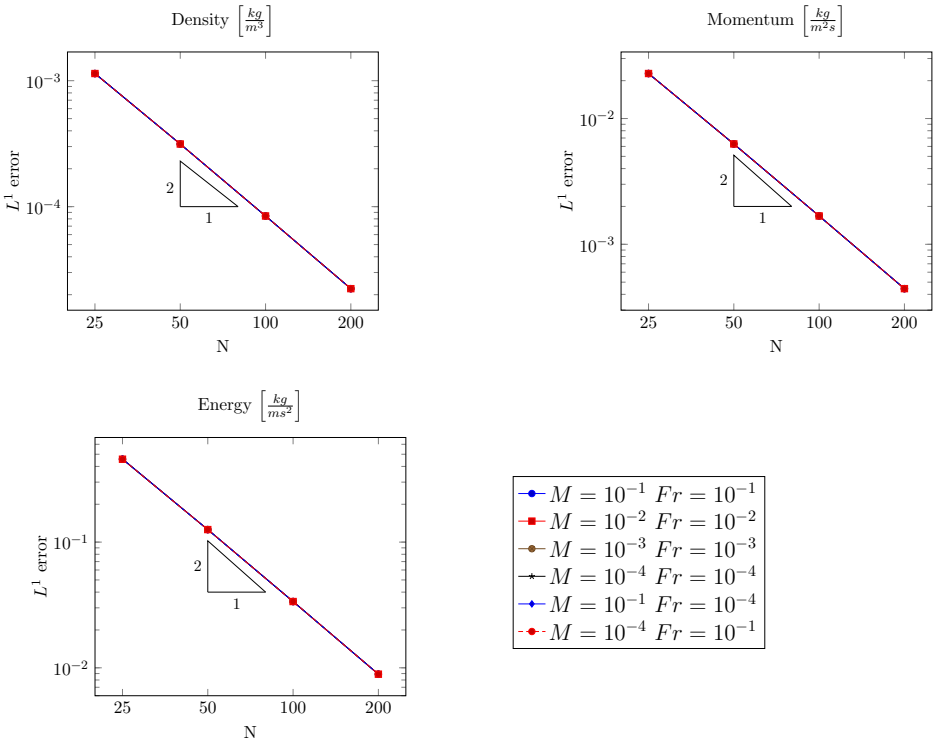


Figure 5.1: L^1 error curves in dependence of Mach and Froude number (dimensional).

The simulations were performed with the first and second order scheme with 100 cells in each space direction on the domain $[0,1]^2$ up to a final time $T_f = 0.1s$. In Figure 5.2, the solution computed with the first and second order scheme is projected onto the x-axis. As can be seen, the pressure and density are close to zero but remain positive throughout the simulation.

5.5.4 A stationary vortex in a gravitational field

With this test-case, we want to demonstrate the low Mach properties of our scheme. For the derivation of a vortex in a gravitational field, we follow the derivation of the Gresho vortex test case for the homogeneous Euler equations [60] as well as [88]. The velocity field in the initial condition is chosen such that it fulfils the divergence free property $\nabla \cdot \mathbf{u} = 0$ and the orthogonality property $\mathbf{u} \cdot \nabla\Phi = 0$. The vortex is placed on top of a hydrostatic equilibrium solution and therefore the initial condition is contained in the set of the well-prepared data Ω_{wp}.

To derive the vortex, we consider the non-dimensional Euler equations (5.3) in radial coordinates (r, θ). The vortex is constructed such that it is axisymmetric, stationary and has zero radial velocity. A solution has to satisfy

$$\frac{1}{M^2}\partial_r p = \frac{\rho u_\theta^2}{r} - \rho\frac{\partial_r \Phi}{Fr^2},$$

where u_θ is the angular velocity. The pressure is split into a hydrostatic pressure p_0 and a pressure p_2 associated with the centrifugal forces and in total is given by $p = p_0 + M^2 p_2$

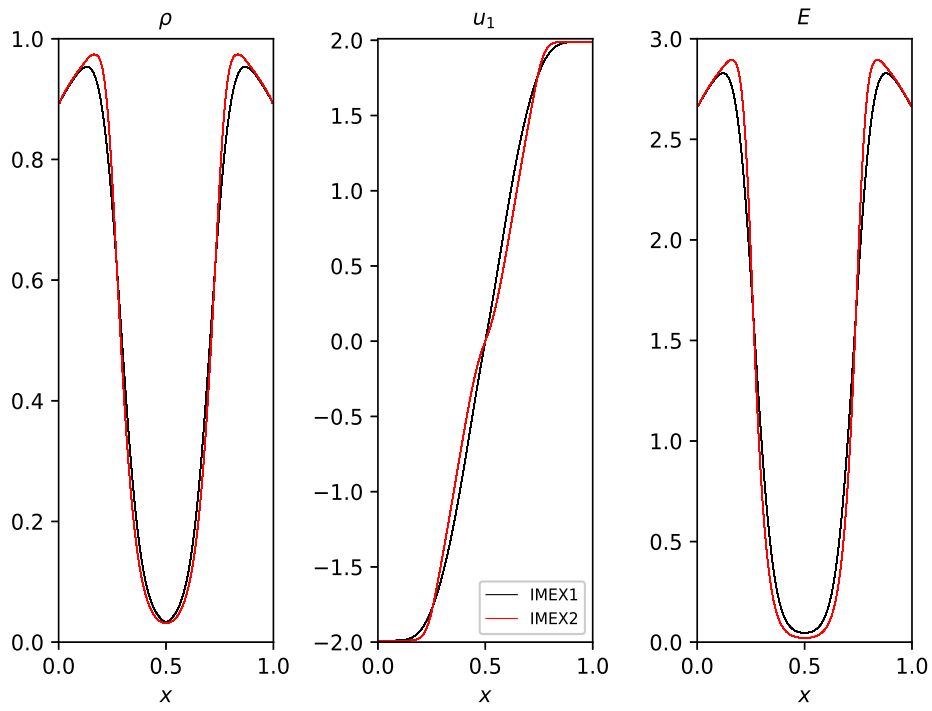


Figure 5.2: Density, velocity and energy for the strong rarefaction test at $T_f = 0.1s$.

and has to satisfy

$$\partial_r p_0 = -\frac{M^2}{Fr^2}\rho\partial_r\Phi, \quad \partial_r p_2 = \rho\frac{u_\theta^2(r)}{r}.$$

We choose an isothermal hydrostatic pressure $p_0 = RT\rho$ and the density is given according to (5.6) by

$$\rho = \exp\left(-\frac{M^2}{Fr^2}\frac{\Phi}{RT}\right).$$

The pressure p_2 is then given as

$$p_2 = \int_0^r \exp\left(-\frac{M^2}{Fr^2}\frac{\Phi(s)}{\chi}\right)\frac{u_\theta(s)^2}{s}ds. \tag{5.67}$$

The velocity profile u_θ is defined piecewise as in the Gresho vortex test case as

$$u_\theta(r) = \frac{1}{u_r}\begin{cases} 5r & \text{if} \quad r \leq 0.2, \\ 2 - 5r & \text{if} \quad 0.2 < r \leq 0.4, \\ 0 & \text{if} \quad r > 0.4. \end{cases}$$

To fully determine p_2 a continuously differentiable gravitational potential has to be given.

We define it piecewise as

$$\Phi(r) = \begin{cases} 12.5r^2 & \text{if} \quad r \leq 0.2 \\ 0.5 - \ln(0.2) + \ln(r) & \text{if} \quad 0.2 < r \leq 0.4 \\ \ln(2) - 0.5\frac{r_c}{r_c - 0.4} + 2.5\frac{r_c}{r_c - 0.4}r - 1.25\frac{1}{r_c - 0.4}r^2 & \text{if} \quad 0.4 < r \leq r_c \\ \ln(2) - 0.5\frac{r_c}{r_c - 0.4} + 1.25\frac{r_c^2}{r_c - 0.4} & \text{if} \quad r > r_c \end{cases}.$$

This choice of Φ ensures the use of periodic boundary conditions since Φ is constant at the boundary and thus we can simulate a closed system. Then we can compute the pressure p_2 according to (5.67) and it is piecewise defined as

$$p_2(r) = \frac{Fr^2 RT}{M^2\ u_r^2} \begin{cases} p_{21}(r) & \text{if} \quad r \leq 0.2 \\ p_{21}(0.2) + p_{22}(r) & \text{if} \quad 0.2 < r \leq 0.4 \\ p_{21}(0.2) + p_{22}(0.4) & \text{if} \quad r > 0.4 \end{cases}$$

with

$$\begin{aligned} p_{21}(r) =& \left(1 - \exp\left(-12.5\frac{M^2}{Fr^2 RT}r^2\right)\right) \\ p_{22}(r) =& \frac{1}{(Fr^2 RT - M^2)\,(Fr^2 RT - 0.5M^2)} \exp\left(\frac{(-0.5 + \ln(0.2))M^2}{Fr^2 RT}\right) \\ & \times \left(r^{-\frac{M^2}{Fr^2 RT}}\ \left(M^4(r(10 - 12.5r) - 2) - 4Fr^4\chi^2 + Fr^2M^2(r(12.5r - 20) + 6)RT\right)\right. \\ & \left. +\ \exp\left(\frac{-\ln(0.2)M^2}{Fr^2 RT}\right)\left(4Fr^4RT^2 - 2.5Fr^2M^2RT + 0.5M^4\right)\right). \end{aligned}$$

As reference values we set $x_r = 1m$, $\rho_r = 1\frac{kg}{m^3}$, $u_r = 2 \cdot 0.2\ \pi\frac{m}{s}$, $t_r = 1\frac{m}{u_r}$ and $RT = \frac{1}{M^2}\frac{m^2}{s^2}$. As reference equilibrium expressed by α and β we set the isothermal equilibrium (5.6), since the vortex is derived from an isothermal steady state. The computations are carried out with $\gamma = 5/3$ and $M = Fr$ on the domain $D = [0, 1]^2$. In Figure 5.3 the initial Mach number distribution for the vortex with $M = 0.1$ is given. In Figure 5.4, the Mach number distribution for different maximum Mach numbers are compared for $N = 40$ at $t = 1$ which corresponds to one turn of the vortex. We see that the accuracy of the vortices are comparable independently of the chosen Mach number and they show the same amount of diffusiveness despite of the coarse grid that is used. The periodic boundary conditions allow us to model a closed system and we can monitor the loss of kinetic energy during the simulation which is depicted in Figure 5.5. The graphs for the Mach numbers $M = 10^{-2}$ and $M = 10^{-3}$ are superposed which shows that the loss of kinetic energy is independent of the Mach number. This is in agreement with the AP property and demonstrates the low Mach number properties of the scheme. We remark that although using the second order scheme, we do not expect to get second order convergence due to the lack of smoothness in the velocity profile u_θ and therefore also in the energy.

5.5.5 Rising bubble test case

The final test case is taken from [58] and models a rising bubble which has a higher temperature than the background atmosphere on the domain $D = [0km, 10km] \times [0km, 15km]$.

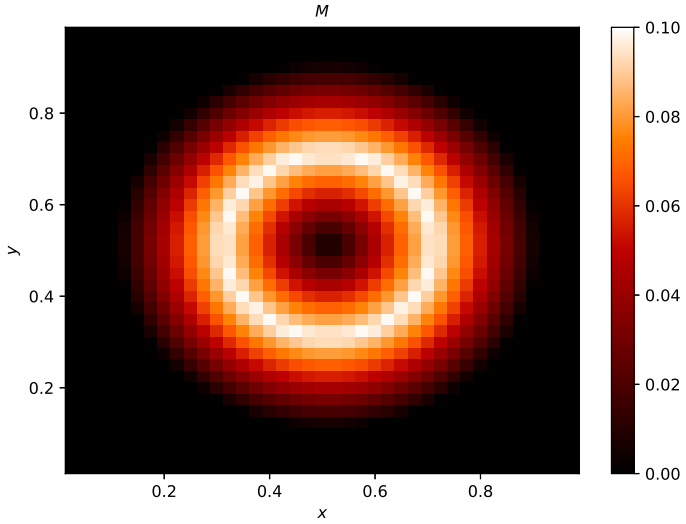


Figure 5.3: Initial Mach number distribution for maximal Mach number $M = 10^{-1}$.

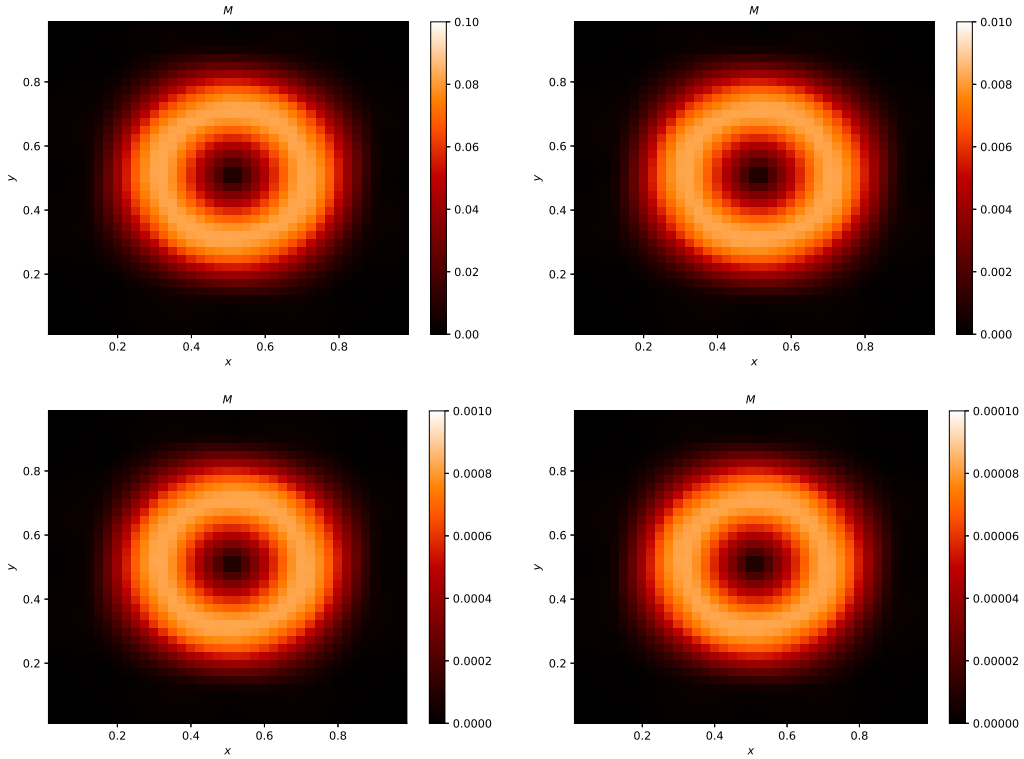


Figure 5.4: Mach number distribution for different maximal Mach numbers at $t = 1$. Top left: $M = 10^{-1}$. Top right: $M = 10^{-2}$, bottom left: $M = 10^{-3}$, bottom right: $M = 10^{-4}$

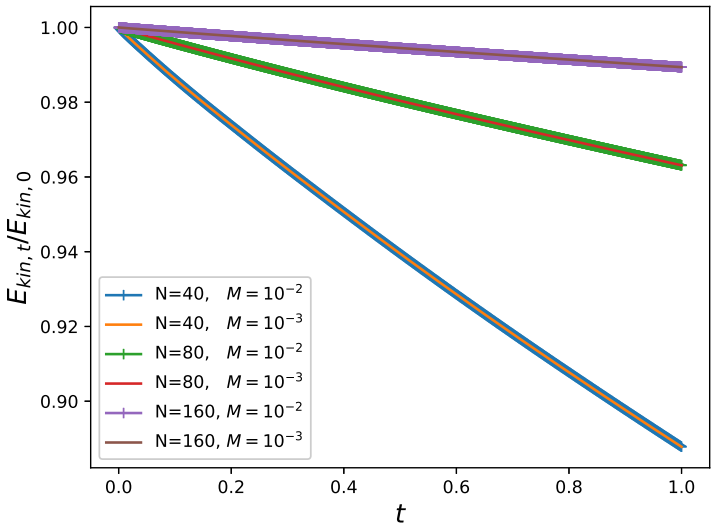


Figure 5.5: Loss of kinetic energy for different grids and Mach numbers after one full turn of the vortex (non-dimensional).

The gravitation acts along the x_2-direction and is given by

$$\Phi(\mathbf{x}) = g x_2 \, \frac{m^2}{s^2},$$

where $g = 9.81 \frac{m}{s^2}$ is the gravitational acceleration. The stratification of the atmosphere is given in terms of the potential temperature θ defined by

$$\theta = T \left(\frac{p_0}{p} \right)^{\frac{R}{c_p}},$$

where c_p is the specific heat at constant pressure and $p_0 = 10^5 \ \frac{kg}{ms^2}$, denotes a reference pressure taken at sea level. Pressure, potential temperature and density are connected by the following relation

$$p = p_0 \left(\frac{\theta R}{p_0} \right)^{\gamma} \rho^{\gamma} = \chi \rho^{\gamma}, \tag{5.68}$$

where c_v is the specific heat at constant volume and $R = c_p - c_v$. Comparing (5.68) to (5.5), the atmosphere is isentropic with the polytropic coefficient $\Gamma = \gamma$. We set $p(\mathbf{x}, 0) = p_0$ and $\theta = 300K$. Therefore we have

$$\rho(\mathbf{x}, 0) = \frac{p_0}{\theta R}$$

and the hydrostatic equilibrium is given by (5.7). To transform the data into non-dimensional quantities, we define the following reference values

$$x_r = 10000 \ m, \ t_r = 10000 \ s, \ u_r = 1 \frac{m}{s}, \ \rho_r = 1 \frac{kg}{m^3}.$$

quantity	SI unit	scaling
x	$[m]$	x_r
t	$[s]$	t_r
ρ	$\left[\frac{kg}{m^3}\right]$	ρ_r
u, c	$\left[\frac{m}{s}\right]$	$u_r = \frac{x_r}{t_r}$, $M = \frac{u_r}{c_r}$
p	$\left[\frac{kg}{m\ s^2}\right]$	$p_r = R_s \rho_r \theta_r$, $p_r = \rho_r c_r^2$
Φ	$\left[\frac{m^2}{s^2}\right]$	$\Phi_r = \frac{u_r^2}{Fr^2}$
R_s	$\left[\frac{m^2}{s^2 K}\right]$	—
T, θ	$[K]$	$\theta_r = \frac{u_r^2}{R_s\ M^2}$

Table 5.4: Overview over units and scaling relations of the physical quantities used in the test cases in Section 6.

The scaling of the remaining variables is given in Table 5.4.

The bubble is modelled as a disturbance in the potential temperature centred at $(x_{1,c}, x_{2,c}) = (5km, 2.75km)$ as

$$\Delta\theta = \begin{cases} \Delta\theta_0 \cos^2\left(\frac{\pi r}{2}\right) & \text{if } r \leq 1 \\ 0 & \text{else} \end{cases}$$

where $\Delta\theta_0 = 6.6K$ and

$$r = \left(\frac{x_1 - x_{1,c}}{r_0}\right)^2 + \left(\frac{x_2 - x_{2,c}}{r_0}\right)^2$$

with the factor $r_0 = 2.0km$. The resulting perturbation in the pressure can be calculated from equation (5.68).

In the simulation, we choose $\gamma = 1.4$ as the air is modelled as a diatomic gas with the corresponding specific gas constant $R_s = 287.058 \frac{m^2}{s^2 K}$. This results in a reference Mach number of $M = 10^{-2}$ and Froude regime of $Fr = 10^{-2}$. Since the bubble is modelled as a perturbation on top of an isentropic atmosphere, we set α and β according to (5.7). The grid consists of 120 cells in x_1-direction and 180 cells in x_2-direction which results into a uniform space discretization. At the boundaries, we have imposed the isentropic background atmosphere. In Figure 5.6, we show the density perturbation at different times t computed with the well-balanced second order scheme. Even though the density perturbation is of order 10^{-5}, the density profile is clearly visible and there are no numerical artefacts in the background stemming from errors in calculating the underlying equilibrium. To demonstrate the importance of accurately capturing the balance between the pressure gradient and the source term, we show in Figure 5.7 the same test case without applying the hydrostatic reconstruction (5.59). In contrast to the evolution depicted in the first frame of Figure 5.6, the bubble seems to have a physically wrong behaviour already

very early in the simulation. Since it vanishes from the computational domain at a later time, we compare the results to the first frame of Figure 5.6. The erroneous result is due to the inaccurate calculation of the pressure at the interface and underlines the necessity of a well-balanced scheme to perform these kind of test cases.

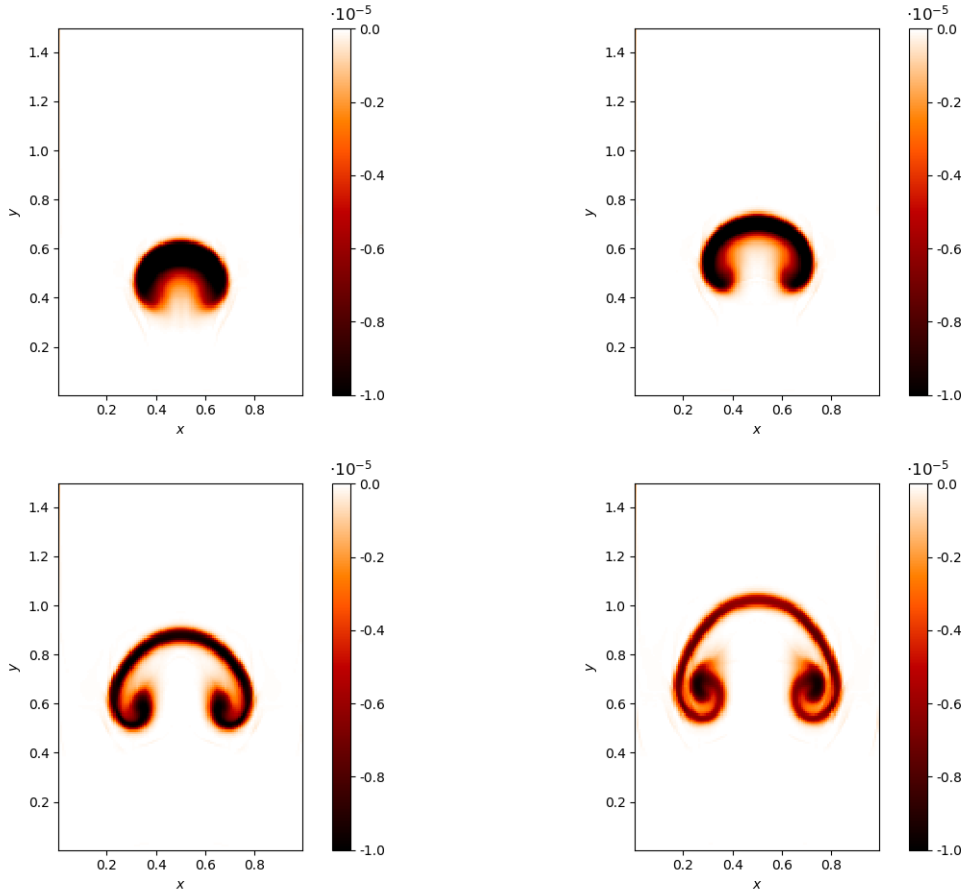


Figure 5.6: Density perturbation from the rising bubble test case from top left to bottom right at times $t = 0.07, 0.09, 0.13, 0.18$.

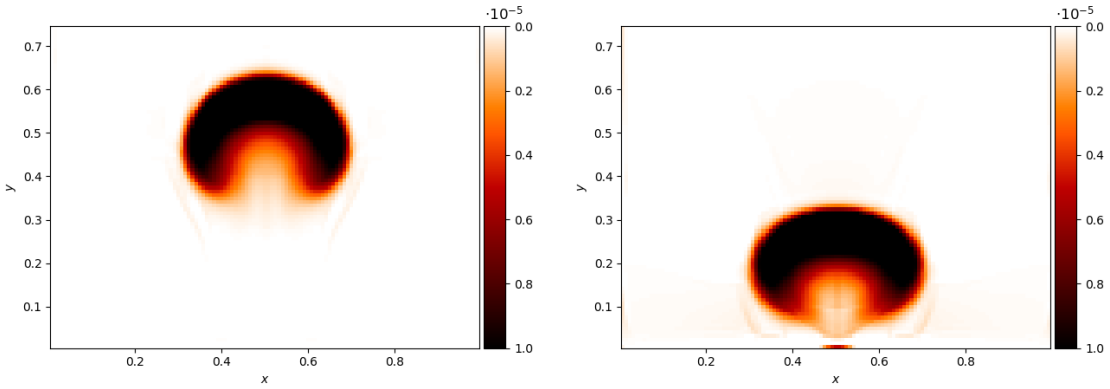


Figure 5.7: Density perturbation from the rising bubble test case at $t = 0.07$. Left: Second order well-balanced scheme. Right: Second order not well-balanced scheme.

Conclusion

In this thesis we have studied, both analytically and numerically, the compressible Euler equations of full gas dynamics in different Mach regimes also in the presence of a gravitational source term.

First, we studied the compressible Euler equations with a gravitational source term, corresponding to a Mach and Froude number of one, with the aim to construct a well-balanced finite volume scheme which is able to preserve arbitrary hydrostatic equilibria on machine precision. The scheme was based on a Suliciu relaxation model where, in addition to the pressure relaxation, also the gravitational potential is relaxed by using a transport equation. Thus the source term was associated with the gas velocity leading to a fixed wave ordering. The resulting approximate Riemann solver took into account also the waves issued by the source term, leading to source term dependent intermediate states. Due to the linear degeneracy of the characteristic fields associated to all eigenvalues of the relaxation model, the approximative Riemann solver could be constructed without much effort leading to a Godunov type first order finite volume scheme. It was proven that in hydrostatic equilibrium the Riemann solver is at rest leading to the well-balancedness of the scheme.

The idea of relaxing the gravitational potential in addition to the pressure was already used in [28] to construct a scheme that is well-balanced for isothermal and polytropic atmospheres. The scheme derived here presented a twofold extension to the therein proposed scheme. The first extension consists of the property to well-balance arbitrary a priori given hydrostatic equilibria. This was done by rewriting the source term with respect to a reference equilibrium which is intended to be numerically preserved on machine precision. Secondly, we have given an extension of the first order scheme to second order accuracy. This is achieved by using a linear reconstruction of the interface values based on the neighbouring cells. The slopes were obtained by using a minmod limiter. To ensure the well-balanced property of the second order scheme, a reconstruction in equilibrium variables was used which was based on a hydrostatic transformation of the pressure.

A second important property of the derived schemes was the preservation of positivity of density and internal energy. For the first order scheme this property was based on the feature of the Riemann solver to provide positive intermediate states for density and internal energy under a sufficiently large relaxation parameter. To ensure that the second order scheme is also positivity preserving, a limiting procedure for the slopes in the linear reconstruction was derived. Numerical test cases were performed to validate the theoretical properties of the first and order scheme. Especially the well-balanced property and the performance of the schemes to resolve small perturbations around a hydrostatic equilibrium were assessed.

Then we turned to the construction of an all-speed scheme for the homogeneous Euler equations. We considered the non-dimensional formulation of the Euler equations where the influence of the Mach number on the flow behaviour could be studied. The scheme was based on a Suliciu type relaxation model which was developed in [7] with the aim to derive an upwind scheme that could be combined with an explicit or implicit time integrator. We modified the model to our purpose to construct an IMEX scheme that uses centred differences in the implicit and an upwind discretization in the explicit part. The model consisted of a splitting of the pressure into a slow compressible and fast acoustic component where both parts were relaxed separately. This enabled the separation of the fast Mach number dependent acoustic parts from the slow Mach number independent dynamics. To ensure stability under a Mach number independent CFL condition, the fast acoustics were integrated implicitly. Due to the relaxation approach it was possible to reformulate the implicit system into one linear equation for the acoustic pressure which was solved with a GMRES algorithm utilizing the PETSC library [4]. As the resulting coefficient matrix in the implicit system was ill conditioned, we used a preconditioner based on a incomplete LU decomposition which increased the precision of the result and drastically decreased the number or iterations in the GMRES algorithm. Due to the linear degeneracy of the eigenvalues, resulting from the relaxation model, the construction of a Godunov type scheme based on an approximate Riemann solver in the explicit part was realizable. The resulting scheme was proven to be asymptotic preserving with a Mach number independent diffusion and positivity preserving under a sufficiently large relaxation parameter.

To increase the accuracy while keeping the above mentioned theoretical properties, a second order extension was given which consisted of a modification of a SSP Runge Kutta time integrator of second order and the linear reconstruction of the interface reconstruction of the conserved variables using a minmod limiter. Numerical tests were performed to verify the properties of the schemes. Especially the performance in the low Mach number regime was tested.

In the last part, we studied the non-dimensional Euler equations equipped with a gravitational source term. The aim was to construct a well-balanced IMEX relaxation scheme which is able to preserve arbitrary hydrostatic equilibria independently of the Mach and Froude regime of the flow. To achieve this we combined the well-balancing technique we used for the compressible equations with the IMEX relaxation approach we used for the all-speed scheme for the homogeneous equations. The well-balanced property of the resulting IMEX scheme relied on the fact that both the implicit part, consisting of solving one linear equation for the fast acoustic pressure, and the explicit part, consisting of a Godunov type scheme based on an approximate Riemann solver, were provably well-balanced. To analyse the asymptotic preserving property of the scheme, we formally derived the low Mach limit of the Euler equations where the Froude number is of the same order as the Mach number. The well-prepared data was found to be a perturbation of an hydrostatic equilibrium. For this set-up the scheme was proven to be AP and due to a Mach number independent CFL condition and numerical diffusion, the scheme was suitable for numerically approximating low Mach, low Froude number flows. An extension to second order was achieved by combining the hydrostatic reconstruction of the interface values with the second order IMEX all-speed scheme for the homogeneous equations. The theoretical properties were numerically verified by several test cases. Especially the performance of the schemes in the presence of small perturbations around an equilibrium state was tested

which requires the well-balanced property as well as the ability to accurately resolve low Mach and low Froude number flows.

Perspectives

The results described in this thesis provide new perspectives and possibilities for further research projects for which partial results have already been obtained.

The schemes presented in Chapters 3 - 5 are second order accurate and current work in progress is the extension of the schemes presented therein to higher order.

The construction of *explicit* higher order well-balanced schemes for the purely compressible Euler equations with gravity is an active field of research, see e.g. [51, 15, 14, 69] and references therein. To obtain a higher order scheme for a balance law, the space reconstruction, the time integration method as well as the source term discretization have to be of higher order. Regarding the explicit higher order interpolation of the interface values, we use a CWENO reconstruction detailed in e.g. [54, 71, 16]. The higher order source term discretization is realized by adopting the source term treatment given in [63]. It leads to a higher order source integration which preserves the well-balanced property when as in our case a reconstruction of the interface values in equilibrium variables is applied. We can use the same type of equilibrium variables as done in Section 3.4.2 for the second order well-balanced scheme. To obtain a fully higher order *explicit* scheme, a Runge Kutta time integration method of sufficient high order can be used, for example SSP-RK3 in case of third order [38].

In case of the all-speed *IMEX* relaxation schemes from Chapters 4 and 5, the straightforward extension of the time integration scheme (4.62) to third order proves to be more difficult. Replacing the underlying SSP-RK2 scheme in (4.61) merely by an SSP-RK3 scheme will not lead to a third order scheme in time. More promising are the higher order IMEX schemes given in [64] which are designed for solving hyperbolic systems with stiff relaxation source terms, where only the relaxation source term is treated implicitly. We will adopt the approaches given in [64] to our IMEX splitting which consists of three parts, the implicit treatment of the fast acoustic waves, the explicit treatment of the slow waves and the implicit treatment of the relaxation source term.

The topic of this thesis was the development of all-speed schemes for the simulation of an ideal gas governed by the inviscid Euler equations. Its flow regime was characterized by a reference Mach number associated to the specific gas under consideration. The situation becomes more challenging when two phases are modelled, for example air and water, which can be described by an ideal gas and a stiffened gas law respectively, as done in [68, 67]. Applications thereof are modelling droplets in air or air bubbles in water. To each phase its phase velocity and pressure is associated which induces in general two different reference Mach numbers that give rise to two scales in the equations. Due to the presence of two phase pressures and velocities, the equations are equipped with a friction and (physical) pressure relaxation source term. We aim to adopt a similar relaxation model as in the one Mach number case, given in equation (4.4), to construct an IMEX relaxation scheme that is suited to simulate two phase flow characterized by two different Mach numbers based on the model given in [68].

IMEX Runge Kutta schemes are widely used to efficiently solve multi-scale problems

arising in the context of hyperbolic conservation and balance laws. The design of IMEX schemes that are total variation diminishing (TVD) to avoid spurious oscillations close to discontinuities or shocks are therefore very interesting in our field. Unfortunately, TVD IMEX RK schemes with scale independent CFL condition can only be of first order which follows directly from the proof given in [38] for implicit unconditionally TVD stable RK schemes. As the construction of higher order TVD IMEX RK schemes with large time steps oriented to the slow scales is impossible, we concentrate on giving a general framework on how to derive highly precise first-order TVD IMEX RK schemes. To obtain theoretical results, we study a linear scalar hyperbolic two-scale equation. The schemes consist of a convex combination of a higher order IMEX and a first order scheme consisting of a forward and backward Euler step, see [59]. They can be easily combined with a MOOD procedure [21] to locally increase the accuracy in regions where the solution is smooth and remains oscillation free when higher order IMEX RK schemes are applied.

In a future work, it would be interesting to investigate this approach further and apply it on two scale non-linear systems of conservation or balance laws in order to study the behaviour of the TVD IMEX RK schemes in the case of non-linear systems. Thereby also the problematic of well-balancing has to be taken into account.

Bibliography

[1] E. Abbate, A. Iollo, and G. Puppo. An all-speed relaxation scheme for gases and compressible materials. *J. Comput. Phys.*, 351:1–24, 2017.

[2] E. Abbate, A. Iollo, and G. Puppo. An implicit scheme for moving walls and multi-material interfaces in weakly compressible materials. *Commun. Comput. Phys.*, 27(1):116–144, 2020.

[3] E. Audusse, F. Bouchut, M-0. Bristeau, R. Klein, and B. Perthame. A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. *SIAM J. Sci. Comput.*, 25(6):2050–2065, 2004.

[4] S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, A. Dener, V. Eijkhout, W. D. Gropp, D. Karpeyev, D. Kaushik, M. G. Knepley, D. A. May, L. Curfman McInnes, R. Tran Mills, T. Munson, K. Rupp, P. Sanan, B. F. Smith, S. Zampini, and H. Zhang. PETSc users manual. Technical Report ANL-95/11 - Revision 3.13, Argonne National Laboratory, 2020.

[5] W. Barsukow, P. VF Edelmann, C. Klingenberg, and F. K. Röpke. A low-Mach Roe-type solver for the Euler equations allowing for gravity source terms. *ESAIM: Proceedings and Surveys*, 58:27–39, 2017.

[6] C. Berthon. Stability of the MUSCL schemes for the Euler equations. *Commun. Math. Sci.*, 3(2):133–157, 2005.

[7] C. Berthon, C. Klingenberg, and M. Zenk. An all Mach number relaxation upwind scheme. *SMAI J. Comput. Math.*, 6:1–31, 2020.

[8] G. Bispen, M. Lukáčová-Medvid'ová, and L. Yelash. Asymptotic preserving IMEX finite volume schemes for low Mach number Euler equations with gravitation. *J. Comput. Phys.*, 335:222–248, 2017.

[9] S. Boscarino, G. Russo, and L. Scandurra. All Mach number second order semi-implicit scheme for the Euler equations of gas dynamics. *J. Sci. Comput.*, 77(2):850–884, 2018.

[10] N Botta, R Klein, S Langenberg, and S Lützenkirchen. Well balanced finite volume methods for nearly hydrostatic flows. *J. Comput. Phys.*, 196(2):539–565, 2004.

[11] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2004.

[12] F. Bouchut, C. Chalons, and S. Guisset. An entropy satisfying two-speed relaxation system for the barotropic Euler equations: application to the numerical approximation of low Mach number flows. *Numer. Math.*, pages 1–42, 2020.

[13] F. Bouchut, E. Franck, and L. Navoret. A low cost semi-implicit low-Mach relaxation scheme for the full Euler equations. *J. Sci. Comput.*, 83(1):24, 2020.

[14] M. Castro, J. M. Gallardo, J. A. López-García, and C. Parés. Well-balanced high order extensions of Godunov's method for semilinear balance laws. *SIAM J. Numer. Anal.*, 46(2):1012–1039, 2008.

[15] M. J. Castro and C. Parés. Well-balanced high-order finite volume methods for systems of balance laws. *J. Sci. Comput.*, 82(2):48, February 2020.

[16] M. J. Castro and M. Semplice. Third-and fourth-order well-balanced schemes for the shallow water equations based on the CWENO reconstruction. *Int. J. Numer. Meth. Fl.*, 89(8):304–325, 2019.

[17] P. Chandrashekar and C. Klingenberg. A second order well-balanced finite volume scheme for Euler equations with gravity. *SIAM J. Sci. Comput.*, 37(3):B382–B402, 2015.

[18] P. Chandrashekar and M. Zenk. Well-balanced nodal discontinuous Galerkin method for Euler equations with gravity. *J. Sci. Comput.*, 71(3):1062–1093, 2017.

[19] G.-Q. Chen, C. D. Levermore, and T.-P. Liu. Hyperbolic conservation laws with stiff relaxation terms and entropy. *Commun. Pur. Appl. Math.*, 47(6):787–830, 1994.

[20] A. Chertock, S. Cui, A. Kurganov, Ş. N. Özcan, and E. Tadmor. Well-balanced schemes for the Euler equations with gravitation: Conservative formulation using global fluxes. *J. Comput. Phys.*, 2018.

[21] S. Clain, S. Diot, and R. Loubère. A high-order finite volume method for systems of conservation laws—Multi-dimensional Optimal Order Detection (MOOD). *J. Comput. Phys.*, 230(10):4028–4050, 2011.

[22] F. Coquel and B. Perthame. Relaxation of energy and approximate Riemann solvers for general pressure laws in fluid dynamics. *SIAM J. on Numer. Anal.*, 35(6):2223–2249, 1998.

[23] F. Cordier, P. Degond, and A. Kumbaro. An asymptotic-preserving all-speed scheme for the Euler and Navier–Stokes equations. *J. Comput. Phys.*, 231(17):5685–5704, 2012.

[24] I. Cravero, G. Puppo, M. Semplice, and G. Visconti. CWENO: uniformly accurate reconstructions for balance laws. *Math. Comput.*, 87(312):1689–1719, 2018.

[25] C. M. Dafermos. *Hyperbolic conservation laws in continuum physics*, volume 325 of *Grundlehren der mathematischen Wissenschaften [Fundamental principles of mathematical sciences]*. Springer-Verlag, Berlin, forth edition, 2016.

[26] S. Dellacherie. Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number. *J. Comput. Phys.*, 229(4):978–1016, 2010.

[27] V. Desveaux. *Contribution à l'approximation numérique des systèmes hyperboliques*. PhD thesis, 2013. Thèse de doctorat dirigée par Berthon, Christophe et Coudière, Yves Mathématiques et leurs interactions Nantes 2013.

[28] V. Desveaux, M. Zenk, C. Berthon, and C. Klingenberg. A well-balanced scheme to capture nonexplicit steady states in the Euler equations with gravity. *Int. J. Numer. Meth. Fl.*, 81(2):104–127, 2016.

[29] V. Desveaux, M. Zenk, C. Berthon, and C. Klingenberg. Well-balanced schemes to capture non-explicit steady states: Ripa model. *Math. Comput.*, 85(300):1571–1602, 2016.

[30] G. Dimarco, R. Loubère, V. Michel-Dansac, and M.-H. Vignal. Second-order implicit-explicit total variation diminishing schemes for the Euler system in the low Mach regime. *J. Comput. Phys.*, 372:178–201, 2018.

[31] G. Dimarco, R. Loubère, and M. Vignal. Study of a new asymptotic preserving scheme for the Euler system in the low Mach number limit. *SIAM J. Sci. Comput.*, 39(5):A2099–A2128, 2017.

[32] B. Einfeldt, C.-D. Munz, P. L. Roe, and B. Sjögreen. On Godunov-type methods near low densities. *J. Comput. Phys.*, 92(2):273–295, 1991.

[33] E. Feireisl, C. Klingenberg, O. Kreml, and S. Markfelder. On oscillatory solutions to the complete Euler system. *J. Differ. Equ.*, 269(2):1521–1543, 2020.

[34] E. Feireisl, C. Klingenberg, and S. Markfelder. On the low Mach number limit for the compressible Euler system. *SIAM J. Math. Anal.*, 51(2):1496–1513, 2019.

[35] E. Gaburro, M. J. Castro, and M. Dumbser. Well-balanced Arbitrary-Lagrangian-Eulerian finite volume schemes on moving nonconforming meshes for the Euler equations of gas dynamics with gravity. *Mon. Not. R. Astron. Soc.*, 477(2):2251–2275, 03 2018.

[36] D. Ghosh and E. M. Constantinescu. *Well-Balanced Formulation of Gravitational Source Terms for Conservative Finite-Difference Atmospheric Flow Solvers*. AIAA Aviation. American Institute of Aeronautics and Astronautics, jun 2015. doi:10.2514/6.2015-2889.

[37] E. Godlewski and P-A. Raviart. *Hyperbolic systems of conservation laws*. Ellipses, 1991.

[38] S. Gottlieb, C.-W. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Rev.*, 43(1):89–112, 2001.

[39] L. Grosheintz-Laval and R. Käppeli. High-order well-balanced finite volume schemes for the Euler equations with gravitation. *J. Comput. Phys.*, 378:324 – 343, 2019.

[40] H. Guillard and C. Viozat. On the behaviour of upwind schemes in the low Mach number limit. *Comput. Fluids*, 28(1):63–86, 1999.

[41] A. Harten, P. D. Lax, and B. Van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Rev.*, 25:35–61, 1983.

[42] S. Jin. Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review. *Riv. Math. Univ. Parma (N.S.)*, 3(2):177–216, 2012.

[43] S. Jin and Z. Xin. The relaxation schemes for systems of conservation laws in arbitrary space dimensions. *Commun. Pur. Appl. Math.*, 48(3):235–276, 1995.

[44] R. Käppeli and S. Mishra. Well-balanced schemes for the Euler equations with gravitation. *J. Comput. Phys.*, 259:199–219, 2014.

[45] R. Käppeli and S. Mishra. A well-balanced finite volume scheme for the Euler equations with gravitation - The exact preservation of hydrostatic equilibrium with arbitrary entropy stratification. *A&A*, 587:A94, 2016.

[46] S. Klainerman and A. Majda. Singular limits of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids. *Commun. Pur. Appl. Math.*, 34(4):481–524, 1981.

[47] S. Klainerman and A. Majda. Compressible and incompressible fluids. *Commun. Pur. Appl. Math.*, 35(5):629–651, 1982.

[48] R. Klein. Semi-implicit extension of a Godunov-type scheme based on low Mach number asymptotics I: One-dimensional flow. *J. Comput. Phys.*, 121(2):213 – 237, 1995.

[49] R. Klein. Scale-dependent models for atmospheric flows. *Annu. Rev. Fluid Mech.*, 42(1):249–274, 2010.

[50] R. Klein, N. Botta, T. Schneider, C.-D. Munz, S. Roller, A. Meister, L. Hoffmann, and T. Sonar. Asymptotic adaptive methods for multi-scale problems in fluid mechanics. *J. Eng. Math.*, 39(1):261–343, 2001.

[51] C. Klingenberg, G. Puppo, and M. Semplice. Arbitrary order finite volume well-balanced schemes for the Euler equations with gravity. *SIAM J. Sci. Comput.*, 41(2):A695–A721, 2019.

[52] R. J. LeVeque. *Finite-Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002.

[53] R. J. LeVeque and D. S. Bale. Wave propagation methods for conservation laws with source terms. In Rolf Jeltsch and Michael Fey, editors, *Hyperbolic Problems: Theory, Numerics, Applications*, volume 130 of *International Series of Numerical Mathematics*, pages 609–618. Birkhäuser Basel, 1999.

[54] D. Levy, G. Puppo, and G. Russo. Compact central WENO schemes for multidimensional conservation laws. *SIAM J. Sci. Comput.*, 22(2):656–672, 2000.

[55] G. Li and Y. Xing. Well-balanced discontinuous Galerkin methods for the Euler equations under gravitational fields. *J. Sci. Comput.*, 67(2):493–513, 2016.

[56] G. Li and Y. Xing. Well-balanced discontinuous Galerkin methods with hydrostatic reconstruction for the Euler equations with gravitation. *J. Comput. Phys.*, 352:445–462, 2018.

[57] T.-P. Liu. Hyperbolic conservation laws with relaxation. *Commun. Math. Phys.*, 108(1):153–175, Mar 1987.

[58] L. R. Mendez-Nunez and J. J. Carroll. Application of the MacCormack scheme to atmospheric nonhydrostatic models. *Mon. Weather Rev.*, 122(5):984–1000, 1994.

[59] V. Michel-Dansac and A. Thomann. TVD IMEX Runge-Kutta schemes based on arbitrarily high order Butcher tableaux. preprint, HAL, February 2020.

[60] F. Miczek, F. K. Röpke, and P.V.F. Edelmann. New numerical solver for flows at various Mach numbers. *Astron. Astrophys.*, 576:A50, 2015.

[61] R. Natalini. Convergence to equilibrium for the relaxation approximations of conservation laws. *Commun. Pur. Appl. Math.*, 49(8):795–823, 1996.

[62] S. Noelle, G. Bispen, K. R. Arun, M. Lukáčová-Medvid'ová, and C.-D. Munz. A weakly asymptotic preserving low Mach number scheme for the Euler equations of gas dynamics. *SIAM J. Sci. Comput.*, 36(6):B989–B1024, 2014.

[63] S. Noelle, N. Pankratz, G. Puppo, and J. R. Natvig. Well-balanced finite volume schemes of arbitrary order of accuracy for shallow water flows. *J. Comput. Phys.*, 213(2):474–499, 2006.

[64] L. Pareschi and G. Russo. Implicit–explicit Runge–Kutta schemes and applications to hyperbolic systems with relaxation. *J. Sci. Comput.*, 25(1):129–155, 2005.

[65] D. Ray, P. Chandrashekar, U. S. Fjordholm, and S. Mishra. Entropy stable scheme on two-dimensional unstructured grids for Euler equations. *Commun. Comput. Phys.*, 19(5):1111–1140, 2016.

[66] P. L. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *J. Comput. Phys.*, 43(2):357–372, 1981.

[67] E. Romenski, D. Drikakis, and E. F. Toro. Conservative models and numerical methods for compressible two-phase flow. *J. Sci. Comput.*, 42(1):68, 2010.

[68] E. Romenski and E. F. Toro. Compressible two-phase flows: two-pressure models and numerical methods. *Comput. Fluid Dyn. J.*, 13:403–416, 2004.

[69] G. Russo and A. Khe. High order well-balanced schemes based on numerical reconstruction of the equilibrium variables. In *Waves and Stability in Continuous Media*, pages 230–241. World Scientific, 2010.

[70] S. Schochet. The mathematical theory of low Mach number flows. *ESAIM-Math. Model. Num.*, 39(3):441–458, 2005.

[71] M. Semplice, A. Coco, and G. Russo. Adaptive mesh refinement for hyperbolic systems based on third-order compact WENO reconstruction. *J. Sci. Comput.*, 66(2):692–724, 2016.

[72] C.-W. Shu. Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In *Advanced numerical approximation of nonlinear hyperbolic equations*, pages 325–432. Springer, 1998.

[73] G. A. Sod. A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws. *J. Comput. Phys.*, 27(1):1–31, 1978.

[74] I. Suliciu. On modelling phase transitions by means of rate-type constitutive equations. Shock wave structure. *Int. J. Eng. Sci.*, 28(8):829–841, 1990.

[75] I. Suliciu. Some stability-instability problems in phase transitions modelled by piecewise linear elastic or viscoelastic constitutive equations. *Int. J. Eng. Sci.*, 30(4):483 – 494, 1992.

[76] P. K. Sweby. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM J. Numer. Anal.*, 21(5):995–1011, 1984.

[77] A. Thomann, G. Puppo, and C. Klingenberg. An all speed second order well-balanced IMEX relaxation scheme for the Euler equations with gravity. *J. Comput. Phys.*, page 109723, 2020.

[78] A. Thomann, M. Zenk, and C. Klingenberg. A second-order positivity-preserving well-balanced finite volume scheme for Euler equations with gravity for arbitrary hydrostatic equilibria. *Int. J. Numer. Meth. Fl.*, 89(11):465–482, 2019.

[79] A. Thomann, M. Zenk, G. Puppo, and C. Klingenberg. An all speed second order IMEX relaxation scheme for the Euler equations. *Commun. Comput. Phys.*, 28(2):591–620, 2020.

[80] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics* . Springer-Verlag: Berlin, 2009.

[81] E. F. Toro. The HLLC Riemann solver. *Shock Waves*, 29(8):1065–1082, November 2019.

[82] R. Touma, U. Koley, and C. Klingenberg. Well-balanced unstaggered central schemes for the Euler equations with gravitation. *SIAM J. Sci. Comput.*, 38(5):B773–B807, 2016.

[83] E. Turkel. Preconditioned methods for solving the incompressible and low speed compressible equations. *J. Comput. Phys.*, 72(2):277 – 298, 1987.

[84] D. Varma and P. Chandrashekar. A second-order, discretely well-balanced finite volume scheme for euler equations with gravity. *Comput. Fluids*, 181:292 – 313, 2019.

[85] J. Vides, B. Braconnier, A. Audit, C. Berthon, and B. Nkonga. A Godunov-type solver for the numerical approximation of gravitational flows. *Comm. Comput. Phys.*, 15:46–75, 2014.

[86] Y. Xing and C.-W. Shu. High order well-balanced WENO scheme for the gas dynamics equations under gravitational fields. *J. Sci. Comput.*, 54(2):645–662, Feb 2013.

[87] J. Zeifang, J. Schütz, K. Kaiser, A. Beck, M. Lukáčová-Medvid'ová, and S. Noelle. A novel full-Euler low Mach number IMEX splitting. *Commun. Comput. Phys.*, 27:292–320, 2020.

[88] M. Zenk. *On Numerical Methods for Astrophysical Applications.* doctoralthesis, Universität Würzburg, 2018.

[89] J. G. Zhou, D. M. Causon, C. G. Mingham, and D. M. Ingram. The surface gradient method for the treatment of source terms in the shallow-water equations. *J. Comput. Phys.*, 168(1):1–25, 2001.