SIAM J. NUMER. ANAL.
Vol. 56, No. 4, pp. 2124–2147

POSITIVITY-PRESERVING ANALYSIS OF NUMERICAL SCHEMES FOR IDEAL MAGNETOHYDRODYNAMICS[*]

KAILIANG WU[†]

Abstract. Numerical schemes *provably* preserving the positivity of density and pressure are highly desirable for ideal magnetohydrodynamics (MHD), but the rigorous positivity-preserving (PP) analysis remains challenging. The difficulties mainly arise from the intrinsic complexity of the MHD equations as well as the indeterminate relation between the PP property and the divergence-free condition on the magnetic field. This paper presents the first rigorous PP analysis of conservative schemes with the Lax–Friedrichs (LF) flux for 1D and multidimensional ideal MHD. The significant innovation is the discovery of the theoretical connection between the PP property and a discrete divergence-free (DDF) condition. This connection is established through the generalized LF splitting properties, which are alternatives to the usually expected LF splitting property that does not hold for ideal MHD. The generalized LF splitting properties involve a number of admissible states strongly coupled by the DDF condition, making their derivation very difficult. We derive these properties via a novel equivalent form of the admissible state set and an important inequality, which is skillfully constructed by technical estimates. Rigorous PP analysis is then presented for finite volume and discontinuous Galerkin schemes with the LF flux on uniform Cartesian meshes. In the 1D case, the PP property is proved for the first-order scheme with proper numerical viscosity, and also for arbitrarily high-order schemes under conditions accessible by a PP limiter. In the 2D case, we show that the DDF condition is necessary and crucial for achieving the PP property. It is observed that even slightly violating the proposed DDF condition may cause failure to preserve the positivity of pressure. We prove that the 2D LF type scheme with proper numerical viscosity preserves both the positivity and the DDF condition. Sufficient conditions are derived for 2D PP high-order schemes, and extension to 3D is discussed. Numerical examples provided in the supplementary material further confirm the theoretical findings.

Key words. compressible magnetohydrodynamics, positivity-preserving, admissible states, discrete divergence-free condition, generalized Lax–Friedrichs splitting, hyperbolic conservation laws

AMS subject classifications. 65M60, 65M08, 65M12, 35L65, 76W05

DOI. 10.1137/18M1168017

1. Introduction.

Magnetohydrodynamics (MHD) play an important role in many fields including astrophysics, space physics, and plasma physics. The d-dimensional ideal compressible MHD equations can be written as

$$\frac{\partial \mathbf{U}}{\partial t} + \sum_{i=1}^{d} \frac{\partial \mathbf{F}_i(\mathbf{U})}{\partial x_i} = \mathbf{0}, \tag{1}$$

together with the divergence-free condition on the magnetic field $\mathbf{B} = (B_1, B_2, B_3)$,

$$\sum_{i=1}^{d} \frac{\partial B_i}{\partial x_i} = 0, \tag{2}$$

where $d = 1$, 2, or 3. In (1), the conservative vector $\mathbf{U} = (\rho, \rho\mathbf{v}, \mathbf{B}, E)^\top$, and $\mathbf{F}_i(\mathbf{U})$ denotes the flux in the x_i-direction, $i = 1, \dots, d$, defined by

$$\mathbf{F}_i(\mathbf{U}) = \Big(\rho v_i,\ \rho v_i \mathbf{v} - B_i \mathbf{B} + p_{tot}\mathbf{e}_i,\ v_i\mathbf{B} - B_i\mathbf{v},\ v_i(E + p_{tot}) - B_i(\mathbf{v}\cdot\mathbf{B})\Big)^\top.$$

[*] Received by the editors January 31, 2018; accepted for publication (in revised form) May 15, 2018; published electronically July 17, 2018.
http://www.siam.org/journals/sinum/56-4/M116801.html

[†] Department of Mathematics, The Ohio State University, Columbus, OH 43210 (wu.3423@osu.edu).

Here ρ is the density, the vector $\mathbf{v} = (v_1, v_2, v_3)$ denotes the fluid velocity, $p_{\rm tot}$ is the total pressure consisting of the gas pressure p and magnetic pressure $p_m = \frac{|\mathbf{B}|^2}{2}$, the vector $\mathbf{e}_i$ represents the ith row of the unit matrix of size 3, and $E = \rho e + \frac{1}{2}\left(\rho|\mathbf{v}|^2 + |\mathbf{B}|^2\right)$ is the total energy consisting of thermal, kinetic, and magnetic energies with e denoting the specific internal energy. The equation of state (EOS) is needed to close the system (1)–(2). For ideal gases it is given by

$$p = (\gamma - 1)\rho e, \tag{3}$$

where the adiabatic index $\gamma > 1$. Although (3) is widely used, there are situations where it is more appropriate to use other EOSs. A general EOS can be expressed as

$$p = p(\rho, e), \tag{4}$$

which is assumed to satisfy the following condition:

$$\text{if} \quad \rho > 0, \quad \text{then} \quad e > 0 \;\Leftrightarrow\; p(\rho, e) > 0. \tag{5}$$

Such a condition is reasonable, holds for the ideal EOS (3), and was also used in [52].

Since (1) involves strong nonlinearity, its analytic treatment is very difficult. Numerical simulation is a primary approach to explore the physical mechanisms in MHD. In the past few decades, the numerical study of MHD has attracted much attention, and various numerical schemes have been developed for (1). Besides the standard difficulty in solving nonlinear hyperbolic conservation laws, an additional numerical challenge for the MHD system comes from the divergence-free condition (2). Although (2) holds for the exact solution as long as it does initially, it cannot be easily preserved by a numerical scheme (for $d \geq 2$). Numerical evidence and some analysis in the literature indicate that negligence in dealing with condition (2) can lead to numerical instabilities or nonphysical features in the computed solutions; see, e.g., [9, 16, 6, 36, 15, 23]. Up to now, many numerical techniques have been proposed to control the divergence error of numerical magnetic field. They include but are not limited to the eight-wave methods [31, 10], the projection method [9], the hyperbolic divergence cleaning methods [15], the locally divergence-free methods [23, 47], and the constrained transport method [16] and its many variants [33, 6, 28, 2, 29, 35, 34, 32, 1, 26, 3, 25, 24, 14]. The readers are also referred to an early survey in [36].

Another numerical challenge in the simulation of MHD is preserving the positivity of density ρ and pressure p. In physics, these two quantities are nonnegative. Numerically their positivity is very critical, but not always satisfied by numerical solutions. In fact, once the negative density or pressure is obtained in the simulations, the discrete problem will become ill-posed due to the loss of hyperbolicity, causing the breakdown of the simulation codes. However, most of the existing MHD schemes are generally not positivity-preserving (PP) and thus may suffer from a large risk of failure when simulating MHD problems with strong discontinuity, low density, low pressure, or low plasma-beta. Several efforts have been made to reduce this risk. Balsara and Spicer [5] proposed a strategy to maintain positive pressure by switching the Riemann solvers for different wave situations. Janhunen [22] designed a new 1D Riemann solver for the modified MHD system and claimed its PP property by numerical experiments. Waagan [37] designed a positive linear reconstruction for the second-order MUSCL-Hancock scheme and conducted some 1D analysis based on the presumed PP property of the first-order scheme. From a relaxation system, Bouchut, Klingenberg, and Waagan [7, 8] derived a multiwave approximate Riemann solver for

1D ideal MHD and deduced sufficient conditions for the solver to satisfy discrete entropy inequalities and the PP property. Recent years have witnessed some significant advances in developing bound-preserving high-order schemes for hyperbolic systems (e.g., [49, 50, 51, 21, 45, 27, 40, 30, 42, 46, 48]). High-order limiting techniques were well developed in [4, 11] for the finite volume or discontinuous Galerkin (DG) methods of MHD to enforce the admissibility[1] of the reconstructed or DG polynomial solutions at certain nodal points. These techniques are based on a presumed proposition that the cell-averaged solutions computed by those schemes are always admissible. Such a proposition has not yet been rigorously proved for those methods, although it could be deduced for the 1D schemes in [11] under some assumptions (see Remark 2.12 of the present paper). With the presumed PP property of the Lax–Friedrichs (LF) scheme, Christlieb et al. [13, 12] developed PP high-order finite difference methods for (1) by extending the parametrized flux limiters [45, 44].

It was demonstrated numerically that the above PP treatments could enhance the robustness of MHD codes. However, as mentioned in [13], *there was no rigorous proof to genuinely and completely show the PP property of those or any other schemes for* (1) *in the multidimensional cases. Even for the simplest first-order schemes, such as the LF scheme, the PP property is still unclear in theory. Moreover, it is also unanswered theoretically whether the divergence-free condition* (2) *is connected with the PP property of schemes for* (1). *Therefore, it is significant to explore provably PP schemes for* (1) *and develop related theories for rigorous PP analysis.*

The aim of this paper is to carry out a rigorous PP analysis of conservative finite volume and DG schemes with the LF flux for the 1D and multidimensional ideal MHD system (1). Such an analysis is extremely nontrivial and technical. The challenges mainly come from the intrinsic complexity of the system (1)–(2), as well as the unclear relation between the PP property and the divergence-free condition on the magnetic field. Fortunately, we find an important novel starting point of the analysis based on an equivalent form of the admissible state set. This form helps us to successfully derive the generalized LF splitting properties, which couple a discrete divergence-free (DDF) condition for the magnetic field with the convex combination of some LF splitting terms. These properties imply a theoretical connection between the PP property and the proposed DDF condition. As the generalized LF splitting properties involve a number of strongly coupled states, their discovery and proofs are extremely technical. With the aid of these properties, we present the rigorous PP analysis for finite volume and DG schemes on uniform Cartesian meshes. Meanwhile, our analysis also reveals that the DDF condition is necessary and crucial for achieving the PP property. This finding is consistent with some existing numerical evidence (violating the divergence-free condition may more easily cause negative pressure; see, e.g., [9, 2, 32, 4]) as well as our previous work on the relativistic MHD [41]. Without considering the relativistic effect, the system (1) yields unboundedness of velocities and poses difficulties essentially different from the relativistic case. It is also worth mentioning that, as will be shown, the 1D LF scheme is not always PP for piecewise constant B_1, making some existing techniques [49] for PP analysis inapplicable in the multidimensional ideal MHD case. Contrary to the usual expectation, we also find that the 1D LF scheme with a standard numerical viscosity parameter is not always PP, no matter how small the CFL number is. A proper viscosity parameter should be estimated, introducing additional difficulties into the analysis. Note that, for the

[1] Throughout this paper, the admissibility of a solution or state $\mathbf{U}$ means that the density and pressure corresponding to the conservative vector $\mathbf{U}$ are both positive; see Definition 2.1.

incompressible flow system in the vorticity-stream function formulation, there is also a divergence-free condition (but) on fluid velocity, i.e., the incompressibility condition, which is crucial in designing schemes that satisfy the maximum principle of vorticity; see, e.g., [49]. An important difference in our MHD case is that our divergence-free quantity (the magnetic field) is also nonlinearly related to defining the concerned positive quantity—the internal energy or pressure; see (6).

The paper is organized as follows. Section 2 gives several important properties of the admissible states for the PP analysis. Sections 3 and 4 respectively study 1D and 2D PP schemes. Numerical verifications and the 3D extension are given in the supplementary material. Section 5 concludes the paper with several remarks.

2. Admissible states.

Under the condition (5), it is natural to define the set of admissible states $\mathbf{U}$ of the ideal MHD as follows.

DEFINITION 2.1. *The set of admissible states of the ideal MHD is defined by*

$$\mathcal{G} = \left\{ \mathbf{U} = (\rho, \mathbf{m}, \mathbf{B}, E)^\top \;\Big|\; \rho > 0,\ \mathcal{E}(\mathbf{U}) := E - \frac{1}{2}\left(\frac{|\mathbf{m}|^2}{\rho} + |\mathbf{B}|^2 \right) > 0 \right\}, \tag{6}$$

where $\mathcal{E}(\mathbf{U}) = \rho e$ denotes the internal energy.

Given that the initial data are admissible, a scheme is defined to be PP if the numerical solutions always stay in the set $\mathcal{G}$. One can see from (6) that it is difficult to numerically preserve the positivity of $\mathcal{E}$, whose computation nonlinearly involves all the conservative variables. In most numerical methods, the conservative quantities are themselves evolved according to their own conservation laws, which are seemingly unrelated to, and numerically do not necessarily guarantee, the positivity of the computed $\mathcal{E}$. In theory, it is indeed a challenge to make an a priori judgment on whether a scheme is always PP under all circumstances or not.

2.1. Basic properties.

To overcome the difficulties arising from the nonlinearity of the function $\mathcal{E}(\mathbf{U})$, we propose the following equivalent definition of $\mathcal{G}$.

LEMMA 2.2 (equivalent definition). *The admissible state set $\mathcal{G}$ is equivalent to*

$$\mathcal{G}_* = \left\{ \mathbf{U} = (\rho, \mathbf{m}, \mathbf{B}, E)^\top \;\Big|\; \rho > 0, \quad \mathbf{U} \cdot \mathbf{n}^* + \frac{|\mathbf{B}^*|^2}{2} > 0\ \forall\ \mathbf{v}^*, \mathbf{B}^* \in \mathbb{R}^3 \right\}, \tag{7}$$

where

$$\mathbf{n}^* = \left(\frac{|\mathbf{v}^*|^2}{2},\ -\mathbf{v}^*,\ -\mathbf{B}^*,\ 1 \right)^\top.$$

Proof. If $\mathbf{U} \in \mathcal{G}$, then $\rho > 0$, and for any $\mathbf{v}^*, \mathbf{B}^* \in \mathbb{R}^3$,

$$\mathbf{U} \cdot \mathbf{n}^* + \frac{|\mathbf{B}^*|^2}{2} = \frac{\rho}{2}\left|\rho^{-1}\mathbf{m} - \mathbf{v}^*\right|^2 + \frac{|\mathbf{B} - \mathbf{B}^*|^2}{2} + \mathcal{E}(\mathbf{U}) \ge \mathcal{E}(\mathbf{U}) > 0,$$

that is, $\mathbf{U} \in \mathcal{G}_*$. Hence $\mathcal{G} \subset \mathcal{G}_*$. On the other hand, if $\mathbf{U} \in \mathcal{G}_*$, then $\rho > 0$, and taking $\mathbf{v}^* = \rho^{-1}\mathbf{m}$ and $\mathbf{B}^* = \mathbf{B}$ gives $0 < \mathbf{U} \cdot \mathbf{n}^* + |\mathbf{B}^*|^2/2 = \mathcal{E}(\mathbf{U})$. This means $\mathbf{U} \in \mathcal{G}$. Therefore, $\mathcal{G}_* \subset \mathcal{G}$. In conclusion, $\mathcal{G} = \mathcal{G}_*$. □

The two constraints in (7) *are both linear with respect to* $\mathbf{U}$*, making it more effective to analytically verify the PP property of numerical schemes for ideal MHD.*

The convexity of admissible state set is very useful in bound-preserving analysis, because it can help reduce the complexity of the analysis if the schemes can be rewritten into certain convex combinations; see, e.g., [50, 52, 38]. For the ideal MHD, the convexity of $\mathcal{G}_*$ or $\mathcal{G}$ can be easily shown by definition, and the proof is omitted here.

Lemma 2.3 (convexity). *The set $\mathcal{G}_*$ is convex. Moreover, $\lambda \mathbf{U}_1 + (1-\lambda)\mathbf{U}_0 \in \mathcal{G}_*$ for any $\mathbf{U}_1 \in \mathcal{G}_*, \mathbf{U}_0 \in \overline{\mathcal{G}}_*$, and $\lambda \in (0,1]$, where $\overline{\mathcal{G}}_*$ is the closure of $\mathcal{G}_*$.*

We also have the following orthogonal invariance, which can be verified directly.

Lemma 2.4 (orthogonal invariance). *Let $\mathbf{T} := \mathrm{diag}\{1, \mathbf{T}_3, \mathbf{T}_3, 1\}$, where $\mathbf{T}_3$ is any orthogonal matrix of size 3. If $\mathbf{U} \in \mathcal{G}$, then $\mathbf{TU} \in \mathcal{G}$.*

We refer to the following property (8) as the *LF splitting property*:

$$\mathbf{U} \pm \frac{\mathbf{F}_i(\mathbf{U})}{\alpha} \in \mathcal{G} \quad \forall\ \mathbf{U} \in \mathcal{G},\ \forall\ \alpha \geq \chi \mathscr{R}_i(\mathbf{U}), \tag{8}$$

where $\chi \geq 1$ is some constant, and $\mathscr{R}_i(\mathbf{U})$ is the spectral radius of the Jacobian matrix in the x_i-direction, $i = 1, 2, 3$. For the ideal MHD system with the EOS (4), one has [31]

$$\mathscr{R}_i(\mathbf{U}) = |v_i| + \mathcal{C}_i,$$

with

$$\mathcal{C}_i := \frac{1}{\sqrt{2}} \left[\mathcal{C}_s^2 + \frac{|\mathbf{B}|^2}{\rho} + \sqrt{\left(\mathcal{C}_s^2 + \frac{|\mathbf{B}|^2}{\rho} \right)^2 - 4 \frac{\mathcal{C}_s^2 B_i^2}{\rho}} \right]^{\frac{1}{2}},$$

where $\mathcal{C}_s = \sqrt{\gamma p / \rho}$ is the sound speed.

If true, the LF splitting property would be very useful in analyzing the PP property of the schemes with the LF flux; see its roles in [50, 40, 38] for the equations of hydrodynamics. Unfortunately, for the ideal MHD, (8) is untrue in general, as evidenced numerically in [11] for ideal gases. In fact, one can disprove (8); see the proof of the following proposition in section SM1.1 of the supplementary material.

Proposition 2.5. *The LF splitting property* (8) *does not hold in general.*

2.2. Generalized LF splitting properties.

Since (8) does not hold, we would like to seek some alternative properties which are weaker than (8). By considering the convex combination of some LF splitting terms, we discover the *generalized LF splitting properties* under some DDF condition for the magnetic field. *As one of the most highlighted points of this paper, the discovery and proofs of such properties are very nontrivial and extremely technical.*

2.2.1. A constructive inequality.

We first construct an inequality which will play a pivotal role in establishing the generalized LF splitting properties.

Lemma 2.6. *If $\mathbf{U}, \tilde{\mathbf{U}} \in \mathcal{G}$, then the inequality*

$$\left(\mathbf{U} - \frac{\mathbf{F}_i(\mathbf{U})}{\alpha} + \tilde{\mathbf{U}} + \frac{\mathbf{F}_i(\tilde{\mathbf{U}})}{\alpha} \right) \cdot \mathbf{n}^* + |\mathbf{B}^*|^2 + \frac{B_i - \tilde{B}_i}{\alpha} (\mathbf{v}^* \cdot \mathbf{B}^*) > 0 \tag{9}$$

holds for any $\mathbf{v}^*, \mathbf{B}^* \in \mathbb{R}^3$ and $|\alpha| > \alpha_i(\mathbf{U}, \tilde{\mathbf{U}})$, where $i \in \{1, 2, 3\}$, and

$$\alpha_i(\mathbf{U}, \tilde{\mathbf{U}}) = \min_{\sigma \in \mathbb{R}} \alpha_i(\mathbf{U}, \tilde{\mathbf{U}}; \sigma), \tag{10}$$

$$\alpha_i(\mathbf{U}, \tilde{\mathbf{U}}; \sigma) = \max \left\{ |v_i| + \mathscr{C}_i, |\tilde{v}_i| + \tilde{\mathscr{C}}_i, |\sigma v_i + (1-\sigma)\tilde{v}_i| + \max\{\mathscr{C}_i, \tilde{\mathscr{C}}_i\} \right\} + f(\mathbf{U}, \tilde{\mathbf{U}}; \sigma),$$

with

$$f(\mathbf{U},\tilde{\mathbf{U}};\sigma)=\frac{|\tilde{\mathbf{B}}-\mathbf{B}|}{\sqrt{2}}\sqrt{\frac{\sigma^2}{\rho}+\frac{(1-\sigma)^2}{\tilde{\rho}}},$$

$$\mathscr{C}_i=\frac{1}{\sqrt{2}}\left[\mathscr{C}_s^2+\frac{|\mathbf{B}|^2}{\rho}+\sqrt{\left(\mathscr{C}_s^2+\frac{|\mathbf{B}|^2}{\rho}\right)^2-4\frac{\mathscr{C}_s^2B_i^2}{\rho}}\right]^{\frac{1}{2}},$$

and $\mathscr{C}_s=\frac{p}{\rho\sqrt{2e}}$.

Proof. (i) We first prove (9) for $i=1$. Let us define

$$\Pi_u=\big(\mathbf{U}+\tilde{\mathbf{U}}\big)\cdot\mathbf{n}^*+|\mathbf{B}^*|^2,\quad \Pi_f=\big(\mathbf{F}_1(\mathbf{U})-\mathbf{F}_1(\tilde{\mathbf{U}})\big)\cdot\mathbf{n}^*-\big(B_1-\tilde{B}_1\big)(\mathbf{v}^*\cdot\mathbf{B}^*).$$

Then we only need to show

$$\frac{|\Pi_f|}{\Pi_u}\le\alpha_1(\mathbf{U},\tilde{\mathbf{U}}),\tag{11}$$

by noting that

$$\Pi_u=|\boldsymbol{\theta}|^2>0,\tag{12}$$

where the nonzero vector $\boldsymbol{\theta}\in\mathbb{R}^{14}$ is defined as

$$\boldsymbol{\theta}=\frac{1}{\sqrt{2}}\Big(\sqrt{\rho}(\mathbf{v}-\mathbf{v}^*),\ \mathbf{B}-\mathbf{B}^*,\ \sqrt{2\rho e},\ \sqrt{\tilde{\rho}}(\tilde{\mathbf{v}}-\mathbf{v}^*),\ \tilde{\mathbf{B}}-\mathbf{B}^*,\ \sqrt{2\tilde{\rho}\tilde{e}}\Big)^\top.$$

The proof of (11) is divided into the following two steps.

Step 1. Reformulate Π_f into a quadratic form in the variables $\theta_j, 1\le j\le 14$. We require that the coefficients of the quadratic form do not depend on $\mathbf{v}^*$ and $\mathbf{B}^*$. This is very nontrivial and becomes the key step of the proof. We first arrange Π_f by a technical decomposition,

$$\Pi_f=\Pi_1+\Pi_2+\Pi_3+(\Pi_4-\tilde{\Pi}_4),\tag{13}$$

where

$$\Pi_j=\frac{1}{2}v_1^*\big(B_j^2-\tilde{B}_j^2\big)-v_1^*B_j^*(B_j-\tilde{B}_j),\quad j=1,2,3,$$

$$\Pi_4=\frac{\rho v_1}{2}|\mathbf{v}-\mathbf{v}^*|^2+v_1\rho e+p(v_1-v_1^*)+\sum_{j=2}^{3}(B_j(v_1-v_1^*)-B_1(v_j-v_j^*))(B_j-B_j^*),$$

$$\tilde{\Pi}_4=\frac{\tilde{\rho}\tilde{v}_1}{2}|\tilde{\mathbf{v}}-\mathbf{v}^*|^2+\tilde{v}_1\tilde{\rho}\tilde{e}+\tilde{p}(\tilde{v}_1-v_1^*)+\sum_{j=2}^{3}(\tilde{B}_j(\tilde{v}_1-v_1^*)-\tilde{B}_1(\tilde{v}_j-v_j^*))(\tilde{B}_j-B_j^*).$$

One can immediately rewrite Π_4 and $\tilde{\Pi}_4$ as

$$\Pi_4=v_1\Big(\sum_{j=1}^{3}\theta_j^2+\theta_7^2\Big)+2\mathscr{C}_s\theta_1\theta_7+\frac{2B_2}{\sqrt{\rho}}\theta_1\theta_5+\frac{2B_3}{\sqrt{\rho}}\theta_1\theta_6-\frac{2B_1}{\sqrt{\rho}}(\theta_2\theta_5+\theta_3\theta_6),$$

$$\tilde{\Pi}_4=\tilde{v}_1\Big(\sum_{j=8}^{10}\theta_j^2+\theta_{14}^2\Big)+2\tilde{\mathscr{C}}_s\theta_8\theta_{14}+\frac{2\tilde{B}_2}{\sqrt{\tilde{\rho}}}\theta_8\theta_{12}+\frac{2\tilde{B}_3}{\sqrt{\tilde{\rho}}}\theta_8\theta_{13}-\frac{2\tilde{B}_1}{\sqrt{\tilde{\rho}}}(\theta_9\theta_{12}+\theta_{10}\theta_{13}).$$

After a careful investigation, we find that Π_j, $j = 1, 2, 3$, can be reformulated as

$$\begin{aligned}\Pi_j = \sigma_j \frac{\tilde{B}_j - B_j}{\sqrt{\rho}} (\theta_1\theta_{j+3} + \theta_1\theta_{j+10}) + (1-\sigma_j)\frac{\tilde{B}_j - B_j}{\sqrt{\tilde{\rho}}}(\theta_8\theta_{j+3} + \theta_8\theta_{j+10}) \\ + \big(\sigma_j v_1 + (1-\sigma_j)\tilde{v}_1\big)\theta_{j+3}^2 - \big(\sigma_j v_1 + (1-\sigma_j)\tilde{v}_1\big)\theta_{j+10}^2,\end{aligned}$$

where σ_1, σ_2, and σ_3 can be taken as any real numbers. In summary, we have reformulated Π_f into a quadratic form in the variables $\theta_j, 1 \le j \le 14$.

Step 2. Estimate the upper bound of $\frac{|\Pi_f|}{\Pi_u}$. There are several approaches to estimating the bound, resulting in different formulas. One sharp upper bound is the spectral radius of the symmetric matrix associated with the above quadratic form but cannot be formulated explicitly and computed easily in practice. An explicit sharp upper bound is $\alpha_1(\mathbf{U}, \tilde{\mathbf{U}})$ in (10). It is estimated as follows. We first notice that

$$\Pi_4 = v_1\Big(\sum_{j=1}^{3}\theta_j^2 + \theta_7^2\Big) + \boldsymbol{\vartheta}_6^\top \mathbf{A}_6 \boldsymbol{\vartheta}_6,$$

where $\boldsymbol{\vartheta}_6 = (\theta_1, \theta_2, \theta_3, \theta_5, \theta_6, \theta_7)^\top$, and

$$\mathbf{A}_6 = \begin{pmatrix} 0 & 0 & 0 & B_2\rho^{-\frac12} & B_3\rho^{-\frac12} & \mathscr{C}_s \\ 0 & 0 & 0 & -B_1\rho^{-\frac12} & 0 & 0 \\ 0 & 0 & 0 & 0 & -B_1\rho^{-\frac12} & 0 \\ B_2\rho^{-\frac12} & -B_1\rho^{-\frac12} & 0 & 0 & 0 & 0 \\ B_3\rho^{-\frac12} & 0 & -B_1\rho^{-\frac12} & 0 & 0 & 0 \\ \mathscr{C}_s & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The spectral radius of $\mathbf{A}_6$ is $\mathscr{C}_1$. This gives the following estimate:

$$\begin{aligned}|\Pi_4| &\le |v_1|\Big(\sum_{j=1}^{3}\theta_j^2 + \theta_7^2\Big) + |\boldsymbol{\vartheta}_6^\top \mathbf{A}_6 \boldsymbol{\vartheta}_6| \le |v_1|\Big(\sum_{j=1}^{3}\theta_j^2 + \theta_7^2\Big) + \mathscr{C}_1|\boldsymbol{\vartheta}_6|^2 \\ &= (|v_1| + \mathscr{C}_1)\Big(\sum_{j=1}^{3}\theta_j^2 + \theta_7^2\Big) + \mathscr{C}_1\big(\theta_5^2 + \theta_6^2\big).\end{aligned} \tag{14}$$

Similarly, we have

$$|\tilde{\Pi}_4| \le (|\tilde{v}_1| + \tilde{\mathscr{C}}_1)\Big(\sum_{j=8}^{10}\theta_j^2 + \theta_{14}^2\Big) + \tilde{\mathscr{C}}_1\big(\theta_{12}^2 + \theta_{13}^2\big). \tag{15}$$

Let us then focus on the first three terms on the right-hand side of (13) and rewrite their summation as

$$\Pi_1 + \Pi_2 + \Pi_3 = \boldsymbol{\vartheta}_8^\top \mathbf{A}_8 \boldsymbol{\vartheta}_8 + \sum_{j=1}^{3}\big(\sigma_j v_1 + (1-\sigma_j)\tilde{v}_1\big)\big(\theta_{j+3}^2 - \theta_{j+10}^2\big), \tag{16}$$

where $\boldsymbol{\vartheta}_8 = (\theta_1, \theta_4, \theta_5, \theta_6, \theta_8, \theta_{11}, \theta_{12}, \theta_{13})^\top$, and

$$\mathbf{A}_8 = \frac{1}{2}\begin{pmatrix} 0 & \boldsymbol{\psi} & 0 & \boldsymbol{\psi} \\ \boldsymbol{\psi}^\top & \mathbf{O} & \tilde{\boldsymbol{\psi}}^\top & \mathbf{O} \\ 0 & \tilde{\boldsymbol{\psi}} & 0 & \tilde{\boldsymbol{\psi}} \\ \boldsymbol{\psi}^\top & \mathbf{O} & \tilde{\boldsymbol{\psi}}^\top & \mathbf{O} \end{pmatrix},$$

with $\mathbf{O}$ denoting the 3×3 null matrix, and

$$\begin{aligned}
&\boldsymbol{\psi} = \rho^{-\frac{1}{2}} \left(\sigma_1(\tilde{B}_1 - B_1), \sigma_2(\tilde{B}_2 - B_2), \sigma_3(\tilde{B}_3 - B_3) \right), \\
&\tilde{\boldsymbol{\psi}} = \tilde{\rho}^{-\frac{1}{2}} \left((1-\sigma_1)(\tilde{B}_1 - B_1), (1-\sigma_2)(\tilde{B}_2 - B_2), (1-\sigma_3)(\tilde{B}_3 - B_3) \right).
\end{aligned}$$

Some algebraic manipulations show that the spectral radius of $\mathbf{A}_8$ is

$$\varrho(\mathbf{A}_8) = \frac{1}{2} \left[|\boldsymbol{\psi}|^2 + |\tilde{\boldsymbol{\psi}}|^2 + \sqrt{(|\boldsymbol{\psi}|^2 - |\tilde{\boldsymbol{\psi}}|^2)^2 + 4(\boldsymbol{\psi} \cdot \tilde{\boldsymbol{\psi}})^2} \right]^{\frac{1}{2}}.$$

It then follows from (16) that, for all $\sigma_1, \sigma_2, \sigma_3 \in \mathbb{R}$,

$$|\Pi_1 + \Pi_2 + \Pi_3| \le \varrho(\mathbf{A}_8) |\boldsymbol{\vartheta}_8|^2 + \sum_{j=1}^{3} \left| \sigma_j v_1 + (1-\sigma_j)\tilde{v}_1 \right| \left| \theta_{j+3}^2 - \theta_{j+10}^2 \right|.$$

For simplicity, we set $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$; then $\varrho(\mathbf{A}_8) = f(\mathbf{U}, \tilde{\mathbf{U}}; \sigma)$, and

$$\begin{aligned}
|\Pi_1 + \Pi_2 + \Pi_3| &\le f(\mathbf{U}, \tilde{\mathbf{U}}; \sigma) |\boldsymbol{\vartheta}_8|^2 + |\sigma v_1 + (1-\sigma)\tilde{v}_1| \sum_{j=1}^{3} \left| \theta_{j+3}^2 - \theta_{j+10}^2 \right| \\
&\le f(\mathbf{U}, \tilde{\mathbf{U}}; \sigma) |\boldsymbol{\theta}|^2 + |\sigma v_1 + (1-\sigma)\tilde{v}_1| \sum_{j=1}^{3} \left(\theta_{j+3}^2 + \theta_{j+10}^2 \right). \qquad (17)
\end{aligned}$$

Combining (13)–(15) and (17), we have

$$\begin{aligned}
|\Pi_f| \le& \; (|v_1| + \mathscr{C}_1) \left(\sum_{j=1}^{3} \theta_j^2 + \theta_7^2 \right) + (|\tilde{v}_1| + \tilde{\mathscr{C}_1}) \left(\sum_{j=8}^{10} \theta_j^2 + \theta_{14}^2 \right) + f(\mathbf{U}, \tilde{\mathbf{U}}; \sigma) |\boldsymbol{\theta}|^2 \\
&+ \mathscr{C}_1 \left(\theta_5^2 + \theta_6^2 \right) + \tilde{\mathscr{C}_1} \left(\theta_{12}^2 + \theta_{13}^2 \right) + |\sigma v_1 + (1-\sigma)\tilde{v}_1| \sum_{j=1}^{3} \left(\theta_{j+3}^2 + \theta_{j+10}^2 \right) \\
\le& \; (|v_1| + \mathscr{C}_1) \left(\sum_{j=1}^{3} \theta_j^2 + \theta_7^2 \right) + (|\tilde{v}_1| + \tilde{\mathscr{C}_1}) \left(\sum_{j=8}^{10} \theta_j^2 + \theta_{14}^2 \right) + f(\mathbf{U}, \tilde{\mathbf{U}}; \sigma) |\boldsymbol{\theta}|^2 \\
&+ \left(|\sigma v_1 + (1-\sigma)\tilde{v}_1| + \mathscr{C}_1 \right) \sum_{j=4}^{6} \theta_j^2 + \left(|\sigma v_1 + (1-\sigma)\tilde{v}_1| + \tilde{\mathscr{C}_1} \right) \sum_{j=11}^{13} \theta_j^2 \\
\le& \; \alpha_1(\mathbf{U}, \tilde{\mathbf{U}}; \sigma) \; |\boldsymbol{\theta}|^2 = \alpha_1(\mathbf{U}, \tilde{\mathbf{U}}; \sigma) \; \Pi_u
\end{aligned}$$

for all $\sigma \in \mathbb{R}$. Hence

$$|\Pi_f| \le \Pi_u \min_{\sigma \in \mathbb{R}} \alpha_1(\mathbf{U}, \tilde{\mathbf{U}}; \sigma) = \Pi_u \alpha_1(\mathbf{U}, \tilde{\mathbf{U}});$$

that is, the inequality (11) holds. The proof for the case of $i = 1$ is completed.

(ii) We then verify the inequality (9) for the cases $i = 2$ and 3 by using the inequality (9) for the case $i = 1$ as well as the orthogonal invariance in Lemma 2.4. For the case of $i = 2$, we introduce an orthogonal matrix $\mathbf{T} = \mathrm{diag}\{1, \mathbf{T}_3, \mathbf{T}_3, 1\}$ with $\mathbf{T}_3 := (\mathbf{e}_2^\top, \mathbf{e}_1^\top, \mathbf{e}_3^\top)$, where $\mathbf{e}_\ell$ is the ℓth row of the unit matrix of size 3. We then have

$\mathbf{TU}, \mathbf{T\tilde{U}} \in \mathcal{G}$ by Lemma 2.4. Let $\mathcal{H}_i(\mathbf{U}, \tilde{\mathbf{U}}, \mathbf{v}^*, \mathbf{B}^*, \alpha)$ denote the left-hand side term of (9). Using (9) with $i = 1$ for $\mathbf{TU}, \mathbf{T\tilde{U}}, \mathbf{v}^*\mathbf{T}_3, \mathbf{B}^*\mathbf{T}_3$, we have

$$\mathcal{H}_1(\mathbf{TU}, \mathbf{T\tilde{U}}, \mathbf{v}^*\mathbf{T}_3, \mathbf{B}^*\mathbf{T}_3, \alpha) > 0 \tag{18}$$

for any $\alpha > \alpha_1(\mathbf{TU}, \mathbf{T\tilde{U}}) = \alpha_2(\mathbf{U}, \tilde{\mathbf{U}})$. Utilizing $\mathbf{F}_1(\mathbf{TU}) = \mathbf{TF}_2(\mathbf{U})$ and the orthogonality of $\mathbf{T}$ and $\mathbf{T}_3$, we find that

$$\mathcal{H}_1(\mathbf{TU}, \mathbf{T\tilde{U}}, \mathbf{v}^*\mathbf{T}_3, \mathbf{B}^*\mathbf{T}_3, \alpha) = \mathcal{H}_2(\mathbf{U}, \tilde{\mathbf{U}}, \mathbf{v}^*, \mathbf{B}^*, \alpha).$$

Thus (18) implies (9) for $i = 2$. Similar arguments hold for $i = 3$. The proof is completed. □

Remark 2.7. In practice, it is not easy to determine the minimum value in (10). Since $\alpha_i(\mathbf{U}, \tilde{\mathbf{U}})$ only plays the role of a lower bound, one can replace it with $\alpha_i(\mathbf{U}, \tilde{\mathbf{U}}; \sigma)$ for a special σ. For example, taking $\sigma = \frac{\rho}{\rho+\tilde{\rho}}$ minimizes $f(\mathbf{U}, \tilde{\mathbf{U}}; \sigma)$ and gives

$$\alpha_i\left(\mathbf{U}, \tilde{\mathbf{U}}; \frac{\rho}{\rho+\tilde{\rho}}\right) = \max\left\{|v_i| + \mathscr{C}_i, |\tilde{v}_i| + \tilde{\mathscr{C}_i}, \frac{|\rho v_i + \tilde{\rho}\tilde{v}_i|}{\rho+\tilde{\rho}} + \max\{\mathscr{C}_i, \tilde{\mathscr{C}_i}\}\right\} + \frac{|\mathbf{B} - \tilde{\mathbf{B}}|}{\sqrt{2(\rho+\tilde{\rho})}}.$$

Taking $\sigma = \frac{\sqrt{\rho}}{\sqrt{\rho}+\sqrt{\tilde{\rho}}}$ gives

$$\alpha_i\left(\mathbf{U}, \tilde{\mathbf{U}}; \frac{\sqrt{\rho}}{\sqrt{\rho}+\sqrt{\tilde{\rho}}}\right) = \max\left\{|v_i| + \mathscr{C}_i, |\tilde{v}_i| + \tilde{\mathscr{C}_i}, \frac{|\sqrt{\rho} v_i + \sqrt{\tilde{\rho}}\tilde{v}_i|}{\sqrt{\rho}+\sqrt{\tilde{\rho}}} + \max\{\mathscr{C}_i, \tilde{\mathscr{C}_i}\}\right\} + \frac{|\mathbf{B} - \tilde{\mathbf{B}}|}{\sqrt{\rho}+\sqrt{\tilde{\rho}}}.$$

Let $a_i := \max\{\mathscr{R}_i(\mathbf{U}), \mathscr{R}_i(\tilde{\mathbf{U}})\}$. For the gamma-law EOS, the following proposition shows that $\alpha_i(\mathbf{U}, \tilde{\mathbf{U}}) < 2a_i$ and $\alpha_i(\mathbf{U}, \tilde{\mathbf{U}}) < a_i + \mathcal{O}(|\mathbf{U} - \tilde{\mathbf{U}}|)$, $i = 1, 2, 3$. When $\mathbf{U} = \tilde{\mathbf{U}}$ with zero magnetic field, $\alpha_i(\mathbf{U}, \tilde{\mathbf{U}}) = |v_i| + \frac{p}{\rho\sqrt{2e}}$, which is consistent with the bound in the LF splitting property for the Euler equations with a general EOS [52].

Proposition 2.8. *For any admissible states* $\mathbf{U}, \tilde{\mathbf{U}}$ *of an ideal gas, it holds that*

$$\alpha_i(\mathbf{U}, \tilde{\mathbf{U}}) < 2a_i, \tag{19}$$

$$\alpha_i(\mathbf{U}, \tilde{\mathbf{U}}) < a_i + \min\left\{\big||v_i| - |\tilde{v}_i|\big|, \big|\mathscr{C}_i - \tilde{\mathscr{C}_i}\big|\right\} + \frac{|\mathbf{B} - \tilde{\mathbf{B}}|}{\sqrt{2(\rho+\tilde{\rho})}}. \tag{20}$$

Proof. The inequality (19) can be shown as follows:

$$\begin{aligned} \alpha_i(\mathbf{U}, \tilde{\mathbf{U}}) &\le \alpha_i\left(\mathbf{U}, \tilde{\mathbf{U}}; \frac{\sqrt{\rho}}{\sqrt{\rho}+\sqrt{\tilde{\rho}}}\right) \\ &\le \max\left\{|v_i| + \mathscr{C}_i,\ |\tilde{v}_i| + \tilde{\mathscr{C}_i}\right\} + \frac{|\sqrt{\rho} v_i + \sqrt{\tilde{\rho}}\tilde{v}_i|}{\sqrt{\rho}+\sqrt{\tilde{\rho}}} + \frac{|\mathbf{B} - \tilde{\mathbf{B}}|}{\sqrt{\rho}+\sqrt{\tilde{\rho}}} \\ &< a_i + \frac{\sqrt{\rho}}{\sqrt{\rho}+\sqrt{\tilde{\rho}}}\left(|v_i| + \frac{|\mathbf{B}|}{\sqrt{\rho}}\right) + \frac{\sqrt{\tilde{\rho}}}{\sqrt{\rho}+\sqrt{\tilde{\rho}}}\left(|\tilde{v}_i| + \frac{|\tilde{\mathbf{B}}|}{\sqrt{\tilde{\rho}}}\right) \\ &\le a_i + \max\left\{|v_i| + \frac{|\mathbf{B}|}{\sqrt{\rho}}, |\tilde{v}_i| + \frac{|\tilde{\mathbf{B}}|}{\sqrt{\tilde{\rho}}}\right\} \\ &\le a_i + \max\{|v_i| + \mathcal{C}_i, |\tilde{v}_i| + \tilde{\mathcal{C}}_i\} = 2a_i, \end{aligned}$$

where we have used $\mathscr{C}_i < \mathcal{C}_i$ because of $\mathscr{C}_s = \sqrt{\frac{(\gamma-1)p}{2\rho}} < \mathcal{C}_s$. We now turn to prove (20). Using the triangle inequality, one can easily show that

$$\begin{aligned} |v_i| + \tilde{\mathscr{C}_i} &\le \min\left\{\big||v_i| - |\tilde{v}_i|\big|, \big|\mathscr{C}_i - \tilde{\mathscr{C}_i}\big|\right\} + \max\left\{|v_i| + \mathscr{C}_i, |\tilde{v}_i| + \tilde{\mathscr{C}_i}\right\}, \\ |\tilde{v}_i| + \mathscr{C}_i &\le \min\left\{\big||v_i| - |\tilde{v}_i|\big|, \big|\mathscr{C}_i - \tilde{\mathscr{C}_i}\big|\right\} + \max\left\{|v_i| + \mathscr{C}_i, |\tilde{v}_i| + \tilde{\mathscr{C}_i}\right\}. \end{aligned}$$

Therefore,

$$\begin{aligned}
&\max\left\{|v_i|+\mathscr{C}_i, |\tilde{v}_i|+\tilde{\mathscr{C}_i}, \frac{|\rho v_i+\tilde{\rho}\tilde{v}_i|}{\rho+\tilde{\rho}}+\max\{\mathscr{C}_i,\tilde{\mathscr{C}_i}\}\right\}\\
&\quad\le \max\left\{|v_i|+\mathscr{C}_i, |\tilde{v}_i|+\tilde{\mathscr{C}_i}, |\tilde{v}_i|+\mathscr{C}_i, |v_i|+\tilde{\mathscr{C}_i}\right\}\\
&\quad\le \max\left\{|v_i|+\mathscr{C}_i, |\tilde{v}_i|+\tilde{\mathscr{C}_i}\right\}+\min\left\{\big||v_i|-|\tilde{v}_i|\big|, \big|\mathscr{C}_i-\tilde{\mathscr{C}_i}\big|\right\}\\
&\quad< a_i+\min\left\{\big||v_i|-|\tilde{v}_i|\big|, \big|\mathscr{C}_i-\tilde{\mathscr{C}_i}\big|\right\}.
\end{aligned}$$

Then using $\alpha_i(\mathbf{U},\tilde{\mathbf{U}})\le\alpha_i\big(\mathbf{U},\tilde{\mathbf{U}};\frac{\rho}{\rho+\tilde{\rho}}\big)$ completes the proof. □

Remark 2.9. It is worth emphasizing the importance of the last term on the left-hand side of (9). This term is extremely technical, necessary, and crucial in deriving the generalized LF splitting properties. Including this term becomes one of the breakthrough points in this paper. The value of this term is not always positive or negative. However, without this term, the inequality (9) does not hold, even if α_i is replaced with $\chi\alpha_i$ for any constant $\chi\ge 1$. More importantly, this term can be canceled out dexterously under the DDF condition (21) or (26); see the proofs of generalized LF splitting properties in the following theorems.

2.2.2. Derivation of generalized LF splitting properties.

We first present the 1D generalized LF splitting property.

THEOREM 2.10 (1D generalized LF splitting). *If $\hat{\mathbf{U}}=(\hat{\rho},\hat{\mathbf{m}},\hat{\mathbf{B}},\hat{E})^\top$ and $\check{\mathbf{U}}=(\check{\rho},\check{\mathbf{m}},\check{\mathbf{B}},\check{E})^\top$ both belong to $\mathcal{G}$ and satisfy the 1D DDF condition*

$$\hat{B}_1-\check{B}_1=0, \tag{21}$$

then for any $\alpha>\alpha_1(\hat{\mathbf{U}},\check{\mathbf{U}})$ it holds that

$$\overline{\mathbf{U}}:=\frac{1}{2}\left(\hat{\mathbf{U}}-\frac{\mathbf{F}_1(\hat{\mathbf{U}})}{\alpha}+\check{\mathbf{U}}+\frac{\mathbf{F}_1(\check{\mathbf{U}})}{\alpha}\right)\in\mathcal{G}. \tag{22}$$

Proof. The first component of $\overline{\mathbf{U}}$ equals $\frac{1}{2}\big(\hat{\rho}\big(1-\frac{\hat{v}_1}{\alpha}\big)+\check{\rho}\big(1+\frac{\check{v}_1}{\alpha}\big)\big)>0$. For any $\mathbf{v}^*,\mathbf{B}^*\in\mathbb{R}^3$, utilizing Lemma 2.6 and the condition (21) gives

$$\overline{\mathbf{U}}\cdot\mathbf{n}^*+\frac{|\mathbf{B}^*|^2}{2}=\frac{1}{2}\left(\hat{\mathbf{U}}-\frac{\mathbf{F}_1(\hat{\mathbf{U}})}{\alpha}+\check{\mathbf{U}}+\frac{\mathbf{F}_1(\check{\mathbf{U}})}{\alpha}\right)\cdot\mathbf{n}^*+\frac{|\mathbf{B}^*|^2}{2}>\frac{\check{B}_1-\hat{B}_1}{2\alpha}(\mathbf{v}^*\cdot\mathbf{B}^*)=0.$$

This implies $\overline{\mathbf{U}}\in\mathcal{G}_*=\mathcal{G}$. □

Remark 2.11. As indicated by Proposition 2.8, the bound $\alpha_1(\hat{\mathbf{U}},\check{\mathbf{U}})$ for α can be very close to $a_1=\max\{\mathscr{R}_1(\hat{\mathbf{U}}),\mathscr{R}_1(\check{\mathbf{U}})\}$, which is the numerical viscosity coefficient in the standard local LF scheme. Nevertheless, (22) does not hold for $\alpha=a_1$ in general. A counterexample can be given by considering the following admissible states of ideal gas with $\gamma=1.4$ and $\hat{B}_1=\check{B}_1$:

$$\begin{cases}\hat{\mathbf{U}}=(0.2,0,0.2,0,10,5,0,62.625)^\top,\\ \check{\mathbf{U}}=(0.32,0,-0.32,0,10,10,0,100.16025)^\top.\end{cases} \tag{23}$$

For (23) and $\alpha=a_1$, one can verify that $\overline{\mathbf{U}}$ in (22) satisfies $\mathcal{E}(\overline{\mathbf{U}})<-0.05$ and $\overline{\mathbf{U}}\notin\mathcal{G}$.

Remark 2.12. The proof of Lemma 2.1 in [11] implies that

$$\mathbf{U}_\lambda := \hat{\mathbf{U}} - \lambda\big(\mathbf{F}_1(\hat{\mathbf{U}}) + a_1\hat{\mathbf{U}} - \mathbf{F}_1(\check{\mathbf{U}}) - a_1\check{\mathbf{U}}\big) \in \mathcal{G} \quad \forall \lambda \in \big(0, 1/(2a_1)\big] \tag{24}$$

holds for all admissible states $\hat{\mathbf{U}}, \check{\mathbf{U}}$ with $\hat{B}_1 = \check{B}_1$. On the contrary, for the special admissible states $\hat{\mathbf{U}}, \check{\mathbf{U}}$ in (23), Remark 2.11 yields that (24) does not always hold when λ is close to $\frac{1}{2a_1}$ because $\lim_{\lambda\to 1/(2a_1)} \mathcal{E}(\mathbf{U}_\lambda) = \mathcal{E}(\overline{\mathbf{U}}) < 0$. This deserves further explanation, as the derivation of (24) in [11] is not mathematically rigorous but based on two assumptions. One assumption is very reasonable (but unproven), stating that the exact solution $\mathbf{U}(x_1, t)$ to the 1D Riemann problem (RP)

$$\begin{cases} \frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}_1(\mathbf{U})}{\partial x_1} = \mathbf{0}, \\ \mathbf{U}(x_1, 0) = \begin{cases} \hat{\mathbf{U}}, & x_1 < 0, \\ \check{\mathbf{U}}, & x_1 > 0, \end{cases} \end{cases} \tag{25}$$

is always admissible if $\hat{\mathbf{U}},\ \check{\mathbf{U}} \in \mathcal{G}$ with $\hat{B}_1 = \check{B}_1$. Another "assumption" (not mentioned but implicitly used in [11]) is that $a_1 = \|\mathscr{R}_1(\mathbf{U}(\cdot, 0))\|_\infty$ is an upper bound of the maximum wave speed in the above RP. In fact, a_1 may not always be such a bound when the fast shocks exist in the RP solution, as indicated in [20] for the gas dynamics system (with zero magnetic field). Hence, the latter assumption may affect some 1D analysis in [11]; see our finding in Theorem 3.1. It is also worth emphasizing that the 1D analysis in [11] could work in general if $\|\mathscr{R}_1(\mathbf{U}(\cdot, 0))\|_\infty$ were replaced with a rigorous upper bound of the maximum wave speed in the RP.

We now present the multidimensional generalized LF splitting properties.

THEOREM 2.13 (2D generalized LF splitting). *If $\bar{\mathbf{U}}^i$, $\tilde{\mathbf{U}}^i$, $\hat{\mathbf{U}}^i$, $\check{\mathbf{U}}^i \in \mathcal{G}$ for $i = 1, \ldots, \mathtt{Q}$ satisfy the 2D DDF condition*

$$\frac{\sum_{i=1}^{\mathtt{Q}} \omega_i(\bar{B}_1^i - \tilde{B}_1^i)}{\Delta x} + \frac{\sum_{i=1}^{\mathtt{Q}} \omega_i(\hat{B}_2^i - \check{B}_2^i)}{\Delta y} = 0, \tag{26}$$

where $\Delta x, \Delta y > 0$, and the sum of the positive numbers $\{\omega_i\}_{i=1}^{\mathtt{Q}}$ equals one, then for any $\alpha_1^{\mathtt{LF}}$ and $\alpha_2^{\mathtt{LF}}$ satisfying $\alpha_1^{\mathtt{LF}} > \max_{1\le i\le \mathtt{Q}} \alpha_1(\bar{\mathbf{U}}^i, \tilde{\mathbf{U}}^i)$, $\alpha_2^{\mathtt{LF}} > \max_{1\le i\le \mathtt{Q}} \alpha_2(\hat{\mathbf{U}}^i, \check{\mathbf{U}}^i)$, it holds that

$$\begin{aligned} \overline{\mathbf{U}} := \frac{1}{2\left(\frac{\alpha_1^{\mathtt{LF}}}{\Delta x} + \frac{\alpha_2^{\mathtt{LF}}}{\Delta y}\right)} \sum_{i=1}^{\mathtt{Q}} \omega_i \Bigg[& \frac{\alpha_1^{\mathtt{LF}}}{\Delta x}\left(\bar{\mathbf{U}}^i - \frac{\mathbf{F}_1(\bar{\mathbf{U}}^i)}{\alpha_1^{\mathtt{LF}}} + \tilde{\mathbf{U}}^i + \frac{\mathbf{F}_1(\tilde{\mathbf{U}}^i)}{\alpha_1^{\mathtt{LF}}} \right) \\ & + \frac{\alpha_2^{\mathtt{LF}}}{\Delta y}\left(\hat{\mathbf{U}}^i - \frac{\mathbf{F}_2(\hat{\mathbf{U}}^i)}{\alpha_2^{\mathtt{LF}}} + \check{\mathbf{U}}^i + \frac{\mathbf{F}_2(\check{\mathbf{U}}^i)}{\alpha_2^{\mathtt{LF}}} \right) \Bigg] \in \mathcal{G}. \end{aligned} \tag{27}$$

Proof. The first component of $\overline{\mathbf{U}}$ equals

$$\frac{1}{2\left(\frac{\alpha_1^{\mathtt{LF}}}{\Delta x} + \frac{\alpha_2^{\mathtt{LF}}}{\Delta y}\right)} \sum_{i=1}^{\mathtt{Q}} \omega_i \left(\frac{\bar{\rho}^i(\alpha_1^{\mathtt{LF}} - \bar{v}_1^i) + \tilde{\rho}^i(\alpha_1^{\mathtt{LF}} + \tilde{v}_1^i)}{\Delta x} + \frac{\hat{\rho}^i(\alpha_2^{\mathtt{LF}} - \hat{v}_2^i) + \check{\rho}^i(\alpha_2^{\mathtt{LF}} + \check{v}_2^i)}{\Delta y} \right),$$

which is positive. For any $\mathbf{v}^*, \mathbf{B}^* \in \mathbb{R}^3$, using Lemma 2.6 and the condition (26) gives

$$\begin{aligned}
&\left(\overline{\mathbf{U}} \cdot \mathbf{n}^* + \frac{|\mathbf{B}^*|^2}{2}\right) \times 2\left(\frac{\alpha_1^{\mathtt{LF}}}{\Delta x} + \frac{\alpha_2^{\mathtt{LF}}}{\Delta y}\right) \\
&\quad = \sum_{i=1}^{\mathtt{Q}} \omega_i \Bigg\{ \frac{\alpha_1^{\mathtt{LF}}}{\Delta x}\left[\left(\bar{\mathbf{U}}^i - \frac{\mathbf{F}_1(\bar{\mathbf{U}}^i)}{\alpha_1^{\mathtt{LF}}} + \tilde{\mathbf{U}}^i + \frac{\mathbf{F}_1(\tilde{\mathbf{U}}^i)}{\alpha_1^{\mathtt{LF}}}\right) \cdot \mathbf{n}^* + |\mathbf{B}^*|^2\right] \\
&\qquad\qquad + \frac{\alpha_2^{\mathtt{LF}}}{\Delta y}\left[\left(\hat{\mathbf{U}}^i - \frac{\mathbf{F}_2(\hat{\mathbf{U}}^i)}{\alpha_2^{\mathtt{LF}}} + \check{\mathbf{U}}^i + \frac{\mathbf{F}_2(\check{\mathbf{U}}^i)}{\alpha_2^{\mathtt{LF}}}\right) \cdot \mathbf{n}^* + |\mathbf{B}^*|^2\right] \Bigg\} \\
&\quad \overset{(9)}{>} \sum_{i=1}^{\mathtt{Q}} \omega_i \left\{ \frac{\alpha_1^{\mathtt{LF}}}{\Delta x}\left[- \frac{\bar{B}_1^i - \tilde{B}_1^i}{\alpha_1^{\mathtt{LF}}} (\mathbf{v}^* \cdot \mathbf{B}^*)\right] + \frac{\alpha_2^{\mathtt{LF}}}{\Delta y}\left[- \frac{\hat{B}_2^i - \check{B}_2^i}{\alpha_2^{\mathtt{LF}}} (\mathbf{v}^* \cdot \mathbf{B}^*)\right] \right\} \\
&\quad = -(\mathbf{v}^* \cdot \mathbf{B}^*) \sum_{i=1}^{\mathtt{Q}} \omega_i \left(\frac{\bar{B}_1^i - \tilde{B}_1^i}{\Delta x} + \frac{\hat{B}_2^i - \check{B}_2^i}{\Delta y} \right) \quad \overset{(26)}{=} 0.
\end{aligned}$$

It follows that $\overline{\mathbf{U}} \cdot \mathbf{n}^* + \frac{|\mathbf{B}^*|^2}{2} > 0$. Thus $\overline{\mathbf{U}} \in \mathcal{G}_* = \mathcal{G}$. □

THEOREM 2.14 (3D generalized LF splitting). *If $\bar{\mathbf{U}}^i$, $\tilde{\mathbf{U}}^i$, $\hat{\mathbf{U}}^i$, $\check{\mathbf{U}}^i$, $\acute{\mathbf{U}}^i$, $\grave{\mathbf{U}}^i \in \mathcal{G}$ for $i = 1, \dots, \mathtt{Q}$, and they satisfy the 3D DDF condition*

$$\frac{\sum_{i=1}^{\mathtt{Q}} \omega_i (\bar{B}_1^i - \tilde{B}_1^i)}{\Delta x} + \frac{\sum_{i=1}^{\mathtt{Q}} \omega_i (\hat{B}_2^i - \check{B}_2^i)}{\Delta y} + \frac{\sum_{i=1}^{\mathtt{Q}} \omega_i (\acute{B}_3^i - \grave{B}_3^i)}{\Delta z} = 0,$$

with $\Delta x, \Delta y, \Delta z > 0$, and the sum of the positive numbers $\{\omega_i\}_{i=1}^{\mathtt{Q}}$ equals one, then for any $\alpha_1^{\mathtt{LF}}$, $\alpha_2^{\mathtt{LF}}$, and $\alpha_3^{\mathtt{LF}}$ satisfying

$$\alpha_1^{\mathtt{LF}} > \max_{1 \le i \le \mathtt{Q}} \alpha_1(\bar{\mathbf{U}}^i, \tilde{\mathbf{U}}^i), \quad \alpha_2^{\mathtt{LF}} > \max_{1 \le i \le \mathtt{Q}} \alpha_2(\hat{\mathbf{U}}^i, \check{\mathbf{U}}^i), \quad \alpha_3^{\mathtt{LF}} > \max_{1 \le i \le \mathtt{Q}} \alpha_3(\acute{\mathbf{U}}^i, \grave{\mathbf{U}}^i),$$

it holds that $\overline{\mathbf{U}} \in \mathcal{G}$, where

$$\begin{aligned}
\overline{\mathbf{U}} := \frac{1}{2\left(\frac{\alpha_1^{\mathtt{LF}}}{\Delta x} + \frac{\alpha_2^{\mathtt{LF}}}{\Delta y} + \frac{\alpha_3^{\mathtt{LF}}}{\Delta z}\right)} \sum_{i=1}^{\mathtt{Q}} \omega_i \Bigg[\frac{\alpha_1^{\mathtt{LF}}}{\Delta x}\left(\bar{\mathbf{U}}^i - \frac{\mathbf{F}_1(\bar{\mathbf{U}}^i)}{\alpha_1^{\mathtt{LF}}} + \tilde{\mathbf{U}}^i + \frac{\mathbf{F}_1(\tilde{\mathbf{U}}^i)}{\alpha_1^{\mathtt{LF}}}\right) \\
+ \frac{\alpha_2^{\mathtt{LF}}}{\Delta y}\left(\hat{\mathbf{U}}^i - \frac{\mathbf{F}_2(\hat{\mathbf{U}}^i)}{\alpha_2^{\mathtt{LF}}} + \check{\mathbf{U}}^i + \frac{\mathbf{F}_2(\check{\mathbf{U}}^i)}{\alpha_2^{\mathtt{LF}}}\right) + \frac{\alpha_3^{\mathtt{LF}}}{\Delta z}\left(\acute{\mathbf{U}}^i - \frac{\mathbf{F}_3(\acute{\mathbf{U}}^i)}{\alpha_3^{\mathtt{LF}}} + \grave{\mathbf{U}}^i + \frac{\mathbf{F}_3(\grave{\mathbf{U}}^i)}{\alpha_3^{\mathtt{LF}}}\right) \Bigg].
\end{aligned}$$

Proof. The proof is similar to that of Theorem 2.13 and is omitted here. □

Remark 2.15. In the above generalized LF splitting properties, the convex combination $\overline{\mathbf{U}}$ depends on a number of strongly coupled states, making it extremely difficult to check the admissibility of $\overline{\mathbf{U}}$. This difficulty is subtly overcome by using the inequality (9) under the DDF condition, which is an approximation to (2). For example, the 2D DDF condition (26) can be derived by using some quadrature rule for the integrals at the left side of

$$\begin{aligned}&\frac{1}{\Delta x}\left(\frac{1}{\Delta y}\int_{\mathrm{y}_0}^{\mathrm{y}_0+\Delta y}\big(B_1(\mathrm{x}_0+\Delta x,\mathrm{y})-B_1(\mathrm{x}_0,\mathrm{y})\big)d\mathrm{y}\right)\\ &+\frac{1}{\Delta y}\left(\frac{1}{\Delta x}\int_{\mathrm{x}_0}^{\mathrm{x}_0+\Delta x}\big(B_2(\mathrm{x},\mathrm{y}_0+\Delta y)-B_2(\mathrm{x},\mathrm{y}_0)\big)d\mathrm{x}\right)\\ &=\frac{1}{\Delta x\Delta y}\int_I\left(\frac{\partial B_1}{\partial \mathrm{x}}+\frac{\partial B_2}{\partial \mathrm{y}}\right)d\mathrm{x}d\mathrm{y}=0,\end{aligned}\tag{28}$$

where $(\mathtt{x},\mathtt{y})=(x_1,x_2)$ and $I=[\mathrm{x}_0,\mathrm{x}_0+\Delta x]\times[\mathrm{y}_0,\mathrm{y}_0+\Delta y]$. It is worth emphasizing that, like the necessity of the last term on the left-hand side of (9), the proposed DDF condition is necessary and crucial for the generalized LF splitting properties. Without this condition, those properties do not hold in general, even if α_i is replaced with $\chi\alpha_i$ or χa_i for any constant $\chi\geq 1$; see the proof of Theorem 4.1.

The above generalized LF splitting properties are important tools in analyzing PP schemes on uniform Cartesian meshes if the numerical flux is taken as the LF flux

$$\hat{\mathbf{F}}_\ell(\mathbf{U}^-,\mathbf{U}^+)=\frac{1}{2}\Big(\mathbf{F}_\ell(\mathbf{U}^-)+\mathbf{F}_\ell(\mathbf{U}^+)-\alpha_{\ell,n}^{\mathtt{LF}}(\mathbf{U}^+-\mathbf{U}^-)\Big),\quad \ell=1,\dots,d.\tag{29}$$

Here $\{\alpha_{\ell,n}^{\mathtt{LF}}\}$ denote the numerical viscosity parameters specified at the nth discretized time level. The extension of the above results on nonuniform or unstructured meshes will be presented in a separate paper.

3. 1D positivity-preserving schemes.

This section applies the above theories to study the provably PP schemes with the LF flux (29) for the system (1) in 1D. In 1D, the divergence-free condition (2) and the fifth equation in (1) yield that $B_1(x_1,t)\equiv$ constant (denoted by $\mathtt{B}_{\mathtt{const}}$) for all x_1 and $t\geq 0$.

To avoid confusing subscripts, we will use the symbol $\mathtt{x}$ to represent the variable x_1 in (1). Assume that the spatial domain is divided into uniform cells $\{I_j=(\mathtt{x}_{j-\frac{1}{2}},\mathtt{x}_{j+\frac{1}{2}})\}$, with a constant spatial step-size Δx. And the time interval is divided into the mesh $\{t_0=0,t_{n+1}=t_n+\Delta t_n,n\geq 0\}$ with the time step-size Δt_n determined by some CFL condition. Let $\bar{\mathbf{U}}_j^n$ denote the numerical cell-averaged approximation of the exact solution $\mathbf{U}(\mathtt{x},t)$ over I_j at $t=t_n$. Assume the discrete initial data $\bar{\mathbf{U}}_j^0\in\mathcal{G}$. A scheme is defined to be PP if its numerical solution $\bar{\mathbf{U}}_j^n$ always stays at $\mathcal{G}$.

3.1. First-order scheme.

The 1D first-order LF scheme reads

$$\bar{\mathbf{U}}_j^{n+1}=\bar{\mathbf{U}}_j^n-\frac{\Delta t_n}{\Delta x}\Big(\hat{\mathbf{F}}_1(\bar{\mathbf{U}}_j^n,\bar{\mathbf{U}}_{j+1}^n)-\hat{\mathbf{F}}_1(\bar{\mathbf{U}}_{j-1}^n,\bar{\mathbf{U}}_j^n)\Big),\tag{30}$$

where the numerical flux $\hat{\mathbf{F}}_1(\cdot,\cdot)$ is defined by (29).

A surprising discovery is that the LF scheme (30) with a standard parameter $\alpha_{1,n}^{\mathtt{LF}}=\max_j\mathscr{R}_1(\bar{\mathbf{U}}_j^n)$ (although it works well in most cases) is not always PP regardless of how small the CFL number is. However, if the parameter $\alpha_{1,n}^{\mathtt{LF}}$ in (29) satisfies

$$\alpha_{1,n}^{\mathtt{LF}}>\max_j\alpha_1(\bar{\mathbf{U}}_{j+1}^n,\bar{\mathbf{U}}_{j-1}^n),\tag{31}$$

then we can rigorously prove that the scheme (30) is PP when the CFL number is less than one. These results are shown in the following two theorems. We remark that the lower bound given in (31) is acceptable in comparison with the standard parameter $\max_j \mathscr{R}_1(\mathbf{U}_j^n)$, because one can derive from Proposition 2.8 that

$$\max_j \alpha_1(\bar{\mathbf{U}}_{j+1}^n, \bar{\mathbf{U}}_{j-1}^n) < 2 \max_j \mathscr{R}_1(\mathbf{U}_j^n),$$

and for smooth problems, $\max_j \alpha_1(\bar{\mathbf{U}}_{j+1}^n, \bar{\mathbf{U}}_{j-1}^n) < \max_j \mathscr{R}_1(\mathbf{U}_j^n) + \mathcal{O}(\Delta x)$.

Theorem 3.1. *Assume that $\bar{\mathbf{U}}_j^0 \in \mathcal{G}$ and $\bar{B}_{1,j}^0 = \mathtt{B}_{\mathtt{const}}$ for all j. Let the parameter $\alpha_{1,n}^{\mathtt{LF}} = \max_j \mathscr{R}_1(\bar{\mathbf{U}}_j^n)$, and*

$$\Delta t_n = \mathtt{C} \frac{\Delta x}{\alpha_{1,n}^{\mathtt{LF}}},$$

where $\mathtt{C}$ is the CFL number. For any constant $\mathtt{C} > 0$, the scheme (30) *is not PP.*

The proof can be found in section SM1.2 of the supplementary material.

Theorem 3.2. *Assume that $\bar{\mathbf{U}}_j^0 \in \mathcal{G}$ and $\bar{B}_{1,j}^0 = \mathtt{B}_{\mathtt{const}}$ for all j, and the parameter $\alpha_{1,n}^{\mathtt{LF}}$ satisfies* (31)*. Then the state $\bar{\mathbf{U}}_j^n$, computed by the scheme* (30) *under the CFL condition*

$$0 < \alpha_{1,n}^{\mathtt{LF}} \Delta t_n / \Delta x \le 1, \tag{32}$$

belongs to $\mathcal{G}$ and satisfies $\bar{B}_{1,j}^n = \mathtt{B}_{\mathtt{const}}$ for all j and $n \in \mathbb{N}$.

Proof. Here the induction argument is used for the time level number n. It is obvious that the conclusion holds for $n = 0$ under the hypothesis on the initial data. We now assume that $\bar{\mathbf{U}}_j^n \in \mathcal{G}$ with $\bar{B}_{1,j}^n = \mathtt{B}_{\mathtt{const}}$ for all j, and we check whether the conclusion holds for $n+1$. For the numerical flux in (29), the fifth equation in (30) gives

$$\bar{B}_{1,j}^{n+1} = \bar{B}_{1,j}^n - \frac{\lambda}{2}\big(2\bar{B}_{1,j}^n - \bar{B}_{1,j+1}^n - \bar{B}_{1,j-1}^n\big) = \mathtt{B}_{\mathtt{const}}$$

for all j, where $\lambda = \alpha_{1,n}^{\mathtt{LF}} \Delta t_n / \Delta x \in (0, 1]$ due to (32). We rewrite the scheme (30) as

$$\bar{\mathbf{U}}_j^{n+1} = (1 - \lambda)\bar{\mathbf{U}}_j^n + \lambda \mathbf{\Xi},$$

with

$$\mathbf{\Xi} := \frac{1}{2}\left(\bar{\mathbf{U}}_{j+1}^n - \frac{\mathbf{F}_1(\bar{\mathbf{U}}_{j+1}^n)}{\alpha_{1,n}^{\mathtt{LF}}} + \bar{\mathbf{U}}_{j-1}^n + \frac{\mathbf{F}_1(\bar{\mathbf{U}}_{j-1}^n)}{\alpha_{1,n}^{\mathtt{LF}}}\right).$$

Under the induction hypothesis $\bar{\mathbf{U}}_{j-1}^n, \bar{\mathbf{U}}_{j+1}^n \in \mathcal{G}$ and $\bar{B}_{1,j-1}^n = \bar{B}_{1,j+1}^n$, we conclude that $\mathbf{\Xi} \in \mathcal{G}$ by the generalized LF splitting property in Theorem 2.10. The convexity of $\mathcal{G}$ further yields $\bar{\mathbf{U}}_j^{n+1} \in \mathcal{G}$. The proof is completed. □

Remark 3.3. If the condition (32) is enhanced to $0 < \alpha_{1,n}^{\mathtt{LF}} \Delta t_n / \Delta x < 1$, then Theorem 3.2 holds for all $\alpha_{1,n}^{\mathtt{LF}} \ge \max_j \alpha_1(\bar{\mathbf{U}}_{j+1}^n, \bar{\mathbf{U}}_{j-1}^n)$, by Lemma 2.3. It is similar for Theorems 3.4, 4.3, 4.6, and 4.7 and will not be repeated.

3.2. High-order schemes. We now study the provably PP high-order schemes for 1D MHD equations (1). With the provenly PP LF scheme (30) as a building block, any high-order finite difference schemes can be modified to be PP by a limiter [13]. The following PP analysis is focused on finite volume and DG schemes. The considered 1D DG schemes are similar to those in [11] but with a different viscosity parameter in the LF flux so that the PP property can be rigorously proved in our case.

For the moment, we use the forward Euler method for time discretization, while high-order time discretization will be discussed later. We consider the high-order finite volume schemes as well as the scheme satisfied by the cell averages of a DG method, which have the following form:

$$\bar{\mathbf{U}}_j^{n+1} = \bar{\mathbf{U}}_j^n - \frac{\Delta t_n}{\Delta x}\Big(\hat{\mathbf{F}}_1(\mathbf{U}^-_{j+\frac{1}{2}}, \mathbf{U}^+_{j+\frac{1}{2}}) - \hat{\mathbf{F}}_1(\mathbf{U}^-_{j-\frac{1}{2}}, \mathbf{U}^+_{j-\frac{1}{2}})\Big), \tag{33}$$

where $\hat{\mathbf{F}}_1(\cdot,\cdot)$ is taken as the LF flux defined in (29). The quantities $\mathbf{U}^-_{j+\frac{1}{2}}$ and $\mathbf{U}^+_{j+\frac{1}{2}}$ are the high-order approximations of the point values $\mathbf{U}\big(\mathtt{x}_{j+\frac{1}{2}}, t_n\big)$ within the cells I_j and I_{j+1}, respectively, computed by

$$\mathbf{U}^-_{j+\frac{1}{2}} = \mathbf{U}_j^n\big(\mathtt{x}_{j+\frac{1}{2}} - 0\big), \quad \mathbf{U}^+_{j+\frac{1}{2}} = \mathbf{U}_{j+1}^n\big(\mathtt{x}_{j+\frac{1}{2}} + 0\big), \tag{34}$$

where the polynomial function $\mathbf{U}_j^n(\mathtt{x})$ is with the cell-averaged value of $\bar{\mathbf{U}}_j^n$, approximates $\mathbf{U}(\mathtt{x}, t_n)$ within the cell I_j, and is either reconstructed in the finite volume methods from $\{\bar{\mathbf{U}}_j^n\}$ or directly evolved in the DG methods with degree $\mathtt{K} \geq 1$. The evolution equations for the high-order "moments" of $\mathbf{U}_j^n(\mathtt{x})$ in the DG methods are omitted because we are only concerned with the PP property of the schemes here.

Generally the high-order scheme (33) is not PP. As proved in the following theorem, the scheme (33) becomes PP if $\mathbf{U}^\pm_{j+\frac{1}{2}}$ are computed by (34) with $\mathbf{U}_j^n(\mathtt{x})$ satisfying

$$B^\pm_{1,j+\frac{1}{2}} = \mathtt{B}_{\mathtt{const}} \quad \forall j, \tag{35}$$

$$\mathbf{U}_j^n(\hat{\mathtt{x}}_j^{(\mu)}) \in \mathcal{G} \quad \forall \mu \in \{1, 2, \dots, \mathtt{L}\},\ \forall j, \tag{36}$$

and $\alpha^{\mathtt{LF}}_{1,n}$ satisfies (37). Here $\{\hat{\mathtt{x}}_j^{(\mu)}\}_{\mu=1}^{\mathtt{L}}$ are the $\mathtt{L}$-point Gauss–Lobatto quadrature nodes in the interval I_j, whose associated quadrature weights are denoted by $\{\hat{\omega}_\mu\}_{\mu=1}^{\mathtt{L}}$ with $\sum_{\mu=1}^{\mathtt{L}} \hat{\omega}_\mu = 1$. We require $2\mathtt{L} - 3 \geq \mathtt{K}$ such that the algebraic precision of corresponding quadrature is at least $\mathtt{K}$, e.g., taking $\mathtt{L}$ as the ceiling part of $\frac{\mathtt{K}+3}{2}$.

THEOREM 3.4. *If the polynomial vectors* $\{\mathbf{U}_j^n(\mathtt{x})\}$ *satisfy* (35)–(36)*, and the parameter* $\alpha^{\mathtt{LF}}_{1,n}$ *in* (29) *satisfies*

$$\alpha^{\mathtt{LF}}_{1,n} > \max_j \alpha_1(\mathbf{U}^\pm_{j+\frac{1}{2}}, \mathbf{U}^\pm_{j-\frac{1}{2}}), \tag{37}$$

then the high-order scheme (33) *is PP under the CFL condition*

$$0 < \alpha^{\mathtt{LF}}_{1,n}\Delta t_n/\Delta x \leq \hat{\omega}_1. \tag{38}$$

Proof. The exactness of the $\mathtt{L}$-point Gauss–Lobatto quadrature rule for the polynomials of degree $\mathtt{K}$ yields

$$\bar{\mathbf{U}}_j^n = \frac{1}{\Delta x}\int_{I_j} \mathbf{U}_j^n(\mathtt{x})d\mathtt{x} = \sum_{\mu=1}^{\mathtt{L}} \hat{\omega}_\mu \mathbf{U}_j^n(\hat{\mathtt{x}}_j^{(\mu)}).$$

Noting $\hat{\omega}_1 = \hat{\omega}_{\mathtt{L}}$ and $\hat{\mathtt{x}}_j^{1,\mathtt{L}} = \mathtt{x}_{j\mp\frac{1}{2}}$, we can then rewrite the scheme (33) into the convex combination form

$$\bar{\mathbf{U}}_j^{n+1} = \sum_{\mu=2}^{\mathtt{L}-1} \hat{\omega}_\mu \mathbf{U}_j^n(\hat{\mathtt{x}}_j^{(\mu)}) + (\hat{\omega}_1 - \lambda)\left(\mathbf{U}^+_{j-\frac{1}{2}} + \mathbf{U}^-_{j+\frac{1}{2}}\right) + \lambda\boldsymbol{\Xi}_- + \lambda\boldsymbol{\Xi}_+, \tag{39}$$

where $\lambda = \alpha_{1,n}^{\mathtt{LF}} \Delta t_n / \Delta x \in (0, \hat{\omega}_1]$, and

$$\mathbf{\Xi}_{\pm} = \frac{1}{2} \left(\mathbf{U}_{j+\frac{1}{2}}^{\pm} - \frac{\mathbf{F}_1(\mathbf{U}_{j+\frac{1}{2}}^{\pm})}{\alpha_{1,n}^{\mathtt{LF}}} + \mathbf{U}_{j-\frac{1}{2}}^{\pm} + \frac{\mathbf{F}_1(\mathbf{U}_{j-\frac{1}{2}}^{\pm})}{\alpha_{1,n}^{\mathtt{LF}}} \right).$$

The condition (35) and (37) yield $\mathbf{\Xi}_{\pm} \in \mathcal{G}$ by the generalized LF splitting property in Theorem 2.10. We therefore have $\bar{\mathbf{U}}_j^{n+1} \in \mathcal{G}$ from (39) by the convexity of $\mathcal{G}$. □

Remark 3.5. The condition (35) is easily ensured in practice, since the exact solution $B_1(x_1, t) \equiv \mathtt{B}_{\mathtt{const}}$ and the flux for B_1 is zero, while the condition (36) can be enforced by a simple scaling limiting procedure, which was well designed in [11] by extending the techniques in [49, 50]. The details of the procedure are omitted here.

The above analysis is focused on first-order time discretization. Actually it is also valid for the high-order explicit time discretization using strong stability-preserving (SSP) methods [19, 17, 18]. This is because of the convexity of $\mathcal{G}$, as well as the fact that an SSP method is a certain convex combination of the forward Euler method.

4. 2D positivity-preserving schemes. This section discusses positivity-preserving (PP) schemes for the MHD system (1) in 2D ($d = 2$). The extension of our analysis to the 3D case ($d = 3$) is straightforward and is displayed in the supplementary material. Our analysis will reveal for the first time that the PP property of a multidimensional MHD scheme is strongly connected with a DDF condition on the numerical magnetic field.

For convenience, the symbols $(\mathtt{x}, \mathtt{y})$ are used to denote the variables (x_1, x_2) in (1). Assume that the 2D spatial domain is divided into a uniform rectangular mesh with cells $\{I_{ij} = (\mathtt{x}_{i-\frac{1}{2}}, \mathtt{x}_{i+\frac{1}{2}}) \times (\mathtt{y}_{j-\frac{1}{2}}, \mathtt{y}_{j+\frac{1}{2}})\}$. The spatial step-sizes in the $\mathtt{x}, \mathtt{y}$ directions are denoted by $\Delta x, \Delta y$, respectively. The time interval is also divided into the mesh $\{t_0 = 0, t_{n+1} = t_n + \Delta t_n, n \geq 0\}$ with the time step-size Δt_n determined by the CFL condition. We use $\bar{\mathbf{U}}_{ij}^n$ to denote the numerical approximation to the cell-averaged value of the exact solution over I_{ij} at time t_n. We aim at seeking numerical schemes whose solution $\bar{\mathbf{U}}_{ij}^n$ is preserved in $\mathcal{G}$.

4.1. First-order scheme. The 2D first-order LF scheme reads

$$\bar{\mathbf{U}}_{ij}^{n+1} = \bar{\mathbf{U}}_{ij}^n - \frac{\Delta t_n}{\Delta x} \Big(\hat{\mathbf{F}}_{1,i+\frac{1}{2},j} - \hat{\mathbf{F}}_{1,i-\frac{1}{2},j} \Big) - \frac{\Delta t_n}{\Delta y} \Big(\hat{\mathbf{F}}_{2,i,j+\frac{1}{2}} - \hat{\mathbf{F}}_{2,i,j-\frac{1}{2}} \Big), \tag{40}$$

where $\hat{\mathbf{F}}_{1,i+\frac{1}{2},j} = \hat{\mathbf{F}}_1(\bar{\mathbf{U}}_{ij}^n, \bar{\mathbf{U}}_{i+1,j}^n)$, $\hat{\mathbf{F}}_{2,i,j+\frac{1}{2}} = \hat{\mathbf{F}}_2(\bar{\mathbf{U}}_{ij}^n, \bar{\mathbf{U}}_{i,j+1}^n)$, and $\hat{\mathbf{F}}_\ell(\cdot,\cdot), \ell = 1, 2$, are the LF fluxes in (29).

As mentioned in [13], there is still no rigorous proof that the LF scheme (40) or any other first-order scheme is PP in the multidimensional cases. For the ideal MHD with the EOS (3), it seems natural to conjecture [11] that

$$\text{given } \bar{\mathbf{U}}_{ij}^n \in \mathcal{G} \ \ \forall i, j, \text{ then } \bar{\mathbf{U}}_{ij}^{n+1} \text{ computed from (40) always belongs to } \mathcal{G}, \tag{41}$$

under a certain CFL condition (e.g., the CFL number is less than 0.5). If (41) holds true, it would be important and very useful for developing PP high-order schemes [11, 12, 13] for (1). Unfortunately, the following theorem shows that (41) does not always hold, no matter how small the specified CFL number is, and even if the parameter $\alpha_{\ell,n}^{\mathtt{LF}}$ is taken as $\chi \max_{ij} \mathscr{R}_\ell(\bar{\mathbf{U}}_{ij}^n)$ with any given constant $\chi \geq 1$. (Note that increasing numerical viscosity can usually enhance the robustness of an LF scheme and increase the possibility of achieving the PP property, and $\alpha_{\ell,n}^{\mathtt{LF}} = \chi \max_{ij} \mathscr{R}_\ell(\bar{\mathbf{U}}_{ij}^n)$ corresponds to the χ times larger numerical viscosity in comparison with the standard one.)

Theorem 4.1. *Let $\alpha^{\texttt{LF}}_{\ell,n} = \chi \max_{ij} \mathscr{R}_\ell(\bar{\mathbf{U}}^n_{ij})$ with the constant $\chi \geq 1$, and*

$$\Delta t_n = \frac{\texttt{C}}{\alpha^{\texttt{LF}}_{1,n}/\Delta x + \alpha^{\texttt{LF}}_{2,n}/\Delta y},$$

where $\texttt{C} > 0$ is the CFL number. For any given constants χ and $\texttt{C}$, there always exists a set of admissible states $\{\bar{\mathbf{U}}^n_{ij}\ \forall i,j\}$ such that the solution $\bar{\mathbf{U}}^{n+1}_{ij}$ of (40) *does not belong to $\mathcal{G}$. In other words, for any given χ and $\texttt{C}$, the admissibility of $\{\bar{\mathbf{U}}^n_{ij}\ \forall i,j\}$ does not always guarantee that $\bar{\mathbf{U}}^{n+1}_{ij} \in \mathcal{G}\ \forall i,j$.*

The proof can be found in section SM1.3 of the supplementary material.

Remark 4.2. The proof of Theorem 4.1 also implies that, for any specified CFL number, the 1D LF scheme (30) is not always PP when B_1 is piecewise constant.

Inspired by Theorem 4.1, we conjecture that, to fully ensure the admissibility of $\bar{\mathbf{U}}^{n+1}_{ij}$, an additional condition is required for the states $\{\bar{\mathbf{U}}^n_{i,j}, \bar{\mathbf{U}}^n_{i\pm1,j}, \bar{\mathbf{U}}^n_{i,j\pm1}\}$ except for their admissibility. Such an additional necessary condition should be a divergence-free condition in the discrete sense for $\{\bar{\mathbf{B}}^n_{ij}\}$, whose importance for robust simulations has been widely realized. The following analysis confirms that a discrete divergence-free (DDF) condition does play an important role in achieving the PP property.

If the states $\{\bar{\mathbf{U}}^n_{i,j}\}$ are all admissible and satisfy the DDF condition

$$\text{div}_{ij}\bar{\mathbf{B}}^n := \frac{\left(\bar{B}_1\right)^n_{i+1,j} - \left(\bar{B}_1\right)^n_{i-1,j}}{2\Delta x} + \frac{\left(\bar{B}_2\right)^n_{i,j+1} - \left(\bar{B}_2\right)^n_{i,j-1}}{2\Delta y} = 0, \tag{42}$$

then we can rigorously prove that the scheme (40) preserves $\bar{\mathbf{U}}^{n+1}_{ij} \in \mathcal{G}$ by using the generalized LF splitting property in Theorem 2.13.

Theorem 4.3. *If, for all i and j, $\bar{\mathbf{U}}^n_{ij} \in \mathcal{G}$ and satisfies the DDF condition* (42)*, then the solution $\bar{\mathbf{U}}^{n+1}_{ij}$ of* (40) *always belongs to $\mathcal{G}$ under the CFL condition*

$$0 < \frac{\alpha^{\texttt{LF}}_{1,n}\Delta t_n}{\Delta x} + \frac{\alpha^{\texttt{LF}}_{2,n}\Delta t_n}{\Delta y} \leq 1, \tag{43}$$

where the parameters $\{\alpha^{\texttt{LF}}_{\ell,n}\}$ satisfy

$$\alpha^{\texttt{LF}}_{1,n} > \max_{i,j} \alpha_1(\bar{\mathbf{U}}^n_{i+1,j}, \bar{\mathbf{U}}^n_{i-1,j}), \quad \alpha^{\texttt{LF}}_{2,n} > \max_{i,j} \alpha_2(\bar{\mathbf{U}}^n_{i,j+1}, \bar{\mathbf{U}}^n_{i,j-1}). \tag{44}$$

Proof. Substituting (29) into (40) gives

$$\bar{\mathbf{U}}^{n+1}_{ij} = \lambda \boldsymbol{\Xi} + (1-\lambda)\bar{\mathbf{U}}^n_{ij},$$

where $\lambda := \Delta t_n\left(\frac{\alpha^{\texttt{LF}}_{1,n}}{\Delta x} + \frac{\alpha^{\texttt{LF}}_{2,n}}{\Delta y}\right) \in (0,1]$ by (43), and

$$\begin{aligned}\boldsymbol{\Xi} := \frac{1}{2\left(\frac{\alpha^{\texttt{LF}}_{1,n}}{\Delta x} + \frac{\alpha^{\texttt{LF}}_{2,n}}{\Delta y}\right)} \Bigg[&\frac{\alpha^{\texttt{LF}}_{1,n}}{\Delta x}\left(\bar{\mathbf{U}}^n_{i+1,j} - \frac{\mathbf{F}_1(\bar{\mathbf{U}}^n_{i+1,j})}{\alpha^{\texttt{LF}}_{1,n}} + \bar{\mathbf{U}}^n_{i-1,j} + \frac{\mathbf{F}_1(\bar{\mathbf{U}}^n_{i-1,j})}{\alpha^{\texttt{LF}}_{1,n}}\right) \\ &+ \frac{\alpha^{\texttt{LF}}_{2,n}}{\Delta y}\left(\bar{\mathbf{U}}^n_{i,j+1} - \frac{\mathbf{F}_2(\bar{\mathbf{U}}^n_{i,j+1})}{\alpha^{\texttt{LF}}_{2,n}} + \bar{\mathbf{U}}^n_{i,j-1} + \frac{\mathbf{F}_2(\bar{\mathbf{U}}^n_{i,j-1})}{\alpha^{\texttt{LF}}_{2,n}}\right)\Bigg].\end{aligned}$$

Using the condition (42) and Theorem 2.13 gives $\boldsymbol{\Xi} \in \mathcal{G}$. The convexity of $\mathcal{G}$ further yields $\bar{\mathbf{U}}^{n+1}_{ij} \in \mathcal{G}$. The proof is completed. ◻

Remark 4.4. The data in (SM2) satisfies $\mathrm{div}_{ij}\bar{\mathbf{B}}^n = \frac{\epsilon}{2\Delta x} > 0$, which can be very small when $0 < \epsilon \ll 1$. Therefore, from the proof of Theorem 4.1, we conclude that violating the condition (42) slightly could lead to inadmissible solution of the scheme (40) if the pressure is sufficiently low. This demonstrates the importance of (42).

We now discuss whether the LF scheme (40) preserves the DDF condition (42).

THEOREM 4.5. *For the LF scheme* (40), *the divergence error*

$$\varepsilon_\infty^n := \max_{ij} \left| \mathrm{div}_{ij}\bar{\mathbf{B}}^n \right|$$

does not grow with n under the condition (43). *Moreover, $\{\bar{\mathbf{U}}_{ij}^n\}$ satisfy* (42) *for all i, j and $n \in \mathbb{N}$ if* (42) *holds for the discrete initial data $\{\bar{\mathbf{U}}_{ij}^0\}$.*

Proof. Using the linearity of the operator div_{ij}, one can deduce from (40) that

$$\begin{aligned}\mathrm{div}_{ij}\bar{\mathbf{B}}^{n+1} =& (1-\lambda)\mathrm{div}_{ij}\bar{\mathbf{B}}^n + \frac{\lambda_1}{2}(\mathrm{div}_{i+1,j}\bar{\mathbf{B}}^n + \mathrm{div}_{i-1,j}\bar{\mathbf{B}}^n) \\ &+ \frac{\lambda_2}{2}(\mathrm{div}_{i,j+1}\bar{\mathbf{B}}^n + \mathrm{div}_{i,j-1}\bar{\mathbf{B}}^n),\end{aligned}$$

where $\lambda_1 = \frac{\alpha_{1,n}^{\mathtt{LF}}\Delta t_n}{\Delta x}, \lambda_2 = \frac{\alpha_{2,n}^{\mathtt{LF}}\Delta t_n}{\Delta y}, \lambda = \lambda_1 + \lambda_2 \in (0,1]$. It follows that

$$\varepsilon_\infty^{n+1} \le (1-\lambda)\varepsilon_\infty^n + \lambda_1\varepsilon_\infty^n + \lambda_2\varepsilon_\infty^n = \varepsilon_\infty^n. \tag{45}$$

This means ε_∞^n does not grow with n. If $\varepsilon_\infty^0 = 0$ for the discrete initial data $\{\bar{\mathbf{U}}_{ij}^0\}$, then $\varepsilon_\infty^n = 0$ by (45), i.e., the condition (42) is satisfied for all i, j and $n \in \mathbb{N}$. □

Finally, we obtain the first provably PP scheme for the 2D MHD system (1), as stated in the following theorem.

THEOREM 4.6. *Assume that the discrete initial data $\{\bar{\mathbf{U}}_{ij}^0\}$ are admissible and satisfy* (42), *which can be met by, e.g., the following second-order approximation:*

$$\begin{aligned}&\left(\bar{\rho}_{ij}^0, \bar{\mathbf{m}}_{ij}^0, \left(\bar{B}_3\right)_{ij}^0, \overline{(\rho e)}_{ij}^0\right) = \frac{1}{\Delta x \Delta y}\iint_{I_{ij}} (\rho, \mathbf{m}, B_3, \rho e)(\mathtt{x},\mathtt{y},0)d\mathtt{x}d\mathtt{y}, \\ &\left(\bar{B}_1\right)_{ij}^0 = \frac{1}{2\Delta y}\int_{\mathtt{y}_{j-1}}^{\mathtt{y}_{j+1}} B_1(\mathtt{x}_i,\mathtt{y},0)d\mathtt{y}, \ \left(\bar{B}_2\right)_{ij}^0 = \frac{1}{2\Delta x}\int_{\mathtt{x}_{i-1}}^{\mathtt{x}_{i+1}} B_2(\mathtt{x},\mathtt{y}_j,0)d\mathtt{x}, \\ &\bar{E}_{ij}^0 = \overline{(\rho e)}_{ij}^0 + \frac{1}{2}\left(\frac{|\bar{\mathbf{m}}_{ij}^0|^2}{\bar{\rho}_{ij}^0} + |\bar{\mathbf{B}}_{ij}^0|^2\right).\end{aligned}$$

If the parameters $\{\alpha_{\ell,n}^{\mathtt{LF}}\}$ satisfy (44), *then under the CFL condition* (43), *the LF scheme* (40) *always preserves both $\bar{\mathbf{U}}_{ij}^{n+1} \in \mathcal{G}$ and* (42) *for all i, j and $n \in \mathbb{N}$.*

Proof. This is a direct consequence of Theorems 4.3 and 4.5. □

4.2. High-order schemes. This subsection discusses the provably PP high-order finite volume or DG schemes for the 2D MHD equations (1). We will focus on the first-order forward Euler method for time discretization, and our analysis also works for high-order explicit time discretization using the SSP methods [19, 17, 18].

Towards achieving high-order [$(\mathtt{K}+1)$th order] spatial accuracy, the approximate solution polynomials $\mathbf{U}_{ij}^n(\mathtt{x},\mathtt{y})$ of degree $\mathtt{K}$ are also built usually, as approximation to the exact solution $\mathbf{U}(\mathtt{x},\mathtt{y},t_n)$ within I_{ij}. Such a polynomial vector $\mathbf{U}_{ij}^n(\mathtt{x},\mathtt{y})$ is either reconstructed in the finite volume methods from the cell averages $\{\bar{\mathbf{U}}_{ij}^n\}$ or evolved in the DG methods. Moreover, the cell average of $\mathbf{U}_{ij}^n(\mathtt{x},\mathtt{y})$ over I_{ij} is $\bar{\mathbf{U}}_{ij}^n$.

Let $\{\mathtt{x}_i^{(\mu)}\}_{\mu=1}^{\mathtt{Q}}$ and $\{\mathtt{y}_j^{(\mu)}\}_{\mu=1}^{\mathtt{Q}}$ denote the $\mathtt{Q}$-point Gauss quadrature nodes in the intervals $[\mathtt{x}_{i-\frac{1}{2}},\mathtt{x}_{i+\frac{1}{2}}]$ and $[\mathtt{y}_{j-\frac{1}{2}},\mathtt{y}_{j+\frac{1}{2}}]$, respectively, and $\{\omega_\mu\}_{\mu=1}^{\mathtt{Q}}$ be the associated weights satisfying $\sum_{\mu=1}^{\mathtt{Q}}\omega_\mu = 1$. With this quadrature rule for approximating the integrals of numerical fluxes on cell interfaces, a finite volume scheme or discrete equation for the cell average in the DG method (see, e.g., [50]) can be written as

$$
\begin{aligned}
\bar{\mathbf{U}}_{ij}^{n+1} = \bar{\mathbf{U}}_{ij}^{n} &- \frac{\Delta t_n}{\Delta x}\sum_{\mu=1}^{\mathtt{Q}}\omega_\mu\left(\hat{\mathbf{F}}_1(\mathbf{U}_{i+\frac{1}{2},j}^{-,\mu},\mathbf{U}_{i+\frac{1}{2},j}^{+,\mu}) - \hat{\mathbf{F}}_1(\mathbf{U}_{i-\frac{1}{2},j}^{-,\mu},\mathbf{U}_{i-\frac{1}{2},j}^{+,\mu})\right)\\
&- \frac{\Delta t_n}{\Delta y}\sum_{\mu=1}^{\mathtt{Q}}\omega_\mu\left(\hat{\mathbf{F}}_2(\mathbf{U}_{i,j+\frac{1}{2}}^{\mu,-},\mathbf{U}_{i,j+\frac{1}{2}}^{\mu,+}) - \hat{\mathbf{F}}_2(\mathbf{U}_{i,j-\frac{1}{2}}^{\mu,-},\mathbf{U}_{i,j-\frac{1}{2}}^{\mu,+})\right),
\end{aligned}
\tag{46}
$$

where $\hat{\mathbf{F}}_1$ and $\hat{\mathbf{F}}_2$ are the LF fluxes in (29), and the limiting values are given by

$$
\begin{aligned}
\mathbf{U}_{i+\frac{1}{2},j}^{-,\mu} &= \mathbf{U}_{ij}^n(\mathtt{x}_{i+\frac{1}{2}},\mathtt{y}_j^{(\mu)}), \qquad & \mathbf{U}_{i-\frac{1}{2},j}^{+,\mu} &= \mathbf{U}_{ij}^n(\mathtt{x}_{i-\frac{1}{2}},\mathtt{y}_j^{(\mu)}),\\
\mathbf{U}_{i,j+\frac{1}{2}}^{\mu,-} &= \mathbf{U}_{ij}^n(\mathtt{x}_i^{(\mu)},\mathtt{y}_{j+\frac{1}{2}}), \qquad & \mathbf{U}_{i,j-\frac{1}{2}}^{\mu,+} &= \mathbf{U}_{ij}^n(\mathtt{x}_i^{(\mu)},\mathtt{y}_{j-\frac{1}{2}}).
\end{aligned}
$$

For the accuracy requirement, $\mathtt{Q}$ should satisfy $\mathtt{Q} \ge \mathtt{K}+1$ for a $\mathbb{P}^{\mathtt{K}}$-based DG method, or $\mathtt{Q} \ge (\mathtt{K}+1)/2$ for a $(\mathtt{K}+1)$th order finite volume scheme.

We denote

$$
\overline{(B_1)}_{i+\frac{1}{2},j}^{\mu} := \frac{1}{2}\left((B_1)_{i+\frac{1}{2},j}^{-,\mu} + (B_1)_{i+\frac{1}{2},j}^{+,\mu}\right),\quad \overline{(B_2)}_{i,j+\frac{1}{2}}^{\mu} := \frac{1}{2}\left((B_2)_{i,j+\frac{1}{2}}^{\mu,-} + (B_2)_{i,j+\frac{1}{2}}^{\mu,+}\right)
$$

and define the discrete divergences of the numerical magnetic field $\mathbf{B}^n(\mathtt{x},\mathtt{y})$ as

$$
\mathrm{div}_{ij}\mathbf{B}^n := \frac{\sum_{\mu=1}^{\mathtt{Q}}\omega_\mu\left(\overline{(B_1)}_{i+\frac{1}{2},j}^{\mu} - \overline{(B_1)}_{i-\frac{1}{2},j}^{\mu}\right)}{\Delta x} + \frac{\sum_{\mu=1}^{\mathtt{Q}}\omega_\mu\left(\overline{(B_2)}_{i,j+\frac{1}{2}}^{\mu} - \overline{(B_2)}_{i,j-\frac{1}{2}}^{\mu}\right)}{\Delta y},
$$

which is an approximation to the left side of (28) with $(\mathtt{x}_0,\mathtt{y}_0)$ taken as $(\mathtt{x}_{i-\frac{1}{2}},\mathtt{y}_{j-\frac{1}{2}})$.

Let $\{\hat{\mathtt{x}}_i^{(\nu)}\}_{\nu=1}^{\mathtt{L}}$ and $\{\hat{\mathtt{y}}_j^{(\nu)}\}_{\nu=1}^{\mathtt{L}}$ be the $\mathtt{L}$-point Gauss–Lobatto quadrature nodes in the intervals $[\mathtt{x}_{i-\frac{1}{2}},\mathtt{x}_{i+\frac{1}{2}}]$ and $[\mathtt{y}_{j-\frac{1}{2}},\mathtt{y}_{j+\frac{1}{2}}]$, respectively, and $\{\hat{\omega}_\nu\}_{\nu=1}^{\mathtt{L}}$ be associated weights satisfying $\sum_{\nu=1}^{\mathtt{L}}\hat{\omega}_\nu = 1$, where $\mathtt{L} \ge \frac{\mathtt{K}+3}{2}$ such that the associated quadrature has algebraic precision of at least degree $\mathtt{K}$. Then we have the following sufficient conditions for the high-order scheme (46) to be PP.

THEOREM 4.7. *If the polynomial vectors* $\{\mathbf{U}_{ij}^n(\mathtt{x},\mathtt{y})\}$ *satisfy*

$$
\mathrm{div}_{ij}\mathbf{B}^n = 0 \quad \forall\ i,j, \tag{47}
$$

$$
\mathbf{U}_{ij}^n(\hat{\mathtt{x}}_i^{(\nu)},\mathtt{y}_j^{(\mu)}),\ \mathbf{U}_{ij}^n(\mathtt{x}_i^{(\mu)},\hat{\mathtt{y}}_j^{(\nu)}) \in \mathcal{G} \quad \forall\ i,j,\mu,\nu, \tag{48}
$$

then the scheme (46) *always preserves* $\bar{\mathbf{U}}_{ij}^{n+1} \in \mathcal{G}$ *under the CFL condition*

$$
0 < \frac{\alpha_{1,n}^{\mathtt{LF}}\Delta t_n}{\Delta x} + \frac{\alpha_{2,n}^{\mathtt{LF}}\Delta t_n}{\Delta y} \le \hat{\omega}_1, \tag{49}
$$

where the parameters $\{\alpha_{\ell,n}^{\mathtt{LF}}\}$ *satisfy*

$$
\alpha_{1,n}^{\mathtt{LF}} > \max_{i,j,\mu}\alpha_1\big(\mathbf{U}_{i+\frac{1}{2},j}^{\pm,\mu},\mathbf{U}_{i-\frac{1}{2},j}^{\pm,\mu}\big), \quad \alpha_{2,n}^{\mathtt{LF}} > \max_{i,j,\mu}\alpha_2\big(\mathbf{U}_{i,j+\frac{1}{2}}^{\mu,\pm},\mathbf{U}_{i,j-\frac{1}{2}}^{\mu,\pm}\big). \tag{50}
$$

Proof. Using the exactness of the Gauss–Lobatto quadrature rule with $\mathtt{L}$ nodes and the Gauss quadrature rule with $\mathtt{Q}$ nodes for the polynomials of degree $\mathtt{K}$, one can derive (cf. [50] for more details) that

$$
\begin{aligned}
\bar{\mathbf{U}}_{ij}^{n} &= \frac{\lambda_1}{\lambda}\sum_{\nu=2}^{\mathtt{L}-1}\sum_{\mu=1}^{\mathtt{Q}} \hat{\omega}_\nu \omega_\mu \mathbf{U}_{ij}^{n}\big(\hat{\mathtt{x}}_i^{(\nu)}, \mathtt{y}_j^{(\mu)}\big) + \frac{\lambda_2}{\lambda}\sum_{\nu=2}^{\mathtt{L}-1}\sum_{\mu=1}^{\mathtt{Q}} \hat{\omega}_\nu \omega_\mu \mathbf{U}_{ij}^{n}\big(\mathtt{x}_i^{(\mu)}, \hat{\mathtt{y}}_j^{(\nu)}\big) \\
&\quad + \frac{\lambda_1 \hat{\omega}_1}{\lambda}\sum_{\mu=1}^{\mathtt{Q}} \omega_\mu \left(\mathbf{U}^{+,\mu}_{i-\frac{1}{2},j} + \mathbf{U}^{-,\mu}_{i+\frac{1}{2},j}\right) + \frac{\lambda_2 \hat{\omega}_1}{\lambda}\sum_{\mu=1}^{\mathtt{Q}} \omega_\mu \left(\mathbf{U}^{\mu,+}_{i,j-\frac{1}{2}} + \mathbf{U}^{\mu,-}_{i,j+\frac{1}{2}}\right),
\end{aligned} \tag{51}
$$

where $\hat{\omega}_1 = \hat{\omega}_{\mathtt{L}}$ is used, and $\lambda_1 = \frac{\alpha^{\mathtt{LF}}_{1,n}\Delta t_n}{\Delta x}$, $\lambda_2 = \frac{\alpha^{\mathtt{LF}}_{2,n}\Delta t_n}{\Delta y}$, $\lambda = \lambda_1 + \lambda_2 \in (0, \hat{\omega}_1]$ by (49). After substituting (29) and (51) into (46), we rewrite the scheme (46) by technical arrangement into the following convex combination form:

$$
\bar{\mathbf{U}}_{ij}^{n+1} = \sum_{\nu=2}^{\mathtt{L}-1} \hat{\omega}_\nu \mathbf{\Xi}_\nu + 2(\hat{\omega}_1 - \lambda)\mathbf{\Xi}_{\mathtt{L}} + 2\lambda \mathbf{\Xi}_1, \tag{52}
$$

where $\mathbf{\Xi}_1 = \frac{1}{2}\left(\mathbf{\Xi}_- + \mathbf{\Xi}_+\right)$, and

$$
\begin{aligned}
\mathbf{\Xi}_\nu &= \frac{\lambda_1}{\lambda}\sum_{\mu=1}^{\mathtt{Q}} \omega_\mu \mathbf{U}_{ij}^{n}\big(\hat{\mathtt{x}}_i^{(\nu)}, \mathtt{y}_j^{\mu}\big) + \frac{\lambda_2}{\lambda}\sum_{\mu=1}^{\mathtt{Q}} \omega_\mu \mathbf{U}_{ij}^{n}\big(\mathtt{x}_i^{(\mu)}, \hat{\mathtt{y}}_j^{(\nu)}\big), \quad 2 \le \nu \le \mathtt{L}-1, \\
\mathbf{\Xi}_{\mathtt{L}} &= \frac{1}{2\lambda}\sum_{\mu=1}^{\mathtt{Q}} \omega_\mu \Big(\lambda_1 \left(\mathbf{U}^{-,\mu}_{i+\frac{1}{2},j} + \mathbf{U}^{+,\mu}_{i-\frac{1}{2},j}\right) + \lambda_2 \left(\mathbf{U}^{\mu,-}_{i,j+\frac{1}{2}} + \mathbf{U}^{\mu,+}_{i,j-\frac{1}{2}}\right)\Big), \\
\mathbf{\Xi}_\pm &= \frac{1}{2\left(\frac{\alpha^{\mathtt{LF}}_{1,n}}{\Delta x} + \frac{\alpha^{\mathtt{LF}}_{2,n}}{\Delta y}\right)}\sum_{\mu=1}^{\mathtt{Q}} \omega_\mu \left[\frac{\alpha^{\mathtt{LF}}_{1,n}}{\Delta x}\left(\mathbf{U}^{\pm,\mu}_{i+\frac{1}{2},j} - \frac{\mathbf{F}_1(\mathbf{U}^{\pm,\mu}_{i+\frac{1}{2},j})}{\alpha^{\mathtt{LF}}_{1,n}} + \mathbf{U}^{\pm,\mu}_{i-\frac{1}{2},j} + \frac{\mathbf{F}_1(\mathbf{U}^{\pm,\mu}_{i-\frac{1}{2},j})}{\alpha^{\mathtt{LF}}_{1,n}}\right)\right. \\
&\quad \left. + \frac{\alpha^{\mathtt{LF}}_{2,n}}{\Delta y}\left(\mathbf{U}^{\mu,\pm}_{i,j+\frac{1}{2}} - \frac{\mathbf{F}_2(\mathbf{U}^{\mu,\pm}_{i,j+\frac{1}{2}})}{\alpha^{\mathtt{LF}}_{2,n}} + \mathbf{U}^{\mu,\pm}_{i,j-\frac{1}{2}} + \frac{\mathbf{F}_2(\mathbf{U}^{\mu,\pm}_{i,j-\frac{1}{2}})}{\alpha^{\mathtt{LF}}_{2,n}}\right)\right].
\end{aligned}
$$

The condition (48) implies $\mathbf{\Xi}_\nu \in \mathcal{G},\ 2 \le \nu \le \mathtt{L}$, because $\mathcal{G}$ is convex. In order to show the admissibility of $\mathbf{\Xi}_1$ by using Theorem 2.13, one has to verify the corresponding DDF condition, which is found to be (47). Hence $\mathbf{\Xi}_1 \in \mathcal{G}$. This means the form (52) is a convex combination of the admissible states $\{\mathbf{\Xi}_k, 1 \le k \le \mathtt{L}\}$. It follows from the convexity of $\mathcal{G}$ that $\bar{\mathbf{U}}_{ij}^{n+1} \in \mathcal{G}$. The proof is completed. ∎

Remark 4.8. For some other hyperbolic systems such as the Euler [50] and shallow water [43] equations, the condition (48) is sufficient to ensure the positivity of 2D high-order schemes. However, contrary to the usual expectation (e.g., [11]), the condition (48) is not sufficient in the ideal MHD case, even if $\mathbf{B}_{ij}^{n}(\mathtt{x}, \mathtt{y})$ is locally divergence-free. This is indicated by Theorem 4.1, is confirmed by the numerical experiments in the supplementary material, and demonstrates the necessity of (47) to some extent.

Remark 4.9. In practice, the condition (48) can be easily met via a simple scaling limiting procedure [11]. It is not easy to meet (47) because it depends on the limiting values of the magnetic field calculated from the four neighboring cells of I_{ij}. If $\mathbf{B}^{n}(\mathtt{x}, \mathtt{y})$ is globally divergence-free, i.e., locally divergence-free in each cell with normal magnetic component continuous across the cell interfaces, then by Green's theorem, (47) is naturally satisfied. However, the PP limiting technique with local scaling may destroy the globally divergence-free property of $\mathbf{B}^n(\mathtt{x},\mathtt{y})$. Hence, it is nontrivial and still open to design a limiting procedure for the polynomials $\{\mathbf{U}_{ij}^n(\mathtt{x},\mathtt{y})\}$ which can enforce the conditions (48) and (47) at the same time. As a continuation of this work, [39] reports our achievement in developing multidimensional probably PP high-order schemes via the discretization of symmetrizable ideal MHD equations.

We now derive a lower bound of the internal energy when the proposed DDF condition (47) is not satisfied, to show that negative internal energy may be more easily computed in the cases with large $|\mathbf{v}\cdot\mathbf{B}|$ and large discrete divergence error.

THEOREM 4.10. *Assume that the polynomial vectors* $\{\mathbf{U}_{ij}^n(\mathtt{x},\mathtt{y})\}$ *satisfy* (48)*, and the parameters* $\{\alpha_{\ell,n}^{\mathtt{LF}}\}$ *satisfy* (50)*. Then under the CFL condition* (49)*, the solution* $\bar{\mathbf{U}}_{ij}^{n+1}$ *of the scheme* (46) *satisfies that* $\bar{\rho}_{ij}^{n+1}>0$*, and*

$$\mathcal{E}(\bar{\mathbf{U}}_{ij}^{n+1}) > -\Delta t_n\big(\bar{\mathbf{v}}_{ij}^{n+1}\cdot\bar{\mathbf{B}}_{ij}^{n+1}\big)\mathrm{div}_{ij}\mathbf{B}^n, \tag{53}$$

where the lower bound dominates the negativity of $\mathcal{E}(\bar{\mathbf{U}}_{ij}^{n+1})$*, and* $\bar{\mathbf{v}}_{ij}^{n+1} := \bar{\mathbf{m}}_{ij}^{n+1}/\bar{\rho}_{ij}^{n+1}$.

Proof. It is seen from (52) that $\bar{\rho}_{ij}^{n+1}$ is a convex combination of the first components of $\boldsymbol{\Xi}_\nu, 1\le\nu\le\mathtt{L}$, which are all positive. Thus $\bar{\rho}_{ij}^{n+1}>0$. For any $\mathbf{v}^*,\mathbf{B}^*\in\mathbb{R}^3$,

$$\left(\boldsymbol{\Xi}_1\cdot\mathbf{n}^*+\frac{|\mathbf{B}^*|^2}{2}\right)\times 2\left(\frac{\alpha_{1,n}^{\mathtt{LF}}}{\Delta x}+\frac{\alpha_{2,n}^{\mathtt{LF}}}{\Delta y}\right) > -(\mathbf{v}^*\cdot\mathbf{B}^*)\mathrm{div}_{ij}\mathbf{B}^n,$$

whose derivation is similar to that of Theorem 2.13. Because $\boldsymbol{\Xi}_\nu\in\mathcal{G},\ 2\le\nu\le\mathtt{L}$, we deduce from (52) that

$$\begin{aligned}\bar{\mathbf{U}}_{ij}^{n+1}\cdot\mathbf{n}^*+\frac{|\mathbf{B}^*|^2}{2} &= \sum_{\nu=2}^{\mathtt{L}-1}\hat{\omega}_\nu\left(\boldsymbol{\Xi}_\nu\cdot\mathbf{n}^*+\frac{|\mathbf{B}^*|^2}{2}\right)\\ &\quad+2(\hat{\omega}_1-\lambda)\left(\boldsymbol{\Xi}_{\mathtt{L}}\cdot\mathbf{n}^*+\frac{|\mathbf{B}^*|^2}{2}\right)+2\lambda\left(\boldsymbol{\Xi}_1\cdot\mathbf{n}^*+\frac{|\mathbf{B}^*|^2}{2}\right)\\ &> 2\lambda\left(\boldsymbol{\Xi}_1\cdot\mathbf{n}^*+\frac{|\mathbf{B}^*|^2}{2}\right) > -\Delta t_n(\mathbf{v}^*\cdot\mathbf{B}^*)\mathrm{div}_{ij}\mathbf{B}^n.\end{aligned}$$

Taking $\mathbf{v}^*=\bar{\mathbf{v}}_{ij}^{n+1}$ and $\mathbf{B}^*=\bar{\mathbf{B}}_{ij}^{n+1}$ gives (53). □

Several numerical examples are provided in the supplementary material to confirm the above PP analysis. The extension of our analysis to 3D is straightforward and for completeness is also given in the supplementary material.

5. Conclusions.

We presented the rigorous PP analysis of conservative schemes with the LF flux for 1D and multidimensional ideal MHD equations. It was based on several important properties of admissible state set, including a novel equivalent form, convexity, orthogonal invariance, and the generalized LF splitting properties. The analysis focused on the finite volume or discontinuous Galerkin schemes on uniform Cartesian meshes. In the 1D case, we proved that the LF scheme with proper numerical viscosity is PP, and the high-order schemes are PP under accessible conditions. In the 2D case, our analysis revealed for the first time that a discrete divergence-free (DDF) condition is crucial for achieving the PP property of schemes for ideal MHD. We proved that the 2D LF scheme with proper numerical viscosity preserves the positivity and the DDF condition. We derived sufficient conditions for achieving 2D PP high-order schemes. The lower bound of the internal energy was derived when the proposed DDF condition was not satisfied, yielding that negative internal energy may be more easily computed in cases with large $|\mathbf{v}\cdot\mathbf{B}|$ and large discrete divergence error. Our analyses were further confirmed by the numerical examples in the supplementary material, where the 3D extension was also presented.

In addition, several usually expected properties were disproved in this paper. Specifically, we rigorously showed that (i) the LF splitting property does not always hold; (ii) the 1D LF scheme with standard numerical viscosity or piecewise constant B_1 is not PP in general, no matter how small the CFL number is; (iii) the 2D LF scheme is not always PP under any CFL condition, unless an additional condition like the DDF condition is satisfied. As a result, some existing techniques for PP analysis become inapplicable in the MHD case. These, together with the technical challenges arising from the solenoidal magnetic field and the intrinsic complexity of the MHD system, make the proposed analysis very nontrivial.

From the viewpoint of preserving positivity, our analyses provided a new understanding of the importance of the divergence-free condition in robust MHD simulations. Our analyses and novel techniques as well as the provenly PP schemes can also be useful for investigating or designing other PP schemes for ideal MHD. In [39], we applied the proposed analysis approach to develop multidimensional probably PP high-order methods for the symmetrizable version of the ideal MHD equations. The extension of the PP analysis to general meshes will be studied in a coming paper.

REFERENCES

[1] R. Artebrant and M. Torrilhon, *Increasing the accuracy in locally divergence-preserving finite volume schemes for MHD*, J. Comput. Phys., 227 (2008), pp. 3405–3427.

[2] D. S. Balsara, *Second-order-accurate schemes for magnetohydrodynamics with divergence-free reconstruction*, Astrophys. J. Suppl. Ser., 151 (2004), pp. 149–184.

[3] D. S. Balsara, *Divergence-free reconstruction of magnetic fields and WENO schemes for magnetohydrodynamics*, J. Comput. Phys., 228 (2009), pp. 5040–5056.

[4] D. S. Balsara, *Self-adjusting, positivity preserving high order schemes for hydrodynamics and magnetohydrodynamics*, J. Comput. Phys., 231 (2012), pp. 7504–7517.

[5] D. S. Balsara and D. Spicer, *Maintaining pressure positivity in magnetohydrodynamic simulations*, J. Comput. Phys., 148 (1999), pp. 133–148.

[6] D. S. Balsara and D. Spicer, *A staggered mesh algorithm using high order Godunov fluxes to ensure solenoidal magnetic fields in magnetohydrodynamic simulations*, J. Comput. Phys., 149 (1999), pp. 270–292.

[7] F. Bouchut, C. Klingenberg, and K. Waagan, *A multiwave approximate Riemann solver for ideal MHD based on relaxation.* I: *Theoretical framework*, Numer. Math., 108 (2007), pp. 7–42.

[8] F. Bouchut, C. Klingenberg, and K. Waagan, *A multiwave approximate Riemann solver for ideal MHD based on relaxation.* II: *Numerical implementation with* 3 *and* 5 *waves*, Numer. Math., 115 (2010), pp. 647–679.

[9] J. U. Brackbill and D. C. Barnes, *The effect of nonzero* $\nabla\cdot\mathbf{B}$ *on the numerical solution of the magnetohydrodynamic equations*, J. Comput. Phys., 35 (1980), pp. 426–430.

[10] P. Chandrashekar and C. Klingenberg, *Entropy stable finite volume scheme for ideal compressible MHD on* 2*-D Cartesian meshes*, SIAM J. Numer. Anal., 54 (2016), pp. 1313–1340, https://doi.org/10.1137/15M1013626.

[11] Y. Cheng, F. Li, J. Qiu, and L. Xu, *Positivity-preserving DG and central DG methods for ideal MHD equations*, J. Comput. Phys., 238 (2013), pp. 255–280.

[12] A. J. Christlieb, X. Feng, D. C. Seal, and Q. Tang, *A high-order positivity-preserving single-stage single-step method for the ideal magnetohydrodynamic equations*, J. Comput. Phys., 316 (2016), pp. 218–242.

[13] A. J. Christlieb, Y. Liu, Q. Tang, and Z. Xu, *Positivity-preserving finite difference weighted ENO schemes with constrained transport for ideal magnetohydrodynamic equations*, SIAM J. Sci. Comput., 37 (2015), pp. A1825–A1845, https://doi.org/10.1137/140971208.

[14] A. J. Christlieb, J. A. Rossmanith, and Q. Tang, *Finite difference weighted essentially non-oscillatory schemes with constrained transport for ideal magnetohydrodynamics*, J. Comput. Phys., 268 (2014), pp. 302–325.

[15] A. Dedner, F. Kemm, D. Kröner, C.-D. Munz, T. Schnitzer, and M. Wesenberg, *Hyperbolic divergence cleaning for the MHD equations*, J. Comput. Phys., 175 (2002), pp. 645–673.

[16] C. R. Evans and J. F. Hawley, *Simulation of magnetohydrodynamic flows: A constrained transport method*, Astrophys. J., 332 (1988), pp. 659–677.

[17] S. Gottlieb, *On high order strong stability preserving Runge-Kutta and multi step time discretizations*, J. Sci. Comput., 25 (2005), pp. 105–128.

[18] S. Gottlieb, D. I. Ketcheson, and C.-W. Shu, *High order strong stability preserving time discretizations*, J. Sci. Comput., 38 (2009), pp. 251–289.

[19] 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.

[20] J.-L. Guermond and B. Popov, *Fast estimation from above of the maximum wave speed in the Riemann problem for the Euler equations*, J. Comput. Phys., 321 (2016), pp. 908–926.

[21] 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.

[22] P. Janhunen, *A positive conservative method for magnetohydrodynamics based on HLL and Roe methods*, J. Comput. Phys., 160 (2000), pp. 649–661.

[23] F. Li and C.-W. Shu, *Locally divergence-free discontinuous Galerkin methods for MHD equations*, J. Sci. Comput., 22 (2005), pp. 413–442.

[24] F. Li and L. Xu, *Arbitrary order exactly divergence-free central discontinuous Galerkin methods for ideal MHD equations*, J. Comput. Phys., 231 (2012), pp. 2655–2675.

[25] F. Li, L. Xu, and S. Yakovlev, *Central discontinuous Galerkin methods for ideal MHD equations with the exactly divergence-free magnetic field*, J. Comput. Phys., 230 (2011), pp. 4828–4847.

[26] S. Li, *High order central scheme on overlapping cells for magneto-hydrodynamic flows with and without constrained transport method*, J. Comput. Phys., 227 (2008), pp. 7368–7393.

[27] C. Liang and Z. Xu, *Parametrized maximum principle preserving flux limiters for high order schemes solving multi-dimensional scalar hyperbolic conservation laws*, J. Sci. Comput., 58 (2014), pp. 41–60.

[28] P. Londrillo and L. Del Zanna, *High-order upwind schemes for multidimensional magnetohydrodynamics*, Astrophys. J., 530 (2000), pp. 508–524.

[29] P. Londrillo and L. Del Zanna, *On the divergence-free condition in Godunov-type schemes for ideal magnetohydrodynamics: The upwind constrained transport method*, J. Comput. Phys., 195 (2004), pp. 17–48.

[30] S. A. Moe, J. A. Rossmanith, and D. C. Seal, *Positivity-preserving discontinuous Galerkin methods with Lax-Wendroff time discretizations*, J. Sci. Comput., 71 (2017), pp. 44–70.

[31] K. G. Powell, *An Approximate Riemann Solver for Magnetohydrodynamics (that Works in More than One Dimension)*, Tech. report, ICASE Report 94-24, NASA, Langley, VA, 1994.

[32] J. A. Rossmanith, *An unstaggered, high-resolution constrained transport method for magnetohydrodynamic flows*, SIAM J. Sci. Comput., 28 (2006), pp. 1766–1797, https://doi.org/10.1137/050627022.

[33] D. Ryu, F. Miniati, T. Jones, and A. Frank, *A divergence-free upwind code for multidimensional magnetohydrodynamic flows*, Astrophys. J., 509 (1998), pp. 244–255.

[34] M. Torrilhon, *Locally divergence-preserving upwind finite volume schemes for magnetohydrodynamic equations*, SIAM J. Sci. Comput., 26 (2005), pp. 1166–1191, https://doi.org/10.1137/S1064827503426401.

[35] M. Torrilhon and M. Fey, *Constraint-preserving upwind methods for multidimensional advection equations*, SIAM J. Numer. Anal., 42 (2004), pp. 1694–1728, https://doi.org/10.1137/S0036142903425033.

[36] G. Tóth, *The $\nabla \cdot \mathbf{B} = 0$ constraint in shock-capturing magnetohydrodynamics codes*, J. Comput. Phys., 161 (2000), pp. 605–652.

[37] K. Waagan, *A positive MUSCL-Hancock scheme for ideal magnetohydrodynamics*, J. Comput. Phys., 228 (2009), pp. 8609–8626.

[38] K. Wu, *Design of provably physical-constraint-preserving methods for general relativistic hydrodynamics*, Phys. Rev. D, 95 (2017), 103001.

[39] K. Wu and C.-W. Shu, *Provably positive discontinuous Galerkin methods for multidimensional ideal magnetohydrodynamics*, SIAM J. Sci. Comput., submitted.

[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] K. Wu and H. Tang, *Admissible states and physical-constraints-preserving schemes for relativistic magnetohydrodynamic equations*, Math. Models Methods Appl. Sci., 27 (2017), pp. 1871–1928.

[42] K. Wu and H. Tang, *Physical-constraint-preserving central discontinuous Galerkin methods for special relativistic hydrodynamics with a general equation of state*, Astrophys. J. Suppl. Ser., 228 (2017), 3.

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

[44] T. Xiong, J.-M. Qiu, and Z. Xu, *Parametrized positivity preserving flux limiters for the high order finite difference WENO scheme solving compressible Euler equations*, J. Sci. Comput., 67 (2016), pp. 1066–1088.

[45] 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.

[46] Z. Xu and X. Zhang, *Bound-preserving high order schemes*, in Handbook of Numerical Methods for Hyperbolic Problems: Applied and Modern Issues, R. Abgrall and C.-W. Shu, eds., Handb. Numer. Anal. 18, Elsevier/North–Holland, Amsterdam, 2017, pp. 81–102.

[47] S. Yakovlev, L. Xu, and F. Li, *Locally divergence-free central discontinuous Galerkin methods for ideal MHD equations*, J. Comput. Sci., 4 (2013), pp. 80–91.

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

[49] 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.

[50] 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.

[51] X. Zhang and C.-W. Shu, *Maximum-principle-satisfying and positivity-preserving high-order schemes for conservation laws: Survey and new developments*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 467 (2011), pp. 2752–2776.

[52] 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.