Low Mach and Well-balanced Numerical Methods for Compressible Euler and Ideal MHD Equations with Gravity

- Doctoral Thesis -

submitted by

Claudius B. Birke

Julius-Maximilians-Universität Würzburg
Faculty of Mathematics and Computer Science

Würzburg, February 2024

Abstract

Physical regimes characterized by low Mach numbers and steep stratifications pose severe challenges to standard *finite volume* (FV) methods. We present three new methods specifically designed to navigate these challenges by being both low Mach compliant and well-balanced. These properties are crucial for numerical methods to efficiently and accurately compute solutions in the regimes considered.
First, we concentrate on the construction of an approximate Riemann solver within Godunov-type FV methods. A new relaxation system gives rise to a two-speed relaxation solver for the Euler equations with gravity. Derived from fundamental mathematical principles, this solver reduces the artificial dissipation in the subsonic regime and preserves hydrostatic equilibria. The solver is particularly stable as it satisfies a discrete entropy inequality, preserves positivity of density and internal energy, and suppresses checkerboard modes. The second scheme is designed to solve the equations of ideal *magnetohydrodynamics* (MHD) and combines different approaches. In order to deal with low Mach numbers, it makes use of a low-dissipation version of the *Harten-Lax-van Leer Disontinuities* (HLLD) solver and a partially implicit time discretization to relax the *Courant-Friedrichs-Lewy* (CFL) time step constraint. A Deviation Well-Balancing method is employed to preserve a priori known magnetohydrostatic equilibria and thereby reduces the magnitude of spatial discretization errors in strongly stratified setups. The third scheme relies on an *implicit-explicit* (IMEX) approach based on a splitting of the MHD equations. The slow scale part of the system is discretized by a time-explicit Godunov-type method, whereas the fast scale part is discretized implicitly by central finite differences. Numerical dissipation terms and CFL time step restriction of the method depend solely on the slow waves of the explicit part, making the method particularly suited for subsonic regimes. Deviation Well-Balancing ensures the preservation of a priori known magnetohydrostatic equilibria. The three schemes are applied to various numerical experiments for the compressible Euler and ideal MHD equations, demonstrating their ability to accurately simulate flows in regimes with low Mach numbers and strong stratification even on coarse grids.

Zusammenfassung

Physikalische Regime mit sehr niedrigen Machzahlen und starken Abschichtungen stellen konventionelle *Finite Volumen* (FV) Verfahren vor erhebliche Herausforderungen. In dieser Arbeit präsentieren wir drei neue Verfahren, die in der Lage sind, die Herausforderungen zu bewältigen. Die neuen Verfahren sind speziell an kleine Machzahlen angepasst und können (magneto-)hydrostatische Gleichgewichte exakt erhalten. Diese Eigenschaften sind essentiell für eine effiziente Berechnung präziser Lösungen in den betrachteten Regimen.
Zunächst konzentrieren wir uns auf die Konstruktion eines approximativen Riemannlösers innerhalb von Godunov-artigen FV Verfahren. Ein neues Relaxationssystem führt zu einem Relaxationslöser für die Euler Gleichungen mit Gravitation, der zwei Relaxationsgeschwindigkeiten verwendet. Abgeleitet von grundlegenden mathematischen Prinzipien reduziert dieser Löser die künstliche Dissipation im subsonischen Bereich und erhält hydrostatische Gleichgewichte. Der Löser ist besonders stabil, da er eine diskrete Entropieungleichung erfüllt, die Positivität von Dichte und interner Energie bewahrt und Schachbrettmuster unterdrückt. Das zweite Verfahren löst die idealen *magnetohydrodynamischen* (MHD) Gleichungen und kombiniert verschiedene Ansätze, um die einzelnen numerischen Herausforderungen zu bewältigen. Für einen effizienten Umgang mit niedrigen Machzahlen wird eine Variante des *Harten-Lax-van Leer Discontinuities* (HLLD) Lösers mit künstlich niedriger Dissipation sowie eine teilweise implizite Zeitdiskretisierung zur Lockerung der *Courant-Friedrichs-Lewy* (CFL) Zeitschrittbeschränkung gewählt. Eine Deviation Well-Balancing Methode wird angewendet, um magnetohydrostatische Gleichgewichte zu bewahren und dadurch das Ausmaß von räumlichen Diskretisierungsfehlern in stark geschichteten Atmosphären zu reduzieren. Das dritte Verfahren verwendet einen *implizit-explizit* (IMEX) Ansatz, welcher auf einer Aufspaltung der MHD Gleichungen basiert. Das Teilsystem mit langsamen Ausbreitungsgeschwindigkeiten wird durch eine zeit-explizite Godunov-artige Methode diskretisiert, während das Teilsystem mit schnellen Ausbreitungsgeschwindigkiten implizit durch zentrale finite Differenzen diskretisiert wird. Numerische Dissipationsterme und die CFL Zeitschrittbeschränkung der Methode hängen somit nur von den langsamen Wellen des expliziten Teils ab, so dass die Methode besonders für subsonische Regime geeignet ist. Deviation Well-Balancing gewährleistet die Erhaltung a priori bekannter magnetohydrostatischer Gleichgewichte. Die drei Verfahren werden auf numerische Experimente für die kompressiblen Euler und idealen MHD Gleichungen angewendet und zeigen darin ihre Fähigkeit, Strömungen in Regimen mit niedrigen Machzahlen und starker Schichtung auch auf groben diskreten Gittern akkurat zu simulieren.

Acknowledgements

I would like to thank a number of people who have supported me along my personal and scientific journey. First and foremost, I am deeply grateful to my supervisor Prof. Dr. Christian Klingenberg for his trust and friendly support over the years. His guidance as well as his dedication to fostering an inspiring scientific atmosphere have been truly invaluable.
I would also like to express my sincere gratitude to Prof. Dr. Friedrich Röpke and his team, especially Giovanni Leidi, for the great cooperation. Engaging discussions in our meetings have enriched my understanding of physics and strengthened the connection between my research work and its practical application. In this context, I also acknowledge the German Research Foundation (DFG) for funding my work within our joint project, enabling me to concentrate fully on my scientific research and to participate in enlightening conferences.
I very much appreciate the fruitful collaborations with Prof. Dr. Christophe Chalons and Prof. Dr. Walter Boscheri, who have contributed with their knowledge and experience to our joint projects. A warm thank goes to Dr. Wasilij Barsukow, a reliable source of advice and a good companion at many conferences.
Furthermore, I consider myself very fortunate to be part of such a lively work group and I want to thank my current and former colleagues for great discussions and experiences.
Finally, I am deeply thankful to my family for their unwavering support in every situation in life, and to Lena for all her love and encouragement.

Claudius Birke

Contents

Chapter 1

Introduction

Fluid dynamics plays a crucial role in many practical applications, e.g. in aerospace for optimizing aircraft designs, in biomedicine for studying blood flows and respiratory mechanisms, and in meteorology for accurately modeling atmospheric flows that form the basis for precise weather predictions. This work focuses on applications in astrophysics, more precisely on fluid flows in the interior of stars. Over large periods of their evolution, stars are close to a (magneto-)hydrostatic equilibrium which is characterized by a balance of pressure gradient and gravitational forces. Dynamical flows in the stellar interior such as convection then constitute (relatively) small perturbations of the equilibrium. The speeds of such flows are typically much smaller than the speed of sound, so that the flows are characterized by low Mach numbers ($\mathcal{M} \lesssim 10^{-2}$) [KM17]. At the same time, these flows have high-β values[1] and moderate Alfvén Mach numbers ($\mathcal{M}_{\text{Alf}} \gtrsim 1$) [Mes99][2].
Gaining knowledge about physical processes in stellar interiors through observations (e.g. from neutrinos or asteroseismology) is only possible to a limited extent. It is therefore essential to also make use of mathematical models that describe these flows. The classic way to model fluid dynamics are hyperbolic systems of *partial differential equations* (PDEs). In this work we focus on the compressible Euler and ideal MHD equations with gravitational source terms as models for describing stellar interiors. Although there is rich literature providing mathematical theory on existence and uniqueness of solutions for these equations and for hyperbolic PDEs in general (see e.g [Gli65, BCP00, BF18, BKK$^+$20]), it is rarely possible to calculate analytical solutions to PDEs in practical applications. Therefore, practitioners have to rely on numerical methods to compute approximate solutions. A very popular approach are FV methods, which discretize space by control volumes and compute fluxes across the boundaries of these volumes. FV methods are appreciated because they are conservative by construction and can handle discontinuous flows. An important subclass of FV methods are *Godunov-type methods*, which approximately solve local Riemann problems at the volume boundaries to compute appropriate approximations of the fluxes. The literature in this field provides a wide variety of approximate Riemann solvers (see e.g. [Rus62, Roe81, HLvL83]). However, these standard FV methods are subject to severe limitations in the considered astrophysical regimes. Most FV methods are designed to work in transonic and supersonic regimes and their approximate Riemann solvers add dissipation terms to the physical flux that scale with the largest wave speed in order to capture shocks. However, in regimes characterized by Mach numbers below 10^{-2} this construction leads to excessive numerical dissipation [GV99, GM04]. Second, explicit time-steppers have to choose a time step in accordance with a CFL condition [CFL28] to

[1] β denotes the ratio of gas pressure to magnetic pressure.
[2] In contrast, the outer layers of active stars (e.g. the solar corona) are characterized by low β values.

be stable. For low Mach number flows, the condition severely reduces the admissible time step. Lastly, a separate discretization of hyperbolic fluxes and gravitational source terms prevents standard schemes to exactly balance equilibrium solutions, generating spurious flows that can dominate the numerical solution.
The aim of this thesis is to present new numerical methods that are capable of solving the compressible Euler and ideal MHD equations at low Mach numbers in strongly stratified setups in an efficient way, in the sense that they

i. reduce the numerical dissipation,

ii. relax the CFL time step condition,

iii. maintain (magneto-)hydrostatic equilibria exactly.

In this context, we first focus on the design of approximate Riemann solvers, since this is the core component of Godunov-type methods. On the basis of a relaxation system, we construct a new Riemann solver for the inhomogeneous Riemann problem of the Euler equations. Physical fluxes and source terms are thereby discretized in a consistent fashion, which enables the solver to maintain hydrostatic equilibria. The numerical dissipation is reduced in the subsonic regime by a Mach number dependent rescaling of the dissipation term. This is done within the construction of the relaxation system, so that the Riemann solver does not have to rely on an artificial low Mach fix with free parameters, but is derived from a fundamental basis. The mathematically closed derivation also helps the solver to satisfy a discrete entropy inequality, to be positivity-preserving and to suppress checkerboard modes, which makes it particularly stable. The spatial discretization is combined with an explicit time-stepper, so that the CFL condition remains very restrictive for low Mach numbers.
The second method is designed for solving the ideal MHD equations. To address the dissipation problem, we use a low-dissipation version of the HLLD solver [MM21]. The CFL condition is relaxed by a time-implicit discretization of the continuity, momentum and energy equation that enables to choose the time step independent of the Mach number dependent wave speeds. The induction equation is handled explicitly and coupled to the rest of the system by *Strang-Splitting* [Str68]. The proposed time-marching algorithm leads to a significant speed-up in subsonic regimes. The scheme is combined with the *Deviation Well-Balancing method* [BCK21], which allows to maintain a priori known equilibrium solutions and considerably reduces numerical errors in strongly stratified setups.
The third method poses an alternative to the second one as it addresses the numerical challenges induced by low Mach numbers through a different approach. Based on the observation that the stiffness of the PDE system in subsonic regimes is only generated by one part of the PDE system, namely the acoustic pressure flux, the system is split into a convective-type and a pressure-type sub-system. The system is then discretized by an IMEX approach: the convective part is discretized explicitly by a Godunov-type method, the pressure part implicitly with central finite difference operators. The resulting implicit part is smaller and easier to invert in comparison with fully implicit methods, making the IMEX method less computationally expensive. The dissipation terms and the CFL condition in the Godunov-type method refer only to the convective sub-system for which all wave speeds are Mach number independent, so that the method becomes particularly suited for subsonic regimes. The approach is coupled with the Deviation Well-Balancing method to preserve magnetohydrostatic equilibria.

The thesis is structured as follows. Chapter 2 provides a brief introduction to conservation and balance laws and discusses the definition and properties of the compressible Euler and ideal MHD equations. In Chapter 3 we introduce basic concepts used in FV methods to discretize given systems of PDEs. In Chapter 4 we then concentrate on deriving a low Mach compliant and well-balanced approximate Riemann solver for the Euler equations. Chapters 5 and 6 are devoted to construct methods for the ideal MHD equations, which make it possible to select relatively large time steps despite small Mach numbers and are therefore particularly efficient in the subsonic regime. In Chapter 5 this is achieved by a partially implicit time-marching scheme, in Chapter 6 by an IMEX approach. Finally, Chapter 7 concludes this thesis by giving a summary and a brief outlook.

Chapter 2

Conservation and Balance Laws

Many problems in technical applications or natural sciences can be modeled by the following principle. Let us consider a spatial domain $\mathcal{I} \subset \mathbb{R}^d$ and an unknown quantity $\mathcal{Q}$, which is defined for all points $\mathbf{x} \in \mathcal{I}$. The evolution of $\mathcal{Q}$ follows a simple principle:

> *The temporal change of $\mathcal{Q}$ in a subset $\omega \subset \mathcal{I}$ is equal to the amount of $\mathcal{Q}$ destroyed or generated within ω, plus the flux balance of $\mathcal{Q}$ across the surface of ω. The quantity $\mathcal{Q}$ thus changes if there is a source or sink within ω or if something flows either in or out over the boundary of ω.*

In the following, we first model the case without considering sources and sinks.

2.1 Conservation Laws

Mathematically, this model can be expressed by the *integral form* of a *conservation law*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega} \mathcal{Q}(\mathbf{x},t)\mathrm{d}\mathbf{x} + \oint_{\partial\omega} \mathcal{F}(\mathcal{Q}(\mathbf{x},t)) \cdot \mathbf{n}\ \mathrm{d}S = \mathbf{0} \tag{2.1}$$

on a time interval $t \in [0, t_f] \subset \mathbb{R}_{\geq 0}$. The vector $\mathcal{Q}(\mathbf{x},t) : \mathbb{R}^d \times \mathbb{R}_{\geq 0} \to \mathbb{R}^m$ contains the conserved variables, while $\mathcal{F} = (\mathcal{F}_1, ..., \mathcal{F}_d)^T$ with $\forall i\ \mathcal{F}_i : \mathbb{R}^m \to \mathbb{R}^m$ denotes the physical flux function and $\mathbf{n}$ the outward pointing normal to the boundary $\partial\omega$. Under the assumption that $\mathcal{Q}$ and $\mathcal{F}(\mathcal{Q})$ are at least continuously differentiable, we can apply the divergence theorem of Gauß to reformulate equation (2.1) into

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega} \mathcal{Q}(\mathbf{x},t)\mathrm{d}\mathbf{x} + \int_{\omega} \nabla \cdot \mathcal{F}(\mathcal{Q}(\mathbf{x},t))\mathrm{d}\mathbf{x} = \mathbf{0}. \tag{2.2}$$

Since equation (2.2) holds for all $\omega \subset \mathcal{I}$, we can consider an infinitesimal ω to obtain the *differential form* of a conservation law

$$\frac{\partial}{\partial t}\mathcal{Q}(\mathbf{x},t) + \nabla \cdot \mathcal{F}(\mathcal{Q}(\mathbf{x},t)) = \mathbf{0}. \tag{2.3}$$

In combination with an initial condition

$$\mathcal{Q}_0(\mathbf{x}) = \mathcal{Q}(\mathbf{x},0), \tag{2.4}$$

equation (2.3) poses a *Cauchy* or *initial value problem* (IVP). The system of conservation laws (2.3) can be written in the quasi-linear form

$$\mathcal{Q}_t + \sum_{i=1}^{d} \mathcal{A}_i(\mathcal{Q})\mathcal{Q}_{x_i} = \mathbf{0}, \tag{2.5}$$

where the $\mathcal{A}_i = \frac{\partial}{\partial \mathcal{Q}}\mathcal{F}_i(\mathcal{Q})$ represent Jacobians. In the remainder of this work we focus on *hyperbolic systems.*

Definition 2.1.1. *The system of conservation laws (2.3) is called* ***hyperbolic*** *if all $\mathcal{A}_i$ have m real eigenvalues $\lambda_1, \dots, \lambda_m$ and a full set of eigenvectors $\mathbf{r}^{(1)}, \dots, \mathbf{r}^{(m)}$. The system is called* ***symmetric hyperbolic*** *if all $\mathcal{A}_i$ are symmetric and* ***strictly hyperbolic*** *if all eigenvalues λ_j are distinct.*

Since the Jacobians of hyperbolic systems can be diagonalized, such systems can be put in *characteristic form*

$$\mathbf{w}_t + diag\left(\lambda_1(\mathbf{w}), \dots, \lambda_m(\mathbf{w})\right)\mathbf{w}_x = \mathbf{0}, \tag{2.6}$$

where $\mathbf{w}$ denotes the vector of *characteristic variables*. The resulting m linear transport equations are decoupled and show that information propagates along characteristic curves defined by the characteristic variables with the speed of the eigenvalues λ_j.

The pair $(\lambda_j(\mathcal{Q}), \mathbf{r}^{(j)})$, consisting of an eigenvalue and its corresponding eigenvector, is called the *j-th characteristic field.*

Definition 2.1.2. *The j-th characteristic field is called* ***genuinely nonlinear*** *if for all $\mathcal{Q}$ it holds*

$$\partial_{\mathcal{Q}}\lambda_j(\mathcal{Q}) \cdot \mathbf{r}^{(j)}(\mathcal{Q}) \neq 0. \tag{2.7}$$

On the other hand, the j-th characteristic field is called ***linearly degenerate*** *if for all $\mathcal{Q}$*

$$\forall \mathbf{r} \in \ker\left(\left(\mathcal{A}(\mathcal{Q}) - \lambda_j(\mathcal{Q})\mathbf{I}\right), \quad \partial_{\mathcal{Q}}\lambda_j(\mathcal{Q}) \cdot \mathbf{r} = 0. \tag{2.8}$$

Based on the eigenstructure, we can determine Riemann invariants.

Definition 2.1.3. *A scalar function $I(\mathcal{Q})$ is called a* ***Riemann invariant*** *to the field j, if for all $\mathcal{Q}$*

$$\forall \mathbf{r} \in \ker\left(\mathcal{A}_i(\mathcal{Q}) - \lambda_j(\mathcal{Q})\mathbf{I}\right), \quad \partial_{\mathcal{Q}} I(\mathcal{Q}) \cdot \mathbf{r} = 0. \tag{2.9}$$

Riemann invariants are a helpful tool for determining the solution of Riemann problems, particularly within the context of relaxation systems as discussed in Sect. 3.4.

Solutions of the system of PDEs (2.3) are called *classical* or *strong solutions.* These solutions must be differentiable. However, this is not guaranteed. Due to the nonlinearity of PDEs, discontinuities can develop in the solution even with smooth initial data. A simple example can be derived from Burgers' equation [LeV02].
In order to allow discontinuities in the solution, we need to introduce a different concept of solutions. Let us assume that there exists a smooth solution of (2.3). We then multiply the differential form (2.3) with a smooth test function $\phi(\mathbf{x}, t)$ with compact support so that it is zero outside of some bounded region of the x-t plane, and integrate over time and space

$$\int_{\mathbb{R}^d \times \mathbb{R}_{\geq 0}} \left[\mathcal{Q}_t + \nabla \cdot \mathcal{F}(\mathcal{Q})\right] \phi(\mathbf{x}, t) \mathrm{d}\mathbf{x}\mathrm{d}t = 0. \tag{2.10}$$

Integration by parts and the compact support of ϕ lead to

$$\int_{\mathbb{R}^d \times \mathbb{R}_{\geq 0}} [\mathcal{Q}\phi_t + \mathcal{F}(\mathcal{Q}) \cdot \nabla \phi] \, \mathrm{d}\mathbf{x}\mathrm{d}t + \int_{\mathbb{R}^d} \mathcal{Q}(\mathbf{x}, 0)\phi(\mathbf{x}, 0)\mathrm{d}\mathbf{x} = 0. \tag{2.11}$$

The advantage is that now all derivatives are on the smooth test function ϕ and not on $\mathcal{Q}$ or $\mathcal{F}(\mathcal{Q})$. Therefore, equation (2.11) also makes sense for discontinuous $\mathcal{Q}$.

Definition 2.1.4. *The function $\mathcal{Q}(\mathbf{x}, t)$ is called a **weak solution** of the conservation law (2.3) with given initial data $\mathcal{Q}(\mathbf{x}, 0)$ if it satisfies (2.11) for all test functions $\phi \in \mathcal{C}_0^1$.*

The class of weak solutions thus includes classical solutions, but is not limited to them [HR15]. In contrast to classical solutions, weak solutions can also be discontinuous. These discontinuities also occur in nature and are called *shock waves* or just *shocks*. Let us consider a one-dimensional *Riemann problem* (RP) defined by

$$\begin{cases} \mathcal{Q}_t + \mathcal{F}(u)_x = 0, \\ \mathcal{Q}(x, 0) = \mathcal{Q}^L, \text{ if } x < 0, \\ \mathcal{Q}(x, 0) = \mathcal{Q}^R, \text{ if } x > 0. \end{cases} \tag{2.12}$$

In this case, the speed of the shock S needs to satisfy the *Rankine-Hugoniot jump condition*

$$\mathcal{F}(\mathcal{Q}^L) - \mathcal{F}(\mathcal{Q}^R) = S\left(\mathcal{Q}^L - \mathcal{Q}^R\right). \tag{2.13}$$

For a scalar conservation law the shock speed can be explicitly computed by

$$S = \frac{\mathcal{F}(\mathcal{Q}^L) - \mathcal{F}(\mathcal{Q}^R)}{\mathcal{Q}^L - \mathcal{Q}^R}. \tag{2.14}$$

One problem about the concept of weak solutions is that they are not unique. Even the application of the Rankine-Hugoniot condition does not lead to unique solutions [GR02]. In order to find physically meaningful solutions, it is therefore advisable to rely on additional physical conditions, such as entropy conditions.

Definition 2.1.5. *The scalar function $\eta(\mathcal{Q}) : \mathbb{R}^m \to \mathbb{R}$ is an **entropy function**, if*

1. *the function η satisfies*

$$\left(\frac{\partial \eta}{\partial \mathcal{Q}}\right)^T \left(\frac{\partial \mathcal{F}_i}{\partial \mathcal{Q}}\right) = \left(\frac{\partial \mathcal{F}_i^{\text{ent}}}{\partial \mathcal{Q}}\right)^T \quad \forall i = 1, \ldots, d, \tag{2.15}$$

   *where $\mathcal{V} = \partial\eta/\partial\mathcal{Q}$ denotes the **entropy variables** and $\mathcal{F}^{\text{ent}} = (\mathcal{F}_1^{\text{ent}}, \ldots, \mathcal{F}_d^{\text{ent}})^T$ the **entropy fluxes**, and*

2. *the function $\eta(\mathcal{Q})$ is convex.*

*The pair $(\eta, \mathcal{F}^{\text{ent}})$ is called **entropy-entropy flux pair**.*

We can derive an evolution equation for the entropy by multiplying the conservation law (2.3) with the entropy variables

$$\mathcal{V}^T \mathcal{Q}_t + \mathcal{V}^T \nabla \cdot \mathcal{F} = 0 \tag{2.16}$$

and using the condition (2.15) to finally derive

$$\eta(\mathcal{Q})_t + \nabla \cdot \mathcal{F}^{\text{ent}}(\mathcal{Q}) = 0. \tag{2.17}$$

Thus, for smooth solutions entropy is conserved. However, for discontinuous solutions, which can always arise for nonlinear equations, the previous steps do not hold. In the case of discontinuities, entropy is dissipated across the shock, which is why (2.17) should be rewritten as an inequality

$$\eta(\mathcal{Q})_t + \nabla \cdot \mathcal{F}^{\text{ent}}(\mathcal{Q}) \leq 0. \tag{2.18}$$

In order to deal with discontinuities, we derive a weak form of the inequality by multiplying with a nonnegative test function ϕ, integrating over space and time and then applying the divergence theorem

$$\int_{\mathbb{R}^d \times \mathbb{R}_{\geq 0}} \phi_t \eta + \nabla \phi \cdot \mathcal{F}^{\text{ent}}(\mathcal{Q}) \mathrm{d}\mathbf{x}\mathrm{d}t \leq -\int_{\mathbb{R}^d} \phi(\mathbf{x}, 0)\eta(\mathbf{x}, 0)\mathrm{d}\mathbf{x}. \tag{2.19}$$

Definition 2.1.6. *A function $\mathcal{Q}(\mathbf{x}, t)$ is called a **weak entropy solution** if it is a weak solution and additionally satisfies (2.19) for all $\phi \in \mathcal{C}_0^1$ and for all entropy-entropy flux pairs $(\eta, \mathcal{F}^{\text{ent}})$.*

This condition on weak solutions is a useful criterion for selecting a unique solution to the system of conservation laws, which otherwise has many weak solutions. Therefore, it is also useful for the construction of numerical methods.

2.2 Balance Laws

Homogeneous conservation laws do not include sources or sinks. In the following, we will now incorporate source terms[1] into the mathematical model. Using similar steps as for conservation laws, we can obtain the following differential form of *balance laws*

$$\frac{\partial}{\partial t}\mathcal{Q}(\mathbf{x}, t) + \nabla \cdot \mathcal{F}(\mathcal{Q}(\mathbf{x}, t)) = \mathcal{S}(\mathcal{Q}(\mathbf{x}, t)). \tag{2.20}$$

The source term $\mathcal{S} : \mathbb{R}^m \to \mathbb{R}^m$ can, for instance, model radiation or chemical reactions. In this work we will consider $\mathcal{S}$ to be a time-independent gravitational source term.
In the solutions of balance laws, as in the case of conservation laws, discontinuities can occur, which is why it makes sense to operate with the concept of weak solutions here as well.

Definition 2.2.1. *The function $\mathcal{Q}(\mathbf{x}, t)$ is called a **weak solution** of the balance law (2.20) with given initial data $\mathcal{Q}(\mathbf{x}, 0)$ if it satisfies*

$$\int_{\mathbb{R}^d \times \mathbb{R}_{\geq 0}} \left[\mathcal{Q}\phi_t + \mathcal{F}(\mathcal{Q}) \cdot \nabla\phi - \mathcal{S}(\mathcal{Q})\phi\right] \mathrm{d}\mathbf{x}\mathrm{d}t + \int_{\mathbb{R}^d} \mathcal{Q}(\mathbf{x}, 0)\phi(\mathbf{x}, 0)\mathrm{d}\mathbf{x} = 0. \tag{2.21}$$

for all test functions $\phi \in \mathcal{C}_0^1$.

For a solution that is piecewise $\mathcal{C}^1$, the Rankine-Hugoniot condition remains the same as for conservation laws. The eigenstructure of the system, on the other hand, changes due to the source term, since it adds a new linearly degenerate eigenvalue $\lambda = 0$. The system of balance laws remains hyperbolic as long as all eigenvalues of the Jacobians of the homogeneous system are nonzero.

[1] Note that the term "source" from here on also includes sinks.

Including source terms in the PDE also has a significant effect on stationary solutions. These solutions are independent of time and satisfy a balance of flux and source term given by

$$\nabla \cdot \mathcal{F}(\mathcal{Q}) = \mathcal{S}(\mathcal{Q}). \tag{2.22}$$

Stationary solutions are of great interest because many physical systems are close to their equilibrium state. This is attributed to the second law of thermodynamics, which states that the (physical) entropy of closed systems tends to increase over time until the system reaches its equilibrium [Wal85].

2.3 Compressible Euler Equations

The compressible Euler equations are a set of fundamental equations in fluid dynamics that describe the behavior of a compressible fluid without considering viscous effects. The homogeneous Euler equations are derived from the physical principles of conservation of mass, momentum and energy (see e.g. [LeV92]). In addition, we include gravitational forces in order to describe atmospheric flows. In consequence, the momentum and energy equations become nonconservative. The resulting hyperbolic system of balance laws writes

$$\frac{\partial}{\partial t}\begin{pmatrix} \rho \\ \rho\mathbf{v} \\ E \end{pmatrix} + \nabla \cdot \begin{pmatrix} \rho\mathbf{v} \\ \rho\mathbf{v}\otimes\mathbf{v} + p\mathbf{I} \\ (E+p)\mathbf{v} \end{pmatrix} = \begin{pmatrix} 0 \\ \rho\mathbf{g} \\ \rho\mathbf{v}\cdot\mathbf{g} \end{pmatrix}, \tag{2.23}$$

where $\rho(\mathbf{x},t) : \mathbb{R}^d \times \mathbb{R}_{\geq 0} \to \mathbb{R}^+$ denotes the density, $\mathbf{v}(\mathbf{x},t) : \mathbb{R}^d \times \mathbb{R}_{\geq 0} \to \mathbb{R}^d$ the velocity vector and $E(\mathbf{x},t) : \mathbb{R}^d \times \mathbb{R}_{\geq 0} \to \mathbb{R}^+$ the total energy. The pressure is given by a pressure law $p(\tau, e) : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$, where $\tau = 1/\rho$ denotes the specific volume and $e > 0$ the internal energy. The total energy can then be expressed by

$$E = \rho e + \frac{1}{2}\rho|\mathbf{v}|^2. \tag{2.24}$$

The source term on the right-hand side of (2.23) contains the gravitational acceleration $\mathbf{g}(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^m$, which can also be expressed by

$$\mathbf{g} = \nabla\Phi, \tag{2.25}$$

where $\Phi(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ denotes a given smooth gravitational potential.

We define the set of physical admissible states, which contains the states $\mathcal{Q} = (\rho, \rho\mathbf{v}, E)$ with positive density and internal energy, by

$$\Omega_{\text{phys}}^{\text{Euler}} = \{\mathcal{Q} \in \mathbb{R}^{d+2}; \rho > 0, e > 0\}. \tag{2.26}$$

This domain is convex invariant in the sense that if $\mathcal{Q}_0(\mathbf{x}) \in \Omega_{\text{phys}}^{\text{Euler}}$ then $\mathcal{Q}(\mathbf{x},t) \in \Omega_{\text{phys}}^{\text{Euler}}$ for all $\mathbf{x} \in \mathcal{I}$, $t > 0$.

2.3.1 Ideal Gas Equation of State

In order to obtain a closed system of equations, one needs in addition to the Euler system in (2.23) an *equation of state* (EoS) that relates the internal energy to pressure and density. The explicit form of the EoS depends on the underlying physics of the gas that shall be modeled. In this work we will consider an ideal gas law given by

$$p = \rho RT \quad \text{and} \quad e = \frac{RT}{\gamma - 1}, \tag{2.27}$$

where R represents the gas constant, which can be derived by dividing the universal gas constant $\mathcal{R}$ by the molecular weight of the gas. The function $T(\tau, e) > 0$ denotes the temperature and γ the adiabatic coefficient. Using (2.27) the pressure can be expressed in terms of the specific volume and internal energy by

$$p(\tau, e) = \frac{(\gamma - 1)e}{\tau}. \tag{2.28}$$

The adiabatic coefficient depends on the nature of the gas and must therefore be adapted to the respective gas under consideration. For mono-atomic gases it can be estimated by $\gamma = 5/3$ and for diatomic gases by $\gamma = 1.4$ [LeV02]. Air on the Earth's surface, for example, is modelled as a diatomic gas, since it consists largely of the diatomic gases nitrogen (N_2) and oxygen (O_2) [Tok02].

2.3.2 Entropy

The Euler equations can have many weak solutions. Therefore, it is necessary to rule out unphysical solutions such as expansion shocks. Essential in this context is the thermodynamic entropy, which can only increase with time according to the second law of thermodynamics. Since we define entropy in Def. 2.1.5 as a convex function, the mathematical entropy can only decrease with time.

Let us assume that the pressure law obeys the second law of thermodynamics so that a specific entropy $s(\tau, e) : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ exists, which satisfies the relation

$$-Tds = de + pd\tau. \tag{2.29}$$

From this relation, we can deduce the conditions

$$s(\tau, e)_\tau = -\frac{p(\tau, e)}{T(\tau, e)} < 0 \quad \text{and} \quad s(\tau, e)_e = -\frac{1}{T(\tau, e)} < 0. \tag{2.30}$$

Lemma 2.3.1. *Smooth solutions of the Euler equations (2.23) satisfy the additional conservation law*

$$\partial_t(\rho\mathcal{G}(s)) + \nabla \cdot (\rho\mathcal{G}(s)\mathbf{v}) = 0. \tag{2.31}$$

We assume

$$\mathcal{G}'(s) > 0 \quad \textit{and} \quad \frac{1}{c_p}\mathcal{G}'(s) + \mathcal{G}''(s) > 0, \tag{2.32}$$

where c_p is the specific heat at constant pressure defined by

$$c_p = -T\left(\frac{\partial s}{\partial T}\right)_p. \tag{2.33}$$

Then $\mathcal{Q} \mapsto \rho\mathcal{G}(s)$ is strictly convex, $(\rho\mathcal{G}(s), \rho\mathcal{G}(s)\mathbf{v})$ defines an entropy-entropy flux pair in the sense of Def. 2.1.5 and weak solutions of (2.23) satisfy

$$\partial_t(\rho\mathcal{G}(s)) + \nabla \cdot (\rho\mathcal{G}(s)\mathbf{v}) \leq 0. \tag{2.34}$$

Proof. From the continuity equation in (2.23), we can derive an equation for the specific volume

$$\partial_t\tau + \mathbf{v} \cdot \nabla\tau - \tau\nabla \cdot \mathbf{v} = 0. \tag{2.35}$$

Multiplying this equation with $-\frac{p}{T}$ and using the first equation in (2.30) yields

$$\partial_\tau s \partial_t \tau + \partial_\tau s \mathbf{v} \cdot \nabla \tau - \partial_t s \tau \nabla \cdot \mathbf{v} = 0. \tag{2.36}$$

From the momentum and total energy equations we can derive the following equation describing the evolution of the internal energy

$$\partial_t e + \mathbf{v} \cdot \nabla e + \tau p \nabla \cdot \mathbf{v} = 0. \tag{2.37}$$

Multiplying this equation with $-\frac{1}{T}$ and using the second equation from (2.30) results in

$$\partial_e s \partial_t e + \partial_e s \mathbf{v} \cdot \nabla e + \tau \partial_t s \nabla \cdot \mathbf{v} = 0. \tag{2.38}$$

Summing up equations (2.36) and (2.38) yields

$$\partial_t s + \mathbf{v} \cdot \nabla s = 0. \tag{2.39}$$

Finally, multiplying (2.39) with $\rho \mathcal{G}'(s)$ and applying the chain rule gives (2.31). The function $\rho \mathcal{G}(s)$ is strictly convex, since its Hessian matrix is positive definite [GR02, Daf09, HLLM98, LeF02]. □

The negatively scaled thermodynamic entropy and the corresponding entropy flux for the Euler equations associated with an ideal gas law are given in [Har83] by

$$\begin{aligned} \eta &= -\frac{\rho s}{\gamma - 1}, \\ \mathcal{F}^{ent} &= -\frac{\rho s}{\gamma - 1} \mathbf{v}, \end{aligned} \tag{2.40}$$

where the specific entropy is defined by

$$s = log\left(\frac{p}{\rho^\gamma}\right). \tag{2.41}$$

2.3.3 Eigenstructure

For a better understanding of the Euler equations and later for the construction of numerical methods to solve them approximately, it is helpful to study the eigenstructure of the matrix $\mathcal{A}$ in the quasi-linear form (2.5). The one-dimensional Euler system with gravitational source term exhibits four waves with the wave speeds

$$\lambda_1^{\text{Euler}} = u - c, \quad \lambda_2^{\text{Euler}} = u, \quad \lambda_3^{\text{Euler}} = u + c, \quad \lambda_4^{\text{Euler}} = 0. \tag{2.42}$$

The quantity c denotes the speed of sound given by

$$c = \tau \sqrt{p \partial_e p - \partial_\tau p}. \tag{2.43}$$

Under the assumption of an ideal gas law (2.28), the sound speed can be computed by

$$c = \sqrt{\frac{\gamma p}{\rho}}. \tag{2.44}$$

The eigenvalue λ_4^{Euler} originates from including the source term. For positive density and pressure the Euler equations have real eigenvalues and a full set of eigenvectors, except for the resonant point $\{u = c\}$, for which two eigenvectors coincide. Apart from this case, the system is hyperbolic.

Remark 2.3.2. *The one-dimensional homogeneous Euler equations are strictly hyperbolic, since they do not exhibit the eigenvalue λ_4^{Euler}. The Euler equations in $d > 1$ dimension fail to be strictly hyperbolic because the eigenvalue u has a multiplicity of d. However, since the corresponding waves are linearly degenerate, their behaviour remains independent [LeV98].*

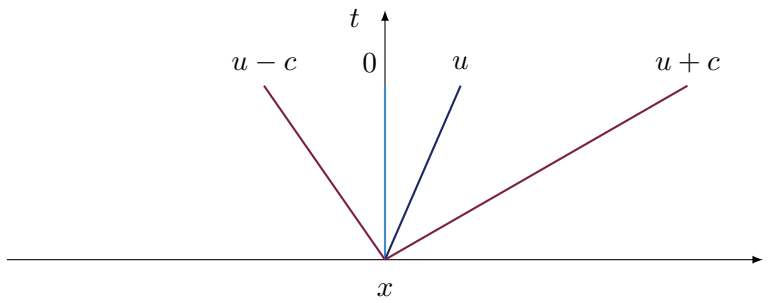

Figure 2.1: Wave structure of the one-dimensional Euler equations with a gravitational source term. The acoustic waves (maroon) are either shock or rarefaction waves, while the contact wave (indigo) and the zero wave (sky blue) are contact discontinuities.

2.3.4 Equilibrium Solutions

We have already seen that stationary solutions satisfy the balance of flux and source term given in (2.22). For applications modeling atmospheric flows or stellar structures, the class of *hydrostatic equilibria* (HSE) is of special interest. These equilibria are stationary states at rest, meaning that the velocities are zero. The balance (2.22) simplifies in these cases to

$$\begin{cases} \mathbf{v} = \mathbf{0}, \\ \nabla p = \rho \mathbf{g}. \end{cases} \tag{2.45}$$

Solutions to system (2.45) are not unique and depend on the pressure law defining the relation between pressure and density. The assumption of additional conditions for the gas law leads to uniqueness and explicit formulas for the solution can be determined. One especially in astrophysics relevant family of EoS assumes

$$p(\mathbf{x}) = \chi \rho(\mathbf{x})^{\Gamma} \tag{2.46}$$

for $\chi > 0$ and $\Gamma \in (0, \infty]$. The parameter Γ denotes the polytropic coefficient, which is to be distinguished from the adiabatic coefficient γ used in the ideal gas law. The HSE equation (2.45) along with (2.46), can be solved explicitly, contingent upon the specific value of Γ. The explicit formulae for three families of equilibria are given below:

- **Isothermal equilibrium:** For $\Gamma = 1$ and a given constant $C \in \mathbb{R}$, the equilibrium is given by

$$\begin{cases} \mathbf{v}(\mathbf{x}) = \mathbf{0}, \\ \rho(\mathbf{x}) = \exp\left(\frac{C - \Phi(\mathbf{x})}{\chi}\right), \\ p(\mathbf{x}) = \chi \exp\left(\frac{C - \Phi(\mathbf{x})}{\chi}\right). \end{cases} \tag{2.47}$$

- **Polytropic equilibrium:** For $\Gamma \in (0,1) \cup (1, \infty)$ and a given constant $C \in \mathbb{R}$, the equilibrium is given by

$$\begin{cases} \mathbf{v}(\mathbf{x}) = \mathbf{0}, \\ \rho(\mathbf{x}) = \left(\frac{\Gamma - 1}{\Gamma \chi}\left(C - \Phi(\mathbf{x})\right)\right)^{\frac{1}{\Gamma - 1}}, \\ p(\mathbf{x}) = \chi^{\frac{1}{1-\Gamma}} \left(\frac{\Gamma - 1}{\Gamma}\left(C - \Phi(\mathbf{x})\right)\right)^{\frac{\Gamma}{\Gamma - 1}}. \end{cases} \tag{2.48}$$

- **Incompressible equilibrium:** For a constant density, the equilibrium is given by

$$\begin{cases} \mathbf{v}(\mathbf{x}) = \mathbf{0}, \\ \rho(\mathbf{x}) = const, \\ p(\mathbf{x}) + \rho(\mathbf{x})\Phi(\mathbf{x}) = const. \end{cases} \tag{2.49}$$

2.3.5 Dimensionless Equations and the Incompressible Limit

In order to analyze the effect of different flow regimes on the gas, it is reasonable to consider the dimensionless form of the Euler equations [BEKR17]:

$$\frac{\partial}{\partial t}\begin{pmatrix} \hat{\rho} \\ \hat{\rho}\hat{\mathbf{v}} \\ \hat{E} \end{pmatrix} + \hat{\nabla}\cdot\begin{pmatrix} \hat{\rho}\hat{\mathbf{v}} \\ \hat{\rho}\hat{\mathbf{v}}\otimes\hat{\mathbf{v}} + \frac{1}{\mathcal{M}^2}\hat{p}\mathbf{I} \\ (\hat{E}+\hat{p})\hat{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{1}{Fr^2}\hat{\rho}\nabla\hat{\Phi} \\ -\frac{\mathcal{M}^2}{Fr^2}\hat{\rho}\hat{\mathbf{v}}\cdot\nabla\hat{\Phi} \end{pmatrix}. \tag{2.50}$$

The rescaled total energy is defined by

$$\hat{E} = \hat{\rho}\hat{e} + \frac{1}{2}\mathcal{M}^2\hat{\rho}|\hat{\mathbf{v}}|^2. \tag{2.51}$$

The different variables have been rescaled by some reference quantity representative of the physical system of interest: $t = t_{\mathrm{r}}\hat{t}$, $\mathbf{x} = x_{\mathrm{r}}\hat{\mathbf{x}}$, $\rho = \rho_{\mathrm{r}}\hat{\rho}$, $\mathbf{v} = v_{\mathrm{r}}\hat{\mathbf{v}}$, $e = c_{\mathrm{r}}^2\hat{e}$, $p = \rho_{\mathrm{r}}c_{\mathrm{r}}^2\hat{p}$, $\Phi = \Phi_r\hat{\Phi}$, $v_{\mathrm{r}} = x_{\mathrm{r}}/t_{\mathrm{r}}$, $c_{\mathrm{r}}^2 = p_{\mathrm{r}}/\rho_{\mathrm{r}}$. $\mathcal{M} = |v_{\mathrm{r}}|/c_{\mathrm{r}}$ represents the characteristic sonic Mach number and $Fr = v_{\mathrm{r}}/\sqrt{\Phi_r}$ the characteristic Froude number of the flow. In this work, we only consider the combined low Mach/low Froude number limit, which is the reason why we set

$$Fr = \mathcal{M}. \tag{2.52}$$

We impose this restriction because in the case $Fr \ll \mathcal{M}$ gravity dominates the complete flow, whereas in the case $Fr \gg \mathcal{M}$ gravity hardly plays a role. The case $Fr \approx \mathcal{M}$, on the other hand, is typically given for atmospheric flows in which we are interested.

The appearance of the Mach number in front of the pressure gradient also effects the acoustic waves:

$$\hat{\lambda}_1^{\text{Euler}} = \hat{u} - \frac{\hat{c}}{\mathcal{M}}, \quad \hat{\lambda}_2^{\text{Euler}} = \hat{u}, \quad \hat{\lambda}_3^{\text{Euler}} = \hat{u} + \frac{\hat{c}}{\mathcal{M}}, \quad \hat{\lambda}_4^{\text{Euler}} = 0. \tag{2.53}$$

At this point it may be noted in particular that the acoustic wave speeds become faster for smaller Mach numbers. For the rest of this subsection, we will omit the hat notation for simplicity.

In the low Mach limit, the solutions of the Euler equations (2.23) tend to the solutions of the incompressible Euler equations [KM82]. We illustrate this behaviour along the lines of [HXX22]. Since it is more convenient, we replace the equation for the total energy in (2.50) by the following equation for the pressure, which can be derived from the total energy equation and the momentum equation:

$$\partial_t p + \mathbf{v}\cdot\nabla p + \gamma p\nabla\cdot\mathbf{v} = 0. \tag{2.54}$$

Defining the pressure p in terms of the potential temperature θ by $p = (\rho\theta)^\gamma$ and inserting this definition into (2.54) yields

$$\partial_t(\rho\theta) + \nabla\cdot(\rho\theta\mathbf{v}) = 0. \tag{2.55}$$

Applying the product rule and using the continuity equation results in an advection equation of the form

$$\partial_t \theta + \mathbf{v} \cdot \nabla \theta = 0. \tag{2.56}$$

For the derivation of the limit equations, we now replace the total energy equation in (2.50) by this transport equation for the potential temperature:

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) &= 0, \\ \partial_t (\rho \mathbf{v}) + \nabla \cdot \left(\rho \mathbf{v} \otimes \mathbf{v} + \frac{1}{\mathcal{M}^2} p \mathbf{I} \right) &= -\frac{1}{\mathcal{M}^2} \rho \nabla \Phi, \\ \partial_t \theta + \mathbf{v} \cdot \nabla \theta &= 0. \end{aligned} \tag{2.57}$$

In this new set of equations, we insert expansions in terms of $\mathcal{M}$ given by

$$\begin{aligned} \rho &= \rho_0 + \mathcal{M}^2 \rho_2 + \mathcal{O}(\mathcal{M}^3), & \mathbf{v} &= \mathbf{v}_0 + \mathcal{M} \mathbf{v}_1 + \mathcal{M}^2 \mathbf{v}_2 + \mathcal{O}(\mathcal{M}^3), \\ p &= p_0 + \mathcal{M}^2 p_2 + \mathcal{O}(\mathcal{M}^3), & \theta &= \theta_0 + \mathcal{M}^2 \theta_2 + \mathcal{O}(\mathcal{M}^3). \end{aligned} \tag{2.58}$$

Collecting all terms of order $\mathcal{O}(\mathcal{M}^{-2})$ yields

$$\nabla p_0 = -\rho_0 \nabla \Phi. \tag{2.59}$$

The couple (ρ_0, p_0) thus fulfills the HSE. We therefore assume a constant background stratification for which θ_0 constitutes a constant background potential temperature. Under these assumptions, the limit equations are

$$\begin{aligned} \nabla \cdot (\rho_0 \mathbf{v}_0) &= 0, \\ \partial_t \mathbf{v}_0 + \mathbf{v}_0 \cdot \nabla \mathbf{v}_0 + \frac{\nabla p_2}{\rho_0} &= -\frac{\rho_2 \nabla \Phi}{\rho_0}, \\ \partial_t \theta_2 + \mathbf{v}_0 \cdot \nabla \theta_2 &= 0. \end{aligned} \tag{2.60}$$

This system was first derived by Ogura and Phillips in [OP62].

Remark 2.3.3. *The limit equations (2.60) contain an unknown ρ_2, to which no conditions seem to be attached that determine its behavior. The system is closed by $p = (\rho\theta)^\gamma$, and making use of (2.58) one can derive the relation*

$$p_2 = \lim_{\mathcal{M} \to 0} \frac{p - p_0}{\mathcal{M}^2} = \lim_{\mathcal{M} \to 0} \frac{(\rho\theta)^\gamma - (\rho_0 \theta_0)^\gamma}{\mathcal{M}^2} = \gamma (\rho_0 \theta_0)^{\gamma - 1} \left(\rho_0 \theta_2 + \rho_2 \theta_0 \right). \tag{2.61}$$

In the next step, we want to analyze to what extent the solutions of the compressible Euler equations correspond to those of the incompressible equations. Under the assumptions that in the density no constant fluctuations occur, i.e.

$$\rho = \rho_0 + \mathcal{O}(\mathcal{M}^2), \tag{2.62}$$

and that the HSE is fulfilled up to errors of order $\mathcal{O}(\mathcal{M}^2)$, i.e.

$$\nabla p + \rho \nabla \Phi = \mathcal{O}(\mathcal{M}^2), \tag{2.63}$$

the Euler equations (2.23) become

$$\begin{aligned} \nabla \cdot (\rho \mathbf{v}) &= \mathcal{O}(\mathcal{M}^2), \\ \partial_t \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{v} + \frac{\nabla p_2}{\rho_0} &= -\frac{\rho_2 \nabla \Phi}{\rho_0} + \mathcal{O}(\mathcal{M}^2), \\ \partial_t \theta + \mathbf{v} \cdot \nabla \theta &= 0. \end{aligned} \tag{2.64}$$

The solutions of (2.23) thus agree with those of the incompressible model up to an error of order $\mathcal{O}(\mathcal{M}^2)$.

It is interesting to investigate the behaviour of the kinetic energy in the incompressible limit $\mathcal{M} \to 0$. We can derive an evolution equation for the kinetic energy $E_{\text{kin}} = \frac{1}{2}\mathcal{M}^2 \rho |\mathbf{v}|^2$ from the rescaled equations in (2.50) given by

$$\partial_t E_{\text{kin}} + \nabla \cdot (E_{\text{kin}} \mathbf{v}) = -\mathbf{v} \cdot (\nabla p + \rho \nabla \Phi)\,. \tag{2.65}$$

In the low Mach number limit, solutions of the compressible Euler equations satisfy the hydrostatic equilibrium up to errors of order $\mathcal{O}(\mathcal{M}^2)$, so that the kinetic energy is conserved in the limit.

2.4 Compressible Ideal MHD Equations

The equations of MHD describe electrically conducting fluids in the presence of a magnetic field. They can be used to model a wide range of physical phenomena, including astrophysical systems such as stars and galaxies, plasma physics, and fusion reactors. Their simplest form is given by the ideal MHD equations, in which the influence of fluid viscosity is neglected. This model consists of a system of nonlinear hyperbolic PDEs that involve the conservation of mass, momentum, and total energy, along with Faraday's law for the magnetic field. A brief derivation can for example be found in [LeV98]. Taking gravitational forces into account, the following balance laws are obtained

$$\frac{\partial}{\partial t}\begin{pmatrix} \rho \\ \rho \mathbf{v} \\ E \\ \mathbf{B} \end{pmatrix} + \nabla \cdot \begin{pmatrix} \rho \mathbf{v} \\ \rho \mathbf{v} \otimes \mathbf{v} + (p + \frac{1}{2}|\mathbf{B}|^2)\mathbf{I} - \mathbf{B} \otimes \mathbf{B} \\ (E + p + \frac{1}{2}|\mathbf{B}|^2)\mathbf{v} - \mathbf{B}\,(\mathbf{B} \cdot \mathbf{v}) \\ \mathbf{v} \otimes \mathbf{B} - \mathbf{B} \otimes \mathbf{v} \end{pmatrix} = \begin{pmatrix} 0 \\ \rho \mathbf{g} \\ \rho \mathbf{v} \cdot \mathbf{g} \\ \mathbf{0} \end{pmatrix}, \tag{2.66}$$

where $\mathbf{B} = (B_x, B_y, B_z)$ represents the magnetic field[2]. All other notations correspond to those of the Euler equations. The total energy does now include the magnetic energy and therefore is defined by

$$E = \rho e + \frac{1}{2}\rho |\mathbf{v}|^2 + \frac{1}{2}|\mathbf{B}|^2. \tag{2.67}$$

The convex invariant set of physical admissible states for the MHD equations with $\mathcal{Q} = (\rho, \rho\mathbf{v}, E, \mathbf{B})$ is denoted by

$$\Omega_{\text{phys}}^{\text{MHD}} = \{\mathcal{Q} \in \mathbb{R}^8; \rho > 0, e > 0\}. \tag{2.68}$$

[2] We use the Lorentz-Heaviside units: $\mathbf{B} = \mathbf{b}/\sqrt{4\pi}$.

2.4.1 Solenoidal Constraint

The ideal MHD equations (2.66) are coupled with an additional solenoidal constraint for the magnetic field

$$\nabla \cdot \mathbf{B} = 0. \tag{2.69}$$

This constraint has its origin in Maxwell's equation and states that physically no magnetic monopoles can exist. Solutions to (2.66) automatically satisfy this condition at all times if the initial magnetic field obeys the constraint. This can easily be illustrated by rewriting the induction equation into the equivalent form

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathcal{E} = 0, \tag{2.70}$$

where $\mathcal{E} = -\mathbf{v} \times \mathbf{B}$ is the *electromotive force*. Applying the divergence to equation (2.70) results in

$$\frac{\partial (\nabla \cdot \mathbf{B})}{\partial t} = 0. \tag{2.71}$$

The solenoidal constraint can also be incorporated into the ideal MHD equations by adding a source term proportional to the divergence of the magnetic field as shown by Godunov [God72]. The homogeneous ideal MHD system then transforms to

$$\frac{\partial}{\partial t}\begin{pmatrix} \rho \\ \rho\mathbf{v} \\ E \\ \mathbf{B} \end{pmatrix} + \nabla \cdot \begin{pmatrix} \rho\mathbf{v} \\ \rho\mathbf{v} \otimes \mathbf{v} + (p + \frac{1}{2}|\mathbf{B}|^2)\mathbf{I} - \mathbf{B} \otimes \mathbf{B} \\ (E + p + \frac{1}{2}|\mathbf{B}|^2)\mathbf{v} - \mathbf{B}\,(\mathbf{B} \cdot \mathbf{v}) \\ \mathbf{v} \otimes \mathbf{B} - \mathbf{B} \otimes \mathbf{v} \end{pmatrix} = -(\nabla \cdot \mathbf{B}) \begin{pmatrix} 0 \\ \mathbf{B} \\ \mathbf{v} \cdot \mathbf{B} \\ \mathbf{v} \end{pmatrix}. \tag{2.72}$$

At the continuous level, the source term only adds a zero to the original system due to the solenoidal constraint (2.69). Mathematically, adding the source term changes the character of the equations. Firstly, the new system of equations can be brought into symmetric hyperbolic form [Bar99, God72] and, secondly, it becomes invariant under the Galilei transformation [PRL+99]. At the same time, even without including gravitational terms, the system is no longer in conservative form, especially in the hydrodynamic equations. If the magnetostatic force density $\mathbf{B}\nabla \cdot \mathbf{B}$ is included in the expression of the Lorentz force when deriving the equations, the source terms in the equations for momentum and energy disappear [Jan00]. The resulting system of equations writes

$$\frac{\partial}{\partial t}\begin{pmatrix} \rho \\ \rho\mathbf{v} \\ E \\ \mathbf{B} \end{pmatrix} + \nabla \cdot \begin{pmatrix} \rho\mathbf{v} \\ \rho\mathbf{v} \otimes \mathbf{v} + (p + \frac{1}{2}|\mathbf{B}|^2)\mathbf{I} - \mathbf{B} \otimes \mathbf{B} \\ (E + p + \frac{1}{2}|\mathbf{B}|^2)\mathbf{v} - \mathbf{B}\,(\mathbf{B} \cdot \mathbf{v}) \\ \mathbf{v} \otimes \mathbf{B} - \mathbf{B} \otimes \mathbf{v} \end{pmatrix} = -(\nabla \cdot \mathbf{B}) \begin{pmatrix} 0 \\ \mathbf{0} \\ 0 \\ \mathbf{v} \end{pmatrix}. \tag{2.73}$$

Now energy and momentum are conserved even for $\nabla \cdot \mathbf{B} \neq 0$. In both systems (2.72) and (2.73), the divergence of the magnetic field is convected as a passive scalar, i.e.

$$\frac{\partial (\nabla \cdot \mathbf{B})}{\partial t} + \nabla \cdot (\mathbf{v} \nabla \cdot \mathbf{B}) = 0. \tag{2.74}$$

2.4.2 Entropy

The solenoidal constraint also has an influence on the entropy equation, as the following lemma shows.

Lemma 2.4.1. *Under the conditions (2.30), smooth solutions of the MHD equations (2.66), which obey the solenoidal constraint (2.69), satisfy an additional conservation law for the entropy of the form (2.31).*

Proof. The equation for the specific volume (2.35) can be derived as for the Euler equations. In order to derive an equation for the internal energy, analogous steps as in the Euler case lead to

$$\partial_t e + \mathbf{v} \cdot \nabla e + \tau p \nabla \cdot \mathbf{v} - \tau(\mathbf{v} \cdot \mathbf{B})(\nabla \cdot \mathbf{B}) = 0. \tag{2.75}$$

Under the assumption that the solenoidal constraint of the magnetic field in (2.69) holds, the last term vanishes from the equation and we end up with (2.37). The further steps are carried out as for Euler and lead to equation (2.31). □

Hence, an additional conservation law for the entropy emerges solely when the solenoidal constraint is fulfilled. Therefore, entropy conservation and solenoidal constraint are linked for the ideal MHD equations. This connection is underlined by the observation that only the eight wave formulation (2.72) can be reformulated in symmetric hyperbolic form [DGWW18]. The Godunov-Mock theorem states the absence of a strictly convex entropy for systems of conservation laws that lack symmetrization [FL71, GR02].

2.4.3 Eigenstructure

In order to analyze the eigenstructure of the MHD system, we consider its one-dimensional version with $B_x = const$. The eigenvalues associated with the Jacobian of the quasilinear system are given by

$$\lambda_{1,8}^{\mathrm{MHD}} = u \mp c_{\mathrm{f},x}, \quad \lambda_{2,7}^{\mathrm{MHD}} = u \mp c_{\mathrm{a},x}, \quad \lambda_{3,6}^{\mathrm{MHD}} = u \mp c_{\mathrm{s},x}, \quad \lambda_{4}^{\mathrm{MHD}} = u, \quad \lambda_{5,9}^{\mathrm{MHD}} = 0. \tag{2.76}$$

The Alfvén wave speed (c_{a}) and the slow (c_{s}) and fast (c_{f}) magnetosonic wave speeds herein are defined by

$$c_{\mathrm{a},x} = \frac{|B_x|}{\sqrt{\rho}}, \tag{2.77}$$

$$c_{\mathrm{f,s},x} = \left[\frac{1}{2} \left(c^2 + \frac{|\mathbf{B}|^2}{\rho} \pm \sqrt{\left(c^2 + \frac{|\mathbf{B}|^2}{\rho} \right)^2 - 4c^2 c_{\mathrm{a},x}^2} \right) \right]^{\frac{1}{2}}. \tag{2.78}$$

The MHD system is hyperbolic, but not strictly hyperbolic (not even in the homogeneous case), because for $\{c_{a,y} + c_{a,z} = 0\}$, $\{c_{a,x} = 0\}$ or $\{c_{a,y} + c_{a,z} = 0 \ \wedge \ c_{a,x} = c\}$ wave speeds do coincide. The analysis of these umbilic points is out of the scope of this work. Interested readers are referred to [LeV98] and references therein.

2.4.4 Equilibrium Solutions

The presence of the magnetic field in the MHD system (2.66) affects the equilibrium solutions. Stationary steady states, for which all time derivatives are zero and the velocity is zero everywhere, are called *magnetohydrostatic equilibria* (MHSE). These states satisfy

$$\begin{cases} \mathbf{v} = \mathbf{0}, \\ \nabla \cdot \left(\left(p + \frac{1}{2}|\mathbf{B}|^2 \right) \mathbf{I} - \mathbf{B} \otimes \mathbf{B} \right) = \rho \mathbf{g}. \end{cases} \tag{2.79}$$

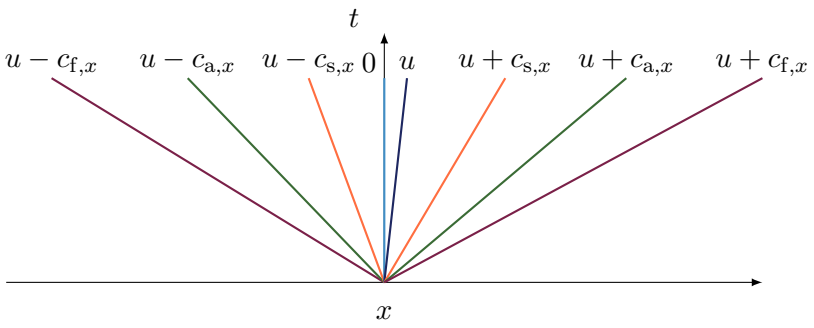

Figure 2.2: Wave structure of the one-dimensional MHD equations including gravity. Fast/slow magnetosonic waves (maroon/orange) are shocks or rarefactions, Alfvén waves (green) are rotational discontinuities, contact wave (indigo) and zero wave (sky blue) are contact discontinuities.

The equilibrium equation is undetermined, so that it offers an entire continuum of magnetohydrostatic solutions. The MHSE plays an important role because it accurately describes the stratification of stars for most of their lifespan. Until later evolutionary stages, the equilibrium underlies only minor perturbations so that the structure remains close to the equilibrium. Substantial changes only occur over longer thermal and nuclear time scales [KWW13, LBA+22].

2.4.5 Dimensionless Equations and Incompressible Limit

In the following we consider the dimensionless form of the ideal MHD equations. All quantities are rescaled as for the Euler equations. In addition, the magnetic field is rescaled by $\mathbf{B} = B_\mathrm{r}\hat{\mathbf{B}}$ and we introduce the Alfvén Mach number $\mathcal{M}_\mathrm{Alf} = |v_\mathrm{r}|/(|B_\mathrm{r}|/\sqrt{\rho_\mathrm{r}})$. The following equations result:

$$\frac{\partial}{\partial\hat{t}}\begin{pmatrix}\hat{\rho}\\ \hat{\rho}\hat{\mathbf{v}}\\ \hat{E}\\ \hat{\mathbf{B}}\end{pmatrix} + \hat{\nabla}\cdot\begin{pmatrix}\hat{\rho}\tilde{\mathbf{v}}\\ \hat{\rho}\hat{\mathbf{v}}\otimes\hat{\mathbf{v}} + (\frac{\hat{p}}{\mathcal{M}^2} + \frac{1}{2}\frac{|\hat{\mathbf{B}}|^2}{\mathcal{M}^2_\mathrm{Alf}})\mathbf{I} - \frac{\hat{\mathbf{B}}\otimes\hat{\mathbf{B}}}{\mathcal{M}^2_\mathrm{Alf}}\\ (\hat{E}+\hat{p}+\frac{1}{2}|\hat{\mathbf{B}}|^2\frac{\mathcal{M}^2}{\mathcal{M}^2_\mathrm{Alf}})\hat{\mathbf{v}} - \hat{\mathbf{B}}\left(\hat{\mathbf{B}}\cdot\hat{\mathbf{v}}\right)\frac{\mathcal{M}^2}{\mathcal{M}^2_\mathrm{Alf}}\\ \hat{\mathbf{v}}\otimes\hat{\mathbf{B}} - \hat{\mathbf{B}}\otimes\hat{\mathbf{v}}\end{pmatrix} = \begin{pmatrix}0\\ -\frac{1}{Fr^2}\hat{\rho}\nabla\hat{\Phi}\\ -\frac{\mathcal{M}^2}{Fr^2}\hat{\rho}\hat{\mathbf{v}}\cdot\nabla\hat{\Phi}\\ \mathbf{0}\end{pmatrix}. \tag{2.80}$$

The eigenvalues corresponding to the one-dimensional dimensionless MHD equations are given by

$$\hat{\lambda}^\mathrm{MHD}_{1,8} = \hat{u}\mp\hat{c}_{f,x}, \quad \hat{\lambda}^\mathrm{MHD}_{2,7} = \hat{u}\mp\hat{c}_{a,x}, \quad \hat{\lambda}^\mathrm{MHD}_{3,6} = \hat{u}\mp\hat{c}_{s,x}, \quad \hat{\lambda}^\mathrm{MHD}_{4} = \hat{u}, \quad \hat{\lambda}^\mathrm{MHD}_{5,9} = 0, \tag{2.81}$$

with the speeds

$$\hat{c}_{a,x} = \frac{1}{\mathcal{M}_\mathrm{Alf}}\frac{|\hat{B}_x|}{\sqrt{\hat{\rho}}}, \tag{2.82}$$

$$\hat{c}_{f,s,x} = \left[\frac{1}{2}\left(\frac{1}{\mathcal{M}^2}\hat{c}^2 + \frac{1}{\mathcal{M}_\mathrm{Alf}}\frac{|\hat{\mathbf{B}}|^2}{\hat{\rho}} \pm \sqrt{\left(\frac{1}{\mathcal{M}^2}\hat{c}^2 + \frac{1}{\mathcal{M}_\mathrm{Alf}}\frac{|\hat{\mathbf{B}}|^2}{\hat{\rho}}\right)^2 - \frac{1}{\mathcal{M}^2}4\hat{c}^2\hat{c}^2_{a,x}}\right)\right]^{\frac{1}{2}}. \tag{2.83}$$

From this point on, we will again omit the hat notation for the rest of this subsection.

In this work we are interested in astrophysical applications, in which the sonic Mach number $\mathcal{M}$ can become very small, while the Alfvén Mach number is $\mathcal{M}_{\text{Alf}} \sim 1$. Therefore, at this point we only consider the limit $\mathcal{M} \to 0$ and keep $\mathcal{M}_{\text{Alf}} = 1$. In this limit, the solutions of the MHD equations (2.23) tend to the solutions of the incompressible MHD equations [MB88]. In the following, we perform analogous steps as in the Euler case to derive the limit equations. Under the assumption of a divergence-free magnetic field, we derive the equation for the pressure given by (2.54) from (2.75). Using the definition $p = (\rho\theta)^\gamma$, we deduce the transport equation for the potential temperature θ and replace the energy equation in the dimensionless MHD equations by this transport equation. The resulting system writes

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) &= 0, \\ \partial_t (\rho \mathbf{v}) + \nabla \cdot \left(\rho \mathbf{v} \otimes \mathbf{v} + \frac{1}{\mathcal{M}^2} p \mathbf{I} \right) &= -\frac{1}{\mathcal{M}^2} \rho \nabla \Phi, \\ \partial_t \theta + \mathbf{v} \cdot \nabla \theta &= 0, \\ \partial_t \mathbf{B} + \nabla \cdot (\mathbf{v} \otimes \mathbf{B} - \mathbf{B} \otimes \mathbf{v}) &= 0. \end{aligned} \tag{2.84}$$

Inserting expansions in terms of $\mathcal{M}$ given by

$$\begin{aligned} &\rho = \rho_0 + \mathcal{M}^2 \rho_2 + \mathcal{O}(\mathcal{M}^3), && \mathbf{v} = \mathbf{v}_0 + \mathcal{M} \mathbf{v}_1 + \mathcal{M}^2 \mathbf{v}_2 + \mathcal{O}(\mathcal{M}^3), \\ &p = p_0 + \mathcal{M}^2 p_2 + \mathcal{O}(\mathcal{M}^3), && \theta = \theta_0 + \mathcal{M}^2 \theta_2 + \mathcal{O}(\mathcal{M}^3), \\ &\mathbf{B} = \mathbf{B}_0 + \mathcal{M} \mathbf{B}_1 + \mathcal{M}^2 \mathbf{B}_2 + \mathcal{O}(\mathcal{M}^3), \end{aligned} \tag{2.85}$$

into the dimensionless MHD equations and collecting terms of order $\mathcal{O}(\mathcal{M}^{-2})$ yields

$$\nabla p_0 = -\rho_0 \nabla \Phi. \tag{2.86}$$

Under the assumption of a background stratification with constant potential temperature θ_0, the limit equations are defined by

$$\begin{aligned} \nabla \cdot (\rho_0 \mathbf{v}_0) &= 0, \\ \partial_t \mathbf{v}_0 + \mathbf{v}_0 \cdot \nabla \mathbf{v}_0 + \frac{1}{\rho_0} \nabla \cdot \left(\left(p_2 + \frac{1}{2} |\mathbf{B}_0|^2 \right) \mathbf{I} - \mathbf{B}_0 \otimes \mathbf{B}_0 \right) &= -\frac{\rho_2 \nabla \Phi}{\rho_0}, \\ \partial_t \theta_2 + \mathbf{v}_0 \cdot \nabla \theta_2 &= 0, \\ \partial_t \mathbf{B}_0 + \nabla \cdot (\mathbf{v}_0 \otimes \mathbf{B}_0 - \mathbf{B}_0 \otimes \mathbf{v}_0) &= \mathbf{0}. \end{aligned} \tag{2.87}$$

Remark 2.4.2. *In order to avoid that no conditions apply to the unknown ρ_2, we close the system by $p = (\rho\theta)^\gamma$ and*

$$p_2 = \lim_{\mathcal{M} \to 0} \frac{p - p_0}{\mathcal{M}^2} = \lim_{\mathcal{M} \to 0} \frac{(\rho\theta)^\gamma - (\rho_0 \theta_0)^\gamma}{\mathcal{M}^2} = \gamma (\rho_0 \theta_0)^{\gamma - 1} (\rho_0 \theta_2 + \rho_2 \theta_0). \tag{2.88}$$

In consequence, all variables can be determined.

Under the assumptions (2.62) and (2.63), the MHD equations (2.66) become

$$\begin{aligned} \nabla \cdot (\rho \mathbf{v}) &= \mathcal{O}(\mathcal{M}^2), \\ \partial_t \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{v} + \frac{1}{\rho_0} \nabla \cdot \left(\left(p_2 + \frac{1}{2} |\mathbf{B}|^2 \right) \mathbf{I} - \mathbf{B} \otimes \mathbf{B} \right) &= -\frac{\rho_2 \nabla \Phi}{\rho_0} + \mathcal{O}(\mathcal{M}^2), \\ \partial_t \theta + \mathbf{v} \cdot \nabla \theta &= 0, \\ \partial_t \mathbf{B} + \nabla \cdot (\mathbf{v} \otimes \mathbf{B} - \mathbf{B} \otimes \mathbf{v}) &= \mathbf{0}. \end{aligned} \tag{2.89}$$

The solutions of the MHD equations (2.66) thus agree with those of the incompressible model up to an error of order $\mathcal{O}(\mathcal{M}^2)$.

As for the Euler equations, we can derive a time evolution equation for the kinetic energy from the rescaled equations (2.80) with $\mathcal{M}_{\mathrm{Alf}} = 1$, which has the form

$$\partial_t E_{\mathrm{kin}} + \nabla \cdot (E_{\mathrm{kin}} \mathbf{v}) = -\mathbf{v} \cdot (\nabla p + \rho \nabla \Phi) - \mathcal{M}^2 \mathbf{v} \cdot \left[\nabla \cdot \left(\frac{1}{2} |\mathbf{B}|^2 \mathbf{I} - \mathbf{B} \otimes \mathbf{B} \right) \right]. \tag{2.90}$$

The kinetic energy is thus conserved in the incompressible limit up to errors of order $\mathcal{O}(\mathcal{M}^2)$.

Chapter 3

Finite Volume Methods

In general, IVPs for conservation and balance laws such as the Euler and ideal MHD equations are difficult to solve exactly. Therefore, it becomes necessary to rely on numerical algorithms that provide approximate solutions. One widely used approach are FV methods, which discretize the integral form of the PDE. This discretization naturally ensures conservation, thereby reflecting the physical nature of the underlying PDE. Additionally, FV methods are valued for their robustness in the face of shock waves and discontinuities and their applicability to a wide range of grid structures, making them a flexible and effective tool in practical applications.
In this chapter, we describe the basic concepts that are used in the development of FV methods. Our presentation is based on the explanations given in standard textbooks on FV methods [LeV02, Tor09, Bou04, GR02]. For more details, the reader is referred to this literature and references therein.

3.1 Finite Volume Approach

In the following, we consider one spatial dimension and derive a numerical scheme for the IVP

$$\begin{cases} \mathcal{Q}(x,t)_t + (\mathcal{F}(\mathcal{Q}(x,t))_x = 0, \\ \mathcal{Q}_0(x) = \mathcal{Q}(x,0), \end{cases} \tag{3.1}$$

on the spatial domain $\mathcal{I} = [x_L, x_R]$ and for $t \in \mathbb{R}_{\geq 0}$. The extension of the numerical concepts to multi-dimensional spaces is given in Sect. 3.7. In order to discretize the spatial domain, we subdivide $\mathcal{I}$ into N_x cells $C_i = (x_{i-1/2}, x_{i+1/2})$ with the uniform cell size Δx. The cell center of cell i is denoted by $x_i = (x_{i+1/2} + x_{i-1/2})/2$. The time is discretized by $t^n = n\Delta t$ for $n = 1, 2, ...$ with a time step Δt.
The core of FV methods is the approximation of the solution on a cell C_i at time t^n by the cell average

$$Q_i^n \approx \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{Q}(x, t^n) \mathrm{d}x \tag{3.2}$$

and the goal is to compute the cell averages at the next time level t^{n+1}. In order to derive an update formula for the cell averages, we integrate the conservation law in (3.1) over $[x_{i-1/2}, x_{i+1/2}) \times [t^n, t^{n+1})$ and get

$$\int_{t^n}^{t^{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{Q}(x,t)_t \mathrm{d}x\mathrm{d}t + \int_{t^n}^{t^{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{F}\left(\mathcal{Q}(x,t)\right)_x \mathrm{d}x\mathrm{d}t = 0. \tag{3.3}$$

Then, applying the fundamental theorem of calculus results in

$$\begin{aligned}&\int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{Q}(x,t^{n+1})\mathrm{d}x - \int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{Q}(x,t^{n})\mathrm{d}x \\ &= -\int_{t^n}^{t^{n+1}} \mathcal{F}\left(\mathcal{Q}(x_{i+1/2},t)\right)\mathrm{d}t + \int_{t^n}^{t^{n+1}} \mathcal{F}\left(\mathcal{Q}(x_{i-1/2},t)\right)\mathrm{d}t.\end{aligned} \tag{3.4}$$

By replacing the integrals on the left-hand side with the cell averages from (3.2), we obtain

$$Q_i^{n+1} = Q_i^n - \frac{1}{\Delta x}\left(\int_{t^n}^{t^{n+1}} \mathcal{F}\left(\mathcal{Q}(x_{i+1/2},t)\right)\mathrm{d}t - \int_{t^n}^{t^{n+1}} \mathcal{F}\left(\mathcal{Q}(x_{i-1/2},t)\right)\mathrm{d}t\right). \tag{3.5}$$

In general, the exact evaluation of the integrals is difficult and expensive. Therefore, we approximate the flux average by a numerical flux

$$F_{i+1/2} \approx \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \mathcal{F}\left(\mathcal{Q}(x_{i+1/2},t)\right)\mathrm{d}t. \tag{3.6}$$

If the numerical flux only depends on data at time t^n, i.e. on Q^n, we derive an explicit formula to compute an approximate solution at time t^{n+1} given by

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^n - F_{i-1/2}^n\right). \tag{3.7}$$

However, if the numerical flux is evaluated based on Q^{n+1}, the update formula becomes implicit and reads

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^{n+1} - F_{i-1/2}^{n+1}\right), \tag{3.8}$$

which in consequence leads to a nonlinear system of equations that must be solved to obtain Q^{n+1}.

In the following, we continue with time-explicit schemes of the form (3.7). Given that information propagates at a finite speed for hyperbolic equations, we calculate the numerical flux at the interface solely based on the cell averages of the two neighboring cells, i.e.

$$F_{i+1/2}^n = F(Q_i^n, Q_{i+1}^n). \tag{3.9}$$

This definition leads to a three-point method, in the sense that the solution Q_i^{n+1} depends on the three cell averages Q_{i-1}^n, Q_i^n and Q_{i+1}^n. As a result, we obtain a first order approximation.

An important property of the method (3.7) is that it is *conservative*. Summing up over $\Delta x Q_i^{n+1}$ shows that all flux contributions cancel out and only the fluxes at the boundaries remain

$$\Delta x \sum_{i=1}^{N_x} Q_i^{n+1} = \Delta x \sum_{i=1}^{N_x} Q_i^n - \Delta t\left(F_{N_x+1/2}^n - F_{1/2}^n\right). \tag{3.10}$$

However, the conservation property does not automatically lead to stability of the numerical method. In order to be stable, it must at least fulfill two further necessary conditions. First, the numerical flux must be *consistent* with the physical flux in the sense of the following definition.

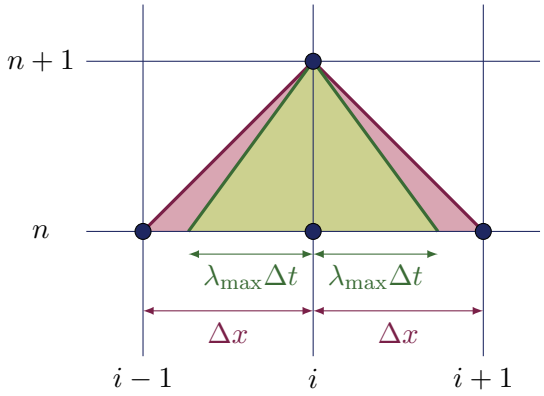

Figure 3.1: The numerical domain of dependence (red) of the three-point method. For a method to be stable, its numerical domain of dependence must be larger than the true domain of dependence (green).

Definition 3.1.1. *A numerical flux is called* ***consistent****, if for* $Q_i^n = Q_{i+1}^n = Q$ *it satisfies the condition*

$$F(Q, Q) = \mathcal{F}(Q). \tag{3.11}$$

The consistency of a numerical flux ensures that its behavior in the simple case of a constant solution exactly matches the behavior of the physical flux.
Second, it is important from which spatial region information can influence a given cell's value. This region is called *numerical domain of dependence* and must necessarily be larger than or equal to the *true domain of dependence* on the level of the PDE. We summarize this constraint in the following lemma.

Lemma 3.1.2. *A numerical method can only be convergent if its numerical domain of dependence contains the true domain of dependence of the PDE, at least in the limit as* Δt *and* Δx *go to zero [LeV02].*

The true domain of dependence for hyperbolic conservation laws depends on the eigenvalues of the flux-Jacobian $\mathcal{F}'(\mathcal{Q})$, since it follows from the characteristic form of the hyperbolic system of conservation laws that information propagates with the speed of these eigenvalues. In the time-explicit three-point method, on the other hand, information may be transported a maximum of one cell width during a time step, because Q_i^{n+1} depends only on Q_{i-1}^n, Q_i^n and Q_{i+1}^n. Thus, the fraction of cell width Δx and time step Δt must be greater than or equal to the maximum absolute wave speed $\lambda_{\max} = \max_j |\lambda_j|$ (see also Fig. 3.1). The CFL condition follows from this consideration, i.e.

$$\Delta t \leq C_{\mathrm{CFL}} \frac{\Delta x}{\lambda_{\max}} \tag{3.12}$$

with a CFL number $C_{\mathrm{CFL}} \leq 1$ for the three-point method. Note that for other methods having a different numerical domain of dependence, a different CFL number is necessary for stability.

Remark 3.1.3. *The implicit method (3.8) is unconditionally stable with respect to the CFL condition, since its numerical domain of dependence includes the whole computational domain. This is the result of having to solve a nonlinear system of equations depending on all cell averages in the computational domain.*

The consistency condition (3.11) and the CFL condition (3.12) are necessary but not always sufficient conditions for stability and convergence of a finite volume method. A useful property that improves the stability of the numerical method is to ensure that it satisfies a discrete version of an entropy inequality

$$\eta(Q_i^{n+1}) \leq \eta(Q_i^n) - \frac{\Delta t}{\Delta x}\left(\mathcal{F}_{i+1/2}^{\text{ent},n} - \mathcal{F}_{i-1/2}^{\text{ent},n}\right). \tag{3.13}$$

Analogous to the entropy inequality (2.18), ensuring (3.13) excludes unphysical discrete solutions for which the mathematical entropy increases.

A crucial theorem on convergence of finite volume methods is provided by Lax and Wendroff [LW60].

Theorem 3.1.4 (Lax-Wendroff theorem, taken from [LeV02])**.** *Consider a sequence of grids* $(\Delta t^{(j)}, \Delta x^{(j)})$ *for which* $\Delta t^{(j)}, \Delta x^{(j)} \to 0$ *as* $j \to \infty$ *and let the function* $Q^{(j)}(x,t)$ *denote a numerical solution computed by a consistent and conservative method on the* j*-th grid. If* $Q^{(j)}$ *converges to a function* $\mathcal{Q}$ *as* $j \to \infty$ *in the sense that*

$$\int_0^{t_f}\int_{x_L}^{x_R} |Q^{(j)}(x,t) - \mathcal{Q}(x,t)|\mathrm{d}x\mathrm{d}t \to 0 \text{ as } j \to \infty \tag{3.14}$$

and that for each t_f *there is a constant* $C > 0$ *such that*

$$TV\left(Q^{(j)}(\cdot,t)\right) < C \quad \text{for all } 0 \leq t \leq t_f, \quad j = 1,2,\ldots, \tag{3.15}$$

then $\mathcal{Q}(x,t)$ *is a weak solution of the conservation law.*

The function $TV(\cdot)$ used in the theorem denotes the total variation function and can be computed by $TV(q) = \int_{x_L}^{x_R} |q'(x)|\mathrm{d}x$. It is important to notice that the Lax-Wendroff theorem does not guarantee convergence. For achieving this, the numerical method must also be stable and even then, different sequences can converge against different weak solutions. Nevertheless, the theorem is helpful because it allows us to assume that a physically reasonable numerical solution, which is derived by a consistent and conservative method, is a good approximation of some weak solution [LeV02].

3.2 Godunov's Method

The method (3.7) provides a way to compute a numerical solution at the next time step. What is still missing is a good approximation for the flux $F_{i+1/2}$ in (3.6). Godunov [God59] has found that the piecewise constant approximation by cell averages gives rise to local Riemann problems at the cell interfaces, which have the form

$$\begin{aligned} &\mathcal{Q}_t + \mathcal{F}(\mathcal{Q})_x = 0, \\ &\mathcal{Q}_0(x) = \begin{cases} Q_i^n, & \text{if } x < x_{i+1/2}, \\ Q_{i+1}^n, & \text{if } x > x_{i+1/2}. \end{cases} \end{aligned} \tag{3.16}$$

Under the assumption that a solution to this Riemann problem exists, he proposes to solve (3.16) exactly, so that a solution $\widetilde{\mathcal{W}}(x, t^{n+1})$ at the new time level t^{n+1} can be constructed, which consists piecewise of the Riemann solutions

$$\mathcal{W}_{i+1/2}\left(\frac{x - x_{i+1/2}}{t - t^n}; Q_i^n, Q_{i+1}^n\right) \quad \text{for } x \in (x_i, x_{i+1}). \tag{3.17}$$

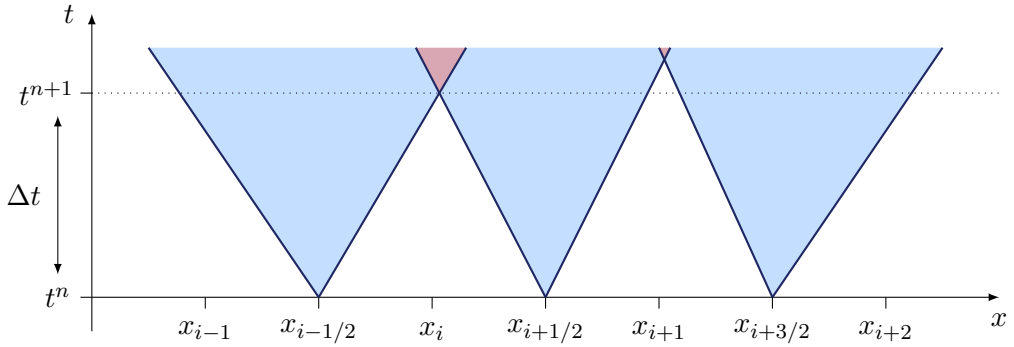

Figure 3.2: Maximum waves of the Riemann problems at $x_{i-1/2}$, $x_{i+1/2}$ and $x_{i+3/2}$. The dashed line marks the maximum time step Δt permitted by the CFL condition (3.19) for Godunov's method.

Finally, the cell averages for the next time level can be obtained by an averaging step over each cell

$$Q_i^{n+1} = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \widetilde{\mathcal{W}}(x, t^{n+1}) \mathrm{d}x. \tag{3.18}$$

When using this approach it is important that the waves of the individual Riemann solutions must not cross (see Fig. 3.2). This results in the following more restrictive CFL condition

$$\Delta t \leq C_{\mathrm{CFL}} \frac{\Delta x}{\lambda_{\max}} \quad \text{with} \quad C_{\mathrm{CFL}} \leq \frac{1}{2}. \tag{3.19}$$

Godunov's method can be summarized by the *Reconstruction-Evolution-Averaging* (REA) algorithm.

Algorithm 3.2.1. *Godunov's method consists of three consecutive steps:*

1. ***Reconstruction:*** *The piecewise constant solution $\widetilde{\mathcal{W}}(x, t^n)$ is constructed from the cell averages Q_i^n.*

2. ***Evolution:*** *The reconstructed function is evolved in time by solving local Riemann problems of the form (3.16) at the cell interfaces by an exact Riemann solver $\mathcal{W}_{i+1/2}$.*

3. ***Averaging:*** *The solution at the new time level $\widetilde{\mathcal{W}}(x, t^{n+1})$ is averaged over each grid cell to determine Q_i^{n+1}.*

Godunov's method can be translated into the conservative form (3.7). For this purpose we reformulate the integral form (3.4) into

$$\begin{aligned} \int_{x_{i-1/2}}^{x_{i+1/2}} \widetilde{\mathcal{W}}(x, t^{n+1}) \mathrm{d}x = & \int_{x_{i-1/2}}^{x_{i+1/2}} \widetilde{\mathcal{W}}(x, t^n) \mathrm{d}x \\ & + \int_0^{\Delta t} \widetilde{\mathcal{W}}(x_{i-1/2}, t) \mathrm{d}t - \int_0^{\Delta t} \widetilde{\mathcal{W}}(x_{i+1/2}, t) \mathrm{d}t. \end{aligned} \tag{3.20}$$

The Riemann solution is self-similar and thus constant along the ray $\frac{x}{t} = const$, which allows us to write

$$\begin{aligned} \widetilde{\mathcal{W}}(x_{i-1/2}, t) &= \mathcal{W}_{i-1/2}(0) = const, \\ \widetilde{\mathcal{W}}(x_{i+1/2}, t) &= \mathcal{W}_{i+1/2}(0) = const. \end{aligned} \tag{3.21}$$

Then, by dividing the integral form (3.20) by Δx and using the definition of the cell averages, we derive the conservative form (3.7) with the numerical flux defined by

$$F^n_{i+1/2} = F(Q^n_i, Q^n_{i+1}) = \mathcal{F}\left(\mathcal{W}_{i+1/2}(0; Q^n_i, Q^n_{i+1})\right). \tag{3.22}$$

Here, only the Riemann solution at the interface is used. Therefore, the waves are allowed to travel over the entire width of a cell, so that a less restrictive CFL number $C_{\text{CFL}} \leq 1$ is sufficient.

The latter approach only uses the Riemann solution at the interface. Nevertheless, one needs to determine the whole Riemann solution, which in general is difficult and costly. Therefore, it is worth considering not solving the Riemann problem exactly, but only approximately with the help of *approximate Riemann solvers*.

3.3 Approximate Riemann Solvers

In general, computing the exact Riemann solution for nonlinear hyperbolic systems is rather cumbersome and computationally inefficient. Roe pointed out that the Riemann problem does not need to be solved exactly, but that an approximate solution is sufficient in many cases. The result of his work was the *Roe solver* [Roe81], which is briefly described in Ex. 3.3.3. Methods that replace the exact Riemann solver by an approximate Riemann solver in the evolution step of the REA algorithm are called *Godunov-type methods*. Since the Riemann solver is the core part in Godunov's method, the efficiency of the scheme can significantly be increased by the use of approximate Riemann solvers. When computing the exact Riemann solution, it is particularly difficult to calculate rarefaction waves. Therefore, we consider approximate Riemann solvers for which the solution only contains shocks or contact discontinuities. In consequence, the solution consists of $K+1$ constant states w_k, $1 \leq k \leq K+1$, separated by waves propagating at speed s_k and has the general structure

$$\mathcal{W}_{\mathcal{R}}\left(\frac{x}{t}; Q^L, Q^R\right) = \begin{cases} w_1 = Q^L, & \text{if } \frac{x}{t} < s_1, \\ w_2, & \text{if } s_1 < \frac{x}{t} < s_2, \\ \vdots & \\ w_K, & \text{if } s_{K-1} < \frac{x}{t} < s_K, \\ w_{K+1} = Q^R, & \text{if } s_K < \frac{x}{t}. \end{cases} \tag{3.23}$$

This solution structure is illustrated in Fig. 3.3. In order to be consistent with the underlying conservation law, the Riemann solver $\mathcal{W}_{\mathcal{R}}$ should satisfy

$$\mathcal{W}_{\mathcal{R}}\left(\frac{x}{t}; Q, Q\right) = Q \tag{3.24}$$

and

$$\int_{x_i}^{x_{i+1}} \mathcal{W}_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}; Q_i, Q_{i+1}\right) \mathrm{d}x = \int_{x_i}^{x_{i+1}} \widetilde{\mathcal{W}}\left(\frac{x - x_{i+1/2}}{\Delta t}; Q_i, Q_{i+1}\right) \mathrm{d}x. \tag{3.25}$$

The approximate solution $\widetilde{\mathcal{W}}_\Delta(x, t^{n+1})$ at time level t^{n+1} then consists piecewise of the Riemann solutions at the interfaces. An advantage of this method is that some properties of the approximate Riemann solver transfer to the overall method, as can be seen in the following theorem.

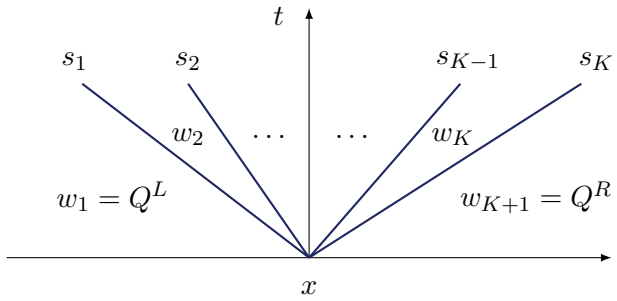

Figure 3.3: Structure of the general approximate Riemann solver (3.23). It consists of $K+1$ constant states separated by shocks or contact discontinuities.

Theorem 3.3.1. *Consider the explicit scheme (3.7) using an approximate Riemann solver of the form (3.23) under the CFL condition (3.19). If all constant states in the solver are in a set Ω_{phys}, then for the updated state it holds $Q_i^{n+1} \in \Omega_{\text{phys}}$.*

Proof. Using the integral form, the scheme (3.7) can be rewritten as follows

$$\begin{aligned} Q_i^{n+1} &= Q_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^n - F_{i-1/2}^n\right) \\ &= \int_{x_{i-1/2}}^{x_{i+1/2}} \widetilde{\mathcal{W}}_\Delta(x, t^{n+1})\mathrm{d}x \\ &= \int_{x_{i-1/2}}^{x_i} \mathcal{W}_{\mathcal{R}}\left(\frac{x - x_{i-1/2}}{\Delta t}; Q_{i-1}^n, Q_i^n\right)\mathrm{d}x + \int_{x_i}^{x_{i+1/2}} \mathcal{W}_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{\Delta t}; Q_i^n, Q_{i+1}^n\right)\mathrm{d}x. \end{aligned}$$

Since the constant states in the approximate Riemann solvers are assumed to be in Ω_{phys} and the set Ω_{phys} is assumed to be convex, both integrals are in Ω_{phys} and thus the statement of the theorem is proven. □

If the Riemann solver (3.23) satisfies (3.24), the CFL condition (3.19) and the Rankine-Hugoniot condition

$$\mathcal{F}(Q^R) - \mathcal{F}(Q^L) = \sum_{k=1}^{K} \lambda_k \left(Q_{k+1} - Q_k\right), \tag{3.26}$$

then it is a Godunov-type method [GR02]. It can be put into conservation form (3.7) with numerical fluxes defined by

$$F(Q^L, Q^R) = \frac{1}{2}\left(\mathcal{F}(Q^L) + \mathcal{F}(Q^R) - \sum_{k=1}^{K} |\lambda_k| \left(Q_{k+1} - Q_k\right)\right). \tag{3.27}$$

In the following, we briefly list some of the most important and most frequently used approximate Riemann solvers.

Example 3.3.2 (Rusanov solver)**.** *A very simple approximate Riemann solver is the Rusanov solver. It consists of three constant states separated by two waves, i.e.*

$$\mathcal{W}_{\mathcal{R}}\left(\frac{x}{t}; Q^L, Q^R\right) = \begin{cases} Q^L, & \text{if } \frac{x}{t} < -\lambda_{\max}^*, \\ Q^{\text{Rus}}, & \text{if } -\lambda_{\max}^* < \frac{x}{t} < \lambda_{\max}^*, \\ Q^R, & \text{if } \lambda_{\max}^* < \frac{x}{t}, \end{cases} \tag{3.28}$$

with the intermediate state

$$Q^{\text{Rus}} = \frac{1}{2}\left(Q^L + Q^R\right) - \frac{1}{2\lambda^*_{\max}}\left(\mathcal{F}(Q^R) - \mathcal{F}(Q^L)\right) \tag{3.29}$$

and the estimate of the maximum wave speed

$$\lambda^*_{\max} = \max\left(\lambda^L_{\max}, \lambda^R_{\max}\right). \tag{3.30}$$

By applying Rankine-Hugoniot conditions of the form (2.13) to the left and right wave, we can derive the well-known Rusanov flux function [Rus62]

$$F^{\text{Rus}}(Q^L, Q^R) = \frac{1}{2}\left(\mathcal{F}(Q^L) + \mathcal{F}(Q^R)\right) - \frac{\lambda^*_{\max}}{2}\left(Q^R - Q^L\right). \tag{3.31}$$

The flux function (3.31) consists of a central flux and an artificial dissipation term. This structure is typical of numerical fluxes, since the central flux alone is unconditionally unstable [LeV02]. The numerical flux resulting from Roe's solver is also structured in this way.

Example 3.3.3 (Roe solver)**.** *Roe's idea is to linearize the flux function by $\mathcal{F}(\mathcal{Q}) = \mathcal{A}\mathcal{Q}$ and to solve the Riemann problem for the simpler system*

$$\mathcal{Q}_t + \mathcal{A}\mathcal{Q}_x = 0. \tag{3.32}$$

The resulting Riemann solver has the form (3.23) with

$$\mathcal{A}(Q^L, Q^R)\left(w_{k+1} - w_k\right) = \lambda_k\left(w_{k+1} - w_k\right), \quad 1 \le k \le K. \tag{3.33}$$

Roe's method can be put in conservation form, in which the numerical flux is defined by

$$F^{\text{Roe}}(Q^L, Q^R) = \frac{1}{2}\left(\mathcal{F}(Q^L) + \mathcal{F}(Q^R)\right) - \left|\mathcal{A}(Q^L, Q^R)\right|\left(Q^R - Q^L\right). \tag{3.34}$$

In practice the Roe matrix can be computed by

$$\left|\mathcal{A}(Q^L, Q^R)\right| = \left|\mathcal{A}(Q^{\text{Roe}})\right| = \mathbf{R}(Q^{\text{Roe}})\left|\mathbf{\Lambda}(Q^{\text{Roe}})\right|\mathbf{R}(Q^{\text{Roe}})^{-1}, \tag{3.35}$$

where $\mathbf{R}(Q)$ denotes the matrix of right eigenvectors of the Jacobian and $\mathbf{\Lambda}(Q)$ is the diagonal matrix of the corresponding eigenvalues. The absolute value operator $|\cdot|$ is applied componentwise. The quantity Q^{Roe} represents an average of the two input states that needs to satisfy

$$\mathcal{A}(Q^{\text{Roe}})\left(Q^R - Q^L\right) = \mathcal{F}(Q^R) - \mathcal{F}(Q^L). \tag{3.36}$$

When using Roe's solver it is possible to compute quite accurate solutions in an efficient way, but the solver lacks some important properties. A major flaw of Roe's original solver is that it does not necessarily satisfy a discrete entropy inequality. Later, modifications to the Roe solver were proposed that overcome this weakness [HH83, Tad84, Osh84, Roe92, PQV01]. Building on Roe's ideas, other approximate Riemann solvers were developed in the following years. One class of such solvers are the *Harten-Lax-van Leer* (HLL) solvers.

Example 3.3.4 (HLL solver)**.** *The HLL Riemann solver [HLvL83] assumes that the Riemann solution consists of three constant states separated by two waves, i.e.*

$$\mathcal{W}_{\mathcal{R}}\left(\frac{x}{t}; Q^L, Q^R\right) = \begin{cases} Q^L, & \text{if } \frac{x}{t} < S^L, \\ Q^{\text{HLL}}, & \text{if } S^L < \frac{x}{t} < S^R, \\ Q^R, & \text{if } S^R < \frac{x}{t}, \end{cases} \tag{3.37}$$

and that the wave speeds S^L and S^R are given by some algorithm. The intermediate state Q^{HLL} is defined as the average of the exact Riemann solution between the slowest and fastest wave. This average is constant and can be computed by

$$Q^{\mathrm{HLL}} = \frac{S^R Q^R - S^L Q^L + \mathcal{F}(Q^L) - \mathcal{F}(Q^R)}{S^R - S^L}. \tag{3.38}$$

For more details on its derivation, see [Tor09]. There are different options to estimate the speeds S^L and S^R. A simple and yet robust choice proposed by Einfeldt *[Ein88] for ordered eigenvalues is given by*

$$S^L = \min\left(\lambda_1(Q^L), \lambda_1(Q^{\mathrm{Roe}})\right) \quad \textit{and} \quad S^R = \max\left(\lambda_m(Q^{\mathrm{Roe}}), \lambda_m(Q^R))\right). \tag{3.39}$$

Remark 3.3.5. *The HLL solver is equivalent to the Rusanov solver, if the wave speeds are estimated by*

$$S^L = -\lambda^*_{\max} \quad \textit{and} \quad S^R = \lambda^*_{\max}. \tag{3.40}$$

As long as the wave speed estimates S^L and S^R are a lower respective upper bound for the wave speeds of the exact Riemann solution, the HLL solver is entropy-satisfying [HLvL83, GR02] and positivity-preserving [ERMS91], leading to a good stability of the method. The HLL solver yields accurate results for systems with only two equations, such as the shallow water equations, but performs poorly for larger systems like the full Euler equations. From the simplification of the solution structure it follows that all intermediate states between Q^L and Q^R, which may exist in the exact solution of the Riemann problem, are approximated by only one intermediate state Q^{HLL}. For the Euler equations, this means that the contact discontinuity associated with the eigenvalue $\lambda_2^{\mathrm{Euler}} = u$ is not resolved. In the case of the MHD equations, the slow magnetosonic as well as the Alfvén waves are not resolved, either. This results in a lower accuracy of the method in these areas. To overcome this drawback for the Euler equations, Toro et al. constructed the *Harten-Lax-van Leer Contact* (HLLC) solver [TSS94].

Example 3.3.6 (HLLC solver)**.** *The HLLC solver for the Euler equations consists of four constant states separated by three waves. i.e.*

$$\mathcal{W}_{\mathcal{R}}\left(\frac{x}{t}; Q^L, Q^R\right) = \begin{cases} Q^L, & \text{if } \frac{x}{t} < S^L, \\ Q^{L*}, & \text{if } S^L < \frac{x}{t} < S^M, \\ Q^{R*}, & \text{if } S^M < \frac{x}{t} < S^R, \\ Q^R, & \text{if } S^R < \frac{x}{t}. \end{cases} \tag{3.41}$$

In order to determine the intermediate states Q^{L*} and Q^{R*} it is necessary to make additional assumptions about the solution. Toro et al. *[TSS94] assume the normal component of the velocity to be constant over the Riemann fan, i.e.*

$$u^{L*} = u^{R*} = S^M. \tag{3.42}$$

Batten et al. *[BCLC97] propose to compute this velocity by the HLL average Q^{HLL}. Then the remaining intermediate states can be calculated using the Rankine-Hugoniot conditions. Details can be found in [MK05].*

Just like the HLL solver, the HLLC solver is positivity-preserving and entropy-satisfying if the wave speeds are chosen appropriately [Bou04]. Thanks to the contact wave in the middle, the HLLC solver is able to resolve isolated contact discontinuities exactly.

There are extensions of the HLLC solver to the MHD equations [Gur04, Li05] and other systems of conservation and balance laws [Tor19]. However, those solvers do not always resolve isolated rotational discontinuities exactly, which may be caused by the two-state approximation in the Riemann solution. This can be cured by the HLLD solver developed by Miyoshi and Kusano for the MHD equations [MK05].

Example 3.3.7 (HLLD solver)**.** *The HLLD solver is a 5-wave solver of the form*

$$\mathcal{W}_{\mathcal{R}}\left(\frac{x}{t};Q^L,Q^R\right)=\begin{cases}Q^L, & \text{if } \frac{x}{t}<S^L,\\ Q^{L*}, & \text{if } S^L<\frac{x}{t}<S^{L*},\\ Q^{L**}, & \text{if } S^{L*}<\frac{x}{t}<S^M,\\ Q^{R**}, & \text{if } S^M<\frac{x}{t}<S^{R*},\\ Q^{R*}, & \text{if } S^{R*}<\frac{x}{t}<S^R,\\ Q^R, & \text{if } S^R<\frac{x}{t}.\end{cases} \tag{3.43}$$

Besides the entropy wave it also includes two Alfvén waves with speeds S^{L*} and S^{R*}, which are estimated by

$$S^{L*}=S^M-\frac{|B_x|}{\sqrt{\rho^{L*}}}\quad \text{and}\quad S^{R*}=S^M+\frac{|B_x|}{\sqrt{\rho^{R*}}}. \tag{3.44}$$

The assumptions for computing the intermediate states are that the normal velocity and the total pressure $p_T=p+\frac{1}{2}|\mathbf{B}|^2$ are constant over the Riemann fan. Then the respective intermediate states can be found by using the Rankine-Hugoniot conditions [MK05]. A possible estimate for the outer wave speeds is given by

$$S^L=\min\left(u^L,u^R\right)-\max(c^L_{f,x},c^R_{f,x})\quad \text{and}\quad S^R=\max\left(u^L,u^R\right)+\max(c^L_{f,x},c^R_{f,x}). \tag{3.45}$$

Remark 3.3.8. *In the case of a zero magnetic field, the Alfvén waves collapse to the contact wave and the HLLD solver reduces to the HLLC solver.*

The HLLD solver is able to resolve all isolated discontinuities formed in the ideal MHD system and is positivity-preserving [MK05].

Of course there exist more approximate Riemann solvers, numerical flux functions and variations of those presented above. For more information the reader is referred to [LeV02, Tor09, GR02] and the references therein.

We have seen that for HLL-type solvers it is not clear how the wave speeds should be estimated in order to obtain a stable method while keeping numerical dissipation as low as possible. In the next section we will take a closer look at another class of approximate Riemann solvers, for which the choice of speeds is more natural and which plays an important role in the further course of this work: *relaxation solvers*.

3.4 Relaxation Systems and Solvers

In the 1990s, the concept of relaxation schemes emerged [JX95, CLL94, CP98, Bou04]. The basic idea is to construct a new enlarged *relaxation system*, including a relaxation term on the right-hand side, which is an approximation of the original system. The numerical scheme then solves the relaxation system in two steps:

1. First solve the left-hand side of the relaxation system, which consists of a linear transport and is therefore numerically easy to solve.
2. Then project the solution of the first step back onto the equilibrium variables, i.e. use only the variables of the original system to solve the next time step.

Thus, the resulting numerical method is simple and yet leads to rather accurate results. Another advantage is that Riemann solvers associated with relaxation systems naturally satisfy a discrete entropy inequality, which results in an increased robustness of the method. Since there is a certain degree of freedom in how to construct the relaxation system, it is moreover possible to equip the approximate Riemann solver with additional desirable properties, as we will see later in Chapter 4. The description of the relaxation concept given in this section follows the lines of [BK23b].

3.4.1 Jin-Xin Relaxation Model

In order to get a deeper understanding of the concept of relaxation, let us first consider the simple case of a scalar conservation law

$$\partial_t q + \partial_x f(q) = 0. \tag{3.46}$$

To solve this equation, Jin and Xin [JX95] introduced the relaxation system

$$\begin{aligned} \partial_t q + \partial_x \nu &= 0, \\ \partial_t \nu + a^2 \partial_x q &= \frac{1}{\varepsilon} \left(f(q) - \nu \right), \end{aligned} \tag{3.47}$$

with a relaxation variable ν, a constant relaxation speed a and a relaxation parameter ε. The relaxation system (3.47) is derived by multiplying the original conservation law (3.46) by $f'(q)$, which yields

$$\partial_t f(q) + f'(q)^2 \partial_x q = 0. \tag{3.48}$$

This equation is linearized by replacing $f'(q)^2$ by a^2. This alone would still lead to a very poor approximation, so $f(q)$ is replaced by the new relaxation variable ν and a relaxation source term connecting $f(q)$ and ν is added. The resulting relaxation system (3.47) is a diffusive approximation of the original scalar conservation law in (3.46). This can be illustrated by a Chapman-Enskog expansion [CC90]. For this procedure we consider a formal expansion of ν in terms of ε

$$\nu = \nu_0 + \varepsilon \nu_1 + \mathcal{O}(\varepsilon^2), \tag{3.49}$$

and insert this expansion into system (3.47)

$$\begin{aligned} \partial_t q + \partial_x \left(\nu_0 + \varepsilon \nu_1 \right) &= 0, \\ \partial_t \left(\nu_0 + \varepsilon \nu_1 \right) + a^2 \partial_x q &= \frac{1}{\varepsilon} \left(f(q) - \nu_0 - \varepsilon \nu_1 \right). \end{aligned} \tag{3.50}$$

From collecting all terms of order $\mathcal{O}(1/\varepsilon)$, we can determine

$$\nu_0 = f(q). \tag{3.51}$$

For the terms with order $\mathcal{O}(1)$, on the other hand, we gain the system

$$\begin{aligned} \partial_t q + \partial_x \nu_0 &= 0, \\ \partial_t \nu_0 + a^2 \partial_x q &= -\nu_1. \end{aligned} \tag{3.52}$$

We can reformulate the second equation using both (3.51) and the chain rule

$$\nu = f(q) - \varepsilon \left(a^2 - f'(q)^2\right) \partial_x q. \tag{3.53}$$

This expression can be plugged into the first equation of (3.47) and we derive

$$\partial_t q + \partial_x f(q) = \varepsilon \partial_x \left(\left(a^2 - f'(q)^2\right) \partial_x q\right). \tag{3.54}$$

Clearly this equation is diffusive as long as the stability criterion

$$-a \leq f'(q) \leq a \tag{3.55}$$

is satisfied. This criterion is called *subcharacteristic condition* [JX95]. The Chapman-Enskog expansion shows that the relaxation system is a suitable approximation of the original conservation law. Therefore it is sufficient to determine the solution of the relaxation system. We do that by applying the following splitting approach. In a first step we solve the left-hand side of (3.47)

$$\begin{aligned} \partial_t q + \partial_x \nu &= 0, \\ \partial_t \nu + a^2 \partial_x q &= 0. \end{aligned} \tag{3.56}$$

The eigenvalues and eigenvectors of this system can be computed to be

$$\lambda_1^{\mathrm{JX}} = -a, \quad \lambda_2^{\mathrm{JX}} = a \tag{3.57}$$

and

$$\mathbf{r}^{(1),\mathrm{JX}} = \begin{pmatrix} -\frac{1}{a} \\ 1 \end{pmatrix}, \quad \mathbf{r}^{(2),\mathrm{JX}} = \begin{pmatrix} \frac{1}{a} \\ 1 \end{pmatrix}. \tag{3.58}$$

It is easy to check that both characteristic fields are linearly degenerate, which makes it is easy to find the solution to the associated Riemann problem. The Riemann invariants associated with the characteristic fields are given by

$$\begin{aligned} \lambda_1^{\mathrm{JX}}: &\quad I_1^{\mathrm{JX}} = \nu + aq, \\ \lambda_2^{\mathrm{JX}}: &\quad I_2^{\mathrm{JX}} = \nu - aq, \end{aligned} \tag{3.59}$$

and give rise to the following intermediate states in the exact Riemann solution of the homogeneous system (3.56)

$$q^* = \frac{1}{2}(q^L + q^R) - \frac{1}{2a}(\nu^R - \nu^L), \tag{3.60a}$$

$$\nu^* = \frac{1}{2}(\nu^L + \nu^R) + \frac{1}{2}a(q^R - q^L). \tag{3.60b}$$

In the second step, the projection step, we solve the system in the limit $\varepsilon \to 0$, i.e.

$$\begin{aligned} \partial_t q &= 0, \\ \partial_t \nu &= \frac{1}{\varepsilon}\left(f(q) - \nu\right). \end{aligned} \tag{3.61}$$

In practice, we use an instantaneous relaxation, meaning that we project on the relaxation equilibrium $\{(q,\nu); \nu = f(q)\}$ and simply take the solution of the first step for q and use this as the initial value when calculating the solution at the next time step. This projection step was first introduced in [Bre84]. Overall, the result is a Godunov-type scheme for the scalar conservation law (3.46) with the approximate Riemann solver

$$\mathcal{W}_{\mathcal{R}}\left(\frac{x}{t}; q^L, q^R\right) = \begin{cases} q^L, & \text{if } \frac{x}{t} < -a, \\ \frac{1}{2}(q^L + q^R) - \frac{1}{2a}(f(q^R) - f(q^L)), & \text{if } -a < \frac{x}{t} < a, \\ q^R, & \text{if } a < \frac{x}{t}. \end{cases} \tag{3.62}$$

Remark 3.4.1. *Clearly, for $a = \lambda^*_{\max}$ the Jin-Xin relaxation solver (3.62) coincides with the Rusanov solver (3.28) and for $S^L = -a$ and $S^R = a$ with the HLL solver (3.37).*

The Jin-Xin relaxation can be extended to systems of hyperbolic conservation laws. In this case, the relaxation system is defined by

$$\begin{aligned} \partial_t \mathcal{Q} + \partial_x V &= \mathbf{0}, \\ \partial_t V + \mathbf{A}\partial_x \mathcal{Q} &= \frac{1}{\varepsilon}(\mathcal{F}(\mathcal{Q}) - V), \end{aligned} \tag{3.63}$$

where $\mathbf{A}$ denotes a constant diagonal matrix with positive entries. The advantage of this approach is that a fixed recipe works for any system of hyperbolic conservation laws. However, the price is that the relaxation system consists of twice as many equations as the original system, which means a relatively high computational effort. As we will see in the following section, there are also smarter approaches with fewer additional relaxation equations.

3.4.2 Suliciu Relaxation Model

Instead of taking a "one-size-fits-all" approach like the Jin-Xin relaxation system, it is more efficient to adapt to the system of conservation laws at hand. To reduce the number of additional equations, it makes sense to approximate only nonlinear terms of the original system by relaxation while keeping linear equations. An example of this approach for the one-dimensional Euler equations is the so-called *Suliciu relaxation model* [Sul90, Sul92, CGP+01, Bou04]. The main idea of this approach is to relax only the nonlinear pressure term in the momentum equation. A new evolution equation for the pressure is derived from the continuity equation in (2.23)

$$\partial_t(\rho p) + \partial_x(\rho u p) + \rho^2 p'(\rho)\partial_x u = 0. \tag{3.64}$$

In this equation one replaces the pressure p by a relaxation variable π and the sound speed $\rho\sqrt{p'(\rho)}$ by a positive constant relaxation speed a

$$\partial_t(\rho\pi) + \partial_x(\rho u\pi + a^2 u) = \rho\frac{p-\pi}{\varepsilon}. \tag{3.65}$$

Finally, this new relaxation equation is added to the original Euler equations and the pressure p is replaced in all equations by π so that the resulting Suliciu relaxation system has the form

$$\begin{aligned} \partial_t \rho + \partial_x(\rho u) &= 0, \\ \partial_t(\rho u) + \partial_x(\rho u^2 + \pi) &= 0, \\ \partial_t E + \partial_x((E+\pi)u) &= 0, \\ \partial_t(\rho\pi) + \partial_x(\rho\pi u + a^2 u) &= \rho\frac{p-\pi}{\varepsilon}. \end{aligned} \tag{3.66}$$

A Chapman-Enskog expansion with similar steps as for the Jin-Xin relaxation leads to the following subcharacteristic stability condition

$$a \geq \rho c, \tag{3.67}$$

where c represents the sound speed. The eigenvalues of the homogeneous system $(3.66)_{\varepsilon=\infty}$ are given by

$$\lambda_1^{\text{Sul}} = u - \frac{a}{\rho}, \quad \lambda_2^{\text{Sul}} = u, \quad \lambda_3^{\text{Sul}} = u + \frac{a}{\rho}, \tag{3.68}$$

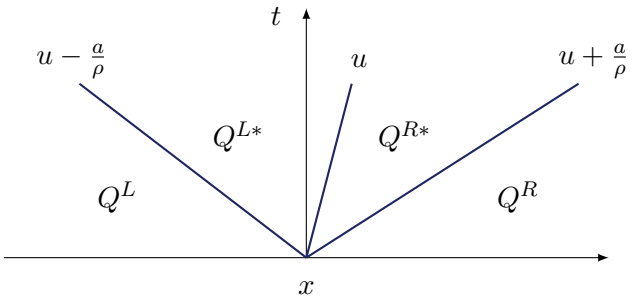

Figure 3.4: Structure of the solution to the Riemann problem associated with the Suliciu relaxation model (3.66).

where λ_2^{Sul} has multiplicity two. It can be checked that all eigenvalues are linearly degenerate, which allows us to exactly solve the Riemann problem associated with the relaxation system. As shown in Fig. 3.4, the solution has four constant states separated by three waves. The intermediate states can then be computed with the help of the Riemann invariants

$$\begin{aligned}
\lambda_1^{\text{Sul}}: \quad & I_1^{\text{Sul}} = u - \tfrac{a}{\rho}, \quad I_2^{\text{Sul}} = \pi + \tfrac{a^2}{\rho}, \quad I_3^{\text{Sul}} = e + \tfrac{\pi}{\rho} + \tfrac{a^2}{2\rho^2}, \\
\lambda_2^{\text{Sul}}: \quad & I_4^{\text{Sul}} = u, \qquad\quad I_5^{\text{Sul}} = \pi, \\
\lambda_3^{\text{Sul}}: \quad & I_6^{\text{Sul}} = u + \tfrac{a}{\rho}, \quad I_7^{\text{Sul}} = \pi + \tfrac{a^2}{\rho}, \quad I_8^{\text{Sul}} = e + \tfrac{\pi}{\rho} + \tfrac{a^2}{2\rho^2}.
\end{aligned} \tag{3.69}$$

We refrain from explicitly stating the definitions of the intermediate states here. Interested readers can find them for example in [BKZ20]. The approximate Riemann solver for the Euler equations thus has the structure

$$\mathcal{W}_{\mathcal{R}}\left(\frac{x}{t}; Q^L, Q^R\right) = \begin{cases} Q^L, & \text{if } \frac{x}{t} < \lambda_1^{\text{Sul}}, \\ Q^{L*}, & \text{if } \lambda_1^{\text{Sul}} < \frac{x}{t} < \lambda_2^{\text{Sul}}, \\ Q^{R*}, & \text{if } \lambda_2^{\text{Sul}} < \frac{x}{t} < \lambda_3^{\text{Sul}}, \\ Q^R, & \text{if } \lambda_3^{\text{Sul}} < \frac{x}{t}. \end{cases} \tag{3.70}$$

It can be proven that for a sufficiently large relaxation speed a and $Q^L, Q^R \in \Omega_{\text{phys}}$, the intermediate states Q^{L*} and Q^{R*} also lie in Ω_{phys} and that the solver satisfies a discrete entropy inequality of the form (3.13) [Bou04, GR02].

Remark 3.4.2. *The structure of the Suliciu Riemann solver strongly resembles that of the HLLC solver. In fact, it can be shown that with a certain choice of wave speeds, both solvers are equivalent [Bou04].*

Based on the Suliciu relaxation system, many other relaxation systems of a similar type have been constructed (see e.g. [CGP+01, BdL09, CGS12, CC14, BL16, BKZ20]). Particularly relevant in the context of this thesis is an extension to the isentropic Euler equations in [BCG20] and to the full Euler equations with gravitational source terms in [DZBK16], whose concepts are discussed in more detail in Chapter 4. Relaxation systems and corresponding solvers for the ideal MHD equations can be found in [BKW07, BKW10, WFK11]. Key idea of these systems is the definition and relaxation of newly defined pressure variables

$$\pi = p + \frac{1}{2}|\mathbf{B}|^2 - B_x^2 \quad \text{and} \quad \pi_\perp = -(B_x B_y, B_x B_z)^\top \tag{3.71}$$

instead of the acoustic pressure.

3.5 Source Terms

Up to this point we have concentrated on the discretization of homogeneous conservation laws. In the following, we also include source terms. Let us consider the one-dimensional system of balance laws

$$\mathcal{Q}_t + \mathcal{F}(\mathcal{Q})_x = \mathcal{S}(\mathcal{Q}). \tag{3.72}$$

In this work we employ two general approaches to discretize this balance law. In the first one, we apply a Godunov-type method to the left-hand side of (3.72) so that the approximate Riemann solver solves the homogeneous Riemann problem. The source term is discretized separately. This results in an unsplit method of the form

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^n - F_{i-1/2}^n\right) + \Delta t S_i^n, \tag{3.73}$$

where $S_i^n = \mathcal{S}(Q_i^n)$ is a suitable choice of a second order accurate discretization. This approach allows us to use standard approximate Riemann solvers from Sect. 3.3-3.4, which are developed for the homogeneous Riemann problem.

For the second approach we assume that the source term can be written in the form

$$\mathcal{S}(\mathcal{Q}(x,t)) = \mathrm{s}(x,t)\mathcal{Z}(x)_x, \tag{3.74}$$

with $\mathrm{s} : \mathbb{R}\times\mathbb{R}_{\geq 0} \to \mathbb{R}^m$ and $\mathcal{Z} : \mathbb{R} \to \mathbb{R}$, so that the balance law (3.72) can be reformulated to

$$\begin{aligned} &\mathcal{Q}_t + \mathcal{F}(\mathcal{Q})_x = \mathrm{s}(\mathcal{Q})\mathcal{Z}_x, \\ &\mathcal{Z}_t = 0. \end{aligned} \tag{3.75}$$

This ansatz is possible for Euler and ideal MHD equations in the case of a time-independent gravity term. In this case $\mathcal{Z}$ corresponds to the gravitational potential Φ. Just like the conservative variables $\mathcal{Q}$, the source $\mathcal{Z}$ is approximated by piecewise constant cell averages

$$Z_i \approx \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{Z}(x)\mathrm{d}x. \tag{3.76}$$

We now aim to design an approximate Riemann solver for the inhomogeneous Riemann problem

$$\begin{aligned} &\mathcal{Q}_t + \mathcal{F}(\mathcal{Q})_x = \mathrm{s}(\mathcal{Q})\mathcal{Z}_x, \\ &\mathcal{Q}_0(x) = \begin{cases} Q_i, & \text{if } x < x_{i+1/2}, \\ Q_{i+1}, & \text{if } x > x_{i+1/2}, \end{cases} \\ &\mathcal{Z}_0(x) = \begin{cases} Z_i, & \text{if } x < x_{i+1/2}, \\ Z_{i+1}, & \text{if } x > x_{i+1/2}. \end{cases} \end{aligned} \tag{3.77}$$

As a consequence, the resulting solver depends on Q and Z and we denote it by

$$\mathcal{W}_{\mathcal{R}}(x/t; Q^L, Z^L, Q^R, Z^R). \tag{3.78}$$

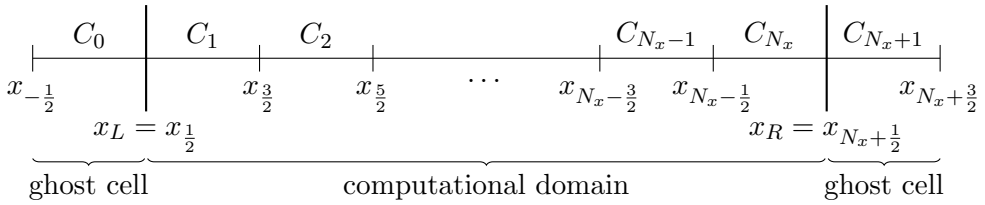

Figure 3.5: One-dimensional computational domain extended by ghost cells.

The cell averages at the new time level t^{n+1} can be computed analogously to the homogeneous case by

$$\begin{aligned} Q_i^{n+1} =& \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \widetilde{\mathcal{W}}_\Delta(x,t^{n+1})\mathrm{d}x \\ =& \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_i} \mathcal{W}_\mathcal{R}\left(\frac{x-x_{i-1/2}}{\Delta t}; Q_{i-1}, Z_{i-1}, Q_i, Z_i\right)\mathrm{d}x \\ &+ \frac{1}{\Delta x}\int_{x_i}^{x_{i+1/2}} \mathcal{W}_\mathcal{R}\left(\frac{x-x_{i+1/2}}{\Delta t}; Q_i, Z_i, Q_{i+1}, Z_{i+1}\right)\mathrm{d}x. \end{aligned} \tag{3.79}$$

3.6 Boundary Conditions

In the previous sections we assume to know the cell averages of cell C_i and its neighboring cells at time t^n when calculating Q_i^{n+1}. In practical applications, however, we operate on finite spatial domains, which is why it becomes necessary to describe the solution at the boundaries. This is done by *boundary conditions*. There are different ways to set the boundary conditions, depending on physically reasonable assumptions or our knowledge about the solution at the boundaries.
In numerical codes, boundary conditions are often implemented with the help of *ghost cells*. In this case, the grid is extended beyond the actual domain by additional cells in which the solution is specified by the boundary conditions. For the one-dimensional case this grid structure is shown in Fig. 3.5. In the following we briefly describe for the one-dimensional case four relevant types of boundary conditions that are used in this work.

- **Periodic boundary conditions:** The simplest choice are periodic boundary conditions. In this case it is assumed that the domain at the right boundary connects to the left boundary. This is implemented numerically by the following definition of the ghost cells
$$Q_0^n = Q_{N_x}^n, \quad Q_{N_x+1}^n = Q_1^n. \tag{3.80}$$

- **Exact boundary conditions:** In this case the solution at the boundary is known, so that we can define the cell averages in the ghost cells by
$$\begin{aligned} Q_0^n = Q_0^{n,exact} &\approx \int_{x_{-1/2}}^{x_{1/2}} \mathcal{Q}(x,t^n)\mathrm{d}x, \\ Q_{N_x+1}^n = Q_{N_x+1}^{n,exact} &\approx \int_{x_{N_x+1/2}}^{x_{N_x+3/2}} \mathcal{Q}(x,t^n)\mathrm{d}x. \end{aligned} \tag{3.81}$$

- **Transmissive conditions:** Especially for small computational domains it may be useful to define boundary conditions that allow the passage of waves without any

effect on them. Such transmissive boundary conditions can be defined by

$$Q_0^n = Q_1^n, \quad Q_{N_x+1}^n = Q_{N_x}^n. \tag{3.82}$$

- **Reflective boundary conditions:** Reflective boundary conditions model a fixed, reflective impermeable wall. They are defined as transmissive boundary conditions with the difference that the sign is switched in the normal velocity component:

$$u_0^n = -u_1^n, \quad u_{N_x+1}^n = -u_{N_x}^n. \tag{3.83}$$

For higher order finite volume methods, more ghost cells are needed. They can be defined in the same way by setting boundary conditions.

3.7 Extension to Multiple Space Dimensions

In Sect. 3.1-3.5 we have constructed a numerical method for one-dimensional systems of conservation and balance laws. This method will now be extended to several spatial dimensions. There are various approaches in the literature, e.g. dimensional splitting [Tor09, LeV02] or multi-dimensional Riemann solvers [Bal10, Bal12, VLW04]. For the numerical methods described in this work, we will rely on so-called *unsplit finite volume methods* [Tor09]. Let us consider the two-dimensional conservation law

$$\mathcal{Q}_t + \mathcal{F}_1(\mathcal{Q})_x + \mathcal{F}_2(\mathcal{Q})_y = 0. \tag{3.84}$$

In this case, the domain is discretized by rectangular cells $C_{i,j} = (x_{i-1/2}, x_{i+1/2}) \times (y_{j-1/2}, y_{j+1/2})$ with uniform space steps Δx and Δy. Thus, the cell averages are defined by

$$Q_{i,j}^n = \frac{1}{\Delta x \Delta y} \int_{y_{j-1/2}}^{y_{j+1/2}} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{Q}(x, y, t^n) \mathrm{d}x \mathrm{d}y. \tag{3.85}$$

In order to evolve the solution to the next time step, we do sweeps in x- and y-directions and solve at each cell edge one-dimensional Riemann problems of the form

$$x\text{-direction:} \quad \begin{cases} \mathcal{Q}_t + \mathcal{F}_1(\mathcal{Q})_x = 0, \\ \mathcal{Q}_0(x, y) = \begin{cases} Q_{i,j}^n, & \text{if } x < x_{i+1/2}, \\ Q_{i+1,j}^n, & \text{if } x > x_{i+1/2}, \end{cases} \end{cases} \tag{3.86}$$

$$y\text{-direction:} \quad \begin{cases} \mathcal{Q}_t + \mathcal{F}_2(\mathcal{Q})_y = 0, \\ \mathcal{Q}_0(x, y) = \begin{cases} Q_{i,j}^n, & \text{if } y < y_{j+1/2}, \\ Q_{i,j+1}^n, & \text{if } y > y_{j+1/2}. \end{cases} \end{cases} \tag{3.87}$$

Hence, one-dimensional (approximate) Riemann solvers can still be applied, which results in Riemann solutions $\mathcal{W}_{i+1/2,j}(x/t)$ and $\mathcal{W}_{i,j+1/2}(y/t)$. The cell averages can then be evolved by

$$Q_{i,j}^{n+1} = Q_{i,j}^n - \frac{\Delta t}{\Delta x} \left(F_{1,i+1/2,j}^n - F_{1,i-1/2,j}^n \right) - \frac{\Delta t}{\Delta y} \left(F_{2,i,j+1/2}^n - F_{2,i,j-1/2}^n \right) \tag{3.88}$$

with the fluxes defined by

$$F_{1,i+1/2,j}^n = \mathcal{F}_1 \left(\mathcal{W}_{i+1/2,j}(0; Q_{i,j}^n, Q_{i+1,j}^n) \right), \tag{3.89a}$$

$$F_{2,i,j+1/2}^n = \mathcal{F}_2 \left(\mathcal{W}_{i,j+1/2}(0; Q_{i,j}^n, Q_{i,j+1}^n) \right). \tag{3.89b}$$

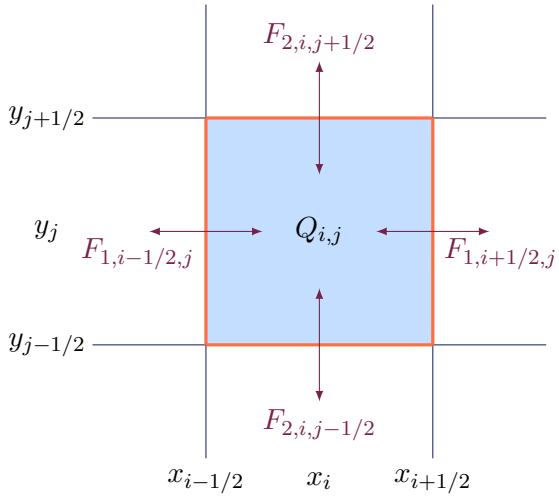

Figure 3.6: In order to evolve the cell average $Q_{i,j}$ on a two-dimensional Cartesian grid, four separate Riemann problems (one at each side of cell $C_{i,j}$) are solved to determine the numerical flux functions.

Fig. 3.6 illustrates the four Riemann problems that need to be solved for the update of the cell average $Q_{i,j}$ and the in- and outflows through the fluxes between cell $C_{i,j}$ and its neighboring cells. Since the method now also solves Riemann problems in y-direction, it must be ensured that the waves in this direction also can only travel one cell width within a time step. Consequently, the CFL condition for the method (3.88) must include the space step Δy and is therefore

$$\Delta t \leq C_{\text{CFL}} \min\left(\frac{\Delta x}{\lambda_{\max,x}}, \frac{\Delta y}{\lambda_{\max,y}}\right) \quad \text{with } C_{\text{CFL}} \leq 1. \tag{3.90}$$

Here, $\lambda_{\max,x}$ denotes the maximum absolute eigenvalue of the Jacobian $\mathcal{A}_1 = \partial\mathcal{F}_1/\partial\mathcal{Q}$ and $\lambda_{\max,y}$ the maximum absolute eigenvalue of $\mathcal{A}_2 = \partial\mathcal{F}_2/\partial\mathcal{Q}$.

The extension of unsplit finite volume methods to a third spatial dimension works according to the same principles and is therefore straightforward.

3.8 Extension to Second Order in Space

The Godunov and Godunov-type methods described so far are first order accurate. Besides the first order temporal discretization, this is due to the fact that the function $\widetilde{\mathcal{W}}(x, t^n)$ in the reconstruction step of the REA algorithm is only piecewise constant. In order to increase the order of the spatial discretization, we rely on a *piecewise linear reconstruction* [vL77]

$$\widetilde{\mathcal{W}}(x, t^n) = Q_i^n + \sigma_i^n (x - x_i) \quad \text{for } x_{i-1/2} \leq x \leq x_{i+1/2}. \tag{3.91}$$

To obtain the initial values for the Riemann problem at each cell interface, the function (3.91) is evaluated in each cell C_i at its boundaries $x_{i-1/2}$ and $x_{i+1/2}$. This clearly gives more accurate initial data for the Riemann problems in the Godunov-type method. It is important to notice that the scheme remains conservative as the linear function satisfies

$$\frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \widetilde{\mathcal{W}}(x, t^n) \mathrm{d}x = Q_i^n. \tag{3.92}$$

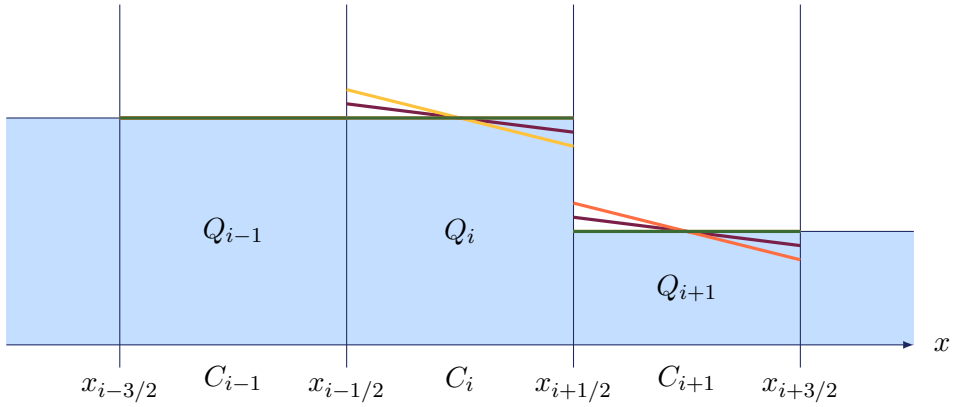

Figure 3.7: Linear reconstruction polynomials derived with different slope definitions near a discontinuity at $x_{i+1/2}$. Maroon: central. Orange: upwind. Yellow: downwind. Green: minmod. All choices except the minmod slope lead to spurious oscillations near the discontinuity.

The slope σ_i^n depends on the neighboring cells and is computed componentwise for each entry in Q_i^n. Obvious choices to compute σ_i^n are

$$\begin{aligned}
\sigma_i^n &= \frac{Q_{i+1}^n - Q_{i-1}^n}{2\Delta x} && \text{(central)},\\
\sigma_i^n &= \frac{Q_i^n - Q_{i-1}^n}{\Delta x} && \text{(upwind)},\\
\sigma_i^n &= \frac{Q_{i+1}^n - Q_i^n}{\Delta x} && \text{(downwind)}.
\end{aligned}$$

Unfortunately, a linear reconstruction that relies on one of the above slopes produces overshoots near discontinuities, as illustrated in Fig. 3.7. These spurious oscillations can lead to unphysical solutions, such as negative values for density or pressure. Numerically, this can result in a premature termination of the scheme, because the root of negative values cannot be computed. We can prevent the phenomenon by using *slope limiters* that recognize extrema on the cell average level and reduce the reconstruction procedure to a first order approximation in such cells, i.e. using a constant reconstruction with $\sigma = 0$. In the literature, one can find different limiters such as the *van Leer* [vL77] or *Superbee* [Roe85] limiter. In this work we choose to use the *minmod limiter* [Roe86] so that the slope is computed by

$$\sigma_i^n = \text{minmod}\left(\frac{Q_{i+1}^n - Q_i^n}{\Delta x}, \frac{Q_i^n - Q_{i-1}^n}{\Delta x}\right). \tag{3.93}$$

The minmod function therein is defined by

$$\text{minmod}(a_1, \ldots, a_n) = \begin{cases} \text{sgn}(a_1) \min_i |a_i|, & \text{if } \text{sgn}(a_1) = \ldots = \text{sgn}(a_n),\\ 0, & \text{else.} \end{cases} \tag{3.94}$$

Thus, the minmod limiter compares the upwind and downwind slope. If the sign is the same, the slope with the smaller magnitude is selected; if the signs are different, the cell average Q_i^n is a local extremum and the slope σ_i^n is set to zero.

In this work we restrict ourselves to second order linear reconstructions. Readers who are interested in higher order methods are referred for example to a *piecewise parabolic reconstruction* [CW84] or *essentially non-oscillatory* (ENO) [Har89, HEOC87, HOEC86], *weighted ENO* (WENO) [JS96, LOC94, Shu03] and *central WENO* (CWENO) [LPR99, LPR00] methods.

3.9 Time Integration Methods

Up to this point, we have considered fully discrete methods, in which we discretized space and time simultaneously. In practice, however, we rely on the *method of lines* approach, in which initially only the spatial discretization is done, leading to a semi-discrete formulation

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_i = -\frac{1}{\Delta x}\left(F_{i+1/2}(Q(t)) - F_{i-1/2}(Q(t))\right). \tag{3.95}$$

This approach thus reduces the PDE to a system of *ordinary differential equations* (ODEs) of general form

$$\frac{\mathrm{d}}{\mathrm{d}t}Q(t) = \mathcal{H}(Q(t)), \tag{3.96}$$

where $\mathcal{H}$ denotes the spatial residual. Now it is possible to apply standard ODE solvers, which makes it easy to achieve higher order discretizations in time. A widely used class of ODE solvers are *Runge-Kutta* (RK) methods [But63, But64, JST81]. If the right-hand side does not explicitly depend on time, a RK method with S stages has the general form

$$\mathbf{k}_s = \mathcal{H}\left(Q^n + \Delta t \sum_{l=1}^{S} a_{sl}\mathbf{k}_l\right), \quad s = 1, ..., S, \tag{3.97}$$

$$Q^{n+1} = Q^n + \Delta t \sum_{s=1}^{S} b_s \mathbf{k}_s. \tag{3.98}$$

The coefficients a_{sl} and b_s define the specific RK method and are typically denoted in Butcher tableaux of the form

$$\begin{array}{c|c} \mathbf{c} & \mathbf{A} \\ \hline & \mathbf{b}^\top \end{array} \tag{3.99}$$

An RK method is *explicit* if $a_{sl} = 0$ whenever $l \geq s$, because then every $\mathbf{k}_s$ can be computed from the previously computed k_l with $l < s$. Explicit RK methods are easy to implement and a single update to the next time level is computationally cheap. The time step size for explicit RK methods depends on a CFL restriction of the form (3.12). It should be noted that the CFL number C_{CFL} required for stability is different for different RK methods [JST81, LT98]. In the following we give two examples of Butcher tableaux for well-known explicit RK methods.

Example 3.9.1. *The Butcher tableau for the first order forward Euler method is given by*

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array} \tag{3.100}$$

and its CFL coefficient for stability is $C_{\mathrm{CFL}} = \frac{1}{2}$.

Example 3.9.2. *The butcher tableau for the third order strong-stability-preserving RK (SSP-RK) method [SO88] is given by*

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1/2 & 1/4 & 1/4 & 0 \\ \hline & 1/6 & 1/6 & 2/3 \end{array} \tag{3.101}$$

Strong-stability-preserving methods are a subclass of RK methods that maintain the total variation diminishing (TVD) property $TV(Q^{n+1}) \leq TV(Q^n)$ of the first order forward Euler method while achieving higher order accuracy in time. The CFL number of SSP-RK3 is $C_{\text{CFL}} = 1$.

RK methods with nonzero entries on and above the diagonal depend on unknown $\mathbf{k}_s$ and are *implicit* therefore. They require to solve a nonlinear system of equations at each time step, which makes a single update to the next time level computationally costly. The advantage, on the other hand, is that the time step can be chosen unconditionally large. The choice then only needs to consider the desired accuracy of the method.

Example 3.9.3. *The Butcher tableau for the first order backward Euler method is given by*

$$\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array} \tag{3.102}$$

Example 3.9.4. *The Butcher tableau of the second order accurate explicit singly diagonal implicit Runge-Kutta (ESDIRK2) method [HS96] is given by*

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & d & d & 0 \\ 1 & w & w & d \\ \hline & w & w & d \\ \hline & \frac{1-w}{3} & \frac{3w+1}{3} & \frac{d}{3} \end{array} \tag{3.103}$$

with $\gamma = 2 - \sqrt{2}$, $w = \sqrt{2}/4$ and $d = \gamma/2$. The coefficients in the last row belong to an embedded low order step, which is used to produce an estimate of the local truncation error of a single RK step and thereby to control the error. For more details on embedded RK methods, the reader is referred to [JKT18].

A third alternative next to explicit and implicit are IMEX-RK methods. They are used when a PDE is partially explicitly and partially implicitly discretized. In this case, the semi-discrete form of the ODEs has the general form

$$\frac{\mathrm{d}}{\mathrm{d}t} Q(t) = \mathcal{H}(Q_E(t), Q_I(t)), \tag{3.104}$$

where the first argument of $\mathcal{H}$ is discretized explicitly, whereas the second argument is discretized implicitly. At the beginning of each time step, the method starts with $Q_E^n = Q_I^n = Q^n$. Then the stage fluxes $\mathbf{k}_s$ for each stage s are computed in the following

way:

$$Q_E^{(s)} = Q_E^n + \Delta t \sum_{l=1}^{s-1} \hat{a}_{sl} \mathbf{k}_l, \qquad 2 \leq s \leq S, \tag{3.105a}$$

$$Q_I^{(s)} = Q_E^n + \Delta t \sum_{l=1}^{s-1} a_{sl} \mathbf{k}_l, \qquad 2 \leq s \leq S, \tag{3.105b}$$

$$\mathbf{k}_s = \mathcal{H}\left(Q_E^{(s)}, Q_I^{(s)} + \Delta t a_{ss} \mathbf{k}_s\right), \qquad 1 \leq s \leq S. \tag{3.105c}$$

Using these stage fluxes, the updated solution at the new time level is computed by

$$Q^{n+1} = Q^n + \Delta t \sum_{s=1}^{S} b_s \mathbf{k}_s. \tag{3.106}$$

The coefficients for IMEX-RK methods are given by a double Butcher tableau of the form

$$\begin{array}{c|c} \hat{\mathbf{c}} & \hat{\mathbf{A}} \\ \hline & \hat{\mathbf{b}}^\top \end{array} \qquad \begin{array}{c|c} \mathbf{c} & \mathbf{A} \\ \hline & \mathbf{b}^\top \end{array} \tag{3.107}$$

Example 3.9.5. *One example of IMEX-RK methods is the L-stable Second-Order Diagonally Implicit Runge Kutta Method (LSDIRK2)(2,2,2) [PR05]. The triplet (2,2,2) refers to the number of stages of the implicit part, the number of stages of the explicit part, and the order of the IMEX scheme, respectively. The double Butcher tableau for LSDIRK2 is given by*

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \beta & \beta & 0 \\ \hline & 1-\gamma & \gamma \end{array} \qquad \begin{array}{c|cc} \gamma & \gamma & 0 \\ 1 & 1-\gamma & \gamma \\ \hline & 1-\gamma & \gamma \end{array} \tag{3.108}$$

with $\gamma = 1 - 1/\sqrt{2}$ and $\beta = 1/(2\gamma)$. The left tableau denotes the coefficients for the explicit part, whereas the right tableau shows the ones for the implicit part.

3.10 Numerical Challenges

The previous sections provide the tools to implement a working method whose numerical solution converges to weak solutions of the compressible Euler or ideal MHD equations. This raises the question of whether it is necessary to design new numerical methods at all. The main reason for the construction of new methods is the varying quality and efficiency of standard methods in different applications. In order to reach an accurate result, a standard method may need to be run on a highly resolved numerical grid, which often is impossible in practice. Given the currently available computing power even of supercomputers, the construction of efficient methods that are adapted to the underlying problem remains essential. Below we consider three numerical challenges we are confronted with in the astrophysical context and for which standard FV methods are not efficient.

3.10.1 Low Mach Numbers

The maximum eigenvalues of the rescaled Euler and MHD equations in (2.53) and (2.81) both scale with $\mathcal{O}(1/\mathcal{M})$. In the case of low Mach numbers, the maximum wave speeds thus become very large. This implies two problems for FV methods:

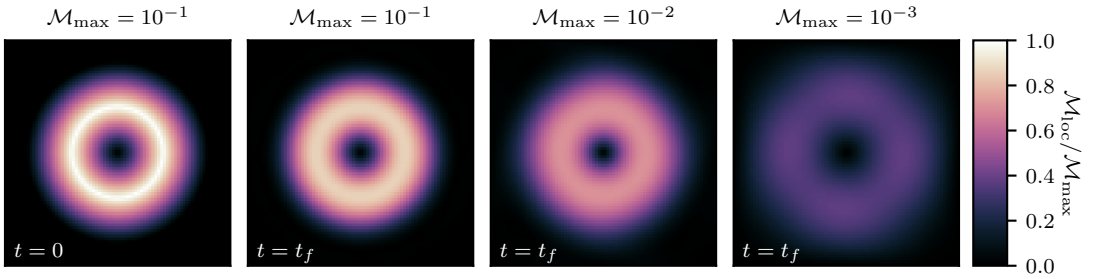

Figure 3.8: Distribution of the local Mach number (normalized by the maximum initial value) of the Gresho vortex. The plot on the left shows the initial vortex for the maximum Mach number $\mathcal{M}_{\max} = 10^{-1}$, the three other the results for the explicit Rusanov FV method after one turnover for different $\mathcal{M}_{\max}$. All results are obtained on a 80×80 grid.

1. Many Riemann solvers used nowadays are designed to work in supersonic regimes. In order to achieve numerical stability, such solvers need to add upwind numerical diffusion terms to the physical fluxes, which smear out any discontinuity present in the flow on a time scale comparable to the cell crossing time of the shock. Thus, the diffusion term scales with the largest wave speed of the underlying PDE and consequently introduces excessive dissipation in the low Mach number regime. This behavior can be analyzed particularly easily in the structure of the Rusanov flux (3.31), but also applies to others such as Roe's solver or HLL-type solvers.

2. The CFL condition in fully time-explicit schemes restricts the time step to the crossing time of the fastest wave resulting from the underlying PDE over a grid cell. Since the fastest wave speed scales with $\mathcal{O}(1/\mathcal{M})$ for both the Euler and MHD equations, small Mach numbers force time-explicit schemes to choose a small time step in order to remain stable. As a consequence, it becomes computationally expensive to resolve slow fluid motions and Alfvén waves.

Example 3.10.1. *The numerical problems induced by low Mach numbers can be illustrated by a simulation of the Gresho vortex, which describes a stationary vortex in which centrifugal forces and pressure gradients are in perfect balance [GC90, MRE15]. The vortex can be set up with different maximum local Mach numbers* $\mathcal{M}_{\max} = \max(\mathcal{M}_{\mathrm{loc},t=0})$. *We solve the IVP with a standard time-explicit FV method that uses the Rusanov solver within the Godunov-type method and an explicit SSP-RK2 method for time integration. The results after one turnover in Fig. 3.8 show that the dissipation increases significantly with decreasing Mach numbers so that the vortex cannot be resolved accurately anymore. An investigation of the evolution of kinetic energy in Fig. 3.9 highlights this behavior. Although the total kinetic energy should be conserved in the incompressible limit (see (2.65)), it decreases more and more for smaller Mach numbers due to the increasing dissipation. Even if the Rusanov solver is known to be very dissipative in general, the results for Roe's and HLL-type solvers are comparable. An evaluation of the wall-clock time of the individual simulations presented in Fig. 3.10 shows that the duration of the simulations scale with the Mach number. This is caused by the acoustic wave speeds in the CFL condition, which become very fast for low Mach numbers.*

These numerical problems as a consequence of low Mach numbers lead to interest in *asymptotic-preserving (AP)* methods.

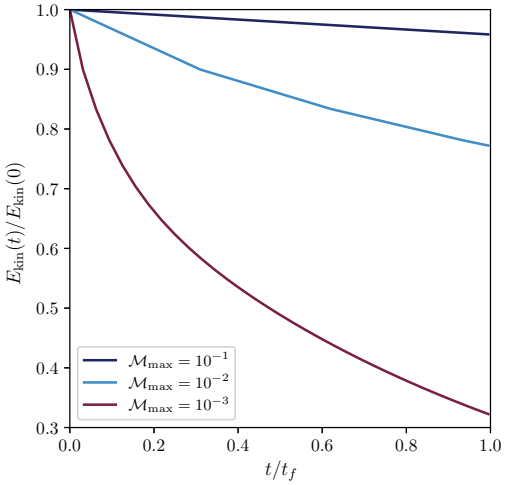

Figure 3.9: Time evolution of the total kinetic energy for different maximum Mach numbers $\mathcal{M}_{\text{max}}$.

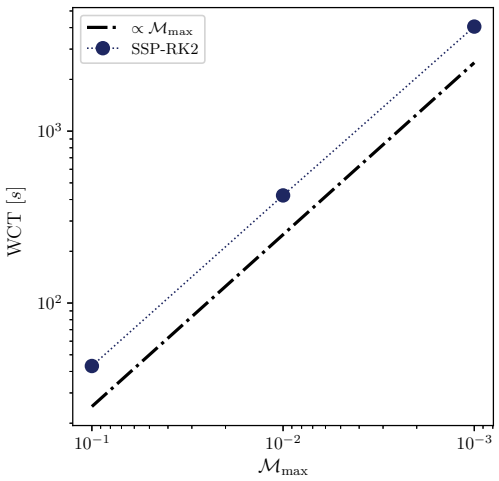

Figure 3.10: Wall-clock time in $[s]$ of the fully time-explicit FV scheme using a SSP-RK2 method for time integration.

Definition 3.10.2. *Let us denote the solutions of the compressible PDE by $\mathcal{Q}^{\mathcal{M}}$. This solution depends on the Mach number $\mathcal{M}$. In the low Mach number limit the solutions tend to the solutions of the incompressible equations denoted by $\mathcal{Q}^0$. A numerical scheme is called **asymptotic-preserving** if its solutions $Q_\Delta^{\mathcal{M}}$ tend to a consistent approximation of $\mathcal{Q}^0$.*

Various approaches have been suggested in the literature to derive low Mach compliant numerical methods. For several approximate Riemann solvers there are low Mach fixes that rescale the dissipation term so that it no longer scales with the inverse of the Mach number [Tur87, Rie11, XC13, MRE15, Lio06, MM21]. These fixes are not derived from a fundamental basis, but artificially incorporated into the one-dimensional solvers, which potentially reduces the method's stability. An alternative strategy is to keep the original one-dimensional method and extend it to multiple dimensions in a particular all-speed way, leading to more stable numerical methods [Bar21].
These approaches can be combined with an implicit time integration as described in Sect. 3.9. Then the time step is not limited by stability conditions but only the desired accuracy. In order to resolve fluid motions and Alfvén waves, it makes sense to use the Alfvén wave speed in the CFL condition. If the Mach number is small enough, the possibility of larger chosen time steps outweighs the disadvantage of higher computational costs for a single time step by using the implicit solver. Combinations of low Mach Riemann solvers and implicit time integration are used for example in [MRE15, BEK+17, HHE+21].

A second approach is given by IMEX methods. These methods rely on a splitting of the underlying system of PDEs into two parts: one containing the slow dynamics and the other containing the stiff acoustic terms. The first one is discretized explicitly in time with a Godunov-type method, whereas the second one is discretized implicitly in time. As a result, the dissipation term in the numerical flux and the CFL condition only refer to the explicit part whose eigenvalues are independent of the Mach number. The IMEX approach has been applied to the homogeneous Euler equations [Kle95, CDK12, DLMDV18, TZPK20, BQRX19, BDL+20], Euler equations with gravity [BLMY17, TPK20] and the

homogeneous MHD equations [LL91, ALJ99, DBTF19, Fam21, CWX23].

AP methods help to resolve low Mach number flows accurately, but they might also exhibit unphysical *checkerboard modes* in their solution [Rie11, NBA+14]. Checkerboard modes describe an odd-even decoupling of the spatial approximation, purely caused by the discretization of the scheme [FP02, LMW02, Del09, Del10]. The discretization of the method thus leads to a unphysical and yet stable solution. Only some of the low Mach fixes suppress the occurrence of checkerboard modes [Rie11, CYX18].

3.10.2 Small Perturbations of Equilibria

In many applications the fluid flows of interest are close to a (magneto-)hydrostatic equilibrium so that they represent only a very small perturbation of the equilibrium. On coarse grids, the magnitude of these perturbations might be smaller than the truncation error of the numerical method, which means that the method cannot resolve the perturbation. The following example gives a simple illustration of this phenomenon.

Example 3.10.3. *Let us consider an isothermal equilibrium for the one-dimensional Euler equations [CK15]. The equilibrium is given by*

$$(\rho, u, p)(x, 0) = (\exp(-\Phi(x)),\ 0,\ \exp(-\Phi(x))) \quad \text{and} \quad \Phi(x) = \frac{1}{2}x^2, \tag{3.109}$$

which clearly satisfies (2.45). We then add a perturbation in the pressure so that the new pressure is defined by

$$p(x, 0) = \exp(-\Phi(x)) + \eta \exp(-100(x - 0.5)^2), \tag{3.110}$$

and control the magnitude of the perturbation by the parameter η. The problem is solved with a standard FV method of the form (3.73) using the Rusanov flux (3.31) first on a coarse grid with $N_x = 100$ and then on a finer grid with $N_x = 2\,000$ cells. Fig. 3.11 presents the initial perturbation Δp_0 and the perturbation Δp at final time $t_f = 0.2$ for a large parameter $\eta = 10^{-1}$ and a small parameter $\eta = 10^{-5}$. The large perturbation is well-resolved on both grids. In case of the small perturbation, the method is only able to resolve the perturbation on the fine grid.

We are interested in efficient numerical methods that can resolve small perturbations already on coarse grids. This can be achieved by *well-balanced* methods.

Definition 3.10.4. *A numerical method is called* ***well-balanced*** *if it exactly preserves the discretization of (certain) equilibria of the form (2.22), e.g. hydrostatic equilibria (2.45) or magnetohydrostatic equilibria (2.79).*

Standard FV methods are typically not well-balanced as the flux and the source term are discretized separately and therefore not coordinated in such a way that they exactly preserve steady states at rest. The literature provides a number of different approaches to construct well-balanced FV methods such as equilibrium preserving reconstructions [KM14, KM16, VC19], path conservative methods [CGLGP08, Par06], global fluxes [CCK+18], special source term discretizations based on given equilibrium states [BCK18, BCKR19], approximate Riemann solvers for the inhomogeneous Riemann problem [DZBK16, TPK20] or methods designed to solve equations for the deviation of the solution from an a priori known equilibrium [BLMY17, BCK21]. A classification and description of different well-balancing approaches can be found in [Ber20].

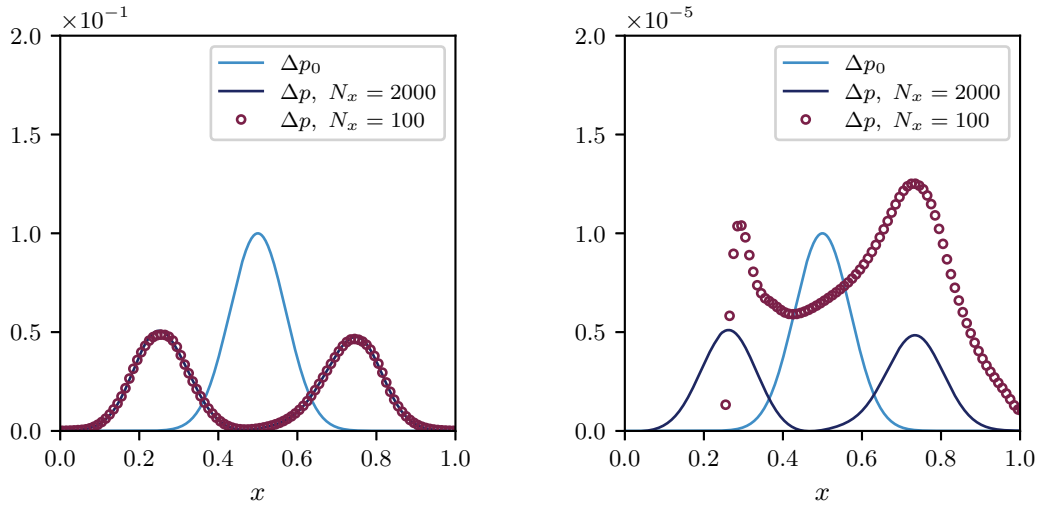

Figure 3.11: Pressure perturbation. Left: $\eta = 10^{-1}$. Right: $\eta = 10^{-5}$.

3.10.3 Solenoidal Constraint

The solenoidal constraint described in Sect. 2.4.1 is not automatically satisfied if the induction equation is solved with standard Godunov-type schemes. As a result, magnetic monopoles are created locally at each time step and tend to accumulate as they cannot be transported away by any of the MHD waves. The following example gives an impression of this behavior.

Example 3.10.5. *We consider the Orszag-Tang vortex which is a two-dimensional MHD test problem. The test is described in more detail in Sect. 6.3.2. Fig. 3.12 shows the time evolution of the maximum and mean relative central discrete divergence*

$$(\nabla \cdot \mathbf{B})_{i,j} = \frac{B_{x,i+1,j} - B_{x,i-1,j}}{2\Delta x} + \frac{B_{y,i,j+1} - B_{y,i,j-1}}{2\Delta y} \tag{3.111}$$

for a standard Godunov-type scheme using the Rusanov flux. Clearly, the discrete divergence increases significantly over time, which poses a unphysical phenomenon on the discrete level.

If not properly treated, these artifacts can accelerate the flow along field lines, generate wrong field topologies, and ultimately lead to severe stability problems [BB80]. Therefore, special care needs to be taken to design accurate and stable Godunov-type methods for solving the MHD equations.
Different strategies have been presented in the literature to handle the divergence constraint. Among these, discretizing the *8-wave formulation* (2.72) relies on including additional source terms that are proportional to $\nabla \cdot \mathbf{B}$, which implies an advection equation for the divergence (see (2.74)). Therefore, magnetic monopoles are advected with the flow and do not accumulate over time [Pow97, PRL+99].
A second approach can be seen in *divergence cleaning schemes* [DKK+02] in which the divergence constraint is coupled to the MHD system using a generalized Lagrangian multiplier ψ. This allows to transport numerical monopoles with the maximum available speed on the grid and reduce divergence errors at the same time. One downside of both the 8-wave formulation and divergence cleaning is that they are not conservative and cannot

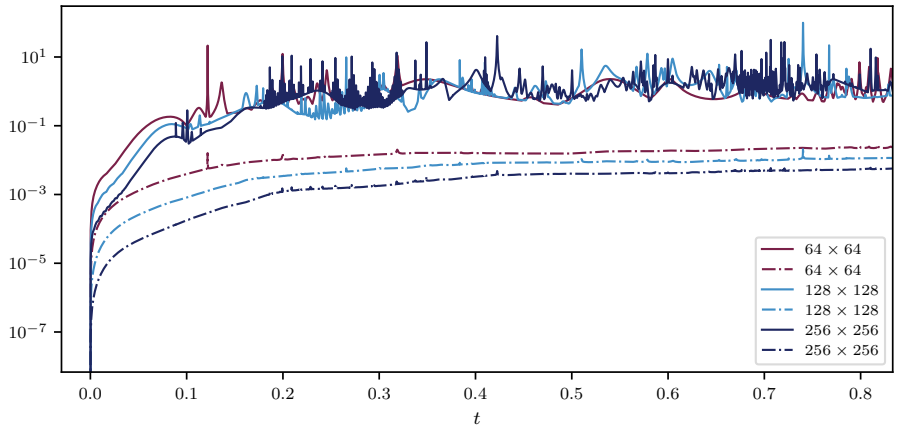

Figure 3.12: Time evolution of the maximum relative divergence $\max[(\nabla\cdot\mathbf{B}\Delta x)/|\mathbf{B}|]$ (solid line) and mean relative divergence $\langle(\nabla\cdot\mathbf{B}\Delta x)/|\mathbf{B}|\rangle$ (dot-dashed) in the simulations of the Orszag-Tang vortex with a standard Rusanov FV method for different grid resolutions.

enforce any discretization of $\nabla\cdot\mathbf{B}$ to zero. Furthermore, these methods are most effective when open boundaries are used, so that the magnetic monopoles can leave the domain. A third strategy, first introduced by Evans and Hawley [EH88], are *constrained transport* (CT) methods based on a staggered formulation. These methods exploit Stokes' theorem to reformulate the induction equations, which results in equations for the face-averaged magnetic field. Thereby, staggered CT methods maintain one particular discretization of $\nabla\cdot\mathbf{B}$ to zero up to machine accuracy [DW98, BS99, Tót00, LZ04, GS08, MZ21]. Although numerical schemes cannot simultaneously be conservative and enforce a discretized Lorentz force orthogonal to the magnetic field lines in every grid cell at the same time [Tót00], CT methods do notably decrease the magnitude of the parallel component of the force acting on the fluid [LBA+22].

Interested readers can find a review of a number of methods for handling the solenoidal constraint discretely in [Tót00].

Chapter 4

A Time-Explicit Two-Speed Relaxation Method

In this chapter we construct a Godunov-type finite volume method for solving the Euler equations with gravity. The main focus lies on reducing the artificial dissipation in the low Mach number limit and preserving steady states at rest. The method though does not address the problem of the restrictive CFL condition resulting from small Mach numbers. In order to design the method's artificial dissipation Mach number independent, we rely on a *two-speed relaxation approach* that is originally constructed for the homogeneous and barotropic Euler equations [BCG20]. Similar to other low Mach modifications, the dissipation term is rescaled in the one-dimensional approximate Riemann solver. While the artificial low Mach fix of other methods might reduce their stability, the two-speed Riemann solver is naturally derived from a new relaxation system. This fundamental basis enables the solver to satisfy a discrete entropy inequality, which improves the stability of the method. The two-speed approach has already been used to construct an IMEX scheme for the homogeneous Euler equations where the two speeds are used to fully linearize the acoustic sub-system so that only an elliptic operator with constant coefficients needs to be inverted [BFN20], and in the original time-explicit spirit for the homogeneous ideal MHD equations [BK23a]. We extend the original explicit method to the full Euler equations with gravity. To obtain a well-balanced method, the source term is included in the approximate Riemann solver by solving the inhomogeneous Riemann problem, which ensures that the solver stays at rest in case of hydrostatic equilibria [DZBK16].

We construct the scheme in the notation of the dimensionless Euler equations given in (2.50) in order to illustrate the effects of low Mach numbers in the scheme. The gravitational source term is written in the form (2.25) using the gravitational potential Φ. The presentation of the method closely follows the results published in [BCK23] and is structured as follows. First, in Sect. 4.1 we define the relaxation system that approximates the original Euler equations and construct the corresponding approximate Riemann solver. On this basis, the Godunov-type method is defined in Sect. 4.2. In Sect. 4.3, we analyze the theoretical properties of the approximate Riemann solver. Then, the method is extended to second order in Sect. 4.4 and multiple space dimensions in Sect. 4.5. Finally, the overall relaxation scheme is investigated in various numerical experiments, including setups near hydrostatic equilibria and in the low Mach number regime.

4.1 Relaxation Model

The following one-dimensional relaxation system is based at its core on the Suliciu relaxation model described in Sect. 3.4.2. The pressure p is approximated by the relaxation variable π and we add an additional equation describing its behavior to the system

$$\partial_t \rho\pi + \partial_x(\rho\pi v) + ab\partial_x v = \rho\frac{p-\pi}{\varepsilon}. \tag{4.1}$$

While only one relaxation speed is used in the classical Suliciu relaxation model, here two speeds $a > 0$ and $b > 0$ appear, as proposed in [BCG20]. This will be useful to control viscosity for pressure and velocity separately. These speeds will be defined later in Sect. 4.3.4 so that they meet stability criteria and keep the viscosity bounded in the low Mach regime. For the velocity we write v instead of u because also the velocity u is approximated by a relaxation variable v. So the following equation is introduced

$$\partial_t(\rho v) + \partial_x\left(\rho v^2\right) + \frac{a}{b}\partial_x\frac{\pi}{\mathcal{M}^2} = \rho\frac{u-v}{\varepsilon} - \frac{a}{b}\frac{1}{\mathcal{M}^2}\rho\partial_x\Phi. \tag{4.2}$$

In the next step, we also want to include the gravitational potential in the approximate Riemann solver. Since we consider a time-independent gravitational force, the intuitive way would be to add

$$\partial_t\Phi = 0 \tag{4.3}$$

to the relaxation model. Unfortunately, adding this equation introduces an additional wave with zero wave speed (see Sect. 2.3.3), which prevents a fixed ordering of the wave speeds and makes it rather cumbersome to determine the Riemann solver for the relaxation system. Instead, we decide to relax the gravitational potential Φ by a relaxation variable Z, as it is done in [DZBK16], and add a transport relaxation equation to the relaxation system

$$\partial_t\rho Z + \partial_x\rho Z v = \rho\frac{\Phi - Z}{\varepsilon}. \tag{4.4}$$

Finally, we add transport equations for the relaxation speeds a and b to the system and derive the following relaxation model

$$\partial_t\rho + \partial_x(\rho v) = 0, \tag{4.5a}$$

$$\partial_t(\rho u) + \partial_x\left(\rho uv + \frac{\pi}{\mathcal{M}^2}\right) = -\frac{1}{\mathcal{M}^2}\rho\partial_x Z, \tag{4.5b}$$

$$\partial_t E + \partial_x((E+\pi)v) = -\rho v\partial_x Z, \tag{4.5c}$$

$$\partial_t(\rho\pi) + \partial_x(\rho\pi v) + ab\partial_x v = \rho\frac{p-\pi}{\varepsilon}, \tag{4.5d}$$

$$\partial_t(\rho v) + \partial_x\left(\rho v^2\right) + \frac{a}{b}\partial_x\frac{\pi}{\mathcal{M}^2} = \rho\frac{u-v}{\varepsilon} - \frac{a}{b}\frac{1}{\mathcal{M}^2}\rho\partial_x Z, \tag{4.5e}$$

$$\partial_t\rho Z + \partial_x\rho Z v = \rho\frac{\Phi - Z}{\varepsilon}, \tag{4.5f}$$

$$\partial_t a + v\partial_x a = 0, \tag{4.5g}$$

$$\partial_t b + v\partial_x b = 0. \tag{4.5h}$$

A Chapman-Enskog expansion as done for the Jin-Xin relaxation system in Sect. 3.4.1 shows that the solutions to this relaxation model can be seen as a viscous approximation of the solutions of the original Euler system (2.50) as long as the subcharacteristic conditions

$$a \geq b \quad \text{and} \quad ab \geq \rho^2 c^2 \tag{4.6}$$

are satisfied.

Remark 4.1.1. *When neglecting gravitational forces ($\Phi = Z = 0$), we can recover the standard Suliciu relaxation model (3.66) by choosing $v = u$ and $b = a$.*

The homogeneous system, denoted by $(4.5)_{\varepsilon=\infty}$, has the following properties.

Lemma 4.1.2. *The relaxation system $(4.5)_{\varepsilon=\infty}$ is hyperbolic and all characteristic fields are linearly degenerate. The eigenvalues of the system are given by*

$$\lambda^- = v - \frac{a}{\mathcal{M}\rho}, \quad \lambda^v = v, \quad \lambda^+ = v + \frac{a}{\mathcal{M}\rho}, \tag{4.7}$$

where λ^v has multiplicity six. The eigenvalues have the fixed ordering

$$\lambda^- < \lambda^v < \lambda^+. \tag{4.8}$$

The Riemann invariants for the different characteristic fields are

$$\begin{aligned}
\lambda^- :\ & I_1 = v - \tfrac{a}{\mathcal{M}\rho}, && I_2 = u - \tfrac{b}{\mathcal{M}\rho}, && I_3 = \tfrac{1}{\rho} + \tfrac{\pi}{ab}, && I_4 = e + \tfrac{(a-b)b + 2\rho(\pi - \mathcal{M}b(v-u))}{2\rho^2},\\
& I_5 = a, && I_6 = b, && I_7 = Z, && \\
\lambda^v :\ & I_8 = v, && && && \\
\lambda^+ :\ & I_9 = v + \tfrac{a}{\mathcal{M}\rho}, && I_{10} = u + \tfrac{b}{\mathcal{M}\rho}, && I_{11} = \tfrac{1}{\rho} + \tfrac{\pi}{ab}, && I_{12} = e + \tfrac{(a-b)b + 2\rho(\pi + \mathcal{M}b(v-u))}{2\rho^2},\\
& I_{13} = a, && I_{14} = b, && I_{15} = Z. &&
\end{aligned} \tag{4.9}$$

Proof. The computations are straightforward and left to the reader. □

Remark 4.1.3. *The relaxation system $(4.5)_{\varepsilon=\infty}$ provides only one Riemann invariant for the contact wave, which is v. Here, in contrast to the standard Suliciu model, the pressure π is not a Riemann invariant. As a result, the associated Riemann problem is under-determined.*

Let us now consider a single Riemann problem associated with the system $(4.5)_{\varepsilon=\infty}$. In order to simplify the notations we introduce the state vector

$$W = (\rho, \rho u, E, \rho\pi, \rho v, \rho Z, a, b)^T \tag{4.10}$$

in the phase space

$$\mathcal{O} = \{W \in \mathbb{R}^8 :\ \rho > 0,\ e > 0\}. \tag{4.11}$$

Additionally, for $\mathcal{Q} \in \Omega^{\text{Euler}}_{\text{phys}}$ and given gravitational potential Φ we denote the state vector at relaxation equilibrium by

$$W^{eq}(\mathcal{Q}) = (\rho, \rho u, E, \rho p(\tau, e), \rho u, \rho\Phi, a, b)^T. \tag{4.12}$$

Then the initial data of the Riemann problem is given by two constant states W^L and W^R separated by one discontinuity located at $x = 0$

$$W_0(x) = \begin{cases} W^L, & \text{if } x < 0, \\ W^R, & \text{if } x > 0. \end{cases} \tag{4.13}$$

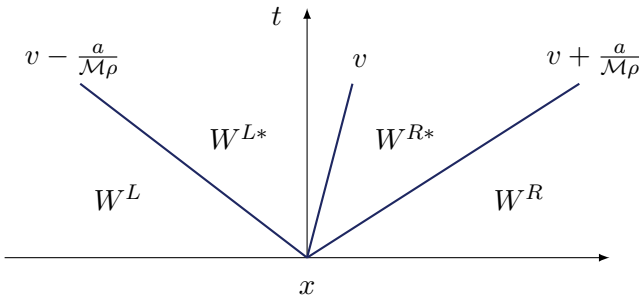

Figure 4.1: Schematic diagram of the Riemann fan for the relaxation system $(4.5)_{\varepsilon=\infty}$. The Riemann solution consists of four constant states W^L, W^{L*}, W^{R*} and W^R. The states are separated by three waves with the wave speeds $v - a/(\mathcal{M}\rho)$, v and $v + a/(\mathcal{M}\rho)$.

The solution to this problem consists of four constant states, each separated by a contact discontinuity. Therefore, the approximate Riemann solver $\mathcal{W}_{\mathcal{R}}(x/t; W^L, W^R)$ has the structure

$$\mathcal{W}_{\mathcal{R}}\left(\frac{x}{t}; W^L, W^R\right) = \begin{cases} W^L, & \text{if } \frac{x}{t} < \lambda^-, \\ W^{L*}, & \text{if } \lambda^- < \frac{x}{t} < \lambda^v, \\ W^{R*}, & \text{if } \lambda^v < \frac{x}{t} < \lambda^+, \\ W^R, & \text{if } \lambda^+ < \frac{x}{t}. \end{cases} \tag{4.14}$$

This structure of the solution is also shown in Fig. 4.1. For the computation of the intermediate states W^{L*} and W^{R*} we can use the Riemann invariants given in Lem. 4.1.2. Since Riemann invariants are constant across their corresponding wave, each Riemann invariant provides one equation. However, only 15 Riemann invariants face 16 unknown intermediate states. Therefore, the Riemann problem (4.13) is, as already stated in Rem. 4.1.3, under-determined. An additional condition is necessary, which takes on the role of the missing Riemann invariant π and connects the left and right intermediate state of the pressure. We follow the approach in [DZBK16] and choose to introduce the relation

$$\pi^{R*} - \pi^{L*} = -\bar{\rho}\left(W^L, W^R\right)\left(Z^R - Z^L\right). \tag{4.15}$$

This equation is a discrete representation of the steady states at rest in (2.45) in one spatial dimension and guarantees that the intermediate states of the pressure fulfill the hydrostatic equilibrium equation for this specific discretization of the source term. This fact will become very useful for the well-balancing of hydrostatic equilibria. The function $\bar{\rho}$ denotes a ρ-average function and depends on the underlying hydrostatic equilibrium. We leave its explicit definition open at this point and will present several possible definitions in Sect. 4.3.5.

Adding the closure equation to the equations resulting from the Riemann invariants enables us to compute the intermediate states in the Riemann solution.

Lemma 4.1.4. *The solution of the Riemann problem* (4.13) *associated with the relaxation*

system $(4.5)_{\varepsilon=\infty}$ *has the structure given in* (4.14) *with the intermediate states*

$$v^* = \frac{\mathcal{M}b^L v^L + \mathcal{M}b^R v^R + \pi^L - \pi^R - \bar{\rho}\left(W^L, W^R\right)\left(Z^R - Z^L\right)}{\mathcal{M}(b^L + b^R)}, \tag{4.16}$$

$$\frac{1}{\rho^{L*}} = \frac{1}{\rho^L} + \frac{\mathcal{M}b^R\left(v^R - v^L\right) + \pi^L - \pi^R - \bar{\rho}\left(W^L, W^R\right)\left(Z^R - Z^L\right)}{a^L\left(b^L + b^R\right)}, \tag{4.17}$$

$$\frac{1}{\rho^{R*}} = \frac{1}{\rho^R} + \frac{\mathcal{M}b^L\left(v^R - v^L\right) + \pi^R - \pi^L + \bar{\rho}\left(W^L, W^R\right)\left(Z^R - Z^L\right)}{a^R\left(b^L + b^R\right)}, \tag{4.18}$$

$$u^{L*} = u^L + \frac{b^L\left(\mathcal{M}b^R\left(v^R - v^L\right) + \pi^L - \pi^R - \bar{\rho}\left(W^L, W^R\right)\left(Z^R - Z^L\right)\right)}{\mathcal{M}a^L\left(b^L + b^R\right)}, \tag{4.19}$$

$$u^{R*} = u^R + \frac{b^R\left(\mathcal{M}b^L\left(v^L - v^R\right) + \pi^L - \pi^R - \bar{\rho}\left(W^L, W^R\right)\left(Z^R - Z^L\right)\right)}{\mathcal{M}a^R\left(b^L + b^R\right)}, \tag{4.20}$$

$$\pi^{L*} = \frac{b^R\pi^L + b^L\pi^R + \mathcal{M}b^L b^R\left(v^L - v^R\right) + b^L\bar{\rho}\left(W^L, W^R\right)\left(Z^R - Z^L\right)}{b^L + b^R}, \tag{4.21}$$

$$\pi^{R*} = \frac{b^R\pi^L + b^L\pi^R + \mathcal{M}b^L b^R\left(v^L - v^R\right) - b^R\bar{\rho}\left(W^L, W^R\right)\left(Z^R - Z^L\right)}{b^L + b^R}, \tag{4.22}$$

$$e^{L*} = e^L + \frac{(\pi^{L*})^2 - (\pi^L)^2}{2a^L b^L} + \frac{(v^* - u^{L*})^2 - (v^L - u^L)^2}{2(\frac{a^L}{b^L} - 1)}, \tag{4.23}$$

$$e^{R*} = e^R + \frac{(\pi^{R*})^2 - (\pi^R)^2}{2a^R b^R} + \frac{(v^* - u^{R*})^2 - (v^R - u^R)^2}{2(\frac{a^R}{b^R} - 1)}, \tag{4.24}$$

$$a^{L*} = a^L,\ a^{R*} = a^R,\ b^{L*} = b^L,\ b^{R*} = b^R,\ Z^{L*} = Z^L,\ Z^{R*} = Z^R. \tag{4.25}$$

Proof. The intermediate states can be computed by solving the system of equations given by the Riemann invariants and the closure equation (4.15). The precise steps are straightforward and therefore left to the reader. □

Including the source term in the Riemann problem means that the gravitational potential is also contained in the intermediate states. In comparison to the states in the standard Suliciu solver for the homogeneous Euler equations, the term $\bar{\rho}(W^L, W^R)(Z^R - Z^L)$ is added to the pressure difference $(\pi^R - \pi^L)$ in each state. As a consequence, each intermediate state contains a discretization of the one-dimensional steady state equation (2.45).

Remark 4.1.5. *At this point, we do not explicitly define the relaxation speeds* a^L, a^R, b^L *and* b^R, *since later, in the proofs of the properties of the Riemann solver, various conditions are placed on these speeds. The explicit definitions are then provided in Sect. 4.3.4.*

Equipped with the approximate Riemann solver, we can now define the overall discretization of the scheme in the next section.

4.2 Relaxation Scheme

At the start of each time step, we assume to be at the relaxation equilibrium. For that reason, the initial data for the relaxation variables at time level n is defined by

$$\pi_i^n = p_i^n, \quad v_i^n = u_i^n, \quad Z_i^n = \Phi_i^n. \tag{4.26}$$

Starting from the equilibrium we solve the homogeneous relaxation system $(4.5)_{\varepsilon=\infty}$ using the Riemann solver $\mathcal{W}_{\mathcal{R}}$ defined in (4.14) and update the cell averages to the next time level t^{n+1} by a Godunov-type method of the form

$$\begin{aligned} Q_i^{n+1} =& Q_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^n - F_{i-1/2}^n\right) \\ &+ \frac{\Delta t}{2}\left(S_{i-1/2}^{+,n}\frac{\Phi_i^n - \Phi_{i-1}^n}{\Delta x} + S_{i+1/2}^{-,n}\frac{\Phi_{i+1}^n - \Phi_i^n}{\Delta x}\right), \\ F_{i-1/2}^n =& F(Q_{i-1}^n, \Phi_{i-1}^n, Q_i^n, \Phi_i^n), \qquad F_{i+1/2}^n = F(Q_i^n, \Phi_i^n, Q_{i+1}^n, \Phi_{i+1}^n), \\ S_{i-1/2}^{+,n} =& S^+(Q_{i-1}^n, \Phi_{i-1}^n, Q_i^n, \Phi_i^n), \quad S_{i+1/2}^{-,n} = S^-(Q_i^n, \Phi_i^n, Q_{i+1}^n, \Phi_{i+1}^n). \end{aligned} \tag{4.27}$$

The numerical flux is defined by

$$F(Q^L, \Phi^L, Q^R, \Phi^R) = \begin{cases} \mathcal{F}(Q^L), & \text{if } \lambda^- > 0, \\ F^{L*}, & \text{if } \lambda^- < 0 \le \lambda^v, \\ F^{R*}, & \text{if } \lambda^v < 0 < \lambda^+, \\ \mathcal{F}(Q^R), & \text{if } \lambda^+ < 0, \end{cases} \tag{4.28}$$

where according to the left-hand sides of the first three equations of (4.5) the intermediate fluxes can be written as

$$\begin{aligned} F^{L*} &= \left(\rho^{L*}v^*, \rho^{L*}u^{L*}v^* + \frac{\pi^{L*}}{\mathcal{M}^2}, (E^{L*} + \pi^{L*})v^*\right), \\ F^{R*} &= \left(\rho^{R*}v^*, \rho^{R*}u^{R*}v^* + \frac{\pi^{R*}}{\mathcal{M}^2}, (E^{R*} + \pi^{R*})v^*\right). \end{aligned} \tag{4.29}$$

The numerical source terms are set as follows

$$\begin{aligned} S^+(Q^L, \Phi^L, Q^R, \Phi^R) &= -(\mathfrak{s}(v^*) + 1)\left(0, \frac{1}{\mathcal{M}^2}\bar{\rho}(W^L, W^R), \bar{\rho}(W^L, W^R)v^*\right)^T, \\ S^-(Q^L, \Phi^L, Q^R, \Phi^R) &= \quad (\mathfrak{s}(v^*) - 1)\left(0, \frac{1}{\mathcal{M}^2}\bar{\rho}(W^L, W^R), \bar{\rho}(W^L, W^R)v^*\right)^T. \end{aligned} \tag{4.30}$$

The function $\mathfrak{s} : \mathbb{R} \to \mathbb{R}$ in the source terms is defined by

$$\mathfrak{s}(v) = \begin{cases} 1, & \text{if } v \ge 0, \\ -1, & \text{else}, \end{cases} \tag{4.31}$$

and is used to adjust the source term to the choice of the numerical flux in (4.28). This is necessary to establish the well-balancing property of the scheme in Sect. 4.3.5.

The Godunov-type scheme (4.27) underlies a CFL time step restriction of the form

$$\frac{\Delta t}{\Delta x}\max_i\left\{\left|v_i - \frac{a_i}{\mathcal{M}\rho_i}\right|, \left|v_i + \frac{a_i}{\mathcal{M}\rho_i}\right|\right\} \le \frac{1}{2}. \tag{4.32}$$

We note that in this procedure only the variables of the original Euler equations (2.50) are updated to the next time level. The relaxation variables π, v and Z are just used for the update of the original variables, but are not actually updated. Instead, for the upcoming time step, we again assume to be at the equilibrium. As a consequence of this projection approach, the relaxation parameter ε does not appear in the relaxation scheme (4.27) and thus does not have to be set explicitly.

4.3 Properties of the Relaxation Scheme

In this section we focus on the properties of the numerical scheme just described. We start with the property of entropy stability.

4.3.1 Entropy Inequality

The final scheme shall seek those correct solutions that satisfy the entropy inequality. In practice, it can be observed that searching for entropy solutions makes a finite volume method more stable. This is partly the case because an entropy inequality can help to ensure the positivity of density and/or internal energy.
We have seen in Sect. 2.3.2 that smooth solutions of the Euler equations satisfy an additional conservation law for the entropy. In one spatial dimension this conservation law is given by

$$\partial_t(\rho\mathcal{G}(s)) + \partial_x(\rho\mathcal{G}(s)u) = 0 \tag{4.33}$$

for all smooth functions $\mathcal{G}$. However, since the Euler equations are nonlinear, discontinuities can arise in the solution in finite time despite of smooth initial conditions. At discontinuities the equation (4.33) is not valid, since it does not consider the entropy dissipation at shocks. Therefore, we replace the equality in (4.33) by an inequality, which leads to the following entropy inequality

$$\partial_t(\rho\mathcal{G}(s) + \partial_x(\rho\mathcal{G}(s)u) \leq 0. \tag{4.34}$$

Our scheme should now mimic this behavior in the sense that its solutions satisfy a discrete version of (4.34).

Theorem 4.3.1. *Let us assume that Q_i^n belongs to $\Omega_{\text{phys}}^{\text{Euler}}$ for all $i \in \mathbb{Q}$. Furthermore, we assume that at each interface with initial left state Q^L and initial right state Q^R the intermediate states for density and internal energy in the Riemann solution are positive, i.e. $\rho^{L*}, \rho^{R*}, e^{L*}, e^{R*} > 0$, and that the relaxation speeds $a^{L,R}$ and $b^{L,R}$ are such that they satisfy the subcharacteristic Whitham conditions*

$$a^L b^L > p(\tau^L, e^L)\partial_e p(\tau^L, e^L) - \partial_\tau p(\tau^L, e^L), \tag{4.35}$$
$$a^L b^L > p(\tau^{L*}, e^{L*})\partial_e p(\tau^{L*}, e^{L*}) - \partial_\tau p(\tau^{L*}, e^{L*}), \tag{4.36}$$
$$a^R b^R > p(\tau^{R*}, e^{R*})\partial_e p(\tau^{R*}, e^{R*}) - \partial_\tau p(\tau^{R*}, e^{R*}), \tag{4.37}$$
$$a^R b^R > p(\tau^R, e^R)\partial_e p(\tau^R, e^R) - \partial_\tau p(\tau^R, e^R). \tag{4.38}$$

Moreover, we assume that the pressure law satisfies Assumption 4.3.3.
Then for all $i \in \mathbb{Q}$, the updated state Q_i^{n+1}, computed with the relaxation scheme (4.27) under the CFL condition (4.32), satisfies the discrete entropy inequality

$$\rho_i^{n+1}\mathcal{G}(s_i^{n+1}) - \rho_i^n\mathcal{G}(s_i^n) - \frac{\Delta t}{\Delta x}\left(\{\rho\mathcal{G}(s)u\}_{i+1/2}^n - \{\rho\mathcal{G}(s)u\}_{i-1/2}^n\right) \leq 0, \tag{4.39}$$

where we define the numerical entropy flux by

$$\{\rho\mathcal{G}u\}_{i-1/2}^n = \{\rho\mathcal{G}(s)u\}\left(W^{eq}(Q_{i-1}^n), W^{eq}(Q_i^n)\right), \tag{4.40}$$

$$\{\rho\mathcal{G}u\}^{L,R} = \{\rho\mathcal{G}(s)u\}\left(W^{eq}(Q^L), W^{eq}(Q^R)\right) = \begin{cases} \rho^L\mathcal{G}(s(\tau^L, e^L))u^L, & \text{if } \lambda^- > 0, \\ \rho^{L*}\mathcal{G}(\hat{s}(W^{L*}))v^*, & \text{if } \lambda^- < 0 \leq \lambda^v, \\ \rho^{R*}\mathcal{G}(\hat{s}(W^{R*}))v^*, & \text{if } \lambda^v < 0 < \lambda^+, \\ \rho^R\mathcal{G}(s(\tau^R, e^R))u^R, & \text{if } \lambda^+ < 0, \end{cases} \tag{4.41}$$

where the function $W \mapsto \hat{s}(W)$ is defined by (4.45).

Remark 4.3.2. *At the beginning of this theorem, we assume the intermediate states of density and internal energy to be positive. In Sect. 4.3.3 we show that the approximate Riemann solver 4.14 satisfies this property for suitably chosen relaxation speeds.*

Proof. **(Proof of Theorem 4.3.1)** The proof of this theorem closely follows the steps of a similar proof in [DZBK16]. Therefore, we only give the basic outline of the proof here and do not prove every intermediate step. For more details see [DZBK16]. First of all, it is easy to check that

$$I(W) = \pi + ab\tau \quad \text{and} \quad J(W) = e - \frac{\mathcal{M}^2(v-u)^2}{2(\frac{a}{b}-1)} - \frac{\pi^2}{2ab} \tag{4.42}$$

are strong Riemann invariants of $(4.5)_{\varepsilon=\infty}$. Therefore, weak solutions of $(4.5)_{\varepsilon=\infty}$ satisfy

$$\partial_t \rho\Psi(I,J) + \partial_x \rho\Psi(I,J)v = 0 \tag{4.43}$$

for all smooth functions $\Psi : \mathbb{R}^2 \to \mathbb{R}$. As a consequence, for a function $W \mapsto \hat{s}(W)$, which only depends on I and J, weak solutions of $(4.5)_{\varepsilon=\infty}$ satisfy the additional conservation law

$$\partial_t \rho\mathcal{G}(\hat{s}) + \partial_x \rho\mathcal{G}(\hat{s})v = 0. \tag{4.44}$$

We define the function $\hat{s}$ by

$$\hat{s}(W) = s(\hat{\tau}(I(W), J(W)), \hat{e}(I(W), J(W))), \tag{4.45}$$

where $\hat{\tau}(I,J)$ is the largest root within $\mathbb{R}^+$ of the function $f_{I,J} : \mathbb{R}^+ \to \mathbb{R}$ defined by

$$f_{I,J}(\tau) = \tau p\left(\tau, e(\tau, I - ab\tau)\right) + ab\tau^2 - I\tau \tag{4.46}$$

and $\hat{e}$ is defined by

$$\hat{e}(I,J) = e(\hat{\tau}(I,J), I - ab\hat{\tau}(I,J)). \tag{4.47}$$

For the further steps, the following assumption is made about the pressure law.

Assumption 4.3.3. *We assume that the pressure law is such that the function $\tau \mapsto f_{I,J}$ is strictly convex for all fixed pairs (I,J).*

This condition is fulfilled by most common pressure laws, including the ideal gas law [DZBK16]. Under this assumption, it can be proven (see [DZBK16]) that for all W, for which the pair $(I(W), J(W))$ is in

$$\begin{aligned}\mathcal{A} = \{&(I,J) \in \mathbb{R}^2, \exists \tau > 0, \exists e > 0, \exists v, \exists u \text{ such that:}\\ &I = p(\tau,e) + ab\tau, &(4.48)\\ &J = e - \frac{p(\tau,e)^2}{2ab}, &(4.49)\\ &ab > p(\tau,e)\partial_e p(\tau,e) - \partial_\tau p(\tau,e)\}, &(4.50)\end{aligned}$$

the function $\hat{s}$ is larger than the specific entropy of the original system, i.e.

$$\hat{s}(W) \geq s(\tau,e) \tag{4.51}$$

and that equality is reached in the relaxation equilibrium, i.e.

$$\hat{s}(W_{|\pi=p(\tau,e),v=u}) = s(\tau, e). \tag{4.52}$$

Let us now go back to the additional conservation law (4.44) and integrate it over $[0, \Delta x/2) \times [0, \Delta t)$

$$\begin{aligned}\int_0^{\Delta x/2} (\rho\mathcal{G}(\hat{s}) \left(\mathcal{W}_{\mathcal{R}}\left(\frac{x}{\Delta t}; W^{eq}(Q^L), W^{eq}(Q^R)\right)\right) \mathrm{d}x = \int_0^{\Delta x/2} (\rho\mathcal{G}(\hat{s}))(W(x,0))\mathrm{d}x \\ -\Delta t(\rho\mathcal{G}(\hat{s})v)\left(\mathcal{W}_{\mathcal{R}}\left(\frac{\Delta x}{2\Delta t}; W^{eq}(Q^L), W^{eq}(Q^R)\right)\right) \\ +\Delta t(\rho\mathcal{G}(\hat{s})v)(\mathcal{W}_{\mathcal{R}}(0; W^{eq}(Q^L), W^{eq}(Q^R))).\end{aligned} \tag{4.53}$$

Under consideration of the CFL condition (4.32) and equality (4.52), this can be rewritten as

$$\begin{aligned}&\frac{1}{\Delta x}\int_0^{\Delta x/2} (\rho\mathcal{G}(\hat{s}) \left(\mathcal{W}_{\mathcal{R}}\left(\frac{x}{\Delta t}; W^{eq}(Q^L), W^{eq}(Q^R)\right)\right) \mathrm{d}x \\ &= \frac{\rho^R\mathcal{G}(s^R)}{2} - \frac{\Delta t}{\Delta x}\left(\rho^R\mathcal{G}(s^R)u^R - \{\rho\mathcal{G}u\}^{L,R}\right).\end{aligned} \tag{4.54}$$

The replacement of v by u in the entropy fluxes is due to the fact that the input values of the approximate Riemann solver are at equilibrium and therefore left and right states of u and v are equal in each case. Just in the intermediate states both velocities differ, which is the reason why we write v^* in (4.41). Due to the inequality (4.51), it follows

$$\hat{s}\left(\mathcal{W}_{\mathcal{R}}\left(\frac{x}{\Delta t}; W^{eq}(Q^L), W^{eq}(Q^R)\right)\right) \geq s\left((\tau^{eq}, e^{eq})\left(\frac{x}{\Delta t}; Q^L, Q^R\right)\right). \tag{4.55}$$

The quantities τ^{eq}, e^{eq} on the right-hand side originate from the approximate Riemann solver $\mathcal{W}_{\mathcal{R}}(x/\Delta t; W^{eq}(Q^L), W^{eq}(Q^R))$. Since we assume $\mathcal{G}$ to be increasing, see (2.32), it in turn follows that

$$\mathcal{G}(\hat{s})\left(\mathcal{W}_{\mathcal{R}}\left(\frac{x}{\Delta t}; W^{eq}(Q^L), W^{eq}(Q^R)\right)\right) \geq \mathcal{G}(s)\left(\mathcal{W}_{\mathcal{R}}^{(\rho,\rho u,E)}\left(\frac{x}{\Delta t}; W^{eq}(Q^L), W^{eq}(Q^R)\right)\right). \tag{4.56}$$

By replacing the content of the integral in (4.54), we obtain the inequality

$$\begin{aligned}&\frac{1}{\Delta x}\int_0^{\Delta x/2} (\rho\mathcal{G}(s)) \left(\mathcal{W}_{\mathcal{R}}^{(\rho,\rho u,E)}\left(\frac{x}{\Delta t}; W^{eq}(Q^L), W^{eq}(Q^R)\right)\right) \mathrm{d}x \\ &\leq \frac{\rho^R\mathcal{G}(s^R)}{2} - \frac{\Delta t}{\Delta x}\left(\rho^R\mathcal{G}(s^R)u^R - \{\rho\mathcal{G}(s)u\}^{L,R}\right).\end{aligned} \tag{4.57}$$

Inserting $Q^L = Q^n_{i-1}$ and $Q^R = Q^n_i$ leads to

$$\begin{aligned}&\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_i} (\rho\mathcal{G}(s)) \left(\frac{x - x_{i-1/2}}{\Delta t}; Q^n_{i-1}, Q^n_i\right) \mathrm{d}x \\ &\leq \frac{\rho^n_i\mathcal{G}(s^n_i)}{2} - \frac{\Delta t}{\Delta x}\left(\rho^n_i\mathcal{G}(s^n_i)u^n_i - \{\rho\mathcal{G}(s)u\}^n_{i-1/2}\right).\end{aligned} \tag{4.58}$$

For the other half of the cell, on the other hand, integrating over $(-\Delta x/2, 0] \times [0, \Delta t)$ and applying similar steps as before results in

$$\begin{aligned}&\frac{1}{\Delta x}\int_{-\Delta x/2}^{0} (\rho\mathcal{G}(s)) \left(\mathcal{W}_{\mathcal{R}}^{(\rho,\rho u,E)}\left(\frac{x}{\Delta t}; W^{eq}(Q^L), W^{eq}(Q^R)\right)\right) \mathrm{d}x \\ &\leq \frac{\rho^L\mathcal{G}(s^L)}{2} - \frac{\Delta t}{\Delta x}\left(\{\rho\mathcal{G}(s)u\}^{L,R} - \rho^L\mathcal{G}(s^L)u^L\right),\end{aligned} \tag{4.59}$$

and inserting $Q^L = Q_i^n$ and $Q^R = Q_{i+1}^n$ leads to

$$\begin{aligned} &\frac{1}{\Delta x}\int_{x_i}^{x_{i+1/2}} (\rho\mathcal{G}(s))\left(\frac{x-x_{i+1/2}}{\Delta t}; Q_i^n, Q_{i+1}^n\right)\mathrm{d}x \\ &\leq \frac{\rho_i^n\mathcal{G}(s_i^n)}{2} - \frac{\Delta t}{\Delta x}\left(\{\rho\mathcal{G}(s)u\}_{i+1/2}^n - \rho_i^n\mathcal{G}(s_i^n)u_i^n\right). \end{aligned} \tag{4.60}$$

Summing up the inequalities (4.58) and (4.60) results in the inequality

$$\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} (\rho\mathcal{G}(s))(Q(x,t^{n+1}))\mathrm{d}x \leq \rho_i^n\mathcal{G}(s_i^n) - \frac{\Delta t}{\Delta x}\left(\{\rho\mathcal{G}(s)u\}_{i+1/2}^n - \{\rho\mathcal{G}(s)u\}_{i-1/2}^n\right). \tag{4.61}$$

Since we assume $\rho\mathcal{G}(s)$ to be strictly convex, by applying Jensen's inequality we get

$$\rho\mathcal{G}(s)\left(\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} Q(x,t^{n+1})\mathrm{d}x\right) \leq \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} (\rho\mathcal{G}(s))(Q(x,t^{n+1}))\mathrm{d}x. \tag{4.62}$$

Finally, we obtain the desired discrete entropy inequality

$$\rho_i^{n+1}\mathcal{G}(s_i^{n+1}) \leq \rho_i^n\mathcal{G}(s_i^n) - \frac{\Delta t}{\Delta x}\left(\{\rho\mathcal{G}(s)u\}_{i+1/2}^n - \{\rho\mathcal{G}(s)u\}_{i-1/2}^n\right). \tag{4.63}$$

□

4.3.2 Prevention of Checkerboard Modes

For asymptotic-preserving methods, stationary and non-constant solutions (checkerboard modes) may occur in the low Mach number regime, jumping between two different values. This behavior can arise from the fact that the divergence or gradient of a variable is supposed to be zero in the limit equations, while the discretization of this term allows a jumping solution. Of course, it is desirable to prevent the occurrence of this unphysical phenomenon.

Theorem 4.3.4. *Under the conditions of Theorem 4.3.1, velocity and pressure of the relaxation solver (4.14) are constant in space for steady periodic solutions.*

Proof. The proof builds on the entropy inequality of the previous subsection and follows the strategy of a similar proof in [BCG20]. First of all, using the notations used in the entropy proof, we can write

$$\begin{aligned} \rho_i^{n+1}\mathcal{G}(s_i^{n+1}) \leq& \rho_i^n\mathcal{G}(s_i^n) - \frac{\Delta t}{\Delta x}\left(\{\rho\mathcal{G}(s)u\}_{i+1/2}^n - \{\rho\mathcal{G}(s)u\}_{i-1/2}^n\right) \\ =& \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} (\rho\mathcal{G}(\hat{s}))\left(\mathcal{W}_{\mathcal{R}}(x,t^{n+1})\right)\mathrm{d}x. \end{aligned} \tag{4.64}$$

Additionally, by applying Jensen's inequality to the left-hand side we get the following inequalities

$$\begin{aligned} \rho_i^{n+1}\mathcal{G}(s_i^{n+1}) \leq& \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} (\rho\mathcal{G}(s))(Q(x,t^{n+1}))\mathrm{d}x \\ \leq& \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} (\rho\mathcal{G}(\hat{s}))\left(\mathcal{W}_{\mathcal{R}}(x,t^{n+1})\right)\mathrm{d}x. \end{aligned} \tag{4.65}$$

We now define the left-hand side of the entropy inequality (4.39) by

$$D_i^n := \rho_i^{n+1}\mathcal{G}(s_i^{n+1}) - \rho_i^n\mathcal{G}(s_i^n) - \frac{\Delta t}{\Delta x}\left(\{\rho\mathcal{G}(s)u\}_{i+1/2}^n - \{\rho\mathcal{G}(s)u\}_{i-1/2}^n\right). \tag{4.66}$$

For steady and space periodic solutions we then have

$$\sum_i D_i^n = 0. \tag{4.67}$$

In combination with the entropy inequality (4.39) we get

$$D_i^n = 0 \quad \forall i. \tag{4.68}$$

From this result follows directly that all the inequalities in (4.65) are replaced by equalities and therefore the entropy is equal to the relaxation entropy

$$(\rho\mathcal{G}(s))(Q(x,t^{n+1})) = (\rho\mathcal{G}(\hat{s}))\left(\mathcal{W}_{\mathcal{R}}(x,t^{n+1})\right). \tag{4.69}$$

In the proof of the entropy inequality it is shown that this is just the case in the relaxation equilibrium, so only if

$$\pi = p(\rho,e),\ u = v,\ \tau = \frac{1}{\rho},\ \hat{s} = s. \tag{4.70}$$

As a consequence, the following relations apply to a single Riemann problem

$$\begin{aligned} \tau^{L*} = \frac{1}{\rho^{L*}},\quad \tau^{R*} = \frac{1}{\rho^{R*}},\quad v^* = u^{L*} = u^{R*},\\ \pi^{L*} = p(\rho^{L*},e^{L*}) = p(\rho^{L*},s^{L*}),\quad \pi^{R*} = p(\rho^{R*},e^{R*}) = p(\rho^{R*},s^{R*}). \end{aligned} \tag{4.71}$$

Since τ is a Riemann invariant for λ^- and λ^+, it holds

$$\tau^{L*} = \tau^L,\ \tau^{R*} = \tau^R. \tag{4.72}$$

We can use this fact to gain more information about the intermediate densities

$$\begin{aligned} \frac{1}{\rho^{L*}} = \tau^{L*} = \frac{1}{\rho^L} \quad &\Rightarrow \quad \rho^{L*} = \rho^L,\\ \frac{1}{\rho^{R*}} = \tau^{R*} = \frac{1}{\rho^R} \quad &\Rightarrow \quad \rho^{R*} = \rho^R. \end{aligned} \tag{4.73}$$

From the explicit definition of the intermediate states in (4.17) and (4.18) we can deduce that

$$\frac{1}{\rho^{L*}} - \frac{1}{\rho^L} = \frac{\mathcal{M}b^R\left(v^R - v^L\right) + \pi^L - \pi^R - \bar{\rho}\left(W^L,W^R\right)\left(Z^R - Z^L\right)}{a^L\left(b^L + b^R\right)} = 0, \tag{4.74}$$

$$\frac{1}{\rho^{R*}} - \frac{1}{\rho^R} = \frac{\mathcal{M}b^L\left(v^R - v^L\right) + \pi^R - \pi^L + \bar{\rho}\left(W^L,W^R\right)\left(Z^R - Z^L\right)}{a^R\left(b^L + b^R\right)} = 0. \tag{4.75}$$

With a look at the intermediate states u^{L*} and u^{R*}, we see that we can use (4.74) and (4.75) to get

$$u^{L*} = u^L + \frac{b^L}{\mathcal{M}}\frac{\mathcal{M}b^R\left(v^R - v^L\right) + \pi^L - \pi^R - \bar{\rho}\left(W^L,W^R\right)\left(Z^R - Z^L\right)}{a^L\left(b^L + b^R\right)} = u^L, \tag{4.76}$$

$$u^{R*} = u^R + \frac{b^R}{\mathcal{M}}\frac{\mathcal{M}b^L\left(v^L - v^R\right) + \pi^L - \pi^R + \bar{\rho}\left(W^L,W^R\right)\left(Z^R - Z^L\right)}{a^R\left(b^L + b^R\right)} = u^R. \tag{4.77}$$

Since we are at equilibrium we can conclude that

$$v^* = u^{L*} = u^{R*} = u^L = u^R = v^L = v^R. \tag{4.78}$$

In the next part we will show that the left and the right state at the interface are equal for π. From the Riemann invariants in (4.9) we take

$$I_3 = I_{11} = \frac{1}{\rho} + \frac{\pi}{2ab}. \tag{4.79}$$

This quantity is constant across the left and right waves in the Riemann fan, which means

$$\begin{aligned} \frac{1}{\rho^{L*}} + \frac{\pi^{L*}}{2a^L b^L} &= \frac{1}{\rho^L} + \frac{\pi^L}{2a^L b^L}, \\ \frac{1}{\rho^{R*}} + \frac{\pi^{R*}}{2a^R b^R} &= \frac{1}{\rho^R} + \frac{\pi^R}{2a^R b^R}. \end{aligned} \tag{4.80}$$

It has already been established in (4.73) that the density has only two states and therefore we can simplify the equations to

$$\begin{aligned} \pi^{L*} &= \pi^L \\ \pi^{R*} &= \pi^R. \end{aligned} \tag{4.81}$$

From the explicit definition of the intermediate states and the closure equation (4.15) we gain

$$\begin{aligned} \pi^{L*} = \pi^L &= \frac{b^R\pi^L + b^L\pi^R + \mathcal{M}b^Lb^R\left(v^L - v^R\right) - b^L\bar{\rho}\left(W^L, W^R\right)\left(Z^R - Z^L\right)}{b^L + b^R} \\ &\overset{(4.78)}{=} \frac{b^R\pi^L + b^L\pi^R - b^L\bar{\rho}\left(W^L, W^R\right)\left(Z^R - Z^L\right)}{b^L + b^R} \\ &\overset{(4.15)}{=} \frac{b^R\pi^L + b^L\pi^R + b^L\left(\pi^R - \pi^L\right)}{b^L + b^R}. \end{aligned} \tag{4.82}$$

Solving for π^L gives

$$\pi^L = \pi^R. \tag{4.83}$$

Thus, we have shown that for both velocities and the pressure, the left and right states at the interface are equal. The solution in these quantities is therefore constant in space. □

Remark 4.3.5. *For pressure laws that depend only on density, it can also be proven that the density and the internal energy are constant for steady and space periodic solutions.*

For steady periodic solutions of the relaxation method the velocity and pressure are constant, which contradicts the non-constant nature of checkerboard modes. The result of the above lemma can thus be interpreted in so far that in velocity and pressure no checkerboard modes can occur.

4.3.3 Positivity-Preserving Property

The two-speed relaxation method shall only compute physical relevant solutions that lie within $\Omega_{\text{phys}}^{\text{Euler}}$. In addition, it is also essential for the robustness of the numerical method that density and internal energy remain positive during the simulation. Otherwise, operations such as taking the root of negative numbers may occur, which lead to the premature termination of the simulation. The following lemma guarantees that the approximate Riemann solver computes only positive values for both density and internal energy.

Lemma 4.3.6. *Given $Q^L, Q^R \in \Omega^{\text{Euler}}_{\text{phys}}$. If the relaxation speeds a^L and a^R are large enough to ensure*

$$v^L - \frac{a^L}{\mathcal{M}\rho^L} < v^* < v^R + \frac{a^R}{\mathcal{M}\rho^R}, \tag{4.84}$$

$$e^L + \frac{(\pi^{L*})^2 - (\pi^L)^2}{2a^L b^L} + \frac{(v^* - u^{L*})^2 - (v^L - u^L)^2}{2(\frac{a^L}{b^L} - 1)} > 0, \tag{4.85}$$

$$e^R + \frac{(\pi^{R*})^2 - (\pi^R)^2}{2a^R b^R} + \frac{(v^* - u^{R*})^2 - (v^R - u^R)^2}{2(\frac{a^R}{b^R} - 1)} > 0, \tag{4.86}$$

then the approximate Riemann solver $\mathcal{W}_{\mathcal{R}}$ preserves the positivity of density and internal energy.

Proof. First, it is trivial that the conditions (4.84), (4.85) and (4.86) are satisfied for sufficiently large a^L and a^R. To prove the positivity of the density intermediate states in a next step, we start with the Riemann invariants I_1 and I_9 from Lem. 4.1.2, which give us

$$v^L - \frac{a^L}{\mathcal{M}\rho^L} = v^* - \frac{a^L}{\mathcal{M}\rho^{L*}} \quad \text{and} \quad v^R + \frac{a^R}{\mathcal{M}\rho^R} = v^* + \frac{a^R}{\mathcal{M}\rho^{R*}}. \tag{4.87}$$

Using these relations, we can rewrite (4.84) by

$$-\rho^{L*} < 0 < \rho^{R*}. \tag{4.88}$$

So the intermediate states for the density are positive. The positivity of the internal energy directly follows from (4.85) and (4.86), since the left-hand sides of these conditions represent the left and right intermediate states of the internal energy. □

Clearly, this lemma is of limited use in practice. It states that in principle it is possible to preserve the positivity, but it does not help to find a suitable definition of the relaxation speeds that works generally. The following lemma gives stricter conditions for the relaxation speeds, which can also be used for their explicit definition. Under these conditions, it can be proven that the density is kept positive.

Lemma 4.3.7. *Consider the relaxation solver (4.14) with intermediate states and speeds defined by (4.16)-(4.25) with the initial data at equilibrium. Assume that the relaxation speeds a^L, a^R, b^L, b^R satisfy*

$$a^L \geq b^L, \quad a^R \geq b^R, \tag{4.89}$$

$$\frac{b^L}{\rho^L} \geq a_q^L, \quad \frac{b^R}{\rho^R} \geq a_q^R, \tag{4.90}$$

$$\frac{\sqrt{a^L b^L}}{\rho^L} \geq c^L\left(1 + \beta X^L\right), \quad \frac{\sqrt{a^R b^R}}{\rho^R} \geq c^R\left(1 + \beta X^R\right), \tag{4.91}$$

for some a_q^L and a_q^R depending on Q^L, Q^R and X^L, X^R defined by (4.94) and (4.95) with a parameter $\beta \geq 1$. The quantities c^L, c^R represent the sound speed. Then the approximate Riemann solver $\mathcal{W}_{\mathcal{R}}$ preserves the positivity of the density.

Proof. We start with the definition of the left intermediate density (4.17)

$$\begin{aligned}\frac{1}{\rho^{L*}} &= \frac{1}{\rho^L} + \frac{\mathcal{M}b^R\left(v^R - v^L\right) + \pi^L - \pi^R - \bar{\rho}\left(W^L, W^R\right)\left(Z^R - Z^L\right)}{a^L\left(b^L + b^R\right)} \\ &\geq \frac{1}{\rho^L} - \frac{\mathcal{M}b^R(v^L - v^R)_+}{a^L(b^L + b^R)} - \frac{\left(\pi^R - \pi^L + \bar{\rho}(W^L, W^R)(Z^R - Z^L)\right)_+}{a^L(b^L + b^R)} \\ &\geq \frac{1}{\rho^L} - \frac{\mathcal{M}(v^L - v^R)_+}{a^L} - \frac{\left(\pi^R - \pi^L + \bar{\rho}(W^L, W^R)(Z^R - Z^L)\right)_+}{a^L(\rho^L a_q^L + \rho^R a_q^R)}. \end{aligned} \tag{4.92}$$

Analogously, for the right intermediate state we get

$$\frac{1}{\rho^{R*}} \geq \frac{1}{\rho^R} - \frac{\mathcal{M}(v^L - v^R)_+}{a^R} - \frac{\left(\pi^L - \pi^R + \bar{\rho}(W^L, W^R)(Z^L - Z^R)\right)_+}{a^R(\rho^L a_q^L + \rho^R a_q^R)}. \tag{4.93}$$

Let us now define the variables

$$X^L = \frac{1}{c^L}\left[\mathcal{M}\left(v^L - v^R\right)_+ + \frac{\left(\pi^R - \pi^L + \bar{\rho}(W^L, W^R)(Z^R - Z^L)\right)_+}{\rho^L a_q^L + \rho^R a_q^R}\right], \tag{4.94}$$

$$X^R = \frac{1}{c^R}\left[\mathcal{M}\left(v^L - v^R\right)_+ + \frac{\left(\pi^L - \pi^R + \bar{\rho}(W^L, W^R)(Z^L - Z^R)\right)_+}{\rho^L a_q^L + \rho^R a_q^R}\right], \tag{4.95}$$

in order to rewrite the former inequalities in the more compact form

$$\frac{1}{\rho^{L*}} \geq \frac{1}{\rho^L}\left(1 - \frac{\rho^L c^L}{a^L} X^L\right), \tag{4.96}$$

$$\frac{1}{\rho^{R*}} \geq \frac{1}{\rho^R}\left(1 - \frac{\rho^R c^R}{a^R} X^R\right). \tag{4.97}$$

From combining the conditions (4.89) and (4.91), it follows that

$$\frac{a^L}{\rho^L} \geq c^L(1 + \beta X^L) \quad \Rightarrow \quad \frac{\rho^L c^L}{a^L} \leq \frac{1}{1 + \beta X^L}, \tag{4.98}$$

$$\frac{a^R}{\rho^R} \geq c^R(1 + \beta X^R) \quad \Rightarrow \quad \frac{\rho^R c^R}{a^R} \leq \frac{1}{1 + \beta X^R}. \tag{4.99}$$

With these inequalities we rewrite (4.96) and (4.97)

$$\frac{1}{\rho^{L*}} \geq \frac{1}{\rho^L}\left(1 - \frac{X^L}{1 + \beta X^L}\right), \tag{4.100}$$

$$\frac{1}{\rho^{R*}} \geq \frac{1}{\rho^R}\left(1 - \frac{X^R}{1 + \beta X^R}\right). \tag{4.101}$$

Because of the definitions in (4.94) and (4.95) we know that $X^L, X^R \geq 0$ and therefore we can conclude that

$$\rho^{L*} > 0, \quad \rho^{R*} > 0. \tag{4.102}$$

□

4.3.4 Asymptotic-Preserving Property

In the low Mach number limit, the solutions of the compressible Euler equations tend to the solutions of the incompressible Euler equations (see Sect. 2.3.5). Following this theoretical result, the numerical scheme should be consistent with the limit behavior as $\mathcal{M}$ tends to zero, in the sense that the discretization for the compressible Euler equations should tend to the incompressible Euler equations if the Mach number tends to zero. The key to achieve this behavior for the presented relaxation scheme is the definition of the relaxation speeds a and b. In the former sections several conditions are imposed on these speeds that have to be satisfied so that the scheme is stable and has the properties presented. A suitable choice that indeed fulfills the so far stated requirements is the classical one, in which a and b are set to be equal

$$\begin{aligned} a_q^\alpha &= c^\alpha, \\ a^\alpha &= b^\alpha = \rho^\alpha c^\alpha (1 + \beta X^\alpha). \end{aligned} \tag{4.103}$$

This definition closely follows the condition (4.91) in Lem. 4.3.7. Unfortunately, this definition does not lead to an appropriate discretization, but to excessive diffusion in the low Mach number limit. In order to change this behaviour the speeds have to be redefined. In this context it is important to ensure that not only the diffusion is reduced, but also that the subcharacteristic conditions (4.6) remain fulfilled. A suitable choice proposed in [BCG20] is given by

$$\begin{aligned} a_q^\alpha &= \min(1, \mathcal{M}) c^\alpha, \\ a^\alpha &= \frac{1}{\min(1, \mathcal{M})} \rho^\alpha c^\alpha (1 + \beta X^\alpha), \\ b^\alpha &= \min(1, \mathcal{M}) \rho^\alpha c^\alpha (1 + \beta X^\alpha). \end{aligned} \tag{4.104}$$

By this definition the speeds are rescaled in the case of low Mach numbers, i.e. for $\mathcal{M} < 1$.

Remark 4.3.8. *In the case of Mach numbers $\mathcal{M} \geq 1$, the relaxation speeds are equal ($a = b$) and so we obtain the standard Suliciu relaxation system with only one relaxation speed.*

Remark 4.3.9. *The new scaling of the relaxation speed a has the effect that the maximum wave speed increases by an order of magnitude $\mathcal{M}$. As a consequence, the CFL condition (4.32) becomes stricter and the time step must be chosen smaller accordingly, i.e.*

$$\Delta t \sim \frac{\mathcal{M}^2 \Delta x}{c}. \tag{4.105}$$

As shown in [BCG20], by replacing $\mathcal{M}$ by $\widehat{\mathcal{M}} = \max\{\mathcal{M}^2, k\Delta x\}$ in the relaxation scheme, the CFL condition can be reduced to the parabolic-type condition

$$\Delta t \sim \frac{\max\{\mathcal{M}^2, k\Delta x\} \Delta x}{c}. \tag{4.106}$$

Theorem 4.3.10. *The two-speed relaxation scheme with the relaxation speeds (4.104) is asymptotic-preserving in the sense that:*

a) it is first-order uniformly with respect to the Mach number $\mathcal{M}$ and

b) for $\mathcal{M} < \sqrt{k\Delta x}$ and k constant it is consistent at first order with the incompressible limit model (2.60).

Proof. In order to prove the first statement of the theorem we evaluate the consistency error by expanding the numerical flux (4.28) in terms of $\mathcal{M}$ and then subtract the central flux $(F(Q^L) + F(Q^R))/2$.
In the low Mach number limit $\mathcal{M} \to 0$, the wave speeds λ^- and λ^+ in (4.7) tend towards infinity. Therefore it is sufficient just to consider the intermediate fluxes F^{L*} and F^{R*} for the numerical flux. In a first step of the analysis we rewrite the relaxation speeds as expansions in terms of $\mathcal{M}$, so we get

$$X^\alpha = \mathcal{O}(\mathcal{M}), \qquad b^\alpha = \mathcal{M}\bar{b}^\alpha + \mathcal{O}(\mathcal{M}^2), \qquad a^\alpha = \frac{\bar{b}^\alpha}{\mathcal{M}}(1 + \mathcal{O}(\mathcal{M})) \tag{4.107}$$

with

$$\bar{b}^\alpha = \rho^\alpha c^\alpha. \tag{4.108}$$

Since

$$\bar{b}^R - \bar{b}^L = \mathcal{O}(\mathcal{M}^2), \tag{4.109}$$

we can write $\bar{b}$ instead of $\bar{b}^L$ and $\bar{b}^R$ up to errors of $\mathcal{O}(\mathcal{M}^2)$. Expanding the intermediate states (4.16)-(4.22) in terms of $\mathcal{M}$ yields

$$\begin{aligned}
v^* &= \frac{u^L + u^R}{2} + \frac{\pi^L - \pi^R}{2\mathcal{M}^2\bar{b}} - \frac{\rho(Z^R - Z^L)}{2\mathcal{M}^2\bar{b}} \\
&\quad + \mathcal{O}(\mathcal{M}(u^L - u^R)) + \mathcal{O}(\frac{\pi^R - \pi^L + \bar{\rho}(Z^R - Z^L)}{\mathcal{M}}), \\
\pi^{L*} &= \frac{\pi^L + \pi^R}{2} + \mathcal{M}^2\bar{b}\frac{u^L - u^R}{2} + \frac{\bar{\rho}(Z^R - Z^L)}{2\bar{b}} \\
&\quad + \mathcal{O}(\mathcal{M}^3(u^L - u^R)) + \mathcal{O}(\mathcal{M}(\pi^R - \pi^L + \bar{\rho}(Z^R - Z^L))), \\
\pi^{R*} &= \frac{\pi^L + \pi^R}{2} + \mathcal{M}^2\bar{b}\frac{u^L - u^R}{2} - \frac{\bar{\rho}(Z^R - Z^L)}{2\bar{b}} \\
&\quad + \mathcal{O}(\mathcal{M}^3(u^L - u^R)) + \mathcal{O}(\mathcal{M}(\pi^R - \pi^L + \bar{\rho}(Z^R - Z^L))), \\
\frac{1}{\rho^{L*}} &= \frac{1}{\rho^L} + \mathcal{O}(\mathcal{M}^2(u^L - u^R)) + \mathcal{O}(\pi^R - \pi^L + \bar{\rho}(Z^R - Z^L)), \\
\frac{1}{\rho^{R*}} &= \frac{1}{\rho^R} + \mathcal{O}(\mathcal{M}^2(u^L - u^R)) + \mathcal{O}(\pi^R - \pi^L + \bar{\rho}(Z^R - Z^L)), \\
u^{L*} &= u^L + \mathcal{O}(\mathcal{M}^2(u^L - u^R)) + \mathcal{O}(\pi^R - \pi^L + \bar{\rho}(Z^R - Z^L)), \\
u^{R*} &= u^R + \mathcal{O}(\mathcal{M}^2(u^L - u^R)) + \mathcal{O}(\pi^R - \pi^L - \bar{\rho}(Z^R - Z^L)).
\end{aligned} \tag{4.110}$$

We can derive these expansions from the intermediate states (4.17)-(4.25) and put the terms $\pi^R - \pi^L + \bar{\rho}(Z^R - Z^L)$ into the error estimates, since, according to (2.59), the hydrostatic equilibrium is satisfied up to terms of order $\mathcal{O}(\mathcal{M}^2)$ in the low Mach limit, i.e.

$$\pi^R - \pi^L + \bar{\rho}(Z^R - Z^L) = \mathcal{O}(\mathcal{M}^2). \tag{4.111}$$

With the help of these expansions, we calculate the flux differences component by component.

i) The difference for the left intermediate flux F^{L*} in the first component writes

$$\begin{aligned}
&\rho^{L*}v^* - \frac{\rho^L u^L + \rho^R u^R}{2} \\
=& -\frac{\rho^L u^L + \rho^R u^R}{2} + \frac{\rho^L}{2\bar{b}}\left(\frac{p^L - p^R}{\mathcal{M}^2} + \frac{\bar{\rho}(\Phi^L - \Phi^R)}{\mathcal{M}^2}\right) \\
&+ \rho^L \frac{u^L + u^R}{2} + \mathcal{O}(\mathcal{M}(u^L - u^R)) + \mathcal{O}\left(\frac{p^R - p^L + \bar{\rho}(\Phi^R - \Phi^L)}{\mathcal{M}}\right).
\end{aligned}$$

This difference can be further simplified. In the low Mach limit, the density is constant up to errors of $\mathcal{O}(\mathcal{M}^2)$. Therefore we can write

$$\rho^R - \rho^L = \mathcal{O}(\mathcal{M}^2) \tag{4.112}$$

and replace ρ^R in the difference by ρ^L. Additionally, we replace the differences between the left and right states by numerical derivatives, i.e.

$$\begin{aligned}
u^L - u^R &= -\Delta x \partial_x u + \mathcal{O}(\Delta x^2), \\
p^L - p^R &= -\Delta x \partial_x p + \mathcal{O}(\Delta x^2), \\
\Phi^L - \Phi^R &= -\Delta x \partial_x \Phi + \mathcal{O}(\Delta x^2).
\end{aligned} \tag{4.113}$$

Applying these simplifications results in

$$\rho^{L*}v^* - \frac{\rho^L u^L + \rho^R u^R}{2} = -\frac{\Delta x}{2}\frac{\rho^L}{\bar{b}}\left(\frac{\partial_x p + \bar{\rho}\partial_x \Phi}{\mathcal{M}^2}\right) + \mathcal{O}(\Delta x^2) + \mathcal{O}(\mathcal{M}\Delta x). \tag{4.114}$$

The denominator $\mathcal{M}^2$ does not lead to excessive dissipation at this point, as again the hydrostatic equilibrium is fulfilled up to $\mathcal{O}(\mathcal{M}^2)$. Analogous calculations for the right intermediate flux F^{R*} lead to

$$\rho^{R*}v^* - \frac{\rho^L u^L + \rho^R u^R}{2} = -\frac{\Delta x}{2}\frac{\rho^R}{\bar{b}}\left(\frac{\partial_x p + \bar{\rho}\partial_x \Phi}{\mathcal{M}^2}\right) + \mathcal{O}(\Delta x^2) + \mathcal{O}(\mathcal{M}\Delta x). \tag{4.115}$$

ii) The second component for the left flux can be expressed by

$$\begin{aligned}
&\rho^{L*}u^{L*}v^* + \frac{\pi^{L*}}{\mathcal{M}^2} - \frac{\rho^L (u^L)^2 + \frac{\pi^L}{\mathcal{M}^2} + \rho^R (u^R)^2 + \frac{\pi^R}{\mathcal{M}^2}}{2} \\
=& \bar{b}\frac{u^L - u^R}{2} + \rho^L u^L \frac{u^L + u^R}{2} - \rho^L u^R \frac{u^L - u^R}{2} + \rho^L u^R \frac{u^L - u^R}{2} \\
&- \frac{\rho^L (u^L)^2 + \rho^R (u^R)^2}{2} + \rho^L u^L \frac{p^L - p^R + \bar{\rho}(\Phi^L - \Phi^R)}{2\bar{b}\mathcal{M}^2} \\
&- \frac{\bar{\rho}(\Phi^L - \Phi^R)}{2\mathcal{M}^2} + \mathcal{O}(\mathcal{M}(u^L - u^R)) + \mathcal{O}\left(\frac{p^R - p^L + \bar{\rho}(\Phi^R - \Phi^L)}{\mathcal{M}}\right) \\
=& \bar{b}\frac{u^L - u^R}{2} + \rho^L u^R \frac{u^L - u^R}{2} + \rho^L u^L \frac{p^L - p^R + \bar{\rho}(\Phi^L - \Phi^R)}{2\bar{b}\mathcal{M}^2} \\
&- \frac{\bar{\rho}(\Phi^L - \Phi^R)}{2\mathcal{M}^2} + \mathcal{O}(\mathcal{M}(u^L - u^R)) + \mathcal{O}\left(\frac{p^R - p^L + \bar{\rho}(\Phi^R - \Phi^L)}{\mathcal{M}}\right) \\
=& -\frac{\Delta x}{2}\left(\bar{b} + \rho^L u^R\right)\partial_x u - \frac{\Delta x}{2}\frac{\rho^L u^L}{\bar{b}}\left(\frac{\partial_x p + \bar{\rho}\partial_x \Phi}{\mathcal{M}^2}\right) + \frac{\Delta x}{2}\bar{\rho}\partial_x \frac{\Phi}{\mathcal{M}^2} \\
&+ \mathcal{O}(\Delta x^2) + \mathcal{O}(\mathcal{M}\Delta x)
\end{aligned}$$

and for the right flux by

$$\begin{aligned}
&\rho^{R*}u^{R*}v^* + \frac{\pi^{R*}}{\mathcal{M}^2} - \frac{\rho^L(u^L)^2 + \frac{\pi^L}{\mathcal{M}^2} + \rho^R(u^R)^2 + \frac{\pi^R}{\mathcal{M}^2}}{2} \\
&= -\frac{\Delta x}{2}\left(\bar{b} + \rho^R u^L\right)\partial_x u - \frac{\Delta x}{2}\frac{\rho^R u^R}{\bar{b}}\left(\frac{\partial_x p + \bar{\rho}\partial_x \Phi}{\mathcal{M}^2}\right) - \frac{\Delta x}{2}\bar{\rho}\partial_x \frac{\Phi}{\mathcal{M}^2} \\
&\quad + \mathcal{O}(\Delta x^2) + \mathcal{O}(\mathcal{M}\Delta x).
\end{aligned}$$

In this flux difference, the new scaling of the relaxation speeds defined in (4.104) unfolds its importance. Clearly, the viscosity on the velocity, represented by the first term, is independent of the Mach number and therefore does not increase in the low Mach limit. With the classical scaling (4.103), on the other hand, this term would have the size $\mathcal{O}(1/\mathcal{M})$ leading to excessive dissipation for low Mach numbers. While a Mach number dependence in the first term would be problematic, it is not in the second term due to (4.111). The remaining third term containing the derivative of the gravitational potential, which also depends on $1/\mathcal{M}^2$, cancels out with the gravitational source term (4.30) in the relaxation scheme.

iii) For the difference in the third component, similar steps for the left flux result in

$$\begin{aligned}
&\left(\left(\frac{1}{2}\mathcal{M}^2\rho^{L*}(u^{L*})^2 + \rho^{L*}e^{L*}\right) + \pi^{L*}\right)v^* - \frac{(E^L + p^L)u^L + (E^R + p^R)u^R}{2} \\
&= \rho^L u^R \frac{e^L - e^R}{2} + u^R \frac{p^L - p^R}{2} + \frac{\rho^L e^L + p^L}{2\bar{b}}\frac{p^L - p^R + \bar{\rho}(\Phi^L - \Phi^R)}{\mathcal{M}^2} \\
&\quad + \mathcal{O}(\mathcal{M}(u^L - u^R)) + \mathcal{O}\left(\frac{p^R - p^L + \bar{\rho}(\Phi^R - \Phi^L)}{\mathcal{M}}\right) \\
&= -\frac{\Delta x}{2}\rho^L u^R \partial_x e - \frac{\Delta x}{2}u^R\partial_x p - \frac{\Delta x}{2}\frac{\rho^L e^L + p^L}{\bar{b}}\left(\frac{\partial_x p + \bar{\rho}\partial_x\Phi}{\mathcal{M}^2}\right) \\
&\quad + \mathcal{O}(\Delta x^2) + \mathcal{O}(\mathcal{M}\Delta x).
\end{aligned}$$

and for the right flux in

$$\begin{aligned}
&\left(\left(\frac{1}{2}\mathcal{M}^2\rho^{R*}(u^{R*})^2 + \rho^{R*}e^{R*}\right) + \pi^{R*}\right)v^* - \frac{(E^L + p^L)u^L + (E^R + p^R)u^R}{2} \\
&= \frac{\Delta x}{2}\rho^R u^L \partial_x e + \frac{\Delta x}{2}u^L\partial_x p - \frac{\Delta x}{2}\frac{\rho^R e^R + p^R}{\bar{b}}\left(\frac{\partial_x p + \bar{\rho}\partial_x\Phi}{\mathcal{M}^2}\right) \\
&\quad + \mathcal{O}(\Delta x^2) + \mathcal{O}(\mathcal{M}\Delta x).
\end{aligned}$$

The expansions for all three components are first-order uniformly in $\mathcal{M}$. It is particularly important that the viscosity on the velocity u in the momentum flux is independent of $\mathcal{M}$.

Remark 4.3.11. *The rescaling of $b^{L,R}$ by the Mach number has not only an effect on the scaling of the intermediate pressures $\pi^{L*,R*}$, where it reduces the dissipation in the low Mach number regime. It also adds a $\mathcal{M}$ in the denominator of the intermediate velocity v^*. Thereby it increases the artificial dissipation there. However, it only acts on the term $\pi^L - \pi^R - \bar{\rho}(W^L, W^R)(Z^R - Z^L)$, which scales with $\mathcal{O}(\mathcal{M}^2)$. Therefore, it does not lead to an increasing dissipation for decreasing Mach numbers. This additional change of scaling in v^* poses a difference to other low Mach fixes, which only concentrate on reducing the dissipation in the intermediate pressure state.*

The result of the first statement can now be used to prove the second statement of the theorem. We have proven that the solution $Q_{\mathcal{M},\Delta x}$ of the relaxation scheme is consistent with the exact solution $Q_{\mathcal{M}}$ of the dimensionless Euler equations (2.50) up to order $\mathcal{O}(\Delta x)$ independent of the Mach number, i.e.

$$Q_{\mathcal{M},\Delta x} - Q_{\mathcal{M}} = \mathcal{O}(\Delta x). \tag{4.116}$$

Additionally, we can deduce from system (2.64) that $Q_{\mathcal{M}}$ is consistent with the solution Q of the incompressible Euler equations up to order $\mathcal{O}(\mathcal{M}^2)$, i.e.

$$Q_{\mathcal{M}} - Q = \mathcal{O}(\mathcal{M}^2). \tag{4.117}$$

Combining (4.116) and (4.117) with the condition $\mathcal{M}^2 = \mathcal{O}(\Delta x)$ finally results in

$$Q_{\mathcal{M},\Delta x} - Q = \mathcal{O}(\Delta x) \tag{4.118}$$

and therefore meets the second statement of the theorem. □

4.3.5 Well-Balanced Property

As described in Sect. 3.10.2, the well-balanced property is important for solving problems close to hydrostatic equilibria. In a first step, we will show that the approximate Riemann solver satisfies this property. Building on this result, we will then prove in the second step that the entire scheme has this property.

Lemma 4.3.12. *Assume two given states at equilibrium W^L and W^R satisfy*

$$u^L = u^R = 0, \tag{4.119}$$
$$p^R - p^L + \bar{\rho}(W^L, W^R)(\Phi^R - \Phi^L) = 0. \tag{4.120}$$

Then the approximate Riemann solver $\mathcal{W}_{\mathcal{R}}$ in (4.14) preserves the steady state, i.e.

$$\mathcal{W}_{\mathcal{R}}(\frac{x}{t}, W^L, W^R) = \begin{cases} W^L, & \text{if } \frac{x}{t} < 0, \\ W^R, & \text{if } \frac{x}{t} > 0. \end{cases} \tag{4.121}$$

Proof. The result directly follows from the definition of the intermediate states given in (4.16)-(4.25). Consider the intermediate state v^*. Since we start at equilibrium, we can replace the relaxation variables by their corresponding original variables. Using the conditions (4.119)-(4.120) results in

$$v^* = \frac{1}{b^L + b^R}\left(\mathcal{M}b^L u^L + \mathcal{M}b^R u^R + p^L - p^R - \bar{\rho}\left(W^L, W^R\right)\left(\Phi^R - \Phi^L\right)\right) = 0.$$

Similar calculations for the other intermediate states complete the proof. □

In this proof the importance of the closure relation (4.15) becomes evident. The newly introduced terms $\bar{\rho}(W^L, W^R)(\Phi^R - \Phi^L)$ in the intermediate states are essential for the approximate Riemann solver preserving steady states, as they cancel out with the pressure difference terms. Otherwise the intermediate states for the velocities would not be zero.

Lem. 4.3.12 is rather general, as it assumes that the conditions in (4.119) and (4.120) are satisfied. Clearly, these conditions depend on the definition of the $\bar{\rho}$-function. For a simple

definition like the arithmetic mean, which is not adjusted to the underlying hydrostatic equilibrium, the scheme maintains the equilibrium to second order [DZBK16]. Since we are free to define $\bar{\rho}$ we can adjust it to the hydrostatic equilibrium and maintain it even up to machine precision. The only limiting requirement for $\bar{\rho}$ that has to be considered is the consistency property

$$\rho^L = \rho^R = \rho \quad \Rightarrow \quad \bar{\rho}(W^L, W^R) = \rho. \tag{4.122}$$

The following lemma describes the adjusted definitions for isothermal, incompressible and polytropic equilibria. These definitions have already been described in [DZBK16].

Lemma 4.3.13. *The approximate Riemann solver $\mathcal{W}_{\mathcal{R}}$ can exactly preserve the following families of hydrostatic equilibria:*

i) Let W^L and W^R be two states satisfying an isothermal equilibrium of the form (2.47). If the function $\bar{\rho}$ is defined by

$$\bar{\rho}(W^L, W^R) = \begin{cases} \frac{\rho^R - \rho^L}{\ln(\rho^R) - \ln(\rho^L)}, & \text{if } \rho^L \neq \rho^R, \\ \rho^L, & \text{if } \rho^L = \rho^R, \end{cases} \tag{4.123}$$

then the approximate Riemann solver $\mathcal{W}_{\mathcal{R}}$ preserves the steady state.

ii) Let W^L and W^R be two states satisfying a polytropic equilibrium of the form (2.48). If the function $\bar{\rho}$ is defined by

$$\bar{\rho}(W^L, W^R) = \begin{cases} \frac{\Gamma-1}{\Gamma} \frac{(\rho^R)^\Gamma - (\rho^L)^\Gamma}{(\rho^R)^{\Gamma-1} - (\rho^L)^{\Gamma-1}}, & \text{if } \rho^L \neq \rho^R, \\ \rho^L, & \text{if } \rho^L = \rho^R, \end{cases} \tag{4.124}$$

then the approximate Riemann solver $\mathcal{W}_{\mathcal{R}}$ preserves the steady state.

iii) Let W^L and W^R be two states satisfying an incompressible equilibrium of the form (2.49). If the function $\bar{\rho}$ satisfies the consistency condition (4.122), then the approximate Riemann solver $\mathcal{W}_{\mathcal{R}}$ preserves the steady state.

Proof. In order to prove this lemma it is sufficient to show that with the explicit definition of $\bar{\rho}$ the conditions (4.119) and (4.120) are satisfied. If so, we can use Lem. 4.3.12 and the proof is complete. Using the definitions of the isothermal equilibrium states, we can determine the following differences

$$\begin{aligned} \Phi^R - \Phi^L &= \chi(\ln(\rho^R) - \ln(\rho^L)), \\ p^R - p^L &= \chi(\rho^R - \rho^L). \end{aligned}$$

By inserting these differences together with $\bar{\rho}$ defined by (4.123) into equation (4.120), it becomes clear that this condition is satisfied. Together with the velocities, which are zero, Lem. 4.3.12 can be applied and the proof of *i)* is complete. The proofs for polytropic and incompressible equilibria work in the same way. For more details we may refer the reader to [DZBK16]. □

In practical applications, e.g. in astrophysics, the hydrostatic states are often only available as discrete data generated by previously performed simulations. The following lemma provides an approach to maintain these hydrostatic equilibria as well.

Lemma 4.3.14. *Let W^L and W^R be two states satisfying some hydrostatic equilibrium*

$$\begin{cases} u^L = u^R = 0, \\ \rho^{L,R} = \rho_{hs}^{L,R}, \\ p^{L,R} = p_{hs}^{L,R}, \end{cases} \tag{4.125}$$

with ρ_{hs} and p_{hs} given hydrostatic states. If the function $\bar{\rho}$ is defined by

$$\bar{\rho}(W^L, W^R) = \frac{1}{2}(\rho^L + \rho^R) \tag{4.126}$$

and the difference of the gravitational potential in the intermediate states is approximated by

$$Z^R - Z^L \approx -\frac{p_{hs}^R - p_{hs}^L}{\frac{1}{2}(\rho_{hs}^L + \rho_{hs}^R)}, \tag{4.127}$$

then the approximate Riemann solver $\mathcal{W}_{\mathcal{R}}$ preserves the steady state.

Proof. As can be seen in the proof of Lem. 4.3.13 it is sufficient to show that the conditions (4.119) and (4.120) are fulfilled so that Lem. 4.3.12 can be applied. In order to do so we insert the states from (4.125) and the approximation (4.127) in (4.120) and use definition (4.126) for $\bar{\rho}$. This results in

$$p_{hs}^R - p_{hs}^L - \frac{1}{2}(\rho_{hs}^L + \rho_{hs}^R)\frac{p_{hs}^R - p_{hs}^L}{\frac{1}{2}(\rho_{hs}^L + \rho_{hs}^R)} = 0. \tag{4.128}$$

□

Remark 4.3.15. *The technique used in Lem. 4.3.14 is related to the α-β method presented in [BCK18]. Both methods approximate the gravitational potential by given hydrostatic states. The difference, however, is that here the approximation is done within the Riemann solver so that the solver stays at rest in equilibrium. In the α-β method the approximation is applied in the numerical source term and combined with a hydrostatic reconstruction.*

Having shown that the approximate Riemann solver $\mathcal{W}_{\mathcal{R}}$ satisfies the well-balanced property, it remains to show that the entire scheme does so as well.

Theorem 4.3.16. *Let us consider cell averages at time level n denoted by Q_{i-1}^n, Q_i^n and assume that for all $i \in \mathbb{Q}$ they satisfy*

$$u_i^n = 0, \tag{4.129}$$

$$\frac{1}{\Delta x}(p_i^n - p_{i-1}^n) + \bar{\rho}(W_{i-1}^n, W_i^n)\frac{\Phi_i - \Phi_{i-1}}{\Delta x} = 0. \tag{4.130}$$

Then the updated state satisfies $Q_i^{n+1} = Q_i^n$ for all $i \in \mathbb{Q}$ and thus the solution stays at rest.

Proof. Since both conditions (4.119) and (4.120) of Lem. 4.3.12 are fulfilled, the approximate Riemann solver stays at rest. In consequence, the Riemann solver returns $W^{L*} = W^L$. Thus, $v^* = 0$ and the numerical flux in (4.28) is chosen to be $F(W^{L*}) = \mathcal{F}(W^L)$. Only the momentum component of the flux is nonzero. Here, the flux contains only the pressure contribution. The source term on the right-hand side is also zero except for the momentum component. The $\mathfrak{s}$ function ensures that it is equal to $S_{i-1/2}^+$.Therefore the update in the momentum equation reads

$$(\rho u)_i^{n+1} = (\rho u)_i^n - \frac{\Delta t}{\Delta x}\left(p_i^n - p_{i-1}^n\right) - \frac{\Delta t}{2}\left(2\bar{\rho}\left(W_{i-1}^n, W_i^n\right)\frac{\Phi_i - \Phi_{i-1}}{\Delta x}\right) \tag{4.131}$$

and with (4.130) the residual in the momentum becomes zero. This means that all residuals are zero and we get the statement of the theorem. □

4.4 Second Order Extension

In this section we give a possible extension of the proposed scheme to second order in space. As a basic strategy we apply a linear reconstruction. However, if we took the standard linear reconstruction described in Sect. 3.8, the scheme would lose the positivity and well-balanced properties.
To ensure that the scheme remains positivity-preserving, we rely on an approach introduced in [TZK19] that builds on the work in [Ber05]. Instead of reconstructing the conservative variables $\mathcal{Q}$, we reconstruct the primitive variables $\mathcal{Q}^P = (\rho, \mathbf{v}, p)$ and therefore evaluate the function

$$Q^P(x) = Q_i^P + \sigma_i(x - x_i) \tag{4.132}$$

in each cell C_i at its boundaries $x_{i-1/2}$ and $x_{i+1/2}$ to obtain initial values of the Riemann problem denoted by $Q_{i-1/2}^{P,R}$ and $Q_{i+1/2}^{P,L}$. The slope σ is computed for each primitive variable separately and is defined by

$$\sigma_i^\rho = \rho_i \max\left(-1, \min\left(1, \frac{\bar{\sigma}_i^\rho}{\rho_i}\right)\right), \tag{4.133}$$

$$\boldsymbol{\sigma}_i^{\mathbf{v}} = \kappa \bar{\boldsymbol{\sigma}}_i^{\mathbf{v}}, \tag{4.134}$$

$$\sigma_i^p = p_i \max\left(-1, \min\left(1, \frac{\bar{\sigma}_i^p}{p_i}\right)\right), \tag{4.135}$$

with

$$\bar{\sigma}_i = \operatorname{minmod}\left(\frac{Q_i^P - Q_{i-1}^P}{\Delta x}, \frac{Q_{i+1}^P - Q_i^P}{\Delta x}\right) \tag{4.136}$$

and

$$\kappa_i = \min(1, \bar{\kappa}_i), \tag{4.137}$$

$$\bar{\kappa}_i = \begin{cases} \frac{-\sigma_i^\rho(\mathbf{v}_i \cdot \bar{\boldsymbol{\sigma}}_i^{\mathbf{v}}) + \sqrt{(\sigma_i^\rho)^2(\mathbf{v}_i \cdot \bar{\boldsymbol{\sigma}}_i^{\mathbf{v}})^2 + \|\bar{\boldsymbol{\sigma}}_i^{\mathbf{v}}\|^2 \frac{\rho_i p_i}{\gamma - 1}}}{\rho_i \|\bar{\boldsymbol{\sigma}}_i^{\mathbf{v}}\|^2}, & \text{if } \bar{\boldsymbol{\sigma}}_i^{\mathbf{v}} \neq 0, \\ 1, & \text{if } \bar{\boldsymbol{\sigma}}_i^{\mathbf{v}} = 0. \end{cases} \tag{4.138}$$

The choice of these slopes provably ensures the positivity of the reconstructed values that serve as initial data for the Riemann problems [TZK19]. This means that the requirements for applying Lem. 4.3.6 and 4.3.7 are met, which ensure that $Q_i^{n+1} \in \Omega_{\text{phys}}^{\text{Euler}}$.
Additionally, we also want to preserve the well-balanced property for the second-order scheme. To achieve this, we adjust the pressure slope by using a hydrostatic reconstruction [KM16, TZK19, TPK20]. Instead of directly using the pressure values of the neighboring cells, one first applies the transformations

$$\begin{aligned} q_{i-1} &= p_{i-1} - \bar{\rho}(W_{i-1}, W_i)(\Phi_i - \Phi_{i-1}), \\ q_{i+1} &= p_{i+1} + \bar{\rho}(W_i, W_{i+1})(\Phi_{i+1} - \Phi_i), \end{aligned} \tag{4.139}$$

and then computes the slope for the pressure by

$$\bar{\sigma}_i^p = \operatorname{minmod}\left(\frac{p_i - q_{i-1}}{\Delta x}, \frac{q_{i+1} - p_i}{\Delta x}\right). \tag{4.140}$$

In the case of a hydrostatic equilibrium, the slope becomes zero and the interface values for the pressure thus reduce to the cell averages. The approximate Riemann solver then stays at rest due to Lem. 4.3.12 and all results of Sect. 4.3.5 about well-balancing remain valid for the second order scheme.

4.5 Multi-Dimensional Extension

For two spatial dimensions the Euler equations (2.50) can be written in the form

$$\mathcal{Q}_t + \mathcal{F}_1(\mathcal{Q})_x + \mathcal{F}_2(\mathcal{Q})_y = \mathcal{S}(\mathcal{Q}, \Phi). \tag{4.141}$$

On a regular Cartesian grid, we extend the numerical scheme described in Sect. 4.2 to two spatial dimensions by applying an unsplit finite volume method as described in Sect. 3.7. The contributions of both directions are used in only one step to update the numerical solution by the formula

$$\begin{aligned} Q_{i,j}^{n+1} = Q_{i,j}^n &- \frac{\Delta t}{\Delta x}\left(F_{1,i+1/2,j}^n - F_{1,i-1/2,j}^n\right) - \frac{\Delta t}{\Delta y}\left(F_{2,i,j+1/2}^n - F_{2,i,j-1/2}^n\right) \\ &+ \frac{\Delta t}{2}\left(S_{i-1/2,j}^{+,n}\frac{\Phi_{i,j}^n - \Phi_{i-1,j}^n}{\Delta x} + S_{i+1/2,j}^{-,n}\frac{\Phi_{i+1,j}^n - \Phi_{i,j}^n}{\Delta x}\right) \\ &+ \frac{\Delta t}{2}\left(S_{i,j-1/2}^{+,n}\frac{\Phi_{i,j}^n - \Phi_{i,j-1}^n}{\Delta y} + S_{i,j+1/2}^{-,n}\frac{\Phi_{i,j+1}^n - \Phi_{i,j}^n}{\Delta y}\right). \end{aligned} \tag{4.142}$$

The definitions of the numerical fluxes and source terms are straightforward extensions of those presented in Sect. 4.2. The numerical fluxes continue to use a one-dimensional approximate Riemann solver so that it is applied separately in x- and y-direction. This Riemann solver corresponds to the one defined in (4.14), in which additionally the intermediate states for the transversal velocity are set by the left and right values at the interface, respectively, since this velocity is a Riemann invariant for the outer waves λ^- and λ^+.
The two-dimensional method relies on the one-dimensional approximate Riemann solver so that the properties proven in Sect. 4.3 for the solver also apply to this method. So we gain the entropy inequality, the absence of checkerboard modes, the positivity of density and internal energy and the asymptotic-preserving property. In addition, the well-balanced property is also preserved, since the approximate Riemann solver is at rest for initial data in hydrostatic equilibrium in both spatial directions and thus in both momentum equations the pressure gradient cancels out with the source term.

4.6 Numerical Results

In this section we numerically investigate the theoretical properties of the *two-speed well-balanced finite volume* (2S-WB-FV) scheme presented in the previous sections. The approximate Riemann solver in the scheme is equipped with the intermediate states defined in (4.16)-(4.25) and the Mach number adjusted relaxation speeds (4.104) with $\beta = 1.1$. Various definitions are used for the $\bar{\rho}$-function. Definition (4.123) is used by default. If a different choice is made, this is indicated in the respective test. The second order spatial scheme is combined with the third order SSP-RK3 method [SO88] for time integration (see Ex. 3.9.2).

We perform eight different numerical tests and start with a convergence study to verify the order of convergence of the method in different Mach number regimes (i). The second test consists of a Sod shock tube with an added gravitational force in order to demonstrate that the scheme can also capture shocks (ii). A strong rarefaction test challenges the ability of the scheme to preserve positive density and internal energy (iii). After that the scheme is applied to different types of hydrostatic equilibria (iv,v) und flows close to

Quantity	SI unit	Scaling
x, y	[m]	x_r
t	[s]	t_r
ρ	$[\frac{\mathrm{kg}}{\mathrm{m}^3}]$	ρ_r
u, c	$[\frac{\mathrm{m}}{\mathrm{s}}]$	$u_r = \frac{x_r}{t_r}$, $\mathcal{M} = \frac{u_r}{c_r}$
p	$[\frac{\mathrm{kg}}{\mathrm{ms}^2}]$	$p_r = \rho_r c_r^2$
Φ	$[\frac{\mathrm{m}^2}{\mathrm{s}^2}]$	$\Phi_r = \frac{u_r^2}{\mathcal{M}^2}$

Table 4.1: Relations between physical quantities, SI units and reference values used in Sect. 4.6.

equilibria (vi) in order to check the well-balanced property and its effect on the simulated solutions. The last two test problems are set up in the low Mach number regime. First, we consider a Kelvin-Helmholtz instability for the homogeneous Euler equations and perform a qualitative comparison between the two-speed relaxation scheme and a one-speed alternative (vii). Second, we simulate a stationary vortex in a gravitational field and do a quantitive measurement of the dissipation introduced by the method in the case of low Mach numbers (viii).

The initial data for the different test problems is given in physical variables and transformed to dimensionless quantities by reference values. The relations between physical variables, *international system of units* (SI) and reference values are given in Tab. 4.1. For all test setups we assume an ideal gas law $p = (\gamma - 1)\rho e$. The computations are performed on a regular Cartesian grid.

4.6.1 Convergence Test

In a first numerical test, which is suggested by [XS13], we investigate the experimental order of convergence of the relaxation scheme presented. For the Euler equations (2.50) on the domain $\mathcal{I} = [0,1]^2$ with a linear gravitational potential $\Phi(x,y) = x + y\,\frac{\mathrm{m}^2}{\mathrm{s}^2}$, one possible exact solution is defined by

$$\begin{aligned}
\rho(x,y,t) &= 1 + 0.2\sin\left(\pi\left(x + y - t(u_0 + v_0)\right)\right)\,\frac{\mathrm{kg}}{\mathrm{m}^3},\\
\mathbf{v}(x,y,t) &= (u_0, v_0)\,\frac{\mathrm{m}}{\mathrm{s}},\\
p(x,y,t) &= 4.5 + (u_0 + v_0)t - (x + y) + 0.2\cos\left(\left(\pi\left(x + y - (u_0 + v_0)t\right)\right)/\pi\right)\frac{\mathrm{kg}}{\mathrm{ms}^2},
\end{aligned} \tag{4.143}$$

with $u_0 = v_0 = 20$. This exact solution is also used for the boundary conditions. The adiabatic coefficient in the EoS is set to $\gamma = 5/3$. We solve the equations in different regimes and therefore transform the initial data into dimensionless quantities using the reference values

$$x_r = 1\,\mathrm{m}, \quad u_r = 1\,\frac{\mathrm{m}}{\mathrm{s}}, \quad \rho_r = 1\,\frac{\mathrm{kg}}{\mathrm{m}^3}, \quad p_r = \frac{1}{\mathcal{M}^2}\,\frac{\mathrm{kg}}{\mathrm{ms}^2}, \quad \Phi_r = \frac{1}{\mathcal{M}^2}\,\frac{\mathrm{m}^2}{\mathrm{s}^2}. \tag{4.144}$$

The numerical solution computed with the 2S-WB-FV scheme is computed on a $N \times N$ grid and compared to the exact solution at final time $t_f = 0.01\,\mathrm{s}$. The resulting L^1-error and the *experimental order of convergence* (EOC) can be found in Table 4.2. As expected, we obtain orders of convergence of nearly 2.0 independently of the Mach number regime. Without the use of limiters full second order is reached.

$\mathcal{M}$	N	$L^1(\rho)$	$EOC(\rho)$	$L^1(\rho u)$	$EOC(\rho u)$	$L^1(\rho v)$	$EOC(\rho v)$	$L^1(E)$	$EOC(E)$
1	32	7.26E-04	-	1.45E-02	-	1.45E-02	-	2.90E-01	-
	64	1.97E-04	1.88	3.93E-03	1.88	3.93E-03	1.88	7.87E-02	1.88
	128	5.22E-05	1.92	1.04E-03	1.92	1.04E-03	1.92	2.08E-02	1.92
	256	1.37E-05	1.92	2.73E-04	1.93	2.73E-04	1.93	5.47E-03	1.93
10^{-1}	32	7.29E-04	-	1.46E-02	-	1.46E-02	-	2.92E-01	-
	64	1.98E-04	1.88	3.94E-03	1.89	3.94E-03	1.89	7.90E-02	1.88
	128	5.24E-05	1.91	1.05E-03	1.92	1.05E-03	1.92	2.09E-02	1.92
	256	1.38E-05	1.92	2.75E-04	1.93	2.75E-04	1.93	5.51E-03	1.93
10^{-2}	32	7.34E-04	-	1.47E-02	-	1.47E-02	-	2.94E-01	-
	64	2.00E-04	1.87	4.01E-03	1.87	4.01E-03	1.87	8.03E-02	1.87
	128	5.40E-05	1.90	1.08E-03	1.90	1.08E-03	1.90	2.15E-02	1.90
	256	1.45E-05	1.90	2.87E-04	1.91	2.87E-04	1.91	5.74E-03	1.91
10^{-3}	32	7.32E-04	-	1.46E-02	-	1.46E-02	-	2.93E-01	-
	64	2.06E-04	1.83	4.12E-03	1.83	4.12E-03	1.83	8.24E-02	1.83
	128	5.77E-05	1.84	1.15E-03	1.84	1.15E-03	1.84	2.30E-02	1.84
	256	1.59E-05	1.86	3.17E-04	1.86	3.17E-04	1.86	6.31E-03	1.86

Table 4.2: L^1-errors and experimental orders of convergence.

4.6.2 Shock Tube under Gravitational Field

The next test case is the standard Sod shock tube for the one-dimensional Euler equations, to which a gravitational source term with linear gravitational potential $\Phi(x) = x$ is added [CK15]. The initial conditions in the domain $\mathcal{I} = [0, 1]$ are given by

$$(\rho, u, p)(x, 0) = \begin{cases} (1, 0, 1), & \text{if } x \leq 0.5, \\ (0.125, 0, 0.1), & \text{if } x > 0.5, \end{cases} \tag{4.145}$$

with solid wall boundary conditions. We choose a compressible regime and therefore set $\mathcal{M} = 1$. The ratio of specific heats is $\gamma = 1.4$. The solution at final time $t_f = 0.2$ is computed by the 2S-WB-FV scheme on 100 cells. The numerical solution is compared to a reference solution, which is computed by a fully explicit second order finite volume method on 20 000 cells. The results in Fig. 4.2 show a good agreement with the reference solution and are also consistent with solutions in the literature [CK15]. This test demonstrates the capability of the relaxation scheme to deal with shocks, i.e. flows which are not in the low Mach number regime.

4.6.3 Strong Rarefaction Test

The second order relaxation scheme shall preserve the positivity of density and internal energy if the relaxation speeds are chosen properly. The following 1-2-0-3 strong rarefaction test is challenging, as density and pressure become very small [TPK20]. In this test setup, two rarefaction waves are launched in x-direction on top of an isothermal atmosphere. Therefore, on the domain $\mathcal{I} = [0, 1]^2$ the density ρ and pressure p are initially defined by (2.47) with the constants $C = -0.01$ and $\chi = \gamma - 1$, an adiabatic coefficient $\gamma = 1.4$ and a quadratic gravitational potential $\Phi(x, y) = \frac{1}{2}[(x - 0.5)^2 + (y - 0.5)^2]$. The initial velocities are set to

$$u(x, y, 0) = \begin{cases} -2, & \text{if } x < 0.5, \\ 2, & \text{if } x \geq 0.5, \end{cases} \quad \text{and} \quad v(x, y, 0) = 0. \tag{4.146}$$

The reference Mach number is set to $\mathcal{M} = 1$ so that the setup is in the compressible regime. One slice along the x-axis of the numerical solution at final time $t_f = 0.1$ computed on a 128×128 grid by our relaxation scheme is presented in Fig. 4.3. Even though the values for

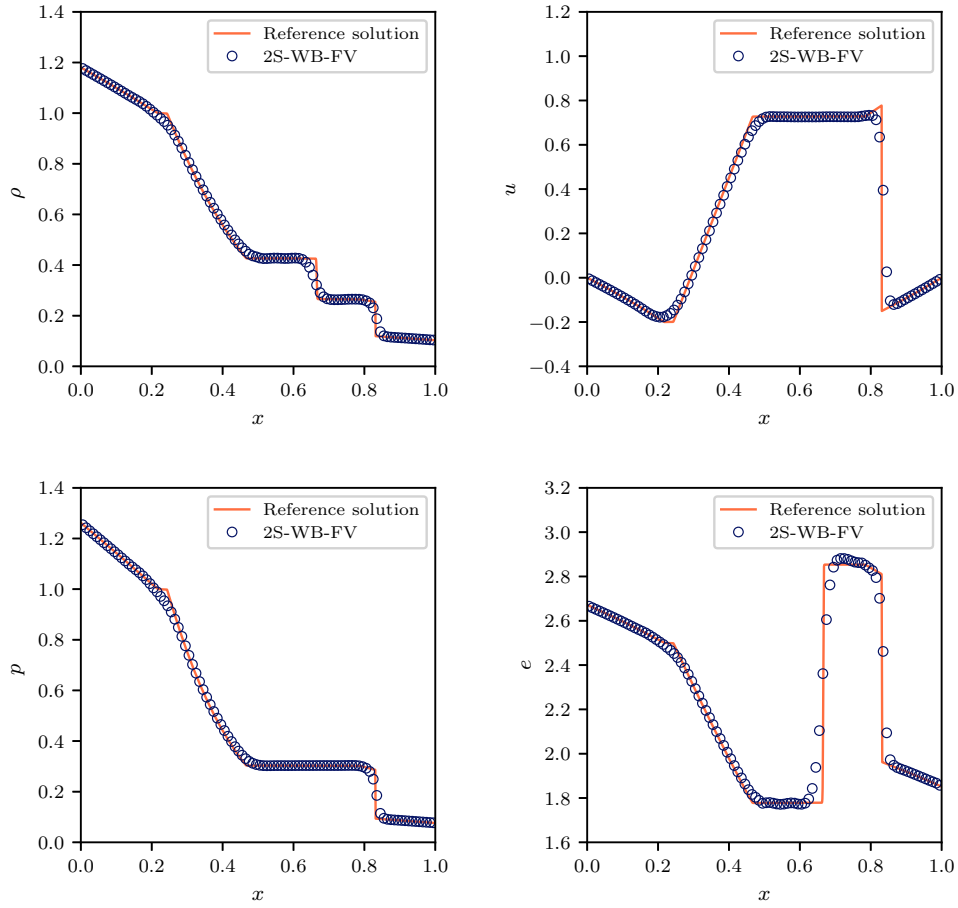

Figure 4.2: Shock tube under gravitational field at time $t_f = 0.2$. Comparison to the reference solution for density (top left), horizontal velocity (top right), pressure (bottom left) and internal energy (bottom right).

density and total pressure become very small during the simulation, they always remain positive. This outcome underlines the theoretical results stated in Lem. 4.3.6 and 4.3.7. Overall, the results of the 2S-WB-FV scheme agree with those presented in the literature [TPK20].

4.6.4 Isothermal Atmosphere

The relaxation scheme equipped with the $\bar{\rho}$-average (4.123) is designed to exactly preserve isothermal equilibria. The following initial data is taken from [CK15] and fulfills the isothermal equilibrium. On the domain $\mathcal{I} = [0, 1]^2$, we consider the gravitational potential

$$\Phi(x, y) = x + y \tag{4.147}$$

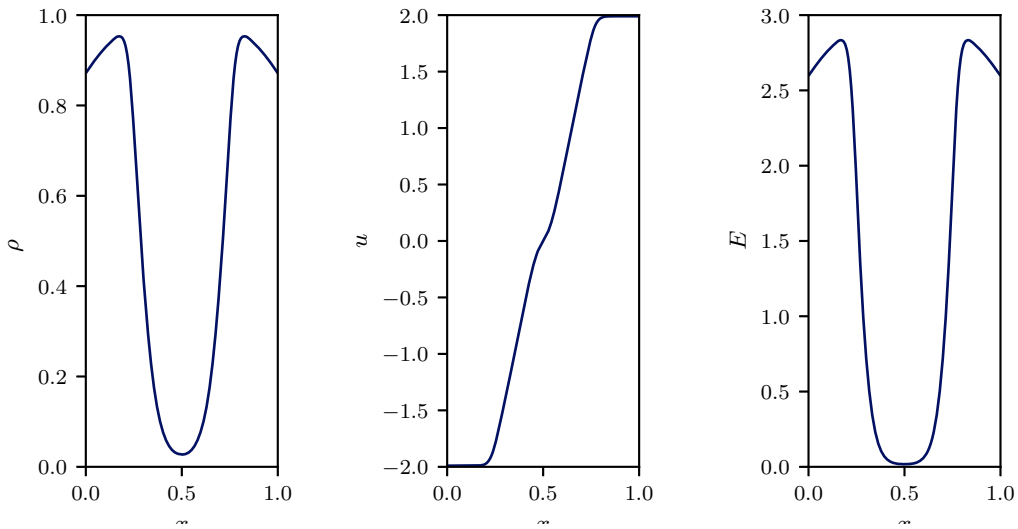

Figure 4.3: Numerical solution for density ρ, velocity u and total energy E at final time $t_f = 0.1$.

and initial conditions

$$\begin{aligned} \rho(x, y, 0) &= \rho_0 \exp(-\rho_0 g(x + y)/p_0), \\ \mathbf{v}(x, y, 0) &= \mathbf{0}, \\ p(x, y, 0) &= p_0 \exp(-\rho_0 g(x + y)/p_0), \end{aligned} \tag{4.148}$$

with the parameters $\rho_0 = 1.21$, $p_0 = 1$ and $g = 1$. The problem is solved in the compressible regime with $\mathcal{M} = 1$ and the adiabatic coefficient is set to $\gamma = 1.4$. The solution should be preserved up to any final time. Here we choose $t_f = 1$. The L^1-error between the approximated solution and the exact solution is given in Table 4.3. It is in the order of magnitude of the machine accuracy, which underlines that the 2S-WB-FV scheme is well-balanced for isothermal equilibria.

N	$L^1(\rho)$	$L^1(\rho u)$	$L^1(\rho v)$	$L^1(E)$
32	8.95E-17	5.21E-16	5.21E-16	4.18E-16
64	1.73E-16	1.62E-16	1.62E-16	7.24E-16
128	3.40E-16	3.47E-16	3.47E-16	1.63E-15
256	6.30E-16	6.89E-16	6.89E-16	3.46E-15
512	1.22E-15	1.54E-15	1.54E-15	7.43E-15

Table 4.3: L^1-errors for an isothermal atmosphere.

4.6.5 General Steady State

In practice, steady states at rest do not always belong to the class of isothermal, polytropic or incompressible equilibria. In order to investigate the behaviour of the well-balancing mechanism for such cases, we now apply the scheme to a general steady state. We take the initial conditions from the setup in Sect. 4.6.1 with $\mathcal{M} = 1$ and set the initial velocities u_0 and v_0 to zero. It is easy to check that the initial data then poses a hydrostatic equilibrium.

In a first step, we use the $\bar{\rho}$-average tuned to isothermal equilibria given by (4.123) and compute the solution at final time $t_f = 1$. As expected, the L^1-error shown in Table 4.4 is not in the order of magnitude of the machine accuracy, but the hydrostatic equilibrium is still preserved up to second order. This result remains true even if we use a constant reconstruction and consequently a first order scheme. As the convergence rates in Table 4.5 show, the hydrostatic equilibrium is maintained up to second order despite the constant reconstruction. Mathematically, this can be explained by the fact that equation (4.130) is satisfied up to second order.

N	$L^1(\rho)$	$EOC(\rho)$	$L^1(\rho u)$	$EOC(\rho u)$	$L^1(\rho v)$	$EOC(\rho v)$	$L^1(E)$	$EOC(E)$
32	9.43E-06	-	1.36E-05	-	1.36E-05	-	5.08E-05	-
64	2.35E-06	2.01	3.43E-06	1.99	3.43E-06	1.99	1.26E-05	2.01
128	5.88E-07	2.00	8.60E-07	2.00	8.60E-07	2.00	3.14E-06	2.01
256	1.47E-07	2.00	2.16E-07	1.99	2.16E-07	1.99	7.85E-07	2.00
512	3.69E-08	2.00	5.42E-08	2.00	5.42E-08	2.00	1.97E-07	2.00

Table 4.4: L^1-error and experimental order of convergence of the second order 2S-WB-FV scheme for a general steady state using the $\bar{\rho}$-average (4.123).

N	$L^1(\rho)$	$EOC(\rho)$	$L^1(\rho u)$	$EOC(\rho u)$	$L^1(\rho v)$	$EOC(\rho v)$	$L^1(E)$	$EOC(E)$
32	9.74E-06	-	1.40E-05	-	1.40E-05	-	5.15E-05	-
64	2.39E-06	2.03	3.48E-06	2.01	3.48E-06	2.01	1.27E-05	2.02
128	5.93E-07	2.01	8.67E-07	2.01	8.67E-07	2.01	3.15E-06	2.01
256	1.48E-07	2.00	2.17E-07	2.00	2.17E-07	2.00	7.86E-07	2.00
512	3.70E-08	2.00	5.43E-08	2.00	5.43E-08	2.00	1.97E-07	2.00

Table 4.5: L^1-error and experimental order of convergence of the first order 2S-WB-FV scheme for a general steady state using the $\bar{\rho}$-average (4.123).

Let us now assume that we a priori know the hydrostatic equilibrium and it is given as discrete data for the density and pressure. In this case, the approach described in Lem. 4.3.14 should be able to maintain this particular hydrostatic equilibrium up to machine precision. In order to check this, we set the values ρ_{hs} and p_{hs} equal to the initial values for density respective pressure. The L^1-error in Table 4.6 shows that the hydrostatic equilibrium is indeed maintained up to machine precision. This result illustrates that the 2S-WB-FV scheme is well-balanced for every a priori known hydrostatic equilibrium.

N	$L^1(\rho)$	$L^1(\rho u)$	$L^1(\rho v)$	$L^1(E)$
32	6.54E-17	9.10E-16	9.10E-16	1.33E-15
64	1.85E-16	1.95E-15	1.95E-15	4.78E-15
128	2.98E-16	4.78E-15	4.78E-15	8.97E-15
256	6.25E-16	8.32E-16	8.32E-16	2.04E-14
512	1.25E-15	1.83E-14	1.83E-14	4.24E-14

Table 4.6: L^1-error for a general steady state using the approach for a priori known hydrostatic equilibria from Lem. 4.3.14.

4.6.6 Perturbation of an Isothermal Atmosphere

One main advantage of well-balanced schemes is their ability to resolve small perturbations of the hydrostatic equilibrium even on coarse grids. It is precisely this effect that we are investigating with the following test. For this purpose, we take the initial values from

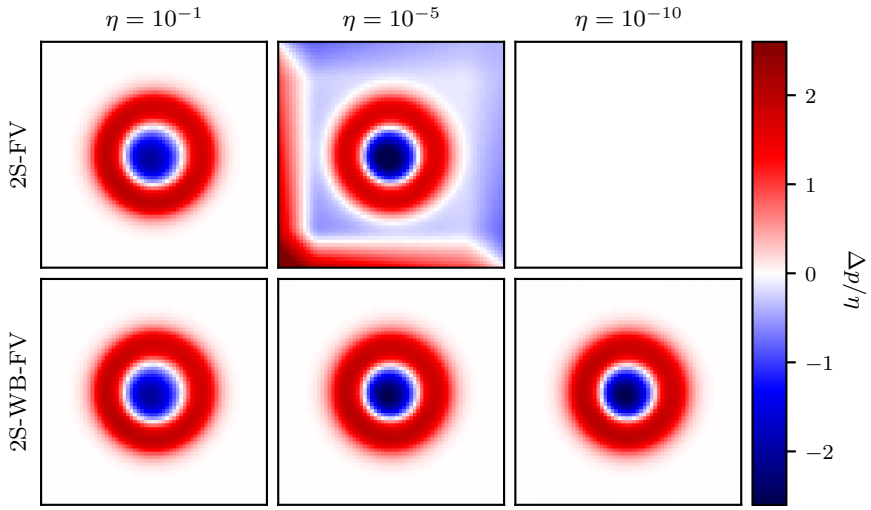

Figure 4.4: Pressure perturbations of varying strength η of an isothermal atmosphere at $t_f = 0.15$ resolved by the 2S-FV (top row) and 2S-WB-FV (bottom row) scheme. The perturbations are rescaled by the corresponding value of η.

the isothermal atmosphere in Sect. 4.6.4, which are in hydrostatic equilibrium, and add a perturbation on the pressure by redefining its initial value by

$$p(x, y, 0) = p_0 \exp\left(-\rho_0 g(x + y)/p_0\right) + \eta \exp\left(-100\rho_0 g((x - 0.5)^2 + (y - 0.5)^2)/p_0\right).$$

The strength of the perturbation is controlled by the parameter η. The numerical solutions are computed on a 64×64 mesh up to a final time $t_f = 0.15$. In order to investigate the well-balancing effect, we compare the results of the new 2S-WB-FV scheme with a non-well-balanced *two-speed finite volume* (2S-FV) scheme that uses the arithmetic mean of left and right density for $\bar{\rho}$.
The numerical solutions of the two schemes for three differently strong perturbations are displayed in Fig. 4.4. For the largest perturbation with $\eta = 10^{-1}$, a qualitative comparison shows no difference between the two solutions. In the case of the medium-sized perturbation, the non-well-balanced 2S-FV scheme can still resolve the perturbation, but the underlying hydrostatic equilibrium is no longer preserved. For the smallest perturbation, the method does not resolve the perturbation at all and the scale of the perturbation lies far outside of the scale of the plot. The well-balanced 2S-WB-FV method, on the other hand, manages to resolve the medium and very small perturbation and also preserves the underlying equilibrium. This underlines the functionality of the well-balancing mechanism and also demonstrates the importance of this property for problems close to hydrostatic equilibria.

4.6.7 Kelvin-Helmholtz Instability

In the following part, we run simulations of a Kelvin-Helmholtz instability. This is the primary instability that arises when there is a velocity shear within a continuous fluid, and it is the main source of vorticity that leads to the energy cascade in 3D turbulent flows [LBA+22]. Let us consider a spatial domain $\mathcal{I} = [0, 2] \times [-0.5, 0.5]$. The initial

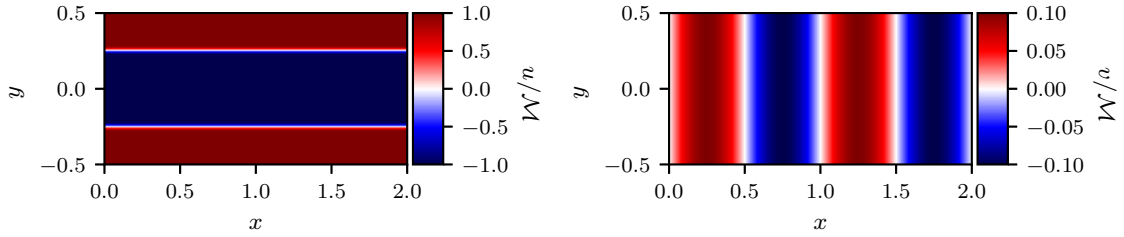

Figure 4.5: Initial setups of u and v (here rescaled by $\mathcal{M}$) used for simulating the growth of the Kelvin-Helmholtz instability.

density, horizontal velocity and pressure are given by

$$\begin{aligned} \rho(x,y,0) &= \gamma \,\frac{\mathrm{kg}}{\mathrm{m}^3}, \\ u(x,y,0) &= \mathcal{M}[1-2\eta(x,y)]\,\frac{\mathrm{m}}{\mathrm{s}}, \\ p(x,y,0) &= 1\,\frac{\mathrm{kg}}{\mathrm{ms}^2}, \end{aligned} \tag{4.149}$$

with

$$\eta(x,y) = \begin{cases} \frac{1}{2}\{1+\sin\left[16\pi(y+0.25)\right]\}, & \text{if } -\frac{9}{32} < y < -\frac{7}{32}, \\ 1, & \text{if } -\frac{7}{32} < y < \frac{7}{32}, \\ \frac{1}{2}\{1-\sin\left[16\pi(y-0.25)\right]\}, & \text{if } \frac{7}{32} < y < \frac{9}{32}, \\ 0, & \text{else.} \end{cases} \tag{4.150}$$

The ratio of specific heat is chosen to be $\gamma = 1.4$. The instability is started by adding a perturbation to the y-velocity component in form of

$$v(x,y,0) = 0.1\mathcal{M}\sin(2\pi x)\,\frac{\mathrm{m}}{\mathrm{s}}. \tag{4.151}$$

The initial velocity profiles are also illustrated in Fig. 4.5. Periodic boundary conditions are imposed in both directions. The initial data is transformed to dimensionless quantities by the reference values

$$x_r = 1\,\mathrm{m}, \quad t_r = \frac{1}{u_r}\,\mathrm{s}, \quad \rho_r = 1\,\frac{\mathrm{kg}}{\mathrm{m}^3}, \quad u_r = \mathcal{M}\,\frac{\mathrm{m}}{\mathrm{s}}, \quad p_r = 1\,\frac{\mathrm{kg}}{\mathrm{ms}^2}. \tag{4.152}$$

The final time is set to $t_f = 0.8$. The chosen initial conditions are such that the interface across the shear flow is smooth and resolved, which leads to convergent results at least in the early stages of the evolution of the flow.
We perform a qualitative comparison between the low Mach compliant 2S-WB-FV scheme and its counterpart, the *one-speed well-balanced finite volume* (1S-WB-FV) scheme that uses the speeds (4.103) in order to investigate the effect of using two different relaxation speeds. In Fig. 4.6, we present the numerical results for the local Mach number $\mathcal{M}_{\mathrm{loc}}$ relative to $\mathcal{M}$ at final time t_f computed on a 128×64 grid. For all Mach numbers $\mathcal{M}$ there is a clear difference in the quality of the results of the two methods. The 1S-WB-FV scheme is not able to resolve the vortexes properly because the dissipation in the momentum flux scales with $\mathcal{O}(1/\mathcal{M})$. The 2S-WB-FV scheme, on the other hand,

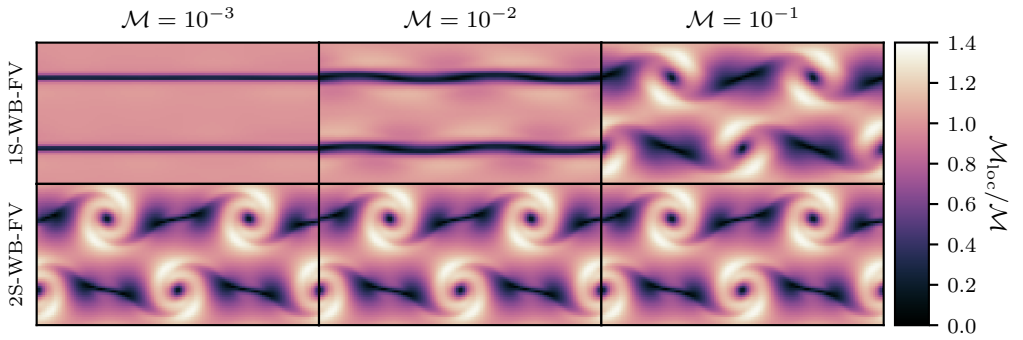

Figure 4.6: Numerical results of the local Mach number rescaled by the corresponding $\mathcal{M}$ at t_f for the 1S-WB-FV (top row) and 2S-WB-FV (bottom row) scheme on a 128×64 grid.

manages to resolve the structures of the flows well even at low Mach numbers. The fact that the results do not deteriorate for smaller Mach numbers shows that the dissipation in the 2S-WB-FV scheme becomes Mach number independent by using the two relaxation speeds and thus underlines the theoretical results from Sect. 4.3.4.

4.6.8 Stationary Vortex in a Gravitational Field

In this section, we investigate whether the positive effect of two relaxation speeds is also visible when including gravitational forces. As a test, we use a version of the Gresho vortex modified for the Euler equations with a gravitational source term that was already given in [TPK20]. The density in this setup is defined by

$$\rho(x, y, 0) = \exp\left(-\frac{\Phi(x, y)}{RT}\right). \tag{4.153}$$

The rest of the initial data is given in radial coordinates (r, θ). The velocity field has the form

$$u_\theta(r, 0) = \begin{cases} 5r, & \text{if } r \leq 0.2, \\ 2 - 5r, & \text{if } 0.2 < r \leq 0.4, \\ 0, & \text{else}, \end{cases} \tag{4.154}$$

and the gravitational potential is defined by

$$\Phi(r) = \begin{cases} 12r^2, & \text{if } r \leq 0.2, \\ 0.5 - \ln(0.2) + \ln(r), & \text{if } 0.2 < r \leq 0.4, \\ \ln(2) - 0.5\frac{r_c}{r_c - 0.4} + 2.5\frac{r_c}{r_c - 0.4}r - 1.25\frac{1}{r_c - 0.4}r^2, & \text{if } 0.4 < r \leq r_c, \\ \ln(2) - 0.5\frac{r_c}{r_c - 0.4} + 1.25\frac{r_c^2}{r_c - 0.4}, & \text{else}. \end{cases} \tag{4.155}$$

The pressure p is split into a hydrostatic pressure p_0 and a pressure p_2 associated with the centrifugal forces and given by $p = p_0 + \mathcal{M}^2 p_2$, where $p_0 = RT\rho$ and

$$p_2(r, 0) = RT \begin{cases} p_{21}(r), & \text{if } r \leq 0.2, \\ p_{21}(0.2) + p_{22}(r), & \text{if } 0.2 < r \leq 0.4, \\ p_{21}(0.2) + p_{22}(0.4), & \text{else}, \end{cases} \tag{4.156}$$

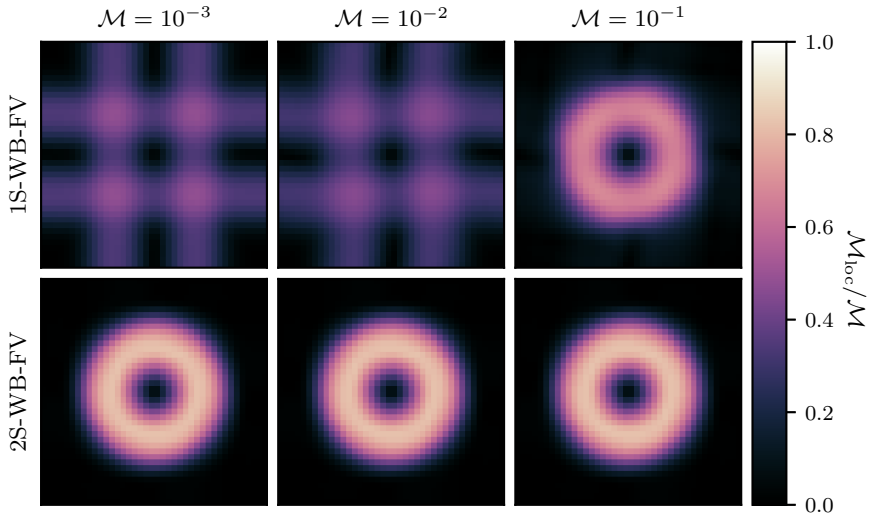

Figure 4.7: Numerical results of the local Mach number rescaled by the corresponding $\mathcal{M}$ at t_f for the 1S-WB-FV (top row) and 2S-WB-FV (bottom row) scheme on a 40×40 grid.

with

$$
\begin{aligned}
p_{21}(r) =& \left(1 - \exp\left(-12.5\frac{r^2}{RT}\right)\right), \\
p_{22}(r) =& \frac{1}{(1-\mathcal{M}^2)(1-0.5\mathcal{M}^2)} \exp\left(\frac{-0.5 + \ln(0.2))}{RT}\right) \\
& \left(r^{-\frac{1}{RT}}\left(\mathcal{M}^4(r(10-12.5r)-2) - 4 + \mathcal{M}^4(r(12.5r-20)+6)RT\right)\right. \\
& \left. + \exp\left(\frac{-\ln(0.2)}{RT}\right)\left(4 - 2.5\mathcal{M}^4 RT + 0.5\mathcal{M}^4\right)\right).
\end{aligned}
$$

The initial data is described in physical quantities. The reference values for the transformation in dimensionless quantities are given by

$$
x_r = 1\,\mathrm{m}, \quad t_r = \frac{1}{u_r}\,\mathrm{s} \quad \rho_r = 1\,\frac{\mathrm{kg}}{\mathrm{m}^3}, \quad u_r = 0.4\pi\,\frac{\mathrm{m}}{\mathrm{s}}, \quad RT = \frac{1}{\mathcal{M}^2}\,\frac{\mathrm{m}^2}{\mathrm{s}^2}. \tag{4.157}
$$

We choose $\gamma = 5/3$ for the adiabatic coefficient. The spatial domain is $\mathcal{I} = [0,1]^2$ and has periodic boundary conditions. The computations are carried out on a 40×40 grid until a final time $t_f = 0.8s$, which corresponds to one turn of the vortex. We solve this initial value problem for various Mach numbers $\mathcal{M}$ using the 1S-WB-FV and the 2S-WB-FV scheme. The solutions generated by the one-speed relaxation scheme are depicted in the top row of Fig. 4.7, while the solutions computed by the two-speed relaxation scheme are shown in the bottom row. The results support the findings of the Kelvin-Helmholtz instability. The one-speed scheme introduces a Mach number dependent dissipation that smears the structure of the vortex. The vortexes produced by the two-speed scheme, on the other hand, retain their shape regardless of the Mach number so that no qualitative difference is discernible.

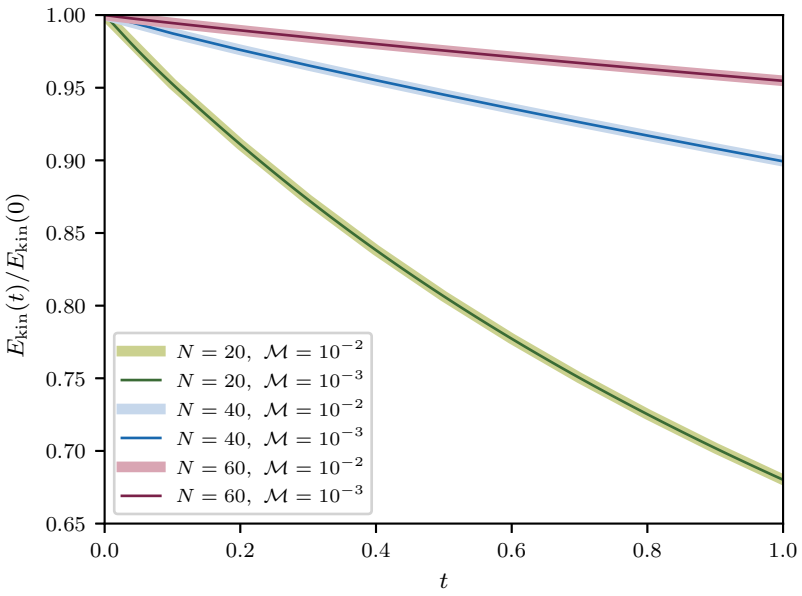

Figure 4.8: Time evolution of the total kinetic energy for different Mach numbers $\mathcal{M}$ and grid resolutions $N = N_x = N_y$ for the 2S-WB-FV scheme.

Quantitative evidence for this behavior can be found in the analysis of kinetic energy. In (2.65) it is shown, that kinetic energy is conserved in the low Mach number limit. A numerical scheme should mimic this behaviour discretely. Fig. 4.8 presents the loss of total kinetic energy in time in the results of the 2S-WB-FV scheme. While the curves for different grid resolutions $N \times N$ differ, the curves for different Mach numbers $\mathcal{M}$ but the same grid resolution match. Consequently, the loss of kinetic energy, which can be interpreted as a measure for the artificial dissipation of the numerical method, only depends on the grid resolution but not on the Mach number.

4.7 Summary and Conclusions

The proposed scheme extends the two-speed relaxation approach to the full Euler equations with a time-independent gravitational source term. The resulting approximate Riemann solver is designed in the way that it solves the inhomogeneous Riemann problem. The two-speed ansatz reduces the artificial dissipation in the low Mach number regime, making the scheme in consequence provably asymptotic-preserving. In numerical tests, the two-speed method shows a significantly better performance in resolving nearly incompressible flows in comparison to a classical one-speed Suliciu relaxation scheme. In the case of supersonic flows, the method is reduced to the one-speed method so that enough dissipation is introduced to capture shocks.
Solving the inhomogeneous Riemann problem helps to find a consistent discretization of the fluxes and the source term. By incorporating a discretization of the hydrostatic equilibrium equation into all intermediate states of the Riemann solver, the method becomes well-balanced for certain families and a priori known hydrostatic equilibria. Numerical simulations show that the well-balanced method can maintain the equilibria up to machine

precision. This property also helps to resolve small perturbations of such a background equilibrium.
The method is particularly stable since the approximate Riemann solver is positivity-preserving, entropy-satisfying and prevents the occurrence of checkerboard modes in the velocity and pressure variables. The price to be paid for this stability is a more restrictive CFL condition. Due to the fully explicit discretization, the CFL condition already depends on the Mach number. By rescaling the relaxation velocity a to comply with the subcharacteristic condition, the CFL condition for the two-speed method scales with $\mathcal{O}(\mathcal{M}^2)$ instead of $\mathcal{O}(\mathcal{M})$. In practice, the method can therefore only be used for moderately low Mach numbers. Numerical experiments moreover showed that if the two-speed solver is combined with an implicit time-marching scheme, the nonlinear solver only converges if the time step has the order $\mathcal{O}(\mathcal{M})$. It is therefore necessary to further develop the two-speed method itself in order to relax the CFL condition.

Chapter 5

An Implicit-Explicit Strang Splitting Method

The two-speed approach successfully reduces the artificial dissipation in the approximate Riemann solver in the low Mach number regime. Though, it does not address the Mach number dependent CFL restriction on the time step. Instead, the CFL condition becomes more restrictive (even with implicit time integration to reach convergence of the nonlinear solver). Therefore, in practice, there is a need for an alternative low Mach number strategy that makes both the dissipation and the time step condition Mach number independent. In this chapter, we present a new FV method for the ideal MHD equations that addresses both problems induced by low Mach numbers. The Godunov-type method relies on a low Mach version of the HLLD solver [MM21] in which only the dissipation term in the intermediate state of the pressure is rescaled while the wave speeds remain unaltered. In order to bypass the restrictive CFL condition, a new time-marching scheme is introduced. Since we only consider small sonic Mach numbers $\mathcal{M}$, but not small Alfvén Mach numbers $\mathcal{M}_{\mathrm{Alf}}$, the CFL condition is only tightened by the acoustic pressure term in the flux. For his reason, we split the MHD system into two parts following the approach in [FMR09], and solve only the subset of continuity, momentum and energy equations implicitly, whereas the induction equation is integrated using an explicit time-stepper. The two separate steps are coupled with second-order accuracy by Strang splitting [Str68]. The time step is then limited by the fastest fluid/Alfvén speeds on the grid, and it is approximately $1/\mathcal{M}$ longer than what is allowed by the standard CFL condition. This leads to a considerable speed-up when the Mach number of the flow is low.
Since the update on the induction equation is performed in a separate step, the flux-Jacobian in the time-implicit part of the algorithm does not need to be evaluated with respect to the magnetic field components. This allows us more flexibility when choosing the method that evolves the magnetic field. In particular, we use a staggered formulation of constrained transport [GS05] to satisfy the solenoidal constraint up to machine precision, at least for a specific discretization of the divergence of the magnetic field.
The method is coupled with the *Deviation Well-Balancing method* [BCK21, EHB+21], which allows to preserve the a priori known background stratification in MHSE, dramatically reducing the magnitude of numerical errors and the strength of spurious flows.

This chapter is structured as follows. In Sect. 5.1 we briefly define the set of MHD equations that shall be discretized. In Sect. 5.2 and 5.3 we provide the details on the numerical methods for space and time discretization. In Sect. 5.4 several numerical experiments are run with the new MHD scheme in order to check its accuracy and efficiency in simulat-

ing flows at low Mach numbers, even in the presence of a steep stratification. Finally, in Sect. 5.5 we draw the conclusions and summarize the fundamental aspects of the proposed algorithm. The results of this chapter closely follow the presentation in [LBA+22].

5.1 Governing Equations

In this chapter we rewrite the MHD equations (2.66) by including the time-independent source term of the energy equation in the definition of the total energy

$$E_\Phi = \rho e + \frac{1}{2}\rho|\mathbf{v}|^2 + \frac{1}{2}|\mathbf{B}|^2 + \rho\Phi. \tag{5.1}$$

Using this definition, we can write the MHD equations in the following form:

$$\frac{\partial}{\partial t}\begin{pmatrix} \rho \\ \rho\mathbf{v} \\ E_\Phi \\ \mathbf{B} \end{pmatrix} + \nabla\cdot\begin{pmatrix} \rho\mathbf{v} \\ \rho\mathbf{v}\otimes\mathbf{v} + (p+\frac{1}{2}|\mathbf{B}|^2)\mathbf{I} - \mathbf{B}\otimes\mathbf{B} \\ (E_\Phi + p + \frac{1}{2}|\mathbf{B}|^2)\mathbf{v} - \mathbf{B}\,(\mathbf{B}\cdot\mathbf{v}) \\ \mathbf{v}\otimes\mathbf{B} - \mathbf{B}\otimes\mathbf{v} \end{pmatrix} = \begin{pmatrix} 0 \\ \rho\mathbf{g} \\ 0 \\ \mathbf{0} \end{pmatrix}. \tag{5.2}$$

Numerical experiments show that solving the equation for E_Φ leads to more accurate results and better entropy and energy conservation properties in simulations of gas dynamics with gravity [Mü20, EHB+21].

5.2 Spatial Discretization

The MHD system (5.2) takes the general form

$$\frac{\partial \mathcal{Q}}{\partial t} + \frac{\partial \mathcal{F}_1(\mathcal{Q})}{\partial x} + \frac{\partial \mathcal{F}_2(\mathcal{Q})}{\partial y} + \frac{\partial \mathcal{F}_3(\mathcal{Q})}{\partial z} = \mathcal{S}(\mathcal{Q}), \tag{5.3}$$

with the respective vector of conservative variables $\mathcal{Q}$, physical fluxes $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ and source term $\mathcal{S}$. For the discretization we use a FV method of the form

$$\begin{aligned} \frac{\partial Q_{i,j,k}}{\partial t} = &-\frac{1}{\Delta x}\left(F_{1,i+1/2,j,k} - F_{1,i-1/2,j,k}\right) \\ &-\frac{1}{\Delta y}\left(F_{2,i,j+1/2,k} - F_{2,i,j-1/2,k}\right) \\ &-\frac{1}{\Delta z}\left(F_{3,i,j,k+1/2} - F_{3,i,j,k-1/2}\right) \\ &+ S_{i,j,k}, \end{aligned} \tag{5.4}$$

where F_1, F_2 and F_3 represent numerical flux functions and S a numerical source term. For the latter, we simply take the value of the physical source in the center of the cell, which is accurate to second order:

$$S_{i,j,k} \simeq \mathcal{S}_{i,j,k}. \tag{5.5}$$

The computation of numerical fluxes, in contrast, needs more care, and is subject of the following section.

5.2.1 Numerical Flux Function

One suitable way to derive a proper estimate of the fluxes in a FV method for the MHD equations is to use the HLLD approximate Riemann solver that is briefly described in Ex. 3.3.7. The numerical fluxes of the original HLLD solver in [MK05] are computed from Rankine-Hugoniot conditions of the form

$$F^{L*} = \mathcal{F}(Q^L) + S^L(Q^{L*} - Q^L), \tag{5.6}$$
$$F^{L**} = F^{L*} + S^{L*}(Q^{L**} - Q^{L*}), \tag{5.7}$$
$$F^{R*} = \mathcal{F}(Q^R) + S^R(Q^{R*} - Q^R), \tag{5.8}$$
$$F^{R**} = F^{R*} + S^{R*}(Q^{R**} - Q^{R*}). \tag{5.9}$$

However, we decide to only use the intermediate states from the original solver and plug those states into the physical flux. Then the numerical flux is defined by

$$F(Q^L, Q^R) = \begin{cases} \mathcal{F}(Q^L), & \text{if } S^L > 0, \\ \mathcal{F}(Q^{L*}), & \text{if } S^L < 0 < S^{L*}, \\ \mathcal{F}(Q^{L**}), & \text{if } S^{L*} < 0 < S^M, \\ \mathcal{F}(Q^{R**}), & \text{if } S^M < 0 < S^{R*}, \\ \mathcal{F}(Q^{R*}), & \text{if } S^{R*} < 0 < S^R, \\ \mathcal{F}(Q^R), & \text{if } S^R < 0. \end{cases} \tag{5.10}$$

Either way, the approximate Riemann solver introduces a large amount of dissipation in the low Mach number regime. We can investigate this behaviour by analyzing the intermediate flux as it is done for the two-speed relaxation solver in Sect. 4.3.4. It is sufficient to concentrate on the momentum flux, as this is the only flux component in which a term is divided by the Mach number $\mathcal{M}$. For the following part, we switch to the dimensionless form of the MHD equations in order to show the Mach number's influence. The intermediate states for the velocity and the pressure are constant across the entropy and Alfvén waves and are given by

$$S_M = u^* = \frac{(S^R - u^R)\rho^R u^R - (S^L - u^L)\rho^L u^L - \frac{1}{\mathcal{M}^2}\left(p_T^R + p_T^L\right)}{\rho^L(u^L - S^L) - \rho^R(S^R - u^R)}, \tag{5.11}$$
$$p_T^* = \frac{(S^R - u^R)\rho^R p_T^L - (S^L - u^L)\rho^L p_T^R + \mathcal{M}^2\rho^L\rho^R(S^R - u^R)(S^L - u^L)(u^R - u^L)}{(S^R - u^R)\rho^R - (S^L - u^L)\rho^L}, \tag{5.12}$$

where S^L and S^R can be defined by (3.45) with the dimensionless fast magnetosonic speeds given in (2.81). The total pressure in this dimensionless version is given by $p_T = p + \frac{1}{2}\mathcal{M}^2|\mathbf{B}|^2$. We can rewrite these intermediate states in more convenient forms that strongly resemble those in Suliciu-type relaxation solvers:

$$u^* = \frac{a^L u^L + a^R u^R + \frac{1}{\mathcal{M}^2}\left(p_T^L - p_T^R\right)}{a^L + a^R}, \tag{5.13}$$
$$p_T^* = \frac{a^R p_T^L + a^L p_T^R + \mathcal{M}^2 a^L a^R (u^L - u^R)}{a^L + a^R}, \tag{5.14}$$

with the speeds

$$a^L = \rho^L(S^L - u^L) \quad \text{and} \quad a^R = \rho^R(S^R - u^R). \tag{5.15}$$

In the low Mach number limit, these speeds are dominated by the sound speed, i.e.

$$a^L = \frac{1}{\mathcal{M}}\left(\rho^L c^L + \mathcal{O}(\mathcal{M})\right) \quad \text{and} \quad a^R = \frac{1}{\mathcal{M}}\left(\rho^R c^R + \mathcal{O}(\mathcal{M})\right). \tag{5.16}$$

We define

$$\bar{a}^L = \rho^L c^L \quad \text{and} \quad \bar{a}^R = \rho^R c^R, \tag{5.17}$$

which satisfy

$$\bar{a}^R - \bar{a}^L = \mathcal{O}(\mathcal{M}^2), \tag{5.18}$$

so that we can write $\bar{a}$ for both $\bar{a}^L$ and $\bar{a}^R$. Expansions of the intermediate states (5.13) and (5.14) then yield

$$u^* = \frac{u^L + u^R}{2} + \frac{p_T^L - p_T^R}{2\bar{a}\mathcal{M}} + \mathcal{O}(\mathcal{M}), \tag{5.19}$$

$$p_T^* = \frac{p_T^L + p_T^R}{2} + \mathcal{M}\bar{a}\frac{u^L - u^R}{2} + \mathcal{O}(\mathcal{M}^2). \tag{5.20}$$

For the sake of simplicity, we do not write out the $\mathcal{O}(\mathcal{M}^2)$ terms in the pressure because they remain bounded in the flux and do not contribute to the excessive dissipation in the low Mach number regime. We can now analyze the intermediate momentum flux, which has the form

$$\begin{aligned} F_{\rho u}^{L*} &= \rho^{L*}(u^*)^2 + \frac{p_T^*}{\mathcal{M}^2} - \left(\frac{B_x^L + B_x^R}{2}\right)^2 \\ &= \frac{\rho^L (u^L)^2 + \rho^R (u^R)^2}{2} + \frac{p_T^L + p_T^R}{2\mathcal{M}^2} - \frac{(B_x^L)^2 + (B_x^R)^2}{2} \\ &\quad - \left(\frac{u^L - u^R}{2}\right)^2 + 2\left(\frac{p_T^L - p_T^R}{2\bar{a}\mathcal{M}}\right)\left(\frac{u^L + u^R}{2}\right) + \left(\frac{p_T^L - p_T^R}{2\bar{a}\mathcal{M}}\right)^2 + \bar{a}\frac{u^L - u^R}{2\mathcal{M}} \\ &\quad + \left(\frac{B_x^L - B_x^R}{2}\right)^2 + \mathcal{O}(1). \end{aligned} \tag{5.21}$$

Here, we made use of the fact that in the low Mach limit the density is constant up to errors of order $\mathcal{O}(\mathcal{M}^2)$. The numerical flux consists of a central flux and a dissipation term. Since the pressure difference $p_T^L - p_T^R$ scales with $\mathcal{O}(\mathcal{M}^2)$, these dissipation terms remain bounded despite the Mach number in the denominator. In contrast, the penultimate dissipation term containing the velocity difference $u^L - u^R$ scales with $\mathcal{O}(1/\mathcal{M})$ and thus causes an increasing dissipation of the original version of the HLLD solver in the low Mach number limit.

In order to cure this problem, we rely on a modification of the solver proposed in [MM21]. In this modification, a Mach number dependent parameter $\phi \propto \mathcal{M}$ is inserted in the intermediate state of the total pressure:

$$\begin{aligned} p_{\mathrm{T}}^* = &\frac{(S^R - u^R)\rho^R p_{\mathrm{T}}^L + (S^L - u^L)\rho^L p_{\mathrm{T}}^R}{(S^R - u^R)\rho_R - (S^L - u^L)\rho^L} \\ &+ \phi\frac{\mathcal{M}^2 \rho^L \rho^R (S^R - u^R)(S^L - u^L)(u^R - u^L)}{(S^R - u^R)\rho^R - (S^L - u^L)\rho^L}. \end{aligned} \tag{5.22}$$

This parameter is computed according to the following formulas:

$$
\begin{aligned}
c_{\mathrm{u}}^{L} &= \left[\frac{1}{2}\left(\frac{|\mathbf{B}^{L}|^2}{\rho^{L}} + |\mathbf{v}^{L}|^2 + \sqrt{\left(\frac{|\mathbf{B}^{L}|^2}{\rho^{L}} + |\mathbf{v}^{L}|^2\right)^2 - 4\frac{|\mathbf{v}^{L}|^2 B_x^2}{\rho^{L}}}\right)\right]^{\frac{1}{2}}, \\
c_{\mathrm{u}}^{R} &= \left[\frac{1}{2}\left(\frac{|\mathbf{B}^{R}|^2}{\rho^{R}} + |\mathbf{v}^{R}|^2 + \sqrt{\left(\frac{|\mathbf{B}^{R}|^2}{\rho^{R}} + |\mathbf{v}^{R}|^2\right)^2 - 4\frac{|\mathbf{v}^{R}|^2 B_x^2}{\rho^{R}}}\right)\right]^{\frac{1}{2}}, \\
\chi &= \max\left\{\frac{c_{\mathrm{u}}^{L}}{c_{\mathrm{f},x}^{L}}, \frac{c_{\mathrm{u}}^{R}}{c_{\mathrm{f},x}^{R}}\right\}, \\
\phi &= \chi(2-\chi).
\end{aligned}
\tag{5.23}
$$

The modification of the pressure changes the scaling in the momentum flux (5.21) into

$$
\begin{aligned}
F_{\rho u}^{L*} = &\frac{\rho^{L}(u^{L})^2 + \rho^{R}(u^{R})^2}{2} + \frac{p_T^{L} + p_T^{R}}{2\mathcal{M}^2} - \frac{(B_x^{L})^2 + (B_x^{R})^2}{2} \\
&- \left(\frac{u^{L} - u^{R}}{2}\right)^2 + 2\left(\frac{p_T^{L} - p_T^{R}}{2\bar{a}\mathcal{M}}\right)\left(\frac{u^{L} + u^{R}}{2}\right) + \left(\frac{p_T^{L} - p_T^{R}}{2\bar{a}\mathcal{M}}\right)^2 + \bar{a}\frac{u^{L} - u^{R}}{2} \\
&+ \left(\frac{B_x^{L} - B_x^{R}}{2}\right)^2 + \mathcal{O}(\mathcal{M}),
\end{aligned}
\tag{5.24}
$$

so that the Mach number in the velocity term vanishes and in consequence the artificial dissipation becomes independent of the sonic Mach number. The modified solver is called *low-dissipation HLLD* (LHLLD) solver.

Remark 5.2.1. *If the Alfvén Mach number $\mathcal{M}_{\mathrm{Alf}}$ is not equal to one, we note that the combined dissipation in (5.24) has a residual scaling $\mathcal{O}(1/\mathcal{M}_{\mathrm{Alf}})$. Therefore, the solver still introduces too much dissipation in sub-Alfvén regimes.*

Remark 5.2.2. *The low Mach fix in LHLLD works very similarly to the one in the two-speed relaxation solver. Here too, an additional $\mathcal{M}$ is inserted into the intermediate pressure to change the scaling of the velocity term. In contrast to the two-speed approach, however, there is only this one local change and no additional dissipation is introduced in the pressure difference dissipation term of the intermediate velocity (see Rem. 4.3.11). Therefore, it is at least questionable whether the LHLLD solver (or a low-dissipation HLLC (LHLLC) solver) fulfills a discrete entropy inequality as the relaxation solver does.*

Remark 5.2.3. *Due to the close relationship between HLL-type and relaxation solvers, it is clear that the low Mach fix proposed in [MM21] can also be transferred to Suliciu-type relaxation solvers.*

5.2.2 Well-Balancing Method

In our scheme the hyperbolic fluxes and gravitational source terms are discretized separately with different methods. As a consequence, the scheme does not automatically preserve magnetohydrostatic solutions on a discrete grid exactly. Therefore, whenever a stratification needs to be enforced to be in MHSE on the computational grid, we use the Deviation Well-Balancing method [BCK21, EHB+21]. The main ingredient of this

method is an a priori known target state $\tilde{\mathcal{Q}}$ that is a magnetohydrostatic solution to (5.3) and consequently satisfies

$$\frac{\partial \mathcal{F}_1(\tilde{\mathcal{Q}})}{\partial x} + \frac{\partial \mathcal{F}_2(\tilde{\mathcal{Q}})}{\partial y} + \frac{\partial \mathcal{F}_3(\tilde{\mathcal{Q}})}{\partial z} = \mathcal{S}(\tilde{\mathcal{Q}}), \tag{5.25}$$

with $\tilde{\mathbf{v}} = 0$. Subtracting (5.25) from the original balance law in (5.3) yields a system of PDEs for the deviations from the target solution $\Delta\mathcal{Q} = \mathcal{Q} - \tilde{\mathcal{Q}}$:

$$\begin{aligned} \frac{\partial(\Delta\mathcal{Q})}{\partial t} &+ \left(\frac{\partial \mathcal{F}_1(\tilde{\mathcal{Q}} + \Delta\mathcal{Q})}{\partial x} - \frac{\partial \mathcal{F}(\tilde{\mathcal{Q}})}{\partial x} \right) \\ &+ \left(\frac{\partial \mathcal{F}_2(\tilde{\mathcal{Q}} + \Delta\mathcal{Q})}{\partial y} - \frac{\partial \mathcal{F}_2(\tilde{\mathcal{Q}})}{\partial y} \right) \\ &+ \left(\frac{\partial \mathcal{F}_3(\tilde{\mathcal{Q}} + \Delta\mathcal{Q})}{\partial z} - \frac{\partial \mathcal{F}_3(\tilde{\mathcal{Q}})}{\partial z} \right) = \mathcal{S}(\tilde{\mathcal{Q}} + \Delta\mathcal{Q}) - \mathcal{S}(\tilde{\mathcal{Q}}). \end{aligned} \tag{5.26}$$

Now, to obtain a well-balanced method, (5.26) is discretized according to the FV method described in Sect. 5.2, which leads to the semi-discrete form

$$\begin{aligned} \frac{\partial(\Delta Q)_{i,j,k}}{\partial t} = &- \frac{1}{\Delta x}\left(F^{dev}_{1,i+1/2,j,k} - F^{dev}_{1,i-1/2,j,k} \right) \\ &- \frac{1}{\Delta y}\left(F^{dev}_{2,i,j+1/2,k} - F^{dev}_{2,i,j-1/2,k} \right) \\ &- \frac{1}{\Delta z}\left(F^{dev}_{3,i,j,k+1/2} - F^{dev}_{3,i,j,k-1/2} \right) \\ &+ S^{dev}_{i,j,k}. \end{aligned} \tag{5.27}$$

In this formulation, the deviation fluxes and source terms are defined by

$$F^{dev}_{1,i+1/2,j,k} = F_{1,i+1/2,j,k} - \mathcal{F}_1\left(\tilde{Q}_{i+1/2,j,k}\right), \tag{5.28}$$

$$F^{dev}_{2,i,j+1/2,k} = F_{2,i,j+1/2,k} - \mathcal{F}_2\left(\tilde{Q}_{i,j+1/2,k}\right), \tag{5.29}$$

$$F^{dev}_{3,i,j,k+1/2} = F_{3,i,j,k+1/2} - \mathcal{F}_3\left(\tilde{Q}_{i,j,k+1/2}\right), \tag{5.30}$$

$$S^{dev}_{i,j,k} = S_{i,j,k} - \mathcal{S}(\tilde{\mathcal{Q}}_{i,j,k}), \tag{5.31}$$

where $F_{1,i+1/2,j,k}$ denotes the LHLLD flux evaluated in the states

$$Q^{L,R}_{i+1/2,j,k} = \Delta Q^{L,R}_{i+1/2,j,k} + \tilde{Q}_{i+1/2,j,k}, \tag{5.32}$$

while $\mathcal{F}_1\left(\tilde{Q}_{i+1/2,j,k}\right)$ corresponds to the physical flux of the MHD system evaluated in the target solution at the cell boundary. The deviations $\Delta Q_{i,j,k}$, rather than the states $Q_{i,j,k}$, are reconstructed to the boundary of the cell[1]. In the case of $\Delta Q_{i,j,k} = \mathbf{0}$, this means that the left and right state at the cell interface, which are used as initial data for the Riemann problem, are equal. A consistent Riemann solver then guarantees

$$F_{1,i+1/2,j,k} = F_1\left(\tilde{Q}_{i+1/2,j,k}, \tilde{Q}_{i+1/2,j,k}\right) = \mathcal{F}_1\left(\tilde{Q}_{i+1/2,j,k}\right), \tag{5.33}$$

[1] Deviations in the primitive variables can also be reconstructed if the corresponding equilibrium values are provided at the cell centers and at the cell boundaries.

so that the deviation fluxes and source terms become zero. Thus, the resulting method preserves magnetohydrostatic solutions and is well-balanced. Moreover, by removing the numerical errors arising from the magnetohydrostatic stratification, this method allows to simulate low Mach flows in stratified setups, which only cause small deviations from the MHSE state and would be completely dominated by spurious flows otherwise.

5.2.3 Constrained Transport Method

As described in Sect. 3.10.3, FV methods do not automatically satisfy the solenoidal constraint. This is also the case for the method at hand. In order to prevent the scheme from creating unphysical magnetic monopoles, we rely on a staggered constrained transport method. The key point of these methods is to compute the surface integral of

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathcal{E} = 0$$

over cell boundaries using Stokes' theorem, which leads to an equation for the magnetic field at the cell face[2]

$$\begin{aligned} \frac{\partial B_{x,i+1/2,j,k}}{\partial t} = \quad & \frac{1}{\Delta y}\left(\mathcal{E}_{z,i+1/2,j+1/2,k} - \mathcal{E}_{z,i+1/2,j-1/2,k}\right) \\ & -\frac{1}{\Delta z}\left(\mathcal{E}_{y,i+1/2,j,k+1/2} - \mathcal{E}_{y,i+1/2,j,k-1/2}\right). \end{aligned} \tag{5.34}$$

Analogous formulas can be derived for B_y and B_z. As a consequence of (5.34), it becomes necessary to store the magnetic field at the cell faces (see Fig. 5.1). However, the reconstruction of the initial data for the Riemann problems that are solved by the approximate Riemann solver still requires the cell-centered values of the magnetic field. These can be derived after the update (5.34) from the therein calculated face-centered magnetic field by an arithmetic mean

$$\begin{aligned} B_{x,i,j,k} =& \frac{1}{2}\left(B_{x,i-1/2,j,k} + B_{x,i+1/2,j,k}\right), \\ B_{y,i,j,k} =& \frac{1}{2}\left(B_{y,i,j-1/2,k} + B_{y,i,j+1/2,k}\right), \\ B_{z,i,j,k} =& \frac{1}{2}\left(B_{z,i,j,k-1/2} + B_{z,i,j,k+1/2}\right). \end{aligned} \tag{5.35}$$

The computation of the face-centered magnetic field in (5.34) relies on the electromotive force at the cell edges, which can be computed by the *Contact-CT algorithm* of [GS05]. In this method, the electric field at cell edges is computed as a simple arithmetic average of the four neighboring face-centered electromotive force components. The average is combined with a diffusion term that helps removing spurious oscillations when the magnetic field is advected. For instance, $\mathcal{E}_{z,i+1/2,j+1/2,k}$ is approximated to second order accuracy by

$$\begin{aligned} \mathcal{E}_{z,i+1/2,j+1/2,k} = & \frac{1}{4}\left(\bar{\mathcal{E}}_{z,i+1/2,j,k} + \bar{\mathcal{E}}_{z,i+1/2,j+1,k} + \bar{\mathcal{E}}_{z,i,j+1/2,k} + \bar{\mathcal{E}}_{z,i+1,j+1/2,k}\right) \\ & + \frac{\Delta y}{8}\left\{\left(\frac{\partial \mathcal{E}_z}{\partial y}\right)_{i+1/2,j+1/4,k} - \left(\frac{\partial \mathcal{E}_z}{\partial y}\right)_{i+1/2,j+3/4,k}\right\} \\ & + \frac{\Delta x}{8}\left\{\left(\frac{\partial \mathcal{E}_z}{\partial x}\right)_{i+1/4,j+1/2,k} - \left(\frac{\partial \mathcal{E}_z}{\partial x}\right)_{i+3/4,j+1/2,k}\right\}, \end{aligned} \tag{5.36}$$

[2] Here the calculation is made over the cell boundary $(i+1/2, j, k)$.

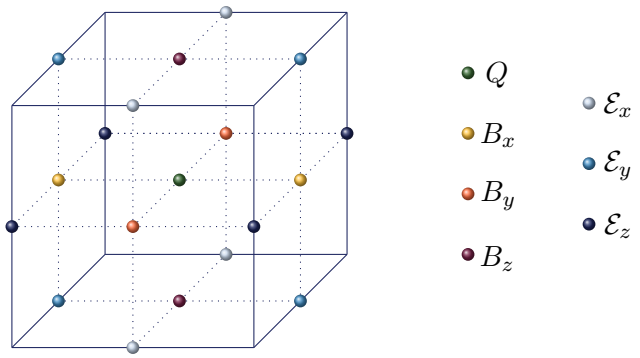

Figure 5.1: Locations within the three-dimensional cell $C_{i,j,k}$ at which the different quantities need to be stored in the Contact-CT method. The vector Q contains all conservative variables.

where $\bar{\mathcal{E}}_z$ can be computed from the solution given by the approximate Riemann solver. The calculation for the x- and y-component is again analogous. The upwind diffusion term enters in the derivatives of the electromotive force in (5.36), which are obtained according to the sign $s_{i+1/2,j,k}$ of the entropy waves at the cell interfaces:

$$\begin{aligned}\left(\frac{\partial \mathcal{E}_z}{\partial y}\right)_{i+1/2,j+1/4,k} =& \frac{1+s_{i+1/2,j,k}}{2}\left(\frac{\bar{\mathcal{E}}_{z,i,j+1/2,k}-\mathcal{E}^{cc}_{z,i,j,k}}{\Delta y/2}\right)+\\ & \frac{1-s_{i+1/2,j,k}}{2}\left(\frac{\bar{\mathcal{E}}_{z,i+1,j+1/2,k}-\mathcal{E}^{cc}_{z,i+1,j,k}}{\Delta y/2}\right).\end{aligned} \tag{5.37}$$

Here $\mathcal{E}^{cc}_{z,i,j,k} = (-\mathbf{v}_{i,j,k}\times\mathbf{B}_{i,j,k})_z$ represents the z-component of the cell-centered electromotive force. The discretization of the electromotive force leads to a semi-discrete form of (5.34) that can be integrated numerically in time. Any time-stepper that solves the resulting system of ODEs can keep the cell volume average of $\nabla\cdot\mathbf{B}$ defined by

$$\begin{aligned}(\nabla\cdot\mathbf{B})_{i,j,k} =& \frac{B_{x,i+1/2,j,k}-B_{x,i-1/2,j,k}}{\Delta x}+\\ & \frac{B_{y,i,j+1/2,k}-B_{y,i,j-1/2,k}}{\Delta y}+\\ & \frac{B_{z,i,j,k+1/2}-B_{z,i,j,k-1/2}}{\Delta z},\end{aligned} \tag{5.38}$$

within rounding errors.

5.3 Time Integration Algorithm

In addition to the spatial discretization, the scheme needs a suitable time discretization. A fully time-explicit method is computationally costly in practice due to its Mach number dependent CFL condition (see Sect. 3.10.1). In order to relax the CFL condition by making it independent of the Mach number, at least a part of the PDE system needs to be discretized time-implicitly. In regimes of low sonic Mach numbers, the stiffness is mostly generated by the pressure flux in the momentum equation, while the nondimensional form of the induction equation in (2.80) does not depend on the Mach number of the flow. This suggests that implicit time discretization only needs to be applied to the subset of continuity, momentum and energy equations, whereas the induction equation can be solved

with explicit time-steppers. Therefore, we split the induction equation from the continuity, momentum and energy equations, based on the approach described in [FMR09]. This allows us to use different spatial and temporal discretizations depending on the problem at hand. For the continuity, momentum and energy equations we decide to use an ESDIRK2 scheme [HS96]. The resulting nonlinear system of equations is solved iteratively with a root-finding Raphson-Newton algorithm, which relies on the analytic formulation of the flux-Jacobian. The *Biconjugate Gradient Stabilized Method* (BiCGSTAB(l)) [SF93] is used to solve each sub-step of the nonlinear solver. In contrast, the semi-discrete form of the induction equation (see (5.34)) is solved with the time-explicit SSP-RK2 method of [SO88].
These two updates can be combined to second order accuracy with Strang splitting [Str68]:

$$\mathcal{Q}^{n+1} = \mathcal{L}^{(\frac{1}{2}\Delta t)}\mathcal{H}^{(\Delta t)}\mathcal{L}^{(\frac{1}{2}\Delta t)}\mathcal{Q}^{n}. \tag{5.39}$$

Here, $\mathcal{L}$ represents a linear operator that updates only the magnetic field with the explicit marching scheme, while the nonlinear operator $\mathcal{H}$ updates density, momentum and total energy (including source terms) using the implicit stepper. In each sub-step of Strang splitting, the discretization of the fluxes, source terms, and electromotive force is performed according to the methods described in Sect. 5.2. From here on, we refer to this type of time discretization as *implicit-explicit Strang splitting* (IESS). The implicit-explicit terminus refers to the different discretization of continuity, momentum and energy equations in contrast to the induction equation. Despite the similarity in name, this approach should not be confused with IMEX methods such as the one in Chapter 6.

Numerical experiments performed with the IESS approach suggest that the maximum time step allowed for stability is approximately determined by

$$\Delta t = \min_{\Omega=(i,j,k)} \left\{ \frac{\Delta x}{|u_\Omega| + c_{\mathrm{a},x,\Omega}}, \frac{\Delta y}{|v_\Omega| + c_{\mathrm{a},y,\Omega}}, \frac{\Delta z}{|w_\Omega| + c_{\mathrm{a},z,\Omega}} \right\}, \tag{5.40}$$

so that the propagation of fluid motions and Alfvén waves is well-resolved in time. This time step is approximately $1/\mathcal{M}$ larger than that allowed by the conventional CFL condition if the plasma-β is high. Therefore, the computational effort is considerably reduced when simulating low Mach number flows. The price one has to pay is that the propagation of fast magnetosonic waves is not well-resolved in time.

A single step of the described time-marching scheme can be summarized in the following way:

1. Get Δt from (5.40) given ρ^n, $\rho\mathbf{v}^n$, E_Φ^n and $\mathbf{B}^n$.

2. Use SSP-RK2 and Contact-CT to solve the induction equation over the first half of the time step $\Delta t/2$. This results in an intermediate solution for the magnetic field $\mathbf{B}^{n+1/2}$.

3. Use this intermediate solution $\mathbf{B}^{n+1/2}$ to solve the continuity, momentum and energy equations over the full time step Δt with ESDIRK2. If gravity is present and a target state $\tilde{\mathcal{Q}}$ is a priori known, then the well-balancing method described in Sect. 5.2.2 can be used. The results are the solutions for density, momentum and energy at the next step, ρ^{n+1}, $\rho\mathbf{v}^{n+1}$ and E_Φ^{n+1}.

4. Use ρ^{n+1}, $\rho\mathbf{v}^{n+1}$, E_Φ^{n+1} and $\mathbf{B}^{n+1/2}$ for solving the induction equation over $\Delta t/2$. This yields the magnetic field at the final step $\mathbf{B}^{n+1}$.

The proposed MHD scheme is extremely modular meaning that time-steppers as well as spatial reconstruction schemes and approximate Riemann solvers as proposed here can in principle be used in each sub-step of the algorithm. Additionally, the well-balancing method can easily be switched off by setting $\tilde{\mathcal{Q}} = 0$ if required.

5.4 Numerical Results

In order to assess the accuracy and performance of the newly proposed IESS method for solving the ideal MHD equations, we rely on a set of numerical experiments. Since the main purpose of the scheme is to be able to simulate flows at low sonic Mach numbers in strong stratifications, we decide not to show the typical tests commonly run by other MHD codes. These usually include shock tubes, supersonic vortices and magnetic blasts, which, however, are designed to test the shock-capturing capabilities of a numerical scheme. Instead, we run a series of verification benchmarks that are more suited for testing the low Mach properties of an MHD code.
In the first two tests (i,ii), we solve the homogeneous MHD equations. The convergence and scaling of the method is analyzed for the advection of a stable MHD vortex (i). We compare the results for the standard HLLD and the low-dissipation LHLLD solver in order to investigate the effect of the low Mach fix on the artificial dissipation. The simulations with the LHLLD solver are repeated in fully explicit mode using a SSP-RK2 method for time integration, which allows to quantify the speed-up of IESS as a function of the Mach number.
The ability of accurately evolving shear instabilities is fundamental in the context of simulations of turbulence, as they generate additional vorticity which leads to a cascade of energy. For this reason, we run simulations of a magnetized Kelvin-Helmholtz instability (ii). We follow the growth and evolution of the instability in a resolution study from low Mach to slightly subsonic regimes. A comparison between the standard HLLD and the low-dissipation LHLLD solver is performed to show the advantage of using low-dissipation fluxes over conventional methods in regimes of low Mach numbers.
In the third numerical experiment, we consider the influence of gravity (iii). The test setup consists of a MHSE to which we add a small perturbation in the pressure in order to validate the well-balancing method. To check the entropy conservation properties of the scheme, we model the rise of a parcel of fluid with higher entropy content than the (isentropic) background stratification, i.e. a "hot bubble". By changing the magnitude of the entropy perturbation, we simulate different rise velocities of the bubble, down to maximum Mach numbers of $\mathcal{M}_{\max} \sim 10^{-4}$. To quantify the magnitude of the numerical errors generated by an unbalanced stratification, we also simulate the rise of the bubble at $\mathcal{M}_{\max} \sim 10^{-2}$ without well-balancing.

For all of the following tests, an ideal gas EoS is used with $\gamma = 5/3$ except when specified otherwise. The time step for the IESS scheme is chosen according to the CFL condition (5.40) and reduced by 20% to get a more conservative stability criterion. Finally, unlimited linear reconstruction, which is second order accurate in space, is applied to primitive variables. Overall, the proposed scheme is (globally) second order accurate.

The method presented is implemented in the Seven-League Hydro (SLH) code [Mic13, Ede14] and all following tests are performed with this code.

5.4.1 Balsara Vortex

In order to check the convergence and to test the low Mach capabilities of the scheme, we consider a MHD vortex first described by [Bal04]. This is an exact stationary solution of the two-dimensional homogeneous ideal MHD equations in which the distribution of the centrifugal acceleration, magnetic tension, gas and magnetic pressure gradients is such that the vortex is stable. The spatial domain is $\mathcal{I} = [-5, 5]^2$ with periodic boundaries in both directions. The initial conditions are given by

$$\begin{aligned}
&\rho(x, y, 0) = 1, \\
&(u, v)(x, y, 0) = \tilde{V} e^{\frac{1-r^2}{2}} (-y, x), \\
&p(x, y, 0) = 1 + \left[\frac{\tilde{B}^2}{2}(1 - r^2) - \frac{\tilde{V}^2}{2} \right] e^{1-r^2}, \\
&(B_x, B_y)(x, y, 0) = \tilde{B} e^{\frac{1-r^2}{2}} (-y, x),
\end{aligned} \tag{5.41}$$

with $r^2 = x^2 + y^2$. $\tilde{V}$ is the maximum rotational velocity of the vortex and $\tilde{B}$ sets the value of the maximum Alfvén speed on the grid. The ratio $\beta_{\mathrm{K}} = \tilde{B}^2/\tilde{V}^2$ represents the ratio of magnetic to rotational kinetic energy, which is constant across the domain. To make this problem numerically more challenging, the vortex is advected along the diagonal of the computational grid with $|\mathbf{v}_{\mathrm{adv}}| = \tilde{V}$. The vortex is evolved for one advective crossing time $t_f = 10\sqrt{2}/\tilde{V}$ after which it returns to the initial position. In this time interval, the vortex rotates 2.25 times.

In order to check the convergence of the scheme in 2D, we run the vortex with $\tilde{V} = 10^{-3}$ (corresponding to $\mathcal{M}_{\mathrm{max}} = \max(\mathcal{M}_{\mathrm{loc}})_{t=0} = 1.55 \times 10^{-3}$) and $\beta_{\mathrm{K}} = 1$ at different resolutions. At the end of the simulation, the L^1-error is computed for each primitive variable $\mathcal{Q}_k^P$ as

$$L^1(\mathcal{Q}_k^P) = \frac{1}{N^2} \sum_{i,j} |Q_{k,i,j}^P(t = t_f) - Q_{k,i,j}^P(t = 0)|, \tag{5.42}$$

where i, j are the spatial indices. Fig. 5.2 shows the convergence of the L^1-error for different grids from $N = 32$ up to $N = 512$ cells per dimension. Convergence is second order for all primitive variables.

In a next step, we investigate the effect of the low Mach modification in LHLLD. Therefore, we re-run this last set of simulations with the standard version of HLLD. In Fig. 5.4 we show the final rotational kinetic energy distribution obtained with the two methods:

$$E_{\mathrm{R}} = \frac{1}{2}\rho \left[\left(u - \tilde{V}/\sqrt{2} \right)^2 + \left(v - \tilde{V}/\sqrt{2} \right)^2 \right]. \tag{5.43}$$

At low resolution, HLLD considerably stretches the vortex and a large fraction of kinetic energy is dissipated into internal energy. In contrast, simulations run with LHLLD show mild dissipation and dispersion errors are only visible at the lowest resolutions. All simulations converge with increasing resolution, but the kinetic energy conservation in the vortex simulated with HLLD is still two orders of magnitude worse than that obtained with LHLLD at the highest resolution considered in this study.

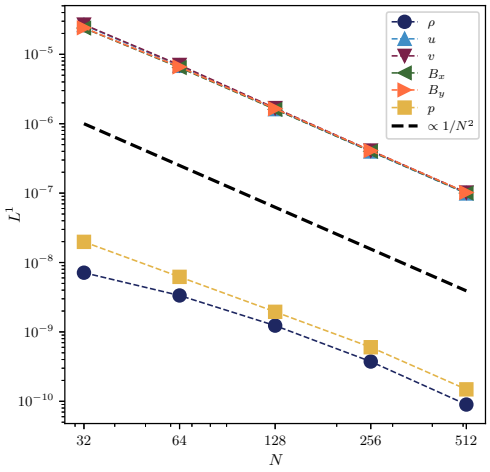

Figure 5.2: Convergence of the L^1-error in the Balsara vortex for each primitive variable as a function of resolution. For these simulations, the initial data are set such that the parameters are $\tilde{V} = 10^{-3}$ and $\beta_{\rm K} = 1$. The dashed black line is the second order scaling.

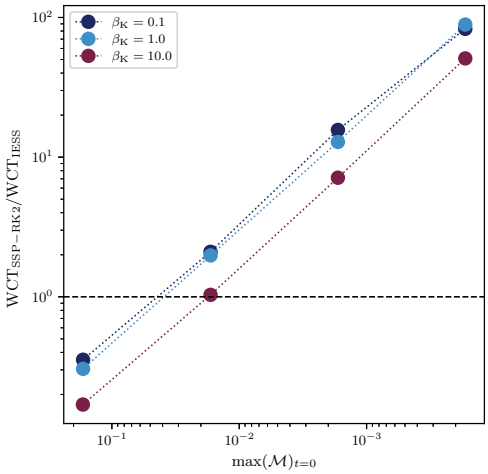

Figure 5.3: Ratio of the wall-clock times obtained with SSP-RK2 and IESS as a function of initial maximum sonic Mach number for different magnetic to (rotational) kinetic energy ratios. The dashed black line is drawn to represent same relative efficiency.

In order to study the behavior of the LHLLD solver in a wider range of subsonic regimes in both weakly and strongly magnetized fluids, we run the following grid of models

$$(\tilde{V}) \times (\beta_{\rm K}) = \left(10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\right) \times \left(10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}\right). \tag{5.44}$$

Given this choice of parameters, the initial maximum Mach number $\mathcal{M}$ ranges from 1.55×10^{-5} to 1.55×10^{-1}. Fig. 5.5 shows the magnetic energy distribution after one advective crossing time t_f computed on a 64 grid. Numerical dissipation converts a fraction of kinetic and magnetic energy into internal energy, but the shape of the vortex is well-preserved in all runs. The dissipation rate is virtually independent of $\mathcal{M}$. In contrast, dissipation of magnetic energy depends on the value of $\beta_{\rm K}$. As already pointed out in Rem. 5.2.1, the pressure-diffusion coefficient in LHLLD has a residual scaling $\mathcal{O}(1/\mathcal{M}_{\rm Alf})$. A larger value of $\beta_{\rm K}$ corresponds to lower $\mathcal{M}_{\rm Alf}$, which then increases the magnitude of the numerical dissipation. The velocity field is progressively more diffused out and becomes less efficient in sustaining the magnetic field through induction against numerical resistivity.

As explained in Sect. 5.3, one advantage of IESS is that it can employ longer time steps than fully time-explicit methods without sacrificing stability. However, a single step of the proposed scheme is much more expensive than a single step of a more standard time-explicit marching scheme, as a large nonlinear system has to be solved iteratively with a Raphson-Newton method. Because of these competing effects, we expect the IESS scheme to be more efficient than an explicit time-stepper below a certain Mach number. To determine this threshold, we run sets of simulations with the parameters

$$(\tilde{V}) \times (\beta_{\rm K}) = (10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}) \times (10^{-1}, 1, 10^{1}), \tag{5.45}$$

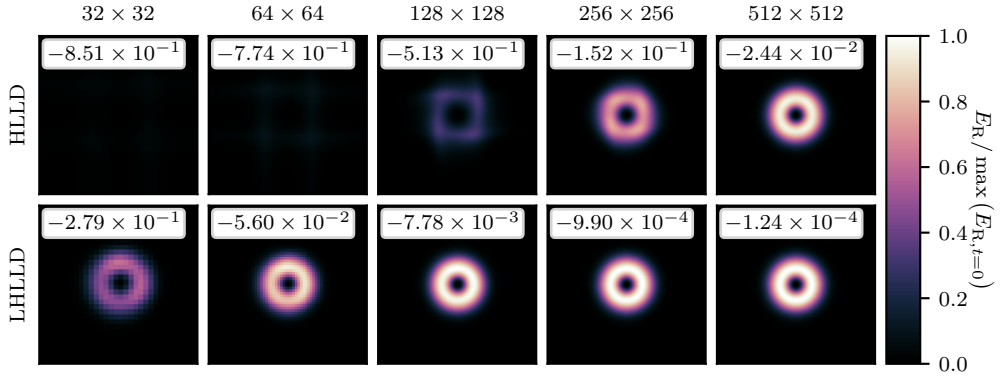

Figure 5.4: Distribution of the rotational kinetic energy (normalized by the maximum initial value) of the Balsara vortex with $\tilde{V} = 10^{-3}$ and $\beta_K = 1$ after one advective time t_f. The top panels show the vortices obtained with the HLLD flux function as a function of resolution, while the plots in the bottom panels are obtained with LHLLD. The insets show the fraction of rotational kinetic energy that has been dissipated by the end of the simulation: $(E_{R,t=t_f})_{tot}/(E_{R,t=0})_{tot} - 1$.

using both IESS and the explicit SSP-RK2[3] on 40×40 grid cells. Every other substep of the Godunov-type method (like the spatial reconstruction, the LHLLD flux function and constrained transport) remains unchanged, so the only difference is in the time discretization. At the end of each simulation, the ratio of the wall-clock times $\mathrm{WCT}_{\mathrm{SSP-RK2}}/\mathrm{WCT}_{\mathrm{IESS}}$ is taken as a measure of the relative efficiency between the marching schemes. The results are shown in Fig. 5.3. As expected, the speed-up of IESS increases as the Mach number of the vortex is decreased. The simulations with $\beta_K = 10$ are slower than the other cases, as the larger Alfvén speed considerably reduces the time step estimate in (5.40), while no significant difference is seen between $\beta_K = 0.1$ and $\beta_K = 1.0$. IESS overtakes SSP-RK2 at $\mathcal{M}_{max} \simeq 4 \times 10^{-2}$ for $\beta_K = (0.1, 1)$ and $\mathcal{M}_{max} \simeq 2 \times 10^{-2}$ for $\beta_K = 10$. At $\mathcal{M}_{max} = 10^{-3}$, IESS is ten to twenty times faster than SSP-RK2. This justifies the implementation efforts of a partially implicit time discretization algorithm for modeling slow flows.

5.4.2 Magnetized Kelvin-Helmholtz Instability

In this section, we run simulations of a magnetized version of the Kelvin-Helmholtz instability presented in Sect. 4.6.7. The spatial domain is again $\mathcal{I} = [0, 2] \times [-0.5, 0.5]$ with periodic boundary conditions and the initial conditions for density, velocity and pressure remain as given in (4.149)-(4.151). At this point, however, we do not transform the values into dimensionless quantities, but solve the dimensional system. Thus, the final time now depends on $\mathcal{M}$ and is reached for $t_{max} = 4.8/\mathcal{M}$. For the MHD equations, a uniform and horizontal initial magnetic field is added:

$$B_x(x, y, 0) = 0.1\mathcal{M}, \quad B_y(x, y, 0) = 0. \tag{5.46}$$

The initial data results in a minimum Alfvén Mach number $\mathcal{M}_{Alf,min} = 11.82$ for all values of $\mathcal{M}$. It is well-known that magnetic fields aligned with the shear flow have a stabilizing effect because they exert a restoring force on the perturbed interface [Cha61].

[3] For the time-explicit simulations, the CFL time step is reduced by 20%.

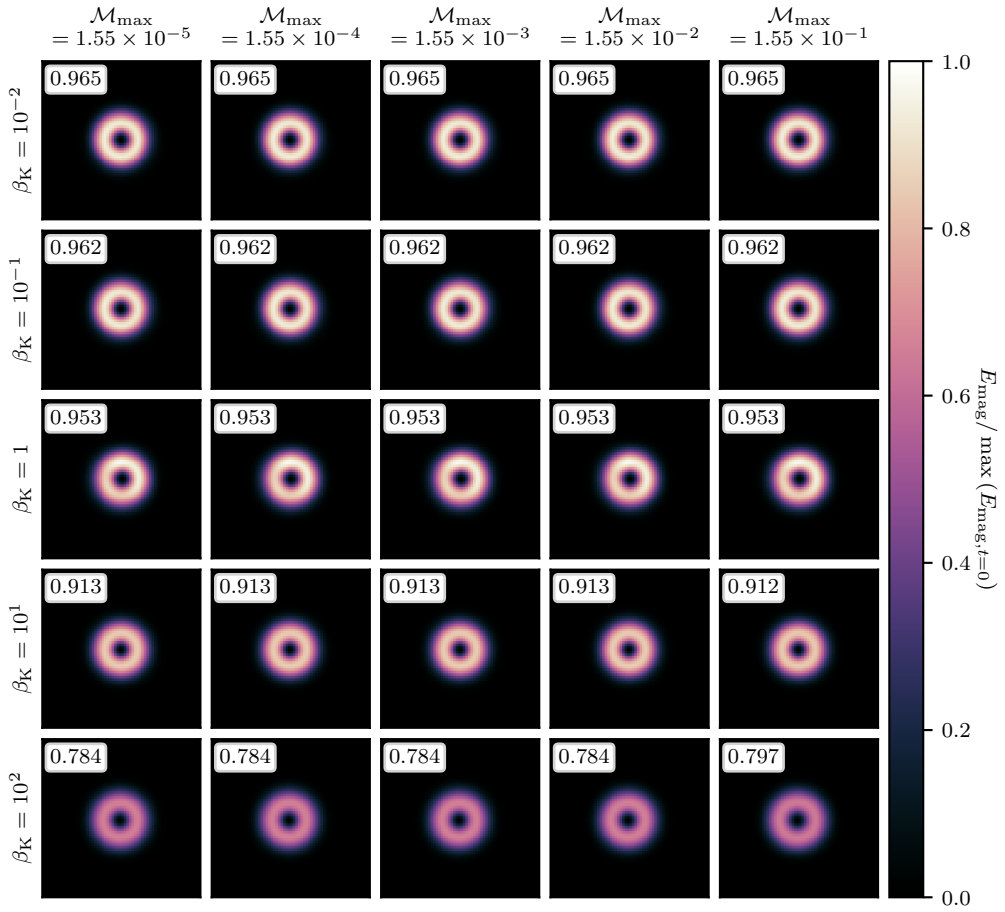

Figure 5.5: Magnetic energy distribution of the Balsara vortex after one advective crossing time t_f, normalized by the maximum magnetic energy at $t = 0$. The ratio of the magnetic to the (rotational) kinetic energy of the vortex is varied along the y-axis (in descending order), while the initial maximum rotational velocity $\tilde{V}$ varies along the x-axis. The inset in each subplot shows the ratio of the final to the initial magnetic energy. The vortex run with $\tilde{V} = 10^{-1}$ and $\beta_K = 10^2$ (bottom right corner) has a maximum Mach number of $\mathcal{M}_{\max} = 1.65 \times 10^{-1}$. In that system, the gas pressure drops in the regions around the center of the vortex to balance the large magnetic and centrifugal forces, which ultimately decreases the sound speed where the velocity is maximum.

With a field too strong, the instability may reach saturation when the flow is still essentially laminar or it may be suppressed completely. Instead, weak magnetic stresses do not considerably affect the initial growth of the instability, so the flow can develop the typical vortex structures present in the pure hydrodynamic case. This leads to a much more complex evolution in the nonlinear phase [FJRG96]. For this weak field regime, nearly laminar flows are expected only if $\min(\mathcal{M}_{\mathrm{Alf}})_{t=0} \lesssim 1.1$, as shown in Fig. 7.2.

The evolution of the Kelvin-Helmholtz instability is studied for a wide range of Mach numbers and grid resolutions:

$$(\mathcal{M}) \times (N) = \left(10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\right) \times (32, 64, 128, 256, 512, 1024)\,. \tag{5.47}$$

As done in the previous test, we compare the results obtained with both the HLLD and LHLLD solver. Fig. 5.6 and Fig. 5.7 show the time evolution of the y-direction kinetic energy $E_{\mathrm{kin},y} = \sum_{ij}(\rho v^2)_{ij}/2$ and the total magnetic energy $E_{\mathrm{mag}} = \sum_{ij} |\mathbf{B}_{ij}|^2/2$ for all the simulations considered in this study. Because of stretching and wrapping of the field lines within the vortices, the magnetic energy slowly increases with time at the expense of the kinetic energy content of the flow. After the primary rolls have reached the top and bottom boundaries ($t/t_{\mathrm{max}} \simeq 0.25$), $E_{\mathrm{kin},y}$ saturates due to the periodicity of the grid and starts to decrease. The secondary vortices keep winding up the magnetic field lines until Lorentz forces start to feedback on the velocity field, breaking down these inner structures. The two original shear interfaces get closer to each other (see Fig. 7.1) until a strong numerical reconnection event happens at $t/t_{\mathrm{max}} \simeq 0.45$, which violently decouples the primary rolls and causes a secondary peak in $E_{\mathrm{kin},y}$ at $t/t_{\mathrm{max}} \simeq 0.5$. After this time span, other reconnection events break up the flow into smaller structures, and both the magnetic and the kinetic energy are slowly dissipated away by the action of numerical resistivity and viscosity.

Since in this case we solve the ideal MHD equations, there is no characteristic scale on which magnetic and kinetic energy are dissipated into heat. So numerical effects play a significant role on progressively smaller scales at higher resolution. Thus, the amplification and dissipation of magnetic energy hardly converge for the resolutions considered in this study. The initial growth of $E_{\mathrm{kin},y}$, by contrast, is not much influenced by the initial weak field, and it is mostly determined by the strength of the shear flows and the width of the shear interface which is resolved. As a consequence, $E_{\mathrm{kin},y}$ converges until the major numerical reconnection event affects the velocity field. As shown in Fig. 5.6, the HLLD solver requires more resolution to reach convergence as the setup is run at progressively lower sonic Mach numbers. Eventually, the Mach number dependent dissipation term in the momentum flux (5.21) completely dominates the evolution of the flow and deteriorates the numerical solution. For this reason, at $\mathcal{M} = 10^{-4}$ we are able to successfully run with HLLD only the 64×32 and 128×64 grids, while for higher resolutions the nonlinear solver fails to converge.

The effects of numerical dissipation are also shown in Fig. 5.8, where the distributions of the sonic Mach number obtained with HLLD and LHLLD are compared at fixed resolution (128×64 cells) for different values of $\mathcal{M}$ at $t/t_{\mathrm{max}} = 1/6$. While in moderately subsonic regimes the large-scale structures in the flow are qualitatively similar, for lower Mach numbers HLLD introduces progressively more dissipation and the instability is eventually halted. When LHLLD is used instead, the morphology of the flow seems to be independent of the Mach number.

Finally, we perform a quantitative convergence study by computing the L^1-error associated with $E_{\mathrm{kin},y}$ at $t/t_{\mathrm{max}} = 1/6$. At this time, the first rolls have developed to considerable vertical wavelengths (see Fig. 5.8) so that the instability has already entered the nonlinear regime, and the flow is expected to converge as shown in Fig. 5.6. The L^1-error is computed against a reference solution that we take from the highest grid resolution runs considered in this test ($N = 1024$) using the LHLLD solver. All simulations (including the reference solutions) are down-sampled to a 64×32 grid so that the errors can directly be computed for different resolutions. This analysis is repeated for different values of $\mathcal{M}$ using both HLLD and LHLLD. The results are shown in Fig. 5.9. The errors are rescaled by $\mathcal{M}^2$ so that curves corresponding to different sonic Mach numbers lie on the same scale. Overall, the convergence is second order with N for all simulations. The LHLLD

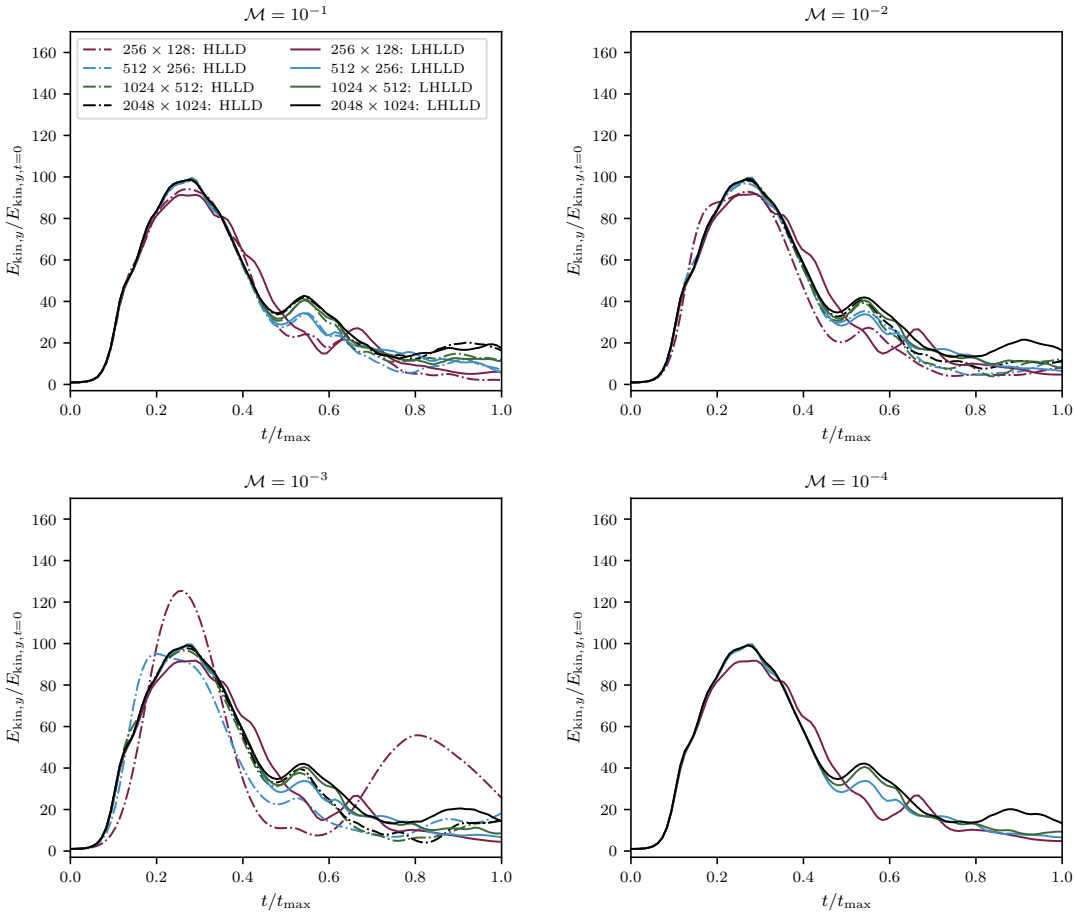

Figure 5.6: Time evolution of the y-direction kinetic energy rescaled by its initial value in the magnetized Kelvin-Helmholtz instability test problem. Each panel corresponds to a different initial Mach number $\mathcal{M}$. Different colors are used for different grid resolutions (the 64×32 and 128×64 grids cells have been left out for clarity). Results obtained with the HLLD solver are represented by dot-dashed lines, while solid lines are used for LHLLD. The black solid line in each panel is the reference solution. As explained in the text, the nonlinear solver does not converge when using HLLD at $\mathcal{M} = 10^{-4}$ for $N > 64$.

solver provides almost identical (rescaled) errors at given resolution in different regimes of Mach numbers. This is expected because the numerical dissipation introduced by this solver does not depend on $\mathcal{M}$, thanks to the low Mach modification in (5.23). Instead, the errors computed for the HLLD runs show a clear dependence on the sonic Mach number, and the errors get larger for slower flows. In particular, at $\mathcal{M} = 0.1$, HLLD needs approximately 1.2 times the resolution of LHLLD to achieve the same accuracy, which justifies the use of HLLD in this regime of Mach numbers. Instead, when $\mathcal{M} = 10^{-2}$ and $\mathcal{M} = 10^{-3}$, HLLD needs respectively twice or four times the resolution to be as accurate as the low-dissipation flux, which increases the amount of computing time by 8 or 64. Thus, the use of a low Mach approximate Riemann solver becomes indispensable for providing accurate results in regimes of low sonic Mach numbers with moderate grid resolutions, which would be unfeasible with more standard solvers.

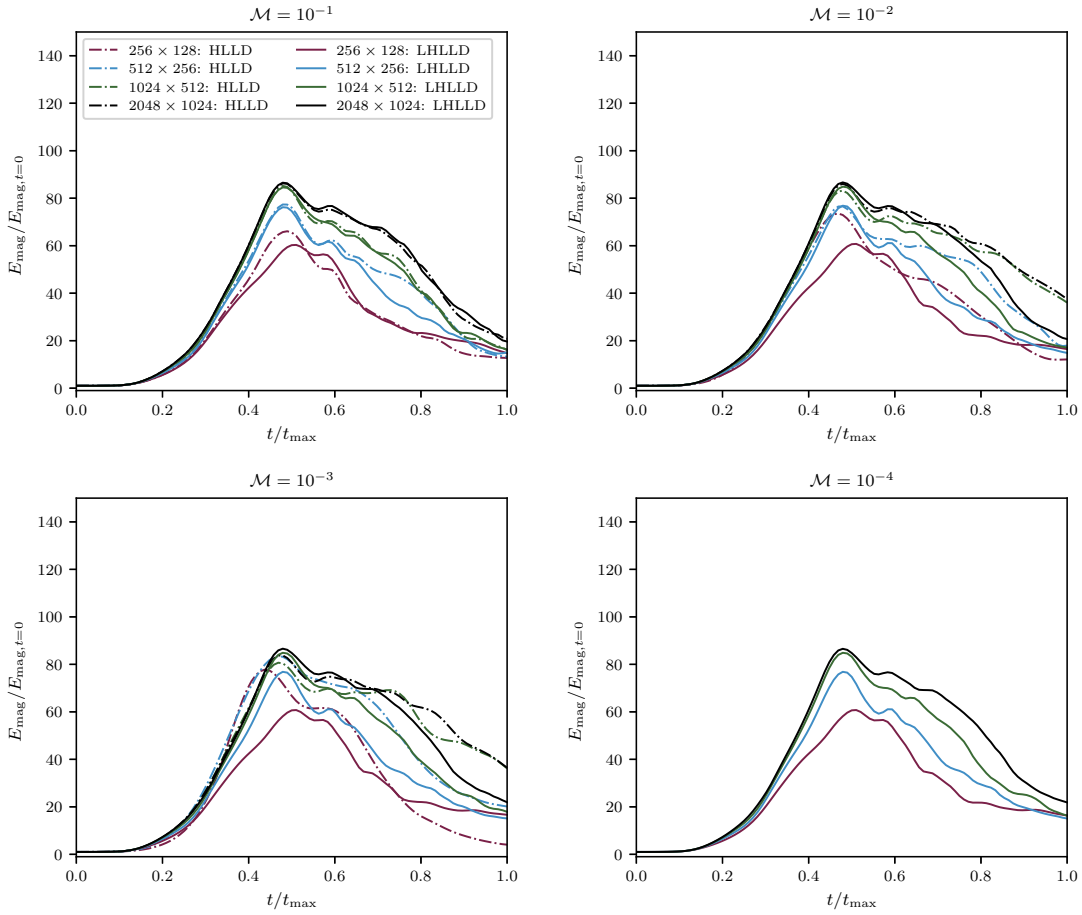

Figure 5.7: Same as Fig. 5.6 but showing the total magnetic energy divided by its initial value.

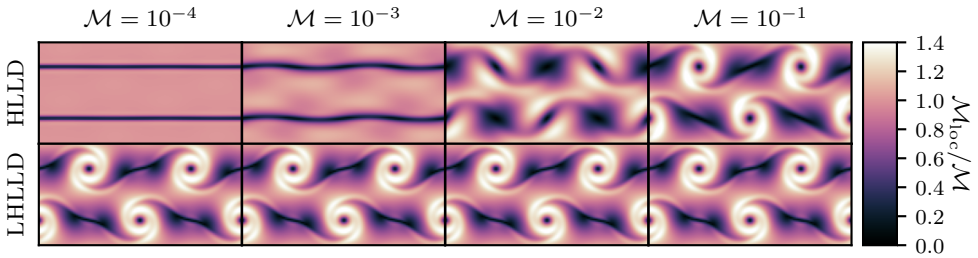

Figure 5.8: Distribution of the local sonic Mach number in the Kelvin-Helmholtz instability test at $t/t_{\rm max} = 1/6$ obtained with the HLLD (top panels) and the LHLLD (bottom panels) solvers on a 128×64 grid for different values of $\mathcal{M}$. All panels are rescaled by the corresponding value of $\mathcal{M}$.

5.4.3 Hot Bubble

Flows in deep stellar convection zones are usually characterized by the presence of slow parcels of fluid which move in a stratification that is unstable against convection. In the absence of volume heating and cooling processes, these packets of fluid preserve their

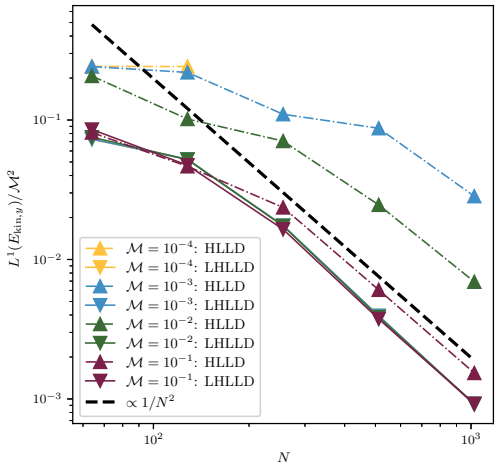

Figure 5.9: Convergence with resolution N of the L^1-error associated with $E_{\mathrm{kin},y}$ rescaled by $\mathcal{M}^2$ in the simulations of the Kelvin-Helmholtz instability. Different colors are used for different initial Mach numbers $\mathcal{M}$ using the LHLLD (solid lines) and HLLD (dot-dashed lines) solvers. The black dashed line is the second order scaling.

entropy content until they mix with the surroundings. Therefore, a numerical scheme designed to simulate such flows should have good entropy conservation properties. However, entropy conservation is hard to achieve if the density, temperature and pressure stratifications span several orders of magnitude and if the flows are very slow, since their entropy content would only be slightly higher/lower than the adiabatic surroundings[4]. Under these conditions, discretization errors caused by an imperfect balance of the background MHSE stratification can dominate the dynamics and deteriorate the numerical solution. The magnitude of such errors can be drastically reduced by using well-balancing techniques.

In this section, we check the entropy conservation properties of the new MHD scheme by running simulations of the "hot bubble" setup described by [EHB+21], where a bubble of higher entropy content with respect to the surroundings buoyantly rises in an adiabatic stratification. The physical domain is mapped on a 2D Cartesian grid ($N_x = 2/3 \times N_y$), and the background stratification is in MHSE. Boundary conditions are periodic everywhere and the gravitational acceleration takes the form

$$g_y(x, y) = g_0 \sin(k_y y), \tag{5.48}$$

where $g_0 = -1.09904373 \times 10^5 \, \frac{\mathrm{cm}}{\mathrm{s}^2}$, $k_y = 2\pi y / L_y$ is the maximum vertical wavelength and L_y is the vertical extent of the grid. The value of g_0 is set such that the ratio of the maximum to the minimum gas pressure[5] $p(x, y)$ is 100, which corresponds to 4.6 pressure scale heights. The entropy profile inside the bubble is given by

$$A = A_0 \left\{ 1 + \left(\frac{\Delta A}{A} \right)_{t=0} \cos\left(\frac{\pi}{2} \frac{r}{r_0} \right)^2 \right\}, \tag{5.49}$$

[4] Better entropy conservation properties can be achieved by directly evolving the specific entropy instead of E_Φ. However, this approach does not conserve the total energy.

[5] More details on how to compute the pressure profile can be found in [EHB+21].

where A_0 is background entropy, r_0 is the radius of the bubble, r is the distance from the center of the bubble and $(\Delta A/A)_{t=0}$ is the initial entropy perturbation. Fig. 5.10 shows exemplarily the initial perturbation for $(\Delta A/A)_{t=0} = 10^{-1}$. The density is

$$\rho(x, y, 0) = \left(\frac{p(x, y, 0)}{A}\right)^{1/\gamma}, \tag{5.50}$$

so that the (initial) buoyant acceleration of the bubble is proportional to the entropy perturbation,

$$a_\mathrm{b} = \frac{\Delta\rho}{\rho} g_y \propto \left(\frac{\Delta A}{A}\right)_{t=0}. \tag{5.51}$$

We run the models for the set of parameters

$$\left(\frac{\Delta A}{A}\right)_{t=0} \times (N_y) = \left(10^{-7}, 10^{-5}, 10^{-3}, 10^{-1}\right) \times (96, 192, 384, 768)\,, \tag{5.52}$$

and we set the maximum time such that in each run the bubble rises approximately the same distance l. For our simulations we use

$$t_f = \frac{30}{\sqrt{10\left(\frac{\Delta A}{A}\right)_{t=0}}}. \tag{5.53}$$

This allows us to simulate different regimes of sonic Mach numbers, as the velocity V reached by the bubble over a length l scales as

$$V \propto (a_\mathrm{b} l)^{1/2}. \tag{5.54}$$

This ultimately leads to the relation

$$\mathcal{M} \propto \left(\frac{\Delta A}{A}\right)^{1/2}_{t=0}. \tag{5.55}$$

We add a uniform horizontal magnetic field such that its strength is rescaled depending on the entropy perturbation,

$$B_x(x, y, 0) = B_0 \left(\frac{\Delta A}{A}\right)^{1/2}_{t=0}. \tag{5.56}$$

This ensures that the relative magnitude of magnetic stresses compared to the ram pressure of the bubble remains the same for all simulations, and that the morphology of the flow is unaltered. $B_0 = 47.3$ is chosen in a way that the final Alfvén Mach number at the position of largest entropy is in the range $\mathcal{M}_\mathrm{Alf} \simeq 2 - 3$ depending on the grid resolution. Thus, magnetic fields are dynamically important but not strong enough to suppress buoyancy.

In Fig. 7.4 we show the final entropy excess for all the simulations run in the parameter study. The center of the bubble accelerates faster than other regions as it is the point with maximum entropy, and the acceleration profile across the bubble leads to the development of shear at its outer edges. As the bubble rises in the stratification, the magnetic field lines are stretched into thin tubes, which locally amplifies the magnetic energy (see Fig. 7.5). The amount of amplification depends on the numerical resistivity and so on resolution. In contrast to the pure hydrodynamic case studied by [EHB+21], here the

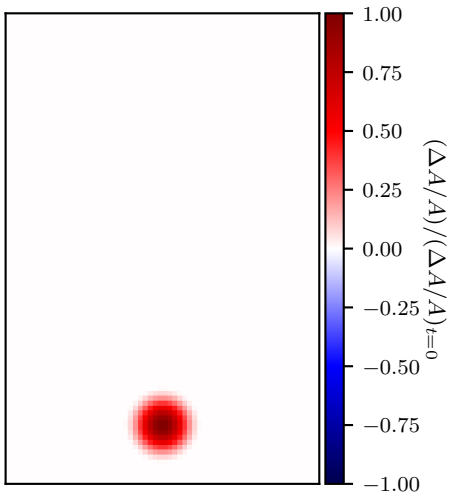

Figure 5.10: Initial entropy perturbation in the hot bubble setup exemplarily illustrated for $(\Delta A/A)_{t=0} = 10^{-1}$ on a 64×96 grid.

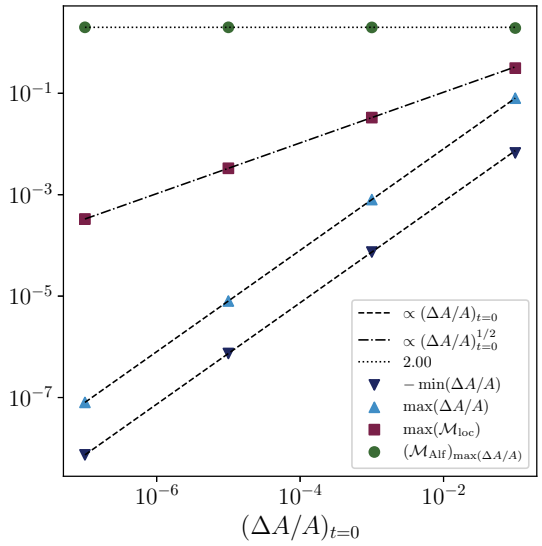

Figure 5.11: Maximum Mach number, minimum and maximum entropy fluctuations and Alfvén speed of the hot bubble as a function of the initial entropy perturbation. The black lines represent the physical scalings.

presence of a magnetic field suppresses the formation of vortices at the sides of the bubble. Overall, the entropy content of the bubble is well-preserved even on the coarsest grid, but some negative entropy fluctuations are present at the very top of the bubble. These negative fluctuations are numerical artifacts. In fact, the entropy fluctuations may locally increase, as a fraction of magnetic and kinetic energy is dissipated into internal energy, but the fluctuations cannot become negative physically. These artifacts do not depend on the entropy perturbation, and they are limited to a very narrow region in the spatial domain which tends to shrink as the resolution is increased. All models converge upon grid refinement.

According to equation (5.55), the sonic Mach number of the bubble is expected to scale as the square root of the initial entropy perturbation. Any deviation from this relation, which has been obtained on the basis of physical arguments, can be due to difficulties in modeling slow flows in a stratified setup and the build-up of significant numerical errors. In Fig. 5.11 we show this scaling for the coarsest grid resolution. All data points overlap with the theoretical curve, and the minimum local Mach number $\mathcal{M}_{\rm loc}$ achieved in this parameter study is 3.32×10^{-4} (see also Fig. 7.6). The ratio of the rising velocity of the bubble to the Alfvén speed (in the point of maximum entropy) does not depend on the amplitude of the entropy perturbation. Since the initial magnetic field is proportional to $(\Delta A/A)_{t=0}^{1/2}$, the amount of amplification due to induction only depends on the velocity of the bubble V and the time scale over which magnetic induction operates ($\propto 1/V$).

Finally, to quantify the strength of the spurious flows that are expected to arise if the stratification is left unbalanced, in Fig. 5.12 we show a comparison between simulations obtained with and without Deviation Well-Balancing, where the vertical resolution N_y

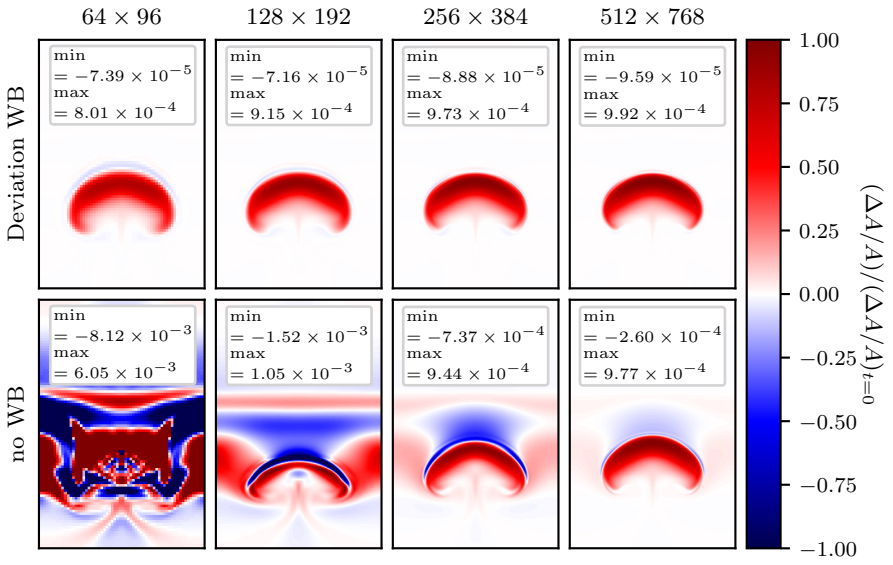

Figure 5.12: Final distribution of the entropy fluctuations of the hot bubble at time $t_f = 300$ for $(\Delta A/A)_{t=0} = 10^{-3}$ at different grid resolutions. The entropy fluctuations are rescaled by $(\Delta A/A)_{t=0}$. The top row shows the results obtained with Deviation Well-Balancing, whereas no well-balancing method was used in the simulations shown in the bottom row. The insets show the minimum and maximum values of the entropy fluctuation in each panel.

ranges from 96 to 768. For this comparison, we fix $(\Delta A/A)^{1/2}_{t=0} = 10^{-3}$ such that the maximum sonic Mach number of the bubble is approximately 3×10^{-2}. The well-balanced method manages to accurately resolve the rise of the bubble already on very coarse grids and its solutions exhibit only narrow regions on top of the bubble with negative entropy fluctuations. The unbalanced simulations, however, develop large entropy fluctuations, both negative and positive, which strongly deteriorate the numerical solution. The shape of the bubble can hardly be identified on the coarsest grid with 64×96 cells. As the grid is refined, the simulations tend to converge, but wide regions of negative entropy fluctuations are still present even on the finest grid. As a result, this test demonstrates that well-balancing techniques are fundamental to correctly simulate the evolution of small entropy perturbations in steep isentropic stratifications and to reduce the effects of numerical errors when using moderately coarse grids.

5.5 Summary and Conclusions

In this chapter we have presented a new FV scheme to solve the fully compressible MHD equations with gravity in regimes of low sonic Mach numbers and high-β environments. This method relies on a modified version of the HLLD Riemann solver called LHLLD [MM21] to avoid the excessive numerical dissipation typical of high-resolution, shock-capturing solvers in the low Mach number regime. The strict Mach number dependent CFL condition on the time step is overcome by using an implicit-explicit time discretization algorithm for which the induction equation is integrated by using an explicit time-

stepper while the rest of the MHD system is integrated implicitly. The solutions to the two subsets of equations are coupled through Strang splitting following the prescription of [FMR09]. The combined marching scheme IESS has a less restrictive CFL condition. The time step is limited only by the fastest fluid and Alfvén speeds instead of the fast magnetosonic speed, and therefore does not depend on the sonic Mach number. Whenever required, a magnetohydrostatic solution can be enforced on the discrete grid with the Deviation Well-Balancing method [BCK21, EHB+21]. This technique leads to better entropy conservation properties of the numerical scheme, even in cases where the pressure and density stratifications span several orders of magnitude across the computational domain. Finally, the solenoidal constraint is enforced using the Contact-CT method [GS05]. Numerical experiments show that the new method is second order accurate. The results of the Balsara vortex and the Kelvin-Helmholtz instability underline the positive effect of the low-dissipation LHLLD Riemann solver, as it does not introduce excessive artificial dissipation in the simulations characterized by low Mach numbers. Measuring the wall-clock time shows that the IESS time-marching approach leads to efficiency advantages over fully time-explicit methods in regimes with Mach numbers below $(2 - 5) \times 10^{-2}$. The Deviation Well-Balancing method significantly reduces discretization errors arising from the background stratification in the "hot bubble" setup. Compared to a non-well-balanced method, the rise of the bubble can be simulated more accurately and unphysical negative entropy fluctuations have a smaller magnitude and are limited to narrow regions. Overall, the results obtained in these tests demonstrate that the proposed numerical method can accurately and efficiently cope with a variety of MHD processes that are relevant in stellar interiors, but are in regimes that are inaccessible to conventional FV methods.

Chapter 6

A Semi-Implicit IMEX Method

The stiffness of the MHD system in the low Mach number regime is essentially caused by the acoustic pressure term in the flux. Therefore, it is not necessary to use an implicit discretization for the whole system. Instead, in this chapter we only treat terms associated with the acoustic pressure implicitly. All remaining terms are part of an explicit sub-system. For this purpose, a time integration inspired by the class of IMEX schemes is used [BFR16, BP21]. The underlying CFL condition of the method is only coupled to the convective sub-system and therefore independent of the Mach number. Moreover, the implicit part is small, easy to invert and does not require to solve large nonlinear systems of equations as in fully implicit methods, which reduces computational costs. Furthermore, no numerical dissipation is embedded in the implicit solver, preventing excessive numerical dissipation in the low Mach number regime. This procedure builds on comparable methods for the Euler equations [BDL+20] and Navier-Stokes equations [BDT21]. Within a finite difference framework, it has also been applied to the homogeneous MHD equations [CWX23]. Our work is strictly related to what is presented in [CWX23], however there are some important differences: i) we use a finite volume discretization for the convective terms; ii) no numerical dissipation is added to the implicit part even in the case of shock waves; iii) we also include gravitational source terms.
In the splitting, the source term is added to the explicit sub-system, because it has no direct effect on the CFL condition and the numerical dissipation. More challenging is its impact on (magneto-)hydrostatic solutions. We combine the semi-implicit scheme with the *Deviation Well-Balancing method* [BCK21] which helps the scheme to exactly preserve a priori known equilibria and significantly reduces discretization errors arising from background stratifications. A similar approach has been recently used in [GCD21] in the context of general relativity. The solenoidal constraint is numerically treated by the *Contact CT method* [GS05]. The CT method corrects the magnetic field directly after the explicit step as the induction equations are entirely assigned to the explicit sub-system of the splitting. By correcting the magnetic field, a specific discrete definition of divergence is kept within machine precision, thereby increasing the stability of the overall scheme.

This chapter is structured as follows. In Sect. 6.1, we rewrite the MHD equations and propose a splitting into convective and pressure sub-systems. On the basis of this splitting, we develop a second order accurate semi-implicit and well-balanced numerical scheme in Sect. 6.2. The properties of this scheme are investigated in numerical experiments in Sect. 6.3, including setups in the low Mach number regime and near (magneto-)hydrostatic equilibria. Finally, Sect. 6.4 concludes with a summary of the method and an outlook on future developments. The results of this chapter can also be found in [BBK24].

6.1 Governing Equations

In order to enforce the flux splitting, it is useful to rewrite the MHD system (2.66) in a more convenient form. To do so we split the total energy E into internal (ρe), kinetic (ρk) and magnetic energy (m), i.e.

$$E = \rho e + \rho k + m, \qquad \rho e = \frac{p}{\gamma - 1}, \quad \rho k = \frac{1}{2}\rho \mathbf{v}^2, \quad m = \frac{1}{2}|\mathbf{B}|^2. \tag{6.1}$$

Then the flux in the energy equation is reformulated in terms of the specific enthalpy $h = e + p/\rho$ by

$$\left(E + p + \frac{1}{2}|\mathbf{B}|^2\right)\mathbf{v} - \mathbf{B}\,(\mathbf{B}\cdot\mathbf{v}) = \mathbf{v}(\rho k + \rho h + 2m) - \mathbf{B}(\mathbf{B}\cdot\mathbf{v}). \tag{6.2}$$

Using this new form of the energy equation, we can write the MHD equations in the equivalent form

$$\frac{\partial \mathcal{Q}}{\partial t} + \frac{\partial \mathcal{F}_1(\mathcal{Q})}{\partial x} + \frac{\partial \mathcal{F}_2(\mathcal{Q})}{\partial y} + \frac{\partial \mathcal{F}_3(\mathcal{Q})}{\partial z} = \mathcal{S}(\mathcal{Q}), \tag{6.3}$$

with the state vector $\mathcal{Q}$, the flux in x-direction $\mathcal{F}_1(\mathcal{Q})$ and the source $\mathcal{S}(\mathcal{Q})$ that explicitly write

$$\mathcal{Q} = \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ \rho w \\ E \\ B_x \\ B_y \\ B_z \end{pmatrix}, \quad \mathcal{F}_1(\mathcal{Q}) = \begin{pmatrix} \rho u \\ \rho u^2 + p + m - B_x^2 \\ \rho uv - B_x B_y \\ \rho uw - B_x B_z \\ u(\rho k + \rho h + 2m) - B_x(\mathbf{v}\cdot\mathbf{B}) \\ 0 \\ uB_y - vB_x \\ uB_z - wB_x \end{pmatrix}, \quad \mathcal{S}(\mathcal{Q}) = \begin{pmatrix} 0 \\ \rho g_x \\ \rho g_y \\ \rho g_z \\ \rho \mathbf{v}\cdot\mathbf{g} \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{6.4}$$

The fluxes $\mathcal{F}_2(\mathcal{Q})$ and $\mathcal{F}_3(\mathcal{Q})$ can be expressed in similar forms.

6.1.1 Flux Splitting

In the low Mach number limit, the sound speed becomes very high compared to the fluid velocity, hence the terms related to the pressure are dominant. Consequently, large values of the fast and slow magnetosonic wave speeds are retrieved (see also Sect. 2.4.5). As analyzed in Sect. 3.10.1, fully explicit FV methods suffer from both an excessive amount of numerical dissipation, which is proportional to the largest absolute eigenvalue, and a drastic reduction of the admissible time step Δt to ensure stability under a classical CFL condition of the type

$$\Delta t \leq \mathrm{CFL} \min_{\mathcal{I}} \frac{\min(\Delta x, \Delta y, \Delta z)}{\max|\lambda^{MHD}|}. \tag{6.5}$$

Therefore, we propose to discretize the pressure gradient in the momentum equation and the enthalpy term in the energy equation implicitly, while keeping an explicit discretization for the nonlinear convective fluxes and the terms related to the magnetic field. To achieve that aim, let the fluxes in x-direction be split into a convective-type flux $\mathcal{F}^c(\mathcal{Q})$ and a

pressure-type flux $\mathcal{F}^p(\mathcal{Q})$, that is

$$\mathcal{F}_1^c(\mathcal{Q}) = \begin{pmatrix} \rho u \\ \rho u^2 + m - B_x^2 \\ \rho uv - B_x B_y \\ \rho uw - B_x B_z \\ u(\rho k + 2m) - B_x(\mathbf{v}\cdot\mathbf{B}) \\ 0 \\ uB_y - vB_x \\ uB_z - wB_x \end{pmatrix}, \quad \mathcal{F}_1^p(\mathcal{Q}) = \begin{pmatrix} 0 \\ p \\ 0 \\ 0 \\ h\rho u \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{6.6}$$

We obtain the following two sub-systems with the corresponding eigenvalues:

- Convective sub-system:

$$\frac{\partial \mathcal{Q}}{\partial t} + \frac{\partial \mathcal{F}_1^c}{\partial x} = \mathcal{S}, \tag{6.7a}$$

$$\lambda_{1,8}^c = u \pm \sqrt{\frac{\mathbf{B}^2}{\rho}}, \quad \lambda_{2,7}^c = u \pm \frac{B_x}{\sqrt{\rho}}, \quad \lambda_{3,4,5,6}^c = 0. \tag{6.7b}$$

- Pressure sub-system:

$$\frac{\partial \mathcal{Q}}{\partial t} + \frac{\partial \mathcal{F}_1^p}{\partial x} = \mathbf{0}, \tag{6.8a}$$

$$\lambda_1^p = \frac{1}{2}\left(u - \sqrt{u^2 + 4c^2}\right), \quad \lambda_{2,3,4,5,6,7}^p = 0, \quad \lambda_8^p = \frac{1}{2}\left(u + \sqrt{u^2 + 4c^2}\right). \tag{6.8b}$$

The fluxes $\mathcal{F}_2$ and $\mathcal{F}_3$ can be split in a similar way with analogous sub-systems and eigenvalues. In the above splitting we have added the source term to the convective sub-system, since it does not introduce a Mach number dependence in the eigenvalues and thus does not pose a numerical problem in the low Mach limit. By treating the pressure sub-system implicitly, the maximum admissible time step of the scheme only depends on the explicit sub-system

$$\Delta t \leq \mathrm{CFL} \min_{\mathcal{I}} \frac{\min(\Delta x, \Delta y, \Delta z)}{\max|\lambda^c|}, \tag{6.9}$$

hence making the scheme particularly well-suited for low Mach number flows ($\mathcal{M} \ll 1$). On the other hand, for strongly convected flows with shocks, the convective eigenvalues in the computation of the time step ensure stability.

6.2 Numerical Scheme

For reasons of clarity and comprehensibility we restrict ourselves in the following part to one spatial dimension before returning to three spatial dimensions in Sect. 6.2.4.

6.2.1 First Order Semi-Discrete Scheme in Time

In order to construct a numerical method that is well-balanced in the sense that it discretely preserves MHSE of the form (2.79) in an exact way, we rely on the Deviation Well-Balancing method described in Sect. 5.2.2. We thus assume that we know a priori an equilibrium solution $\tilde{\mathcal{Q}}$ that satisfies (2.79). As usual in this approach we then do not

discretize (6.3), but do so for the following equation for the deviation of the equilibrium solution

$$\frac{\partial \Delta\mathcal{Q}}{\partial t} + \frac{\partial \mathcal{F}(\Delta\mathcal{Q}+\tilde{\mathcal{Q}})}{\partial x} - \frac{\partial \mathcal{F}(\tilde{\mathcal{Q}})}{\partial x} = \mathcal{S}(\Delta\mathcal{Q}+\tilde{\mathcal{Q}}) - \mathcal{S}(\tilde{\mathcal{Q}}). \tag{6.10}$$

As a consequence, we apply the splitting into convective and pressure parts defined in (6.6) to this equation, i.e.

$$\begin{aligned}&\frac{\partial \Delta\mathcal{Q}}{\partial t} + \frac{\partial \mathcal{F}^c(\Delta\mathcal{Q}+\tilde{\mathcal{Q}})}{\partial x} - \frac{\partial \mathcal{F}^c(\tilde{\mathcal{Q}})}{\partial x} + \frac{\partial \mathcal{F}^p(\Delta\mathcal{Q}+\tilde{\mathcal{Q}})}{\partial x} - \frac{\partial \mathcal{F}^p(\tilde{\mathcal{Q}})}{\partial x} \\ &= \mathcal{S}(\Delta\mathcal{Q}+\tilde{\mathcal{Q}}) - \mathcal{S}(\tilde{\mathcal{Q}}).\end{aligned} \tag{6.11}$$

The semi-discrete form of the convective sub-system writes

$$\Delta\mathcal{Q}^* = \Delta\mathcal{Q}^n - \Delta t\frac{\partial}{\partial x}\mathcal{F}^c(\Delta\mathcal{Q}^n+\tilde{\mathcal{Q}}) + \Delta t\frac{\partial}{\partial x}\mathcal{F}^c(\tilde{\mathcal{Q}}) + \Delta t\mathcal{S}(\Delta\mathcal{Q}^n+\tilde{\mathcal{Q}}) - \Delta t\mathcal{S}(\tilde{\mathcal{Q}}). \tag{6.12}$$

To simplify the notation, we introduce $\bar{\mathcal{Q}}^n = \Delta\mathcal{Q}^n + \tilde{\mathcal{Q}}$. Broken up into the individual components, system (6.12) results in

$$\Delta\rho^* = \Delta\rho^n - \Delta t\frac{\partial}{\partial x}\left(\bar{\rho}\bar{u}\right)^n + \Delta t\frac{\partial}{\partial x}(\tilde{\rho}\tilde{u}), \tag{6.13a}$$

$$\begin{aligned}(\Delta\rho u)^* = (\Delta\rho u)^n &- \Delta t\frac{\partial}{\partial x}\left(\bar{\rho}\bar{u}^2 + \bar{m} - \bar{B}_x^2\right)^n + \Delta t\frac{\partial}{\partial x}\left(\tilde{\rho}\tilde{u}^2 + \tilde{m} - \tilde{B}_x^2\right) \\ &+ \Delta t\left(\bar{\rho}g_x\right)^n - \Delta t\left(\tilde{\rho}g_x\right),\end{aligned} \tag{6.13b}$$

$$(\Delta\rho v)^* = (\Delta\rho v)^n - \Delta t\frac{\partial}{\partial x}\left(\bar{\rho}\bar{u}\bar{v} - \bar{B}_x\bar{B}_y\right)^n + \Delta t\frac{\partial}{\partial x}\left(\tilde{\rho}\tilde{u}\tilde{v} - \tilde{B}_x\tilde{B}_y\right), \tag{6.13c}$$

$$(\Delta\rho w)^* = (\Delta\rho w)^n - \Delta t\frac{\partial}{\partial x}\left(\bar{\rho}\bar{u}\bar{w} - \bar{B}_x\bar{B}_z\right)^n + \Delta t\frac{\partial}{\partial x}\left(\tilde{\rho}\tilde{u}\tilde{w} - \tilde{B}_x\tilde{B}_z\right), \tag{6.13d}$$

$$\begin{aligned}(\Delta E)^* = (\Delta E)^n &- \Delta t\frac{\partial}{\partial x}\left(\bar{u}(\bar{\rho}\bar{k} + 2\bar{m}) - \bar{B}_x(\bar{\mathbf{v}}\cdot\bar{\mathbf{B}})\right)^n \\ &+ \Delta t\frac{\partial}{\partial x}\left(\tilde{u}(\tilde{\rho}\tilde{k} + 2\tilde{m}) - \tilde{B}_x(\tilde{\mathbf{v}}\cdot\tilde{\mathbf{B}})\right) + \Delta t\left(\bar{\rho}\bar{u}g_x\right)^n - \Delta t\left(\tilde{\rho}\tilde{u}g_x\right),\end{aligned} \tag{6.13e}$$

$$\Delta B_x^* = 0, \tag{6.13f}$$

$$\Delta B_y^* = \Delta B_y^n - \Delta t\frac{\partial}{\partial x}\left(\bar{u}\bar{B}_y - \bar{v}\bar{B}_x\right)^n + \Delta t\frac{\partial}{\partial x}\left(\tilde{u}\tilde{B}_y - \tilde{v}\tilde{B}_x\right), \tag{6.13g}$$

$$\Delta B_z^* = \Delta B_z^n - \Delta t\frac{\partial}{\partial x}\left(\bar{u}\bar{B}_z - \bar{w}\bar{B}_x\right)^n + \Delta t\frac{\partial}{\partial x}\left(\tilde{u}\tilde{B}_z - \tilde{w}\tilde{B}_x\right). \tag{6.13h}$$

In the above equations, we include terms containing $\tilde{u}$, $\tilde{v}$ or $\tilde{w}$. In the following, however, we will omit these terms since the equilibrium velocities are zero. The definitions presented above are employed to obtain a first order semi-implicit time discretization [BFR16, BP21,

BT22] of (6.10), which is defined by

$$\Delta\rho^{n+1} = \Delta\rho^*, \tag{6.14a}$$

$$(\Delta\rho u)^{n+1} = (\Delta\rho u)^* - \Delta t\frac{\partial}{\partial x}\left(p^{n+1} - \tilde{p}\right), \tag{6.14b}$$

$$(\Delta\rho v)^{n+1} = (\Delta\rho v)^*, \tag{6.14c}$$

$$(\Delta\rho w)^{n+1} = (\Delta\rho w)^*, \tag{6.14d}$$

$$(\Delta E)^{n+1} = (\Delta E)^* - \Delta t\frac{\partial}{\partial x}\left(h^n(\rho u)^{n+1}\right), \tag{6.14e}$$

$$\Delta B_x^{n+1} = \Delta B_x^*, \tag{6.14f}$$

$$\Delta B_y^{n+1} = \Delta B_y^*, \tag{6.14g}$$

$$\Delta B_z^{n+1} = \Delta B_z^*. \tag{6.14h}$$

Here, we use the known enthalpy h^n at time level n in the energy equation to avoid nonlinear terms in the implicit part, which is different from the schemes proposed in [DBTF19, Fam21]. Additionally, we have made use of the fact that in the implicit fluxes we only work with point values at the cell centers, i.e. we use an implicit finite difference discretization. Thus, we can write

$$\left(\Delta Q + \tilde{Q}\right)_i = Q_i - \tilde{Q}_i + \tilde{Q}_i = Q_i. \tag{6.15}$$

Therefore, the implicit fluxes are simplified by

$$\frac{\partial}{\partial x}\left(\Delta p^{n+1} + \tilde{p}\right) - \Delta t\frac{\partial}{\partial x}\tilde{p} = \frac{\partial}{\partial x}\left(p^{n+1} - \tilde{p}\right), \tag{6.16}$$

$$\frac{\partial}{\partial x}\left(\left(\Delta h^n + \tilde{h}\right)\left((\Delta\rho u)^{n+1} + \tilde{\rho}\tilde{u}\right)\right) = \frac{\partial}{\partial x}\left(h^n(\rho u)^{n+1}\right). \tag{6.17}$$

According to the definition given in the energy equation (6.1), we split the deviation of the total energy at the new time level as follows

$$(\Delta E)^{n+1} = (\Delta\rho e)^{n+1} + (\Delta\rho k)^{n+1} + (\Delta m)^{n+1}. \tag{6.18}$$

The deviation of the kinetic energy therein is set to

$$(\Delta\rho k)^{n+1} = (\rho k)^{n+1} - (\tilde{\rho}\tilde{k}) = \frac{1}{2}\frac{(\rho u)^n}{\rho^{n+1}}(\rho u)^{n+1} - \frac{1}{2}\tilde{\rho}\tilde{u}^2, \tag{6.19}$$

where a semi-implicit strategy is adopted for the term $(\rho k)^{n+1}$. Indeed, the splitting of the momentum contribution into an explicit and an implicit part is again done in order to avoid nonlinear implicit terms. The density ρ^{n+1} and the deviation in the magnetic energy Δm^{n+1} are known because both continuity and induction equations are fully explicit. With the help of the ideal gas law, the deviation in the internal energy can be expressed in terms of the pressure

$$(\Delta\rho e)^{n+1} = \frac{p^{n+1} - \tilde{p}}{\gamma - 1}. \tag{6.20}$$

In order to derive a preliminary discretization of the total energy equation we transform (6.14b) into

$$(\rho u)^{n+1} = (\Delta\rho u)^* + \tilde{\rho}\tilde{u} - \Delta t\frac{\partial}{\partial x}\left(p^{n+1} - \tilde{p}\right), \tag{6.21}$$

and insert the result into the energy equation (6.14e). This leads to an elliptic equation for the pressure

$$\frac{p^{n+1}}{\gamma-1} - \Delta t \frac{(\rho u)^n}{2\rho^{n+1}} \frac{\partial}{\partial x}\left(p^{n+1} - \tilde{p}\right) - \Delta t^2 \frac{\partial}{\partial x}\left(h^n \frac{\partial}{\partial x}\left(p^{n+1} - \tilde{p}\right)\right) = b^n, \tag{6.22}$$

with the known right-hand side given by

$$\begin{aligned} b^n = \frac{\tilde{p}}{\gamma-1} + \frac{1}{2}\tilde{\rho}\tilde{u}^2 + (\Delta E)^* - \frac{(\rho u)^n}{2\rho^{n+1}}\left((\Delta \rho u)^* + \tilde{\rho}\tilde{u}\right) - \Delta m^{n+1} \\ - \Delta t \frac{\partial}{\partial x}\left(h^n\left((\Delta \rho u)^* + \tilde{\rho}\tilde{u}\right)\right). \end{aligned} \tag{6.23}$$

The pressure equation (6.22) constitutes a linear system for the scalar unknown p^{n+1} that is solved using the iterative *Generalized Minimal Residual Method* (GMRES) solver [SS86] up to a prescribed tolerance (we typically set tol $= 10^{-12}$). Different from [DBTF19, Fam21], this approach does not need any fixed point method thanks to the semi-implicit splitting of the enthalpy flux and the kinetic energy in the energy equation. Once the new pressure is known, the deviation in the momentum $(\Delta\rho u)^{n+1}$ is updated with (6.14b), and then the deviation in the total energy is updated using (6.14e). Notice that the scheme is written in flux form, therefore it is locally and globally conservative.

6.2.2 Discrete Spatial Operators

The convective sub-system (6.12) is discretized by a Godunov-type FV method which writes

$$\begin{aligned} \Delta Q_i^* = &\Delta Q_i^n - \frac{\Delta t}{\Delta x}(F^c_{i+1/2} - F^c_{i-1/2}) + \frac{\Delta t}{\Delta x}(\mathcal{F}^c(\tilde{Q}_{i+1/2}) - \mathcal{F}^c(\tilde{Q}_{i-1/2})) \\ &+ \Delta t S_i - \Delta t \mathcal{S}(\tilde{Q}_i), \end{aligned} \tag{6.24}$$

where $F^c_{i+1/2}$ denotes a numerical flux function and S_i a numerical source. For the numerical flux, we decide to use a simple and robust Rusanov-type flux of the form

$$\begin{aligned} F^c_{i+1/2} &= F^c(\bar{Q}^L_{i+1/2}, \bar{Q}^R_{i+1/2}) \\ &= \frac{1}{2}\left(\mathcal{F}^c(\bar{Q}^L_{i+1/2}) + \mathcal{F}^c(\bar{Q}^R_{i+1/2})\right) - \frac{1}{2}\lambda_{\max}\left(\bar{Q}^R_{i+1/2} - \bar{Q}^L_{i+1/2}\right). \end{aligned} \tag{6.25}$$

The numerical dissipation term herein is defined by

$$\lambda_{\max} = \max\left(|(\lambda^c)^L_{i+1/2}|, |(\lambda^c)^R_{i+1/2}|\right) \tag{6.26}$$

and thus only depends on the convective eigenvalues. This is a crucial difference to the definition in (3.31) and its application to the Gresho vortex in Ex. 3.10.1, where the maximum eigenvalue of the complete PDE system is used. The convective eigenvalues used in (6.26) are independent of the Mach number $\mathcal{M}$, which is why the flux does not introduce excessive dissipation for low Mach numbers. The numerical flux is evaluated in the states

$$\bar{Q}^{L,R}_{i+1/2} = \Delta Q^{L,R}_{i+1/2} + \tilde{Q}_{i+1/2}, \tag{6.27}$$

where the superscripts L, R denote the left and right extrapolated data at the interface. It is essential for the well-balancing that only the deviation ΔQ is reconstructed, while the equilibrium solution $\tilde{Q}$ is evaluated at the cell interface. The third term on the right-hand

side of (6.24) simply computes the physical fluxes based on the equilibrium solution at the cell interface that is $\mathcal{F}^c(\tilde{Q}_{i\pm1/2})$. For the source term, we substitute the volume-averaged quantity with its cell-centered value, which is accurate up to second order:

$$S_i = \mathcal{S}(\bar{Q}_i). \tag{6.28}$$

The implicit terms that appear in the pressure sub-system (6.8) are approximated by means of finite difference operators without any numerical dissipation [BP21]:

$$\left.\frac{\partial p}{\partial x}\right|_i^{n+1} = \frac{p_{i+1}^{n+1} - p_{i-1}^{n+1}}{2\,\Delta x} + \mathcal{O}(\Delta x^2), \tag{6.29a}$$

$$\left.\frac{\partial}{\partial x}\left(h\frac{\partial p}{\partial x}\right)\right|_i^{n,n+1} = \frac{1}{\Delta x^2}\begin{bmatrix} h_{i-1}^n & h_i^n & h_{i+1}^n \end{bmatrix}\begin{bmatrix} 3/4 & -1 & 1/4 \\ 0 & 0 & 0 \\ 1/4 & -1 & 3/4 \end{bmatrix}\begin{bmatrix} p_{i-1}^{n+1} \\ p_i^{n+1} \\ p_{i+1}^{n+1} \end{bmatrix} + \mathcal{O}(\Delta x^2). \tag{6.29b}$$

6.2.3 Second Order Extension

The method that has been described so far is only first order accurate in time and space. In order to increase the accuracy of the time-marching scheme, we apply the second order accurate IMEX-RK method LSDIRK2(2,2,2) from Ex. 3.9.5 [PR05]. In the terminology of IMEX-RK methods, the discretization of the convective sub-system constitutes the explicit operator Q_E, while the discretization of the pressure sub-system is represented by Q_I.

The spatial discretization within the Godunov-type method in the explicit update is extended to second order accuracy by a simple piecewise linear reconstruction of the conservative variables (see Sect. 3.8). The finite difference operators (6.29) in the implicit update are already second order accurate and can therefore be retained unchanged.

6.2.4 Multi-Dimensional Extension

In three spatial dimensions we discretize the domain $\mathcal{I} = [x_L, x_R] \times [y_L, y_R] \times [z_L, z_R]$ by a Cartesian grid with $N_x \times N_y \times N_z$ cells having the uniform cell size $\Delta x \times \Delta y \times \Delta z$. The discretization of the explicit part is extended to three spatial dimensions by using an unsplit finite volume method according to [Tor09] (see also Sect. 3.7) given by

$$\begin{aligned}
\Delta Q_{i,j,k}^* = \Delta Q_{i,j,k}^n
&- \frac{\Delta t}{\Delta x}\left(F_{1,i+1/2,j,k}^c - F_{1,i-1/2,j,k}^c\right) + \frac{\Delta t}{\Delta x}\left(\mathcal{F}_1^c(\tilde{Q}_{i+1/2,j,k}) - \mathcal{F}_1^c(\tilde{Q}_{i-1/2,j,k})\right) \\
&- \frac{\Delta t}{\Delta y}\left(F_{2,i,j+1/2,k}^c - F_{2,i,j-1/2,k}^c\right) + \frac{\Delta t}{\Delta y}\left(\mathcal{F}_2^c(\tilde{Q}_{i,j+1/2,k}) - \mathcal{F}_2^c(\tilde{Q}_{i,j-1/2,k})\right) \\
&- \frac{\Delta t}{\Delta z}\left(F_{3,i,j,k+1/2}^c - F_{3,i,j,k-1/2}^c\right) + \frac{\Delta t}{\Delta z}\left(\mathcal{F}_3^c(\tilde{Q}_{i,j,k+1/2}) - \mathcal{F}_3^c(\tilde{Q}_{i,j,k-1/2})\right) \\
&+ \Delta t S_{i,j,k} - \Delta t \mathcal{S}(\tilde{Q}_{i,j,k}).
\end{aligned} \tag{6.30}$$

The numerical fluxes F_1^c, F_2^c and F_3^c have the form of the Rusanov flux and are constructed as in (6.25), while the source term S uses the cell-centered value as in (6.28). The updated density $\Delta\rho_{i,j,k}^{n+1}$ and magnetic field $\Delta\mathbf{B}_{i,j,k}^{n+1}$ are equal to their explicit update since in the splitting the complete flux of these components is explicit. The update of the momentum

components, on the other hand, contains implicit parts and reads in the fully discrete and three-dimensional form as

$$(\Delta\rho u)^{n+1}_{i,j,k} = (\Delta\rho u)^{*}_{i,j,k} - \frac{\Delta t}{2\Delta x}\left(p^{n+1}_{i+1,j,k} - \tilde{p}_{i+1,j,k} - p^{n+1}_{i-1,j,k} + \tilde{p}_{i-1,j,k}\right), \tag{6.31a}$$

$$(\Delta\rho v)^{n+1}_{i,j,k} = (\Delta\rho v)^{*}_{i,j,k} - \frac{\Delta t}{2\Delta y}\left(p^{n+1}_{i,j+1,k} - \tilde{p}_{i,j+1,k} - p^{n+1}_{i,j-1,k} + \tilde{p}_{i,j-1,k}\right), \tag{6.31b}$$

$$(\Delta\rho w)^{n+1}_{i,j,k} = (\Delta\rho w)^{*}_{i,j,k} - \frac{\Delta t}{2\Delta z}\left(p^{n+1}_{i,j,k+1} - \tilde{p}_{i,j,k+1} - p^{n+1}_{i,j,k-1} + \tilde{p}_{i,j,k-1}\right). \tag{6.31c}$$

Implicit terms also appear in the update of the total energy:

$$\begin{aligned}(\Delta E)^{n+1}_{i,j,k} = (\Delta E)^{*}_{i,j,k} &- \frac{\Delta t}{2\Delta x}\left(h^{n}_{i+1,j,k}(\rho u)^{n+1}_{i+1,j,k} - h^{n}_{i-1,j,k}(\rho u)^{n+1}_{i-1,j,k}\right)\\ &- \frac{\Delta t}{2\Delta y}\left(h^{n}_{i,j+1,k}(\rho v)^{n+1}_{i,j+1,k} - h^{n}_{i,j-1,k}(\rho v)^{n+1}_{i,j-1,k}\right)\\ &- \frac{\Delta t}{2\Delta z}\left(h^{n}_{i,j,k+1}(\rho w)^{n+1}_{i,j,k+1} - h^{n}_{i,j,k-1}(\rho w)^{n+1}_{i,j,k-1}\right).\end{aligned} \tag{6.32}$$

The pressure p^{n+1}, which is needed for the updates (6.31) and (6.32), can be determined by solving the following elliptic equation:

$$\begin{aligned}&\frac{p^{n+1}_{i,j,k}}{\gamma-1}\\ &- \frac{\Delta t}{2\Delta x}\frac{(\rho u)^{n}_{i,j,k}}{2\rho^{n+1}_{i,j,k}}\left(p^{n+1}_{i+1,j,k} - \tilde{p}_{i+1,j,k} - p^{n+1}_{i-1,j,k} + \tilde{p}_{i-1,j,k}\right)\\ &- \frac{\Delta t}{2\Delta y}\frac{(\rho v)^{n}_{i,j,k}}{2\rho^{n+1}_{i,j,k}}\left(p^{n+1}_{i,j+1,k} - \tilde{p}_{i,j+1,k} - p^{n+1}_{i,j-1,k} + \tilde{p}_{i,j-1,k}\right)\\ &- \frac{\Delta t}{2\Delta z}\frac{(\rho w)^{n}_{i,j,k}}{2\rho^{n+1}_{i,j,k}}\left(p^{n+1}_{i,j,k+1} - \tilde{p}_{i,j,k+1} - p^{n+1}_{i,j,k-1} + \tilde{p}_{i,j,k-1}\right)\\ &- \frac{\Delta t^2}{\Delta x^2}\Bigg[\left(\frac{3}{4}h^{n}_{i-1,j,k} + \frac{1}{4}h^{n}_{i+1,j,k}\right)\left(p^{n+1}_{i-1,j,k} - \tilde{p}_{i-1,j,k}\right) - \left(h^{n}_{i-1,j,k} + h^{n}_{i+1,j,k}\right)\left(p^{n+1}_{i,j,k} - \tilde{p}_{i,j,k}\right)\\ &\qquad + \left(\frac{1}{4}h^{n}_{i-1,j,k} + \frac{3}{4}h^{n}_{i+1,j,k}\right)\left(p^{n+1}_{i+1,j,k} - \tilde{p}_{i+1,j,k}\right)\Bigg]\\ &- \frac{\Delta t^2}{\Delta y^2}\Bigg[\left(\frac{3}{4}h^{n}_{i,j-1,k} + \frac{1}{4}h^{n}_{i,j+1,k}\right)\left(p^{n+1}_{i,j-1,k} - \tilde{p}_{i,j-1,k}\right) - \left(h^{n}_{i,j-1,k} + h^{n}_{i,j+1,k}\right)\left(p^{n+1}_{i,j,k} - \tilde{p}_{i,j,k}\right)\\ &\qquad + \left(\frac{1}{4}h^{n}_{i,j-1,k} + \frac{3}{4}h^{n}_{i,j+1,k}\right)\left(p^{n+1}_{i,j+1,k} - \tilde{p}_{i,j+1,k}\right)\Bigg]\\ &- \frac{\Delta t^2}{\Delta z^2}\Bigg[\left(\frac{3}{4}h^{n}_{i,j,k-1} + \frac{1}{4}h^{n}_{i,j,k+1}\right)\left(p^{n+1}_{i,j,k-1} - \tilde{p}_{i,j,k-1}\right) - \left(h^{n}_{i,j,k-1} + h^{n}_{i,j,k+1}\right)\left(p^{n+1}_{i,j,k} - \tilde{p}_{i,j,k}\right)\\ &\qquad + \left(\frac{1}{4}h^{n}_{i,j,k-1} + \frac{3}{4}h^{n}_{i,j,k+1}\right)\left(p^{n+1}_{i,j,k+1} - \tilde{p}_{i,j,k+1}\right)\Bigg]\\ &= b^{n}_{i,j,k}\end{aligned} \tag{6.33}$$

with the right-hand side

$$\begin{aligned} b^n_{i,j,k} = &\frac{\tilde{p}_{i,j,k}}{\gamma-1} + (\Delta E)^*_{i,j,k} - \Delta m^{n+1}_{i,j,k} \\ &- \frac{(\rho u)^n_{i,j,k}}{2\rho^{n+1}_{i,j,k}}(\Delta\rho u)^*_{i,j,k} - \frac{\Delta t}{2\Delta x}\left(h^n_{i+1,j,k}(\Delta\rho u)^*_{i+1,j,k} - h^n_{i-1,j,k}(\Delta\rho u)^*_{i-1,j,k}\right) \\ &- \frac{(\rho v)^n_{i,j,k}}{2\rho^{n+1}_{i,j,k}}(\Delta\rho v)^*_{i,j,k} - \frac{\Delta t}{2\Delta y}\left(h^n_{i,j+1,k}(\Delta\rho v)^*_{i,j+1,k} - h^n_{i,j-1,k}(\Delta\rho v)^*_{i,j-1,k}\right) \\ &- \frac{(\rho w)^n_{i,j,k}}{2\rho^{n+1}_{i,j,k}}(\Delta\rho w)^*_{i,j,k} - \frac{\Delta t}{2\Delta z}\left(h^n_{i,j,k+1}(\Delta\rho w)^*_{i,j,k+1} - h^n_{i,j,k-1}(\Delta\rho w)^*_{i,j,k-1}\right). \end{aligned} \tag{6.34}$$

To set up this equation for the pressure, we have made use of the finite difference operators (6.29).

6.2.5 Constrained Transport Method

The numerical scheme described up to this point does not obey the solenoidal constraint (2.69). As a result, the divergence of the magnetic field can increase significantly during the simulation leading to a reduced stability and unphyiscal solutions. Therefore, an additional correction of the magnetic field is required after each time step. In order to restore a solenoidal magnetic field at the discrete level, we rely on the second order accurate *Contact CT method* [GS05] that is described in Sect. 5.2.3. The scheme thus operates on a staggered grid and stores the magnetic field additionally at the cell faces. As the induction equation is handled completely in the explicit part, the update of the magnetic field in x-direction at the faces given by

$$\begin{aligned} \frac{\partial}{\partial t}\Delta B_{x,i+1/2,j,k} = &\frac{1}{\Delta y}\left(\Delta\mathcal{E}_{z,i+1/2,j+1/2,k} - \Delta\mathcal{E}_{z,i+1/2,j-1/2,k}\right) \\ &-\frac{1}{\Delta z}\left(\Delta\mathcal{E}_{y,i+1/2,j,k+1/2} - \Delta\mathcal{E}_{y,i+1/2,j,k-1/2}\right) \end{aligned} \tag{6.35}$$

is done directly after the explicit step. The formulas to evolve the face-centered values of B_y and B_z are analogous. The new cell-centered magnetic field is then calculated by determining the arithmetic mean of the respective face-entered values as done in (5.35). The definition of the elliptic equation for the pressure in (6.33) then relies on the corrected and therefore divergence-free magnetic field $\mathbf{B}^{n+1}$.

6.2.6 Well-Balanced Property

The following theorem states that the presented fully discrete scheme preserves MHSE of the form (2.79) up to rounding errors.

Theorem 6.2.1. *Let us assume that the numerical solution Q^n is equal to the discrete magnetohydrostatic equilibrium solution $\tilde{Q}$, i.e.*

$$Q^n_{i,j,k} = \tilde{Q}_{i,j,k} \qquad \forall (i,j,k) \in \{1,...,N_x\} \times \{1,...,N_y\} \times \{1,...,N_z\}. \tag{6.36}$$

Then the numerical method described in Sect. 6.2.1-6.2.5 preserves the solution up to machine precision.

Proof. From the assumption (6.36) one can conclude that $\Delta Q^n_{i,j,k} = 0$ at every grid point. Thus, the input (6.27) of the Rusanov fluxes in the explicit sub-system is reduced to the equilibrium solution at the interface, i.e.

$$\bar{Q}^L_{i+1/2,j,k} = \bar{Q}^R_{i+1/2,j,k} = \tilde{Q}_{i+1/2,j,k}, \tag{6.37a}$$
$$\bar{Q}^L_{i,j+1/2,k} = \bar{Q}^R_{i,j+1/2,k} = \tilde{Q}_{i,j+1/2,k}, \tag{6.37b}$$
$$\bar{Q}^L_{i,j,k+1/2} = \bar{Q}^R_{i,j,k+1/2} = \tilde{Q}_{i,j,k+1/2}. \tag{6.37c}$$

The Rusanov flux is a consistent flux in the sense of Def. 3.1.1. For that reason, the numerical fluxes are equal to the physical fluxes so that all flux terms in (6.30) cancel each other out. Likewise, the source terms are truncated away. Therefore, the contributions from the explicit part (6.30) are zero:

$$\begin{aligned} \Delta\rho^*_{i,j,k} = (\Delta E)^*_{i,j,k} = 0 \quad \text{and} \quad (\Delta\rho\mathbf{v})^*_{i,j,k} = \Delta\mathbf{B}^*_{i,j,k} = \mathbf{0} \\ \forall (i,j,k) \in \{1,...,N_x\} \times \{1,...,N_y\} \times \{1,...,N_z\}. \end{aligned} \tag{6.38}$$

Consequently, also the deviation in the magnetic energy $\Delta m^{n+1}_{i,j,k}$ is zero. The constrained transport step that is performed after the explicit update keeps the deviations in the magnetic field at zero, because the electric field at the corners in the update (6.35) is equal to zero due to the zero velocity in the equilibrium. Under these conditions, the elliptic equation for the pressure (6.33) simplifies to

$$\begin{aligned} &\frac{p^{n+1}_{i,j,k}}{\gamma - 1} \\ &-\frac{\Delta t^2}{\Delta x^2}\left[\left(\frac{3}{4}h^n_{i-1,j,k} + \frac{1}{4}h^n_{i+1,j,k}\right)p^{n+1}_{i-1,j,k} - \left(h^n_{i-1,j,k} + h^n_{i+1,j,k}\right)p^{n+1}_{i,j,k}\right. \\ &\qquad \left. + \left(\frac{1}{4}h^n_{i-1,j,k} + \frac{3}{4}h^n_{i+1,j,k}\right)p^{n+1}_{i+1,j,k}\right] \\ &-\frac{\Delta t^2}{\Delta y^2}\left[\left(\frac{3}{4}h^n_{i,j-1,k} + \frac{1}{4}h^n_{i,j+1,k}\right)p^{n+1}_{i,j-1,k} - \left(h^n_{i,j-1,k} + h^n_{i,j+1,k}\right)p^{n+1}_{i,j,k}\right. \\ &\qquad \left. + \left(\frac{1}{4}h^n_{i,j-1,k} + \frac{3}{4}h^n_{i,j+1,k}\right)p^{n+1}_{i,j+1,k}\right] \\ &-\frac{\Delta t^2}{\Delta z^2}\left[\left(\frac{3}{4}h^n_{i,j,k-1} + \frac{1}{4}h^n_{i,j,k+1}\right)p^{n+1}_{i,j,k-1} - \left(h^n_{i,j,k-1} + h^n_{i,j,k+1}\right)p^{n+1}_{i,j,k}\right. \\ &\qquad \left. + \left(\frac{1}{4}h^n_{i,j,k-1} + \frac{3}{4}h^n_{i,j,k+1}\right)p^{n+1}_{i,j,k+1}\right] \end{aligned} \tag{6.39}$$

$$
\begin{aligned}
&= \frac{\tilde{p}_{i,j,k}}{\gamma - 1} \\
&\quad - \frac{\Delta t^2}{\Delta x^2} \Bigg[\left(\frac{3}{4} h^n_{i-1,j,k} + \frac{1}{4} h^n_{i+1,j,k} \right) \tilde{p}_{i-1,j,k} - \left(h^n_{i-1,j,k} + h^n_{i+1,j,k} \right) \tilde{p}_{i,j,k} \\
&\qquad\qquad + \left(\frac{1}{4} h^n_{i-1,j,k} + \frac{3}{4} h^n_{i+1,j,k} \right) \tilde{p}_{i+1,j,k} \Bigg] \\
&\quad - \frac{\Delta t^2}{\Delta y^2} \Bigg[\left(\frac{3}{4} h^n_{i,j-1,k} + \frac{1}{4} h^n_{i,j+1,k} \right) \tilde{p}_{i,j-1,k} - \left(h^n_{i,j-1,k} + h^n_{i,j+1,k} \right) \tilde{p}_{i,j,k} \\
&\qquad\qquad + \left(\frac{1}{4} h^n_{i,j-1,k} + \frac{3}{4} h^n_{i,j+1,k} \right) \tilde{p}_{i,j+1,k} \Bigg] \\
&\quad - \frac{\Delta t^2}{\Delta z^2} \Bigg[\left(\frac{3}{4} h^n_{i,j,k-1} + \frac{1}{4} h^n_{i,j,k+1} \right) \tilde{p}_{i,j,k-1} - \left(h^n_{i,j,k-1} + h^n_{i,j,k+1} \right) \tilde{p}_{i,j,k} \\
&\qquad\qquad + \left(\frac{1}{4} h^n_{i,j,k-1} + \frac{3}{4} h^n_{i,j,k+1} \right) \tilde{p}_{i,j,k+1} \Bigg],
\end{aligned}
$$

which admits the unique solution

$$
p^{n+1}_{i,j,k} = \tilde{p}_{i,j,k}. \tag{6.40}
$$

This means that the right-hand side of the momentum update (6.31) becomes zero and consequently the right-hand side of the energy update (6.32) also becomes zero. Thus the updated deviations for all components of the state vector remain zero, which means that the numerical solution does not change during one time step, i.e.

$$
Q^{n+1}_{i,j,k} = Q^n_{i,j,k} = \tilde{Q}_{i,j,k} \qquad \forall (i,j,k) \in \{1,...,N_x\} \times \{1,...,N_y\} \times \{1,...,N_z\}. \tag{6.41}
$$

□

6.2.7 Summary of the Scheme

Since various methods are combined in the *semi-implicit well-balanced finite volume* (SI-WB-FV) scheme presented, we provide an overview of the individual steps of the scheme at this point. Let us assume that we know the time-independent equilibrium solution $\tilde{Q}$ at every point in the computational domain $\mathcal{I}$. Then a single time step in the method from time $t = t^n$ to time $t = t^{n+1}$ consists of the following sub-steps:

1. Start with the values at the current time level: $Q^n_{i,j,k}$, $B^n_{x,i+1/2,j,k}$, $B^n_{y,i,j+1/2,k}$, $B^n_{z,i,j,k+1/2}$.
2. Compute the deviation: $\Delta Q^n_{i,j,k} = Q^n_{i,j,k} - \tilde{Q}_{i,j,k}$.
3. Reconstruct the values at the cell interfaces for the deviation: $\Delta Q^L_{i+1/2,j,k}$, $\Delta Q^R_{i+1/2,j,k}$, $\Delta Q^L_{i,j+1/2,k}$, $\Delta Q^R_{i,j+1/2,k}$, $\Delta Q^L_{i,j,k+1/2}$, $\Delta Q^R_{i,j,k+1/2}$.
4. Perform the explicit update (6.24): $\Delta Q^*_{i,j,k}$.
5. Start the CT routine, update the staggered magnetic field at the face centers and compute the updated cell-centered magnetic field: $B^{n+1}_{x,i+1/2,j,k}$, $B^{n+1}_{y,i,j+1/2,k}$, $B^{n+1}_{z,i,j,k+1/2}$, $\mathbf{B}^{n+1}_{i,j,k}$.

6. Use the computed quantities from the explicit part to solve the elliptic equation for the pressure (6.33) via GMRES: $p_{i,j,k}^{n+1}$.

7. Compute the updated deviation in the momentum: $(\Delta\rho u)_{i,j,k}^{n+1}$.

8. Compute the updated deviation in the total energy: $(\Delta E)_{i,j,k}^{n+1}$.

9. Recompute the actual solution at the new time level: $Q_{i,j,k}^{n+1} = \Delta Q_{i,j,k}^{n+1} + \tilde{Q}_{i,j,k}$.

In the case that we do not want to balance a MHSE, the equilibrium solution $\tilde{Q}$ can be set to zero.

6.2.8 Modified Density Update

Recently, it has been noted that semi-implicit methods of the form presented lead to relatively high magnitudes of density fluctuations in the incompressible limit [ALB]. Numerical results for the Balsara vortex in Sect. 6.3.3 support this finding. This phenomenon is rooted in two factors: firstly, it is proven for the homogeneous case that the divergence of the velocity field within this type of semi-implicit method has the order of the time step in the incompressible limit [BP21], i.e.

$$\nabla \cdot \mathbf{v}^{n+1} = \mathcal{O}(\Delta t). \tag{6.42}$$

As the time step is chosen independently of the Mach number, the divergence does not disappear in the incompressible limit so that an initially constant density is evolved by a nonzero flux that produces new fluctuations of Mach number independent magnitude. Secondly, the continuity equation is discretized by a Godunov-type method with an upwinding technique. Therefore, the density is subject to compression and decompression [Bar], which intensifies the density fluctuations. In order to avoid the upwind discretization and thereby reduce the magnitude of the fluctuations, we propose to evolve the density in the implicit part [ALB]. After computing the updated deviation in the momentum in step seven, the density deviation is updated by central finite differences of the form

$$\begin{aligned}\Delta\rho_{i,j,k}^{n+1} = \Delta\rho_{i,j,k}^{n} &+ \frac{\Delta t}{2\Delta x}\left(\Delta\rho u_{i+1,j,k}^{n+1} - \Delta\rho u_{i-1,j,k}^{n+1}\right) \\ &+ \frac{\Delta t}{2\Delta y}\left(\Delta\rho u_{i,j+1,k}^{n+1} - \Delta\rho u_{i,j-1,k}^{n+1}\right) \\ &+ \frac{\Delta t}{2\Delta z}\left(\Delta\rho u_{i,j,k+1}^{n+1} - \Delta\rho u_{i,j,k-1}^{n+1}\right).\end{aligned} \tag{6.43}$$

In the following sections, we call the method with this new density update *modified semi-implicit well-balanced finite volume* (SI$^+$-WB-FV) method.

6.3 Numerical Results

In this section, we apply the second order SI-WB-FV method to various numerical experiments for the compressible Euler and ideal MHD equations. First, we assess the method's capability to handle shocks through the examination of two standard tests: a 1D shock tube under gravity (i) and the 2D Orszag-Tang vortex (ii). Subsequently, we solve the Balsara vortex (iii) and a magnetized Kelvin-Helmholtz instability (iv) to investigate the

method's performance in low Mach number environments. Our investigation includes the analysis of the order of convergence and of the influence of Mach numbers on the scheme's dissipation.
Moving forward, we check the well-balancing property of Sect. 6.2.6. As a start, we ascertain whether the method exactly preserves hydrostatic equilibria for the Euler equations (v). Then, we introduce a small pressure perturbation to find out whether the method can resolve it on a coarse grid (vi). Following this principle, we examine the well-balanced property for magnetohydrostatic solutions in the context of the MHD equations (vii, viii). Finally, we apply the method to a stationary vortex for the Euler equations, evaluating its effectiveness in resolving low Mach flows under the influence of gravity (ix).
In all numerical experiments, the method's time step is computed using the CFL condition (6.9) with a prescribed CFL value of 0.9. In the tests studying the impact of the well-balancing method on resolving small perturbations (vi, viii), the SI-WB-FV results are juxtaposed with those of a non-well-balanced counterpart called SI-FV. While sharing the same semi-implicit scheme, the SI-FV method sets the equilibrium solution $\tilde{\mathcal{Q}}$ to zero. All tests are performed on a uniform Cartesian grid.

6.3.1 Shock Tube under Gravitational Field

The first test case is the Sod shock tube for the one-dimensional Euler equations with an added gravitational source. The test is designed as described in Sect. 4.6.2. Since we define the source term directly on the basis of the gravitational acceleration $\mathbf{g}$ and not on the potential Φ, we use $g_x = -1$. The magnetic field $\mathbf{B}$ is set to zero. The solution at final time $t_f = 0.2$ is computed by the second order SI-WB-FV scheme on 100 cells. Since we have no flow around an equilibrium in this test, we set the equilibrium solution $\tilde{\mathcal{Q}}$ to zero. The numerical solution is compared to a reference solution, which is computed by a fully explicit second order FV method on 20 000 cells. The results in Fig. 6.1 show good agreement with the reference solution and are also consistent with the solution in Sect. 4.6.2 and solutions in the literature [CK15]. The test demonstrates the capability of the semi-implicit scheme to deal with shocks, meaning flows which are not in the low Mach number regime.

6.3.2 Orszag-Tang Vortex

A well-known test for the two-dimensional homogeneous MHD equations is the Orszag-Tang vortex [CRT14, DBTF19, OT79]. Starting with smooth initial data, over time shocks develop along the diagonal direction in combination with a vortex located at the center of the computational domain. The initial conditions for the primitive variables on the spatial domain $\mathcal{I} = [0,1]^2$ are given by

$$\mathcal{Q}_0^P = \left(\frac{25}{36\pi},\ -\sin(2\pi y),\ \sin(2\pi x),\ 0,\ \frac{5}{12\pi},\ -\frac{1}{\sqrt{4\pi}}\sin(2\pi y),\ \frac{1}{\sqrt{4\pi}}\sin(4\pi x), 0 \right). \tag{6.44}$$

The gravitational acceleration $\mathbf{g}$ and the equilibrium solution $\tilde{\mathcal{Q}}$ are set to zero. Periodic boundary conditions are imposed on all sides. We discretize the computational domain by 128×128 control volumes. In Fig. 6.2 results for the pressure p at different times computed by the SI-WB-FV method are presented. The numerical method successfully captures the shocks that occur as time evolves and the results align qualitatively with those reported in the literature. [CRT14, DBTF19].

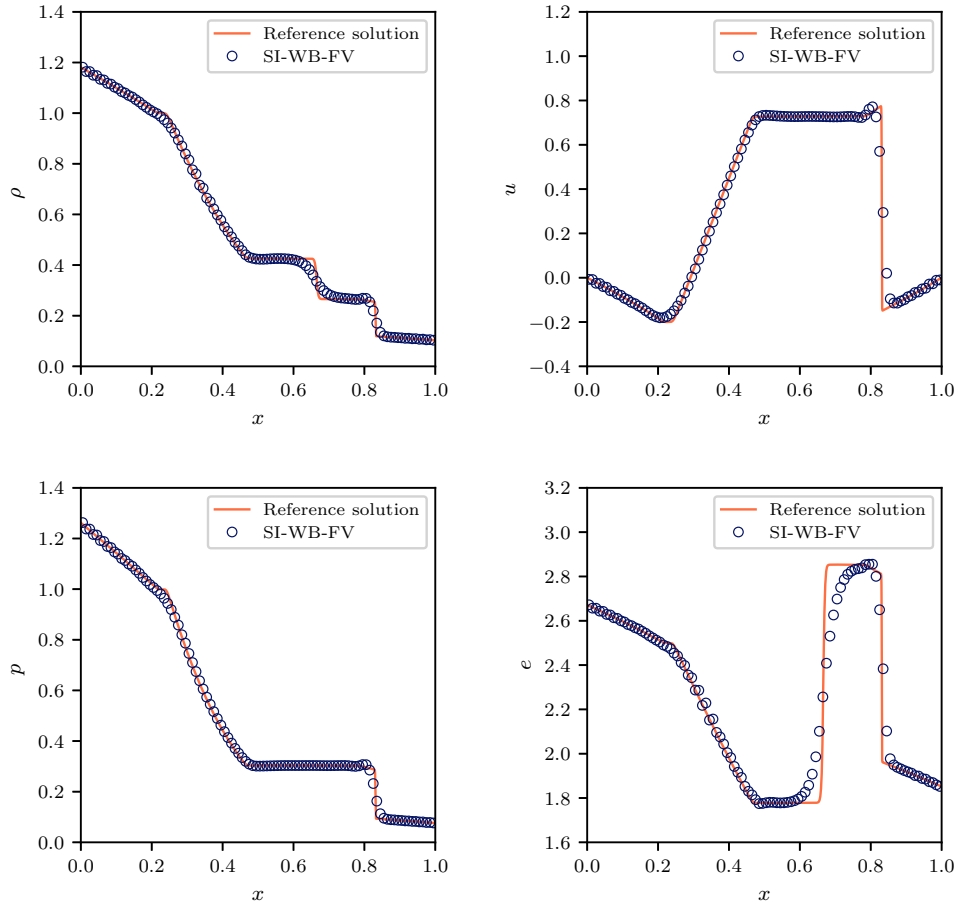

Figure 6.1: Shock tube under gravitational field at time $t_f = 0.2$. Comparison with the reference solution for density (top left), horizontal velocity (top right), pressure (bottom left) and internal energy (bottom right).

As the Orszag-Tang vortex is a two-dimensional problem, the solenoidal constraint must be taken into account. The time evolution of the discrete divergence (5.38) in Fig. 6.3 shows that the Contact CT method ensures that the divergence remains very small during the entire simulation and has no significant influence on the results.

6.3.3 Balsara Vortex

In order to investigate the scaling and the low Mach capabilities of the SI-WB-FV scheme, we consider the Balsara vortex described in Sect. 5.4.1. Since this setup is purely subsonic, we do not use a limiter in the linear reconstruction of the method.
As done before, the knowledge of the exact solution is used for a convergence study. In the initial data of the vortex, we set $\tilde{V} = 10^{-3}$ and $\beta_K = 1$, which corresponds to a maximum initial Mach number of $\mathcal{M}_{\max} = \max(\mathcal{M}_{\text{loc}})_{t=0} = 1.55 \times 10^{-3}$. This test setup is solved with different grid resolutions: $N = 32, 64, 128, 256$. The resulting L^1-errors for the primitive variables are presented in Fig. 6.4 and confirm the expected second order

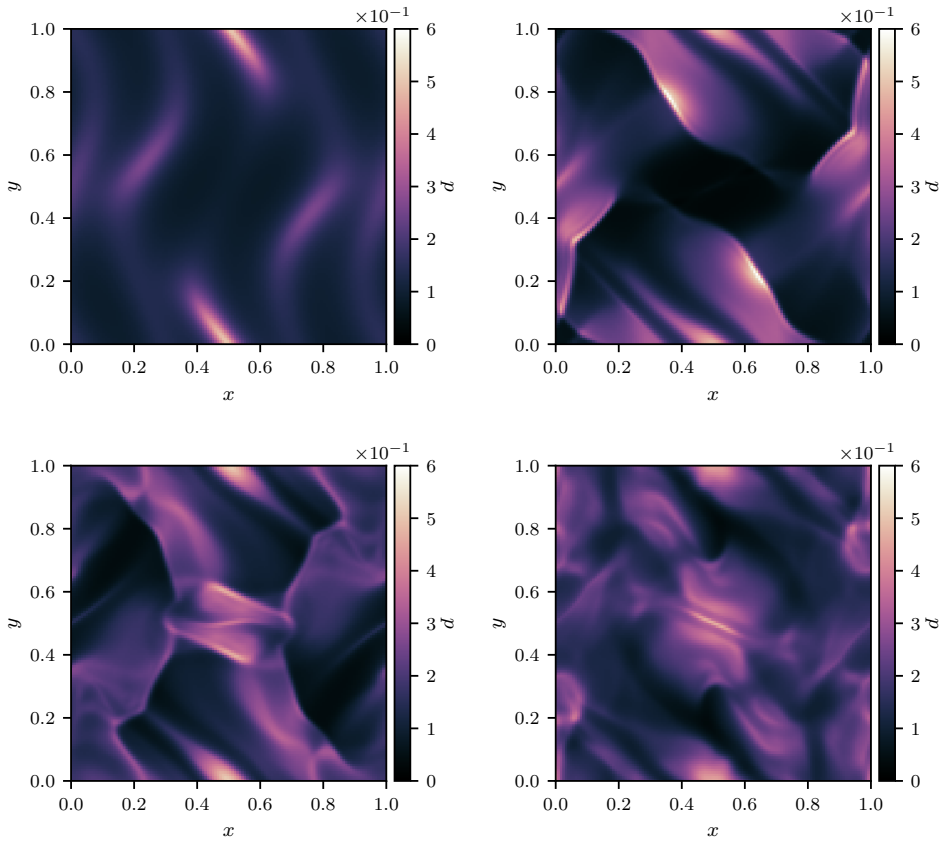

Figure 6.2: Orszag-Tang vortex. Numerical solution of the pressure at output time $t_f = 1/12$ (top left), $t_f = 1/3$ (top right), $t_f = 1/2$ (bottom left) and $t_f = 5/6$ (bottom right).

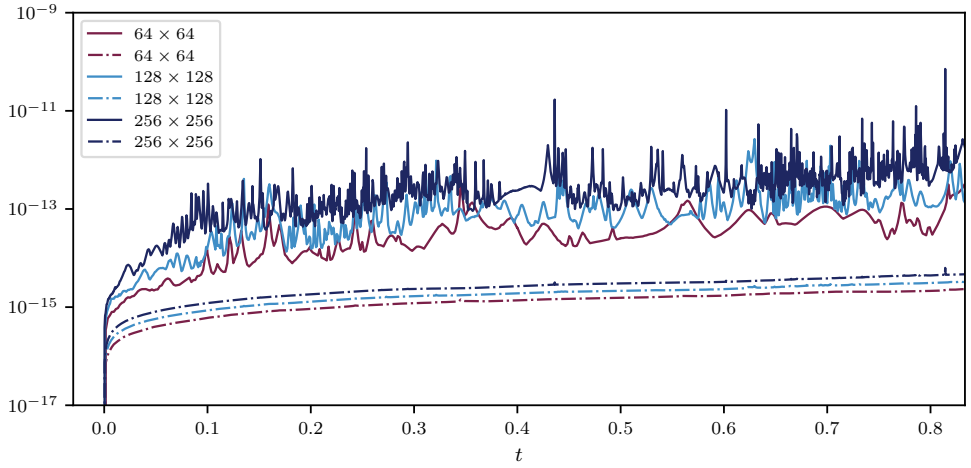

Figure 6.3: Time evolution of the maximum relative divergence $\max[(\nabla \cdot \mathbf{B}\Delta x)/|\mathbf{B}|]$ (solid line) and mean relative divergence $\langle(\nabla \cdot \mathbf{B}\Delta x)/|\mathbf{B}|\rangle$ (dot-dashed) in the simulations of the Orszag-Tang vortex.

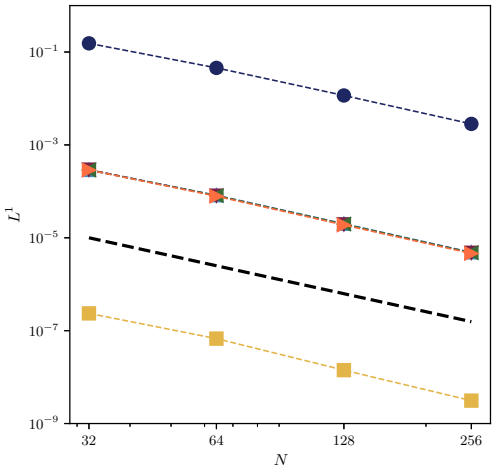

Figure 6.4: Convergence of the L^1-error for the SI-WB-FV method in the Balsara vortex with $\tilde{V} = 10^{-3}$ and $\beta_{\mathrm{K}} = 1$ for each primitive variable as a function of resolution. The dashed black line is the second order scaling.

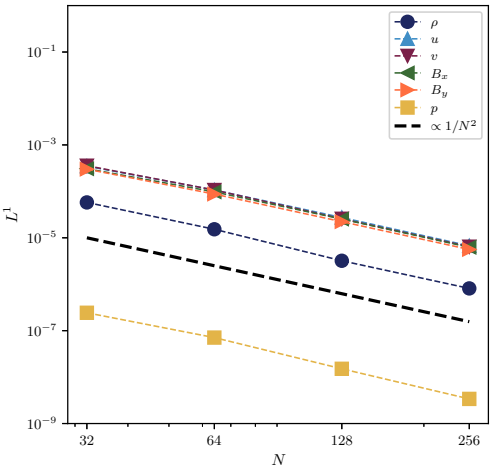

Figure 6.5: Convergence of the L^1-error for the SI$^+$-WB-FV method in the Balsara vortex with $\tilde{V} = 10^{-3}$ and $\beta_{\mathrm{K}} = 1$ for each primitive variable as a function of resolution. The dashed black line is the second order scaling.

accuracy of the scheme. The errors for the density though have a significantly greater magnitude compared to the other quantities and also compared to the results of the IESS scheme in Chapter 5. The reason lies in the increased density fluctuations produced by the semi-implicit IMEX method in the low Mach number regime (see Sect. 6.2.8). We therefore repeat the convergence study with the modified density update (6.43). The results of the SI$^+$-WB-FV method in Fig. 6.5 show that this modification decreases the error in the density variable by several orders of magnitude. This occurs without a significant effect on the error of the other primitive variables. The smaller error is due to lower fluctuations in density, as an evaluation in Appendix B.1 shows. It therefore makes sense to use the modified version of the method in purely subsonic regimes.
In a second step, we investigate the effect of the IMEX approach on the numerical diffusion in the low Mach number regime. In Fig. 6.6, we present the distributions of the rotational kinetic energy

$$E_{\mathrm{R}} = \frac{1}{2}\rho\left[\left(u - \tilde{V}/\sqrt{2}\right)^2 + \left(v - \tilde{V}/\sqrt{2}\right)^2\right] \tag{6.45}$$

after one advective time t_f that result from the runs in the convergence study. The SI$^+$-WB-FV method is able to resolve the vortex already on rather coarse grids, which can be attributed to the low Mach compliant dissipation terms of the spatial operators in Sect. 6.2.2. The loss of rotational kinetic energy is kept small and has a comparable magnitude with that in the IESS method.

To analyze the amount of numerical dissipation for different Mach number regimes, we run a set of simulations

$$(\tilde{V}) \times (N) = (10^{-4}, 10^{-3}) \times (32, 64, 128) \tag{6.46}$$

with $\beta_K = 1$. The corresponding maximum initial Mach number in the setup is either

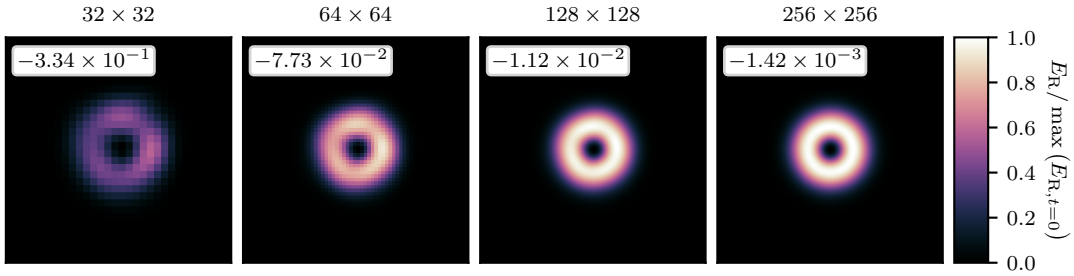

Figure 6.6: Distribution of the rotational kinetic energy (normalized by the maximum initial value) of the Balsara vortex with $\tilde{V} = 10^{-3}$ and $\beta_{\mathrm{K}} = 1$ after one advective time t_f obtained by the SI$^+$-WB-FV scheme. The insets show the fraction of rotational kinetic energy that has been dissipated by the end of the simulation: $(E_{\mathrm{R},t=t_f})_{\mathrm{tot}}/(E_{\mathrm{R},t=0})_{\mathrm{tot}} - 1$.

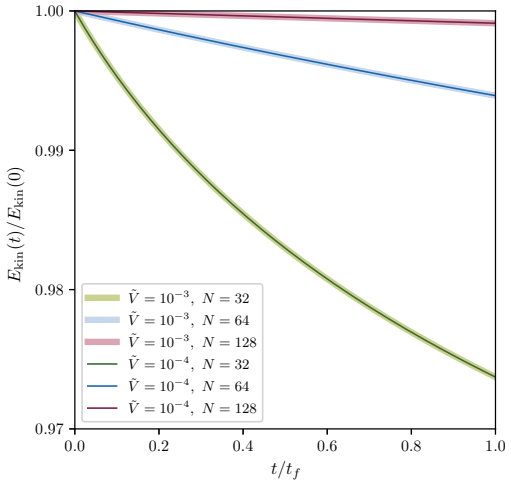

Figure 6.7: Time evolution of the total kinetic energy E_{kin} for different maximum rotational velocities $\tilde{V}$ (and therefore different $\mathcal{M}_{\mathrm{max}}$) and grid resolutions $N = N_x = N_y$ for the SI$^+$-WB-FV scheme.

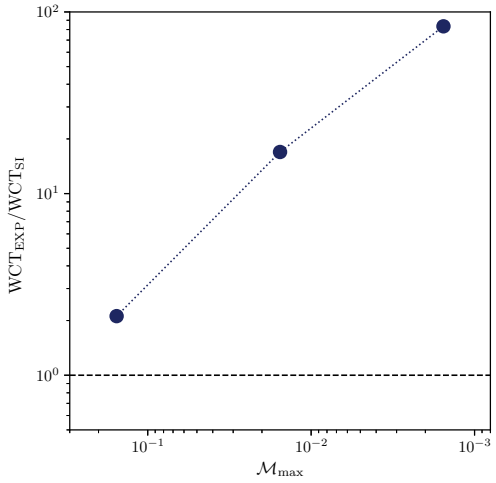

Figure 6.8: Ratio of the wall-clock times obtained with a MUSCL-Hancock scheme using the LHLLD solver and the SI$^+$-WB-FV scheme as a function of initial maximum sonic Mach number. The dashed black line is drawn to represent same relative efficiency.

$\mathcal{M}_{\mathrm{max}} = 1.55 \times 10^{-3}$ for $\tilde{V} = 10^{-3}$ or $\mathcal{M}_{\mathrm{max}} = 1.55 \times 10^{-4}$ for $\tilde{V} = 10^{-4}$. The loss of total kinetic energy during the simulations serves as measure for the dissipation. The results in Fig. 6.7 show that the curves only differ for different grid resolutions $N \times N$, but not for different Mach numbers $\mathcal{M}_{\mathrm{max}}$. This finding underlines that the dissipation in the SI$^+$-WB-FV scheme does not increase in the low Mach number regime.

In order to measure the efficiency of the semi-implicit approach in subsonic regimes, we compare its wall-clock time for different $\mathcal{M}_{\mathrm{max}}$ with those of a fully time-explicit alternative. In order to reach second order in time, the explicit scheme relies on a MUSCL-Hancock approach [vL84]. The LHLLD solver described in Sect. 5.2.1 is used in its

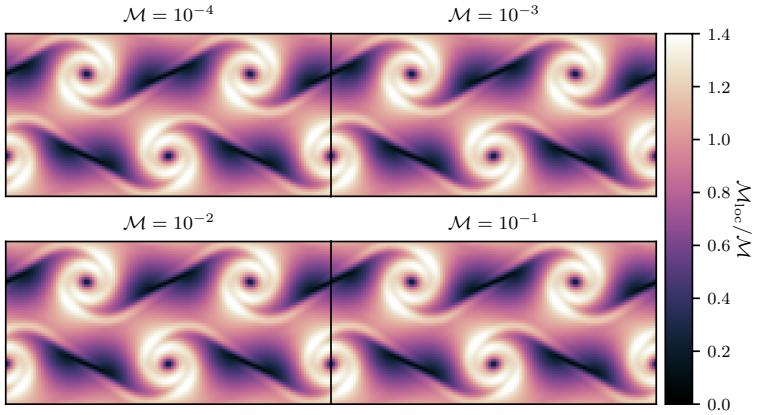

Figure 6.9: Distribution of the local sonic Mach number in the Kelvin-Helmholtz instability test at $t/t_{\max} = 1/6$ obtained with the SI$^+$-WB-FV scheme on a 128×64 grid for different values of $\mathcal{M}$. All panels are rescaled by the corresponding value of $\mathcal{M}$.

Godunov-type method to obtain accurate results in subsonic regimes. The linear reconstruction and the constrained transport method remain unaltered. We run the following set of simulations

$$(\tilde{V}) \times (\beta_{\mathrm{K}}) = (10^{-3}, 10^{-2}, 10^{-1}) \times (1), \tag{6.47}$$

on a grid with 40×40 cells. The results in Fig. 6.8 show that the semi-implicit scheme is faster than the fully explicit method in all considered setups. It is already more than twice as fast for a moderate Mach number of 10^{-1}, and the efficiency gap is widening further for smaller Mach numbers. The threshold above which the semi-implicit approach is worthwhile is therefore at even higher Mach numbers. Thus, the semi-implicit approach leads to efficiency advantages very quickly due to its small implicit part. This property makes it particularly suitable for applications that include both regimes with very low and regimes with moderate Mach numbers.

6.3.4 Magnetized Kelvin-Helmholtz Instability

In this section we consider a magnetized Kelvin-Helmholtz instability. The test setup is exactly defined as in Sect. 5.4.2. Since it is purely subsonic, we do not apply limiters in the reconstruction method and use the modified density update (6.43). We perform simulations for the following range of Mach numbers and grid resolutions:

$$(\mathcal{M}) \times (N) = \left(10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\right) \times (32, 64, 128, 256)\,. \tag{6.48}$$

For a start, it is worth investigating whether the SI$^+$-WB-FV method evolves the Kelvin-Helmholtz instability properly. In order to do that, we check the distribution of the local Mach number $\mathcal{M}_{\mathrm{loc}}$ at time $t/t_{\max} = 1/6$ computed on 126×64 control volumes. The results presented in Fig. 6.9 show that the method can resolve the vortexes independently of the Mach number $\mathcal{M}$ in the respective setup and that there is no visible qualitative difference between the solutions.

These findings are supported by the time evolution of the y-direction kinetic energy $E_{\mathrm{kin},y} = \sum_{ij} (\rho v^2)_{ij}/2$ and the total magnetic energy $E_{\mathrm{mag}} = \sum_{ij} |\mathbf{B}_{ij}|^2/2$, which are

N	$L^1(\rho)$	$L^1(u)$	$L^1(p)$	Iterations	Residual
20	0.0000E+00	0.0000E+00	0.0000E+00	0	0.0000E+00
40	0.0000E+00	0.0000E+00	0.0000E+00	0	0.0000E+00
80	0.0000E+00	0.0000E+00	0.0000E+00	0	0.0000E+00
160	0.0000E+00	0.0000E+00	0.0000E+00	0	0.0000E+00

Table 6.1: Isothermal atmosphere. The errors are measured in L^1 norm at time $t_f = 1$ for the density, velocity and pressure using different numbers of cells ($N = N_x = N_y$). The maximum number of the iterations and the corresponding residuals needed in the linear solver for the pressure system are also reported.

presented in Fig. 7.9 and Fig. 7.10 in Appendix B.2. The respective time evolutions are independent of the Mach number $\mathcal{M}$. Moreover, they converge very quickly towards the reference solution, which is given by the 1024×512 run of the IESS scheme, and the results for the 256×128 grid are qualitatively comparable to the respective result of the IESS scheme.

6.3.5 Isothermal Atmosphere for Euler

In order to verify the well-balanced property of the SI-WB-FV scheme, we simulate the isothermal atmosphere described in Sect. 4.6.4. The initial data for the hydrodynamic variables is again given by

$$(\rho, \mathbf{v}, p)(x, y, 0) = \left(1.21e^{-1.21(x+y)}, \mathbf{0}, e^{-1.21(x+y)}\right),$$

the magnetic field $\mathbf{B}$ is set to zero and the gravitational acceleration is defined by $\mathbf{g} = (-1, -1, 0)$. We run this test until a final time $t_f = 1$ on a domain $\mathcal{I} = [0, 1]^2$ with different numbers of cells ($N = N_x = N_y$). As boundary conditions we enforce the exact solution. The numerical results of this test in Tab. 6.1 show that no iterations are needed to solve the pressure linear system (6.33) because the right-hand side is perfectly balanced by the pressure terms. As a consequence, the residual has the order of the machine accuracy and the solution is exactly preserved throughout the simulation.

6.3.6 Perturbation of an Isothermal Atmosphere for Euler

The well-balanced property shall enable the numerical method to resolve small perturbations of equilibria even on coarse grids. Keeping this in mind, we add a pressure perturbation to the isothermal equilibrium, i.e.

$$p(x, y, 0) = e^{-1.21(x+y)} + \eta e^{-121\left((x-0.5)^2+(y-0.5)^2\right)}. \tag{6.49}$$

The parameter η regulates the size of the perturbation. We perform the test on a 64×64-grid with the non-well-balanced method SI-FV as well as with the well-balanced method SI-WB-FV. For the well-balanced scheme, the unperturbed initial data is used for $\tilde{\mathcal{Q}}$.

We carry out the test with three different levels of perturbation: $\eta = (10^{-1}, 10^{-5}, 10^{-10})$. In Fig. 6.10, we present the perturbations rescaled by η at final time $t_f = 0.15$. The largest perturbation is resolved equally well by both methods. For smaller perturbations, however, only the well-balanced method is able to resolve the perturbation on this coarse grid. The truncation error is too large for the method without well-balancing, so that its amplitude is larger than the perturbation itself. This result is consistent with those in Sect. 4.6.6 and in the literature [BCK23, CK15, KPS19].

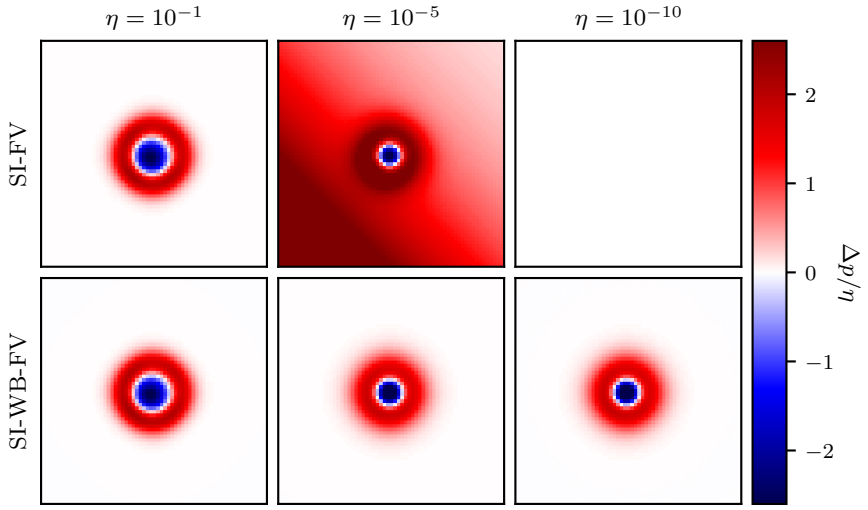

Figure 6.10: Pressure perturbations of varying strength η of an isothermal atmosphere at $t_f = 0.15$ resolved by the SI-FV (top row) and SI-WB-FV (bottom row) scheme. The perturbations are rescaled by the corresponding value of η.

6.3.7 MHD Steady State

The following test is intended to show that the method can also maintain equilibria exactly, which have a nonzero magnetic field. For this purpose, we modify the isothermal equilibrium from the former sections by adding a magnetic field. The initial values on the domain $\mathcal{I} = [0,1]^2$ are then given by

$$(\rho, \mathbf{v}, p, \mathbf{B})(x, y, 0) = \left(2.21e^{-(x+y)},\ \mathbf{0},\ 1.21e^{-(x+y)},\ e^{-\frac{1}{2}(x+y)},\ -e^{-\frac{1}{2}(x+y)},\ 0\right), \quad (6.50)$$

and we set $\gamma = 1.4$. The boundary conditions rely on the exact solution. The L^1-errors for the well-balanced scheme at time $t_f = 1$ are zero just as for the HSE before, which shows that the MHSE is preserved exactly. Consequently, the method is also well-balanced for MHSE as stated in Theorem 6.2.1.

6.3.8 Perturbation of an MHD Steady State

As already done for the isothermal Euler equilibrium, we also add a perturbation to the MHD equilibrium (6.50). The initial pressure is then given by

$$p(x, y, 0) = 1.21e^{-(x+y)} + \eta e^{-100\left((x-0.5)^2+(y-0.5)^2\right)}. \quad (6.51)$$

This test is also carried out with perturbations of different strength: $\eta = 10^{-1}, 10^{-5}, 10^{-10}$. Fig. 6.11 shows the results on a 64×64 grid for the well-balanced and the non-well-balanced method at final time $t_f = 0.15$. The behavior is similar to the one for the Euler equilibrium in Sect. 6.3.6. The non-well-balanced method can only resolve the large perturbation, but already fails for the medium perturbation. The well-balanced method, on the other hand, reproduces all three perturbations accurately. It should be noted that the altered shape of the Gaussian perturbation in comparison to the one of the Euler equilibrium results from the newly introduced force of the magnetic field.

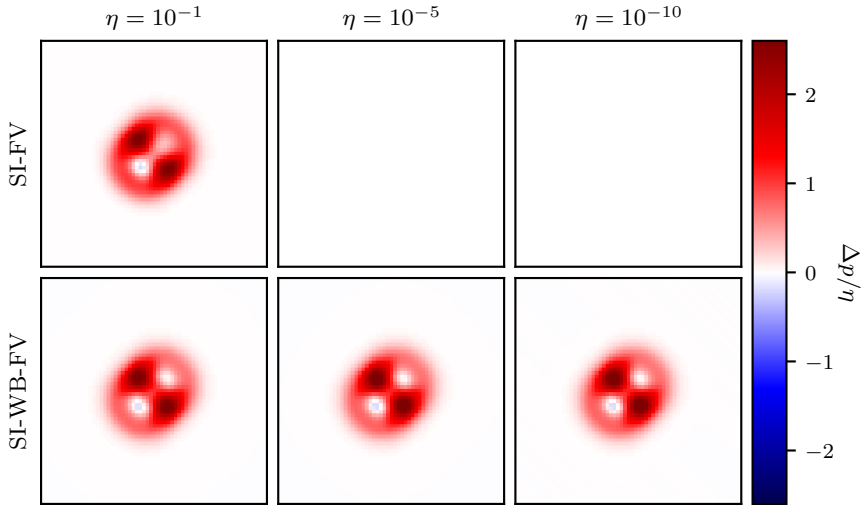

Figure 6.11: Pressure perturbations of varying strength η of an MHD equilibrium at $t_f = 0.15$ resolved by the SI-FV (top row) and SI-WB-FV (bottom row) scheme. The perturbations are rescaled by the corresponding value of η.

6.3.9 Euler Vortex in a Gravitational Field

After demonstrating low Mach capabilities for tests without gravitational source terms in Sect. 6.3.3 and 6.3.4, this final test is carried out to show the low Mach property of the SI$^+$-WB-FV scheme for the inhomogeneous case. For this purpose, we use the vortex for the Euler equations described in Sect. 4.6.8. Since the vortex is placed on top of a hydrostatic equilibrium solution, we set $\tilde{Q}$ to this equilibrium state.
We compute the solution after one full turn of the vortex for different initial maximum Mach numbers $\mathcal{M}_{\max}$ on a 40×40 grid. Fig. 6.12 shows the distributions of the local Mach number $\mathcal{M}_{\text{loc}}$ at final time t_f and additionally the initial distribution for $\mathcal{M}_{\max} = 10^{-1}$ (top left). It is evident that the numerical scheme adds a certain amount of numerical dissipation to the approximate solution in comparison with the initial data, which dampens the local Mach number. The vortex, however, is well-resolved in each case and there is no qualitative difference between the solutions for different $\mathcal{M}_{\max}$. This again shows that the numerical dissipation is independent of the Mach number due to the implicit discretization of the acoustic terms in the flux splitting.

This finding is underlined by the time evolution of the total kinetic energy. Since the domain is a closed setup due to the periodic boundary conditions, the total kinetic energy should be conserved over time in the low Mach number limit. Fig. 6.13 shows the time evolution of the kinetic energy for different $\mathcal{M}_{\max}$ and grid resolutions $N = N_x = N_y$. It turns out that the loss of kinetic energy only depends on the grid resolution and not on the Mach number. The results are consistent with those in Sect. 4.6.8 and results in the literature [TPK20].

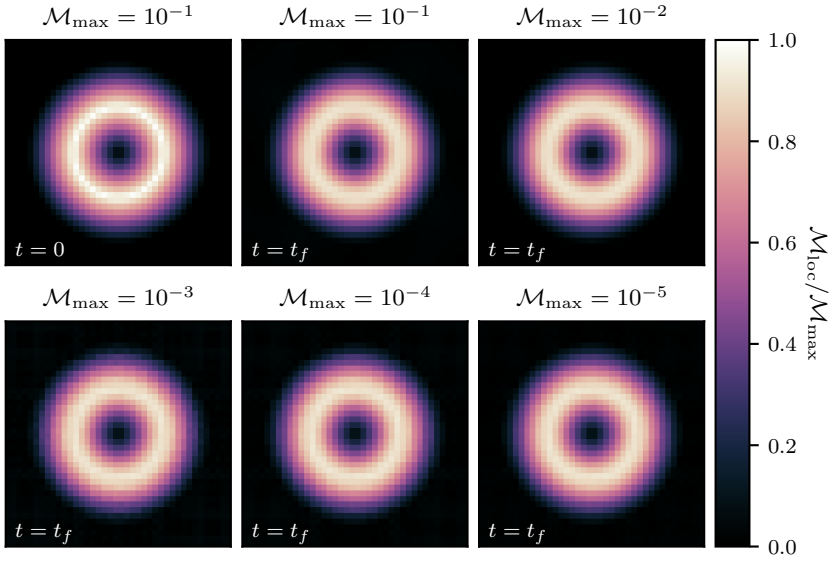

Figure 6.12: Distribution of the local Mach number $\mathcal{M}_{\rm loc}$ at $t_f = 0$ (top left) and after one turn for different maximum Mach numbers $\mathcal{M}_{\max}$. All panels are rescaled by the corresponding value of $\mathcal{M}_{\max}$.

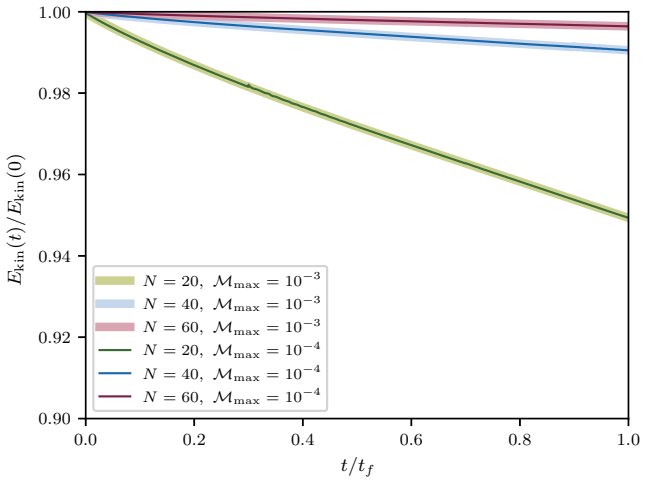

Figure 6.13: Time evolution of the total kinetic energy for different maximum Mach numbers $\mathcal{M}_{\max}$ and grid resolutions $N = N_x = N_y$ for the SI$^+$-WB-FV scheme.

6.4 Summary and Conclusions

In this chapter we have presented a new semi-implicit and well-balanced IMEX FV scheme for the compressible Euler and ideal MHD equations with gravitational source terms. The method is based on splitting the equations into its convective-type part, which is discretized explicitly in time, and the acoustic part, which is discretized implicitly. In

consequence, the CFL time step condition only depends on the slow dynamics in the explicit part. The implicit discretization generates an elliptic equation for the pressure, which is solved via GMRES. The resulting new pressure is used to update momentum and energy by finite difference operators without any numerical dissipation. This and the new CFL condition make the method particularly suitable for problems in the low Mach regime. However, numerical tests show that the method introduces density fluctuations of high magnitude in subsonic regimes, caused by an inconsistent update of density and momentum. The magnitude of the fluctuations can be significantly reduced by introducing a modified density update that relies on a central difference of the momentum at the new time level. Numerical results of the Balsara vortex and the Kelvin-Helmholtz instability show indeed that the modified method resolves low Mach flows accurately in different regimes. This finding is also valid under the influence of gravity, as shown in the results of the stationary vortex. The analysis of the kinetic energy loss numerically proves that the numerical diffusion of the new method is independent of the Mach number.

At the same time, the explicit Godunov-type method for the nonlinear flux terms ensures stability in the presence of shocks, as shown by the results of the Sod shock tube and the Orszag-Tang vortex.

The semi-implicit approach is combined with the Deviation Well-Balancing method to maintain (magneto-)hydrostatic equilibria exactly and to resolve flows around the equilibrium state even on coarse grids. The resulting scheme solves the equations for the deviation of the solution of an a priori known equilibrium solution. We have proven that, in the equilibrium, the elliptic pressure equation has only one solution, which is the equilibrium pressure itself. As a result, the solution is preserved exactly in the case of a (magneto-)hydrostatic equilibrium so that the scheme is well-balanced. Numerical tests demonstrate that arbitrary equilibria are exactly preserved and even very small perturbations are resolved accurately.

In order to ensure the solenoidal property of the magnetic field at the discrete level, a Contact-CT method is embedded in the scheme. In all numerical tests concerning the MHD equations, it could be checked that the magnetic field remains divergence-free up to rounding errors.

The new semi-implicit IMEX method thus represents an interesting alternative to fully time-implicit methods for simulating flows in low Mach number regimes. The development of density fluctuations in the incompressible limit, though, deserves more attention and offers potential for further improvements.

Chapter 7

Conclusion and Outlook

The simulation of fluid flows in the interior of stars is challenging and computationally expensive. Therefore, numerical methods must be particularly efficient in the considered physical regime in the sense that they (i) do not introduce excessive dissipation, (ii) allow a large time step and (iii) preserve (magneto-)hydrostatic equilibria. In this context, we have presented three different numerical methods which are based on very different approaches.

When searching for a method that fulfills all three requirements, we first focused on the design of approximate Riemann solvers. Within the concept of relaxation solvers, we have constructed a new Suliciu-type relaxation system whose solutions are approximations of the solutions of the Euler equations with gravity as long as the subcharacteristic condition is met. The characteristic fields of the derived relaxation system are linearly degenerate so that the corresponding Riemann problem is easy to solve. The resulting solver is an approximate Riemann solver for the inhomogeneous Riemann problem of the Euler equations. By relaxing the gravitational potential in the relaxation system, a discrete formulation of the hydrostatic equilibrium equation can be incorporated into the solver, ensuring that certain families of as well as a priori known hydrostatic equilibria can exactly be preserved. Furthermore, we employ the two-speed approach in the relaxation system. The different scaling of the two relaxation speeds ensures that the numerical dissipation does not increase for low Mach numbers and that the method is asymptotic-preserving in the incompressible limit. The key to this positive result is the influence of the second relaxation speed on the intermediate pressure state of the Riemann solver by which the Mach number is canceled out in the dissipation term. A side effect of the rescaling is that the dissipation is increased within the intermediate velocity. However, this dissipation term remains bounded and does not increase with decreasing Mach numbers. The choice of relaxation speeds respects the subcharacteristic condition so that the solver satisfies an entropy inequality and is positivity-preserving. From the entropy inequality follows that the solver suppresses checkerboard modes in velocity and pressure variables. These properties guarantee that the solver is particularly stable. The price that must be paid for this stability is a CFL condition that is more restrictive than the classic one. The resulting scheme thus only fulfills requirements (i) and (iii), but not requirement (ii).

The second scheme in this thesis solves the ideal MHD equations and addresses all three requirements. The dissipation is reduced by a low Mach fix in the HLLD solver, which acts only on the intermediate pressure. Within this intermediate state, the mechanism of the fix is very similar to that of the two-speed approach. The fix, however, is artificially built

into the solver and not based on a fundamental mathematical derivation. It is therefore not entirely clear what impact the fix has on the stability of the solver. However, numerical experiments in the low Mach number regime show no negative stability effects by using the fix. The CFL condition is relaxed by using time-implicit discretization techniques. Within the IESS algorithm, the continuity, momentum and energy equations are solved implicitly in time, while the induction equation is treated time-explicitly. This means that the CFL condition only depends on the Alfvén wave speed and is therefore independent of the Mach number. Numerical tests indicate that this approach is computationally cheaper in comparison with fully explicit schemes for Mach numbers smaller than 10^{-2}. The separate explicit discretization of the induction equation offers the advantage of being able to easily add a staggered Contact CT method to the scheme. The CT method ensures that the cell volume-averaged divergence of the magnetic field remains within the order of the machine accuracy, suppressing the evolution of unphysical magnetic monopoles and increasing the stability of the scheme. A priori known magnetohydrostatic solutions can be preserved by the Deviation Well-Balancing method, which prevents spurious flows and reduces numerical errors in simulations with steep stratifications. Numerical experiments show that the new-formed method is indeed capable of accurately resolving flows characterized by low Mach numbers in strongly stratified setups, and is significantly more efficient in such regimes than standard explicit methods are, due to the partially implicit time integration.

The third method relies on an IMEX approach to address the low Mach requirements. The MHD system is splitted into a slow scale convective-type and a fast scale pressure-type sub-system. The splitting ensures that only the wave speeds of the pressure-type sub-system depend on the Mach number. The convective-type sub-system is discretized time-explicitly by a Godunov-type method using the Rusanov solver, whereas the pressure-type sub-system is discretized implicitly and uses finite difference operators without any dissipation. The implicit part constitutes a linear system for the pressure, which is solved iteratively via GMRES. Then, based on the updated pressure, momentum and energy can be evolved in time. As a consequence of this discretization, the dissipation terms of the method are independent of the Mach number and do not introduce excessive dissipation. The CFL condition refers exclusively to the convective sub-system and thus allows time step sizes independent of the Mach number, while remaining restrictive enough to allow the scheme to capture shocks in supersonic regimes. Since the induction equation is completely contained in the explicit part, the staggered Contact CT method can be applied to keep a divergence-free magnetic field. The third requirement is satisfied by using the Deviation Well-Balancing method. It can be proven that the pressure equation only has one solution in the equilibrium case, which is the equilibrium pressure itself. We can deduce from this result that the other conservative variables are also exactly preserved. The IMEX method presented is applied to a number of numerical experiments in the subsonic regime and near magnetohydrostatic equilibria, which underline that the method fulfills the requirements (i)-(iii).

Overall, we can conclude that the two-speed scheme in its current form is not efficient enough to be applied in very subsonic regimes because of its restrictive CFL condition. Nevertheless, the numerical analysis of the method gives us interesting insights into the relation between low Mach compliant dissipation terms and stability properties of approximate Riemann solvers. One interesting aspect is that the stability of low-dissipation solvers such as LHLLD can potentially be increased by also incorporating the low Mach fix factor in the intermediate state of the velocity without leading to excessive dissipation

in the low Mach number regime.
The second method that uses a partially implicit time integration in combination with the low Mach compliant LHLLD solver has proven to be a very good option in the numerical experiments presented and can currently be considered as the best option in practice. However, the approach requires the laborious implementation of implicit solvers, which then have to solve large nonlinear systems of equations. At this point, the IMEX approach offers the potential to increase the efficiency in the low Mach number regime, since only a smaller implicit part has to be solved. In the future, it would be interesting to compare the computational costs of the two methods under identical conditions in order to compare their efficiency. Such a study could be done similarly to the one that is carried out in [LAB+23] for different approximate Riemann solvers.
In addition, the construction of the IMEX method itself leaves room for further improvement. The development of density fluctuations in the low Mach number regime needs a more detailed investigation in order to revise the design of the method such that it mimics the limit behaviour on the discrete level. An alternative strategy that might help in this context could be an extension of the IMEX method presented in [TPK20] from the Euler to the ideal MHD equations. The method therein relies on a *Klein splitting* [Kle95] of the pressure within a relaxation system. Implicit discretized terms only appear in the additional relaxation equations, whereas the conservative variables of the original PDE system are completely updated in the closed form of a Godunov-type method. Thus, density and momentum are evolved simultaneously and in a consistent way within the explicit part of the IMEX method, which potentially reduces the fluctuations.

Appendices

A An Implicit-Explicit Strang Splitting Method

A.1 Magnetized Kelvin-Helmholtz Instability

This appendix explores the effects of the grid resolution and strength of the initial magnetic field on the evolution of the Kelvin-Helmholtz instability shown in Sect. 5.4.2. Firstly, Fig. 7.1 illustrates the evolution of B_x in time on a fine grid with 2048×1024 control volumes. In Fig. 7.2, we present the results of nine different simulations with initial magnetic fields of different strength. It shows that strong magnetic fields prevent the growth of the instability. The simulations with different resolutions show in Fig 7.3 that secondary instabilities arise in the magnetic field for low resolution runs, which are purely caused by discretization errors and disappear in the solutions generated on finer grids ($N > 128$).

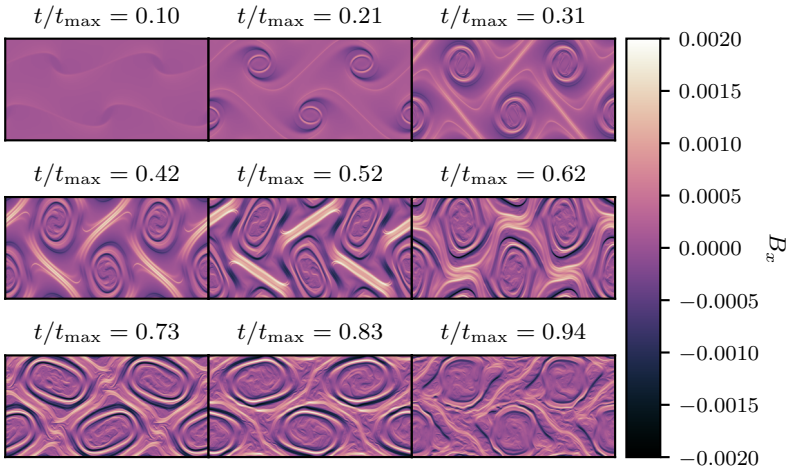

Figure 7.1: Time evolution of B_x in the simulations of the Kelvin-Helmholtz instability starting from the initial conditions described in Sect. 5.4.2 with $\mathcal{M} = 10^{-3}$. The grid consists of 2048×1024 cells.

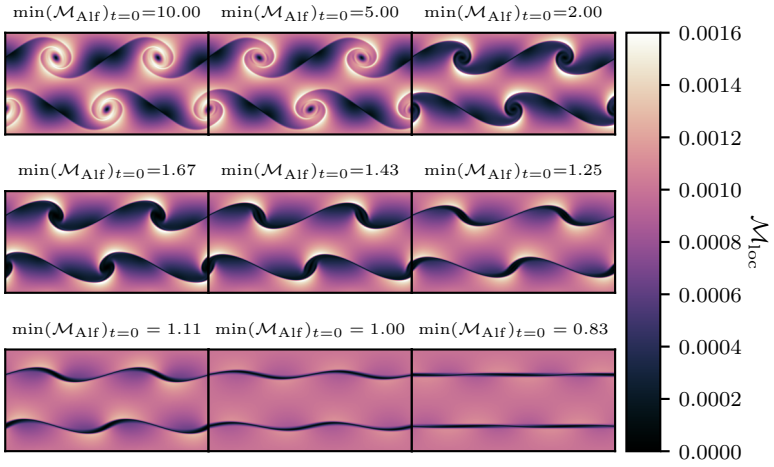

Figure 7.2: Distribution of the local sonic Mach number $\mathcal{M}_{\rm loc}$ in the simulations of the Kelvin-Helmholtz instability at $t/t_{\rm max} = 1/6$ for different values of the initial magnetic field B_x, computed as $B_x = \sqrt{\gamma}\alpha$, with $\alpha = (0.1, 0.2, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.2)$. These simulations are performed on a 512×256 grid with $\mathcal{M} = 10^{-3}$. The title in each panel is the corresponding minimum Alfvén Mach number of the flow at $t = 0$. For an initial magnetic field that is strong enough, the magnetic stresses prevent the growth of the instability.

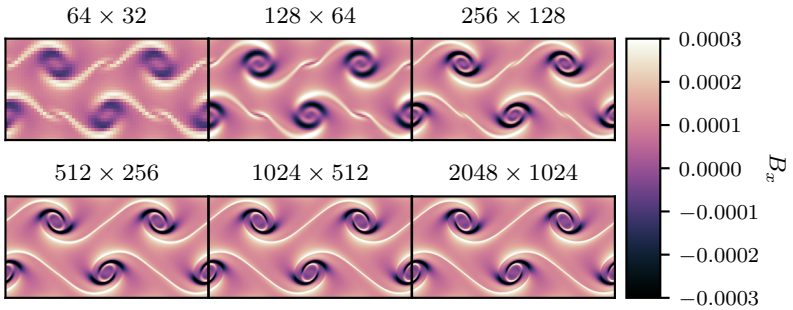

Figure 7.3: Distribution of B_x in the simulations of the Kelvin-Helmholtz instability at $t/t_{\rm max} = 1/6$ for different grid resolutions, starting from the initial conditions described in Sect. 5.4.2 with $\mathcal{M} = 10^{-3}$. On grids with $N \leq 128$, numerical discretization errors generate grid-scale vorticity, which leads to the growth of secondary instabilities in the regions between the primary rolls. This effect does not appear in better converged simulations.

A.2 Hot Bubble

In this section, we extend the study of the "hot bubble" described in Sect. 5.4.3. In particular, we show the dependence of the entropy fluctuations, p_B and $\mathcal{M}$ on the magnitude of the initial entropy perturbation $(\Delta A/A)_{t=0}$.

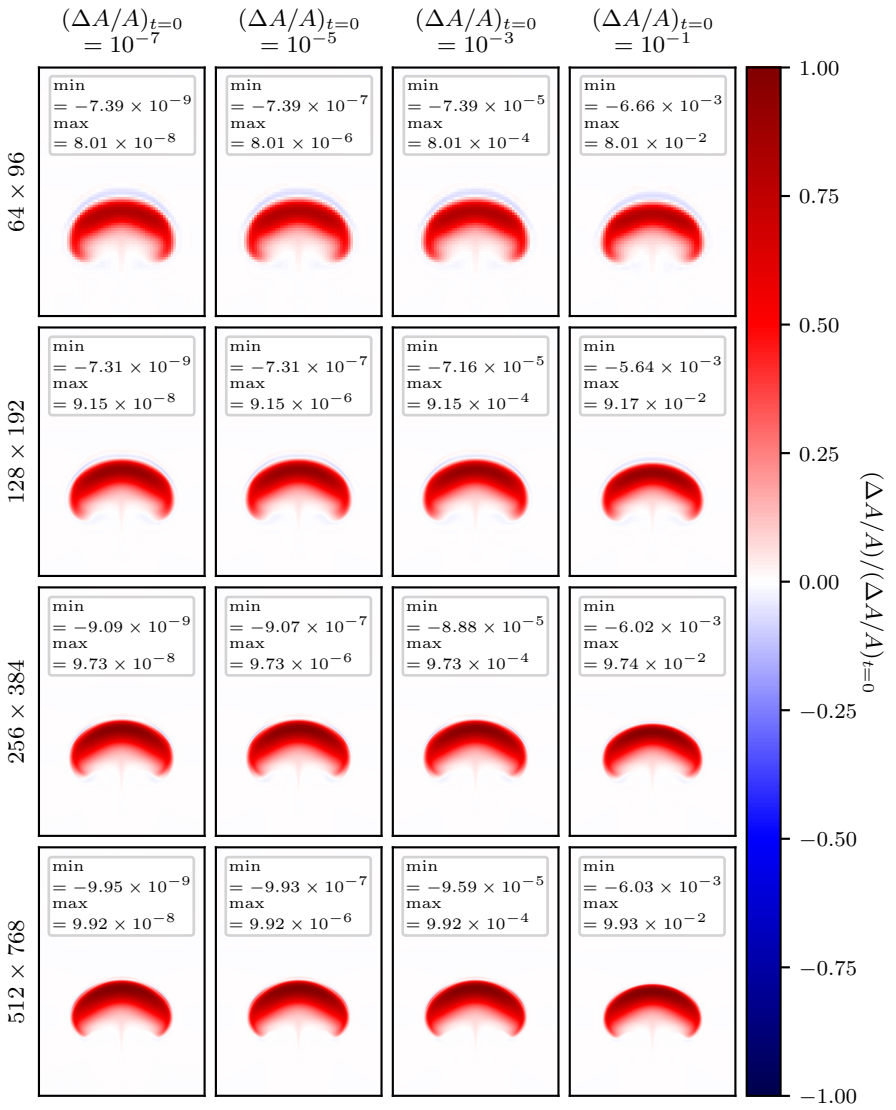

Figure 7.4: Final distribution of the entropy fluctuations $\Delta A/A$ of the hot bubble for different values of $(\Delta A/A)_{t=0}$ and grid resolutions. Each panel is rescaled by the corresponding value of $(\Delta A/A)_{t=0}$. The insets provide the minimum and maximum values of the entropy fluctuations in each plot.

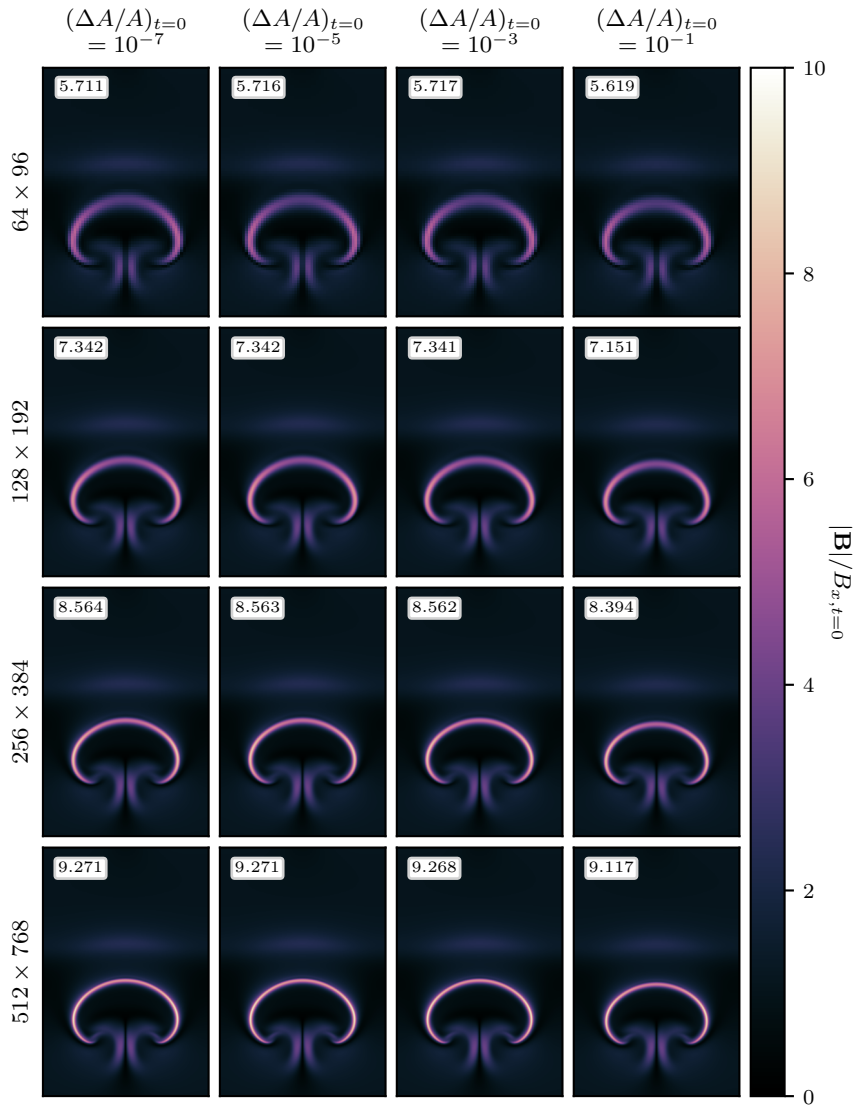

Figure 7.5: Final distribution of $|\mathbf{B}|/B_{x,t=0}$ for different values of $(\Delta A/A)_{t=0}$ and grid resolutions in the simulations of the hot bubble. The insets show the maximum ratio in each panel. The amount of numerical resistivity decreases upon grid refinement, which leads to the generation of narrower stripes with stronger magnetic fields.

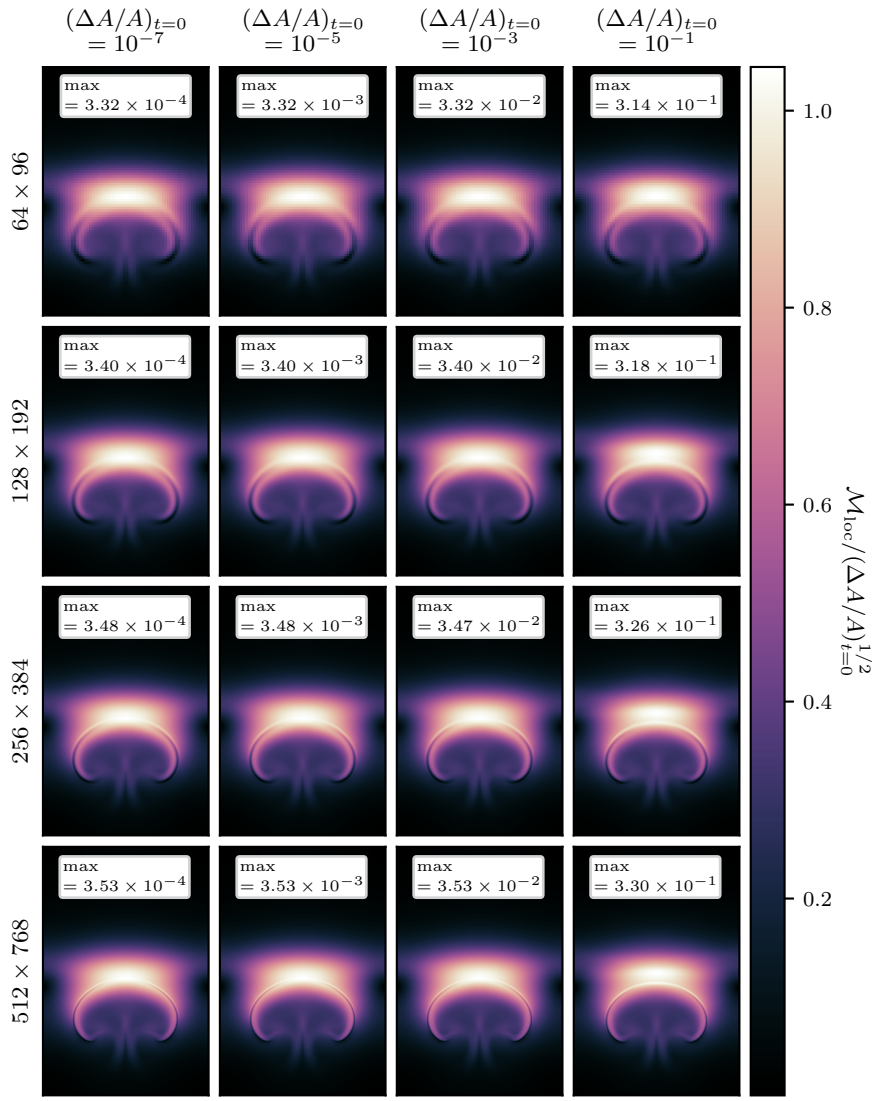

Figure 7.6: Final distribution of the sonic Mach number of the hot bubble for different values of $(\Delta A/A)_{t=0}$ and different grid resolutions. Each panel is rescaled by the corresponding value of $(\Delta A/A)^{1/2}_{t=0}$. The insets show the maximum sonic Mach number. An entropy perturbation of $(\Delta A/A)_{t=0} = 0.1$ drives flows that are far from the low Mach number regime. In this case, effects of compressibility caused by the high ram pressure of the bubble are large enough to cause a $6 - 7\%$ deviation from the theoretical scaling discussed in Sect. 5.4.3.

B A Semi-Implicit IMEX Method

B.1 Balsara Vortex

In this part, we briefly investigate the intensity of density fluctuations within the solutions of the Balsara vortex. The left plot in Fig. 7.7 shows that the fluctuations decrease with second order in terms of the grid resolution for both versions of the semi-implicit method. The fluctuations for the modified scheme SI$^+$-WB-FV are several orders smaller in comparison to the ones produced by SI-WB-FV, which is a positive effect of the modified density update (6.43). The right plot in Fig. 7.7 illustrates the magnitude of the fluctuations in relation to the maximum Mach number of the different runs. The maximum of the absolute value of the fluctuations does not change at all with the Mach number and remains on a high level for the SI-WB-FV method. For the modified version, we can observe a small decrease in the results from $\mathcal{M}_{\max} = 1.55 \times 10^{-1}$ and $\mathcal{M}_{\max} = 1.55 \times 10^{-2}$, but for smaller Mach numbers the magnitude is constant and therefore independent of the maximum Mach number. Since the divergence of the velocity field for this type of semi-implicit method has the order of the time step in the incompressible limit [BP21], and the time step is chosen independently of the Mach number due to the IMEX approach, the divergence does not decrease for smaller Mach numbers. This has the effect that the divergence does not vanish in the discretization of the continuity equation and Mach number independent pile-ups of density fluctuations arise. Further investigations into this process will need to be carried out in the future.
For a comparison, we show the scaling for the IESS scheme using the LHLLD solver in Fig. 7.8. The maximum fluctuations for this method scale with $\mathcal{O}(\mathcal{M}^2)$.

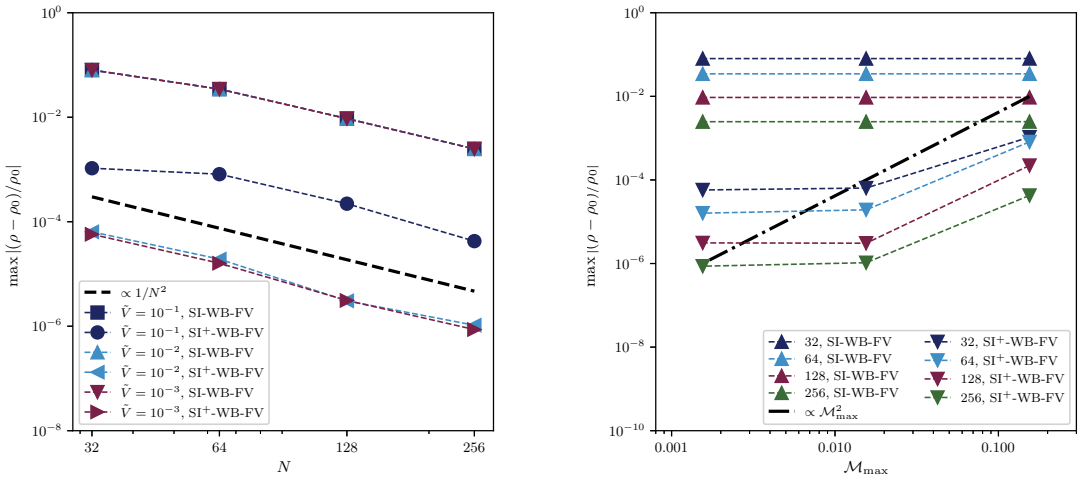

Figure 7.7: Maximum absolute density fluctuation over the grid for the Balsara vortex at final time t_f in the results of two versions of the semi-implicit method for grid resolutions $N \times N$ with $N = 32, 64, 128, 256$ (left) and maximum rotational velocities $\tilde{V} = 10^{-1}, 10^{-2}, 10^{-3}$ (right).

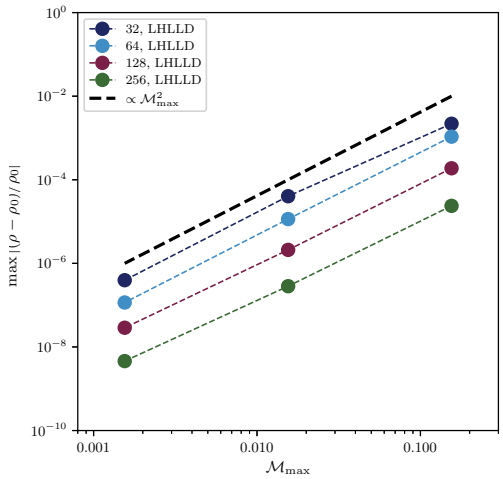

Figure 7.8: Maximum absolute density fluctuation over the grid for the Balsara vortex at final time t_f of the IESS method using the LHLLD solver for different grid resolutions $N \times N$ and maximum rotational velocities $\tilde{V}$.

B.2 Magnetized Kelvin-Helmholtz Instability

In order to underline the low Mach capabilities of the semi-implicit method, we analyze the time evolution of the y-direction kinetic energy and the magnetic energy. The test is run with the SI$^+$-WB-FV scheme on the following range of initial Mach numbers $\mathcal{M}$ and grid resolutions $(N_x \times N_y)$:

$$(\mathcal{M}) \times (N_x \times N_y) = (10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}) \times (64 \times 32, 128 \times 64, 256 \times 128). \quad (7.1)$$

For comparison, we include the solutions generated by the IESS scheme presented in Chapter 5 in Fig. 7.9 and 7.10. The time evolutions for both kinetic and magnetic energy are independent of the Mach number $\mathcal{M}$, which underlines the low Mach number compliance of the semi-implicit method. Additionally, the results show good agreement with those of the IESS scheme on the 256×128 grid.

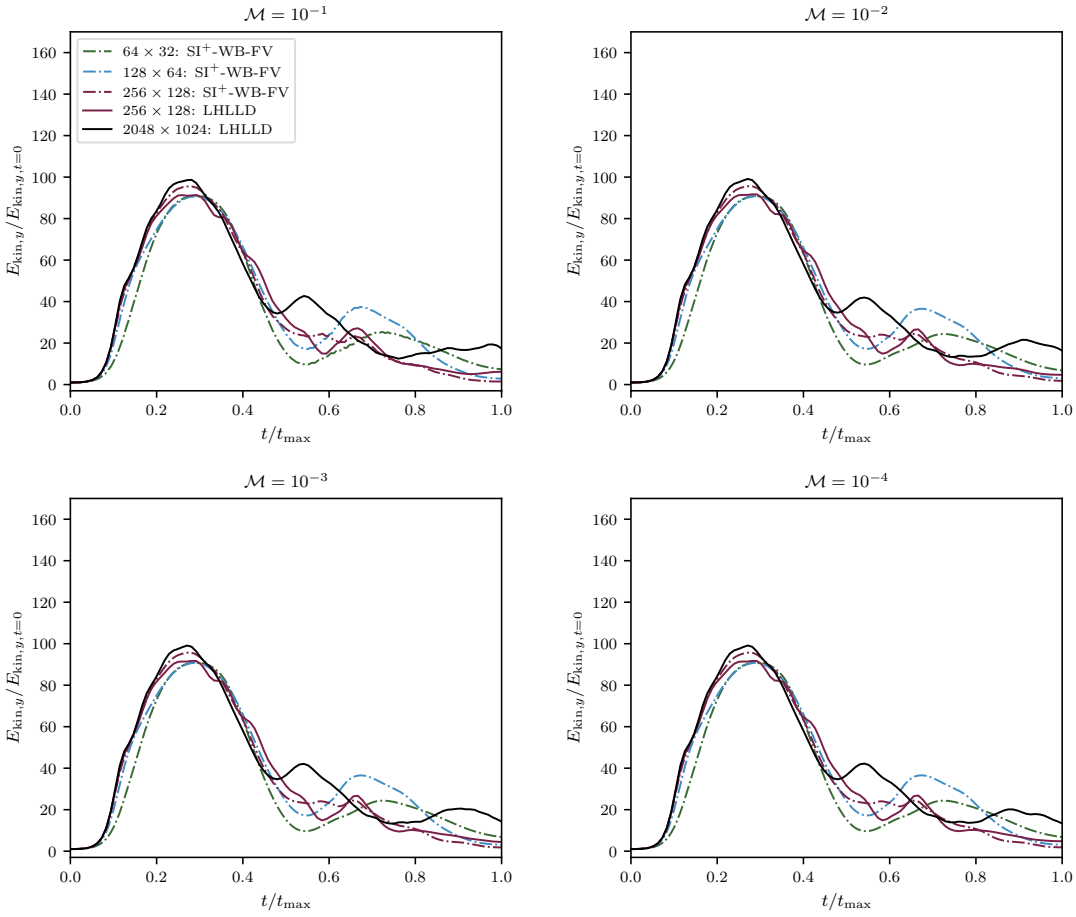

Figure 7.9: Time evolution of the y-direction kinetic energy rescaled by its initial value in the magnetized Kelvin-Helmholtz instability test problem. Each panel corresponds to a different initial Mach number $\mathcal{M}$. Different colors are used for different grid resolutions. Results obtained with the SI$^+$-WB-FV are represented by dot-dashed lines. The results are compared to those of the IESS scheme with the LHLLD solver, represented by solid lines.

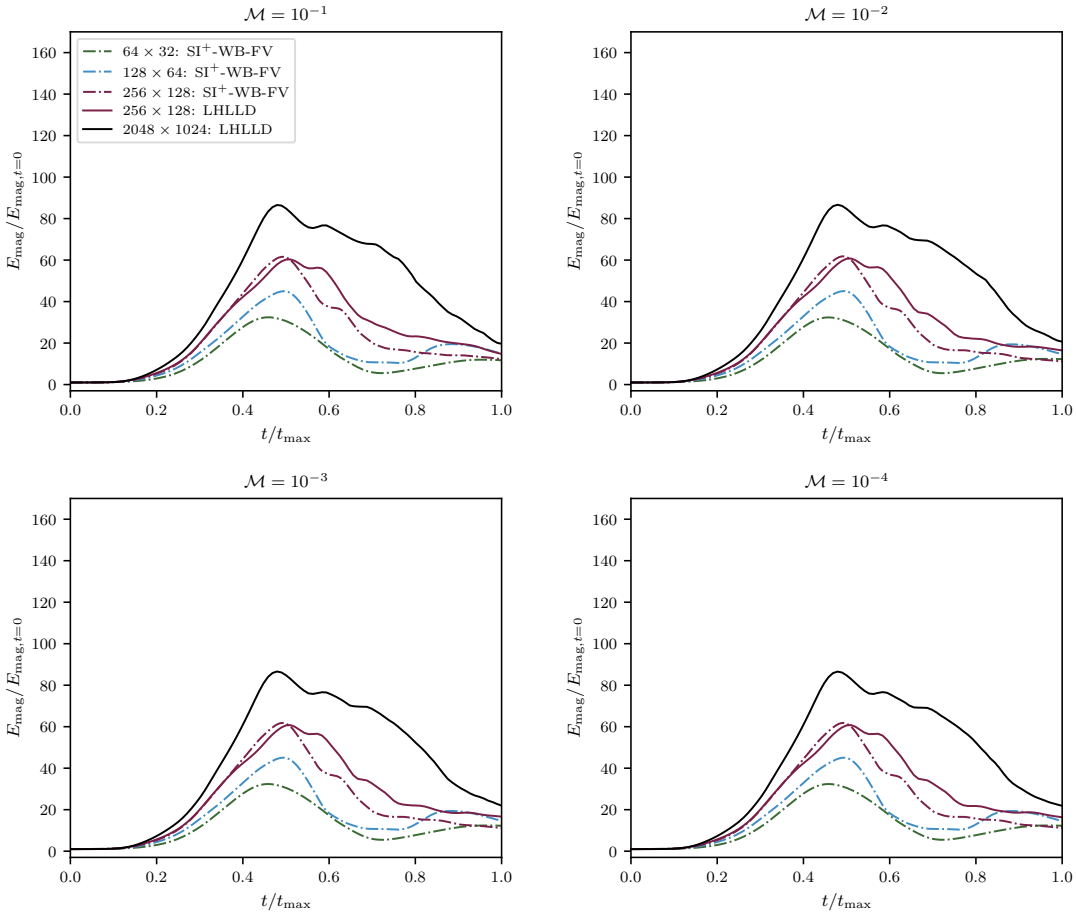

Figure 7.10: Time evolution of the magnetic energy rescaled by its initial value in the magnetized Kelvin-Helmholtz instability test problem. Each panel corresponds to a different initial Mach number $\mathcal{M}$. Different colors are used for different grid resolutions. Results obtained with the SI$^+$-WB-FV are represented by dot-dashed lines. The results are compared to those of the IESS scheme with the LHLLD solver, represented by solid lines.

Bibliography

[ALB] R. Andrássy, G. Leidi, and W. Barsukow. Private communication.

[ALJ99] T. Amari, J. F. Luciani, and P. Joly. A Preconditioned Semi-Implicit Method for Magnetohydrodynamics Equations. *SIAM Journal on Scientific Computing*, 21(3):970–986, 1999.

[Bal04] D. S. Balsara. Second-Order-accurate Schemes for Magnetohydrodynamics with Divergence-free Reconstruction. *The Astrophysical Journal Supplement Series*, 151(1):149–184, 2004.

[Bal10] D. S. Balsara. Multidimensional HLLE Riemann solver: Application to Euler and magnetohydrodynamic flows. *Journal of Computational Physics*, 229(6):1970–1993, 2010.

[Bal12] D. S. Balsara. A two-dimensional HLLC Riemann solver for conservation laws: Application to Euler and magnetohydrodynamic flows. *Journal of Computational Physics*, 231(22):7476–7503, 2012.

[Bar] W. Barsukow. Private communication.

[Bar99] T. J. Barth. *Numerical Methods for Gasdynamic Systems on Unstructured Meshes*, pages 195–285. Springer, Berlin Heidelberg, 1999.

[Bar21] W. Barsukow. Truly multi-dimensional all-speed schemes for the Euler equations on Cartesian grids. *Journal of Computational Physics*, 435:110216, 2021.

[BB80] J. U. Brackbill and D. C. Barnes. The Effect of Nonzero $\nabla \cdot \mathbf{B}$ on the numerical solution of the magnetohydrodynamic equations. *Journal of Computational Physics*, 35(3):426–430, 1980.

[BBK24] C. Birke, W. Boscheri, and C. Klingenberg. A Well-Balanced Semi-implicit IMEX Finite Volume Scheme for Ideal Magnetohydrodynamics at All Mach Numbers. *Journal of Scientific Computing*, 98(2):34, 2024.

[BCG20] 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. *Numerische Mathematik*, 145(1):35–76, 2020.

[BCK18] J. P. Berberich, P. Chandrashekar, and C. Klingenberg. A General Well-Balanced Finite Volume Scheme for Euler Equations with Gravity. In

C. Klingenberg and M. Westdickenberg, editors, *Theory, Numerics and Applications of Hyperbolic Problem I*, volume 236 of *Springer Proceedings in Mathematics & Statistics*, pages 151–163, Springer, Cham, 2018.

[BCK21] J. P. Berberich, P. Chandrashekar, and C. Klingenberg. High order well-balanced finite volume methods for multi-dimensional systems of hyperbolic balance laws. *Computers & Fluids*, 219:104858, 2021.

[BCK23] C. Birke, C. Chalons, and C. Klingenberg. A low Mach two-speed relaxation scheme for the compressible Euler equations with gravity. *Communications in Mathematical Sciences*, 21(8):2213–2246, 2023.

[BCKR19] J. P. Berberich, P. Chandrashekar, C. Klingenberg, and F. K. Röpke. Second order finite volume scheme for Euler equations with gravity which is well-balanced for general equations of state and grid systems. *Communications in Computational Physics*, 26:599–630, 2019.

[BCLC97] P. Batten, N. Clarke, C. Lambert, and D. M. Causon. On the Choice of Wavespeeds for the HLLC Riemann Solver. *SIAM Journal on Scientific Computing*, 18(6):1553–1570, 1997.

[BCP00] A. Bressan, G. Crasta, and B. Piccoli. Well-posedness of the Cauchy problem for $n \times n$ systems of conservation laws. *Memoirs of the American Mathematical Society*, 146(694):1–134, 2000.

[BdL09] F. Bouchut and T. Morales de Luna. Semi-discrete Entropy Satisfying Approximate Riemann Solvers. The Case of the Suliciu Relaxation Approximation. *Journal of Scientific Computing*, 41(3):483–509, 2009.

[BDL+20] W. Boscheri, G. Dimarco, R. Loubère, M. Tavelli, and M. H. Vignal. A second order all Mach number IMEX finite volume solver for the three dimensional Euler equations. *Journal of Computational Physics*, 415:109486, 2020.

[BDT21] W. Boscheri, G. Dimarco, and M. Tavelli. An efficient second order all Mach finite volume solver for the compressible Navier-Stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 374:113602, 2021.

[BEK+17] W. Barsukow, P. V. F. Edelmann, C. Klingenberg, F. Miczek, and F. K. Röpke. A Numerical Scheme for the Compressible Low-Mach Number Regime of Ideal Fluid Dynamics. *Journal of Scientific Computing*, 72(2):623–646, 2017.

[BEKR17] W. Barsukow, P. V. F. 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.

[Ber05] C. Berthon. Stability of the MUSCL Schemes for the Euler Equations. *Communications in Mathematical Sciences*, 3(2):133–157, 2005.

[Ber20] J. P. Berberich. *Fluids in Gravitational Fields - Well-Balanced Modifications for Astrophysical Finite-Volume Codes*. Dissertation, Julius-Maximilians-Universität Würzburg, 2020.

[BF18] J. Březina and E. Feireisl. Measure-valued solutions to the complete Euler system. *Journal of the Mathematical Society of Japan*, 70(4):1227–1245, 2018.

[BFN20] F. Bouchut, E. Franck, and L. Navoret. A Low Cost Semi-implicit Low-Mach Relaxation Scheme for the Full Euler Equations. *Journal of Scientific Computing*, 83(1):24, 2020.

[BFR16] S. Boscarino, F. Filbet, and G. Russo. High Order Semi-implicit Schemes for Time Dependent Partial Differential Equations. *Journal of Scientific Computing*, 68(3):975–1001, 2016.

[BK23a] C. Birke and C. Klingenberg. A Low Mach Number Two-Speed Relaxation Scheme for Ideal MHD Equations. In E. Franck, J. Fuhrmann, V. Michel-Dansac, and L. Navoret, editors, *Finite Volumes for Complex Applications X—Volume 2, Hyperbolic and Related Problems*, pages 43–51, Springer Nature Switzerland, Cham, 2023.

[BK23b] C. Birke and C. Klingenberg. Finding an Approximate Riemann Solver via Relaxation: Concept and Advantages. In *Proceedings of HYP XVIII*. SEMA SIMAI Springer Series, 2023.

[BKK+20] H. Al Baba, C. Klingenberg, O. Kreml, V. Mácha, and S. Markfelder. Nonuniqueness of Admissible Weak Solution to the Riemann Problem for the Full Euler System in Two Dimensions. *SIAM Journal on Mathematical Analysis*, 52(2):1729–1760, 2020.

[BKW07] F. Bouchut, C. Klingenberg, and K. Waagan. A multiwave approximate Riemann solver for ideal MHD based on relaxation. I: theoretical framework. *Numerische Mathematik*, 108(1):7–42, 2007.

[BKW10] F. Bouchut, C. Klingenberg, and K. Waagan. A multiwave approximate Riemann solver for ideal MHD based on relaxation II: numerical implementation with 3 and 5 waves. *Numerische Mathematik*, 115(4):647–679, 2010.

[BKZ20] C. Berthon, C. Klingenberg, and M. Zenk. An all Mach number relaxation upwind scheme. *The SMAI Journal of Computational Mathematics*, 6:1–31, 2020.

[BL16] F. Bouchut and X. Lhébrard. A 5-wave relaxation solver for the shallow water MHD system. *Journal of Scientific Computing*, 68:92–115, 2016.

[BLMY17] G. Bispen, M. Lukáčová-Medvid'ová, and L. Yelash. Asymptotic preserving IMEX finite volume schemes for low Mach number Euler equations with gravitation. *Journal of Computational Physics*, 335:222–248, 2017.

[Bou04] F. Bouchut. *Nonlinear Stability of Finite Volume Methods for Hyperbolic Conservation Laws, and well-balanced schemes for sources.* Frontiers in Mathematics. Birkhäuser Verlag, 2004.

[BP21] W. Boscheri and L. Pareschi. High order pressure-based semi-implicit IMEX schemes for the 3D Navier-Stokes equations at all Mach numbers. *Journal of Computational Physics*, 434:110206, 2021.

[BQRX19] S. Boscarino, J.-M. Qiu, G. Russo, and T. Xiong. A high order semi-implicit IMEX WENO scheme for the all-Mach isentropic Euler system. *Journal of Computational Physics*, 392:594–618, 2019.

[Bre84] Y. Brenier. Averaged Multivalued Solutions for Scalar Conservation Laws. *SIAM Journal on Numerical Analysis*, 21(6):1013–1037, 1984.

[BS99] D. S. Balsara and D. S. Spicer. A Staggered Mesh Algorithm Using High Order Godunov Fluxes to Ensure Solenoidal Magnetic Fields in Magnetohydrodynamic Simulations. *Journal of Computational Physics*, 149(2):270–292, 1999.

[BT22] W. Boscheri and M. Tavelli. High order semi-implicit schemes for viscous compressible flows in 3D. *Applied Mathematics and Computation*, 434:127457, 2022.

[But63] J. C. Butcher. Coefficients for the study of Runge-Kutta integration processes. *Journal of the Australian Mathematical Society*, 3(2):185–201, 1963.

[But64] J. C. Butcher. Implicit Runge-Kutta Processes. *Mathematics of Computation*, 18(85):50–64, 1964.

[CC90] S. Chapman and T. G. Cowling. *The Mathematical Theory of Non-uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction and Diffusion in Gases*. Cambridge Mathematical Library. Cambridge University Press, 3rd edition, 1990.

[CC14] C. Chalons and F. Coquel. Modified Suliciu relaxation system and exact resolution of isolated shock waves. *Mathematical Models and Methods in Applied Sciences*, 24(5):937–971, 2014.

[CCK+18] A. Chertock, S. Cui, A. Kurganov, S. N. Özcan, and E. Tadmor. Well-balanced schemes for the Euler equations with gravitation: Conservative formulation using global fluxes. *Journal of Computational Physics*, 358:36–52, 2018.

[CDK12] F. Cordier, P. Degond, and A. Kumbaro. An Asymptotic-Preserving all-speed scheme for the Euler and Navier-Stokes equations. *Journal of Computational Physics*, 231(17):5685–5704, 2012.

[CFL28] R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Mathematische Annalen*, 100:32–74, 1928.

[CGLGP08] M. J. 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 Journal of Numerical Analysis*, 46(2):1012–1039, 2008.

[CGP+01] F. Coquel, E. Godlewski, B. Perthame, A. In, and P. Rascle. *Some New Godunov and Relaxation Methods for Two-Phase Flow Problems*, pages 179–188. Springer US, New York, NY, 2001.

[CGS12] F. Coquel, E. Godlewski, and N. Seguin. Relaxation of fluid systems. *Mathematical Models and Methods in Applied Sciences*, 22(8):1250014, 2012.

[Cha61] S. Chandrasekhar. *Hydrodynamic and hydromagnetic stability.* 1961.

[CK15] P. Chandrashekar and C. Klingenberg. A Second Order Well-Balanced Finite Volume Scheme for Euler Equations with Gravity. *SIAM Journal on Scientific Computing*, 37(3):B382–B402, 2015.

[CLL94] G.-Q. Chen, C.-D. Levermore, and T.-P. Liu. Hyprbolic Conservation Laws with stiff Relaxation Terms and Entropy. *Communications on Pure and Applied Mathemathics*, 47(6):787–830, 1994.

[CP98] F. Coquel and B. Perthame. Relaxation of Energy and Approximate Riemann Solvers for General Pressure Laws in Fluid Dynamics. *SIAM Journal on Numerical Analysis*, 35(6):2223–2249, 1998.

[CRT14] A. J. Christlieb, J. A. Rossmanith, and Q. Tang. Finite difference weighted essentially non-oscillatory schemes with constrained transport for ideal magnetohydrodynamics. *Journal of Computational Physics*, 268:302–325, 2014.

[CW84] P. Colella and P. R. Woodward. The Piecewise Parabolic Method (PPM) for Gas-Dynamical Simulations. *Journal of Computational Physics*, 54(1):174–201, 1984.

[CWX23] W. Chen, K. Wu, and T. Xiong. High Order asymptotic preserving finite difference WENO Schemes with constrained transport for MHD equations in all sonic Mach numbers. *Journal Of Computational Physics*, 488:112240, 2023.

[CYX18] S. Chen, C. Yan, and X. Xiang. Effective low-Mach number improvement for upwind schemes. *Computers & Mathematics with Applications*, 75(10):3737–3755, 2018.

[Daf09] C. M. Dafermos. *Hyperbolic Conservation Laws in Continuum Physics.* Grundlehren der mathematischen Wissenschaften. Springer, 3rd edition, 2009.

[DBTF19] M. Dumbser, D. S. Balsara, M. Tavelli, and F. Fambri. A divergence-free semi-implicit finite volume scheme for ideal, viscous, and resistive magnetohydrodynamics. *International Journal for Numerical Methods in Fluids*, 89:16–42, 2019.

[Del09] S. Dellacherie. Checkerboard modes and wave equation. *Proceedings of the 18th Conference on Scientific Computing*, pages 71–80, 2009.

[Del10] S. Dellacherie. Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number. *Journal of Computational Physics*, 229(4):978–1016, 2010.

[DGWW18] D. Derigs, G. J. Gassner, S. Walch, and A. R. Winters. Entropy Stable Finite Volume Approximations for Ideal Magnetohydrodynamics. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 120(3):153–219, 2018.

[DKK+02] A. Dedner, F. Kemm, D. Kröner, C.-D. Munz, T. Schnitzer, and M. Wesenberg. Hyperbolic Divergence Cleaning for the MHD Equations. *Journal of Computational Physics*, 175(2):645–673, 2002.

[DLMDV18] 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. *Journal of Computational Physics*, 372:178–201, 2018.

[DW98] W. Dai and P. R. Woodward. On the Divergence-free Condition and Conservation Laws in Numerical Simulations for Supersonic Magnetohydrodynamical Flows. *The Astrophysical Journal*, 494(1):317–335, 1998.

[DZBK16] V. Desveaux, M. Zenk, C. Berthon, and C. Klingenberg. Well-balanced schemes to capture non-explicit steady states in the Euler equations with gravity. *International Journal for Numerical Methods in Fluids*, 81(2):104–127, 2016.

[Ede14] P. V. F. Edelmann. *Coupling of Nuclear Reaction Networks and Hydrodynamics for Application in Stellar Astrophysics*. Dissertation, Technische Universität München, 2014.

[EH88] C. R. Evans and J. F. Hawley. Simulation of Magnetohydrodynamic Flows: A Constrained Transport Model. *The Astrophysical Journal*, 332:659–677, 1988.

[EHB+21] P. V. F. Edelmann, L. Horst, J. P. Berberich, R. Andrássy, J. Higl, G. Leidi, C. Klingenberg, and F. K. Röpke. Well-balanced treatment of gravity in astrophysical fluid dynamics simulations at low Mach numbers. *Astronomy & Astrophysics*, 652:A53, 2021.

[Ein88] B. Einfeldt. On Godunov-Type Methods for Gas Dynamics. *SIAM Journal on Numerical Analysis*, 25(2):294–318, 1988.

[ERMS91] B. Einfeldt, P. L. Roe, C.-D. Munz, and B. Sjögreen. On Godunov-Type Methods near Low Densities. *Journal of Computational Physics*, 92(2):273–295, 1991.

[Fam21] F. Fambri. A novel structure preserving semi-implicit finite volume method for viscous and resistive magnetohydrodynamics. *International Journal for Numerical Methods in Fluids*, 93(12):3447–3489, 2021.

[FJRG96] A. Frank, T. W. Jones, D. Ryu, and J. B. Gaalaas. The Magnetohydrodynamic Kelvin-Helmholtz Instability: A Two-dimensional Numerical Study. *The Astrophysical Journal*, 460:777–793, 1996.

[FL71] K. O. Friedrichs and P. D. Lax. Systems of Conservation Equations with a Convex Extension. *Proceedings of the National Academy of Sciences of the United States of America*, 68(8):1686–1688, 1971.

[FMR09] F. G. Fuchs, S. Mishra, and N. H. Risebro. Splitting based finite volume schemes for ideal MHD equations. *Journal of Computational Physics*, 228(3):641–660, 2009.

[FP02] J. H. Ferziger and M. Perić. *Computational Methods for Fluid Dynamics*. Springer Berlin Heidelberg, 3rd edition, 2002.

[GC90] P. M. Gresho and S. T. Chan. On the theory of semi-implicit projection methods for viscous incompressible flow and its implementation via a finite element method that also introduces a nearly consistent mass matrix. Part 2: Implementation. *International Journal for Numerical Methods in Fluids*, 11:621–659, 1990.

[GCD21] E. Gaburro, M. J. Castro, and M. Dumbser. A Well Balanced Finite Volume Scheme for General Relativity. *SIAM Journal on Scientific Computing*, 43(6):B1226–B1251, 2021.

[Gli65] J. Glimm. Solutions in the large for nonlinear hyperbolic systems of equations. *Communications on Pure and Applied Mathematics*, 18(4):697–715, 1965.

[GM04] H. Guillard and A. Murrone. On the behavior of upwind schemes in the low Mach number limit : II. Godunov type schemes. *Computers & Fluids*, 33:655–675, 2004.

[God59] S. K. Godunov. Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics. *Matematičeskij Sbornik*, 47(3):271–306, 1959.

[God72] S. K. Godunov. Symmetric form of the magnetohydrodynamic equation. *Numerical Methods for Mechanics of Continuum Medium*, 3(1):26–34, 1972.

[GR02] E. Godlewski and P.-A. Raviart. Coupling nonlinear Hyperbolic Systems: mathematical and numerical analysis. In R. Herbin and D. Kröner, editors, *Finite Volumes for Complex Applications III*, pages 211–218. Hermes Penton Science, 2002.

[GS05] T. A. Gardiner and J. M. Stone. An unsplit Godunov method for ideal MHD via constrained transport. *Journal of Computational Physics*, 205(2):509–539, 2005.

[GS08] T. A. Gardiner and J. M. Stone. An unsplit Godunov method for ideal MHD via constrained transport in three dimensions. *Journal of Computational Physics*, 227(8):4123–4141, 2008.

[Gur04] K. F. Gurski. An HLLC-Type Approximate Riemann Solver for Ideal Magnetohydrodynamics. *SIAM Journal on Scientific Computing*, 25(6):2165–2187, 2004.

[GV99] H. Guillard and C. Viozat. On the behavior of upwind schemes in the low Mach limit. *Computers & Fluids*, 28:63–86, 1999.

[Har83] A. Harten. On the symmetric form of systems of conservation laws with entropy. *Journal of Computational Physics*, 49(1):151–164, 1983.

[Har89] A. Harten. ENO schemes with subcell resolution. *Journal of Computational Physics*, 83(1):148–184, 1989.

[HEOC87] A. Harten, B. Engquist, S. Osher, and S. R. Chakravarthy. Uniformly High Order Accurate Essentially Non-oscillatory Schemes III. *Journal of Computational Physics*, 71(2):231–303, 1987.

[HH83] A. Harten and J. M. Hyman. Self adjusting grid methods for one-dimensional hyperbolic conservation laws. *Journal of Computational Physics*, 50(2):235–269, 1983.

[HHE+21] L. Horst, R. Hirschi, P. V. F. Edelmann, R. Andrássy, and F. K. Röpke. Multidimensional low-Mach number time-implicit hydrodynamic simulations of convective helium shell burning in a massive star. *Astronomy & Astrophysics*, 653:A55, 2021.

[HLLM98] A. Harten, P. D. Lax, C. D. Levermore, and W. J. Morokoff. Convex Entropies and Hyperbolicity for General Euler Equations. *SIAM Journal on Numerical Analysis*, 35(6):2117–2127, 1998.

[HLvL83] A. Harten, P. D. Lax, and B. van Leer. On Upstream Differencing and Godunov-Type Schemes for Hyperbolic Conservation Laws. *SIAM Review*, 25(1):35–61, 1983.

[HOEC86] A. Harten, S. Osher, B. Engquist, and S. R. Chakravarthy. Some results on uniformly high-order accurate essentially nonoscillatory schemes. *Applied Numerical Mathematics*, 2(3-5):347–377, 1986.

[HR15] H. Holden and N. H. Risebro. *Front Tracking for Hyperbolic Conservation Laws*. Applied Mathematical Sciences. Springer, 2nd edition, 2015.

[HS96] M. E. Hosea and L. F. Shampine. Analysis and implementation of TR-BDF2. *Applied Numerical Mathematics*, 20(1-2):21–37, 1996.

[HXX22] G. Huang, Y. Xing, and T. Xiong. High order asymptotic preserving well-balanced finite difference WENO schemes for all Mach full Euler equations with gravity. 2022. arXiv:2211.16673.

[Jan00] P. Janhunen. A Positive Conservative Method for Magnetohydrodynamics Based on HLL and Roe Methods. *Journal of Computational Physics*, 160(2):649–661, 2000.

[JKT18] J. B. Jørgensen, M. R. Kristensen, and P. G. Thomsen. A Family of ESDIRK Integration Methods. 2018. arXiv:1803.01613.

[JS96] G.-S. Jiang and C.-W. Shu. Efficient Implementation of Weighted ENO Schemes. *Journal of Computational Physics*, 126(1):202–228, 1996.

[JST81] A. Jameson, W. Schmidt, and E. Turkel. Numerical Solution of the Euler Equations by Finite Volume Methods Using Runge-Kutta Time-Stepping Schemes. *AIAA 14th Fluid and Plasma Dynamic Conference*, 1981.

[JX95] S. Jin and Z. Xin. The relaxation schemes for systems of conservation laws in arbitrary space dimensions. *Communications on Pure and Applied Mathematics*, 48(3):235–276, 1995.

[Kle95] R. Klein. Semi-implicit extension of a Godunov-type scheme based on low Mach number asymptotics I: One-dimensional flow. *Journal of Computational Physics*, 121(2):213–237, 1995.

[KM82] S. Klainermann and A. Majda. Compressible and incompressible fluids. *Communications on Pure and Applied Mathematics*, 35(5):629–651, 1982.

[KM14] R. Käpelli and S. Mishra. Well-balanced schemes for the Euler equations with gravitation. *Journal Of Computational Physics*, 259:199–219, 2014.

[KM16] R. Käpelli 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. *Astronomy & Astrophysics*, 587:A94, 2016.

[KM17] F. Kupka and H. J. Muthsam. Modelling of stellar convection. *Living Reviews in Computational Astrophysics*, 3(1):1, 2017.

[KPS19] C. Klingenberg, G. Puppo, and M. Semplice. Arbitrary Order Finite Volume Well-Balanced Schemes for the Euler Equations with Gravity. *SIAM Journal on Scientific Computing*, 41(2):A695–A721, 2019.

[KWW13] R. Kippenhahn, A. Weigert, and A. Weiss. *Stellar Structure and Evolution*. Springer, 2013.

[LAB+23] G. Leidi, R. Andrássy, W. Barsukow, J. Higl, P. V. F. Edelmann, and F. K. Röpke:. Performance of high-order Godunov-type methods in simulations of astrophysical low Mach number flows. 2023.

[LBA+22] G. Leidi, C. Birke, R. Andrássy, J. Higl, P. V. F. Edelmann, G. Wiest, C. Klingenberg, and F. K. Röpke. A finite-volume scheme for modeling compressible magnetohydrodynamic flows at low Mach numbers in stellar interiors. *Astronomy & Astrophysics*, 668:A143, 2022.

[LeF02] P. G. LeFloch. *Hyperbolic Systems of Conservation Laws: The Theory of Classical and Nonclassical Shock Waves*. Lectures in Mathematics. ETH Zürich. Birkhäuser Verlag, 2002.

[LeV92] R. J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhäuser Basel, 2nd edition, 1992.

[LeV98] R. J. LeVeque. *Nonlinear Conservation Laws and Finite Volume Methods*, pages 1–159. Springer, Berlin Heidelberg, 1998.

[LeV02] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002.

[Li05] S. Li. An HLLC Riemann solver for magneto-hydrodynamics. *Journal of Computational Physics*, 203(1):344–357, 2005.

[Lio06] M.-S. Liou. A sequel to AUSM, Part II: AUSM$^+$-up for all speeds. *Journal of Computational Physics*, 214(1):137–170, 2006.

[LL91] K. Lerbinger and J. F. Luciani. A New Semi-implicit Method for MHD Computations. *Journal of Computational Physics*, 97(2):444–459, 1991.

[LMW02] H. P. Langtangen, K.-A. Mardal, and R. Winther. Numerical methods for incompressible viscous flow. *Advances in Water Resources*, 25(8):1125–1146, 2002.

[LOC94] X.-D. Liu, S. Osher, and T. Chan. Weighted Essentially Non-oscillatory Schemes. *Journal of Computational Physics*, 115(1):200–212, 1994.

[LPR99] D. Levy, G. Puppo, and G. Russo. Central WENO schemes for hyperbolic systems of conservation laws. *ESAIM M2AN. Mathematical Modelling and Numerical Analysis*, 33(3):547–571, 1999.

[LPR00] D. Levy, G. Puppo, and G. Russo. A third order central WENO scheme for 2D conservation laws. *Applied Numerical Mathematics*, 33(1):415–421, 2000.

[LT98] D. Levy and E. Tadmor. From Semidiscrete to Fully Discrete: Stability of Runge-Kutta Schemes by The Energy Method. *SIAM Review*, 40(1):40–73, 1998.

[LW60] P. D. Lax and B. Wendroff. Systems of conservation laws. *Communications on Pure and Applied Mathematics*, 13(2):217–237, 1960.

[LZ04] P. Londrillo and L. Del Zanna. On the divergence-free condition in Godunov-type schemes for ideal magnetohydrodynamics: the upwind constrained transport method. *Journal of Computational Physics*, 195(1):17–48, 2004.

[MB88] W. H. Matthaeus and M. R. Brown. Nearly incompressible magnetohydrodynamics at low Mach number. *Physics of Fluids*, 31(12):3634–3644, 1988.

[Mes99] L. Mestel. *Stellar magnetism*. Oxford University Press, 1999.

[Mic13] F. Miczek. *Simulation of low Mach number astrophysical flows*. Dissertation, Technische Universität München, 2013.

[MK05] T. Miyoshi and K. Kusano. A multi-state HLL approximate Riemann solver for ideal magnetohydrodynamics. *Journal of Computational Physics*, 208(1):315–344, 2005.

[MM21] T. Minoshima and T. Miyoshi. A low-dissipation HLLD approximate Riemann solver for a very wide range of Mach numbers. *Journal of Computational Physics*, 446:110639, 2021.

[MRE15] F. Miczek, F. K. Röpke, and P. V. F. Edelmann. New numerical solver for flows at various Mach numbers. *Astronomy & Astrophysics*, 576:A50, 2015.

[MZ21] A. Mignone and L. Del Zanna. Systematic construction of upwind constrained transport schemes for MHD. *Journal of Computational Physics*, 424:109748, 2021.

[Mü20] B. Müller. Hydrodynamics of core-collapse supernovae and their progenitors. *Living Reviews in Computational Astrophysics*, 6:3, 2020.

[NBA+14] S. Noelle, G. Bispen, K. R. Arun, M. Lukáčová-Medviďová, and C.-D. Munz. A Weakly Asymptotic Preserving Low Mach Number Scheme for the Euler Equations of Gas Dynamics. *SIAM Journal on Scientific Computing*, 36(6):B989–B1024, 2014.

[OP62] Y. Ogura and N. A. Phillips. Scale Analysis of Deep and Shallow Convection in the Atmosphere. *Journal of the Atmospheric Sciences*, 19(2):173–179, 1962.

[Osh84] S. Osher. Riemann Solvers, the Entropy Condition, and Difference Approximations. *SIAM Journal on Numerical Analysis*, 21(2):217–235, 1984.

[OT79] S. A. Orszag and C.-M. Tang. Small-scale structure of two-dimensional magnetohydrodynamic turbulence. *Journal of Fluid Mechanics*, 90(1):129–1143, 1979.

[Par06] C. Parés. Numerical methods for nonconservative hyperbolic systems: a theoretical framework. *SIAM Journal on Numerical Analysis*, 44(1):300–321, 2006.

[Pow97] K. G. Powell. *An Approximate Riemann Solver for Magnetohydrodynamics (That Works in More than One Dimension)*, pages 570–583. Springer, Berlin Heidelberg, 1997.

[PQV01] M. Pelanti, L. Quartapelle, and L. Vigevano. *Low Dissipation Entropy Fix for Positivity Preserving Roe's Scheme*, pages 685–690. Springer US, New York, NY, 2001.

[PR05] L. Pareschi and G. Russo. Implicit-Explicit Runge-Kutta Schemes and Applications to Hyperbolic Systems with Relaxation. *Journal of Scientific Computing*, 25:129–155, 2005.

[PRL+99] K. G. Powell, P. L. Roe, T. J. Linde, T. I. Gombosi, and D. L. De Zeeuw. A Solution-Adaptive Upwind Scheme for Ideal Magnetohydrodynamics. *Journal of Computational Physics*, 154(2):284–309, 1999.

[Rie11] F. Rieper. A low-Mach number fix for Roe's approximate Riemann solver. *Journal of Computational Physics*, 230(13):5263–5287, 2011.

[Roe81] P. L. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics*, 43(2):357–372, 1981.

[Roe85] P. L. Roe. Some contributions to the modeling of discontinuous flows. *Lecture Notes in Applied Mathematics*, 22:163–193, 1985.

[Roe86] P. L. Roe. Characteristic-Based Schemes for the Euler Equations. *Annual Review of Fluid Mechanics*, 18(1):337–365, 1986.

[Roe92] P. L. Roe. Sonic Flux Formulae. *SIAM Journal on Scientific and Statistical Computing*, 13(2):611–630, 1992.

[Rus62] V. V. Rusanov. The calculation of the interaction of non-stationary shock waves and obstacles. *USSR Computational Mathematics and Mathematical Physics*, 1(2):304–320, 1962.

[SF93] G. L. G. Sleijpen and D. R. Fokkema. BiCGstab(l) for linear equations involving unsymmetric matrices with complex spectrum. *ETNA. Electronic Transactions on Numerical Analysis*, 1:11–32, 1993.

[Shu03] C.-W. Shu. High-order Finite Difference and Finite Volume WENO Schemes and Discontinuous Galerkin Methods for CFD. *International Journal of Computational Fluid Dynamics*, 17(2):107–118, 2003.

[SO88] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77(2):439–471, 1988.

[SS86] Y. Saad and M. Schultz. GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems. *SIAM Journal on Scientific and Statistical Computing*, 7(3):856–869, 1986.

[Str68] G. Strang. On the Construction and Comparison of Difference Schemes. *SIAM Journal on Numerical Analysis*, 5(3):506–517, 1968.

[Sul90] I. Suliciu. On modelling phase transitions by means of rate-type constitutive equations. Shock wave structure. *International Journal of Engineering Science*, 28(8):829–841, 1990.

[Sul92] I. Suliciu. Some stability-instability problems in phase transitions modelled by piecewise linear elastic or viscoelastic constitutive equations. *International Journal of Engineering Science*, 30(4):483–494, 1992.

[Tad84] E. Tadmor. Numerical Viscosity and the Entropy Condition for Conservative Difference Schemes. *Mathematics of Computation*, 43(168):369–381, 1984.

[Tok02] A. T. Tokunaga. *Infrared Astronomy*, pages 143–167. Springer, New York, NY, 2002.

[Tor09] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*. Springer, 3rd edition, 2009.

[Tor19] E. F. Toro. The HLLC Riemann solver. *Shock Waves*, 29(8):1065–1082, 2019.

[Tót00] G. Tóth. The $\nabla\cdot \mathbf{B}$=0 Constraint in Shock-Capturing Magnetohydrodynamics Codes. *Journal of Computational Physics*, 161(2):605–652, 2000.

[TPK20] A. Thomann, G. Puppo, and C. Klingenberg. An all speed second order well-balanced IMEX relaxation scheme for the Euler equations with gravity. *Journal of Computational Physics*, 420:109723, 2020.

[TSS94] E. F. Toro, M. Spruce, and W. Speares. Restoration of the contact surface in the HLL-Riemann solver. *Shock Waves*, 4(1):25–34, 1994.

[Tur87] E. Turkel. Preconditioned methods for solving the incompressible and low speed compressible equations. *Journal of Computational Physics*, 72(2):277–298, 1987.

[TZK19] 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. *International Journal for Numerical Methods in Fluids*, 89(11):465–482, 2019.

[TZPK20] A. Thomann, M. Zenk, G. Puppo, and C. Klingenberg. An All Speed Second Order IMEX Relaxation Scheme for the Euler Equations. *Communications in Computational Physics*, 28(2):591–620, 2020.

[VC19] D. Varma and P. Chandrashekar. A second-order, discretely well-balanced finite volume scheme for Euler equations with gravity. *Computers & Fluids*, 181:292–313, 2019.

[vL77] B. van Leer. Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection. *Journal of Computational Physics*, 23(3):276–299, 1977.

[vL84] B. van Leer. On the Relation Between the Upwind-Differencing Schemes of Godunov, Engquist–Osher and Roe. *SIAM Journal on Scientific and Statistical Computing*, 5(1):1–20, 1984.

[VLW04] P. Váchal, R. Liska, and B. Wendroff. Fully Two-dimensional HLLEC Riemann Solver and Associated Difference Schemes. In M. Feistauer, V. Dolejší, P. Knobloch, and K. Najzar, editors, *Numerical Mathematics and Advanced Applications*, pages 815–824, Springer, Berlin Heidelberg, 2004.

[Wal85] J. R. Waldram. *The Theory of Thermodynamics*. Cambridge University Press, 1985.

[WFK11] K. Waagan, C. Federrath, and C. Klingenberg. A robust numerical scheme for highly compressible magnetohydrodynamics: Nonlinear stability, implementation and tests. *Journal of Computational Physics*, 230(9):3331–3351, 2011.

[XC13] X.-s. Li and C.-w. Gu. Mechanism of Roe-type schemes for all-speed flows and its application. *Computers & Fluids*, 86:56–70, 2013.

[XS13] Y. Xing and C.-W. Shu. High Order Well-Balanced WENO Scheme for the Gas Dynamics Equations Under Gravitational Fields. *Journal of Scientific Computing*, 54(2):645–662, 2013.