Journal of Computational Physics 295 (2015) 715–739

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Entropy-bounded discontinuous Galerkin scheme for Euler equations

Yu Lv *, Matthias Ihme

Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA

ARTICLE INFO

Article history:
Received 17 November 2014
Received in revised form 17 March 2015
Accepted 17 April 2015
Available online 28 April 2015

Keywords:
Discontinuous Galerkin
Shock-capturing
Limiter
Entropy principle

ABSTRACT

An entropy-bounded Discontinuous Galerkin (EBDG) scheme is proposed in which the solution is regularized by constraining the entropy. The resulting scheme is able to stabilize the solution in the vicinity of discontinuities and retains the optimal accuracy for smooth solutions. The properties of the limiting operator according to the entropy-minimum principle are proofed, and an optimal CFL-criterion is derived. We provide a rigorous description for locally imposing entropy constraints to capture multiple discontinuities. Significant advantages of the EBDG-scheme are the general applicability to arbitrary high-order elements and its simple implementation for multi-dimensional configurations. Numerical tests confirm the properties of the scheme, and particular focus is attributed to the robustness in treating discontinuities on arbitrary meshes.

1. Introduction

The stabilization of solutions near flow-field discontinuities remains an open problem to the discontinuous Galerkin (DG) community. Considerable progress has been made on the development of limiters for two-dimensional quadrilateral and triangular elements. These limiters can be categorized into three classes. Methods that limit the solution using information about the slope along certain spatial directions [1,2] fall in the first class. The second class of limiters extends this idea by limiting based on the moments of the solution [3,4], and schemes in which the DG-solution is projected onto a WENO [5–7] or Hermit WENO (HWENO) [8] representation fall in the last category. Although these limiters show promising results for canonical test cases on regular elements and structured mesh partitions, the following two issues related to practical applications have not been clearly answered:

- How can discontinuous solutions be regularized on multi-dimensional curved high-order elements?
- How can non-physical solutions that are triggered by strong discontinuities and geometric singularities be avoided?

The present work attempts to simultaneously address both of these questions.

Recently, positivity-preserving DG-schemes have been developed for the treatment of flow-field discontinuities, and relevant contributions are by Zhang and Shu [9–11]. The positivity preserving method provides a robust framework with provable L_1-stability, preventing the appearance of negative pressure and density. Resulting algorithmic modifications are minimal, and these schemes have been used in simulations of detonation systems with complex reaction chemistry [12,13].

* Corresponding author.
E-mail addresses: ylv@stanford.edu (Y. Lv), mihme@stanford.edu (M. Ihme).

http://dx.doi.org/10.1016/j.jcp.2015.04.026
0021-9991/

Motivated by these attractive properties, the present work aims at developing an algorithm that avoids non-physical solutions on arbitrary elements and multi-dimensional spatial representations. The resulting scheme that will be developed in this work has the following properties: First, by invoking the entropy principle, solutions are constrained by a local entropy bound. Second, a general implementation on arbitrary elements is proposed without restriction to a specific quadrature rule. Third, the entropy constraint is imposed on the solutions through few algebraic operations, thereby avoiding the computationally expensive inversion of a nonlinear system. Fourth, a method for the evaluation of an optimal CFL-criterion is derived, which is applicable to general polynomial orders and arbitrary element types.

The remainder of this paper has the following structure. The governing equations and the discretization are summarized in the next two sections. The entropy-bounded DG (EBDG) formulation is presented in Section 4, and the derivation of the CFL-constraint and the limiting operator are presented. This analysis is performed by considering a one-dimensional setting, and the generalization to multi-dimensional and arbitrary elements is presented in Section 5. Section 6 is concerned with the evaluation of the entropy-bounded DG-scheme, and a detailed description of the algorithmic implementation is given in Section 7. The EBDG-method is demonstrated by considering several test cases, and the accuracy and stability are examined in Section 8. The paper finishes with conclusions.

2. Governing equations

We consider a system of conservation equations,

$$\frac{\partial \mathsf{U}}{\partial t} + \nabla \cdot \mathsf{F} = 0 \qquad \text{in } \Omega\,, \tag{1}$$

where the solution variable $\mathsf{U} : \mathbb{R}^{N_d} \times \mathbb{R} \to \mathbb{R}^{N_v}$ and the flux term $\mathsf{F} : \mathbb{R}^{N_v} \to \mathbb{R}^{N_v \times N_d}$. Here, N_d denotes the spatial dimension and N_v is the dimension of the solution vector. For the Euler equations, U and F take the form:

$$\mathsf{U}(x,t) = (\rho, \rho u, \rho e)^T\,, \tag{2a}$$

$$\mathsf{F}(\mathsf{U}) = (\rho u, \rho u \otimes u + p\mathrm{I}, u(\rho e + p))^T\,, \tag{2b}$$

where t is the time, $x \in \mathbb{R}^{N_d}$ is the spatial coordinate vector, ρ is the density, $u \in \mathbb{R}^{N_d}$ is the velocity vector, e is the specific total energy, and p is the pressure. Eq. (1) is closed with the ideal gas law:

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

in which γ is the ratio of specific heats, which, for the present work, is set to a constant value of $\gamma = 1.4$. Here and in the following, we use $|\cdot|$ to represent the Euclidean norm. With this, we define the local maximum characteristic speed as:

$$\nu = |u| + c \qquad \text{with} \qquad c = \sqrt{\frac{\gamma p}{\rho}}\,, \tag{4}$$

where c is the speed of sound.

Because of the presence of discontinuities in the solution of Eq. (1), we seek a weak solution that satisfies physical principles. This is the so-called entropy solution. By introducing $\mathcal{U}$ as a convex function of U with $\mathcal{U} : \mathbb{R}^{N_v} \to \mathbb{R}$, Lax [14] showed that the entropy solution of Eq. (1) satisfies the following inequality:

$$\frac{\partial \mathcal{U}}{\partial t} + \nabla \cdot \mathcal{F} \leq 0\,, \tag{5}$$

where $\mathcal{F} : \mathbb{R}^{N_v} \to \mathbb{R}^{N_d}$ is the corresponding flux of $\mathcal{U}$. The consistency condition between Eqs. (1) and (5) requires [14]:

$$\left(\frac{\partial \mathcal{U}}{\partial \mathsf{U}}\right)^T \frac{\partial \mathsf{F}}{\partial \mathsf{U}} = \frac{\partial \mathcal{F}}{\partial \mathsf{U}}\,. \tag{6}$$

The weak solution of Eq. (1) that satisfies this additional condition for the entropy–entropy flux pair $(\mathcal{U}, \mathcal{F})$ is called an entropy solution. With this definition, Eq. (5) is commonly called entropy inequality or entropy condition, and $\mathcal{U}$ is called entropy function. A familiar example for gas-dynamic applications is to relate $\mathcal{U}$ to the physical entropy s with:

$$s = \ln(p) - \gamma \ln(\rho) + s_0\,, \tag{7}$$

where s_0 is the reference entropy. The corresponding definitions for the entropy function and entropy flux in the context of the Euler system are [15]:

$$(\mathcal{U}, \mathcal{F}) = (-\rho s, -\rho s u)\,. \tag{8}$$

Note that Eq. (7) directly provides a constraint on the positivity of pressure p and density ρ.

3. Discontinuous Galerkin discretization

We consider the problem to be posed on the domain Ω with boundary $\partial\Omega$. A mesh partition is defined as $\Omega = \cup_{e=1}^{N_e}\Omega_e$, where Ω_e corresponds to a discrete element of this partition. The edge of element Ω_e is defined as $\partial\Omega_e$. In order to distinguish different sides of the edge, the superscripts "+" and "−" are used to denote the interior and exterior, respectively. We define a global space of test functions as

$$\mathcal{V} = \oplus_{e=1}^{N_e}\mathcal{V}_e\,, \qquad \mathcal{V}_e = \mathrm{span}\{\varphi_n(\Omega_e)\}_{n=1}^{N_p}\,, \tag{9}$$

where φ_n is the nth polynomial basis, and N_p is the number of bases. On the space $\mathcal{V}_e$ we seek an approximate solution to Eq. (1) of the form:

$$\mathsf{U} \simeq U = \oplus_{e=1}^{N_e}U_e\,, \qquad U_e \in \mathcal{V}_e\,, \tag{10}$$

where the solution vector U_e on each individual element takes the general form

$$U_e(x,t) = \sum_{m=1}^{N_p}\widetilde{U}_{e,m}(t)\varphi_m(x)\,, \tag{11}$$

and the unknown vector of basic coefficients $\widetilde{U}_{e,m} \in \mathbb{R}^{N_v\times N_p}$ is obtained from the discretized weak solution of Eq. (1):

$$\frac{d\widetilde{U}_{e,m}}{dt}\int_{\Omega_e}\varphi_n\varphi_m d\Omega - \int_{\Omega_e}\nabla\varphi_n\cdot F(U_e)d\Omega + \int_{\partial\Omega_e}\varphi_n^+\widehat{F}(U_e^+,U_e^-,\widehat{\mathrm{n}})d\Gamma = 0\,, \tag{12}$$

$\forall\varphi_n$ with $n = 1,\ldots,N_p$. The numerical Riemann flux $\widehat{F}$ is evaluated based on the states at both sides of the interface $\partial\Omega_e$ and the outward-pointing normal vector $\widehat{\mathrm{n}}$. It is of interest to note that for the particular case of $N_p = 1$ and $\varphi_1 = 1$, the weak form reduces to the classical first-order finite-volume (FV) discretization. It can also be seen that the DG-scheme does not rely on a specific type of basis functions. Since the following derivation is based on this mathematical property, we introduce the lemma and its proof is obvious.

Lemma 1. *A polynomial P,*

$$P(x) = \sum_{m=1}^{N_p}\widetilde{P}_m\varphi_m(x) \qquad \textit{for } x\in\Omega_e\,, \tag{13}$$

with a set of polynomial bases $\{\varphi_m(x), m = 1,\ldots,N_p\}$, can be exactly interpolated by a Lagrangian polynomial of N_p points $\{y_n\in\Omega_e,\ n = 1,\ldots,N_p\}$ under the condition that $\left[\varphi_m(y_n)\right]$ is non-singular:

$$P(x) = \sum_{n=1}^{N_p}P(y_n)\phi_n(x) \qquad \textit{for } x\in\Omega_e\,. \tag{14}$$

Remark 1. The significance of this lemma is that it provides a description to convert any basis set to a Lagrangian basis set with N_p interpolation points, as long as they are located at general positions. To facilitate the following derivation, we choose the points y_n with $n = 1,\ldots,N_p$ from the N_q quadrature points [16]. According to the accuracy requirement of the quadrature scheme for Eq. (12), $N_q \geq N_p$ is alway true.

4. Entropy principle and entropy-bounded discontinuous Galerkin method

In this section, we review the entropy principle by considering a three-point FV-setting. Then, we will explore how to extend this principle to a DG-scheme, which leads to the concept of entropy boundedness. In order to enable the implementation of this concept, two important ingredients will be discussed, namely a time-step constraint and a limiting operator. After conducting numerical analyses by considering a one-dimensional configuration, we will extend the entropy boundedness to multi-dimensional and arbitrary element types. The dimensional generality, geometric adaptability and simple implementation are major advantages of the resulting entropy-bounded DG-method.

4.1. Preliminaries and related work

To illustrate the entropy principle, we consider a local Lax–Friedrichs flux, which can be written as:

$$\widehat{F}(U_L,U_R,\widehat{\mathrm{n}}) = \frac{1}{2}\left(F(U_L)+F(U_R)\right)\cdot\widehat{\mathrm{n}} - \frac{1}{2}\lambda(U_R-U_L)\,, \tag{15}$$

and

$$\lambda \geq \max_{k\in\{L,R\}} \nu(U_k)$$

is the dissipation coefficient. Note that this flux function satisfies consistency: $\widehat{F}(U,U,\widehat{\mathrm{n}}) = F(U)\cdot\widehat{\mathrm{n}}$, conservation: $\widehat{F}(U_L,U_R,\widehat{\mathrm{n}}) = -\widehat{F}(U_R,U_L,-\widehat{\mathrm{n}})$, and Lipschitz-continuity. In the following, we consider the simplest case of a DGP0 scheme, with $N_p = 1$, in a one-dimensional setting. This formulation is consistent with the classical three-point FV-discretization. For $x\in\Omega_e = [x_{e-1/2},\ x_{e+1/2}]$, the discretized solution to Eq. (1) can be written as:

$$\widetilde{U}_e(t+\Delta t) = \widetilde{U}_e - \frac{\Delta t}{h}\left(\widehat{F}(\widetilde{U}_e,\widetilde{U}_{e-1},-1) + \widehat{F}(\widetilde{U}_e,\widetilde{U}_{e+1},1)\right), \tag{16}$$

where $\widetilde{U}_e$ is the basis coefficient, which is identical to the piecewise constant approximation to the exact solution in Ω_e. In the following, we introduce $\widetilde{U}_e^{\Delta t}$ to denote the solution vector $\widetilde{U}_e(t+\Delta t)$, and use the superscript Δt to denote a temporally updated quantity at $t+\Delta t$. With the numerical flux given in Eq. (15), this discretization preserves the positivity of pressure and density under the CFL-condition [9,17]:

$$\frac{\Delta t\lambda}{h} \leq \frac{1}{2}. \tag{17}$$

In addition, it was discussed in [17] that Eq. (16) satisfies the discrete minimum entropy principle proposed by Tadmor [15]:

$$s(\widetilde{U}_e^{\Delta t}) \geq s_e^0(t) = \min_{j\in\{e-1,e,e+1\}} s(\widetilde{U}_j). \tag{18}$$

To show this property, we recall the discussion presented in [17] and rewrite Eq. (16) by splitting $\widetilde{U}_e(t+\Delta t)$ into two parts. For $x\in\Omega_e$, this is written as:

$$\widetilde{U}_e(t+\Delta t) = \frac{1}{2}\left(\widetilde{U}_{e,p1}^{\Delta t} + \widetilde{U}_{e,p2}^{\Delta t}\right), \tag{19a}$$

$$\widetilde{U}_{e,p1}^{\Delta t} = \widetilde{U}_e - \frac{\Delta t}{h}\left(F(\widetilde{U}_{e+1}) - \lambda_{e+1/2}\widetilde{U}_{e+1} - F(\widetilde{U}_e) + \lambda_{e+1/2}\widetilde{U}_e\right), \tag{19b}$$

$$\widetilde{U}_{e,p2}^{\Delta t} = \widetilde{U}_e + \frac{\Delta t}{h}\left(F(\widetilde{U}_{e-1}) + \lambda_{e-1/2}\widetilde{U}_{e-1} - F(\widetilde{U}_e) - \lambda_{e-1/2}\widetilde{U}_e\right), \tag{19c}$$

where $\widetilde{U}_{e,p1}^{\Delta t}$ and $\widetilde{U}_{e,p2}^{\Delta t}$ can be viewed as the P0-approximations to the solutions of the hyperbolic systems (under the CFL constraint of Eq. (17)):

$$\frac{\partial \mathsf{U}}{\partial t} + \left(\mathsf{F}'(\mathsf{U}) - \lambda_{e+1/2}\mathsf{I}\right)\frac{\partial \mathsf{U}}{\partial x} = 0, \tag{20a}$$

$$\frac{\partial \mathsf{U}}{\partial t} + \left(\mathsf{F}'(\mathsf{U}) + \lambda_{e-1/2}\mathsf{I}\right)\frac{\partial \mathsf{U}}{\partial x} = 0, \tag{20b}$$

with the exact (Godunov) flux. If we denote the exact solutions to Eqs. (20a) and (20b) as $\mathsf{U}_{p1}(x,t+\Delta t)$ and $\mathsf{U}_{p2}(x,t+\Delta t)$, respectively, then their P0-approximations in Ω_e yield $\widetilde{U}_{e,p1}^{\Delta t} = \frac{1}{h}\int_{x_{e-1/2}}^{x_{e+1/2}} \mathsf{U}_{p1}(x,t+\Delta t)dx$ and $\widetilde{U}_{e,p2}^{\Delta t} = \frac{1}{h}\int_{x_{e-1/2}}^{x_{e+1/2}} \mathsf{U}_{p2}(x,t+\Delta t)dx$. Both equation systems are obtained by imposing a constant shift on the characteristic speeds without modifying the characteristic variables. With these modifications, all characteristics in Eq. (20a) are right-running while those in Eq. (20b) are left-running. The corresponding entropy inequalities take the form:

$$\frac{\partial \mathcal{U}}{\partial t} + \frac{\partial}{\partial x}\left(\mathcal{F} - \lambda_{e+1/2}\mathcal{U}\right) \leq 0, \tag{21a}$$

$$\frac{\partial \mathcal{U}}{\partial t} + \frac{\partial}{\partial x}\left(\mathcal{F} + \lambda_{e-1/2}\mathcal{U}\right) \leq 0. \tag{21b}$$

Without loss of generality, we now consider Eq. (21a) and integrate over $[t,t+\Delta t]\times[x_{e-1/2},x_{e+1/2}]$, resulting in the following expression:

$$\int_{x_{e-1/2}}^{x_{e+1/2}} \mathcal{U}\big(\mathsf{U}_{p1}(x,t+\Delta t)\big)dx - \int_{x_{e-1/2}}^{x_{e+1/2}} \mathcal{U}(\widetilde{U}_e)dx + \int_t^{t+\Delta t}\left(\mathcal{F}(\mathsf{U}(x_{e+1/2},t)) - \lambda_{e+1/2}\mathcal{U}(\mathsf{U}(x_{e+1/2},t))\right)dt$$
$$- \int_t^{t+\Delta t}\left(\mathcal{F}(\mathsf{U}(x_{e-1/2},t)) - \lambda_{e-1/2}\mathcal{U}(\mathsf{U}(x_{e-1/2},t))\right)dt \leq 0. \tag{22}$$

Recognizing that all characteristics are right-running, the temporal integral can be evaluated exact since $\mathsf{U}(x_{e-1/2},t) = \widetilde{U}_e$ and $\mathsf{U}(x_{e+1/2},t) = \widetilde{U}_{e+1}$ under the condition of Eq. (17). Then by utilizing the convexity of $\mathcal{U}$ with respect to U, the following estimate is obtained:

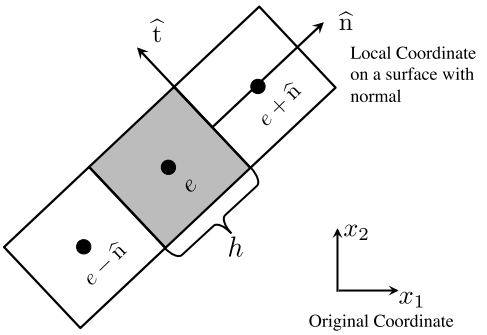


Fig. 1. Discretization for an arbitrarily oriented face in a quasi-1D configuration.

$$\mathcal{U}(\widetilde{U}^{\Delta t}_{e,p1}) = \mathcal{U}\left(\frac{1}{h}\int_{x_{e-1/2}}^{x_{e+1/2}} \mathsf{U}_{p1}(x,t+\Delta t)dx\right) \le \frac{1}{h}\int_{x_{e-1/2}}^{x_{e+1/2}} \mathcal{U}(\mathsf{U}_{p1}(x,t+\Delta t))dx$$
$$\le \mathcal{U}(\widetilde{U}_e) + \frac{\Delta t}{h}\left(\mathcal{F}(\widetilde{U}_e) - \lambda_{e+1/2}\mathcal{U}(\widetilde{U}_e)\right) - \frac{\Delta t}{h}\left(\mathcal{F}(\widetilde{U}_{e+1}) - \lambda_{e+1/2}\mathcal{U}(\widetilde{U}_{e+1})\right) .$$

With the definition of $(\mathcal{U},\mathcal{F})$, given in Eq. (8), it follows

$$s(\widetilde{U}^{\Delta t}_{e,p1}) \ge \frac{\rho_e}{\rho^{\Delta t}_{e,p1}}\left[1 - \frac{\Delta t}{h}\left(\lambda_{e+1/2} - u_e\right)\right] s(\widetilde{U}_e) + \frac{\rho_{e+1}}{\rho^{\Delta t}_{e,p1}}\frac{\Delta t}{h}\left(\lambda_{e+1/2} - u_{e+1}\right) s(\widetilde{U}_{e+1}) . \tag{23}$$

The constraint (17) ensures that the coefficients in front of $s(\widetilde{U}_e)$ and $s(\widetilde{U}_{e+1})$ are positive and sum to unity according to Eq. (19b). From these arguments directly follows:

$$s(\widetilde{U}^{\Delta t}_{e,p1}) \ge \min\{s(\widetilde{U}_e), s(\widetilde{U}_{e+1})\} , \tag{24}$$

and

$$s(\widetilde{U}^{\Delta t}_{e,p2}) \ge \min\{s(\widetilde{U}_{e-1}), s(\widetilde{U}_e)\} . \tag{25}$$

Combining these two relations with the quasi-concavity of the entropy s (Lemma 2.1 of [11]), the discrete minimum entropy principle of Eq. (18) is obtained.

The result above is obtained for a one-dimensional setting. To analyze the general multi-dimensional setting in the following, a similar conclusion for an arbitrarily oriented face has to be drawn. For this, we consider a general three-point system, which is shown in Fig. 1. A face with a normal $\hat{n}$ is considered and the tangential direction is defined by $\hat{t}$. Three FV cells are aligned along $\hat{n}$ with the indices $e-\hat{n}$, e and $e+\hat{n}$. With the first-order FV scheme, the solution vector in Ω_e is updated as

$$\widetilde{U}^{\Delta t}_e = \widetilde{U}_e - \frac{\Delta t}{h}\left(\widehat{F}(\widetilde{U}_e, \widetilde{U}_{e-\hat{n}}, -\hat{n}) + \widehat{F}(\widetilde{U}_e, \widetilde{U}_{e+\hat{n}}, \hat{n})\right) , \tag{26}$$

which is an augmented form of the one-dimensional system, Eq. (16). To show the entropy-boundedness of the solution $\widetilde{U}^{\Delta t}_e$ in this formula, we first define an auxiliary system in the local rotated coordinate frame $(\hat{n},\ \hat{t})$ with initial conditions at time t: $\forall j \in \{e-\hat{n}, e, e+\hat{n}\}$,

$${}^{\mathcal{R}}\widetilde{U}_j \equiv \left(\rho_j,\ \rho_j u_j\cdot\hat{n},\ \frac{p_j}{\gamma-1} + \frac{\rho_j(u_j\cdot\hat{n})^2}{2}\right)^T ,$$

where the momentum is taken by projecting the multi-dimensional solution onto the normal $\hat{n}$, and the total energy is defined by excluding the kinetic energy corresponding to the tangential velocity component, so that by construction we have $\rho({}^{\mathcal{R}}\widetilde{U}_j) = \rho(\widetilde{U}_j)$, $p({}^{\mathcal{R}}\widetilde{U}_j) = p(\widetilde{U}_j)$ and $s({}^{\mathcal{R}}\widetilde{U}_j) = s(\widetilde{U}_j)$. This auxiliary system is essentially a one-dimensional system along $\hat{n}$, which can be solved by advancing the time to $t+\Delta t$ under the CFL-constraint, Eq. (17), with $\lambda \ge \max \nu(U_j)$. Built on the above discussion, the updated solution preserves positivity of density and pressure and retains a bounded entropy value. For entropy, this means that

$$s({}^{\mathcal{R}}\widetilde{U}^{\Delta t}_e) \ge \min_{j\in\{e-\hat{n},e,e+\hat{n}\}} s({}^{\mathcal{R}}\widetilde{U}_j) = \min_{j\in\{e-\hat{n},e,e+\hat{n}\}} s(\widetilde{U}_j) . \tag{27}$$

To examine the properties of the solution $\widetilde{U}^{\Delta t}_e$ in the multi-dimensional system, the link between $\widetilde{U}^{\Delta t}_e$ and ${}^{\mathcal{R}}\widetilde{U}^{\Delta t}_e$ needs to be established. With some algebraic manipulation, we can obtain the following identities:

$$\rho(\widetilde{U}_e^{\Delta t}) = \rho({}^{\mathcal{R}}\widetilde{U}_e^{\Delta t}) \,, \tag{28a}$$

$$p(\widetilde{U}_e^{\Delta t}) \geq p({}^{\mathcal{R}}\widetilde{U}_e^{\Delta t}) \,, \tag{28b}$$

$$s(\widetilde{U}_e^{\Delta t}) \geq s({}^{\mathcal{R}}\widetilde{U}_e^{\Delta t}) \,. \tag{28c}$$

Combining Eqs. (27) and (28), we now conclude the above analysis with the following lemma as a critical building block for the subsequent derivation.

Lemma 2. *For a three-point system defined on $\mathbb{R}^{N_d}$, the solution along an arbitrary direction $\widehat{\mathrm{n}}$,*

$$\widetilde{U}_e^{\Delta t} = \widetilde{U}_e + \frac{\Delta t}{h}\left(\widehat{F}(\widetilde{U}_e, \widetilde{U}_{e-\widehat{\mathrm{n}}}, -\widehat{\mathrm{n}}) + \widehat{F}(\widetilde{U}_e, \widetilde{U}_{e+\widehat{\mathrm{n}}}, \widehat{\mathrm{n}})\right) \,, \tag{29}$$

with the flux function $\widehat{F}$ specified in Eq. (15), preserves the positivity of density and pressure, and satisfies the entropy principle:

$$s(\widetilde{U}_e^{\Delta t}) \geq \min_{j\in\{e-\widehat{\mathrm{n}},e,e+\widehat{\mathrm{n}}\}} s(\widetilde{U}_j) \,,$$

under the CFL condition:

$$\frac{\Delta t\lambda}{h} \leq \frac{1}{2}, \qquad \lambda \geq \max_{j\in\{e-\widehat{\mathrm{n}},e,e+\widehat{\mathrm{n}}\}} \nu(\widetilde{U}_j) \,.$$

This three-point system is consistent with that used in [9,11]. By considering specific quadrature rules, Zhang and Shu [11] showed the entropy boundedness of the cell-averaged solution of a high-order approximation in one-dimensional and two-dimensional rectangular and triangular elements. The resulting scheme was implemented using a global entropy bound that is derived from the initial condition $\min_{x\in\Omega} s(x,0)$. The focus of the present work is to explore if some of the implementation constraints can be relaxed, how to extend the entropy-principle to high-order solution approximations or elements with general geometric representations, and if it is possible to utilize a local entropy bound $s_e^0(t)$.

Although a local Lax–Friedrichs flux was used for illustrative purposes, other Riemann solvers that preserve positivity and entropy stability are equally suitable, for example, the Roe-type solver with entropy fix [18], the kinetic-type solver [19] and the exact Godunov solver.

4.2. Entropy-bounded DG-scheme

To robustly capture shocks while retaining the high-order benefit of the DG-scheme, sub-cell shock resolution is required [20]. We now extend the discussion by considering a high-order DG-solution with sub-cell representation. In each DG-cell, the whole solution is approximated by a function space. However, there is no guarantee that the high-order ($N_p > 1$) solution obeys the physical entropy principle. This is the reason that DG suffers from numerical instability in the vicinity of discontinuities. To suppress these instabilities, one approach is to consider imposing constraints based on the behavior of the entropy solution. The positivity-preserving DG-method [9,21] is a successful example for this approach. Based on the entropy principle, Eq. (18), Zhang and Shu [11] extended their implementation to an entropy-based constraint. Here, we propose a general framework that is based on the entropy principle, and major differences and advantages have been highlighted in Section 1.

We define the constraint for the high-order DG-scheme as follows:

$$\forall x \in \Omega_e, \;\; s(U_e^{\Delta t}(x)) \geq \min\{s(U(y))|\; y \in \Omega_e \cup \partial\Omega_e^-\} \equiv s_e^0(t) \,. \tag{30}$$

In this equation, the right-hand-side sets an entropy bound for an element-local solution in Ω_e; with this, we refer to a DG-solution as *entropy-bounded* if it satisfies this principle. $s_e^0(t)$ is a local estimate for the true entropy minimum in Ω_e, $|s_e^0(t) - \min_{x\in\Omega_e} s(\mathsf{U}(x))| \sim \mathcal{O}(h^k)$, where k is the local order of accuracy. Besides that, $s_e^0(t)$ is bounded if the entropy is bounded at the domain boundaries, $s_e^0(t) \geq \min_{x\in\Omega} s(U(x,t=0)) = s^0$, where s^0 is the minimum entropy at the initial condition. By imposing this constraint, we expect that the sub-cell DG-solution is regularized, avoiding the appearance of non-physical solutions. This idea is illustrated in Fig. 2. At time level t, $s_e^0(t)$ is calculated and used to set a reference bound for the solution at the next step, $U_e^{\Delta t}$. If $U_e^{\Delta t}$ yields entropy undershoot with respect to $s_e^0(t)$, it will be modified to satisfy the constraint of Eq. (30). In order to implement this regularization for a high-order DG-scheme, the following aspects require addressing:

(i) To impose Eq. (30) on the DG-solution, we introduce a limiting operator $\mathcal{L}$. The regularized solution, denoted by ${}^{\mathcal{L}}U_e^{\Delta t}$, requires that $s({}^{\mathcal{L}}U_e^{\Delta t}(x)) \geq s_e^0(t)\;\; \forall x \in \Omega_e$. In the following, we relax this condition, and impose Eq. (30) only on the set of quadrature points, $\mathcal{D}$, that are involved in solving the weak form in Eq. (12).

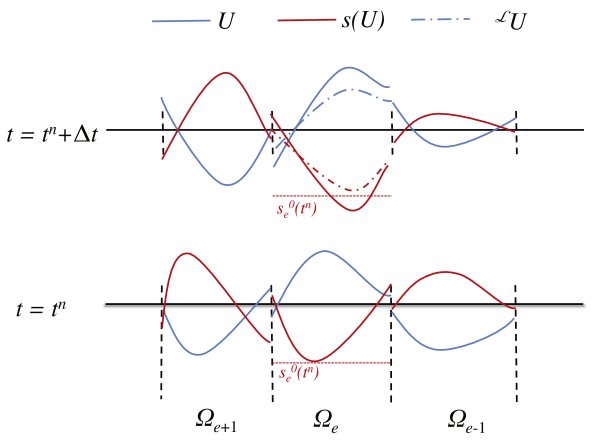


Fig. 2. (Color online.) Schematic of entropy-bounding of the EBDG-scheme.

(ii) Guaranteeing that the constraint (30) is always imposed requires the existence of the operator $\mathcal{L}$. A sufficient condition for this is that the element-averaged solution is entropy-bounded, $s(\overline{U}_e^{\Delta t}) \geq s_e^0(t)$. Enforcing this condition relies on the selection of a proper CFL-condition, and this analysis will be developed in Section 4.3 for a one-dimensional system. Subsequently, this analysis is then extended in Section 5.2 to general multi-dimensional elements.

(iii) Algorithmic details on the implementation of the operator $\mathcal{L}$ constraining the element-local DG-solution are discussed in Section 4.4.

(iv) The evaluation of the lower bound $s_e^0(t)$ that is necessary to constrain the entropy solution is given in Section 6.

4.3. CFL-constraint for one-dimensional entropy-bounded DG

The objective now is to extend the analysis for DGP0 to a DG-scheme with high-order polynomial representations. Consider a one-dimensional domain in which the element Ω_e is centered at x_e, and a quadrature rule with weights w_q and $\sum_{q=1}^{N_q} w_q = 1$. These quadrature weights are evaluated at the quadrature points $x_q \in [x_{e+1/2},\ x_{e-1/2}]$. The discretized cell-averaged solution $\overline{U}_e$ is defined as:

$$\overline{U}_e = \frac{1}{h}\int_{\Omega_e} U_e dx, \tag{31}$$

(for the P0-case discussed above, $\overline{U}_e = \widetilde{U}_e$), which can be further expanded by a quadrature rule with sufficient accuracy:

$$\begin{aligned}
\overline{U}_e &= \sum_{q=1}^{N_q} w_q U_e(x_q)\,, \\
&= \sum_{q=1}^{N_q} \left(w_q - \theta_l \phi_q(x_{e-1/2}) - \theta_r \phi_q(x_{e+1/2})\right) U_e(x_q) + \theta_l U_e(x_{e-1/2}) + \theta_r U_e(x_{e+1/2})\,, \\
&= \sum_{q=1}^{N_q} \theta_q U_e(x_q) + \theta_l U_e(x_{e-1/2}) + \theta_r U_e(x_{e+1/2})\,,
\end{aligned} \tag{32}$$

where the first line utilizes the exactness of the quadrature rule, the second line utilizes Lemma 1, and the third line defines $\theta_q = w_q - \theta_l \phi_q(x_{e-1/2}) - \theta_r \phi_q(x_{e+1/2})$. Under the condition that $\theta_{r,l} > 0$ and $\theta_q \geq 0$, the last line of Eq. (32) is a convex combination. Since the quadrature weights w_q are positive, the existence of $\theta_{r,l}$ is guaranteed through the condition $w_q \geq \theta_l \phi_q(x_{e-1/2}) + \theta_r \phi_q(x_{e+1/2})$. If $\phi_q(x_{e\pm 1/2}) > 0$, $\theta_{r,l}$ is constrained as $(0, \min_q w_q / \max\{\phi_q(x_{e\pm 1/2})\}]$. If some of $\phi_q(x_{e\pm 1/2})$ are negative, they are not essential in setting the upper bound for $\theta_{r,l}$.

Remark 2. In the following, $\theta_{r,l}$ will be related to a CFL-constraint. To obtain an optimal CFL-number, the largest value of $\theta_{r,l}$ needs to be found. This can be formulated as a maximization problem subject to the constraints, $\theta_{r,l} > 0$ and $\theta_q \geq 0$.

For illustration, we fully discretize Eq. (12) using a forward Euler time integration scheme and insert the results from Eq. (32). The element-averaged solution in Ω_e is then updated as:

$$\begin{aligned}\overline{U}_e^{\Delta t} &= \overline{U}_e - \frac{\Delta t}{h}\left(\widehat{F}(U_e(x_{e-1/2}), U_{e-1}(x_{e-1/2}), -1) + \widehat{F}(U_e(x_{e+1/2}), U_{e+1}(x_{e+1/2}), 1)\right) ,\\ &= \sum_{q=1}^{N_q}\theta_q U_e(x_q) + \theta_l U_e(x_{e-1/2}) - \frac{\Delta t}{h}\left(\widehat{F}(U_e(x_{e-1/2}), U_{e-1}(x_{e-1/2}), -1) + \widehat{F}(U_e(x_{e-1/2}), U_e^*, 1)\right)\\ &\quad + \theta_r U_e(x_{e+1/2}) - \frac{\Delta t}{h}\left(\widehat{F}(U_e(x_{e+1/2}), U_e^*, -1) + \widehat{F}(U_e(x_{e+1/2}), U_{e+1}(x_{e+1/2}), 1)\right) ,\end{aligned} \tag{33}$$

where

$$U_e^* = \frac{1}{2}\left(U_e(x_{e-1/2}) + U_e(x_{e+1/2})\right) - \frac{1}{2\lambda}\left(F(U_e(x_{e+1/2})) - F(U_e(x_{e-1/2}))\right) \tag{34}$$

is introduced to simplify subsequent analyses. Note that U_e^* ensures the validity of the second equality in Eq. (33) with a λ that is defined in the following lemma. We can see that Eq. (33) contains two three-point systems discussed in Section 4.1. To guarantee that $\overline{U}_e^{\Delta t}$ is entropy-bounded, it is necessary that these systems conform to the entropy principle of Eq. (18). This leads to the following lemma.

Lemma 3. *For a one-dimensional DG-system, the element-averaged solution satisfies the entropy principle*

$$s(\overline{U}_e^{\Delta t}) \geq s_e^0(t) = \min\{s(U(y))|\ y \in \Omega_e \cup \partial\Omega_e^-\} , \tag{35}$$

under the CFL-constraint

$$\frac{\Delta t\lambda}{h} \leq \frac{1}{2}\min\{\theta_l, \theta_r\} , \tag{36}$$

where λ is the maximum wave speed that is evaluated over the set of point-wise solutions

$$\lambda \geq \max_U\ \nu(U) \qquad with \qquad U \in \{U_{e-1}(x_{e-1/2}), U_e(x_{e\pm 1/2}), U_{e+1}(x_{e+1/2})\} , \tag{37}$$

given the conditions $\theta_{r,l} > 0$ and $\theta_q \geq 0$.

Proof. First, we need to show that U_e^* satisfies the discretized entropy principle of Eq. (18). This is indeed the case since U_e^* is essentially the left-hand-side of Eq. (19b) with time step taking $\Delta t = h/(2\lambda)$, corresponding to the upper bound of the CFL-constraint. Since U_e^* is an entropy solution, $\nu(U_e^*)$ is bounded by λ, which makes the Lax–Friedrichs flux $\widehat{F}$ involving U_e^* in Eq. (33) valid according to the definition in Eq. (15). Therefore, we have

$$s(U_e^*) \geq \min\left\{s(U_e(x_{e-1/2})), s(U_e(x_{e+1/2}))\right\} ,$$

Second, we reformulate Eq. (33) as

$$\overline{U}_e^{\Delta t} = \sum_{q=1}^{N_q}\theta_q U_e(x_q) + \theta_l \overline{U}_{e,p1}^{\Delta t} + \theta_r \overline{U}_{e,p2}^{\Delta t} , \tag{38}$$

in which $\overline{U}_{e,p1}^{\Delta t}$ and $\overline{U}_{e,p2}^{\Delta t}$ are the two updated solutions of the three-point system. Their definitions are readily obtained by comparing Eqs. (38) and (33). The given constraints, $\theta_{r,l} > 0$ and $\theta_q \geq 0$, guarantee that the form of the convex combination in Eq. (38) always holds. According to Lemma 2, it follows

$$\begin{aligned} s(\overline{U}_{e,p1}^{\Delta t}) &\geq \min\left\{s(U_e^*), s(U_e(x_{e-1/2})), s(U_{e-1}(x_{e-1/2}))\right\} ,\\ s(\overline{U}_{e,p2}^{\Delta t}) &\geq \min\left\{s(U_e^*), s(U_e(x_{e+1/2})), s(U_{e+1}(x_{e+1/2}))\right\} ,\end{aligned}$$

under the given CFL-constraint, Eq. (36). Combining this with the quasi-concavity of entropy [11], it follows

$$\begin{aligned} s(\overline{U}_e^{\Delta t}) &\geq \min\left\{s(U_e(x_q)), s(\overline{U}_{e,p1}^{\Delta t}), s(\overline{U}_{e,p2}^{\Delta t})\right\} ,\\ &\geq \min\{s(U(y))|\ y \in \Omega_e \cup \partial\Omega_e^-\} . \qquad \square\end{aligned}$$

Remark 3. Eq. (36) requires the positivity of $p(\overline{U}_e^{\Delta t})$ and $\rho(\overline{U}_e^{\Delta t})$.

In this context, we emphasize that the CFL-constraint (36) provides a description for the entropy boundedness and does not conflict with the general CFL-constraint for linear stability, $\mathrm{CFL}^{\mathrm{L}}$. To distinguish both constraints, here we use $\mathrm{CFL}^{\mathrm{EB}}$ to denote the CFL-number for guaranteeing the entropy boundedness. In general, the time step has to be selected to satisfy both criteria. Eq. (36) shows that $\mathrm{CFL}^{\mathrm{EB}}$ depends on the value $\min\{\theta_l, \theta_r\}$, and a rigorous evaluation for this will be given below.

Although we consider the specific case of a forward Euler time discretization scheme, all the derivation and conclusions are directly applicable to any explicit Runge–Kutta (RK) methods with positive coefficients, since the RK-solution is a convex combination of solutions obtained from several forward Euler sub-steps. In practice, RK-methods are preferred as DG time-integration schemes due to their compatible stability properties [22].

4.4. Construction of a limiting operator $\mathcal{L}$

Following Lemma 3, the entropy constraint is imposed on the set of quadrature points, $x \in \mathcal{D} \subset \Omega_e$. For the one-dimensional case, $\mathcal{D}$ is:

$$\mathcal{D} = \{x_{e\pm 1/2}, x_q, q = 1, \ldots, N_q \ (N_q \geq N_p)\} . \tag{39}$$

In the following, we are concerned with the construction of a limiting operator $\mathcal{L}$, such that

$$\forall x \in \mathcal{D}, \qquad s({}^{\mathcal{L}}U_e^{\Delta t}(x)) \geq s_e^0(t) . \tag{40}$$

Since the operator $\mathcal{L}$ is applied at the end of each sub-iteration, we will omit the superscript Δt in the subsequent analysis. According to the entropy definition (7), Eq. (40) can be written as:

$$p({}^{\mathcal{L}}U_e(x)) \geq \exp(s_e^0)\rho^\gamma({}^{\mathcal{L}}U_e(x)) \qquad \forall x \in \mathcal{D} . \tag{41}$$

To define the operator $\mathcal{L}$, we follow the work of Zhang and Shu [9,11], and introduce a linear scaling:

$${}^{\mathcal{L}}U_e = U_e + \varepsilon(\overline{U}_e - U_e) , \tag{42}$$

where ε is the limiting parameter with $0 \leq \varepsilon \leq 1$. The solution reduces to first-order accuracy for $\varepsilon = 1$ and retains the optimal accuracy for $\varepsilon = 0$. To determine ε, we first expand Eq. (41) by applying Jensen's inequality:

$$p\left((1-\varepsilon)U_e + \varepsilon\overline{U}_e\right) \geq (1-\varepsilon)p(U_e) + \varepsilon p(\overline{U}_e) , \tag{43a}$$

$$\left[(1-\varepsilon)\rho^\gamma(U_e) + \varepsilon\rho^\gamma(\overline{U}_e)\right] \geq \rho^\gamma\left((1-\varepsilon)U_e + \varepsilon\overline{U}_e\right) . \tag{43b}$$

Enforcing the validity of Eq. (41) requires the following relation:

$$(1-\varepsilon)p(U_e) + \varepsilon p(\overline{U}_e) \geq \exp(s_e^0)\left[(1-\varepsilon)\rho^\gamma(U_e) + \varepsilon\rho^\gamma(\overline{U}_e)\right] ,$$

from which, ε can be evaluated as:

$$\varepsilon = \frac{\tau}{\tau - [p(\overline{U}_e) - \exp(s_e^0)\rho^\gamma(\overline{U}_e)]} \qquad \text{with} \qquad \tau = \min\left\{0, \ \min_{x\in\mathcal{D}}\{p(U_e(x)) - \exp(s_e^0)\rho^\gamma(U_e(x))\}\right\} , \tag{44}$$

which is subject to the conditions

$$\rho(\overline{U}_e) > 0 \text{ and } p(\overline{U}_e) > \exp(s_e^0)\rho^\gamma(\overline{U}_e) . \tag{45}$$

These conditions are automatically guaranteed through the CFL^{EB}-constraint of Lemma 3. While the positivity condition for pressure is embedded in Eq. (43), the positivity of density must be imposed for all $x \in \mathcal{D}$ before $\mathcal{L}$ is applied, and the methodology for this is presented in [21].

Compared to the limiting operator, presented in [11], the herein proposed method is substantially simplified. Specifically, the step for imposing the positivity of pressure is avoided; in addition, ε is obtained from an algebraic relation, and does not require a computationally expensive Newton iteration. It is also noted that the operator $\mathcal{L}$ contains the positivity-preserving limiter as a special case, which is obtained by setting $s_e^0 \to -\infty$.

4.5. Numerical analysis of the limiting operator $\mathcal{L}$

In this section, numerical properties of the limiting operator are examined.

4.5.1. Conservation

Integrating Eq. (42) over Ω_e,

$$\int_{\Omega_e} {}^{\mathcal{L}}U_e dx = (1-\varepsilon)\int_{\Omega_e} U_e dx + \varepsilon\overline{U}_e\int_{\Omega_i} dx = \int_{\Omega_e} U_e dx ,$$

confirms that the limiting operator preserves the conservation properties of the solution vector.

4.5.2. Stability

Since the positivity of density and pressure is preserved at the quadrature points, $\mathcal{L}$ is L_1-stable, which was shown in [9,11]. Here, we extend this stability analysis and evaluate the L_2-stability. By considering a periodic domain and taking the L_2-norm of Eq. (42) we obtain:

$$||^{\mathcal{L}}U_e||^2 = \int_{\Omega_e} [U_e + \varepsilon(\overline{U}_e - U_e)]^2 dx\,, \tag{46a}$$

$$= (1-\varepsilon)^2 \int_{\Omega_e} U_e^2 dx - \varepsilon(\varepsilon-2)\int_{\Omega_e} \overline{U}_e^2 dx\,, \tag{46b}$$

$$\leq (1-\varepsilon)^2 \int_{\Omega_e} U_e^2 dx - \varepsilon(\varepsilon-2)\int_{\Omega_e} U_e^2 dx\,, \tag{46c}$$

$$\leq ||U_e||^2\,. \tag{46d}$$

After integrating over the entire domain, we obtain

$$||^{\mathcal{L}}U||_\Omega^2 \leq ||U||_\Omega^2\,,$$

which shows that $\mathcal{L}$ does not affect the stability of the DG-discretization. Further, since $\mathcal{L}$ constrains pressure and density, λ in Eq. (36) provides a robust CFL-criterion, without the need for arbitrarily reducing Δt to increase the stability region.

4.5.3. Accuracy

In regions where the solution is smooth, we assume that the weak solution before limiting has optimal accuracy:

$$||U - \mathsf{U}||_\Omega \leq C_1 h^{p+1}\,,$$

and that undershoots in entropy remain small $\tau \sim \mathcal{O}(h^{p+1})$ in Eq. (44). This implies $\varepsilon \sim \mathcal{O}(h^p)$ under the condition $|p(\overline{U}_e) - \exp(s_e^0)\rho^\gamma(\overline{U}_e)| \sim \mathcal{O}(h)$. Thus, the error is estimated as follows:

$$||^{\mathcal{L}}U - \mathsf{U}||_\Omega^2 = \sum_e ||^{\mathcal{L}}U_e - \mathsf{U}_e||^2\,, \tag{47a}$$

$$= \sum_e ||\varepsilon(\overline{U}_e - \mathsf{U}_e) + (1-\varepsilon)(\mathsf{U}_e - U_e)||^2\,, \tag{47b}$$

$$\leq \sum_e \left(2\varepsilon^2 ||(\overline{U}_e - \mathsf{U}_e)||^2 + 2(1-\varepsilon)^2 ||U_e - \mathsf{U}_e||^2\right)\,, \tag{47c}$$

$$\leq C_2 h^{2p+2}\,, \tag{47d}$$

where for simplicity, we introduce U_e to denote the element-wise representation to U. Here we use the fact that $\overline{U}_e$ is locally a first-order approximation to U_e, $\overline{U}_e = \mathsf{U}_e + C_3^e\mathcal{O}(h)$. This analysis applies to general smooth solutions. For the cases that the condition $|p(\overline{U}_e) - \exp(s_e^0)\rho^\gamma(\overline{U}_e)| \sim \mathcal{O}(h)$ cannot be guaranteed, the local accuracy might be impacted by applying $\mathcal{L}$.

In the vicinity of a discontinuity, the DG-solution loses its regularity so that the convergence rate reduces to first-order: $||U - \mathsf{U}||_\Omega \leq C_4 h$. Triggered by spurious sub-cell solutions, the entropy undershoot can be very large, so that $\varepsilon \sim \mathcal{O}(1)$. By repeating the above argument, we obtain an estimate for the accuracy of the discontinuous solution:

$$||^{\mathcal{L}}U - \mathsf{U}||_\Omega \leq C_5 h.$$

The accuracy arguments given here are substantiated through numerical tests in Section 8.

5. Generalization to multi-dimension and arbitrary elements

The entropy-bounded DG scheme that was presented for one-dimensional systems in the previous section can be generalized to arbitrary elements in multi-dimensions. This extension is the subject of the following analysis.

Since EBDG does not rely on a specific quadrature rule, any quadrature method can be used as long as it accurately integrates the problem and ensures the positivity of the quadrature weights. The limiting procedure requires the definition of a new set of quadrature points $\mathcal{D}$ for the general multi-dimensional setting. The selection of these points is given in the next section. The extension to arbitrary elements requires special consideration of the CFL^{EB} number.

5.1. Generalization to multi-dimension and arbitrary elements

To present a general formulation for multi-dimensional configurations, we first introduce necessary notations to describe general elements with curvatures. For this, we define a geometric mapping function $\Phi : \mathbb{R}^{N_d} \to \mathbb{R}^{N_d}$ on a reference element Ω_e^{r}, such that $x = \Phi(r)$ maps any point $r \in \Omega_e^{\text{r}}$ onto $x \in \Omega_e$, and $\mathcal{J} = [\partial x/\partial r]$ is the geometric Jacobian. With these specifications, we can write the discretized state vector as:

$$U_e(x,t) = \sum_{m=1}^{N_p} \widetilde{U}_{e,m}(t)\varphi_m(r)\,, \qquad x = x(r) \in \Omega_e\,, \qquad \forall r \in \Omega_e^{\text{r}}. \tag{48}$$

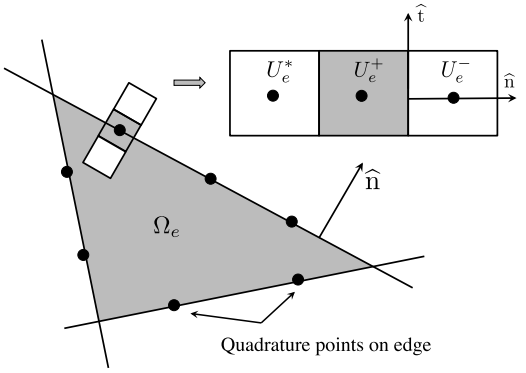


Fig. 3. A graphic interpretation of the multi-dimension convex expansion.

The mapping function is commonly parameterized by a polynomial function $x(r) = \sum_{m=1}^{N_g} \widetilde{x}_m \varphi_m^g(r)$, where $\varphi_m^g(r)$ is a Lagrangian interpolation and N_g is the number of geometric bases used to represent Ω_e. Since the reference element is regular, we can use a subspace of r to parameterize the element edges. Therefore, to parameterize the kth edge of Ω_e we define $g_k = \mathcal{P}_k r \in \mathbb{R}^{N_d-1}$ such that $\forall r \in \partial\Omega_{e,k}^{\mathrm{r}}$, $r = \mathcal{P}_k^{-1} g_k$, in which $\mathcal{P}_k^{-1}$ is the pseudo-inverse of $\mathcal{P}_k$. For the physical element, the edge can be represented as:

$$\partial\Omega_{e,k} = \{x \in \Omega_e \mid x = \Phi(r),\ r = \mathcal{P}_k^{-1}(g_k) \in \partial\Omega_{e,k}^{\mathrm{r}}\}\,. \tag{49}$$

The integral in Eq. (12) is evaluated using multi-dimensional quadrature rules. Considering the complexity of the dimensionality, here we follow the quadrature convention that is $\sum_{v=1}^{N_q} w_v = V_e^{\mathrm{r}}$ (the volume of Ω_e^{r}) and $\sum_{q=1}^{N_q^k} w_{k,q} = S_{e,k}^{\mathrm{r}}$ (the area of $\partial\Omega_{e,k}$). With these preliminaries, we can evaluated any volume integral in Eq. (12) as:

$$\int_{\Omega_e} f(x)dx = \int_{\Omega_e^{\mathrm{r}}} |\mathcal{J}(r)| f(x(r))dr = \sum_{v=1}^{N_q} |\mathcal{J}(r_v)| f(x(r_v)) w_v\,. \tag{50}$$

The surface integral of a scalar function on $\partial\Omega_{e,k}$ can be written as:

$$\int_{\partial\Omega_{e,k}} f(x)d\Gamma = \int_{\partial\Omega_{e,k}^{\mathrm{r}}} f(x(g_k))|\mathcal{J}_k^{\partial}|dg_k = \sum_{q=1}^{N_{q,k}^{\partial}} |\mathcal{J}_k^{\partial}(g_{k,q})| f(x(g_{k,q})) w_{k,q}\,, \tag{51}$$

where the surface Jacobian $\mathcal{J}_k^{\partial}$ is related to the volume Jacobian $\mathcal{J}$ as $\mathcal{J}_k^{\partial} = \mathcal{J}\mathcal{P}_k^{-1}$. The normal $\hat{\mathrm{n}}$ of the surface $\partial\Omega_{e,k}$ is required for the Riemann flux evaluation, and it can be evaluated using $\mathcal{J}_k^{\partial}/|\mathcal{J}_k^{\partial}|$. Note that for polynomial-parameterized geometries, the quadrature exactness still holds as long as the quadrature order is sufficiently high. Hence, for curved elements, the quadrature order has to be improved to account for the geometric complexity.

With the above notation, we are now able to define the set of quadrature points $\mathcal{D}$ for general curved elements:

$$\mathcal{D} = \bigcup_{k=1}^{N_\partial} \{g_{k,q},\ q = 1, \ldots, N_{q,k}^{\partial}\} \bigcup \{r_v,\ v = 1, \ldots, N_q\}\,, \tag{52}$$

where N_∂ is the number of element edges (which is equal to the number of neighbor elements). In this context, it is noted that $\mathcal{D}$ includes all quadrature points that are involved in the integration. With this specification of $\mathcal{D}$, the limiting operator $\mathcal{L}$, developed in Section 4.4, can be directly extended to arbitrary elements on multi-dimensional configurations. In the following, a $\mathrm{CFL}^{\mathrm{EB}}$-constraint is derived that extends the results of Lemma 3, thereby ensuring the existence of the general limiter $\mathcal{L}$.

5.2. CFL-constraint

Following the same approach as for the one-dimensional derivation in Section 4.3, we expand the updated element-averaged solution to a convex combination. To facilitate the derivation, Fig. 3 provides a graphical interpretation, showing the relation between the following expansion and the three-point system that was discussed in Section 4.1. With this, $U_e^{\Delta t}$ is evaluated as:

$$\overline{U}_e^{\Delta t} = \overline{U}_e - \frac{\Delta t}{V_e}\sum_{k=1}^{N_\partial}\int_{\partial\Omega_{e,k}} \widehat{F}\left(U_e^+, U_e^-, \widehat{\mathrm{n}}\right) d\Gamma \,, \tag{53a}$$

$$= \sum_{v=1}^{N_q}\frac{|\mathcal{J}(r_v)|w_v}{V_e}U_e(r_v) - \sum_{k=1}^{N_\partial}\sum_{q=1}^{N_{q,k}^\partial}\frac{\Delta t|\mathcal{J}_k^\partial(g_{k,q})|w_{k,q}}{V_e}\widehat{F}\left(U_e^+(r(g_{k,q})), U_e^-(r(g_{k,q})), \widehat{\mathrm{n}}(g_{k,q})\right) \,, \tag{53b}$$

$$= \sum_{v=1}^{N_q}\theta_v U_e(r_v) + \sum_{k=1}^{N_\partial}\sum_{q=1}^{N_{q,k}^\partial}\Big[\theta_{k,q}U_e^+(r(g_{k,q})) - \Delta t\zeta_{k,q}\left(\widehat{F}\left(U_e^+(r(g_{k,q})), U_e^-(r(g_{k,q})), \widehat{\mathrm{n}}(g_{k,q})\right) + \widehat{F}\left(U_e^+(r(g_{k,q})), U_e^*, -\widehat{\mathrm{n}}(g_{k,q})\right)\right)\Big] \,, \tag{53c}$$

where

$$\theta_v = \frac{|J(r_v)|w_v}{V_e} - \sum_{k=1}^{N_\partial}\sum_{q=1}^{N_{q,k}^\partial}\theta_{k,q}\phi_v\left(r(g_{k,q})\right) \tag{54}$$

is introduced to decompose the volumetric quadrature to obtain $U_e^+(r(g_{k,q}))$. For notational simplification, we defined

$$\zeta_{k,q} = \frac{|\mathcal{J}_k^\partial(g_{k,q})|w_{k,q}}{V_e} \,, \tag{55}$$

such that $S_e = \sum_{k=1}^{N_\partial}\sum_{q=1}^{N_{q,k}^\partial}\zeta_{k,q}$ is equal to the surface area of Ω_e, and $\sum_{k=1}^{N_\partial}\sum_{q=1}^{N_{q,k}^\partial}\zeta_{k,q}\widehat{\mathrm{n}}(g_{k,q}) = 0$ since Ω_e has a closed surface. To apply the results from the three-point system to the multi-dimensional configuration, we introduce the auxiliary variable U_e^*:

$$U_e^* = \sum_{k=1}^{N_\partial}\sum_{q=1}^{N_{q,k}^\partial}\frac{\zeta_{k,q}}{S_e}\left[U_e^+(r(g_{k,q})) - \frac{1}{\lambda^*}F(U_e^+(r(g_{k,q}))\cdot\widehat{\mathrm{n}}(g_{k,q})\right] . \tag{56}$$

It can be shown that U_e^* is essentially the solution to the following equation:

$$\sum_{k=1}^{N_\partial}\sum_{q=1}^{N_{q,k}^\partial}\zeta_{k,q}\widehat{F}\left(U_e^+(r(g_{k,q})), U_e^*, -\widehat{\mathrm{n}}(g_{k,q})\right) = 0 \,, \tag{57}$$

subject to a preselected dissipation coefficient λ^*, so that the equality in Eq. (53c) holds true. Here, we evaluate λ^* from the following relation:

$$\lambda^* = \tau\max\left\{\nu(U) \mid U \in \{U_e^+(r(g_{k,q}))\}\right\}, \qquad \tau = \max\left\{\sqrt{N_d}, \sqrt{2+\gamma(\gamma-1)}\right\} \tag{58}$$

and the rationale for this selection is provided later in Remark 4. To prove that $\overline{U}_e^{\Delta t}$ is entropy bounded, we present the following lemma.

Lemma 4. *U_e^* in Eq. (56) satisfies $s(U_e^*) \geq \min\left\{s(U) \mid U \in \{U_e^+(r(g_{k,q}))\}\right\}$.*

Proof. For notational simplification, we combine the indices k and q into a single index j, and we denote the total number of surface quadrature points on $\partial\Omega_e$ by N_{tot}, $N_{\text{tot}} = \sum_{k=1}^{N_\partial}N_{q,k}^\partial$. Considering $\sum_{j=1}^{N_{\text{tot}}}\zeta_j\widehat{\mathrm{n}}_j^{(d)} = 0$ and $\zeta_j > 0$, the dth components of the surface-normal vectors have different signs. To denote each component, we introduce the superscript (d). By sorting $\widehat{\mathrm{n}}_j^{(d)}$ so that the first $N_{\text{tot}}^>$ vector components are positive. The following statement is true for any d:

$$\sum_{j=1}^{N_{\text{tot}}^>}\zeta_j\widehat{\mathrm{n}}_j^{(d)} = -\sum_{j=N_{\text{tot}}^>+1}^{N_{\text{tot}}}\zeta_j\widehat{\mathrm{n}}_j^{(d)} = \sum_{n=1}^{N_{par}}l_n \,, \tag{59}$$

where l_n introduces a partition as illustrated in Fig. 4 and N_{par} is the dimension of this partition. With this, we are able to introduce a variable mapping,

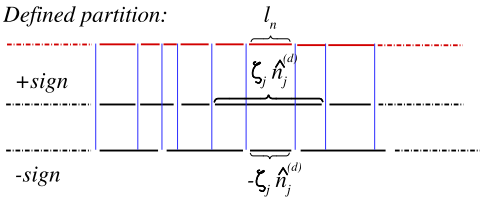


Fig. 4. Illustration of the partition introduced in Lemma 4.

$$U_n^{s+} = U_e^+(r_j), \quad \text{if } \sum_{i=1}^{j-1} \zeta_i \hat{\mathrm{n}}_i^{(d)} < \sum_{i=1}^{n} l_i \le \sum_{i=1}^{j} \zeta_i \hat{\mathrm{n}}_i^{(d)} ,$$

$$U_n^{s-} = U_e^+(r_j), \quad \text{if } -\sum_{i=N_{\mathrm{tot}}^{>}+1}^{j-1} \zeta_i \hat{\mathrm{n}}_i^{(d)} < \sum_{i=1}^{n} l_i \le -\sum_{i=N_{\mathrm{tot}}^{>}+1}^{j} \zeta_i \hat{\mathrm{n}}_i^{(d)} .$$

With this, Eq. (56) is equivalent to:

$$\begin{aligned} U_e^* &= \frac{1}{S_e}\left(\sum_{j=1}^{N_{\mathrm{tot}}} \zeta_j U_e^+(r_j) - \sum_{j=1}^{N_{\mathrm{tot}}} \frac{\zeta_j}{\lambda^*} F(U_e^+(r_j)) \cdot \hat{\mathrm{n}}_j\right) , \\ &= \sum_{j=1}^{N_{\mathrm{tot}}} \frac{\zeta_j}{S_e}\left(1 - \sum_{d=1}^{N_d} \frac{|\hat{\mathrm{n}}_j^{(d)}|}{\sqrt{N_d}}\right) U_e^+(r_j) + \sum_{d=1}^{N_d}\sum_{j=1}^{N_{\mathrm{tot}}} \frac{1}{S_e}\left(\frac{\zeta_j |\hat{\mathrm{n}}_j^{(d)}|}{\sqrt{N_d}} U_e^+(r_j) - \frac{\zeta_j \hat{\mathrm{n}}_j^{(d)}}{\lambda^*} F^{(d)}(U_e^+(r_j))\right) , \\ &= \sum_{j=1}^{N_{\mathrm{tot}}} \frac{\zeta_j}{S_e}\left(1 - \sum_{d=1}^{N_d} \frac{|\hat{\mathrm{n}}_j^{(d)}|}{\sqrt{N_d}}\right) U_e^+(r_j) + \sum_{d=1}^{N_d} \frac{1}{S_e}\left(\frac{2}{\sqrt{N_d}} \sum_{n=1}^{N_{par}} l_n U_{d,n}^{**}\right) , \end{aligned}$$

where we introduce

$$U_{d,n}^{**} = \frac{1}{2}(U_n^{s+} + U_n^{s-}) - \frac{\sqrt{N_d}}{2\lambda^*}\left(F^{(d)}(U_n^{s+}) - F^{(d)}(U_n^{s-})\right) ,$$

which takes the same form as the left-hand-side of Eq. (34). Note that $U_{d,n}^{**}$ is essentially expressed in a one-dimensional setting along $x^{(d)}$. Therefore, one can follow the same argument used for Eq. (34) in Lemma 3 to verify that

$$s(U_{d,n}^{**}) \ge \min\left\{s(U_n^{s\pm})\right\} \ge \min\left\{s(U) \mid U \in \{U_e^+(r_j)\},\ j = 1, \ldots, N_{\mathrm{tot}}\right\} ,$$

with the given form of λ^* in Eq. (58). As given above, U_e^* is a convex combination; by using the quasi-concavity of entropy [11], we conclude that

$$s(U_e^*) > \min\left\{s(U) \mid U \in \{U_e^+(r_j)\},\ j = 1, \ldots, N_{\mathrm{tot}}\right\} . \qquad \square$$

Remark 4. Note that the maximum characteristic speed of $U_{d,n}^{**}$ is bounded, $\nu(U_{d,n}^{**}) \le \nu(U_e^+(r))$. According to the combination law given in Appendix A, we have $\nu(U_e^*) \le \sqrt{2+\gamma(\gamma-1)}\max\left\{\nu(U_e^+(r))\right\} \le \lambda^*$. With this, we are now able to show that the maximum characteristic speed of U_e^* is bounded by the chosen value of λ^*, so that the Lax–Friedrichs flux $\widehat{F}\left(U_e^+(r), U_e^*, -\hat{\mathrm{n}}\right)$ in Eq. (53) is valid according to the definition of Eq. (15).

To enforce the entropy boundedness, the decomposition of $\overline{U}_e^{\Delta t}$ in Eq. (53) is required to be convex. This can be satisfied under the following condition:

$$\begin{cases} \theta_v \ge 0\,, \ \ \forall v = 1, \ldots, N_q, \\ \theta_{k,q} > 0\,, \ \ \forall (k,\ q),\ q = 1, \ldots, N_k^{\partial},\ k = 1, \ldots, N_{\partial}\,, \end{cases} \tag{60}$$

With this, the entropy boundedness of $\overline{U}_e^{\Delta t}$ is shown by the following lemma.

Lemma 5. *For a general DG element, the element-averaged solution is entropy bounded,*

$$s(\overline{U}_e^{\Delta t}) \ge s_e^0(t) = \min\{s(U(y)) \mid y \in \Omega_e \cup \partial\Omega_e^-\}, \tag{61}$$

under the condition that Eq. (60) holds and that the following constraint is fulfilled:

$$\Delta t \lambda \le \frac{1}{2}\min\left\{\frac{\theta_{k,q}}{\zeta_{k,q}}\right\}, \ \ \forall (k,\ q),\ q = 1, \cdots, N_k^{\partial},\ k = 1, \cdots, N_{\partial}\,, \tag{62}$$

where $\lambda \ge \max\{\nu(U) \mid U \in \{U_e^{\pm}(r(g_{k,q}))\}$ *and* $\lambda \ge \lambda^*$.

Proof. The proof follows Lemma 3, utilizing Lemma 2 and the quasi-concavity of entropy. □

Note that Lemma 5 does not rely on any assumption regarding the dimensionality or shape of the finite element, and is therefore general. Another observation is that Eq. (62) essentially provides an estimate for $\mathrm{CFL}^{\mathrm{EB}}$ that is only a function of the geometry of the element. For practical applications, we require the right-hand-side of Eq. (62) to be as large as possible so that larger time steps can be taken. This can be achieved by solving a convex optimization problem:

$$\begin{aligned} &\text{maximize } \left(\min \left\{\frac{\theta_{k,q}}{\zeta_{k,q}}\right\}\right), \\ &\text{subject to Eq. (60)} \end{aligned} \tag{63}$$

where $\theta_{k,q}$, $\zeta_{k,q}$ are properties of the geometry alone. This problem can be solved for each individual element as a pre-processing step prior to the simulation. Another way to interpret the expression is to identify a length scale from the right-hand-side of Eq. (62), for which the $\mathrm{CFL}^{\mathrm{EB}}$ number can be explicitly defined. For this, $L_e = \min V_e/|\mathcal{J}_k^{\partial}(g_{k,q})|$ is used as a characteristic length for Ω_e. Hence,

$$\min \left\{\frac{\theta_{k,q}}{\zeta_{k,q}}\right\} \geq L_e \min \left\{\frac{\theta_{k,q}}{w_{k,q}}\right\}$$

and an alternative expression to Eq. (63) is

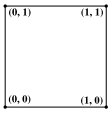
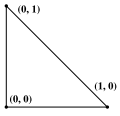
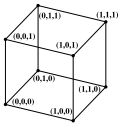
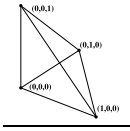
$$\begin{aligned} &\text{maximize } \min \left\{\frac{\theta_{k,q}}{w_{k,q}}\right\}, \\ &\text{subject to Eq. (60)}, \end{aligned} \tag{64}$$

where the optimal solution is the value of $\mathrm{CFL}^{\mathrm{EB}}$. With this, the CFL-constraint can be written as

$$\frac{\Delta t \lambda}{L_e} \leq \frac{1}{2}\mathrm{CFL}^{\mathrm{EB}}, \tag{65}$$

which is used in the following numerical tests. The factor of $1/2$ is a consequence of the Riemann flux formulation. For some of the most relevant element types with regular shapes, the value of $\mathrm{CFL}^{\mathrm{EB}}$ has been calculated and listed in Table 1 for different polynomial orders. In practice, we found that the bound in Eq. (65) leads to a conservative estimate for the time step. Considering the computation of efficiency and the constraint for the linear stability by [22], this condition is relaxed and we consider 0.8 $\mathrm{CFL}^{\mathrm{EB}}$ for the following numerical experiments.

Table 1
Summary of quadrature orders and optimal CFL numbers for different types of elements. Quadrature rule (QR) applied: Line, Quadrilateral and Brick: tensor-product Gauss–Legendre; Triangle: Dunavant [23]; Tetrahedron: Zhang et al. [24]. (Note that Dunavant's triangle rule includes negative weights for 3rd- and 7th-order quadrature, therefore, only quadrature rules with positive weights are used with one extra order.)

Element	Order	QR on $\partial\Omega_e$	QR on Ω_e	$\mathrm{CFL}^{\mathrm{EB}}$
	$p=1$	–	3	0.5
	$p=2$	–	5	0.167
	$p=3$	–	7	0.123
	$p=4$	–	9	0.073
	$p=1$	3	3	0.25
	$p=2$	5	5	0.083
	$p=3$	7	7	0.062
	$p=4$	9	9	0.036
	$p=1$	3	4	0.135
	$p=2$	5	5	0.067
	$p=3$	7	8	0.058
	$p=4$	9	9	0.033
	$p=1$	3	3	0.167
	$p=2$	5	5	0.056
	$p=3$	7	7	0.041
	$p=4$	9	9	0.024
	$p=1$	4	3	0.066
	$p=2$	5	5	0.035
	$p=3$	8	7	0.015
	$p=4$	9	9	0.013

6. Evaluation of entropy bound

In this section, we propose an approach for evaluating the entropy bound $s_e^0(t)$ to answer the fourth implementation problem listed in Section 4.2. Obviously, the most accurate way for evaluating a lower bound of entropy is to use Newton's method. However, this approach can significantly impair the efficiency, since searching the minimum on a multi-dimensional high-order element is intractable in terms of computational cost. To overcome this issue, we propose the following two approaches:

- *User-defined global bound*. The first strategy is to let the user specify a global entropy bound, which is then kept constant and used everywhere in the computational domain. Although this approach is simple and robust, it is not optimal. It is suitable for certain problems with a well-defined entropy bound. As example, for a supersonic flow over an airfoil, the free stream entropy can be used to impose this bound. However, for more complex cases with multiple discontinuities that include several entropy jumps, such a constant bound is not able to enforce the constraint for all elements. Note that this approach recovers the positivity constraint of Zhang and Shu [9,21] in the limit of $s_e^0(t) \to -\infty$.
- *Estimate of local entropy bound*. This strategy imposes an entropy bound for each element and dynamically updates $s_e^0(t)$ during the simulation. Instead of relying on a sophisticated search algorithm, $s_e^0(t)$ can be approximately evaluated by reusing available information on quadrature points that consist of two sets, $\mathcal{D}$ on the local cell and the set of quadrature points on $\partial\Omega_e^-$, denoted as $\mathcal{D}^-$. According to the definition of $s_e^0(t)$ in Eq. (30), a direct approach is to approximate the minimum entropy over the domain of dependence as:

$$\mathcal{M}\left(s_e^0(t)\right) = \min\left\{\min_{x\in\mathcal{D}^-} s(U(x)),\ s_m - \theta\left(s_n - s_m\right)\right\}, \tag{66}$$

where we introduce s_m and s_n to denote the minimum and maximum entropy values, respectively. Hence,

$$s_m = s(U_e(x_m)) = \min_{x\in\mathcal{D}} s(U_e(x)),$$
$$s_n = s(U_e(x_n)) = \max_{x\in\mathcal{D}} s(U_e(x)),$$

where x_m and x_n are the corresponding spatial locations. As given in Eq. (66), a relaxation factor, θ, is introduced to avoid overestimation, and for this study, we use

$$\theta = \frac{\min\limits_{x\in\mathcal{D}, x\neq x_m}\{|x_m - x|\}}{|x_m - x_n|}.$$

Although this estimate is simple and inexpensive, one has to realize that any relaxation in the vicinity of discontinuities becomes dangerous due to the spurious behavior of the sub-cell solution. This, however, can be resolved by referring to the entropy bounds around Ω_e at the last time step,

$$s_e^0(t) \approx {}^{\mathcal{E}}s_e^0(t) = \max\left\{\mathcal{M}\left(s_e^0(t)\right),\ \min_{k\in\mathcal{N}_e\cup\{e\}} s_k^0(t-\Delta t)\right\}, \tag{67}$$

where $\mathcal{N}_e$ refers to the set of indices of all neighbor elements of Ω_e that share a common edge. The performance of this estimation will be tested in Section 8, and will be compared to the global bound in Section 8.3.

As a closing remark, the reader is reminded that as far as the estimation is concerned, the robustness is not possible to be guaranteed for all gas-dynamics calculations. In the case that an overestimation occurs, there are several ways to fix this, for example, by increasing the relaxation factor θ or directly using ${}^{\mathcal{E}}s_e^0(t) = \min\limits_{k\in\mathcal{N}_e\cup\{e\}} s_k^0(t-\Delta t)$.

For practical tests, we found that the above strategy can be applied in a combined way. Specifically, Eq. (66) is used for initializing the simulation, and Eq. (67) is then applied during the subsequent integration step.

7. Algorithmic implementation

Algorithm 1 provides a description of the implementation details of the EBDG-scheme. This algorithm will be tested in the next section. As mentioned above, the recommended CFL number, 0.8 CFL^{EB}, is applied, and found to be sufficiently robust for all tests. However, we remind the reader that CFL^{EB}, derived above, is based on a one-time update of a forward Euler scheme. For RKDG, the algorithm relies on a multi-stage RK scheme. Therefore, a rigorous enforcement of this CFL condition requires that the spectral-radius of the discretized system (or the maximum wave speed over the domain) does not increase during Δt. In practice, this requirement is not of a concern. To further improve the robustness, this algorithm can be combined with a recursive strategy to dynamically reduce the CFL number [10], in case that the condition, Eq. (61) in Lemma 5, is not satisfied at any stage of the applied RK scheme.

Algorithm 1: Implementation of EBDG scheme.

Pre-computation of CFL condition: For each element, solve Eq. (63); alternatively, take $\mathrm{CFL}^{\mathrm{EB}}$ from Table 1 and compute L_e (recommended for simplicity)
Initialization: Initialize solution vector $U(x,0)=U_0$
while $t \le t_{\mathrm{end}}$ **do**
  for *each element* **do**
    Find λ and estimate time step size Δt according to Eq. (62)
  end
  Find the minimum permissible time step Δt_{min} over all elements
  for *each stage k of a Runge–Kutta integration scheme* **do**
    for *each element* **do**
      if *bounding is applied locally* **then**
        Estimate entropy bound $s_e^0(t)$ according to Eqs. (66) and (67)
      else
        Use $s_e^0(t)$ estimated based on U_0
      end
      Update solution vector $U^{k+1}=U^k+\Delta t_{\mathrm{min}} R^k$ (R refers to the residual)
      Apply $\mathcal{L}$ on U^{k+1} with $s_e^0(t)$ according to Eqs. (42) and (44)
    end
  end
  Advance time $t=t+\Delta t_{\mathrm{min}}$
end

Table 2
Convergence test of 1D advection with SSPRK33, showing degradation of convergence order for DGP3 and DGP4 (here we use density to evaluate the error).

h	DGP1		DGP2		DGP3		DGP4	
	L_2-error	rate	L_2-error	rate	L_2-error	rate	L_2-error	rate
1/10	3.074e−3	–	1.274e−4	–	4.716e−6	–	2.036e−7	–
1/20	6.508e−4	2.240	1.513e−5	3.073	3.073e−7	3.940	1.980e−8	3.362
1/40	1.535e−4	2.084	1.891e−6	3.000	2.182e−8	3.816	2.454e−9	3.013
1/80	3.775e−5	2.024	2.364e−7	3.000	1.880e−9	3.537	3.130e−10	2.971
1/160	9.398e−6	2.006	2.955e−8	3.000	2.001e−10	3.232	3.924e−11	2.995
1/320	2.347e−6	2.002	3.694e−9	3.000	2.401e−11	3.059	4.922e−12	2.995

8. Results and numerical test cases

In the following, EBDG is applied to a series of test cases to demonstrate the performance of this method. We begin by considering one-dimensional configurations to confirm the high-order accuracy and essential convergence properties. This is followed by two- and three-dimensional cases with specific emphasis on applications to unstructured meshes and general curved elements.

8.1. One-dimensional smooth solution

The first case considers a one-dimensional periodic domain $x \in [0,1]$ with smooth initial conditions:

$$\rho(x,0) = 1 + 0.1\sin(2\pi x)\,,$$
$$u(x,0) = 1\,,$$
$$p(x,0) = 1\,.$$

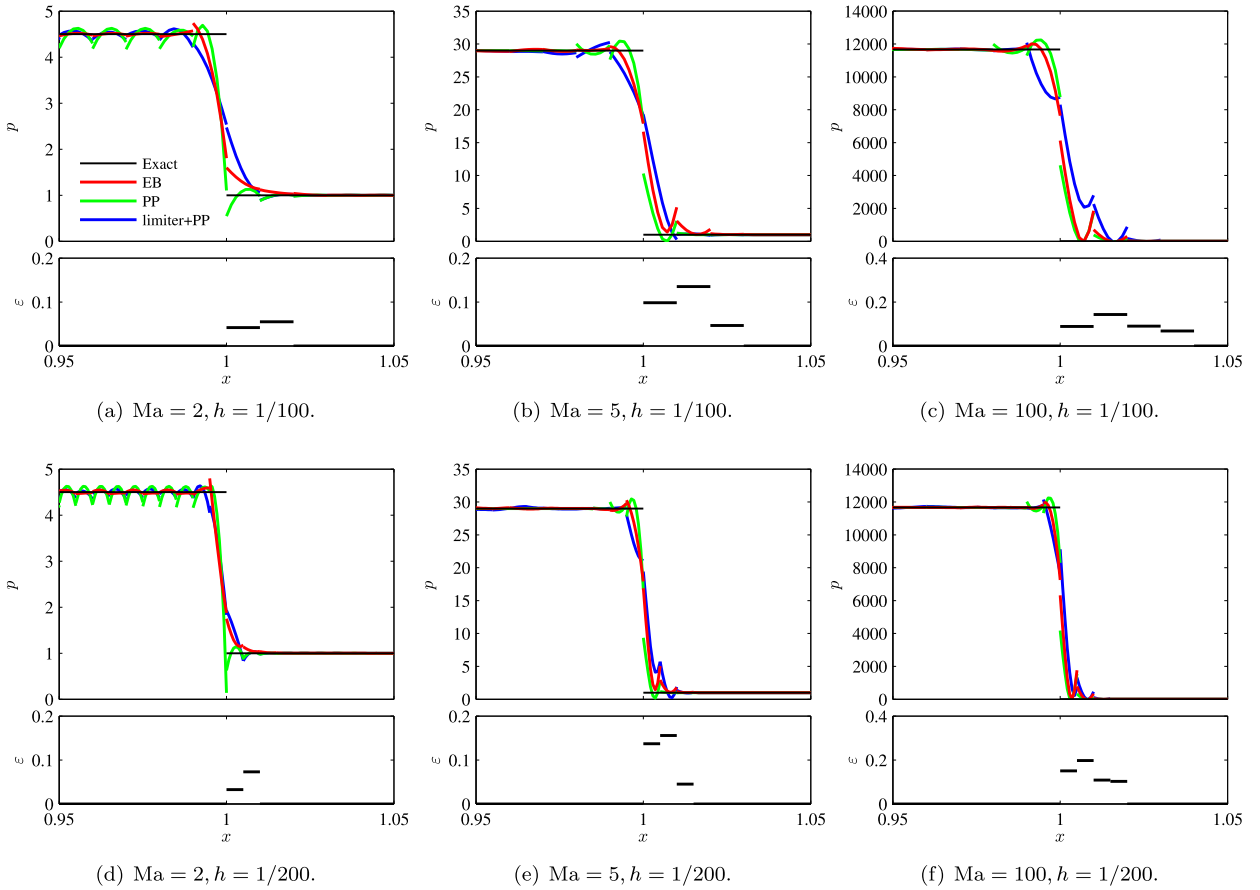
The accuracy is examined by considering different spatial resolutions and polynomial orders. For each polynomial order, the CFL number is assigned to 0.8 $\mathrm{CFL}^{\mathrm{EB}}$, in which $\mathrm{CFL}^{\mathrm{EB}}$ is taken from Table 1. Initially, s_0 is set to $\ln(0.874)$, corresponding to the minimum entropy value of the initial condition. The SSPRK33 time-integration scheme [25] is used, and the convergence rate is given in Table 2. Although the EBDG-scheme remains stable, it can be seen that the solutions do not reach the optimal convergence rates for DGP3 and DGP4. Since this is a time-dependent calculation, temporal errors accumulate with the applied CFL-condition and the convergence rate degrades to third-order (consistent with the SSPRK33 scheme). To demonstrate this, we switch the time-integration scheme to a standard RK45. As can be seen from Table 3, the optimal convergence rates for all cases are achieved, demonstrating that the optimal convergence for smooth solutions is preserved by the EBDG-scheme. In the following, the standard RK45 is used for all other cases.

8.2. One-dimensional moving shock wave

A moving shock-wave in a one-dimensional domain is considered as a test-case for evaluating the robustness and performance of EBDG for shock-capturing. A domain with $x \in [-0.1, 1.1]$ is considered, in which the initial shock front is located at $x=0$. The domain is initialized in $x<0$ with the following pre-shock state:

Table 3
Convergence test of 1D advection with standard RK45 (here we use density to evaluate the error).

h	DGP1		DGP2		DGP3		DGP4	
	L_2-error	rate	L_2-error	rate	L_2-error	rate	L_2-error	rate
1/10	3.494e−3	–	2.140e−4	–	4.650e−6	–	1.438e−7	–
1/20	7.231e−4	2.273	1.513e−5	3.823	2.920e−7	3.993	4.517e−9	4.992
1/40	1.630e−4	2.150	1.891e−6	3.000	1.826e−8	3.999	1.419e−10	4.992
1/80	3.790e−5	2.105	2.364e−7	3.000	1.141e−9	4.000	4.444e−12	4.997
1/160	9.398e−6	2.012	2.955e−8	3.000	7.134e−11	4.000	1.497e−13	4.892
1/320	2.347e−6	2.002	3.694e−9	3.000	4.463e−12	3.999	8.930e−14	7.453e−1



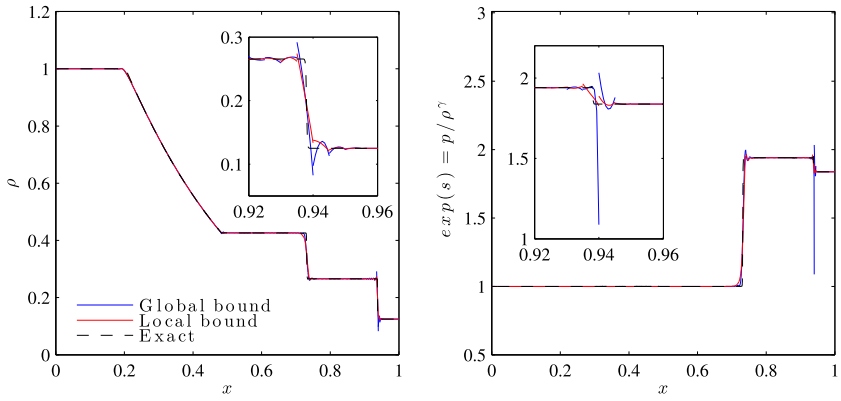
(a) $\mathrm{Ma} = 2, h = 1/100$. (b) $\mathrm{Ma} = 5, h = 1/100$. (c) $\mathrm{Ma} = 100, h = 1/100$.

(d) $\mathrm{Ma} = 2, h = 1/200$. (e) $\mathrm{Ma} = 5, h = 1/200$. (f) $\mathrm{Ma} = 100, h = 1/200$.

Fig. 5. (Color online.) DGP2 simulation of moving shock wave for different Mach numbers (abbreviations: EB – entropy bounding; PP – positivity preserving [9]; Limiter + PP – WENO limiter [7]) with positivity preserving [9].

$$\rho = 1.4\,,$$
$$u = 0\,,$$
$$p = 1\,.$$

Shocks are specified with different Mach numbers ($\mathrm{Ma} = u_s/c$), and $\mathrm{Ma} = \{2, 5, 100\}$ are considered in this case. For all cases considered, the initial value for the entropy, s_0, is set to a value of $\ln(0.620)$, corresponding of the minimum value in the initial condition. The simulation ends when the exact solution of the shock front reaches the location at $x = 1$. Results are illustrated in Fig. 5, showing that the entropy boundedness guarantees the robustness and consistent performance over a wide range of shock strengths. Entropy bounding ($\varepsilon \neq 0$) is only activated in elements that are occupied by flow discontinuities. Compared to the positivity-preserving method, entropy bounding entirely avoids unphysical undershoots in pressure, and provides an improved suppression of oscillations in the post-shock region. Compared to limiting, the entropy bounding shows better robustness in describing shocks at different conditions, introducing lower dissipation in the vicinity of discontinuities.



(a) Density profile (b) Entropy profile

Fig. 6. (Color online.) Comparison of simulation results obtained using different bounding strategies.

8.3. Sod shock-tube: local entropy bound vs. global entropy bound

In this section, we compare simulation results obtained by two different limiting implementations that were discussed in Section 6. We consider the classic Sod shock-tube case with initial conditions given by:

$$(\rho, u, p)^T = \begin{cases} (1.0,\ 0.0,\ 1.0)^T & \text{for } x \le 0.5\,, \\ (0.125,\ 0.0,\ 0.1)^T & \text{for } x > 0.5\,. \end{cases} \tag{68}$$

The one-dimensional domain with $x \in [0, 1]$ is discretized using 200 elements and the DGP2 scheme is used. For limiting with a global entropy bound, the bound is fixed to be $s_0 = \ln(1)$ based on the minimum entropy of the initial condition. In the case that limiting is enforced with a local entropy bound, we first initialize the entropy bound for each cell with s_0. Subsequently, the local bound for each element is estimated according to Eqs. (66) and (67) and updated at every iteration. The simulation runs until $t = 0.25$, and results are shown in Fig. 6. It can be seen that limiting using the local entropy bound outperforms the global limiter, resulting in considerably lower overshoots and undershoots at the shock front. The reason for this can be deduced from the entropy profile in Fig. 6(b). The global minimum entropy in this problem is prescribed at the left side, with a lower magnitude than the entropy value at the right-moving shock. Therefore, the operator $\mathcal{L}$ that is built on such a global minimum loses efficacy in identifying troubled elements around the shock. Except for the solution around the shock, differences in the solution away from the discontinuity are marginal.

8.4. Two-dimensional flow over a cylinder

In this section, we verify the convergence order of the EBDG-scheme for high-order curved elements by considering a two-dimensional flow over a round cylinder. The radius of the cylinder is $R = 1$ and the far-field boundary is a concentric circle with $R = 20$. The condition in the free stream is given as:

$$\rho_\infty = 1.4\,,$$
$$u_\infty = 5.32\,,$$
$$v_\infty = 0.0\,,$$
$$p_\infty = 1\,.$$

The corresponding Mach number is 0.38 and characteristic boundary conditions are imposed at the far-field. The entire domain is initialized with free-stream conditions and $s_0 = \ln(0.620)$. We compare results on quadrilateral and triangular meshes at three levels of refinement. High-order elements are generated using cubic Lagrangian interpolation to accommodate the curvature of the geometry. It is noteworthy to mention that this treatment guarantees the exactness of the computed cell-averaged quantities. The CFL number is set to the CFL^{EB} number from Table 2 for the corresponding shape and polynomial order, multiplied by a factor of 0.8.

A main issue in these simulations is the occurrence of numerical instabilities that are initiated at the leading edge of the cylinder. As a result of this instability, DGP2 and DGP3 without any entropy-bounding diverge (the code blows up) after few iterations. Previously, limiters have been used in this case for stabilizing the transient solutions [8]. However, for high-order polynomials, it is difficult to develop limiters to achieve the optimal convergence rate without a nontrivial implementation. In contrast, EBDG provides a considerably simpler implementation for enabling high-order simulations for such complex geometric configurations.

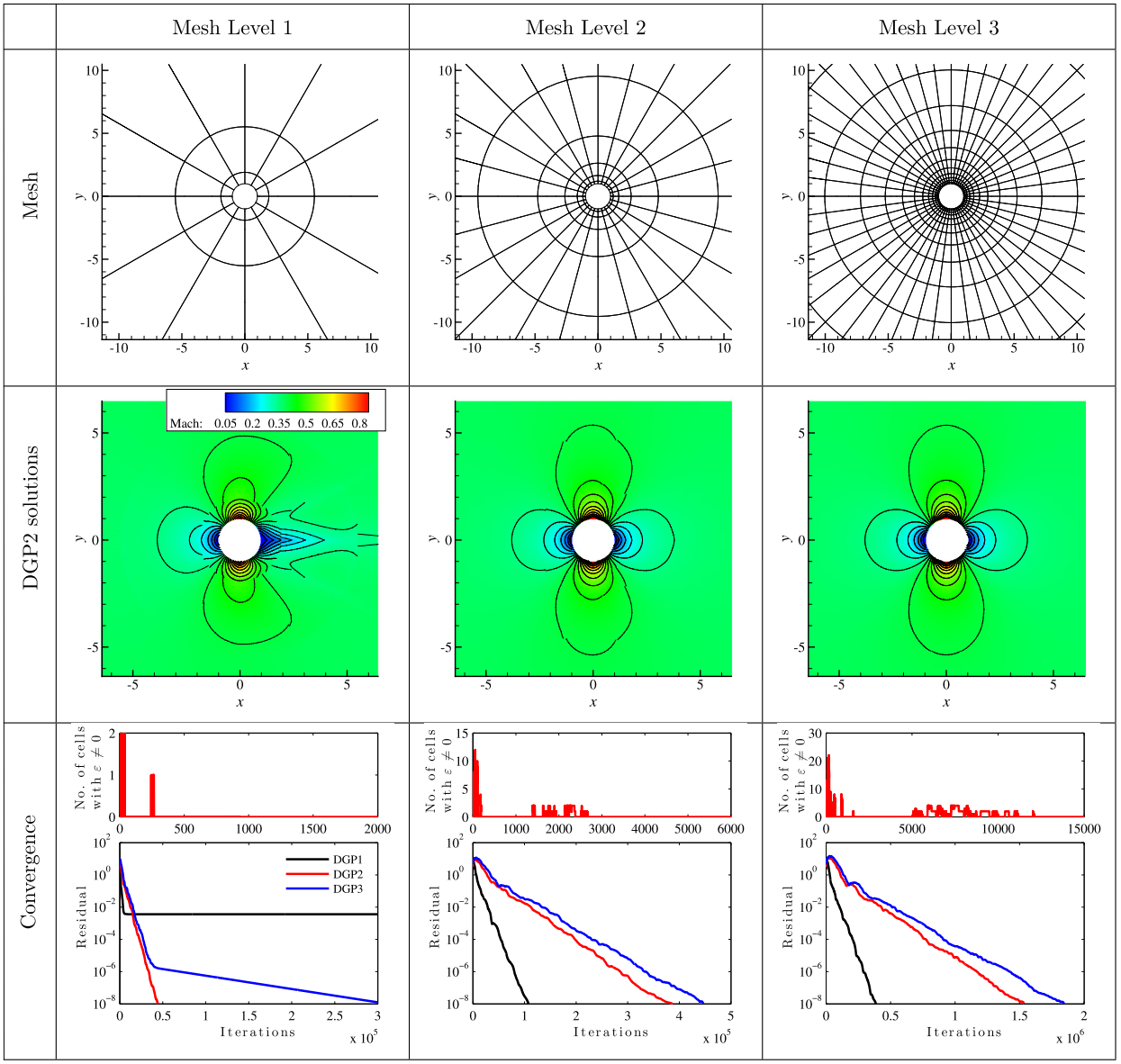


Fig. 7. EBDG-solution of flow over a cylinder on curved quadrilateral meshes with three different refinement levels; top: computational mesh in near-field of the cylinder; middle: Mach number; bottom: convergence history and activation of entropy bounding as a function of iteration.

Comparisons of the computational meshes, simulation results, and convergence properties are presented in Figs. 7 and 8. It is evident that the solution is improved by increasing the mesh resolution. The convergence history of the residual, provided in the last row of both figures, shows that entropy bounding is mostly activated during the start-up phase of the simulation to suppress numerical oscillations and ensure stability. It is interesting to note that the number of elements that require bounding is restricted to the region near the stagnation point upstream of the cylinder, and is confined to less than 8% of the total number of elements. As the solution converges to the steady-state condition, entropy-bounding remains deactivated, retaining the high-order accuracy. Since the solution is smooth the physical entropy production is zero. Therefore, the convergence rates are measured in terms of entropy error using the discrete L_2-norm. A comparison of the convergence rates are presented in Table 4, confirming that the optimal convergence rate is preserved even for complex geometries with curved elements.

8.5. Two-dimensional double Mach reflection

This test case is designed to assess the performance of EBDG for simulations of flows with strong shocks and wave structures. The numerical setup follows this of Woodward and Colella [26], representing a Mach 10 shock over a 30°-wedge.

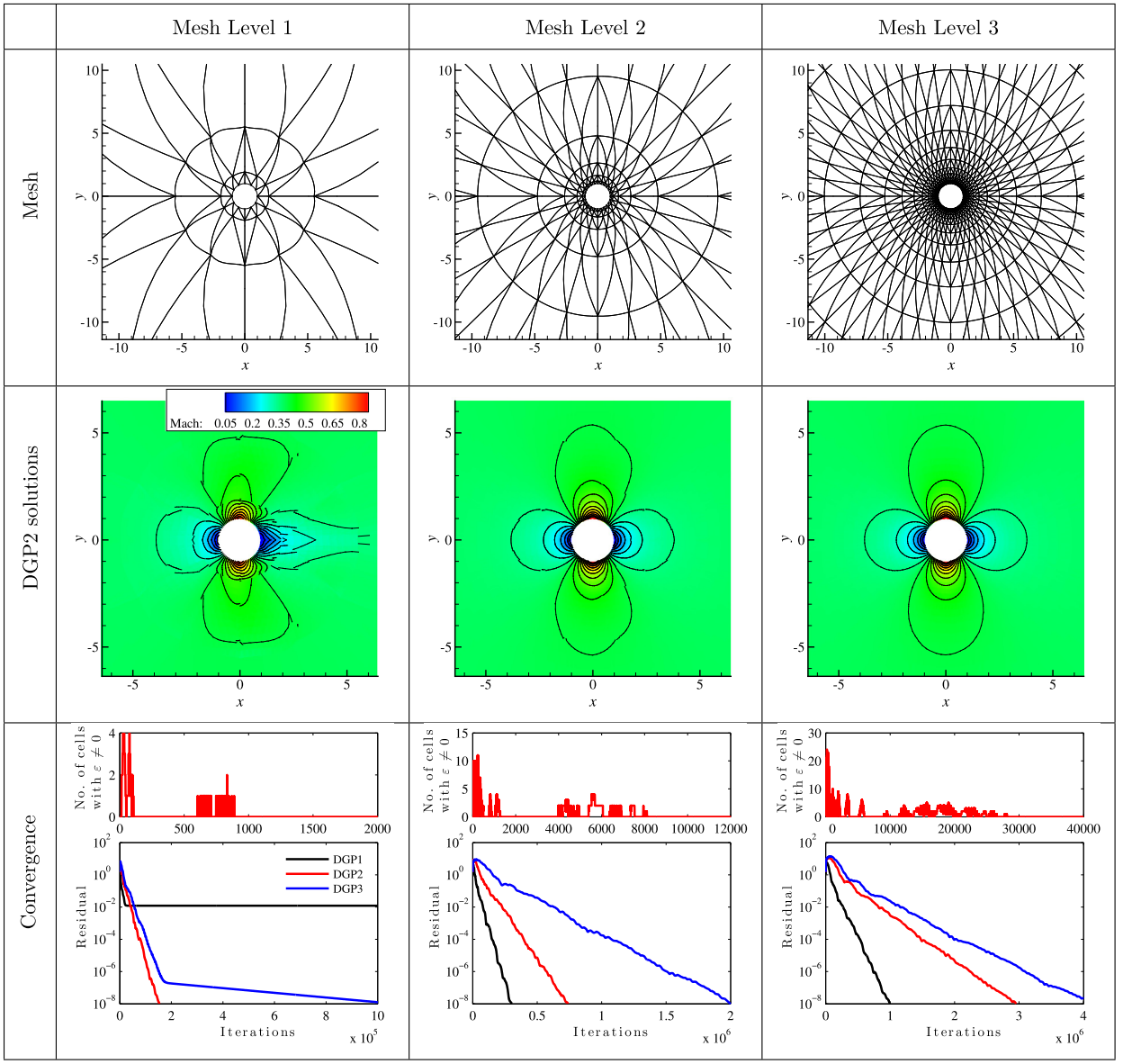
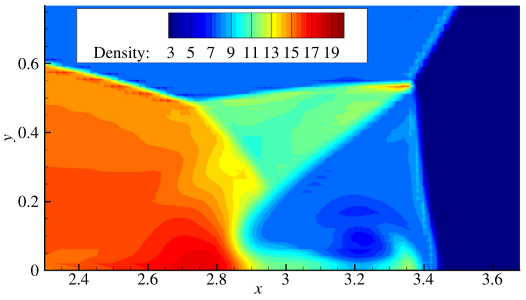


Fig. 8. DG-solution of flow over a cylinder on curved triangular meshes with three different refinement levels; top: computational mesh in the near-field of the cylinder; middle: Mach number; bottom: convergence history and activation of entropy bounding as a function of iteration.

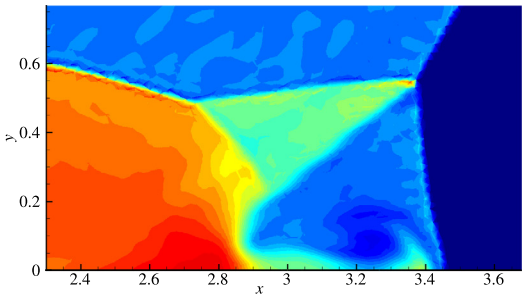
Table 4
Comparisons of convergence rate for 2D flow over a cylinder (here we use entropy to evaluate the error).

Mesh	DGP1		DGP2		DGP3	
	L_2-error	rate	L_2-error	rate	L_2-error	rate
Quadrilateral elements						
Level 1	7.272e−2	–	1.694e−2	–	3.816e−3	–
Level 2	1.318e−2	2.464	7.219e−4	4.552	1.827e−4	4.384
Level 3	2.441e−3	2.433	6.029e−5	3.582	1.036e−5	4.141
Triangular elements						
Level 1	1.137e−1	–	2.590e−2	–	4.086e−3	–
Level 2	1.865e−2	2.608	8.899e−4	4.863	1.291e−4	4.984
Level 3	3.391e−3	2.459	7.222e−5	3.623	6.939e−6	4.217

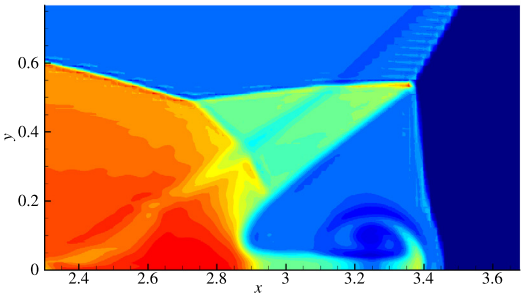
All quantities are non-dimensionalized, and the computational domain is $[0, 4] \times [0, 1]$. In the present study, we consider two different mesh-discretizations, consisting of a Cartesian mesh with quadratic elements ($L_e = h = 0.02$) and a mesh with



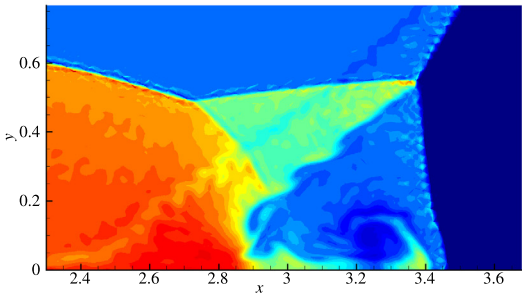
(a) EBDGP1, quadrilateral mesh with $h = 0.02$.



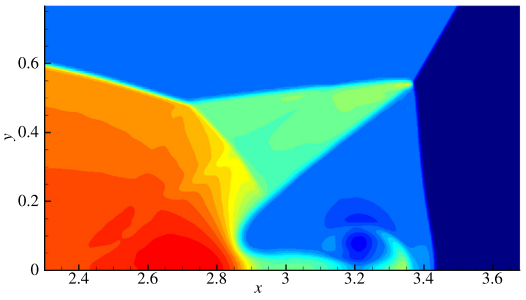
(b) EBDGP1, triangular mesh with $h = 0.02$.



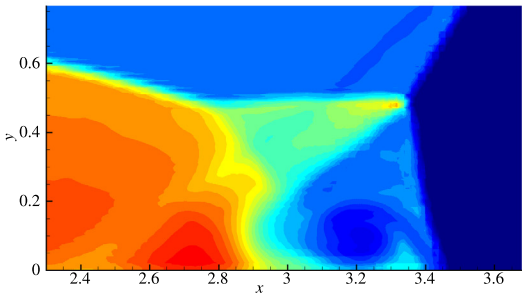
(c) EBDGP2, quadrilateral mesh with $h = 0.02$.



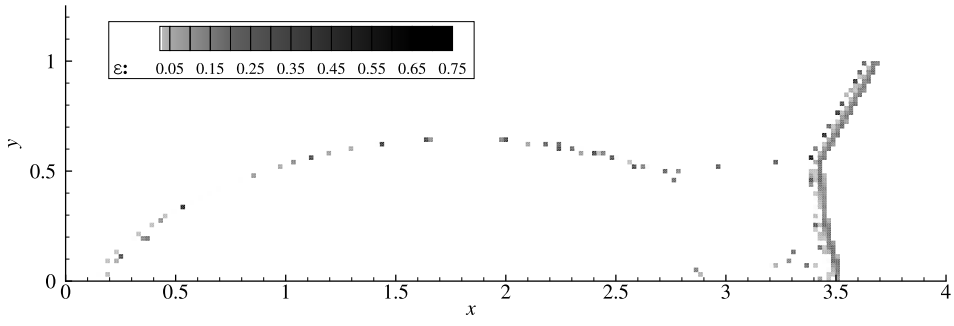
(d) EBDGP2, triangular mesh with $h = 0.02$.



(e) WENO5, quadrilateral mesh with $h = 0.0067$.



(f) DGP2 with WENO limiter [7], quadrilateral mesh with $h = 0.02$.



(g) Instantaneous snapshot of ε for DGP2 on quadrilateral mesh.

Fig. 9. (Color online.) Simulation results of double Mach reflection over a 30°-wedge.

triangular elements ($L_e \approx 0.02$). The pre-shock state is the same as that in Section 8.2 and hence s_0 is set to ln(0.620). The CFL number is prescribed from Table 2 using a safety factor of 0.8.

Simulation results for density contours at time $t = 0.25$ are shown in Fig. 9. The proposed EBDG-method captures all wave-features, and it is found that without enforcing the entropy constraint the solution diverges in the first iteration for

these strong shock conditions. For comparison, a reference solution obtained using a fifth-order WENO-scheme is shown in Fig. 9(e), and results from a DGP2-simulation using a WENO-limiter [7] are presented in Fig. 9(f). Comparisons between EBDGP1 and EBDGP2 results show the benefit of the high-order scheme in providing improved representations of the shock-wave structure. At the same degrees of freedom, the DGP2-solution provides comparable predictions to that of the fifth-order WENO scheme, except for the small oscillations that cannot be removed by the linear scaling procedure. Compared to the DG-simulation with WENO-limiter (Fig. 9(f)), EBDG effectively avoids introducing excessive numerical dissipation since the solution is only entropy-constrained in regions in which the entropy condition is violated.

8.6. Three-dimensional supersonic flow over a sphere

This test case extends the evaluation of the EBDG-method to three-dimensional configurations with complex geometries. Currently, robust approaches for capturing strong shocks in three-dimensional curved elements are still subject to investigation. This test case considers a flow at a Mach number of 6.8 over a sphere. The radius of the sphere is $R = 1$. Due to the geometric symmetry, the computational domain considers only an eighth section of the domain, and it is extended to $3R$ in radial direction. Symmetry boundary conditions are imposed at the planes $y = 0$ and $z = 0$, and outflow boundary conditions are prescribed at $x = 0$. Normal velocity inflow is prescribed at the outer shell with the following specification:

$$\rho_\infty = 1.4\,,$$
$$u_\infty = -6.80\,,$$
$$v_\infty = 0.0\,,$$
$$w_\infty = 0.0\,,$$
$$p_\infty = 1\,.$$

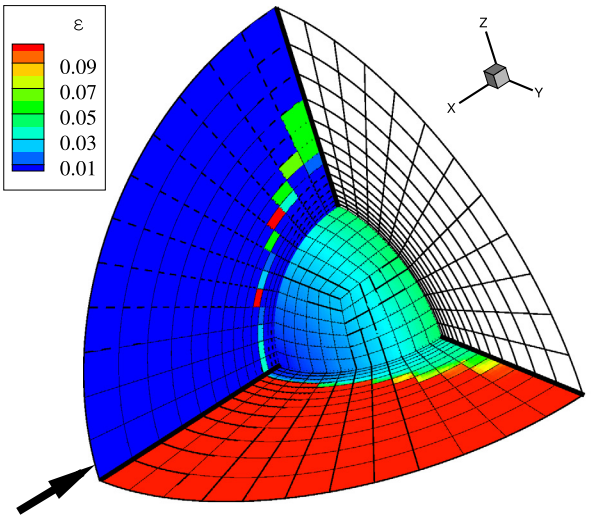
Slip-wall conditions are imposed at the surface of the sphere. The computational domain is discretized with quadratically curved hexahedral elements. For the initial mesh, the radial dimension is partitioned with 14 elements with a linear stretching factor of 1.1 while the azimuthal dimension of the plane at $x = 0$ is partitioned using 12 elements. DGP2 is applied for this case and the CFL number is 0.8 $\mathrm{CFL}^{\mathrm{EB}}$ with $\mathrm{CFL}^{\mathrm{EB}} = 0.056$.

Simulation results are illustrated in Fig. 10(a), showing the surface mesh and isocontours of the Mach number. The bounding parameter ε can be utilized as an indicator for local mesh refinement. We sample the elements with non-zero ε values over few iterations, and then locally refine these elements. Results using one and two levels of refinement are shown in Figs. 10(b) and 10(c), respectively. This direct comparison shows that the shock profiles become sharper with increasing resolution. Since the bounding parameter is sharp, the mesh-refinement is confined to a narrow region in the vicinity of the shock. The flow-field solution behind the shock is smooth, and no entropy bounding is applied in this region. To provide a quantitative analysis, simulation results from the EBDG-method are compared against measurements by Billig [27] in Fig. 10(d), showing good agreement between experiments and computations.

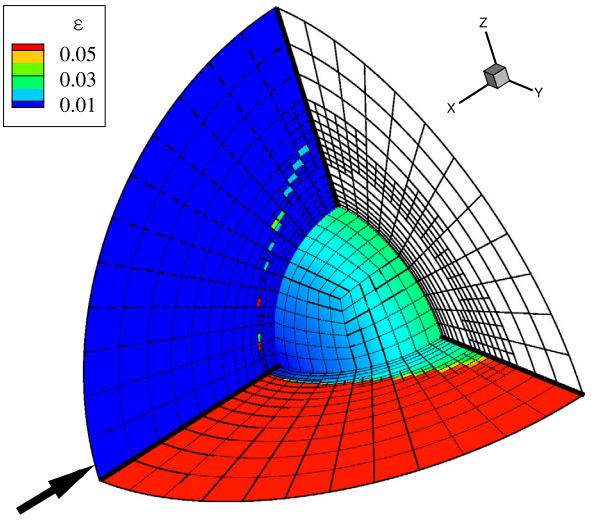
9. Conclusions

A regularization technique for the discontinuous Galerkin scheme was developed using the entropy principle. Motivated by the FV-entropy solution, the high-order DG-scheme is stabilized by constraining the solution to obey the entropy condition. The implementation of the resulting entropy-bounding discontinuous Galerkin scheme relies on two key components, namely a limiting operator and a CFL-constraint. These essential components were derived by considering first a one-dimensional setting and the subsequent extension to multi-dimensional configurations with arbitrary and curved elements. Specifically, utilizing the interpolation basis we were able to extend the entropy bounding (also including positivity preserving) to arbitrarily shaped elements independent of specific quadrature rules. The bounding procedure is obtained from algebraic operations, resulting in a computationally efficient and simple implementation. A sufficient CFL-condition was rigorously derived and proofed to ensure that the entropy constraint can be enforced on different types and orders of elements. By considering different configurations, numerical tests were conducted to examine accuracy and stability of the entropy-bounding DG-scheme. These test cases confirm the efficacy in regularizing solutions in the vicinity of discontinuities, generated either by true flow physics or during the transient solution update. The added benefit of the entropy bounding method is its utilization as a refinement indicator.

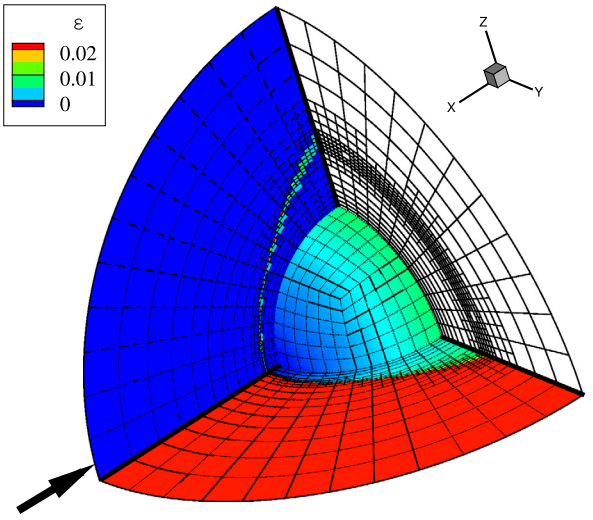
Since the herein proposed entropy bounding scheme relies on a linear scaling operator, it is not capable to remove shock-triggered oscillations of smaller magnitude, although it stabilizes the solution and prevents the solver from diverging. As a final remark, the derivation that was presented in this study is general and extendable to other discontinuous schemes with sub-cell solution representations, such as spectral finite volume schemes [28] and the flux reconstruction scheme [29]. Therefore, entropy-bounding, as an idea, has the potential to improve the robustness of shock-capturing for these emerging high-order numerical methods.



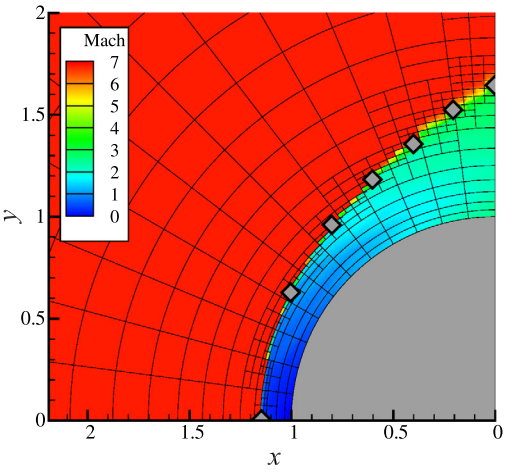
(a) Baseline mesh.



(b) Local refinement at level one.



(c) Local refinement at level two.



(d) Comparison with measurement in [27].

Fig. 10. (Color online.) Simulations of Mach 6.8 flow over a sphere showing the ε profile ($y = 0$ plane) and Mach-number distribution ($z = 0$ plane) on (a) baseline mesh, and simulation results with local refinement with (b) one refinement level and (c) two refinement levels. Comparisons of the shock location with measurements by Billig [27] are shown in (d).

Acknowledgements

Financial support through NSF with Award No. CBET-0844587 is gratefully acknowledged. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ASC130004. Helpful discussions with Yee Chee See on the mathematical analysis are appreciated.

Appendix A. Combination rule

In this section, we derive an estimate for the maximum characteristic speed for convex state solution. For this, we consider a state vector U of Eq. (2a), which is written in the following form:

$$\mathsf{U} = \sum_k \beta_k \mathsf{U}_k \,, \tag{69}$$

where $\beta_k > 0$ and $\sum_k \beta_k = 1$. The maximum characteristic speed of U is:

$$\nu(\mathsf{U}) = |u(\mathsf{U})| + c(\mathsf{U}) = |u(\mathsf{U})| + \sqrt{\gamma(\gamma-1)\left(e(\mathsf{U}) - \frac{1}{2}|u(\mathsf{U})|^2\right)},$$

in which u and E can be calculated according to Eq. (69) as

$$u(\mathsf{U}) = \sum_k \alpha_k u(\mathsf{U}_k), \qquad e(\mathsf{U}) = \sum_k \alpha_k e(\mathsf{U}_k),$$

and

$$\alpha_k = \frac{\beta_k \rho(\mathsf{U}_k)}{\sum_k \beta_k \rho(\mathsf{U}_k)}$$

is a new set of coefficients that is introduced to convert from conservative to primitive variables. Furthermore, because of

$$\begin{aligned}
\gamma(\gamma-1)e(\mathsf{U}) &= \sum_k \alpha_k \left(c^2(\mathsf{U}_k) + \frac{\gamma(\gamma-1)}{2}|u(\mathsf{U}_k)|^2\right), \\
|u(\mathsf{U})| &= \sqrt{|\sum_k \alpha_k u(\mathsf{U}_k)|^2}, \\
&\le \sqrt{\sum_k \alpha_k |u(\mathsf{U}_k)|^2},
\end{aligned}$$

we obtain

$$\begin{aligned}
\nu(\mathsf{U}) &= |u(\mathsf{U})| + \sqrt{\sum_k \alpha_k c^2(\mathsf{U}_k) + \frac{\gamma(\gamma-1)}{2}\left(\sum_k \alpha_k |u(\mathsf{U}_k)|^2 - |u(\mathsf{U})|^2\right)}, \\
&\le \sqrt{\sum_k \alpha_k |u(\mathsf{U}_k)|^2} + \sqrt{\sum_k \alpha_k c^2(\mathsf{U}_k) + \frac{\gamma(\gamma-1)}{2}\sum_k \alpha_k |u(\mathsf{U}_k)|^2}, \\
&\le \sqrt{2\sum_k \alpha_k c^2(\mathsf{U}_k) + (2+\gamma(\gamma-1))\sum_k \alpha_k |u(\mathsf{U}_k)|^2}, \\
&\le \sqrt{2+\gamma(\gamma-1)}\sqrt{\sum_k \alpha_k (c^2(\mathsf{U}_k) + |u(\mathsf{U}_k)|^2)}, \\
&\le \sqrt{2+\gamma(\gamma-1)}\sqrt{\sum_k \alpha_k (c(\mathsf{U}_k) + |u(\mathsf{U}_k)|)^2}, \\
&\le \sqrt{2+\gamma(\gamma-1)}\max_k \{c(\mathsf{U}_k) + |u(\mathsf{U}_k)|\}.
\end{aligned}$$

We used this estimation for preselecting the dissipation coefficient λ^* in Lemma 4.

References

[1] B. Cockburn, C.-W. Shu, TVB Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws III: one dimensional systems, J. Comput. Phys. 84 (1989) 90–113.
[2] B. Cockburn, C.-W. Shu, The Runge–Kutta discontinuous Galerkin method for conservation laws V: multidimensional systems, J. Comput. Phys. 141 (1998) 199–224.
[3] R. Biswas, K. Devine, J.E. Flaherty, Parallel adaptive finite element methods for conservation laws, Appl. Numer. Math. 14 (1994) 255–284.
[4] L. Krivodonova, Limiters for high-order discontinuous Galerkin methods, J. Comput. Phys. 226 (2007) 879–896.
[5] J. Qiu, C.-W. Shu, Runge–Kutta discontinuous Galerkin method using WENO limiters, SIAM J. Sci. Comput. 26 (2005) 907–929.
[6] J. Zhu, J. Qiu, C.-W. Shu, M. Dumbser, Runge–Kutta discontinuous Galerkin method using WENO limiters II: unstructured meshes, J. Comput. Phys. 227 (2008) 4330–4353.
[7] X. Zhong, C.-W. Shu, A simple weighted essentially nonoscillatory limiter for Runge–Kutta discontinuous Galerkin methods, J. Comput. Phys. 232 (2013) 397–415.
[8] H. Luo, J.D. Baum, R. Löhner, A Hermite WENO-based limiter for discontinuous Galerkin method on unstructured grids, J. Comput. Phys. 225 (2007) 686–713.
[9] X. Zhang, C.-W. Shu, On positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes, J. Comput. Phys. 229 (2010) 8918–8934.

[10] C. Wang, X. Zhang, C.-W. Shu, J. Ning, Robust high order discontinuous Galerkin schemes for two-dimensional gaseous detonations, J. Comput. Phys. 231 (2) (2012) 653–665.
[11] X. Zhang, C.-W. Shu, A minimum entropy principle of high order schemes for gas dynamics equations, Numer. Math. 121 (2012) 545–563.
[12] Y. Lv, M. Ihme, Computational analysis of re-ignition and re-initiation mechanisms of quenched detonation waves behind a backward facing step, Proc. Combust. Inst. 35 (2015) 1963–1972.
[13] Y. Lv, M. Ihme, Discontinuous Galerkin method for multicomponent chemically reacting flows and combustion, J. Comput. Phys. 270 (2014) 105–137.
[14] P.D. Lax, Shock waves and entropy, in: E.H. Zarantonello (Ed.), Contributions to Nonlinear Functional Analysis, Academic Press, New York, London, 1971, pp. 603–634.
[15] E. Tadmor, A minimum entropy principle in the gas dynamics equations, Appl. Numer. Math. 2 (1986) 211–219.
[16] S.L. Sobolev, V. Vaskevich, The Theory of Cubature Formulas, Springer, 1997.
[17] B. Perthame, C.-W. Shu, On positivity preserving finite volume schemes for Euler equations, Numer. Math. 73 (1996) 119–130.
[18] E. Tadmor, Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems, Acta Numer. 12 (2003) 451–512.
[19] B. Khobalatte, B. Perthame, Maximum principle on the entropy and second-order kinetic schemes, Math. Comput. 62 (1994) 119–131.
[20] Z.J. Wang, K. Fidkowski, R. Abgrall, F. Bassi, D. Caraeni, A. Cary, H. Deconinck, R. Hartmann, K. Hillewaert, H.T. Huynh, N. Kroll, G. May, P.-O. Persson, B. van Leer, M. Visbal, High-order CFD methods: current status and perspective, Int. J. Numer. Methods Eng. 72 (2013) 811–845.
[21] X. Zhang, C.-W. Shu, Positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations with source terms, J. Comput. Phys. 230 (2011) 1238–1248.
[22] B. Cockburn, C.-W. Shu, Runge–Kutta discontinuous Galerkin methods for convection-dominated problems, J. Sci. Comput. 16 (3) (2001) 173–261.
[23] D. Dunavant, High degree efficient symmetrical Gaussian quadrature rules for the triangle, Int. J. Numer. Methods Eng. 21 (1985) 1129–1148.
[24] L. Zhang, T. Cui, H. Liu, A set of symmetric quadrature rules on triangles and tetrahedra, J. Comput. Math. 21 (2009) 89–96.
[25] S. Gottlieb, C.-W. Shu, E. Tadmor, Strong stability-preserving high-order time discretization methods, SIAM Rev. 43 (2001) 89–112.
[26] P.R. Woodward, P. Colella, The numerical simulation of two-dimensional fluid flow with strong shocks, J. Comput. Phys. 54 (1984) 115–173.
[27] F.S. Billig, Shock-wave shapes around spherical- and cylindrical-nosed bodies, J. Spacecr. Rockets 4 (1967) 822–823.
[28] Z.J. Wang, Y. Liu, Spectral (finite) volume method for conservation laws on unstructured grids V: extension to three-dimensional systems, J. Comput. Phys. 212 (2006) 454–472.
[29] H.T. Huynh, A flux reconstruction approach to high-order schemes including discontinuous Galerkin methods, in: 18th AIAA Computational Fluid Dynamics Conference, Miami, FL, 2007, AIAA 2007-4079.