SIAM J. SCI. COMPUT.
Vol. 43, No. 1, pp. A472–A510

UNIFORMLY HIGH-ORDER STRUCTURE-PRESERVING DISCONTINUOUS GALERKIN METHODS FOR EULER EQUATIONS WITH GRAVITATION: POSITIVITY AND WELL-BALANCEDNESS*

KAILIANG WU† AND YULONG XING‡

Abstract. This paper presents novel high-order accurate discontinuous Galerkin (DG) schemes for the compressible Euler equations under gravitational fields. A notable feature of these schemes is that they are well-balanced for a general known hydrostatic equilibrium state and, at the same time, provably preserve the positivity of density and pressure. In order to achieve the well-balanced and positivity-preserving properties simultaneously, a novel DG spatial discretization is carefully designed with suitable source term reformulation and a properly modified Harten–Lax–van Leer-contact (HLLC) flux. Based on some technical decompositions as well as several key properties of the admissible states and HLLC flux, rigorous positivity-preserving analyses are carried out. It is proven that the resulting well-balanced DG schemes, coupled with strong-stability-preserving time discretizations, satisfy a weak positivity property, which implies that one can apply a simple existing limiter to effectively enforce the positivity-preserving property, without losing high-order accuracy and conservation. The proposed methods and analyses are illustrated with the ideal equation of state (EOS) for notational convenience only, while the extensions to general EOS are straightforward and are discussed in the supplementary material. Extensive one- and two-dimensional numerical tests demonstrate the desired properties of these schemes, including the exact preservation of the equilibrium state, the ability to capture small perturbation of such state, the robustness for solving problems involving low density and/or low pressure, and good resolution for smooth and discontinuous solutions.

Key words. discontinuous Galerkin method, hyperbolic balance laws, positivity-preserving, well-balanced, compressible Euler equations, gravitational field

AMS subject classifications. 65M60, 65M12, 35L60, 35L65

DOI. 10.1137/20M133782X

1. Introduction.

In this paper, we present highly accurate and robust numerical methods for the compressible Euler equations with gravitation, which have wide application in astrophysics and atmospheric science. In the d-dimensional case, this model can be written as the following nonlinear system of balance laws:

$$\mathbf{U}_t + \boldsymbol{\nabla} \cdot \mathbf{F}(\mathbf{U}) = \mathbf{S}(\mathbf{U}, \mathbf{x}), \tag{1}$$

with

$$\mathbf{U} = \begin{pmatrix} \rho \\ \mathbf{m} \\ E \end{pmatrix}, \quad \mathbf{F}(\mathbf{U}) = \begin{pmatrix} \rho\mathbf{u} \\ \rho\mathbf{u} \otimes \mathbf{u} + p\mathbf{I}_d \\ (E+p)\mathbf{u} \end{pmatrix}, \quad \mathbf{S}(\mathbf{U}, \mathbf{x}) = \begin{pmatrix} 0 \\ -\rho\boldsymbol{\nabla}\phi \\ -\mathbf{m} \cdot \boldsymbol{\nabla}\phi \end{pmatrix}. \tag{2}$$

Here $\mathbf{m} = \rho\mathbf{u}$ denotes the momentum vector; ρ, $\mathbf{u}$, and p denote the fluid density, velocity, and pressure, respectively; $\mathbf{I}_d$ is the identity matrix of size d; and

*Submitted to the journal's Methods and Algorithms for Scientific Computing section May 13, 2020; accepted for publication (in revised form) October 28, 2020; published electronically February 1, 2021.

https://doi.org/10.1137/20M133782X

Funding: The work of the second author was partially supported by the NSF grant DMS-1753581.

†Department of Mathematics, Southern University of Science and Technology, Shenzhen, Guangdong 518055, People's Republic of China (wukl@sustech.edu.cn)

‡Department of Mathematics, The Ohio State University, Columbus, OH 43210 USA (xing.205@osu.edu).

$E = \frac{1}{2}\rho\|\mathbf{u}\|^2 + \rho e$ is the total nongravitational energy with e denoting the specific internal energy. The source terms at the right-hand side of (1) represent the effect of the gravitational field, and $\phi(\mathbf{x})$ is the static gravitational potential. An additional thermodynamic equation relating state variables, the so-called equation of state (EOS), is needed to close the system (2). A general EOS can be written as $e = \mathcal{E}(\rho, p)$. For ideal gases it is given by

$$p = (\gamma - 1)\rho e = (\gamma - 1)\left(E - \frac{\|\mathbf{m}\|^2}{2\rho}\right), \tag{3}$$

where the constant $\gamma > 1$ denotes the adiabatic index. Although we will mainly focus on the ideal EOS for better legibility, the methods and analyses presented in this paper are readily extensible to general EOS as shown in the supplementary material.

Equations (1) with (3) form a hyperbolic system of balance laws and admit (nontrivial) hydrostatic equilibrium solutions, in which the gravitational source term is exactly balanced by the flux gradient, with two well-known examples being the isothermal and polytropic equilibria. The astrophysical and atmospheric applications often involve nearly equilibrium flows, which are small perturbations of the hydrostatic equilibrium states. Standard numerical methods may not balance the contribution of the flux and gravitational source terms, and generate large numerical error, especially for a long-time simulation, e.g., in modeling star and galaxy formation. To address the issue, one may need to conduct the simulation on a very refined mesh, which can be time-consuming especially for the multidimensional problems. To save the computational cost, well-balanced methods, which preserve exactly the discrete version of these steady state solutions up to machine accuracy, are designed to effectively capture these nearly equilibrium flows well on relatively coarse meshes. Study of well-balanced methods has attracted much attention over the past few decades. Most of them were proposed for the shallow water equations over a nonflat bottom topology, another prototype example of hyperbolic balance laws; see, e.g., [4, 15, 23, 47, 1, 41, 46, 44] and the references therein. In recent years, well-balanced numerical methods for the Euler equations (1) with gravitation have been designed within several different frameworks, including but not limited to the finite volume methods [24, 5, 19, 6, 25, 20, 21, 17], gas-kinetic schemes [48, 29], finite difference methods [43, 12, 28], and finite element discontinuous Galerkin (DG) methods [26, 7, 27, 35]. Most of these works assume that the target equilibrium is explicitly known, which is also adopted in this paper. Recently, there exist some efforts [11, 20, 8, 34, 3] on designing well-balanced methods for the Euler equation with gravitation, requiring no a priori knowledge of the hydrostatic solution. A numerical comparison between the high-order DG method and well-balanced DG methods was carried out in [35].

Besides maintaining the hydrostatic equilibrium states, another numerical challenge for the system (1) is to preserve the positivity of density and pressure. Such positivity property is not only necessary for the physical nature of the solution, but also crucial for the robustness of numerical computations. In fact, when negative density and/or pressure are produced, numerical instability can develop and cause the breakdown of numerical simulations. However, most high-order accurate schemes for the Euler equations with gravity are generally not positivity-preserving, and thus may suffer from a risk of failure when simulating problems with low density, low pressure, and/or strong discontinuity. In recent years, high-order bound-preserving numerical schemes have been actively studied for hyperbolic systems. Most of them are built upon two types of limiting approaches: a simple scaling limiter [52] for the reconstructed or evolved solution polynomials in finite volume/DG methods (see, e.g.,

[51, 52, 46, 50, 38, 39]) or a flux-correction limiter [49, 18, 40]. For more developments and applications, we refer to the recent review [31] and the references therein. Based on the simple scaling limiter, high-order positivity-preserving DG schemes were constructed for the Euler equations without source term [52, 54] and with source terms including the gravitational source term [53]. The bound-preserving framework was also extended in [37] to the general relativistic Euler equations under strong gravitational fields.

The main objective of this paper is to develop uniformly high-order DG methods, which are well-balanced and at the same time provably positivity-preserving for the Euler equations with gravitation. Most of the existing methods possess only one of these two properties. A recent work to satisfy both properties was studied in [32], based on a new approximate Riemann solver using a relaxation approach. The accuracy of the schemes in [32] was limited to second-order, yet its extension to higher-order is challenging. The framework established in this paper would be the first one, to the best of our knowledge, that achieves this goal with arbitrarily high-order accurate schemes. The efforts in this paper are summarized as follows.

- One key novelty of this work is to devise novel high-order well-balanced DG schemes, with suitable source term treatments and proper well-balanced numerical fluxes, so that the desired positivity-preserving property is also accommodated in the discretization at the same time.
- We use a properly modified Harten–Lax–van Leer-contact (HLLC) numerical flux, instead of the modified Lax–Friedrichs (LF) fluxes employed in the previous well-balanced DG study [26]. Motivated by the contact property of the HLLC flux observed in [6], we will show in our framework that the HLLC flux can be properly modified, in a *unified* way, to be well-balanced with our discrete source terms for an *arbitrary* known hydrostatic equilibrium. The proposed modification to the HLLC flux is novel and very different from the modifications to the LF flux in [26, 25], which were not formulated in a unified way but were presented for two special equilibria (isothermal and polytropic equilibria) in a separate case-by-case way. More importantly, our new modification does not affect the high-order accuracy and also retains the desired positivity-preserving property, which cannot be shown for the existing modified LF fluxes when polytropic equilibrium is considered.
- Our source term discretization is motivated by [43], where the gravitational source is first reformulated into an equivalent special form using the corresponding hydrostatic equilibrium solution. For the well-balancedness, the reformulation was made based on either the cell-centered solution values (in a DG framework [26]) or the cell average of the solution (in a finite volume framework [25]). In this work, we first make the theoretical observation that the latter reformulation is advantageous for establishing the positivity-preserving property under a milder CFL condition; see Remark 3.7 for details. Moreover, for the theoretical positivity-preserving considerations, we also observe that the source term in the energy equation should be discretized in the same fashion as in the momentum equations, which is not used/required for only the well-balancedness consideration.
- Based on some technical decompositions as well as several key properties of the admissible states and HLLC flux, we will rigorously prove that the resulting well-balanced DG schemes satisfy a weak positivity property, which implies that a simple existing limiter [52, 36] can effectively enforce the positivity-preserving property without losing high-order accuracy and conservation. The well-balanced modification of the numerical flux and discretization of source terms lead to additional difficulties in the positivity-preserving analyses, which are more complicated than

the analyses for the standard DG methods in [52, 53].

It is also worth noting that, in the context of shallow water equations, several positivity-preserving well-balanced schemes have been developed in the literature [22, 46, 42]. In that context, the positivity refers to the nonnegativity of the water height. In the Euler equations (1), the density is the analogue of the water height and is evolved only in the continuity equation, which makes it relatively easy to ensure its positivity. However, it is much more difficult to guarantee the positivity of pressure, since it depends nonlinearly on all the conservative variables $\{\rho, \mathbf{m}, E\}$, as shown in (3). More specifically, the pressure (internal energy) is computed by subtracting the kinetic energy $\|\mathbf{m}\|^2/(2\rho)$ from the total energy E. For high Mach flows or very cold flows, when the numerical errors in E and $\|\mathbf{m}\|^2/(2\rho)$ are large enough, negative pressure can be produced easily. Since the conservative quantities $\{\rho, \mathbf{m}, E\}$ are evolved according to their own conservation laws which are seemingly unrelated, the positivity of pressure is not easy to guarantee numerically. In theory, it is indeed a challenge to make an a priori judgment on whether a numerical scheme is always positivity-preserving under all circumstances or not. For these reasons, seeking positivity-preserving well-balanced schemes for the Euler equations (1) with gravitation is quite nontrivial and cannot directly follow any existing frameworks on shallow water equations.

The rest of this paper is organized as follows. In section 2, we will introduce the stationary hydrostatic solutions of (1) and present several useful properties of the admissible state set and the HLLC flux. We first construct the positivity-preserving well-balanced DG schemes for the one-dimensional (1D) system in section 3, and then extend them to the multidimensional cases in section 4. We conduct numerical tests to verify the properties and effectiveness of the proposed schemes in section 5, before concluding the paper in section 6. The extensions of the proposed methods and analyses to general EOSs are presented in the supplementary material, where we also discuss the positivity of the well-balanced DG schemes with a modified LF flux for the isothermal case.

2. Auxiliary results.

This section introduces the stationary hydrostatic solutions of (1) and presents several useful properties of the admissible state set and the HLLC flux.

2.1. Stationary hydrostatic solutions.

Under the time-independent gravitation potential, the system (1) admits zero-velocity stationary hydrostatic solutions of the form

$$\rho = \rho(\mathbf{x}), \quad \mathbf{u} = \mathbf{0}, \quad \boldsymbol{\nabla} p = -\rho \boldsymbol{\nabla} \phi. \tag{4}$$

The relation (4) alone is not complete, since the density and pressure stratifications are not uniquely defined; see [19]. We usually need to specify the profile of another thermodynamic variable, for example, temperature or entropy, to determine a stable equilibrium. Two important special classes of equilibria arising in the applications are the polytropic [19] and isothermal [43] hydrostatic states. For an isothermal hydrostatic state, we have $T(\mathbf{x}) \equiv T_0$, where T denotes the temperature. For an ideal gas, it is given by

$$\rho = \rho_0 \exp\left(-\frac{\phi}{RT_0}\right), \qquad \mathbf{u} = \mathbf{0}, \qquad p = p_0 \exp\left(-\frac{\phi}{RT_0}\right),$$

where R is the gas constant; p_0, ρ_0, and T_0 are positive constants satisfying $p_0 = \rho_0 R T_0$. A polytropic equilibrium is characterized by $p = K_0 p^\gamma$, which leads to the

form of

$$\rho = \left(\frac{\gamma-1}{K_0\gamma}(C-\phi)\right)^{\frac{1}{\gamma-1}}, \quad \mathbf{u} = \mathbf{0}, \quad p = \frac{1}{K_0^{\frac{1}{\gamma-1}}}\left(\frac{\gamma-1}{\gamma}(C-\phi)\right)^{\frac{\gamma}{\gamma-1}},$$

where K_0 and C are both constant.

2.2. Properties of admissible states. In physics, the density ρ and the pressure p are both positive, which is equivalent to the description that the conservative vector $\mathbf{U}$ should stay in the set of physically admissible states, defined by

$$G := \left\{\mathbf{U} = (\rho, \mathbf{m}, E)^\top : \ \rho > 0, \ \mathcal{G}(\mathbf{U}) := E - \frac{\|\mathbf{m}\|^2}{2\rho} > 0\right\}, \tag{5}$$

where $\mathcal{G}(\mathbf{U})$ is a concave function of $\mathbf{U}$ if $\rho \geq 0$. It is easy to show that the admissible state set G satisfies the following properties, which will be useful in our positivity-preserving analysis.

LEMMA 2.1 (convexity). *The set G is a convex set. Moreover, $\lambda\mathbf{U}_1 + (1-\lambda)\mathbf{U}_0 \in G$ for any $\mathbf{U}_1 \in G, \mathbf{U}_0 \in \overline{G}$, and $\lambda \in (0,1]$, where $\overline{G}$ is the closure of G.*

This property can be verified by definition and Jensen's inequality; see [52].

LEMMA 2.2 (scale invariance). *If $\mathbf{U} \in G$, for any $\lambda > 0$, it holds that $\lambda\mathbf{U} \in G$.*

The proof is straightforward. Combining Lemmas 2.1 and 2.2, we immediately obtain the following stronger property.

LEMMA 2.3. *For any $\lambda_1 > 0$, $\lambda_0 \geq 0$, $\mathbf{U}_1 \in G$, and $\mathbf{U}_0 \in \overline{G}$, we have $\widehat{\mathbf{U}} := \lambda_1\mathbf{U}_1 + \lambda_0\mathbf{U}_0 \in G$.*

Proof. Let $\lambda := \frac{\lambda_1}{\lambda_1+\lambda_0} \in (0,1]$. It follows from Lemma 2.1 that $\lambda\mathbf{U}_1 + (1-\lambda)\mathbf{U}_0 \in G$. Thus, we have $\widehat{\mathbf{U}} = (\lambda_1+\lambda_0)(\lambda\mathbf{U}_1 + (1-\lambda)\mathbf{U}_0) \in G$, according to Lemma 2.2. ∎

LEMMA 2.4. *For any $\lambda \geq 0$, $\delta \in \mathbb{R}$, $\mathbf{U} = (\rho, \mathbf{m}, E)^\top \in G$, and $\mathbf{a} \in \mathbb{R}^d$, if $|\delta|\frac{\|\mathbf{a}\|}{\sqrt{2e}} \leq \lambda$, then*

$$\widehat{\mathbf{U}} := \lambda\mathbf{U} + \delta\big(0, \rho\mathbf{a}, \mathbf{m}\cdot\mathbf{a}\big)^\top \in \overline{G}.$$

Proof. If $\lambda = 0$, it then follows from $|\delta|\|\mathbf{a}\|/\sqrt{2e} \leq \lambda$ that $\delta = 0$ or $\mathbf{a} = \mathbf{0}$, which implies $\widehat{\mathbf{U}} = \mathbf{0} \in \overline{G}$. If $\lambda > 0$, the first component of $\widehat{\mathbf{U}}$ equals $\lambda\rho > 0$, and $\widehat{\mathbf{U}} = (\lambda\rho, \lambda\mathbf{m} + \delta\rho\mathbf{a}, \lambda E + \delta\mathbf{m}\cdot\mathbf{a})^\top$ satisfies

$$\mathcal{G}(\widehat{\mathbf{U}}) = \lambda E + \delta\mathbf{m}\cdot\mathbf{a} - \frac{\|\lambda\mathbf{m} + \delta\rho\mathbf{a}\|^2}{2\lambda\rho} = \rho e\left(1 + |\delta|\frac{\|\mathbf{a}\|}{\lambda\sqrt{2e}}\right)\left(\lambda - |\delta|\frac{\|\mathbf{a}\|}{\sqrt{2e}}\right) \geq 0,$$

where the last inequality follows from the condition $|\delta|\|\mathbf{a}\|/\sqrt{2e} \leq \lambda$. Therefore, $\widehat{\mathbf{U}} \in \overline{G}$. ∎

LEMMA 2.5. *For any $\mathbf{U} \in G$ and any unit vector $\mathbf{n} \in \mathbb{R}^d$, we have $\mathbf{U} - \lambda\mathbf{F}(\mathbf{U})\cdot\mathbf{n} \in G$, for any $\lambda \in \mathbb{R}$ satisfying $|\lambda|\alpha_{\mathbf{n}}(\mathbf{U}) \leq 1$, where $\alpha_{\mathbf{n}}(\mathbf{U}) := |\mathbf{u}\cdot\mathbf{n}| + \sqrt{\gamma p/\rho}$.*

The proof of Lemma 2.5 can be found in [52, 50].

2.3. Properties of HLLC flux in one dimension. In this subsection, we introduce several important properties of the HLLC numerical flux, whose properly modified version will be a key ingredient of our numerical schemes presented later. For notational convenience, we focus here on the properties of the HLLC flux in the

1D case ($d = 1$), while the multidimensional extensions will be discussed in section 4.1.1.

In the 1D case, the HLLC flux (see, for example, [2, 33]) is defined by

$$\mathbf{F}^{hllc}(\mathbf{U}_L, \mathbf{U}_R) = \begin{cases} \mathbf{F}(\mathbf{U}_L) & \text{if } 0 \le S_L, \\ \mathbf{F}_{*L} & \text{if } S_L \le 0 \le S_*, \\ \mathbf{F}_{*R} & \text{if } S_* \le 0 \le S_R, \\ \mathbf{F}(\mathbf{U}_R) & \text{if } 0 \ge S_R, \end{cases} \tag{6}$$

where S_L and S_R are the estimated (left and right) fastest signal velocities arising from the solution of the Riemann problem, and the middle wave speed S_* and fluxes are given by

$$S_* = \frac{p_R - p_L + \rho_L u_L(S_L - u_L) - \rho_R u_R(S_R - u_R)}{\rho_L(S_L - u_L) - \rho_R(S_R - u_R)},$$
$$\mathbf{F}_{*i} = \mathbf{F}_i + S_i(\mathbf{U}_{*i} - \mathbf{U}_i), \quad i = L, R,$$

with the intermediate states given by

$$\mathbf{U}_{*i} = \rho_i \left(\frac{S_i - u_i}{S_i - S_*}\right) \begin{pmatrix} 1 \\ S_* \\ \frac{E_i}{\rho_i} + (S_* - u_i)\left(S_* + \frac{p_i}{\rho_i(S_i - u_i)}\right) \end{pmatrix}. \tag{7}$$

With $\alpha_\pm = u \pm \sqrt{\gamma p/\rho}$, the following estimates of S_L and S_R are used in our computation:

$$S_L = \min\{\alpha_-(\mathbf{U}_L), \alpha_-(\mathbf{U}_R)\}, \qquad S_R = \max\{\alpha_+(\mathbf{U}_L), \alpha_+(\mathbf{U}_R)\}. \tag{8}$$

The HLLC flux possesses two important properties, namely the contact property (see, e.g., [6]) and the positivity [2], as outlined below.

LEMMA 2.6. *For any two states* $\mathbf{U}_L = (\rho_L, 0, p/(\gamma-1))^\top$ *and* $\mathbf{U}_R = (\rho_R, 0, p/(\gamma-1))^\top$*, the HLLC flux* (6) *satisfies*

$$\mathbf{F}^{hllc}(\mathbf{U}_L, \mathbf{U}_R) = (0, p, 0)^\top.$$

The proof is straightforward. The importance of this property for the well-balancedness was observed and used in [6].

LEMMA 2.7. *For any two admissible states* $\mathbf{U}_L \in G$ *and* $\mathbf{U}_R \in G$*, the intermediate states defined in* (7) *satisfy*

$$\mathbf{U}_{*L} \in G, \quad \mathbf{U}_{*R} \in G.$$

The proof of this property for the Euler equations can be found in [2, section 5.3]. As a direct consequence of Lemma 2.7, we have the following conclusions, which are relevant to the positivity of the HLLC scheme for the 1D Euler equations without gravitation.

LEMMA 2.8. *For any two admissible states* $\mathbf{U}_0, \mathbf{U}_1 \in G$*, one has*

$$\mathbf{U}_\lambda^{(1)} := \mathbf{U}_1 - \lambda\left(\mathbf{F}(\mathbf{U}_1) - \mathbf{F}^{hllc}(\mathbf{U}_0, \mathbf{U}_1)\right) \in G, \tag{9}$$

$$\mathbf{U}_\lambda^{(0)} := \mathbf{U}_0 - \lambda\left(\mathbf{F}^{hllc}(\mathbf{U}_0, \mathbf{U}_1) - \mathbf{F}(\mathbf{U}_0)\right) \in G \tag{10}$$

if $\lambda > 0$ *and satisfies*

$$\lambda \max_{\mathbf{U}\in\{\mathbf{U}_0,\mathbf{U}_1\}} \alpha_{\max}(\mathbf{U}) \le 1, \tag{11}$$

where

$$\alpha_{\max}(\mathbf{U}) := |u| + \sqrt{\gamma p/\rho} = \max\{|\alpha_-(\mathbf{U})|, |\alpha_+(\mathbf{U})|\}.$$

Proof. Let $S_1 := S_L(\mathbf{U}_0, \mathbf{U}_1)$, which satisfies $\lambda|S_1| \le 1$. According to the definition of the HLLC flux, we derive that

$$\mathbf{U}_\lambda^{(1)} = \int_0^{\lambda \max\{S_1,0\}} \mathcal{R}(x/\lambda, \mathbf{U}_0, \mathbf{U}_1)\mathrm{d}x + (1 - \lambda \max\{S_1, 0\})\mathbf{U}_1,$$

where $\mathcal{R}(x/t, \mathbf{U}_L, \mathbf{U}_R)$ denotes the approximate HLLC solution to the Riemann problem between the states $\mathbf{U}_L$ and $\mathbf{U}_R$, i.e.,

$$\mathcal{R}(x/t, \mathbf{U}_L, \mathbf{U}_R) = \begin{cases} \mathbf{U}_L & \text{if } \frac{x}{t} \le S_L, \\ \mathbf{U}_{*L} & \text{if } S_L \le \frac{x}{t} \le S_*, \\ \mathbf{U}_{*R} & \text{if } S_* \le \frac{x}{t} \le S_R, \\ \mathbf{U}_R & \text{if } \frac{x}{t} \ge S_R. \end{cases}$$

Thanks to Lemma 2.7, we have $\mathcal{R}(x/t, \mathbf{U}_0, \mathbf{U}_1) \in G$ for all $x \in \mathbb{R}$ and $t > 0$. The convexity of G leads to $\mathbf{U}_\lambda^{(1)} \in G$ under the condition (11). A similar argument yields $\mathbf{U}_\lambda^{(0)} \in G$. ∎

LEMMA 2.9. *For any three admissible states* $\mathbf{U}_L, \mathbf{U}_M, \mathbf{U}_R \in G$*, one has*

$$\mathbf{U}_\lambda := \mathbf{U}_M - \lambda\left(\mathbf{F}^{hllc}(\mathbf{U}_M, \mathbf{U}_R) - \mathbf{F}^{hllc}(\mathbf{U}_L, \mathbf{U}_M)\right) \in G$$

if $\lambda > 0$ *satisfies*

$$\lambda \max_{\mathbf{U}\in\{\mathbf{U}_L,\mathbf{U}_M,\mathbf{U}_R\}} \alpha_{\max}(\mathbf{U}) \le \frac{1}{2}. \tag{12}$$

Proof. Under the condition (12), applying Lemma 2.8 leads to

$$\mathbf{U}_M - 2\lambda\left(\mathbf{F}(\mathbf{U}_M) - \mathbf{F}^{hllc}(\mathbf{U}_L, \mathbf{U}_M)\right) \in G, \ \mathbf{U}_M - 2\lambda\left(\mathbf{F}^{hllc}(\mathbf{U}_M, \mathbf{U}_R) - \mathbf{F}(\mathbf{U}_M)\right) \in G.$$

Taking the average of the above two terms and using the convexity of G then yields $\mathbf{U}_\lambda \in G$. ∎

As a generalization of Lemmas 2.8 and 2.9, the following results discuss the positivity of a properly modified HLLC flux, used in the construction of well-balanced methods in section 3.

LEMMA 2.10. *For any parameters* ζ_1, ζ_2, ζ_3, $\zeta_4 \in \mathbb{R}^+$ *and any two admissible states* $\mathbf{U}_0, \mathbf{U}_1 \in G$*, if* $\lambda > 0$ *and satisfies* (11)*, we have*

$$\zeta_2\mathbf{U}_1 - \lambda\left(\mathbf{F}(\zeta_2\mathbf{U}_1) - \mathbf{F}^{hllc}(\zeta_1\mathbf{U}_0, \zeta_2\mathbf{U}_1)\right) \in G, \tag{13}$$

$$\zeta_3\mathbf{U}_0 - \lambda\left(\mathbf{F}^{hllc}(\zeta_3\mathbf{U}_0, \zeta_4\mathbf{U}_1) - \mathbf{F}(\zeta_3\mathbf{U}_0)\right) \in G. \tag{14}$$

This follows from Lemmas 2.8 and 2.2, and noting $\max_{\mathbf{U}\in\{\zeta_1\mathbf{U}_0,\zeta_2\mathbf{U}_1\}} \alpha_{\max}(\mathbf{U}) = \max_{\mathbf{U}\in\{\mathbf{U}_0,\mathbf{U}_1\}} \alpha_{\max}(\mathbf{U})$.

LEMMA 2.11. *For any parameters* $\zeta_1,\ \zeta_2,\ \zeta_3 \in \mathbb{R}^+$ *and any admissible states* $\mathbf{U}_L, \mathbf{U}_M, \mathbf{U}_R \in G$, *if* $\lambda > 0$ *satisfies* (12), *we have*

$$\zeta_2 \mathbf{U}_M - \lambda \left(\mathbf{F}^{hllc}(\zeta_2 \mathbf{U}_M, \zeta_3 \mathbf{U}_R) - \mathbf{F}^{hllc}(\zeta_1 \mathbf{U}_L, \zeta_2 \mathbf{U}_M) \right) \in G.$$

The proof directly follows from Lemma 2.9 by noting that $\zeta_1 \mathbf{U}_L, \zeta_2 \mathbf{U}_M, \zeta_3 \mathbf{U}_R \in G$ (due to Lemma 2.2) and that $\max_{\mathbf{U} \in \{\zeta_1 \mathbf{U}_L, \zeta_2 \mathbf{U}_M, \zeta_3 \mathbf{U}_R\}} \alpha_{\max}(\mathbf{U}) = \max_{\mathbf{U} \in \{\mathbf{U}_L, \mathbf{U}_M, \mathbf{U}_R\}} \alpha_{\max}(\mathbf{U})$.

3. Positivity-preserving well-balanced DG methods in one dimension. In one spatial dimension, the Euler equations (1) take the form of

$$\mathbf{U}_t + (\mathbf{F}(\mathbf{U}))_x = \mathbf{S}(\mathbf{U}, x), \tag{15}$$

with

$$\mathbf{U} = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix}, \quad \mathbf{F}(\mathbf{U}) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ (E+p)u \end{pmatrix}, \quad \mathbf{S}(\mathbf{U}, x) = \begin{pmatrix} 0 \\ -\rho \phi_x \\ -m \phi_x \end{pmatrix}. \tag{16}$$

3.1. Well-balanced DG discretization. Assume that the spatial domain Ω is divided into cells $\{I_j = (x_{j-1/2}, x_{j+1/2})\}$, and the mesh size is denoted by $h_j = x_{j+1/2} - x_{j-1/2}$, with $h = \max_j\{h_j\}$. The center of each cell is $x_j = (x_{j-1/2} + x_{j+1/2})/2$. Denote the DG numerical solutions as $\mathbf{U}_h(x,t)$, and for each $t \in (0, T_f]$, each component of $\mathbf{U}_h$ belongs to the finite-dimensional space of discontinuous piecewise polynomial functions, $\mathbb{V}_h^k$, defined by

$$\mathbb{V}_h^k = \left\{ u(x) \in L^2(\Omega) :\ u(x)\big|_{I_j} \in \mathbb{P}^k(I_j)\ \forall j \right\},$$

where $\mathbb{P}^k(I_j)$ denotes the space of polynomials of degree up to k in cell I_j. Then the semidiscrete DG methods for (15) are given as follows: for any test function $v \in \mathbb{V}_h^k$, $\mathbf{U}_h$ is computed by

$$\int_{I_j} (\mathbf{U}_h)_t v \mathrm{d}x - \int_{I_j} \mathbf{F}(\mathbf{U}_h) v_x \mathrm{d}x + \widehat{\mathbf{F}}_{j+\frac{1}{2}} v(x^-_{j+\frac{1}{2}}) - \widehat{\mathbf{F}}_{j-\frac{1}{2}} v(x^+_{j-\frac{1}{2}}) = \int_{I_j} \mathbf{S} v \mathrm{d}x, \tag{17}$$

where $\widehat{\mathbf{F}}_{j+1/2}$ denotes the numerical flux at $x_{j+1/2}$. The notations $x^-_{j+1/2}$ and $x^+_{j+1/2}$ indicate the associated limits at $x_{j+1/2}$ taken from the left and right sides, respectively, with $\mathbf{U}^{\pm}_{j+1/2} := \mathbf{U}_h(x^{\pm}_{j+1/2})$. For notional convenience, the t dependence of all quantities is suppressed hereafter.

Now, we construct the well-balanced DG methods which preserve a general equilibrium state (4). Assume that the target stationary hydrostatic solutions to be preserved are explicitly known and are denoted by $\{\rho^e(x), p^e(x), u^e(x) = 0\}$. This yields

$$(p^e(x))_x = -\rho^e(x)\phi_x, \qquad u^e(x) = 0. \tag{18}$$

Let $\rho_h^e(x)$ and $p_h^e(x)$ denote the projections of $\rho^e(x)$ and $p^e(x)$ onto the space $\mathbb{V}_h^k$, respectively.

To render the DG methods (17) well-balanced, we consider the modified HLLC numerical flux

$$\widehat{\mathbf{F}}_{j+\frac{1}{2}} = \mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j+\frac{1}{2}}}{p_h^e(x^-_{j+\frac{1}{2}})} \mathbf{U}^-_{j+\frac{1}{2}}, \frac{p^{e,\star}_{j+\frac{1}{2}}}{p_h^e(x^+_{j+\frac{1}{2}})} \mathbf{U}^+_{j+\frac{1}{2}} \right), \tag{19}$$

where $p^{e,\star}_{j+\frac{1}{2}}$ is a suitable approximation to the equilibrium pressure at $x_{j+\frac{1}{2}}$. Here we define it as

$$p^{e,\star}_{j+\frac{1}{2}} = \frac{1}{2}\left(p^e_h(x^-_{j+\frac{1}{2}}) + p^e_h(x^+_{j+\frac{1}{2}})\right), \tag{20}$$

and other choices of $p^{e,\star}_{j+\frac{1}{2}}$, including $\min\left\{p^e_h(x^-_{j+\frac{1}{2}}), p^e_h(x^+_{j+\frac{1}{2}})\right\}$ and $\max\left\{p^e_h(x^-_{j+\frac{1}{2}}), p^e_h(x^+_{j+\frac{1}{2}})\right\}$, also work. This modification does not affect the accuracy, provided that $\rho^e(x)$ and $p^e(x)$ are smooth. The element integral $\int_{I_j}\mathbf{F}(\mathbf{U}_h)v_x\mathrm{d}x$ in (17) is approximated by the standard quadrature rule

$$\int_{I_j}\mathbf{F}(\mathbf{U}_h)v_x\mathrm{d}x \approx h_j\sum_{\mu=1}^{N}\omega_\mu\mathbf{F}\big(\mathbf{U}_h(x_j^{(\mu)})\big)v_x(x_j^{(\mu)}), \tag{21}$$

where $\{x_j^{(\mu)}, \omega_\mu\}_{1\le\mu\le N}$ denote the N-point Gauss quadrature nodes and weights in I_j.

Next, we consider the discretization of the integrals of the source terms in (17) to achieve the well-balanced property. Let $\mathbf{S} =: (0, S^{[2]}, S^{[3]})^\top$. Following the techniques in [43, 26, 25], we reformulate and decompose the integral of the source term in the momentum equation as

$$\begin{aligned}\int_{I_j} S^{[2]}v\mathrm{d}x &= \int_{I_j} -\rho\phi_x v\mathrm{d}x = \int_{I_j}\frac{\rho}{\rho^e}p^e_x v\mathrm{d}x = \int_{I_j}\left(\frac{\rho}{\rho^e} - \frac{\overline{\rho}_j}{\overline{\rho}^e_j} + \frac{\overline{\rho}_j}{\overline{\rho}^e_j}\right)p^e_x v\mathrm{d}x\\ &= \int_{I_j}\left(\frac{\rho}{\rho^e} - \frac{\overline{\rho}_j}{\overline{\rho}^e_j}\right)p^e_x v\mathrm{d}x + \frac{\overline{\rho}_j}{\overline{\rho}^e_j}\left(p^e(x^-_{j+\frac{1}{2}})v(x^-_{j+\frac{1}{2}}) - p^e(x^+_{j-\frac{1}{2}})v(x^+_{j-\frac{1}{2}}) - \int_{I_j}p^e v_x\mathrm{d}x\right),\end{aligned} \tag{22}$$

where (18) has been used in the second identity, and the notation $\overline{(\cdot)}_j$ denotes the cell average of the associated quantity over I_j. We then approximate it by

$$\begin{aligned}\int_{I_j} S^{[2]}v\mathrm{d}x &\approx h_j\sum_{\mu=1}^{N}\omega_\mu\left(\frac{\rho_h(x_j^{(\mu)})}{\rho^e_h(x_j^{(\mu)})} - \frac{\overline{(\rho_h)}_j}{\overline{(\rho^e_h)}_j}\right)(p^e_h)_x(x_j^{(\mu)})v(x_j^{(\mu)})\\ &\quad + \frac{\overline{(\rho_h)}_j}{\overline{(\rho^e_h)}_j}\left(p^{e,\star}_{j+\frac{1}{2}}v(x^-_{j+\frac{1}{2}}) - p^{e,\star}_{j-\frac{1}{2}}v(x^+_{j-\frac{1}{2}}) - h_j\sum_{\mu=1}^{N}\omega_\mu p^e_h(x_j^{(\mu)})v_x(x_j^{(\mu)})\right) =: \left\langle S^{[2]}, v\right\rangle_j.\end{aligned} \tag{23}$$

Similarly, we approximate the integral of the source term in the energy equation by

$$\begin{aligned}\int_{I_j} S^{[3]}v\mathrm{d}x &\approx h_j\sum_{\mu=1}^{N}\omega_\mu\left(\frac{m_h(x_j^{(\mu)})}{\rho^e_h(x_j^{(\mu)})} - \frac{\overline{(m_h)}_j}{\overline{(\rho^e_h)}_j}\right)(p^e_h)_x(x_j^{(\mu)})v(x_j^{(\mu)})\\ &\quad + \frac{\overline{(m_h)}_j}{\overline{(\rho^e_h)}_j}\left(p^{e,\star}_{j+\frac{1}{2}}v(x^-_{j+\frac{1}{2}}) - p^{e,\star}_{j-\frac{1}{2}}v(x^+_{j-\frac{1}{2}}) - h_j\sum_{\mu=1}^{N}\omega_\mu p^e_h(x_j^{(\mu)})v_x(x_j^{(\mu)})\right) =: \left\langle S^{[3]}, v\right\rangle_j.\end{aligned} \tag{24}$$

By combining them, we have the well-balanced DG methods of the form

(25)
$$\int_{I_j}(\mathbf{U}_h)_t v\mathrm{d}x = h_j\sum_{\mu=1}^{N}\omega_\mu \mathbf{F}\big(\mathbf{U}_h(x_j^{(\mu)})\big)v_x(x_j^{(\mu)}) - \Big(\widehat{\mathbf{F}}_{j+\frac{1}{2}}v(x_{j+\frac{1}{2}}^-) - \widehat{\mathbf{F}}_{j-\frac{1}{2}}v(x_{j-\frac{1}{2}}^+)\Big) + \Big(0, \big\langle S^{[2]}, v\big\rangle_j, \big\langle S^{[3]}, v\big\rangle_j\Big)^\top \qquad \forall v\in\mathbb{V}_h^k.$$

Remark 3.1. We choose here the modified HLLC flux (19), instead of the modified LF fluxes as in [26], due to the following two considerations. First, the HLLC flux satisfies the contact property (Lemma 2.6), which provides a unified modification approach to make the HLLC flux well-balanced for an arbitrary hydrostatic equilibrium; whereas the modifications of the LF flux [26] have to be done separately for different types of equilibria. Second, we will show that our modified HLLC flux (19) also meets the positivity-preserving requirements, whereas the modification to the LF fluxes in the polytropic equilibrium case may lose the positivity-preserving property. We can prove the positivity of the well-balanced DG methods with the modified LF fluxes, only when isothermal equilibria are considered (see the supplementary material).

Remark 3.2. Here, we approximate the integral $\int_{I_j} S^{[3]}vdx$ in (24) in a way consistent with the term $\int_{I_j} S^{[2]}vdx$, while in [26] $\int_{I_j} S^{[3]}vdx$ was approximated by the standard quadrature rule. For the well-balancedness only, either approach is fine, and the standard one is even simpler. However, our analysis will indicate that it is important to use a "consistent" approach for the purpose of accommodating the theoretical positivity-preserving property at the same time.

THEOREM 3.3. *For the* 1*D Euler equations* (15) *with gravitation, the semidiscrete DG schemes* (25) *are well-balanced for a general known stationary hydrostatic solution* (18).

Proof. At the equilibrium state (18), we have $\rho_h = \rho_h^e,\ u_h = u_h^e = 0,\ E_h = \frac{p_h^e}{\gamma-1}$, which leads to

$$\frac{p_{j+\frac{1}{2}}^{e,\star}}{p_h^e(x_{j+\frac{1}{2}}^{\pm})}\mathbf{U}_{j+\frac{1}{2}}^{\pm} = \left(\rho_h^e(x_{j+\frac{1}{2}}^{\pm})\frac{p_{j+\frac{1}{2}}^{e,\star}}{p_h^e(x_{j+\frac{1}{2}}^{\pm})},\ 0,\ \frac{p_{j+\frac{1}{2}}^{e,\star}}{\gamma-1}\right)^\top.$$

Thanks to the contact property (Lemma 2.6), the modified HLLC numerical flux (19) reduces to

$$\widehat{\mathbf{F}}_{j+\frac{1}{2}} = \big(0,\ p_{j+\frac{1}{2}}^{e,\star},\ 0\big)^\top. \tag{26}$$

It is easy to observe that the well-balanced property holds for the mass and energy equations of (25), as the first and third components of both the flux and source term approximations become zero. For the momentum equation, thanks to $\rho_h(x_j^{(\mu)})/\rho_h^e(x_j^{(\mu)}) = \overline{(\rho_h)}_j/\overline{(\rho_h^e)}_j = 1$, we have

$$\big\langle S^{[2]}, v\big\rangle_j = p_{j+\frac{1}{2}}^{e,\star}v(x_{j+\frac{1}{2}}^-) - p_{j-\frac{1}{2}}^{e,\star}v(x_{j-\frac{1}{2}}^+) - h_j\sum_{\mu=1}^{N}\omega_\mu p_h^e(x_j^{(\mu)})v_x(x_j^{(\mu)}).$$

Let $F^{[2]}$ denote the second component of $\mathbf{F}$. Since $u_h = 0$, the flux term $F^{[2]}(\mathbf{U}_h(x_j^{(\mu)}))$

reduces to $p_h^e(x_j^{(\mu)})$. This, together with (26), implies

$$h_j \sum_{\mu=1}^{N} \omega_\mu F^{[2]}\big(\mathbf{U}_h(x_j^{(\mu)})\big) v_x(x_j^{(\mu)}) - \Big(\widehat{F}_{2,j+\frac{1}{2}} v(x_{j+\frac{1}{2}}^-) - \widehat{F}_{2,j-\frac{1}{2}} v(x_{j-\frac{1}{2}}^+)\Big)$$
$$= h_j \sum_{\mu=1}^{N} \omega_\mu p_h^e(x_j^{(\mu)}) v_x(x_j^{(\mu)}) - \Big(p_{j+\frac{1}{2}}^{e,\star} v(x_{j+\frac{1}{2}}^-) - p_{j-\frac{1}{2}}^{e,\star} v(x_{j-\frac{1}{2}}^+)\Big),$$

which is exactly equal to $-\langle S^{[2]}, v\rangle_j$. Therefore, the flux and source term approximations balance each other, which leads to the well-balanced property of our methods (25). ∎

The weak form (25) can be rewritten in the ODE form as

$$\frac{\mathrm{d}\mathbf{U}_h(t)}{\mathrm{d}t} = \mathbf{L}(\mathbf{U}_h), \tag{27}$$

after choosing a suitable basis of $\mathbb{V}_h^k$ and representing $\mathbf{U}_h$ as a linear combination of the basis functions; see [9] for details. The semidiscrete DG schemes (27) can be further discretized in time by some explicit strong-stability-preserving (SSP) Runge–Kutta (RK) methods [14]. For example, with Δt being the time step size, the third-order accurate SSP RK method is given by

$$\begin{aligned}
\mathbf{U}_h^{(1)} &= \mathbf{U}_h^n + \Delta t \mathbf{L}(\mathbf{U}_h^n),\\
\mathbf{U}_h^{(2)} &= \frac{3}{4}\mathbf{U}_h^n + \frac{1}{4}\Big(\mathbf{U}_h^{(1)} + \Delta t \mathbf{L}(\mathbf{U}_h^{(1)})\Big),\\
\mathbf{U}_h^{n+1} &= \frac{1}{3}\mathbf{U}_h^n + \frac{2}{3}\Big(\mathbf{U}_h^{(2)} + \Delta t \mathbf{L}(\mathbf{U}_h^{(2)})\Big).
\end{aligned} \tag{28}$$

3.2. Positivity of first-order well-balanced DG scheme. In this and the next subsections, we shall analyze the positivity of the well-balanced DG schemes (25). The well-balanced modification of the numerical flux and discretization of source terms leads to additional difficulties in the positivity-preserving analyses, which are more complicated than the analyses for the standard DG methods.

Denote the cell average of $\mathbf{U}_h$ over I_j by

$$\overline{\mathbf{U}}_j(t) = \frac{1}{h_j}\int_{I_j} \mathbf{U}_h(x,t)\mathrm{d}x.$$

Taking the test function $v = 1$ in (25), one can obtain the semidiscrete evolution equations satisfied by the cell average as

$$\frac{\mathrm{d}\overline{\mathbf{U}}_j(t)}{\mathrm{d}t} = \mathbf{L}_j(\mathbf{U}_h) := -\frac{1}{h_j}\left(\widehat{\mathbf{F}}_{j+\frac{1}{2}} - \widehat{\mathbf{F}}_{j-\frac{1}{2}}\right) + \overline{\mathbf{S}}_j, \tag{29}$$

where $\overline{\mathbf{S}}_j = \big(0, \overline{S}_j^{[2]}, \overline{S}_j^{[3]}\big)^\top$ with $\overline{S}_j^{[\ell]} := \frac{1}{h_j}\left\langle S^{[\ell]}, 1\right\rangle_j$, $\ell = 2,\ 3$.

When the polynomial degree $k = 0$, we have $\mathbf{U}_h(x,t) \equiv \overline{\mathbf{U}}_j(t)$ for all $x \in I_j$, and the above DG methods (29) reduce to the corresponding first-order scheme with

$$\widehat{\mathbf{F}}_{j+\frac{1}{2}} = \mathbf{F}^{hllc}\left(\frac{p_{j+\frac{1}{2}}^{e,\star}}{\overline{p}_j^e}\overline{\mathbf{U}}_j,\ \frac{p_{j+\frac{1}{2}}^{e,\star}}{\overline{p}_{j+1}^e}\overline{\mathbf{U}}_{j+1}\right). \tag{30}$$

We start by showing the positivity property of the homogeneous case.

LEMMA 3.4. *If the DG polynomial degree $k = 0$ and $\overline{\mathbf{U}}_j \in G$ for all j, we have*

$$\overline{\mathbf{U}}_j - \frac{\Delta t}{h_j}\left(\widehat{\mathbf{F}}_{j+\frac{1}{2}} - \widehat{\mathbf{F}}_{j-\frac{1}{2}}\right) \in G \quad \forall j, \tag{31}$$

under the CFL-type condition

$$\frac{\Delta t}{h_j}\left(\frac{p^{e,\star}_{j+\frac{1}{2}} + p^{e,\star}_{j-\frac{1}{2}}}{\overline{p}^e_j} \max_{\mathbf{U}\in\{\overline{\mathbf{U}}_{j-1},\overline{\mathbf{U}}_j,\overline{\mathbf{U}}_{j+1}\}} \alpha_{\max}(\mathbf{U})\right) \le \frac{1}{2}. \tag{32}$$

Proof. Using (30), we have

$$\begin{aligned}
&\overline{\mathbf{U}}_j - \frac{\Delta t}{h_j}\left(\widehat{\mathbf{F}}_{j+\frac{1}{2}} - \widehat{\mathbf{F}}_{j-\frac{1}{2}}\right)\\
&= \overline{\mathbf{U}}_j - \frac{\Delta t}{h_j}\left[\mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j+\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j, \frac{p^{e,\star}_{j+\frac{1}{2}}}{\overline{p}^e_{j+1}}\overline{\mathbf{U}}_{j+1}\right) - \mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j-\frac{1}{2}}}{\overline{p}^e_{j-1}}\overline{\mathbf{U}}_{j-1}, \frac{p^{e,\star}_{j-\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j\right)\right].
\end{aligned}$$

Note that the well-balanced modification leads to

$$\frac{p^{e,\star}_{j+\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j \neq \frac{p^{e,\star}_{j-\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j,$$

so that the positivity of the standard HLLC scheme cannot be used directly. To address this issue, we make the following decomposition:

$$\overline{\mathbf{U}}_j - \frac{\Delta t}{h_j}\left(\widehat{\mathbf{F}}_{j+\frac{1}{2}} - \widehat{\mathbf{F}}_{j-\frac{1}{2}}\right) = \beta_j\left(\mathbf{W}_1 + \mathbf{W}_2\right),$$

with

$$\beta_j := \frac{\overline{p}^e_j}{p^{e,\star}_{j+\frac{1}{2}} + p^{e,\star}_{j-\frac{1}{2}}} > 0,$$

and

$$\begin{aligned}
\mathbf{W}_1 &= \frac{p^{e,\star}_{j+\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j - \frac{\Delta t}{\beta_j h_j}\left[\mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j+\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j, \frac{p^{e,\star}_{j+\frac{1}{2}}}{\overline{p}^e_{j+1}}\overline{\mathbf{U}}_{j+1}\right) - \mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j-\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j, \frac{p^{e,\star}_{j+\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j\right)\right],\\
\mathbf{W}_2 &= \frac{p^{e,\star}_{j-\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j - \frac{\Delta t}{\beta_j h_j}\left[\mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j-\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j, \frac{p^{e,\star}_{j+\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j\right) - \mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j-\frac{1}{2}}}{\overline{p}^e_{j-1}}\overline{\mathbf{U}}_{j-1}, \frac{p^{e,\star}_{j-\frac{1}{2}}}{\overline{p}^e_j}\overline{\mathbf{U}}_j\right)\right].
\end{aligned}$$

Applying Lemma 2.11 leads to $\mathbf{W}_1, \mathbf{W}_2 \in G$ under the condition (32). We can conclude (31) by using Lemma 2.3, which completes the proof. ∎

For all j, we define $\overline{e}_j := \frac{1}{\overline{\rho}_j}\left(\overline{E}_j - \frac{\overline{m}_j^2}{2\overline{\rho}_j}\right)$ and $\widehat{\alpha}_j := \widehat{\alpha}_j^F + \widehat{\alpha}_j^S$ with

$$\widehat{\alpha}_j^F := 2\frac{p^{e,\star}_{j+\frac{1}{2}} + p^{e,\star}_{j-\frac{1}{2}}}{\overline{p}^e_j} \max_{\mathbf{U}\in\{\overline{\mathbf{U}}_{j-1},\overline{\mathbf{U}}_j,\overline{\mathbf{U}}_{j+1}\}} \alpha_{\max}(\mathbf{U}), \qquad \widehat{\alpha}_j^S := \frac{\left|p^{e,\star}_{j+\frac{1}{2}} - p^{e,\star}_{j-\frac{1}{2}}\right|}{\overline{\rho}^e_j\sqrt{2\overline{e}_j}}.$$

THEOREM 3.5. *If the DG polynomial degree $k = 0$ and $\overline{\mathbf{U}}_j \in G$ for all j, we have*

$$\overline{\mathbf{U}}_j + \Delta t\mathbf{L}_j(\mathbf{U}_h) \in G \quad \forall j, \tag{33}$$

under the CFL-type condition

$$\widehat{\alpha}_j\Delta t \le h_j. \tag{34}$$

Proof. When $k=0$, one has $\overline{\rho}_j^e = \frac{1}{h_j}\int_{I_j}\rho_h^e(x)dx > 0$, $\overline{p}_j^e = \frac{1}{h_j}\int_{I_j} p_h^e(x)dx > 0$, and

$$\overline{\mathbf{S}}_j = \frac{\left|p_{j+\frac{1}{2}}^{e,\star} - p_{j-\frac{1}{2}}^{e,\star}\right|}{h_j\overline{\rho}_j^e}\Big(0,\overline{\rho}_j,\overline{m}_j\Big)^\top. \tag{35}$$

If $\left|p_{j+\frac{1}{2}}^{e,\star} - p_{j-\frac{1}{2}}^{e,\star}\right| = 0$, we have $\overline{\mathbf{S}}_j = \mathbf{0}$, and $\overline{\mathbf{U}}_j + \Delta t\mathbf{L}_j(\mathbf{U}_h) = \overline{\mathbf{U}}_j - \frac{\Delta t}{h_j}\big(\widehat{\mathbf{F}}_{j+\frac{1}{2}} - \widehat{\mathbf{F}}_{j-\frac{1}{2}}\big) \in G$, according to Lemma 3.4. Otherwise, decompose the scheme as

$$\overline{\mathbf{U}}_j + \Delta t\mathbf{L}_j(\mathbf{U}_h) = \overline{\mathbf{U}}_j - \frac{\Delta t}{h_j}\left(\widehat{\mathbf{F}}_{j+\frac{1}{2}} - \widehat{\mathbf{F}}_{j-\frac{1}{2}}\right) + \Delta t\overline{\mathbf{S}}_j = \frac{\widehat{\alpha}_j^F}{\widehat{\alpha}_j}\mathbf{W}_F + \frac{1}{\widehat{\alpha}_j}\mathbf{W}_S, \tag{36}$$

where

$$\mathbf{W}_F := \overline{\mathbf{U}}_j - \frac{\Delta t\widehat{\alpha}_j}{h_j\widehat{\alpha}_j^F}\left(\widehat{\mathbf{F}}_{j+\frac{1}{2}} - \widehat{\mathbf{F}}_{j-\frac{1}{2}}\right),$$

$$\mathbf{W}_S := \widehat{\alpha}_j^S\overline{\mathbf{U}}_j + \widehat{\alpha}_j\Delta t\overline{\mathbf{S}}_j = \widehat{\alpha}_j^S\overline{\mathbf{U}}_j + \Delta t\widehat{\alpha}_j\widehat{\alpha}_j^S\frac{\sqrt{2\overline{e}_j}}{h_j}\Big(0,\overline{\rho}_j,\overline{m}_j\Big)^\top.$$

The condition (34) implies $\left|\Delta t\widehat{\alpha}_j\widehat{\alpha}_j^S\frac{\sqrt{2e_j}}{h_j}\right|\frac{1}{\sqrt{2\overline{e}_j}} \le \widehat{\alpha}_j^S$, which leads to, based on Lemma 2.4, $\mathbf{W}_S \in \overline{G}$. With the aid of Lemma 3.4, we obtain $\mathbf{W}_F \in G$ under the condition (34). Finally, the combination of (36) and Lemma 2.3 yields (33). ∎

Theorem 3.5 indicates that the first-order ($k=0$) well-balanced DG method (25), coupled with a forward Euler time discretization, is positivity-preserving under the CFL-type condition (34).

3.3. Positivity-preserving high-order well-balanced DG schemes. When the polynomial degree $k \ge 1$, the high-order well-balanced DG schemes (25) are not positivity-preserving in general. Fortunately, a weak positivity property can be proven for the schemes (25); see Theorem 3.6. As we will see, such weak positivity is crucial and implies that a simple limiter can enforce the positivity-preserving property without losing conservation and high-order accuracy.

3.3.1. Theoretical positivity-preserving analysis. Let $\{\widehat{x}_j^{(\nu)}\}_{1\le\nu\le L}$ be the Gauss–Lobatto nodes transformed into the interval I_j, and let $\{\widehat{\omega}_\nu\}_{1\le\nu\le L}$ be the associated quadrature weights satisfying $\sum_{\nu=1}^L\widehat{\omega}_\nu = 1$ and $\widehat{\omega}_1 = \widehat{\omega}_L = \frac{1}{L(L-1)}$, with $L \ge (k+3)/2$ to ensure that the algebraic precision of the corresponding quadrature rule is at least k. For each cell I_j, we define the point set

$$\mathbb{S}_j := \{\widehat{x}_j^{(\nu)}\}_{\nu=1}^L \cup \{x_j^{(\mu)}\}_{\mu=1}^N, \tag{37}$$

and define $\widetilde{\alpha}_j$ as

$$\widetilde{\alpha}_j := \widetilde{\alpha}_j^F + \widetilde{\alpha}_j^S + \overline{\alpha}_j^S,\ \ \widetilde{\alpha}_j^F := 2\max\left\{\frac{p_{j+\frac{1}{2}}^{e,\star}}{p_h^e(x_{j+\frac{1}{2}}^-)}, \frac{p_{j-\frac{1}{2}}^{e,\star}}{p_h^e(x_{j-\frac{1}{2}}^+)}\right\}\max_{\mathbf{U}\in\{\mathbf{U}_{j-\frac{1}{2}}^\pm,\mathbf{U}_{j+\frac{1}{2}}^\pm\}}\alpha_{\max}(\mathbf{U}),$$

$$\widetilde{\alpha}_j^S := \widehat{\omega}_1 h_j\max_{1\le\mu\le N}\left\{\frac{\left|\big(p_h^e\big)_x(x_j^{(\mu)})\right|}{\rho_h^e(x_j^{(\mu)})\sqrt{2e_h(x_j^{(\mu)})}}\right\}, \qquad \overline{\alpha}_j^S := \widehat{\omega}_1\frac{\left|[\![p_h^e]\!]_{j+\frac{1}{2}} + [\![p_h^e]\!]_{j-\frac{1}{2}}\right|}{2\overline{\rho}_j^e\sqrt{2\overline{e}_j}}, \tag{38}$$

with $[\![p_h^e]\!]_{j+\frac{1}{2}} := p_h^e(x^+_{j+\frac{1}{2}}) - p_h^e(x^-_{j+\frac{1}{2}})$, where $\widetilde{\alpha}_j^S + \overline{\alpha}_j^S = \mathcal{O}(h_j)$ and

$$\max\left\{\frac{p^{e,\star}_{j+\frac{1}{2}}}{p_h^e(x^-_{j+\frac{1}{2}})}, \frac{p^{e,\star}_{j-\frac{1}{2}}}{p_h^e(x^+_{j-\frac{1}{2}})}\right\} = 1 + \mathcal{O}(h^{k+1})$$

for smooth $p^e(x)$. Then we have the following sufficient condition for the high-order scheme (27) to be positivity-preserving.

THEOREM 3.6. *Assume that the projected stationary hydrostatic solutions satisfy*

$$\rho_h^e(x) > 0, \quad p_h^e(x) > 0 \quad \forall x \in \mathbb{S}_j, \quad \forall j, \tag{39}$$

and the numerical solution $\mathbf{U}_h$ *satisfies*

$$\mathbf{U}_h(x) \in G \quad \forall x \in \mathbb{S}_j, \quad \forall j; \tag{40}$$

then we have the weak positivity property

$$\overline{\mathbf{U}}_j + \Delta t \mathbf{L}_j(\mathbf{U}_h) \in G \quad \forall j, \tag{41}$$

under the CFL-type condition

$$\widetilde{\alpha}_j \Delta t \le \widehat{\omega}_1 h_j. \tag{42}$$

Proof. The exactness of the L-point Gauss–Lobatto quadrature rule for polynomials of degree up to k implies

$$\overline{\mathbf{U}}_j = \frac{1}{h_j}\int_{I_j} \mathbf{U}_h(x)\mathrm{d}x = \sum_{\nu=1}^{L} \widehat{\omega}_\nu \mathbf{U}_h(\widehat{x}_j^{(\nu)}),$$

with $\widehat{x}_j^{(1)} = x_{j-\frac{1}{2}}$, $\widehat{x}_j^{(L)} = x_{j+\frac{1}{2}}$, and $\widehat{\omega}_1 = \widehat{\omega}_L$. We consider, for an arbitrary parameter $\eta \in (0,1]$, the following decomposition:

$$\begin{aligned}
\overline{\mathbf{U}}_j + \Delta t \mathbf{L}_j(\mathbf{U}_h) &= \eta\overline{\mathbf{U}}_j - \frac{\Delta t}{h_j}\left(\widehat{\mathbf{F}}_{j+\frac{1}{2}} - \widehat{\mathbf{F}}_{j-\frac{1}{2}}\right) + (1-\eta)\overline{\mathbf{U}}_j + \Delta t\overline{\mathbf{S}}_j \\
&= \eta\sum_{\nu=1}^{L}\widehat{\omega}_\nu \mathbf{U}_h(\widehat{x}_j^{(\nu)}) - \frac{\Delta t}{h_j}\left(\widehat{\mathbf{F}}_{j+\frac{1}{2}} - \widehat{\mathbf{F}}_{j-\frac{1}{2}}\right) + (1-\eta)\overline{\mathbf{U}}_j + \Delta t\overline{\mathbf{S}}_j \\
&= \left[\eta\sum_{\nu=2}^{L-1}\widehat{\omega}_\nu \mathbf{U}_h(\widehat{x}_j^{(\nu)})\right] + \left[\eta\widehat{\omega}_1\left(\mathbf{U}^+_{j-\frac{1}{2}} + \mathbf{U}^-_{j+\frac{1}{2}}\right) - \frac{\Delta t}{h_j}\left(\widehat{\mathbf{F}}_{j+\frac{1}{2}} - \widehat{\mathbf{F}}_{j-\frac{1}{2}}\right)\right] \\
&\quad + \left[(1-\eta)\overline{\mathbf{U}}_j + \Delta t\overline{\mathbf{S}}_j\right] \\
&=: \mathbf{W}_1 + \mathbf{W}_2 + \mathbf{W}_3,
\end{aligned} \tag{43}$$

where $\mathbf{W}_1 \in G \cup \{\mathbf{0}\} \subset \overline{G}$ according to Lemma 2.3. The parameter η could be simply taken as $1/2$, but this will lead to a restrictive condition for Δt. In the following we would like to determine a suitable parameter η in $(0,1]$ such that $\mathbf{W}_2 \in G$ and $\mathbf{W}_3 \in \overline{G}$.

Let us first consider $\mathbf{W}_2$ and reformulate it as follows:

$$
\begin{aligned}
\mathbf{W}_2 &= \eta\widehat{\omega}_1\mathbf{U}^+_{j-\frac{1}{2}} + \eta\widehat{\omega}_1\mathbf{U}^-_{j+\frac{1}{2}} \\
&\quad - \frac{\Delta t}{h_j}\left[\mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j+\frac{1}{2}}}{p^e_h(x^-_{j+\frac{1}{2}})}\mathbf{U}^-_{j+\frac{1}{2}}, \frac{p^{e,\star}_{j+\frac{1}{2}}}{p^e_h(x^+_{j+\frac{1}{2}})}\mathbf{U}^+_{j+\frac{1}{2}}\right)\right. \\
&\quad \left. - \mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j-\frac{1}{2}}}{p^e_h(x^-_{j-\frac{1}{2}})}\mathbf{U}^-_{j-\frac{1}{2}}, \frac{p^{e,\star}_{j-\frac{1}{2}}}{p^e_h(x^+_{j-\frac{1}{2}})}\mathbf{U}^+_{j-\frac{1}{2}}\right)\right] \\
&= \eta\widehat{\omega}_1\mathbf{U}^-_{j+\frac{1}{2}} - \frac{\Delta t}{h_j}\left[\mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j+\frac{1}{2}}}{p^e_h(x^-_{j+\frac{1}{2}})}\mathbf{U}^-_{j+\frac{1}{2}}, \frac{p^{e,\star}_{j+\frac{1}{2}}}{p^e_h(x^+_{j+\frac{1}{2}})}\mathbf{U}^+_{j+\frac{1}{2}}\right)\right. \\
&\quad \left. - \mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j-\frac{1}{2}}}{p^e_h(x^+_{j-\frac{1}{2}})}\mathbf{U}^+_{j-\frac{1}{2}}, \frac{p^{e,\star}_{j+\frac{1}{2}}}{p^e_h(x^-_{j+\frac{1}{2}})}\mathbf{U}^-_{j+\frac{1}{2}}\right)\right] \\
&\quad + \eta\widehat{\omega}_1\mathbf{U}^+_{j-\frac{1}{2}} - \frac{\Delta t}{h_j}\left[\mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j-\frac{1}{2}}}{p^e_h(x^+_{j-\frac{1}{2}})}\mathbf{U}^+_{j-\frac{1}{2}}, \frac{p^{e,\star}_{j+\frac{1}{2}}}{p^e_h(x^-_{j+\frac{1}{2}})}\mathbf{U}^-_{j+\frac{1}{2}}\right)\right. \\
&\quad \left. - \mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j-\frac{1}{2}}}{p^e_h(x^-_{j-\frac{1}{2}})}\mathbf{U}^-_{j-\frac{1}{2}}, \frac{p^{e,\star}_{j-\frac{1}{2}}}{p^e_h(x^+_{j-\frac{1}{2}})}\mathbf{U}^+_{j-\frac{1}{2}}\right)\right] \\
&=: \eta\widehat{\omega}_1\frac{p^e_h(x^-_{j+\frac{1}{2}})}{p^{e,\star}_{j+\frac{1}{2}}}\mathbf{W}^+_2 + \eta\widehat{\omega}_1\frac{p^e_h(x^+_{j-\frac{1}{2}})}{p^{e,\star}_{j-\frac{1}{2}}}\mathbf{W}^-_2, \qquad (44)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{W}^+_2 &= \frac{p^{e,\star}_{j+\frac{1}{2}}}{p^e_h(x^-_{j+\frac{1}{2}})}\mathbf{U}^-_{j+\frac{1}{2}} - \frac{\Delta t p^{e,\star}_{j+\frac{1}{2}}}{\eta\widehat{\omega}_1 h_j p^e_h(x^-_{j+\frac{1}{2}})} \\
&\quad \times \left[\mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j+\frac{1}{2}}}{p^e_h(x^-_{j+\frac{1}{2}})}\mathbf{U}^-_{j+\frac{1}{2}}, \frac{p^{e,\star}_{j+\frac{1}{2}}}{p^e_h(x^+_{j+\frac{1}{2}})}\mathbf{U}^+_{j+\frac{1}{2}}\right)\right. \\
&\quad \left. - \mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j-\frac{1}{2}}}{p^e_h(x^+_{j-\frac{1}{2}})}\mathbf{U}^+_{j-\frac{1}{2}}, \frac{p^{e,\star}_{j+\frac{1}{2}}}{p^e_h(x^-_{j+\frac{1}{2}})}\mathbf{U}^-_{j+\frac{1}{2}}\right)\right], \\
\mathbf{W}^-_2 &= \frac{p^{e,\star}_{j-\frac{1}{2}}}{p^e_h(x^+_{j-\frac{1}{2}})}\mathbf{U}^+_{j-\frac{1}{2}} - \frac{\Delta t p^{e,\star}_{j-\frac{1}{2}}}{\eta\widehat{\omega}_1 h_j p^e_h(x^+_{j-\frac{1}{2}})} \\
&\quad \times \left[\mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j-\frac{1}{2}}}{p^e_h(x^+_{j-\frac{1}{2}})}\mathbf{U}^+_{j-\frac{1}{2}}, \frac{p^{e,\star}_{j+\frac{1}{2}}}{p^e_h(x^-_{j+\frac{1}{2}})}\mathbf{U}^-_{j+\frac{1}{2}}\right)\right. \\
&\quad \left. - \mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{j-\frac{1}{2}}}{p^e_h(x^-_{j-\frac{1}{2}})}\mathbf{U}^-_{j-\frac{1}{2}}, \frac{p^{e,\star}_{j-\frac{1}{2}}}{p^e_h(x^+_{j-\frac{1}{2}})}\mathbf{U}^+_{j-\frac{1}{2}}\right)\right].
\end{aligned}
$$

Thanks to Lemma 2.11, we have $\mathbf{W}^\pm_2 \in G$ if

$$
\frac{\Delta t p^{e,\star}_{j\pm\frac{1}{2}}}{\eta\widehat{w}_1 h_j p^e_h(x^\mp_{j\pm\frac{1}{2}})} \max_{\mathbf{U}\in\{\mathbf{U}^-_{j-\frac{1}{2}},\mathbf{U}^+_{j-\frac{1}{2}},\mathbf{U}^-_{j+\frac{1}{2}},\mathbf{U}^+_{j+\frac{1}{2}}\}} \alpha_{\max}(\mathbf{U}) \le \frac{1}{2},
$$

or, equivalently,

$$\Delta t\widetilde{\alpha}_j^F \le \eta\widehat{w}_1 h_j. \tag{45}$$

By applying Lemma 2.3 on (44), we obtain $\mathbf{W}_2 \in G$ under the condition (45).

Next, the term $\mathbf{W}_3$ is analyzed. Note that, for an arbitrary parameter $\lambda \in [0,1]$, we have

$$\begin{aligned}
&(1-\eta)\overline{m}_j + \Delta t\overline{S}_j^{[2]} \\
&= (1-\eta)\overline{m}_j + \Delta t\sum_{\mu=1}^N \omega_\mu \left(\frac{\rho_h(x_j^{(\mu)})}{\rho_h^e(x_j^{(\mu)})} - \frac{\overline{\rho}_j}{\overline{\rho}_j^e}\right)(p_h^e)_x(x_j^{(\mu)}) + \frac{\Delta t}{h_j}\frac{\overline{\rho}_j}{\overline{\rho}_j^e}\left(p_{j+\frac12}^{e,\star} - p_{j-\frac12}^{e,\star}\right) \\
&= (1-\eta)\left[(1-\lambda)\overline{m}_j + \lambda\sum_{\mu=1}^N \omega_\mu m_h(x_j^{(\mu)})\right] + \Delta t\sum_{\mu=1}^N \omega_\mu \frac{\rho_h(x_j^{(\mu)})}{\rho_h^e(x_j^{(\mu)})}(p_h^e)_x(x_j^{(\mu)}) \\
&\quad + \frac{\Delta t}{h_j}\frac{\overline{\rho}_j}{\overline{\rho}_j^e}\left(p_{j+\frac12}^{e,\star} - p_{j-\frac12}^{e,\star} - \int_{I_j}(p_h^e)_x \mathrm{d}x\right) \\
&= (1-\eta)\lambda\sum_{\mu=1}^N \omega_\mu m_h(x_j^{(\mu)}) + \Delta t\sum_{\mu=1}^N \omega_\mu \frac{\rho_h(x_j^{(\mu)})}{\rho_h^e(x_j^{(\mu)})}(p_h^e)_x(x_j^{(\mu)}) \\
&\quad + (1-\eta)(1-\lambda)\overline{m}_j + \frac{\Delta t}{h_j}\frac{\overline{\rho}_j}{\overline{\rho}_j^e}\frac12\left(\llbracket p_h^e\rrbracket_{j+\frac12} + \llbracket p_h^e\rrbracket_{j-\frac12}\right),
\end{aligned}$$

and similarly,

$$\begin{aligned}
(1-\eta)\overline{E}_j + \Delta t\overline{S}_j^{[3]} &= (1-\eta)\lambda\sum_{\mu=1}^N \omega_\mu E_h(x_j^{(\mu)}) + \Delta t\sum_{\mu=1}^N \omega_\mu \frac{m_h(x_j^{(\mu)})}{\rho_h^e(x_j^{(\mu)})}(p_h^e)_x(x_j^{(\mu)}) \\
&\quad + (1-\eta)(1-\lambda)\overline{E}_j + \frac{\Delta t}{h_j}\frac{\overline{m}_j}{\overline{\rho}_j^e}\frac12\left(\llbracket p_h^e\rrbracket_{j+\frac12} + \llbracket p_h^e\rrbracket_{j-\frac12}\right).
\end{aligned}$$

Therefore, we have

$$\mathbf{W}_3 = \sum_{\mu=1}^N \omega_\mu \mathbf{W}_3^{(\mu)} + \overline{\mathbf{W}}_3, \tag{46}$$

$$\mathbf{W}_3^{(\mu)} := (1-\eta)\lambda\mathbf{U}_h(x_j^{(\mu)}) + \Delta t\frac{(p_h^e)_x(x_j^{(\mu)})}{\rho_h^e(x_j^{(\mu)})}\left(0,\ \rho_h(x_j^{(\mu)}),\ m_h(x_j^{(\mu)})\right)^\top, \tag{47}$$

$$\overline{\mathbf{W}}_3 := (1-\eta)(1-\lambda)\overline{\mathbf{U}}_j + \Delta t\frac{\llbracket p_h^e\rrbracket_{j+\frac12} + \llbracket p_h^e\rrbracket_{j-\frac12}}{2h_j\overline{\rho}_j^e}\left(0,\ \overline{\rho}_j,\ \overline{m}_j\right)^\top. \tag{48}$$

Thanks to Lemma 2.4, we have $\overline{\mathbf{W}}_3 \in G$ and $\mathbf{W}_3^{(\mu)} \in \overline{G}$ for all μ if

$$\Delta t\max_{1\le\mu\le N}\left\{\frac{\left|\left(p_h^e\right)_x(x_j^{(\mu)})\right|}{\rho_h^e(x_j^{(\mu)})\sqrt{2e_h(x_j^{(\mu)})}}\right\} \le (1-\eta)\lambda, \quad \Delta t\frac{\left|\llbracket p_h^e\rrbracket_{j+\frac12} + \llbracket p_h^e\rrbracket_{j-\frac12}\right|}{2h_j\overline{\rho}_j^e\sqrt{2\overline{e}_j}} \le (1-\eta)(1-\lambda),$$

or, equivalently,

$$\Delta t\widetilde{\alpha}_j^S \le \widehat{\omega}_1 h_j(1-\eta)\lambda, \qquad \Delta t\overline{\alpha}_j^S \le \widehat{\omega}_1 h_j(1-\eta)(1-\lambda). \tag{49}$$

By applying Lemma 2.3 on (46), we obtain $\mathbf{W}_3 \in \overline{G}$ under the condition (49).

Combining these results, we conclude that if Δt satisfies

(50)
$$\Delta t \in \Omega^{(j)}_{\eta,\lambda} := \left\{\tau \in \mathbb{R}^+ :\ \tau\widetilde{\alpha}^F_j \le \eta\widehat{w}_1 h_j,\ \tau\widetilde{\alpha}^S_j \le \widehat{\omega}_1 h_j(1-\eta)\lambda,\ \tau\overline{\alpha}^S_j \le \widehat{\omega}_1 h_j(1-\eta)(1-\lambda)\right\},$$

then

$$\mathbf{W}_1 \in \overline{G}, \quad \mathbf{W}_2 \in G, \quad \mathbf{W}_3 \in \overline{G},$$

which implies (41), i.e., $\overline{\mathbf{U}}_j + \Delta t\mathbf{L}_j(\mathbf{U}_h) = \sum_{i=1}^3 \mathbf{W}_i \in G$, following Lemma 2.3. Since the two parameters η and λ can be chosen arbitrarily in this proof, we would like to specify the "best" η and λ that maximize $\sup\Omega^{(j)}_{\eta,\lambda} =: g(\eta,\lambda)$. Solving such an optimization problem gives

$$\max_{\eta\in(0,1],\lambda\in[0,1]} g(\eta,\lambda) = g(\eta_*,\lambda_*) = \frac{\widehat{\omega}_1 h_j}{\widetilde{\alpha}^F_j + \widetilde{\alpha}^S_j + \overline{\alpha}^S_j} = \frac{\widehat{\omega}_1 h_j}{\widetilde{\alpha}_j},$$

which is reached at $\eta_* = \widetilde{\alpha}^F_j/\widetilde{\alpha}_j$, $\lambda_* = \frac{\widetilde{\alpha}^S}{\widetilde{\alpha}^S+\overline{\alpha}^S}$. Therefore, the condition (50) reduces to

$$\Delta t \le g(\eta_*,\lambda_*),$$

which is equivalent to (42). This finishes the proof. ∎

Theorem 3.6 gives a sufficient condition for the proposed high-order well-balanced DG schemes (27) to ensure that the cell averages $\overline{\mathbf{U}}_j$ are in G, when combined with the forward Euler time discretization. Since any high-order SSP-RK time discretization can be written as a convex combination of the forward Euler method, the same conclusion also holds when SSP-RK time discretization is used.

Remark 3.7. The well-balanced source term reformulation (22) involves the cell average $\{\overline{\rho}_j, \overline{\rho}^e_j\}$, instead of the midpoint values $\{\rho(x_j), \rho^e(x_j)\}$ used in [26], which also works for the purpose of the well-balanced property. However, in the latter case, the vector $\overline{\mathbf{W}}_3$ in (48) would become

$$\overline{\mathbf{W}}_3 := (1-\eta)(1-\lambda)\overline{\mathbf{U}}_j + \Delta t\frac{[\![p^e_h]\!]_{j+\frac{1}{2}} + [\![p^e_h]\!]_{j-\frac{1}{2}}}{2h_j\rho^e_h(x_j)}\left(0, \rho_h(x_j), m_h(x_j)\right)^\top,$$

and a more restrictive condition on Δt is required to ensure $\overline{\mathbf{W}}_3 \in \overline{G}$, because, in general, $\rho_h(x_j)$ and $m_h(x_j)$ are not necessarily components of $\overline{\mathbf{U}}_j$.

3.3.2. Positivity-preserving limiter. A simple positivity-preserving limiter (cf. [52, 36]) can be applied to enforce the condition (40). Denote

$$\overline{\mathbb{G}}^k_h := \left\{\mathbf{u} \in [\mathbb{V}^k_h]^3 :\ \ \frac{1}{h_j}\int_{I_j}\mathbf{u}(x)\mathrm{d}x \in G\ \forall j\right\},$$
$$\mathbb{G}^k_h := \left\{\mathbf{u} \in [\mathbb{V}^k_h]^3 :\ \ \mathbf{u}\big|_{I_j}(x) \in G\ \forall x \in \mathbb{S}_j, \forall j\right\},$$

where $\mathbb{S}_j$ is defined in (37). For any $\mathbf{U}_h \in \overline{\mathbb{G}}^k_h$ with $\mathbf{U}_h\big|_{I_j} =: \mathbf{U}_j(x)$, we define the positivity-preserving limiting operator $\mathbf{\Pi}_h : \overline{\mathbb{G}}^k_h \to \mathbb{G}^k_h$ as

$$\mathbf{\Pi}_h\mathbf{U}_h\big|_{I_j} = \theta^{(2)}_j(\widehat{\mathbf{U}}_j(x) - \overline{\mathbf{U}}_j) + \overline{\mathbf{U}}_j \quad \forall j, \tag{51}$$

with $\theta_j^{(2)} = \min\big\{1, \frac{\mathcal{G}(\overline{\mathbf{U}}_j)-\epsilon_2}{\mathcal{G}(\overline{\mathbf{U}}_j)-\min_{x\in\mathbb{S}_j}\mathcal{G}\left(\widehat{\mathbf{U}}_j(x)\right)}\big\}$, $\mathcal{G}(\mathbf{U})$ defined in (5), $\widehat{\mathbf{U}}_j(x) := (\widehat{\rho}_j(x), \mathbf{m}_j(x), E_j(x))^\top$, and

$$\widehat{\rho}_j(x) = \theta_j^{(1)}(\rho_j(x) - \overline{\rho}_j) + \overline{\rho}_j, \quad \theta_j^{(1)} = \min\left\{1, \frac{\overline{\rho}_j - \epsilon_1}{\overline{\rho}_j - \min_{x\in\mathbb{S}_j}\rho_j(x)}\right\}. \tag{52}$$

Here ϵ_1 and ϵ_2 are two sufficiently small positive numbers, introduced to avoid the effect of the round-off error. In the computation, one can take $\epsilon_1 = \min\{10^{-13}, \overline{\rho}_j\}$ and $\epsilon_2 = \min\{10^{-13}, \mathcal{G}(\overline{\mathbf{U}}_j)\}$. Note that the positivity-preserving limiter keeps the mass conservation $\int_{I_j}\mathbf{\Pi}_h(\mathbf{u})\mathrm{d}x = \int_{I_j}\mathbf{u}\mathrm{d}x\ \forall\mathbf{u}\in\overline{\mathbb{G}}_h^k$ and does not destroy the high-order accuracy; see [51, 52, 50] for details.

Define the initial numerical solutions as $\mathbf{U}_h^0(x) := \mathbf{\Pi}_h\mathbf{P}_h\mathbf{U}(x,0)$. For the well-balanced DG schemes (27) coupled with an SSP-RK method, if the positivity-preserving limiter (51) is used at each RK stage, the resulting fully discrete DG methods are positivity-preserving, namely the numerical solutions $\mathbf{U}_h^n$ always satisfy (40), i.e., $\mathbf{U}_h^n\in\mathbb{G}_h^k$. For example, when the third-order method (28) is adopted, the proposed high-order positivity-preserving well-balanced DG schemes of the form

$$\begin{aligned}
\mathbf{U}_h^{(1)} &= \mathbf{\Pi}_h\left[\mathbf{U}_h^n + \Delta t\mathbf{L}(\mathbf{U}_h^n)\right],\\
\mathbf{U}_h^{(2)} &= \mathbf{\Pi}_h\left[\frac{3}{4}\mathbf{U}_h^n + \frac{1}{4}\Big(\mathbf{U}_h^{(1)} + \Delta t\mathbf{L}(\mathbf{U}_h^{(1)})\Big)\right],\\
\mathbf{U}_h^{n+1} &= \mathbf{\Pi}_h\left[\frac{1}{3}\mathbf{U}_h^n + \frac{2}{3}\Big(\mathbf{U}_h^{(2)} + \Delta t\mathbf{L}(\mathbf{U}_h^{(2)})\Big)\right]
\end{aligned} \tag{53}$$

are positivity-preserving under the CFL-type condition (42).

Remark 3.8. If the projected stationary hydrostatic solutions ρ_h^e and p_h^e do not satisfy the condition (39) in Theorem 3.6, we can redefine $\rho_h^e, p_h^e\in\mathbb{V}_h^k$ as

$$\left(\rho_h^e(x), 0, \frac{p_h^e(x)}{\gamma-1}\right)^\top := \mathbf{\Pi}_h\mathbf{P}_h\left(\rho^e(x), 0, \frac{p^e(x)}{\gamma-1}\right)^\top, \tag{54}$$

where $\mathbf{P}_h$ denotes the L^2-projection onto the space $[\mathbb{V}_h^k]^3$. One can verify that ρ_h^e and p_h^e defined by (54) always satisfy (39). In practice, if the exact stationary hydrostatic solutions ρ^e and p^e do not involve low density or low pressure, the operator $\mathbf{\Pi}_h$ in (54) would not be turned on. We remark that the positivity-preserving DG schemes also retain the well-balanced property, if (54) is used.

Remark 3.9. Note that the CFL constraint (42) is sufficient rather than necessary for preserving positivity. Also, for an RK time discretization, to enforce the CFL condition rigorously, we need to obtain an accurate estimation of $\widetilde{\alpha}_j$ for all the stages of RK based only on the numerical solution at time level n, which is very difficult in most of the test examples. An efficient implementation (cf. [45]) may be, if a preliminary calculation to the next time step produces negative density or pressure, we restart the computation from the time step n with half of the time step size. Our numerical tests demonstrate that the proposed methods always work robustly with a CFL number slightly smaller than $\widehat{\omega}_1$ and the restart is yet never encountered.

4. Positivity-preserving well-balanced DG methods in multiple dimensions. In this section, we extend the proposed 1D positivity-preserving well-balanced

DG methods to the multidimensional cases. For the sake of clarity, we shall focus on the two-dimensional (2D) case with $d = 2$ in the remainder of this section, and the extension of our numerical methods and analyses to the three-dimensional (3D) case ($d = 3$) follows similar lines.

4.1. Well-balanced DG discretization. Assume that the 2D spatial domain Ω is partitioned into a mesh $\mathcal{T}_h$, which may be unstructured and consist of polygonal cells. Throughout this section, the lowercase k is used to denote the DG polynomial degree, while the capital K always represents a cell in $\mathcal{T}_h$. Denote the DG numerical solutions as $\mathbf{U}_h(\mathbf{x}, t)$, and for any $t \in (0, T_f]$, each component of $\mathbf{U}_h$ belongs to the finite-dimensional space of discontinuous piecewise polynomial functions, $\mathbb{V}_h^k$, defined by

$$\mathbb{V}_h^k = \left\{ u(\mathbf{x}) \in L^2(\Omega) : \ u(\mathbf{x})\big|_K \in \mathbb{P}^k(K) \ \forall K \in \mathcal{T}_h \right\},$$

where $\mathbb{P}^k(K)$ is the space of polynomials of total degree up to k in cell K. The semi-discrete DG methods for (1) are given as follows: for any test function $v \in \mathbb{V}_h^k$, $\mathbf{U}_h$ is computed by

$$\int_K (\mathbf{U}_h)_t v \mathrm{d}\mathbf{x} - \int_K \mathbf{F}(\mathbf{U}_h) \cdot \boldsymbol{\nabla} v \mathrm{d}\mathbf{x} + \sum_{\mathscr{E} \in \partial K} \int_{\mathscr{E}} \widehat{\mathbf{F}}_{\mathbf{n}_{\mathscr{E},K}} v^{\mathtt{int}(K)} \mathrm{d}s = \int_K \mathbf{S} v \mathrm{d}\mathbf{x} \quad \forall v \in \mathbb{V}_h^k, \tag{55}$$

where ∂K denotes the boundary of the cell K, $\widehat{\mathbf{F}}_{\mathbf{n}_{\mathscr{E},K}}$ denotes the numerical flux on edge $\mathscr{E}$, $\mathbf{n}_{\mathscr{E},K}$ is the outward unit normal to the edge $\mathscr{E}$ of K, and the superscripts "$\mathtt{int}(K)$" or "$\mathtt{ext}(K)$" indicate that the associated limit of $v(\mathbf{x})$ at the cell interfaces is taken from the interior or the exterior of K.

Assume that the target stationary hydrostatic solutions to be preserved are explicitly known and are denoted by $\{\rho^e(\mathbf{x}), p^e(\mathbf{x}), u^e(\mathbf{x}) = 0\}$. Let $\rho_h^e(\mathbf{x})$ and $p_h^e(\mathbf{x})$ be the projections of $\rho^e(\mathbf{x})$ and $p^e(\mathbf{x})$ onto the space $\mathbb{V}_h^k$, respectively. The design of our multidimensional well-balanced DG methods is similar to the 1D case. More specifically, it is based on the well-balanced numerical flux and source term approximation given as follows.

4.1.1. The modified HLLC numerical fluxes. For any unit vector $\mathbf{n} \in \mathbb{R}^d$, let $\mathbf{F}^{hllc}(\mathbf{U}_L, \mathbf{U}_R; \mathbf{n})$ denote the standard HLLC numerical flux in the direction $\mathbf{n}$ for the 2D Euler equations. Details of the standard HLLC flux in the multidimensional cases can be found in [2], and note that this HLLC numerical flux does not refer to any genuinely multidimensional Riemann solver. Analogous to the 1D HLLC flux, the 2D HLLC flux satisfies the following properties, whose proofs are similar to the 1D case and are omitted.

LEMMA 4.1. *For any two states* $\mathbf{U}_L = (\rho_L, 0, 0, p/(\gamma-1))^\top$ *and* $\mathbf{U}_R = (\rho_R, 0, 0, p/(\gamma-1))^\top$, *the 2D HLLC flux satisfies*

$$\mathbf{F}^{hllc}(\mathbf{U}_L, \mathbf{U}_R; \mathbf{n}) = (0, p\mathbf{n}^\top, 0)^\top.$$

LEMMA 4.2. *For any parameters* $\zeta_1, \ \zeta_2 \in \mathbb{R}^+$ *and any two admissible states* $\mathbf{U}_0, \mathbf{U}_1 \in G$, *one has*

$$\zeta_1 \mathbf{U}_0 - \lambda \left[\mathbf{F}^{hllc}(\zeta_1 \mathbf{U}_0, \zeta_2 \mathbf{U}_1; \mathbf{n}) - \mathbf{F}(\zeta_1 \mathbf{U}_0) \cdot \mathbf{n} \right] \in G$$

if $\lambda > 0$ *and satisfies*

$$\lambda \max_{\mathbf{U} \in \{\mathbf{U}_0, \mathbf{U}_1\}} \alpha_{\mathbf{n}}(\mathbf{U}) \le 1, \qquad \textit{with} \quad \alpha_{\mathbf{n}}(\mathbf{U}) := |\mathbf{u} \cdot \mathbf{n}| + \sqrt{\gamma p/\rho}.$$

Based on the above properties, our well-balanced numerical fluxes are chosen as the modified HLLC flux

$$\widehat{\mathbf{F}}_{\mathbf{n}_{\mathscr{E},K}} = \mathbf{F}^{hllc}\left(\frac{p_h^{e,\star}}{p_h^{e,\mathtt{int}(K)}}\mathbf{U}_h^{\mathtt{int}(K)},\ \frac{p_h^{e,\star}}{p_h^{e,\mathtt{ext}(K)}}\mathbf{U}_h^{\mathtt{ext}(K)};\ \mathbf{n}_{\mathscr{E},K}\right), \tag{56}$$

with $p_h^{e,\star} := \frac{1}{2}\left(p_h^{e,\mathtt{int}(K)} + p_h^{e,\mathtt{ext}(K)}\right)$. Using the N-point Gauss quadrature with $N = k+1$, we obtain the following approximation to the edge integral of numerical flux in (55):

$$\int_{\mathscr{E}} \widehat{\mathbf{F}}_{\mathbf{n}_{\mathscr{E},K}} v^{\mathtt{int}(K)} \mathrm{d}s \approx |\mathscr{E}| \sum_{\mu=1}^{N} \omega_\mu \widehat{\mathbf{F}}_{\mathbf{n}_{\mathscr{E},K}}(\mathbf{x}_{\mathscr{E}}^{(\mu)}) v^{\mathtt{int}(K)}(\mathbf{x}_{\mathscr{E}}^{(\mu)}), \tag{57}$$

where $|\mathscr{E}|$ is the length of the edge $\mathscr{E}$, and $\{\mathbf{x}_{\mathscr{E}}^{(\mu)},\ \omega_\mu\}_{1\le\mu\le N}$ denote the set of 1D N-point Gauss quadrature nodes and weights on the edge $\mathscr{E}$.

4.1.2. Source term approximations. Let $\mathbf{S} =: (0, \mathbf{S}^{[2]}, S^{[3]})^\top$ with $\mathbf{S}^{[2]} := -\rho\boldsymbol{\nabla}\phi$. We decompose the integral of the source terms in the momentum equations as

$$\begin{aligned}\int_K \mathbf{S}^{[2]} v \mathrm{d}\mathbf{x} &= \int_{I_j} -\rho\boldsymbol{\nabla}\phi v \mathrm{d}\mathbf{x} = \int_K \frac{\rho}{\rho^e}\boldsymbol{\nabla}p^e v \mathrm{d}\mathbf{x} = \int_K \left(\frac{\rho}{\rho^e} - \frac{\overline{\rho}_K}{\overline{\rho}_K^e} + \frac{\overline{\rho}_K}{\overline{\rho}_K^e}\right)\boldsymbol{\nabla}p^e v \mathrm{d}\mathbf{x}\\ &= \int_K \left(\frac{\rho}{\rho^e} - \frac{\overline{\rho}_K}{\overline{\rho}_K^e}\right)\boldsymbol{\nabla}p^e v \mathrm{d}x + \frac{\overline{\rho}_K}{\overline{\rho}_K^e}\left(\sum_{\mathscr{E}\in\partial K}\int_{\mathscr{E}} p^e v^{\mathtt{int}(K)}\mathbf{n}_{\mathscr{E},K}\mathrm{d}s - \int_K p^e\boldsymbol{\nabla}v \mathrm{d}\mathbf{x}\right),\end{aligned}$$

where $\boldsymbol{\nabla}p^e = -\rho^e\boldsymbol{\nabla}\phi$ has been used in the second identity, and the notation $\overline{(\cdot)}_K$ denotes the cell average of the associated quantity over the cell K. This source term is then approximated by

$$\begin{aligned}\int_K \mathbf{S}^{[2]} v \mathrm{d}x \approx |K| \sum_{q=1}^{Q} \varpi_q &\left(\frac{\rho_h(\mathbf{x}_K^{(q)})}{\rho_h^e(\mathbf{x}_K^{(q)})} - \frac{\overline{(\rho_h)}_K}{\overline{(\rho_h^e)}_K}\right)\boldsymbol{\nabla}p_h^e(\mathbf{x}_K^{(q)}) v(\mathbf{x}_K^{(q)})\\ &+ \frac{\overline{(\rho_h)}_K}{\overline{(\rho_h^e)}_K}\left[\sum_{\mathscr{E}\in\partial K}\left(|\mathscr{E}|\sum_{\mu=1}^{N}\omega_\mu p_h^{e,\star}(\mathbf{x}_{\mathscr{E}}^{(\mu)}) v^{\mathtt{int}(K)}(\mathbf{x}_{\mathscr{E}}^{(\mu)})\mathbf{n}_{\mathscr{E},K}\right)\right.\\ &\left. - |K|\sum_{q=1}^{Q}\varpi_q p_h^e(\mathbf{x}_K^{(q)})\boldsymbol{\nabla}v(\mathbf{x}_K^{(q)})\right] =: \left\langle \mathbf{S}^{[2]}, v\right\rangle_K,\end{aligned} \tag{58}$$

where $|K|$ is the area of the cell K, and $\{\mathbf{x}_K^{(q)},\ \varpi_K^{(q)}\}_{1\le q\le Q}$ denote a set of 2D quadrature nodes and weights in K. Similarly, we approximate the integral of the source term in the energy equation by

$$\begin{aligned}\int_K S^{[3]} v \mathrm{d}x \approx |K| \sum_{q=1}^{Q} \varpi_q &\left(\frac{\mathbf{m}_h(\mathbf{x}_K^{(q)})}{\rho_h^e(\mathbf{x}_K^{(q)})} - \frac{\overline{(\mathbf{m}_h)}_K}{\overline{(\rho_h^e)}_K}\right)\cdot\boldsymbol{\nabla}p_h^e(\mathbf{x}_K^{(q)}) v(\mathbf{x}_K^{(q)})\\ &+ \frac{\overline{(\mathbf{m}_h)}_K}{\overline{(\rho_h^e)}_K}\left[\sum_{\mathscr{E}\in\partial K}\left(|\mathscr{E}|\sum_{\mu=1}^{N}\omega_\mu p_h^{e,\star}(\mathbf{x}_{\mathscr{E}}^{(\mu)}) v^{\mathtt{int}(K)}(\mathbf{x}_{\mathscr{E}}^{(\mu)})\mathbf{n}_{\mathscr{E},K}\right)\right.\\ &\left. - |K|\sum_{q=1}^{Q}\varpi_q p_h^e(\mathbf{x}_K^{(q)})\boldsymbol{\nabla}v(\mathbf{x}_K^{(q)})\right] =: \left\langle S^{[3]}, v\right\rangle_K.\end{aligned} \tag{59}$$

4.1.3. Well-balanced DG methods. The element integral $\int_K \mathbf{F}(\mathbf{U}_h)\cdot\boldsymbol{\nabla} v\mathrm{d}\mathbf{x}$ should be approximated by the same 2D quadrature set

$$\int_K \mathbf{F}(\mathbf{U}_h)\cdot\boldsymbol{\nabla} v\mathrm{d}\mathbf{x} \approx |K|\sum_{q=1}^{Q}\varpi_q\mathbf{F}\big(\mathbf{U}_h(\mathbf{x}_K^{(q)})\big)\cdot\boldsymbol{\nabla} v(\mathbf{x}_K^{(q)}). \tag{60}$$

Substituting the approximations (56)–(60) into (55) gives the following DG formulation:

$$\begin{aligned}\int_K (\mathbf{U}_h)_t v\mathrm{d}x = {}& |K|\sum_{q=1}^{Q}\varpi_q\mathbf{F}\big(\mathbf{U}_h(\mathbf{x}_K^{(q)})\big)\cdot\boldsymbol{\nabla} v(\mathbf{x}_K^{(q)}) + \Big(0,\big\langle \mathbf{S}^{[2]},v\big\rangle_j,\big\langle S^{[3]},v\big\rangle_j\Big)^\top \\ & - \sum_{\mathscr{E}\in\partial K}\left(|\mathscr{E}|\sum_{\mu=1}^{N}\omega_\mu\widehat{\mathbf{F}}_{\mathbf{n}_{\mathscr{E},K}}(\mathbf{x}_{\mathscr{E}}^{(\mu)})v^{\mathtt{int}(K)}(\mathbf{x}_{\mathscr{E}}^{(\mu)})\right) \qquad \forall v\in\mathbb{V}_h^k.\end{aligned} \tag{61}$$

THEOREM 4.3. *For the 2D Euler equations* (1) *with gravitation, the semidiscrete DG schemes* (61) *are well-balanced for a general known stationary hydrostatic solution* (4).

The proof is similar to that of Theorem 3.3 and is thus omitted.

4.2. Positivity of first-order well-balanced DG scheme. Denote the cell average of $\mathbf{U}_h(\mathbf{x},t)$ over K by $\overline{\mathbf{U}}_K(t)$, and take the test function $v=1$ in (61). We obtain the semidiscrete evolution equations satisfied by the cell average as

$$\frac{\mathrm{d}\overline{\mathbf{U}}_K(t)}{\mathrm{d}t} = \mathbf{L}_K(\mathbf{U}_h) := -\frac{1}{|K|}\sum_{\mathscr{E}\in\partial K}\left(|\mathscr{E}|\sum_{\mu=1}^{N}\omega_\mu\widehat{\mathbf{F}}_{\mathbf{n}_{\mathscr{E},K}}(\mathbf{x}_{\mathscr{E}}^{(\mu)})\right)+\overline{\mathbf{S}}_K, \tag{62}$$

where $\overline{\mathbf{S}}_K=\big(0,\overline{\mathbf{S}}_K^{[2]},\overline{S}_K^{[3]}\big)^\top$ with $\overline{\mathbf{S}}_K^{[\ell]}:=\frac{1}{|K|}\big\langle\mathbf{S}^{[\ell]},1\big\rangle_K$ for $\ell=2,3$.

We start with showing the positivity of the first-order ($k=0$) well-balanced DG scheme (61). For each $K\in\mathcal{T}_h$, let $K_{\mathscr{E}}$ denote the adjacent cell that shares the edge $\mathscr{E}$ with K, and define

$$\widehat{\alpha}_K^F := \max\left\{\max_{\mathscr{E}\in\partial K}\alpha_{\mathbf{n}_{\mathscr{E},K}}(\overline{\mathbf{U}}_K),\max_{\mathscr{E}\in\partial K}\alpha_{\mathbf{n}_{\mathscr{E},K}}(\overline{\mathbf{U}}_{K_{\mathscr{E}}})\right\},\ \widehat{\alpha}_K^S := \frac{\Big\|\sum_{\mathscr{E}\in\partial K}|\mathscr{E}|p_{\mathscr{E},K}^{e,\star}\mathbf{n}_{\mathscr{E},K}\Big\|}{|K|\overline{\rho}_K^e\sqrt{2\overline{e}_K}},$$

where $p_{\mathscr{E},K}^{e,\star}:=(\overline{p}_K^e+\overline{p}_{K_{\mathscr{E}}}^e)/2$.

THEOREM 4.4. *If the DG polynomial degree $k=0$ and $\overline{\mathbf{U}}_K\in G$ for all $K\in\mathcal{T}_h$, we have*

$$\overline{\mathbf{U}}_K+\Delta t\mathbf{L}_K(\mathbf{U}_h)\in G\quad\forall K\in\mathcal{T}_h, \tag{63}$$

under the CFL-type condition

$$\Delta t\left(2\frac{\widehat{\alpha}_K^F}{|K|}\sum_{\mathscr{E}\in\partial K}|\mathscr{E}|\frac{p_{\mathscr{E},K}^{e,\star}}{\overline{p}_K^e}+\widehat{\alpha}_K^S\right)\le 1. \tag{64}$$

Proof. Note that, for $k=0$, $\mathbf{U}_h(\mathbf{x},t)\equiv\overline{\mathbf{U}}_K(t)$ for all $\mathbf{x}\in K$. We have

$$\begin{aligned}\overline{\mathbf{U}}_K+\Delta t\mathbf{L}_K(\mathbf{U}_h)&=\overline{\mathbf{U}}_K-\frac{\Delta t}{|K|}\sum_{\mathscr{E}\in\partial K}|\mathscr{E}|\mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_K}\overline{\mathbf{U}}_K,\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_{K_{\mathscr{E}}}}\overline{\mathbf{U}}_{K_{\mathscr{E}}};\mathbf{n}_{\mathscr{E},K}\right)+\Delta t\overline{\mathbf{S}}_K\\&=\overline{\mathbf{U}}_K-\frac{\Delta t}{|K|}\sum_{\mathscr{E}\in\partial K}\left(|\mathscr{E}|\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_K}\mathbf{F}(\overline{\mathbf{U}}_K)\cdot\mathbf{n}_{\mathscr{E},K}\right)+\Delta t\overline{\mathbf{S}}_K\\&\quad+\frac{\Delta t}{|K|}\sum_{\mathscr{E}\in\partial K}|\mathscr{E}|\left[\mathbf{F}\left(\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_K}\overline{\mathbf{U}}_K\right)\cdot\mathbf{n}_{\mathscr{E},K}\right.\\&\quad\left.-\mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_K}\overline{\mathbf{U}}_K,\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_{K_{\mathscr{E}}}}\overline{\mathbf{U}}_{K_{\mathscr{E}}};\mathbf{n}_{\mathscr{E},K}\right)\right],\end{aligned}$$

where the homogeneous property $\mathbf{F}(a\mathbf{U})=a\mathbf{F}(\mathbf{U})$ for any $a\in\mathbb{R}^+$ has been used. We further split $\overline{\mathbf{U}}_K+\Delta t\mathbf{L}_K(\mathbf{U}_h)$ into four parts as

$$\overline{\mathbf{U}}_K+\Delta t\mathbf{L}_K(\mathbf{U}_h)=\mathbf{W}_1+\mathbf{W}_2+\mathbf{W}_3+\mathbf{W}_4, \tag{65}$$

with

$$\mathbf{W}_1:=\left[1-\Delta t\left(2\frac{\widehat{\alpha}^F_K}{|K|}\sum_{\mathscr{E}\in\partial K}|\mathscr{E}|\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_K}+\widehat{\alpha}^S_K\right)\right]\overline{\mathbf{U}}_K,$$

$$\mathbf{W}_2:=\frac{\Delta t}{|K|}\sum_{\mathscr{E}\in\partial K}|\mathscr{E}|\widehat{\alpha}^F_K\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_K}\left(\overline{\mathbf{U}}_K-\frac{1}{\widehat{\alpha}^F_K}\mathbf{F}(\overline{\mathbf{U}}_K)\cdot\mathbf{n}_{\mathscr{E},K}\right),\ \mathbf{W}_3:=\Delta t\left(\widehat{\alpha}^S_K\overline{\mathbf{U}}_K+\overline{\mathbf{S}}_K\right),$$

$$\begin{aligned}\mathbf{W}_4:=\frac{\Delta t}{|K|}\sum_{\mathscr{E}\in\partial K}|\mathscr{E}|\widehat{\alpha}^F_K\left\{\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_K}\overline{\mathbf{U}}_K-\frac{1}{\alpha^F_K}\left[\mathbf{F}^{hllc}\left(\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_K}\overline{\mathbf{U}}_K,\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_{K_{\mathscr{E}}}}\overline{\mathbf{U}}_{K_{\mathscr{E}}};\mathbf{n}_{\mathscr{E},K}\right)\right.\right.\\\left.\left.-\mathbf{F}\left(\frac{p^{e,\star}_{\mathscr{E},K}}{\overline{p}^e_K}\overline{\mathbf{U}}_K\right)\cdot\mathbf{n}_{\mathscr{E},K}\right]\right\}.\end{aligned}$$

By using Lemma 2.2, it is easy to observe that $\mathbf{W}_1\in\overline{G}$ under the condition (64). Lemma 2.5 leads to $\overline{\mathbf{U}}_K-\frac{1}{\widehat{\alpha}^F_K}\mathbf{F}(\overline{\mathbf{U}}_K)\cdot\mathbf{n}_{\mathscr{E},K}\in G$, which implies $\mathbf{W}_2\in G$ with the aid of Lemma 2.3. Note that

$$\widehat{\alpha}^S_K\overline{\mathbf{U}}_K+\overline{\mathbf{S}}_K=\widehat{\alpha}^S_K\overline{\mathbf{U}}_K+\frac{1}{|K|\overline{\rho}^e_K}\big(0,\ \overline{\rho}_K\mathbf{a},\ \overline{\mathbf{m}}_K\cdot\mathbf{a}\big)^\top,\qquad\mathbf{a}:=\sum_{\mathscr{E}\in\partial K}|\mathscr{E}|p^{e,\star}_{\mathscr{E},K}\mathbf{n}_{\mathscr{E},K},$$

and $\frac{1}{|K|\overline{\rho}^e_K}\frac{\|\mathbf{a}\|}{\sqrt{2\overline{e}_K}}=\widehat{\alpha}^S_K$. This yields $\widehat{\alpha}^S_K\overline{\mathbf{U}}_K+\overline{\mathbf{S}}_K\in\overline{G}$ by Lemma 2.4. Thus $\mathbf{W}_3\in\overline{G}$. Sequentially, using Lemmas 4.2 and 2.3 yields $\mathbf{W}_4\in G$. Because $\mathbf{W}_1,\mathbf{W}_3\in\overline{G}$ and $\mathbf{W}_2,\mathbf{W}_4\in G$, we conclude from (65) that $\overline{\mathbf{U}}_K+\Delta t\mathbf{L}_K(\mathbf{U}_h)\in G$, which completes the proof. ∎

Theorem 4.4 indicates that the first-order ($k=0$) well-balanced DG method (61), coupled with the forward Euler time discretization, is positivity-preserving under the CFL-type condition (64).

4.3. Positivity-preserving high-order well-balanced DG schemes. When the DG polynomial degree $k\geq1$, the high-order well-balanced DG schemes (61) are not positivity-preserving in general. Similar to the 1D case, we can prove that our

schemes satisfy a weak positivity property, which is crucial and implies that a simple limiter can enforce the positivity-preserving property without losing conservation and high-order accuracy.

4.3.1. Theoretical positivity-preserving analysis. Assume that there exists a special 2D quadrature on each cell $K \in \mathcal{T}_h$ satisfying the following:

(i) The quadrature rule has positive weights and is exact for integrals of polynomials of degree up to k on the cell K;

(ii) The set of the quadrature points, denoted by $\mathbb{S}_K^{(1)}$, must include all the Gauss quadrature points $\mathbf{x}_{\mathscr{E}}^{(\mu)}$, $\mu = 1, \ldots, N$, on all the edges $\mathscr{E} \in \partial K$.

In other words, we would like to have a special quadrature such that

$$\frac{1}{|K|}\int_K u(\mathbf{x})\mathrm{d}\mathbf{x} = \sum_{\mathscr{E}\in\partial K}\sum_{\mu=1}^{N} \widehat{\varpi}_{\mathscr{E}}^{(\mu)} u(\mathbf{x}_{\mathscr{E}}^{(\mu)}) + \sum_{q=1}^{\widetilde{Q}} \widetilde{\varpi}_q u(\widetilde{\mathbf{x}}_K^{(q)}) \quad \forall u \in \mathbb{P}^k(K), \tag{66}$$

where $\{\widetilde{\mathbf{x}}_K^{(q)}\}$ are the other (possible) quadrature points in K, and the quadrature weights $\widehat{\varpi}_{\mathscr{E}}^{(\mu)}$ and $\widetilde{\varpi}_q$ are positive. For rectangular cells, this quadrature was constructed in [51, 52] by tensor products of Gauss quadrature and Gauss–Lobatto quadrature. For triangular cells and more general polygons, see [54, 10] for how to construct such quadrature. We remark that this special quadrature is only used in the proof and the positivity-preserving limiter presented later, and will not be used to evaluate any integral in the numerical implementation. With this, we can define the point set

$$\begin{aligned}\mathbb{S}_K &:= \mathbb{S}_K^{(1)} \cup \mathbb{S}_K^{(2)} \\ &= \big\{\mathbf{x}_{\mathscr{E}}^{(\mu)} : \mathscr{E} \in \partial K, 1 \le \mu \le N\big\} \cup \big\{\widetilde{\mathbf{x}}_K^{(q)} : 1 \le q \le \widetilde{Q}\big\} \cup \big\{\mathbf{x}_K^{(q)} : 1 \le q \le Q\big\},\end{aligned} \tag{67}$$

where $\mathbb{S}_K^{(2)} := \{\mathbf{x}_K^{(q)}\}_{1\le q\le Q}$ are the 2D quadrature points involved in the approximations (58)–(60).

For convenience we will frequently use the following shortened notations:

$$\begin{aligned}&\mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)} := \mathbf{U}_h^{\mathtt{int}(K)}(\mathbf{x}_{\mathscr{E}}^{(\mu)}), \qquad \mathbf{U}_{\mathscr{E},\mu}^{\mathrm{ext}(K)} := \mathbf{U}_h^{\mathtt{ext}(K)}(\mathbf{x}_{\mathscr{E}}^{(\mu)}), \qquad p_{\mathscr{E},\mu}^{e,\star} := p_h^{e,\star}(\mathbf{x}_{\mathscr{E}}^{(\mu)}),\\ &p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)} := p_h^{e,\mathtt{int}(K)}(\mathbf{x}_{\mathscr{E}}^{(\mu)}), \qquad p_{\mathscr{E},\mu}^{e,\mathrm{ext}(K)} := p_h^{e,\mathtt{ext}(K)}(\mathbf{x}_{\mathscr{E}}^{(\mu)}),\\ &\qquad\qquad [\![p_h^e(\mathbf{x}_{\mathscr{E}}^{(\mu)})]\!] := p_{\mathscr{E},\mu}^{e,\mathtt{ext}(K)} - p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}.\end{aligned}$$

THEOREM 4.5. *Assume that the projected stationary hydrostatic solution satisfies*

$$\rho_h^e(\mathbf{x}) > 0, \quad p_h^e(\mathbf{x}) > 0 \quad \forall \mathbf{x} \in \mathbb{S}_K, \quad \forall K \in \mathcal{T}_h, \tag{68}$$

and the numerical solution $\mathbf{U}_h$ *satisfies*

$$\mathbf{U}_h(\mathbf{x}) \in G \quad \forall \mathbf{x} \in \mathbb{S}_K, \quad \forall K \in \mathcal{T}_h; \tag{69}$$

then we have

$$\overline{\mathbf{U}}_K + \Delta t \mathbf{L}_K(\mathbf{U}_h) \in G \quad \forall K \in \mathcal{T}_h, \tag{70}$$

under the CFL-type condition

$$\Delta t\left(\widetilde{\alpha}_K^F \frac{2|\mathscr{E}|p_{\mathscr{E},\mu}^{e,\star}}{|K|p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}} + \widetilde{\alpha}_K^S \frac{\widehat{\varpi}_{\mathscr{E}}^{(\mu)}}{\omega_\mu}\right) \le \frac{\widehat{\varpi}_{\mathscr{E}}^{(\mu)}}{\omega_\mu}, \qquad 1 \le \mu \le N \ \forall \mathscr{E} \in \partial K, \ \forall K \in \mathcal{T}_h, \tag{71}$$

where

$$\widetilde{\alpha}_K^F := \max\Big\{ \max_{\mathscr{E}\in\partial K, 1\le\mu\le N} \alpha_{\mathbf{n}_{\mathscr{E},K}}(\mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)}), \max_{\mathscr{E}\in\partial K, 1\le\mu\le N} \alpha_{\mathbf{n}_{\mathscr{E},K}}(\mathbf{U}_{\mathscr{E},\mu}^{\mathtt{ext}(K)})\Big\},$$

$$\widetilde{\alpha}_K^S = \widetilde{\alpha}_K^{S,1} + \widetilde{\alpha}_K^{S,2},$$

$$\widetilde{\alpha}_K^{S,1} := \max_{1\le q\le Q} \left\{ \frac{\left\| \boldsymbol{\nabla} p_h^e(\mathbf{x}_K^{(q)}) \right\|}{\rho_h^e(\mathbf{x}_K^{(q)}) \sqrt{2 e_h(\mathbf{x}_K^{(q)})}} \right\},$$

$$\widetilde{\alpha}_K^{S,2} := \frac{\left\| \sum_{\mathscr{E}\in\partial K} \left(|\mathscr{E}| \sum_{\mu=1}^N \omega_\mu [\![p_h^e(\mathbf{x}_{\mathscr{E}}^{(\mu)})]\!] \right) \mathbf{n}_{\mathscr{E},K} \right\|}{2|K| \overline{\rho}_K^e \sqrt{2\overline{e}_K}}.$$

Proof. For the modified HLLC flux, applying Lemmas 4.2 and 2.3 yields

$$\mathbf{W}_1 := \frac{\Delta t}{|K|} \widetilde{\alpha}_K^F \sum_{\mathscr{E}\in\partial K} |\mathscr{E}| \sum_{\mu=1}^N \omega_\mu \left(\frac{p_{\mathscr{E},\mu}^{e,\star}}{p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}} \mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)} - \frac{1}{\widetilde{\alpha}_K^F} \left[\widehat{\mathbf{F}}_{\mathbf{n}_{\mathscr{E},K}}(\mathbf{x}_{\mathscr{E}}^{(\mu)}) - \mathbf{F}\left(\frac{p_{\mathscr{E},\mu}^{e,\star}}{p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}} \mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)} \right) \cdot \mathbf{n}_{\mathscr{E},K} \right] \right) \in G. \tag{72}$$

Using the formulas of $\mathbf{W}_1$ and $\mathbf{L}_K(\mathbf{U}_h)$ in (72) and (62), respectively, we deduce that

$$\begin{aligned}
&\overline{\mathbf{U}}_K + \Delta t \mathbf{L}_K(\mathbf{U}_h) - \mathbf{W}_1 - \Delta t \overline{\mathbf{S}}_K \\
&= \overline{\mathbf{U}}_K - \frac{\Delta t}{|K|} \widetilde{\alpha}_K^F \sum_{\mathscr{E}\in\partial K} \left[|\mathscr{E}| \sum_{\mu=1}^N \omega_\mu \left(\frac{p_{\mathscr{E},\mu}^{e,\star}}{p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}} \mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)} \right.\right. \\
&\quad \left.\left. + \frac{1}{\widetilde{\alpha}_K^F} \mathbf{F}\left(\frac{p_{\mathscr{E},\mu}^{e,\star}}{p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}} \mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)} \right) \cdot \mathbf{n}_{\mathscr{E},K} \right) \right] \\
&= \overline{\mathbf{U}}_K - 2 \frac{\Delta t}{|K|} \widetilde{\alpha}_K^F \sum_{\mathscr{E}\in\partial K} \left[|\mathscr{E}| \sum_{\mu=1}^N \omega_\mu \left(\frac{p_{\mathscr{E},\mu}^{e,\star}}{p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}} \mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)} \right) \right] \\
&\quad + \frac{\Delta t}{|K|} \widetilde{\alpha}_K^F \sum_{\mathscr{E}\in\partial K} \left[|\mathscr{E}| \sum_{\mu=1}^N \omega_\mu \left(\frac{p_{\mathscr{E},\mu}^{e,\star}}{p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}} \mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)} - \frac{1}{\widetilde{\alpha}_K^F} \frac{p_{\mathscr{E},\mu}^{e,\star}}{p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}} \mathbf{F}\left(\mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)} \right) \cdot \mathbf{n}_{\mathscr{E},K} \right) \right],
\end{aligned} \tag{73}$$

where the homogeneous property $\mathbf{F}(a\mathbf{U}) = a\mathbf{F}(\mathbf{U})$ for any $a \in \mathbb{R}^+$ is used. Applying Lemmas 2.5 and 2.3 implies that

$$\mathbf{W}_2 := \frac{\Delta t}{|K|} \widetilde{\alpha}_K^F \sum_{\mathscr{E}\in\partial K} \left(|\mathscr{E}| \sum_{\mu=1}^N \omega_\mu \frac{p_{\mathscr{E},\mu}^{e,\star}}{p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}} \left(\mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)} - \frac{1}{\widetilde{\alpha}_K^F} \mathbf{F}\left(\mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)} \right) \cdot \mathbf{n}_{\mathscr{E},K} \right) \right) \in G. \tag{74}$$

Based on (73) and the definition of $\mathbf{W}_2$, we rewrite $\overline{\mathbf{U}}_K + \Delta t \mathbf{L}_K(\mathbf{U}_h)$ as

$$\overline{\mathbf{U}}_K + \Delta t \mathbf{L}_K(\mathbf{U}_h) = \mathbf{W}_1 + \mathbf{W}_2 + \mathbf{W}_3, \tag{75}$$

with

$$\mathbf{W}_3 := \overline{\mathbf{U}}_K - 2 \frac{\Delta t}{|K|} \widetilde{\alpha}_K^F \sum_{\mathscr{E}\in\partial K} \left[|\mathscr{E}| \sum_{\mu=1}^N \omega_\mu \left(\frac{p_{\mathscr{E},\mu}^{e,\star}}{p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}} \mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)} \right) \right] + \Delta t \overline{\mathbf{S}}_K.$$

Recall that $\overline{\mathbf{S}}_K = \left(0, \overline{\mathbf{S}}_K^{[2]}, \overline{S}_K^{[3]}\right)^\top$ with $\overline{\mathbf{S}}_K^{[\ell]} = \frac{1}{|K|}\left\langle \mathbf{S}^{[\ell]}, 1\right\rangle_K$, $\ell = 2, 3$. We can reformulate $\overline{\mathbf{S}}_K^{[2]}$ as

$$\begin{aligned}
\overline{\mathbf{S}}_K^{[2]} &= \sum_{q=1}^{Q} \varpi_q \left(\frac{\rho_h(\mathbf{x}_K^{(q)})}{\rho_h^e(\mathbf{x}_K^{(q)})} - \frac{\overline{\rho}_K}{\overline{\rho}_K^e}\right) \boldsymbol{\nabla} p_h^e(\mathbf{x}_K^{(q)}) \\
&\quad + \frac{\overline{\rho}_K}{\overline{\rho}_K^e}\left[\frac{1}{|K|}\sum_{\mathscr{E}\in\partial K}\left(|\mathscr{E}|\sum_{\mu=1}^{N}\omega_\mu p_h^{e,\star}(\mathbf{x}_{\mathscr{E}}^{(\mu)})\mathbf{n}_{\mathscr{E},K}\right)\right] \\
&= \sum_{q=1}^{Q} \varpi_q \frac{\rho_h(\mathbf{x}_K^{(q)})}{\rho_h^e(\mathbf{x}_K^{(q)})} \boldsymbol{\nabla} p_h^e(\mathbf{x}_K^{(q)}) \\
&\quad + \frac{\overline{\rho}_K}{\overline{\rho}_K^e|K|}\left[\sum_{\mathscr{E}\in\partial K}\left(|\mathscr{E}|\sum_{\mu=1}^{N}\omega_\mu p_h^{e,\star}(\mathbf{x}_{\mathscr{E}}^{(\mu)})\mathbf{n}_{\mathscr{E},K}\right) - \int_K \boldsymbol{\nabla} p_h^e(\mathbf{x})\mathrm{d}\mathbf{x}\right] \\
&= \sum_{q=1}^{Q} \varpi_q \frac{\rho_h(\mathbf{x}_K^{(q)})}{\rho_h^e(\mathbf{x}_K^{(q)})} \boldsymbol{\nabla} p_h^e(\mathbf{x}_K^{(q)}) \\
&\quad + \frac{\overline{\rho}_K}{\overline{\rho}_K^e|K|}\sum_{\mathscr{E}\in\partial K}\left(|\mathscr{E}|\sum_{\mu=1}^{N}\omega_\mu p_h^{e,\star}(\mathbf{x}_{\mathscr{E}}^{(\mu)}) - \int_{\mathscr{E}} p_h^e \mathrm{d}s\right)\mathbf{n}_{\mathscr{E},K} \\
&= \sum_{q=1}^{Q} \varpi_q \frac{\rho_h(\mathbf{x}_K^{(q)})}{\rho_h^e(\mathbf{x}_K^{(q)})} \boldsymbol{\nabla} p_h^e(\mathbf{x}_K^{(q)}) + \frac{\overline{\rho}_K}{2\overline{\rho}_K^e|K|}\mathbf{a},
\end{aligned}$$

with $\mathbf{a} := \sum_{\mathscr{E}\in\partial K}\left(|\mathscr{E}|\sum_{\mu=1}^{N}\omega_\mu[\![p_h^e(\mathbf{x}_{\mathscr{E}}^{(\mu)})]\!]\right)\mathbf{n}_{\mathscr{E},K}$, where we used the divergence theorem and the exactness of the quadrature rules for polynomials of degree up to k. Similarly, $\overline{S}_K^{[3]}$ can be written as

$$\overline{S}_K^{[3]} = \sum_{q=1}^{Q} \varpi_q \left(\frac{\mathbf{m}_h(\mathbf{x}_K^{(q)})}{\rho_h^e(\mathbf{x}_K^{(q)})}\right) \cdot \boldsymbol{\nabla} p_h^e(\mathbf{x}_K^{(q)}) + \frac{\overline{\mathbf{m}}_K}{2\overline{\rho}_K^e|K|}\cdot\mathbf{a}.$$

Therefore, $\widetilde{\alpha}_K^S\overline{\mathbf{U}}_K + \overline{\mathbf{S}}_K = (\widetilde{\alpha}_K^{S,1} + \widetilde{\alpha}_K^{S,2})\overline{\mathbf{U}}_K + \overline{\mathbf{S}}_K$ can be reformulated as

$$\begin{aligned}
&\sum_{q=1}^{Q}\varpi_q\left[\widetilde{\alpha}_K^{S,1}\mathbf{U}_h(\mathbf{x}_K^{(q)}) + \frac{1}{\rho_h^e(\mathbf{x}_K^{(q)})}\begin{pmatrix}0\\ \rho_h(\mathbf{x}_K^{(q)})\boldsymbol{\nabla}p_h^e(\mathbf{x}_K^{(q)})\\ \mathbf{m}_h(\mathbf{x}_K^{(q)})\cdot\boldsymbol{\nabla}p_h^e(\mathbf{x}_K^{(q)})\end{pmatrix}\right] \\
&\qquad + \left[\widetilde{\alpha}_K^{S,2}\overline{\mathbf{U}}_K + \frac{1}{2\overline{\rho}_K^e|K|}\begin{pmatrix}0\\ \overline{\rho}_K\mathbf{a}\\ \overline{\mathbf{m}}_h\cdot\mathbf{a}\end{pmatrix}\right].
\end{aligned}$$

Since

$$\frac{1}{\rho_h^e(\mathbf{x}_K^{(q)})}\frac{\left\|\boldsymbol{\nabla}p_h^e(\mathbf{x}_K^{(q)})\right\|}{\sqrt{2e_h(\mathbf{x}_K^{(q)})}} \le \widetilde{\alpha}_K^{S,1}, \qquad \frac{1}{2\overline{\rho}_K^e|K|}\frac{\|\mathbf{a}\|}{\sqrt{2\overline{e}_K}} = \widetilde{\alpha}_K^{S,2},$$

we conclude that $\widetilde{\alpha}_K^S\overline{\mathbf{U}}_K + \overline{\mathbf{S}}_K \in \overline{G}$, according to Lemmas 2.4 and 2.3. It follows that

$$\mathbf{W}_4 := \Delta t\left(\widetilde{\alpha}_K^S\overline{\mathbf{U}}_K + \overline{\mathbf{S}}_K\right) \in \overline{G}. \tag{76}$$

Subtracting $\mathbf{W}_4$ from $\mathbf{W}_3$ gives

$$\mathbf{W}_3-\mathbf{W}_4=(1-\Delta t\widetilde{\alpha}_K^S)\overline{\mathbf{U}}_K-2\frac{\Delta t}{|K|}\widetilde{\alpha}_K^F\sum_{\mathscr{E}\in\partial K}\left[|\mathscr{E}|\sum_{\mu=1}^N\omega_\mu\left(\frac{p_{\mathscr{E},\mu}^{e,\star}}{p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}}\mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)}\right)\right]. \tag{77}$$

Note that the exactness of the quadrature rule (66) for polynomials of degree up to k leads to

$$\overline{\mathbf{U}}_K=\sum_{\mathscr{E}\in\partial K}\sum_{\mu=1}^N\widehat{\varpi}_{\mathscr{E}}^{(\mu)}\mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)}+\sum_{q=1}^{\widetilde{Q}}\widetilde{\varpi}_q\mathbf{U}_h^{\mathtt{int}(K)}(\widetilde{\mathbf{x}}_K^{(q)})=:\sum_{\mathscr{E}\in\partial K}\sum_{\mu=1}^N\widehat{\varpi}_{\mathscr{E}}^{(\mu)}\mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)}+\mathbf{W}_5, \tag{78}$$

and obviously we have $\mathbf{W}_5\in\overline{G}$. Substituting (78) into (77) yields

$$\begin{aligned}\mathbf{W}_3-\mathbf{W}_4&=(1-\Delta t\widetilde{\alpha}_K^S)\mathbf{W}_5\\&\quad+\sum_{\mathscr{E}\in\partial K}\sum_{\mu=1}^N\omega_\mu\left[\frac{\widehat{\varpi}_{\mathscr{E}}^{(\mu)}}{\omega_\mu}-\Delta t\left(\frac{2\widetilde{\alpha}_K^F|\mathscr{E}|p_{\mathscr{E},\mu}^{e,\star}}{|K|p_{\mathscr{E},\mu}^{e,\mathtt{int}(K)}}+\widetilde{\alpha}_K^S\frac{\widehat{\varpi}_{\mathscr{E}}^{(\mu)}}{\omega_\mu}\right)\right]\mathbf{U}_{\mathscr{E},\mu}^{\mathtt{int}(K)},\end{aligned}$$

which belongs to $\overline{G}$, by Lemma 2.3, under the CFL condition (71). Recall that we have shown in (76) that $\mathbf{W}_4\in\overline{G}$. It then follows that $\mathbf{W}_3=(\mathbf{W}_3-\mathbf{W}_4)+\mathbf{W}_4\in\overline{G}$. Recalling $\mathbf{W}_1\in G$ and $\mathbf{W}_2\in G$ in (72) and (74), and from (75) and Lemma 2.3, we finally conclude (70). This completes the proof. ∎

Theorem 4.5 provides a sufficient condition (69) for the proposed high-order well-balanced DG schemes (61) to be positivity-preserving, when an SSP-RK time discretization is used. The condition (69) can again be enforced by a simple positivity-preserving limiter similar to the 1D case; see (51)–(52) with the 1D point set $\mathbb{S}_j$ replaced by the 2D point set (67) accordingly. With the limiter applied at each stage of the SSP-RK time steps, the fully discrete DG schemes are positivity-preserving.

4.3.2. Illustration of some details on Cartesian meshes. Assume that the mesh is rectangular with cells $\{[x_{i-1/2},x_{i+1/2}]\times[y_{\ell-1/2},y_{\ell+1/2}]\}$ and spatial step-sizes $\Delta x_i=x_{i+1/2}-x_{i-1/2}$ and $\Delta y_\ell=y_{\ell+1/2}-y_{\ell-1/2}$ in the x- and y-directions, respectively, where (x,y) denotes the 2D spatial coordinate variables. Let $\mathbb{S}_i^x=\{x_i^{(\mu)}\}_{\mu=1}^N$ and $\mathbb{S}_\ell^y=\{y_\ell^{(\mu)}\}_{\mu=1}^N$ denote the N-point Gauss quadrature nodes in the intervals $[x_{i-1/2},x_{i+1/2}]$ and $[y_{\ell-1/2},y_{\ell+1/2}]$, respectively. For the cell $K=[x_{i-1/2},x_{i+1/2}]\times[y_{\ell-1/2},y_{\ell+1/2}]$, the point sets $\mathbb{S}_K^{(1)}$ and $\mathbb{S}_K^{(2)}$ in (67) are given by (cf. [51])

$$\mathbb{S}_K^{(1)}=\big(\widehat{\mathbb{S}}_i^x\otimes\mathbb{S}_\ell^y\big)\cup\big(\mathbb{S}_i^x\otimes\widehat{\mathbb{S}}_\ell^y\big),\qquad\mathbb{S}_K^{(2)}=\mathbb{S}_i^x\otimes\mathbb{S}_\ell^y, \tag{79}$$

where $\widehat{\mathbb{S}}_i^x=\{\widehat{x}_i^{(\nu)}\}_{\nu=1}^L$ and $\widehat{\mathbb{S}}_\ell^y=\{\widehat{y}_\ell^{(\nu)}\}_{\nu=1}^L$ denote the L-point ($L\ge\frac{k+3}{2}$) Gauss–Lobatto quadrature nodes in the intervals $[x_{i-1/2},x_{i+1/2}]$ and $[y_{\ell-1/2},y_{\ell+1/2}]$, respectively. With $\mathbb{S}_K^{(1)}$ in (79), a special 2D quadrature [51] satisfying (66) can be

constructed:

$$
\begin{aligned}
&\frac{1}{|K|}\int_K u(\mathbf{x})d\mathbf{x}\\
&=\sum_{\mu=1}^{N}\frac{\Delta x_i\widehat{\omega}_1\omega_\mu}{\Delta x_i+\Delta y_\ell}\left(u\big(x_i^{(\mu)},y_{\ell-\frac{1}{2}}\big)+u\big(x_i^{(\mu)},y_{\ell+\frac{1}{2}}\big)\right)\\
&+\sum_{\mu=1}^{N}\frac{\Delta y_\ell\widehat{\omega}_1\omega_\mu}{\Delta x_i+\Delta y_\ell}\left(u\big(x_{i-\frac{1}{2}},y_\ell^{(\mu)}\big)+u\big(x_{i+\frac{1}{2}},y_\ell^{(\mu)}\big)\right)\\
&+\sum_{\nu=2}^{L-1}\sum_{\mu=1}^{N}\frac{\widehat{\omega}_\nu\omega_\mu}{\Delta x_i+\Delta y_\ell}\left(\Delta x_i u\big(x_i^{(\mu)},\widehat{y}_\ell^{(\nu)}\big)+\Delta y_\ell u\big(\widehat{x}_i^{(\nu)},y_\ell^{(\mu)}\big)\right)\qquad\forall u\in\mathbb{P}^k(K),
\end{aligned}
\tag{80}
$$

where $\{\widehat{w}_\mu\}_{\mu=1}^{L}$ are the weights of the L-point Gauss–Lobatto quadrature. If labeling the bottom, right, top, and left edges of K as $\mathscr{E}_1$, $\mathscr{E}_2$, $\mathscr{E}_3$, and $\mathscr{E}_4$, respectively, then (80) implies, for $1\le\mu\le N$, that $\varpi_{\mathscr{E}_j}^{(\mu)}=\frac{\Delta x_i\widehat{\omega}_1\omega_\mu}{\Delta x_i+\Delta y_\ell},\ j=1,3;\ \varpi_{\mathscr{E}_j}^{(\mu)}=\frac{\Delta y_\ell\widehat{\omega}_1\omega_\mu}{\Delta x_i+\Delta y_\ell},\ j=2,4$. According to Theorem 4.5, the CFL condition (71) for our positivity-preserving DG schemes on Cartesian meshes is

$$
\Delta t\left[2\widetilde{\alpha}_K^F\frac{p_{\mathscr{E}_j,\mu}^{e,\star}}{p_{\mathscr{E}_j,\mu}^{e,\mathtt{int}(K)}}\left(\frac{1}{\Delta x_i}+\frac{1}{\Delta y_\ell}\right)+\widetilde{\alpha}_K^S\widehat{\omega}_1\right]\le\widehat{\omega}_1\quad\forall K\in\mathcal{T}_h,\ 1\le j\le 4,
\tag{81}
$$

where $\widehat{\omega}_1=\frac{1}{L(L-1)}$. Assume the mesh is regular and define $h=\max_{i,\ell}\{\Delta x_i,\Delta y_\ell\}$; then for smooth $p^e(\mathbf{x})$, it holds that

$$
\frac{p_{\mathscr{E}_j,\mu}^{e,\star}}{p_{\mathscr{E}_j,\mu}^{e,\mathtt{int}(K)}}=\frac{1}{2}+\frac{p_h^{e,\mathtt{ext}(K)}(\mathbf{x}_{\mathscr{E}_j}^{(\mu)})}{2p_h^{e,\mathtt{int}(K)}(\mathbf{x}_{\mathscr{E}_j}^{(\mu)})}=1+\mathcal{O}(h^{k+1}),
$$

whose effect in the CFL condition (81) can be ignored.

5. Numerical tests. This section presents several 1D and 2D examples to demonstrate the well-balanced and positivity-preserving properties of the proposed DG methods on uniform Cartesian meshes. For the sake of comparison, we will also show the numerical results of the traditional non-well-balanced (denoted as "non-WB") DG schemes with the straightforward source term discretization and the original HLLC flux. Unless otherwise stated, we use the explicit third-order SSP-RK time discretization (28) and the ideal equation of state (3) with $\gamma=1.4$, and the CFL numbers C_{cfl} for the third-order, fourth-order, and fifth-order DG methods are taken as 0.2, 0.12, and 0.1, respectively. In all the tests, the method is implemented by using C++ language with double precision.

5.1. Example 1: One-dimensional polytropic equilibrium. This test is used to investigate the performance of the proposed schemes near the polytropic equilibrium states [19]. Under the gravitational field $\phi(x)=gx$, the stationary hydrostatic solutions are

$$
\rho^e(x)=\left(\rho_0^{\gamma-1}-\frac{1}{K_0}\frac{\gamma-1}{\gamma}gx\right)^{\frac{1}{\gamma-1}},\qquad u^e(x)=0,\qquad p^e(x)=K_0\left(\rho^e(x)\right)^\gamma,
\tag{82}
$$

with $g=1$, $\gamma=5/3$, $\rho_0=p_0=1$, and $K_0=p_0/\rho_0^\gamma$ on a computational domain $[0,2]$.

We first use this example to check the well-balancedness of our DG methods. The initial data are taken as the stationary hydrostatic solutions (82). We simulate this problem up to $t = 4$ by using our third-order well-balanced DG scheme with different mesh points, and list the l^1-errors of numerical solutions in Table 1. These errors are evaluated between the numerical solutions and the projected stationary hydrostatic solutions. It is clearly observed that the numerical errors are all at the level of round-off error, which verify the desired well-balanced property.

TABLE 1
Example 1: l^1*-errors on different meshes of* M *uniform cells.*

M	Errors in ρ	Errors in m	Errors in E
50	1.0682e-14	1.0332e-14	4.4756e-16
100	3.6074e-14	4.5115e-14	6.5160e-15
200	5.2993e-14	4.9258e-14	7.8335e-15

Next, a small perturbation is imposed to the stationary hydrostatic state (82), so as to compare the performance of well-balanced and non-WB DG schemes in simulating the evolution of such small perturbation. More specifically, we add a periodic velocity perturbation

$$u(x,t) = A\sin(4\pi t),$$

with $A = 10^{-6}$, to the system on the left boundary $x = 0$. The solutions are computed until $t = 1.5$, before the waves propagate to the right boundary $x = 2$. Figure 1 displays the pressure perturbation and the velocity at $t = 1.5$, computed by the proposed third-order well-balanced DG scheme on a mesh of 100 uniform cells, against the reference solutions computed on a much-refined mesh of 1000 cells. For comparison, we also perform the third-order non-WB DG method and show its results in the same figure. As we can see, the results by the well-balanced DG method agree well with the reference ones, while the results by the non-WB DG method do not match the reference ones especially in the region where $x > 1.5$. This demonstrates that the well-balanced methods are advantageous and more accurate for resolving small amplitude perturbations to equilibrium states.

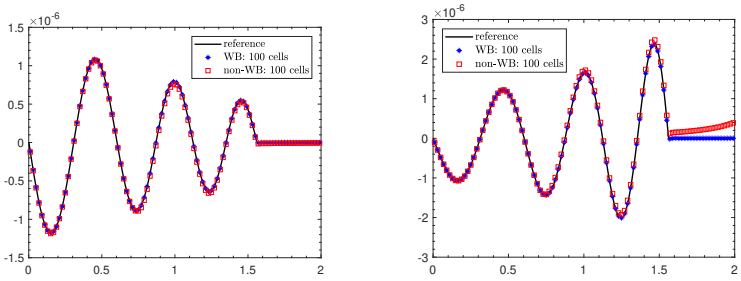


FIG. 1. *Example* 1: *Small amplitude waves with* $A = 10^{-6}$ *traveling up the polytropic hydrostatic atmosphere. The numerical solutions of the well-balanced method (denoted by "WB") and the non-WB method are obtained on the mesh of* 100 *uniform cells. The reference solutions are computed by the well-balanced method using* 1000 *mesh points. Left: Pressure perturbation. Right: Velocity.*

In the last test case of this example, we conduct the same simulation but with a large perturbation $A = 0.1$. We again evolve the simulation until $t = 1.5$. Because the discontinuities are formed in the final solution, the WENO limiter [30] is implemented right before the positivity-preserving limiting procedure with the aid of the local

characteristic decomposition within a few "trouble" cells detected adaptively. The numerical solutions by both the well-balanced and non-WB DG methods are shown in Figure 2, against the reference solutions. One can see that both DG methods produce satisfactory results. This agrees with the normal expectation that the well-balanced methods perform similarly as non-WB methods in capturing solutions far away from steady states.

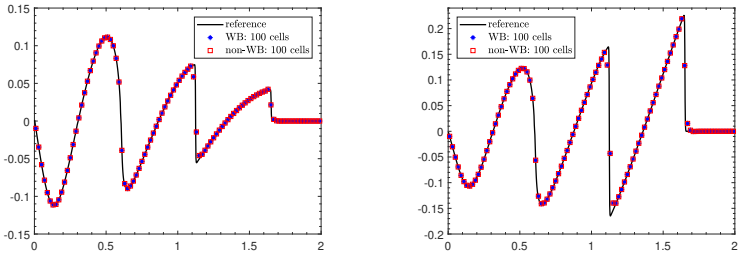


FIG. 2. *Same as Figure* 1 *except for large amplitude waves with* $A = 0.1$ *traveling up the polytropic hydrostatic atmosphere. Left: Pressure perturbation. Right: Velocity.*

5.2. Example 2: Rarefaction test with low density and low pressure. To demonstrate the positivity-preserving property, we consider an extreme rarefaction test under a quadratic gravitational potential $\phi(x) = x^2/2$ centered around $x = 0$. The computational domain is taken as $[-1, 1]$, and the initial state is the same as a Riemann problem in [52], given by

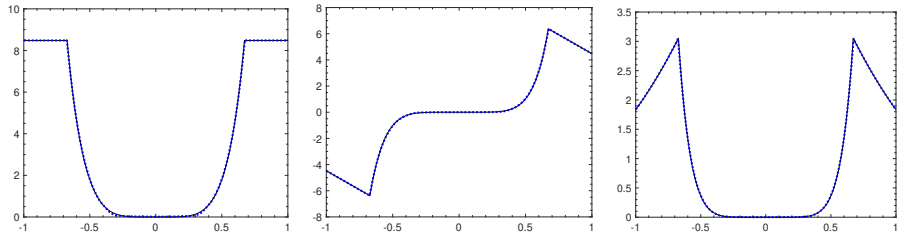
$$\rho(x,0) = 7, \qquad p(x,0) = 0.2, \qquad u(x,0) = \begin{cases} -1, & x < 0, \\ 1, & x > 0, \end{cases}$$

with outflow boundary conditions at $x = -1$ and $x = 1$. This problem involves extremely low density and pressure, so that the positivity-preserving limiter should be employed. The CFL number is set as 0.15, which is slightly smaller than $\widehat{\omega}_1 = \frac{1}{6}$. Figure 3 gives the numerical results at $t = 0.6$, obtained by our positivity-preserving third-order well-balanced DG scheme, on a mesh with 800 cells, compared with reference solutions obtained with much refined 128000 cells. It is seen that the low density and low pressure wave structures are well captured by the proposed method. During the whole simulation, our scheme exhibits good robustness. We observe that it is necessary to enforce the condition (40); otherwise the DG code will break down due to nonphysical solution.

5.3. Example 3: Leblanc problem in linear gravitational field. In this test, we consider an extension of the standard 1D Leblanc shock tube problem to the gravitational case with $\phi(x) = gx$ and $g = 1$. The initial condition of this problem is given by

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

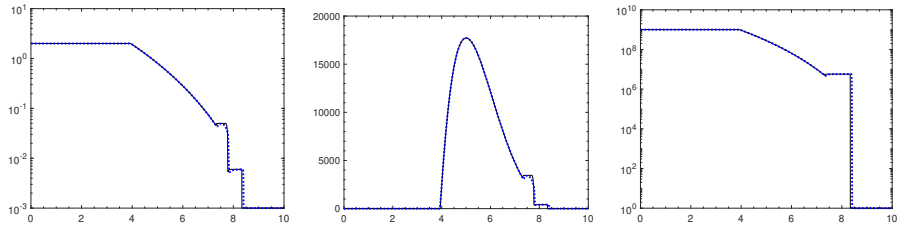
This problem is highly challenging due to the presence of the strong jumps in the initial density and pressure. The computational domain is taken as $[0, 10]$ with reflection boundary conditions at $x = 0$ and $x = 10$. To fully resolve the wave structure, a fine mesh is required for such a test. In the computations, the CFL number is taken as 0.15. As the exact solution contains strong discontinuities, the WENO limiter [30] is



(a) ρ (b) m (c) E

FIG. 3. *Example* 2: *Density, momentum, and energy for the rarefaction test at* $t = 0.6$ *obtained by the positivity-preserving well-balanced DG scheme with* 800 *cells (dotted lines) and* 128000 *cells (solid lines).*

implemented right before the positivity-preserving limiting procedure with the aid of the local characteristic decomposition within the adaptively detected "trouble" cells. Figure 4 displays our numerical results at $t = 0.00004$, obtained by the third-order positivity-preserving well-balanced DG scheme, on a mesh with 1600 cells, compared with reference solutions obtained with much refined 6400 cells. We see that the strong discontinuities are captured by the proposed method with high resolution.



(a) ρ (b) m (c) p

FIG. 4. *Example* 3: *The log plot of density (left), the velocity (middle), and the log plot of pressure (right) for the extended Leblanc problem at* $t = 0.00004$ *obtained by the positivity-preserving well-balanced DG scheme with* 1600 *cells (dotted lines) and* 6400 *cells (solid lines), respectively.*

5.4. Example 4: Two-dimensional accuracy test. In this example, we examine the accuracy of the proposed schemes on a 2D smooth problem [43] with a linear gravitational field $\phi_x = \phi_y = 1$ in the domain $\Omega = [0, 2]^2$. The exact solution takes the following form:

$$\rho(x, y, t) = 1 + 0.2\sin(\pi(x + y - t(u_0 + v_0))), \quad \mathbf{u}(x, y, t) = (u_0, v_0),$$
$$p(x, y, t) = p_0 + t(u_0 + v_0) - x - y + 0.2\cos(\pi(x + y - t(u_0 + v_0)))/\pi,$$

where the parameters are taken as $u_0 = v_0 = 1$ and $p_0 = 4.5$. The adiabatic index γ is taken as $5/3$. The domain Ω is divided into $M \times M$ uniform cells, and the boundary condition is specified by the exact solution on $\partial\Omega$. To match the accuracy of DG spatial discretization, we use (only in this accuracy test) the classical fourth-order explicit RK time discretization (cf. [16, p. 131]) in the fourth-order and fifth-order DG schemes, and $\Delta t = 0.1h^{\frac{5}{4}}/(2\widetilde{\alpha}_K^F)$ for the fifth-order DG scheme in order to match the temporal and spatial accuracy. Tables 2, 3, and 4, respectively, list the l^1-errors at $t = 0.1$ and the corresponding orders obtained by the proposed third-order, fourth-order, and fifth-order well-balanced DG schemes at different grid resolutions. The

results show that the expected convergence orders are achieved. Our modification of the numerical flux and the nontrivial source term approximation do not affect the accuracy of the DG methods.

TABLE 2
Example 4: l^1-errors at $t = 0.1$ in $\rho, \mathbf{m} = (m_1, m_2), E$, and corresponding convergence rates for the third-order ($\mathbb{P}^2$-based) well-balanced DG method at different grid resolutions.

Mesh	ρ		m_1		m_2		E	
	Error	Order	Error	Order	Error	Order	Error	Order
8×8	4.20e-3	–	4.29e-3	–	4.29e-3	–	4.67e-3	–
16×16	5.25e-4	3.00	5.42e-4	2.98	5.42e-4	2.98	5.76e-4	3.02
32×32	6.62e-5	2.99	6.86e-5	2.98	6.86e-5	2.98	7.28e-5	2.98
64×64	8.31e-6	2.99	8.61e-6	2.99	8.61e-6	2.99	9.17e-6	2.99
128×128	1.04e-6	3.00	1.08e-6	3.00	1.08e-6	3.00	1.15e-6	3.00
256×256	1.30e-7	3.00	1.35e-7	3.00	1.35e-7	3.00	1.44e-7	3.00
512×512	1.63e-8	3.00	1.69e-8	3.00	1.69e-8	3.00	1.80e-8	3.00

TABLE 3
Same as Table 2 except for our fourth-order accurate ($\mathbb{P}^3$-based) DG method.

Mesh	ρ		m_1		m_2		E	
	Error	Order	Error	Order	Error	Order	Error	Order
8×8	4.28e-4	–	4.39e-4	–	4.39e-4	–	4.81e-4	–
16×16	2.46e-5	4.12	2.59e-5	4.09	2.59e-5	4.09	2.78e-5	4.11
32×32	1.56e-6	3.98	1.65e-6	3.97	1.65e-6	3.97	1.74e-6	4.00
64×64	9.86e-8	3.98	1.05e-7	3.98	1.05e-7	3.98	1.09e-7	4.00
128×128	6.09e-9	4.02	6.48e-9	4.01	6.48e-9	4.01	6.76e-9	4.01
256×256	3.81e-10	4.00	4.06e-10	4.00	4.06e-10	4.00	4.23e-10	4.00
512×512	2.38e-11	4.00	2.54e-11	4.00	2.54e-11	4.00	2.65e-11	4.00

TABLE 4
Same as Table 2 except for our fifth-order accurate ($\mathbb{P}^4$-based) DG method.

Mesh	ρ		m_1		m_2		E	
	Error	Order	Error	Order	Error	Order	Error	Order
4×4	1.11e-3	–	1.13e-3	–	1.13e-3	–	1.14e-3	–
8×8	3.26e-5	5.09	3.34e-5	5.08	3.34e-5	5.08	3.57e-5	4.99
16×16	1.09e-6	4.91	1.12e-6	4.90	1.12e-6	4.90	1.16e-6	4.95
32×32	3.60e-8	4.91	3.72e-8	4.91	3.72e-8	4.91	3.77e-8	4.94
64×64	1.15e-9	4.96	1.20e-9	4.96	1.20e-9	4.96	1.21e-9	4.96
128×128	3.63e-11	4.99	3.79e-11	4.98	3.79e-11	4.98	3.84e-11	4.98
256×256	1.13e-12	5.00	1.19e-12	5.00	1.19e-12	5.00	1.21e-12	4.99

5.5. Example 5: Two-dimensional isothermal equilibrium. This example is used to demonstrate the well-balanced property and the capability of the proposed methods in capturing the small perturbation of a 2D isothermal equilibrium solution [43]. We consider a linear gravitational field with $\phi_x = \phi_y = g$ and take $g = 1$. The computational domain is taken as the unit square $[0, 1]^2$. The isothermal equilibrium state under consideration takes the following form:

(83)
$$\rho(x, y) = \rho_0 \exp\left(-\frac{\rho_0 g}{p_0}(x + y)\right), \quad \mathbf{u}(x, y) = \mathbf{0}, \quad p(x, y) = p_0 \exp\left(-\frac{\rho_0 g}{p_0}(x + y)\right),$$

with the parameters $\rho_0 = 1.21$ and $p_0 = 1$.

We first validate the well-balanced property of the proposed DG method. To this end, we take the initial data as the equilibrium solution (83) and conduct the simulation up to $t = 1$ on the three different uniform meshes. The l^1-errors in ρ, $\mathbf{m} = (m_1, m_2)$, and E are shown in Table 5. One can clearly see that the steady state solution is indeed maintained up to the round-off error, which confirms the well-balancedness of the proposed DG method.

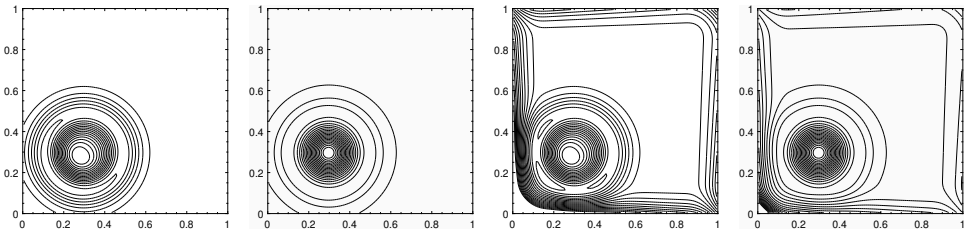
We then investigate the capability of the proposed well-balanced method in capturing small perturbations of the hydrostatic equilibrium. Initially, a small Gaussian hump perturbation centered at $(0.3, 0.3)$ is imposed in the pressure to the equilibrium solution (83) as follows:

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

where η is set as 0.001. We evolve the solution up to $t = 0.15$ on a mesh of 100×100 uniform cells with transmissive boundary conditions. The contour plots of the pressure perturbation and density perturbation are displayed in Figure 5, obtained via the well-balanced and the non-WB DG schemes, respectively. It is observed that the non-WB DG method cannot capture such small perturbation well on the relatively coarse mesh, while the well-balanced one can resolve it accurately.

TABLE 5
Example 5: l^1*-errors for the steady state solution in section* 5.5 *at different grid resolutions.*

Mesh	Errors in ρ	Errors in m_1	Errors in m_2	Errors in E
50×50	2.0615e-15	1.8301e-15	1.8527e-15	7.3921e-15
100×100	4.5131e-15	3.7141e-15	3.7649e-15	1.5167e-14
200×200	9.6832e-15	7.3142e-15	7.3067e-15	3.0940e-14



(a) WB method: Pressure perturbation (b) WB method: Density perturbation (c) Non-WB method: Pressure perturbation (d) Non-WB method: Density perturbation

FIG. 5. *Example* 5: *The contour plots of the pressure perturbation and the density perturbation of the hydrostatic solution at time* $t = 0.15$ *obtained by the third-order WB and non-WB DG schemes with* 100×100 *cells. Twenty equally spaced contour lines are displayed: from* -0.0003 *to* 0.0003 *for pressure perturbation; from* -0.001 *to* 0.0002 *for density perturbation.*

5.6. Example 6: Two-dimensional polytropic equilibrium. In this example, we verify the performance of the proposed methods on a 2D polytropic test case [19] arising from astrophysics. We consider a static adiabatic gaseous sphere, which is held together by self-gravitation and can be constructed from the hydrostatic equilibrium $\frac{\mathrm{d}p}{\mathrm{d}r} = -\rho\frac{\mathrm{d}\phi}{\mathrm{d}r}$, with $\gamma = 2$. One equilibrium solution of this model is given by

$$\rho(r) = \rho_c \frac{\sin(\alpha r)}{\alpha r}, \quad u(r) = 0, \quad v(r) = 0, \quad p(r) = K_0 \rho(r)^2, \tag{84}$$

under the gravitational field

$$\phi(r) = -2K_0\rho_c \frac{\sin(\alpha r)}{\alpha r}, \tag{85}$$

where $\alpha = \sqrt{2\pi g/K_0}$ with $K_0 = g = \rho_c = 1$, and $r := \sqrt{x^2+y^2}$ denotes the radial variable. The computational domain is taken as $[-0.5, 0.5]^2$.

We first demonstrate the well-balanced property of our DG scheme. The initial condition is taken as the equilibrium solution (84), which should be exactly preserved. The computation is performed until $t = 14.8$ on three different uniform meshes. The l^1-errors in the numerical solutions are presented in Table 6. It shows that the steady state is preserved up to the round-off error, as expected from the well-balancedness of the proposed method.

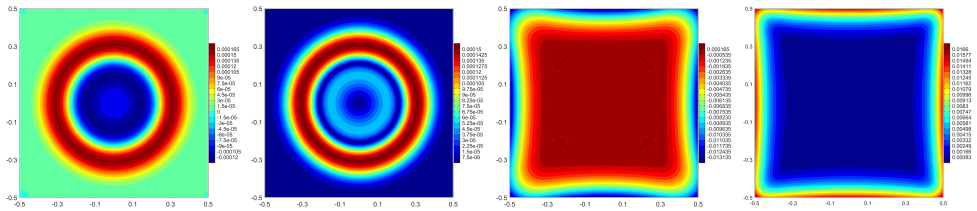
TABLE 6
Example 6: l^1*-errors for the steady state solution in section* 5.6 *at different grid resolutions.*

Mesh	Errors in ρ	Errors in m_1	Errors in m_2	Errors in E
50×50	3.9099e-14	1.0132e-13	1.0312e-13	8.5883e-15
100×100	7.4068e-14	1.8519e-13	1.8328e-13	1.7675e-14
200×200	1.4237e-13	3.3540e-13	3.3567e-13	3.5853e-14

We now impose a small perturbation to the initial pressure state

$$p(x, y, 0) = K_0\rho(r)^2 + \eta \exp(-100r^2),$$

and then compute the solution up to $t = 0.2$ on a mesh of 200×200 uniform cells with transmissive boundary conditions. Figure 6 shows the contour plots of the pressure perturbation and the velocity magnitude $\|\mathbf{u}\|$, obtained by using our well-balanced DG method and the non-WB DG method, respectively. We observe that the well-balanced DG scheme captures the small perturbation very well and preserves the axial symmetry, while the non-WB DG method cannot accurately resolve the small perturbation and maintain the axial symmetry on the relatively coarse mesh.



(a) WB method: Pressure perturbation (b) WB method: Velocity magnitude (c) Non-WB method: Pressure perturbation (d) Non-WB method: Velocity magnitude

FIG. 6. *Example* 6: *The contour plots of the pressure perturbation and the velocity magnitude at time* $t = 0.2$ *obtained by using our well-balanced and non-WB DG schemes on* 200×200 *cells.*

5.7. Example 7: Two-dimensional blast problem. To further verify the positivity-preserving property and the capability of the proposed DG method in resolving strong discontinuities, we consider a 2D blast problem under the gravitational field (85). The initial data is obtained by adding a huge jump to the pressure term of the equilibrium (84), and the initial pressure is

$$p(x, y, 0) = K_0\rho(r)^2 + \begin{cases} 100, & r < 0.1, \\ 0, & r \geq 0.1. \end{cases}$$

We set the parameters $K_0 = g = 1$ and $\gamma = 2$ as those in Example 6, and $\rho_c = 0.01$ so that low pressure and low density appear in the solution. This, along with the presence of the strong discontinuities, makes this test challenging.

Figure 7 displays the contour plots of ρ and $\log(p)$ at $t = 0.005$ computed by the positivity-preserving third-order well-balanced DG method with 400×400 uniform cells. We also show the plot of p along the line $y = x$, from which we can clearly observe a strong shock at $\sqrt{x^2+y^2}/\sqrt{2} \approx 0.28$. In this test, the CFL number of 0.15 is used, and the WENO limiter is implemented. We observe that the discontinuities are well captured with high resolution, and the proposed DG method preserves the positivity of density and pressure as well as the axisymmetric structure of the solution. In this extreme test, it is necessary to use the positivity-preserving limiting technique; otherwise we observe that the DG code would start to produce negative numerical pressure at $t \approx 0.00267$.

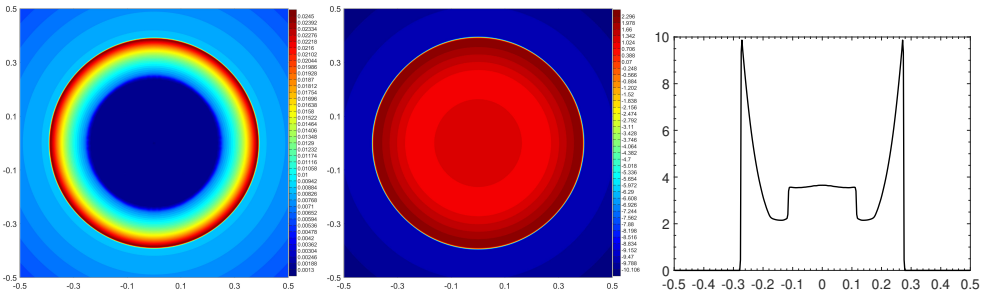


FIG. 7. *Example* 7: *The contour plots of the density* ρ *(left) and the pressure logarithm* $\log(p)$ *(middle) at* $t = 0.005$*, and the plot of* p *(right) along the line* $y = x$ *within the* scaled *interval* $[-0.5, 0.5]$*, obtained by the positivity-preserving well-balanced DG scheme with* 400×400 *cells.*

5.8. Example 8: Inertia-gravity wave. This is a benchmark test problem arising from atmospheric flows. The setup is adopted from [13, 12]. The computational domain is a $[0, 300000] \times [0, 10000]$ m^2 channel, with inviscid wall boundary conditions on the bottom and top boundaries, and periodic boundary conditions on the left and right boundaries. The gravitational field of this problem is linear with $\phi_x = 0$ and $\phi_y = g = 9.8$ m/s^2. Consider a uniformly stratified atmosphere with a constant velocity $\mathbf{u} = (20\text{ m/s}, 0\text{ m/s})$. The potential temperature and Exner pressure are, respectively, given by

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

where the Brunt–Väisälä frequency $\mathcal{N} = 0.01$/s, the reference temperature $T_0 = 300$ K at $y = 0$ m, and the gas constant $R = 287.058$ J/kg K. Initially, a small perturbation is added to the potential temperature:

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

where $\theta_c = 0.01$ K, $h_c = 10000$ m, $x_c = 100000$ m, and $a_c = 5000$ m. The pressure and density are computed by Θ and Π via

$$p = p_0 \Pi^{\frac{\gamma}{\gamma-1}}, \qquad \rho = \frac{p_0}{R\Theta}\Pi^{\frac{1}{\gamma-1}}, \tag{86}$$

with the reference pressure $p_0 = 10^5$ N/m^2 at $y = 0$ m.

We simulate this problem up to $t = 3000$ s on a mesh of 1200×40 uniform cells, by using the proposed fourth-order accurate ($\mathbb{P}^3$-based) and fifth-order accurate ($\mathbb{P}^4$-based) DG methods, respectively. The left of Figure 8 shows the contours of the potential temperature perturbation $\Delta\Theta(x, y, t = 3000\text{ s})$ for the solutions obtained by our methods. (The specified contour values are the same as in [13].) The right side of Figure 8 displays the profiles of $\Delta\Theta(x, y = 5000\text{ m}, t = 3000\text{ s})$. We observe that the evolution of potential temperature perturbation is correctly resolved and that the solution structures agree well with those presented in [13, 12].

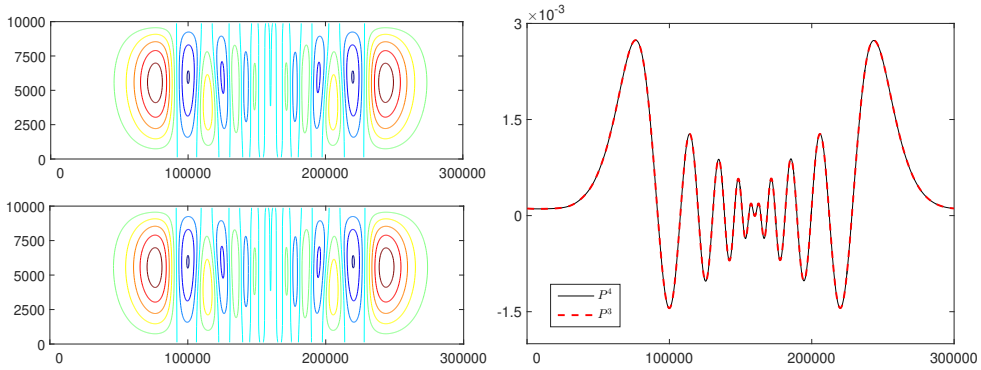


FIG. 8. *Example* 8: *Potential temperature perturbation* $\Delta\Theta$ *at* $t = 3000$ s. *Left: The contours of* $\Delta\Theta$ *obtained with our fourth-order (top-left) and fifth-order (bottom-left) well-balanced DG schemes on* 200×200 *cells (*10 *equally spaced contour lines from* -0.0015 *to* 0.003*). Right: Profiles of* $\Delta\Theta$ *along the line* $y = 5000$ m.

5.9. Example 9: Rising thermal bubble. The last example, also a benchmark test problem for atmospheric flows, simulates the dynamics of a warm bubble. The setup is the same as in [13, 12]. The computational domain is $[0, 1000] \times [0, 1000]$ m^2, with inviscid wall boundary conditions. As in Example 8, the gravitational field is linear with $\phi_x = 0$ and $\phi_y = g = 9.8$ m/s^2. Consider a stratified atmosphere, with zero velocity $\mathbf{u} = \mathbf{0}$, a constant potential temperature $\Theta = T_0 = 300$ K, and Exner pressure $\Pi = 1 - \frac{(\gamma-1)gy}{\gamma R T_0}$, where $R = 287.058$ J/kg K is the gas constant. Initially, the warm bubble is added as a potential temperature perturbation to the hydrostatic balance,

$$\Delta\Theta(x, y, t = 0) = \begin{cases} 0, & r > r_c, \\ \frac{\theta_c}{2}\left(1 + \cos(\pi r / r_c)\right), & r \le r_c, \end{cases} \qquad r = \sqrt{(x - x_c)^2 + (y - y_c)^2},$$

where $\theta_c = 0.5$ K, $(x_c, y_c) = (500, 350)$ m, and $r_c = 250$ m. The pressure and density are computed by Θ and Π via the formulas in (86), with the reference pressure $p_0 = 10^5$ N/m^2. Figure 9 shows the evolution of potential temperature perturbation $\Delta\Theta$ obtained by the proposed fifth-order accurate DG method on two different meshes of 100×100 cells (10 m resolution) and 200×200 cells (5 m resolution), respectively. It is seen that the initial circular bubble is deformed to a mushroom-like cloud. The flow structures are well resolved, and the solutions are in good agreement with those presented in [13, 12]. It can be observed that our solutions are comparable with the one reported in [12] using a fifth-order WENO scheme with a fine resolution of 2.5 m.

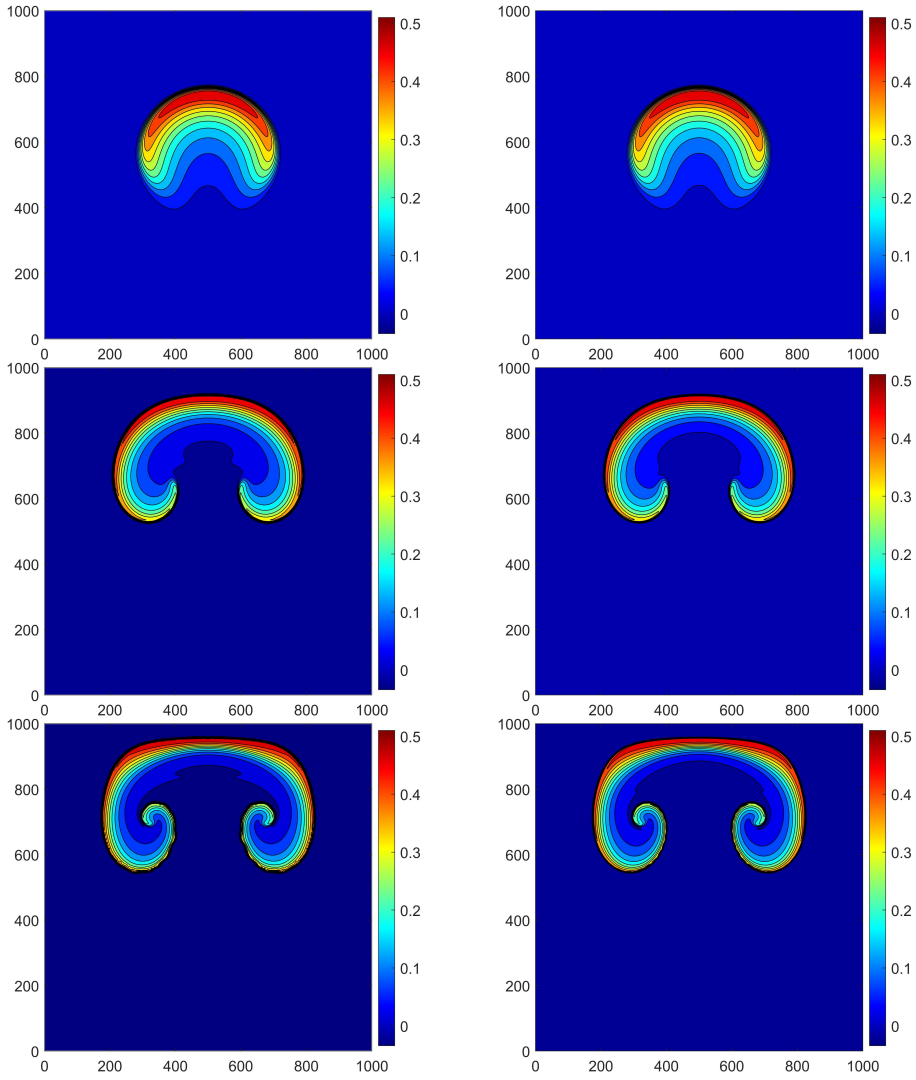


FIG. 9. *Example* 9: *The contour plots of potential temperature perturbation* $\Delta\Theta$ *at* $t = 400$ s *(top),* $t = 600$ s *(middle), and* $t = 700$ s *(bottom), respectively, computed by our fifth-order DG method. Ten equally spaced contour lines are displayed. Left: Mesh of* 100×100 *cells. Right: Mesh of* 200×200 *cells.*

6. Conclusion. In this paper, we constructed high-order accurate positivity-preserving well-balanced DG methods for the compressible Euler equations with gravitation. A novel well-balanced spatial discretization was specially designed with suitable source term treatments and a properly modified HLLC flux, while the desired positivity property was also achieved in the discretization at the same time. Based on some technical decompositions as well as several key properties of the admissible states and HLLC flux, rigorous positivity-preserving analyses were carried out in theory. It was proven that the resulting well-balanced DG schemes with SSP time discretization satisfy a weak positivity property, which implies that a simple existing limiter can effectively enforce the positivity-preserving property without losing conservation and

high-order accuracy. Extensive 1D and 2D numerical tests were provided to demonstrate the accuracy, well-balancedness, positivity preservation, and high resolution of the proposed schemes. It is worth noting that the proposed numerical framework is also readily applicable for designing positivity-preserving well-balanced high-order accurate finite volume methods.

REFERENCES

[1] E. Audusse, F. Bouchut, M.-O. Bristeau, R. Klein, and B. Perthame, *A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows*, SIAM J. Sci. Comput., 25 (2004), pp. 2050–2065, https://doi.org/10.1137/S1064827503431090.

[2] 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 (1997), pp. 1553–1570, https://doi.org/10.1137/S1064827593260140.

[3] J. P. Berberich, R. Käppeli, P. Chandrashekar, and C. Klingenberg, *High Order Discretely Well-balanced Methods for Arbitrary Hydrostatic Atmospheres*, preprint, https://arxiv.org/abs/2005.01811v3, 2020.

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

[5] N. Botta, R. Klein, S. Langenberg, and S. Lützenkirchen, *Well-balanced finite volume methods for nearly hydrostatic flows*, J. Comput. Phys., 196 (2004), pp. 539–565.

[6] P. Chandrashekar and C. Klingenberg, *A second order well-balanced finite volume scheme for Euler equations with gravity*, SIAM J. Sci. Comput., 37 (2015), pp. B382–B402, https://doi.org/10.1137/140984373.

[7] P. Chandrashekar and M. Zenk, *Well-balanced nodal discontinuous Galerkin method for Euler equations with gravity*, J. Sci. Comput., 71 (2017), pp. 1062–1093.

[8] A. Chertock, S. Cui, A. Kurganov, Ş. N. Özcan, and E. Tadmor, *Well-balanced schemes for the Euler equations with gravitation: Conservative formulation using global fluxes*, J. Comput. Phys., 358 (2018), pp. 36–52.

[9] B. Cockburn and C.-W. Shu, *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws.* II. *General framework*, Math. Comp., 52 (1989), pp. 411–435.

[10] J. Du and C.-W. Shu, *Positivity-preserving high-order schemes for conservation laws on arbitrarily distributed point clouds with a simple WENO limiter*, Internat. J. Numer. Anal. Model., 15 (2018), pp. 1–25.

[11] E. Franck and L. S. Mendoza, *Finite volume scheme with local high order discretization of the hydrostatic equilibrium for the Euler equations with external forces*, J. Sci. Comput., 69 (2016), pp. 314–354.

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

[13] F. X. Giraldo and M. Restelli, *A study of spectral element and discontinuous Galerkin methods for the Navier–Stokes equations in nonhydrostatic mesoscale atmospheric modeling: Equation sets and test cases*, J. Comput. Phys., 227 (2008), pp. 3849–3877.

[14] S. Gottlieb, C.-W. Shu, and E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Rev., 43 (2001), pp. 89–112, https://doi.org/10.1137/S003614450036757X.

[15] J. M. Greenberg and A. Y. LeRoux, *A well-balanced scheme for the numerical processing of source terms in hyperbolic equations*, SIAM J. Numer. Anal., 33 (1996), pp. 1–16, https://doi.org/10.1137/0733001.

[16] D. F. Griffiths and D. J. Higham, *Numerical Methods for Ordinary Differential Equations: Initial Value Problems*, Springer-Verlag, London, 2010.

[17] L. Grosheintz-Laval and R. Käppeli, *High-order well-balanced finite volume schemes for the Euler equations with gravitation*, J. Comput. Phys., 378 (2019), pp. 324–343.

[18] X. Y. Hu, N. A. Adams, and C.-W. Shu, *Positivity-preserving method for high-order conservative schemes solving compressible Euler equations*, J. Comput. Phys., 242 (2013), pp. 169–180.

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

[20] R. Käppeli and S. Mishra, *A well-balanced finite volume scheme for the Euler equations with gravitation: The exact preservation of hydrostatic equilibrium with arbitrary entropy*

stratification, Astron. Astrophys., 587 (2016), A94.
[21] 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 (2019), pp. A695–A721, https://doi.org/10.1137/18M1196704.
[22] A. Kurganov and G. Petrova, *A second-order well-balanced positivity preserving central-upwind scheme for the Saint–Venant system*, Commun. Math. Sci., 5 (2007), pp. 133–160.
[23] R. J. LeVeque, *Balancing source terms and flux gradients on high-resolution Godunov methods: The quasi-steady wave-propagation algorithm*, J. Comput. Phys., 146 (1998), pp. 346–365.
[24] R. J. LeVeque and D. S. Bale, *Wave propagation methods for conservation laws with source terms*, in Proceedings of the 7th International Conference on Hyperbolic Problems, Hyperbolic Problems: Theory, Numerics, Applications, Vol. II (Zürich, 1998), Internat. Ser. Numer. Math. 130, Birkhäuser, Basel, 1998, pp. 609–618.
[25] G. Li and Y. Xing, *High order finite volume WENO schemes for the Euler equations under gravitational fields*, J. Comput. Phys., 316 (2016), pp. 145–163.
[26] G. Li and Y. Xing, *Well-balanced discontinuous Galerkin methods for the Euler equations under gravitational fields*, J. Sci. Comput., 67 (2016), pp. 493–513.
[27] G. Li and Y. Xing, *Well-balanced discontinuous Galerkin methods with hydrostatic reconstruction for the Euler equations with gravitation*, J. Comput. Phys., 352 (2018), pp. 445–462.
[28] G. Li and Y. Xing, *Well-balanced finite difference weighted essentially non-oscillatory schemes for the Euler equations with static gravitational fields*, Comput. Math. Appl., 75 (2018), pp. 2071–2085.
[29] J. Luo, K. Xu, and N. Liu, *A well-balanced symplecticity-preserving gas-kinetic scheme for hydrodynamic equations under gravitational field*, SIAM J. Sci. Comput., 33 (2011), pp. 2356–2381, https://doi.org/10.1137/100803699.
[30] J. Qiu and C.-W. Shu, *Runge–Kutta discontinuous Galerkin method using WENO limiters*, SIAM J. Sci. Comput., 26 (2005), pp. 907–929, https://doi.org/10.1137/S1064827503425298.
[31] C.-W. Shu, *Bound-preserving high-order schemes for hyperbolic equations: Survey and recent developments*, in Theory, Numerics and Applications of Hyperbolic Problems II, C. Klingenberg and M. Westdickenberg, eds., Springer, Cham, 2018, pp. 591–603.
[32] A. Thomann, M. Zenk, and C. Klingenberg, *A second-order positivity-preserving well-balanced finite volume scheme for Euler equations with gravity for arbitrary hydrostatic equilibria*, Int. J. Numer. Methods Fluids, 89 (2019), pp. 465–482.
[33] E. F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*, Springer-Verlag, Berlin, Heidelberg, 2013.
[34] D. Varma and P. Chandrashekar, *A second-order, discretely well-balanced finite volume scheme for Euler equations with gravity*, Comput. Fluids, 181 (2019), pp. 292–313.
[35] M. Veiga, D. A. Velasco-Romero, R. Abgrall, and R. Teyssier, *Capturing near-equilibrium solutions: A comparison between high-order discontinuous Galerkin methods and well-balanced schemes*, Commun. Comput. Phys., 26 (2019), pp. 1–34.
[36] C. Wang, X. Zhang, C.-W. Shu, and J. Ning, *Robust high order discontinuous Galerkin schemes for two-dimensional gaseous detonations*, J. Comput. Phys., 231 (2012), pp. 653–665.
[37] K. Wu, *Design of provably physical-constraint-preserving methods for general relativistic hydrodynamics*, Phys. Rev. D, 95 (2017), 103001.
[38] K. Wu, *Positivity-preserving analysis of numerical schemes for ideal magnetohydrodynamics*, SIAM J. Numer. Anal., 56 (2018), pp. 2124–2147, https://doi.org/10.1137/18M1168017.
[39] K. Wu and C.-W. Shu, *Provably positive high-order schemes for ideal magnetohydrodynamics: Analysis on general meshes*, Numer. Math., 142 (2019), pp. 995–1047.
[40] K. Wu and H. Tang, *High-order accurate physical-constraints-preserving finite difference WENO schemes for special relativistic hydrodynamics*, J. Comput. Phys., 298 (2015), pp. 539–564.
[41] Y. Xing and C.-W. Shu, *High order finite difference WENO schemes with the exact conservation property for the shallow water equations*, J. Comput. Phys., 208 (2005), pp. 206–227.
[42] Y. Xing and C.-W. Shu, *High-order finite volume WENO schemes for the shallow water equations with dry states*, Adv. Water. Resour., 34 (2011), pp. 1026–1038.
[43] Y. Xing and C.-W. Shu, *High order well-balanced WENO scheme for the gas dynamics equations under gravitational fields*, J. Sci. Comput., 54 (2013), pp. 645–662.
[44] Y. Xing and C.-W. Shu, *A survey of high order schemes for the shallow water equations*, J. Math. Study, 47 (2014), pp. 221–249.
[45] Y. Xing and X. Zhang, *Positivity-preserving well-balanced discontinuous Galerkin methods*

for the shallow water equations on unstructured triangular meshes, J. Sci. Comput., 57 (2013), pp. 19–41.

[46] Y. Xing, X. Zhang, and C.-W. Shu, *Positivity-preserving high order well-balanced discontinuous Galerkin methods for the shallow water equations*, Adv. Water Resour., 33 (2010), pp. 1476–1493.

[47] K. Xu, *A well-balanced gas-kinetic scheme for the shallow-water equations with source terms*, J. Comput. Phys., 178 (2002), pp. 533–562.

[48] K. Xu, J. Luo, and S. Chen, *A well-balanced kinetic scheme for gas dynamic equations under gravitational field*, Adv. Appl. Math. Mech., 2 (2010), pp. 200–210.

[49] Z. Xu, *Parametrized maximum principle preserving flux limiters for high order schemes solving hyperbolic conservation laws: One-dimensional scalar problem*, Math. Comp., 83 (2014), pp. 2213–2238.

[50] X. Zhang, *On positivity-preserving high order discontinuous Galerkin schemes for compressible Navier-Stokes equations*, J. Comput. Phys., 328 (2017), pp. 301–343.

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

[52] 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 (2010), pp. 8918–8934.

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

[54] X. Zhang, Y. Xia, and C.-W. Shu, *Maximum-principle-satisfying and positivity-preserving high order discontinuous Galerkin schemes for conservation laws on triangular meshes*, J. Sci. Comput., 50 (2012), pp. 29–62.