SIAM Review
Vol. 65, No. 4, pp. 1031–1073

Geometric Quasilinearization Framework for Analysis and Design of Bound-Preserving Schemes*

Kailiang Wu[†]
Chi-Wang Shu[‡]

Abstract. Solutions to many partial differential equations satisfy certain bounds or constraints. For example, the density and pressure are positive for equations of fluid dynamics, and in the relativistic case the fluid velocity is upper bounded by the speed of light, etc. As widely realized, it is crucial to develop bound-preserving numerical methods that preserve such intrinsic constraints. Exploring provably bound-preserving schemes has attracted much attention and has been actively studied in recent years. This is, however, still a challenging task for many systems, especially those involving nonlinear constraints. Based on some key insights from geometry, we systematically propose an innovative and general framework, referred to as geometric quasilinearization (GQL), which paves a new effective way for studying bound-preserving problems with nonlinear constraints. The essential idea of GQL is to *equivalently* transform all nonlinear constraints to *linear* ones, by properly introducing some free auxiliary variables. We establish the fundamental principle and general theory of GQL via the geometric properties of convex regions and propose three simple effective methods for constructing GQL. We apply the GQL approach to a variety of partial differential equations and demonstrate its effectiveness and remarkable advantages for studying bound-preserving schemes, using diverse challenging examples and applications which cannot be easily handled by direct or traditional approaches.

Key words. geometric quasilinearization, nonlinear constraints, bound-preserving numerical schemes, time-dependent PDE systems, convex invariant regions, hyperbolic conservation laws

MSC codes. 65M08, 65M60, 65M12, 65M06, 35L65

DOI. 10.1137/21M1458247

Contents

*Received by the editors November 9, 2021; accepted for publication (in revised form) August 16, 2022; published electronically November 7, 2023.
https://doi.org/10.1137/21M1458247

Funding: The work of the first author was supported by NSFC grant 12171227. The work of the second author was supported by NSF grant DMS-2010107 and AFOSR grant FA9550-20-1-0055.

[†]Department of Mathematics, SUSTech International Center for Mathematics, Southern University of Science and Technology, and National Center for Applied Mathematics Shenzhen (NCAMS), Shenzhen, Guangdong 518055, China (wukl@sustech.edu.cn).

[‡]Division of Applied Mathematics, Brown University, Providence, RI 02912 USA (chi-wang_shu@brown.edu).

1. Introduction.

Solutions to many partial differential equations (PDEs) satisfy certain algebraic constraints that are usually derived from some (physical) bound principles, for example, the positivity of density and pressure. Consider such time-dependent PDE systems in the general form

$$\partial_t \mathbf{u} + \boldsymbol{\mathcal{L}}(\mathbf{u}) = \mathbf{0}, \qquad \mathbf{u}(\boldsymbol{x}, 0) = \mathbf{u}_0(\boldsymbol{x}), \tag{1.1}$$

where $\boldsymbol{\mathcal{L}}$ denotes the differential operator associated with the spatial coordinates $\boldsymbol{x}$, and suppose the system (1.1) is defined in a bounded domain with suitable boundary conditions. Assume that the algebraic constraints (bound principles) can be expressed by either the positivity or the nonnegativity of several (linear or nonlinear) functions of $\mathbf{u}$ as

$$g_i(\mathbf{u}) > 0 \quad \forall i \in \mathbb{I}, \qquad g_i(\mathbf{u}) \geq 0 \quad \forall i \in \widehat{\mathbb{I}}, \tag{1.2}$$

where $\mathbb{I} \cup \widehat{\mathbb{I}} = \{1, \dots, I\}$ with the positive integer I denoting the total number of the constraints. In other words, the evolved variables $\mathbf{u} = (u_1, \dots, u_N)^\top$ are constrained in the admissible state set:

$$G = \left\{ \mathbf{u} \in \mathbb{R}^N : \; g_i(\mathbf{u}) > 0 \;\; \forall i \in \mathbb{I}, \quad g_i(\mathbf{u}) \geq 0 \;\; \forall i \in \widehat{\mathbb{I}} \right\}. \tag{1.3}$$

Throughout this paper, we assume that G is a *convex* set with nonempty interior, while the functions $\{g_i(\mathbf{u})\}_{i=1}^{I}$ are *not* necessarily concave (and are *not* required to be concave in this paper). Moreover, we suppose that G is an *invariant region* for the exact solution of system (1.1):

- If $\mathbf{u}(\boldsymbol{x}, 0) \in G$ for all $\boldsymbol{x}$, then $\mathbf{u}(\boldsymbol{x}, t) \in G$ for all $\boldsymbol{x}$ and $t > 0$.

The PDE models with such convex invariant regions appear widely in many science and engineering problems from a variety of fields such as fluid mechanics [86, 109, 119], materials science [13, 73], astrophysics [103, 108], hydraulic engineering [113], chemistry [29, 25], biology [19, 92], etc. Typical examples of such PDEs (1.1) include the semilinear parabolic equations with maximum principles [28, 2] and the reaction-diffusion and convection-diffusion systems [22, 29, 95]. Another important class of such systems is that of the hyperbolic conservation laws $\partial_t \mathbf{u} + \nabla \cdot \mathbf{f}(\mathbf{u}) = \mathbf{0}$ and other related hyperbolic systems; see [48, 90, 121, 122, 44, 94] and a few examples listed in section 2.

A basic goal in the design of numerical methods for (1.1) is that the numerical solutions can inherit as far as possible the intrinsic properties of system (1.1). The constraints (1.2) and the associated invariant region G carry important meanings or properties of the system. It is natural to explore *bound-preserving* schemes that keep the numerical solutions within the region G:

- If $\mathbf{u}_h(\cdot, t_0) \in G$, then $\mathbf{u}_h(\cdot, t_n) \in G$ for all $n \in \mathbb{N}$,

where $\mathbf{u}_h(\cdot, t_n)$ denotes the numerical solutions at nth time level. In fact, preserving such constraints is not only necessary for physical considerations, but also very essential for numerical stability and theoretical (e.g., convergence [1]) analysis. If the intrinsic physical constraints (1.2) are violated numerically, often the PDE system (1.1) and its discrete equations may become ill-posed. For example, when negative density and/or pressure are produced when numerically solving the compressible Euler equations, the fundamental hyperbolicity of the system is lost. As such, failure to preserve such physically relevant constraints may cause serious numerical problems such as nonlinear instability, nonphysical solutions or phenomena, blowups of the code, etc. Therefore, it is significant and highly desirable to develop bound-preserving schemes.

In the past few decades, the exploration of bound-preserving numerical methods has attracted much attention and has been actively studied. The extensive research in this direction includes, for example, maximum-principle-preserving schemes [23, 9, 32, 121, 116, 24, 33, 28], positivity-preserving schemes [86, 54, 60, 82, 122, 36, 13, 11, 102, 16], invariant-region-preserving schemes [58, 29, 44, 42, 56, 57], and other bound-preserving type schemes [109, 111, 103, 47, 78, 67, 12, 69, 15] based on various kinds of discretizations. The developments of bound-preserving schemes have also covered a wide variety of differential equations in various fields, for example, ordinary differential equations (ODEs) [98, 53, 97, 31, 83], differential-algebraic equations (DAEs) [10], and various time-dependent PDEs in diverse applications including but not limited to conservation laws [121, 116], hydrodynamic systems [63, 122, 50, 115, 103, 106, 41, 70], phase-field models [91, 13, 27, 73, 34], reaction-diffusion equations [29, 55, 36], convection-diffusion equations [4, 67], radiative transfer models [118], Boltzmann transport equations [49, 17], Vlasov–Poisson equations [89], Fokker–Planck equations [76], the Kerr–Debye model [52], Keller–Segel equations [19, 71, 77, 92, 51], and Poisson–Nernst–Planck equations [93, 75], to name a few. In particular, the discrete maximum principle of numerical schemes has been addressed for a class of linear or nonlinear scalar parabolic equations; see, for example, [32, 33, 59, 28] and the references therein. In [29], Estep, Larson, and Williams systematically studied the posteriori error estimation and the preservation of invariant regions for numerical solutions of a general class of reaction-diffusion equations with adaptive error control; see also [30] for an application to the analysis of shear layers. Recently, Lagrange multiplier approaches were proposed for constructing positivity/bound-preserving schemes for parabolic-type equations in [102, 15, 16]. There are also many other bound-preserving techniques for parabolic or dissipative equations, such as cut-off [65], convex splitting [13, 77], and reformulation [51], to name a few. While these bound-preserving approaches have been proven effective for parabolic problems, they are not quite suitable for hyperbolic type systems as they may not retain other important properties such as conservation and the nonoscillatory shock-capturing property.

In [121, 122], Zhang and Shu established a framework for constructing bound-preserving high-order finite volume and discontinuous Galerkin schemes for hyperbolic conservation laws. The loss of the bound-preserving property may be classified into two cases: one case is that the point values of the numerical solutions may go outside the bounds, and the other is that the cell averages evolved to the next time step become out of bounds during the updating process. A key step in the Zhang–Shu framework [121, 122] is to look for the high-order schemes that have a provable "weak" bound-preserving property keeping the updated cell averages in the region G. Achieving such a weak bound-preserving property is essential, but its proof needs subtle analysis. Once this property is proven, a simple scaling limiter can be used to enforce the pointwise bounds for the numerical solutions [121, 122, 125]. Zhang and Shu's approach has been applied to many other hyperbolic systems [113, 123, 17, 18, 14, 87, 118, 119, 111, 112] as well as the convection-diffusion or reaction equations [120, 71, 25]. Another bound-preserving approach [116, 50, 72] for hyperbolic systems is to use flux-correction limiters, which modify high-order numerical fluxes to enforce the bounds by using a prepared provably bound-preserving (lower-order) scheme as the building block; see also [20, 21, 114, 109, 115]. Continuous finite element approximations with convex limiting were developed in [44, 45, 42, 46] to preserve invariant regions for hyperbolic equations. Some thorough reviews on bound-preserving type schemes can be found in [117, 94, 28]. While our work may be extended to other systems, in the remainder of this paper we will mainly focus on hyperbolic type equations.

Due to the lack of a general theory, rigorously analyzing or proving the bound-preserving property of a numerical scheme remains a challenging task. Take the hyperbolic systems, for example. Despite the success of the limiter-based frameworks (cf. [121, 122, 116, 50]) in constructing high-order bound-preserving schemes, the validity of those limiters is actually based on some (weak) bound-preserving properties of the cell-average schemes and/or of the lower-order numerical fluxes as the key building blocks. Rigorously proving such properties is essential but often very difficult [94, 104, 108]. To illustrate the challenges in rigorous bound-preserving analysis, we suppose that a numerical scheme for (1.1) may be written as

$$\mathbf{u}_j^{n+1} = \boldsymbol{\mathcal{E}}_h(\mathbf{u}_{j-k}^n, \mathbf{u}_{j-k+1}^n, \ldots, \mathbf{u}_j^n, \ldots, \mathbf{u}_{j+s-1}^n, \mathbf{u}_{j+s}^n), \tag{1.4}$$

where $\boldsymbol{\mathcal{E}}_h$ is the discretization operator, the superscripts on $\mathbf{u}$ denote the time levels, and the subscripts on $\mathbf{u}$ indicate the indices of the spatial grid or nodal points.

The bound-preserving problem for the scheme (1.4) can boil down to answering the question of

$$\text{whether} \quad \mathbf{u}_j^n \in G \ \ \forall j \quad \text{implies} \quad \mathbf{u}_j^{n+1} \in G \ \ \forall j.$$

In essence, the problem is to explore whether or not the range of the high-dimensional function $\boldsymbol{\mathcal{E}}_h$ is always contained in G: $\boldsymbol{\mathcal{E}}_h(G^{s+k+1}) \subseteq G$. For some scalar PDEs with linear constraints, e.g., when the bounds are linearly defined by maximum principles, a general approach for analyzing and designing bound-preserving schemes is to exploit certain monotonicity in schemes; see, e.g., [121, 26, 68]. Yet, for PDE systems with nonlinear constraints, there is no universal tool like monotonicity; as such, direct and complicated algebraic verifications are usually performed in a case-by-case manner for different schemes and different PDEs; see, e.g., [122, 84, 109, 87, 119, 79, 106]. Therefore, the design and analysis of bound-preserving schemes involving nonlinear constraints are highly nontrivial, even for first-order schemes; cf. [101, 6, 39, 100, 85, 61, 74, 80, 110, 104].

Nonlinear constraints exist in many physical PDE systems; see several representative examples in section 2. For instance, the physical constraints for solutions of the special relativistic magnetohydrodynamic (MHD) equations (2.14) include the positivity of density D and thermal pressure p, and the upper bounding of fluid velocity field $\boldsymbol{v}$ by the speed of light c, namely,

$$D > 0, \qquad p(\mathbf{u}) > 0, \qquad c - \|\boldsymbol{v}(\mathbf{u})\| > 0, \tag{1.5}$$

where the evolved variables $\mathbf{u} = (D, \boldsymbol{m}, \mathbf{B}, E)^\top$ with the momentum vector $\boldsymbol{m} \in \mathbb{R}^3$, the magnetic field $\mathbf{B} \in \mathbb{R}^3$, and the total energy E; see Example 2.7 and [110, 108] for more details. *The second and third constraints in* (1.5) *are highly nonlinear with respect to* $\mathbf{u}$*, because* $p(\mathbf{u})$ *and* $\boldsymbol{v}(\mathbf{u})$ *cannot be explicitly formulated in terms of* $\mathbf{u}$. These implicit functions $p(\mathbf{u})$ and $\boldsymbol{v}(\mathbf{u})$ are often expressed via another implicit function $\hat{\phi}(\mathbf{u})$ as

$$p(\mathbf{u}) = \frac{\Gamma - 1}{\Gamma \Upsilon_{\mathbf{u}}^2(\hat{\phi})}\Big(\hat{\phi} - D\Upsilon_{\mathbf{u}}(\hat{\phi})\Big), \quad \boldsymbol{v}(\mathbf{u}) = \Big(\boldsymbol{m} + (\boldsymbol{m}\cdot\mathbf{B})\mathbf{B}/\hat{\phi}\Big)/(\hat{\phi} + |\mathbf{B}|^2), \tag{1.6}$$

where $\hat{\phi} = \hat{\phi}(\mathbf{u})$ is implicitly defined by the positive root of the nonlinear function $F(\phi;\mathbf{u}) := \phi - E + \|\mathbf{B}\|^2 - \frac{1}{2}\left(\frac{(\boldsymbol{m}\cdot\mathbf{B})^2}{\phi^2} + \frac{\|\mathbf{B}\|^2}{\Upsilon_{\mathbf{u}}^2(\phi)}\right) + \frac{\Gamma-1}{\Gamma}\left(\frac{D}{\Upsilon_{\mathbf{u}}(\phi)} - \frac{\phi}{\Upsilon_{\mathbf{u}}^2(\phi)}\right)$, the constant Γ is the ratio of specific heats, and $\Upsilon_{\mathbf{u}}(\phi) := \left(\frac{\phi^2(\phi+\|\mathbf{B}\|^2)^2 - \left[\phi^2\|\boldsymbol{m}\|^2 + (2\phi+\|\mathbf{B}\|^2)(\boldsymbol{m}\cdot\mathbf{B})^2\right]}{\phi^2(\phi+\|\mathbf{B}\|^2)^2}\right)^{-\frac{1}{2}}$. In traditional approaches to studying the preservation of bounds (1.5), we typically substitute the numerical solution, evolved by a scheme like (1.4), into the functions $p(\mathbf{u})$ and $\boldsymbol{v}(\mathbf{u})$, then evaluate these *implicit* functions and try to verify the nonlinear constraints (1.5). The whole calculation and analysis can be too complicated, if not impossible, to enable us to reach any conclusion.

In this paper we discover that, by introducing some extra auxiliary variables *independent* of the system variables $\mathbf{u}$, nonlinear constraints can be *equivalently* represented by a set of *linear* constraints, given the region G is convex. For example, the nonlinear constraint

$$g(\mathbf{u}) = u_2 - u_1^2 > 0 \tag{1.7}$$

is exactly equivalent to[1]

$$\varphi(\mathbf{u};\theta_*) := u_2 - 2u_1\theta_* + \theta_*^2 > 0 \qquad \forall \theta_* \in \mathbb{R}, \tag{1.8}$$

[1] The equivalence of (1.7) and (1.8) can be easily proven using $\min_{\theta_*\in\mathbb{R}} \varphi(\mathbf{u};\theta_*) = g(\mathbf{u})$.

where the extra parameter θ_* is *independent* of $\mathbf{u}$ and called the *free auxiliary variable* in this paper. Clearly, the new constraints (1.8) are linear[2] with respect to $\mathbf{u}$. As we will show, an equivalent linear representation such as (1.8) can be found for general nonlinear constraints, even for implicitly formulated nonlinear constraints like (1.5). For instance, as will be shown in Theorem 6.6, the highly nonlinear implicit constraints (1.5) can be equivalently represented as

$$(1.9) \qquad D > 0, \qquad \mathbf{u} \cdot \mathbf{n}_* + p_m^* > 0 \quad \forall \mathbf{B}_* \in \mathbb{R}^3, \quad \forall \boldsymbol{v}_* \in \mathbb{B}_1(\mathbf{0}),$$

where $\{\mathbf{B}_*, \boldsymbol{v}_*\}$ are the free auxiliary variables; the vector $\mathbf{n}_*$ and scalar p_m^* are functions of $\{\mathbf{B}_*, \boldsymbol{v}_*\}$ defined by (6.6)–(6.7); and $\mathbb{B}_1(\mathbf{0}) := \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\| < 1\}$. Note that the equivalent constraints (1.9) are all linear with respect to $\mathbf{u}$. Owing to such linearity, this novel equivalent form (1.9) has significant advantages over the original nonlinear form (1.5) in designing and analytically analyzing bound-preserving schemes [110, 108]. Several important questions naturally arise: Are there any intrinsic mechanisms behind such an equivalent linear representation? What is the condition for its existence? In general, how do we find or construct it?

The aim of this article is to establish a universal framework, termed as geometric quasilinearization (GQL), for constructing equivalent linear representations for general nonlinear constraints. It will be based on some geometric insights about recharacterizing a convex region G in an alternative way. The GQL framework will shed new light on the challenging bound-preserving problems involving nonlinear constraints. The novelty and significance of the proposed GQL framework include the following:

- A distinctive innovation of GQL lies in a novel geometric point of view on the nonlinear algebraic constraints and the convex invariant region G. The GQL framework reveals that *nonlinear constraints are actually "linear" if viewed in higher dimensions*, as long as all the constraints form a convex region.
- Through introducing some extra free auxiliary variables, this framework provides a simple yet universal approach to deriving the equivalent linear representation (termed the GQL representation) for a general convex region G.
- GQL offers a highly effective tool for analyzing and designing bound-preserving schemes for problems with nonlinear constraints.
- The GQL representations have concise formulations and can be constructed easily. We will propose three effective methods for constructing GQL.

The idea of GQL is motivated by a series of our recent works on seeking bound-preserving schemes for the (single-component) compressible MHD systems [110, 104, 106, 107, 108]. For the invariant region of the ideal nonrelativistic MHD equations, its equivalent linear representation was first established by technical algebraic manipulations [104]. Such a representation played a critical role in obtaining the first rigorous positivity-preserving analysis of numerical schemes for the ideal MHD system [104], and also in designing the provably positivity-preserving multidimensional MHD schemes [106, 107]. The success of the GQL idea in these special cases strongly encourages us to explore its essential mechanisms and universal framework for general systems.

Our efforts in this article include the following:

- We interpret, from a geometric viewpoint, the fundamental principle behind the GQL representations for general nonlinear algebraic constraints.
- We establish the universal GQL framework and its mathematical theory.

[2] This paper broadly uses the word "linear," which means "affine" for functions or constraints with respect to $\mathbf{u}$.

- We propose three simple, effective methods for constructing GQL representations using extra free auxiliary variables in exchange for linearity. As examples, the GQL representations are derived for the invariant regions of various physical systems.
- We illustrate the GQL methodology and related techniques for nonlinear bound-preserving analysis and design, and we demonstrate its effectiveness and remarkable advantages using diverse examples that cannot be easily handled by direct or traditional approaches.

We emphasize that the GQL framework is applicable to general convex invariant regions with nonlinear constraints and is not restricted to the specific forms of the equations (1.1).

The paper is organized as follows. Section 2 presents several examples of physical PDE systems with convex invariant regions and nonlinear constraints. Section 3 explores the fundamental principle and the general theory for the GQL framework. We propose in section 4 three simple, effective methods for constructing GQL representations. Sections 5 and 6 illustrate the GQL approach and its advantages for bound-preserving study using diverse examples covering different schemes of various PDE systems in one and two dimensions. In section 7 we apply the GQL approach to design bound-preserving schemes for the multicomponent MHD system and further demonstrate its powerful capabilities in addressing challenging bound-preserving problems that could not be handled by direct or traditional approaches. Several experimental results are given in section 8 to verify the performance of the bound-preserving schemes developed via GQL. The conclusions follow in section 9. Throughout this paper, we will use $\mathrm{cl}(G)$, $\mathrm{int}(G)$, and ∂G to denote the closure, the interior, and the boundary of a region G, respectively. We employ $\|\mathbf{a}\|$ to denote the 2-norm of vector $\mathbf{a}$. We use $\mathbf{a}\cdot\mathbf{b}$ to denote the inner product of two vectors $\mathbf{a}$ and $\mathbf{b}$, and $\mathbf{a}\otimes\mathbf{b}$ to denote the outer product, i.e., in index notation, $(\mathbf{a}\otimes\mathbf{b})_{ij} = a_i b_j$.

2. Examples of PDE Systems with Nonlinear Constraints.

This section presents several examples of PDEs (1.1) with nonlinear constraints and convex invariant regions. For convenience, the ideal equation of state $p = (\Gamma - 1)\rho e$ is used to close the systems in Examples 2.1, 2.2, 2.4, 2.6, and 2.7, with p denoting the thermal pressure, ρ the (rest-mass) density, e the specific internal energy, and the constant $\Gamma > 1$ denoting the ratio of specific heats. For the relativistic models in Examples 2.3, 2.4, and 2.7, normalized units are employed such that the speed of light $c = 1$.

Example 2.1 (Euler system). Consider the one-dimensional (1D) compressible Euler equations

$$\partial_t \mathbf{u} + \partial_x \mathbf{f}(\mathbf{u}) = \mathbf{0}, \qquad \mathbf{u} = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix}, \qquad \mathbf{f}(\mathbf{u}) = \begin{pmatrix} m \\ mv + p \\ (E+p)v \end{pmatrix}, \tag{2.1}$$

where ρ, m, $v = m/\rho$, and p denote the fluid density, momentum, velocity, and pressure, respectively. The quantity $E = \rho e + \frac{1}{2}\rho v^2$ is the total energy, with e being the specific internal energy. For this system, the density ρ and the internal energy ρe are positive, namely, $\mathbf{u}$ should stay in the region

$$G = \left\{ \mathbf{u} = (\rho, m, E)^\top \in \mathbb{R}^3 : \ \rho > 0, \ g(\mathbf{u}) := E - \frac{m^2}{2\rho} > 0 \right\}, \tag{2.2}$$

which is a convex invariant region [122] of the system (2.1). If we further consider Tadmor's minimum entropy principle [99], $S(\mathbf{u}) \geq S_{min} := \min_{\boldsymbol{x}} S(\mathbf{u}_0(\boldsymbol{x}))$, for the specific entropy $S = p\rho^{-\Gamma}$, then we obtain another convex invariant region [124]:

$$\widetilde{G} = \left\{ \mathbf{u} = (\rho, m, E)^\top \in \mathbb{R}^3 : \ \rho > 0, \ \widetilde{g}(\mathbf{u}) \geq 0 \right\} \tag{2.3}$$

with $\widetilde{g}(\mathbf{u}) := S(\mathbf{u}) - S_{min} = \frac{\Gamma-1}{\rho^\Gamma}\left(E - \frac{m^2}{2\rho}\right) - S_{min}$.

Example 2.2 (Navier–Stokes system). Consider the 1D dimensionless compressible Navier–Stokes equations (see, for example, [119])

$$\partial_t \mathbf{u} + \partial_x \mathbf{f}(\mathbf{u}) = \frac{\eta}{\mathsf{Re}} \partial_{xx} \mathbf{r}(\mathbf{u}), \qquad \mathbf{r}(\mathbf{u}) = \begin{pmatrix} 0 \\ v \\ \frac{v^2}{2} + \frac{\Gamma}{\mathtt{Pr}\ \eta} e \end{pmatrix}, \tag{2.4}$$

where $\{\eta, \mathtt{Re}, \mathtt{Pr}\}$ are positive constants, and the definitions of $\mathbf{u}$ and $\mathbf{f}(\mathbf{u})$ are the same as in Example 2.1. Both sets in (2.2) and (2.3) are also invariant regions for system (2.4).

Example 2.3 (M1 model of radiative transfer). For the solutions of the gray M1 moment system of radiative transfer (see, for example, [84]), a convex invariant region is

$$G = \left\{ \mathbf{u} = (E_r, \boldsymbol{\mathcal{F}}_r)^\top \in \mathbb{R}^4 : \ g(\mathbf{u}) := E_r - \|\boldsymbol{\mathcal{F}}_r\| \geq 0 \right\}, \tag{2.5}$$

where E_r is the radiation energy, and $\boldsymbol{\mathcal{F}}_r$ is the radiation energy flux.

Example 2.4 (relativistic hydrodynamic system). Consider the 1D governing equations of special relativistic hydrodynamics (RHD) system [109, 87]:

$$\partial_t \mathbf{u} + \partial_x \mathbf{f}(\mathbf{u}) = \mathbf{0}, \qquad \mathbf{u} = \begin{pmatrix} D \\ m \\ E \end{pmatrix}, \qquad \mathbf{f}(\mathbf{u}) = \begin{pmatrix} Dv \\ mv + p \\ m \end{pmatrix}, \tag{2.6}$$

with density $D = \rho\gamma$, momentum $m = \rho h \gamma^2 v$, and energy $E = \rho h \gamma^2 - p$. Here, ρ, v, p, and $\gamma = (1 - v^2)^{-\frac{1}{2}}$ denote the rest-mass density, velocity, pressure, and Lorentz factor, respectively. The quantity $h = 1 + e + p/\rho$ represents the specific enthalpy, with e being the specific internal energy. For this system, the density and the pressure are positive, and the magnitude of v must be smaller than the speed of light ($c = 1$). These physical constraints define the invariant region $G = \left\{ \mathbf{u} \in \mathbb{R}^3 : D > 0, \ p(\mathbf{U}) > 0, \ 1 - |v(\mathbf{U})| > 0 \right\}$. It was proven in [109] that this region G is convex and can be equivalently represented as

$$G = \left\{ \mathbf{u} \in \mathbb{R}^3 : \ D > 0, \ g(\mathbf{u}) := E - \sqrt{D^2 + m^2} > 0 \right\}. \tag{2.7}$$

As shown in [105], the minimum entropy principle $S(\mathbf{u}) \geq S_{min}$ also holds for the RHD system (2.6), yielding another invariant region

$$\widetilde{G} = \left\{ \mathbf{u} \in \mathbb{R}^3 : \ D > 0, \ g(\mathbf{u}) > 0, \ \widetilde{g}(\mathbf{u}) \geq 0 \right\}, \tag{2.8}$$

where $\widetilde{g}(\mathbf{u}) := p(\mathbf{u})(\rho(\mathbf{u}))^{-\Gamma} - S_{min}$ is a highly nonlinear implicit function. In the RHD case, the functions $p(\mathbf{u})$ and $\rho(\mathbf{u})$ cannot be explicitly expressed in terms of $\mathbf{u}$. Specifically, $p(\mathbf{u})$ is implicitly defined by the positive root of the nonlinear function $F(p; \mathbf{u}) := \frac{m^2}{E+p} + D\left(1 - \frac{m^2}{(E+p)^2}\right)^{\frac{1}{2}} + \frac{p}{\Gamma-1} - E$, and then $\rho(\mathbf{u}) = D\sqrt{1 - m^2/(E + p(\mathbf{u}))^2}$.

Example 2.5 (ten-moment Gaussian closure system). In 2D, this system [79, 80] reads

$$\partial_t \mathbf{u} + \partial_x \mathbf{f}_1(\mathbf{u}) + \partial_y \mathbf{f}_2(\mathbf{u}) = \mathbf{0}, \tag{2.9}$$
$$\mathbf{u} = \begin{pmatrix} \rho \\ m_1 \\ m_2 \\ E_{11} \\ E_{12} \\ E_{22} \end{pmatrix}, \qquad \mathbf{f}_j(\mathbf{u}) = \begin{pmatrix} m_j \\ m_1 v_j + p_{1j} \\ m_2 v_j + p_{2j} \\ E_{11} v_j + p_{1j} v_1 \\ E_{12} v_j + \frac{1}{2}(p_{1j} v_2 + p_{2j} v_1) \\ E_{22} v_j + p_{2j} v_2 \end{pmatrix}, \quad j = 1, 2.$$

Here, ρ, $\boldsymbol{m} = (m_1, m_2)$, $\boldsymbol{v} = \boldsymbol{m}/\rho$, $\mathbf{E} = (E_{ij})_{1\le i,j\le 2}$, and $\mathbf{p} = (p_{ij})_{1\le i,j\le 2}$ are, respectively, the density, momentum vector, velocity, symmetric energy tensor, and symmetric anisotropic pressure tensor. The system (2.9) is closed by $\mathbf{p} = 2\mathbf{E} - \rho \boldsymbol{v} \otimes \boldsymbol{v}$. The density ρ is positive and the pressure tensor $\mathbf{p}$ is positive-definite, namely, the evolved variables $\mathbf{u}$ should stay in the region

$$G = \left\{ \mathbf{u} \in \mathbb{R}^6 : \ \rho > 0, \ \mathbf{E} - \frac{\boldsymbol{m} \otimes \boldsymbol{m}}{2\rho} \text{ is positive-definite} \right\} \tag{2.10}$$
$$= \left\{ \mathbf{u} \in \mathbb{R}^6 : \ \rho > 0, \ \boldsymbol{z}^\top \left(\mathbf{E} - \frac{\boldsymbol{m} \otimes \boldsymbol{m}}{2\rho} \right) \boldsymbol{z} > 0 \ \ \forall \boldsymbol{z} \in \mathbb{R}^2 \setminus \{\mathbf{0}\} \right\}. \tag{2.11}$$

Example 2.6 (ideal MHD system). This system [104, 106] can be written as

$$\partial_t \begin{pmatrix} \rho \\ \boldsymbol{m} \\ \mathbf{B} \\ E \end{pmatrix} + \nabla \cdot \begin{pmatrix} \boldsymbol{m} \\ \boldsymbol{m} \otimes \boldsymbol{v} - \mathbf{B} \otimes \mathbf{B} + p_{tot} \mathbf{I} \\ \boldsymbol{v} \otimes \mathbf{B} - \mathbf{B} \otimes \boldsymbol{v} \\ (E + p_{tot})\, \boldsymbol{v} - (\boldsymbol{v} \cdot \mathbf{B}) \mathbf{B} \end{pmatrix} = \mathbf{0} \tag{2.12}$$

with ρ being the density, $\boldsymbol{m}$ the momentum vector, $\boldsymbol{v} = \boldsymbol{m}/\rho$ the velocity, $E = \rho e + \frac{1}{2}(\rho \|\boldsymbol{v}\|^2 + \|\mathbf{B}\|^2)$ the total energy, $p_{tot} = p + \frac{1}{2}\|\mathbf{B}\|^2$ the total pressure, p the thermal pressure, and $\mathbf{B}$ the magnetic field that satisfies the extra divergence-free condition $\nabla \cdot \mathbf{B} = 0$. For this system, the density ρ and the internal energy ρe are positive, namely, $\mathbf{u}$ should stay in the invariant region

$$G = \left\{ \mathbf{u} = (\rho, \boldsymbol{m}, \mathbf{B}, E)^\top \in \mathbb{R}^8 : \ \rho > 0, \ g(\mathbf{u}) := E - \frac{\|\boldsymbol{m}\|^2}{2\rho} - \frac{\|\mathbf{B}\|^2}{2} > 0 \right\}. \tag{2.13}$$

Example 2.7 (relativistic MHD system). This system [110, 108] takes the form of

$$\partial_t \begin{pmatrix} D \\ \boldsymbol{m} \\ \mathbf{B} \\ E \end{pmatrix} + \nabla \cdot \begin{pmatrix} D \boldsymbol{v} \\ \boldsymbol{m} \otimes \boldsymbol{v} - \mathbf{B} \otimes \left(\gamma^{-2} \mathbf{B} + (\boldsymbol{v} \cdot \mathbf{B}) \boldsymbol{v} \right) + p_{tot} \mathbf{I} \\ \boldsymbol{v} \otimes \mathbf{B} - \mathbf{B} \otimes \boldsymbol{v} \\ \boldsymbol{m} \end{pmatrix} = \mathbf{0} \tag{2.14}$$

with mass density $D = \rho\gamma$, momentum vector $\boldsymbol{m} = (\rho h \gamma^2 + \|\mathbf{B}\|^2)\boldsymbol{v} - (\boldsymbol{v} \cdot \mathbf{B})\mathbf{B}$, and energy $E = \rho h \gamma^2 - p_{tot} + \|\mathbf{B}\|^2$, and the magnetic field $\mathbf{B}$ satisfies $\nabla \cdot \mathbf{B} = 0$ as the ideal MHD case. The total pressure p_{tot} consists of the magnetic pressure $p_m := \frac{1}{2}\left(\gamma^{-2}\|\mathbf{B}\|^2 + (\boldsymbol{v} \cdot \mathbf{B})^2 \right)$ and the thermal pressure p. Analogously to Example 2.4, the quantities ρ, $\boldsymbol{v}$, h, and $\gamma = (1 - \|\boldsymbol{v}\|^2)^{-\frac{1}{2}}$ are, respectively, the rest-mass density, velocity, specific enthalpy, and Lorentz factor. The positivity of density and pressure as well as the subluminal constraint $\|\boldsymbol{v}\| < c = 1$ constitute the invariant region

$$G = \left\{ \mathbf{u} = (D, \boldsymbol{m}, \mathbf{B}, E)^\top \in \mathbb{R}^8 : D > 0,\ p(\mathbf{u}) > 0,\ 1 - \|\boldsymbol{v}(\mathbf{u})\| > 0 \right\}, \tag{2.15}$$

where $p(\mathbf{u})$ and $\boldsymbol{v}(\mathbf{u})$ are highly nonlinear and cannot be explicitly formulated, as discussed in (1.6).

3. Theoretical Framework of Geometric Quasilinearization.

This section establishes the universal GQL framework, with geometric insights into understanding the fundamental principle behind the GQL representations.

Let $G \subset \mathbb{R}^N$ be an invariant region or admissible state set of a physical system. Assume that G can be formulated in the general form (1.3). For notational convenience, we represent G as

$$G = \left\{ \mathbf{u} \in \mathbb{R}^N :\ g_i(\mathbf{u}) \succ 0,\ 1 \le i \le I \right\}, \tag{3.1}$$

where the symbol "$\succ$" denotes "$>$" if $i \in \mathbb{I}$ or "$\ge$" if $i \in \widehat{\mathbb{I}}$. Let $G_L = \{\mathbf{u} \in \mathbb{R}^N : g_i(\mathbf{u}) \succ 0\ \forall i \in \mathbb{I}_L\}$ be the region formed by all the linear constraints in G, i.e., the function g_i is linear for $i \in \mathbb{I}_L$. If $\mathbb{I}_L = \emptyset$, then we define $G_L = \mathbb{R}^N$.

We consider the nontrivial case that at least one of the functions $\{g_i(\mathbf{u})\}$ is nonlinear, namely, $G \subset G_L$ and $G \neq G_L$. The goal of our GQL methodology is to use some extra free auxiliary variables in exchange for linearity and, more precisely, to equivalently represent G by using only *linear* constraints with the help of free auxiliary variables.

Definition 3.1. *We define a set G_* as an equivalent linear representation (termed the GQL representation) of the region G if $G_* = G$ and G_* takes the form*

$$G_* = \left\{ \mathbf{u} \in \mathbb{R}^N : \varphi_i(\mathbf{u}; \boldsymbol{\theta}_{i*}) \succ 0 \quad \forall \boldsymbol{\theta}_{i*} \in \Theta_i,\ 1 \le i \le I \right\}, \tag{3.2}$$

where the functions $\{\varphi_i\}$ are all linear *(affine) with respect to $\mathbf{u}$; the parameters $\boldsymbol{\theta}_{i*}$ are independent of $\mathbf{u}$ and represent the (possible) extra free auxiliary variables with Θ_i denoting their ranges.*

Based on Definition 3.1, we immediately have the following theorem.

Theorem 3.2. *Assume that a set G_* is of the form* (3.2) *with φ_i being linear with respect to $\mathbf{u}$ and satisfying*

$$\min_{\boldsymbol{\theta}_{i*} \in \Theta_i} \varphi_i(\mathbf{u}; \boldsymbol{\theta}_{i*}) = \lambda_i(\mathbf{u}) g_i(\mathbf{u}) \tag{3.3}$$

with $\lambda_i(\mathbf{u}) > 0$ for all $\mathbf{u} \in G_L$. Then $G_* = G$ and G_* is the GQL representation of G.

Remark 3.3. For $i \in \mathbb{I}_L$, the function $g_i(\mathbf{u})$ is already linear, and thus we can simply take $\varphi_i(\mathbf{u}; \boldsymbol{\theta}_{i*}) = g_i(\mathbf{u})$ without the free auxiliary variable $\boldsymbol{\theta}_{i*}$. That is, all the linear constraints remain unchanged in the GQL representation.

Theorem 3.2 indicates a way to seek the GQL representation, namely, by constructing linear functions $\{\varphi_i\}$ such that (3.3) holds. We have used this approach in [104] to establish the GQL representation of the invariant region (2.13) for the ideal MHD equations. However, this constructive approach needs some empirical observations or trial-and-error procedures, as Theorem 3.2 does not provide any intuition into how to find the qualified $\{\varphi_i\}$. In the following, we explore a simpler yet universal approach from the geometric point of view.

Given that $\{\varphi_i\}$ in (3.2) are all linear with respect to $\mathbf{u}$, the set G_* is always convex. This means that if the region G has GQL representation (3.2), then G must also be convex. Hence, we should make the following basic (minimal) assumption.

Assumption 3.4. *The invariant region G is convex, and* $\operatorname{int}(G) \neq \emptyset$.

This basic assumption is valid for many physical systems including all those introduced in section 2. Again, we emphasize that the functions $\{g_i(\mathbf{u})\}$ are *not* necessarily concave.

3.1. A Heuristic Example.

Before deriving the general theory, let us look at an example to gain some intuition which inspired us to develop the GQL framework.

Example 3.5. Consider the simple example mentioned in (1.7)–(1.8), i.e., $G = \{\mathbf{u} = (u_1, u_2)^\top \in \mathbb{R}^2 : \; g(\mathbf{u}) = u_2 - u_1^2 > 0\}$. According to Theorem 3.2, the GQL representation of G is

$$G_* = \left\{\mathbf{u} = (u_1, u_2)^\top \in \mathbb{R}^2 : \; \varphi(\mathbf{u}; \theta_*) = u_2 - 2u_1\theta_* + \theta_*^2 > 0 \quad \forall \theta_* \in \mathbb{R}\right\}. \tag{3.4}$$

In this way, we gain linearity by introducing the extra free auxiliary variable θ_*. To understand the intrinsic mechanisms, we draw the graph of the region G and its boundary curve $\partial G = \{\mathbf{u} : g(\mathbf{u}) = 0\}$ on the u_1-u_2 plane in Figure 1. We also plot the graphs of $\{\mathbf{u} : \varphi(\mathbf{u}; \theta_*) = 0\}$ for several special values of $\theta_* \in \{\pm 2, \pm 1, 0\}$ in the left subfigure of Figure 1. It is observed that all the lines $\{\mathbf{u} : \varphi(\mathbf{u}; \theta_*) = 0\}$ are tangent to the parabolic curve ∂G, which exactly forms an envelope containing G.

Let $\mathbf{u}_* = (\theta_*, \theta_*^2)^\top$ denote an arbitrary point on ∂G. One can verify that $\mathbf{n}_* = (-2\theta_*, 1)^\top$ is an inward-pointing normal vector of ∂G at $\mathbf{u}_*$ and

$$\varphi(\mathbf{u}; \theta_*) = \mathbf{u} \cdot \mathbf{n}_* - \mathbf{u}_* \cdot \mathbf{n}_* = \overrightarrow{\mathbf{u}_*\mathbf{u}} \cdot \mathbf{n}_* > 0 \quad \forall \mathbf{u} \in G.$$

Imagine we are walking along the boundary ∂G in the direction shown in the right subfigure of Figure 1; then the region G always lie entirely on the left side of the tangent lines, namely, the angle between the two vectors $\overrightarrow{\mathbf{u}_*\mathbf{u}}$ and $\mathbf{n}_*$ is always less than 90° for all $\mathbf{u} \in G$ and all $\mathbf{u}_* \in \partial G$. This geometric viewpoint intuitively illustrates the GQL representation G_* in (3.4) and its equivalence to the original G.

GEOMETRIC QUASILINEARIZATION FRAMEWORK 11

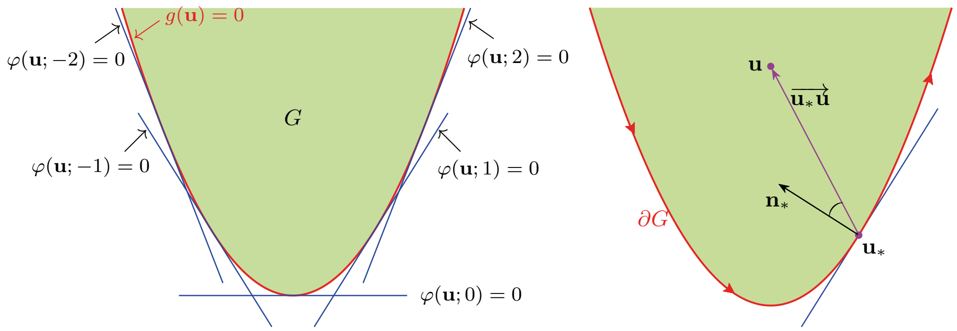
$g(\mathbf{u}) = 0$
$\varphi(\mathbf{u}; -2) = 0$
$\varphi(\mathbf{u}; 2) = 0$
G
$\varphi(\mathbf{u}; -1) = 0$
$\varphi(\mathbf{u}; 1) = 0$
$\varphi(\mathbf{u}; 0) = 0$
$\mathbf{u}$
$\overrightarrow{\mathbf{u}_*\mathbf{u}}$
$\mathbf{n}_*$
∂G
$\mathbf{u}_*$

Fig. 2: xxx.

Fig. 1 *Illustrations for Example* 3.5.

3.2. Concepts from Geometry and Convex Sets. Let us recall some concepts and results from the theory of geometry and convex analysis [64, 88, 40].

A hyperplane in $\mathbb{R}^N$ is a plane of dimension $N-1$. Let $\mathbf{n}_* \neq \mathbf{0}$ denote a normal vector of a hyperplane H, and let $\mathbf{u}_*$ be a point on H. Then H can be expressed as $H = \{\mathbf{u} \in \mathbb{R}^N : (\mathbf{u}-\mathbf{u}_*)\cdot\mathbf{n}_* = 0\}$, and it divides $\mathbb{R}^N$ into two halfspaces: $H^+ = \{\mathbf{u} \in \mathbb{R}^N : (\mathbf{u}-\mathbf{u}_*)\cdot\mathbf{n}_* \geq 0\}$ and $H^- = \{\mathbf{u} \in \mathbb{R}^N : (\mathbf{u}-\mathbf{u}_*)\cdot\mathbf{n}_* \leq 0\}$.

Definition 3.6 (supporting hyperplane and halfspace). *The hyperplane* $H = \{\mathbf{u} \in \mathbb{R}^N : (\mathbf{u}-\mathbf{u}_*)\cdot\mathbf{n}_* = 0\}$ *through* $\mathbf{u}_* \in \partial G$ *is called a supporting hyperplane to* G *at* $\mathbf{u}_*$ *if* G *lies in one of the two closed halfspaces determined by* H. *Furthermore, if the normal vector* $\mathbf{n}_*$ *points toward* G, *then the closed halfspace containing* G *is* $H^+ = \{\mathbf{u} \in \mathbb{R}^N : (\mathbf{u}-\mathbf{u}_*)\cdot\mathbf{n}_* \geq 0\}$ *and is called a closed supporting halfspace to* G.

Lemma 3.7 (supporting hyperplane theorem [64]). *If* G *is a convex set and* $\mathrm{int}(G) \neq \emptyset$, *then, for any* $\mathbf{u}_* \in \partial G$, *there exists a supporting hyperplane to* G *at* $\mathbf{u}_*$.

Remark 3.8. If the boundary ∂G is smooth at a point $\mathbf{u}_*$, then the supporting hyperplane to G at $\mathbf{u}_*$ is unique and coincides with the tangent [88, 40].

3.3. GQL Framework. We are now in a position to establish the GQL framework.

3.3.1. A Special Case. Inspired by Example 3.5, we first consider the special case that G is either open or closed with differentiable boundary. The general case will be discussed in subsection 3.3.2.

Proposition 3.9. *Suppose that Assumption* 3.4 *holds, the region* G *is either open or closed, and* ∂G *is differentiable. Then* G *has the following GQL representation:*

$$G_* = \left\{\mathbf{u} \in \mathbb{R}^N : \ (\mathbf{u}-\mathbf{u}_*)\cdot\mathbf{n}_* \succ 0 \quad \forall \mathbf{u}_* \in \partial G\right\}, \tag{3.5}$$

where the symbol "$\succ$" is taken as "$>$" if G *is open, or as "$\geq$" if* G *is closed, and* $\mathbf{n}_*$ *is only dependent on* $\mathbf{u}_*$ *and denotes the inward-pointing normal vector of* ∂G *at* $\mathbf{u}_*$.

A proof of Proposition 3.9 is presented in Appendix A. Following the proof, one can further extend the above result to any closed convex region G whose boundary is typically not everywhere smooth; in this case, the supporting hyperplanes at a nonsmooth boundary point are not unique. Let $\mathcal{N}(\mathbf{u}_*)$ denote the set of the inward-pointing unit normal vectors of all the supporting hyperplanes to G at $\mathbf{u}_* \in \partial G$. Then one can prove that

$$G = \left\{\mathbf{u} \in \mathbb{R}^N : \ (\mathbf{u}-\mathbf{u}_*)\cdot\mathbf{n} \geq 0 \ \ \forall \mathbf{n} \in \mathcal{N}(\mathbf{u}_*), \ \ \forall \mathbf{u}_* \in \partial G\right\}. \tag{3.6}$$

This means that any closed convex region is the intersection of all its closed supporting halfspaces [64]. However, the representation (3.6) is *not* applicable to a general convex region that is neither closed nor open (e.g., the invariant regions in (2.3) and (2.8)). Moreover, the representation (3.6) requires the information for *all* the supporting hyperplanes at each nonsmooth boundary point, which can be difficult to explicitly formulate or verify, so that (3.6) is not desirable in a bound-preserving study. A practical GQL representation for more general regions will be derived in subsection 3.3.2.

Example 3.10. To further illustrate the above-mentioned limitations of (3.6), we consider the simple example $G = \{(u_1, u_2)^\top : u_2 \geq |u_1|\}$ whose boundary corresponds to the graph of the convex nonsmooth function $y = |x|$. In this example, the representation (3.6) requires all the supporting hyperplanes to G, i.e., infinitely many subtangent lines (with slopes being any number between -1 and 1) for the graph of $y = |x|$ at $x = 0$. We will derive a more desirable GQL representation (3.8) in Theorem 3.11, which only requires two subtangent lines with slopes ± 1 for this example.

3.3.2. General Case. Consider a general convex region G that may be *not necessarily* open or closed and whose boundary may be not everywhere smooth. Note that the boundary of a convex region can be partitioned into several pieces, each of which can be *locally* represented as the graph of a convex function (with respect to a suitable supporting hyperplane). According to the classical theorems of Rademacher and Alexandrov (cf. [81]), any convex function is locally Lipschitz continuous and twice differentiable almost everywhere. Based on these facts and for convenience, we make a considerably mild assumption on the convex invariant region G. We assume that the boundary of G is piecewise C^1 and without loss of generality, for each $i \in \{1, \dots, I\}$, the function $g_i(\mathbf{u})$ in (3.1) is C^1 at any point on

$$\mathcal{S}_i := \partial G \cap \partial G_i \qquad \text{with} \quad G_i := \left\{\mathbf{u} \in \mathbb{R}^N : \ g_i(\mathbf{u}) \succ 0\right\},$$

where $\{\mathcal{S}_i\}$ are C^1 hypersurfaces in $\mathbb{R}^N$ and constitute the smooth pieces of ∂G, i.e., $\partial G = \cup_{1 \le i \le I} \mathcal{S}_i$. Notice that, in general, $\mathcal{S}_i$ may not equal ∂G_i, the region G_i may be not convex, and G may be neither open nor closed; see Figure 2 for a schematic illustration. These observations make our following discussions nontrivial.

We remark that $G_i = \{\mathbf{u} : g_i(\mathbf{u}) \ge 0\}$ is closed for $i \in \widehat{\mathbb{I}}$ and $G_i = \{\mathbf{u} : g_i(\mathbf{u}) > 0\}$ is open for $i \in \mathbb{I}$. Since, for each $i \in \mathbb{I}$, the set G_i is not necessarily convex, there is a possibility that G might not be entirely contained in an *open* supporting halfspace at $\mathbf{u}_* \in \mathcal{S}_i$. This issue is avoided if the open region $\cap_{i \in \mathbb{I}} G_i$ is convex, which is satisfied by all the examples in section 2. As such, we have

$$G \cap \left(\cup_{\mathbf{u}_* \in \mathcal{S}_i} \{\mathbf{u} \in \mathbb{R}^N : \ (\mathbf{u} - \mathbf{u}_*) \cdot \mathbf{n}_{i*} = 0\}\right) = \emptyset \quad \forall i \in \mathbb{I}, \tag{3.7}$$

where $\mathbf{n}_{i*}$ is an inward-pointing normal vector of $\mathcal{S}_i$ at $\mathbf{u}_*$.

18 KAILIANG WU AND CHI-WANG SHU

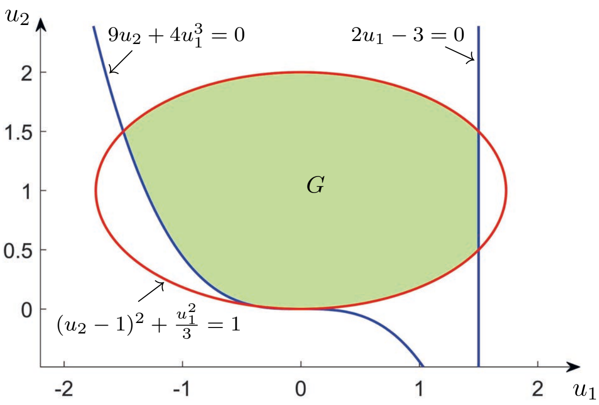

Fig. 6: xxx.

Fig. 2 *A convex region G involving nonlinear constraints.* $G = \{\mathbf{u} \in \mathbb{R}^2 : g_1(\mathbf{u}) \ge 0,\ g_2(\mathbf{u}) \ge 0,\ g_3(\mathbf{u}) > 0\}$ *with* $g_1(\mathbf{u}) = 3 - 2u_1$, $g_2(\mathbf{u}) = 9u_2 + 4u_1^3$, *and* $g_3(\mathbf{u}) = 1 - u_1^2/3 - (u_2 - 1)^2$.

Theorem 3.11. *Suppose that Assumption* 3.4 *holds, condition* (3.7) *is satisfied when* $\mathbb{I} \neq \emptyset$, *and the boundary of* G *is piecewise* C^1. *Then the region* G *has the following GQL representation:*

$$G_* = \Big\{ \mathbf{u} \in \mathbb{R}^N : \ \big(\mathbf{u} - \mathbf{u}_*\big) \cdot \mathbf{n}_{i*} \succ 0 \ \ \forall \mathbf{u}_* \in \mathcal{S}_i, \ 1 \leq i \leq I \Big\}, \tag{3.8}$$

where the symbol "$\succ$*" is taken as "*$>$*" if* $i \in \mathbb{I}$, *or as "*$\geq$*" if* $i \in \widehat{\mathbb{I}}$; *the nonzero vector* $\mathbf{n}_{i*}$ *denotes an inward-pointing normal vector of* $\mathcal{S}_i$ *at* $\mathbf{u}_*$.

Proof. The proof is divided into three steps.

(i) *Prove that* $G \subseteq G_*$. Let $\partial G =: \widetilde{\partial G} \cup \widehat{\partial G}$ with $\widetilde{\partial G}$ denoting the set of smooth boundary points and $\widehat{\partial G}$ the set of nonsmooth boundary points. For any $\mathbf{u}_* \in \widetilde{\partial G} \cap \mathcal{S}_i$, the hyperplane $(\mathbf{u} - \mathbf{u}_*) \cdot \mathbf{n}_{i*} = 0$ supports the region G, implying that

$$G \subseteq \big\{ \mathbf{u} \in \mathbb{R}^N : \ (\mathbf{u} - \mathbf{u}_*) \cdot \mathbf{n}_{i*} \geq 0 \big\} \quad \forall \mathbf{u}_* \in \widetilde{\partial G} \cap \mathcal{S}_i, \ 1 \leq i \leq I. \tag{3.9}$$

Next, we consider an arbitrary nonsmooth boundary point $\mathbf{u}_* \in \widehat{\partial G} \cap \mathcal{S}_i$. There exists a sequence of smooth boundary points $\{\mathbf{u}_*^{(j)}\}_{j \in \mathbb{N}} \subset \widetilde{\partial G} \cap \mathcal{S}_i$ such that $\lim_{j \to \infty} \mathbf{u}_*^{(j)} = \mathbf{u}_*$. For every $\mathbf{u}_*^{(j)}$, it follows from (3.9) that

$$\big(\mathbf{u} - \mathbf{u}_*^{(j)}\big) \cdot \mathbf{n}_{i, \mathbf{u}_*^{(j)}} \geq 0 \qquad \forall \mathbf{u} \in G, \tag{3.10}$$

where $\mathbf{n}_{i, \mathbf{u}_*^{(j)}}$ is the inward-pointing normal vector of $\mathcal{S}_i$ at $\mathbf{u}_*^{(j)}$ satisfying $\lim_{j \to \infty} \mathbf{n}_{i, \mathbf{u}_*^{(j)}} = \mathbf{n}_{i*}$. Taking the limit $j \to +\infty$ in (3.10) gives

$$(\mathbf{u} - \mathbf{u}_*) \cdot \mathbf{n}_{i*} \geq 0 \quad \forall \mathbf{u} \in G, \quad \forall \mathbf{u}_* \in \widehat{\partial G} \cap \mathcal{S}_i, \ 1 \leq i \leq I,$$

which along with (3.9) yields

$$G \subseteq \big\{ \mathbf{u} \in \mathbb{R}^N : \ \big(\mathbf{u} - \mathbf{u}_*\big) \cdot \mathbf{n}_{i*} \geq 0 \ \ \forall \mathbf{u}_* \in \mathcal{S}_i, \ 1 \leq i \leq I \big\}. \tag{3.11}$$

Based on (3.7), we then conclude that $G \subseteq G_*$.

(ii) *Prove that* $G_* \subseteq \mathrm{cl}(G)$ *by contradiction.* Assume that $G_* \nsubseteq \mathrm{cl}(G)$, namely, there exists $\mathbf{u}_0 \in G_*$ but $\mathbf{u}_0 \notin \mathrm{cl}(G)$. According to the theory of convex optimization [8], the minimum of the convex function $\zeta(\mathbf{u}) := \frac{1}{2}\|\mathbf{u} - \mathbf{u}_0\|^2$ over the closed convex region $\mathrm{cl}(G)$ is attained at a certain boundary point $\mathbf{u}_{*0} \in \partial G$. In other words, $\mathbf{u}_{*0}$ is a solution to the optimization problem

$$\begin{aligned} &\underset{\mathbf{u} \in \mathrm{cl}(G)}{\text{minimize}} \quad \zeta(\mathbf{u}) \\ &\text{subject to} \quad -g_i(\mathbf{u}) \leq 0 \ \ \forall i \in \mathbb{I}; \quad -g_i(\mathbf{u}) \leq 0 \ \ \forall i \in \widehat{\mathbb{I}}. \end{aligned} \tag{3.12}$$

Since the function $-g_i(\mathbf{u})$ is not necessarily convex, the problem (3.12) is not generally the standard form of convex optimization. Note that the assumption $\mathrm{int}(G) \neq \emptyset$ ensures the Slater condition [8, 7]. The Karush–Kuhn–Tucker conditions [8, 7] tell us that there exist $\{\lambda_i\}_{1 \leq i \leq I}$ such that

$$0 = \nabla \zeta(\mathbf{u}_{*0}) - \sum_{i=1}^{I} \lambda_i \nabla g_i(\mathbf{u}_{*0}), \tag{3.13}$$

$$0 = \lambda_i g_i(\mathbf{u}_{*0}), \qquad 1 \leq i \leq I, \tag{3.14}$$

$$\lambda_i \geq 0, \qquad 1 \leq i \leq I. \tag{3.15}$$

Define $\mathbb{I}_+ := \{1 \le i \le I : \lambda_i > 0\}$. Obviously $\mathbb{I}_+ \neq \emptyset$; otherwise $\lambda_i = 0$ for all $1 \le i \le I$, so that $\mathbf{u}_{*0} - \mathbf{u}_0 = \nabla\zeta(\mathbf{u}_{*0}) = \mathbf{0}$, which leads to the contradiction $\partial G \ni \mathbf{u}_{*0} = \mathbf{u}_0 \notin \mathrm{cl}(G)$. This also implies $\mathbf{u}_{*0} \neq \mathbf{u}_0$. Let $\mathbf{n}_{i*0}$ be the inward-pointing normal vector of $\mathcal{S}_i$ at $\mathbf{u}_{*0}$. Since there exists $\mu_i \ge 0$ such that $\nabla g_i(\mathbf{u}_{*,0}) = \mu_i \mathbf{n}_{i*0}$, condition (3.13) can be rewritten as

$$\mathbf{u}_{*0} - \mathbf{u}_0 = \sum_{i \in \mathbb{I}_+} \lambda_i \mu_i \mathbf{n}_{i*0}. \tag{3.16}$$

Thanks to (3.14), we obtain $g_i(\mathbf{u}_{*0}) = 0$ for all $i \in \mathbb{I}_+$, which along with $\mathbf{u}_{*0} \in \partial G$ leads to

$$\mathbf{u}_{*0} \in \mathcal{S}_i = \partial G_i \cap \partial G \qquad \forall i \in \mathbb{I}_+.$$

Because $\mathbf{u}_0 \in G_*$, we then have $(\mathbf{u}_0 - \mathbf{u}_{*0}) \cdot \mathbf{n}_{i*0} \succ 0$ for all $i \in \mathbb{I}_+$. This, together with (3.16) and $\mathbf{u}_{*0} \neq \mathbf{u}_0$, leads to a contradiction:

$$\begin{aligned} 0 > -\|\mathbf{u}_0 - \mathbf{u}_{*0}\|_2^2 &= (\mathbf{u}_0 - \mathbf{u}_{*0}) \cdot (\mathbf{u}_{*0} - \mathbf{u}_0) \\ &= (\mathbf{u}_0 - \mathbf{u}_{*0}) \cdot \left(\sum_{i \in \mathbb{I}_+} \lambda_i \mu_i \mathbf{n}_{i*0} \right) = \sum_{i \in \mathbb{I}_+} \lambda_i \mu_i \left(\mathbf{u}_0 - \mathbf{u}_{*0}\right) \cdot \mathbf{n}_{i*0} \ge 0. \end{aligned}$$

Hence the assumption $G_* \nsubseteq \mathrm{cl}(G)$ is incorrect. We have $G_* \subseteq \mathrm{cl}(G)$.

(iii) *Prove that* $G_* \subseteq G$. If $\mathbb{I} = \emptyset$, then G is a closed region and $G = \mathrm{cl}(G)$. We immediately obtain $G_* \subseteq G$ from step (ii). In the following, we focus on $\mathbb{I} \neq \emptyset$ and prove $G_* \subseteq G$ by contradiction. Assume that there exists $\mathbf{u}_0 \in G_*$, but $\mathbf{u}_0 \notin G$. As we have shown $G_* \subseteq \mathrm{cl}(G)$ in step (ii), we then get $\mathbf{u}_0 \in \mathrm{cl}(G) \setminus G = \partial G$. Note that $\mathbf{u}_0 \in G_*$ implies

$$(\mathbf{u}_0 - \mathbf{u}_*) \cdot \mathbf{n}_{i*} > 0 \qquad \forall \mathbf{u}_* \in \mathcal{S}_i, \quad \forall i \in \mathbb{I},$$

which leads to $\mathbf{u}_0 \notin \mathcal{S}_i = \partial G_i \cap \partial G$ for all $i \in \mathbb{I}$. It follows that $\mathbf{u}_0 \notin \partial G_i$ for all $i \in \mathbb{I}$. For $i \in \mathbb{I}$, one has $\mathbf{u}_0 \in \mathrm{cl}(G) \subseteq \mathrm{cl}(G_i)$, which gives

$$\mathbf{u}_0 \in \mathrm{cl}(G_i) \setminus \partial G_i = G_i \quad \forall i \in \mathbb{I}. \tag{3.17}$$

On the other hand, $\mathbf{u}_0 \in \mathrm{cl}(G) \subseteq \cap_{i \in \widehat{\mathbb{I}}} G_i$, which along with (3.17) implies $\mathbf{u}_0 \in (\cap_{i \in \mathbb{I}} G_i) \cap (\cap_{i \in \widehat{\mathbb{I}}} G_i) = G$. This contradicts the assumption that $\mathbf{u}_0 \notin G$. Hence the assumption is incorrect, and we have $G_* \subseteq G$.

Combining the conclusions proven in steps (i) and (iii), we obtain $G = G_*$. ◻

Remark 3.12. If we replace $\mathcal{S}_i$ with $\mathcal{S}_i \cap \widetilde{\partial G}$ for $i \in \widehat{\mathbb{I}}$ in (3.8), Theorem 3.11 remains valid, because for $i \in \widehat{\mathbb{I}}$ we have $\{\mathbf{u} : (\mathbf{u} - \mathbf{u}_*) \cdot \mathbf{n}_{i*} \ge 0 \ \forall \mathbf{u}_* \in \mathcal{S}_i\} = \{\mathbf{u} : (\mathbf{u} - \mathbf{u}_*) \cdot \mathbf{n}_{i*} \ge 0 \ \forall \mathbf{u}_* \in \mathcal{S}_i \cap \widetilde{\partial G}\}$.

Remark 3.13. An illustration of the GQL representation (3.8) for $N = 2$ is shown in Figure 3. In contrast to (3.6), the GQL representation (3.8) involves only at most N rather than *all* the supporting halfspaces at each nonsmooth "junction" point. This feature makes the GQL representation (3.8) much easier to formulate and construct. Besides, Theorem 3.11 does not require G to be closed or open.

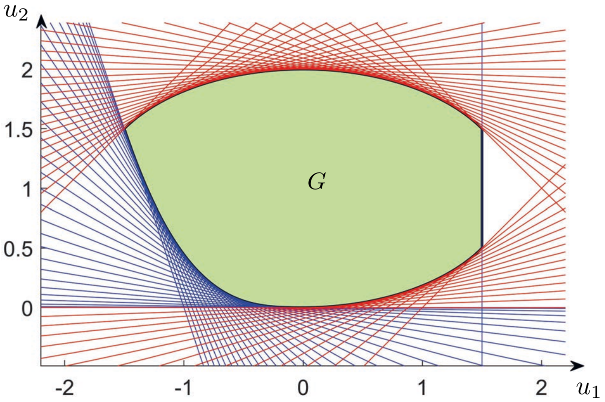

Fig. 7: xxx.

Fig. 3 *Illustration of the GQL representation for the convex region G given in Figure 2. The blue (resp., red) lines correspond to closed (resp., open) supporting halfspaces.*

Remark 3.14. The *support function* [88] for G is defined by $\psi(\mathbf{n}_*) := \sup_{\mathbf{u}\in G}\{\mathbf{u}\cdot \mathbf{n}_*\}$, whose effective domain is denoted as $\mathcal{N} := \{\mathbf{n}_* \in \mathbb{R}^N : \psi(\mathbf{n}_*) < +\infty\}$. If the convex region G is *either closed or open*, then by the support function [88, Theorem 13.1], one can derive $G = \{\mathbf{u} \in \mathbb{R}^N : \psi(\mathbf{n}_*) - \mathbf{u}\cdot\mathbf{n}_* \succ 0 \ \ \forall \mathbf{n}_* \in \mathcal{N}\setminus\{\mathbf{0}\}\}$. This representation has the same limitations as (3.6), involving all the supporting halfspaces at each nonsmooth point. Moreover, it is in general quite difficult to compute the support function, e.g., for the invariant regions in section 2.

Remark 3.15 (relation to Legendre transform). In some special cases, it is also possible to interpret the GQL representation geometrically via the *Legendre transform*. For example, consider $G = \{\mathbf{u} \in \mathbb{R}^2 : g(\mathbf{u}) > 0\}$ with $g(\mathbf{u}) = u_2 - f(u_1)$ being concave. That is, the function f is convex, with its Legendre transform [3] defined as $\widehat{f}(s_*) := \sup_{u_1}\{s_* u_1 - f(u_1)\}$. For any fixed s_* with $\widehat{f}(s_*) < +\infty$, the line $u_2 = s_* u_1 - \widehat{f}(s_*)$ is tangent to the graph of f with slope s_*. By convexity of f, the envelope of all these tangent lines has the equation $u_2 = f(u_1)$ (see [3, pages 63–64]), which defines the boundary of G. Hence, the set G has the equivalent representation $G_* = \{\mathbf{u} : u_2 > s_* u_1 - \widehat{f}(s_*)\ \forall s_*\}$. Note that here the free auxiliary variable s_* plays the role of slope, which is different from $\mathbf{u}_*$ in Theorem 3.11. Similar to (3.8), the representation based on the Legendre transform also has the limitation discussed in Example 3.10. Moreover, it seems that the Legendre transform may not apply in our general cases, especially when $\{g_i\}$ are not concave.

Remark 3.16 (significance of GQL). Compared to the original nonlinear form (3.1) of the invariant region G, its equivalent GQL representation G_* in (3.8) is described with only linear constraints. Such linearity has great advantages over the original nonlinear form (3.1) in analyzing and designing bound-preserving schemes; see sections 5 to 7.

4. Construction of Geometric Quasilinearization.

Based on Theorems 3.11 and 3.2, we introduce three simple effective methods for constructing the GQL representation of G. Several specific examples and applications will be given in sections 5 to 7.

4.1. Gradient-Based Method.

The first method is based on the following result, which is a direct consequence of Theorem 3.11.

THEOREM 4.1. *Assume that the hypotheses of Theorem* 3.11 *hold and*

$$\nabla g_i(\mathbf{u}_*) \neq \mathbf{0} \quad \forall \mathbf{u}_* \in \mathcal{S}_i,\ 1 \leq i \leq I; \tag{4.1}$$

then the invariant region G is exactly equivalent to

$$G_* = \Big\{ \mathbf{u} \in \mathbb{R}^N :\ \varphi_i(\mathbf{u}; \mathbf{u}_*) \succ 0 \quad \forall \mathbf{u}_* \in \mathcal{S}_i,\ 1 \leq i \leq I \Big\}, \tag{4.2}$$

where the function φ_i is linear with respect to $\mathbf{u}$, defined by

$$\varphi_i(\mathbf{u}; \mathbf{u}_*) := (\mathbf{u} - \mathbf{u}_*) \cdot \nabla g_i(\mathbf{u}_*). \tag{4.3}$$

Theorem 4.1 says that if $\{\nabla g_i\}$ are computable and satisfy (4.1), then we can directly obtain the GQL representation in the form (4.2) with (4.3).

In some cases, however, it is difficult to calculate the gradients of nonlinear functions $\{g_i\}$, e.g., the implicit functions in (2.8) and (2.15). This motivates us to propose the following cross-product method based on a suitable parametrization of the hypersurface $\mathcal{S}_i$. The use of parametrization can also reduce or decouple the free auxiliary variables, which is highly desirable for bound-preserving applications; see Remark 5.2 and the examples in sections 5 and 6.

4.2. Cross-Product Method.

Assume that for each i the hypersurface $\mathcal{S}_i$ has the parametric expression

$$\mathcal{S}_i = \left\{ \mathbf{u}_* = \mathbf{U}_i(\boldsymbol{\theta}_{i*}) :\ \boldsymbol{\theta}_{i*} \in \Theta_i \subseteq \mathbb{R}^{N-1} \right\}, \tag{4.4}$$

where $\mathbf{U}_i$ is a C^1 vector function defined on the parameter domain Θ_i with $\mathcal{S}_i$ being the function range. Denote $\theta_{i*}^{(k)}$ as the kth component of $\boldsymbol{\theta}_{i*}$. For each i, we define

$$\boldsymbol{\tau}_{i,k}(\boldsymbol{\theta}_{i*}) := \frac{\partial \mathbf{U}_i}{\partial \theta_{i*}^{(k)}}, \qquad 1 \leq k \leq N-1.$$

The vectors $\{\boldsymbol{\tau}_{i,k}(\boldsymbol{\theta}_{i*}) : 1 \leq k \leq N-1\}$ are $(N-1)$ tangent vectors of the hypersurface $\mathcal{S}_i$ and generate its local tangent space at $\mathbf{u}_*$. Then, the normal vector of $\mathcal{S}_i$ at $\mathbf{u}_*$ can be constructed using the $(N-1)$-ary analogue of the cross product (cf. [96, pages 83–85]) in $\mathbb{R}^N$,

$$\mathbf{n}_{i*} = \delta_{i*} \bigwedge_{k=1}^{N-1} \boldsymbol{\tau}_{i,k}(\boldsymbol{\theta}_{i*}) := \delta_{i*} \boldsymbol{\tau}_{i,1}(\boldsymbol{\theta}_{i*}) \times \boldsymbol{\tau}_{i,2}(\boldsymbol{\theta}_{i*}) \times \cdots \times \boldsymbol{\tau}_{i,N-1}(\boldsymbol{\theta}_{i*}),$$

where δ_{i*} is a nonzero factor, which may be used to simplify the final formula or to adjust the sign such that $\mathbf{n}_{i*}$ is directed toward the interior of G.

As a direct consequence of Theorem 3.11, the following result holds.

THEOREM 4.2. *Suppose the hypotheses of Theorem* 3.11 *hold and*

$$\bigwedge_{k=1}^{N-1} \boldsymbol{\tau}_{i,k}(\boldsymbol{\theta}_{i*}) \neq \mathbf{0} \quad \forall \boldsymbol{\theta}_{i*} \in \Theta_i,\ 1 \leq i \leq I;$$

then the region G is exactly equivalent to

$$G_* = \Big\{ \mathbf{u} \in \mathbb{R}^N :\ \varphi_i(\mathbf{u}; \boldsymbol{\theta}_{i*}) \succ 0 \quad \forall \boldsymbol{\theta}_{i*} \in \Theta_i,\ 1 \leq i \leq I \Big\}, \tag{4.5}$$

where the function φ_i is linear with respect to $\mathbf{u}$*, defined by*

$$\varphi_i(\mathbf{u};\boldsymbol{\theta}_{i*}) := \Big(\mathbf{u}-\mathbf{U}_i(\boldsymbol{\theta}_{i*})\Big)\cdot\left(\delta_{i*}\bigwedge_{k=1}^{N-1}\boldsymbol{\tau}_{i,k}(\boldsymbol{\theta}_{i*})\right). \tag{4.6}$$

Remark 4.3 (parametrization). The choices of free auxiliary variables are not unique. They can depend on the parametrization of the region boundary if the above cross-product method is used. Different sets of free auxiliary variables may lead to possibly different formulations of the GQL representation, which are all equivalent to G. However, a suitable set of free auxiliary variables facilitates the derivation of the GQL representation and the bound-preserving analysis. In many cases, there exists a natural physics-based parametrization of the hypersurface $\mathcal{S}_i$; see the examples in sections 5 and 6. The advantages of using the parametric form (4.4) in the GQL representation will become more clear in the examples and bound-preserving applications in sections 5 to 7.

4.3. Constructive Method.

For completeness, we also summarize the constructive approach and its variant as our third method. Recall that Theorem 3.2 has told us that if we can construct linear functions $\varphi_i(\mathbf{u};\boldsymbol{\theta}_{i*})$, $1\le i\le I$, such that (3.3) holds, then the GQL representation of G is (3.2). The constructive approach does not require the assumptions in Theorems 4.1 and 4.2, but it often needs some empirical trial-and-error techniques to find the qualified $\{\varphi_i\}$. In practice, one may use the proposed three methods in a hybrid way: first formulate $\{\varphi_i\}$ formally via either (4.3) or (4.6) and then verify (3.3). Such a hybrid approach is efficient, as it may exempt the assumptions in Theorems 4.1 and 4.2 and also avoid the trial-and-error procedure.

5. GQL for Bound-Preserving Study on Euler and Navier–Stokes Systems.

As an example, this section discusses in detail the GQL approach for the preservation of a particular invariant region (2.2), which is shared by the Euler and Navier–Stokes systems. We will illustrate how to use the methods in section 4 to derive the GQL representation and show the advantages of the GQL approach for bound-preserving study. More challenging examples and applications of GQL will be discussed in sections 6 and 7 for various PDE systems.

5.1. GQL Representation of Invariant Region.

We first derive the GQL representation of the invariant region G in (2.2).

THEOREM 5.1. *For the* 1*D Euler and Navier–Stokes systems, the GQL representation of the invariant region* G *in* (2.2) *is given by*

$$G_* = \left\{\mathbf{u}=(\rho,m,E)^\top:\ \rho>0,\quad \varphi(\mathbf{u};v_*)>0\ \ \forall v_*\in\mathbb{R}\right\} \tag{5.1}$$

with $\varphi(\mathbf{u};v_*) := E - mv_* + \rho\frac{v_*^2}{2}$ *being linearly dependent on* $\mathbf{u}$*.*

Proof. For illustrative purposes, we use the three methods proposed in section 4 to derive the GQL representation for this example. Note that the first constraint in (2.2) is linear.

(i) *Gradient-based method.* For the second constraint in (2.2), the gradient of the internal energy $g(\mathbf{u}) = E-\frac{m^2}{2\rho}$ is $\nabla g(\mathbf{u}) = (\frac{m^2}{2\rho^2},-\frac{m}{\rho},1)^\top$, and the associated boundary hypersurface $\mathcal{S}=\{\mathbf{u}_*=(\rho_*,m_*,E_*)^\top:\rho_*>0,\ g(\mathbf{u}_*)=0\}$ can be parameterized as

$$\mathcal{S}=\left\{\mathbf{u}_*=\left(\rho_*,\rho_*v_*,\frac{\rho_*}{2}v_*^2\right)^\top:\rho_*>0,\ v_*\in\mathbb{R}\right\}. \tag{5.2}$$

For $\mathbf{u}_* \in \mathcal{S}$ and $\mathbf{u} = (\rho, m, E)^\top$, we have

$$(\mathbf{u} - \mathbf{u}_*) \cdot \nabla g(\mathbf{u}_*) = (\rho - \rho_*)\frac{v_*^2}{2} + (m - \rho_* v_*)(-v_*) + E - \frac{\rho_*}{2} v_*^2 = \varphi(\mathbf{u}; v_*).$$

By Theorem 4.1, we obtain the GQL representation (5.1) of G.

(ii) *Cross-product method.* Based on the parametrization of $\mathcal{S}$ in (5.2), we can compute the normal vector of $\mathcal{S}$ at $\mathbf{u}_*$ by cross product,

$$\frac{\partial \mathbf{u}_*}{\partial \rho_*} \times \frac{\partial \mathbf{u}_*}{\partial v_*} = \left(1, v_*, \frac{1}{2} v_*^2\right)^\top \times (0, \rho_*, \rho_* v_*)^\top = \rho_* \left(\frac{1}{2} v_*^2, -v_*, 1\right)^\top =: \frac{1}{\delta_*}\mathbf{n}_*,$$

where $\delta_* = 1/\rho_*$ is a nonzero factor. By Theorem 4.2 and $(\mathbf{u} - \mathbf{u}_*) \cdot \mathbf{n}_* = \varphi(\mathbf{u}; v_*)$, we get (5.1).

(iii) *Constructive method.* Observe that

$$\varphi(\mathbf{u}; v_*) = E - m v_* + \rho \frac{v_*^2}{2} = \frac{\rho}{2}\left(v_* - \frac{m}{\rho}\right)^2 + g(\mathbf{u}) \geq g(\mathbf{u}),$$

which implies $\min_{v_* \in \mathbb{R}} \varphi(\mathbf{u}; v_*) = g(\mathbf{u})$ for $\rho > 0$. By Theorem 3.2, one also obtains (5.1). □

Remark 5.2. Note that only one free auxiliary variable v_* appears explicitly in the GQL representation (5.1). This is due to the use of the parametric form (5.2).

Remark 5.3 (physical interpretation of GQL). It seems that the linear function $\varphi(\mathbf{u}; v_*)$ plays an energy-like role from a physical point of view. For the present example, $\varphi(\mathbf{u}; v_*) = \frac{1}{2}\rho(v - v_*)^2 + \rho e$, which represents the total energy in the reference frame moving at a velocity of v_*.

One can similarly utilize the cross-product method to construct the GQL representation of the invariant region $\widetilde{G}$ in (2.3), where the minimum entropy principle $S(\mathbf{u}) := p\rho^{-\Gamma} \geq S_{min}$ is also included as a constraint.

THEOREM 5.4. *For the* 1D *Euler and Navier–Stokes systems, the GQL representation of the invariant region* $\widetilde{G}$ *in* (2.3) *is given by*

$$\widetilde{G}_* = \left\{\mathbf{u} = (\rho, m, E)^\top :\ \rho > 0,\ \ \widetilde{\varphi}(\mathbf{u}; \rho_*, v_*) \geq 0 \ \ \forall \rho_* \in \mathbb{R}^+, \ \ \forall v_* \in \mathbb{R}\right\} \tag{5.3}$$

with $\widetilde{\varphi}(\mathbf{u}; \rho_*, v_*) := \mathbf{u} \cdot \mathbf{n}_* + S_{min}\rho_*^\Gamma$ *and* $\mathbf{n}_* := \left(\frac{v_*^2}{2} - \frac{S_{min}\Gamma\rho_*^{\Gamma-1}}{\Gamma-1},\ -v_*,\ 1\right)^\top$.

Proof. We only need to handle the nonlinear constraint $\widetilde{g}(\mathbf{u}) > 0$ in (2.3), with the boundary hypersurface $\widetilde{\mathcal{S}} := \{\mathbf{u}_* = (\rho_*, m_*, E_*) : \rho_* > 0,\ \widetilde{g}(\mathbf{u}_*) = 0\}$. Motivated by the equivalence of $\widetilde{g}(\mathbf{u}) = 0$ and $p = S_{min}\rho^\Gamma$, we find a natural physics-based parametrization of $\widetilde{\mathcal{S}}$ given by

$$\widetilde{\mathcal{S}} = \left\{\mathbf{u}_* = \left(\rho_*,\ \rho_* v_*,\ \frac{1}{2}\rho_* v_*^2 + \frac{S_{min}\rho_*^\Gamma}{\Gamma - 1}\right)^\top \ :\ \rho_* > 0,\ v_* \in \mathbb{R}\right\}.$$

Then we can derive the normal vector $\mathbf{n}_*$ of $\mathcal{S}$ at $\mathbf{u}_*$ by cross product,

$$\frac{\partial \mathbf{u}_*}{\partial \rho_*} \times \frac{\partial \mathbf{u}_*}{\partial v_*} = \left(1,\ v_*,\ \frac{S_{min}\Gamma\rho_*^{\Gamma-1}}{\Gamma - 1} + \frac{v_*^2}{2}\right)^\top \times (0,\ \rho_*,\ \rho_* v_*)^\top = \rho_* \mathbf{n}_*.$$

By Theorem 4.2 and $(\mathbf{u} - \mathbf{u}_*) \cdot \mathbf{n}_* = \widetilde{\varphi}(\mathbf{u}; \rho_*, v_*)$, we obtain the GQL representation (5.3). □

5.2. GQL for Bound-Preserving Analysis. Consider a finite volume scheme

$$\bar{\mathbf{u}}_j^{n+1} = \bar{\mathbf{u}}_j^n - \sigma\left(\hat{\mathbf{f}}_{j+\frac{1}{2}} - \hat{\mathbf{f}}_{j-\frac{1}{2}}\right) \tag{5.4}$$

for solving the 1D Euler system (2.1) on a uniform spatial mesh $\{[x_{j-1/2}, x_{j+1/2}]\}$ with $\sigma := \Delta t/\Delta x$ denoting the ratio of the temporal step-size Δt to the spatial step-size Δx. Here $\bar{\mathbf{u}}_j^n$ is an approximation to the average of $\mathbf{u}(x, t_n)$ on cell $[x_{j-1/2}, x_{j+1/2}]$, and $\hat{\mathbf{f}}_{j+1/2}$ is a numerical flux at $x_{j+1/2}$. For system (2.1), $\mathbf{f}(\mathbf{u}) = v\mathbf{u} + p(0,1,v)^\top$, which will be used in the following analysis.

We apply the GQL approach to analyze the bound-preserving property of the scheme (5.4) with the invariant region G defined in (2.2). Thanks to the GQL representation in Theorem 5.1, we have

$$G = G_* = \{\mathbf{u} : \ \mathbf{u}\cdot\mathbf{e}_1 > 0, \ \mathbf{u}\cdot\mathbf{n}_* > 0 \ \ \forall v_* \in \mathbb{R}\} \tag{5.5}$$

with $\mathbf{e}_1 := (1,0,0)^\top$ and $\mathbf{n}_* := \left(\frac{v_*^2}{2}, -v_*, 1\right)^\top$. GQL converts the nonlinear bound-preserving problem into preserving the positivity of $\mathbf{u}\cdot\mathbf{e}_1$ and $\mathbf{u}\cdot\mathbf{n}_*$, which are both *linear scalar* functions with respect to $\mathbf{u}$, and thus simplifies the bound-preserving study.

5.2.1. Lax–Friedrichs Scheme for Euler System. To clearly illustrate the basic idea, we begin with the simple Lax–Friedrichs scheme with the numerical flux $\hat{\mathbf{f}}_{j+1/2}$ taken to be

$$\hat{\mathbf{f}}^{\rm LF}(\bar{\mathbf{u}}_j^n, \bar{\mathbf{u}}_{j+1}^n) := \frac{1}{2}\Big(\mathbf{f}(\bar{\mathbf{u}}_j^n) + \mathbf{f}(\bar{\mathbf{u}}_{j+1}^n) - \alpha_n(\bar{\mathbf{u}}_{j+1}^n - \bar{\mathbf{u}}_j^n)\Big), \tag{5.6}$$

where $\alpha_n := \max_j \alpha(\bar{\mathbf{u}}_j^n)$ with $\alpha(\mathbf{u}) := |v| + \sqrt{\Gamma p/\rho}$ being the spectral radius of the Jacobian matrix $\partial\mathbf{f}/\partial\mathbf{u}$. Given that $\bar{\mathbf{u}}_j^n \in G$ for all j, we wish that $\bar{\mathbf{u}}_j^{n+1} \in G$.

For $\mathbf{n} = \mathbf{e}_1$ and $\mathbf{n} = \mathbf{n}_*$, thanks to the linearity of $\mathbf{u}\cdot\mathbf{n}$, we obtain

$$\bar{\mathbf{u}}_j^{n+1}\cdot\mathbf{n} = (1-\sigma\alpha_n)\,\bar{\mathbf{u}}_j^n\cdot\mathbf{n} + \frac{\sigma}{2}\Big(\alpha_n\bar{\mathbf{u}}_{j+1}^n\cdot\mathbf{n} - \mathbf{f}(\bar{\mathbf{u}}_{j+1}^n)\cdot\mathbf{n} + \alpha_n\bar{\mathbf{u}}_{j-1}^n\cdot\mathbf{n} + \mathbf{f}(\bar{\mathbf{u}}_{j-1}^n)\cdot\mathbf{n}\Big).$$

The problem boils down to controlling the effect of $\mathbf{f}(\bar{\mathbf{u}}_\pm^n)\cdot\mathbf{n}$ by using the positivity of $\bar{\mathbf{u}}_{j\pm1}^n\cdot\mathbf{n}$. For any $\mathbf{u}\in G$, we have $\mathbf{u}\cdot\mathbf{n} > 0$ and

$$\begin{aligned}
\pm\mathbf{f}(\mathbf{u})\cdot\mathbf{e}_1 &= \pm v(\mathbf{u}\cdot\mathbf{e}_1) < \alpha(\mathbf{u})\mathbf{u}\cdot\mathbf{e}_1,\\
\pm\mathbf{f}(\mathbf{u})\cdot\mathbf{n}_* &= \pm v(\mathbf{u}\cdot\mathbf{n}_*) \pm p(v-v_*)\\
&\le |v|(\mathbf{u}\cdot\mathbf{n}_*) + \left(\frac{1}{2}\rho(v-v_*)^2 + \rho e\right)\frac{p}{\rho\sqrt{2e}} \qquad (5.7)\\
&= \left(|v| + \frac{p}{\rho\sqrt{2e}}\right)\mathbf{u}\cdot\mathbf{n}_* < \alpha(\mathbf{u})\mathbf{u}\cdot\mathbf{n}_*,
\end{aligned}$$

which yield $\alpha_n\bar{\mathbf{u}}_{j\pm1}^n\cdot\mathbf{n} \mp \mathbf{f}(\bar{\mathbf{u}}_{j\pm1}^n)\cdot\mathbf{n} > 0$. Here the step (5.7) follows from the AM-GM inequality $\frac{1}{2}\rho(v-v_*)^2 + \rho e \ge 2\sqrt{\frac{1}{2}\rho(v-v_*)^2\rho e} = |v-v_*|\rho\sqrt{2e}$. Thus, we obtain $\bar{\mathbf{u}}_j^{n+1}\cdot\mathbf{n} > (1-\sigma\alpha_n)\bar{\mathbf{u}}_j^n\cdot\mathbf{n} \ge 0$ provided that $\sigma\alpha_n \le 1$. This proves that the scheme (5.4) with the Lax–Friedrichs flux (5.6) is bound-preserving under the standard Courant–Friedrichs–Lewy (CFL) condition $\sigma\alpha_n \le 1$.

Remark 5.5. As we have seen, unlike the direct approaches that require substituting the target scheme into the original nonlinear constraint of G in (2.2), the GQL approach skillfully transforms all the constraints to linear ones which can be investigated in a unified and easier way. The above analysis also applies to a general equation of state if we redefine $\alpha(\mathbf{u}) := |v| + \max\{c_s, p/(\rho\sqrt{2e})\}$, where c_s is the sound speed.

Remark 5.6. A classic approach showing the positivity of the Lax–Friedrichs scheme is based on a certain average of the exact Riemann solutions to two split equations; see [86, page 129] for details. To avoid the interaction of waves, the proof in [86] actually requires α_n to be a rigorous upper bound of the maximum wave speed (this is not tractable when fast shocks exist, as discussed in [43]), and it also needs a stricter CFL condition $\sigma\alpha_n \leq \frac{1}{2}$. The approach in [86] is not applicable to multidimensional MHD systems [104, 106, 108], whose exact Riemann solutions are generally *not* bound-preserving when there is a jump in the magnetic component normal to the cell interface.

5.2.2. Gas-Kinetic Scheme for Euler System. In order to demonstrate the significant advantages of the GQL approach in bound-preserving analysis, we consider a challenging example—the gas-kinetic scheme [101] with the numerical flux $\hat{\mathbf{f}}_{j+1/2}$ taken to be

$$\hat{\mathbf{f}}^{\mathrm{GK}}(\bar{\mathbf{u}}_j^n, \bar{\mathbf{u}}_{j+1}^n) := \mathbf{f}^+(\bar{\mathbf{u}}_j^n) + \mathbf{f}^-(\bar{\mathbf{u}}_{j+1}^n), \tag{5.8}$$

$$\mathbf{f}^\pm(\mathbf{u}) := \int_{\mathbb{R}^\pm}\int_{\mathbb{R}^M} \begin{pmatrix} w \\ w^2 \\ \frac{w}{2}(w^2 + \boldsymbol{\xi}^2) \end{pmatrix} F(w, \boldsymbol{\xi}; \mathbf{u}) \mathrm{d}\boldsymbol{\xi} \mathrm{d}w, \tag{5.9}$$

where w is the particle velocity, $\boldsymbol{\xi} \in \mathbb{R}^M$ denotes the internal variables with the degrees of freedom $M = (3 - \Gamma)/(\Gamma - 1)$, and the equilibrium distribution function F is

$$F(w, \boldsymbol{\xi}; \mathbf{u}) := \rho \left(\frac{\lambda}{\pi}\right)^{\frac{M+1}{2}} \mathrm{e}^{-\lambda\left((w-v)^2 + \|\boldsymbol{\xi}\|^2\right)}, \tag{5.10}$$

with ρ the fluid velocity, v the fluid velocity, and $\lambda = \rho/(2p)$.

The bound-preserving property of this scheme was studied in, for example, [39, 101]. A traditional approach in [101] consists of three steps: (i) evaluate the integration (5.9) as

$$\mathbf{f}^\pm(\mathbf{u}) = \rho \begin{pmatrix} \frac{v}{2}\mathrm{erfc}(\mp\sqrt{\lambda}v) \pm \frac{1}{2}\frac{\mathrm{e}^{-\lambda v^2}}{\sqrt{\pi\lambda}} \\ \left(\frac{v^2}{2} + \frac{1}{4\lambda}\right)\mathrm{erfc}(\mp\sqrt{\lambda}v) \pm \frac{v}{2}\frac{\mathrm{e}^{-\lambda v^2}}{\sqrt{\pi\lambda}} \\ \left(\frac{v^3}{4} + \frac{M+3}{8\lambda}v\right)\mathrm{erfc}(\mp\sqrt{\lambda}v) \pm \left(\frac{v^2}{4} + \frac{M+2}{8\lambda}\right)\frac{\mathrm{e}^{-\lambda v^2}}{\sqrt{\pi\lambda}} \end{pmatrix} \tag{5.11}$$

with $\mathrm{erfc}(x) := \frac{2}{\sqrt{\pi}}\int_x^{+\infty} \mathrm{e}^{-w^2} dw$; (ii) plug the numerical flux (5.8) with (5.11) into (5.4) and split the scheme (5.4) into two steps; and (iii) check the bound-preserving properties of the split schemes by verifying the original nonlinear constraints of G in (2.2). For this scheme, verifying the nonlinear constraint in (2.2) is very difficult and complicated.

Owing to the linear feature, the GQL approach is highly effective for this challenging case. For $\mathbf{n} = \mathbf{e}_1$ or $\mathbf{n} = \mathbf{n}_*$, thanks to the linearity of $\mathbf{u} \cdot \mathbf{n}$, we obtain

$$\bar{\mathbf{u}}_j^{n+1} \cdot \mathbf{n} = \bar{\mathbf{u}}_j^n \cdot \mathbf{n} - \sigma\left(\mathbf{f}^+(\bar{\mathbf{u}}_j^n) - \mathbf{f}^-(\bar{\mathbf{u}}_j^n)\right) \cdot \mathbf{n} - \sigma\mathbf{f}^-(\bar{\mathbf{u}}_{j+1}^n) \cdot \mathbf{n} + \sigma\mathbf{f}^+(\bar{\mathbf{u}}_{j-1}^n) \cdot \mathbf{n}.$$

Note that, for any $\mathbf{u} \in G$, we have $F(w, \boldsymbol{\xi}; \mathbf{u}) > 0$ and

$$\pm\mathbf{f}^{\pm}(\mathbf{u}) \cdot \mathbf{e}_1 = \int_{\mathbb{R}^{\pm}} \int_{\mathbb{R}^M} |w| F(w, \boldsymbol{\xi}; \mathbf{u}) \mathrm{d}\boldsymbol{\xi} \mathrm{d}w > 0, \tag{5.12}$$

$$\pm\mathbf{f}^{\pm}(\mathbf{u}) \cdot \mathbf{n}_* = \int_{\mathbb{R}^{\pm}} \int_{\mathbb{R}^M} \frac{|w|}{2} \Big((w - v_*)^2 + \|\boldsymbol{\xi}\|^2 \Big) F(w, \boldsymbol{\xi}; \mathbf{u}) \mathrm{d}\boldsymbol{\xi} \mathrm{d}w > 0. \tag{5.13}$$

It follows, for $\mathbf{n} = \mathbf{e}_1$ and $\mathbf{n} = \mathbf{n}_*$, respectively, that

$$\bar{\mathbf{u}}_j^{n+1} \cdot \mathbf{n} > \bar{\mathbf{u}}_j^n \cdot \mathbf{n} - \sigma \left(\mathbf{f}^+(\bar{\mathbf{u}}_j^n) - \mathbf{f}^-(\bar{\mathbf{u}}_j^n) \right) \cdot \mathbf{n}. \tag{5.14}$$

Next, we use the positivity of $\mathbf{u} \cdot \mathbf{n}$ to bound the effect of $(\mathbf{f}^+(\mathbf{u}) - \mathbf{f}^-(\mathbf{u})) \cdot \mathbf{n}$ as follows:

$$\begin{aligned}
\big(\mathbf{f}^+(\mathbf{u}) - \mathbf{f}^-(\mathbf{u})\big) \cdot \mathbf{e}_1 &= (\mathbf{u} \cdot \mathbf{e}_1) \left(\frac{\lambda}{\pi} \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}} |w| \mathrm{e}^{-\lambda(w-v)^2} \mathrm{d}w \right) \\
&\le (\mathbf{u} \cdot \mathbf{e}_1) \left(\frac{\lambda}{\pi} \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}} (|v| + |w - v|) \mathrm{e}^{-\lambda(w-v)^2} \mathrm{d}w \right) \\
&= (\mathbf{u} \cdot \mathbf{e}_1) \left(|v| + 1/\sqrt{\pi\lambda} \right) < \alpha(\mathbf{u}) \mathbf{u} \cdot \mathbf{e}_1, \qquad (5.15)
\end{aligned}$$

$$\begin{aligned}
\big(\mathbf{f}^+(\mathbf{u}) - \mathbf{f}^-(\mathbf{u})\big) \cdot \mathbf{n}_* &= \int_{\mathbb{R}} \int_{\mathbb{R}^M} \frac{|w|}{2} \Big((w - v_*)^2 + \|\boldsymbol{\xi}\|^2 \Big) F(w, \boldsymbol{\xi}; \mathbf{u}) \mathrm{d}\boldsymbol{\xi} \mathrm{d}w \\
&\le \int_{\mathbb{R}} \int_{\mathbb{R}^M} \frac{|v| + |w - v|}{2} \Big((w - v_*)^2 + \|\boldsymbol{\xi}\|^2 \Big) F(w, \boldsymbol{\xi}; \mathbf{u}) \mathrm{d}\boldsymbol{\xi} \mathrm{d}w \\
&= |v| (\mathbf{u} \cdot \mathbf{n}_*) + \frac{\rho}{2\sqrt{\pi\lambda}} \left((v - v_*)^2 + \frac{M+2}{2\lambda} \right) \\
&\le |v| (\mathbf{u} \cdot \mathbf{n}_*) + \frac{\rho}{2\sqrt{\pi\lambda}} \left((v - v_*)^2 + \frac{M+1}{2\lambda} \right) \frac{M+2}{M+1} \\
&= \left(|v| + \frac{M+2}{M+1} (\pi\lambda)^{-\frac{1}{2}} \right) (\mathbf{u} \cdot \mathbf{n}_*) < \alpha(\mathbf{u}) (\mathbf{u} \cdot \mathbf{n}_*). \qquad (5.16)
\end{aligned}$$

This implies $\left(\mathbf{f}^+(\bar{\mathbf{u}}_j^n) - \mathbf{f}^-(\bar{\mathbf{u}}_j^n) \right) \cdot \mathbf{n} < \alpha(\bar{\mathbf{u}}_j^n) \bar{\mathbf{u}}_j^n \cdot \mathbf{n} \le \alpha_n \bar{\mathbf{u}}_j^n \cdot \mathbf{n}$. It then follows from (5.14) that $\bar{\mathbf{u}}_j^{n+1} \cdot \mathbf{n} > (1 - \sigma\alpha_n) \bar{\mathbf{u}}_j^n \cdot \mathbf{n} \ge 0$, provided that $\sigma\alpha_n \le 1$. This proves that the scheme (5.4) with the gas-kinetic flux (5.8) is bound-preserving under the standard CFL condition $\sigma\alpha_n \le 1$.

Remark 5.7. The linearity of GQL brought about by introducing the free auxiliary variable v_* provides remarkable advantages in our above analysis. Because v_* is independent of all the system variables $\mathbf{u}$, it can freely move across the integrals. We no longer need to substitute a complicated scheme into the nonlinear function $g(\mathbf{u})$ in (2.2) to verify $g(\mathbf{u}) > 0$. Instead, we work on the simpler but equivalent linear constraint $\mathbf{u} \cdot \mathbf{n} > 0$. The interested reader can compare the above analysis based on GQL and the traditional analysis in [101].

5.2.3. High-Order Schemes. The GQL approach can also be used in conjunction with the limiter-based framework [122] to explore high-order bound-preserving schemes. Consider a $(K+1)$-order scheme (5.4) with the numerical flux

$$\hat{\mathbf{f}}_{j+\frac{1}{2}}^n = \hat{\mathbf{f}} \left(\mathbf{u}_{j+\frac{1}{2}}^-, \mathbf{u}_{j+\frac{1}{2}}^+ \right), \tag{5.17}$$

where $\hat{\mathbf{f}}(\cdot, \cdot)$ is a bound-preserving numerical flux, and we take the gas-kinetic flux (5.8) as the example. Here $\mathbf{u}_{j+\frac{1}{2}}^{\pm} := \lim_{\varepsilon \to 0^+} \mathbf{u}_h^n(x_{j+\frac{1}{2}} \pm \varepsilon)$, in which $\mathbf{u}_h^n(x)$ is a reconstructed piecewise polynomial vector of degree K as a $(K+1)$-order approximation to $\mathbf{u}(x,t_n)$. Note that the scheme (5.4) can also be regarded as the discrete equations satisfied by cell averages of the discontinuous Galerkin schemes [122], for which the following analysis is also applicable.

Suppose $\mathbf{u}^{\pm}_{j+\frac{1}{2}} \in G$ for all j. For $\mathbf{n} = \mathbf{e}_1$ or $\mathbf{n} = \mathbf{n}_*$, thanks to the linearity of $\mathbf{u}\cdot\mathbf{n}$, we have for the scheme (5.4) with (5.17) that

$$\begin{aligned}\bar{\mathbf{u}}^{n+1}_j \cdot \mathbf{n} &= \bar{\mathbf{u}}^n_j \cdot \mathbf{n} - \sigma\left(\mathbf{f}^+(\mathbf{u}^-_{j+\frac{1}{2}}) + \mathbf{f}^-(\mathbf{u}^+_{j+\frac{1}{2}})\right)\cdot\mathbf{n} + \sigma\left(\mathbf{f}^+(\mathbf{u}^-_{j-\frac{1}{2}}) + \mathbf{f}^-(\mathbf{u}^+_{j-\frac{1}{2}})\right)\cdot\mathbf{n} \\ &> \bar{\mathbf{u}}^n_j \cdot \mathbf{n} - \sigma\mathbf{f}^+(\mathbf{u}^-_{j+\frac{1}{2}})\cdot\mathbf{n} + \sigma\mathbf{f}^-(\mathbf{u}^+_{j-\frac{1}{2}})\cdot\mathbf{n}, \end{aligned} \tag{5.18}$$

where we have used (5.12)–(5.13). *As clearly revealed by the GQL approach, the preservation of region G essentially hinges on using the positive term $\bar{\mathbf{u}}^n_j\cdot\mathbf{n}$ to dominate the potentially negative terms in* (5.18)*, similar to the linear bound-preserving analysis in the scalar case.* Recall that for any $\mathbf{u}\in G$, the estimates (5.15)–(5.16) give $\mathbf{f}^+(\mathbf{u})\cdot\mathbf{n} - \mathbf{f}^-(\mathbf{u})\cdot\mathbf{n} < \alpha(\mathbf{u})\mathbf{u}\cdot\mathbf{n}$, which, along with $\pm\mathbf{f}^{\pm}(\mathbf{u})\cdot\mathbf{n} > 0$, leads to $\pm\mathbf{f}^{\pm}(\mathbf{u})\cdot\mathbf{n} < \alpha(\mathbf{u})\mathbf{u}\cdot\mathbf{n}$. Thus we have from (5.18) that

$$\bar{\mathbf{u}}^{n+1}_j \cdot \mathbf{n} \geq \bar{\mathbf{u}}^n_j\cdot\mathbf{n} - \sigma\alpha_n\mathbf{u}^-_{j+\frac{1}{2}}\cdot\mathbf{n} - \sigma\alpha_n\mathbf{u}^+_{j-\frac{1}{2}}\cdot\mathbf{n} \tag{5.19}$$

with $\alpha_n := \max_j \alpha(\mathbf{u}^{\pm}_{j+\frac{1}{2}})$. As observed in [121], the cell average $\bar{\mathbf{u}}^n_j\cdot\mathbf{n}$ can be used to bound the values $\mathbf{u}^-_{j+\frac{1}{2}}\cdot\mathbf{n}$ and $\mathbf{u}^+_{j-\frac{1}{2}}\cdot\mathbf{n}$ at two endpoints inside the cell $[x_{j-1/2}, x_{j+1/2}]$, due to the exactness of L-point Gauss–Lobatto quadrature with $L = \lceil \frac{K+3}{2} \rceil$ for polynomials of degree k; namely,

$$\bar{\mathbf{u}}^n_j\cdot\mathbf{n} = \sum_{\mu=1}^{L} \widehat{\omega}_\mu \mathbf{u}^n_h(\widehat{x}^{(\mu)}_j)\cdot\mathbf{n} = \sum_{\mu=2}^{L-1} \widehat{\omega}_\mu \mathbf{u}^n_h(\widehat{x}^{(\mu)}_j)\cdot\mathbf{n} + \widehat{\omega}_1\left(\mathbf{u}^+_{j-\frac{1}{2}}\cdot\mathbf{n} + \mathbf{u}^-_{j+\frac{1}{2}}\cdot\mathbf{n}\right), \tag{5.20}$$

where $\{\widehat{\omega}_\mu, x^{(\mu)}_j\}_{1\leq\mu\leq L}$ are the quadrature weights and nodes on $[x_{j-1/2}, x_{j+1/2}]$. If

$$\mathbf{u}^n_h(\widehat{x}^{(\mu)}_j) \in G \qquad \forall j, \mu, \tag{5.21}$$

then $\mathbf{u}^n_h(\widehat{x}^{(\mu)}_j)\cdot\mathbf{n} > 0$, and combining (5.19) with (5.20) gives $\bar{\mathbf{u}}^{n+1}_j\cdot\mathbf{n} > (\widehat{\omega}_1 - \sigma\alpha_n)(\mathbf{u}^+_{j-\frac{1}{2}}\cdot\mathbf{n} + \mathbf{u}^-_{j+\frac{1}{2}}\cdot\mathbf{n}) \geq 0$ under the CFL condition $\sigma\alpha_n \leq \widehat{\omega}_1$.

Remark 5.8. The above analysis indicates that the high-order scheme (5.4) with (5.17) is bound-preserving under the CFL condition $\sigma\alpha_n \leq \widehat{\omega}_1$, if (5.21) is satisfied. This conclusion is consistent with [122], while (5.21) can be enforced by the scaling limiter in [122]. The theoretical CFL condition $\sigma\alpha_n \leq \widehat{\omega}_1$ suggests that a CFL number not exceeding $\widehat{\omega}_1$ is sufficient for bound preservation. For the $(K+1)$th-order discontinuous Galerkin scheme, such a CFL number is comparable to the standard one $\mathtt{C_{std}} := \frac{1}{2K+1}$; for example, $\widehat{\omega}_1 = 1/6$ and $\mathtt{C_{std}} = 1/5$ for $K = 2$; $\widehat{\omega}_1 = 1/6$ and $\mathtt{C_{std}} = 1/7$ for $K = 3$. However, for some finite volume schemes, the theoretical CFL condition is stricter than the standard condition. A practical implementation is that if the cell averages evolved with the standard CFL go outside G, we restart the computation *from the last time step* with half of Δt and then proceed. The theoretical proof guarantees that one only needs to restart for at most a fixed number of times.

Remark 5.9. Based on the GQL approach, the preservation of nonlinear constraints is cast into a simple linear positivity-preserving problem as in the scalar case.

This is different from the standard analysis approach in [122, 123], which typically relies on decomposing high-order schemes into convex combinations of some bound-preserving subterms (this strategy may fail in some cases such as the MHD systems [106, 107, 108]). Since the linear feature of GQL has already naturally incorporated the convexity of G into the GQL representation, the technical role of convex decomposition is no longer essential.

5.2.4. A Bound-Preserving Scheme for Navier–Stokes System. Consider the scheme

$$\bar{\mathbf{u}}_j^{n+1} = \bar{\mathbf{u}}_j^n - \sigma\big(\hat{\mathbf{f}}_{j+\frac{1}{2}} - \hat{\mathbf{f}}_{j-\frac{1}{2}}\big) + \frac{\Delta t}{\Delta x^2}\frac{\eta}{\mathtt{Re}}\mathbf{H}_j \tag{5.22}$$

with $\mathbf{H}_j := \mathbf{r}(\bar{\mathbf{u}}_{j+1}^n) - 2\mathbf{r}(\bar{\mathbf{u}}_j^n) + \mathbf{r}(\bar{\mathbf{u}}_{j-1}^n)$ for solving the 1D Navier–Stokes equations (2.4); see [119]. Here $\hat{\mathbf{f}}_{j+1/2}$ is taken as a bound-preserving numerical flux for the Euler system (2.1), for example, the Lax–Friedrichs flux (5.6) or the gas-kinetic flux (5.8), which satisfy the following: if $\bar{\mathbf{u}}_j^n \in G$ for all j, then

$$\big(\hat{\mathbf{f}}_{j+\frac{1}{2}} - \hat{\mathbf{f}}_{j-\frac{1}{2}}\big)\cdot\mathbf{n} < \alpha_n \bar{\mathbf{u}}_j^n\cdot\mathbf{n} \qquad \forall j$$

holds for $\mathbf{n} = \mathbf{e}_1$ and $\mathbf{n} = \mathbf{n}_*$, according to the analysis in subsections 5.2.1 and 5.2.2. Thus we have

$$\bar{\mathbf{u}}_j^{n+1}\cdot\mathbf{n} > (1-\sigma\alpha_n)\bar{\mathbf{u}}_j^n\cdot\mathbf{n} + \frac{\Delta t}{\Delta x^2}\frac{\eta}{\mathtt{Re}}\mathbf{H}_j\cdot\mathbf{n}. \tag{5.23}$$

Thanks to GQL, we clearly see that the bound-preserving essence is to control the potentially negative term $\mathbf{H}_j\cdot\mathbf{n}$ by the positive term $\bar{\mathbf{u}}_j^n\cdot\mathbf{n}$. Note that $\mathbf{H}_j\cdot\mathbf{e}_1 = 0$, and so $\bar{\mathbf{u}}_j^{n+1}\cdot\mathbf{e}_1 > (1-\sigma\alpha_n)\bar{\mathbf{u}}_j^n\cdot\mathbf{e}_1 \ge 0$ if $\sigma\alpha_n \le 1$. For any $\mathbf{u}\in G$ and $v_*\in\mathbb{R}$, we have

$$-\frac{v_*^2}{2} < \mathbf{r}(\mathbf{u})\cdot\mathbf{n}_* = \frac{1}{2}(v-v_*)^2 + \frac{\Gamma}{\mathtt{Pr}\ \eta}e - \frac{v_*^2}{2} \le \max\left\{1, \frac{\Gamma}{\mathtt{Pr}\ \eta}\right\}\frac{1}{\rho}(\mathbf{u}\cdot\mathbf{n}_*) - \frac{v_*^2}{2}.$$

This gives

$$\begin{aligned}
\mathbf{H}_j\cdot\mathbf{n}_* &= \Big(\mathbf{r}(\bar{\mathbf{u}}_{j+1}^n)\cdot\mathbf{n}_* + \mathbf{r}(\bar{\mathbf{u}}_{j-1}^n)\cdot\mathbf{n}_*\Big) - 2\mathbf{r}(\bar{\mathbf{u}}_j^n)\cdot\mathbf{n}_* \\
&\ge \left(-\frac{v_*^2}{2} - \frac{v_*^2}{2}\right) - 2\left(\max\left\{1, \frac{\Gamma}{\mathtt{Pr}\ \eta}\right\}\frac{1}{\bar{\rho}_j^n}(\bar{\mathbf{u}}_j^n\cdot\mathbf{n}_*) - \frac{v_*^2}{2}\right) \\
&= -\frac{2}{\bar{\rho}_j^n}\max\left\{1, \frac{\Gamma}{\mathtt{Pr}\ \eta}\right\}(\bar{\mathbf{u}}_j^n\cdot\mathbf{n}_*).
\end{aligned}$$

It then follows from (5.23) that

$$\bar{\mathbf{u}}_j^{n+1}\cdot\mathbf{n}_* > (1-\sigma\alpha_n)\bar{\mathbf{u}}_j^n\cdot\mathbf{n}_* - \frac{\Delta t}{\Delta x^2}\frac{\eta}{\mathtt{Re}}\frac{2}{\bar{\rho}_j^n}\max\left\{1, \frac{\Gamma}{\mathtt{Pr}\ \eta}\right\}(\bar{\mathbf{u}}_j^n\cdot\mathbf{n}_*).$$

We then immediately have $\bar{\mathbf{u}}_j^{n+1}\cdot\mathbf{n}_* > 0$, provided that

$$\alpha_n\frac{\Delta t}{\Delta x} + \frac{\Delta t}{\Delta x^2}\frac{2}{\bar{\rho}_j^n\mathtt{Re}}\max\left\{\eta, \frac{\Gamma}{\mathtt{Pr}}\right\} \le 1. \tag{5.24}$$

In conclusion, the scheme (5.22) is bound-preserving under condition (5.24).

Remark 5.10. A standard approach for handling bound-preserving problems with multiple terms (e.g., convection term and diffusion term [119], or convection term and source term [123]) is based on decomposing the schemes into a convex combination of some subterms and then enforcing all the subterms in G. This may lead to stricter conditions on the time step-size Δt.

6. More Examples of GQL.

This section provides more examples on GQL representations of invariant regions and further illustrates the advantages of GQL for bound-preserving study. The application of GQL to designing high-order bound-preserving schemes will also be explored in section 7 for the multicomponent MHD system, to further demonstrate the capability of GQL in addressing challenging bound-preserving problems that could not be handled by direct approaches.

6.1. Example 1: Ideal MHD System.

THEOREM 6.1. *For the ideal MHD system* (2.12)*, the GQL representation of the invariant region G in* (2.13) *is given by*

$$G_* = \left\{ \mathbf{u} = (\rho, \boldsymbol{m}, \mathbf{B}, E)^\top \in \mathbb{R}^8 : \ \rho > 0, \ \varphi(\mathbf{u}; \boldsymbol{v}_*, \mathbf{B}_*) > 0 \quad \forall \boldsymbol{v}_*, \mathbf{B}_* \in \mathbb{R}^3 \right\} \tag{6.1}$$

with $\varphi(\mathbf{u}; \boldsymbol{v}_*, \mathbf{B}_*) := \mathbf{u} \cdot \mathbf{n}_* + \frac{\|\mathbf{B}_*\|^2}{2}$ *and* $\mathbf{n}_* := \left(\frac{\|\boldsymbol{v}_*\|^2}{2}, -\boldsymbol{v}_*, -\mathbf{B}_*, 1 \right)^\top$.

Proof. The representation (6.1) was first established in [104] by technical algebraic manipulations. We illustrate a different derivation using the gradient-based method and Theorem 4.1. For the nonlinear constraint in (2.13), the gradient of $g(\mathbf{u})$ is $\nabla g(\mathbf{u}) = \left(\frac{\|\boldsymbol{m}\|^2}{2\rho^2}, -\frac{\boldsymbol{m}}{\rho}, -\mathbf{B}, 1 \right)^\top$, and the corresponding boundary hypersurface is $\mathcal{S} := \{ \mathbf{u}_* = (\rho_*, \boldsymbol{m}_*, \mathbf{B}_*, E_*)^\top : \rho_* > 0, g(\mathbf{u}_*) = 0 \}$. Based on the equivalence of $g(\mathbf{u}) = 0$ and $p = 0$, we obtain a natural physics-based parametrization of $\mathcal{S}$:

$$\mathcal{S} = \left\{ \mathbf{u}_* = \left(\rho_*, \rho_* \boldsymbol{v}_*, \mathbf{B}_*, \frac{1}{2} \left(\rho_* \|\boldsymbol{v}_*\|^2 + \|\mathbf{B}_*\|^2 \right) \right)^\top : \rho_* > 0, \ \boldsymbol{v}_* \in \mathbb{R}^3, \ \mathbf{B}_* \in \mathbb{R}^3 \right\}.$$

For $\mathbf{u}_* \in \mathcal{S}$ and $\mathbf{u} = (\rho, \boldsymbol{m}, \mathbf{B}, E)^\top$, we have $(\mathbf{u} - \mathbf{u}_*) \cdot \nabla g(\mathbf{u}_*) = \varphi(\mathbf{u}; \boldsymbol{v}_*, \mathbf{B}_*)$. By Theorem 4.1, we obtain the GQL representation (6.1). □

Remark 6.2 (bound-preserving applications). The GQL approach with (6.1) opened a door to studying provably bound-preserving schemes for MHD [104, 106, 107]. Most notably, it led to the discovery of the intrinsic relations between the bound-preserving and magnetic divergence-free properties, which had been unclear for a long time [104]. In fact, before the works [104, 106, 107] motivated by GQL, there was no rigorous proof for the bound-preserving property of any (even first-order) scheme for multidimensional MHD, as mentioned in [21, page A1835].

6.2. Example 2: Relativistic Hydrodynamic System.

THEOREM 6.3. *For the* 1*D RHD system* (2.6)*, the GQL representation of the invariant region G in* (2.7) *is given by*

$$G_* = \left\{ \mathbf{u} = (D, m, E)^\top : D > 0, \ \varphi(\mathbf{u}; v_*) > 0 \quad \forall v_* \in (-1, 1) \right\}, \tag{6.2}$$

with $\varphi(\mathbf{u}; v_*) := E - m v_* - D\sqrt{1 - v_*^2}$ *being a linear function of* $\mathbf{u}$.

Proof. The first constraint in (2.7) is linear. We deal with the second one using the constructive method. The Cauchy–Schwarz inequality implies

$$\varphi(\mathbf{u};v_*) \geq E - \sqrt{D^2+m^2}\sqrt{v_*^2 + \left(\sqrt{1-v_*^2}\right)^2} = g(\mathbf{u}),$$

where equality holds if $v_* = m/\sqrt{D^2+m^2}$, namely, $\min_{v_*\in(-1,1)} \varphi(\mathbf{u};v_*) = g(\mathbf{u})$. According to Theorem 3.2, we get the GQL representation (6.2). □

We now utilize the cross-product method to construct the GQL representation of the invariant region $\widetilde{G}$ in (2.8), where the minimum entropy principle is also included as a constraint.

THEOREM 6.4. *For the* 1*D RHD system* (2.6)*, the GQL representation of the invariant region* $\widetilde{G}$ *in* (2.8) *is given by*

$$\widetilde{G}_* = \left\{\mathbf{u} = (D,m,E)^\top :\ \rho > 0,\ \ \widetilde{\varphi}(\mathbf{u};\rho_*,v_*) \geq 0 \ \ \forall \rho_* \in \mathbb{R}^+, \ \ \forall v_* \in (-1,1)\right\}, \tag{6.3}$$

with $\widetilde{\varphi}(\mathbf{u};\rho_*,v_*) := \mathbf{u}\cdot\mathbf{n}_* + S_{min}\rho_*^\Gamma$ *and* $\mathbf{n}_* := \left(-\sqrt{1-v_*^2}\left(1+\frac{S_{min}\Gamma\rho_*^{\Gamma-1}}{\Gamma-1}\right), -v_*, 1\right)^\top$.

Proof. We only need to tackle the second and third constraints in (2.8). For the third one, $\widetilde{g}(\mathbf{u}) \geq 0$, the corresponding boundary hypersurface is $\widetilde{\mathcal{S}} := \{\mathbf{u}_* = (\rho_*, m_*, E_*) : \rho_* > 0,\ g(\mathbf{u}_*) > 0,\ \widetilde{g}(\mathbf{u}_*) = 0\}$. Based on the equivalence of $\widetilde{g}(\mathbf{u}) = 0$ and $p = S_{min}\rho^\Gamma$, we obtain a natural physics-based parametrization of $\widetilde{\mathcal{S}}$, namely,

$$\widetilde{\mathcal{S}} = \left\{ \mathbf{u}_* = \left(\frac{\rho_*}{\sqrt{1-v_*^2}}, \frac{\left(\rho_* + \frac{S_{min}\Gamma\rho_*^\Gamma}{\Gamma-1}\right)v_*}{1-v_*^2}, \frac{\rho_* + \frac{S_{min}\Gamma\rho_*^\Gamma}{\Gamma-1}}{1-v_*^2} - S_{min}\rho_*^\Gamma \right)^\top : \rho_* > 0,\ v_* \in (-1,1) \right\}.$$

We can then derive the normal vector $\mathbf{n}_*$ of $\mathcal{S}$ at $\mathbf{u}_*$ by cross product $\frac{\partial \mathbf{u}_*}{\partial \rho_*} \times \frac{\partial \mathbf{u}_*}{\partial v_*}$ with $\delta_* := (1-v_*)^{5/2}\left(\rho_* + \frac{S_{min}\Gamma}{\Gamma-1}\rho_*^\Gamma(1+v_*^2-\Gamma v_*^2)\right)^{-1}$. By Theorem 4.2 and $(\mathbf{u}-\mathbf{u}_*)\cdot\mathbf{n}_* = \widetilde{\varphi}(\mathbf{u};\rho_*,v_*)$, the GQL representation for $\widetilde{g}(\mathbf{u}) \geq 0$ is

$$\widetilde{\varphi}(\mathbf{u};\rho_*,v_*) \geq 0 \qquad \forall \rho_* \in \mathbb{R}^+, \quad \forall v_* \in \mathbb{R}. \tag{6.4}$$

Note that $S_{min} > 0$ and

$$g(\mathbf{u}) > g(\mathbf{u}) - \frac{S_{min}}{\Gamma-1}\left(\frac{D^2}{\sqrt{D^2+m^2}}\right)^\Gamma = \widetilde{\varphi}\left(\mathbf{u};\frac{D^2}{\sqrt{D^2+m^2}},\frac{m}{\sqrt{D^2+m^2}}\right),$$

which means that (6.4) also implies $g(\mathbf{u}) > 0$ in (2.8). That is, the second and third constraints in (2.8) can be equivalently represented by (6.4). We therefore obtain the GQL representation (6.3). □

Remark 6.5 (bound-preserving application). The GQL approach with (6.3) played a critical role in developing provably invariant-region-preserving schemes for RHD in [105].

6.3. Example 3: Relativistic MHD System.

THEOREM 6.6. *For the relativistic MHD system* (2.14)*, the GQL representation of the invariant region* G *in* (2.15) *is given by*

$$G_* = \left\{\mathbf{u} \in \mathbb{R}^8 :\ D > 0,\ \varphi(\mathbf{u};\boldsymbol{v}_*,\mathbf{B}_*) > 0 \ \ \forall \mathbf{B}_* \in \mathbb{R}^3, \ \ \forall \boldsymbol{v}_* \in \mathbb{B}_1(\mathbf{0})\right\}, \tag{6.5}$$

where $\mathbf{u} = (D, \boldsymbol{m}, \mathbf{B}, E)^\top$, $\mathbb{B}_1(\mathbf{0}) := \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\| \leq 1\}$ *is a unit* 3*D ball, and the linear function* $\varphi(\mathbf{u}; \boldsymbol{v}_*, \mathbf{B}_*) := \mathbf{u} \cdot \mathbf{n}_* + p_m^*$ *with*

$$p_m^* := \frac{1}{2}\left((1 - \|\boldsymbol{v}_*\|^2)\|\mathbf{B}_*\|^2 + (\boldsymbol{v}_* \cdot \mathbf{B}_*)^2\right), \tag{6.6}$$

$$\mathbf{n}_* := \left(-\sqrt{1 - \|\boldsymbol{v}_*\|^2},\ -\boldsymbol{v}_*,\ -(1 - \|\boldsymbol{v}_*\|^2)\mathbf{B}_* - (\boldsymbol{v}_* \cdot \mathbf{B}_*)\boldsymbol{v}_*,\ 1\right)^\top. \tag{6.7}$$

Note that p_m^* *and* $\mathbf{n}_*$ *only depend on the free auxiliary variables* $(\boldsymbol{v}_*, \mathbf{B}_*)$.

Proof. As shown in [110], the region G in (2.15) can be equivalently represented as

$$G = \left\{\mathbf{u} \in \mathbb{R}^8 :\ D > 0,\ g_2(\mathbf{u}) > 0,\ p(\mathbf{u}) > 0\right\}, \tag{6.8}$$

with $g_2(\mathbf{u}) := E - \sqrt{D^2 + \|\boldsymbol{m}\|^2}$. Although the implicit function $p(\mathbf{u})$ defined in (1.6) cannot be explicitly formulated, the corresponding boundary hypersurface $\mathcal{S} := \{\mathbf{u}_* = (D_*, \boldsymbol{m}_*, \mathbf{B}_*, E_*)^\top : D_* > 0, g_2(\mathbf{u}_*) > 0, p(\mathbf{u}_*) = 0\}$ has an explicit physics-based parametrization

$$\mathcal{S} = \Big\{\mathbf{u}_* = \Big(\rho_*\gamma_*,\ \rho_*\gamma_*^2\boldsymbol{v}_* + \|\mathbf{B}_*\|^2\boldsymbol{v}_* - (\boldsymbol{v}_* \cdot \mathbf{B}_*)\mathbf{B}_*,\ \mathbf{B}_*,$$
$$\rho_*\gamma_*^2 + \|\mathbf{B}_*\|^2 - p_m^*\Big)^\top : \rho_* > 0,\ \mathbf{B}_* \in \mathbb{R}^3,\ \boldsymbol{v}_* \in \mathbb{B}_1(\mathbf{0})\Big\}$$

with p_m^* defined in (6.6) and $\gamma_* := (1 - \|\boldsymbol{v}_*\|^2)^{\frac{1}{2}}$. This parametrization is helpful for dealing with the highly nonlinear constraint $p(\mathbf{u}) > 0$ by the cross-product method. For $1 \leq i \leq 3$, denote $\mathbf{e}_i := (\delta_{1i}, \delta_{2i}, \delta_{3i})$ with δ_{ij} the Kronecker delta. Taking the partial derivatives of $\mathbf{u}_*$ with respect to the parametric variables $\{\rho_*, \boldsymbol{v}_*, \mathbf{B}_*\}$, we can obtain the expressions (omitted here) of $\frac{\partial \mathbf{u}_*}{\partial \rho_*}$, $\frac{\partial \mathbf{u}_*}{\partial v_{i*}}$, and $\frac{\partial \mathbf{u}_*}{\partial B_{i*}}$, which are all perpendicular to the nonzero vector $\mathbf{n}_*$ defined in (6.7). This means $\mathbf{n}_*$ is parallel to the cross product $\frac{\partial \mathbf{u}_*}{\partial \rho_*} \times \big(\bigwedge_{i=1}^{3} \frac{\partial \mathbf{u}_*}{\partial v_{i*}}\big) \times \big(\bigwedge_{i=1}^{3} \frac{\partial \mathbf{u}_*}{\partial B_{i*}}\big)$, implying that $\mathbf{n}_*$ is a normal vector of $\mathcal{S}$ at $\mathbf{u}_*$. It can be verified that $\mathbf{n}_*$ is always directed toward the concave side of $\mathcal{S}$. By Theorem 4.2 and $(\mathbf{u} - \mathbf{u}_*) \cdot \mathbf{n}_* = \varphi(\mathbf{u}; \boldsymbol{v}_*, \mathbf{B}_*)$, we know that the GQL representation for $p(\mathbf{u}) > 0$ is

$$\varphi(\mathbf{u}; \boldsymbol{v}_*, \mathbf{B}_*) > 0 \quad \forall \mathbf{B}_* \in \mathbb{R}^3, \quad \forall \boldsymbol{v}_* \in \mathbb{B}_1(\mathbf{0}). \tag{6.9}$$

By taking $\boldsymbol{v}_* = \boldsymbol{m}/\sqrt{D^2 + \|\boldsymbol{m}\|^2}$ and $\mathbf{B}_* = \mathbf{0}$, we obtain $\varphi(\mathbf{u}; \boldsymbol{v}_*, \mathbf{B}_*) = g_2(\mathbf{u})$, which means that (6.9) also implies $g_2(\mathbf{u}) > 0$ in (6.8). In other words, the second and third constraints in (6.8) can be equivalently represented by (6.9). Therefore, we obtain the GQL representation (6.5). □

Remark 6.7 (bound-preserving application). In the original form (2.15) of G, two constraints are *highly nonlinear and cannot be explicitly* formulated, making the bound-preserving study *extremely difficult*. However, all the constraints in the GQL representation (6.5) are *explicit and linear*, greatly facilitating the analysis and design of bound-preserving relativistic MHD schemes; see [108].

6.4. Example 4: Ten-Moment Gaussian Closure System.

THEOREM 6.8. *For the* 2*D ten-moment Gaussian closure system* (2.9)*, the GQL representation of the invariant region* G *in* (2.11) *is given by*

$$G_* = \Big\{\mathbf{u} \in \mathbb{R}^6 : \mathbf{u} \cdot \mathbf{e}_1 > 0,\ \varphi(\mathbf{u}; \boldsymbol{z}, \boldsymbol{v}_*) > 0\ \ \forall \boldsymbol{v}_* \in \mathbb{R}^2, \quad \forall \boldsymbol{z} \in \mathbb{R}^2 \setminus \{\mathbf{0}\}\Big\}, \tag{6.10}$$

where $\mathbf{e}_1 := (1, 0, \dots, 0)^\top$, $\mathbf{u} := (\rho, \boldsymbol{m}, E_{11}, E_{12}, E_{22})^\top$, *and the function* $\varphi(\mathbf{u}; \boldsymbol{z}, \boldsymbol{v}_*)$ *is linear with respect to* $\mathbf{u}$*:*

$$\varphi(\mathbf{u}; \boldsymbol{z}, \boldsymbol{v}_*) := \boldsymbol{z}^\top \left(\mathbf{E} - \boldsymbol{m} \otimes \boldsymbol{v}_* + \rho \frac{\boldsymbol{v}_* \otimes \boldsymbol{v}_*}{2} \right) \boldsymbol{z}. \tag{6.11}$$

Proof. We only need to deal with the nonlinear constraint in (2.11). Note that

$$\varphi(\mathbf{u}; \boldsymbol{z}, \boldsymbol{v}_*) = \boldsymbol{z}^\top \left(\mathbf{E} - \frac{\boldsymbol{m} \otimes \boldsymbol{m}}{2\rho} \right) \boldsymbol{z} + \frac{\rho}{2} \left| \boldsymbol{z} \cdot \left(\boldsymbol{v}_* - \frac{\boldsymbol{m}}{\rho} \right) \right|^2,$$

which implies $\min_{\boldsymbol{v}_* \in \mathbb{R}^2} \varphi(\mathbf{u}; \boldsymbol{z}, \boldsymbol{v}_*) = \boldsymbol{z}^\top \big(\mathbf{E} - \frac{\boldsymbol{m} \otimes \boldsymbol{m}}{2\rho} \big) \boldsymbol{z}$. By Theorem 3.2, we obtain (6.10). □

Remark 6.9 (bound-preserving application). In the original form (2.10) of G, the second constraint is the positive-definiteness of a matrix $\mathbf{E} - \frac{\boldsymbol{m} \otimes \boldsymbol{m}}{2\rho}$, which depends nonlinearly on $\mathbf{u}$. This leads to the challenges in the bound-preserving study. Thanks to Theorem 6.8, the invariant region G is equivalently represented as (6.10) with only linear constraints.

To illustrate the advantages of the GQL approach, we use it to investigate the bound-preserving property of the scheme

$$\bar{\mathbf{u}}_{ij}^{n+1} = \bar{\mathbf{u}}_{ij}^{n} - \sigma_1 \big(\hat{\mathbf{f}}_{1, i+\frac{1}{2}, j} - \hat{\mathbf{f}}_{1, i-\frac{1}{2}, j} \big) - \sigma_2 \big(\hat{\mathbf{f}}_{2, i, j+\frac{1}{2}} - \hat{\mathbf{f}}_{2, i, j-\frac{1}{2}} \big) \tag{6.12}$$

for solving the Gaussian closure equations (2.9) on a uniform Cartesian mesh $\{[x_{i-1/2}, x_{i+1/2}] \times [y_{j-1/2}, y_{j+1/2}]\}$, with $\sigma_1 = \frac{\Delta t}{\Delta x}$, $\sigma_2 = \frac{\Delta t}{\Delta y}$. Here $\bar{\mathbf{u}}_{ij}^n$ denotes an approximation to the average of $\mathbf{u}(x, y, t_n)$ on each cell, and the Lax–Friedrichs numerical fluxes are considered, i.e.,

$$\hat{\mathbf{f}}_{1, i+1/2, j} = \hat{\mathbf{f}}_1^{\rm LF}(\bar{\mathbf{u}}_{ij}^n, \bar{\mathbf{u}}_{i+1, j}^n), \quad \hat{\mathbf{f}}_{2, i, j+1/2} = \hat{\mathbf{f}}_2^{\rm LF}(\bar{\mathbf{u}}_{ij}^n, \bar{\mathbf{u}}_{i, j+1}^n), \tag{6.13}$$

$$\hat{\mathbf{f}}_\ell^{\rm LF}(\mathbf{u}^L, \mathbf{u}^R) := \frac{1}{2} \Big(\mathbf{f}_\ell(\mathbf{u}^L) + \mathbf{f}_\ell(\mathbf{u}^R) - \alpha_{\ell, n} (\mathbf{u}^R - \mathbf{u}^L) \Big), \quad \ell = 1, 2, \tag{6.14}$$

where $\alpha_{\ell, n} = \max_{ij} \alpha_\ell(\bar{\mathbf{u}}_{ij}^n)$ and $\alpha_\ell(\mathbf{u}) := |v_\ell| + \sqrt{p_{\ell\ell}/\rho}$.

For system (2.9) and any $\mathbf{u} \in G$, we have $\mathbf{f}_\ell(\mathbf{u}) \cdot \mathbf{e}_1 = v_\ell (\mathbf{u} \cdot \mathbf{e}_1)$ and

$$\pm \mathbf{f}_\ell(\mathbf{u}) \cdot \mathbf{e}_1 \le |v_\ell| (\mathbf{u} \cdot \mathbf{e}_1) < \alpha_\ell(\mathbf{u}) (\mathbf{u} \cdot \mathbf{e}_1),$$

which gives $\bar{\mathbf{u}}_{ij}^{n+1} \cdot \mathbf{e}_1 > 0$ under the CFL condition $\sigma_1 \alpha_{1,n} + \sigma_2 \alpha_{2,n} < 1$. In the following, we focus on the second constraint in (6.10). Thanks to the linearity of $\varphi(\cdot; \boldsymbol{z}, \boldsymbol{v}_*)$, we obtain

$$\varphi(\mathbf{f}_1(\mathbf{u}); \boldsymbol{z}, \boldsymbol{v}_*) = v_1 \varphi(\mathbf{u}; \boldsymbol{z}, \boldsymbol{v}_*) + [\boldsymbol{z} \cdot (\boldsymbol{v} - \boldsymbol{v}_*)] (\boldsymbol{p}_1 \cdot \boldsymbol{z}) \tag{6.15}$$

with the vector $\boldsymbol{p}_1 := (p_{11}, p_{12})^\top$. For any $\mathbf{u} \in G$, using the AM-GM inequality gives

$$\begin{aligned} \Big| [\boldsymbol{z} \cdot (\boldsymbol{v} - \boldsymbol{v}_*)] (\boldsymbol{p}_1 \cdot \boldsymbol{z}) \Big| &\le \frac{1}{2} \sqrt{\rho p_{11}} \, |\boldsymbol{z} \cdot (\boldsymbol{v} - \boldsymbol{v}_*)|^2 + \frac{1}{2\sqrt{\rho p_{11}}} \, |\boldsymbol{p}_1 \cdot \boldsymbol{z}|^2 \\ &= \sqrt{\frac{p_{11}}{\rho}} \varphi(\mathbf{u}; \boldsymbol{z}, \boldsymbol{v}_*) - \frac{z_2^2 \det(\mathbf{p})}{2\sqrt{\rho p_{11}}} \le \sqrt{\frac{p_{11}}{\rho}} \varphi(\mathbf{u}; \boldsymbol{z}, \boldsymbol{v}_*), \end{aligned}$$

which together with the identity (6.15) yields

$$\pm \varphi(\mathbf{f}_1(\mathbf{u}); \boldsymbol{z}, \boldsymbol{v}_*) \le \Big(|v_1| + \sqrt{p_{11}/\rho} \Big) \varphi(\mathbf{u}; \boldsymbol{z}, \boldsymbol{v}_*) = \alpha_1(\mathbf{u}) \varphi(\mathbf{u}; \boldsymbol{z}, \boldsymbol{v}_*). \tag{6.16}$$

Using the linearity of $\varphi(\cdot;\boldsymbol{z},\boldsymbol{v}_*)$ again and (6.16), we obtain

$$\begin{aligned}\varphi\left(\hat{\mathbf{f}}_{1,i+\frac{1}{2},j}-\hat{\mathbf{f}}_{1,i-\frac{1}{2},j};\boldsymbol{z},\boldsymbol{v}_*\right)&=\frac{1}{2}\left[\varphi(\mathbf{f}_1(\bar{\mathbf{u}}^n_{i+1,j});\boldsymbol{z},\boldsymbol{v}_*)-\alpha_{1,n}\varphi(\bar{\mathbf{u}}^n_{i+1,j};\boldsymbol{z},\boldsymbol{v}_*)\right]\\&\quad+\frac{1}{2}\left[-\varphi(\mathbf{f}_1(\bar{\mathbf{u}}^n_{i-1,j});\boldsymbol{z},\boldsymbol{v}_*)-\alpha_{1,n}\varphi(\bar{\mathbf{u}}^n_{i-1,j};\boldsymbol{z},\boldsymbol{v}_*)\right]\\&\quad+\alpha_{1,n}\varphi(\bar{\mathbf{u}}^n_{ij};\boldsymbol{z},\boldsymbol{v}_*)\leq\alpha_{1,n}\varphi(\bar{\mathbf{u}}^n_{ij};\boldsymbol{z},\boldsymbol{v}_*).\end{aligned}$$

Similarly, we have $\varphi\big(\hat{\mathbf{f}}_{2,i,j+\frac{1}{2}}-\hat{\mathbf{f}}_{2,i,j-\frac{1}{2}};\boldsymbol{z},\boldsymbol{v}_*\big)\leq\alpha_{2,n}\varphi(\bar{\mathbf{u}}^n_{ij};\boldsymbol{z},\boldsymbol{v}_*)$. It then follows that

$$\varphi\left(\bar{\mathbf{u}}^{n+1}_{ij};\boldsymbol{z},\boldsymbol{v}_*\right)\geq(1-\sigma_1\alpha_{1,n}-\sigma_2\alpha_{2,n})\,\varphi\left(\bar{\mathbf{u}}^n_{ij};\boldsymbol{z},\boldsymbol{v}_*\right)>0 \tag{6.17}$$

under the CFL condition $\sigma_1\alpha_{1,n}+\sigma_2\alpha_{2,n}<1$. This, along with $\bar{\mathbf{u}}^{n+1}_{ij}\cdot\mathbf{e}_1>0$, implies $\bar{\mathbf{u}}^{n+1}_{ij}\in G_*=G$ and the bound-preserving property of the scheme (6.12) with (6.13).

7. GQL for Design of Bound-Preserving Schemes for Multicomponent MHD.

This section applies the GQL approach to developing bound-preserving high-order finite volume and discontinuous Galerkin schemes for the multicomponent MHD system. We mainly focus on the 2D case, while our discussions are extensible to the 3D case. The 2D multicomponent compressible MHD system for an ideal fluid mixture with N_c components can be written as

$$\partial_t\mathbf{u}+\partial_x\mathbf{f}_1(\mathbf{u})+\partial_y\mathbf{f}_2(\mathbf{u})=\mathbf{0}, \tag{7.1a}$$

$$\mathbf{u}=\begin{pmatrix}\rho\mathbf{Y}\\\rho\\\boldsymbol{m}\\\mathbf{B}\\E\end{pmatrix},\quad \mathbf{f}_\ell(\mathbf{u})=\begin{pmatrix}\rho\mathbf{Y}v_\ell\\\rho v_\ell\\\boldsymbol{m}v_\ell-\mathbf{B}B_\ell+p_{tot}\mathbf{e}_\ell\\\mathbf{B}v_\ell-\boldsymbol{v}B_\ell\\v_\ell(E+p_{tot})-B_\ell(\boldsymbol{v}\cdot\mathbf{B})\end{pmatrix},\quad \ell=1,2, \tag{7.1b}$$

along with the extra divergence-free condition on the magnetic field $\mathbf{B}$:

$$\nabla\cdot\mathbf{B}:=\partial_xB_1+\partial_yB_2=0. \tag{7.2}$$

In (7.1b), ρ denotes the total density, $\boldsymbol{m}=\rho\boldsymbol{v}$ is the momentum with $\boldsymbol{v}$ the fluid velocity, $\mathbf{Y}=(Y_1,\dots,Y_{n_c-1})^\top$ denotes the mass fractions of the first (n_c-1) components, the mass fraction of the n_cth component is $Y_{n_c}:=1-\sum_{k=1}^{n_c-1}Y_k$, and $p_{tot}=p+\frac{\|\mathbf{B}\|^2}{2}$ is the total pressure, with the thermal pressure p calculated by

$$p=(\Gamma(\mathbf{u})-1)\left(E-\frac{\|\boldsymbol{m}\|^2}{2\rho}-\frac{\|\mathbf{B}\|^2}{2}\right),\qquad\Gamma(\mathbf{u}):=\frac{\sum_{k=1}^{n_c}\Gamma_kC_{v_k}Y_k}{\sum_{k=1}^{n_c}C_{v_k}Y_k}, \tag{7.3}$$

where $C_{v_k}>0$ and $\Gamma_k>1$, respectively, denote the heat capacity at constant volume and the ratio of specific heats for species k.

7.1. GQL Representation of Invariant Region.

For the system (7.1), the total density ρ and the thermal pressure p are both positive, and the mass fractions $\{Y_k\}_{k=1}^{n_c}$ are between 0 and 1. These constraints constitute the invariant region

$$G=\left\{\mathbf{u}\in\mathbb{R}^{n_c+7}:\ 0\leq Y_k\leq1,\ 1\leq k\leq n_c,\ \rho>0,\ p(\mathbf{u})>0\right\}, \tag{7.4}$$

where $p(\mathbf{u})$ is a highly nonlinear function defined by (7.3). Due to the strong nonlinearity and the underlying connections between the bound-preserving and divergence-free properties, the design and analysis of bound-preserving schemes for system (7.1) are highly challenging.

Following the GQL framework, the convex region G in (7.4) can be equivalently represented as

$$G_* = \left\{ \mathbf{u} \in \mathbb{R}^{n_c+7} : \mathbf{u}\cdot\mathbf{e}_k \geq 0, 0 \leq k < n_c,\ \mathbf{u}\cdot\mathbf{e}_{n_c} > 0,\ \varphi(\mathbf{u};\boldsymbol{v}_*,\mathbf{B}_*) > 0\ \forall \boldsymbol{v}_*,\mathbf{B}_* \in \mathbb{R}^3 \right\}, \tag{7.5}$$

where $\mathbf{e}_0 := \mathbf{e}_{n_c} - \sum_{k=1}^{n_c-1}\mathbf{e}_k$, the vector $\mathbf{e}_k$ for $k \geq 1$ has a 1 in the kth component and zeros elsewhere, and $\varphi(\mathbf{u};\boldsymbol{v}_*,\mathbf{B}_*) := \mathbf{u}\cdot\mathbf{n}_* + \frac{\|\mathbf{B}_*\|^2}{2}$ with $\mathbf{n}_* = (\mathbf{0}_{n_c-1}, \frac{\|\boldsymbol{v}_*\|^2}{2}, -\boldsymbol{v}_*, -\mathbf{B}_*, 1)^\top$. In what follows, we will derive bound-preserving schemes for (7.1) based on the GQL representation (7.5). *The GQL approach will not only help overcome the difficulties arising from the nonlinearity, but it will also play a crucial role in establishing the key relations between the bound-preserving property and a discrete divergence-free (DDF) condition on the numerical magnetic field.*

7.2. GQL Bridges Bound-Preserving Property and DDF Condition.

We focus on the Euler forward method for time discretization, while all our discussions are directly extensible to high-order strong-stability-preserving time discretizations [38] that are formally convex combinations of Euler forward. Consider the finite volume methods and the scheme of the cell averages of the discontinuous Galerkin method, which can be written in a unified form as

$$\bar{\mathbf{u}}_{ij}^{n+1} = \bar{\mathbf{u}}_{ij}^{n} - \sigma_1\big(\hat{\mathbf{f}}_{1,i+\frac{1}{2},j} - \hat{\mathbf{f}}_{1,i-\frac{1}{2},j}\big) - \sigma_2\big(\hat{\mathbf{f}}_{2,i,j+\frac{1}{2}} - \hat{\mathbf{f}}_{2,i,j-\frac{1}{2}}\big) \tag{7.6}$$

for solving (7.1) on a uniform Cartesian mesh $\{\mathcal{I}_{ij} := [x_{i-1/2}, x_{i+1/2}]\times[y_{j-1/2}, y_{j+1/2}]\}$, with $\sigma_1 = \frac{\Delta t}{\Delta x}$ and $\sigma_2 = \frac{\Delta t}{\Delta y}$. Here $\bar{\mathbf{u}}_{ij}^n$ denotes the approximate cell average of $\mathbf{u}(x,y,t_n)$ on $\mathcal{I}_{ij}$. For a $(K+1)$th-order accurate scheme, in each cell $\mathcal{I}_{ij}$ a polynomial vector of degree K, denoted by $\mathbf{U}_{ij}^n(x,y)$, is also constructed as the approximate solution, which is either the reconstructed polynomial solution in a finite volume scheme or the discontinuous Galerkin polynomial solution. Denote $\{\omega_q, x_i^{(q)}\}_{q=1}^Q$ and $\{\omega_q, y_j^{(q)}\}_{q=1}^Q$ as the Gauss quadrature weights and nodes in $[x_{i-1/2}, x_{i+1/2}]$ and $[y_{j-1/2}, y_{j+1/2}]$, respectively. Let $\mathbf{u}_{i\mp\frac{1}{2},j}^{\pm,q} = \mathbf{U}_{ij}^n(x_{i\mp\frac{1}{2}}, y_j^{(q)})$, $\mathbf{u}_{i,j\mp\frac{1}{2}}^{q,\pm} = \mathbf{U}_{ij}^n(x_i^{(q)}, y_{j\mp\frac{1}{2}})$. The numerical fluxes in (7.6) are then given by

$$\hat{\mathbf{f}}_{1,i+\frac{1}{2},j} = \sum_{q=1}^{Q}\omega_q\hat{\mathbf{f}}_1^{\mathrm{LF}}(\mathbf{u}_{i+\frac{1}{2},j}^{-,q}, \mathbf{u}_{i+\frac{1}{2},j}^{+,q}),\quad \hat{\mathbf{f}}_{2,i,j+\frac{1}{2}} = \sum_{q=1}^{Q}\omega_q\hat{\mathbf{f}}_2^{\mathrm{LF}}(\mathbf{u}_{i,j+\frac{1}{2}}^{q,-}, \mathbf{u}_{i,j+\frac{1}{2}}^{q,+}), \tag{7.7}$$

where $\hat{\mathbf{f}}_\ell^{\mathrm{LF}}(\cdot,\cdot)$ is taken as the Lax–Friedrichs flux (6.14) with the numerical viscosity parameters

$$\alpha_{1,n} \geq \max_{i,j,\mu}\widehat{\alpha}_1\big(\mathbf{u}_{i+\frac{1}{2},j}^{\mp,q}, \mathbf{u}_{i-\frac{1}{2},j}^{\pm,q}\big),\quad \alpha_{2,n} \geq \max_{i,j,q}\widehat{\alpha}_2\big(\mathbf{u}_{i,j+\frac{1}{2}}^{q,\mp}, \mathbf{u}_{i,j-\frac{1}{2}}^{q,\pm}\big). \tag{7.8}$$

Here, $\widehat{\alpha}_\ell(\mathbf{u},\tilde{\mathbf{u}}) = \max\big\{|v_\ell| + \mathcal{C}_\ell, |\tilde{v}_\ell| + \tilde{\mathcal{C}}_\ell, \frac{|\sqrt{\rho}v_\ell + \sqrt{\tilde{\rho}}\tilde{v}_\ell|}{\sqrt{\rho}+\sqrt{\tilde{\rho}}} + \max\{\mathcal{C}_\ell, \tilde{\mathcal{C}}_\ell\}\big\} + \frac{\|\mathbf{B}-\tilde{\mathbf{B}}\|}{\sqrt{\rho}+\sqrt{\tilde{\rho}}}$, $\ell = 1, 2$, and $\mathcal{C}_1$ and $\mathcal{C}_2$ are the fast magnetoacoustic speeds in the x- and y-directions, respectively.

Seeking a condition for the scheme (7.6) to be bound-preserving is very challenging, due to the complexity of the system (7.1) and the region (7.4) as well as the intrinsic relations between the bound-preserving property and the DDF condition. On one hand, it is very difficult to establish such relations, since the bound-preserving property is a pointwise *algebraic* property, while the DDF condition is a discrete *differential* property. In fact, their relations remained unclear for a long time until the recent work [104, 110] on the single-component MHD case. On the other hand, the DDF condition strongly couples the states $\{\mathbf{u}^{\pm,q}_{i\mp\frac{1}{2},j}, \mathbf{u}^{q,\pm}_{i,j\mp\frac{1}{2}}\}$, making the traditional or standard analysis approaches (which typically rely on decomposing high-order or/and multidimensional schemes into convex combinations of first-order 1D schemes [121, 122, 125]) *inapplicable* to the present case.

First, let us consider the first-order scheme to gain some insights. In this case, the polynomial degree $K = 0$, so that $\mathbf{U}^n_{ij}(x, y) \equiv \bar{\mathbf{u}}^n_{ij}$ for all $(x, y) \in \mathcal{I}_{ij}$, and we can reformulate the scheme (7.6) as

$$\bar{\mathbf{u}}^{n+1}_{ij} = (1-\lambda)\bar{\mathbf{u}}^n_{ij} + \sigma_1\alpha_{1,n}\mathbf{\Pi}_1 + \sigma_2\alpha_{2,n}\mathbf{\Pi}_2, \tag{7.9}$$

with $\lambda := \sigma_1\alpha_{1,n} + \sigma_2\alpha_{2,n}$, and

$$\begin{aligned}
\mathbf{\Pi}_1 &= \frac{1}{2}\left(\bar{\mathbf{u}}^n_{i+1,j} - \frac{\mathbf{f}_1(\bar{\mathbf{u}}^n_{i+1,j})}{\alpha_{1,n}} + \bar{\mathbf{u}}^n_{i-1,j} + \frac{\mathbf{f}_1(\bar{\mathbf{u}}^n_{i-1,j})}{\alpha_{1,n}}\right),\\
\mathbf{\Pi}_2 &= \frac{1}{2}\left(\bar{\mathbf{u}}^n_{i,j+1} - \frac{\mathbf{f}_2(\bar{\mathbf{u}}^n_{i,j+1})}{\alpha_{2,n}} + \bar{\mathbf{u}}^n_{i,j-1} + \frac{\mathbf{f}_2(\bar{\mathbf{u}}^n_{i,j-1})}{\alpha_{2,n}}\right).
\end{aligned}$$

LEMMA 7.1. *If* $\mathbf{u}, \tilde{\mathbf{u}} \in G$, *then for any* $\boldsymbol{v}_*, \mathbf{B}_* \in \mathbb{R}^3$ *it holds that*

$$\frac{1}{2}\left(\mathbf{u} - \frac{\mathbf{f}_\ell(\mathbf{u})}{\alpha} + \tilde{\mathbf{u}} + \frac{\mathbf{f}_\ell(\tilde{\mathbf{u}})}{\alpha}\right)\cdot\mathbf{n}_* + \frac{\|\mathbf{B}_*\|^2}{2} > \frac{\boldsymbol{v}_*\cdot\mathbf{B}_*}{2\alpha}(\tilde{B}_\ell - B_\ell),$$

where $|\alpha| > \widehat{\alpha}_\ell(\mathbf{u}, \tilde{\mathbf{u}})$ *and* $\ell \in \{1, 2\}$.

The proof of Lemma 7.1 follows from [104, Lemma 2.6] and is omitted.

THEOREM 7.2. *If* $\bar{\mathbf{u}}^n_{ij} \in G$ *for all* i *and* j, *then, under the standard CFL condition* $\lambda \le 1$, *the solution* $\bar{\mathbf{u}}^{n+1}_{ij}$ *of* (7.9) *satisfies*

$$\bar{\mathbf{u}}^{n+1}_{ij}\cdot\mathbf{e}_k \ge 0, \quad 0 \le k < n_c, \qquad \bar{\mathbf{u}}^{n+1}_{ij}\cdot\mathbf{e}_{n_c} > 0, \tag{7.10}$$

$$\varphi(\bar{\mathbf{u}}^{n+1}_{ij}; \boldsymbol{v}_*, \mathbf{B}_*) > -\Delta t(\boldsymbol{v}_*\cdot\mathbf{B}_*)\mathrm{div}_{ij}\bar{\mathbf{B}} \qquad \forall \boldsymbol{v}_*, \mathbf{B}_* \in \mathbb{R}^3, \tag{7.11}$$

where $\mathrm{div}_{ij}\bar{\mathbf{B}} := \frac{\bar{B}^n_{1,i+1,j} - \bar{B}^n_{1,i-1,j}}{2\Delta x} + \frac{\bar{B}^n_{2,i,j+1} - \bar{B}^n_{2,i,j-1}}{2\Delta y}$ *is a discrete divergence. Furthermore, if the states* $\{\bar{\mathbf{u}}^n_{ij}\}$ *satisfy the DDF condition* $\mathrm{div}_{ij}\bar{\mathbf{B}} = 0$, *then* (7.10)–(7.11) *imply* $\bar{\mathbf{u}}^{n+1}_{ij} \in G_* = G$.

Proof. For $0 \le k < n_c$ and any $\mathbf{u} \in \{\bar{\mathbf{u}}^n_{ij}\}$, we have $\pm\mathbf{f}_\ell(\mathbf{u})\cdot\mathbf{e}_k = \pm v_\ell(\mathbf{u}\cdot\mathbf{e}_k) \le \alpha_{\ell,n}(\mathbf{u}\cdot\mathbf{e}_k)$, which implies $\mathbf{\Pi}_\ell\cdot\mathbf{e}_k \ge 0$. Similarly, $\mathbf{\Pi}_\ell\cdot\mathbf{e}_{n_c} > 0$. These lead to (7.10). Lemma 7.1 implies

$$\begin{aligned}
\varphi(\mathbf{\Pi}_1; \boldsymbol{v}_*, \mathbf{B}_*) &> \frac{\boldsymbol{v}_*\cdot\mathbf{B}_*}{2\alpha_{1,n}}\left(\bar{B}^n_{1,i-1,j} - \bar{B}^n_{1,i+1,j}\right),\\
\varphi(\mathbf{\Pi}_2; \boldsymbol{v}_*, \mathbf{B}_*) &> \frac{\boldsymbol{v}_*\cdot\mathbf{B}_*}{2\alpha_{2,n}}\left(\bar{B}^n_{2,i,j-1} - \bar{B}^n_{2,i,j+1}\right).
\end{aligned}$$

Thanks to the linearity of $\varphi(\cdot; \boldsymbol{v}_*, \mathbf{B}_*)$, it then follows from (7.9) that

$$\begin{aligned}
\varphi(\bar{\mathbf{u}}^{n+1}_{ij}; \boldsymbol{v}_*, \mathbf{B}_*) &= (1-\lambda)\varphi(\bar{\mathbf{u}}^n_{ij}; \boldsymbol{v}_*, \mathbf{B}_*) + \sigma_1\alpha_{1,n}\varphi(\mathbf{\Pi}_1; \boldsymbol{v}_*, \mathbf{B}_*) + \sigma_2\alpha_{2,n}\varphi(\mathbf{\Pi}_2; \boldsymbol{v}_*, \mathbf{B}_*)\\
&> (1-\lambda)\varphi(\bar{\mathbf{u}}^n_{ij}; \boldsymbol{v}_*, \mathbf{B}_*) - \Delta t(\boldsymbol{v}_*\cdot\mathbf{B}_*)\mathrm{div}_{ij}\bar{\mathbf{B}},
\end{aligned}$$

which yields (7.11) under the CFL condition $\lambda \le 1$. □

Theorem 7.2 shows the connection between the bound-preserving property and a DDF condition, which is bridged by (7.11) with the help of the free auxiliary variables $\{\boldsymbol{v}_*, \mathbf{B}_*\}$ in the GQL representation (7.5). *This demonstrates the essential importance of the GQL approach in establishing this connection and its significant advantages for bound-preserving analysis and design.* It seems to be very challenging (if not impossible) to draw such a connection without using the GQL approach.

Now, we use the GQL approach to explore bound-preserving high-order schemes with $K \geq 1$. Let $\{\widehat{x}_i^{(\beta)}\}_{\beta=1}^L$ and $\{\widehat{y}_j^{(\beta)}\}_{\beta=1}^L$ be the Gauss–Lobatto quadrature points in $[x_{i-1/2}, x_{i+1/2}]$ and $[y_{j-1/2}, y_{j+1/2}]$, respectively, and let $\{\widehat{\omega}_\beta\}_{\beta=1}^L$ be the weights, with $L = \lceil \frac{K+3}{2} \rceil$. Similar to Theorem 7.2 and [104, Theorem 4.7], the following result can be derived (with the proof omitted here).

THEOREM 7.3. *If, for all i and j, $\bar{\mathbf{u}}_{ij}^n \in G$ and the polynomial vector $\mathbf{U}_{ij}^n(x,y)$ satisfies*

$$\mathbf{U}_{ij}^n(\widehat{x}_i^{(\beta)}, y_j^{(q)}), \mathbf{U}_{ij}^n(x_i^{(q)}, \widehat{y}_j^{(\beta)}) \in G \qquad \forall \beta, q, \tag{7.12}$$

then the solution $\bar{\mathbf{u}}_{ij}^{n+1}$ of the scheme (7.6) *satisfies*

$$\varphi(\bar{\mathbf{u}}_{ij}^{n+1}; \boldsymbol{v}_*, \mathbf{B}_*) > 2(\widehat{\omega}_1 - \lambda)\varphi(\boldsymbol{\Pi}; \boldsymbol{v}_*, \mathbf{B}_*) - \Delta t(\boldsymbol{v}_* \cdot \mathbf{B}_*)\mathrm{div}_{ij}\mathbf{B} \tag{7.13}$$

with $\boldsymbol{\Pi} := \frac{1}{2\lambda}\sum_q \omega_q \big[\sigma_1 \alpha_{1,n}\big(\mathbf{u}_{i+\frac{1}{2},j}^{-,q} + \mathbf{u}_{i-\frac{1}{2},j}^{+,q}\big) + \sigma_2 \alpha_{2,n}\big(\mathbf{u}_{i,j+\frac{1}{2}}^{q,-} + \mathbf{u}_{i,j-\frac{1}{2}}^{q,+}\big)\big] \in G$. Furthermore, under the CFL condition $\lambda \leq \widehat{\omega}_1$, we have

$$\bar{\mathbf{u}}_{ij}^{n+1} \cdot \mathbf{e}_k \geq 0, \quad 0 \leq k < n_c, \qquad \bar{\mathbf{u}}_{ij}^{n+1} \cdot \mathbf{e}_{n_c} > 0, \tag{7.14}$$

$$\varphi(\bar{\mathbf{u}}_{ij}^{n+1}; \boldsymbol{v}_*, \mathbf{B}_*) > -\Delta t(\boldsymbol{v}_* \cdot \mathbf{B}_*)\mathrm{div}_{ij}\mathbf{B} \qquad \forall \boldsymbol{v}_*, \mathbf{B}_* \in \mathbb{R}^3, \tag{7.15}$$

where the discrete divergence is defined as $\mathrm{div}_{ij}\mathbf{B} := \frac{1}{2}\big(\mathrm{div}_{ij}^{-}\mathbf{B} + \mathrm{div}_{ij}^{+}\mathbf{B}\big)$ *with*

$$\mathrm{div}_{ij}^{\mp}\mathbf{B} := \frac{1}{\Delta x}\sum_{q=1}^{Q} \omega_q \Big(B_{1,i+\frac{1}{2},j}^{\mp,q} - B_{1,i-\frac{1}{2},j}^{\pm,q}\Big) + \frac{1}{\Delta y}\sum_{q=1}^{Q} \omega_q \Big(B_{2,i,j+\frac{1}{2}}^{q,\mp} - B_{2,i,j-\frac{1}{2}}^{q,\pm}\Big).$$

Remark 7.4. The condition (7.12) in Theorem 7.3 is standard (cf. [121, 122]) and can be easily enforced by a simple scaling limiter (see Appendix B). Thanks to the GQL representation (7.5), we observe from (7.14)–(7.15) that, in order to ensure $\bar{\mathbf{u}}_{ij}^{n+1} \in G_* = G$, an extra DDF condition

$$\mathrm{div}_{ij}\mathbf{B} := \frac{1}{2}\big(\mathrm{div}_{ij}^{-}\mathbf{B} + \mathrm{div}_{ij}^{+}\mathbf{B}\big) = 0 \tag{7.16}$$

is also required. Unfortunately, the high-order schemes (7.6) do not preserve the DDF condition (7.16), which depends on the numerical solution information from adjacent cells. Although a few *globally* divergence-free techniques (e.g., [35]) have been developed and could meet the condition (7.16), the local scaling limiter for (7.12) will destroy the globally divergence-free property. Notice that the *locally* divergence-free technique (e.g., [66]) is compatible with the local scaling limiter, but can only guarantee $\mathrm{div}_{ij}^{-}\mathbf{B} = 0$. In subsection 7.3, we will use the GQL approach to explore how to eliminate the effect of the remaining part $\mathrm{div}_{ij}^{+}\mathbf{B}$ by properly modifying the scheme (7.6).

7.3. Seek High-Order Provably Bound-Preserving Schemes via GQL. We have established the relations between the bound-preserving and divergence-free properties at the discrete level. Interestingly, at the continuous level, bound preservation is also closely related to the continuous divergence-free condition (7.2): If (7.2) is slightly violated, then even the exact solution of system (7.1) may not stay in G; see [106, page B1326] for a discussion which is also valid for system (7.1). To address this issue, we consider a modified formulation of the multicomponent MHD equations

$$\partial_t \mathbf{u} + \partial_x \mathbf{f}_1(\mathbf{u}) + \partial_y \mathbf{f}_2(\mathbf{u}) + (\nabla \cdot \mathbf{B})\mathbf{S}(\mathbf{u}) = \mathbf{0} \tag{7.17}$$

by adding an extra source term to (7.1a) with $\mathbf{S}(\mathbf{u}) = (\mathbf{0}_{n_c}, \mathbf{B}, \boldsymbol{v}, \boldsymbol{v}\cdot\mathbf{B})^\top$. Such a formulation was first proposed by Godunov [37] for the purpose of entropy symmetrization in the single-component MHD case. Notice that, for divergence-free initial conditions, the exact solutions of the modified form (7.17) and the standard form (7.1) are the same. However, if the divergence-free condition (7.2) is violated, the extra source term in the modified form (7.17) becomes beneficial and helps keep the exact solutions in G; see [107, pages 1042–1043] for an analysis which also works for system (7.17). This finding motivates us to explore bound-preserving schemes based on suitable discretization of the modified form (7.17). Thus we consider

$$\bar{\mathbf{u}}_{ij}^{n+1} = \bar{\mathbf{u}}_{ij}^{n} - \sigma_1\big(\hat{\mathbf{f}}_{1,i+\frac{1}{2},j} - \hat{\mathbf{f}}_{1,i-\frac{1}{2},j}\big) - \sigma_2\big(\hat{\mathbf{f}}_{2,i,j+\frac{1}{2}} - \hat{\mathbf{f}}_{2,i,j-\frac{1}{2}}\big) - \widehat{\mathbf{S}}_{ij} \tag{7.18}$$

by adding a properly discretized source term $\widehat{\mathbf{S}}_{ij}$ into the standard finite volume or discontinuous Galerkin schemes (7.6). As discussed in Remark 7.4, we can adopt a locally divergence-free technique for the magnetic components of $\mathbf{U}_{ij}^n(x,y)$ such that $\mathrm{div}_{ij}^{-}\mathbf{B} = 0$. This gives $2\mathrm{div}_{ij}\mathbf{B} = \mathrm{div}_{ij}^{+}\mathbf{B} = \mathrm{div}_{ij}^{+}\mathbf{B} - \mathrm{div}_{ij}^{-}\mathbf{B}$, thereby leading to

$$\mathrm{div}_{ij}\mathbf{B} = \frac{1}{2\Delta x}\sum_{q=1}^{Q}\omega_q\Big([\![B_1]\!]^q_{i+\frac{1}{2},j} + [\![B_1]\!]^q_{i-\frac{1}{2},j}\Big) + \frac{1}{2\Delta y}\sum_{q=1}^{Q}\omega_q\Big([\![B_2]\!]^q_{i,j+\frac{1}{2}} + [\![B_2]\!]^q_{i,j-\frac{1}{2}}\Big), \tag{7.19}$$

where $[\![B_1]\!]^q_{i+\frac{1}{2},j} = B^{+,q}_{1,i+\frac{1}{2},j} - B^{-,q}_{1,i+\frac{1}{2},j}$ and $[\![B_2]\!]^q_{i,j+\frac{1}{2}} = B^{q,-}_{2,i,j+\frac{1}{2}} - B^{q,-}_{2,i,j+\frac{1}{2}}$ are the jumps of the normal magnetic component across the cell interfaces. Using the GQL approach with the linearity of $\varphi(\cdot;\boldsymbol{v}_*,\mathbf{B}_*)$ and the estimate (7.13) under the hypothesis of Theorem 7.3, we obtain

$$\varphi(\bar{\mathbf{u}}_{ij}^{n+1};\boldsymbol{v}_*,\mathbf{B}_*) > 2(\widehat{\omega}_1 - \lambda)\varphi(\mathbf{\Pi};\boldsymbol{v}_*,\mathbf{B}_*) - \Big[\Delta t(\boldsymbol{v}_*\cdot\mathbf{B}_*)\mathrm{div}_{ij}\mathbf{B} + \widehat{\mathbf{S}}_{ij}\cdot\mathbf{n}_*\Big]. \tag{7.20}$$

Then the key is to carefully design $\widehat{\mathbf{S}}_{ij}$ to exactly offset the effect of $\mathrm{div}_{ij}\mathbf{B}$ in (7.20), so that the resulting schemes (7.18) become bound-preserving. Observing that for any $b \in \mathbb{R}$ and any $\mathbf{u} \in G$,

$$b(\boldsymbol{v}_*\cdot\mathbf{B}_* + \mathbf{S}(\mathbf{u})\cdot\mathbf{n}_*) = b(\boldsymbol{v} - \boldsymbol{v}_*)\cdot(\mathbf{B} - \mathbf{B}_*) \le |b|\rho^{1/2}\varphi(\mathbf{u};;\boldsymbol{v}_*,\mathbf{B}_*), \tag{7.21}$$

we devise

$$\begin{aligned}\widehat{\mathbf{S}}_{ij} = {} & \frac{\sigma_1}{2}\sum_{q=1}^{Q}\omega_q\Big[[\![B_1]\!]^q_{i+\frac{1}{2},j}\mathbf{S}(\mathbf{u}^{-,q}_{i+\frac{1}{2},j}) + [\![B_1]\!]^q_{i-\frac{1}{2},j}\mathbf{S}(\mathbf{u}^{+,q}_{i-\frac{1}{2},j})\Big] \\ & + \frac{\sigma_2}{2}\sum_{q=1}^{Q}\omega_q\Big[[\![B_2]\!]^q_{i,j+\frac{1}{2}}\mathbf{S}(\mathbf{u}^{q,-}_{i,j+\frac{1}{2}}) + [\![B_2]\!]^q_{i,j-\frac{1}{2}}\mathbf{S}(\mathbf{u}^{q,+}_{i,j-\frac{1}{2}})\Big],\end{aligned} \tag{7.22}$$

such that the last term in (7.20) satisfies

$$\begin{aligned}
&\Delta t(\boldsymbol{v}_* \cdot \mathbf{B}_*)\mathrm{div}_{ij}\mathbf{B} + \widehat{\mathbf{S}}_{ij} \cdot \mathbf{n}_* \\
&= \frac{\sigma_1}{2} \sum_{q=1}^{Q} \omega_q \bigg[[\![B_1]\!]^q_{i+\frac{1}{2},j} \Big(\boldsymbol{v}_* \cdot \mathbf{B}_* + \mathbf{S}(\mathbf{u}^{-,q}_{i+\frac{1}{2},j}) \cdot \mathbf{n}_* \Big) \\
&\qquad + [\![B_1]\!]^q_{i-\frac{1}{2},j} \Big(\boldsymbol{v}_* \cdot \mathbf{B}_* + \mathbf{S}(\mathbf{u}^{+,q}_{i-\frac{1}{2},j}) \cdot \mathbf{n}_* \Big) \bigg] \\
&\quad + \frac{\sigma_2}{2} \sum_{q=1}^{Q} \omega_q \bigg[[\![B_2]\!]^q_{i,j+\frac{1}{2}} \Big(\boldsymbol{v}_* \cdot \mathbf{B}_* + \mathbf{S}(\mathbf{u}^{q,-}_{i,j+\frac{1}{2}}) \cdot \mathbf{n}_* \Big) \\
&\qquad + [\![B_2]\!]^q_{i,j-\frac{1}{2}} \Big(\boldsymbol{v}_* \cdot \mathbf{B}_* + \mathbf{S}(\mathbf{u}^{q,+}_{i,j-\frac{1}{2}}) \cdot \mathbf{n}_* \Big) \bigg] \\
&\le \varepsilon \lambda \varphi(\mathbf{\Pi}; \boldsymbol{v}_*, \mathbf{B}_*),
\end{aligned} \tag{7.23}$$

where we use (7.19) in the equality and (7.21) in the inequality, and $\varepsilon = \max\{\beta_1/\alpha_{1,n}, \beta_2/\alpha_{2,n}\}$ with

$$\beta_1 = \max_{i,j,q} \left\{ \big| [\![B_1]\!]^q_{i+\frac{1}{2},j} \big| (\rho^{\pm,q}_{i+\frac{1}{2},j})^{-1/2} \right\}$$

and

$$\beta_2 = \max_{i,j,q} \left\{ \big| [\![B_2]\!]^q_{i,j+\frac{1}{2}} \big| (\rho^{q,\pm}_{i,j+\frac{1}{2}})^{-1/2} \right\}.$$

Combining (7.20) with (7.23), we obtain

$$\varphi(\bar{\mathbf{u}}^{n+1}_{ij}; \boldsymbol{v}_*, \mathbf{B}_*) > 2(\widehat{\omega}_1 - \lambda - \varepsilon\lambda)\varphi(\mathbf{\Pi}; \boldsymbol{v}_*, \mathbf{B}_*) \ge 0$$

under the CFL condition $(1+\varepsilon)\lambda \le \widehat{\omega}_1$. Notice that the first n_c components of $\widehat{\mathbf{S}}_{ij}$ are zeros, which implies that (7.14) in Theorem 7.3 also holds for the modified schemes (7.18). In summary, we obtain the following result.

THEOREM 7.5. *If, for all i and j, $\bar{\mathbf{u}}^n_{ij} \in G$ and the polynomial vector $\mathbf{U}^n_{ij}(x,y)$ satisfies* (7.12) *and* $\mathrm{div}^-_{ij}\mathbf{B} = 0$, *then, under the CFL condition* $(1+\varepsilon)\lambda \le \widehat{\omega}_1$, *the solution* $\bar{\mathbf{u}}^{n+1}_{ij}$ *of* (7.18) *is always preserved in* G_*.

Theorem 7.5 indicates that if we use the scaling limiter in Appendix B to enforce (7.12) and a locally divergence-free technique to ensure $\mathrm{div}^-_{ij}\mathbf{B} = 0$, then the scheme (7.18) with (7.22) is bound-preserving. The bounds are also preserved if a high-order strong-stability-preserving time discretization [38] is used to replace the Euler forward method.

8. Experimental Results. This section gives two highly demanding numerical examples to further demonstrate our theoretical analysis as well as the robustness and effectiveness of the bound-preserving schemes designed via GQL in subsection 7.3 for the 2D multicomponent MHD. We use the proposed bound-preserving third-order locally divergence-free discontinuous Galerkin method for spatial discretization. As the tests involve strong discontinuities, the locally divergence-free WENO limiter [126] is also employed in some trouble cells adaptively detected by the indicator of [62]. The third-order strong-stability-preserving Runge–Kutta method [38] is adopted for time discretization, with the CFL number set as 0.15.

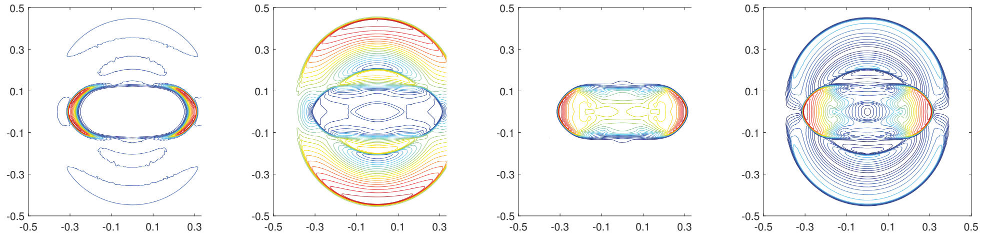

Fig. 4 *The contour plots of ρ, p_m, p, and $\|\boldsymbol{v}\|$ (from left to right) for the blast problem at $t = 0.01$.*

Example 8.1 (blast problem). This test simulates a benchmark MHD problem in the domain $[-0.5, 0.5]^2$ with outflow boundary conditions. The setup is similar to that in [5] except for a fluid mixture with $n_c = 2$, $C_{v_1} = 2.42$, $C_{v_2} = 0.72$, $\Gamma_1 = 5/3$, and $\Gamma_2 = 1.4$. Initially, the fluid is stationary, with $(\rho, p, Y_1, Y_2) = (1, 1000, 1, 0)$ in the explosion region $(x^2 + y^2 \leq 0.01)$ and $(1, 0.1, 0, 1)$ in the ambient region $(x^2 + y^2 > 0.01)$. The magnetic field $\mathbf{B}$ is initialized as $(100/\sqrt{4\pi}, 0, 0)$. Due to the large jump in p and the strong magnetic field, negative numerical p can be easily produced and often causes failure of the numerical simulations. Figure 4 presents the contour plots of the density ρ, the magnetic pressure $p_m = \frac{1}{2}\|\mathbf{B}\|^2$, the thermal pressure p, and the velocity magnitude $\|\boldsymbol{v}\|$ computed by the proposed bound-preserving discontinuous Galerkin method with 400×400 uniform cells. We observe that the flow structures are well captured and our method is highly robust and always preserves the bound principles (7.4) in the whole simulation.

Example 8.2 (astrophysical jet). This test simulates a high-speed MHD jet flow in the domain $[-0.5, 0.5] \times [0, 1.5]$ with $n_c = 2$, $C_{v_1} = 0.72$, $C_{v_2} = 2.42$, $\Gamma_1 = 1.4$, and $\Gamma_2 = 5/3$. The domain is initially filled with static fluid with $(\rho, p, Y_1, Y_2) = (0.14, 1, 0, 1)$. The inflow jet condition is fixed on boundary $\{|x| < 0.05, y = 0\}$ with $(\rho, p, Y_1, Y_2) = (1.4, 1, 1, 0)$ and $\boldsymbol{v} = (0, 800, 0)$, while the outflow conditions are specified on the other boundaries. There is a strong magnetic field $\mathbf{B}$ initialized as $(0, \sqrt{4000}, 0)$, which makes this test more challenging. Our simulation is based on the proposed bound-preserving method with 200×600 uniform cells in $[0, 0.5] \times [0, 1.5]$. The numerical results are shown in Figure 5. The flow pattern is captured with high resolution and is similar to the single-component MHD case reported in [106, 107]. In such an extreme test, our bound-preserving method exhibits good robustness. However, if the proposed scaling limiter is not used to enforce (7.12), or if the locally divergence-free technique is not employed to ensure $\mathrm{div}_{ij}^{-}\mathbf{B} = 0$, or if the proposed source term (7.22) is dropped, the resulting method even with the WENO limiter is not bound-preserving and would fail quickly due to nonphysical numerical solutions out of the bounds. This confirms our theoretical analyses and the importance of the proposed conditions and techniques.

9. Conclusions. We have systematically proposed a novel and general framework, called geometric quasilinearization (GQL), for studying bound-preserving problems with nonlinear constraints. GQL skillfully converts all nonlinear constraints into linear ones, by properly introducing some free auxiliary variables independent of the system variables. We have established the fundamental principle and general theory of GQL and provided three simple methods for constructing GQL representations. The GQL approach equivalently casts the nonlinear bound-preserving problems into

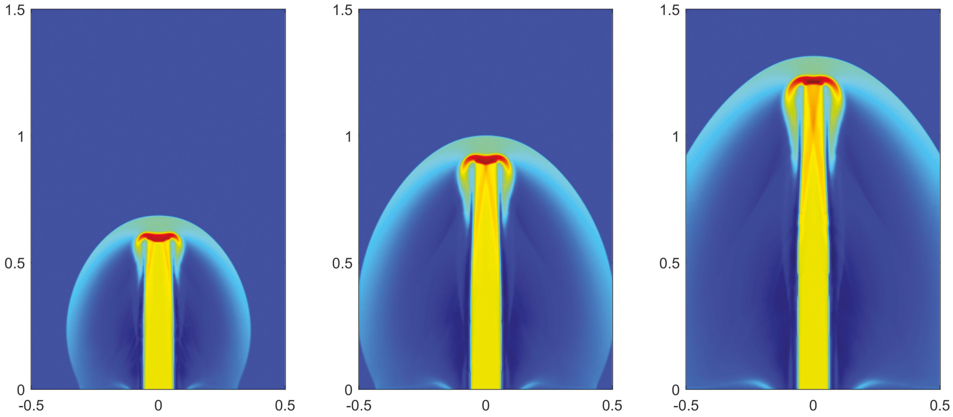

Fig. 5 *The plots of* $\log(\rho)$ *for the jet problem. From left to right:* $t = 0.001$*,* 0.0015*, and* 0.002*.*

preserving the positivity of linear scalar functions, thereby opening up a new effective way for bound-preserving study. Several examples have been provided to demonstrate the effectiveness and advantages of the GQL approach in addressing nonlinear bound-preserving problems that are highly challenging and could not be easily handled by direct or traditional approaches. Recently, the GQL approach has also achieved success in finding high-order bound-preserving schemes for several complicated PDE systems in [104, 106, 110, 105, 107, 108]. The GQL approach applies to the bound-preserving study for the general d-dimensional hyperbolic system $\partial_t \mathbf{u} + \nabla \cdot \mathbf{f}(\mathbf{u}) = \mathbf{0}$, and in particular, for a general scheme with the Lax–Friedrichs flux on general polytope meshes, the central task is to establish the estimate

$$(\boldsymbol{\xi} \cdot \mathbf{f}(\mathbf{u})) \cdot \mathbf{n}_{i*} \leq \alpha(\mathbf{u}, \boldsymbol{\xi})(\mathbf{u} - \mathbf{u}_*) \cdot \mathbf{n}_{i*} - \boldsymbol{\xi} \cdot \mathbf{h}(\mathbf{u}_*) \quad \forall \mathbf{u}_* \in \mathcal{S}_i$$

for any unit vector $\boldsymbol{\xi} \in \mathbb{R}^d$ by seeking a suitable $\mathbf{h}(\mathbf{u}_*)$ and an estimated maximum wave speed $\alpha(\mathbf{u}, \boldsymbol{\xi})$ in the direction of $\boldsymbol{\xi}$. This will be detailed in a future paper.

The GQL framework is applicable to general convex invariant regions with nonlinear constraints and is not restricted to the specific forms of the PDEs. However, this does not claim that our approach addresses the bound-preserving problems in all cases: It is just a promising framework that simplifies the bound-preserving study of many important hyperbolic type systems, and we hope this paper will motivate further study of applying the GQL approach to a wider class of problems.

Appendix A. Proof of Proposition 3.9. The proof is divided into two steps.

(i) *Prove that* $G \subseteq G_*$. For any $\mathbf{u}_* \in \partial G$, the hyperplane $(\mathbf{u} - \mathbf{u}_*) \cdot \mathbf{n}_* = 0$ supports the convex region G at $\mathbf{u}_*$. Thus we have

$$G \subseteq \{\mathbf{u} \in \mathbb{R}^N : (\mathbf{u} - \mathbf{u}_*, \mathbf{n}_*) \geq 0 \ \ \forall \mathbf{u}_* \in \partial G\}. \tag{A.1}$$

If G is closed, then (A.1) means $G \subseteq G_*$. Next, we assume G is open and show $G \subseteq G_*$ by contradiction. Assume that

$$\text{there exists } \mathbf{u}_0 \in G \text{ but } \mathbf{u}_0 \notin G_*. \tag{A.2}$$

Then, according to (A.1), there exists $\mathbf{u}_* \in \partial G$ such that $(\mathbf{u}_0 - \mathbf{u}_*) \cdot \mathbf{n}_* = 0$. Since G is open, there exists $\delta > 0$ such that $\Omega_\delta := \{\mathbf{u} \in \mathbb{R}^N : \|\mathbf{u} - \mathbf{u}_0\| < \delta\} \subset G$. We take

$\mathbf{u}_\delta := \mathbf{u}_0 - \frac{\delta}{2\|\mathbf{n}_*\|}\mathbf{n}_* \in \Omega_\delta$. Then $\mathbf{u}_\delta \in G$. However, using $(\mathbf{u}_0 - \mathbf{u}_*) \cdot \mathbf{n}_* = 0$ gives

$$(\mathbf{u}_\delta - \mathbf{u}_*) \cdot \mathbf{n}_* = \left(\mathbf{u}_0 - \mathbf{u}_* - \frac{\delta}{2\|\mathbf{n}_*\|}\mathbf{n}_* \right) \cdot \mathbf{n}_* = -\frac{\delta}{2}\|\mathbf{n}_*\| < 0,$$

which contradicts (A.1) and $\mathbf{u}_\delta \in G$. Thus the assumption (A.2) is incorrect, and we have $G \subseteq G_*$.

(ii) *Prove that* $G_* \subseteq G$. We first show that $G_* \subseteq \mathrm{cl}(G)$ by contradiction. Assume that

$$\text{(A.3)} \qquad \text{there exists } \mathbf{u}_0 \in G_* \text{ but } \mathbf{u}_0 \notin \mathrm{cl}(G).$$

According to the theory of convex optimization [8], the minimum of the convex function $\zeta(\mathbf{u}) := \|\mathbf{u} - \mathbf{u}_0\|^2$ over the closed convex region $\mathrm{cl}(G)$ is attained at a certain boundary point $\mathbf{u}_* \in \partial G$. Let $\widehat{\mathbf{u}}$ be an arbitrary interior point of G. Thanks to the convexity of $\mathrm{cl}(G)$, one has $\mathbf{u}_\lambda := \lambda \widehat{\mathbf{u}} + (1-\lambda)\mathbf{u}_* \in \mathrm{cl}(G)$ for any $\lambda \in [0,1]$. We then know that the quadratic function

$$\widehat{\zeta}(\lambda) := \zeta(\mathbf{u}_\lambda) = \lambda^2 \left\|\widehat{\mathbf{u}} - \mathbf{u}_*\right\|^2 + 2\lambda \left(\widehat{\mathbf{u}} - \mathbf{u}_*\right) \cdot (\mathbf{u}_* - \mathbf{u}_0) + \|\mathbf{u}_* - \mathbf{u}_0\|^2$$

attains its minimum over $[0,1]$ at $\lambda = 0$. This implies $(\widehat{\mathbf{u}} - \mathbf{u}_*) \cdot (\mathbf{u}_* - \mathbf{u}_0) \geq 0$ for an arbitrary interior point $\widehat{\mathbf{u}}$ of G. Thus $\mathrm{int}(G) \subseteq \{\mathbf{u} : (\mathbf{u} - \mathbf{u}_*) \cdot (\mathbf{u}_* - \mathbf{u}_0) \geq 0\} =: H_*^+$, where H_*^+ is a closed halfspace. It follows that H_*^+ is a supporting halfspace to G, and $\mathbf{u}_* - \mathbf{u}_0$ is an inward-pointing normal vector of G at $\mathbf{u}_*$. Because ∂G is smooth, there exists $\mu > 0$ such that $\mathbf{n}_* = \mu(\mathbf{u}_* - \mathbf{u}_0)$, which implies $(\mathbf{u}_0 - \mathbf{u}_*) \cdot \mathbf{n}_* = -\mu \left\|\mathbf{u}_0 - \mathbf{u}_*\right\|^2 < 0$. This contradicts the assumption $\mathbf{u}_0 \in G_*$. Thus the assumption (A.3) is incorrect, and we have $G_* \subseteq \mathrm{cl}(G)$. If G is closed, then we obtain $G_* \subseteq G$. If G is open, then $\partial G \cap G_* = \emptyset$, which along with $G_* \subseteq \mathrm{cl}(G)$ yields $G_* \subseteq G$.

In summary, we have $G = G_*$, and the proof is completed. □

Appendix B. A Simple Scaling Limiter to Enforce (7.12). Condition (7.12) is not always automatically satisfied by the polynomial vector $\mathbf{U}_{ij}^n(x,y)$ in the high-order schemes. If this occurs, the following limiter is used to modify $\mathbf{U}_{ij}^n(x,y)$ into $\widetilde{\mathbf{U}}_{ij}^n(x,y)$ such that $\widetilde{\mathbf{U}}_{ij}^n(x,y)$ satisfies (7.12). Define $\mathbb{Q}_{ij} = \{(\widehat{x}_i^{(\beta)}, y_j^{(q)}), (x_i^{(q)}, \widehat{y}_j^{(\beta)}) \; \forall \beta, q\}$ as the set of all the points involved in (7.12). Since the limiter is performed separately for each cell, the subscripts ij and superscript n of all quantities are omitted below for convenience. First, modify the density as

$$\widehat{\rho}(x,y) = \overline{\rho} + \theta_1(\rho(x,y) - \overline{\rho}), \qquad \theta_1 := (\overline{\rho} - \epsilon_1)/\big(\overline{\rho} - \min_{(x,y)\in\mathbb{Q}_{ij}} \rho(x,y)\big),$$

where ϵ_1 is a small positive number and may be taken as $\min\{10^{-13}, \overline{\rho}\}$. Define $\mathbb{S}_k = \{(x,y) \in \mathbb{Q}_{ij} : \rho Y_k(x,y) \leq 0\}$. Then, modify the mass fractions [25] as

$$\widehat{\rho Y_k}(x,y) = \rho Y_k(x,y) + \theta_2 \left(\overline{\rho Y_k} \widehat{\rho}(x,y)/\overline{\rho} - \rho Y_k(x,y) \right), \qquad 1 \leq k \leq n_c - 1,$$

where $\theta_2 = \max_{1\leq k \leq n_c} \max_{(x,y)\in\mathbb{S}_k} \{ \frac{-\rho Y_k(x,y)}{\overline{\rho Y_k}\widehat{\rho}(x,y)/\overline{\rho} - \rho Y_k(x,y)} \}$ with $\rho Y_{n_c} = \widehat{\rho} - \sum_{k=1}^{n_c-1} \rho Y_k$. Denote $\widehat{\mathbf{U}} = (\widehat{\rho \mathbf{Y}}, \widehat{\rho}, \boldsymbol{m}, \mathbf{B}, E)^\top$. Finally, modify $\widehat{\mathbf{U}}$ to enforce the positivity of $g(\mathbf{U}) = E - \frac{1}{2}(\|\boldsymbol{m}\|^2/\rho + \|\mathbf{B}\|^2)$ by

$$\widetilde{\mathbf{U}}(x,y) = \overline{\mathbf{U}} + \theta_3(\widehat{\mathbf{U}}(x,y) - \overline{\mathbf{U}}), \qquad \theta_3 := (g(\overline{\mathbf{U}}) - \epsilon_2)/\big(g(\overline{\mathbf{U}}) - \min_{(x,y)\in\mathbb{Q}_{ij}} g(\widehat{\mathbf{U}}(x,y))\big),$$

where ϵ_2 is a small positive number and may be taken as $\min\{10^{-13}, g(\overline{\mathbf{U}})\}$. Note that the pressure function $p(\mathbf{U})$ in (7.3) is generally not concave so we use the concave function $g(\mathbf{U})$ instead of $p(\mathbf{U})$. It can be verified that the limited solution $\widetilde{\mathbf{U}}(x, y) \in G$ for all $(x, y) \in \mathbb{Q}_{ij}$ and its cell average equals $\overline{\mathbf{U}}$. Such types of limiters do not lose the high-order accuracy, as demonstrated in [121, 122, 119].

REFERENCES

[1] R. Abgrall, M. Lukácova-Medvid'ová, and P. Öffner, *On the convergence of residual distribution schemes for the compressible Euler equations via dissipative weak solutions*, Math. Models Methods Appl. Sci., 33 (2023), pp. 139–173. (Cited on p. 1033)

[2] H. Amann, *Invariant sets and existence theorems for semilinear parabolic and elliptic systems*, J. Math. Anal. Appl., 65 (1978), pp. 432–467. (Cited on p. 1033)

[3] V. I. Arnol'd, *Mathematical Methods of Classical Mechanics*, Grad. Texts Math. 60, Springer Science & Business Media, 1989. (Cited on p. 1046)

[4] S. Badia and A. Hierro, *On discrete maximum principles for discontinuous Galerkin methods*, Comput. Methods Appl. Mech. Engrg., 286 (2015), pp. 107–122. (Cited on p. 1033)

[5] D. S. Balsara and D. Spicer, *A staggered mesh algorithm using high order Godunov fluxes to ensure solenoidal magnetic fields in magnetohydrodynamic simulations*, J. Comput. Phys., 149 (1999), pp. 270–292. (Cited on p. 1065)

[6] P. Batten, N. Clarke, C. Lambert, and D. M. Causon, *On the choice of wavespeeds for the HLLC Riemann solver*, SIAM J. Sci. Comput., 18 (1997), pp. 1553–1570, https://doi.org/10.1137/S1064827593260140. (Cited on p. 1035)

[7] J. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples*, Springer Science & Business Media, 2010. (Cited on p. 1044)

[8] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. (Cited on pp. 1044, 1067)

[9] E. Burman and A. Ern, *Stabilized Galerkin approximation of convection-diffusion-reaction equations: Discrete maximum principle and convergence*, Math. Comp., 74 (2005), pp. 1637–1652. (Cited on p. 1033)

[10] S. L. Campbell and E. Moore, *Constraint preserving integrators for general nonlinear higher index DAEs*, Numer. Math., 69 (1995), pp. 383–399. (Cited on p. 1033)

[11] M. Campos-Pinto, F. Charles, and B. Després, *Algorithms for positive polynomial approximation*, SIAM J. Numer. Anal., 57 (2019), pp. 148–172, https://doi.org/10.1137/17M1131891. (Cited on p. 1033)

[12] M. Campos Pinto, F. Charles, B. Després, and M. Herda, *A projection algorithm on the set of polynomials with two bounds*, Numer. Algorithms, 85 (2020), pp. 1475–1498. (Cited on p. 1033)

[13] W. Chen, C. Wang, X. Wang, and S. M. Wise, *Positivity-preserving, energy stable numerical schemes for the Cahn–Hilliard equation with logarithmic potential*, J. Comput. Phys. X, 3 (2019), art. 100031. (Cited on pp. 1033, 1034)

[14] J. Cheng and C.-W. Shu, *Positivity-preserving Lagrangian scheme for multi-material compressible flow*, J. Comput. Phys., 257 (2014), pp. 143–168. (Cited on p. 1034)

[15] Q. Cheng and J. Shen, *A new Lagrange multiplier approach for constructing structure preserving schemes,* II. *Bound preserving*, SIAM J. Numer. Anal., 60 (2022), pp. 970–998, https://doi.org/10.1137/21M144877X. (Cited on pp. 1033, 1034)

[16] Q. Cheng and J. Shen, *A new Lagrange multiplier approach for constructing structure preserving schemes,* I. *Positivity preserving*, Comput. Methods Appl. Mech. Engrg., 391 (2022), art. 114585. (Cited on pp. 1033, 1034)

[17] Y. Cheng, I. Gamba, and J. Proft, *Positivity-preserving discontinuous Galerkin schemes for linear Vlasov-Boltzmann transport equations*, Math. Comp., 81 (2012), pp. 153–190. (Cited on pp. 1033, 1034)

[18] Y. Cheng, F. Li, J. Qiu, and L. Xu, *Positivity-preserving DG and central DG methods for ideal MHD equations*, J. Comput. Phys., 238 (2013), pp. 255–280. (Cited on p. 1034)

[19] A. Chertock and A. Kurganov, *A second-order positivity preserving central-upwind scheme for chemotaxis and haptotaxis models*, Numer. Math., 111 (2008), pp. 169–205. (Cited on p. 1033)

[20] A. J. Christlieb, Y. Liu, Q. Tang, and Z. Xu, *High order parametrized maximum-principle-preserving and positivity-preserving WENO schemes on unstructured meshes*, J. Comput. Phys., 281 (2015), pp. 334–351. (Cited on p. 1034)

[21] A. J. Christlieb, Y. Liu, Q. Tang, and Z. Xu, *Positivity-preserving finite difference weighted ENO schemes with constrained transport for ideal magnetohydrodynamic equations*, SIAM J. Sci. Comput., 37 (2015), pp. A1825–A1845, https://doi.org/10.1137/140971208. (Cited on pp. 1034, 1055)

[22] K. N. Chueh, C. C. Conley, and J. A. Smoller, *Positively invariant regions for systems of nonlinear diffusion equations*, Indiana Univ. Math. J., 26 (1977), pp. 373–392. (Cited on p. 1033)

[23] P. G. Ciarlet and P.-A. Raviart, *Maximum principle and uniform convergence for the finite element method*, Comput. Methods Appl. Mech. Engrg., 2 (1973), pp. 17–31. (Cited on p. 1033)

[24] J. Droniou and C. Le Potier, *Construction and convergence study of schemes preserving the elliptic local maximum principle*, SIAM J. Numer. Anal., 49 (2011), pp. 459–490, https://doi.org/10.1137/090770849. (Cited on p. 1033)

[25] J. Du, C. Wang, C. Qian, and Y. Yang, *High-order bound-preserving discontinuous Galerkin methods for stiff multispecies detonation*, SIAM J. Sci. Comput., 41 (2019), pp. B250–B273, https://doi.org/10.1137/18M122265X. (Cited on pp. 1033, 1034, 1067)

[26] Q. Du, Z. Huang, and P. G. LeFloch, *Nonlocal conservation laws. A new class of monotonicity-preserving models*, SIAM J. Numer. Anal., 55 (2017), pp. 2465–2489, https://doi.org/10.1137/16M1105372. (Cited on p. 1035)

[27] Q. Du, L. Ju, X. Li, and Z. Qiao, *Maximum principle preserving exponential time differencing schemes for the nonlocal Allen–Cahn equation*, SIAM J. Numer. Anal., 57 (2019), pp. 875–898, https://doi.org/10.1137/18M118236X. (Cited on p. 1033)

[28] Q. Du, L. Ju, X. Li, and Z. Qiao, *Maximum bound principles for a class of semilinear parabolic equations and exponential time-differencing schemes*, SIAM Rev., 63 (2021), pp. 317–359, https://doi.org/10.1137/19M1243750. (Cited on pp. 1033, 1034)

[29] D. J. Estep, M. G. Larson, and R. D. Williams, *Estimating the error of numerical solutions of systems of reaction-diffusion equations*, Mem. Amer. Math. Soc., 146 (696) (2000). (Cited on pp. 1033, 1034)

[30] D. J. Estep, S. M. V. Lunel, and R. D. Williams, *Analysis of shear layers in a fluid with temperature-dependent viscosity*, J. Comput. Phys., 173 (2001), pp. 17–60. (Cited on p. 1034)

[31] D. J. Estep and A. M. Stuart, *The dynamical behavior of the discontinuous Galerkin method and related difference schemes*, Math. Comp., 71 (2002), pp. 1075–1103. (Cited on p. 1033)

[32] I. Faragó and R. Horváth, *Discrete maximum principle and adequate discretizations of linear parabolic problems*, SIAM J. Sci. Comput., 28 (2006), pp. 2313–2336, https://doi.org/10.1137/050627241. (Cited on pp. 1033, 1034)

[33] I. Faragó, J. Karátson, and S. Korotov, *Discrete maximum principles for nonlinear parabolic PDE systems*, IMA J. Numer. Anal., 32 (2012), pp. 1541–1573. (Cited on pp. 1033, 1034)

[34] F. Frank, A. Rupp, and D. Kuzmin, *Bound-preserving flux limiting schemes for DG discretizations of conservation laws with applications to the Cahn–Hilliard equation*, Comput. Methods Appl. Mech. Engrg., 359 (2020), art. 112665. (Cited on p. 1033)

[35] P. Fu, F. Li, and Y. Xu, *Globally divergence-free discontinuous Galerkin methods for ideal magnetohydrodynamic equations*, J. Sci. Comput., 77 (2018), pp. 1621–1659. (Cited on p. 1062)

[36] Z. Gao and J. Wu, *A second-order positivity-preserving finite volume scheme for diffusion equations on general meshes*, SIAM J. Sci. Comput., 37 (2015), pp. A420–A438, https://doi.org/10.1137/140972470. (Cited on p. 1033)

[37] S. K. Godunov, *Symmetric form of the equations of magnetohydrodynamics*, Numer. Methods Mech. Continuum Med., 1 (1972), pp. 26–34. (Cited on p. 1063)

[38] S. Gottlieb, C.-W. Shu, and E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Rev., 43 (2001), pp. 89–112, https://doi.org/10.1137/S003614450036757X. (Cited on pp. 1060, 1064)

[39] J. Gressier, P. Villedieu, and J.-M. Moschetta, *Positivity of flux vector splitting schemes*, J. Comput. Phys., 155 (1999), pp. 199–220. (Cited on pp. 1035, 1051)

[40] P. M. Gruber, *Convex and Discrete Geometry*, Grundlehren Math. Wiss. 336, Springer Science & Business Media, 2007. (Cited on p. 1042)

[41] J.-L. Guermond, M. Maier, B. Popov, and I. Tomas, *Second-order invariant domain preserving approximation of the compressible Navier–Stokes equations*, Comput. Methods Appl. Mech. Engrg., 375 (2021), art. 113608. (Cited on p. 1033)

[42] J.-L. Guermond, M. Nazarov, B. Popov, and I. Tomas, *Second-order invariant domain*

preserving approximation of the Euler equations using convex limiting, SIAM J. Sci. Comput., 40 (2018), pp. A3211–A3239, https://doi.org/10.1137/17M1149961. (Cited on pp. 1033, 1034)

[43] J.-L. Guermond and B. Popov, *Fast estimation from above of the maximum wave speed in the Riemann problem for the Euler equations*, J. Comput. Phys., 321 (2016), pp. 908–926. (Cited on p. 1051)

[44] J.-L. Guermond and B. Popov, *Invariant domains and first-order continuous finite element approximation for hyperbolic systems*, SIAM J. Numer. Anal., 54 (2016), pp. 2466–2489, https://doi.org/10.1137/16M1074291. (Cited on pp. 1033, 1034)

[45] J.-L. Guermond and B. Popov, *Invariant domains and second-order continuous finite element approximation for scalar conservation equations*, SIAM J. Numer. Anal., 55 (2017), pp. 3120–3146, https://doi.org/10.1137/16M1106560. (Cited on p. 1034)

[46] J.-L. Guermond, B. Popov, and I. Tomas, *Invariant domain preserving discretization-independent schemes and convex limiting for hyperbolic systems*, Comput. Methods Appl. Mech. Engrg., 347 (2019), pp. 143–175. (Cited on p. 1034)

[47] H. Guo and Y. Yang, *Bound-preserving discontinuous Galerkin method for compressible miscible displacement in porous media*, SIAM J. Sci. Comput., 39 (2017), pp. A1969–A1990, https://doi.org/10.1137/16M1101313. (Cited on p. 1033)

[48] D. Hoff, *Invariant regions for systems of conservation laws*, Trans. Amer. Math. Soc., 289 (1985), pp. 591–610. (Cited on p. 1033)

[49] J. Hu, R. Shu, and X. Zhang, *Asymptotic-preserving and positivity-preserving implicit-explicit schemes for the stiff BGK equation*, SIAM J. Numer. Anal., 56 (2018), pp. 942–973, https://doi.org/10.1137/17M1144362. (Cited on p. 1033)

[50] X. Y. Hu, N. A. Adams, and C.-W. Shu, *Positivity-preserving method for high-order conservative schemes solving compressible Euler equations*, J. Comput. Phys., 242 (2013), pp. 169–180. (Cited on pp. 1033, 1034)

[51] F. Huang and J. Shen, *Bound/positivity preserving and energy stable scalar auxiliary variable schemes for dissipative systems: Applications to Keller–Segel and Poisson–Nernst–Planck equations*, SIAM J. Sci. Comput., 43 (2021), pp. A1832–A1857, https://doi.org/10.1137/20M1365417. (Cited on pp. 1033, 1034)

[52] J. Huang and C.-W. Shu, *A second-order asymptotic-preserving and positivity-preserving discontinuous Galerkin scheme for the Kerr–Debye model*, Math. Models Methods Appl. Sci., 27 (2017), pp. 549–579. (Cited on p. 1033)

[53] A. R. Humphries and A. M. Stuart, *Runge–Kutta methods for dissipative and gradient dynamical systems*, SIAM J. Numer. Anal., 31 (1994), pp. 1452–1485, https://doi.org/10.1137/0731075. (Cited on p. 1033)

[54] A. Jameson, *Positive schemes and shock modelling for compressible flows*, Internat. J. Numer. Methods Fluids, 20 (1995), pp. 743–776. (Cited on p. 1033)

[55] J. W. Jerome, *A trapping principle and convergence result for finite element approximate solutions of steady reaction/diffusion systems*, Numer. Math., 109 (2008), pp. 121–142. (Cited on p. 1033)

[56] Y. Jiang and H. Liu, *Invariant-region-preserving DG methods for multi-dimensional hyperbolic conservation law systems, with an application to compressible Euler equations*, J. Comput. Phys., 373 (2018), pp. 385–409. (Cited on p. 1033)

[57] Y. Jiang and H. Liu, *An invariant-region-preserving limiter for DG schemes to isentropic Euler equations*, Numer. Methods Partial Differential Equations, 35 (2019), pp. 5–33. (Cited on p. 1033)

[58] D. A. Jones, A. M. Stuart, and E. S. Titi, *Persistence of invariant sets for dissipative evolution equations*, J. Math. Anal. Appl., 219 (1998), pp. 479–502. (Cited on p. 1033)

[59] L. Ju, X. Li, Z. Qiao, and J. Yang, *Maximum bound principle preserving integrating factor Runge–Kutta methods for semilinear parabolic equations*, J. Comput. Phys., 439 (2021), art. 110405. (Cited on p. 1034)

[60] A. Jüngel and R. Pinnau, *A positivity-preserving numerical scheme for a nonlinear fourth order parabolic system*, SIAM J. Numer. Anal., 39 (2001), pp. 385–406, https://doi.org/10.1137/S0036142900369362. (Cited on p. 1033)

[61] B. Khobalatte and B. Perthame, *Maximum principle on the entropy and second-order kinetic schemes*, Math. Comp., 62 (1994), pp. 119–131. (Cited on p. 1035)

[62] 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), pp. 323–338. (Cited on p. 1064)

[63] D. Kuzmin, M. Möller, J. N. Shadid, and M. Shashkov, *Failsafe flux limiting and constrained data projections for equations of gas dynamics*, J. Comput. Phys., 229 (2010),

pp. 8766–8779. (Cited on p. 1033)

[64] I. E. Leonard and J. E. Lewis, *Geometry of Convex Sets*, John Wiley & Sons, Hoboken, NJ, 2015. (Cited on p. 1042)

[65] B. Li, J. Yang, and Z. Zhou, *Arbitrarily high-order exponential cut-off methods for preserving maximum principle of parabolic equations*, SIAM J. Sci. Comput., 42 (2020), pp. A3957–A3978, https://doi.org/10.1137/20M1333456. (Cited on p. 1034)

[66] F. Li and C.-W. Shu, *Locally divergence-free discontinuous Galerkin methods for MHD equations*, J. Sci. Comput., 22 (2005), pp. 413–442. (Cited on p. 1062)

[67] H. Li, S. Xie, and X. Zhang, *A high order accurate bound-preserving compact finite difference scheme for scalar convection diffusion equations*, SIAM J. Numer. Anal., 56 (2018), pp. 3308–3345, https://doi.org/10.1137/18M1208551. (Cited on p. 1033)

[68] H. Li and X. Zhang, *On the monotonicity and discrete maximum principle of the finite difference implementation of C^0-Q^2 finite element method*, Numer. Math., 145 (2020), pp. 437–472. (Cited on p. 1035)

[69] J. Li, X. Li, L. Ju, and X. Feng, *Stabilized integrating factor Runge–Kutta method and unconditional preservation of maximum bound principle*, SIAM J. Sci. Comput., 43 (2021), pp. A1780–A1802, https://doi.org/10.1137/20M1340678. (Cited on p. 1033)

[70] M. Li, Y. Cheng, J. Shen, and X. Zhang, *A bound-preserving high order scheme for variable density incompressible Navier–Stokes equations*, J. Comput. Phys., 425 (2021), art. 109906. (Cited on p. 1033)

[71] X. H. Li, C.-W. Shu, and Y. Yang, *Local discontinuous Galerkin method for the Keller–Segel chemotaxis model*, J. Sci. Comput., 73 (2017), pp. 943–967. (Cited on pp. 1033, 1034)

[72] C. Liang and Z. Xu, *Parametrized maximum principle preserving flux limiters for high order schemes solving multi-dimensional scalar hyperbolic conservation laws*, J. Sci. Comput., 58 (2014), pp. 41–60. (Cited on p. 1034)

[73] H.-L. Liao, T. Tang, and T. Zhou, *On energy stable, maximum-principle preserving, second-order BDF scheme with variable steps for the Allen–Cahn equation*, SIAM J. Numer. Anal., 58 (2020), pp. 2294–2314, https://doi.org/10.1137/19M1289157. (Cited on p. 1033)

[74] D. Ling, J. Duan, and H. Tang, *Physical-constraints-preserving Lagrangian finite volume schemes for one- and two-dimensional special relativistic hydrodynamics*, J. Comput. Phys., 396 (2019), pp. 507–543. (Cited on p. 1035)

[75] C. Liu, C. Wang, S. Wise, X. Yue, and S. Zhou, *A positivity-preserving, energy stable and convergent numerical scheme for the Poisson–Nernst–Planck system*, Math. Comp., 90 (2021), pp. 2071–2106. (Cited on p. 1033)

[76] H. Liu and H. Yu, *Maximum-principle-satisfying third order discontinuous Galerkin schemes for Fokker–Planck equations*, SIAM J. Sci. Comput., 36 (2014), pp. A2296–A2325, https://doi.org/10.1137/130935161. (Cited on p. 1033)

[77] J.-G. Liu, L. Wang, and Z. Zhou, *Positivity-preserving and asymptotic preserving method for 2D Keller–Segal equations*, Math. Comp., 87 (2018), pp. 1165–1189. (Cited on pp. 1033, 1034)

[78] S. Mabuza, J. N. Shadid, and D. Kuzmin, *Local bounds preserving stabilization for continuous Galerkin discretization of hyperbolic systems*, J. Comput. Phys., 361 (2018), pp. 82–110. (Cited on p. 1033)

[79] A. K. Meena, H. Kumar, and P. Chandrashekar, *Positivity-preserving high-order discontinuous Galerkin schemes for ten-moment Gaussian closure equations*, J. Comput. Phys., 339 (2017), pp. 370–395. (Cited on pp. 1035, 1039)

[80] A. K. Meena, R. Kumar, and P. Chandrashekar, *Positivity-preserving finite difference WENO scheme for ten-moment equations with source term*, J. Sci. Comput., 82 (2020), pp. 1–37. (Cited on pp. 1035, 1039)

[81] C. Niculescu and L.-E. Persson, *Convex Functions and Their Applications*, Springer, 2006. (Cited on p. 1043)

[82] R. Nochetto and L. Wahlbin, *Positivity preserving finite element approximation*, Math. Comp., 71 (2002), pp. 1405–1419. (Cited on p. 1033)

[83] P. Öffner and D. Torlo, *Arbitrary high-order, conservative and positivity preserving Patankar-type deferred correction schemes*, Appl. Numer. Math., 153 (2020), pp. 15–34. (Cited on p. 1033)

[84] E. Olbrant, C. D. Hauck, and M. Frank, *A realizability-preserving discontinuous Galerkin method for the M1 model of radiative transfer*, J. Comput. Phys., 231 (2012), pp. 5612–5639. (Cited on pp. 1035, 1038)

[85] B. Perthame, *Second-order Boltzmann schemes for compressible Euler equations in one and two space dimensions*, SIAM J. Numer. Anal., 29 (1992), pp. 1–19, https://doi.org/10.

1137/0729001. (Cited on p. 1035)

[86] B. Perthame and C.-W. Shu, *On positivity preserving finite volume schemes for Euler equations*, Numer. Math., 73 (1996), pp. 119–130. (Cited on pp. 1033, 1051)

[87] T. Qin, C.-W. Shu, and Y. Yang, *Bound-preserving discontinuous Galerkin methods for relativistic hydrodynamics*, J. Comput. Phys., 315 (2016), pp. 323–347. (Cited on pp. 1034, 1035, 1038)

[88] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 2015. (Cited on pp. 1042, 1046)

[89] J. A. Rossmanith and D. C. Seal, *A positivity-preserving high-order semi-Lagrangian discontinuous Galerkin scheme for the Vlasov–Poisson equations*, J. Comput. Phys., 230 (2011), pp. 6203–6232. (Cited on p. 1033)

[90] D. Serre, *Systems of Conservation Laws. 2. Geometric Structures, Oscillations, and Initial-Boundary Value Problems*, Cambridge University Press, Cambridge, UK, 2000; translated from the 1996 French original by I. N. Sneddon. (Cited on p. 1033)

[91] J. Shen, T. Tang, and J. Yang, *On the maximum principle preserving schemes for the generalized Allen–Cahn equation*, Commun. Math. Sci., 14 (2016), pp. 1517–1534. (Cited on p. 1033)

[92] J. Shen and J. Xu, *Unconditionally bound preserving and energy dissipative schemes for a class of Keller–Segel equations*, SIAM J. Numer. Anal., 58 (2020), pp. 1674–1695, https://doi.org/10.1137/19M1246705. (Cited on p. 1033)

[93] J. Shen and J. Xu, *Unconditionally positivity preserving and energy dissipative schemes for Poisson–Nernst–Planck equations*, Numer. Math., 148 (2021), pp. 671–697. (Cited on p. 1033)

[94] C.-W. Shu, *Bound-preserving high-order schemes for hyperbolic equations: Survey and recent developments*, in Theory, Numerics and Applications of Hyperbolic Problems II, C. Klingenberg and M. Westdickenberg, eds., Springer International, Cham, 2018, pp. 591–603. (Cited on pp. 1033, 1034)

[95] J. Smoller, *Shock Waves and Reaction-Diffusion Equations*, Grundlehren Math. Wiss. 258, Springer Science & Business Media, 2012. (Cited on p. 1033)

[96] M. Spivak, *Calculus on Manifolds*, W. A. Benjamin, New York, 1965. (Cited on p. 1047)

[97] A. Stuart and A. R. Humphries, *Dynamical Systems and Numerical Analysis*, Vol. 2, Cambridge University Press, 1998. (Cited on p. 1033)

[98] A. M. Stuart, *Numerical analysis of dynamical systems*, Acta Numer., 3 (1994), pp. 467–572. (Cited on p. 1033)

[99] E. Tadmor, *A minimum entropy principle in the gas dynamics equations*, Appl. Numer. Math., 2 (1986), pp. 211–219. (Cited on p. 1037)

[100] H.-Z. Tang and K. Xu, *Positivity-preserving analysis of explicit and implicit Lax–Friedrichs schemes for compressible Euler equations*, J. Sci. Comput., 15 (2000), pp. 19–28. (Cited on p. 1035)

[101] T. Tang and K. Xu, *Gas-kinetic schemes for the compressible Euler equations: Positivity-preserving analysis*, Z. Angew. Math. Phys., 50 (1999), pp. 258–281. (Cited on pp. 1035, 1051, 1052)

[102] J. J. W. van der Vegt, Y. Xia, and Y. Xu, *Positivity preserving limiters for time-implicit higher order accurate discontinuous Galerkin discretizations*, SIAM J. Sci. Comput., 41 (2019), pp. A2037–A2063, https://doi.org/10.1137/18M1227998. (Cited on pp. 1033, 1034)

[103] K. Wu, *Design of provably physical-constraint-preserving methods for general relativistic hydrodynamics*, Phys. Rev. D, 95 (2017), art. 103001. (Cited on p. 1033)

[104] K. Wu, *Positivity-preserving analysis of numerical schemes for ideal magnetohydrodynamics*, SIAM J. Numer. Anal., 56 (2018), pp. 2124–2147, https://doi.org/10.1137/18M1168017. (Cited on pp. 1034, 1035, 1036, 1039, 1040, 1051, 1055, 1061, 1062, 1066)

[105] K. Wu, *Minimum principle on specific entropy and high-order accurate invariant-region-preserving numerical methods for relativistic hydrodynamics*, SIAM J. Sci. Comput., 43 (2021), pp. B1164–B1197, https://doi.org/10.1137/21M1397994. (Cited on pp. 1038, 1056, 1066)

[106] K. Wu and C.-W. Shu, *A provably positive discontinuous Galerkin method for multidimensional ideal magnetohydrodynamics*, SIAM J. Sci. Comput., 40 (2018), pp. B1302–B1329, https://doi.org/10.1137/18M1168042. (Cited on pp. 1033, 1035, 1036, 1039, 1051, 1054, 1055, 1063, 1065, 1066)

[107] K. Wu and C.-W. Shu, *Provably positive high-order schemes for ideal magnetohydrodynamics: Analysis on general meshes*, Numer. Math., 142 (2019), pp. 995–1047. (Cited on pp. 1036, 1054, 1055, 1063, 1065, 1066)

[108] K. Wu and C.-W. Shu, *Provably physical-constraint-preserving discontinuous Galerkin methods for multidimensional relativistic MHD equations*, Numer. Math., 148 (2021), pp. 699–741. (Cited on pp. 1033, 1034, 1035, 1036, 1039, 1051, 1054, 1057, 1066)

[109] K. Wu and H. Tang, *High-order accurate physical-constraints-preserving finite difference WENO schemes for special relativistic hydrodynamics*, J. Comput. Phys., 298 (2015), pp. 539–564. (Cited on pp. 1033, 1034, 1035, 1038)

[110] K. Wu and H. Tang, *Admissible states and physical-constraints-preserving schemes for relativistic magnetohydrodynamic equations*, Math. Models Methods Appl. Sci., 27 (2017), pp. 1871–1928. (Cited on pp. 1035, 1036, 1039, 1057, 1061, 1066)

[111] K. Wu and H. Tang, *Physical-constraint-preserving central discontinuous Galerkin methods for special relativistic hydrodynamics with a general equation of state*, Astrophys. J. Suppl. Ser., 228 (2017), art. 3. (Cited on pp. 1033, 1034)

[112] K. Wu and Y. Xing, *Uniformly high-order structure-preserving discontinuous Galerkin methods for Euler equations with gravitation: Positivity and well-balancedness*, SIAM J. Sci. Comput., 43 (2021), pp. A472–A510, https://doi.org/10.1137/20M133782X. (Cited on p. 1034)

[113] Y. Xing, X. Zhang, and C.-W. Shu, *Positivity-preserving high order well-balanced discontinuous Galerkin methods for the shallow water equations*, Adv. Water Resour., 33 (2010), pp. 1476–1493. (Cited on pp. 1033, 1034)

[114] T. Xiong, J.-M. Qiu, and Z. Xu, *High order maximum-principle-preserving discontinuous Galerkin method for convection-diffusion equations*, SIAM J. Sci. Comput., 37 (2015), pp. A583–A608, https://doi.org/10.1137/140965326. (Cited on p. 1034)

[115] T. Xiong, J.-M. Qiu, and Z. Xu, *Parametrized positivity preserving flux limiters for the high order finite difference WENO scheme solving compressible Euler equations*, J. Sci. Comput., 67 (2016), pp. 1066–1088. (Cited on pp. 1033, 1034)

[116] Z. Xu, *Parametrized maximum principle preserving flux limiters for high order schemes solving hyperbolic conservation laws: One-dimensional scalar problem*, Math. Comp., 83 (2014), pp. 2213–2238. (Cited on pp. 1033, 1034)

[117] Z. Xu and X. Zhang, *Bound-preserving high-order schemes*, in Handbook of Numerical Analysis, Handb. Numer. Anal. 18, Elsevier, 2017, pp. 81–102. (Cited on p. 1034)

[118] D. Yuan, J. Cheng, and C.-W. Shu, *High order positivity-preserving discontinuous Galerkin methods for radiative transfer equations*, SIAM J. Sci. Comput., 38 (2016), pp. A2987–A3019, https://doi.org/10.1137/16M1061072. (Cited on pp. 1033, 1034)

[119] X. Zhang, *On positivity-preserving high order discontinuous Galerkin schemes for compressible Navier-Stokes equations*, J. Comput. Phys., 328 (2017), pp. 301–343. (Cited on pp. 1033, 1034, 1035, 1038, 1054, 1055, 1068)

[120] X. Zhang, Y. Liu, and C.-W. Shu, *Maximum-principle-satisfying high order finite volume weighted essentially nonoscillatory schemes for convection-diffusion equations*, SIAM J. Sci. Comput., 34 (2012), pp. A627–A658, https://doi.org/10.1137/110839230. (Cited on p. 1034)

[121] X. Zhang and C.-W. Shu, *On maximum-principle-satisfying high order schemes for scalar conservation laws*, J. Comput. Phys., 229 (2010), pp. 3091–3120. (Cited on pp. 1033, 1034, 1035, 1053, 1061, 1062, 1068)

[122] X. Zhang and C.-W. Shu, *On positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes*, J. Comput. Phys., 229 (2010), pp. 8918–8934. (Cited on pp. 1033, 1034, 1035, 1037, 1052, 1053, 1054, 1061, 1062, 1068)

[123] X. Zhang and C.-W. Shu, *Positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations with source terms*, J. Comput. Phys., 230 (2011), pp. 1238–1248. (Cited on pp. 1034, 1054, 1055)

[124] X. Zhang and C.-W. Shu, *A minimum entropy principle of high order schemes for gas dynamics equations*, Numer. Math., 121 (2012), pp. 545–563. (Cited on p. 1038)

[125] X. Zhang, Y. Xia, and C.-W. Shu, *Maximum-principle-satisfying and positivity-preserving high order discontinuous Galerkin schemes for conservation laws on triangular meshes*, J. Sci. Comput., 50 (2012), pp. 29–62. (Cited on pp. 1034, 1061)

[126] J. Zhao and H. Tang, *Runge-Kutta discontinuous Galerkin methods for the special relativistic magnetohydrodynamics*, J. Comput. Phys., 343 (2017), pp. 33–72. (Cited on p. 1064)