ASYMPTOTIC PRESERVING NUMERICAL SCHEMES FOR TRANSPORT AND FLUID EQUATIONS

By

Jeffrey Robert Haack

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
(Mathematics)

at the

UNIVERSITY OF WISCONSIN – MADISON
2010
Abstract

In certain asymptotic regimes many physical models can be accurately approximated by another model. However, the two models often have very different properties and require different strategies for numerical simulation. Asymptotic preserving (AP) schemes are schemes that can capture at the discrete level the transition from one model to the other, while only resolving time and space scales relevant to the regime at interest. One example of a model with such an asymptotic limit is kinetic models of particle transport through a background material. As the mean free path between particle collisions with the medium becomes small, over long time scales the predominant macroscopic behavior of the system is diffusive. However, when simulating multi-scale problems with regions of both high and low collisionality, such as multiple materials, a standard kinetic solver will require expensive numerical resolution of the mean free path in the diffusive regime, while the relevant scales of the underlying diffusive dynamics are independent of the mean free path. Another example is the low Mach number limit of the compressible Euler and Navier-Stokes equations. As the characteristic Mach number of the system becomes small solutions of the compressible equations can be approximated by solutions to the incompressible equations. However, in the low Mach number limit compressible solvers require resolution of the acoustic wave time scales of the system, despite the fact that these waves do not exist in incompressible regimes.

In this dissertation I investigated some problems arising in asymptotic preserving schemes for a linear model of diffusive transport. What we found was that with certain initial data, there is too little diffusion which causes unphysical oscillations to arise. We
develop a program for correcting this diffusion that is still consistent with the asymptotic limit. We also develop an asymptotic preserving scheme for the isentropic Euler and Navier-Stokes equations that is suitable for both compressible and incompressible regimes. We present numerical results in compressible regimes that demonstrate its shock capturing ability and results in the incompressible regime that demonstrate that it captures the asymptotics without requiring resolution of the Mach number.
Acknowledgements

I am very lucky to have so many people who have helped me reach this point in my academic career. I cannot thank my advisor, Shi Jin, enough for all of his insight, guidance, patience, and support, which have been invaluable. Further thanks go to our research collaborator Jian-Guo Liu, whose help and knowledge of fluid dynamics have been crucial to this project. I thank Cory Hauck and Rob Lowrie for their insight on the fluid dynamics work and especially for sparking my interest in kinetic transport problems. I would like to thank my doctoral committee - Fabian Waleffe, Paul Milewski, Mikhail Feldman, and Mike Graham. Special thanks are due to Gloria Mari-Beffa, whose advice and love of mathematics inspired me to become a mathematician instead of just another programmer in the first place.

I was supported by many grants and programs during my graduate career. I had the pleasure of receiving multiple NSF VIGRE fellowships while I was here, and I would also like to thank Paul Milewski for research support granted early in my career. I did much of the work on transport problems while a summer student at the Center for Nonlinear Studies at Los Alamos National Laboratory in 2008 and 2009. My experience at the lab was inspirational, and I’d like to thank the vibrant community there — especially Cory and Rob for arranging my placement there and Ryan McClarren for his encyclopedic knowledge of transport problems.

Finally, and most importantly, I would like to thank my wife, Jennette Cross, whose love and patience through this entire process have been beyond measure.
## List of Figures

1. Standard upwind schemes with periodic boundary conditions and initial condition (2.6). In all four figures, the triangles represent the solution with $\Delta x = 0.04$, the circles represent the solution with $\Delta x = 0.01$, and the solid line is a highly resolved reference solution that uses $\Delta x = 10^{-4}$. The timestep is $\Delta t = 0.1\sqrt{3\varepsilon\Delta x}$.

2. The splitting scheme, with periodic boundary conditions and initial condition (2.6). In all four figures, the triangles represent the solution with $\Delta x = 0.04$, the circles represent the solution with $\Delta x = 0.01$, and the solid line is a reference solution with $\Delta x = 10^{-4}$. The timesteps are chosen using the CFL condition (2.14) with $C = 0.1$.

3. First-order splitting scheme results. Left: discontinuous cross-section given by (2.16) and initial condition given by (2.6), $\varepsilon = 0.1$, and $t = 0.075$. Right: Constant cross-section of $\sigma = 0.02$, initial condition given by (2.6), $\varepsilon = 0.1$, and $t = 0.1$. In both figures, circles represent the numerical solution with $\Delta x = 0.01$ and time step determined by (2.14) with $C = 0.1$. The solid line is a reference solution computed with a second-order upwind scheme with $\Delta x = 0.01$ and $\Delta t = 0.1\sqrt{3\varepsilon\Delta x}$. 

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Standard upwind schemes</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>The splitting scheme</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>First-order splitting scheme results</td>
<td>27</td>
</tr>
</tbody>
</table>
4 On the left, a recalculation of Figure 2(c), with the black-red diffusion discretization replaced with a standard discretization, as described in (2.21). On the right, the black-red discretization is kept, but the smooth initial condition (2.22) is used. In both cases, triangles represent the solution with $\Delta x = 0.04$ and the solid line is a highly resolved reference solution.

5 Repeat of calculation from Figure 3(a). Left: First-order splitting scheme with dispersive term (2.20) removed. Right: First-order splitting scheme with black-red diffusion term modified according to (2.21).

6 Left: $P_1$ solution at time $t = 0.1$ with constant cross-section $\sigma = 0.02$. Right: Solution at time $t = 0.075$ with the discontinuous cross section given by (2.16). In both figures, the solid line is the reference solution and the circles are the solution using the splitting method with the corrected numerical diffusion. The timestep is chosen using the CFL conditions (2.14) with $C = 0.1$.

7 1D Riemann problem with initial data (3.31). $\varepsilon = 1$, $\gamma = 1.4$, $\alpha = 0.1$ and 0.5. The solid line is the true solution.

8 1D Riemann problem with initial data (3.31). $\varepsilon = 1$, $\gamma = 1.4$, $\alpha = 0.5$ and 0.75. The solid line is the true solution.

9 Solution to initial data (3.32) at $T = 0.001$ with $\varepsilon = 1$ and $\alpha = 0.5$. The solid line is the true solution.

10 Density plots of 2D Riemann problem with initial data (3.33). $\varepsilon = 1$, $\gamma = 1.4$, $\alpha = 0.01$ and 0.5.
Density plots of 2D Riemann problem with initial data (3.33). \( \varepsilon = 1, \gamma = 1.4, \alpha = 0.75 \) and \( 0.99. \)

Results for periodic flow test case (3.34). Left column: \( \varepsilon = 0.1. \) Right column: \( \varepsilon = 10^{-4}. \) Top row: stream function of solution with \( \Delta x = 1/16. \) Middle row: error in \( u. \) Bottom row: error in \( v. \)

Results for Vortex in a box flow (3.35) at \( T = 0.125. \) Left column: \( \varepsilon = 0.1. \) Right column: \( \varepsilon = 10^{-4}. \) Top row: stream function of solution with \( \Delta x = 1/16. \) Middle row: error in \( u. \) Bottom row: error in \( v. \)

Left: Contour plot of streamfunction for backwards step flow with \( \varepsilon = 0.01, \Delta x = 1/N = 1/16 = 0.0625, \Delta t = 9.765 \times 10^{-4}, \text{Re} = 100, \gamma = 1.4, \alpha = \varepsilon^2 \) and \( \theta = 1 \) at \( T = 20. \) Right: Contour plot of streamfunction from [51]. The AP scheme captures the reattachment point of the circulation region.
## Contents

Abstract i

Acknowledgements iii

1 Introduction 1

1.1 Introduction to multi-scale diffusive transport 2

1.2 The $P_N$ system 3

1.3 Introduction to the low Mach number limit of the Isentropic Euler and Navier-Stokes equations 8

1.3.1 Low Mach number limit of the isentropic Euler and Navier-Stokes equations 11

1.3.2 The low Mach number limit 13

1.3.3 Numerical difficulties in the low Mach number limit 15

1.4 Chapter Organization 17

2 Numerical Diffusion for AP schemes in the diffusion limit 18

2.1 Upwind discretization 18

2.2 Splitting method 21

2.3 A closer look at operator splitting 25

3 The all-speed asymptotic-preserving method for the isentropic Navier-Stokes equations 33

3.1 All-speed asymptotic-preserving schemes 33
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>Hyperbolic splitting</td>
<td>34</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Time discretization of the split systems</td>
<td>35</td>
</tr>
<tr>
<td>3.3</td>
<td>$L^2$ stability</td>
<td>37</td>
</tr>
<tr>
<td>3.4</td>
<td>Spatial discretization of the split systems</td>
<td>38</td>
</tr>
<tr>
<td>3.5</td>
<td>Boundary and initial conditions</td>
<td>39</td>
</tr>
<tr>
<td>3.6</td>
<td>The discrete low Mach number limit</td>
<td>40</td>
</tr>
<tr>
<td>3.7</td>
<td>The fast solver for $\rho^{n+1}$</td>
<td>46</td>
</tr>
<tr>
<td>3.7.1</td>
<td>Computational cost of the method</td>
<td>47</td>
</tr>
<tr>
<td>3.7.2</td>
<td>The CFL condition</td>
<td>48</td>
</tr>
<tr>
<td>3.8</td>
<td>Numerical results</td>
<td>48</td>
</tr>
<tr>
<td>3.8.1</td>
<td>Compressible flow examples ($\varepsilon = O(1)$)</td>
<td>48</td>
</tr>
<tr>
<td>3.8.2</td>
<td>Low Mach number limit examples ($\varepsilon &lt;&lt; 1$)</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>Concluding remarks and future work</td>
<td>59</td>
</tr>
<tr>
<td>4.1</td>
<td>Diffusive transport</td>
<td>59</td>
</tr>
<tr>
<td>4.2</td>
<td>Low Mach number limit</td>
<td>60</td>
</tr>
<tr>
<td>A</td>
<td>Appendix</td>
<td>62</td>
</tr>
<tr>
<td>A.1</td>
<td>A few basic tools</td>
<td>62</td>
</tr>
<tr>
<td>A.2</td>
<td>The schemes</td>
<td>64</td>
</tr>
</tbody>
</table>

Bibliography 68
Chapter 1

Introduction

In certain asymptotic regimes many physical models can be accurately approximated by another model. However, the two models often have very different properties and require different strategies for numerical simulation. Asymptotic preserving (AP) schemes are schemes that can capture at the discrete level the transition from one model to the other, while only resolving time and space scales relevant to the regime at interest. Specifically, if the time and spatial steps \( \Delta t \) and \( \Delta x \) are kept fixed, as the small scale parameter goes to zero the method automatically transforms to a stable discretization of the limiting model. The asymptotic-preserving approach was first introduced in the context of linear transport in diffusive regimes [33, 20, 44], and was later extended to discrete-velocity kinetic models [37], radiative transfer [38], kinetic semiconductor equations [35], and fluid limits of the Boltzmann equation [32], to name a few. A complete review of this methodology can be found in [36].

In this dissertation, I will investigate asymptotic-preserving approaches for the computation of diffusion limit of a kinetic transport model. I will also develop an asymptotic preserving scheme for the isentropic Euler and Navier Stokes equations that is suitable for both the compressible and incompressible regimes. A detailed introduction of those topics is introduced in the following sections.
1.1 Introduction to multi-scale diffusive transport

In kinetic models of particle transport, diffusive relaxation is a common phenomena that occurs when the mean free path between particle collisions with a material medium is small when compared to the macroscopic scales of interest. In such cases, particles undergo frequent collisions with the material so that, over long time scales, the predominant macroscopic behavior of the system is diffusive [45, 49, 63].

Even though particle transport in collisional regimes can be accurately approximated with a diffusive model, a kinetic model is still needed for regimes in which collisions are less frequent. Thus, when simulating multi-scale problems with regions of both high and low collisionality, such as multiple materials, a reasonable approach is to use a kinetic model everywhere, since it is valid for all regimes. The challenge of such an approach is to develop numerical methods that mimic the asymptotics of the diffusion limit at the discrete level without resolving temporal and spatial scales associated with the mean free path, which tends to zero in the diffusion limit. In particular, a numerical scheme for a kinetic model should be consistent with the diffusion equation in the diffusion limit. This is the so-called asymptotic preserving property cited above [31].

Conventional Godunov-type solvers for kinetic models, based on spatial reconstructions and upwinding, are typically not asymptotic preserving. Like almost all hyperbolic solvers, these methods use numerical dissipation to maintain stability around discontinuities, and in most cases, the dissipation increases as the system approaches the diffusion limit. Eventually, the numerical dissipation will actually dominate the physical diffusion in the system. Consequently, one may generate results which appear to resolve the solution profile, but are far from accurate.
1.2 The $P_N$ system

The $P_N$ system is an type of \textit{moment method} used to approximate the solution of a transport equation of the form

$$\partial_t F + v \cdot \nabla_x F = C(F),$$  \hspace{1cm} (1.1)

where $F(x, v, t)$ is the positive particle distribution function in phase space that describes the dynamics of particles at each point $x$ traveling with velocity $v$ at time $t$. The left hand side of the equation describes the evolution of $F$ along these trajectories, and the right hand side describes the effects of interactions between the particles with each other and/or the background through scattering and other similar processes. Typically these problems are very difficult to solve due to the fact that the collision term $C(F)$ is frequently an integral operator.

As the physical observables of these systems are generally the moments themselves, one straightforward approach for solving these types of equations is to integrate a set of test functions $m(v)$ against (1.1) to obtain the system (where $\langle \cdot \rangle$ denotes integration over $v$)

$$\partial_t \langle mF \rangle + \partial_x \frac{1}{\varepsilon} \langle \mu mF \rangle = -\frac{\sigma}{\varepsilon^2} \langle (mF - \frac{1}{2} \langle mF \rangle) \rangle.$$  \hspace{1cm} (1.2)

The $P_N$ approximation [49, 63] consists of assuming (in one dimensional slab geometry)

$$F(x, \mu, t) = \sum_{n=0}^{\infty} \frac{2n + 1}{2} u_n(x, t) p_n(\mu),$$

where $\mu \in [-1, 1]$ is the cosine of the angle between $v$ and the coordinate axis, and $p_n(v)$ are the Legendre polynomials on $[-1, 1]$. The "$P$" stands for this polynomial basis, and "$N$" stands for the number of basis polynomials chosen. Inserting this into (1.2)
and again using the Legendre polynomials as the test functions, we obtain the linear, constant coefficient hyperbolic system

\[
\partial_t u_n + \frac{1}{\varepsilon} \left( \frac{n}{2n+1} \partial_x u_{n-1} + \frac{n+1}{2n+1} \partial_x u_{n+1} \right) = -\frac{\sigma}{\varepsilon^2} u_n (1 - \delta_{0,n}),
\]

where

\[
\delta_{i,j} = \begin{cases} 
1 & i = j \\
0 & \text{otherwise}
\end{cases}
\]

The coefficients in the flux terms are a consequence of the Legendre polynomial identity \( \mu P_n = \frac{n}{2n+1} P_{n-1} + \frac{n+1}{2n+1} P_{n+1} \). The system is closed by setting \( u_n = 0 \) for \( n > N \) (essentially, truncating the expansion in \( F \)). The Legendre polynomials are attractive because of their orthogonality, but in principle any basis set of functions could be chosen.

In higher dimensions, \( F \) is expanded in terms of spherical harmonics (and thus \( P_N \) is sometimes referred to as the spherical harmonic method).

One application of these systems, and a major driving force for the research behind them, lies in neutron transport [48, 57], which models the flow of neutrons in e.g. nuclear reactors. For simplicity, we will assume that the neutrons move with a single fixed speed (and hence energy), and interact with the background medium through an isotropic scattering process. Furthermore, we will also assume for ease of exposition a one-dimensional ‘slab’ geometry. We take \( N = 1 \), and these equations form a simple \( 2 \times 2 \) linear hyperbolic system which serves as a rough approximation of more complicated kinetic descriptions of particle transport through a material medium. In practice, the \( P_1 \) system serves as a prototype for more complicated moment systems and transport models, such as multi-group equations (where the velocity is not constant) or non-isotropic scattering, absorption, etc.
In one space dimension, and with a diffusive scaling, the nondimensionalized $P_1$ equations are

\begin{align}
\partial_t \rho + \partial_x m &= 0, \hspace{1cm} (1.3a) \\
\partial_t m + \frac{1}{3\varepsilon^2} \partial_x \rho &= -\frac{\sigma}{\varepsilon^2} m. \hspace{1cm} (1.3b)
\end{align}

The variables $\rho$ and $m$ are the first two moments of the underlying kinetic distribution function $F$. Physically, $\rho$ is the concentration of particles and $\varepsilon m$ is the bulk momentum, where the parameter $\varepsilon$ is the magnitude of the macroscopic reference velocity relative to the microscopic particle velocity (i.e. the scaled mean free path). Finally, $\sigma = \sigma(x)$ is the non-dimensional scattering cross-section. Given a macroscopic length scale $L$, the physical cross-section (the inverse of the mean free path) is $\sigma/\varepsilon L$.

The mathematical aspects of the $P_1$ system and its nonlinear variants—including the diffusive asymptotic behavior—have been studied both theoretically [50, 43, 56, 53, 54, 58] and numerically [21, 52, 37]. Studies of related models can be found in the context of radiation and neutron transport [57, 48, 4, 22, 55] and also in drift-diffusion systems such as charge transport in semiconductors [29, 38, 60] and chemotaxis [17, 16].

Upon diagonalization, the $P_1$ equations take the form of a Goldstein-Taylor model [19, 66] with wave speeds $\pm (\sqrt{3}\varepsilon)^{-1}$. Indeed, if we set $\phi^\pm = \rho \pm \sqrt{3}\varepsilon m$, then

\begin{align}
\partial_t \phi^+ + \frac{1}{\varepsilon \sqrt{3}} \partial_x \phi^+ &= -\frac{\sigma}{2\varepsilon^2} (\phi^+ - \phi^-), \hspace{1cm} (1.4a) \\
\partial_t \phi^- - \frac{1}{\varepsilon \sqrt{3}} \partial_x \phi^- &= -\frac{\sigma}{2\varepsilon^2} (\phi^- - \phi^+). \hspace{1cm} (1.4b)
\end{align}

Meanwhile, the diffusive character of (1.3) is evident upon formally balancing powers of
ε in (1.3b) to specify a closure for (1.3a) that is accurate up to \( O(\varepsilon^2) \):

\[
m = -\frac{1}{3\sigma} \partial_x \rho, \tag{1.5a}
\]

\[
\partial_t \rho = \partial_x \left( \frac{1}{3\sigma} \partial_x \rho \right). \tag{1.5b}
\]

Several methods have been developed for the \( P_1 \) equations and related systems which are AP in the diffusive limit. Many different approaches have been made to tackle this problem — we mention in passing the well-balanced approach taken in [21, 22, 4] and the discontinuous Galerkin formulation used in [44, 52, 55]. However, our focus here is the AP operator splitting methodology that has been developed over the last ten years, specifically the work on a modified version of the \( P_1 \) system that was first introduced by Jin in [31] and later generalized by Hauck and Lowrie in [30]. Unlike naive splitting approaches that are based solely on simplifying the numerical implementation of a model, AP splitting is done in such a way as to preserve asymptotic balances. In the context of diffusive relaxation, this means separating a given kinetic model into two component systems with the following properties:

1. One component contains fast dynamics related to the mean free path. It is numerically stiff and therefore updated in time with an implicit method. Although implicit, the update is much simpler than an implicit update of the full model and can often be implemented in an explicit fashion.

2. The second component contains slower, macroscopic dynamics and can be updated with a standard explicit method.

3. As an isolated system, each component is well-posed in some sense.
The evolution of the complete system is performed by alternately updating each of the two components. Such an approach is first order in time, but can be improved with more complex time integrators, such as implicit-explicit IMEX schemes [2, 61, 62].

Although successfully implemented in a number of applications, the operator splitting methodology has some subtle issues. One of these is the onset on spurious oscillations which are stable, but lead to unphysical results. Often, such oscillations arise in so-called transition regimes, where collisions play an important role in the dynamics of the model, but are not frequent enough to validate the diffusion limit. Moment equations for simulating transport are most useful in these regimes, either as stand-alone models or as preconditioners for more complex models.

The cause of spurious oscillations in a scheme based on operator splitting is not easily deduced from inspection of the two component systems individually. Rather, one must look at the composition of the two components as a single method. Previously, it has been conjectured that oscillations and instabilities might be attributed to numerical dispersion terms [31] or to numerical diffusion terms defined on staggered grids [30], both of which are numerical artifacts of the splitting that appear in the full scheme. We confirm here that binary oscillations can occur in low-resolution schemes as a result of the staggered grid. However, we also determine that oscillations of a dispersive nature can be attributed to too little numerical dissipation in the scheme, as opposed to standard upwind methods which are usually overly dissipative.
1.3 Introduction to the low Mach number limit of the Isentropic Euler and Navier-Stokes equations

We are interested in the efficient numerical simulation of unsteady compressible flows in all ranges of Mach numbers. These flows arise in many physical applications, including atmospheric modeling, magnetohydrodynamics, and combustion. When the Mach number is of order one, modern shock capturing methods provide high resolution numerical approximations to shocks and other complex flow structures. However, when the Mach number is small, near the so-called incompressible regime, there is a wide gap between the flow speed and the acoustic waves in the system, which are often unimportant in the incompressible regime. In the incompressible regime, standard explicit shock-capturing methods require the time step to scale inversely with the maximum wave speed in the system for stability, which greatly overresolves the solution in time. Furthermore, these shock capturing methods will introduce numerical diffusions that scale with the inverse of the wave speeds around discontinuities, which requires overresolution in space in order to ensure that the numerical diffusion does not dominate the solution or physical viscosity for high Reynolds number flows.

Our goal is to develop all-speed flow simulators that work in all regimes of Mach number, including both compressible and incompressible regimes and their mixture. As a first step, we focus on the compressible isentropic Euler and Navier-Stokes equations of gas dynamics. It was shown by Klainerman and Majda [40] that solutions to these equations converge to solutions of the incompressible equations in the limit where the Mach
number goes to zero. The major difference between compressible and incompressible systems lies in the pressure term. In the compressible case, the pressure is determined by the equation of state of the system and plays an important role in the flux terms of the conservation law, and is the source of the acoustic waves in the system. However, in the limiting incompressible equations the pressure term acts as a Lagrange multiplier to enforce the incompressibility condition and is in fact an asymptotic perturbation of the physical pressure from the compressible equations.

The development of computational methods for nearly incompressible (small Mach number flows) has attracted great attention for many years. Much of the early literature in this area focused on preconditioning techniques for steady state problems. In fact, Chorin’s artificial compressibility approach [7] sought to avoid the difficulties of the pressure term in the incompressible equations entirely by solving a form of the compressible low Mach number system, which has much clearer boundary conditions. It was later recognized [68] that these ideas could be used to calculate steady states of incompressible flows. Later studies applied these ideas to compute solutions to low Mach number flow by introducing preconditioning matrices to symmetrize the system in terms of a set of non-conservative variables [1, 26]. However, these methods assume that the flow is already in the low Mach number regime and thus cannot accurately compute problems where the Mach number is of order unity. Guillard and Viozat [24] followed the asymptotic analysis of Klainerman and Majda [40] to show that the artificial numerical dissipation in upwind methods for the Euler equation are what causes the method to perform poorly in the low Mach number limit, and use the preconditioner of Turkel [68] to alter the dissipation terms in their method to capture the correct limit. Colella and Pao [9] used the Hodge decomposition to split the method into incompressible and
irrotational components and obtained a method that is applicable to the broader regime of flows with Mach number less than one. However, the accuracy breaks down at the higher Mach numbers and it does not capture the correct speeds of the acoustic waves in regimes where they become important. This work was later extended by Gatti-Bono and Colella in the context of atmospheric flows [18].

Other approaches have sought to develop all-speed methods, which are suited to both fully compressible and low Mach number regimes. Harlow and Amsden [28] sought to extend the staggered-mesh MAC scheme for incompressible flows [27] to compressible flows. Their method, the Implicit Continuous-fluid Eulerian (ICE) method, iteratively solves an implicit second-order wave equation to update the density and pressure terms. However, this scheme is not conservative and has difficulties in capturing strong shocks. Inspired by this work, Degond and Tang [14] split the stiff pressure term with a numerical parameter to derive a nonlinear elliptic equation for the density updates that acts similar to a classical incompressible projection on the system. Klein [41] presents a predictor-corrector type method based on pressure variables at each order in the asymptotic expansion of the pressure. Osher et al. [39] developed a method using a second order preconditioner algorithm that captured the correct shock speed and suppressed oscillations in multi-fluid systems. However, in low Mach regimes this method still requires temporal resolution of the Mach number for stability. In our case here, the limiting scheme, when the Mach number goes to zero, becomes a good incompressible solver similar to a second order projection method. By adequately splitting the compressible Euler/Navier-Stokes equations into a compressible, non-stiff nonlinear hyperbolic system and a stiff linear acoustic wave system which can be easily handled by a fast Fourier transform based Poission solver, our method allows the use of sound speed (essentially
the reciprocal of the Mach number) independent time and spatial steps.

In Chapter 3 we present a new numerical method for the solution of the isentropic Euler and Navier-Stokes equations that is valid for all Mach numbers (namely the all-speed property). It allows the use of standard conservative shock capturing methods that are necessary for the compressible regime, yet the method is shown to be Asymptotic Preserving (AP)[31] in the zero Mach number limit.

In [12], Degond, Jin, and Liu studied the time discretization of Asymptotic Preserving methods for several compressible flow problems by means of a Hodge-like decomposition. While our work was inspired by this research, we capture the low Mach number limit in a different fashion.

1.3.1 Low Mach number limit of the isentropic Euler and Navier-Stokes equations

The two-dimensional isentropic Navier-Stokes equations are given by (for Euler, remove the viscous term on the right hand side):

\[ \rho_t + \nabla \cdot (\rho u) = 0 \]

\[ (\rho u)_t + \nabla \cdot (\rho u \otimes u) + \nabla p(\rho) = \mu \Delta u \]

\[ p(\rho) = A \rho^\gamma \]

Here, \( \rho \) is the density of the fluid, \( \mathbf{m} = \rho \mathbf{u} \) is the momentum of the fluid, \( \mu \) is the dynamic viscosity of the fluid, and \( p(\rho) \) is the pressure. Typically air is composed of \( N_2 \) and \( O_2 \), which gives \( \gamma = 1.4 \), and we take \( A = 1 \) for simplicity. To obtain the Euler equations, set \( \mu = 0 \).
To describe the low Mach number (incompressible) limit, one scales the equations in the following manner. Let \( x_0, t_0, \rho_0, p_0, u_0 \) be a set of characteristic scales for the variables in the equation. The dimensionless variables are then given by \( \hat{x} = x/x_0, \hat{t} = t/t_0, \) etc. Inserting these into the equations (and dropping the hats), one obtains the nondimensionalized equations

\[
\frac{\rho_0}{t_0} \rho_t + \frac{\rho_0 u_0}{x_0} \nabla \cdot (\rho u) = 0
\]

\[
\frac{\rho_0 u_0}{t_0} (\rho u)_t + \frac{u_0^2 \rho_0}{x_0} \nabla \cdot (\rho u \otimes u) + \frac{p_0}{x_0} \nabla p = \frac{u_0 \mu}{x_0^2} \Delta u
\]

\[
p_0 p = \tilde{\rho}_0 \rho^\gamma.
\]

Using the fact that \( u_0 = x_0/t_0 \), one has

\[
\rho_t + \nabla \cdot (\rho u) = 0
\]

\[
(\rho u)_t + \nabla \cdot (\rho u \otimes u) + \frac{p_0}{\rho_0 u_0^2} \nabla p = \frac{1}{Re} \Delta u
\]

\[
\frac{p_0}{\rho_0^\gamma} = \rho^\gamma.
\]

Here \( Re = \frac{\rho_0 u_0 x_0}{\mu} \) is the dimensionless Reynolds number, which measures the ratio between the inertial and diffusive forces in the system. \( p_0 \) scales as \( \tilde{\rho}_0 \), as expected from the equation of state. What remains to determine is the term in front of the pressure. Since the speed of sound is given by \( c^2 = \gamma p/\rho \), one has \( c_0^2 \rho = \frac{\gamma p_0}{p_0} \), and thus one defines the dimensionless reference Mach number to be \( \varepsilon^2 = \frac{\gamma u_0^2}{c_0^2} = \frac{\rho_0 u_0^2}{p_0} \). The nondimensionalized equations then take the form

\[
\rho_t + \nabla \cdot (\rho u) = 0
\]

\[
(\rho u)_t + \nabla \cdot (\rho u \otimes u) + \frac{1}{\varepsilon^2} \nabla p = \frac{1}{Re} \Delta u
\]

\[
p = \rho^\gamma.
\]
One thing to note is that when \( \gamma = 2 \) and without the viscous term, this is just the shallow water system (replace the density \( \rho \) with the height \( h \)).

### 1.3.2 The low Mach number limit

To determine the asymptotic behavior as \( \varepsilon \to 0 \), one takes an asymptotic expansion of the variables as

\[
\rho = \rho^{(0)} + \varepsilon^2 \rho^{(2)} + \ldots
\]

for small \( \varepsilon \), and look at the balances within the equations. At \( O \left( \frac{1}{\varepsilon^2} \right) \), one has the balance

\[
\nabla p^{(0)} = 0 \quad \Rightarrow \quad p^{(0)}(x, t) = p^{(0)}(t) \quad \Rightarrow \quad \rho^{(0)} = \rho^{(0)}(t),
\]

i.e., the leading order pressure (and hence density) are constant in space.

Next, use this fact to enforce incompressibility. The \( O(1) \) equations are

\[
\begin{align*}
\partial_t \rho^{(0)} + \nabla \cdot (\rho^{(0)} u^{(0)}) &= 0 \quad (1.7) \\
\partial_t (\rho^{(0)} u^{(0)}) + \nabla \cdot (\rho^{(0)} u^{(0)} \otimes u^{(0)}) + \nabla p^{(2)} &= 0. \quad (1.8)
\end{align*}
\]

Incompressibility is enforced using the *boundary conditions*. Three types of boundary conditions that give incompressibility are listed below.

1. **Wall boundary condition.** The problem takes place in a bounded domain \( \Omega \) with fixed walls. In this case \( u \cdot n = 0 \) on \( \partial \Omega \). Thus, if one integrates the density equation over the domain one obtains
\[ |\Omega| \partial_t \rho^{(0)} + \rho^{(0)} \int_{\Omega} \nabla \cdot \mathbf{u}^{(0)} = 0 \]
\[ \partial_t \rho^{(0)} = -\frac{\rho^{(0)}}{|\Omega|} \int_{\Omega} \nabla \cdot \mathbf{u}^{(0)} \]
\[ \partial_t \rho^{(0)} = -\frac{\rho^{(0)}}{|\Omega|} \int_{\partial\Omega} \mathbf{u}^{(0)} \cdot \mathbf{n} \]
\[ \partial_t \rho^{(0)} = 0. \]

The last step comes from applying the boundary condition. Thus the density does not change and one sets \( \nabla \cdot \mathbf{u}^{(0)} = 0 \).

2. **Periodic boundary condition.** This result is similar to above. Integrating over the domain gives

\[ |\Omega| \partial_t \rho^{(0)} + \rho^{(0)} \int_{\Omega} \nabla \cdot \mathbf{u}^{(0)} = 0 \]
\[ \partial_t \rho^{(0)} = -\frac{\rho^{(0)}}{|\Omega|} \int_{\Omega} \nabla \cdot \mathbf{u}^{(0)} \]
\[ \partial_t \rho^{(0)} = -\frac{\rho^{(0)}}{|\Omega|} \int_{\partial\Omega} \mathbf{u}^{(0)} \cdot \mathbf{n} \]
\[ \partial_t \rho^{(0)} = 0. \]

The last step cancels out because the velocities are the same at opposite ends of the box, while the normal vectors point in opposite directions, thus giving cancellation. Again one gets that \( \nabla \cdot \mathbf{u}^{(0)} = 0 \).

3. **Open boundary.** In this case, one needs to have some sort of far-field boundary condition — usually this is enforced on the pressure. By assuming that this
does not change with time (which makes sense, as it is just a general background pressure), one gets that $P^{(0)}(t) = P_0 = \rho^{(0)}$, and thus $\partial_t \rho^{(0)} = 0$, giving $\nabla \cdot \mathbf{u}^{(0)} = 0$.

One consequence of these results is that incompressibility comes from a global procedure — a numerical scheme that hopes to capture it will have to take this into account.

Thus, assuming one of these boundary conditions one has that $\rho^{(0)}$ is constant in space and time and the incompressibility condition is satisfied for the leading order velocity. Next, looking at the $O(1)$ momentum equation

$$\rho^{(0)} \partial_t \mathbf{u}^{(0)} + \rho^{(0)} \nabla \cdot (\mathbf{u}^{(0)} \otimes \mathbf{u}^{(0)}) + \nabla p^{(2)} = \frac{1}{Re} \Delta \mathbf{u}^{(0)},$$

using the incompressibility found above gives

$$\nabla \cdot \mathbf{u}^{(0)} = 0$$

$$\partial_t \mathbf{u}^{(0)} + (\mathbf{u}^{(0)} \cdot \nabla) \mathbf{u}^{(0)} + \frac{1}{\rho^{(0)}} \nabla p^{(2)} = \frac{1}{Re} \Delta \mathbf{u}^{(0)},$$

which is the incompressible Navier-Stokes equation.

### 1.3.3 Numerical difficulties in the low Mach number limit

Standard finite volume shock-capturing hyperbolic solvers have difficulties in low Mach number regimes. The compressible equations have acoustic waves that scale as $O(\frac{1}{\varepsilon})$ which require temporal resolution for stability. Furthermore, artificial viscosity on the order of the wave speeds is also introduced to suppress numerical oscillations across shocks and contact discontinuities. Thus, for a desired spatial accuracy, one must also resolve the waves in space. This is prohibitively expensive to undertake as the underlying incompressible system does not admit these acoustic waves, and the relevant time scale of interest is that of the fluid velocity $\mathbf{u}$. We illustrate these problems by looking at the
one-dimensional Euler system with a simple local Lax-Friedrichs finite volume method, for simplicity, though these same issues will arise with other standard choices for the numerical flux.

\[
\frac{\rho_j^{n+1} - \rho_j^n}{\Delta t} + \frac{m_j^{n+1} - m_j^n}{2\Delta x} - \frac{\lambda_{j+1/2}}{2\Delta x} (\rho_{j+1}^n - \rho_j^n) + \frac{\lambda_{j-1/2}}{2\Delta x} (\rho_j^n - \rho_{j-1}^n) = 0
\]

\[
\frac{m_j^{n+1} - m_j^n}{\Delta t} + \left( \frac{m^2/n}{\rho} \right)_{j+1}^n - \left( \frac{m^2/n}{\rho} \right)_{j-1}^n + \frac{1}{\varepsilon^2} \frac{p_{j+1}^n - p_{j-1}^n}{2\Delta x} - \frac{\lambda_{j+1/2}}{2\Delta x} (m_{j+1}^n - m_j^n) + \frac{\lambda_{j-1/2}}{2\Delta x} (m_j^n - m_{j-1}^n) = 0.
\]

Here \(\lambda_{j+1/2} = \max\{\lambda_j, \lambda_{j+1}\}\), where \(\lambda_j\) is the maximum wave speed based on the values in cell \(j\).

The maximum wave speed of the system is \(\lambda_{\max} = |u_{\max}| \pm \frac{1}{\varepsilon} \sqrt{p'(\rho_{\max})}\) (where \(u_{\max}\) and \(\rho_{\max}\) are the values at the point where the maximum wave speed is reached); thus we need

\[
\Delta t = \nu \frac{\Delta x}{\lambda_{\max}} = \nu \frac{\Delta x \varepsilon}{\varepsilon |u_{\max}| + \sqrt{p'(\rho_{\max})}} = O(\varepsilon \Delta x)
\]

for some \(0 \leq \nu \leq 1\) to satisfy the CFL condition for stability. Furthermore, the numerical diffusion, while local, will at least at one point be given by

\[
\frac{\Delta x \lambda_{\max}}{2} = \frac{\Delta x (\varepsilon |u_{\max}| + \sqrt{p'(\rho_{\max})})}{\varepsilon 2\nu}.
\]

Thus we need \(\Delta x = o(\varepsilon)\) for accuracy purposes, ensuring that the numerical diffusion does not dominate the solution. A further consequence of this is that we now need \(\Delta t = o(\varepsilon \Delta x) = o(\varepsilon^2)\). This is unacceptable, as the limiting incompressible equations only have a timestep restriction of \(\Delta t = O(\Delta x)\), independent of \(\varepsilon\).
1.4 Chapter Organization

In Chapter 2, approaches for computing numerical solutions to the $P_1$ equations are introduced. In particular, the challenges associated with upwind methods are identified and we review how a splitting approach can be used to attack these challenges. Some cases which exhibit unphysical oscillations within this framework are explored. We then investigate the first order version of the splitting scheme in more detail, which includes diagnosing the source of oscillations and implementing corrections. Numerical results are presented to exhibit the improvement affected by our corrections. The details of numerical schemes are given in Appendix A. This is joint work with Cory Hauck of Oak Ridge National Laboratory, and a large portion of it was done while visiting the Center for Nonlinear Studies at Los Alamos National Laboratory.

In Chapter 3, we review the details of the low Mach number limit of the isentropic Euler and Navier-Stokes equations and make a close study of the difficulties encountered by standard hyperbolic shock capturing methods in this regime. We propose a hyperbolic splitting of the system to separate the fast acoustic waves from the low-speed hyperbolic flow. The fast acoustic system is a linear system with constant coefficients, which can be solved implicitly, while the relatively slow system contains the flow dynamics and is solved using an explicit shock capturing central scheme. We then perform an asymptotic analysis to show that the scheme becomes an incompressible scheme in the low Mach number limit. We then provide numerical results on a number of problems in both compressible and incompressible regimes. This is joint work with Shi Jin and Jian-Guo Liu.
Chapter 2

Numerical Diffusion for AP schemes in the diffusion limit

In the following subsections, we review schemes for the $P_1$ system based on both up-winding and splitting. Because these schemes are not new, many of the details of their implementation have been moved to the appendix.

2.1 Upwind discretization

Upwind discretizations of (1.3) and similar equations are known to have difficulties in diffusive regimes [52, 34, 37, 60, 59]. For fixed mesh spacing $\Delta x$, there are two related issues that arise in the limit $\varepsilon \to 0$: (i) excessive numerical dissipation and (ii) a restrictive time step. Both are easily understood by way of a semi-discrete, first-order, upwind scheme for (1.4):

\begin{align}
\frac{d}{dt} \phi_j^+ + \frac{1}{\varepsilon} \frac{1}{\sqrt{3}} \frac{\phi_{j+1}^+ - \phi_j^-}{\Delta x} &= -\frac{\sigma_j}{2\varepsilon^2} (\phi_j^+ - \phi_j^-), \\
\frac{d}{dt} \phi_j^- - \frac{1}{\varepsilon} \frac{1}{\sqrt{3}} \frac{\phi_{j-1}^- - \phi_j^+}{\Delta x} &= -\frac{\sigma_j}{2\varepsilon^2} (\phi_j^- - \phi_j^+),
\end{align}

(2.1a) (2.1b)
which, in terms of \( \rho \) and \( m \), takes the form

\[
d_t \rho_j + \frac{m_{j+1} - m_{j-1}}{2\Delta x} = \frac{\Delta x}{2\varepsilon} \frac{1}{\sqrt{3}} \frac{\rho_{j+1} - 2\rho_j + \rho_{j-1}}{\Delta x^2}, \tag{2.2a}
\]

\[
d_t m_j + \frac{1}{\varepsilon^2 \frac{1}{3}} \frac{\rho_{j+1} - \rho_{j-1}}{2\Delta x} + \frac{\sigma_j}{\varepsilon^2} m_j = \frac{\Delta x}{2\varepsilon} \frac{1}{\sqrt{3}} \frac{m_{j+1} - 2m_j + m_{j-1}}{\Delta x^2}. \tag{2.2b}
\]

Here, and for the remainder of this paper, the subscript \( j \) for state variables \( \rho \) and \( m \) and the cross-section \( \sigma \) denotes the average value of the respective quantity over a computational cell \( I_j \) of width \( \Delta x \). Later, we will also introduce composite parameters \( \gamma_j, \kappa_j, \) and \( \beta_j \), which depend on \( \sigma_j \).

For \( \varepsilon \) small, the dominant balance in (2.2b) is

\[
m_j = -\frac{1}{3\sigma_j} \frac{\rho_{j+1} - \rho_{j-1}}{2\Delta x} + O(\varepsilon), \tag{2.3}
\]

which when substituted into (2.2a), gives

\[
d_t \rho_j = \frac{1}{3} \frac{\sigma_{j+1}(\rho_{j+2} - \rho_j) - \sigma_{j-1}(\rho_j - \rho_{j-2})}{4\Delta x^2} + \frac{\Delta x}{2\varepsilon} \frac{1}{\sqrt{3}} \frac{\rho_{j+1} - 2\rho_j + \rho_{j-1}}{\Delta x^2} + O(\varepsilon). \tag{2.4}
\]

As \( \varepsilon \to 0 \), the first term on the right hand side of (2.4) yields a consistent discretization of the diffusive flux on the right hand side of (1.5b). However, the second term on the right hand side of (2.4)—the numerical dissipation term—will clearly affect the accuracy of the solution unless the mesh spacing \( \Delta x \) is chosen much smaller than \( \varepsilon \)—an expensive undertaking given that one needs not resolve such small scales when discretizing the diffusion equation (1.5b) directly. The expense of resolving \( \varepsilon \) is exacerbated by a stiff hyperbolic CFL condition which requires that the time step \( \Delta t \) in any temporal discretization of (2.2) satisfies

\[
\Delta t < \varepsilon C \Delta x \tag{2.5}
\]

for some \( O(1) \) constant \( C \).
The dissipative nature of the upwind scheme (2.2) is exhibited by the numerical solutions of the $P_1$ equations presented in Figures 1(a) and 1(c), where temporal integration of the semi-discrete scheme is achieved using first-order, semi-implicit method: the fluxes are evaluated explicitly, but the source term in (2.2b) is handled implicitly. Solutions are computed for $x \in [0, 2]$, with a constant cross-section $\sigma = 1.0$, periodic boundary conditions, and initial conditions given by

$$
\rho = \begin{cases} 
2.0, & x \in [0.8, 1.2], \\
0.0, & x \in [0.0, 0.8) \cup (1.2, 2], 
\end{cases} \quad (2.6)
$$

Solution profiles for $\rho$ are given for $\varepsilon = 0.8$, $t = 0.5$ and $\varepsilon = 10^{-4}$, $t = 0.05$. These profiles make clear that, while the upwind scheme is satisfactory for $O(1)$ values of $\varepsilon$, it lacks any reasonable sense of accuracy when $\varepsilon$ is small. Indeed, the numerical diffusion is so dominant when $\varepsilon = 10^{-4}$, that the profiles for $\rho$ in Figure 1(a) appear as a flat line. For comparison, a highly resolved reference solution is computed with a second-order, upwind scheme with 20,000 computational cells.

Higher-order Godunov-type schemes will decrease the numerical dissipation in (2.4) with respect to $\Delta x$, but the factor of $\varepsilon^{-1}$ remains. For example, in smooth regions where a slope-limiter is not needed, the modified system for the second order scheme of Van Leer [69] is [52]

$$
\partial_t \rho + \partial_x m = \frac{1}{12} \Delta x^2 \partial_x^3 m - \frac{1}{8\sqrt{3}} \frac{\Delta x^3}{\varepsilon} \partial_x^4 \rho, \quad (2.7a)
$$

$$
\partial_t m + \frac{1}{3} \frac{1}{\varepsilon^2} \partial_x \rho + \frac{\sigma}{\varepsilon^2} m = \frac{1}{36} \frac{\Delta x^3}{\varepsilon^2} \partial_x^3 m - \frac{1}{8\sqrt{3}} \frac{\Delta x^3}{\varepsilon} \partial_x^4 m. \quad (2.7b)
$$
Meanwhile the modified system for the piecewise parabolic method [10], again with no limiter, is [34]

\[\begin{align*}
\partial_t \rho + \partial_x m &= -\frac{1}{12\sqrt{3}} \frac{\Delta x^3}{\varepsilon} \partial_x^4 \rho + \frac{1}{30} \Delta x^4 \partial_x^5 m, \\
\partial_t m + \frac{1}{3\varepsilon^2} \partial_x \rho + \frac{\sigma}{\varepsilon^2} m &= -\frac{1}{12\sqrt{3}} \frac{\Delta x^3}{\varepsilon} \partial_x^4 m + \frac{1}{90} \frac{\Delta x^4}{\varepsilon^2} \partial_x^5 \rho.
\end{align*}\] (2.8a)

(2.8b)

In both cases, the numerical dissipation term in the equation for \(\rho\) is \(O(\Delta x^3/\varepsilon)\).

Profiles for \(\rho\) computed with a second-order upwind method are displayed in Figures 1(b) and 1(d). When \(\varepsilon = 0.8\), the scheme performs well, but when \(\varepsilon = 10^{-4}\), the results are strongly dependent on the mesh spacing. For example, with 200 computational cells, \(\Delta x^3/\varepsilon = 10^{-2}\). Hence the dissipation is relatively small and the solution in Figure 1(d) captures the profile of the reference solution fairly well. However, with 50 computational cells, \(\Delta x^3/\varepsilon = 0.64\), which means the dissipation term in (2.7) contributes significantly. As a result the profile is quite smeared.

### 2.2 Splitting method

The splitting approach introduced in [37, 60] was designed to eliminate the problem of excessive numerical dissipation and to improve the restrictive CFL condition in (2.5). The idea is to separate the \(P_1\) system into the following components:

1. Stiff Component:

\[\begin{align*}
\partial_t \rho &= 0, \\
\partial_t m + \frac{1 - \varepsilon^2}{3\varepsilon^2} \partial_x \rho &= -\frac{\sigma}{\varepsilon^2} m.
\end{align*}\] (2.9a)

(2.9b)
Figure 1: Standard upwind schemes with periodic boundary conditions and initial condition (2.6). In all four figures, the triangles represent the solution with $\Delta x = 0.04$, the circles represent the solution with $\Delta x = 0.01$, and the solid line is a highly resolved reference solution that uses $\Delta x = 10^{-4}$. The timestep is $\Delta t = 0.1\sqrt{3\varepsilon \Delta x}$. 

(a) $\varepsilon = 0.8$, first order in space and time  
(b) $\varepsilon = 0.8$, second order in space and time  
(c) $\varepsilon = 10^{-4}$, first order in space and time  
(d) $\varepsilon = 10^{-4}$, second order in space and time
2. Non-Stiff Component:

\[
\begin{align*}
\partial_t \rho + \partial_x m &= 0, \\
\partial_t m + \frac{1}{3} \partial_x \rho &= 0.
\end{align*}
\]

(2.9c) (2.9d)

For simplicity, we are assuming that \( \varepsilon < 1 \), although such a restriction is easily lifted [37]. With this splitting, each component is formally well defined. In particular, the non-stiff component is a hyperbolic system with wave speeds \( \pm 1/\sqrt{3} \).

The splitting scheme is implemented by updating the stiff component (2.9a-b) first, using the backward Euler method for the time integration and central differencing to evaluate the spatial derivative of \( \rho \). This gives the intermediate state

\[
\begin{align*}
\rho^*_j &= \rho^n_j, \\
m^*_j &= \gamma_j m^n_j - \frac{\kappa_j \rho^n_{j+1} - \rho^n_{j-1}}{2\Delta x},
\end{align*}
\]

(2.10a) (2.10b)

where

\[
\gamma_j := \frac{\varepsilon^2}{\varepsilon^2 + \sigma_j \Delta t} \quad \text{and} \quad \kappa_j := \frac{(1 - \varepsilon^2) \Delta t}{\varepsilon^2 + \sigma_j \Delta t}.
\]

(2.11)

The non-stiff component is then evaluated with a first-order upwind method, using \((\rho^*, m^*)\) for an initial condition:

\[
\begin{align*}
\frac{\rho^{n+1}_j - \rho^n_j}{\Delta t} + \frac{m^{n+1}_j - m^{n-1}_j}{2\Delta x} &= \frac{1}{\sqrt{3}} \frac{\rho^n_{j+1} - 2\rho^n_j + \rho^n_{j-1}}{2\Delta x}, \\
\frac{m^{n+1}_j - m^n_j}{\Delta t} + \frac{1}{3} \frac{\rho^n_{j+1} - \rho^n_{j-1}}{2\Delta x} &= \frac{1}{\sqrt{3}} \frac{m^n_{j+1} - 2m^n_j + m^n_{j-1}}{2\Delta x}.
\end{align*}
\]

(2.12a) (2.12b)

Because \( \rho \) does not evolve in the stiff step (2.10), the implementation of the two steps is essentially explicit.

Alternating evaluation of the two steps in (2.10) and (2.12) produces a first order scheme. Second-order accuracy in space can be easily obtained using a MUSCL approach.
to update the non-stiff component. However, second-order temporal accuracy requires a more advanced time integrator such as the one used in [37] or one of the IMEX methods discussed in [61, 62].

It is straightforward to show that the splitting method produces an AP scheme. Indeed, as \( \varepsilon \to 0 \), (2.10b) becomes a consistent discretization of the balance in (1.5a):

\[
m^*_j = -\frac{1}{3\sigma_j} \frac{\rho^n_{j+1} - \rho^n_{j-1}}{2\Delta x}.
\]

When substituted into (2.12a), the relation in (2.13) yields a consistent discretization of the diffusive flux in (1.5b). Moreover, unlike the semidiscrete upwind discretization (2.4), the numerical dissipation term in (2.12a) is independent of \( \varepsilon \).

The splitting scheme also has a less restrictive CFL condition than the upwind scheme. In particular, one need only satisfy the two CFL conditions

\[
\frac{\gamma \Delta t}{\varepsilon} < C \Delta x \quad \text{and} \quad \kappa \Delta t < C \frac{\Delta x^2}{2},
\]

for some \( O(1) \) constant \( C \). These two conditions correspond to the hyperbolic and diffusive behavior, respectively, of the \( P_1 \) system. When \( \varepsilon < C \sigma \Delta x \), the hyperbolic condition on the left will be satisfied independently of the choice of \( \Delta t \) and, as \( \varepsilon \to 0 \), the condition on the right becomes the standard time step condition for an explicit diffusion equation:

\[
\Delta t < \frac{C \sigma h^2}{2}.
\]

Using the splitting approach, we repeat the same set of numerical experiments that were performed with the upwind schemes: with \( \sigma = 1.0 \), periodic boundary conditions, and initial conditions given by (2.6). Results are presented in Figure 2. When \( \varepsilon = 0.8 \), the splitting method and upwind method give similar results at both first and
second order. (Compare to Figure 1.) When $\varepsilon = 10^{-4}$, the splitting method is more accurate than the upwind method, particularly at first order and with the coarser mesh. Unfortunately, the profile for the splitting method on the coarse mesh contains what appear to be binary oscillations. Such oscillations are a drawback of the splitting method.

In addition to the binary oscillations observed in Figure 2, other types of oscillations appear in other problems. In Figure 3(a), for example, oscillations appear in the $\rho$ profile of an experiment that uses the discontinuous cross-section

$$\sigma = \begin{cases} 
0.02, & x \in [0.35, 0.65] \cup [1.35, 1.65], \\
1.0, & x \in [0, 0.35) \cup (0.65, 1.35) \cup (1.65, 2]. 
\end{cases} \quad (2.16)$$

This profile is computed using a first-order splitting scheme with $\varepsilon = 0.1$ and the same initial condition (2.6). Clearly, non-physical oscillations occur in the regions where $\sigma$ is smaller. These oscillations appear to be dispersive and are not restricted to discontinuous cross-sections. For example, similar oscillations also occur when the cross-section is set to the constant value $\sigma = 0.02$, as is depicted Figure 3(b). These results are of particular concern because the oscillations cause the profile to go negative. The reference solution in these figures is generated with a second-order in space and time upwind scheme with $\Delta x = 0.01$. This spatial step resolves the mean free path, and the numerical diffusion is $O(h^3/\varepsilon) = O(10^{-5})$.

### 2.3 A closer look at operator splitting

In this section, we take a closer look at the operator splitting method, in an attempt to diagnose and correct oscillations discussed at the end of the previous section. For simplicity, we focus on the first order implementation, in which case updating the state
Figure 2: The splitting scheme, with periodic boundary conditions and initial condition (2.6). In all four figures, the triangles represent the solution with $\Delta x = 0.04$, the circles represent the solution with $\Delta x = 0.01$, and the solid line is a reference solution with $\Delta x = 10^{-4}$. The timesteps are chosen using the CFL condition (2.14) with $C = 0.1$. 

(a) $\varepsilon = 0.8$, first order in space and time  
(b) $\varepsilon = 0.8$, second order in space and time  
(c) $\varepsilon = 10^{-4}$, first order in space and time  
(d) $\varepsilon = 10^{-4}$, second order in space and time
Figure 3: First-order splitting scheme results. Left: discontinuous cross-section given by (2.16) and initial condition given by (2.6), $\varepsilon = 0.1$, and $t = 0.075$. Right: Constant cross-section of $\sigma = 0.02$, initial condition given by (2.6), $\varepsilon = 0.1$, and $t = 0.1$. In both figures, circles represent the numerical solution with $\Delta x = 0.01$ and time step determined by (2.14) with $C = 0.1$. The solid line is a reference solution computed with a second-order upwind scheme with $\Delta x = 0.01$ and $\Delta t = 0.1 \sqrt{3\varepsilon \Delta x}$.

of the $P_1$ system from $(\rho^n, m^n)$ to $(\rho^{n+1}, m^{n+1})$ requires one iteration of the two steps in (2.10) and (2.12). The complete scheme is

\[
\frac{\rho_j^{n+1} - \rho_j^n}{\Delta t} + \frac{\gamma_j m_{j+1}^n - \gamma_j m_{j-1}^n}{2\Delta x} = \frac{1}{3} \frac{\kappa_{j+1}(\rho_{j+2}^n - \rho_j^n) - \kappa_{j-1}(\rho_j^n - \rho_{j-2}^n)}{4\Delta x^2}
\]

\[
+ \frac{1}{\sqrt{3}} \frac{\rho_{j+1}^n - 2\rho_j^n + \rho_{j-1}^n}{2\Delta x},
\]

\[
\frac{m_j^{n+1} - m_j^n}{\Delta t} + \frac{\beta_j \rho_j^n}{3\varepsilon^2} + \frac{\gamma_j \sigma_j m_j^n}{\varepsilon^2} = \frac{1}{\sqrt{3}} \frac{\gamma_j m_{j+1}^n - 2\gamma_j m_j^n + \gamma_j m_{j-1}^n}{2\Delta x}
\]

\[- \frac{1}{3\sqrt{3}} \frac{\kappa_{j+1}(\rho_{j+2}^n - \rho_j^n) - 2\kappa_j(\rho_{j+1}^n - \rho_j^n) + \kappa_{j-1}(\rho_j^n - \rho_{j-2}^n)}{4\Delta x^2},
\]

where $\gamma_j$ and $\kappa_j$ are given in (2.11) and

\[
\beta_j := \gamma_j + \varepsilon^2(1 - \gamma_j).
\]

Two terms in (2.17) immediately stand out as possible sources of oscillations. First
is one of the diffusive terms appearing in (2.17a):

\[
\frac{\kappa_{j+1}(\rho_{j+2}^n - \rho_j^n) - \kappa_{j-1}(\rho_j^n - \rho_{j-2}^n)}{4\Delta x^2} = \partial_x(\kappa \partial_x \rho) + O(h),
\]

which is based on a wide stencil that uses every other cell—a so-called *black-red discretization*. In [30], it is speculated that this stencil makes the scheme susceptible to binary oscillations. The other obvious candidate is the dispersive term in (2.17b):

\[
\frac{\kappa_{j+1}(\rho_{j+2}^n - \rho_j^n) - 2\kappa_j(\rho_{j+1}^n - \rho_{j-1}^n) + \kappa_{j-1}(\rho_j^n - \rho_{j-2}^n)}{4\Delta x^2} = \frac{\Delta x}{2} \partial_x(\kappa \partial_x^2 \rho) + O(\Delta x^2).
\]

The presence of this term and the subtle numerical issues it poses has been noted in [31].

We first address the oscillations in Figures 2(c) and 2(d). Because the oscillations appear to be binary, we replace the black-red diffusion discretization in (2.17) with a more standard discretization:

\[
\frac{\kappa_{j+1}(\rho_{j+2}^n - \rho_j^n) - \kappa_{j-1}(\rho_j^n - \rho_{j-2}^n)}{4\Delta x^2} \rightarrow \frac{\kappa_{j+1/2}(\rho_{j+1}^n - \rho_j^n) - \kappa_{j-1/2}(\rho_j^n - \rho_{j-1}^n)}{\Delta x^2}.
\]

where \( \kappa_{j+1/2} \) is the harmonic average of \( \kappa_j \) and \( \kappa_{j+1} \). It is clear from Figure 2.4(a) that this change has fixed the problem.

It should also be noted that for the *smooth* initial condition

\[
\rho = e^{-10(x-1)^2} \quad \text{and} \quad m = 0,
\]

there are no oscillations. See Figure 2.4(b). Thus we conclude that these binary oscillations are associated with the discontinuity in the initial condition (2.6).

The next task is to determine the cause of oscillations depicted in Figure 3. These oscillations appear to be dispersive, so our first inclination is to remove the dispersive
Figure 4: On the left, a recalculation of Figure 2(c), with the black-red diffusion discretization replaced with a standard discretization, as described in (2.21). On the right, the black-red discretization is kept, but the smooth initial condition (2.22) is used. In both cases, triangles represent the solution with $\Delta x = 0.04$ and the solid line is a highly resolved reference solution.

term (2.20) from (2.17). Because this term is $O(h)$, removing it does not affect the consistency of the scheme. Unfortunately, it is clear from Figure 2.5(a) that the oscillatory behavior does not improve. We also try replacing the black-red diffusion discretization with a more standard discretization, as described in (2.21). However, as Figure 2.5(b) shows, the oscillations remain.

It turns out that one way to suppress the oscillations in Figure 3 can be motivated by the modified equations for (2.17), which have the form

\begin{align}
\partial_t \rho + \partial_x (\gamma m) &= \frac{1}{3} \partial_x (\kappa \partial_x \rho) + \frac{\Delta x}{2\sqrt{3}} \partial_x^2 \rho, \\
\partial_t m + \frac{1}{3} \beta \partial_x \rho + \frac{\gamma \sigma}{\varepsilon^2} m &= \frac{\Delta x}{2\sqrt{3}} \partial_x^2 (\gamma m) - \frac{\Delta x}{6\sqrt{3}} \partial_x (\kappa \partial_x^2 \rho),
\end{align}

(2.23)

where $\kappa$, $\beta$, and $\gamma$ are now functions of $x$. These modified equations are an $O(\Delta x, \Delta t)$
approximation of the following hyperbolic-parabolic system:

\begin{align}
\partial_t \rho + \partial_x (\gamma m) &= \frac{1}{3} \partial_x (\kappa \partial_x \rho) , \\
\partial_t m + \frac{1}{3} \beta \varepsilon^2 \partial_x \rho &= -\frac{\gamma \sigma}{\varepsilon^2} m ,
\end{align}

which, to our knowledge, was first derived for a constant cross-section in [31]. Following the work in [31], a similar set of modified equations was derived in [30] for general \( P_N \) systems [49, 63] in one dimension, but with spatially varying cross-sections.

For a constant cross-section (and thus constant \( \kappa, \beta, \) and \( \gamma \)), a first-order semi-discrete method for (2.24) is given by

\begin{align}
\partial_t \rho_j + \gamma \frac{m_{j+1} - m_{j-1}}{2\Delta x} &= \frac{\Delta x}{2\varepsilon} \sqrt{\beta \gamma} \frac{\rho_{j+1} - 2\rho_j + \rho_{j-1}}{3 \Delta x^2} , \\
\partial_t m_j + \frac{1}{\varepsilon^2} \beta \frac{\rho_{j+1} - \rho_{j-1}}{3 \Delta x} + \frac{\sigma}{\varepsilon^2} m_j &= \frac{\Delta x}{2\varepsilon} \sqrt{\beta \gamma} \frac{m_{j+1} - 2m_j + m_{j-1}}{3 \Delta x^2} .
\end{align}

The numerical dissipation terms, which are necessary for stability and suppression of oscillations, are on the right. Meanwhile, the discretization for (2.24) given by the
splitting scheme (2.17) has numerical dissipation terms of the form

\[
\frac{1}{\sqrt{3}} \frac{\rho_{j+1}^n - 2\rho_j^n + \rho_{j-1}^n}{2\Delta x} \quad \text{and} \quad \frac{\gamma}{\sqrt{3}} \frac{m_{j+1}^n - 2m_j^n + m_{j-1}^n}{2\Delta x}.
\] (2.26)

We conclude that the splitting scheme in (2.17) suffers from too little numerical dissipation, as opposed to a standard upwind solver, which uses too much. Indeed, if we replace the dissipation terms in (2.17) by the corresponding dissipation terms in (2.24), the oscillations in Figure 3 disappear and we obtain the correct behavior. These results are presented in Figures 6(a) and (b) for both the constant and non-constant cross-section.

Rather than repeat our analysis for the second-order version of the splitting scheme (a much more tedious task), we work directly with the modified system (2.24). For a non-constant cross-section, discretization of this system is a challenge because it is not in conservative form. Here the splitting scheme provides some guidance. In particular, the form of the dissipation terms in (2.17) suggests the appropriate discretization for (2.24). In terms of the variables \( \rho \) and \( q := \beta^{-1}m \), the modified system (2.24) takes the form

\[
\partial_t \rho + \partial_x (\gamma \beta q) = \frac{1}{3} \partial_x (\kappa \partial_x \rho), \quad (2.27a)
\]

\[
\partial_t q + \frac{1}{3\varepsilon^2} \partial_x \rho = \frac{\gamma \sigma}{\varepsilon^2} q. \quad (2.27b)
\]

This system is now in conservative form, but with spatially varying fluxes. It therefore requires a generalization of the upwind treatment. Details of the scheme are given in the appendix. Results that are first-order in time, but second-order in space are presented in Figures 6(b) and (d). The profiles in these figures are quite accurate, even though the time integration is only first-order.
Figure 6: Left: $P_1$ solution at time $t = 0.1$ with constant cross-section $\sigma = 0.02$. Right: Solution at time $t = 0.075$ with the discontinuous cross section given by (2.16). In both figures, the solid line is the reference solution and the circles are the solution using the splitting method with the corrected numerical diffusion. The timestep is chosen using the CFL conditions (2.14) with $C = 0.1$.
Chapter 3

The all-speed asymptotic-preserving method for the isentropic Navier-Stokes equations

3.1 All-speed asymptotic-preserving schemes

Our goal is to develop a method that preserves, at the discrete level, the asymptotic passage from the compressible equations to the incompressible equations without resolving the spatial and temporal scales associated with the acoustic waves, which are unimportant in the limit. In particular, the numerical method should demonstrate this discrete asymptotic limit by being consistent with a method for the incompressible equations in the low Mach number limit. This discrete limit is taken with spatial and temporal steps fixed, with only the reference Mach number parameter driven to zero. A method that satisfies these properties is said to be asymptotic-preserving (AP) [31].

We stress that we want to obtain an all-speed scheme, in other words, a scheme that gives correct solutions for any reference Mach number \( \varepsilon \), not just in the low Mach number regime. Therefore, we need to take care in the construction of our scheme that we use structures that are well-suited for each regime. To this end, we work in
conservative variables, to ensure that the scheme can capture shocks in compressible regimes, (i.e., Mach number of order unity) and track these values at the center of each computational cell as in finite volume schemes for hyperbolic systems. On the other hand, incompressible solvers typically have an implicit global projection step that calculates an intermediate velocity field and pressure to update the velocity equations. This suggests that our overall scheme should be semi-implicit, weaving together these features of the two systems.

### 3.2 Hyperbolic splitting

Looking at the compressible system (1.6), there are two scales that we need to resolve: the (fast) acoustic wave scale and the (slow) convection scale, which contains the underlying incompressible dynamics. Therefore, we split the system into two systems

\[
\begin{align*}
\partial_t \rho + \alpha \nabla \cdot (\rho u) + (1 - \alpha) \nabla \cdot (\rho u) &= 0 \\
\partial_t (\rho u) + \nabla \cdot (\rho u \otimes u) + \nabla \left( \frac{p(\rho) - a(t)\rho}{\varepsilon^2} \right) + \frac{a(t)}{\varepsilon^2} \nabla \rho &= 1 \frac{1}{Re} \Delta u
\end{align*}
\]

The slow dynamics are the system

\[
\begin{align*}
\partial_t \rho + \alpha \nabla \cdot (\rho u) &= 0 \\
\partial_t (\rho u) + \nabla \cdot (\rho u \otimes u) + \nabla \left( \frac{p(\rho) - a(t)\rho}{\varepsilon^2} \right) &= 1 \frac{1}{Re} \Delta u,
\end{align*}
\]

and the fast dynamics are contained in the system

\[
\begin{align*}
\partial_t \rho + (1 - \alpha) \nabla \cdot (\rho u) &= 0 \\
\partial_t (\rho u) + \frac{a(t)}{\varepsilon^2} \nabla \rho &= 0
\end{align*}
\]
The key splitting here is of the stiff pressure term. We subtract off a linear piece $a(t)\rho$, to be determined below, and add it back in fast system. The splitting parameter $0 < \alpha < 1$ determines how much of the momentum is seen by each system. As we will see, some momentum is necessary in the fast system to ensure incompressibility.

The choice of the splitting $a(t)$ is motivated by hyperbolicity. The wave speeds of (3.2) in the $x$ direction are

$$\lambda = u, u \pm \sqrt{\frac{(1 - \alpha)u^2 + \alpha(p'(\rho) - a(t))}{\varepsilon^2}},$$

where $u$ is the first component of $\mathbf{u}$, so if we choose

$$a(t) = \min_x p'(\rho)$$

for our pressure splitting term, we ensure that the wave speeds will always be real, and thus the slow system is hyperbolic. As we will see in the asymptotic analysis later, in the low Mach number limit the fast system (3.3) will force variations in $\rho$ to be small, and thus the wave speeds of the slow system will be $O(1)$. We can discretize this system using any shock-capturing hyperbolic solver, noting that the wave speeds are no longer stiff which avoids the dissipation and time step problems seen in the original system (1.6). The fast dynamics, contained in (3.3), are simply a linear system with constant coefficients, which leads to a straightforward implicit implementation.

### 3.2.1 Time discretization of the split systems

For a first order in time scheme, we can write the system in a simple semi-implicit form

$$\frac{\rho^{n+1} - \rho^n}{\Delta t} + \alpha \nabla \cdot (\rho \mathbf{u})^n + (1 - \alpha) \nabla \cdot (\rho \mathbf{u})^{n+1} = 0 \quad (3.4)$$

$$\frac{(\rho \mathbf{u})^{n+1} - (\rho \mathbf{u})^n}{\Delta t} + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u})^n + \nabla \left( \frac{p'(\rho) - a(t)\rho}{\varepsilon^2} \right)^n \nabla \rho^{n+1} + \frac{a(t)}{\varepsilon^2} \Delta \mathbf{u}^n = \frac{1}{Re} \Delta \mathbf{u}^n \quad (3.5)$$
To obtain second order in time, we use a two-level Adams-Bashfort discretization for the explicit, slow terms and Crank-Nicholson for the fast, implicit terms

\[
\frac{\rho^{n+1} - \rho^n}{\Delta t} + \alpha \nabla \cdot \left( \frac{3}{2} (\rho u)^n - \frac{1}{2} (\rho u)^{n-1} \right) + (1 - \alpha) \nabla \cdot \left( \frac{(\rho u)^{n+1} + (\rho u)^n}{2} \right) = 0
\]

(3.6)

\[
\frac{(\rho u)^{n+1} - (\rho u)^n}{\Delta t} + \nabla \cdot \left( \frac{3}{2} (\rho u \otimes u)^n - \frac{1}{2} (\rho u \otimes u)^{n-1} \right) + \nabla \cdot \left( \frac{3}{2} \left( \frac{p(\rho) - a(t)\rho}{\varepsilon^2} \right)^n - \frac{1}{2} \left( \frac{p(\rho) - a(t)\rho}{\varepsilon^2} \right)^{n-1} \right) + \frac{a(t)}{\varepsilon^2} \nabla \rho^{n+1} + \rho^n = \frac{1}{Re} \Delta \left( \frac{3}{2} u^n - \frac{1}{2} u^{n-1} \right).
\]

(3.7)

Degond and Tang [14] noted that one can rewrite the momentum equation in (3.7) in terms of \((\rho u)^{n+1}\) and insert it into the density equation, obtaining an elliptic equation for \(\rho^{n+1}\):

\[
\frac{\rho^{n+1} - \rho^n}{\Delta t} + \alpha \nabla \cdot \left( \frac{3}{2} (\rho u)^n - \frac{1}{2} (\rho u)^{n-1} \right) + (1 - \alpha) \Delta t \nabla \cdot \left( \frac{3}{2} (\rho u \otimes u)^n - \frac{1}{2} (\rho u \otimes u)^{n-1} \right) - (1 - \alpha) \Delta t \nabla \cdot \left( \frac{3}{2} \left( \frac{p(\rho) - a(t)\rho}{\varepsilon^2} \right)^n - \frac{1}{2} \left( \frac{p(\rho) - a(t)\rho}{\varepsilon^2} \right)^{n-1} \right) - (1 - \alpha) \Delta t \nabla \cdot \frac{a(t)}{\varepsilon^2} \nabla \rho^{n+1} + \rho^n = (1 - \alpha) \Delta t \Delta \left( \frac{3}{2} \nabla \cdot u^n - \frac{1}{2} \nabla \cdot u^{n-1} \right).
\]

(3.8)

This system is now a Helmholtz equation for the unknown variable \(\rho^{n+1}\), and the terms from the previous steps can be pushed to the right hand side as source terms. If the right spatial discretization is chosen, this system can be solved efficiently for \(\rho^{n+1}\) using fast Fourier transform techniques. An important feature of this Helmholtz equation is
that it is uniformly elliptic for any $\varepsilon$ \cite{11,13}. The updated momentum $(\rho^{n+1}u^{n+1})$ is then obtained from the momentum equation (3.7).

Note that, while popular in many incompressible solvers, we do not discretize the diffusion terms using Crank-Nicholson. This is done for two reasons: first of all, because it involves the velocity rather than the momentum, and thus it is a nonlinear function of the conservative variables if this term is calculated fully implicitly, which would require iteration. More importantly, if treated implicitly this term will appear in the elliptic equation (3.8) for the updated density and thus disallows the application of the fast spectral solver for the updated density and pressure. Furthermore, for high Reynolds number flow, where $Re >> 1$, an explicit diffusion term is clearly adequate \cite{15}.

3.3 $L^2$ stability

To check the stability, we write this as a semi-implicit method

$$\frac{U^{n+1} - U^n}{\Delta t} + AU^n + BU^{n+1} = 0.$$ 

We have the following Lemma

**Lemma 3.1** If both methods

$$\frac{U^{n+1} - U^n}{\Delta t} + A(U^n) = 0, \quad \frac{U^{n+1} - U^n}{\Delta t} + BU^{n+1} = 0$$

are stable, then the original method is also stable.

**Proof**

Method $U^{n+1} = (I - \Delta t A)U^n$ is stable iff

$$\|I - \Delta t A\| \leq 1 + c_1 \Delta t.$$
Method $U^{n+1} = (I + \Delta tB)^{-1}U^n$ is stable iff

$$
\| (I + \Delta tB)^{-1} \| \leq 1 + c_2 \Delta t.
$$

Thus the combined method $U^{n+1} = (I + \Delta tB)^{-1}(I - \Delta tA)U^n$ is stable iff

$$
\| (I + \Delta tB)^{-1}(I - \Delta tA) \| \leq \| (I + \Delta tB)^{-1} \| \| (I - \Delta tA) \| \leq 1 + C \Delta t.
$$

Therefore, so long as the fast slow systems are *individually* stable, the combined scheme will also be stable.

### 3.4 Spatial discretization of the split systems

For simplicity, assume a uniform grid with spacing $\Delta x$ and define $\phi_{i,j} = \phi(x_i, y_j)$ for any variable $\phi$, where $(x_i, y_j) = (\frac{1}{2}\Delta x + i\Delta x, \frac{1}{2}\Delta x + j\Delta x)$. We also assume for ease of explanation a rectangular domain.

The fast system (3.3) is discretized in space using central differences:

$$
\nabla \cdot (\rho \mathbf{u})_{i,j}^{n+1} = (1 - \alpha)D_x^0(\rho u)_{i,j}^{n+1} + (1 - \alpha)D_y^0(\rho v)_{i,j}^{n+1}
$$

$$
\frac{a^n}{\varepsilon^2} \nabla \rho_{i,j}^{n+1} = \frac{a^n}{\varepsilon^2} \nabla_0 \rho_{i,j}^{n+1}
$$

Here $D_x^0 \phi_{i,j} = \frac{\phi_{i+1,j} - \phi_{i-1,j}}{2\Delta x}$, $D_y^0 \phi = \frac{\phi_{i,j+1} - \phi_{i,j-1}}{2\Delta x}$, is the central difference operator, $\nabla_0 \phi_{i,j} = \left(\begin{array}{c}
D_x^0 \phi_{i,j} \\
D_y^0 \phi_{i,j}
\end{array}\right)$ is the natural extension to the central difference gradient, and $a^n$ is the value of $a(t) = \min p'(\rho)$ at time $t = t^n$.

We discretize the convective flux terms in (3.2) using a second order central scheme [42], which is a higher-order extension of the Lax-Friedrichs scheme. This choice is
by no means unique — any standard shock-capturing scheme will be sufficient. However this choice leads to the most straightforward analysis. We take the conservative discretization,

$$F(U_{i,j}^n) = \frac{H_{i+1/2,j}^n - H_{i-1/2,j}^n}{\Delta x} + \frac{H_{i,j+1/2}^n - H_{i,j-1/2}^n}{\Delta x},$$

where

$$H_{i+1/2,j}^n = \frac{1}{2}(f(U_{i+1/2,j,+}^n) + f(U_{i+1/2,j,-}^n) - \lambda_{i+1/2,j}^n(U_{i+1/2,j,+}^n - U_{i+1/2,j,-}^n))$$

$$\lambda_{i+1/2,j}^n = \max \left\{ \sigma \left( \frac{\partial F}{\partial U}(U_{i+1/2,j,+}) \right), \sigma \left( \frac{\partial F}{\partial U}(U_{i+1/2,j,-}) \right) \right\}.$$  

$\sigma$ is the spectral radius of the Jacobians in (3.12), i.e., the maximum wave speed. Here $f$ is the relevant flux function from the slow system (3.2) chosen for $x$ and $y$ fluxes as needed. The edge values $U_{i+1/2,j}$ at each interface are reconstructed component-wise using the generalized minmod limiter, with $\theta \in [1, 2]$,

$$\sigma_{i,j}^n = \minmod \left( \theta \frac{u_{i+1,j,+}^n - u_{i,j}^n}{\Delta x}, \theta \frac{u_{i,j,+}^n - u_{i-1,j}^n}{\Delta x}, \frac{u_{i+1,j}^n - u_{i-1,j}^n}{2\Delta x} \right)$$

$$u_{i+1/2,j,+}^n = u_{i+1,j}^n - \frac{\Delta x}{2} \sigma_{i+1,j}^n, \quad u_{i+1/2,j,-}^n = u_{i,j}^n + \frac{\Delta x}{2} \sigma_{i,j}^n.$$  

Note here that in central schemes, the slope limiter is based on conserved variables rather than a local characteristic decomposition.

### 3.5 Boundary and initial conditions

Our example boundary conditions listed in section 2 only gave the boundary conditions for $u$, not $\rho$. To remedy this, we artificially enforce a boundary condition of $\partial \rho / \partial n = 0$ on $\partial \Omega$. In the following analysis we will also assume a solid wall boundary with $u \cdot n = 0$
on $\partial \Omega$. For the Navier-Stokes case, we extend this to the no-slip boundary condition $u = 0$ on $\partial \Omega$.

We enforce these boundary conditions by using ghost cells, setting

$$\rho_{-1} = \rho_0, \quad u_{-1} = -u_0, \quad (\rho u)_{-1} = -(\rho u)_0,$$  \hspace{1cm} (3.15)

as well as

$$\rho_{-2} = \rho_1, \quad u_{-2} = -u_1, \quad (\rho u)_{-2} = -(\rho u)_1,$$  \hspace{1cm} (3.16)

for calculating the reconstructions within the ghost cells. This results in a second order approximation of the boundary condition.

We only consider initial data of the form

$$\rho(0, x) = \rho_0 + \varepsilon^2 \rho^{(2)}(x) + \ldots$$

$$\nabla \cdot (u)(0, x) = O(\varepsilon).$$  \hspace{1cm} (3.17)

This initial data converges to admissible initial data of the limiting incompressible equation.

This scheme is not expected to work for general initial data. In fact, general initial data will generate $O(1/\varepsilon)$ acoustic waves that require resolution in space and time for accuracy and stability. This situation reflects a poor choice of terms in nondimensionalization, as the behavior of the solution is compressible, not incompressible.

### 3.6 The discrete low Mach number limit

Next, we show that in the limit $\varepsilon << 1$ this solver automatically transforms into an incompressible solver. In the following, we will write $\nabla_0$ to be the standard centered
difference gradient operator and $\tilde{\nabla}$ to be the shock-capturing difference operator from (3.11).

We write

$$
\rho_{i,j}^n = \rho_{i,j}^{n(0)} + \varepsilon^2 \rho_{i,j}^{n(2)} + \ldots \tag{3.18}
$$

$$
(\rho \mathbf{u})_{i,j}^n = (\rho \mathbf{u})_{i,j}^{n(0)} + \varepsilon^2 (\rho \mathbf{u})_{i,j}^{n(2)} + \ldots \tag{3.19}
$$

$$
p_{i,j}^n = p_{i,j}^{n(0)} + \varepsilon^2 p_{i,j}^{n(2)} + \ldots = (\rho_{i,j}^{n(0)})^\gamma + \varepsilon^2 (\rho_{i,j}^{n(0)})^{\gamma-1} p_{i,j}^{n(2)} + \ldots \tag{3.20}
$$

We skip $\varepsilon^1$ in the expansion because there are no $O\left(\frac{1}{\varepsilon}\right)$ terms in the discrete equations.

We now look at how the terms balance at each order in $\varepsilon$. For small $\varepsilon$, the $O\left(\frac{1}{\varepsilon^2}\right)$ terms are given by

$$
\frac{a^n}{4\Delta x} (\rho_{i+1,j}^{n+1} - \rho_{i,j}^{n+1} + \rho_{i,j}^{n} - \rho_{i-1,j}^{n}) = 0
$$

$$
\frac{a^n}{4\Delta x} (\rho_{i,j+1}^{n+1} - \rho_{i,j}^{n+1} + \rho_{i,j}^{n} - \rho_{i,j-1}^{n}) = 0
$$

$$
\Rightarrow \rho_{i+1,j}^{n+1} - \rho_{i-1,j}^{n+1} = -(\rho_{i+1,j}^{n} - \rho_{i-1,j}^{n}) = 0, \quad \rho_{i+1,j}^{n+1} = \rho_{i-1,j}^{n+1}
$$

$$
\Rightarrow \rho_{i,j+1}^{n+1} - \rho_{i,j-1}^{n+1} = -(\rho_{i,j+1}^{n} - \rho_{i,j-1}^{n}) = 0, \quad \rho_{i,j+1}^{n+1} = \rho_{i,j-1}^{n+1}
$$

This result comes from the fact that the leading order density was constant at the previous timestep, consistent with the initial condition (3.17).

A straightforward application of the boundary conditions (3.15) gives $\rho_{i,j}^{n+1} = \rho_{i,j}^{0+n+1}$, $\forall i, j$, a constant in space but not necessarily in time. Next we need to show incompressibility, and this is where the $\alpha$ terms in the splitting become important. The
\( O(1) \) equation for the density is given by

\[
\frac{\rho_{i,j}^{n+1,(0)} - \rho_{i,j}^{n,(0)}}{\Delta t} + \alpha \nabla \cdot \left( \frac{3}{2} (\rho u^n) - \frac{1}{2} (\rho u)^{n-1} \right) + \frac{1 - \alpha}{2} \nabla_0 \cdot (\rho_{i,j}^{n+1,(0)} u_{i,j}^{n+1,(0)}) + \nabla_0 (\rho_{i,j}^{n+1,(0)} u_{i,j}^{n,(0)}) = 0
\]

Note that in \( \nabla \), the density jump at the interface is zero due to the constant profile of \( \rho^{(0)} \), resulting in a numerical dissipation term of the form \( \lambda_{i+1/2,j}^n (\rho_{i+1/2,j,+}^{(0),n} - \rho_{i+1/2,j,-}^{(0),n}) \) (note that the dissipation would not be of this form if a slope limiter based on a local characteristic decomposition is used). The operator is essentially the central difference operator that uses the reconstructed edge values. For a first order scheme it is exactly the central difference operator.

We now do the discrete analog of the integrals done in section 2.1 by summing this equation over all \( i, j \). Recalling that the leading order density is a constant and noting that the flux terms telescope and the resulting boundary terms cancel out, we result in

\[
N^2 \rho_0^{n+1} - \sum_{i,j} \rho_{i,j}^{n,(0)} = 0
\]

\[
\Rightarrow \rho_0^{n+1} = \frac{1}{N} \sum_{i,j} \rho_{i,j}^{n,(0)}.
\]

where \( N \) is the total number of grid points in each direction. This merely says that the new density is a constant and is simply equal to the average value of the previous timestep’s density variable. Furthermore, as the previous step was also constant to leading order in space this says that the two coincide, so the density is also constant in time as was seen in the continuous case.

Using this result, the density terms cancel out and we are left with

\[
\frac{3}{2} \alpha \nabla_0 \cdot (\rho^{(0)} u^{(0)})^n - \frac{1}{2} \alpha \nabla_0 \cdot (\rho^{(0)} u^{(0)})^n + (1 - \alpha) \rho^{n+1,(0)} \frac{1}{2} (\nabla_0 \cdot (u^{n+1,(0)} + u^{n,(0)})) = 0.
\]
As we assumed the initial velocity field was incompressible to $O(\varepsilon)$, the terms from the $t^n$ step drop out and we are left with $\nabla_0 \cdot u^{n+1,(0)} = 0$, the discrete incompressibility condition for $u^{n+1,(0)}$.

Finally we also note that we can derive an equation for the density correction term $\rho^{n+1,(2)}$ (and hence the incompressible pressure $p^{(2)}$) by looking at the $O(1)$ terms in the elliptic equation reformulation (3.8).

\[
\frac{\rho^{n+1,(0)} - \rho^{n,(0)}}{\Delta t} + \alpha \nabla_0 \cdot \left( \frac{3}{2}(\rho u)^{n,(0)} - \frac{1}{2}(\rho u)^{n-1,(0)} \right) \\
+ (1 - \alpha)\Delta t \nabla_0 \cdot (\rho u)^{n,(0)} \\
- \Delta t(1 - \alpha)\nabla_0 \cdot \hat{\nabla} \cdot \left( \frac{3}{2}(\rho u \otimes u)^{n,(0)} - \frac{1}{2}(\rho u \otimes u)^{n-1,(0)} \right) \\
- (1 - \alpha)\Delta t \nabla_0 \cdot \hat{\nabla} \left( \frac{3}{2} \left( \frac{p(\rho) - a^n \rho}{\varepsilon^2} \right)^{n,(0)} - \frac{1}{2} \left( \frac{p(\rho) - a^n \rho}{\varepsilon} \right)^{n-1,(0)} \right) \\
- (1 - \alpha)a^n \Delta t \nabla_0 \cdot \nabla_0 \rho^{n+1,(2)} + \rho^{n,(2)} \\
= \frac{1}{Re} (1 - \alpha)\Delta t \nabla_0 \cdot \Delta \left( \frac{3}{2}u^{n,(0)} - \frac{1}{2}u^{n-1,(0)} \right)
\]

(3.22)

Again, we have used the fact that $\hat{\nabla} = \nabla_0$ for the explicit flux terms in the leading order density equation. We can rewrite this as

\begin{align*}
- \frac{a^n}{4\Delta x^2} \left( \rho^{n+1,(2)}_{i+1,j} + \rho^{n+1,(2)}_{i-1,j} + \rho^{n+1,(2)}_{i,j+1} + \rho^{n+1,(2)}_{i,j-1} - 4\rho^{n+1,(2)}_{i,j} \right) &= \phi(U^{n,(0)}, U^{n+1,(0)})
\end{align*}

where $\phi$ collects all the explicit terms. This is simply a Poisson equation for $\rho^{n+1,(2)}$.

Using the expansion of the pressure (3.20) and the definition of $a^n$, we have that $a^n \rho^{(2)} = p^{(2)}$. Using this fact and knowledge of the previous steps’ leading order density and
incompressibility we have a discretization of a pressure Poisson equation

\[-\Delta_2 p^{(2)} = -\nabla_0 \cdot \tilde{\nabla}_0 \cdot \left( \frac{3}{2} (\rho \mathbf{u} \otimes \mathbf{u})^{n,(0)} - \frac{1}{2} (\rho \mathbf{u} \otimes \mathbf{u})^{n-1,(0)} \right) \]

\[+ \frac{1}{Re} \nabla_0 \cdot \Delta_0 \left( \frac{3}{2} \mathbf{u}^{n,(0)} + \frac{1}{2} \mathbf{u}^{n-1,(0)} \right). \]  

(3.23)

(3.24)

Here, \(\Delta_0\) is the standard centered second order Laplacian, and \(\Delta_2\) is the second order centered Laplacian with stencil of size \(2\Delta x\) generated by \(\nabla_0 \cdot \nabla_0\). In the continuous case, the divergence and Laplacian would commute and the diffusion-type term would drop out due to incompressibility, but this is not necessarily true at the discrete level.

We also note that the explicit pressure term drops out, as the modified explicit pressure \(\frac{1}{\varepsilon^2} (p(\rho) - a(t)\rho)\) has a simple Taylor expansion for small \(\varepsilon\) of

\[\tilde{p}_\varepsilon = \frac{1}{\varepsilon^2} (\rho_0^{\gamma} + \gamma \varepsilon^2 (\rho_0^{\gamma} - 1) \rho_{i,j}^{n+1,(2)} - a^n (\rho_0 + \varepsilon^2 \rho_{i,j}^{n+1,(2)})) \]

For convex equations of state (such as the one we are using), \(a^n\) is found at \(\rho_{\text{min}} := \rho^\dagger\), which will not deviate much from \(\rho_0\). Thus we have

\[a^n := p'(\rho^\dagger) = \gamma \rho_0^{\gamma-1} + \gamma (\gamma - 1) \varepsilon^2 \rho_0^{\gamma-2} (\rho^\dagger - \rho_0) = \gamma \rho_0^{\gamma-1} + O(\varepsilon^4), \]

as \(\rho^\dagger - \rho = O(\varepsilon^2)\). Therefore,

\[\frac{1}{\varepsilon^2} ((\rho_0)^\gamma + \gamma \varepsilon^2 (\rho_0^{\gamma} - 1) \rho_{i,j}^{n+1,(2)} - a^n (\rho_0 + \varepsilon^2 \rho_{i,j}^{n+1,(2)})) = \frac{1}{\varepsilon^2} (1 - \gamma) \rho_0^{\gamma-1}. \]

Thus the explicit pressure becomes a constant in the low Mach number limit. In principle we can also subtract a constant derived from the density (such as the average or minimum density) to ensure that this constant background pressure does not become too large, but in practice it does not really matter because this pressure term is only seen as a derivative, so the background constant value does not matter. Thus we have found that,
to $O(1)$, the modified pressure in the slow system (3.2) becomes constant in space in the low Mach number limit.

For the momentum equations, we then have at $O(1)$

$$\frac{u_{i,j}^{n+1,0} - u_{i,j}^{n,0}}{\Delta t} + \nabla \left( \frac{3}{2} u_{i,j}^{n+1,0} \otimes u_{i,j}^{n,0} - \frac{1}{2} u_{i,j}^{n-1,0} \otimes u_{i,j}^{n+1,0} \right)$$

$$+ \nabla_0 \frac{1}{2} (p_{i,j}^{n+1,2} + p_{i,j}^{n,2}) = \frac{1}{Re} \Delta_0 \left( \frac{3}{2} u_{i,j}^{n+1,0} - \frac{1}{2} u_{i,j}^{n,0} \right)$$

This is an equivalent (conservative) formulation of the incompressible momentum equation.

We note that when the second-order formulation (3.4)–(3.5) is used, the implicit terms contribute no numerical dissipation to the system, i.e. the only numerical dissipation is the ($O(1)$) contribution from the explicit, slow flux terms. We sketch the proof of this in one dimension; the proof for higher dimensions is similar.

The terms of the semi-implicit formulation from the fast system are

$$\frac{\rho_j^{n+1} - \rho_j^n}{\Delta t} + (1 - \alpha) \frac{1}{2} \left( \frac{(pu)_j^{n+1} - (pu)_{j-1}^{n+1}}{\Delta x} + \frac{(pu)_j^{n} - (pu)_{j-1}^{n}}{\Delta x} \right) = 0 \quad (3.25)$$

$$\frac{(pu)_j^{n+1} - (pu)_j^n}{\Delta t} + \frac{a}{\varepsilon^2} \frac{1}{2} \left( \frac{\rho_j^{n+1} - \rho_j^{n}}{\Delta x} + \frac{\rho_j^{n+1} - \rho_j^{n}}{\Delta x} \right) = 0. \quad (3.26)$$

Note: for $\varepsilon << 1$ we can assume that $a$ is constant to leading order. We want to make sure that numerical dissipation does not dominate when $\varepsilon$ is small.

Define the discrete energy functional

$$\mathcal{L}^{n+1} = \sum_{j=1}^{N} \left( \frac{\rho_j^{n+1} - \rho_j^n}{\Delta t} \right)^2 + (1 - \alpha) a \frac{1}{\varepsilon^2} \left( \frac{\rho_j^{n+1} - \rho_j^{n}}{2\Delta x} + \frac{\rho_j^{n+1} - \rho_j^{n}}{2\Delta x} \right)^2.$$

By taking the time difference of (3.25) and the divergence of (3.26), summing over the domain, and using summation by parts, one can derive that

$$\mathcal{L}^{n+1} - \mathcal{L}^{n} = 0,$$
which implies that the implicit terms introduce no numerical dissipation into the solution.

In summary, the limiting incompressible scheme is

\[ \rho_{i,j}^{(0)} = \rho_0 \]  
\[ \nabla_0 \cdot u_{i,j}^{(0)} = 0 \]

\[ -\Delta_2 p^{(2)} = -\nabla_0 \cdot \nabla \cdot \left( \frac{3}{2}(\rho u \otimes u)^{n,(0)} - \frac{1}{2}(\rho u \otimes u)^{n-1,(0)} \right) \]

\[ + \frac{1}{Re} \nabla_0 \cdot \Delta_0 \left( \frac{3}{2}u^{n,(0)} + \frac{1}{2}u^{n-1,(0)} \right) \]  

\[ \frac{u_{i,j}^{n+1,(0)} - u_{i,j}^{n,(0)}}{\Delta t} + \nabla \left( \frac{3}{2}u_{i,j}^{n,(0)} \otimes u_{i,j}^{n,(0)} - \frac{1}{2}u_{i,j}^{n-1,(0)} \otimes u_{i,j}^{n-1,(0)} \right) \]

\[ + \nabla_0 \frac{1}{2}(p_{i,j}^{n+1,(2)} + p_{i,j}^{n,(2)}) = \frac{1}{Re} \Delta_0 \left( \frac{3}{2}u_{i,j}^{n,(0)} - \frac{1}{2}u_{i,j}^{n,(0)} \right) . \]  

This is a second order version of a projection type method [8, 67].

### 3.7 The fast solver for \( \rho^{n+1} \)

We can efficiently solve the elliptic equation (3.8) using fast Fourier transform (FFT) based solvers. Here we will sketch the basic idea for homogeneous Neumann boundary conditions; for further details on the development of these methods see Swarztrauber [65], and for further information on using FFT techniques on different gridding systems and boundary conditions, see Bradford [3].

The basic idea, demonstrated here in one dimension, is to expand the solution in trigonometric functions that are consistent with the boundary conditions. The expansion derived below is for a cell-centered grid with homogeneous Neumann boundary
conditions, but it is relatively straighforward to find an expansion for typical boundaries. For a system with \( N \) grid points placed at \( x_{j+1/2} = \frac{1}{2}\Delta x + j/N \) on \((0, 1)\), we expand using \( \cos(\pi k(j - 1/2)/N) \), \( k = 0, \ldots, N - 1 \). Therefore, we assume \( \rho_{j}^{n+1} = \sum_{k=0}^{N-1} \hat{\rho}_{k} \cos(\pi k(j - 1/2)/N) \). Sticking this into the elliptic equation (3.8) for \( \rho_{j}^{n+1} \) gives

\[
\hat{\rho}_{k} \cos(\pi k(j - 1/2)/N)(\Delta x^2 + -4 \sin^2(\pi k/N)) = \Delta x^2 \hat{\phi}_{k}.
\]

Here, \( \hat{\phi}_{k} \) is the transformed right-hand side in the expansion functions chosen. This gives a simple system for solving \( \hat{\rho}_{k} \), which are then transformed back onto the grid. Determination of the coefficients \( \hat{\phi}_{k} \) is done through the use of a modified Fourier transform [3], as is the inverse transform back to the physical grid. The boundary conditions are automatically built into this framework by the choice of basis function.

### 3.7.1 Computational cost of the method

As the method combines both an incompressible and compressible solver, it simply adds the computational cost of the two of them together. Most of the computation is performed in the elliptic solve in the incompressible solver. In two dimensions, this transform only needs to be done along one grid dimension. The resulting system will be tridiagonal, and can thus be solved in \( O(N) \) steps. Therefore, the overall computational cost is \( O(N^2 \log N) \), which scales exactly the same as an incompressible solver. In compressible (\( \varepsilon = O(1) \)) regimes, it is more expensive to solve than just a simple hyperbolic solver. However, there is a small advantage as the implicit solve subtracts the slowest acoustic wave speed from all of the acoustic waves in the system, allowing for a larger timestep from the CFL condition.
3.7.2 The CFL condition

Following [42], the CFL condition for a first order in time central scheme of this form is

$$\frac{\Delta t \lambda}{\Delta x} \leq \frac{1}{8}.$$  

This follows from the fact that there are 8 terms that have to be evaluated in the flux calculation of each timestep.

For our second order Adams-Bashforth-Crank-Nicolson scheme, we have a further restriction of

$$\frac{\Delta t \lambda}{\Delta x} \leq \frac{1}{16}.  \frac{1}{2}.$$  

We now have 16 terms involved in the evaluation of the flux terms, and the additional $\frac{1}{2}$ factor is a restriction from the Adams-Bashforth stability region. While this CFL constant is restrictive, it is important to notice that it is independent of $\varepsilon$. In fact, as we will see below from the numerical results, the method performs best in the low Mach number limit when we take $\alpha = O(\varepsilon^2)$. In this case the wave speeds $\lambda$ are given by $\lambda = u, u \pm \sqrt{u^2}$ and we have the CFL condition

$$\Delta t \leq \frac{1}{64} \max |u|\Delta x.$$  

3.8 Numerical results

3.8.1 Compressible flow examples ($\varepsilon = O(1)$)

Experimentation with the artificial splitting parameter $\alpha$ in compressible regimes (i.e. $\varepsilon = O(1)$) showed that there is little effect on the solution unless $\alpha$ is chosen close to 0. In the compressible examples below, we show results for $\alpha = 0.5$.  

First, we will demonstrate the method in a compressible regime, i.e., where the Mach number is $O(1)$. We start with a 1-d Riemann problem with the initial condition

$$
\rho(x, 0) = \begin{cases} 
3.0 & x < 1/2 \\
1.0 & x \geq 1/2 
\end{cases} \quad u(x, 0) = 0
$$

We discretize this with 100 points and set $\varepsilon = 1, \gamma = 1.4, \theta = 1$, so we are in the fully compressible regime and are using a spatially second order scheme (using $\theta = 1$ in the minmod limiter). The results are found in Fig. 7.

Next, we test a strong shock to check that the scheme captures the correct wave speed. Inspired by the example from [39], we take the initial data

$$
\rho(x, 0) = \begin{cases} 
10.0 & x < 1/2 \\
20.0 & x \geq 1/2 
\end{cases} \quad u(x, 0) = \begin{cases} 
2000.0 & x < 1/2 \\
0.0 & x \geq 1/2 
\end{cases}
$$

We use 500 points and set $\varepsilon = 1, \gamma = 1.4, \theta = 1$, so we are again in the fully compressible regime. The results are given in Fig. 9.

We also run a test on a 2D shock problem inspired by the initial data in [47] for the full Euler equations. Here we take the initial data

$$
\rho(x, y, 0) = \begin{cases} 
0.5323 & x < 1/2, y \geq 1/2, \\
0.138 & x < 1/2, y < 1/2, \\
1.05323 & x \geq 1/2, y < 1/2 
\end{cases} \quad u(x, y, 0) = \begin{cases} 
1.206 & x < 1/2, y \geq 0, \\
1.206 & x < 1/2, y < 1/2, \\
0 & x \geq 1/2, y \geq 1/2 
\end{cases} \quad v(x, y, 0) = \begin{cases} 
0 & x < 1/2, y \geq 0, \\
1.206 & x < 1/2, y < 1/2, \\
1.206 & x \geq 1/2, y \geq 1/2 
\end{cases}
$$
Figure 7: 1D Riemann problem with initial data (3.31). $\varepsilon = 1$, $\gamma = 1.4$, $\alpha = 0.1$ and 0.5. The solid line is the true solution.
Figure 8: 1D Riemann problem with initial data (3.31). $\varepsilon = 1$, $\gamma = 1.4$, $\alpha = 0.5$ and $0.75$. The solid line is the true solution.
This initial data results in four shock waves, as seen in figure 3.8.1. As in the 1D case, we set $\varepsilon = 1$, $\gamma = 1.4$, and we take 50 points in each direction. The results are given in Fig. 3.8.1.

### 3.8.2 Low Mach number limit examples ($\varepsilon \ll 1$)

In all of the examples above, numerical experimentation has revealed that choosing $\alpha = \varepsilon^2$ provides good results in small $\varepsilon$ regimes.

First, we test a simple problem with periodic boundary conditions, to divorce the AP property from any boundary peculiarities. This example was used in [14]. The initial
Figure 10: Density plots of 2D Riemann problem with initial data (3.33). $\varepsilon = 1, \gamma = 1.4, \alpha = 0.01$ and 0.5.
Figure 11: Density plots of 2D Riemann problem with initial data (3.33). \( \varepsilon = 1, \gamma = 1.4, \alpha = 0.75 \) and 0.99.
conditions and constant $\gamma$ are

$$\begin{align*}
\rho(0, x, y) &= 1 + \varepsilon^2 \sin^2(2\pi(x + y)) \\
u(0, x, y) &= \sin(2\pi(x - y)) \\
v(0, x, y) &= \sin(2\pi(x - y)) \\
\gamma &= 2
\end{align*}$$

The initial velocity field is divergence free, and the density field is constant at leading order in the incompressible limit. We fix the Reynolds number at $Re = 100$, the spatial step at $\Delta x = 1/32$, and the temporal step at $2.5 \times 10^{-4}$, and look at $\varepsilon$ values of $0.1$ and $10^{-4}$. Figure 3.8.2 compares the solution given by the AP scheme at $T = 1$ to a highly resolved solution. In both cases we see that there is little error between the two, especially in the $\varepsilon = 10^{-4}$ case, where $\Delta x$ grossly underresolves $\varepsilon$.

Next, we test a similar case that was used in [9]. This is the so-called "vortex in a box", given by

$$\begin{align*}
\rho(0, x, y) &= 1 - \frac{\varepsilon^2}{2} \tanh(y - 1/2) \\
u(0, x, y) &= 2 \sin^2(\pi x) \sin(\pi y) \cos(\pi y) \\
v(0, x, y) &= -2 \sin(\pi x) \cos(\pi x) \sin^2(y) \\
\gamma &= 1.4
\end{align*}$$

The boundary condition is now a no-slip boundary condition ($u = 0$ on $\partial \Omega$). We again fix the Reynolds number at $Re = 100$, the spatial step at $\Delta x = 1/64$, and the temporal step at $2.5 \times 10^{-4}$, and drive $\varepsilon$ to zero. The solution at $T = 0.125$ using the AP scheme is compared to a highly resolved solution in Figure 3.8.2.

Next, we test an problem with non-zero velocity boundary conditions. We examine
Figure 12: Results for periodic flow test case (3.34). Left column: $\varepsilon = 0.1$. Right column: $\varepsilon = 10^{-4}$. Top row: stream function of solution with $\Delta x = 1/16$. Middle row: error in $u$. Bottom row: error in $v$
Figure 13: Results for Vortex in a box flow (3.35) at $T = 0.125$. Left column: $\varepsilon = 0.1$. Right column: $\varepsilon = 10^{-4}$. Top row: stream function of solution with $\Delta x = 1/16$. Middle row: error in $u$. Bottom row: error in $v$. 

the case of flow over a backward facing step, as found in [51]. The computational domain for this problem is \( \Omega = [0, L] \times [-0.5, 0.5] \). A no-flow boundary condition \((u, v) \cdot \hat{n} = 0\) is given for the step \((x = 0, -0.5 \leq y \leq 0)\) and the top and bottom walls. The velocity in the left, inflow boundary is given by \((u, v) = (12y(1 - 2y), 0)\) on \((x = 0, 0 \leq y \leq 0.5)\) and the outflow velocity is given as \((u, v) = (-3y^2 + 3/4, 0)\), \((x = L, -0.5 \leq y \leq 0.5)\) These boundary conditions are slowly ramped up from time 0 to time 1 by the function \(\frac{1}{2}(1 - \cos(\pi t))\). A Neumann boundary condition for the density (and thus the pressure) is enforced on all of the boundaries.

In Figure 14, we compare the AP scheme solution for \(\varepsilon = 0.01\) and a Reynolds number of 100 with the incompressible solution computed by Liu et al in [51]. We take \(\Delta x = 1/16, \Delta t = 9.765 \times 10^{-4}\) and a channel of length \(L = 8\).

![Figure 14: Left: Contour plot of streamfunction for backwards step flow with \(\varepsilon = 0.01, \Delta x = 1/N = 1/16 = 0.0625, \Delta t = 9.765 \times 10^{-4}, Re = 100, \gamma = 1.4, \alpha = \varepsilon^2\) and \(\theta = 1\) at \(T = 20\). Right: Contour plot of streamfunction from [51]. The AP scheme captures the reattachment point of the circulation region.](image)

In comparing the solution in Figure 14, we see that the reattachment point of the circulation region behind the step matches with the results found in [51].
Chapter 4

Concluding remarks and future work

4.1 Diffusive transport

We have presented a computational study of an operator splitting method for the $P_1$ equations. Our results show that small modifications are sometimes needed to suppress numerical oscillations in the method. In some cases, we have found that discretization of diffusion terms on a black-red stencil can introduce binary oscillations into a numerical solution. In other cases, there is simply not enough numerical dissipation to suppress oscillations.

The examples presented here are not exhaustive. Indeed, a more complete analysis of the splitting scheme would be needed to conclude that the scheme is completely free of oscillations. For example, one could attempt to establish some notion of a total variation diminishing (TVD) property or, more generally, a positivity property in the sense of Lax and Liu [46].

While oscillations of an isolated system may be stable, they may introduce instabilities when coupled to large multi-physics codes. For example, it is important to ensure that a numerical scheme does not produce false minima or maxima. Moving forward, it
seems that a more robust approach is to use modified systems like (2.24). This approach has seen some initial success in [31, 30], but the presence of non-conservative products in these systems presents a major obstacle. In this respect, operator splitting schemes may continue to provide insight into how to appropriately discretize modified systems.

4.2 Low Mach number limit

We proposed a new numerical method for solution of the compressible isentropic Euler (and Navier-Stokes) equations that is stable and accurate for any Mach number. The method is based on a hyperbolic splitting that splits the compressible Euler equations into a slowly moving nonlinear conservative hyperbolic system and a fast moving stiff linear acoustic system. The slow part is suitable for modern shock capturing methods, while the stiff acoustic system is solved implicitly with a fast Poisson solver as in a typical projection type method for incompressible flows. This scheme allows the use of time step and space mesh size independent of the Mach number. When the Mach number goes to zero it effectively becomes a second order projection type method for incompressible flows, a property called asymptotic-preserving. Numerical results in one and two space dimensions demonstrate that the scheme is adequate in both compressible and incompressible regimes, capturing shocks with a high resolution in the compressible regime and the incompressible features for small Mach numbers.

In future work, we will seek to extend this approach to the full Euler (and Navier-Stokes) equations. A similar splitting can be made that results in a linear implicit system, and in this case the hyperbolicity requirement enforces bounds on the choices of the splitting parameters (i.e., the $\alpha$ we saw in the isentropic case). This work is a
stepping stone to the full equations. One important thing to note is that the limiting incompressible equations for this system allow for variable density, as opposed to constant density in the isentropic case. This could prove to be an important development in the simulation of two-phase flows such as bubbles in water, mostly incompressible flows with regions of high compressibility such as underwater explosions, or atmospheric flows.
Appendix A

Appendix

The purpose of this appendix is to provide complete details of the schemes used to produce the results presented in the paper. As in the main text, all state variables and the cross-section $\sigma$ are adorned with the subscript $j$ to denote an average value over a computational cell $I_j$ of width $\Delta x$. The parameters $\gamma_j$, $\beta_j$, and $\kappa_j$ are composite parameters that depend on $\sigma_j$ via the relations given in (2.11) and (2.18). Supscripts of the form $j \pm 1/2$ denote pointwise values, or approximations thereof, at cell edges.

A.1 A few basic tools

Many of the schemes in the paper use the following tools.

**Discretization of Linear Flux Gradients.** Let $Au_x$ be a linear flux in a hyperbolic system, with $A$ a constant matrix. A finite volume, upwind discretization of $Au_x$ is given by

\[
\int_{x_{j-1/2}}^{x_{j+1/2}} Au_x \, dx = A(u_{j+1/2} - u_{j-1/2})
\]

\[
\simeq \frac{1}{2} A (u_{j+1/2,\ell} + u_{j+1/2,r}) + \frac{1}{2} |A| (u_{j+1/2,\ell} - u_{j+1/2,r}) . \tag{A.1}
\]

Here the $t$-dependence of $u$ has been suppressed and the matrix $|A| := R|\Lambda|R^{-1}$ is calculated using the eigenvectors and eigenvalues in the diagonal decomposition $A = \ldots$
The right and left edge values of $u$ in (A.1) are given by
\begin{align*}
u_{j+1/2, \ell} &= R w_{j+1/2, \ell} \quad \text{and} \quad \nu_{j+1/2, r} = R w_{j+1/2, r}, \quad (A.2)
\end{align*}
and the characteristic edge values are determined by linear reconstructions on adjacent cells; that is,
\begin{align*}
w_{j+1/2, \ell} &= w_j + \frac{h}{2} w_j' \quad \text{and} \quad w_{j+1/2, r} = w_{j+1} - \frac{h}{2} w_{j+1}', \quad (A.3)
\end{align*}
where $w_j := R^{-1} u_j$ and the slopes $w_j'$ approximate derivatives in each cell. We use three different approximations for the slope: (i) zero slope; (ii) central difference; and (iii) the double minmod limiter:
\begin{align*}
w_j' &= \frac{1}{h} \minmod \left( 2(w_{j+1} - w_j), \frac{1}{2}(w_{j+1} - w_{j-1}), 2(w_j - w_{j-1}) \right), \quad (A.4)
\end{align*}
which is applied to vectors component-wise.

**Second-Order Time Integration.** For second-order time integration, we use two methods: a TVD Runge-Kutta method and an IMEX method. For an ODE of the form
\begin{align*}
\dot{y} = L(y), \quad (A.5)
\end{align*}
with $L$ linear, a second-order Runge Kutta scheme is [23, 64]
\begin{align*}
y^{(1)} &= y^n + \Delta t L(y^n), \quad (A.6a) \\
y^{n+1} &= 0.5[y^n + y^{(1)} + \Delta t L(y^{(1)})]. \quad (A.6b)
\end{align*}
Given $L = L_1 + L_2$, the following IMEX scheme [62, 61] is used when an explicit update of $L_1$ and an implicit update of $L_2$ is desired:

\[
y^{(1)} = y^n + \Delta t \eta L_2(y^{(1)}) , \quad (A.7a)
\]

\[
y^{(2)} = y^n + \Delta t [L_1(y^{(1)}) + (1 - 2\eta) L_2(y^{(1)} + \eta L_2(y^{(2)})] , \quad (A.7b)
\]

\[
y^{n+1} = y^n + \Delta t \frac{1}{2} [L_1(y^{(1)}) + L_1(y^{(2)}) + L_2(y^{(1)}) + L_2(y^{(2)})] . \quad (A.7c)
\]

where

\[
\eta = 1 - \frac{1}{\sqrt{2}} . \quad (A.8)
\]

### A.2 The schemes

We now present the details for the upwind and splitting schemes used to produce results in the main paper. We also write out the details for discretizing the regularized system (2.27).

**Upwind scheme.** Let \( u = (\rho, m)^T \). Then the upwind discretization of (1.3) is given by

\[
d_t u_j + A_\varepsilon \frac{(u_{j+1/2} - u_{j-1/2})}{\Delta x} = -\sigma_j \varepsilon u_j , \quad (A.9)
\]

where

\[
A_\varepsilon = \begin{pmatrix}
1 & 0 \\
0 & \frac{1}{\varepsilon^2}
\end{pmatrix} \quad \text{and} \quad Q_\varepsilon = \begin{pmatrix}
0 & 0 \\
0 & \frac{1}{\varepsilon^2}
\end{pmatrix} . \quad (A.10)
\]

Note that the time dependence of \( u \) has been suppressed. The evaluation of the fluxes \( A_\varepsilon u_{j\pm 1/2} \) is given by (A.1) with \( A = A_\varepsilon \). For the test cases in Figure 1, a first-order spatial discretization is accomplished by setting the slope in (A.3) to zero. This is exactly
Meanwhile, the second order version uses the double minmod limiter (A.4). The time integration of (A.9) is accomplished at first order with the forward Euler method, and at second order by an IMEX scheme of the form given in (A.7) that treats the flux terms explicitly and the source term implicitly. In both cases, the time step is determined by (2.14) with $C = 0.1$. The second-order reference solutions in Figures 3, 5, and 6 are all computed using in the same way, with 200 computational cells.

The highly resolved reference solutions in Figures 1, 2, and 4 are also computed with an upwind method, but with 20000 cells. The exact implementation of the method depends on the value $\varepsilon$. For $\varepsilon = 0.8$, the spatial reconstruction uses the double minmod limiter (A.4) and the Runge-Kutta method (A.6a) with a time step $\Delta t = 0.1 \varepsilon \sqrt{3} \Delta x$. Thus the scheme is fully second order. For $\varepsilon = 10^{-4}$ the solution is presumably smooth, so the spatial derivative in (A.3) is computed using a central difference—with no limiting—in order to obtain second-order spatial accuracy. However, only a first order time integrator is used, with the source term treated implicitly and the flux term explicitly. Here we use a time step $\Delta t = 0.8 \varepsilon \sqrt{3} \Delta x$, which is very restrictive when $\varepsilon$ is so small. In particular $\Delta t \sim \Delta x^2$, so a second-order time integrator is not really needed.

**Splitting Scheme.** For the splitting scheme, the implicit step (2.9a) is accomplished by using (2.10) for both first and second-order as

$$d_t u_j + A_1 \frac{(u_{j+1/2} - u_{j-1/2})}{dx} = 0,$$  \quad (A.11)

where $u_j = (\rho_j^*, m_j^*)$ (see (2.10)) and

$$A_1 = A_{2|\varepsilon=1} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{pmatrix}.$$  \quad (A.12)
As in the upwind discretization above, the evaluation of the fluxes \( A u_j + \frac{1}{2} \) is determined by (A.1), except this time with \( A = A_1 \). At first order, the slope approximation in (A.3) is set to zero; at second order, the double minmod limiter (A.4) is used. Time integration at first order involves alternating evaluation of the two steps in (2.10) and (2.12). To get second order in time, the IMEX scheme (A.7) is used, where \( L_1 \) represents the non-stiff step and \( L_2 \) represents the stiff step. The time step is set according to (2.14) with \( C = 0.1 \).

The first order version of the splitting scheme is given explicitly in (2.17). First-order modifications to this scheme, which were introduced to suppress the non-physical oscillations, are described explicitly in Section 3. Rather than directly implement similar modifications to the second-order splitting method, we instead work with the regularized system (2.27).

**Regularized Equations.** The numerical scheme for regularized system (2.27) (which was used to generate the profile in Figures 6(c) and 6(d)) can be formulated as a generalization of the upwind method for spatially varying fluxes. However, it is simpler to write out the scheme in terms of the components \( \rho \) and \( q \). In semi-discrete form, the scheme is

\[
\begin{align*}
\partial_t \rho_j + \frac{(\gamma \beta q)_{j+1/2} - (\gamma \beta q)_{j-1/2}}{\Delta x} &= \frac{1}{3} \frac{\kappa_{j+1/2} \rho_{j+1}^n - \rho_j^n - \kappa_{j-1/2} \rho_j^n}{\Delta x^2}, \quad (A.13a) \\
\partial_t q_j + \frac{1}{3 \varepsilon^2} \frac{\rho_{j+1/2} - \rho_{j-1/2}}{\Delta x} &= -\frac{\gamma_j \sigma_j}{\varepsilon^2} q_j. \quad (A.13b)
\end{align*}
\]

where \( \kappa_{j+1/2} \) is the harmonic average of surrounding cell values:

\[
\kappa_{j+1/2} = \frac{2 \kappa_j \kappa_{j+1}}{\kappa_j + \kappa_{j+1}}. \quad (A.14)
\]
To compute edge values in (A.13), we define the diagonal variables

$$w^\pm := \rho \pm \sqrt{3\gamma\beta\varepsilon q},$$

so that

$$\rho_{j+1/2} = \frac{1}{2}(w^+_{j+1/2} + w^-_{j+1/2}) \quad \text{and} \quad (\gamma\beta q)_{j+1/2} = \frac{1}{2}\varepsilon \sqrt{3(\gamma\beta_{j+1/2})(w^+_{j+1/2} - w^-_{j+1/2})}.$$ 

(A.16)

Because of the spatial variation in $\gamma$ and $\beta$, two upwinding steps are needed. First the edge values $w^\pm_{j+1/2}$ are determined in standard fashion:

$$w^+_{j+1/2} = w^+_{j+1/2,l} \quad \text{and} \quad w^-_{j+1/2} = w^-_{j+1/2,r},$$

(A.17)

where the left and right values are based on linear reconstructions:

$$w^+_{j+1/2,l} = w^+_j + \frac{h}{2}s^+_j \quad \text{and} \quad w^-_{j+1/2,r} = w^-_{j+1} - \frac{h}{2}s^-_j,$$

(A.18)

and the slopes $s^\pm$ are approximated with the double minmod limiter (A.4). Next, the products $(\sqrt{\gamma\beta w^\pm})_{j+1/2}$ in (A.16) are computed as

$$(\sqrt{\gamma\beta w^+})_{j+1/2} = (\gamma\beta)_{j+1/2,l} w^+_{j+1/2,l}$$

(A.19)

$$(\sqrt{\gamma\beta w^-})_{j+1/2} = (\gamma\beta)_{j+1/2,r} w^-_{j+1/2,r}$$

(A.20)

where $(\gamma\beta)_{j+1/2,l}$ and $(\gamma\beta)_{j+1/2,r}$ are calculated with the same type of linear reconstruction as in (A.18), again with the double minmod limiter.

For constant cross-section, the approach described here reduces to a standard second-order (in space) upwind method. We use this method with a forward Euler time integrator for (A.13) to compute the solutions presented in Figures 6(c) and (d). Thus the scheme is second order in space, but only first order in time. The time step is set by (2.14) with $C = 0.1$. 
Bibliography


