Summer 2021

High-Order Positivity-Preserving L_2-Stable Spectral Collocation Schemes for the 3-D Compressible Navier-Stokes Equations

Johnathon Keith Upperman
Old Dominion University, JKUPPERMAN@EMAIL.WM.EDU

HIGH-ORDER POSITIVITY-PRESERVING L_2-STABLE SPECTRAL COLLOCATION SCHEMES FOR THE 3-D COMPRESSIBLE NAVIER-STOKES EQUATIONS

by

Johnathon Keith Upperman
B.S. Mathematics May 2012, College of William & Mary
M.A. Education May 2013, College of William & Mary

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTATIONAL AND APPLIED MATHEMATICS

OLD DOMINION UNIVERSITY
August 2021

Approved by:

Nail Yamaleev (Director)

Mark Carpenter (Member)

Fang Hu (Member)

John Adam (Member)

ABSTRACT

HIGH-ORDER POSITIVITY-PRESERVING L_2-STABLE SPECTRAL COLLOCATION SCHEMES FOR THE 3-D COMPRESSIBLE NAVIER-STOKES EQUATIONS

Johnathon Keith Upperman
Old Dominion University, 2021
Director: Dr. Nail Yamaleev

High-order entropy stable schemes are a popular method used in simulations with the compressible Euler and Navier-Stokes equations. The strength of these methods is that they formally satisfy a discrete entropy inequality which can be used to guarantee L_2 stability of the numerical solution. However, a fundamental assumption that is explicitly or implicitly used in all entropy stability proofs available in the literature for the compressible Euler and Navier-Stokes equations is that the thermodynamic variables (e.g., density and temperature) are strictly positive in the entire space–time domain considered. Without this assumption, any entropy stability proof for a numerical scheme solving the compressible Navier-Stokes equations is incomplete. Unfortunately, if the solution loses regularity the positivity assumption may fail to hold for a high-order entropy stable scheme unless special care is taken. To address this problem, we present a new class of positivity-preserving, entropy stable spectral collocation schemes for the 3-D compressible Navier-Stokes equations. The key distinctive property of our method is that it is proven to guarantee the pointwise positivity of density and temperature for compressible viscous flows. The new schemes are constructed by combining a positivity-violating entropy stable method of arbitrary order of accuracy and a novel first-order positivity-preserving entropy stable method discretized on the same Legendre-Gauss-Lobatto (LGL) collocation points used for the high-order counterpart. The proposed framework is general and can be directly extended to other SBP-SAT-type schemes. Numerical results demonstrating accuracy and positivity-preserving properties of the new spectral collocation schemes are presented for viscous and inviscid flows with nearly vacuum regions, very strong shocks, and contact discontinuities.

ACKNOWLEDGEMENTS

I would like to thank the Department of Defense for funding me through the Science, Mathematics and Research for Transformation (SMART) program. The associated internships through the SMART program were invaluable experiences for my professional development. In particular, I want to thank Dr. White and Dr. Garrett for mentoring me during those internships. I was also supported by the Virginia Space Grant Consortium (VSGC) Graduate STEM Research Fellowship during my studies. The annual conferences hosted by VSGC served as excellent opportunities for me to present my research and learn from other scholars.

I am thankful for the patience Professor Yamaleev has demonstrated as he has guided me throughout this project. His insight from experience has been an indispensable resource to me. I have grown as a direct result of his guidance. I want to also thank the other members of my committee and the feedback they have provided along the way–especially at my pre-defense.

I would like to also thank all of my professors at Old Dominion University (ODU) for the excellent instruction I received. In particular, the knowledge I gained from Professor Melrose's courses on tensor calculus and continuum mechanics have proved indispensable for conducting my project. Professor Hu's course on parallel computing was also essential for my project. I have also benefited from excellent graduate program directors. Thank you Professor Zhou and Cheng, you both were always incredibly helpful and professional.

This research was supported by the Research Computing clusters at Old Dominion University. Not only did I have access to excellent computing resources, I also greatly benefited from the excellent service provided by the Old Dominion University ITS help desk. In particular, without Min Dong's quick and effective help I would probably still be chasing my tail.

A journey this long is best undertaken with great traveling companions and I have certainly been blessed with some amazing fellow students to share this experience with. Charles Armstrong and Kumudu Gamage have been great friends and life lines for me throughout my time at Old Dominion University. The journey would have been much more difficult and

less meaningful without your help and friendship. I wish you both the best of luck on your future endeavors and thank you for your friendship.

Finally, I want to thank my best friend and companion, my wife Amy. I greatly appreciate your patience and encouragement throughout my years as a student. You have stepped in tremendously in raising our two children during my studies and I look forward to sharing that joy more evenly with you soon.

TABLE OF CONTENTS

Page

LIST OF TABLES viii

LIST OF FIGURES ix

Chapter

1. INTRODUCTION 1
   - 1.1 BACKGROUND 1
   - 1.2 THESIS ORGANIZATION AND SUMMARY OF RESULTS 5

2. THE 3-D COMPRESSIBLE NAVIER-STOKES EQUATIONS 8
   - 2.1 CARTESIAN COORDINATES 8
   - 2.2 CURVILINEAR COORDINATES 9
   - 2.3 ENTROPY STABILITY 10
   - 2.4 3-D BRENNER-NAVIER-STOKES EQUATIONS 19

3. OPERATORS AND NOTATION FOR REPRESENTING DISCRETE NUMERICAL SCHEMES 22
   - 3.1 HIGH-ORDER SBP OPERATORS 22
   - 3.2 NOTATION FOR DISCRETE TERMS 25

4. ENTROPY STABLE HIGH-ORDER DISCRETIZATIONS 30
   - 4.1 BASELINE 3-D SPECTRAL COLLOCATION SCHEME OF ARBITRARY ORDER OF ACCURACY 30
   - 4.2 BASELINE 3-D SPECTRAL COLLOCATION SCHEME WITH HIGH-ORDER ARTIFICIAL DISSIPATION 36

5. ARTIFICIAL VISCOSITY 38
   - 5.1 ENTROPY RESIDUAL 39
   - 5.2 RESIDUAL-BASED SENSOR 39
   - 5.3 COMPRESSION AND PRESSURE GRADIENT SENSORS 41
   - 5.4 ARTIFICIAL VISCOSITY COEFFICIENT 42

6. PRESERVING POSITIVITY OF THE THERMODYNAMIC VARIABLES 47
   - 6.1 FIRST-ORDER POSITIVITY-PRESERVING SCHEME 47
   - 6.2 ENTROPY STABLE VELOCITY AND TEMPERATURE LIMITERS FOR VISCOUS FLOWS 63
   - 6.3 HIGH-ORDER POSITIVITY–PRESERVING SCHEME 72
   - 6.4 IMPLEMENTATION DETAILS 89

Chapter Page

7. NUMERICAL RESULTS 93
7.1 NON-DIMENSIONAL 3-D COMPRESSIBLE NAVIER-STOKES EQUATIONS 94
7.2 1-D NUMERICAL RESULTS 95
7.3 2-D AND 3-D NUMERICAL RESULTS 105

8. SUMMARY AND CONCLUSIONS 139

REFERENCES 142

APPENDICES 149

A. ENTROPY STABILITY PROOFS 149
A.1 ENTROPY STABILITY OF FIRST-ORDER SYMMETRIC POSITIVE (SEMI-)DEFINITE FLUXES 149
A.2 ENTROPY STABILITY OF HIGH-ORDER VISCOUS FLUXES 151
A.3 ENTROPY STABLE BRENNER-NAVIER-STOKES FLUXES 154
A.4 ENTROPY STABILITY FOR GENERALIZED ENTROPIES 155
A.5 ENTROPY CONSERVATIVE INVISCID FLUXES 160

B. BOUNDARY CONDITIONS 162
B.1 FORM OF INVISCID BOUNDARY PENALTIES 162
B.2 FORM OF HIGH-ORDER VISCOUS BOUNDARY PENALTIES 163
B.3 FORM OF BOUNDARY PENALTIES FOR THE GRADIENT OF THE ENTROPY VARIABLES 163
B.4 FORM OF FIRST-ORDER BRENNER BOUNDARY PENALTIES 164
B.5 PENALTIES FOR SPECIFIC BOUNDARY CONDITIONS 164

C. CHOLESKY DECOMPOSITION BASED IDENTITIES FOR $\frac{\partial^2 \mathcal{S}}{\partial U^2}$ 167

VITA 169

LIST OF TABLES

Table Page

1. Final L_∞ and L_2 errors and their convergence rates obtained with the ESSC and PPESAD schemes for $p = 4, 5, 6$ for the viscous shock problem on uniform grids with K^3 number of elements. 106

2. Final L_∞ and L_2 errors and their convergence rates obtained with the ESSC and PPESAD schemes for $p = 4, 5, 6$ for the viscous shock problem on non-uniform grids with K^3 number of elements. 107

LIST OF FIGURES

Figure Page

1. Density (first row) and pressure (second row) profiles computed with the PPESAD-p6 scheme on uniform grids with 64, 128, 256 elements for the inviscid (left column) and viscous (right column) blast wave flows at $t = 0.038$. 96

2. The low- and high-order ($p = 6$) artificial viscosities obtained on the 256-element grid for the inviscid (left panel) and viscous blast wave flows at $t = 0.038$. 97

3. Time step histories for the inviscid (left panel) and viscous ($Re = 10^3$) blast wave flows computed with the PPESAD-p6 scheme on 256-element uniform grid. 98

4. Density (first row) and pressure (second row) profiles computed with the PPESAD-p5 scheme on uniform grids with 64, 128, 256 elements for the inviscid (left column) and viscous (right column) double rarefaction wave problems at $t = 0.15$. 99

5. The low- and high-order ($p = 5$) artificial viscosities obtained on the 256-element grid at $t = 0.0075$ for the inviscid (left panel) and viscous ($Re = 10^3$) double rarefaction wave problems. 100

6. Time step histories for the inviscid (left panel) and viscous ($Re = 10^3$) double rarefaction wave flows computed with the PPESAD-p5 scheme on 256-element uniform grid. 100

7. Density (first row) and pressure (second row) profiles computed with the PPESAD-p4 scheme on uniform grids with 64, 128, 256 elements for the inviscid (left column) and viscous (right column) LeBlanc flows. 102

8. The low- and high-order ($p = 4$) artificial viscosities obtained on the 256-element grid at $t = 0.4$ for the inviscid (left panel) and viscous LeBlanc flows. 103

9. Time step histories for the inviscid (left panel) and viscous ($Re = 10^5$) LeBlanc flows computed with the PPESAD-p4 scheme on 256-element uniform grid. 104

10. Initial density for the 3-D viscous shock on the 3^3 (left) and 6^3 (right) non-uniform grids. 108

11. Randomly generated low-order artificial viscosity (top-left), high-order artificial viscosity (top-right), and flux limiter (bottom-left) are displayed for the PPESAD-p4 solution of the freestream preservation problem at $t = 10$. 110

12. Time series plot (left) of the total entropy for the modified ESSC-p4 and PPES-p4 solutions of the isentropic vortex simulation. 111

Figure Page

13. The computational domain for the shock diffraction problem is bounded by the boundary lines 1-6. 112

14. The computational grid for the viscous shock diffraction problem. 113

15. Density (top row) and pressure (bottom row) are shown for the viscous shock diffraction problem with shock of Mach number 5.09. 114

16. Density (top row) and pressure (bottom row) are shown for the inviscid shock diffraction problem with shock of Mach number 5.09. 115

17. Flux limiter plot for the PPES-p4 solution of the inviscid (left) and viscous (right) shock diffraction problem with shock of Mach number 5.09. 116

18. Density (top row) and pressure (bottom row) are shown for the viscous (left) and inviscid (right) shock diffraction problem with shock of Mach number 200. 117

19. High-order (left column) and low-order (right column) artificial viscosity ($\log_{10}$) of the PPESAD-p5 solution of the inviscid (top row) and viscous (bottom row) shock diffraction problem with shock of Mach number 200. 118

20. The cumulative usage of the temperature (left) and V_1 (right) entropy stable limiters are shown for the PPESAD-p5 solution of the viscous shock diffraction problem with shock of Mach number 200. 119

21. The computational domain for the shock boundary layer interaction problem. 121

22. The computational grid used for the $Ma = 2.15$ SBLI problem. 122

23. Skin friction (left) and relative pressure (right) profiles at the solid wall boundary ($y = 0$) for the $Ma = 2.15$ oblique SBLI problem. 123

24. Density (top row), relative pressure (middle row), and Mach number (bottom row) are shown for the $Ma = 2.15$ oblique SBLI problem. 124

25. High-order aritficial viscosity is shown for the PPESAD-p4 solution of the $Ma = 2.15$ SBLI problem. 125

26. The medium resolution computational grid used for the $Ma = 6.85$ SBLI problem. 125

27. Skin friction (left) and relative pressure (right) profiles at the solid wall boundary ($y = 0$) for the $Ma = 6.85$ oblique SBLI problem. 126

Figure Page

28. Density (top), relative pressure (middle), and Mach number (bottom) of the PPESAD-p6 fine grid solution are shown for the $Ma = 6.85$ oblique SBLI problem. 127

29. High-order (top) and low-order (bottom) artificial viscosity of the PPESAD-p6 fine grid solution are shown for the $Ma = 6.85$ oblique SBLI problem. 128

30. Contour plot of $\log_{10} \|\hat{\mathbf{U}}_t\|_{L_2,k}$ for the PPESAD-p4 fine grid solution of the $Ma = 6.85$ oblique SBLI problem. 129

31. The coarse grid used for the hypersonic cylinder problem. 130

32. Time-averaged pressure (left) and skin friction (right) coefficients obtained with the PPESAD scheme on the no-slip boundary cylinder wall for the hypersonic cylinder simulation. 132

33. Density (top left), pressure (top right), vorticity (bottom left), and Mach number (bottom right) are shown for the PPESAD-p5 fine grid solution of the hypersonic cylinder problem. 133

34. High-order (left) and low-order (right) artificial viscosity ($\log_{10}$) of the PPESAD-p5 fine grid solution are shown for the hypersonic cylinder problem. 134

35. Time series plot of the total kinetic energy (left) and total entropy residual (right) for the ESSC-p4 and PPESAD-p4 solutions of the $Ma = 2$ TGV problem on grids 4^3, 16^3 and 64^3. 135

36. Time series plot of the total kinetic energy (left) and total entropy residual (right) for the PPES-p6 and PPESAD-p6 solutions of the $Ma = 10$ TGV problem on grids 4^3, 16^3 and 64^3. 135

37. Density (top row), pressure (middle row), and velocity component $V_1 = U$ (bottom row) are plotted for the PPESAD-p4 and ESSC-p4 solutions of the $Ma = 2$ TGV problem on the 16^3 (left column) and 64^3 (right column) grids. 137

38. Density (top row), pressure (middle row), and velocity component $V_1 = U$ (bottom row) are plotted for the $p = 5$ (left column) and $p = 6$ (right column) PPESAD and PPES solutions of the $Ma = 10$ TGV problem on the 64^3 grid. 138

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Computational fluid dynamics (CFD) is used to gain insight into many physical phenomena ranging from applications in aerospace, automobiles, microe-electronics, ships, and astrophysics [1, 2]. In many instances, researchers look for agreement between results obtained with CFD simulations and experimental results gathered from physical measurements; thus, CFD can play a key role in validating experimental results. Furthermore, CFD can be used to estimate physical information related to a given phenomenon that is not easily measured experimentally, by using more easily measured and theoretically derived quantities as conditions for a CFD simulation.

The majority of CFD codes used in industry applications are first- or second-order accurate Reynolds averaged Navier-Stokes (RANS) solvers [1, 2, 3]. Thus, third-order or higher accurate schemes are considered high-order in the aerospace community [2] and we adopt this convention. Although RANS simulations have found successful application in modeling steady viscous transonic and supersonic flows in aerospace applications, they lack the ability to reliably predict turbulent-separated flows where wind tunnel testing is still the preferred method of obtaining reliable design related data [1, 4]. In particular, first- and second-order methods tend to over dissipate unsteady vortices and hence perform poorly in tracking them over long periods of time unless computationally cost prohibitive meshes are used. The long term behavior of unsteady vortices are not negligible either. For example, they play a significant role in the aerodynamic forces experienced by helicopters [1, 2]. The shortcomings of RANS methods suggest that large-eddy simulations (LES) and direct numerical simulations (DNS) in conjunction with high-order methods are needed for more accurate and reliable simulations of complex turbulent flows [1, 2, 3]. However, there are still major obstacles preventing high-order LES and DNS from realizing their potential as robust CFD tools used

throughout industry. On the hardware side, high-order LES and DNS methods are currently too computationally expensive for current computer hardware. Indeed, the current and foreseeable hardware landscape places significant constraints on the design of numerical schemes beyond simply being amenable to parallel computing [1]. Other obstacles for high-order LES and DNS methods include: they are more complicated to implement than low-order methods, they are typically less robust and slower to converge to steady state, they require more memory for implicit time stepping, and there is a lack of robust high-order mesh generators [2].

High-order numerical algorithms have the potential to greatly improve the simulation accuracy of time dependent flows given their increased accuracy per degree of freedom, faster error convergence rate, and smaller numerical errors in terms of both dispersion and dissipation [2, 5, 6]. Unfortunately, high-order methods perform poorly in the presence of discontinuities or under-resolved features in the flow. In particular, large-magnitude features such as shock waves, contact discontinuities, strong thermal gradients, and thin shear layers (collectively referred to as 'sharp features') can lead to Gibbs oscillations that destroy the accuracy of the solution and may also lead to simulation breakdown [3].

Many numerical methods have been developed to stabilize high-order numerical schemes in the presence of sharp features. The methods are commonly referred to as 'shock capturing' methods (we adopt this convention), despite the fact that they are typically designed to stabilize the numerical solution for all sharp features. The literature on shock capturing methods is extensive with origins dating back over seventy years [7]. Methods for detecting regions with sharp features include employing some combination of physics-based sensors that look for strong compression (shock waves), or other high-gradient features such as shear and thermal layers (e.g., see [3, 8, 9, 10]). Other methods detect non-smooth features by inspecting the smoothness of the numerical solution (e.g., see [11, 12, 13, 14, 15, 16, 17]). Once a sharp feature has been detected, most numerical methods use some combination of filtering [18], limiters (e.g., see [19, 20, 21]) or artificial viscosity (e.g., see [11, 12, 16, 17, 22, 23]) to stabilize the solution. A typical pitfall of many stabilization methods is that they rely on heuristics and parameters that need to be tuned for individual problems and hence they lack sufficient robustness for industry adaptation.

A robust method for stabilizing a numerical scheme should have mathematically provable properties that are consistent with physical properties of the continuous equations. In this regard, numerical schemes for the compressible Navier-Stokes equations have been developed that discretely mimic the non-linear entropy stability properties (i.e. the second law of thermodynamics) of the continuous equations. Schemes that discretely (or, at least semi-discretely) possess a physical entropy inequality are called entropy stable. Assuming positive density and temperature, L_2 bounds on the conservative variables can be derived from entropy stability (see Section 2.3.2 and [24, 25, 26, 27]). Notice that the development of entropy stable schemes is not new for low-order methods (e.g., see [28, 29, 30]). However, in the last two decades much has been accomplished towards producing robust high-order entropy stable schemes (e.g., see [5, 12, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44] and references therein).

Although high-order entropy stable schemes formally satisfy the discrete entropy inequality, the entropy stability alone is not enough to guarantee L_2 stability of the numerical solution. A fundamental assumption that is explicitly or implicitly used in all entropy stability proofs available in the literature for the compressible Euler and Navier-Stokes equations is that the thermodynamic variables (e.g., density and temperature) are positive in the entire space–time domain considered. Note that this assumption is critical and without this assumption any discrete entropy stability proof for the compressible Navier-Stokes equations is incomplete. What makes the problem even more difficult is that no theoretical results on positivity of weak solutions of the compressible Navier-Stokes equations are currently available in the literature. The Navier-Stokes equations themselves do not guarantee positivity of density and temperature or impossibility of existence of vacuum regions. The lack of theoretical results on positivity of thermodynamic variables for the compressible Navier-Stokes equations hinders the development of robust high-order accurate numerical methods, thus indicating that new positivity-preserving numerical schemes must be developed for this class of problems.

Despite the numerous papers on high-order entropy stable methods for the compressible Navier-Stokes equations, papers on positivity-preserving methods for the compressible

Navier-Stokes equations are very rare. In [45], a positivity-preserving first-order finite difference scheme based on the Rusanov artificial dissipation has been developed for the 3-D compressible Navier-Stokes equations on Cartesian uniform grids. Note that the positivity proof in this paper relies on some special memetic properties of the 1st-order finite difference operators on uniform grids, which are not available for other discretizations or unstructured grids. Another first-order positivity-preserving scheme for the compressible Navier-Stokes equations was developed in [46]. This pressure correction scheme is based on staggered-in-space discretizations and solves the internal energy balance instead of the total energy conservation equation. The scheme is unconditionally stable and reduces to a projection method in the limit of the vanishing Mach number. Recently, Zhang presented a positivity-preserving high-order discontinuous Galerkin (DG) scheme for the compressible Navier-Stokes equations in [47]. This method provides only so-called weak positivity of the thermodynamic variables, where the positivity is guaranteed for element averages but not the individual collocation points that are directly used for approximation of the governing equations. A limiting procedure is applied in [47] to filter out negative values at the collocation points. The entropy stability of this filtering process is not clear and obscured by the presence of collocation points with potentially undefined entropy prior to limiting. Furthermore, the positivity-preserving DG scheme developed in [47] imposes very severe constraints on the time step, which is about an order of magnitude less than that of the baseline method for high degree polynomial bases. Note that the actual time step constraint may be much stiffer, because the lower bound on the artificial viscosity coefficient, which is required for providing the positivity, may grow dramatically, as the velocity gradients increase. Recently, a positivity-preserving scheme for the compressible Navier-Stokes equations has been proposed by Guermond et al. [48]. This approach relies on the invariant domain preserving [49] approximation of the Euler equations and the Strang's operator splitting technique that is at most 2nd-order accurate.

To our knowledge, there are no formally high-order numerical schemes that provide both entropy stability and pointwise positivity of the thermodynamic variables (e.g., density and temperature) for the 3-D compressible Navier-Stokes equations. This discouraging fact also implies that there are no L_2–stable high-order schemes for the compressible 3-D Navier-Stokes

equations. This lack of stability results is yet one more reason for why there are no robust high-order numerical methods for solving the compressible 3-D Navier-Stokes equations at high Mach and Reynolds numbers in realistic geometries.

1.2 THESIS ORGANIZATION AND SUMMARY OF RESULTS

In this thesis, we construct new, entropy stable, high-order spectral collocation flux-limiting schemes and artificial dissipation operators that provide pointwise positivity of density and temperature for the 3-D unsteady compressible Navier-Stokes equations at high Mach and Reynolds numbers on unstructured curvilinear grids. Much work has already been done to develop entropy stable spectral collocation element methods of arbitrary order of accuracy (e.g, see [5, 12, 31, 34, 36, 37, 38, 39]). These methods are similar to strong form, nodal discontinuous Galerkin spectral element methods which currently show the most promise for overcoming the aforementioned hardware challenges involved in solving the compressible Navier-Stokes equations on complex grids [1]. Hence, the specific high-order method we have modified serves as a good candidate for future DNS and LES models; however, we emphasize that many of the tools developed herein can be generally applied.

The key idea of the new methodology is to construct a first-order positivity-preserving entropy stable scheme defined on the same Legendre-Gauss-Lobatto collocation points used for the high-order operators and combine it with the high-order entropy stable spectral collocation scheme that does not in general guarantee positivity of the thermodynamic variables. Both the low- and high-order schemes use artificial dissipation operators that are based on the Brenner regularization of the compressible Navier-Stokes equations. In troubled elements where the density and/or temperature become negative, the proposed scheme combines the high-order fluxes with the corresponding 1st-order inviscid and artificial dissipation fluxes and imposes an appropriate constraint on the time step size to guarantee the pointwise positivity of both density and temperature. In contrast to the existing positivity-preserving schemes that rely on the monotonicity properties of the Rusanov-type artificial dissipation, the proposed method minimizes the amount of artificial dissipation required for pointwise positivity of the thermodynamic variables, which is critical for accurate prediction of viscous flows.

In Chapter 2, we will review the 3-D compressible Navier-Stokes equations in both Cartesian and curvilinear coordinates. In the review, we will cover entropy stability and the L_2 stability of the conservative variables that is implied by entropy stability. In particular, we present a new form of the bounds for the L_2 stability of the conservative variables. Lastly, we introduce the 3-D Brenner-Navier-Stokes equations, their entropy stability properties, and the general form of the Brenner regularization we use.

In Chapter 3, we introduce high-order diagonal-norm summation-by-parts operators and their key properties that are used in developing high-order entropy stable numerical schemes. In this chapter, we also introduce the notation used in this thesis for discrete terms.

In Chapter 4, we introduce an entropy stable 3-D spectral collocation scheme of arbitrary order of accuracy that lacks positivity properties. We also discuss how this scheme can be regularized using a high-order discretization of the Brenner viscous flux.

In Chapter 5, we construct the artificial viscosity coefficient that is used to control the amount of artificial dissipation added in each element. Critically, this artificial viscosity is built so that it depends on a combination of physics-based and residual-based sensors.

In Chapter 6, we present the tools we developed for preserving positivity while maintaining all desired properties: high-order accuracy in smooth regions, conservation, entropy stability, and freestream preservation on curvilinear grids. We end by discussing implementation details for the proposed high-order positivity-preserving flux-limiting scheme.

In Chapter 7, we present numerical results obtained from simulations using our proposed scheme. In particular, we show that our proposed scheme is more accurate for an under-resolved 3-D viscous shock than its non-regularized counterpart; furthermore, once enough resolution is added, the regularization in our scheme vanishes. We demonstrate the robustness of our scheme by simulating both an inviscid and viscous shock diffraction problem involving a shock of Mach number 200. The ability of our scheme to solve steady state problems is demonstrated by solving a shock wave / laminar boundary layer interaction problem with inflow Mach number of 6.85. A 2-D hypersonic cylinder problem with inflow Mach number of 17.605 is also solved on a curvilinear mesh, demonstrating the ability of the proposed scheme to maintain stability on curvilinear meshes. Finally, we solve the 3-D viscous Taylor-Green vortex problem with Mach number 10 to demonstrate how our numerical

scheme may perform for under-resolved turbulent flows.

CHAPTER 2

THE 3-D COMPRESSIBLE NAVIER-STOKES EQUATIONS

2.1 CARTESIAN COORDINATES

The 3-D compressible Navier-Stokes equations in conservation law form in the Cartesian coordinates (x_1, x_2, x_3) are given by

$$\begin{aligned}
&\frac{\partial \boldsymbol{U}}{\partial t} + \sum_{m=1}^{3} \frac{\partial \boldsymbol{F}_{x_m}}{\partial x_m} = \sum_{m=1}^{3} \frac{\partial \boldsymbol{F}^{(v)}_{x_m}}{\partial x_m}, \quad \forall\, (x_1, x_2, x_3) \in \Omega, \quad t \geq 0, \\
&\boldsymbol{U}\,(x_1, x_2, x_3, t) = \boldsymbol{G}^{(B)}\,(x_1, x_2, x_3, t)\,, \quad \forall\, (x_1, x_2, x_3) \in \Gamma, \quad t \geq 0, \\
&\boldsymbol{U}\,(x_1, x_2, x_3, 0) = \boldsymbol{G}^{(0)}\,(x_1, x_2, x_3, 0)\,, \quad \forall\, (x_1, x_2, x_3) \in \Omega,
\end{aligned} \tag{1}$$

where $\boldsymbol{U}$ is a vector of conservative variables, and $\boldsymbol{F}_{x_m}$, and $\boldsymbol{F}^{(v)}_{x_m}$ are the inviscid and viscous fluxes associated with the x_m coordinate, respectively. The boundary data, $\boldsymbol{G}^{(B)}$ and the initial condition, $\boldsymbol{G}^{(0)}$, are assumed to be bounded in $L_2 \cap L_\infty$. In addition, $\boldsymbol{G}^{(B)}$ is assumed to contain boundary data that are entropy stable in the sense that the corresponding boundary conditions satisfy the entropy inequality.

The vector of conservative variables is given as

$$\boldsymbol{U} = \begin{bmatrix} \rho & \rho V_1 & \rho V_2 & \rho V_3 & \rho E \end{bmatrix}^\top, \tag{2}$$

where ρ denotes the density, $\boldsymbol{V} = \begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix}^\top$ is the velocity vector, and E is the specific total energy. The specific total energy obeys $E = \frac{1}{\rho}\,(\mathrm{IE} + \mathrm{KE})$ where $\mathrm{IE} = \frac{P}{\gamma - 1}$ is the internal energy, P is the pressure, and $\mathrm{KE} = \rho \frac{\|\boldsymbol{V}\|^2}{2}$ is the kinetic energy. The inviscid fluxes, $\boldsymbol{F}_{x_m}, m = 1, 2, 3$, are given by

$$\boldsymbol{F}_{x_m} = \begin{bmatrix} \rho V_m & \rho V_m V_1 + \delta_{m,1} P & \rho V_m V_2 + \delta_{m,2} P & \rho V_m V_3 + \delta_{m,3} P & \rho V_m H \end{bmatrix}^\top, \tag{3}$$

where H is the specific total enthalpy and $\delta_{i,j}$ is the Kronecker delta.

The viscous fluxes, $\boldsymbol{F}^{(v)}_{x_m}, m = 1, 2, 3$, are defined as

$$\boldsymbol{F}^{(v)}_{x_m} = \begin{bmatrix} 0 & \tau_{1,m} & \tau_{2,m} & \tau_{3,m} & \sum_{i=1}^{3} \tau_{i,m} V_i + \kappa \frac{\partial T}{\partial x_m} \end{bmatrix}^\top . \tag{4}$$

The viscous stresses are given by

$$\tau_{i,j} = \mu \left(\frac{\partial V_i}{\partial x_j} + \frac{\partial V_j}{\partial x_i} - \delta_{i,j} \frac{2}{3} \sum_{n=1}^{3} \frac{\partial V_n}{\partial x_n} \right), \tag{5}$$

where $\mu(T)$ is the dynamic viscosity and $\kappa(T)$ is the thermal conductivity.

To close the Navier-Stokes equations, Eq. (1), the following constituent relations are used:

$$h = c_P T, \quad H = h + \frac{1}{2} \boldsymbol{V}^\top \boldsymbol{V}, \quad P = \rho R T, \quad R = \frac{R_u}{M_w},$$

where T is the temperature, R_u is the universal gas constant, M_w is the molecular weight of the gas, and c_P is the specific heat capacity at constant pressure. Finally, the specific thermodynamic entropy is given as

$$s = \frac{R}{\gamma - 1} \log \left(\frac{T}{T_\infty} \right) - R \log \left(\frac{\rho}{\rho_\infty} \right), \quad \gamma = \frac{c_P}{c_P - R}, \tag{6}$$

where T_∞ and ρ_∞ are reference temperature and density, respectively.

2.2 CURVILINEAR COORDINATES

To solve the Navier-Stokes equations in complex geometries, we recast these equations in curvilinear coordinates. An unstructured grid in the physical domain is generated by individually mapping a reference domain $(\xi_1, \xi_2, \xi_3) \in \hat{\Omega} = [-1, 1]^3$ onto each grid element in the physical domain $(x_1, x_2, x_3) \in \Omega$. Assuming that each individual transformation

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{\xi}) \tag{7}$$

is a diffeomorphism, it can be described by the following Jacobian matrix:

$$\frac{\partial(\boldsymbol{x})}{\partial(\boldsymbol{\xi})} = \begin{bmatrix} \frac{\partial x_1}{\partial \xi_1} & \frac{\partial x_1}{\partial \xi_2} & \frac{\partial x_1}{\partial \xi_3} \\ \frac{\partial x_2}{\partial \xi_1} & \frac{\partial x_2}{\partial \xi_2} & \frac{\partial x_2}{\partial \xi_3} \\ \frac{\partial x_3}{\partial \xi_1} & \frac{\partial x_3}{\partial \xi_2} & \frac{\partial x_3}{\partial \xi_3} \end{bmatrix}, \quad J = \left| \frac{\partial(\boldsymbol{x})}{\partial(\boldsymbol{\xi})} \right|.$$

In the present analysis, only static curvilinear unstructured grids are considered. For possible generalization of the proposed methodology to dynamic grids, we refer the reader to [34].

Taking into account that the following identity, which is called the geometric conservation law (GCL), holds [50]

$$\sum_{l=1}^{3} \frac{\partial}{\partial \xi_l} \left(J \frac{\partial \xi_l}{\partial x_m} \right) = 0, \quad m = 1, 2, 3, \tag{8}$$

the Navier-Stokes equations can be recast in the curvilinear coordinates (ξ_1, ξ_2, ξ_3) as follows:

$$\begin{aligned} &\frac{\partial J \boldsymbol{U}}{\partial t} + \sum_{m,l=1}^{3} \frac{\partial}{\partial \xi_l} \left(\boldsymbol{F}_{\xi_l} - \boldsymbol{F}_{\xi_l}^{(v)} \right) = 0, \\ &\boldsymbol{F}_{\xi_l} \equiv \sum_{m=1}^{3} J \frac{\partial \xi_l}{\partial x_m} \boldsymbol{F}_{x_m}, \quad \boldsymbol{F}_{\xi_l}^{(v)} \equiv \sum_{m=1}^{3} J \frac{\partial \xi_l}{\partial x_m} \boldsymbol{F}_{x_m}^{(v)}. \end{aligned} \tag{9}$$

Note that the GCL equation (8) guarantees that any physically meaningful constant vector of conservative variables $\boldsymbol{U} = \boldsymbol{const}$ is a solution of the Navier-Stokes equations (9). Though, the GCL equation (8) are satisfied exactly at the continuous level, this is not necessarily the case at the discrete level [50]. A discussion on how the corresponding metric coefficients should be discretized to satisfy the GCL equation is presented elsewhere (e.g., see [36, 50, 51, 52]).

2.3 ENTROPY STABILITY

A necessary condition for selecting a unique, physically relevant solution among possibly many weak solutions of Eq. (1) is the entropy inequality. It is well known that the entropy inequality holds for the Navier-Stokes equations in the Cartesian and curvilinear coordinates (e.g., see [34, 36]). For convenience, we repeat here the derivation of the entropy inequality for the Navier-Stokes equations for the case of time-independent curvilinear coordinates.

The compressible Navier-Stokes equations are equipped with a convex scalar entropy function $\mathcal{S}$ and the corresponding entropy flux $\mathcal{F}$, which are given by

$$\begin{aligned} \mathcal{S} &= -\rho s, \\ \boldsymbol{\mathcal{F}} &= -\rho s \boldsymbol{V}, \end{aligned} \tag{10}$$

where s is the thermodynamic entropy defined by Eq. (6) and $\boldsymbol{V}$ is the velocity vector. Note that the mathematical entropy $\mathcal{S}$ has the opposite sign from the thermodynamic entropy. Thus, the mathematical entropy across a shock decreases rather than increases. This nomenclature is used throughout the paper.

The entropy function $\mathcal{S}$ satisfies the following properties:

1. $\mathcal{S}(\boldsymbol{U})$ is strictly convex and its Hessian matrix, $\frac{\partial^2 \mathcal{S}}{\partial U^2}$, is positive definite provided that $\rho > 0$ and $T > 0$ $\forall \boldsymbol{x} \in \Omega$, thus yielding a one-to-one mapping from the conservative to entropy variables that are defined as follows:

$$\boldsymbol{W}^\top \equiv \frac{\partial \mathcal{S}}{\partial \boldsymbol{U}} = \begin{bmatrix} \frac{h}{T} - s - \frac{\boldsymbol{V}^\top \boldsymbol{V}}{2T} & \frac{V_1}{T} & \frac{V_2}{T} & \frac{V_3}{T} & -\frac{1}{T} \end{bmatrix}^\top. \tag{11}$$

2. The entropy variables satisfy the following compatibility relations for all inviscid fluxes of the compressible Navier-Stokes equations:

$$\boldsymbol{W}^\top \frac{\partial \boldsymbol{F}_{x_m}}{\partial x_m} = \boldsymbol{W}^\top \frac{\partial \boldsymbol{F}_{x_m}}{\partial \boldsymbol{U}} \frac{\partial \boldsymbol{U}}{\partial x_m} = \frac{\partial \mathcal{F}_{x_m}}{\partial \boldsymbol{U}} \frac{\partial \boldsymbol{U}}{\partial x_m} = \frac{\partial \mathcal{F}_{x_m}}{\partial x_m}, \quad m = 1, 2, 3, \tag{12}$$

   where $\mathcal{F}_{x_m}$ is the entropy flux in the m-th spatial direction.

3. The entropy variables symmetrize the compressible Navier-Stokes equations, which can be recast in terms of $\boldsymbol{W}$ as follows:

$$\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{W}} \frac{\partial \boldsymbol{W}}{\partial t} + \sum_{m=1}^{3} \frac{\partial \boldsymbol{F}_{x_m}}{\partial \boldsymbol{W}} \frac{\partial \boldsymbol{W}}{\partial x_m} = \sum_{l,m=1}^{3} \frac{\partial}{\partial x_l} \left(\mathsf{C}_{l,m} \frac{\partial \boldsymbol{W}}{\partial x_m} \right), \tag{13}$$

   with the symmetry conditions $\frac{\partial U}{\partial W} = \left(\frac{\partial U}{\partial W}\right)^\top$, $\frac{\partial \boldsymbol{F}_{x_m}}{\partial W} = \left(\frac{\partial \boldsymbol{F}_{x_m}}{\partial W}\right)^\top$, and $\mathsf{C}_{l,m} = (\mathsf{C}_{m,l})^\top$.

Furthermore, $\frac{\partial U}{\partial W} = \left(\frac{\partial^2 \mathcal{S}}{\partial U^2}\right)^{-1}$ is positive definite, and the matrices $\mathsf{C}_{l,m}$ satisfy the following inequality:

$$\sum_{l,m=1}^{3} \frac{\partial \boldsymbol{W}^\top}{\partial x_l} \mathsf{C}_{l,m} \frac{\partial \boldsymbol{W}}{\partial x_m} \geq 0, \quad \forall \frac{\partial \boldsymbol{W}}{\partial x_m} \in \mathbb{R}^5, \tag{14}$$

provided that $\rho > 0$ and $T > 0$ $\forall \boldsymbol{x} \in \Omega$. Note that the term on the right–hand side of Eq. (13) is a recast form of the viscous fluxes in terms of entropy variables, that is,

$$\boldsymbol{F}^{(v)}_{x_m} = \sum_{j=1}^{3} \mathsf{C}_{m,j} \frac{\partial \boldsymbol{W}}{\partial x_j}. \tag{15}$$

It has been proven by Godunov in [53] that if (1) is symmetrized by introducing new variables $\boldsymbol{W}$ and φ is a convex function of $\boldsymbol{W}$, then the entropy function and the corresponding entropy flux satisfy the following equations:

$$\varphi = \boldsymbol{W}^\top \boldsymbol{U} - \mathcal{S}, \tag{16}$$

$$\psi_m = \boldsymbol{W}^\top \boldsymbol{F}_{x_m} - \mathcal{F}_{x_m}, \quad m = 1, 2, 3, \tag{17}$$

where the functions φ and ψ_{x_m} are called the entropy potential and entropy potential flux, respectively.

2.3.1 ENTROPY INEQUALITY

We now show that, if temperature and density remain positive, the entropy inequality holds for the compressible Navier-Stokes equations in the time-independent curvilinear coordinates. Contracting Eq. (9) with the entropy variables given by Eq. (11) yields

$$\overbrace{\boldsymbol{W}^\top \frac{\partial J \boldsymbol{U}}{\partial \tau}}^{I} + \sum_{m,l=1}^{3} \overbrace{\boldsymbol{W}^\top \frac{\partial}{\partial \xi_l} \left(J \frac{\partial \xi_l}{\partial x_m} \boldsymbol{F}_{x_m} \right)}^{II} = \sum_{l,n=1}^{3} \overbrace{\boldsymbol{W}^\top \frac{\partial}{\partial \xi_l} \left(\hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n} \right)}^{III}. \tag{18}$$

The matrices $\hat{\mathsf{C}}_{l,n}$ on the right-hand side of Eq. (18) are given by

$$\hat{\mathsf{C}}_{l,n} \equiv \sum_{m,j=1}^{3} J \frac{\partial \xi_l}{\partial x_m} \mathsf{C}_{m,j} \frac{\partial \xi_n}{\partial x_j}. \tag{19}$$

For further details on how $\mathsf{C}_{m,j}$ and $\hat{\mathsf{C}}_{l,n}$ are constructed, see [52].

Using $\boldsymbol{W}^\top = \frac{\partial \mathcal{S}}{\partial \boldsymbol{U}}$, the term I in Eq. (18) can be manipulated as follows:

$$I = J \frac{\partial \mathcal{S}}{\partial \boldsymbol{U}} \frac{\partial \boldsymbol{U}}{\partial \tau} = \frac{\partial (J\mathcal{S})}{\partial \tau}. \tag{20}$$

Using the compatibility relations (Eq. (12)), the term II is reduced to

$$\begin{aligned} II = &\sum_{l,m=1}^{3} J \frac{\partial \xi_l}{\partial x_m} \frac{\partial \mathcal{S}}{\partial \boldsymbol{U}} \frac{\partial \boldsymbol{F}_{x_m}}{\partial \xi_l} + \boldsymbol{W}^\top \boldsymbol{F}_{x_m} \frac{\partial}{\partial \xi_l} \left(J \frac{\partial \xi_l}{\partial x_m} \right) = \\ &\sum_{l,m=1}^{3} \frac{\partial}{\partial \xi_l} \left(J \frac{\partial \xi_l}{\partial x_m} \mathcal{F}_{x_m} \right) + \sum_{m=1}^{3} \boldsymbol{W}^\top \boldsymbol{F}_{x_m} \sum_{l=1}^{3} \frac{\partial}{\partial \xi_l} \left(J \frac{\partial \xi_l}{\partial x_m} \right). \end{aligned} \tag{21}$$

The last term in Eq. (18) can be manipulated as follows:

$$III = \sum_{l,n=1}^{3} \frac{\partial}{\partial \xi_l} \left(\boldsymbol{W}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n} \right) - \frac{\partial \boldsymbol{W}}{\partial \xi_l}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n}. \tag{22}$$

Integrating Eq. (18) over the computational domain and taking into account Eqs. (20–22), we have

$$\begin{aligned} &\int_{\hat{\Omega}} \frac{\partial (J\mathcal{S})}{\partial \tau} \mathrm{d}\hat{\Omega} + \int_{\hat{\Omega}} \sum_{l,m=1}^{3} \frac{\partial}{\partial \xi_l} \left(J \frac{\partial \xi_l}{\partial x_m} \mathcal{F}_{x_m} \right) \mathrm{d}\hat{\Omega} = \\ &\int_{\hat{\Omega}} \left[\sum_{l,n=1}^{3} \frac{\partial}{\partial \xi_l} \left(\boldsymbol{W}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n} \right) - \frac{\partial \boldsymbol{W}}{\partial \xi_l}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n} \right] \mathrm{d}\hat{\Omega}, \end{aligned} \tag{23}$$

where we have used the GCL equations given by Eq. (8). Using the integration-by-parts

(IBP) formula, the above equation can be recast in the following form:

$$\begin{aligned}\int_{\hat{\Omega}} \frac{\partial(J\mathcal{S})}{\partial\tau}\mathrm{d}\hat{\Omega} = \sum_{l,m=1}^{3} \oint_{\Gamma} \left(\boldsymbol{W}^{\top}\hat{\mathsf{C}}_{l,m}\frac{\partial \boldsymbol{W}}{\partial\xi_m} - J\frac{\partial\xi_l}{\partial x_m}\mathcal{F}_{x_m} \right) n_{\xi_l}\mathrm{d}\hat{\Gamma} \\ - \sum_{l,n=1}^{3} \int_{\hat{\Omega}} \frac{\partial \boldsymbol{W}}{\partial\xi_l}^{\top}\hat{\mathsf{C}}_{l,n}\frac{\partial \boldsymbol{W}}{\partial\xi_n}\mathrm{d}\hat{\Omega},\end{aligned} \tag{24}$$

where n_{ξ_l} is the ξ_l component of the outward facing unit normal of the reference element.

Taking into account that the matrices $\hat{\mathsf{C}}_{l,m}$ satisfy Eq. (14) and assuming that the boundary conditions are entropy stable, Eq. (24) becomes

$$\int_{\hat{\Omega}} \frac{\partial(J\mathcal{S})}{\partial\tau}\mathrm{d}\hat{\Omega} = \frac{d}{d\tau}\int_{\hat{\Omega}} J\mathcal{S}\mathrm{d}\hat{\Omega} \leq 0. \tag{25}$$

Equation (25) represents the entropy inequality in the domain which is only valid under the assumption of positive density and temperature. Note that for the Euler equations with smooth solutions that satisfy the positivity assumption, Eq. (25) becomes an equality. Although the entropy inequality depends on the positivity assumption, there are no general positivity proofs for the 3-D compressible Navier-Stokes equations; hence, a proof that the entropy inequality holds generally is incomplete. The entropy inequality (25) is only a necessary condition, which is not by itself sufficient to guarantee convergence to a physically relevant weak solution of the Navier-Stokes equations.

2.3.2 L_2 BOUND ON $\boldsymbol{U}$

Not only is the entropy inequality (25) a necessary condition for the solution, it can also provide an L_2 bound on $\boldsymbol{U}$ (e.g., see [25, 26, 27]). Indeed, in Chapter 5 of [25] Dafermos shows that if a system of conservation laws possesses a convex entropy function, $\mathcal{S}(\boldsymbol{U})$, then a global bound on $\mathcal{S}$ implies an L_2 bound on the solution $\boldsymbol{U}$ [31].

Minimum eigenvalue of $\mathcal{S}_{UU}$ bound

We first present the derivation of the L_2 bound on $\boldsymbol{U}$ for the 3-D Navier-Stokes equations in curvilinear coordinates by following closely the derivation for the 1-D Navier-Stokes equations presented in [24]. The derived bound is written in terms of the minimum eigenvalue of $\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}$.

Define a new entropy $\bar{\mathcal{S}} = \mathcal{S} - \mathcal{S}(\boldsymbol{U}_0) - \mathcal{S}_{\boldsymbol{U}}(\boldsymbol{U}_0)^\top(\boldsymbol{U} - \boldsymbol{U}_0)$ where $\boldsymbol{U}_0$ is a user-defined constant non-zero state. In this derivation, we leave $\boldsymbol{U}_0$ unspecified, but assume that it is constructed so that certain constraints are satisfied–and we discuss example $\boldsymbol{U}_0$ which satisfy these constraints. Note that the associated entropy variables are

$$\bar{\boldsymbol{W}} \equiv \bar{\mathcal{S}}_{\boldsymbol{U}} = \mathcal{S}_{\boldsymbol{U}} - \mathcal{S}_{\boldsymbol{U}}(\boldsymbol{U}_0) = \boldsymbol{W} - \boldsymbol{W}_0 \tag{26}$$

and hence the compatability relation for the associated entropy flux is

$$\frac{\partial \bar{\mathcal{F}}_{x_m}}{\partial \boldsymbol{U}} = \bar{\mathcal{S}}_{\boldsymbol{U}} \frac{\partial \boldsymbol{F}_{x_m}}{\partial \boldsymbol{U}} = \frac{\partial \mathcal{F}_{x_m}}{\partial \boldsymbol{U}} - \boldsymbol{W}_0 \frac{\partial \boldsymbol{F}_{x_m}}{\partial \boldsymbol{U}}. \tag{27}$$

Contracting Eq. (9) with the new entropy variables given by Eq. (26) yields

$$\overbrace{\bar{\boldsymbol{W}}^\top \frac{\partial J\boldsymbol{U}}{\partial \tau}}^{I} + \sum_{m,l=1}^{3} \overbrace{\bar{\boldsymbol{W}}^\top \frac{\partial}{\partial \xi_l}\left(J\frac{\partial \xi_l}{\partial x_m}\boldsymbol{F}_{x_m}\right)}^{II} = \sum_{l,n=1}^{3} \overbrace{\bar{\boldsymbol{W}}^\top \frac{\partial}{\partial \xi_l}\left(\hat{\mathsf{C}}_{l,n}\frac{\partial \boldsymbol{W}}{\partial \xi_n}\right)}^{III}. \tag{28}$$

Using $\boldsymbol{W}^\top = \frac{\partial \bar{\mathcal{S}}}{\partial \boldsymbol{U}}$, the term I in Eq. (28) can be manipulated as follows:

$$I = J\frac{\partial \bar{\mathcal{S}}}{\partial \boldsymbol{U}}\frac{\partial \boldsymbol{U}}{\partial \tau} = \frac{\partial (J\bar{\mathcal{S}})}{\partial \tau}. \tag{29}$$

Using Eq. (26) and the GCL equations given by Eq. (8), the term II is reduced to

$$\begin{aligned} II &= \sum_{l,m=1}^{3} \frac{\partial}{\partial \xi_l}\left(J\frac{\partial \xi_l}{\partial x_m}\mathcal{F}_{x_m}\right) - \boldsymbol{W}_0^\top \frac{\partial}{\partial \xi_l}\left(J\frac{\partial \xi_l}{\partial x_m}\boldsymbol{F}_{x_m}\right) \\ &= \sum_{l,m=1}^{3} \frac{\partial}{\partial \xi_l}\left(J\frac{\partial \xi_l}{\partial x_m}\left(\mathcal{F}_{x_m} - \boldsymbol{W}_0^\top \boldsymbol{F}_{x_m}\right)\right). \end{aligned} \tag{30}$$

The last term in Eq. (28) can be manipulated as follows:

$$III = \sum_{l,n=1}^{3} \frac{\partial}{\partial \xi_l} \left(\bar{\boldsymbol{W}}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n} \right) - \frac{\partial \boldsymbol{W}}{\partial \xi_l}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n}. \tag{31}$$

Integrating Eq. (28) over the computational domain, taking into account Eqs. (29–31), using the integration-by-parts (IBP) formula, and comparing with Eq. (24) we have:

$$\begin{aligned}
\int_{\hat{\Omega}} \frac{\partial (J\bar{\mathcal{S}})}{\partial \tau} \mathrm{d}\hat{\Omega} &= \sum_{l,m=1}^{3} \oint_{\Gamma} \left(\bar{\boldsymbol{W}}^\top \hat{\mathsf{C}}_{l,m} \frac{\partial \boldsymbol{W}}{\partial \xi_m} - J \frac{\partial \xi_l}{\partial x_m} \left(\mathcal{F}_{x_m} - \boldsymbol{W}_0^\top \boldsymbol{F}_{x_m} \right) \right) n_{\xi_l} \mathrm{d}\hat{\Gamma} \\
&\quad - \sum_{l,n=1}^{3} \int_{\hat{\Omega}} \frac{\partial \boldsymbol{W}}{\partial \xi_l}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n} \mathrm{d}\hat{\Omega} \\
&= \int_{\hat{\Omega}} \frac{\partial (J\mathcal{S})}{\partial \tau} \mathrm{d}\hat{\Omega} \\
&\quad + \sum_{l,m=1}^{3} \oint_{\Gamma} \left(- \boldsymbol{W}_0^\top \hat{\mathsf{C}}_{l,m} \frac{\partial \boldsymbol{W}}{\partial \xi_m} + J \frac{\partial \xi_l}{\partial x_m} \boldsymbol{W}_0^\top \boldsymbol{F}_{x_m} \right) n_{\xi_l} \mathrm{d}\hat{\Gamma} \\
&= \int_{\hat{\Omega}} \frac{\partial (J\mathcal{S})}{\partial \tau} \mathrm{d}\hat{\Omega} + \sum_{l,m=1}^{3} \oint_{\Gamma} J \frac{\partial \xi_l}{\partial x_m} \boldsymbol{W}_0^\top \left(- \boldsymbol{F}_{x_m}^{(v)} + \boldsymbol{F}_{x_m} \right) n_{\xi_l} \mathrm{d}\hat{\Gamma},
\end{aligned} \tag{32}$$

where n_{ξ_l} is the ξ_l component of the outward facing unit normal of the reference element.

Assume that we have entropy stable boundary conditions so that Eq. (24) implies

$$\int_{\hat{\Omega}} \frac{\partial (J\mathcal{S})}{\partial \tau} \mathrm{d}\hat{\Omega} \leq - \sum_{l,n=1}^{3} \int_{\hat{\Omega}} \frac{\partial \boldsymbol{W}}{\partial \xi_l}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n} \mathrm{d}\hat{\Omega}. \tag{33}$$

Assume further that Eq. (32) implies $\int_{\hat{\Omega}} \frac{\partial (J\bar{\mathcal{S}})}{\partial \tau} \mathrm{d}\hat{\Omega} \leq \int_{\hat{\Omega}} \frac{\partial (J\mathcal{S})}{\partial \tau} \mathrm{d}\hat{\Omega}$. For example, consider thermally insulated no slip boundary walls:

$$\boldsymbol{V}|_{\Gamma} = 0, \quad \left. \sum_{l,m=1}^{3} \frac{\partial \xi_l}{\partial x_m} \frac{\partial T}{\partial x_m} n_{\xi_l} \right|_{\Gamma} = 0. \tag{34}$$

In this case, Eq. (24) implies

$$\int_{\hat{\Omega}} \frac{\partial (J\mathcal{S})}{\partial \tau} \mathrm{d}\hat{\Omega} = - \sum_{l,n=1}^{3} \int_{\hat{\Omega}} \frac{\partial \boldsymbol{W}}{\partial \xi_l}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n} \mathrm{d}\hat{\Omega}. \tag{35}$$

and if we select $\boldsymbol{V}_0 = 0$, $\rho_0 > 0$, and $T_0 > 0$, then we also have $\int_{\hat{\Omega}} \frac{\partial(J\bar{\mathcal{S}})}{\partial\tau}\mathrm{d}\hat{\Omega} = \int_{\hat{\Omega}} \frac{\partial(J\mathcal{S})}{\partial\tau}\mathrm{d}\hat{\Omega}$.

Through Taylor expansion of $\mathcal{S}$ around $\boldsymbol{U}_0$ we have

$$\mathcal{S}(\boldsymbol{U}) = \mathcal{S}(\boldsymbol{U}_0) + \mathcal{S}_{\boldsymbol{U}}(\boldsymbol{U}_0)^\top (\boldsymbol{U} - \boldsymbol{U}_0) + \frac{1}{2} (\boldsymbol{U} - \boldsymbol{U}_0)^\top \mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}(\boldsymbol{U}(\theta)) (\boldsymbol{U} - \boldsymbol{U}_0), \tag{36}$$

for some state $\boldsymbol{U}(\theta) = \boldsymbol{U}_0(1-\theta) + \theta\boldsymbol{U}$ where $\theta \in (0,1)$. Note that since density is additive and internal energy is concave, if we assume $\rho, T > 0$ then $\rho_0, T_0 > 0$ implies that $\rho(\theta), T(\theta) > 0$. Therefore, $\mathcal{S}^{\min}_{\boldsymbol{U}\boldsymbol{U}}(t) > 0$ where $\mathcal{S}^{\min}_{\boldsymbol{U}\boldsymbol{U}}(t)$ is the minimal eigenvalue of $\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}(\boldsymbol{U}(\theta), t)$ in space at time t. We should note that Eq. (36) is not a Taylor expansion in space (in which case, the necessary smoothness of the corresponding spatial derivatives would be highly questionable near discontinuous features such as shocks), but instead the partial derivatives are with respect to the conserved variables.

Notice that by definition $\bar{\mathcal{S}} = \mathcal{S} - \mathcal{S}(\boldsymbol{U}_0) - \mathcal{S}_{\boldsymbol{U}}(\boldsymbol{U}_0)^\top(\boldsymbol{U} - \boldsymbol{U}_0) = \frac{1}{2} (\boldsymbol{U} - \boldsymbol{U}_0)^\top \mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}(\boldsymbol{U}(\theta)) (\boldsymbol{U} - \boldsymbol{U}_0)$. Hence, if we integrate in time to $t = \mathcal{T}$ we have

$$\begin{aligned}
&\int_{\hat{\Omega}} \frac{\partial(J\bar{\mathcal{S}})}{\partial\tau}\mathrm{d}\hat{\Omega} \le \int_{\hat{\Omega}} \frac{\partial(J\mathcal{S})}{\partial\tau}\mathrm{d}\hat{\Omega} \le -\sum_{l,n=1}^{3} \int_{\hat{\Omega}} \frac{\partial \boldsymbol{W}}{\partial \xi_l}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n}\mathrm{d}\hat{\Omega}, \\
&\int_{\hat{\Omega}} J\bar{\mathcal{S}}(\boldsymbol{U}(\cdot,\mathcal{T}))\mathrm{d}\hat{\Omega} \le \int_{\hat{\Omega}} J\bar{\mathcal{S}}(\boldsymbol{U}(\cdot,0))\mathrm{d}\hat{\Omega} - \int_0^{\mathcal{T}} \sum_{l,n=1}^{3} \int_{\hat{\Omega}} \frac{\partial \boldsymbol{W}}{\partial \xi_l}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n}\mathrm{d}\hat{\Omega}, \\
&\int_{\hat{\Omega}} J\frac{1}{2} (\boldsymbol{U} - \boldsymbol{U}_0)^\top \mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}(\boldsymbol{U}(\theta(\mathcal{T}))) (\boldsymbol{U} - \boldsymbol{U}_0)\,\mathrm{d}\hat{\Omega} \le \int_{\hat{\Omega}} J\bar{\mathcal{S}}(\boldsymbol{U}(\cdot,0))\mathrm{d}\hat{\Omega} - \\
&\qquad \int_0^{\mathcal{T}} \sum_{l,n=1}^{3} \int_{\hat{\Omega}} \frac{\partial \boldsymbol{W}}{\partial \xi_l}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n}\mathrm{d}\hat{\Omega}.
\end{aligned} \tag{37}$$

Let

$$\mathcal{C}(\mathcal{T}) = \int_{\hat{\Omega}} J\bar{\mathcal{S}}(\boldsymbol{U}(\cdot,0))\mathrm{d}\hat{\Omega} - \int_0^{\mathcal{T}} \sum_{l,n=1}^{3} \int_{\hat{\Omega}} \frac{\partial \boldsymbol{W}}{\partial \xi_l}^\top \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n}\mathrm{d}\hat{\Omega} \le \int_{\hat{\Omega}} J\bar{\mathcal{S}}(\boldsymbol{U}(\cdot,0))\mathrm{d}\hat{\Omega}.$$

It follows from (37), that

$$2\mathcal{S}^{\min}_{\boldsymbol{U}\boldsymbol{U}}(\mathcal{T}) \int_{\hat{\Omega}} J\,(\boldsymbol{U} - \boldsymbol{U}_0)^\top (\boldsymbol{U} - \boldsymbol{U}_0)\,\mathrm{d}\hat{\Omega} \le 4\mathcal{C}(\mathcal{T}).$$

Since

$$\begin{aligned}\boldsymbol{U}^\top \boldsymbol{U} &= (\boldsymbol{U}-\boldsymbol{U}_0+\boldsymbol{U}_0)^\top(\boldsymbol{U}-\boldsymbol{U}_0+\boldsymbol{U}_0)\\ &= (\boldsymbol{U}-\boldsymbol{U}_0)^\top(\boldsymbol{U}-\boldsymbol{U}_0)+\boldsymbol{U}_0^\top\boldsymbol{U}_0+2(\boldsymbol{U}-\boldsymbol{U}_0)^\top\boldsymbol{U}_0\\ &\leq 2(\boldsymbol{U}-\boldsymbol{U}_0)^\top(\boldsymbol{U}-\boldsymbol{U}_0)+2\boldsymbol{U}_0^\top\boldsymbol{U}_0,\end{aligned}$$

we have that

$$\int_{\hat{\Omega}} J\,\boldsymbol{U}^\top(\mathcal{T})\,\boldsymbol{U}(\mathcal{T})\mathrm{d}\hat{\Omega} \leq 4\frac{\mathcal{C}(\mathcal{T})}{\mathcal{S}^{\min}_{\boldsymbol{U}\boldsymbol{U}}(\mathcal{T})}+2\int_{\hat{\Omega}} J\,\boldsymbol{U}_0^\top\boldsymbol{U}_0\mathrm{d}\hat{\Omega}, \tag{38}$$

which is the desired L_2 bound on the solution $\boldsymbol{U}$ at time $\mathcal{T}$.

An alternative approach using $LDL^\top$ decomposition

Given that the bound in Eq. (38) depends on the minimum eigenvalue of $\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}$ it would be nice to know what the minimum eigenvalue of $\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}$ is. To our knowledge, the full set of eigenvalues of $\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}$ are unknown (we are only aware of the eigenvalue $\frac{R}{P}$ of multiplicity 2). Hence, we present an alternative bound based on identities obtained from the $LDL^\top$ decomposition of $\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}$ (see Appendix C). We make all the same assumptions that led us to the bound given by (38).

It follows from (254) and (255) of Appendix C that

$$\begin{aligned}&(\boldsymbol{U}-\boldsymbol{U}_0)^\top\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}(\boldsymbol{U}(\theta(\mathcal{T})))\,(\boldsymbol{U}-\boldsymbol{U}_0) \geq \frac{(\rho-\rho_0)^2}{b_1(\theta)},\\ &(\boldsymbol{U}-\boldsymbol{U}_0)^\top\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}(\boldsymbol{U}(\theta(\mathcal{T})))\,(\boldsymbol{U}-\boldsymbol{U}_0) \geq \frac{(\rho V_i)^2}{b_{i+1}(\theta)}, \quad i=1,2,3,\\ &(\boldsymbol{U}-\boldsymbol{U}_0)^\top\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}(\boldsymbol{U}(\theta(\mathcal{T})))\,(\boldsymbol{U}-\boldsymbol{U}_0) \geq \frac{(\rho E-\rho_0 E_0)^2}{b_5(\theta)},\\ &b_1(\theta)=\frac{\rho(\theta)}{R}, \quad b_{i+1}(\theta)=\frac{P(\theta)+\rho(\theta)V_i^2(\theta)}{R}, \quad i=1,2,3,\\ &b_5(\theta)=\frac{P^2(\theta)\gamma+P(\theta)\rho(\theta)\|\boldsymbol{V}(\theta)\|^2\gamma+\left(\rho(\theta)\frac{\|\boldsymbol{V}(\theta)\|^2}{2}\right)^2}{R\rho(\theta)}.\end{aligned} \tag{39}$$

Let $b_i^{\max}(t)>0$ be the maximal value of $b_i(\boldsymbol{U}(\theta),t)$ in space at time t. It follows from (37),

that

$$\begin{aligned}
\frac{2}{b_1^{\max}(\mathcal{T})}\int_{\hat{\Omega}} J(\rho-\rho_0)^2 \mathrm{d}\hat{\Omega} \leq & 4\mathcal{C}(\mathcal{T}), \\
\frac{2}{b_{i+1}^{\max}(\mathcal{T})}\int_{\hat{\Omega}} J(\rho V_i)^2 \mathrm{d}\hat{\Omega} \leq & 4\mathcal{C}(\mathcal{T}), \quad i=1,2,3, \\
\frac{2}{b_5^{\max}(\mathcal{T})}\int_{\hat{\Omega}} J(\rho E-\rho_0 E_0)^2 \mathrm{d}\hat{\Omega} \leq & 4\mathcal{C}(\mathcal{T}).
\end{aligned} \tag{40}$$

Therefore,

$$\begin{aligned}
\int_{\hat{\Omega}} J\rho^2 \mathrm{d}\hat{\Omega} \leq & 4b_1^{\max}(\mathcal{T})\mathcal{C}(\mathcal{T}) + 2\int_{\hat{\Omega}} J\rho_0^2 \mathrm{d}\hat{\Omega}, \\
\int_{\hat{\Omega}} J(\rho V_i)^2 \mathrm{d}\hat{\Omega} \leq & 2b_{i+1}^{\max}(\mathcal{T})\mathcal{C}(\mathcal{T}), \qquad i=1,2,3, \\
\int_{\hat{\Omega}} J(\rho E)^2 \mathrm{d}\hat{\Omega} \leq & 4b_5^{\max}(\mathcal{T})\mathcal{C}(\mathcal{T}) + 2\int_{\hat{\Omega}} J(\rho_0 E_0)^2 \mathrm{d}\hat{\Omega}.
\end{aligned} \tag{41}$$

It is not ideal to have the $b_i^{\max}$ terms present in the bound, since e.g. the L_2 bound of density in Eq. (41) depends on the current L_∞ norm of density through $b_1^{\max}(\mathcal{T})$. The minimum eigenvalue of $\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}$ bound given by Eq. (38) likely has a similar issue though since $\frac{R}{P}$ is an eigenvalue of $\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}$. However, the presence of the $b_i^{\max}$ terms in the bounds are counter-balanced by the fact that

$$\mathcal{C}(\mathcal{T}) = \int_{\hat{\Omega}} J\bar{\mathcal{S}}(\boldsymbol{U}(\cdot,0))\mathrm{d}\hat{\Omega} - \int_0^{\mathcal{T}} \sum_{l,n=1}^{3} \int_{\hat{\Omega}} \frac{\partial \boldsymbol{W}}{\partial \xi_l}^{\top} \hat{\mathsf{C}}_{l,n} \frac{\partial \boldsymbol{W}}{\partial \xi_n} \mathrm{d}\hat{\Omega} \leq \int_{\hat{\Omega}} J\bar{\mathcal{S}}(\boldsymbol{U}(\cdot,0))\mathrm{d}\hat{\Omega}$$

can only decrease with time. Hence, the cumulative dissipation in time acts to pull down the weight of the current $b_i^{\max}$ value in the bound.

2.4 3-D BRENNER-NAVIER-STOKES EQUATIONS

In the present analysis, the Navier-Stokes equations are regularized by adding artificial dissipation in the form of the diffusion operator of the Brenner-Navier-Stokes equations introduced in [54, 55]. The Brenner-Navier-Stokes equations can be obtained from Eq. (1) by replacing the mass velocity $\boldsymbol{V}$ of the inviscid fluxes, $\boldsymbol{F}_{x_m}$, with the volume velocity $\boldsymbol{V}_{\mathrm{v}}$

which is given by $\boldsymbol{V}_{\mathrm{v}} = \boldsymbol{V} + \sigma\nabla\rho/\rho$, thus leading to

$$\frac{\partial \boldsymbol{U}}{\partial t} + \sum_{m=1}^{3} \frac{\partial \boldsymbol{F}_{x_m}}{\partial x_m} = \sum_{m=1}^{3} \frac{\partial \boldsymbol{F}^{(B)}_{x_m}}{\partial x_m}, \quad \forall\,(x_1, x_2, x_3) \in \Omega, \quad t \geq 0, \tag{42}$$

where σ is the volume diffusivity and the viscous fluxes, $\boldsymbol{F}^{(B)}_{x_m}, m = 1, 2, 3$, are defined as

$$\boldsymbol{F}^{(B)}_{x_m} = \boldsymbol{F}^{(v)}_{x_m} + \sigma \frac{\partial \rho}{\partial x_m} \begin{bmatrix} 1 & \boldsymbol{V} & E \end{bmatrix}^{\top}. \tag{43}$$

The entropy stability of the Brenner-Navier-Stokes equations is proven in a manner identical to what was was done in Section 2.3. The only difference is that the viscosity matrices $\mathsf{C}_{l,l},\ l = 1, 2, 3$ of Eq. (15) have the Brenner contribution but Eq. (14) still holds provided that $\rho > 0$ and $T > 0\ \forall \boldsymbol{x} \in \Omega$. We label the Brenner-Navier-Stokes viscosity matrices $\mathsf{C}^{(B)}_{m,j}$ and note that

$$\begin{aligned} &\boldsymbol{F}^{(B)}_{x_m} = \sum_{j=1}^{3} \mathsf{C}^{(B)}_{m,j} \frac{\partial \boldsymbol{W}}{\partial x_j}, \\ &\mathsf{C}^{(B)}_{l,m} = (\mathsf{C}^{(B)}_{m,l})^{\top}, \\ &\sum_{l,m=1}^{3} \frac{\partial \boldsymbol{W}^{\top}}{\partial x_l} \mathsf{C}^{(B)}_{l,m} \frac{\partial \boldsymbol{W}}{\partial x_m} \geq 0, \quad \forall \frac{\partial \boldsymbol{W}}{\partial x_m} \in \mathbb{R}^5. \end{aligned} \tag{44}$$

Despite that Eqs. (1) and (42) are very similar to each other, the Brenner-Navier-Stokes equations possess some remarkable properties that are not available for the Navier-Stokes equations. In contrast to Eq. (1), the Brenner-Navier-Stokes equations guarantee existence of a weak solution and uniqueness of a strong solution, ensure global-in-time positivity of the density and temperature, satisfy a large class of entropy inequalities, and is compatible with a minimum entropy principle [56, 24, 57, 58]. For further discussion on satisfying the large class of entropy inequalities, see appendix Section A.4.

Capitalizing on these remarkable properties of the Brenner-Navier-Stokes equations, we propose to regularize the Navier-Stokes equations by adding the following dissipation term

to Eq. (1):

$$\frac{\partial \boldsymbol{U}}{\partial t} + \sum_{m=1}^{3} \frac{\partial \boldsymbol{F}_{x_m}}{\partial x_m} = \sum_{m=1}^{3} \left[\frac{\partial \boldsymbol{F}_{x_m}^{(v)}}{\partial x_m} + \frac{\partial \boldsymbol{F}_{x_m}^{(AD)}}{\partial x_m} \right], \tag{45}$$

where the the artificial dissipation flux $\boldsymbol{F}_{x_m}^{(AD)}$ can be obtained from the viscous flux of the Brenner-Navier-Stokes equations,$\boldsymbol{F}_{x_m}^{(B)}$, by setting $\mu = \mu^{AD}$, $\sigma = c_\rho \mu^{AD}/\rho$, and $\kappa = c_T \mu^{AD}$. The coefficient μ^{AD} is an artificial viscosity and c_T and c_ρ are positive tunable coefficients. In this paper, we used $c_\rho = 0.9$ and $c_T = c_\rho \frac{c_P}{\gamma}$ in order to satisfy the necessary and sufficient condition given by Eq. (232) for the Brenner-Navier-Stokes viscosity flux (43) to satisfy (44) for a much larger classs of entropies [59, 60] as discussed in Section A.4.

By construction, Eq. (45) preserves some key properties of the Brenner-Navier-Stokes equations including conservation, entropy stability, and positivity of thermodynamic variables. Herein, we propose to develop a new numerical scheme that replicates these properties of the regularized Navier-Stokes equations (45) at the discrete level.

CHAPTER 3

OPERATORS AND NOTATION FOR REPRESENTING DISCRETE NUMERICAL SCHEMES

3.1 HIGH-ORDER SBP OPERATORS

3.1.1 HIGH–ORDER DIAGONAL-NORM SUMMATION-BY-PARTS OPERATORS

The derivatives in (9) are discretized by spectral collocation operators that satisfy the summation-by-parts (SBP) property [37, 61]. In the one-dimensional setting, this mimetic property is achieved by approximating the first derivative with a discrete operator, D, in the form:

$$D = \mathcal{P}^{-1}\mathcal{Q}. \tag{46}$$

The local mass $\mathcal{P}$ and stiffness $\mathcal{Q}$ matrices satisfy the following properties:

$$\begin{aligned} \mathcal{P} = \mathcal{P}^\top, \quad \mathbf{v}^\top \mathcal{P}\mathbf{v} > 0, \quad \forall \mathbf{v} \neq \mathbf{0}, \\ \mathcal{Q} = B - \mathcal{Q}^\top, \quad B = \mathrm{diag}(-1, 0, \ldots, 0, 1). \end{aligned} \tag{47}$$

Only diagonal-norm SBP operators are considered herein, which is critical for proving the entropy inequality at the discrete level.

In the one-dimensional setting, the physical domain is divided into K non-overlapping elements $[x_1^k, x_{N_p}^k]$ with $K+1$ nonuniformly distributed points, so that $x_1^k = x_{N_p}^{(k-1)}$. The discrete solution inside each element is defined on Legendre-Gauss-Lobatto (LGL) points, $\mathbf{x}_k = \left[x_1^k, \ldots, x_{N_p}^k\right]^\top$. These local points $\mathbf{x}_k$ are referred to as solution points. We represent non-discrete variables on this grid through projection $\mathbf{u}_k(t) = \left[\boldsymbol{u}_k(x_1^k, t), \ldots, \boldsymbol{u}_k(x_{N_p}^k, t)\right]^\top$. Using Eqs. (46, 47), it can be shown that the one-dimensional discrete derivative operator,

D, satisfies the following SBP property:

$$\mathbf{v}^T\mathcal{P}D\mathbf{u} = \mathbf{v}^T\mathcal{Q}\mathbf{u} = \mathbf{v}^T(B - \mathcal{Q}^T)\mathbf{u} = v_{N_p}u_{N_p} - v_1u_1 - (D\mathbf{v})^T\mathcal{P}\mathbf{u}. \tag{48}$$

3.1.2 TELESCOPIC FLUX FORM

Along with the solution points, we also define a set of intermediate points $\bar{\mathbf{x}}_k = \left[\bar{x}_0^k, \ldots, \bar{x}_{N_p}^k\right]^\top$ prescribing bounding control volumes around each solution point. These points referred to as flux points form a complementary grid whose spacing is precisely equal to the diagonal elements of the positive definite matrix $\mathcal{P}$ in Eq. (47), i.e.,

$$\Delta\bar{\mathbf{x}} = \mathcal{P}\mathbf{1}, \tag{49}$$

where $\bar{\mathbf{x}} = \left[\bar{x}_0, \ldots, \bar{x}_{N_p}\right]^\top$ is a vector of flux points, $\mathbf{1} = [1, \ldots, 1]^\top$, and Δ is an $N_p \times (N_p+1)$ matrix corresponding to the two-point backward difference operator [37, 33].

As has been proven in [62], all discrete SBP derivative operators can be recast into the following telescopic flux form:

$$\mathcal{P}^{-1}\mathcal{Q}\mathbf{f} = \mathcal{P}^{-1}\Delta\bar{\mathbf{f}}, \tag{50}$$

where $\bar{\mathbf{f}}$ is a pth-order flux vector defined at the flux points. The above telescopic flux form satisfies the following generalized SBP property:

$$\mathbf{v}^\top\mathcal{P}\mathcal{P}^{-1}\Delta\bar{\mathbf{f}} = \bar{f}_{N_p}v_{N_p} - \bar{f}_0v_1 - \sum_{j=1}^{N_p-1}\bar{f}_j(v_{j+1} - v_j) = \mathbf{v}^\top\tilde{B}\bar{\mathbf{f}} - \mathbf{v}^\top\tilde{\Delta}\bar{\mathbf{f}}, \tag{51}$$

where

$$\tilde{\Delta} = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

3.1.3 EXTENSION TO THREE DIMENSIONS

In the general three dimensional case on unstructured grids, simulations are performed on the union of piecewise smooth subdomains/elements $\mathbf{\Omega}_k(x^i)$ through the use of the reference domain $\hat{\mathbf{\Omega}}(\xi^i) = [-1,1]^3$ as described in section 2.2. The discrete solution points on $\hat{\mathbf{\Omega}}(\xi^i)$ are formed by taking tensor products of the one-dimensional LGL points on $[-1,1]$. We adopt the convention that a superscript "i" for the computational ξ^i or physical x^i coordinates refers to the associated direction and the subscript indexes the solution point on the element. For example, the $N_p = (p+1)^3$ solution points on the reference domain can be written as the tuples $(N = p+1)$

$$\left(\xi_1^1, \xi_1^2, \xi_1^3\right), \left(\xi_2^1, \xi_1^2, \xi_1^3\right), \ldots, \left(\xi_N^1, \xi_1^2, \xi_1^3\right), \left(\xi_1^1, \xi_2^2, \xi_1^3\right), \ldots, \left(\xi_N^1, \xi_N^2, \xi_N^3\right).$$

Letting only the ith component of $\vec{\xi}$ vary, we recover the one-dimensional LGL grid. From this perspective, we see how the flux points as defined in Section 3.1.2 for 1-D elements generalize to the 3-D reference element. For example, the flux points corresponding to letting only the ξ^1 component vary can be written as the tuples

$$\left(\bar{\xi}_0^1, \xi_1^2, \xi_1^3\right), \left(\bar{\xi}_1^1, \xi_1^2, \xi_1^3\right), \ldots, \left(\bar{\xi}_N^1, \xi_1^2, \xi_1^3\right), \left(\bar{\xi}_0^1, \xi_2^2, \xi_1^3\right), \ldots, \left(\bar{\xi}_N^1, \xi_N^2, \xi_N^3\right).$$

Every solution point has 6 surrounding flux points that again prescribe a bounding control volume. The 6 flux points surrounding the solution point $\vec{\xi}_{ijk}$ are $\vec{\xi}_{\overline{i-1}jk} = \left(\bar{\xi}_{i-1}^1, \xi_j^2, \xi_k^3\right)$ and $\vec{\xi}_{\bar{i},j,k} = \left(\bar{\xi}_i^1, \xi_j^2, \xi_k^3\right)$ in the ξ^1 direction and similarly written in the other two directions. The solution and flux points on the physical element $\mathbf{\Omega}_k(x^i)$ are obtained directly through the mapping $\vec{x}\left(\vec{\xi}\right)$ for that element. We adopt the convention of using the reference domain index for the physical domain when there is no ambiguity about which element we are on: i.e. we may write $\vec{x}\left(\vec{\xi}_{ijk}\right) = \vec{x}_{ijk}$.

Discretizing the curvilinear form of the Navier-Stokes equations given in Eq. (9) requires differentiating in the computational directions. For this purpose, we can extend the

one-dimensional SBP operators in Section 3.1.1 via tensor products to multiple spatial dimensions. The multidimensional tensor product operators are

$$\begin{aligned} D_{\xi^1} &= (D_N \otimes I_N \otimes I_N \otimes I_5)\,, \qquad \mathcal{P}_{\xi^1} = (\mathcal{P}_N \otimes I_N \otimes I_N \otimes I_5)\,, \\ \mathcal{P}_{\xi^1,\xi^2} &= (\mathcal{P}_N \otimes \mathcal{P}_N \otimes I_N \otimes I_5)\,, \qquad \mathcal{P} = (\mathcal{P}_N \otimes \mathcal{P}_N \otimes \mathcal{P}_N \otimes I_5)\,, \\ \widehat{\mathcal{P}} &= (\mathcal{P}_N \otimes \mathcal{P}_N \otimes \mathcal{P}_N)\,, \qquad \mathcal{P}_{\perp,\xi^1} = (I_N \otimes \mathcal{P}_N \otimes \mathcal{P}_N \otimes I_5)\,, \end{aligned} \tag{52}$$

with similar definitions for other directions and operators $\mathcal{Q}_{\xi^i}$, Δ_{ξ^i} and B_{ξ^i}. We will also use the following notation $\mathcal{P}_{ijk} = \mathcal{P}_{i,i}\mathcal{P}_{j,j}\mathcal{P}_{k,k}$ and $\mathcal{P}_{ij} = \mathcal{P}_{i,i}\mathcal{P}_{j,j}$ where $\mathcal{P}_{i,i}$ is the scalar ith diagonal entry of $\mathcal{P}_N$.

To simplify notation when dealing with domain boundaries, on the ath element we use the indicator function

$$\chi_a^{(BC)}(\vec{\xi}_{ijk}) = \begin{cases} 1, & \text{if } \vec{\xi}_{ijk} \text{ is on a domain boundary} \\ 0, & \text{otherwise} \end{cases}. \tag{53}$$

To pick off only terms at domain boundary faces we use

$$B_{\xi^1,a}^{(BC)} = \left(\text{diag}\left[\begin{array}{ccccc} -\chi_a^{(BC)}(\vec{\xi}_{1jk}) & 0 & \dots & 0 & \chi_a^{(BC)}(\vec{\xi}_{Njk}) \end{array}\right] \otimes I_N \otimes I_N \otimes I_5\right), \tag{54}$$

with identical definitions in the other computational directions.

Similarly, we define the indicator function

$$\chi_a^{(Int)}(\vec{\xi}_{ijk}) = \begin{cases} 1, & \text{if } \vec{\xi}_{ijk} \text{ is on an interior element face} \\ 0, & \text{otherwise} \end{cases}. \tag{55}$$

3.2 NOTATION FOR DISCRETE TERMS

Here we introduce the notation that is used to express the discrete terms in this dissertation.

3.2.1 2-D ARRAYS

Frequently, we need to describe a term which has a list of components (of fixed length) at every solution point or some subset of flux points on an element. Such 2-D arrays, we represent using bold font. For example, all of the following represent 2-D arrays used in the dissertation: $\mathbf{U}$, $\hat{\bar{\mathbf{f}}}_l$, $\hat{\mathbf{f}}_l^{(v)}$, $\hat{\mathbf{g}}_l$, etc. We use the convention that $\hat{\mathbf{g}}_l(\vec{\xi}_{ijk})$ refers to the list of components of $\hat{\mathbf{g}}_l$ at the solution point $\vec{\xi}_{ijk}$ and $\hat{\bar{\mathbf{f}}}_l(\vec{\xi}_{\bar{i}jk})$ refers to the components of $\hat{\bar{\mathbf{f}}}_l$ at the flux point $\vec{\xi}_{\bar{i}jk}$.

Although a term which has only a scalar value at each solution point (or some subset of the flux points) could be called a 1-D array, we treat them as 2-D arrays (with component list length of 1).

2-D arrays at flux points

We continue the convention of using an over bar (e.g. $\hat{\bar{\mathbf{f}}}_l$) to indicate quantities stored at some subset (possibly all) of the flux points. However, many 2-D arrays with an over bar in this dissertation are not stored at all of the flux points on an element. The ambiguity arises from the interpolation step of the 3-D equivalent of Eq. (50). The ambiguity for such quantities is removed since we only use the over bar for a flux quantity when we have differentiated it and the differentation always happens in the computational direction corresponding to the flux direction e.g. $\mathcal{P}_{\xi^1}^{-1}\Delta_{\xi^1}\hat{\bar{\mathbf{f}}}_{\mathbf{1}}$ where the quantity $\hat{\bar{\mathbf{f}}}_{\mathbf{1}}$ has values only at flux points of the form $\vec{\xi}_{\bar{i}jk}$ and $\hat{\bar{\mathbf{f}}}_{\mathbf{1}}$ represents a flux in the 1st computational direction i.e. the "1" tells you the direction of interpolation and the flux.

Operations on 2-D arrays

We combine 2-D arrays using two different operations. To illustrate these, we consider two arbitrary 2-D arrays stored at the solution points such that the components $\mathbf{A}(\vec{\xi}_{ijk})$ and $\mathbf{B}(\vec{\xi}_{ijk})$ or of equal length:

$$\begin{aligned}\mathbf{A} &= \begin{bmatrix} \mathbf{A}(\vec{\xi}_{111}) & \mathbf{A}(\vec{\xi}_{211}) & \dots & \mathbf{A}(\vec{\xi}_{NNN}) \end{bmatrix}, \\ \mathbf{B} &= \begin{bmatrix} \mathbf{B}(\vec{\xi}_{111}) & \mathbf{B}(\vec{\xi}_{211}) & \dots & \mathbf{B}(\vec{\xi}_{NNN}) \end{bmatrix}.\end{aligned} \tag{56}$$

The first operation is implied by juxtaposition and produces a scalar quantity:

$$\mathbf{A}^\top \mathbf{B} = \mathbf{B}^\top \mathbf{A} = \sum_{i,j,k=1}^{N} \mathbf{A}(\vec{\xi}_{ijk}) \cdot \mathbf{B}(\vec{\xi}_{ijk}), \tag{57}$$

where $\mathbf{A}(\vec{\xi}_{ijk}) \cdot \mathbf{B}(\vec{\xi}_{ijk})$ is the usual dot product. If the 2-D array $\mathbf{C}$ has components $\mathbf{C}(\vec{\xi}_{ijk})$ of length 1 at every solution point, then

$$\mathbf{C}^\top \mathbf{B} = \mathbf{B}^\top \mathbf{C} = \sum_{i,j,k=1}^{N} \mathbf{C}(\vec{\xi}_{ijk}) \mathbf{B}(\vec{\xi}_{ijk}), \tag{58}$$

where now $\mathbf{C}(\vec{\xi}_{ijk})$ simply scales the array $\mathbf{B}(\vec{\xi}_{ijk})$ and $\mathbf{C}^\top \mathbf{B}$ is a 1-D array with the same length as the length of $\mathbf{B}(\vec{\xi}_{ijk})$.

The second operation produces a new 2-D array with scalar components at each solution point:

$$\begin{aligned} \mathbf{C} &= \mathbf{A} \odot \mathbf{B} = \mathbf{B} \odot \mathbf{A}, \\ \mathbf{C}(\vec{\xi}_{ijk}) &= \mathbf{A}(\vec{\xi}_{ijk}) \cdot \mathbf{B}(\vec{\xi}_{ijk}). \end{aligned} \tag{59}$$

2-D arrays with multidimensional tensor product operators

The set of 2-D arrays where one dimension indexes the solution (or flux) points can be mapped through a bijection to a set of 4-D arrays. The mapping is

$$(\mathbf{A}(\vec{\xi}_{ijk}))_a \mapsto A^a_{ijk}, \tag{60}$$

where $(\mathbf{A}(\vec{\xi}_{ijk}))_a$ is the ath component of $\mathbf{A}(\vec{\xi}_{ijk})$. Since this mapping is a bijection, we use both representations interchangeably. Through this mapping, it is easiest to understand the action of the multidimensional tensor product operators. For example,

$$((D_{\xi^1} \mathbf{A})(\vec{\xi}_{ijk}))_a = \sum_{n=1}^{N_p} D_{in} A^a_{njk}, \tag{61}$$

where D_{in} is the (i, n) component of the 1-D derivative operator D of Eq. (46).

3.2.2 BLOCK DIAGONAL MATRICES

The 2-D arrays discussed in Section 3.2.1 store 1-D arrays at each solution point. When a term has a 2-D matrix at each solution point, we enclose the term in square brackets. For example, a term which is the 5 by 5 identity matrix at every solution point is written $[I]$. Note that $[I]$ is understood to be a block diagonal matrix with N_p blocks of dimension 5 by 5.

Assume that the block diagonal matrix $[I]$ stores the 5 by 5 matrix $I(\vec{\xi}_{ijk})$ at the solution point $\vec{\xi}_{ijk}$. Assume that the 2-D array $\mathbf{A}$ has $\mathbf{A}(\vec{\xi}_{ijk})$ of length 5. If we write $\mathbf{B} = [I]\mathbf{A}$, then that means $\mathbf{B}$ is the 2-D array such that $\mathbf{B}(\vec{\xi}_{ijk}) = I(\vec{\xi}_{ijk})\mathbf{A}(\vec{\xi}_{ijk})$ where the usual vector matrix multiplication is implied.

3.2.3 THE $\mathbf{1}_N$ AND $\mathbf{0}_N$ 2-D ARRAYS

The 2-D arrays $\mathbf{1}_n$ and $\mathbf{0}_n$ are useful in forming many identities where $\mathbf{1}_n(\vec{\xi}_{ijk}) = \begin{bmatrix} 1 & 1 & \dots \end{bmatrix}^\top$ and $\mathbf{0}_n(\vec{\xi}_{ijk}) = \begin{bmatrix} 0 & 0 & \dots \end{bmatrix}^\top$. The subscript n determines the length of $\mathbf{1}_n(\vec{\xi}_{ijk})$ and $\mathbf{0}_n(\vec{\xi}_{ijk})$ so that e.g. $\mathbf{1}_3(\vec{\xi}_{ijk}) = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$.

3.2.4 1-D ARRAYS

Section 3.2.1 describes the notation for 2-D arrays which contain 1-D arrays (e.g. vector of conserved variables) at each solution point. Sometimes, we wish to speak generally about a particular type of 1-D array or scalar without referencing solution points. In that case, we use bold italics for 1-D arrays and non-bold for scalars. For example, $\boldsymbol{u}$, $\boldsymbol{\nu}$, and $\boldsymbol{w}$ will be used to denote a vector of conserved, primitive, and entropy variables, respectively. The vector of primitive variables are: $\boldsymbol{\nu} = \begin{bmatrix} \rho & \vec{\boldsymbol{V}} & T \end{bmatrix}^\top$.

3.2.5 USE OF $\vec{}$

Since many of the arrays of interest are of length 5, we use the over arrow symbol (e.g. $\vec{x}$) to distinguish those 2-D arrays that store length 3 arrays at each solution point and 1-D arrays of length 3.

CHAPTER 4

ENTROPY STABLE HIGH-ORDER DISCRETIZATIONS

4.1 BASELINE 3-D SPECTRAL COLLOCATION SCHEME OF ARBITRARY ORDER OF ACCURACY

In this section, we present the baseline 3-D spectral collocation scheme [36, 37, 52]. While this scheme has multiple important properties that we will discuss, it is prone to spurious oscillations in the presence of discontinuities and lacks positivity properties. The purpose of this dissertation is to present a modification of this baseline scheme that addresses both of these shortcomings.

4.1.1 METRIC TERMS AND SATISFYING THE DISCRETE GCL EQUATION

The non-discrete GCL equation given in Eq. (8) can be enforced at the discrete level [50, 51]. To implement this method, we first calculate the arrays $\widehat{D}_{\xi^i}\mathbf{x}^j$ from which we compute the array of discrete pointwise Jacobians, $\mathbf{J}$, directly. Although one could then use matrix inversion or exact formulas to obtain the discrete $J\frac{\partial \xi^l}{\partial x^m}$, these will not satisfy the discrete GCL equation exactly. The discrete approximation of the scalar $J\frac{\partial \xi^l}{\partial x^m}$ at the solution point $\vec{\xi}_{ijk}$ is denoted $\hat{\mathbf{a}}^l_m(\vec{\xi}_{ijk})$. The block diagonal matrix $[\hat{a}^l_m]$ contains blocks with entries $\hat{\mathbf{a}}^l_m(\vec{\xi}_{ijk})I_{5\times 5}$ where $I_{5\times 5}$ is the identity matrix of size 5. The specific formulas for $\hat{\mathbf{a}}^l_m(\vec{\xi}_{ijk})$ are recorded elsewhere (e.g., see [36, 50, 51, 52]). Note that $\hat{\mathbf{a}}^l_m(\vec{\xi}_{ijk})$ is continuous at element interfaces and satisfies the following GCL equation

$$\sum_{l=1}^{3} D_{\xi^l}[\hat{a}^l_m]\mathbf{1}_5 = \mathbf{0}_5, \quad m = 1, 2, 3, \tag{62}$$

where $\mathbf{1}_5$ serves the purpose of transforming the diagonal matrices into vectors at each solution point. Equation (62) is the typical manner of writing the discrete GCL equation

(e.g., see [36, 52]), but we also make use of the equivalent statement

$$\sum_{l=1}^{3} \widehat{D}_{\xi^l} \hat{\mathbf{a}}_m^l = \mathbf{0}_1, \quad m = 1, 2, 3. \tag{63}$$

We use the notation $\hat{\vec{\mathbf{a}}}^i$ to refer to the array with components

$$\hat{\vec{\mathbf{a}}}^l(\vec{\xi}_{ijk}) = \left[\begin{array}{ccc} \hat{\mathbf{a}}_1^l(\vec{\xi}_{ijk}) & \hat{\mathbf{a}}_2^l(\vec{\xi}_{ijk}) & \hat{\mathbf{a}}_3^l(\vec{\xi}_{ijk}) \end{array} \right]^\top.$$

The discrete metric terms, $[\hat{a}_m^l]$, are used to discretely transform fluxes in the Cartesian coordinate system to contravariant form. Let $\mathbf{f}_{x^m}$ represent an array containing the numerical approximations to a given flux in the Cartesian direction x^m at the solution points. We write $\hat{\mathbf{f}}_l = \hat{\mathbf{f}}_{\xi^l} = \sum_{m=1}^{3} [\hat{a}_m^l] \mathbf{f}_{x^m}$ to denote the contravariant form of the flux in the ξ^l direction.

4.1.2 BASELINE SEMI-DISCRETE SCHEME AND ENTROPY ANALYSIS

Now we discuss the baseline 3-D spectral collocation scheme of arbitrary order of accuracy [36, 37, 52] that we intend to regularize. The semi-discrete form of the baseline scheme to solve Eq. (9) can be written

$$\hat{\mathbf{U}}_t + \sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\tilde{\mathbf{f}}}_l - D_{\xi^l} \hat{\mathbf{f}}_l^{(v)} = \sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \hat{\mathbf{g}}_l^{(BC)} + \mathcal{P}_{\xi^l}^{-1} \hat{\mathbf{g}}_l^{(Int)}, \tag{64}$$

where $\hat{\mathbf{U}} = [J]\,\mathbf{U}$, and the discontinuous formulation requires the use of the penalty $\hat{\mathbf{g}}_l^{(BC)}$ to enforce boundary conditions, and $\hat{\mathbf{g}}_l^{(Int)}$ to couple element interfaces. The block diagonal matrix $[J]$ contains blocks with entries $\mathbf{J}(\vec{\xi}_{ijk}) I_{5\times5}$ where $I_{5\times5}$ is the identity matrix of size 5. We write $\hat{\mathbf{g}}_l^{(BC)} = \hat{\mathbf{g}}_l^{(BC,I)} + \hat{\mathbf{g}}_l^{(BC,v)}$ where $\hat{\mathbf{g}}_l^{(BC,I)}$ represents the boundary penalties related to the inviscid terms and $\hat{\mathbf{g}}_l^{(BC,v)}$ the boundary penalties related to the viscous terms. Similarly, $\hat{\mathbf{g}}_l^{(Int)} = \hat{\mathbf{g}}_l^{(Int,I)} + \hat{\mathbf{g}}_l^{(Int,v)}$. In the following subsections, we will present the basic details of how the terms in Eq. (64) are constructed and conclude with a discussion of entropy stability. Since the exact form of $\hat{\mathbf{g}}_l^{(BC)}$ depends on the boundary condition in question, we will only discuss $\hat{\mathbf{g}}_l^{(Int)}$ here. Constructing entropy stable domain boundary penalties through $\hat{\mathbf{g}}_l^{(BC)}$ is non-trivial and an open area of research (e.g., see [5, 36]). In Appendix B, we discuss the

form of $\hat{\mathbf{g}}_l^{(BC)}$ that we used for obtaining the numerical results in Chapter 7.

Baseline viscous terms

The entropy stability of all high-order viscous terms in this dissertation depend on the discretization of the gradient of the entropy variables which can be written as [31, 36, 37]

$$\begin{aligned}
\boldsymbol{\Theta}_{x^j} &= \sum_{l=1}^{3} [\hat{a}_j^l][J^{-1}] \left(D_{\xi^l}\mathbf{w} + \mathcal{P}_{\xi^l}^{-1}\hat{\mathbf{g}}_l^{\boldsymbol{\Theta}} \right) \\
&= \mathbf{w}_{x^j} + \mathbf{g}_j^{\boldsymbol{\Theta}}, \\
\hat{\mathbf{g}}_1^{\boldsymbol{\Theta}}(\vec{\xi}_{ijk}) &= \frac{1}{2}\left(\delta_{1i}\Delta_1\mathbf{w}(\vec{\xi}_{0jk}) + \delta_{Ni}\Delta_1\mathbf{w}(\vec{\xi}_{Njk}) \right), \\
\Delta_1\mathbf{w}(\vec{\xi}_{ijk}) &= \mathbf{w}(\vec{\xi}_{i+1jk}) - \mathbf{w}(\vec{\xi}_{ijk}),
\end{aligned} \tag{65}$$

where we use similar definitions in each computational direction. The value $\mathbf{w}(\vec{\xi}_{N+1jk})$ is the value collocated with $\mathbf{w}(\vec{\xi}_{Njk})$ on the neighboring element or associated with a boundary condition (similar definition for $\mathbf{w}(\vec{\xi}_{0jk})$ and other directions). We use the Kronecker Delta function δ_{ij}.

The contravariant viscous fluxes, $\hat{\mathbf{f}}_l^{(v)}$, are constructed as follows

$$\hat{\mathbf{f}}_l^{(v)} = \sum_{m=1}^{3} [\hat{a}_m^l]\mathbf{f}_{x^m}^{(v)}, \quad \mathbf{f}_{x^m}^{(v)} = \sum_{j=1}^{3} [c_{m,j}^{(v)}]\boldsymbol{\Theta}_{x^j}. \tag{66}$$

For each $1 \le a, b \le 3$, $[c_{a,b}^{(v)}]$ is a block diagonal matrix with blocks that are 5×5, $[\left(c_{a,b}^{(v)}\right)^T] = [c_{b,a}^{(v)}]$, and $\sum_{a=1}^{3}\sum_{b=1}^{3} \mathbf{v}^T[c_{a,b}^{(v)}]\mathbf{v} \ge 0, \forall \mathbf{v}$ i.e. the full viscous tensor is symmetric positive semi-definite (SPSD). See [52] for the exact form of the $c_{a,b}^{(v)}$ matrices.

We further decompose the viscous interface penalties as the sum of an entropy conservative term and entropy dissipative term $\hat{\mathbf{g}}_l^{(Int,v)} = \hat{\mathbf{g}}_l^{(Int,v,C)} + \hat{\mathbf{g}}_l^{(Int,v,D)}$ that are now given. For all fixed $1 \le j, k \le N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$\begin{aligned}
\hat{\mathbf{g}}_1^{(Int,v,C)}(\vec{\xi}_i) &= \frac{\chi^{(Int)}(\vec{\xi}_i)}{2} \sum_{m=1}^{3} \hat{a}_m^1(\vec{\xi}_i) \left(\delta_{1i}\Delta_1\mathbf{f}_{x^m}^{(v)}(\vec{\xi}_0) + \delta_{Ni}\Delta_1\mathbf{f}_{x^m}^{(v)}(\vec{\xi}_N) \right) \\
\hat{\mathbf{g}}_1^{(Int,v,D)}(\vec{\xi}_i) &= \chi^{(Int)}(\vec{\xi}_i) \left(-\delta_{1i}\Lambda^{(v)}(\vec{\xi}_0, \vec{\xi}_1)\Delta_1\mathbf{w}(\vec{\xi}_0) + \delta_{Ni}\Lambda^{(v)}(\vec{\xi}_{N+1}, \vec{\xi}_N)\Delta_1\mathbf{w}(\vec{\xi}_N) \right)
\end{aligned} \tag{67}$$

where identical definitions hold for other computational directions. Note that the local discontinuous Galerkin (LDG) penalties, $\hat{\mathbf{g}}_l^{(Int,v,C)}$ and $\hat{\mathbf{g}}_l^{\mathbf{\Theta}}$, can be written in a slightly more general form involving an extra parameter (often denoted "α") [31, 37], but in this dissertation we simply use the symmetric LDG value of $\alpha = 0$ and avoid referencing α to reduce complexity.

For the exact form of the matrix $\Lambda^{(v)}$ that we used, see [31]. Here we note that $\Lambda^{(v)}$ is SPSD and is scaled by the physical viscosity so that it is zero for inviscid flows. Note that the entropy stability of the physical viscous terms is well established (see e.g. [36, 37, 52]) and for convenience we have included a proof in Section A.2 of the appendix given by Lemma 19.

Baseline inviscid terms

We begin by noting that a two point matrix $\bar{f}_{(S)}(\cdot,\cdot)$ is said to satisfy the entropy consistency condition [30] if for any two physical states $\boldsymbol{u}_a$ and $\boldsymbol{u}_b$ it satisfies:

$$(\boldsymbol{w}_a - \boldsymbol{w}_b)^\top \bar{f}_{(S)}(\boldsymbol{u}_a, \boldsymbol{u}_b) = \vec{\boldsymbol{\psi}}_a - \vec{\boldsymbol{\psi}}_b \quad . \tag{68}$$

The contravariant inviscid fluxes at the flux points, $\hat{\bar{\mathbf{f}}}_l$, are constructed as follows for all fixed $1 \le j,k \le N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$\begin{aligned}
&\text{for } 1 \le i \le N-1,\\
&\hat{\bar{\mathbf{f}}}_1(\vec{\xi}_i) = \sum_{R=i+1}^{N}\sum_{L=1}^{i} 2q_{L,R}\bar{f}_{(S)}(\mathbf{U}(\vec{\xi}_L),\mathbf{U}(\vec{\xi}_R))\frac{\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_L)+\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_R)}{2},\\
&\text{and for } \bar{i} \in \{0, N\},\\
&\hat{\bar{\mathbf{f}}}_1(\vec{\xi}_{\bar{i}}) = \bar{f}_{(S)}(\mathbf{U}(\vec{\xi}_{\bar{i}}),\mathbf{U}(\vec{\xi}_{\bar{i}}))\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_{\bar{i}}),
\end{aligned} \tag{69}$$

with similiar definitions for the other computational directions and for $\bar{f}_{(S)}(\cdot,\cdot)$ any two-point, consistent, entropy consistent inviscid interface flux can be used. We used the flux of Chandrashekar [35]; however, there are multiple options [44] and we show in Section 6.1.5 that our method for ensuring positivity is not dependent on a particular choice. The $\hat{\bar{\mathbf{f}}}_l$ fluxes in the other directions are handled similarly.

We further decompose the inviscid interface penalties as the sum of an entropy conservative term and entropy dissipative term $\hat{\mathbf{g}}_l^{(Int,I)} = \hat{\mathbf{g}}_l^{(Int,I,C)} + \hat{\mathbf{g}}_l^{(Int,I,D)}$ that are now given. For all fixed $1 \leq j,k \leq N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$\begin{aligned}
\hat{\mathbf{g}}_1^{(Int,I,C)}(\vec{\xi}_i) &= \chi^{(Int)}(\vec{\xi}_i)\left(-\delta_{1i}\left(\hat{\bar{\mathbf{f}}}_1(\vec{\xi}_1) - \bar{f}_{(S)}(\mathbf{U}(\vec{\xi}_1),\mathbf{U}(\vec{\xi}_0))\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_1)\right)\right. \\
&\quad \left. + \delta_{Ni}\left(\hat{\bar{\mathbf{f}}}_1(\vec{\xi}_N) - \bar{f}_{(S)}(\mathbf{U}(\vec{\xi}_{N+1}),\mathbf{U}(\vec{\xi}_N))\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_N)\right)\right), \\
\hat{\mathbf{g}}_1^{(Int,I,D)}(\vec{\xi}_i) &= \chi^{(Int)}(\vec{\xi}_i)\left(-\delta_{1i}M^{\mathcal{Y}}(\mathbf{U}(\vec{\xi}_0),\mathbf{U}(\vec{\xi}_1),\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_1))\Delta_1\mathbf{w}(\vec{\xi}_0)\right. \\
&\quad \left. + \delta_{Ni}M^{\mathcal{Y}}(\mathbf{U}(\vec{\xi}_{N+1}),\mathbf{U}(\vec{\xi}_N),\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_N))\Delta_1\mathbf{w}(\vec{\xi}_N)\right),
\end{aligned} \tag{70}$$

with identical definitions for the other computational directions. The SPSD matrix $M^{\mathcal{Y}}$ we used is the entropy dissipative characteristic flux proposed by Merriam [26] that dissipates each characteristic wave based on the magnitude of its eigenvalue. The exact form of this flux is discussed in Section 6.1.4.

The entropy stability of the high-order inviscid flux is proven by way of the following theorem found in [36, 52].

Theorem 1. *The contravariant inviscid flux defined in Eq.* (69) *is entropy conservative and satisfies the identity*

$$\sum_{l=1}^{3} \mathbf{w}^\top \mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\bar{\mathbf{f}}}_l = \sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\bar{\mathbf{F}}}_l^{(S)}, \tag{71}$$

where the local entropy flux $\hat{\bar{\mathbf{F}}}_l^{(S)}$ *is constructed as follows. For all fixed* $1 \leq j,k \leq N$ *and* $\vec{\xi}_i = \vec{\xi}_{ijk}$*, we have*

$$\begin{aligned}
&\textit{for } 1 \leq i \leq N-1, \\
&\hat{\bar{\mathbf{F}}}_1^{(S)}(\vec{\xi}_{\bar{i}}) = \sum_{R=i+1}^{N}\sum_{L=1}^{i} 2q_{L,R}\bar{F}_{(S)}(\mathbf{U}(\vec{\xi}_L),\mathbf{U}(\vec{\xi}_R))\frac{\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_L)+\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_R)}{2}, \\
&\textit{and for } \bar{i} \in \{0,N\}, \\
&\hat{\bar{\mathbf{F}}}_1(\vec{\xi}_{\bar{i}}) = \bar{F}_{(S)}(\mathbf{U}(\vec{\xi}_{\bar{i}}),\mathbf{U}(\vec{\xi}_{\bar{i}}))\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_{\bar{i}}),
\end{aligned} \tag{72}$$

with identical definitions in other directions and

$$\bar{F}_{(S)}(\boldsymbol{u}_a, \boldsymbol{u}_b) = \frac{(\boldsymbol{w}_a + \boldsymbol{w}_b)^\top}{2} \bar{f}_{(S)}(\boldsymbol{u}_a, \boldsymbol{u}_b) - \frac{\boldsymbol{\psi}_a + \boldsymbol{\psi}_b}{2}. \tag{73}$$

Furthermore, if the two-point nondissipative flux $\bar{f}_{(S)}(\cdot,\cdot)$ satisfies Tadmor's [30] criterion

$$\bar{f}_{(S)}(\boldsymbol{u}_a, \boldsymbol{u}_b) = \int_0^1 g(\boldsymbol{w}_a + \eta(\boldsymbol{w}_b - \boldsymbol{w}_a))\, d\eta, \quad g(\boldsymbol{w}_a) = f(\boldsymbol{u}_a), \tag{74}$$

then both $\sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\bar{\mathbf{f}}}_l$ and $\sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\bar{\mathbf{F}}}_l^{(S)}$ are design-order accurate.

Proof. See [36, 52]. □

Remark 1. The two-point entropy conservative flux of Chandrashekar [35] satisfies Eq. (68) but has not been shown to satisfy Eq. (74). A posteriori accuracy tests demonstrate design-order convergence for smooth problems.

From Theorem 1, it follows that

$$\sum_{l=1}^{3} \mathbf{w}^\top \mathcal{P} \mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\bar{\mathbf{f}}}_l = \sum_{l=1}^{3} \mathbf{1}_1^\top \widehat{\mathcal{P}}_{\perp,\xi^l} \widehat{B}_{\xi^l} \hat{\mathbf{F}}_l. \tag{75}$$

As has been shown in [36, 37, 52], entropy conservation follows quickly from Eq. (75) and the definition of $\hat{\mathbf{g}}_l^{(Int,I)}$. For reference, we give a proof of this claim in Lemma 21.

Entropy stability of the baseline scheme

The semi-discrete entropy stability analysis is performed by contracting the entropy variables, $\mathbf{w}^\top$, with the semi-discrete equation (64) and integrating over the entire domain. Given that $\boldsymbol{W}^\top \equiv \frac{\partial \mathcal{S}}{\partial \boldsymbol{U}}$, the semi-discrete time derivative is manipulated for diagonal-norm SBP operators as $\mathbf{w}^\top \mathcal{P} \hat{\mathbf{U}}_t = \mathbf{1}_1^\top \mathcal{P} \hat{\mathbf{S}}_t$ where $\hat{\mathbf{S}} = [J]\,\mathbf{S}$. The details of this analysis have been dissected elsewhere (e.g., see [31, 36, 37, 52]). Summing over the K elements in the global

domain and applying Lemmas 18, 19, and 21, we have

$$\begin{aligned}\sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\hat{\mathbf{S}}_t^k &= \sum_{k=1}^{K}\sum_{l=1}^{3}\mathbf{w}_k^\top\mathcal{P}_{\perp,\xi^l}\left[-\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_{l,k}+Q_{\xi^l}\hat{\mathbf{f}}_{l,k}^{(v)}+\hat{\mathbf{g}}_{l,k}^{(BC)}+\hat{\mathbf{g}}_{l,k}^{(Int)}\right]\\ &= \sum_{k=1}^{K}\sum_{l=1}^{3}\left[\mathbf{w}_k^\top\mathcal{P}_{\perp,\xi^l}\left[B_{\xi^l,k}^{(BC)}\hat{\mathbf{f}}_{l,k}^{(v)}+\hat{\mathbf{g}}_{l,k}^{(BC,v)}+\hat{\mathbf{g}}_{l,k}^{(BC,I)}\right]\right.\\ &+\left.\left(\hat{\mathbf{g}}_{l,k}^{(BC,\boldsymbol{\Theta})}\right)^\top\mathcal{P}_{\perp,\xi^l}\hat{\mathbf{f}}_{l,k}^{(v)}+\mathbf{1}_1^\top\widehat{\mathcal{P}}_{\perp,\xi^l}\widehat{B}_{\xi^l,k}^{(BC)}\hat{\mathbf{F}}_{l,k}\right]\\ &-\sum_{k=1}^{K}\left[H_k^{(v,D)}+L_k^{(Int,v,D)}+L_k^{(Int,I,D)}\right],\end{aligned}\tag{76}$$

where $H_k^{(v,D)}$, $L_k^{(Int,v,D)}$, and $L_k^{(Int,I,D)}$ are all non-negative and $L_k^{(Int,I,D)}$ is the entropy contribution from $\hat{\mathbf{g}}_l^{(Int,I,D)}$ as described in Lemma 18. Hence, we see that the baseline spectral collocation scheme of arbitrary order of accuracy is entropy stable up to entropy stable boundary conditions. Furthermore, the entropy stability at boundary faces depends solely on how $\hat{\mathbf{g}}_{l,k}^{(BC,v)}$, $\hat{\mathbf{g}}_{l,k}^{(BC,I)}$, and $\hat{\mathbf{g}}_{l,k}^{(BC,\boldsymbol{\Theta})}$ are chosen.

4.2 BASELINE 3-D SPECTRAL COLLOCATION SCHEME WITH HIGH-ORDER ARTIFICIAL DISSIPATION

The baseline 3-D spectral collocation scheme given by Eq. (64) discussed in Section 4.1 performs poorly in under-resolved smooth regions and at discontinuities such as shock waves. In such regions, the Gibbs oscillations generated by the scheme not only destroy the accuracy of the solution, but they can also lead to negative densities and temperatures. One approach to alleviate this problem is to regularize the scheme by an additional high-order viscous term controlled by an artificial viscosity. In [11], high-order artificial Brenner dissipation was used to regularize the 1-D form of the spectral collocation scheme given by Eq. (64). In this section, we generalize the method in [11] to the 3-D Navier-Stokes equations.

The baseline 3-D spectral collocation scheme with high-order Brenner dissipation is given by

$$\hat{\mathbf{U}}_t+\sum_{l=1}^{3}\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l-D_{\xi^l}\left[\hat{\mathbf{f}}_l^{(v)}+\hat{\mathbf{f}}_l^{(AD_p)}\right]=\sum_{l=1}^{3}\mathcal{P}_{\xi^l}^{-1}\left[\hat{\mathbf{g}}_l^{(AD_p)}+\hat{\mathbf{g}}_l\right],\tag{77}$$

where $\hat{\mathbf{g}}_l=\hat{\mathbf{g}}_l^{(BC)}+\hat{\mathbf{g}}_l^{(Int)}$ represents all of the penalties from Eq. (64). The high-order Brenner

dissipation terms, $\hat{\mathbf{f}}_l^{(AD_p)}$ and $\hat{\mathbf{g}}_l^{(AD_p)}$, are constructed in a manner identical to the viscous terms of Eq. (64) as discussed in Section 4.1.2. The only difference is that the contravariant Brenner fluxes, $\hat{\mathbf{f}}_l^{(AD_p)}$, are constructed as follows

$$\hat{\mathbf{f}}_l^{(AD_p)} = \sum_{m=1}^{3} [\hat{a}_m^l] \mathbf{f}_{x^m}^{(AD_p)}, \quad \mathbf{f}_{x^m}^{(AD_p)} = \sum_{j=1}^{3} [c_{m,j}^{(B)}] \boldsymbol{\Theta}_{x^j}, \tag{78}$$

where again we have that for each $1 \leq a, b \leq 3$, $[c_{a,b}^{(B)}]$ is a block diagonal matrix with blocks that are 5×5, $[\left(c_{a,b}^{(B)}\right)^T] = [c_{b,a}^{(B)}]$, and $\sum_{a=1}^{3}\sum_{b=1}^{3} \mathbf{v}^T [c_{a,b}^{(B)}] \mathbf{v} \geq 0, \forall \mathbf{v}$ i.e. the full viscous tensor is symmetric positive semi-definite (SPSD). The form of $[c_{a,b}^{(B)}]$ at a given solution point is straightforwardly obtained from $[c_{a,b}^{(v)}]$ (found in [52]) and Eq. (43). See appendix Section A.3.

To ensure consistency, maintain design-order accuracy in smooth resolved regions, and control the amount of dissipation added in under-resolved or discontinuous regions, we use $\boldsymbol{\mu} = \boldsymbol{\mu}^{AD}$ where the artificial viscosity, $\boldsymbol{\mu}^{AD}$, is described in Chapter 5. The mass and heat viscosity at each solution point are set as $\boldsymbol{\sigma}(\vec{\xi}_{ijk}) = c_\rho \boldsymbol{\mu}^{AD}(\vec{\xi}_{ijk}) / \boldsymbol{\rho}(\vec{\xi}_{ijk})$, and $\boldsymbol{\kappa}(\vec{\xi}_{ijk}) = c_T \boldsymbol{\mu}^{AD}(\vec{\xi}_{ijk})$ (see Section 2.4).

In Section 2.4, we discussed how Brenner's modification to the 3-D compressible Navier-Stokes Equations preserves the entropy stability estimate of the Navier-Stokes Equations (25). The same is true at the discrete level as follows from Lemma 19. The total entropy of the discrete scheme (77) evolves as follows

$$\begin{aligned}
\sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\mathbf{S}}_t^k = & \sum_{k=1}^{K}\sum_{l=1}^{3} \Big[\mathbf{w}_k^\top \mathcal{P}_{\perp,\xi^l} \left[B_{\xi^l,k}^{(BC)} \hat{\mathbf{f}}_{l,k}^{(v+AD_p)} + \hat{\mathbf{g}}_{l,k}^{(BC,v+AD_p)} + \hat{\mathbf{g}}_{l,k}^{(BC,I)} \right] \\
& + \left(\hat{\mathbf{g}}_{l,k}^{(BC,\boldsymbol{\Theta})} \right)^\top \mathcal{P}_{\perp,\xi^l} \hat{\mathbf{f}}_{l,k}^{(v+AD_p)} + \mathbf{1}_1^\top \widehat{\mathcal{P}}_{\perp,\xi^l} \widehat{B}_{\xi^l,k}^{(BC)} \hat{\mathbf{F}}_{l,k} \Big] \\
& - \sum_{k=1}^{K} \left[H_k^{(v+AD_p,D)} + L_k^{(Int,v+AD_p,D)} + L_k^{(Int,I,D)} \right],
\end{aligned} \tag{79}$$

where $\hat{\mathbf{f}}_{l,k}^{(v+AD_p)} = \hat{\mathbf{f}}_{l,k}^{(AD_p)} + \hat{\mathbf{f}}_{l,k}^{(v)}$ etc. We have simply added the entropy contribution (as described by Lemma 19) of the artificial Brenner dissipation to Eq. (76).

CHAPTER 5

ARTIFICIAL VISCOSITY

The pointwise, scalar artificial viscosity, $\boldsymbol{\mu}^{AD}$, controls the artificial dissipation and is used for both the high- and low-order artificial dissipation operators. The artificial viscosity is constructed based on the smoothness of the numerical solution and the physical behavior of the fluid. The steps for constructing the artificial viscosity are

1. Use Eq. (81) to form the entropy residual, $\mathbf{R}$, on every element.
2. Use Eq. (85) to form the entropy residual sensor, Sn, using the entropy residual on every element.
3. For elements where $Sn > 0$, compute the compression sensor, Cn, using Eq. (87) and pressure sensor, Pn, using (89).
4. For elements where $Sn > 0$, obtain the local reference length for viscosity, h^k, as described in Section 5.4.2.
5. Finally, obtain $\boldsymbol{\mu}^{AD}$ as described in Section 5.4.3 by first forming $\mu^k_{\max}$ where $\mu^k_{\max} = 0$ if $Sn = 0$.

Notice that for most simulations $Sn > 0$ for only a small subset of the elements in the domain at a given time step; hence, only a small fraction of elements will need to compute Cn, Pn, h^k, or $\mu^k_{\max}$ which greatly reduces the computational demand of the proposed algorithm.

In this section, we make use of the following globally defined parameter

$$\delta = \left(\frac{1}{K}\right)^{\frac{1}{d}}, \tag{80}$$

where d is the dimensionality of the grid and K is the total number of elements used. We also use L^* which is a grid-dependent length parameter. For simple grids, $L^* = \left(\sum_{i=1}^{K} V_i\right)^{\frac{1}{d}}$ where V_i is the volume on the ith element.

5.1 ENTROPY RESIDUAL

To make the discrete entropy residual consistent with the entropy stability properties of the scheme, we approximate the finite element residual of the entropy equation on the kth element as follows:

$$\mathbf{R} = \mathbf{w} \odot \hat{\mathbf{U}}_t^{base} - \sum_{l=1}^{3} \left(-\widehat{D}_{\xi^l} \hat{\mathbf{F}}_l + \widehat{D}_{\xi^l} \left(\mathbf{w} \odot \hat{\mathbf{f}}_l^{(v)} \right) - (\mathbf{\Theta}_{\xi^l}) \odot \hat{\mathbf{f}}_l^{(v)} \right), \tag{81}$$

where $\hat{\mathbf{U}}_t^{base}$ is $\hat{\mathbf{U}}_t$ in Eq. (64). Note that no alternative form of the penalty terms are included in $\mathbf{R} - \mathbf{w} \odot \hat{\mathbf{U}}_t^{base}$ because the interface penalties are already design-order and serve as a good measure of the error present in the numerical solution.

There are several advantages of this approach. First of all, Eq. (81) does not explicitly involve the time derivative term, thus eliminating the spurious entropy production due to an approximation of $\mathcal{S}_t$ which is usually not entropy stable. Also, if a high-order Runge-Kutta method is used to discretize $\mathcal{S}_t$, the above entropy residual is design-order accurate at any Runge-Kutta stage. Furthermore, the entropy residual given by Eq. (81) measures spurious entropy production due to the spatial discretization of the continuous governing equations and reaches its maximum values in regions where the numerical solution contains a discontinuity or is under-resolved. Another attractive feature of Eq. (81) is its computational efficiency. Half of the residual, $\mathbf{w} \odot \hat{\mathbf{U}}_t^{base}$, is essentially free since we always calculate $\hat{\mathbf{U}}_t^{base}$. The other half of the residual involves calculating the divergence of the entropy flux which is relatively cheap and for the viscous part we already have at our disposal $\mathbf{\Theta}_{\xi^l}$ and $\hat{\mathbf{f}}_l^{(v)}$ (from calculating $\hat{\mathbf{U}}_t^{base}$).

5.2 RESIDUAL-BASED SENSOR

Once we are in possession of the entropy residual for a given element, we form the entropy residual-based sensor. Before forming the element wise sensor, we form an intermediate point wise sensor as follows:

$$\mathbf{r}(\vec{\xi}_{ijk}) = \frac{\frac{\mathbf{R}(\vec{\xi}_{ijk})}{\mathbf{J}(\vec{\xi}_{ijk})}}{\max(\frac{\mathbf{R}(\vec{\xi}_{ijk})}{\mathbf{J}(\vec{\xi}_{ijk})}, \mathbf{d}(\vec{\xi}_{ijk}))}, \tag{82}$$

where at each solution point we use

$$\mathbf{d}(\vec{\xi}_{ijk}) = \left[\kappa\|\nabla\boldsymbol{\Theta}^5\|\mathbf{T} + \mu\sqrt{\mathbf{T}}\sqrt{\sum_{n=2}^{4}\|\nabla\boldsymbol{\Theta}^n\|} + \|\mathbf{F}\| + \boldsymbol{\rho}\delta\mathbf{c}\right]_{\vec{\xi}_{ijk}} \times \left[\frac{1}{\mathcal{P}_{i,i}} + \frac{1}{\mathcal{P}_{j,j}} + \frac{1}{\mathcal{P}_{k,k}}\right]\frac{2}{L^*}. \tag{83}$$

All quantities inside $[\ldots]_{\vec{\xi}_{ijk}}$ are evaluated at the solution point $\vec{\xi}_{ijk}$. We used the notation $\nabla\boldsymbol{\Theta}^a(\vec{\xi}_{ijk}) = \left[\begin{array}{ccc}\boldsymbol{\Theta}^a_{x^1} & \boldsymbol{\Theta}^a_{x^2} & \boldsymbol{\Theta}^a_{x^3}\end{array}\right]^\top_{\vec{\xi}_{ijk}}$ where $\boldsymbol{\Theta}^a_{x^j}(\vec{\xi}_{ijk})$ is the ath component of $\boldsymbol{\Theta}_{x^j}$ (see Eq. (65)) at $\vec{\xi}_{ijk}$. The quantity c is the speed of sound.

If we ignore the δ term which is present partly to avoid division by zero and vanishes with grid resolution, $\mathbf{d}(\vec{\xi}_{ijk})$ is directly related to the entropy residual $\mathbf{R}(\vec{\xi}_{ijk})$ defined in Eq. (81) since e.g. the identity

$$\mathbf{w}^\top(\vec{\xi}_{ijk})\mathbf{f}^{(v)}_{x^m}(\vec{\xi}_{ijk}) = -\kappa\boldsymbol{\Theta}^5_{x^m}(\vec{\xi}_{ijk})\mathbf{T}(\vec{\xi}_{ijk}) \tag{84}$$

holds discretely and in general the residual of the physical viscous terms is in direct proportion to the magnitude of the gradient of the entropy variable–except for the gradient of the first component of the entropy variables.

The entropy residual sensor for the kth element, Sn^k, is formed as follows:

$$\begin{aligned} Sn^k_0 &= \max(\mathbf{r}^k)^{\max(1,\frac{p-1}{p-1.5})}, \\ Sn^k &= \begin{cases} Sn^k_0, & \text{if } Sn^k_0 \geq \max(0.2,\delta) \\ 0, & \text{otherwise} \end{cases}, \end{aligned} \tag{85}$$

where p is the polynomial order. Note that Sn^k_0 is built so that if $\max(\mathbf{r}^k) < 1$ (which happens for under-resolved or discontinuous features that are not strong shocks), then the sensor becomes smaller as the polynomial order grows. In our experience, away from strong shocks, less dissipation is needed as the polynomial order grows. Also, note that although the sensor is discontinuous across elements, the artificial viscosity constructed from the sensor is continuous across elements.

5.3 COMPRESSION AND PRESSURE GRADIENT SENSORS

If $Sn^k > 0$ for the kth element, then we test to see whether the element is in a region of compression by obtaining the compression sensor, Cn^k, and we also investigate the behavior of the pressure gradient by obtaining the pressure sensor, Pn^k. The purpose of these sensors is to carefully identify regions where we can reduce the artificial viscosity used. Rather than directly computing all of the necessary gradients, we make use of the already calculated $\boldsymbol{\Theta}_{x^m}$ terms by forming

$$\boldsymbol{\nu}_{x^m}^{(\Theta)} = \left[\frac{\partial \nu}{\partial W}\right] \boldsymbol{\Theta}_{x^m}, \tag{86}$$

where $\nu = \begin{bmatrix} \rho & \vec{\boldsymbol{V}} & T \end{bmatrix}^\top$ is the array of primitive variables.

5.3.1 COMPRESSION SENSOR

The approximation of the velocity divergence obtained from $\boldsymbol{\nu}_{x^m}^{(\Theta)}$ is denoted $\left(\nabla \cdot \vec{\boldsymbol{V}}\right)^{(\Theta)}$. The compression sensor on the kth element, Cn^k, is calculated using the integral of the divergence over the element

$$\begin{aligned} Cn^k &= (Cn_0^k)^b \frac{\arctan\left[s(Cn_0^k - x_s)\right] + \frac{\pi}{2}}{\arctan\left[s(1 - x_s)\right] + \frac{\pi}{2}}, \\ Cn_0^k &= \max\left(\frac{-\mathbf{J}\widehat{\mathcal{P}}\left(\nabla \cdot \vec{\boldsymbol{V}}\right)^{(\Theta)}}{\mathbf{J}\widehat{\mathcal{P}}|\left(\nabla \cdot \vec{\boldsymbol{V}}\right)^{(\Theta)}| + \epsilon}, 0\right), \end{aligned} \tag{87}$$

where we found that $b = 0.1$, $x_s = 0.2$, and $s = 50$ worked well for all problems we considered. The goal is to keep Cn^k close to 1 unless the compression in the element is relatively weak ($Cn_0^k \lessapprox 0.2$) at which point we want Cn^k to decrease fairly rapidly.

5.3.2 PRESSURE SENSOR

It is well known that the momentum equations of the Navier-Stokes equations contracted with the velocity can be manipulated into an expression for the kinetic energy (e.g., see [56]).

In that expression, we see the following relationship

$$\frac{\partial \mathrm{KE}}{\partial t} = -\vec{\boldsymbol{V}} \cdot \nabla P - \mathrm{KE} \nabla \cdot \vec{\boldsymbol{V}} + \ldots, \tag{88}$$

where we have only written how the pressure gradient and velocity divergence change the kinetic energy, "KE." At the shock, we always have $-\mathrm{KE}\nabla \cdot \vec{\boldsymbol{V}} \geq 0$ but the pressure gradient term may have either sign. We have found that for shocks where $-\vec{\boldsymbol{V}} \cdot \nabla P \leq 0$, we can use less dissipation; hence, the pressure sensor aims to distinguish such regions.

Let $(\nabla \mathbf{P})^{(\Theta)}$ denote the array containing the gradient of the pressure at the solution points as obtained from the arrays $\boldsymbol{\nu}_{x^m}^{(\Theta)}$. Then, we define the pressure sensor, Pn^k, as follows

$$\begin{aligned} Pn^k &= \max\left(0, \frac{Pn_0^k}{Pn_d^k}\right), \\ Pn_0^k &= -\sum_{i,j,k=1}^{N} \mathcal{P}_{ijk} \mathbf{J}(\vec{\xi}_{ijk})\, \vec{\boldsymbol{V}}(\vec{\xi}_{ijk}) \cdot (\nabla \mathbf{P})^{(\Theta)}(\vec{\xi}_{ijk}), \\ Pn_d^k &= \epsilon + \sum_{i,j,k=1}^{N} \mathcal{P}_{ijk} \mathbf{J}(\vec{\xi}_{ijk}) \| \vec{\boldsymbol{V}}(\vec{\xi}_{ijk}) \| \| (\nabla \mathbf{P})^{(\Theta)}(\vec{\xi}_{ijk}) \|. \end{aligned} \tag{89}$$

5.4 ARTIFICIAL VISCOSITY COEFFICIENT

In this section, we discuss all the remaining steps for forming $\boldsymbol{\mu}^{AD}$. To minimize the amount of artificial dissipation added at strong discontinuities and under-resolved flow features, we follow the approach developed in [11] and determine the upper bound of the artificial viscosity, $\mu_{\max}$, based on the physics of the problem rather than numerics. $\mu_{\max}$ is only nonzero in elements where $Sn > 0$ and is constant for each element. As has been shown in [11, 63], the physical viscosity coefficient for the 1-D compressible Navier-Stokes equations at the Prandtl number $Pr = 3/4$ is related to a velocity jump across a shock wave and its thickness as follows:

$$\mu_* = \frac{3(\gamma+1)}{32\gamma} \rho_* \Delta v \delta_{\mathrm{sh}}, \tag{90}$$

where Δv is a velocity jump across the shock, the subscript $*$ denotes a value of the corresponding quantity at the sonic point, and δ_{sh} is the shock wave thickness. In [11], the shock

thickness δ_{sh} is replaced with an averaged grid spacing h^x/p and the velocity jump in Eq. (90) is estimated as the difference of velocities at neighboring collocation points $\Delta v = |v_{j+1} - v_j|$. Note, however, that this approach may not provide enough dissipation at strong discontinuities especially if the velocity field is nearly zero, which is quite common at the beginning of time integration. To overcome this problem, in [12] the velocity jump Δv in Eq. (90) was replaced by the jump in the maximum eigenvalue of the inviscid flux Jacobian. Here, we generalize the approach of [12] to the 3-D case.

5.4.1 LOCAL DERIVATIVES

The first step in forming $\mu_{\max}$ is to form locally defined derivatives $\frac{d_1}{d_1\xi^j}$ where the subscript in "d_1" is to indicate that these derivatives are taken using only nearest neighbors. Explicitly, for the ath velocity component, for all fixed $1 \le j,k \le N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$\frac{d_1 \boldsymbol{V}_a}{d_1\xi^1}(\vec{\xi}_i) = \begin{cases} \frac{\boldsymbol{V}_a(\vec{\xi}_{i+1}) - \boldsymbol{V}_a(\vec{\xi}_i)}{\xi^1_{i+1} - \xi^1_i}, & \text{if } i \le \frac{p+1}{2} \\ \frac{\boldsymbol{V}_a(\vec{\xi}_{i+1}) - \boldsymbol{V}_a(\vec{\xi}_{i-1})}{\xi^1_{i+1} - \xi^1_{i-1}}, & \text{else if } i \le \frac{p+1}{2} + 1 \\ \frac{\boldsymbol{V}_a(\vec{\xi}_i) - \boldsymbol{V}_a(\vec{\xi}_{i-1})}{\xi^1_i - \xi^1_{i-1}}, & \text{otherwise} \end{cases}, \tag{91}$$

where at the end points we also consider jumps formed using the collocated states (from the element that shares the entire face) and take whichever has a larger magnitude. Identical definitions follow for the other computational directions.

As in [12], we again use local derivatives of the square root of pressure:

$$\frac{d_1\sqrt{\mathbf{P}}}{d_1\xi^1}(\vec{\xi}_i) = \begin{cases} \sqrt{\gamma}\frac{\sqrt{\mathbf{P}(\vec{\xi}_{i+1})} - \sqrt{\mathbf{P}(\vec{\xi}_i)}}{\xi^1_{i+1} - \xi^1_i}, & \text{if } i \le \frac{p+1}{2} \\ \sqrt{\gamma}\frac{\sqrt{\mathbf{P}(\vec{\xi}_{i+1})} - \sqrt{\mathbf{P}(\vec{\xi}_{i-1})}}{\xi^1_{i+1} - \xi^1_{i-1}}, & \text{else if } i \le \frac{p+1}{2} + 1 \\ \sqrt{\gamma}\frac{\sqrt{\mathbf{P}(\vec{\xi}_i)} - \sqrt{\mathbf{P}(\vec{\xi}_{i-1})}}{\xi^1_i - \xi^1_{i-1}}, & \text{otherwise} \end{cases}. \tag{92}$$

These computational derivatives are transformed using the inverse metrics, $\hat{\mathbf{a}}^l_m$, to obtain $\frac{d_1 \boldsymbol{V}_a}{d_1 x^j}$ and $\frac{d_1\sqrt{\mathbf{P}}}{d_1 x^j}$.

5.4.2 LOCAL REFERENCE LENGTH FOR VISCOSITY

On the kth element where $Sn^k > 0$, we need to define a reference length, h^k, that will be used in forming $\mu^k_{\max}$. For nonuniform grids, defining a reference length to scale the artificial viscosity by is nontrivial especially if one seeks a viscosity that is continuous across elements. A length that is suitable for one element might be too small or large for a neighbor when adjacent elements have relatively large differences in sizes. Furthermore, large aspect ratios within a single element can make defining a good length for a given element difficult. We have found that the following approach works well for all the grids and problems we have tried.

Let $D_i\mathbf{x}$ contain the tangential derivative in the ξ^i direction at a given point: $D_i\mathbf{x}(\vec{\xi}_{ijk}) = \left[\begin{array}{ccc} \widehat{D}_{\xi^i}\mathbf{x}^1 & \widehat{D}_{\xi^i}\mathbf{x}^2 & \widehat{D}_{\xi^i}\mathbf{x}^3 \end{array}\right]^\top_{\vec{\xi}_{ijk}}$. We also define $\left[\frac{d_1\vec{\boldsymbol{V}}}{d_1\boldsymbol{x}}\right]$ as the block diagonal matrix with entries

$$\left[\frac{d_1\vec{\boldsymbol{V}}}{d_1\boldsymbol{x}}\right](\vec{\xi}_{ijk}) = \left[\begin{array}{ccc} \frac{d_1\boldsymbol{V}_1}{d_1x^1} & \frac{d_1\boldsymbol{V}_1}{d_1x^2} & \frac{d_1\boldsymbol{V}_1}{d_1x^3} \\ \frac{d_1\boldsymbol{V}_2}{d_1x^1} & \frac{d_1\boldsymbol{V}_2}{d_1x^2} & \frac{d_2\boldsymbol{V}_2}{d_1x^3} \\ \frac{d_1\boldsymbol{V}_3}{d_1x^1} & \frac{d_1\boldsymbol{V}_3}{d_1x^2} & \frac{d_3\boldsymbol{V}_3}{d_1x^3} \end{array}\right]_{\vec{\xi}_{ijk}} . \tag{93}$$

We begin by forming an array of reference lengths, $\mathbf{L}^b$, on the bth element defined at each solution point as follows

$$\begin{aligned} \mathbf{L}^b(\vec{\xi}_{ijk}) &= 2\left[\prod_{a=1}^{3}\|D_a\mathbf{x}\|^{\bar{\mathbf{E}}_a}\right]_{\vec{\xi}_{ijk}}, \\ \bar{\mathbf{E}}_a(\vec{\xi}_{ijk}) &= \mathbf{E}_a(\vec{\xi}_{ijk})\Big/\sum_{m=1}^{3}\mathbf{E}_a(\vec{\xi}_{ijk}), \\ \mathbf{E}_a(\vec{\xi}_{ijk}) &= \left\|\left[\frac{d_1\vec{\boldsymbol{V}}}{d_1\boldsymbol{x}}\right]\frac{D_a\mathbf{x}}{\|D_a\mathbf{x}\|}\right\|_{\vec{\xi}_{ijk}} + \epsilon, \quad a = 1,2,3, \end{aligned} \tag{94}$$

where the factor of 2 accounts for the fact that the computational domain has edges of length 2. Next, we use $\mathbf{L}^k$ to find an average length on the kth element

$$\hat{h}^k = \frac{\mathbf{1}_1^\top\widehat{\mathcal{P}}\mathbf{L}^k}{\mathbf{1}_1^\top\widehat{\mathcal{P}}\mathbf{1}_1}, \tag{95}$$

where the computational domain volume is fixed: $\mathbf{1}_1^\top \widehat{\mathcal{P}} \mathbf{1}_1 = 8$. From $\hat{h}^k$, we obtain h^k used for forming the artificial viscosity as follows

$$
\begin{aligned}
h^k &= \begin{cases} \left[\prod_{j \in N_k} h_i^v \right]^{\frac{1}{|N_k|}}, & \text{if } |N_k| > 0 \\ 0, & \text{otherwise} \end{cases}, \\
h_i^v &= \begin{cases} \left[\prod_{j \in I_i} \hat{h}^j \right]^{\frac{1}{|I_i|}}, & \text{if } |I_i| > 0 \\ 0, & \text{otherwise} \end{cases},
\end{aligned}
\tag{96}
$$

where I_i contains the element indices of all elements that touch the ith global vertex and have a nonzero $\hat{h}^k$, and N_k contains the global vertex indices of all vertices that touch the kth element and have a nonzero h_i^v.

5.4.3 OBTAINING $\boldsymbol{\mu}^{AD}$

If $Sn^k = 0$ on the kth element, we set $\mu_{\max}^k = 0$. If $Sn^k > 0$, we form $\mu_{\max}^k$ as follows

$$
\begin{aligned}
\mu_{\max}^k &= \frac{\left(h^k\right)^2}{p} \frac{3(\gamma+1)}{32\gamma} z_{Sn^k} \max_{1 \le i,j,k \le N} \Bigg[z_{Pn^k,Cn^k} \sqrt{\bar{\boldsymbol{\rho}} \sum_{b=1}^{3} \left(\frac{d_1 \sqrt{\mathbf{P}}}{d_1 x^b} \right)^2} \\
&\quad + \bar{\boldsymbol{\rho}} \left(\frac{\mathcal{P}_{1,1}}{2} \sqrt{\sum_{\substack{a,b=1, \\ a \neq b}}^{3} \left(\frac{d_1 \boldsymbol{V}_a}{d_1 x^b} \right)^2} + \min(\mathbf{Ma}, z_{Cn^k}) \left| \sum_{a=1}^{3} \frac{d_1 \boldsymbol{V}_a}{d_1 x^a} \right| \right) \Bigg]_{\vec{\xi}_{ijk}}, \\
\bar{\boldsymbol{\rho}}(\vec{\xi}_{ijk}) &= \left(\boldsymbol{\rho}(\vec{\xi}_{ijk}) \boldsymbol{\rho}(\vec{\xi}_{i+1jk}) \boldsymbol{\rho}(\vec{\xi}_{i-1jk}) \boldsymbol{\rho}(\vec{\xi}_{ij+1k}) \boldsymbol{\rho}(\vec{\xi}_{ij-1k}) \boldsymbol{\rho}(\vec{\xi}_{ijk+1}) \boldsymbol{\rho}(\vec{\xi}_{ijk-1}) \right)^{\frac{1}{7}}, \\
\mathbf{Ma}(\vec{\xi}_{ijk}) &= \frac{\| \vec{\boldsymbol{V}}(\vec{\xi}_{ijk}) \|}{\mathbf{c}(\vec{\xi}_{ijk})}, \\
z_{Sn^k} &= \min(0.5, 1.25(Sn^k - 0.2)) \ge 0, \quad z_{Cn^k} = g(Cn^k), \\
z_{Pn^k,Cn^k} &= \min(g(Pn^k), g(Cn^k)), \qquad g(x) = \frac{\mathcal{P}_{1,1}}{2}(1-x) + x,
\end{aligned}
\tag{97}
$$

where $\mathcal{P}_{1,1}$ is the smallest distance between flux points on the 1-D computational element, for the density average collocated states are used from elements sharing the same entire face, and $\mathbf{c}(\vec{\xi}_{ijk})$ is the speed of sound at the solution point $\vec{\xi}_{ijk}$. Note that $\mu_{\max}^k$ is built to be

most dissipative in the presence of shocks.

From $\mu^k_{\max}$, we obtain the following vertex viscosities

$$\mu^v_i = \max_{k \in I_i} \mu^k_{\max}, \tag{98}$$

where I_i contains the element indices of all elements that touch the ith global vertex. Hence, all elements that share the ith global vertex have the same viscosity, μ^v_i, stored at the ith global vertex. Therefore, the globally continuous artificial viscosity can be constructed by obtaining $\boldsymbol{\mu}^{AD}_k$ through tri-linear interpolation on the kth element using the 8 vertex viscosities, μ^v_i, obtained for the kth element.

CHAPTER 6

PRESERVING POSITIVITY OF THE THERMODYNAMIC VARIABLES

6.1 FIRST-ORDER POSITIVITY-PRESERVING SCHEME

We now present a positivity-preserving, entropy stable first-order scheme for the regularized 3-D compressible Navier-Stokes equations (45). This section is organized as follows. First, we present the first-order scheme and discuss the first-order inviscid terms. Then, we construct new Brenner and Merriam–Roe first-order fluxes used for density positivity and semi-discrete entropy stability. At the end of this section, we present a new constraint on the time step that ensures positivity of internal energy and show that the first-order scheme is entropy stable.

Note that $\boldsymbol{\nu}_i = \begin{bmatrix} \rho_i & \vec{\boldsymbol{V}}_i & T_i \end{bmatrix}^\top$ is used to denote the vector of primitive variables at the ith solution point. We also make substantial use of the logarithmic, harmonic, arithmetic, and geometric averages, which are denoted for quantities z_1 and z_2 by using the following subscript notation: z_L, z_H, z_A, and z_G, respectively. Note that the following inequalities hold for any $z_1 > 0$, $z_2 > 0$:

$$\begin{gathered} \min(z_1, z_2) \leq z_H \leq z_G \leq z_L \leq z_A \leq \max(z_1, z_2), \\ z_H < 2\min(z_1, z_2). \end{gathered} \tag{99}$$

6.1.1 FIRST-ORDER SCHEME

The first-order scheme on a given element is constructed on the same LGL points used for the high-order scheme. The first-order element treats solution points in a finite volume manner with the flux points acting as control volume edges. The first-order scheme can be written as

$$\hat{\mathbf{U}}_t + \sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \left[\hat{\tilde{\mathbf{f}}}_l^{(MR)} - \hat{\tilde{\mathbf{f}}}_{\hat{\sigma},l}^{(AD_1)} - \hat{\tilde{\mathbf{f}}}_l^{(AD_1)} \right] - D_{\xi^l} \hat{\mathbf{f}}_l^{(v)} = \sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \left[\hat{\mathbf{g}}_l + \hat{\mathbf{g}}_l^{(AD_1)} \right], \tag{100}$$

where if we compare to the baseline scheme in Eq. (64) we notice that the physical viscosity, $\hat{\mathbf{f}}_l^{(v)}$, is high-order and the high-order penalties, $\hat{\mathbf{g}}_l = \hat{\mathbf{g}}_l^{(BC)} + \hat{\mathbf{g}}_l^{(Int)}$, are still present. The inviscid terms have been replaced by the first-order approximation $\hat{\tilde{\mathbf{f}}}_l^{(MR)}$. We have also added first-order artificial dissipation. We discuss the new terms in the following sections.

6.1.2 FIRST-ORDER INVISCID TERM

We write the inviscid term as the sum of entropy conservative and entropy dissipative terms: $\hat{\tilde{\mathbf{f}}}_l^{(MR)} = \hat{\tilde{\mathbf{f}}}_l^{(EC)} - \hat{\tilde{\mathbf{f}}}_l^{(ED)}$. The entropy dissipative term, $\hat{\tilde{\mathbf{f}}}_l^{(ED)}$, acts to extend $\hat{\mathbf{g}}_l^{(Int,I,D)}$ in Eq. (70) to the interior flux points and we will discuss its exact form in Section 6.1.4.

The role of $\hat{\tilde{\mathbf{f}}}_l^{(EC)}$ is to replace the high-order inviscid fluxes, $\hat{\tilde{\mathbf{f}}}_l$, of Eq. (69) with a first-order approximation. The $\hat{\tilde{\mathbf{f}}}_l^{(EC)}$ contribution is formed as follows for all fixed $1 \leq j, k \leq N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$
\begin{aligned}
&\text{for } 1 \leq i \leq N-1, \\
&\hat{\tilde{\mathbf{f}}}_1^{(EC)}(\vec{\xi}_{\bar{i}}) = \bar{f}_{(S)}(\mathbf{U}(\vec{\xi}_i), \mathbf{U}(\vec{\xi}_{i+1}))\hat{\tilde{\vec{\mathbf{a}}}}^1(\vec{\xi}_{\bar{i}}), \\
&\hat{\tilde{\vec{\mathbf{a}}}}^1(\vec{\xi}_{\bar{i}}) = \sum_{R=i+1}^{N} \sum_{L=1}^{i} 2 q_{L,R} \frac{\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_L) + \hat{\vec{\mathbf{a}}}^1(\vec{\xi}_R)}{2}, \\
&\text{and for } \bar{i} \in \{0, N\}, \\
&\hat{\tilde{\mathbf{f}}}_1^{(EC)}(\vec{\xi}_{\bar{i}}) = \bar{f}_{(S)}(\mathbf{U}(\vec{\xi}_{\bar{i}}), \mathbf{U}(\vec{\xi}_{\bar{i}}))\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_{\bar{i}}),
\end{aligned}
\tag{101}
$$

where for $\bar{f}_{(S)}(\cdot, \cdot)$ any two-point, consistent, entropy consistent inviscid interface flux can be used. We used the flux of Chandrashekar [35]. In other computational directions, $\hat{\tilde{\mathbf{f}}}_l^{(EC)}$ has an identical definition. Comparing Eq. (101) with Eq. (69) we note that they are equivalent at the element faces ($\bar{i} \in \{0, N\}$) and only differ at the interior interface fluxes.

In Eq. (101), we used the high-order interpolation of the metric terms to the flux points which comes from the $\mathcal{Q}$ matrix. This was the only choice we could find that allowed us to prove entropy conservation and freestream preservation of $\hat{\tilde{\mathbf{f}}}_l^{(EC)}$ in Lemma 2. For the high-order inviscid term of the scheme given by Eq. (64), $\hat{\tilde{\mathbf{f}}}_l$, freestream preservation and entropy conservation follow directly from the metric terms satisfying the discrete GCL (see Eq. (62) and [36, 52]). In the following lemma, we show that the same holds for $\hat{\tilde{\mathbf{f}}}_l^{(EC)}$.

Lemma 2. *The inviscid term* $\hat{\tilde{\mathbf{f}}}_l^{(EC)}$ *given by Eq.* (101) *is freestream preserving and*

$$\begin{aligned}\sum_{l=1}^{3} \mathbf{w}^\top \mathcal{P}\mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\tilde{\mathbf{f}}}_l^{(EC)} &= \sum_{l=1}^{3} \mathbf{1}_1^\top \widehat{\mathcal{P}}_{\perp,\xi^l} \widehat{B}_{\xi^l} \hat{\mathbf{F}}_l \\ &= \sum_{l=1}^{3} \mathbf{w}^\top \mathcal{P}\mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\tilde{\mathbf{f}}}_l,\end{aligned} \tag{102}$$

where $\hat{\tilde{\mathbf{f}}}_l$ *is the high-order inviscid term of the scheme given by Eq.* (64) *and* $\sum_{l=1}^{3} \mathbf{1}_1^\top \widehat{\mathcal{P}}_{\perp,\xi^l} \widehat{B}_{\xi^l} \hat{\mathbf{F}}_l = \sum_{l=1}^{3} \mathbf{w}^\top \mathcal{P}\mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\tilde{\mathbf{f}}}_l$ *comes from Eq.* (75).

Proof. For freestream preservation, we assume a constant state on an element and want to show that this implies $\sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\tilde{\mathbf{f}}}_l^{(EC)} = \mathbf{0}_5$. Let $\boldsymbol{u}_0$ be the constant state on the element and note that for any pairing of solution points $\vec{\xi}_{ijk}$ and $\vec{\xi}_{abc}$ we have

$$\bar{f}_{(S)}(\mathbf{U}(\vec{\xi}_{ijk}), \mathbf{U}(\vec{\xi}_{abc})) = \bar{f}_{(S)}(\boldsymbol{u}_0, \boldsymbol{u}_0) = f(\boldsymbol{u}_0). \tag{103}$$

We look at the contribution of $\sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\tilde{\mathbf{f}}}_l^{(EC)}$ at a single point $\vec{\xi}_{ijk}$ on the element:

$$\begin{aligned}&\left[\sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\tilde{\mathbf{f}}}_l^{(EC)}\right](\vec{\xi}_{ijk}) \\ &= f(\boldsymbol{u}_0) \left[\frac{\hat{\tilde{\mathbf{a}}}^1(\vec{\xi}_{ijk}) - \hat{\tilde{\mathbf{a}}}^1(\vec{\xi}_{\overline{i-1}jk})}{\mathcal{P}_{i,i}} + \frac{\hat{\tilde{\mathbf{a}}}^2(\vec{\xi}_{ijk}) - \hat{\tilde{\mathbf{a}}}^2(\vec{\xi}_{i\overline{j-1}k})}{\mathcal{P}_{j,j}} + \frac{\hat{\tilde{\mathbf{a}}}^3(\vec{\xi}_{ijk}) - \hat{\tilde{\mathbf{a}}}^3(\vec{\xi}_{ij\overline{k-1}})}{\mathcal{P}_{k,k}}\right] \\ &= f(\boldsymbol{u}_0) \left[\frac{\sum_{n=1}^{N} q_{i,n} \hat{\tilde{\mathbf{a}}}^1(\vec{\xi}_{njk})}{\mathcal{P}_{i,i}} + \frac{\sum_{n=1}^{N} q_{j,n} \hat{\tilde{\mathbf{a}}}^2(\vec{\xi}_{ink})}{\mathcal{P}_{j,j}} + \frac{\sum_{n=1}^{N} q_{k,n} \hat{\tilde{\mathbf{a}}}^3(\vec{\xi}_{ijn})}{\mathcal{P}_{k,k}}\right] \\ &= \begin{bmatrix} 0 & \dots & 0 \end{bmatrix}^\top,\end{aligned} \tag{104}$$

where the second equality follows from an argument identical to the one used to prove Eq. (50) and the last equality follows from the metric terms satisfying the discrete GCL given by Eq. (63).

The entropy identity follows from scaling by the mass matrix and contracting with the

entropy variables:

$$\begin{aligned}\sum_{l=1}^{3}\mathbf{w}^\top\mathcal{P}\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l^{(EC)} &= \sum_{l=1}^{3}\mathbf{w}^\top\mathcal{P}_{\perp,\xi^l}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l^{(EC)}\\ &= \sum_{l=1}^{3}\widehat{\mathbf{1}}_1^\top\widehat{\mathcal{P}}_{\perp,\xi^l}\left[\mathbf{w}^\top\odot\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l^{(EC)}\right]\\ &= \sum_{l=1}^{3}\widehat{\mathbf{1}}_1^\top\widehat{\mathcal{P}}_{\perp,\xi^l}\left[\mathbf{w}^\top\odot\tilde{B}_{\xi^l}\hat{\tilde{\mathbf{f}}}_l^{(EC)}-\mathbf{w}^\top\odot\tilde{\Delta}_{\xi^l}\hat{\tilde{\mathbf{f}}}_l^{(EC)}\right]\\ &= \sum_{l=1}^{3}\widehat{\mathbf{1}}_1^\top\widehat{\mathcal{P}}_{\perp,\xi^l}\left[\widehat{B}_{\xi^l}\left(\hat{\boldsymbol{\psi}}_l+\hat{\mathbf{F}}_l\right)-\mathbf{w}^\top\odot\tilde{\Delta}_{\xi^l}\hat{\tilde{\mathbf{f}}}_l^{(EC)}\right],\end{aligned} \tag{105}$$

where we made use of Eq. (17) relating the entropy potential flux ($\hat{\boldsymbol{\psi}}_l$), entropy flux ($\hat{\mathbf{F}}_l$), and the inviscid flux. Hence, it only remains to show that $\sum_{l=1}^{3}\widehat{\mathbf{1}}_1^\top\widehat{\mathcal{P}}_{\perp,\xi^l}\left[\widehat{B}_{\xi^l}\hat{\boldsymbol{\psi}}_l-\mathbf{w}^\top\odot\tilde{\Delta}_{\xi^l}\hat{\tilde{\mathbf{f}}}_l^{(EC)}\right]=0$. Notice that

$$\begin{aligned}&\sum_{l=1}^{3}\widehat{\mathbf{1}}_1^\top\widehat{\mathcal{P}}_{\perp,\xi^l}\left[\widehat{B}_{\xi^l}\hat{\boldsymbol{\psi}}_l-\mathbf{w}^\top\odot\tilde{\Delta}_{\xi^l}\hat{\tilde{\mathbf{f}}}_l^{(EC)}\right]=\sum_{j,k=1}^{N}\widehat{\mathcal{P}}_{\perp,\xi^1}(\vec{\xi}_{ijk})\\ &\left[\hat{\boldsymbol{\psi}}_1(\vec{\xi}_N)-\hat{\boldsymbol{\psi}}_1(\vec{\xi}_1)-\sum_{i=1}^{N-1}\hat{\tilde{\mathbf{f}}}_1^{(EC)}(\vec{\xi}_i)\left(\mathbf{w}(\vec{\xi}_{i+1})-\mathbf{w}(\vec{\xi}_i)\right)^\top\right]_{\vec{\xi}_{\cdot jk}}+\cdots\\ &=\sum_{j,k=1}^{N}\widehat{\mathcal{P}}_{\perp,\xi^1}(\vec{\xi}_{ijk})\left[\hat{\boldsymbol{\psi}}_1(\vec{\xi}_N)-\hat{\boldsymbol{\psi}}_1(\vec{\xi}_1)-\sum_{i=1}^{N-1}\left(\vec{\boldsymbol{\psi}}(\vec{\xi}_{i+1})-\vec{\boldsymbol{\psi}}(\vec{\xi}_i)\right)\hat{\bar{\tilde{\mathbf{a}}}}^1(\vec{\xi}_{\bar{i}})\right]_{\vec{\xi}_{\cdot jk}}+\cdots\\ &=\sum_{j,k=1}^{N}\widehat{\mathcal{P}}_{\perp,\xi^1}(\vec{\xi}_{ijk})\left[\sum_{i=1}^{N}\vec{\boldsymbol{\psi}}(\vec{\xi}_i)\hat{\bar{\tilde{\mathbf{a}}}}^1(\vec{\xi}_{\bar{i}})-\sum_{i=1}^{N}\vec{\boldsymbol{\psi}}(\vec{\xi}_i)\hat{\bar{\tilde{\mathbf{a}}}}^1(\vec{\xi}_{\overline{i-1}})\right]_{\vec{\xi}_{\cdot jk}}+\cdots\\ &=\sum_{i,j,k=1}^{N}\widehat{\mathcal{P}}(\vec{\xi}_{ijk})\vec{\boldsymbol{\psi}}(\vec{\xi}_{ijk})\left[\frac{\hat{\bar{\tilde{\mathbf{a}}}}^1(\vec{\xi}_{\bar{i}jk})-\hat{\bar{\tilde{\mathbf{a}}}}^1(\vec{\xi}_{\overline{i-1}jk})}{\mathcal{P}_{i,i}}+\frac{\hat{\bar{\tilde{\mathbf{a}}}}^2(\vec{\xi}_{i\bar{j}k})-\hat{\bar{\tilde{\mathbf{a}}}}^2(\vec{\xi}_{i\overline{j-1}k})}{\mathcal{P}_{j,j}}+\cdots\right]\\ &=0,\end{aligned} \tag{106}$$

where $\vec{\boldsymbol{\psi}}(\vec{\xi}_i)=\begin{bmatrix}\boldsymbol{\psi}_{x^1}(\vec{\xi}_i) & \boldsymbol{\psi}_{x^2}(\vec{\xi}_i) & \boldsymbol{\psi}_{x^3}(\vec{\xi}_i)\end{bmatrix}^\top$, we made use of Eq. (68) for entropy conservatives fluxes, and again we use the discrete GCL Eq. (63). In light of Eq. (75), which shows that $\sum_{l=1}^{3}\mathbf{1}_1^\top\widehat{\mathcal{P}}_{\perp,\xi^l}\widehat{B}_{\xi^l}\hat{\mathbf{F}}_l=\sum_{l=1}^{3}\mathbf{w}^\top\mathcal{P}\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l$, it now follows that Eq. (102) holds: $\sum_{l=1}^{3}\mathbf{w}^\top\mathcal{P}\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l^{(EC)}=\sum_{l=1}^{3}\mathbf{1}_1^\top\widehat{\mathcal{P}}_{\perp,\xi^l}\widehat{B}_{\xi^l}\hat{\mathbf{F}}_l=\sum_{l=1}^{3}\mathbf{w}^\top\mathcal{P}\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l$. □

6.1.3 FIRST-ORDER ARTIFICIAL DISSIPATION

Here, we discuss the first-order Brenner dissipation added through $\hat{\hat{\mathbf{f}}}^{(AD_1)}_{\hat{\sigma},l}$, $\hat{\hat{\mathbf{f}}}^{(AD_1)}_{l}$ and $\hat{\mathbf{g}}^{(AD_1)}_{l}$ to the first-order scheme given by Eq. (100). To facilitate proving both density positivity and entropy stability, we begin by presenting a useful matrix based on the average of two states.

Lemma 3. *For two vectors of conservative variables $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ with positive density and temperature, consider the matrix*

$$\nu_w(\boldsymbol{u}_1, \boldsymbol{u}_2) = \begin{bmatrix} \frac{\rho_L}{R} & \frac{\rho_L}{R}(V_1)_A & \frac{\rho_L}{R}(V_2)_A & \frac{\rho_L}{R}(V_3)_A & \frac{\rho_L}{R}E_{avg} \\ 0 & T_H & 0 & 0 & T_H(V_1)_A \\ 0 & 0 & T_H & 0 & T_H(V_2)_A \\ 0 & 0 & 0 & T_H & T_H(V_3)_A \\ 0 & 0 & 0 & 0 & T_G^2 \end{bmatrix}, \tag{107}$$

where

$$E_{avg} = \frac{T_G^2}{T_L}\frac{R}{\gamma - 1} + \frac{\vec{\boldsymbol{V}}(\boldsymbol{u}_1)\cdot\vec{\boldsymbol{V}}(\boldsymbol{u}_2)}{2}, \tag{108}$$

and $\vec{\boldsymbol{V}}(\boldsymbol{u}_i)$ is the velocity associated with state $\boldsymbol{u}_i$. The matrix is 1) consistent with $\frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{w}}$, 2) invertible and 3) satisfies the exact algebraic relation $\nu_w(\boldsymbol{u}_1, \boldsymbol{u}_2)\,(\boldsymbol{w}_2 - \boldsymbol{w}_1) = (\boldsymbol{\nu}_2 - \boldsymbol{\nu}_1)$.

Proof. We verified these claims in Mathematica. □

Remark 2. We label the inverse of $\nu_w(\boldsymbol{u}_1, \boldsymbol{u}_2)$ as $w_\nu(\boldsymbol{u}_1, \boldsymbol{u}_2)$ and note that by Lemma 107 the following equality holds $w_\nu(\boldsymbol{u}_1, \boldsymbol{u}_2)\,(\boldsymbol{\nu}_2 - \boldsymbol{\nu}_1) = (\boldsymbol{w}_2 - \boldsymbol{w}_1)$.

The matrix $\nu_w(\boldsymbol{u}_1, \boldsymbol{u}_2)$ simplifies the process of finding 2-point approximations of other matrices as shown by the next lemma.

Lemma 4. *Let $\vec{\boldsymbol{n}}$ be a non-zero direction vector and $\vec{\boldsymbol{n}} = \frac{\vec{\boldsymbol{n}}}{\|\vec{\boldsymbol{n}}\|} = \begin{bmatrix} \bar{n}_1 & \bar{n}_2 & \bar{n}_3 \end{bmatrix}^\top$. For two*

admissible states $\boldsymbol{u}_1$ *and* $\boldsymbol{u}_2$ *with positive density and temperature, consider the matrix*

$$c_\nu^{(B)}(\boldsymbol{u}_1,\boldsymbol{u}_2,\vec{\boldsymbol{n}}) = \|\vec{\boldsymbol{n}}\|^2 \begin{bmatrix} \sigma & 0 & 0 & 0 & 0 \\ \sigma(V_1)_A & 0 & 0 & 0 & 0 \\ \sigma(V_2)_A & 0 & 0 & 0 & 0 \\ \sigma(V_3)_A & 0 & 0 & 0 & 0 \\ \sigma E_{avg} & 0 & 0 & 0 & \kappa \end{bmatrix} + \mu\|\vec{\boldsymbol{n}}\|^2 \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\bar{n}_1^2}{3}+1 & \frac{\bar{n}_2\bar{n}_1}{3} & \frac{\bar{n}_3\bar{n}_1}{3} & 0 \\ 0 & \frac{\bar{n}_2\bar{n}_1}{3} & \frac{\bar{n}_2^2}{3}+1 & \frac{\bar{n}_2\bar{n}_3}{3} & 0 \\ 0 & \frac{\bar{n}_3\bar{n}_1}{3} & \frac{\bar{n}_2\bar{n}_3}{3} & \frac{\bar{n}_3^2}{3}+1 & 0 \\ 0 & (V_1)_A + \frac{\bar{n}_1}{3}\vec{\boldsymbol{V}}_A\cdot\vec{\bar{\boldsymbol{n}}} & (V_2)_A + \frac{\bar{n}_2}{3}\vec{\boldsymbol{V}}_A\cdot\vec{\bar{\boldsymbol{n}}} & (V_3)_A + \frac{\bar{n}_3}{3}\vec{\boldsymbol{V}}_A\cdot\vec{\bar{\boldsymbol{n}}} & 0 \end{bmatrix}, \tag{109}$$

where E_{avg} *was defined in Eq.* (108)*,* $\vec{\boldsymbol{V}}_A$ *is the arithmetic average of the velocities, and* σ*,* μ*, and* κ *are the positive diffusion coefficients of the Brenner-Navier-Stokes flux (see Eq.* (43)*). We have that*

$$\begin{aligned} c^{(B)}(\boldsymbol{u}_1,\boldsymbol{u}_2,\vec{\boldsymbol{n}}) &= c_\nu^{(B)}(\boldsymbol{u}_1,\boldsymbol{u}_2,\vec{\boldsymbol{n}})\nu_w(\boldsymbol{u}_1,\boldsymbol{u}_2), \\ c^{(B)}(\boldsymbol{u}_1,\boldsymbol{u}_1,\vec{\boldsymbol{n}}) &= \sum_{l,m=1}^{3} n_l \mathsf{C}_{l,m}^{(B)}(\boldsymbol{u}_1) n_m, \\ c^{(B)}(\boldsymbol{u}_1,\boldsymbol{u}_2,\vec{\boldsymbol{n}}) &= (c^{(B)}(\boldsymbol{u}_1,\boldsymbol{u}_2,\vec{\boldsymbol{n}}))^\top, \\ \boldsymbol{V}^\top c^{(B)}(\boldsymbol{u}_1,\boldsymbol{u}_2,\vec{\boldsymbol{n}})\boldsymbol{V} &> 0, \quad \forall \boldsymbol{V}\in\mathbb{R}^5 - \{\boldsymbol{0}\}, \end{aligned} \tag{110}$$

where $\mathsf{C}_{l,m}^{(B)}$ *are the viscosity matrices from* (44)*.*

Proof. We checked all of this using Mathematica. To show positive definiteness of $c^{(B)}(\boldsymbol{u}_1,\boldsymbol{u}_2,\vec{\boldsymbol{n}})$ for positive diffusion coefficients, we used the Cholesky decomposition

$c^{(B)}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}}) = LDL^\top$ where

$$
\begin{aligned}
L &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ (V_1)_A & 1 & 0 & 0 & 0 \\ (V_2)_A & \frac{\bar{n}_1\bar{n}_2}{d_2} & 1 & 0 & 0 \\ (V_3)_A & \frac{\bar{n}_1\bar{n}_3}{d_2} & \frac{\bar{n}_2\bar{n}_3}{d_3} & 1 & 0 \\ E_{avg} & \frac{3(V_1)_A+\bar{n}_1\vec{\boldsymbol{V}}_A\cdot\vec{\boldsymbol{n}}}{d_2} & \frac{\left(4-\bar{n}_3^2\right)(V_2)_A+\bar{n}_2\bar{n}_3(V_3)_A}{d_3} & (V_3)_A & 1 \end{bmatrix}, \\
D &= \operatorname{diag}\left(\|\vec{\boldsymbol{n}}\|^2 \begin{bmatrix} \rho_L\frac{\sigma}{R} & T_H\mu\frac{d_2}{3} & T_H\mu\frac{d_3}{3+\bar{n}_1^2} & T_H\mu\frac{4}{4-\bar{n}_3^2} & T_G^2\kappa \end{bmatrix}\right),
\end{aligned}
\tag{111}
$$

$d_2 = \bar{n}_1^2+3$ and $d_3 = 4\bar{n}_1^4+4\bar{n}_2^4+7\bar{n}_2^2\bar{n}_3^2+3\bar{n}_3^4+\bar{n}_1^2(8\bar{n}_2^2+7\bar{n}_3^2)$. Since D has only positive entries (when the diffusion coefficients are all positive), $c^{(B)}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}})$ is positive definite. □

Remark 3. Let $\Delta\boldsymbol{w} = \boldsymbol{w}_2 - \boldsymbol{w}_1$ and $\Delta\boldsymbol{\nu} = \boldsymbol{\nu}_2 - \boldsymbol{\nu}_1$. Assume that the two states represent the cell averages of two adjacent volumes connected by a shared face. Let $\vec{\boldsymbol{n}}$ be a direction vector normal to the face connecting the two adjacent states. Locally on the shared face, assume change only in the normal direction so that $\nabla \boldsymbol{W} = \sum_{j=1}^{3} \frac{\partial \boldsymbol{W}}{\partial x_j}\bar{n}_j\vec{\boldsymbol{n}}$. Then it follows that near the interface

$$
\begin{aligned}
\sum_{m=1}^{3} \boldsymbol{F}_{x_m}^{(B)}\bar{n}_m &= \sum_{m,j=1}^{3} \bar{n}_m \mathsf{C}_{m,j}^{(B)}\bar{n}_j \left(\sum_{l=1}^{3} \frac{\partial \boldsymbol{W}}{\partial x_l}\bar{n}_l\right) \\
&\approx c^{(B)}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\bar{\boldsymbol{n}}}) \left(\sum_{l=1}^{3} \frac{\partial \boldsymbol{W}}{\partial x_l}\bar{n}_l\right) \\
&\approx c^{(B)}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\bar{\boldsymbol{n}}}) \left(\frac{\Delta\boldsymbol{w}}{\Delta\boldsymbol{x}}\right) \\
&= c_{\nu}^{(B)}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\bar{\boldsymbol{n}}}) \left(\frac{\Delta\boldsymbol{\nu}}{\Delta\boldsymbol{x}}\right),
\end{aligned}
\tag{112}
$$

where $\Delta\boldsymbol{x}$ is an appropriate length term. In this sense we say that $c_{\nu}^{(B)}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}})\frac{\Delta\boldsymbol{\nu}}{\Delta\boldsymbol{x}}$ is a consistent approximation of the normal Brenner flux at this interface. For the shear terms, the approximation of $\nabla \boldsymbol{W}$ is perhaps not ideal; however, we only use this approximation as a method of introducing artificial dissipation.

Remark 4. If we use the entropy consistent flux of Chandrashekar [35] we have

$$\bar{f}_{(S)}(\boldsymbol{u}_1,\boldsymbol{u}_2)\vec{\boldsymbol{n}} = \begin{bmatrix} \rho_L \\ \rho_L \vec{\boldsymbol{V}}_A \\ \rho_L E_{avg} \end{bmatrix} \vec{\boldsymbol{V}}_A \cdot \vec{\boldsymbol{n}} + P_{avg} \begin{bmatrix} 0 \\ \vec{\boldsymbol{n}} \\ \vec{\boldsymbol{V}}_A \cdot \vec{\boldsymbol{n}} \end{bmatrix}, \tag{113}$$

where $P_{avg} = R\rho_A T_H$. Hence if we replace the mass velocity $\vec{\boldsymbol{V}}_A \cdot \vec{\boldsymbol{n}}$ with the normal volume velocity $\vec{\boldsymbol{V}}_A \cdot \vec{\boldsymbol{n}} + \sigma \frac{\rho_2 - \rho_1}{\rho_L * \Delta \boldsymbol{x}}$ which is consistent with Brenner's modification of the Navier-Stokes equations at the discrete level (see Section 2.4), we obtain the mass diffusion from Eq. (109)

$$\begin{aligned} \bar{f}^{\sigma}_{(S)}(\boldsymbol{u}_1,\boldsymbol{u}_2)\vec{\boldsymbol{n}} &= \begin{bmatrix} \rho_L \\ \rho_L \vec{\boldsymbol{V}}_A \\ \rho_L E_{avg} \end{bmatrix} \vec{\boldsymbol{V}}_A \cdot \vec{\boldsymbol{n}} + \begin{bmatrix} 1 \\ \vec{\boldsymbol{V}}_A \\ E_{avg} \end{bmatrix} \sigma \frac{\rho_2 - \rho_1}{\Delta \boldsymbol{x}} + P_{avg} \begin{bmatrix} 0 \\ \vec{\boldsymbol{n}} \\ \vec{\boldsymbol{V}}_A \cdot \vec{\boldsymbol{n}} \end{bmatrix} \\ &= \bar{f}_{(S)}(\boldsymbol{u}_1,\boldsymbol{u}_2)\vec{\boldsymbol{n}} + c_{\nu}^{(B)}(\boldsymbol{u}_1,\boldsymbol{u}_2,\vec{\boldsymbol{n}}) \begin{bmatrix} 1 \\ 0 \\ \vdots \end{bmatrix} \frac{\rho_2 - \rho_1}{\Delta \boldsymbol{x}}. \end{aligned} \tag{114}$$

The $\hat{\vec{\mathbf{f}}}^{(AD_1)}_{\hat{\sigma},l}$, $\hat{\vec{\mathbf{f}}}^{(AD_1)}_{l}$ and $\hat{\mathbf{g}}^{(AD_1)}_{l}$ contributions are formed as follows for all fixed $1 \le j,k \le N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$\begin{aligned} &\text{for } 1 \le i \le N-1, \\ &\hat{\vec{\mathbf{f}}}^{(AD_1)}_{1}(\vec{\xi_i}) = \frac{c_{\nu}^{(B)}(\mathbf{U}(\vec{\xi}_i),\mathbf{U}(\vec{\xi}_{i+1}),\hat{\vec{\mathbf{a}}}^1(\vec{\xi_i}))}{\sqrt{\boldsymbol{J}(\vec{\xi}_i)\boldsymbol{J}(\vec{\xi}_{i+1})}} \frac{\boldsymbol{\nu}(\vec{\xi}_{i+1}) - \boldsymbol{\nu}(\vec{\xi}_i)}{\xi_{i+1} - \xi_i}, \\ &\hat{\vec{\mathbf{f}}}^{(AD_1)}_{\hat{\sigma},1}(\vec{\xi_i}) = \frac{c_{\nu}^{(B)}(\mathbf{U}(\vec{\xi}_i),\mathbf{U}(\vec{\xi}_{i+1}),\hat{\vec{\mathbf{a}}}^1(\vec{\xi_i}),\hat{\vec{\boldsymbol{\sigma}}}_1(\vec{\xi_i}),\mu=\kappa=0)}{\sqrt{\boldsymbol{J}(\vec{\xi}_i)\boldsymbol{J}(\vec{\xi}_{i+1})}} \frac{\boldsymbol{\nu}(\vec{\xi}_{i+1}) - \boldsymbol{\nu}(\vec{\xi}_i)}{\xi_{i+1} - \xi_i}, \\ &\hat{\vec{\mathbf{f}}}^{(AD_1)}_{1}(\vec{\xi_0}) = \hat{\vec{\mathbf{f}}}^{(AD_1)}_{1}(\vec{\xi_N}) = \hat{\vec{\mathbf{f}}}^{(AD_1)}_{\hat{\sigma},1}(\vec{\xi_0}) = \hat{\vec{\mathbf{f}}}^{(AD_1)}_{\hat{\sigma},1}(\vec{\xi_N}) = \mathbf{0}, \\ &\hat{\mathbf{g}}^{(AD_1)}_{1}(\vec{\xi_i}) = \left(\hat{\mathbf{g}}^{(AD_1)}_{1}(\vec{\xi_1})\delta_{1i} + \hat{\mathbf{g}}^{(AD_1)}_{1}(\vec{\xi_N})\delta_{Ni}\right), \\ &\hat{\mathbf{g}}^{(AD_1)}_{1}(\vec{\xi_1}) = \frac{c_{\nu}^{(B)}(\mathbf{U}(\vec{\xi}_0),\mathbf{U}(\vec{\xi}_1),\hat{\vec{\mathbf{a}}}^1(\vec{\xi_0}))}{\sqrt{\boldsymbol{J}(\vec{\xi}_0)\boldsymbol{J}(\vec{\xi}_1)}} \frac{\boldsymbol{\nu}(\vec{\xi}_0) - \boldsymbol{\nu}(\vec{\xi}_1)}{\mathcal{P}_{1,1}}, \end{aligned} \tag{115}$$

with identical definitions in other computational directions. As discussed in Section 2.4, for

$\hat{\hat{\mathbf{f}}}_l^{(AD_1)}$ one replaces μ, σ, and κ with artificial viscosity coefficients that depend on $\boldsymbol{\mu}^{AD}$; however, $\boldsymbol{\mu}^{AD}$ is stored at the solution points and $\hat{\hat{\mathbf{f}}}_l^{(AD_1)}$ is formed at the flux points. In Section 6.3.5, we give explicit details of how the viscosity coefficients for $\hat{\hat{\mathbf{f}}}_l^{(AD_1)}$ are formed. The term $\hat{\hat{\mathbf{f}}}_{\hat{\hat{\sigma}},l}^{(AD_1)}$ is added solely for the purpose of ensuring that the total Brenner first-order mass diffusion at the flux point is sufficient for density positivity when needed and hence uses only mass diffusion as specified by the variable $\hat{\hat{\boldsymbol{\sigma}}}_1$. See Sections 6.3.5 and 6.4 for full details on how the artificial viscous coefficients are handled.

Lemma 5. *The terms* $\hat{\mathbf{g}}_l^{(AD_1)}$, $\hat{\hat{\mathbf{f}}}_{\hat{\hat{\sigma}},l}^{(AD_1)}$ *and* $\hat{\hat{\mathbf{f}}}_l^{(AD_1)}$ *given by* (115) *are entropy stable.*

Proof. This follows directly from Lemmas 4 and 18. □

6.1.4 FIRST-ORDER MERRIAM–ROE FLUX

Recall that the inviscid term $\hat{\hat{\mathbf{f}}}_l^{(MR)}$ of Eq. (100) is written as the sum $\hat{\hat{\mathbf{f}}}_l^{(MR)} = \hat{\hat{\mathbf{f}}}_l^{(EC)} - \hat{\hat{\mathbf{f}}}_l^{(ED)}$. Here, we discuss the entropy dissipative term $\hat{\hat{\mathbf{f}}}_l^{(ED)}$. Often, two-point entropy conservative fluxes are stabilized through the use of Rusanov-type fluxes (e.g., see [47, 64]). Note, however, that the Rusanov-type fluxes dissipate each characteristic wave regardless of the magnitude of the corresponding eigenvalue associated with this wave, thus making them too dissipative. A less dissipative and more refined approach is to use an entropy dissipative characteristic flux proposed by Merriam in [26], which is herein referred to as the Merriam–Roe (MR) flux and given by

$$\boldsymbol{f}^{(MR)}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}}) = \bar{f}_{(S)}(\boldsymbol{u}_1, \boldsymbol{u}_2)\vec{\boldsymbol{n}} - M^{\mathcal{Y}}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}})\Delta\boldsymbol{w}. \tag{116}$$

For $\bar{f}_{(S)}(\cdot,\cdot)$ any two-point, consistent, entropy consistent inviscid interface flux can be used, $\Delta\boldsymbol{w} = \boldsymbol{w}_2 - \boldsymbol{w}_1$, $M^{\mathcal{Y}}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}})$ is a two-point consistent average of the matrix $\frac{1}{2}\mathcal{Y}|\lambda|\mathcal{Y}^T$. The matrix $\mathcal{Y}$ is a matrix composed out of normalized eigenvectors of the flux Jacobian $f'(\mathbf{W}, \vec{\boldsymbol{n}}) = f'(\mathbf{U}, \vec{\boldsymbol{n}})\frac{\partial \mathbf{U}}{\partial \mathbf{W}}$ which can be decomposed as follows:

$$\begin{aligned} &f'(\mathbf{W}, \vec{\boldsymbol{n}}) = \mathcal{Y}\lambda\mathcal{Y}^T, \frac{\partial \boldsymbol{U}}{\partial \boldsymbol{W}} = \mathcal{Y}\mathcal{Y}^T, \\ &\lambda = \text{diag}\left(\; -c\|\vec{\boldsymbol{n}}\| + \vec{\boldsymbol{V}}\cdot\vec{\boldsymbol{n}} \quad c\|\vec{\boldsymbol{n}}\| + \vec{\boldsymbol{V}}\cdot\vec{\boldsymbol{n}} \quad \vec{\boldsymbol{V}}\cdot\vec{\boldsymbol{n}} \quad \vec{\boldsymbol{V}}\cdot\vec{\boldsymbol{n}} \quad \vec{\boldsymbol{V}}\cdot\vec{\boldsymbol{n}} \; \right), \end{aligned} \tag{117}$$

where the exact form of $\mathcal{Y}$ can be found in [52]. The matrix $M^{\mathcal{Y}}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}})$ is SPSD if the density and temperature values used to build the matrix are positive. For two admissible states $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, there are many options for building $M^{\mathcal{Y}}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}})$ at an interface. In the present analysis, we use the following average:

$$\boldsymbol{\nu}(\boldsymbol{u}_1, \boldsymbol{u}_2) = \begin{bmatrix} \rho_L \\ \frac{\vec{\boldsymbol{V}}(\boldsymbol{u}_1)T_2 + \vec{\boldsymbol{V}}(\boldsymbol{u}_2)T_1}{T1+T2} \\ T_H \end{bmatrix}, \tag{118}$$

where $\vec{\boldsymbol{V}}(\boldsymbol{u}_1)$ is the velocity vector of $\boldsymbol{u}_1$. With this average, we can write the first component of $M^{\mathcal{Y}}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}})\Delta\boldsymbol{w}$ in a form that facilitates proving density positivity:

$$\begin{aligned} \left(M^{\mathcal{Y}}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}})\Delta\boldsymbol{w}\right)_\rho &= \rho_L \mathcal{V}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}}) + \Delta\rho\lambda_c, \\ \mathcal{V}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}}) &= -\left(\frac{\Delta(\log T)}{\gamma - 1} + \frac{\Delta T \left\| \Delta \vec{\boldsymbol{V}} \right\|^2}{8R_g T_A^2}\right)\lambda_c + \frac{\Delta T}{4T_A(\gamma-1)}(\lambda_2 + \lambda_3) \\ &\quad + (\lambda_3 - \lambda_2)\frac{\Delta \vec{\boldsymbol{V}} \cdot \frac{\vec{\boldsymbol{n}}}{\|\vec{\boldsymbol{n}}\|}\sqrt{T_H}}{2T_A\sqrt{R_g\gamma}}, \end{aligned} \tag{119}$$

where $\lambda_1 = |\,\vec{\boldsymbol{V}}_{avg} \cdot \vec{\boldsymbol{n}}|$, $\lambda_2 = |\,\vec{\boldsymbol{V}}_{avg} \cdot \vec{\boldsymbol{n}} - c_{avg}\,\|\vec{\boldsymbol{n}}\|\,|$, $\lambda_3 = |\,\vec{\boldsymbol{V}}_{avg} \cdot \vec{\boldsymbol{n}} + c_{avg}\,\|\vec{\boldsymbol{n}}\|\,|$, $\lambda_c = \frac{\lambda_1(\gamma-1)}{2\gamma} + \frac{\lambda_2+\lambda_3}{4\gamma}$, $\vec{\boldsymbol{V}}_{avg} = \frac{\vec{\boldsymbol{V}}_1 T_2 + \vec{\boldsymbol{V}}_2 T_1}{T1+T2}$, $c_{avg} = \sqrt{R_g T_H \gamma}$, and $\Delta T = T_2 - T_1$. This method is significantly less dissiptive than the Rusanov–type dissipation, but a positivity proof based on $\boldsymbol{f}^{(MR)}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}})$ alone is not currently available–the issue being that one cannot increase the mass diffusion term in Eq. (119) without changing the term $\mathcal{V}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}})$. Of course, for forward Euler time integration, one can always simply solve the inequality for density positivity to obtain a time step constraint that guarantees positivity, but then density remains in the time step constraint and it is difficult to compare the time step constraint with the standard CFL condition. Our approach is to use $\boldsymbol{f}^{(MR)}(\boldsymbol{u}_1, \boldsymbol{u}_2, \vec{\boldsymbol{n}})$ and supplement it with the Brenner mass diffusion to minimize the amount of dissipation introduced into the numerical solution and ensure density positivity with a time step constraint comparable to the standard CFL condition.

The $\hat{\tilde{\mathbf{f}}}_l^{(ED)}$ contribution to $\hat{\tilde{\mathbf{f}}}_l^{(MR)}$ of Eq. (100) is formed as follows for all fixed $1 \leq j, k \leq N$

and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$\begin{aligned}
&\text{for } 1 \le i \le N-1, \\
&\hat{\tilde{\mathbf{f}}}_1^{(ED)}(\vec{\xi}_{\bar{i}}) = M^{\mathcal{Y}}(\mathbf{U}(\vec{\xi}_i), \mathbf{U}(\vec{\xi}_{i+1}), \hat{\tilde{\vec{\mathbf{a}}}}^1(\vec{\xi}_{\bar{i}}))\left(\mathbf{w}(\vec{\xi}_{i+1}) - \mathbf{w}(\vec{\xi}_i)\right), \\
&\hat{\tilde{\mathbf{f}}}_1^{(ED)}(\vec{\xi}_{\overline{0}}) = \hat{\tilde{\mathbf{f}}}_1^{(ED)}(\vec{\xi}_{\overline{N}}) = \mathbf{0},
\end{aligned} \tag{120}$$

with identical definitions in other computational directions.

6.1.5 POSITIVITY OF DENSITY

We now discuss how positivity of density can be maintained for the first-order discretization given by Eq. (100).

Theorem 6. *Assume that explicit Euler in time is used for the discretization given by Eq.* (100) *where we only assume that* $\hat{\tilde{\mathbf{f}}}_l^{(MR)}$ *is some consistent inviscid interface flux. Let* $J_{ijk} = \mathbf{J}(\vec{\xi}_{ijk})$, $\rho_{ijk} = \boldsymbol{\rho}(\vec{\xi}_{ijk})$, *and* $\mathbf{U}_{ijk} = \mathbf{U}(\vec{\xi}_{ijk})$. *Consider the update of the density* $J_{ijk}\rho_{ijk} = \hat{\rho}$ *at the solution point* $\vec{\xi}_{ijk}$. *The density update depends only on the nearest neighbors and we write the interface fluxes as* $\hat{\tilde{\boldsymbol{f}}}_1^{\rho}(\mathbf{U}_{ijk}, \mathbf{U}_{i+1jk}) = \hat{\tilde{\boldsymbol{f}}}_1^{\rho,+}$, $\hat{\tilde{\boldsymbol{f}}}_1^{\rho}(\mathbf{U}_{ijk}, \mathbf{U}_{i-1jk}) = \hat{\tilde{\boldsymbol{f}}}_1^{\rho,-}$ *and similarly in other directions. Hence, we have*

$$\hat{\rho}^{n+1} = \hat{\rho}^n - \tau\left[\frac{\hat{\tilde{\boldsymbol{f}}}_1^{\rho,+} - \hat{\tilde{\boldsymbol{f}}}_1^{\rho,-}}{P_{ii}} + \frac{\hat{\tilde{\boldsymbol{f}}}_2^{\rho,+} - \hat{\tilde{\boldsymbol{f}}}_2^{\rho,-}}{P_{jj}} + \frac{\hat{\tilde{\boldsymbol{f}}}_3^{\rho,+} - \hat{\tilde{\boldsymbol{f}}}_3^{\rho,-}}{P_{kk}}\right]. \tag{121}$$

The numerical density fluxes can be written generally as the sum of a diffusive and non-diffusive term: $\hat{\tilde{\boldsymbol{f}}}_l^{\rho,+/-} = \hat{\tilde{m}}_l^{+/-} - K_l^{+/-}\Delta_l^{+/-}\rho$ *where* $\Delta_1^+\rho = \rho_{i+1jk} - \rho_{ijk}$, $\Delta_1^-\rho = \rho_{ijk} - \rho_{i-1jk}$ *etc. Let* $\rho_{1,A}^+ = \frac{\rho_{ijk}+\rho_{i+1jk}}{2}$ *and* $\rho_{1,A}^- = \frac{\rho_{ijk}+\rho_{i-1jk}}{2}$ *with identical definitions in other directions. If* $K_l^{+/-} \ge K_{l,\min}^{+/-} = \frac{|\hat{\tilde{m}}_l^{+/-}|}{2\rho_{l,A}^{+/-}}$, *then the above first–order FV scheme preserves the positivity of the density* ρ *under the following CFL condition:*

$$\tau < \frac{J_{ijk}}{2\sum_{l=1}^{3}\frac{K_l^+ + K_l^-}{P_{ll}}} = \tau_\rho^s. \tag{122}$$

Proof. We split the discrete density equation as follows:

$$\hat{\rho}^{n+1} = \left(\frac{\hat{\rho}^n}{6} - \tau\frac{\hat{\tilde{\boldsymbol{f}}}_1^{\rho,+}}{P_{ii}}\right) + \left(\frac{\hat{\rho}^n}{6} + \tau\frac{\hat{\tilde{\boldsymbol{f}}}_1^{\rho,-}}{P_{ii}}\right) + \dots \tag{123}$$

Since all the interfaces are handled identically, we look at the first term:

$$\left(\frac{\hat{\rho}^n}{6} - \tau\frac{\hat{\tilde{\boldsymbol{f}}}_1^{\rho,+}}{P_{ii}}\right) = \frac{\hat{\rho}^n}{6} - \frac{\tau}{P_{ii}}(\hat{\tilde{m}}_1^+ - K_1^+\Delta_1^+\rho).$$

First, we assume that despite $K_1^+ \geq K_{1,\min}^+ = \frac{|\hat{\tilde{m}}_1^+|}{2\rho_{1,A}^+}$, $K_1^+ = 0$. Then since $K_1^+ \geq \frac{|\hat{\tilde{m}}_1^+|}{2\rho_{1,A}^+}$ we have $\left(\frac{\hat{\rho}^n}{6} - \tau\frac{\hat{\tilde{\boldsymbol{f}}}_1^{\rho,+}}{P_{ii}}\right) = \frac{\hat{\rho}^n}{6} = \rho^n\left[\frac{J_{ijk}}{6} - \frac{2\tau K_1^+}{P_{ii}}\right]$. Now, assume that $K_1^+ > 0$. Then we have

$$\begin{aligned}
\frac{\hat{\rho}^n}{6} - \frac{\tau}{P_{ii}}(\hat{\tilde{m}}_1^+ - K_1^+\Delta_1^+\rho) &= \frac{\hat{\rho}^n}{6} - \frac{\tau K_1^+}{P_{ii}}\left(\frac{\hat{\tilde{m}}_1^+}{K_1^+} - \Delta_1^+\rho\right) \\
&\geq \frac{\hat{\rho}^n}{6} - \frac{\tau K_1^+}{P_{ii}}\left(2\rho_{1,A}^+ - \Delta_1^+\rho\right) \\
&= \rho^n\left[\frac{J_{ijk}}{6} - \frac{2\tau K_1^+}{P_{ii}}\right].
\end{aligned} \tag{124}$$

Therefore, summing over all interfaces, we have

$$\hat{\rho}^{n+1} \geq \rho^n\left[J_{ijk} - 2\tau\sum_{l=1}^{3}\frac{K_l^+ + K_l^-}{P_{ll}}\right] > 0. \tag{125}$$

□

Remark 5. While (122) is sufficient for density positivity, we have found that in practice the following stricter time step constraint is better suited for ensuring stability and positivity for multi-stage Runge-Kutta time integration

$$\tau < \frac{J_{ijk}}{12}\min_l\left(\frac{P_{ll}}{\max\left(K_l^+, K_l^-\right)}\right) = \tau_\rho^I, \tag{126}$$

where the superscript I indicates that this time step condition preserves positivity of (124) at each interface, but τ_ρ^s only preserves positivity of the solution point. We use (126) in the remainder of this dissertation.

Remark 6. Since we require $K_l^{+/-} \geq \frac{|\hat{\bar{m}}_l^{+/-}|}{2\rho_{l,A}^{+/-}}$, density is present in the time step constraints (122) and (126). However, this does not impose a stricter constraint on the time step when the density jump increases, because the ratio $\frac{2\rho_{l,A}^{+/-}}{|\hat{\bar{m}}_l^{+/-}|}$ can be bounded from below by a positive nonzero value that is independent of density for all inviscid interface fluxes we have checked (see below).

Remark 7. Note that we didn't need to assume that $\hat{\bar{\mathbf{f}}}_l^{(MR)}$ was first-order or had a specific stencil. However, if we use a high-order flux for $\hat{\bar{\mathbf{f}}}_l^{(MR)}$, then the time step will contain a ratio of density that cannot be removed and in discontinuous regions $|\hat{\bar{m}}_l^{+/-}|$ can be much larger than one would obtain with a first-order stencil.

We now give three examples demonstrating how Theorem 6 and (126) can be used to preserve density positivity of the scheme given by Eq. (100) when only minimum mass diffusion is used i.e. $K_l^{+/-} = K_{l,\min}^{+/-} = \frac{|\hat{\bar{m}}_l^{+/-}|}{2\rho_{l,A}^{+/-}}$.

Positivity of density: Chandrashekar EC flux

Assume that for $\bar{f}_{(S)}(\cdot,\cdot)$ in (101) we use the flux of Chandrashekar [35] given by Eq. (113). Assume further that $\hat{\bar{\mathbf{f}}}_l^{(MR)} = \hat{\bar{\mathbf{f}}}_l^{(EC)}$ i.e. $\hat{\bar{\mathbf{f}}}_l^{(ED)} = 0$ so that $\hat{\bar{m}}_l^{+/-} = \rho_L \vec{\boldsymbol{V}}_A \cdot \vec{\boldsymbol{n}}$ where for readability we suppress the $+/-$ and l flux point identifiers. Then, we have that

$$K_{l,\min}^{+/-} = \frac{|\rho_L \vec{\boldsymbol{V}}_A \cdot \vec{\boldsymbol{n}}|}{2\rho_A} \leq \frac{|\vec{\boldsymbol{V}}_A \cdot \vec{\boldsymbol{n}}|}{2}, \tag{127}$$

so that the time step constraints of Theorem 6 and (126) are comparable with the regular CFL condition for supersonic flows and impose even weaker restrictions on the time step for transonic and subsonic flows.

Positivity of density: Ismail and Roe EC flux

Assume that for $\bar{f}_{(S)}(\cdot,\cdot)$ in (101) we use the flux of Ismail and Roe [43]. Assume further that $\hat{\bar{\mathbf{f}}}_l^{(MR)} = \hat{\bar{\mathbf{f}}}_l^{(EC)}$ i.e. $\hat{\bar{\mathbf{f}}}_l^{(ED)} = 0$ so that $\hat{\bar{m}}_l^{+/-} = \gamma(\rho c)_L \left(\frac{\vec{\boldsymbol{V}}}{c}\right)_A \cdot \vec{\boldsymbol{n}}$. where c is the speed of sound and for readability we suppress the $+/-$ and l flux point identifiers. Then, we have

that

$$K_{l,\min}^{+/-} = \frac{|\gamma(\rho c)_L \left(\frac{\vec{\boldsymbol{V}}}{c}\right)_A \cdot \vec{\boldsymbol{n}}|}{2\rho_A} \leq \frac{\gamma(\rho c)_A}{2\rho_A} \left| \left(\frac{\vec{\boldsymbol{V}}}{c}\right)_A \cdot \vec{\boldsymbol{n}} \right| \tag{128}$$

$$\leq \frac{\gamma \rho_{\max} c_A}{2\rho_A} \left| \left(\frac{\vec{\boldsymbol{V}}}{c}\right)_A \cdot \vec{\boldsymbol{n}} \right| \tag{129}$$

$$\leq \gamma c_A \left| \left(\frac{\vec{\boldsymbol{V}}}{c}\right)_A \cdot \vec{\boldsymbol{n}} \right|, \tag{130}$$

so that again we see the time step constraints of Theorem 6 and (126) are comparable with the regular CFL condition and are not adversely influenced by the presence of density.

Positivity of density: Merriam–Roe flux

Here, we use Theorem 6 to prove density positivity for the full first-order scheme given by Eq. (100).

Corollary 6.1. *Assume that the EC flux of Chandrashekar [35] is used for* (100)*. Assume the notation of Theorem 6 and let* $\hat{\vec{\mathbf{a}}}^l_{+/-}$ *represent the metric term at the* $+/-$ *interface in the* l *direction. The semi-discrete scheme given by* (100) *preserves density positivity of the solution point* $\boldsymbol{u}$ *when the explicit Euler in time method is used if at each interface of* $\boldsymbol{u}$ *we have*

$$\begin{aligned} \left[\lambda_c + \frac{\sigma \|\hat{\vec{\mathbf{a}}}\|^2}{J_G \Delta\xi}\right]_l^{+/-} = K_l^{+/-} &\geq K_{l,\min}^{+/-} \\ &= \frac{\rho_{l,L}^{+/-}}{2\rho_{l,A}^{+/-}} \left| \vec{\boldsymbol{V}}_A \cdot \hat{\vec{\mathbf{a}}}^l_{+/-} - \mathcal{V}(\boldsymbol{u}, \boldsymbol{u}_l^{+/-}, \hat{\vec{\mathbf{a}}}^l_{+/-}) \right|, \end{aligned} \tag{131}$$

and the corresponding time step restriction given by (122) *is satisfied. In particular, the constraint on* $\sigma_l^{+/-}$ *is*

$$\sigma_l^{+/-} \geq \sigma_{l,\min}^{+/-} = \left[\max\left(0, \frac{\rho_L}{2\rho_A} \left| \vec{\boldsymbol{V}}_A \cdot \hat{\vec{\mathbf{a}}} - \mathcal{V}(\boldsymbol{u}, \boldsymbol{u}_l^{+/-}, \hat{\vec{\mathbf{a}}}) \right| - \lambda_c \right) \frac{J_G \Delta\xi}{\|\hat{\vec{\mathbf{a}}}\|^2} \right]_l^{+/-}, \tag{132}$$

where all terms are understood to come from the $+/-$*,* l *interface in question based on the*

definitions already given in this section.

Proof. This follows directly from Theorem (6). □

Remark 8. Note that the velocity and temperature averages in Eq. (118) were specifically chosen to make $\mathcal{V}(\boldsymbol{u}, \boldsymbol{u}_l^{+/-}, \hat{\vec{\mathbf{a}}}_{+/-}^l)$ density independent and to grow slowly in the case of large temperature jumps so that (131) would not impose an unnecessarily strict time step constraint. In particular, note that for the eigenvalues since $T_H < 2T_{\min}$ we have $c_{avg} < \sqrt{2}c_{\min}$. For the coefficients of the eigenvalues, only the logarithmic jump in temperature doesn't permit a formal upper bound–but for practical purposes we consider it bounded. Hence, for moderate Mach numbers $\mathcal{V}(\boldsymbol{u}, \boldsymbol{u}_l^{+/-}, \hat{\vec{\mathbf{a}}}_{+/-}^l)$ is bounded. As the Mach number increases, $\mathcal{V}(\boldsymbol{u}, \boldsymbol{u}_l^{+/-}, \hat{\vec{\mathbf{a}}}_{+/-}^l)$ grows without bound for the case of large velocity jumps at high Mach numbers. However, one would already expect to need a strict time step constraint for such flows; furthermore, the direct diffusion between neighbors in the scheme given by Eq. (100) always acts to alleviate sharp two point jumps and, when necessary, the discretely entropy stable velocity and temperature limiters of Section 6.2 can be used.

6.1.6 POSITIVITY OF INTERNAL ENERGY

When the explicit first-order Euler scheme is used to advance the solution in time, i.e.

$$\hat{\mathbf{U}}^{n+1} = \hat{\mathbf{U}}^n + \tau \hat{\mathbf{U}}_t, \tag{133}$$

so that τ is on the interval that preserves the positivity of $\boldsymbol{\rho}^{n+1}(\vec{\xi}_{ijk})$, the sign of the internal energy at the next time level for the solution point $\vec{\xi}_{ijk}$ is determined by the following polynomial:

$$\mathrm{IE}(\boldsymbol{u}^{n+1})\rho^{n+1} = \left(\frac{\tau}{J}\right)^2 \left(\frac{dE}{dt}\frac{d\rho}{dt} - \frac{1}{2}\left\|\frac{d\boldsymbol{m}}{dt}\right\|^2\right) + \frac{\tau}{J}(\boldsymbol{u}^n)^\top \begin{bmatrix} \frac{dE}{dt} \\ -\frac{d\boldsymbol{m}}{dt} \\ \frac{d\rho}{dt} \end{bmatrix} + \mathrm{IE}(\boldsymbol{u}^n)\rho^n, \tag{134}$$

where $\hat{\mathbf{U}}^n(\vec{\xi}_{ijk}) = J\boldsymbol{u}^n$, $\mathbf{J}(\vec{\xi}_{ijk}) = J$, $\boldsymbol{\rho}^{n+1}(\vec{\xi}_{ijk}) = \rho^{n+1}$, $\hat{\mathbf{U}}_t(\vec{\xi}_{ijk}) = \begin{bmatrix} \frac{d\rho}{dt} & \frac{d\boldsymbol{m}}{dt} & \frac{dE}{dt} \end{bmatrix}^\top$ and $\mathrm{IE}(\boldsymbol{u}^n)$ is the internal energy of $\boldsymbol{u}^n$. The above quadratic trinomial can be recast in the

following form that illustrates the role of $(\boldsymbol{w}^n)^T \hat{\mathbf{U}}_t$:

$$\begin{aligned} \mathrm{IE}(\boldsymbol{u}^{n+1})\rho^{n+1} = \left(\frac{\tau}{J}\right)^2 \left(\frac{dE}{dt}\frac{d\rho}{dt} - \frac{1}{2}\left\|\frac{d\boldsymbol{m}}{dt}\right\|^2\right) - \\ \frac{\tau}{J}\mathrm{IE}(\boldsymbol{u}^n)\left[\frac{\gamma-1}{R_g}(\boldsymbol{w}^n)^T\hat{\mathbf{U}}_t + \frac{d\rho}{dt}\left(s^n\frac{\gamma-1}{R_g} - (\gamma+1)\right)\right] + \mathrm{IE}(\boldsymbol{u}^n)\rho^n. \end{aligned} \tag{135}$$

The following lemma follows from these identities.

Lemma 7. *Assume that $\boldsymbol{\rho}^n(\vec{\xi}_{ijk}), IE(\mathbf{U}^n(\vec{\xi}_{ijk})) > 0$ for all solution points in the domain. Assume that the explicit Euler scheme in time given by Eq. (133) guarantees $\boldsymbol{\rho}^{n+1}(\vec{\xi}_{ijk}) > 0$ for all solution points for all $0 \leq \tau < \tau^\rho$. Then, there exists $0 < \tau^{\min} \leq \tau^\rho$ such that for all time steps $0 < \tau < \tau^{\min}$ the scheme preserves the positivity of internal energy, i.e., $IE(\mathbf{U}^{n+1}(\vec{\xi}_{ijk})) > 0$ for every solution point.*

Proof. Since for all solution points $\mathrm{IE}(\mathbf{U}^n(\vec{\xi}_{ijk}))\boldsymbol{\rho}^n(\vec{\xi}_{ijk}) > 0$, the above quadratic trinomial in τ is either strictly positive, i.e., $\mathrm{IE}(\mathbf{U}^{n+1}(\vec{\xi}_{ijk}))\boldsymbol{\rho}^{n+1}(\vec{\xi}_{ijk}) > 0 \ \forall \tau > 0$ (thus imposing no time step constraint for positivity of temperature), or a minimum positive root $\boldsymbol{\tau}^{\min}(\vec{\xi}_{ijk})$ of the quadratic equation $\mathrm{IE}(\mathbf{U}(\vec{\xi}_{ijk})^{n+1})\boldsymbol{\rho}(\vec{\xi}_{ijk})^{n+1} = 0$ exists for which positivity of $\mathrm{IE}(\mathbf{U}(\vec{\xi}_{ijk})^{n+1})\boldsymbol{\rho}(\vec{\xi}_{ijk})^{n+1}$ is guaranteed for all $\tau < \boldsymbol{\tau}^{\min}(\vec{\xi}_{ijk})$. Hence, a sufficient condition for internal energy positivity at the next time level for a scheme given by Eq. (133) is that $\tau < \tau^{\min} = \min(\tau^\rho, \min_{ijk}(\boldsymbol{\tau}^{\min}(\vec{\xi}_{ijk})))$ (note that if τ^ρ is sharp, then $\tau < \tau^{\min}$ is also a necessary condition). □

To bound the internal energy at each solution point $\mathrm{IE}(\mathbf{U}^{n+1}(\vec{\xi}_{ijk}))$ from below, we can choose $\tau \leq \tau^{\min} = \min_{ijk}(\boldsymbol{\tau}^{\min}(\vec{\xi}_{ijk}))$ where $\boldsymbol{\tau}^{\min}(\vec{\xi}_{ijk})$ are redefined as follows. Let c_{IE} be a user-defined parameter $0 < c_{\mathrm{IE}} < 1$. Then, $\boldsymbol{\tau}^{\min}(\vec{\xi}_{ijk})$ is defined such that $\mathrm{IE}(\mathbf{U}^{n+1}(\vec{\xi}_{ijk})) \geq c_{\mathrm{IE}}\mathrm{IE}(\mathbf{U}^n(\vec{\xi}_{ijk}))$. Hence, the upper bound of $\boldsymbol{\tau}^{\min}(\vec{\xi}_{ijk})$ is the minimum positive root of the following quadratic equation:

$$0 = \left(\frac{\tau}{J}\right)^2 \left(\frac{dE}{dt}\frac{d\rho}{dt} - \frac{1}{2}\left\|\frac{d\boldsymbol{m}}{dt}\right\|^2\right) + \frac{\tau}{J}\left(\widetilde{\boldsymbol{u}}^n\right)^\top \begin{bmatrix} \frac{dE}{dt} \\ -\frac{d\boldsymbol{m}}{dt} \\ \frac{d\rho}{dt} \end{bmatrix} + \mathrm{IE}(\widetilde{\boldsymbol{u}}^n)\rho^n, \tag{136}$$

where $\widetilde{\boldsymbol{u}}_i^n$ is $\boldsymbol{u}_i^n$ with the temperature scaled by $1 - c_{\mathrm{IE}}$. If no positive roots exist for this equation, then for all $\tau > 0$, $\mathrm{IE}(\boldsymbol{u}_i^{n+1}) > c_{\mathrm{IE}}\mathrm{IE}(\boldsymbol{u}_i^n)$; otherwise, there exists the minimum positive root $\tau_i^{\min}$ such that for all $\tau \leq \tau_i^{\min}$ the following inequality holds: $\mathrm{IE}(\boldsymbol{u}_i^{n+1}) \geq c_{\mathrm{IE}}\mathrm{IE}(\boldsymbol{u}_i^n)$.

We have stated above the necessary and sufficient condition for the positivity of internal energy at all solution points when the explicit forward Euler scheme is used for time integration. This condition is used to enforce positivity of temperature as discussed in Section 6.4.

6.1.7 ENTROPY STABILITY OF THE FIRST-ORDER SCHEME

We summarize the entropy stability property of the first-order scheme given by Eq. (100).

Theorem 8. *The semi-discrete first-order scheme given by Eq. (100) is entropy stable.*

Proof. The first-order scheme given in Eq. (100) is

$$\hat{\mathbf{U}}_t + \sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \left[\hat{\bar{\mathbf{f}}}_l^{(MR)} - \hat{\bar{\mathbf{f}}}_{\hat{\sigma},l}^{(AD_1)} - \hat{\bar{\mathbf{f}}}_l^{(AD_1)} \right] - D_{\xi^l} \hat{\mathbf{f}}_l^{(v)} = \sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \left[\hat{\mathbf{g}}_l + \hat{\mathbf{g}}_l^{(AD_1)} \right].$$

Entropy stability of the entire scheme follows from the entropy stability of the individual terms. Recall that $\hat{\bar{\mathbf{f}}}_l^{(MR)} = \hat{\bar{\mathbf{f}}}_l^{(EC)} - \hat{\bar{\mathbf{f}}}_l^{(ED)}$ and Lemma 2 showed that $\hat{\bar{\mathbf{f}}}_l^{(EC)}$ has the same element-wise contribution to the total entropy as the high-order EC flux of the baseline scheme (64). Hence, the entropy stability of $\hat{\bar{\mathbf{f}}}_l^{(EC)}$, $\hat{\mathbf{f}}_l^{(v)}$ and $\hat{\mathbf{g}}_l$ follow directly from the baseline scheme. The additional terms–$\hat{\bar{\mathbf{f}}}_l^{(ED)}$, $\hat{\bar{\mathbf{f}}}_l^{(AD_1)}$,$\hat{\bar{\mathbf{f}}}_{\hat{\sigma},l}^{(AD_1)}$, and $\hat{\mathbf{g}}_l^{(AD_1)}$–are all formed from 2-point SPSD matrices and hence are entropy stable according to Lemma 18. □

6.2 ENTROPY STABLE VELOCITY AND TEMPERATURE LIMITERS FOR VISCOUS FLOWS

As discussed in the previous section, the physical viscous terms of the first-order scheme are discretized by using the same high-order SBP operators used for the high-order scheme, which improves the accuracy when the solution of the Navier-Stokes equations has enough regularity. Note, however, that the high-order discretization of the viscous terms can significantly increase the stiffness of the temperature positivity time step constraint. To overcome

this problem, we construct new conservative, discretely entropy stable limiters that bound the magnitude of the velocity and temperature gradients in troubled elements. The proposed approach differs from the limiter in [64] in two distinct ways: 1) density is not altered at any solution point and 2) we apply the limiter before negativity is encountered. The benefit of (2) is that one can then prove discrete entropy stability.

6.2.1 BOUNDS ON VELOCITY AND TEMPERATURE

To bound the viscous fluxes in troubled elements, we propose to limit the deviation of velocity and temperature values at solution points from the corresponding arithmetic averages computed on the same high-order element. Note that this same limiting procedure can be used to bound the deviation from other convex averages as well (e.g., cell averages). Taking into account the contribution of velocity and temperature terms to the high-order approximation of the gradient of entropy variables and consequently to the viscous fluxes, we propose to impose the following bounds on $(V_l)_i$ and T_i at each solution point of the troubled element:

$$|(V_l)_i - \bar{\bar{V}}_l| \leq \frac{\bar{\bar{\rho}}_H h \bar{\bar{T}}_H}{\mu}, \;\; \tilde{\lambda}_i \frac{|T_i - \bar{\bar{T}}|}{T_i \bar{\bar{T}}} \leq \frac{\bar{\bar{\rho}}_H h}{\mu}, \tag{137}$$

where

$$\tilde{\lambda}_i = \frac{\| \vec{\boldsymbol{V}}_i + \overline{\overline{\vec{\boldsymbol{V}}}} \|}{2} + \frac{c_i + c(\bar{\bar{T}})}{2}, \tag{138}$$

$$\bar{\bar{q}} = \frac{1}{N_p} \sum_{j=1}^{N_p} q_j \tag{139}$$

is the arithmetic average of a quantity q on a high-order element, $\bar{\bar{\rho}}_H$ is the harmonic average of ρ_i and $\bar{\bar{\rho}}$, μ is the physical viscosity coefficient, $c(\bar{\bar{T}})$ is the speed of sound associated with the average temperature and h is a reference length for the element e.g. cubed root of the volume. Note that the corresponding cell averages on a given high-order element can also be used instead of the arithmetic averages in Eqs. (137) and (138).

We now construct velocity and temperature limiters such that they ensure the bounds

given by Eq. (137) without changing the density at any solution point, preserve the conservation of mass, momentum and energy, and can only decrease the discrete integral of the mathematical entropy on a given element. The limiting procedure is broken into two steps. The first step enforces the velocity bound while altering the temperature field in a pointwise discretely entropy stable manner. The second step enforces the temperature bound by only altering the energy equation in an elementwise entropy stable manner.

6.2.2 A LIMITER TO ENFORCE THE VELOCITY BOUND

First, we modify the velocity at each solution point on a given high-order element, so that it satisfies Eq. (137). Let $\vec{\xi}_{ijk} = \vec{\xi}_a$ be some solution point on the element. To enforce this velocity bound, we propose the following limiter:

$$\hat{\mathbf{U}}_a^v = \hat{\mathbf{U}}_a + \frac{1}{\mathcal{P}_a} \boldsymbol{f}_v(\mathbf{U}_a, \boldsymbol{\theta}^v), \tag{140}$$

where $\hat{\mathbf{U}}(\vec{\xi}_{ijk})^v = \hat{\mathbf{U}}_a^v$, $\mathcal{P}_{ijk} = \mathcal{P}_a$, $\boldsymbol{\theta}^v = \begin{bmatrix} \theta_1^v & \theta_2^v & \theta_3^v \end{bmatrix}^\top$,

$$\boldsymbol{f}_v(\mathbf{U}_a, \boldsymbol{\theta}^v) = \rho_{\min} \begin{bmatrix} 0 & 0 & 0 \\ \theta_1^v & 0 & 0 \\ 0 & \theta_2^v & 0 \\ 0 & 0 & \theta_3^v \\ \theta_1^v \bar{\bar{V}}_1 & \theta_2^v \bar{\bar{V}}_2 & \theta_3^v \bar{\bar{V}}_3 \end{bmatrix} \left(\bar{\bar{\vec{\boldsymbol{V}}}} - \vec{\boldsymbol{V}}_a \right), \tag{141}$$

$\rho_{\min}$ is the minimum density on the element and $\bar{\bar{\vec{\boldsymbol{V}}}}$ is the arithmetic average of velocity on the high-order element.

Note that the temperature after applying the velocity limiter is given by

$$T(\mathbf{U}_a^v) = T(\mathbf{U}_a) + \Delta \vec{\boldsymbol{V}} M(\mathbf{U}_a, \boldsymbol{\theta}^v) \Delta \vec{\boldsymbol{V}}, \tag{142}$$

where

$$M(\mathbf{U}_a, \boldsymbol{\theta}^v) = \frac{\gamma - 1}{R_g} \frac{\rho_{\min}}{\rho_a J_a \mathcal{P}_a} \\ \text{diag}\left[\ \theta_1^v \left(1 - \frac{\theta_1^v \rho_{\min}}{2\rho_a J_a \mathcal{P}_a}\right) \quad \theta_2^v \left(1 - \frac{\theta_2^v \rho_{\min}}{2\rho_a J_a \mathcal{P}_a}\right) \quad \theta_3^v \left(1 - \frac{\theta_3^v \rho_{\min}}{2\rho_a J_a \mathcal{P}_a}\right) \ \right], \tag{143}$$

and $\Delta \vec{\boldsymbol{V}} = \overline{\overline{\vec{\boldsymbol{V}}}} - \vec{\boldsymbol{V}}_a$. Hence, if $0 \leq \theta_l^v \leq 2\mathcal{P}_a J_a \frac{\rho_a}{\rho_{\min}} \quad \forall\, l$ then $T(\mathbf{U}_a^v) \geq T(\mathbf{U}_a)$ and $S(\mathbf{U}_a^v) \leq S(\mathbf{U}_a)$. As follows from Eqs. (140– 141), the velocity components of $\hat{\mathbf{U}}_a^v$ obey:

$$V_l(\hat{\mathbf{U}}_a^v) - \bar{\bar{V}}_l = \left(V_l(\hat{\mathbf{U}}_a) - \bar{\bar{V}}_l\right) \left(1 - \frac{\theta_l^v}{J_a \mathcal{P}_a} \frac{\rho_{\min}}{\rho_a}\right). \tag{144}$$

Since $\overline{\overline{\vec{\boldsymbol{V}}}}$ may be changed by the limiting procedure, enforcing the velocity bound at each solution point on a given element should in principle be done iteratively, i.e., Eq. (144) can be recast in the following form:

$$(V_l)_a^{(m)} - \bar{\bar{V}}_l^{(m-1)} = \left((V_l)_a^{(m-1)} - \bar{\bar{V}}_l^{(m-1)}\right) \left(1 - \frac{(\theta_l^v)^{(m)}}{J_a \mathcal{P}_a} \frac{\rho_{\min}}{\rho_a}\right), \tag{145}$$

where the superscript is the iteration number and $\bar{\bar{V}}_l^{(m)} = \frac{1}{N_p} \sum_{j=1}^{N_p} (V_l)_j^{(m)}$. Each iteration begins by finding $\boldsymbol{\theta}_a^v$ for all solution points on the element. If the lth velocity component of $\hat{\mathbf{U}}_a$ violates the velocity bound given by Eq. (137), then we solve Eqs. (137, 144) for θ_l^v and set

$$\left(\theta_l^v\right)_a = \mathcal{P}_a J_a \frac{\rho_a}{\rho_{\min}} \left(1 - \frac{\bar{\bar{\rho}}_H h \bar{\bar{T}}_H}{\mu \left| \left(V_l\right)_a - \bar{\bar{V}}_l \right|}\right), \tag{146}$$

otherwise we set $\left(\theta_l^v\right)_a = 0$. Finally, we calculate θ_l^v as follows:

$$\theta_l^v = \min(\max_a((\theta_l^v)_a), \min_a(\mathcal{P}_a J_a \frac{\rho_a}{\rho_{\min}})), \tag{147}$$

alter the vector of conservative variables at each point on the element according to Eq. (140), update the velocity average, and repeat this iterative process until convergence. The key properties of the proposed velocity limiter are given in Theorem 10. First, we prove the following lemma.

Lemma 9. *After the mth iteration of the method given by Eqs. (140– 141, 145–147), for*

any lth component of velocity there exist two solution points $i_{l,\max}^{(m)}$ and $i_{l,\min}^{(m)}$ such that for all $1 \leq j \leq N_p$ on a given element

$$(V_l)_{i_{l,\min}^{(m)}}^{(m)} \leq (V_l)_j^{(m)} \leq (V_l)_{i_{l,\max}^{(m)}}^{(m)}, \tag{148}$$

$$(V_l)_{i_{l,\min}^{(m)}}^{(m-1)} \leq \bar{\bar{V}}_l^{(m-1)} \leq (V_l)_{i_{l,\max}^{(m)}}^{(m-1)}, \tag{149}$$

where $(V_l)_{i_{l,\min}^{(m)}}^{(m-1)}$ is the velocity at solution point $i_{l,\min}^{(m)}$ before the mth iteration.

Proof. We prove the existence of an $i_{l,\max}^{(m)}$ satisfying both inequalities. Let $(V_l)_a^{(m)} = \max_{1\leq j\leq N_p} (V_l)_j^{(m)}$ so that the index "a" satisfies the role of $i_{l,\max}^{(m)}$ in (148). If a also satisfies (149), then we can set $a = i_{l,\max}^{(m)}$ and hence such an $i_{l,\max}^{(m)}$ exists. Suppose that $(V_l)_a^{(m-1)} < \bar{\bar{V}}_l^{(m-1)}$. Then, there exists $(V_l)_b^{(m-1)} > \bar{\bar{V}}_l^{(m-1)}$ and by Eqs. (145) and (147) we must have $(V_l)_b^{(m)} \geq \bar{\bar{V}}_l^{(m-1)}$. Note that we cannot have $(V_l)_a^{(m)} > (V_l)_b^{(m)} \geq \bar{\bar{V}}_l^{(m-1)}$ since then by Eq. (145) $(V_l)_a^{(m-1)} > \bar{\bar{V}}_l^{(m-1)}$. Thus, $(V_l)_a^{(m)} = (V_l)_b^{(m)}$ so that the bth solution point satisfies (148) and (149). Hence, we can take $b = i_{l,\max}^{(m)}$ so that again we have found a solution point $i_{l,\max}^{(m)}$ satisfying both (148) and (149). An identical argument holds for $i_{l,\min}^{(m)}$. □

Theorem 10. *The iterative method given by Eqs. (140– 141, 145–147) is conservative and pointwise entropy dissipative. Also, the maximum possible velocity variation after m iterations is bounded as follows:*

$$\begin{aligned}\max_a((V_l)_a^{(m)}) - \min_a((V_l)_a^{(m)}) \leq (\max_a((V_l)_a^{(0)}) - \min_a((V_l)_a^{(0)})) \\ \prod_{n=1}^{m}\left(1 - \frac{(\theta_l^v)^{(n)}}{\max_a(\mathcal{P}_a J_a \frac{\rho_a}{\rho_{\min}})}\right) \quad \forall l.\end{aligned} \tag{150}$$

Furthermore, this iterative method converges, so that the velocity at all solution points satisfy the bound given by Eq. (137) upon convergence.

Proof. For each iteration, θ_l^v is computed using Eq. (147), so that $\theta_l^v \leq \min_a(\mathcal{P}_a J_a \frac{\rho_a}{\rho_{\min}})$ and the temperature at each solution point may only increase as follows from Eq. (142). Since the density at each solution point remains unchanged during this limiting procedure, the mathematical entropy can only decrease. Therefore, this iterative method is pointwise

entropy dissipative. Conservation follows from the fact that at each iteration $(\theta_l^v)^{(m)}$ is a constant on each high-order element and $\sum_{a=1}^{N_p} \mathcal{P}_a(\frac{1}{\mathcal{P}_a}\boldsymbol{f}_v(\mathbf{U}_a, \boldsymbol{\theta}^v)) = 0$.

Convergence follows from the fact the iteration given by Eq. ((140– 141, 145–147)) is contractive. Indeed, let $i_{l,\min}^{(m)}$ and $i_{l,\max}^{(m)}$ be defined as in Lemma 9. Taking into account that $(\theta^v)_l^{(m)} \leq \min_a(\mathcal{P}_a J_a \frac{\rho_a}{\rho_{\min}})$ $\forall m$ and using Eq. (144), we have

$$
\begin{aligned}
\max_a((V_l)_a^{(m)}) - \min_a((V_l)_a^{(m)}) = & \qquad (151)\\
&= (V_l)^{(m)}_{i^{(m)}_{l,\max}} - \bar{\bar{V}}_l^{(m-1)} + \bar{\bar{V}}_l^{(m-1)} - (V_l)^{(m)}_{i^{(m)}_{l,\min}}\\
&= ((V_l)^{(m-1)}_{i^{(m)}_{l,\max}} - \bar{\bar{V}}_l^{(m-1)}) \left(1 - \frac{(\theta_l^v)^{(m)}}{\mathcal{P}_{i^{(m)}_{l,\max}} J_{i^{(m)}_{l,\max}}} \frac{\rho_{\min}}{\rho_{i^{(m)}_{l,\max}}}\right)\\
&+ (\bar{\bar{V}}_l^{(m-1)} - (V_l)^{(m-1)}_{i^{(m)}_{l,\min}}) \left(1 - \frac{(\theta_l^v)^{(m)}}{\mathcal{P}_{i^{(m)}_{l,\min}} J_{i^{(m)}_{l,\min}}} \frac{\rho_{\min}}{\rho_{i^{(m)}_{l,\min}}}\right)\\
&\leq ((V_l)^{(m-1)}_{i^{(m)}_{l,\max}} - (V_l)^{(m-1)}_{i^{(m)}_{l,\min}}) \left(1 - \frac{(\theta_l^v)^{(m)}}{\max_a(\mathcal{P}_a J_a \frac{\rho_a}{\rho_{\min}})}\right)\\
&\leq \left(\max_a((V_l)_a^{(m-1)}) - \min_a((V_l)_a^{(m-1)})\right) \left(1 - \frac{(\theta_l^v)^{(m)}}{\max_a(\mathcal{P}_a J_a \frac{\rho_a}{\rho_{\min}})}\right)\\
&\vdots\\
&\leq \left(\max_a((V_l)_a^{(0)}) - \min_a((V_l)_a^{(0)})\right) \prod_{n=1}^{m} \left(1 - \frac{(\theta_l^v)^{(n)}}{\max_a(\mathcal{P}_a J_a \frac{\rho_a}{\rho_{\min}})}\right).
\end{aligned}
$$

□

Remark 9. The convergence rate of this iterative method depends on $\frac{\min_a(\mathcal{P}_a J_a \rho_a)}{\max_a(\mathcal{P}_a J_a \rho_a)}$. For LGL grids with $p > 1$, this ratio is usually small. If large variations in density are present and causing slow convergence, one could limit the entire vector of conserved variables which can also be done in a discretely entropy stable manner. Note, however, for all test problems presented in this dissertation, only one iteration of the above iterative method per Runge-Kutta stage is sufficient to eliminate the stiffness of the time step constraint for temperature positivity for each troubled element.

6.2.3 A LIMITER TO ENFORCE THE TEMPERATURE BOUND

The second step is to enforce the bound on temperature, which is given by Eq. (137). Similar to the velocity limiter, we modify the temperature at each solution point by using the following limiter:

$$\hat{\mathbf{U}}_a^t = \hat{\mathbf{U}}_a + \frac{\theta^t}{\mathcal{P}_a} \boldsymbol{f}_t(\hat{\mathbf{U}}_a), \tag{152}$$

where

$$\boldsymbol{f}_t(\hat{\mathbf{U}}_a) = \rho_{\min} \begin{bmatrix} 0 & 0 & 0 & 0 & (\bar{\bar{T}} - T_a) \end{bmatrix}^\top, \tag{153}$$

and $\bar{\bar{T}}$ is the arithmetic average of temperature on a given high-order element. After applying the limiter, the modified temperature is given by

$$T(\hat{\mathbf{U}}_a^t) = T_a + \frac{\gamma - 1}{R_g} (\bar{\bar{T}} - T_a) \frac{\theta^t \rho_{\min}}{J_a \mathcal{P}_a \rho_a}, \tag{154}$$

and obeys:

$$T(\hat{\mathbf{U}}_a^t) - \bar{\bar{T}} = (T_a - \bar{\bar{T}}) \left(1 - \frac{\gamma - 1}{R_g} \frac{\theta^t \rho_{\min}}{J_a \mathcal{P}_a \rho_a} \right). \tag{155}$$

If $\hat{\mathbf{U}}_a$ violates the temperature bound given by Eq. (137), then we set

$$\theta_a^t = J_a \mathcal{P}_a \frac{\rho_a}{\rho_{\min}} \frac{R_g}{\gamma - 1} \left(1 - \frac{T_a \bar{\bar{T}}}{|T_a - \bar{\bar{T}}|} \frac{\bar{\bar{\rho}}_H h}{\tilde{\lambda}_a \mu} \right), \tag{156}$$

otherwise we set $\theta_a^t = 0$. Note that by construction, $0 \leq \theta_a^t \leq 1$, $\forall a$. Finally, the temperature limiter is defined as follows:

$$\theta^t = \min \left(\max_a(\theta_a^t), \frac{R_g}{\gamma - 1} \min_a (J_a \mathcal{P}_a \frac{\rho_a}{\rho_{\min}}) \right) \tag{157}$$

and the vector of conservative variables at all solution points on the element is modified according to Eq. (152). Similar to the velocity limiter, the temperature limiting procedure should in general be performed iteratively. The key properties of the proposed temperature limiter are presented in the following theorem.

Theorem 11. *The iterative temperature limiting procedure given by Eqs. (152, 153, 156,*

157) is conservative and elementwise entropy dissipative. Also, the maximum possible temperature variation after m iterations is bounded as follows:

$$\max_a(T_a^{(m)}) - \min_a(T_a^{(m)}) \leq (\max_a(T_a^{(0)}) - \min_a(T_a^{(0)})) \prod_{n=1}^{m} \left(1 - \frac{\gamma - 1}{R_g} \frac{(\theta^t)^{(n)}}{\max_a(J_a \mathcal{P}_a \frac{\rho_a}{\rho_{\min}})}\right). \tag{158}$$

Furthermore, this iterative method converges, so that the temperature at all solution points satisfies the bound given by Eq. (137) upon convergence.

Proof. Conservation follows from the fact that $(\theta^t)^{(m)}$ is a constant on each high-order element and $\sum_{a=1}^{N_p} \mathcal{P}_a(\frac{(\theta^t)^{(m)}}{\mathcal{P}_a} \boldsymbol{f}_t(\hat{\mathbf{U}}_a^{(m-1)})) = 0$.

Let us show that the temperature limiting procedure is elementwise entropy dissipative, i.e.,

$$\sum_{a=1}^{N_p} \mathcal{P}_a J_a S(\mathbf{U}_a^t) \leq \sum_{a=1}^{N_p} \mathcal{P}_a J_a S(\mathbf{U}_a). \tag{159}$$

Note that it is sufficient to show that the entropy dissipates at the first iteration, because the same argument holds for all other iterations as well. Let I_L, I_E and I_G be the following index sets: $T_a < \bar{\bar{T}}\ \forall a \in I_L$, $T_a > \bar{\bar{T}}\ \forall a \in I_G$, and $T_a = \bar{\bar{T}}\ \forall a \in I_E$. Taking into account that

$$\frac{dS(\mathbf{U}_a(\theta^t))}{d\theta^t} = -\frac{\rho_{\min}}{J_a \mathcal{P}_a} \frac{\bar{\bar{T}} - T_a}{T(\mathbf{U}_a(\theta^t))}, \tag{160}$$

we have

$$\begin{aligned}
\frac{d}{d\theta^t}\sum_{a=1}^{N_p}\mathcal{P}_a J_a S(\mathbf{U}_a(\theta^t)) &= \sum_{a=1}^{N_p}\mathcal{P}_a J_a \frac{dS(\mathbf{U}_a(\theta^t))}{d\theta^t} = -\rho_{\min}\sum_{a=1}^{N_p}\frac{\bar{\bar{T}}-T_a}{T(\mathbf{U}_a(\theta^t))} \\
&= \rho_{\min}\left(-\sum_{a\in I_L}\frac{\bar{\bar{T}}-T_a}{T(\mathbf{U}_a(\theta^t))}+\sum_{a\in I_G}\frac{T_a-\bar{\bar{T}}}{T(\mathbf{U}_a(\theta^t))}\right) \\
&\leq \rho_{\min}\left(-\sum_{a\in I_L}\frac{\bar{\bar{T}}-T_a}{\bar{\bar{T}}}+\sum_{a\in I_G}\frac{T_a-\bar{\bar{T}}}{\bar{\bar{T}}}\right) \\
&= \frac{\rho_{\min}}{\bar{\bar{T}}}\left(-\sum_{a\in I_L}\bar{\bar{T}}-T_a+\sum_{a\in I_G}T_a-\bar{\bar{T}}\right) \\
&= \frac{\rho_{\min}}{\bar{\bar{T}}}\left(-N_p\bar{\bar{T}}+\sum_{a=1}^{N_p}T_a\right)=0, \qquad (161)
\end{aligned}$$

so long as $0 \leq \theta^t \leq \frac{R_g}{\gamma-1}\min_a(J_a\mathcal{P}_a\frac{\rho_a}{\rho_{\min}})$ which is the case when θ^t is selected according to Eq. (157). Thus, $\sum_{a=1}^{N_p}\mathcal{P}_a J_a S(\mathbf{U}_a(\theta^t))$ is non-increasing as a function of θ^t on $\leq \theta^t \leq \frac{R_g}{\gamma-1}\min_a(J_a\mathcal{P}_a\frac{\rho_a}{\rho_{\min}})$ and satisfies Eq. (159). The proof of the temperature bound given by Eq. (158) relies on Eqs. (155, 157) and is nearly identical to the proof of the velocity bound (Eq. (150)) and therefore not presented herein. Together Eq. (158) and Eq. (157) imply that the temperature variation decreases with each iteration. Hence, the bound in (137) is met after a finite number of iterations, because $\min_a(T_a) \leq T_a^{(m)} \leq \max_a(T_a)\ \forall m$ and the following lower bound holds:

$$\frac{T_a^{(m)}\bar{\bar{T}}^{(m)}\bar{\bar{\rho}}_H h}{\tilde{\lambda}_a^{(m)}\mu} \geq \frac{(\min_a(T_a))^2\bar{\bar{\rho}}_H h}{\left[\max_a(\|\boldsymbol{V}\|_a)+c(\max_a(T_a))\right]\mu} > 0, \qquad (162)$$

where $\bar{\bar{T}}^{(m)} = \frac{1}{N_p}\sum_{a=1}^{N_p}T_a^{(m)}$ represents the temperature average after the mth iteration. □

We would like to emphasize again that only one iteration per each troubled element per Runge–Kutta stage was sufficient to control the temperature positivity time constraint for all test problems considered in this dissertation.

6.2.4 CONSISTENCY OF THE VELOCITY AND TEMPERATURE LIMITING PROCEDURE

As has been mentioned above, the velocity and temperature limiters are only applied in troubled elements. We use two criteria for determining troubled elements. The limiting procedure is applied only when both criteria are met. The first criterion is that the time step restriction for pointwise temperature positivity for the element must be stricter than the global time step chosen based on density positivity and the CFL condition. To define the second criterion for determining troubled elements where this limiting procedure should be used, we note that the bounds in Eq. (137) require that the velocity and temperature gradients are bounded from above by a quantity that is of the order of the Reynolds number $O(Re)$. Taking into account that such gradients occur at strong discontinuities, we flag all troubled elements that satisfy the following condition:

$$Sn^k > C_{\text{tol}}, \tag{163}$$

where $0 \leq Sn^k \leq 1$ (85) is the entropy residual sensor for the kth element and C_{tol} is a user-defined parameter that is set equal to 0.9 for all test problems considered in this dissertation. As follows from Eq. (163), this condition is satisfied only in those elements where the residual of the entropy equation is O(1), which occurs only if the solution is discontinuous or fully unresolved. For smooth solutions, the entropy residual given by Eq. (81) is of the order of $O(h^{p-1})$, which implies that these elements will never be flagged for the velocity and temperature limiting. Hence, the above limiting procedure is design-order accurate. It should be pointed out that the velocity and temperature bounds given by Eq. (137) can in principle be violated even for smooth solutions if the Reynolds number is $O(1)$. Note, however, that in this case, the condition (163) is not satisfied.

6.3 HIGH-ORDER POSITIVITY–PRESERVING SCHEME

In contrast to the limiting approach developed in [64], for which an entropy stability proof is not available, we propose a novel limiting scheme that is design-order accurate (for smooth solutions), entropy stable, and pointwise positive for the thermodynamic variables.

This high-order positivity–preserving flux-limiting scheme is constructed by using a convex combination of the positivity-violating high-order spectral collocation scheme (Eq. (64)) and the first-order positivity–preserving finite volume scheme (Eq. (100)). The key properties of this scheme are presented next.

6.3.1 POSITIVITY

We begin with some notation. Assume that the time derivative term in Eq. (9) is approximated by using the 1st-order explicit Euler scheme, so that on a given element we have

$$\hat{\mathbf{U}}_p^{n+1} = \hat{\mathbf{U}}^n + \tau \left(\frac{d\hat{\mathbf{U}}}{dt} \right)_p ,$$

$$\hat{\mathbf{U}}_1^{n+1} = \hat{\mathbf{U}}^n + \tau \left(\frac{d\hat{\mathbf{U}}}{dt} \right)_1 ,$$

where $\hat{\mathbf{U}}_p^{n+1}$ and $\hat{\mathbf{U}}_1^{n+1}$ are pth-order and first-order numerical solutions defined on the same grid element with the same metric terms so that $\hat{\mathbf{U}}_p^{n+1} = [J]\,\mathbf{U}_p^{n+1}$ and $\hat{\mathbf{U}}_1^{n+1} = [J]\,\mathbf{U}_1^{n+1}$. Since the first-order scheme presented in Section 6.1 is positivity preserving, we assume that at every ith solution point on the element $\mathrm{IE}((\hat{\mathbf{U}}_1^{n+1})_i) > 0$ and $(\rho_1^{n+1})_i > 0$, where $\mathrm{IE}((\hat{\mathbf{U}}_1^{n+1})_i)$ is the internal energy associated with the 1st-order solution $(\hat{\mathbf{U}}_1^{n+1})_i$.

To combine the 1st- and pth-order schemes, we use a flux-limiting approach, which is in fact equivalent to limiting the low- and high-order solution vectors of the conservative variables. Indeed, the solution vector obtained using the flux-limiting approach can be represented as follows:

$$\begin{aligned} \hat{\mathbf{U}}^{n+1}(\theta_f) &= \hat{\mathbf{U}}^n + \tau \left[(1-\theta_f) \left(\frac{d\hat{\mathbf{U}}}{dt} \right)_1 + \theta_f \left(\frac{d\hat{\mathbf{U}}}{dt} \right)_p \right] \\ &= (1-\theta_f)\hat{\mathbf{U}}_1^{n+1} + \theta_f \hat{\mathbf{U}}_p^{n+1} \\ &= \hat{\mathbf{U}}_1^{n+1} + \theta_f [\hat{\mathbf{U}}_p^{n+1} - \hat{\mathbf{U}}_1^{n+1}], \end{aligned} \tag{164}$$

where the flux limiter θ_f, $0 \le \theta_f \le 1$, is a constant on a given high-order element.

Let us define a function $\aleph$, $0 < \aleph < 1$, so that it approaches to zero and is bounded from

below by a small positive number (e.g., 10^{-8}) for elements where the solution is smooth and goes to 1 for elements where the solution loses its regularity. In the present analysis, $\aleph$ is defined as follows:

$$\aleph^k = \max(10^{-8}, L^k), \qquad L^k = Sn^k \max_i \left(\frac{|\Delta P|}{2P_A} \right), \tag{165}$$

where $0 \leq Sn^k \leq 1$ is the residual-based sensor given by Eq. (85) and $0 \leq \max_i \left(\frac{|\Delta P|}{2P_A} \right) < 1$ is one half the maximum relative two–point pressure jump (including jumps at the interfaces) on the kth element.

At each solution point, we define local lower bounds for density and internal energy as follows:

$$\epsilon_i^\rho = (\rho_1)_i^{n+1} \aleph, \qquad \epsilon_i^{\text{IE}} = \text{IE}((\hat{\mathbf{U}}_1)_i^{n+1}) \aleph. \tag{166}$$

Note that since $0 \leq L^k < 1$, $0 < \epsilon_i^\rho < (\rho_1)_i^{n+1}$ and $0 < \epsilon_i^{\text{IE}} < \text{IE}((\hat{\mathbf{U}}_1)_i^{n+1})$. We now prove the following two lemmas.

Lemma 12. *For every ith solution point, define the set*

$$H_i^\rho = \{\theta_f \in [0,1] \mid \rho_i^{n+1}(\theta_f) \geq \epsilon_i^\rho\}.$$

The set H_i^ρ can be written as $H_i^\rho = [0, \theta_i^\rho]$ where $0 < \theta_i^\rho \leq 1$. Furthermore, we have the following statements: (1) if $0 \leq \theta_f < \theta_i^\rho$, then $\rho_i^{n+1}(\theta_f) > \epsilon_i^\rho$ and (2) if $\theta_i^\rho < 1$, then $\rho_i^{n+1}(\theta_i^\rho) = \epsilon_i^\rho$.

Proof. This follows directly from the fact that $\rho_i^{n+1}(\theta_f)$ given by Eq. (164) is a linear equation in the variable θ_f with $\rho_i^{n+1}(0) > \epsilon_i^\rho$. □

A similar statement can also be proven for the internal energy.

Lemma 13. *For every ith solution point, define the set*

$$H_i^{IE} = \{\theta_f \in H_i^\rho \mid IE(\hat{\mathbf{U}}_i^{n+1}(\theta_f)) \geq \epsilon_i^{IE}\},$$

where $H_i^\rho = [0, \theta_i^\rho]$ was defined in Lemma 12. The set H_i^{IE} can be written as $H_i^{IE} = [0, \theta_i^{IE}]$

where $0 < \theta_i^{IE} \leq \theta_i^{\rho}$. *Furthermore, we have the following statements: (1) if* $0 \leq \theta_f < \theta_i^{IE}$, *then* $IE(\hat{\mathbf{U}}_i^{n+1}(\theta_f)) > \epsilon_i^{IE}$ *and (2) if* $\theta_i^{IE} < \theta_i^{\rho}$, *then* $IE(\hat{\mathbf{U}}_i^{n+1}(\theta_i^{IE})) = \epsilon_i^{IE}$.

Proof. For each ith solution point, if $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\rho})) \geq \epsilon_i^{\text{IE}}$, then we set $\theta_i^{\text{IE}} = \theta_i^{\rho}$. Assume that there is a solution point such that $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\rho})) < \epsilon_i^{\text{IE}}$. Since $\rho_i^{n+1}(\theta_f) \geq \epsilon_i^{\rho} > 0 \; \forall \theta_f \in [0, \theta_i^{\rho}]$, it follows from Eq. (134) that $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_f))$ is a continuous function with respect to θ_f for $\theta_f \in [0, \theta_i^{\rho}]$. Since $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(0)) = \text{IE}((\hat{\mathbf{U}}_1^{n+1})_i) > \epsilon_i^{\text{IE}}$ and $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\rho})) < \epsilon_i^{\text{IE}}$, it follows by the intermediate value theorem that there exists $\theta_i^* \in (0, \theta_i^{\rho})$ such that $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^*)) = \epsilon_i^{\text{IE}}$. Let $\theta_i^{\text{IE}} = \theta_i^*$ (notice that once we establish (1) from the lemma statement we will have shown that there is only one $\theta_i^* \in (0, \theta_i^{\rho})$ such that $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^*)) = \epsilon_i^{\text{IE}}$).

Now we show that for all $0 \leq \theta_f < \theta_i^{\text{IE}}$, we have $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_f)) > \epsilon_i^{\text{IE}}$. By definition of ϵ_i^{IE}, $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(0)) > \epsilon_i^{\text{IE}}$. For $0 < \theta_f < \theta_i^{\text{IE}}$, we have

$$\begin{aligned} \hat{\mathbf{U}}_i^{n+1}(\theta_f) &= (1-\theta_f)(\hat{\mathbf{U}}_1)_i^{n+1} + \theta_f(\hat{\mathbf{U}}_p)_i^{n+1} \\ &= \frac{\theta_f}{\theta_i^{\text{IE}}}\left[\theta_i^{\text{IE}}\left((\hat{\mathbf{U}}_p)_i^{n+1} - (\hat{\mathbf{U}}_1)_i^{n+1}\right) + (\hat{\mathbf{U}}_1)_i^{n+1}\right] + \left(1 - \frac{\theta_f}{\theta_i^{\text{IE}}}\right)(\hat{\mathbf{U}}_1)_i^{n+1} \\ &= \frac{\theta_f}{\theta_i^{\text{IE}}}\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\text{IE}}) + \left(1 - \frac{\theta_f}{\theta_i^{\text{IE}}}\right)(\hat{\mathbf{U}}_1)_i^{n+1}. \end{aligned} \tag{167}$$

Hence, due to the concavity of internal energy

$$\begin{aligned} \text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_f)) &\geq \frac{\theta_f}{\theta_i^{\text{IE}}}\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\text{IE}})) + \left(1 - \frac{\theta_f}{\theta_i^{\text{IE}}}\right)\text{IE}((\hat{\mathbf{U}}_1)_i^{n+1}) \\ &> \frac{\theta_f}{\theta_i^{\text{IE}}}\epsilon_i^{\text{IE}} + \left(1 - \frac{\theta_f}{\theta_i^{\text{IE}}}\right)\epsilon_i^{\text{IE}} = \epsilon_i^{\text{IE}}. \end{aligned} \tag{168}$$

□

Remark 10. Note that θ_i^{IE} in Lemma 13 can readily be found by solving the quadratic equation analogous to Eq. (136).

For a given element, we define $\theta_{\text{IE}} = \min_i\{\theta_i^{\text{IE}}\} > 0$. By construction, $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_{\text{IE}})) \geq \epsilon_i^{\text{IE}}$ and $\rho(\hat{\mathbf{U}}_i^{n+1}(\theta_{\text{IE}})) \geq \epsilon_i^{\rho}$ for every solution point on the element. The solution at the $(n+1)$th time level is set equal to $\hat{\mathbf{U}}^{n+1}(\theta_{\text{IE}})$, which preserves pointwise positivity of density and internal energy.

Remark 11. The above limiting is not immediately conservative for general $\hat{\mathbf{U}}_1^{n+1}$ and $\hat{\mathbf{U}}_p^{n+1}$. We refer the reader to Section 6.3.3 which presents an implementation of this limiting procedure in a way that preserves conservation.

6.3.2 DESIGN ORDER OF ACCURACY

We now show that the proposed limiting preserves the design order of accuracy for smooth solutions and sufficient grid resolution. For simplicity, we assume that the grid resolution depends on one parameter $0 < h^x \leq 1$ such that all element edges are linearly proportional to h^x with an h^x-independent constant of proportionality. In this section, $\|\cdot\|$ denotes the Euclidean norm. Let $\hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1})$ denote the smooth exact solution at the ith solution point when $t = t_{n+1}$. For each solution point, we define a local admissible set

$$\mathcal{A}_i^\epsilon = \{\boldsymbol{u}_i = \begin{bmatrix} \rho & \rho\vec{\boldsymbol{V}} & \rho E \end{bmatrix}^\top \mid \text{IE}(\boldsymbol{u}_i) \geq \epsilon_i^{\text{IE}}, \rho_i \geq \epsilon_i^\rho\}$$

and assume that $\hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1}) \in \mathcal{A}_i^\epsilon$. Note that ϵ_i^{IE} and ϵ_i^ρ are positive user-defined parameters that can be made arbitrarily small by selecting a sufficiently small value of the parameter $\aleph$ for a given element. In the present analysis, $\aleph$, which is given by Eq. (165), is set such that it becomes smaller when the regularity of the numerical solution increases. We also assume that the solution is sufficiently smooth such that $\|(\hat{\mathbf{U}}_1^{n+1})_i - (\hat{\mathbf{U}}_p^{n+1})_i\| \leq \|(\hat{\mathbf{U}}_1^{n+1})_i - \hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1})\| + \|\hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1}) - (\hat{\mathbf{U}}_p^{n+1})_i\| = \mathcal{O}(h^x)$, as $h^x \to 0$.

Let us show that $\|\hat{\mathbf{U}}_i^{n+1}(\theta_{\text{IE}}) - \hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1})\| = \mathcal{O}((h^x)^p)$ for all solution points. If $\theta_i^{\text{IE}} = 1\ \forall i$ on a given element, then $\theta_{\text{IE}} = \min_i\{\theta_i^{\text{IE}}\} = 1$, $\hat{\mathbf{U}}^{n+1}(\theta_{\text{IE}}) = \hat{\mathbf{U}}_p^{n+1}$ and the result follows.

We now assume that $\theta_{\text{IE}} < 1$. In this case, to prove the consistency of the limiting procedure, it is sufficient to show that $1 - \theta_{\text{IE}} = \mathcal{O}((h^x)^{p-1})$. Indeed, if $1 - \theta_{\text{IE}} = \mathcal{O}((h^x)^{p-1})$, then for every solution point we have

$$\begin{aligned} \|\hat{\mathbf{U}}_i^{n+1}(\theta_{\text{IE}}) - \hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1})\| &\leq (1-\theta_{\text{IE}})\|(\hat{\mathbf{U}}_1)_i^{n+1} - \hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1})\| \\ &\quad + \theta_{\text{IE}}\|(\hat{\mathbf{U}}_p)_i^{n+1} - \hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1})\| \\ &= (1-\theta_{\text{IE}})O(h^x) + \theta_{\text{IE}}O((h^x)^p) = O((h^x)^p). \end{aligned} \tag{169}$$

To prove that $1-\theta_{\text{IE}} = 1-\min_i\{\theta_i^{\text{IE}}\} = \mathcal{O}((h^x)^{p-1})$, it is sufficient to show that if $\theta_i^{\text{IE}} < 1$ (which is only possible if $(\hat{\mathbf{U}}_p)_i^{n+1} \notin \mathcal{A}_i^{\epsilon}$), then $1-\theta_i^{\text{IE}} = \mathcal{O}((h^x)^{p-1})\ \forall i$. Assume that at the ith solution point $\theta_i^{\text{IE}} < 1$. Since $\theta_i^{\text{IE}} \leq \theta_i^{\rho}$, we only have to consider the following two cases: 1) $\theta_i^{\text{IE}} = \theta_i^{\rho}$ and $\theta_i^{\rho} < 1$, 2) $\theta_i^{\text{IE}} < \theta_i^{\rho}$.

Case 1. Since $\theta_i^{\rho} < 1$, the following inequalities hold $(\rho_p)_i^{n+1} < \epsilon_i^{\rho} \leq \rho_i^{\text{ex}}(t_{n+1})$, which lead to $(\rho_p)_i^{n+1} = \epsilon_i^{\rho} + \mathcal{O}((h^x)^p)$. From Lemma 12 it follows that θ_i^{ρ} satisfies

$$\rho_i^{n+1}(\theta_i^{\rho}) = (\rho_1)_i^{n+1} + \theta_i^{\rho}((\rho_p)_i^{n+1} - (\rho_1)_i^{n+1}) = \epsilon_i^{\rho}. \tag{170}$$

Thus,

$$1-\theta_i^{\rho} = \frac{\epsilon_i^{\rho} - (\rho_p)_i^{n+1}}{(\rho_1)_i^{n+1} - (\rho_p)_i^{n+1}} = \frac{\mathcal{O}((h^x)^p)}{\mathcal{O}(h^x)} = \mathcal{O}((h^x)^{p-1}). \tag{171}$$

Taking into account that $\theta_i^{\text{IE}} = \theta_i^{\rho}$, we also have $1-\theta_i^{\text{IE}} = \mathcal{O}((h^x)^{p-1})$. Using $0 < \theta_i^{\rho} < 1$ and Eq. (171) yield

$$\begin{aligned} \|\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\rho}) - \hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1})\| &\leq \|\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\rho}) - (\hat{\mathbf{U}}_p)_i^{n+1}\| + \|(\hat{\mathbf{U}}_p)_i^{n+1} - \hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1})\| \\ &= (1-\theta_i^{\rho})\|(\hat{\mathbf{U}}_1)_i^{n+1} - (\hat{\mathbf{U}}_p)_i^{n+1}\| + \mathcal{O}((h^x)^p) \\ &= \mathcal{O}((h^x)^p). \end{aligned} \tag{172}$$

Case 2. Now, we assume that $\theta_i^{\text{IE}} < \theta_i^{\rho}$. First, it should be noted that the internal energy $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\rho}))$ is defined at i, because $\rho_i^{n+1}(\theta_i^{\rho}) \geq \epsilon_i^{\rho} > 0$. As in Case 1, Eq. (172) holds in this case as well, because it has been proven by only assuming $\theta_i^{\rho} < 1$. Furthermore, if $\theta_i^{\rho} = 1$, then $\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\rho}) = (\hat{\mathbf{U}}_p)_i^{n+1}$, which again implies that Eq. (172) holds. Using Eq. (172) yields

$$\begin{aligned} \text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\rho})) &= \rho_i^{n+1}(\theta_i^{\rho})E_i^{n+1}(\theta_i^{\rho}) - \frac{\rho_i^{n+1}(\theta_i^{\rho})}{2}\|\vec{\boldsymbol{V}}_i^{n+1}(\theta_i^{\rho})\|^2 \\ &= \text{IE}(\hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1})) + \mathcal{O}((h^x)^p), \end{aligned} \tag{173}$$

where $E_i^{n+1}(\theta_i^{\rho})$ is the specific total energy of $\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\rho})$. Since $\theta_i^{\text{IE}} < \theta_i^{\rho}$, $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\rho})) < \epsilon_i^{\text{IE}} \leq \text{IE}(\hat{\mathbf{U}}_i^{\text{ex}}(t_{n+1}))$. Therefore, from Eq. (173) it follows that $\text{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^{\rho})) = \epsilon_i^{\text{IE}} + \mathcal{O}((h^x)^p)$.

Using Eq. (167) for $0 < \theta < \theta_i^\rho$, we have

$$\hat{\mathbf{U}}_i^{n+1}(\theta) = \frac{\theta}{\theta_i^\rho}\hat{\mathbf{U}}_i^{n+1}(\theta_i^\rho) + \left(1 - \frac{\theta}{\theta_i^\rho}\right)(\hat{\mathbf{U}}_1)_i^{n+1}. \tag{174}$$

Again, $\hat{\mathbf{U}}_i^{n+1}(\theta)$ may have non-positive internal energy, but it has positive density. Hence, for all $\theta \in (0, \theta_i^\rho)$, $\mathrm{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta))$ is defined at i and the following bound holds:

$$\begin{aligned}\mathrm{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta)) &= \frac{\theta}{\theta_i^\rho}\mathrm{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^\rho)) + \left(1 - \frac{\theta}{\theta_i^\rho}\right)\mathrm{IE}((\hat{\mathbf{U}}_1^{n+1})_i) \\ &\quad + \frac{\rho_i^{n+1}(\theta_i^\rho)(\rho_1^{n+1})_i \left\| (\vec{\boldsymbol{V}}_1^{n+1})_i - \vec{\boldsymbol{V}}_i^{n+1}(\theta_i^\rho) \right\|^2 \frac{\theta}{\theta_i^\rho}\left(1 - \frac{\theta}{\theta_i^\rho}\right)}{2\rho_i^{n+1}(\theta)} \\ &\geq \frac{\theta}{\theta_i^\rho}\mathrm{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^\rho)) + \left(1 - \frac{\theta}{\theta_i^\rho}\right)\mathrm{IE}((\hat{\mathbf{U}}_1^{n+1})_i).\end{aligned} \tag{175}$$

Note that there exists a unique $\theta_i^* \in (0, \theta_i^\rho)$ such that

$$\frac{\theta_i^*}{\theta_i^\rho}\mathrm{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^\rho)) + \left(1 - \frac{\theta_i^*}{\theta_i^\rho}\right)\mathrm{IE}((\hat{\mathbf{U}}_1)_i^{n+1}) = \epsilon_i^{\mathrm{IE}}. \tag{176}$$

From Eq. (175) it follows that $\mathrm{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^*)) \geq \epsilon_i^{\mathrm{IE}}$ and according to Lemma 13, $\theta_i^* \leq \theta_i^{\mathrm{IE}}$. Using Eq. (176) and Eq. (173), we have

$$1 - \frac{\theta_i^*}{\theta_i^\rho} = \frac{\epsilon_i^{\mathrm{IE}} - \mathrm{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^\rho))}{\mathrm{IE}((\hat{\mathbf{U}}_1)_i^{n+1}) - \mathrm{IE}(\hat{\mathbf{U}}_i^{n+1}(\theta_i^\rho))} = \frac{\mathcal{O}((h^x)^p)}{\mathcal{O}((h^x))} = \mathcal{O}((h^x)^{p-1}). \tag{177}$$

Equations (171) and (177) yield $1 - \theta_i^* = \mathcal{O}((h^x)^{p-1})$. Since $\theta_i^* \leq \theta_i^{\mathrm{IE}} < 1$, it follows that $1 - \theta_i^{\mathrm{IE}} = \mathcal{O}((h^x)^{p-1})\ \forall i$ and Eq. (169) holds.

6.3.3 HIGH-ORDER POSITIVITY-PRESERVING FLUX-LIMITING SCHEME

We now present the semi-discrete form of the high-order positivity-preserving scheme.

For the kth element, we have

$$\begin{aligned}
\frac{d\hat{\mathbf{U}}}{dt} &= \theta_f^k \left(\frac{d\hat{\mathbf{U}}}{dt}\right)_p + (1-\theta_f^k)\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_1 + \left(\frac{d\hat{\mathbf{U}}}{dt}\right)_{AD}, \\
\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_p &= \sum_{l=1}^{3} -\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l + D_{\xi^l}\hat{\mathbf{f}}_l^{(v)} + \mathcal{P}_{\xi^l}^{-1}\hat{\mathbf{g}}_l, \\
\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_1 &= \sum_{l=1}^{3} -\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l^{(MR)} + D_{\xi^l}\hat{\mathbf{f}}_l^{(v)} + \mathcal{P}_{\xi^l}^{-1}\hat{\mathbf{g}}_l, \\
\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_{AD} &= \sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\left[(1-\theta_f^k)\hat{\tilde{\mathbf{f}}}_{\hat{\sigma},l}^{(AD_1)} + \hat{\tilde{\mathbf{f}}}_l^{(AD_1)}\right] + D_{\xi^l}\hat{\mathbf{f}}_l^{(AD_p)} \\
&\qquad + \mathcal{P}_{\xi^l}^{-1}\left[\hat{\mathbf{g}}_l^{(AD_1)} + \hat{\mathbf{g}}_l^{(AD_p)}\right],
\end{aligned} \tag{178}$$

where $0 \leq \theta_f^k \leq 1$ is the flux limiter computed independently in each element as described in Section 6.4. Note that the flux limiting is only applied to the inviscid terms and the mass diffusion term required for positivity of density. The term $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_p$ is the baseline high-order scheme with no artificial dissipation, where $\hat{\mathbf{g}}_l$ represents both the inviscid and viscous penalties (see Section 4.1). The remaining terms from $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_1$ and $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_{AD}$ were described in Sections 4.2 and 6.1.

The artificial dissipation terms are proportional to the residual-based sensor given by (85). Therefore, in regions where the solution is sufficiently smooth and resolved the scheme given by Eq. (178) becomes design-order accurate as described in Section 6.3.2.

6.3.4 CONSERVATION

Since θ_f^k is set independently on each element, it is not immediately clear that the scheme given by Eq. (178) is conservative for all $0 \leq \theta_f^k \leq 1$. Let us show that the scheme is indeed conservative.

Theorem 14. *The high-order positivity–preserving flux-limiting scheme given by* (178) *is conservative for all* $0 \leq \theta_f^k \leq 1$.

Proof. Collecting like terms in Eq. (178), shows that θ_f^k only affects the amount of $\hat{\tilde{\mathbf{f}}}_l$, $\hat{\tilde{\mathbf{f}}}_l^{(MR)}$, and $\hat{\tilde{\mathbf{f}}}_{\hat{\sigma},l}^{(AD_1)}$ used on the element. $\hat{\tilde{\mathbf{f}}}_{\hat{\sigma},l}^{(AD_1)}$ is only defined at the interior flux points (see Eq. (115))

and hence the flux differencing form immediately implies $\hat{\tilde{\mathbf{f}}}^{(AD_1)}_{\hat{\sigma},l}$ is conservative. Therefore, we need only note that we have

$$
\begin{aligned}
&\sum_{l=1}^{3} \mathbf{1}_1^\top \mathcal{P}\mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \left[\theta_f^k \hat{\tilde{\mathbf{f}}}_l + (1-\theta_f^k) \hat{\tilde{\mathbf{f}}}_l^{(MR)} \right] \\
&= \sum_{j,k=1}^{N} \mathcal{P}_{jk} \sum_{i=1}^{N} \Big[\theta_f^k \left(\hat{\tilde{\mathbf{f}}}_1(\vec{\xi}_{ijk}) - \hat{\tilde{\mathbf{f}}}_1(\vec{\xi}_{i-1jk}) \right) \\
&\qquad\qquad + (1-\theta_f^k) \left(\hat{\tilde{\mathbf{f}}}_1^{(MR)}(\vec{\xi}_{ijk}) - \hat{\tilde{\mathbf{f}}}_1^{(MR)}(\vec{\xi}_{i-1jk}) \right) \Big] + \cdots \\
&= \sum_{j,k=1}^{N} \mathcal{P}_{jk} \Big[\theta_f^k \left(\hat{\tilde{\mathbf{f}}}_1(\vec{\xi}_{Njk}) - \hat{\tilde{\mathbf{f}}}_1(\vec{\xi}_{1jk}) \right) \\
&\qquad\qquad + (1-\theta_f^k) \left(\hat{\tilde{\mathbf{f}}}_1^{(MR)}(\vec{\xi}_{Njk}) - \hat{\tilde{\mathbf{f}}}_1^{(MR)}(\vec{\xi}_{1jk}) \right) \Big] + \cdots \\
&= \sum_{j,k=1}^{N} \mathcal{P}_{jk} \left[\hat{\tilde{\mathbf{f}}}_1(\vec{\xi}_{Njk}) - \hat{\tilde{\mathbf{f}}}_1(\vec{\xi}_{1jk}) \right] \\
&+ \sum_{i,k=1}^{N} \mathcal{P}_{ik} \left[\hat{\tilde{\mathbf{f}}}_2(\vec{\xi}_{iNk}) - \hat{\tilde{\mathbf{f}}}_2(\vec{\xi}_{i1k}) \right] + \sum_{i,j=1}^{N} \mathcal{P}_{ij} \left[\hat{\tilde{\mathbf{f}}}_3(\vec{\xi}_{ijN}) - \hat{\tilde{\mathbf{f}}}_3(\vec{\xi}_{ij1}) \right] \\
&= \sum_{l=1}^{3} \mathbf{1}_1^\top \mathcal{P}\mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\tilde{\mathbf{f}}}_l,
\end{aligned}
\tag{179}
$$

where the second to last equality follows by comparing (101), (69), and (120). Hence, conservation for the flux-limiting scheme given by Eq. (178) follows directly from conservation of the baseline scheme given by Eq. (64). □

6.3.5 ARTIFICIAL VISCOSITY FOR THE FLUX-LIMITING SCHEME

The construction of the artificial viscosity, $\boldsymbol{\mu}^{AD}$, was discussed in Chapter 5. Here, we present how the viscosity is set for the flux-limiting scheme given by Eq. (178). The artificial viscosity controls the artificial dissipation terms of $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_{AD}$ in Eq. (178). If an element is considered for flux limiting (i.e, $\theta_f^k < 1$) for positivity or dissipation purposes, then only the first-order dissipation is used for that element, even if it is later determined that $\theta_f^k = 1$. High-order elements considered for flux limiting are herein referred to as "limited elements."

The artificial viscosity coefficient $\boldsymbol{\mu}^{AD}$ presented in Chapter 5 is used to construct the

first-order Brenner dissipation defined at element flux points, $\bar{\boldsymbol{\mu}}_1^{AD}$, and the pth-order Brenner dissipation calculated at element solution points, $\boldsymbol{\mu}_p^{AD}$. Let $V_1^k, V_2^k, \ldots, V_8^k$ be the 8 vertices of the kth element. Define an indicator function, $\chi(\cdot)$, such that $\chi(V_a^k) = 1$ if V_a^k is collocated with or on a limited element; otherwise, $\chi(V_a^k) = 0$. Then, set $\boldsymbol{\mu}_p^{AD}(V_a^k) = \boldsymbol{\mu}^{AD}(V_a^k)(1 - \chi(V_a^k))$ and use tri-linear interpolation to obtain $\boldsymbol{\mu}_p^{AD}$ at the remaining solution points. Notice that for elements with limiting, $\boldsymbol{\mu}_p^{AD} = 0$ and only first-order Brenner dissipation is used. The first-order dissipation is formed as follows for all fixed $1 \leq j, k \leq N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$
\begin{aligned}
&\text{for } 1 \leq i \leq N-1, \\
&\bar{\boldsymbol{\mu}}_1^{AD}(\vec{\xi}_{\bar{i}}) = \frac{\boldsymbol{\mu}^{AD}(\vec{\xi}_i) + \boldsymbol{\mu}^{AD}(\vec{\xi}_{i+1}) - \left(\boldsymbol{\mu}_p^{AD}(\vec{\xi}_i) + \boldsymbol{\mu}_p^{AD}(\vec{\xi}_{i+1})\right)}{2}, \\
&\bar{\boldsymbol{\mu}}_1^{AD}(\vec{\xi}_{\bar{0}}) = \boldsymbol{\mu}^{AD}(\vec{\xi}_1) - \boldsymbol{\mu}_p^{AD}(\vec{\xi}_1), \quad \bar{\boldsymbol{\mu}}_1^{AD}(\vec{\xi}_{\bar{N}}) = \boldsymbol{\mu}^{AD}(\vec{\xi}_N) - \boldsymbol{\mu}_p^{AD}(\vec{\xi}_N),
\end{aligned}
\tag{180}
$$

with identical definitions in the other computational directions. For the first-order artificial dissipation, the c_ρ and c_T coefficients are set equal to those of the pth-order counterpart (see Section 2.4).

We would like for the first-order mass viscosity, $\bar{\boldsymbol{\sigma}}_1^{AD}$, to be in proportion to $\bar{\boldsymbol{\mu}}_1^{AD}$ with the density scaling removed and to preserve density positivity when needed. The density scaling in $\bar{\boldsymbol{\mu}}_1^{AD}$ is removed by dividing through by the geometric average of density at the interface in question. That is, for all fixed $1 \leq j, k \leq N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$
\begin{aligned}
&\text{for } 0 \leq i \leq N, \\
&\bar{\boldsymbol{\sigma}}_1^{AD}(\vec{\xi}_{\bar{i}}) = \max\left(\chi(\vec{\xi}_{\bar{i}})\,(\delta_{0,i} + \delta_{N,i})\,\bar{\boldsymbol{\sigma}}_{\min}(\vec{\xi}_{\bar{i}}), c_\rho \frac{\bar{\boldsymbol{\mu}}_1^{AD}(\vec{\xi}_{\bar{i}})}{\sqrt{\boldsymbol{\rho}(\vec{\xi}_i)\boldsymbol{\rho}(\vec{\xi}_{i+1})}}\right),
\end{aligned}
\tag{181}
$$

where $\boldsymbol{\rho}(\vec{\xi}_0)$ and $\boldsymbol{\rho}(\vec{\xi}_{N+1})$ come from the collocated numerical or boundary state and identical definitions are given in other directions. At every interface collocated with a limited element, the first-order artificial mass viscosity is set so that density positivity is guaranteed through $\bar{\boldsymbol{\sigma}}_{\min}$ which is the minimum mass diffusion for density positivity for the first-order scheme with the explicit Euler discretization in time given by $\sigma_{l,\min}^{+/-}$ in Corollary 6.1.

If there exists one solution point on the element that would otherwise not have positive density, we also require that the mass diffusion for all interior flux points be sufficient for positivity. The total mass diffusion is increased through $\hat{\bar{\boldsymbol{\sigma}}}_1$ which is the mass diffusion used by $\hat{\mathbf{f}}_{\hat{\sigma},l}^{(AD_1)}$ in Eq. (178). Specficially, for all fixed $1 \leq j,k \leq N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we use

$$\begin{aligned} &\text{for } 0 \leq i \leq N, \\ &\hat{\bar{\boldsymbol{\sigma}}}_1(\vec{\xi}_i) = \max\left(\bar{\boldsymbol{\sigma}}_{\min}(\vec{\xi}_i) - \bar{\boldsymbol{\sigma}}_1^{AD}(\vec{\xi}_i), 0\right), \end{aligned} \tag{182}$$

with identical definitions in the other computational directions.

6.3.6 ENTROPY STABILITY

The entropy stability of the high-order positivity–preserving flux-limiting scheme given by Eq. (178) is given by the following theorem.

Theorem 15. *The total entropy of the high-order positivity–preserving flux-limiting scheme given by Eq.* (178) *obeys the following semi-discrete statement:*

$$\begin{aligned} \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\mathbf{S}}_t^k = &\sum_{k=1}^{K} \sum_{l=1}^{3} \Bigg[\mathbf{w}_k^\top \mathcal{P}_{\perp,\xi^l} \left[B_{\xi^l,k}^{(BC)} \hat{\mathbf{f}}_{l,k}^{(v+AD_p)} + \hat{\mathbf{g}}_{l,k}^{(BC,v+AD_p+AD_1)} + \hat{\mathbf{g}}_{l,k}^{(BC,I)} \right] \\ &+ \left(\hat{\mathbf{g}}_{l,k}^{(BC,\boldsymbol{\Theta})} \right)^\top \mathcal{P}_{\perp,\xi^l} \hat{\mathbf{f}}_{l,k}^{(v+AD_p)} + \mathbf{1}_1^\top \widehat{\mathcal{P}}_{\perp,\xi^l} \widehat{B}_{\xi^l,k}^{(BC)} \hat{\mathbf{F}}_{l,k} \Bigg] \\ &- \sum_{k=1}^{K} \Big[H_k^{(v+AD_p+AD_1,D)} + (1-\theta_f^k) \left(H_k^{(\hat{\bar{\sigma}},D)} + H_k^{(MR,D)} \right) \\ &\qquad + L_k^{(Int,v+AD_p+AD_1,D)} + L_k^{(Int,I,D)} \Big], \end{aligned} \tag{183}$$

where we have

1. $H_k^{(v+AD_p+AD_1,D)}, H_k^{(MR,D)}, H_k^{(\hat{\bar{\sigma}},D)}, L_k^{(Int,v+AD_p+AD_1,D)}, L_k^{(Int,I,D)} \geq 0.$
2. $H_k^{(v+AD_p+AD_1,D)} = H_k^{(v,D)} + H_k^{(AD_p,D)} + H_k^{(AD_1,D)}$ *where all of the non-negative* $H_k^{(\cdot,D)}$ *terms are described in Lemmas 18 and 19.*
3. $L_k^{(Int,v+AD_p+AD_1,D)} = L_k^{(Int,v,D)} + L_k^{(Int,AD_p,D)} + L_k^{(Int,AD_1,D)}$ *where all of the non-negative* $L_k^{(Int,\cdot,D)}$ *terms are described in Lemmas 18 and 19.*

4. $\hat{\mathbf{g}}_{l,k}^{(BC,v+AD_p+AD_1)} = \hat{\mathbf{g}}_{l,k}^{(BC,v)} + \hat{\mathbf{g}}_{l,k}^{(BC,AD_p)} + \hat{\mathbf{g}}_{l,k}^{(BC,AD_1)}$ *where* $\hat{\mathbf{g}}_{l,k}^{(BC,\cdot)}(\vec{\xi}_{abc}) = \hat{\mathbf{g}}_{l,k}^{(\cdot)}(\vec{\xi}_{abc})\chi_k^{(BC)}(\vec{\xi}_{abc})$.

5. $\hat{\mathbf{f}}_{l,k}^{(v+AD_p)} = \hat{\mathbf{f}}_{l,k}^{(v)} + \hat{\mathbf{f}}_{l,k}^{(AD_p)}$.

Proof. The theorem follows by directly applying Lemmas 18, 19, or 21 to each term. In particular, notice that Lemma 2 equates the entropy contributions of $\hat{\tilde{\mathbf{f}}}_l$ and $\hat{\tilde{\mathbf{f}}}_l^{(EC)}$ where $\hat{\tilde{\mathbf{f}}}_l^{(MR)} = \hat{\tilde{\mathbf{f}}}_l^{(EC)} - \hat{\tilde{\mathbf{f}}}_l^{(ED)}$. Therefore,

$$\sum_{l=1}^{3} \mathbf{w}^\top \mathcal{P}\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\left[\theta_f^k \hat{\tilde{\mathbf{f}}}_l + (1-\theta_f^k)\hat{\tilde{\mathbf{f}}}_l^{(EC)}\right] = \mathbf{w}^\top \mathcal{P}\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l \tag{184}$$

for all $0 \leq \theta_f^k \leq 1$. Thus, we can apply Lemma 21 to $\theta_f^k \hat{\tilde{\mathbf{f}}}_l + (1-\theta_f^k)\hat{\tilde{\mathbf{f}}}_l^{(EC)}$ together with the inviscid penalties from $\hat{\mathbf{g}}_l$. □

6.3.7 FREESTREAM PRESERVATION

For curvilinear meshes, freestream preservation is an important property that is not guaranteed and easily over looked.

Theorem 16. *The high-order positivity–preserving flux-limiting scheme given by Eq.* (178) *is freestream preserving.*

Proof. For freestream preservation, we assume a globally constant state (including boundary conditions) and want to show that this implies $\frac{d\hat{\mathbf{U}}}{dt} = \mathbf{0}_5$.

Notice that all viscous terms in this dissertation–including artificial dissipation terms–depend directly on two-point jumps in states or high-order computational derivatives of states on an element. Hence, all viscous terms preserve freestream.

Thus, only the inviscid terms remain. The inviscid penalty given by Eq. (70) is also zero since the single state and two state fluxes are equivalent at the interface. Finally, $\hat{\tilde{\mathbf{f}}}_l$ has been proven to be freestream preserving [36, 52] and Lemma 2 proved that $\hat{\tilde{\mathbf{f}}}_l^{(EC)}$ is also freestream preserving.

□

6.3.8 STABILITY PROPERTIES

The high-order positivity-preserving flux-limiting scheme given by Eq. (178) admits both L_1 and L_2 stability statements which we discuss in this section.

L_1 stability

Notice that having a conservative scheme implies that at the nth time level

$$\begin{aligned}\sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\boldsymbol{\rho}}_k^n &= \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\boldsymbol{\rho}}_k^0 + \sum_{i=0}^{n-1} B_\rho^i, \\ \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \widehat{\mathbf{Et}}_k^n &= \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \widehat{\mathbf{Et}}_k^0 + \sum_{i=0}^{n-1} B_{Et}^i,\end{aligned} \tag{185}$$

where we have summed over the K elements in the domain, $\hat{\boldsymbol{\rho}}_k^n$ and $\widehat{\mathbf{Et}}_k^n$ are the arrays of density and total energy (scaled by the Jacobian) on the kth element at time level n and the B^i terms represent the effect from the boundaries. In particular, if all boundary faces are periodic we have $B^i = 0$ for all i. If we also have point-wise positivity, we get the following theorem.

Theorem 17. *Assume that the initial numerical solution at every solution point in the domain has positive density and temperature. Furthermore, assume for all $n \in \mathbb{N}$ we have*

$$c_{\min}^{\rho} \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\boldsymbol{\rho}}_k^0 \leq \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\boldsymbol{\rho}}_k^0 + \sum_{i=0}^{n-1} B_\rho^i \leq c_{\max}^{\rho} \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\boldsymbol{\rho}}_k^0,$$

$$c_{\min}^{Et} \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \widehat{\mathbf{Et}}_k^0 \leq \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \widehat{\mathbf{Et}}_k^0 + \sum_{i=0}^{n-1} B_{Et}^i \leq c_{\max}^{Et} \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \widehat{\mathbf{Et}}_k^0,$$

for fixed positive constants $0 < c_{\min}^{Et} \leq c_{\max}^{Et}$ and $0 < c_{\min}^{\rho} \leq c_{\max}^{\rho}$. Then, the high-order positivity-preserving flux-limiting scheme given by Eq. (178) *admits the following discrete L_1*

bounds on the solution at the nth time level

$$\begin{aligned}
c_{\min}^{\rho}\sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\hat{\boldsymbol{\rho}}_k^0 \leq &\quad \sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\,|\hat{\boldsymbol{\rho}}_k^n| &&\leq c_{\max}^{\rho}\sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\hat{\boldsymbol{\rho}}_k^0,\\
c_{\min}^{Et}\sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\widehat{\mathbf{Et}}_k^0 \leq &\quad \sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\left|\widehat{\mathbf{Et}}_k^n\right| &&\leq c_{\max}^{Et}\sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\widehat{\mathbf{Et}}_k^0,\\
c_{\min}^{Et}\sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\widehat{\mathbf{Et}}_k^0 \leq &\quad \sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\left|\widehat{\mathbf{IE}}_k^n\right| &&\leq c_{\max}^{Et}\sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\widehat{\mathbf{Et}}_k^0,\\
c_{\min}^{Et}\sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\widehat{\mathbf{Et}}_k^0 \leq &\quad \sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\left|\widehat{\mathbf{KE}}_k^n\right| &&\leq c_{\max}^{Et}\sum_{k=1}^{K}\mathbf{1}_1^\top\widehat{\mathcal{P}}\widehat{\mathbf{Et}}_k^0,
\end{aligned} \tag{186}$$

where $|\hat{\boldsymbol{\rho}}_k^n|$, $\left|\widehat{\mathbf{Et}}_k^n\right|$, $\left|\widehat{\mathbf{IE}}_k^n\right|$, *and* $\left|\widehat{\mathbf{KE}}_k^n\right|$ *are the arrays of the absolute values of the discrete density, total energy, internal energy, and kinetic energy on the kth element at time level n.*

Proof. Pointwise positivity implies that $|\hat{\boldsymbol{\rho}}_k^n| = \hat{\boldsymbol{\rho}}_k^n$, $\left|\widehat{\mathbf{Et}}_k^n\right| = \widehat{\mathbf{Et}}_k^n$, $\left|\widehat{\mathbf{IE}}_k^n\right| = \widehat{\mathbf{IE}}_k^n$, and $\left|\widehat{\mathbf{KE}}_k^n\right| = \widehat{\mathbf{KE}}_k^n$ on every element. Hence, the bounds for density and total energy are an immediate consequence of Eq. (185). Furthermore, the positivity of internal energy and kinetic energy at every solution point in the domain implies the remaining bounds since at every solution point the internal energy and kinetic energy are each bounded from above by the total energy at that solution point. □

L_2 stability

We now discuss the discrete form of the L_2 bound on the conservatives variables presented in Section 2.3.2. We do not formulate the following as a theorem about the high-order positivity-preserving flux-limiting scheme given by Eq. (178) because we did not use the relaxation methods in [65, 66] to strictly enforce the condition in Eq. (191). This was mostly a choice made out of practical considerations concerning time and code complexity. Furthermore, from our numerical experiments on the test cases we considered, we have observed that our scheme produces sufficiently non-oscillatory solutions and a temporal evolution of the total entropy that is monotonically decreasing when plotted for problems with entropy stable boundary conditions (with possible increases between time steps on the order of the discretization error). The following discussion is to show how for a scheme that guarantees pointwise-positivity such as the one given by Eq. (178), the remaining steps for a

fully discrete L_2 bound on the conservatives variables are entropy stable boundary conditions and strictly enforcing the condition in Eq. (191).

Again, we define a new convex entropy $\bar{S} = S - S(\boldsymbol{u}_0) - S_{\boldsymbol{U}}(\boldsymbol{u}_0)^\top(\boldsymbol{u} - \boldsymbol{u}_0)$ where $\boldsymbol{u}_0$ is a constant non-zero state with zero velocity and the associated entropy variables

$$\bar{\boldsymbol{w}} \equiv \bar{S}_{\boldsymbol{U}} = S_{\boldsymbol{U}} - S_{\boldsymbol{U}}(\boldsymbol{u}_0) = \boldsymbol{w} - \boldsymbol{w}_0. \tag{187}$$

From this new entropy, we form the 2-D array $\bar{\mathbf{w}}_k = \mathbf{w}_k - \mathbf{w}_0$ on the kth element where $\mathbf{w}_0(\vec{\xi}_{abc}) = \boldsymbol{w}_0$ for every solution point on every element. Contracting Eq. (178) with the new entropy variables given by Eq. (187) yields

$$\sum_{k=1}^{K} \bar{\mathbf{w}}_k^\top \mathcal{P} \frac{d\hat{\mathbf{U}}_k}{dt} = \sum_{k=1}^{K} \mathbf{w}_k^\top \mathcal{P} \frac{d\hat{\mathbf{U}}_k}{dt} - \mathbf{w}_0^\top \mathcal{P} \frac{d\hat{\mathbf{U}}_k}{dt}. \tag{188}$$

From Theorem 15, we know that

$$\sum_{k=1}^{K} \mathbf{w}_k^\top \mathcal{P} \frac{d\hat{\mathbf{U}}_k}{dt} = B - D,$$

where B contains terms related to the domain boundary and $D \geq 0$. Assume that we have entropy stable boundary conditions so that

$$\sum_{k=1}^{K} \mathbf{w}_k^\top \mathcal{P} \frac{d\hat{\mathbf{U}}_k}{dt} \leq -D.$$

Notice that we have

$$\sum_{k=1}^{K} \mathbf{w}_0^\top \mathcal{P} \frac{d\hat{\mathbf{U}}_k}{dt} = \boldsymbol{w}_0 \sum_{k=1}^{K} \mathbf{1}_1^\top \mathcal{P} \frac{d\hat{\mathbf{U}}_k}{dt} \tag{189}$$

and since the scheme given by Eq. (178) is conservative (see Theorem 14), we know that $\sum_{k=1}^{K} \mathbf{1}_1^\top \mathcal{P} \frac{d\hat{\mathbf{U}}_k}{dt}$ depends only on the boundary. Again, we assume that we have boundary

conditions so that

$$\sum_{k=1}^{K} \bar{\mathbf{w}}_k^\top \mathcal{P} \frac{d\hat{\mathbf{U}}_k}{dt} \leq \sum_{k=1}^{K} \mathbf{w}_k^\top \mathcal{P} \frac{d\hat{\mathbf{U}}_k}{dt} \leq -D. \tag{190}$$

For example, periodic boundary conditions would guarantee this.

We now introduce the superscript "n" to denote the time level. To obtain the fully discrete L_2 bound we need the semi-discrete statement of Eq. (190) to imply that

$$\sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\bar{\mathbf{S}}}_k^{n+1} \leq \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\bar{\mathbf{S}}}_k^{n} - \tau^n D^n, \tag{191}$$

where $D^n \geq 0$ represents the cumulative entropy dissipation over the time step (e.g. over all the Runge–Kutta stages) and $\tau^n > 0$ is the time step used to advance the solution from the nth to the $(n+1)$th time level. Obtaining an inequality like this is dependent on the time discretization used. However, since $\bar{S}$ is convex, one can apply the relaxation methods in [65, 66] to enforce this condition for Runge-Kutta methods (explicit or implicit) or multistep methods. Thus, we assume that Eq. (191) holds discretely. Hence, it follows that at the nth time level we have

$$\sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\bar{\mathbf{S}}}_k^{n} \leq \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\bar{\mathbf{S}}}_k^{1} - \sum_{i=1}^{n-1} \tau^i D^i. \tag{192}$$

Taylor expansion of S around $\boldsymbol{u}_0$ implies

$$\begin{aligned} \mathbf{S}_k^n(\vec{\xi}_a) = S(\mathbf{U}_k^n(\vec{\xi}_a)) = S(\boldsymbol{u}_0) + S_{\boldsymbol{U}}(\boldsymbol{u}_0)^\top \left(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0\right) \\ + \frac{1}{2}\left(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0\right)^\top S_{\boldsymbol{U}\boldsymbol{U}}(\tilde{\mathbf{U}}_k^n(\theta(\vec{\xi}_a)))\left(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0\right), \end{aligned} \tag{193}$$

where the state $\tilde{\mathbf{U}}_k^n(\theta(\vec{\xi}_a))$ has positive density and temperature since $\boldsymbol{u}_0$ and $\mathbf{U}_k^n(\vec{\xi}_a)$ both do. Notice that by definition $\bar{\mathbf{S}}_k^n(\vec{\xi}_a) = \mathbf{S}_k^n(\vec{\xi}_a) - S(\boldsymbol{u}_0) - S_{\boldsymbol{U}}(\boldsymbol{u}_0)^\top(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0) = \frac{1}{2}\left(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0\right)^\top S_{\boldsymbol{U}\boldsymbol{U}}(\tilde{\mathbf{U}}_k^n(\theta(\vec{\xi}_a)))\left(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0\right)$. Let $S_{\boldsymbol{U}\boldsymbol{U}}^{\min,n}$ be the minimum eigenvalue

of all terms $S_{\boldsymbol{U}\boldsymbol{U}}(\tilde{\mathbf{U}}_k^n(\theta(\vec{\xi}_a)))$ in the domain. Then, if we set

$$\mathcal{C}^{n-1} = \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\tilde{\mathbf{S}}}_k^1 - \sum_{i=1}^{n-1} \tau^i D^i \leq \sum_{k=1}^{K} \mathbf{1}_1^\top \widehat{\mathcal{P}} \hat{\tilde{\mathbf{S}}}_k^1,$$

we have

$$2S_{\boldsymbol{U}\boldsymbol{U}}^{\min,n} \sum_{k=1}^{K} \left(\mathbf{U}_k^n - \mathbf{U}_0\right)^\top \mathcal{P} \left[J\right]_k \left(\mathbf{U}_k^n - \mathbf{U}_0\right) \leq 4\mathcal{C}^{n-1} \tag{194}$$

and by pointwise application of Eq. (38) we have

$$\sum_{k=1}^{K} \left(\mathbf{U}_k^n\right)^\top \mathcal{P} \left[J\right]_k \mathbf{U}_k^n \leq 4\frac{\mathcal{C}^{n-1}}{S_{\boldsymbol{U}\boldsymbol{U}}^{\min,n}} + 2\sum_{k=1}^{K} \mathbf{U}_0^\top \mathcal{P} \left[J\right]_k \mathbf{U}_0, \tag{195}$$

which is the fully-discrete analogue of the L_2 bound on the solution given by Eq. (38).

For the fully-discrete analogue of the L_2 bound on the solution given by Eq. (41) which is based on the bounds obtained from the $LDL^\top$ decomposition of $\mathcal{S}_{\boldsymbol{U}\boldsymbol{U}}$, we note that it follows from (254) and (255) of Appendix C that

$$\begin{aligned}
&\left(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0\right)^\top S_{\boldsymbol{U}\boldsymbol{U}}(\tilde{\mathbf{U}}_k^n(\theta(\vec{\xi}_a))) \left(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0\right) \geq \frac{(\boldsymbol{\rho}_k^n(\vec{\xi}_a) - \rho_0)^2}{b_1(\tilde{\mathbf{U}}_k^n(\theta(\vec{\xi}_a)))}, \\
&\left(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0\right)^\top S_{\boldsymbol{U}\boldsymbol{U}}(\tilde{\mathbf{U}}_k^n(\theta(\vec{\xi}_a))) \left(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0\right) \geq \frac{((\boldsymbol{m}_i)_k^n(\vec{\xi}_a))^2}{b_{i+1}(\tilde{\mathbf{U}}_k^n(\theta(\vec{\xi}_a)))}, \quad i = 1,2,3, \\
&\left(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0\right)^\top S_{\boldsymbol{U}\boldsymbol{U}}(\tilde{\mathbf{U}}_k^n(\theta(\vec{\xi}_a))) \left(\mathbf{U}_k^n(\vec{\xi}_a) - \boldsymbol{u}_0\right) \geq \frac{(\mathbf{Et}_k^n(\vec{\xi}_a) - Et_0)^2}{b_5(\tilde{\mathbf{U}}_k^n(\theta(\vec{\xi}_a)))}, \\
&b_1(\boldsymbol{u}_a) = \frac{\rho_a}{R}, \quad b_{i+1}(\boldsymbol{u}_a) = \frac{P_a + \rho_a(V_i^2)_a}{R}, \quad i = 1,2,3\ , \\
&b_5(\boldsymbol{u}_a) = \frac{P_a^2\gamma + P_a\rho_a\|\boldsymbol{V}_a\|^2\gamma + \left(\rho_a\frac{\|\boldsymbol{V}_a\|^2}{2}\right)^2}{R\rho_a},
\end{aligned} \tag{196}$$

where $(\boldsymbol{m}_i)_k^n(\vec{\xi}_a) = \boldsymbol{\rho}_k^n(\vec{\xi}_a)(\boldsymbol{V}_i)_k^n(\vec{\xi}_a)$ represents the ith component of momentum and $\mathbf{Et}_k^n(\vec{\xi}_a) = \boldsymbol{\rho}_k^n(\vec{\xi}_a)\mathbf{E}_k^n(\vec{\xi}_a)$ is the total energy. Let $b_i^{\max,n} = \max\limits_{1\leq k\leq K} \max\limits_{1\leq a\leq N_p} b_i(\tilde{\mathbf{U}}_k^n(\theta(\vec{\xi}_a)))$. Then

we have

$$\begin{aligned}
\sum_{k=1}^{K} \left(\boldsymbol{\rho}_k^n\right)^\top \widehat{\mathcal{P}} \hat{\boldsymbol{\rho}}_k^n &\leq 4 b_1^{\max,n} \mathcal{C}^{n-1} + 2 \sum_{k=1}^{K} \boldsymbol{\rho}_0^\top \widehat{\mathcal{P}} \hat{\boldsymbol{\rho}}_0, \\
\sum_{k=1}^{K} \left(\left(\boldsymbol{m}_i\right)_k^n\right)^\top \widehat{\mathcal{P}} \left(\hat{\boldsymbol{m}}_i\right)_k^n &\leq 2 b_{i+1}^{\max,n} \mathcal{C}^{n-1}, \qquad i = 1, 2, 3, \\
\sum_{k=1}^{K} \left(\mathbf{Et}_k^n\right)^\top \widehat{\mathcal{P}} \widehat{\mathbf{Et}}_k^n &\leq 4 b_5^{\max,n} \mathcal{C}^{n-1} + 2 \sum_{k=1}^{K} \mathbf{Et}_0^\top \widehat{\mathcal{P}} \widehat{\mathbf{Et}}_0,
\end{aligned} \tag{197}$$

which is the fully-discrete analogue of the L_2 bound on the solution given by Eq. (41).

6.4 IMPLEMENTATION DETAILS

In this section, we discuss implementation details of the positivity preserving entropy stable flux-limiting scheme given by Eq. (178).

6.4.1 FLUX LIMITER DETAILS

The flux limiter, θ_f, in Eq. (178) becomes less than one only on troubled elements for which at least one of the thermodynamic variables at any solution point is negative or the residual-based sensor given by Eq. (85) is nonzero and the two-point relative jump in pressure exceeds its threshold value which is equivalent to the pressure jump across a Mach 1.75 shock. For each troubled element, the limiter θ_f^k in Eq. (178) is determined such that it satisfies the positivity constraints described in Section 6.3.1 and the following inequality: $\theta_f^k \leq 1 - Sn^k \max_k \left(\frac{|\Delta P|}{2P_A}\right)$, where $0 \leq Sn^k \leq 1$ is the residual-based sensor given by Eq. (85) on the kth element and $0 \leq \max_k \left(\frac{|\Delta P|}{2P_A}\right) \leq 1$ is one half of the maximum relative two–point pressure jump (including jumps at the interfaces) on the same element.

6.4.2 TIME STEP DETAILS

The time step constraint required for pointwise positivity of density is very similar to the conventional CFL condition, as discussed in Section 6.1.5. As mentioned already, we use the slightly stricter time step constraint of Eq. (126) instead of the one given by Theorem 6. Applying Corollary 6.1, we see that the time step constraint of Eq. (126) at the $\vec{\xi}_{ijk}$ solution

point on the element is given by:

$$\tau_{ijk} < \frac{J_{ijk}}{12} \min_{l} \left(\frac{P_{ll}}{\max\left[\lambda_c + \frac{\sigma \|\hat{\vec{\mathbf{a}}}\|^2}{J_G \Delta\xi}\right]_l^{+/-}} \right)_{\vec{\xi}_{ijk}} = (\tau_\rho^I)_{ijk}. \tag{198}$$

For viscous flows, we augment the time step constraint of Eq. (198) to satisfy the linear stability condition for diffusion as follows:

$$\tau_{ijk} \leq \frac{\text{CFL}}{\frac{1}{(\tau_\rho^I)_{ijk}} + \left(\frac{\mu}{J^2 \mathcal{P}_{ijk}^{2/3}} \sum_{b=1}^{3} \|\hat{\vec{\mathbf{a}}}^b\|^2 \right)_{\vec{\xi}_{ijk}}}. \tag{199}$$

For all test problems considered, we set CFL = 1. Furthermore, the safety parameter c_{IE}, which is used to determine the time step constraint required for positivity of internal energy in Section 6.1.6, is set equal to 0.9.

6.4.3 ALGORITHM AND POSITIVITY

In Sections 6.1 and 6.3, we have proven that the baseline first-order scheme and the corresponding high-order flux-limiting scheme preserve the positivity of density and internal energy under suitable time step constraints when the explicit forward Euler method is used to discretize the time derivative terms. To generalize the proposed positivity-preserving methods to high-order temporal discretizations, we use the third-order strong stability preserving (SSP) Runge-Kutta scheme developed in [67], which can be represented as a convex combination of forward Euler schemes. At each Runge-Kutta stage, the high-order positivity-preserving entropy stable scheme is implemented according to the following algorithm.

Algorithm (Iterative positivity-preserving explicit SSP Runge-Kuta spectral collocation method)

1. Compute $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_p$ using Eq. (64).
2. Compute Sn^k for all k as described in Chapter 5.
3. For those elements where $Sn^k > 0$, we compute the artificial viscosity, $\boldsymbol{\mu}^{AD}$, defined in

Chapter 5 and maximum relative pressure jump defined in Section 6.3.1.

4. At the first Runge-Kutta stage, compute the time step, τ^n, given by Eq. (199). Furthermore, we require that $\tau^n \leq 1.01\tau^{n-1}$, where τ^{n-1} is the previous time step.

5. For a given element, if $Sn^k = 1$ and $\max_k \left(\frac{|\Delta P|}{2P_A}\right) > threshold$ then obtain $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_*$ using $\frac{d\hat{\mathbf{U}}}{dt}$ from Eq. (178) with $\theta_f^k = 0$ and only first-order artificial dissipation. Otherwise, obtain $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_*$ using $\theta_f^k = 1$ and only high-order artificial dissipation.

6. If $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_*$ does not preserve positivity and $\theta_f^k = 1$, then obtain $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_*$ using $\frac{d\hat{\mathbf{U}}}{dt}$ from Eq. (178) with $\theta_f^k = 0$ and only first-order artificial dissipation.

7. If on the second or later Runge-Kutta stage, $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_*$ does not preserve positivity and $\theta_f^k = 0$, then set $\tau_{\text{new}}^n = 0.5\tau^n$ and restart from the first Runge-Kutta stage. For viscous flows, if only temperature positivity was violated and Eq. (163) holds then apply one iteration of the velocity and temperature limiting procedure described in Section 6.2 on that element at the beginning of every Runge-Kutta stage until the whole time step is complete.

8. For the first Runge-Kutta stage, use the current $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_*$ to compute the time-step constraint required for positivity of internal energy, which is given by Eq. (136) with $C_{\text{IE}} = 0.9$. For viscous flows, if this time step constraint is stricter than the current global time step for some element and Eq. (163) holds for that element, then apply one iteration of the velocity and temperature limiting procedure described in Section 6.2 on that element at the beginning of every Runge-Kutta stage until the whole time step is complete.

9. For elements where $\left(\frac{d\hat{\mathbf{U}}}{dt}\right)_*$ uses $\theta_f^k = 0$, adjust θ_f^k as described in Section 6.3 to obtain $\frac{d\hat{\mathbf{U}}}{dt}$.

10. Advance to the next Runge-Kutta stage.

By construction, the proposed scheme guarantees positivity of density and temperature at the first Runge-Kutta stage. For subsequent Runge-Kutta stages, the new scheme, which can

be represented as forward Euler steps, preserves the positivity of thermodynamic variables, if the time step chosen at the first stage satisfies the time-step positivity constraint at the remaining stages. If the scheme fails to preserve positivity on a later Runge-Kutta stage, one can update the time step that meets the positivity constraint and repeat iterations until the positivity constraint is met for all stages. Note, however, that for all test problems presented in this dissertation, failing positivity on a later Runge-Kutta stage was extremely rare and never required restarting the time step more than once. This potential issue can be avoided by using an SSP multi-time step discretization as discussed in [64], but the above method has been chosen herein because of its simplicity.

CHAPTER 7

NUMERICAL RESULTS

We test the proposed positivity-preserving high-order limiting scheme on standard benchmark problems with smooth and discontinuous solutions. In all numerical experiments presented herein, the 3rd-order strong stability preserving (SSP) Runge-Kutta scheme developed in [67] is used to advance the semi-discretization in time. Note that this scheme violates the entropy stability property of the semi-discrete operator by a factor proportional to the local temporal truncation error. As discussed in Section 6.4.3, the time step in our numerical experiments is selected by using the Courant-Friedrich-Levy (CFL) condition given by Eq. (199) and the density and temperature positivity constraints presented in Section 6.1.

In Section 7.2, we present 1-D results that we obtained using the method in [12]. In Section 7.3, we present the results of 2-D and 3-D simulations using the proposed method in this dissertation.

We use the following acronyms for the numerical schemes presented in this dissertation:

- **ESSC-pW** Solutions obtained using only the scheme of Eq. (64) with polynomial order W will be denoted ESSC-pW ("Entropy Stable Spectral Collocation").

- **PPESAD-pW** Solutions obtained using the proposed scheme of this dissertation (Eq. (178)) with polynomial order W will be denoted PPESAD-pW ("Positivity Preserving Entropy Stable Artificial Dissipation").

- **PPES-pW** Solutions obtained using the proposed scheme of this dissertation (Eq. (178)) with $\boldsymbol{\mu}^{AD}$ artificially set to zero and polynomial order W will be denoted PPES-pW ("Positivity Preserving Entropy Stable").

We use the PPES-pW scheme to see the effects of the artificial dissipation introduced through $\boldsymbol{\mu}^{AD}$, while maintaining positivity for a simulation where ESSC-pW fails to maintain positivity.

7.1 NON-DIMENSIONAL 3-D COMPRESSIBLE NAVIER-STOKES EQUATIONS

The numerical results presented in this chapter were obtained by simulating a non-dimensional form of Eq. (45). The non-dimensional equations are obtained by introducing a characteristic length L, velocity V_∞, time L/V_∞, density ρ_∞, temperature T_∞, dynamic viscosity μ_∞, thermal conductivity κ_∞, and artificial viscosity $L\rho_\infty V_\infty$. The dimensionless variables with the subscript asterisk are given by

$$\begin{aligned} &x^i = x^i_* L,\ V_i = (V_i)_* V_\infty,\ t = t_* L/V_\infty,\ \rho = \rho_* \rho_\infty,\ T = T_* T_\infty,\ \mu = \mu_* \mu_\infty, \\ &\kappa = \kappa_* \kappa_\infty,\ \mu^{AD} = \mu^{AD}_* L\rho_\infty V_\infty. \end{aligned} \tag{200}$$

The characteristic Mach number is defined as $Ma \equiv V_\infty/\sqrt{\gamma R T_\infty}$. The non-dimensional form of Eq. (45) that we used, can be obtained by multiplying the first equation of Eq. (45) by $L/(\rho_\infty V_\infty)$, the momentum equations by $L/(\rho_\infty V_\infty^2)$, and the fifth equation by $L/(\rho_\infty V_\infty T_\infty c_P)$. Hence, the non-dimensional form Eq. (45) can be written as (for the remainder of this chapter, we omit the asterisks for clarity)

$$\begin{aligned} &\frac{\partial \boldsymbol{U}}{\partial t} + \sum_{m=1}^{3} \frac{\partial \boldsymbol{F}_{x_m}}{\partial x_m} = \sum_{m=1}^{3} \left[\frac{\partial \boldsymbol{F}^{(v)}_{x_m}}{\partial x_m} + \frac{\partial \boldsymbol{F}^{(AD)}_{x_m}}{\partial x_m} \right], \\ &\boldsymbol{U} = \left[\begin{array}{cccccc} \rho & \rho V_1 & \rho V_2 & \rho V_3 & \rho E \end{array} \right]^\top, \\ &\boldsymbol{F}_{x_m} = \left[\begin{array}{cccc} \rho V_m & \rho V_m V_1 + \delta_{m,1} P & \rho V_m V_2 + \delta_{m,2} P & \rho V_m V_3 + \delta_{m,3} P & \rho V_m H \end{array} \right]^\top, \\ &\boldsymbol{F}^{(v)}_{x_m} = \frac{1}{Re} \boldsymbol{F}^{(visc)}_{x_m}(\mu, \frac{\kappa}{Pr}), \\ &\boldsymbol{F}^{(AD)}_{x_m} = \mu^{AD} \left(\boldsymbol{F}^{(visc)}_{x_m}(1, \frac{c_T}{c_P}) + \frac{c_\rho}{\rho} \frac{\partial \rho}{\partial x_m} \left[\begin{array}{ccc} 1 & \vec{\boldsymbol{V}} & E \end{array} \right]^\top \right), \\ &\boldsymbol{F}^{(visc)}_{x_m}(a,b) = \left[\begin{array}{ccccc} 0 & \tau_{1,m}(a) & \tau_{2,m}(a) & \tau_{3,m}(a) & \sum_{i=1}^{3} \tau_{i,m}(a) V_i (\gamma-1) Ma^2 + b \frac{\partial T}{\partial x_m} \end{array} \right]^\top, \\ &\tau_{i,j}(a) = a \left(\frac{\partial V_i}{\partial x_j} + \frac{\partial V_j}{\partial x_i} - \delta_{i,j} \frac{2}{3} \sum_{n=1}^{3} \frac{\partial V_n}{\partial x_n} \right), \\ &P = \frac{\rho T}{\gamma Ma^2}, \quad H = T + \frac{(\gamma-1)Ma^2}{2} \| \vec{\boldsymbol{V}} \|^2, \quad E = \frac{T}{\gamma} + \frac{(\gamma-1)Ma^2}{2} \| \vec{\boldsymbol{V}} \|^2, \end{aligned} \tag{201}$$

where we see that Eq. (201) contains the dimensionless parameters: $Re \equiv L\rho_\infty V_\infty/\mu_\infty$

the Reynolds number, $Pr \equiv \mu_\infty c_P/\kappa_\infty$ the Prandtl number, $Ma \equiv V_\infty/\sqrt{\gamma RT_\infty}$ the Mach number, and the adiabatic exponent of gas γ. For all numerical simulations, we used $\gamma = 1.4$.

7.2 1-D NUMERICAL RESULTS

We now present some of the 1-D numerical results that we obtained in [12]. The approach in [12] differed from that discussed here in several ways involving mostly the artificial viscosity and dissipation. Most significantly, in [12] we did not use the Mach number, compression sensor, or pressure sensor when constructing $\mu^k_{\max}$ for the artificial viscosity (see Eq. (97)). Furthermore, the parameters c_ρ and c_T were set equal to 0.25 and $0.25\frac{c_P}{\gamma}$, respectively. Despite these differences, the main components of the algorithm are the same. For the viscous flows in this section, we used a Prandtl number of $Pr = 0.75$ and we did not use Sutherland's law. For all results in this section we used $Ma = 1/\sqrt{\gamma}$ so that $P = \frac{\rho T}{\gamma Ma^2} = \rho T$.

7.2.1 BLAST WAVE

To demonstrate the performance of the new high-order positivity-preserving entropy stable scheme for flows with very strong shocks and contact discontinuities, we solve the inviscid and viscous blast wave flows with the initial conditions proposed by Woodward and Colella [68]. For both the 1-D Euler and Navier-Stokes equations, the initial conditions are as follows:

$$(\rho, v, P) = \begin{cases} (1, 0, 1000), & \text{for } -0.5 \le x < -0.4 \\ (1, 0, 0.01), & \text{for } -0.4 \le x \le 0.4 \\ (1, 0, 100), & \text{for } \quad 0.4 \le x \le 0.5, \end{cases}$$

and the reflection boundary conditions are imposed on both the left and right boundaries. The final time is set equal to $t = 0.038$. This test problem is characterized by the presence of very large pressure and density jumps that may lead to discrete solutions with negative densities and temperatures if the high-order scheme alone is used for capturing these strong shock waves and contact discontinuities. Density and pressure profiles computed with the new high-order ($p = 6$) positivity-preserving spectral collocation scheme on 64, 128, 256–element grids and the reference solution obtained with the third-order finite difference ESWENO scheme developed in [69] on a uniform grid with $4,000$ cells for the inviscid blast wave flow

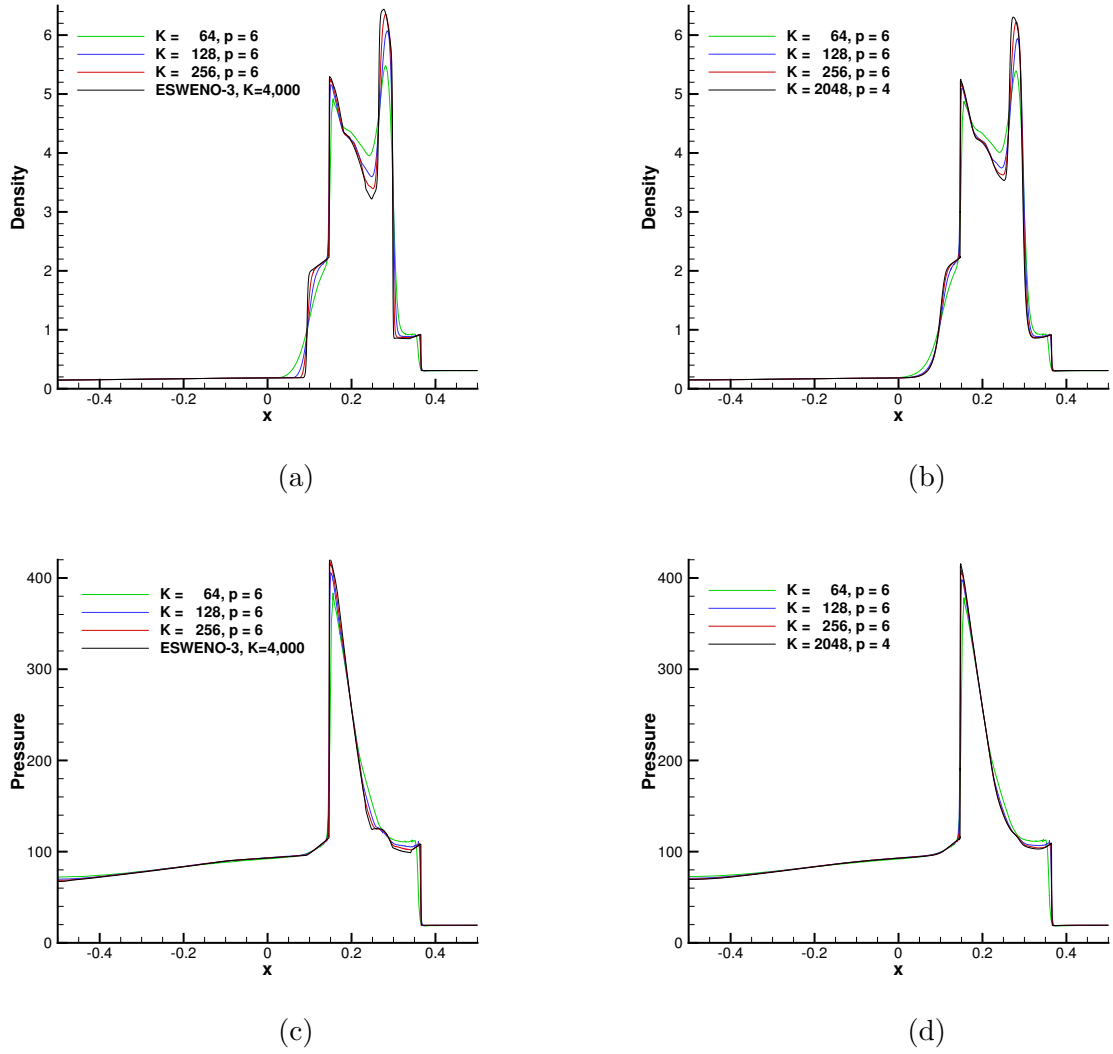
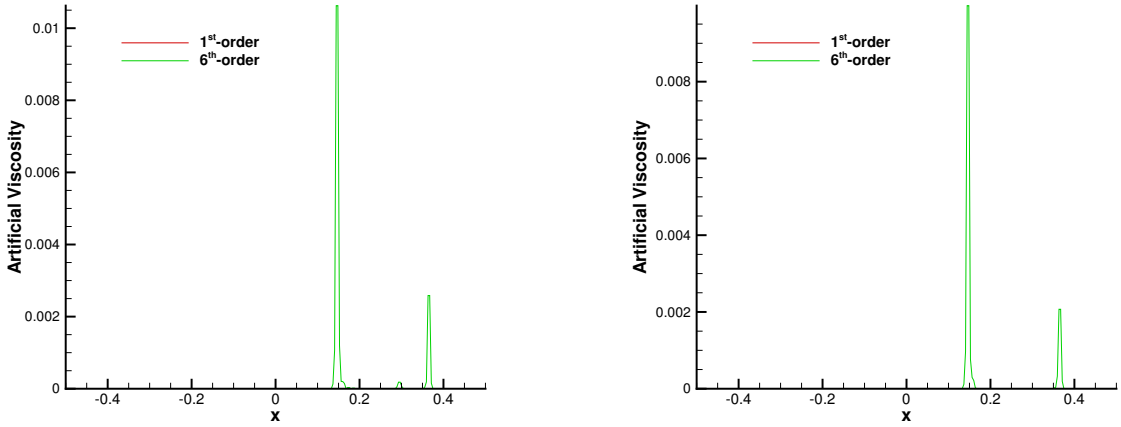


Fig. 1: Density (first row) and pressure (second row) profiles computed with the PPESAD-p6 scheme on uniform grids with 64, 128, 256 elements for the inviscid (left column) and viscous (right column) blast wave flows at $t = 0.038$.

are presented in the left panel of Fig. 1. Along with the conventional inviscid blast wave flow, we also consider the corresponding viscous counterpart with the Reynolds number of 10^3, which is solved by using the same high-order positivity-preserving flux-limiting scheme. The right panel of Figure 1 shows the density and pressure profiles computed with new high-order ($p = 6$) limiting scheme on the same grids used for the inviscid flow. These results are compared with the reference solution obtained using the forth-order ($p = 4$)

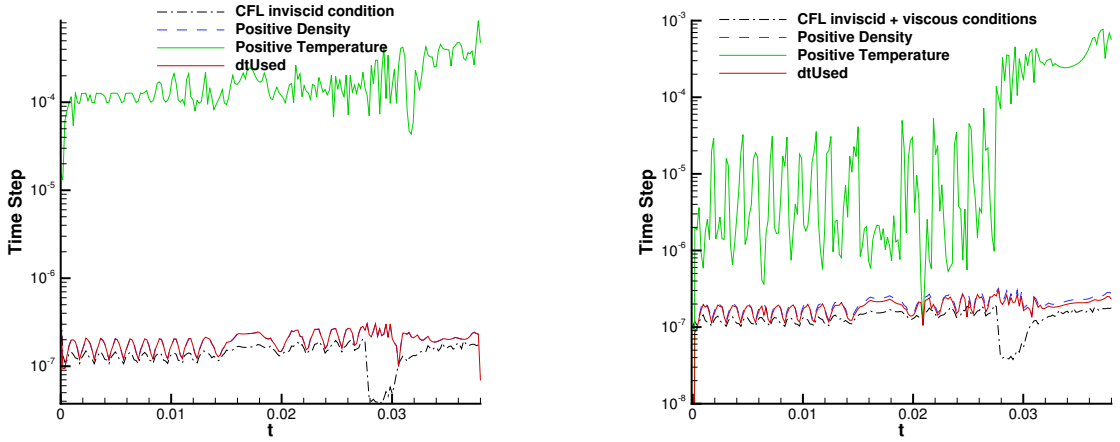


(a) (b)

Fig. 2: The low- and high-order ($p = 6$) artificial viscosities obtained on the 256-element grid for the inviscid (left panel) and viscous blast wave flows at $t = 0.038$.

positivity-preserving scheme on a very fine mesh with 2048 grid elements. As follows from these numerical results, the new high-order ($p = 6$) spectral collocation limiting scheme provides not only positivity of the thermodynamic variables, but also excellent dissipation properties that allow us to capture the strong shocks within one element for both the viscous and inviscid flows on all grids considered. Numerical solutions obtained with other polynomial bases ($p = 4, 5$) demonstrate similar discontinuity-capturing properties and therefore are not presented herein. For both the inviscid and viscous blast wave flows, the 1st-order scheme is used only at the beginning of computation and only in elements containing the strong shocks. The high-order artificial dissipation is used during the entire time interval considered and is nonzero only at the strong shocks, as one can see in Figure 2. Another attractive feature of the new entropy-based artificial dissipation method is that practically no dissipation is added at the contact discontinuity.

The time step histories obtained using the 1-D form of the density and temperature positivity constraints given by Eqs. (126) and (134) as well as the conventional CFL condition (formed like Eq. (199) but $\frac{1}{(\tau_\rho^I)_{ijk}}$ is replaced by a term proportional to the pointwise maximum eigenvalue) with the CFL number set equal to 0.5 on the 256–element grid are shown in Figure 3 for the inviscid (left panel) and viscous flows. For the inviscid blast wave flow, the



(a) (b)

Fig. 3: Time step histories for the inviscid (left panel) and viscous ($Re = 10^3$) blast wave flows computed with the PPESAD-p6 scheme on 256-element uniform grid.

time step required for positivity of temperature (Eq. (134)) is 2–3 orders of magnitude higher than that imposed by the density positivity and conventional CFL conditions, as one can see in Figure 3. Note, however, that this behavior is qualitatively different for viscous flows. As evident from Figure 3 (right panel), the time step required for positivity of temperature for the Reynolds number $Re = 10^3$ is orders of magnitude smaller at the beginning of the simulation. This stiffness of the time step constraint is caused by the presence of large solution gradients at the initial instant in time. If not limited, the velocity and temperature gradients are extremely large near strong discontinuities, thus making the temperature positivity time constraint (Eq. (134)) very stiff. The new velocity and temperature gradient limiters presented in Section 6.2 weaken this constraint and allow the artificial dissipation method to capture these strong shock waves and contact discontinuities without producing negative densities and temperatures, while not imposing severe constraints on the time step during the entire time interval considered as seen in Figure 3.

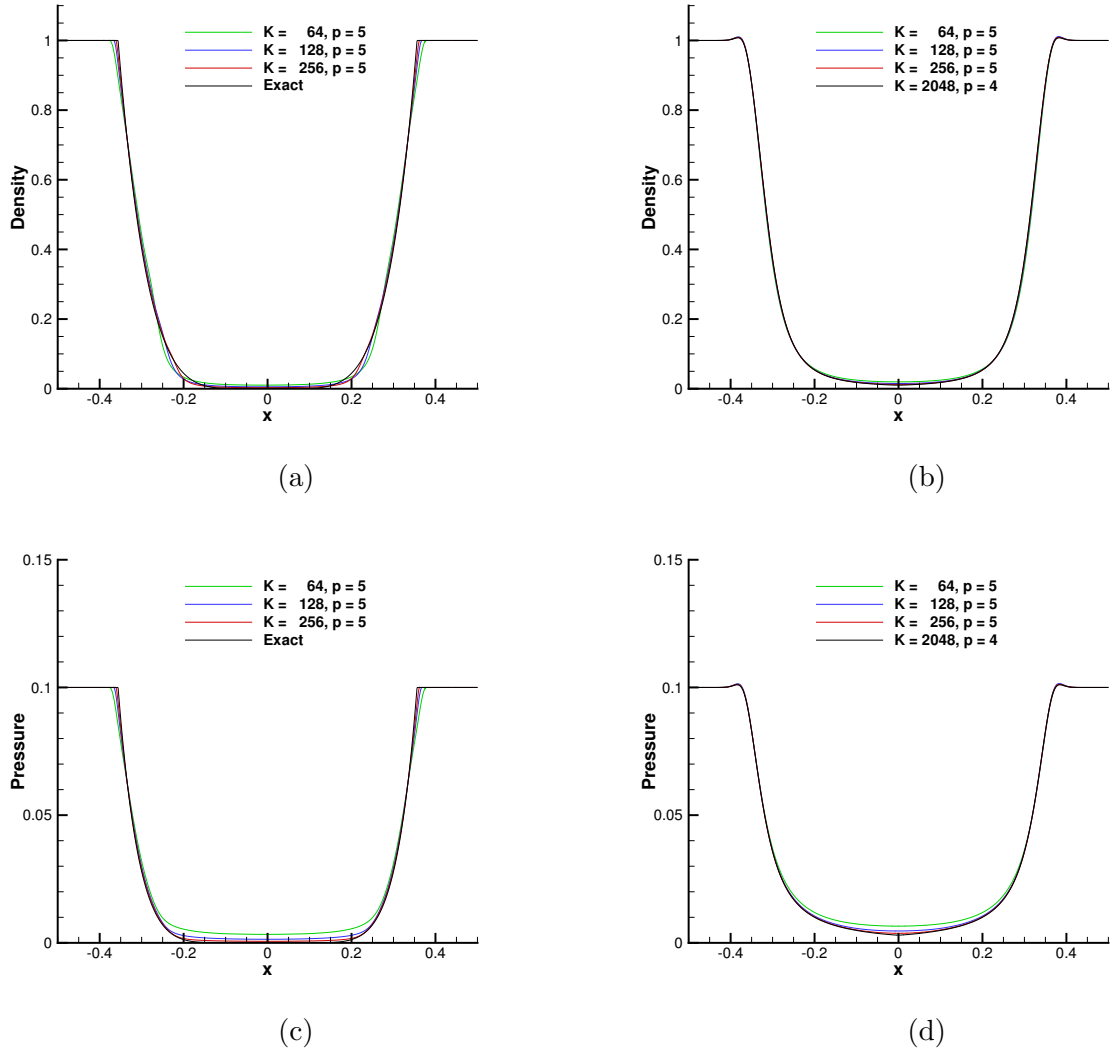
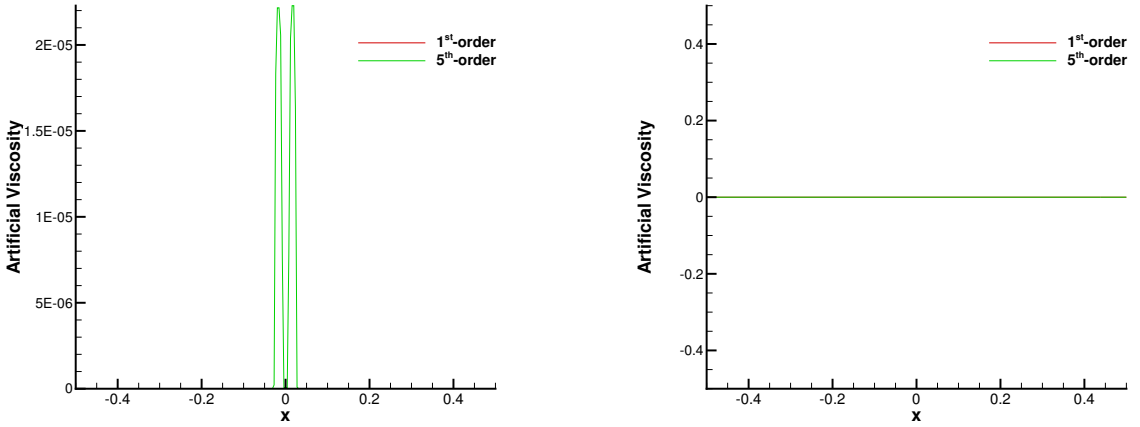


Fig. 4: Density (first row) and pressure (second row) profiles computed with the PPESAD-p5 scheme on uniform grids with 64, 128, 256 elements for the inviscid (left column) and viscous (right column) double rarefaction wave problems at $t = 0.15$.

7.2.2 TWO RAREFACTION WAVES

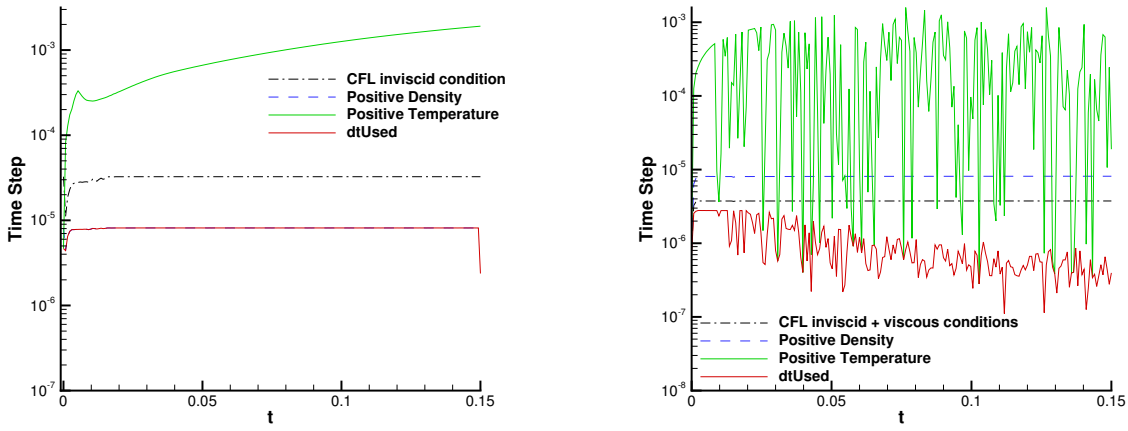
The next test problem is a Riemann problem with two identical rarefaction waves. The initial condition for this test flow is as follows:

$$(\rho, v, P) = \begin{cases} (1, -2, 0.1), & \text{if } -0.5 \leq x \leq 0 \\ (1, 2, 0.1), & \text{if } 0 < x \leq 0.5. \end{cases}$$



(a) (b)

Fig. 5: The low- and high-order ($p = 5$) artificial viscosities obtained on the 256-element grid at $t = 0.0075$ for the inviscid (left panel) and viscous ($Re = 10^3$) double rarefaction wave problems.



(a) (b)

Fig. 6: Time step histories for the inviscid (left panel) and viscous ($Re = 10^3$) double rarefaction wave flows computed with the PPESAD-p5 scheme on 256-element uniform grid.

The initial jump in the velocity profile leads to the development of two rarefaction waves that move in opposite directions. As a result, a vacuum zone forms in the middle of the domain. This is a very challenging problem especially for high-order schemes, because any

spurious oscillations in the numerical solution may generate negative values of density and pressure. The comparison of density and pressure profiles computed with the proposed high-order ($p = 5$) positivity-preserving flux-limiting scheme and the exact solution of this inviscid Riemann problem is presented in Figure 4. In addition to the 1-D Euler equations, we also solve the corresponding Navier-Stokes equations at $Re = 10^3$ with the same initial conditions. Figure 4 (right column) shows density and pressure profiles obtained with the new high-order ($p = 5$) positivity-preserving spectral collocation scheme on a sequence of uniform grids with 64, 128, 256 elements and references solutions computed using the high-order ($p = 4$) entropy stable scheme on a very fine 2048–element uniform grid. For this test problem, the low- and high-order artificial dissipations are added only at the beginning of computation, while no artificial dissipation (except the Merriam-Roe dissipation added at element interfaces) is used in the remainder of the simulation. As follows from these comparisons, the discrete solutions are free of spurious oscillations for all grids considered and converge to the exact and reference solutions for both the Euler and Navier-Stokes equations, respectively.

Similar to the previous test case, we also compare histories of time steps that satisfy the upper bound of the density and temperature positivity constraints and the standard CFL conditions for inviscid and viscous flows with the CFL number set equal to 0.5. For the inviscid flow, the time step required for positivity of temperature is greater than that imposed by the inviscid CFL condition and the time step is solely determined by the density positivity constraint over the entire time interval considered. Note, however, that for the viscous flow at $Re = 10^3$, the time step imposed by the temperature positivity condition (Eq. (134)) varies dramatically and from time to time becomes less than the time steps defined by the density positivity and conventional CFL conditions. Another distinct feature of the proposed high-order entropy stable scheme is that for the viscous flow, the time steps required for positivity of density and temperature (Eqs. (126) and (134)) are in general greater than the time step imposed by the standard CFL condition, as on can see in Figure 6.

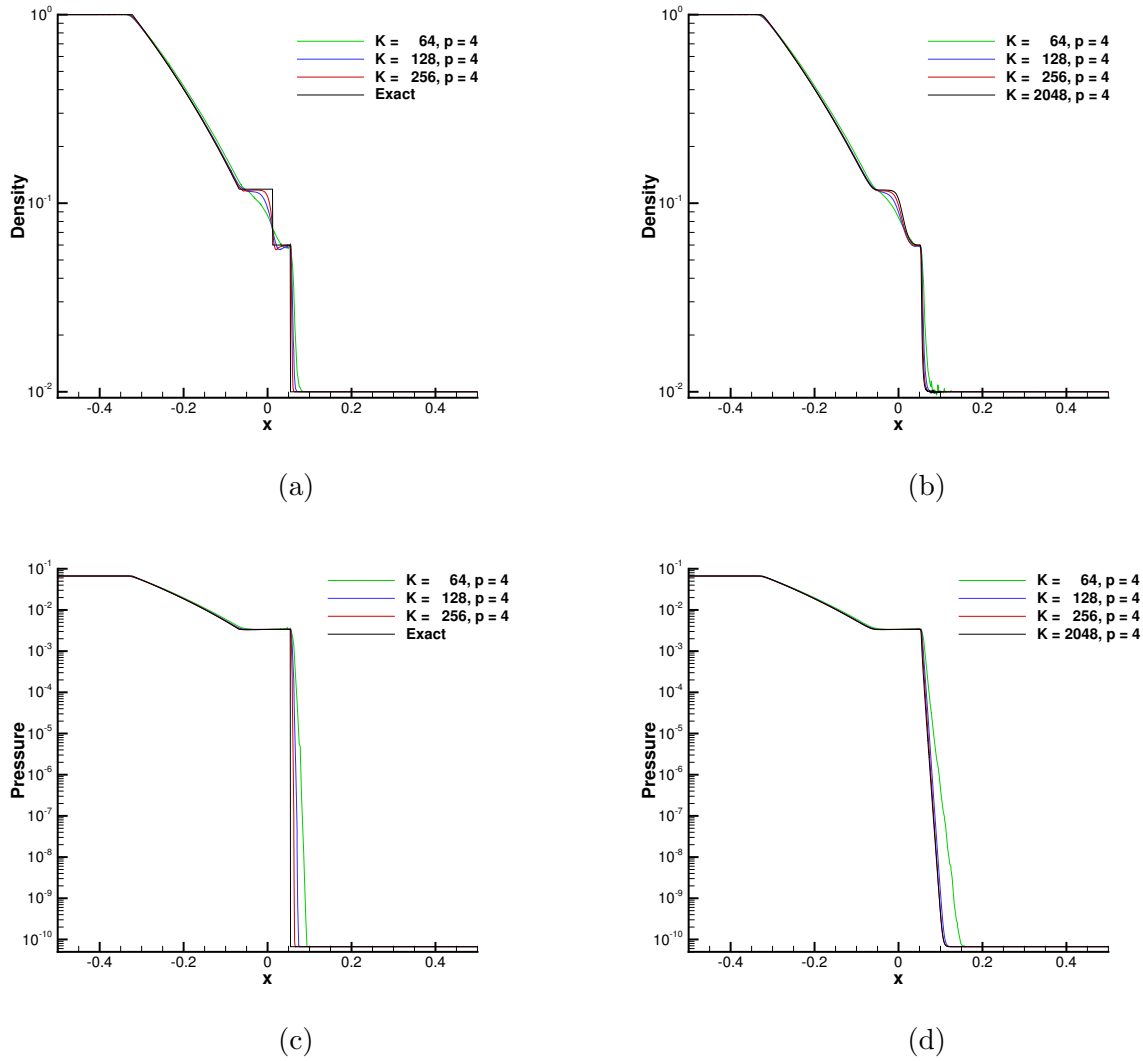


Fig. 7: Density (first row) and pressure (second row) profiles computed with the PPESAD-p4 scheme on uniform grids with 64, 128, 256 elements for the inviscid (left column) and viscous (right column) LeBlanc flows.

7.2.3 LEBLANC SHOCK TUBE PROBLEM

The last 1-D test problem considered is the LeBlanc shock tube problem with the following initial condition:

$$(\rho, v, P) = \begin{cases} (1, 0, 6.666667 \times 10^{-2}), & \text{for } -0.5 \leq x < -0.2 \\ (10^{-2}, 0, 6.666667 \times 10^{-11}), & \text{for } -0.2 \leq x \leq 0.5. \end{cases} \tag{202}$$

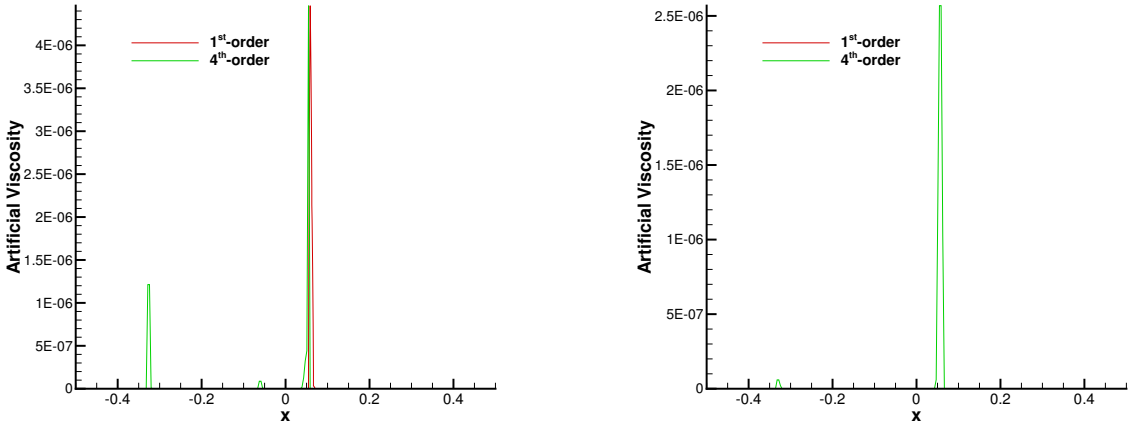
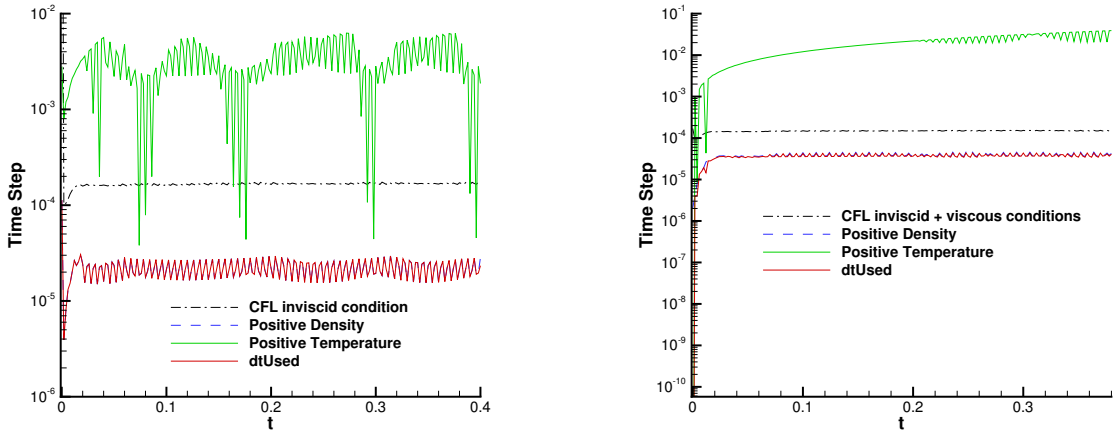


Fig. 8: The low- and high-order ($p = 4$) artificial viscosities obtained on the 256-element grid at $t = 0.4$ for the inviscid (left panel) and viscous LeBlanc flows.

Note that in contrast to the conventional LeBlanc shock tube problem, for which the ratio of specific heats γ is $5/3$, we use $\gamma = 7/5$. To demonstrate the performance of the new family of high-order positivity-preserving artificial dissipation schemes, along with the 1-D Euler equations, we also solve the Navier-Stokes equations with the Reynolds number of 10^5 and the same initial condition given by Eq. (202). As one can see from Eq. (202), the initial pressure and density values across the discontinuity drop down by nine and two orders of magnitude, respectively. The presence of a very strong discontinuity and very low values of the thermodynamic variables at the shock front make this shock tube flow a very challenging problem, because even small amplitude oscillations may lead to negative values of density or pressure. The density and pressure profiles computed by the positivity-preserving high-order ($p = 4$) spectral collocation limiting scheme for the inviscid (left column) and viscous LeBlanc flows on uniform grids with 64, 128, and 256 elements are shown in Figure 7. For both the Euler and Navier-Stokes equations, the proposed high-order flux-limiting scheme provides excellent discontinuity-capturing capabilities and leads to numerical solutions that are nearly free of spurious oscillations, so that density and pressure at each solution point always remain positive. Figure 8 shows the low- and high-order artificial viscosity coefficients for the inviscid (left panel) and viscous flows at the final moment of time, $t = 0.4$. For the



(a) (b)

Fig. 9: Time step histories for the inviscid (left panel) and viscous ($Re = 10^5$) LeBlanc flows computed with the PPESAD-p4 scheme on 256-element uniform grid.

inviscid flow, the first-order artificial dissipation is used only in the element containing the shock, while the high-order dissipation is added in the neighboring element, thus damping oscillation in the transition region. For the viscous flow, no first-order artificial dissipation is used and the high-order dissipation is added only at the shock wave. Note that both the low- and high-order artificial viscosities are nearly equal to zero at the contact discontinuity. For all grids considered, the first- and high-order dissipation operators suppress high-amplitude oscillations near the shock wave and contact discontinuity and provide the positivity of density and pressure during the entire time interval of interest.

Histories of time steps that satisfy the upper bound of the density and temperature positivity constraints and the standard CFL condition for the inviscid (left panel) and viscous ($Re = 10^5$) LeBlanc flows computed with the positivity-preserving high-order ($p = 4$) limiting scheme on 256-element uniform grid are compared in Figure 9. Overall, the time step histories for the inviscid and viscous LeBlanc flows demonstrate a similar behavior as those obtained for the previous test problems. For the inviscid flow, the density positivity condition imposes the most strict constraint on the time step as compared with the temperature positivity and CFL conditions during the entire time interval considered. A similar behavior is observed for the viscous flow except that at the beginning of computation, the time step

imposed by the temperature positivity condition is several orders of magnitude less than that set by the density positivity and conventional CFL conditions. It should be noted that the new velocity and temperature gradient limiters presented in Section 6.2 allow us to eliminate this time step constraint and integrate the discretized Navier-Stokes equations with the time step comparable to that used for the Euler equations.

7.3 2-D AND 3-D NUMERICAL RESULTS

We now present 2-D and 3-D numerical results obtained using the high-order positvity-preserving flux-limiting scheme presented in this dissertation. While the slightly different method used to obtain the 1-D results presented in Section 7.2 worked well for 1-D problems, we found that the form presented in this dissertation was less dissipative for smooth features and had a better convergence rate for steady state problems. In this section, we used $c_\rho = 0.9$ and $c_T = c_\rho \frac{c_P}{\gamma}$ (see Eq. (232)). See Appendix B for an explicit discussion concerning the implementation of boundary conditions used for the 2-D and 3-D numerical sections.

For the simulations that converge to a steady state, the following element-wise norm is used to measure convergence

$$\|\hat{\mathbf{U}}_t\|_{L_2,k} = \sqrt{\frac{\frac{d\hat{\mathbf{U}}}{dt}_k^\top \mathcal{P} \left[J^{-1}\right]_k \frac{d\hat{\mathbf{U}}}{dt}_k}{\mathbf{1}_5^\top \mathcal{P} \left[J\right]_k \mathbf{1}_5}} \tag{203}$$

for the kth element. To measure global convergence of the K total elements in the domain, we use

$$\|\hat{\mathbf{U}}_t\|_{L_2} = \sqrt{\frac{\sum_{k=1}^{K} \frac{d\hat{\mathbf{U}}}{dt}_k^\top \mathcal{P} \left[J^{-1}\right]_k \frac{d\hat{\mathbf{U}}}{dt}_k}{\sum_{k=1}^{K} \mathbf{1}_5^\top \mathcal{P} \left[J\right]_k \mathbf{1}_5}}. \tag{204}$$

7.3.1 3-D VISCOUS SHOCK

We now consider the propagation of a 3-D viscous shock on uniform and non-uniform grids. This problem possesses a smooth analytical solution; however, for insufficient grid resolution the problem can appear to possess a shock discontinuity. Hence, we use this

TABLE 1: Final L_∞ and L_2 errors and their convergence rates obtained with the ESSC and PPESAD schemes for $p = 4, 5, 6$ for the viscous shock problem on uniform grids with K^3 number of elements. Bold numbers indicate simulations where PPESAD used artificial dissipation. For all non-bold entries, the PPESAD simulation was identical to the ESSC simulation.

	ESSC				PPESAD			
K	L_∞ error	rate	L_2 error	rate	L_∞ error	rate	L_2 error	rate
				$p = 4$				
3	**1.96**	–	**3.99e-2**	–	**0.76**	–	**3.99e-2**	–
6	**0.66**	1.57	**7.11e-3**	2.49	**0.66**	0.21	**7.11e-3**	2.49
12	2.40e-2	4.78	5.50e-4	3.69	2.40e-2	4.78	5.50e-4	3.69
24	1.15e-3	4.39	2.38e-5	4.53	1.15e-3	4.39	2.38e-5	4.53
48	4.54e-5	4.66	9.27e-7	4.68	4.54e-5	4.66	9.27e-7	4.68
				$p = 5$				
3	**1.93**	–	**2.69e-2**	–	**0.67**	–	**2.16e-2**	–
6	9.95e-2	4.28	2.63e-3	3.36	9.95e-2	2.75	2.63e-3	3.03
12	5.07e-3	4.29	1.41e-4	4.22	5.07e-3	4.29	1.41e-4	4.22
24	1.40e-4	5.18	3.23e-6	5.44	1.40e-4	5.18	3.23e-6	5.44
48	2.74e-6	5.67	4.52e-8	6.16	2.74e-6	5.67	4.52e-8	6.16
				$p = 6$				
3	**0.44**	–	**1.32e-2**	–	**0.42**	–	**1.31e-2**	–
6	0.11	2.01	1.36e-3	3.27	0.11	1.97	1.36e-3	3.27
12	1.31e-3	6.37	3.27e-5	5.38	1.31e-3	6.37	3.27e-5	5.38
24	2.28e-5	5.85	2.95e-7	6.79	2.28e-5	5.85	2.95e-7	6.79
48	2.13e-7	6.74	3.04e-9	6.60	2.13e-7	6.74	3.04e-9	6.60

problem to test the ability of the proposed PPESAD scheme to detect and dissipate under-resolved and discontinuous features in the flow, while not destroying accuracy or the error convergence properties of the underlying ESSC scheme. The derivation of the analytical solution and initial conditions can be found in [42, 52]. We rotated the planar shock so that it propagates along the direction $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$ and is initially centered at the origin. We used the following simulation parameters: $Re = 50$, $Ma = 2.5$, and $Pr = 3/4$. The simulation was run from $t_{\text{initial}} = 0$ to $t_{\text{final}} = 0.1$. We penalized against the exact solution at all domain boundaries. We used two sets of grids. The first set of grids consisted of equal sized cubes partitioning the domain $-0.5 \leq x, y, z \leq 0.5$. The second set of grids was formed from the first set by randomly perturbing the coordinates of each vertex in the domain. Specifically,

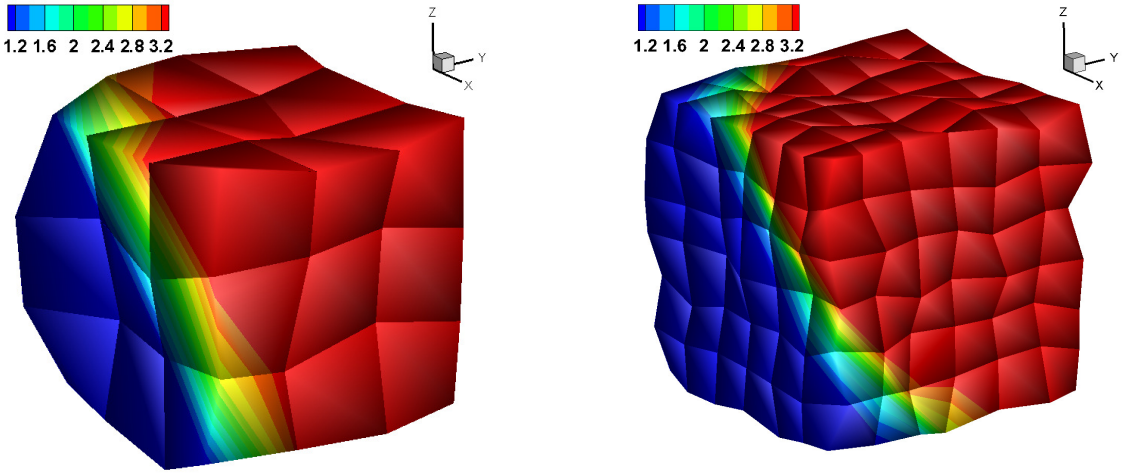
TABLE 2: Final L_∞ and L_2 errors and their convergence rates obtained with the ESSC and PPESAD schemes for $p = 4, 5, 6$ for the viscous shock problem on non-uniform grids with K^3 number of elements. Bold numbers indicate simulations where PPESAD used artificial dissipation. For all non-bold entries, the PPESAD simulation was identical to the ESSC simulation.

	ESSC				PPESAD			
K	L_∞ error	rate	L_2 error	rate	L_∞ error	rate	L_2 error	rate
				$p = 4$				
3	**1.24**	–	**4.75e-2**	–	**0.68**	–	**4.03e-2**	–
6	**0.80**	0.63	**8.50e-3**	2.48	**0.56**	0.27	**8.20e-3**	2.30
12	0.11	2.89	8.51e-4	3.32	0.11	2.37	8.51e-4	3.27
24	6.93e-3	3.96	4.96e-5	4.10	6.93e-3	3.96	4.96e-5	4.10
48	3.09e-4	4.49	1.54e-6	5.01	3.09e-4	4.49	1.54e-6	5.01
				$p = 5$				
3	**3.15**	–	**3.30e-2**	–	**0.88**	–	**2.99e-2**	–
6	0.34	3.20	4.13e-3	3.00	0.34	1.36	4.13e-3	2.86
12	4.37e-2	2.97	2.49e-4	4.05	4.37e-2	2.97	2.49e-4	4.05
24	2.30e-3	4.25	7.77e-6	5.00	2.30e-3	4.25	7.77e-6	5.00
48	3.50e-5	6.04	9.99e-8	6.28	3.50e-5	6.04	9.99e-8	6.28
				$p = 6$				
3	**1.27**	–	**2.11e-2**	–	**0.52**	–	**1.99e-2**	–
6	0.12	3.35	1.92e-3	3.46	0.12	2.07	1.92e-3	3.38
12	1.44e-2	3.11	7.33e-5	4.71	1.44e-2	3.11	7.33e-5	4.71
24	3.27e-4	5.46	1.20e-6	5.94	3.27e-4	5.46	1.20e-6	5.94
48	3.06e-6	6.74	7.56e-9	7.31	3.06e-6	6.74	7.56e-9	7.31

for a uniform grid with K^3 total elements, the corresponding non-uniform grid with K^3 total elements was formed by adding r/K to each coordinate of each vertex in the domain where the variable r is a random number that is generated for each coordinate of each vertex and is in the set $[0, 0.4)$. See Figure 10 for the 3^3 and 6^3 non-uniform grids.

For each simulation, we recorded the error of the numerical solution at $t = t_{\text{final}}$. The global L_2 error on the K^3 elements is calculated as

$$\|\mathbf{U} - \mathbf{U}^{\text{ex}}\|_{L_2} = \sqrt{\frac{\sum_{k=1}^{K^3} \left(\hat{\mathbf{U}}_k - \hat{\mathbf{U}}_k^{\text{ex}}\right)^\top \mathcal{P} \left[J^{-1}\right]_k \left(\hat{\mathbf{U}}_k - \hat{\mathbf{U}}_k^{\text{ex}}\right)^\top}{5 \sum_{k=1}^{K^3} \mathbf{1}_5^\top \mathcal{P} \left[J\right]_k \mathbf{1}_5}}, \tag{205}$$



(a) (b)

Fig. 10: Initial density for the 3-D viscous shock on the 3^3 (left) and 6^3 (right) non-uniform grids.

where $\hat{\mathbf{U}}_k^{\text{ex}}$ is the array of conservatives variables for the exact solution on the kth element (scaled by $[J]_k$). The global L_∞ error is calculated as

$$\|\mathbf{U} - \mathbf{U}^{\text{ex}}\|_{L_\infty} = \max_{1\leq k\leq K^3} \max_{1\leq i\leq N_p} \|\mathbf{U}_k^{\text{ex}}(\vec{\xi}_i) - \mathbf{U}_k(\vec{\xi}_i)\|_{L_\infty}, \tag{206}$$

where N_p is the total number of solution points on an element. The final errors are presented in Tables 1 and 2.

In Tables 1 and 2, bold numbers indicate simulations where PPESAD used artificial dissipation. For all non-bold entries, the PPESAD simulation was identical to the ESSC simulation. Recall that the PPESAD scheme is designed so that if $\theta_f^k = 1$ for all elements, $\boldsymbol{\mu}^{AD}$ is zero for all elements, and the velocity and temperature limiters of Section 6.2 are not used, then the PPESAD scheme is equivalent to the ESSC scheme (see Eq. (178)). We chose the Reynolds number large enough that the PPESAD scheme differed from the ESSC scheme on the coarsest meshes, but the Reynolds number is also small enough that we can see the PPESAD scheme reverting to the baseline ESSC scheme as grid resolution increases. In particular, looking at Tables 1 and 2 we see that for all simulations on the 3^3 grids and all $p = 4$ simulations on the 6^3 grids, the PPESAD scheme differs from the ESSC scheme.

Furthermore, every time the PPESAD scheme simulation differed from the ESSC scheme it acted in a way that either reduced or did not increase the error. Once the grid resolution was sufficient for the ESSC scheme error to begin converging near design order accuracy, the PPESAD scheme reverted to the ESSC scheme as it is supposed to do. Hence, these results indicate that the proposed PPESAD scheme detects and dissipates under-resolved and discontinuous features in the flow in a manner that is error reducing, while not destroying accuracy or the error convergence properties of the underlying ESSC scheme.

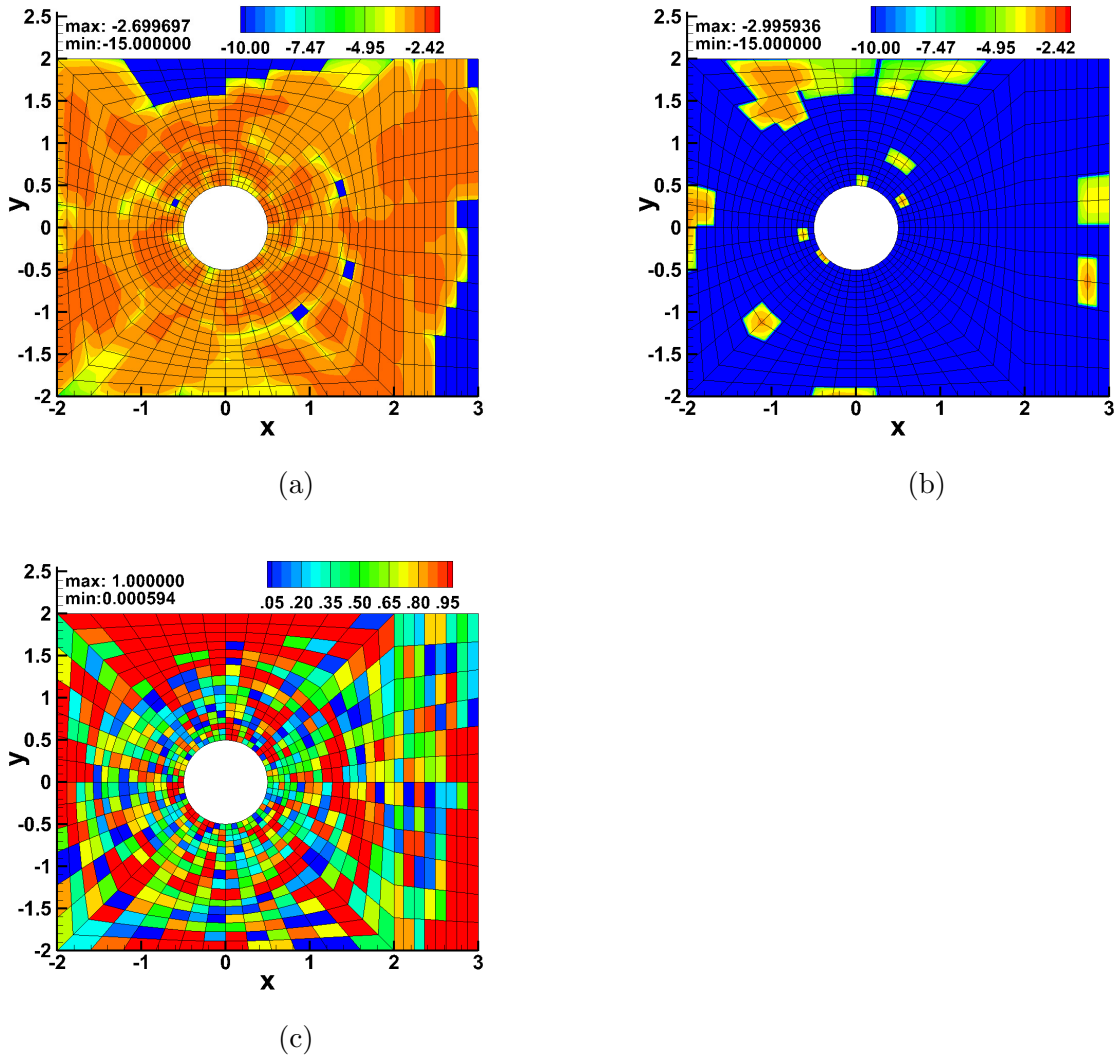
7.3.2 FREESTREAM PRESERVATION ON 2-D CYLINDER

In Theorem 16, we prove that the high-order positivity–preserving flux-limiting scheme given by Eq. (178) is freestream preserving. To demonstrate this, we simulate a uniform state on a 2-D grid with a cylinder and elements that are genuinely curvilinear in a region surrounding the cylinder. The grid has a total of 864 elements and is constructed in a manner similar to the grids used in Section 7.3.6 for the hypersonic cylinder. See Figure 11. All boundaries use the initial state for forming penalties and we simulate till $t_{\text{final}} = 10$. The initial state is $\rho = 1$, $T = 1$, and $\vec{\boldsymbol{V}} = \begin{bmatrix} \cos(10^\circ) & \sin(10^\circ) & 0 \end{bmatrix}^\top$. We used the following simulation parameters: $Re = 500$, $Ma = 3.5$, and $Pr = 0.7$. To ensure that all terms in the high-order positivity–preserving flux-limiting scheme given by Eq. (178) turn on during the simulation, at every Runge-Kutta stage we randomly set the artificial viscosities $\boldsymbol{\mu}_p^{AD}$ and $\bar{\boldsymbol{\mu}}_1^{AD}$ (see Section 6.3.5) up to a maximum near $1/Re$, and the flux limiter (see Section 6.3.3) in the range $0 \leq \theta_f < 1$. See Figure 11 for the randomly generated values at $t = 10$.

As can be see from Figure 11, all terms in the high-order positivity–preserving flux-limiting scheme given by Eq. (178) were used throughout the simulation. Nonetheless, the final global L_2 error (see Eq. (205)) was 2.84e−15 and the global L_∞ error (see Eq. (206)) was 1.46e−13. Hence, this example confirms what we have proven in Theorem 16.

7.3.3 ENTROPY CONSERVATION FOR ISENTROPIC VORTEX

In Lemma 2, we prove that the first-order inviscid term is entropy conservative. To demonstrate this, we simulate the rightward propagation of an inviscid isentropic vortex on a randomly perturbed coarse grid (see Figure 12). This problem is an exact solution for



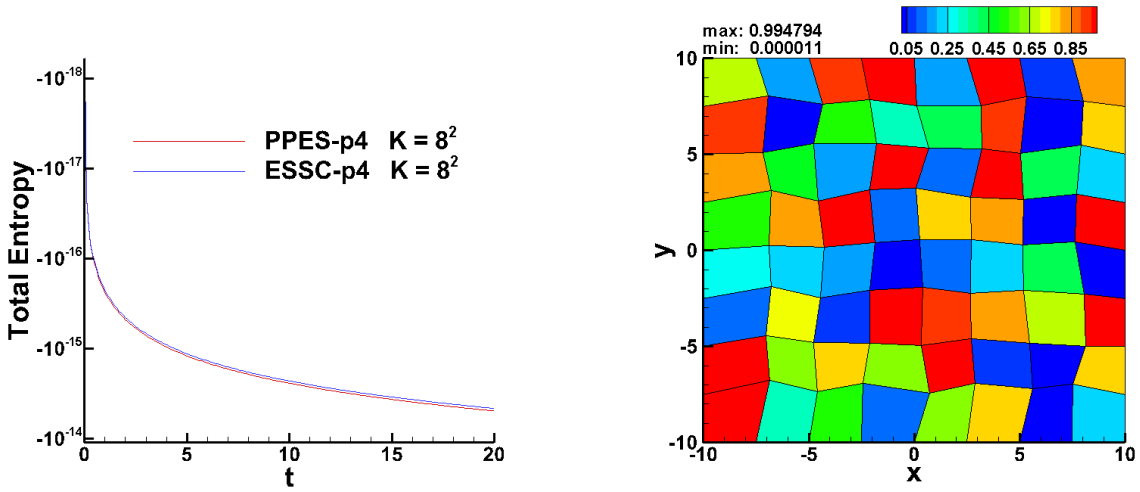
(a) (b)

(c)

Fig. 11: Randomly generated low-order artificial viscosity (top-left), high-order artificial viscosity (top-right), and flux limiter (bottom-left) are displayed for the PPESAD-p4 solution of the freestream preservation problem at $t = 10$. The $\log_{10}$ of the artificial viscosities are plotted. Element edges are displayed.

the Euler equations (e.g., see [34, 52]). For this inviscid problem, we used $Ma = 0.3$. All boundaries are periodic. The vortex is initially centered at $(0, 0)$, propagates to the right and returns to the center by $t_{\text{final}} = 20$.

Since this problem consists of a smooth inviscid flow with periodic boundary conditions, the ESSC and PPES schemes semi-discretely conserve the total entropy in the domain–which is initially zero–if all dissipation terms are turned off. Semi-discrete entropy conservation



(a) (b)

Fig. 12: Time series plot (left) of the total entropy for the modified ESSC-p4 and PPES-p4 solutions of the isentropic vortex simulation. All entropy dissipative terms were turned off (Brenner, interface penalties, etc.) and the PPES-p4 simulation used a randomly generated flux limiter coefficient (right) in each element. Simulations were run on a $K = 8^2$ mesh with interior vertices randomly perturbed (right).

implies that the evolution of the total entropy is proportional to the truncation error of the temporal discretization. Hence, we turned off all dissipative terms for the ESSC and PPES schemes and for the PPES scheme we randomly generated a flux limiter in each element (see Section 6.3.3) in the range $0 < \theta_f < 1$. With this setup, the ESSC and PPES schemes only differ by the presence of the first-order inviscid terms in the PPES scheme. Despite this difference and the coarse randomly perturbed grid, both schemes conserve the discrete total entropy up the to the order of the round off error for sufficiently small time steps–see Figure 12. For both schemes, we used a constant $\Delta t = 2\text{e}{-4}$.

7.3.4 2-D SHOCK DIFFRACTION

We now consider the diffraction of a rightward moving shock of Mach numbers 5.09 and 200 for viscous and inviscid flows. High speed shocks diffracting over sharp corners are well known for producing negative densities and pressures in numerical simulations; hence, this problem serves as an excellent example of the robustness of the proposed scheme.

The computational domain is shown in Figure 13. For all shock diffraction simulations,

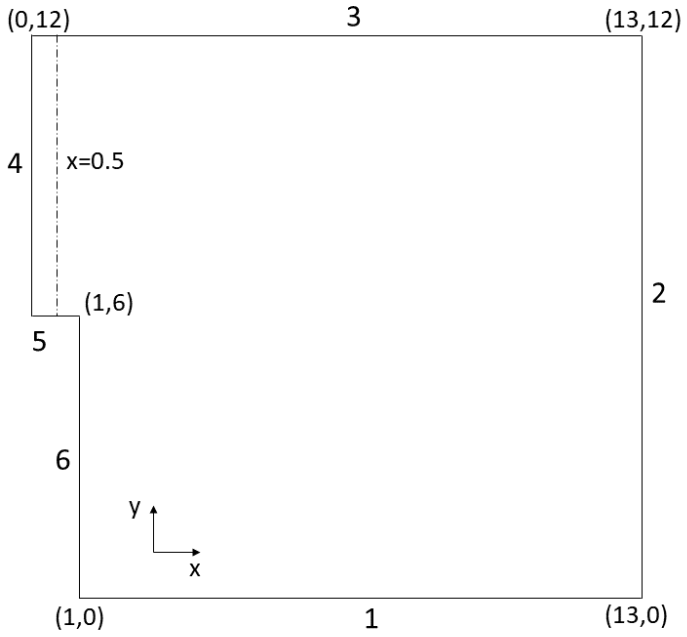
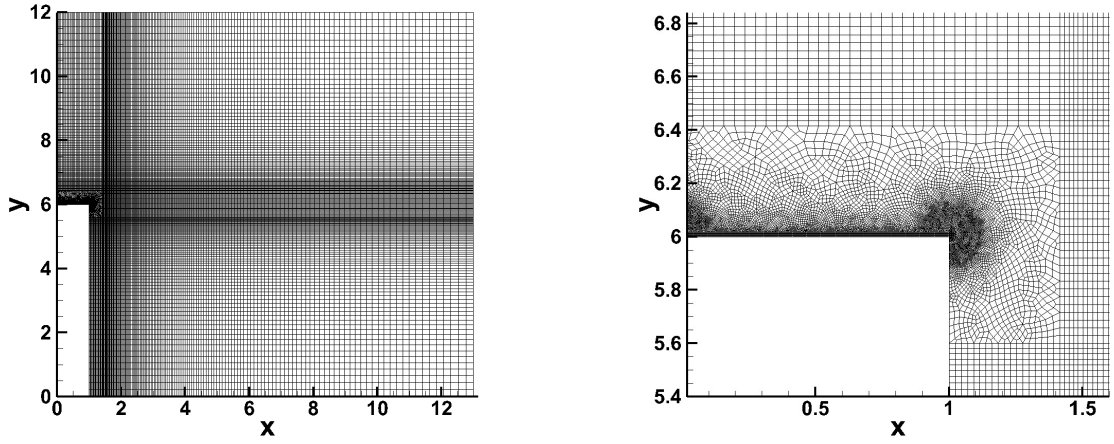


Fig. 13: The computational domain for the shock diffraction problem is bounded by the boundary lines 1-6. The dashed line shows the initial location of the rightward moving shock.

the boundary conditions are: outflow at boundaries 1 and 2, inflow at boundary 4, and slip walls at boundaries 3 and 6. For the viscous flows, we use entropy stable adiabatic no-slip wall boundary conditions at boundary 5. For the inviscid flows, we use slip wall boundary conditions at boundary 5. See Appendix B for an explicit discussion concerning the implementation of boundary conditions used for the 2-D and 3-D numerical sections.

For all shock diffraction simulations, $Ma = 1/\sqrt{\gamma}$ so that $P = \frac{\rho T}{\gamma Ma^2} = \rho T$. The initial conditions consist of a rightward moving shock of a given Mach number located at x = 0.5 and $6 \leq y \leq 12$. On the right side of the shock, the initial conditions are $\rho = 1.4$, $P = 1$, and $\vec{\boldsymbol{V}} = 0$. The left side of the shock is determined using the Rankine–Hugoniot conditions and the given shock speed. For the viscous flows, we use the Blasius boundary layer solution near the wall on the left side of the shock with freestream conditions corresponding to the Mach number of the shock. Let Ma_s be the Mach number of the initial shock for a given simulation. The simulations are integrated over the time interval $0 \leq t \leq \frac{2.3*5.09}{Ma_s}$. For the

viscous simulations, Sutherland's law is used, $Pr = 0.75$ and $Re = 10^4$.



(a) (b)

Fig. 14: The computational grid for the viscous shock diffraction problem. Note that in the boundary layer along the no-slip wall and at the corner, a structured grid is used. Element edges are displayed.

For all inviscid shock diffraction simulations, we used a uniform rectangular mesh with $\Delta x = \Delta y$ and $40,000$ total elements. For the viscous shock diffraction simulations, we used a grid with a total of $36,027$ elements. The grid for the viscous shock diffraction problem has three regions: 1) the region that is a distance of 0.4 or more from boundary line 5, 2) the boundary layer region which contains all points that are within a distance of 0.016 of boundary line 5, and 3) the unstructured grid region which connects regions (1) and (2). Region (2), the boundary layer region, used a uniform rectangular grid with $\Delta x = 1.5\Delta y$ and had 6 elements in the wall normal direction; hence, in the boundary layer $\Delta y = 0.016/6$.

Shock of Mach number 5.09

We begin by comparing the PPES-p4 and PPESAD-p4 solutions for a Mach number 5.09 shock. The ESSC scheme fails to preserve positivity for shocks of Mach number greater than ≈ 3 (depending on the polynomial order) for this shock diffraction problem. However,

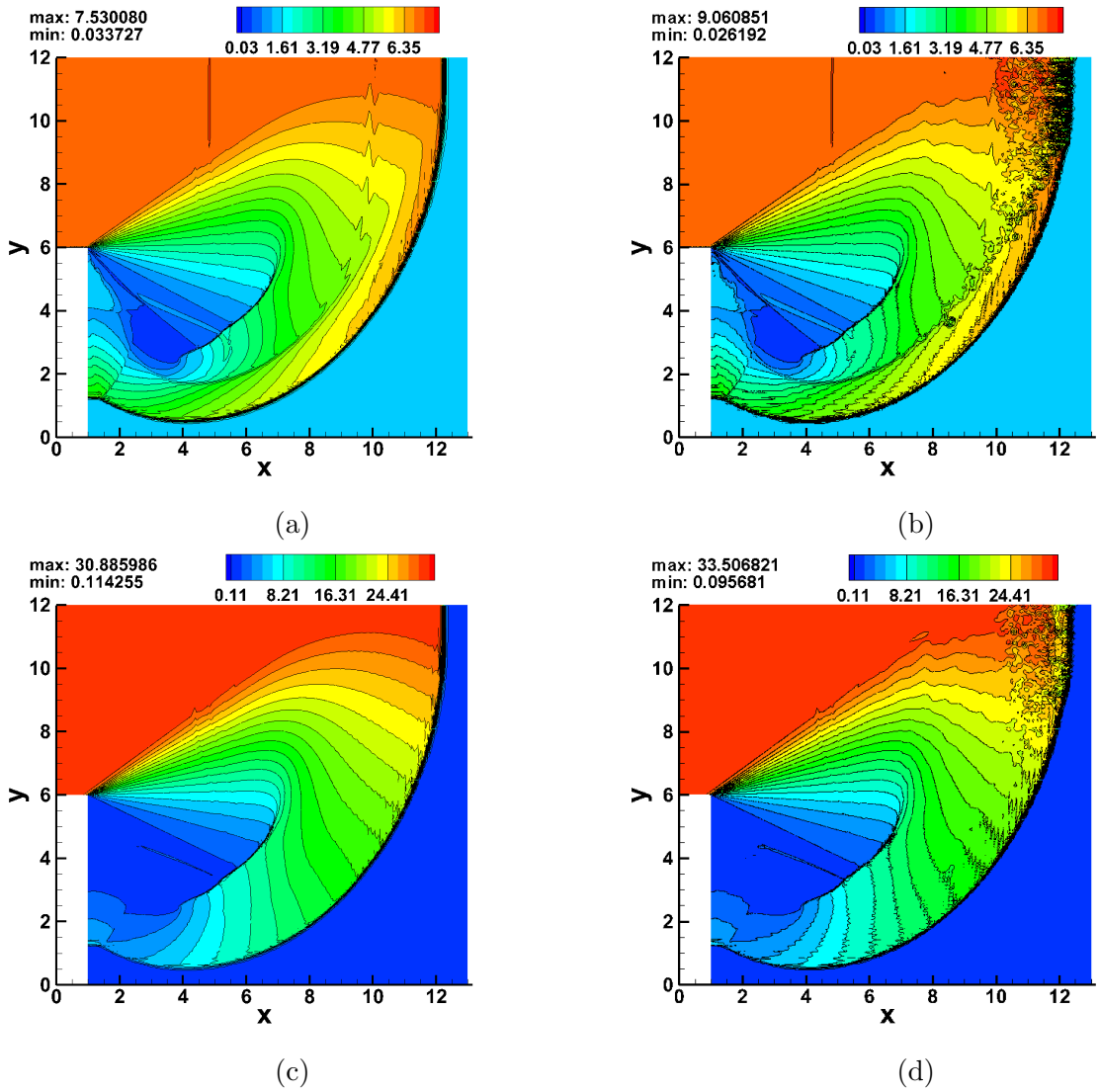


Fig. 15: Density (top row) and pressure (bottom row) are shown for the viscous shock diffraction problem with shock of Mach number 5.09. The left column shows the PPESAD-p4 solution. The right column shows the PPES-p4 solution.

looking at the contour plots in Figure 17 of the flux limiter, θ_f^k, for the inviscid and viscous PPES-p4 solutions, we see that $\theta_f^k = 1$ almost everywhere except for a relatively small number of elements including the shock regions and the corner. Recall, from Eq. (178), that when $\theta_f^k = 1$ the PPES-p4 method is equivalent to the ESSC-p4 method for the kth element for inviscid problems. For viscous problems, they are equivalent if the velocity and temperature limiters of Section 6.2 are not used. The velocity and temperature limiters were

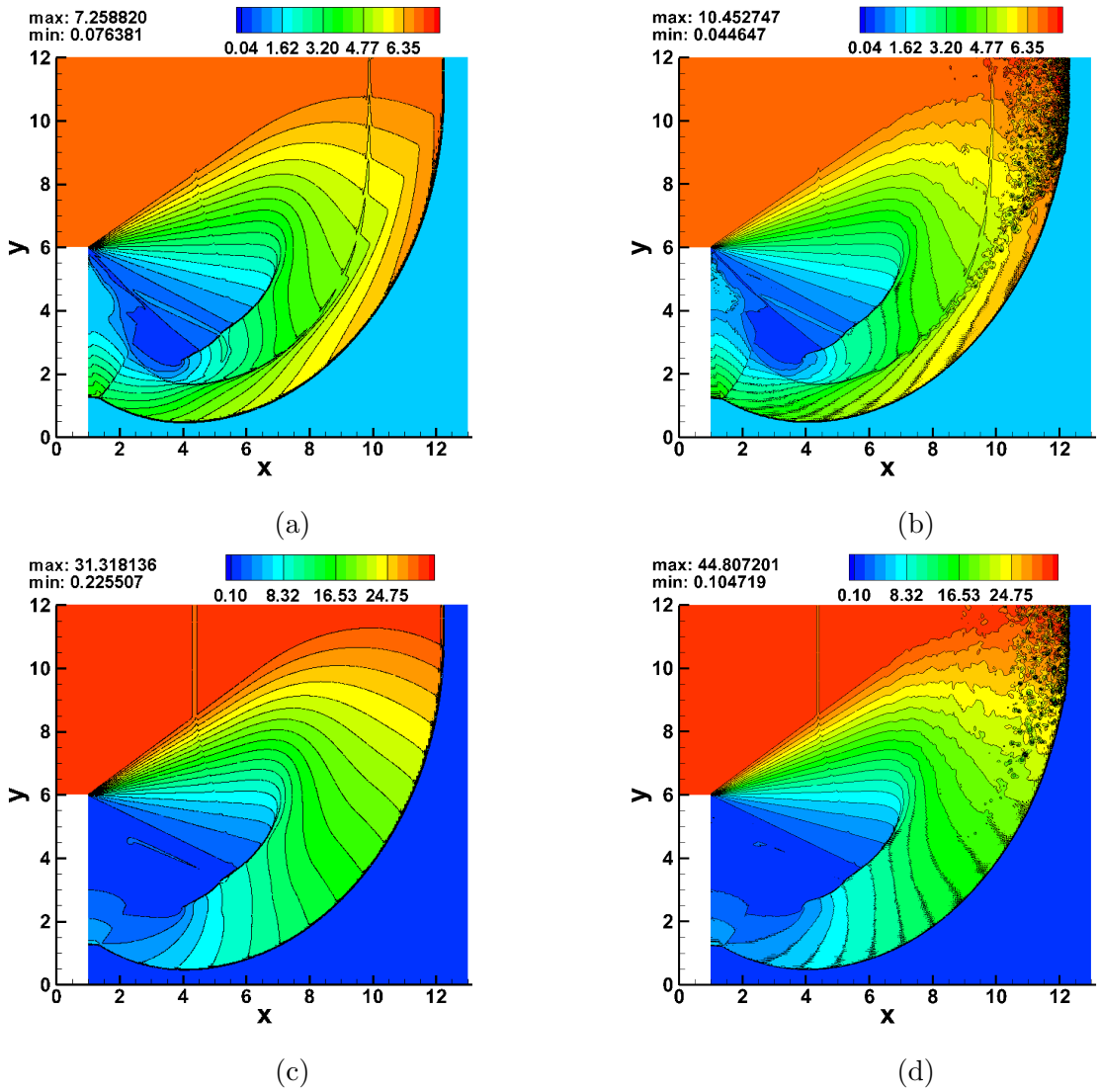
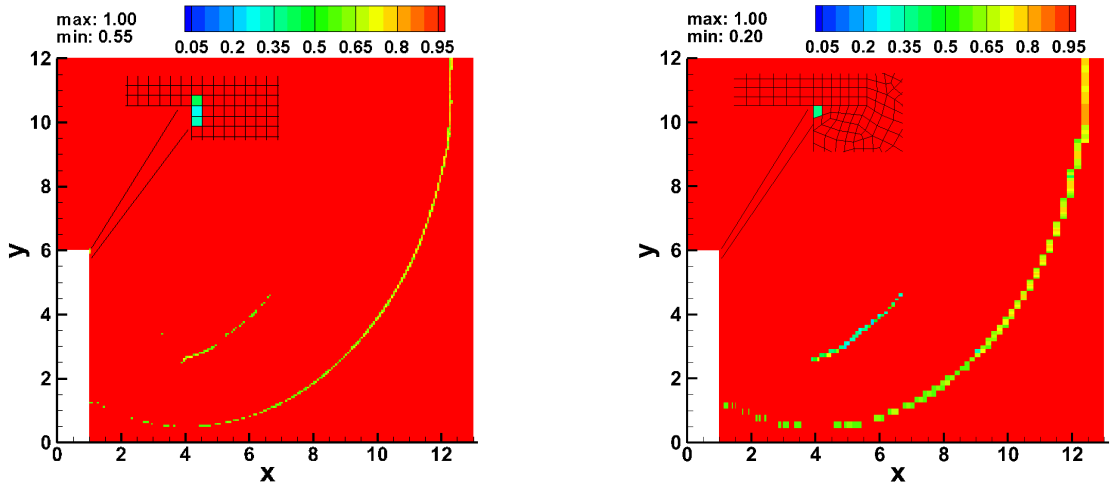


Fig. 16: Density (top row) and pressure (bottom row) are shown for the inviscid shock diffraction problem with shock of Mach number 5.09. The left column shows the PPESAD-p4 solution. The right column shows the PPES-p4 solution.

used for the PPES-p4 solution of the viscous shock diffraction problem but only two elements near $(x, y) = (1, 5.94)$ used the limiters throughout the entire simulation. Therefore, we can reasonably interpret the PPES-p4 solution as a close approximation to the ESSC-p4 solution.

Looking at Figure 15, we see that the PPESAD-p4 solution retains the features present in the PPES-p4 solution for the viscous shock diffraction problem. In Figure 16, we see the same result for the inviscid shock diffraction problem. For both comparisons, the solutions



(a) (b)

Fig. 17: Flux limiter plot for the PPES-p4 solution of the inviscid (left) and viscous (right) shock diffraction problem with shock of Mach number 5.09. Sub box in each figure shows flux limiter for corner elements.

differ most in the shock region where $y \geq 6$, where the shock strength is largest. The lack of sufficient dissipation in this region causes the PPES-p4 solution to produce spurious oscillations that pollute the surrounding regions and serves to illustrate the important stabilization role that $\boldsymbol{\mu}^{AD}$ plays in forming the PPESAD-p4 solution.

Shock of Mach number 200

Now, we consider the shock diffraction problem for the case with an initial shock of Mach number 200. The density, pressure, and artificial viscosity results for the inviscid and viscous case of this problem can be seen in Figures 18 and 19. Notice that the contour color range for the density plots goes up to ≈ 9, but the maximum values are ≈ 20. For both the inviscid and viscous simulations, the maximum density values are obtained in the region near $(x, y) \approx (1, 2.5)$. Everywhere else, the density is no greater than ≈ 9; hence, the maximum color contour was chosen to be ≈ 9. Notice that the artificial viscosity near the shock for the viscous solution is spread out over a wider area. This is partly due to the fact that the viscous grid has significantly less resolution than the inviscid grid (outside the region near the solid wall).

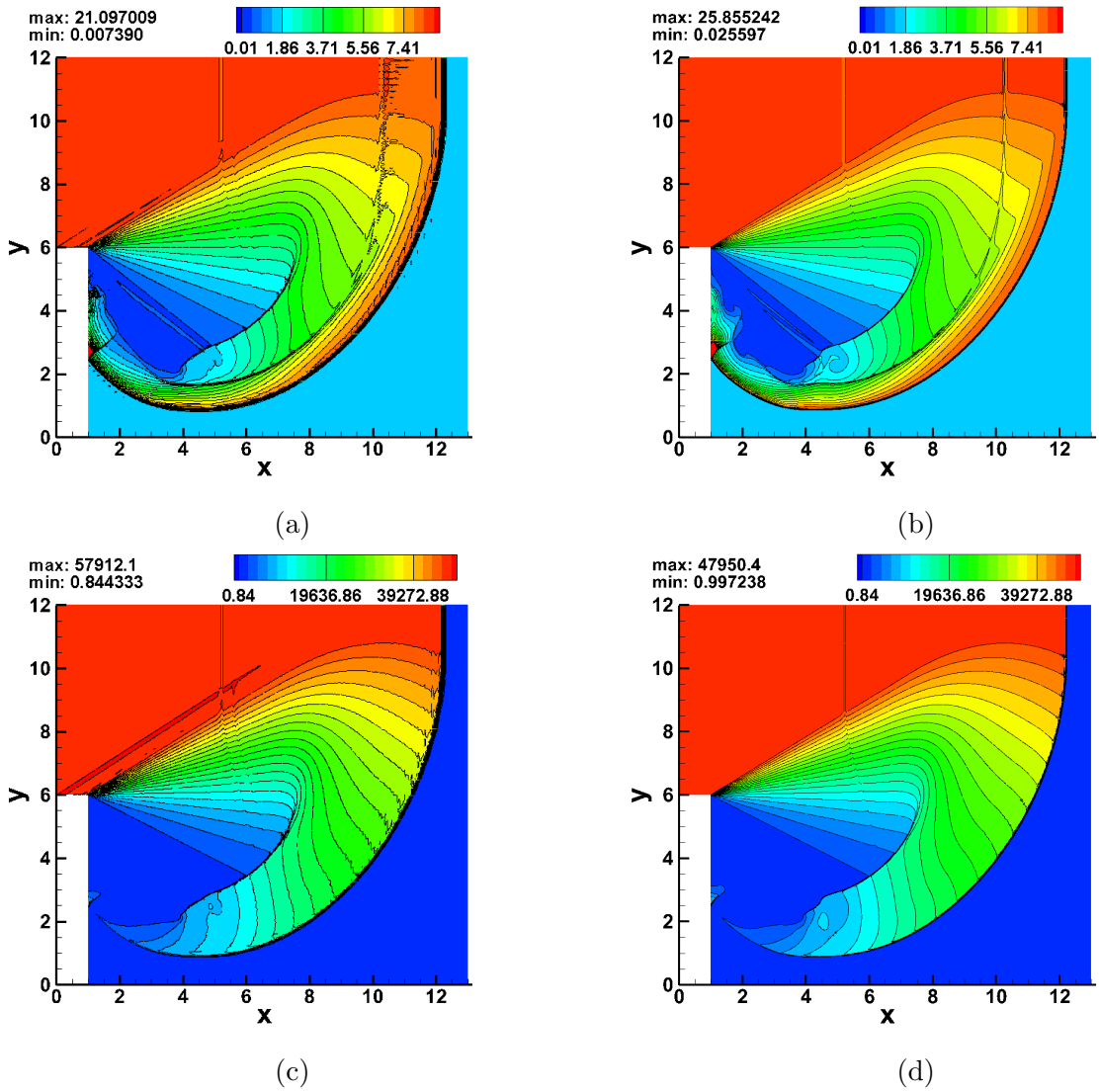
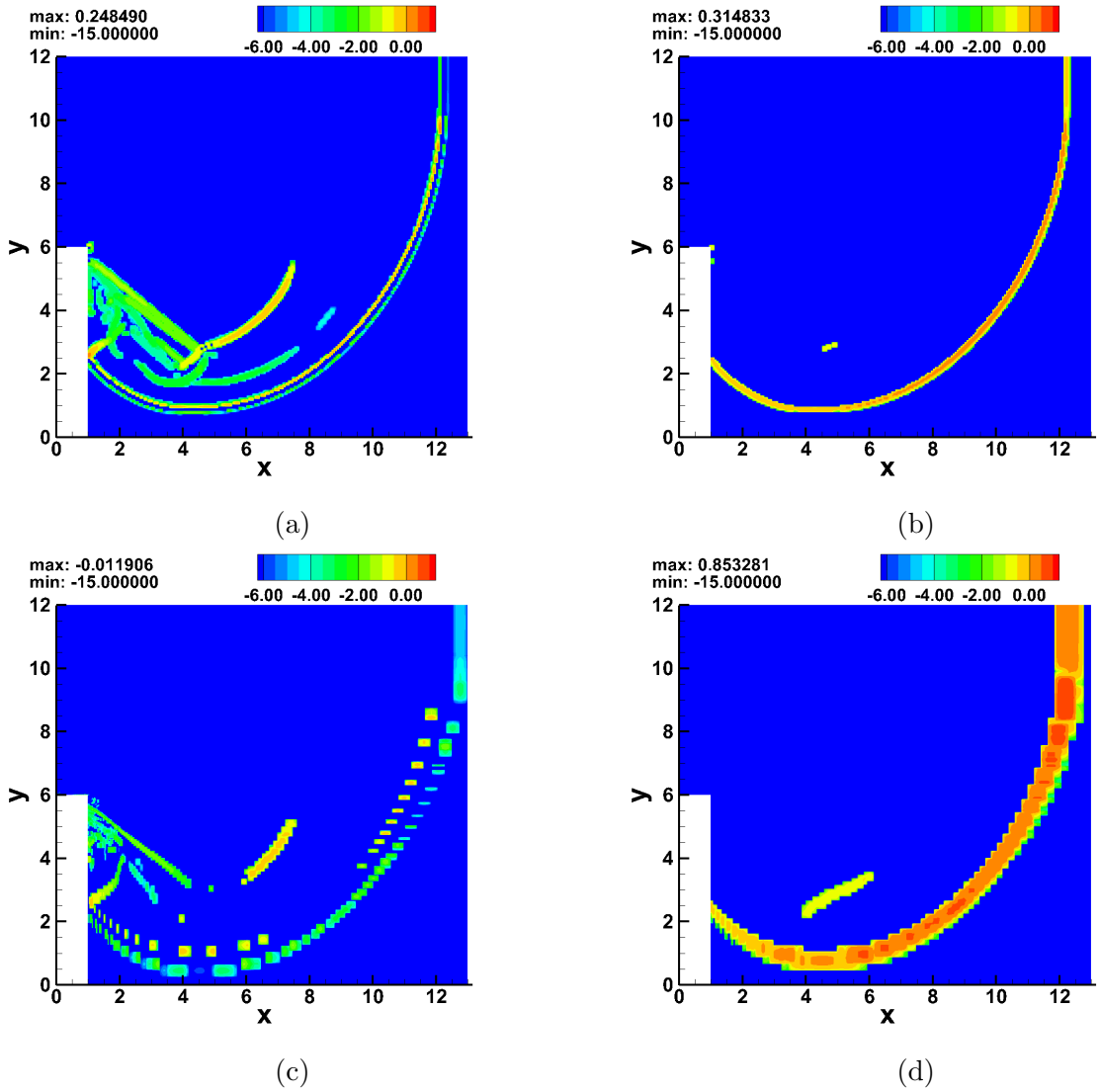


Fig. 18: Density (top row) and pressure (bottom row) are shown for the viscous (left) and inviscid (right) shock diffraction problem with shock of Mach number 200. All solutions were obtained with the PPESAD-p5 scheme.

The inviscid simulation was not significantly more difficult to run (in terms of issues such as stiffness) than the case of the inviscid shock of Mach number 5.09. We attribute this largely to the fact that for the inviscid case the flux limiter can switch the scheme to fully first-order when necessary (see Eq. (178)). Using a fully first-order scheme reduces the stencil width which contributes to a less oscillatory solution in troubled regions such as at the shock.

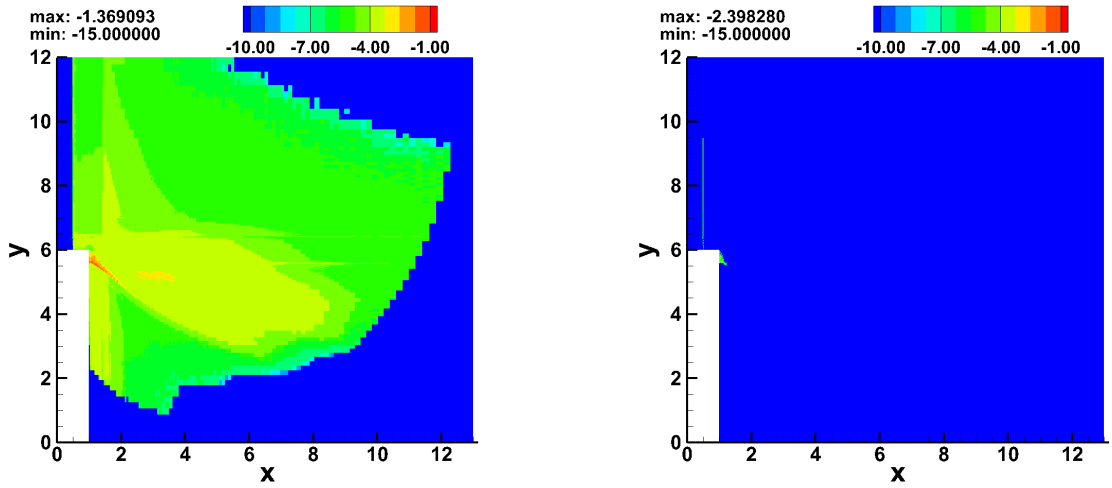


(a) (b)

(c) (d)

Fig. 19: High-order (left column) and low-order (right column) artificial viscosity ($\log_{10}$) of the PPESAD-p5 solution of the inviscid (top row) and viscous (bottom row) shock diffraction problem with shock of Mach number 200.

However, for the viscous case we always have the high-order physical viscous term. The high-order physical viscous term for this problem creates significant stiffness immediately and throughout the simulation (e.g., the initial temperature positivity constraint is $\Delta t \lessapprox 10^{-14}$) if not dealt with. Using the first-order artificial dissipation and inviscid terms alone is not sufficient to reduce this stiffness adequately. Hence, the discretely entropy stable velocity and temperature limiters of Section 6.2 must be used for this simulation. Unlike in the



(a) (b)

Fig. 20: The cumulative usage of the temperature (left) and V_1 (right) entropy stable limiters are shown for the PPESAD-p5 solution of the viscous shock diffraction problem with shock of Mach number 200. The Θ_k^t variable is plotted in the left sub figure and the $(\Theta_1^v)_k$ variable is plotted in the right sub figure, see Eq. (207).

case of the Mach 5.09 viscous shock diffraction problem and all the viscous 1-D problems of Section 7.2, the bounds in Eq. (137) were not sufficiently small enough to cause the velocity and temperature limiters to be used. Hence, if the bounds in Eq. (137) are used for the viscous shock diffraction with shock of Mach number 200, the limiters never turn on and the problem remains stiff. To fix this issue, we scaled the bounds in Eq. (137) by $\max(1/h, \sqrt{Re})/Re$. This brought the bounds down sufficiently low enough for the limiters to be used.

To visualize the usage of the velocity (θ_1^v and θ_2^v, see Eq. (140)) and temperature (θ^t, see Eq. (152)) limiters which are element-wise constants for each iteration, we consider the variables

$$\begin{aligned}(\Theta_1^v)_k &= \sum_{i=1}(\theta_1^v)_k(\mathrm{RK}_i),\\ (\Theta_2^v)_k &= \sum_{i=1}(\theta_2^v)_k(\mathrm{RK}_i),\\ \Theta_k^t &= \sum_{i=1}(\theta^t)_k(\mathrm{RK}_i),\end{aligned} \tag{207}$$

where for the kth element Θ_k^t is the sum of the θ_k^t values where the sum is taken over all Runge-Kutta stages over the entire simulation. Recall that we used no more than one iteration of the velocity and temperature limiters per Runge-Kutta stage. Looking at Figure 20, we see that θ^t is used substantially more often than θ_1^v. This implies that the variation in V_1 was smaller than the modified upper bound given by Eq. (137) (and multiplied by $\max(1/h, \sqrt{Re})/Re$) for most places in the domain for a majority of the simulation, but the same was not true for the variation in temperature. Indeed, Figure 20 indicates that θ_1^v was used immediately when the simulation started to reduce the V_1 variation of the initial shock. Then, θ_1^v was used again near the corner. Not surprisingly, θ_2^v was used in the same region near the corner and in a similar amount, but was not used anywhere else.

7.3.5 2-D SHOCK WAVE / LAMINAR BOUNDARY LAYER INTERACTION

Shock boundary layer interactions (SBLI) occur in many physical applications that involve transonic, supersonic, and hypersonic flows. The boundary-layer separation that results from a SBLI can lead to adverse effects such as, for example, reduced performance in engine inlets, increased drag on airfoils, and surface heating especially for hypersonic flows [70]. Given that SBLI are a significant source of performance degradation and shocks are usually unavoidable in high speed flows, various techniques have been developed to try and control the negative side effects [71]. Hence, a numerical scheme simulating the compressible Navier-Stokes equations should be robust enough to produce accurate predictions for SBLI problems with high Mach number shocks given their relevance in applications. Furthermore, the computational setup we adapted from [72] results in an eventual steady state. Hence, this test problem not only serves the purpose of testing the shock capturing and positivity preserving capabilities of the proposed scheme, but it also tests the ability of the proposed scheme to converge to a steady state.

We now consider the 2-D case of an oblique shock wave impinging on a flat plate over which a laminar boundary layer is forming. The interaction of the shock with the boundary layer produces separation of the flow and a subsequent recirculation bubble [72]. The flow was originally studied experimentally and numerically in [73]. The particular computational setup we use is from [72] and is shown in Figure 21. The initial conditions consist of a

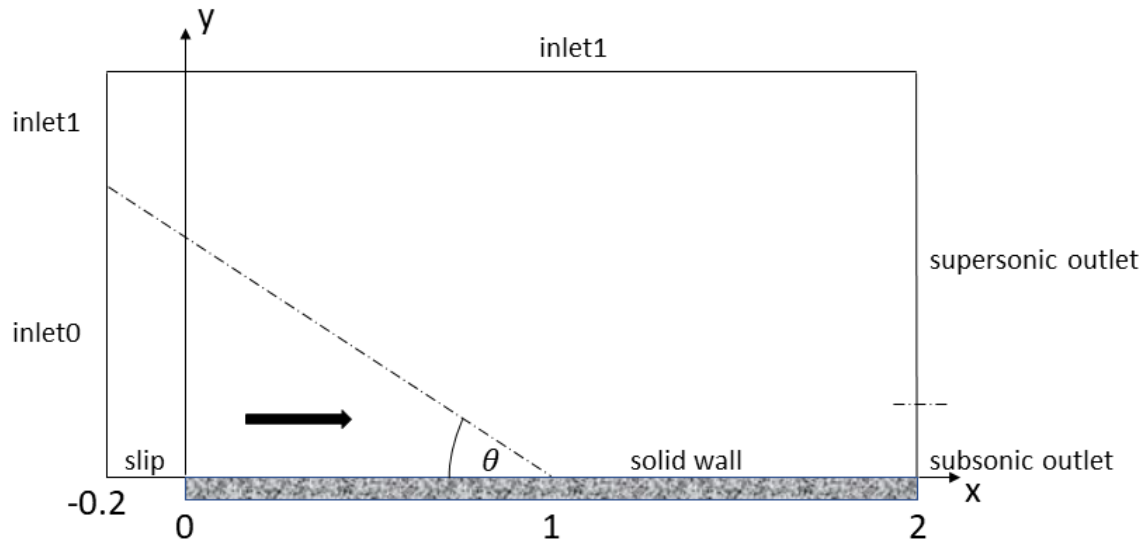


Fig. 21: The computational domain for the shock boundary layer interaction problem. The incident shock makes an angle of θ with the solid wall. Slip wall boundary conditions are used for the boundary $y = 0,\ -0.2 \leq x \leq 0$. For the $Ma = 2.15$ simulations, the maximum y-value of the domain is 1 and for the $Ma = 6.85$ simulations it is 0.45.

uniform state where $\rho = 1$, $T = 1$, and $\vec{\boldsymbol{V}} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$. This initial state is also the supersonic inflow state of inlet0 for the entire simulation. For inlet1, the inflow state is defined so as to satisfy the Rankine-Hugoniot relations through the shock. Notice that θ is the angle between the incident shock wave and the x-axis if the solid wall was an inviscid wall and the shock was reflected at $x = 1$. Thus, the boundary between inlet1 and inlet0 (the diagonal dashed line in Figure 21) is determined by the line $y = (1 - x)\tan(\theta)$. On the right boundary of the domain (see Figure 21), a portion of the outlet boundary is subsonic which can lead to instabilities in the numerical simulation. Hence, for that subsonic region of the right boundary we penalize against a state with a specified constant pressure (see appendix Section B.5.4). Before the shock reflects off the solid wall, we use the average pressure on a subset of $x = 2$ as the constant pressure. After the shock reflects, we use the pressure predicted by the oblique shock wave theory as the constant pressure for the subsonic outlet. The supersonic outlet uses no boundary condition. For all SBLI simulations, we used

Sutherland's law, $Pr = 0.72$, and $Re = 10^5$.

The $Ma = 2.15$ case

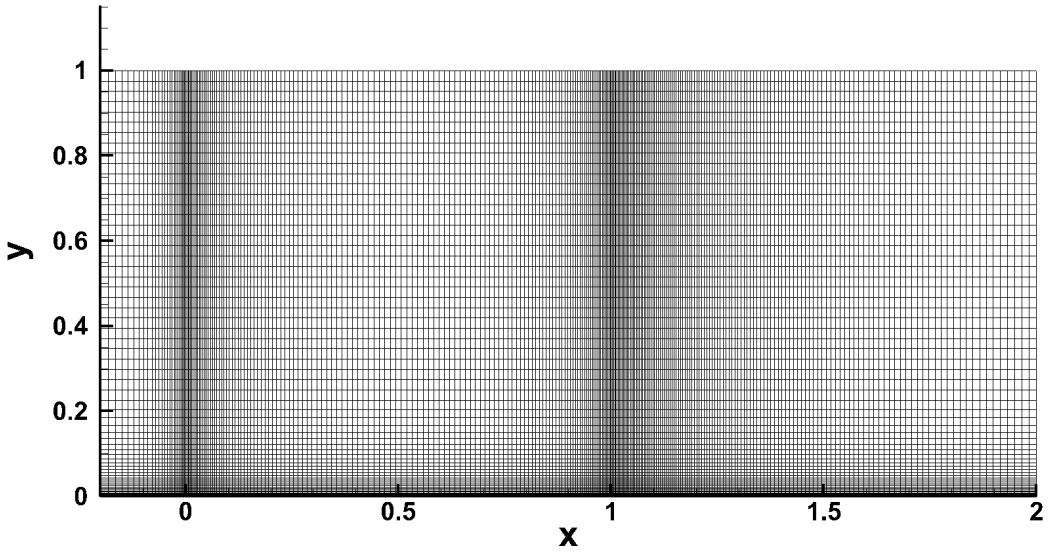
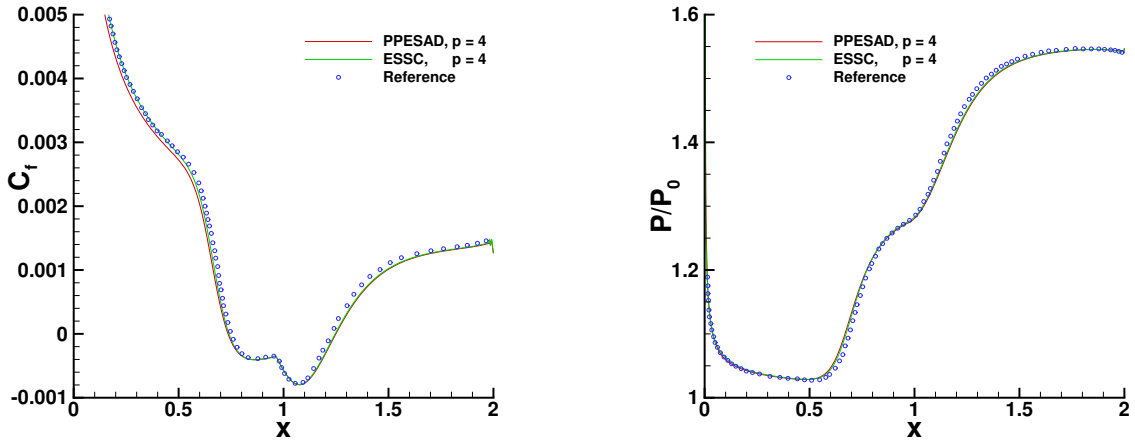


Fig. 22: The computational grid used for the $Ma = 2.15$ SBLI problem. Element edges are displayed.

We begin by considering the case where $Ma = 2.15$ and $\theta = 30.8°$. For this choice of Ma and θ, our setup is identical to that consider in [72]; hence, in Figure 23 we compare skin friction and relative pressure ($P/P_0 = P\gamma Ma^2$) results with those found in [72]. The results in [72] were obtained using a weak-form DG method that adds Godunov-type dissipation at element interfaces with implicit time integration. For $Ma = 2.15$, the shock at the leading edge is weak enough that, with p-restarting, the ESSC scheme can also be used for comparison. We ran all $Ma = 2.15$ simulations on the grid presented in Figure 22 which is comparable to the resolution of the fine grid in [72]. For the grid in Figure 22, the average Δy of the first four elements near the solid wall in the normal direction is 0.0016 and the smallest Δx is 0.0026. As can be seen, the grid is stretched to provide more resolution in the boundary layer and at $x = 0$ and $x = 1$. The grid in Figure 22 used a total of $17,050$



(a) (b)

Fig. 23: Skin friction (left) and relative pressure (right) profiles at the solid wall boundary ($y = 0$) for the $Ma = 2.15$ oblique SBLI problem. The PPESAD-p4 and ESSC-p4 methods used the grid in Figure 22. The reference [72] used $p = 6$ and $K = 11,041$.

elements.

The PPESAD-p4 and ESSC-p4 simulations were run until the elements in the boundary layer reached $\|\hat{\mathbf{U}}_t\|_{L_2,k} \lessapprox 10^{-6}$. For the PPESAD-p4 solution, $\|\hat{\mathbf{U}}_t\|_{L_2} = 6.51\text{e}{-4}$. For the ESSC-p4 solution, $\|\hat{\mathbf{U}}_t\|_{L_2} = 5.30\text{e}{-4}$. Notice that the PPESAD-p4 and ESSC-p4 solutions are nearly indistinguishable for the skin friction and pressure plots at the wall in Figure 23; thus, indicating that PPESAD-p4 does not over-dissipate. The slight variation from the reference solution [72] in Figure 23 may be accounted for by the differences in the grid resolution in the boundary layer.

In Figure 24, we compare the density, relative pressure, and Mach number plots for the ESSC-p4 and PPESAD-p4 solutions. The results are similar which is reasonable given that only high-order elements are used and the artificial dissipation used is small and nonzero in relatively few elements (see Figure 25).

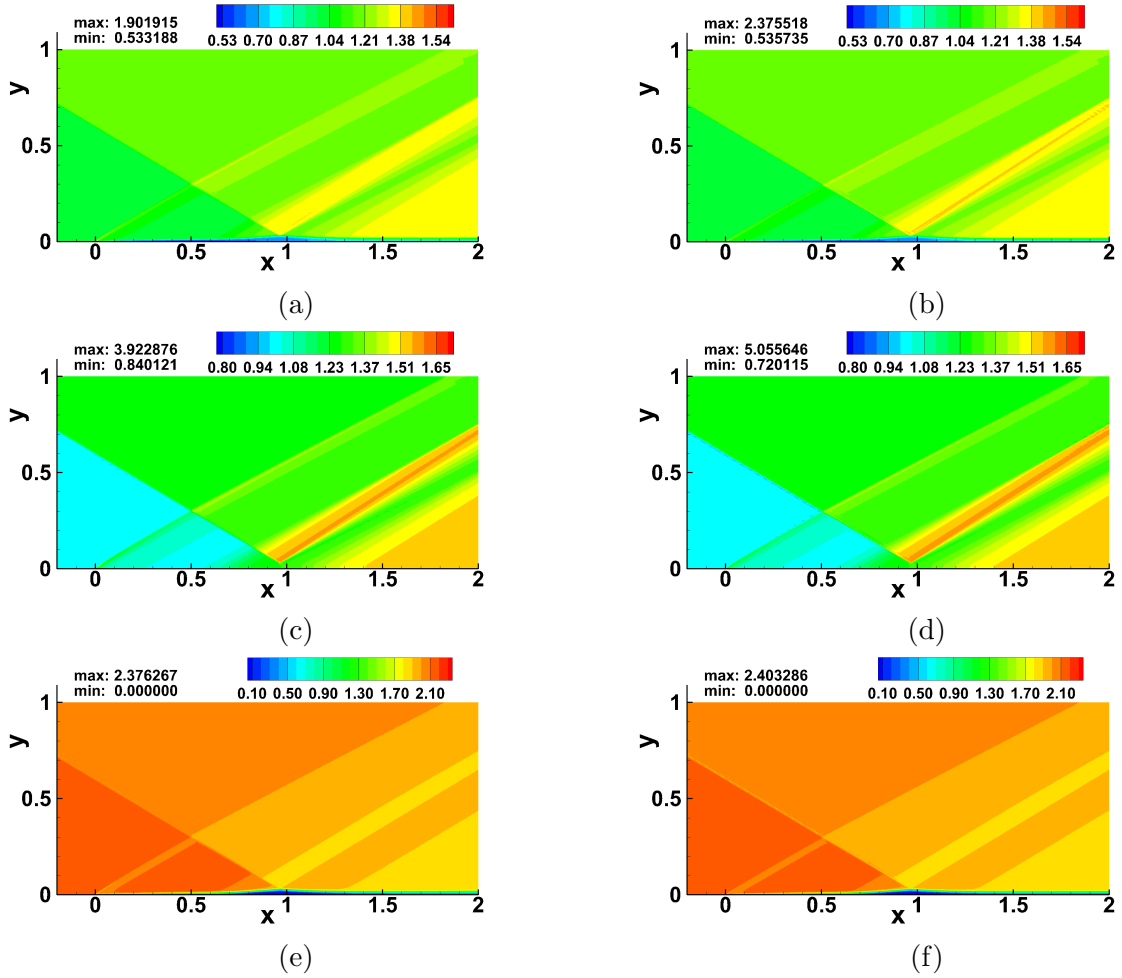


Fig. 24: Density (top row), relative pressure (middle row), and Mach number (bottom row) are shown for the $Ma = 2.15$ oblique SBLI problem. The left column shows the PPESAD-p4 solution. The right column shows the ESSC-p4 solution. Maximum values not visible in the plot occur at the compression corner, $(x, y) = (0, 0)$.

The $Ma = 6.85$ case

Next, we consider the case where $Ma = 6.85$ and $\theta = 11.8°$ which also leads to a steady state. For this case, the ESSC scheme was unable to maintain positivity beyond $p = 2$. Hence, we compare the solution on two different grids for polynomial orders $p = 4$ and $p = 6$. The medium grid used a total of $17,920$ elements and is shown in Figure 26. As can be seen the grid is stretched in the x-direction so as to provide more resolution between $x = 0$ and $x = 1$ where Δx is constant and uniformly equal to ≈ 0.0034. In the y-direction, $\Delta y \approx 0.017$ above $y = 0.2$. Below $y = 0.2$, Δy decreases to an average of $\Delta y \approx 0.0034$

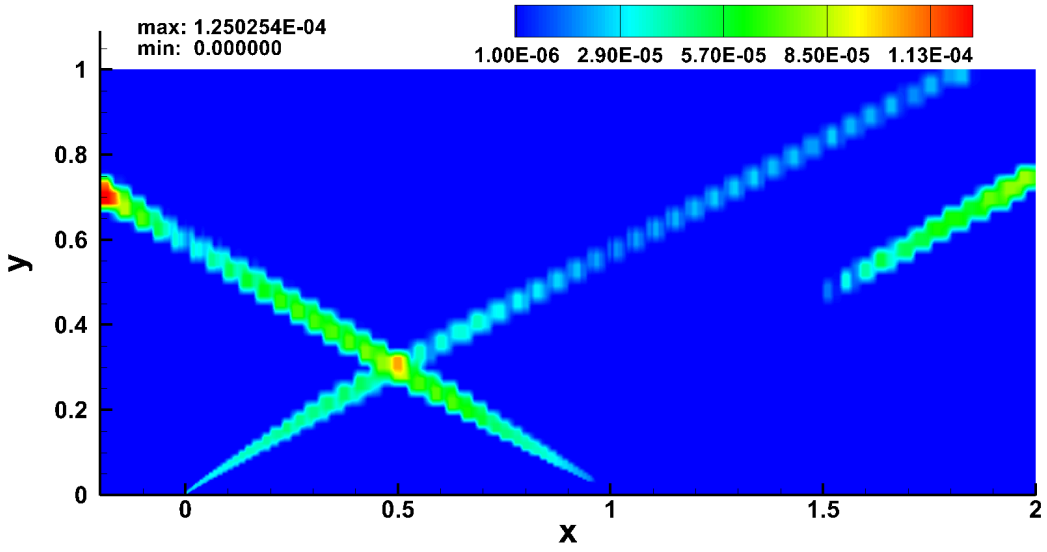


Fig. 25: High-order aritficial viscosity is shown for the PPESAD-p4 solution of the $Ma = 2.15$ SBLI problem. Low-order artificial viscosity was globally zero for the PPESAD-p4 solution.

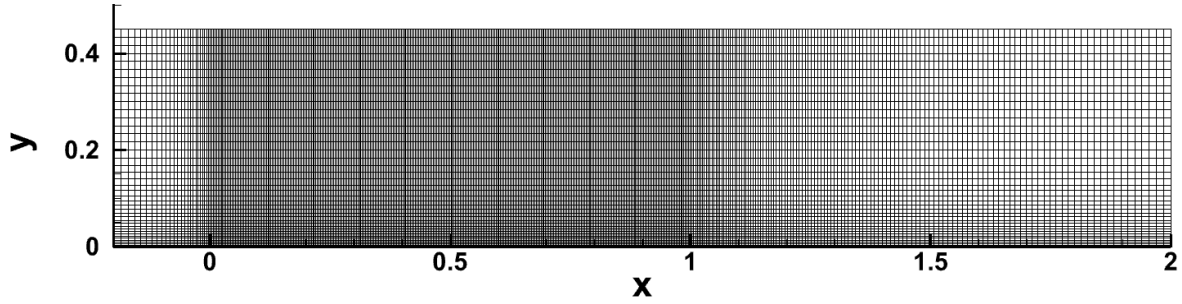
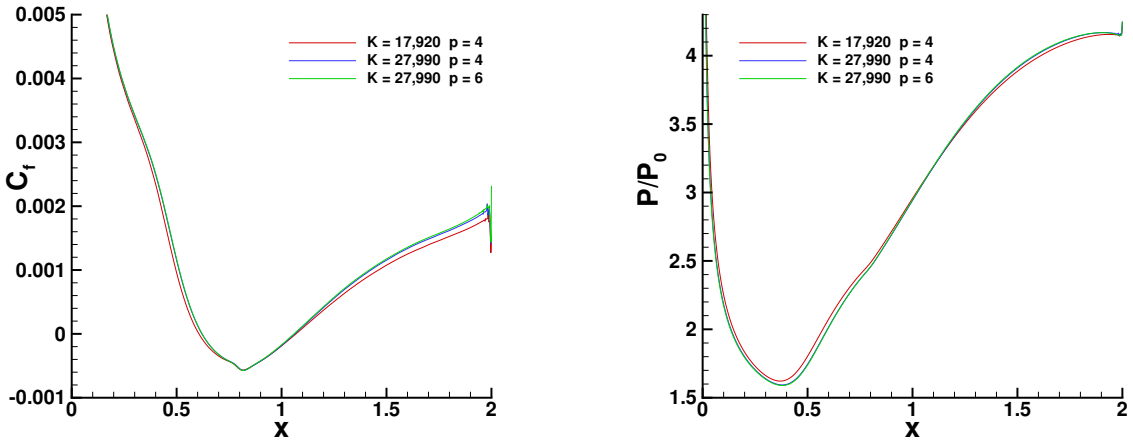


Fig. 26: The medium resolution computational grid used for the $Ma = 6.85$ SBLI problem. Element edges are displayed.

for the 5 elements closest to the wall in the normal direction. The fine grid uses a total of $27,990$ elements and is stretched in the same manner as the medium grid. For $0 \leq x \leq 1$, $\Delta x \approx 0.002$. Above $y = 0.2$, $\Delta y \approx 0.017$. Below $y = 0.2$, Δy decreases to an average of $\Delta y \approx 0.002$ for the 5 elements closest to the wall in the normal direction.

The results from three different simulations are presented in this section. For one simulation, we used the medium grid and $p = 4$. The other two simulations used the fine grid and $p = 4, 6$. All simulations were run until the elements in the boundary layer reached

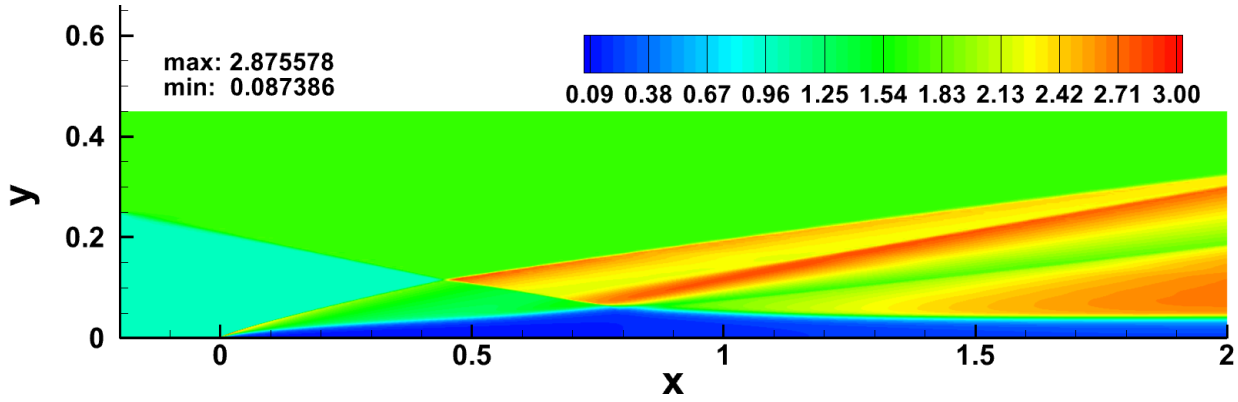


(a) (b)

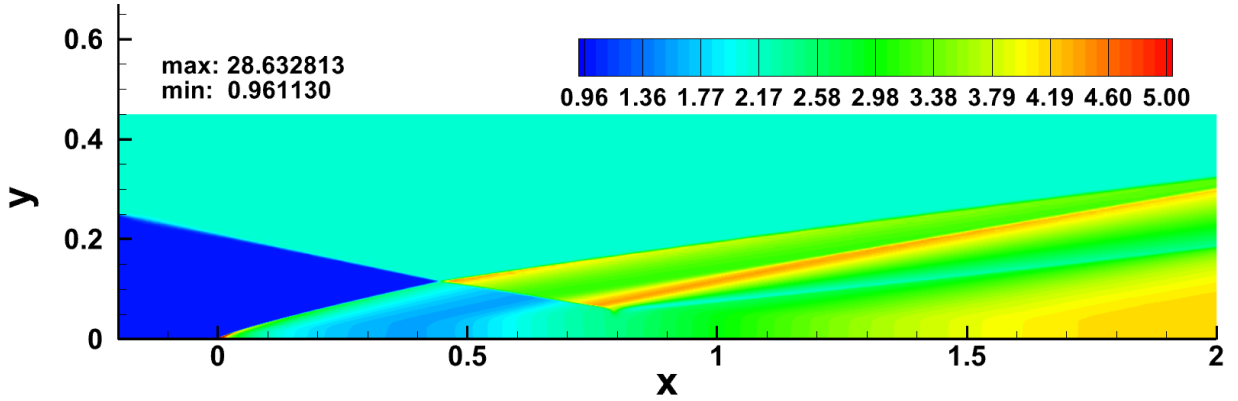
Fig. 27: Skin friction (left) and relative pressure (right) profiles at the solid wall boundary ($y = 0$) for the $Ma = 6.85$ oblique SBLI problem. Solutions were obtained with the PPESAD scheme on the medium ($K = 17,920$) and fine grid for polynomial orders $p = 4$ and $p = 6$.

$\|\hat{\mathbf{U}}_t\|_{L_2,k} \lessapprox 10^{-6}$. For the medium grid simulation, $\|\hat{\mathbf{U}}_t\|_{L_2} = 9.481\text{e}{-4}$. For the fine grid simulation with $p = 4$, $\|\hat{\mathbf{U}}_t\|_{L_2} \approx 0.1$; however, looking at the global contour plot of $\|\hat{\mathbf{U}}_t\|_{L_2,k}$ (see Figure 30) shows that $\|\hat{\mathbf{U}}_t\|_{L_2,k}$ is $\lessapprox 10^{-6}$ globally besides at the compression corner where about 6 elements have $0.1 \lessapprox \|\hat{\mathbf{U}}_t\|_{L_2,k} \lessapprox 20$. The slow convergence of $\|\hat{\mathbf{U}}_t\|_{L_2}$ for this simulation may be partially explained by the the fact that the fine grid changes Δx more rapidly than the medium grid outside of $0 \leq x \leq 1$ (e.g., for the fine grid Δx changes by a factor of 8 in the 8 elements leading up to $x = 0$, but the medium grid only changes by a factor of 1.5). The other explanation is the lack of smoothness of the switches used in the scheme. This latter issue is one we intend to address when generalizing our method to implicit time integration. We also used the PPES-p4 method for the fine grid, but it had the same convergence issue for $\|\hat{\mathbf{U}}_t\|_{L_2}$; hence, we do not think it is the artificial dissipation alone that causes the convergence issue. The fine grid PPESAD-p6 solution was obtained using the PPESAD-p4 fine grid solution as initial conditions and the PPESAD-p6 solution quickly obtained $\|\hat{\mathbf{U}}_t\|_{L_2} = 4.5\text{e}{-5}$.

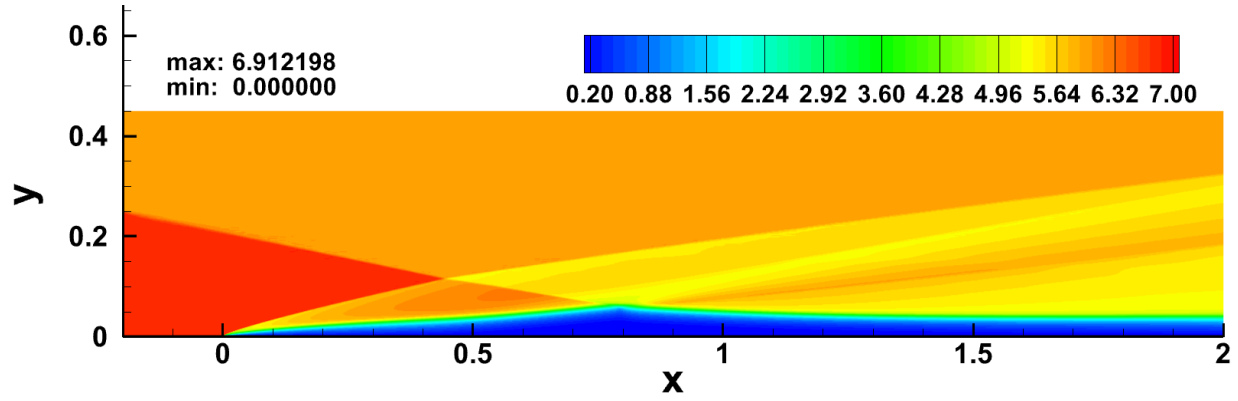
In Figure 27, we compare the skin friction and relative pressure profiles of the three simulations ($P_0 = 1/(\gamma Ma^2)$). Notice that the fine grid and medium grid solutions are close



(a)



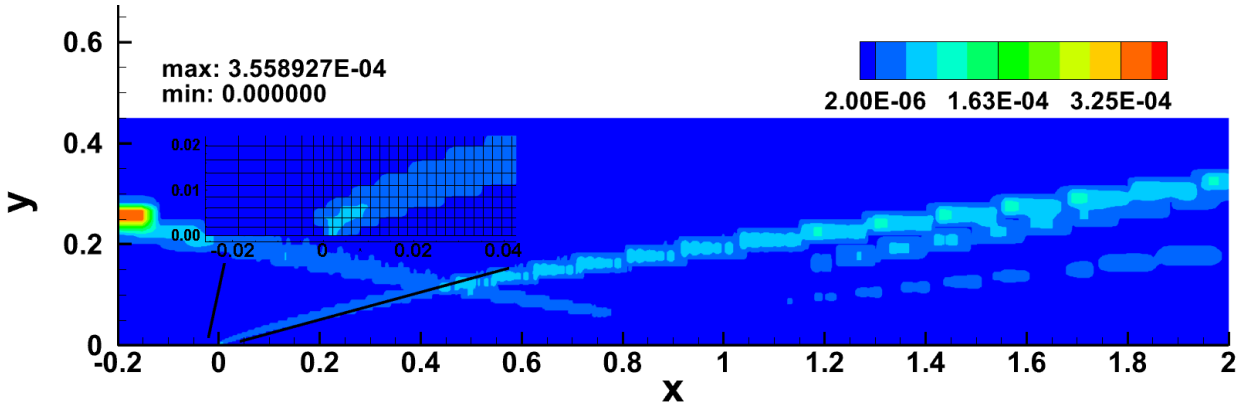
(b)



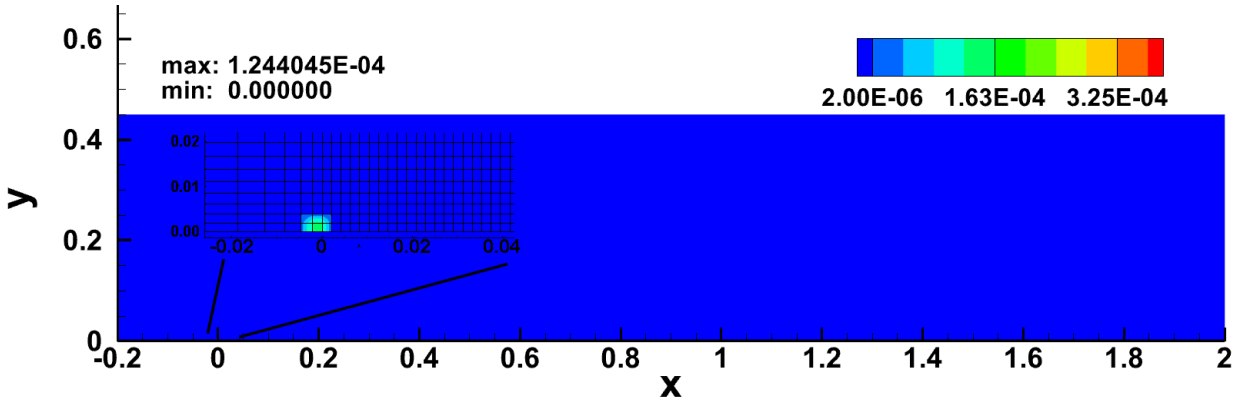
(c)

Fig. 28: Density (top), relative pressure (middle), and Mach number (bottom) of the PPESAD-p6 fine grid solution are shown for the $Ma = 6.85$ oblique SBLI problem. Maximum values not visible in the plot occur at the compression corner, $(x, y) = (0, 0)$.

to each other and the two solutions on the fine grid are nearly identical. Since the PPESAD-p4 and PPESAD-p6 solutions were nearly identical, in Figure 28 we present the density,



(a)



(b)

Fig. 29: High-order (top) and low-order (bottom) artificial viscosity of the PPESAD-p6 fine grid solution are shown for the $Ma = 6.85$ oblique SBLI problem. The low-order viscosity is zero except for in six elements near $(x, y) = (0, 0)$.

relative pressure, and Mach number results only for PPESAD-p6. Notice that the relative pressure near $(x, y) = (0, 0)$ is about six times larger than anywhere else in the domain. Also, notice how the circulation bubble has shifted to the left as compared to the $Ma = 2.15$ results. in Figure 29, we plot the high-order and low-order artificial viscosities. In contrast to the case for $Ma = 2.15$ where the low-order viscosity was globally zero, the low-order artificial viscosity is nonzero when $Ma = 6.85$ but only at the compression corner for six elements. Furthermore, the flux limiter ($0 \leq \theta_f^k \leq 1$ from Eq. (178)) is equal to 1 everywhere except for the single element whose bottom right corner touches $(x, y) = (0, 0)$. For this element, $\theta_f^k = 0.415$. For the high-order artificial viscosity, we see a result similar to the $Ma = 2.15$ simulation where the high-order artificial viscosity for the $Ma = 6.85$ simulation

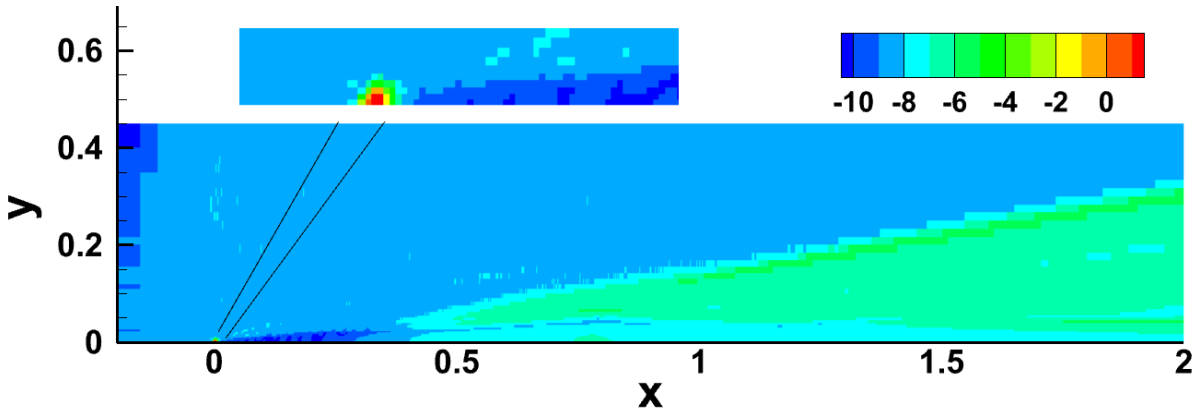


Fig. 30: Contour plot of $\log_{10} \|\hat{\mathbf{U}}_t\|_{L_2,k}$ for the PPESAD-p4 fine grid solution of the $Ma = 6.85$ oblique SBLI problem. For the fine grid PPESAD-p6 solution, the spike in $\log_{10} \|\hat{\mathbf{U}}_t\|_{L_2,k}$ at the corner is not present.

is mostly zero everywhere except for at the shocks. The velocity and temperature limiters of Section 6.2 were never used for this simulation.

7.3.6 2-D HYPERSONIC CYLINDER

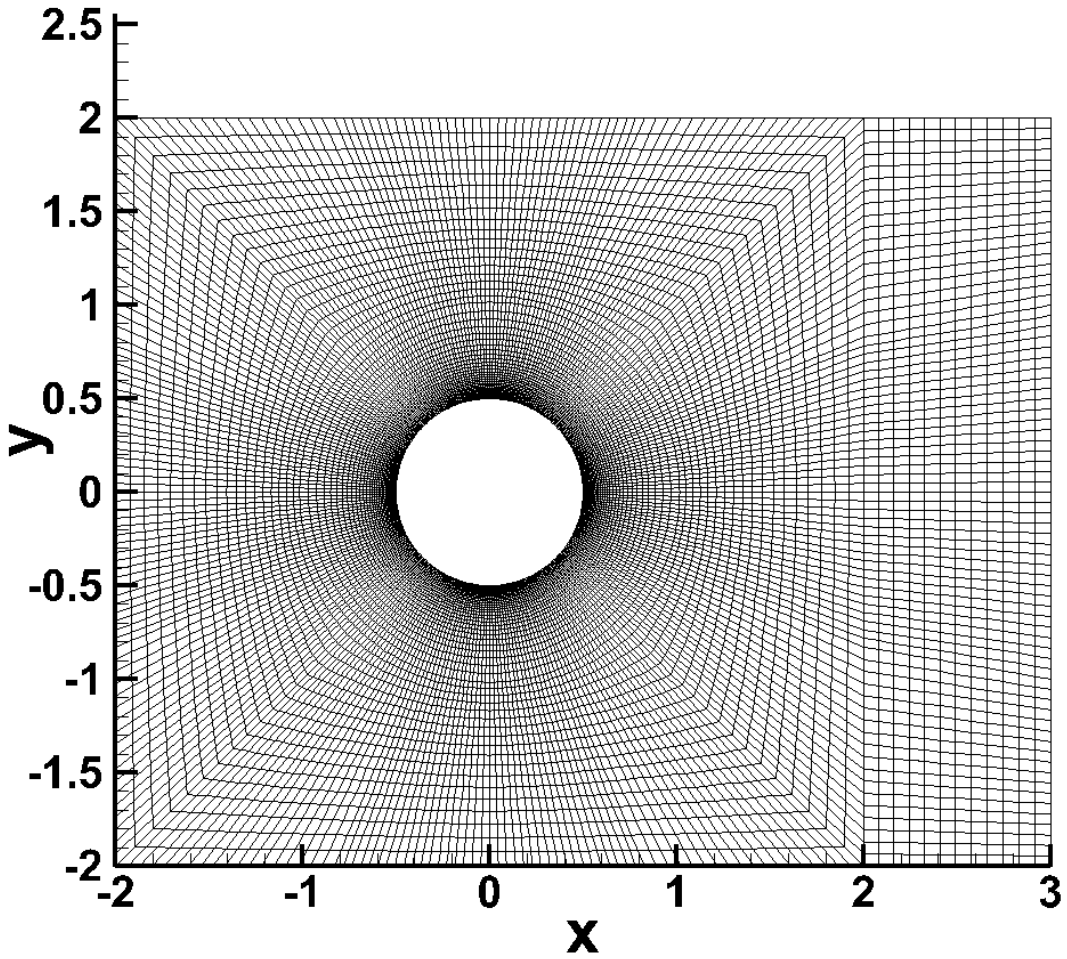
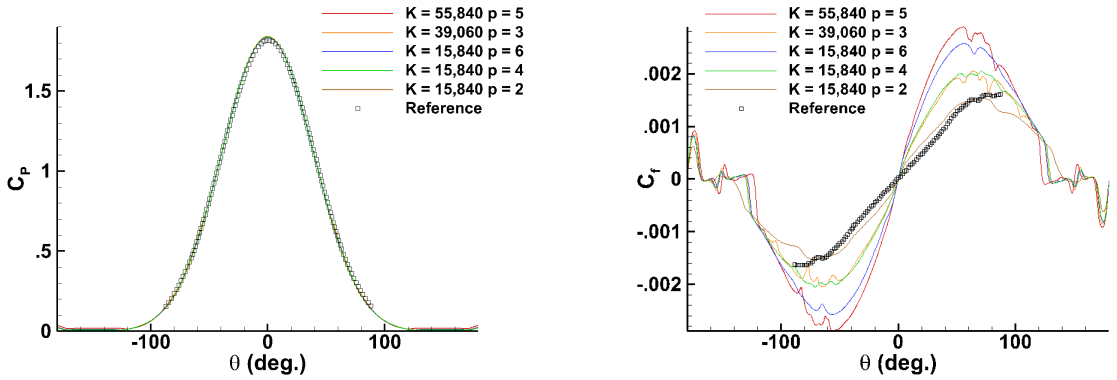


Fig. 31: The coarse grid used for the hypersonic cylinder problem. Element edges are displayed.

In this section, we consider the hypersonic flow around a two-dimensional adiabatic cylinder of diameter 1. We use the same parameters used in [3]: $Re = 376,930$, $Ma = 17.605$, and $Pr = 0.71$. They do not appear to specify if they used Sutherland's law in [3], but based on our numerical results (we ran both cases), we believe that they did not. Hence,

we present only results for the case with no Sutherland's law. In [3], they used a slightly larger domain than we used; however, our domain includes the essential region of the flow and is nearly identical to the part of the domain they include in their figures. See Figure 31 for the domain we used. The initial conditions are a uniform state where $\rho = 1$, $T = 1$, and $\vec{\boldsymbol{V}} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$. The boundary of the cylinder used adiabatic no-slip wall boundary conditions. The left and right boundaries penalized against the initial state. For the left boundary, this is an inflow boundary condition. For the right boundary, this was non-problematic since the flow was supersonic. The top and bottom boundaries used no boundary conditions. We used a coarse, medium, and fine grid for this problem. See Figure 31 for the coarse grid we used. All grids were designed in the following manner. Rectangular elements were used for $2 \leq x \leq 3$. Region 1 consists of all points within 0.008 distance from the wall. Region 1 used elements with two curved edges each of different constant radius with respect to $(x, y) = (0, 0)$. In region 1, Δr was kept constant. For the coarse grid, there are 3 elements in the normal direction in region 1 and hence $\Delta r = 0.008/3$. For the medium grid, $\Delta r = 0.008/4$ and for the fine grid $\Delta r = 0.008/6$ in region 1. The tangential resolution in region 1 is determined by the number of edges radiating from the cylinder boundary. The coarse grid used a total of 288 radial lines, the medium grid used 560, and the fine grid used 720. The grid is then stretched to be coarser closer to the boundaries. Elements more than a distance of 1 away from $(0, 0)$ are no longer curved. The coarse grid used $15,840$ total elements, the medium grid used $39,060$, and the fine grid used $55,260$. We should note that in [3] they used both a larger mesh and only $16,000$ elements total. The authors in [3] used fourth-order hybridized discontinuous Galerkin (HDG) and third-order DIRK(3,3) schemes with a novel form of artificial dissipation. They did not specify the resolution they used in the boundary layer.

All solutions presented were obtained from $(p-1)$-restarting, beginning with $p = 2$. Time averaging windows were chosen independently for each simulation based on the steadiness of the upstream skin friction coefficient. For most $p > 2$ simulations, $t_{\text{final}} - t_{\text{initial}} = 5$ and the time averaging window was about 3/4 of that time. In Figure 32, we compare time-averaged pressure and skin friction coefficient results with those obtained in [3]. Notice that there is only a significant difference for the skin friction plots. Based on our numerical results,



(a) (b)

Fig. 32: Time-averaged pressure (left) and skin friction (right) coefficients obtained with the PPESAD scheme on the no-slip boundary cylinder wall for the hypersonic cylinder simulation. The point $(-0.5, 0)$ corresponds to $\theta = 0$ and positive angles are associated with clockwise rotation from the point $(-0.5, 0)$. The reference solution used a fourth-order HDG method on $16,000$ elements as described in [3].

the skin friction changes significantly with resolution; hence, we estimate that the boundary layer resolution in [3] was comparable to the $p = 2$ coarse grid which is reasonable given that they used a larger grid with only $16,000$ elements. Additionally, it is possible that in [3] more artificial dissipation is added in the boundary region. Notice that the PPESAD-p5 solution in Figure 34 adds no artificial dissipation in the boundary layer. The same was observed for the oblique SBLI problems in Figures 25 and 29. The velocity and temperature limiters discussed in Section 6.2 were not used for the PPEASAD-p5 fine grid simulation.

Contour plots of the density, pressure, vorticity and Mach number are shown for the PPESAD-p5 fine grid solution in Figure 33. The extremum vorticity values are obtained on the cylinder wall. The contour level range is chosen over the smaller range of $[-20, 20]$ so that other features can be observed.

7.3.7 TAYLOR-GREEN VORTEX

We now present numerical results for the viscous, compressible Taylor-Green vortex (TGV) problem at Mach numbers $Ma = 2$ and $Ma = 10$. Often, this test problem is solved at low Mach numbers (e.g., the $Ma = 0.1$ case was considered in [52, 74] and the

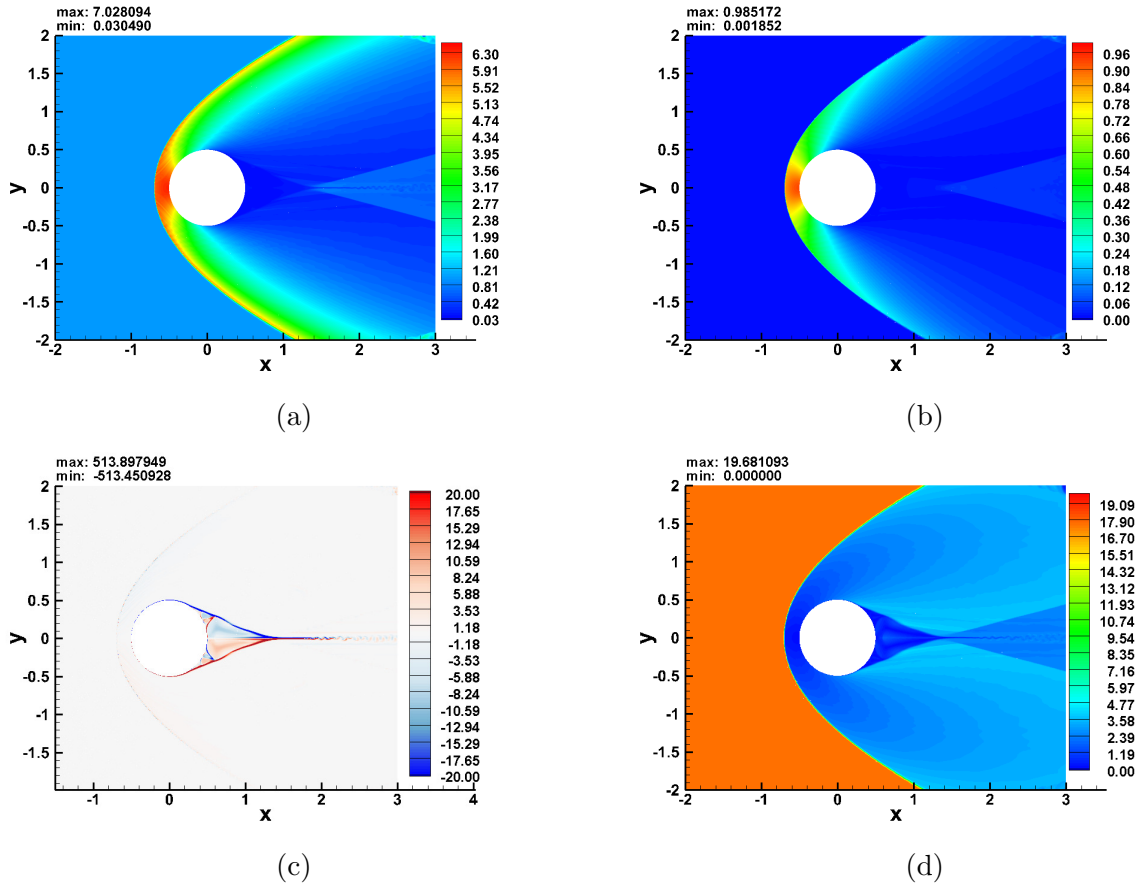
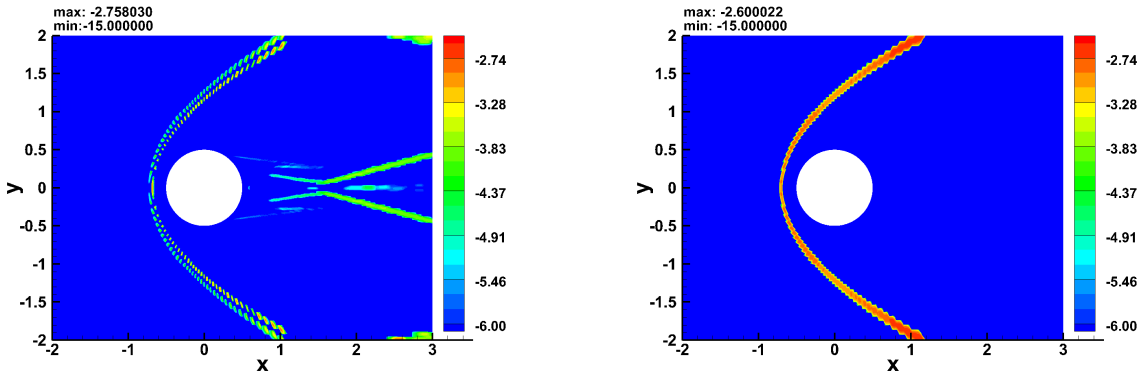


Fig. 33: Density (top left), pressure (top right), vorticity (bottom left), and Mach number (bottom right) are shown for the PPESAD-p5 fine grid solution of the hypersonic cylinder problem.

$Ma = 0.08$ case was considered in [36]) and is used as a test case for comparing how different numerical schemes perform for under-resolved turbulent flows. However, in [75], simulations were performed for Mach numbers in the range $Ma = 0.5$ to $Ma = 2$. We adopt the settings used in [75] and compare our results for the $Ma = 2$ case. The settings are: $Re = 400$, $Pr = 0.7$, and Sutherland's law is used. The problem is solved on the periodic box defined



(a) (b)

Fig. 34: High-order (left) and low-order (right) artificial viscosity ($\log_{10}$) of the PPESAD-p5 fine grid solution are shown for the hypersonic cylinder problem.

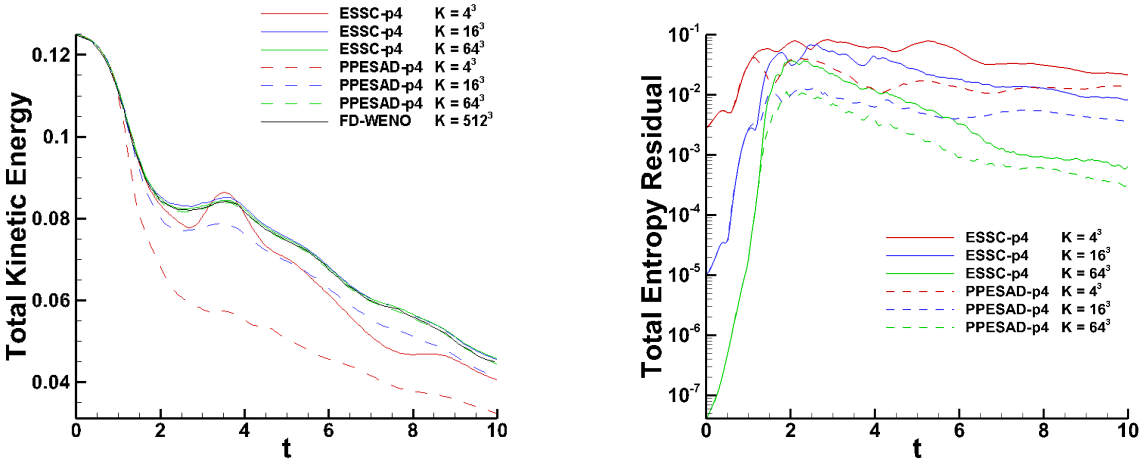
by $0 \leq x, y, z \leq 2\pi$ with the following initial conditions

$$\begin{aligned} \rho &= 1 + \frac{1}{16}(\cos 2x + \cos 2y)(\cos 2z + 2), \\ V_1 &= \sin x \cos y \cos z, \\ V_2 &= -\cos x \sin y \cos z, \\ V_3 &= 0, \\ T &= 1. \end{aligned} \tag{208}$$

Uniform, rectangular grids with $\Delta x = \Delta y = \Delta z$ are used for all test cases; hence, we refer to the grid with 8 total elements such that $\Delta x = \pi$ for all elements as the "2^3 grid", for example.

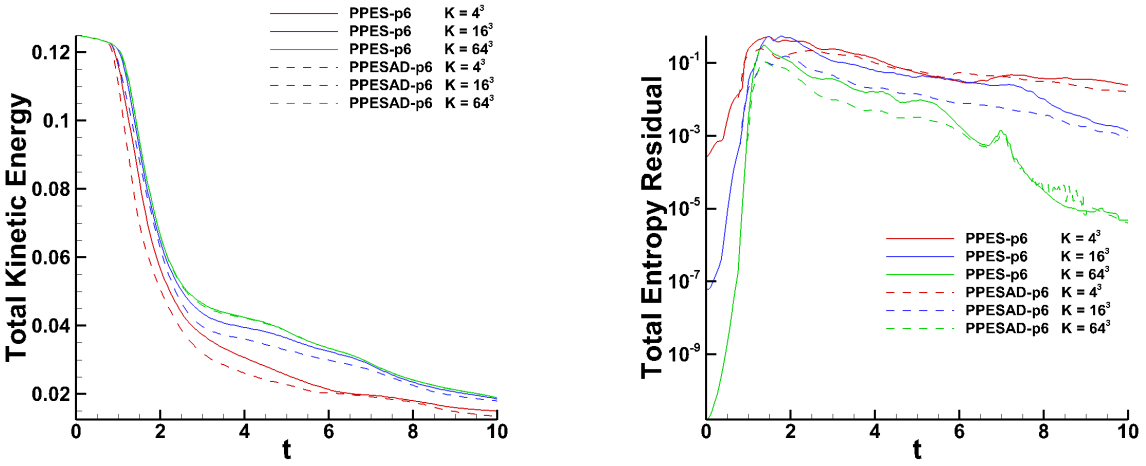
The $Ma = 2$ case

In [75], they used a hybrid compact eighth-order finite difference and seventh-order weighted essentially non-oscillatory (FD-WENO) scheme with hyperviscosity for uniform grids 128^3, 256^3, and 512^3. In Figure 35, we compare the temporal evolution of the total kinetic energy for the ESSC-p4 and PPESAD-p4 solutions to the results obtained in [75]. The 128^3, 256^3, and 512^3 results for the time series plot of the total kinetic energy in [75]



(a) (b)

Fig. 35: Time series plot of the total kinetic energy (left) and total entropy residual (right) for the ESSC-p4 and PPESAD-p4 solutions of the $Ma = 2$ TGV problem on grids 4^3, 16^3 and 64^3. The FD-WENO reference solution is from [75].



(a) (b)

Fig. 36: Time series plot of the total kinetic energy (left) and total entropy residual (right) for the PPES-p6 and PPESAD-p6 solutions of the $Ma = 10$ TGV problem on grids 4^3, 16^3 and 64^3.

were indistinguishable. As can be seen, the 4^3 and 16^3 PPESAD-p4 solutions dissipate the total kinetic energy significantly more than their ESSC-p4 counterparts; however, this does

not imply that the ESSC-p4 solution is overall more accurate. To see this, we look at Figure 37 where we see that for density and pressure the ESSC-p4 solution on 16^3 contains large overshoots that aren't present in the 64^3 solution; furthermore, it is clear from Figure 37 that the ESSC-p4 solutions on 16^3 and 64^3 for pressure, density, and V_1 possess non-physical oscillations. The PPESAD-p4 solution, on the other hand, recognizes that the resolution is insufficient for the coarse grids and adds artificial dissipation to maintain a smooth, non-oscillatory solution. The PPESAD-p4 solution recognizes the lack of sufficient resolution via the entropy residual ($\mathbf{R}$, Eq. (81)) and the residual-based sensor (Sn, Eq. (85)), then adds artificial dissipation to reduce the entropy residual. As can be seen in Figure 35, for a given grid the PPESAD-p4 scheme keeps the total entropy residual in the domain lower than the ESSC-p4 scheme does.

The $Ma = 10$ case

For $Ma \gtrapprox 3$, the ESSC scheme fails to preserve positivity (depending on the polynomial order) for the viscous TGV problem; hence, we compare the PPESAD and PPES $Ma = 10$ solutions. In Figure 36, we see that the decay rate of the total kinetic energy for the two methods for the 4^3 and 16^3 grids is fairly similar, but the PPES method is less dissipative. As we would expect, this corresponds to the total entropy residual being typically larger for the PPES method on those same grids. For the more resolved 64^3 grid, the total kinetic energy decay rates are nearly identical. In Figure 38, we see that both the PPES and PPESAD solutions appear to be free of spurious oscillations.

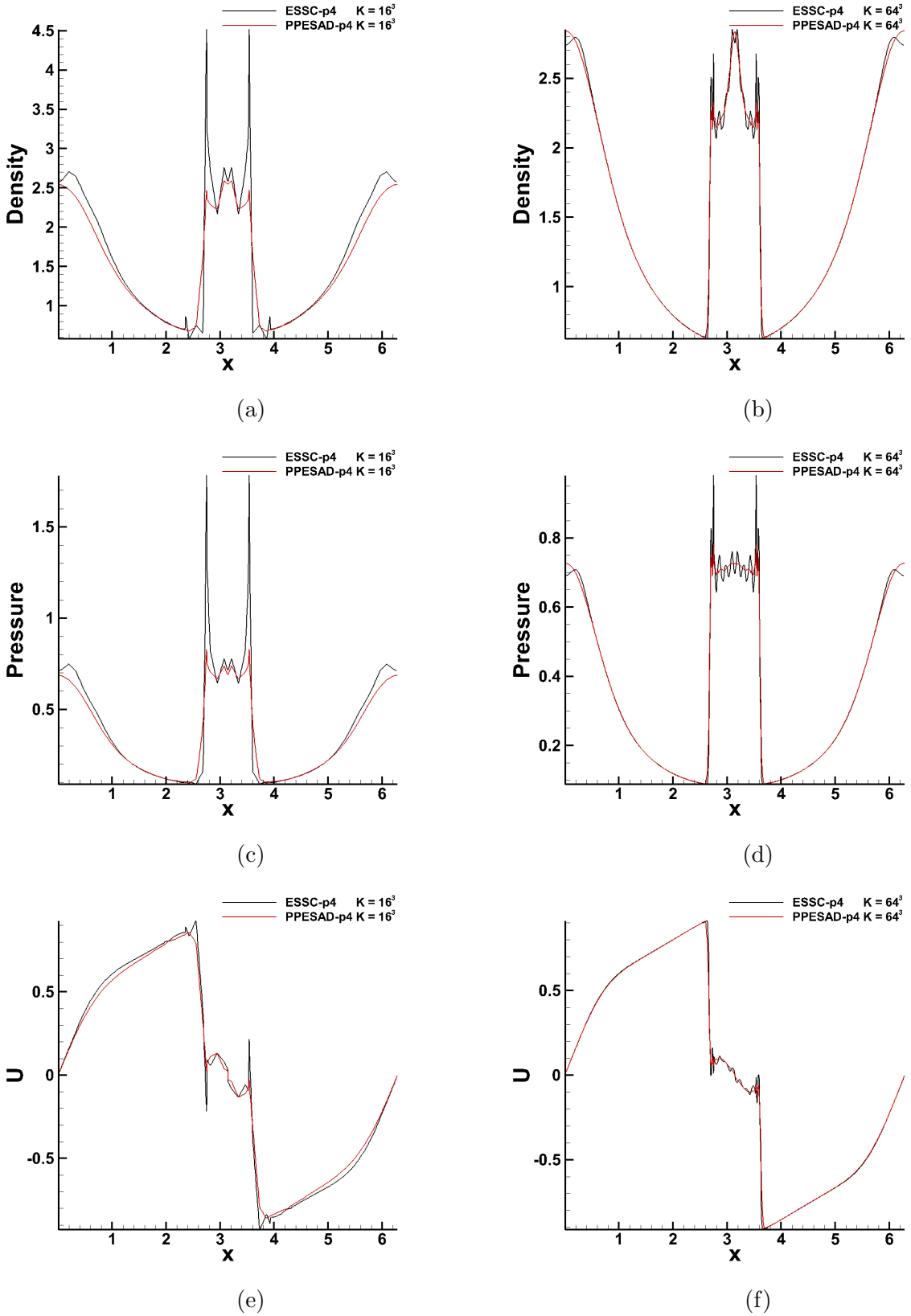


Fig. 37: Density (top row), pressure (middle row), and velocity component $V_1 = U$ (bottom row) are plotted for the PPESAD-p4 and ESSC-p4 solutions of the $Ma = 2$ TGV problem on the 16^3 (left column) and 64^3 (right column) grids. Data is obtained at time $t = 2.5$ from the line intersected by the planes $y = \pi$ and $z = \pi$.

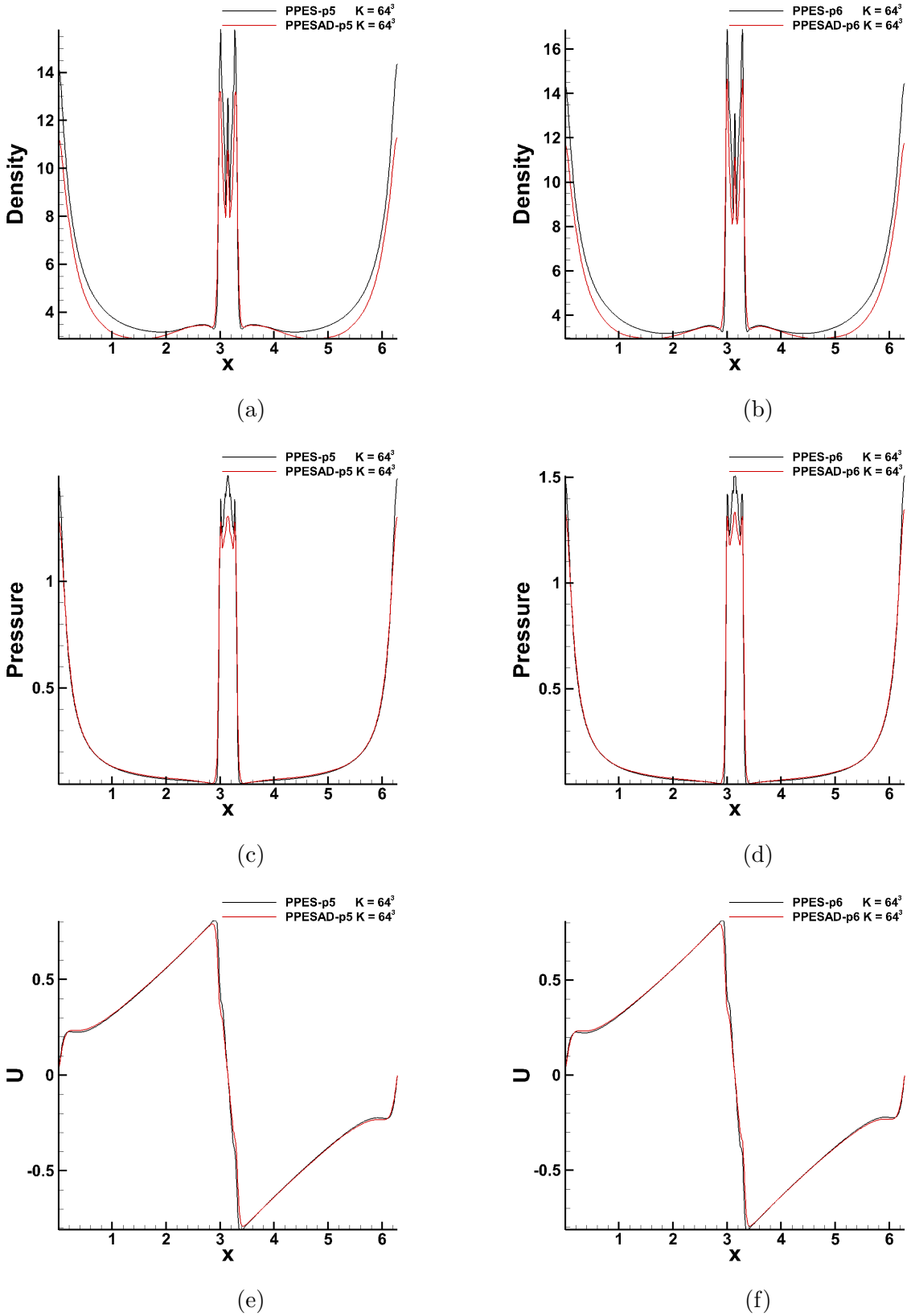


Fig. 38: Density (top row), pressure (middle row), and velocity component $V_1 = U$ (bottom row) are plotted for the $p = 5$ (left column) and $p = 6$ (right column) PPESAD and PPES solutions of the $Ma = 10$ TGV problem on the 64^3 grid. Data is obtained at time $t = 2.5$ from the line intersected by the planes $y = \pi$ and $z = \pi$.

CHAPTER 8

SUMMARY AND CONCLUSIONS

We have constructed a new class of positivity-preserving, entropy stable, spectral collocation schemes of arbitrary order of accuracy for the 3-D compressible Navier–Stokes equations on unstructured curvilinear grids. To our knowledge, the proposed spectral collocation methods are the first class of high-order schemes that provide both the pointwise positivity preservation of thermodynamic variables and entropy stability property for the compressible Navier-Stokes equations. In Chapter 7, the robustness of our method was tested on several problems for which maintaining positivity is extremely non-trivial including the viscous Leblanc problem and the diffraction of a viscous shock of Mach number 200. Furthermore, we demonstrated through the 3-D viscous shock and Taylor-Green vortex problems that our new method can increase the accuracy of a high-order scheme for under-resolved simulations and does not destroy accuracy for sufficiently resolved simulations. We also demonstrated the robustness of the method by simulating problems with steady state solutions (shock boundary layer interaction problem) and problems with sharp features on curvilinear grids (hypersonic cylinder problem).

Pivotal to the success of our proposed method is the residual-based sensor developed in Chapter 5. This sensor plays many key roles in our scheme: it is the first step in screening for under-resolved and discontinuous features, it scales the artificial viscosity, it controls the flux-limiting on troubled elements, and it is one of the quantities used to determine when the entropy stable velocity and temperature limiters for viscous flows can be used. Without the residual-based sensor to tell the scheme when regularization is not necessary, we would have certainly over-dissipated the 3-D viscous shock problem on sufficiently resolved grids. Furthermore, the residual-based sensor is relatively cheap to obtain as compared to constructing a host of physics-based sensors; hence, by being the first step in screening for under-resolved features the residual-based sensor serves to save computational time.

The artificial viscosity coefficient constructed in Chapter 5 also involves several non-trivial choices. While physics-based sensors can be useful for building an artificial viscosity,

we chose to rely on them sparingly for constructing the artificial viscosity coefficient. The physics-based sensors can only change the amount of dissipation within a range, but they do not decide when dissipation is used. Additionally, we decided to construct the dissipation in proportion to the regularity of the velocity and pressure fields instead of using more common choices such as the max eigenvalue. This choice ensures that the artificial viscosity is smaller for smooth regions and non-shock discontinuities, e.g. contact discontinuities. Lastly, the choice to scale part of the artificial viscosity by the mach number led to a significant reduction in the amount of dissipation added at solid walls and consequently more accurate results for the shock boundary layer interaction problem.

The first-order positivity-preserving scheme developed in Chapter 6 relied on several new contributions. While it is somewhat intuitive that mixing the first-order and high-order inviscid terms can stabilize the numerical scheme at sharp features, developing first-order inviscid terms that have the same element-wise entropy contribution and preserve freestream for general curvilinear grids is not as straightforward, but highly necessary for building a robust scheme. Hence, we consider Lemma 2 a significant contribution of this work. In our numerical tests, Rusanov-type fluxes performed poorly as compared to the Merriam–Roe flux. Yet, finding a Merriam–Roe flux for which the density contribution was as well behaved as in Eq. (119) required substantial work. Hence, we consider Eq. (119) a significant contribution. Although density positivity for first-order schemes is certainly not new, we believe that Theorem 6 presents a sufficiently general and sharp requirement for developing density positivity-preserving schemes. Furthermore, the two-point matrix ν_w in Lemma 107 greatly simplifies the process of moving between the primitive form of a proposed two-point flux and the entropy variable form that is useful for proving entropy stability.

The discretely entropy stable velocity and temperature limiters presented in Section 6.2 are essential for viscous simulations at high mach numbers when only high-order viscous terms are used. Using the necessary temperature positivity time step restriction given in Section 6.1.6 and beginning a viscous simulation with the initial conditions of the Blastwave, Leblanc or the Mach 200 shock diffraction problem immediately requires time steps smaller than 10^{-14} and the situation does not significantly improve with first-order dissipation and inviscid terms. However, the velocity and temperature limiters quickly act to reduce the

strictness of the temperature positivity constraint. Indeed, the ability to enforce any velocity and temperature variation constraint in a conservative and discretely entropy stable manner over any set of points (we used elements) without changing the density field is quite powerful and we anticipate that these new limiters may have additional utility beyond what we used them for.

In Section 6.3, we present a flux-limiting method for combining a high-order positivity-violating entropy stable spectral collocation scheme and a first-order positivity-preserving entropy stable scheme defined on the same collocation points used for the high-order counterpart. The positivity preservation and entropy stability properties are obtained by introducing the low- and high-order artificial dissipation operators that mimic the corresponding diffusion operators of the Brenner-Navier-Stokes equations. Since both schemes are defined on the same set of collocation points, no interpolation is required between high- and low-order elements. The low- and high-order schemes are coupled by using the flux limiter that preserves the conservation, positivity preservation, and entropy stability properties, thus facilitating the rigorous L_2-stability proof for the symmetric form of the discretized 3-D compressible Navier-Stokes equations on curvilinear grids. An additional attractive property of the proposed class of schemes is that the 1st-order artificial dissipation is only added in troubled elements where the density or temperature becomes negative or the shock strength exceeds the user-defined threshold, while in the rest of the computational domain the high-order entropy stable scheme is used. Our numerical experiment show that the new flux-limiting schemes demonstrate the high-order error convergence for smooth solutions and provide the positivity of thermodynamic variables and excellent shock-capturing capabilities for discontinuous flows.

While there are certainly still many roadblocks to overcome in developing next-generation high-order numerical algorithms for LES and DNS, we believe that we have developed significantly general tools that can be used to stabilize and preserve positivity properties for other high-order algorithms and may perhaps inspire future, improved methods for more general settings.

REFERENCES

[1] F. D. Witherden and A. Jameson, "Future directions in computational fluid dynamics," in *23rd AIAA Computational Fluid Dynamics Conference*, 2017, p. 3791.

[2] Z. J. Wang, K. Fidkowski, R. Abgrall, F. Bassi, D. Caraeni, A. Cary, H. Deconinck, R. Hartmann, K. Hillewaert, H. T. Huynh, *et al.*, "High-order CFD methods: Current status and perspective," *International Journal for Numerical Methods in Fluids*, vol. 72, no. 8, pp. 811–845, 2013.

[3] P. Fernandez, N.-C. Nguyen, and J. Peraire, "A physics-based shock capturing method for large-eddy simulation," *arXiv:1806.06449*, 2018.

[4] J. Slotnick, A. Khodadoust, J. Alonso, D. Darmofal, W. Gropp, E. Lurie, and D. Mavriplis, "CFD vision 2030 study: A path to revolutionary computational aerosciences," National Aeronautics and Space Administration, Langley Research Center, Tech. Rep., 2014.

[5] L. Dalcin, D. Rojas, S. Zampini, D. C. D. R. Fernández, M. H. Carpenter, and M. Parsani, "Conservative and entropy stable solid wall boundary conditions for the compressible Navier–Stokes equations: Adiabatic wall and heat entropy transfer," *Journal of Computational Physics*, vol. 397, p. 108 775, 2019.

[6] J. S. Hesthaven, *Numerical methods for conservation laws: From analysis to algorithms*. SIAM, 2017.

[7] J. VonNeumann and R. D. Richtmyer, "A method for the numerical calculation of hydrodynamic shocks," *Journal of Applied Physics*, vol. 21, no. 3, pp. 232–237, 1950.

[8] C. Nguyen and J. Peraire, "An adaptive shock-capturing HDG method for compressible flows," in *20th AIAA Computational Fluid Dynamics Conference*, 2011, p. 3060.

[9] D. Moro, N. C. Nguyen, and J. Peraire, "Dilation-based shock capturing for high-order methods," *International Journal for Numerical Methods in Fluids*, vol. 82, no. 7, pp. 398–416, 2016.

[10] G. E. Barter and D. L. Darmofal, "Shock capturing with PDE-based artificial viscosity for DGFEM: Part I. Formulation," *Journal of Computational Physics*, vol. 229, no. 5, pp. 1810–1827, 2010.

[11] J. Upperman and N. K. Yamaleev, "Entropy stable artificial dissipation based on Brenner regularization of the Navier-Stokes equations," *Journal of Computational Physics*, vol. 393, pp. 74–91, 2019.

[12] ——, "Positivity-preserving entropy stable spectral collocation schemes for the 1-D compressible Navier-Stokes equations," submitted to Journal of Computational Physics.

[13] A. W. Cook and W. H. Cabot, "Hyperviscosity for shock-turbulence interactions," *Journal of Computational Physics*, vol. 203, no. 2, pp. 379–385, 2005.

[14] S. Kawai and S. K. Lele, "Localized artificial diffusivity scheme for discontinuity capturing on curvilinear meshes," *Journal of Computational Physics*, vol. 227, no. 22, pp. 9498–9526, 2008.

[15] S. Kawai, S. K. Shankar, and S. K. Lele, "Assessment of localized artificial diffusivity scheme for large-eddy simulation of compressible turbulent flows," *Journal of Computational Physics*, vol. 229, no. 5, pp. 1739–1762, 2010.

[16] V. Zingan, J.-L. Guermond, J. Morel, and B. Popov, "Implementation of the entropy viscosity method with the discontinuous Galerkin method," *Computer Methods in Applied Mechanics and Engineering*, vol. 253, pp. 479–490, 2013.

[17] J.-L. Guermond, R. Pasquetti, and B. Popov, "Entropy viscosity method for nonlinear conservation laws," *Journal of Computational Physics*, vol. 230, no. 11, pp. 4248–4267, 2011.

[18] J. S. Hesthaven and T. Warburton, *Nodal discontinuous Galerkin methods: algorithms, analysis, and applications*. Springer Science & Business Media, 2007.

[19] M. Sonntag and C.-D. Munz, "Efficient parallelization of a shock capturing for discontinuous Galerkin methods using finite volume sub-cells," *Journal of Scientific Computing*, vol. 70, no. 3, pp. 1262–1289, 2017.

[20] H. Luo, J. D. Baum, and R. Löhner, "A Hermite WENO-based limiter for discontinuous Galerkin method on unstructured grids," *Journal of Computational Physics*, vol. 225, no. 1, pp. 686–713, 2007.

[21] B. Cockburn and C.-W. Shu, "The Runge–Kutta discontinuous Galerkin method for conservation laws V: Multidimensional systems," *Journal of Computational Physics*, vol. 141, no. 2, pp. 199–224, 1998.

[22] R. Hartmann, "Higher-order and adaptive discontinuous Galerkin methods with shock-capturing applied to transonic turbulent delta wing flow," *International Journal for Numerical Methods in Fluids*, vol. 72, no. 8, pp. 883–894, 2013.

[23] H. Abbassi, F. Mashayek, and G. B. Jacobs, "Shock capturing with entropy-based artificial viscosity for staggered grid discontinuous spectral element method," *Computers & Fluids*, vol. 98, pp. 152–163, 2014.

[24] M. Svärd, "Weak solutions and convergent numerical schemes of modified compressible Navier–Stokes equations," *Journal of Computational Physics*, vol. 288, pp. 19–51, 2015.

[25] C. M. Dafermos and C. M. Dafermos, *Hyperbolic conservation laws in continuum physics*. Springer, 2005, vol. 3.

[26] M. L. Merriam, "An entropy-based approach to nonlinear stability," National Aeronautics and Space Administration, Ames Research Center, Tech. Rep., 1989.

[27] P. Dutt, "Stable boundary conditions and difference schemes for Navier–Stokes equations," *SIAM Journal on Numerical Analysis*, vol. 25, no. 2, pp. 245–267, 1988.

[28] T. J. Hughes, L. P. Franca, and M. Mallet, "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," *Computer Methods in Applied Mechanics and Engineering*, vol. 54, no. 2, pp. 223–234, 1986.

[29] E. Tadmor, "The numerical viscosity of entropy stable schemes for systems of conservation laws. I," *Mathematics of Computation*, vol. 49, no. 179, pp. 91–103, 1987.

[30] ——, "Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems," *Acta Numerica*, vol. 12, no. 1, pp. 451–512, 2003.

[31] M. Parsani, M. H. Carpenter, and E. J. Nielsen, "Entropy stable discontinuous interfaces coupling for the three-dimensional compressible Navier-Stokes equations," *Journal of Computational Physics*, vol. 290, no. 1, pp. 132–138, 2015.

[32] A. R. Winters, D. A. Kopriva, G. J. Gassner, and F. Hindenlang, "Construction of modern robust nodal discontinuous Galerkin spectral element methods for the compressible Navier–Stokes equations," in *Efficient High-Order Discretizations for Computational Fluid Dynamics*, Springer, 2021, pp. 117–196.

[33] N. K. Yamaleev and M. H. Carpenter, "A family of fourth-order entropy stable nonoscillatory spectral collocation schemes for the 1-D Navier–Stokes equations," *Journal of Computational Physics*, vol. 331, pp. 90–107, 2017.

[34] N. K. Yamaleev, D. C. D. R. Fernandez, J. Lou, and M. H. Carpenter, "Entropy stable spectral collocation schemes for the 3-D Navier-Stokes equations on dynamic unstructured grids," *Journal of Computational Physics*, vol. 399, p. 108 897, 2019.

[35] P. Chandrashekar, "Kinetic energy preserving and entropy stable finite volume schemes for compressible Euler and Navier-Stokes equations," *Communications in Computational Physics*, vol. 14, no. 5, pp. 1252–1286, 2013.

[36] M. Carpenter, T. Fisher, E. Nielsen, M. Parsani, M Svärd, and N Yamaleev, "Entropy stable summation-by-parts formulations for compressible computational fluid dynamics," in *Handbook of Numerical Analysis*, vol. 17, Elsevier, 2016, pp. 495–524.

[37] M. H. Carpenter, T. C. Fisher, E. J. Nielsen, and S. H. Frankel, "Entropy stable spectral collocation schemes for the Navier–Stokes equations: Discontinuous interfaces," *SIAM Journal on Scientific Computing*, vol. 36, no. 5, B835–B867, 2014.

[38] T. C. Fisher and M. H. Carpenter, "High-order entropy stable finite difference schemes for nonlinear conservation laws: Finite domains," *Journal of Computational Physics*, vol. 252, pp. 518–557, 2013.

[39] M. H. Carpenter and T. C. Fisher, "High-order entropy stable formulations for computational fluid dynamics," in *21st AIAA Computational Fluid Dynamics Conference*, 2013, p. 2868.

[40] L. Friedrich, A. R. Winters, D. C. D. R. Fernández, G. J. Gassner, M. Parsani, and M. H. Carpenter, "An entropy stable h/p non-conforming discontinuous Galerkin method with the summation-by-parts property," *Journal of Scientific Computing*, vol. 77, no. 2, pp. 689–725, 2018.

[41] J. Chan, "On discretely entropy conservative and entropy stable discontinuous Galerkin methods," *Journal of Computational Physics*, vol. 362, pp. 346–374, 2018.

[42] H. Ranocha, L. Dalcin, and M. Parsani, "Fully discrete explicit locally entropy-stable schemes for the compressible Euler and Navier–Stokes equations," *Computers & Mathematics with Applications*, vol. 80, no. 5, pp. 1343–1359, 2020.

[43] F. Ismail and P. L. Roe, "Affordable, entropy-consistent Euler flux functions II: Entropy production at shocks," *Journal of Computational Physics*, vol. 228, no. 15, pp. 5410–5436, 2009.

[44] H. Ranocha, "Comparison of some entropy conservative numerical fluxes for the Euler equations," *Journal of Scientific Computing*, vol. 76, no. 1, pp. 216–242, 2018.

[45] M. Svärd, "A convergent numerical scheme for the compressible Navier–Stokes equations," *SIAM Journal on Numerical Analysis*, vol. 54, no. 3, pp. 1484–1506, 2016.

[46] D. Grapsas, R. Herbin, W. Kheriji, and J.-C. Latché, "An unconditionally stable staggered pressure correction scheme for the compressible Navier-Stokes equations," *The SMAI Journal of Computational Mathematics*, vol. 2, pp. 51–97, 2016.

[47] X. Zhang, "On positivity-preserving high order discontinuous Galerkin schemes for compressible Navier–Stokes equations," *Journal of Computational Physics*, vol. 328, pp. 301–343, 2017.

[48] J.-L. Guermond, M. Maier, B. Popov, and I. Tomas, "Second-order invariant domain preserving approximation of the compressible Navier–Stokes equations," *Computer Methods in Applied Mechanics and Engineering*, vol. 375, p. 113 608, 2021.

[49] K. N. Chueh, C. C. Conley, and J. A. Smoller, "Positively invariant regions for systems of nonlinear diffusion equations," *Indiana University Mathematics Journal*, vol. 26, no. 2, pp. 373–392, 1977.

[50] P. Thomas and C. Lombard, "Geometric conservation law and its application to flow computations on moving grids," *AIAA Journal*, vol. 17, no. 10, pp. 1030–1037, 1979.

[51] M. R. Visbal and D. V. Gaitonde, "On the use of higher-order finite-difference schemes on curvilinear and deforming meshes," *Journal of Computational Physics*, vol. 181, no. 1, pp. 155–185, 2002.

[52] T. C. Fisher, "High-order L2 stable multi-domain finite difference method for compressible flows," Ph.D. dissertation, Purdue University, 2012.

[53] S. K. Godunov, "An interesting class of quasilinear systems," in *Dokl. Acad. Nauk SSSR*, vol. 139, 1961, pp. 521–523.

[54] H. Brenner, "Navier–Stokes revisited," *Physica A: Statistical Mechanics and its Applications*, vol. 349, no. 1-2, pp. 60–132, 2005.

[55] ——, "Beyond Navier–Stokes," *International Journal of Engineering Science*, vol. 54, pp. 67–98, 2012.

[56] E. Feireisl and A. Vasseur, "New perspectives in fluid dynamics: Mathematical analysis of a model proposed by Howard Brenner," in *New directions in Mathematical Fluid Mechanics*, Springer, 2009, pp. 153–179.

[57] J.-L. Guermond and B. Popov, "Viscous regularization of the Euler equations and entropy principles," *SIAM Journal on Applied Mathematics*, vol. 74, no. 2, pp. 284–305, 2014.

[58] E. Feireisl and A. Novotnỳ, "Weak–strong uniqueness property for the full Navier–Stokes–Fourier system," *Archive for Rational Mechanics and Analysis*, vol. 204, no. 2, pp. 683–706, 2012.

[59] A. Harten, "On the symmetric form of systems of conservation laws with entropy," *Journal of Computational Physics*, vol. 49, pp. 151–164, 1983.

[60] A. Harten, P. D. Lax, C. D. Levermore, and W. J. Morokoff, "Convex entropies and hyperbolicity for general Euler equations," *SIAM Journal on Numerical Analysis*, vol. 35, no. 6, pp. 2117–2127, 1998.

[61] M. Svärd and J. Nordström, "Review of summation-by-parts schemes for initial–boundary-value problems," *Journal of Computational Physics*, vol. 268, pp. 17–38, 2014.

[62] T. C. Fisher, M. H. Carpenter, J. Nordström, N. K. Yamaleev, and C. Swanson, "Discretely conservative finite-difference formulations for nonlinear conservation laws in split form: Theory and boundary conditions," *Journal of Computational Physics*, vol. 234, pp. 353–375, 2013.

[63] A. Puckett and H. Stewart, "The thickness of a shock wave in air," *Quarterly of Applied Mathematics*, vol. 7, no. 4, pp. 457–463, 1950.

[64] X. Zhang and C.-W. Shu, "On positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes," *Journal of Computational Physics*, vol. 229, no. 23, pp. 8918–8934, 2010.

[65] H. Ranocha, M. Sayyari, L. Dalcin, M. Parsani, and D. I. Ketcheson, "Relaxation Runge–Kutta methods: Fully discrete explicit entropy-stable schemes for the compressible Euler and Navier–Stokes equations," *SIAM Journal on Scientific Computing*, vol. 42, no. 2, A612–A638, 2020.

[66] H. Ranocha, L. Lóczi, and D. I. Ketcheson, "General relaxation methods for initial-value problems with application to multistep schemes," *Numerische Mathematik*, vol. 146, no. 4, pp. 875–906, 2020.

[67] C.-W. Shu, "Total-variation-diminishing time discretizations," *SIAM Journal on Scientific and Statistical Computing*, vol. 9, no. 6, pp. 1073–1084, 1988.

[68] P. Woodward and P. Colella, "The numerical simulation of two-dimensional fluid flow with strong shocks," *Journal of Computational Physics*, vol. 54, no. 1, pp. 115–173, 1984.

[69] N. K. Yamaleev and M. H. Carpenter, "Third-order energy stable WENO scheme," *Journal of Computational Physics*, vol. 228, no. 8, pp. 3025–3047, 2009.

[70] Y. Yao, L Krishnan, N. Sandham, and G. Roberts, "The effect of Mach number on unstable disturbances in shock/boundary-layer interactions," *Physics of Fluids*, vol. 19, no. 5, p. 054 104, 2007.

[71] J Délery and R Bur, "The physics of shock wave/boundary layer interaction control: Last lessons learned," *OFFICE NATIONAL D ETUDES ET DE RECHERCHES AEROSPATIALES ONERA-PUBLICATIONS-TP*, vol. 181, 2000.

[72] F. Renac, "BL2 - Laminar shock-boundary layer interaction," in *4th International Workshop on High-Order CFD Methods*, 2016.

[73] G. Degrez, C. Boccadoro, and J. F. Wendt, "The interaction of an oblique shock wave with a laminar boundary layer revisited. An experimental and numerical study," *Journal of Fluid Mechanics*, vol. 177, pp. 247–263, 1987.

[74] G.-S. Jiang and C.-W. Shu, "Efficient implementation of weighted ENO schemes," *Journal of Computational Physics*, vol. 126, no. 1, pp. 202–228, 1996.

[75] N. Peng and Y. Yang, "Effects of the Mach number on the evolution of vortex-surface fields in compressible Taylor-Green flows," *Physical Review Fluids*, vol. 3, no. 1, p. 013 401, 2018.

[76] M. Parsani, M. H. Carpenter, and E. J. Nielsen, "Entropy stable wall boundary conditions for the three-dimensional compressible Navier–Stokes equations," *Journal of Computational Physics*, vol. 292, pp. 88–113, 2015.

APPENDIX A

ENTROPY STABILITY PROOFS

A.1 ENTROPY STABILITY OF FIRST-ORDER SYMMETRIC POSITIVE (SEMI-)DEFINITE FLUXES

The simplest manner of ensuring that a first-order flux is entropy stable is to write it in terms of a symmetric positive (semi-)definite matrix multiplied by the jump in the entropy variables. Since only semi-definiteness is required for entropy stability, we will assume symmetric positive semi-definite matrices (SPSD), but the same statements hold for SPD matrices as well. Since the entropy stability proofs of all such fluxes are essentially identical, we record it here for a general flux for reference.

Lemma 18. *Assume that* $\hat{\mathbf{U}}_t = \sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\tilde{\mathbf{f}}}_l^{(dis_1)} + \dots$ *where for all fixed* $1 \leq j,k \leq N$, $0 \leq i \leq N$, *and* $\vec{\xi}_i = \vec{\xi}_{ijk}$ *we have*

$$\hat{\tilde{\mathbf{f}}}_1^{(dis_1)}(\vec{\xi}_i) = M^{(dis_1)}(\vec{\xi}_i) \left(\mathbf{w}(\vec{\xi}_{i+1}) - \mathbf{w}(\vec{\xi}_i) \right) = M^{(dis_1)}(\vec{\xi}_i) \Delta_1 \mathbf{w}(\vec{\xi}_i), \tag{209}$$

where $\mathbf{w}(\vec{\xi}_0)$ *and* $\mathbf{w}(\vec{\xi}_{N+1})$ *are taken from the collocated state (numerical or boundary condition),* $M^{(dis_1)}(\vec{\xi}_i)$ *is SPSD, and we use identical definitions in the other computational directions.*

Let

$$H_a^{(dis_1)} = \mathbf{w}_a^\top \mathcal{P} \sum_{l=1}^{3} \mathcal{P}_{\xi^l}^{-1} \Delta_{\xi^l} \hat{\tilde{\mathbf{f}}}_{l,a}^{(dis_1)} \tag{210}$$

denote the total entropy contribution of $\hat{\tilde{\mathbf{f}}}_l^{(dis_1)}$ *on the* a*th element. Then, summing over the* K *total elements in the domain we have*

$$\sum_{a=1}^{K} H_a^{(dis_1)} = \sum_{a=1}^{K} \left[\sum_{l=1}^{3} \mathbf{w}_a^\top \mathcal{P}_{\perp,\xi^l} \tilde{B}_{\xi^l,a}^{(BC)} \hat{\tilde{\mathbf{f}}}_{l,a}^{(dis_1)} - H_a^{(dis_1,D)} - L_a^{(Int,dis_1,D)} \right], \tag{211}$$

where

$$\begin{aligned} H_a^{(dis_1,D)} &= \sum_{j,k=1}^{N} \mathcal{P}_{jk} \sum_{i=1}^{N-1} \Delta_1 \mathbf{w}_a(\vec{\xi}_{ijk}) \hat{\tilde{\mathbf{f}}}_{1,a}^{(dis_1)}(\vec{\xi}_{\bar{i}jk}) \\ &+ \sum_{i,k=1}^{N} \mathcal{P}_{ik} \sum_{j=1}^{N-1} \Delta_2 \mathbf{w}_a(\vec{\xi}_{ijk}) \hat{\tilde{\mathbf{f}}}_{2,a}^{(dis_1)}(\vec{\xi}_{i\bar{j}k}) \\ &+ \sum_{i,j=1}^{N} \mathcal{P}_{ij} \sum_{k=1}^{N-1} \Delta_3 \mathbf{w}_a(\vec{\xi}_{ijk}) \hat{\tilde{\mathbf{f}}}_{3,a}^{(dis_1)}(\vec{\xi}_{ij\bar{k}}) \end{aligned} \tag{212}$$

is non-negative and the entropy contribution of the interior interfaces is expressed by

$$\begin{aligned} L_a^{(Int,dis_1,D)} &= \sum_{j,k=1}^{N} \frac{\mathcal{P}_{jk}}{2} \Big[\Delta_1 \mathbf{w}_a(\vec{\xi}_{0jk}) \hat{\tilde{\mathbf{f}}}_{1,a}^{(dis_1)}(\vec{\xi}_{\bar{0}jk}) \chi_a^{(Int)}(\vec{\xi}_{0jk}) \\ &\qquad + \Delta_1 \mathbf{w}_a(\vec{\xi}_{Njk}) \hat{\tilde{\mathbf{f}}}_{1,a}^{(dis_1)}(\vec{\xi}_{\bar{N}jk}) \chi_a^{(Int)}(\vec{\xi}_{Njk}) \Big] \\ &+ \sum_{i,k=1}^{N} \frac{\mathcal{P}_{ik}}{2} \Big[\Delta_2 \mathbf{w}_a(\vec{\xi}_{i0k}) \hat{\tilde{\mathbf{f}}}_{2,a}^{(dis_1)}(\vec{\xi}_{i\bar{0}k}) \chi_a^{(Int)}(\vec{\xi}_{i0k}) \\ &\qquad + \Delta_2 \mathbf{w}_a(\vec{\xi}_{iNk}) \hat{\tilde{\mathbf{f}}}_{2,a}^{(dis_1)}(\vec{\xi}_{i\bar{N}k}) \chi_a^{(Int)}(\vec{\xi}_{iNk}) \Big] \\ &+ \sum_{i,j=1}^{N} \frac{\mathcal{P}_{ij}}{2} \Big[\Delta_3 \mathbf{w}_a(\vec{\xi}_{ij0}) \hat{\tilde{\mathbf{f}}}_{3,a}^{(dis_1)}(\vec{\xi}_{ij\bar{0}}) \chi_a^{(Int)}(\vec{\xi}_{ij0}) \\ &\qquad + \Delta_3 \mathbf{w}_a(\vec{\xi}_{ijN}) \hat{\tilde{\mathbf{f}}}_{3,a}^{(dis_1)}(\vec{\xi}_{ij\bar{N}}) \chi_a^{(Int)}(\vec{\xi}_{ijN}) \Big], \end{aligned} \tag{213}$$

which is also non-negative.

Proof. The generalized SBP property (51) gives us

$$\begin{aligned} H_a^{(dis_1)} &= \mathbf{w}_a^\top \sum_{l=1}^{3} \mathcal{P}_{\perp,\xi^l} \Delta_{\xi^l} \hat{\tilde{\mathbf{f}}}_{l,a}^{(dis_1)} \\ &= \mathbf{w}_a^\top \sum_{l=1}^{3} \mathcal{P}_{\perp,\xi^l} \tilde{B}_{\xi^l} \hat{\tilde{\mathbf{f}}}_{l,a}^{(dis_1)} - \mathbf{w}_a^\top \sum_{l=1}^{3} \mathcal{P}_{\perp,\xi^l} \tilde{\Delta}_{\xi^l} \hat{\tilde{\mathbf{f}}}_{l,a}^{(dis_1)} \\ &= \mathbf{w}_a^\top \sum_{l=1}^{3} \mathcal{P}_{\perp,\xi^l} \tilde{B}_{\xi^l} \hat{\tilde{\mathbf{f}}}_{l,a}^{(dis_1)} - H_a^{(dis_1,D)}, \end{aligned} \tag{214}$$

which proves the claim in the absence of interior element faces. Summing the entropy contribution of $\mathbf{w}_a^\top \sum_{l=1}^{3} \mathcal{P}_{\perp,\xi^l} \tilde{B}_{\xi^l} \hat{\tilde{\mathbf{f}}}_{l,a}^{(dis_1)}$ from each element at an interior face and then allocating half to each element sharing the face we obtain the contribution of $L_a^{(Int,dis_1,D)}$ given in

(213). □

A.2 ENTROPY STABILITY OF HIGH-ORDER VISCOUS FLUXES

In this section, we present a general proof of entropy stability for high-order viscous terms on curvilinear grids. We closely follow similar proofs that have been given for the Navier-Stokes viscous terms (e.g. see [36, 37, 52]).

Lemma 19. *Assume that* $\hat{\mathbf{U}}_t = \sum_{l=1}^{3} D_{\xi^l}\hat{\mathbf{f}}_l^{(visc)} + \mathcal{P}_{\xi^l}^{-1}\hat{\mathbf{g}}_l^{(visc)} \ldots$ *where the dissipative term* $\hat{\mathbf{f}}_l^{(visc)}$ *can be written as*

$$\hat{\mathbf{f}}_l^{(visc)} = \sum_{m=1}^{3}[\hat{a}_m^l]\mathbf{f}_{x^m}^{(visc)}, \quad \mathbf{f}_{x^m}^{(visc)} = \sum_{j=1}^{3}[c_{m,j}^{(visc)}]\mathbf{\Theta}_{x^j}. \tag{215}$$

Assume that for each $1 \leq a, b \leq 3$, $[c_{a,b}^{(visc)}]$ *is a block diagonal matrix with blocks that are* 5×5, $[\left(c_{a,b}^{(visc)}\right)^T] = [c_{b,a}^{(visc)}]$, *and* $\sum_{a=1}^{3}\sum_{b=1}^{3}\mathbf{v}^T[c_{a,b}^{(visc)}]\mathbf{v} \geq 0, \forall \mathbf{v}$ *i.e. the full viscous tensor is symmetric positive semi-definite (SPSD). Assume that* $\hat{\mathbf{g}}_l^{(visc)} = \hat{\mathbf{g}}_l^{(BC,visc)} + \hat{\mathbf{g}}_l^{(Int,visc)}$ *where* $\hat{\mathbf{g}}_l^{(BC,visc)}$ *is nonzero only at domain boundary faces and hence enforces the boundary conditions while* $\hat{\mathbf{g}}_l^{(Int,visc)}$ *is only nonzero at all interior faces collocated with neighboring elements. Assume that* $\hat{\mathbf{g}}_l^{(Int,visc)}$ *can be decomposed as the sum of an entropy conservative term and entropy dissipative term* $\hat{\mathbf{g}}_l^{(Int,visc)} = \hat{\mathbf{g}}_l^{(Int,visc,C)} + \hat{\mathbf{g}}_l^{(Int,visc,D)}$ *where for all fixed* $1 \leq j, k \leq N$ *and* $\vec{\xi}_i = \vec{\xi}_{ijk}$ *we have*

$$\begin{aligned}
\hat{\mathbf{g}}_1^{(Int,visc,C)}(\vec{\xi}_i) &= \frac{\chi^{(Int)}(\vec{\xi}_i)}{2}\sum_{m=1}^{3}\hat{a}_m^1(\vec{\xi}_i)\left(\delta_{1i}\Delta_1\mathbf{f}_{x^m}^{(visc)}(\vec{\xi}_{i-1}) + \delta_{Ni}\Delta_1\mathbf{f}_{x^m}^{(visc)}(\vec{\xi}_i)\right), \\
\hat{\mathbf{g}}_1^{(Int,visc,D)}(\vec{\xi}_i) &= \chi^{(Int)}(\vec{\xi}_i)\Big(-\delta_{1i}\Lambda^{(visc)}(\vec{\xi}_{i-1},\vec{\xi}_i)\Delta_1\mathbf{w}(\vec{\xi}_{i-1}) \\
&\qquad + \delta_{Ni}\Lambda^{(visc)}(\vec{\xi}_{i+1},\vec{\xi}_i)\Delta_1\mathbf{w}(\vec{\xi}_i)\Big),
\end{aligned} \tag{216}$$

where identical definitions hold for other computational directions and $\Lambda^{(visc)}$ *is SPSD. Let*

$$H_k^{(visc)} = \mathbf{w}_k^\top\mathcal{P}\left[\sum_{l=1}^{3}D_{\xi^l}\hat{\mathbf{f}}_{l,k}^{(visc)} + \mathcal{P}_{\xi^l}^{-1}\hat{\mathbf{g}}_{l,k}^{(visc)}\right] \tag{217}$$

denote the total entropy contribution of $\hat{\mathbf{f}}_l^{(visc)}$ *and* $\hat{\mathbf{g}}_l^{(visc)}$ *on the* k*th element. Then, summing over the* K *total elements in the domain we have*

$$\begin{aligned}\sum_{k=1}^{K} H_k^{(visc)} = \sum_{k=1}^{K}\sum_{l=1}^{3} \Big[\mathbf{w}_k^\top \mathcal{P}_{\perp,\xi^l} \left[B_{\xi^l,k}^{(BC)} \hat{\mathbf{f}}_{l,k}^{(visc)} + \hat{\mathbf{g}}_{l,k}^{(BC,visc)} \right] \\ + \left(\hat{\mathbf{g}}_{l,k}^{(BC,\mathbf{\Theta})} \right)^\top \mathcal{P}_{\perp,\xi^l} \hat{\mathbf{f}}_{l,k}^{(visc)} \Big] - \sum_{k=1}^{K} \left[H_k^{(visc,D)} + L_k^{(Int,visc,D)} \right],\end{aligned} \tag{218}$$

where $\hat{\mathbf{g}}_{l,k}^{(BC,\mathbf{\Theta})}(\vec{\xi}_{abc}) = \hat{\mathbf{g}}_{l,k}^{\mathbf{\Theta}}(\vec{\xi}_{abc})\chi_k^{(BC)}(\vec{\xi}_{abc})$,

$$H_k^{(visc,D)} = \sum_{m,j=1}^{3} \left(\mathbf{\Theta}_{x^m,k}\right)^\top \mathcal{P}[J]_k [c_{m,j}^{(visc)}]_k \mathbf{\Theta}_{x^j,k} \tag{219}$$

is non-negative, and the entropy contribution of $\hat{\mathbf{g}}_1^{(Int,visc,D)}$ *is expressed by* $L_k^{(Int,visc,D)}$ *given in Eq.* (213) *which is also non-negative.*

Proof. We begin by inspecting the entropy contribution on a single element:

$$\begin{aligned} H_k^{(visc)} &= \sum_{l=1}^{3} \mathbf{w}_k^\top \mathcal{P}_{\perp,\xi^l} \left[Q_{\xi^l} \hat{\mathbf{f}}_{l,k}^{(visc)} + \hat{\mathbf{g}}_{l,k}^{(visc)} \right] \\ &= \sum_{l=1}^{3} \left[\mathbf{w}_k^\top \mathcal{P}_{\perp,\xi^l} \left[B_{\xi^l} \hat{\mathbf{f}}_{l,k}^{(visc)} + \hat{\mathbf{g}}_{l,k}^{(visc)} \right] - \mathbf{w}_k^\top \mathcal{P}_{\perp,\xi^l} \mathcal{Q}_{\xi^l}^\top \hat{\mathbf{f}}_{l,k}^{(visc)} \right] \\ &= \sum_{l=1}^{3} \left[\mathbf{w}_k^\top \mathcal{P}_{\perp,\xi^l} \left[B_{\xi^l} \hat{\mathbf{f}}_{l,k}^{(visc)} + \hat{\mathbf{g}}_{l,k}^{(visc)} \right] - \left(D_{\xi^l} \mathbf{w}_k \right)^\top \mathcal{P} \hat{\mathbf{f}}_{l,k}^{(visc)} \right]. \end{aligned} \tag{220}$$

Note that

$$\begin{aligned}
\sum_{l=1}^{3} \left(D_{\xi^l}\mathbf{w}_k\right)^\top \mathcal{P}\hat{\mathbf{f}}_{l,k}^{(visc)} &= \sum_{l=1}^{3} \left(D_{\xi^l}\mathbf{w}_k\right)^\top \mathcal{P} \sum_{m=1}^{3} [\hat{a}_m^l]_k \mathbf{f}_{x^m,k}^{(visc)} \\
&= \sum_{m=1}^{3}\sum_{l=1}^{3} \left(D_{\xi^l}\mathbf{w}_k\right)^\top [\hat{a}_m^l]_k \mathcal{P}\mathbf{f}_{x^m,k}^{(visc)} \\
&= \sum_{m=1}^{3} \left(\mathbf{w}_{x^m,k}\right)^\top \mathcal{P}[J]_k \mathbf{f}_{x^m,k}^{(visc)} \\
&= \sum_{m,j=1}^{3} \left(\mathbf{\Theta}_{x^m,k} - \mathbf{g}_{m,k}^{\mathbf{\Theta}}\right)^\top \mathcal{P}[J]_k [c_{m,j}^{(v)}]_k \mathbf{\Theta}_{x^j,k} \\
&= H_k^{(visc,D)} - \sum_{m=1}^{3} \left(\mathbf{g}_{m,k}^{\mathbf{\Theta}}\right)^\top \mathcal{P}[J]_k \mathbf{f}_{x^m,k}^{(visc)} \\
&= H_k^{(visc,D)} - \sum_{l=1}^{3} \left(\hat{\mathbf{g}}_{l,k}^{\mathbf{\Theta}}\right)^\top \mathcal{P}_{\perp,\xi^l} \hat{\mathbf{f}}_{l,k}^{(visc)}.
\end{aligned} \tag{221}$$

Hence,

$$\begin{aligned}
H_k^{(visc)} &= \sum_{l=1}^{3} \left[\mathbf{w}_k^\top \mathcal{P}_{\perp,\xi^l} \left[B_{\xi^l}\hat{\mathbf{f}}_{l,k}^{(visc)} + \hat{\mathbf{g}}_{l,k}^{(visc)}\right] - \left(D_{\xi^l}\mathbf{w}_k\right)^\top \mathcal{P}\hat{\mathbf{f}}_{l,k}^{(visc)}\right] \\
&= \sum_{l=1}^{3} \left[\mathbf{w}_k^\top \mathcal{P}_{\perp,\xi^l} \left[B_{\xi^l}\hat{\mathbf{f}}_{l,k}^{(visc)} + \hat{\mathbf{g}}_{l,k}^{(visc)}\right] + \left(\hat{\mathbf{g}}_{l,k}^{\mathbf{\Theta}}\right)^\top \mathcal{P}_{\perp,\xi^l} \hat{\mathbf{f}}_{l,k}^{(visc)}\right] - H_k^{(visc,D)}.
\end{aligned} \tag{222}$$

If all faces were domain boundary faces, then Eq. (222) would directly imply the result we intend to prove. Hence, we inspect the sum of the boundary terms in Eq. (222) at interior domain faces. Given that $\hat{\mathbf{g}}_l^{\mathbf{\Theta}}$, $B_{\xi^l}\hat{\mathbf{f}}_l^{(visc)}$ and $\hat{\mathbf{g}}_l^{(visc)}$ are only nonzero at element faces, it is sufficient to consider a single point on one interior face for two general elements. We use $\vec{\xi}_L$ and $\vec{\xi}_R$ to denote the computational coordinates on two different elements that map to the same physical coordinate at a shared element interface. Furthermore, since the element wise defined computational directions may differ, we let $\hat{\mathbf{f}}_l^{(visc)}(\vec{\xi}_L) = \hat{f}^{(visc)}(\vec{\xi}_L)$ represent the outward (relative to the $\vec{\xi}_L$ state element) normal flux for the $\vec{\xi}_L$ state. We split the sum of the terms in Eq. (222) (not including $H_k^{(visc,D)}$) from each element at each shared point as

$C + D$ where

$$\begin{aligned}
D &= \mathcal{P}_\perp (w(\vec{\xi}_L) - w(\vec{\xi}_R))^\top \Lambda^{(visc)}(\vec{\xi}_L, \vec{\xi}_R)(w(\vec{\xi}_L) - w(\vec{\xi}_R))^\top, \\
C &= w(\vec{\xi}_L)^\top \mathcal{P}_\perp \left[\hat{f}^{(v)}(\vec{\xi}_L) - \frac{1}{2}\left(\hat{f}^{(v)}(\vec{\xi}_R) + \hat{f}^{(v)}(\vec{\xi}_L) \right) \right] \\
&\quad + \frac{\mathcal{P}_\perp}{2} \left(\left(w(\vec{\xi}_R) - w(\vec{\xi}_L) \right)^\top \hat{f}^{(v)}(\vec{\xi}_L) \right) + \\
&\quad w(\vec{\xi}_R)^\top \mathcal{P}_\perp \left[\hat{f}^{(v)}(\vec{\xi}_R) - \frac{1}{2}\left(\hat{f}^{(v)}(\vec{\xi}_R) + \hat{f}^{(v)}(\vec{\xi}_L) \right) \right] \\
&\quad + \frac{\mathcal{P}_\perp}{2} \left(\left(w(\vec{\xi}_L) - w(\vec{\xi}_R) \right)^\top \hat{f}^{(v)}(\vec{\xi}_R) \right) \\
&\quad = 0,
\end{aligned} \tag{223}$$

where the scalar $\mathcal{P}_\perp$ is the scaling from $\mathcal{P}_{\perp,l}$ at the given point and since $C = 0$ we see that only the dissipative term D is present in Eq. (218) through the sum of $L_k^{(Int,visc,D)}$. □

A.3 ENTROPY STABLE BRENNER-NAVIER-STOKES FLUXES

Brenner's modification to the Navier-Stokes equations [54, 55] can be viewed as changing the Navier-Stokes viscous fluxes, $\boldsymbol{F}_{x_m}^{(v)}$, to the Brenner-Navier-Stokes viscous fluxes, $\boldsymbol{F}_{x_m}^{(B)}, m = 1, 2, 3$, where

$$\boldsymbol{F}_{x_m}^{(B)} = \boldsymbol{F}_{x_m}^{(v)} + \sigma \frac{\partial \rho}{\partial x_m} \begin{bmatrix} 1 & \vec{\boldsymbol{V}} & E \end{bmatrix}^\top. \tag{224}$$

This change can also be viewed as changing the viscosity matrices $\mathsf{C}_{m,j}$ of Eq. (15) to the Brenner-Navier-Stokes viscosity matrices $\mathsf{C}_{m,j}^{(B)}$ where $\mathsf{C}_{m,j}^{(B)} = \mathsf{C}_{m,j},\ m \neq j$ and $\mathsf{C}_{j,j}^{(B)} = \mathsf{C}_{j,j} +$

$M^{(B)},\ j = 1, 2, 3$. The matrix $M^{(B)}$ is given by

$$
\begin{aligned}
M^{(B)} &= \sigma \begin{bmatrix} 1 & 0 & \dots \\ \vec{\boldsymbol{V}} & 0 & \dots \\ E & 0 & \dots \end{bmatrix} \frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{W}} \\
&= \frac{\rho\sigma}{R} \begin{bmatrix} 1 & V_1 & V_2 & V_3 & E \\ V_1 & V_1^2 & V_1V_2 & V_1V_3 & V_1E \\ V_2 & V_1V_2 & V_2^2 & V_2V_3 & V_2E \\ V_3 & V_1V_3 & V_2V_3 & V_3^2 & V_3E \\ E & V_1E & V_2E & V_3E & E^2 \end{bmatrix} \\
&= \frac{\rho\sigma}{R} \begin{bmatrix} 1 & \vec{\boldsymbol{V}} & E \end{bmatrix} \otimes \begin{bmatrix} 1 & \vec{\boldsymbol{V}} & E \end{bmatrix},
\end{aligned}
\tag{225}
$$

where $\boldsymbol{\nu} = \begin{bmatrix} \rho & \vec{\boldsymbol{V}} & T \end{bmatrix}^T$ are the primitive variables.

The entropy stability property of the Navier-Stokes (NS) viscosity matrices, $\mathsf{C}_{m,j}$, that was discussed in Section 2.3 also holds for the the Brenner-Navier-Stokes (BNS) viscosity matrices, $\mathsf{C}^{(B)}_{m,j}$, as well. In [59, 60] a much larger class of entropies were developed for the Euler equations, but the NS viscosity matrices are only SPSD for one of them, the physical entropy given by Eq. (10). In [57], a general viscous regularization of the Euler equations were derived that was entropy dissipative for all the generalized entropies of [59, 60] and in [57] the authors mention that their general viscous regularization is connected to the BNS viscous term. Here, we explicitly give the conditions for $\mathsf{C}^{(B)}_{m,j}$ to be SPSD for all of the generalized entropies of [59, 60] and the corresponding viscosity matrices.

A.4 ENTROPY STABILITY FOR GENERALIZED ENTROPIES

Let s (the specific thermodynamic entropy given by Eq. (6)) be twice differentiable for an admissible state $\boldsymbol{u}$ with positive density and temperature and assume that f is a twice differentiable functtion of a real variable. The generalized entropies in [60] are those functions $\mathcal{S}_f = -\rho f(s)$ which are strictly convex and in [60] it was shown that strict convexity holds if and only if

$$
f'(s) > 0, \qquad f'(s)\frac{1}{c_P} - f''(s) > 0. \tag{226}
$$

Convexity of $\mathcal{S}_f$ gives us a one-to-one mapping from the conservative to generalized entropy variables that are defined as follows:

$$\boldsymbol{W}_f^\top \equiv \frac{\partial \mathcal{S}_f}{\partial \boldsymbol{U}} = f'(s)\left[\frac{h}{T} - \frac{f(s)}{f'(s)} - \frac{\boldsymbol{V}^\top \boldsymbol{V}}{2T}, \frac{V_1}{T}, \frac{V_2}{T}, \frac{V_3}{T}, -\frac{1}{T}\right]^\top. \tag{227}$$

Using the generalized entropy variables, we can attempt to symmetrize the BNS viscosity matrices with respect to the generalized entropy variables by forming:

$$\mathsf{C}_{m,j}^{(B),f} = \mathsf{C}_{m,j}^{(B)} \frac{\partial \boldsymbol{W}}{\partial \boldsymbol{W}_f}, \tag{228}$$

where

$$\frac{\partial \boldsymbol{W}}{\partial \boldsymbol{W}_f} = \frac{1}{T}\begin{bmatrix} T\frac{\gamma-1}{c_1^f} - \frac{c_2^f \|\vec{\boldsymbol{V}}\|^2}{2} & c_4^f V_1 & c_4^f V_2 & c_4^f V_3 & \frac{\|\vec{\boldsymbol{V}}\|^2 c_3^f c_2^f}{2(\gamma-1)} \\ c_2^f V_1 & \frac{T}{f'(s)} + V_1^2 c_2^f & c_2^f V_1 V_2 & c_2^f V_1 V_3 & c_2^f V_1 \frac{\|\vec{\boldsymbol{V}}\|^2}{2} \\ c_2^f V_2 & c_2^f V_2 V_1 & \frac{T}{f'(s)} + V_2^2 c_2^f & c_2^f V_2 V_3 & c_2^f V_2 \frac{\|\vec{\boldsymbol{V}}\|^2}{2} \\ c_2^f V_3 & c_2^f V_3 V_1 & c_2^f V_3 V_2 & \frac{T}{f'(s)} + V_3^2 c_2^f & c_2^f V_3 \frac{\|\vec{\boldsymbol{V}}\|^2}{2} \\ -c_2^f & -c_2^f V_1 & -c_2^f V_2 & -c_2^f V_3 & \frac{T}{f'(s)} - c_2^f \frac{\|\vec{\boldsymbol{V}}\|^2}{2} \end{bmatrix}, \tag{229}$$

$$c_1^f = f'(s)\frac{R\gamma}{c_P} - f''(s)R\gamma > 0, \qquad c_2^f = f''(s)\frac{(\gamma-1)}{c_1^f f'(s)},$$
$$c_3^f = R\gamma T - \frac{\|\vec{\boldsymbol{V}}\|^2}{2}(\gamma-1), \qquad c_4^f = \frac{f''(s)}{f'(s)}\frac{c_3^f}{c_1^f}.$$

Notice that for the case $f(s) = s$ (i.e. the physical entropy) $\frac{\partial \boldsymbol{W}}{\partial \boldsymbol{W}_f} = \frac{\partial \boldsymbol{W}}{\partial \boldsymbol{W}}$ is the identity since $f''(s), c_2^f, c_4^f = 0$, $c_1^f = \frac{R\gamma}{c_P} = \gamma - 1$, and $f'(s) = 1$.

We wish to derive conditions for which $\mathsf{C}_{m,j}^{(B),f}$ satisfy $\mathsf{C}_{m,j}^{(B),f} = (\mathsf{C}_{j,m}^{(B),f})^\top$

$$\begin{aligned} &\mathsf{C}_{m,j}^{(B),f} = (\mathsf{C}_{j,m}^{(B),f})^\top, \\ &\sum_{l,m=1}^{3} \boldsymbol{A}_l \mathsf{C}_{l,m}^{(B),f} \boldsymbol{A}_m \geq 0, \quad \forall \boldsymbol{A}_i \in \mathbb{R}^5. \end{aligned} \tag{230}$$

By comparing the entries $(\mathsf{C}_{1,1}^{(B),f})_{1,5}$ and $(\mathsf{C}_{1,1}^{(B),f})_{5,1}$ of the matrix $\mathsf{C}_{1,1}^{(B),f}$ we immediately see

a necessary condition for symmetry relating the mass and heat diffusion coefficients

$$\begin{aligned}
&\frac{T}{f'(s)}\frac{\sigma\rho}{\gamma-1}+\sigma_c\frac{\|\vec{\boldsymbol{V}}\|^2}{2}=(\mathsf{C}_{1,1}^{(B),f})_{1,5}=(\mathsf{C}_{1,1}^{(B),f})_{5,1}=\kappa_c+\sigma_c E,\\
&\sigma_c=\frac{\sigma\rho}{c_1^f}\left(\frac{\gamma-1}{R}-\frac{f''(s)}{f'(s)}\right),\qquad\qquad \kappa_c=-c_2^f\kappa T.
\end{aligned}\tag{231}$$

The necessary symmetry condition of (231) is satisfied for all physical states and $f(s)$ satisfying (226) if and only if

$$\sigma=\frac{\gamma-1}{R}\frac{\kappa}{\rho}=\frac{\gamma}{c_P}\frac{\kappa}{\rho}.\tag{232}$$

We denote the matrices $\mathsf{C}_{m,j}^{(B),f}$ where $\sigma=\frac{\gamma-1}{R}\frac{\kappa}{\rho}$ as $\mathsf{C}_{m,j}^{(B_s),f}$ where the subscript s is added because it can be verified that Eq. (232) is also sufficient for the symmetry property $\mathsf{C}_{m,j}^{(B_s),f}=(\mathsf{C}_{j,m}^{(B_s),f})^\top$. The matrices $\mathsf{C}_{m,j}^{(B_s),f}$ can be written as

for i= 1,2,3,

$$\mathsf{C}_{i,i}^{(B_s),f}=\begin{bmatrix}
\kappa_d & \kappa_d V_1 & \kappa_d V_2 & \kappa_d V_3 & \kappa_e\\
\kappa_d V_1 & \mu_d L_{i,1}+\kappa_d V_1^2 & \kappa_d V_1V_2 & \kappa_d V_1V_3 & (\mu_d L_{i,1}+\kappa_e)V_1\\
\kappa_d V_2 & \kappa_d V_1V_2 & \mu_d L_{i,2}+\kappa_d V_2^2 & \kappa_d V_2V_3 & (\mu_d L_{i,2}+\kappa_e)V_2\\
\kappa_d V_3 & \kappa_d V_1V_3 & \kappa_d V_2V_3 & \mu_d L_{i,3}+\kappa_d V_3^2 & (\mu_d L_{i,3}+\kappa_e)V_3\\
\kappa_e & (\mu_d L_{i,1}+\kappa_e)V_1 & (\mu_d L_{i,2}+\kappa_e)V_2 & (\mu_d L_{i,3}+\kappa_e)V_3 & \mu_d M_i+\kappa_f
\end{bmatrix},$$

$$\begin{aligned}
&\kappa_d=\frac{\kappa}{f'(s)}\frac{\gamma-1}{R^2}\frac{f'(s)\frac{R\gamma}{c_P}-Rf''(s)}{c_1^f}\geq 0,\quad \kappa_d=0 \text{ iff } \kappa=0,\\
&\kappa_e=\frac{\kappa}{f'(s)}\frac{T}{R}+\kappa_d\frac{\|\vec{\boldsymbol{V}}\|^2}{2},\\
&\kappa_f=\frac{\kappa}{f'(s)}\frac{\gamma}{\gamma-1}T^2+\frac{\kappa}{f'(s)}\frac{T}{R}\|\vec{\boldsymbol{V}}\|^2+\kappa_d\left(\frac{\|\vec{\boldsymbol{V}}\|^2}{2}\right)^2,\\
&L_{i,j}=1+\frac{1}{3}\delta_{i,j},\qquad\qquad M_i=L_{i,1}V_1^2+L_{i,2}V_2^2+L_{i,3}V_3^2,\\
&\mu_d=\frac{\mu}{f'(s)}T,
\end{aligned}\tag{233}$$

and

$$
\begin{aligned}
\mathsf{C}_{1,2}^{(B_s),f} &= \left(\mathsf{C}_{2,1}^{(B_s),f}\right)^{\top} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{2}{3}\mu_d & 0 & -\frac{2}{3}\mu_d V_2 \\ 0 & \mu_d & 0 & 0 & \mu_d V_1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \mu_d V_2 & -\frac{2}{3}\mu_d V_1 & 0 & \frac{1}{3}\mu_d V_1 V_2 \end{bmatrix}, \\
\mathsf{C}_{1,3}^{(B_s),f} &= \left(\mathsf{C}_{3,1}^{(B_s),f}\right)^{\top} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{2}{3}\mu_d & -\frac{2}{3}\mu_d V_3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \mu_d & 0 & 0 & \mu_d V_1 \\ 0 & \mu_d V_3 & 0 & -\frac{2}{3}\mu_d V_1 & \frac{1}{3}\mu_d V_1 V_3 \end{bmatrix}, \\
\mathsf{C}_{2,3}^{(B_s),f} &= \left(\mathsf{C}_{3,2}^{(B_s),f}\right)^{\top} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{2}{3}\mu_d & -\frac{2}{3}\mu_d V_3 \\ 0 & 0 & \mu_d & 0 & \mu_d V_2 \\ 0 & 0 & \mu_d V_3 & -\frac{2}{3}\mu_d V_2 & \frac{1}{3}\mu_d V_2 V_3 \end{bmatrix}.
\end{aligned} \tag{234}
$$

Less trivially, we can show that Eq. (232) is also sufficient for the SPSD statement of (230).

Theorem 20. *For $\kappa, \mu \geq 0$, states where $\rho, T > 0$, and $f(s)$ satisfying* (226)*, the viscosity matrices $\mathsf{C}_{i,j}^{(B_s),f}$ $i, j = 1, 2, 3$ given by* (233) *and* (234) *have the property*

$$
\begin{aligned}
&\mathsf{C}_{m,j}^{(B_s),f} = (\mathsf{C}_{j,m}^{(B_s),f})^{\top}, \\
&\sum_{l,m=1}^{3} \boldsymbol{A}_l \mathsf{C}_{l,m}^{(B_s),f} \boldsymbol{A}_m \geq 0, \quad \forall \boldsymbol{A}_i \in \mathbb{R}^5.
\end{aligned} \tag{235}
$$

Proof. The symmetry property $\mathsf{C}^{(B_s),f}_{m,j} = (\mathsf{C}^{(B_s),f}_{j,m})^\top$ was checked in Mathematica. To establish that $\sum_{l,m=1}^{3} \boldsymbol{A}_l \mathsf{C}^{(B_s),f}_{l,m} \boldsymbol{A}_m \geq 0, \forall \boldsymbol{A}_i \in \mathbb{R}^5$, we create a larger matrix

$$C^{(B_s),f} = \begin{bmatrix} \mathsf{C}^{(B_s),f}_{1,1} & \mathsf{C}^{(B_s),f}_{1,2} & \mathsf{C}^{(B_s),f}_{1,3} \\ \mathsf{C}^{(B_s),f}_{2,1} & \mathsf{C}^{(B_s),f}_{2,2} & \mathsf{C}^{(B_s),f}_{2,3} \\ \mathsf{C}^{(B_s),f}_{3,1} & \mathsf{C}^{(B_s),f}_{3,2} & \mathsf{C}^{(B_s),f}_{3,3} \end{bmatrix} \tag{236}$$

and show that $C^{(B_s),f}$ is positive semi-definite, i.e. that

$$\begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \boldsymbol{A}_3 \end{bmatrix} C^{(B_s),f} \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \boldsymbol{A}_3 \end{bmatrix} \geq 0, \quad \forall \boldsymbol{A}_i \in \mathbb{R}^5. \tag{237}$$

Using Mathematica, we obtain the Cholesky decomposition, $C^{(B_s),f} = LDL^\top$, where

$$\begin{aligned} D &= \operatorname{diag}\begin{bmatrix} \kappa_d & \frac{4}{3}\mu_d & \mu_d & \mu_d & \kappa_g & \kappa_d & 0 & \mu_d & \mu_d & \kappa_g & \kappa_d & 0 & 0 & 0 & \kappa_g \end{bmatrix}, \\ \kappa_g &= \frac{(\gamma-1)\kappa T^2}{f'(s)\frac{R\gamma}{c_P} - f''(s)R} \geq 0, \quad \kappa_g = 0 \text{ iff } \kappa = 0, \end{aligned} \tag{238}$$

where $f'(s)\frac{R\gamma}{c_P} - f''(s)R > 0$ follows from (226). Since D has only non-negative entries it follows that $C^{(B_s),f}$ is positive semi-definite. □

Remark 12. Notice that no condition was required for the dynamic viscosity μ besides non-negativity. This implies that shear stress already satisfies (235) for all of the generalized entropies of [60]. However, the thermal conductivity κ and mass diffusion σ only satisfy (235) for all the generalized entropies of [60] if Eq. (232) is satisfied. In particular, it is only the heat diffusion terms that prevent the Navier-Stokes equations from satisfying (235) for all of the generalized entropies of [60].

A.5 ENTROPY CONSERVATIVE INVISCID FLUXES

The high-order and first-order inviscid fluxes discussed in Sections 4.1.2 and 6.1.2, respectively, have the same element-wise entropy contributions (see (75) of Theorem 1 and (102) of Lemma 2). Since they also use the same penalties (70), they have the same global entropy contribution which we present here for reference. This result is already known and can be found in (or, is a straightforward consequence of results in) [36, 37, 52].

Lemma 21. *Assume that* $\hat{\mathbf{U}}_t = \sum_{l=1}^{3} -\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l + \mathcal{P}_{\xi^l}^{-1}\hat{\mathbf{g}}_l \ldots$ *for some flux* $\hat{\tilde{\mathbf{f}}}_l$ *such that*

$$\sum_{l=1}^{3} \mathbf{w}^\top \mathcal{P}\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_l = \sum_{l=1}^{3} \mathbf{1}_1^\top \widehat{\mathcal{P}}_{\perp,\xi^l}\widehat{B}_{\xi^l}\hat{\mathbf{F}}_l. \tag{239}$$

Furthermore, assume that $\hat{\mathbf{g}}_l = \hat{\mathbf{g}}_l^{(BC)} + \hat{\mathbf{g}}_l^{(Int)}$ *where* $\hat{\mathbf{g}}_l^{(BC)}$ *is nonzero only at domain boundary faces and hence enforces the boundary conditions while* $\hat{\mathbf{g}}_l^{(Int)}$ *is only nonzero at all interior faces collocated with neighboring elements. Assume that for all fixed* $1 \leq j,k \leq N$ *and* $\vec{\xi}_i = \vec{\xi}_{ijk}$ *we have*

$$\begin{aligned}\hat{\mathbf{g}}_1^{(Int)}(\vec{\xi}_i) = \chi^{(Int)}(\vec{\xi}_i)\Big(&-\delta_{1i}\left(\hat{\tilde{\mathbf{f}}}_1(\vec{\xi}_1) - \bar{f}_{(S)}(\mathbf{U}(\vec{\xi}_1),\mathbf{U}(\vec{\xi}_0))\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_1)\right)\\ &+ \delta_{Ni}\left(\hat{\tilde{\mathbf{f}}}_1(\vec{\xi}_N) - \bar{f}_{(S)}(\mathbf{U}(\vec{\xi}_{N+1}),\mathbf{U}(\vec{\xi}_N))\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_N)\right)\Big),\end{aligned} \tag{240}$$

for $\bar{f}_{(S)}(\cdot,\cdot)$ *any two-point, consistent, entropy consistent inviscid interface flux and identical definitions hold for the other computational directions. Let*

$$H_k^{(I)} = \mathbf{w}_k^\top \mathcal{P}\left[\sum_{l=1}^{3} -\mathcal{P}_{\xi^l}^{-1}\Delta_{\xi^l}\hat{\tilde{\mathbf{f}}}_{l,k} + \mathcal{P}_{\xi^l}^{-1}\hat{\mathbf{g}}_{l,k}\right] \tag{241}$$

denote the total entropy contribution of $\hat{\tilde{\mathbf{f}}}_{l,k}$ *and* $\hat{\mathbf{g}}_{l,k}$ *on the* k*th element. Then, summing over the* K *total elements in the domain we have*

$$\sum_{k=1}^{K} H_k^{(I)} = \sum_{k=1}^{K}\sum_{l=1}^{3}\left[\mathbf{1}_1^\top \widehat{\mathcal{P}}_{\perp,\xi^l}\widehat{B}_{\xi^l,k}^{(BC)}\hat{\mathbf{F}}_{l,k} + \mathbf{w}_k^\top \mathcal{P}_{\perp,\xi^l}\hat{\mathbf{g}}_{l,k}^{(BC)}\right]. \tag{242}$$

Hence, $\hat{\tilde{\mathbf{f}}}_l$ *and* $\hat{\mathbf{g}}_l$ *discretely conserve the entropy in the domain up to the boundary conditions.*

Proof. The element-wise entropy contribution of $\hat{\tilde{\mathbf{f}}}_l$ is assumed to be given by Eq. (239). Hence, we need only look at the entropy contribution of $\hat{\mathbf{g}}_l$. Notice that

$$\mathbf{w}_k^\top \mathcal{P} \mathcal{P}_{\xi^l}^{-1} \hat{\mathbf{g}}_{l,k} = \mathbf{w}_k^\top \mathcal{P}_{\perp,\xi^l} \hat{\mathbf{g}}_{l,k}. \tag{243}$$

If all faces were domain boundaries, then Eq. (242) would be proven. Hence, we inspect $\hat{\mathbf{g}}_{l,k}$ at interior domain faces. Equation (243) is the weighted sum of the pointwise entropy variables multiplied by the pointwise penalty, $\hat{\mathbf{g}}_{l,k}$. Hence, it is sufficient to consider the sum of this product for two general elements at a single shared point. We use $\vec{\xi}_L$ and $\vec{\xi}_R$ to denote the computational coordinates on two different elements that map to the same physical coordinate at a shared element interface. The single state fluxes in $\hat{\mathbf{g}}_{l,k}$ are denoted $\hat{\tilde{f}}(\vec{\xi}_L)$ and $\hat{\tilde{f}}(\vec{\xi}_R)$ and have the outward normal sign–relative to their respective elements. For the two point flux, we write $f_{(S)}^{L,R}$ and arbitrarily choose the sign that makes it outward for the "L" element. We then have (note that the $\mathcal{P}$ scaling is the same for both elements at the shared point so we ignore this scaling)

$$\begin{aligned}
&\mathbf{w}(\vec{\xi}_L)^\top \hat{\mathbf{g}}_l^{(Int)}(\vec{\xi}_L) + \mathbf{w}(\vec{\xi}_R)^\top \hat{\mathbf{g}}_l^{(Int)}(\vec{\xi}_R) \\
=&\mathbf{w}(\vec{\xi}_L)^\top \left(\hat{\tilde{f}}(\vec{\xi}_L) - f_{(S)}^{L,R} \right) + \mathbf{w}(\vec{\xi}_R)^\top \left(\hat{\tilde{f}}(\vec{\xi}_R) + f_{(S)}^{L,R} \right) \\
=&\psi(\vec{\xi}_L) + F(\vec{\xi}_L) + \psi(\vec{\xi}_R) + F(\vec{\xi}_R) + \left(\mathbf{w}(\vec{\xi}_R) - \mathbf{w}(\vec{\xi}_L) \right)^\top f_{(S)}^{L,R} \\
=&F(\vec{\xi}_L) + F(\vec{\xi}_R),
\end{aligned} \tag{244}$$

where $\psi(\vec{\xi}_L)$, $\psi(\vec{\xi}_R)$, $F(\vec{\xi}_L)$, and $F(\vec{\xi}_R)$ are the outward (relative to their respective elements) entropy potential fluxes and entropy fluxes (respectively) and we made use of Eqs. (17) and (68). Given that we also have Eq. (239), it follows that at every interior interface, the sum of the entropy contributions of $\hat{\mathbf{g}}_l$ and $\hat{\tilde{\mathbf{f}}}_l$ is zero and hence Eq. (242) now follows. □

APPENDIX B

BOUNDARY CONDITIONS

Here, we address the treatment of domain boundaries. We begin by writing the general form of the boundary penalties for the various terms in the high-order positivity-preserving scheme given by Eq. (178); then, we will discuss specific boundary conditions used in obtaining the results in Section 7.3.

B.1 FORM OF INVISCID BOUNDARY PENALTIES

The general form we used for the inviscid boundary penalties are identical to those given in Eq. (70) for the interior interface penalties. The inviscid boundary penalties are decomposed as the sum of two terms $\hat{\mathbf{g}}_l^{(BC,I)} = \hat{\mathbf{g}}_l^{(BC,I,C)} + \hat{\mathbf{g}}_l^{(BC,I,D)}$. Notice that, depending on the implementation of the boundary conditions, $\hat{\mathbf{g}}_l^{(BC,I,C)}$ and $\hat{\mathbf{g}}_l^{(BC,I,D)}$ may not be entropy conservative or entropy dissipative, but we maintain the "C" and "D" superscripts to highlight the parallel between these terms and those used at the interfaces (see Eq. (70)). For all fixed $1 \leq j,k \leq N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$\begin{aligned}
\hat{\mathbf{g}}_1^{(BC,I,C)}(\vec{\xi}_i) &= \chi^{(BC)}(\vec{\xi}_i)\left(-\delta_{1i}\left(\hat{\bar{\mathbf{f}}}_1(\vec{\xi}_1) - \bar{f}_{(S)}(\mathbf{U}(\vec{\xi}_1),\mathbf{U}(\vec{\xi}_0))\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_1)\right)\right. \\
&\quad \left. + \delta_{Ni}\left(\hat{\bar{\mathbf{f}}}_1(\vec{\xi}_N) - \bar{f}_{(S)}(\mathbf{U}(\vec{\xi}_{N+1}),\mathbf{U}(\vec{\xi}_N))\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_N)\right)\right), \\
\hat{\mathbf{g}}_1^{(BC,I,D)}(\vec{\xi}_i) &= \chi^{(BC)}(\vec{\xi}_i)\left(-\delta_{1i}M^{\mathcal{Y}}(\mathbf{U}(\vec{\xi}_0),\mathbf{U}(\vec{\xi}_1),\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_1))\Delta_1\mathbf{w}(\vec{\xi}_0)\right. \\
&\quad \left. + \delta_{Ni}M^{\mathcal{Y}}(\mathbf{U}(\vec{\xi}_{N+1}),\mathbf{U}(\vec{\xi}_N),\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_N))\Delta_1\mathbf{w}(\vec{\xi}_N)\right),
\end{aligned} \tag{245}$$

with identical definitions for the other computational directions. The $\mathbf{U}(\vec{\xi}_{N+1})$ and $\mathbf{U}(\vec{\xi}_0)$ states are specified by the boundary condition and are the only means of enforcing the boundary conditions for the inviscid terms.

B.2 FORM OF HIGH-ORDER VISCOUS BOUNDARY PENALTIES

The boundary penalties for $\hat{\mathbf{f}}_l^{(v)}$ and $\hat{\mathbf{f}}_l^{(AD_p)}$ are enforced through $\hat{\mathbf{g}}_l^{(BC,v)}$ and $\hat{\mathbf{g}}_l^{(BC,AD_p)}$ respectively. In general form, the penalty $\hat{\mathbf{g}}_l^{(BC,AD_p)}$ is identical to $\hat{\mathbf{g}}_l^{(BC,v)}$; hence, we write $\hat{\mathbf{g}}_l^{(BC,v_p)}$ as the general high-order viscous flux domain boundary penalty. The boundary penalty is decomposed as the sum of two terms $\hat{\mathbf{g}}_l^{(BC,v_p)} = \hat{\mathbf{g}}_l^{(BC,v_p,C)} + \hat{\mathbf{g}}_l^{(BC,v_p,D)}$. As just discussed for the inviscid case, depending on the implementation of the boundary conditions, $\hat{\mathbf{g}}_l^{(BC,v_p,C)}$ and $\hat{\mathbf{g}}_l^{(BC,v_p,D)}$ may not be entropy conservative or entropy dissipative, but we maintain the "C" and "D" superscripts to emphasize the similarity between these terms and those used at the interior interfaces (see Eq. (67)). For all fixed $1 \leq j,k \leq N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$\begin{aligned}
\hat{\mathbf{g}}_1^{(BC,v_p,C)}(\vec{\xi}_i) &= \frac{\chi^{(BC)}(\vec{\xi}_i)}{2}\left[\delta_{1i}\left(\hat{\mathbf{f}}_1^{(v_p)}(\vec{\xi}_1) - \hat{\mathbf{f}}_1^{(v_p,BC)}(\vec{\xi}_0)\right) + \delta_{Ni}\left(\hat{\mathbf{f}}_1^{(v_p,BC)}(\vec{\xi}_{N+1}) - \hat{\mathbf{f}}_1^{(v_p)}(\vec{\xi}_N)\right)\right], \\
\hat{\mathbf{g}}_1^{(BC,v_p,D)}(\vec{\xi}_i) &= \chi^{(BC)}(\vec{\xi}_i)\left(-\delta_{1i}\Lambda^{(v_p)}(\vec{\xi}_0,\vec{\xi}_1)\Delta_1\mathbf{w}(\vec{\xi}_0) + \delta_{Ni}\Lambda^{(v_p)}(\vec{\xi}_{N+1},\vec{\xi}_N)\Delta_1\mathbf{w}(\vec{\xi}_N)\right),
\end{aligned} \tag{246}$$

where identical definitions hold for other computational directions. Again, the $\mathbf{U}(\vec{\xi}_{N+1})$ and $\mathbf{U}(\vec{\xi}_0)$ states are specified by the boundary condition at the face and are not necessarily the same as those used for the inviscid boundary conditions. The boundary fluxes, $\hat{\mathbf{f}}_1^{(v_p,BC)}$, are boundary condition dependent as well.

B.3 FORM OF BOUNDARY PENALTIES FOR THE GRADIENT OF THE ENTROPY VARIABLES

The gradient of the entropy variables is penalized at domain boundaries through $\hat{\mathbf{g}}_l^{(BC,\boldsymbol{\Theta})}$. Since $\hat{\mathbf{g}}_l^{(BC,\boldsymbol{\Theta})}(\vec{\xi}_a) = \hat{\mathbf{g}}_l^{\boldsymbol{\Theta}}(\vec{\xi}_a)\chi^{(BC)}(\vec{\xi}_a)$, we already discussed $\hat{\mathbf{g}}_l^{(BC,\boldsymbol{\Theta})}$ in Eq. (247), but we repeat it here for convenience. For all fixed $1 \leq j,k \leq N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$\begin{aligned}
\hat{\mathbf{g}}_1^{(BC,\boldsymbol{\Theta})}(\vec{\xi}_i) &= \frac{\chi^{(BC)}(\vec{\xi}_i)}{2}\left(\delta_{1i}\Delta_1\mathbf{w}(\vec{\xi}_0) + \delta_{Ni}\Delta_1\mathbf{w}(\vec{\xi}_N)\right), \\
\Delta_1\mathbf{w}(\vec{\xi}_{ijk}) &= \mathbf{w}(\vec{\xi}_{i+1jk}) - \mathbf{w}(\vec{\xi}_{ijk}),
\end{aligned} \tag{247}$$

where we use similar definitions in each computational direction. The values $\mathbf{w}(\vec{\xi}_{N+1})$ and

$\mathbf{w}(\vec{\xi}_0)$ are determined by the boundary condition.

B.4 FORM OF FIRST-ORDER BRENNER BOUNDARY PENALTIES

The only remaining boundary penalty term in the high-order positivity-preserving scheme given by Eq. (178) is the $\hat{\mathbf{g}}_l^{(BC,AD_1)}$ contribution. Since $\hat{\mathbf{g}}_l^{(BC,AD_1)}(\vec{\xi}_i) = \chi^{(BC)}(\vec{\xi}_i)\hat{\mathbf{g}}_l^{(AD_1)}(\vec{\xi}_i)$, $\hat{\mathbf{g}}_l^{(BC,AD_1)}$ was already specified by Eq. (115), but we repeat it here for convenience. For all fixed $1 \leq j,k \leq N$ and $\vec{\xi}_i = \vec{\xi}_{ijk}$ we have

$$\begin{aligned}
\hat{\mathbf{g}}_l^{(BC,AD_1)}(\vec{\xi}_i) &= \chi^{(BC)}(\vec{\xi}_i)\left(\hat{\mathbf{g}}_l^{(BC,AD_1)}(\vec{\xi}_1)\delta_{1i} + \hat{\mathbf{g}}_l^{(BC,AD_1)}(\vec{\xi}_N)\delta_{Ni}\right), \\
\hat{\mathbf{g}}_l^{(BC,AD_1)}(\vec{\xi}_1) &= \frac{c_\nu^{(B)}(\mathbf{U}(\vec{\xi}_0),\mathbf{U}(\vec{\xi}_1),\hat{\vec{\mathbf{a}}}^1(\vec{\xi}_0))}{\boldsymbol{J}(\vec{\xi}_1)}\frac{\boldsymbol{\nu}(\vec{\xi}_0)-\boldsymbol{\nu}(\vec{\xi}_1)}{\mathcal{P}_{1,1}},
\end{aligned} \tag{248}$$

with identical definitions in other computational directions. Note that the boundary conditions are imposed by the states $\mathbf{U}(\vec{\xi}_{N+1})$ and $\mathbf{U}(\vec{\xi}_0)$.

B.5 PENALTIES FOR SPECIFIC BOUNDARY CONDITIONS

Now that we have discussed the general form of the boundary penalties, we will discuss specific boundary conditions. Except for at the no-slip wall boundaries, we used $\hat{\mathbf{g}}_l^{(BC,\boldsymbol{\Theta})}(\vec{\xi}_a) = 0$. Although $\hat{\mathbf{g}}_l^{(BC,AD_p)}$ is identical to $\hat{\mathbf{g}}_l^{(BC,v)}$ in general form, they are not formed identically for every boundary condition. Except for at the no-slip wall boundaries, $\hat{\mathbf{g}}_l^{(BC,AD_p)}$ uses $\hat{\mathbf{f}}_l^{(AD_p,BC)}(\vec{\xi}_a) = 0$ and the same $\mathbf{U}(\vec{\xi}_{N+1})$ and $\mathbf{U}(\vec{\xi}_0)$ states used by $\hat{\mathbf{g}}_l^{(BC,v)}$.

B.5.1 NO BOUNDARY CONDITION

When "no boundary condition" penalties are used, we have $\mathbf{U}(\vec{\xi}_{N+1}) = \mathbf{U}(\vec{\xi}_N)$ and $\mathbf{U}(\vec{\xi}_0) = \mathbf{U}(\vec{\xi}_1)$ for the inviscid penalties (see Eq. (245)), high-order viscous penalties (see Eq. (246)), and for the first-order Brenner penalties (see Eq. (248)). Furthermore, $\hat{\mathbf{f}}_l^{(v,BC)}(\vec{\xi}_a) = 0$ at such boundaries.

B.5.2 SLIP WALL

Fix $1 \leq j,k \leq N$ and let $\vec{\xi}_i = \vec{\xi}_{ijk}$. We use the entropy stable inviscid penalties described in [5, 76]. The slip wall boundary condition uses the state $\boldsymbol{\nu}(\vec{\xi}_{N+1}) = \left[\begin{array}{ccc} \boldsymbol{\rho} & \vec{\boldsymbol{V}} - 2(\vec{\boldsymbol{V}} \cdot \frac{\hat{\vec{\mathbf{a}}}^1}{\|\hat{\vec{\mathbf{a}}}^1\|})\frac{\hat{\vec{\mathbf{a}}}^1}{\|\hat{\vec{\mathbf{a}}}^1\|} & \mathbf{T} \end{array} \right]^\top_{\vec{\xi}_N}$ with a similar definition for $\boldsymbol{\nu}(\vec{\xi}_0)$. The states $\boldsymbol{\nu}(\vec{\xi}_{N+1})$ and $\boldsymbol{\nu}(\vec{\xi}_0)$ are used for the inviscid penalties (see Eq. (245)), high-order viscous penalties (see Eq. (246)), and for the first-order Brenner penalties (see Eq. (248)). The physical viscous boundary penalties use

$$\begin{aligned} \hat{\mathbf{f}}_1^{(v,BC)}(\vec{\xi}_0) &= \sum_{m=1}^{3} \hat{a}_m^1(\vec{\xi}_1) \sum_{j=1}^{3} [c_{m,j}^{(v)}]_{\boldsymbol{\nu}(\vec{\xi}_0)} \boldsymbol{\Theta}_{x^j}(\vec{\xi}_1), \\ \hat{\mathbf{f}}_1^{(v,BC)}(\vec{\xi}_{N+1}) &= -\sum_{m=1}^{3} \hat{a}_m^1(\vec{\xi}_N) \sum_{j=1}^{3} [c_{m,j}^{(v)}]_{\boldsymbol{\nu}(\vec{\xi}_{N+1})} \boldsymbol{\Theta}_{x^j}(\vec{\xi}_N). \end{aligned} \tag{249}$$

B.5.3 ENTROPY STABLE ADIABATIC NO-SLIP WALL

Fix $1 \leq j,k \leq N$ and let $\vec{\xi}_i = \vec{\xi}_{ijk}$. The adiabatic no-slip wall boundary condition is formed in the entropy stable manner described in [5]. The inviscid boundary penalties are formed in exactly the same manner described in Section B.5.2 for slip walls. The viscous penalties for the no-slip wall boundary condition use the states $\boldsymbol{\nu}(\vec{\xi}_{N+1}) = \left[\begin{array}{ccc} \boldsymbol{\rho} & -\vec{\boldsymbol{V}} & \mathbf{T} \end{array} \right]^\top_{\vec{\xi}_N}$ and $\boldsymbol{\nu}(\vec{\xi}_0) = \left[\begin{array}{ccc} \boldsymbol{\rho} & -\vec{\boldsymbol{V}} & \mathbf{T} \end{array} \right]^\top_{\vec{\xi}_1}$. These states are used in forming the penalties for the gradient of the entropy variables (see Eq. (247)), the high-order viscous boundary penalties (see Eq. (246)) and the first-order Brenner penalties (see Eq. (248)). The high-order viscous boundary penalties also make use of manufactured gradients of the entropy variables at the wall face given by

$$\boldsymbol{\Theta}_{x^j}(\vec{\xi}_{N+1}) = \left[\frac{\partial W}{\partial \nu} \right]_{\boldsymbol{\nu}(\vec{\xi}_{N+1})} \text{diag} \left[\begin{array}{ccccc} -1 & 1 & 1 & 1 & -1 \end{array} \right] \left[\frac{\partial \nu}{\partial W} \right]_{\boldsymbol{\nu}(\vec{\xi}_N)} \boldsymbol{\Theta}_{x^j}(\vec{\xi}_N), \tag{250}$$

with a similar definition for $\mathbf{\Theta}_{x^j}(\vec{\xi}_0)$. Using these boundary states, we have

$$\begin{aligned}\hat{\mathbf{f}}_1^{(v_p,BC)}(\vec{\xi}_0) &= \sum_{m=1}^{3} \hat{a}_m^1(\vec{\xi}_1) \sum_{j=1}^{3} [c_{m,j}^{(v_p)}]_{\boldsymbol{\nu}(\vec{\xi}_0)} \mathbf{\Theta}_{x^j}(\vec{\xi}_0), \\ \hat{\mathbf{f}}_1^{(v_p,BC)}(\vec{\xi}_{N+1}) &= \sum_{m=1}^{3} \hat{a}_m^1(\vec{\xi}_N) \sum_{j=1}^{3} [c_{m,j}^{(v_p)}]_{\boldsymbol{\nu}(\vec{\xi}_{N+1})} \mathbf{\Theta}_{x^j}(\vec{\xi}_{N+1}).\end{aligned} \tag{251}$$

B.5.4 CONSTANT PRESSURE FACE

Assume we know that on the face defined by fixed $i = N$ and $1 \leq j, k \leq N$ the pressure should be a constant value P_c. let $\vec{\xi}_{N+1} = \vec{\xi}_{N+1jk}$. Then we use the state $\boldsymbol{\nu}(\vec{\xi}_{N+1}) = \left[\begin{array}{ccc} \frac{P_c}{R\mathbf{T}} & \vec{\boldsymbol{V}} & \mathbf{T} \end{array} \right]_{\vec{\xi}_N}^{\top}$ for the inviscid penalties (see Eq. (245)), high-order viscous penalties (see Eq. (246)), and for the first-order Brenner penalties (see Eq. (248)). Furthermore, $\hat{\mathbf{f}}_l^{(v,BC)}(\vec{\xi}_{N+1}) = 0$ at such boundaries.

APPENDIX C

CHOLESKY DECOMPOSITION BASED IDENTITIES FOR $\frac{\partial^2 \mathcal{S}}{\partial U^2}$

We record here for reference some identities that can be derived from the Cholesky decomposition of $\frac{\partial^2 \mathcal{S}}{\partial U^2}$:

$$
\begin{aligned}
\frac{\partial^2 \mathcal{S}}{\partial U^2} =& LDL^\top \\
L =& \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
-2(\gamma-1)^2 V_1 \frac{\|\vec{\boldsymbol{V}}\|^2}{c_1} & 1 & 0 & 0 & 0 \\
-2(\gamma-1)^2 V_2 \frac{\|\vec{\boldsymbol{V}}\|^2}{c_1} & \frac{4R^2(\gamma-1)\gamma T^2 V_1 V_2}{c_1 c_2} & 1 & 0 & 0 \\
-2(\gamma-1)^2 V_3 \frac{\|\vec{\boldsymbol{V}}\|^2}{c_1} & \frac{4R^2(\gamma-1)\gamma T^2 V_1 V_3}{c_1 c_2} & \frac{4R(\gamma-1)\gamma T V_2 V_3}{c_3} & 1 & 0 \\
2(\gamma-1)\frac{-2RT+(\gamma-1)\|\vec{\boldsymbol{V}}\|^2}{c_1} & \frac{-2R(\gamma-1)TV_1 c_6}{c_1 c_2} & -2(\gamma-1)V_2 \frac{c_6}{c_3} & -\frac{c_5}{c_4} & 1
\end{bmatrix}, \\
D =& \operatorname{diag}\left[\frac{c_1}{4R(\gamma-1)T^2\rho} \quad \frac{c_1+(\gamma-1)V_1^2 4R\gamma T}{T\rho c_1} \quad \frac{Rc_3}{c_1 c_2 \rho} \quad R\frac{c_1+\left(\frac{c_4-c_1 c_3}{c_3}\right)}{c_1 c_2 \rho} \quad \frac{R}{\gamma-1}\frac{\rho}{(\rho E)^2 \frac{\gamma}{\gamma-1}-\left(\rho\frac{\|\vec{\boldsymbol{V}}\|^2}{2}\right)^2} \right], \\
c_1 =& 4R^2\gamma T^2 + (\gamma-1)^2\|\vec{\boldsymbol{V}}\|^4 > 0, \quad c_2 = RT\left(1+\frac{4R(\gamma-1)\gamma T V_1^2}{c_1}\right) > 0, \\
c_3 =& c_1 + 4R(\gamma-1)\gamma T(V_1^2+V_2^2) > 0, \\
c_4 =& c_1 c_3 + 4R(\gamma-1)\gamma T\left(c_1(V_1^2+V_3^2) + 4R(\gamma-1)\gamma T V_1^2\|\vec{\boldsymbol{V}}\|^2\right) > 0, \\
c_5 =& 2(\gamma-1)V_3 c_6\left(c_1 + 4R(\gamma-1)\gamma T V_1^2\right), \quad c_6 = 2R\gamma T + (\gamma-1)\|\vec{\boldsymbol{V}}\|^2 > 0,
\end{aligned}
\tag{252}
$$

where we have assumed positive density and temperature and it is clear that the diagonal entries of D are all strictly positive. We label the diagonal entries of D as $d_1 = \frac{c_1}{4R(\gamma-1)T^2\rho}$, $d_2 = \frac{c_1+(\gamma-1)V_1^2 4R\gamma T}{T\rho c_1}$, $\ldots$ etc.

Let $A = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \end{bmatrix}^\top$ be an arbitrary real array. Denote $L^\top A = \begin{bmatrix} La_1 & La_2 & La_3 & La_4 & La_5 \end{bmatrix}^\top$ and notice that $a_5 = La_5$. The following inequality is an

immediate consequence of Eq. (252):

$$A^\top \frac{\partial^2 \mathcal{S}}{\partial U^2} A = (L^\top A)^\top D L^\top A = \sum_{k=1}^{5} L a_k^2 d_k \geq d_5 a_5^2 = a_5^2 \frac{R}{\gamma - 1} \frac{\rho}{(\rho E)^2 \frac{\gamma}{\gamma - 1} - \left(\rho \frac{\| \vec{\boldsymbol{V}} \|^2}{2}\right)^2}. \quad (253)$$

We would like to have bounds similar to (253) for a_i $i = 1, 2, 3, 4$. Unfortunately, such bounds are not obvious for the decomposition given by Eq. (252). However, this can easily be remedied by simply changing the basis ordering of $\frac{\partial^2 \mathcal{S}}{\partial U^2}$ and obtaining a new $LDL^\top$ decomposition. This is the approach we now take.

We briefly recall some ideas from linear algebra. Let $\frac{\partial^2 \mathcal{S}}{\partial U^2}^c{}_{ab}$ be the matrix obtained from $\frac{\partial^2 \mathcal{S}}{\partial U^2}$ after interchanging columns a and b. Furthermore, let $\frac{\partial^2 \mathcal{S}}{\partial U^2}_{ab}$ be the matrix obtained from $\frac{\partial^2 \mathcal{S}}{\partial U^2}^c{}_{ab}$ after interchanging rows a and b (the order of interchanging does not matter). Let A_{ab} be the array obtained from A by interchanging the a and b components. For example, $A_{45} = \begin{bmatrix} a_1 & a_2 & a_3 & a_5 & a_4 \end{bmatrix}^\top$. Given that the ith component of the matrix multiplication Mx can be written as $\sum_j M_{ij} x_j$, it is not difficult to see that $\frac{\partial^2 \mathcal{S}}{\partial U^2}^c{}_{ab} A_{ab} = \frac{\partial^2 \mathcal{S}}{\partial U^2} A$. Hence, $\frac{\partial^2 \mathcal{S}}{\partial U^2}_{ab} A_{ab}$ has the same collection of components as $\frac{\partial^2 \mathcal{S}}{\partial U^2} A$ but in a different order and interchanging the a component with the b component of $\frac{\partial^2 \mathcal{S}}{\partial U^2}_{ab} A_{ab}$ will give $\frac{\partial^2 \mathcal{S}}{\partial U^2} A$. Furthermore, $A_{ab}^\top \frac{\partial^2 \mathcal{S}}{\partial U^2}_{ab} A_{ab} = A^\top \frac{\partial^2 \mathcal{S}}{\partial U^2} A$ for all a and b interchanges $(a, b = 1, 2, 3, 4, 5)$.

We label the matrices in the Cholesky decomposition of $\frac{\partial^2 \mathcal{S}}{\partial U^2}_{ab}$ as $D_{ab} = \text{diag} \begin{bmatrix} d_1^{ab} & d_2^{ab} & \dots \end{bmatrix}$ and L_{ab}. It would be tedious to explicitly write all D_{ab} and L_{ab}; hence, we only list the following:

$$d_5^{15} = \frac{R}{\rho}, \quad d_5^{25} = \frac{R}{P + \rho V_1^2}, \quad d_5^{35} = \frac{R}{P + \rho V_2^2}, \quad d_5^{45} = \frac{R}{P + \rho V_3^2}, \quad (254)$$

from which we obtain the full set of bounds

$$\begin{aligned} A^\top \frac{\partial^2 \mathcal{S}}{\partial U^2} A &= (L_{15}^\top A_{15})^\top D_{15} L_{15}^\top A_{15} = \sum_{k=1}^{5} (L a_{15})_k^2 d_k^{15} \geq d_5^{15} a_1^2 = a_1^2 \frac{R}{\rho}, \\ A^\top \frac{\partial^2 \mathcal{S}}{\partial U^2} A &\geq a_{i+1}^2 \frac{R}{P + \rho V_i^2}, \quad i = 1, 2, 3, \\ A^\top \frac{\partial^2 \mathcal{S}}{\partial U^2} A &\geq a_5^2 \frac{R}{\gamma - 1} \frac{\rho}{(\rho E)^2 \frac{\gamma}{\gamma - 1} - \left(\rho \frac{\| \vec{\boldsymbol{V}} \|^2}{2}\right)^2} = a_5^2 \frac{R\rho}{P^2 \gamma + P \rho \| \vec{\boldsymbol{V}} \|^2 \gamma + \left(\rho \frac{\| \vec{\boldsymbol{V}} \|^2}{2}\right)^2}. \end{aligned} \quad (255)$$

VITA

Johnathon Keith Upperman
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529

PREVIOUS DEGREES

M.A. Education, College of William & Mary, Williamsburg, VA (2013)
B.S. Mathematics, College of William & Mary, Williamsburg, VA (2012)

PUBLICATIONS

J. Upperman and N. K. Yamaleev, "Entropy stable artificial dissipation based on Brenner regularization of the Navier-Stokes equations," *Journal of Computational Physics*, Vol. 393, pp. 74–91, 2019.

J. Upperman and C. R. Vinroot, "A weight statistic and partial order on products of m-cycles," *Discrete Mathematics*, Vol. 315-316, pp. 9–17, 2014.

Typeset using LaTeX.