Structure-preserving active flux method for the compressible Euler equations with gravity

Jiahui Zhang* Christian Klingenberg†

Abstract

In this paper, we propose a structure-preserving active flux method for the compressible Euler equations under gravitational fields, ensuring exactness for the generalized hydrostatic equilibrium state and the preservation of positivity for both density and pressure. This approach entails the update of the cell average and the point values simultaneously. To obtain the equilibrium preservation of the scheme for the cell average, the strategy pursued is to reformulate the gravitational term using the steady state solution, followed by the application of a decomposition algorithm. For the point values, we introduce minor adjustments to the Lax-Friedrichs flux vector splitting, coupled with the same finite difference operator performed on the flux and source term to restore the well-balancedness. Furthermore, a posteriori Multi-dimensional Optimal Order Detection (MOOD) paradigm is adopted for achieving positivity preservation and oscillation-control near strong discontinuities. A series of numerical experiments are performed to verify the high-order accuracy of our proposed method, along with its exact equilibrium and positivity preservation, and its ability to capture small perturbations at or close to steady flow without spurious oscillations.

Keywords: Euler equations with gravity; active flux method; well-balanced; positivity-preserving; source term treatment; hydrostatic equilibrium.

1 Introduction

This paper concentrates on discussing the compressible Euler equations which incorporate the influence of the gravitational force. Many interesting phenomena in astrophysical and atmospheric are modelled by this system, such as core-collapse supernova simulation, stellar evolution, numerical weather forecasting, etc. In the d-dimensional case, the governing equations for the conservation of mass, momentum and energy, together with

*Institute for Mathematics, Johannes Gutenberg University of Mainz, Mainz, 55128, Germany. E-mail: jizhang@uni-mainz.de.

†Corresponding author. Institute of Mathematics, University of Würzburg, Würzburg, 97074, Germany. E-mail: christian. klingenberg@uni-wuerzburg.de

the gravitational field, are expressed as follows,

$$\begin{cases} \rho_t + \nabla \cdot (\rho \boldsymbol{u}) = 0, \\ (\rho \boldsymbol{u})_t + \nabla \cdot (\rho \boldsymbol{u} \otimes \boldsymbol{u} + p\mathbf{I}_d) = -\rho \nabla \phi, \\ E_t + \nabla \cdot ((E+p)\boldsymbol{u}) = -\rho \boldsymbol{u} \cdot \nabla \phi, \end{cases} \tag{1.1}$$

where $\rho, \boldsymbol{u}, p$ denote the mass density, the velocity of the gas, and the pressure, respectively. $\rho\boldsymbol{u}$ corresponds to the momentum vector. The total fluid energy E is defined as the sum of the internal and kinetic energy, i.e. $E = \rho e + \frac{1}{2}\rho\|\boldsymbol{u}\|^2$, where e denotes the internal energy. $\phi = \phi(\boldsymbol{x})$ represents the time-independent gravitational potential. $\mathbf{I}_d$ is the identity matrix and the operators ∇, $\nabla\cdot$, $\otimes$ signify the gradient, divergence and tensor product. To fully close this system of equations, it is essential to incorporate an equation of state (EoS) that establishes a relationship between the specific internal energy, density, and pressure. It is given by

$$p = p(\rho, e) = (\gamma - 1)\rho e = (\gamma - 1)\left(E - \frac{1}{2}\rho\|\boldsymbol{u}\|^2\right), \tag{1.2}$$

with γ being the ratio of specific heats.

As a typical example of hyperbolic balance laws, one significant aspect of such models is their tendency to exhibit non-trivial steady state solutions, where the non-zero flux gradient precisely balances the source term. In particular, they admit hydrostatic stationary solutions with $\boldsymbol{u} = 0$, corresponding to a mechanical equilibrium where pressure balances external forces such as gravity,

$$\rho = \rho(\boldsymbol{x}), \quad \boldsymbol{u} = 0, \quad \nabla p = -\rho \nabla \phi. \tag{1.3}$$

Since this system of equations is underdetermined, additional assumptions are required to fully characterize the equilibrium. One special class of equilibrium is the isothermal hydrostatic state, where the temperature is a constant, i.e. $T(x) = T_0$. For an ideal gas, it is given by

$$\rho = \rho_0 \exp\left(-\frac{\phi}{RT_0}\right), \quad \boldsymbol{u} = 0, \quad p = p_0 \exp\left(-\frac{\phi}{RT_0}\right), \tag{1.4}$$

where R is the gas constant, ρ_0, p_0 are constants satisfying $p_0 = \rho_0 RT_0$. Another class of equilibrium state is polytropic hydrostatic state given by

$$\rho = \left(\frac{\gamma - 1}{K\gamma}(c - \phi)\right)^{\frac{1}{\gamma-1}}, \quad \boldsymbol{u} = 0, \quad p = \frac{1}{K^{\frac{1}{\gamma-1}}}\left(\frac{\gamma - 1}{\gamma}(c - \phi)\right)^{\frac{\gamma}{\gamma-1}}, \tag{1.5}$$

with c being a constant, K is a constant defined by the polytropic relation $p = K\rho^\gamma$.

The existence of source terms presents a well-known challenge for the standard numerical schemes, which often fail to preserve the steady state even when highly refined meshes are utilized. This promotes the development of methods that can exactly maintain steady state solutions at the discrete level while accurately resolving small perturba-

tions on relatively coarse grids. Such methods are referred to as well-balanced schemes, first introduced by Bermudez and Vazquez [11]. In recent years, considerable attention has been devoted to the design of well-balanced schemes for the Euler equations with gravitation. The first work was constructed by LeVeque and Bale [25], where a quasi-steady wave-propagation algorithm was extended to an ideal gas under a static gravitational field. Later, many efforts have been made in this direction, including the finite difference methods [20, 33], finite volume methods [23, 24, 26], and discontinuous Galerkin schemes [27,32,37]. Another class of methods [10,30] were designed to maintain consistency between the discrete discretization and the steady-state partial differential equations of interest. It is free of additional assumptions and a priori knowledge of the hydrostatic solution. In addition, several studies [22, 34–36] have explored a generalized more complicated moving equilibrium with the non-zero velocity.

Another challenging in developing structure-preserving method is the property of positivity preservation. Preserving the positivity of the density and pressure is essential for the robustness of numerical algorithms. Otherwise, the computation might break down, especially for low density problems such as blast waves and low pressure problems such as high Mach number astrophysical jets. In recent years, positivity-preserving techniques have gained significant popularity for high-order schemes. The most commonly used is a simple linear scaling limiter proposed by Zhang and Shu [38] for the reconstructed polynomials in finite volume or discontinuous Galerkin (DG) schemes. It can be generalized to Euler equations without the source term [39] and with source term [40]. An alternative strategy is Multi-dimensional Optimal Order Detection (MOOD) paradigm, which was first introduced by Clain et al. [14] and further developed by Diot et al. [15]. Instead of relying on a priori scaling of the solution, MOOD adopts a posteriori detection and recomputation strategy. When problematic solutions are detected after each time update, the scheme will reduce the local polynomial degree before recomputing the solution.

Recently, the active flux method is a relatively new variant of a finite volume method which is attracting increasing interest. Both cell averages and point values at cell interfaces are used as degrees of freedom. This approach is characterized by a compact, globally continuous reconstruction, the avoidance of Riemann solvers at cell interfaces, while maintaining high-order accuracy and conservation properties. In the early developments of the active flux method, exact or approximate evolution operators are employed to evolve the point values in conjunction with Simpson's rule to compute the flux quadrature in time. Such techniques have been applied to a variety of hyperbolic conservation laws, for instance, the linear equations [19,29], Burgers' equation [6,12], one-dimensional shallow water equations [7], and the compressible Euler equations [6, 13, 18], etc. These resulting schemes are fully-discrete. However, some limitations exist since it is difficult to obtain evolution operators for nonlinear systems of equations, especially in multi-dimensional case. This has led [1, 3, 4, 16] to consider semi-discrete formulations, where the cell average and point values are used to discretize the spatial derivatives and subsequently advanced in time using suitable time integration methods. As outlined in [2], this approach applies readily to nonlinear problems free of evolution operators, at the

expense of a reduced CFL condition. Abgrall et.al. [5] developed a novel well-balanced active flux method for the one-dimensional shallow water equations. Furthermore, recent work by Barsukow et.al. [8] has shown that the semi-discrete active flux and discontinuous Galerkin are the same scheme for linear problems in one and several dimensions, and in a qualified sense for nonlinear one-dimensional systems. Because of the higher polynomial degree per degree of freedom for DG, the authors show that active flux is more economical than DG for a target error threshold. Nevertheless, the two schemes exhibit significant differences in their representation of the solution, implementation, and computational cost.

In this paper, we develop a structure-preserving scheme for solving the Euler equations with gravitation, utilizing the semi-discrete active flux approach. The designed scheme aims to simultaneously maintain the generalized hydrostatic equilibrium state (1.3) with high-order accuracy and preserve the positivity of the density and pressure. For the evolution of the cell average, we adopt special source term treatment, inspired by [26], to preserve the well-balanced property. The gravitational potential is reformulated into an equivalent form based on the corresponding hydrostatic equilibrium solution. For the evolution of the point values, we employ the strategy of Lax-Friedrichs flux vector splitting developed in [16], which shows better performance than the Jacobian splitting [3, 4] for strong discontinuities. Drawing inspiration from [33], where the source term mimics the same weighted essentially non-oscillatory approximation to the flux term, we apply an identical finite difference operator to both the flux term and the source term. The well-balanced property is also achieved by making minor modification to the Lax-Friedrichs flux vector splitting. Besides the well-balancedness, the constructed scheme is also positivity-preserving. To address nonphysical solutions or spurious oscillations near discontinuities generated by the high-order scheme, we apply a posteriori MOOD paradigm introduced in [5, 14] to both the cell average and point values. Theoretical analysis and numerical experiments demonstrate the well-balanced and positivity-preserving nature of the proposed scheme, as well as its ability to achieve high resolution.

An outline of this paper is as follows. Section 2 introduces the well-balanced active flux method for the Euler equations with gravity, which can preserve both the isothermal and polytropic equilibrium simultaneously. In addition to the well-balanced property, the proposed scheme also preserves the positivity of the density and pressure. In Section 3, numerical experiments in different circumstances are illustrated to validate all the mathematical properties of the active flux method. In the end, some concluding remarks are provided in Section 4.

2 Well-balanced active flux method

In this section, we present the well-balanced and positivity-preserving active flux method for the Euler equations with gravitation, with the aim of preserving the general hydrostatic states (1.3). Here we restrict ourselves to two dimensions, the Euler equations

(1.1) take the form of

$$\boldsymbol{U}_t + \boldsymbol{F}(\boldsymbol{U})_x + \boldsymbol{G}(\boldsymbol{U})_y = \boldsymbol{S}(\boldsymbol{U}, \phi), \tag{2.1}$$

with

$$\boldsymbol{F}(\boldsymbol{U}) = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ (E+p)u \end{bmatrix}, \ \boldsymbol{G}(\boldsymbol{U}) = \begin{bmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ (E+p)v \end{bmatrix}, \ \boldsymbol{S}(\boldsymbol{U}, \phi) = \begin{bmatrix} 0 \\ -\rho\phi_x \\ -\rho\phi_y \\ -\rho u\phi_x - \rho v\phi_y \end{bmatrix}. \tag{2.2}$$

Assuming that the computational domain Ω is divided into $Nx \times Ny$ rectangular cells I_{ij}, i.e.

$$\Omega = \bigcup_{i=1}^{Nx}\bigcup_{j=1}^{Ny} I_{ij} = \bigcup_{i=1}^{Nx}\bigcup_{j=1}^{Ny} [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}].$$

Without loss of generality, uniform meshes are considered in this work. The cell lengths are denoted by $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$, $\Delta y = y_{j+\frac{1}{2}} - y_{j-\frac{1}{2}}$, with corresponding cell centers located at $(x_i, y_j) = (\frac{1}{2}(x_{i-\frac{1}{2}} + x_{i+\frac{1}{2}}), \frac{1}{2}(y_{j-\frac{1}{2}} + y_{j+\frac{1}{2}}))$. The degrees of freedom involves the cell average $\overline{\boldsymbol{U}}_{ij}(t) = \frac{1}{\Delta x \Delta y}\int_{I_{ij}} \boldsymbol{U}(x, y, t)\, dxdy$ as well as the point values $\boldsymbol{U}_\sigma$ located at the cell interfaces. Here $\boldsymbol{U}(x, y, t)$ is the high-order numerical solution, $\sigma = (i+\frac{1}{2}, j), (i, j+\frac{1}{2}), (i+\frac{1}{2}, j+\frac{1}{2})$. Specifically, the face-centered values $\boldsymbol{U}_{i+\frac{1}{2},j}(t) = \boldsymbol{U}(x_{i+\frac{1}{2}}, y_j, t), \boldsymbol{U}_{i,j+\frac{1}{2}} = \boldsymbol{U}(x_i, y_{j+\frac{1}{2}}, t)$ and the values at the corner $\boldsymbol{U}_{i+\frac{1}{2},j+\frac{1}{2}} = \boldsymbol{U}(x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}, t)$ are included for the third-order well-balanced active flux method.

2.1 Well-balanced evolution of cell average

Following the procedure of [16], the well-balanced evolution of the cell average can be expressed as:

$$\frac{d}{dt}\overline{\boldsymbol{U}}_{ij} = -\frac{1}{\Delta x}(\widehat{\boldsymbol{F}}_{i+\frac{1}{2},j} - \widehat{\boldsymbol{F}}_{i-\frac{1}{2},j}) - \frac{1}{\Delta y}(\widehat{\boldsymbol{G}}_{i,j+\frac{1}{2}} - \widehat{\boldsymbol{G}}_{i,j-\frac{1}{2}}) + \frac{1}{\Delta x_i \Delta y_j}\overline{\boldsymbol{S}}_{ij}, \tag{2.3}$$

with the numerical fluxes defined as follows,

$$\begin{aligned} \widehat{\boldsymbol{F}}_{i+\frac{1}{2},j} &= \frac{1}{\Delta y}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \boldsymbol{F}(\boldsymbol{U}(x_{i+\frac{1}{2}}, y))\, dy, \\ \widehat{\boldsymbol{G}}_{i,j+\frac{1}{2}} &= \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \boldsymbol{G}(\boldsymbol{U}(x, y_{j+\frac{1}{2}}))\, dx. \end{aligned} \tag{2.4}$$

To attain third-order accuracy, Simpson's formula is employed to evaluate (2.4), yielding

$$\widehat{\boldsymbol{F}}_{i+\frac{1}{2},j} = \frac{1}{6}(\boldsymbol{F}(\boldsymbol{U}_{i+\frac{1}{2},j-\frac{1}{2}}) + 4\boldsymbol{F}(\boldsymbol{U}_{i+\frac{1}{2},j}) + \boldsymbol{F}(\boldsymbol{U}_{i+\frac{1}{2},j+\frac{1}{2}})). \tag{2.5}$$

The numerical flux $\widehat{\boldsymbol{G}}_{i,j+\frac{1}{2}}$ can be obtained in an analogous manner.

Moreover, $\overline{\boldsymbol{S}}_{ij}$ is the high-order approximation of $\int_{I_{ij}} \boldsymbol{S}(\boldsymbol{U}(x,y,t),\phi(x,y))\,dxdy$. Direct computation of this quantity generally leads to a loss of the well-balanced property of the scheme. Therefore, a special treatment of the source term is necessary to maintain the equilibrium state. Suppose that the target hydrostatic equilibrium $\{\rho^e(x,y),\boldsymbol{u}^e(x,y),p^e(x,y)\}$ is known a priori, which implies the following relation holds:

$$\nabla p^e = \rho^e \nabla \phi, \quad \boldsymbol{u}^e = 0. \tag{2.6}$$

Taking the second component of the source term as an instance, which is denoted by $\boldsymbol{S}_{ij}^{(2)}$. We follow the similar technique in [26,32] to reformulate and decompose its integral as

$$\begin{aligned}\overline{\boldsymbol{S}}_{ij}^{(2)} &= \int_{I_{ij}} -\rho\phi_x\,dxdy = \int_{I_{ij}} \frac{\rho}{\rho^e}p_x^e\,dxdy = \int_{I_{ij}} \left(\frac{\rho}{\rho^e} - \frac{\bar{\rho}_{ij}}{\bar{\rho}_{ij}^e} + \frac{\bar{\rho}_{ij}}{\bar{\rho}_{ij}^e}\right) p_x^e\,dxdy \\ &= \int_{I_{ij}} \left(\frac{\rho}{\rho^e} - \frac{\bar{\rho}_{ij}}{\bar{\rho}_{ij}^e}\right) p_x^e\,dxdy + \frac{\bar{\rho}_{ij}}{\bar{\rho}_{ij}^e}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} (p^e_{i+\frac{1}{2},j} - p^e_{i-\frac{1}{2},j})\,dy.\end{aligned} \tag{2.7}$$

Here $\bar{\cdot}_{ij}$ denotes the cell average of the corresponding quantity in cell I_{ij}. The first integral of the final identity can be approximated by the standard Gauss quadrature rule, with $\{(x_{i_\mu},y_{j_\nu}),\omega_{\mu\nu}\}$ being the Gauss quadrature nodes and associated weights. The second integral is evaluated using a Simpson formula similar to that in (2.5). For ease of understanding, we have

$$\begin{aligned}\overline{\boldsymbol{S}}_{ij}^{(2)} \approx &\sum_{\mu,\nu}\omega_{\mu\nu}\left(\frac{\rho_{i_\mu,j_\nu}}{\rho^e_{i_\mu,j_\nu}} - \frac{\bar{\rho}_{ij}}{\bar{\rho}_{ij}^e}\right) p_x^e(x_{i_\mu},y_{j_\nu}) \\ &+ \frac{\bar{\rho}_{ij}}{\bar{\rho}_{ij}^e}\frac{\Delta y}{6}\left((p^e_{i+\frac{1}{2},j-\frac{1}{2}} + 4p^e_{i+\frac{1}{2},j} + p^e_{i+\frac{1}{2},j+\frac{1}{2}}) - (p^e_{i-\frac{1}{2},j-\frac{1}{2}} + 4p^e_{i-\frac{1}{2},j} + p^e_{i-\frac{1}{2},j+\frac{1}{2}})\right).\end{aligned} \tag{2.8}$$

The third and forth part of the source term $\overline{\boldsymbol{S}}_{ij}^{(3)}$, $\overline{\boldsymbol{S}}_{ij}^{(4)}$ can be treated in a similar manner.

2.2 Well-balanced evolution of point values

Apart from the evolution of the cell average, third-order active flux method also requires the evolution of the point values at the cell interfaces, which yields a consistent approximation of the flux derivative. In this paper, we update the point values by utilizing the flux vector splitting proposed in [16], which are given by

$$\frac{d}{dt}\boldsymbol{U}_{i+\frac{1}{2},j+\frac{1}{2}} = -[\mathcal{D}_1^+\boldsymbol{F}^+ + \mathcal{D}_1^-\boldsymbol{F}^-]_{i+\frac{1}{2},j+\frac{1}{2}} - [\mathcal{D}_2^+\boldsymbol{G}^+ + \mathcal{D}_2^-\boldsymbol{G}^-]_{i+\frac{1}{2},j+\frac{1}{2}} + \boldsymbol{S}_{i+\frac{1}{2},j+\frac{1}{2}}, \tag{2.9}$$

$$\frac{d}{dt}\boldsymbol{U}_{i+\frac{1}{2},j} = -[\mathcal{D}_1^+\boldsymbol{F}^+ + \mathcal{D}_1^-\boldsymbol{F}^-]_{i+\frac{1}{2},j} - [\mathcal{D}_2\boldsymbol{G}]_{i+\frac{1}{2},j} + \boldsymbol{S}_{i+\frac{1}{2},j}, \tag{2.10}$$

$$\frac{d}{dt}\boldsymbol{U}_{i,j+\frac{1}{2}} = -[\mathcal{D}_1\boldsymbol{F}]_{i,j+\frac{1}{2}} - [\mathcal{D}_2^+\boldsymbol{G}^+ + \mathcal{D}_2^-\boldsymbol{G}^-]_{i,j+\frac{1}{2}} + \boldsymbol{S}_{i,j+\frac{1}{2}}. \tag{2.11}$$

Here the flux $\boldsymbol{F} = \boldsymbol{F}^+ + \boldsymbol{F}^-$ and satisfies $\lambda(\frac{\partial \boldsymbol{F}^+}{\partial \boldsymbol{U}}) \geq 0, \lambda(\frac{\partial \boldsymbol{F}^-}{\partial \boldsymbol{U}}) \leq 0$. Considering the well-balanced property, we adopt the Lax-Friedrichs flux vector splitting with slight modification, where $\boldsymbol{F}^\pm$ is defined by

$$\boldsymbol{F}^\pm = \frac{1}{2}(\boldsymbol{F}(\boldsymbol{U}) \pm \alpha \widetilde{\boldsymbol{U}}), \quad \widetilde{\boldsymbol{U}} = \boldsymbol{U} - \boldsymbol{U}^e. \tag{2.12}$$

$\boldsymbol{U}^e$ is the conservative variables at the equilibrium state, which can be obtained from (2.6) and α is taken as the global maximum of the spectral radius of $\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{U}}$ over the computation domain. The flux $\boldsymbol{G}^\pm$ can be defined analogously. The cell-centered point value is obtained through the bi-parabolic reconstruction using the cell average and point values at the cell interfaces,

$$\boldsymbol{U}_{ij} = \frac{1}{16}\left[36\overline{\boldsymbol{U}}_{ij} - 4(\boldsymbol{U}_{i-\frac{1}{2},j} + \boldsymbol{U}_{i+\frac{1}{2},j} + \boldsymbol{U}_{i,j-\frac{1}{2}} + \boldsymbol{U}_{i,j+\frac{1}{2}}) - (\boldsymbol{U}_{i-\frac{1}{2},j-\frac{1}{2}} + \boldsymbol{U}_{i-\frac{1}{2},j+\frac{1}{2}} + \boldsymbol{U}_{i+\frac{1}{2},j-\frac{1}{2}} + \boldsymbol{U}_{i+\frac{1}{2},j+\frac{1}{2}})\right] \tag{2.13}$$

More details can be found in [4]. We can easily obtain the fluxes on the edges and the cell-centered point value with the definition of (2.12). Furthermore, $\mathcal{D}_1^\pm, \mathcal{D}_2^\pm$ are the finite difference operators acquired from differentiating the bi-parabolic interpolation of $\boldsymbol{F}^\pm, \boldsymbol{G}^\pm$ in the cell I_{ij},

$$\begin{aligned}
[\mathcal{D}_1^+ \boldsymbol{F}^+]_{i+\frac{1}{2},j_\beta} &= \frac{1}{\Delta x}(\boldsymbol{F}^+_{i-\frac{1}{2},j_\beta} - 4\boldsymbol{F}^+_{i,j_\beta} + 3\boldsymbol{F}^+_{i+\frac{1}{2},j_\beta}),\\
[\mathcal{D}_1^- \boldsymbol{F}^-]_{i+\frac{1}{2},j_\beta} &= \frac{1}{\Delta x}(-3\boldsymbol{F}^-_{i+\frac{1}{2},j_\beta} + 4\boldsymbol{F}^-_{i+1,j_\beta} - \boldsymbol{F}^-_{i+\frac{3}{2},j_\beta}),\\
[\mathcal{D}_1 \boldsymbol{F}]_{i,j_\beta} &= \frac{1}{\Delta x}(\boldsymbol{F}_{i+\frac{1}{2},j_\beta} - \boldsymbol{F}_{i-\frac{1}{2},j_\beta}),\\
[\mathcal{D}_2^+ \boldsymbol{G}^+]_{i_\beta,j+\frac{1}{2}} &= \frac{1}{\Delta y}(\boldsymbol{G}^+_{i_\beta,j-\frac{1}{2}} - 4\boldsymbol{G}^+_{i_\beta,j} + 3\boldsymbol{G}^+_{i_\beta,j+\frac{1}{2}}),\\
[\mathcal{D}_2^- \boldsymbol{G}^-]_{i_\beta,j+\frac{1}{2}} &= \frac{1}{\Delta y}(-3\boldsymbol{G}^-_{i_\beta,j+\frac{1}{2}} + 4\boldsymbol{G}^-_{i_\beta,j+1} - \boldsymbol{G}^-_{i_\beta,j+\frac{3}{2}}),\\
[\mathcal{D}_2 \boldsymbol{G}]_{i_\beta,j} &= \frac{1}{\Delta y}(\boldsymbol{G}_{i_\beta,j+\frac{1}{2}} - \boldsymbol{G}_{i_\beta,j-\frac{1}{2}}),
\end{aligned} \tag{2.14}$$

where $i_\beta = \{i - \frac{1}{2}, i, i + \frac{1}{2}\}$, $j_\beta = \{j - \frac{1}{2}, j, j + \frac{1}{2}\}$.

For the source term approximation, we employ the relation (2.6) and reformulate the second component of the source term as

$$\boldsymbol{S}^{(2)}_{i+\frac{1}{2},j+\frac{1}{2}} = [-\rho\phi_x]_{i+\frac{1}{2},j+\frac{1}{2}} = \left(\frac{\rho}{\rho^e}\right)_{i+\frac{1}{2},j+\frac{1}{2}} (p^e_x)_{i+\frac{1}{2},j+\frac{1}{2}}. \tag{2.15}$$

Inspired by [33] and in order to preserve the well-balanced property, we split the derivatives in the source term as $\nabla p^e = \nabla(\frac{1}{2}p^e) + \nabla(\frac{1}{2}p^e)$, and apply the same finite difference

operator to approximate them. We have

$$
\begin{aligned}
\boldsymbol{S}^{(2)}_{i+\frac{1}{2},j+\frac{1}{2}} &= \left(\frac{\rho}{\rho^e}\right)_{i+\frac{1}{2},j+\frac{1}{2}} \left[\mathcal{D}_1^+\left(\frac{1}{2}p^e\right) + \mathcal{D}_1^-\left(\frac{1}{2}p^e\right)\right]_{i+\frac{1}{2},j+\frac{1}{2}}, \\
\boldsymbol{S}^{(3)}_{i+\frac{1}{2},j+\frac{1}{2}} &= \left(\frac{\rho}{\rho^e}\right)_{i+\frac{1}{2},j+\frac{1}{2}} \left[\mathcal{D}_2^+\left(\frac{1}{2}p^e\right) + \mathcal{D}_2^-\left(\frac{1}{2}p^e\right)\right]_{i+\frac{1}{2},j+\frac{1}{2}}, \\
\boldsymbol{S}^{(4)}_{i+\frac{1}{2},j+\frac{1}{2}} &= \left(\frac{\rho u}{\rho^e}\right)_{i+\frac{1}{2},j+\frac{1}{2}} \left[\mathcal{D}_1^+\left(\frac{1}{2}p^e\right) + \mathcal{D}_1^-\left(\frac{1}{2}p^e\right)\right]_{i+\frac{1}{2},j+\frac{1}{2}} \\
&\quad + \left(\frac{\rho v}{\rho^e}\right)_{i+\frac{1}{2},j+\frac{1}{2}} \left[\mathcal{D}_2^+\left(\frac{1}{2}p^e\right) + \mathcal{D}_2^-\left(\frac{1}{2}p^e\right)\right]_{i+\frac{1}{2},j+\frac{1}{2}}.
\end{aligned}
\tag{2.16}
$$

Such choice of splitting corresponds to the Lax-Friedrichs flux vector splitting and plays a crucial role in a well-balanced manner. The face-centered source term in (2.10) is written as

$$
\begin{aligned}
\boldsymbol{S}^{(2)}_{i+\frac{1}{2},j} &= \left(\frac{\rho}{\rho^e}\right)_{i+\frac{1}{2},j} \left[\mathcal{D}_1^+\left(\frac{1}{2}p^e\right) + \mathcal{D}_1^-\left(\frac{1}{2}p^e\right)\right]_{i+\frac{1}{2},j}, \\
\boldsymbol{S}^{(3)}_{i+\frac{1}{2},j} &= \left(\frac{\rho}{\rho^e}\right)_{i+\frac{1}{2},j} \left[\mathcal{D}_2\left(p^e\right)\right]_{i+\frac{1}{2},j}, \\
\boldsymbol{S}^{(4)}_{i+\frac{1}{2},j} &= \left(\frac{\rho u}{\rho^e}\right)_{i+\frac{1}{2},j} \left[\mathcal{D}_1^+\left(\frac{1}{2}p^e\right) + \mathcal{D}_1^-\left(\frac{1}{2}p^e\right)\right]_{i+\frac{1}{2},j} + \left(\frac{\rho v}{\rho^e}\right)_{i+\frac{1}{2},j} \left[\mathcal{D}_2\left(p^e\right)\right]_{i+\frac{1}{2},j},
\end{aligned}
\tag{2.17}
$$

and $\boldsymbol{S}_{i,j+\frac{1}{2}}$ in (2.11) can be obtained similarly.

2.3 Time discretization

For hyperbolic conservation law, the semi-discrete scheme is usually paired with high-order strong stability preserving Runge-Kutta (SSP-RK) or multistep time discretization methods [21, 28]. Here, the third-order SSP-RK method is taken for our time discretization,

$$
\begin{aligned}
\boldsymbol{U}^{(1)} &= \boldsymbol{U}^n + \Delta t \mathrm{RHS}(\boldsymbol{U}^n), \\
\boldsymbol{U}^{(2)} &= \frac{3}{4}\boldsymbol{U}^n + \frac{1}{4}(\boldsymbol{U}^{(1)} + \Delta t \mathrm{RHS}(\boldsymbol{U}^{(1)})), \\
\boldsymbol{U}^{n+1} &= \frac{1}{3}\boldsymbol{U}^n + \frac{2}{3}(\boldsymbol{U}^{(2)} + \Delta t \mathrm{RHS}(\boldsymbol{U}^{(2)})),
\end{aligned}
\tag{2.18}
$$

where RHS are the spatial operators given in the right hand side of the schemes (2.3), (2.9)-(2.11).

2.4 Well-balanced property

We present the well-balanced property for the semi-discrete and fully-discrete active flux method in the following theorems.

Theorem 2.1. *The semi-discrete active flux method given by (2.3), (2.9)-(2.11) is well-balanced for the hydrostatic equilibrium state (1.3).*

Proof. Assuming the initial conditions correspond to an exact hydrostatic equilibrium state, which means

$$\rho(x,y)=\rho^e(x,y),\ \boldsymbol{u}(x,y)=0,\ p(x,y)=p^e(x,y),\ E(x,y)=\frac{1}{\gamma-1}p^e(x,y). \tag{2.19}$$

Since $\boldsymbol{u}=0$, the fluxes term reduce to

$$\boldsymbol{F}(\boldsymbol{U})=(0,p,0,0)^T,\quad \boldsymbol{G}(\boldsymbol{U})=(0,0,p,0)^T. \tag{2.20}$$

It is straightforward to verify that for the cell average (2.3), the well-balanced property holds for the first and forth equations, since both the flux and the source term approximations becomes zero in equilibrium. From the steady state (2.19), we have $\bar{\rho}_{ij}=\bar{\rho}^e_{ij}$. Therefore, the first term of (2.8) vanishes in equilibrium, and

$$\overline{\boldsymbol{S}}^{(2)}_{ij}=\frac{\Delta y}{6}\left((p^e_{i+\frac12,j-\frac12}+4p^e_{i+\frac12,j}+p^e_{i+\frac12,j+\frac12})-(p^e_{i-\frac12,j-\frac12}+4p^e_{i-\frac12,j}+p^e_{i-\frac12,j+\frac12})\right). \tag{2.21}$$

By substituting (2.20) and (2.21) into (2.3), we get

$$\begin{aligned}\frac{d}{dt}\overline{\boldsymbol{U}}^{(2)}_{ij}&=-\frac{1}{\Delta x}(\widehat{\boldsymbol{F}}^{(2)}_{i+\frac12,j}-\widehat{\boldsymbol{F}}^{(2)}_{i-\frac12,j})-\frac{1}{\Delta y}(\widehat{\boldsymbol{G}}^{(2)}_{i,j+\frac12}-\widehat{\boldsymbol{G}}^{(2)}_{i,j-\frac12})+\frac{1}{\Delta x\Delta y}\overline{\boldsymbol{S}}^{(2)}_{ij}\\&=-\frac{1}{\Delta x}\frac16\left((p_{i+\frac12,j-\frac12}+4p_{i+\frac12,j}+p_{i+\frac12,j+\frac12})-(p_{i-\frac12,j-\frac12}+4p_{i-\frac12,j}+p_{i-\frac12,j+\frac12})\right)\\&+\frac{1}{\Delta x\Delta y}\frac{\Delta y}{6}\left((p^e_{i+\frac12,j-\frac12}+4p^e_{i+\frac12,j}+p^e_{i+\frac12,j+\frac12})-(p^e_{i-\frac12,j-\frac12}+4p^e_{i-\frac12,j}+p^e_{i-\frac12,j+\frac12})\right)\\&=0.\end{aligned}$$

The last equality follows from the equilibrium state (2.19). Likewise, we can obtain $\frac{d}{dt}\overline{\boldsymbol{U}}^{(3)}_{ij}=0$. Then $\frac{d}{dt}\overline{\boldsymbol{U}}_{ij}=0$ follows.

For the point values, thanks to the steady state (2.19), we have $\widetilde{\boldsymbol{U}}=0$, thus the fluxes become

$$\boldsymbol{F}^{\pm}=\frac12\boldsymbol{F}(\boldsymbol{U})=(0,\frac12p,0,0)^T,\quad \boldsymbol{G}^{\pm}=\frac12\boldsymbol{G}(\boldsymbol{U})=(0,0,\frac12p,0)^T.$$

Analogously, it is trivial that the well-balanced property holds for the mass and energy equation, we mainly focus on the momentum equation. Combined with the fact that $(\rho/\rho^e)_{i+\frac12,j+\frac12}=1$ and the same finite difference operator performed on the source term, we have

$$\frac{d}{dt}\boldsymbol{U}^{(2)}_{i+\frac12,j+\frac12}=-[\mathcal{D}_1^+\boldsymbol{F}^{(2),+}+\mathcal{D}_1^-\boldsymbol{F}^{(2),-}]_{i+\frac12,j+\frac12}-[\mathcal{D}_2^+\boldsymbol{G}^{(2),+}+\mathcal{D}_2^-\boldsymbol{G}^{(2),-}]_{i+\frac12,j+\frac12}+\boldsymbol{S}^{(2)}_{i+\frac12,j+\frac12}$$

$$= -\left[\mathcal{D}_1^+\left(\frac{1}{2}p\right) + \mathcal{D}_1^-\left(\frac{1}{2}p\right)\right]_{i+\frac{1}{2},j+\frac{1}{2}} + \left[\mathcal{D}_1^+\left(\frac{1}{2}p^e\right) + \mathcal{D}_1^-\left(\frac{1}{2}p^e\right)\right]_{i+\frac{1}{2},j+\frac{1}{2}}$$
$$= 0.$$

The third component, as well as the remaining face-centered point values (2.10) and (2.11), can be acquired in a similar manner. This completes the proof of the well-balanced property of our scheme. □

A straightforward induction yields the well-balancedness of the full-discrete scheme, and we omit the detailed proof in this context.

Theorem 2.2. *The fully-discrete active flux method, given by* (2.3), (2.9)-(2.11) *coupled with third-order SSP-RK time discretization, is well-balanced for the hydrostatic equilibrium state* (1.3).

2.5 Positivity-preserving property

In physics, the conservative variables should belong to the set of admissible state, which is defined by

$$\mathcal{G} := \left\{ \boldsymbol{U} = (\rho, \rho\boldsymbol{u}, E)^T |\ \rho > 0,\ p = p(\boldsymbol{U}) = (\gamma - 1)\left(E - \frac{||\rho\boldsymbol{u}||^2}{2\rho}\right) > 0 \right\}. \tag{2.22}$$

This property, however, is not automatically guaranteed by numerical schemes, which may produce negative density or pressure if we do not take special treatment in low density and low pressure regions. It is also known that high-order schemes are prone to numerical oscillations when containing discontinuous solutions. Such numerical instability might cause the breakdown of numerical simulations. Therefore, in addition to achieving well-balanced property, it is of vital importance to preserve the positivity of the density and pressure while effectively controlling spurious oscillations.

For the proposed active flux method, we should keep both the cell average $\overline{\boldsymbol{U}}_{ij} \in \mathcal{G}$ and the point values $\boldsymbol{U}_\sigma \in \mathcal{G}$. To this end, we apply a simplified version of the MOOD paradigm which was used in [1, 5, 14, 15]. The MOOD procedure is divided into three parts. First, the candidate solution is computed by the well-balanced third-order active flux method (2.3), (2.9)-(2.11). Then we check whether the recomputation is needed in each cell according to a set of detection criteria. Finally, if a cell is flagged as a 'trouble' cell, we use a more robust lower-order scheme for this specific cell.

2.5.1 Detection procedure

Let $(\overline{\boldsymbol{U}}_{ij}^n, \boldsymbol{U}_\sigma^n)$ be the solution at time t_n. After each inner stage of Runge-Kutta method, the candidate solution is denoted as $(\overline{\boldsymbol{U}}_{ij}^*, \boldsymbol{U}_\sigma^*)$. The following criteria are applied to detect trouble cells for the cell average,

1. Computer admissible detector: We check if all the variables are numbers, i.e. not NaN or Inf. If not, we flag the element .FALSE. and proceed to the next element. Otherwise, we set the element .TRUE..
2. Physically admissible detector: We check if $\overline{\boldsymbol{U}}_{ij}^* \in \mathcal{G}$. If this holds, we set this element .TRUE. and proceed to check the next criteria, otherwise the element is flagged .FALSE..
3. Locally constant detection: We check whether the solution is not locally constant at time t_n. To be specific, we compare the values of the degrees of freedom in the element I_{ij} with those in the neighboring elements sharing a face. A procedure similar to that in [5] is followed to avoid detecting a fake maximum principle. If the solution is locally constant, no modification of the element is required and the corresponding flag is set to .TRUE..
4. Numerical admissible detection: We check whether a new extremum is produced or not via the discrete maximum principle. Here we test for the density, and possibly the pressure. Let ξ be the function on which we perform the test, and $\mathcal{V}_{ij}$ be the set containing I_{ij} together with its neighboring cells that sharing a face. Following the procedure in [17], if

$$\bar{\xi}_{ij}^* \in [\min_{l\in\mathcal{V}_{ij}} \xi_l^n - \delta, \max_{l\in\mathcal{V}_{ij}} \xi_l^n + \delta], \quad \delta = \max(\delta_0, \epsilon(\max_{l\in\mathcal{V}_{ij}} \xi_l^n - \min_{l\in\mathcal{V}_{ij}} \xi_l^n)),$$

with $\delta_0 = 10^{-4}, \epsilon = 10^{-3}$, we set the element .TRUE.. Otherwise, we employ the technique introduced in [31] to avoid misidentifying smooth extrema as spurious ones, thereby preserving the accuracy of the scheme.

The similar detection procedure is applied to the point values as well.

2.5.2 Recomputation procedure

For any cell in which the candidate numerical high-order solution fails to respect the constraints, we switch to a parachute first-order scheme. Notably, not all first-order schemes possess both the well-balanced and positivity-preserving properties simultaneously. In order not to destroy these properties, the parachute scheme used here is a first-order scheme specially tailored for the Euler equations with gravity.

For the cell average, we employ the piecewise constant structure-preserving DG scheme proposed in [32], where the test function is taken to be one. More details can be found in [32]. Analogous treatments are applied to the point values, which read as

$$\frac{d}{dt}\boldsymbol{U}_{i+\frac{1}{2},j} = -\frac{1}{\Delta x}(\widehat{\boldsymbol{F}}^L_{i+1,j} - \widehat{\boldsymbol{F}}^L_{i,j}) - \frac{1}{\Delta y}(\widehat{\boldsymbol{G}}^L_{i+\frac{1}{2},j+\frac{1}{2}} - \widehat{\boldsymbol{G}}^L_{i+\frac{1}{2},j-\frac{1}{2}}) + \boldsymbol{S}^L_{i+\frac{1}{2},j}, \tag{2.23}$$

$$\frac{d}{dt}\boldsymbol{U}_{i,j+\frac{1}{2}} = -\frac{1}{\Delta x}(\widehat{\boldsymbol{F}}^L_{i+\frac{1}{2},j+\frac{1}{2}} - \widehat{\boldsymbol{F}}^L_{i-\frac{1}{2},j+\frac{1}{2}}) - \frac{1}{\Delta y}(\widehat{\boldsymbol{G}}^L_{i,j+1} - \widehat{\boldsymbol{G}}^L_{i,j}) + \boldsymbol{S}^L_{i,j+\frac{1}{2}}, \tag{2.24}$$

$$\frac{d}{dt}\boldsymbol{U}_{i+\frac{1}{2},j+\frac{1}{2}} = -\frac{1}{\Delta x}(\widehat{\boldsymbol{F}}^L_{i+1,j+\frac{1}{2}} - \widehat{\boldsymbol{F}}^L_{i,j+\frac{1}{2}}) - \frac{1}{\Delta y}(\widehat{\boldsymbol{G}}^L_{i+\frac{1}{2},j+1} - \widehat{\boldsymbol{G}}^L_{i+\frac{1}{2},j}) + \boldsymbol{S}^L_{i+\frac{1}{2},j+\frac{1}{2}}. \tag{2.25}$$

Here the superscript ^{L} means the lower-order numerical fluxes or source term.

$$\widehat{\boldsymbol{F}}^L_{i+1,j} = \boldsymbol{F}^{hllc}\left(\frac{p^{e,*}_{i+1,j}}{p^e_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}, \frac{p^{e,*}_{i+1,j}}{p^e_{i+\frac{3}{2},j}}\boldsymbol{U}_{i+\frac{3}{2},j}\right), \widehat{\boldsymbol{G}}^L_{i+\frac{1}{2},j+\frac{1}{2}} = \boldsymbol{G}^{hllc}\left(\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}}{p^e_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}, \frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}}{p^e_{i+\frac{1}{2},j+1}}\boldsymbol{U}_{i+\frac{1}{2},j+1}\right), \tag{2.26}$$

are the modified HLLC numerical fluxes [9] and

$$p^{e,*}_{i+1,j} = \frac{1}{2}(p^e_{i+\frac{1}{2},j} + p^e_{i+\frac{3}{2},j}), \quad p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}} = \frac{1}{2}(p^e_{i+\frac{1}{2},j} + p^e_{i+\frac{1}{2},j+1}). \tag{2.27}$$

The source term is reformulated in the same manner as in (2.15), and then the first derivative of p^e is approximated by the central difference scheme, which means

$$\begin{aligned}
\boldsymbol{S}^{L,(2)}_{i+\frac{1}{2},j} &= \left(\frac{\rho}{\rho^e}\right)_{i+\frac{1}{2},j}(p^e_x)_{i+\frac{1}{2},j} = \left(\frac{\rho}{\rho^e}\right)_{i+\frac{1}{2},j}\frac{p^e_{i+\frac{3}{2},j} - p^e_{i-\frac{1}{2},j}}{2\Delta x},\\
\boldsymbol{S}^{L,(3)}_{i+\frac{1}{2},j} &= \left(\frac{\rho}{\rho^e}\right)_{i+\frac{1}{2},j}(p^e_y)_{i+\frac{1}{2},j} = \left(\frac{\rho}{\rho^e}\right)_{i+\frac{1}{2},j}\frac{p^e_{i+\frac{1}{2},j+1} - p^e_{i+\frac{1}{2},j-1}}{2\Delta y}.
\end{aligned} \tag{2.28}$$

The numerical fluxes and source terms for the face-centered points in y-direction and those at the corner points can be defined analogously.

The well-balanced and positivity-preserving properties of the cell averages for the first-order scheme have already established in [32]. Some useful properties for the admissible state and HLLC numerical flux are presented in Appendix A. For completeness, the corresponding results for the point values are summarized in the following two theorems.

Theorem 2.3. *The semi-discrete first-order scheme given by (2.23)-(2.25) for the point values is well-balanced for the hydrostatic equilibrium state (1.3).*

Proof. The proof is shown in Appendix B. □

Theorem 2.4. *Let $\boldsymbol{U}^n_\sigma$ be the point values of the first-order scheme (2.23)-(2.25) at time t_n. If $\boldsymbol{U}^n_\sigma \in \mathcal{G}$ for all i, j, then we have $\boldsymbol{U}^{n+1}_\sigma \in \mathcal{G}$ under the CFL-type condition $\Delta t\widehat{\alpha}_\sigma \le 1$, $\widehat{\alpha}_\sigma = \max\{\widehat{\alpha}_{i+\frac{1}{2},j}, \widehat{\alpha}_{i,j+\frac{1}{2}}, \widehat{\alpha}_{i+\frac{1}{2},j+\frac{1}{2}}\}$. Here*

$$\widehat{\alpha}_{i+\frac{1}{2},j} = 2\frac{\widehat{\alpha}_F}{p^e_{i+\frac{1}{2},j}}\left(\frac{p^{e,*}_{i+1,j} + p^{e,*}_{i,j}}{\Delta x} + \frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}} + p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{\Delta y}\right) + \widehat{\alpha}_S,$$

with

$$\widehat{\alpha}_F = \max_{\boldsymbol{U}_{i\pm\frac{1}{2},j}, \boldsymbol{U}_{i+\frac{3}{2},j}, \boldsymbol{U}_{i+\frac{1}{2},j\pm1}} \varrho(\boldsymbol{U}), \widehat{\alpha}_S = \frac{1}{\rho^e_{i+\frac{1}{2},j}\sqrt{2e_{i+\frac{1}{2},j}}}\left(\left|\frac{p^{e,*}_{i+1,j} - p^{e,*}_{i,j}}{\Delta x}\right| + \left|\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}} - p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{\Delta y}\right|\right).$$

ϱ is the spectral radius of $\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{U}}$, $\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{U}}$. The other two variables $\widehat{\alpha}_{i,j+\frac{1}{2}}, \widehat{\alpha}_{i+\frac{1}{2},j+\frac{1}{2}}$ can be defined analogously.

Proof. The proof is shown in Appendix B. □

Remark 2.1. *The cell-centered point value (2.13), which is used to update the point values, is not obtained as a convex combination of admissible states. Consequently, it is possible that* $\boldsymbol{U}_{ij} \notin \mathcal{G}$. *To address this issue, scaling limiter [40] should be applied to* $\boldsymbol{U}_{ij}$ *at the beginning of each inner stage of SSP-RK method. Specifically, we first replace* ρ_{ij} *by*

$$\widehat{\rho}_{ij} = \theta_{ij}^1(\rho_{ij} - \overline{\rho}_{ij}) + \bar{\rho}_{ij}, \quad \theta_{ij}^1 = \min\left\{1, \frac{\bar{\rho}_{ij} - \epsilon_1}{\bar{\rho}_{ij} - \rho_{ij}}\right\}, \quad \epsilon_1 = \min\{10^{-11}, \bar{\rho}_{ij}, \rho_\sigma\}.$$

We denote $\widehat{\boldsymbol{U}}_{ij} = (\widehat{\rho}_{ij}, \rho\boldsymbol{u}_{ij}, E_{ij})^T$, *the final limiting procedure is defined as follows*

$$\boldsymbol{U}_{ij}^{new} = \theta_{ij}^2(\widehat{\boldsymbol{U}}_{ij} - \overline{\boldsymbol{U}}_{ij}) + \overline{\boldsymbol{U}}_{ij}, \quad \theta_{ij}^2 = \min\left\{1, \frac{p(\overline{\boldsymbol{U}}_{ij}) - \epsilon_2}{p(\overline{\boldsymbol{U}}_{ij}) - p(\widehat{\boldsymbol{U}}_{ij})}\right\},$$

with $\epsilon_2 = \min\{10^{-11}, p(\overline{\boldsymbol{U}}_{ij}), p(\boldsymbol{U}_\sigma)\}$.

3 Numerical examples

This section presents the implementation of our proposed structure-preserving active flux method for one- and two-dimensional Euler equations with gravitation. Unless otherwise specified, the CFL number is taken as 0.3 for one-dimensional case, and 0.2 for two-dimensional case. In two dimensions, except for the cell average, only the results at the corner points are reported, as those at the face-centered points exhibit similar behavior and are therefore omitted for brevity. We also compare our performance with that of the non-well-balanced scheme, which employs a straightforward discretization for the source term and the original Lax-Friedrichs flux vector splitting.

3.1 One-dimensional tests

Example 3.1. Test for one-dimensional polytropic equilibrium

To demonstrate the well-balanced property of the proposed structure-preserving active flux method for the polytropic equilibrium, we consider the benchmark example introduced in [25]. The corresponding steady state solution is given by

$$\rho(x) = \left(\rho_0^{\gamma-1} - \frac{1}{K_0}\frac{\gamma-1}{\gamma}\phi(x)\right)^{\frac{1}{\gamma-1}}, \; u(x) = 0, \; p(x) = K_0\rho(x)^\gamma, \tag{3.1}$$

where $\phi(x) = x$ is the linear gravitational potential. The ratio of specific heats is taken as $\gamma = 5/3$. The simulation is performed up to the final time $t = 2$. The L^1 and L^∞ errors of both the cell averages and the point values with different mesh sizes, computed in double precision, are presented in Table 3.1. These results confirm that our proposed method is well balanced, preserving the polytropic steady state up to machine precision.

Table 3.1: Example 3.1: L^1 and L^∞ errors for one-dimensional polytropic hydrostatic equilibrium flow (3.1).

	N	L^1 error			L^∞ error		
		ρ	ρu	E	ρ	ρu	E
average	50	3.73E-16	3.60E-16	3.80E-16	7.77E-16	4.06E-16	6.66E-16
	100	8.37E-16	3.52E-16	5.99E-16	3.33E-15	6.36E-16	1.55E-15
	200	1.75E-15	4.12E-16	6.35E-16	9.88E-15	8.95E-16	1.55E-15
point	50	4.01E-16	3.60E-16	3.77E-16	9.99E-16	4.21E-16	4.44E-16
	100	7.38E-16	3.70E-16	5.71E-16	3.00E-15	6.68E-16	1.44E-15
	200	1.77E-15	4.24E-16	6.11E-16	9.33E-15	8.55E-16	1.33E-15

Example 3.2. Small perturbation of the polytropic equilibrium

In this example, we evaluate the ability of our proposed method to accurately propagate small perturbations near the hydrostatic equilibrium, a property which cannot be realized by a non-well-balanced scheme. Almost the same setup as in Example 3.1, we add a small disturbance to the pressure,

$$(\rho, u, p)(x, 0) = (\rho_{eq}(x), u_{eq}(x), p_{eq}(x) + A\exp(-100(x-1)^2)). \tag{3.2}$$

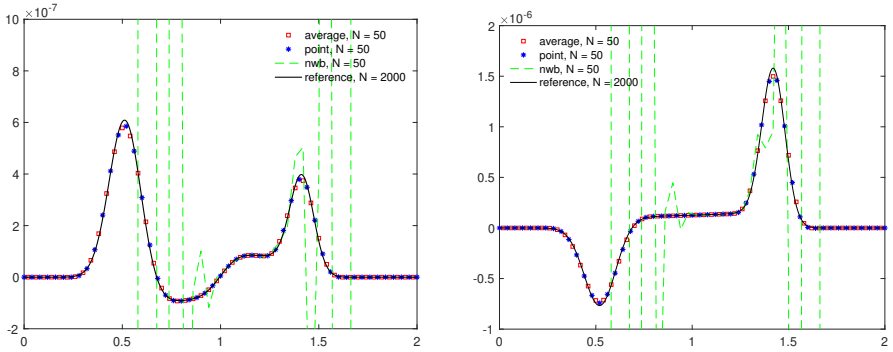
Here $(\rho_{eq}, u_{eq}, p_{eq})$ are the equilibrium functions given in (3.1). We first consider a small amplitude wave propagation $A = 10^{-6}$ and take the stopping time $t = 0.45$. The numerical results for the pressure perturbation and velocity computed on a coarse mesh consisting of 50 uniform cells are displayed in Fig. 3.1. For comparison, the same test is also performed using a non-well-balanced scheme. All results are compared with the reference solutions computed by third-order well-balanced DG scheme [27] with much refined 2000 mesh. As expected, the non-well-balanced scheme fails to resolve the solution accurately, while our proposed structure-preserving active flux method maintains excellent agreement with the reference solutions on a relatively coarse mesh.

To further assess the performance of the proposed method for problems involving discontinuities, we consider the case with a large amplitude $A = 1$. The stopping time is kept the same as before. Fig. 3.2 shows the simulation results with 200 and 2000 uniform cells for comparison. It can be seen that our proposed method accurately captures the propagation of the perturbation while effectively controlling spurious oscillations near the discontinuity.

Example 3.3. Shock tube problem

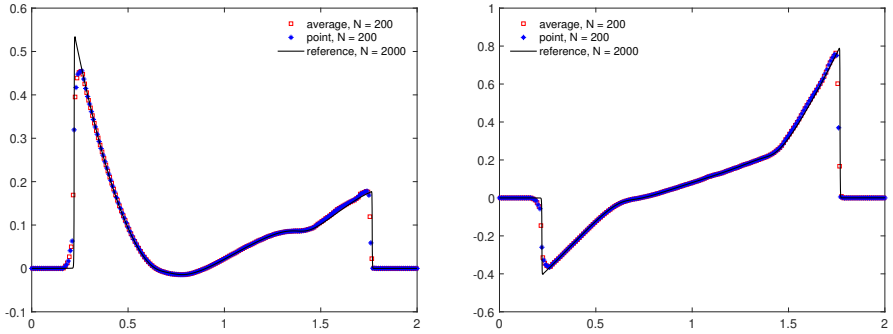
We consider the standard Sod test coupled with a linear gravitational field $\phi(x) = x$. The initial conditions are given by

$$(\rho, u, p)(x, 0) = \begin{cases} (1, 0, 1), & \text{if } x \le 0.5, \\ (0.125, 0, 0.1), & \text{otherwise}, \end{cases} \tag{3.3}$$



(a) pressure perturbation (b) velocity

Fig 3.1: Example 3.2: Pressure perturbation and velocity of the polytropic equilibrium for the small amplitude wave propagation $A = 10^{-6}$.



(a) pressure perturbation (b) velocity

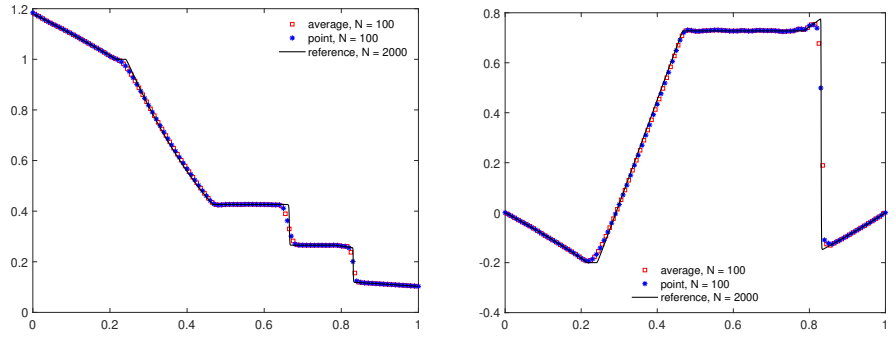
Fig 3.2: Example 3.2: Pressure perturbation and velocity of the hydrostatic atmosphere for the large amplitude wave propagation $A = 1$.

with $\gamma = 1.4$ on the computational domain $[0, 1]$. Solid wall boundary conditions are imposed, and the final time is set to $t = 0.2$. Fig. 3.3 shows the corresponding numerical results for the density, velocity, energy and pressure with a mesh of 100 cells. For comparison, the solutions obtained by the third-order well-balanced DG scheme [27] on 2000 uniform cells are treated as the reference solution. It is evident that our developed method achieves excellent agreement with the reference solution, and provides oscillation-free solutions.

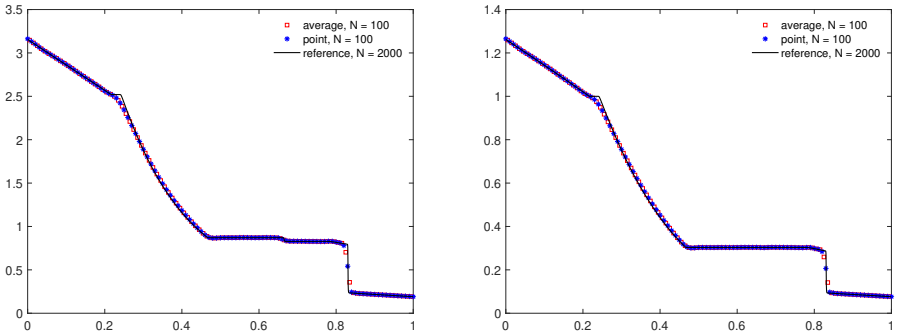
Example 3.4. Rarefaction test with low density and low pressure

To test the positivity-preserving property of our constructed active flux method, we take an extreme rarefaction test on the computational domain $[-1, 1]$. The initial state is the Riemann problem given by [39],

$$\rho(x, 0) = 7, \quad p(x, 0) = 0.2, \quad u(x, 0) = \begin{cases} -1, & x \leq 0, \\ 1, & x > 0, \end{cases} \tag{3.4}$$



(a) density (b) velocity



(c) energy (d) pressure

Fig 3.3: Example 3.3: The numerical density, velocity, energy and pressure for the shock tube problem at time $t = 0.2$.

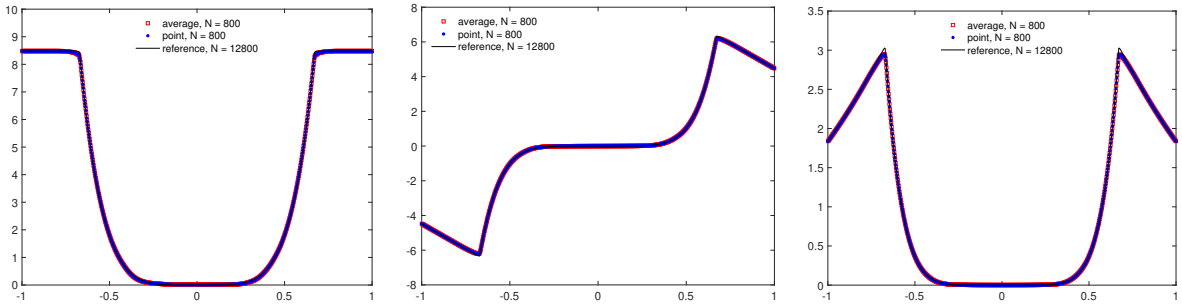
coupled with a quadratic gravitational field $\phi(x) = x^2/2$ centered around $x = 0$. Outflow boundary conditions are imposed at $x = -1$ and $x = 1$. A MOOD paradigm is necessary since the presence of extremely low density and pressure. Without special positivity-preserving treatment, the scheme will blow up numerically in a few time steps. Fig. 3.4 plots the numerical results at time $t = 0.6$, calculated by our method with 800 and 12800 uniform cells for comparison. It is clear that the structure of the low density and low pressure are well captured with the MOOD paradigm, which indicates the effectiveness of our proposed structure-preserving active flux method.

Example 3.5. Leblanc problem in linear gravitational field

We consider another highly challenging test problem, which is an extension of one-dimensional Leblanc shock tube problem with linear gravitational potential $\phi(x) = x$. The initial conditions are given by

$$(\rho, u, p)(x, 0) = \begin{cases} (2, 0, 10^9), & x \leq 5, \\ (10^{-3}, 0, 1), & x > 5. \end{cases} \tag{3.5}$$

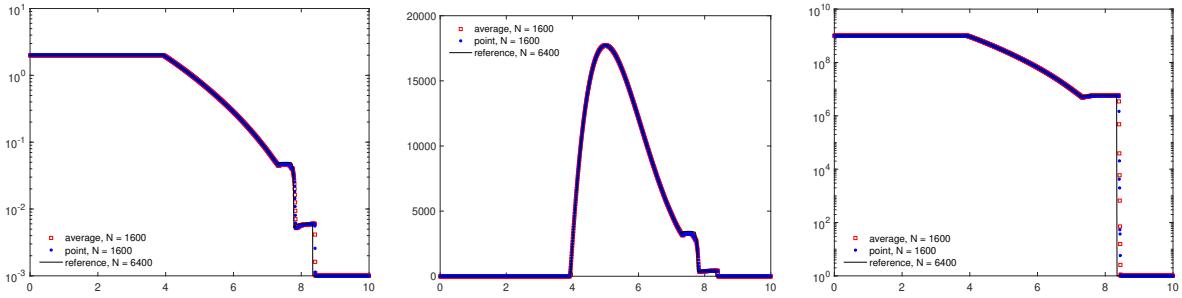
We can see that the density and pressure involve strong jumps initially. Solid wall boundary conditions are imposed at the end points $x = 0$ and $x = 10$. A sufficiently fine mesh is required to fully resolve the wave structure in this example. To this end,



(a) density (b) momentum (c) energy

Fig 3.4: Example 3.4: The numerical density, momentum and energy for the rarefaction test at time $t = 0.6$.

we treat the solution obtained on a uniform mesh of 6400 cells as the reference solution. The numerical results at time $t = 0.00004$ with 1600 uniform cells are also presented in Fig. 3.5. We can observe that the discontinuity are well captured without introducing spurious oscillations, while exhibiting excellent resolution of the wave structures.



(a) density (b) momentum (c) pressure

Fig 3.5: Example 3.5: The log plot of density, momentum and the log plot of pressure for the leblanc problem at time $t = 0.00004$.

3.2 Two-dimensional tests

Example 3.6. Accuracy test

To verify the high-order accuracy of our structure-preserving active flux method for a smooth solution in two-dimensional Euler equations, we consider the following exact solutions on the computational domain $[0, 2] \times [0, 2]$,

$$\begin{aligned} \rho(x, y, t) &= 1 + 0.2 \sin(\pi(x + y - t(u_0 + v_0))), \\ u(x, y, t) &= u_0, \\ v(x, y, t) &= v_0, \\ p(x, y, t) &= p_0 + t(u_0 + v_0) - x - y + 0.2 \cos(\pi(x + y - t(u_0 + v_0)))/\pi, \end{aligned} \tag{3.6}$$

where the parameters $u_0 = v_0 = 1, p_0 = 4.5$ and $\gamma = 1.4$. The gravity is given by a linear gravitational potential $\phi_x = \phi_y = 1$. The boundary condition is specified by the exact solution. The final time is set as $t = 0.1$. The L^1 errors of the cell average and point values, along with their corresponding convergence rates are presented in Tables 3.2 and 3.3, respectively. It is observed that our proposed structure-preserving active flux method attains third-order accuracy.

Table 3.2: Example 3.6: L^1 errors of the cell average and numerical orders of accuracy with initial condition (3.6), $t = 0.1$.

$Nx \times Ny$	ρ		ρu		ρv		E	
	L^1 error	order	L^1 error	order	L^1 error	order	L^1 error	order
10×10	2.68E-04	–	2.95E-04	–	2.95E-04	–	9.32E-04	–
20×20	2.73E-05	3.30	3.37E-05	3.13	3.37E-05	3.13	1.34E-04	2.79
40×40	3.47E-06	2.98	4.46E-06	2.92	4.46E-06	2.92	1.80E-05	2.90
80×80	4.52E-07	2.94	5.81E-07	2.94	5.81E-07	2.94	2.33E-06	2.95
160×160	5.80E-08	2.96	7.45E-08	2.96	7.45E-08	2.96	2.95E-07	2.98

Table 3.3: Example 3.6: L^1 errors of the point values and numerical orders of accuracy with initial condition (3.6), $t = 0.1$.

$Nx \times Ny$	ρ		ρu		ρv		E	
	L^1 error	order	L^1 error	order	L^1 error	order	L^1 error	order
10×10	1.48E-03	–	1.42E-03	–	1.42E-03	–	2.38E-03	–
20×20	1.32E-04	3.48	1.32E-04	3.43	1.32E-04	3.43	2.67E-04	3.16
40×40	1.49E-05	3.15	1.52E-05	3.11	1.52E-05	3.11	3.26E-05	3.03
80×80	1.82E-06	3.03	1.89E-06	3.01	1.89E-06	3.01	4.07E-06	3.00
160×160	2.27E-07	3.00	2.37E-07	2.99	2.37E-07	2.99	5.09E-07	3.00

Example 3.7. Test for two-dimensional isothermal equilibrium

To illustrate the capability of our developed method for the preservation and perturbation of the isothermal hydrostatic equilibrium state in two dimensions, we test the same example as in [26], where the isothermal equilibrium state is given by

$$\begin{aligned} \rho(x,y) &= \rho_0 \exp\left(-\frac{\rho_0 g}{p_0}(x+y)\right), \\ u(x,y) &= v(x,y) = 0, \\ p(x,y) &= p_0 \exp\left(-\frac{\rho_0 g}{p_0}(x+y)\right), \end{aligned} \tag{3.7}$$

where the parameters $\rho_0 = 1.21, p_0 = 1, g = 1$. The linear gravitation potential $\phi(x, y) = x + y$ is considered on the computational domain $[0, 1] \times [0, 1]$. The ratio of specific heats is $\gamma = 1.4$. We simulate the solutions up to $t = 0.5$ with different mesh sizes and present the L^1 and L^∞ errors of the cell average and point values at double precision in Table 3.4. It can be seen clearly that the isothermal equilibrium state is exactly preserved even on coarse meshes in two-dimensional case.

Table 3.4: Example 3.7: L^1 and L^∞ errors for two-dimensional isothermal hydrostatic equilibrium flow (3.7).

	$Nx \times Ny$	error	ρ	ρu	ρv	E
average	50×50	L^1	1.98E-16	2.58E-16	2.38E-16	7.49E-16
		L^∞	3.77E-15	6.23E-15	5.92E-15	1.42E-14
	100×100	L^1	2.61E-16	1.19E-16	1.31E-16	3.98E-16
		L^∞	7.99E-15	2.21E-15	2.14E-15	6.22E-15
point	50×50	L^1	2.17E-16	2.86E-16	2.67E-16	8.09E-16
		L^∞	4.22E-15	6.94E-15	6.41E-15	1.42E-14
	100×100	L^1	2.61E-16	1.27E-16	1.39E-16	4.13E-16
		L^∞	5.22E-15	2.48E-15	2.36E-15	6.22E-15

Subsequently, a small perturbation is added to the initial pressure profile

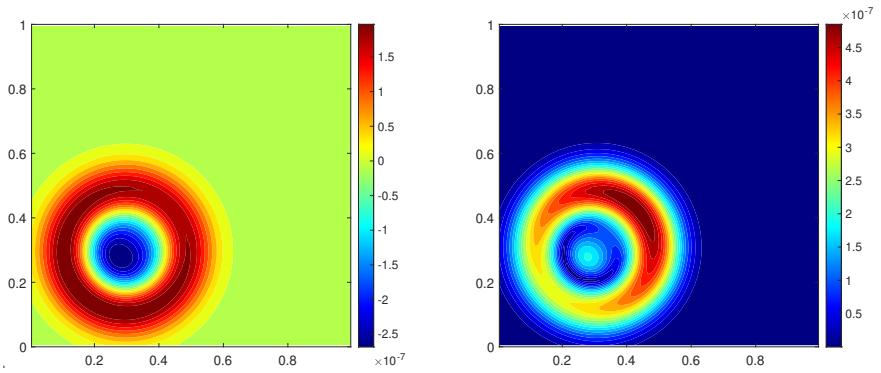
$$p(x, y, 0) = p(x, y) + A \exp\left(-\frac{100\rho_0}{p_0}((x - 0.3)^2 + (y - 0.3)^2)\right), \tag{3.8}$$

with the density and velocities remain unchanged. We consider the small amplitude $A = 10^{-6}$ and impose transmission boundary conditions. We simulate the solution up to $t = 0.15$ on a uniform grid of 100×100 cells, and display the cell average contours of the pressure perturbation and velocity in Fig. 3.6. For comparison, the corresponding results obtained by the non-well-balanced method are also illustrated. The results indicate that our proposed well-balanced scheme outperforms the non-well-balanced counterpart. It can achieve high resolution and is capable of capturing the small features of the disturbance accurately.

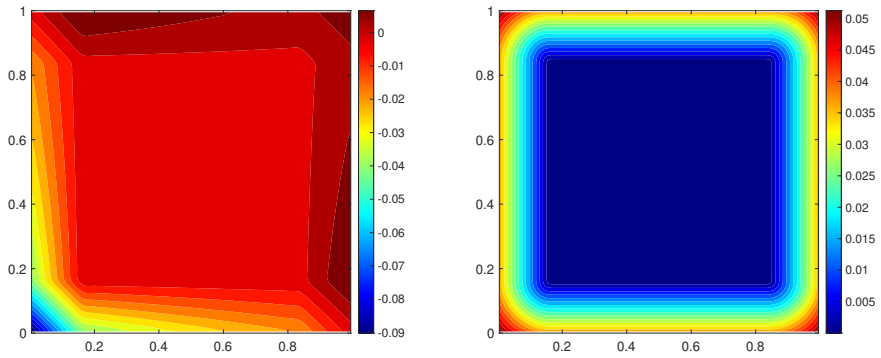
Example 3.8. Test for two-dimensional polytropic equilibrium

We use this example to assess the performance of our developed scheme near the two-dimensional polytropic equilibrium. An astrophysical problem studied in [23] is considered on the computational domain $[-0.5, 0.5] \times [-0.5, 0.5]$. The analytical hydrostatic equilibrium and the gravitation potential are given by

$$\begin{aligned} &\rho(r) = \rho_c \frac{\sin(\alpha r)}{\alpha r}, \ u(r) = v(r) = 0, \ p(r) = \rho(r)^2, \\ &\phi(r) = -2\rho_c \frac{\sin(\alpha r)}{\alpha r}, \end{aligned} \tag{3.9}$$



(a) pressure perturbation (b) velocity



(c) pressure perturbation (d) velocity

Fig 3.6: Example 3.7: The cell average contours of the pressure perturbation and velocity $\sqrt{u^2+v^2}$ for the small amplitude $A = 10^{-6}$ in (3.8), with 20 uniformly spaced contour lines at time $t = 0.15$. Top: results by our well-balanced active flux method. Bottom: results by the non-well-balanced method.

where $r = \sqrt{x^2+y^2}$ is the radial variable, the parameters are taken as $\alpha = \sqrt{2\pi}$, $\rho_c = 1$ and $\gamma = 2$. To demonstrate that the steady state is indeed preserved up to machine error, we calculate the L^1 and L^∞ errors at double precision up to $t = 2$ with different mesh sizes. The corresponding results both for the cell average and point values are listed in Table 3.5, from which the desired well-balanced property can be observed.

We keep the density and velocity unchanged, and then add a small Gaussian hump perturbation to the initial pressure profile

$$p(x, y, 0) = \rho(r)^2 + A\exp(-100r^2), \tag{3.10}$$

with a small amplitude $A = 10^{-6}$. We impose transmission boundary conditions and simulate the solution to the final time $t = 0.2$ before the excited waves reach the boundary. The contours of the pressure perturbation and velocity, computed using our proposed well-balanced active flux method and a non-well-balanced method, are shown in Fig. 3.7 for comparison. The non-well-balanced method yields an inaccurate solution, while our proposed method performs robustly and maintains axial symmetry. This underscores

Table 3.5: Example 3.8: L^1 and L^∞ errors for two-dimensional polytropic hydrostatic equilibrium flow (3.9).

	$Nx \times Ny$	error	ρ	ρu	ρv	E
average	50×50	L^1	9.89E-17	1.59E-16	1.60E-16	6.86E-17
		L^∞	1.44E-15	1.05E-15	1.36E-15	5.00E-16
	100×100	L^1	1.53E-16	2.27E-16	2.27E-16	7.21E-17
		L^∞	3.00E-15	1.45E-15	1.49E-15	6.66E-16
point	50×50	L^1	1.38E-16	1.79E-16	1.80E-16	8.23E-17
		L^∞	8.88E-16	1.12E-15	1.17E-15	4.44E-16
	100×100	L^1	1.79E-16	2.45E-16	2.45E-16	8.18E-17
		L^∞	2.11E-15	1.68E-15	1.75E-15	6.66E-16

the importance of the well-balanced scheme for accurately capturing small perturbations near the steady state.

Example 3.9. Two-dimensional blast problem

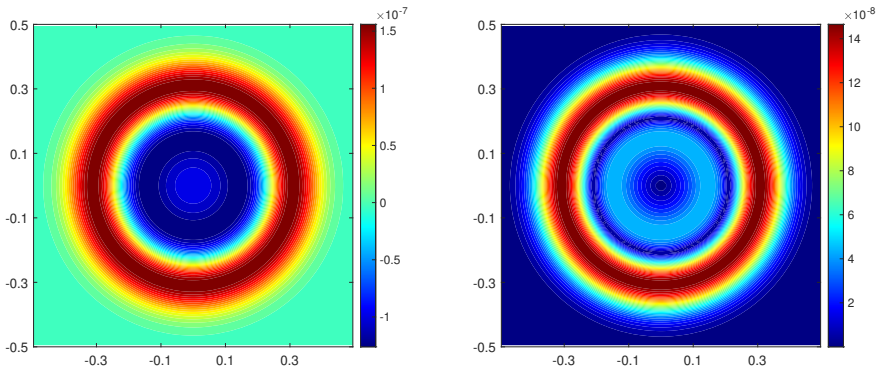
To further illustrate the positivity-preserving property of our scheme, we test a two-dimensional blast problem, which is introduced in [32]. The gravitational field, density, and velocity are chosen to be the same as those in (3.9), we only introduce a huge jump in the initial pressure, that is

$$p(x, y, 0) = \rho(r)^2 + \begin{cases} 100, & r < 0.1, \\ 0, & r \geq 0.1. \end{cases} \tag{3.11}$$

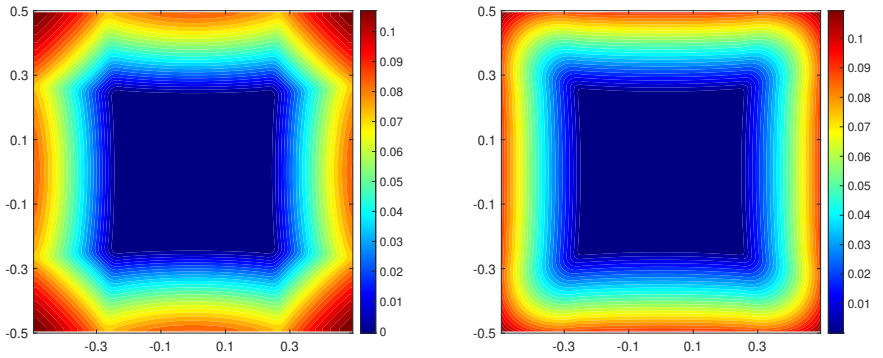
The same set of parameters as in Example 3.8 is taken, except that ρ_c is changed to 0.01, yielding the appearance of low density and low pressure in the solution. Transmissive boundary conditions are considered here. The computation is carried out on a 400×400 mesh up to the final time $t = 0.0005$, and we show the contours of numerical density and pressure logarithm in Fig. 3.8. We also display the pressure profiles along the line $y = x$. It shows that the solution develops strong discontinuities, with a shock located at $\sqrt{x^2 + y^2}/\sqrt{2} \approx 0.28$. With the MOOD paradigm, our method is successful in maintaining the positivity-preserving property and suppressing numerical oscillations at the same time.

Example 3.10. Inertia-gravity wave

The last example we take is given by [20], which is a benchmark problem arising from atmospheric flows. We consider a two-dimensional channel defined on the domain $[0, 300000] \times [0, 10000]$ m^2, with periodic boundary conditions on the left and right boundaries, and solid wall boundary conditions on the bottom and top boundaries. The flow evolves under a linear gravitational field with $\phi_x = 0$ and $\phi_y = g = 9.8$ m/s^2. The

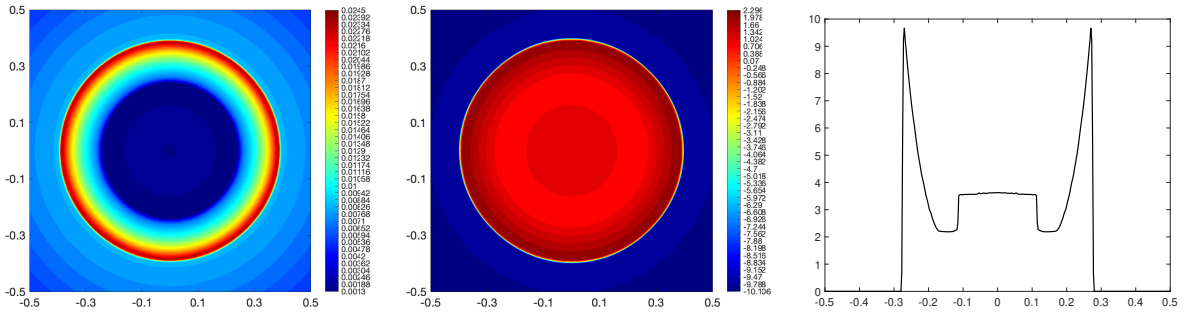


(a) pressure perturbation (b) velocity



(c) pressure perturbation (d) velocity

Fig 3.7: Example 3.8: The cell average contours of the pressure perturbation and velocity $\sqrt{u^2+v^2}$ for the small amplitude $A = 10^{-6}$ in (3.10), with 30 uniformly spaced contour lines at time $t = 0.2$. Top: results by our well-balanced active flux method. Bottom: results by the non-well-balanced method.



(a) density (b) pressure logarithm (c) pressure along $y = x$

Fig 3.8: Example 3.9: The point values contours of the density, the pressure logarithm, and the plot of pressure along $y = x$ at time $t = 0.0005$.

initial state consists of a uniformly stratified atmosphere with a constant velocity field

$\boldsymbol{u} = (20\ \text{m/s}, 0\ \text{m/s})$. The potential temperature and Exner pressure are specified by

$$\Theta = T_0 \exp\left(\frac{\mathcal{N}^2}{g} y\right), \quad \Pi = 1 + \frac{(\gamma-1)g^2}{\gamma R T_0 \mathcal{N}^2}\left[\exp\left(-\frac{\mathcal{N}^2}{g} y\right) - 1\right]. \tag{3.12}$$

where the Brunt–Väisälä frequency $\mathcal{N} = 0.01/\text{s}$, the gas constant $R = 287.058$ J/kg K, and the reference temperature $T_0 = 300$ K at $y = 0$ m. A small perturbation is introduced into the initial potential temperature,

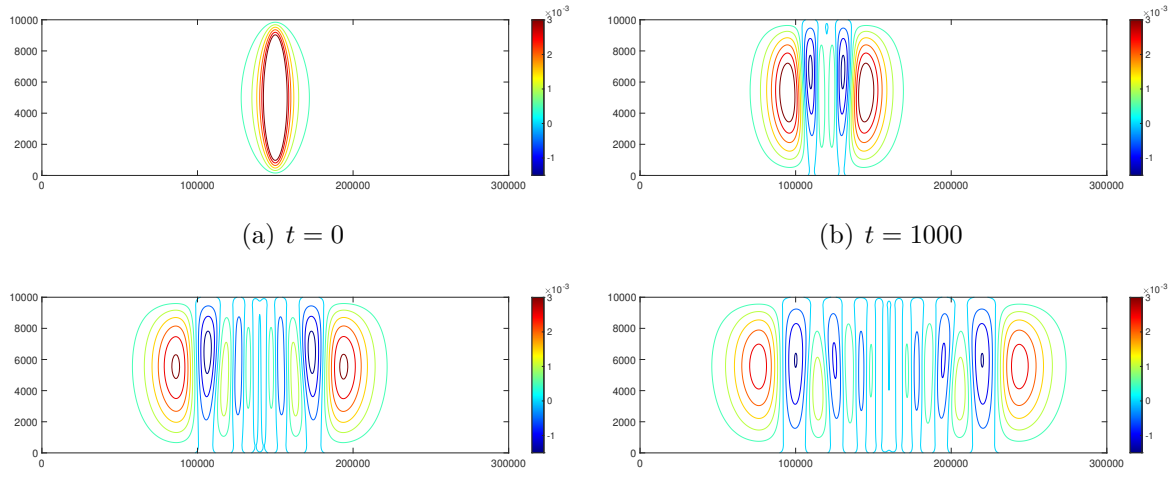
$$\Delta\Theta(x, y, 0) = \theta_c \sin\left(\frac{\pi y}{h_c}\right)\left[1 + (x - x_c)^2/a_c^2\right]^{-1},$$

with $\theta_c = 0.01$ K, $h_c = 10000$ m, $x_c = 10000$ m, $a_c = 5000$ m. The pressure and density are obtained from Θ and Π through

$$p = p_0 \Pi^{\frac{\gamma}{\gamma-1}}, \quad \rho = \frac{p_0}{R\Theta}\Pi^{\frac{1}{\gamma-1}},$$

where the reference pressure is taken as $p_0 = 10^5$ N/m^2 at $y = 0$ m.

We run the simulation until $t = 3000$ s with 1200×40 mesh cells, seeing the contours of the potential temperature perturbation $\Delta\Theta$ at different times $t = 0, 1000, 2000, 3000$ s in Fig. 3.9, and the profiles of $\Delta\Theta$ at $t = 3000$ s along the line $y = 5000$ m in Fig. 3.10. We can observe a well-resolved evolution of the potential temperature perturbation, which demonstrates excellent agreement with the reference structures in [20, 32].



(a) $t = 0$ (b) $t = 1000$

(c) $t = 2000$ (d) $t = 3000$

Fig 3.9: Example 3.10: The point values contours of the potential temperature perturbation $\Delta\Theta$ at different times $t = 0, 1000, 2000, 3000$ s. 10 uniformly spaced contour lines from -0.0015 to 0.003.

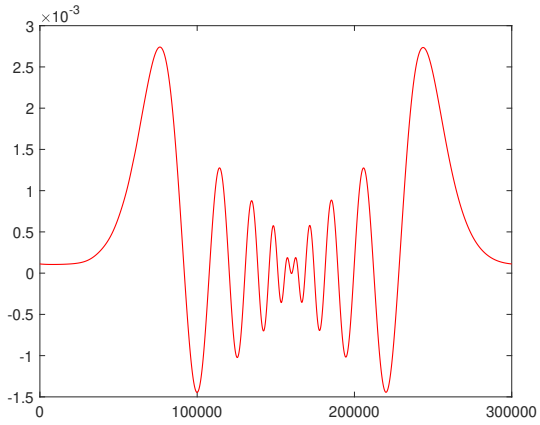


Fig 3.10: Example 3.10: The potential temperature perturbation $\Delta\Theta$ at time $t = 3000$ s along the line $y = 5000$ m.

4 Conclusion

In this work, the high-order structure-preserving active flux method was constructed for the compressible Euler equations under gravitational fields. Accounting for the well-balanced property, the spatial discretization for both the cell average and the point values was specially designed, with the source term reformulated based on the target hydrostatic steady state solutions and a properly modified Lax-Friedrichs flux vector splitting for the flux term. Positivity-preserving property was achieved by applying a posteriori MOOD paradigm equipped with a structure-preserving first-order scheme. Theoretical analysis and numerical experiments have demonstrated our schemes' high-order accuracy, positivity preservation and robustness in capturing complex small features at or near the equilibrium flow.

A Useful properties

Some important properties for the admissible state and the one-dimensional HLLC numerical flux are listed here, two-dimensional case is similar. All the lemmas can be found in [32] as well.

Lemma A.1 ([32]). *For any constant $\lambda \geq 0$, $\delta \in \mathbb{R}$, $\boldsymbol{U} = (\rho, \rho\boldsymbol{u}, E)^T \in \mathcal{G}$, and $\boldsymbol{a} \in \mathbb{R}^d$, if $\delta \frac{||a||}{\sqrt{2e}} \leq \lambda$, then $\lambda\boldsymbol{U} + \delta(0, \rho\boldsymbol{a}, \rho\boldsymbol{u} \cdot \boldsymbol{a})^T \in \overline{\mathcal{G}}$.*

Lemma A.2 ([32]). *Let $\boldsymbol{H}(\boldsymbol{U}) = [\boldsymbol{F}(\boldsymbol{U}), \boldsymbol{G}(\boldsymbol{U})]$ be the flux functions. For any $\boldsymbol{U} \in \mathcal{G}$ and unit vector $\boldsymbol{n} \in \mathbb{R}^d$, we have $\boldsymbol{U} - \lambda\boldsymbol{H}(\boldsymbol{U}) \cdot \boldsymbol{n} \in \mathcal{G}$, for any $\lambda \in \mathbb{R}$ satisfying $|\lambda|\varrho(\boldsymbol{U}) \leq 1$.*

Lemma A.3 (Contact property [32]). *For any two states $\boldsymbol{U}_l = (\rho_l, 0, p/(\gamma-1))^T$, and $\boldsymbol{U}_r = (\rho_r, 0, p/(\gamma-1))^T$, HLLC flux satisfies $\boldsymbol{F}^{hllc}(\boldsymbol{U}_l, \boldsymbol{U}_r) = (0, p, 0)^T$.*

Lemma A.4 ([32]). *For any parameters $\zeta_1, \zeta_2, \zeta_3, \zeta_4 \geq 0$ and any admissible states $\boldsymbol{U}_l$, $\boldsymbol{U}_r \in \mathcal{G}$, if $\lambda > 0$ satisfies $\lambda \max_{\boldsymbol{U}\in\{\boldsymbol{U}_l,\boldsymbol{U}_r\}} \varrho(\boldsymbol{U}) \leq \frac{1}{2}$, then we have*

$$\zeta_2 \boldsymbol{U}_r - \lambda(\boldsymbol{F}(\zeta_2 \boldsymbol{U}_r) - \boldsymbol{F}^{hllc}(\zeta_1 \boldsymbol{U}_l, \zeta_2 \boldsymbol{U}_r)) \in \mathcal{G},$$
$$\zeta_3 \boldsymbol{U}_l - \lambda(\boldsymbol{F}^{hllc}(\zeta_3 \boldsymbol{U}_l, \zeta_4 \boldsymbol{U}_r) - \boldsymbol{F}(\zeta_3 \boldsymbol{U}_l)) \in \mathcal{G}.$$

B Proof of Theorem 2.3 and 2.4

Proof of the well-balanced property 2.3. For brevity, we prove only (2.23) here, as the proofs for (2.24) and (2.25) proceed analogously.

At the equilibrium state, (2.19) is satisfied, and we have

$$\frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{1}{2},j}} \boldsymbol{U}_{i+\frac{1}{2},j} = \left(\frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{1}{2},j}} \rho_{i+\frac{1}{2},j}, 0, 0, \frac{p^{e,*}_{i+1,j}}{\gamma - 1} \right)^T, \quad \frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{3}{2},j}} \boldsymbol{U}_{i+\frac{3}{2},j} = \left(\frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{3}{2},j}} \rho_{i+\frac{3}{2},j}, 0, 0, \frac{p^{e,*}_{i+1,j}}{\gamma - 1} \right)^T.$$

According to the contract property Lemma A.3, the modified HLLC flux becomes $\widehat{\boldsymbol{F}}^L_{i+1,j} = (0, p^{e,*}_{i+1,j}, 0, 0)^T$. Similarly, we have $\widehat{\boldsymbol{G}}^L_{i+\frac{1}{2},j+\frac{1}{2}} = (0, 0, p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}, 0)^T$. Owing to the fact $(\rho/\rho^e)_{i+\frac{1}{2},j} = 1$ in equilibrium, and the central difference approximation for the source term, we have

$$\frac{1}{2}(p^{e}_{i+\frac{3}{2},j} - p^{e}_{i-\frac{1}{2},j}) = p^{e,*}_{i+1,j} - p^{e,*}_{i,j}, \quad \frac{1}{2}(p^{e}_{i+\frac{1}{2},j+1} - p^{e}_{i+\frac{1}{2},j-1}) = p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}} - p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}.$$

Thus the momentum equation in (2.23) equals

$$\frac{d}{dt}\boldsymbol{U}^{(2)}_{i+\frac{1}{2},j} = -\frac{1}{\Delta x}(p^{e,*}_{i+1,j} - p^{e,*}_{i,j}) + \frac{1}{2\Delta x}(p^{e}_{i+\frac{3}{2},j} - p^{e}_{i-\frac{1}{2},j}) \overset{(2.27)}{=} 0,$$
$$\frac{d}{dt}\boldsymbol{U}^{(3)}_{i+\frac{1}{2},j} = -\frac{1}{\Delta y}(p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}} - p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}) + \frac{1}{2\Delta y}(p^{e}_{i+\frac{1}{2},j+1} - p^{e}_{i+\frac{1}{2},j-1}) \overset{(2.27)}{=} 0.$$

Moreover, the well-balanced property for the mass and energy equations in (2.23) follows directly, since the numerical flux and source term both reduce to zero. This completes the proof of Theorem 2.3. □

Proof of the positivity-preserving property 2.4. In fact, the proof follows essentially the same strategy as that in [32]. The scheme (2.23) coupled with Euler forward time discretization yields

$$\begin{aligned}
\boldsymbol{U}^{n+1}_{i+\frac{1}{2},j} &= \boldsymbol{U}_{i+\frac{1}{2},j} - \frac{\Delta t}{\Delta x}(\widehat{\boldsymbol{F}}^L_{i+1,j} - \widehat{\boldsymbol{F}}^L_{i,j}) - \frac{\Delta t}{\Delta y}(\widehat{\boldsymbol{G}}^L_{i+\frac{1}{2},j+\frac{1}{2}} - \widehat{\boldsymbol{G}}^L_{i+\frac{1}{2},j-\frac{1}{2}}) + \Delta t \boldsymbol{S}_{i+\frac{1}{2},j} \\
&= \boldsymbol{U}_{i+\frac{1}{2},j} - \frac{\Delta t}{\Delta x}\frac{p^{e,*}_{i+1,j} - p^{e,*}_{i,j}}{p^{e}_{i+\frac{1}{2},j}} \boldsymbol{F}(\boldsymbol{U}_{i+\frac{1}{2},j}) - \frac{\Delta t}{\Delta y}\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}} - p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}} \boldsymbol{G}(\boldsymbol{U}_{i+\frac{1}{2},j}) + \Delta t \boldsymbol{S}_{i+\frac{1}{2},j}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\Delta t}{\Delta x}\left[\boldsymbol{F}\left(\frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)-\boldsymbol{F}^{hllc}\left(\frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j},\frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{3}{2},j}}\boldsymbol{U}_{i+\frac{3}{2},j}\right)\right]\\
&-\frac{\Delta t}{\Delta x}\left[\boldsymbol{F}\left(\frac{p^{e,*}_{i,j}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)-\boldsymbol{F}^{hllc}\left(\frac{p^{e,*}_{i,j}}{p^{e}_{i-\frac{1}{2},j}}\boldsymbol{U}_{i-\frac{1}{2},j},\frac{p^{e,*}_{i,j}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)\right]\\
&+\frac{\Delta t}{\Delta y}\left[\boldsymbol{G}\left(\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)-\boldsymbol{G}^{hllc}\left(\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j},\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j+1}}\boldsymbol{U}_{i+\frac{1}{2},j+1}\right)\right]\\
&-\frac{\Delta t}{\Delta y}\left[\boldsymbol{G}\left(\frac{p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)-\boldsymbol{G}^{hllc}\left(\frac{p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j-1}}\boldsymbol{U}_{i+\frac{1}{2},j-1},\frac{p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)\right].
\end{aligned}
$$

Here the superscript ^{n} is omitted from the right hand side. Subsequently, we split $\boldsymbol{U}^{n+1}_{i+\frac{1}{2},j}$ into four parts, i.e. $\boldsymbol{U}^{n+1}_{i+\frac{1}{2},j}:=\boldsymbol{W}_1+\boldsymbol{W}_2+\boldsymbol{W}_3+\boldsymbol{W}_4$, with

$$
\begin{aligned}
\boldsymbol{W}_1&=\left[1-\Delta t\left(2\frac{\widehat{\alpha}_F}{p^{e}_{i+\frac{1}{2},j}}\left(\frac{p^{e,*}_{i+1,j}+p^{e,*}_{i,j}}{\Delta x}+\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}+p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{\Delta y}\right)+\widehat{\alpha}_S\right)\right]\boldsymbol{U}_{i+\frac{1}{2},j},\\
\boldsymbol{W}_2&=\Delta t\left[\frac{\widehat{\alpha}_F}{\Delta x}\frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{1}{2},j}}\left(\boldsymbol{U}_{i+\frac{1}{2},j}-\frac{1}{\widehat{\alpha}_F}\boldsymbol{F}(\boldsymbol{U}_{i+\frac{1}{2},j})\right)+\frac{\widehat{\alpha}_F}{\Delta x}\frac{p^{e,*}_{i,j}}{p^{e}_{i+\frac{1}{2},j}}\left(\boldsymbol{U}_{i+\frac{1}{2},j}+\frac{1}{\widehat{\alpha}_F}\boldsymbol{F}(\boldsymbol{U}_{i+\frac{1}{2},j})\right)\right.\\
&\quad\left.+\frac{\widehat{\alpha}_F}{\Delta y}\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\left(\boldsymbol{U}_{i+\frac{1}{2},j}-\frac{1}{\widehat{\alpha}_F}\boldsymbol{G}(\boldsymbol{U}_{i+\frac{1}{2},j})\right)+\frac{\widehat{\alpha}_F}{\Delta y}\frac{p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\left(\boldsymbol{U}_{i+\frac{1}{2},j}+\frac{1}{\widehat{\alpha}_F}\boldsymbol{G}(\boldsymbol{U}_{i+\frac{1}{2},j})\right)\right],\\
\boldsymbol{W}_3&=\Delta t\left(\widehat{\alpha}_S\boldsymbol{U}_{i+\frac{1}{2},j}+\boldsymbol{S}_{i+\frac{1}{2},j}\right),\\
\boldsymbol{W}_4&=\Delta t\left\{\frac{\widehat{\alpha}_F}{\Delta x}\left[\frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}-\frac{1}{\widehat{\alpha}_F}\left[\boldsymbol{F}^{hllc}\left(\frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j},\frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{3}{2},j}}\boldsymbol{U}_{i+\frac{3}{2},j}\right)-\boldsymbol{F}\left(\frac{p^{e,*}_{i+1,j}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)\right]\right]\right.\\
&\quad+\frac{\widehat{\alpha}_F}{\Delta x}\left[\frac{p^{e,*}_{i,j}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}-\frac{1}{\widehat{\alpha}_F}\left[\boldsymbol{F}\left(\frac{p^{e,*}_{i,j}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)-\boldsymbol{F}^{hllc}\left(\frac{p^{e,*}_{i,j}}{p^{e}_{i-\frac{1}{2},j}}\boldsymbol{U}_{i-\frac{1}{2},j},\frac{p^{e,*}_{i,j}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)\right]\right]\\
&\quad+\frac{\widehat{\alpha}_F}{\Delta y}\left[\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}-\frac{1}{\widehat{\alpha}_F}\left[\boldsymbol{G}^{hllc}\left(\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j},\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j+1}}\boldsymbol{U}_{i+\frac{1}{2},j+1}\right)-\boldsymbol{G}\left(\frac{p^{e,*}_{i+\frac{1}{2},j+\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)\right]\right]\\
&\quad\left.+\frac{\widehat{\alpha}_F}{\Delta y}\left[\frac{p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}-\frac{1}{\widehat{\alpha}_F}\left[\boldsymbol{G}\left(\frac{p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)-\boldsymbol{G}^{hllc}\left(\frac{p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j-1}}\boldsymbol{U}_{i+\frac{1}{2},j-1},\frac{p^{e,*}_{i+\frac{1}{2},j-\frac{1}{2}}}{p^{e}_{i+\frac{1}{2},j}}\boldsymbol{U}_{i+\frac{1}{2},j}\right)\right]\right]\right\}.
\end{aligned}
$$

Using the Lemma A.1, A.2, A.4 given in Appendix A and following the arguments in [32], we can establish that $\boldsymbol{W}_i\in\mathcal{G}, i=1,\cdots,4$. Here we only prove the positivity-preserving property for (2.23), (2.24) and (2.25) can be treated in a similar manner. □

References

[1] 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):370–402, 2023.

[2] R. Abgrall and W. Barsukow. A hybrid finite element–finite volume method for conservation laws. *Appl. Math. Comput.*, 447:127846, 2023.

[3] R. Abgrall and W. Barsukow. Extensions of active flux to arbitrary order of accuracy. *ESAIM: Math. Model. Numer. Anal.*, 57(2):991–1027, 2023.

[4] R. Abgrall, W. Barsukow, and C. Klingenberg. A semi-discrete active flux method for the Euler equations on Cartesian grids. *J. Sci. Comput.*, 102(2):36, 2025.

[5] R. Abgrall and Y. Liu. A new approach for designing well-balanced schemes for the shallow water equations: a combination of conservative and primitive formulations. *SIAM J. Sci. Comput.*, 46(6):A3375–A3400, 2024.

[6] W. Barsukow. The active flux scheme for nonlinear problems. *J. Sci. Comput.*, 86(1):3, 2021.

[7] W. Barsukow and J. P. Berberich. A well-balanced active flux method for the shallow water equations with wetting and drying. *Commun. Appl. Math. Comput.*, 6(4):2385–2430, 2024.

[8] W. Barsukow, C. Klingenberg, and S. Krotsch. On the equivalence of semi-discrete Active Flux and Discontinuous Galerkin methods and a comparison of their performance. *arXiv preprint arXiv:2603.18160*, 2026.

[9] P. Batten, N. Clarke, C. Lambert, and D. M. Causon. On the choice of wavespeeds for the HLLC Riemann solver. *SIAM J. Sci. Comput.*, 18(6):1553–1570, 1997.

[10] J. P. Berberich, R. Käppeli, P. Chandrashekar, and C. Klingenberg. High order discretely well-balanced methods for arbitrary hydrostatic atmospheres. *Commun. Comput. Phys.*, 30(3):666–708, 2021.

[11] A. Bermudez and M. E. Vazquez. Upwind methods for hyperbolic conservation laws with source terms. *Comput. Fluids*, 23(8):1049–1071, 1994.

[12] E. Chudzik, C. Helzel, and D. Kerkmann. The Cartesian grid active flux method: linear stability and bound preserving limiting. *Appl. Math. Comput.*, 393:125501, 2021.

[13] E. Chudzik, C. Helzel, and M. Lukáčová-Medvid'ová. Active flux methods for hyperbolic systems using the method of bicharacteristics. *J. Sci. Comput.*, 99(1):16, 2024.

[14] S. Clain, S. Diot, and R. Loubère. A high-order finite volume method for systems of conservation laws—Multi-dimensional Optimal Order Detection (MOOD). *J. Comput. Phys.*, 230(10):4028–4050, 2011.

[15] S. Diot, S. Clain, and R. Loubère. Improved detection criteria for the multi-dimensional optimal order detection (MOOD) on unstructured meshes with very high-order polynomials. *Comput. Fluids*, 64:43–63, 2012.

[16] J. Duan, W. Barsukow, and C. Klingenberg. Active flux methods for hyperbolic conservation laws—flux vector splitting and bound-preservation. *SIAM J. Sci. Comput.*, 47(2):A811–A837, 2025.

[17] M. Dumbser, O. Zanotti, R. Loubère, and S. Diot. A posteriori subcell limiting of the discontinuous galerkin finite element method for hyperbolic conservation laws. *J. Comput. Phys.*, 278:47–75, 2014.

[18] T. Eymann and P. Roe. Active flux schemes for systems. In *20th AIAA computational fluid dynamics conference*, page 3840, 2011.

[19] T. A. Eymann and P. L. Roe. Multidimensional active flux schemes. In *21st AIAA computational fluid dynamics conference*, page 2940, 2013.

[20] D. Ghosh and E. M. Constantinescu. Well-balanced, conservative finite difference algorithm for atmospheric flows. *AIAA Journal*, 54(4):1370–1385, 2016.

[21] S. Gottlieb, C.-W. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Rev.*, 43(1):89–112, 2001.

[22] L. Grosheintz-Laval and R. Käppeli. Well-balanced finite volume schemes for nearly steady adiabatic flows. *J. Comput. Phys.*, 423:109805, 2020.

[23] R. Käppeli and S. Mishra. Well-balanced schemes for the Euler equations with gravitation. *J. Comput. Phys.*, 259:199–219, 2014.

[24] C. Klingenberg, G. Puppo, and M. Semplice. Arbitrary order finite volume well-balanced schemes for the Euler equations with gravity. *SIAM J. Sci. Comput.*, 41(2):A695–A721, 2019.

[25] R. J. LeVeque and D. S. Bale. Wave Propagation Methods for Conservation Laws with Source Terms. In R. Jeltsch and M. Fey, editors, *Hyperbolic Problems: Theory, Numerics, Applications*, pages 609–618. Birkhäuser Basel, 1999.

[26] G. Li and Y. Xing. High order finite volume WENO schemes for the Euler equations under gravitational fields. *J. Comput. Phys.*, 316:145–163, 2016.

[27] G. Li and Y. Xing. Well-balanced discontinuous Galerkin methods with hydrostatic reconstruction for the Euler equations with gravitation. *J. Comput. Phys.*, 352:445–462, 2018.

[28] C.-W. Shu. Total-variation-diminishing time discretizations. *SIAM J. Sci. Statist. Comput.*, 9(6):1073–1084, 1988.

[29] B. Van Leer. Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection. *J. Comput. Phys.*, 23(3):276–299, 1977.

[30] D. Varma and P. Chandrashekar. A second-order, discretely well-balanced finite volume scheme for Euler equations with gravity. *Comput. Fluids*, 181:292–313, 2019.

[31] F. Vilar. A posteriori correction of high-order discontinuous galerkin scheme through subcell finite volume formulation and flux reconstruction. *J. Comput. Phys.*, 387:245–279, 2019.

[32] K. Wu and Y. Xing. Uniformly high-order structure-preserving discontinuous Galerkin methods for Euler equations with gravitation: Positivity and well-balancedness. *SIAM J. Sci. Comput.*, 43(1):A472–A510, 2021.

[33] Y. Xing and C.-W. Shu. High order well-balanced WENO scheme for the gas dynamics equations under gravitational fields. *J. Sci. Comput.*, 54(2):645–662, 2013.

[34] Z. Xu and C.-W. Shu. A high-order well-balanced discontinuous Galerkin method for hyperbolic balance laws based on the Gauss-Lobatto quadrature rules. *J. Sci. Comput.*, 2024. accepted.

[35] J. Zhang, Y. Xia, and Y. Xu. Moving water equilibria preserving discontinuous Galerkin method for the shallow water equations. *J. Sci. Comput.*, 95:48, 2023.

[36] J. Zhang, Y. Xia, and Y. Xu. Equilibrium preserving space in discontinuous Galerkin methods for hyperbolic balance laws. *Commun. Comput. Phys.*, 38(1):109–155, 2025.

[37] W. Zhang, Y. Xing, Y. Xia, and Y. Xu. High-order positivity-preserving well-balanced discontinuous Galerkin methods for Euler equations with gravitation on unstructured meshes. *Commun. Comput. Phys.*, 31(3):771–815, 2022.

[38] X. Zhang and C.-W. Shu. On maximum-principle-satisfying high order schemes for scalar conservation laws. *J. Comput. Phys.*, 229(9):3091–3120, 2010.

[39] X. Zhang and C.-W. Shu. On positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes. *J. Comput. Phys.*, 229(23):8918–8934, 2010.

[40] X. Zhang and C.-W. Shu. Positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations with source terms. *J. Comput. Phys.*, 230(4):1238–1248, 2011.