Mathematical Models and Methods in Applied Sciences
Vol. 27, No. 10 (2017) 1871–1928

DOI: 10.1142/S0218202517500348

World Scientific
www.worldscientific.com

Admissible states and physical-constraints-preserving schemes for relativistic magnetohydrodynamic equations

Kailiang Wu

School of Mathematical Sciences,
Peking University, Beijing 100871, P. R. China
wukl@pku.edu.cn

Huazhong Tang*

HEDPS, CAPT and LMAM, School of Mathematical Sciences,
Peking University, Beijing 100871, P. R. China

School of Mathematics and Computational Science,
Xiangtan University, Xiantan 411105,
Hunan Province, P. R. China
hztang@math.pku.edu.cn

Received 18 April 2016
Revised 16 April 2017
Accepted 14 May 2017
Published 27 July 2017
Communicated by F. Bouchut

This paper first studies the admissible state set $\mathcal{G}$ of relativistic magnetohydrodynamics (RMHD). It paves a way for developing physical-constraints-preserving (PCP) schemes for the RMHD equations with the solutions in $\mathcal{G}$. To overcome the difficulties arising from the extremely strong nonlinearities and no explicit formulas of the primitive variables and the flux vectors with respect to the conservative vector, two equivalent forms of $\mathcal{G}$ with explicit constraints on the conservative vector are skillfully discovered. The first is derived by analyzing roots of several polynomials and transferring successively them, and further used to prove the convexity of $\mathcal{G}$ with the aid of semi-positive definiteness of the second fundamental form of a hypersurface. While the second is derived based on the convexity, and then used to show the orthogonal invariance of $\mathcal{G}$. The Lax–Friedrichs (LxF) splitting property does not hold generally for the nonzero magnetic field, but by a constructive inequality and pivotal techniques, we discover the generalized LxF splitting properties, combining the convex combination of some LxF splitting terms with a discrete divergence-free condition of the magnetic field. Based on the above analyses, several 1D and 2D PCP schemes are then studied. In the 1D case, a first-order accurate LxF-type scheme is first proved to be PCP under the Courant–Friedrichs–Lewy (CFL) condition, and then the high-order accurate PCP schemes are proposed via a PCP limiter. In the 2D case, the discrete divergence-free condition and PCP property are analyzed for a first-order accurate LxF-type scheme, and two sufficient conditions are derived for high-order accurate PCP schemes. *Our analysis reveals in theory for*

*Corresponding author

the first time that the discrete divergence-free condition is closely connected with the PCP property. Several numerical examples demonstrate the theoretical findings and the performance of numerical schemes.

Keywords: Relativistic magnetohydrodynamics; physical-constraints-preserving schemes; admissible state set; convexity; generalized Lax–Friedrichs splitting; discrete divergence-free condition.

AMS Subject Classification: 65N30, 76M10, 76Y05

1. Introduction

The paper is concerned with establishing mathematical properties on the admissible state set and developing physical-constraints-preserving (PCP) numerical schemes (which preserve the positivity of the density and pressure, and the bound of the fluid velocity) for special relativistic magnetohydrodynamics (RMHD). The d-dimensional governing equations of the special RMHDs is a first-order quasilinear hyperbolic system, see e.g. Ref. 18, and in the laboratory frame, it can be written in the divergence form

$$\frac{\partial \mathbf{U}}{\partial t} + \sum_{i=1}^{d} \frac{\partial \mathbf{F}_i(\mathbf{U})}{\partial x_i} = \mathbf{0}, \tag{1.1}$$

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

$$\sum_{i=1}^{d} \frac{\partial B_i}{\partial x_i} = 0, \tag{1.2}$$

where $d = 1$, or 2 or 3, and $\mathbf{U}$ and $\mathbf{F}_i(\mathbf{U})$ denote the conservative vector and the flux in the x_i-direction, respectively, defined by

$$\begin{aligned} \mathbf{U} &= (D, \mathbf{m}, \mathbf{B}, E)^\top, \\ \mathbf{F}_i(\mathbf{U}) &= (Dv_i, v_i\mathbf{m} - B_i(W^{-2}\mathbf{B} + (\mathbf{v}\cdot\mathbf{B})\mathbf{v}) \\ &\quad + p_{\rm tot}\mathbf{e}_i, v_i\mathbf{B} - B_i\mathbf{v}, m_i)^\top, \quad i = 1, \ldots, d, \end{aligned}$$

with the mass density $D = \rho W$, the momentum density (row) vector $\mathbf{m} = \rho h W^2 \mathbf{v} + |\mathbf{B}|^2\mathbf{v} - (\mathbf{v}\cdot\mathbf{B})\mathbf{B}$, the energy density $E = \rho h W^2 - p_{\rm tot} + |\mathbf{B}|^2$, and the row vector $\mathbf{e}_i$ denoting the ith row of the unit matrix of size 3. Here ρ is the rest-mass density, the row vector $\mathbf{v} = (v_1, v_2, v_3)$ denotes the fluid velocity, $p_{\rm tot}$ is the total pressure containing the gas pressure p and magnetic pressure $p_m := \frac{1}{2}(W^{-2}|\mathbf{B}|^2 + (\mathbf{v}\cdot\mathbf{B})^2)$, $W = 1/\sqrt{1-v^2}$ is the Lorentz factor with $v := (v_1^2 + v_2^2 + v_3^2)^{1/2}$, h is the specific enthalpy defined by

$$h = 1 + e + \frac{p}{\rho},$$

with units in which the speed of light c is equal to one, and e is the specific internal energy. It can be seen that the conservative variables $\mathbf{m}$ and E depend on the magnetic field $\mathbf{B}$ nonlinearly.

The system (1.1) takes into account the relativistic description for the dynamics of electrically-conducting fluid (plasma) at nearly speed of light in vacuum in the presence of magnetic fields. The relativistic magneto-fluid flow appears in investigating numerous astrophysical phenomena from stellar to galactic scales, e.g. core collapse supernovae, coalescing neutron stars, X-ray binaries, active galactic nuclei, formation of black holes, superluminal jets and gamma-ray bursts, etc. However, due to relativistic effect, especially the appearance of the Lorentz factor, the system (1.1) involves strong nonlinearity, making its analytic treatment extremely difficult. A primary and powerful approach to improve our understanding of the physical mechanisms in the RMHDs is through numerical simulations. In comparison with the non-relativistic MHD case, the numerical difficulties are coming from strongly nonlinear coupling between the RMHD equations (1.1), which leads to no explicit expression of the primitive variable vector $\mathbf{V} = (\rho, \mathbf{v}, \mathbf{B}, p)^\top$ and the flux $\mathbf{F}_i$ in terms of $\mathbf{U}$, and some physical constraints such as $\rho > 0$, $p > 0$, and $v < c = 1$, etc.

Since nearly the 2000s, the numerical study of the RMHDs has attracted considerable attention, and various modern shock-capturing methods have been developed for the RMHD equations, e.g. the Godunov-type scheme based on the linear Riemann solver,[25] the total variation diminishing scheme,[3] the third-order accurate central-type scheme based on two-speed approximate Riemann solver,[14] the high-order kinetic flux-splitting method,[34] the HLLC (Harten–Lax-van Leer–Contact)-type schemes,[22,24,30] the adaptive methods with mesh refinement,[1,38] the adaptive moving mesh method,[21] the locally divergence-free Runge–Kutta discontinuous Galerkin (RKDG) method and exactly divergence-free central RKDG method with the weighted essentially non-oscillatory limiters,[48] the ADER (Arbitrary high-order schemes using DERivatives) DG method,[44] and the ADER-WENO-type schemes with subluminal reconstruction,[7] etc. The readers are also referred to the early review papers, Refs. 17 and 29.

To our best knowledge, up to now, no work shows in theory that those existing numerical methods for RMHDs can preserve the positivity of the rest-mass density and the pressure and the bounds of the fluid velocity at the same time, although those schemes have been used to simulate some RMHD flows successfully. There exists a large and long-standing risk of failure when a numerical scheme is applied to the RMHD problems with large Lorentz factor, low density or pressure, or strong discontinuity. This is because as soon as the negative density or pressure, or the superluminal fluid velocity may be obtained, the eigenvalues of the Jacobian matrix become imaginary, such that the discrete problem is ill-posed. It is of great significance to develop high-order accurate PCP numerical schemes, in the sense of that the solution of numerical scheme always belongs to the (physical) *admissible state set*

$$\mathcal{G} := \{\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \in \mathbb{R}^8 \,|\, \rho(\mathbf{U}) > 0, p(\mathbf{U}) > 0, v(\mathbf{U}) < c = 1\}. \quad (1.3)$$

Because the functions $\rho(\mathbf{U})$, $p(\mathbf{U})$ and $v(\mathbf{U})$ in (1.3), and $\mathbf{F}_i(\mathbf{U})$ are highly nonlinear and cannot be expressed explicitly in $\mathbf{U}$, it is extremely hard to check whether

a given state $\mathbf{U}$ is admissible, or whether the numerical scheme is PCP. For this reason, developing the PCP schemes for the RMHDs is highly challenging. It is still an unsolved problem. In fact, it is also a blank in developing the positivity-preserving scheme with strictly and completely theoretical proof for the (non-relativistic) ideal compressible MHD.[a] Studying the intrinsic mathematical properties of the admissible state set $\mathcal{G}$ may open a window for such unsolved problem, see e.g. the recent works[40,41,39] on the PCP schemes for the RHDs.

Besides three physical constraints (1.3) on the admissible state $\mathbf{U}$, another difficulty for the RMHD system (1.1) comes from the divergence-free condition (1.2). Numerically preserving such condition is very non-trivial (for $d \geq 2$) but important for the robustness of numerical scheme, and has to be respected. In physics, numerically incorrect magnetic field topologies may lead to nonphysical plasma transport orthogonal to the magnetic field, see e.g. Ref. 9. The condition (1.2) is also very crucial for the stability of induction equation.[43] The existing numerical experiments in the non-relativistic MHD case have also indicated that violating the divergence-free condition of magnetic field may lead to numerical instability and nonphysical or inadmissible solutions.[9,4,36,6] Up to now, several numerical treatments have been proposed to reduce such risk, see e.g. Refs. 16, 4, 27, 5 and 28 and references therein. *However, it is still unknown in theory why violating the divergence-free condition of magnetic field does more easily cause inadmissible solution.*

The aim of the paper is to do the first attempt in studying the properties of $\mathcal{G}$ and the PCP numerical schemes for the special RMHD equations (1.1). The main contributions are outlined as follows:

(1) Deriving the first equivalent form of $\mathcal{G}$ by analyzing polynomial roots and transferring successively. The constraints in this equivalent form of $\mathcal{G}$ depend explicitly on the value of $\mathbf{U}$ so that the judgment of the admissible state becomes direct and it is useful to develop the PCP limiter for the high-order accurate PCP schemes for the RMHDs.

(2) Proving the convexity of $\mathcal{G}$. The convexity of $\mathcal{G}$ seems natural from the physical point of view and is critical for studying the PCP schemes. However, its proof is non-trivial and suffers from the difficulty arising from the strongly nonlinear constraints. The key point is to utilize the semi-positive definiteness of the second fundamental form of a hypersurface, which is discovered to have a proper parametric equation.

(3) Discovery of the second equivalent form of $\mathcal{G}$ based on its convexity. This equivalent form of $\mathcal{G}$ is simple and beautiful with the constraints depending

[a] Although several efforts were made to enforce positivity of the reconstructed or DG polynomial solutions based on the assumption that the cell average values calculated by the numerical schemes are admissible, see e.g. Refs. 6, 10 and 7. However, there is no rigorous proof, especially in the multi-dimensional case, to genuinely show that those schemes can preserve the positivity. In fact, the 2D first-order accurate (LxF) scheme is still possibly not PCP, see Example 3.1.

linearly on **U** and plays a pivotal role in verifying the PCP property of the numerical scheme for the RMHDs. Moreover, it also implies the orthogonal invariance of $\mathcal{G}$.

(4) Establishment of the generalized Lax–Friedrichs (LxF) splitting properties of $\mathcal{G}$. An analytic counterexample shows that $\mathcal{G}$ does not have the LxF splitting property in general. Luckily, we discover an alternative, the so-called generalized LxF splitting property, which is coupling the convex combination of some LxF splitting terms with a "discrete divergence-free" condition for the magnetic field. *Since the generalized LxF splitting properties involve lots of states with strongly coupling condition, their discovery and proofs are extremely technical and become the most highlighted point of this paper.*

(5) Close connection between the discrete divergence-free condition and PCP property is revealed in theory for the first time. Analytic example indicates that first-order accurate LxF-type scheme violating the divergence-free condition may produce inadmissible solution. Our theoretical analysis clearly shows the importance of discrete divergence-free condition in proving the PCP properties of numerical schemes.

(6) Theoretical analysis on several 1D and 2D PCP schemes. The 1D first-order accurate LxF-type scheme is proved to be PCP under the Courant–Friedrichs–Lewy (CFL) condition and the PCP limiter is developed to propose the 1D high-order accurate PCP schemes. The discrete divergence-free condition and PCP property are analyzed for the 2D first-order accurate LxF-type scheme, and two sufficient conditions are derived for the 2D high-order accurate PCP schemes. Several numerical examples are given to demonstrate the theoretical analyses and the performance of numerical schemes.

The rest of the paper is organized as follows. Section 2 derives the two equivalent definitions of $\mathcal{G}$, proves its convexity, and establishes the generalized LxF splitting properties under the "discrete divergence-free" condition. They play pivotal roles in analyzing the PCP property of the numerical methods based on the LxF-type flux for the RMHD equations (1.1), see Sec. 3, where the PCP properties of the 1D and 2D first-order accurate LxF schemes are proved, the PCP limiting procedure and the high-order accurate PCP schemes for the 1D RMHD equations (1.1) are presented, and sufficient conditions for the 2D high-order accurate PCP schemes are also proposed. Section 4 conducts several numerical experiments to demonstrate the theoretical analyses and the performance of the proposed schemes. Section 5 concludes the paper with several remarks.

2. Properties of the Admissible State Set

Throughout the paper, the equation of state (EOS) will be restricted to the Γ-law

$$p = (\Gamma - 1)\rho e, \tag{2.1}$$

where the adiabatic index $\Gamma \in (1, 2]$. The restriction of $\Gamma \leq 2$ is required for the compressibility assumptions[12] and the causality in the theory of relativity (the sound speed does not exceed the speed of light $c = 1$). All results in this paper can be extended to the general EOS case by the similar discussion presented in Ref. 41.

Lemma 2.1. *The admissible state* $\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \in \mathcal{G}$ *must satisfy*

$$D > 0, \quad q(\mathbf{U}) := E - \sqrt{D^2 + |\mathbf{m}|^2} > 0. \tag{2.2}$$

Proof. Under three conditions of $\mathcal{G}$ in (1.3), i.e. $\rho(\mathbf{U}) > 0$, $p(\mathbf{U}) > 0$, and $v(\mathbf{U}) < 1$, one has

$$\begin{aligned} D &= \rho W > 0, \\ E &= \rho h W^2 - p + \frac{1+v^2}{2}|\mathbf{B}|^2 - \frac{1}{2}(\mathbf{v}\cdot\mathbf{B})^2 \\ &\geq \rho h W^2 - p > \rho h - p = \rho + \frac{p}{\Gamma - 1} > 0, \end{aligned}$$

and

$$\begin{aligned} E^2 - (D^2 + |\mathbf{m}|^2) &= \left[\rho h W^2 - p + \frac{1+v^2}{2}|\mathbf{B}|^2 - \frac{1}{2}(\mathbf{v}\cdot\mathbf{B})^2\right]^2 \\ &\quad - (\rho W)^2 - |(\rho h W^2 + |\mathbf{B}|^2)\mathbf{v} - (\mathbf{v}\cdot\mathbf{B})\mathbf{B}|^2 \\ &= [(\rho h W^2 - p)^2 - (\rho W)^2 - (\rho h W^2 + |\mathbf{B}|^2)^2 v^2] \\ &\quad + \left[\frac{1+v^2}{2}|\mathbf{B}|^2 - \frac{1}{2}(\mathbf{v}\cdot\mathbf{B})^2\right]^2 + (\rho h W^2 - p)[(1+v^2)|\mathbf{B}|^2 \\ &\quad - (\mathbf{v}\cdot\mathbf{B})^2] - (\mathbf{v}\cdot\mathbf{B})^2|\mathbf{B}|^2 + 2(\rho h W^2 + |\mathbf{B}|^2)(\mathbf{v}\cdot\mathbf{B})^2. \end{aligned}$$

The first term on the right-hand side of the above identity should be larger than

$$-(2\rho h W^2 |\mathbf{B}|^2 + |\mathbf{B}|^4) v^2,$$

because of the inequality

$$(\rho h W^2 - p)^2 > |\rho h W^2 \mathbf{v}|^2 + (\rho W)^2,$$

which has been proved in Ref. 40. Thus, one has

$$\begin{aligned} E^2 - (D^2 + |\mathbf{m}|^2) &> 2\rho h W^2 [(\mathbf{v}\cdot\mathbf{B})^2 - |\mathbf{B}|^2 v^2] \\ &\quad + (\rho h W^2 - p)[(1+v^2)|\mathbf{B}|^2 - (\mathbf{v}\cdot\mathbf{B})^2] \\ &\quad + |\mathbf{B}|^2(\mathbf{v}\cdot\mathbf{B})^2 + \left[\frac{1+v^2}{2}|\mathbf{B}|^2 - \frac{1}{2}(\mathbf{v}\cdot\mathbf{B})^2\right]^2 - |\mathbf{B}|^4 v^2 \end{aligned}$$

$$
\begin{aligned}
&= \rho h W^2[|\mathbf{B}|^2 - v^2|\mathbf{B}|^2 + (\mathbf{v}\cdot\mathbf{B})^2] - p[(1+v^2)|\mathbf{B}|^2 - (\mathbf{v}\cdot\mathbf{B})^2] \\
&\quad + |\mathbf{B}|^2[(\mathbf{v}\cdot\mathbf{B})^2 - |\mathbf{B}|^2 v^2] + \left[\frac{|\mathbf{B}|^2}{2} + \frac{v^2|\mathbf{B}|^2 - (\mathbf{v}\cdot\mathbf{B})^2}{2}\right]^2 \\
&\geq (\rho h - p(1+v^2))|\mathbf{B}|^2 + \left[\frac{|\mathbf{B}|^2}{2} - \frac{v^2|\mathbf{B}|^2 - (\mathbf{v}\cdot\mathbf{B})^2}{2}\right]^2 \\
&\geq (\rho h - 2p)|\mathbf{B}|^2 = \left(\rho + \frac{2-\Gamma}{\Gamma-1}p\right)|\mathbf{B}|^2 \geq 0,
\end{aligned}
$$

which along with $E > 0$ yield $q(\mathbf{U}) = E - \sqrt{D^2 + |\mathbf{m}|^2} > 0$. The proof is completed. □

If the magnetic field $\mathbf{B}$ is zero, then (2.2) is also sufficient for $\mathbf{U} \in \mathcal{G}$, see Ref. 40, and $q(\mathbf{U})$ is a concave function in terms of $\mathbf{U}$. Those results have played pivotal roles in the analysis and constructions of the PCP schemes for the RHDs.[40] Unfortunately, (2.2) is only necessary (not sufficient) for $\mathbf{U} \in \mathcal{G}$ if $\mathbf{B} \neq \mathbf{0}$. In spite of this, (2.2) is still important and essential in the coming analysis.

Since there is no explicit expression of $\rho(\mathbf{U}), p(\mathbf{U})$ and $\mathbf{v}(\mathbf{U})$ for the RMHDs, the value of $\mathbf{V}$ should be derived from given $\mathbf{U}$ by solving some nonlinear algebraic equation, see e.g. Refs. 3, 14, 25, 30, 32 and 33. This paper considers the nonlinear algebraic equation used in Ref. 30:

$$
f_{\mathbf{U}}(\xi) := \xi - \frac{\Gamma-1}{\Gamma}\left(\frac{\xi}{W^2} - \frac{D}{W}\right) + |\mathbf{B}|^2 - \frac{1}{2}\left[\frac{|\mathbf{B}|^2}{W^2} + \frac{(\mathbf{m}\cdot\mathbf{B})^2}{\xi^2}\right] - E = 0, \tag{2.3}
$$

for the unknown $\xi \in \mathbb{R}^+$, where the Lorentz factor W has been expressed as a function of ξ by

$$
W(\xi) = (\xi^{-2}(\xi + |\mathbf{B}|^2)^{-2} f_\Omega(\xi))^{-1/2}, \tag{2.4}
$$

with

$$
f_\Omega(\xi) := \xi^2(\xi + |\mathbf{B}|^2)^2 - [\xi^2|\mathbf{m}|^2 + (2\xi + |\mathbf{B}|^2)(\mathbf{m}\cdot\mathbf{B})^2]. \tag{2.5}
$$

It is reasonable to find the solution of (2.3) within the interval

$$
\Omega_f := \mathbb{R}^+ \cap \{\xi \,|\, f_\Omega(\xi) > 0\}, \tag{2.6}
$$

otherwise, $f_\Omega(\xi) \leq 0$ such that $W(\xi)$ takes the value of 0 or the imaginary number. If denote the solution of Eq. (2.3) by $\xi_* = \xi_*(\mathbf{U})$, then $\xi_* = \rho(\mathbf{U})h(\mathbf{U})W^2(\xi_*) = \rho(\mathbf{U})h(\mathbf{U})/(1 - v^2(\mathbf{U}))$ and the values of the primitive variables $\rho(\mathbf{U})$, $p(\mathbf{U})$ and $v(\mathbf{U})$ in (1.3) can be calculated by

$$
\mathbf{v}(\mathbf{U}) = (\mathbf{m} + \xi_*^{-1}(\mathbf{m}\cdot\mathbf{B})\mathbf{B})/(\xi_* + |\mathbf{B}|^2), \tag{2.7}
$$

$$\rho(\mathbf{U}) = \frac{D}{W(\xi_*)}, \tag{2.8}$$

$$p(\mathbf{U}) = \frac{\Gamma - 1}{\Gamma W^2(\xi_*)}(\xi_* - DW(\xi_*)). \tag{2.9}$$

The above procedure clearly shows the strong nonlinearity of the functions $\mathbf{v}(\mathbf{U})$, $\rho(\mathbf{U})$ and $p(\mathbf{U})$, and the difficulty in verifying whether $\mathbf{U}$ is in the admissible state set $\mathcal{G}$. To overcome such difficulty, two equivalent definitions of the admissible state set $\mathcal{G}$ will be given in the following. The first is very suitable to check whether a given state $\mathbf{U}$ is admissible and constructs the PCP limiter for the development of high-order accurate PCP schemes for 1D RMHD equations, while the second is very effective in verifying the PCP property of a numerical scheme. Moreover, the convexity of $\mathcal{G}$ will also be analyzed.

2.1. *First equivalent definition*

This subsection introduces the first equivalent definition of the admissible state set $\mathcal{G}$.

Theorem 2.1. (First equivalent definition) *The admissible state set $\mathcal{G}$ is equivalent to the following set*

$$\mathcal{G}_0 := \{\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \mid D > 0, q(\mathbf{U}) > 0, \Psi(\mathbf{U}) > 0\}, \tag{2.10}$$

where

$$\Psi(\mathbf{U}) := (\Phi(\mathbf{U}) - 2(|\mathbf{B}|^2 - E))\sqrt{\Phi(\mathbf{U}) + |\mathbf{B}|^2 - E} - \sqrt{\frac{27}{2}(D^2|\mathbf{B}|^2 + (\mathbf{m}\cdot\mathbf{B})^2)},$$

with $\Phi(\mathbf{U}) := \sqrt{(|\mathbf{B}|^2 - E)^2 + 3(E^2 - D^2 - |\mathbf{m}|^2)}$.

Proof. The proof of Theorem 2.1 is very technical, and will be built on several lemmas given behind it.

(1) Lemma 2.2 tells us that three constraints $\rho > 0$, $p > 0$, and $v < 1$ in $\mathcal{G}$ can be equivalently replaced with

$$\xi_*(\mathbf{U}) > 0, \quad f_4(\xi_*(\mathbf{U})) > 0, \quad D > 0, \quad q(\mathbf{U}) > 0, \tag{2.11}$$

where the existence and uniqueness of $\xi_*(\mathbf{U})$ have been required, and $f_4(\xi)$ is a quartic polynomial defined by

$$f_4(\xi) := f_\Omega(\xi) - D^2(\xi + |\mathbf{B}|^2)^2. \tag{2.12}$$

For $\xi \in \Omega_f$, $f_4(\xi) = (W(\xi))^{-2}(\xi^2 - D^2(W(\xi))^2)(\xi + |\mathbf{B}|^2)^2$. The subsequent task is to prove the equivalence between first two conditions in (2.11) and the third one in $\mathcal{G}_0$ under (2.2).

(2) Lemma 2.3 shows that Ω_f is an open interval and can be equivalently expressed as $\Omega_f = (\xi_\Omega, +\infty)$, where $\xi_\Omega = \xi_\Omega(\mathbf{U})$ denotes the biggest non-negative root of $f_\Omega(\xi)$ in (2.5).

(3) Lemma 2.4 shows that the polynomial $f_4(\xi)$ has unique positive root in Ω_f, denoted by ξ_4, and first two constraints in (2.11) are equivalently replaced with $\xi_* > \xi_4$, that is to say, (2.11) is equivalent to

$$\xi_*(\mathbf{U}) > \xi_4(\mathbf{U}), \quad D > 0, \quad q(\mathbf{U}) > 0. \tag{2.13}$$

(4) Lemma 2.5 states that the function $f_{\mathbf{U}}(\xi)$ defined in (2.3) is strictly monotone increasing in Ω_f, and $\lim_{\xi\to+\infty} f_{\mathbf{U}}(\xi) = +\infty$. Hence the first inequality in (2.13) holds if and only if

$$f_{\mathbf{U}}(\xi_4) = \xi_4 - \frac{D^2|\mathbf{B}|^2 + (\mathbf{m}\cdot\mathbf{B})^2}{2\xi_4^2} + |\mathbf{B}|^2 - E < 0 = f_{\mathbf{U}}(\xi_*).$$

Here we have used that $\xi_4 \in \Omega_f$ and $\xi_4 = DW(\xi_4)$ for the left equal sign.

If defining a cubic polynomial $f_3(\xi)$ by

$$f_3(\xi) := \xi^3 + (|\mathbf{B}|^2 - E)\xi^2 - \frac{|\mathbf{B}|^2 D^2 + (\mathbf{m}\cdot\mathbf{B})^2}{2}, \tag{2.14}$$

then $f_3(\xi_4) = \xi_4^2 f_{\mathbf{U}}(\xi_4)$ and (2.13) is equivalent to

$$f_3(\xi_4(\mathbf{U})) < 0, \quad D > 0, \quad q(\mathbf{U}) > 0. \tag{2.15}$$

(5) Let us reduce the degree of polynomial in the constraints by transferring successively the lower-order constraint on the root of a high-degree polynomial into the higher-order constraint on the root of a low-degree polynomial. Lemma 2.6 shows that the polynomial $f_3(\xi)$ has unique positive root, denoted by ξ_3. The continuity of $f_3(\xi)$ implies that for any $\xi > 0$, one has

$$f_3(\xi) < 0 \Leftrightarrow \xi < \xi_3 \quad \text{or} \quad f_3(\xi) > 0 \Leftrightarrow \xi > \xi_3. \tag{2.16}$$

Thus the first inequality in (2.15) is equivalent to

$$\xi_4(\mathbf{U}) < \xi_3(\mathbf{U}). \tag{2.17}$$

Lemma 2.4 yields

$$f_4(\xi) > 0 \Leftrightarrow \xi_4 < \xi,$$

for any $\xi > 0$. Therefore, (2.17) is equivalent to

$$f_4(\xi_3(\mathbf{U})) > 0. \tag{2.18}$$

If defining a quadratic polynomial $f_2(\xi)$ by

$$f_2(\xi) := 3\xi^2 + 4(|\mathbf{B}|^2 - E)\xi + |\mathbf{B}|^4 + D^2 + |\mathbf{m}|^2 - 2|\mathbf{B}|^2 E, \tag{2.19}$$

then one gets

$$\begin{aligned} f_4(\xi_3) &= \xi_3^2(\xi_3 + |\mathbf{B}|^2)^2 - [D^2(\xi_3 + |\mathbf{B}|^2)^2 + \xi_3^2|\mathbf{m}|^2 + (2\xi_3 + |\mathbf{B}|^2)(\mathbf{m}\cdot\mathbf{B})^2] \\ &= \xi_3^2(\xi_3 + |\mathbf{B}|^2)^2 - \xi_3^2(D^2 + |\mathbf{m}|^2) - [D^2|\mathbf{B}|^2 + (\mathbf{m}\cdot\mathbf{B})^2](2\xi_3 + |\mathbf{B}|^2) \\ &= \xi_3^2(\xi_3 + |\mathbf{B}|^2)^2 - \xi_3^2(D^2 + |\mathbf{m}|^2) - 2(\xi_3^3 + (|\mathbf{B}|^2 - E)\xi_3^2)(2\xi_3 + |\mathbf{B}|^2) \\ &= -\xi_3^2 f_2(\xi_3). \end{aligned}$$

Here the identity $f_3(\xi_3) = 0$ has been used in the third equal sign. Hence, (2.18) becomes

$$f_2(\xi_3(\mathbf{U})) < 0. \tag{2.20}$$

(6) Lemma 2.7 tells us that the polynomial $f_2(\xi)$ has two real roots, denoted by $\xi_{2,L}$ and $\xi_{2,R}$ with $\xi_{2,L} < \xi_{2,R}$. Because the graph of $f_2(\xi)$ opens upward, (2.20) is equivalent to

$$\xi_{2,L}(\mathbf{U}) < \xi_3(\mathbf{U}) < \xi_{2,R}(\mathbf{U}), \tag{2.21}$$

which implies

$$\xi_{2,R}(\mathbf{U}) > 0, \quad f_3(\xi_{2,R}(\mathbf{U})) > 0, \tag{2.22}$$

because of (2.16) and $\xi_3 > 0$. Conversely, one can show that (2.22) also implies (2.21), thus they are equivalent to each other. In fact, if (2.22) holds, one has $\xi_3 < \xi_{2,R}$ by using (2.16). Assume that (2.22) holds but (2.21) does not hold, then $\xi_{2,R} > \xi_{2,L} \geq \xi_3$. By using *Vieta's formula* for the quadratic polynomial that relate the coefficients of a polynomial to sums and products of its roots, $\xi_{2,M} := \frac{1}{2}(\xi_{2,L} + \xi_{2,R}) = -\frac{2}{3}(|\mathbf{B}|^2 - E) > \xi_3 > 0$. Due to (2.16), one has

$$f_3(\xi_{2,M}) > 0. \tag{2.23}$$

On the other hand, because

$$f_3'(\xi) = 3\xi^2 + 2(|\mathbf{B}|^2 - E)\xi = 3\xi(\xi - \xi_{2,M}),$$

the function $f_3(\xi)$ is strictly monotone decreasing in the interval $(0, \xi_{2,M})$, and thus

$$f_3(\xi_{2,M}) < f_3(0) = -\frac{|\mathbf{B}|^2 D^2 + (\mathbf{m}\cdot\mathbf{B})^2}{2} \leq 0,$$

which leads to a contradiction with (2.23). Therefore, (2.21) is equivalent to (2.22) under (2.2).

Because

$$\xi_{2,R} = \frac{\Phi(\mathbf{U}) - 2(|\mathbf{B}|^2 - E)}{3},$$

two inequalities in (2.22) become

$$\begin{aligned} &\Phi(\mathbf{U}) - 2(|\mathbf{B}|^2 - E) > 0, \\ &(\Phi(\mathbf{U}) - 2(|\mathbf{B}|^2 - E))^2(\Phi(\mathbf{U}) + (|\mathbf{B}|^2 - E)) > \frac{27}{2}(D^2|\mathbf{B}|^2 + (\mathbf{m}\cdot\mathbf{B})^2), \end{aligned} \tag{2.24}$$

which are equivalent to $\Psi(\mathbf{U}) > 0$ by noting that

$$\Phi(\mathbf{U}) + (|\mathbf{B}|^2 - E) > ||\mathbf{B}|^2 - E| + (|\mathbf{B}|^2 - E) \geq 0,$$

under $q(\mathbf{U}) > 0$. The proof is complete. □

The rest of this subsection gives all lemmas used in the proof of Theorem 2.1 and two remarks on Theorem 2.1 as well as a corollary.

Lemma 2.2. $\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \in \mathcal{G}$ *if and only if* $f_{\mathbf{U}}(\xi)$ *has unique zero* $\xi_*(\mathbf{U})$ *in* Ω_f *and satisfies*

$$D > 0, \quad q(\mathbf{U}) > 0, \quad \xi_*(\mathbf{U}) > 0, \quad f_4(\xi_*(\mathbf{U})) > 0, \tag{2.25}$$

where $f_4(\xi)$ *is a quartic polynomial defined in* (2.12).

Proof. (i) Assume $\mathbf{U} \in \mathcal{G}$. Lemma 2.1 shows that the first two inequalities in (2.25) hold. Because $\rho(\mathbf{U}) > 0$, $p(\mathbf{U}) > 0$, and $v(\mathbf{U}) < 1$, one has

$$\xi_* = \rho h W^2 = \frac{\rho(\mathbf{U})h(\mathbf{U})}{1 - v^2(\mathbf{U})} = \frac{\rho(\mathbf{U}) + \frac{\Gamma}{\Gamma - 1}p(\mathbf{U})}{1 - v^2(\mathbf{U})} > 0.$$

On the other hand, because of (2.9), and the facts that $\Gamma > 1$ and $v < 1$, one has $\xi_* > DW(\xi_*)$, which implies $f_4(\xi_*) > 0$.

(ii) Assume that four inequalities in (2.25) hold. Because of (2.12) and $D > 0$, $\xi_* > 0$, one has

$$f_\Omega(\xi_*) > f_\Omega(\xi_*) - D^2(\xi_* + |\mathbf{B}|^2)^2 = f_4(\xi_*) > 0,$$

which implies

$$W^{-2} = 1 - v^2(\mathbf{U}) = \frac{f_\Omega(\xi_*)}{\xi_*^2(\xi_* + |\mathbf{B}|^2)^2} > 0.$$

Thus $v(\mathbf{U}) < 1$ and $W(\xi_*) \geq 1$. Thanks to (2.8) and $D > 0$, one has $\rho(\mathbf{U}) = D/W(\xi_*) > 0$. Using (2.9) and $\Gamma > 1$ gives

$$\begin{aligned} p(\mathbf{U}) &= \frac{\Gamma - 1}{\Gamma W(\xi_*)}\left(\frac{\xi_*}{W(\xi_*)} - D\right) \\ &= \frac{\Gamma - 1}{\Gamma W(\xi_*)}\left(\frac{\xi_*}{W(\xi_*)} + D\right)^{-1}(\xi_*^2 W^{-2}(\xi_*) - D^2) \\ &= \frac{\Gamma - 1}{\Gamma W(\xi_*)}\left(\frac{\xi_*}{W(\xi_*)} + D\right)^{-1}\frac{f_4(\xi_*)}{(\xi_* + |\mathbf{B}|^2)^2} > 0. \end{aligned}$$

The proof is complete. □

Lemma 2.3. *For any* $\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \in \mathbb{R}^8$, *the set* Ω_f *in* (2.6) *is an open interval and can be expressed as*

$$\Omega_f = (\xi_\Omega, +\infty), \tag{2.26}$$

where $\xi_\Omega = \xi_\Omega(\mathbf{U})$ *is the biggest non-negative root of* $f_\Omega(\xi)$.

Proof. If $\mathbf{m} \cdot \mathbf{B} = 0$, then (2.5) gives $f_\Omega(\xi) = \xi^2(\xi + |\mathbf{B}|^2 + |\mathbf{m}|)(\xi + |\mathbf{B}|^2 - |\mathbf{m}|)$, whose biggest non-negative root is $\xi_\Omega = \max\{0, |\mathbf{m}| - |\mathbf{B}|^2\}$. Thus (2.26) holds.

Assume that $\mathbf{m}\cdot\mathbf{B}\neq 0$ and $\xi>0$. In this case, $|\mathbf{B}|\neq 0$ such that the expression of $f_\Omega(\xi)$ in (2.5) is reformulated as follows:

$$\begin{aligned} f_\Omega(\xi) &= \xi^2(\xi+|\mathbf{B}|^2)^2-\left[\xi^2|\mathbf{m}|^2+\frac{(\mathbf{m}\cdot\mathbf{B})^2}{|\mathbf{B}|^2}(2\xi|\mathbf{B}|^2+|\mathbf{B}|^4)\right] \\ &= \xi^2(\xi+|\mathbf{B}|^2)^2-\xi^2\left(|\mathbf{m}|^2-\frac{(\mathbf{m}\cdot\mathbf{B})^2}{|\mathbf{B}|^2}\right)-\frac{(\mathbf{m}\cdot\mathbf{B})^2}{|\mathbf{B}|^2}(\xi+|\mathbf{B}|^2)^2. \end{aligned} \tag{2.27}$$

Define

$$g_\Omega(\xi):=\left(1-\frac{(\mathbf{m}\cdot\mathbf{B})^2}{\xi^2|\mathbf{B}|^2}\right)(\xi+|\mathbf{B}|^2)^2-\left(|\mathbf{m}|^2-\frac{(\mathbf{m}\cdot\mathbf{B})^2}{|\mathbf{B}|^2}\right), \tag{2.28}$$

which implies $f_\Omega(\xi)=\xi^2 g_\Omega(\xi)$ and $g_\Omega(\xi)\leq 0$ for $0<\xi\leq|\mathbf{m}\cdot\mathbf{B}|/|\mathbf{B}|=:\zeta_0$. It is also easy to verify that $g_\Omega(\xi)$ satisfies

$$g_\Omega(\zeta_0)\leq 0,\quad \lim_{\xi\to+\infty} g_\Omega(\xi)=+\infty,$$

and is also strictly monotone increasing in the interval $[\zeta_0,+\infty)$, because the first term at the right-hand side of (2.28) is a product of two non-negative and strictly monotone increasing functions in $[\zeta_0,+\infty)$. The *intermediate value theorem* shows that $g_\Omega(\xi)$ has unique positive root $\xi_\Omega(\mathbf{U})$ in $[\zeta_0,+\infty)$, which is the biggest positive root of $f_\Omega(\xi)$ because of the relationship $f_\Omega(\xi)=\xi^2 g_\Omega(\xi)$. Therefore, the domain Ω_f can be equivalently replaced with (2.26). The proof is completed. □

Lemma 2.4. *If* $\mathbf{U}=(D,\mathbf{m},\mathbf{B},E)^\top\in\mathbb{R}^8$ *with* $D>0$, *then the quartic polynomial* $f_4(\xi)$ *defined in* (2.12) *has unique positive root* ξ_4, *satisfying* $\xi_4>\xi_\Omega$. *Moreover*, $f_4(\xi)>0$ *is equivalent to* $\xi_4<\xi$ *for any* $\xi\in\mathbb{R}^+$.

Proof. If $|\mathbf{B}|=0$, then $f_4(\xi)=\xi^2(\xi^2-(D^2+|\mathbf{m}|^2))$ has unique positive root $\xi_4(\mathbf{U})=\sqrt{D^2+|\mathbf{m}|^2}$, which satisfies $\xi_4(\mathbf{U})>|\mathbf{m}|=\xi_\Omega$. If $|\mathbf{B}|\neq 0$, then $f_4(\xi)$ is rewritten as follows:

$$f_4(\xi)=\xi^2 g_4(\xi),\quad \xi>0,$$

where the rational polynomial

$$g_4(\xi):=(1-\xi^{-2}\xi_0^2)(\xi+|\mathbf{B}|^2)^2-\left(|\mathbf{m}|^2-\frac{(\mathbf{m}\cdot\mathbf{B})^2}{|\mathbf{B}|^2}\right), \tag{2.29}$$

with

$$\xi_0:=\sqrt{D^2+(\mathbf{m}\cdot\mathbf{B})^2/|\mathbf{B}|^2}.$$

Obviously, if $\xi\in(0,\xi_0)$, then one has

$$g_4(\xi)<-\left(|\mathbf{m}|^2-\frac{(\mathbf{m}\cdot\mathbf{B})^2}{|\mathbf{B}|^2}\right)\leq 0.$$

Thus, the positive zero of $g_4(\xi)$ may be in the interval $[\xi_0, +\infty)$. The existence of the positive zero of $g_4(\xi)$ is verified as follows. It is easy to get that

$$g_4(\xi_0) \le 0, \quad \lim_{\xi \to +\infty} g_4(\xi) = +\infty.$$

On the other hand, the function $g_4(\xi)$ is strictly monotone increasing in the interval $[\xi_0, +\infty)$, because the first term at the right-hand side of (2.29) is a product of two positive and strictly monotone increasing functions in $[\xi_0, +\infty)$. The *intermediate value theorem* shows that $g_4(\xi)$ has unique positive root in $[\xi_0, +\infty)$, equivalently, $f_4(\xi)$ has unique positive root ξ_4. It satisfies

$$f_\Omega(\xi_4) = f_4(\xi_4) + D^2(\xi_4 + |\mathbf{B}|^2)^2 = D^2(\xi_4 + |\mathbf{B}|^2)^2 > 0,$$

which implies $\xi_4 \in \Omega_f$. Using Lemma 2.3 completes the proof. □

Lemma 2.5. *For any* $\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \in \mathbb{R}^8$ *with* $D > 0$, *the function* $f_{\mathbf{U}}(\xi)$ *defined in* (2.3) *is strictly monotone increasing in the interval* $\Omega_f = (\xi_\Omega, +\infty)$, *and* $\lim_{\xi \to +\infty} f_{\mathbf{U}}(\xi) = +\infty$.

Proof. From (2.3) and (2.4), the derivatives of $f_{\mathbf{U}}(\xi)$ and $W(\xi)$ with respect to ξ are calculated as follows:

$$f'_{\mathbf{U}}(\xi) = \Xi_\xi - \frac{\Gamma - 1}{\Gamma}\left(\frac{1}{W^2} - \frac{2\xi}{W^3}W'(\xi) + \frac{D}{W^2}W'(\xi)\right) \tag{2.30}$$

and

$$W'(\xi) = -W^3 \frac{(\mathbf{m}\cdot\mathbf{B})^2(3\xi^2 + 3\xi|\mathbf{B}|^2 + |\mathbf{B}|^4) + |\mathbf{m}|^2\xi^3}{\xi^3(\xi + |\mathbf{B}|^2)^3},$$

where

$$\Xi_\xi := 1 + \frac{|\mathbf{B}|^2}{W^3}W'(\xi) + \frac{(\mathbf{m}\cdot\mathbf{B})^2}{\xi^3}.$$

Let us prove that $\Xi_\xi > 0$ for any $\mathbf{B} \in \mathbb{R}^3$ and $\xi \in \Omega_f$. If $\mathbf{B} = \mathbf{0}$, then $\Xi_\xi = 1 > 0$. Assume that $\mathbf{B} \neq \mathbf{0}$ and thus (2.27) is available. Using (2.27) and $f_\Omega(\xi) > 0$ gives $|\mathbf{B}|^2\xi^2 - (\mathbf{m}\cdot\mathbf{B})^2 > 0$ and

$$(\xi + |\mathbf{B}|^2)^2 > \frac{\xi^2(|\mathbf{m}|^2|\mathbf{B}|^2 - (\mathbf{m}\cdot\mathbf{B})^2)}{|\mathbf{B}|^2\xi^2 - (\mathbf{m}\cdot\mathbf{B})^2}.$$

It follows that:

$$\begin{aligned}
\Xi_\xi &= \frac{(\xi + |\mathbf{B}|^2)^3 - (|\mathbf{m}|^2|\mathbf{B}|^2 - (\mathbf{m}\cdot\mathbf{B})^2)}{(\xi + |\mathbf{B}|^2)^3} \\
&> \frac{(\xi + |\mathbf{B}|^2)\frac{\xi^2(|\mathbf{m}|^2|\mathbf{B}|^2 - (\mathbf{m}\cdot\mathbf{B})^2)}{|\mathbf{B}|^2\xi^2 - (\mathbf{m}\cdot\mathbf{B})^2} - (|\mathbf{m}|^2|\mathbf{B}|^2 - (\mathbf{m}\cdot\mathbf{B})^2)}{(\xi + |\mathbf{B}|^2)^3} \\
&= \frac{(\xi^3 + (\mathbf{m}\cdot\mathbf{B})^2)(|\mathbf{m}|^2|\mathbf{B}|^2 - (\mathbf{m}\cdot\mathbf{B})^2)}{(\xi + |\mathbf{B}|^2)^3(|\mathbf{B}|^2\xi^2 - (\mathbf{m}\cdot\mathbf{B})^2)} \ge 0.
\end{aligned}$$

Because $\Gamma \in (1,2]$, $\frac{\Gamma}{\Gamma-1} \geq 2$. Noting that $W'(\xi) \leq 0$ for $\xi \in \Omega_f$ and using (2.30) give

$$\begin{aligned}\frac{\Gamma}{\Gamma-1} f'_{\mathbf{U}}(\xi) &\geq 2\Xi_\xi - \left(\frac{1}{W^2} - \frac{2\xi}{W^3}W'(\xi) + \frac{D}{W^2}W'(\xi)\right) \\ &\geq 2\Xi_\xi - \left(\frac{1}{W^2} - \frac{2\xi}{W^3}W'(\xi)\right) \\ &= 2\left[1 + \frac{(\mathbf{m}\cdot\mathbf{B})^2}{\xi^3} - \frac{(\mathbf{m}\cdot\mathbf{B})^2(3\xi^2 + 3\xi|\mathbf{B}|^2 + |\mathbf{B}|^4) + |\mathbf{m}|^2\xi^3}{\xi^3(\xi+|\mathbf{B}|^2)^2}\right] - \frac{1}{W^2} \\ &= \frac{2}{W^2} - \frac{1}{W^2} = \frac{f_\Omega(\xi)}{\xi^2(\xi+|\mathbf{B}|^2)^2} > 0,\end{aligned}$$

which implies $f'_{\mathbf{U}}(\xi) > 0$ and $f_{\mathbf{U}}(\xi)$ is strictly monotone increasing in the interval Ω_f. Note that

$$f_{\mathbf{U}}(\xi) > \left(1 - \frac{\Gamma-1}{\Gamma W^2}\right)\xi - \frac{1}{2}\left[\frac{|\mathbf{B}|^2}{W^2} + \frac{(\mathbf{m}\cdot\mathbf{B})^2}{\xi^2}\right] - E \to +\infty, \quad \text{as } \xi \to +\infty,$$

where $\lim_{\xi\to+\infty} W(\xi) = 1$ has been used. This implies $\lim_{\xi\to+\infty} f_{\mathbf{U}}(\xi) = +\infty$ and the proof is complete. □

Lemma 2.6. *If* $\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \in \mathbb{R}^8$ *satisfying* (2.2), *then the cubic polynomial* $f_3(\xi)$ *defined in* (2.14) *has unique positive root* ξ_3.

Proof. If $\mathbf{B} = \mathbf{0}$, then $f_3(\xi) = \xi^2(\xi - E)$ with unique positive root $\xi_3 = E$. If $\mathbf{B} \neq \mathbf{0}$, then $f_3(\xi)$ is rewritten as follows:

$$f_3(\xi) = \xi^2 g_3(\xi), \quad \xi > 0,$$

with the rational polynomial

$$g_3(\xi) := \xi - \frac{D^2|\mathbf{B}|^2 + (\mathbf{m}\cdot\mathbf{B})^2}{2\xi^2} + |\mathbf{B}|^2 - E,$$

which is strictly monotone increasing in $\mathbb{R}^+$ and satisfies

$$\lim_{\xi\to 0^+} g_3(\xi) = -\infty, \quad \lim_{\xi\to+\infty} g_3(\xi) = +\infty.$$

According to the *intermediate value theorem*, $g_3(\xi)$ has unique positive root, and thus $f_3(\xi)$ has unique positive root in $\mathbb{R}^+$. The proof is complete. □

Lemma 2.7. *If* $q(\mathbf{U}) > 0$, *then the quadratic polynomial* $f_2(\xi)$ *defined in* (2.19) *has two real roots.*

Proof. Because the discriminant of the quadratic polynomial $f_2(\xi)$ is

$$\begin{aligned}\Delta &= 16(|\mathbf{B}|^2 - E)^2 - 12(|\mathbf{B}|^4 + D^2 + |\mathbf{m}|^2 - 2|\mathbf{B}|^2 E)\\ &= 4(|\mathbf{B}|^2 - E)^2 + 12(E^2 - (D^2 + |\mathbf{m}|^2))\\ &\geq 12q(\mathbf{U})(E + \sqrt{D^2 + |\mathbf{m}|^2}) \geq 12q^2(\mathbf{U}) > 0,\end{aligned}$$

the function $f_2(\xi)$ has two real roots. □

Remark 2.1. Using (2.24) and some algebraic manipulations, one can verify that the constraint $\Psi(\mathbf{U}) > 0$ is equivalent to two constraints $\hat{q}(\mathbf{U}) > 0$ and $\tilde{q}(\mathbf{U}) > 0$, where

$$\begin{aligned}\hat{q}(\mathbf{U}) &:= \Phi(\mathbf{U}) - 2(|\mathbf{B}|^2 - E)\\ &= \sqrt{(E - |\mathbf{B}|^2)^2 + 3(E^2 - D^2 - |\mathbf{m}|^2)} + 2(E - |\mathbf{B}|^2),\\ \tilde{q}(\mathbf{U}) &:= \Phi^6(\mathbf{U}) - \bigg((E - |\mathbf{B}|^2)^3 + \frac{27}{2}(|\mathbf{B}|^2 D^2 + |\mathbf{m}\cdot\mathbf{B}|^2)\\ &\qquad - 9(E^2 - D^2 - |\mathbf{m}|^2)(E - |\mathbf{B}|^2)\bigg)^2.\end{aligned}$$

Moreover, $\Psi(\mathbf{U}) = 0$ if and only if $\hat{q}(\mathbf{U}) \geq 0$ and $\tilde{q}(\mathbf{U}) = 0$.

Remark 2.2. The first equivalent definition of $\mathcal{G}$ is very important in the following aspects:

- To guide the initial guess in numerically solving the nonlinear algebraic equation (2.3), because the proof of Theorem 2.1 has shown that $\xi_* > \xi_4$ for $\mathbf{U} \in \mathcal{G}$, where ξ_4 is discussed in Lemma 2.4, and

$$\begin{aligned}\Gamma E - \xi_*(\mathbf{U}) &= \Gamma\bigg(\rho h W^2 - p + \frac{1+v^2}{2}|\mathbf{B}|^2 - \frac{1}{2}(\mathbf{v}\cdot\mathbf{B})^2\bigg) - \rho h W^2\\ &\geq \Gamma(\rho h W^2 - p) - \rho h W^2 \geq (\Gamma - 1)\rho h - \Gamma p = (\Gamma - 1)\rho > 0.\end{aligned}$$

- To develop the PCP limiter and high-order accurate PCP schemes for the 1D RMHD equations (1.1), see Sec. 3.1.2.
- To prove the convexity of $\mathcal{G}$, see Sec. 2.2, and the scaling invariance.

Corollary 2.1. (Scaling invariance) *If the state* $\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \in \mathcal{G}_0$, *then for any* $\lambda \in \mathbb{R}^+$, *the state* $\mathbf{U}_\lambda := (\lambda D, \lambda\mathbf{m}, \sqrt{\lambda}\mathbf{B}, \lambda E)^\top \in \mathcal{G}_0$.

Proof. It can be verified that $q(\mathbf{U}_\lambda) = \lambda q(\mathbf{U}) > 0$ and $\Psi(\mathbf{U}_\lambda) = \lambda^{3/2}\Psi(\mathbf{U}) > 0$. The proof is complete. □

2.2. *Convexity*

This section will prove the convexity of admissible state set $\mathcal{G}_0 = \mathcal{G}$ for the RMHDs. It will play a pivotal role in analyzing the PCP property of numerical schemes.

Theorem 2.2. *The admissible state set $\mathcal{G}_0$ is a convex set.*

Proof. It is not trivial and cannot be completed by using the convexity definition of the set because of the strong nonlinearity of the function $\Psi(\mathbf{U})$ used in (2.10). Instead, it will be done with the aid of the close connection between the set convexity in $\mathbb{R}^N$ and the concave–convex character of the region boundary corresponding to the set, see e.g. Ref. 23.

It is easy to show by the proof of Lemma 2.2 in Ref. 40 that the set

$$\mathcal{G}_2 := \{\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \in \mathbb{R}^8 \mid D > 0, q(\mathbf{U}) > 0\}$$

is convex. Therefore, the subsequent task is to prove that the hypersurface $\mathcal{S}$ in $\mathbb{R}^8$ described by

$$\Psi(\mathbf{U}) = 0 \tag{2.31}$$

is convex within the region $\mathcal{G}_2$, and the points in $\mathcal{G}_0$ are all located in the concave side of the hypersurface $\mathcal{S}$. Unfortunately, it is impractical to check the convexity of the hypersurface $\mathcal{S}$ by directly using the highly nonlinear equation (2.31) via the theory of geometry. To overcome this difficulty, we try to give a parameter equation for the hypersurface $\mathcal{S}$.

An important discovery is that (2.31) is equivalent to $p(\mathbf{U}) = 0$ for $\mathbf{U} \in \mathcal{G}_2$. In fact, on the one hand, it can be seen from the proof of Theorem 2.1 that (2.31) implies $f_3(\xi_{2,R}(\mathbf{U})) = 0$. It means that $\xi_3(\mathbf{U}) = \xi_{2,R}(\mathbf{U})$ and satisfies $f_2(\xi_3(\mathbf{U})) = 0$, which yields $f_4(\xi_3(\mathbf{U})) = 0$. It follows that $\xi_4(\mathbf{U}) = \xi_3(\mathbf{U})$ and satisfies $f_3(\xi_4(\mathbf{U})) = 0$ and $f_{\mathbf{U}}(\xi_4) = 0$. This further implies $\xi_*(\mathbf{U}) = \xi_4(\mathbf{U})$, and thus one has that $\xi_* = DW(\xi_*)$ and $p(\mathbf{U}) = 0$. On the other hand, if $p(\mathbf{U}) = 0$, then $h = 1 + e + p/\rho = 1$, and

$$\begin{cases} D = \dfrac{\rho}{\sqrt{1-v^2}}, \\ \mathbf{m} = \dfrac{\rho\mathbf{v}}{1-v^2} + |\mathbf{B}|^2\mathbf{v} - (\mathbf{v}\cdot\mathbf{B})\mathbf{B}, \\ E = \rho W^2 - p_m + |\mathbf{B}|^2 = \dfrac{\rho}{1-v^2} + \dfrac{1+v^2}{2}|\mathbf{B}|^2 - \dfrac{(\mathbf{v}\cdot\mathbf{B})^2}{2}. \end{cases} \tag{2.32}$$

Thus one has

$$\Phi^2(\mathbf{U}) = (E - |\mathbf{B}|^2)^2 + 3(E^2 - D^2 - |\mathbf{m}|^2) = (\rho W^2 + 2p_m)^2,$$

$$(E - |\mathbf{B}|^2)^3 + \frac{27}{2}(|\mathbf{B}|^2 D^2 + |\mathbf{m}\cdot\mathbf{B}|^2) - 9(E^2 - D^2 - |\mathbf{m}|^2)(E - |\mathbf{B}|^2)$$
$$= (\rho W^2 + 2p_m)^3,$$

which imply that $\hat{q}(\mathbf{U})$ and $\tilde{q}(\mathbf{U})$ in Remark 2.1 satisfy

$$\hat{q}(\mathbf{U}) = \sqrt{(E-|\mathbf{B}|^2)^2 + 3(E^2 - D^2 - |\mathbf{m}|^2)} + 2(E-|\mathbf{B}|^2) = 3\rho W^2 > 0$$

and $\tilde{q}(\mathbf{U}) = 0$. The conclusion in Remark 2.1 yields (2.31).

Based on the above discovery, the hypersurface $\mathcal{S}$ defined in (2.31) can be represented by the parametric equations (2.32) through seven parameters $\boldsymbol{\mathcal{V}} := (\rho, \mathbf{v}, \mathbf{B})^\top$ with $\rho > 0, |\mathbf{v}| < 1$ and $\mathbf{B} \in \mathbb{R}^3$. Obviously, the hypersurface $\mathcal{S}$ is not 6-cylindrical. Based on the theorem in Ref. 23, one only needs to show that its second fundamental form is positive semi-definite, i.e. prove that the matrix

$$\mathbf{\Pi} := \left[\sum_{l=1}^{8} \frac{\partial^2 U_{\langle l\rangle}}{\partial \mathcal{V}_{\langle i\rangle} \partial \mathcal{V}_{\langle j\rangle}} n_l\right]_{7\times 7}$$

is positive semi-definite, where $U_{\langle l\rangle}$ and $\mathcal{V}_{\langle i\rangle}$ denote the lth component of the vector $\mathbf{U}$ and the ith component of the vector $\boldsymbol{\mathcal{V}}$, respectively, and $\mathbf{n} := (n_1, n_2, \ldots, n_8)^\top$ represents the inward-pointing (to the region $\mathcal{G}_0$) normal vector of the hypersurface $\mathcal{S}$. Taking partial derivatives of $\mathbf{U}$ with respect to $\mathcal{V}_{\langle i\rangle}$ gives

$$\begin{aligned}
\partial_\rho \mathbf{U} &= (W, W^2 v_1, W^2 v_2, W^2 v_3, 0, 0, 0, W^2)^\top,\\
\partial_{v_1} \mathbf{U} &= (\rho W^3 v_1, \rho W^2(1+2W^2 v_1^2) + B_2^2 + B_3^2, 2\rho W^4 v_1 v_2 - B_1 B_2,\\
&\quad 2\rho W^4 v_1 v_3 - B_1 B_3, 0, 0, 0, |\mathbf{B}|^2 v_1 - B_1(\mathbf{v}\cdot\mathbf{B}) + 2\rho W^4 v_1)^\top,\\
\partial_{v_2} \mathbf{U} &= (\rho W^3 v_2, 2\rho W^4 v_1 v_2 - B_1 B_2, \rho W^2(1+2W^2 v_2^2) + B_1^2 + B_3^2,\\
&\quad 2\rho W^4 v_2 v_3 - B_2 B_3, 0, 0, 0, |\mathbf{B}|^2 v_2 - B_2(\mathbf{v}\cdot\mathbf{B}) + 2\rho W^4 v_2)^\top,\\
\partial_{v_3} \mathbf{U} &= (\rho W^3 v_3, 2\rho W^4 v_1 v_3 - B_1 B_3, 2\rho W^4 v_2 v_3 - B_2 B_3,\\
&\quad \rho W^2(1+2W^2 v_3^2) + B_1^2 + B_2^2, 0, 0, 0, |\mathbf{B}|^2 v_3 - B_3(\mathbf{v}\cdot\mathbf{B}) + 2\rho W^4 v_3)^\top,\\
\partial_{B_1} \mathbf{U} &= (0, -B_2 v_2 - B_3 v_3, 2B_1 v_2 - B_2 v_1, 2B_1 v_3 - B_3 v_1,\\
&\quad 1, 0, 0, B_1(1+v^2) - v_1(\mathbf{v}\cdot\mathbf{B}))^\top,\\
\partial_{B_2} \mathbf{U} &= (0, 2B_2 v_1 - B_1 v_2, -B_1 v_1 - B_3 v_3, 2B_2 v_3 - B_3 v_2,\\
&\quad 0, 1, 0, B_2(1+v^2) - v_2(\mathbf{v}\cdot\mathbf{B}))^\top,\\
\partial_{B_3} \mathbf{U} &= (0, 2B_3 v_1 - B_1 v_3, 2B_3 v_2 - B_2 v_3, -B_1 v_1 - B_2 v_2,\\
&\quad 0, 0, 1, B_3(1+v^2) - v_3(\mathbf{v}\cdot\mathbf{B}))^\top.
\end{aligned}$$

These are seven tangent vectors of the hypersurface $\mathcal{S}$ and generate the local tangent space. Because they are perpendicular to the normal vector $\mathbf{n}$, their inner products with $\mathbf{n}$ should be equal to zero, and thus a linear system of seven algebraic equations for $(n_1, n_2, \ldots, n_8)^\top$ is formed. Solving this linear system gives

$$\mathbf{n} = (-\sqrt{1-v^2}, -\mathbf{v}, -(1-v^2)\mathbf{B} - (\mathbf{v}\cdot\mathbf{B})\mathbf{v}, 1)^\top. \tag{2.33}$$

First, let us check the positive semi-definiteness of $\mathbf{\Pi}$. Taking the second-order partial derivatives of $\mathbf{U}$ with respect to $\boldsymbol{\mathcal{V}}$, and then calculating their inner products with $\mathbf{n}$ give the expression of the matrix $\mathbf{\Pi}$ as follows:

$$\mathbf{\Pi} = \mathrm{diag}\{0, \mathbf{\Pi}_1, \mathbf{\Pi}_2\},$$

where

$$\mathbf{\Pi}_1 = \rho W^4[(1-v^2)\mathbf{I}_3 + \mathbf{v}^\top\mathbf{v}] + |\mathbf{B}|^2\mathbf{I}_3 - \mathbf{B}^\top\mathbf{B} = \rho W^4\mathbf{\Pi}_2 + |\mathbf{B}|^2\mathbf{I}_3 - \mathbf{B}^\top\mathbf{B},$$
$$\mathbf{\Pi}_2 = (1-v^2)\mathbf{I}_3 + \mathbf{v}^\top\mathbf{v}.$$

Here $\mathbf{I}_3$ denotes a unit matrix of size 3. The matrix $\mathbf{v}^\top\mathbf{v}$ has rank of 1 and eigenvalues of $\{0, 0, |\mathbf{v}|^2\}$, so the eigenvalues of $\mathbf{\Pi}_2$ are $\{1-|\mathbf{v}|^2, 1-|\mathbf{v}|^2, 1\}$, which imply the positive definiteness of $\mathbf{\Pi}_2$. Similarly, one can show that the eigenvalues of $|\mathbf{B}|^2\mathbf{I}_3 - \mathbf{B}^\top\mathbf{B}$ are $\{|\mathbf{B}|^2, |\mathbf{B}|^2, 0\}$. Because $\mathbf{\Pi}_1$ is the sum of a positive definite matrix and a positive semi-definite matrix, it is positive semi-definite. In conclusion, $\mathbf{\Pi}$ is a positive semi-definite matrix with positive inertia index of 6 so that the hypersurface $\mathcal{S}$ described in (2.31) is a convex surface in $\mathcal{G}_2$.

Next, let us prove that all the points in $\mathcal{G}_0 = \mathcal{G}$ are located at the concave side of the hypersurface $\mathcal{S}$, that is to say, the normal vector $\mathbf{n}$ in (2.33) is the inward-pointing vector to the region $\mathcal{G}_0$. For this purpose, we need to show that, for any $\tilde{\mathbf{U}} \in \mathcal{G}_0 = \mathcal{G}$ and $\mathbf{U} \in \mathcal{S}$, it holds that

$$(\tilde{\mathbf{U}} - \mathbf{U}) \cdot \mathbf{n} > 0,$$

which is equivalent to

$$\tilde{\mathcal{F}}(\tilde{\rho}, \tilde{p}, \tilde{\mathbf{v}}, \tilde{\mathbf{B}}, \mathbf{v}, \mathbf{B}) := \tilde{\mathbf{U}} \cdot \mathbf{n} + p_m > 0,$$

because of (2.32) and (2.33). By defining $\tilde{\mathcal{F}}_0(\tilde{\mathbf{v}}, \tilde{\mathbf{B}}, \mathbf{v}, \mathbf{B}) := \tilde{\mathcal{F}}(0, 0, \tilde{\mathbf{v}}, \tilde{\mathbf{B}}, \mathbf{v}, \mathbf{B})$, one can infer that

$$\begin{aligned}
\tilde{\mathcal{F}}_0(\tilde{\mathbf{v}}, \tilde{\mathbf{B}}, \mathbf{v}, \mathbf{B}) &= (|\tilde{\mathbf{B}}|^2\tilde{\mathbf{v}} - (\tilde{\mathbf{v}}\cdot\tilde{\mathbf{B}})\tilde{\mathbf{B}}) \cdot (-\mathbf{v}) + (W^{-2}\mathbf{B} + (\mathbf{v}\cdot\mathbf{B})\mathbf{v}) \cdot (-\tilde{\mathbf{B}}) \\
&\quad + \frac{(1+\tilde{v}^2)|\tilde{\mathbf{B}}|^2 - (\tilde{\mathbf{v}}\cdot\tilde{\mathbf{B}})^2}{2} + \frac{(1-v^2)|\mathbf{B}|^2 + (\mathbf{v}\cdot\mathbf{B})^2}{2} \\
&= \frac{(1-v^2)|\mathbf{B}-\tilde{\mathbf{B}}|^2}{2} + \frac{|\mathbf{v}-\tilde{\mathbf{v}}|^2|\tilde{\mathbf{B}}|^2}{2} \\
&\quad + \frac{(\mathbf{v}\cdot\mathbf{B})^2}{2} - (\mathbf{v}\cdot\mathbf{B})(\mathbf{v}\cdot\tilde{\mathbf{B}}) - \frac{(\tilde{\mathbf{v}}\cdot\tilde{\mathbf{B}})^2}{2} + (\tilde{\mathbf{v}}\cdot\tilde{\mathbf{B}})(\mathbf{v}\cdot\tilde{\mathbf{B}}) \\
&\geq \frac{|\mathbf{v}-\tilde{\mathbf{v}}|^2|\tilde{\mathbf{B}}|^2}{2} + \frac{[(\mathbf{v}\cdot\mathbf{B}) - (\mathbf{v}\cdot\tilde{\mathbf{B}})]^2}{2} - \frac{[(\tilde{\mathbf{v}}\cdot\tilde{\mathbf{B}}) - (\mathbf{v}\cdot\tilde{\mathbf{B}})]^2}{2} \\
&= \frac{|\mathbf{v}-\tilde{\mathbf{v}}|^2|\tilde{\mathbf{B}}|^2}{2} - \frac{((\mathbf{v}-\tilde{\mathbf{v}})\cdot\tilde{\mathbf{B}})^2}{2} + \frac{(\mathbf{v}\cdot(\mathbf{B}-\tilde{\mathbf{B}}))^2}{2} \geq 0.
\end{aligned}$$

Thus for any given $\mathbf{U}$ on the hypersurface $\mathcal{S}$, one has

$$\begin{aligned}\tilde{\mathbf{U}}\cdot\mathbf{n}+p_m &= \tilde{\rho}\tilde{W}^2(1-\tilde{\mathbf{v}}\cdot\mathbf{v}-\tilde{W}^{-1}W^{-1})\\ &\quad+\tilde{p}\left(\frac{\Gamma}{\Gamma-1}\tilde{W}^2(1-\tilde{\mathbf{v}}\cdot\mathbf{v})-1\right)+\tilde{\mathcal{F}}_0(\tilde{\mathbf{v}},\tilde{\mathbf{B}},\mathbf{v},\mathbf{B})\\ &\geq \tilde{\rho}\tilde{W}^2(1-(\tilde{v}_1,\tilde{v}_2,\tilde{v}_3,\tilde{W}^{-1})\cdot(v_1,v_2,v_3,W^{-1}))\\ &\quad+\tilde{p}(2\tilde{W}^2(1-\tilde{\mathbf{v}}\cdot\mathbf{v})-1)\\ &\geq \tilde{\rho}\tilde{W}^2(1-|(\tilde{v}_1,\tilde{v}_2,\tilde{v}_3,\tilde{W}^{-1})||(v_1,v_2,v_3,W^{-1})|)\\ &\quad+\tilde{p}(2\tilde{W}^2(1-|\tilde{\mathbf{v}}||\mathbf{v}|)-1)\\ &\geq \tilde{p}(2\tilde{W}^2(1-|\tilde{\mathbf{v}}|)-1)=\frac{\tilde{p}(1-|\tilde{\mathbf{v}}|)}{1+|\tilde{\mathbf{v}}|}>0.\end{aligned}$$

The proof is complete. □

2.3. *Second equivalent definition*

The convexity of the admissible state set $\mathcal{G}$ can give its second equivalent form, whose importance lies in that all constraints are linear with respect to $\mathbf{U}$ so that it will be very effective in verifying theoretically the PCP property of the numerical schemes for the RMHD equations (1.1).

Theorem 2.3. (Second equivalent definition) *The admissible state set $\mathcal{G}$ or $\mathcal{G}_0$ is equivalent to the set*

$$\begin{aligned}\mathcal{G}_1 := \{\mathbf{U}=(D,\mathbf{m},\mathbf{B},E)^\top\in\mathbb{R}^8 \,|\, D>0, \mathbf{U}\cdot\mathbf{n}^*+p_m^*>0,\\ \text{for any } \mathbf{B}^*,\mathbf{v}^*\in\mathbb{R}^3 \text{ with } |\mathbf{v}^*|<1\},\end{aligned}\tag{2.34}$$

where

$$\mathbf{n}^*=(-\sqrt{1-|\mathbf{v}^*|^2},-\mathbf{v}^*,-(1-|\mathbf{v}^*|^2)\mathbf{B}^*-(\mathbf{v}^*\cdot\mathbf{B}^*)\mathbf{v}^*,1)^\top,\tag{2.35}$$

$$p_m^*=\frac{(1-|\mathbf{v}^*|^2)|\mathbf{B}^*|^2+(\mathbf{v}^*\cdot\mathbf{B}^*)^2}{2}.\tag{2.36}$$

Here $\mathbf{U}^*$ denotes any point on the hypersurface $\mathcal{S}$, and $p_m^*=-\mathbf{U}^*\cdot\mathbf{n}^*$ and $\mathbf{n}^*$ are corresponding magnetic pressure and inward-pointing vector to the region $\mathcal{G}_0$, respectively.

Proof. Theorem 2.2 and its proof have shown that $\mathcal{G}_0=\mathcal{G}\subseteq\mathcal{G}_1$. The subsequent task is to prove $\mathcal{G}_1\subseteq\mathcal{G}_0$. For any $\mathbf{U}\in\mathcal{G}_1$, the convexity of the hypersurface $\mathcal{S}$ in (2.31) implies the constraint $\Psi(\mathbf{U})>0$ in $\mathcal{G}_0$. Thus it needs to prove that the state $\mathbf{U}\in\mathcal{G}_1$ satisfies $q(\mathbf{U})>0$. If taking the vectors $\mathbf{B}^*,\mathbf{v}^*\in\mathbb{R}^3$ as

$$\mathbf{B}^*=\mathbf{0},\quad \mathbf{v}^*=\frac{1}{\sqrt{D^2+|\mathbf{m}|^2}}\mathbf{m},$$

and substituting them into the second inequality in $\mathcal{G}_1$, one has

$$\begin{aligned} 0 < \mathbf{U}\cdot\mathbf{n}^* + p_m^* &= E - \mathbf{m}\cdot\mathbf{v}^* - D\sqrt{1-|\mathbf{v}^*|^2} \\ &= E - \frac{|\mathbf{m}|^2}{\sqrt{D^2+|\mathbf{m}|^2}} - \frac{D^2}{\sqrt{D^2+|\mathbf{m}|^2}} \\ &= E - \sqrt{D^2+|\mathbf{m}|^2} = q(\mathbf{U}). \end{aligned}$$

The proof is complete. □

Remark 2.3. It is seen that $\mathbf{n}^*$ in (2.35) can be rewritten as

$$\mathbf{n}^* = -\sqrt{1-|\mathbf{v}^*|^2}(1, u_1^*, u_2^*, u_3^*, b_1^*, b_2^*, b_3^*, u_0^*)^\top,$$

where u_α^* and b_α^* denote the velocity and magnetic field in 4D space-time, respectively.

Remark 2.4. Theorems 2.1 and 2.3 indicate that $\mathcal{G} = \mathcal{G}_0 = \mathcal{G}_1$. Thus they will not be deliberately distinguished henceforth.

Theorem 2.3 implies the following orthogonal invariance of the admissible state set $\mathcal{G}_1$.

Corollary 2.2. (Orthogonal invariance) *Let* $\mathbf{T} := \mathrm{diag}\{1, \mathbf{T}_3, \mathbf{T}_3, 1\}$, *where* $\mathbf{T}_3$ *denotes any orthogonal matrix of size* 3. *If* $\mathbf{U} \in \mathcal{G}_1$, *then* $\mathbf{TU} \in \mathcal{G}_1$.

Proof. For any $\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \in \mathcal{G}_1$, if denoting $\bar{\mathbf{U}} = \mathbf{TU} =: (\bar{D}, \bar{\mathbf{m}}, \bar{\mathbf{B}}, \bar{E})^\top$, then $\bar{D} = D > 0$. For any $\mathbf{B}^*, \mathbf{v}^* \in \mathbb{R}^3$ with $|\mathbf{v}^*| < 1$, if denoting $\hat{\mathbf{B}}^* := \mathbf{B}^*\mathbf{T}_3$, $\hat{\mathbf{v}}^* := \mathbf{v}^*\mathbf{T}_3$, then $|\hat{\mathbf{v}}^*| = |\mathbf{v}^*| < 1$, $\hat{p}_m^* = p_m^*$, and $\hat{\mathbf{n}}^* = \mathbf{T}^{-1}\mathbf{n}^*$. Using Theorem 2.3 for $\mathbf{U} \in \mathcal{G}_1$ gives

$$0 < \mathbf{U}\cdot\hat{\mathbf{n}}^* + \hat{p}_m^* = (\mathbf{T}^{-1}\bar{\mathbf{U}})\cdot(\mathbf{T}^{-1}\mathbf{n}^*) + p_m^* = \bar{\mathbf{U}}\cdot\mathbf{n}^* + p_m^*,$$

where the orthogonality of $\mathbf{T}^{-1}$ has been used in the last equality. Hence using Theorem 2.3 again yields $\bar{\mathbf{U}} \in \mathcal{G}_1$. The proof is complete. □

Remark 2.5. Corollary 2.2 implies the rotational or symmetric invariance of the admissible state set $\mathcal{G}$ if $\mathbf{T}_3$ is taken as a rotational or symmetric matrix of size 3.

2.4. *Generalized Lax–Friedrichs splitting properties*

The section utilizes the second equivalent definition of $\mathcal{G}$ in Theorem 2.3 to present the generalized LxF splitting properties of the admissible state set $\mathcal{G}$ for the special RMHD equations (1.1).

Lemma 2.8. (LxF splitting) *If* $\mathbf{B} = \mathbf{0}$, *then the special RMHD equations* (1.1) *satisfy the* LxF splitting property:

$$\mathbf{U} \pm \alpha^{-1}\mathbf{F}_i(\mathbf{U}) \in \mathcal{G} \quad \text{for } \mathbf{U} \in \mathcal{G},$$

where $\alpha \geq \varrho_i$ *and* ϱ_i *denotes a proper upper bound of the spectral radius of the Jacobian matrix* $\partial\mathbf{F}_i/\partial\mathbf{U}$, $i = 1, 2, 3$. *If* $\mathbf{B} \neq \mathbf{0}$, *then the* LxF splitting property *does not always hold.*

Proof. The first part has been proved in Ref. 40, while the second part is proved by contradiction as follows.

Assume that the LxF splitting property holds for $\mathbf{U} \in \mathcal{G}$ and $\Gamma \in (1, 2]$. For any $\mathbf{V} = (\rho, \mathbf{v}, \mathbf{B}, p)^\top$ satisfying $\rho > 0$, $p > 0$, and $v < 1$, one has

$$\mathbf{U}(\mathbf{V}) \pm \alpha^{-1}\mathbf{F}_i(\mathbf{U}(\mathbf{V})) \in \mathcal{G}, \quad \forall\, \alpha \geq \varrho_i, \quad i = 1, 2, 3.$$

Because the speed of light $c = 1$ is a rigorous bound of the spectral radius of $\partial\mathbf{F}_i/\partial\mathbf{U}$, one can specially take $\rho = p = \epsilon > 0$, $\mathbf{v} = (0.5, 0, 0)$, $\mathbf{B} = (1, 0, 0)$, $\alpha = 1/\theta$ for $\theta \in (0, 1]$, and $\Gamma = 5/3$, such that:

$$\begin{aligned}\mathbf{U}^{\pm}(\epsilon, \theta) &:= \mathbf{U} \pm \alpha^{-1}\mathbf{F}_1(\mathbf{U}) \\ &= \left(\frac{\sqrt{3}}{3}(2+\theta)\epsilon, \frac{14 \pm 13\theta}{6}\epsilon \mp \frac{\theta}{2}, 0, 0, 1, 0, 0, \frac{11 \pm 7\theta}{3}\epsilon + \frac{1}{2}\right)^\top \in \mathcal{G} = \mathcal{G}_0.\end{aligned}$$

According to Remark 2.1, one has $\tilde{q}(\mathbf{U}^{\pm}(\epsilon, \theta)) > 0$, for all $\epsilon > 0$ and $\theta \in (0, 1]$. The continuity of $\tilde{q}(\mathbf{U})$ with respect to $\mathbf{U}$ further implies that for any fixed θ, $\mathbf{U}^{\pm}(\epsilon, \theta)$ is also continuous with respect to ϵ. Therefore

$$0 \leq \lim_{\epsilon \to 0^+} \tilde{q}(\mathbf{U}^{\pm}(\epsilon, \theta)) = \tilde{q}(\mathbf{U}^{\pm}(0, \theta)) = -\frac{27}{64}\theta^2(\theta^2 + 4)^2 < 0,$$

which leads to the contradiction. Hence the LxF splitting property does not hold for the admissible state set $\mathcal{G}$ for the RMHD equations (1.1) in general. □

Although the LxF splitting property may not hold for the nonzero magnetic field, we discover the generalized LxF splitting properties which are coupling the convex combination of some LxF splitting terms with a "discrete divergence-free" condition for the magnetic field vector $\mathbf{B}$. However, it is extremely difficult and technical because of the "discrete divergence-free" condition for the magnetic field $\mathbf{B}$ and the strong nonlinearity in the constraints of the admissible state set and $\mathbf{F}_i(\mathbf{U})$, etc. Their breakthrough is made by a constructive inequality in the following lemma.

Lemma 2.9. *If* $\mathbf{U} \in \mathcal{G}$, *then for any* $\theta \in [-1, 1]$ *and* $\mathbf{B}^*$, $\mathbf{v}^* \in \mathbb{R}^3$ *with* $|\mathbf{v}^*| < 1$ *it holds*

$$(\mathbf{U} + \theta\mathbf{F}_i(\mathbf{U})) \cdot \mathbf{n}^* + p_m^* + \theta(v_i^* p_m^* - B_i(\mathbf{v}^* \cdot \mathbf{B}^*)) > 0, \tag{2.37}$$

where $i \in \{1, 2, 3\}$, *and* $\mathbf{n}^*$ *and* p_m^* *are defined in* (2.35) *and* (2.36), *respectively.*

Proof. (i) First let us prove the inequality (2.37) for the case of $i = 1$, i.e.

$$\mathcal{H}(\rho, p, \mathbf{v}, \mathbf{B}, \mathbf{v}^*, \mathbf{B}^*, \theta) := (\mathbf{U} + \theta \mathbf{F}_1(\mathbf{U})) \cdot \mathbf{n}^* + (1 + \theta v_1^*) p_m^* - \theta B_1 (\mathbf{v}^* \cdot \mathbf{B}^*) > 0. \tag{2.38}$$

Taking partial derivatives of $\mathcal{H}$ with respect to ρ and p respectively gives

$$\begin{aligned}
\frac{\partial \mathcal{H}}{\partial \rho} &= (1 + \theta v_1) W^2 (1 - \mathbf{v} \cdot \mathbf{v}^* - W^{-1} (W^{-1})^*) \\
&\geq (1 + \theta v_1) W^2 (1 - (v_1, v_2, v_3, W^{-1}) \cdot (v_1^*, v_2^*, v_3^*, (W^{-1})^*)) \\
&\geq (1 + \theta v_1) W^2 (1 - |(v_1, v_2, v_3, W^{-1})| |(v_1^*, v_2^*, v_3^*, (W^{-1})^*)|) = 0, \\
\frac{\partial \mathcal{H}}{\partial p} &= \frac{\Gamma}{\Gamma - 1} (1 + \theta v_1) W^2 (1 - \mathbf{v} \cdot \mathbf{v}^*) - (1 + \theta v_1^*) \\
&\geq 2 (1 + \theta v_1) W^2 (1 - \mathbf{v} \cdot \mathbf{v}^*) - (1 + \theta v_1^*) \\
&\geq \min\{H_p^+, H_p^-\} > 0,
\end{aligned}$$

where

$$\begin{aligned}
H_p^{\pm} &:= 2(1 \pm v_1) W^2 (1 - \mathbf{v} \cdot \mathbf{v}^*) - (1 \pm v_1^*) \\
&= 2(1 \pm v_1) W^2 - 1 - 2(1 \pm v_1) W^2 \left[v_1^* \left(v_1 \pm \frac{1}{2(1 \pm v_1) W^2} \right) + v_2^* v_2 + v_3^* v_3 \right] \\
&\geq 2(1 \pm v_1) W^2 - 1 - 2(1 \pm v_1) W^2 |\mathbf{v}^*| \sqrt{\left(v_1 \pm \frac{1}{2(1 \pm v_1) W^2} \right)^2 + v_2^2 + v_3^2} \\
&= (2(1 \pm v_1) W^2 - 1)(1 - |\mathbf{v}^*|) \geq (2(1 \pm v_1)/(1 - v_1^2) - 1)(1 - |\mathbf{v}^*|) \\
&= 2(1 \pm v_1)^2 (1 - |\mathbf{v}^*|)/(1 - v_1^2) > 0.
\end{aligned}$$

Here we have used that $|\mathbf{v}| < 1$ because of $\mathbf{U} \in \mathcal{G}$, and the Cauchy–Schwarz inequality. Thus, together with $\rho > 0$, $p > 0$, one has

$$\mathcal{H}(\rho, p, \mathbf{v}, \mathbf{B}, \mathbf{v}^*, \mathbf{B}^*, \theta) > \mathcal{H}(0, 0, \mathbf{v}, \mathbf{B}, \mathbf{v}^*, \mathbf{B}^*, \theta) =: \mathcal{H}_0(\mathbf{v}, \mathbf{B}, \mathbf{v}^*, \mathbf{B}^*, \theta).$$

The subsequent task is to show that $\mathcal{H}_0 \geq 0$. This is equivalent to the positive semi-definiteness of a symmetric matrix $\mathcal{A}^H(\mathbf{v}, \mathbf{v}^*, \theta)$ for any $\theta \in [-1, 1]$ and $\mathbf{v}, \mathbf{v}^* \in \mathbb{R}^3$ with $|\mathbf{v}| < 1$ and $|\mathbf{v}^*| < 1$, because $\mathcal{H}_0(\mathbf{v}, \mathbf{B}, \mathbf{v}^*, \mathbf{B}^*, \theta)$ can be reformulated into a quadratic form of $(\mathbf{B}, \mathbf{B}^*)$, i.e.

$$\mathcal{H}_0(\mathbf{v}, \mathbf{B}, \mathbf{v}^*, \mathbf{B}^*, \theta) = \frac{1}{2} (\mathbf{B}, \mathbf{B}^*) \mathcal{A}^H(\mathbf{v}, \mathbf{v}^*, \theta) (\mathbf{B}, \mathbf{B}^*)^\top.$$

Here the diagonal and the upper triangular elements of the symmetric matrix $\mathcal{A}^H = [\mathcal{A}_{jk}^H(\mathbf{v}, \mathbf{v}^*, \theta)]_{6 \times 6}$ are

$$\begin{aligned}
\mathcal{A}_{11}^H &= 2(1 - v_2^* v_2 - v_3^* v_3) + (1 - \theta v_1^*)(v_2^2 + v_3^2 - 1), \\
\mathcal{A}_{12}^H &= v_1 v_2^* + v_2 v_1^* - v_1 v_2 + \theta (v_3^* v_2 v_3 - v_2^* v_3^2 + v_2^* - v_2 + v_1^* v_1 v_2),
\end{aligned}$$

$$\mathcal{A}^H_{13} = v_1v_3^* + v_3v_1^* - v_1v_3 + \theta(v_2^*v_2v_3 - v_3^*v_2^2 + v_3^* - v_3 + v_1^*v_1v_3),$$
$$\mathcal{A}^H_{14} = \theta(v_1^*v_2v_2^* + v_1^*v_3v_3^* - v_1^*) + {v_2^*}^2 + {v_3^*}^2 - 1,$$
$$\mathcal{A}^H_{15} = \theta(v_2^*v_3v_3^* - v_2(v_3v_3^* + v_1v_1^* - 1) - v_2^*) - v_1^*v_2^*,$$
$$\mathcal{A}^H_{16} = \theta(v_3^*v_2v_2^* - v_3(v_1v_1^* + v_2v_2^* - 1) - v_3^*) - v_1^*v_3^*,$$
$$\mathcal{A}^H_{22} = \theta(-v_1^*v_1^2 - 2v_1v_3v_3^* + 2v_1 + v_1^*v_3^2 - v_1^*) + v_1^2 - 2v_1v_1^* + v_3^2 - 2v_3v_3^* + 1,$$
$$\mathcal{A}^H_{23} = \theta(v_3^*v_1v_2 + v_2^*v_1v_3 - v_1^*v_2v_3) + v_2^*v_3^* - (v_2 - v_2^*)(v_3 - v_3^*),$$
$$\mathcal{A}^H_{24} = -(1+\theta v_1)v_1^*v_2^*, \quad \mathcal{A}^H_{25} = (1+\theta v_1)({v_1^*}^2 + {v_3^*}^2 - 1),$$
$$\mathcal{A}^H_{26} = -(1+\theta v_1)v_2^*v_3^*,$$
$$\mathcal{A}^H_{33} = \theta(-v_1^*v_1^2 - 2v_1v_2v_2^* + 2v_1 + v_1^*v_2^2 - v_1^*) + v_1^2 - 2v_1v_1^* + v_2^2 - 2v_2v_2^* + 1,$$
$$\mathcal{A}^H_{34} = -(1+\theta v_1)v_1^*v_3^*, \quad \mathcal{A}^H_{35} = -(1+\theta v_1)v_2^*v_3^*,$$
$$\mathcal{A}^H_{36} = (1+\theta v_1)({v_1^*}^2 + {v_2^*}^2 - 1),$$
$$\mathcal{A}^H_{44} = -(1+\theta v_1^*)({v_2^*}^2 + {v_3^*}^2 - 1), \quad \mathcal{A}^H_{45} = (1+\theta v_1^*)v_1^*v_2^*,$$
$$\mathcal{A}^H_{46} = (1+\theta v_1^*)v_1^*v_3^*, \quad \mathcal{A}^H_{55} = -(1+\theta v_1^*)({v_1^*}^2 + {v_3^*}^2 - 1),$$
$$\mathcal{A}^H_{56} = (1+\theta v_1^*)v_2^*v_3^*, \quad \mathcal{A}^H_{66} = -(1+\theta v_1^*)({v_1^*}^2 + {v_2^*}^2 - 1).$$

If taking the upper triangular matrix

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 1 & \dfrac{\theta(v_2^* - v_2)}{1+\theta v_1^*} & \dfrac{\theta(v_3^* - v_3)}{1+\theta v_1^*} \\ & 1 & 0 & 0 & \dfrac{1+\theta v_1}{1+\theta v_1^*} & 0 \\ & & 1 & 0 & 0 & \dfrac{1+\theta v_1}{1+\theta v_1^*} \\ & & & 1 & 0 & 0 \\ & & & & 1 & 0 \\ & & & & & 1 \end{pmatrix},$$

one has

$$\mathbf{P}\mathcal{A}^H(\mathbf{v}, \mathbf{v}^*, \theta)\mathbf{P}^\top = \operatorname{diag}\left\{ \frac{1}{1+\theta v_1^*}\boldsymbol{\mathcal{B}}^H(\mathbf{v}, \mathbf{v}^*, \theta), \boldsymbol{\mathcal{C}}^H(\mathbf{v}^*, \theta) \right\},$$

where $\boldsymbol{\mathcal{B}}^H$ and $\boldsymbol{\mathcal{C}}^H$ are two symmetric matrices respectively defined by

$$\boldsymbol{\mathcal{B}}^H = \begin{pmatrix} \mathcal{B}^H_{11} & \mathcal{B}^H_{12} & \mathcal{B}^H_{13} \\ \mathcal{B}^H_{21} & \mathcal{B}^H_{22} & \mathcal{B}^H_{23} \\ \mathcal{B}^H_{31} & \mathcal{B}^H_{32} & \mathcal{B}^H_{33} \end{pmatrix}, \quad \boldsymbol{\mathcal{C}}^H = (1+\theta v_1^*)\Big[(1 - |\mathbf{v}^*|^2)\mathbf{I} + {\mathbf{v}^*}^\top \mathbf{v}^*\Big],$$

with

$$\mathcal{B}_{11}^{H} = (1, v_2, v_3)\begin{pmatrix} (1-\theta^2)(v_2^{*2}+v_3^{*2}) & (\theta^2-1)v_2^* & (\theta^2-1)v_3^* \\ (\theta^2-1)v_2^* & 1-\theta^2+\theta^2 v_3^{*2} & -\theta^2 v_2^* v_3^* \\ (\theta^2-1)v_3^* & -\theta^2 v_2^* v_3^* & 1-\theta^2+\theta^2 v_2^{*2} \end{pmatrix} \times (1, v_2, v_3)^\top =: (1, v_2, v_3)\hat{\boldsymbol{\mathcal{B}}}_{11}(1, v_2, v_3)^\top,$$

$$\mathcal{B}_{22}^{H} = (1, v_1, v_3)\begin{pmatrix} (1-\theta^2)v_1^{*2}+v_3^{*2} & (\theta^2-1)v_1^*+\theta v_3^{*2} & -(1+\theta v_1^*)v_3^* \\ (\theta^2-1)v_1^*+\theta v_3^{*2} & 1-\theta^2+\theta^2 v_3^{*2} & -\theta(1+\theta v_1^*)v_3^* \\ -(1+\theta v_1^*)v_3^* & -\theta(1+\theta v_1^*)v_3^* & (1+\theta v_1^*)^2 \end{pmatrix} \times (1, v_1, v_3)^\top =: (1, v_1, v_3)\hat{\boldsymbol{\mathcal{B}}}_{22}(1, v_1, v_3)^\top,$$

$$\mathcal{B}_{33}^{H} = (1, v_1, v_2)\begin{pmatrix} (1-\theta^2)v_1^{*2}+v_2^{*2} & (\theta^2-1)v_1^*+\theta v_2^{*2} & -(1+\theta v_1^*)v_2^* \\ (\theta^2-1)v_1^*+\theta v_2^{*2} & 1-\theta^2+\theta^2 v_2^{*2} & -\theta(1+\theta v_1^*)v_2^* \\ -(1+\theta v_1^*)v_2^* & -\theta(1+\theta v_1^*)v_2^* & (1+\theta v_1^*)^2 \end{pmatrix} \times (1, v_1, v_2)^\top =: (1, v_1, v_2)\hat{\boldsymbol{\mathcal{B}}}_{33}(1, v_1, v_2)^\top,$$

$$\mathcal{B}_{12}^{H} = \mathcal{B}_{21}^{H} = (1, \mathbf{v})\begin{pmatrix} (\theta^2-1)v_1^*v_2^* & (1-\theta^2)v_2^* & (1-\theta^2)v_1^*-\theta v_3^{*2} & \theta v_2^* v_3^* \\ & 0 & \theta^2(1-v_3^{*2})-1 & \theta^2 v_2^* v_3^* \\ & & 0 & (1+\theta v_1^*)\theta v_3^* \\ & & & -(1+\theta v_1^*)\theta v_2^* \end{pmatrix} \times (1, \mathbf{v})^\top,$$

$$\mathcal{B}_{13}^{H} = \mathcal{B}_{31}^{H} = (1, \mathbf{v})\begin{pmatrix} (\theta^2-1)v_1^*v_3^* & (1-\theta^2)v_3^* & \theta v_2^* v_3^* & (1-\theta^2)v_1^*-\theta v_2^{*2} \\ & 0 & \theta^2 v_2^* v_3^* & \theta^2(1-v_2^{*2})-1 \\ & & -(1+\theta v_1^*)\theta v_3^* & (1+\theta v_1^*)\theta v_2^* \\ & & & 0 \end{pmatrix} \times (1, \mathbf{v})^\top,$$

$$\mathcal{B}_{23}^{H} = \mathcal{B}_{32}^{H} = -(v_2 - v_2^* - \theta v_1 v_2^* + \theta v_2 v_1^*)(v_3 - v_3^* - \theta v_1 v_3^* + \theta v_3 v_1^*).$$

Under the hypothesis, one has that $1+\theta v_1^* \geq 1-|v_1^*| > 0$ and thus the matrix $\boldsymbol{\mathcal{C}}^H(\mathbf{v}^*,\theta)$ is positive definite. Therefore, the subsequent task is to show the positive semi-definiteness of the symmetric matrix $\boldsymbol{\mathcal{B}}^H$, or equivalently, the non-negativity of all principal minors of $\boldsymbol{\mathcal{B}}^H$. It is observed that these minors can be estimated through the quadratic forms of $(1,\mathbf{v})^\top$. First check the first-order principal minors of $\boldsymbol{\mathcal{B}}^H$. If taking

$$\mathbf{P}_1 = \begin{pmatrix} 1 & v_2^* & v_3^* \\ & 1 & 0 \\ & & 1 \end{pmatrix},$$

one has

$$\mathbf{P}_1\hat{\boldsymbol{\mathcal{B}}}_{\mathbf{11}}\mathbf{P}_1^\top = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1-\theta^2+\theta^2 {v_3^*}^2 & -\theta^2 v_2^* v_3^* \\ 0 & -\theta^2 v_2^* v_3^* & 1-\theta^2+\theta^2 {v_2^*}^2 \end{pmatrix}.$$

Then using

$$1-\theta^2+\theta^2 {v_3^*}^2 \geq 0, \quad 1-\theta^2+\theta^2 {v_2^*}^2 \geq 0,$$
$$\det\begin{pmatrix} 1-\theta^2+\theta^2 {v_3^*}^2 & -\theta^2 v_2^* v_3^* \\ -\theta^2 v_2^* v_3^* & 1-\theta^2+\theta^2 {v_2^*}^2 \end{pmatrix} = (1-\theta^2)[1-\theta^2+\theta^2({v_2^*}^2+{v_3^*}^2)] \geq 0,$$

yields the positive semi-definiteness of the matrix $\hat{\boldsymbol{\mathcal{B}}}_{\mathbf{11}}$, which follows that $\mathcal{B}_{11}^H \geq 0$. Similarly, one has $\mathcal{B}_{22}^H \geq 0$ and $\mathcal{B}_{33}^H \geq 0$. Next we consider the second-order principal minors of $\boldsymbol{\mathcal{B}}^H$. Some algebraic manipulations yield

$$\det\begin{pmatrix} \mathcal{B}_{11}^H & \mathcal{B}_{12}^H \\ \mathcal{B}_{21}^H & \mathcal{B}_{22}^H \end{pmatrix} = (v_3-v_3^*)^2\Xi, \quad \det\begin{pmatrix} \mathcal{B}_{11}^H & \mathcal{B}_{13}^H \\ \mathcal{B}_{31}^H & \mathcal{B}_{33}^H \end{pmatrix} = (v_2-v_2^*)^2\Xi,$$
$$\det\begin{pmatrix} \mathcal{B}_{22}^H & \mathcal{B}_{23}^H \\ \mathcal{B}_{32}^H & \mathcal{B}_{33}^H \end{pmatrix} = (v_1-v_1^*)^2\Xi,$$

where

$$\Xi = (1-\theta^2)\mathbf{z}^\top \operatorname{diag}\left\{(1-\theta^2)\begin{pmatrix} {v_1^*}^2 & -v_1^* \\ -v_1^* & 1 \end{pmatrix}, (1+\theta v_1^*)^2\mathbf{I}_2\right\}\mathbf{z},$$

with

$$\mathbf{z} = \left(1, v_1, v_2 - \frac{(1+\theta v_1)v_2^*}{1+\theta v_1^*}, v_3 - \frac{(1+\theta v_1)v_3^*}{1+\theta v_1^*}\right)^\top.$$

It is not difficult to know that $\Xi \geq 0$ by noting the positive semi-definiteness of

$$\begin{pmatrix} {v_1^*}^2 & -v_1^* \\ -v_1^* & 1 \end{pmatrix}.$$

Therefore, all three second-order principal minors of $\boldsymbol{\mathcal{B}}^H$ are non-negative. Finally we consider the third-order principal minor of $\boldsymbol{\mathcal{B}}^H$, i.e. $\det(\boldsymbol{\mathcal{B}}^H)$. Some algebraic manipulations yield

$$\det\begin{pmatrix}\mathcal{B}_{21}^H & \mathcal{B}_{23}^H\\ \mathcal{B}_{31}^H & \mathcal{B}_{33}^H\end{pmatrix} = (v_1^* - v_1)(v_2 - v_2^*)\Xi,$$

$$\det\begin{pmatrix}\mathcal{B}_{21}^H & \mathcal{B}_{22}^H\\ \mathcal{B}_{31}^H & \mathcal{B}_{32}^H\end{pmatrix} = (v_1 - v_1^*)(v_3 - v_3^*)\Xi,$$

$$\det\begin{pmatrix}\mathcal{B}_{11}^H & \mathcal{B}_{13}^H\\ \mathcal{B}_{21}^H & \mathcal{B}_{23}^H\end{pmatrix} = (v_2^* - v_2)(v_3 - v_3^*)\Xi.$$

Based on those first- and second-order principal minors of the symmetric matrix $\boldsymbol{\mathcal{B}}^H$, one obtains the adjoint matrix of $\boldsymbol{\mathcal{B}}^H$:

$$\begin{aligned}\operatorname{adj}(\boldsymbol{\mathcal{B}}^H) &= \Xi\begin{pmatrix}(v_1 - v_1^*)^2 & (v_1 - v_1^*)(v_2 - v_2^*) & (v_1 - v_1^*)(v_3 - v_3^*)\\ (v_1 - v_1^*)(v_2 - v_2^*) & (v_2 - v_2^*)^2 & (v_2 - v_2^*)(v_3 - v_3^*)\\ (v_1 - v_1^*)(v_3 - v_3^*) & (v_2 - v_2^*)(v_3 - v_3^*) & (v_3 - v_3^*)^2\end{pmatrix}\\ &= \Xi(\mathbf{v} - \mathbf{v}^*)^\top(\mathbf{v} - \mathbf{v}^*),\end{aligned}$$

which is also a symmetric matrix of size 3 and has rank of at most 1, such that $\operatorname{adj}(\boldsymbol{\mathcal{B}}^H)$ and $\boldsymbol{\mathcal{B}}^H$ are irreversible and $\det(\boldsymbol{\mathcal{B}}^H) = 0$. In conclusion, the matrix $\boldsymbol{\mathcal{B}}^H$ is positive semi-definite, and the inequality (2.37) for the case of $i = 1$, i.e. (2.38), does hold.

(ii) The inequality (2.37) for the case of $i = 2$ or 3 can be verified by using (2.38) and the orthogonal invariance in Corollary 2.2.

For the case of $i = 2$, we introduce a symmetric matrix $\mathbf{T} = \operatorname{diag}\{1, \mathbf{T}_3, \mathbf{T}_3, 1\}$ with the orthogonal matrix

$$\mathbf{T}_3 = \begin{pmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}.$$

Regarding the conservative vector $\mathbf{U}$ as a vector function of the primitive variables $\mathbf{V}$, i.e. $\mathbf{U}(\mathbf{V})$, then one has $\mathbf{U}(\mathbf{TV}) = \mathbf{TU} =: \tilde{\mathbf{U}}$, $\mathbf{F}_1(\mathbf{U}(\mathbf{TV})) = \mathbf{TF}_2(\mathbf{U})$, and

$$\mathbf{F}_1(\tilde{\mathbf{U}}) = \mathbf{F}_1(\mathbf{TU}) = \mathbf{F}_1(\mathbf{U}(\mathbf{TV})) = \mathbf{TF}_2(\mathbf{U}).$$

Thanks to Corollary 2.2, one obtains $\tilde{\mathbf{U}} \in \mathcal{G}$. Let $\tilde{\mathbf{v}}^* = \mathbf{v}^*\mathbf{T}_3$ and $\tilde{\mathbf{B}}^* = \mathbf{B}^*\mathbf{T}_3$, then

$$|\tilde{\mathbf{v}}^*| = |\mathbf{v}| < 1, \quad |\tilde{\mathbf{B}}^*| = |\mathbf{B}^*|, \quad \tilde{\mathbf{v}}^* \cdot \tilde{\mathbf{B}}^* = \mathbf{v}^* \cdot \mathbf{B}^*.$$

It follows from (2.35) and (2.36) that $\tilde{p}_m^* = p_m^*$ and $\tilde{\mathbf{n}}^* = \mathbf{T}\mathbf{n}^*$. Using the inequality (2.38) with $\tilde{\mathbf{U}} \in \mathcal{G}$, $\tilde{\mathbf{v}}^*$, and $\tilde{\mathbf{B}}^*$ gives

$$\begin{aligned} 0 &< (\tilde{\mathbf{U}} + \theta \mathbf{F}_1(\tilde{\mathbf{U}})) \cdot \tilde{\mathbf{n}}^* + \tilde{p}_m^* + \theta(\tilde{v}_1^* \tilde{p}_m^* - \tilde{B}_1(\tilde{\mathbf{v}}^* \cdot \tilde{\mathbf{B}}^*)) \\ &= (\mathbf{T}\mathbf{U} + \theta \mathbf{T}\mathbf{F}_2(\mathbf{U})) \cdot (\mathbf{T}\mathbf{n}^*) + p_m^* + \theta(v_2^* p_m^* - B_2(\mathbf{v}^* \cdot \mathbf{B}^*)) \\ &= (\mathbf{U} + \theta \mathbf{F}_2(\mathbf{U})) \cdot \mathbf{n}^* + p_m^* + \theta(v_2^* p_m^* - B_2(\mathbf{v}^* \cdot \mathbf{B}^*)), \end{aligned}$$

where the orthogonality of the matrix $\mathbf{T}$ has been used in the last equality. This verifies the inequality (2.37) for the case of $i = 2$. The case of $i = 3$ can be similarly derived by taking the orthogonal matrix

$$\mathbf{T}_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

The proof is complete. □

Remark 2.6. Thanks to the second equivalent form of $\mathcal{G}$, Lemma 2.8 tells us that the inequality

$$(\mathbf{U} + \theta \mathbf{F}_i(\mathbf{U})) \cdot \mathbf{n}^* + p_m^* > 0, \tag{2.39}$$

does not always hold for any $\theta \in [-1, 1]$ and $\mathbf{B}^*, \mathbf{v}^* \in \mathbb{R}^3$ with $|\mathbf{v}^*| < 1$, where $i \in \{1, 2, 3\}$. Compare (2.39) to the inequality (2.37), the third term at the left-hand side of (2.37) is extremely technical and crucial in deriving the generalized LxF splitting properties. Although this term is not always positive or negative, it can be canceled out dexterously with the help of the following "discrete divergence-free" condition (2.40) or (2.42), see the proofs of following theorems.

Based on the above lemma, we derive the following generalized LxF splitting properties.

Theorem 2.4. (1D generalized LxF splitting) *If* $\tilde{\mathbf{U}} = (\tilde{D}, \tilde{\mathbf{m}}, \tilde{\mathbf{B}}, \tilde{E})^\top \in \mathcal{G}$ *and* $\hat{\mathbf{U}} = (\hat{D}, \hat{\mathbf{m}}, \hat{\mathbf{B}}, \hat{E})^\top \in \mathcal{G}$ *satisfy* 1D "*discrete divergence-free*" *condition*

$$\tilde{B}_i - \hat{B}_i = 0, \tag{2.40}$$

for a given $i \in \{1, 2, 3\}$, *then for any* $\alpha \geq c = 1$ *it holds*

$$\bar{\mathbf{U}} := \frac{1}{2}(\tilde{\mathbf{U}} - \alpha^{-1}\mathbf{F}_i(\tilde{\mathbf{U}}) + \hat{\mathbf{U}} + \alpha^{-1}\mathbf{F}_i(\hat{\mathbf{U}})) \in \mathcal{G}. \tag{2.41}$$

Proof. It is obvious that

$$\frac{1}{2}(\tilde{D}(1 - \tilde{v}_i/\alpha) + \hat{D}(1 + \hat{v}_i/\alpha)) > 0,$$

that is to say, the first component of $\bar{\mathbf{U}}$ satisfies the first constraint in $\mathcal{G}_1$, see Theorem 2.3.

Next, let us check the second constraint in $\mathcal{G}_1$. For any $\mathbf{B}^*, \mathbf{v}^* \in \mathbb{R}^3$ with $v^* < 1$, using Lemma 2.9 gives

$$\begin{aligned}\bar{\mathbf{U}} \cdot \mathbf{n}^* + p_m^* &= \frac{1}{2}((\tilde{\mathbf{U}} - \alpha^{-1}\mathbf{F}_i(\tilde{\mathbf{U}})) \cdot \mathbf{n}^* + p_m^*) + \frac{1}{2}((\hat{\mathbf{U}} + \alpha^{-1}\mathbf{F}_i(\hat{\mathbf{U}})) \cdot \mathbf{n}^* + p_m^*) \\ &\overset{(2.37)}{>} -\frac{1}{2}\alpha^{-1}(\tilde{B}_i(\mathbf{v}^* \cdot \mathbf{B}^*) - v_i^* p_m^*) + \frac{1}{2}\alpha^{-1}(\hat{B}_i(\mathbf{v}^* \cdot \mathbf{B}^*) - v_i^* p_m^*) \\ &\overset{(2.40)}{=} 0,\end{aligned}$$

where $\mathbf{n}^*$ and p_m^* are defined in (2.35) and (2.36), respectively. Using Theorem 2.3 completes the proof. □

Theorem 2.5. (2D generalized LxF splitting) *If* $\tilde{\mathbf{U}}^i$, $\hat{\mathbf{U}}^i$, $\bar{\mathbf{U}}^i$, $\breve{\mathbf{U}}^i \in \mathcal{G}$ *for* $i = 1, 2, \ldots, L$ *satisfy* 2D "*discrete divergence-free*" *condition*

$$\frac{\sum_{i=1}^L \omega_i(\tilde{B}_1^i - \hat{B}_1^i)}{\Delta x} + \frac{\sum_{i=1}^L \omega_i(\bar{B}_2^i - \breve{B}_2^i)}{\Delta y} = 0, \tag{2.42}$$

where $\Delta x, \Delta y > 0$, *and the sum of all positive numbers* $\{\omega_i\}_{i=1}^L$ *is equal to* 1, *then for all* $\alpha \geq c = 1$ *it holds*

$$\begin{aligned}\bar{\mathbf{U}} := \frac{1}{2(\frac{1}{\Delta x} + \frac{1}{\Delta y})} \sum_{i=1}^L \omega_i \Big[&\frac{1}{\Delta x}(\tilde{\mathbf{U}}^i - \alpha^{-1}\mathbf{F}_1(\tilde{\mathbf{U}}^i) + \hat{\mathbf{U}}^i + \alpha^{-1}\mathbf{F}_1(\hat{\mathbf{U}}^i)) \\ &+ \frac{1}{\Delta y}(\bar{\mathbf{U}}^i - \alpha^{-1}\mathbf{F}_2(\bar{\mathbf{U}}^i) + \breve{\mathbf{U}}^i + \alpha^{-1}\mathbf{F}_2(\breve{\mathbf{U}}^i))\Big] \in \mathcal{G}.\end{aligned} \tag{2.43}$$

Proof. The first component of $\bar{\mathbf{U}}$ satisfies the first constraint in $\mathcal{G}_1$, i.e.

$$\begin{aligned}\frac{1}{2(\frac{1}{\Delta x} + \frac{1}{\Delta y})} \sum_{i=1}^L \omega_i \Big[&\frac{1}{\Delta x}(\tilde{D}^i(1 - \alpha^{-1}\tilde{v}_1^i) + \hat{D}^i(1 + \alpha^{-1}\hat{v}_1^i)) \\ &+ \frac{1}{\Delta y}\Big(\bar{D}^i(1 - \alpha^{-1}\bar{v}_2^i) + \breve{D}^i(1 + \alpha^{-1}\breve{v}_2^i)\Big)\Big] > 0.\end{aligned}$$

For any $\mathbf{B}^*, \mathbf{v}^* \in \mathbb{R}^3$ with $|\mathbf{v}^*| < 1$, utilizing Lemma 2.9 and (2.42) gives

$$\begin{aligned}\bar{\mathbf{U}} \cdot \mathbf{n}^* + p_m^* = \frac{1}{2(\frac{1}{\Delta x} + \frac{1}{\Delta y})} \sum_{i=1}^L \omega_i \Big[&\frac{1}{\Delta x}((\tilde{\mathbf{U}}^i - \alpha^{-1}\mathbf{F}_1(\tilde{\mathbf{U}}^i)) \cdot \mathbf{n}^* + p_m^* \\ &+ (\hat{\mathbf{U}}^i + \alpha^{-1}\mathbf{F}_1(\hat{\mathbf{U}}^i)) \cdot \mathbf{n}^* + p_m^*) + \frac{1}{\Delta y}((\bar{\mathbf{U}}^i - \alpha^{-1}\mathbf{F}_2(\bar{\mathbf{U}}^i)) \cdot \mathbf{n}^* \\ &+ p_m^* + (\breve{\mathbf{U}}^i + \alpha^{-1}\mathbf{F}_2(\breve{\mathbf{U}}^i)) \cdot \mathbf{n}^* + p_m^*)\Big]\end{aligned}$$

$$
\begin{aligned}
&\overset{(2.37)}{>} \frac{1}{2(\frac{1}{\Delta x}+\frac{1}{\Delta y})}\sum_{i=1}^{L}\omega_i\left[\frac{1}{\Delta x}(-\alpha^{-1}(\tilde{B}_1^i(\mathbf{v}^*\cdot\mathbf{B}^*)-v_1^*p_m^*)\right.\\
&\quad+\alpha^{-1}(\hat{B}_1^i(\mathbf{v}^*\cdot\mathbf{B}^*)-v_1^*p_m^*))+\frac{1}{\Delta y}(-\alpha^{-1}(\bar{B}_2^i(\mathbf{v}^*\cdot\mathbf{B}^*)-v_2^*p_m^*)\\
&\quad\left.+\alpha^{-1}(\breve{B}_2^i(\mathbf{v}^*\cdot\mathbf{B}^*)-v_2^*p_m^*))\right]\\
&= -\frac{\mathbf{v}^*\cdot\mathbf{B}^*}{2\alpha(\frac{1}{\Delta x}+\frac{1}{\Delta y})}\sum_{i=1}^{L}\omega_i\left(\frac{\tilde{B}_1^i-\hat{B}_1^i}{\Delta x}+\frac{\bar{B}_2^i-\breve{B}_2^i}{\Delta y}\right)\\
&\overset{(2.42)}{=} 0.
\end{aligned}
$$

Thus $\bar{\mathbf{U}}$ also satisfies the second constraint in $\mathcal{G}_1$. Using Theorem 2.3 completes the proof. □

Theorem 2.6. (3D generalized LxF splitting) *If* $\tilde{\mathbf{U}}^i$, $\hat{\mathbf{U}}^i$, $\bar{\mathbf{U}}^i$, $\breve{\mathbf{U}}^i$, $\bar{\bar{\mathbf{U}}}^i$, $\breve{\breve{\mathbf{U}}}^i \in \mathcal{G}$ *for* $i=1,2,\ldots,L$, *and they satisfy the* 3D "*discrete divergence-free*" *condition*

$$
\frac{\sum_{i=1}^{L}\omega_i(\tilde{B}_1^i-\hat{B}_1^i)}{\Delta x}+\frac{\sum_{i=1}^{L}\omega_i(\bar{B}_2^i-\breve{B}_2^i)}{\Delta y}+\frac{\sum_{i=1}^{L}\omega_i\left(\bar{\bar{B}}_3^i-\breve{\breve{B}}_3^i\right)}{\Delta z}=0,
$$

where $\Delta x,\Delta y,\Delta z>0$, *and the sum of all positive numbers* $\{\omega_i\}_{i=1}^{L}$ *is equal to* 1, *then for any* $\alpha\geq c=1$ *it holds*

$$
\begin{aligned}
&\frac{1}{2(\frac{1}{\Delta x}+\frac{1}{\Delta y}+\frac{1}{\Delta z})}\sum_{i=1}^{L}\omega_i\left[\frac{1}{\Delta x}(\tilde{\mathbf{U}}^i-\alpha^{-1}\mathbf{F}_1(\tilde{\mathbf{U}}^i)+\hat{\mathbf{U}}^i+\alpha^{-1}\mathbf{F}_1(\hat{\mathbf{U}}^i))\right.\\
&\quad+\frac{1}{\Delta y}(\bar{\mathbf{U}}^i-\alpha^{-1}\mathbf{F}_2(\bar{\mathbf{U}}^i)+\breve{\mathbf{U}}^i+\alpha^{-1}\mathbf{F}_2(\breve{\mathbf{U}}^i))\\
&\quad\left.+\frac{1}{\Delta z}\left(\bar{\bar{\mathbf{U}}}^i-\alpha^{-1}\mathbf{F}_3(\bar{\bar{\mathbf{U}}}^i)+\breve{\breve{\mathbf{U}}}^i+\alpha^{-1}\mathbf{F}_3\left(\breve{\breve{\mathbf{U}}}^i\right)\right)\right]\in\mathcal{G}.
\end{aligned}
$$

Proof. The proof is similar to that of Theorem 2.5 and omitted here. □

Remark 2.7. Because the convex combination $\bar{\mathbf{U}}$ in the above generalized LxF splitting properties depends on several states, it is very difficult to directly check whether $\bar{\mathbf{U}}$ belongs to the set $\mathcal{G}$. It is subtly and fortunately overcame by using the inequality (2.37) in Lemma 2.9 and the "discrete divergence-free" condition, which is an approximation to (1.2). For example, the "discrete divergence-free" condition (2.42) can be derived by using some quadrature rule for the integrals at

the left-hand side of

$$\frac{1}{\Delta x}\left(\frac{1}{\Delta y}\int_{y_0}^{y_0+\Delta y}(B_1(x_0+\Delta x,y)-B_1(x_0,y))dy\right)$$
$$+\frac{1}{\Delta y}\left(\frac{1}{\Delta x}\int_{x_0}^{x_0+\Delta x}(B_2(x,y_0+\Delta y)-B_2(x,y_0))dx\right)$$
$$=\frac{1}{\Delta x\Delta y}\int_I\left(\frac{\partial B_1}{\partial x}+\frac{\partial B_2}{\partial y}\right)dxdy=0, \tag{2.44}$$

where $(x,y)=(x_1,x_2)$ and $I:=[x_0,x_0+\Delta x]\times[y_0,y_0+\Delta y]$.

The above generalized LxF splitting properties are important tools in developing and analyzing the PCP numerical schemes on uniform meshes if the numerical flux is taken as the LxF-type flux (3.2). Moreover, it is easy to extend them on non-uniform or unstructured meshes. For example, the following theorem shows an extension to the case of 2D arbitrarily convex polygon mesh.

Theorem 2.7. *If for* $i=1,2,\ldots,L$ *and* $j=1,2,\ldots,J$, $\mathbf{U}^{ij}\in\mathcal{G}$ *and satisfy* 2D "*discrete divergence-free*" *condition over an* J*-sided convex polygon*

$$\sum_{j=1}^{J}\left[\sum_{i=1}^{L}\omega_i(B_1^{ij}\mathcal{N}_1^j+B_2^{ij}\mathcal{N}_2^j)\right]\ell_j=0, \tag{2.45}$$

where $\ell_j>0$ *and* $(\mathcal{N}_1^j,\mathcal{N}_2^j)$ *are the length and the unit outward normal vector of the* j*th edge of the polygon, respectively, and the sum of all positive numbers* $\{\omega_i\}_{i=1}^L$ *is equal to* 1, *then for all* $\alpha\geq c=1$ *it holds*

$$\bar{\mathbf{U}}:=\frac{1}{\sum_{j=1}^J\ell_j}\sum_{j=1}^{J}\left[\sum_{i=1}^{L}\omega_i(\mathbf{U}^{ij}-\alpha^{-1}(\mathbf{F}_1(\mathbf{U}^{ij})\mathcal{N}_1^j+\mathbf{F}_2(\mathbf{U}^{ij})\mathcal{N}_2^j))\right]\ell_j\in\mathcal{G}.$$

Proof. The rotational invariance property of the 2D RMHD equations (1.1) yields

$$\mathbf{F}_1(\mathbf{U}^{ij})\mathcal{N}_1^j+\mathbf{F}_2(\mathbf{U}^{ij})\mathcal{N}_2^j=\mathbf{T}_j^{-1}\mathbf{F}_1(\mathbf{T}_j\mathbf{U}^{ij}),$$

where $\mathbf{T}_j:=\mathrm{diag}\{1,\mathbf{T}_{3,j},\mathbf{T}_{3,j},1\}$ with the rotational matrix $\mathbf{T}_{3,j}$ defined by

$$\mathbf{T}_{3,j}:=\begin{pmatrix}\mathcal{N}_1^j & \mathcal{N}_2^j & 0\\ -\mathcal{N}_2^j & \mathcal{N}_1^j & 0\\ 0 & 0 & 1\end{pmatrix}.$$

For each j and any $\mathbf{B}^*,\mathbf{v}^*\in\mathbb{R}^3$ with $|\mathbf{v}^*|<1$, let $\hat{\mathbf{v}}^*=\mathbf{v}^*\mathbf{T}_{3,j}$ and $\hat{\mathbf{B}}^*=\mathbf{B}^*\mathbf{T}_{3,j}$, one has $|\hat{\mathbf{v}}^*|=|\mathbf{v}|<1$, $\hat{\mathbf{v}}^*\cdot\hat{\mathbf{B}}^*=\mathbf{v}^*\cdot\mathbf{B}^*$, $\hat{p}_m^*=p_m^*$, and $\hat{\mathbf{n}}^*=\mathbf{T}_j\mathbf{n}^*$. Utilizing

Lemma 2.9 for $\mathbf{T}_j\mathbf{U}^{ij}$, $\hat{\mathbf{v}}^*$, and $\hat{\mathbf{B}}^*$ gives

$$\begin{aligned}0 &< (\mathbf{T}_j\mathbf{U}^{ij} - \alpha^{-1}\mathbf{F}_1(\mathbf{T}_j\mathbf{U}^{ij}))\cdot\hat{\mathbf{n}}^* + \hat{p}_m^* \\ &\quad - \alpha^{-1}(\hat{v}_1^*\hat{p}_m^* - (B_1^{ij}\mathcal{N}_1^j + B_2^{ij}\mathcal{N}_2^j)(\hat{\mathbf{v}}^*\cdot\hat{\mathbf{B}}^*)) \\ &= (\mathbf{U}^{ij} - \alpha^{-1}\mathbf{T}_j^{-1}\mathbf{F}_1(\mathbf{T}_j\mathbf{U}^{ij}))\cdot\mathbf{n}^* + p_m^* \\ &\quad - \alpha^{-1}((v_1^*\mathcal{N}_1^j + v_2^*\mathcal{N}_2^j)p_m^* - (B_1^{ij}\mathcal{N}_1^j + B_2^{ij}\mathcal{N}_2^j)(\mathbf{v}^*\cdot\mathbf{B}^*)),\end{aligned} \tag{2.46}$$

where the orthogonality of $\mathbf{T}_j$ has been used. Hence, one has

$$\begin{aligned}\bar{\mathbf{U}}\cdot\mathbf{n}^* + p_m^* &= \frac{1}{\sum_{j=1}^J \ell_j}\sum_{j=1}^J\left[\sum_{i=1}^L \omega_i\Big((\mathbf{U}^{ij} - \alpha^{-1}\mathbf{T}_j^{-1}\mathbf{F}_1(\mathbf{T}_j\mathbf{U}^{ij})\Big)\cdot\mathbf{n}^* + p_m^*)\right]\ell_j \\ &\overset{(2.46)}{>} \frac{1}{\alpha\sum_{j=1}^J \ell_j}\sum_{j=1}^J\left[\sum_{i=1}^L \omega_i((v_1^*\mathcal{N}_1^j + v_2^*\mathcal{N}_2^j)p_m^*\right. \\ &\qquad \left. - (B_1^{ij}\mathcal{N}_1^j + B_2^{ij}\mathcal{N}_2^j)(\mathbf{v}^*\cdot\mathbf{B}^*))\right]\ell_j \\ &\overset{(2.45)}{=} \frac{p_m^*}{\alpha\sum_{j=1}^J \ell_j}\sum_{j=1}^J\left[\sum_{i=1}^L \omega_i(v_1^*\mathcal{N}_1^j + v_2^*\mathcal{N}_2^j)\right]\ell_j \\ &= \frac{p_m^*}{\alpha\sum_{j=1}^J \ell_j}\sum_{j=1}^J(v_1^*\mathcal{N}_1^j + v_2^*\mathcal{N}_2^j)\ell_j = 0,\end{aligned}$$

which implies that $\bar{\mathbf{U}}$ satisfies the second constraint in $\mathcal{G}_1$. On the other hand, $\bar{\mathbf{U}}$ satisfies the first constraint in $\mathcal{G}_1$ because

$$\begin{aligned}&\frac{1}{\sum_{j=1}^J \ell_j}\sum_{j=1}^J\left[\sum_{i=1}^L \omega_i D^{ij}(1 - \alpha^{-1}(v_1^{ij}\mathcal{N}_1^j + v_2^{ij}\mathcal{N}_2^j))\right]\ell_j \\ &\quad \geq \frac{1}{\sum_{j=1}^J \ell_j}\sum_{j=1}^J\left[\sum_{i=1}^L \omega_i D^{ij}(1 - \alpha^{-1}\sqrt{(v_1^{ij})^2 + (v_2^{ij})^2})\right]\ell_j > 0.\end{aligned}$$

Thus, the proof is completed by using Theorem 2.3. □

3. Physical-Constraints-Preserving Schemes

This section applies the previous theoretical results on the admissible state set $\mathcal{G}$ to develop PCP numerical schemes for the 1D and 2D special RMHD equations (1.1).

3.1. 1D *PCP schemes*

For the sake of convenience, this subsection will use the symbol x to replace the independent variable x_1 in (1.1). Assume that the spatial domain is divided into a uniform mesh with a constant spatial step-size Δx and the jth cell $I_j = (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$,

and the time interval is also divided into the (non-uniform) grid $\{t_0 = 0, t_{n+1} = t_n + \Delta t_n, n \geq 0\}$ with the time step-size Δt_n determined by the CFL-type condition. Let $\bar{\mathbf{U}}_j^n$ be the numerical approximation to the cell average value of the exact solution $\mathbf{U}(x,t)$ over the cell I_j at $t = t_n$. Our aim is to seek numerical schemes of the 1D RMHD equations (1.1), whose solution $\bar{\mathbf{U}}_j^n$ belongs to the set $\mathcal{G}$ if $\bar{\mathbf{U}}_j^0 \in \mathcal{G}$.

3.1.1. *First-order accurate schemes*

Consider the first-order accurate LxF-type scheme

$$\bar{\mathbf{U}}_j^{n+1} = \bar{\mathbf{U}}_j^n - \frac{\Delta t_n}{\Delta x}(\hat{\mathbf{F}}_1(\bar{\mathbf{U}}_j^n, \bar{\mathbf{U}}_{j+1}^n) - \hat{\mathbf{F}}_1(\bar{\mathbf{U}}_{j-1}^n, \bar{\mathbf{U}}_j^n), \tag{3.1}$$

where the numerical flux $\hat{\mathbf{F}}_1$ is defined by

$$\hat{\mathbf{F}}_i(\mathbf{U}^-, \mathbf{U}^+) = \frac{1}{2}(\mathbf{F}_i(\mathbf{U}^-) + \mathbf{F}_i(\mathbf{U}^+) - \varrho_i(\mathbf{U}^+ - \mathbf{U}^-)), \quad i = 1, 2, 3. \tag{3.2}$$

Here ϱ_i is an appropriate upper bound of the spectral radius of the Jacobian matrix $\partial \mathbf{F}_i(\mathbf{U})/\partial \mathbf{U}$ and may be taken as $\varrho_i = c = 1$.

Thanks to the generalized LxF splitting property shown in Theorem 2.4, one can prove that the scheme (3.1) is PCP under a CFL-type condition.

Theorem 3.1. *If $\bar{\mathbf{U}}_j^0 \in \mathcal{G}$ and $\bar{B}_{1,j}^0 = B_{\text{const}}$ for all j, then $\bar{\mathbf{U}}_j^n$, calculated by using* (3.1) *under the CFL-type condition*

$$0 < \Delta t_n \leq \Delta x / c \tag{3.3}$$

belongs to $\mathcal{G}$ and satisfies $\bar{B}_{1,j}^n = B_{\text{const}}$ for all j and $n \in \mathbb{N}$, where $c = 1$ is the speed of light.

Proof. Here the induction argument is used for the time-level number n. It is obvious that the conclusion holds for $n = 0$ because of the hypothesis on the initial data. Now assume that $\bar{\mathbf{U}}_j^n \in \mathcal{G}$ with $\bar{B}_{1,j}^n = B_{\text{const}}$ for all j, and check whether the conclusion holds for $n+1$. Thanks to the numerical flux in (3.2), the fifth equation in (3.1) gives

$$\begin{aligned} \bar{B}_{1,j}^{n+1} &= \bar{B}_{1,j}^n - \frac{\Delta t_n}{2\Delta x}(2\bar{B}_{1,j}^n - \bar{B}_{1,j+1}^n - \bar{B}_{1,j-1}^n) \\ &= B_{\text{const}} - \frac{\Delta t_n}{2\Delta x}(2B_{\text{const}} - B_{\text{const}} - B_{\text{const}}) = B_{\text{const}}, \end{aligned}$$

for all j. If substituting (3.2) into (3.1), one can rewrite (3.1) in the following form

$$\begin{aligned} \bar{\mathbf{U}}_j^{n+1} &= (1-\lambda)\bar{\mathbf{U}}_j^n + \frac{\lambda}{2}(\bar{\mathbf{U}}_{j+1}^n - \mathbf{F}_1(\bar{\mathbf{U}}_{j+1}^n) + \bar{\mathbf{U}}_{j-1}^n + \mathbf{F}_1(\bar{\mathbf{U}}_{j-1}^n)) \\ &=: (1-\lambda)\bar{\mathbf{U}}_j^n + \lambda \Xi, \end{aligned}$$

where $\lambda := \Delta t_n/\Delta x \in (0, 1]$ due to (3.3). With the induction hypothesis and Theorem 2.4, one has $\Xi \in \mathcal{G}$. The convexity of $\mathcal{G}$ further yields $\bar{\mathbf{U}}_j^{n+1} \in \mathcal{G}$. The proof is complete. □

3.1.2. *High-order accurate schemes*

This subsection discusses the high-order accurate PCP schemes for the 1D RMHD equations (1.1).

Let us consider the high-order (spatially) accurate scheme for the 1D RMHD equations (1.1):

$$\bar{\mathbf{U}}_j^{n+1} = \bar{\mathbf{U}}_j^n - \frac{\Delta t_n}{\Delta x}\Big(\hat{\mathbf{F}}_1\Big(\mathbf{U}^-_{j+\frac{1}{2}}, \mathbf{U}^+_{j+\frac{1}{2}}\Big) - \hat{\mathbf{F}}_1\Big(\mathbf{U}^-_{j-\frac{1}{2}}, \mathbf{U}^+_{j-\frac{1}{2}}\Big)\Big), \tag{3.4}$$

where the numerical flux $\hat{\mathbf{F}}_1$ is defined by (3.2). Equation (3.4) may be derived from high-order accurate finite volume schemes or the discrete evolution equation for the cell average value $\{\bar{\mathbf{U}}_j^n\}$ in the DG methods. The quantities $\mathbf{U}^-_{j+\frac{1}{2}}$ and $\mathbf{U}^+_{j+\frac{1}{2}}$ are the $(K+1)$th-order accurate approximations of the point values $\mathbf{U}\big(x_{j+\frac{1}{2}}, t_n\big)$ within the cells I_j and I_{j+1} respectively, and given by

$$\mathbf{U}^-_{j+\frac{1}{2}} = \mathbf{U}_j^n(x_{j+\frac{1}{2}} - 0), \quad \mathbf{U}^+_{j+\frac{1}{2}} = \mathbf{U}_{j+1}^n(x_{j+\frac{1}{2}} + 0),$$

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

Generally, the solution $\bar{\mathbf{U}}_j^{n+1}$ of the high-order accurate scheme (3.4) may not belong to $\mathcal{G}$ even if $\bar{\mathbf{U}}_j^n \in \mathcal{G}$ for all j. Thus if the scheme (3.4) is used to solve some ultra-relativistic problems with low density or pressure, or very large velocity, it may break down after some time steps due to the nonphysical numerical solutions generated by (3.4). To cure such defect, the positivity-preserving limiters devised in Refs. 42, 45 and 46 will be extended to our RMHD case and $\mathbf{U}_j(x)$ is limited as $\tilde{\mathbf{U}}_j(x)$ such that the values of $\tilde{\mathbf{U}}_j(x)$ at some critical points in I_j belong to $\mathcal{G}$. Let $\{\hat{x}_j^\alpha\}_{\alpha=1}^L$ be the Gauss–Lobatto nodes transformed into the interval I_j, and $\{\hat{\omega}_\alpha\}_{\alpha=1}^L$ be associated Gaussian quadrature weights satisfying $\sum_{\alpha=1}^L \hat{\omega}_\alpha = 1$. Here we take $2L - 3 \geq K$ in order that the algebraic precision of corresponding quadrature is at least K. In particular, one can take L as the smallest integer not less than $\frac{K+3}{2}$.

Theorem 3.2. *If the polynomial vector* $\mathbf{U}_j^n(x) =: (D_j(x), \mathbf{m}_j(x), \mathbf{B}_j(x), E_j(x))^\top$ *satisfy*:

(i) $B_{1,j}(x) = B_{\text{const}}$ *for any* $x \in I_j$ *and all* j, *and*
(ii) $\mathbf{U}_j^n(\hat{x}_j^\alpha) \in \mathcal{G}$ *for all* j *and* $\alpha = 1, 2, \ldots, L$,

then under the CFL-type condition

$$0 < \Delta t_n \leq \hat{\omega}_1 \Delta x, \tag{3.5}$$

it holds that $\bar{\mathbf{U}}_j^{n+1} \in \mathcal{G}$ *for the numerical scheme* (3.4).

Proof. The exactness of the Gauss–Lobatto quadrature rule with L nodes for the polynomials of degree K yields

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

Because $\hat{\omega}_1 = \hat{\omega}_L$, one has

$$\begin{aligned}
\bar{\mathbf{U}}_j^{n+1} &= \sum_{\alpha=1}^{L} \hat{\omega}_\alpha \mathbf{U}_j^n(\hat{x}_j^\alpha) - \lambda\Big(\hat{\mathbf{F}}_1\Big(\mathbf{U}^-_{j+\frac{1}{2}}, \mathbf{U}^+_{j+\frac{1}{2}}\Big) - \hat{\mathbf{F}}_1\Big(\mathbf{U}^-_{j-\frac{1}{2}}, \mathbf{U}^+_{j-\frac{1}{2}}\Big)\Big) \\
&= \sum_{\alpha=2}^{L-1} \hat{\omega}_\alpha \mathbf{U}_j^n(\hat{x}_j^\alpha) + \frac{\lambda}{2}\Big(\mathbf{U}^+_{j+\frac{1}{2}} - \mathbf{F}_1\Big(\mathbf{U}^+_{j+\frac{1}{2}}\Big) + \mathbf{U}^-_{j-\frac{1}{2}} + \mathbf{F}_1\Big(\mathbf{U}^-_{j-\frac{1}{2}}\Big)\Big) \\
&\quad + \hat{\omega}_1 \mathbf{U}^+_{j-\frac{1}{2}} + \hat{\omega}_L \mathbf{U}^-_{j+\frac{1}{2}} - \frac{\lambda}{2}\left(\mathbf{U}^-_{j+\frac{1}{2}} + \mathbf{F}_1\left(\mathbf{U}^-_{j+\frac{1}{2}}\right) + \mathbf{U}^+_{j-\frac{1}{2}} - \mathbf{F}_1\left(\mathbf{U}^+_{j-\frac{1}{2}}\right)\right) \\
&= \sum_{\alpha=2}^{L-1} \hat{\omega}_\alpha \mathbf{U}_j^n\left(\hat{x}_j^\alpha\right) + \frac{\lambda}{2}\left(\mathbf{U}^+_{j+\frac{1}{2}} - \mathbf{F}_1\left(\mathbf{U}^+_{j+\frac{1}{2}}\right) + \mathbf{U}^-_{j-\frac{1}{2}} + \mathbf{F}_1\left(\mathbf{U}^-_{j-\frac{1}{2}}\right)\right) \\
&\quad + \left(\hat{\omega}_1 - \frac{\lambda}{2}\right)\left[\mathbf{U}^+_{j-\frac{1}{2}} - \left(\frac{2\hat{\omega}_1}{\lambda} - 1\right)^{-1} \mathbf{F}_1\left(\mathbf{U}^+_{j-\frac{1}{2}}\right)\right. \\
&\quad \left. + \mathbf{U}^-_{j+\frac{1}{2}} + \left(\frac{2\hat{\omega}_1}{\lambda} - 1\right)^{-1} \mathbf{F}_1\left(\mathbf{U}^-_{j+\frac{1}{2}}\right)\right] \\
&=: \sum_{\alpha=2}^{L-1} \hat{\omega}_\alpha \mathbf{U}_j^n\left(\hat{x}_j^\alpha\right) + \lambda\Xi_1 + \left(\hat{\omega}_1 + \hat{\omega}_L - \lambda\right)\Xi_2, \qquad (3.6)
\end{aligned}$$

where $\lambda = \Delta t_n/\Delta x \in (0, \hat{\omega}_1]$, and

$$\Xi_1 := \frac{1}{2}\left(\mathbf{U}^+_{j+\frac{1}{2}} - \mathbf{F}_1\left(\mathbf{U}^+_{j+\frac{1}{2}}\right) + \mathbf{U}^-_{j-\frac{1}{2}} + \mathbf{F}_1\left(\mathbf{U}^-_{j-\frac{1}{2}}\right)\right),$$

$$\Xi_2 := \frac{1}{2}\left[\mathbf{U}^+_{j-\frac{1}{2}} - \left(\frac{2\hat{\omega}_1}{\lambda} - 1\right)^{-1} \mathbf{F}_1\left(\mathbf{U}^+_{j-\frac{1}{2}}\right) + \mathbf{U}^-_{j+\frac{1}{2}} + \left(\frac{2\hat{\omega}_1}{\lambda} - 1\right)^{-1} \mathbf{F}_1\Big(\mathbf{U}^-_{j+\frac{1}{2}}\Big)\right].$$

Since $B_{1,j}(x)$ is a constant and $2\hat{\omega}_1/\lambda - 1 \geq 1$, the generalized LxF property in Theorem 2.4 tell us that $\Xi_1, \Xi_2 \in \mathcal{G}$. Using $\hat{\omega}_1 + \hat{\omega}_L - \lambda > 0$, (3.6), and the convexity of $\mathcal{G}$ gives $\bar{\mathbf{U}}_j^{n+1} \in \mathcal{G}$. The proof is complete. □

Theorem 3.2 gives two sufficient conditions on the approximate function $\mathbf{U}_j^n(x)$ for that the scheme (3.4) is PCP. The first condition is easily ensured in practice since the flux for B_1 is zero and the divergence-free condition (1.2) in the case of $d = 1$ implies that B_1 is always a constant for the exact solution to (1.1). To meet the second condition, we need a PCP limiting procedure, in which $\mathbf{U}_j^n(x)$ is limited as $\tilde{\mathbf{U}}_j(x)$ satisfying $\tilde{\mathbf{U}}_j(\hat{x}_j^\alpha) \in \mathcal{G}$.

To avoid the effect of the rounding error, we introduce a sufficiently small positive number[b] ϵ such that $\bar{\mathbf{U}}_j^n \in \mathcal{G}_\epsilon$, where

$$\mathcal{G}_\epsilon = \{\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \mid D \geq \epsilon, q(\mathbf{U}) \geq \epsilon, \Psi_\epsilon(\mathbf{U}) \geq 0\}, \tag{3.7}$$

with

$$\Psi_\epsilon(\mathbf{U}) := \Psi(\mathbf{U}_\epsilon), \quad \mathbf{U}_\epsilon := (D, \mathbf{m}, \mathbf{B}, E - \epsilon)^\top.$$

Then the `1D PCP limiting procedure` is divided into the following three steps.

Step (i): Enforce the positivity of $D(\mathbf{U})$. Let $D_{\min} = \min_{x\in\mathcal{S}_j} D_j(x)$, where $\mathcal{S}_j := \{\hat{x}_j^\alpha\}_{\alpha=1}^L$. If $D_{\min} < \epsilon$, then $D_j(x)$ is limited as

$$\hat{D}_j(x) = \theta_1(D_j(x) - \bar{D}_j^n) + \bar{D}_j^n,$$

where $\theta_1 = (\bar{D}_j^n - \epsilon)/(\bar{D}_j^n - D_{\min})$. Otherwise, take $\hat{D}_j(x) = D_j(x)$ and $\theta_1 = 1$. Denote $\hat{\mathbf{U}}_j(x) := (\hat{D}_j(x), \mathbf{m}_j(x), \mathbf{B}_j(x), E_j(x))^\top$.

Step (ii): Enforce the positivity of $q(\mathbf{U})$. Let $q_{\min} = \min_{x\in\mathcal{S}_j} q(\hat{\mathbf{U}}_j(x))$. If $q_{\min} < \epsilon$, then $\hat{\mathbf{U}}_j(x)$ is limited as

$$\breve{\mathbf{U}}_j(x) = (\theta_2(\hat{D}_j(x) - \bar{D}_j^n) + \bar{D}_j^n, \theta_2(\hat{\mathbf{m}}_j(x) - \bar{\mathbf{m}}_j^n) + \bar{\mathbf{m}}_j^n,$$
$$\hat{\mathbf{B}}_j(x), \theta_2(\hat{E}_j(x) - \bar{\mathbf{E}}_j^n) + \bar{E}_j^n)^\top,$$

where $\theta_2 = (q(\bar{\mathbf{U}}_j^n) - \epsilon)/(q(\bar{\mathbf{U}}_j^n) - q_{\min})$. Otherwise, set $\breve{\mathbf{U}}_j(x) = \hat{\mathbf{U}}_j(x)$ and $\theta_2 = 1$.

Step (iii): Enforce the positivity of $\Psi(\mathbf{U})$. For each $x \in \mathcal{S}_j$, if $\Psi_\epsilon(\breve{\mathbf{U}}_j(x)) < 0$, then define $\tilde{\theta}(x)$ by solving the nonlinear equation

$$\Psi_\epsilon((1 - \tilde{\theta})\bar{\mathbf{U}}_j^n + \tilde{\theta}\breve{\mathbf{U}}_j(x)) = 0, \quad \tilde{\theta} \in [0, 1). \tag{3.8}$$

Otherwise, set $\tilde{\theta}(x) = 1$. Let $\theta_3 = \min_{x\in\mathcal{S}_j}\{\tilde{\theta}(x)\}$ and

$$\tilde{\mathbf{U}}_j(x) = \theta_3(\breve{\mathbf{U}}_j(x) - \bar{\mathbf{U}}_j^n) + \bar{\mathbf{U}}_j^n. \tag{3.9}$$

Remark 3.1. For some high-order finite volume methods, it only needs to reconstruct the limiting values $\mathbf{U}^\pm_{j+\frac{1}{2}}$, instead of the polynomial vector $\mathbf{U}_j(x)$. For this case, due to the proof of Theorem 3.2, it is sufficient that the limiting values satisfy

$$\mathbf{U}^+_{j-\frac{1}{2}}, \mathbf{U}^-_{j+\frac{1}{2}}, \frac{\bar{\mathbf{U}}_j^n - \hat{\omega}_1\mathbf{U}^+_{j-\frac{1}{2}} - \hat{\omega}_L\mathbf{U}^-_{j+\frac{1}{2}}}{1 - 2\hat{\omega}_1} \in \mathcal{G},$$

for all j. Similar to the discussions in Sec. 5 of Ref. 47, the previous PCP limiting procedure can be easily revised to meet such condition.

[b]In practice, ϵ can be chosen as 10^{-13}, and certainly it may be different for three constraints in (3.7). However, for the extreme problems with $E \gg 1$, $\epsilon = 10^{-13}\bar{E}_j^n$ is a good choice for the last constraint.

If replacing $\mathbf{U}^{\pm}_{j+\frac{1}{2}}$ in (3.4) respectively by

$$\tilde{\mathbf{U}}^{-}_{j+\frac{1}{2}} = \tilde{\mathbf{U}}_j(x_{j+\frac{1}{2}}), \quad \tilde{\mathbf{U}}^{+}_{j+\frac{1}{2}} = \tilde{\mathbf{U}}_{j+1}(x_{j+\frac{1}{2}}),$$

then the resulting scheme is PCP under the CFL-type condition (3.5), according to the conclusion (ii) of the coming Lemma 3.1. The above PCP limiter satisfies

$$\begin{aligned}\bar{\mathbf{U}}^n_j &= \frac{1}{\Delta x}\int_{I_j} \mathbf{U}_j(x)dx = \frac{1}{\Delta x}\int_{I_j} \hat{\mathbf{U}}_j(x)dx \\ &= \frac{1}{\Delta x}\int_{I_j} \breve{\mathbf{U}}_j(x)dx = \frac{1}{\Delta x}\int_{I_j} \tilde{\mathbf{U}}_j(x)dx,\end{aligned}$$

and that $\tilde{B}_{1,j}(x)$ remains constant for any $x \in I_j$ and j if $B_{1,j}(x)$ is constant for any $x \in I_j$ and j. It also preserves the accuracy for smooth solutions, similar to Ref. 46. The scheme (3.4) is only first-order accurate in time. To achieve high-order accurate PCP scheme in time and space, one can replace the forward Euler time discretization in (3.4) with high-order accurate strong stability-preserving (SSP) methods.[20] For example, utilizing the third-order accurate SSP Runge–Kutta method obtains:

$$\begin{aligned}\bar{\mathbf{U}}^*_j &= \bar{\mathbf{U}}^n_j + \Delta t_n \mathcal{L}(\tilde{\mathbf{U}}_j(x); j), \\ \bar{\mathbf{U}}^{**}_j &= \frac{3}{4}\bar{\mathbf{U}}^n_j + \frac{1}{4}(\bar{\mathbf{U}}^*_j + \Delta t_n \mathcal{L}(\tilde{\mathbf{U}}^*_j(x); j)), \\ \bar{\mathbf{U}}^{n+1}_j &= \frac{1}{3}\bar{\mathbf{U}}^n_j + \frac{2}{3}(\bar{\mathbf{U}}^{**}_j + \Delta t_n \mathcal{L}(\tilde{\mathbf{U}}^{**}_j(x); j)),\end{aligned} \tag{3.10}$$

where $\tilde{\mathbf{U}}_j(x)$, $\tilde{\mathbf{U}}^*_j(x)$ and $\tilde{\mathbf{U}}^{**}_j(x)$ denote the PCP limited versions of the reconstructed or evolved polynomial vector function at each Runge–Kutta stage, and

$$\mathcal{L}(\mathbf{U}_j(x); j) := -\frac{1}{\Delta x}\left(\hat{\mathbf{F}}_1\left(\mathbf{U}^{-}_{j+\frac{1}{2}}, \mathbf{U}^{+}_{j+\frac{1}{2}}\right) - \hat{\mathbf{F}}_1\left(\mathbf{U}^{-}_{j-\frac{1}{2}}, \mathbf{U}^{+}_{j-\frac{1}{2}}\right)\right).$$

Since such SSP method is a convex combination of the forward Euler method, the resulting high-order scheme is still PCP under the CFL-type condition (3.5) by the convexity of $\mathcal{G}$. Moreover, similar to Ref. 41, the PCP schemes hold a discrete L^1-type stability for the solution components $\tilde{D}_j(x)$, $\tilde{\mathbf{m}}_j(x)$ and $\tilde{E}_j(x)$. It is worth noting that the set $\mathcal{G}_\epsilon$ in (3.7) is convex thanks to the convexity of $\mathcal{G}_0$ so that the solution to (3.8) is unique. This allows that one can use some root-finding methods such as the bisection method to numerically solve (3.8). Moreover, one can show $\mathcal{G}_\epsilon \subset \mathcal{G}_0$ and $\lim_{\epsilon\to 0^+} \mathcal{G}_\epsilon = \bar{\mathcal{G}}_0$.

Lemma 3.1. (i) $\mathcal{G}_\epsilon \subset \mathcal{G}_0$. (ii) *If* $\bar{\mathbf{U}}^n_j \in \mathcal{G}_\epsilon$, *then* $\tilde{\mathbf{U}}_j(x)$ *defined in* (3.9) *belongs to* $\mathcal{G}_\epsilon$ *for all* $x \in \mathcal{S}_j$.

Proof. (i) Let us first prove $\mathcal{G}_\epsilon \subset \mathcal{G}_0$. For any $\mathbf{U} = (D, \mathbf{m}, \mathbf{B}, E)^\top \in \mathcal{G}_\epsilon$, one has $D \geq \epsilon > 0,\ q(\mathbf{U}) \geq \epsilon > 0$, and $\Psi(\mathbf{U}_\epsilon) \geq 0$ with $\mathbf{U}_\epsilon := (D, \mathbf{m}, \mathbf{B}, E - \epsilon)^\top$. Taking

partial derivative of $\Psi(\mathbf{U})$ with respect to E gives

$$\begin{aligned}\frac{\partial\Psi}{\partial E} &= \left(\frac{\partial\Phi(\mathbf{U})}{\partial E}+2\right)\sqrt{\Phi(\mathbf{U})+|\mathbf{B}|^2-E}+\frac{\Phi(\mathbf{U})-2(|\mathbf{B}|^2-E)}{2\sqrt{\Phi(\mathbf{U})+|\mathbf{B}|^2-E}}\left(\frac{\partial\Phi(\mathbf{U})}{\partial E}-1\right)\\ &= \frac{3}{2\sqrt{\Phi(\mathbf{U})+|\mathbf{B}|^2-E}}\left(\Phi(\mathbf{U})\frac{\partial\Phi(\mathbf{U})}{\partial E}+\Phi(\mathbf{U})+2(|\mathbf{B}|^2-E)\right)\\ &= \frac{3}{2\sqrt{\Phi(\mathbf{U})+|\mathbf{B}|^2-E}}\left(\Phi(\mathbf{U})\frac{4E-|\mathbf{B}|^2}{\Phi(\mathbf{U})}+\Phi(\mathbf{U})+2(|\mathbf{B}|^2-E)\right)\\ &= \frac{3(2E+|\mathbf{B}|^2+\Phi(\mathbf{U}))}{2\sqrt{\Phi(\mathbf{U})+|\mathbf{B}|^2-E}}>0,\end{aligned}$$

for any $\mathbf{U}\in\mathcal{G}_2$. This implies $\Psi(\mathbf{U})>\Psi(\mathbf{U}_\epsilon)\geq 0$, and concludes that $\mathbf{U}\in\mathcal{G}_0$. Therefore, $\mathcal{G}_\epsilon\subseteq\mathcal{G}_0$. Because $(\frac{\epsilon}{2},\mathbf{0},\mathbf{0},\epsilon)^\top$ belongs to $\mathcal{G}_0$, but it does not in $\mathcal{G}_\epsilon$, one has $\mathcal{G}_\epsilon\subset\mathcal{G}_0$.

(ii) Next we prove $\tilde{\mathbf{U}}_j(x)\in\mathcal{G}_\epsilon$ for any $x\in\mathcal{S}_j$. The above PCP limiting procedure yields

$$\hat{D}_j(x)\geq\epsilon,\quad q(\breve{\mathbf{U}}_j(x))\geq\epsilon,\quad \Psi_\epsilon(\tilde{\mathbf{U}}_j(x))\geq 0,$$

for $x\in\mathcal{S}_j$. For any $x\in\mathcal{S}_j$, one has

$$\begin{aligned}\tilde{D}_j(x) &= \theta_3(\breve{D}_j(x)-\bar{D}_j^n)+\bar{D}_j^n=\theta_2\theta_3(\hat{D}_j(x)-\bar{D}_j^n)+\bar{D}_j^n\\ &\geq \theta_2\theta_3(\epsilon-\bar{D}_j^n)+\bar{D}_j^n\geq\epsilon,\end{aligned}$$

by noting $\theta_2,\theta_3\in[0,1]$. Similarly, making use of the concavity of $q(\mathbf{U})$ gives

$$\begin{aligned}q(\tilde{\mathbf{U}}_j(x)) &= q(\theta_3\breve{\mathbf{U}}_j(x)+(1-\theta_3)\bar{\mathbf{U}}_j^n)\geq\theta_3 q(\breve{\mathbf{U}}_j(x))+(1-\theta_3)q(\bar{\mathbf{U}}_j^n)\\ &\geq\theta_3\epsilon+(1-\theta_3)\epsilon=\epsilon.\end{aligned}$$

The proof is complete. □

3.2. 2D *PCP schemes*

For the sake of convenience, this subsection will use the symbols (x,y) to replace the independent variables (x_1,x_2) in (1.1). Assume that the spatial domain is divided into a uniform rectangular mesh with cells $\{I_{ij}=(x_{i-\frac{1}{2}},x_{i+\frac{1}{2}})\times(y_{j-\frac{1}{2}},y_{j+\frac{1}{2}})\}$ and the spatial step-sizes Δx and Δy in x- and y-directions respectively, and the time interval is also divided into the (non-uniform) 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-type condition. Let $\bar{\mathbf{U}}_{ij}^n$ be the numerical approximation to the cell average value of the exact solution $\mathbf{U}(x,y,t)$ over I_{ij} at $t=t_n$. Our aim is to seek numerical schemes of the 2D RMHD equations (1.1), whose solution $\bar{\mathbf{U}}_{ij}^n$ stays at $\mathcal{G}$ if $\bar{\mathbf{U}}_{ij}^0\in\mathcal{G}$.

3.2.1. *First-order accurate schemes*

Consider the first-order accurate LxF-type scheme

$$\bar{\mathbf{U}}_{ij}^{n+1} = \bar{\mathbf{U}}_{ij}^{n} - \frac{\Delta t_n}{\Delta x}(\hat{\mathbf{F}}_1(\bar{\mathbf{U}}_{ij}^{n}, \bar{\mathbf{U}}_{i+1,j}^{n}) - \hat{\mathbf{F}}_1(\bar{\mathbf{U}}_{i-1,j}^{n}, \bar{\mathbf{U}}_{ij}^{n}) - \frac{\Delta t_n}{\Delta y}(\hat{\mathbf{F}}_2(\bar{\mathbf{U}}_{ij}^{n}, \bar{\mathbf{U}}_{i,j+1}^{n}) - \hat{\mathbf{F}}_2(\bar{\mathbf{U}}_{i,j-1}^{n}, \bar{\mathbf{U}}_{ij}^{n}), \tag{3.11}$$

where the x- and y-directional numerical fluxes $\hat{\mathbf{F}}_1$ and $\hat{\mathbf{F}}_2$ are defined as (3.2). If $\bar{\mathbf{U}}_{ij}^{n}$ belongs to $\mathcal{G}$ for all i, j, but the magnetic field $\bar{\mathbf{B}}_{ij}^{n}$ is not divergence-free in the discrete sense, then the solution $\bar{\mathbf{U}}_{ij}^{n+1}$ of (3.11) may not belong to $\mathcal{G}$, see Example 3.1. It means that the scheme (3.11) is not PCP in general when the divergence of magnetic field is nonzero.

Example 3.1. For any $\epsilon > 0$, take the primitive variable vectors $\hat{\mathbf{V}} = (\epsilon, 0.5, 0, 0, 0, 0, 0, \epsilon)^\top$ and $\tilde{\mathbf{V}} = (\epsilon, 0.5, 0, 0, 1, 0, 0, \epsilon)^\top$, and let $\hat{\mathbf{U}} = \hat{\mathbf{U}}(\hat{\mathbf{V}})$ and $\tilde{\mathbf{U}} = \tilde{\mathbf{U}}(\tilde{\mathbf{V}})$ be corresponding conservative vectors. If taking $\mathbf{U}_{i+1,j}^{n} = \tilde{\mathbf{U}} \in \mathcal{G}$ and $\mathbf{U}_{ij}^{n} = \mathbf{U}_{i,j\pm1}^{n} = \mathbf{U}_{i-1,j}^{n} = \hat{\mathbf{U}} \in \mathcal{G}$, then substituting them into (3.11) gives

$$\bar{\mathbf{U}}_{ij}^{n+1}(\epsilon) = \left(\frac{2\sqrt{3}}{3}\epsilon, \left(\frac{4}{3} + \frac{2}{3(\Gamma-1)}\right)\epsilon + \frac{\Delta t_n}{4\Delta x}, 0, 0, \frac{\Delta t_n}{2\Delta x}, 0, 0, \left(\frac{5}{3} + \frac{4}{3(\Gamma-1)}\right)\epsilon + \frac{\Delta t_n}{4\Delta x}\right)^\top.$$

Because of the continuity of $\tilde{q}(\bar{\mathbf{U}}_{ij}^{n+1}(\epsilon))$ with respect to ϵ, one has

$$\lim_{\epsilon\to 0^+} \tilde{q}\Big(\bar{\mathbf{U}}_{ij}^{n+1}(\epsilon)\Big) = \tilde{q}\Big(\bar{\mathbf{U}}_{ij}^{n+1}(0)\Big) = 27\left(\frac{\Delta t_n}{4\Delta x}\right)^7\left(\frac{2\Delta t_n}{\Delta x} + 1\right)^2\left(\frac{\Delta t_n}{\Delta x} - 4\right) < 0,$$

for any time step-size Δt_n satisfying the linear stability condition $\frac{\Delta t_n}{\Delta x} + \frac{\Delta t_n}{\Delta y} \le 1$. The locally sign-preserving property for continuous functions implies that there is a small positive number ϵ_0 such that $\tilde{q}(\bar{\mathbf{U}}_{ij}^{n+1}(\epsilon_0)) < 0$. Hence $\bar{\mathbf{U}}_{ij}^{n+1}(\epsilon_0) \notin \mathcal{G}$ thanks to Remark 2.1.

The above example shows clearly that it is necessary for a PCP RMHD code to preserve the discrete divergence-free condition, and the locally divergence-free condition of magnetic field within the cell cannot ensure the PCP property even for a first-order accurate scheme. The divergence-free MHD code is very important in the MHDs, see e.g. Refs. 9, 16, 37, etc. The nonzero divergence of the numerical magnetic field may lead to the generation of nonphysical wave or the negative pressure or density.[9,36] Although some works, e.g. Refs. 6, 10, 11 and 31, have discussed the positivity-preserving schemes for the non-relativistic MHD equations, up to now no any multi-dimensional MHD numerical scheme is rigorously proved to be PCP in theory.

If the scheme (3.11) satisfies a discrete divergence-free condition, then one can use the generalized LxF splitting property in Theorem 2.5 to prove that the scheme (3.11) is PCP.

Theorem 3.3. *The solution* $\bar{\mathbf{U}}_{ij}^n$ *of* (3.11) *satisfies the discrete divergence-free condition*

$$\mathrm{div}_{ij}\,\bar{\mathbf{B}}^n := \frac{(\bar{B}_1)_{i+1,j}^n - (\bar{B}_1)_{i-1,j}^n}{2\Delta x} + \frac{(\bar{B}_2)_{i,j+1}^n - (\bar{B}_2)_{i,j-1}^n}{2\Delta y} = 0, \tag{3.12}$$

for all $n \in \mathbb{N}$ *and* i, j, *if* (3.12) *holds for the discrete initial data* $\{\bar{\mathbf{U}}_{ij}^0\}$.

Proof. It is proved by the induction argument for the time level number n. The conclusion is true for $n = 0$ due to the hypothesis. Now assume that (3.12) holds for a non-negative integer n and all i, j, and then check whether the conclusion holds for $n+1$. Using (3.2) and noting that the fifth component of $\mathbf{F}_1$ and the sixth component of $\mathbf{F}_2$ are zero, one can rewrite the fifth and sixth equations in (3.11) as

$$\begin{aligned}(\bar{B}_1)_{i,j}^{n+1} &= (1-\lambda)(\bar{B}_1)_{i,j}^n + \frac{\Delta t_n}{2\Delta x}((\bar{B}_1)_{i+1,j}^n + (\bar{B}_1)_{i-1,j}^n) \\ &\quad + \frac{\Delta t_n}{2\Delta y}((\bar{B}_1)_{i,j+1}^n + (\bar{B}_1)_{i,j-1}^n) + \frac{\Delta t_n}{2\Delta y}(\Omega_{i,j+1} - \Omega_{i,j-1}), \end{aligned}\tag{3.13}$$

$$\begin{aligned}(\bar{B}_2)_{i,j}^{n+1} &= (1-\lambda)(\bar{B}_2)_{i,j}^n + \frac{\Delta t_n}{2\Delta x}((\bar{B}_2)_{i+1,j}^n + (\bar{B}_2)_{i-1,j}^n) \\ &\quad + \frac{\Delta t_n}{2\Delta y}((\bar{B}_2)_{i,j+1}^n + (\bar{B}_2)_{i,j-1}^n) + \frac{\Delta t_n}{2\Delta x}(-\Omega_{i+1,j} + \Omega_{i-1,j}), \end{aligned}\tag{3.14}$$

where Ω_{ij} denotes the sixth component of $\mathbf{F}_1(\bar{\mathbf{U}}_{ij}^n)$, and the fact that Ω_{ij} is equal to the opposite number of the fifth component of $\mathbf{F}_2(\bar{\mathbf{U}}_{ij}^n)$ has been used. Since the operator div_{ij} in (3.12) is linear, using (3.13) and (3.14) gives

$$\begin{aligned}\mathrm{div}_{ij}\,\bar{\mathbf{B}}^{n+1} &= (1-\lambda)\mathrm{div}_{ij}\,\bar{\mathbf{B}}^n + \frac{\Delta t_n}{2\Delta x}(\mathrm{div}_{i+1,j}\,\bar{\mathbf{B}}^n + \mathrm{div}_{i-1,j}\,\bar{\mathbf{B}}^n) \\ &\quad + \frac{\Delta t_n}{2\Delta y}(\mathrm{div}_{i,j+1}\,\bar{\mathbf{B}}^n + \mathrm{div}_{i,j-1}\,\bar{\mathbf{B}}^n) \\ &\quad + \frac{\Delta t_n}{2\Delta x\Delta y}[(\Omega_{i+1,j+1} - \Omega_{i+1,j-1}) - (\Omega_{i-1,j+1} - \Omega_{i-1,j-1})] \\ &\quad + \frac{\Delta t_n}{2\Delta x\Delta y}[(-\Omega_{i+1,j+1} + \Omega_{i-1,j+1}) - (-\Omega_{i+1,j-1} + \Omega_{i-1,j-1})] \\ &= (1-\lambda)\mathrm{div}_{ij}\,\bar{\mathbf{B}}^n + \frac{\Delta t_n}{2\Delta x}(\mathrm{div}_{i+1,j}\,\bar{\mathbf{B}}^n + \mathrm{div}_{i-1,j}\,\bar{\mathbf{B}}^n) \\ &\quad + \frac{\Delta t_n}{2\Delta y}(\mathrm{div}_{i,j+1}\,\bar{\mathbf{B}}^n + \mathrm{div}_{i,j-1}\,\bar{\mathbf{B}}^n) = 0, \end{aligned}\tag{3.15}$$

where the induction hypothesis has been used in the last equal sign. Hence (3.12) holds for all $n \in \mathbb{N}$ and j. □

Theorem 3.4. *If* $\bar{\mathbf{U}}_{ij}^n =: (\bar{D}_{ij}^n, \bar{\mathbf{m}}_{ij}^n, \bar{\mathbf{B}}_{ij}^n, \bar{E}_{ij}^n)^\top \in \mathcal{G}$ *satisfies the discrete divergence-free condition* (3.12) *for all* i *and* j*, then under the CFL-type condition*

$$0 < \frac{c\Delta t_n}{\Delta x} + \frac{c\Delta t_n}{\Delta y} \le 1, \tag{3.16}$$

the solution $\bar{\mathbf{U}}_{ij}^{n+1}$ *of* (3.11) *belongs to* $\mathcal{G}$ *for all* i *and* j*, where* $c = 1$ *is the speed of light.*

Proof. Substituting (3.2) into (3.11) gives

$$\begin{aligned}\bar{\mathbf{U}}_{ij}^{n+1} &= \frac{\lambda}{2(\frac{1}{\Delta x} + \frac{1}{\Delta y})}\left[\frac{1}{\Delta x}(\bar{\mathbf{U}}_{i+1,j}^n - \mathbf{F}_1(\bar{\mathbf{U}}_{i+1,j}^n) + \bar{\mathbf{U}}_{i-1,j}^n + \mathbf{F}_1(\bar{\mathbf{U}}_{i-1,j}^n))\right.\\ &\quad \left. + \frac{1}{\Delta y}(\bar{\mathbf{U}}_{i,j+1}^n - \mathbf{F}_2(\bar{\mathbf{U}}_{i,j+1}^n) + \bar{\mathbf{U}}_{i,j-1}^n + \mathbf{F}_2(\bar{\mathbf{U}}_{i,j-1}^n))\right] + (1-\lambda)\bar{\mathbf{U}}_j^n\\ &=: \lambda\Xi + (1-\lambda)\bar{\mathbf{U}}_{ij}^n,\end{aligned}$$

where $\lambda := \Delta t_n(\frac{1}{\Delta x} + \frac{1}{\Delta y}) \in (0, 1]$ due to (3.16). Using the condition (3.12) and Theorem 2.5 gives $\Xi \in \mathcal{G}$. The convexity of $\mathcal{G}$ further yields $\bar{\mathbf{U}}_{ij}^{n+1} \in \mathcal{G}$. The proof is completed. □

Let us discuss how to get the discrete initial data which are admissible, i.e. $\bar{\mathbf{U}}_{ij}^0 \in \mathcal{G}$, and satisfy the condition (3.12) for all i and j. After giving initial data $(\rho, \mathbf{v}, \mathbf{B}, p)(x, y, 0)$, calculate the cell average values of the initial primitive variables $(\rho, \mathbf{v}, \mathbf{B}, p)$ by

$$(\bar{\rho}_{ij}^0, \bar{\mathbf{v}}_{ij}^0, (\bar{B}_3)_{ij}^0, \bar{p}_{ij}^0) = \frac{1}{\Delta x\Delta y}\iint_{I_{ij}} (\rho, \mathbf{v}, B_3, p)(x, y, 0)dxdy, \tag{3.17}$$

and

$$(\bar{B}_1)_{ij}^0 = \frac{1}{2\Delta y}\int_{y_{j-1}}^{y_{j+1}} B_1(x_i, y, 0)dy, \quad (\bar{B}_2)_{ij}^0 = \frac{1}{2\Delta x}\int_{x_{i-1}}^{x_{i+1}} B_2(x, y_j, 0)dx, \tag{3.18}$$

for each i and j, then $\bar{\mathbf{U}}_{ij}^0 = \mathbf{U}(\bar{\mathbf{V}}_{ij}^0)$ belongs to $\mathcal{G}$ and satisfies the condition (3.12) for all i and j. In fact, one has $\bar{\rho}_{ij}^0 > 0$, $\bar{p}_{ij}^0 > 0$, and

$$\begin{aligned}|\bar{\mathbf{v}}_{ij}^0|^2 &= \sum_{k=1}^3 \left(\iint_{I_{ij}} \frac{1}{\Delta x\Delta y} \cdot v_k(x, y, 0)dxdy\right)^2\\ &\le \sum_{k=1}^3 \left(\iint_{I_{ij}} \left(\frac{1}{\Delta x\Delta y}\right)^2 dxdy\right)\left(\iint_{I_{ij}} v_k^2(x, y, 0)dxdy\right)\\ &= \frac{1}{\Delta x\Delta y}\left(\iint_{I_{ij}} v^2(x, y, 0)dxdy\right) < 1,\end{aligned}$$

where the Cauchy–Schwarz inequality has been used in the penultimate inequality. Moreover, with (3.18), it holds

$$\begin{aligned}\mathrm{div}_{ij}\,\bar{\mathbf{B}}^0 &= \frac{1}{2\Delta x}\left(\frac{1}{2\Delta y}\int_{y_{j-1}}^{y_{j+1}} B_1(x_{i+1},y,0)dy - \frac{1}{2\Delta y}\int_{y_{j-1}}^{y_{j+1}} B_1(x_{i-1},y,0)dy\right)\\ &\quad+\frac{1}{2\Delta y}\left(\frac{1}{2\Delta x}\int_{x_{i-1}}^{x_{i+1}} B_2(x,y_{j+1},0)dx - \frac{1}{2\Delta x}\int_{x_{i-1}}^{x_{i+1}} B_2(x,y_{j-1},0)dx\right)\\ &= \frac{1}{4\Delta x\Delta y}\int_{x_{i-1}}^{x_{i+1}}\int_{y_{j-1}}^{y_{j+1}}\left(\frac{\partial B_1}{\partial x}+\frac{\partial B_2}{\partial y}\right)dxdy = 0,\end{aligned}$$

where the divergence theorem and (1.2) for $t = 0$ have been used. In practical computations, the integrals in (3.17) and (3.18) can be approximately calculated by some numerical quadratures so that the condition (3.12) may not hold exactly due to the numerical error. Fortunately, the discrete divergence error $\mathcal{E}_\infty^n := \max_{ij}\{|\mathrm{div}_{ij}\,\bar{\mathbf{B}}^n|\}$ does not grow with n under the condition (3.16), because using (3.15) and the triangular inequality gives

$$\mathcal{E}_\infty^{n+1} \le \mathcal{E}_\infty^n.$$

3.2.2. *High-order accurate schemes*

This subsection discusses the high-order accurate PCP schemes for the 2D RMHD equations (1.1).

Assume that the approximate solution $\mathbf{U}_{ij}^n(x,y)$ at time $t = t_n$ within the cell I_{ij} is either reconstructed in the finite volume methods from the cell average values $\{\bar{\mathbf{U}}_{ij}^n\}$ or evolved in the DG methods. The function $\mathbf{U}_{ij}^n(x,y)$ is a vector of the polynomial of degree K, and its cell average value over the cell I_{ij} is equal to $\bar{\mathbf{U}}_{ij}^n$. Moreover, let $\mathbf{U}_{i\pm\frac{1}{2},j}^{\mp}(y)$ and $\mathbf{U}_{i,j\pm\frac{1}{2}}^{\mp}(x)$ denote the traces of $\mathbf{U}_{ij}^n(x,y)$ on the four edges $\{x_{i\pm\frac{1}{2}}\}\times(y_{j-\frac{1}{2}},y_{j+\frac{1}{2}})$ and $(x_{i-\frac{1}{2}},x_{i+\frac{1}{2}})\times\{y_{j\pm\frac{1}{2}}\}$ of the cell I_{ij}, respectively.

For the 2D RMHD equations (1.1), the finite volume scheme or discrete equation for the cell average value in the DG method may be given by

$$\begin{aligned}\bar{\mathbf{U}}_{ij}^{n+1} &= \bar{\mathbf{U}}_{ij}^n - \frac{\Delta t_n}{\Delta x}\sum_{\beta=1}^{Q}\omega_\beta\left(\hat{\mathbf{F}}_1\left(\mathbf{U}_{i+\frac{1}{2},\beta}^-,\mathbf{U}_{i+\frac{1}{2},\beta}^+\right)-\hat{\mathbf{F}}_1\left(\mathbf{U}_{i-\frac{1}{2},\beta}^-,\mathbf{U}_{i-\frac{1}{2},\beta}^+\right)\right)\\ &\quad-\frac{\Delta t_n}{\Delta y}\sum_{\alpha=1}^{Q}\omega_\alpha\left(\hat{\mathbf{F}}_2\left(\mathbf{U}_{\alpha,j+\frac{1}{2}}^-,\mathbf{U}_{\alpha,j+\frac{1}{2}}^+\right)-\hat{\mathbf{F}}_2\left(\mathbf{U}_{\alpha,j-\frac{1}{2}}^-,\mathbf{U}_{\alpha,j-\frac{1}{2}}^+\right)\right),\end{aligned}\tag{3.19}$$

which is an approximation of the equation

$$\begin{aligned}\bar{\mathbf{U}}_{ij}^{n+1} &= \bar{\mathbf{U}}_{ij}^n - \frac{\Delta t_n}{\Delta x\Delta y}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}\hat{\mathbf{F}}_1\left(\mathbf{U}_{i+\frac{1}{2},j}^-(y),\mathbf{U}_{i+\frac{1}{2},j}^+(y)\right)\\ &\quad-\hat{\mathbf{F}}_1\left(\mathbf{U}_{i-\frac{1}{2},j}^-(y),\mathbf{U}_{i-\frac{1}{2},j}^+(y)\right)dy\end{aligned}$$

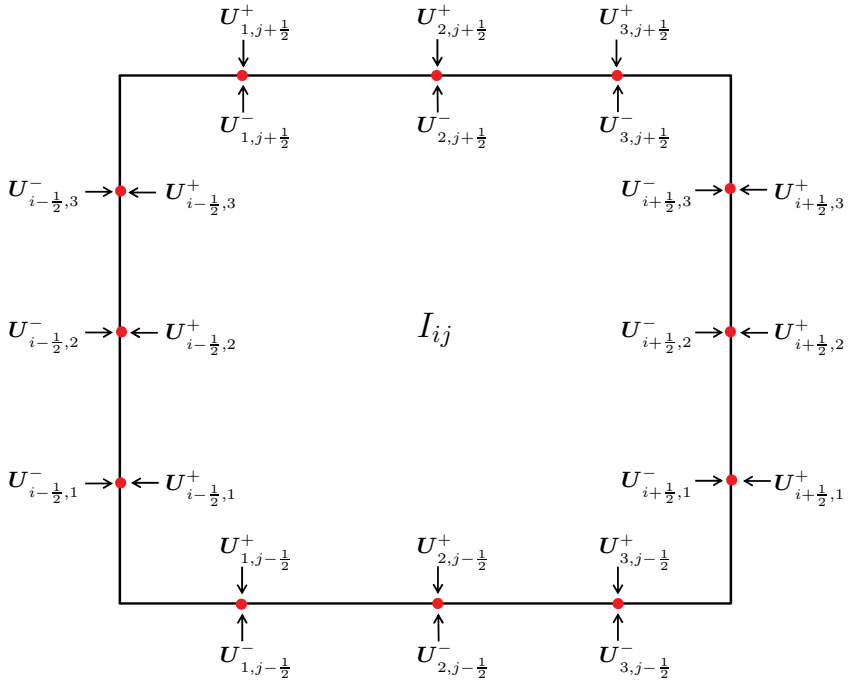


Fig. 1. The limiting values at Q Gaussian nodes on four edges of the cell I_{ij} with $Q = 3$.

$$-\frac{\Delta t_n}{\Delta x \Delta y}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \hat{\mathbf{F}}_2\left(\mathbf{U}^-_{i,j+\frac{1}{2}}(x), \mathbf{U}^+_{i,j+\frac{1}{2}}(x)\right)$$

$$-\hat{\mathbf{F}}_2\left(\mathbf{U}^-_{i,j-\frac{1}{2}}(x), \mathbf{U}^+_{i,j-\frac{1}{2}}(x)\right) dx,$$

by using the Gaussian quadrature for each integral with Q nodes and the weights $\{\omega_\alpha\}_{\alpha=1}^Q$ satisfying $\sum_{\alpha=1}^Q \omega_\alpha = 1$. In (3.19), $\hat{\mathbf{F}}_1$ and $\hat{\mathbf{F}}_2$ denote the numerical fluxes in x- and y-directions respectively, and are taken as the LxF flux defined in (3.2). Moreover, as shown schematically in Fig. 1, the limiting values $\mathbf{U}^\pm_{i+\frac{1}{2},\beta} = \mathbf{U}^\pm_{i+\frac{1}{2},j}(y_j^\beta)$ and $\mathbf{U}^\pm_{\alpha,j+\frac{1}{2}} = \mathbf{U}^\pm_{i,j+\frac{1}{2}}(x_i^\alpha)$, where $\{x_i^\alpha\}_{\alpha=1}^Q$ and $\{y_j^\alpha\}_{\alpha=1}^Q$ denote the Gaussian nodes transformed into the intervals $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ and $[y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$, respectively. For the accuracy requirement, Q should satisfy: $Q \geq K+1$ for a $\mathbb{P}^K$-based DG method,[13] or $Q \geq (K+1)/2$ for a $(K+1)$th order accurate finite volume scheme.

Let $\{\hat{x}_i^\alpha\}_{\alpha=1}^L$ and $\{\hat{y}_j^\alpha\}_{\alpha=1}^L$ be the Gauss–Lobatto nodes transformed into the intervals $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ and $[y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$ respectively, and $\{\hat{\omega}_\alpha\}_{\alpha=1}^L$ be associated weights satisfying $\sum_{\alpha=1}^L \hat{\omega}_\alpha = 1$. Here $L \geq (K+3)/2$ such that the algebraic precision degree of corresponding quadrature is at least K. Similar to Theorem 3.2

for the 1D case, we have the following sufficient conditions for that the high-order accurate scheme (3.19) is PCP.

Theorem 3.5. *If* $\mathbf{U}_{ij}^n(x,y) =: \big(D_{ij}(x,y), \mathbf{m}_{ij}(x,y), \mathbf{B}_{ij}(x,y), E_{ij}(x,y)\big)^\top$ *satisfy*:

(i) *the discrete divergence-free conditions*:

$$\begin{aligned}\mathrm{div}_{ij}^{\mathrm{in}}\,\mathbf{B} \triangleq\ & \frac{1}{\Delta x}\sum_{\beta=1}^{Q}\omega_\beta\left((B_1)^-_{i+\frac12,\beta}-(B_1)^+_{i-\frac12,\beta}\right)\\ &+\frac{1}{\Delta y}\sum_{\beta=1}^{Q}\omega_\beta\left((B_2)^-_{\beta,j+\frac12}-(B_2)^+_{\beta,j-\frac12}\right)=0,\end{aligned}\tag{3.20}$$

$$\begin{aligned}\mathrm{div}_{ij}^{\mathrm{out}}\,\mathbf{B} \triangleq\ & \frac{1}{\Delta x}\sum_{\beta=1}^{Q}\omega_\beta\left((B_1)^+_{i+\frac12,\beta}-(B_1)^-_{i-\frac12,\beta}\right)\\ &+\frac{1}{\Delta y}\sum_{\beta=1}^{Q}\omega_\beta\left((B_2)^+_{\beta,j+\frac12}-(B_2)^-_{\beta,j-\frac12}\right)=0,\end{aligned}\tag{3.21}$$

for all i *and* j, *and*

(ii) $\mathbf{U}_{ij}^n(\hat{x}_i^\alpha, y_j^\beta)$, $\mathbf{U}_{ij}^n(x_i^\beta, \hat{y}_j^\alpha) \in \mathcal{G}$, *for all* i,j,α,β, *then under the CFL-type condition*

$$0 < \frac{\Delta t_n}{\Delta x}+\frac{\Delta t_n}{\Delta y} \le \hat{\omega}_1,\tag{3.22}$$

the solution $\bar{\mathbf{U}}_{ij}^{n+1}$ *of the scheme* (3.19) *belongs to* $\mathcal{G}$.

Proof. The exactness of the Gauss–Lobatto quadrature rule with L nodes and the Gauss quadrature rule with Q nodes for the polynomials of degree K yields

$$\begin{aligned}\bar{\mathbf{U}}_{ij}^n &= \frac{1}{\Delta x\Delta y}\iint_{I_{ij}}\mathbf{U}_{ij}^n(x,y)dxdy = \frac{1}{\Delta x}\int_{x_{i-\frac12}}^{x_{i+\frac12}}\left(\sum_{\beta=1}^{Q}\omega_\beta\mathbf{U}_{ij}^n(x,y_j^\beta)\right)dx\\ &= \sum_{\beta=1}^{Q}\omega_\beta\left(\frac{1}{\Delta x}\int_{x_{i-\frac12}}^{x_{i+\frac12}}\mathbf{U}_{ij}^n(x,y_j^\beta)dx\right) = \sum_{\beta=1}^{Q}\omega_\beta\left(\sum_{\alpha=1}^{L}\hat{\omega}_\alpha\mathbf{U}_{ij}^n(\hat{x}_i^\alpha,y_j^\beta)\right)\\ &= \sum_{\alpha=1}^{L}\sum_{\beta=1}^{Q}\hat{\omega}_\alpha\omega_\beta\mathbf{U}_{ij}^n(\hat{x}_i^\alpha,y_j^\beta).\end{aligned}\tag{3.23}$$

Similarly, one has

$$\bar{\mathbf{U}}_{ij}^n = \sum_{\alpha=1}^{L}\sum_{\beta=1}^{Q}\hat{\omega}_\alpha\omega_\beta\mathbf{U}_{ij}^n(x_i^\beta,\hat{y}_j^\alpha).\tag{3.24}$$

By combining (3.23) and (3.24), and using $\mathbf{U}_{ij}^n(\hat{x}_i^1, y_j^\beta) = \mathbf{U}_{i-\frac{1}{2},\beta}^+$, $\mathbf{U}_{ij}^n(\hat{x}_i^L, y_j^\beta) = \mathbf{U}_{i+\frac{1}{2},\beta}^-$, $\mathbf{U}_{ij}^n(x_i^\beta, \hat{y}_j^1) = \mathbf{U}_{\beta,j-\frac{1}{2}}^+$, $\mathbf{U}_{ij}^n(x_i^\beta, \hat{y}_j^L) = \mathbf{U}_{\beta,j+\frac{1}{2}}^-$, and $\hat{\omega}_1 = \hat{\omega}_L$, one has

$$
\begin{aligned}
\bar{\mathbf{U}}_{ij}^n &= \frac{\lambda_x}{\lambda_x+\lambda_y}\bar{\mathbf{U}}_{ij}^n + \frac{\lambda_y}{\lambda_x+\lambda_y}\bar{\mathbf{U}}_{ij}^n \\
&= \frac{\lambda_x}{\lambda_x+\lambda_y}\sum_{\alpha=1}^{L}\sum_{\beta=1}^{Q}\hat{\omega}_\alpha\omega_\beta\mathbf{U}_{ij}^n(\hat{x}_i^\alpha, y_j^\beta) + \frac{\lambda_y}{\lambda_x+\lambda_y}\sum_{\alpha=1}^{L}\sum_{\beta=1}^{Q}\hat{\omega}_\alpha\omega_\beta\mathbf{U}_{ij}^n(x_i^\beta, \hat{y}_j^\alpha) \\
&= \frac{\lambda_x}{\lambda_x+\lambda_y}\sum_{\alpha=2}^{L-1}\sum_{\beta=1}^{Q}\hat{\omega}_\alpha\omega_\beta\mathbf{U}_{ij}^n(\hat{x}_i^\alpha, y_j^\beta) + \frac{\lambda_y}{\lambda_x+\lambda_y}\sum_{\alpha=2}^{L-1}\sum_{\beta=1}^{Q}\hat{\omega}_\alpha\omega_\beta\mathbf{U}_{ij}^n(x_i^\beta, \hat{y}_j^\alpha) \\
&\quad + \frac{\lambda_x\hat{\omega}_1}{\lambda_x+\lambda_y}\sum_{\beta=1}^{Q}\omega_\beta\left(\mathbf{U}_{i-\frac{1}{2},\beta}^+ + \mathbf{U}_{i+\frac{1}{2},\beta}^-\right) + \frac{\lambda_y\hat{\omega}_1}{\lambda_x+\lambda_y}\sum_{\beta=1}^{Q}\omega_\beta\left(\mathbf{U}_{\beta,j-\frac{1}{2}}^+ + \mathbf{U}_{\beta,j+\frac{1}{2}}^-\right),
\end{aligned}
$$

where $\lambda_x := \Delta t_n/\Delta x$ and $\lambda_y := \Delta t_n/\Delta y$. Substituting the above identity and (3.2) into (3.19) gives

$$
\begin{aligned}
\bar{\mathbf{U}}_{ij}^{n+1} &= \frac{\lambda_x}{\lambda_x+\lambda_y}\sum_{\alpha=2}^{L-1}\sum_{\beta=1}^{Q}\hat{\omega}_\alpha\omega_\beta\mathbf{U}_{ij}^n\left(\hat{x}_i^\alpha, y_j^\beta\right) + \frac{\lambda_y}{\lambda_x+\lambda_y}\sum_{\alpha=2}^{L-1}\sum_{\beta=1}^{Q}\hat{\omega}_\alpha\omega_\beta\mathbf{U}_{ij}^n\left(x_i^\beta, \hat{y}_j^\alpha\right) \\
&\quad + \frac{\lambda_x\hat{\omega}_1}{\lambda_x+\lambda_y}\sum_{\beta=1}^{Q}\omega_\beta\left(\mathbf{U}_{i-\frac{1}{2},\beta}^+ + \mathbf{U}_{i+\frac{1}{2},\beta}^-\right) + \frac{\lambda_y\hat{\omega}_1}{\lambda_x+\lambda_y}\sum_{\beta=1}^{Q}\omega_\beta\left(\mathbf{U}_{\beta,j-\frac{1}{2}}^+ + \mathbf{U}_{\beta,j+\frac{1}{2}}^-\right) \\
&\quad - \lambda_x\sum_{\beta=1}^{Q}\omega_\beta\left(\hat{\mathbf{F}}_1\left(\mathbf{U}_{i+\frac{1}{2},\beta}^-, \mathbf{U}_{i+\frac{1}{2},\beta}^+\right) - \hat{\mathbf{F}}_1\left(\mathbf{U}_{i-\frac{1}{2},\beta}^-, \mathbf{U}_{i-\frac{1}{2},\beta}^+\right)\right) \\
&\quad - \lambda_y\sum_{\beta=1}^{Q}\omega_\beta\left(\hat{\mathbf{F}}_2\left(\mathbf{U}_{\beta,j+\frac{1}{2}}^-, \mathbf{U}_{\beta,j+\frac{1}{2}}^+\right) - \hat{\mathbf{F}}_2\left(\mathbf{U}_{\beta,j-\frac{1}{2}}^-, \mathbf{U}_{\beta,j-\frac{1}{2}}^+\right)\right) \\
&= (1-2\hat{\omega}_1)\,\Xi_1 + (\lambda_x+\lambda_y)\,\Xi_2 + (2\hat{\omega}_1 - (\lambda_x+\lambda_y))\,\Xi_3, \qquad (3.25)
\end{aligned}
$$

with

$$
\Xi_1 := \sum_{\alpha=2}^{L-1}\frac{\hat{\omega}_\alpha}{1-2\hat{\omega}_1}\left[\frac{\lambda_x}{\lambda_x+\lambda_y}\sum_{\beta=1}^{Q}\omega_\beta\mathbf{U}_{ij}\left(\hat{x}_i^\alpha, y_j^\beta\right) + \frac{\lambda_y}{\lambda_x+\lambda_y}\sum_{\beta=1}^{Q}\omega_\beta\mathbf{U}_{ij}\left(x_i^\beta, \hat{y}_j^\alpha\right)\right],
$$

$$
\begin{aligned}
\Xi_2 := \frac{1}{2\left(\frac{1}{\Delta x}+\frac{1}{\Delta y}\right)}\sum_{\beta=1}^{Q}\omega_\beta\Bigg[&\frac{1}{\Delta x}\left(\mathbf{U}_{i+\frac{1}{2},\beta}^+ - \mathbf{F}_1\left(\mathbf{U}_{i+\frac{1}{2},\beta}^+\right) + \mathbf{U}_{i-\frac{1}{2},\beta}^- + \mathbf{F}_1\left(\mathbf{U}_{i-\frac{1}{2},\beta}^-\right)\right) \\
&+ \frac{1}{\Delta y}\left(\mathbf{U}_{\beta,j+\frac{1}{2}}^+ - \mathbf{F}_2\left(\mathbf{U}_{\beta,j+\frac{1}{2}}^+\right) + \mathbf{U}_{\beta,j-\frac{1}{2}}^- + \mathbf{F}_2\left(\mathbf{U}_{\beta,j-\frac{1}{2}}^-\right)\right)\Bigg],
\end{aligned}
$$

$$\Xi_3 := \frac{1}{2\left(\frac{1}{\Delta x}+\frac{1}{\Delta y}\right)}\sum_{\beta=1}^{Q}\omega_\beta\Bigg[\frac{1}{\Delta x}\left(\mathbf{U}^-_{i+\frac{1}{2},\beta}-\theta^{-1}\mathbf{F}_1\left(\mathbf{U}^-_{i+\frac{1}{2},\beta}\right)\right.$$
$$\left.+\mathbf{U}^+_{i-\frac{1}{2},\beta}+\theta^{-1}\mathbf{F}_1\left(\mathbf{U}^+_{i-\frac{1}{2},\beta}\right)\right)+\frac{1}{\Delta y}\left(\mathbf{U}^-_{\beta,j+\frac{1}{2}}-\theta^{-1}\mathbf{F}_2\left(\mathbf{U}^-_{\beta,j+\frac{1}{2}}\right)\right.$$
$$\left.+\mathbf{U}^+_{\beta,j-\frac{1}{2}}+\theta^{-1}\mathbf{F}_2\left(\mathbf{U}^+_{\beta,j-\frac{1}{2}}\right)\right)\Bigg],$$

where $\theta := \frac{2\hat{\omega}_1}{\lambda_x+\lambda_y}-1\geq 1$. Thanks to the convexity of $\mathcal{G}$ and the condition (ii), $\Xi_1\in\mathcal{G}$. With $\mathbf{U}^{\mp}_{i\pm\frac{1}{2},\beta}, \mathbf{U}^{\mp}_{\beta,j\pm\frac{1}{2}}\in\mathcal{G}$, $\theta\geq 1$, and the condition (3.20), one has $\Xi_3\in\mathcal{G}$ by the generalized LxF splitting property in Theorem 2.5. Similarly, utilizing the condition (3.21) gives $\Xi_2\in\mathcal{G}$. Thus using (3.25) and the convexity of $\mathcal{G}$ yields $\bar{\mathbf{U}}^{n+1}_{ij}\in\mathcal{G}$, and completes the proof. □

Remark 3.2. Both (3.20) and (3.21) in the condition (i) are approximating (2.44) by replacing x_0 and y_0 with $x_{i-\frac{1}{2}}$ and $y_{j-\frac{1}{2}}$, respectively.

Remark 3.3. For some high-order finite volume methods, it only needs to reconstruct the limiting values $\mathbf{U}^+_{i-\frac{1}{2},\beta}$, $\mathbf{U}^-_{i+\frac{1}{2},\beta}$, $\mathbf{U}^+_{\beta,j-\frac{1}{2}}$, $\mathbf{U}^-_{\beta,j+\frac{1}{2}}$, instead of the polynomial vector $\mathbf{U}^n_{ij}(x,y)$. In this case, the condition (ii) in Theorem 3.5 can be replaced with the following condition:

$$\mathbf{U}^+_{i-\frac{1}{2},\beta},\mathbf{U}^-_{i+\frac{1}{2},\beta},\mathbf{U}^+_{\beta,j-\frac{1}{2}},\mathbf{U}^-_{\beta,j+\frac{1}{2}}\in\mathcal{G},\quad \beta=1,2,\ldots,Q,$$
$$\Xi_1=\frac{1}{1-2\hat{\omega}_1}\left(\bar{\mathbf{U}}^n_{ij}-\hat{\omega}_1\sum_{\beta=1}^{Q}\frac{\omega_\beta}{\lambda_x+\lambda_y}\Big(\lambda_x\Big(\mathbf{U}^+_{i-\frac{1}{2},\beta}+\mathbf{U}^-_{i+\frac{1}{2},\beta}\Big)\right.$$
$$\left.+\lambda_y\Big(\mathbf{U}^+_{\beta,j-\frac{1}{2}}+\mathbf{U}^-_{\beta,j+\frac{1}{2}}\Big)\Big)\right)\in\mathcal{G},$$

for all i and j.

It is worth emphasizing that the above discussions can be extended to non-uniform or unstructured meshes by using Theorem 2.7. Theorem 3.5 provides two sufficient conditions (i) and (ii) on the function $\mathbf{U}^n_{ij}(x,y)$ reconstructed in the finite volume method or evolved in the DG method in order to ensure that the numerical schemes (3.19) are PCP. The condition (ii) can be easily met by using the PCP limiter similar to that in Sec. 3.1.2, but Eqs. (3.20) and (3.21) in the condition (i) are two constraints on the discrete divergence. By using the *divergence theorem*, it can be seen that the discrete divergence-free condition (3.20) may be met if the reconstructed or evolved polynomial vector $(B_1,B_2)_{ij}(x,y)$ is locally divergence-free, see e.g. Refs. 27 and 48. The locally divergence-free property of $(B_1,B_2)_{ij}(x,y)$ is not destroyed in the PCP limiting procedure since the PCP limiter modifies

the vectors $\mathbf{U}_{ij}^n(x,y)$ only with a simple scaling. The condition (3.21) is necessary for a PCP numerical scheme for the RMHD equations, see Example 3.1, where the magnetic vector satisfies (3.20) and the condition (ii). Numerical results in Sec. 4 will further demonstrate the importance of condition (3.21). However, it is not easy to meet the condition (3.21) because (3.21) depends on the limiting values of the magnetic field calculated from the neighboring cells $I_{i\pm1,j}$ and $I_{i,j\pm1}$ of I_{ij}. If the polynomials $(B_1,B_2)_{ij}(x,y)$ are globally or exactly divergence-free, in other words, it is locally divergence-free in I_{ij} with normal magnetic component continuous across the cell interface, then (3.20) and (3.21) are satisfied. But the PCP limiter with local scaling may destroy the globally or exactly divergence-free property of $(B_1,B_2)_{ij}(x,y)$. Hence, it is non-trivial and still open to design a limiting procedure for the polynomial vector $\mathbf{U}_{ij}^n(x,y)$ satisfying two sufficient conditions in Theorem 3.5 at the same time.

Remark 3.4. As the mesh is refined, it can be weakened that violating the condition (3.21) impacts on the PCP property, if the reconstructed or evolved polynomial vector $(B_1,B_2)_{ij}(x,y)$ is locally divergence-free, i.e. $\mathrm{div}_{ij}^{\mathrm{in}}\mathbf{B}=0$. In fact, the proof of Theorem 3.5 shows that the condition (3.21) is only related to $\Xi_2\in\mathcal{G}$. If assuming that $(B_1,B_2)_{ij}(x,y)$ approximates the exact solution $(B_1,B_2)(x,y,t_n)$ with at least first order, then the continuity of $B_1(x,y,t_n)$ across the edge $\{x_{i+\frac12}\}\times(y_{j-\frac12},y_{j+\frac12})$ implies

$$(B_1)^-_{i+\frac12,\beta}=B_1\Big(x_{i+\frac12},y_j^\beta,t_n\Big)+\mathcal{O}(\Delta),\quad (B_1)^+_{i+\frac12,\beta}=B_1\Big(x_{i+\frac12},y_j^\beta,t_n\Big)+\mathcal{O}(\Delta),$$

where $\Delta=\min\{\Delta x,\Delta y\}$. Similarly, one has

$$(B_2)^-_{\beta,j+\frac12}=B_2\Big(x_i^\beta,y_{j+\frac12},t_n\Big)+\mathcal{O}(\Delta),\quad (B_2)^-_{\beta,j+\frac12}=B_2\Big(x_i^\beta,y_{j+\frac12},t_n\Big)+\mathcal{O}(\Delta).$$

It follows that $\mathrm{div}_{ij}^{\mathrm{out}}\mathbf{B}=\mathrm{div}_{ij}^{\mathrm{in}}\mathbf{B}+\mathcal{O}(1)=\mathcal{O}(1)$, so that Ξ_2 may not belong to $\mathcal{G}$. However, Ξ_2 is very close to $\mathcal{G}$ in the sense of that the first component of Ξ_2 is positive, and for any $\mathbf{B}^*,\mathbf{v}^*\in\mathbb{R}^3$ with $|\mathbf{v}^*|<1$, it holds

$$\Xi_2\cdot\mathbf{n}^*+p_m^*\geq-\frac{\mathbf{v}^*\cdot\mathbf{B}^*}{2(\frac{1}{\Delta x}+\frac{1}{\Delta y})}(\mathrm{div}_{ij}^{\mathrm{out}}\mathbf{B})=-\mathcal{O}(\Delta),$$

whose derivation is similar to that of Theorem 2.5. Therefore, as Δ approaches 0, the convex combination in (3.25) becomes more possibly in $\mathcal{G}$.

4. Numerical Experiments

This section conducts numerical experiments on several 1D and 2D challenging RMHD problems with either large Lorentz factors, strong discontinuities, low plasma-beta $\beta:=p/p_m$, or low rest-mass density or pressure, to demonstrate our theoretical analyses and the performance of the proposed PCP limiter. Without loss of generality, we take the (third-order accurate) $\mathbb{P}^2$-based, locally divergence-free DG methods,[48] together with the third-order SSP Runge–Kutta time disretization (3.10), as our base schemes. According to the analysis in Sec. 3.1.2, the 1D-base

scheme with the proposed PCP limiter results in a PCP DG scheme. As discussed in Sec. 3.2.2, the 2D base scheme with such limiter may not be PCP in general, because the discrete divergence-free condition (3.21) in Theorem 3.5 is not strictly satisfied even though the locally divergence-free property can ensure the condition (3.20). However, it will be shown in the following that the PCP limiter can still improve the robustness of 2D DG method. To meet the conditions (3.5) and (3.22), the time step-sizes in 1D and 2D will be taken as $0.15\Delta x$ and $0.15(1/\Delta x + 1/\Delta y)^{-1}$, respectively. Unless otherwise stated, all the computations are restricted to the EOS (2.1) with the adiabatic index $\Gamma = 5/3$.

Example 4.1. (Smooth problems) A 1D and a 2D smooth problems are respectively solved within the domain $[0,1]^d$ on the uniform meshes of N^d cells to test the accuracy of the $\mathbb{P}^2$-based DG methods with the proposed PCP limiter.

The 1D problem describes Alfvén waves propagating periodically with large velocity of 0.99 and low pressure, and has the exact solution

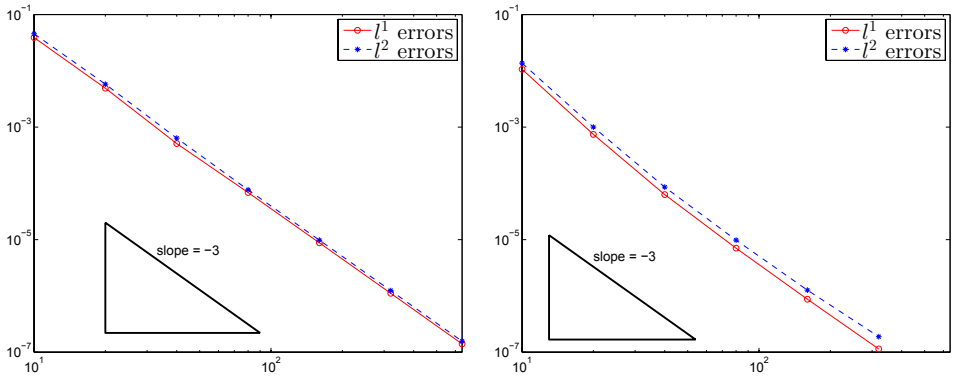
$$\mathbf{V}(x,t) = (1, 0, v_2, v_3, 1, \kappa v_2, \kappa v_3, 10^{-2})^\top, \quad (x,t) \in [0,1] \times \mathbb{R}^+,$$

where $v_2 = 0.99\sin(2\pi(x + t/\kappa))$, $v_3 = 0.99\cos(2\pi(x + t/\kappa))$, and $\kappa = \sqrt{1 + \rho h W^2}$. While the 2D problem's exact solution is given by

$$\mathbf{V}(x,y,t) = (1 + 0.99999999\sin(2\pi(x+y)), 0.9, 0.2, 0, 1, 1, 1, 10^{-2})^\top, \\ (x,y,t) \in [0,1]^2 \times \mathbb{R}^+,$$

which describes a RMHD sine wave fast propagating with low density and pressure.

Figure 2 displays the l^1- and l^2-errors at $t = 1$ and corresponding orders obtained by using the proposed DG methods, respectively. The results show that the theoretical orders are obtained in both 1D and 2D cases, and the PCP limiting procedure does not destroy the accuracy.



(a) Errors in v_2 for 1D smooth problem (b) Errors in ρ for 2D smooth problem

Fig. 2. Example 4.1: Numerical l^1- and l^2-errors at $t = 1$. The horizontal axis represents the value of N.

To verify the capability of DG methods with PCP limiter in resolving 1D and 2D ultra-relativistic wave configurations, three 1D Riemann problems (RPs), a 2D shock and cloud interaction problem, and several 2D blast problems will be solved. For those problems, before using the PCP limiting procedure, the WENO limiter[35,48] will be implemented with the aid of the local characteristic decomposition[2] to enhance the numerical stability of high-order DG methods in resolving the strong discontinuities as well as their interactions. Different from Ref. 48, the improved WENO proposed in Ref. 8 and the "trouble" cell indicator in Ref. 26 are used here.

Example 4.2. (1D RPs) This example verifies the robustness and effectiveness of the PCP DG method by simulating three 1D RPs, whose initial data comprise two different constant states separated by the initial discontinuity at $x = 0$, see Table 1. The computational domain is $[-0.5, 0.5]$ and divided into 1000 uniform cells.

The first two RPs are similar to but more ultra than those 1D blast wave problems in Refs. 3 and 19. Specifically, the stronger magnetic field ($|\mathbf{B}| \approx 37.108$) appears in the left state of the first problem, while a very strong initial jump in pressure ($\Delta p := |p_R - p_L|/p_R \approx 10^{12}$) and extremely low gas pressure (the minimum plasma-beta $\beta := p/p_m \approx 1.98 \times 10^{-10}$) in the second problem. The numerical results of those problem at $t = 0.4$ obtained by using the $\mathbb{P}^2$-based PCP DG method are displayed by symbols "$\circ$" in Figs. 3 and 4 respectively, where and hereafter the solid lines denote the reference solutions obtained by a second-order MUSCL scheme with PCP limiter over the uniform mesh of 20,000 cells. It is seen that the PCP DG method exhibits good resolution and strong robustness, and the results agree well with the reference ones. Without employing the PCP limiting procedure, the high-order accurate DG methods will break down quickly within few time steps due to nonphysical numerical solutions.

The third RP describes the strong collision between two high-speed RMHD flows with a Lorentz factor of about 223.61. As a result, it is a very strongly relativistic test problem. Figure 5 gives the numerical results at $t = 0.4$ obtained by using the $\mathbb{P}^2$-based PCP DG method. As the time increases, we see that two fast and two slow reflected shock waves are produced, and a very high pressure region appears between the two slow shock waves. Those shock waves are well resolved robustly, even though there exists the well-known wall-heating-type phenomenon around

Table 1. Initial data of the three 1D RPs in Example 4.2.

		ρ	v_1	v_2	v_3	B_1	B_2	B_3	p
RP I	Left state	1	0	0	0	5	26	26	30
	Right state	1	0	0	0	5	0.7	0.7	1
RP II	Left state	1	0	0	0	10	7	7	10^4
	Right state	1	0	0	0	10	0.7	0.7	10^{-8}
RP III	Left state	1	0.99999	0	0	100	70	70	0.1
	Right state	1	−0.99999	0	0	100	−70	−70	0.1

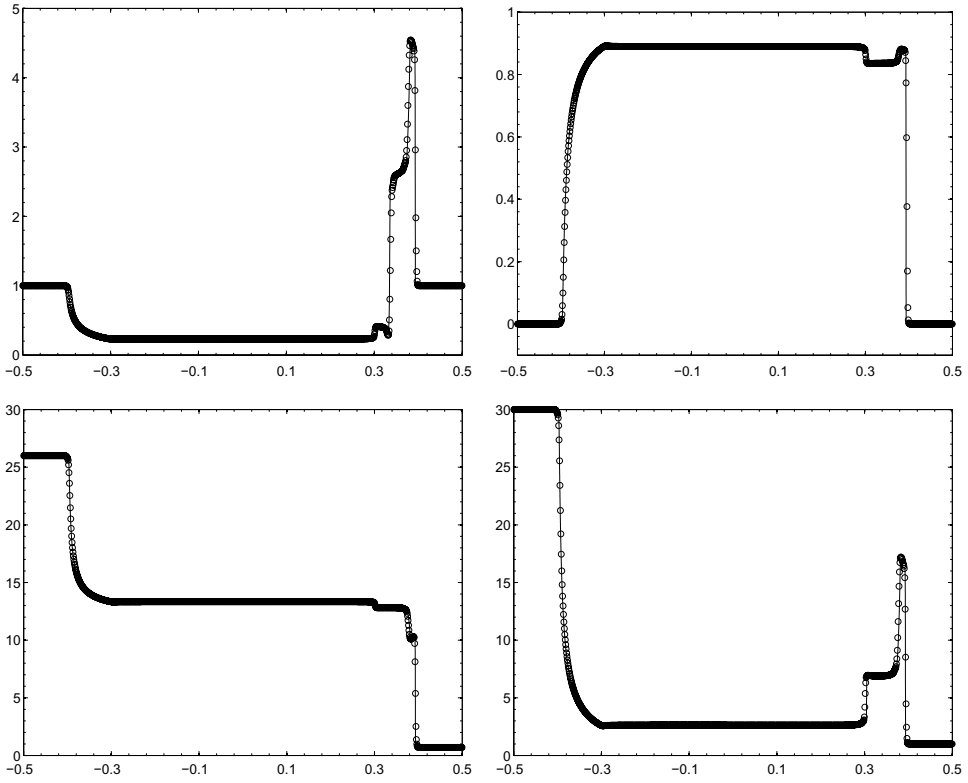


Fig. 3. RP I in Example 4.2: The density ρ (top-left), velocity v_1 (top-right), magnetic-field B_2 (bottom-left), and pressure p (bottom-right) at $t = 0.4$ obtained by the PCP DG method. The solid lines denote the reference solutions.

$x = 0$, which is often observed in the literatures e.g. Refs. 3 and 21. It is worth mentioning that the $\mathbb{P}^2$-based DG method fails in the first time step if the PCP limiting procedure is not employed.

Example 4.3. (Shock and cloud interaction problem) This problem describes the disruption of a high density cloud by a strong shock wave. The setup is the same as that in Ref. 21. Different from the setup in Ref. 30, the magnetic field is not orthogonal to the slab plane so that the magnetic divergence-free treatment has to be imposed. The computational domain is $[-0.2, 1.2] \times [0, 1]$, with the left boundary specified as inflow condition and the others as outflow conditions. Initially, a shock wave moves to the right from $x = 0.05$, with the left and right states $\mathbf{V}_L = (3.86859, 0.68, 0, 0, 0, 0.84981, -0.84981, 1.25115)^\top$ and $\mathbf{V}_R = (1, 0, 0, 0, 0, 0.16106, 0.16106, 0.05)^\top$, respectively. There exists a rest circular cloud centered at the point $(0.25, 0.5)$ with radius 0.15. The cloud has the same states to the surrounding fluid except for a higher density 30.

Figure 6 displays the Schlieren images of rest-mass density logarithm $\ln \rho$ and magnetic pressure logarithm $\ln p_m$ at $t = 1.2$ obtained by using the $\mathbb{P}^2$-based DG

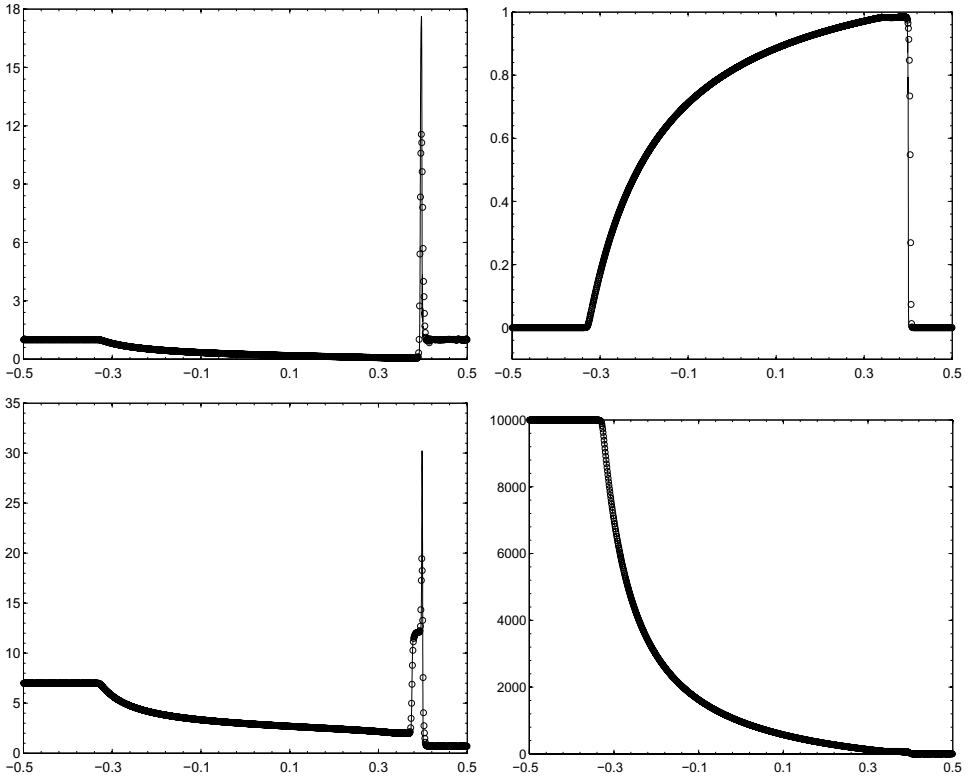


Fig. 4. Same as Fig. 3 except for RP II.

method with the PCP limiter over the uniform mesh of 560×400 cells. One can see that the discontinuities are captured with high resolution, and the results agree well with those in Ref. 21. In this test, it is also necessary to use the PCP limiting procedure for the successful performance of high-order accurate DG methods. The $\mathbb{P}^2$-based DG method without the PCP limiter will fail at $t \approx 0.05$ due to inadmissible numerical solutions.

Example 4.4. (Blast problems) Blast problem has become a standard test for 2D RMHD numerical schemes. If the low gas pressure, strong magnetic field or low plasma-beta $\beta := p/p_m$ is involved, then simulating those ultra-RMHD blast problems becomes very challenging.[29] Several different setups have been used in the literature, see e.g. Refs. 25, 30, 15, 44 and 29. Our setups are similar to that in Refs. 30, 15, 7 and 44. Initially, the computational domain $[-6, 6]^2$ is filled with a homogeneous gas at rest with adiabatic index $\Gamma = \frac{4}{3}$. The explosion zone ($r < 0.8$) has a density of 10^{-2} and a pressure of 1, while the ambient medium ($r > 1$) has a density of 10^{-4} and a pressure of $p_a = 5 \times 10^{-4}$, where $r = \sqrt{x^2 + y^2}$. A linear taper is applied to the density and pressure for $r \in [0.8, 1]$. The magnetic field is initialized in the x-direction as B_a. As B_a is set larger, the initial ambient

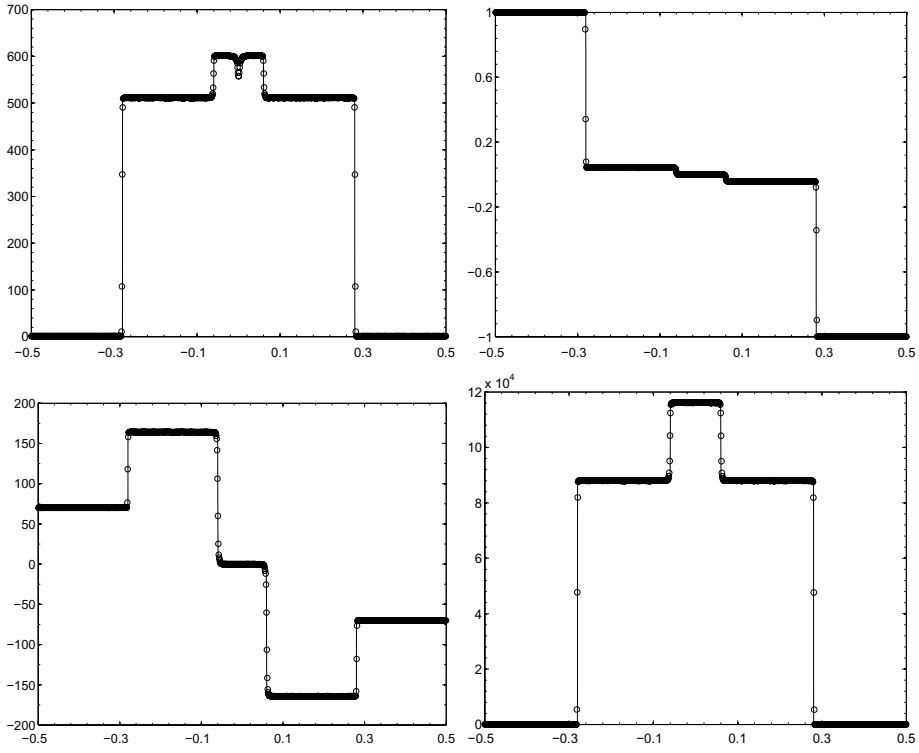


Fig. 5. Same as Fig. 3 except for RP III.

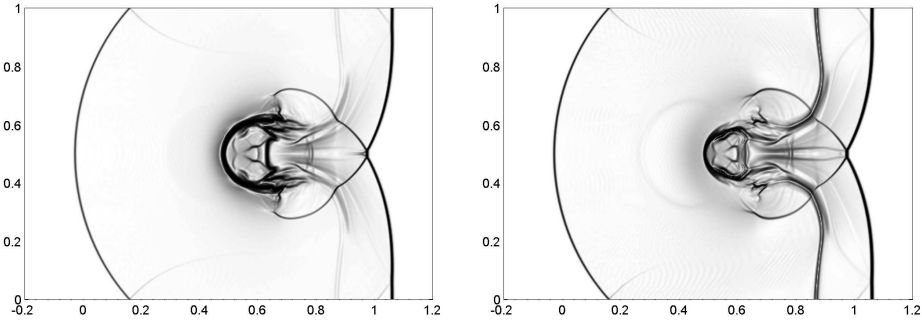


Fig. 6. Example 4.3: The Schlieren images of rest-mass density logarithm (left) and magnetic pressure logarithm (right) at $t = 1.2$.

magnetization becomes higher ($\beta_a := p_a/p_m$ becomes lower) and this test becomes more challenging. In the literatures,[30,15,7] B_a is usually specified as 0.1, which corresponds to a moderate magnetized case ($\beta_a = 0.1$). A more strongly magnetized case with $B_a = 0.5$ is tested in Ref. 44, corresponding to a lower plasma-beta $\beta_a = 4 \times 10^{-3}$. Most existing methods in literature need some artificial treatments

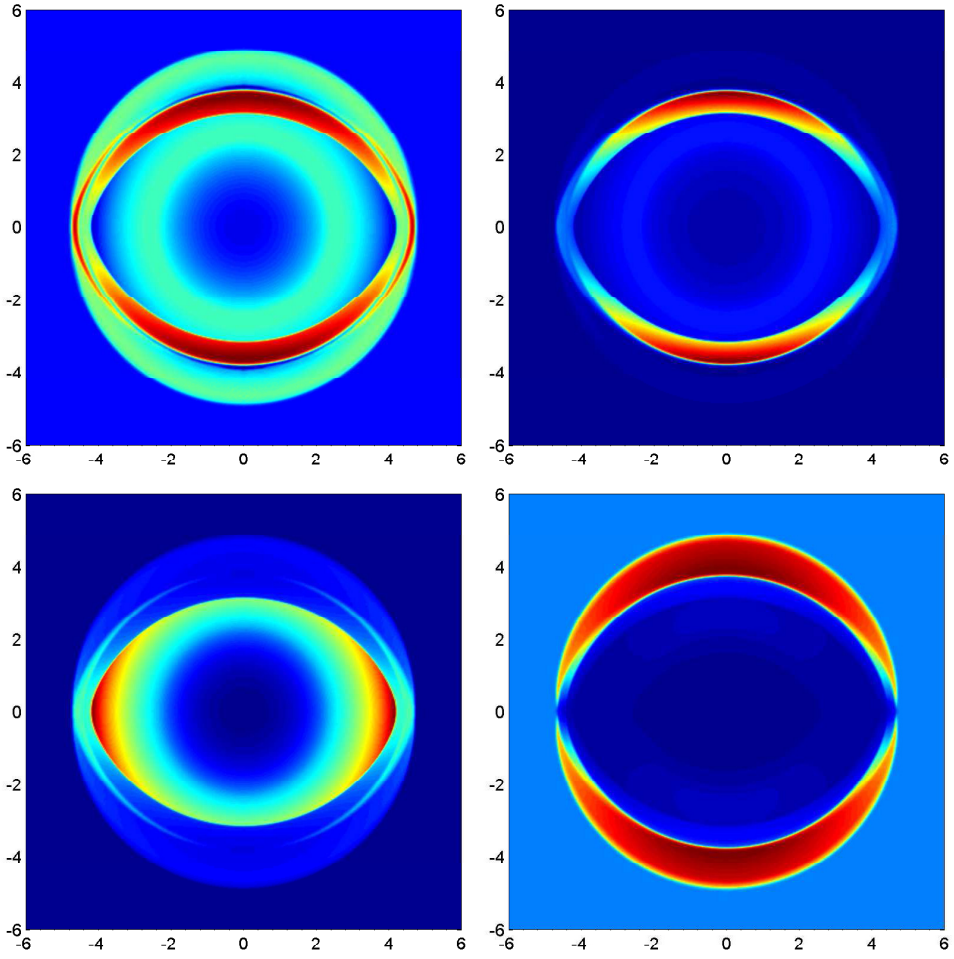


Fig. 7. Example 4.4 with $B_a = 0.1$: The Schlieren images of rest-mass density logarithm (top-left), gas pressure (top-right), Lorentz factor (bottom-left) and magnetic field strength (bottom-right) at $t = 4$.

for the strongly magnetized case, see e.g. Refs. 25 and 30. It is reported in Ref. 15 that the RMHD code ECHO is not able to run this test with $B_a > 0.1$ if no *ad hoc* numerical strategy is introduced.

Our numerical results of this test at $t = 4$ are shown in Fig. 7 for the moderately magnetized case with $B_a = 0.1$, in Fig. 8 for the relatively strongly magnetized case with $B_a = 0.5$, and in Fig. 9 for the strongly magnetized case with $B_a = 20$ (corresponding $\beta_a = 2.5 \times 10^{-6}$). All of them are obtained by using the $\mathbb{P}^2$-based DG method with the PCP limiter over the uniform mesh of 400×400 cells. During those simulations, the present method exhibits very good robustness without any artificial treatment. For the first two cases, our results agree quite well with those

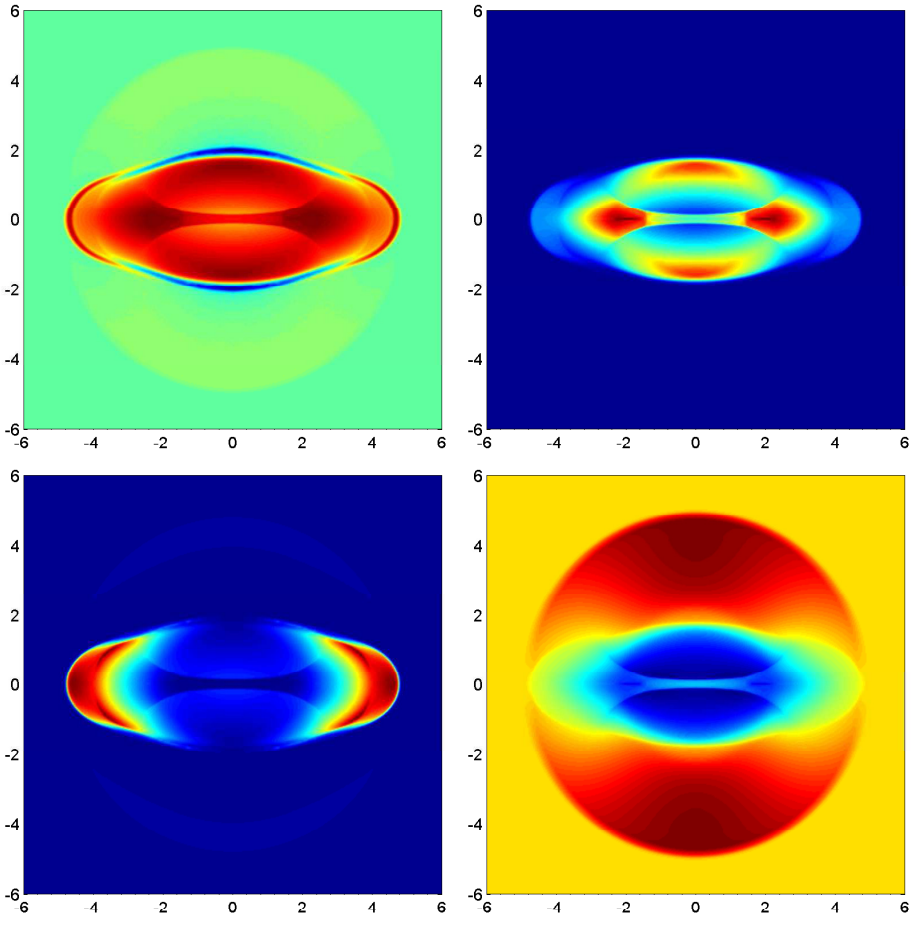


Fig. 8. Same as Fig. 7 except for $B_a = 0.5$.

reported in Refs. 44 and 7. From Fig. 7, it is observed that the wave pattern of the configuration is composed by two main waves, an external fast and a reverse shock wave. The former is almost circular, while the latter is somewhat elliptic. The magnetic field is essentially confined between them, while the inner region is almost devoid of magnetization. In the case of $B_a = 0.5$, the external circular fast shock is clearly visible in the rest-mass density and in the magnetic field, but very weak. When B_a is increased to 20, the external circular fast shock becomes much weaker and is only visible in the magnetic field in Fig. 9. As the magnetization is increased, the blast wave is confined to propagate along the magnetic field lines, creating a structure elongated in the x-direction.

To investigate the importance of discrete divergence-free condition (3.21) in Theorem 3.5, we now try to test a much lower plasma-beta case $\beta_a = 10^{-7}$ (i.e. $B_a = 100$) on the mesh of 400×400 cells. For such extreme case, our method breaks down

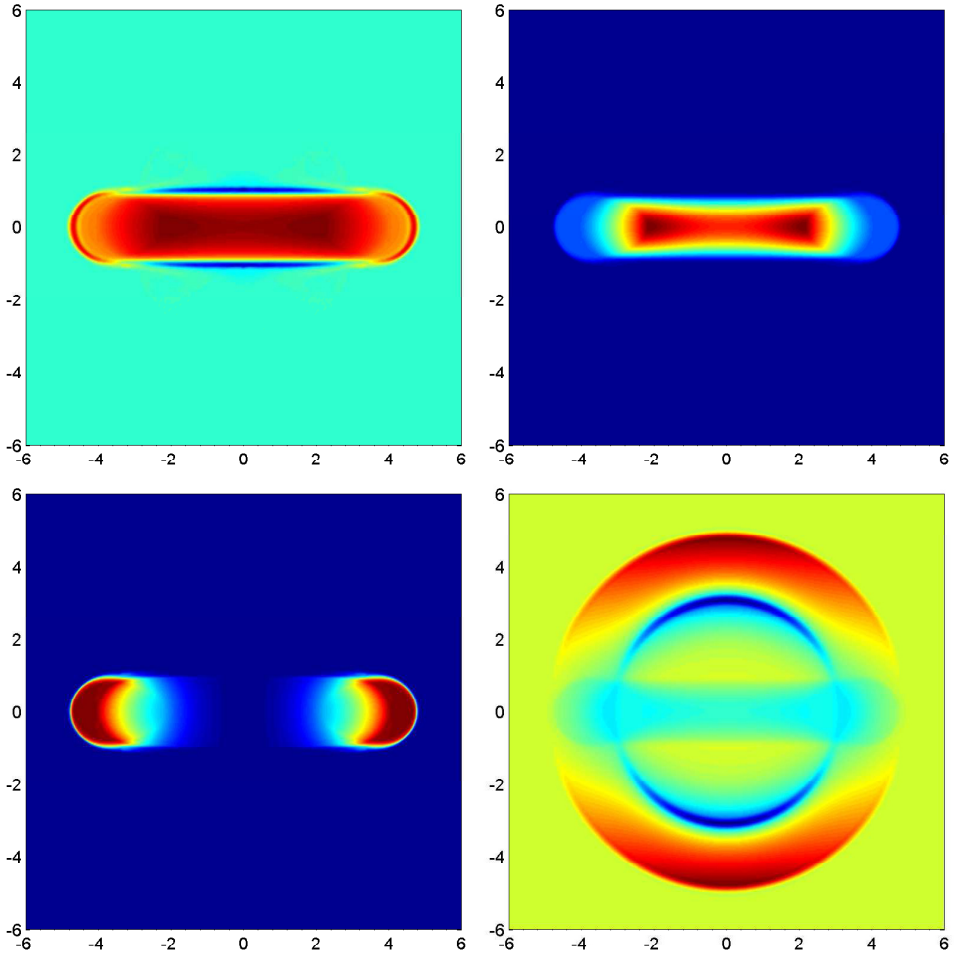


Fig. 9. Same as Fig. 7 except for $B_a = 20$.

at $t \approx 0.783$. This failure results from the computed inadmissible cell averages of conservative variables, detected in the three cells centered at points $(-0.225, 1.095)$, $(-0.165, 1.095)$ and $(-0.165, -1.095)$, respectively. Figure 10 displays the Schlieren image of $|\mathrm{div}_{ij}^{\mathrm{out}}\,\mathbf{B}|$ at the moment of failure. It clearly shows the subregions with large values of $|\mathrm{div}_{ij}^{\mathrm{out}}\,\mathbf{B}|$, where the condition (3.21) is violated most seriously, and the three detected cells are exactly located in those subregions. This further demonstrates that the condition (3.21) is really crucial in achieving completely PCP schemes in 2D. As mentioned in Remark 3.4, for the purpose of numerical simulation, it is possible to weaken the impact of violating (3.21) by refining the mesh. By numerical experiments, we find that our method can work successfully on a refined mesh of 600×600 cells for the case of $B_a = 100$ and an extremely strongly magnetized case with $B_a = 1500$ ($\beta_a \approx 4.444 \times 10^{-10}$). The flow structures

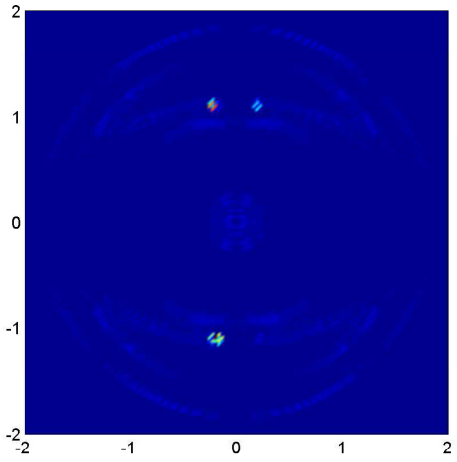


Fig. 10. Example 4.4 with $B_a = 100$: Close-up of the Schlieren image of $|\mathrm{div}_{ij}^{\mathrm{out}}\,\mathbf{B}|$ at $t = 0.783$ on the uniform mesh of 400×400 cells.

in those two cases are similar to the case of $B_a = 20$ and omitted here. To our best knowledge, the 2D blast test with so low plasma-beta is rarely considered in the literature.

5. Conclusions

The paper studied mathematical properties of the admissible state set $\mathcal{G}$ defined in (1.3) of the RMHD equations (1.1). In comparison with the non-relativistic and relativistic hydrodynamical cases (with the zero magnetic field), the difficulties mainly came from the extremely strong nonlinearities, no explicit formulas of the primitive variables and the flux vectors with respect to the conservative variables, and the solenoidal magnetic field. To overcome those difficulties, the first equivalent form of $\mathcal{G}$ with explicit constraints on the conservative vector was first skillfully discovered with the aid of polynomial root properties, and followed by the scaling invariance. The convexity of $\mathcal{G}$ was proved by utilizing the semi-positive definiteness of the second fundamental form of a hypersurface, and then the second equivalent form of $\mathcal{G}$ and the orthogonal invariance were obtained. It was verified that the LxF splitting property did not hold in general when the magnetic field was nonzero. However, by combining the convex combination of some LxF splitting terms with a "discrete divergence-free" condition for the magnetic field, the generalized LxF splitting properties were subtly discovered with a constructive inequality and some pivotal techniques. This revealed in theory for the first time the close connection between the "discrete divergence-free" condition and the PCP property of numerical schemes.

The above mathematical properties were footstone of studying PCP numerical schemes for RMHDs. Based on the resulting theoretical results, several 1D and

2D PCP schemes were studied for the first time. In the 1D case, a first-order accurate LxF-type scheme was first proved to be PCP under the CFL condition. Then, the high-order accurate 1D PCP schemes were proposed via a PCP limiter, which was designed by using the first equivalent form of $\mathcal{G}$. In the 2D case, the "discrete divergence-free" condition and PCP property were analyzed for a first-order accurate LxF-type scheme, and followed by two sufficient conditions for high-order accurate PCP schemes. Several numerical experiments were conducted to demonstrate the theoretical analyses and the performance of numerical schemes as well as the importance of discrete divergence-free condition in achieving genuinely PCP scheme in 2D. The studies on the PCP schemes may be easily extended to the 3D case by Theorem 2.6, the non-uniform or unstructured meshes by Theorem 2.7, and the general EOS case by the similar discussions in Ref. 41.

Acknowledgments

This work was partially supported by the Special Project on High-Performance Computing under the National Key R&D Program (No. 2016YFB0200603), Science Challenge Project (No. JCKY2016212A502), and the National Natural Science Foundation of China (Nos. 91330205, 91630310 and 11421101).

References

1. M. Anderson, E. W. Hirschmann, S. L. Liebling and D. Neilsen, Relativistic MHD with adaptive mesh refinement, *Class. Quantum Grav.* **23** (2006) 6503–6524.
2. L. Antón, J. A. Miralles, J. M. Martí, J. M. Ibáñez, M. A. Aloy and P. Mimica, Relativistic magnetohydrodynamics: Renormalized eigenvectors and full wave decomposition Riemann solver, *Astrophys. J. Suppl. Ser.* **188** (2010) 1–31.
3. D. S. Balsara, Total variation diminishing scheme for relativistic magnetohydrodynamics, *Astrophys. J. Suppl. Ser.* **132** (2001) 83–101.
4. D. S. Balsara, Second-order-accurate schemes for magnetohydrodynamics with divergence-free reconstruction, *Astrophys. J. Suppl. Ser.* **151** (2004) 149–184.
5. D. S. Balsara, Divergence-free reconstruction of magnetic fields and WENO schemes for magnetohydrodynamics, *J. Comput. Phys.* **228** (2009) 5040–5056.
6. D. S. Balsara, Self-adjusting, positivity preserving high order schemes for hydrodynamics and magnetohydrodynamics, *J. Comput. Phys.* **231** (2012) 7504–7517.
7. D. S. Balsara and J. Kim, A subluminal relativistic magnetohydrodynamics scheme with ADER–WENO predictor and multidimensional Riemann solver-based corrector, *J. Comput. Phys.* **312** (2016) 357–384.
8. R. Borges, M. Carmona, B. Costa and W. S. Don, An improved weighted essentially nonoscillatory scheme for hyperbolic conservation laws, *J. Comput. Phys.* **227** (2008) 3101–3211.
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) 426–430.
10. Y. Cheng, F. Y. Li, J. X. Qiu and L. W. Xu, Positivity-preserving DG and central DG methods for ideal MHD equations, *J. Comput. Phys.* **238** (2013) 255–280.

11. A. J. Christlieb, Y. Liu, Q. Tang and Z. F. Xu, Positivity-preserving finite difference weighted ENO schemes with constrained transport for ideal magnetohydrodynamic equations, *SIAM J. Sci. Comput.* **37** (2015) A1825–A1845.
12. M. Cissoko, Detonation waves in relativistic hydrodynamics, *Phys. Rev. D* **45** (1992) 1045–1052.
13. B. Cockburn, S. C. Hu and C.-W. Shu, The Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws IV: The multidimensional case, *Math. Comput.* **54** (1990) 545–581.
14. L. Del Zanna, N. Bucciantini and P. Londrillo, An efficient shock-capturing central-type scheme for multidimensional relativistic flows II. Magnetohydrodynamics, *Astron. Astrophys.* **400** (2003) 397–413.
15. L. Del Zanna, O. Zanotti, N. Bucciantini and P. Londrillo, ECHO: A Eulerian conservative high-order scheme for general relativistic magnetohydrodynamics and magnetodynamics, *Astron. Astrophys.* **473** (2007) 11–30.
16. C. R. Evans and J. F. Hawley, Simulation of magnetodydrodynamic flows: A constrained transport method, *Astrophys. J.* **332** (1988) 659–677.
17. J. A. Font, Numerical hydrodynamics and magnetohydrodynamics in general relativity, *Living Rev. Relativ.* **11** (2008) 7.
18. K. O. Friedrichs, On the laws of relativistic electro-magneto-fluid dynamics, *Commun. Pure Appl. Math.* **27** (1974) 749–808.
19. B. Giacomazzo and L. Rezzolla, The exact solution of the Riemann problem in relativistic magnetohydrodynamics, *J. Fluid Mech.* **562** (2006) 223–259.
20. S. Gottlieb, D. J. Ketcheson and C.-W. Shu, High order strong stability preserving time discretizations, *J. Sci. Comput.* **38** (2009) 251–289.
21. P. He and H. Z. Tang, An adaptive moving mesh method for two-dimensional relativistic magnetohydrodynamics, *Comput. Fluids* **60** (2012) 1–20.
22. V. Honkkila and P. Janhunen, HLLC solver for ideal relativistic MHD, *J. Comput. Phys.* **223** (2007) 643–656.
23. L. D. Jonker, Immersions with semi-definite second fundamental forms, *Canad. J. Math.* **27** (1975) 610–617.
24. J. Kim and D. S. Balsara, A stable HLLC Riemann solver for relativistic magnetohydrodynamics, *J. Comput. Phys.* **270** (2014) 634–639.
25. S. S. Komissarov, A Godunov-type scheme for relativistic magnetohydrodynamics, *Mon. Not. Roy. Astron. Soc.* **303** (1999) 343–366.
26. L. Krivodonova, J. Xin, J.-F. Remacle, N. Chevaugeon and J. E. Flaherty, Shock detection and limiting with discontinuous Galerkin methods for hyperbolic conservation laws, *Appl. Numer. Math.* **48** (2004) 323–338.
27. F. Y. Li and C.-W. Shu, Locally divergence-free discontinuous Galerkin methods for MHD equations, *J. Sci. Comput.* **22** (2005) 413–442.
28. F. Y. Li, L. W. Xu and S. Yakovlev, Central discontinuous Galerkin methods for ideal MHD equations with the exactly divergence-free magnetic field, *J. Comput. Phys.* **230** (2011) 4828–4847.
29. J. M. Martí and E. Müller, Grid-based methods in relativistic hydrodynamics and magnetohydrodynamics, *Living Rev. Comput. Astrophys.* **1** (2015) 3.
30. A. Mignone and G. Bodo, An HLLC Riemann solver for relativistic flows — II. Magnetohydrodynamics, *Mon. Not. Roy. Astron. Soc.* **368** (2006) 1040–1054.
31. T. Miyoshi and K. Kusano, A multi-state HLL approximate Riemann solver for ideal magnetohydrodynamics, *J. Comput. Phys.* **208** (2005) 315–344.

32. W. I. Newman and N. D. Hamlin, Primitive variable determination in conservative relativistic magnetohydrodynamic simulations, *SIAM J. Sci. Comput.* **36** (2014) B661–B683.
33. S. C. Noble, C. F. Gammie, J. C. McKinney and L. D. Zanna, Primitive variable solvers for conservative general relativistic magnetohydrodynamics, *Astrophys. J. Suppl. Ser.* **641** (2006) 626–637.
34. S. Qamar and G. Warnecke, A high-order kinetic flux-splitting method for the relativistic magnetohydrodynamics, *J. Comput. Phys.* **205** (2005) 182–204.
35. J. Qiu and C.-W. Shu, Runge–Kutta discontinuous Galerkin method using WENO limiters, *SIAM J. Sci. Comput.* **26** (2005) 907–929.
36. J. A. Rossmanith, An unstaggered, high-resolution constrained transport method for magnetohydrodynamic flows, *SIAM J. Sci. Comput.* **28** (2006) 1766–1797.
37. G. Tóth, The $\nabla \cdot \mathbf{B} = 0$ constraint in shock-capturing magnetohydrodynamics codes, *J. Comput. Phys.* **161** (2000) 605–652.
38. B. van der Holst, R. Keppens and Z. Meliani, A multidimensional grid-adaptive relativistic magnetofluid code, *Comput. Phys. Comm.* **179** (2008) 617–627.
39. K. L. Wu, Design of provably physical-constraint-preserving methods for general relativistic hydrodynamics, *Phys. Rev. D* **95** (2017) 103001.
40. K. L. Wu and H. Z. Tang, High-order accurate physical-constraints-preserving finite difference WENO schemes for special relativistic hydrodynamics, *J. Comput. Phys.* **298** (2015) 539–564.
41. K. L. Wu and H. Z. Tang, Physical-constraints-preserving central discontinuous Galerkin methods for special relativistic hydrodynamics with a general equation of state, *Astrophys. J. Suppl. Ser.* **228** (2017) 3.
42. 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) 1476–1493.
43. H. Yang and F. Y. Li, Stability analysis and error estimates of an exactly divergence-free method for the magnetic induction equations, *ESAIM*: *Math. Model. Numer. Anal.* **50** (2016) 965–993.
44. O. Zanotti, F. Fambri and M. Dumbser, Solving the relativistic magnetohydrodynamics equations with ADER discontinuous Galerkin methods, *a posteriori* subcell limiting and adaptive mesh refinement, *Mon. Not. Roy. Astron. Soc.* **452** (2015) 3010–3029.
45. X. Zhang and C.-W. Shu, On maximum-principle-satisfying high order schemes for scalar conservation laws, *J. Comput. Phys.* **229** (2010) 3091–3120.
46. 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) 8918–8934.
47. X. Zhang and C.-W. Shu, Maximum-principle-satisfying and positivity-preserving high-order schemes for conservation laws: Survey and new developments, *Proc. Roy. Soc. A* **467** (2011) 2752–2776.
48. J. Zhao and H. Z. Tang, Runge–Kutta discontinuous Galerkin methods for the special relativistic magnetohydrodynamics, *J. Comput. Phys.* **343** (2017) 33–72.