Contributions to the Development of Entropy-Stable Schemes for Compressible Flows

by

Ayoub Gouasmi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Aerospace Engineering)
of the University of Michigan
2020

Doctoral Committee:

Associate Professor Karthik Duraisamy, Chair
Professor Smadar Karni
Dr. Scott M. Murman, NASA Ames Research Center
Professor Philip L. Roe
Professor Eitan Tadmor, University of Maryland

Ayoub Gouasmi

gouasmia@umich.edu

ORCID iD: 0000-0003-0647-6723

To my parents Abderrahmane and Saïda, my brother Mohamed and my sisters Imaine, Rayane, Roumaïssa and Kaouthar.

ACKNOWLEDGEMENTS

First off, I would like to thank my colleagues in the Aerospace Engineering department and all the friends I made in Ann Arbor for making the past four years the most enriching ones of my life. Mentioning them all would take eons, but I feel compelled to give a few shout-outs in the very least. I am grateful to my labmates, past and present, in particular Anand, Eric, Nicholas, Vishal and Christopher for being fun, engaging and inspiring people. I would have gotten nowhere without you guys. I am grateful to my close friend Yash, for his unwavering support and trademark composure during complicated times. My gratitude also extends to longtime friends such as Florian, Félix, Olivier and Meik who make me miss the old days and my home country every time I visit them.

This amazing journey would have never taken place without my advisor Professor Karthik Duraisamy. I owe him for introducing me to the fascinating field of CFD and for giving me the opportunity to carry out PhD research in an amazingly diverse group of people. I am grateful for his trust, for the tremendous amount of freedom I have been granted, for his tireless (and sometimes tiring) enthusiasm at every little step and finally for his tolerance and patience towards hot-blooded researchers in the making. Our relationship took all sorts of turns over the past three years, but all in all I reckon that my research work would not have been better anywhere else in the world.

I feel extremely lucky to have landed where Professor Philip Roe was working and teaching. I am thankful for all the time he spent answering my questions whenever I would barge into his office (“random is better” is what he replied when I suggested a regular time slot to not perturb his schedule). I am also thankful for the honesty and humility of his answers. I think researchers in general do not realize that sometimes, simply saying “I don’t know” will open up more doors than improvising an answer. Furthermore, I think Professor Roe exemplifies the notion that it is crucial to cultivate your own ways of thinking while respecting and acknowledging those of your fellows.

Talking about luck, I must say that my encounter with Dr. Scott Murman, which I owe to my advisor, was nothing short of a miracle. It came right at a time when I was struggling to come to terms with the reality of the research world and seriously considering a different path in life. I am grateful to him for constantly throwing nasty problems, none of which I solved, at me and beautifully tearing my ego to shreds. Scott strongly influenced my thesis and, for better or for worse, strongly impacted the way I now look at research. I hope that I will be able to give back as much as he contributed to my research and education.

My fortune does not stop there. I have also had the utmost privilege to learn and interact with the person who laid the foundations supporting my work. Professor Eitan Tadmor helped me realize that developing consistent numerical algorithms goes hand in hand with understanding the structure of the equations we work with. I am deeply grateful for him sharing historical perspectives and walking me through some of his proofs. What struck me the most about him, besides his terrifyingly sharp mathematical skill and intuition, is his kindness and consideration for those who just began their adventure. It reassures me to no end to know that there are and have been people like him in this world.

I would like to address special thanks to Professor Smadar Karni for serving as the cognate member of the PhD committee. I am also grateful for her many contributions to the field of CFD, some of which are relevant to the thesis work and could have been followed-up upon.

As of matter of course, this thesis would not have been possible without the financial support of several entities. This thesis work was supported by the Air Force Office of Scientific Research (AFOSR) through grant number FA9550-16-1-030 (Tech. monitor: Fariba Fahroo). Computational resources were also provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. I also extend my appreciation to the NASA Space Technology Mission Directorate (STMD) through the Entry Systems Modeling (ESM) project for supporting my visits to NASA Ames Research Center over the past summers.

On a non-academic but nonetheless important note, I would like to thank Lionel Messi from FC Barcelona for contributing to 50% of my happiness during the PhD, my friends Nicholas and Yash for helping me buy my first car, my friend Raphaël for making me run my first marathon, and coach Casey Bantle from Ann Arbor FC for making sure that this great city always gets its fill of football[1] opportunities.

As I will soon be stripped of my title of student, I would like to express my profound gratitude to all my teachers, from the Oberlin kindergarten in Strasbourg all the way to the University of Michigan in Ann Arbor, who carry out one of the most important duties in the world. I feel particularly indebted to my middle school and high school mathematics teachers Mr. Nicolas Méziane and Mr. Yoann Letang, respectively. Mr. Méziane got me hooked on mathematics and made me proud of this vocation. In the same vein, Mr. Letang's thorough and ambitious teaching, together with his constant encouragements, built up the conviction that hard work and mathematics can get you anywhere you want. I am also grateful to my physics teachers from *classes préparatoires* Mr. Sylvain Soufflet and Mr. Frédéric Paviet-Salomon. Their calm and composure in approaching both dauntingly complex problems and

[1] Americans call it soccer

the sight of a student who would *occasionally* and *unintentionally* make up for his lack of sleep during class will never be forgotten.

Ultimately, I am of the son of two hard-working and loving parents, who always emphasized the importance of education and integrity. They made every possible sacrifice so that I would never be in need of anything while pursuing my passions. My mother bought me my first pair of football cleats and my father would drive me to football practice at 9 am on Sundays, only a couple of hours after the end of his second job's night shift. I feel guilty about my younger self diligently and shamelessly waking him up every time, all the more so considering that I did not manage to play professionally in the end. Broken dreams aside, I have five strong siblings who always stood their grounds and fought for what they believed is right. I am especially admiring of my older sister Kaouthar, who will relentlessly keep moving forward in the face of all upsets. To sum it up in a few words, *if you don't fight, you can't win* [2].

[2] *Shingeki No Kiojin*, chapter 106.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

ABSTRACT

Entropy-Stable (ES) schemes have gathered considerable attention over the last decade, especially in the context of under-resolved simulations of compressible turbulent flows, where achieving both high-order accuracy and robustness is difficult. ES schemes provide stability in a nonlinear and integral sense: the total entropy of the discrete solution can be made non-decreasing, in agreement with the second principle of thermodynamics. Additionally, the amount of entropy produced by the scheme is known and can be modified, making room for analysis and improvements. This thesis delves into some of the challenges currently limiting their use in practice.

The current state-of-the-art solves the compressible Navier-Stokes equations for a single-component perfect gas in chemical and thermal equilibrium. This model is inappropriate in aerospace engineering applications such as hypersonics and combustion, which typically involve chemically reacting gas mixtures far from equilibrium. As a first step towards enabling their use for these applications, we formulated ES schemes for the multicomponent compressible Euler equations. Special care had to be taken as we found out that the theoretical foundations of ES schemes begin to crumble in the limit of vanishing partial densities.

The realization that ES schemes can only go as far as their theory led us to review some of it. A fundamental result supporting the development of limiting strategies for high-order methods is the minimum entropy principle for the compressible Euler equations. It states that the specific entropy of the physically relevant weak solution does not decrease. We proved that the same result holds for the specific entropy of the gas mixture in the multicomponent case.

While entropy-stability is a valuable property, it does not imply a well-behaved solution. One must recall that the second principle is a prescription on the correct behavior of a system at the *global* level only. To better understand how ES schemes may or may not improve the quality of the numerical solution, we revisited two classical problems encountered in the development of shock-capturing techniques.

First, we studied the receding flow problem, which is a simple setup used to study the anomalous temperature rise, termed "overheating", typically observed in shock reflection and shock interaction calculations. Previous studies showed that the anomaly can be cured if conservation of entropy is enforced, but at the considerable price of total energy conservation. Entropy-Conservative (EC) schemes, a particular instance of ES schemes, can achieve both simultaneously and therefore appeared as a potential solution. We showed that while the overheating is correlated to entropy production, entropy conservation does not necessarily prevent it.

Second, we studied the behavior of ES schemes in the low Mach number regime, where shock-capturing schemes are known to suffer from severe accuracy degradation issues. A classic remedy to this problem is the flux-preconditioning technique, which consists in modifying artificial dissipation terms to enforce consistent low Mach behavior. We showed that ES schemes suffer from the same issues and that the flux-preconditioning technique can improve their behavior without interfering with entropy-stability. Furthermore, we demonstrated analytically that these issues stem from an acoustic entropy production field which scales improperly with the Mach number, generating spatial fluctuations that are inconsistent with the equations. An important outgrowth of this effort is the discovery that skew-symmetric dissipation operators can alter the way entropy is produced or conserved locally.

CHAPTER I

Introduction

Together with the advent of High-Performance Computing, Computational Fluid Dynamics (CFD) simulations have significantly contributed to the advancement of Aerospace Engineering over the last several decades. CFD simulations offer the prospect of a more efficient alternative to the costly ground-based and in-flight experiments commonly used in Aircraft and Spacecraft Design, which to this date remain immovable. They can also provide engineers with a more detailed description of the physics of interest, enabling a more thorough investigation of the design space, making more room for creativity and innovation, and fostering fundamental research.

Implementation aspects aside, CFD comprises two fundamental components. The first component is a physical model, that is a mathematical representation of the physics of interest. In fluid mechanics and many other fields, the model typically consists of a system of Partial Differential Equations (PDE) such as the Potential flow equations, the Navier-Stokes equations or the Euler equations, which for compressible flows in one dimension write:

$$\frac{\partial}{\partial t}\begin{bmatrix}\rho\\ \rho u\\ \rho e^t\end{bmatrix}+\frac{\partial}{\partial x}\begin{bmatrix}\rho u\\ \rho u^2+p\\ (\rho e^t+p)u\end{bmatrix}=0. \tag{1.1}$$

In the above system, ρ is the density, u is the velocity, e^t is the total energy and the pressure p is determined by an equation of state. Analytical solutions to these equations are not available in practice, hence the need to develop numerical schemes (that is the second component) which can solve the chosen model with satisfying accuracy and speed. In more concrete terms, the governing equations are *discretized* in space and time using a variety of techniques such as the finite difference, finite volume and finite element methods.

There is a plethora of reasons why developing reliable numerical schemes is an extremely complex task. Many of them can be traced back to the plague of *under-resolution*. On one hand, the governing equations are continuous, i.e. they involve the flow field at every single point in the domain. On the other hand, the numerical scheme typically seeks a *discrete*[1] solution by solving a finite set of equations for the flow field at finite locations in space and time with no knowledge of the solution anywhere else. This limitation comes from cost considerations. The finer the grid, the more equations to solve and the more expensive and lengthy the simulation becomes.

For a number of discretization techniques[2], under-resolution manifests through the appearance of *discontinuities* or abrupt changes in the discrete flow field. They threaten the *robustness* of the scheme, that is its ability to always predict a flow field with finite and physical values (the velocity u for instance can be of any sign, but the density ρ should not be negative). They also challenge its *accuracy*, that is its ability to provide the best answer possible for a given resolution. Without getting into details, the importance of handling discontinuities properly can be comprehended by simply recalling that our physical models are PDEs which relate the local variations of the variables of interest in time to their local variations in space. Hence the root

[1] We acknowledge that the very idea of assuming a discontinuous flow field can be challenged [156].

[2] Continuous finite-element methods represent the flow field everywhere. Instead of discontinuities, they have to deal with oscillations or "wiggles" [2] which pose the same threats as discontinuities.

problem is not that discontinuities exist on the grid, because nothing interesting would happen otherwise. The problem is that the continuous equations do not tell us how *discrete* gradients should drive *discrete* physics.

1.1 Challenges in compressible flow simulations

1.1.1 Shock-capturing

A lot of the research effort on numerical schemes has been spent on the prediction of shock and contact waves in compressible inviscid flows. While they are not completely grasped at the moment, there is some consensus about the properties such *shock-capturing*[3] schemes should have. The leading thread is that the numerical scheme should remain physically consistent with the governing equations despite under-resolution. While shocks and contacts discontinuities are not smooth flow phenomena (to the human eye in the very least), they satisfy the Rankine-Hugoniot conditions which are obtained by leveraging the fact that the model equations describe conservation of mass, momentum and energy. One of the first requirements is therefore that the scheme should be *conservative* [21]. Another requirement stems from the observation that shocks and contacts are basically disturbances created by a moving body. These disturbances propagate in the fluid at finite speeds. This observation is consistent with the governing equations (1.1) which, through their *hyperbolic* nature, naturally describe wave motion. Hence, the scheme should be able to propagate waves accurately. The linear advection equation for a scalar field u in one-dimension:

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0, \tag{1.2}$$

which describes solutions which propagate at a constant speed a, is one of the test-bed equations to assess the wave propagation capability of a scheme[4]. It is also one

[3] *Shock-fitting* [3, 4] techniques fall outside the scope of this thesis.

[4] In one dimension that is [1].

of the systems from which some of the early foundations of CFD, such as Von Neumann analysis [7] and modified PDE analysis [6] and the Courant-Friedrichs-Lewy (CFL) condition [5] were built. The stability analysis of numerical schemes for linear systems is now well-established [8, 9], and several methods such as the Lax-Friedrichs scheme, the Lax-Wendroff scheme, the MacCormack scheme and the upwind scheme [11], which were developed in that context are now part of the foundations of modern-shock capturing techniques.

Linear stability analysis shows limits for nonlinear PDEs, where discontinuous solutions can form from smooth initial conditions and are not uniquely determined by them. Considerable progress towards nonlinear stability was achieved for nonlinear scalar PDEs, where monotone schemes [16, 17] were shown to always converge to the physically relevant solution, as a consequence of being consistent with entropy inequalities [18, 19]. Their limitation to first-order accuracy [15, 16, 17] motivated the Total-Variation-Diminishing (TVD) algorithms introduced by Harten [13] which can achieve 2nd-order accuracy and satisfy entropy inequalities. Extensions of TVD schemes from scalar PDEs to nonlinear PDE systems were obtained using exact Riemann solvers [12, 26], approximate Riemann solvers such as Roe's [10] and the flux-vector splitting technique developed by Steger [28] and Van Leer [29]. The TVD property is no longer provable, but consistency with entropy inequalities can be retained, depending on the Riemann solver [27, 22]. Other TVD-type high-resolution schemes include the MUSCL scheme of Van Leer [24], the flux-corrected transport (FCT) scheme of Boris and Book [25] and the Piecewise Parabolic Method of Colella and Woodward [23]. We refer the interested reader to the review of Yee [20] and the book of Levecque [14] for more details on the development and applications of TVD schemes.

The accuracy degradation problems of TVD schemes near extrema motivated the development of uniformly high-order schemes such as the Essentially Non-Oscillatory

(ENO) schemes of Harten *et al.* [30]. The idea is to achieve high-order accuracy by wisely choosing the reconstruction stencil, avoiding interpolation across discontinuities. Further improvements in accuracy and efficiency were brought with the class of Weighted ENO (WENO) schemes introduced by Liu *et al.* [31] and further developed by Jiang & Shu [32]. Despite the lack of stability property in the nonlinear systems case, ENO and WENO schemes have been successfully applied to a wide range of problems [33].

Another shock-capturing approach that finds its roots in the seminal work of Von Neumann and Richtmyer [34] is the Artificial Viscosity (AV) method. It consists in using centered schemes for the convective terms and adding artificial dissipation terms to stabilize the solution. The magnitude of these dissipation operators is determined by adjustable artificial viscosity coefficients. The AV approach has been further developed in the spectral context by Tadmor [36] and Guo *et al.* [37] and in the finite-difference context by Cook & Cabot [38, 39], Cook [40] and Kawai & Lele [41]. The foundations of the AV method are not as theoretical as those of monotone or TVD schemes, but their simplicity and flexibility are attractive. A different perspective on the roots of AV methods is given by Mattsson & Rider [35].

While high-order accuracy and robustness are desirable attributes of a shock-capturing scheme, they do not guarantee a well-behaved solution. There is a number of situations where the numerical solution exhibits anomalous behavior which cannot be resolved by either. To give a few examples: -carbuncles in blunt-body flows [148, 55, 54, 149]; -pressure oscillations in slowly moving shocks [150] and material interfaces in multicomponent flows [185, 184, 152]; -wall heating in shock reflection and shock interaction [167, 162]; -excessive damping of acoustic waves [151] and vortical structures [236]. Other anomalies are documented in the work of Quirk [155]. These errors can further complicate the simulation of compressible turbulent flows as shown by Johnsen *et al.* [153], Larsson & Lele [154] and Cook [152].

1.1.2 Turbulence

In addition to properly capturing sharp flow features such as shocks, another considerable obstacle that has stalled the advance of CFD over the last two decades is the simulation of turbulent flows. These are characterized by a very wide range of spatial and temporal scales which cannot be fully solved in acceptable turn around times for engineering applications. This brings back to the table the problem of under-resolution, but with the added complexity that there is no a priori knowledge about their inherent structure. Unlike shocks, they are not observable and they unfortunately do not come with their own "Rankine-Hugoniot conditions". The only thing we know with certainty is that with high-enough resolution these discontinuities would not exist[5].

1.1.2.1 High-resolution methods

The high resolution requirements posed by turbulent flows motivated the development of spectral [89, 90] and pseudo-spectral [91] methods. They enabled some of the first numerical simulations of incompressible turbulent flows by Smagorinsky [94], Fox & Lilly [95], Rogallo & Moin [97], Kim *et al.* [98], and Spalart [99], providing a valuable complement to experimental data in support of the longtime effort to grasp their physics. When the spatial and temporal scales of the flow are not completely resolved, the stability of these algorithms is threatened by the appearance of aliasing errors [96, 91, 92], which induce non-physical kinetic energy transfers between scales. Several de-aliasing approaches such as high-wavenumber filtering [93] were proposed to alleviate these issues.

Given the limitation of spectral methods to simple geometries and boundary conditions, various high-order finite difference schemes were developed. Using Fourier anal-

[5] similar statements could actually be made about shocks, whose thickness is of the order of the mean free path.

ysis, Lele [100] developed high-order compact finite difference schemes, which combined with filters can achieve resolution levels comparable to those of spectral methods. Rai & Moin [101] developed high-order upwind finite difference schemes and highlighted their robustness. Their work motivated further use for compressible turbulent configurations [102, 103]. However, these schemes were also shown to be overly dissipative in more under-resolved configurations [104, 105]. A compromise between accuracy and robustness was found in finite-difference schemes which discretize the convective terms of the governing equations in split forms [106, 107, 108, 109, 110] which can enable discrete consistency with conservation of kinetic energy in incompressible flows and conservation of entropy [113, 114] in compressible flows. Summation-by-Parts finite difference operators [79, 80] are key components of these schemes. A comprehensive review of these techniques can be found in Zang [111] and Morinishi [115].

The Discontinuous Galerkin (DG) method [117, 118, 116] is currently one of the flagship high-order discretization techniques [119] under development for turbulent flow simulations [138, 139, 122, 119]. DG methods combine the attractive resolution properties of spectral methods with the geometric flexibility of finite-element methods, which gives them an edge over finite difference methods in flow configurations which require unstructured meshes. Their compact nature also makes them more amenable to high-performance computing [122], and adaptive h/p refinement strategies [122, 119].

Despite their geometric flexibility and resolution properties, DG methods suffer from aliasing issues as well. And while many of the techniques (de-aliasing, filtering, kinetic energy preservation) used in the spectral and finite difference context have been successfully adapted to the DG framework [120, 121, 123, 124, 125], they fail to overcome the robustness issues of DG schemes for compressible turbulent flows at high Reynolds numbers. That is to say that for such flows, aliasing is not the primary cause of instability. The interested reader is referred to the recent studies of Moura *et al.* [140], Fernandez *et al.* [141, 143] and Mengaldo *et al.* [142].

1.1.2.2 Modeling and numerics

In parallel to developing high-resolution schemes, considerable efforts have been put into reconsidering the physical model itself. This is where the field of turbulence modeling begins. Its primary goal is to develop alternatives to classical models such as the Navier-Stokes equations, which in addition to providing a faithful representation of the flow physics, should account for the reality of under-resolution and prescribe some structure to it. These are generally obtained by formal decomposition of the 'exact' flow field into resolved and unresolved components and by manipulating the baseline governing equations in such a manner that new equations governing the resolved flow field components only are obtained. These equations will also ineluctably introduce some closure problem to solve. The two most established turbulence approaches in CFD are Reynolds Averaged Navier-Stokes (RANS) [126] models and Large Eddy Simulation (LES) [94] models. The former enabled the first affordable turbulent flow calculations for engineering purposes and is now routinely used in industry. The latter, enabled by more powerful computers, has been emerging as both a complement and an alternative to RANS models which show limits in predicting turbulent flows where unsteady flow features driving design-critical phenomena such as flow separation must be captured with accuracy. Going into the specifics of each approach, their developments and applications would take us far beyond the scope of this thesis. We refer the reader to the comprehensive book of Sagaut *et al.* [127].

The emphasis turbulence modeling puts on physical model development does not reduce the importance of numerical scheme development. Much to the contrary, it amplifies it, and most of the aforementioned high-resolution methods were actually developed as potential baseline schemes to “realize” these new models. However, the successful combination of both, outside the ideal but impractical spectral context [133, 134, 135], remains elusive to this date.

For the validity of the turbulence model to be properly assessed, the discretization errors introduced by the scheme need to be small enough. Quantifying these errors is a complicated task [130, 129]. Spectral analysis techniques are typically used for that purpose [129, 140, 141, 142] but the insights they provide is limited. Controlling these errors is even harder. Shock-capturing schemes are known to introduce numerical errors which make turbulence models inactive [105, 104, 128], but they cannot be easily discarded because the dissipation introduced by turbulence models alone struggles to stabilize the solution in under-resolved configurations.

Perhaps as a consequence of these complications, the view that physical models should be rethought in turbulent flow simulations does not have unanimous support among practitioners. Part of the community actually advocates a no-model or Implicit LES (ILES) approach [136, 137], where the baseline physical model is left untouched and the artificial dissipation the shock-capturing scheme introduces is viewed as an implicit turbulence model. This stance stems from past studies which showed good results without turbulence models [131, 138, 139]. There has also been work showing through modified PDE analysis [132, 136] that the equations the scheme effectively solves do have the markings of a turbulence model.

1.2 Entropy-Stable schemes

1.2.1 A brief overview

A number of systems of conservation laws imply additional conservation equations for *mathematical entropies*, namely scalar convex functions of the conserved variables. For instance, the compressible Euler equations (1.1) imply:

$$\frac{\partial}{\partial t}(-\rho s) + \frac{\partial}{\partial x}(-\rho u s) = 0, \tag{1.3}$$

where $s = \ln(p) - \gamma \ln(\rho)$ is the specific entropy. In shock calculations, another well-established guideline is that entropy should be produced across shocks. In more formal terms, this is equivalent to requiring that the numerical scheme should be consistent, in an integral sense, with the inequality:

$$\frac{\partial}{\partial t}(-\rho s) + \frac{\partial}{\partial x}(-\rho u s) < 0, \tag{1.4}$$

Building from extensive theoretical work [42, 43, 44, 45, 46, 47, 48] on the structure of such systems, Tadmor [49] introduced finite-volume discretizations which are consistent with either the conservation equation for entropy or the entropy inequality at the semi-discrete level. The scheme is termed Entropy-Conservative (EC) in the first case and Entropy-Stable (ES) in the second case. EC fluxes, which are defined by a scalar EC condition, and ES fluxes, which are obtained by combining an EC flux with appropriate dissipation operators, are the main two ingredients for this purpose.

Despite some developments in the following years, which include the high-order fully-discrete entropy-conservative schemes of LeFloch *et al.* [50] and the fully-discrete entropy-stability analysis of Tadmor [51], ES schemes did not receive much attention partially because of the complex form of the first EC fluxes Tadmor proposed, which did not foster practical applications. Important contributions were subsequently brought by Roe and Ismail [52, 55, 54] who derived a simple EC flux by solving the scalar EC condition algebraically, and provided an upwind-type entropy-stable dissipation operator, leading to the first 'affordable' entropy-stable scheme. Their contributions were made as part of an attempt to cure the longtime carbuncle problem in hypersonic flows by trying to produce the right amount of entropy across the shock structure. The techniques used in the construction of this affordable scheme can be used for other systems such as the shallow water equations [56] and the ideal and relativistic MHD equations [57, 58].

ES schemes further gained attention when high-order formulations such as the TecNO schemes of Fjordholm *et al.* [60] and the space-time DG scheme of Mishra & Hiltebrand [67] started to emerge. TecNO schemes leverage the previous work of LeFloch *et al.*, the entropy-stable dissipation operator of Roe [52] and the sign property of the ENO reconstruction proved by Fjordholm *et al.* [61]. The scheme of Mishra & Hiltebrand established some of the missing links between Tadmor's framework [49] and the space-time DG scheme of Barth [59], which achieves entropy-stability at the fully-discrete level in a different way. A significant contribution was brought by Fisher & Carpenter [62] who generalized the constructs of entropy-conservation and entropy-stability to Summation-By-Parts operators [79, 80, 81, 82, 83, 84] and developed the first high-order entropy-stable finite-difference scheme for the compressible Navier-Stokes equations. They also made the first comparisons with existing high-order schemes, such as the kinetic energy preserving scheme of Subbareddy & Candler [112], which highlighted the superior robustness of ES schemes in under-resolved simulations of compressible turbulent flows. Their contribution is largely to credit for the recognition that, quoting from the NASA CFD Vision 2030 Study [66], "*Longer term, high-risk research should focus on the development of truly enabling technologies such as monotone or entropy stable schemes in combination with innovative solvers on large-scale HPC hardware.*" Their contribution also set the stage for what is currently the most actively developed high-order entropy-stable numerical framework [85, 86, 87, 88, 69].

The superior robustness properties of ES schemes in the severely under-resolved high-order setting have been further established by the work Diosady & Murman [68], Pazner & Persson [70] and Fernandez *et al.* [71], and spurred considerable efforts on the development of code infrastructures which build upon this numerical framework. An example is the *eddy* [74, 73, 75, 76, 77, 78] solver developed at NASA Ames Research Center for the simulation of turbulent separated flows.

1.2.2 Stance and goals of the thesis

The naming 'entropy-stable' does not exclusively apply to Tadmor's family of schemes, even though we use it as such throughout this thesis. It generally applies to any scheme which is consistent with entropy inequalities such as (1.4) at the discrete level. Godunov-type schemes, which include the well-known Roe scheme, are built upon the knowledge of exact and approximate solutions to Riemann problems at interfaces. These schemes can achieve entropy-stability under conditions laid out in the seminal work of Van Leer, Harten and Lax [27]. The same goes for other schemes such as the Lax-Friedrichs scheme and the E-schemes of Osher [72]. We should also mention the schemes recently developed by Guermond and collaborators [176, 177], which on top of entropy-stability can ensure the positivity of density and pressure.

In view of the challenges posed by the simulation of compressible flows, we find Tadmor's approach to be more appealing. This is mainly because entropy-stability is achieved in a sharper way. Indeed, the user can enforce entropy-conservation, that is no production of entropy, if deemed necessary, and when entropy-stability is enforced, it comes with the precise knowledge of how much entropy is being produced at the discrete level. This contrasts with the other entropy-stable schemes for which the user only knows that entropy is being produced. This knowledge can aid in better understanding some of the fundamental problems faced by shock-capturing schemes (last paragraph of section 1.1.1), as entropy production is often found at the center of the discussion [52, 55, 35]. This knowledge can also be leveraged in the context of turbulence, where one of the main challenges is to strike a balance between high-order accuracy and robustness. Entropy production can serve as a metric for how much information is lost because of stabilization, as well as for how unstable the scheme becomes when the stabilization is modified. What's more, entropy turns out to be an important physical quantity in compressible turbulent flow theory as well [157, 158, 159, 160]. Viewed from this angle, entropy could also provide a common language enabling the fields of turbulence modeling and numerical analysis to work together[6]. In this regard, the fact that Tadmor's framework leaves to the user the responsibility of figuring out what is the right way to conserve or produce entropy is beneficial. The standard entropy-conservative flux and dissipation operator in use as of now are Roe's [52], but nothing is set in stone. Upon understanding how entropy-stable schemes manage entropy at the discrete level, turbulence theoreticians for instance could propose their own EC/ES discretization. All in all, the stance of this thesis is that ES schemes could make for a good *foundation*[7] supporting the development of CFD algorithms for compressible flows.

This thesis has two primary goals. The first one is to further develop this foundation, not towards more advanced high-order discretizations, but towards enabling applications to more complex physical models. The second goal is to better understand how entropy-stable schemes work. That is not only understanding how they can be constructed, but also understanding how they may or may not improve the quality of the numerical solution. We make a conscious effort to use entropy as a lens in this effort. Not doing so would be counter-intuitive in view of the framework we chose to work with.

It is important to recognize that entropy-stable schemes have garnered attention mostly for providing a potential solution to the stability issues of high-order methods in under-resolved compressible turbulent flow simulations. Putting aside the fact that

[6] The ILES/no-model approach was mentioned earlier for completeness only. We do not advocate for it.

[7] We do not see entropy-stability as a finality or "the holy grail of numerical analysis" [65]. We see it as a good starting point.

entropy-stability does not make for an unbreakable code[8], entropy-stable schemes do not enjoy the same popularity and practical interest elsewhere. The reason for that is simple. The reader will quickly realize that entropy-stability does not guarantee a well-behaved solution. It is for this precise reason that we decided to investigate some of the shock-capturing problems mentioned at the end of section 1.1.1, *where stability is not the problem.* We believe that there is more to learn about local behavior and numerical error from these problems than from the canonical turbulent flow configurations typically used to highlight the robustness of entropy-stable schemes.

1.3 Contributions and outline

In chapter 2, the numerical crafts of entropy-stable schemes and the theory behind them are covered. The remaining chapters report on the individual contributions of the thesis.

In chapter 3, we study the receding flow problem, which is a simple setup used to study the anomalous temperature rise, termed "overheating", typically observed in shock reflection and shock interaction predictions. Previous studies showed that the anomaly can be cured if conservation of entropy is enforced, but at the considerable price of total energy conservation. Entropy-Conservative (EC) schemes can achieve both simultaneously and therefore appeared as a potential solution. We show that while the overheating is correlated to entropy production, entropy conservation does not necessarily prevent it.

The current state-of-the-art solves the compressible Navier-Stokes equations for a single-component perfect gas in chemical and thermal equilibrium. This model is inappropriate in aerospace engineering applications such as hypersonics and combustion, which typically involve chemically reacting gas mixtures far from equilibrium. As a first step towards enabling their use for these applications, we formulate, in

[8] entropy-stability does not guarantee that density for instance will remain positive

chapter 4, ES schemes for the multicomponent compressible Euler equations. Special care had to be taken as we found out that the theoretical foundations of ES schemes begin to crumble in the limit of vanishing partial densities.

The realization that ES schemes can only go as far as their theory led us to review some of it. A fundamental result supporting the development of limiting strategies for high-order methods is the minimum entropy principle proved by Tadmor for the compressible Euler equations. It states that the specific entropy of the physically relevant weak solution does not decrease. In chapter 5, we prove a minimum entropy principle for the mixture's specific entropy in the multicomponent case, which implies that the aforementioned limiting strategies could be extended to this system.

In chapter 6, we study the behavior of ES schemes in the low Mach number regime, where shock-capturing schemes are known to suffer from severe accuracy degradation issues. A classic remedy to this problem is the flux-preconditioning technique, which consists in modifying artificial dissipation terms to enforce consistent low Mach behavior. We showed that ES schemes suffer from the same issues and that the flux-preconditioning technique can improve their behavior without interfering with entropy-stability. Furthermore, we demonstrated analytically that these issues stem from an acoustic entropy production field which scales improperly with the Mach number, generating spatial fluctuations that are inconsistent with the equations. An important outgrowth of this effort is the discovery that skew-symmetric dissipation operators can alter the way entropy is conserved or produced locally.

CHAPTER II

Theoretical and Numerical Background

While in this thesis we only work with the compressible Euler equations and some of its variants, the concepts and techniques proper to entropy-stable schemes apply to a broader range of systems. Therefore, we consider general hyperbolic systems of conservation laws in one dimension as follows:

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}}{\partial x} = 0, \tag{2.1}$$

where $\mathbf{u} = \mathbf{u}(x,t)$ is the vector of conserved variables and $\mathbf{f} = \mathbf{f}(\mathbf{u})$ is the vector of fluxes.

2.1 Weak solutions, entropy and symmetrization

It is well known that discontinuous solutions to (2.1) can develop from smooth initial conditions $\mathbf{u}(x,0)$ when the fluxes are nonlinear functions of the conserved variables. *Weak* solutions must therefore be sought. Unfortunately, these are not uniquely defined and one needs additional conditions to hopefully find the right one, or at least discard the non-physical ones.

It is common practice to view physical solutions as those arising as vanishing viscosity limits, $\mathbf{u}(x,t) = \lim_{\epsilon\to 0} \mathbf{u}^\epsilon(x,t)$, of solutions $\mathbf{u}^\epsilon(x,t)$ to the regularized system.

$$\frac{\partial \mathbf{u}^\epsilon}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u}^\epsilon)}{\partial x} = \epsilon \frac{\partial^2 \mathbf{u}^\epsilon}{\partial x^2}, \;\; \epsilon > 0. \tag{2.2}$$

A number of hyperbolic systems admit a convex extension [46], in the sense that they imply an additional conservation equation for a mathematical entropy U:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = 0, \tag{2.3}$$

where the entropy pair $(U, F) = (U(\mathbf{u}), F(\mathbf{u}))$ is such that $U(\mathbf{u})$ is strictly convex and the compatibility relation

$$\frac{\partial U}{\partial \mathbf{u}} \frac{\partial \mathbf{f}}{\partial \mathbf{u}} = \frac{\partial F}{\partial \mathbf{u}}, \tag{2.4}$$

necessary for the system (2.1) to imply (2.3), is satisfied. The compressible Euler equations qualify with $(U, F) = (-\rho s, -\rho u s)$ for instance. For systems endowed with such an entropy structure, one has:

$$\left(\frac{\partial U}{\partial \mathbf{u}^\epsilon}\right) \times (2.2) \quad \Longrightarrow \quad \frac{\partial U(\mathbf{u}^\epsilon)}{\partial t} + \frac{\partial F(\mathbf{u}^\epsilon)}{\partial x} = \epsilon \left(\frac{\partial U}{\partial \mathbf{u}^\epsilon}\right) \frac{\partial^2 \mathbf{u}^\epsilon}{\partial x^2}$$

Using a chain rule on the right hand side term and using the convexity of U, it can be showed that:

$$\frac{\partial U(\mathbf{u}^\epsilon)}{\partial t} + \frac{\partial F(\mathbf{u}^\epsilon)}{\partial x} \leq \epsilon \frac{\partial^2 U(\mathbf{u}^\epsilon)}{\partial x^2}.$$

In the limit $\epsilon \to 0$, this leads to the well-known entropy inequality:

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} \leq 0, \tag{2.5}$$

which makes for an indirect way to establish whether a weak solution is physical. For the compressible Euler equations with $(U, F) = (-\rho s, \; -\rho u s)$ the above inequality leads to the well-known entropy conditions which should be satisfied across shock discontinuities [47].

Besides, Mock [45] showed that the existence of an entropy equation (2.3) also implies the existence of a one-to-one mapping $\mathbf{u} \to \mathbf{v}$, with:

$$\mathbf{v} := \left(\frac{\partial U}{\partial \mathbf{u}} \right)^T, \tag{2.6}$$

which *symmetrizes* the original system. This means that, starting from the quasi-linear form of the original system (2.1):

$$\frac{\partial \mathbf{u}}{\partial t} + A \frac{\partial \mathbf{u}}{\partial x} = 0, \; A := \frac{\partial \mathbf{f}}{\partial \mathbf{u}}, \tag{2.7}$$

applying the change of variables $\mathbf{u} \to \mathbf{v}$ turns (2.7) into a system of the form:

$$H \frac{\partial \mathbf{v}}{\partial t} + B \frac{\partial \mathbf{u}}{\partial x} = 0, \; H := \frac{\partial \mathbf{u}}{\partial \mathbf{v}}, \; B := \frac{\partial \mathbf{f}}{\partial \mathbf{v}} = AH, \tag{2.8}$$

where B and H are symmetric, and H is symmetric positive definite. Such systems are called symmetric hyperbolic and are appreciated in the analysis of PDEs [46, 172]. The vector $\mathbf{v}$ is commonly referred to as the vector of *entropy variables.* For completeness, we mention that Godunov showed the converse of Mock's result [43].

At this point, we draw the attention of the reader to the difference between *thermodynamic* entropy ρs and *mathematical* entropy U, which is a more general concept. For the Burgers equation ($\mathbf{u} = u, \; \mathbf{f} = u^2/2$) for instance, $U = \frac{1}{2}u^2$ is an entropy. In our context, the mathematical entropy is the opposite of the thermodynamic entropy, and the statement of integral entropy-stability can be interpreted either as dissipation of (mathematical) entropy or as production of (thermodynamic) entropy. We adopt the latter throughout this thesis.

Numerical schemes consistent with the system (2.1) are not necessarily consistent with the entropy equation (2.3) or the entropy inequality (2.5). The pioneering work of Tadmor [49] introduced a class of finite-volume schemes which can achieve such consistency at the semi-discrete level (and at the fully-discrete level as well, this is covered in chapter III). In addition to the entropy variables, the construction of these schemes involve potential functions $(\mathcal{U}, \mathcal{F})$ defined by:

$$\mathcal{U} := \mathbf{v}^T\mathbf{u} - U, \ \mathcal{F} := \mathbf{v}^T\mathbf{f} - F.$$

The potential functions satisfy the relationships:

$$\mathbf{u} = \left(\frac{\partial \mathcal{U}}{\partial \mathbf{v}}\right)^T, \ \mathbf{f} = \left(\frac{\partial \mathcal{F}}{\partial \mathbf{v}}\right)^T.$$

For the Euler equations, the entropy pair we usually work with is the one introduced by Hughes *et. al* [44]:

$$U = -\frac{\rho s}{\gamma - 1}, \ F = -\frac{\rho u s}{\gamma - 1}. \tag{2.9}$$

This pair belongs to the more general class of entropy pairs derived by Harten [48] (which is discussed in chapter V). The $\gamma - 1$ factor in the denominator is introduced so that the corresponding potential functions write $\mathcal{U} = \rho, \mathcal{F} = \rho u$. The corresponding entropy variables are given by:

$$\mathbf{v} = \left[\frac{\gamma - s}{\gamma - 1} - \frac{1}{2}\frac{\rho u^2}{p}, \ \frac{\rho u}{p}, \ -\frac{\rho}{p}\right]^T. \tag{2.10}$$

.

2.2 Formulation

In this section we detail the main steps to follow in order to construct semi-discrete entropy-stable schemes. Consider the finite volume scheme:

$$\frac{d}{dt}\mathbf{u}_j(t) + \frac{1}{\Delta x}\big(\mathbf{f}_{j+\frac{1}{2}} - \mathbf{f}_{j-\frac{1}{2}}\big) = 0, \tag{2.11}$$

where $\mathbf{f}_{j+\frac{1}{2}}$ is a consistent interface flux. The subscripts j and $j+\frac{1}{2}$ refer to cell and interface indices, respectively. The first and trademark step of Tadmor's framework consists in seeking entropy conservation. The finite volume scheme (2.11) is termed entropy-conservative if it satisfies the equation:

$$\frac{d}{dt}U(\mathbf{u}_j(t)) + \frac{1}{\Delta x}\big(F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}}\big) = 0, \tag{2.12}$$

where $F_{j+\frac{1}{2}}$ is a consistent entropy interface flux. The second step consists in adding carefully designed dissipation terms on top of the entropy-conservative scheme to achieve entropy-stability, that is meeting the inequality:

$$\frac{d}{dt}U(\mathbf{u}_j(t)) + \frac{1}{\Delta x}\big(F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}}\big) < 0. \tag{2.13}$$

2.2.1 Entropy-Conservative Fluxes

By definition of the entropy variables, one has:

$$\mathbf{v}_j^T\frac{d}{dt}\mathbf{u}_j = \frac{d}{dt}U(\mathbf{u}_j),$$

therefore the scheme (2.11) is entropy conservative if and only if the interface flux $\mathbf{f}_{j+\frac{1}{2}}$ is such that there exists a consistent entropy interface flux $F_{j+\frac{1}{2}}$ that satisfies:

$$\mathbf{v}_j^T\big(\mathbf{f}_{j+\frac{1}{2}} - \mathbf{f}_{j-\frac{1}{2}}\big) = F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}}.$$

Tadmor [49] showed that this holds if and only if the entropy conservation condition:

$$[\mathbf{v}_{j+1}-\mathbf{v}_j]^T\mathbf{f}_{j+\frac{1}{2}}=\mathcal{F}_{j+1}-\mathcal{F}_j, \tag{2.14}$$

where $\mathcal{F}_j=\mathcal{F}(\mathbf{u}_j)$, is satisfied. In this case, $\mathbf{f}_{j+\frac{1}{2}}$ is called an entropy-conservative flux and the corresponding entropy flux $F_{j+\frac{1}{2}}$ is given by the formula:

$$F_{j+\frac{1}{2}}=\frac{1}{2}(\mathbf{v}_j+\mathbf{v}_{j+1})^T\mathbf{f}_{j+\frac{1}{2}}-\frac{1}{2}(\mathcal{F}_j+\mathcal{F}_{j+1}). \tag{2.15}$$

For scalar PDEs, $\mathbf{v}$ is a scalar and the entropy conservation condition (2.14) has only one solution, namely $\mathbf{f}_{j+\frac{1}{2}}=[\mathcal{F}]_{j+\frac{1}{2}}/[\mathbf{v}]_{j+\frac{1}{2}}$. For systems, equation (2.14) does not uniquely determine the interface flux. The first entropy-conservative flux was introduced by Tadmor [49]:

$$\mathbf{f}_{j+\frac{1}{2}}=\int_0^1 f(\mathbf{v}_{j+\frac{1}{2}}(\xi))d\xi,\ \mathbf{v}_{j+\frac{1}{2}}(\xi)=\mathbf{v}_j+\xi[\mathbf{v}]_{j+\frac{1}{2}}. \tag{2.16}$$

This elegant flux has the inconvenient property of not having a closed form. Its evaluation requires quadrature rules. Later on, Tadmor [51] proposed to use piecewise-constant paths instead. The resulting flux has an explicit form which depends on the path decomposition but it did not get much attention.

An entropy-conservative flux that has been more popular for its simplicity compared to the previous two is the one derived by Roe [52] for the compressible Euler equations. The method used to derive it is general enough to be applied to other systems (in chapter IV we use it for the multicomponent system). It is also central to the work of chapter III. We cover it here.

Denote $\mathbf{f}^*=[f_1,\ f_2,\ f_3]$ the interface flux separating two adjacent cells. Using compact jump notation, the condition (2.14) can be rewritten as:

$$[\mathbf{v}]^T\mathbf{f}^* = [\mathcal{F}].$$

For the entropy pair (2.9), the jump terms in the entropy variables write:

$$[\mathbf{v}]^T = \left[\frac{-[s]}{\gamma-1} - \frac{1}{2}\left[\frac{\rho u^2}{p}\right],\ \left[\frac{\rho u}{p}\right],\ -\left[\frac{\rho}{p}\right]\right].$$

Define the set of independent variables $(z_1, z_2, z_3) = (\sqrt{\frac{\rho}{p}}, \sqrt{\frac{\rho}{p}}u, \sqrt{\rho p})$. Then

$$\rho = z_1 z_3,\ p = \frac{z_3}{z_1},\ \frac{\rho}{p} = z_1^2,\ \frac{\rho u}{p} = z_1 z_2,\ \frac{\rho u^2}{p} = z_2^2,\ \rho u = z_2 z_3,$$
$$S = (1-\gamma)ln(z_3) - (1+\gamma)ln(z_1).$$

Using the identities $[ab] = \bar{a}[b] + \bar{b}[a]$ and $[ln(a)] = [a]/a^{ln}$, where $\bar{a}$ and a^{ln} denote the arithmetic and logarithmic averages, respectively, one can show that:

$$[s] = \frac{(1-\gamma)}{z_3^{ln}}[z_3] - \frac{(1+\gamma)}{z_1^{ln}}[z_1],\ \left[\frac{\rho u^2}{p}\right] = 2\bar{z_2}[z_2],\ \left[\frac{\rho u}{p}\right] = \bar{z_1}[z_2] + \bar{z_2}[z_1],$$
$$\left[\frac{\rho}{p}\right] = 2\bar{z_1}[z_1],\ [\rho u] = \bar{z_2}[z_3] + \bar{z_3}[z_2].$$

The motivation behind the introduction of the variables z_i is to “exactly linearize” all the jump terms involved in the entropy conservation condition, which now writes:

$$f_1\left(\frac{1}{z_3^{ln}}[z_3] - \frac{1+\gamma}{1-\gamma}\frac{1}{z_1^{ln}}[z_1] - \bar{z_2}[z_2]\right) + f_2(\bar{z_1}[z_2] + \bar{z_2}[z_1]) + f_3(-2\bar{z_1}[z_1]) = \bar{z_2}[z_3] + \bar{z_3}[z_2]. \tag{2.17}$$

Regrouping, equation (2.17) becomes:

$$[z_1]\left(-f_1\frac{1+\gamma}{1-\gamma}\frac{1}{z_1^{ln}} + f_2\bar{z_2} - 2f_3\bar{z_1}\right) + [z_2]\big(-f_1\bar{z_2} + f_2\bar{z_1}\big) + [z_3]\big(\frac{1}{z_3^{ln}}f_1\big) = [z_2]\bar{z_3} + [z_3]\bar{z_2}.$$

The jumps in z_i are independent, therefore:

$$-f_1\frac{1+\gamma}{1-\gamma}\frac{1}{z_1^{ln}} + f_2\bar{z}_2 - 2f_3\bar{z}_1 = 0,\; -f_1\bar{z}_2 + f_2\bar{z}_1 = \bar{z}_3,\; \frac{1}{z_3^{ln}}f_1 = \bar{z}_2.$$

The variables z_i basically enabled the conversion of the scalar condition (2.14) into a system of 3 equations that can easily be solved:

$$f_1 = \bar{z}_2 z_3^{ln},\; f_2 = (\bar{z}_3 + f_1\bar{z}_2)/(\bar{z}_1),\; f_3 = \frac{1}{2\bar{z}_1}(-f_1\frac{1+\gamma}{1-\gamma}\frac{1}{z_1^{ln}} + f_2\bar{z}_2).$$

This is the entropy-conservative flux of Roe. The choice of independent variables z_i is open. Using the same method with the set $(z_1, z_2, z_3) = (\rho, u, \frac{\rho}{2p})$ instead, Chandrasekhar [63] derived the following entropy-conservative flux:

$$f_1 = z_1^{ln}\bar{z}_2,\; f_2 = \frac{\bar{z}_1}{2\bar{z}_3} + \bar{z}_2 f_1,\; f_3 = \left[\frac{1}{2(\gamma-1)z_3^{ln}} - \frac{1}{2}\overline{z_2^2}\right]f_1 + \bar{z}_2 f_2. \tag{2.18}$$

Jameson [64] showed that an interface flux preserves the kinetic energy of the system at the semi-discrete level provided that the density flux f_1 and momentum flux f_2 satisfy $f_2 = \tilde{p} + \overline{u}f_1$, where $\tilde{p}$ is any consistent average pressure. The entropy-conservative flux given by equation (2.18) satisfies this property as well. In contrast to the conclusions of [63] (section 4.6.), such a flux is not unique. With the set $(z_1, z_2, z_3) = (p, u, \frac{\rho}{2p})$, the resulting entropy-conservative flux:

$$f_1 = 2\bar{z}_3\bar{z}_2 z_1^{ln},\; f_2 = \frac{\bar{\rho}}{2\bar{z}_3} + \bar{u}f_1,\; f_3 = f_1(\frac{\gamma}{\gamma-1}\frac{1}{2z_3^{ln}} - \frac{1}{2}\overline{z_2^2}) + f_2\bar{z}_2 - \bar{z}_1\bar{z}_2,$$

is also kinetic energy preserving. The term $-\bar{z}_1\bar{z}_2$ of the energy flux f_3 is missing in [63].

2.2.2 Entropy-Stable Fluxes

The finite-volume scheme (2.11) with the interface flux $\mathbf{f}_{j+\frac{1}{2}}$ defined as:

$$\mathbf{f}_{j+\frac{1}{2}} = \mathbf{f}^*_{j+\frac{1}{2}} - D_{j+\frac{1}{2}}[\mathbf{v}]_{j+\frac{1}{2}}.$$

where $f^*_{j+\frac{1}{2}}$ satisfies the entropy conservation condition (2.14), and $D_{j+\frac{1}{2}}$ is a positive definite matrix, satisfies

$$\frac{d}{dt}U(\mathbf{u}_j(t)) + \frac{1}{\Delta x}[F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}}] = -\mathcal{E}_j. \tag{2.19}$$

with $\mathcal{E}_j$ given by:

$$\mathcal{E}_j = \frac{1}{2\Delta x}\big([\mathbf{v}]^T_{j+\frac{1}{2}} D_{j+\frac{1}{2}}[\mathbf{v}]_{j+\frac{1}{2}} + [\mathbf{v}]^T_{j-\frac{1}{2}} D_{j-\frac{1}{2}}[\mathbf{v}]_{j-\frac{1}{2}}\big) > 0, \tag{2.20}$$

and is therefore entropy-stable. In this case, the interface entropy flux $F_{j+\frac{1}{2}}$ is given by:

$$F_{j+\frac{1}{2}} = F^*_{j+\frac{1}{2}} - \frac{1}{2}(\mathbf{v}_j + \mathbf{v}_{j+1})D_{j+\frac{1}{2}}[\mathbf{v}]_{j+\frac{1}{2}}, \tag{2.21}$$

where $F^*_{j+\frac{1}{2}}$ is the entropy flux associated with $f^*_{j+\frac{1}{2}}$.

Summing over all cells and assuming periodic boundary conditions leads to the integral entropy-stability statement:

$$\frac{d}{dt}\sum_j U(\mathbf{u}_j) = -\sum_j \mathcal{E}_j < 0. \tag{2.22}$$

For the compressible Euler equations with the pair (2.9), this writes:

$$\frac{d}{dt}\sum_j (\rho s)_j = (\gamma - 1)\sum_j \mathcal{E}_j > 0. \tag{2.23}$$

This integral stability statement is obtained as a consequence of the local statement (2.19), which in itself is not a stability statement. It does not necessarily imply for instance that in every cell:

$$\frac{d}{dt}U(\mathbf{u}_j) < 0 \ \Leftrightarrow \ \frac{d}{dt}(\rho s)(\mathbf{u}_j) > 0. \tag{2.24}$$

Looking at equation (2.19), we see that the local variation of U in time depends on $\mathcal{E}_j$, which is a positive quantity, but also on the interface entropy flux $F_{j+\frac{1}{2}}$, which according to equation (2.21) is determined by both the entropy-conservative flux and the dissipation operator. To the best of the author's knowledge, conditions under which (2.24) is met have not been examined.

Upwind-type dissipation operator

Let $R\Lambda R^{-1}$ be an eigendecomposition of A. A popular choice for the dissipation operator consists of recasting the upwind operator of Roe [10] $\rightarrow \frac{1}{2}R|\Lambda|R^{-1}[\mathbf{u}]$ in terms of the entropy variables. With the differential relation $d\mathbf{u} = Hd\mathbf{v}$, this leads to:

$$D[\mathbf{v}] = \frac{1}{2}R|\Lambda|R^{-1}H[\mathbf{v}].$$

For the compressible Euler equations, Merriam [65] (section 7.3) pointed out that there exists a scaling of the columns of R such that $H = RR^T$, which leads to a dissipation matrix $D = R|\Lambda|R^T$ that has the desirable property of being positive definite. Later on, Barth [59] showed that Merriam's finding is a direct consequence of the fact that the entropy variables symmetrize the system. Merriam's result is therefore more general, and is recast by Barth as an eigenscaling theorem: for any diagonalizable matrix A symmetrized on the right by a symmetric positive definite matrix H, there exists a symmetric positive definite block diagonal matrix T that block scales the eigenvectors R of A in such a manner that:

$$A = (RT)\Lambda(RT)^T \quad \text{and} \quad H = (RT)(RT)^T. \tag{2.25}$$

The dimensions of the blocks of T correspond to the multiplicities of the eigenvalues of A. The second identity in equation (2.25) provides an explicit expression for the squared scaling matrix $T^2 = R^{-1}HR^{-T}$.

More details about this dissipation operator can be found in chapters IV and VI.

2.2.3 High-order discretizations

In this section, we cover two high-order entropy-stable formulations: TecNO schemes [60] and Discontinuous Galerkin [117, 118] (DG) schemes discretizing the entropy variables [59, 44, 67]. High-order ES schemes are not limited to these two options. Formulations based on Summation-By-Parts operators [62, 85, 86, 87, 88, 69] for instance are being actively developed.

Developing high-order entropy-stable schemes is not the purpose of this thesis. We mostly worked with TecNO schemes as they are relatively easy to implement. The DG formulation based on entropy variables is discussed to offer some perspective on what is, to the best of the author's knowledge, the first high-order entropy-stable formulation.

2.2.3.1 TecNO schemes

TecNO schemes (Fjordholm *et al.* [60]) are high-order entropy-stable finite volume schemes that combine the high-order entropy-conservative flux formulation of LeFloch *et al.* [50], the stencil selection procedure of ENO schemes [30] and entropy-stable dissipation operators [49, 52].

The scheme still writes as (2.11) but differs in the interface flux. The two components, entropy-conservative flux and dissipation operator, are altered to achieve high-resolution in such a manner that entropy-stability is retained.

The entropy-conservative flux $\mathbf{f}^*$, is replaced with a high-order entropy-conservative flux $\mathbf{f}^*_{\mathbf{2p}}$ defined over a centered stencil of $2p$ points $(v_{j-p+1}, \ \ldots \ , v_{j+p})$ by:

$$\mathbf{f}^*_{\mathbf{2p}}(v_{j-p+1}, \ \ldots \ , v_{j+p}) = \sum_{i=1}^{p} \alpha_{i,p} \sum_{s=0}^{i-1} \mathbf{f}^*(v_{j-s}, v_{j-s+i}).$$

This flux essentially consists of a weighted combination of second-order entropy-conservative fluxes. The coefficients $\alpha_{i,p}$ need to satisfy [50]:

$$\sum_{i=1}^{p} i\alpha_{i,p} = 1, \ \sum_{i=1}^{p} i^{2s-1}\alpha_{i,p} = 0, \ s = 2 \ldots p.$$

The first equation is for consistency, the second is for 2p-th order accuracy. For $p = 2$ (4-th order) and $p = 3$ (6-th order) the coefficients are:

$$\alpha_{1,2} = \frac{4}{3}, \ \alpha_{2,2} = -\frac{1}{6}$$
$$\alpha_{1,3} = \frac{3}{2}, \ \alpha_{2,3} = -\frac{3}{10}, \ \alpha_{3,3} = \frac{1}{30}.$$

As we saw earlier, the dissipation term of a first order ES flux typically takes the form:

$$D[\mathbf{v}] = (RT)|\Lambda|(RT)^T[\mathbf{v}],$$

where R is the matrix of right eigenvectors of A and T is a scaling matrix. The reconstruction used by TecNO schemes is motivated by the fact that the ENO reconstruction satisfies a *sign property*. It was shown [61] that for any vector $\mathbf{w} \in \mathbb{R}^N$, the TecNO reconstruction satisfies:

$$\langle \mathbf{w} \rangle = B[\mathbf{w}], \ B = diag([b_0, \ \ldots, \ b_{N-1}]), \ b_i \geq 0,$$

where $\langle \mathbf{w} \rangle$ and $[\mathbf{w}]$ are the reconstructed and initial jumps, respectively. Let $\mathbf{w}$ be such that $[\mathbf{w}] = (RT)^T[\mathbf{v}]$, then the dissipation operator $\tilde{D}[\mathbf{w}] = (RT)|\Lambda| \langle \mathbf{w} \rangle$ is ES because:

$$(RT)|\Lambda| \langle \mathbf{w} \rangle = (RT)|\Lambda|B[\mathbf{w}] = (RT)(|\Lambda|B)(RT)^T[\mathbf{v}]$$

$|\Lambda|B$ is a positive diagonal matrix. The high-order dissipation operator $\tilde{D}[\mathbf{w}]$ thus achieves entropy-stability by reconstructing jumps in a specific set of variables. Fjordholm *et al.* refer to them as the *scaled entropy variables*. We will see in chapter VI that $[\mathbf{w}]$ can be interpreted as a vector of wave strengths.

We conclude this section with a reminder that while TecNO schemes do use non-oscillatory reconstructions, they are not *de facto* embedded with *provable* non-oscillatory properties (see figures 6 and 7 in [60]). In fact, this remark applies to any similarly termed scheme in the context of nonlinear systems of conservation laws. TVD schemes [13] are well-grounded in the scalar case and in the linear constant coefficients case because there is a clear definition of what quantity "total-variation" points to. Even though many good results have been produced by these schemes (there have even been efforts to accommodate the TVD framework in the finite-element context [242, 243]), the theoretical bearings of non-oscillatory schemes are currently missing for nonlinear systems [13, 20].

2.2.3.2 Discontinuous Galerkin

Tadmor's pioneering work came out around the same time as when Hughes *et. al* [44] showed, under the assumption of exact numerical quadrature, that Continuous finite element solutions to the compressible Euler equations become consistent with the entropy equation [1] when the entropy variables are discretized instead of conservative variables.

[1] the result was actually proved for the more general compressible Navier-Stokes equations, which we do not cover in this thesis.

For DG discretizations of the Compressible Euler Equations, the same result immediately follows if entropy-stable fluxes, in the sense of Tadmor [49], are used. Yet, that is not how Barth approached it [59], even though parallels can be drawn [67]. Consider the more general case of a multi-dimensional system of conservation laws:

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f}_j}{\partial x_j} = 0, \tag{2.26}$$

which implies the multi-dimensional entropy equation:

$$\frac{\partial U(\mathbf{u})}{\partial t} + \frac{\partial F_j}{\partial x_j} = 0.$$

The weak form associated with a semi-discrete DG discretization of (2.26) typically writes, for each element K of the mesh and each degree of freedom (dof) i:

$$R_{K,i} = 0, \tag{2.27}$$

$$R_{K,i} = \int_K \phi_{K,i} \frac{d\mathbf{u}(\mathbf{q}^K)}{dt}\, dV \; - \int_K \mathbf{f}_j(\mathbf{q}^K) \frac{\partial \phi_{K,i}}{\partial x_j}\, dV \; + \int_{\delta K} \phi_{K,i} \mathbf{f}^*\, dS. \tag{2.28}$$

$R_{K,i}$ denotes the residual associated with the i-th dof in element K. The interface flux is given by $\mathbf{f}^* = \mathbf{f}^*(\mathbf{q}^K, \mathbf{q}^{K'}, \mathbf{n})$ where K' denotes the neighboring element on δK and $\mathbf{n} = (n_j)_{1 \leq j \leq dim}$ is the vector normal to δK. $\mathbf{q}^K$ denotes the discrete solution in element K which consists of a linear combination of polynomial basis functions $\phi_{K,i}$:

$$\mathbf{q}^K = \sum_{i=1}^{N_d} \phi_{K,i} \mathbf{q}_{K,i}.$$

Using integration by parts on the second term of (2.28), we have:

$$R_{K,i} = \int_K \phi_{K,i} \frac{\partial \mathbf{u}(\mathbf{q}^K)}{\partial t}\, dV \; + \int_K \phi_{K,i} \frac{\partial \mathbf{f}_j(\mathbf{q}^K)}{\partial x_j}\, dV \; + \int_{\delta K} \phi_{K,i} \big(\, \mathbf{f}^* - \mathbf{f}_n(\mathbf{q}^K) \,\big)\, dS.$$

where $\mathbf{f}_n = \sum_{j=1}^{d} n_j \mathbf{f}_j$. The solution in each element is typically represented as a polynomial expansion in terms of the conserved variables:

$$\mathbf{q}^K := \mathbf{u}^K = \sum_{i}^{N_d} \phi_{K,i} \mathbf{u}_{K,i}.$$

Discretizing the entropy variables means that it is the entropy variables that are represented using polynomials instead:

$$\mathbf{q}^K := \mathbf{v}^K = \sum_{i}^{N_d} \phi_{K,i} \mathbf{v}_{K,i}. \tag{2.29}$$

From equation (2.27), it is clear that any linear combination of the residuals $R_{i,K}$ is zero. When the entropy variables are discretized, the particular combination $\sum_i \mathbf{v}_{K,i}^T R_{K,i}$ gives:

$$\begin{aligned} \int_K \Big(\sum_i \mathbf{v}_{K,i} \phi_{K,i} \Big) \frac{\partial \mathbf{u}(\mathbf{q}^K)}{\partial t} \, dV \; + \int_K \Big(\sum_i \mathbf{v}_{K,i} \phi_{K,i} \Big) \frac{\partial \mathbf{f}_j(\mathbf{q}^K)}{\partial x_j} \, dV \\ + \int_{\delta K} \Big(\sum_i \mathbf{v}_{K,i} \phi_{K,i} \Big) \big(\ \mathbf{f}^* - \mathbf{f}_n(\mathbf{q}^K) \ \big) \, dS = 0, \end{aligned}$$

which by definition (2.29) writes:

$$\begin{aligned} &\int_K (\mathbf{v}^K)^T \frac{\partial \mathbf{u}(\mathbf{v}^K)}{\partial t} \, dV \; + \int_K (\mathbf{v}^K)^T \frac{\partial \mathbf{f}_j(\mathbf{v}^K)}{\partial x_j} \, dV \; + \int_{\delta K} (\mathbf{v}^K)^T (\ \mathbf{f}^* - \mathbf{f}_n(\mathbf{v}^K) \ ) \, dS = 0. \\ \Leftrightarrow &\int_K \frac{\partial U(\mathbf{v}^K)}{\partial t} \, dV \; + \int_K \frac{\partial F_j(\mathbf{v}^K)}{\partial x_j} \, dV \; + \int_{\delta K} (\mathbf{v}^K)^T (\ \mathbf{f}^* - \mathbf{f}_n(\mathbf{v}^K) \ ) \, dS = 0. \\ \Leftrightarrow &\int_K \frac{\partial U(\mathbf{v}^K)}{\partial t} \, dV \; + \int_{\delta K} (\mathbf{v}^K)^T \mathbf{f}^* - \big(\ (\mathbf{v}^K)^T \mathbf{f}_n(\mathbf{v}^K) - F_n(\mathbf{v}^K) \ \big) \, dS = 0. \\ \Leftrightarrow &\int_K \frac{\partial U(\mathbf{v}^K)}{\partial t} \, dV \; + \int_{\delta K} \big(\ (\mathbf{v}^K)^T \mathbf{f}^* - \mathcal{F}(\mathbf{v}^K) \ \big) \, dS = 0. \end{aligned} \tag{2.30}$$

Decomposing the integrand of the second term:

$$(\mathbf{v}^K)^T \mathbf{f}^* \ - \ \mathcal{F}(\mathbf{v}^K) = \frac{1}{2} (\mathbf{v}^K + \mathbf{v}^{K'})^T \mathbf{f}^* \ - \ \frac{1}{2} (\mathbf{v}^{K'} - \mathbf{v}^K)^T \mathbf{f}^* \ - \ \mathcal{F}(\mathbf{v}^K),$$

it follows that if the interface flux satisfies what is a multi-dimensional version of Tadmor's EC condition (2.14):

$$(\mathbf{v}^{K'} - \mathbf{v}^{K})^T \mathbf{f}^* = \mathcal{F}(\mathbf{v}^{K'}) - \mathcal{F}(\mathbf{v}^{K}), \tag{2.31}$$

then the second integrand in (2.30) makes for a consistent entropy flux $F^*(\mathbf{v}^K, \mathbf{v}^{K'}, \mathbf{n})$ given by:

$$F^* = (\mathbf{v}^K)^T \mathbf{f}^* \ - \ \mathcal{F}(\mathbf{v}^K) = \frac{1}{2}(\mathbf{v}^K + \mathbf{v}^{K'})^T \mathbf{f}^* \ - \ \frac{1}{2}(\mathcal{F}(\mathbf{v}^K) + \mathcal{F}(\mathbf{v}^{K'})), \tag{2.32}$$

and equation (2.30) becomes a semi-discrete *finite volume* discretization of the entropy equation:

$$\int_K \frac{\partial U(\mathbf{v}^K)}{\partial t}\, dV \ + \ \int_{\delta K} F^*\, dS = 0. \tag{2.33}$$

If the interface flux is of the form:

$$\mathbf{f}^*(\mathbf{v}^K, \mathbf{v}^{K'}, \mathbf{n}) = \mathbf{f}^{\mathbf{EC}}(\mathbf{v}^K, \mathbf{v}^{K'}, \mathbf{n}) - D(\mathbf{v}^K, \mathbf{v}^{K'}, \mathbf{n})(\mathbf{v}^{K'} - \mathbf{v}^K),$$

where $\mathbf{f}^{\mathbf{EC}}$ satisfies (2.31) and D is a positive definite matrix, then equation (2.30) writes instead:

$$\int_K \frac{\partial U(\mathbf{v}^K)}{\partial t}\, dV \ + \ \int_{\delta K} F^*\, dS = -\int_{\delta K} \mathcal{E}\, dS = 0. \tag{2.34}$$

where F^* is also given by (2.32) and:

$$\mathcal{E}(\mathbf{v}^K, \mathbf{v}^{K'}, \mathbf{n}) = \frac{1}{2}(\mathbf{v}^{K'} - \mathbf{v}^K)^T D(\mathbf{v}^K, \mathbf{v}^{K'}, \mathbf{n})(\mathbf{v}^{K'} - \mathbf{v}^K) > 0, \tag{2.35}$$

leading to a discretization that is consistent with the entropy inequality:

$$\frac{\partial U}{\partial t} + \frac{\partial F_i}{\partial x_i} < 0. \tag{2.36}$$

Conservation is ensured as long as the constant unity function belongs to the span of the basis functions $\phi_{i,K}$. In this regard, it does not matter which variables are discretized.

This shows that Tadmor's framework for first-order finite-volume schemes naturally extends to the DG setting when the entropy variables are discretized. It also naturally extends to the fully-discrete setting, we refer the reader to [51, 145] for more details.

In addition to providing more perspective on what discretizing the entropy variables means and entails, these derivations will allow us to make a clear point that the entropy production breakdown we introduce in chapter VI in the context of finite-volume schemes naturally fits in the DG setting. This breakdown can only take place in Tadmor's setting where entropy-stability is approached as "entropy conservation" + "entropy production".

CHAPTER III

Entropy Conservative Schemes and the Receding Flow Problem

One of the first questions that arise when studying entropy-stable schemes is: how much entropy should the scheme produce in the presence of discontinuities? This question is usually asked with shock discontinuities in mind, because they are the primary physical phenomenon that needs to be accurately predicted in compressible flows and also because they are entropy-producing discontinuities. But what about discontinuous flow configurations where entropy is expected to be conserved, not produced? Should the entropy-stable scheme revert to its entropy-conservative foundation to make for a good solution in these conditions?

The latter questions arose while investigating a simple one-dimensional Riemann problem: the receding flow problem (similar to the *123 problem* of Toro's book [168]) extensively studied by Liou [165, 166]. The problem consists of a flow undergoing rarefaction caused by two flows receding from each other. It is identified by Liou as an open numerical problem as many well-known finite-volume/finite-difference schemes produce an anomalous temperature rise, termed "overheating" or "Wall heating" [167, 162], at the origin that cannot be fixed by refining the mesh, decreasing the time-step, or increasing the solution order. He first established a connection between the overheating and a spurious entropy rise after the first time step [166]. In a more recent paper [165], he connected the entropy rise with the pressure component of the momentum flux. A cure he eventually proposed consists of replacing the conservation equation for total energy with either the transport equation for the specific entropy s or the conservation equation for the entropy ρs. The main liability of this approach is that conservation of total energy is no longer guaranteed and that conserving entropy is incompatible with shock discontinuities.

In view of these factors, entropy-conservative and entropy-stable schemes appear as an interesting option, as they enable the conservation or production of entropy without giving up conservation of mass, momentum and total energy. What's more, they were not considered in Liou's endeavours. An analysis similar to that of Liou [165] within the framework of Tadmor is therefore worth carrying out. While the semi-discrete analysis of Liou [165] suggests that the overheating could be avoided with an EC flux, numerical results say otherwise and this led us to carry out a fully-discrete analysis of the entropy behavior in this problem.

This chapter is organized as follows. In section 3.1 we introduce the problem. In section 3.2, we discuss Liou's analysis and its limits. In section 3.3, we carry out a fully-discrete analysis of the problem to better understand the generation of specific entropy and complement it with numerical results in section 3.4.

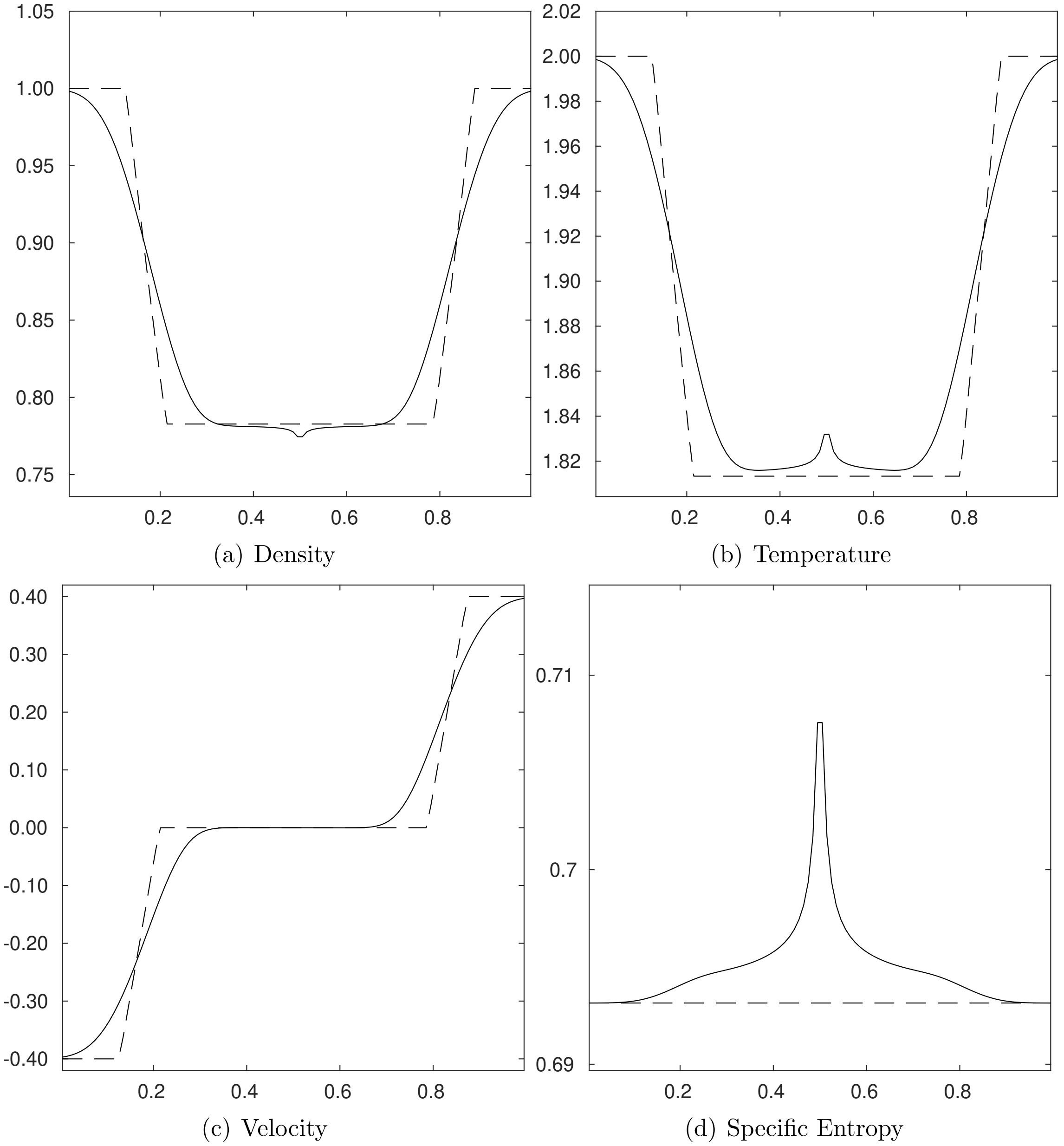
3.1 Problem description

The receding flow problem [165, 166] is a 1D Riemann problem defined by the following initial conditions:

$$u_L < 0,\ u_R > 0, \rho_L = \rho_R = \rho^0,\ p_L = p_R = p^0. \tag{3.1}$$

where the subscripts L and R refer to the left and right sides of the domain, respectively. Liou considered the case of equal velocity magnitudes $|u_L| = |u_R| = u^0$.

Liou describes this problem as “fundamental” in the sense that the overheating cannot be overcome by refining the mesh or changing the time step (it is independent of the CFL number). One of the main findings of his study is that the overheating originates from an *ab initio* entropy production at the beginning of the run. Figure 3.1 (Roe flux in space, forward euler in time) illustrates the numerical behavior that is typically observed with the wide range of fluxes Liou [165, 166] considered. The pressure is well resolved whereas the density is slightly under-estimated at the center (see figure 3.1-(a)). That the overheating is generated at the very first instant is intuitive given the nature of rarefaction waves (discontinuities that should vanish after some time) in contrast to shock waves (discontinuities that persist in time).



(a) Density

(b) Temperature

(c) Velocity

(d) Specific Entropy

Figure 3.1: Receding flow problem: Numerical solution (full line) and exact solution (dotted line) at $t = 0.18s$ with the Roe flux and Forward Euler in time. 100 cells and $\Delta t = 10^{-3}s$.

3.2 Liou's semi-discrete analysis

To investigate how entropy is initially produced in the discretized conservation laws, Liou [165] begins with the following equation:

$$\frac{p}{R_g}\frac{\partial s}{\partial t} = -\left(\frac{a^2}{\gamma - 1} - \frac{u^2}{2}\right)\frac{\partial \rho}{\partial t} - u\frac{\partial \rho u}{\partial t} + \frac{\partial \rho e^t}{\partial t},$$

where R_g is the gas constant. This equation relates the temporal variation of the specific entropy to that of mass, momentum and total energy.

Denote cell "R" as the cell immediately to the right of the interface with index 1 (its interfaces have therefore indices 1/2 and 3/2, respectively). Then integration over cell R at $t = 0$ gives:

$$\oint_R \frac{p}{R_g}\frac{\partial s}{\partial t} dV = \left(\frac{(a^0)^2}{\gamma - 1} - \frac{(u^0)^2}{2}\right)[(\rho u)_{3/2} - (\rho u)_{1/2}] + u^0[(\rho u^2 + p)_{3/2} - (\rho u^2 + p)_{1/2}] - [(\rho u h^t)_{3/2} - (\rho u h^t)_{1/2}]. \quad (3.2)$$

The fluxes at interface 3/2 are determined by the initial conditions:

$$(\rho u)_{3/2} = \rho^0 u^0,\ (\rho u^2 + p)_{3/2} = \rho^0 (u^0)^2 + p^0,\ (\rho u h^t)_{3/2} = \rho^0 u^0 (h^t)^0. \quad (3.3)$$

For all the fluxes tested by Liou [165], the values at the interface 1/2 are given by:

$$(\rho u)_{1/2} = 0,\ (\rho u^2 + p)_{1/2} = m_{1/2} + p_{1/2},\ (\rho u h^t)_{1/2} = 0. \quad (3.4)$$

$m_{1/2}$ and $p_{1/2}$ are the momentum and pressure fluxes [1], respectively. Combining equations (3.2), (3.3) and (3.4) results in the following:

$$\oint_R \frac{p}{R_g}\frac{\partial s}{\partial t} dV = u_0[(p^0 - p_{1/2}) - m_{1/2}]. \tag{3.5}$$

The right-hand side term remains non-zero for all the fluxes Liou considered [165]. An identical result is obtained for the "L" cell immediately to the left of the interface, meaning that the entropy rise occurs symmetrically about the interface. In light of equation (3.5), Liou attributed the *ab initio* generation of entropy to the pressure and momentum components of the numerical flux. He concluded his study by showing that replacing the energy equation with the transport equation for specific entropy or equivalently the conservation equation for entropy cures the overheating. This cure is not compatible with the simulation of shock discontinuities.

Liou's study did not consider Tadmor's family of schemes. If the EC Roe flux is used, the flux values at interface 1/2 take the values:

$$(\rho u)_{1/2} = 0,\ (\rho u^2 + p)_{1/2} = p^0,\ (\rho u h^t)_{1/2} = 0.$$

and we obtain, for both the R and L cells:

$$\oint \frac{p}{R_g}\frac{\partial s}{\partial t} dV = 0.$$

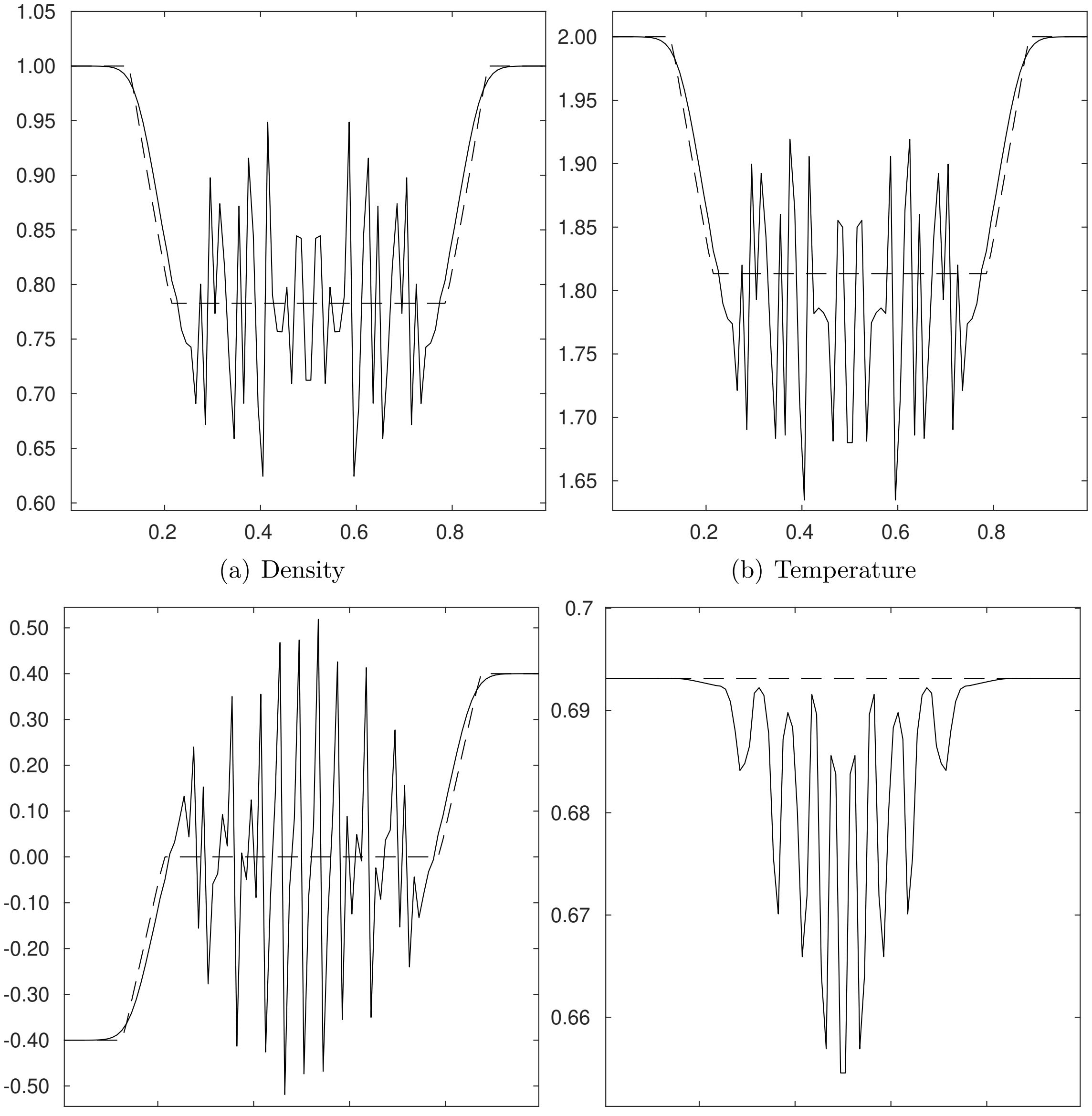
This suggests that the spurious entropy production would be avoided.

Unfortunately, these analytical results are not supported by numerical tests. In figure 3.2, we show the numerical solution with an EC flux in space together with a first order explicit time-integration scheme. The solution is oscillatory in all components and we note (figure 3.2-(d)) that the specific entropy profile is going downwards. In

[1] this breakdown of the second component of the interface flux is possible for all the schemes Liou considered. For EC/ES fluxes, it is not always possible but this does not matter in our analysis

figure 3.3, we show the numerical solution with a first order implicit time-integration scheme instead. The solution is better than with the explicit time scheme, but it is still oscillatory. We note that this time the specific entropy profile is going upwards.

These first results suggest that time-integration has a clear impact on the solution quality. A fully-discrete analysis is necessary.



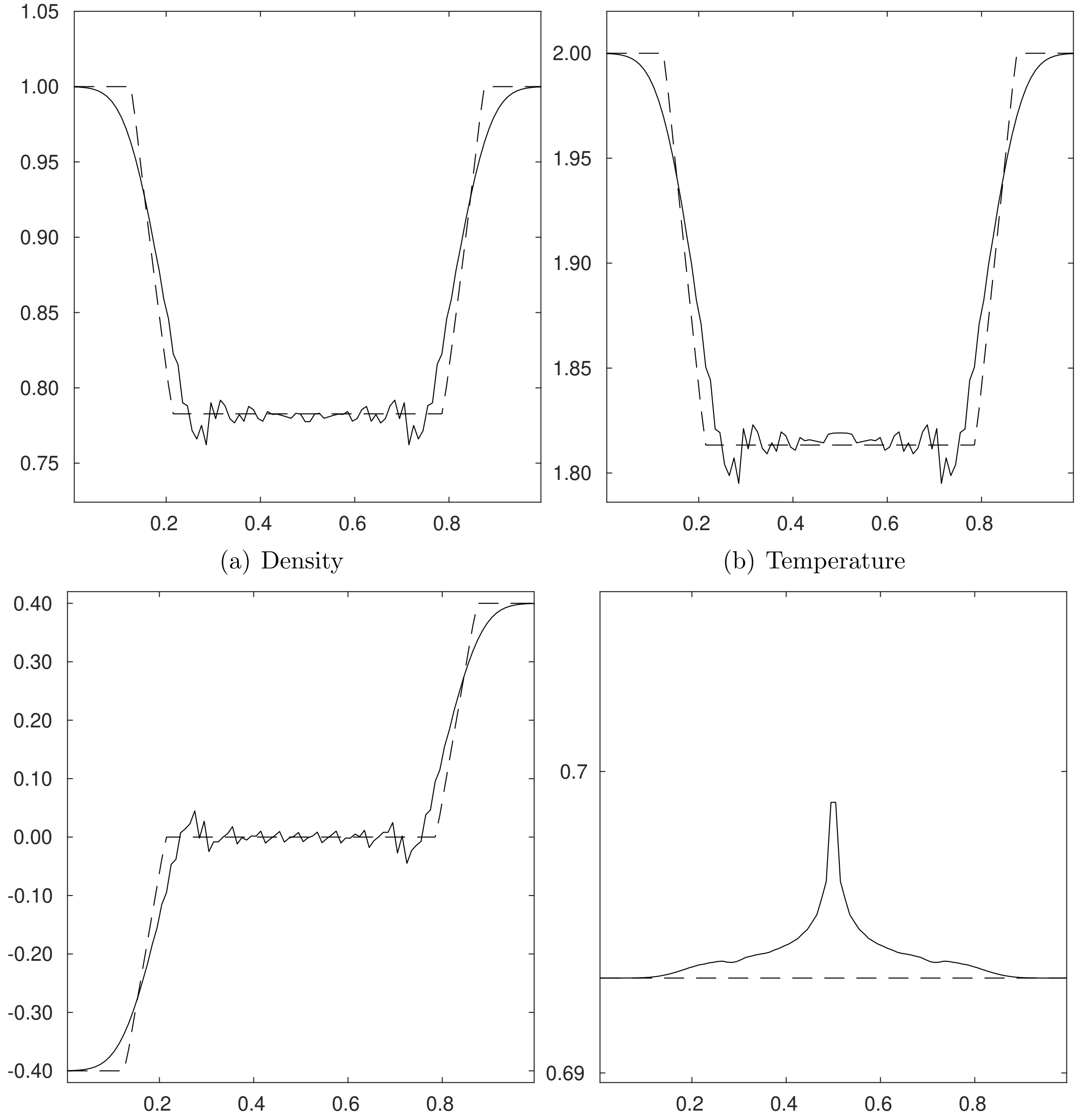
(a) Density

(b) Temperature

(c) Velocity

(d) Specific Entropy

Figure 3.2: Receding flow problem: Numerical solution (full line) at $t = 0.18s$ with the EC Roe flux and Forward Euler in time. 100 cells and $\Delta t = 10^{-3}s$.



(a) Density

(b) Temperature

(c) Velocity

(d) Specific Entropy

Figure 3.3: Receding flow problem: Numerical solution (full line) at $t = 0.18s$ with the EC Roe flux and Backward Euler in time. 100 cells and $\Delta t = 10^{-3}s$.

3.3 A fully-discrete analysis

3.3.1 Entropy stability of time schemes

Let's assume that an EC flux is used in space. What happens at the fully-discrete level, when time is discretized? If we evolve in time using Forward Euler (FE) for instance:

$$\mathbf{u}_j^{n+1} - \mathbf{u}_j^n + \lambda[\mathbf{f}_{j+\frac{1}{2}}^n - \mathbf{f}_{j-\frac{1}{2}}^n] = 0, \;\; \lambda = \frac{\Delta t}{\Delta x}, \tag{3.6}$$

are we simultaneously evolving in time the entropy equation with Forward Euler? In other words, is the fully-discrete scheme (3.6) also solving

$$U(\mathbf{u}_j^{n+1}) - U(\mathbf{u}_j^n) + \lambda[F_{j+\frac{1}{2}}^n - F_{j-\frac{1}{2}}^n] = 0 \ ? \tag{3.7}$$

In equation (3.7) and what follows, the superscript n refers to the time instant. We know that

$$(\mathbf{v}_j^n)^T[\mathbf{f}_{j+\frac{1}{2}}^n - \mathbf{f}_{j-\frac{1}{2}}^n] = F_{j+\frac{1}{2}}^n - F_{j-\frac{1}{2}}^n,$$

therefore the answer depends on whether

$$(\mathbf{v}_j^n)^T[\mathbf{u}_j^{n+1} - \mathbf{u}_j^n] = U(\mathbf{u}_j^{n+1}) - U(\mathbf{u}_j^n)$$

holds. Tadmor [51] showed that, for Forward Euler:

$$(\mathbf{v}_j^n)^T[\mathbf{u}_j^{n+1} - \mathbf{u}_j^n] = U(\mathbf{u}_j^{n+1}) - U(\mathbf{u}_j^n) - \mathcal{E}^{FE}(\mathbf{v}_j^n, \mathbf{v}_j^{n+1}), \tag{3.8a}$$

$$\mathcal{E}^{FE}(\mathbf{v}_j^n, \mathbf{v}_j^{n+1}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} (\frac{1}{2} + \xi)(\Delta \mathbf{v}_j^{n+\frac{1}{2}})^T H(\mathbf{v}_j^{n+\frac{1}{2}}(\xi)) \Delta \mathbf{v}_j^{n+\frac{1}{2}} d\xi > 0, \tag{3.8b}$$

where $\mathbf{v}_j^{n+\frac{1}{2}}(\xi) = \frac{\mathbf{v}_j^{n+1} + \mathbf{v}_j^n}{2} + \xi(\mathbf{v}_j^{n+1} - \mathbf{v}_j^n)$ and $\Delta \mathbf{v}_j^{n+\frac{1}{2}} = \mathbf{v}_j^{n+1} - \mathbf{v}_j^n$. This means that at the fully-discrete level:

$$U(\mathbf{u}_j^{n+1}) - U(\mathbf{u}_j^n) + \lambda[F_{j+\frac{1}{2}}^n - F_{j+\frac{1}{2}}^n] = \mathcal{E}^{FE} > 0. \tag{3.9}$$

This makes Forward Euler unconditionally entropy unstable. For Backward Euler, Tadmor [51] showed that:

$$(\mathbf{v}_j^{n+1})^T[\mathbf{u}_j^{n+1} - \mathbf{u}_j^n] = U(\mathbf{u}_j^{n+1}) - U(\mathbf{u}_j^n) + \mathcal{E}^{BE}(\mathbf{v}_j^n, \mathbf{v}_j^{n+1}), \tag{3.10a}$$

$$\mathcal{E}^{BE}(\mathbf{v}_j^n, \mathbf{v}_j^{n+1}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} (\frac{1}{2} - \xi)(\Delta \mathbf{v}_j^{n+\frac{1}{2}})^T H(\mathbf{v}_j^{n+\frac{1}{2}}(\xi))\Delta \mathbf{v}_j^{n+\frac{1}{2}} d\xi > 0. \tag{3.10b}$$

This means that at the fully-discrete level:

$$U(\mathbf{u}_j^{n+1}) - U(\mathbf{u}_j^n) + \lambda[F_{j+\frac{1}{2}}^{n+1} - F_{j+\frac{1}{2}}^{n+1}] = -\mathcal{E}^{BE} < 0 \tag{3.11}$$

This makes Backward Euler unconditionally entropy stable. One may wonder if all implicit and explicit time-integration schemes are unconditionally entropy stable and unstable, respectively. To the best of the author's knowledge, this is an open question. To support this statement, let's use the two main results of Tadmor's analysis, i.e. eqns. (3.8a) and (3.10a), to derive the entropy production of some well-known time-integration schemes. Define:

$$\mathbf{R}_j^f(u) = -\frac{1}{\Delta x}(\mathbf{f}_{j+\frac{1}{2}} - \mathbf{f}_{j-\frac{1}{2}}), \ \mathbf{R}_j^F(u) = \mathbf{v}_j^T \mathbf{R}_j^f(u) = -\frac{1}{\Delta x}(F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}}).$$

The implicit 2nd-order backward difference (BDF2) scheme is given by:

$$\mathbf{u}_j^{n+2} - \frac{4}{3}\mathbf{u}_j^{n+1} + \frac{1}{3}\mathbf{u}_j^n = \frac{2}{3}\Delta t \mathbf{R}_j^f(\mathbf{u}^{n+2}). \tag{3.12}$$

If we rewrite the scheme (3.12) as:

$$\frac{4}{3}(\mathbf{u}_j^{n+2} - \mathbf{u}_j^{n+1}) - \frac{1}{3}(\mathbf{u}_j^{n+2} - \mathbf{u}_j^n) = \frac{2}{3}\Delta t \mathbf{R}_j^f(\mathbf{u}^{n+2}),$$

left-multiply by $(\mathbf{v}_j^{n+2})^T$ and use equation (3.10a), we obtain the following for the discrete entropy:

$$U(\mathbf{u}_j^{n+2}) - \frac{4}{3}U(\mathbf{u}_j^{n+2}) + \frac{1}{3}U(\mathbf{u}_j^{n}) = \frac{2}{3}\Delta t\mathbf{R}_j^F(\mathbf{u}^{n+2}) - \mathcal{E}^{BDF2}$$

This is basically BDF2 for the discrete entropy, with an additional entropy production term $\mathcal{E}^{BDF2}$ given by:

$$\mathcal{E}^{BDF2} = \frac{4}{3}\mathcal{E}^{BE}(\mathbf{v}_j^{n+1}, \mathbf{v}_j^{n+2}) - \frac{1}{3}\mathcal{E}^{BE}(\mathbf{v}_j^{n}, \mathbf{v}_j^{n+2})$$

The production term $\mathcal{E}^{BE}(\mathbf{v}_j^{n}, \mathbf{v}_j^{n+2})$ can determine the entropy stability of BDF2. However, its sign is hard to establish. The explicit Leap-Frog Method is given by:

$$\mathbf{u}_j^{n+1} = \mathbf{u}_j^{n-1} + 2\Delta t\mathbf{R}_j^f(\mathbf{u}^n).$$

Subtracting $\mathbf{u}_j^n$ on both sides, left-multiplying by $(\mathbf{v}_j^n)^T$, and using eqns. (3.8a) and (3.10a), we get a Leap-Frog of the entropy

$$U(\mathbf{u}_j^{n+1}) = U(\mathbf{u}_j^{n-1}) + 2\Delta t\mathbf{R}_j^F(\mathbf{u}^n) - \mathcal{E}^{LF}$$

with an entropy production term $\mathcal{E}^{LF}$ given by:

$$\mathcal{E}^{LF} = \mathcal{E}^{BE}(\mathbf{v}_j^{n-1}, \mathbf{v}_j^{n}) - \mathcal{E}^{FE}(\mathbf{v}_j^{n}, \mathbf{v}_j^{n+1}).$$

Here again, it is hard to make a statement about the sign of the entropy production term $\mathcal{E}^{LF}$. We could derive similar results for other schemes and reach the same conclusion.

Going back to the receding flow problem, the results observed with EC fluxes can be now better explained. Let's assume that Forward Euler is used. We are interested in the jump of specific entropy $s^1 - s^0$ in the cell R after the first time step. We have the following discrete equation for density:

$$\rho^1 - \rho^0 = \frac{\Delta t}{\Delta x}[(\rho u)_{1/2} - (\rho u)_{3/2}] = -\frac{\Delta t}{\Delta x}\rho^0 u^0. \tag{3.13}$$

Using equation (3.9) with $U = -\rho s/(\gamma - 1)$ we obtain a discrete equation for the entropy ρs :

$$(\rho s)^1 - (\rho s)^0 = \frac{\Delta t}{\Delta x}[(\rho us)_{1/2} - (\rho us)_{3/2}] + (1-\gamma)\mathcal{E}^{FE}. \tag{3.14}$$

The interface flux for the entropy is given by equation (2.15). With the EC Roe flux, we have :

$$(\rho us)_{1/2} - (\rho us)_{3/2} = -\rho^0 u^0 s^0.$$

Eq. (3.14) then becomes:

$$(\rho s)^1 - (\rho s)^0 = -\frac{\Delta t}{\Delta x}\rho^0 u^0 s^0 + (1-\gamma)\mathcal{E}^{FE}. \tag{3.15}$$

Combining equations (3.15) and (3.13) gives:

$$(\rho s)^1 - (\rho s)^0 = s^0(\rho^1 - \rho^0) + (1-\gamma)\mathcal{E}^{FE}.$$

Regrouping, one obtains:

$$s^1 - s^0 = (1-\gamma)\mathcal{E}^{FE}/\rho^1 < 0. \tag{3.16}$$

Equation (3.16) is exact and shows that when the EC Roe flux is combined with Forward Euler in time, the specific entropy at the center is going to *decrease* after the first time instant.

With Backward Euler in time, the fluxes are evaluated at the next state in time so we cannot derive an exact and concise expression like (3.16). However, if the time step is small enough we can make reasonable approximations, namely:

$$\rho^1 - \rho^0 = \frac{\Delta t}{\Delta x}[(\rho u)_{1/2} - (\rho u)_{3/2}] \approx -\frac{\Delta t}{\Delta x}\rho^1 u^1, \quad \text{and} \quad (\rho us)_{1/2} - (\rho us)_{3/2} \approx -\rho^1 u^1 s^1.$$

Eq. (3.11) with $U = -\rho s/(\gamma - 1)$, then gives a discrete equation for entropy:

$$(\rho s)^1 - (\rho s)^0 = \frac{\Delta t}{\Delta x}[(\rho us)_{1/2} - (\rho us)_{3/2}] + (\gamma-1)\mathcal{E}^{BE} \approx s^1(\rho^1 - \rho^0) + (\gamma-1)\mathcal{E}^{BE}. \quad (3.17)$$

Regrouping, equation (3.17)writes:

$$s^1 - s^0 \approx (\gamma - 1)\mathcal{E}^{BE}/\rho^0 > 0, \quad (3.18)$$

and states that when the EC Roe flux is combined with Backward Euler in time, the specific entropy at the center is going to *increase* after the first time instant.

Eqs. (3.16) and (3.18) only describe the evolution of the numerical solution at the very first instant, but along with the specific entropy profiles (3.2)-(d) and (3.3) they seem to suggest that whether the specific entropy s increases or decreases at the center is correlated to the discrete entropy (ρs) production caused by the time scheme.

Ultimately, we should not forget the primary goal of this study that is (attempting) to prevent the overheating without giving up on conservation of total energy. In view of equations (3.16) and (3.18), the following question arises: What if a time-integration scheme that conserves entropy is used?

3.3.2 Conserving entropy in time

LeFloch *et al.* [50] (Theorem 3.1.) showed that the following scheme:

$$\frac{\mathbf{u}_j^{n+1} - \mathbf{u}_j^n}{\Delta t} = \mathbf{R}_j^f(\mathbf{u}(\mathbf{v}^{n+\frac{1}{2}})), \tag{3.19}$$

with $\mathbf{v}^{n+\frac{1}{2}}$ an intermediate state in time given by:

$$\mathbf{v}^{n+\frac{1}{2}} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbf{v}\bigg(\frac{\mathbf{u}^n + \mathbf{u}^{n+1}}{2} + \xi \Delta \mathbf{u}^{n+\frac{1}{2}}\bigg) d\xi,\ \Delta \mathbf{u}^{n+\frac{1}{2}} = \mathbf{u}^{n+1} - \mathbf{u}^n, \tag{3.20}$$

is entropy conservative in the sense that the scheme satisfies:

$$\frac{U(\mathbf{u}_j^{n+1}) - U(\mathbf{u}_j^n)}{\Delta t} = \mathbf{R}_j^F(\mathbf{u}(\mathbf{v}^{n+\frac{1}{2}})).$$

The scheme was also shown to be second-order accurate, in the sense that equation (3.19) is a second-order approximation to the system (2.1) evaluated at $t = \frac{t^{n+1}+t^n}{2}$.

Tadmor refers to this scheme as a Generalized Crank-Nicolson scheme in [51]. In the general nonlinear case, this scheme is impractical to the same extent as the first EC flux. The intermediate state does not have a closed form and requires quadrature, just like the first EC flux 2.16 Tadmor proposed.

The similarity between the intermediate state in time (3.20) and the first EC flux (2.16) is no coincidence. The condition on the intermediate state for the proposed scheme to be entropy conservative is (see Assumption 2.1. in [50] for $q = 1$):

$$(\mathbf{v}_j^{n+\frac{1}{2}})^T[\mathbf{u}_j^{n+1} - \mathbf{u}_j^n] = U(\mathbf{u}_j^{n+1}) - U(\mathbf{u}_j^n). \tag{3.21}$$

This equation is a time analog of the Entropy conservation condition in space (2.14). We can therefore apply the technique used to derive the EC Roe flux to derive an affordable intermediate state in time. Denote $\mathbf{v}^{n+\frac{1}{2}} = [v_1,\ v_2,\ v_3]$. Let's consider $\rho,\ u$

and p as the independent variables. The jumps in time can be written as:

$$[\rho u] = \bar{\rho}[u] + \bar{u}[\rho],\ [\rho e^t] = \frac{[p]}{\gamma - 1} + \frac{1}{2}[\rho u^2] = \frac{[p]}{\gamma - 1} + \frac{1}{2}\bar{u^2}[\rho] + \bar{\rho}\bar{u}[u], \tag{3.22}$$

$$[\rho s] = \bar{\rho}[s] + \bar{s}[\rho] = \bar{\rho}\frac{[p]}{p^{ln}} - \gamma\bar{\rho}\frac{[\rho]}{\rho^{ln}} + \bar{s}[\rho]. \tag{3.23}$$

Injecting equations (3.22) and (3.23) in equation (3.21) and regrouping we obtain:

$$[\rho](v_1 + \bar{u}v_2 + \frac{1}{2}\bar{u^2}v_3) + [u](\bar{\rho}v_2 + \bar{u}\bar{\rho}v_3) + [p](\frac{v_3}{\gamma - 1}) = \frac{-1}{\gamma - 1}([\rho](\bar{s} - \gamma\frac{\bar{\rho}}{\rho^{ln}}) + [p]\frac{\bar{\rho}}{p^{ln}}).$$

The jumps are independent, therefore:

$$v_1 + \bar{u}v_2 + \frac{1}{2}\bar{u^2}v_3 = \frac{1}{\gamma - 1}(\gamma - \bar{s}\frac{\bar{\rho}}{\rho^{ln}}),\ \bar{\rho}v_2 + \bar{u}\bar{\rho}v_3 = 0,\ v_3 = -\frac{\bar{\rho}}{p^{ln}}.$$

The solution is:

$$v_1 = \frac{1}{\gamma - 1}(\gamma\frac{\bar{\rho}}{\rho^{ln}} - \bar{s}) - \bar{u}v_2 - \frac{1}{2}\bar{u^2}v_3,\ v_2 = -\bar{u}v_3,\ v_3 = -\frac{\bar{\rho}}{p^{ln}}. \tag{3.24}$$

This intermediate state satisfies condition (3.21) and is consistent. Let's show that the resulting scheme is second-order as well. A Taylor expansion about $t = \frac{t^n + t^{n+1}}{2}$ gives:

$$\frac{\mathbf{u}_j^{n+1} - \mathbf{u}_j^n}{\Delta t} = \partial_t \mathbf{u}\left(x_j, \frac{t^n + t^{n+1}}{2}\right) + \mathcal{O}(\Delta t^2).$$

To conclude, let's show that:

$$\mathbf{v}^{n+\frac{1}{2}} = \mathbf{v}\left(\frac{t^n + t^{n+1}}{2}\right) + \mathcal{O}(\Delta t^2).$$

Let's establish a few results first. Let $a(t)$ be a strictly positive time-dependent quantity and denote $a^n = a(t^n)$, $a^{n+1} = a(t^{n+1})$ and $a^* = a(\frac{t^n + t^{n+1}}{2})$. Using a Taylor analysis about $t = \frac{t^n + t^{n+1}}{2} = t^n + \frac{\Delta t}{2} = t^{n+1} - \frac{\Delta t}{2}$, one can show that:

$$\bar{a} = a^* + \mathcal{O}(\Delta t^2),\ \bar{a^2} = (a^*)^2 + \mathcal{O}(\Delta t^2), \tag{3.25}$$

$$a^{n+1} - a^n = \Delta t a^*_{,t} + \mathcal{O}(\Delta t^3),\ log(a^{n+1}) - log(a^n) = \Delta t \frac{a^*_{,t}}{a^*} + \mathcal{O}(\Delta t^3). \tag{3.26}$$

Therefore:

$$a^{ln} = \frac{a^{n+1} - a^n}{log(a^{n+1}) - log(a^n)} = \frac{\Delta t a^*_{,t} + \mathcal{O}(\Delta t^3)}{\Delta t \frac{a^*_{,t}}{a^*} + \mathcal{O}(\Delta t^3)} = \frac{a^* + \mathcal{O}(\Delta t^2)}{1 + \mathcal{O}(\Delta t^2)} = a^* + \mathcal{O}(\Delta t^2)$$

Likewise we show another useful identity:

$$\frac{\bar{a}}{b^{ln}} = \frac{a^*}{b^*} + \mathcal{O}(\Delta t^2), \tag{3.27}$$

where $b(t)$ is another strictly positive quantity. With equations (3.25) and (3.27) we can show that the nonlinear intermediate state $\mathbf{v}_{n+\frac{1}{2}} = [v_1,\ v_2,\ v_3]$ defined by equation (3.24) satisfies:

$$v_1 = \frac{\gamma - s^*}{\gamma - 1} - \frac{1}{2}\frac{\rho^*(u^*)^2}{p^*} + \mathcal{O}(\Delta t^2),\ v_2 = \frac{\rho^* u^*}{p^*} + \mathcal{O}(\Delta t^2),\ v_3 = -\frac{\rho^*}{p^*} + \mathcal{O}(\Delta t^2).$$

This is not the first time scheme of this type. Ray [161] developed pretty much the same scheme[2] in his thesis, building from Subbareddy & Candler's work [112] on fully-discrete kinetic energy preserving schemes

3.4 Numerical results

Figure 3.4 shows the numerical solution with the EC time scheme for a grid of 100 elements and a time-step $\Delta t = 10^{-3}s$. Unfortunately, completely conserving entropy does not solve the problem. Oscillations are observed again and their magnitude is

[2] much to the disillusionment of the author, who thought that he came up with something new.

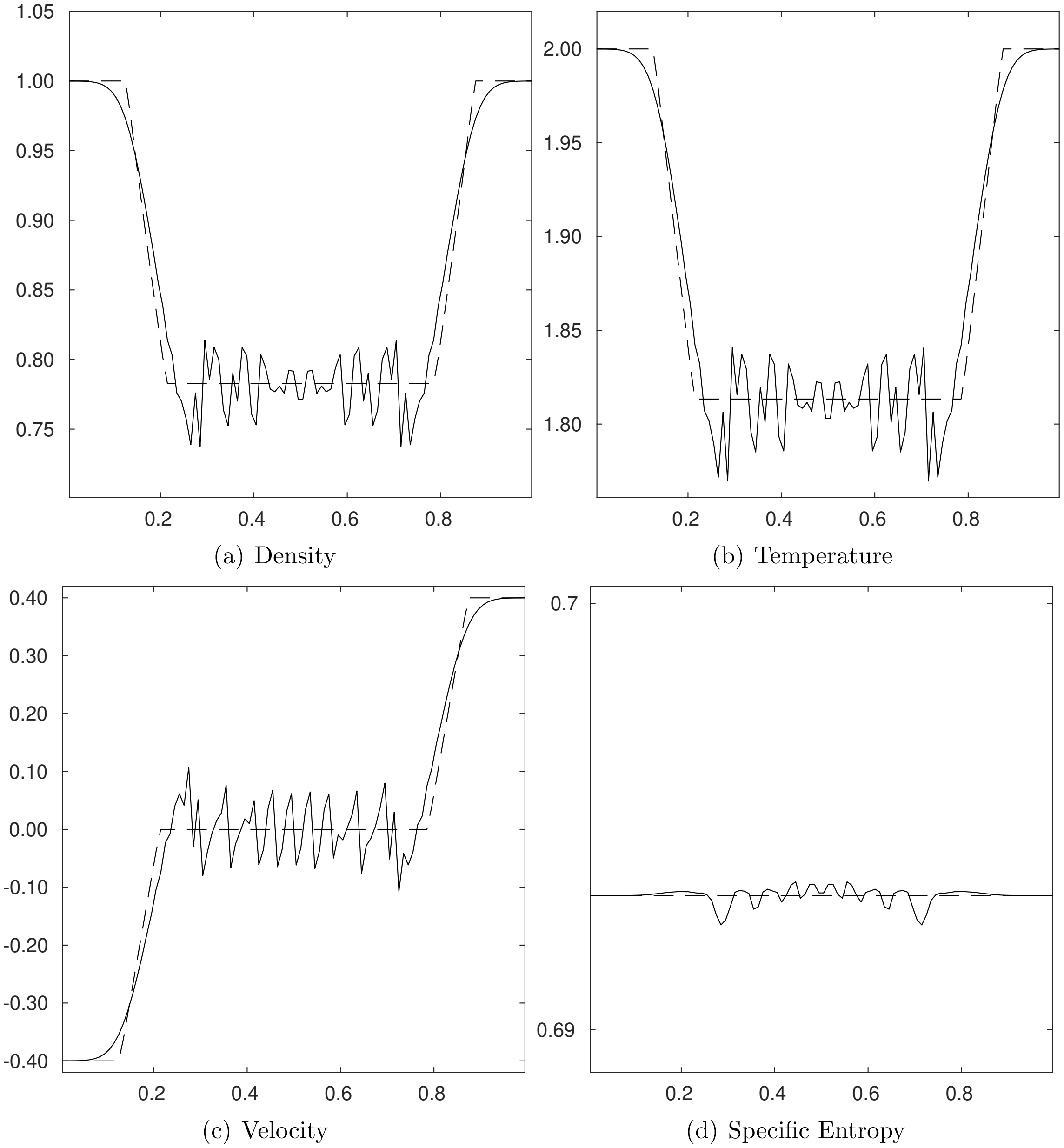
higher than when Backward Euler in time is used (fig. (3.3)), but lower than with Forward Euler in time (fig. (3.2)).

Figure 3.5 shows the production of entropy over time in all three cases. It confirms the entropy stability properties of Forward Euler and Backward Euler covered in section 3.3 and the entropy conservation property of the present time scheme. The rate at which Backward Euler produces entropy decreases with time. This is because, as time goes by, the oscillations caused by the EC flux in space are damped by the dissipation of Backward Euler and the numerical solution becomes smoother. On the other hand, the entropy losses incurred by using Forward Euler in time keep growing with time. The oscillations keep growing and will eventually lead to unphysical values of pressure/density. In the fully EC case, the oscillations persist in time, but they do not grow in magnitude.

If we look at the density, velocity and temperature profiles, the better results are obtained with Backward Euler, but if we look at the specific entropy profile, the entropy conservative scheme produces the most accurate result.

Figures (3.6) and (3.7) show the numerical solution with Backward Euler and the EC time scheme, respectively, with finer resolution but same CFL as before. The Forward Euler calculation crashed. With Backward Euler, pretty much all of the oscillations have disappeared. The overheating is visible, and overshoots/undershoots in the velocity, density and temperature fields can be seen behind the rarefactions. With the EC time scheme, the magnitude of the oscillations is smaller but their frequency is higher. What is striking is the specific entropy profile (fig. 3.7-(d)) which displays an intriguing structure. There is a spike at the center which links to two localized regions behind the rarefaction waves where the specific entropy drops. Figure 3.8, which features snapshots of the specific entropy profile over time, shows that this structure is conserved over time.

All in all, we conclude that enforcing conservation of entropy, even when it is a property of the exact solution, does not necessarily lead to a better behaved numerical solution.



(a) Density

(b) Temperature

(c) Velocity

(d) Specific Entropy

Figure 3.4: Receding flow problem: Numerical solution (full line) at $t = 0.18s$ with the EC Roe flux and the EC scheme in time. 100 cells and $\Delta t = 10^{-3}s$.

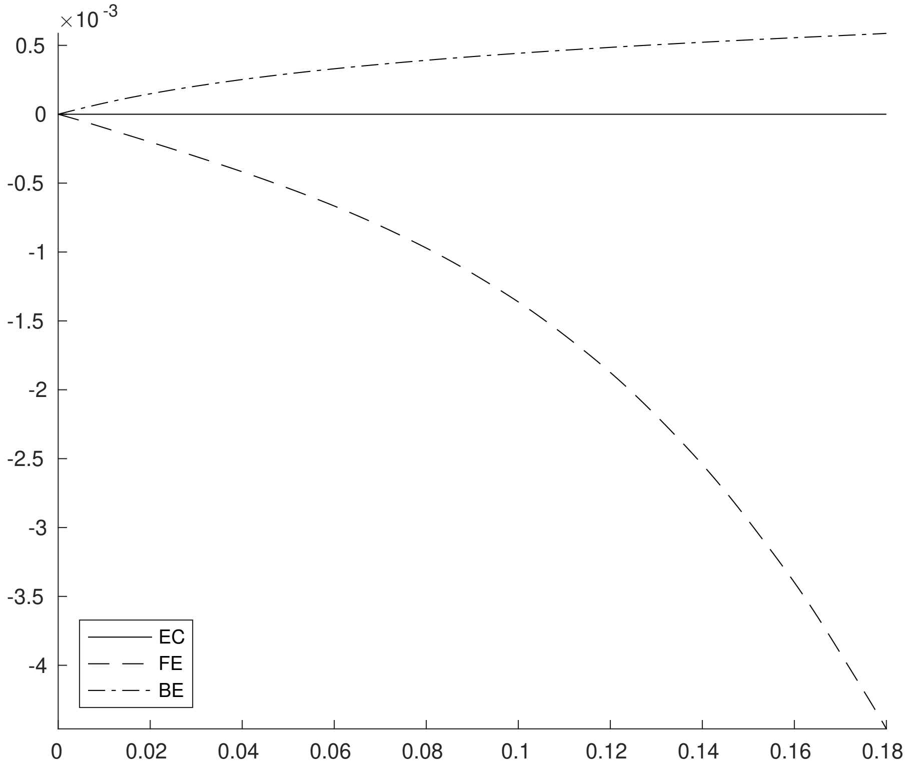
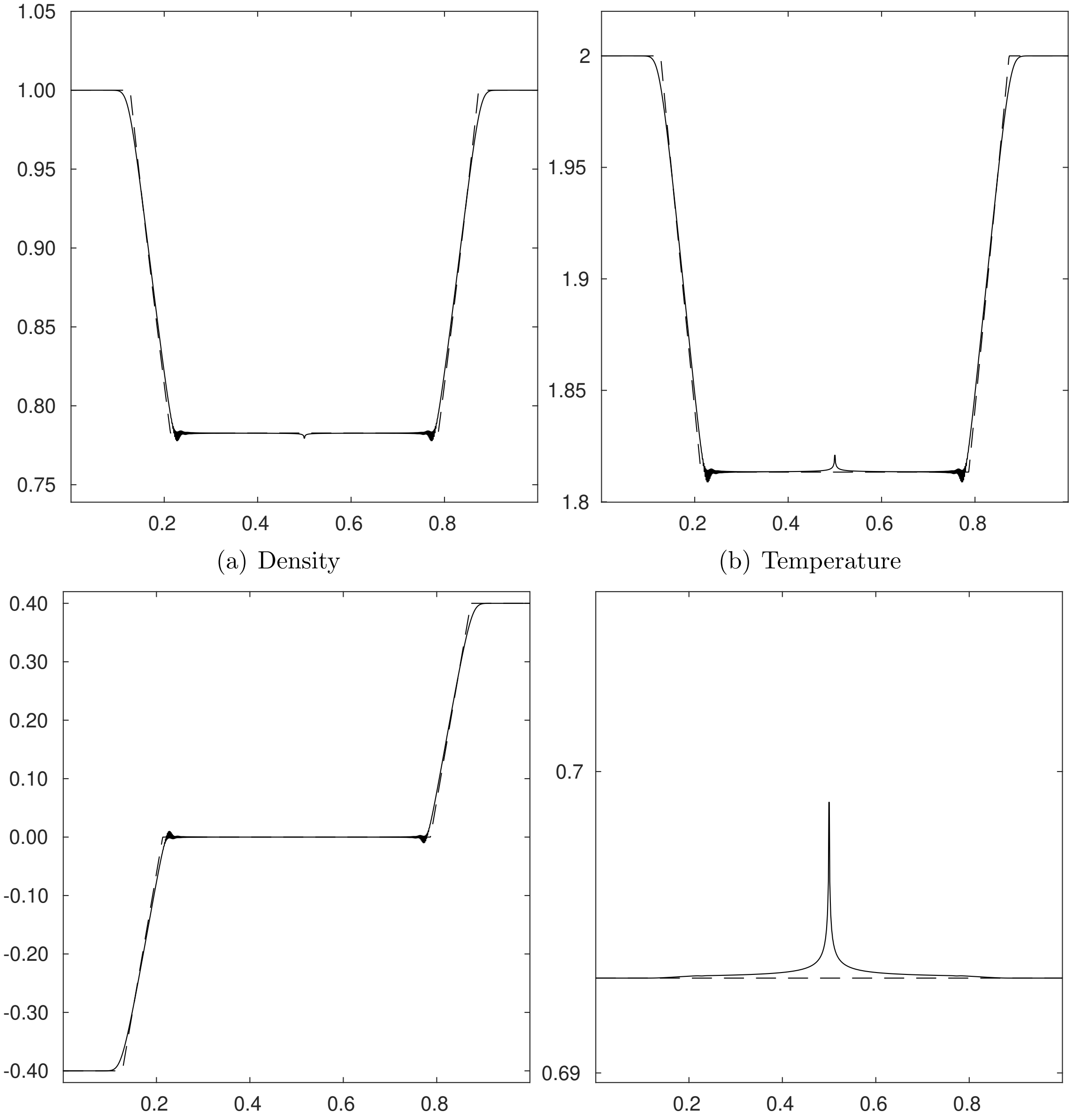


Figure 3.5: Receding flow problem: Total Entropy (ρs) production over time. The EC Roe flux in used in space. EC: Our EC scheme in time. FE: Forward Euler in time. BE: Backward Euler in time. 100 cells and $\Delta t = 10^{-3}s$.



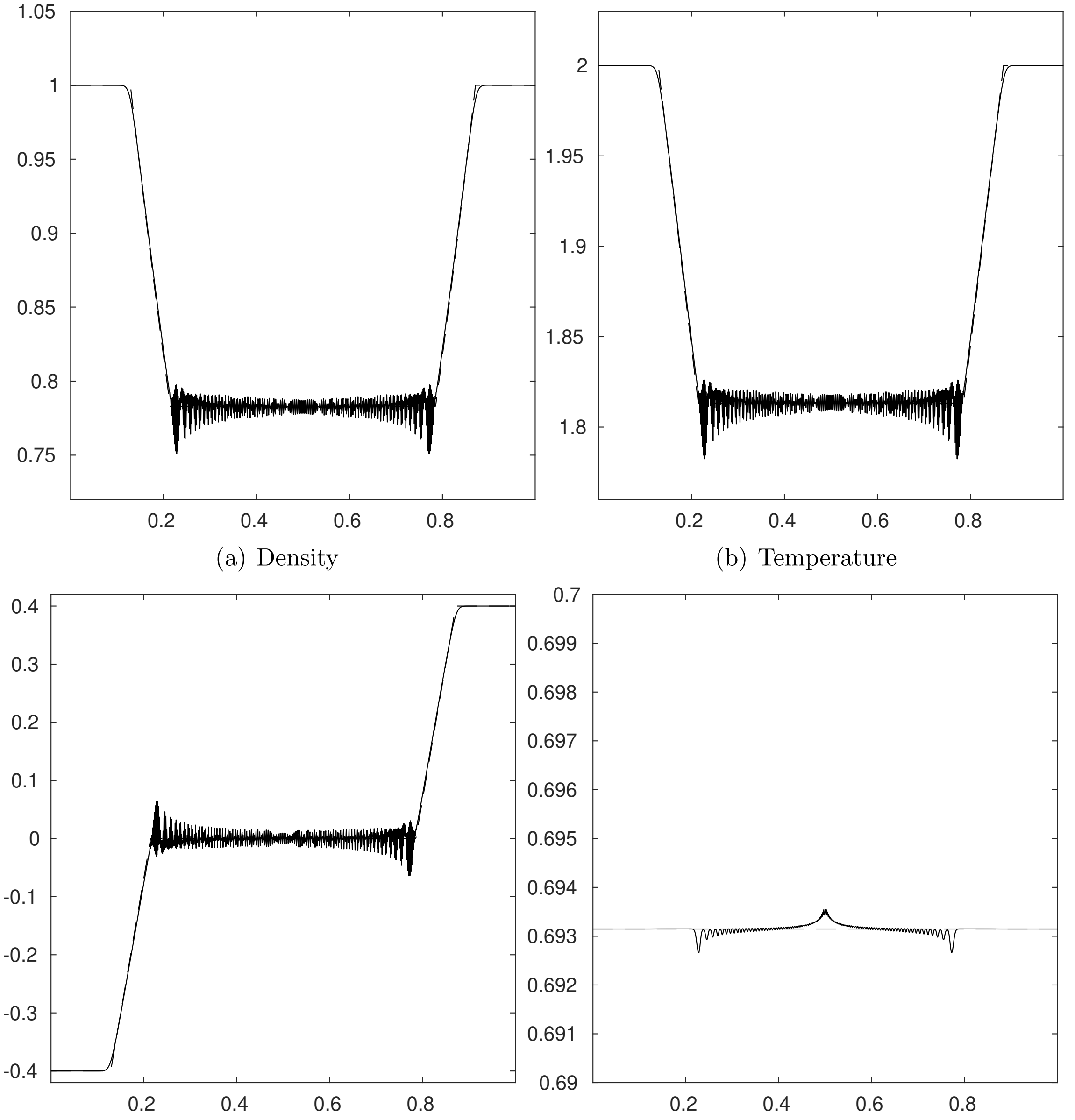
(a) Density

(b) Temperature

(c) Velocity

(d) Specific Entropy

Figure 3.6: Receding flow problem: Numerical solution (full line) at $t = 0.18s$ with the EC Roe flux and Backward Euler in time. 1000 cells and $\Delta t = 10^{-4}s$.



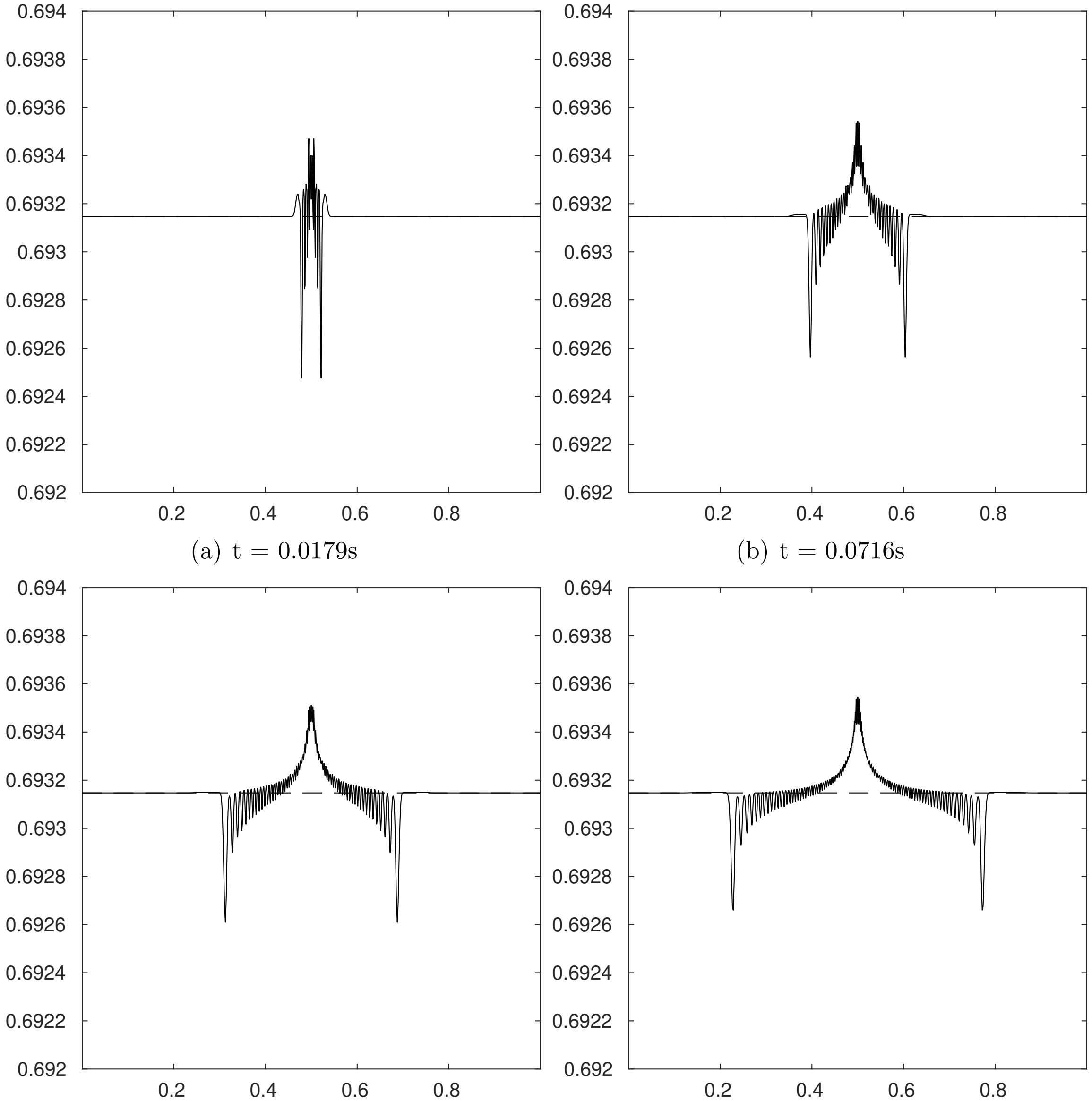
(a) Density

(b) Temperature

(c) Velocity

(d) Specific Entropy

Figure 3.7: Receding flow problem: Numerical solution (full line) at $t = 0.18s$ with the EC Roe flux and our EC scheme in time. 1000 cells and $\Delta t = 10^{-4}s$.



(a) t = 0.0179s

(b) t = 0.0716s

(c) t = 0.1253s

(d) t = 0.179s

Figure 3.8: Receding flow problem: Snapshots of the specific entropy profile when the EC Roe flux is used in space and our EC scheme is used in time. 1000 cells and $\Delta t = 10^{-4}s$.

3.5 Summary

In this chapter, we investigated entropy conservative schemes as a possible remedy in the receding flow problem. This was motivated by Liou's latest study that showed the connection between the overheating and a spurious entropy production *ab initio*.

Liou's analysis suggested that the EC flux of Roe would prevent the overheating, but this was not the case because time-integration was not accounted for. Indeed, a fully-discrete analysis leveraging entropy-stability theory confirmed it. It appears that whether the specific entropy spuriously increases or decreases depends on whether the scheme produces entropy at the fully-discrete level. The analysis also brought the question of how a fully-discrete entropy conservative scheme would perform. Building on the analogy between the entropy conservation condition for the spatial fluxes and the entropy conservation condition of a class of time-integration schemes introduced by LeFloch *et. al* [50], we derived an entropy conservative time-integration scheme which combined with an EC flux in space achieves fully-discrete entropy conservation. However we observed that it does not cure the overheating problem either. To be more specific, a better specific entropy profile is obtained but the oscillatory nature of the numerical solution does not make it a practical option.

Whether all entropy conservative discretizations would have the same unsatisfactory behavior on this type of problem, where one expects the continuous solution to conserve entropy, is a question that requires further investigation. The EC scheme that has been developed in this paper is just one way among many to conserve entropy in addition to mass, momentum and energy at the fully-discrete level. Recall that for systems ($N > 1$) there is more than one possible EC flux. Likewise, the intermediate state in time we used in our EC time scheme is just one choice among many. The time scheme given by equation (3.19) is part of a more general class of entropy conservative scheme that LeFloch *et al.* introduced in [50].

While one can arguably take the stance that fully-discrete EC schemes will produce a similar behavior to that in figures 3.4 and 3.7, it is known from past work [63, 163] that all EC fluxes do not perform equally. Besides simplicity, one of the reasons why the EC Roe flux is preferred over the first EC flux of Tadmor is that the latter does not preserve stationary contact discontinuities. Chandrasekhar [63] introduced an EC flux that has the additional property of discretely preserving, in the sense of Jameson [64], the kinetic energy of the system. This type of property is often sought when turbulent flows are simulated [112].

Another metric is how good of a foundation an EC flux constitutes in an ES scheme. The dissipation component of entropy-stable fluxes is often seen as the complement needed by EC fluxes in the presence of shocks. An EC scheme will produce non-physical solutions (oscillations) in the presence of shocks because entropy is not produced. This picture is correct but incomplete. In the presence of rarefaction waves and moving contact discontinuities, which do not physically require any production of entropy, EC schemes have the same oscillatory behavior. The receding flow problem is an illustration. This places an additional burden on the dissipation term which has to make up for the flaws of its foundation. A case in point can be found in Derigs *et al.* [163] where it was showed that entropy-stable schemes perform better on high-pressure shock problems if Chandrasekhar's EC flux is used instead of Roe's.

CHAPTER IV

Formulation of an Entropy-Stable Scheme for the Compressible Multicomponent Compressible Euler Equations

In this chapter, we consider the multicomponent (N species) compressible Euler equations consisting of the conservation equations for partial densities, momentum and total energy. This system can be viewed as the Euler equations (conservation of total mass, momentum and total energy) complemented with $N-1$ species conservation equations. This observation motivated early multicomponent schemes such as the one by Habbal *et al.* [183], where the Roe scheme [10] is applied to the Euler part of the equations and the $N-1$ remaining equations are treated separately. In Larrouturou [182], such approaches are termed uncoupled as opposed to fully coupled approaches which treat the multicomponent system as a whole. An example of a fully coupled approach is the extension of the Roe scheme by Fernandez and Larroutouru [181] and Abgrall [179]. It might be then tempting to use an existing entropy-stable scheme for the Euler part and evolve the remaining $N-1$ species equations with another scheme. A case in point can be found in Derigs *et al.* [196] (section 3.8). While this approach has the benefit of simplicity (minimal programming and computational effort), it is lacking from a theoretical viewpoint. This approach implicitly assumes that the entropy of the single component system is an admissible entropy for the multicomponent system, meaning that it is a conserved quantity in the smooth regime and a convex function of the conserved variables. That is not the case. The necessity of a fully coupled approach to develop entropy-stable schemes, in the sense of Tadmor [49], for multicomponent flows is motivated by both mathematical and physical arguments.

At this juncture, it is important to recall that the schemes we are interested in this thesis achieve entropy-stability in a specific way. That is to say that there is more than one way that a scheme can be made stable in an entropy sense, and hence be called 'entropy-stable'. Osher's family of E-schemes [72] and Barth's space-time discontinuous galerkin schemes [59] are conservative schemes which satisfy an entropy inequality, but their construction is different. There are also non-conservative schemes which can be designed to conserve or produce entropy [198]. Entropy stability can be understood in a different way as well. The scheme introduced by Ma *et al.* [197] for multicomponent flows is termed entropy-stable because it enforces a minimum principle of the specific entropy, proved by Tadmor for the Euler equations [175]. In their work, stability is sought in a point-wise sense (a scheme which preserves the positivity of density and satisfies the minimum principle cannot crash in principle), not in an integral sense. Integral stability and point-wise stability are both important concepts, and in principle they do not imply each other. In either case, it is important to emphasize that *the correct formulation of these schemes depends on the structure of the equations they are applied to.* Chalot *et al.* [174] and Giovangigli [172] demonstrated that the multicomponent compressible Euler equations do possess the structure that ES schemes require. It is not clear to the author whether these results extend to the systems considered in [197].

This chapter is organized as follows. In section 4.1, we present the modeling assumptions, the governing equations and their symmetrization using the entropy variables [174, 172]. Section 4.2 is dedicated to the construction of ES schemes for multicomponent flows. We formulate an EC flux and an ES interface flux based on upwinding [52] and discuss their definition in the limit of vanishing partial densities. In section 4.3, we discuss how this limit impacts standard high-order ES discretizations. In section 4.4, the numerical scheme is tested on one-dimensional and two-dimensional interface and shock-interface problems.

4.1 Governing equations, entropy variables and symmetrization

The governing equations describe the conservation of species mass, momentum and total energy. In 1D, that is system (2.1) with the vector of conserved variables $\mathbf{u}$ and the vector of fluxes $\mathbf{f}$ defined by:

$$\mathbf{u} := \begin{bmatrix} \rho_1 & \dots & \rho_N & \rho u & \rho e^t \end{bmatrix}^T, \ \mathbf{f} := \begin{bmatrix} \rho_1 u & \dots & \rho_N u & \rho u^2 + p & (\rho e^t + p)u \end{bmatrix}^T.$$

ρ_k denotes the partial density of species k, $\rho := \sum_{k=1}^{N} \rho_k$ denotes the total density, u denotes the velocity and e^t denotes the specific total energy. The pressure p is given by the ideal gas law:

$$p := \sum_{k=1}^{N} \rho_k r_k T, \ r_k = \frac{R}{m_k},$$

where m_k is the molar mass of species k, R is the gas constant. The temperature T is determined by the internal energy $\rho e := \rho e^t - (\rho u)^2/(2\rho)$ which in this work is modeled following a calorically perfect gas assumption:

$$\rho e = \sum_k \rho_k e_k, \ e_k := e_{0k} + c_{vk} T. \tag{4.1}$$

For species k, e_{0k} is a constant and c_{vk} is the constant volume specific heat. T is computed by solving equation (4.1). For later use, we introduce the species mass fraction $Y_k := \rho_k/\rho$, the species constant pressure specific heats $c_{pk} := c_{vk} + r_k$ and the specific heat ratio $\gamma := \big(\sum_k Y_k c_{pk}\big)/\big(\sum_k Y_k c_{vk}\big)$.

An additional conservation equation [174, 172] can be derived from the governing equations:

$$\frac{\partial \rho s}{\partial t} + \frac{\partial \rho u s}{\partial x} = 0. \tag{4.2}$$

ρs is the thermodynamic entropy of the mixture given by:

$$\rho s := \sum_{k=1}^{N} \rho_k s_k, \;\; s_k := c_{vk} \ln(T) - r_k \ln(\rho_k),$$

s denotes the specific entropy of the mixture and s_k denotes the specific entropy of species k. Equation (4.2) can be rewritten in the form of equation (2.3) with $(U,\ F) = (-\rho s,\ -\rho u s)$. For $\rho_k > 0$ and $T > 0$, U is a convex function of the conserved variables and $(U,\ F)$ is a valid entropy pair for the multicomponent system [174, 172] (more details can be found in the next chapter).

In order to derive the flux potentials $\mathcal{U}$ and $\mathcal{F}$, we first derive the entropy variables. Following [174, 172], we use a chain rule:

$$\frac{\partial U}{\partial \mathbf{u}} = \frac{\partial U}{\partial Z}\left(\frac{\partial \mathbf{u}}{\partial Z}\right)^{-1},\; Z := \begin{bmatrix} \rho_1 & \dots & \rho_N & u & T \end{bmatrix}^T,$$

where Z denotes the vector of primitive variables. The Gibbs identity can be written as:

$$Td(\rho s) = d(\rho e) - \sum_{k=1}^{N} g_k d\rho_k, \tag{4.3}$$

where $g_k := h_k - Ts_k$ is the Gibbs function of species k and $h_k := e_k + r_k T$ is the specific enthalpy of species k. We have:

$$d(\rho e) = \sum_{k=1}^{N} e_k d\rho_k + \rho c_v dT,\;\; \rho c_v := \sum_{k=1}^{N} \rho_k c_{vk}. \tag{4.4}$$

Combining eqs. (4.4) and (4.3), one obtains:

$$d(\rho s) = \frac{1}{T}\bigg(\sum_{k=1}^{N} (e_k - g_k) d\rho_k + \rho c_v dT \bigg).$$

This gives:

$$\frac{\partial U}{\partial Z} = \frac{1}{T} \Big[(g_1 - e_1) \quad \dots \quad (g_N - e_N) \quad 0 \quad -\rho c_v \Big]. \tag{4.5}$$

The Jacobian of the mapping $Z \to \mathbf{u}$ is given by:

$$\frac{\partial \mathbf{u}}{\partial Z} = \begin{bmatrix} 1 & & 0 & 0 & 0 \\ & \ddots & & \vdots & \vdots \\ 0 & & 1 & 0 & 0 \\ u & \dots & u & \rho & 0 \\ e_1 + \frac{1}{2}u^2 & \dots & e_N + \frac{1}{2}u^2 & \rho u & \rho c_v \end{bmatrix}. \tag{4.6}$$

The inverse of this matrix is given by:

$$\left(\frac{\partial \mathbf{u}}{\partial Z} \right)^{-1} = \begin{bmatrix} 1 & & 0 & 0 & 0 \\ & \ddots & & \vdots & \vdots \\ 0 & & 1 & 0 & 0 \\ -u\rho^{-1} & \dots & -u\rho^{-1} & \rho^{-1} & 0 \\ (\frac{1}{2}u^2 - e_1)(\rho c_v)^{-1} & \dots & (\frac{1}{2}u^2 - e_N)(\rho c_v)^{-1} & -u(\rho c_v)^{-1} & (\rho c_v)^{-1} \end{bmatrix} \tag{4.7}$$

Combining eqs. (4.7) and (4.5) yields the entropy variables [174, 172]:

$$\mathbf{v} = \left(\frac{\partial U}{\partial \mathbf{u}} \right)^T = \frac{1}{T} \Big[g_1 - \tfrac{1}{2}u^2 \quad \dots \quad g_N - \tfrac{1}{2}u^2 \quad u \quad -1 \Big]^T. \tag{4.8}$$

From this expression, the potential functions ($\mathcal{U},\ \mathcal{F}$) can be derived:

$$\mathcal{U} = \sum_{k=1}^{N} r_k \rho_k,\ \mathcal{F} = \sum_{k=1}^{N} r_k \rho_k u. \tag{4.9}$$

We conclude this section with two remarks:

- Denote $\mathbf{v} = [v_{1,k}\ \dots\ v_{1,N}\ v_2\ v_3]^T$. In order to derive the mapping from entropy variables to primitive variables, one can first compute the temperature, velocity, gibbs functions and specific entropies as:

  $$T := -\frac{1}{v_3},\ u := -\frac{v_2}{v_3},\ g_k := -\frac{v_{1,k}}{v_3} + \frac{1}{2}\left(\frac{v_2}{v_3}\right)^2,\ s_k(T, \rho_k) = c_{pk} - v_{1,k} + \frac{1}{2}\frac{v_2^2}{v_3}$$

  The partial densities are inferred from the specific entropies:

  $$\rho_k := \exp\left(\frac{m_k}{R}(c_{vk} ln(T) - s_k)\right) = \exp\left(\frac{m_k}{R}\left(- c_{vk} ln(v_3) - c_{pk} + v_{1,k} - \frac{1}{2}\frac{v_2^2}{v_3}\right)\right).$$

  The requirement that $\rho_k > 0$ and $T > 0$ manifests in the definition of the entropy variables, which require the evaluation of $\ln(\rho_k)$ and $\ln(T)$. On the other hand, it is interesting to note that if one works with the entropy variables instead of the conservative variables, the requirement that $\rho_k > 0$ and $T > 0$ boils down to the single requirement that $v_3 < 0$. The remaining entropy variables can be of any sign in principle. The author is not aware of any physical consideration which would impose the sign of $g_k - \frac{1}{2}u^2$, namely the difference between gibbs energy and kinetic energy.

- In the compressible Euler case with $e_{0k} = 0$, it is easy to show, using the ideal gas law $p = \rho r T$ and the Mayer relation $c_p - c_v = r$ that the vector of entropy variables (4.8) simplifies to :

  $$\mathbf{v} = r \begin{bmatrix} \frac{\gamma - \bar{s}}{\gamma - 1} - \frac{\rho u^2}{2p} & \frac{\rho u}{p} & -\frac{\rho}{p} \end{bmatrix}^T,\ \bar{s} = \ln p - \gamma \ln \rho - \ln r.$$

This differs by a constant factor r from the expression that is usually used when designing entropy-stable schemes for the compressible Euler equations. The trivial difference comes from a different choice of entropy pair $(U,\ F)$.

4.2 Formulation

4.2.1 Entropy-conservative flux

In compact notation, the entropy conservation condition (2.14) writes:

$$[\mathbf{v}]^T\mathbf{f}^* = [\mathcal{F}], \tag{4.10}$$

where $\mathbf{f}^* = [f_{1,1}\ \dots\ f_{1,N}\ f_2\ f_3]$ denotes the interface flux. Define the set of algebraic variables:

$$\mathbf{z} = \begin{bmatrix}\rho_1 & \dots & \rho_N & u & \frac{1}{T}\end{bmatrix} = \begin{bmatrix}z_{1,1} & \dots & z_{1,N} & z_2 & z_3\end{bmatrix}.$$

The jump in the potential function can be rewritten as:

$$[\mathcal{F}] = \sum_{k=1}^{N} r_k[\rho_k u] = \left(\sum_{k=1}^{N} r_k\overline{z_{1,k}}\right)[z_2] + \sum_{k=1}^{N} r_k\overline{z_2}[z_{1,k}]. \tag{4.11}$$

For the jump in entropy variables, we need to examine the first N components. For $1 \leq k \leq N$:

$$\begin{aligned}\frac{1}{T}(g_k - \frac{1}{2}u^2) = {} & \frac{e_{0k}}{T} + c_{vk} + r_k - c_{vk}\ln T + r_k\ln(\rho_k) - \frac{1}{2}\frac{u^2}{T}\\ & = e_{0k}z_3 + c_{vk} + r_k + c_{vk}\ln(z_3) + r_k\ln(z_{1,k}) - \frac{1}{2}z_2^2 z_3.\end{aligned}$$

The corresponding jumps then write:

$$\left[\frac{1}{T}(g_k - \frac{1}{2}u^2)\right] = [z_{1,k}]\frac{r_k}{z_{1,k}^{ln}} - [z_2]\overline{z_2}\ \overline{z_3} + [z_3](e_{0k} + \frac{c_{vk}}{z_3^{ln}} - \frac{1}{2}\overline{z_2^2}) \tag{4.12}$$

The remaining jumps are given by:

$$\left[\frac{u}{T}\right] = \overline{z_3}[z_2] + \overline{z_2}[z_3], -\left[\frac{1}{T}\right] = -[z_3]. \tag{4.13}$$

Using equations (4.11), (4.12) and (4.13), the entropy conservation condition (4.10) can be rewritten as the requirement that a linear combination of the jumps in the algebraic variables equals zero:

$$\sum_{k=1}^{N}[z_{1,k}]\left(\frac{r_k}{z_{1,k}^{ln}}f_{1,k}\right) + [z_2]\left((-\overline{z_3}\ \overline{z_2})\sum_{k=1}^{N} f_{1,k} + \overline{z_3}f_2\right) +$$

$$[z_3]\left(\sum_{k=1}^{N}(e_{0k} + c_{vk}\frac{1}{z_3^{ln}} - \frac{1}{2}\overline{z_2^2})f_{1,k} + \overline{z_2}f_2 - f_3\right) = [z_2]\left(\sum_{k=1}^{N} r_k\overline{z_{1,k}}\right) + \sum_{k=1}^{N}[z_{1,k}]r_k\overline{z_2}.$$

This scalar condition leads to the system of $N+3$ equations:

$$\begin{aligned}
r_k\frac{1}{z_{1,k}^{ln}} &= r_k\overline{z_2}f_{1,k}, \quad 1 \le k \le N,\\
(-\overline{z_3}\ \overline{z_2})\sum_{k=1}^{N} f_{1,k} + \overline{z_3}f_2 &= \left(\sum_{k=1}^{N} r_k\overline{z_{1,k}}\right),\\
\sum_{k=1}^{N}(e_{0k} + c_{vk}\frac{1}{z_3^{ln}} - \frac{1}{2}\overline{z_2^2})f_{1,k} + \overline{z_2}f_2 - f_3 &= 0.
\end{aligned}$$

The solution under these assumptions is:

$$\begin{aligned}
f_{1,k} &= \overline{z_2}z_{1,k}^{ln},\\
f_2 &= \frac{1}{\overline{z_3}}\left(\sum_{k=1}^{N} r_k\overline{z_{1,k}}\right) + \overline{z_2}\sum_{k=1}^{N} f_{1,k},\\
f_3 &= \sum_{k=1}^{N}(e_{0k} + c_{vk}\frac{1}{z_3^{ln}} - \frac{1}{2}\overline{z_2^2})f_{1,k} + \overline{z_2}f_2.
\end{aligned}$$

Therefore we obtain::

$$
\begin{aligned}
f_{1,k} &= \rho_k^{ln}\overline{u},\ 1 \le k \le N \\
f_2 &= \frac{1}{\overline{1/T}}\left(\sum_{k=1}^{N} r_k\overline{\rho_k}\right) + \overline{u}\sum_{k=1}^{N} f_{1,k}, \\
f_3 &= \sum_{k=1}^{N}(e_{0k} + c_{vk}\frac{1}{(1/T)^{ln}} - \frac{1}{2}\overline{u^2})f_{1,k} + \overline{u}f_2.
\end{aligned}
\tag{4.14}
$$

This EC flux is the multicomponent version of Chandrasekhar's EC flux [63] for the compressible Euler equations. Chandrasekhar's flux was designed with the additional property of being Kinetic-Energy Preserving (KEP) in the sense of Jameson [64], meaning that the kinetic energy equation is satisfied by the finite volume scheme in a semi-discrete sense (in the same spirit as EC schemes). This property can be useful in turbulent flow simulations [112]. For the compressible Euler equations, Jameson [64] showed that this is achieved if the momentum flux $f^{\rho u}$ has the form $f^{\rho u} = \tilde{p} + \overline{u}f^{\rho}$, where f^{ρ} is the mass flux and $\tilde{p}$ is any consistent pressure average. The extension of Jameson's analysis to the multicomponent case is straightforward and it can be showed that if the momentum flux has the same form as in the single component case (with f^{ρ} denoting the *total* mass flux), the KEP property is achieved. The EC flux we derived qualifies, with $\tilde{p}$ given by:

$$
\tilde{p} = \frac{1}{\overline{(1/T)}}\sum_{k=1}^{N} r_k\overline{\rho_k}.
$$

Note in passing that if $[p] = 0$ across the discontinuity, then $\tilde{p}$ defined above is exactly the pressure on each side.

The flux (4.14) is well-defined in the limit case $\rho_k = 0$, where the entropy variable $v_{1,k}$ is undefined, but is the entropy conservation condition still met? Equation (2.14) writes

$$\sum_{k=1}^{N}[v_{1,k}]f_{1,k} + [v_2]f_2 + [v_3]f_3 = [\mathcal{F}].$$

The jumps in $v_{1,k}$ are undefined because they involve jumps in $\ln(\rho_k)$. However, we note that the first term:

$$\begin{aligned}
\sum_{k=1}^{N}[v_{1,k}]f_{1,k} &= \sum_{k=1}^{N}\left[\frac{h_k}{T} - \frac{1}{2}\frac{u^2}{T}\right](\rho_k^{ln}\overline{u}) \;-\; \sum_{k}[s_k](\rho_k^{ln}\overline{u}) \\
&= \sum_{k=1}^{N}\left[\frac{h_k}{T} - \frac{1}{2}\frac{u^2}{T}\right](\rho_k^{ln}\overline{u}) \;-\; \sum_{k} c_{vk}[\ln T](\rho_k^{ln}\overline{u}) + \sum_{k=1}^{N} r_k[\ln\rho_k](\rho_k^{ln}\overline{u}) \\
&= \sum_{k=1}^{N}\left[\frac{h_k}{T} - \frac{1}{2}\frac{u^2}{T}\right](\rho_k^{ln}\overline{u}) \;-\; \sum_{k} c_{vk}[\ln T](\rho_k^{ln}\overline{u}) + \sum_{k=1}^{N} r_k[\rho_k]\overline{u}
\end{aligned}$$

is well-defined at $\rho_k = 0$, because the logarithmic averages ρ_k^{ln} compensate for the logarithmic jumps $[\ln\rho_k]$. We thus find that the entropy conservation condition is satisfied even in the limit $\rho_k = 0$.

Note that the EC flux does not transfer mass across an interface separating two different species.

4.2.2 Entropy-stable flux

4.2.2.1 Upwind-type dissipation operator

We now proceed to derive the scaling matrix for the multicomponent Euler system. First, the flux Jacobian is given by:

$$A = \begin{bmatrix}
u(1-Y_1) & \dots & -uY_1 & Y_1 & 0 \\
 & \ddots & & \vdots & 0 \\
-uY_N & \dots & u(1-Y_N) & Y_N & 0 \\
\frac{(\gamma-3)}{2}u^2 + d_1 & \dots & \frac{(\gamma-3)}{2}u^2 + d_N & u(3-\gamma) & \gamma-1 \\
u(d_1 - h^t + \frac{u^2}{2}(\gamma-1)) & \dots & u(d_N - h^t + \frac{u^2}{2}(\gamma-1)) & h^t - u^2(\gamma-1) & u\gamma
\end{bmatrix},$$

where $d_k = h_k - \gamma e_k$ and $h^t = h + \frac{1}{2}u^2$. We have $A = R\Lambda R^{-1}$ with:

$$R = \begin{bmatrix} 1 & & 0 & Y_1 & Y_1 \\ & \ddots & & \vdots & \vdots \\ 0 & & 1 & Y_N & Y_N \\ u & \dots & u & u+a & u-a \\ k - \frac{d_1}{\gamma-1} & \dots & k - \frac{d_N}{\gamma-1} & h^t + ua & h^t - ua \end{bmatrix},$$

$$\Lambda = diag(u,\ \dots,\ u,\ u+a,\ u-a),\ a^2 = \gamma r T,$$

where $k = \frac{u^2}{2}$. The Jacobian H of the mapping $\mathbf{v} \to \mathbf{u}$ is given by [172]:

$$H = \begin{bmatrix} \rho_1/r_1 & & 0 & u\rho_1/r_1 & \rho_1 e_1^t/r_1 \\ & \ddots & & \vdots & \vdots \\ 0 & & \rho_N/r_N & u\rho_N/r_N & \rho_N e_N^t/r_1 \\ u\rho_1/r_1 & \dots & u\rho_N/r_N & \rho T + u^2 S_1 & u(\rho T + S_2) \\ \rho_1 e_1^t/r_1 & \dots & \rho_N e_N^t/r_N & u(\rho T + S_2) & \rho T(u^2 + c_v T) + S_3 \end{bmatrix}, \tag{4.15}$$

$$S_1 = \sum_k \frac{\rho_k}{r_k},\ S_2 = \sum_k \frac{\rho_k}{r_k}(e_k^t),\ S_3 = \sum_k \frac{\rho_k}{r_k}(e_k^t)^2$$

where $e_k^t = e_k + k$. The squared scaling matrix is given by:

$$T^2 = R^{-1} H R^{-T} = \frac{\rho}{\gamma r} diag(T^{2Y},\ 1/2,\ 1/2), \tag{4.16}$$

where $T^{2Y} \in \mathbb{R}^{N \times N}$ is given by:

$$\begin{aligned} T_{ii}^{2Y} &= (\gamma - 1)Y_i^2 + \sum_{k \neq i} (\gamma r_k / r_i) Y_k Y_i,\ 1 \leq i \leq N, \\ T_{ij}^{2Y} &= -Y_i Y_j,\ 1 \leq i \neq j \leq N. \end{aligned}$$

At this point, the dissipation operator writes:

$$D[\mathbf{v}] = \frac{1}{2} R|\Lambda|T^2 R^T[\mathbf{v}], \tag{4.17}$$

and qualifies for the ES scheme because the matrix $R|\Lambda|T^2R^T$ is positive definite (T^2 and $|\Lambda|$ commute). However, as will be seen in section 4.3.2, a scaled form (2.25) of the dissipation operator is necessary.

For $N = 2$, we have:

$$T^{2Y} = \begin{bmatrix} (\gamma-1)Y_1^2 + (\gamma r_2/r_1)Y_1Y_2 & -Y_1Y_2 \\ -Y_1Y_2 & (\gamma-1)Y_2^2 + (\gamma r_1/r_2)Y_1Y_2 \end{bmatrix}.$$

T^{2Y} is symmetric, real-valued with non-negative eigenvalues therefore there exists T^Y with the same properties such that $T^{2Y} = (T^Y)^2 = T^Y(T^Y)^T$ (T^Y is the square root of T^{2Y}). This matrix can be derived using an eigenvalue decomposition. However the expression of T^Y is quite complicated. The square root of T^{2Y} is not necessary to proceed. For $N = 2$, T^{2Y} can be rewritten as:

$$T^{2Y} = \mathcal{T}^Y(\mathcal{T}^Y)^T,\ \mathcal{T}^Y = \begin{bmatrix} -\sqrt{Y_1Y_2}\sqrt{\gamma r_2/r_1} & Y_1\sqrt{\gamma-1} \\ \sqrt{Y_1Y_2}\sqrt{\gamma r_1/r_2} & Y_2\sqrt{\gamma-1} \end{bmatrix}. \tag{4.18}$$

$\mathcal{T}^Y$ is not the square root of T^{2Y}, however it is enough to obtain a scaled formulation (2.25) because:

$$T^2 = \mathcal{T}\mathcal{T}^T,\ \mathcal{T} = \sqrt{\frac{\rho}{\gamma r}} diag(\mathcal{T}^Y,\ 1/\sqrt{2},\ 1/\sqrt{2}),$$

and $\mathcal{T}$ commutes with $|\Lambda|$ therefore the dissipation operator can be rewritten as:

$$D[\mathbf{v}] = \frac{1}{2}(R\mathcal{T})|\Lambda|(R\mathcal{T})^T[\mathbf{v}]. \tag{4.19}$$

For $N > 2$, the expression for T^Y becomes even more complicated. For $N = 3$, the alternative (4.18) to T^Y we proposed for $N = 2$ becomes:

$$T^{2Y} = \mathcal{T}^Y(\mathcal{T}^Y)^T,$$

$$\mathcal{T}^Y = \begin{bmatrix} -\sqrt{Y_1Y_2}\sqrt{\gamma r_2/r_1} & -\sqrt{Y_1Y_3}\sqrt{\gamma r_3/r_1} & 0 & -Y_1\sqrt{\gamma-1} \\ \sqrt{Y_1Y_2}\sqrt{\gamma r_1/r_2} & 0 & -\sqrt{Y_2Y_3}\sqrt{\gamma r_3/r_2} & -Y_2\sqrt{\gamma-1} \\ 0 & \sqrt{Y_1Y_3}\sqrt{\gamma r_1/r_3} & \sqrt{Y_2Y_3}\sqrt{\gamma r_2/r_3} & -Y_3\sqrt{\gamma-1} \end{bmatrix}.$$

There is one more column compared to the $N = 2$ case. The form (4.19) still holds except that $R\mathcal{T} \in \mathbb{R}^{3\times4}$ instead of $\mathbb{R}^{3\times3}$ and $|\Lambda| \in \mathbb{R}^{4\times4}$ diagonal with an extra $|u|$ term. For N species, the "pseudo" scaling matrix $\mathcal{T}^Y$ we described will be in $\mathbb{R}^{N\times(N(N-1)/2+1)}$.

4.2.2.2 Average state

To complete the definition of the dissipation operator, an average state (referred to with the superscript $*$) needs to be specified.

We showed in section 4.2.1 that the EC flux (4.14) is well defined in the limit $\rho_k = 0$. What about the dissipation operator $R|\Lambda|T^2R^T[\mathbf{v}]$? At first glance, the presence of the jump term $[\mathbf{v}]$ is problematic, because $[v_{1,k}]$ is undefined. For $N = 2$, the squared scaling matrix can be rewritten as:

$$T^2 = \overline{T}^2\mathcal{R},\ \overline{T}^2 = \frac{1}{\gamma r}\begin{bmatrix} (\gamma-1)Y_1 + \gamma r_2/r_1 Y_2 & -Y_2 & 0 & 0 \\ -Y_1 & (\gamma-1)Y_2 + \gamma r_1/r_2 Y_1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}, \quad (4.20)$$

$$\mathcal{R} = \begin{bmatrix} \rho_1 & 0 & 0 & 0 \\ 0 & \rho_2 & 0 & 0 \\ 0 & 0 & \rho & 0 \\ 0 & 0 & 0 & \rho \end{bmatrix}. \quad (4.21)$$

Denoting $D_k = k - \frac{d_k}{\gamma-1}$, we have:

$$\mathcal{R}R^T[\mathbf{v}] = \begin{bmatrix} \rho_1^* & 0 & 0 & 0 \\ 0 & \rho_2^* & 0 & 0 \\ 0 & 0 & \rho^* & 0 \\ 0 & 0 & 0 & \rho^* \end{bmatrix} \begin{bmatrix} 1 & 0 & u^* & D_1^* \\ 0 & 1 & u^* & D_2^* \\ Y_1^* & Y_2^* & u^*+a^* & (h^t)^* + u^*a^* \\ Y_1^* & Y_2^* & u^*-a^* & (h^t)^* - u^*a^* \end{bmatrix} \begin{bmatrix} \big[v_{1,1}\big] \\ \big[v_{1,2}\big] \\ \big[v_2\big] \\ \big[v_3\big] \end{bmatrix} =$$

$$\begin{bmatrix} \rho_1^*\big[v_{1,1}\big] + \rho_1^* u\big[v_2\big] + \rho_1^* D_1^*\big[v_3\big] \\ \rho_2^*\big[v_{1,2}\big] + \rho_2^* u\big[v_2\big] + \rho_2^* D_2^*\big[v_3\big] \\ \rho_1^*\big[v_{1,1}\big] + \rho_2^*\big[v_{1,2}\big] + \rho^*(u^*+a^*)\big[v_2\big] + \rho^*((h^t)^* + u^*a^*)\big[v_3\big] \\ \rho_1^*\big[v_{1,1}\big] + \rho_2^*\big[v_{1,2}\big] + \rho^*(u^*-a^*)\big[v_2\big] + \rho^*((h^t)^* - u^*a^*)\big[v_3\big] \end{bmatrix} \quad (4.22)$$

We can see that $\mathcal{R}R^T[\mathbf{v}]$ is well-defined if $\rho_k^*[v_{1,k}]$ is well-defined as well. Given that:

$$\rho_k^*[v_{1,k}] = \Bigg(- c_{vk}[ln(T)] - \left[\frac{u^2}{2T}\right]\Bigg)\rho_k^* + \frac{R}{m_k}[\ln \rho_k]\rho_k^*,$$

we see that with $\rho_k^* = \rho_k^{ln}$ the dissipation operator is well-defined. For total density, one might be tempted to take $\rho^* = \sum_{k=1}^{N} \rho_k^* = \sum_{k=1}^{N} \rho_k^{ln}$. This definition makes $\rho^* = 0$

possible, which is undesirable given that $Y_k^* = \rho_k^*/\rho^*$. $\rho^* = \rho^{\ln}$ is a safer choice.

The exact resolution of stationary contact discontinuities is a desirable property in the calculation of boundary and shear layers (even though it might produce carbuncles on blunt-body calculations, see [155] paragraph 2.4.). In this case, $[p] = 0,\ [u] = \overline{u} = 0$ and the EC flux we derived reduces to:

$$f_{1,k} = 0,\ f_2 = \frac{1}{\overline{1/T}}\left(\sum_{k=1}^{N} \frac{R}{m_k}\overline{\rho_k}\right),\ f_3 = 0.$$

From the ideal gas law we can state that f_2 is exactly the pressure on both sides of the contact. The dissipation term must therefore cancel out if we want the ES scheme to exactly preserve stationary contact discontinuities. Since $u = 0$, we have:

$$R^T = \begin{bmatrix} 1 & & 0 & 0 & d_1^*/(\gamma-1) \\ & \ddots & & \vdots & \vdots \\ 0 & & 1 & 0 & d_N^*/(\gamma-1) \\ Y_1^* & \dots & Y_N^* & a^* & h^* \\ Y_1^* & \dots & Y_N^* & -a^* & h^* \end{bmatrix},\ [\mathbf{v}] = \begin{bmatrix} \left[\frac{g_1}{T}\right] & \dots & \left[\frac{g_N}{T}\right] & 0 & -\left[\frac{1}{T}\right] \end{bmatrix}^T,$$

therefore:

$$R^T[\mathbf{v}] = \begin{bmatrix} \left[\frac{g_1}{T}\right] - d_1^*/(\gamma-1)\left[\frac{1}{T}\right] \\ \vdots \\ \left[\frac{g_N}{T}\right] - d_N^*/(\gamma-1)\left[\frac{1}{T}\right] \\ \sum_{k=1}^N Y_k^*\left[\frac{g_k}{T}\right] - \left[\frac{1}{T}\right]h^* \\ \sum_{k=1}^N Y_k^*\left[\frac{g_k}{T}\right] - \left[\frac{1}{T}\right]h^* \end{bmatrix}.$$

$|u^*| = 0$ so the product of the eigenvalue matrix $|\Lambda|$ and the squared scaling matrix T^2 simplifies to:

$$|\Lambda|T^2 = \frac{\rho^* a^*}{2\gamma^* r^*} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Therefore, the dissipation term $R|\Lambda|T^2R^T[\mathbf{v}]$ cancels out if:

$$\sum_{k=1}^{N} Y_k^* \left[\frac{g_k}{T}\right] - \left[\frac{1}{T}\right] h^* = 0.$$

This is equivalent to:

$$\sum_{k=1}^{N} \rho_k^* \Big(e_{0k} \left[\frac{1}{T}\right] - [s_k] \Big) - \left[\frac{1}{T}\right] \rho^* h^* = 0 \tag{4.23}$$

Using:

$$-[s_k] = c_{vk} \left[\frac{1}{T}\right] \frac{1}{(1/T)^{ln}} + r_k \frac{[\rho_k]}{\rho^{ln}},$$

equation (4.23) can then be rewritten as:

$$\sum_{k=1}^{N} \left(\frac{\rho_k^*}{\rho_k^{ln}}\right) r_k[\rho_k] - \left[\frac{1}{T}\right] \Bigg(\rho^* h^* - \sum_{k=1}^{N} \rho_k^* \Big(e_{0k} + c_{vk} \frac{1}{(1/T)^{ln}} \Big) \Bigg) = 0 \tag{4.24}$$

The ideal gas law along with the assumption of constant pressure allows us to relate the jumps in partial densities and temperature:

$$\sum_{k=1}^{N} r_k[\rho_k] = \overline{p} \left[\frac{1}{T}\right]. \tag{4.25}$$

If $\rho_k^* = \rho_k^{ln}$, then using equation (4.25), equation (4.24) simplifies to:

$$\left[\frac{1}{T}\right]\left(\bar{p} + \sum_{k=1}^{N} \rho_k^* \left(e_{0k} + c_{vk}\frac{1}{(1/T)^{ln}}\right) - \rho^* h^*\right) = 0.$$

This leads to the condition:

$$\rho^* h^* = \sum_{k=1}^{N} \rho_k^* \left(e_{0k} + c_{vk}\frac{1}{(1/T)^{ln}}\right) + \overline{p},\ \rho_k^* = \rho_k^{ln}. \tag{4.26}$$

The remaining averages are taken as $u^* = \overline{u},\ T^* = 1/(1/T)^{ln},\ r^* = \bar{r},\ \gamma^* = \overline{\gamma}$ and $a^* = \sqrt{\gamma^* r^* T^*}$.

4.2.3 Additional considerations

4.2.3.1 Time integration

The first-order finite volume scheme we derived is ES at the semi-discrete level only. Entropy stability or entropy conservation at the fully-discrete level can be obtained using a variety of techniques [59, 50, 51, 161, 144, 145, 69] which can be applied to the multicomponent compressible Euler system. However, this typically requires implicit time-integration schemes. For simplicity, we use explicit Runge-Kutta schemes in time, which do not guarantee entropy stability at the fully-discrete level.

4.2.3.2 Positivity

We showed in sections 4.2.1 and 4.2.2 that despite the fact that the entropy variables are undefined in the limit $\rho_k \to 0$, the EC flux is well-defined and that the ES flux remains well-defined if the averaged partial densities are properly chosen (we show a similar result for high-order TecNO schemes in section 4.3.2). This does not guarantee that the resulting scheme will not produce negative densities and/or negative pressures. In one-dimensional test problems, the first order semi-discrete ES scheme fortunately did not produce negative densities or pressure but the high-order TecNO scheme (section 4.3.2) systematically did. In the original setup of the two-dimensional shock-bubble interaction problem [194, 191], the first-order scheme had the same issue. This lack of positivity proof is not unique to ES schemes, but applies to all schemes which do not strictly enforce local wave propagation.

There are schemes which are conservative, entropy stable and can preserve the correct sign of density and pressure. The Godunov scheme as well as the Lax-Friedrichs scheme qualify if an appropriate CFL condition is met [175, 147]. Note also the recent work of Guermond *et al.* [176, 177]. A common trait of these schemes is that they take root in the notion of a Riemann problem and the assumption that there exists a solution satisfying all entropy inequalities. These schemes also require an algorithm capable of computing the maximum speed of propagation (or an upper bound) for the Riemann problem at each interface. One such algorithm is discussed in Guermond & Popov [178]. We could in principle adopt a hybrid approach where these schemes are used in areas where the present one fails to maintain positive densities and pressure. Whether this can effectively be accomplished is left for future work.

4.2.3.3 Construction for thermally perfect gases

Here we briefly discuss the construction of an ES scheme in the case where the specific heats are not constant but functions of temperature. In this configuration, the specific internal energies and entropies are defined as:

$$e_k := e_{0k} + \int_0^T c_{vk}(\tau)d\tau,\ s_k := \int_0^T \frac{c_{vk}(\tau)}{\tau}d\tau - r_k \ln \rho_k.$$

The structure that ES schemes build on [172, 174] is still present. The multicomponent system still admits an additional conservation equation for the thermodynamic entropy of the mixture ρs and $(U,F) = (-\rho s,\ -\rho u s)$ is a valid entropy pair for $\rho_k > 0,\ T > 0$. In practice, the specific heats are represented using polynomial interpolation:

$$c_{vk} := c_{k0} + \sum_{j=1}^{P} c_{kj} T^j,$$

where $c_{kj},\ 0 \leq j \leq P$ are constants. This gives:

$$e_k = e_{0k} + c_{k0} T + \sum_{j=1}^{p} \frac{c_{kj}}{j+1} T^{j+1},\ s_k = c_{k0} \ln T + \sum_{j=1}^{P} \frac{c_{kj}}{j} T^j - r_k \ln \rho_k.$$

The expressions of the entropy variables and potential functions remain unchanged. Regarding the construction of the EC flux, we have for $1 \leq k \leq N$:

$$\begin{aligned}[v_{1,k}] = \left[\frac{1}{T}\left(g_k - \frac{u^2}{2}\right)\right] = \left[\frac{e_k}{T} - s_k\right] - \left[\frac{u^2}{2T}\right] = \\ e_{0k}\left[\frac{1}{T}\right] - c_{k0}[\ln T] - \sum_{j=1}^{P} \frac{c_{kj}}{j(j+1)}[T^j] + r_k[\ln \rho_k] - \frac{\overline{u^2}}{2}\left[\frac{1}{T}\right] - \overline{\frac{1}{T}}\overline{u}[u] \quad (4.27)\end{aligned}$$

For a given quantity a, let's define the product operator $a^\times = a_L a_R$. For $a \neq 0$, we have the following identity:

$$[a] = -a^\times \left[\frac{1}{a}\right]$$

It can be easily shown, by induction for instance, that for $j \geq 1$, there exists an averaging operator $f_j(a)$, consistent with a^{j-1}, such that $[a^j] = j f_j(a)[a]$ ($f_1(a) = 1$, $f_2(a) = \overline{a}$, $f_3(a) = (2/3)\overline{a}\,\overline{a} + (1/3)\overline{a^2}$ and so on using product rules). Equation (4.27) can therefore be rewritten as:

$$\begin{aligned}[v_{1,k}] &= e_{0k}\left[\frac{1}{T}\right] - c_{k0}[\ln T] + \sum_{j=1}^{P} \frac{c_{kj}}{(j+1)}(f_j(T)T^\times)\left[\frac{1}{T}\right] + r_k[\ln \rho_k] - \frac{\overline{u^2}}{2}\left[\frac{1}{T}\right] - \overline{\frac{1}{T}}\overline{u}[u] \\ &= \left[\frac{1}{T}\right]\left(e_{0k} + c_{k0}\frac{1}{(1/T)^{ln}} + \sum_{j=1}^{P} \frac{c_{kj}}{(j+1)}(f_j(T)T^\times) - \frac{\overline{u^2}}{2}\right) + [\rho_k]\frac{r_k}{\rho_k^{ln}} - [u]\overline{\left(\frac{1}{T}\right)}\overline{u}\end{aligned}$$

Repeating the procedure outlined in section 4.2.1, we obtain an EC flux:

$$
\begin{aligned}
f_{1,k} &= \rho_k^{ln}\overline{u}, \\
f_2 &= \frac{1}{\overline{1/T}}\left(\sum_{k=1}^{N} r_k\overline{\rho_k}\right) + \overline{u}\sum_{k=1}^{N} f_{1,k}, \\
f_3 &= \sum_{k=1}^{N}\left(e_{0k} + c_{k0}\frac{1}{(1/T)^{ln}} + \sum_{j=1}^{P}\frac{c_{kj}}{(j+1)}(f_j(T)T^{\times}) - \frac{1}{2}\overline{u^2}\right) f_{1,k} + \overline{u}f_2.
\end{aligned}
\tag{4.28}
$$

which is consistent ($f_j(T)T^{\times}$ is consistent with T^{j+1}) and differs from the expression we obtained in the calorically perfect gas case (equation (4.14)) in the total energy component only.

We will not go into the details of the upwind dissipation operator. Barth's eigenscaling theorem applies because H still symmetrizes A from the right. Therefore the scaling matrix exists ($T^2 = R^{-1}HR^{-T}$) and the dissipation operator constructed in section 4.2.2 can be constructed for thermally perfect gases as well.

4.2.3.4 Total mass form

Instead of solving the conservation equations for the N partial densities ρ_k, one might want to solve for the conservation of the total density ρ and N-1 partial densities. The state and flux vectors are then:

$$
\tilde{\mathbf{u}} := \begin{bmatrix}\rho & \rho_2 & \dots & \rho_N & \rho u & \rho e^t\end{bmatrix}^T, \; \tilde{\mathbf{f}} := \begin{bmatrix}\rho u & \rho_2 u & \dots & \rho_N u & \rho u^2 + p & (\rho e^t + p)u\end{bmatrix}^T.
$$

The entropy variables in this configuration can be easily obtained using a chain rule:

$$
\tilde{\mathbf{v}}^T = -\frac{\partial \rho s}{\partial \tilde{\mathbf{u}}} = -\frac{\partial \rho s}{\partial \mathbf{u}}\left(\frac{\partial \tilde{\mathbf{u}}}{\partial \mathbf{u}}\right)^{-1} = \mathbf{v}^T\left(\frac{\partial \tilde{\mathbf{u}}}{\partial \mathbf{u}}\right)^{-1} = \mathbf{v}^T\left(\frac{\partial \mathbf{u}}{\partial \tilde{\mathbf{u}}}\right)
$$

$\tilde{\mathbf{u}}$ and $\mathbf{u}$ only differ in the first component, $\rho_1 = \rho - \sum_{k=2}^{N} \rho_k$ therefore:

$$\frac{\partial \mathbf{u}}{\partial \tilde{\mathbf{u}}} = \begin{bmatrix} 1 & -1 & \dots & -1 & 0 & 0 \\ & 1 & & 0 & 0 & 0 \\ & & \ddots & 0 & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix},$$

$$\tilde{\mathbf{v}} = \frac{1}{T} \begin{bmatrix} (g_1 - \frac{1}{2}u^2) & (g_2 - g_1) & \dots & (g_N - g_1) & u & -1 \end{bmatrix}^T.$$

The corresponding potential functions are unchanged because:

$$\tilde{\mathbf{v}}^T \tilde{\mathbf{f}} = \mathbf{v}^T \left(\frac{\partial \mathbf{u}}{\partial \tilde{\mathbf{u}}} \right) \left(\frac{\partial \tilde{\mathbf{u}}}{\partial \mathbf{u}} \right) \mathbf{f} = \mathbf{v}^T \mathbf{f},\ \tilde{\mathbf{v}}^T \tilde{\mathbf{u}} = \mathbf{v}^T \left(\frac{\partial \mathbf{u}}{\partial \tilde{\mathbf{u}}} \right) \left(\frac{\partial \tilde{\mathbf{u}}}{\partial \mathbf{u}} \right) \mathbf{u} = \mathbf{v}^T \mathbf{u}$$

Accordingly, an EC flux $\tilde{\mathbf{f}}_{\mathbf{EC}}$ for the total mass form can simply be obtained by mapping an EC flux in the first form $\mathbf{f}_{\mathbf{EC}}$:

$$\tilde{\mathbf{f}}_{\mathbf{EC}} = \left(\frac{\partial \tilde{\mathbf{u}}}{\partial \mathbf{u}} \right) \mathbf{f}_{EC}.$$

Likewise, applying the same mapping to an ES flux $\mathbf{f}^*$ in the first form given by:

$$\mathbf{f}^* = \mathbf{f}_{\mathbf{EC}} - D[\mathbf{v}]$$

where D is positive definite, results in a flux $\tilde{\mathbf{f}}^*$ given by:

$$\tilde{\mathbf{f}}^* = \left(\frac{\partial \tilde{\mathbf{u}}}{\partial \mathbf{u}} \right) \mathbf{f}^* = \tilde{\mathbf{f}}_{\mathbf{EC}} - \left(\frac{\partial \tilde{\mathbf{u}}}{\partial \mathbf{u}} \right) D[\mathbf{v}] = \tilde{\mathbf{f}}_{\mathbf{EC}} - \left(\frac{\partial \tilde{\mathbf{u}}}{\partial \mathbf{u}} \right) D \left(\frac{\partial \tilde{\mathbf{u}}}{\partial \mathbf{u}} \right)^T [\tilde{\mathbf{v}}] = \tilde{\mathbf{f}}_{\mathbf{EC}} - \tilde{D}[\tilde{\mathbf{v}}].$$

$\tilde{D}$ is positive definite by congruence therefore $\tilde{\mathbf{f}}^*$ is entropy stable.

4.3 High-order discretizations

In this section, we are essentially interested in how the fundamental issues highlighted in the previous sections manifest in a high-order setting. We discuss two high-order ES formulations: TecNO schemes [60] and Discontinuous Galerkin [117, 118] (DG) schemes discretizing the entropy variables [59, 44, 67]. High-order ES schemes are not limited to these two options, but the issues their formulation raises are no different.

4.3.1 Discontinuous Galerkin

The fact that the entropy variables are undefined in the limit $\rho_k \to 0$ poses a daunting problem, unless the flow configuration is such that one can expect $\rho_k > 0$ at all times. A way around this issue (other than not discretizing the entropy variables) has not been found by the author. There might be other entropy functions for which the corresponding entropy variables are well-defined in this limit. However, in the context of the multicomponent Navier-Stokes equations (including viscous stresses, heat conduction, multicomponent diffusion), the *a priori* entropy stability resulting from discretizing the entropy variables might be lost [44, 173].

4.3.2 TecNO schemes

The TecNO approach does not need to discretize the entropy variables but it does require the knowledge of the scaled eigenvectors (RT). If the dissipation operator is only known in the form:

$$D[\mathbf{v}] = R|\Lambda|T^2R^T[\mathbf{v}],$$

then choosing $\mathbf{w}$ such that $[\mathbf{w}] = (T^2 R^T)[\mathbf{v}]$ for the ENO reconstruction will result in a high-order dissipation operator:

$$\tilde{D}[\mathbf{w}] = R|\Lambda| \langle \mathbf{w} \rangle = R(|\Lambda| B T^2) R^T [\mathbf{v}].$$

$|\Lambda|$, B and T^2 are all symmetric and at least positive semi-definite, however their product is not necessarily positive semi-definite. If the squared scaling matrix is diagonal then entropy stability is preserved. In the general case of a block-diagonal scaling matrix, entropy stability is no longer ensured. The same problem would arise if $\mathbf{w}$ was chosen such that $[\mathbf{w}] = R^T[\mathbf{v}]$ (the question would be whether $(|\Lambda| T^2 B)$ is positive definite).

If the dissipation operator is expressed according to equation (4.19) with the pseudo-scaling matrix $\mathcal{T}$ given by equation (4.18), the vector of reconstructed variables $\mathbf{w}$ such that $= [\mathbf{w}] = (R\mathcal{T})^T[\mathbf{v}]$ will have as many components as the number of rows of $\mathcal{T}$. For $N > 2$, the number of rows of $\mathcal{T}$ grows as $\mathcal{O}(N^2)$, so for a large enough N, one might prefer working with T instead of $\mathcal{T}$ and avoid having to reconstruct too many variables.

In section 4.2.2, we showed that the dissipation term expressed as $R|\Lambda|T^2R^T$ is well defined in the limit $\rho_k \to 0$ provided that $\rho_k^* = \rho_k^{ln}$. This was possible because one could extract from T^2 a diagonal matrix $\mathcal{R}$ (see equation (4.20)) of partial densities. The TecNO algorithm requires the isolated evaluation of the scaled entropy variables defined by the jump relation $[\mathbf{w}] = (R\mathcal{T})^T[\mathbf{v}]$ or $[\mathbf{w}] = (RT)^T[\mathbf{v}]$. The matrix $\mathcal{R}$ that was extracted from the squared scaling matrix, might not be extracted from T or $\mathcal{T}$ without leaving $1/\rho_k^*$ terms behind. However, $\mathcal{R}^{1/2}$ can be extracted from the pseudo-scaling matrix $\mathcal{T}$ and since:

$$[\ln(\rho_k)] = [2 \; \ln(\sqrt{\rho_k})] = 2 \; \frac{[\sqrt{\rho_k}]}{(\sqrt{\rho_k})^{ln}},$$

it can be shown that $[\mathbf{w}] = (R\mathcal{T})^T[\mathbf{v}]$ is well-defined in the limit $\rho_k \to 0$, provided that $\rho_k^* = ((\sqrt{\rho_k})^{ln})^2$. The decomposition:

$$T^2 = \mathcal{R}^{1/2}\widetilde{T}^2(\mathcal{R}^{1/2})^T,\ \widetilde{T}^2 = \frac{1}{\gamma r}\begin{bmatrix} (\gamma-1)Y_1 + \gamma r_2/r_1 Y_2 & -\sqrt{Y_1Y_2} & 0 & 0 \\ -\sqrt{Y_1Y_2} & (\gamma-1)Y_2 + \gamma r_1/r_2 Y_1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}. \tag{4.29}$$

suggests that a TecNO reconstruction based on the scaling matrix T would still be defined in the limit with the same averaging. Note that this average is not compatible with the stationary contact preservation condition (4.26) we derived in section 4.2.2. That is because equation (4.26) requires $\rho_k^* = \rho_k^{ln}$.

4.4 Numerical experiments

In this section, we present and discuss numerical results on 1D (section 4.4.1) and 2D (section 4.4.2) test problems involving interfaces and shocks. A 3D formulation of the scheme (entropy variables, EC flux, ES flux) is provided in appendix A.

4.4.1 One-dimensional cases

In all three problems, the first-order finite volume scheme with the ES flux in space and forward Euler in time with a CFL of 0.3 on three grids with 100, 300 and 1000 cells was applied. All figures in this section use the same legend as in figure 4.1-(a).

4.4.1.1 Moving interface

The first test problem is the advection of a contact discontinuity (constant velocity and constant pressure) separating two different species. The initial conditions are given by:

$$\begin{cases} (\rho_1,\ \rho_2,\ u,\ p) = & (0.1,\ 0.0,\ 1.0,\ 1.0),\ 0 \leq x \leq 0.5, \\ (\rho_1,\ \rho_2,\ u,\ p) = & (0.0,\ 1.0,\ 1.0,\ 1.0),\ 0.5 < x \leq 1. \end{cases}$$

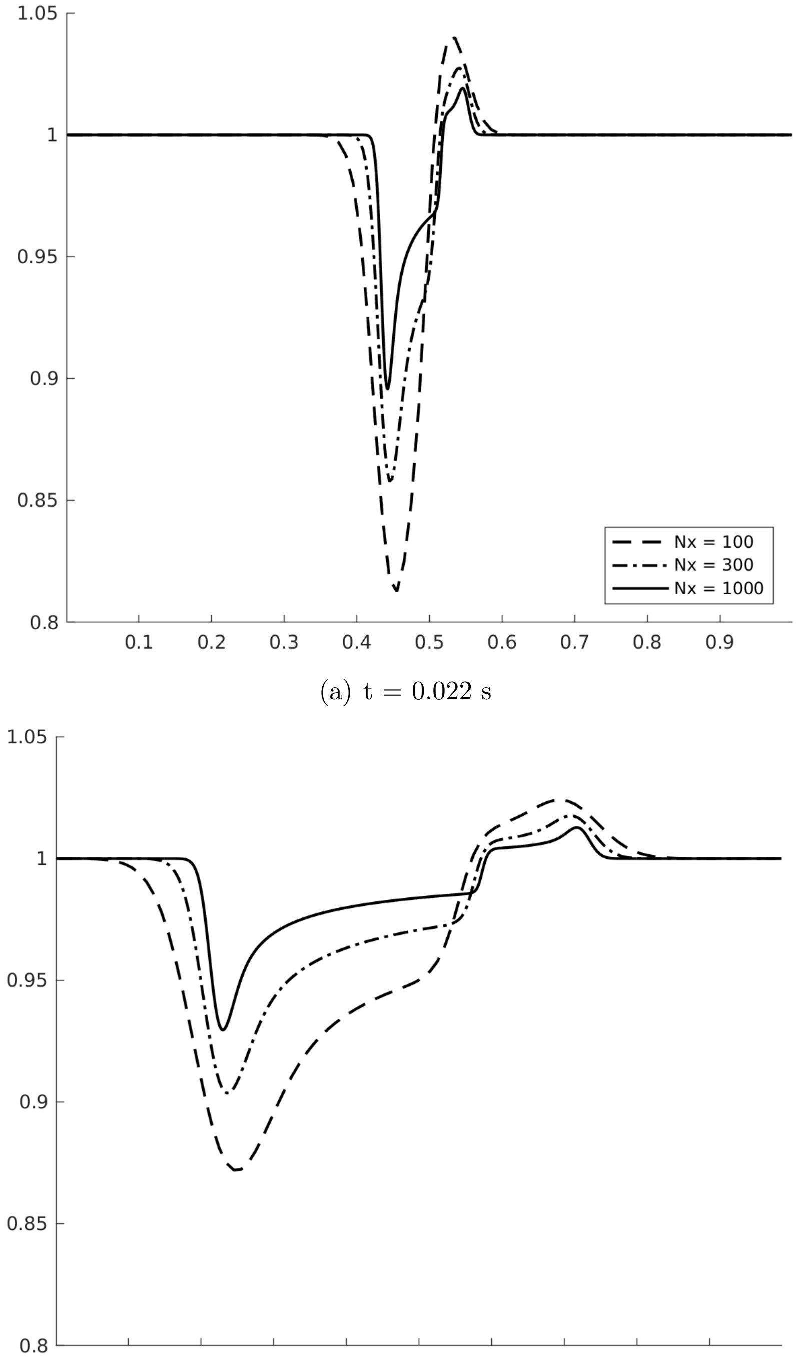
with $c_{p1} = 1.6,\ c_{p2} = 1.4,\ c_{v1} = c_{v2} = 1$. The velocity and pressure profiles at $t = 0.022s$ and $t = 0.1s$ are shown in figures 4.1 and 4.2. These profiles show overshoots and undershoots that are typically observed with conservative schemes (see figure 4 in Abgrall & Karni [184] for instance). These anomalies are often termed oscillations in the literature [184, 185, 186, 194, 195]. Upon closer examination, we can see three waves propagating at different speeds. The one moving to the left has the fastest propagation speed, roughly -3. There are two moving to the right, one with a propagation speed close to 1 (the speed of the contact) and another moving at a speed that is roughly 2. The speed of sound on the left-hand side of the contact is $a_1 = 4$. The speed of sound on the right-hand side of the contact is $a_2 = \sqrt{1.4} \approx 1.18$. The left-moving and (fastest) right-moving pressure waves propagate at speeds close to $u - a_1$ and $u + a_2$ respectively (note the dependence on the grid resolution).

Figure 4.3 shows the total entropy $\rho s = \rho_1 s_1 + \rho_2 s_2$ profiles at $t = 0.022s$ and $t = 0.1s$. The wave structure of the pressure and velocity anomalies is more apparent, and we can see that each wave is carrying an spurious increase in entropy. Figure 4.5 shows the evolution of the total entropy over time (the contributions from the boundaries were removed). This shows that the anomalies observed do not violate entropy stability. On the contrary, it appears that inappropriate production of entropy is the issue. Additionally, we find that these anomalies are still present when the upwind dissipation operator is discarded and the Backward Euler scheme is used to ensure entropy stability (see appendix B).

In the same vein, we find that these anomalies do not violate a minimum principle of the specific entropy (see figure 4.4.1.1).

That these anomalies are not linked to entropy stability or the minimum entropy principle should not come as a surprise. These anomalies were already observed with the Godunov scheme [184, 185], which is both entropy stable and satisfies a minimum entropy principle [27, 175, 147].

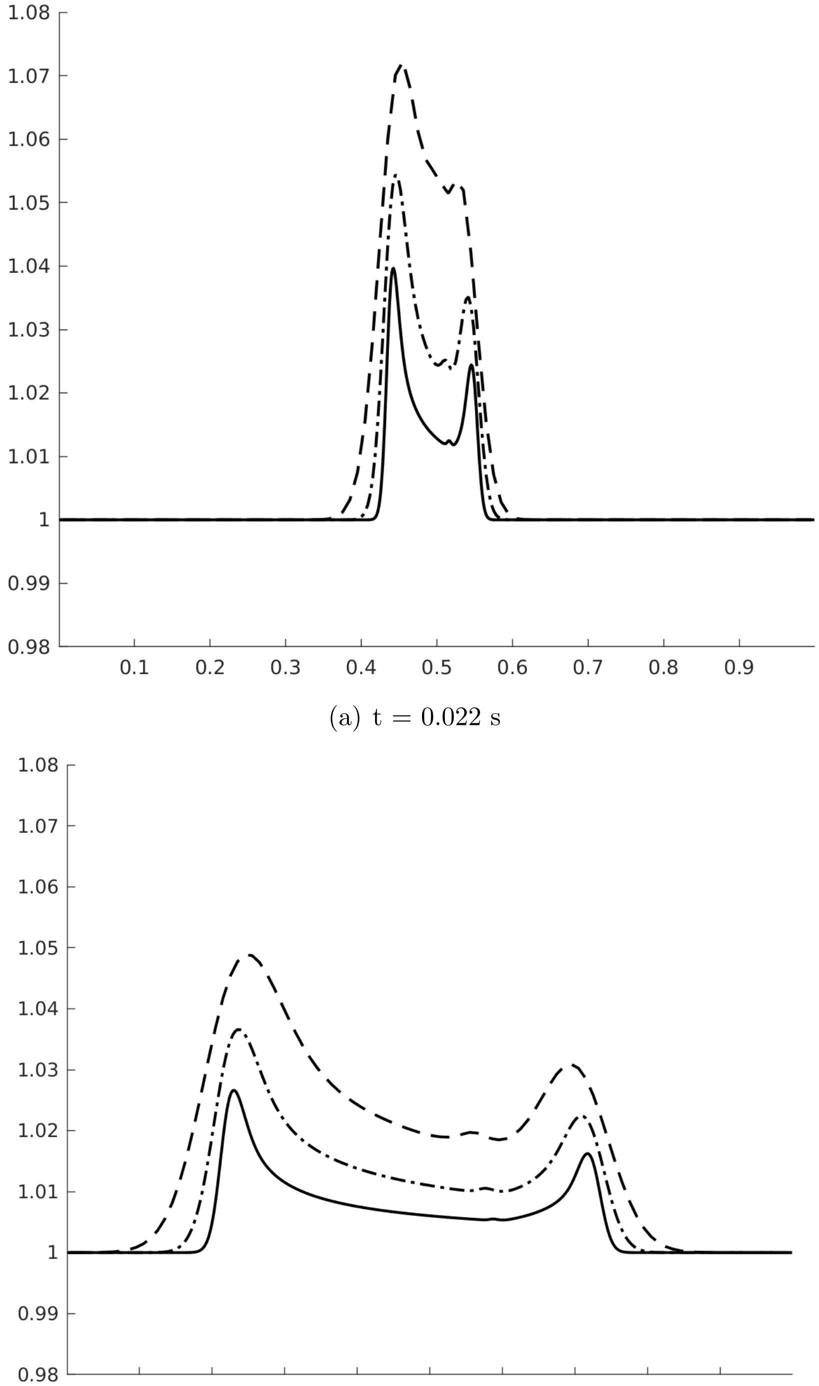
It is important to understand that this problem is intrinsic to multicomponent flows. There are no such anomalies on moving contacts in the compressible Euler equations (B). Furthermore, these anomalies are produced by first-order schemes. Intrigued readers are strongly encouraged to read early studies of this problem [185, 184, 186].



(a) t = 0.022 s

(b) t = 0.1 s

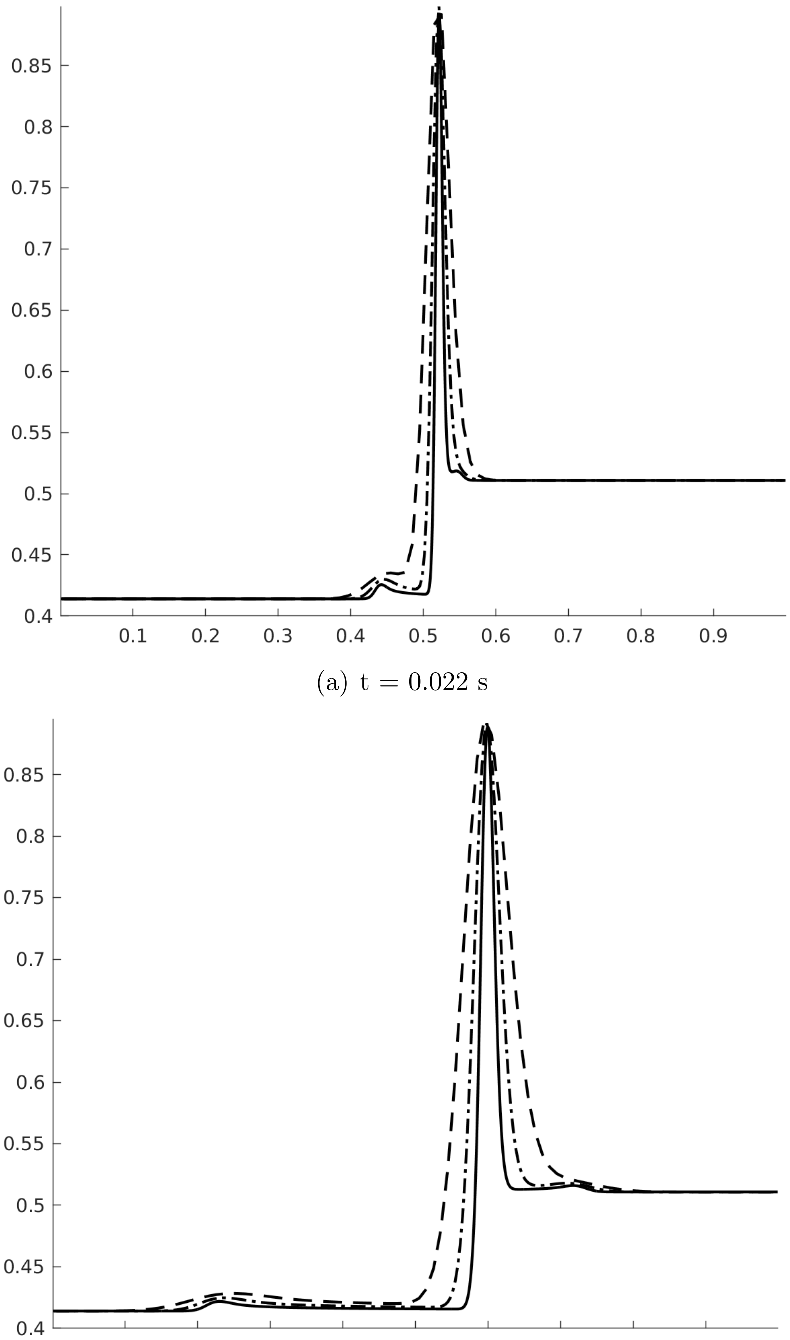
Figure 4.1: Moving Interface: velocity profiles.



(a) t = 0.022 s

(b) t = 0.1 s

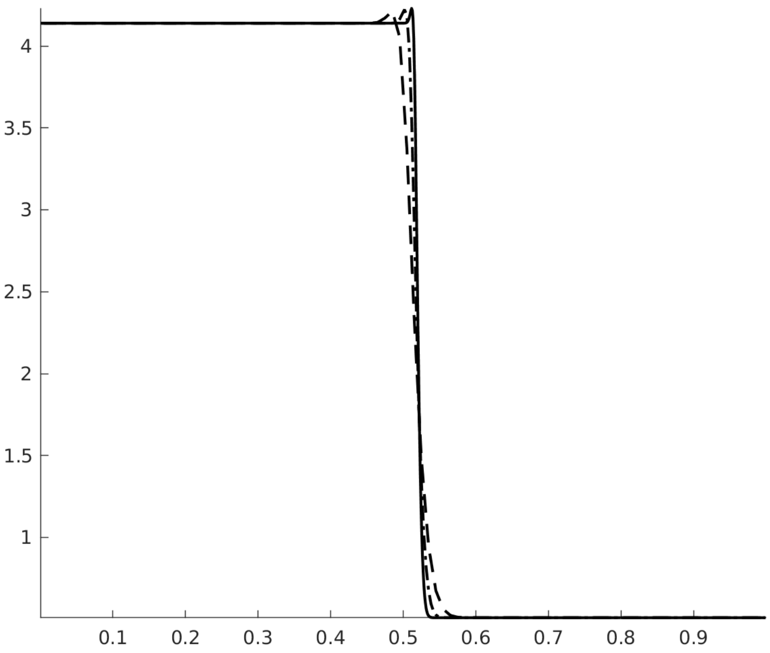
Figure 4.2: Moving Interface: pressure profiles.



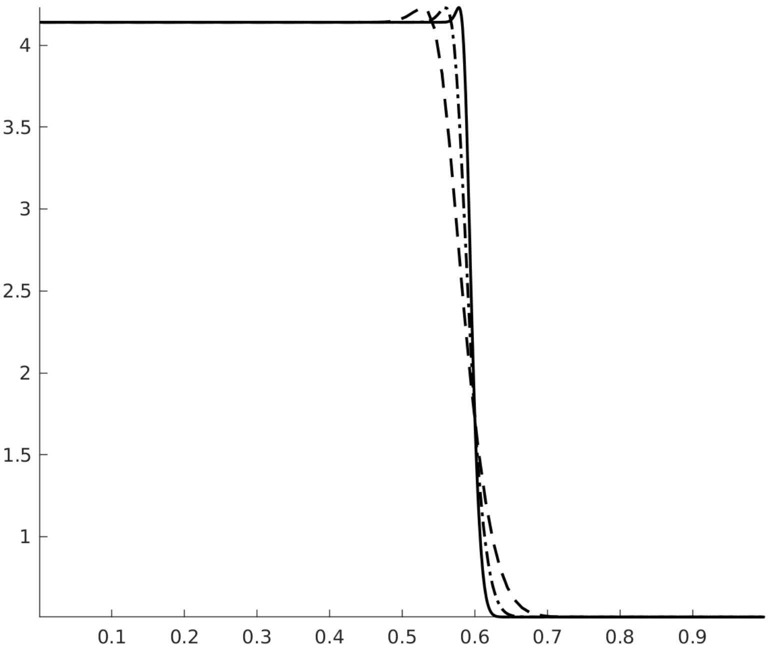
(a) t = 0.022 s

(b) t = 0.1 s

Figure 4.3: Moving Interface: entropy ($\rho s = \rho_1 s_1 + \rho_2 s_2$) profiles.



(a) t = 0.022 s



(b) t = 0.1 s

Figure 4.4: Moving Interface: specific entropy ($s = Y_1 s_1 + Y_2 s_2$) profiles.

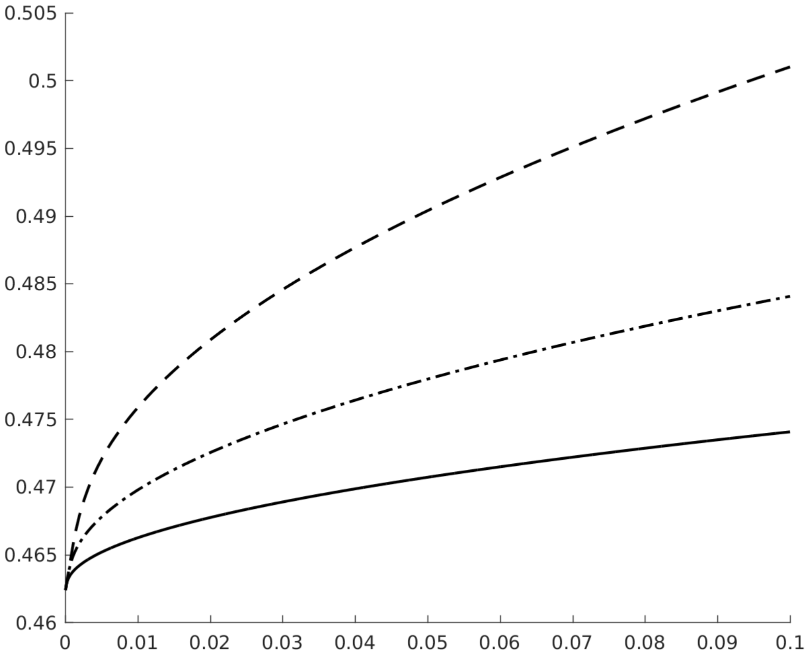


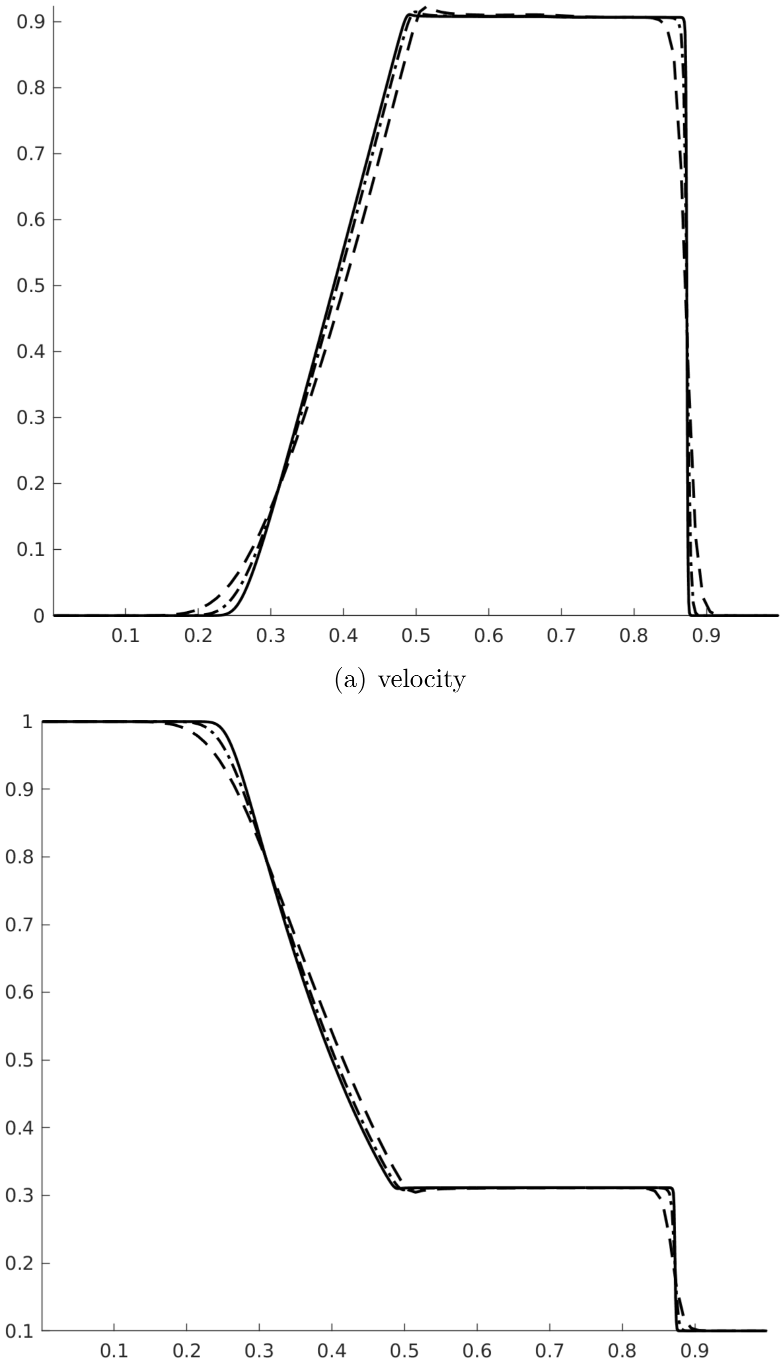
Figure 4.5: Moving Interface: total entropy ρs over time.

4.4.1.2 Shock-tube problem

We simulate the shock tube problem with two different species across the initial discontinuity. The initial conditions are given by:

$$\begin{cases} (\rho_1,\ \rho_2,\ u,\ p) = & (1,\ 0,\ 0,\ 1),\ 0 \leq x \leq 0.5, \\ (\rho_1,\ \rho_2,\ u,\ p) = & (0,\ 0.125,\ 0,\ 0.1),\ 0.5 < x \leq 1.0, \end{cases}$$

with $\gamma_1 = 1.4,\ \gamma_2 = 1.6$ and $c_{v1} = c_{v2} = 1$. Figure 4.6 shows the velocity and pressure profiles at $t = 0.2s$.



(a) velocity

(b) pressure

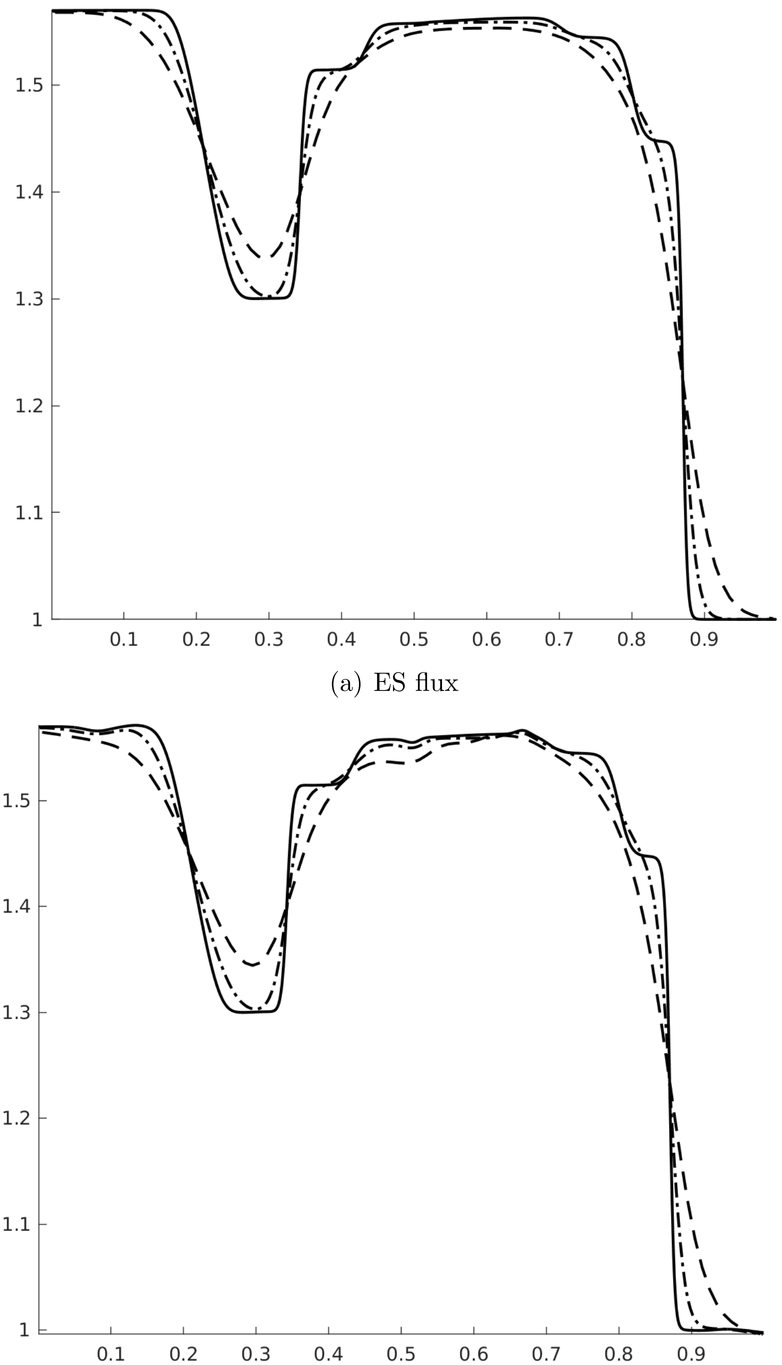
Figure 4.6: 2 species shock tube problem: solution at t = 0.2 s.

4.4.1.3 Shock-interface interaction

We simulate a test problem from Quirk & Karni [185] which consists of a shock tube filled with air, where a shock wave moves to the right and eventually meets a stationary bubble of helium at pressure equilibrium. The initial conditions are given by:

$$\begin{cases} (\rho_1,\ \rho_2,\ u,\ p) = & (1.3765,\ 0,\ 0.3948,\ 1.57),\ 0 \leq x \leq 0.25,\ \text{Post-shock, air,} \\ (\rho_1,\ \rho_2,\ u,\ p) = & (1.,\ 0,\ 0.,\ 1.),\ 0.25 \leq x \leq 0.4,\ \text{Pre-shock, air,} \\ (\rho_1,\ \rho_2,\ u,\ p) = & (0.,\ 0.139,\ 0.,\ 1.),\ 0.4 \leq x \leq 0.6,\ \text{Pre-shock, helium bubble,} \\ (\rho_1,\ \rho_2,\ u,\ p) = & (1.,\ 0,\ 0.,\ 1.),\ 0.6 \leq x \leq 1,\ \text{Post-shock, air.} \end{cases}$$

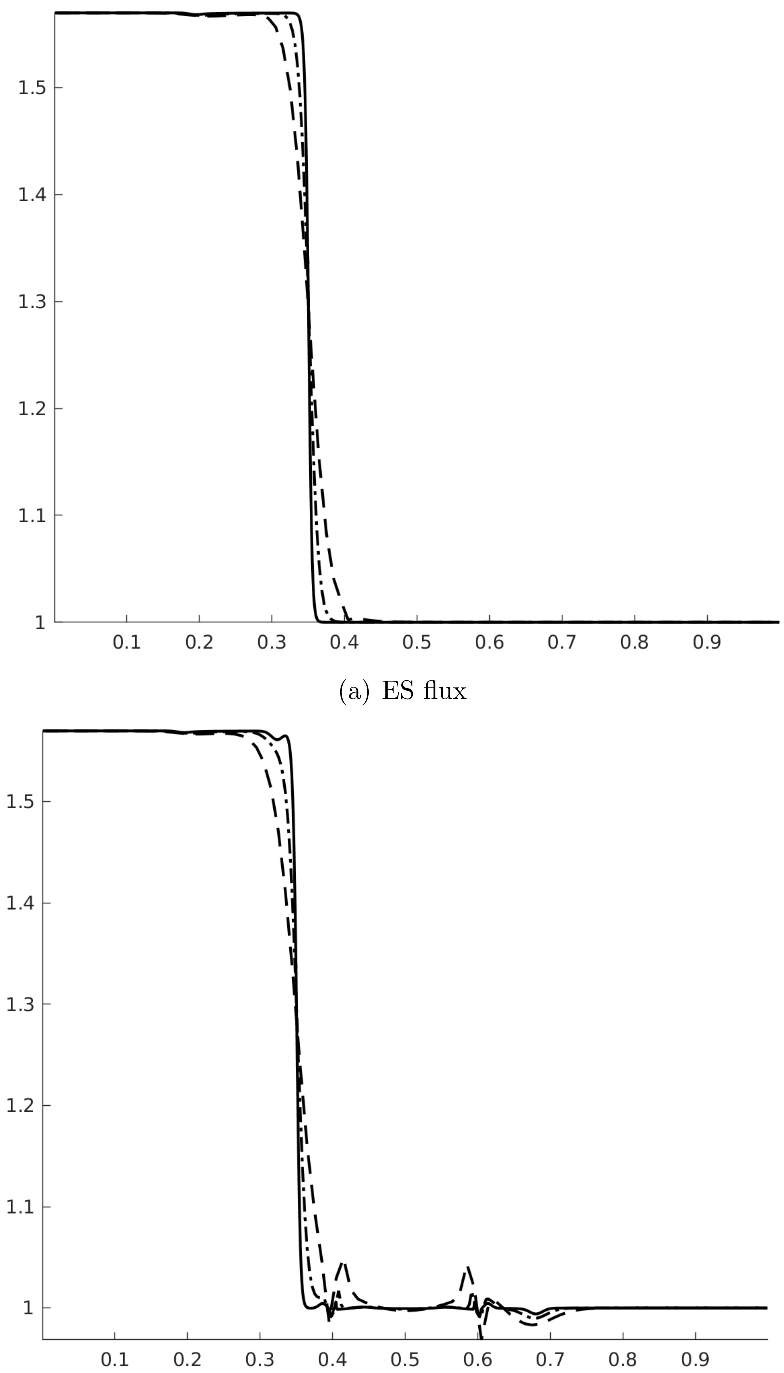
For air $c_{v1} = 0.72,\ \gamma_1 = 1.4$. For helium $c_{v2} = 2.42,\ \gamma_2 = 1.67$. In [191], this is problem is used to highlight the superiority of Karni's non-conservative scheme [185] over a conservative scheme using the Roe flux (see figure 2 in [191]). In a similar spirit, we compared our semi-discrete ES scheme with the Roe scheme. Figure 4.7 shows the pressure profile at $t = 0.35s$ obtained with each scheme. As expected, the solution with Roe's scheme is rife with oscillations unlike the solution with the present ES scheme which converges to Karni's solution without a single oscillation. The cause of this improvement is not entropy stability, but the property of preserving stationary contact discontinuities. Figure 4.8 shows the pressure profile before the right-moving shock a couple of instants before it meets the helium bubble. Roe's scheme does not preserve stationary contacts and therefore produces pressure anomalies which eventually pollute the solution at $t = 0.35$ (figure 4.7). This problem is a good illustration of the importance of treating interfaces properly in the simulation of multicomponent compressible flows. The results also suggest that the current will not produce oscillations on shock-interface interaction problems if the interface is stationary. This encouraged us to consider the next problem.



(a) ES flux

(b) Roe flux

Figure 4.7: 1D shock-bubble interaction: Solution at $t = 0.35$ s.



(a) ES flux

(b) Roe flux

Figure 4.8: 1D shock-bubble interaction: numerical solutions before the shock reaches the bubble, $t = 0.069$ s.

4.4.2 A two-dimensional case

Shock-bubble interaction

A test case that is commonly used in the development of numerical schemes for compressible multicomponent flows [192, 194, 195, 191] is the interaction of a shock wave with a cylindrical gas inhomogeneity. This problem is a two-dimensional analog of the three-dimensional shock-induced mixing concept proposed by Marble *et al.* [190] in the context of supersonic scramjet design. This problem is also used in experimental and computational investigations of the Richtmyer-Meshkov instability [187, 188].

Validating the present ES scheme against experimental data is beyond the scope of the present work. In this section, we are essentially interested in the ability of the scheme to simulate the physics relevant to this classic problem. For this purpose, we tried to reproduce the results of Marquina & Mulet [194]. The computational domain (ABCD) is shown in figure 4.9. A Mach $M_S = 1.22$ shock wave, positioned at $x = 275\ mm$, moves to the left through quiescent air (species 1, $\gamma_1 = 1.4$ and $r_1 = 0.287\ 10^3\ J.kg.^{-1}.K^{-1}$) and eventually meets a cylindrical bubble, centered at $(x, y) = [225,\ 0]\ mm$, filled with helium contaminated with 28% of air ($\gamma_2 = 1.647, r_2 = 1.578\ 10^3\ J.kg^{-1}.K^{-1}$). The flow is assumed to be symmetric about the shock-tube axis (BC), therefore only the upper half of the physical domain is considered. Reflecting boundary conditions are applied on the top (AD) and bottom (BC) boundaries. The boundary conditions upstream (AB) and downstream (CD) the shock are not crucial in this problem [185] so we simply extrapolate the flow, as in [194].

Since the current scheme is unable to guarantee positive partial densities and pressure, the initial conditions from [194] had to be modified:

$$\begin{aligned}
\text{Region I: } & (\rho_1,\ \rho_2,\ u,\ v,\ p) = (\delta\rho,\ 1.225(r_1/r_2) - \delta\rho,\ 0.,\ 0.,\ 101325), \\
\text{Region II: } & (\rho_1,\ \rho_2,\ u,\ v,\ p) = (1.225 - \delta\rho,\ \delta\rho,\ 0.,\ 0.,\ 101325), \\
\text{Region III: } & (\rho_1,\ \rho_2,\ u,\ v,\ p) = (1.6861 - \delta\rho,\ \delta\rho,\ -113.5243,\ 0.,\ 159060).
\end{aligned}$$

with $\delta\rho = 0.03$ (units for density, velocity and pressure are kg/m^3, m/s and Pa, respectively). This setup differs from the original in two aspects. First, the composition of the gas in regions I, II and III will not be the same. Second, regions I and II are in pressure and temperature equilibrium in the original setup whereas in ours, temperature equilibrium is lost. These differences make quantitative comparisons, notably with the experimental data of Haas & Sturtevant [193], difficult to carry out. However, we expect the results to be remain similar qualitatively (see Picone & Boris [189] who studied this problem using a single gas flow model).

For this simulation, we used a 4-th order TecNO scheme in space with a 4-th order explicit Runge-Kutta scheme in time. We used a 4000 × 400 grid and set the CFL number to 0.3. Figure 4.10 shows pseudo-schlieren images of the density gradients ($\phi = \exp(-\psi|\nabla\rho|/|\nabla\rho|_{\max})$, $\psi = 10Y_1 + 150Y_2$) at different times after the shock reached the bubble. These are in good agreement with those produced by Marquina & Mulet, figure 7 in [194] (see also Quirk & Karni [191], figure 9). We refer to these two references for a detailed discussion of the physical mechanisms at work.

In figure 4.12-(a), we show an x–t diagram of the position of the key features of the shock-bubble interaction. These features are explained in figure 4.12-(b). The positions of these features are obtained by looking at inflection points of horizontal sections of the shading function ϕ used in figure 4.10. The upstream bubble interface is tracked on a section at a height 20 mm from the axis. The incident shock is tracked on a section at 5 mm from the top wall. The remaining features are tracked on a

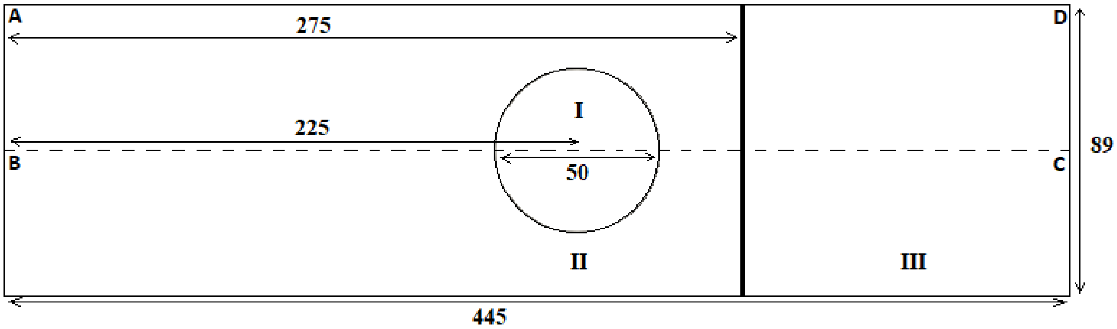
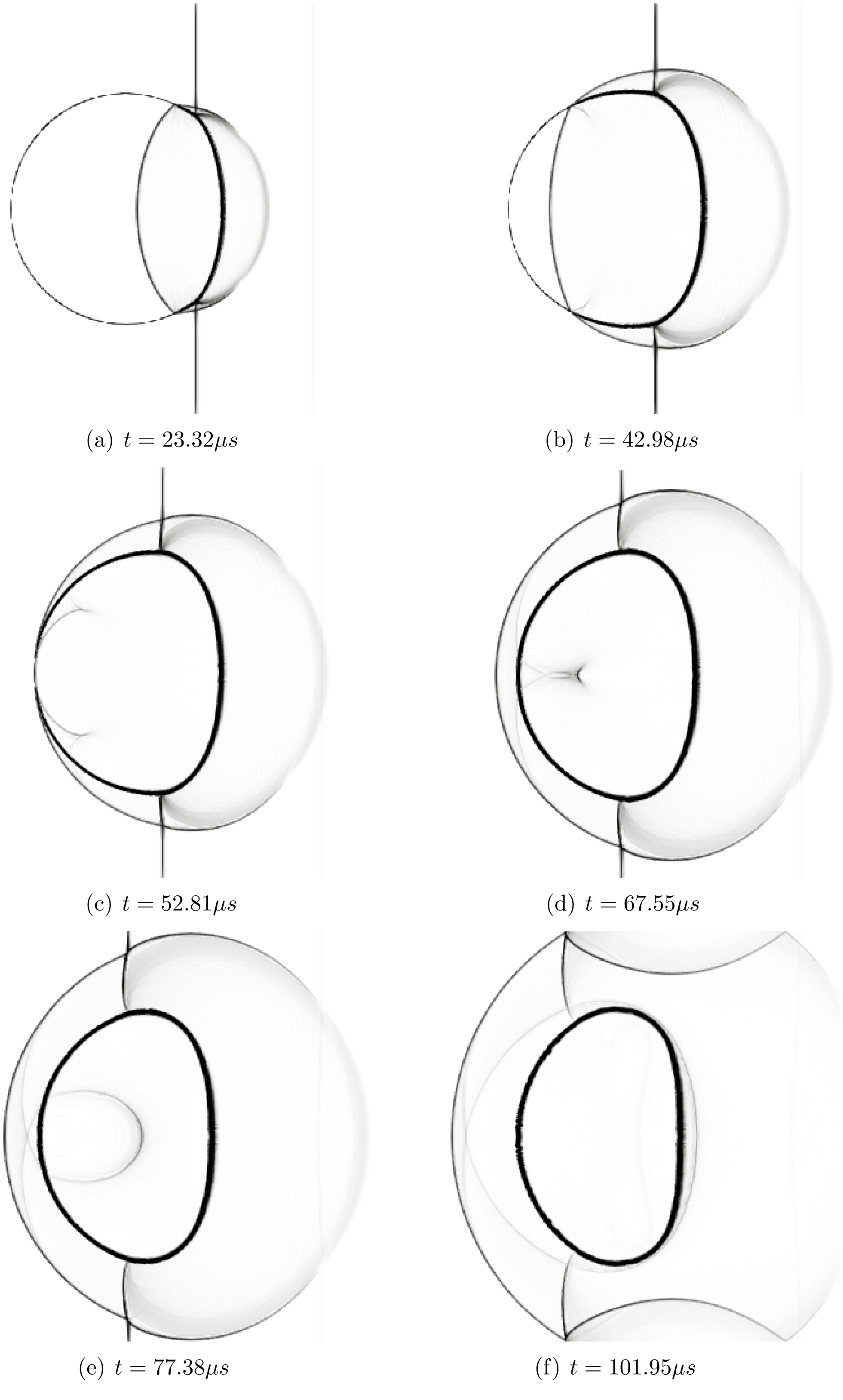


Figure 4.9: 2D Shock-bubble interaction: Computational domain (not to scale). Only the top half of the domain (ABCD) is simulated. Lengths in millimeters. Region I: Bubble. Region II: Pre-shock. Region III: Post-shock.

	VS	VR	VT	Vui1	Vui2	Vdi	Vj
Gouasmi *et al.*	422	954	377	185	105	138	228
Marquina and Mulet [194]	414	943	373	176	111	153	229

Table 4.1: Velocities in m/s of the features explained in figure 4.12. The time intervals in μs for computing each velocity are: VS [3.66, 62.64], VR [3.66, 52.81], VT [52.81, 141.26], Vui1 [3.66, 141.26], Vui2 [146.18, 254.29], Vdi [141.26, 254.29], Vj [146.18, 254.29].

section along the symmetry line. The x-t diagram from Marquina & Mulet, figure 5 in [194], shows similar trends. The mean velocities of these features are calculated from their visually straight trajectories using linear regression, and displayed in Table 4.1.



(a) $t = 23.32\mu s$

(b) $t = 42.98\mu s$

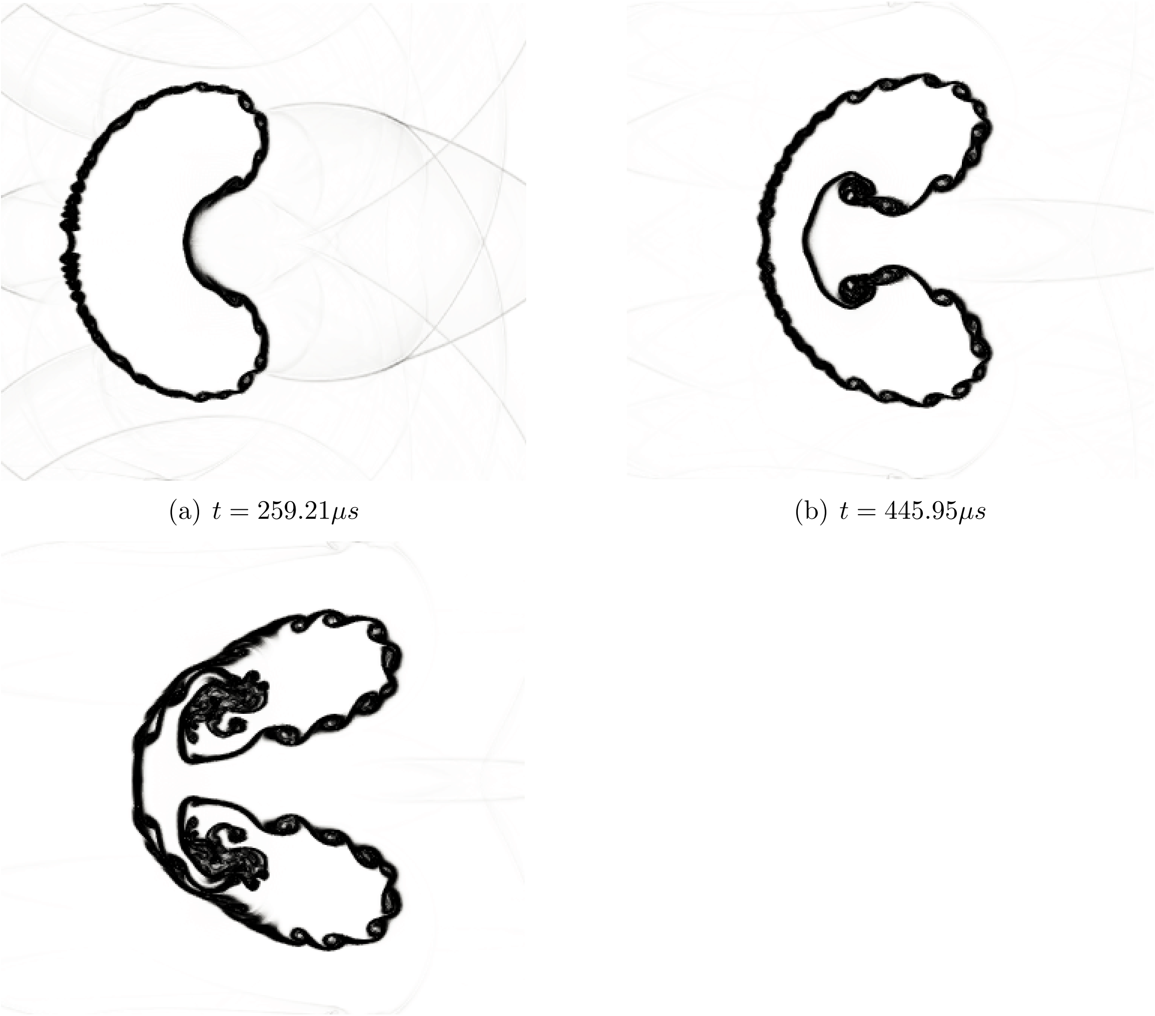
(c) $t = 52.81\mu s$

(d) $t = 67.55\mu s$

(e) $t = 77.38\mu s$

(f) $t = 101.95\mu s$

Figure 4.10: 2D Shock-bubble interaction: Pseudo-Schlieren images of density gradient.

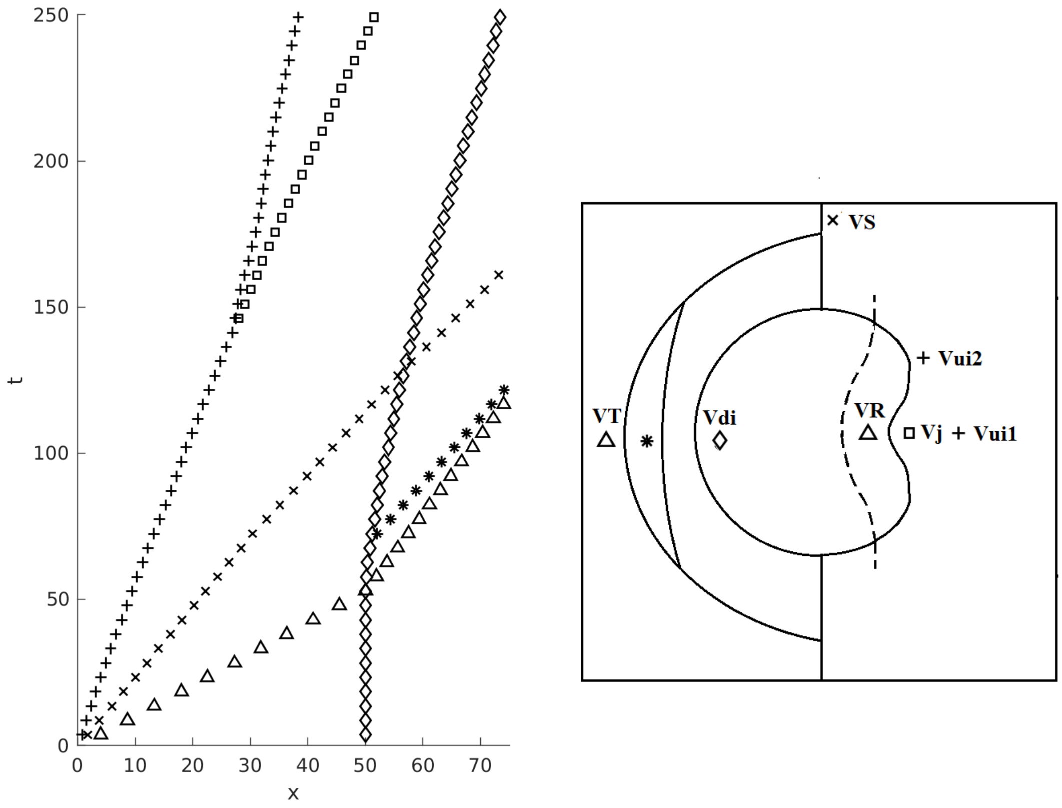


(a) $t = 259.21\mu s$

(b) $t = 445.95\mu s$

(c) $t = 676.91\mu s$

Figure 4.11: Figure 4.10 continued



(a) (b)

Figure 4.12: 2D Shock-bubble interaction: (a) x–t diagram of the key features explained in (b); (b) VS: incident shock, VR: refracted shock VT: transmitted shock, Vui: upstream border of the bubble, Vdi: downstream border of the bubble, Vj: air jet head

4.5 Summary

We have come across a few obstacles during the course of this work. The first one is theoretical: the structure required by ES schemes collapses when one of the partial densities is zero. The entropy U is no longer convex and the entropy variables, which are key in constructing ES schemes, are no longer defined. Upon closer examination, we observed that the EC flux we derived is well-defined in this limit and still satisfies the Entropy Conservation condition. We also found that the dissipation operator remains defined provided that the averaged partial densities involved in the dissipation matrix are evaluated in a certain way.

The second obstacle is that while the overall scheme is always defined, there is no guarantee that it will not produce negative densities or pressure, even at first order, at the next time step. This lack of positivity proof is not unique to ES schemes, but applies to all schemes which do not strictly enforce local wave propagation. Last but not least, numerical experiments showed that while the ES scheme can handle shocks and stationary contact discontinuities correctly, it fails to preserve pressure equilibrium and constant velocity when a moving interface is simulated.

Third and last, it is a well-known issue that conservative schemes are subject to pressure oscillations on moving interfaces. Numerical experiments showed that the ES scheme we constructed is no exception. In addition, we stress that these anomalies, which are not present in the single component case, violate neither entropy stability nor a minimum principle on the specific entropy of the mixture. Most of the early remedies to the pressure oscillation problems consist in partially giving up on conservation of total energy [186, 185, 180, 184] and possibly the ability to properly capture shocks. A compromise between ensuring entropy stability and the proper treatment of moving interfaces could perhaps be achieved with the EC/ES schemes for non-conservative hyperbolic systems developed by Castro *et al.* [198]. That being said, non-conservative schemes have their own lot of issues [199, 200].

CHAPTER V

A Minimum Entropy Principle in the Multicomponent Compressible Euler Equations

At this point in the thesis, it should be clear that satisfying an entropy inequality is not a strong enough condition to ensure the quality of the approximate solution. For good measure, we add [169, 170, 171] to the list of counter-examples.

Recall from chapter II that the entropy inequality (2.5) is only a *necessary condition* for the weak solution to be a limit solution to the regularized system (2.2). A more stringent condition one can set is that the weak solution satisfies *all entropy inequalities*. Such solutions are called *entropy solutions*.

For the compressible Euler equations, Harten [48] showed that the pairs $(U,\ F) = (-\rho h(s), -\rho u h(s))$ with $h^{'} > 0,\ h^{'} - \gamma h^{''} > 0$ are entropy pairs, and building from these, Tadmor [175] proved that entropy solutions, whether smooth or discrete, must satisfy a *minimum entropy principle*[1], namely that the spatial minimum of the specific entropy is an increasing function of time.

This result is also a necessary condition, but unlike the entropy inequality (2.5), it makes a clear statement about the *local* behavior of the physical solution. Limiting procedures for high-order schemes have been designed around this property

[1] This is neither a statement about entropy production nor a statement which applies to any PDE system with a convex extension. A reviewer suggested referring to this property as a “minimum principle of the specific entropy” to avoid the confusion.

[212, 213, 177, 209] for the compressible Euler equations. Schemes which preserve the positivity of density and the minimum entropy principle automatically preserve the positivity of pressure.

In this chapter, we seek to extend this result to entropy solutions of the multicomponent compressible Euler equations. In section 5.1, we review the system at hand. In section 5.2, we recall the original proof and motivate the two families of entropy functions we investigate in section 5.3. We end up showing a minimum entropy principle for the mixture's specific entropy. In section 5.4, we review numerical schemes which satisfy this property.

5.1 Governing equations

We consider the compressible multicomponent Euler equations [172] which consist of the conservation of species mass, momentum and total energy. In one dimension, that is equation (2.1) with the state vector $\mathbf{u}$ and flux vector $\mathbf{f}$ defined by:

$$\mathbf{u} := \begin{bmatrix} \rho_1 & \dots & \rho_N & \rho u & \rho e^t \end{bmatrix}^\top, \ \mathbf{f} := \begin{bmatrix} \rho_1 u & \dots & \rho_N u & \rho u^2 + p & (\rho e^t + p)u \end{bmatrix}^\top,$$

where ρ_k is the partial density of species k, $\rho := \sum_{k=1}^N \rho_k$ is the total density, e^t is the specific total energy, and u is the fluid velocity. The pressure p is given by the perfect gas law:

$$p := \sum_{k=1}^N \rho_k r_k T, \ r_k = \frac{R}{m_k},$$

where m_k is the molar mass of species k and R is the gas constant. The temperature T is determined by the internal energy $\rho e := \rho e^t - (\rho u)^2/(2\rho)$ which in this work is modeled following a thermally perfect gas assumption:

$$\rho e := \sum_{k=1}^N \rho_k e_k, \ e_k := e_{0k} + \int_0^T c_{vk}(\tau) d\tau.$$

For species k, e_k is the specific internal energy of species k, e_{0k} is a constant and $c_{vk} = c_{vk}(T) > 0$ is the constant volume specific heat. Other quantities which will be used in this work are given by:

$$h_k := e_k + r_k T,\ \rho c_v := \sum_{k=1}^{N} \rho_k c_{vk},\ c_{pk} := c_{vk} + r_k, \rho c_p := \sum_{k=1}^{N} \rho_k c_{pk},\ \gamma := \frac{c_p}{c_v},\ Y_k := \frac{\rho_k}{\rho}.$$

h_k is the specific enthalpy of species k, c_v is the constant volume specific heat of the gas mixture, c_p is the constant pressure specific heat of the gas mixture, γ is the specific heat ratio and Y_k is the mass fraction of species k. The thermodynamic entropy of the gas mixture is given by:

$$\rho s := \sum_{k=1}^{N} \rho_k s_k,\ s_k := \int_0^T \frac{c_{vk}(\tau)}{\tau} d\tau - r_k \ln(\rho_k)$$

Combining the transport equations for total density, species fractions and internal energy:

$$D_t \rho = -\rho \partial_x u,\ D_t Y_k = 0,\ D_t e = -\frac{p}{\rho} \partial_x u, \tag{5.1}$$

with the Gibbs relation:

$$T ds = de - \frac{p}{\rho^2} d\rho - \sum_{k=1}^{N} g_k dY_k, \tag{5.2}$$

leads to a transport equation for the specific entropy s:

$$D_t s = 0. \tag{5.3}$$

With total mass conservation, this leads to the conservation equation:

$$\partial_t(\rho s) + \partial_x(\rho s u) = 0. \tag{5.4}$$

For $\rho_k > 0,\ T > 0$, $(U, F) = (-\rho s, -\rho u s)$ is a valid entropy-entropy flux pair [174, 172]. The condition (2.4) is met as a consequence of the conservation equation. The convexity of U is established by looking at the entropy Hessian $\mathbf{G}$ given by:

$$\mathbf{G} := \frac{\partial^2 U}{\partial \mathbf{u}^2} = \frac{\partial \mathbf{v}}{\partial \mathbf{u}} = \frac{\partial \mathbf{v}}{\partial Z}\left(\frac{\partial \mathbf{u}}{\partial Z}\right)^{-1}.$$

The entropy variables $\mathbf{v}$ for the multicomponent system can be easily derived using variable changes. Define the vector of primitive variables $Z = \begin{bmatrix} \rho_1 & \dots & \rho_N & u & T \end{bmatrix}^\top$. The chain rule gives:

$$\frac{\partial U}{\partial \mathbf{u}} = \frac{\partial U}{\partial Z}\left(\frac{\partial \mathbf{u}}{\partial Z}\right)^{-1}.$$

The Gibbs identity (5.2) can be written as:

$$T dU = -d\rho e + \sum_{k=1}^{N} g_k d\rho_k, \tag{5.5}$$

where $g_k = h_k - T s_k$ is the Gibbs function of species k. From the definition of ρe we have:

$$d\rho e = \sum_{k=1}^{N} e_k d\rho_k + \rho c_v dT. \tag{5.6}$$

Combining eqs. (5.6) and (5.5), one obtains:

$$dU = \frac{1}{T}\left(\sum_{k=1}^{N}(g_k - e_k) d\rho_k - \rho c_v dT\right).$$

This gives:

$$\frac{\partial U}{\partial Z} = \frac{1}{T}\begin{bmatrix} (g_1 - e_1) & \dots & (g_N - e_N) & 0 & -\rho c_v \end{bmatrix}. \tag{5.7}$$

The Jacobian of the mapping $Z \to \mathbf{u}$ is given by:

$$\frac{\partial \mathbf{u}}{\partial Z} = \begin{bmatrix} 1 & & 0 & 0 & 0 \\ & \ddots & & \vdots & \vdots \\ 0 & & 1 & 0 & 0 \\ u & \dots & u & \rho & 0 \\ e_1 + k & \dots & e_N + k & \rho u & \rho c_v \end{bmatrix}, \tag{5.8}$$

where $k = \frac{1}{2}u^2$. The inverse of this matrix is given by:

$$\left(\frac{\partial \mathbf{u}}{\partial Z}\right)^{-1} = \begin{bmatrix} 1 & & 0 & 0 & 0 \\ & \ddots & & \vdots & \vdots \\ 0 & & 1 & 0 & 0 \\ -u\rho^{-1} & \dots & -u\rho^{-1} & \rho^{-1} & 0 \\ (k - e_1)(\rho c_v)^{-1} & \dots & (k - e_N)(\rho c_v)^{-1} & -u(\rho c_v)^{-1} & (\rho c_v)^{-1} \end{bmatrix}. \tag{5.9}$$

Combining eqs. (5.9) and (5.7) yields the entropy variables [174, 172]:

$$\mathbf{v} = \left(\frac{\partial U}{\partial \mathbf{u}}\right)^{\top} = \frac{1}{T}\begin{bmatrix} g_1 - k & \dots & g_N - k & u & -1 \end{bmatrix}^{\top}. \tag{5.10}$$

We have:

$$\frac{\partial \mathbf{v}}{\partial Z} = \begin{bmatrix} r_1/\rho_1 & & 0 & -u/T & (k - e_1)/T^2 \\ & \ddots & & \vdots & \vdots \\ 0 & & r_N/\rho_N & -u/T & (k - e_N)/T^2 \\ 0 & \dots & 0 & 1/T & -u/T^2 \\ 0 & \dots & 0 & 0 & 1/T^2 \end{bmatrix}. \tag{5.11}$$

therefore the Hessian is given by:

$$\mathbf{G} = \frac{1}{\rho c_v T^2} \begin{bmatrix} & & & & sym \\ & (\zeta_{ij}) & & & \\ & & & & \\ -u(k-(e_1-c_vT)) & \dots & -u(k-(e_N-c_vT)) & (u^2+c_vT) & \\ -(e_1-k) & \dots & -(e_N-k) & -u & 1 \end{bmatrix}, \tag{5.12}$$

with $\zeta_{ij} = (\rho c_v T^2)\big(\delta_{ij} r_i/\rho_i + u^2 c_v T\big) + (e_i - k)(e_j - k)$ for $1 \leq i, j \leq N$. The positive definiteness of the Hessian matrix $\mathbf{G}$ is not immediately visible because it is dense. However the matrix $\mathbf{H}$ defined by the congruence relation:

$$\mathbf{H} := \left(\frac{\partial \mathbf{u}}{\partial Z}\right)^{\top} \mathbf{G} \left(\frac{\partial \mathbf{u}}{\partial Z}\right) = \left(\frac{\partial \mathbf{u}}{\partial Z}\right)^{\top} \frac{\partial \mathbf{v}}{\partial Z} = \begin{bmatrix} r_1/\rho_1 & & 0 & 0 & 0 \\ & \ddots & & \vdots & \vdots \\ 0 & & r_N/\rho_N & 0 & 0 \\ 0 & \dots & 0 & \rho/T & 0 \\ 0 & \dots & 0 & 0 & \rho c_v/T^2 \end{bmatrix}, \tag{5.13}$$

is positive definite, therefore G is positive definite. This congruence relation, which we picked up from [203], will be used as well in section 5.3.

5.2 The minimum entropy principle

In this section, we review the proof of Tadmor [175] for the compressible Euler equations then discuss how to apply it to the multicomponent system.

5.2.1 Review

Integrating the inequality (2.5) over any domain Ω which induces no entropy influx across its boundaries gives:

$$\frac{d}{dt}\int_\Omega U(\mathbf{u}(x,t))dx \leq 0 \tag{5.14}$$

Integrating the above in time gives [202]:

$$\int_\Omega U(\mathbf{u}(x,t))dx \leq \int_\Omega U(\mathbf{u}(x,0))dx \tag{5.15}$$

Tadmor [205] showed that a sharper, more local version of the above inequality can be obtained:

$$\int_{|x|\leq R} U(\mathbf{u}(x,t))dx \leq \int_{|x|\leq R+t\cdot q_{max}} U(\mathbf{u}(x,0))dx, \tag{5.16}$$

where q_{max} is the maximum velocity in the domain at $t = 0$. For the Euler equations, Harten [48] sought pairs of the form $(U^h, F^h) = (-\rho h(s), -\rho u h(s))$ where $s = \ln(p) - \gamma \ln(\rho)$ is the dimensionless specific entropy (divided by the c_v, we will use the letter f instead of h in section 5.3) and h is a smooth function of S. Harten showed that the pair (U^h, F^h) is admissible if and only if h satisfies:

$$h^{'} - \gamma\ h^{''} > 0,\ h^{'} > 0. \tag{5.17}$$

For any such function h, the inequality (5.16) with $U = U^h$ gives:

$$\int_{|x|\leq R} \rho(x,t)\cdot h(s(x,t))\ dx \geq \int_{|x|\leq R+t\cdot q_{max}} \rho(x,0)\cdot h(s(x,0))\ dx. \tag{5.18}$$

Tadmor makes a special choice h_0 for the function h:

$$h_0(s) = \min[s - s_0,\ 0],\ s_0 = \underset{|x|\leq R+t\cdot q_{max}}{\text{Ess inf}}\ s(x,0).$$

s_0 is the essential infimum of the specific entropy in the domain $\Omega = \{x : |x| < R + t \cdot q_{max}\}$. From inequality (5.18), we get:

$$\int_{|x|\leq R} \rho(x,t) \cdot \min[s(x,t) - s_0,\ 0]\ dx \geq \int_{|x|\leq R+t\cdot q_{max}} \rho(x,0) \cdot \min[s(x,0) - s_0,\ 0]\ dx. \tag{5.19}$$

The right-hand side drops by definition of s_0, so equation (5.19) simplifies to:

$$\int_{|x|\leq R} \rho(x,t) \cdot \min[s(x,t) - s_0,\ 0]\ dx \geq 0. \tag{5.20}$$

The integrand on the left-hand side is negative, therefore inequality (5.20) imposes for $|x| \leq R$:

$$\min[s(x,t) - s_0,\ 0] = 0 \Leftrightarrow s(x,t) \geq \underset{|x|\leq R+t\cdot q_{max}}{\text{Ess inf}}\ s(x,0). \tag{5.21}$$

This is the minimum entropy principle satisfied by *entropy solutions to the compressible Euler equations.* A similar result holds for discrete solutions $\mathbf{u}_i^n$ (the subscript i and the superscript n refer to the cell index and time instant, respectively) which satisfy the fully-discrete entropy inequality:

$$\sum_i U(\mathbf{u}_i^{n+1}) \leq \sum_i U(\mathbf{u}_i^n), \tag{5.22}$$

for all entropies U. Taking $U = -\rho h_0(s)$ with s_0 defined as the minimum specific entropy at time instant n leads to:

$$\sum_i \rho(\mathbf{u}_i^{n+1}) \cdot \min[s(\mathbf{u}_i^{n+1}) - s_0,\ 0] \geq 0.$$

If $\rho(\mathbf{u}_i^{n+1}) > 0$, this imposes in every cell:

$$\min[s(\mathbf{u}_i^{n+1}) - s_0,\ 0] = 0\ \Leftrightarrow s(\mathbf{u}_i^{n+1}) \geq \min_i s(\mathbf{u}_i^n). \tag{5.23}$$

At first glance, injecting $U = -\rho h_0(s)$ in inequalities (5.16) and (5.22) should not be allowed because h_0 is not a smooth function of s. What makes this step valid nonetheless is the fact that h_0 *can be written as the limit of a sequence of smooth functions which satisfy Harten's conditions*. Without loss of generality, let's assume $s_0 = 0$ and consider the convolution defined as:

$$h(s) = \int_{-\infty}^{+\infty} h_0(s - \overline{s})\phi(\overline{s})d\overline{s}.$$

where ϕ is a smooth function satisfying:

$$\int_{-\infty}^{+\infty} \phi(\overline{s})d\overline{s} = 1,\ \phi(\overline{s}) > 0.$$

ϕ should also be such that the convolution is well-defined everywhere. $\phi(\overline{s}) = \exp(-\overline{s}^2)/\sqrt{\pi}$ is a valid choice. By definition of h_0, we have:

$$h(s) = \int_{s}^{+\infty} (s - \overline{s})\phi(\overline{s})d\overline{s} = s\int_{s}^{+\infty} \phi(\overline{s})d\overline{s} - \int_{s}^{+\infty} \overline{s}\phi(\overline{s})d\overline{s}.$$

h is smooth and satisfies Harten's conditions because:

$$h^{'}(s) = \int_{s}^{+\infty} \phi(\overline{s})d\overline{s} > 0,\ h^{''}(s) = -\phi(s) < 0.$$

$\forall\varepsilon > 0$, the function h_ε defined by:

$$h_\varepsilon(s) = \int_{-\infty}^{+\infty} h_0(s - \overline{s})\phi_\varepsilon(\overline{s})d\overline{s},\ \phi_\varepsilon(\overline{s}) = \frac{1}{\varepsilon}\phi\left(\frac{\overline{s}}{\varepsilon}\right), \tag{5.24}$$

is smooth and satisifies Harten's conditions as well. What is more, ϕ_ε converges, in the sense of distributions, to the Dirac delta function when $\varepsilon \to 0$ (classic result). Therefore, inequality (5.19) is obtained $h_0 = \lim_{\varepsilon\to} h_\varepsilon$.

The main takeaway of this review is that *not all entropy inequalities need to be* *satisfied for a minimum entropy principle to hold in the compressible Euler equations.* Those involving the "convolution entropies" $U = -\rho h_\varepsilon(s), \forall\varepsilon > 0$ defined by equation (5.24) are enough.

<u>Remark 1</u>: This proof and Harten's characterization (5.17) are both independent of the number of spatial dimensions [48, 175]. Throughout this manuscript, we worked in one dimension for the sake of simplicity only.

<u>Remark 2</u>: Kroner *et al.* [207] use a different approach to demonstrate that bounded entropy solutions to the quasi-1D Euler equations with discontinuous cross-section satisfy a minimum entropy principle. The inequality (5.18) is used with $h(s) = -(s_0 - s)^p,\ p > 1, s_0 > s$ (s_0 denotes an upper bound in this context), raised to the power $1/p$ and passed to the limit $p \to \infty$.

<u>Remark 3</u>: A minimum entropy principle for smooth solutions to well-designed regularizations of the Euler equations was proved by Guermond and Popov [208] (see also Delchini *et al.* [210, 211] for other systems). In this work, we are interested in the minimum entropy principle as a property of entropy solutions, *whether smooth or discrete*, to the multicomponent compressible Euler equations.

5.2.2 Elements of proof for the multicomponent compressible Euler equations

We need to formulate *what a minimum entropy principle would be* in the multicomponent case. The first option is a minimum entropy principle involving *the specific entropy of each species*:

$$s_k(x,t) \geq s_{0k} = \underset{|x|\leq R+t\cdot q_{max}}{\text{Ess inf}}\ s_k(x,0),\ 1 \leq k \leq N.$$

Working Tadmor's proof backwards, this is obtained if we can show that entropy solutions satisfy the inequality:

$$\int_{|x|\leq R} \sum_{k=1}^{N} \rho_k(x,t) \cdot f_k(s_k(x,t))\; dx \geq \int_{|x|\leq R+t\cdot q_{max}} \sum_{k=1}^{N} \rho_k(x,0) \cdot f_k(s_k(x,0))\; dx, \quad (5.25)$$

and that f_k can be taken as $f_{0k}(s_k) = \min[s_k - s_{0k},\ 0]$. This leads us to examine entropy pairs (U_I^f, F_I^f) of the form:

$$(U_I^f, F_I^f) = \Big(-\sum_{k=1}^{N} \rho_k f_k,\ -\sum_{k=1}^{N} \rho_k u f_k \Big),\ \ f_k = f_k(s_k), \quad (5.26)$$

and attempt to show that those with f_k defined as the convolution (5.24) are valid entropy pairs. The second option is a minimum entropy principle involving *the specific entropy of the gas mixture*:

$$s(x,t) \geq s_0 = \operatorname*{Ess\ inf}_{|x|\leq R+t\cdot q_{max}} s(x,0).$$

In the same vein, this is obtained if we can show that entropy solutions satisfy the inequality:

$$\int_{|x|\leq R} \rho(x,t) \cdot f(s(x,t))\; dx \geq \int_{|x|\leq R+t\cdot q_{max}} \rho(x,0) \cdot f(s(x,0))\; dx, \quad (5.27)$$

and that f can be taken as $f_0(s) = \min[s - s_0,\ 0]$. This leads us to examine entropy pairs (U_{II}^f, F_{II}^f) of the form:

$$(U_{II}^f, F_{II}^f) = (-\rho f(s), -\rho u f(s)), \quad (5.28)$$

and attempt show that those with f defined as the convolution (5.24) are valid entropy pairs.

These two families are investigated in the next section. The admissibility conditions will take the form of constraints on the first and second derivatives of f_k (first case) and f (second case). If the first and second derivatives are allowed to be strictly positive and negative, respectively, then the convolution (5.24) qualifies and a minimum entropy principle follows.

5.3 Entropy functions in the multicomponent case

For each candidate family of entropy functions, we must check for conservation and convexity with respect to the conservative variables. For a candidate entropy U^f, convexity is equivalent to the positive definiteness of its Hessian matrix $\mathbf{G}$:

$$\mathbf{G} = \frac{\partial^2 U^f}{\partial \mathbf{u}^2} = \frac{\partial \mathbf{v}^f}{\partial \mathbf{u}}, \ \mathbf{v}^f = \left(\frac{\partial U^f}{\partial \mathbf{u}}\right)^\top .$$

$\mathbf{v}^f$ is the vector of entropy variables associated with the candidate entropy.

5.3.1 Candidate I

Conservation

Equation (2.3) with $(U,F) = (U_I^f, F_I^f)$ holds if and only if $\sum_{k=1}^{N} Y_k f_k$ satisfies a transport equation. We have:

$$
\begin{aligned}
d\Big(\sum_{k=1}^{N} Y_k f_k\Big) &= \sum_{k=1}^{N} Y_k df_k + \sum_{k=1}^{N} f_k dY_k \\
&= \sum_{k=1}^{N} Y_k f_k^{'} ds_k + \sum_{k=1}^{N} f_k dY_k \\
&= \sum_{k=1}^{N} Y_k f_k^{'} \Big(\frac{c_{vk}}{T} dT - \frac{r_k}{\rho_k} d\rho_k\Big) + \sum_{k=1}^{N} f_k dY_k \\
&= \Big(\sum_{k=1}^{N} Y_k f_k^{'} c_{vk}\Big)\frac{dT}{T} - \frac{1}{\rho}\sum_{k=1}^{N} f_k^{'} r_k d\rho_k + \sum_{k=1}^{N} f_k dY_k \\
&= \Big(\sum_{k=1}^{N} Y_k f_k^{'} c_{vk}\Big)\frac{dT}{T} - \Big(\sum_{k=1}^{N} f_k^{'} Y_k r_k\Big)\frac{d\rho}{\rho} + \sum_{k=1}^{N}(f_k - r_k f_k^{'}) dY_k.
\end{aligned}
$$

From the differential relation:

$$
de = \sum_{k=1}^{N} dY_k e_k + \sum_{k=1}^{N} Y_k c_{vk} dT = \sum_{k=1}^{N} dY_k e_k + c_v dT,
$$

we obtain the following equation for temperature:

$$
D_t T = -\frac{p}{\rho c_v}\partial_x u = \frac{p}{\rho^2 c_v} D_t \rho. \tag{5.29}
$$

Using equations (5.1) and (5.29), we can show that U_I^f is conserved if and only if:

$$
\frac{1}{T}\Big(\sum_{k=1}^{N} Y_k f_k^{'} c_{vk}\Big) D_t T - \frac{1}{\rho}\Big(\sum_{k=1}^{N} f_k^{'} Y_k r_k\Big) D_t \rho = 0 \;\Leftrightarrow
$$

$$
\frac{p}{\rho T}\Big(\frac{\sum_{k=1}^{N} Y_k f_k^{'} c_{vk}}{\sum_{k=1}^{N} Y_k c_{vk}}\Big) - \Big(\sum_{k=1}^{N} f_k^{'} Y_k r_k\Big) = 0 \tag{5.30}
$$

Using the ideal gas law, this condition rewrites:

$$\frac{\sum_{k=1}^{N}\rho_k c_{vk} f_k^{'}}{\sum_{k=1}^{N}\rho_k c_{vk}} = \frac{\sum_{k=1}^{N}\rho_k r_k f_k^{'}}{\sum_{k=1}^{N}\rho_k r_k}. \tag{5.31}$$

Convexity

We have:

$$\frac{\partial s_k}{\partial \rho_k} = -\frac{r_k}{\rho_k}, \quad \frac{\partial s_k}{\partial T} = \frac{c_{vk}}{T}, \quad \frac{\partial f_k}{\partial \rho_k} = -\frac{r_k}{\rho_k} f_k^{'}, \quad \frac{\partial f_k}{\partial T} = \frac{c_{vk}}{T} f_k^{'}.$$

Therefore

$$\frac{\partial U_I^f}{\partial Z} = \begin{bmatrix} -f_1 + r_1 f_1^{'} & \dots & -f_N + r_N f_N^{'} & 0 & -\frac{1}{T}\left(\sum_{k=1}^{N}\rho_k c_{vk} f_k^{'}\right) \end{bmatrix},$$

and the entropy variables (chain rule) are given by:

$$\mathbf{v}_I^f = \begin{bmatrix} -f_1 + r_1 f_1^{'} - \beta\frac{k-e_1}{T} & \dots & -f_N + r_N f_N^{'} - \beta\frac{k-e_N}{T} & \beta\frac{u}{T} & -\beta\frac{1}{T} \end{bmatrix}^\top,$$

$$\beta = \frac{\sum_{k=1}^{N}\rho_k c_{vk} f_k^{'}}{\sum_{k=1}^{N}\rho_k c_{vk}}.$$

For simplicity, let's assume calorically perfect gases (c_{vk} and c_{pk} constants) and drop the standard formation constants. To proceed with the Hessian calculation we need the following:

$$\frac{\partial \beta}{\partial \rho_k} = \frac{c_{vk}}{\rho c_v}(f_k^{'} - r_k f_k^{''} - \beta), \quad \frac{\partial \beta}{\partial T} = \frac{\eta}{T}, \ \eta = \frac{\sum_{k=1}^{N}\rho_k c_{vk}^2 f_k^{''}}{\sum_{k=1}^{N}\rho_k c_{vk}}.$$

Denote $\xi_k = f_k^{'} - r_k f_k^{''}$ and $\mathbf{v}_I^f = [v_{1,1}^f \ \dots \ v_{1,N}^f \ v_2^f \ v_3^f]^\top$. The gradients of the last component are given by:

$$\frac{\partial v_3^f}{\partial \rho_k} = -\frac{1}{T}\frac{c_{vk}}{\rho c_v}(\xi_k - \beta), \ \frac{\partial v_3^f}{\partial u} = 0, \ \frac{\partial v_3^f}{\partial T} = \frac{\beta - \eta}{T^2}. \tag{5.32}$$

The gradients of the before-last component are given by:

$$\frac{\partial v_2^f}{\partial \rho_k} = \frac{u}{T}\frac{c_{vk}}{\rho c_v}(\xi_k - \beta), \; \frac{\partial v_2^f}{\partial u} = \frac{\beta}{T}, \; \frac{\partial v_2^f}{\partial T} = u\frac{\eta - \beta}{T^2}. \tag{5.33}$$

The gradient of the l-th component is given by:

$$\frac{\partial v_{1,l}^f}{\partial \rho_k} = \delta_{kl}\frac{r_k}{\rho_k}\xi_k - (\frac{k}{T} - c_{vl})\frac{c_{vk}}{\rho c_v}(\xi_k - \beta), \; \frac{\partial v_{1,l}^f}{\partial u} = -u\frac{\beta}{T}, \; \frac{\partial v_{1,l}^f}{\partial T} = -\frac{c_{vl}}{T}\xi_l + \frac{(\beta - \eta)k}{T^2} + c_{vl}\frac{\eta}{T}. \tag{5.34}$$

The chain rule gives for the Hessian $\mathbf{G}_I$:

$$\mathbf{G}_I = \frac{\partial \mathbf{v}_I^f}{\partial Z}\left(\frac{\partial \mathbf{u}}{\partial Z}\right)^{-1}.$$

The coefficients of the first Jacobian matrix are given by equations (5.32)-(5.34). The Hessian $\mathbf{G}_I$ is dense. We establish convexity conditions (G positive definite) by looking at the congruent matrix $\mathbf{H}_I$ defined by:

$$\mathbf{H}_I = \left(\frac{\partial \mathbf{u}}{\partial Z}\right)^{\top}\mathbf{G}_I\left(\frac{\partial \mathbf{u}}{\partial Z}\right) = \left(\frac{\partial \mathbf{u}}{\partial Z}\right)^{\top}\frac{\partial \mathbf{v}_I^f}{\partial Z},$$

instead. $\mathbf{H}_I$ is given by:

$$\mathbf{H}_I = \begin{bmatrix} \frac{r_1}{\rho_1}\xi_1 & 0 & 0 & -\frac{c_{v1}}{T}(\xi_1 - \beta) \\ 0 & \frac{r_2}{\rho_2}\xi_2 & 0 & -\frac{c_{v2}}{T}(\xi_2 - \beta) \\ 0 & 0 & \frac{\rho\beta}{T} & 0 \\ -\frac{c_{v1}}{T}(\xi_1 - \beta) & -\frac{c_{v2}}{T}(\xi_2 - \beta) & 0 & \rho c_v\frac{\beta - \eta}{T^2} \end{bmatrix}$$

$\mathbf{H}_I$ is positive definite if and only if the determinants of the major blocks of $\mathbf{H}_I$ are all positive (from Harten [48]). For the first three major blocks, this is equivalent to the requirement that $\xi_1 > 0$, $\xi_2 > 0$ and $\beta > 0$ are positive. Last:

$$\begin{aligned}
det(\mathbf{H}_I) &= \frac{\rho\beta}{T^3} r_1 r_2 \Bigg(\rho c_v (\beta - \eta) \frac{\xi_1 \xi_2}{\rho_1 \rho_2} - \frac{c_{v1}}{\gamma_1 - 1}(\xi_1 - \beta)^2 \frac{\xi_2}{\rho_2} - \frac{c_{v2}}{\gamma_2 - 1}(\xi_2 - \beta)^2 \frac{\xi_1}{\rho_1} \Bigg) \\
&= \frac{\rho\beta r_1 r_2 \xi_1 \xi_2}{\rho_1 \rho_2 T^3} \Bigg(\rho c_v (\beta - \eta) - \frac{\rho_1 c_{v1}}{\gamma_1 - 1} \frac{(\xi_1 - \beta)^2}{\xi_1} - \frac{\rho_2 c_{v2}}{\gamma_2 - 1} \frac{(\xi_2 - \beta)^2}{\xi_2} \Bigg) \\
&= \frac{\rho\beta r_1 r_2 \xi_1 \xi_2}{\rho_1 \rho_2 T^3} \Bigg(\rho_1 c_{v1} \Bigg((\beta - \eta) - \frac{1}{\gamma_1 - 1} \frac{(\xi_1 - \beta)^2}{\xi_1} \Bigg) \\
&\qquad + \rho_2 c_{v2} \Bigg((\beta - \eta) - \frac{1}{\gamma_2 - 1} \frac{(\xi_2 - \beta)^2}{\xi_2} \Bigg) \Bigg) \\
&= \frac{\rho\beta r_1 r_2 \xi_1 \xi_2}{\rho_1 \rho_2 T^3} \Bigg(\frac{\rho_1 c_{v1}}{\xi_1(\gamma_1 - 1)} \Bigg((\beta - \eta)\xi_1(\gamma_1 - 1) - (\xi_1 - \beta)^2 \Bigg) \\
&\qquad + \frac{\rho_2 c_{v2}}{\xi_2(\gamma_2 - 1)} \Bigg((\beta - \eta)\xi_2(\gamma_2 - 1) - (\xi_2 - \beta)^2 \Bigg) \Bigg) \\
&= \frac{\rho\beta r_1 r_2 \xi_1 \xi_2}{\rho_1 \rho_2 T^3} \Bigg(\frac{\rho_1 c_{v1}}{\xi_1(\gamma_1 - 1)} \Delta_1 + \frac{\rho_2 c_{v2}}{\xi_2(\gamma_2 - 1)} \Delta_2 \Bigg),
\end{aligned}$$

where $\Delta_k = (\beta - \eta)\xi_k(\gamma_k - 1) - (\xi_k - \beta)^2$. For an arbitrary number of species:

$$\mathbf{H}_I = \begin{bmatrix} \frac{r_1}{\rho_1}\xi_1 & & & 0 & -\frac{c_{v1}}{T}(\xi_1 - \beta) \\ & \ddots & & \vdots & \vdots \\ & & \frac{r_N}{\rho_N}\xi_N & 0 & -\frac{c_{vN}}{T}(\xi_N - \beta) \\ 0 & \dots & 0 & \frac{\rho\beta}{T} & 0 \\ -\frac{c_{v1}}{T}(\xi_1 - \beta) & \dots & -\frac{c_{vN}}{T}(\xi_N - \beta) & 0 & \rho c_v \frac{\beta - \eta}{T^2} \end{bmatrix}, \tag{5.35}$$

and one can easily show that:

$$det(\mathbf{H}_I) = \frac{\rho\beta}{T^3} \Bigg(\prod_{k=1}^{N} \frac{r_k \xi_k}{\rho_k} \Bigg) \Bigg(\sum_{k=1}^{N} \frac{\rho_k c_{vk}}{\xi_k(\gamma_k - 1)} \Delta_k \Bigg). \tag{5.36}$$

Overall, U^f is an admissible entropy for the multicomponent Euler equations if and only if:

$$\frac{\sum_{k=1}^N \rho_k c_{vk} f_k'}{\sum_{k=1}^N \rho_k c_{vk}} = \frac{\sum_{k=1}^N \rho_k r_k f_k'}{\sum_{k=1}^N \rho_k r_k}, \quad \xi_k > 0, \quad \beta > 0, \quad \sum_{k=1}^{N} \frac{\rho_k c_{vk}}{\xi_k(\gamma_k - 1)} \Delta_k > 0. \tag{5.37}$$

While the sufficient conditions $f_k^{'} > 0,\ f_k^{''} < 0$ for a minimum entropy principle are compatible with $\xi_k > 0$ and $\beta > 0$, it is not clear whether they are compatible with the last inequality of (5.37) (Δ_k being the difference of two positive terms). Additionally, the equality constraint (5.31) which came from the requirement of conservation does not seem to offer any option other than $f_k^{'}$ constant. Note that if $f_k^{'} > 0,\ f_k^{''} < 0$ were to violate any of the conditions derived here, it would only mean that we cannot prove a minimum entropy principle with the approach exposed in section 5.2.1. Disproving a minimum entropy principle would require a counterexample.

For the compressible Euler equations, $\mathbf{H}_I$ simplifies to:

$$\mathbf{H}_I = \begin{bmatrix} \frac{r}{\rho}\xi & 0 & -\frac{c_v}{T}(\xi-\beta) \\ 0 & \frac{\rho\beta}{T} & 0 \\ -\frac{c_v}{T}(\xi-\beta) & 0 & \rho c_v \frac{\beta-\eta}{T^2} \end{bmatrix},\ \xi = f^{'} - rf^{''},\ \beta = f^{'},\ \eta = c_v f^{''}.$$

The determinants of the three major blocks are:

$$det(H_{11}) = \frac{r}{\rho}\xi,\ det(H_{22}) = \frac{\rho}{T}\beta,\ \det(\mathbf{H}_I) = \frac{\rho r c_v \beta}{T^3(\gamma-1)}\Big((\beta-\eta)\xi(\gamma-1) - (\xi-\beta)^2\Big).$$

Using $(\gamma-1)(\beta-\eta) = (\gamma-1)f^{'} - rf^{''}$ and $\xi - \beta = -rf^{''}$, the determinant simplifies to:

$$det(\mathbf{H}_I) = \frac{\rho r c_v \beta^2}{T^3}\Big(f^{'} - c_p f^{''}\Big)$$

The necessary conditions for $\mathbf{H}_I$ to be positive definite are then:

$$f^{'} - rf^{''} > 0,\ f^{'} > 0,\ f^{'} - c_p f^{''} > 0. \qquad (5.38)$$

Since $f^{'} > 0$, the first and third inequality of (5.38) can be rewritten as:

$$\frac{f^{''}}{f^{'}} < \frac{1}{r},\ \frac{f^{''}}{f^{'}} < \frac{1}{c_p}.$$

Since $c_p > r$, the first inequality is implied by the second. Therefore, the necessary conditions (5.38) simplify to:

$$f' > 0,\ f' - c_p f'' > 0. \tag{5.39}$$

These are the well-known conditions (5.17) for the Euler equations (note that the function f in this section and the function h in section 5.2.1 are related by $f(s) = h(s/c_v)$). The conditions (5.37) are therefore consistent with Harten's in the Euler case.

5.3.2 Candidate II

Conservation

Multiplying the transport equation for the specific entropy (5.3) with f' leads to a transport equation for $f(s)$. Conservation of U^f_{II} with the entropy flux F^f_{II} then follows from the total mass conservation equation.

Convexity

We have:

$$\frac{\partial Y_j}{\partial \rho_k} = \frac{\delta_{jk}}{\rho} - \frac{\rho_j}{\rho^2},\ \frac{\partial s}{\partial \rho_k} = \frac{1}{\rho}(s_k - r_k - s),\ \frac{\partial s}{\partial T} = \frac{c_v}{T}.$$

This gives:

$$\frac{\partial U^f_{II}}{\partial Z} = \left[f'(-s_1 + r_1 + s) - f \quad \dots \quad f'(-s_N + r_N + s) - f \quad 0 \quad -\tfrac{\rho c_v}{T} f' \right], \tag{5.40}$$

and the entropy variables:

$$\begin{aligned}\mathbf{v}^f_{II} &= \begin{bmatrix} f^{'}\frac{g_1-k}{T}+f^{'}s-f & \dots & f^{'}\frac{g_N-k}{T}+f^{'}s-f & f^{'}\frac{u}{T} & -f^{'}\frac{1}{T}\end{bmatrix}^\top \\ &= f^{'}\mathbf{v}+(f^{'}s-f)\begin{bmatrix}1 & \cdots & 1 & 0 & 0\end{bmatrix}^\top. \end{aligned} \tag{5.41}$$

Again, the conditions for convexity are established by looking at the congruent matrix $\mathbf{H}_{II}$ defined by:

$$\mathbf{H}_{II} = \left(\frac{\partial \mathbf{u}}{\partial Z}\right)^\top \mathbf{G}_{II}\left(\frac{\partial \mathbf{u}}{\partial Z}\right) = \left(\frac{\partial \mathbf{u}}{\partial Z}\right)^\top \frac{\partial \mathbf{v}^f_{II}}{\partial Z}.$$

We have:

$$\frac{\partial \mathbf{v}^f_{II}}{\partial Z} = f^{'}\frac{\partial \mathbf{v}}{\partial Z} + \frac{f^{''}}{\rho}\begin{bmatrix}(g_1-k)/T+s \\ \vdots \\ (g_N-k)/T+s \\ u/T \\ -1/T\end{bmatrix}\begin{bmatrix} s_1-r_1-s & \dots & s_N-r_N-s & 0 & \frac{\rho c_v}{T}\end{bmatrix}$$

and

$$\left(\frac{\partial \mathbf{u}}{\partial Z}\right)^{\top}\begin{bmatrix}(g_1-k)/T+s\\ \vdots\\ (g_N-k)/T+s\\ u/T\\ -1/T\end{bmatrix}=\begin{bmatrix}-s_1+r_1+s\\ \vdots\\ -s_N+r_N+s\\ 0\\ -\frac{\rho c_v}{T}\end{bmatrix},$$

$$\left(\frac{\partial \mathbf{u}}{\partial Z}\right)^{\top}\frac{\partial \mathbf{v}}{\partial Z}=\begin{bmatrix}r_1/\rho_1 & & 0 & 0 & 0\\ & \ddots & & \vdots & \vdots\\ 0 & & r_N/\rho_N & 0 & 0\\ 0 & \dots & 0 & \rho/T & 0\\ 0 & \dots & 0 & 0 & \rho c_v/T^2\end{bmatrix}.$$

Therefore:

$$\mathbf{H}_{II}=f^{'}\begin{bmatrix}r_1/\rho_1 & & 0 & 0 & 0\\ & \ddots & & \vdots & \vdots\\ 0 & & r_N/\rho_N & 0 & 0\\ 0 & \dots & 0 & \rho/T & 0\\ 0 & \dots & 0 & 0 & \rho c_v/T^2\end{bmatrix}-\frac{f^{''}}{\rho}\begin{bmatrix}R_1\\ \vdots\\ R_N\\ 0\\ -\frac{\rho c_v}{T}\end{bmatrix}\begin{bmatrix}R_1 & \dots & R_N & 0 & -\frac{\rho c_v}{T}\end{bmatrix}, \tag{5.42}$$

where $R_i = -s_i + r_i + s$. We recover Harten's conditions in the compressible Euler case. At this point, we immediately note that if $f^{'} > 0,\ f^{''} < 0$ then $\mathbf{H}_{II}$ is positive definite (as the sum of a positive definite matrix and a positive semi-definite matrix). Therefore a minimum entropy principle for the gas mixture's specific entropy holds.

Continuing on the characterization of convexity, $\mathbf{H}_{II}$ writes:

$$\mathbf{H}_{II} = \frac{f'}{\rho}\begin{bmatrix} r_1/Y_1 & & 0 & 0 & 0 \\ & \ddots & & \vdots & \vdots \\ 0 & & r_N/Y_N & 0 & 0 \\ 0 & \dots & 0 & \rho^2/T & 0 \\ 0 & \dots & 0 & 0 & \rho^2 c_v/T^2 \end{bmatrix}$$

$$- \frac{f''}{\rho}\begin{bmatrix} R_1^2 & & R_1 R_N & 0 & -\frac{\rho c_v}{T} R_1 \\ & \ddots & & \vdots & \vdots \\ R_1 R_N & & R_N^2 & 0 & -\frac{\rho c_v}{T} R_N \\ 0 & \dots & 0 & 0 & 0 \\ -\frac{\rho c_v}{T} R_1 & \dots & -\frac{\rho c_v}{T} R_N & 0 & \frac{\rho^2 c_v^2}{T^2} \end{bmatrix}.$$

Let $\overline{r}_i = r_i/Y_i$ and $\eta = f' - c_v f''$, for two species we have:

$$\mathbf{H}_{II} = \frac{1}{\rho}\begin{bmatrix} f'\overline{r}_1 - f'' R_1^2 & -R_1 R_2 f'' & 0 & \frac{\rho c_v}{T} R_1 f'' \\ -R_1 R_2 f'' & f'\overline{r}_2 - f'' R_2^2 & 0 & \frac{\rho c_v}{T} R_2 f'' \\ 0 & 0 & \rho^2 f'/T & 0 \\ \frac{\rho c_v}{T} R_1 f'' & \frac{\rho c_v}{T} R_2 f'' & 0 & \frac{\rho^2 c_v}{T^2}\eta \end{bmatrix}$$

The determinants of the first three major blocks of $\mathbf{H}$ are:

$$H_{11} = \overline{r}_1\left(f' - f''\frac{R_1^2}{\overline{r}_1}\right),\ H_{22} = \overline{r}_1\overline{r}_2 f'\left(f' - f''\left(\frac{R_1^2}{\overline{r}_1} + \frac{R_2^2}{\overline{r}_2}\right)\right),\ H_{33} = \frac{\rho^2 f'}{T} H_{22}. \quad (5.43)$$

Last:

$$
\begin{aligned}
det(\rho\mathbf{H}_{II}) &= \frac{\rho^2 f'}{T}\begin{vmatrix} f'\overline{r}_1 - f''R_1^2 & -R_1R_2f'' & \frac{\rho c_v}{T}R_1f'' \\ -R_1R_2f'' & f'\overline{r}_2 - f''R_2^2 & \frac{\rho c_v}{T}R_2f'' \\ \frac{\rho c_v}{T}R_1f'' & \frac{\rho c_v}{T}R_2f'' & \frac{\rho^2 c_v}{T^2}\eta \end{vmatrix} \\
&= \frac{\rho^4 c_v f'}{T^3}\begin{vmatrix} f'\overline{r}_1 - f''R_1^2 & -R_1R_2f'' & R_1f'' \\ -R_1R_2f'' & f'\overline{r}_2 - f''R_2^2 & R_2f'' \\ c_vR_1f'' & c_vR_2f'' & \eta \end{vmatrix} \\
&= \frac{\rho^4 c_v f'}{T^3}\left(\eta H_{22} - c_vR_2^2f''\begin{vmatrix} f'\overline{r}_1 - f''R_1^2 & -R_1f'' \\ R_1 & 1 \end{vmatrix} - \right. \\
&\qquad\qquad \left. c_vR_1^2f''\begin{vmatrix} f'\overline{r}_2 - f''R_2^2 & -R_2f'' \\ R_2 & 1 \end{vmatrix}\right) \\
&= \frac{\rho^4 c_v f'}{T^3}\Big(\eta H_{22} - c_vf''f'(R_2^2\overline{r}_1 + R_1^2\overline{r}_2)\Big) \\
&= \frac{\rho^4 c_v (f')^2}{T^3}\overline{r}_1\overline{r}_2\left(\eta f' - (\eta + c_v)f''\left(\frac{R_1^2}{\overline{r}_1} + \frac{R_2^2}{\overline{r}_2}\right)\right).
\end{aligned}
$$

We obtain conditions on f involving terms of the form $f' - \alpha f''$, but unlike in the Euler case, α is not a constant. In section 4.1, the simple structure of the mapped Hessian $\mathbf{H}_I$, given by equation (5.35), allowed us to easily derive the necessary and sufficient conditions (5.37) for convexity for an arbitrary number of species. Nevertheless, we were not able to conclude on a minimum entropy principle on the specific entropy of each species. Here, the mapped Hessian $\mathbf{H}_{II}$, given by equation (5.42), is mostly dense, which complicates the task of establishing convexity conditions for an arbitrary number of species. However, we know from equation (5.42) that $f' > 0$ and $f'' < 0$ are sufficient conditions for admissibility, independently of the number of species, which is enough to conclude on a minimum entropy principle on the gas mixture's specific entropy.

5.4 Numerical schemes satisfying a minimum entropy principle

In this section, we review schemes which, by virtue of satisfying all entropy inequalities under some assumptions, satisfy a minimum entropy principle for the compressible multicomponent Euler equations.

We only discuss first-order schemes in one dimension. Extensions to high-order and multiple dimensions (including unstructured grids) can be found in [212, 213, 176, 177, 209]. These schemes are typically constructed as composite *convex* combinations of one-dimensional first-order updates. Since entropies are convex functions, any entropy inequality satisfied by the baseline one-dimensional first-order update will be satisfied by the whole scheme as well.

5.4.1 Godunov-type schemes

Let $\mathbf{w}(x/t; \mathbf{u}_L, \mathbf{u}_R)$ be the solution of the Riemann problem:

$$\partial_t \mathbf{u} + \partial_x \mathbf{f} = 0,\ \mathbf{u}(x, 0) = \begin{cases} \mathbf{u}_L, & x < 0, \\ \mathbf{u}_R, & x > 0, \end{cases} \tag{5.44}$$

where $\mathbf{u}_L$ and $\mathbf{u}_R$ are constant states. Let a_L and a_R be the smallest and largest signal velocities. Then $\mathbf{w}$ satisfies:

$$\mathbf{w}(x/t; \mathbf{u}_L, \mathbf{u}_R) = \begin{cases} \mathbf{u}_L, & x/t \leq a_L \\ \mathbf{u}_R, & x/t \geq a_R \end{cases} \tag{5.45}$$

In the Godunov scheme [12], each discontinuity in the discrete field $\mathbf{u}_i^n$ gives rise to a local Riemann problem (5.45). If $\lambda |a_{max}| < 1/2$, where a_{max} is the largest signal speed in the domain, then there is no interaction between neighboring Riemann problems and the exact solution $\mathbf{w}_{n+1}(x)$ at the next time instant writes:

$$\mathbf{w}_{n+1}(x) = \mathbf{w}((x - x_{i+\frac{1}{2}})/\Delta t; \mathbf{u}_i^n, \mathbf{u}_{i+1}^n), \text{ for } |x - x_{i+\frac{1}{2}}| \leq \Delta x/2,$$

where $x_{i+\frac{1}{2}}$ is the position of the interface between cells i and $i+1$. The Godunov scheme is obtained by averaging $\mathbf{w}_{n+1}$ in each cell:

$$\begin{aligned}\mathbf{u}_i^{n+1} &= \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathbf{w}_{n+1}(x)\ dx \\ &= \frac{1}{\Delta x}\int_0^{\Delta x/2} \mathbf{w}(x/\Delta t; \mathbf{u}_{i-1}^n, \mathbf{u}_i^n)\ dx + \frac{1}{\Delta x}\int_{-\Delta x/2}^{0} \mathbf{w}(x/\Delta t; \mathbf{u}_i^n, \mathbf{u}_{i+1}^n)\ dx.\end{aligned}$$

This update can be rewritten in conservative form:

$$\mathbf{u}_i^{n+1} = \mathbf{u}_i^n - \lambda\big(\mathbf{f}(\hat{\mathbf{w}}_{i+\frac{1}{2}}) - \mathbf{f}(\hat{\mathbf{w}}_{i-\frac{1}{2}})\big),\ \hat{\mathbf{w}}_{i+\frac{1}{2}} = \mathbf{w}(0; \mathbf{u}_i^n, \mathbf{u}_{i+1}^n),$$

with $\lambda = \Delta t/\Delta x$. An important assumption from there [27, 176] is that the exact Riemann solution is an entropy solution. This implies, for all entropies:

$$\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} U(\mathbf{w}_{n+1}(x))\ dx \leq U(\mathbf{u}_i^n) - \lambda\big(F(\hat{\mathbf{w}}_{i+\frac{1}{2}}) - F(\hat{\mathbf{w}}_{i-\frac{1}{2}})\big).$$

With Jensen's inequality:

$$U\bigg(\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathbf{w}_{n+1}(x)\ dx\bigg) \leq \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} U(\mathbf{w}_{n+1}(x))\ dx,$$

it follows that the Godunov scheme satisfies:

$$U(\mathbf{u}_i^{n+1}) \leq U(\mathbf{u}_i^n) - \lambda\big(F(\hat{\mathbf{w}}_{i+\frac{1}{2}}) - F(\hat{\mathbf{w}}_{i-\frac{1}{2}})\big). \tag{5.46}$$

This shows that the Godunov scheme inherits, by construction, all the entropy inequalities that the exact Riemann solution satisfies. This result also applies to schemes based on approximate Riemann solutions provided that they remain consistent with the integral forms of the conservation law and the entropy inequality (see Theorem 3.1 in [27]). The bottom line is that *full knowledge of the Riemann solution is not necessary.* For instance, the HLL scheme [27] qualifies if the maximum right and left wave speeds are correctly estimated (from above).

The Godunov scheme satisfies a sharper version of (5.23). Taking $U = -\rho f_0(s)$ with $s_0 = \min[s(\mathbf{u}_{i-1}^n),\ s(\mathbf{u}_i^n),\ s(\mathbf{u}_{i+1}^n)]$ in (5.46), and using the fact that the exact solution $\mathbf{w}$ is an entropy solution satisfying (5.21), it follows that the Godunov scheme satisfies:

$$s(\mathbf{u}_i^{n+1}) \geq \min[s(\mathbf{u}_{i-1}^n),\ s(\mathbf{u}_i^n),\ s(\mathbf{u}_{i+1}^n)], \tag{5.47}$$

For the compressible Euler equations, procedures for calculating the exact solution (see Toro [168]) and estimating the maximum wave speed (see Guermond & Popov [178]) are available and can be extended to the multicomponent case (a follow-up to [178] is proposed by Frolov in [204], section 4.5).

It is unclear whether the assumption that the exact Riemann solution satisfies *all* entropy inequalities is valid. To the best of the author's knowledge, there is no proof that Harten's entropies [48] are the only entropies of the compressible Euler equations. The same can be said about the entropies that we explored in section 5.3 for the multicomponent case. This precludes a direct proof where entropy inequalities are evaluated for the exact Riemann solution. Another way of proving this would be to show that the exact Riemann solution can be written as a limit solution to the regularized system (2.2) or any other system which implies all entropy inequalities. As far as the minimum entropy principle is concerned, showing that the exact Riemann solution satisfies all entropy inequalities associated with Harten's family or with the convolution entropies of section 5.2.1 would be enough.

5.4.2 The Lax-Friedrichs scheme

The Lax-Friedrichs (LxF) scheme writes:

$$\mathbf{u}_i^{n+1} = \frac{\mathbf{u}_{i-1}^n + \mathbf{u}_{i+1}^n}{2} + \frac{\lambda}{2}\big(\mathbf{f}(\mathbf{u}_{i-1}^n) - \mathbf{f}(\mathbf{u}_{i+1}^n)\big).$$

Harten (private communication in [206], section 4) observed that if the time step is small enough, the LxF scheme coincides with the Godunov scheme over a staggered grid. The solution thus inherits the entropy inequalities that the Riemann solution satisfies:

$$U(\mathbf{u}_i^{n+1}) \leq \frac{U(\mathbf{u}_{i-1}^n) + U(\mathbf{u}_{i+1}^n)}{2} + \frac{\lambda}{2}\big(F(\mathbf{u}_{i-1}^n) - F(\mathbf{u}_{i+1}^n)\big). \tag{5.48}$$

As in section 5.2.1, inequality (5.48) with $U = -\rho f_0(s)$ and $s_0 = \min[s(\mathbf{u}_{i-1}^n),\ s(\mathbf{u}_{i+1}^n)]$ leads to a minimum entropy principle:

$$s(\mathbf{u}_i^{n+1}) \geq \min[s(\mathbf{u}_{i-1}^n),\ s(\mathbf{u}_{i+1}^n)], \tag{5.49}$$

that is sharper than (5.23).

On the other hand, Lax [201] proved, without invoking Riemann solutions, that the LxF scheme can be made to satisfy (5.48) for *any* given entropy pair. We recall his proof here.

Denote $\mathfrak{u} = \mathbf{u}_i^{n+1}$, $\mathfrak{v} = \mathbf{u}_{i-1}^n$ and $\mathfrak{w} = \mathbf{u}_{i+1}^n$. The LxF scheme writes:

$$\mathfrak{u}(\mathfrak{v}, \mathfrak{w}) = \frac{\mathfrak{v} + \mathfrak{w}}{2} + \frac{\lambda}{2}(\mathbf{f}(\mathfrak{v}) - \mathbf{f}(\mathfrak{w})),$$

and the entropy inequality (5.48) can be studied by looking at the sign of the difference function:

$$\Delta\mathcal{S}(\mathfrak{v}, \mathfrak{w}) = \frac{U(\mathfrak{v}) + U(\mathfrak{w})}{2} + \frac{\lambda}{2}(F(\mathfrak{v}) - F(\mathfrak{w})) - U(\mathfrak{u}).$$

Lax [201] used a homotopy approach. Let $s \in [0\ 1]$, and define:

$$\overline{\mathfrak{v}}(s) = s\mathfrak{v} + (1-s)\mathfrak{w},\ \overline{\mathfrak{u}}(s) = \mathfrak{u}(\overline{\mathfrak{v}}(s), \mathfrak{w}).$$

Since $\overline{\mathfrak{v}}(1) = \mathfrak{v},\ \overline{\mathfrak{v}}(0) = \mathfrak{w}$, and $\Delta\mathcal{S}(\mathfrak{w}, \mathfrak{w}) = 0$, the fundamental theorem of calculus gives:

$$\Delta\mathcal{S}(\mathfrak{v}, \mathfrak{w}) = \Delta\mathcal{S}(\overline{\mathfrak{v}}(1), \mathfrak{w}) - \Delta\mathcal{S}(\overline{\mathfrak{v}}(0), \mathfrak{w}) = \int_0^1 \frac{d}{ds}\Big(\Delta\mathcal{S}(\overline{\mathfrak{v}}(s), \mathfrak{w})\Big)\ ds. \tag{5.50}$$

$\overline{\mathfrak{u}}$ and $\overline{\mathfrak{v}}$ satisfy:

$$\frac{d\overline{\mathfrak{v}}}{ds} = \mathfrak{v} - \mathfrak{w},\ \frac{d\overline{\mathfrak{u}}}{ds} = \frac{\mathfrak{v} - \mathfrak{w}}{2} + \frac{\lambda}{2}A(\overline{\mathfrak{v}})(\mathfrak{v} - \mathfrak{w}) = \frac{1}{2}\Big(I + \lambda A(\overline{\mathfrak{v}})\Big)(\mathfrak{v} - \mathfrak{w}),$$

where A is the flux Jacobian. Using chain rules and the constitutive relation (2.4), the integrand in equation (5.50) writes:

$$\frac{d}{ds}\Big(\Delta\mathcal{S}(\overline{\mathfrak{v}}(s), \mathfrak{w})\Big) = \frac{1}{2}\left(\frac{dU}{d\mathbf{u}}(\overline{\mathfrak{v}}) - \frac{dU}{d\mathbf{u}}(\overline{\mathfrak{u}})\right)\Big(I + \lambda A(\overline{\mathfrak{v}})\Big)(\mathfrak{v} - \mathfrak{w}).$$

Again, let $r \in [0\ 1]$, and define:

$$\overline{\mathfrak{w}}(r, s) = r\overline{\mathfrak{v}}(s) + (1-r)\mathfrak{w} = rs\mathfrak{v} + (1-rs)\mathfrak{w},\ \bar{\bar{\mathfrak{u}}}(r, s) = \mathfrak{u}(\overline{\mathfrak{v}}(s), \overline{\mathfrak{w}}(r)).$$

Since $\bar{\bar{\mathfrak{u}}}(1, s) = \overline{\mathfrak{v}}(s),\ \bar{\bar{\mathfrak{u}}}(0, s) = \overline{\mathfrak{u}}(s)$, the fundamental theorem of calculus gives:

$$\frac{dU}{d\mathbf{u}}(\overline{\mathfrak{v}}) - \frac{dU}{d\mathbf{u}}(\overline{\mathfrak{u}}) = \int_0^1 \frac{d}{dr}\left(\frac{dU}{d\mathbf{u}}(\bar{\bar{\mathfrak{u}}})\right)\ dr = \int_0^1 \left(\frac{d\bar{\bar{\mathfrak{u}}}}{dr}\right)^T G(\bar{\bar{\mathfrak{u}}})\ dr, \tag{5.51}$$

where G is the entropy Hessian. With:

$$\frac{d\bar{\bar{\mathfrak{u}}}}{dr} = \frac{s}{2}\Big(I - \lambda A(\overline{\mathfrak{w}})\Big)(\mathfrak{v} - \mathfrak{w})$$

and equations (5.50) - (5.51), the difference function $\Delta\mathcal{S}$ can finally be rewritten as:

$$\begin{aligned}\Delta\mathcal{S}(\mathfrak{v},\mathfrak{w}) = & \int_0^1\int_0^1 \frac{s}{4}\Big(\big(I-\lambda A(\overline{\mathfrak{w}})\big)(\mathfrak{v}-\mathfrak{w})\Big)^T G(\overline{\overline{\mathfrak{u}}})\Big(\big(I+\lambda A(\overline{\mathfrak{v}})\big)(\mathfrak{v}-\mathfrak{w})\Big)\, dsdr. \\ & = \langle\mathfrak{z},\ \mathfrak{z}\rangle_G - \lambda(\langle A(\overline{\mathfrak{w}})\mathfrak{z},\ \mathfrak{z}\rangle_G + \langle\mathfrak{z},\ A(\overline{\mathfrak{v}})\mathfrak{z}\rangle_G) - \lambda^2\langle A(\overline{\mathfrak{w}})\mathfrak{z},\ A(\overline{\mathfrak{v}})\mathfrak{z}\rangle_G.\end{aligned}$$

where $\mathfrak{z} = (\mathfrak{v}-\mathfrak{w})$ and $\langle\ ,\ \rangle_G$ is the inner product defined by:

$$\langle\mathfrak{a},\ \mathfrak{b}\rangle_G = \int_0^1\int_0^1 \frac{s}{4}\mathfrak{a}^T G(\overline{\overline{u}})\mathfrak{b}\ dsdr.$$

Since G is symmetric positive definite, $\langle\mathfrak{z},\ \mathfrak{z}\rangle_G > 0$ and one can expect the entropy inequality (5.48) to be met if λ is small enough. Within the vector space spanned by $(r,\ s)$, let c be the maximum matrix norm of A, m be the minimum eigenvalue of G and M be the maximum eigenvalue of G. Then, for $||\mathfrak{v}|| \neq 0$, if λ satisfies:

$$m - 2c\lambda M - c^2\lambda^2 M > 0 \ \Leftrightarrow\ \lambda c < \sqrt{1+(m/M)} - 1. \tag{5.52}$$

then the inequality (5.48) is met. Since U is strictly convex, $(m/M) > 0$ and the right-hand side of (5.52) is strictly positive. In other words, for *any* entropy U, there will always exist a time step small enough such that the condition (5.52) is met.

While Lax's proof does not invoke Riemann solutions, it does not completely support the statement [175] that the LxF scheme can be made to satisfy *all entropy inequalities*. The factor m/M in (5.52) is strictly positive, but also depends on the entropy at hand. The fact that we do not know all the entropies of a hyperbolic system in general leaves open the possibility that m/M can be arbitrarily small. One needs to show that there exists a strictly positive and entropy-independent lower bound K on m/M, so that under the condition:

$$\lambda c < \sqrt{1+K} - 1 \tag{5.53}$$

the LxF scheme will effectively satisfy all entropy inequalities. As far as the minimum entropy principle is concerned however, we recalled in section 5.2.1 that not all entropy inequalities need to be satisfied.

5.5 Summary

We proved a minimum entropy principle for entropy solutions to the multicomponent compressible Euler equations, extending Tadmor's result [175]. The proof was carried out in one dimension but easily follows in two and three dimensions (the characterization of the two families in section 5.3 is independent of the number of dimensions). This principle was proven for the gas mixture's specific entropy only. It would be interesting to establish whether this also holds for the specific entropy of each species. We assumed a mixture of thermally perfect gases governed by an ideal gas law. The methodology outlined here and in the work of Harten *et al.* [203], which extended Harten's characterization [48] to gases with an arbitrary equation of state, should provide helpful guidelines for those interested in taking this result farther.

While numerical schemes consistent with the entropy condition (2.5) for a given pair (U, F) can be constructed (we constructed one in the previous chapter for the pair $(-\rho s, -\rho u s)$), designing numerical schemes which lead to discrete entropy solutions is more challenging. A common trait of these schemes [27, 212, 176, 213] is that they take root in the notion of a Riemann problem and the existence of solutions satisfying all entropy inequalities.

While the minimum entropy principle is only a property of entropy solutions, it provides valuable information about the *local* behavior of the physical solution. Limiting procedures for high-order schemes have been designed around this property [212, 213, 177, 209] for the Euler equations and may henceforth prove useful in multicomponent flow simulations.

Finally, we emphasize that the present work is not meant to provide a comprehensive review of the symmetrizability of the multicomponent system. We refer the interested reader to Giovangigli & Matuszewski [173] for instance. The investigation of entropy functions carried out in section 5.3 was driven by the prospect of proving a minimum entropy principle. Harten's pioneering work [48] had broader motivations.

CHAPTER VI

The Low Mach Regime

A well known issue with numerical schemes for compressible flows is that their performance degrades in the low Mach number regime [219, 216, 231], most notably in the incompressible limit, despite the fact that the incompressible Euler equations are a particular occurrence of the more general compressible Euler equations [215, 214]. The research effort in adapting compressible flow codes to incompressible flow problems is motivated by two factors. First, it is sometimes more convenient to expand an existing and validated compressible flow solver instead of developing and validating an incompressible flow solver and juggling between the two. Second, there are many flow configurations of engineering interest that exhibit both compressible and incompressible flow phenomena at the same time. Examples include transonic flow, subsonic combustion, nozzle flows, and shock-induced shear instabilities.

Steady state calculations, which typically consist in evolving the unsteady system until a stationary solution is found, require a number of iterations which grows as the Mach number decreases. This can be explained by the CFL condition which becomes more stringent because of the acoustic eigenvalues. *Preconditioning* methods [216, 217, 218, 226, 228, 227, 230, 231] have been developed to address this specific issue. The key idea is to modify the temporal scales of the unsteady system that is iterated by pre-multiplying the time derivative by a well-chosen preconditioning matrix P. Ideally, the new CFL condition that arises from this modification should allow for faster convergence. In addition to stiffness, the converged solution also becomes less accurate. The root cause of this issue lies in the artificial viscosity or artificial dissipation introduced by upwind fluxes. Turkel [216, 217, 218] showed that the dissipation matrix of upwind fluxes contains terms which prevent the discrete equations solving the compressible system to converge to a set of discrete equations solving the incompressible system. A different perspective is provided in the work of Guillard & Viozat [226] who recalled that in the incompressible regime, pressure fluctuations in space typically scale as the square of the Mach number. By writing the discrete equations for the compressible system, and using asymptotic expansions [229] they were able to rigorously demonstrate that certain terms in the dissipation matrix of the upwind flux impose pressure fluctuations in space which scale as the Mach number instead of its square. The accuracy degradation problems can be alleviated with *Flux-Preconditioning*, which consists in modifying the upwind dissipation using matrix operations. Other remedies to this problem, which do not involve preconditioning matrices, have also been proposed [235, 236, 237, 238, 239]. The stiffness and accuracy issues are also present in the simulation of unsteady flows. The time-step restriction imposed by the CFL condition prohibits the use of explicit schemes and motivates the development of implicit schemes and nonlinear solvers. The losses in accuracy manifest by an excessive dissipation low Mach vortical structures [223, 224, 225, 235, 236], which in certain configurations need to be properly captured.

In this chapter, we build upon previous theoretical work on this topic to study how entropy-stable schemes behave in the low Mach regime. We focus on the accuracy degradation problem in unsteady calculations and analyze it from an entropy production perspective. Since entropy-stability is typically achieved by using upwind-type dissipation operators, we expect the same accuracy degradation to take place. We also consider the acoustic low Mach limit [228, 240], which has received perhaps less attention than the incompressible one but remains of practical interest.

This chapter is organized as follows. Section 6.1 introduces the non-dimensional compressible Euler equations and the two low Mach limits following [226, 228]. Section 6.2 recaps the root of the accuracy degradation problems and the flux-preconditioning technique. In section 6.3, we begin analyzing entropy-stable schemes in this context. We seek to establish whether the flux-preconditioning approach, taking the preconditioner of Miczek *et. al* [223, 224] and one of Turkel's [218, 226] as examples, is compatible with our base requirement of entropy-stability. Numerical experiments are carried out in section 6.4 and further analyzed in section 6.5, where the ideas of Guillard & Viozat [226] are used to revisit the accuracy problems from the angle of entropy production. Section 6.6 offers additional perspectives on those developments.

6.1 The compressible Euler equations in the low Mach limit

In this chapter, we write the dimensional form of the compressible Euler equations as [1]:

$$\begin{aligned}
&\frac{\partial}{\partial \hat{t}}\,(\hat{\rho}) + \hat{\nabla}^T(\hat{\rho}\hat{\mathbf{u}}) = 0,\\
&\frac{\partial}{\partial \hat{t}}\,(\hat{\rho}\hat{\mathbf{u}}) + \hat{\nabla}\cdot(\hat{\rho}\hat{\mathbf{u}}\otimes\hat{\mathbf{u}}) + \hat{\nabla}\hat{p} = 0,\\
&\frac{\partial}{\partial \hat{t}}\,\hat{\rho}(\hat{e}+\hat{k}) + \hat{\nabla}\cdot(\hat{\mathbf{u}}(\hat{\rho}(\hat{e}+\hat{k})+\hat{p})) = 0.
\end{aligned} \tag{6.1}$$

$\hat{\rho}$ is the density, $\hat{\mathbf{u}}$ is the velocity vector, $\hat{k} = \frac{1}{2}|\hat{\mathbf{u}}|^2$ is the kinetic energy, $\hat{e}$ is the internal energy and $\hat{p}$ is the pressure. The equation of state writes $\hat{p} = (\gamma-1)\hat{\rho}\hat{e}$.

To derive the incompressible system, we follow the same procedure as in Guillard & Viozat [226]. We first rewrite the equations in non-dimensional form. Let ρ_r, p_r, u_r be reference values for density, pressure and velocity magnitude, respectively, and

[1] Most the derivations carried out in this work involve the non-dimensional system. We use the hat notation for the dimensional system to enable a lighter notation for the non-dimensional one.

define a reference speed of sound $a_r = \sqrt{p_r/\rho_r}$. Let l_r be a reference length scale and define the reference time scale as $t_r = l_r/u_r$. We define the non-dimensional variables as:

$$\rho = \hat{\rho}/\rho_r,\ \mathfrak{u} = \hat{\mathfrak{u}}/u_r,\ e = \hat{e}/(a_r)^2,\ t = \hat{t}/t_r, \tag{6.2}$$

Introducing the non-dimensional differential operator $\nabla = l_r\ \hat{\nabla}$, we obtain the non-dimensional system:

$$\begin{aligned}
&\frac{\partial}{\partial t}\ (\rho) + \nabla \cdot (\rho\mathfrak{u}) = 0,\\
&\frac{\partial}{\partial t}\ (\rho\mathfrak{u}) + \nabla \cdot (\rho\mathfrak{u} \otimes \mathfrak{u}) + \frac{1}{M_r^2}\nabla p = 0,\\
&\frac{\partial}{\partial t}\ \rho(e + M_r^2 k) + \nabla \cdot (\mathfrak{u}(\rho(e + M_r^2 k) + p)) = 0,
\end{aligned} \tag{6.3}$$

where $M_r = u_r/a_r$ denotes the reference Mach number. The corresponding equation of state writes $p = (\gamma - 1)\rho e$. The second step towards the incompressible system is to consider asymptotic expansions [229] of the flow variables in powers of the reference Mach number:

$$p = p_0 + M_r p_1 + M_r^2 p_2 + \mathcal{O}(M_r^3), \tag{6.4}$$

$$\mathfrak{u} = \mathfrak{u}_0 + M_r \mathfrak{u}_1 + M_r^2 \mathfrak{u}_2 + \mathcal{O}(M_r^3),$$

$$\rho = \rho_0 + M_r \rho_1 + M_r^2 \rho_2 + \mathcal{O}(M_r^3).$$

Injecting these expressions into (6.3) and collecting terms of same order, we get:

1. Order $1/M_r^2$:

$$\nabla p_0 = 0. \tag{6.5}$$

2. Order $1/M$:

$$\nabla p_1 = 0. \tag{6.6}$$

3. Order 1:

$$\frac{\partial}{\partial t}\left(\rho_0\right) + \nabla \cdot \left(\rho_0 \mathfrak{u}_0\right) = 0, \tag{6.7}$$

$$\frac{\partial}{\partial t}\left(\rho_0 \mathfrak{u}_0\right) + \nabla \cdot \left(\rho_0 \mathfrak{u}_0 \otimes \mathfrak{u}_0\right) + \nabla p_2 = 0, \tag{6.8}$$

$$\frac{\partial}{\partial t}\left(\rho_0 e_0\right) + \nabla \cdot \left(\mathfrak{u}_0 \rho_0 e_0 + p_0\right) = 0. \tag{6.9}$$

Equations (6.5) and (6.6) imply that pressure variations in space scale as M_r^2 at least. The second order expansion thus writes: $p(x,t) = p_0(t) + M_r p_1(t) + M_r^2 p_2(x,t) = P_0(t) + M_r^2 p_2(x,t)$. If P_0 is constant then from the equation of state $\rho_0 e_0$ is constant as well and equation (6.9) implies the divergence constraint $\nabla \cdot \mathfrak{u}_0 = 0$. Injecting the divergence constraint into equation (6.7) implies that the material derivative of density is zero. Assuming that all particle paths come from regions of same density ρ_0, we get that density is constant everywhere and equations (6.7), (6.8) and (6.9) finally reduce to:

$$\rho_0 \left(\frac{\partial}{\partial t}(\mathfrak{u}_0) + \nabla \cdot \left(\mathfrak{u}_0 \otimes \mathfrak{u}_0\right) \right) + \nabla p_2 = 0, \tag{6.10}$$

$$\nabla \cdot \mathfrak{u}_0 = 0. \tag{6.11}$$

The divergence constraint also implies that the kinetic energy is conserved.

As recalled in Guillard & Nkonga [228], the incompressible system is not the only low Mach limit. In the process leading to (6.3) it was assumed that the relevant time scale was determined by the reference velocity u_r. If the time scale is defined in terms of the reference speed of sound a_r instead, that is $t_r = l_r/a_r$, then instead of (6.3), we have:

$$\begin{aligned}
&\frac{1}{M_r}\frac{\partial}{\partial t}\,(\rho) + \nabla\cdot(\rho\mathbf{u}) = 0,\\
&\frac{1}{M_r}\frac{\partial}{\partial t}\,(\rho\mathbf{u}) + \nabla\cdot(\rho\mathbf{u}\otimes\mathbf{u}) + \frac{1}{M_r^2}\nabla p = 0,\\
&\frac{1}{M_r}\frac{\partial}{\partial t}\,\rho(e + M_r^2 k) + \nabla\cdot(\mathbf{u}(\rho(e + M_r^2 k) + p)) = 0,
\end{aligned} \tag{6.12}$$

Introducing the expansion (6.4) into (6.12) and collecting terms, one gets:

1. Order $1/M_r^2$:

$$\nabla p_0 = 0. \tag{6.13}$$

2. Order $1/M_r$:

$$\frac{\partial}{\partial t}\,(\rho_0) = 0, \tag{6.14}$$

$$\frac{\partial}{\partial t}\,(\rho_0\mathbf{u}_0) + \nabla p_1 = 0, \tag{6.15}$$

$$\frac{\partial}{\partial t}\,p_0 = 0. \tag{6.16}$$

Equations (6.13) and (6.16) imply that the pressure variations in space scale as M_r, that is one order of magnitude bigger than those in the incompressible limit. With further assumptions and manipulations (the interested reader is referred to Guillard & Nkonga [228]), it can be shown that the first order pressure p_1 satisfies the wave equation with propagation speed a_0, that is the first order speed of sound.

Overall, there are at least two low Mach limits to the compressible Euler equations, which can be characterized by how pressure fluctuations scale with the reference Mach number M_r.

At a later stage in this work, we will need a clear definition of the entropy in the non-dimensional context. We first recall the dimensional conservation equation for entropy $\hat{\rho}\hat{s}$:

$$\frac{\partial(\hat{\rho}\hat{s})}{\partial\hat{t}} + \hat{\nabla}\cdot(\hat{\rho}\hat{\mathbf{u}}\hat{s}) = 0,\ \ \hat{s} = \ln\hat{p} - \gamma\ln\hat{\rho}. \tag{6.17}$$

Since we already have non-dimensional density and pressure variables, we define the non-dimensional entropy ρs accordingly from:

$$\hat{\rho}\hat{s} = \rho_r(\rho s) + \rho_r s_r,\ \ s_r = \ln p_r - \gamma\ln s_r.$$

Injecting this expression into (6.17) gives:

$$\begin{aligned}\frac{\partial(\hat{\rho}\hat{s})}{\partial\hat{t}} + \hat{\nabla}\cdot(\hat{\rho}\hat{\mathbf{u}}\hat{s}) &= \rho_r\Bigg(\frac{\partial(\rho s)}{\partial t} + \nabla\cdot(\rho\mathbf{u}s)\Bigg)\ +\ \rho_r s_r\Bigg(\frac{\partial}{\partial t}\,(\rho) + \nabla\cdot(\rho\mathbf{u})\Bigg)\\ &= \rho_r\Bigg(\frac{\partial(\rho s)}{\partial t} + \nabla\cdot(\rho\mathbf{u}s)\Bigg)\end{aligned}$$

This gives the non-dimensional equation for entropy:

$$\frac{\partial(\rho s)}{\partial t} + \nabla\cdot(\rho\mathbf{u}s) = 0, \tag{6.18}$$

If the time scale is taken as $t_r = l_r/u_r$, and:

$$\frac{1}{M_r}\frac{\partial(\rho s)}{\partial t} + \nabla\cdot(\rho\mathbf{u}s) = 0, \tag{6.19}$$

if the time scale is taken as $t_r = l_r/a_r$ instead.

6.2 Discrete analysis and flux-preconditioning

To introduce the discretization, we rewrite the non-dimensional compressible Euler equations (6.3) in conservative form:

$$\frac{\partial\mathbf{u}}{\partial t} + \frac{\partial\mathbf{f_x}}{\partial x} + \frac{\partial\mathbf{f_y}}{\partial y} + \frac{\partial\mathbf{f_z}}{\partial z} = 0, \tag{6.20}$$

where the state vector $\mathbf{u}$ and the flux vectors $\mathbf{f_x}$, $\mathbf{f_y}$ and $\mathbf{f_z}$ are defined by:

$$
\begin{aligned}
\mathbf{u} &= \begin{bmatrix} \rho & \rho u & \rho v & \rho w & \rho e^t \end{bmatrix}^T , \\
\mathbf{f_x} &= \begin{bmatrix} \rho u & \rho u^2 + p/M_r^2 & \rho uv & \rho uw & (\rho e^t + p)u \end{bmatrix}^T , \\
\mathbf{f_y} &= \begin{bmatrix} \rho v & \rho uv & \rho v^2 + p/M_r^2 & \rho vw & (\rho e^t + p)v \end{bmatrix}^T , \\
\mathbf{f_z} &= \begin{bmatrix} \rho w & \rho uw & \rho vw & \rho w^2 + p/M_r^2 & (\rho e^t + p)w \end{bmatrix}^T ,
\end{aligned}
$$

with $e^t = e + M_r^2 k$.

The general form of a finite-volume discretization is, in cell Ω_i:

$$\frac{d\mathbf{u}_i}{dt} + \frac{1}{V_i} \int_{\delta\Omega_i} \mathbf{f}^* dS = 0, \tag{6.21}$$

where $\mathbf{f}^* = \mathbf{f}^*(\mathbf{u}_i, \mathbf{u}_j, \mathbf{n})$ denotes the numerical flux along the interface $\delta\Omega_i$ ($\mathbf{u}_j$ is the neighboring cell state value) whose normal vector writes $\mathbf{n} = [n_x,\ n_y,\ n_z]$. A classic choice is the Roe flux:

$$\mathbf{f}^*(\mathbf{u}_L, \mathbf{u}_R, \mathbf{n}) = \frac{1}{2}(\mathbf{f}(\mathbf{u}_L) + \mathbf{f}(\mathbf{u}_R)) - \frac{1}{2}|A|(\mathbf{u}_R - \mathbf{u}_L),\ |A| = R|\Lambda|R^{-1}. \tag{6.22}$$

where $\mathbf{f}(\mathbf{u}) = n_x\mathbf{f_x}(\mathbf{u}) + n_y\mathbf{f_y}(\mathbf{u}) + n_z\mathbf{f_z}(\mathbf{u})$ is the flux in the direction normal to the interface and A is the corresponding Jacobian:

$$A = n_x \frac{\partial \mathbf{f_x}}{\partial \mathbf{u}} + n_y \frac{\partial \mathbf{f_y}}{\partial \mathbf{u}} + n_z \frac{\partial \mathbf{f_z}}{\partial \mathbf{u}}. \tag{6.23}$$

In the low-Mach number regime, the accuracy of such a scheme, deteriorates as the Mach number goes to zero. Turkel [216, 217] explained that it is because that the dissipation matrix $|A|$ contains terms which prevent the discrete equations solving the compressible system to converge to a set of discrete equations for the incompressible system in the low Mach limit. To illustrate, Turkel considers [216] the simple case of a 2-by-2 hyperbolic system with the following Jacobian:

$$A = \begin{bmatrix} u & a/M_r \\ a/M_r & u \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} u + a/M_r & 0 \\ 0 & u - a/M_r \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1}. \tag{6.24}$$

It has eigenvalues $u + a/M_r$ and $u - a/M_r$. In the subsonic regime, $|u + a/M_r| = u + a/M_r$ and $|u - a/M_r| = -(u - a/M_r)$. This change of sign leads to a dissipation matrix which does not possess the same scaling behavior as the flux Jacobian:

$$|A| = \begin{bmatrix} a/M_r & u \\ u & a/M_r \end{bmatrix} = \begin{bmatrix} \mathcal{O}(1/M_r) & \mathcal{O}(1) \\ \mathcal{O}(1) & \mathcal{O}(1/M_r) \end{bmatrix} \neq \begin{bmatrix} \mathcal{O}(1) & \mathcal{O}(1/M_r) \\ \mathcal{O}(1/M_r) & \mathcal{O}(1) \end{bmatrix}.$$

This difference in scaling behavior is the root cause of the accuracy degradation issues. The dissipation term is an important component of the flux. It cannot be discarded because the scheme would be less robust.

The flux-preconditioning technique can be seen a compromise between stability and correct low-Mach behavior. It consists in replacing the dissipation matrix $|A|$ with $P^{-1}|PA|$ where P is a preconditioning matrix. The resulting interface flux then writes:

$$\mathbf{f}^*(\mathbf{u}_L, \mathbf{u}_R, \mathbf{n}) = \frac{1}{2}(\mathbf{f}(\mathbf{u}_L) + \mathbf{f}(\mathbf{u}_R)) - \frac{1}{2}P^{-1}|PA|(\mathbf{u}_R - \mathbf{u}_L). \tag{6.25}$$

P should correct the asympotic behavior of the dissipation term in the low-Mach regime and only be active in this regime ($P \to I$ as $M_r \to 1$).

The design of such a matrix P is not straightforward, even though it is clear that the acoustic eigenspace $\lambda = u \pm a/M_r$ of the dissipation matrix should be targeted. To simplify the analysis, similarity transformations are typically used. They amount to considering the compressible Euler equations expressed in a alternative set of variables

$\mathbf{z}$. Define:

$$A_{\mathbf{z}} = Q^{-1}AQ, \; Q = \left(\frac{\partial \mathbf{u}}{\partial \mathbf{z}}\right).$$

First, a preconditioning matrix $P_{\mathbf{z}}$ is sought so that $P_{\mathbf{z}}^{-1}|P_{\mathbf{z}}A_{\mathbf{z}}|$ has appropriate Mach number scalings. The preconditioning matrix P in terms of the conservative variables $\mathbf{u}$ is then derived from the similarity relation $P = QP_{\mathbf{z}}Q^{-1}$. Indeed, one has:

$$P^{-1}|PA_{\mathbf{u}}| = QP_{\mathbf{z}}^{-1}Q^{-1}|QP_{\mathbf{z}}Q^{-1}QA_{\mathbf{z}}Q^{-1}| = QP_{\mathbf{z}}^{-1}(|P_{\mathbf{z}}A_{\mathbf{z}}|)Q^{-1}$$

As a matter of course, this strategy is efficient only if $A_{\mathbf{z}}$ has a simpler structure than A. With the *differential entropy variables* [2] defined by:

$$d\mathbf{z} = (dp/(\rho a M_r), \; du, \; dv, \; dw, \; dp - a^2 d\rho),$$

the Jacobian writes:

$$A_{\mathbf{z}} = \begin{bmatrix} u_n & n_x a/M_r & n_y a/M_r & n_z a/M_r & 0 \\ n_x a/M_r & u_n & 0 & 0 & 0 \\ n_y a/M_r & 0 & u_n & 0 & 0 \\ n_z a/M_r & 0 & 0 & u_n & 0 \\ 0 & 0 & 0 & 0 & u_n \end{bmatrix}, \; u_n = n_x u + n_y v + n_z w, \quad (6.26)$$

[2] These variables are referred to as the "entropy variables" in the literature [218, 225]. The naming "differential entropy variables" is introduced to avoid confusion with the entropy variables $\mathbf{v}$ that ES schemes are centered around.

and has the eigendecomposition $A_{\mathbf{z}} = R_{\mathbf{z}}\Lambda R_{\mathbf{z}}^{-1}$ with

$$R_{\mathbf{z}} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & -n_z & n_y & n_x & n_x \\ n_z & 0 & -n_x & n_y & n_y \\ -n_y & n_x & 0 & n_z & n_z \\ -n_x & -n_y & -n_z & 0 & 0 \end{bmatrix}, \ \Lambda = diag(u_n,\ u_n,\ u_n,\ u_n,\ u_n{+}a/M_r,\ u_n{-}a/M_r]). \tag{6.27}$$

The 2-by-2 hyperbolic system (6.24) of Turkel in [216] is a specific case of system (6.26). For this system, Turkel *et al.* [218, 216] established the following necessary condition on $P_{\mathbf{z}}$ for convergence in the low Mach limit:

$$P_{\mathbf{z}}^{-1}|P_{\mathbf{z}}A_{\mathbf{z}}| = \begin{bmatrix} \mathcal{O}(1/M^2) & \mathcal{O}(1/M) & \mathcal{O}(1/M) & \mathcal{O}(1/M) & 0 \\ \mathcal{O}(1/M) & \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(1) & 0 \\ \mathcal{O}(1/M) & \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(1) & 0 \\ \mathcal{O}(1/M) & \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(1) & 0 \\ 0 & 0 & 0 & 0 & \mathcal{O}(1) \end{bmatrix}. \tag{6.28}$$

They also showed that this is achieved with the Turkel preconditioning matrix:

$$P_{\mathbf{z}} = \begin{bmatrix} p^2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{6.29}$$

where $p = \min(\max(M_r, M_{cut}), 1)$. The parameter p is defined in such a way that $P_{\mathbf{z}}$ is always invertible (the cut-off Mach number M_{cut} prevents $p \to 0$ and $P_{\mathbf{z}}$ singular) and $P_{\mathbf{z}}$ approaches the identity matrix when $M_r \to 1$. Indeed, one has $P_{\mathbf{z}}A_{\mathbf{z}} = R_p\Lambda_p R_p^{-1}$

with

$$R_p = \left[\begin{bmatrix} 0 & 0 & 0 \\ 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \\ -n_x & -n_y & -n_z \end{bmatrix} \tfrac{1}{(K1-K2)} \begin{bmatrix} K_1 & K_2 \\ n_x & n_x \\ n_y & n_y \\ n_z & n_z \\ 0 & 0 \end{bmatrix}\right],$$

$$\Lambda_p = diag([u_n,\ u_n,\ u_n,\ u_n,\ u_{np}+a_p,\ u_{np}-a_p]),$$

$$u_{np} = \frac{1}{2}u_n(p^2+1),\ a_p = (u_{np}^2 + p^2((a/M)^2 - u_n^2))^{1/2},$$

$$K_1 = (u_{np} - u_n + a_p)M_r/a,\ K_2 = (u_{np} - u_n - a_p)M_r/a.$$

The modified dissipation matrix thus writes:

$$P_{\mathbf{z}}^{-1}|P_{\mathbf{z}}A_{\mathbf{z}}|$$

$$= \begin{bmatrix} C_0 & n_xC_1 & n_yC_1 & n_zC_1 & 0 \\ n_xC_2 & n_x^2C_3 + (1-n_x^2)u_n & n_xn_yC_4 & n_xn_zC_4 & 0 \\ n_yC_2 & n_yn_xC_4 & n_y^2C_3 + (1-n_y^2)u_n & n_yn_zC_4 & 0 \\ n_zC_2 & n_zn_xC_4 & n_zn_yC_4 & n_x^3C_3 + (1-n_z^2)u_n & 0 \\ 0 & 0 & 0 & 0 & u_n \end{bmatrix},$$

$$C_0 = (a_p^2 + u_{np}^2 - u_nu_{np})/(a_pp^2),\ C_1 = M_ru_{np}(a_p + u_n - u_{np})(a_p - u_n + u_{np})/(aa_pp^2),$$

$$C_2 = (au_{np})/(M_ra_p),\ C_3 = a_p + u_{np}(u_n - u_{np})/a_p,$$

$$C_4 = (a_p - u_{np})(a_p - u_n + u_{np})/a_p,$$

and meets condition (6.28). A different perspective is provided in the work of Guillard & Viozat [226] who observed that in the incompressible regime, pressure fluctuations in space typically scale as M_r^2. By applying the process described in section 6.1 to the discrete equations, they were able to rigorously demonstrate that certain terms in the dissipation matrix of the upwind flux impose pressure fluctuations in space which scale as M_r instead. Additionally, they show that with Turkel's preconditioner (6.29), the proper scaling of pressure fluctuations is recovered.

The second preconditioning matrix we consider in this chapter was more recently introduced by Miczek *et al.* [223, 224] for insteady calculations. It writes:

$$P_{\mathbf{z}} = \begin{bmatrix} 1 & n_x p & n_y p & n_z p & 0 \\ -n_x p & 1 & 0 & 0 & 0 \\ -n_y p & 0 & 1 & 0 & 0 \\ -n_z p & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} . \tag{6.30}$$

with $p = 1 - 1/\delta,\ \delta = \min(\max(M_r, M_{cut}), 1)$. One has $P_{\mathbf{z}} A_{\mathbf{z}} = R_p \Lambda_p R_p^{-1}$ with

$$R_p = \left[\begin{bmatrix} 0 & 0 & 0 \\ 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \\ -n_x & -n_y & -n_z \end{bmatrix} \tfrac{1}{(K1-K2)} \begin{bmatrix} K_1 & K_2 \\ n_x & n_x \\ n_y & n_y \\ n_z & n_z \\ 0 & 0 \end{bmatrix} \right] ,$$

$$\Lambda_p = diag(u_n,\ u_n,\ u_n,\ u_n,\ u_n + a_p,\ u_n - a_p]),$$

$$a_p = \sqrt{(p^2+1)a^2/M_r^2 - p^2 u_n^2},$$

$$K_1 = (a + M_r p u_n)/(M_r a_p - ap),\ K_2 = -(a + M_r p u_n)/(M_r a_p + ap).$$

The modified dissipation matrix writes:

$$
\begin{gathered}
P_{\mathbf{z}}^{-1}|P_{\mathbf{z}}A_{\mathbf{z}}| \\
= \begin{bmatrix}
C_0 & n_xC_1 & n_yC_1 & n_zC_1 & 0 \\
n_xC_2 & n_x^2C_3 + (1-n_x^2)u_n & n_xn_yC_4 & n_xn_zC_4 & 0 \\
n_yC_2 & n_yn_xC_4 & n_y^2C_3 + (1-n_y^2)u_n & n_yn_zC_4 & 0 \\
n_zC_2 & n_zn_xC_4 & n_zn_yC_4 & n_x^3C_3 + (1-n_z^2)u_n & 0 \\
0 & 0 & 0 & 0 & u_n
\end{bmatrix},
\end{gathered}
$$

$$
\begin{gathered}
C_0 = (a_p + u_n)/(p^2+1) + 2u_n(K_2 - p)/((p^2+1)(K_1 - K_2)), \\
C_1 = (p((K_2 - K_1)a_p + (K_1 + K_2)u_n) - 2K_1K_2u_n)/((p^2+1)(K_1 - K_2)), \\
C_2 = (2u_n + p((K_1 - K_2)a_p + (K_1 + K_2)u_n))/((p^2+1)(K_1 - K_2)), \\
C_3 = ((K_1 - K_2)a_p - (K_1 + K_2)u_n - 2K_1K_2pu_n)/((p^2+1)(K_1 - K_2)), \\
C_4 = -u_n + ((K_1 - K_2)a_p - (K_1 + K_2)u_n - 2K_1K_2pu_n)/((p^2+1)(K_1 - K_2)),
\end{gathered}
$$

and satisfies Turkel's necessary condition (6.28) as well:

$$
P_{\mathbf{z}}^{-1}|P_{\mathbf{z}}A_{\mathbf{z}}| = \begin{bmatrix}
\mathcal{O}(1) & \mathcal{O}(1/M) & \mathcal{O}(1/M) & \mathcal{O}(1/M) & 0 \\
\mathcal{O}(1/M) & \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(1) & 0 \\
\mathcal{O}(1/M) & \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(1) & 0 \\
\mathcal{O}(1/M) & \mathcal{O}(1) & \mathcal{O}(1) & \mathcal{O}(1) & 0 \\
0 & 0 & 0 & 0 & \mathcal{O}(1)
\end{bmatrix}. \tag{6.31}
$$

This flux-preconditioning matrix was designed to meet the more stringent condition (6.31) that $P_{\mathbf{z}}^{-1}|P_{\mathbf{z}}A_{\mathbf{z}}|$ has the same Mach number scalings as $A_{\mathbf{z}}$. It is argued [223, 224, 225] that meeting this condition, which implies Turkel's (6.28), improves the accuracy of the scheme in the low Mach limit, in particular in the acoustic limit. It was recently shown by Bruel *et al.* [240] that while flux-preconditioning with the

Turkel matrix improves the accuracy in the incompressible limit, it also leads to a numerical scheme which handles acoustic waves poorly. Acoustic waves are damped significantly faster than without any flux-preconditioning.

Many other preconditioning matrices have been proposed in the literature [227, 230, 231] with the acceleration of steady state calculations as the primary focus. We do not cover them in this work.

6.3 Flux-preconditioning and Entropy-stability

We now return to the numerical framework of this thesis. At first-order, the only difference between conventional finite-volume schemes and entropy-stable schemes is the choice of interface flux $\mathbf{f}^*$, whose general form is:

$$\mathbf{f}^*(\mathbf{u}_L, \mathbf{u}_R, \mathbf{n}) = \mathbf{f}^{\mathbf{EC}}(\mathbf{u}_L, \mathbf{u}_R, \mathbf{n}) - \frac{1}{2} D(\mathbf{v}_R - \mathbf{v}_L), \tag{6.32}$$

where $\mathbf{f}^{\mathbf{EC}}$ denotes an entropy-conservative flux and D is a positive definite matrix. The dissipation matrix used in practice is $D = R|\Lambda|R^T$ where the eigenvectors R of the flux jacobian A are scaled so that $RR^T = H$.

6.3.1 Preliminaries

Since we are working with non-dimensional equations, we need to redefine the fundamental components of an entropy-stable scheme. The entropy equation writes:

$$\begin{gathered}\frac{\partial U}{\partial t} + \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} = 0, \\ U = -\rho s/(\gamma - 1),\ F_x = -\rho u s/(\gamma - 1),\ F_y = -\rho v s/(\gamma - 1),\ F_z = -\rho w s/(\gamma - 1).\end{gathered} \tag{6.33}$$

While the expression of the entropy function $U = -\rho s/(\gamma - 1)$ does not differ from the usual, the entropy variables do because the vector of conserved variables $\mathbf{u}$ now contains a M_r factor in the total energy component. We have:

$$\mathbf{v} = \left(\frac{\partial U}{\partial \mathbf{u}}\right)^T = \begin{bmatrix} \frac{\gamma - s}{\gamma - 1} - M_r^2 \frac{\rho k}{p} & M_r^2 \frac{\rho u}{p} & M_r^2 \frac{\rho v}{p} & M_r^2 \frac{\rho w}{p} & -\frac{\rho}{p} \end{bmatrix}^T . \tag{6.34}$$

The mapping from $\mathbf{v}$ to $\mathbf{u}$ is given by:

$$\mathbf{u} = \rho \begin{bmatrix} 1 & -\frac{v_2}{v_5}\frac{1}{M_r^2} & -\frac{v_3}{v_5}\frac{1}{M_r^2} & -\frac{v_4}{v_5}\frac{1}{M_r^2} & -\frac{1}{(\gamma-1)v_5} + \frac{1}{2M_r^2}\Big(\big(\frac{v_2}{v_5}\big)^2 + \big(\frac{v_3}{v_5}\big)^2 + \big(\frac{v_4}{v_5}\big)^2\Big) \end{bmatrix}^T ,$$
$$\rho = \exp\left(v_1 - \frac{\gamma}{\gamma - 1} - \frac{\ln(-v_5)}{\gamma - 1} - \frac{1}{2M_r^2 v_5}\big(v_2^2 + v_3^2 + v_4^2\big)\right).$$

Let $F = n_x F_x + n_y F_y + n_z F_z$. The potential function $\mathcal{F}$ in space writes:

$$\mathcal{F} = \mathbf{v}^T \mathbf{f} - F = \begin{bmatrix} \frac{\gamma - s}{\gamma - 1} - M_r^2 \frac{\varrho}{p} k \\ M_r^2 \frac{\rho u}{p} \\ M_r^2 \frac{\rho v}{p} \\ M_r^2 \frac{\rho w}{p} \\ -\frac{\varrho}{p} \end{bmatrix} \cdot \begin{bmatrix} \rho u_n \\ \rho u u_n + n_x (p/M_r^2) \\ \rho v u_n + n_y (p/M_r^2) \\ \rho w u_n + n_z (p/M_r^2) \\ u_n \left(\frac{\gamma p}{(\gamma - 1)} + M_r^2 \rho k \right) \end{bmatrix} + \frac{\rho u_n s}{\gamma - 1} = \rho u_n .$$

Last, the temporal Jacobian writes:

$$H = \frac{\partial \mathbf{u}}{\partial \mathbf{v}} =$$
$$\begin{bmatrix} \rho & \rho u & \rho v & \rho w & \rho e^t \\ & \rho u^2 + \frac{p}{M_r^2} & \rho u v & \rho u w & \big(\rho e^t + p\big) u \\ & & \rho v^2 + \frac{p}{M_r^2} & \rho v w & \big(\rho e^t + p\big) v \\ & & & \rho w^2 + \frac{p}{M_r^2} & \big(\rho e^t + p\big) w \\ sym & & & & \rho (e^t)^2 + p\big(\frac{p}{(\gamma-1)\rho} + M_r^2 (u^2 + v^2 + w^2)\big) \end{bmatrix} .$$

6.3.2 Entropy-Conservative fluxes

In conventional finite-volume methods, only the dissipation part of the flux is modified because the central flux does not introduce terms that introduce inappropriate Mach number scalings. What about an entropy-conservative flux? The first entropy-conservative flux of Tadmor is given by the straight path integral:

$$\mathbf{f}^{\mathbf{EC}} = \int_0^1 \mathbf{f}(\mathbf{v}(\xi))\, d\xi,$$

therefore it has the same scaling as $\mathbf{f}$. The entropy-conservative flux of Chandrasekhar [63] has also the correct scaling as it writes, in non-dimensional variables, $\mathbf{f}^{\mathbf{EC}} = \begin{bmatrix} f_1 & f_2 & f_3 & f_4 & f_5 \end{bmatrix}$ with:

$$\begin{aligned}
f_1 &= \rho^{ln} u_n, \\
f_2 &= n_x \frac{1}{M_r^2} \frac{\overline{\rho}}{\overline{\rho/p}} + \overline{u} f_1, \\
f_3 &= n_y \frac{1}{M_r^2} \frac{\overline{\rho}}{\overline{\rho/p}} + \overline{v} f_1, \\
f_4 &= n_z \frac{1}{M_r^2} \frac{\overline{\rho}}{\overline{\rho/p}} + \overline{w} f_1, \\
f_5 &= \left(\frac{1}{(\gamma - 1)(\rho/p)^{ln}} - M_r^2 \overline{k} \right) f_1 + M_r^2 \overline{u} f_2 + M_r^2 \overline{v} f_3 + M_r^2 \overline{w} f_4.
\end{aligned}$$

We found the same scaling with the non-dimensional form of Roe's EC flux [52, 53]. Looking at the entropy conservation condition in non-dimensional form:

$$[\mathbf{v}]^T \mathbf{f}^{\mathbf{EC}} = [\mathcal{F}]$$

$$\Leftrightarrow \left[\frac{\gamma - s}{\gamma - 1} - M_r^2 \frac{\rho}{p} k \right] f_1 \ + \ M_r^2 \left(\left[\frac{\rho u}{p} \right] f_2 \ + \ \left[\frac{\rho v}{p} \right] f_3 \ + \ \left[\frac{\rho w}{p} \right] f_4 \right) \ - \ \left[\frac{\rho}{p} \right] f_5 = [\rho u_n],$$

it is not clear whether an entropy-conservative flux would always have the same Mach number scaling as $\mathbf{f}$. Perhaps one could argue that those constructed in the same way as Roe's and Chandrasekhar's will always meet this condition.

6.3.3 Entropy-Stable Upwind Dissipation

As explained in chapter 2, the classic upwind dissipation term $R|\Lambda|R^{-1}[\mathbf{u}]$ and its entropy-stable variant $R|\Lambda|R^T[\mathbf{v}]$ are not different in essence. For infinitesimal variations, one has:

$$R|\Lambda|R^{-1}d\mathbf{u} = |A|d\mathbf{u} = |A|(Hd\mathbf{v}) = (R|\Lambda|R^{-1})(RR^Td\mathbf{v}) = R|\Lambda|R^Td\mathbf{v}$$

From this relation, it is fair to assume that $|A|[\mathbf{u}]$ (conventional Roe-type dissipation) and $|A|H[\mathbf{v}]$ (conventional ES dissipation) have the same scaling hence the same accuracy issues in the low Mach limit. Additionally, from the understanding that the dissipation matrix in (6.32) also writes $D = |A|H$, it is fair to define the preconditioned matrix as $D_P = P^{-1}|PA|H$. The modified entropy-stable flux therefore writes:

$$\mathbf{f}^*(\mathbf{u}_L, \mathbf{u}_R, \mathbf{n}) = \mathbf{f}^{\mathbf{EC}}(\mathbf{u}_L, \mathbf{u}_R, \mathbf{n}) - \frac{1}{2}P^{-1}|PA|H(\mathbf{v}_R - \mathbf{v}_L). \tag{6.35}$$

The main question that arises from here is whether this correction is compatible with the requirement of entropy stability. In other words, can one find P invertible such that $D_P = P^{-1}|PA|H$ is positive definite and $P^{-1}|PA|$ has the same Mach number scalings as A?

We first seek conditions on P for D_P to be positive definite. If $P = I$, positive definiteness follows from the eigenscaling theorem because H is symmetric positive definite and symmetrizes A from the right. Writing $H = RR^T$ as before is not helpful unless the eigenvectors of $|PA|$ are related to the eigenvalues of $|A|$ in a convenient way. If H symmetrizes PA from the right, then $|PA|H$ is symmetric positive definite but it is not clear if this matrix would remain positive definite upon multiplication on the left by P^{-1}. In addition, the condition that H symmetrizes PA might be too stringent to work with. Since H symmetrizes A, HP^T symmetrizes PA and we can rewrite D_P as:

$$D_P = P^{-1}|PA|H = P^{-1}|PA|HP^T P^{-T} = P^{-1}(|PA|HP^T)(P^{-1})^T \tag{6.36}$$

Equation (6.36) shows by congruence that D_P is positive definite if and only if $|PA|HP^T$ is positive definite. For the eigenscaling theorem to apply, we need HP^T to be symmetric positive definite. Given that P and H are full matrices, this also appears as a complicated a condition to work with.

In section 6.2, we recalled that preconditioners are typically developed for a mapped system first. Let $A_{\mathbf{z}} = Q^{-1}AQ$ and $P_{\mathbf{z}}$ be the associated preconditioner, then $P^{-1}|PA| = QP_{\mathbf{z}}^{-1}|P_{\mathbf{z}}A_{\mathbf{z}}|Q^{-1}$. From there, we note that since H symmetrizes A from the right, then $H_{\mathbf{z}} = Q^{-1}HQ^{-T}$ symmetrizes $A_{\mathbf{z}}$ from the right as well. We can then further decompose D_P as:

$$D_P = P^{-1}|PA|H = QP_{\mathbf{z}}^{-1}|P_{\mathbf{z}}A_{\mathbf{z}}|H_{\mathbf{z}}Q^T = (QP_{\mathbf{z}}^{-1})|P_{\mathbf{z}}A_{\mathbf{z}}|H_{\mathbf{z}}P_{\mathbf{z}}^T(QP_{\mathbf{z}}^{-1})^T \tag{6.37}$$

Equation (6.37) shows, again by congruence, that D_P is positive definite if and only if $|P_{\mathbf{z}}A_{\mathbf{z}}|H_{\mathbf{z}}P_{\mathbf{z}}^T$ is positive definite. Since $H_{\mathbf{z}}P_{\mathbf{z}}^T$ symmetrizes $P_{\mathbf{z}}A_{\mathbf{z}}$ from the right, then the eigenscaling theorem applies if $H_{\mathbf{z}}P_{\mathbf{z}}^T$ is symmetric positive definite. With the differential entropy variables $d\mathbf{z} = (dp/(\rho a M_r),\ du,\ dv,\ dw,\ dp - a^2 d\rho)$, the Jacobian $A_{\mathbf{z}}$ given by equation (6.26) is symmetric and from its structure, we can assume that $P_{\mathbf{z}}$ will have the general form:

$$P_{\mathbf{z}} = \begin{bmatrix} P_{5\times5} & O_{5\times1} \\ O_{1\times5} & 1 \end{bmatrix}.$$

Interestingly, the matrix:

$$H_{\mathbf{z}} = \frac{(a/M_r)^2}{\gamma r} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & (aM_r)^2(\gamma - 1) \end{bmatrix},$$

happens to commute with $P_{\mathbf{z}}$, $A_{\mathbf{z}}$ and with $|P_{\mathbf{z}}A_{\mathbf{z}}|$, in addition to being symmetric positive definite. This allows us to ultimately rewrite D_P as:

$$D_P = (QP_{\mathbf{z}}^{-1})|P_{\mathbf{z}}A_{\mathbf{z}}|H_{\mathbf{z}}P_{\mathbf{z}}^T(QP_{\mathbf{z}}^{-1})^T = (QH_{\mathbf{z}}^{1/2}P_{\mathbf{z}}^{-1})|P_{\mathbf{z}}A_{\mathbf{z}}|P_{\mathbf{z}}^T(QH_{\mathbf{z}}^{1/2}P_{\mathbf{z}}^{-1})^T, \quad (6.38)$$

and demonstrate that a *sufficient* condition for D_P to be positive definite is that $P_{\mathbf{z}}$ (equivalently $P_{5\times5}$) is symmetric positive definite. The Turkel preconditioner (6.29) qualifies. The Miczek preconditioner does not meet this condition.

We now have to integrate in this discussion the requirement that $P_{\mathbf{z}}^{-1}|P_{\mathbf{z}}A_{\mathbf{z}}|$ has the same Mach number scaling as $A_{\mathbf{z}}$. Surprisingly, we have not managed to find a symmetric positive definite matrix $P_{\mathbf{z}}$ which satisfies the scaling requirements. To simplify the analysis, let's consider the scenario where the flow and the interface normal are aligned with the x-direction. This brings us back to the 2-by-2 system (6.24). Can we find $p(M_r)$, $p_1(M_r)$ and $p_2(M_r)$ such that:

$$P_{\mathbf{z}}^{-1}|P_{\mathbf{z}}A_{\mathbf{z}}| = \begin{bmatrix} \mathcal{O}(1) & \mathcal{O}(1/M_r) \\ \mathcal{O}(1/M_r) & \mathcal{O}(1) \end{bmatrix} \quad \text{and} \quad P_{\mathbf{z}} = \begin{bmatrix} p_1 & p \\ p & p_2 \end{bmatrix} \quad \text{positive definite?}$$

We have $P_{\mathbf{z}}A_{\mathbf{z}} = R_p \Lambda_p R_p^{-1}$ with:

$$\Lambda_p = diag([0.5(u_p + a_p),\ 0.5(u_p - a_p)]),\ u_p = (p_1 + p_2)u + 2ap/M_r,$$

$$a_p = \sqrt{u_p^2 + 4det(P)(a^2/M_r^2 - u^2)},$$

$$R_p = \begin{bmatrix} r_1 & r_2 \\ 1 & 1 \end{bmatrix},$$

$$r_1 = (M_r a_p - ap_2(p_1 - p_2)/p)/(2ap_2 + 2M_r pu) + (p_1 - p_2)/(2p),$$

$$r_2 = (-M_r a_p - ap_2(p_1 - p_2)/p)/(2ap_2 + 2M_r pu) + (p_1 - p_2)/(2p).$$

$det(P) = p_1 p_2 - p^2 > 0$ and $trace(P) = p_1 + p_2 > 0$ impose p_1 and p_2 to be positive. $det(P) > 0$ and $a^2/M_r^2 - u^2 > 0$ in the subsonic regime, therefore $u_p < a_p \implies |u_p - a_p| = -(u_p - a_p)$ and we have:

$$P_{\mathbf{z}}^{-1}|P_{\mathbf{z}}A_{\mathbf{z}}| = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$$

$$a_{11} = \frac{1}{a_p}\big(u^2(p_1 - p_2) + 2a^2 p_2/M_r^2 + 2apu/M_r,\ a_{12} = \frac{1}{a_p}2pu^2 + a(p_1 + p_2)u/M_r\big),$$

$$a_{21} = a_{12},\ a_{22} = \frac{1}{a_p}\big(-u^2(p_1 - p_2) + 2a^2 p_1/M_r^2 + 2apu/M_r\big).$$

Looking at the expression of a_p, we see that in the limit $M_r \to 0$, a_p can scale either as p/M_r, $\sqrt{p_1 p_2}/M_r$, p_1 or p_2.

- If $a_p \approx p/M_r$: $a_{11} \approx u^2(p_1 - p_2)M_r/p + 2a^2 p_2/(pM_r) + 2au = \mathcal{O}(1)$ requires p_2/p to scale as M_r at most. Likewise, p_1/p must scale as M_r at most for a_{22} to be $\mathcal{O}(1)$. But then $a_{12} \approx 2M_r u^2 + a(p_1/p + p_2/p)u = \mathcal{O}(M_r)$ does not scale as $1/M_r$.

- If $a_p \approx p_1$: the second term in a_{22} scales as $1/M_r^2$.

- If $a_p \approx p_2$: the second term in a_{11} scales as $1/M_r^2$.

- If $a_p \approx \sqrt{p_1 p_2}/M_r$: Denote $X = \sqrt{p_1/p_2}$. Then $a_{11} \approx u^2(X - 1/X)M_r + 2a^2/(XM_r) + 2apu/(\sqrt{p_1 p_2}) = \mathcal{O}(1)$ imposes that X scales as $1/M_r$. But then the second term in a_{22} scales at $1/M_r^2$ instead of 1.

In each case, it seems that the Mach number scaling requirements cannot be met.

It is important to recall that we posed $P_{\mathbf{z}}$ symmetric positive definite as a sufficient condition only. Miczek's flux-preconditioner can be found by seeking $P_{\mathbf{z}}$ in the form:

$$P_{\mathbf{z}} = \begin{bmatrix} 1 & p \\ -p & 1 \end{bmatrix}.$$

$P_{\mathbf{z}}$ is not symmetric but it is positive definite for any p since its symmetric part is the identity matrix. We have $P_{\mathbf{z}} A_{\mathbf{z}} = R_p \Lambda_p R_p^{-1}$ with:

$$\Lambda_p = diag([u + a_p,\ u - a_p]),\ a_p = \sqrt{u^2 + det(P)(a^2/M_r^2 - u^2)} > u_p,$$
$$R_p = \begin{bmatrix} (-M_r a_p + ap)/(a - M_r pu) & (M_r a_p + ap)/(a - M_r pu) \\ 1 & 1 \end{bmatrix}.$$

In the subsonic regime, this gives:

$$P_{\mathbf{z}}^{-1}|P_{\mathbf{z}} A_{\mathbf{z}}| = \frac{1}{a_p} \begin{bmatrix} a^2/M_r^2 & pu^2 + a(M_r u - ap)/M_r^2 \\ -pu^2 + a(M_r u - ap)/M_r^2 & a^2/M_r^2 \end{bmatrix}.$$

For the first term to be $\mathcal{O}(1)$ we need $a_p = \mathcal{O}(1/M_r^2)$ which imposes $p = \mathcal{O}(1/M_r)$. The scaling of $A_{\mathbf{z}}$ is completely recovered. Miczek takes $p = 1 - 1/M_r$ so that in the limit $M_r \to 1$, $P_{\mathbf{z}} \to I$. Furthermore, this matrix does conserve entropy stability. For the 2×2 system, the symmetric part of $|P_{\mathbf{z}} A_{\mathbf{z}}| P_{\mathbf{z}}^T$ has a determinant $a^2(a^2 - M_r^2 a^2) det(P)^2/(M_r^4 a_p^2)$ and a trace $2a^2 det(P)/(M_r^2 a_p)$ that are both positive. For the general system, this can be proved by constructing a scaled form for D_P.

6.4 Numerical experiments

In this section, we examine four different first-order entropy-stable schemes in two simple flow configurations which are representative of the incompressible and acoustic limits. In section 6.4.1, we consider the Gresho vortex (incompressible limit). In section 6.4.2, we consider a right-moving sound wave (acoustic limit). Periodic boundary conditions are used in both problems.

In space, we use the classic ES Roe flux, the ES Turkel flux, the ES Miczek flux, with the EC flux of Chandrasekhar as the base (we observed the same results with the EC flux of Roe). We also consider the EC flux of Chandrasekhar alone. These four fluxes are used in conjunction with Backward Euler in time with a CFL of 1.

The calculations are made using a code which solves the dimensional form of the compressible Euler equations. This does not impact the interpretation of the results as most of the metrics used are non-dimensional.

6.4.1 Gresho vortex

The Gresho vortex [221, 222, 224] is an exact steady-state solution of the incompressible Euler equations in two dimensions. Let R be the radius of the vortex and r be the radial coordinate. Density is constant $\hat{\rho} = \rho_r$. The velocity field is given by:

$$\hat{\mathrm{u}} = u_\phi \boldsymbol{e}_\phi,\ u_\phi = u_r \begin{cases} r/R, & 0 \le r < R \\ 2 - r/R, & R \le r \le 2R \\ 0, & 2R \le r \end{cases} \tag{6.39}$$

u_ϕ denotes the tangential velocity, $r = \sqrt{x^2 + y^2}$, $\boldsymbol{e}_\phi = -sin(\phi)\boldsymbol{u_x} + cos(\phi)\boldsymbol{u_y} = -y/r\boldsymbol{u_x} + x/r\boldsymbol{u_y}$. The period of the vortex and reference time scale is defined as

$t_r = 2\pi R/u_\phi(R) = 2\pi R/u_r$. The pressure $\hat{p}$ must provide the centripetal force:

$$\begin{aligned}\hat{p} &= p_r + \int_0^r \rho_r \frac{u_\phi^2(\overline{r})}{\overline{r}}\, d\overline{r} \\ &= p_r + \rho_r u_r^2 \begin{cases} (r/R)^2/2, & 0 \leq r < R \\ (r/R)^2/2 + 4(1-(r/R)+\ln(r/R)), & R \leq r \leq 2R \\ -2 + 4\ln 2, & 2R \leq r \end{cases}\end{aligned}$$

p_r is a strictly positive constant. The reference Mach number M_r for this setup is defined as the one at $r = R$:

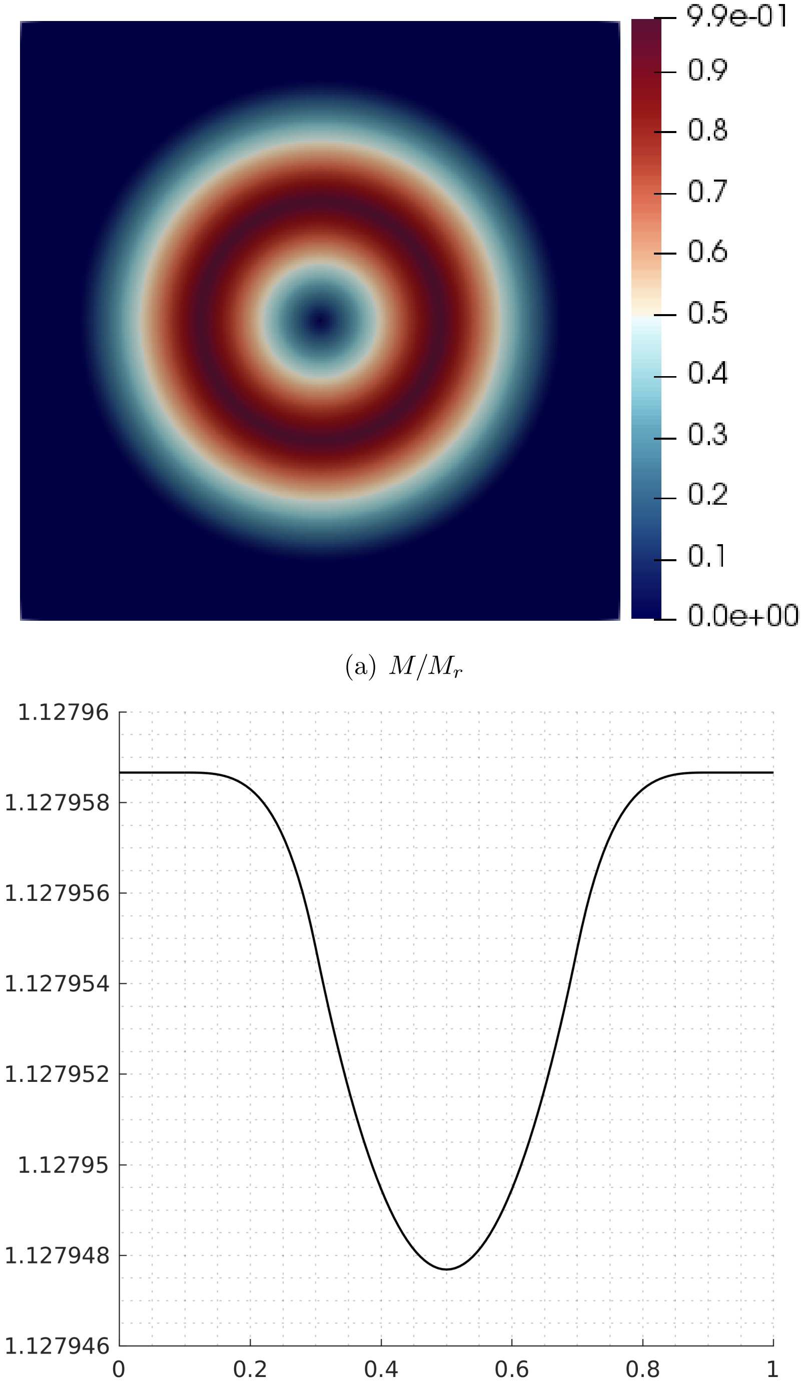
$$M_r = \frac{u_r}{\sqrt{\gamma(p_r/\rho_r + u_r^2/2)}}.$$

This equation can be rewritten as:

$$p_r = \rho_r u_r^2 \left(\frac{1}{\gamma M_r^2} - \frac{1}{2} \right), \tag{6.40}$$

and shows how the free constants $(\rho_r,\ u_r,\ p_r)$ relate to the reference Mach number M_r. Following Miczek *et al.* [224], we fix the grid size (150 × 150 cells) and run the schemes at different reference Mach numbers $M_r \in \{3 \times 10^{-1},\ 3 \times 10^{-2},\ 3 \times 10^{-3}\}$ until the vortex completes one revolution, that is until $t = 1$. We take $R = 0.2$, $\rho_r = 1.0$ and $u_r = 2\pi R M_r \implies t_r = 1/M_r$. The domain is $(x, y) \in [0\ 1]^2$. The pressure is determined by equation (6.40). Figure 6.1 shows the initial conditions for $M_r = 3 \times 10^{-3}$.

The central flux alone is not a viable option in the low Mach regime because it is not stable, even when combined with Backward Euler in time [223, 224]. An illustration is provided in figure 6.2-(b), which shows losses in total entropy, and in figure 6.2-(a), which shows that the central flux leads to total kinetic energy production. With the EC flux, the total entropy increases and the total kinetic energy fluctuates



(a) M/M_r

(b) $\hat{p}(x, 0.5)$

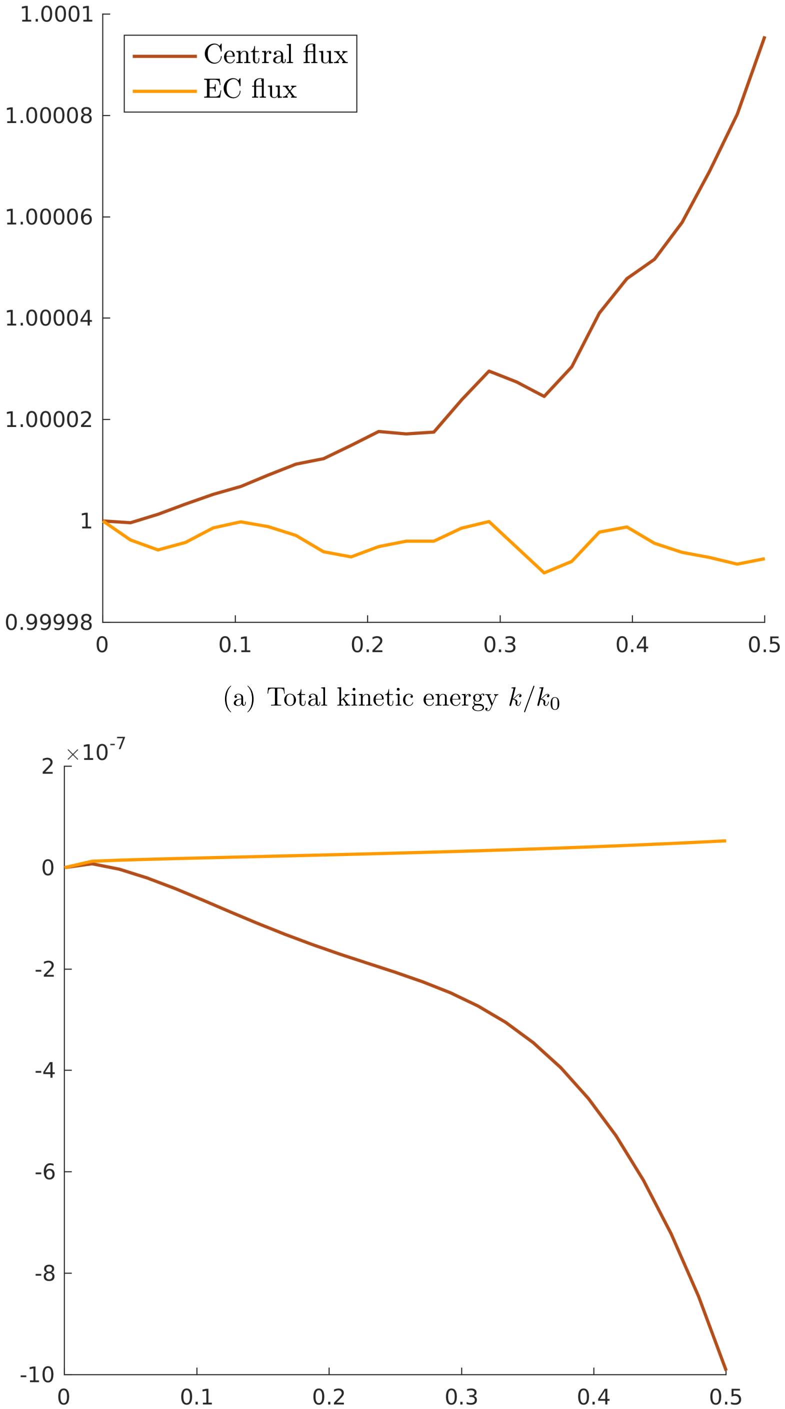
Figure 6.1: Gresho Vortex: Initial solution for $M_r = 3 \times 10^{-3}$.

but does not become bigger than what it was initially.

Figure 6.3 shows snapshots of the solution with each scheme at different Mach numbers. Figures 6.3-(j), 6.3-(k) and 6.3-(l) provide a clear illustration of the accuracy degradation issues in the low Mach regime. The other three schemes do not show a visible dependency on the reference Mach number. The best results are obtained with the EC flux, which is hardly surprising considering that the flow configuration is smooth. The difference between the ES Turkel and ES Miczek fluxes is not clearly visible from these plots.

Figure 6.4 shows the normalized kinetic energy evolution for all four schemes at different Mach numbers. For the ES Roe flux, we see that the rate at which the kinetic energy decays increases with the Mach number. We can also see that the normalized kinetic energy for $M_r = 3 \times 10^{-2}$ becomes bigger than for $M_r = 3 \times 10^{-3}$. This was slightly visible in the previous figure already. Given that the Gresho vortex is a *stationary* solution, it is not surprising that numerical solution would eventually reach a steady state. For the ES Miczek and ES Turkel fluxes, the kinetic energy decay appears to be independent of the Mach number. In each case, we see that the $M_r = 3 \times 10^{-1}$ curve is not matching exactly with the $M_r = \{3 \times 10^{-2},\ 3 \times 10^{-3}\}$ ones. This figure suggests that the ES Miczek flux performs better than the ES Turkel flux. This is also supported by figures 6.5(a)-(c) which shows that the ES Turkel flux produces more entropy than the ES Miczek flux.

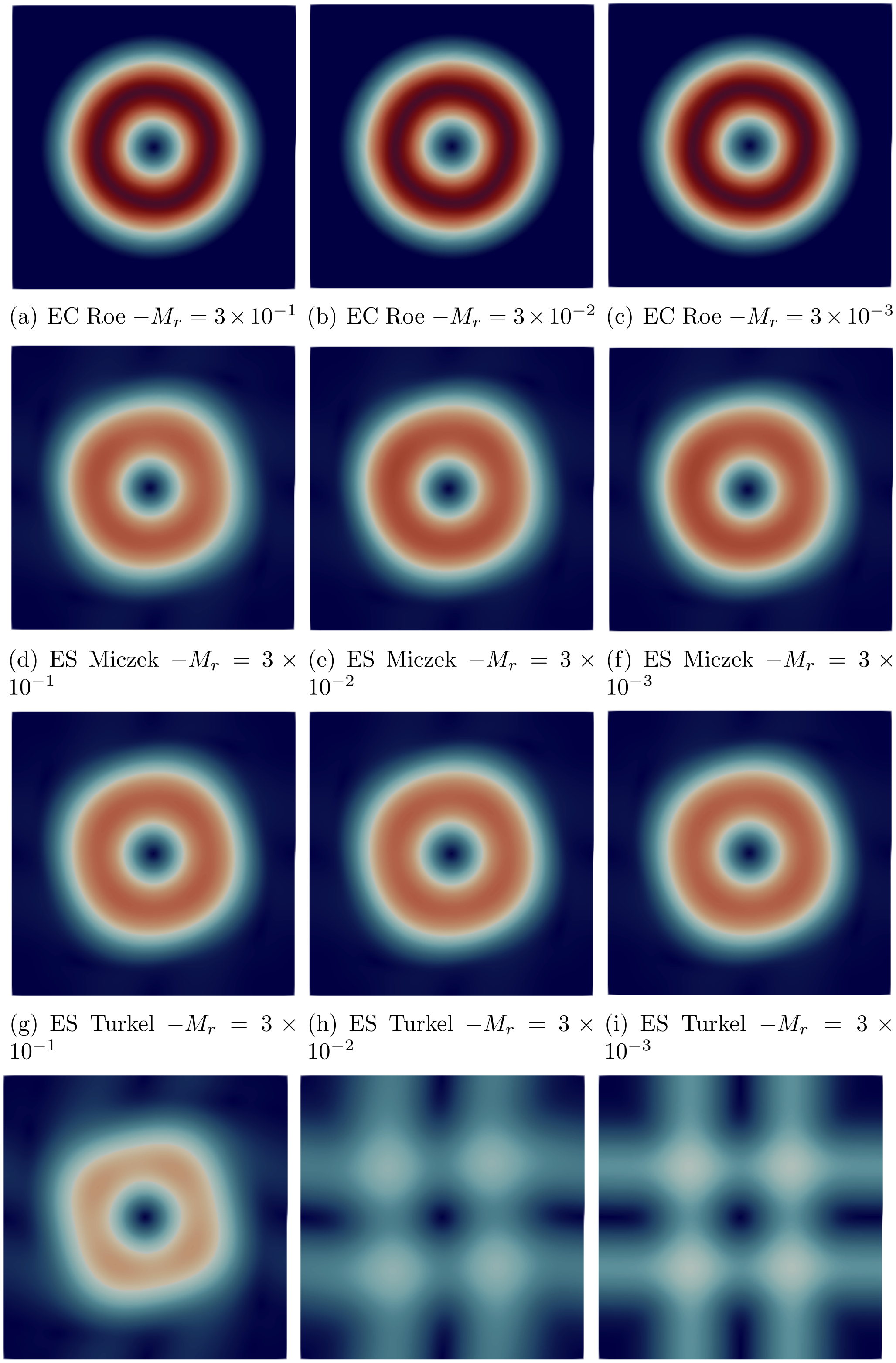
Figure 6.6 shows the pressure distribution along the centerline $y = 0.5$, after one revolution at $M_r = 3 \times 10^{-3}$. The solution with the ES Miczek flux is clearly not in phase with the exact solution. The same anomaly is observed at different Mach numbers. Figures 6.7(a)-(b) suggest that this anomaly is the consequence of a spurious transient in the early stages of the vortex rotation. We found that the duration of this transient decreases with the Mach number.



(a) Total kinetic energy k/k_0

(b) Total Entropy $(\rho s) - (\rho s)_0$

Figure 6.2: Kinetic energy and entropy evolution with the central and EC fluxes in space and Backward Euler in time for half a rotation of the Gresho vortex at $M_r = 3 \times 10^{-1}$. k_0 and $(\rho s)_0$ are the kinetic energy and entropy at $t = 0$.



(a) EC Roe $-M_r = 3 \times 10^{-1}$ (b) EC Roe $-M_r = 3 \times 10^{-2}$ (c) EC Roe $-M_r = 3 \times 10^{-3}$

(d) ES Miczek $-M_r = 3 \times 10^{-1}$ (e) ES Miczek $-M_r = 3 \times 10^{-2}$ (f) ES Miczek $-M_r = 3 \times 10^{-3}$

(g) ES Turkel $-M_r = 3 \times 10^{-1}$ (h) ES Turkel $-M_r = 3 \times 10^{-2}$ (i) ES Turkel $-M_r = 3 \times 10^{-3}$

(j) ES Roe $-M_r = 3 \times 10^{-1}$ (k) ES Roe $-M_r = 3 \times 10^{-2}$ (l) ES Roe $-M_r = 3 \times 10^{-3}$

Figure 6.3: Gresho Vortex: M/M_r profiles at $t = 1$. Same legend as figure 6.1-(a).

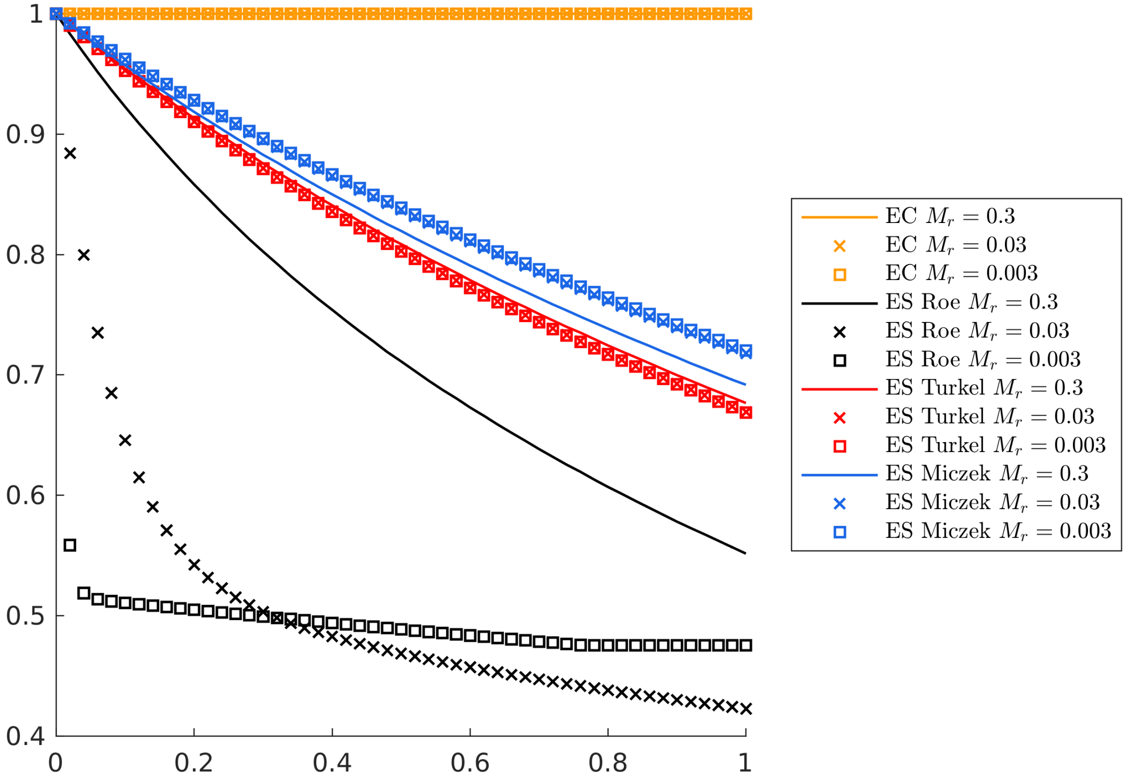
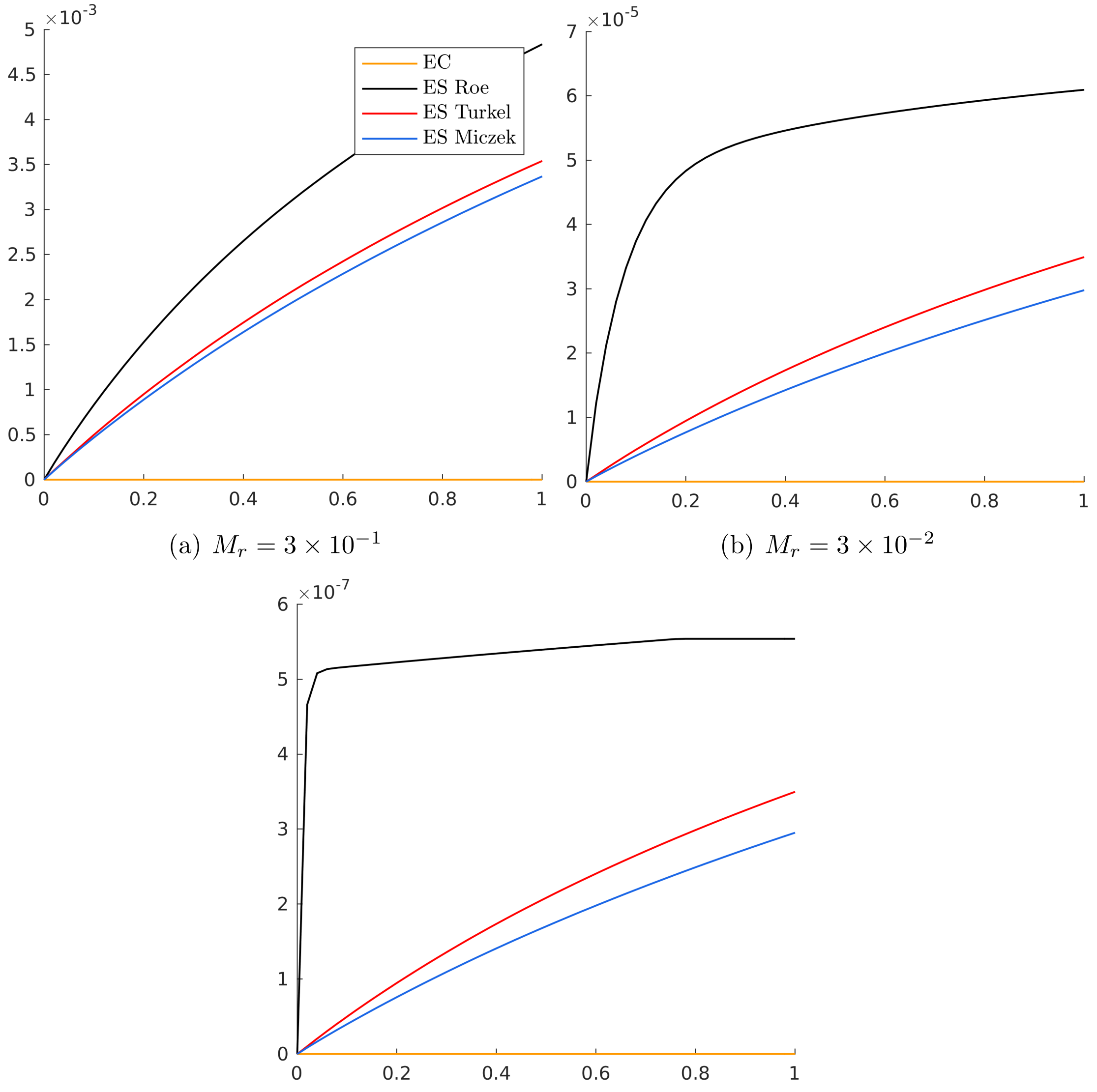


Figure 6.4: Gresho vortex: Total kinetic energy k/k_0 evolution over time for different ES fluxes at different Mach numbers.



(a) $M_r = 3 \times 10^{-1}$

(b) $M_r = 3 \times 10^{-2}$

(c) $M_r = 3 \times 10^{-3}$

Figure 6.5: Gresho Vortex: Total entropy $(\rho s) - (\rho s)_0$ evolution for different fluxes at different Mach numbers.

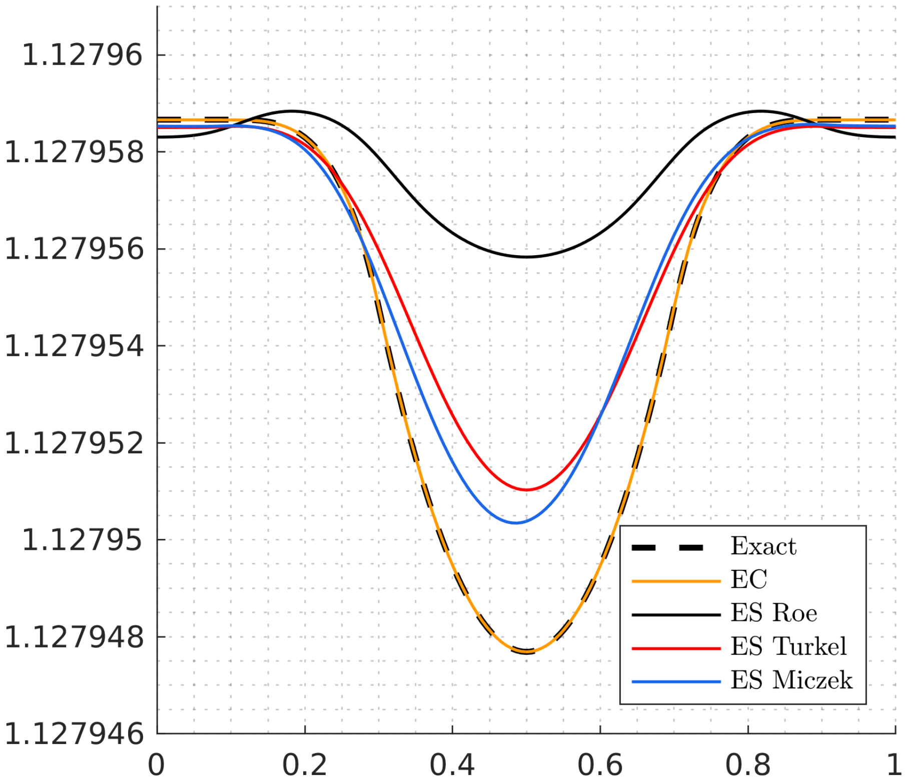
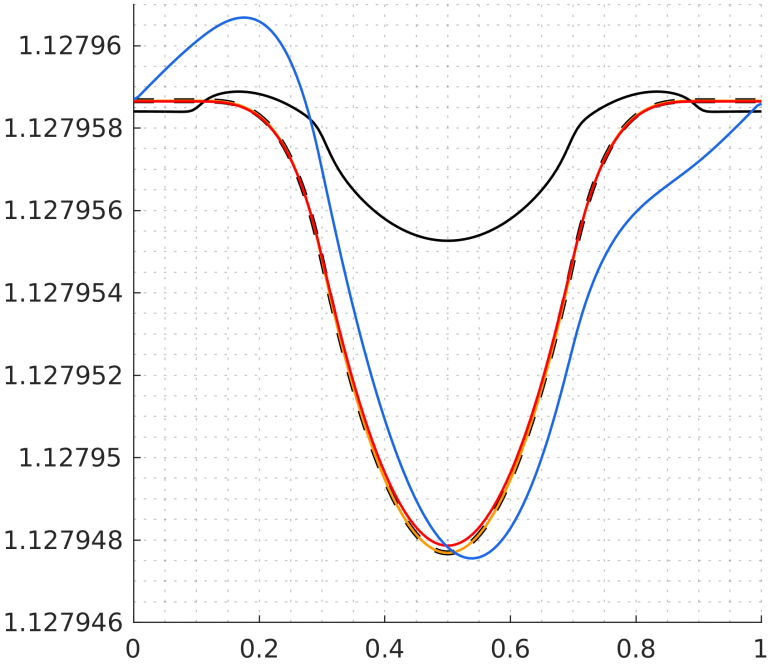
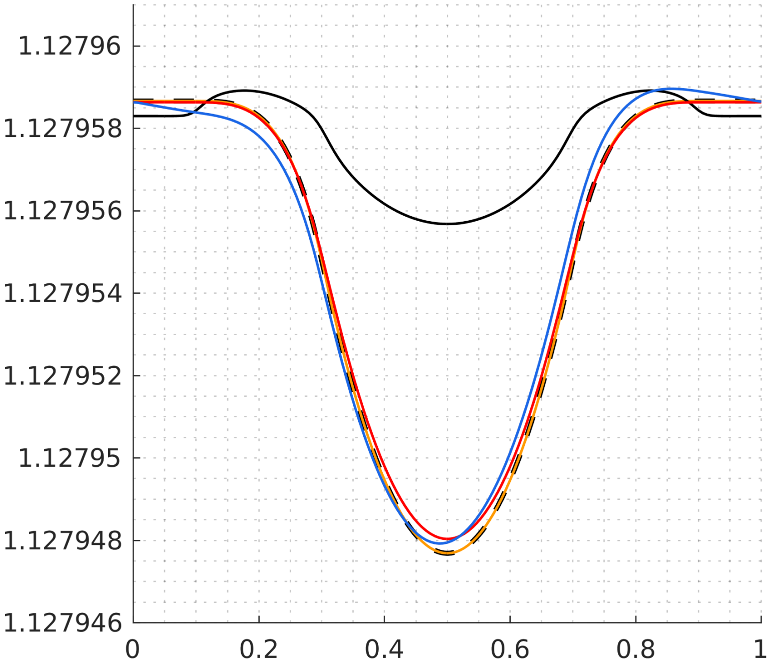


Figure 6.6: Gresho vortex: Centerline pressure profile $\hat{p}(x, 0.5)$ after one rotation at $M_r = 3 \times 10^{-3}$.



(a) $t = 0.04$



(b) $t = 0.08$

Figure 6.7: Gresho Vortex: Centerline pressure profiles at early instants highlighting the spurious transient observed with the ES Miczek flux. $M_r = 3 \times 10^{-3}$. Same legend as figure 6.6.

6.4.2 Acoustic wave

One way to set up a right-moving acoustic wave is to consider, as in Bruel *et al.* [240], a free stream (ρ_∞, u_∞) and add density and velocity fluctuations such that the Riemann invariants associated with the left moving acoustic wave and the entropy wave are unchanged. In a domain $\Omega = [-0.5,\ 0.5]$, we perturb the density as follows:

$$\hat{\rho}(x,0) = \rho_\infty\big(1 + M_r\psi(x)\big),$$

where $\psi(x) = \exp(-\alpha x^2)$ defines a gaussian pulse centered at the center of the domain. We set $\alpha = \ln(10^3)/0.15^2$ so that $\psi(x) < 10^{-3}$ for $|x| < 0.15$. For isentropic flow we take $\hat{p} = (\hat{\rho})^\gamma \implies \hat{a} = \sqrt{\gamma}(\hat{\rho})^{\frac{\gamma-1}{2}}$. The corresponding velocity perturbation must satisfy

$$\big(\hat{u}(x,0) - u_\infty\big) - \frac{2\big(\hat{a}(x,0) - a_\infty\big)}{\gamma - 1} = 0 \implies \hat{u}(x,0) = u_\infty + \frac{2\big(\hat{a}(x,0) - a_\infty\big)}{\gamma - 1}.$$

$\hat{a}(x,0)$ and $\hat{p}(x,0)$ are imposed by the density. If the reference Mach number M_r is small enough, we can write:

$$\hat{u}(x,0) = u_\infty + \frac{2a_\infty}{\gamma - 1}\left(\big(1 + M_r\psi\big)^{\frac{\gamma-1}{2}} - 1\right) = u_\infty + a_\infty M_r\psi + \mathcal{O}(M_r^2). \qquad (6.41)$$

We set $u_\infty = 0$ and $a_\infty = 1$, so that the speed of propagation of the acoustic wave is one. We take the reference time scale $t_r = 1$, that is the time it takes for the acoustic wave to do one period. We have $\rho_r = \rho_\infty$ and $u_r = M_r$. Changing the reference Mach number changes the amplitude of the velocity, density and pressure perturbations.

We tested the four schemes on a grid of 500 cells for $M_r \in \{10^{-2},\ 10^{-3},\ 10^{-4}\}$. Figure 6.8 shows the initial pressure profile for the $M_r = 10^{-2}$ case. Figures 6.9(a)-(c) show the numerical solution at $t = 1$ for different Mach numbers. The reference

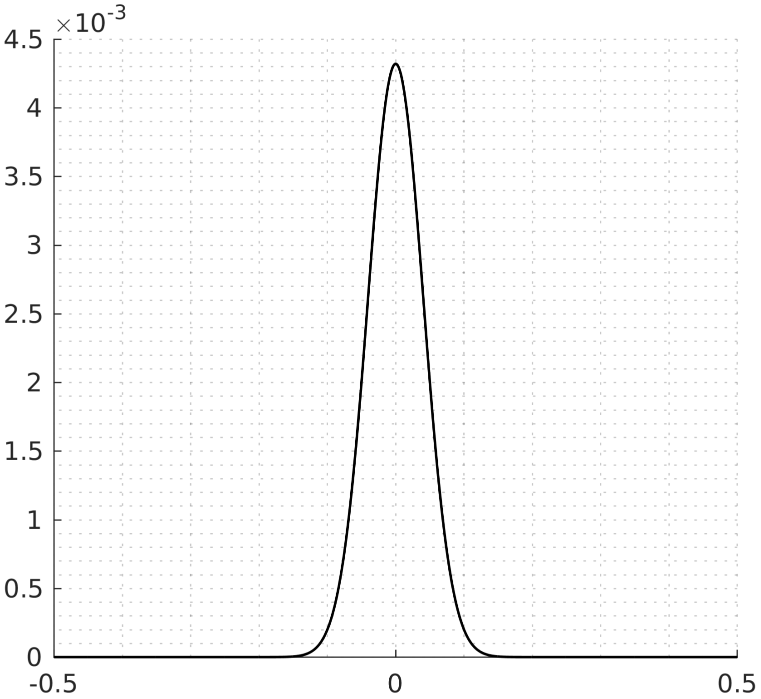


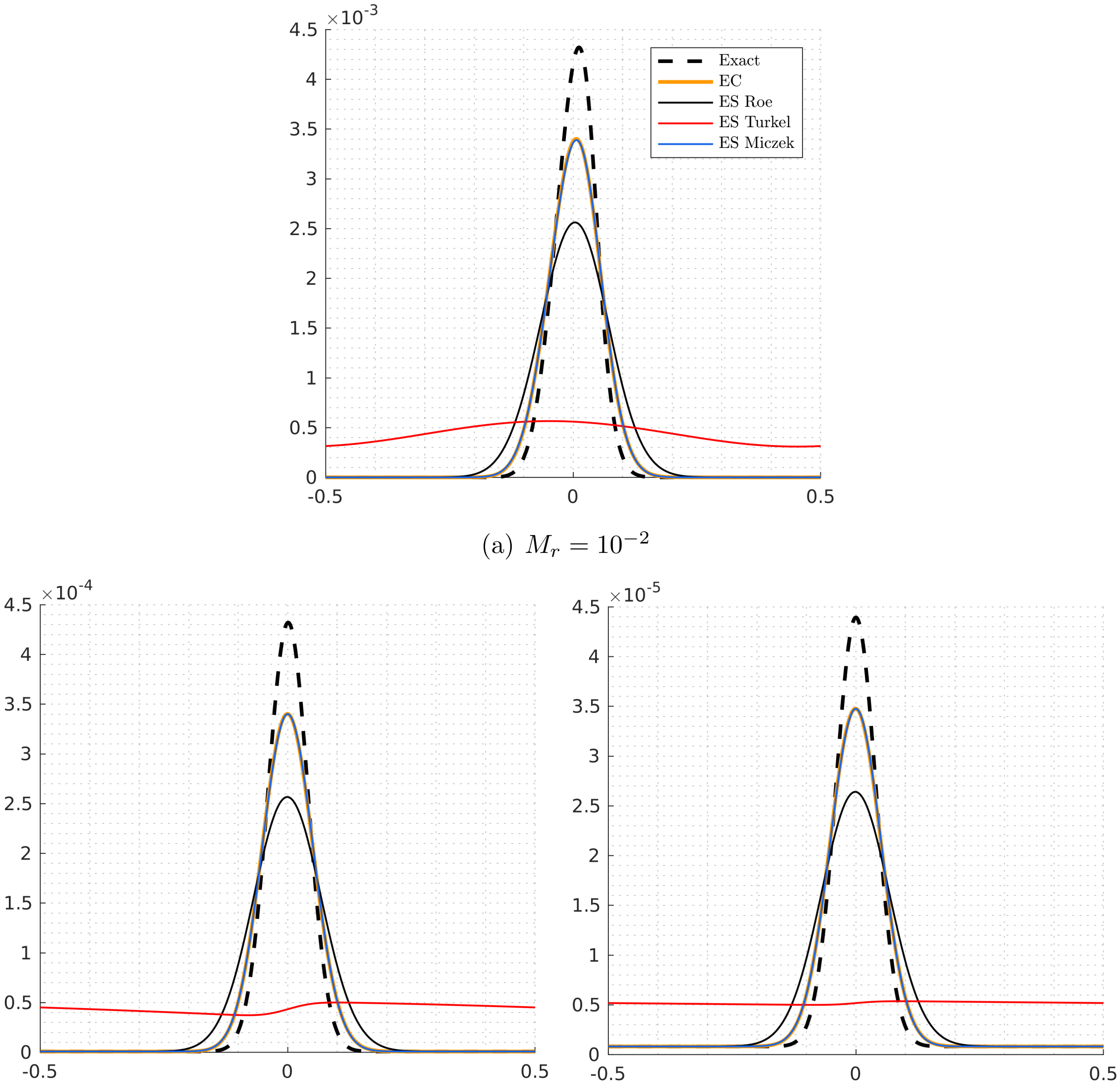
Figure 6.8: Sound wave: Initial pressure distribution for $M_r = 10^{-2}$.

solution[3] is obtained using a 4-th order TecNO scheme on a grid of 1000 cells with a 4-th order Runge-Kutta time integration and a CFL of 0.5. We see that the ES Roe flux, ES Miczek flux and EC flux lead to a self-similar numerical solution. We can see that the acoustic wave is almost completely gone with the ES Turkel flux. This is in agreement with the analysis and results of Bruel *et al.* [240] for the barotropic Euler equations. The ES Miczek flux does not have this problem. Furthermore, it seems to perform just as well as the EC flux. This is also supported by figures (6.10)(a)-(c) where the total entropy evolution in time is shown.

Figure 6.11 shows the temporal evolution of a normalized sound wave amplitude which we define as:

$$A(t) = \frac{\max_x \hat{p}(x,t)}{\max_x \hat{p}(x,0)}. \tag{6.42}$$

[3] An exact solution can be calculated using the method of characteristics, which requires a nonlinear solver [240]. The solution for this problem is simple enough for a fine numerical solution to be trusted.

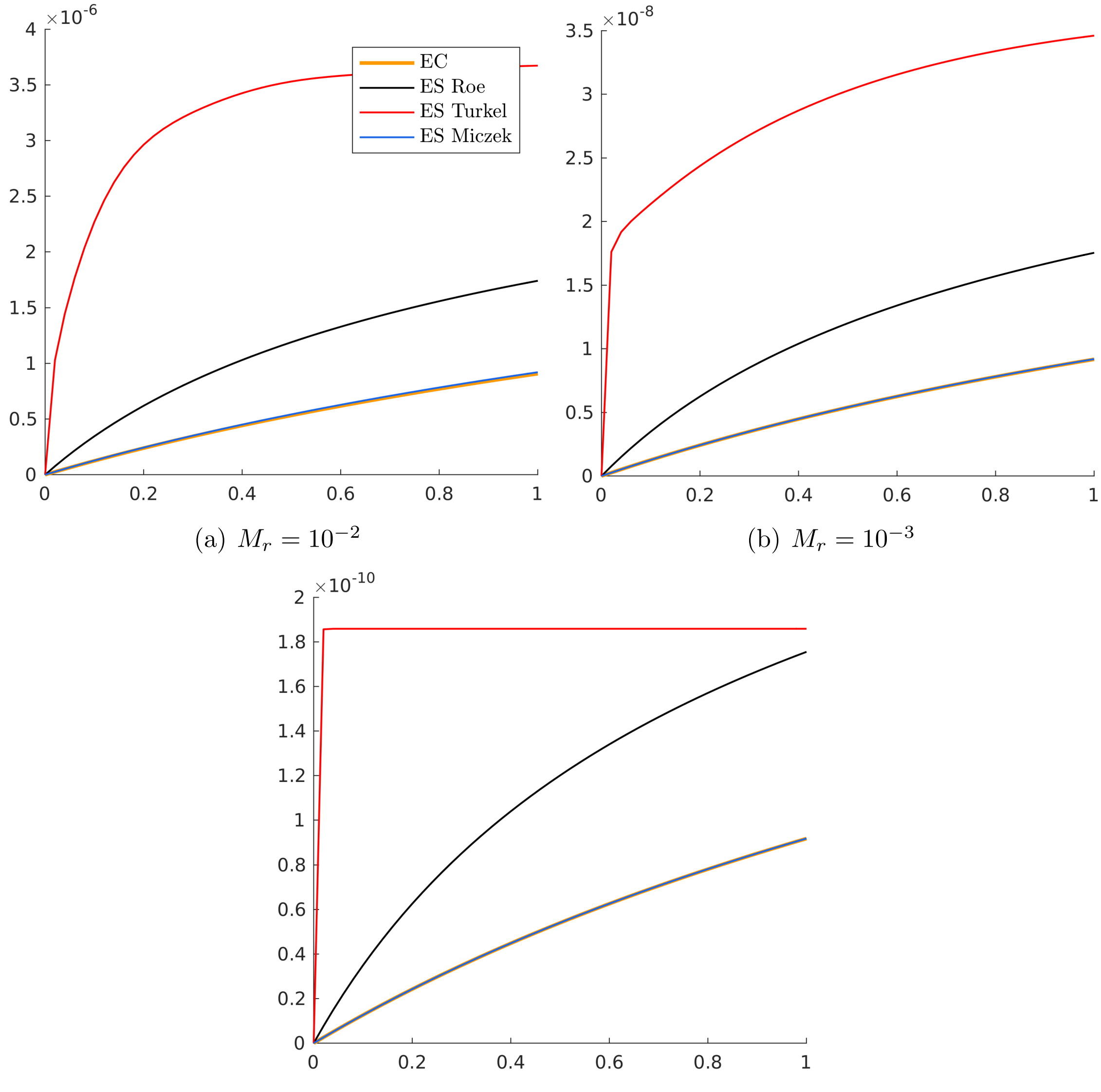


(a) $M_r = 10^{-2}$

(b) $M_r = 10^{-3}$

(c) $M_r = 10^{-4}$

Figure 6.9: Sound wave: Pressure profiles at $t = 1$ for different Mach numbers.



(a) $M_r = 10^{-2}$

(b) $M_r = 10^{-3}$

(c) $M_r = 10^{-4}$

Figure 6.10: Sound wave: Total entropy $(\rho s) - (\rho s)_0$ over time for different Mach numbers.

We can see that the rate at which the ES Turkel flux damps the sound wave increases as the Mach number decreases, while all other fluxes show a self-similar behavior. For the ES Miczek flux, we notice slight perturbations in A which seem to occur around $t = \{0,\ 0.5,\ 1.0\}$. Figures 6.12 and 6.13 suggest that these perturbations are caused by a spurious left-moving acoustic wave created at $t = 0$. It will meet the right-moving acoustic wave when it reaches the periodic boundary ($t \approx 0.5$) and when it reaches the center of the domain ($t \approx 1$).

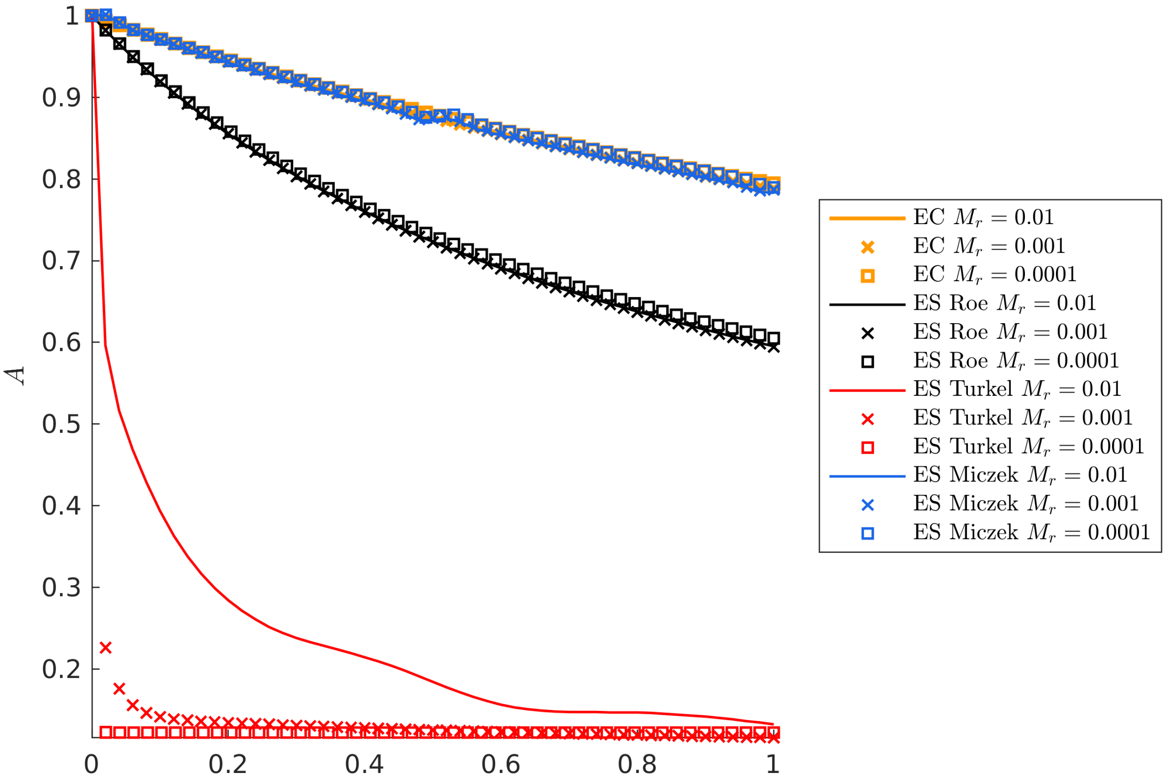
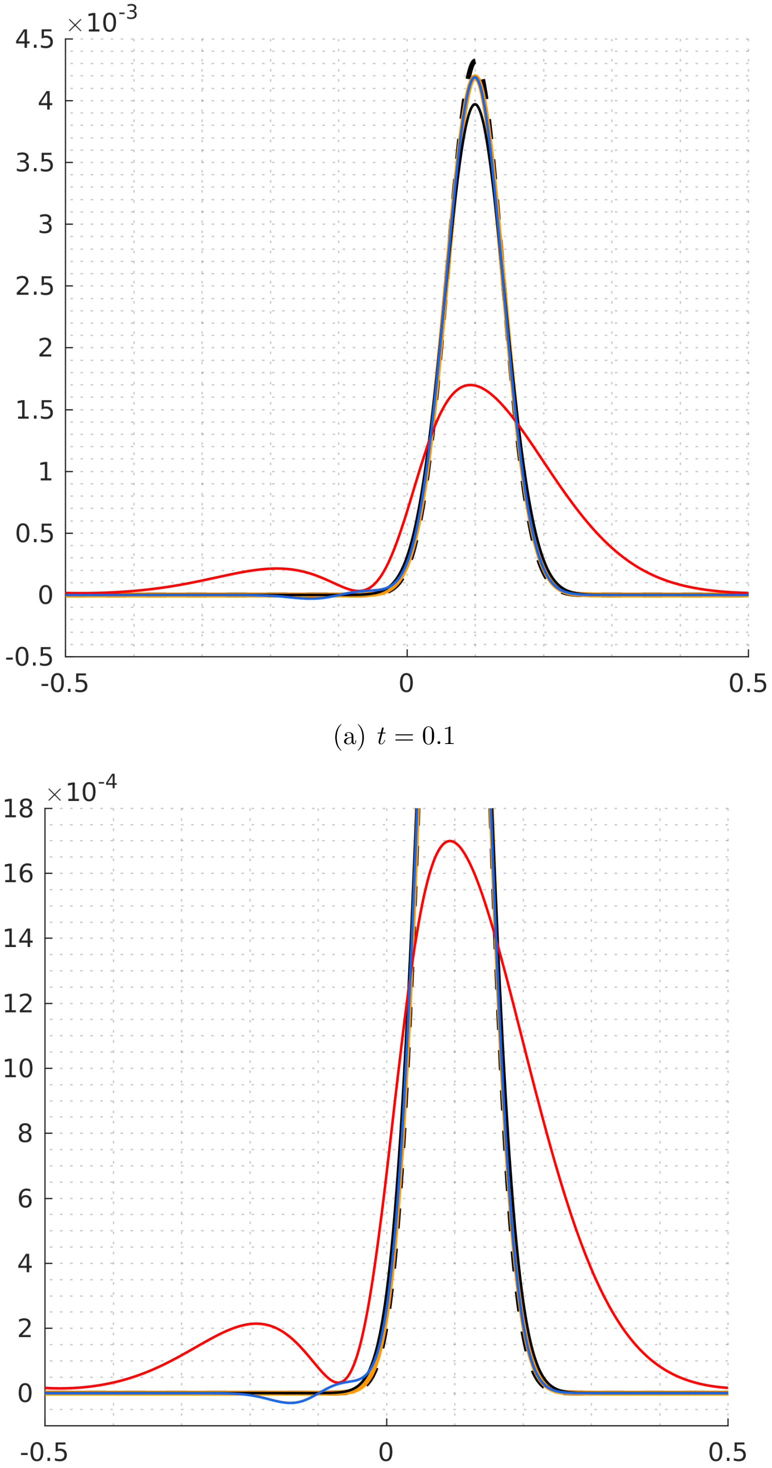


Figure 6.11: Sound wave: Normalized amplitude evolution for all fluxes at different Mach numbers.

6.4.3 Summary

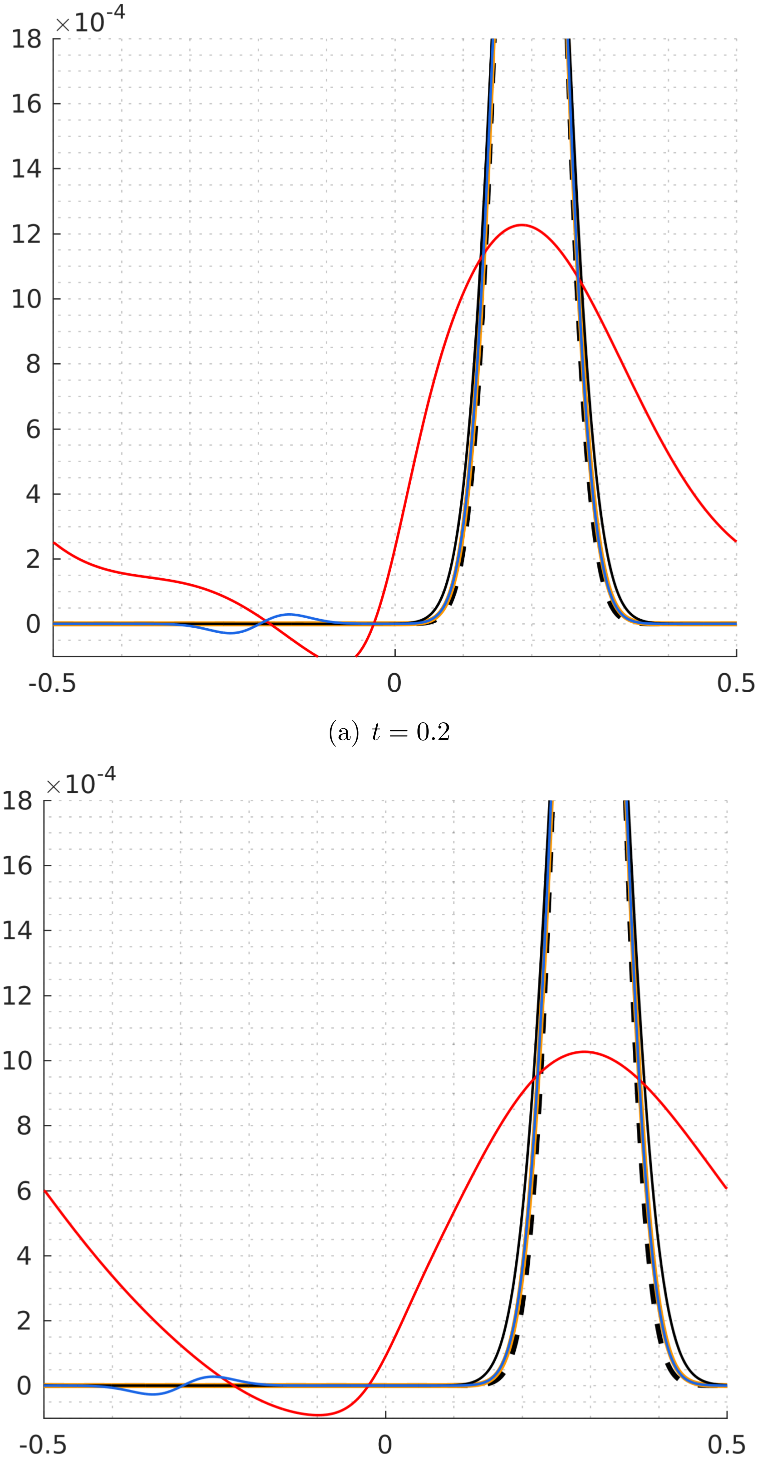
So far we have demonstrated, analytically and numerically, that flux-preconditioning is compatible with entropy-stability to an extent. Numerical tests show that:



(a) $t = 0.1$

(b) $t = 0.1$ (zoom)

Figure 6.12: Sound wave: Pressure profiles showing a spurious wave propagating to the left is created by the Miczek flux. $M_r = 10^{-2}$. Same legend as figure 6.9(a).



(a) $t = 0.2$

(b) $t = 0.3$

Figure 6.13: Sound wave: Pressure profiles showing that the spurious wave reported in figure 6.12 is an acoustic wave propagating at a speed of one. Same legend as figure 6.9(a).

S.1 The ES Roe flux suffers from the same accuracy issues in the incompressible low Mach limit as those previously reported with the classic Roe flux [223, 224, 225]. This does not come as a surprise considering that the dissipation operators are not fundamentally different. Recall that for infinitesimal variations, we have:

$$R|\Lambda|R^{-1}d\mathbf{u} = R|\Lambda|R^{T}d\mathbf{v}.$$

These issues are not observed with the sound wave. The temporal variation of the wave amplitude (6.42) is independent of the Mach number.

S.2 The ES Turkel flux has a more consistent behavior in the incompressible limit, but at the price of damping acoustic waves harder as the Mach number decreases.

S.3 The ES Miczek flux performs well in both limits if we ignore the spurious transient in the Gresho vortex and the spurious left-moving acoustic wave, none of which were reported in [223, 224, 225].

S.4 The EC flux performs the best in both cases. This confirms that for ES schemes, it is the dissipation component that causes the accuracy issues. We note that there is more dissipation in the acoustic case. This is due to the entropy produced by the Backward Euler time scheme.

We could have concluded the effort here with the stance that in the low Mach limit, an entropy-stable scheme should simply revert to an entropy-conservative one. There are two reasons for not doing so:

1. In both problems, there clearly appears to be a correlation between the amount of entropy produced by the scheme and its ability to preserve kinetic energy of the Gresho vortex or the amplitude of the sound wave. Given the framework that we work with, we should be able to explain (**S.1**), (**S.2**) and (**S.3**)

analytically.

2. The errors observed with the ES Miczek flux are intriguing. For the sound wave in particular, the spurious left-moving acoustic wave is reminiscent of what we observed with material interfaces in multicomponent flows. These errors are also similar to the overheating, which can be described as an anomalous entropy wave standing at the center ($\lambda = u = 0$).

6.5 The accuracy degradation from an entropy production perspective

Given that the schemes we are interested in are designed to let us control entropy to an extent, we begin by reinterpreting the theoretical results of section 6.1 on the low Mach limit in terms of entropy. In the incompressible limit, Guillard & Viozat [226] showed that pressure fluctuations in space are of order M^2. Assuming constant density $\rho = \rho_0$, we can write:

$$\begin{aligned}
\rho s &= \rho_0 \Big(\ln(p) - \gamma \ln(\rho_0) \Big) \\
&= \rho_0 \Big(\ln\left(p_0 + M_r^2 p_2 + \mathcal{O}(M_r^3)\right) - \gamma \ln(\rho_0) \Big) \\
&= \rho_0 \Big(\ln(p_0) - \gamma \ln(\rho_0) + \ln\left(1 + M_r^2 (p_2/p_0) + \mathcal{O}(M_r^3)\right) \Big) \\
&= \rho_0 \Big(\ln(p_0) - \gamma \ln(\rho_0) + M_r^2 (p_2/p_0) + \mathcal{O}(M_r^3) \Big) \\
&= \rho_0 s_0 + M_r^2 \rho_0 (p_2/p_0) + \mathcal{O}(M_r^3).
\end{aligned}$$

Therefore, we state:

E.1 In the incompressible limit, entropy ρs fluctuations in space should be of order M_r^2.

Similarly [228]:

E.2 In the acoustic limit, entropy ρs fluctuations in space should be of order M_r.

In the incompressible limit, there is the additional requirement that kinetic energy should be conserved. To precisely and rigorously relate discrete changes in kinetic energy to discrete changes in entropy is not straightforward. Figures 6.2(a)-(b) suggest that if there's any, it is not a straightforward one. Let's assume periodic boundary conditions so that discrete conservation of total energy implies that it remains constant globally. We can write:

$$\Delta(\rho e + M_r^2 \rho k) = 0$$
$$\Leftrightarrow \Delta(\rho k) = -\frac{1}{M_r^2}\Delta(\rho e) = -\frac{1}{(\gamma-1)M_r^2}\Delta(p) = -\frac{1}{(\gamma-1)M_r^2}\Delta\Big(\exp\left(\frac{\rho s - \gamma\rho\ln(\rho)}{\rho}\right)\Big),$$

where Δ refers to the global change, that is the sum $\sum_i \Delta_i$ of local changes in each cell i. Assuming constant density, this relation simplifies to:

$$\Delta(\rho k) = \frac{-1}{\exp(\rho)(\gamma-1)M_r^2}\Delta\big(\exp(\rho s)\big). \tag{6.43}$$

Equation (6.43) relates the global change in kinetic energy ρk to the global change in the *exponential of* the entropy, which entropy-stable schemes do not explicitly control. It is certainly tempting to say that since the exponential function is monotonically increasing, $\Delta(\rho s) > 0 \implies \Delta\big(\exp(\rho s)\big) > 0$. This statement is true locally, but entropy-stable schemes are not explicitly designed to achieve $\Delta_i(\rho s) > 0$, because there are also fluxes (see equation (6.44)). We therefore refrain from making any hasty interpretation of (6.43). If anything, the minus sign in (6.43) supports the general intuition that production of entropy implies losses in kinetic energy.

6.5.1 Breaking down the entropy production of upwind ES fluxes

A remarkable feature of entropy-stable schemes is the relation that holds at the semi-discrete level for entropy in each cell:

$$\frac{dU(\mathbf{u}_i)}{dt} + \frac{1}{V_i}\int_{\delta\Omega_i} F^* dS = -\mathcal{E}_i,\ \mathcal{E}_i = \frac{1}{V_i}\int_{\delta\Omega_i} \mathcal{E} dS. \tag{6.44}$$

where the entropy flux $F^*(\mathbf{u}_i, \mathbf{u}_j, \mathbf{n})$ and the entropy production $\mathcal{E}(\mathbf{u}_i, \mathbf{u}_j, \mathbf{n})$ write:

$$F^*(\mathbf{u}_i, \mathbf{u}_j, \mathbf{n}) = \frac{1}{2}(\mathbf{v}_i + \mathbf{v}_j)^T \mathbf{f}^*(\mathbf{u}_i, \mathbf{u}_j, \mathbf{n}) - \frac{1}{2}(\mathcal{F}_i + \mathcal{F}_j), \tag{6.45}$$

$$\mathcal{E}(\mathbf{u}_i, \mathbf{u}_j, \mathbf{n}) = \frac{1}{4}(\mathbf{v}_j - \mathbf{v}_i)^T D(\mathbf{v}_j - \mathbf{v}_i) > 0. \tag{6.46}$$

Summing over all cells and assuming periodic boundary conditions lead to the semi-discrete global entropy stability statement:

$$\frac{d}{dt}\left(\sum_i U(\mathbf{u}_i)\right) = -\sum_i \mathcal{E}_i < 0 \tag{6.47}$$

The cell valued field $\mathcal{E}_i$ tells us how much entropy is produced in space in response to the jumps in entropy variables across interfaces. It can also give us an idea of the magnitude of the entropy fluctuations the scheme creates. To this end, we proceed to derive a more detailed expression for $\mathcal{E}$. In compact notation, and ignoring the 1/4 factor, we have:

$$\mathcal{E} = [\mathbf{v}]^T D[\mathbf{v}] = [\mathbf{v}]^T R|\Lambda|R^T[\mathbf{v}].$$

Let $R = [\mathbf{r_1}\ \dots\ \mathbf{r_N}]$, $|\Lambda| = diag(|\lambda_1|,\ \dots |\lambda_N|)$ and $\boldsymbol{\mu}^T = [\mu_1,\ \dots,\ \mu_N] = [\mathbf{v}]^T R$. Then:

$$\mathcal{E} = \boldsymbol{\mu}^T|\Lambda|\boldsymbol{\mu} = \sum_{i=1}^{N} |\lambda_i|\mu_i^2,$$

and we can see how the total entropy production is the sum of the positive contributions from each mode $\mathbf{r_i}$. This decomposition is inspired by how Roe & Pike [241] rewrote the Roe flux:

$$R|\Lambda|R^{-1}[\mathbf{u}] = \sum_i |\lambda_i|\alpha_i \mathbf{r_i},\ \boldsymbol{\alpha} = R^{-1}[\mathbf{u}]. \tag{6.48}$$

It is also inspired by the family of closed-form entropy-conservative fluxes Tadmor proposed in [51] (section 6). In that work, Tadmor sought to derive entropy-conservative and entropy-stable schemes for which qualitative statements about the treatment of waves could be made. In a way, we are following up on his efforts.

The vector $\boldsymbol{\alpha}$ in equation (6.48) is typically referred to as a vector of wave strengths. We can interpret $\boldsymbol{\mu} = R^T[\mathbf{v}]$ in our decomposition as a vector of wave strengths as well, because for infinitesimal disturbances we have:

$$d\mathbf{u} = Hd\mathbf{v} = RR^T d\mathbf{v} \implies R^{-1}d\mathbf{u} = R^T d\mathbf{v}.$$

For the non-dimensional compressible Euler equations, we have:

$$\mathcal{E} = |u_n|(\mu_1^2 + \mu_2^2 + \mu_3^2) + |u_n + (a/M_r)|\mu_4^2 + |u_n - (a/M_r)|\mu_5^2. \tag{6.49}$$

It breaks down into entropy production due to convective modes (first three terms) and entropy production due to acoustic modes (remaining two terms). We expect the latter to be the key in understanding the low Mach problems.

At this point, we remind the reader that all the derivations of this work are made with the non-dimensional variables. The entropy production field $\hat{\mathcal{E}}$ that the code solving the dimensional system computes is related to $\mathcal{E}$ by the simple relation:

$$\hat{\mathcal{E}} = (\rho_r u_r) \times \mathcal{E}. \tag{6.50}$$

Given that in both test problems the reference Mach number M_r is adjusted by changing the the reference velocity u_r only, we simply use:

$$\hat{\mathcal{E}} = M_r \times \mathcal{E}. \tag{6.51}$$

We also define the global quantity:

$$\langle \mathcal{E} \rangle = \sum_i V_i \mathcal{E}_i, \tag{6.52}$$

which we will subsequently use to visualize the influence of each entropy production field in (6.49) on the total entropy production in space.

Convective modes

The scaled eigenvectors associated with $\lambda = u_n$ are given by $\mathbf{r_1} = K_q(n_x K_0 \mathbf{r_0} + (a/M_r)\mathbf{r_{sx}})$, $\mathbf{r_2} = K_q(n_y K_0 \mathbf{r_0} + (a/M_r)\mathbf{r_{sy}})$ and $\mathbf{r_3} = K_q(n_z K_0 \mathbf{r_0} + (a/M_r)\mathbf{r_{sz}})$ where:

$$\mathbf{r_0} = \begin{bmatrix} 1 \\ u \\ v \\ w \\ \frac{M_r^2}{2}(u^2+v^2+w^2) \end{bmatrix}, \mathbf{r_{sx}} = \begin{bmatrix} 0 \\ 0 \\ n_z \\ -n_y \\ M_r^2(n_z v - n_y w) \end{bmatrix}, \mathbf{r_{sy}} = \begin{bmatrix} 0 \\ -n_z \\ 0 \\ n_x \\ M_r^2(n_x w - n_z u) \end{bmatrix},$$

$$\mathbf{r_{sz}} = \begin{bmatrix} 0 \\ n_y \\ -n_x \\ 0 \\ M_r^2(n_y u - n_x v) \end{bmatrix}, K_q = (\rho/\gamma)^{1/2},\ K_0 = (\gamma-1)^{1/2}.$$

$\mathbf{r_0}$ is an entropy wave. Let $\mu_0 = \mathbf{r_0}^T[\mathbf{v}]$ be the corresponding wave strength. It is then given by:

$$
\begin{aligned}
\mu_0 &= \frac{-[s]}{\gamma - 1} - M_r^2\left[\frac{\rho}{p}k\right] + M_r^2 u\left[\frac{\rho u}{p}\right] + M_r^2 v\left[\frac{\rho v}{p}\right] + M_r^2 w\left[\frac{\rho w}{p}\right] - \frac{M_r^2}{2}(u^2+v^2+w^2)\left[\frac{\rho}{p}\right] \\
&= \frac{-[s]}{\gamma - 1} - M_r^2\overline{\left(\frac{\rho}{p}\right)}\Big([k] - (u[u] + v[v] + w[w])\Big) + \\
&\qquad M_r^2\left[\frac{\rho}{p}\right]\Big(u\overline{u} + v\overline{v} + w\overline{w} - (\frac{1}{2}(u^2+v^2+w^2) + \overline{k})\Big)
\end{aligned}
$$

Arithmetic averages are typically used for the velocities. This gives:

$$
\mu_0 = \frac{-[s]}{\gamma - 1} + M_r^2\left[\frac{\rho}{p}\right]\left(\frac{1}{2}(\overline{u}^2 + \overline{v}^2 + \overline{w}^2) - \overline{k}\right) = \frac{-[s]}{\gamma - 1} - \frac{M_r^2}{8}\left[\frac{\rho}{p}\right]\Big([u]^2 + [v]^2 + [w]^2\Big). \tag{6.53}
$$

$\mathbf{r_{sx}}, \mathbf{r_{sy}}$ and $\mathbf{r_{sz}}$ are shear waves (they satisfy $n_x\mathbf{r_{sx}} + n_y\mathbf{r_{sy}} + n_z\mathbf{r_{sz}} = 0$). The corresponding wave strengths $\mu_{sx} = \mathbf{r_{sx}}^T[\mathbf{v}]$, $\mu_{sy} = \mathbf{r_{sy}}^T[\mathbf{v}]$ and $\mu_{sz} = \mathbf{r_{sz}}^T[\mathbf{v}]$ are given by:

$$
\mu_{sx} = M_r^2\overline{\left(\frac{\rho}{p}\right)}[\mathcal{V}_x],\ \mu_{sy} = M_r^2\overline{\left(\frac{\rho}{p}\right)}[\mathcal{V}_y],\ \mu_{sz} = M_r^2\overline{\left(\frac{\rho}{p}\right)}[\mathcal{V}_z],
$$

$$
\mathcal{V}_x = n_z v - n_y w,\ \mathcal{V}_z = n_x w - n_z u,\ \mathcal{V}_z = n_y u - n_x v,
$$

and the entropy produced by the convective modes is given by:

$$
\begin{aligned}
\mathcal{E}_{u_n} &= |u_n|(\mu_1^2 + \mu_2^2 + \mu_3^2) \\
&= |u_n|K_q^2\Big(\ (n_x K_0\mu_0 + (a/M_r)\mu_{sx})^2\ +\ (n_y K_0\mu_0 + (a/M_r)\mu_{sy})^2\ + \\
&\qquad\qquad (n_z K_0\mu_0 + (a/M_r)\mu_{sz})^2\ \Big) \\
&= |u_n|K_q^2\Big(\ K_0^2\mu_0^2\ +\ (a/M_r)^2(\mu_{sx}^2 + \mu_{sy}^2 + \mu_{sz}^2)\ \Big) \\
&= |u_n|\Big(\ \frac{(\gamma-1)}{\gamma}\rho\mu_0^2\ +\ \frac{\rho a^2}{\gamma M_r^2}(\mu_{sx}^2 + \mu_{sy}^2 + \mu_{sz}^2)\Big).
\end{aligned}
$$

That is:

$$\mathcal{E}_{u_n} = |u_n|\bigg(\frac{\gamma - 1}{\gamma}\rho\mu_0^2 \;+\; \alpha\rho M_r^2([\mathcal{V}_x]^2 + [\mathcal{V}_y]^2 + [\mathcal{V}_z]^2)\bigg),\ \alpha = \frac{a^2}{\gamma}\overline{\left(\frac{\rho}{p}\right)}^2, \tag{6.54}$$

which we rewrite as the sum of a contribution due to entropy waves $\mathcal{E}_{u_n,s}$ (first term) and a contribution due to shear waves $\mathcal{E}_{u_n,\tau}$ (remaining three terms).

$$\mathcal{E}_{u_n} = \mathcal{E}_{u_n,s} + \mathcal{E}_{u_n,\tau}.$$

Acoustic modes

The scaled acoustic eigenvectors $\mathbf{r}_{4,5} = \mathbf{r}_{u_n \pm a}$ are given by :

$$\mathbf{r_{u_n \pm a}} = K_a \begin{bmatrix} 1 \\ u \pm n_x(a/M_r) \\ v \pm n_y(a/M_r) \\ w \pm n_z(a/M_r) \\ h + M_r^2 k \pm u_n a M_r \end{bmatrix},\ K_a = \left(\frac{\rho}{2\gamma}\right)^{1/2}.$$

The acoustic wave strengths $\mu_{u_n \pm a}$ are given by:

$$\mu_{u_n \pm a} = K_a\bigg(\mu_0 - h\left[\frac{\rho}{p}\right] \pm M_r a\overline{\left(\frac{\rho}{p}\right)}[u_n]\bigg).$$

The entropy production field due to acoustic modes therefore writes:

$$\mathcal{E}_{u_n \pm a} = |u_n \pm (a/M_r)| K_a^2\bigg(\mu_0 - h\left[\frac{\rho}{p}\right] \pm M_r a\overline{\left(\frac{\rho}{p}\right)}[u_n]\bigg)^2.$$

Summary

Overall, the discrete entropy production field $\mathcal{E} = [\mathbf{v}]^T D[\mathbf{v}]$ can be decomposed as:

$$\mathcal{E} = \mathcal{E}_{u_n,s} + \mathcal{E}_{u_n,\tau} + \mathcal{E}_{u_n+a} + \mathcal{E}_{u_n-a} \tag{6.55}$$

For cartesian grids, which we work with, the normal vectors are along the unitary directions and the entropy production due to shear can be broken down into 6 terms:

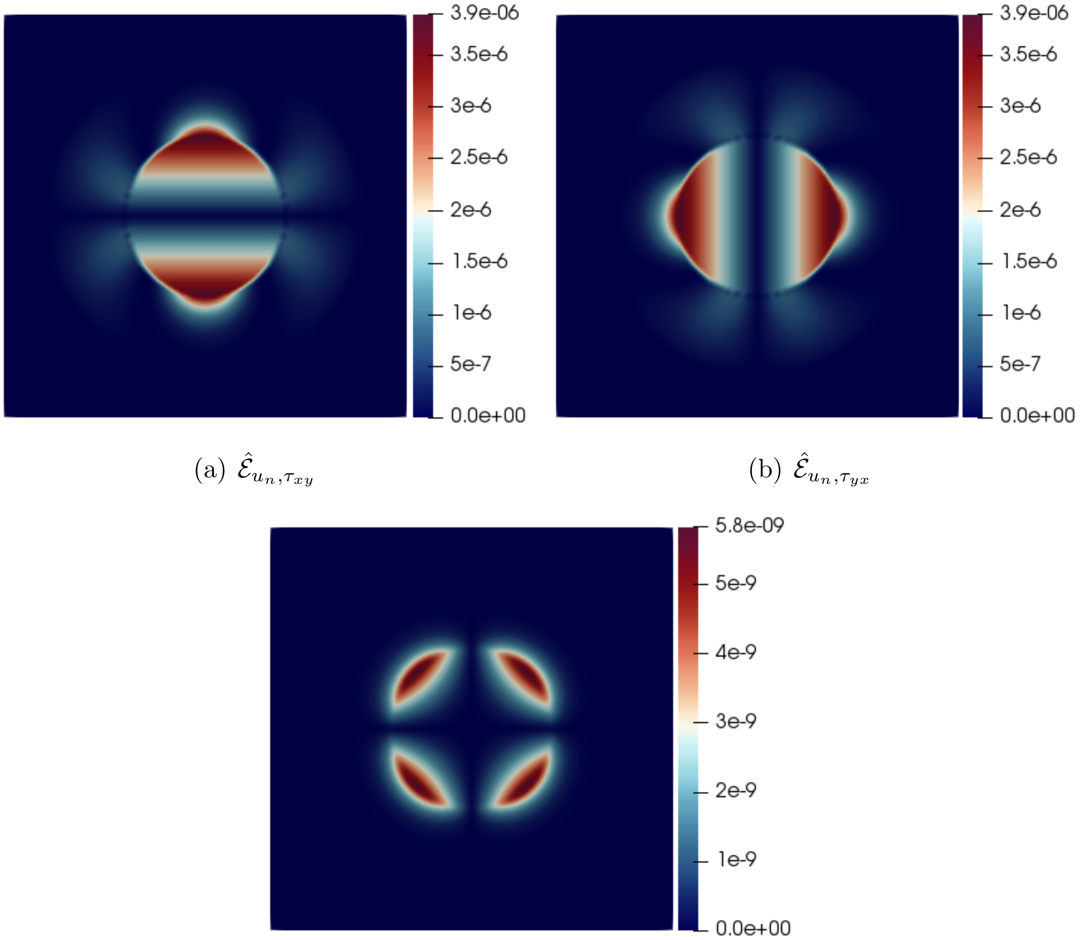
- Along x, shear in y ($\mathcal{E}_{u_n,\tau_{xy}} = |u_n|\mu_3^2$) and shear in z ($\mathcal{E}_{u_n,\tau_{xz}} = |u_n|\mu_2^2$).
- Along y, shear in z ($\mathcal{E}_{u_n,\tau_{yz}} = |u_n|\mu_1^2$) and shear in x ($\mathcal{E}_{u_n,\tau_{yx}} = |u_n|\mu_3^2$).
- Along z, shear in y ($\mathcal{E}_{u_n,\tau_{zy}} = |u_n|\mu_1^2$) and shear in x ($\mathcal{E}_{u_n,\tau_{zx}} = |u_n|\mu_2^2$).

This gives:

$$\mathcal{E} = \left(\mathcal{E}_{u_n,s} + \left(\mathcal{E}_{u_n,\tau_{xy}} + \mathcal{E}_{u_n,\tau_{xz}} + \mathcal{E}_{u_n,\tau_{yx}} + \mathcal{E}_{u_n,\tau_{yz}} + \mathcal{E}_{u_n,\tau_{zx}} + \mathcal{E}_{u_n,\tau_{zy}}\right)\right) + \mathcal{E}_{u_n+a} + \mathcal{E}_{u_n-a}. \tag{6.56}$$

Each one of these entropy production fields can be visualized. Figures 6.14 and 6.15 show them for the Gresho vortex at $t = 0$. This is, to the best of our knowledge, the first time that such a concrete view on how an entropy-stable scheme produces entropy locally is given. We will not delve into the details of why each field is the way is it (the analytical formulas we just derived would be used for that purpose). What is striking is that the acoustic entropy production fields are 2 to 5 orders of magnitudes bigger than the convective ones. Figures 6.16(a)-(c) show that the acoustic entropy production fields make for most of the entropy produced by the ES Roe flux.

Similarly, figures 6.17(a)-(c) show the entropy production fields at $t = 0$ for the acoustic wave. The entropy production field associated with entropy waves (there is no shear in this one-dimensional setup) and the entropy production field associated with left-moving acoustic waves are negligible compared to the entropy production field associated with right-moving acoustic waves. This makes sense, and over time, this power difference is sustained as shown in figures 6.18(a)-(c).



(a) $\hat{\mathcal{E}}_{u_n,\tau_{xy}}$

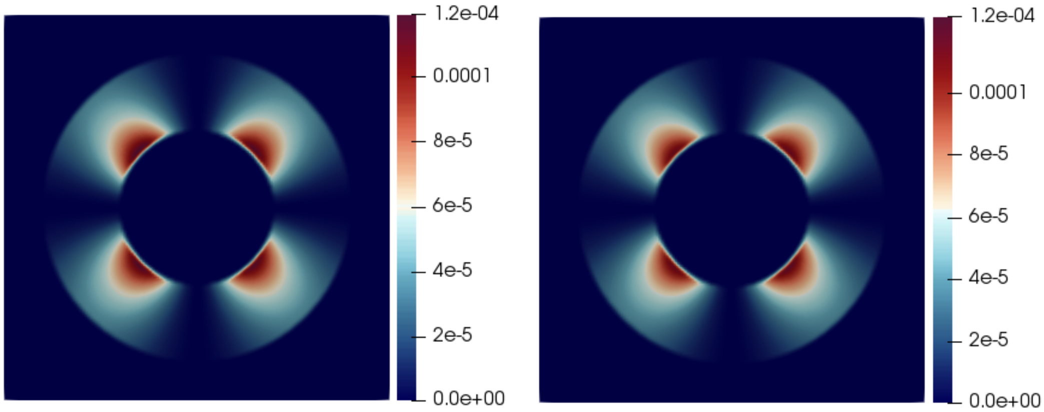
(b) $\hat{\mathcal{E}}_{u_n,\tau_{yx}}$

(c) $\hat{\mathcal{E}}_{u_n,s}$

Figure 6.14: Gresho vortex: Entropy production fields associated with the convective modes at $t = 0$ and $M_r = 3 \times 10^{-2}$. These are common to all ES fluxes.

6.5.1.1 Preconditioned operators

A similar decomposition can be obtained with the preconditioned dissipation operators $D_P[\mathbf{v}]$ of Turkel and Miczek, but it requires that they be written in a scaled form $R|\Lambda|R^T$. We were not far from one such form in section 6.3. Indeed, from



(a) $\hat{\mathcal{E}}_{u_n+a}$

(b) $\hat{\mathcal{E}}_{u_n-a}$

Figure 6.15: Gresho vortex: Acoustic Entropy production fields at $t = 0$ of the ES Roe flux at $M_r = 3 \times 10^{-2}$.

equation (6.38), which we rewrite here:

$$D_P = P^{-1}|PA|[\mathbf{v}] = (QH_{\mathbf{z}}^{1/2}P_{\mathbf{z}}^{-1})|P_{\mathbf{z}}A_{\mathbf{z}}|P_{\mathbf{z}}^T(QH_{\mathbf{z}}^{1/2}P_{\mathbf{z}}^{-1})^T,$$

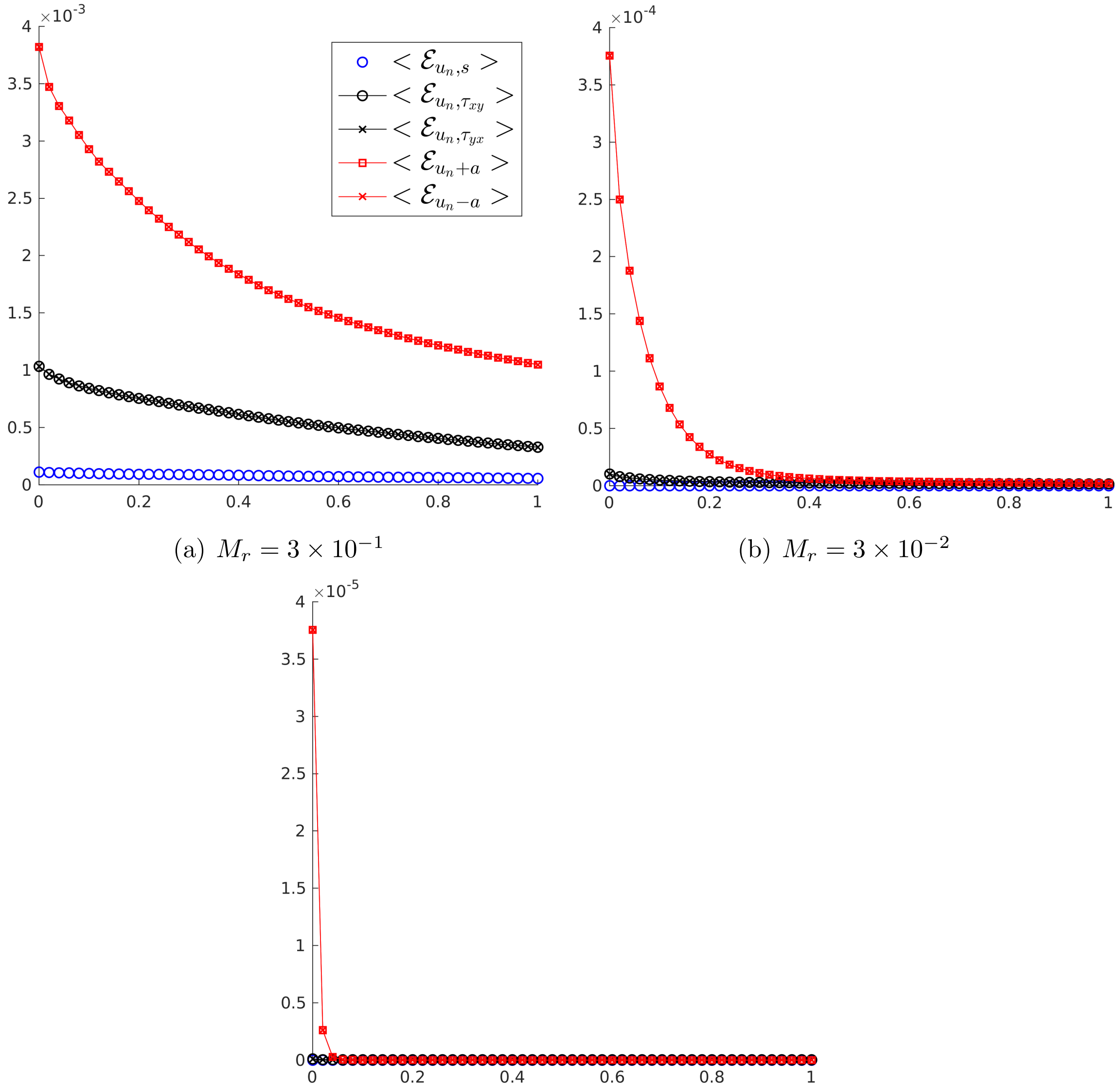
it is clear that finding a scaled form for D_P is equivalent to finding one for the congruent matrix $|P_{\mathbf{z}}A_{\mathbf{z}}|P_{\mathbf{z}}^T$. A simple trick to proceed, picked up from Diosady & Murman [68] (section IV), consists in "forcing the eigenscaling theorem" by introducing a matrix T_p defined by $P_{\mathbf{z}}^T = R_pT_pR_p^T$. This gives $|P_{\mathbf{z}}A_{\mathbf{z}}|P_{\mathbf{z}}^T = R_p|\Lambda_p|T_pR_p^T$ and ultimately:

$$D_P = \mathcal{R}_p(|\Lambda_p|T_p)\mathcal{R}_p^T, \ \mathcal{R}_p = QH_{\mathbf{z}}^{1/2}P_{\mathbf{z}}^{-1}R_p. \tag{6.57}$$

From here, one hopes that T_p is diagonal positive.

For Turkel's flux-preconditioner, that is the case because the eigenscaling theorem applies. T_p is given by:

$$T_p = diag([1, \ 1, \ 1, \ K_2^2 + p^2, \ K_1^2 + p^2]).$$

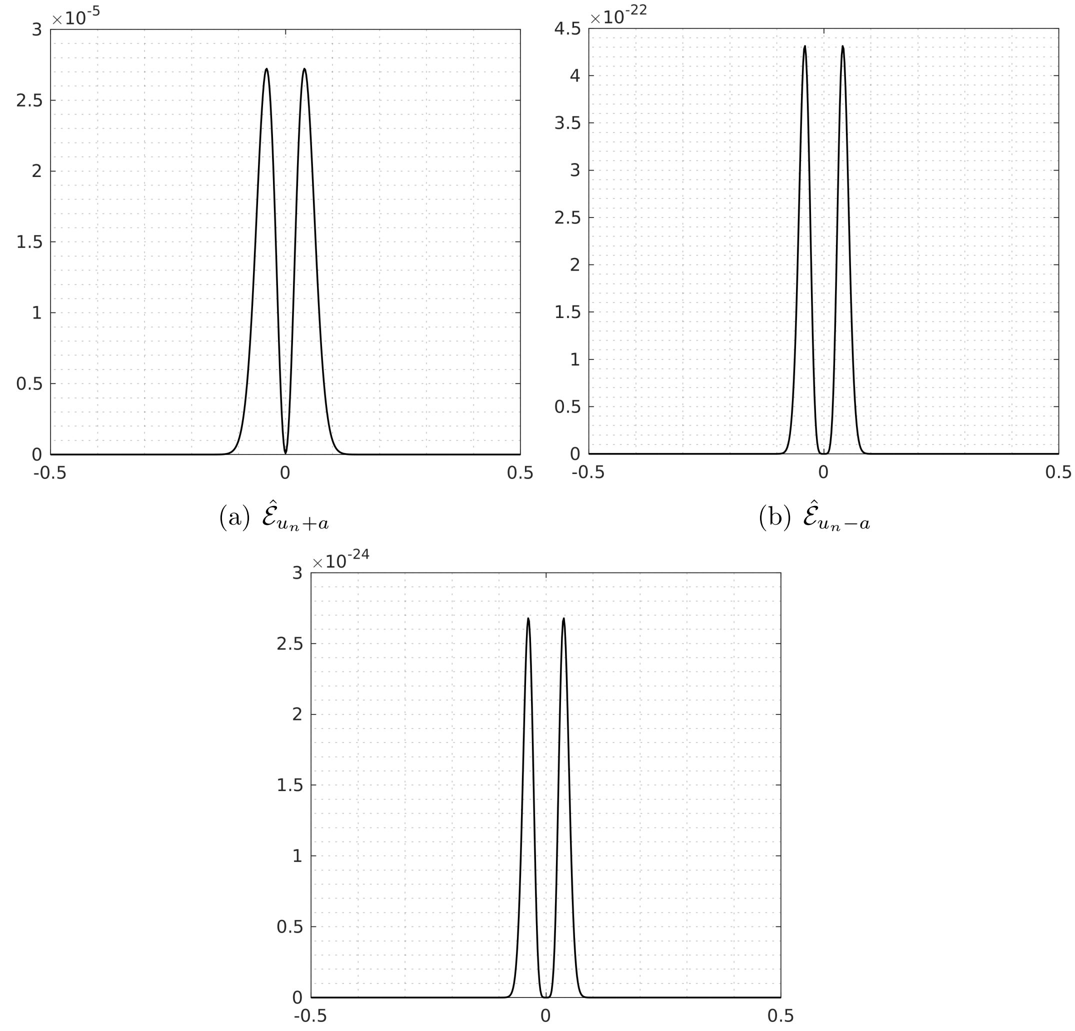


(a) $M_r = 3 \times 10^{-1}$

(b) $M_r = 3 \times 10^{-2}$

(c) $M_r = 3 \times 10^{-3}$

Figure 6.16: Gresho Vortex: Integral of each entropy production field with time at different Mach numbers for the ES Roe flux.

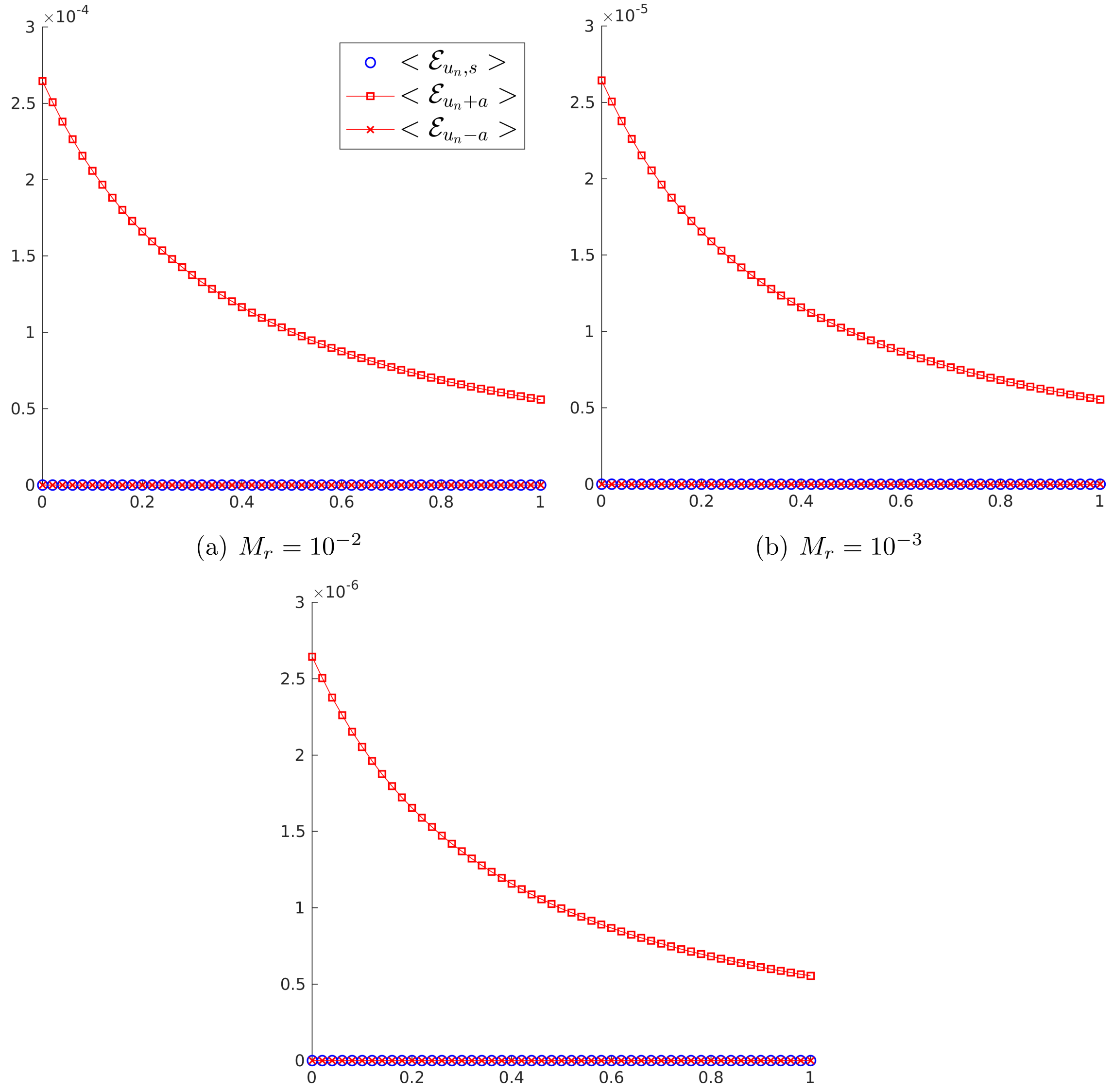


(a) $\hat{\mathcal{E}}_{u_n+a}$

(b) $\hat{\mathcal{E}}_{u_n-a}$

(c) $\hat{\mathcal{E}}_{u_n,s}$

Figure 6.17: Gresho vortex: Entropy production fields at different Mach numbers at $t = 0$ and $M_r = 10^{-2}$ for the ES Roe flux.



(a) $M_r = 10^{-2}$

(b) $M_r = 10^{-3}$

(c) $M_r = 10^{-4}$

Figure 6.18: Sound wave: Integral of each entropy production field with time at different Mach numbers for the ES Roe flux.

The corresponding entropy production fields will differ from (6.55)-(6.56) in the acoustic part. We have:

$$\mathcal{E}_p = [\mathbf{v}]^T \mathcal{R}_p (\Lambda_p T_p) \mathcal{R}_p^T [\mathbf{v}] = \mathcal{E}_{u_n,s} + \mathcal{E}_{u_n,\tau} + \mathcal{E}_{u_{np}+a_p} + \mathcal{E}_{u_{np}-a_p}, \tag{6.58}$$

$$\mathcal{E}_{u_{np}+a_p} = \mu^2_{u_{np}+a_p} (K_2^2 + p^2)|u_{np} + a_p|,\ \mathcal{E}_{u_{np}-a_p} = \mu^2_{u_{np}-a_p} (K_1^2 + p^2)|u_{np} - a_p|,$$

where $\mu_{u_{np}\pm a_p} = \mathbf{r}^T_{u_{np}\pm a_p}[\mathbf{v}]$ and

$$\mathbf{r}_{u_{np}+a_p} = \sqrt{\frac{\rho}{\gamma}} \frac{K_1}{p^2(K_1 - K_2)} \begin{bmatrix} 1 \\ u + n_x(a/M_r)(p^2/K_1) \\ v + n_y(a/M_r)(p^2/K_1) \\ w + n_z(a/M_r)(p^2/K_1) \\ h^t + u_n a M_r (p^2/K_1) \end{bmatrix},$$

$$\mathbf{r}_{u_{np}-a_p} = \sqrt{\frac{\rho}{\gamma}} \frac{K_2}{p^2(K_1 - K_2)} \begin{bmatrix} 1 \\ u + n_x(a/M_r)(p^2/K_2) \\ v + n_y(a/M_r)(p^2/K_2) \\ w + n_z(a/M_r)(p^2/K_2) \\ h^t + u_n a M_r (p^2/K_2) \end{bmatrix},$$

$$u_{np} = 0.5u_n(p^2 + 1),\ a_p = (u_{np}^2 + p^2(a^2/M_r^2 - u_n^2))^{1/2},$$

$$K_1 = (u_{np} - u_n + a_p)M_r/a,\ K_2 = (u_{np} - u_n - a_p)M_r/a.$$

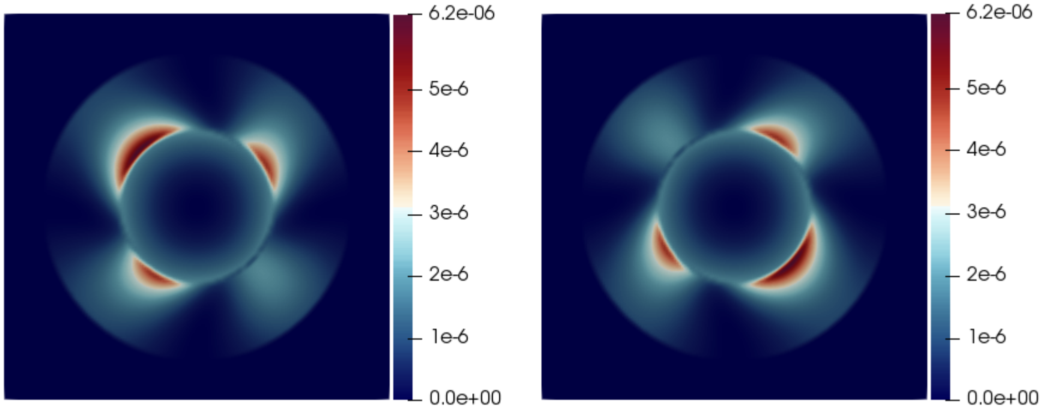
The modified acoustic wave strengths write:

$$\mu_{u_{np}+a_p} = \mathbf{r}^T_{u_{np}+a_p}[\mathbf{v}] = \sqrt{\frac{\rho}{\gamma}} \frac{K_1}{p^2(K_1 - K_2)} \left(\mu_0 - h\left[\frac{\rho}{p}\right] + \frac{aM_r p^2}{K_1} \overline{\left(\frac{\rho}{p}\right)} [u_n] \right),$$

$$\mu_{u_{np}-a_p} = \mathbf{r}^T_{u_{np}-a_p}[\mathbf{v}] = \sqrt{\frac{\rho}{\gamma}} \frac{K_2}{p^2(K_1 - K_2)} \left(\mu_0 - h\left[\frac{\rho}{p}\right] + \frac{aM_r p^2}{K_2} \overline{\left(\frac{\rho}{p}\right)} [u_n] \right).$$

Figures 6.19(a)-(b) show the modified acoustic entropy production fields for the Gresho vortex at $t = 0$. The two fields are of the same magnitude and they appear to be in some sort of symmetry. Figures 6.20(a)-(c) show that the total acoustic entropy production fields are of the same order as those associated with the shear waves over time.

For the sound wave, the initial entropy production fields showed in figures 6.21(a)-(b) illustrate why we refrain from describing the entropy production decomposition in terms of "waves". The preconditioning leads to modified acoustic eigenvectors which can no longer be tied to right-moving and left-moving moving waves. The flow consists of a right-moving acoustic wave, yet we see that both entropy production fields $\mathcal{E}_{u_{np} \pm a_p}$ are active. Figures 6.22(a)-(c) show the overwhelming domination of both the acoustic entropy production fields.

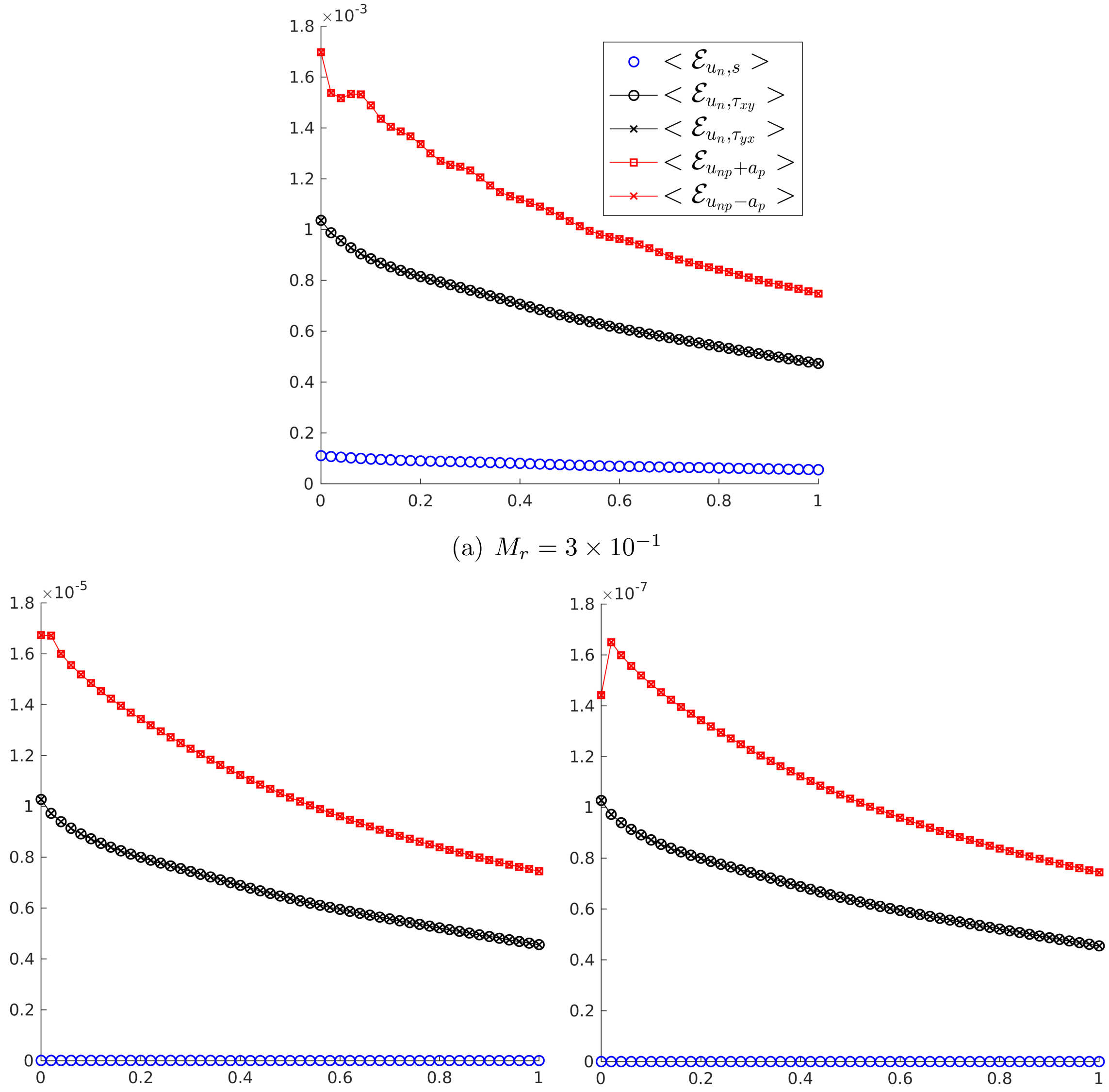


(a) $\hat{\mathcal{E}}_{u_{np}+a_p}$

(b) $\hat{\mathcal{E}}_{u_{np}-a_p}$

Figure 6.19: Gresho Vortex: Entropy production fields at $t = 0$ for the ES Turkel flux $M_r = 3 \times 10^{-2}$.

For Miczek's flux-preconditioner, the eigenscaling theorem does not apply, but we

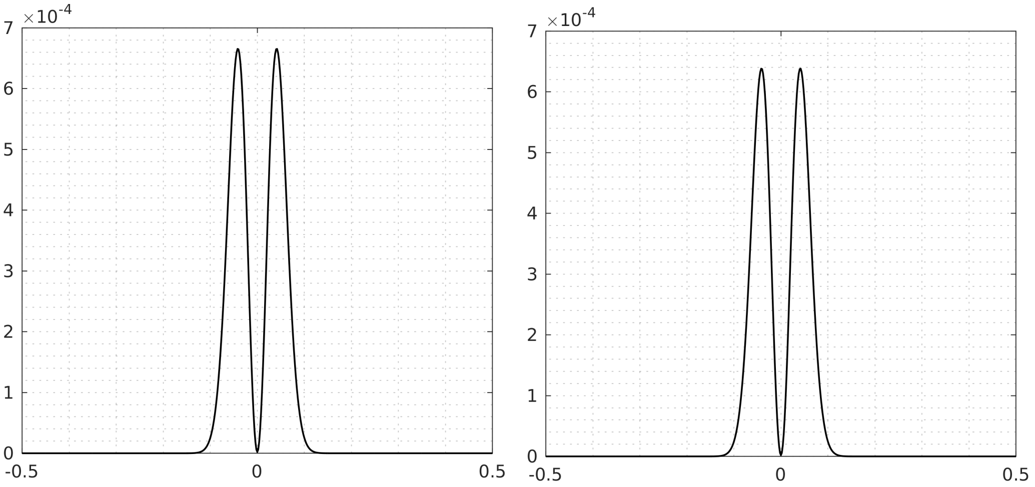


(a) $M_r = 3 \times 10^{-1}$

(b) $M_r = 3 \times 10^{-2}$

(c) $M_r = 3 \times 10^{-3}$

Figure 6.20: Gresho Vortex: Integral of each entropy production field with time at different Mach numbers for the ES Turkel flux.



(a) $\hat{\mathcal{E}}_{u_{np}+a_p}$ (b) $\hat{\mathcal{E}}_{u_{np}-a_p}$

Figure 6.21: Sound wave: Entropy production fields at $t = 0$ for the ES Turkel flux $M_r = 10^{-2}$.

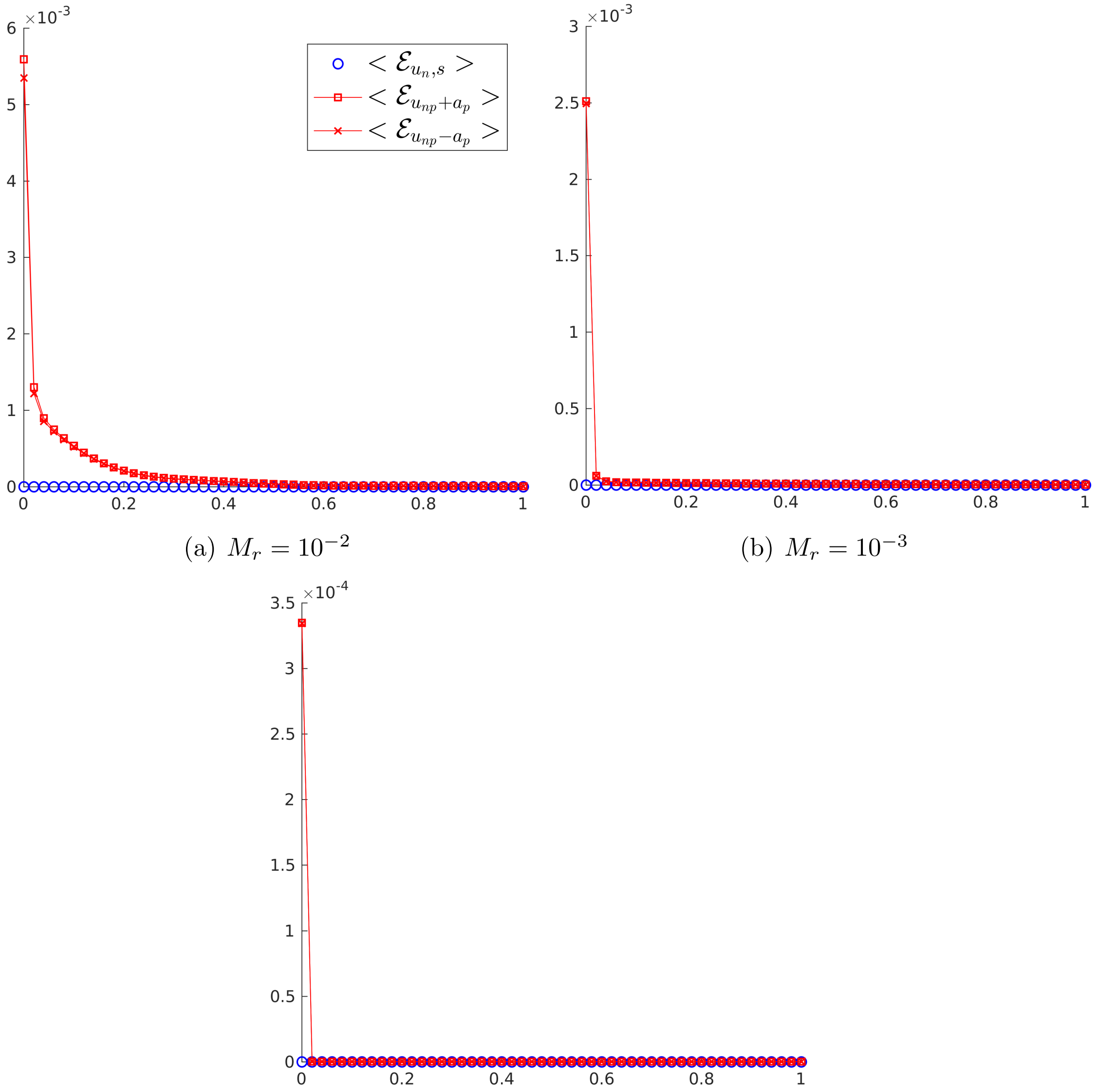
know that T_p is positive definite by congruence. It is given by:

$$T_p = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & K_2^2+1 & (K_2-K_1)p - K_1K_2 - 1 \\ 0 & 0 & 0 & (K_1-K_2)p - K_1K_2 - 1 & K_1^2+1 \end{bmatrix}.$$

Upon closer examination, D_P can be further reduced by observing that in the subsonic regime, the last 2-by-2 bloc of $|\Lambda_p|T_p$ can be rewritten as:

$$\begin{bmatrix} |u_n+a_p|(K_2^2+1) & 0 \\ 0 & |u_n-a_p|(K_1^2+1) \end{bmatrix} + \begin{bmatrix} 0 & -\delta_p \\ \delta_p & 0 \end{bmatrix},$$

$$\delta_p = 2p(p^2+1)(a^2 - M_r^2u_n^2)/(M_r(a - M_rpu_n)).$$



(a) $M_r = 10^{-2}$

(b) $M_r = 10^{-3}$

(c) $M_r = 10^{-4}$

Figure 6.22: Sound wave: Integral of each entropy production field with time at different Mach numbers for the ES Turkel flux.

We therefore have $D_P = \mathcal{R}_p(|\Lambda_p|\overline{T}_p + \Delta_p)\mathcal{R}_p^T$ with:

$$\overline{T}_p = diag([1,\ 1,\ 1,\ K_2^2+1,\ K_1^2+1]),\ \Delta_p = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\delta_p \\ 0 & 0 & 0 & \delta_p & 0 \end{bmatrix}$$

Since $\overline{T}_p$ is diagonal positive and Δ_p is skew-symmetric, it follows that D_P is positive definite for the Miczek flux-preconditioner as well. The entropy production breakdown is more subtle than for Turkel's because of the skew symmetric matrix Δ_p. The acoustic part $D_P^A[\mathbf{v}]$ of the dissipation operator $\mathcal{R}_p(\Lambda_p\overline{T}_p + \Delta_p)\mathcal{R}_p^T[\mathbf{v}]$ writes:

$$D_P^A[\mathbf{v}] = \begin{bmatrix} \mathbf{r}_{u_n+a_p} & \mathbf{r}_{u_n-a_p} \end{bmatrix} \begin{bmatrix} |u_n+a_p|(1+K_2^2) & -\delta_p \\ \delta_p & |u_n-a_p|(1+K_1^2) \end{bmatrix} \begin{bmatrix} \mu_{u_n+a_p} \\ \mu_{u_n-a_p} \end{bmatrix} \quad (6.59)$$

where:

$$\mathbf{r}_{u_n+a_p} = \sqrt{\frac{\rho}{\gamma}}\frac{1}{(p^2+1)(K_1-K_2)}\begin{bmatrix}(K_1-p)\\ u(K_1-p)+(a/M_r)n_x(K_1p+1)\\ v(K_1-p)+(a/M_r)n_y(K_1p+1)\\ w(K_1-p)+(a/M_r)n_z(K_1p+1)\\ h^t(K_1-p)+u_n(aM_r)(K_1p+1)\end{bmatrix},$$

$$\mathbf{r}_{u_n-a_p} = \sqrt{\frac{\rho}{\gamma}}\frac{1}{(p^2+1)(K_1-K_2)}\begin{bmatrix}(K_2-p)\\ u(K_2-p)+(a/M_r)n_x(K_2p+1)\\ v(K_2-p)+(a/M_r)n_y(K_2p+1)\\ w(K_2-p)+(a/M_r)n_z(K_2p+1)\\ h^t(K_2-p)+u_n(aM_r)(K_2p+1)\end{bmatrix},$$

$$a_p = \sqrt{(p^2+1)a^2/M_r^2 - p^2u_n^2},$$

$$K_1 = (a+M_rpu_n)/(M_ra_p-ap),\;\; K_2 = -(a+M_rpu_n)/(M_ra_p+ap),$$

and $\mu_{u_n\pm a_p} = \mathbf{r}^T_{u_n\pm a_p}[\mathbf{v}]$ given by

$$\mu_{u_n+a_p} = \sqrt{\frac{\rho}{\gamma}}\frac{1}{(p^2+1)(K_1-K_2)}\Bigg((K_1-p)\bigg(\mu_0 - h\left[\frac{\rho}{p}\right]\bigg) + aM_r(K_1p+1)\overline{\left(\frac{\rho}{p}\right)}[u_n]\Bigg),$$
$$\mu_{u_n-a_p} = \sqrt{\frac{\rho}{\gamma}}\frac{1}{(p^2+1)(K_1-K_2)}\Bigg((K_2-p)\bigg(\mu_0 - h\left[\frac{\rho}{p}\right]\bigg) + aM_r(K_2p+1)\overline{\left(\frac{\rho}{p}\right)}[u_n]\Bigg).$$

Expanding (6.59), we have:

$$D_P^A[\mathbf{v}] = \mathbf{r}_{u_n+a_p}\Big(\mu_{u_n+a_p}|u_n+a_p|(1+K_2^2) - \delta_p\mu_{u_n-a_p}\Big) + \mathbf{r}_{u_n-a_p}\Big(\mu_{u_n-a_p}|u_n-a_p|(1+K_1^2) + \delta_p\mu_{u_n+a_p}\Big). \quad (6.60)$$

Multiplying on the left by $[\mathbf{v}]^T$ gives an acoustic entropy production field:

$$[\mathbf{v}]^T D_P^A [\mathbf{v}] = \mathcal{E}_{u_n+a_p} + \mathcal{E}_{u_n-a_p}, \tag{6.61}$$

where the fields $\mathcal{E}_{u_n \pm a_p}$ break down into contributions $\{\mathcal{E}^S_{u_n \pm a_p},\ \mp \Delta\mathcal{E}^p\}$ from the symmetric and skew-symmetric parts of the dissipation operator:

$$\mathcal{E}_{u_n+a_p} = \mathcal{E}^S_{u_n+a_p} - \Delta\mathcal{E}_p,\ \mathcal{E}_{u_n-a_p} = \mathcal{E}^S_{u_n-a_p} + \Delta\mathcal{E}_p, \tag{6.62}$$

$$\mathcal{E}^S_{u_n+a_p} = \mu^2_{u_n+a_p}|u_n + a_p|(1 + K_2^2),\ \mathcal{E}^S_{u_n-a_p} = \mu^2_{u_n-a_p}|u_n - a_p|(1 + K_1^2), \tag{6.63}$$

$$\Delta\mathcal{E}_p = \delta_p \mu_{u_n+a_p} \mu_{u_n-a_p}. \tag{6.64}$$

$\mathcal{E}_{u_n+a_p}$ and $\mathcal{E}_{u_n-a_p}$ are no longer positive in principle but their addition is always positive.

Equations (6.60) and (6.62) suggest that *while Δ_p does not change the amount of entropy $\mathcal{E}$ produced at an interface, it effects how this amount is distributed locally among the modes*. In this case, the skew-symmetric terms appear to redistribute the entropy produced through the acoustic modes.

Figures 6.23(a)-(b) show the modified acoustic entropy production fields for the Gresho vortex at $t = 0$. They resemble those of the ES Turkel flux. Figures 6.23(c)-(e) show the contributions of the symmetric and skew-symmetric terms. The skew-symmetric component is not negligible. Figures 6.24(a)-(c) show that the total acoustic entropy production fields are of the same order as those associated with the shear waves over time. We also see that the total contribution $\langle \Delta\mathcal{E}_p \rangle$ from the skew-symmetric matrix evolves in time like a damped oscillator, with a characteristic time that decreases with the Mach number. This suggests that the spurious transient causing the phase errors we observed earlier has something to do with the skew-symmetric matrix. A simple way to confirm this is to multiply the skew-symmetric term by a factor and see how it impacts the solution. This is illustrated in figures 6.25(a)-(b).

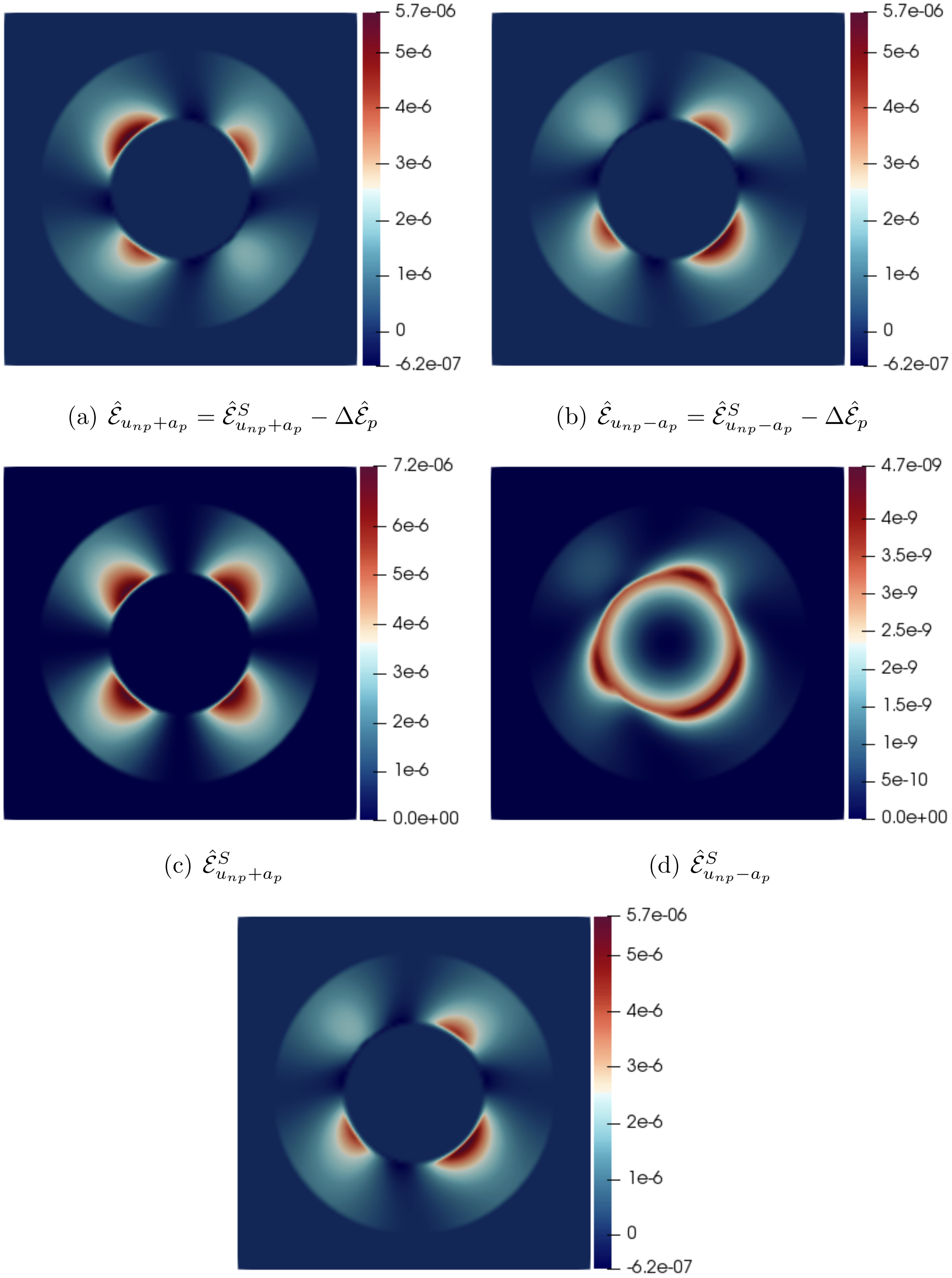
Taking out the skew-symmetric indeed removes the transient and phase errors. Making the skew-symmetric term stronger amplifies them. What's more, figure 6.26 shows that the skew-symmetric term does not have a visible impact on the ability of the scheme to conserve the kinetic energy of the system.

For the sound wave, the initial entropy production fields are showed in figures 6.27(a)-(e). The contribution from the skew-symmetric part is two orders of magnitude bigger than the contribution from the symmetric part. This is why, for visibility, we show the integrated entropy production fields in two parts (figures 6.28 and 6.29). While the perturbations we observed in figure 6.11 appear in the symmetric parts $\mathcal{E}^S_{u_n \pm a_p}$ of the acoustic entropy production fields, it turns out from figures 6.30(a)-(b) that it is the skew-symmetric term again that is causing the appearance of a spurious left-moving acoustic wave.

6.5.2 The discrete low Mach regime revisited

Using the analytical expressions we just derived, we can now determine how the entropy produced by each ES flux scales with the Mach number in the low Mach limit and establish whether (**E.1**) and (**E.2**) are satisfied. This effort, similar in spirit to the analysis of Guillard & Viozat [226], Guillard & Nkonga [228] and Bruel *et al.* [240], will provide an explanation to (**S.1**), (**S.2**) and (**S.3**) in terms of entropy production.

To verify the scaling analysis, we computed, for each ES flux, the integrated (6.52) entropy production fields at $t = 0$ for the Gresho vortex and the sound wave at different Mach numbers. These are given in figures 6.31, 6.32 and 6.33.



(a) $\hat{\mathcal{E}}_{u_{np}+a_p} = \hat{\mathcal{E}}^S_{u_{np}+a_p} - \Delta\hat{\mathcal{E}}_p$

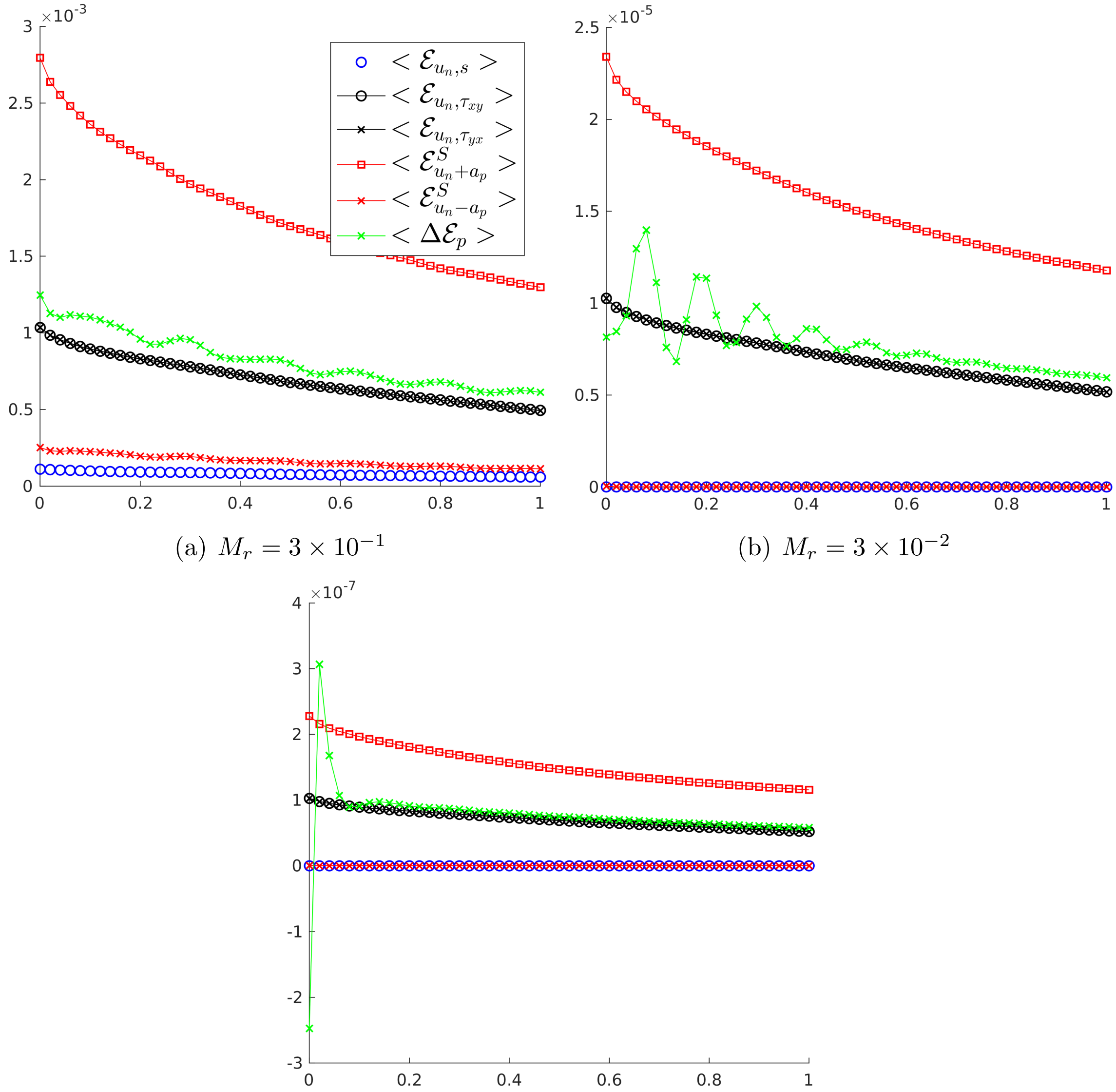
(b) $\hat{\mathcal{E}}_{u_{np}-a_p} = \hat{\mathcal{E}}^S_{u_{np}-a_p} - \Delta\hat{\mathcal{E}}_p$

(c) $\hat{\mathcal{E}}^S_{u_{np}+a_p}$

(d) $\hat{\mathcal{E}}^S_{u_{np}-a_p}$

(e) $\Delta\hat{\mathcal{E}}_p$

Figure 6.23: Gresho Vortex: Entropy production fields at $t = 0$ for the ES Miczek flux $M_r = 3 \times 10^{-2}$.

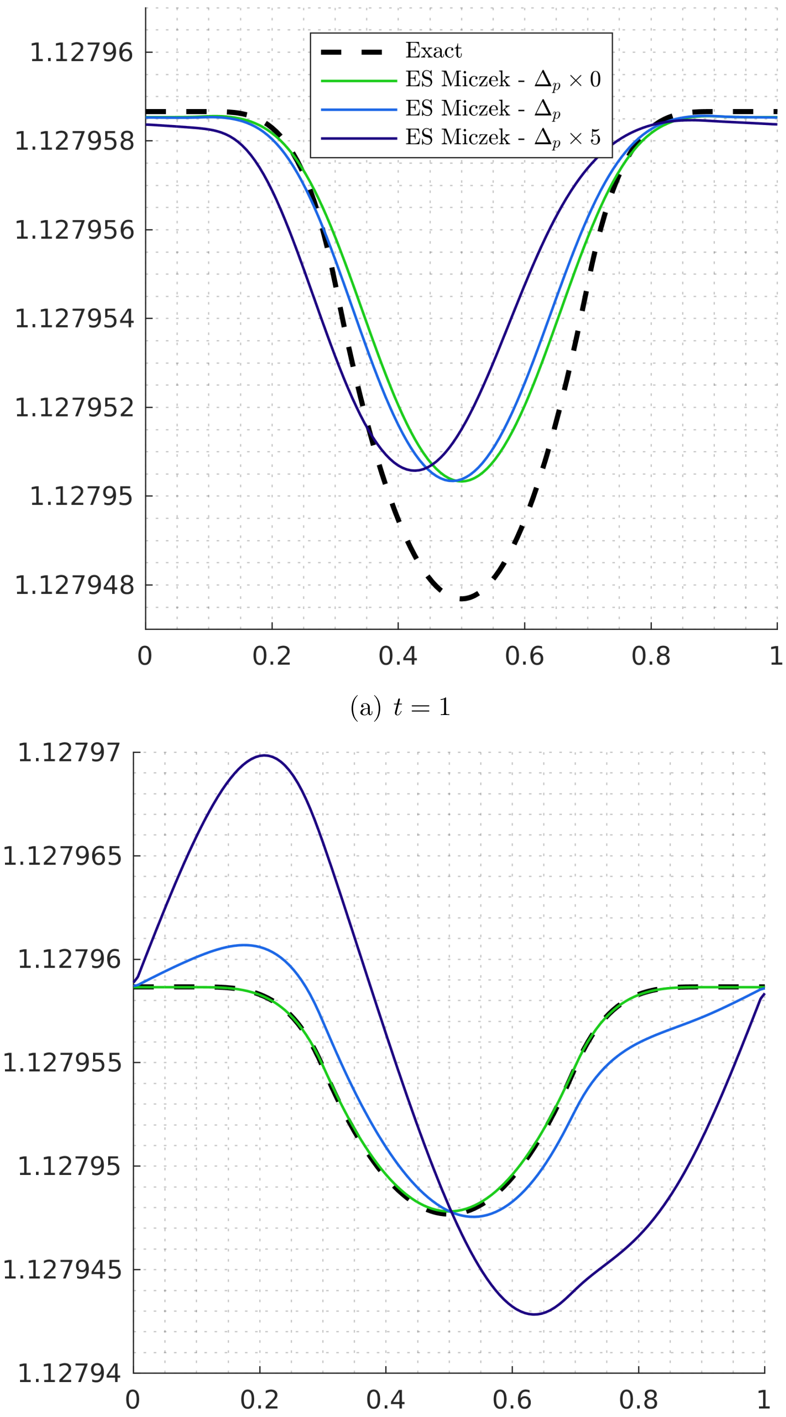


(a) $M_r = 3 \times 10^{-1}$

(b) $M_r = 3 \times 10^{-2}$

(c) $M_r = 3 \times 10^{-3}$

Figure 6.24: Gresho Vortex: Integral of each entropy production field with time at different Mach numbers for the ES Miczek flux.



(a) $t = 1$

(b) $t = 0.04$

Figure 6.25: Gresho Vortex: Pressure field at $M_r = 3 \times 10^{-3}$ for the Miczek flux when the skew-symmetric term is multiplied by a factor. The phase errors observed in figures 6.6 and 6.7 disappear if the skew-symmetric term is removed, and amplified if the skew-symmetric term is made bigger.

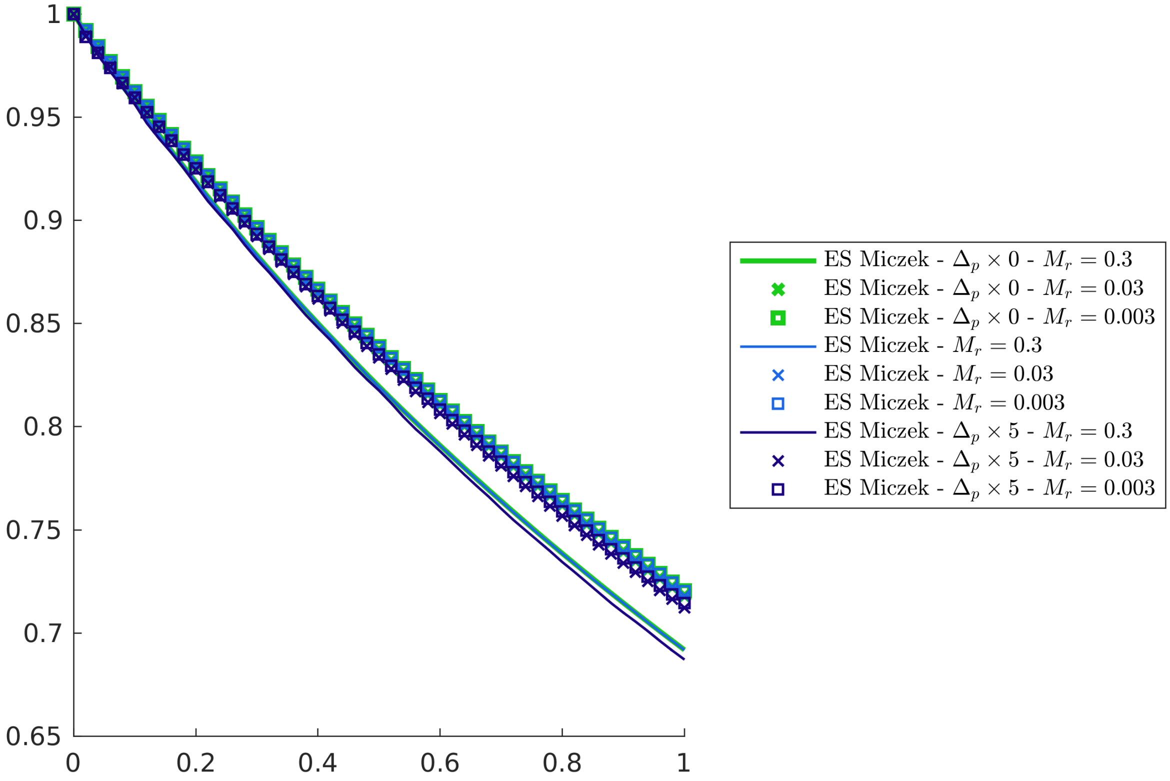
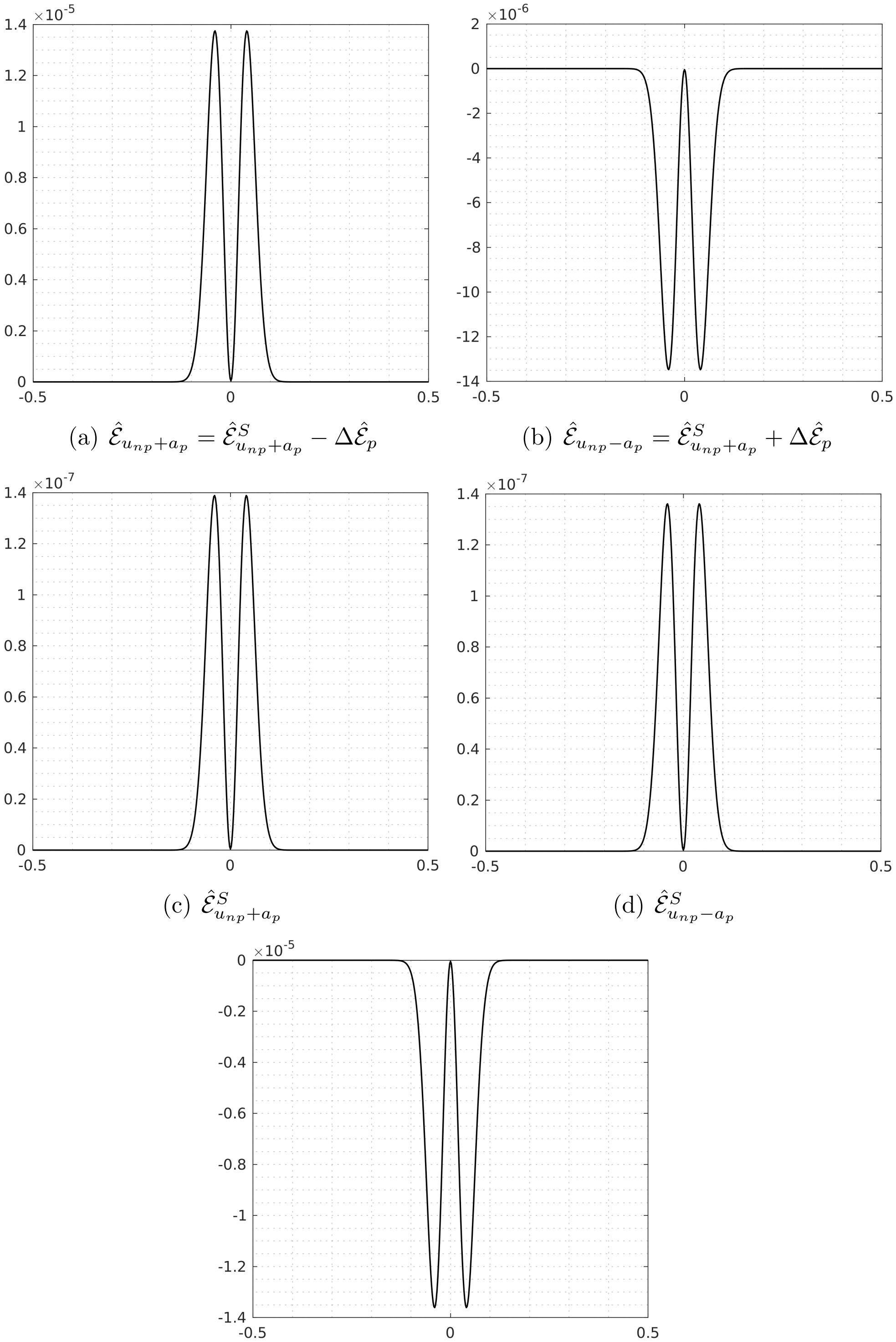


Figure 6.26: Gresho vortex: Total kinetic energy over time. The skew-symmetric matrix does not strongly impact the ability of the scheme to preserve the kinetic energy.



(a) $\hat{\mathcal{E}}_{u_{np}+a_p} = \hat{\mathcal{E}}^S_{u_{np}+a_p} - \Delta\hat{\mathcal{E}}_p$

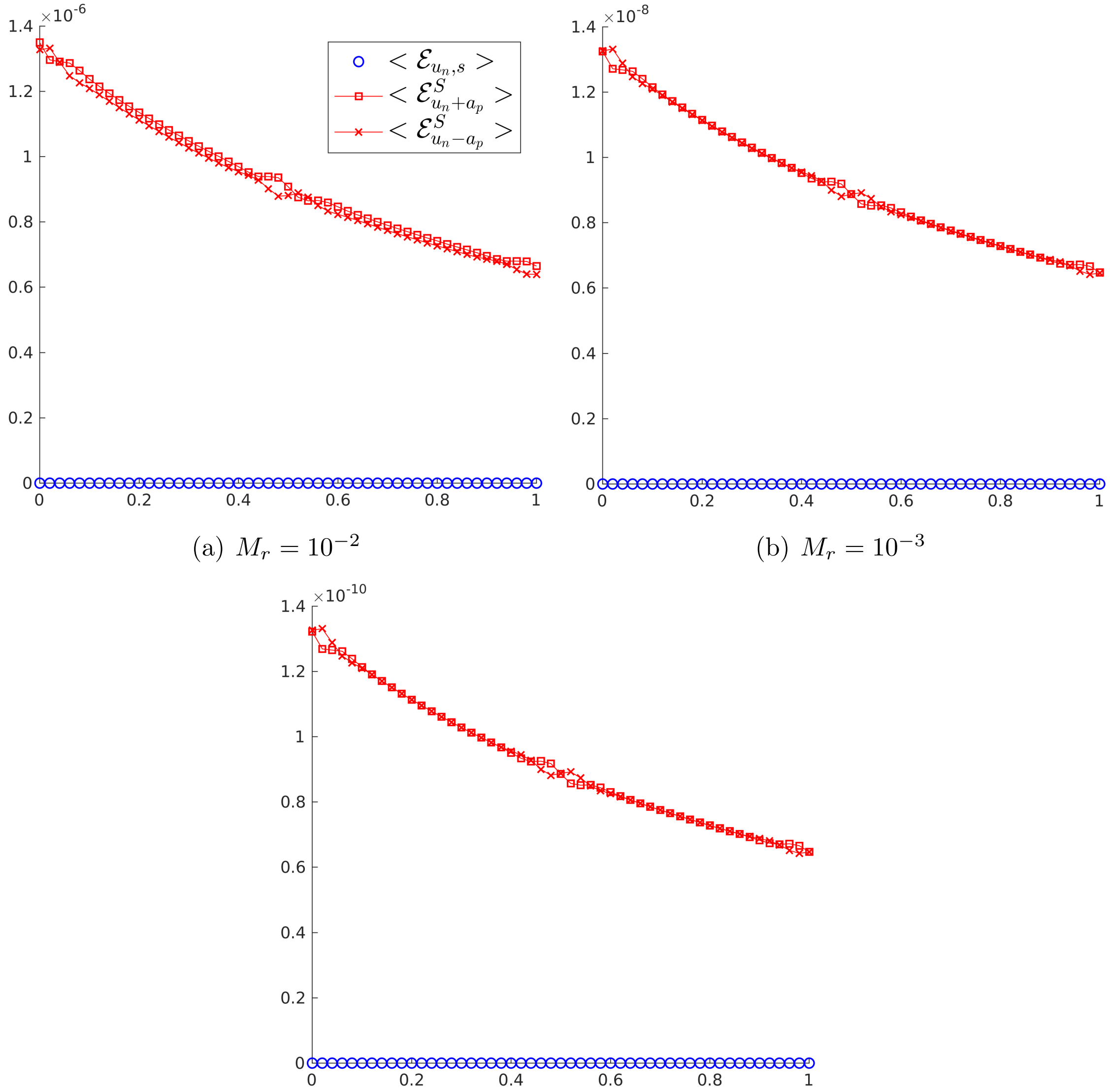
(b) $\hat{\mathcal{E}}_{u_{np}-a_p} = \hat{\mathcal{E}}^S_{u_{np}+a_p} + \Delta\hat{\mathcal{E}}_p$

(c) $\hat{\mathcal{E}}^S_{u_{np}+a_p}$

(d) $\hat{\mathcal{E}}^S_{u_{np}-a_p}$

(e) $\Delta\hat{\mathcal{E}}_p$

Figure 6.27: Sound wave: Entropy production fields at $t = 0$ for the ES Miczek flux. $M_r = 10^{-2}$.

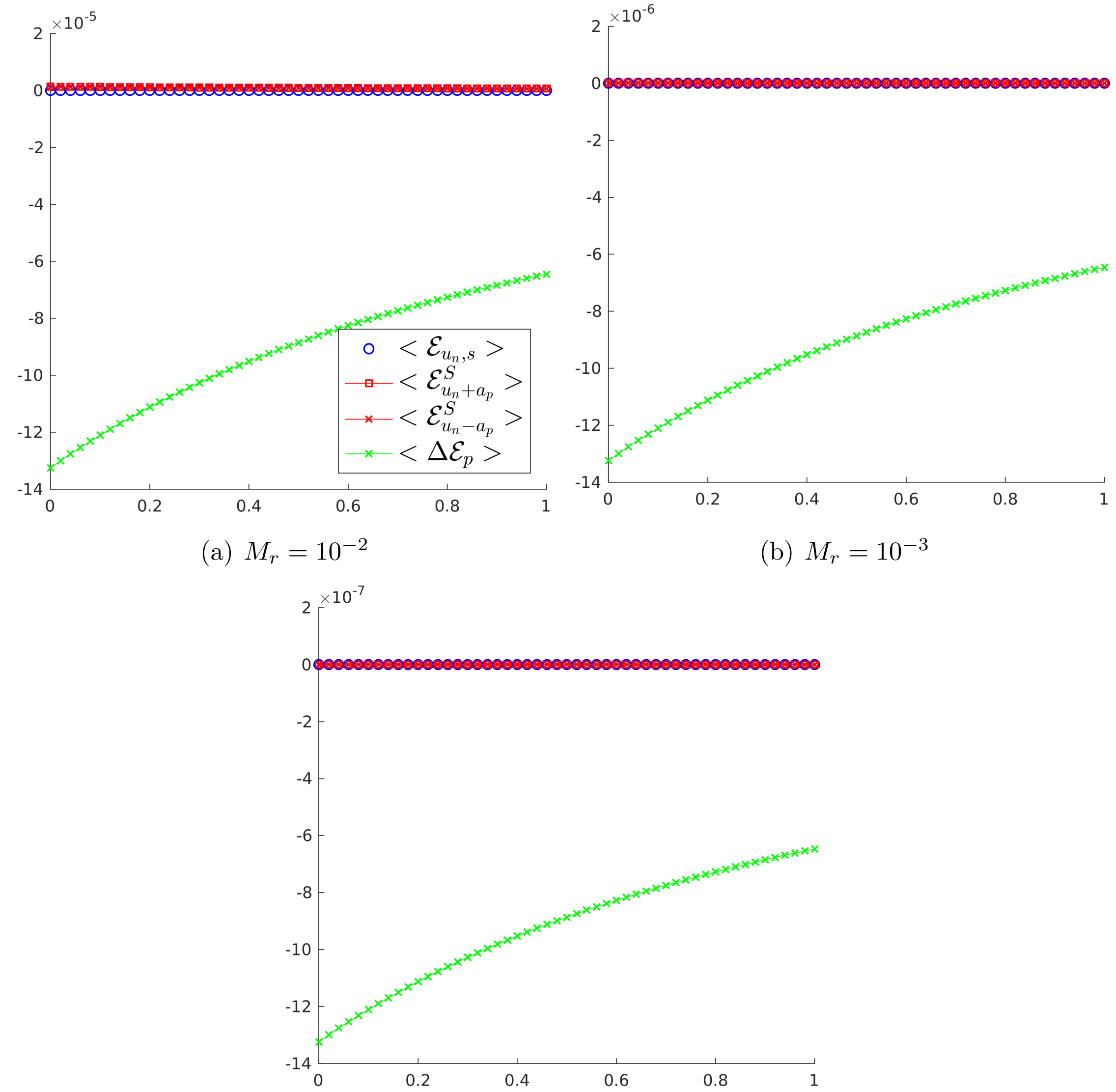


(a) $M_r = 10^{-2}$

(b) $M_r = 10^{-3}$

(c) $M_r = 10^{-4}$

Figure 6.28: Sound wave: Integral of entropy production fields, omitting the contribution of the skew-symmetric matrix, with time at different Mach numbers for the ES Miczek flux.

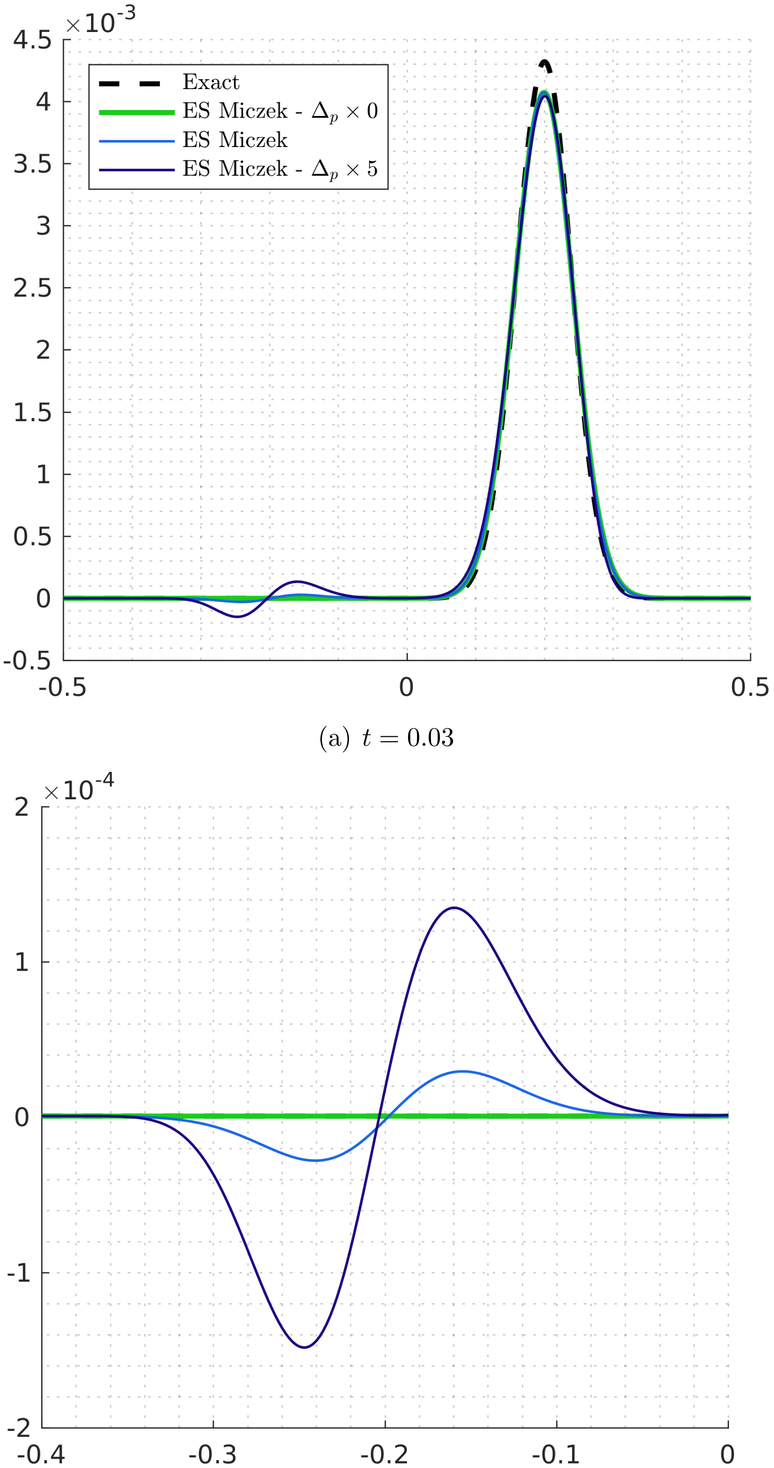


(a) $M_r = 10^{-2}$

(b) $M_r = 10^{-3}$

(c) $M_r = 10^{-4}$

Figure 6.29: Sound wave: Integral of entropy production fields, including the contribution of the skew-symmetric matrix, with time at different Mach numbers for the ES Miczek flux.



(a) $t = 0.03$

(b) $t = 0.03$, zoomed

Figure 6.30: Sound wave: Pressure snapshots showing that the spurious left-moving acoustic wave is due to the skew-symmetric term in the ES Miczek flux. $M_r = 10^{-2}$

Incompressible limit

For the Gresho vortex, density is constant, pressure and velocity variations are of order M_r^2 and 1, respectively. At the discrete level, this translates into:

$$\begin{aligned}&[\rho] = 0, [p] = \mathcal{O}(M_r^2),\ [u] = \mathcal{O}(1), [v] = \mathcal{O}(1) \implies \\ &\qquad [s] = \mathcal{O}(M_r^2),\ \left[\frac{\rho k}{p}\right] = \mathcal{O}(1),\ \left[\frac{\rho u}{p}\right] = \mathcal{O}(1),\ \left[\frac{\rho v}{p}\right] = \mathcal{O}(1),\ \left[\frac{\rho}{p}\right] = \mathcal{O}(M_r^2).\end{aligned} \tag{6.65}$$

For the classic entropy-stable upwind dissipation, this gives:

$$\mu_0 = \mathcal{O}(M_r^2) \text{ and } \mu_{sx} = \mu_{sy} = 0,\ \mu_{sz} = \mathcal{O}(1) \implies \mathcal{E}_{u_n,s} = \mathcal{O}(M_r^4),\ \mathcal{E}_{u_n,\tau} = \mathcal{O}(M_r^2). \tag{6.66}$$

$$\mu_{u_n \pm a} = \mathcal{O}(M_r) \implies \mathcal{E}_{u_n \pm a} = \mathcal{O}(M_r). \tag{6.67}$$

This implies that the overall discrete entropy production scales as M_r, that is one order of magnitude above what is expected (**E.1**). This can explain the accuracy degradation observed.

With the flux-preconditioner of Turkel, we have $p = \mathcal{O}(M_r)$, $a_p, u_{np} = \mathcal{O}(1)$ and $K_1, K_2, K_1 - K_2 = \mathcal{O}(M_r)$, therefore:

$$\mu_{u_{np} \pm a_p} = \mathcal{O}(1) \implies \mathcal{E}_{u_{np} \pm a_p} = \mathcal{O}(M_r^2).$$

That is the correct scaling. Hence the consistent behavior.

With the preconditioner of Miczek, we have $p = \mathcal{O}(1/M_r)$, $a_p = \mathcal{O}(1/M_r^2)$, $K_2 =$

$\mathcal{O}(M_r)$ and $K_1 = \mathcal{O}(1/M_r)$ because its denominator writes:

$$\begin{aligned} M_r a_p - ap &= M_r \sqrt{(p^2+1)a^2/M_r^2 - p^2 u_n^2} - ap \\ &= ap\left(\sqrt{1 + 1/p^2 - M_r^2 u_n^2/a^2} - 1\right) = \mathcal{O}(M_r). \end{aligned}$$

Therefore:

$$\begin{gathered} \mu_{u_n+a_p} = \mathcal{O}(M_r^2),\ \mu_{u_n-a_p} = \mathcal{O}(M_r^4) \implies \mathcal{E}^S_{u_n+a_p} = \mathcal{O}(M_r^2),\ \mathcal{E}^S_{u_n-a_p} = \mathcal{O}(M_r^4), \\ \delta_p = \mathcal{O}(1/M_r^4) \implies \Delta\mathcal{E}_p = \mathcal{O}(M_r^2). \end{gathered}$$

Here again, the discrete entropy production has the correct scaling.

Acoustic limit

For the sound wave configuration, density, velocity and pressure gradients are of order M_r. The specific entropy is constant. At the discrete level, this translates into:

$$\begin{aligned} &[\rho] = \mathcal{O}(M_r), [p] = \mathcal{O}(M_r),\ [u] = \mathcal{O}(1),\ [s] = 0, \\ &\qquad\qquad \implies \left[\frac{\rho k}{p}\right] = \mathcal{O}(1),\ \left[\frac{\rho u}{p}\right] = \mathcal{O}(1),\ \left[\frac{\rho}{p}\right] = \mathcal{O}(M_r). \end{aligned} \tag{6.68}$$

For the classic entropy-stable upwind dissipation, this gives:

$$\mu_0 = \mathcal{O}(M_r^3) \text{ and } \mu_{sx} = \mu_{sy} = \mu_{sz} = 0 \implies \mathcal{E}_{u_n,s} = \mathcal{O}(M_r^6),\ \mathcal{E}_{u_n,\tau} = 0. \tag{6.69}$$

$$\mu_{u_n \pm a} = \mathcal{O}(M_r) \implies \mathcal{E}_{u_n \pm a} = \mathcal{O}(M_r). \tag{6.70}$$

This implies that the overall discrete entropy production scales as M_r, in agreement with (**E.2**).

With Turkel's preconditioner, we have:

$$\mu_{u_{np} \pm a_p} = \mathcal{O}(1/M_r) \implies \mathcal{E}_{u_{np} \pm a_p} = \mathcal{O}(1),$$

meaning that entropy fluctuations will be one order of magnitude stronger than what is expected. This can explain why sound waves are severely damped with this flux.

With Miczek's preconditioner, we have:

$$\mu_{u_n + a_p} = \mathcal{O}(M_r^2),\ \mu_{u_n - a_p} = \mathcal{O}(M_r^3)$$
$$\implies \mathcal{E}^S_{u_n + a_p} = \mathcal{O}(M_r^2),\ \mathcal{E}^S_{u_n - a_p} = \mathcal{O}(M_r^2),\ \Delta\mathcal{E}_p = \mathcal{O}(M_r)$$

The discrete entropy production is one order of magnitude weaker than what is expected. This can explain why the sound wave is less damped than with the ES Roe flux.

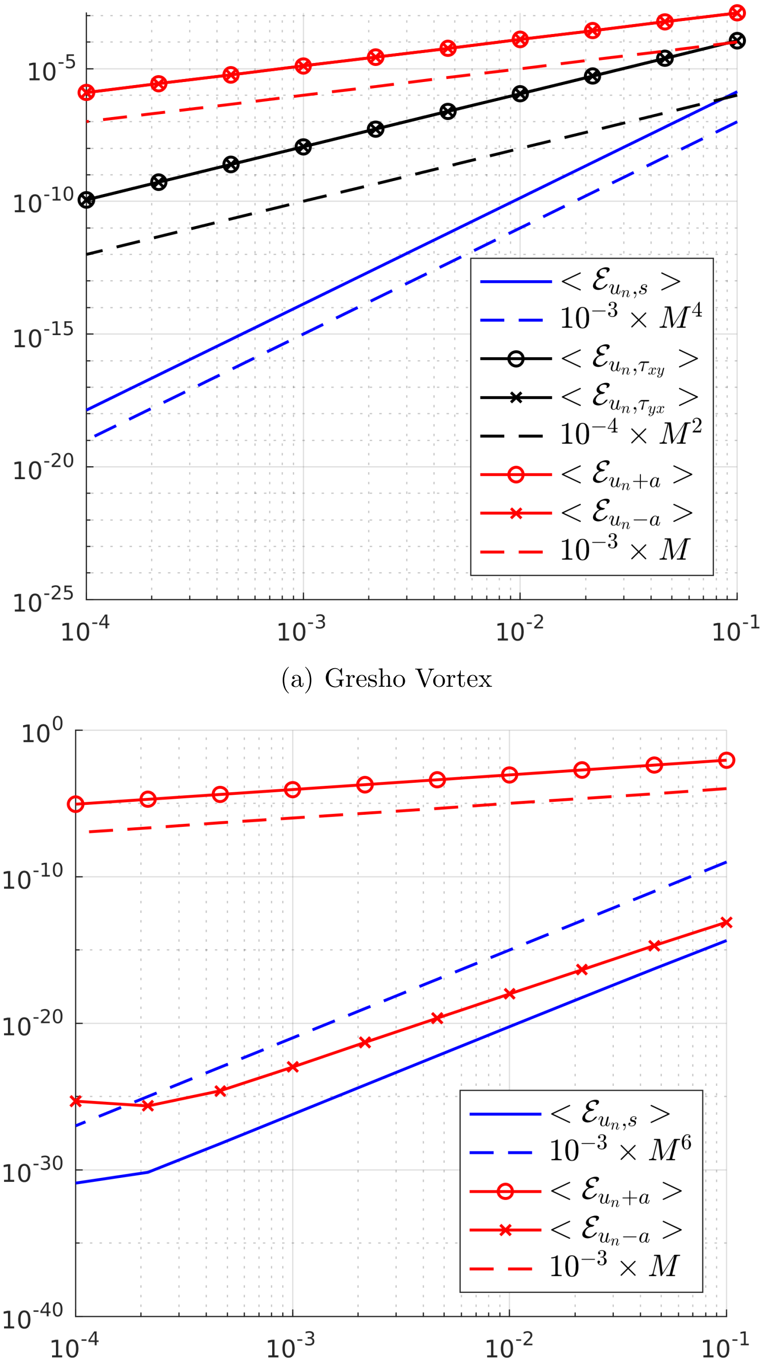
6.5.3 Connections with other low-Mach fixes

Several alternatives to the flux-preconditioning technique have been proposed, some of which are discussed in Guillard & Nkonga [228] and can be broken down into two categories.

There are corrections based on the idea that the excessive dissipation in the low Mach limit is due to the acoustic eigenvalues scaling as $\mathcal{O}(1/M_r)$ and therefore becoming infinitely large. Li & Gu [232, 233] introduced an all-speed Roe-type scheme where the eigenvalues are modified as:

$$u_n \pm (a/M_r) \ \rightarrow \ u_n \pm f(M_r)(a/M_r),$$

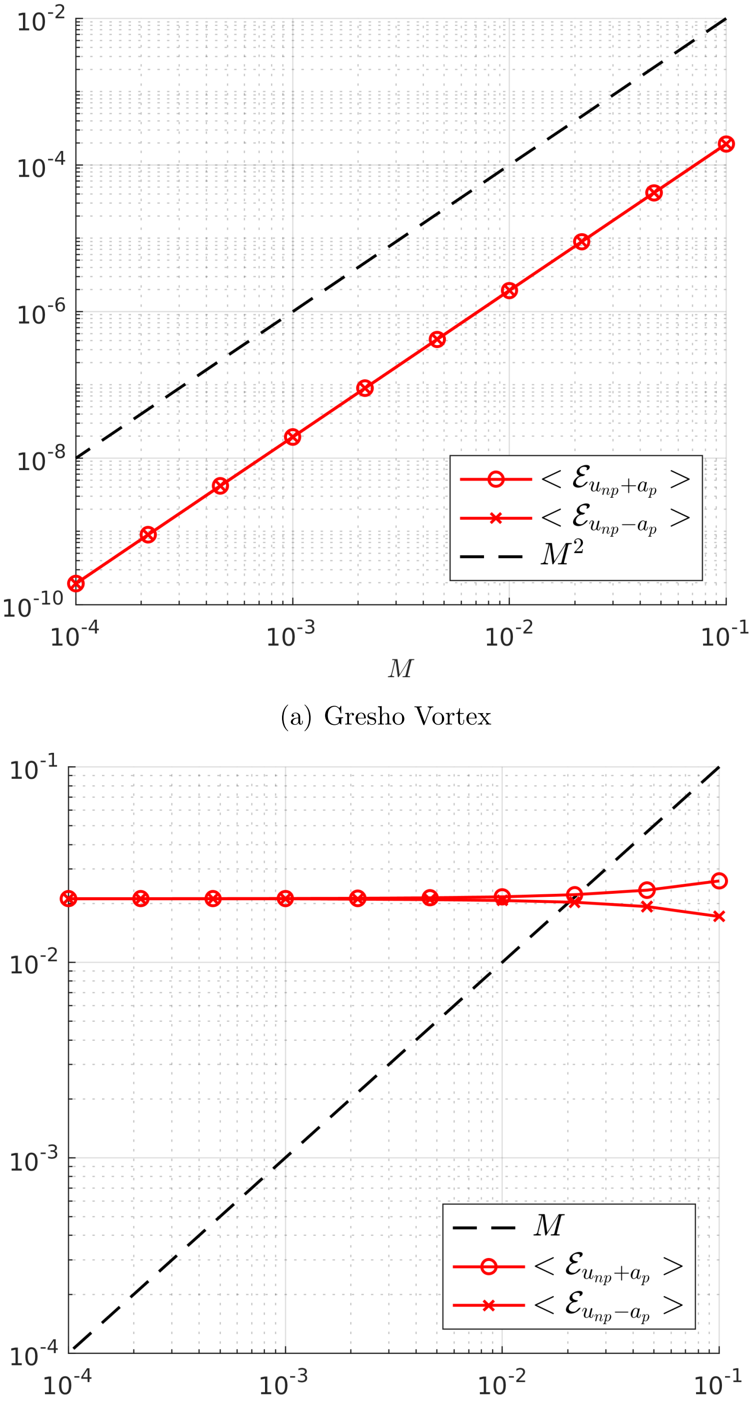
and $f(M_r)$ is a correction introduced so that $f(M_r)(a/M_r)$ is bounded in the low Mach limit. This approach can be easily applied in our setting and it is easy to show



(a) Gresho Vortex

(b) Sound wave

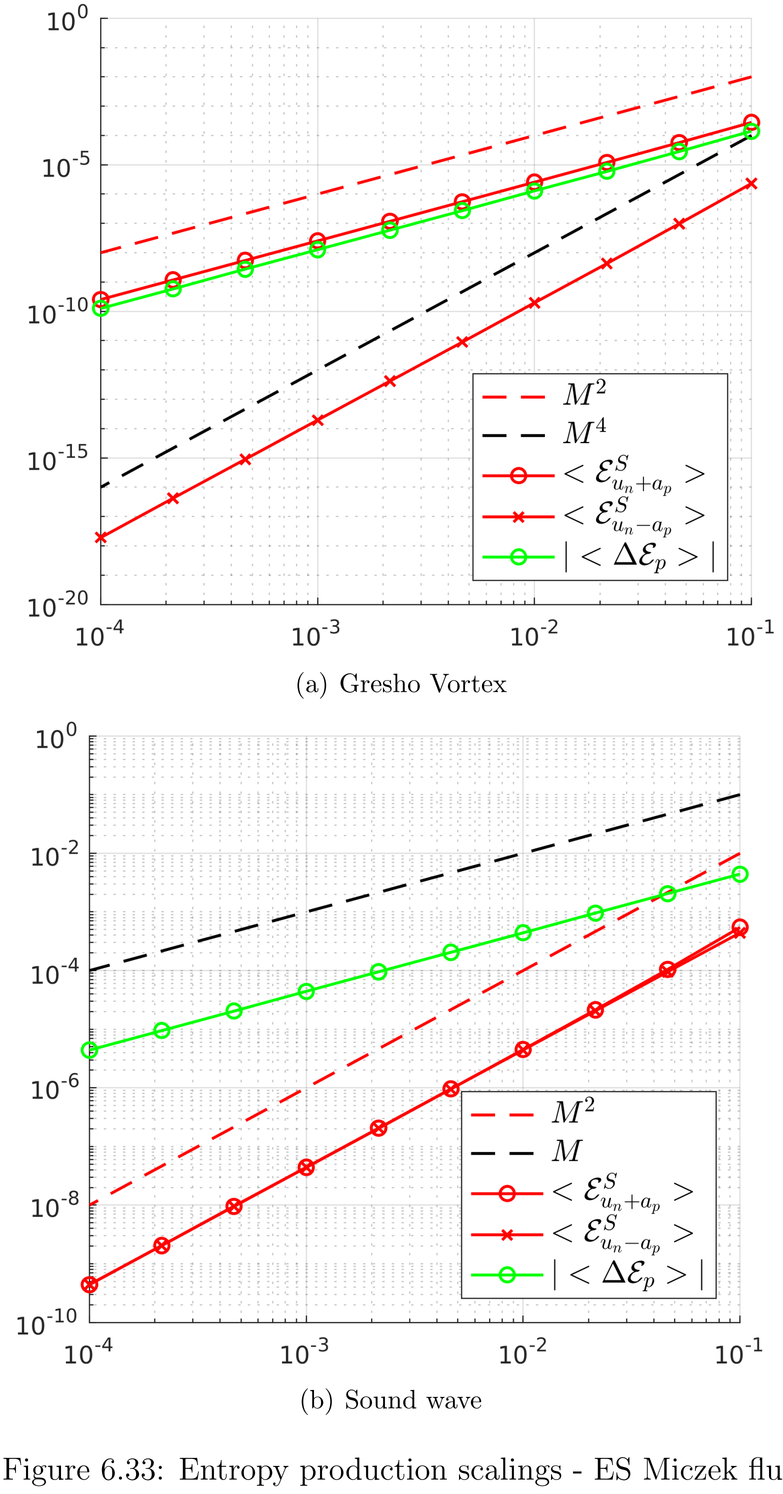
Figure 6.31: Entropy production scalings - ES Roe flux



(a) Gresho Vortex

(b) Sound wave

Figure 6.32: Entropy production scalings - ES Turkel flux



(a) Gresho Vortex

(b) Sound wave

Figure 6.33: Entropy production scalings - ES Miczek flux

that it will lead to a discrete entropy production that has appropriate Mach number scaling.

Starting with Thornber *et al.* [235, 236], a number of fixes [238, 237, 239] consisting in modifying the jump terms in the normal velocity $[u_n]$ have been proposed. By and large, they multiply $[u_n]$ by a correction term of order M_r. These fixes are motivated in part by the work of Birken & Meister [234], who showed that the flux-preconditioner of Turkel enforce a more stringent (by a factor M_r) CFL condition. The same result was proved by Barsukow *et al.* [225] for the flux-preconditioner of Miczek.

For the ES Roe flux, the acoustic part $D^A[\mathbf{v}]$ of the dissipation operator writes:

$$D^A[\mathbf{v}] = \begin{bmatrix} \mathbf{r}_{u_n+a} & \mathbf{r}_{u_n-a} \end{bmatrix} \begin{bmatrix} |u_n + (a/M_r)| & 0 \\ 0 & |u_n - (a/M_r)| \end{bmatrix} \begin{bmatrix} \mu_{u_n+a} \\ \mu_{u_n-a} \end{bmatrix} \tag{6.71}$$

where

$$\mu_{u_n \pm a} = K_a \left(\mu_0 - h \left[\frac{\rho}{p} \right] \pm M_r a \overline{\left(\frac{\rho}{p} \right)} [u_n] \right).$$

Let $\tilde{\mu}_{u_n \pm a}$ be the wave strength obtained after applying the correction $[u_n] \to M_r[u_n]$:

$$\tilde{\mu}_{u_n \pm a} = K_a \left(\mu_0 - h \left[\frac{\rho}{p} \right] \pm M_r^2 a \overline{\left(\frac{\rho}{p} \right)} [u_n] \right).$$

The resulting acoustic entropy production field $\tilde{\mathcal{E}}^A$ becomes:

$$\tilde{\mathcal{E}}^A = [\mathbf{v}]^T D^A[\mathbf{v}] = |u_n + (a/M_r)| \mu_{u_n+a} \tilde{\mu}_{u_n+a} \; + \; |u_n - (a/M_r)| \mu_{u_n-a} \tilde{\mu}_{u_n-a}.$$

It is not clear whether $\tilde{\mathcal{E}}^A > 0$, which we would need to maintain entropy-stability. In the very least, it can be showed that in the incompressible limit, $\tilde{\mathcal{E}}^A = \mathcal{O}(M_r^2)$, in agreement with (**E.1**).

6.6 Discussion

6.6.1 The origin of the skew-symmetric term

Given that the Miczek flux-preconditioner was constructed so that $P^{-1}|PA|$ has the same scaling as A, the appearance of a skew-symmetric term in the scaled form of the Miczek ES dissipation operator could be explained by examining the acoustic entropy production field *without upwinding*, that is with Λ instead of $|\Lambda|$. We have:

$$\mathcal{E}_{u_n+a} = \mu^2_{u_n+a}(u_n + (a/M_r)),\ \mathcal{E}_{u_n-a} = \mu^2_{u_n-a}(u_n - (a/M_r)). \tag{6.72}$$

If we assume $\mu^2_{u_n+a} \approx \mu^2_{u_n-a} \approx \mu_{u_n+a}\mu_{u_n-a}$, then we can write something similar to (6.62):

$$\mathcal{E}_{u_n+a} \approx \mathcal{E}^S_{u_n+a} + \Delta\mathcal{E}_a,\ \mathcal{E}_{u_n-a} \approx \mathcal{E}^S_{u_n-a} - \Delta\mathcal{E}_a, \tag{6.73}$$

where:

$$\mathcal{E}^S_{u_n+a} = \mu^2_{u_n+a}u_n,\ \Delta\mathcal{E}_a = \mu_{u_n+a}\mu_{u_n-a}(a/M_r),\ \mathcal{E}^S_{u_n-a} = \mu^2_{u_n-a}u_n.$$

Of course, the resulting dissipation operator is no longer guaranteed to be entropy stable. The point is that the entropy production balance between acoustic fields described by equation (6.73) might be what the skew-symmetric matrix Δ_p of the ES Miczek flux tries to reproduce, as a consequence of the requirement that $P^{-1}|PA|$ should behave like A in the low Mach limit. Whether recovering this balance is key in ensuring a good low Mach behavior is a different story. The numerical results advise against it.

6.6.2 A simple equivalent to the ES Miczek flux

Consider the dissipation operator $D_P = R(|\Lambda_p| + \Delta_p)R^T$ where:

$$|\Lambda_p| = \begin{bmatrix} |u_n| & 0 & 0 & 0 & 0 \\ 0 & |u_n| & 0 & 0 & 0 \\ 0 & 0 & |u_n| & 0 & 0 \\ 0 & 0 & 0 & f_1|u_n + (a/M_r)| & 0 \\ 0 & 0 & 0 & 0 & f_2|u_n - (a/M_r)| \end{bmatrix},$$

$$\Delta_p = g \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & +\delta_p \\ 0 & 0 & 0 & -\delta_p & 0 \end{bmatrix}, \tag{6.74}$$

and f_1, f_2 and g are factors. The eigenvectors R are left untouched. Similarly to (6.62) and (6.73), we have:

$$\mathcal{E}_{u_n-a} = \mathcal{E}^S_{u_n-a} - g\Delta\mathcal{E}_p,\ \mathcal{E}_{u_n+a} = \mathcal{E}^S_{u_n+a} + g\Delta\mathcal{E}_p$$
$$\mathcal{E}^S_{u_n-a} = \mu^2_{u_n-a} f_2|u_n - a|,\ \Delta\mathcal{E}_p = \delta_p \mu_{u_n-a}\mu_{u_n+a},\ \mathcal{E}^S_{u_n+a} = \mu^2_{u_n+a} f_1|u_n + a|).$$

This dissipation operator is ES as long as $f_1, f_2 \geq 0$, and we can emulate equation (6.73) by taking $f_1 = |u_n|/|u_n + a/M|$, $f_2 = |u_n|/|u_n - a/M|$, $g = 1$ and $\delta_p = a/M$. This gives:

$$\mathcal{E}^S_{u_n-a} = \mu^2_{u_n-a}|u_n|,\ \Delta\mathcal{E}_p = \mu_{u_n-a}\mu_{u_n+a}(a/M_r),\ \mathcal{E}^S_{u_n+a} = \mu^2_{u_n+a}|u_n|.$$

In the incompressible and acoustic limits, we have:

$$\mathcal{E}^S_{u_n-a} = \mathcal{O}(M_r^2),\ \Delta\mathcal{E}_p = \mathcal{O}(M_r),\ \mathcal{E}^S_{u_n+a} = \mathcal{O}(M_r^2),$$

which meets the low Mach requirements (**E.1**) and (**E.2**). This dissipation operator also meets Miczek's requirement that the dissipation matrix should have exactly the same Mach number scalings as A. Indeed, we have:

$$R_{\mathbf{z}}(|\Lambda_p| + \Delta_p)R_{\mathbf{z}}^{-1} = \begin{bmatrix} |u_n| & -n_x a/M_r & -n_y a/M_r & -n_z a/M_r & 0 \\ n_x a/M_r & |u_n| & 0 & 0 & 0 \\ n_y a/M_r & 0 & |u_n| & 0 & 0 \\ n_z a/M_r & 0 & 0 & |u_n| & 0 \\ 0 & 0 & 0 & 0 & |u_n| \end{bmatrix}.$$

To have $D_P \to R|\Lambda|R^T[\mathbf{v}]$ as $M_r \to 1$, we can set $f_1 = M_r + (1-M_r)|u_n|/|u_n + (a/M_r)|$, $f_2 = M_r + (1-M_r)|u_n|/|u_n - (a/M_r)|$, and $g = 1 - M_r$ for instance.

Numerical results (figures 6.34, 6.35, 6.36 and 6.37) show that this "skewed" dissipation operator behaves just like the ES Miczek dissipation operator for both the Gresho vortex and the sound wave.

The skewed ES dissipation operator (6.6.2) has not been introduced to compete with Miczek's flux or any of the aforementioned schemes (we remind the reader that the best results were observed with an EC flux in space). We introduced it to assess our intuition that skew-symmetric dissipation operators change the way entropy is locally produced.

In appendix C, we show that the skewed dissipation operator induces a $\mathcal{O}(M_r^2)$ CFL condition, like the Turkel [234] and Miczek [225] dissipation operators.

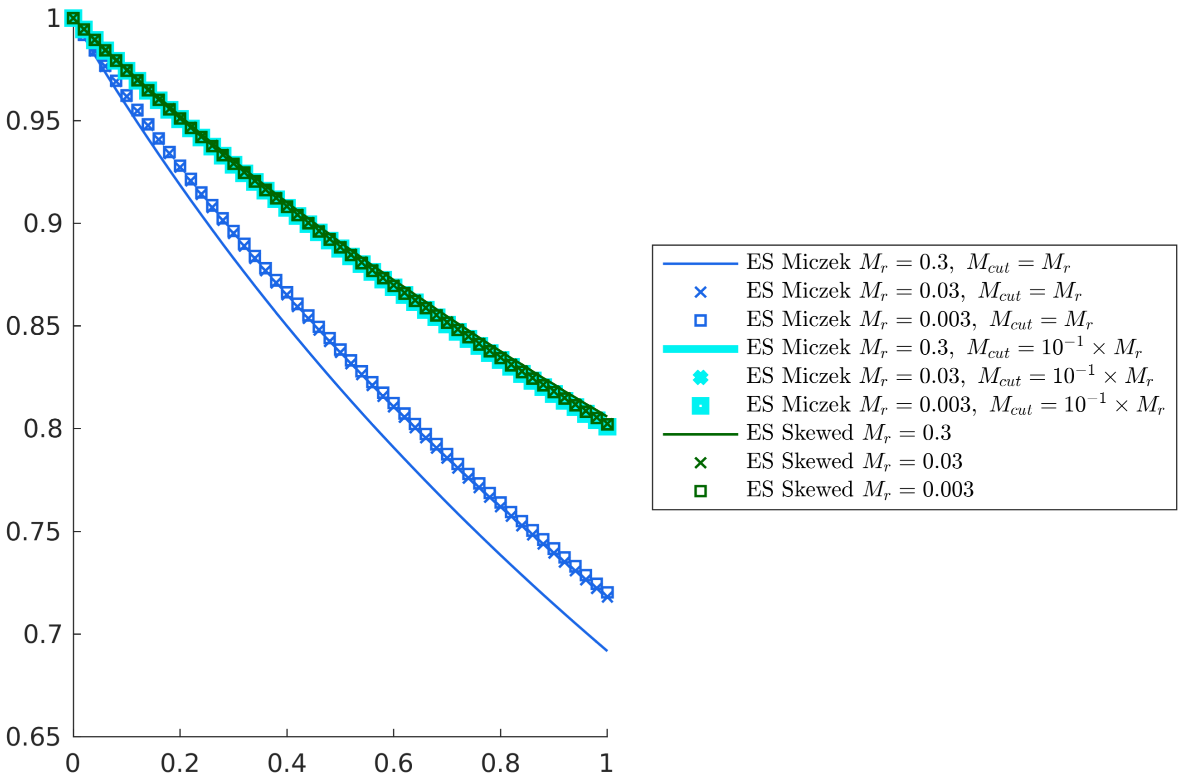
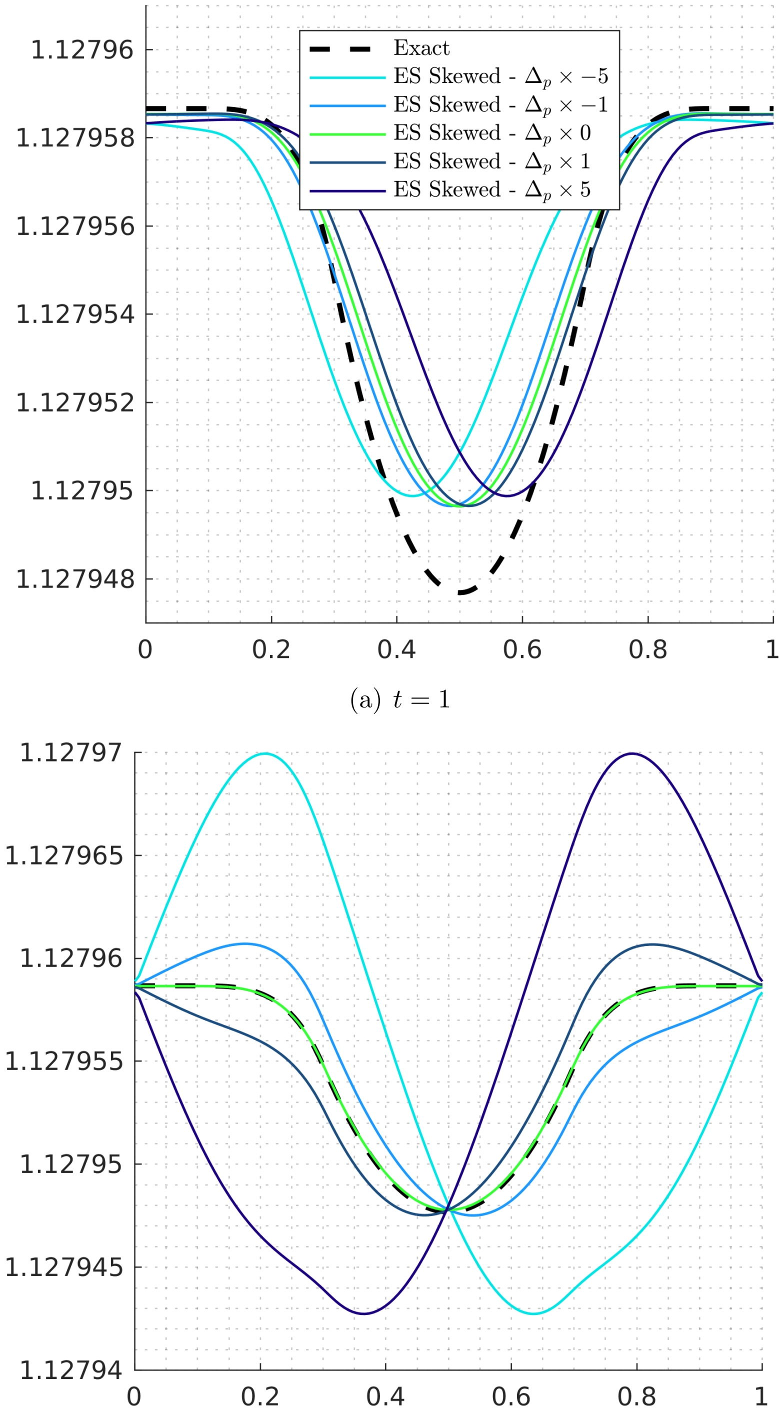


Figure 6.34: Gresho vortex: Total kinetic energy k/k_0 evolution over time at different Mach numbers. The skewed ES flux and the Miczek ES flux (with two cut-off Mach numbers) are compared.



(a) $t = 1$

(b) $t = 0.04$

Figure 6.35: Gresho Vortex: Pressure field at $M_r = 3 \times 10^{-3}$ for the skewed ES flux when the skew-symmetric term is multiplied by different factors.

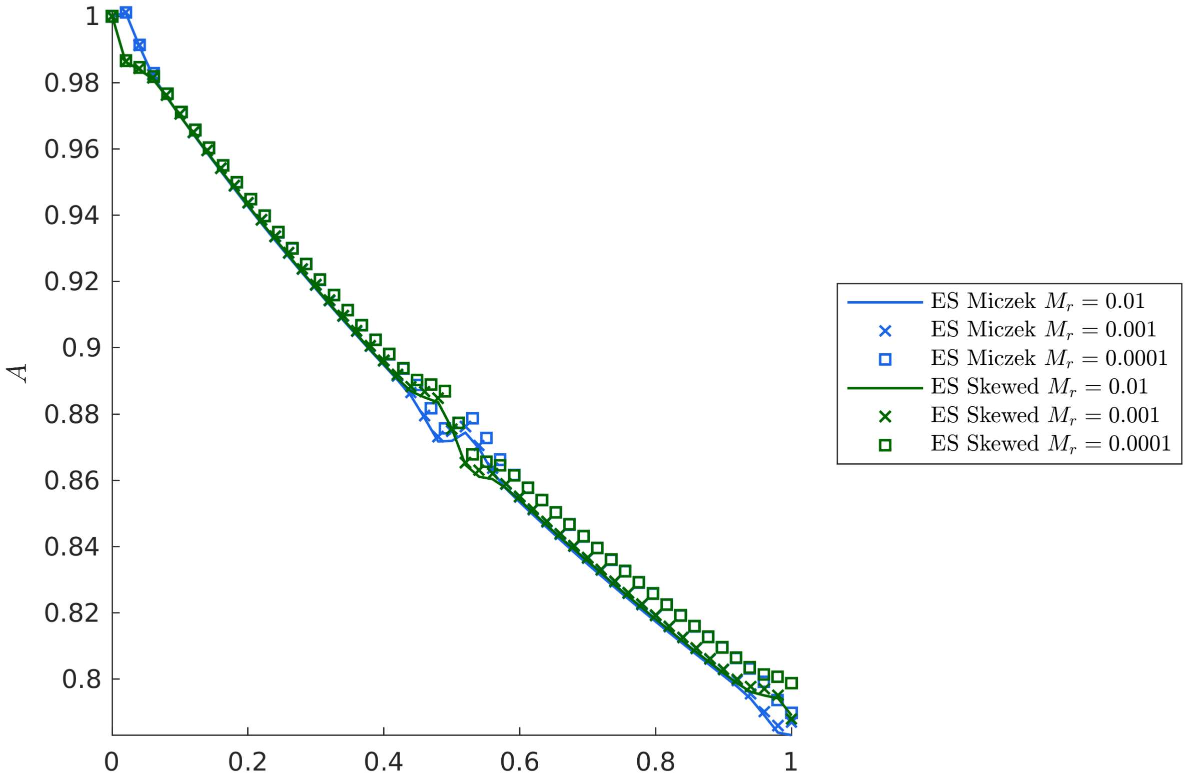
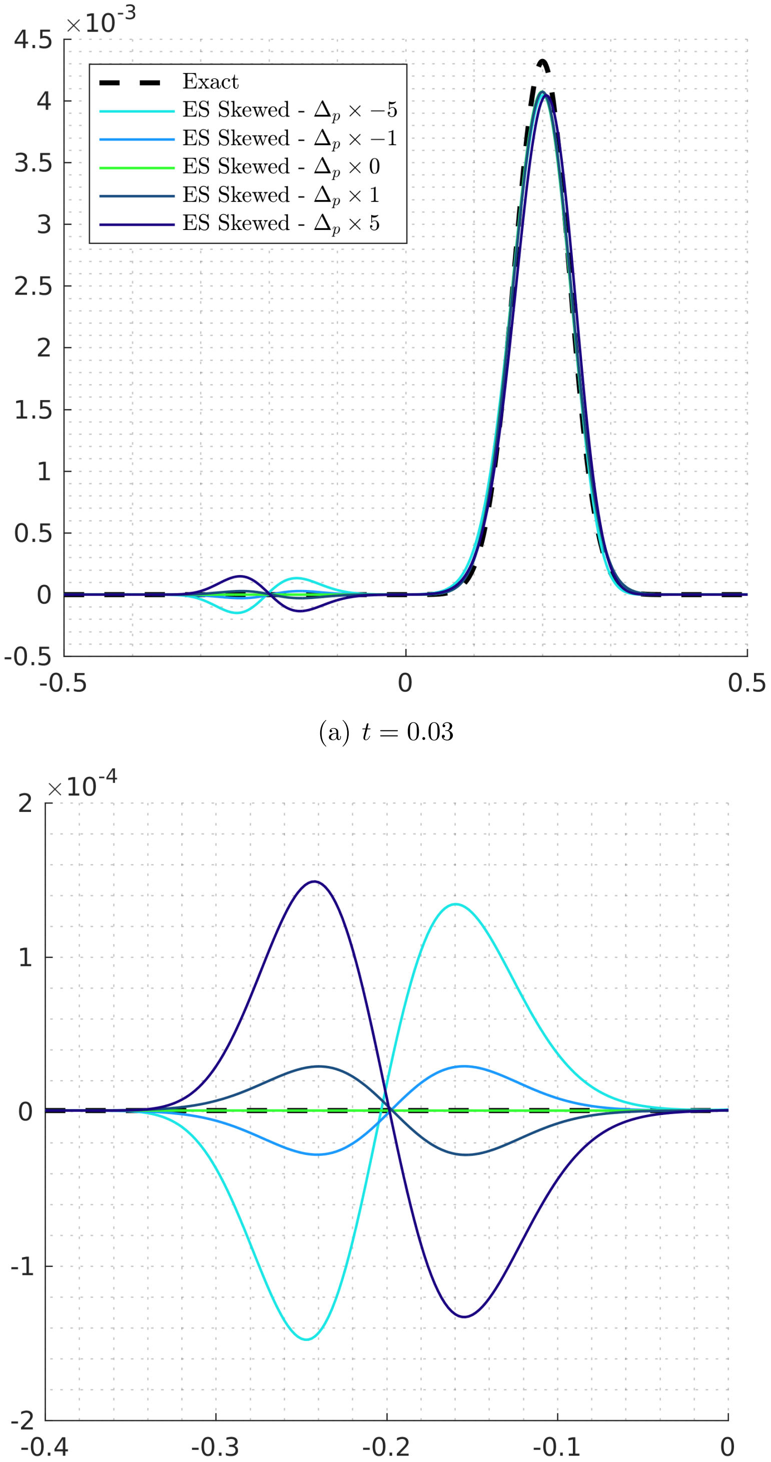


Figure 6.36: Sound wave: Normalized amplitude evolution for the skewed ES and Miczek ES ($M_{cut} = M_r$, lowering it did not improve the results) fluxes at different Mach numbers.



(a) $t = 0.03$

(b) $t = 0.03$, zoomed

Figure 6.37: Sound wave: Same pressure profiles as in figure 6.30 with the skewed ES flux instead. $M_r = 10^{-2}$.

6.6.3 Different ways of breaking down the entropy production

It is important to recognize that the entropy production breakdown we introduced in section 6.5.1 is not unique. Consider an ES dissipation operator consisting of two modes $R = [\mathbf{r_1},\ \mathbf{r_2}]$:

$$R|\Lambda|R^T[\mathbf{v}] = |\lambda_1|\mu_1\mathbf{r_1} + |\lambda_2|\mu_2\mathbf{r_2} \implies \mathcal{E} = [\mathbf{v}]^T R|\Lambda|R^T[\mathbf{v}] = \mathcal{E}_1 + \mathcal{E}_2,\ \mathcal{E}_i = \mu_i^2|\Lambda_i|.$$

Let $\overline{R} = \{\bar{\mathbf{r}}_\mathbf{1},\ \bar{\mathbf{r}}_\mathbf{2}\}$ be an alternative pair of modes defined by the linear mapping:

$$\begin{cases} \mathbf{r_1} = \alpha_{11}\bar{\mathbf{r}}_\mathbf{1} + \alpha_{12}\bar{\mathbf{r}}_\mathbf{2}, \\ \mathbf{r_2} = \alpha_{21}\bar{\mathbf{r}}_\mathbf{1} + \alpha_{22}\bar{\mathbf{r}}_\mathbf{2}. \end{cases} \implies \begin{cases} \mu_1 = \alpha_{11}\bar{\mu}_1 + \alpha_{12}\bar{\mu}_2, \\ \mu_2 = \alpha_{21}\bar{\mu}_1 + \alpha_{22}\bar{\mu}_2. \end{cases}$$

Then the dissipation operator can be expressed in terms of these modes:

$$\begin{aligned} R|\Lambda|R^T[\mathbf{v}] &= |\lambda_1|\mu_1\mathbf{r_1} + |\lambda_2|\mu_2\mathbf{r_2} \\ &= |\lambda_1|(\alpha_{11}\bar{\mu}_1 + \alpha_{12}\bar{\mu}_2)(\alpha_{11}\bar{\mathbf{r}}_\mathbf{1} + \alpha_{12}\bar{\mathbf{r}}_\mathbf{2}) \\ &\qquad\qquad + |\lambda_2|(\alpha_{21}\bar{\mu}_1 + \alpha_{22}\bar{\mu}_2)(\alpha_{21}\bar{\mathbf{r}}_\mathbf{1} + \alpha_{22}\bar{\mathbf{r}}_\mathbf{2}) \\ &= \big(|\lambda_1|\alpha_{11}(\alpha_{11}\bar{\mu}_1 + \alpha_{12}\bar{\mu}_2) + |\lambda_2|\alpha_{21}(\alpha_{21}\bar{\mu}_1 + \alpha_{22}\bar{\mu}_2)\big)\ \bar{\mathbf{r}}_\mathbf{1} \\ &\qquad\qquad + \big(|\lambda_1|\alpha_{12}(\alpha_{11}\bar{\mu}_1 + \alpha_{12}\bar{\mu}_2) + |\lambda_2|\alpha_{22}(\alpha_{21}\bar{\mu}_1 + \alpha_{22}\bar{\mu}_2)\big)\ \bar{\mathbf{r}}_\mathbf{2} \\ &= \big((|\lambda_1|\alpha_{11}^2 + |\lambda_2|\alpha_{21}^2)\bar{\mu}_1 + (|\lambda_1|\alpha_{11}\alpha_{12} + |\lambda_2|\alpha_{21}\alpha_{22})\bar{\mu}_2\big)\ \bar{\mathbf{r}}_\mathbf{1} \\ &\qquad\qquad + \big((|\lambda_1|\alpha_{11}\alpha_{12} + |\lambda_2|\alpha_{21}\alpha_{22})\bar{\mu}_1 + (|\lambda_1|\alpha_{12}^2 + |\lambda_2|\alpha_{22}^2)\bar{\mu}_2\big)\ \bar{\mathbf{r}}_\mathbf{2} \end{aligned}$$

In more compact notation, this writes:

$$R|\Lambda|R^T[\mathbf{v}] = |\lambda_1|\mu_1\mathbf{r_1}+|\lambda_2|\mu_2\mathbf{r_2} = \big(S_{11}\bar{\mu}_1+S_{12}\bar{\mu}_2\big)\bar{\mathbf{r}}_\mathbf{1}+\big(S_{22}\bar{\mu}_2+S_{21}\bar{\mu}_1\big)\bar{\mathbf{r}}_\mathbf{2} = \overline{R}S\overline{R}^T[\mathbf{v}],$$

where $S = \left(S_{ij}\right)_{1\leq(i,j)\leq 2}$ with:

$$S_{11} = (|\lambda_1|\alpha_{11}^2 + |\lambda_2|\alpha_{21}^2) > 0,\ S_{22} = (|\lambda_1|\alpha_{12}^2 + |\lambda_2|\alpha_{22}^2) > 0, \tag{6.75}$$

$$S_{12} = S_{21} = (|\lambda_1|\alpha_{11}\alpha_{12} + |\lambda_2|\alpha_{21}\alpha_{22}). \tag{6.76}$$

This gives the alternative decomposition:

$$\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2 = \overline{\mathcal{E}}_1 + \overline{\mathcal{E}}_2,\ \overline{\mathcal{E}}_i = \left(\sum_j S_{ij}\bar{\mu}_j\right)\bar{\mu}_i. \tag{6.77}$$

Unlike the initial decomposition, each individual field $\overline{\mathcal{E}}_i$ is no longer guaranteed to be positive in principle. This does not matter much given that the sign of $\mathcal{E}$ will not change. We also see that depending on the choice of modes, coupling terms within each field $\overline{\mathcal{E}}_i$ may appear. We can write:

$$\overline{\mathcal{E}}_i = \overline{\mathcal{E}}_i^S + \sum_{j\neq i}\overline{\mathcal{E}}_{ij},\ \overline{\mathcal{E}}_i^S = S_{ii}\bar{\mu}_i^2,\ \overline{\mathcal{E}}_{ij} = S_{ij}\bar{\mu}_j\bar{\mu}_i. \tag{6.78}$$

Note that $\overline{\mathcal{E}}_i^S > 0$ hence the coupling terms, just like the skew-symmetric terms, can, if needed, be discarded without losing entropy-stability. Also note that the coupling terms do not cancel each other, i.e. $\overline{\mathcal{E}}_{ij} + \overline{\mathcal{E}}_{ji} = 0,\ j \neq i$. In fact, they are equal $\overline{\mathcal{E}}_{ij} = \overline{\mathcal{E}}_{ji}$.

If the dissipation operator has a skew-symmetric component, which we can write:

$$R\Delta R^T[\mathbf{v}] = \delta_p(\mu_2\mathbf{r_1} - \mu_1\mathbf{r_2}) \implies [\mathbf{v}]^T R\Delta R^T[\mathbf{v}] = \Delta\mathcal{E} - \Delta\mathcal{E} = 0,\ \Delta\mathcal{E} = \delta\mu_1\mu_2,$$

we can also rewrite it in terms of $\{\bar{\mathbf{r}}_{\mathbf{1}},\ \bar{\mathbf{r}}_{\mathbf{2}}\}$ instead:

$$
\begin{aligned}
R\Delta R^T[\mathbf{v}] &= \delta(\mu_2\mathbf{r}_{\mathbf{1}} - \mu_1\mathbf{r}_{\mathbf{2}}) \\
&= \delta\big((\alpha_{21}\bar{\mu}_1 + \alpha_{22}\bar{\mu}_2)(\alpha_{11}\bar{\mathbf{r}}_{\mathbf{1}} + \alpha_{12}\bar{\mathbf{r}}_{\mathbf{2}}) - (\alpha_{11}\bar{\mu}_1 + \alpha_{12}\bar{\mu}_2)(\alpha_{21}\bar{\mathbf{r}}_{\mathbf{1}} + \alpha_{22}\bar{\mathbf{r}}_{\mathbf{2}}\big) \\
&= \delta(\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21})\big(\bar{\mu}_2\bar{\mathbf{r}}_{\mathbf{1}} - \bar{\mu}_1\bar{\mathbf{r}}_{\mathbf{2}}\big) \\
&= \bar{\delta}\big(\bar{\mu}_2\bar{\mathbf{r}}_{\mathbf{1}} - \bar{\mu}_1\bar{\mathbf{r}}_{\mathbf{2}}\big) \\
&= \overline{R}\ \overline{\Delta}\ \overline{R}^T[\mathbf{v}]
\end{aligned}
$$

If the mapping is one-to-one, the skew-symmetric operator $R\Delta R^T[\mathbf{v}]$ which transfers entropy between modes $\{\mathbf{r}_{\mathbf{1}},\ \mathbf{r}_{\mathbf{2}}\}$ is equivalent to a skew-symmetric operator $\overline{R}\ \overline{\Delta}\ \overline{R}^T[\mathbf{v}]$ which transfers entropy between modes $\{\bar{\mathbf{r}}_{\mathbf{1}},\ \bar{\mathbf{r}}_{\mathbf{2}}\}$:

$$
[\mathbf{v}]^T\overline{R}\ \overline{\Delta}\ \overline{R}^T[\mathbf{v}] = \overline{\Delta\mathcal{E}} - \overline{\Delta\mathcal{E}} = 0,\ \overline{\Delta\mathcal{E}} = \bar{\delta}\bar{\mu}_1\bar{\mu}_2.
$$

We can now rewrite the entropy production breakdowns of the ES Turkel and ES Miczek fluxes in terms of the original acoustic eigenvectors instead of the modified ones. For Turkel's flux-preconditioner, we map from $\{\mathbf{r}_{\mathbf{1}},\ \mathbf{r}_{\mathbf{2}}\} = \{\mathbf{r}_{u_{np}+a_p},\ \mathbf{r}_{u_{np}-a_p}\}$ to $\{\bar{\mathbf{r}}_{\mathbf{1}},\ \bar{\mathbf{r}}_{\mathbf{2}}\} = \{\mathbf{r}_{u_n+a},\ \mathbf{r}_{u_n-a}\}$ following:

$$
\begin{aligned}
\mathbf{r}_{u_{np}+a_p} &= \frac{1}{\sqrt{2}p^2(K_1-K_2)}\Big((K_1+p^2)\mathbf{r}_{u_n+a} + (K_1-p^2)\mathbf{r}_{u_n-a}\Big), \\
\mathbf{r}_{u_{np}-a_p} &= \frac{1}{\sqrt{2}p^2(K_1-K_2)}\Big((K_2+p^2)\mathbf{r}_{u_n+a} + (K_2-p^2)\mathbf{r}_{u_n-a}\Big).
\end{aligned}
$$

The new decomposition (6.77)-(6.78) writes:

$$
\mathcal{E}_{u_{np}+a_p} + \mathcal{E}_{u_{np}-a_p} = \big(\mathcal{E}^S_{u_n+a} + \mathcal{E}_{u_n+a,u_n-a}\big) + \big(\mathcal{E}^S_{u_n-a} + \mathcal{E}_{u_n-a,u_n+a}\big). \qquad (6.79)
$$

The left-hand side term recalls the previous breakdown (equation (6.58)) along the modified acoustic eigenvectors. The right-hand side term is the new breakdown along the acoustic eigenvectors. Figures 6.38 and 6.39 show the initial entropy production fields for the Gresho vortex and the sound wave using the decomposition (6.79). These entropy production fields are more similar to those of the classic ES Roe flux (figures 6.15 and 6.17) than the ones along the modified acoustic eigenvectors (figures 6.19 and 6.21). We also see that the coupling term $\mathcal{E}_{u_n+a,u_n-a}$ can be negative.

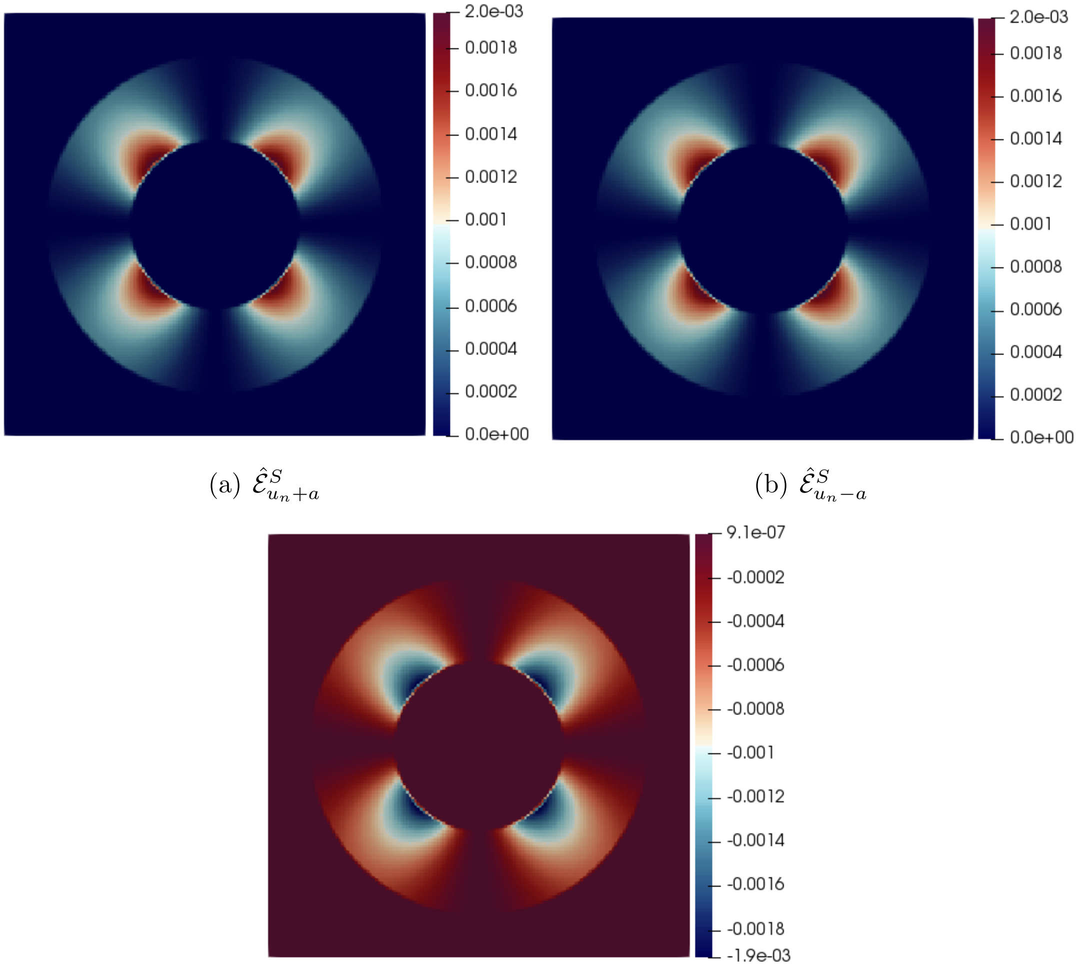
For Miczek's flux-preconditioner, using:

$$\begin{aligned}
\mathbf{r}_{u_n+a_p} &= \frac{1}{\sqrt{2}(p^2+1)(K_1-K_2)}\Bigg(\big(K_1(1+p)+1-p\big)\mathbf{r}_{u_n+a}+ \\
&\qquad\qquad\qquad\qquad\qquad\big(K_1(1-p)-(1+p)\big)\mathbf{r}_{u_n-a}\Bigg), \\
\mathbf{r}_{u_n-a_p} &= \frac{1}{\sqrt{2}(p^2+1)(K_1-K_2)}\Bigg(\big(K_2(1+p)+1-p\big)\mathbf{r}_{u_n+a}+ \\
&\qquad\qquad\qquad\qquad\qquad\big(K_2(1-p)-(1+p)\big)\mathbf{r}_{u_n-a}\Bigg).
\end{aligned}$$

The new decomposition (6.77)-(6.78) writes:

$$\begin{aligned}
&\big(\mathcal{E}^S_{u_n+a_p}-\Delta\mathcal{E}_p\big)+\big(\mathcal{E}^S_{u_n-a_p}+\Delta\mathcal{E}_p\big)= \\
&\qquad\big(\mathcal{E}^S_{u_n+a}+\mathcal{E}_{u_n+a,u_n-a}-\Delta\mathcal{E}_a\big)+\big(\mathcal{E}^S_{u_n-a}+\mathcal{E}_{u_n-a,u_n+a}+\Delta\mathcal{E}_a\big). \quad (6.80)
\end{aligned}$$

The left-hand side term recalls the previous breakdown (equations (6.61) and (6.62)) along the modified acoustic eigenvectors. The right-hand side term is the new breakdown along the acoustic eigenvectors. Figures 6.40 and 6.41 show the initial entropy production fields for the Gresho vortex and the sound wave using the decomposition (6.80). The coupling term $\mathcal{E}_{u_n+a,u_n-a}$ is negligible. For the sound wave, figure 6.42 shows the integrated entropy production fields according to (6.80). The culpability of the skew-symmetric part of the ES Miczek flux in the spurious behavior is more apparent than with the previous decomposition (figures 6.28-(a) and 6.29-(a)).

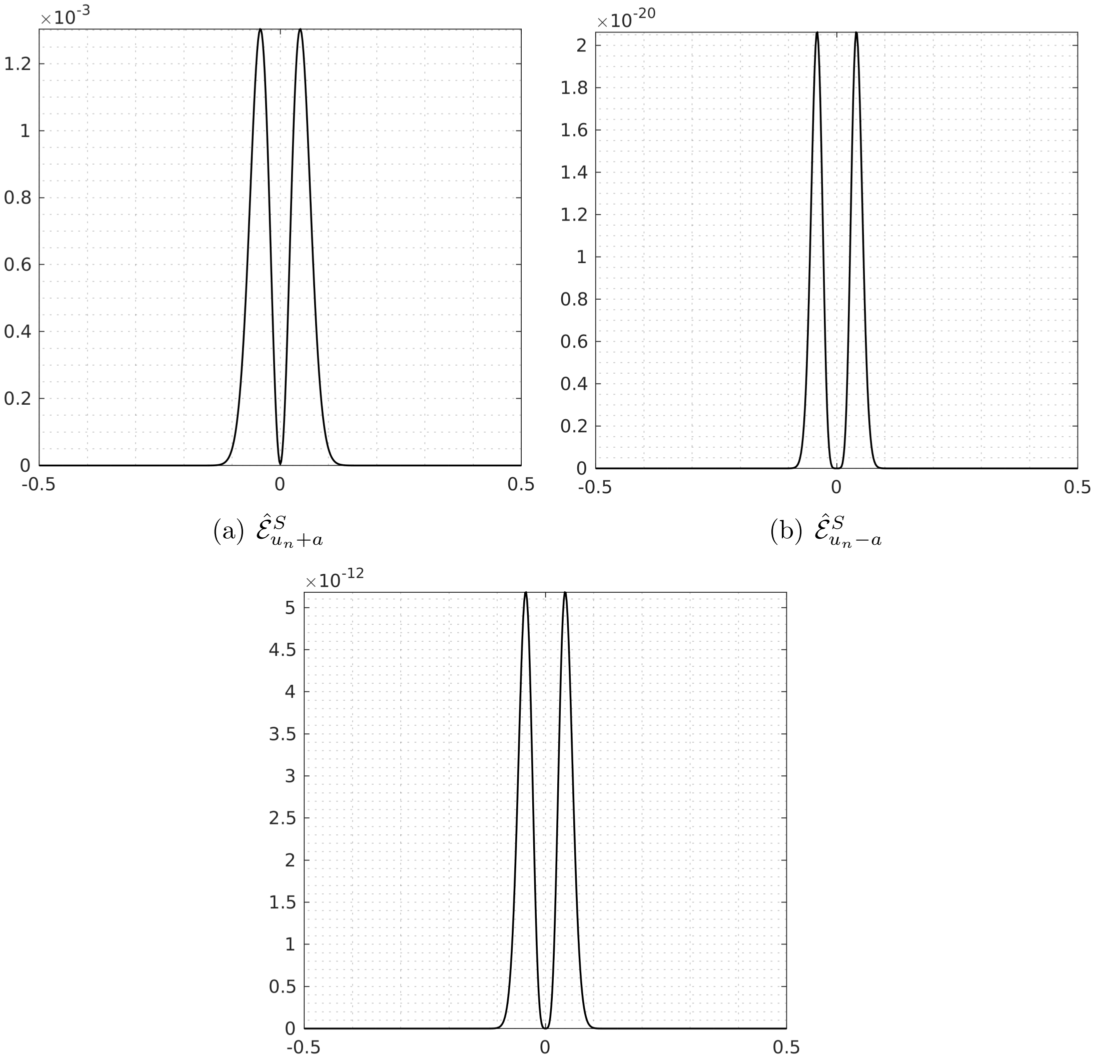


(a) $\hat{\mathcal{E}}^{S}_{u_n+a}$

(b) $\hat{\mathcal{E}}^{S}_{u_n-a}$

(c) $\hat{\mathcal{E}}_{u_n-a,u_n+a}$

Figure 6.38: Gresho Vortex: Entropy production fields along the acoustic eigenvectors at $t = 0$ for the ES Turkel flux. $M_r = 3 \times 10^{-2}$.

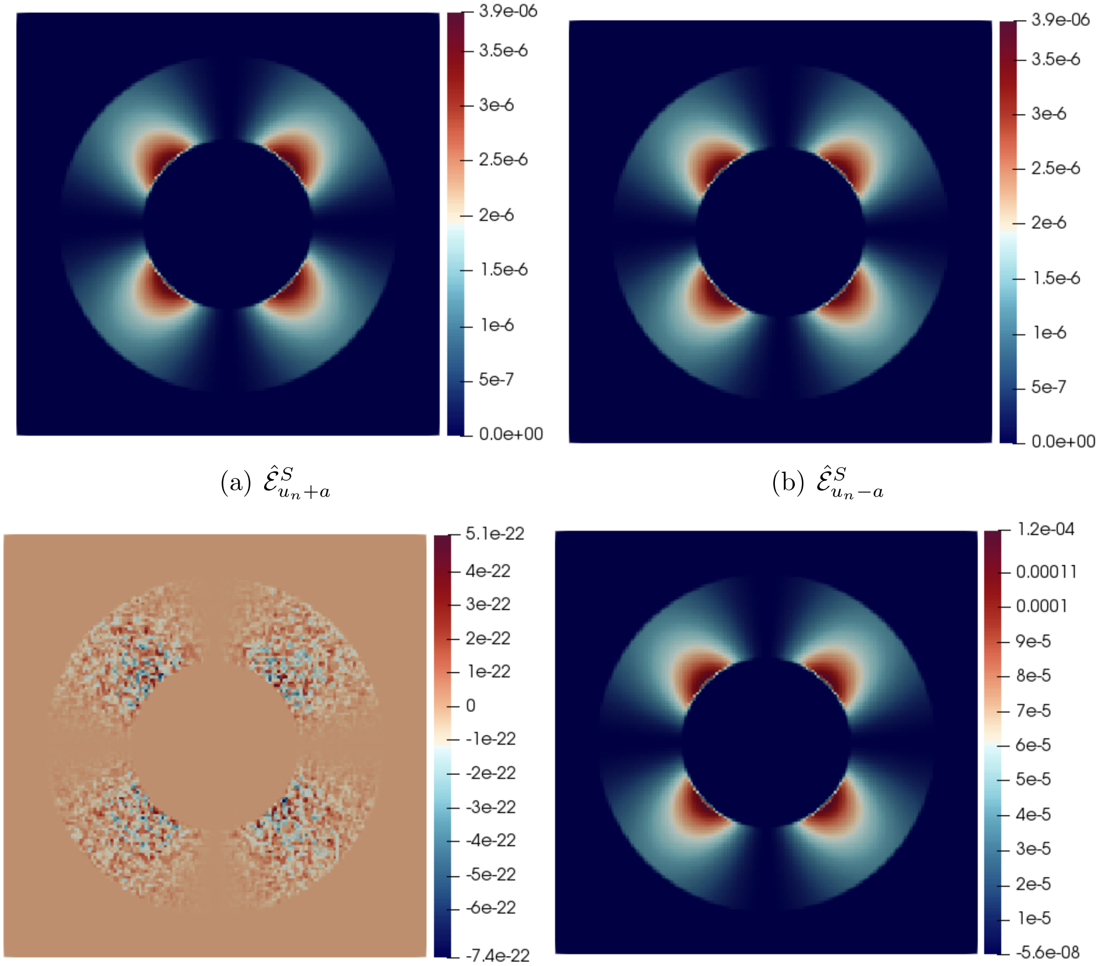


(a) $\hat{\mathcal{E}}^S_{u_n+a}$

(b) $\hat{\mathcal{E}}^S_{u_n-a}$

(c) $\hat{\mathcal{E}}_{u_n-a,u_n+a}$

Figure 6.39: Sound wave: Entropy production fields along the acoustic eigenvectors at $t = 0$ for the ES Turkel flux $M_r = 10^{-2}$.



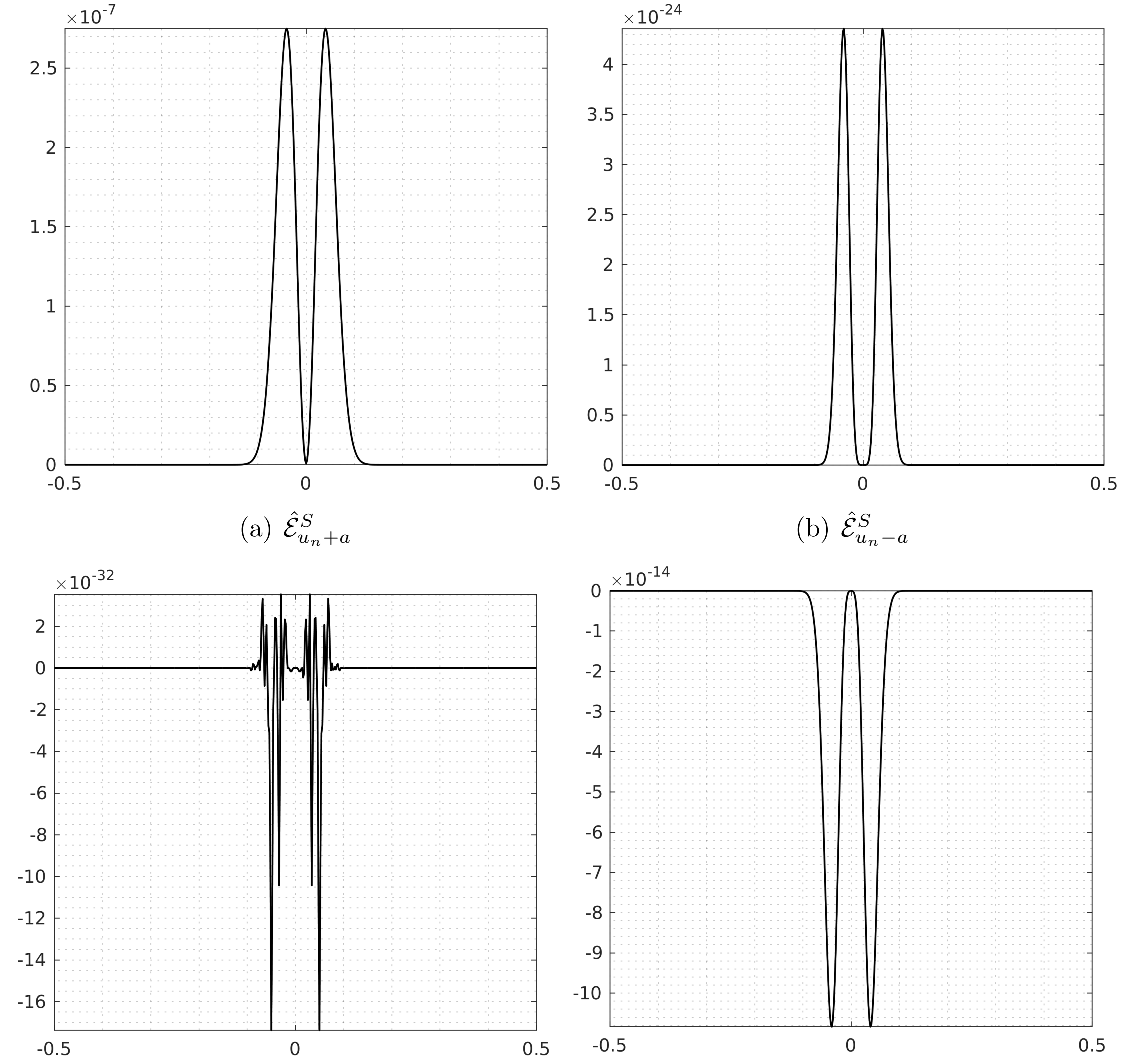
(a) $\hat{\mathcal{E}}^S_{u_n+a}$

(b) $\hat{\mathcal{E}}^S_{u_n-a}$

(c) $\hat{\mathcal{E}}_{u_n-a,u_n+a}$

(d) $\Delta\hat{\mathcal{E}}_a$

Figure 6.40: Gresho Vortex: Entropy production fields along the acoustic eigenvectors at $t = 0$ for the ES Miczek flux. $M_r = 3 \times 10^{-2}$.



(a) $\hat{\mathcal{E}}^S_{u_n+a}$

(b) $\hat{\mathcal{E}}^S_{u_n-a}$

(c) $\hat{\mathcal{E}}_{u_n-a,u_n+a}$

(d) $\Delta\hat{\mathcal{E}}_a$

Figure 6.41: Sound wave: Entropy production fields along the acoustic eigenvectors at $t = 0$ for the ES Miczek flux $M_r = 10^{-2}$.

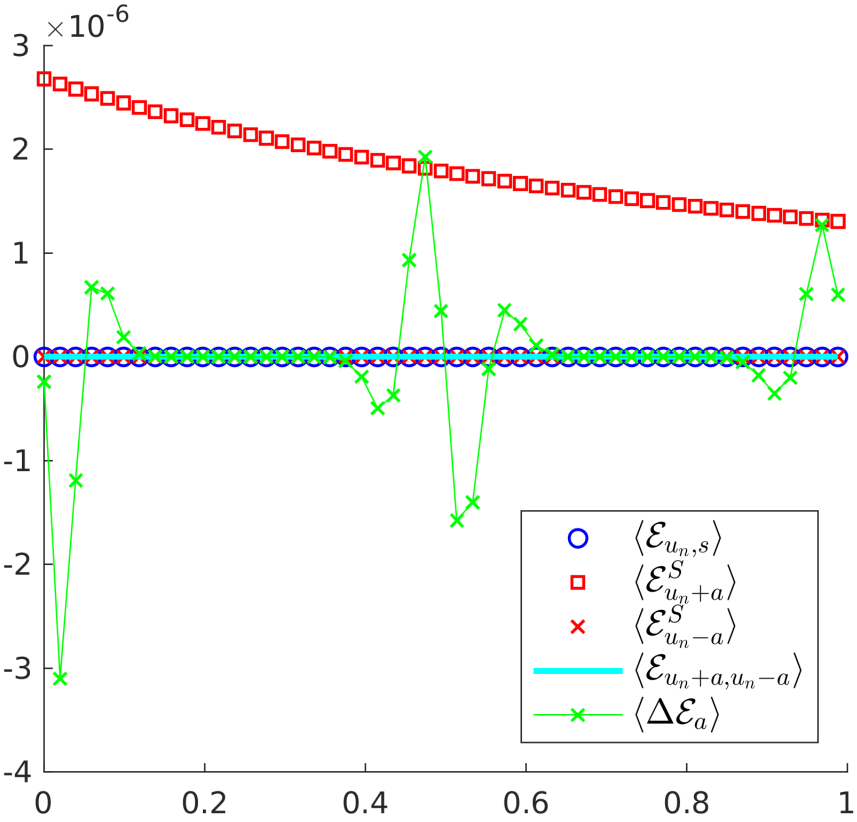


Figure 6.42: Sound wave: Integral of entropy production fields along the acoustic eigenvectors with time at $M_r = 10^{-2}$ for the ES Miczek flux.

6.6.4 An incomplete picture

We should also recognize that our analysis does not take entropy conservation into account. One could see the skew-symmetric term in Miczek's flux as a modification of the baseline entropy-conservative flux instead of a modification of the dissipation operator. Roe [53] observed that if $\mathbf{f}^*$ is entropy-conservative then for any skew-symmetric matrix B, we have:

$$[\mathbf{v}]^T\Big(\mathbf{f}^* - B[\mathbf{v}]\Big) = [\mathbf{v}]^T\mathbf{f}^* - [\mathbf{v}]^T B[\mathbf{v}] = [\mathbf{v}]^T\mathbf{f}^* = [\mathcal{F}].$$

In other words, an EC flux with the addition a the skew-symmetric dissipation term remains EC. However, the corresponding interface entropy flux changes. Equation (6.45) writes:

$$F^*(\mathbf{u}_i, \mathbf{u}_j, \mathbf{n}) = \frac{1}{2}(\mathbf{v}_i + \mathbf{v}_j)^T\Big(\mathbf{f}^*(\mathbf{u}_i, \mathbf{u}_j, \mathbf{n}) - B(\mathbf{v}_i - \mathbf{v}_j)\Big) - \frac{1}{2}(\mathcal{F}_i + \mathcal{F}_j).$$

It could be that the entropy flux (not the entropy production) is responsible for the spurious transients. Figures 6.43 and 6.44 show the results with the skewed scheme (6.6.2) without the symmetric part of the dissipation operator ($|\Lambda| \leftarrow |\Lambda| \times 0$). What is left is a "skewed" EC flux. The same anomalies are observed.

To complete the picture the entropy production breakdown begins to draw, we will need to get a grasp of how entropy is conserved as well. The closed form entropy-conservative fluxes Tadmor developed in [51] (section 6) could provide a good starting point for this purpose.

The fact that skew-symmetric dissipation operators preserve entropy-conservation is proof that entropy-conservative fluxes do not necessarily have the correct Mach number scaling (section 6.3.2).

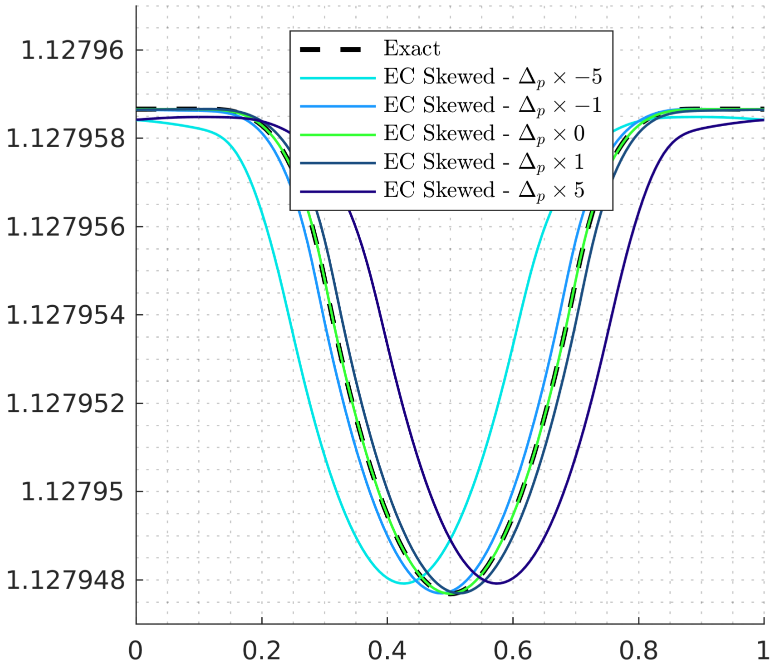
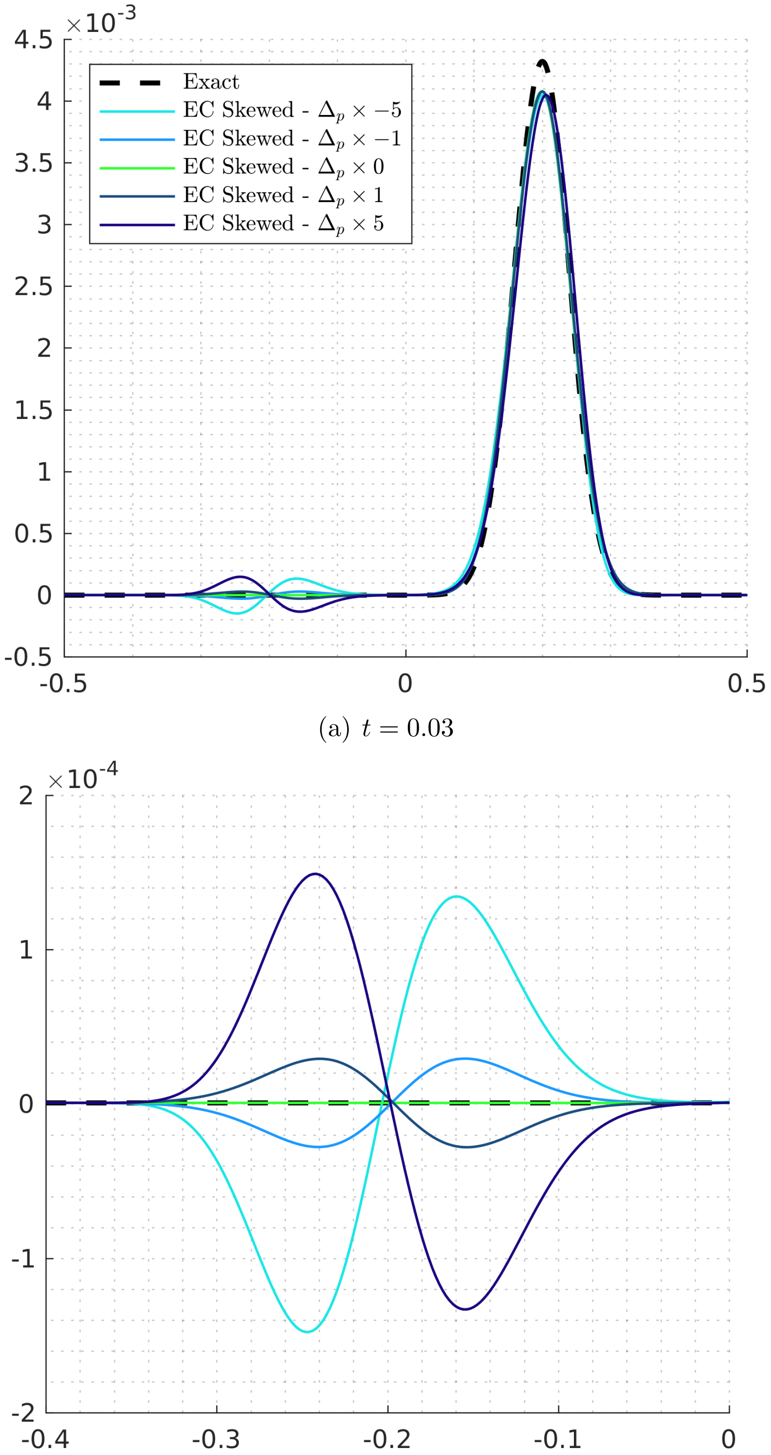


Figure 6.43: Gresho Vortex: Pressure field at $t = 1$ and $M_r = 3 \times 10^{-3}$ for the skewed EC flux when the skew-symmetric term is multiplied by different factors.

6.7 Summary

In this chapter, the behavior of entropy-stable schemes in the low Mach regime was investigated. We showed that standard entropy-stable schemes suffer from the same accuracy degradation issues as standard upwind schemes. The root cause is in the upwind-type dissipation operator as well. Using appropriate similarity and congruence transforms, we were able to show analytically that the flux-preconditioning technique is compatible with entropy-stability to an extent. We derived entropy-stable counterparts of the modified upwind fluxes of Turkel and Miczek. Numerical results confirmed the analysis but also highlighted spurious transients with the flux-preconditioner of Miczek which were not reported until now.

These unexpected anomalies, together with the recent work of Bruel *et al.* [240] on the acoustic limit and the failure of the Turkel flux-preconditioner, led us to further investigate the matter. Leveraging Tadmor's framework, we introduced an entropy



(a) $t = 0.03$

(b) $t = 0.03$, zoomed

Figure 6.44: Sound wave: Same pressure profiles as in figure 6.30 with the skewed EC flux instead. $M_r = 10^{-2}$.

production breakdown of upwind-type dissipation operators that allowed us to revisit the accuracy degradation issue in terms of entropy. In the same spirit as Guillard & Viozat [226] (incompressible limit) and Bruel *et al.* [240] (acoustic limit), we showed that the accuracy degradation problems at the discrete level are caused by discrete entropy fluctuations that are inconsistent with the continuous system. An outgrowth of the overall effort is the discovery that the spurious transients observed with the ES Miczek flux are caused by a skew-symmetric dissipation term which appeared when a scaled form $R|\Lambda|R^T[\mathbf{v}]$ of the modified dissipation operator was sought. Analytical and numerical arguments suggest that this term induces entropy transfers between acoustic waves, but as discussed in the previous section, the role played by skew-symmetric terms remains to be fully understood.

Ultimately, the present work only considered first-order entropy-stable schemes on cartesian grids with two very simple flow configurations. Future work will continue the analysis in a more complex setting, including unstructured grids [220], high-order discretizations and mixed flow configurations [236] of practical interest. Efficient time-integration and preconditioning will be paramount.

CHAPTER VII

Conclusions

The first goal of this thesis was to extend the field of application of ES schemes to more complex physical models. We achieved this objective through the formulation of ES schemes for the compressible multicomponent Euler equations and in this process, we realized that ES schemes can only go as far as their underlying theory. We were fortunate that Chalot *et al.* [174] and Giovangigli [172] already laid some of the ground for us to proceed with the derivations. The limit of vanishing partial densities is intriguing. While the thermodynamic entropy is no longer strictly concave, the ES scheme can be made well-defined and executed without crashing in certain configurations. Pursuing the formulation towards the full compressible Navier-Stokes system, including viscous stresses, heat conduction, and multicomponent diffusion is the logical next step. It can be shown [172] that each of these term will lead to entropy production at the continuous level. One of the challenges ahead will thus be to discretize these terms in a way that is consistent with these results. In the same vein, the full scope of the minimum entropy principle we proved for the multicomponent system remains to be established. It also remains to be seen whether limiting strategies can effectively and rigorously be applied to the multicomponent system.

The second goal of this thesis was to better understand how entropy-stable schemes behave locally. The receding flow problem was more of a toy problem to make ourselves familiar with ES schemes. This problem showed us the limits of semi-discrete analysis. Indeed, Liou's latest work lured us into thinking that entropy-conservative schemes could solve this longtime problem. It turned out that the impact of time-integration was not taken into account. In our setting, the overheating appeared to be strongly correlated to entropy-stability, and entropy-conservation does not necessarily cure this problem. We believe that there is still insight to be gained in this problem. In a different context, Noh [167] showed that the overheating can be cured by introducing an artificial heat flux in the scheme. As previously mentioned, heat conduction is an entropy-producing physical process and first-order entropy-stable discretizations have been derived [164]. If Noh's statement applies in our context, this would mean that the overheating is not simply about producing or not producing entropy but rather about how entropy is produced. This should be investigated.

Without a doubt, our endeavours with the low Mach regime have been the most enriching ones of this thesis. The wealth of past work on this topic is largely to credit for the progress we made, in particular the work of Turkel [218], Guillard & Viozat [226], Miczek *et al.* [224] and Bruel *et al.* [240]. Using congruence, we were able to demonstrate that flux-preconditioning and entropy-stability are compatible to some extent. By leveraging some of the ideas of Roe & Pike [241] and Tadmor [51], we introduced an entropy production breakdown for the upwind dissipation operator of Roe which allowed us to revisit the accuracy degradation issues in entropy terms. In addition, it enabled us to remove the numerical artifacts polluting the numerical solution obtained with the Miczek ES flux. The finding that skew-symmetric dissipation operators can alter the way entropy is produced among waves is an important contribution of this thesis. These operators could be used to probe the local behavior of ES schemes in more details. We could for instance go back to the moving interface problem and see if the same mechanism is behind the spurious acoustic waves generated. The same goes for the overheating, which can be seen as a spurious entropy wave.

Key questions remain to be answered, such as the integration of entropy-conservation in the analysis, but as it currently stands, the entropy production breakdown can already provide valuable information guiding the design of more accurate entropy-stable schemes.

APPENDICES

APPENDIX A

Entropy-conservative and Entropy-stable fluxes in three dimensions for the multicomponent system

The three-dimensional multicomponent Euler equations are given by:

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{f_1}}{\partial x_1} + \frac{\partial \mathbf{f_2}}{\partial x_2} + \frac{\partial \mathbf{f_3}}{\partial x_3} = 0. \tag{A.1}$$

The state vector $\mathbf{u}$ and flux vectors $\mathbf{f_1}$, $\mathbf{f_2}$ and $\mathbf{f_3}$ are defined by:

$$\begin{aligned}
\mathbf{u} &= \begin{bmatrix} \rho_1 & \dots & \rho_N & \rho u & \rho v & \rho w & \rho e^t \end{bmatrix}^T, \\
\mathbf{f_1} &= \begin{bmatrix} \rho_1 u & \dots & \rho_N u & \rho u^2 + p & \rho uv & \rho uw & (\rho e^t + p)u \end{bmatrix}^T, \\
\mathbf{f_2} &= \begin{bmatrix} \rho_1 v & \dots & \rho_N v & \rho uv & \rho v^2 + p & \rho vw & (\rho e^t + p)v \end{bmatrix}^T, \\
\mathbf{f_3} &= \begin{bmatrix} \rho_1 w & \dots & \rho_N v & \rho uw & \rho vw & \rho w^2 + p & (\rho e^t + p)w \end{bmatrix}^T.
\end{aligned}$$

The conservation equation for entropy writes:

$$\frac{\partial U}{\partial t} + \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{\partial F_3}{\partial x_3} = 0,\ U = -\rho s,\ F_1 = -u\rho s,\ F_2 = -v\rho s, F_3 = -w\rho s. \tag{A.2}$$

The vector of entropy variables is:

$$\begin{aligned}\mathbf{v} &= \frac{1}{T}\begin{bmatrix} g_1 - \frac{1}{2}(u^2+v^2+w^2) & \dots & g_N - \frac{1}{2}(u^2+v^2+w^2) & u & v & w & -1 \end{bmatrix} \\ &= \begin{bmatrix} v_{1,1} & \dots & v_{1,N} & v_2 & v_3 & v_4 & v_5 \end{bmatrix}. \end{aligned} \tag{A.3}$$

The potential flux in space $\mathcal{U}$ is unchanged. There are now three spatial potential functions $\mathcal{F}_i = \mathbf{v}^T\mathbf{f_i} - F_i$ to work with. They are given by:

$$\mathcal{F}_1 = \sum_{k=1}^{N} \frac{R}{m_k}\rho_k u, \;\; \mathcal{F}_2 = \sum_{k=1}^{N} \frac{R}{m_k}\rho_k v, \;\; \mathcal{F}_3 = \sum_{k=1}^{N} \frac{R}{m_k}\rho_k w. \tag{A.4}$$

An entropy-conservative flux across an interface of normal $\mathbf{n} = (n_1, n_2, n_3)$, denoted $\mathbf{f}^*$ must satisfy:

$$[\mathbf{v}]^T\mathbf{f}^* = [n_1\mathcal{F}_1 + n_2\mathcal{F}_2 + n_3\mathcal{F}_3].$$

Using the same method as in the 1D case, one obtains $\mathbf{f}^* = [f_{1,1} \; \dots \; f_{1,N} \; f_2 \; f_3 \; f_4 \; f_5]$ with:

$$\begin{aligned} f_{1,k} &= \rho_k^{ln}\overline{u_n}, \\ f_2 &= \frac{n_1}{\overline{1/T}}\left(\sum_{k=1}^{N} r_k\overline{\rho_k}\right) + \overline{u}\sum_{k=1}^{N} f_{1,k}, \\ f_3 &= \frac{n_2}{\overline{1/T}}\left(\sum_{k=1}^{N} r_k\overline{\rho_k}\right) + \overline{v}\sum_{k=1}^{N} f_{1,k}, \\ f_4 &= \frac{n_3}{\overline{1/T}}\left(\sum_{k=1}^{N} r_k\overline{\rho_k}\right) + \overline{w}\sum_{k=1}^{N} f_{1,k}, \\ f_5 &= \sum_{k=1}^{N}(e_{0k} + c_{vk}\frac{1}{(1/T)^{ln}} - \frac{1}{2}\overline{u^2+v^2+w^2})f_{1,k} + \overline{u}f_2 + \overline{v}f_3 + \overline{w}f_4. \end{aligned} \tag{A.5}$$

$u_n = n_1u + n_2v + n_3w$ is the velocity normal to the interface. The temporal Jacobian

[172] is given by:

$$H = \begin{bmatrix} \frac{\rho_1}{r_1} & 0 & \frac{\rho_1}{r_1}u & \frac{\rho_1}{r_1}v & \frac{\rho_1}{r_1}w & \frac{\rho_1}{r_1}(e_1^t) \\ \ddots & & \vdots & \vdots & \vdots & \vdots \\ & \frac{\rho_N}{r_N} & \frac{\rho_N}{r_N}u & \frac{\rho_N}{r_N}v & \frac{\rho_N}{r_N}w & \frac{\rho_N}{r_N}(e_N^t) \\ & & \rho T + u^2 S_1 & uvS_1 & uwS_1 & u(\rho T + S_2) \\ & & & \rho T + v^2 S_1 & vwS_1 & v(\rho T + S_2) \\ & & & & \rho T + w^2 S_1 & w(\rho T + S_2) \\ sym & & & & & \rho T(2k + c_v T) + S_3. \end{bmatrix},$$

$$S_1 = \sum_k \frac{\rho_k}{r_k},\ S_2 = \sum_k \frac{\rho_k}{r_k}(e_k^t),\ S_3 = \sum_k \frac{\rho_k}{r_k}(e_k^t)^2$$

Next is the scaling matrix. Let A_n be the flux Jacobian in the normal direction:

$$A_n = n_1 \frac{\partial \mathbf{f_1}}{\partial \mathbf{u}} + n_2 \frac{\partial \mathbf{f_2}}{\partial \mathbf{u}} + n_3 \frac{\partial \mathbf{f_3}}{\partial \mathbf{u}}.$$

A general expression for the eigenvector matrix R such that $A_n = R \Lambda R^{-1}$ is the following:

$$R = \begin{bmatrix} 1 & & & 0 & 0 & Y_1 & Y_1 \\ & \ddots & & \vdots & \vdots & \vdots & \vdots \\ & & 1 & 0 & 0 & Y_N & Y_N \\ u & \dots & u & & & u + a n_1 & u - a n_1 \\ v & \dots & v & \mathbf{r_I} & \mathbf{r_{II}} & v + a n_2 & v - a n_2 \\ w & \dots & w & & & w + a n_3 & w - a n_3 \\ k - \frac{d_1}{\gamma - 1} & \dots & k - \frac{d_N}{\gamma - 1} & & & h^t + u_n a & h^t - u_n a \end{bmatrix},$$

$$\Lambda = diag(\begin{bmatrix} u_n & \dots & u_n & u_n & u_n + a & u_n - a \end{bmatrix}),$$

where $\mathbf{r_I}, \mathbf{r_{II}} \in \mathbb{R}^{4\times 1}$ are such that $\mathbf{r_I}, \mathbf{r_{II}} \in span\{\mathbf{e_1},\ \mathbf{e_2},\ \mathbf{e_3}\}$ with:

$$\mathbf{e_1} \ = \begin{bmatrix} 0 \\ -an_3 \\ an_2 \\ -a(n_3v - n_2w) \end{bmatrix}, \ \mathbf{e_2} \ = \begin{bmatrix} -an_3 \\ 0 \\ an_1 \\ -a(n_3u - n_1w) \end{bmatrix}, \ \mathbf{e_3} \ = \begin{bmatrix} -an_2 \\ an_1 \\ 0 \\ -a(n_2u - n_1v) \end{bmatrix}$$

and $rank(\mathbf{r_I},\ \mathbf{r_{II}}) = 2$. If $n_1 \neq 0$, take $(\mathbf{r_I}, \mathbf{r_{II}}) = (\mathbf{e_3}, \mathbf{e_2})$. The squared scaling matrix is given by:

$$T^2 = R^{-1}HR^{-T} = \frac{\rho}{\gamma r} diag(\begin{bmatrix} T^{2Y} & T^{2N} & 1/2 & 1/2 \end{bmatrix}), \tag{A.6}$$

where $T^Y = \mathcal{T}^Y(\mathcal{T}^Y)^T \in \mathbb{R}^{N\times N}$ is the same as in the 1D setting, and $T^{2N} \in \mathbb{R}^2$ is given by:

$$T^{2N} = \frac{1}{n_1^2} \begin{bmatrix} n_1^2 + n_3^2 & -n_2n_3 \\ -n_2n_3 & n_1^2 + n_2^2 \end{bmatrix}.$$

If $n_2^2 + n_3^2 = 0$, then $T^{2N} = I_{2\times 2}$, otherwise $T^{2N} = T^N(T^N)^T$ with:

$$T^N = \frac{1}{n_1\sqrt{n_2^2 + n_3^2}} \begin{bmatrix} -n_3 & n_1n_2 \\ n_2 & n_1n_3 \end{bmatrix}.$$

If $n_1 = 0$, then if $n_2 \neq 0$, take $(\mathbf{r_I}, \mathbf{r_{II}}) = (\mathbf{e_3}, -\mathbf{e_1})$. equation (A.6) holds with:

$$T^{2N} = \frac{1}{n_2^2} \begin{bmatrix} n_2^2 + n_3^2 & -n_1n_3 \\ -n_1n_3 & n_2^2 + n_1^2 \end{bmatrix}.$$

Since $n_1 = 0$, this simplifies to $T^{2N} = diag([1 + (n_3/n_2)^2\ 1]) = (T^N)^2$. If $n_1 = 0$ and $n_2 = 0$, then $n_3 \neq 0$. In this last case, take $(\mathbf{r_I}, \mathbf{r_{II}}) = (\mathbf{e_2}, \mathbf{e_1})$. This leads to $T^{2N} = n_3^2 diag([1\ 1])$ and $T^N = n_3 diag([1\ 1])$.

In each case, we can write $T^2 = \mathcal{T}\mathcal{T}^T$ with:

$$\mathcal{T} = \sqrt{\frac{\rho}{\gamma r}} diag(\begin{bmatrix} \mathcal{T}^Y & T^N & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}),$$

and the entropy-stable flux $\mathbf{f}^*$ writes:

$$\mathbf{f}^* = \mathbf{f_{EC}} - \frac{1}{2}(R\mathcal{T})|\Lambda|(R\mathcal{T})^T[\mathbf{v}].$$

APPENDIX B

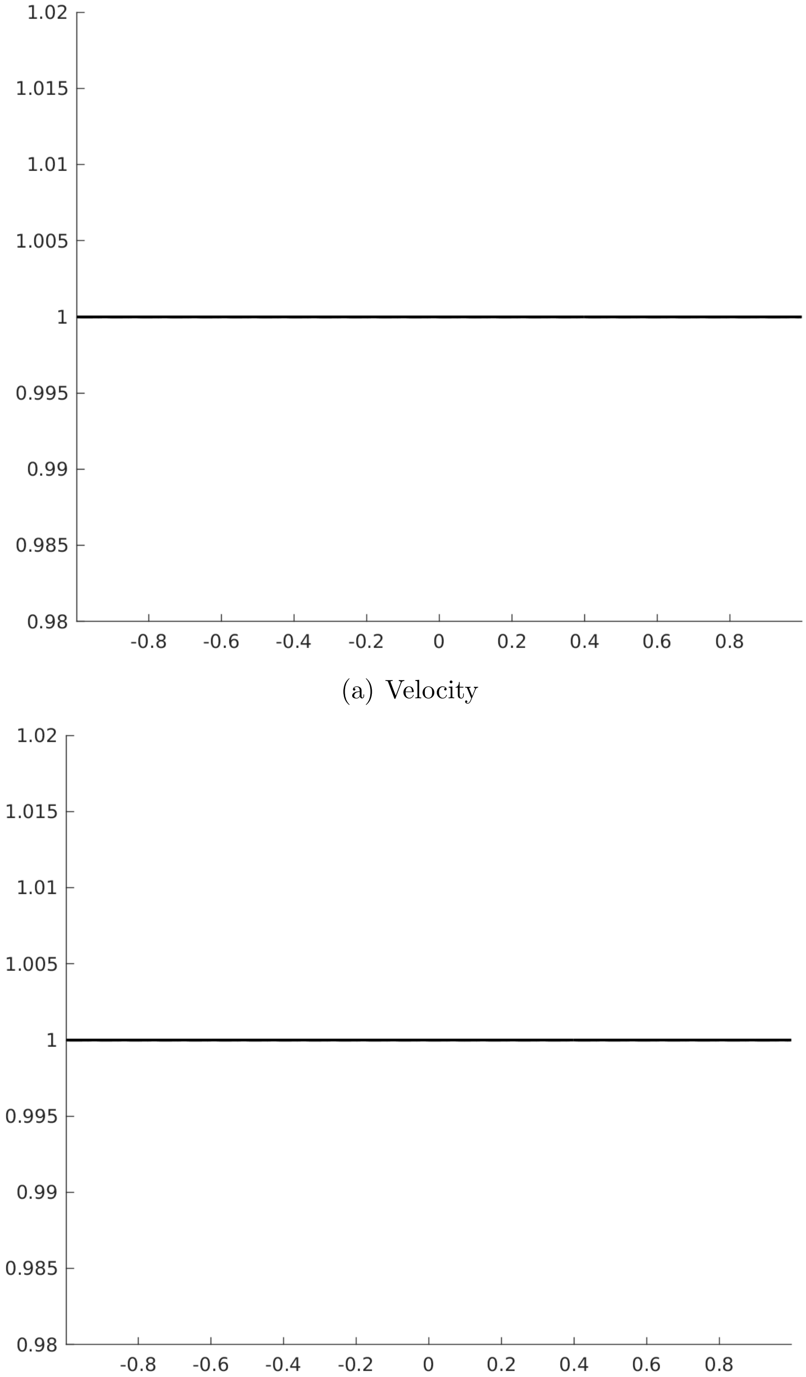
Additional results and discussion for the moving interface problem

Single component case

Here we present numerical results for a moving contact discontinuity in the compressible Euler equations. The initial conditions are given by:

$$\begin{cases} (\rho,\ u,\ p) = & (0.1,\ 1.,\ 1.),\ 0 \leq x \leq 0.5, \\ (\rho,\ u,\ p) = & (1.,\ 1.,\ 1.),\ 0.5 < x \leq 1.0, \end{cases}$$

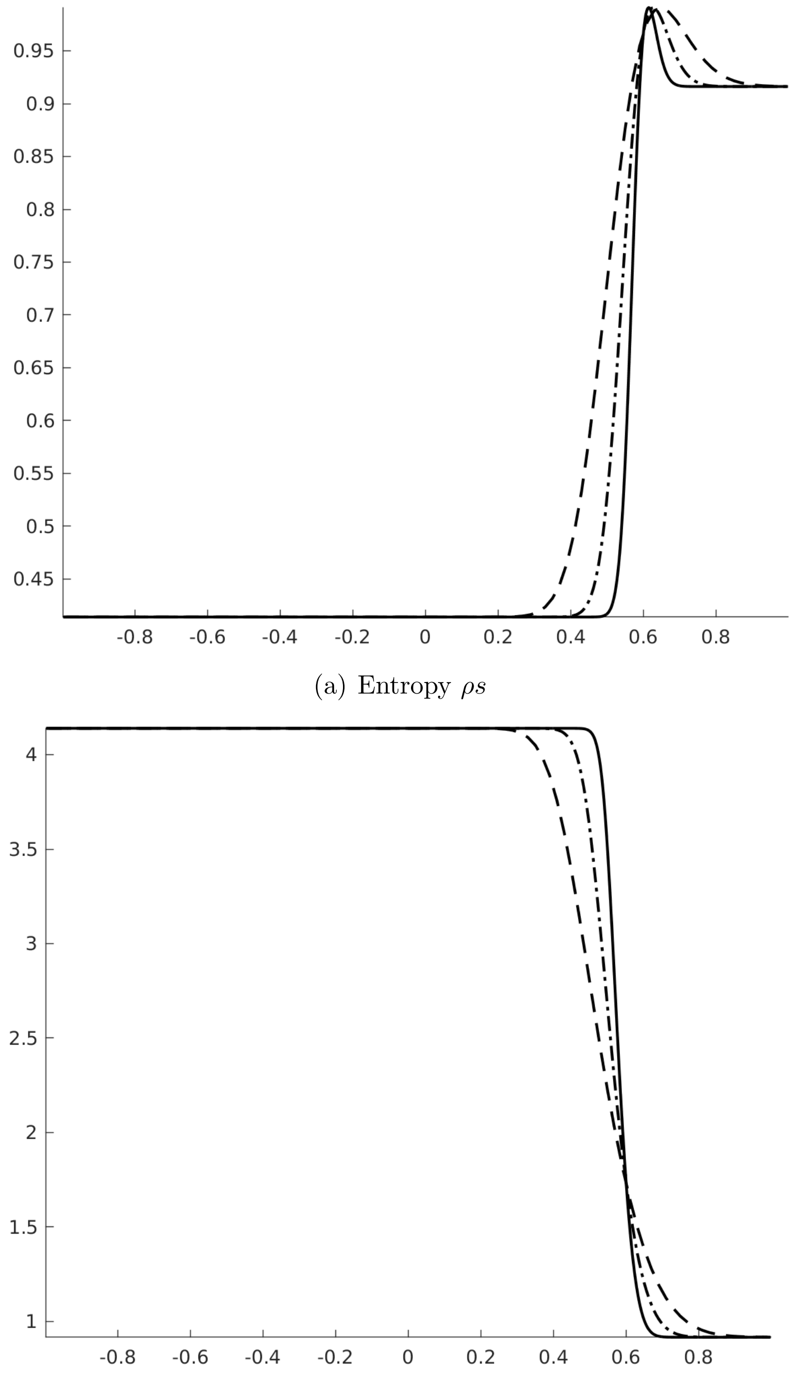
with $\gamma = 1.4$ and $c_v = 1$. The anomalies observed in the multicomponent case are not present. The velocity and pressure remain constant at all times.



(a) Velocity

(b) Pressure

Figure B.1: Single-Component Moving Contact: velocity, pressure, entropy and specific entropy at $t = 0.6$s.



(a) Entropy ρs

(b) Specific Entropy s

Figure B.2: Single-Component Moving Contact: Entropy and specific entropy at $t = 0.6$s.

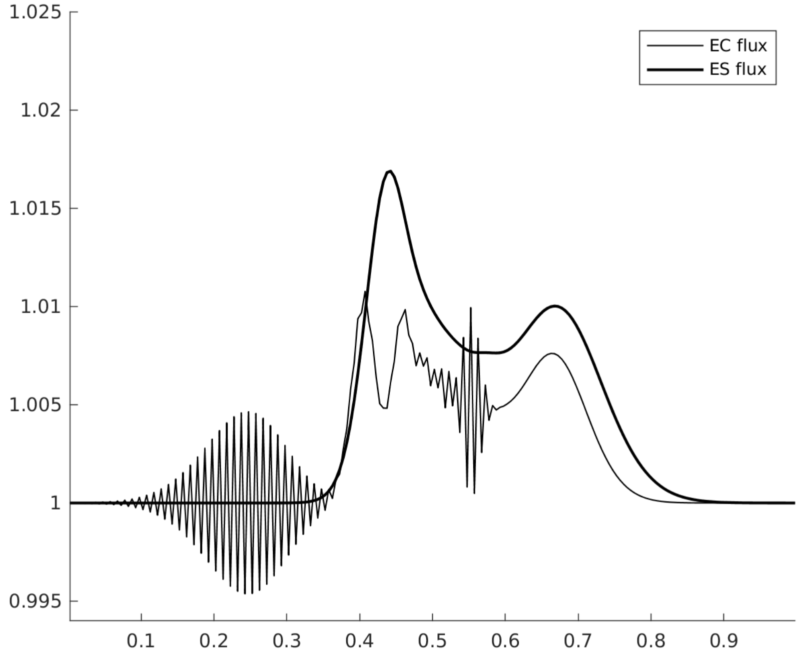
Conserving entropy in space, producing entropy in time

The implicit scheme did not converge in the original setup (due to the generation of negative densities). We therefore considered a different setup for which the pressure oscillations problem is still present.

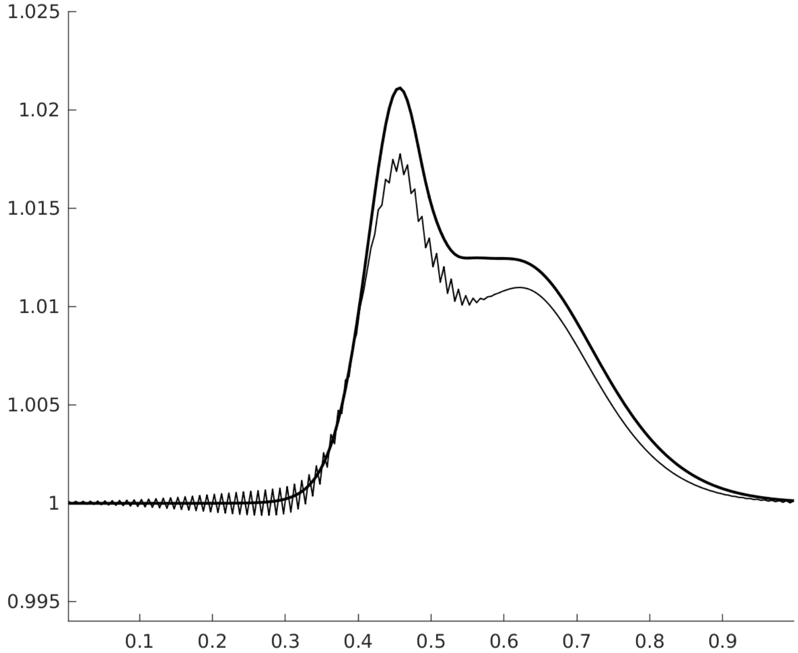
$$\begin{cases} (\rho_1,\ \rho_2,\ u,\ p) = & (0.3,\ 0.15,\ 1.,\ 1.),\ 0 \leq x \leq 0.5, \\ (\rho_1,\ \rho_2,\ u,\ p) = & (0.15,\ 1.,\ 1.,\ 1.),\ 0.5 < x \leq 1.0, \end{cases}$$

with $\gamma_1 = 1.4,\ \gamma_2 = 1.6$ and $c_{v1} = c_{v2} = 1$. A computational grid of 200 cells is used. Figure B.3 show the pressure profiles obtained with an EC flux and an ES flux in space, respectively, for two different CFL numbers. In the first case, the entropy production of the scheme only comes from the stabilization of the Backward Euler time scheme, which grows with Δt. The high frequency oscillations observed are typically observed when too little dissipation is added to EC schemes [144, 164]. In the event that the scheme is EC at the fully discrete level (see [144] for an example), these oscillations will be present but will not increase in magnitude over time. This is not desirable.

The main conclusion we draw is that the pressure anomalies remain present when the upwind dissipation typically used in space is replaced by the dissipation of Backward Euler. The magnitude of the pressure anomalies is higher when both the interface flux and time scheme are ES.



(a) CFL = 2



(b) CFL = 8

Figure B.3: Moving Interface: pressure profiles at $t = 0.1$ s with BE in time and either an EC flux or an ES flux in space

APPENDIX C

Linear stability analysis of the skewed scheme

For a first-order explicit scheme, linear stability is achieved [234] if the complex amplification matrix H defined by:

$$H = I - \nu((\cos(\phi) - 1)D - i\sin(\phi)A),\ \nu = \frac{\Delta t}{\Delta x},\ i^2 = -1,$$

where D is the dissipation matrix, has a spectral radius $\rho(H) \leq 1,\ \forall\phi$. For simplicity, we work with (6.24):

$$A = \begin{bmatrix} u & a/M_r \\ a/M_r & u \end{bmatrix} = R\Lambda R^{-1},\ \Lambda = \begin{bmatrix} u + a/M_r & 0 \\ 0 & u - a/M_r \end{bmatrix},\ R = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

The dissipation operator (6.6.2) writes:

$$D_P = R|\Lambda_P|R^{-1},\ |\Lambda_P| = \begin{bmatrix} f_1(u + a/M_r) & 0 \\ 0 & f_2(a/M_r - u) \end{bmatrix} + \begin{bmatrix} 0 & \delta \\ -\delta & 0 \end{bmatrix}.$$

Taking $f_1 = |u|/(u + a/M_r)$, $|u|/(a/M_r - u)$, and $\delta = a/M_r$ gives:

$$D_P = \begin{bmatrix} |u| & -a/M_r \\ a/M_r & |u| \end{bmatrix},$$

which scales like A. To inspect the spectral radius of H we consider the matrix $\widetilde{H}$ defined by the similarity transformation:

$$\begin{aligned} \widetilde{H} &= R^{-1}HR = I - \nu((c-1)\Lambda_P - is\Lambda) \\ &= \begin{bmatrix} 1 - \nu(c-1)|u| + i\nu s(u + a/M_r) & -\nu(c-1)a/M_r \\ \nu(c-1)a/M_r & 1 - \nu(c-1)|u| + i\nu s(u - a/M_r) \end{bmatrix}, \end{aligned}$$

where $c = \cos(\phi)$ and $s = \sin(\phi)$. $\widetilde{H}$ has two complex eigenvalues λ_1 and λ_2 given by:

$$\begin{aligned} \lambda_1 &= 1 + \nu|u|(c-1) + i\nu((a/M_r)\sqrt{2(1-c)} + su), \\ \lambda_2 &= 1 + \nu|u|(c-1) - i\nu((a/M_r)\sqrt{2(1-c)} - su). \end{aligned}$$

If $c = 1$, then $s = 0$ and $\Lambda_{1,2} = 1$. For $c \neq 1$, we have:

$$\lambda_{1,2}^2 = 1 - 2(1-c)\nu\big(|u| - \nu(u^2 + (a/M_r)^2 \pm s(a/M_r)u\sqrt{2/(1-c)})\big).$$

$\lambda_{1,2}^2 \leq 1$ imposes:

$$\begin{aligned} &2(1-c)\nu\big(|u| - \nu(u^2 + (a/M_r)^2 \pm s(a/M_r)u\sqrt{2/(1-c)})\big) \geq 0 \\ &\Leftrightarrow \nu \leq \frac{M_r^2|u|}{M_r^2u^2 + a^2 \pm sM_rau\sqrt{2/(1-c)}} = \mathcal{O}(M_r^2). \end{aligned}$$

This shows that the linearized scheme with D_P has a more restrictive CFL condition ($\nu = \mathcal{O}(M_r^2)$) than with the classic upwind operator ($\nu = \mathcal{O}(M_r)$). If the skew-symmetric component is removed, a similar CFL condition is obtained:

$$\nu < \frac{M_r^2|u|}{M_r^2u^2 + a^2(c+1)/2 \pm (c+1)M_r a u} = \mathcal{O}(M_r^2).$$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Roe, P.L. : Multidimensional Upwinding. Handbook of Numerical Methods for Hyperbolic Problems - Applied and Modern Issues, pp. 53–80, 2017.

[2] Gresho, P.M., and Lee, R.L. : Don't suppress the wiggles — They're telling you something! Comput. Fluids 9(2), pp. 223-253, 1981.

[3] Salas, M.D. : A Shock-Fitting Primer, CRC Press, 2010.

[4] Moretti, G.: On the matter of shock fitting. In Lecture Notes in Physics 35, pp. 287-292, 1974.

[5] Courant, R., Friedrichs, K.O., and Lewy, H., On the Partial Difference Equations of Mathematical Physics, IBM. J. of Res. and Dev. 11, 215-234, 1967.

[6] Warming, R.F, Hyett, B.J. : The modified equation approach to the stability and accuracy analysis of finite-difference methods, J. Comput. Phys., 14(2), pp. 159-179, 1974.

[7] Crank, J., Nicolson, P.A : Practical Method for Numerical Evaluation of Solutions of Partial Differential Equations of Heat Conduction Type, Proc. Camb. Phil. Soc., 43 pp. 50–56, 1947.

[8] Richtmyer, R.D., and Morton, K.W. : Difference Methods for Initial-Value Problems, Interscience-Wiley, New York, 1967.

[9] Mitchell, A.R., Computational Methods in Partial Differential Equations, Wiley and Son, New York, 1976.

[10] Roe, P.L. : Approximate Riemann Solvers, Parameter Vectors, and Difference Schemes, J. Comput. Phys., 43, pp.357-372, 1981.

[11] Courant, R., Isaacson, E., and Rees, M. : On the solution of non-linear hyperbolic differential equations, Commun. Pure Appl. Maths, 5 pp. 243-255, 1952.

[12] Godunov, S.K. : A difference scheme for numerical computation of discontinuous solution of hydrodynamics equations, Math. Sbornik 47 pp. 271-306, 1959.

[13] Harten, A., : High resolution schemes for hyperbolic conservation laws, J. Comput. Phys. 49, pp. 357-393, 1983.

[14] Levecque, R. : Numerical Methods for Conservation Laws, Birkhauser-Verlag, 1990.

[15] Godunov, S. K.: Difference Methods for Shock Waves, Moscow State University, Ph.D. Dissertation, 1954.

[16] Harten, A., Hyman, J.M., Lax, P.D., Keyfitz, B. : On finite-difference approximations and entropy conditions for shocks, Commun. Pure Appl. Math., 29(3), pp. 294-322, 1986.

[17] Crandall, M.G., and Majda, A. : Monotone Difference Approximations for Scalar Conservation Laws, Math. Comp. 34(149), pp.1-21, 1980.

[18] LeRoux, A.Y. : A Numerical Conception of Entropy for Quasi-Linear Equations, Math. Comput., 31 (140) , pp.848-872, 1977.

[19] Majda, A., and Osher, S. : Numerical Viscosity and the Entropy Condition, Commun. Pure Appl. Math., 32, pp.797-838, 1979.

[20] Yee, H.C. : A class of high-resolution explicit and implicit shock-capturing methods, Tech. Memo. 101088, NASA, 1989.

[21] Lax, P.D., and, Wendroff, B. : Systems of Conservation Laws, Commun. Pure Appl. Math., 13, pp.217-237, 1960.

[22] Lax, P.D., Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves, SIAM, Philadelphia, 1972.

[23] Colella, P., and, Woodward, P.R. : The Piecewise Parabolic Method (PPM) for Gas-Dynamical Simulations, J. Comput. Phys. 54, pp.174-201, 1984.

[24] van Leer, B. : Towards the Ultimate Conservative Difference Scheme, II. Monotonicity and Conservation Combined in a Second-Order Scheme, J. Comput. Phys. (14), pp.361-370, 1974.

[25] Boris, J.P., and Book, D.L. : Flux-Corrected Transport, I. SHASHTA, A Fluid Transport Algorithm That Works, J. Comput. Phys. 11, pp. 38-69, 1973.

[26] Colella, P. : Glimm's Method for Gas Dynamics, SIAM J. Sci. Stat. Comput. (30), pp.76-110, 1982.

[27] Harten, A., Lax, P.D., and van Leer B., On upstream Difference and Godunov-Type Schemes for Hyperbolic Conservative Laws, SIAM Rev. 25(1), pp.35-61, 1983.

[28] Steger, J.L., and Warming, R.F. : Flux Vector Splitting of the Inviscid Gas dynanmic Equations with Application to Finite Difference Methods, J. Comput. Phys. 40, pp.263-293, 1981.

[29] van Leer, B. : Flux-vector splitting for the Euler Equations, ICASE Report 82-30, 1982.

[30] Harten, A., Engquist, B., Osher, S., and Chakravarthy, S.R. : Uniformly high order accurate essentially non-oscillatory schemes, III, J. Comput. Phys. 71, pp. 231-303, 1987.

[31] Liu, X.D., Osher, S., Chan, T. : Weighted essentially non-oscillatory schemes, J. Comput. Phys. 115, pp. 200–212, 1994.

[32] Jiang, G.S., Shu, C.-W. : Efficient implementation of weighted ENO schemes, J. Comput. Phys. 126, pp.202–228, 1996.

[33] Shu, C.-W. : High order weighted essentially nonoscillatory schemes for convection dominated schemes, SIAM Rev. 51, pp. 82–126, 2009.

[34] Von Neumann, J., and Richtmyer, R.D. : A method for the calculation of hydrodynamical shocks, J. Appl. Phys. 21, pp. 232-237, 1950.

[35] Mattsson, A.E., and Rider, W.J. : Artificial Viscosity: Back to the basics, Int. J. Numer. Meth. Fl. 77, pp. 400–417, 2015.

[36] Tadmor, E. : Convergence of spectral methods for nonlinear conservation laws, SIAM J. Numer. Anal. 26, pp. 30–44, 1989.

[37] Guo, B.Y., Ma, H.P., Tadmor, E. : Spectral vanishing viscosity method for nonlinear conservation laws, SIAM J. Numer. Anal. 39, pp.1254–1268, 2001.

[38] Cook, A.W., and Cabot, W.H. : A high-wavenumber viscosity for high-resolution numerical methods, J. Comput. Phys. 195, pp. 594–601, 2004.

[39] Cook, A.W., and Cabot, W.H. : Hyperviscosity for shock-turbulence interactions, J. Comput. Phys. 203, pp. 379–385, 2005.

[40] Cook, A.W. : Artificial fluid properties for large-eddy simulation of compressible turbulent mixing, Phys.Fluids, 19:055103, 2007.

[41] Kawai, S., Lele, S.K. : Localized artificial diffusivity scheme for discontinuity capturing on curvilinearmeshes, J. Comput. Phys. 227, pp.9498–9526, 2007.

[42] Kruzkov, S.N., First-order quasilinear equations in several independent variables, Mathematics of USSR-Sbornik, 10(2), 217, 1970.

[43] Godunov, S.K., An interesting class of quasilinear system, Dokl. Acad. Nauk. SSSR 139, pp. 521-523, 1961.

[44] Hughes, T.J.R, Franca, L.P., and Mallet, M. : A new finite element formulation for computational fluid dynamics: I. Symmetric forms of the compressible Euler and Navier-Stokes equations and the second law of thermodynamics, Comput. Method. Appl. M. 54(2), pp. 223-234, 1986.

[45] Mock, M. S. : Systems of conservation laws of mixed type, J. Differ. Equations 70 (1), pp. 70-88, 1980.

[46] K. O. Friedrichs, and Lax, P.D. : Systems of Conservation Equations with a Convex Extension, Proc. Natl. Acad. Sci. U.S.A. 68 (8), pp. 1686-1688, 1971.

[47] Lax, P. D. : Hyperbolic systems of conservation laws and the Mathematical Theory of Shock Waves, SIAM Regional Conference Lecturers in Applied Mathematics, 11, 1972.

[48] Harten, A. : On the symmetric form of systems of conservation laws with entropy, J. Comput. Phys. 49 (1), pp. 151-164, 1983.

[49] Tadmor, E. : The numerical viscosity of entropy stable schemes for systems of conservation laws. I, Math. Comput. 49, pp. 91-103, 1987.

[50] LeFloch, P.G., Mercier, J.M., and Rohde, C. : Fully Discrete, Entropy Conservative Schemes of Arbitrary Order, SIAM J. Numer. Anal. 40(5), pp. 1968-1992, 2002.

[51] Tadmor, E. : Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems, Acta Numer., 12, pp. 451-512, 2003.

[52] Roe, P.L. : Affordable, entropy consistent flux functions. In: Eleventh International Conference on Hyperbolic Problems: Theory, Numerics and Applications, 2006.

[53] Roe, P.L. : Affordable, Entropy-consistent, Euler Flux Functions: I. Analytical Results, unpublished, 2007.

[54] Ismail, F. : Toward a Reliable Prediction of Shocks in Hypersonic Flow: Resolving Carbuncle With Entropy and Vorticity Control, Phd Thesis, University of Michigan, 2006.

[55] Ismail, F., and Roe, P.L. : Affordable, entropy-consistent Euler flux functions II: Entropy production at shocks, J. Comput. Phys., 228, pp. 5410-5436, 2009.

[56] Fjordholm, U.S., Mishra, S., Tadmor, E., Energy preserving and energy stable schemes for the shallow water equations, in: F. Cucker, A. Pinkus, M. Todd (Eds.), Foundations of Computational Mathematics, Proceedings of the FoCM held in Hong Kong 2008, London Math. Soc. Lecture Notes Ser., vol. 363, 2009, pp. 93–139.

[57] Winters, A.R., and Gassner, G. J. : Affordable, entropy conserving and entropy stable flux functions for the ideal MHD equations, J. Comput. Phys., 304, pp 72-108, 2015.

[58] Wu, K., and Shu, C.-W. : Entropy symmetrization and high-order accurate entropy stable numerical schemes for relativistic MHD equations, arXiv:1907.07467, 2019.

[59] Barth, T.J. : Numerical Methods for Gasdynamic Systems on Unstructured Meshes. In: Kröner D., Ohlberger M., Rohde C. (eds) An Introduction to Recent Developments in Theory and Numerics for Conservation Laws. Lecture Notes in Computational Science and Engineering, vol 5. Springer, Berlin, Heidelberg, 1999.

[60] Fjordholm, U.S., Mishra, S., and Tadmor, E. : Arbitrarily High-order Accurate Entropy Stable Essentially Nonoscillatory Schemes for Systems of Conservation Laws, SIAM J. Numer. Anal., 50 (2) pp. 544-573, 2012.

[61] Fjordholm, U.S., Mishra, S., and Tadmor, E. : ENO reconstruction and ENO interpolation are stable, Found. Comput. Math., 13 (2), pp. 139-159, 2013.

[62] Fisher, T.C. and Carpenter, M.H. : High-order entropy stable finite difference schemes for nonlinear conservation laws: Finite domains, J. Comput. Phys., 252(1) pp 518-557, 2013.

[63] Chandrashekar, P. : Kinetic Energy Preserving and Entropy Stable Finite Volume Schemes for Compressible Euler and Navier-Stokes Equations, Commun. Comput. Phys., 14(5), pp. 1252-1286, 2013.

[64] Jameson, A. : Formulation of Kinetic Energy Preserving Conservative Schemes for Gas Dynamics and Direct Numerical Simulation of One-Dimensional Viscous Compressible Flow in a Shock Tube Using Entropy and Kinetic Energy Preserving Schemes, J. Sci. Comput., 34(2), pp. 188-208, 2008.

[65] Merriam, M. : An Entropy-Based Approach to Nonlinear Stability, NASA Technical Memorandum, 1989.

[66] Slotnick J, Khodadoust A, Alonso J, Darmofal D, Gropp W, et al. : CFD Vision 2030 study: a path to revolutionary computational aerosciences. NASA Tech. Rep. CR-2014-218178, Langley Res. Cent., Hampton, VA.

[67] Hiltebrand, A., and Mishra, S. : Entropy stable shock capturing space–time discontinuous Galerkin schemes for systems of conservation laws, Numer. Math., 126(1), pp. 103-151, 2014.

[68] Diosady, L. T., and Murman, S. M. : Higher-Order Methods for Compressible Turbulent Flows Using Entropy Variables, 53rd AIAA Aerospace Sciences Meeting, 2015.

[69] Friedrich, L., Schnücke, G., Winters, A.R., Del Rey Fernández, D.C., Gassner, G. J., and Carpenter, M.H. : Entropy Stable Space–Time Discontinuous Galerkin Schemes with Summation-by-Parts Property for Hyperbolic Conservation Laws, J. Sci. Comput., 2019.

[70] Pazner, W., and Persson, P-O : Analysis and Entropy Stability of the Line Based Discontinuous Galerkin Method, J. Sci. Comput., 80 (1) pp. 376-402, 2019.

[71] Fernandez, P., Nguyen, N-C, and Peraire, J. : Entropy-stable hybridized discontinuous Galerkin methods for the compressible Euler and Navier-Stokes equations, arXiv:1808.05066, 2018.

[72] Osher, S. : Riemann Solvers, the Entropy Condition, and Difference, SIAM J. Numer. Anal., 21 (2) pp. 217–235, 1984.

[73] Diosady, L., and Murman, S.M. : Design of a Variational Multiscale Method for Turbulent Compressible Flows, AIAA Paper 2013-2870.

[74] Carton de Wiart, C., Diosady, L.T., Garai, A., Burgess, N., Blonigan, P., and Murman, S.M. : Design of a modular monolithic implicit solver for multi-physics applications, AIAA SciTech Forum, 2018.

[75] Murman, S.M., Diosady, L.T., Garai, A., and Ceze, M. : A Space-Time Discontinuous-Galerkin Approach for Separated Flows, 54th AIAA Aerospace Sciences Meeting, 2016.

[76] Diosady, L.T., Murman, S.M. and Carton de Wiart, C. : A higher-order space-time finite-element method for moving-body and fluid-structure interaction problems, ICCFD 10-2018-0310, 2018.

[77] Garai, A., Diosady, L.T., Murman, S.M., and Madavan, N. : Scale-resolving Simulations of Bypass Transition in a High-pressure Turbine Cascade Using a Spectral-element Discontinuous-Galerkin Method, J. Turbomach., 1400 (3), 2017.

[78] Garai, A., Diosady, L.T., Murman, S.M., and Madavan, N., Scale-resolving Simulations of Low-pressure Turbine Cascades with Wall Roughness Using a Spectral-element Method, in Proceedings of ASME Turbo Expo 2018, no. GTS2018-75982, 2018.

[79] Olsson P. : Summation by parts, projections, and stability, RIACS Tech. Rep. TR-93-04, NASA Ames Res.Center, 1993.

[80] Strand, B. : Summation by Parts for Finite Difference Approximations for d/dx, J. Comput. Phys., 110 (1), pp. 47-67, 1994.

[81] Olsson, P., and Oliger, J. : Energy and maximum norm estimates for nonlinear conservation laws, Tech. Rep., Research Institute of Advanced Computer Science, 1994.

[82] Olsson, P., and Gerritsen, M. : Designing an efficient solution strategy for fluid flows, J. Comput. Phys., 129 (2), pp. 245-262, 1996.

[83] Yee, H.C., Sandham, N., and Djomehri, M. : Low-dissipative high-order shock-capturing methods using characteristic-based filters, J. Comput. Phys., 150 (1), pp. 199-238, 1999.

[84] Fisher, T.C., Carpenter, M.H., Nordström, J., Yamaleev, N.K., and Swanson, R.C. : Discretely conservative finite-difference formulations for nonlinear conservation laws in split form: theory and boundary conditions, NASA Tech. Rep. TM 2011-217307.

[85] Gassner, G.J., Winters, A.R., and Kopriva, D.A. : Split form nodal discontinuous Galerkin schemes with summation-by-parts property for the compressible Euler equations, J. Comput. Phys., 327, pp. 39-66, 2016.

[86] Carpenter, M.H., Fisher, T.C., Nielsen, E.J., and Frankel, S.H. : Entropy Stable Spectral Collocation Schemes For The Navier-Stokes Equations: Discontinuous Interfaces, SIAM J. Sci. Comput., 36(5), pp. B835-B867, 2014.

[87] Chen, T., and Shu, C.-W. : Entropy stable high order discontinuous Galerkin methods with suitable quadrature rules for hyperbolic conservation laws, J. Comput. Phys., 345, pp. 427-461, 2017.

[88] Crean, J., Hicken, J.E., Fernandez, D.C.D.R, Zingg, D.W., and Carpenter M.H., Entropy-stable summation-by-parts discretization of the Euler equations on general curved elements, J. Comput. Phys., 356, pp. 410-438, 2018.

[89] Gottlieb, D., Orszag S.A. : Numerical Analysis of Spectral Methods: Theory and Applications. SIAM: Philadelphia, 1977.

[90] Canuto, C., Hussaini, M.Y., Quarteroni, A., and Zang, T.A. : Spectral Methods in Fluid Dynamics, Springer-Verlag, New York, 1987.

[91] Orszag, S. A. : Numerical simulation of incompressible flows within simple boundaries: accuracy, J. Fluid Mech., 49, pp. 75-112, 1971.

[92] Orszag, S.A. : Comparison of pseudospectral and spectral approximation, Stud. Appl. Math., 51 pp. 253-259, 1972.

[93] Orszag, S.A. : On the Elimination of Aliasing in Finite-Difference Schemes by Filtering High-Wavenumber Components, J. Atmos. Sci., 28(6), 1971.

[94] Smagorinsky, J. : General circulation experiments with the primitive equations. I. The basic experiment. Mon. Weather Rev. 91, pp. 99-164, 1963.

[95] Fox, D.G., and Lilly, D.K. : Numerical Simulation of Turbulent Flows, Rev. Geophys., 10(1), 1972.

[96] Philips, N.A. : An examples of nonlinear computational instability. In The Atmosphere and Sea in Motion, ed. B. Bolin, pp. 501-504, New York: Rockefeller Inst. Press, 1959.

[97] Rogallo, R.S., and Moin, P. : Numerical Simulation of Turbulent Flows, Ann. Rev. Fluid Mech. 16, pp. 99-137, 1984.

[98] Kim, J., Moin, P., and Moser, R. : Turbulence statistics in fully developed channel flow at low Reynolds number, J. Fluid Mech., 177 pp. 133-166, 1987.

[99] Spalart, P.R. : Direct simulation of a turbulent boundary layer up to $R_\theta = 1410$, J. Fluid Mech., 187 pp. 61-98, 2019.

[100] Lele, S.K. : Compact Finite Difference Schemes with Spectral-like Resolution, J. Comput. Phys. 103, pp. 16-42, 1992.

[101] Rai, M.M., and Moin, P. : Direct simulations of turbulent flow using finite-difference schemes, J. Comput. Phys., 96(1) pp. 15-53, 1991.

[102] Rai, M.M., and Moin, P. : Direct numerical simulation of transition and turbulence in a spatially evolving boundary layer, J. Comput. Phys., 109 pp. 169–92, 1993.

[103] Foysi, H., Sarkar, S., and Friedrich R. : Compressibility effects and turbulence scalings in supersonic channel flow, J. Fluid Mech., 509 pp. 207–216, 2004.

[104] Mittal, R. and Moin, P. : Suitability of upwind-biased finite difference schemes for large-eddy simulation of turbulent flows, J. Comput. Phys., 35 pp. 1415-1417, 1997.

[105] Park, N., Yoo, J.Y. and Choi, H. : Discretization errors in large-eddy simulation: on the suitability of centered and upwind-biased compact difference schemes, J. Comput. Phys, 198 pp. 580–616, 2004.

[106] Mansour, N.N., Moin, P., Reynolds, W.C., Ferziger, J.H. : Improved methods for large-eddy simulations of turbulence. InTurbulent Shear Flows I, ed. BF Launder, FW Schmidt, HH Whitelaw, pp. 386–401. Berlin:Springer-Verlag, 1979.

[107] Feiereisen, W.J., Reynolds, W.C., Ferziger, J.H., Numerical simulation of a compressible, homogeneous,turbulent shear flow, Rep. TF 13, Thermosci. Div., Mech. Eng., Stanford Univ, 1981.

[108] Blaisdell, G.A., Spyropoulos, E.T., Qin, J.H. : The effect of the formulation of non-linear terms on aliasing errors in spectral methods, Appl. Numer. Math, 21 pp. 207–219, 1996.

[109] Ducros, F., Laporte, F., Soulères, T., Guinot, V., Moinat, P., and Caruelle, B. : High-order fluxes for conservative skew-symmetric-like schemes in structured meshes: application to compressible flows, J. Comput. Phys., pp. 161 pp. 114–139, 2000.

[110] Kennedy, C.A., and Gruber, A. : Reduced aliasing formulations of the convective terms within the Navier-Stokes equations, J. Comput. Phys., 227 pp. 1676–1700, 2008.

[111] Zang, T.A. : On the rotation and skew-symmetric forms for incompressible flow simulations, Appl. Numer. Math., 7 (1) pp. 27-40, 1991.

[112] Subbareddy, P.K., and Candler, G.V. : A fully discrete, kinetic energy consistent finite-volume scheme for compressible flows, J. Comput. Phys., 228, pp. 1347-1364, 2009.

[113] Sandham ND, Li Q, Yee HC. : Entropy splitting for high-order numerical simulation of compressible turbulence, J. Comput. Phys., 178 pp. 307–322, 2002.

[114] Honein, A.E., and Moin, P. : Higher entropy conservation and numerical stability of compressible turbulence simulations, J. Comput. Phys, 201 pp. 531–545, 2004.

[115] Morinishi, Y. : Skew-symmetric form of convective terms and fully conservative finite difference schemes for variable density low-Mach number flows, J. Comput. Phys, 229 pp. 276–300, 2010.

[116] Karniadakis, G.E.M., and Sherwin, S.J. : Spectral/HP Element Methods for Computational Fluid Dynamics, Oxford University Press, Oxford, 2005.

[117] Reed, W. H., and Hill, T. R. : Triangular mesh methods for the neutron transport equation, Technical Report LA-UR-73-479, Los Alamos Scientific Laboratory, 1973.

[118] Cockburn B., Karniadakis G.E., and Shu C.W. : The Development of Discontinuous Galerkin Methods. In: Discontinuous Galerkin Methods. Lecture Notes in Computational Science and Engineering, vol 11. Springer, Berlin, Heidelberg, 2000.

[119] Wang, Z.J., Fidkowski, K., Abgrall, R., Bassi, F., Caraeni, D., Cary, A., Deconinck, H., Hartmann, R., Hillewaert, K., Huynh, H.T., Kroll, N., May, G., Persson, P-0, Van Leer, B., Visbal, M. : High-Order CFD Methods: Current Status and Perspective, Int. J. Numer. Meth. Fluids, 72 pp. 811-845, 2013.

[120] Kirby, R.M., and Karniadakis, G.E.: De-aliasing on non-uniform grids: algorithms and applications, J. Comput. Phys. 191, pp. 249–264, 2003.

[121] Gassner, G.J., and Beck, A.D. : On the accuracy of underresolved turbulence simulations, Theor. Comput. Fluid dyn. 22, pp. 2560-2579, 2011.

[122] Kroll, N., Hirsch, C., Bassi, F., Johnston, C., and Hillewaert K. : IDIHOM: Industrialization of High-Order Methods - A Top Down Approach. Results of a Collaborative Research Project Funded by the European Union, 2010-2014, Springer, 2015.

[123] Gassner, G.J., Winters, A.R., and Kopriva, D.A. : Split form nodal discontinuous Galerkin schemes with summation-by-parts property for the compressible Euler equations, J. Comput. Phys., 327 pp. 39-66, 2016.

[124] Flad, D., Beck, A., and Munz, C.-D. : Simulation of underresolved turbulent flows by adaptive filtering using the high order discontinuous Galerkin spectral element method, J. Comput. Phys., 313 pp. 1-12, 2016.

[125] Flad, D., and Gassner, G.J. : On the use of kinetic energy preserving DG-schemes for large eddy simulation, J. Comput. Phys., 350 pp. 782-795, 2017.

[126] Reynolds, O. : On the Dynamical Theory of Incompressible Viscous Fluids and the Determination of the Criterion, Philosophical Transactions of the Royal Society of London A., 186 pp. 123–164, 1895.

[127] Sagaut, P., Deck, S., and Terracol, M. : Multiscale and Multiresolution Approaches in Turbulence, Imperial College Press, 2013.

[128] Garnier, E., Mossi, M., Sagaut, P., Comte, P., and Deville, M. : On the Use of Shock-Capturing Schemes for Large-Eddy Simulation, J. Comput. Phys., 153 (2), pp. 273-311, 1999.

[129] Ghosal, S. : An analysis of numerical errors in large-eddy simulations of turbulence, J. Comput. Phys., 125 (1), pp. 187-206, 1996.

[130] Vreman, B., Geurts, and H. Kuerten, Discretization error dominance over subgrid terms in large eddy simulation of compressible shear layers in 2D, Comm. Numer. Methods Eng, 10, pp. 785 , 1994.

[131] Boris, J.P., Grinstein, F.F., Oran, E.S., and Kolbe, R.J.: New insights into Large Eddy Simulation, Fluid Dyn. Res., 10, 199, 1992.

[132] Grinstein, F.F., and Fureby, C. : On Flux-Limiting-Based Implicit Large Eddy Simulation, J. Fluids Eng. Dec, 129(12), pp. 1483-1492, 2007.

[133] Germano, M., Piomelli, U., Moin, P., and Cabot, W. : A Dynamic Subgrid-Scale Eddy Viscosity Model, Phys. Fluids A, 3(7) pp. 1760-1765, 1991.

[134] Ghosal, S., Lund, T.S., Moin, P., and Akselvoll, K. : A Dynamic Localization Model for Large-Eddy Simulation of Turbulent Flows, J. Fluid Mech., 286 pp. 229-255, 1995.

[135] Piomelli, U. : High Reynolds Number Calculations Using the Dynamic Subgrid-Scale Stress Model, Phys. Fluids A, 5(6) pp.1484-1490, 1993.

[136] Margolin, L.G., and Rider, W.J. : A rationale for implicit turbulence modelling Int. J. Numer. Methods Fluids, 39 (9) pp. 821-841, 2002.

[137] Grinstein, F.F., and Margolin, L.G., and Rider, W.J. : Implicit Large Eddy Simulation: Computing Turbulent Fluid Dynamics, Cambridge University Press, 2007.

[138] Beck, A.D., Bolemann, T., Flad, D., Frank, H., Gassner, G.J., Hindenlang, F., and Munz, C.D. : High-order discontinuous Galerkin spectral element methods for transitional and turbulent flow simulations, Int. J. Numer. Methods Fluids, 76 (8) pp. 522-548, 2014.

[139] Uranga, A., Persson, P.O., Drela, M., and Peraire, J. : Implicit large eddy simulation of transition to turbulence at low Reynolds numbers using a discontinuous Galerkin method, Int. J. Numer. Methods Eng., 87 (1–5) pp. 232-261, 2011.

[140] Moura, R.C., Mengaldo, G., Peiró, J., and Sherwin, S.J. : On the eddy-resolving capability of high-order discontinuous Galerkin approaches to implicit LES/under-resolved DNS of Euler turbulence, J. Comput. Phys., 330 pp. 615-623, 2017.

[141] Fernandez, P., Nguyen, N-C, and Peraire, J.: Subgrid-scale modeling and implicit numerical dissipation in DG-based Large-Eddy Simulation, AIAA Aviation Forum, 2017.

[142] Mengaldo, G., Moura, R.C., Giralda, B., Peiró, J. and Sherwin S.J. : Spatial eigensolution analysis of discontinuous Galerkin schemes with practical insights for under-resolved computations and implicit LES, Comput. Fluids, 169 pp. 349-364, 2018.

[143] Fernandez, P., Moura, R.C., Mengaldo, G., and Peraire, J. : Non-modal analysis of spectral element methods: Towards accurate and robust large-eddy simulations, Comput. Method. Appl. M., 346 pp. 43-62, 2019.

[144] Gouasmi, A., Murman, S.M., and Duraisamy, K. : Entropy conservative schemes and the receding flow problem, J. Sci. Comput., 78 (2) pp. 971-994, 2018.

[145] Gouasmi, A., Duraisamy, K. and Murman, S.M. : On entropy stable temporal fluxes, arXiv:1807.03483, 2018.

[146] Gouasmi, A., Duraisamy, K.D. and Murman, S.M. : Formulation of Entropy-Stable schemes for the compressible multicomponent Euler equations, arxiv:1904.00972, 2019.

[147] Gouasmi, A., Duraisamy, K.D., Murman, S.M., and Tadmor, E. : A minimum entropy principle in the multicomponent compressible Euler equations, ESAIM-Math. Model. Num., 2019.

[148] Peery, K. M., and Imlay, S. T. : Blunt-Body Flow Simulations, AIAA Paper 88-2904, 1988.

[149] Kemm, F. : Heuristical and numerical considerations for the carbuncle phenomenon, Appl. Math. Comput., 320 C, pp. 596-613, 2018.

[150] Arora, M., and Roe, P.L. : On Postshock Oscillations Due to Shock Capturing Schemes in Unsteady Flows, J. Comput. Phys., 130 (1), pp. 25-40, 1997.

[151] Casper, J., and Carpenter, M.H. : Computational considerations for the simulation of shock-induced sound, SIAM J. Sci. Comput., 19 pp. 813–828, 1998.

[152] Cook, A.W. : Enthalpy diffusion in multicomponent flows, Phys. Fluids 21, 055109, 2009.

[153] Johnsen, E., Larsson, J., Bhagatwala, A.V., Cabot, W.H., Moin, P., Olson, B.J., Rawat, P.S., Shankar, S.J., Sjögreen, b., Yee, H.C., Zhong, X., and Lele, S.K.: Assessment of high-resolution methods for numerical simulations of compressible turbulence with shock waves, J. Comput. Phys, 229 pp. 1213–1237, 2010.

[154] Larsson, J., and Lele, S.K. : Direct numerical simulation of canonical shock/turbulence interaction, Phys. Fluids, 21:126101, 2009.

[155] Quirk, J.J. : A Contribution to the Great Riemann solver debate, Int. J. Numer. Fl., 18, 555-574, 1994.

[156] Roe, P.L. : Is Discontinuous Reconstruction Really a Good Idea?, J. Sci. Comput., 73 (2-3) pp. 1094-1114, 2017.

[157] Eyink, G.L., and Drivas, T.D. : and An Onsager Singularity Theorem for Turbulent Solutions of Compressible Euler Equations, Commun. Math. Phys., 359(2), pp 733–763, 2017.

[158] Eyink, G.L., and Drivas, T.D. : Cascades and Dissipative Anomalies in Compressible Fluid Turbulence, Phys. Rev. X 8, 2018.

[159] Wang, J., Minping, W., Chen, S., Xie, C., Wang, L.-P., and Chen, S. : Cascades of temperature and entropy fluctuations in compressible turbulence, J. Fluid Mech., 867, pp. 195–215, 2019.

[160] Bayly, B.J., Levermore, C.D., and Passot, T. : Density variations in weakly compressible flows, Phys. Fluids A 4, pp. 945–954, 1992.

[161] Ray, D. : Entropy-stable finite difference and finite volume schemes for compressible flows, Ph.D. Thesis, TIFR, Bangalore, 2017.

[162] Rider, W.J. : Revisiting Wall-Heating, J. Comput. Phys., 162(2), pp. 395-410, 2000.

[163] Derigs, D., Winters, A.R., Gassner, G.J., and Walch, S. : A novel averaging technique for discrete entropy-stable dissipation operators for ideal MHD, J. Comput. Phys., 330 pp. 624-632, 2017.

[164] Tadmor, E., and Zhong, W. : Entropy stable approximations of Navier-Stokes equations with no artificial numerical viscosity, J. Hyperbol. Differ. Eq., 3(3), pp. 529-559, 2006.

[165] M. S. Liou, The Root Cause of the Overheating Problem, 23rd AIAA Computational Fluid Dynamics Conference, 2017.

[166] M. S. Liou, Why is the overheating problem difficult: The role of entropy, 21st AIAA Computational Fluid Dynamics Conference, 2017.

[167] W. F. Noh, Errors for calculations of strong shocks using an artificial viscosity and an artifical heat flux, J. Comput. Phys., 72, pp. 78-120 (1987).

[168] E. F. Toro, Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction, Third Edition, Springer-Verlag Heidelberg, 2009.

[169] Chiodaroli, E., A counterexample to well-posedness of entropy solutions to the compressible Euler system, J. Hyperbol. Differ. Eq., 11, pp. 493-519, 2014.

[170] de Lellis, D., and Szekelyhidi, L. : On admissibility criteria for weak solutions of the euler equations, Arch. Ration. Mech. An., 195, pp. 225-260, 2010.

[171] Elling, V., The carbuncle phenomenon is incurable, Acta Math. Sci., 29 (6), pp. 1647-1656, 2009.

[172] Giovangigli, V. : Multicomponent flow modeling, Chapters 1 and 8, Birkhauser, Boston, 1999.

[173] Giovangigli, V., and Matuszewski, L. : Structure of Entropies in Dissipative Multicomponent Fluids, Kin. Rel. Models, 6 pp. 373-406, 2013.

[174] Chalot, F., Hughes, T.J.R. and Shakib, F. : Symmetrization of conservation laws with entropy for high-temperature hypersonic computations, Computing systems in engineering, 1 (2-4) pp. 495-521, 1990.

[175] Tadmor, E. : A minimum entropy principle in the gas dynamics equations, Appl. Numer. Math., 2 (3-5) pp. 211-219, 1986.

[176] Guermond, J.L., Popov, B. : Invariant Domain and First-Order Continuous Finite Element Approximation for hyperbolic systems, SIAM J. Numer. Anal., 54 (4), pp. 2466-2489, 2016.

[177] Guermond, J.L., Nazarov, M., Popov, B., Tomas, I. : Second-order invariant domain preserving approximation of the Euler equations using convex limiting, SIAM J. Sci. Comput., 40 (5), pp. 3211–3239, 2018.

[178] Guermond, J.L., and Popov, B. : Fast estimation from above of the maximum wave speed in the Riemann problem for the Euler equations, J. Comput. Phys., 328 pp. 908-926, 2016.

[179] Abgrall, R. : Généralisation du solveur de Roe pour le calcul d'écoulements de mélanges de gaz parfaits à concentrations variables. La Recherche Aérospatiale, 6, pp 31–43, 1988.

[180] Billet, G., and Abgrall, R. : An adaptive shock-capturing algorithm for solving unsteady reactive flows, Comput. Fluids, 32 (10) 1473-1495, 2003.

[181] Fernandez, G., and Larrouturou, B. : Hyperbolic Schemes for Multi-Component Euler Equations. In: Ballmann J., Jeltsch R. (eds) Nonlinear Hyperbolic Equations — Theory, Computation Methods, and Applications. Notes on Numerical Fluid Mechanics, vol 24. Vieweg-Teubner Verlag, 1989.

[182] Larrouturou, B. : How to preserve the mass fractions positivity when computing compressible multi-component flows, J. Comput. Phys., 95 (1) 59-84, 2001.

[183] Habbal, A., Dervieux, A., Guillard, H., and Larrouturou, B. : Explicit calculation of reactive flows with an upwind finite element hydro-dynamical code. Technical Report 690, INRIA, 1987.

[184] Abgrall, R., and Karni, S. : Computations of Compressible Multifluids, J. Comput. Phys., 169 (2) pp. 594-623, 2000.

[185] Karni, S. : Multicomponent flow calculation by a consistent primitive algorithm, J. Comput. Phys., 112 (1) pp. 31-43, 1994.

[186] Abgrall, R. : How to prevent pressure oscillations in multicomponent flow calculations: a quasi-conservative approach, J. Comput. Phys., 125 (1) pp. 150-160, 1996.

[187] Richtmyer, R.D. : Taylor instability in shock acceleration of compressible fluids, Commun. Pure Appl. Math, 13 (2) pp. 297-319, 1960.

[188] Meshkov, E. E. : Instability of the interface of two gases accelerated by a shock wave, Fluid Dyn., 4 (5) pp. 101-104, 1969.

[189] Picone, J. M., and Boris, J. P. : Vorticity generation by shock propagation through bubbles in a gas, J. Fluid Mech., 189 pp. 23-51, 1988.

[190] Marble, F. E, Hendricks, G.J., and Zukoski, E. : Progress towards shock enhancement of supersonic combustion processes, 23rd Joint Propulsion Conference, 1987.

[191] Karni, S., and Quirk, J.J. : On the dynamics of a shock-bubble interaction, J. Fluid Mech., 318 pp. 129-163, 1996.

[192] Kawai, S., and Terashima, H. : A high-resolution scheme for compressible multicomponent flows with shock waves, Int. J. Numer. Fl., 66 (10) pp. 1207-1225, 2011.

[193] Haas, J.F., and Sturtevant, B. : Interactions of weak shock waves with cylindrical and spherical gas inhomogeneities, J. Fluid Mech., 181 pp. 41-76, 1987.

[194] Marquina, A., and Mulet, P. : A flux-split algorithm applied to conservative models for multicomponent compressible flows, J. Comput. Phys., 185 (1) pp. 120-138, 2003.

[195] Johnsen, E., and Colonius, T. : Implementation of WENO schemes in compressible multicomponent flow problems, J. Comput. Phys., 219 (2) pp. 715-732, 2006.

[196] Derigs, D., Winters, A. R., Gassner, G. J., and Walch, S. : A novel high-order, entropy stable, 3D AMR MHD solver with guaranteed positive pressure, J. Comput. Phys., 317 pp. 223-256, 2016.

[197] Ma, P.C., Lv, Y., and Ihme, M. : An entropy-stable hybrid scheme for simulations of transcritical real-fluid flows, J. Comput. Phys., 340 pp. 330-357, 2017.

[198] Castro, M.J., Fjordholm, U.S., Mishra, S. and Parés, C. : Entropy conservative and entropy stable schemes for nonconservative hyperbolic systems, SIAM J. Numer. Anal., 51(3) pp. 1371-1391, 2012.

[199] Hou, T.Y., and Le Floch : Why nonconservative schemes converge to wrong solutions: error analysis, Math. Comput., 62 (206), pp. 497-530, 1994.

[200] Abgrall, R., and Karni, S. : A comment on the computation of non-conservative products, J. Comput. Phys., 229(8), pp. 2759-2763, 2010.

[201] P. D. Lax, Shock Waves and Entropy, In: E. Zarantonello, Ed., Contributions to Nonlinear Functional Analysis, Academia Press, New York, 1971, pp. 603-634.

[202] Lax, P. D. : Hyperbolic systems of conservation laws. II, Commun. Pure Appl. Math, 10 pp. 537–566, 1957.

[203] Harten, A., Lax, P. D., Levermore C. D., and Morokoff, W. J. : Convex entropies and hyperbolicity for general Euler equations, SIAM J. Numer. Anal., 33 (6) pp. 2117-2127, 1998.

[204] Frolov, R. : An efficient algorithm for the multicomponent compressible Navier–Stokes equations in low- and high-Mach number regimes, Comput. Fluids, 178 pp. 15–40, 2019.

[205] Tadmor, E. : Skew-Self adjoint Form for Systems of Conservation Laws, J. Math. Anal. Appl., 103 pp. 428-442, 1984.

[206] Tadmor, E. : Numerical Viscosity and the Entropy Condition for Conservative Difference Schemes, Math. Comput., 43 (168) pp. 369-381, 1984.

[207] Kroner, D., LeFloch, P.G. and Thanh, M. : The minimum entropy principle for compressible fluid flows in a nozzle with discontinuous cross-section, ESAIM-Math. Model. Num., 42 pp. 425-442, 2008.

[208] Guermond, J.L., Popov, B. : Viscous Regularization of the Euler equations and entropy principles, SIAM J. Appl. Math., 74 (2), pp. 284-305, 2014.

[209] Guermond, J.-L., Popov, B., and Tomas, I. : Invariant domain preserving discretization-independent schemes and convex limiting for hyperbolic systems, Comput. Method. Appl. M., 347, 2019, pp. 143-175.

[210] Delchini, M. O., Ragusa, J.C., Berry, R. A. : Viscous Regularization for the Non-equilibrium Seven-Equation Two-Phase Flow Model, J. Sci. Comput., 69 pp.764-804, 2016.

[211] Delchini, M.O., Ragusa, J.C., and Ferguson, J. : Viscous Regularization of the full set of nonequilibrium-diffusion Grey Radiation-Hydrodynamic equations, Int. J. Numer. Fl., 85(1), pp. 30-47, 2017.

[212] Zhang, X., Shu, C.-W. : A minimum entropy principle of high order schemes for gas dynamics equations Numer. Math., 121 pp. 545-563, 2012.

[213] Lv, Y., Ihme, M. : Entropy-bounded discontinuous Galerkin scheme for Euler equations, J. Comput. Phys., 295 pp. 715–73, 2015.

[214] Klainerman, S., and Majda, A. : Compressible and Incompressible Fluids, Commun. Pure Appl. Math, 35, pp. 629-651, 1982.

[215] Schochet, S. : The mathematical theory of low Mach number flows, ESAIM-Math. Model. Num., 39 (3), pp. 441-458, 2005.

[216] Turkel, E. : Preconditioning Techniques in Computational Fluid Dynamics, Annu. Rev. Fluid Mech., 31, pp.385-416, 1999.

[217] Turkel, Preconditioned Methods for Solving the Incompressible and Low Speed Compressible Equations, J. Comput. Phys., 72(2), pp. 277-298, 1987.

[218] Turkel, E., Fiterman, A., and van Leer B. : Pre-conditioning and the limit to the incompress-ible flow equations, In: *Computing the Future: Frontiers of Computational Fluid Dynamics*, ed. DA Caughey, MM Hafez, pp. 215–34. New York: Wiley, 1994.

[219] G. Volpe, Performance of compressible flow codes at low Mach numbers, AIAA, 1993.

[220] Rieper, F., and Bader, G. : The influence of cell geometry on the accuracy of upwind schemes in the low mach number regime, J. Comput. Phys. 228 (8), pp. 2918-2933, 2009.

[221] Gresho, P.M., and Chan, S.T. : On the theory of semi-implicit projection methods for viscous incompressible flow and its implementation via a finite element method that also introduces a nearly consistent mass matrix. Part 2: Implementation, Int. J. Numer. Fl., 11 (5), pp. 1990.

[222] Liska, R., and Wendroff, B. : Comparison of several difference schemes on 1D and 2D test problems for the Euler equations, SIAM J. Sci. Comput. 25 (3), pp. 995–1017, 2003.

[223] Miczek, F. : Simulation of low Mach number astrophysical flows, PhD Thesis, Technical University of Munich, 2013.

[224] Miczek, F., Ropke, F.K., and Edelmann, P.V.F. : New numerical solver for flows at various Mach numbers, A&A, 576, A50 (2015).

[225] Barsukow, W., Edelmann, P.V.F., Klingenberg, C., Miczek, F., and Röpke, F.K. : A numerical scheme for the compressible low-Mach number regime of ideal fluid dynamics, J. Sci. Comput., 72(2), pp 623–646, 2017.

[226] Guillard, H. and Viozat, C.: On the Behavior of Upwind Schemes in the Low Mach Number Limit, Comput. Fluids 28, pp. 63-86, 1999.

[227] Weiss, J.M., and Smith, W.A. : Preconditioning applied to variable and constant density flows, AIAA Journal, 33(11), pp. 2050-2057, 1995.

[228] Guillard, H., and Nkonga, B. : On the Behavior of Upwind Schemes in the Low Mach Number limit: A Review, chapter 8 in *Handbook of Numerical Analysis 18*, 2017.

[229] Klein, R. : Semi-implicit extension of a godunov-type scheme based on low mach number asymptotics I: One-dimensional flow, J. Comput. Phys. 121 (2), pp. 213-237, 1995.

[230] Van Leer, B., Lee, W-T., and Roe, P. L. : Characteristic time-stepping or local preconditioning of the Euler equations, AIAA-1991-1552, AIAA Computational Fluid Dynamics Conference, 1991.

[231] Merkle, C. and Choi, Y. : Computation of Low Speed Compressible Flows with Time-Marching Methods, Int. J. Numer. Meth. Eng., 25, pp. 293–311, 1985.

[232] Li, X.S., and Gu, C.W. : An all-speed Roe-type scheme and its asymptotic analysis of low Mach number behavior, J. Comput. Phys. 227 (10), pp. 5144-5159, 2008.

[233] Li, X.S., and Gu, C.W. : Mechanism of Roe-type schemes for all-speed flows and its application, Comput. Fluids 38 (4), pp. 810-817, 2009.

[234] Birken, B., and Meister, A. : Stability of Preconditioned Finite Volume schemes at low Mach numbers, BIT Numer. Math., 45(3), pp. 463-480, 2005.

[235] Thornber, B., Drirakis, D., Williams, R.J.R, and Youngs, D. : On entropy generation and dissipation of kinetic energy in high-resolution shock-capturing schemes, J. Comput. Phys., 227 (10) pp. 4853-4872, 2008.

[236] Thornber, B., Mosedale, A., Drirakis, D., Williams, R.J.R, and Youngs, D. : An improved reconstruction method for compressible flows with low Mach number features, J. Comput. Phys., 227 (10) pp. 4873-4894, 2008.

[237] Dellacherie, S. : Analysis of Godunov type schemes applied to compressible Euler system at low Mach numbers, J. Comput. Phys., 229 (4), pp. 978-1016, 2010.

[238] Rieper, F. : A low-Mach number fix for Roe's approximate Riemann solver, J. Comput. Phys., 230 (13), pp. 5263-5287, 2011.

[239] Oßwald, K., Siegmund, A., Birken, P., Hannemann, V., and Meister, A. : L2Roe: a low dissipation version of Roe's approximate Riemann solver for low Mach numbers, Int. J. Numer. Meth. Fl., 81, pp. 71–86, 2016.

[240] Bruel, P., Delmas, S., Jung, J., and Perrier, V. : A low Mach correction able to deal with low Mach acoustics, J. Comput. Phys., 378 pp. 723-759, 2019.

[241] Roe, P. L., and Pike, J. : Efficient Construction and Utilisation of Approximate Riemann Solutions, Computing Methods in Applied Science and Engineering, INRIA North-Holland, pp. 499-518, 1984.

[242] Evans, J.A., and Hughes, T.J.R. : Variational Multiscale Analysis: A New Link Between Flux Correction, Total Variation, and Constrained Optimization, ICES Report 10-35, 2010.

[243] ten Eikelder, M.F.P. Akkerman I. : Variation entropy: a continuous local generalization of the TVD property using entropy principles, Comput. Method. Appl. M., 355, pp. 261-283, 2019.