

A compact high-order and positivity-preserving active flux method for compressible Navier–Stokes equations on Cartesian meshes

Junming Duan* Wasilij Barsukow[†] Christian Klingenberg[‡]

Abstract

This paper develops a compact fourth-order positivity-preserving active flux (AF) method for the one- and two-dimensional compressible Navier–Stokes equations on Cartesian meshes. The method retains the cell averages and shared point values of the standard third-order AF method as its degrees of freedom. To avoid the order reduction that can arise when diffusion is discretized using operators from the standard third-order AF method, while maintaining compactness, the divergence of the viscous flux is discretized directly using compact fourth-order operators. For the inviscid part, incorporating a downwind point value into the biased stencil yields fourth-order accuracy. A monolithic flux limiting blends high-order total numerical fluxes with low-order positivity-preserving counterparts, treating the inviscid and viscous fluxes jointly while maintaining local conservation. Together with a scaling limiter for point values, this procedure preserves density and pressure positivity for both cell averages and point values. Numerical experiments demonstrate fourth-order convergence, positivity preservation, and accurate resolution of shocks and viscous flow structures. For the two-dimensional viscous shock tube, a comparison with a discontinuous Galerkin method shows improved computational efficiency in resolving complex interaction of shock waves and boundary layers.

Keywords: Active flux; Navier–Stokes equations; high-order accuracy; finite volume method; compact scheme; positivity-preserving

1 Introduction

Computational fluid dynamics is central to weather prediction, pollutant transport, biomedical applications, and aerodynamic design [1], with accurate modeling of viscous effects essential for resolving boundary layers and predicting aerodynamic drag. The development of high-order numerical methods, which can provide enhanced accuracy and computational

*Corresponding author. School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Guangdong, 518172, P.R. China. Email: duanjunming@cuhk.edu.cn.

[†]Institut de Mathématiques de Bordeaux (IMB), CNRS UMR 5251, University of Bordeaux, Talence, 33405, France. Email: wasilij.barsukow@math.u-bordeaux.fr.

[‡]Institute of Mathematics, University of Würzburg, Emil-Fischer-Straße 40, Würzburg, 97074, Germany. Email: christian.klingenberg@uni-wuerzburg.de.

efficiency [2], is therefore of fundamental importance in both scientific research and engineering practice. In this work, we focus on the development of a compact high-order numerical method for solving the d -dimensional compressible Navier–Stokes (NS) equations

$$\partial_t \mathbf{U} + \sum_{\ell=1}^d \partial_{x_\ell} (\mathbf{F}_\ell^c(\mathbf{U}) - \mathbf{F}_\ell^v(\mathbf{U}, \nabla \mathbf{U})) = \mathbf{0}, \quad (1)$$

where $\mathbf{U} = (\rho, \rho \mathbf{u}, E)^\top$ denotes the vector of conservative variables, and $\mathbf{F}_\ell^c$ and $\mathbf{F}_\ell^v$ denote, respectively, the inviscid and viscous fluxes in the x_ℓ -direction,

$$\mathbf{F}_\ell^c = \begin{pmatrix} \rho u_\ell \\ \rho \mathbf{u} u_\ell + p \mathbf{e}_\ell \\ (E + p) u_\ell \end{pmatrix}, \quad \mathbf{F}_\ell^v = \begin{pmatrix} 0 \\ \boldsymbol{\tau} \mathbf{e}_\ell \\ \mathbf{u} \cdot (\boldsymbol{\tau} \mathbf{e}_\ell) - q_\ell \end{pmatrix}, \quad q_\ell = \mathbf{q} \cdot \mathbf{e}_\ell. \quad (2)$$

Here, ρ is the density, $\mathbf{u} = (u_1, \dots, u_d)^\top$ is the velocity vector, p is the pressure, and $E = \rho e + \frac{1}{2} \rho \|\mathbf{u}\|^2$ is the total energy, with e the specific internal energy. The vector $\mathbf{e}_\ell$ is the unit vector with all zeros except for the ℓ th component. For a Newtonian fluid satisfying Fourier’s law of heat conduction, the viscous stress tensor and Fourier heat flux are given as follows,

$$\boldsymbol{\tau} = \mu \left(\nabla \mathbf{u} + \nabla \mathbf{u}^\top - \frac{2}{3} (\nabla \cdot \mathbf{u}) \mathbf{I}_d \right), \quad \mathbf{q} = -\kappa \nabla T, \quad (3)$$

with $\mu > 0$ the constant dynamic viscosity. In the nondimensional form, the temperature is $T = p/\rho$ and the thermal conductivity is $\kappa = \frac{\gamma \mu}{(\gamma-1) \text{Pr}}$, where $\text{Pr} > 0$ is the Prandtl number, which characterizes the relative importance of momentum and thermal diffusion. To close the system (1), the ideal-gas equation of state $p = (\gamma - 1)\rho e$ is used, with $\gamma > 1$ the ratio of specific heats.

A variety of high-order accurate numerical methods have been developed for the compressible NS equations, including compact finite difference methods [3], $P_N P_M$ methods [4], discontinuous Galerkin (DG) methods [5], and flux reconstruction (FR) methods [6, 7], see also the review [2]. High-order methods can provide high resolution for smooth and multiscale flow structures on relatively coarse meshes. For large-scale simulations, the locality and compactness of the discretization are also important, since localized data dependencies reduce communication and facilitate parallel implementation on modern high-performance computing architectures, which is one of the reasons why DG methods are attractive. However, the number of local degrees of freedom (DoFs) in DG methods grows rapidly with the polynomial degree. For example, a tensor-product polynomial space of degree k in each coordinate direction contains $(k+1)^d$ coefficients per conservative variable in d dimensions, leading to quadratic and cubic growth in memory requirements and computational costs in 2D and 3D, respectively. These considerations motivate the development of compact high-order methods with fewer DoFs.

The active flux (AF) method provides a compact finite volume framework, with fewer DoFs than DG methods. Although its number of DoFs still grows quadratically (cubically) with the polynomial degree, the comparison in [8] shows higher computational efficiency. Originally developed for hyperbolic conservation laws [9, 10, 11], inspired by [12], AF methods evolve cell averages together with point values shared by neighboring cells, which reduces the number of DoFs. The cell average follows a conservative finite volume update, and the point value provides additional interface information for high-order approximations. As

described below, the AF method is local because point values are evolved using compact finite difference operators, without quadrature or inverting a mass matrix. All these factors contribute to the efficiency of the AF methods.

In the original fully-discrete formulation [9, 10, 11], the point values are evolved by exact or approximate evolution operators, which are difficult to construct for general cases, especially to achieve high-order accuracy [13, 14, 15, 16]. A method-of-lines formulation was subsequently introduced in [17], where the evolution equations for both cell averages and point values are first discretized in space and then integrated in time using Runge–Kutta (RK) methods. Upwinding is incorporated in the point value update using Jacobian splitting [17] or flux-vector splitting [18], which has been extended to higher-order accuracy [19, 20], polygonal meshes [21], and magnetohydrodynamics [22, 23].

Compared with the hyperbolic case, AF discretizations of diffusion terms are relatively less developed. One approach reformulates the diffusion terms as a first-order hyperbolic or relaxation system with auxiliary variables [24, 25]. The pseudo-time formulation in [24] uses inner iterations, whereas [25] uses a diagonally implicit RK method. Another discretizes the diffusion terms directly using auxiliary gradient approximations [26], but it is second-order accurate. In [27], a compact fourth-order AF method is constructed for solving convection–diffusion equations with degenerate diffusion coefficients, where the diffusion term is discretized based on a rewriting into a doubly conservative form of the diffusion potential.

A further challenge is to preserve physically admissible states in the presence of strong shocks or near-vacuum states, where high-order methods are prone to producing negative density or pressure due to unphysical oscillations. This issue has been considered in the AF methods [28, 18, 29, 30, 31, 23] and other methods for the NS equations [32, 33, 34, 35], to name a few. For the compressible NS equations, the viscous flux depends on velocity and temperature gradients and contains coupled stress components in multiple dimensions, making positivity-preserving (PP) limiting more delicate.

This work develops a compact fourth-order PP AF method for the one- and two-dimensional compressible NS equations on Cartesian meshes. The choice of fourth-order operators is motivated by the order reduction to second-order accuracy in two dimensions, as analyzed and numerically verified in [25] for the AF method using the standard third-order operators. This paper discretizes the viscous flux divergence directly using compact fourth-order operators, while retaining the DoFs of the standard third-order AF method. For the inviscid part, a biased stencil incorporating a downwind point value achieves fourth-order accuracy with the same DoFs as the standard third-order AF method. To preserve density and pressure positivity, we construct a monolithic convex limiter for the numerical flux. The high-order inviscid and viscous contributions are limited jointly against a low-order PP flux based on the local Lax–Friedrichs (LLF) scheme in [36]. A decomposition into directional intermediate states yields a convex representation of the cell average update and a sufficient CFL condition for positivity preservation in the explicit strong-stability-preserving (SSP)-RK method. A scaling limiter enforces the PP property for point values. Numerical experiments assess the accuracy and PP property of the proposed method. Its computational efficiency is examined through a comparison with the DG method [34] for the viscous shock tube. In this test, AF is approximately ten times faster at a comparable level of accuracy and with the same number of DoFs.

The remainder of this paper is organized as follows. Section 2 presents the 1D and 2D compact AF discretizations. Section 3 presents the first-order PP LLF scheme and the PP

limiting. Numerical experiments are conducted in Section 4, followed by concluding remarks in Section 5.

2 The compact AF formulation

2.1 The 1D formulation

The 1D NS equations can be written as

$$\partial_t \mathbf{U} + \partial_x \mathbf{F}^c(\mathbf{U}) = \partial_x \mathbf{F}^v(\mathbf{U}, \partial_x \mathbf{Z}) = \partial_x (\mathbf{K}(\mathbf{W}) \partial_x \mathbf{Z}), \quad (4)$$

where

$$\mathbf{F}^c = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ (E + p)u \end{pmatrix}, \quad \mathbf{K}(\mathbf{W}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{4}{3}\mu & 0 \\ 0 & \frac{4}{3}\mu u & \frac{\gamma\mu}{(\gamma-1)\text{Pr}} \end{pmatrix},$$

with

$$\mathbf{W} = (\rho, u, p)^\top, \quad \mathbf{Z} = (0, u, T)^\top. \quad (5)$$

Assume that a 1D computational domain is divided into N uniform cells $I_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ with the cell size $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$, $i = 1, \dots, N$. The cell centers are denoted by $x_i = (x_{i-\frac{1}{2}} + x_{i+\frac{1}{2}})/2$. The DoFs are the cell averages and shared interface point values of conservative variables,

$$\bar{\mathbf{U}}_i(t) = \frac{1}{\Delta x} \int_{I_i} \mathbf{U}_h(x, t) dx, \quad \mathbf{U}_{i+1/2}(t) = \mathbf{U}_h(x_{i+\frac{1}{2}}, t),$$

with $\mathbf{U}_h(x, t)$ the numerical solution. The AF method for solving (4) can be constructed as follows,

$$\frac{d\bar{\mathbf{U}}_i}{dt} + \frac{1}{\Delta x} \left(\mathbf{F}_{i+\frac{1}{2}}^c - \mathbf{F}_{i-\frac{1}{2}}^c \right) = \frac{1}{\Delta x} \left(\mathbf{K}(\mathbf{W}_{i+\frac{1}{2}})(\mathbf{Z}_x)_{i+\frac{1}{2}} - \mathbf{K}(\mathbf{W}_{i-\frac{1}{2}})(\mathbf{Z}_x)_{i-\frac{1}{2}} \right), \quad (6a)$$

$$\frac{d\mathbf{U}_{i+\frac{1}{2}}}{dt} + \mathcal{D}_+(\mathbf{F}^{c,+})_{i+\frac{1}{2}} + \mathcal{D}_-(\mathbf{F}^{c,-})_{i+\frac{1}{2}} = \mathcal{D}^2(\mathbf{K}(\mathbf{W}), \mathbf{Z})_{i+\frac{1}{2}}, \quad (6b)$$

$$(\mathbf{Z}_x)_{i+\frac{1}{2}} = \mathcal{D}_c(\mathbf{Z})_{i+\frac{1}{2}}, \quad (6c)$$

where the cell average update (6a) follows the standard finite volume method, which ensures conservation. The corresponding $\mathbf{W}_i$, $\mathbf{Z}_i$, and $\mathbf{F}_i^{c,\pm} = \mathbf{F}^{c,\pm}(\mathbf{U}_i)$ are then evaluated from $\mathbf{U}_i$. In the point value update (6b), the inviscid part uses the flux vector splitting introduced in [18], and the viscous part is based on a compact discretization. To be specific, the inviscid flux is split using the local Lax–Friedrichs flux vector splitting (FVS), given by

$$\mathbf{F}^{c,\pm} = \frac{1}{2} \left(\mathbf{F}^c(\mathbf{U}) \pm \alpha_{i+\frac{1}{2}} \mathbf{U} \right), \quad \lambda \left(\frac{\partial \mathbf{F}^{c,+}}{\partial \mathbf{U}} \right) \geq 0, \quad \lambda \left(\frac{\partial \mathbf{F}^{c,-}}{\partial \mathbf{U}} \right) \leq 0,$$

where λ denotes the eigenvalue, and the numerical viscosity coefficient $\alpha_{i+\frac{1}{2}}$ is fixed across the spatial stencil and determined by

$$\alpha_{i+\frac{1}{2}} = \max\{\varrho(\mathbf{U}_{i-\frac{1}{2}}), \varrho(\mathbf{U}_i), \varrho(\mathbf{U}_{i+\frac{1}{2}}), \varrho(\mathbf{U}_{i+1}), \varrho(\mathbf{U}_{i+\frac{3}{2}})\},$$

with ϱ the spectral radius of the Jacobian matrix $\partial \mathbf{F}^c / \partial \mathbf{U}$. Overbars denote cell averages, whereas quantities without overbars are point values. The cell-centered conservative variables are reconstructed through Simpson's rule

$$\mathbf{U}_i = \frac{1}{4}(6\overline{\mathbf{U}}_i - (\mathbf{U}_{i-\frac{1}{2}} + \mathbf{U}_{i+\frac{1}{2}})),$$

which is a 4th-order approximation for smooth solutions. Each component of the split flux is discretized using the following biased operators. Here z denotes a generic scalar component of the quantity being differentiated, distinct from the vector $\mathbf{Z}$. Note that z_i is a cell-center point value, not a cell average. The operators are obtained by differentiating the corresponding cubic reconstructions based on $\{z_{i-\frac{1}{2}}, z_i, z_{i+\frac{1}{2}}, z_{i+1}\}$ or $\{z_i, z_{i+\frac{1}{2}}, z_{i+1}, z_{i+\frac{3}{2}}\}$ at $x_{i+\frac{1}{2}}$,

$$\mathcal{D}_+(z)_{i+\frac{1}{2}} = \frac{1}{3\Delta x} \left(z_{i-\frac{1}{2}} - 6z_i + 3z_{i+\frac{1}{2}} + 2z_{i+1} \right), \quad (7a)$$

$$\mathcal{D}_-(z)_{i+\frac{1}{2}} = \frac{1}{3\Delta x} \left(-2z_i - 3z_{i+\frac{1}{2}} + 6z_{i+1} - z_{i+\frac{3}{2}} \right). \quad (7b)$$

Biased stencils provide upwinding for the split inviscid fluxes, while their symmetric average gives a central difference approximation for the viscous gradients as there is no upwinding mechanism for the diffusion part. The first-order derivative in (6c) is obtained by using the central operator defined by

$$\mathcal{D}_c(z)_{i+\frac{1}{2}} = \frac{1}{2} (\mathcal{D}_+(z) + \mathcal{D}_-(z))_{i+\frac{1}{2}} = \frac{1}{6\Delta x} \left(z_{i-\frac{1}{2}} - 8z_i + 8z_{i+1} - z_{i+\frac{3}{2}} \right). \quad (8)$$

The 4th-order compact operator of the second derivative $\mathcal{D}^2$ is adapted from [37],

$$\mathcal{D}^2(\mathbf{K}, \mathbf{Z})_{i+\frac{1}{2}} = \frac{4}{\Delta x^2} \sum_{m=-2}^2 \sum_{n=-2}^2 b_{mn} \mathbf{K} \left(\mathbf{W}_{i+\frac{m+1}{2}} \right) \mathbf{Z}_{i+\frac{n+1}{2}}. \quad (9)$$

Here the coefficients b_{mn} are obtained from the order conditions, a sparsity requirement, and symmetry. More precisely, let

$$B = \begin{bmatrix} 0 & 0 \\ 0 & M \end{bmatrix} - \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}, \quad M = \begin{bmatrix} \frac{1}{8} & -\frac{1}{6} & \frac{1}{24} & 0 \\ -\frac{1}{6} & -\frac{3}{8} & \frac{2}{3} & -\frac{1}{8} \\ \frac{1}{8} & -\frac{2}{3} & \frac{3}{8} & \frac{1}{6} \\ 0 & -\frac{1}{24} & \frac{1}{6} & -\frac{1}{8} \end{bmatrix}, \quad (10)$$

and the rows and columns of B are ordered by $\{-2, -1, 0, 1, 2\}$, then b_{mn} denotes the entry associated with row index m and column index n in this ordering. Note that the matrix $\mathbf{K}(\mathbf{W}_{i+\frac{m+1}{2}})$ is evaluated separately at every stencil point, and the right-hand side in (6b) discretizes $\partial_x(\mathbf{K} \partial_x \mathbf{Z})$ directly.

2.2 The 2D formulation

The 2D compressible NS equations can be written as

$$\partial_t \mathbf{U} + \sum_{\ell=1}^2 \partial_{x_\ell} \mathbf{F}_\ell^c = \sum_{\ell=1}^2 \partial_{x_\ell} \mathbf{F}_\ell^v = \sum_{\ell, m=1}^2 \partial_{x_\ell} (\mathbf{K}_{\ell m}(\mathbf{W}) \partial_{x_m} \mathbf{Z}), \quad (11)$$

with $\mathbf{U} = (\rho, \rho u, \rho v, E)^\top$, $\mathbf{W} = (\rho, u, v, p)^\top$, $\mathbf{Z} = (0, u, v, T)^\top$, and $(x_1, x_2) = (x, y)$. The viscous fluxes can be expressed as

$$\mathbf{F}_1^v = \mathbf{K}_{11}(\mathbf{W})\partial_x \mathbf{Z} + \mathbf{K}_{12}(\mathbf{W})\partial_y \mathbf{Z}, \quad \mathbf{F}_2^v = \mathbf{K}_{21}(\mathbf{W})\partial_x \mathbf{Z} + \mathbf{K}_{22}(\mathbf{W})\partial_y \mathbf{Z}.$$

The four 4×4 matrices are

$$\begin{aligned} \mathbf{K}_{11}(\mathbf{W}) &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{4}{3}\mu & 0 & 0 \\ 0 & 0 & \mu & 0 \\ 0 & \frac{4}{3}\mu u & \mu v & \kappa \end{pmatrix}, & \mathbf{K}_{22}(\mathbf{W}) &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 \\ 0 & 0 & \frac{4}{3}\mu & 0 \\ 0 & \mu u & \frac{4}{3}\mu v & \kappa \end{pmatrix}, \\ \mathbf{K}_{12}(\mathbf{W}) &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{2}{3}\mu & 0 \\ 0 & \mu & 0 & 0 \\ 0 & \mu v & -\frac{2}{3}\mu u & 0 \end{pmatrix}, & \mathbf{K}_{21}(\mathbf{W}) &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \mu & 0 \\ 0 & -\frac{2}{3}\mu & 0 & 0 \\ 0 & -\frac{2}{3}\mu v & \mu u & 0 \end{pmatrix}. \end{aligned}$$

Consider a uniform Cartesian mesh with $N_1 \times N_2$ cells $I_{ij} = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$ with cell centers $((x_{i-\frac{1}{2}} + x_{i+\frac{1}{2}})/2, (y_{j-\frac{1}{2}} + y_{j+\frac{1}{2}})/2)$ and cell sizes $\Delta x, \Delta y$. The DoFs consist of the cell averages, edge-centered, and corner point values of the numerical solution $\mathbf{U}_h(\mathbf{x}, t)$, defined as

$$\bar{\mathbf{U}}_i(t) = \frac{1}{\Delta x \Delta y} \int_{I_i} \mathbf{U}_h(\mathbf{x}, t) \, d\mathbf{x}, \quad \mathbf{U}_\zeta(t) = \mathbf{U}_h(\mathbf{x}_\zeta, t).$$

Here $\mathbf{i} = (i, j)$, $\mathbf{x} = (x, y) = (x_1, x_2)$, and $\zeta \in \{(i + \frac{1}{2}, j), (i, j + \frac{1}{2}), (i + \frac{1}{2}, j + \frac{1}{2})\}$ denotes the locations of the point values. The proposed AF method is as follows,

$$\begin{aligned} \frac{d\bar{\mathbf{U}}_i}{dt} + \sum_{\ell=1}^2 \frac{1}{\Delta x_\ell} \left(\widehat{\mathbf{F}}_{\ell, i+\frac{1}{2}\mathbf{e}_\ell}^c - \widehat{\mathbf{F}}_{\ell, i-\frac{1}{2}\mathbf{e}_\ell}^c \right) &= \sum_{\ell=1}^2 \frac{1}{\Delta x_\ell} \left(\widehat{\mathbf{F}}_{\ell, i+\frac{1}{2}\mathbf{e}_\ell}^v - \widehat{\mathbf{F}}_{\ell, i-\frac{1}{2}\mathbf{e}_\ell}^v \right), \\ \widehat{\mathbf{F}}_{\ell, i+\frac{1}{2}\mathbf{e}_\ell}^c &= \overline{\mathbf{F}_\ell^c(\mathbf{U})}_{i+\frac{1}{2}\mathbf{e}_\ell}, \quad \widehat{\mathbf{F}}_{\ell, i+\frac{1}{2}\mathbf{e}_\ell}^v = \overline{\mathbf{F}_\ell^v(\mathbf{U}, \nabla \mathbf{Z})}_{i+\frac{1}{2}\mathbf{e}_\ell}, \\ \left(\frac{\partial \mathbf{Z}}{\partial x_\ell} \right)_{i+\frac{1}{2}\mathbf{e}_\ell} &= \mathcal{D}_c^{(\ell)}(\mathbf{Z})_{i+\frac{1}{2}\mathbf{e}_\ell}, \\ \frac{d\mathbf{U}_\zeta}{dt} + \sum_{\ell=1}^2 \left(\mathcal{D}_+^{(\ell)}(\mathbf{F}_\ell^{c,+})_\zeta + \mathcal{D}_-^{(\ell)}(\mathbf{F}_\ell^{c,-})_\zeta \right) &= \sum_{\ell=1}^2 \mathcal{D}^{(\ell),2}(\mathbf{K}_{\ell\ell}(\mathbf{W}), \mathbf{Z})_\zeta \\ + \sum_{\ell, m=1, \ell \neq m}^2 \mathcal{D}_c^{(\ell)}(\mathbf{K}_{\ell m}(\mathbf{W}) \mathcal{D}_c^{(m)} \mathbf{Z})_\zeta, \end{aligned} \tag{12}$$

where the overbar on each flux denotes its average along the corresponding edge, approximated by Simpson's rule, and the mixed derivatives in the point value update are discretized by using the operators in each direction sequentially. Here, the finite difference operators $\mathcal{D}_\pm^{(\ell)}$, $\mathcal{D}_c^{(\ell)}$, and $\mathcal{D}^{(\ell),2}$ are applied along the x_ℓ -direction, based on (7a)-(7b), (8), and (9), respectively. For example,

$$\begin{aligned} (\mathcal{D}_+^{(1)} z)_{i+\frac{1}{2},j} &= \frac{1}{3\Delta x} \left(z_{i-\frac{1}{2},j} - 6z_{i,j} + 3z_{i+\frac{1}{2},j} + 2z_{i+1,j} \right), \\ (\mathcal{D}_-^{(1)} z)_{i+\frac{1}{2},j} &= \frac{1}{3\Delta x} \left(-2z_{i,j} - 3z_{i+\frac{1}{2},j} + 6z_{i+1,j} - z_{i+\frac{3}{2},j} \right), \end{aligned}$$

$$\begin{aligned}
(\mathcal{D}_+^{(2)}z)_{i,j+\frac{1}{2}} &= \frac{1}{3\Delta y} \left(z_{i,j-\frac{1}{2}} - 6z_{i,j} + 3z_{i,j+\frac{1}{2}} + 2z_{i,j+1} \right), \\
(\mathcal{D}_-^{(2)}z)_{i,j+\frac{1}{2}} &= \frac{1}{3\Delta y} \left(-2z_{i,j} - 3z_{i,j+\frac{1}{2}} + 6z_{i,j+1} - z_{i,j+\frac{3}{2}} \right).
\end{aligned}$$

For each conservative component z , the cell-centered point value is obtained through

$$\begin{aligned}
z_{i,j} &= \frac{1}{16} \left[36\bar{z}_{i,j} - 4 \left(z_{i-\frac{1}{2},j} + z_{i+\frac{1}{2},j} + z_{i,j-\frac{1}{2}} + z_{i,j+\frac{1}{2}} \right) \right. \\
&\quad \left. - \left(z_{i-\frac{1}{2},j-\frac{1}{2}} + z_{i+\frac{1}{2},j-\frac{1}{2}} + z_{i-\frac{1}{2},j+\frac{1}{2}} + z_{i+\frac{1}{2},j+\frac{1}{2}} \right) \right].
\end{aligned}$$

The central finite difference operator is

$$(\mathcal{D}_c^{(\ell)}z)_{i,j} = \frac{1}{2} \left((\mathcal{D}_+^{(\ell)}z)_{i,j} + (\mathcal{D}_-^{(\ell)}z)_{i,j} \right).$$

The compact discretization for $\partial_x(\mathbf{K}_{11}(\mathbf{W})\partial_x\mathbf{Z})$, can be expressed as

$$\mathcal{D}^{(1),2}(\mathbf{K}, \mathbf{Z})_{i+\frac{1}{2},j} = \frac{4}{\Delta x^2} \sum_{m=-2}^2 \sum_{n=-2}^2 b_{mn} \mathbf{K} \left(\mathbf{W}_{i+\frac{m+1}{2},j} \right) \mathbf{Z}_{i+\frac{n+1}{2},j}.$$

The remaining formulas follow by half-index shifts in i and/or j .

2.3 Time discretization

For the semi-discretizations (6) and (12), this paper uses the explicit five-stage, fourth-order strong-stability-preserving (SSP) RK method SSP-RK54 for time integration. The CFL condition is

$$\Delta t \leq \frac{1}{\max_{i,\ell} \frac{\varrho_\ell(\bar{\mathbf{U}}_i)}{C_{\text{CFL}}^c \Delta x_\ell} + \max_{i,\ell} \frac{\tilde{\varrho}(\bar{\mathbf{U}}_i)}{C_{\text{CFL}}^v \Delta x_\ell^2}}, \quad (13)$$

where $\varrho_\ell = |u_\ell| + c$ and $\tilde{\varrho} = \max\{\frac{4\mu}{3\rho}, \frac{\gamma\mu}{\text{Pr}\rho}\}$ are the maximum absolute values of the eigenvalues of the inviscid and viscous parts, respectively. The CFL numbers $C_{\text{CFL}}^c = 1.04/d$ and $C_{\text{CFL}}^v = 0.33/d$ are chosen based on the von Neumann analysis in [27] for the inviscid and viscous parts, respectively. Note that, for the 1D linear advection equation, the stable CFL limit is higher than that of the standard 3rd-order AF method [19], which is another advantage of using the additional downwind point in the biased stencil.

3 Positivity-preserving limiting and oscillation control

The high-order AF method may produce negative density or pressure, leading to unphysical solutions or even causing the simulations to blow up. This section first presents a monolithic convex limiting approach that constrains individual numerical fluxes to preserve density and pressure positivity in the cell average update, adapting the approach proposed in [38] to the NS equations.

Definition 3.1. An AF method is called positivity-preserving (PP) if starting from cell averages and point values in the admissible state set

$$\mathcal{G} = \left\{ \mathbf{U} = (\rho, \rho \mathbf{u}, E) \mid \rho > 0, \ p = (\gamma - 1) \left(E - \frac{\|\rho \mathbf{u}\|^2}{2\rho} \right) > 0 \right\}, \quad (14)$$

the cell averages and point values remain in $\mathcal{G}$ at the next time step.

The PP properties of the limiting are shown using representations in terms of intermediate states that stay in the convex admissible state set (14). Positivity is proved separately for the inviscid and viscous split states, which are then combined in the final limiting procedure.

3.1 Preliminaries

The following statement for the inviscid part is the usual one for the Euler equations.

Lemma 3.1 (LF splitting for the inviscid part). Let $\mathbf{U} \in \mathcal{G}$ and the sound speed $c = \sqrt{\gamma p / \rho}$, the following statement holds,

$$\mathbf{U} \pm \frac{\mathbf{F}_\ell^c(\mathbf{U})}{\alpha_\ell} \in \mathcal{G}, \quad \text{if } \alpha_\ell \geq \varrho_\ell = |u_\ell| + c.$$

The following useful PP property of the directional LF splitting can be found in [36].

Lemma 3.2 (LF splitting for the viscous part). Let $\mathbf{U} \in \mathcal{G}$, $\boldsymbol{\tau}_\ell = \boldsymbol{\tau} \mathbf{e}_\ell$, $q_\ell = \mathbf{q} \cdot \mathbf{e}_\ell$, and introduce an auxiliary variable $\mathbf{S}$ as an approximation to $\nabla \mathbf{Z}$. Define

$$b_\ell(\mathbf{U}, \mathbf{S}) = \frac{\sqrt{\rho^2 q_\ell^2 + 2\rho^2 e \|\boldsymbol{\tau}_\ell\|^2 + \rho |q_\ell|}}{2\rho^2 e},$$

then

$$\mathbf{U} \pm \frac{\mathbf{F}_\ell^v(\mathbf{U}, \mathbf{S})}{\beta_\ell} \in \mathcal{G}, \quad \text{if } \beta_\ell > b_\ell.$$

Proof. The mass component of $\mathbf{F}_\ell^v$ is zero, so density positivity follows from $\mathbf{U} \in \mathcal{G}$. Using $\mathbf{F}_\ell^v = (0, \boldsymbol{\tau}_\ell, \mathbf{u} \cdot \boldsymbol{\tau}_\ell - q_\ell)^\top$ gives

$$\rho e \left(\mathbf{U} \pm \frac{\mathbf{F}_\ell^v}{\beta_\ell} \right) = \rho e(\mathbf{U}) \mp \frac{q_\ell}{\beta_\ell} - \frac{\|\boldsymbol{\tau}_\ell\|^2}{2\rho\beta_\ell^2}.$$

It is positive if $\rho e > \frac{|q_\ell|}{\beta_\ell} + \frac{\|\boldsymbol{\tau}_\ell\|^2}{2\rho\beta_\ell^2}$, which yields the condition $\beta_\ell > b_\ell$. □

Lemma 3.3 (LF splitting for the total flux). Let

$$\mathbf{F}_\ell(\mathbf{U}, \mathbf{S}) = \mathbf{F}_\ell^c(\mathbf{U}) - \mathbf{F}_\ell^v(\mathbf{U}, \mathbf{S}), \quad \lambda_\ell = \alpha_\ell + \beta_\ell, \quad (15)$$

where α_ℓ and β_ℓ satisfy the conditions in Lemmas 3.1 and 3.2, respectively. Then

$$\mathbf{U} \pm \frac{\mathbf{F}_\ell(\mathbf{U}, \mathbf{S})}{\lambda_\ell} \in \mathcal{G}. \quad (16)$$

Proof. The identity

$$\mathbf{U} \pm \frac{\mathbf{F}_\ell}{\lambda_\ell} = \frac{\alpha_\ell}{\lambda_\ell} \left(\mathbf{U} \pm \frac{\mathbf{F}_\ell^c}{\alpha_\ell} \right) + \frac{\beta_\ell}{\lambda_\ell} \left(\mathbf{U} \mp \frac{\mathbf{F}_\ell^v}{\beta_\ell} \right)$$

is a convex combination, so the proof follows from the convexity of $\mathcal{G}$. $\square$

For $d \in \{1, 2\}$, consider the following low-order LLF scheme for the cell average,

$$\begin{aligned} \bar{\mathbf{U}}_i^L &= \bar{\mathbf{U}}_i^n - \sum_{\ell=1}^d \mu_\ell \left(\hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^L - \hat{\mathbf{F}}_{i-\frac{1}{2}\mathbf{e}_\ell}^L \right), \\ \hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^L &= \hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{c,L} - \hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{v,L}, \quad \mu_\ell = \Delta t^n / \Delta x_\ell, \end{aligned} \quad (17)$$

where

$$\hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{c,L} = \frac{1}{2} \left(\mathbf{F}_\ell^c(\bar{\mathbf{U}}_i) + \mathbf{F}_\ell^c(\bar{\mathbf{U}}_{i+\mathbf{e}_\ell}) \right) - \frac{\alpha_{i+\frac{1}{2}\mathbf{e}_\ell}}{2} (\bar{\mathbf{U}}_{i+\mathbf{e}_\ell} - \bar{\mathbf{U}}_i), \quad (18a)$$

$$\hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{v,L} = \frac{1}{2} \left(\mathbf{F}_\ell^v(\bar{\mathbf{U}}_i, \mathbf{S}_i) + \mathbf{F}_\ell^v(\bar{\mathbf{U}}_{i+\mathbf{e}_\ell}, \mathbf{S}_{i+\mathbf{e}_\ell}) \right) + \frac{\beta_{i+\frac{1}{2}\mathbf{e}_\ell}}{2} (\bar{\mathbf{U}}_{i+\mathbf{e}_\ell} - \bar{\mathbf{U}}_i). \quad (18b)$$

Here the low-order gradient $\mathbf{S}_i$ is computed using the central difference of $\mathbf{Z}(\bar{\mathbf{U}}_i)$. Note that the superscript n is omitted for brevity. Define the total flux

$$\mathbf{F}_{\ell,i} = \mathbf{F}_\ell^c(\bar{\mathbf{U}}_i) - \mathbf{F}_\ell^v(\bar{\mathbf{U}}_i, \mathbf{S}_i),$$

then the total numerical flux can also be expressed as

$$\hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^L = \frac{1}{2} (\mathbf{F}_{\ell,i} + \mathbf{F}_{\ell,i+\mathbf{e}_\ell}) - \frac{\lambda_{i+\frac{1}{2}\mathbf{e}_\ell}}{2} (\bar{\mathbf{U}}_{i+\mathbf{e}_\ell} - \bar{\mathbf{U}}_i), \quad (19)$$

where

$$\lambda_{i+\frac{1}{2}\mathbf{e}_\ell} = \alpha_{i+\frac{1}{2}\mathbf{e}_\ell} + \beta_{i+\frac{1}{2}\mathbf{e}_\ell}.$$

The numerical viscosity coefficients are chosen as

$$\begin{aligned} \alpha_{i+\frac{1}{2}\mathbf{e}_\ell} &= \max \{ \varrho_\ell(\bar{\mathbf{U}}_i), \varrho_\ell(\bar{\mathbf{U}}_{i+\mathbf{e}_\ell}) \}, \\ \beta_{i+\frac{1}{2}\mathbf{e}_\ell} &= \max \{ b_\ell(\bar{\mathbf{U}}_i, \mathbf{S}_i), b_\ell(\bar{\mathbf{U}}_{i+\mathbf{e}_\ell}, \mathbf{S}_{i+\mathbf{e}_\ell}) \}. \end{aligned}$$

With the above notations, the LLF scheme (17) can be written as a convex decomposition

$$\bar{\mathbf{U}}_i^L = \frac{1}{2d} \sum_{\ell=1}^d (\mathbf{U}_{\ell,i}^{L,+} + \mathbf{U}_{\ell,i}^{L,-}),$$

where

$$\begin{aligned} \mathbf{U}_{\ell,i}^{L,+} &= \bar{\mathbf{U}}_i - 2d\mu_\ell \left(\hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^L - \mathbf{F}_{\ell,i} \right), \\ \mathbf{U}_{\ell,i}^{L,-} &= \bar{\mathbf{U}}_i - 2d\mu_\ell \left(\mathbf{F}_{\ell,i} - \hat{\mathbf{F}}_{i-\frac{1}{2}\mathbf{e}_\ell}^L \right). \end{aligned} \quad (20)$$

To show the PP property of the LLF scheme, it suffices to show that the intermediate states stay in $\mathcal{G}$ due to the convexity of $\mathcal{G}$.

Proposition 3.1. Assume that $\bar{\mathbf{U}}_{\mathbf{i}} \in \mathcal{G}$ for all $\mathbf{i}$, and the numerical viscosity coefficients satisfy the conditions in Lemmas 3.1 and 3.2. Then the intermediate states in (20) satisfy $\mathbf{U}_{\ell,\mathbf{i}}^{\mathbf{L},\pm} \in \mathcal{G}$, under the following time-step size constraint

$$\Delta t \leq \min_{1 \leq \ell \leq d} \frac{\Delta x_\ell}{2d \max_{\mathbf{i}} \{\lambda_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}\}}. \quad (21)$$

Proof. For the two interfaces, let $\nu_{\ell,\pm} = \mu_\ell \lambda_{\mathbf{i} \pm \frac{1}{2}\mathbf{e}_\ell}$, respectively. The intermediate states can be expressed as convex decompositions

$$\begin{aligned} \mathbf{U}_{\ell,\mathbf{i}}^{\mathbf{L},+} &= (1 - 2d\nu_{\ell,+}) \bar{\mathbf{U}}_{\mathbf{i}} + d\nu_{\ell,+} \left(\bar{\mathbf{U}}_{\mathbf{i}} + \frac{\mathbf{F}_{\ell,\mathbf{i}}}{\lambda_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}} \right) + d\nu_{\ell,+} \left(\bar{\mathbf{U}}_{\mathbf{i}+\mathbf{e}_\ell} - \frac{\mathbf{F}_{\ell,\mathbf{i}+\mathbf{e}_\ell}}{\lambda_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}} \right), \\ \mathbf{U}_{\ell,\mathbf{i}}^{\mathbf{L},-} &= (1 - 2d\nu_{\ell,-}) \bar{\mathbf{U}}_{\mathbf{i}} + d\nu_{\ell,-} \left(\bar{\mathbf{U}}_{\mathbf{i}} - \frac{\mathbf{F}_{\ell,\mathbf{i}}}{\lambda_{\mathbf{i}-\frac{1}{2}\mathbf{e}_\ell}} \right) + d\nu_{\ell,-} \left(\bar{\mathbf{U}}_{\mathbf{i}-\mathbf{e}_\ell} + \frac{\mathbf{F}_{\ell,\mathbf{i}-\mathbf{e}_\ell}}{\lambda_{\mathbf{i}-\frac{1}{2}\mathbf{e}_\ell}} \right). \end{aligned}$$

Under (21), the coefficients $1 - 2d\nu_{\ell,\pm}$, $d\nu_{\ell,\pm}$, and $d\nu_{\ell,\pm}$ are nonnegative and sum to one. Using Lemma 3.3 and the convexity of $\mathcal{G}$, one has $\mathbf{U}_{\ell,\mathbf{i}}^{\mathbf{L},\pm} \in \mathcal{G}$. $\square$

It is known that the low-order scheme is PP and non-oscillatory, thus the basic idea of the limiting for cell average is to blend the high-order AF numerical flux with the low-order LLF flux, which guarantees local conservation. To be specific, at the cell interface $\mathbf{i} + \frac{1}{2}\mathbf{e}_\ell$, define the total flux in the high-order AF method as

$$\hat{\mathbf{F}}_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}^{\mathbf{H}} = \hat{\mathbf{F}}_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}^{\mathbf{c},\mathbf{H}} - \hat{\mathbf{F}}_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}^{\mathbf{v},\mathbf{H}},$$

and denote the anti-diffusion term by

$$\Delta \hat{\mathbf{F}}_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell} = \hat{\mathbf{F}}_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}^{\mathbf{H}} - \hat{\mathbf{F}}_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}^{\mathbf{L}}.$$

Then the modified limited flux can be generally expressed as

$$\hat{\mathbf{F}}_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}^{\text{modified}} = \hat{\mathbf{F}}_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}^{\mathbf{L}} + \theta_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}^{\text{modified}} \Delta \hat{\mathbf{F}}_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell},$$

where $\theta_{\mathbf{i}+\frac{1}{2}\mathbf{e}_\ell}^{\text{modified}} \in [0, 1]$ is determined by a smoothness indicator or by density and pressure positivity requirements. The blending parameters are determined sequentially below.

3.2 Smoothness indicator based limiting

This paper adopts the smoothness indicator in [18], which combines Jameson's shock sensor in [39] and a modified Ducros shock sensor [40]. For $d \in \{1, 2\}$, let

$$(\varphi_1)_{\mathbf{i}} = \max_{1 \leq m \leq d} \frac{|\bar{p}_{\mathbf{i}+\mathbf{e}_m} - 2\bar{p}_{\mathbf{i}} + \bar{p}_{\mathbf{i}-\mathbf{e}_m}|}{|\bar{p}_{\mathbf{i}+\mathbf{e}_m} + 2\bar{p}_{\mathbf{i}} + \bar{p}_{\mathbf{i}-\mathbf{e}_m}|},$$

and

$$(\varphi_2)_{\mathbf{i}} = \max \left\{ \frac{-(\nabla \cdot \bar{\mathbf{u}})_{\mathbf{i}}}{\sqrt{(\nabla \cdot \bar{\mathbf{u}})_{\mathbf{i}}^2 + \|(\nabla \times \bar{\mathbf{u}})_{\mathbf{i}}\|^2 + 10^{-13}}}, 0 \right\}.$$

where $(\nabla \cdot \bar{\mathbf{u}})_i$ and $(\nabla \times \bar{\mathbf{u}})_i$ are obtained by using central differences, with $\bar{\mathbf{u}}_i$ and $\bar{p}_i$ the velocity and pressure recovered from the cell average $\bar{\mathbf{U}}_i$. The blending coefficient is designed as

$$\begin{aligned}\theta_{i+\frac{1}{2}\mathbf{e}_\ell}^s &= \exp(-C_s(\varphi_1)_{i+\frac{1}{2}\mathbf{e}_\ell}(\varphi_2)_{i+\frac{1}{2}\mathbf{e}_\ell}) \in [0, 1], \\ (\varphi_s)_{i+\frac{1}{2}\mathbf{e}_\ell} &= \max\{(\varphi_s)_i, (\varphi_s)_{i+\mathbf{e}_\ell}\}, \quad s = 1, 2,\end{aligned}$$

where the parameter C_s adjusts the strength of the limiting, which is fixed as 1 throughout this paper. The limited numerical flux after this limiting is

$$\hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim},s} = \hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{L}} + \theta_{i+\frac{1}{2}\mathbf{e}_\ell}^s \Delta \hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}, \quad (22)$$

and the corresponding anti-diffusion term is updated as

$$\Delta \hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim},s} = \hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim},s} - \hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{L}}.$$

3.3 Positivity-preserving limiting

Similar to the low-order scheme (17), the full cell average update in the AF method can be decomposed into $2d$ intermediate states as follows,

$$\bar{\mathbf{U}}_i^{\text{H}} = \frac{1}{2d} \sum_{\ell=1}^d (\mathbf{U}_{\ell,i}^{\text{H},+} + \mathbf{U}_{\ell,i}^{\text{H},-}),$$

where

$$\begin{aligned}\mathbf{U}_{\ell,i}^{\text{H},+} &= \bar{\mathbf{U}}_i - 2d\mu_\ell \left(\hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim},s} - \mathbf{F}_{\ell,i} \right), \\ \mathbf{U}_{\ell,i}^{\text{H},-} &= \bar{\mathbf{U}}_i - 2d\mu_\ell \left(\mathbf{F}_{\ell,i} - \hat{\mathbf{F}}_{i-\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim},s} \right),\end{aligned}$$

with $\hat{\mathbf{F}}_{i\pm\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim},s}$ given by (22). The superscript H denotes the high-order candidate after the limiting based on the smoothness indicator. Define the limited cell average update as

$$\bar{\mathbf{U}}_i^{\text{Lim,pp}} = \bar{\mathbf{U}}_i^n - \sum_{\ell=1}^d \mu_\ell \left(\hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim,pp}} - \hat{\mathbf{F}}_{i-\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim,pp}} \right), \quad (23)$$

where the limited numerical flux is given by

$$\hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim,pp}} = \hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{L}} + \theta_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{pp}} \Delta \hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim},s}.$$

Rewrite the limited cell average as

$$\bar{\mathbf{U}}_i^{\text{Lim,pp}} = \frac{1}{2d} \sum_{\ell=1}^d (\mathbf{U}_{\ell,i}^{\text{Lim,pp},+} + \mathbf{U}_{\ell,i}^{\text{Lim,pp},-}),$$

with the limited intermediate states

$$\begin{aligned}\mathbf{U}_{\ell,i}^{\text{Lim,pp},+} &= \mathbf{U}_{\ell,i}^{\text{L},+} - 2d\mu_\ell \theta_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{pp}} \Delta \hat{\mathbf{F}}_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim},s} = (1 - \theta_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{pp}}) \mathbf{U}_{\ell,i}^{\text{L},+} + \theta_{i+\frac{1}{2}\mathbf{e}_\ell}^{\text{pp}} \mathbf{U}_{\ell,i}^{\text{H},+}, \\ \mathbf{U}_{\ell,i}^{\text{Lim,pp},-} &= \mathbf{U}_{\ell,i}^{\text{L},-} + 2d\mu_\ell \theta_{i-\frac{1}{2}\mathbf{e}_\ell}^{\text{pp}} \Delta \hat{\mathbf{F}}_{i-\frac{1}{2}\mathbf{e}_\ell}^{\text{Lim},s} = (1 - \theta_{i-\frac{1}{2}\mathbf{e}_\ell}^{\text{pp}}) \mathbf{U}_{\ell,i}^{\text{L},-} + \theta_{i-\frac{1}{2}\mathbf{e}_\ell}^{\text{pp}} \mathbf{U}_{\ell,i}^{\text{H},-},\end{aligned}$$

where the blending parameter $\theta_{i\pm\frac{1}{2}\mathbf{e}_\ell}^{\text{pp}} \in [0, 1]$. With the above construction, one can verify the following proposition.

Proposition 3.2. If the limited intermediate states $\mathbf{U}_{\ell,i}^{\text{Lim,pp},\pm} \in \mathcal{G}$ for all i and $\ell = 1, \dots, d$, then the limited average update (23) is PP, i.e., $\bar{\mathbf{U}}_i^{\text{Lim,pp}} \in \mathcal{G}$, under the CFL condition (21). If the limiting is applied to every forward-Euler stage, the SSP-RK method preserves the PP property through its convex stage combinations.

The remaining task is to determine the parameter $\theta_{i \pm \frac{1}{2}e_\ell}^{\text{pp}} \in [0, 1]$ at each interface, such that $\mathbf{U}_{\ell,i}^{\text{Lim,pp},\pm} \in \mathcal{G}$ and $\theta_{i \pm \frac{1}{2}e_\ell}^{\text{pp}}$ is as large as permitted. Take the computation of $\theta_{i + \frac{1}{2}e_\ell}$ as an example. Define

$$\Theta(z^L, z^H, \varepsilon) = \begin{cases} \frac{z^L - \varepsilon}{z^L - z^H}, & z^H < \varepsilon, \\ 1, & \text{otherwise,} \end{cases}$$

for $z^L \geq \varepsilon$. By construction, it holds $(1 - \vartheta)z^L + \vartheta z^H \geq \varepsilon$ for all $\vartheta \in [0, \Theta(z^L, z^H, \varepsilon)]$.

The first step is to impose the density positivity

$$\rho(\tilde{\mathbf{U}}_{\ell,i}^{\text{Lim},\rho,+}) > \varepsilon_{\ell,i}^\rho, \quad \rho(\tilde{\mathbf{U}}_{\ell,i+e_\ell}^{\text{Lim},\rho,-}) > \varepsilon_{\ell,i+e_\ell}^\rho,$$

such that the following intermediate states have positive density component,

$$\tilde{\mathbf{U}}_{\ell,i}^{\text{Lim},\rho,\pm} = (1 - \theta_{i \pm \frac{1}{2}e_\ell}^\rho) \mathbf{U}_{\ell,i}^{L,\pm} + \theta_{i \pm \frac{1}{2}e_\ell}^\rho \mathbf{U}_{\ell,i}^{H,\pm},$$

where $\varepsilon_{\ell,i}^\rho$ is a small positive number defined by $\varepsilon_{\ell,i}^\rho := \min\{10^{-13}, \rho(\mathbf{U}_{\ell,i}^{L,\pm})\}$. The blending parameter for the density is chosen as

$$\theta_{i + \frac{1}{2}e_\ell}^\rho = \min \left\{ \Theta(\rho(\mathbf{U}_{\ell,i}^{L,+}), \rho(\mathbf{U}_{\ell,i}^{H,+}), \varepsilon_{\ell,i}^\rho), \Theta(\rho(\mathbf{U}_{\ell,i+e_\ell}^{L,-}), \rho(\mathbf{U}_{\ell,i+e_\ell}^{H,-}), \varepsilon_{\ell,i+e_\ell}^\rho) \right\},$$

which considers the two cells sharing the same interface.

The second step is to impose the pressure positivity

$$p(\mathbf{U}_{\ell,i}^{\text{Lim,pp},+}) > \varepsilon_{\ell,i}^p, \quad p(\mathbf{U}_{\ell,i+e_\ell}^{\text{Lim,pp},-}) > \varepsilon_{\ell,i+e_\ell}^p,$$

where $\varepsilon_{\ell,i}^p$ is a small positive number defined by $\varepsilon_{\ell,i}^p := \min\{10^{-13}, p(\mathbf{U}_{\ell,i}^{L,\pm})\}$. The blending parameter for the pressure is chosen as

$$\theta_{i + \frac{1}{2}e_\ell}^p = \min \left\{ \Theta(p(\mathbf{U}_{\ell,i}^{L,+}), p(\tilde{\mathbf{U}}_{\ell,i}^{\text{Lim},\rho,+}), \varepsilon_{\ell,i}^p), \Theta(p(\mathbf{U}_{\ell,i+e_\ell}^{L,-}), p(\tilde{\mathbf{U}}_{\ell,i+e_\ell}^{\text{Lim},\rho,-}), \varepsilon_{\ell,i+e_\ell}^p) \right\},$$

which uses the concavity of the pressure function. Thus the final limited intermediate states and numerical flux are

$$\begin{aligned} \mathbf{U}_{\ell,i}^{\text{Lim,pp},\pm} &= (1 - \theta_{i \pm \frac{1}{2}e_\ell}^\rho \theta_{i \pm \frac{1}{2}e_\ell}^p) \mathbf{U}_{\ell,i}^{L,\pm} + \theta_{i \pm \frac{1}{2}e_\ell}^\rho \theta_{i \pm \frac{1}{2}e_\ell}^p \mathbf{U}_{\ell,i}^{H,\pm}, \\ \hat{\mathbf{F}}_{i + \frac{1}{2}e_\ell}^{\text{Lim,pp}} &= \hat{\mathbf{F}}_{i + \frac{1}{2}e_\ell}^L + \theta_{i + \frac{1}{2}e_\ell}^\rho \theta_{i + \frac{1}{2}e_\ell}^p \Delta \hat{\mathbf{F}}_{i + \frac{1}{2}e_\ell}^{\text{Lim,s}}. \end{aligned}$$

Remark 3.1. Note that the cell-centered value should also be limited using the scaling limiter with the help of the cell average, see [18] for the details.

3.4 Limiting for the point value

To achieve the PP property, a simple scaling limiter is adopted. One can easily write down a low-order PP LLF method for the point value based on the spatial stencil used in [18] and the numerical flux defined in (18). Then the high-order AF solution is blended with the low-order one to preserve positivity, with the blending parameters determined similarly to those for cell averages.

Let us summarize the main results of the proposed PP AF method.

Theorem 3.1. If the initial numerical solution $\bar{U}_i^0, U_\zeta^0 \in \mathcal{G}$ for all i and ζ , and the time-step size satisfies (21) for the cell average and also the one for the point value, then the AF method (6a)–(6b) equipped with an SSP Runge–Kutta method and the PP limiters for cell averages (23) and point values preserves density and pressure positivity.

Remark 3.2. In the actual implementation, the time-step size is first chosen based on the CFL condition (13). If the low-order counterparts needed in the proposed limiting are not PP, then the solution is reset and the full RK step is recomputed with a halved time-step size.

4 Numerical experiments

This section presents numerical experiments for the compressible NS equations (1), including smooth accuracy tests, problems with strong shocks and low pressure, and viscous flows at low Mach numbers. The ratio of specific heats is $\gamma = 1.4$ unless specified otherwise.

Example 4.1 (1D viscous shock wave). Consider the moving viscous shock used in [34] on the domain $[-1, 1.5]$. Let u_∞ denote the translation velocity and $\xi = x - u_\infty t$. Let $w(\xi)$ denote the velocity relative to the moving shock. The exact primitive variables are $\rho(\xi)$, $u(x, t) = u_\infty + w(\xi)$, and $p(\xi)$, where

$$\rho(\xi) = \frac{m_0}{w(\xi)}, \quad p(\xi) = (\gamma - 1)\rho(\xi)e(\xi), \quad e(\xi) = \frac{1}{2\gamma} \left(\frac{\gamma + 1}{\gamma - 1} w_0^2 - w(\xi)^2 \right).$$

Let w_L and w_R be the velocities relative to the shock at $-\infty$ and $+\infty$, respectively. The velocity $w = w(\xi)$ is determined by

$$\xi = \frac{2\mu\gamma}{(\gamma + 1)m_0 \text{Pr}} \left[\frac{w_L}{w_L - w_R} \log \left(\frac{w_L - w}{w_L - w_0} \right) - \frac{w_R}{w_L - w_R} \log \left(\frac{w - w_R}{w_0 - w_R} \right) \right],$$

which was derived by Becker [41]. Here, $m_0 = \rho_L w_L$, $w_0 = \sqrt{w_L w_R}$. The ratio of the upstream and downstream relative velocities is determined by the upstream Mach number in the shock frame,

$$\frac{w_R}{w_L} = \frac{(\gamma - 1) + 2/M_0^2}{\gamma + 1}, \quad M_0 = \frac{w_L}{\sqrt{\gamma p_L / \rho_L}}.$$

Take $\text{Pr} = 3/4$, $u_\infty = 0.2$, $\rho_L = w_L = m_0 = 1$, and the initial shock is located at $x = 0$. For the first case, take $M_0 = 3$ and $\mu = 10^{-2}$ for the accuracy test without limiting, whereas the second case uses $M_0 = 20$ and $\mu = 10^{-3}$ for the accuracy test with limiting. Both cases are evolved to $t_{\text{final}} = 0.2$. Boundary values at both ends of the domain are prescribed using the time-dependent exact solution.

Figures 1 and 2 show the numerical solutions and the ℓ_1 errors and convergence rates for the first and second cases, respectively. The AF method accurately captures the exact solution and achieves fourth-order convergence under mesh refinement.

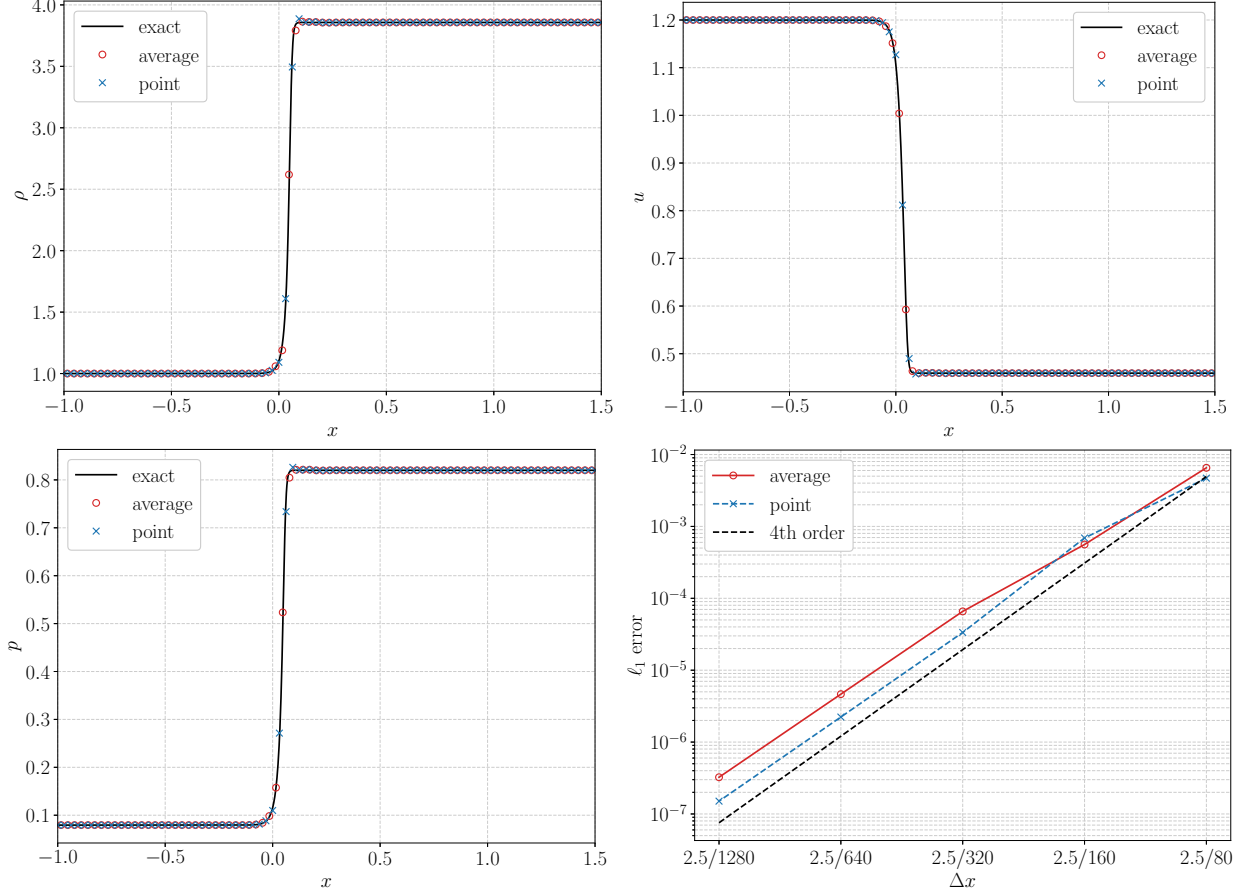


Figure 1: Example 4.1: viscous shock with $M_0 = 3$. The numerical solutions with 80 cells and the ℓ_1 errors and convergence rates of the AF method.

Example 4.2 (2D manufactured Taylor–Green vortex). Consider the periodic domain $[0, 2\pi] \times [0, 2\pi]$. Let $a(t) = \exp(-2\mu t)$ and $b(x, y) = \cos(2x) + \cos(2y)$. The prescribed solution is

$$\rho = 1, \quad \mathbf{u} = (a \sin x \cos y - a \cos x \sin y)^\top, \quad p = \frac{p_0}{\gamma - 1} + \frac{a^2 b}{4},$$

with $p_0 = 10^2$. The mass and momentum equations are satisfied automatically. For the energy equation, one has to add the following manufactured energy source

$$S_E = \frac{p_t + up_x + vp_y}{\gamma - 1} - 4\mu a^2 \cos^2 x \cos^2 y - \kappa \Delta p,$$

where $p_t = -\mu a^2 b$, $p_x = -a^2 \sin(2x)/2$, $p_y = -a^2 \sin(2y)/2$, $\Delta p = -a^2 b$. The viscosity is $\mu = 10^{-5}$, with $\text{Pr} = 1$ and $t_{\text{final}} = 0.2$.

The numerical solutions obtained by the proposed AF method with 80×80 cells are shown in Figure 3. It is observed that the results agree well with the exact solution. The errors and convergence rates are presented in Figure 4, which clearly show fourth-order convergence.

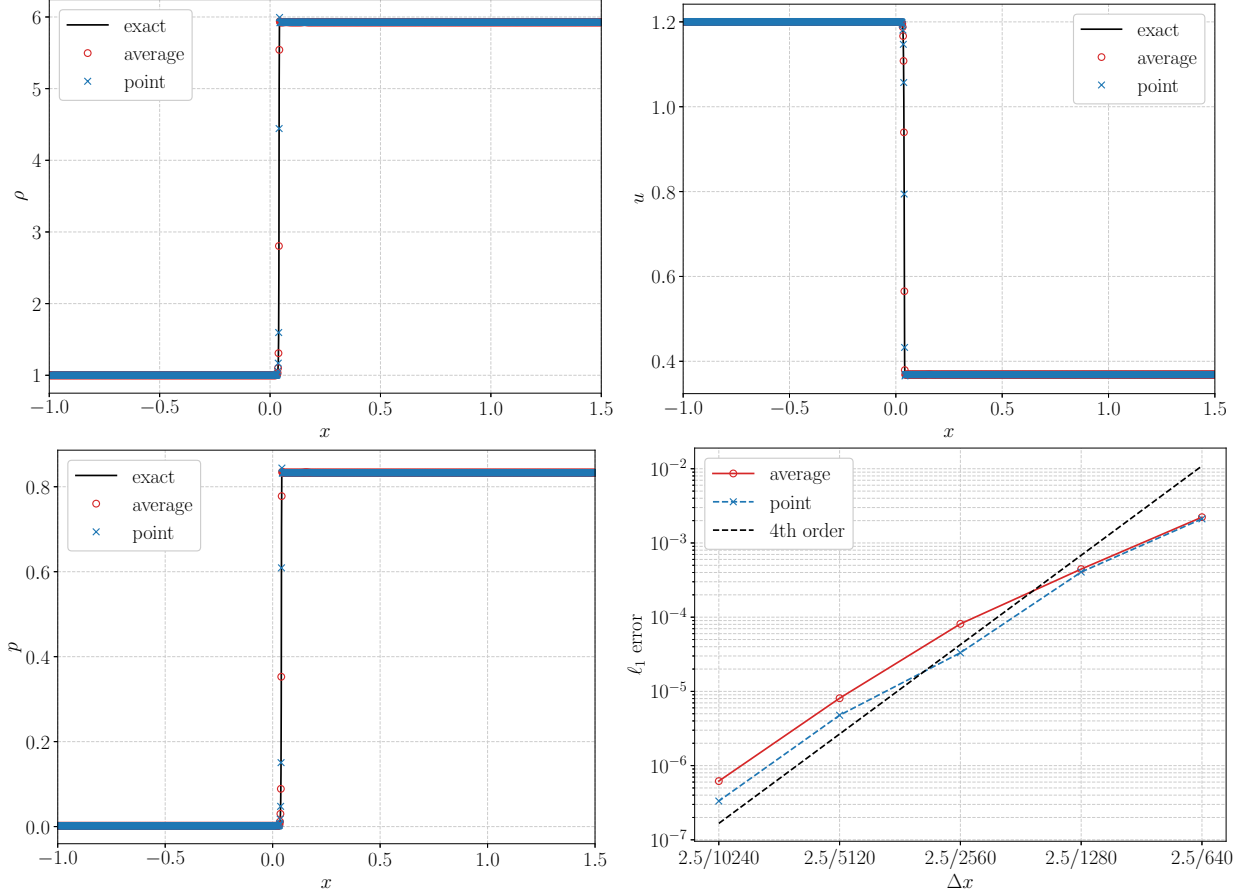


Figure 2: Example 4.1: viscous shock with $M_0 = 20$. The numerical solutions with 1280 cells and the ℓ_1 errors and convergence rates of the AF method.

Example 4.3 (First problem of Stokes). This test considers the diffusion of a tangential velocity discontinuity, where the solution depends only on x , while the transverse velocity is retained. The computational domain is $[-0.5, 0.5]$ with outflow boundary conditions, and the initial condition is

$$(\rho, u, v, p) = \begin{cases} (1, 0, -0.1, 1/\gamma), & \text{if } x < 0, \\ (1, 0, 0.1, 1/\gamma), & \text{otherwise.} \end{cases}$$

We consider $\mu = 10^{-2}$, 10^{-3} , and 10^{-4} , with $\kappa = 0$ in all three cases. The test is solved until $t_{\text{final}} = 1$. In the incompressible limit, this test has an analytical solution given by $v = \frac{1}{10} \text{erf}\left(\frac{x}{2\sqrt{\mu t}}\right)$.

Figure 5 presents the numerical solutions for the three viscosity values $\mu = 10^{-2}$, 10^{-3} , and 10^{-4} using 100 cells. The results agree well with the reference solutions without oscillations.

Example 4.4 (One-dimensional LeBlanc problem). We consider the one-dimensional LeBlanc problem using the initial data associated with the test in [42]. On the domain $[-0.5, 0.5]$, the initial condition is

$$(\rho, u, p)(x, 0) = \begin{cases} (1, 0, 6.666667 \times 10^{-2}), & x < -0.2, \\ (10^{-2}, 0, 6.666667 \times 10^{-11}), & x \geq -0.2. \end{cases}$$

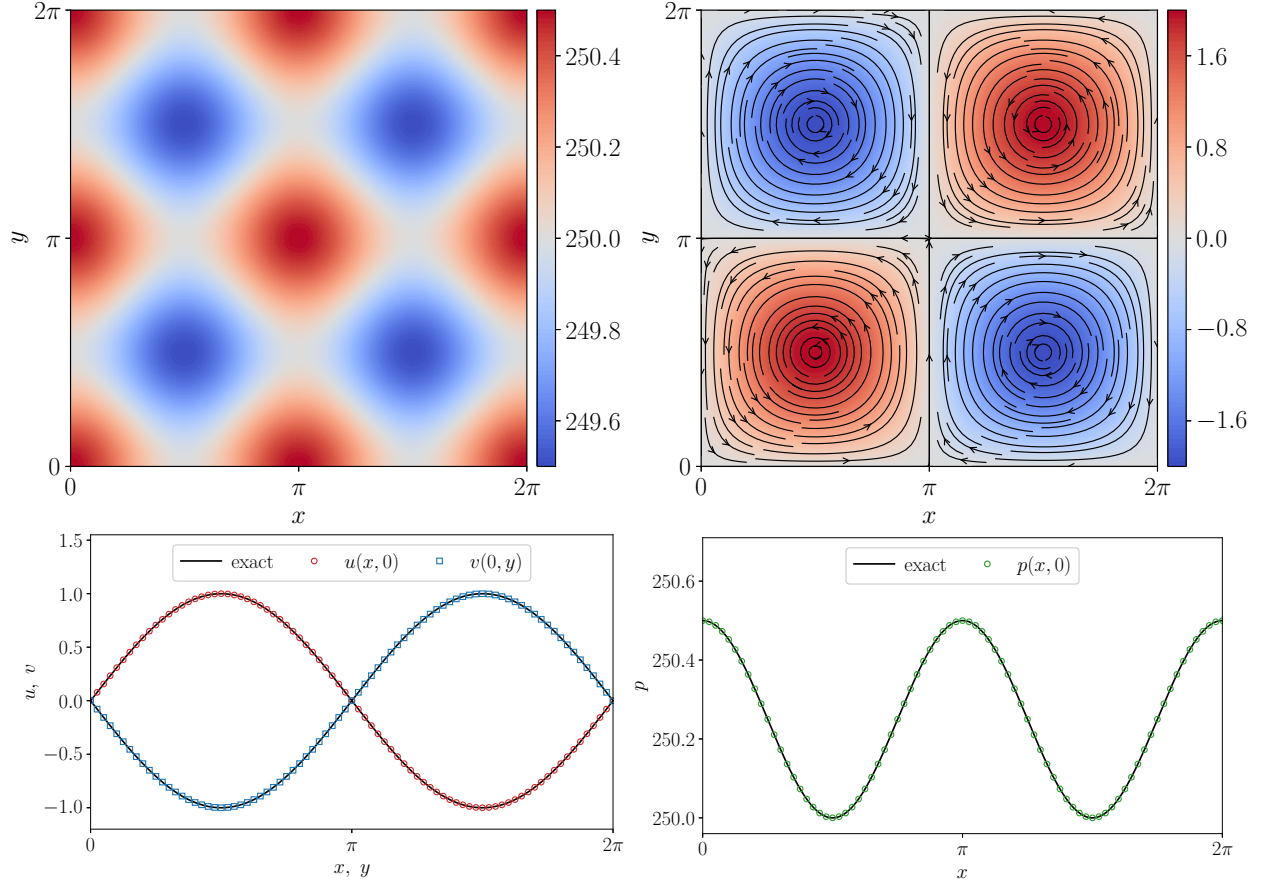


Figure 3: Example 4.2: manufactured Taylor–Green vortex. The pressure, streamlines colored by the vorticity, velocity cut-lines, and pressure cut-line (from upper left to lower right).

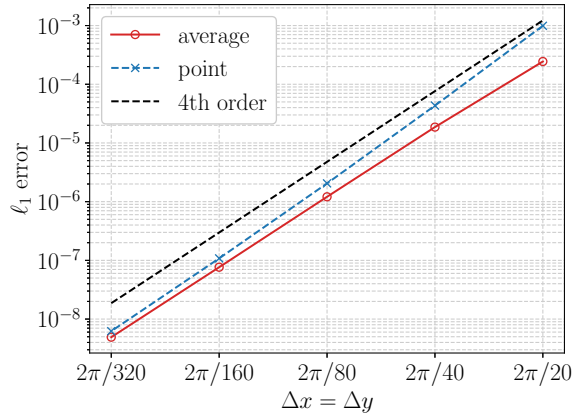


Figure 4: Example 4.2: manufactured Taylor–Green vortex. The errors and convergence rates.

We consider the inviscid case $\mu = 0$ and the viscous case $\mu = 10^{-5}$ with $\text{Pr} = 0.75$. Both cases use outflow boundary conditions and are evolved to $t_{\text{final}} = 0.4$.

The numerical solutions obtained with 800 cells are shown in Figure 6. The inviscid results are compared with the exact Euler solution, and the viscous results are compared with an AF reference solution computed using 2000 cells. Viscosity smooths the discontinuities. A small density undershoot appears near the tail of the rarefaction in the inviscid case;

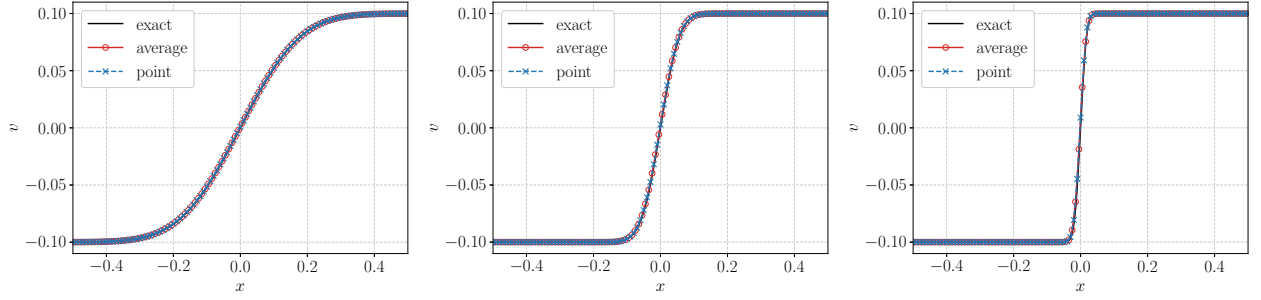


Figure 5: Example 4.3: the first problem of Stokes. The tangential velocity obtained by the AF method with $\mu = 10^{-2}, 10^{-3}, 10^{-4}$ (from left to right).

otherwise, the numerical solutions agree well with the reference solutions.

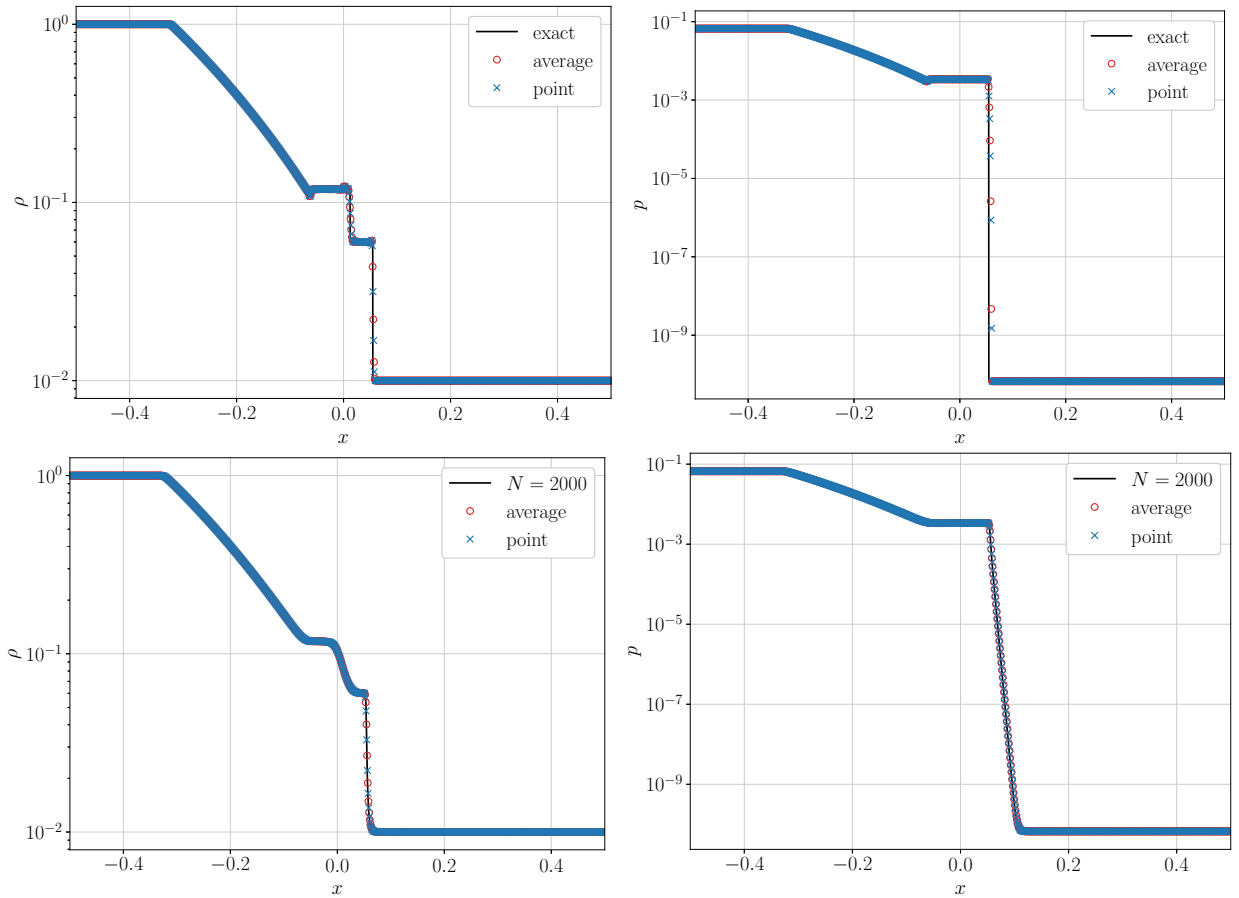


Figure 6: Example 4.4: LeBlanc test. The numerical solutions obtained with 800 cells for $\mu = 0$ (upper) and $\mu = 10^{-5}$ (lower).

Example 4.5 (Shock diffraction). Consider the diffraction of a Mach 5.09 shock around a corner, following [36]. The computational domain is $[0, 13] \times [0, 11]$ with the solid region $[0, 1] \times [0, 6]$ excluded. Slip adiabatic wall conditions are imposed on the exposed solid boundaries. The post-shock state is prescribed at the left inlet for $y \geq 6$, and outflow conditions are imposed on the remaining external boundaries. Initially, the right-moving

shock is located at $x = 0.5$. The initial condition is

$$(\rho, u, v, p) = \begin{cases} (1.4, 0, 0, 1), & \text{if } x > 0.5, \\ (7.041132907, 4.077946955, 0, 30.05945), & \text{otherwise.} \end{cases}$$

We solve the inviscid Euler equations with $\mu = 0$ and the NS equations with $\text{Re} = 200$ and $\text{Pr} = 0.72$. Both cases are evolved to $t_{\text{final}} = 2.3$.

The numerical solutions obtained by the proposed AF method are shown in Figure 7. Our results agree well with those in [36], and the viscosity makes the solution much smoother near the corner.

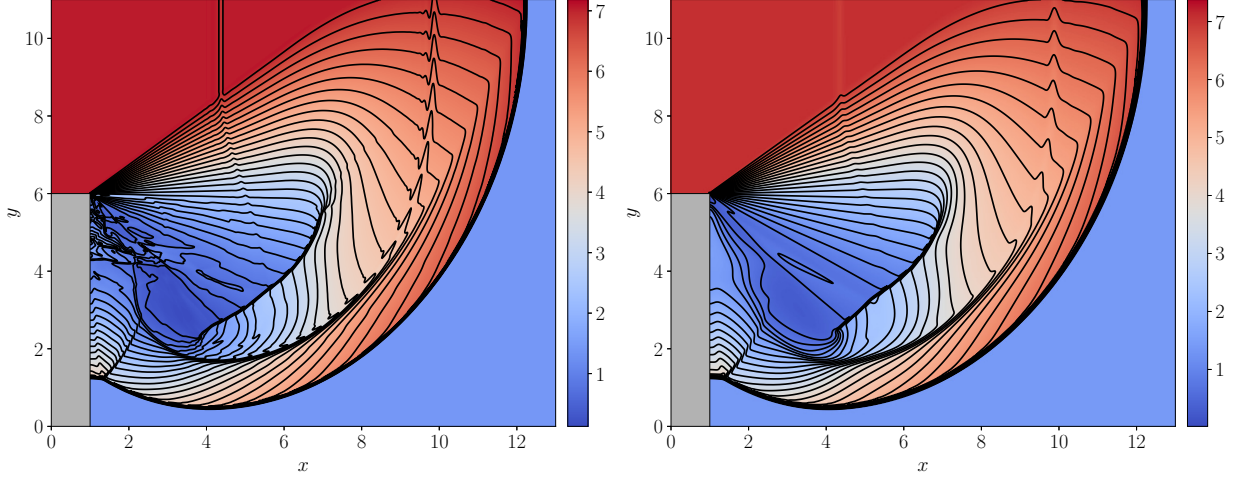


Figure 7: Example 4.5: shock diffraction. The density obtained on a 520×440 mesh for $\mu = 0$ (left) and $\text{Re} = 200$ (right).

Example 4.6 (Two-dimensional viscous shock tube with shock–boundary-layer interaction). To examine the interaction of shocks with boundary layers, consider the shock tube problem in [43]. The computational domain is the lower half of the square tube $[0, 1] \times [0, 0.5]$, with a symmetry boundary condition for the top boundary, while no-slip conditions are imposed elsewhere. The initial condition is

$$(\rho, u, v, p) = \begin{cases} (120, 0, 0, 120/\gamma), & \text{if } x < 0.5, \\ (1.2, 0, 0, 1.2/\gamma), & \text{otherwise,} \end{cases}$$

and $\mu = 10^{-3}$ and $\text{Pr} = 0.73$. The output time is $t_{\text{final}} = 1$. The reflected shock wave interacts with the boundary layers, leading to complex flow patterns consisting of a lambda shock, a main vortex, and other features.

Let $g = |\nabla \rho|$ be the magnitude of the density gradient. Figure 8 shows the indicator $\phi = \exp(-10(g - g_{\min})/(g_{\max} - g_{\min}))$, obtained by the AF method and the Q^3 entropy stable DG method implemented in the Julia code of [34]. For the DG results shown here, additional shock capturing is disabled, while positivity limiting is retained. Smaller ϕ (red) indicates a larger normalized density gradient, while larger ϕ (blue) indicates a smaller one. It is observed that the proposed AF method can capture the shock waves, vortices, and other small features with good resolution. When using the same total number of DoFs, the AF method generally captures the flow structures better than the DG method. For

example, comparing the results obtained by the AF method with 800×400 cells and by the DG method with 400×200 cells, the former captures the shock waves emerging from the lambda shock better, and also the structures between the wedge-shaped region located in $[0.5, 0.7] \times [0, 0.05]$. The results obtained by the AF method also agree better with the reference solution provided in [44], which is obtained by a second-order gas kinetic scheme on a 5000×2500 mesh.

The blending parameter of the proposed limiting for the AF is shown in Figure 8, i.e., $1 - \theta^{PP}\theta^S$. The plot indicates that the limiting is mainly activated near strong discontinuities, thus the AF method can give high-resolution solutions.

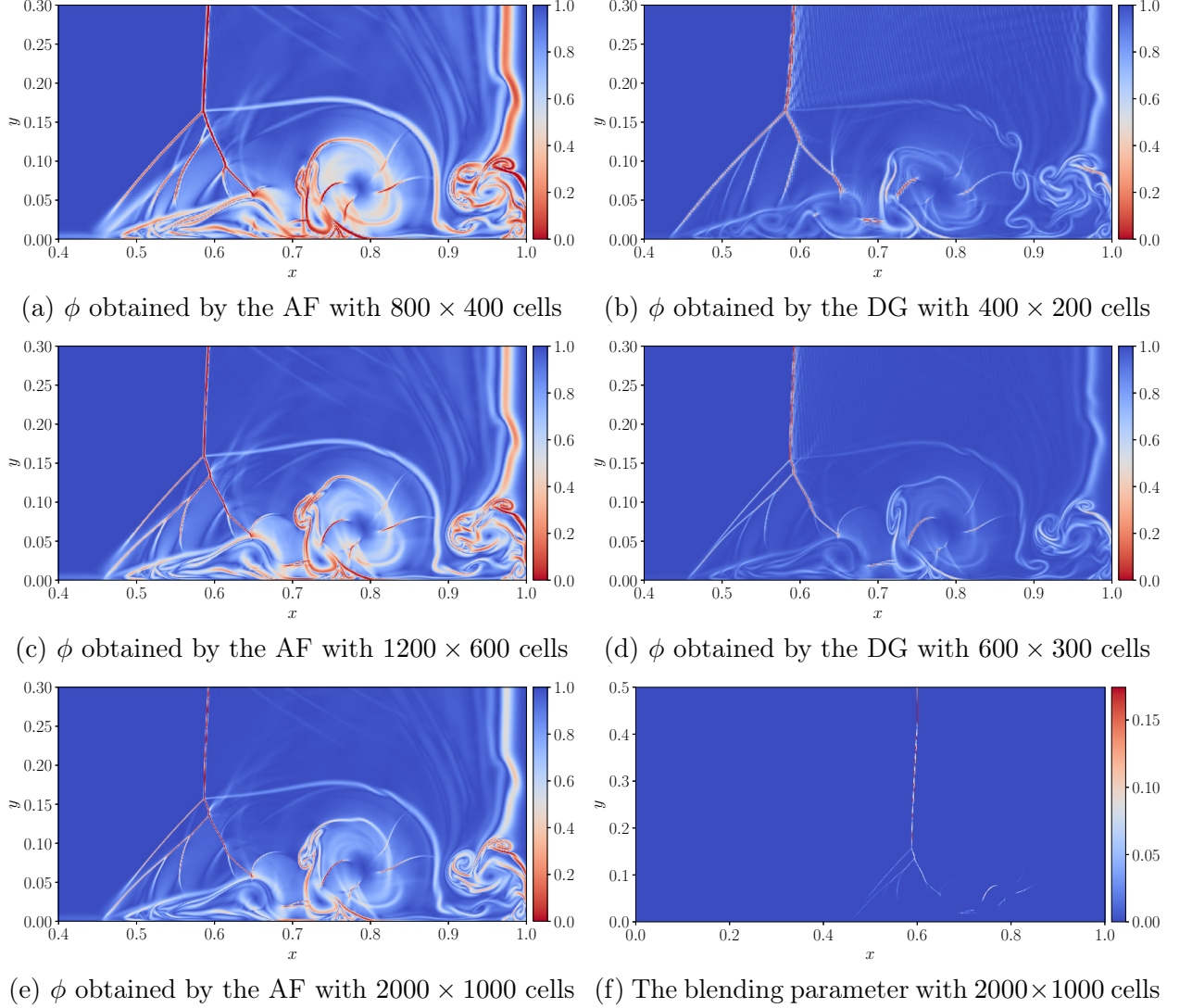


Figure 8: Example 4.6: two-dimensional viscous shock tube with shock–boundary-layer interaction with $\text{Re} = 1000$. All panels correspond to $t_{\text{final}} = 1$. Panels (a)–(e) show a close-up of the region $[0.4, 1] \times [0, 0.3]$, including the lambda shock, vortices, and shock interaction with the bottom boundary layer, especially the structures in $[0.5, 0.7] \times [0, 0.05]$. Panel (f) shows the limiting activity over the full computational domain $[0, 1] \times [0, 0.5]$.

For a quantitative comparison, the wall density profiles are also presented in Figure 9 with the reference data in [44]. The wall density converges with mesh refinement, and the

AF method matches the reference solution better with the same number of DoFs.

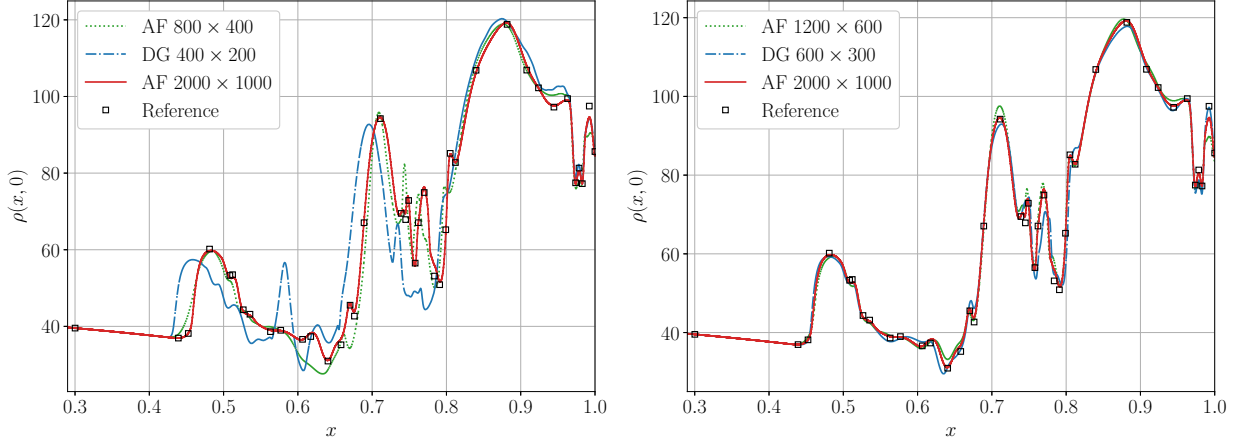


Figure 9: Example 4.6: viscous shock tube with $\text{Re} = 1000$. The density profile on the bottom wall $\rho(x, 0)$ compared with the reference data.

To see the advantage of the AF method in terms of efficiency, the errors of the wall density in the ℓ_1 norm versus wall-clock time are shown in Figure 10. Both methods are run using a Mac Studio with M3 Ultra chip with 24 cores. It is clearly seen that the AF method is more efficient with the same number of DoFs, and gives better resolution with the same wall-clock time.

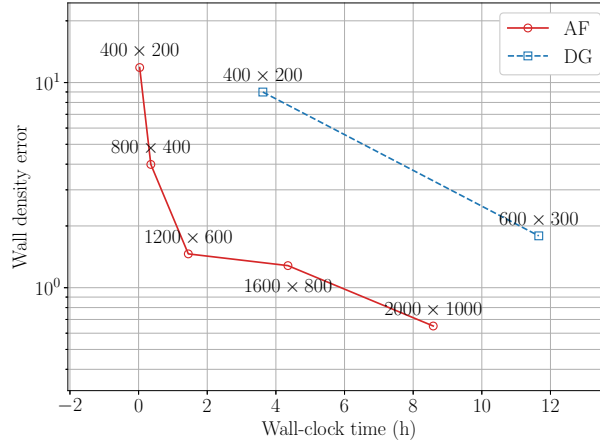


Figure 10: Example 4.6: two-dimensional viscous shock tube with shock–boundary-layer interaction with $\text{Re} = 1000$. The error of the wall density in the ℓ_1 norm versus the wall-clock time.

The last two examples consider nearly incompressible viscous flows, and heat conduction is neglected by setting $\kappa = 0$.

Example 4.7 (Lid-driven cavity). This classical benchmark describes flow in a square cavity $[0, 1] \times [0, 1]$ driven by a lid moving with velocity $(1, 0)$. The initial condition is $(\rho, u, v, p)(x, y, 0) = (1, 0, 0, 10^4/\gamma)$, and no-slip adiabatic conditions are imposed on all walls. The top lid moves with velocity $(1, 0)$, while the other three walls are stationary. The reference Mach number is $\text{Ma} = 0.01$ so this test falls in the incompressible limit. We consider

Reynolds numbers $\text{Re} = 100$ and 1000 . The test is run until $t_{\text{final}} = 40$ with a 128×128 mesh.

The horizontal velocity and streamlines are shown in Figure 11, together with the velocity cut-lines along $x = 0.5$ and $y = 0.5$. It is observed that the main recirculation and secondary corner vortices are well captured by the proposed AF method, which are comparable to the results in [45], and the velocity profiles match the reference solutions given in [46].

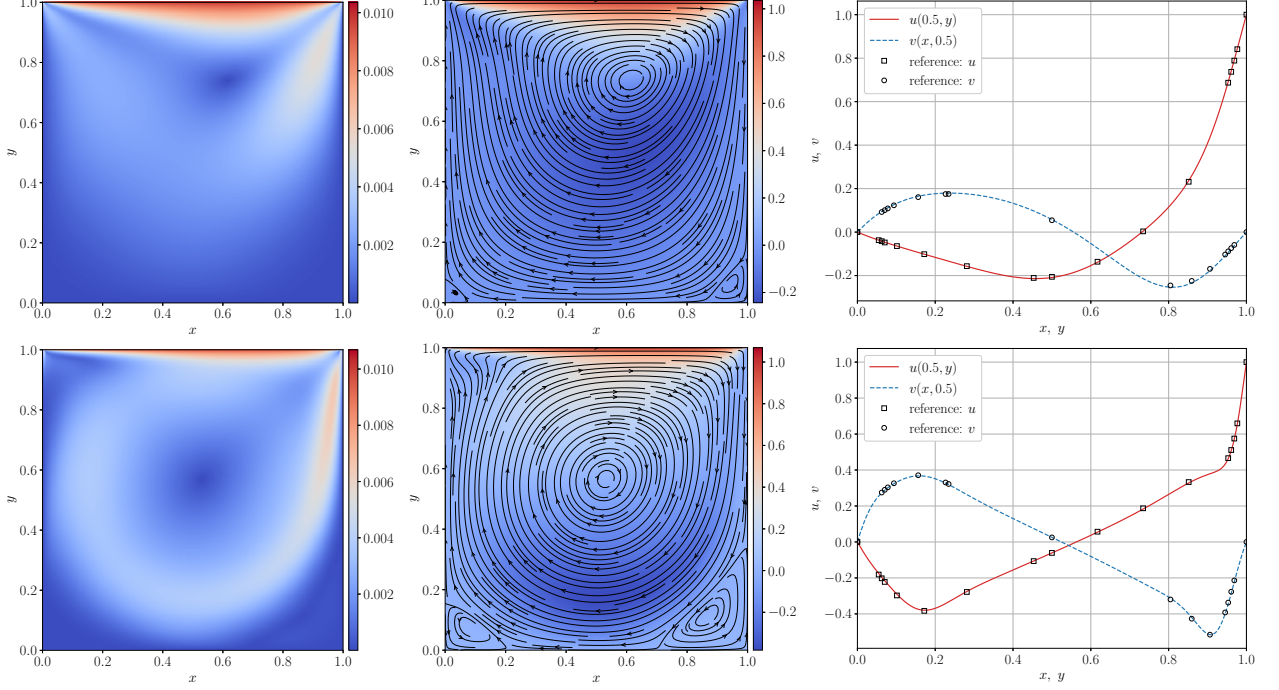


Figure 11: Lid-driven cavity Example 4.7 with $\text{Re} = 100$ (top) and 1000 (bottom). Left: Mach number, middle: horizontal velocity and streamlines, right: $u(0.5, y)$ and $v(x, 0.5)$ compared with the reference solutions.

Example 4.8 (Double shear layer). This test considers the evolution of thin shear layers at low Mach numbers, following the setup in [47]. The computational domain is $[0, 1] \times [0, 1]$ with periodic boundary conditions. The initial density and pressure are 1 and $10^4/\gamma$, and the velocity is

$$u = \begin{cases} \tanh(30(y - 0.25)), & \text{if } y \leq 0.5, \\ \tanh(30(0.75 - y)), & \text{otherwise,} \end{cases}$$

$$v = 0.05 \sin(2\pi x).$$

The dynamic viscosity is taken as $\mu = 2 \times 10^{-4}$ and the test is solved until $t_{\text{final}} = 1.8$. The Mach number is $\text{Ma} = 0.01$ and the Reynolds number is $\text{Re} = 10000$.

The vorticity $\partial_x v - \partial_y u$ at $t = 0.4, 0.8, 1.2, 1.8$ obtained with 200×200 cells is shown in Figure 12, which agrees well with the results in [47], thus the proposed AF method seems to work well in the low-Mach regime.

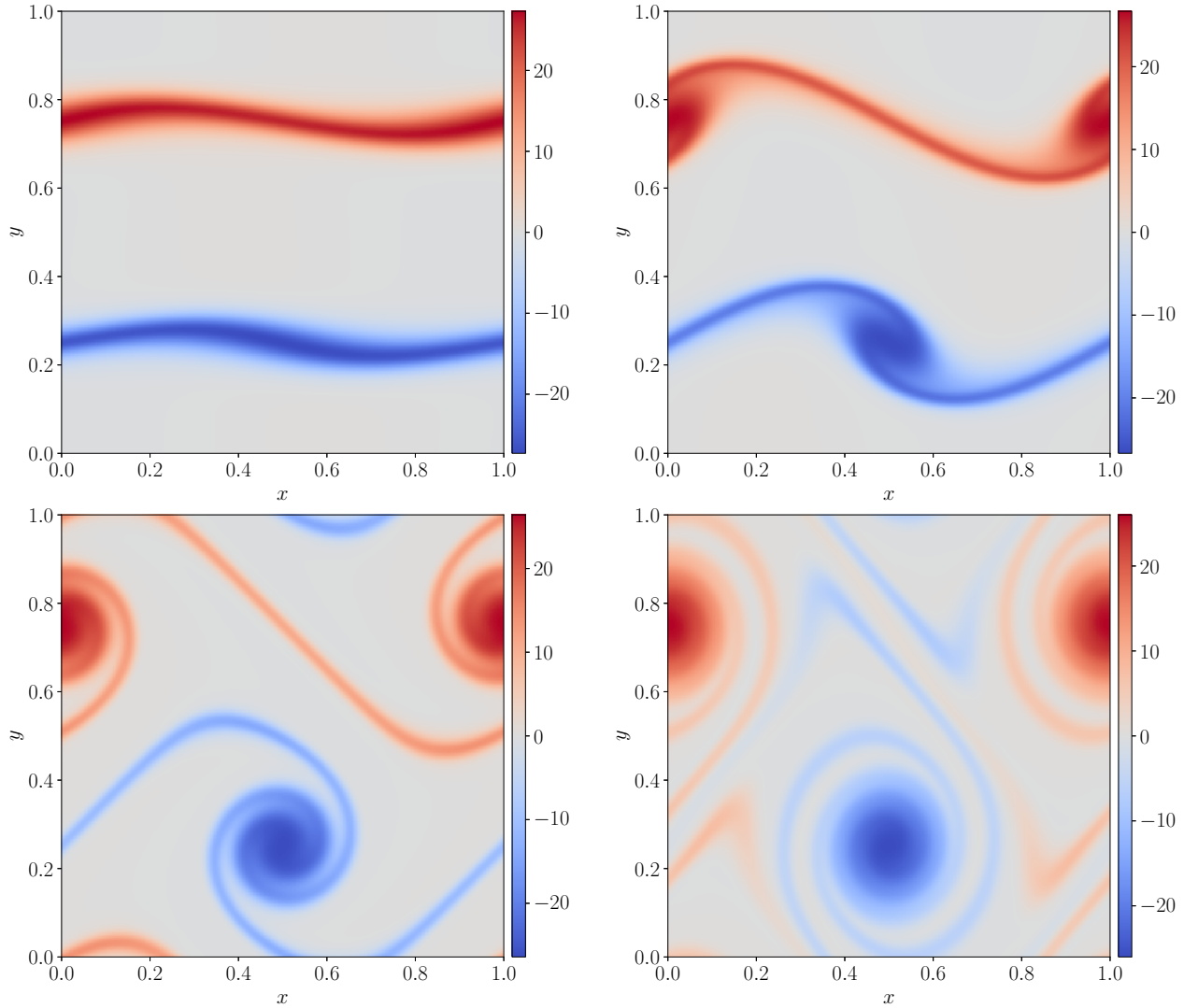


Figure 12: The vorticity of Example 4.8 with $\text{Re} = 10000$ at $t = 0.4, 0.8, 1.2, 1.8$ (from top left to bottom right).

5 Conclusions

We have developed a compact fourth-order positivity-preserving AF method for the one- and two-dimensional compressible Navier–Stokes equations on Cartesian meshes. The viscous flux divergence is discretized directly using compact fourth-order operators. The inviscid discretization achieves fourth-order accuracy while retaining the degrees of freedom of the standard third-order AF method. The CFL condition is also less restrictive than that of the standard third-order AF method due to the additional downwind point included in the biased stencil. Monolithic flux limiting treats the inviscid and viscous fluxes jointly in the cell-average update and maintains local conservation. Together with a scaling limiter for point values, it preserves density and pressure positivity and suppresses nonphysical oscillations near strong discontinuities. The numerical experiments demonstrate fourth-order convergence for smooth viscous flows and positivity preservation in problems involving strong discontinuities. In the two-dimensional viscous shock tube with shock–boundary-layer interaction, comparisons with the DG method show that AF attains comparable accuracy

with fewer degrees of freedom and less wall-clock time.

Acknowledgements

The first author was partially supported by National Natural Science Foundation of China (No. 12671481), 1+1+1 CUHK-CUHK(SZ)-GDSTC Joint Collaboration Fund (No. 2025A0505000079), University Development Fund (No. UDF01004388), and the Alexander von Humboldt Foundation Research Fellowship (No. CHN-1234352-HFST-P).

References

- [1] J. Slotnick, A. Khodadoust, J. Alonso, D. Darmofal, W. Gropp, E. Lurie, D. Mavriplis, CFD vision 2030 study: A path to revolutionary computational aerosciences, Tech. Rep. NF1676L-18332, NASA (2014).
- [2] Z. J. Wang, K. Fidkowski, R. Abgrall, F. Bassi, D. Caraeni, A. Cary, H. Deconinck, R. Hartmann, K. Hillewaert, H. T. Huynh, N. Kroll, G. May, P. Persson, B. van Leer, M. Visbal, High-order CFD methods: Current status and perspective, *Int. J. Numer. Meth. Fl.* 72 (8) (2013) 811–845. doi:10.1002/fld.3767.
- [3] S. K. Lele, Compact finite difference schemes with spectral-like resolution, *J. Comput. Phys.* 103 (1) (1992) 16–42. doi:10.1016/0021-9991(92)90324-R.
- [4] M. Dumbser, Arbitrary high order PNPM schemes on unstructured meshes for the compressible Navier-Stokes equations, *Comput. Fluids* 39 (1) (2010) 60–76. doi:10.1016/j.compfluid.2009.07.003.
- [5] F. Bassi, S. Rebay, A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier–Stokes equations, *J. Comput. Phys.* 131 (2) (1997) 267–279. doi:10.1006/jcph.1996.5572.
- [6] H. Huynh, A flux reconstruction approach to high-order schemes including discontinuous galerkin methods, in: 18th AIAA Computational Fluid Dynamics Conference AIAA 2007-4079 25 - 28 June 2007, Miami, FL, 2007.
- [7] P. E. Vincent, P. Castonguay, A. Jameson, A new class of high-order energy stable flux reconstruction schemes, *J. Sci. Comput.* 47 (1) (2011) 50–72. doi:10.1007/s10915-010-9420-z.
- [8] W. Barsukow, C. Klingenberg, S. Krotsch, On the equivalence of semi-discrete active flux and discontinuous Galerkin methods and a comparison of their performance (2026). doi:10.48550/ARXIV.2603.18160.
- [9] T. A. Eymann, P. L. Roe, Active flux schemes, in: 49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, American Institute of Aeronautics and Astronautics, Orlando, Florida, 2011. doi:10.2514/6.2011-382.

- [10] T. A. Eymann, P. L. Roe, Active flux schemes for systems, in: 20th AIAA Computational Fluid Dynamics Conference, Fluid Dynamics and Co-located Conferences, American Institute of Aeronautics and Astronautics, 2011. doi:10.2514/6.2011-3840.
- [11] T. A. Eymann, P. L. Roe, Multidimensional active flux schemes, in: 21st AIAA Computational Fluid Dynamics Conference, American Institute of Aeronautics and Astronautics, 2013. doi:10.2514/6.2013-2940.
- [12] B. Van Leer, Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection, *J. Comput. Phys.* 23 (3) (1977) 276–299. doi:10.1016/0021-9991(77)90095-X.
- [13] C. Helzel, D. Kerkmann, L. Scandurra, A new ADER method inspired by the active flux method, *J. Sci. Comput.* 80 (3) (2019) 1463–1497. doi:10.1007/s10915-019-00988-1.
- [14] E. Chudzik, C. Helzel, M. Lukáčová-Medvid’ová, Active flux methods for hyperbolic systems using the method of bicharacteristics, *J. Sci. Comput.* 99 (1) (2024) 16. doi:10.1007/s10915-024-02462-z.
- [15] W. Barsukow, An active flux method for the Euler equations based on the exact acoustic evolution operator, *J. Sci. Comput.* 108 (3) (2026). doi:10.1007/s10915-026-03405-6.
- [16] E. Chudzik, C. Helzel, A. Porfetye, A fully discrete truly multidimensional active flux method for the two-dimensional Euler equations, *SIAM J. Sci. Comput.* 48 (5) (2026) A2572–A2598. doi:10.1137/25m1770928.
- [17] R. Abgrall, A combination of residual distribution and the active flux formulations or a new class of schemes that can combine several writings of the same hyperbolic problem: Application to the 1D Euler equations, *Commun. Appl. Math. Comput.* 5 (1) (2023) 370–402. doi:10.1007/s42967-021-00175-w.
- [18] J. Duan, W. Barsukow, C. Klingenberg, Active flux methods for hyperbolic conservation laws – flux vector splitting and bound-preservation, *SIAM J. Sci. Comput.* 47 (2) (2025) A811–A837. doi:10.1137/24M1658887.
- [19] R. Abgrall, W. Barsukow, Extensions of active flux to arbitrary order of accuracy, *ESAIM: Math. Model. Numer. Anal.* 57 (2) (2023) 991–1027. doi:10.1051/m2an/2023004.
- [20] W. Barsukow, P. Chandrashekar, C. Klingenberg, L. Lechner, A generalized active flux method of arbitrarily high order in two dimensions, *Comput. Fluids* 307 (2026) 106886. doi:10.1016/j.compfluid.2025.106886.
- [21] R. Abgrall, W. Boscheri, Y. Liu, Virtual finite element and hyperbolic problems: The PAMPA algorithm, *J. Comput. Phys.* 546 (2026) 114521. doi:10.1016/j.jcp.2025.114521.
- [22] J. Duan, P. Chandrashekar, C. Klingenberg, Active flux for ideal magnetohydrodynamics: A positivity-preserving scheme with the Godunov-Powell source term (Jun. 2025). doi:10.48550/arxiv.2506.04857.

- [23] M. Liu, D. Pang, R. Abgrall, K. Wu, An active-flux-type scheme for ideal MHD with provable positivity and a point-value-level discrete divergence-free constraint, *J. Comput. Phys.* 567 (2026) 115319. doi:10.1016/j.jcp.2026.115319.
- [24] H. Nishikawa, P. L. Roe, Third-order active-flux scheme for advection diffusion: Hyperbolic diffusion, boundary condition, and Newton solver, *Comput. Fluids* 125 (2016) 71–81. doi:10.1016/j.compfluid.2015.10.020.
- [25] J. Duan, W. Barsukow, C. Klingenberg, An asymptotic-preserving active flux scheme for the hyperbolic heat equation in the diffusive scaling (2025). doi:10.48550/arxiv.2508.05166.
- [26] F. He, Towards a New-Generation Numerical Scheme for the Compressible Navier-Stokes Equations with the Active Flux Method, Ph.D. thesis, University of Michigan, Ann Arbor, Michigan (2021). doi:10.7302/2732.
- [27] J. Duan, A compact fourth-order doubly conservative active flux method with maximum-principle-preserving limiting for degenerate convection–diffusion equations (2026). doi:10.48550/arxiv.2510.11527v2.
- [28] E. Chudzik, C. Helzel, D. Kerkmann, The Cartesian grid active flux method: Linear stability and bound preserving limiting, *Appl. Math. Comput.* 393 (2021) 125501. doi:10.1016/j.amc.2020.125501.
URL <https://www.sciencedirect.com/science/article/pii/S0096300320304598>
- [29] R. Abgrall, M. Jiao, Y. Liu, K. Wu, Bound-preserving point-average-moment polynomial-interpreted (PAMPA) scheme: One-dimensional case, *Commun. Comput. Phys.* 39 (1) (2025) 29–58. doi:10.4208/cicp.oa-2024-0266.
- [30] R. Abgrall, M. Jiao, Y. Liu, K. Wu, A novel and simple invariant-domain-preserving framework for pampa scheme: 1d case, *SIAM J. Sci. Comput.* 47 (6) (2025) A3536–A3565. doi:10.1137/24m1715015.
- [31] R. Abgrall, Y. Liu, Robust PAMPA scheme in the DG formulation on unstructured triangular meshes: Bound preservation, oscillation elimination, and boundary conditions, *J. Comput. Phys.* 561 (2026) 114962. doi:10.1016/j.jcp.2026.114962.
- [32] C. Fan, X. Zhang, J. Qiu, Positivity-preserving high order finite volume hybrid Hermite WENO schemes for compressible Navier–Stokes equations, *J. Comput. Phys.* 445 (2021) 110596. doi:10.1016/j.jcp.2021.110596.
- [33] C. Fan, X. Zhang, J. Qiu, Positivity-preserving high order finite difference WENO schemes for compressible Navier–Stokes equations, *J. Comput. Phys.* 467 (2022) 111446. doi:10.1016/j.jcp.2022.111446.
- [34] Y. Lin, J. Chan, I. Tomas, A positivity preserving strategy for entropy stable discontinuous Galerkin discretizations of the compressible Euler and Navier-Stokes equations, *J. Comput. Phys.* 475 (2023) 111850. arXiv:2201.11816, doi:10.1016/j.jcp.2022.111850.

- [35] N. K. Yamaleev, J. Upperman, High-order positivity-preserving entropy stable schemes for the 3-D compressible Navier–Stokes equations, *J. Sci. Comput.* 95 (1) (2023) 11. doi:10.1007/s10915-023-02136-2.
- [36] X. Zhang, On positivity-preserving high order discontinuous Galerkin schemes for compressible Navier–Stokes equations, *J. Comput. Phys.* 328 (2017) 301–343. doi:10.1016/j.jcp.2016.10.002.
- [37] R. Kamakoti, C. Pantano, High-order narrow stencil finite-difference approximations of second-order derivatives involving variable coefficient, *SIAM J. Sci. Comput.* 31 (6) (2010) 4222–4243.
- [38] D. Kuzmin, Monolithic convex limiting for continuous finite element discretizations of hyperbolic conservation laws, *Comput. Methods Appl. Mech. Engrg.* 361 (2020) 112804. doi:10.1016/j.cma.2019.112804.
- [39] A. Jameson, W. Schmidt, E. Turkel, Solutions of the Euler equations by finite volume methods using Runge-Kutta time-stepping schemes, *AIAA J.* 1259 (1981).
- [40] F. Ducros, V. Ferrand, F. Nicoud, C. Weber, D. Darracq, C. Gacherieu, T. Poinso, Large-eddy simulation of the shock/turbulence interaction, *J. Comput. Phys.* 152 (2) (1999) 517–549. doi:10.1006/jcph.1999.6238.
- [41] R. Becker, Stosswelle und Detonation, *Physik* 8 (1923) 321.
- [42] J. Upperman, N. K. Yamaleev, Positivity-preserving entropy stable schemes for the 1-D compressible Navier-Stokes equations: First-order approximation, *J. Comput. Phys.* 466 (2022) 111355. doi:10.1016/j.jcp.2022.111355.
- [43] V. Daru, C. Tenaud, Evaluation of TVD high resolution schemes for unsteady viscous shocked flows, *Comput. Fluids* 30 (1) (2001) 89–113. doi:10.1016/s0045-7930(00)00006-2.
- [44] G. Zhou, K. Xu, F. Liu, Grid-converged solution and analysis of the unsteady viscous flow in a two-dimensional shock tube, *Phys. Fluids* 30 (1) (2018). doi:10.1063/1.4998300.
- [45] M. Tavelli, M. Dumbser, A pressure-based semi-implicit space-time discontinuous Galerkin method on staggered unstructured meshes for the solution of the compressible Navier-Stokes equations at all Mach numbers, *J. Comput. Phys.* 341 (2017) 341–376. doi:10.1016/j.jcp.2017.03.030.
- [46] U. Ghia, K. N. Ghia, C. T. Shin, High-Re solutions for incompressible flow using the Navier-Stokes equations and a multigrid method, *J. Comput. Phys.* 48 (3) (1982) 387–411. doi:10.1016/0021-9991(82)90058-4.
- [47] W. Boscheri, S. Busto, M. Dumbser, An all Mach number semi-implicit hybrid finite volume/virtual element method for compressible viscous flows on Voronoi meshes, *Comput. Methods Appl. Mech. Engrg.* 433 (2025) 117502. doi:10.1016/j.cma.2024.117502.