Numerische Mathematik (2019) 141:1079–1120
https://doi.org/10.1007/s00211-019-01027-9

Numerische
Mathematik

An analysis of discontinuous Galerkin methods for the compressible Euler equations: entropy and L_2 stability

David M. Williams[1]

Received: 4 August 2017 / Revised: 25 August 2018 / Published online: 9 February 2019

Abstract

The objective of this article is to characterize the entropy and L_2 stability of several representative discontinuous Galerkin (DG) methods for solving the compressible Euler equations. Towards this end, three DG methods are constructed: one DG method with entropy variables as its unknowns, and two DG methods with conservative variables as their unknowns. These methods are employed in order to discretize the compressible Euler equations in space. Thereafter, the resulting semi-discrete formulations are analyzed, and the entropy and L_2 stability characteristics are evaluated. It is shown that the semi-discrete formulation of the DG method with entropy variables is entropy and L_2 stable. Furthermore, it is shown that the semi-discrete formulations of the DG methods with conservative variables are only guaranteed to be entropy and L_2 stable under the following assumptions: the entropy projection errors vanish, or the terms containing the entropy projection errors are non-positive. Thereafter, the semi-discrete formulation with entropy variables, and one of the semi-discrete formulations with conservative variables, are discretized in time with an 'algebraically stable' Runge–Kutta (RK) scheme. The resulting formulations are fully-discrete and can be immediately applied to practical problems. In this article, they are employed to simulate a vortex propagating for long distances. It is shown that temporal stability is maintained by the DG method with entropy variables, but the DG method with conservative variables exhibits instability.

Mathematics Subject Classification 65M12 · 65M60 · 76N99

✉ David M. Williams
david.m.williams@psu.edu

[1] Department of Mechanical Engineering, The Pennsylvania State University, University Park,
PA 16802, USA

1 Introduction

The compressible Euler equations for gas dynamics are a non-linear system of hyperbolic conservation laws that are inherently difficult to solve. They are remarkably complex, as their solutions allow the propagation of waves over large distances, and the formation of sharp gradients and discontinuities. Naturally, the numerical methods for solving the compressible Euler equations must possess a high degree of robustness and flexibility in order to accurately handle these phenomena. In addition, it is desirable for the numerical methods to be capable of operating on unstructured grids in order to facilitate the treatment of complex geometries. Arguably, these requirements narrow the range of acceptable methods down to two distinct categories: (1) Finite Volume methods (FVMs), and (2) Finite Element methods (FEMs). The existence of robust, high-order FVMs that are capable of operating on (at least) hexahedron and tetrahedron elements in 3D is well-known. The most popular FVMs of this type are the 'essentially non-oscillatory' (ENO) schemes [34–37,58,59] and 'weighted-essentially non-oscillatory' (WENO) schemes [3,41,50,57]. These schemes are able to achieve high-order accuracy while simultaneously controlling spurious oscillations that arise due to aliasing errors and discontinuities. Unfortunately, high-order FVMs generally require the construction of large stencils that may artificially link elements that are in disparate parts of the mesh. In order to avoid this issue, many researchers have pursued the development of FEMs (i.e., schemes in the second category). It is common knowledge that FEMs, and in particular discontinuous FEMs (DFEMs), can achieve high-order accuracy while maintaining a compact stencil by employing a high-order polynomial basis that is locally constructed on each element [15,25,38]. The resulting schemes are flexible and can operate on many element types, including hexahedrons, tetrahedrons, prisms, and pyramids in 3D. Unfortunately, DFEMs tend to be less robust when dealing with aliasing errors and discontinuities relative to their FVM counterparts. For this reason, it is desirable to obtain a better understanding of the stability of DFEMs for solving the compressible Euler equations. Currently, the stability of DG methods [14,16–18,20] is of interest, as they are arguably the most popular DFEMs for solving the compressible Euler equations—although other alternatives exist, including the well-known discontinuous Petrov Galerkin (DPG) methods [10,11,22–24,27,70]. Regrettably, it is not possible to cover *all* existing methods in this work, and therefore the main focus will be on the stability of the more popular DG methods.

Before proceeding further, it important to clarify the intended meaning of the word 'stability' when the stability of a numerical scheme, or more specifically the stability of a DG scheme, is discussed. There are three distinct types of stability that should generally be considered when the compressible Euler equations are discretized via numerical schemes: (1) the stability of the semi-discrete formulation obtained by discretizing in space utilizing a particular numerical scheme, (2) the stability of the fully-discrete formulation obtained by discretizing in space and time utilizing a particular numerical scheme, and (3) the stability of the fully-discrete formulation obtained by discretizing in space and time utilizing two different numerical schemes. This work will primarily focus on establishing theoretical results that govern the first type of stability for DG schemes. The second type of stability is not covered for the sake of brevity, although, it is believed that an extension of the analysis in this work to space-time DG schemes is straightforward. Finally, the third type of stability is briefly covered by numerical experiments in the latter part of this work.

In order to assess the first type of stability for a DG scheme, i.e. stability of the semi-discrete formulation, it is useful to employ the criteria of 'entropy stability' and 'L_2 stability'. Entropy stability is considered by many to be an essential requirement for a numerical scheme, as it ensures that the scheme satisfies a specific set of entropy inequalities. In turn, these inequalities ensure that the scheme will converge to the appropriate 'entropy solutions' (cf. [21,65]) of the compressible Euler equations. This is of particular importance for FEMs, as their construction utilizes the weak form of the governing equations, and as a result, it is possible to obtain a wide class of weak solutions, some of which may not be physically realizable. Naturally, if an FEM is entropy stable, then it immediately follows that only the realizable entropy solutions will be obtained from amongst these weak solutions [21,39,65]. Note that these entropy solutions are not necessarily unique, however, they are (at minimum) guaranteed to satisfy the second law of thermodynamics at the discrete level.

Now, in order to establish entropy stability for an FEM, it is necessary to first symmetrize the compressible Euler equations by rewriting them in terms of so-called 'entropy variables'. Note that these variables have been carefully obtained by a number of researchers for the compressible Euler equations [32,33,48,51] and also for the compressible Navier–Stokes equations [42,56]. In order to complete the proof of entropy stability, it is necessary to discretize the resulting symmetrized equations and to establish an inequality that governs an entropy measure of the solution in time. This procedure is carried out for a continuous FEM in [56], as entropy stability for the compressible Navier–Stokes equations is proven for a space-time streamline upwind Petrov Galerkin (SUPG) scheme which utilizes entropy variables as its unknowns, in place of the usual conservative variables. The interested reader is referred to [42–45,56] for additional details on the SUPG scheme and its stability. In [5,6] a similar result is obtained for space-time DG schemes for the compressible Euler equations. In [39, 40,46] similar efforts were undertaken to prove entropy inequalities for DG schemes applied to general hyperbolic systems of conservation laws. Finally, in [69] entropy stability for the compressible Navier–Stokes equations is proven for space-time hybridizable DG (HDG) methods which utilize entropy variables as their unknowns.

As a side note, entropy stability has also been established for certain classes of Finite Difference (FD) schemes. These schemes are robust, but are limited to structured grids. For more information, the reader is encouraged to consult [28,29,49,64,65].

L_2 stability is another essential requirement for a numerical scheme. The classical L_2 stability condition governs the L_2 norm of the conservative variables, and is the only known a priori global bound that can be obtained for the solution [21,39]. The idea of L_2 stability is very closely related to the idea of entropy stability. In fact, if a scheme is entropy stable, it is straightforward to demonstrate (under relatively mild additional assumptions) that it is also L_2 stable [21]. This is a remarkable fact that further emphasizes the importance of entropy stability.

The previous research in this area (see above) has usually focused on proving the entropy and L_2 stability of space-time DG methods (and similar FEMs), all of which utilize entropy variables as their unknowns. This is valuable groundwork, however it is incomplete, as the vast majority of researchers solve the compressible Euler

Table 1 A summary of the stability properties of DG methods for the compressible Euler equations

	DG method with entropy variables Eq. (4.1)	DG method with conservative variables Eq. (4.3)	DG method with conservative variables Eq. (4.4)
Entropy stable	✓	–	–
Entropy stable*	✓	$\checkmark^{\dagger}$	$\checkmark^{\dagger}$
L_2 stable	$\checkmark^{\dagger}$	–	–
L_2 stable*	$\checkmark^{\dagger}$	$\checkmark^{\dagger}$	$\checkmark^{\dagger}$

The significance of the symbols in the table are explained in the text

Table 2 A summary of the theorems that establish the results in Table 1

	DG method with entropy variables Eq. (4.1)	DG method with conservative variables Eq. (4.3)	DG method with conservative variables Eq. (4.4)
Entropy stable	Thm. 5.5	–	–
Entropy stable*	Thm. 5.5	Thm. $6.2^{\dagger}$	Thm. $6.8^{\dagger}$
L_2 stable	Thms. 7.4, $7.7^{\dagger}$	–	–
L_2 stable*	Thms. 7.4, $7.7^{\dagger}$	Thms. $8.2^{\dagger}$, $8.3^{\dagger}$	Thms. $8.2^{\dagger}$, $8.3^{\dagger}$

The significance of the symbols in the table are explained in the text

equations by spatially discretizing them with DG methods which utilize *conservative* variables as their unknowns. While there is significant numerical evidence that would suggest that these schemes are stable [15,38,67], the stability of these schemes has not been thoroughly investigated mathematically. The purpose of this work is to precisely identify the circumstances under which entropy and L_2 stability are ensured for DG methods which utilize conservative variables as their unknowns. In addition, this work summarizes and extends a number of well-known results that govern the stability of DG methods which utilize entropy variables as their unknowns. As mentioned previously, all the theoretical results in this work are given for the semi-discrete formulations of these schemes. The stability of the fully-discrete formulations are explored through some numerical experiments at the end of this work.

2 Summary of the main theoretical results and layout

In this work, three DG methods are introduced for the purpose of spatially discretizing the compressible Euler equations. Thereafter, the entropy stability, L_2 stability, and closely related concepts are investigated for each scheme. The results of this investigation are summarized in Tables 1 and 2: Table 1 contains an overview of the results and Table 2 contains an enumeration of the associated theorems. In the tables, a check mark ✓ indicates that a standard result was obtained, and a dash mark – indicates that a result could not be obtained. In addition, a checkmark with a dagger $\checkmark^{\dagger}$ (or simply a dagger †) indicates that a new result, or an important extension of a existing result was obtained. Finally, an asterisk $*$ means that a statement holds true when entropy projection errors vanish pointwise in each element, or when the entropy projection terms are non-positive.

The terminology utilized in Tables 1 and 2, and the assumptions under which the results were obtained, are fully elucidated in the remaining sections of this work. In Sect. 3, the compressible Euler equations are introduced, and in Sect. 4, the associated DG schemes are constructed. In Sects. 5 and 6, the entropy stability characteristics of the schemes are evaluated. In Sects. 7 and 8, the L_2 stability characteristics of the schemes are evaluated. Finally, in Sect. 9, the fully-discrete, temporal stability characteristics of the schemes are evaluated with numerical experiments.

In order to conclude this section, it is appropriate to comment on the relative strengths and weaknesses of the DG schemes. Based on Tables 1 and 2, it is clear that the DG scheme with entropy variables is more stable than the other schemes in the sense that it satisfies the entropy stability and L_2 stability criteria directly, whereas the DG schemes with conservative variables are only entropy and L_2 stable when entropy projection errors vanish, or when the entropy projection terms are non-positive.

3 The compressible Euler equations and entropy variables

The compressible Euler equations can be written as follows

$$\boldsymbol{u}_{,t} + \boldsymbol{f}^{i}_{,\boldsymbol{x}_i} = \boldsymbol{0} \quad \text{in } \Omega \times T, \tag{3.1}$$

where t and $\boldsymbol{x}_i$ are the temporal and spatial coordinates in the one-dimensional temporal domain T and the d-dimensional bounded spatial domain Ω. In addition, $\boldsymbol{u}$ is the m-valued solution, $\boldsymbol{u} : \Omega \times T \rightarrow \mathbb{R}^m$, where $m = d + 2$. Furthermore, each $\boldsymbol{f}^i = \boldsymbol{f}^i (\boldsymbol{u})$ contains the m-valued components of the inviscid fluxes in the ith spatial direction, $\boldsymbol{f}^i : \mathbb{R}^m \rightarrow \mathbb{R}^m$, where $1 \leq i \leq d$. The reader is encouraged to consult "Appendix A" for a useful explanation of the notation utilized here and in the remainder of this work.

On its own, Eq. (3.1) is not complete, but must be combined with an initial condition $\boldsymbol{u} = \boldsymbol{u}(t_0, \boldsymbol{x}) = \boldsymbol{u}(t_0)$ and boundary conditions of the following type

$$\mathcal{L}(\boldsymbol{u}, \boldsymbol{u}^{\partial}) = \boldsymbol{0} \quad \text{on } \partial\Omega \times T,$$

where $\mathcal{L} : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a linear function of the solution and information specified on the boundary (denoted by $\boldsymbol{u}^{\partial}$).

In order to fix ideas, one may briefly review the following specific definitions of $\boldsymbol{u}$ and $\boldsymbol{f}^i$ for the compressible Euler equations

$$\boldsymbol{u} = \begin{bmatrix} \rho \\ \rho \left\{\mathbf{V}^j\right\} \\ \rho \left(e + \frac{1}{2}\mathbf{V}^k\mathbf{V}^k\right) \end{bmatrix}, \qquad \boldsymbol{f}^i = \begin{bmatrix} \rho\mathbf{V}^i \\ \left\{\rho\mathbf{V}^i\mathbf{V}^j + p\delta_{ij}\right\} \\ \rho\mathbf{V}^i \left(e + \frac{1}{2}\mathbf{V}^k\mathbf{V}^k + \frac{p}{\rho}\right) \end{bmatrix}.$$

Here, the standard fluid mechanical properties are denoted as follows: ρ is the density, $\mathbf{V} = \left\{\mathbf{V}^i\right\}$ is the velocity vector ($\mathbf{V} = \{u, v, w\}^T$ for $\Omega \subset \mathbb{R}^3$), e is the internal energy, and p is the pressure. One may assume that the temperature T, pressure p, density ρ, and internal energy e are related via the following equations of state

$$p = \rho R T, \qquad e = \frac{p}{\rho\left(\gamma - 1\right)},$$

where R denotes the specific gas constant and γ denotes the ratio of specific heats.

Having introduced several standard definitions, one may now define the well-known entropy variables $\boldsymbol{v} : \varOmega \times T \rightarrow \mathbb{R}^m$. A convenient definition for these variables (due to [42] and [56]) is as follows

$$\boldsymbol{v} = \frac{1}{e\left(\gamma - 1\right)} \begin{bmatrix} -\frac{\mathbf{V}^k \mathbf{V}^k - 2e\left(\gamma - \mathrm{In}\left(e(\gamma-1)\rho^{1-\gamma}\right)\right)}{2e(\gamma-1)} \\ \left\{\mathbf{V}^j\right\} \\ -1 \end{bmatrix}.$$

It is interesting to note that the entropy variables $\boldsymbol{v}$, the solution variables $\boldsymbol{u}$, and the inviscid fluxes $\boldsymbol{f}^i$ are related to each other by entropy functionals, which act from $\mathbb{R}^m \rightarrow \mathbb{R}$, and are denoted by $U\left(\boldsymbol{u}\right)$, $F^i\left(\boldsymbol{u}\right)$, $\mathcal{U}\left(\boldsymbol{v}\right)$, and $\mathcal{F}^i\left(\boldsymbol{v}\right)$. These functionals are implicitly defined such that

$$\mathcal{U}_{,\boldsymbol{v}} = \boldsymbol{u}^T, \qquad \mathcal{U}_{,\boldsymbol{v},\boldsymbol{v}} = \boldsymbol{u}_{,\boldsymbol{v}}, \tag{3.2}$$

$$\mathcal{F}^i_{,\boldsymbol{v}} = \left(\boldsymbol{f}^i\right)^T, \qquad \mathcal{F}^i_{,\boldsymbol{v},\boldsymbol{v}} = \boldsymbol{f}^i_{,\boldsymbol{v}}, \tag{3.3}$$

$$U_{,\boldsymbol{u}} = \boldsymbol{v}^T, \qquad F^i_{,\boldsymbol{u}} = \boldsymbol{v}^T \boldsymbol{f}^i_{,\boldsymbol{u}}, \tag{3.4}$$

$$U\left(\boldsymbol{u}\right) = \boldsymbol{v}^T\left(\boldsymbol{u}\right)\boldsymbol{u} - \mathcal{U}\left(\boldsymbol{v}\left(\boldsymbol{u}\right)\right), \qquad F^i\left(\boldsymbol{u}\right) = \boldsymbol{v}^T\left(\boldsymbol{u}\right)\boldsymbol{f}^i\left(\boldsymbol{u}\right) - \mathcal{F}^i\left(\boldsymbol{v}\left(\boldsymbol{u}\right)\right). \tag{3.5}$$

Equivalently, the entropy functionals $U\left(\boldsymbol{u}\right)$ and $F^i\left(\boldsymbol{u}\right)$ can be *explicitly* defined as follows

$$U\left(\boldsymbol{u}\right) = -\rho\left(\frac{s - s_\infty}{R}\right) = -\frac{\rho}{\left(\gamma - 1\right)}\ln\left(\frac{p/\rho^\gamma}{p_\infty/\rho_\infty^\gamma}\right),$$

$$F^i\left(\boldsymbol{u}\right) = -\rho\mathbf{V}^i\left(\frac{s - s_\infty}{R}\right) = -\frac{\rho\mathbf{V}^i}{\left(\gamma - 1\right)}\ln\left(\frac{p/\rho^\gamma}{p_\infty/\rho_\infty^\gamma}\right),$$

where $s = s\left(p, \rho\right)$ is the entropy, and s_∞, p_∞, and ρ_∞ are the reference entropy, pressure, and density at an arbitrary, predetermined state. In a similar fashion, explicit definitions for $\mathcal{U}\left(\boldsymbol{v}\right)$ and $\mathcal{F}^i\left(\boldsymbol{v}\right)$ can be obtained by substituting the previous explicit expressions for $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{f}^i$, $U\left(\boldsymbol{u}\right)$, and $F^i\left(\boldsymbol{u}\right)$ into Eq. (3.5).

One may set Eqs. (3.2)–(3.5) aside for the moment and return attention to Eq. (3.1). Upon rewriting the solution variables in Eq. (3.1) in terms of entropy variables [i.e. upon defining $\boldsymbol{u} = \boldsymbol{u}\left(\boldsymbol{v}\right)$ in this equation], one obtains the following

$$\boldsymbol{u}_{,\boldsymbol{v}}\boldsymbol{v}_{,t} + \boldsymbol{f}^i_{,\boldsymbol{u}}\boldsymbol{u}_{,\boldsymbol{v}}\boldsymbol{v}_{,\boldsymbol{x}_i} = 0. \tag{3.6}$$

Next, Eq. (3.6) can be rewritten in terms of several Jacobian matrix definitions due to Barth [5]. The matrix definitions are as follows

$$\widetilde{A}_0 = \boldsymbol{u}_{,\boldsymbol{v}}, \quad \widetilde{A}_0^{-1} = \boldsymbol{v}_{,\boldsymbol{u}}, \quad A_i = \boldsymbol{f}^i_{,\boldsymbol{u}}, \quad \widetilde{A}_i = A_i \widetilde{A}_0, \tag{3.7}$$

where it can be shown that $\widetilde{A}_0 \in \mathbb{R}^{m\times m}$ and $\widetilde{A}_0^{-1} \in \mathbb{R}^{m\times m}$ are symmetric positive-definite (SPD) matrices under mild assumptions (cf. Lemma B.1), and each $\widetilde{A}_i \in \mathbb{R}^{m\times m}$ matrix is symmetric [5]. On substituting two of the four definitions from Eq. (3.7) into Eq. (3.6), one obtains

$$\widetilde{A}_0\, \boldsymbol{v}_{,t} + \widetilde{A}_i\, \boldsymbol{v}_{,\boldsymbol{x}_i} = 0, \tag{3.8}$$

or equivalently,

$$\boldsymbol{v}_{,t} + \widetilde{A}_0^{-1} \widetilde{A}_i\, \boldsymbol{v}_{,\boldsymbol{x}_i} = 0. \tag{3.9}$$

These equations are convenient for certain types of analysis (cf. [5]). However, for the purpose of constructing a discretization (as will be done in the next section), it is easier to work with Eq. (3.8). Furthermore, in preparation for the next section, it is convenient to initially omit the Jacobian definitions from Eq. (3.8) and to simply rewrite Eq. (3.8) as follows

$$\boldsymbol{u}_{,t}\,(\boldsymbol{v}) + \boldsymbol{f}^i_{,\boldsymbol{x}_i}\,(\boldsymbol{u}\,(\boldsymbol{v})) = 0. \tag{3.10}$$

4 Discontinuous Galerkin methods

In this section, the standard DG approach is presented on simplex elements. As a preliminary step, the notation associated with a mesh of simplex elements is clearly defined. Thereafter, the DG method is presented in two distinct forms: first it is expressed in terms of entropy variables as the unknowns, and second it is expressed in terms of conservative variables as the unknowns.

4.1 Preliminaries

The implementation of the DG approach requires division of the domain Ω into elements T_k. For simplicity, it is assumed that each element T_k is a d-dimensional simplex, and therefore, the collection of all elements is referred to as the triangulation (denoted by $\mathcal{T}_h$, where h is some representative element size). Furthermore, it is assumed that the mesh is conforming, i.e. the mesh does not contain hanging nodes. By definition, one requires that the simplexes that make up the mesh have boundaries ∂T_k composed of $(d-1)$-dimensional simplicial faces denoted by $\mathcal{F}_l$ where $1 \leq l \leq (d+1)$. In turn, each face is associated with a normal vector $\boldsymbol{n} \in \mathbb{R}^d$. One further requires that the elements are non-overlapping and that the domain Ω is polygonal such that

$$\bigcup_{T_k \in \mathcal{T}_h} \overline{T_k} = \Omega, \qquad \overline{T_k} = T_k \cup \partial T_k.$$

For the sake of clarity, one may make a distinction between faces of the triangulation (F_ℓ) and faces of the individual elements that belong to the triangulation ($\mathcal{F}_l$). The latter faces were already discussed. The former faces can be classified as follows: the union of all internal faces of the triangulation (i.e. the union of faces F_ℓ of the triangulation that do not coincide with $\partial\Omega$) can be denoted by $\mathcal{E}_h^0$, and the union of all boundary faces of the triangulation (i.e. the union of faces that coincide with $\partial\Omega$) can be denoted by $\mathcal{E}_h^\partial$. The total collection of triangulation faces is denoted by $\mathcal{E}_h$, and can be simply defined as $\mathcal{E}_h = \mathcal{E}_h^0 \cup \mathcal{E}_h^\partial$. For each face in the triangulation, one may associate a normal vector $\widehat{\boldsymbol{n}} \in \mathbb{R}^d$. The vector $\widehat{\boldsymbol{n}}$ coincides in magnitude and angle with $\boldsymbol{n}$ on the faces of the individual elements such that: $\boldsymbol{n}_- = \widehat{\boldsymbol{n}}$ and $\boldsymbol{n}_+ = -\widehat{\boldsymbol{n}}$.

After dividing the domain Ω into elements T_k in the fashion described above, one may introduce the standard DG polynomial space as follows,

$$\boldsymbol{\mathcal{W}}^h = \left\{ \boldsymbol{w}^h : \boldsymbol{w}^h \in \boldsymbol{L}_2(\Omega), \boldsymbol{w}^h|_{T_k} \in \boldsymbol{\mathcal{P}}^p(T_k), \forall T_k \in \mathcal{T}_h \right\},$$

where $\boldsymbol{L}_2 = (L_2(\Omega))^m$ is a vector-valued L_2 function space, and $\boldsymbol{\mathcal{P}}^p(T_k) = (\mathcal{P}^p(T_k))^m$ is a vector-valued space of polynomials of order $\leq p$ on T_k.

4.2 The DG method with entropy variables

The DG method for Eq. (3.10) can be written in terms of entropy variables $\boldsymbol{v}^h$ as follows

$$\begin{aligned}\sum_{T_k \in \mathcal{T}_h} \int_{T_k} \left[\left(\boldsymbol{w}^h\right)^T \boldsymbol{u}_{,t}\left(\boldsymbol{v}^h\right) - \left(\boldsymbol{w}^h_{,\boldsymbol{x}_i}\right)^T \boldsymbol{f}^i\left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)\right)\right] dx \\ + \sum_{T_k \in \mathcal{T}_h} \int_{\partial T_k} \left(\boldsymbol{w}^h\right)^T \boldsymbol{f}^\star\left(\boldsymbol{v}^h_-, \boldsymbol{v}^h_+; \boldsymbol{n}\right) d\hat{x} = 0, \qquad (4.1)\end{aligned}$$

where $\boldsymbol{w}^h \in \boldsymbol{\mathcal{W}}^h$, $\boldsymbol{v}^h \in \boldsymbol{\mathcal{W}}^h$, and where the numerical flux $\boldsymbol{f}^\star\left(\boldsymbol{v}^h_-, \boldsymbol{v}^h_+; \boldsymbol{n}\right) : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^m$ takes the following form

$$\boldsymbol{f}^\star\left(\boldsymbol{v}^h_-, \boldsymbol{v}^h_+; \boldsymbol{n}\right) = \frac{1}{2}\left(\boldsymbol{f}^i\left(\boldsymbol{u}\left(\boldsymbol{v}^h_-\right)\right) + \boldsymbol{f}^i\left(\boldsymbol{u}\left(\boldsymbol{v}^h_+\right)\right)\right)\boldsymbol{n}^i + \frac{1}{2}\boldsymbol{h}^f\left(\boldsymbol{v}^h_-, \boldsymbol{v}^h_+; \boldsymbol{n}\right). \qquad (4.2)$$

The numerical flux is required to satisfy $\boldsymbol{f}^\star\left(\boldsymbol{v}^h, \boldsymbol{v}^h; \boldsymbol{n}\right) = \boldsymbol{f}^i\left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)\right)\boldsymbol{n}^i$. In addition, note that the function $\boldsymbol{h}^f\left(\boldsymbol{v}^h_-, \boldsymbol{v}^h_+; \boldsymbol{n}\right)$ in Eq. (4.2) provides necessary dissipation. A precise definition for $\boldsymbol{h}^f\left(\boldsymbol{v}^h_-, \boldsymbol{v}^h_+; \boldsymbol{n}\right)$ will be discussed later on in this work.

4.3 The DG methods with conservative variables

The DG method for Eq. (3.10) can be written in terms of conservative variables $\boldsymbol{u}^h$ as follows

$$\sum_{T_k \in \mathcal{T}_h} \int_{T_k} \left[\left(\boldsymbol{w}^h\right)^T \boldsymbol{u}^h_{,t} - \left(\boldsymbol{w}^h_{,\boldsymbol{x}_i}\right)^T \boldsymbol{f}^i\left(\boldsymbol{u}^h\right) \right] dx + \sum_{T_k \in \mathcal{T}_h} \int_{\partial T_k} \left(\boldsymbol{w}^h\right)^T \boldsymbol{f}^{\star}\left(\boldsymbol{v}\left(\boldsymbol{u}^h_-\right), \boldsymbol{v}\left(\boldsymbol{u}^h_+\right); \boldsymbol{n}\right) d\hat{x} = 0, \tag{4.3}$$

where $\boldsymbol{u}^h \in \mathcal{W}^h$. It is important to note that Eqs. (4.1) and (4.3) describe fundamentally different methods, as in general $\boldsymbol{u}^h \neq \boldsymbol{u}\left(\boldsymbol{v}^h\right)$ and $\boldsymbol{v}^h \neq \boldsymbol{v}\left(\boldsymbol{u}^h\right)$.

One should note that Eq. (4.3) is not the standard way to write the DG method with conservative variables as the unknowns. There is another more direct way to write the DG method in terms of $\boldsymbol{u}^h$ as follows

$$\sum_{T_k \in \mathcal{T}_h} \int_{T_k} \left[\left(\boldsymbol{w}^h\right)^T \boldsymbol{u}^h_{,t} - \left(\boldsymbol{w}^h_{,\boldsymbol{x}_i}\right)^T \boldsymbol{f}^i\left(\boldsymbol{u}^h\right) \right] dx + \sum_{T_k \in \mathcal{T}_h} \int_{\partial T_k} \left(\boldsymbol{w}^h\right)^T \boldsymbol{f}^{\diamond}\left(\boldsymbol{u}^h_-, \boldsymbol{u}^h_+; \boldsymbol{n}\right) d\hat{x} = 0, \tag{4.4}$$

where the numerical flux $\boldsymbol{f}^{\diamond}\left(\boldsymbol{u}^h_-, \boldsymbol{u}^h_+; \boldsymbol{n}\right) : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^m$ takes the following form

$$\boldsymbol{f}^{\diamond}\left(\boldsymbol{u}^h_-, \boldsymbol{u}^h_+; \boldsymbol{n}\right) = \frac{1}{2}\left(\boldsymbol{f}^i\left(\boldsymbol{u}^h_-\right) + \boldsymbol{f}^i\left(\boldsymbol{u}^h_+\right)\right)\boldsymbol{n}^i + \frac{1}{2}\boldsymbol{h}^f\left(\boldsymbol{u}^h_-, \boldsymbol{u}^h_+; \boldsymbol{n}\right), \tag{4.5}$$

and one requires that $\boldsymbol{f}^{\diamond}\left(\boldsymbol{u}^h, \boldsymbol{u}^h; \boldsymbol{n}\right) = \boldsymbol{f}^i\left(\boldsymbol{u}^h\right)\boldsymbol{n}^i$. Although Eqs. (4.3) and (4.4) are both written in terms of conservative variables $\boldsymbol{u}^h$, the numerical fluxes that they employ are generally fundamentally different, as roughly speaking, the numerical flux in Eq. (4.3) adds dissipation that is proportional to interfacial jumps in $\boldsymbol{v}(\boldsymbol{u}^h)$ and the numerical flux in Eq. (4.4) adds dissipation that is proportional to interfacial jumps in $\boldsymbol{u}^h$. This point will become more clear when the numerical fluxes are precisely defined later in this work.

5 Entropy stability for the DG method with entropy variables

It is well known that Eq. (4.1) is an entropy stable formulation for solving the compressible Euler equations. In order to demonstrate this, one must substitute $\boldsymbol{v}^h$ in place of $\boldsymbol{w}^h$ in Eq. (4.1) and follow the procedures described in [5,69]. For the sake of completeness, the main result of these procedures is summarized in the following theorem. The theorem is prefaced with several necessary definitions in order to facilitate its presentation.

Definition 5.1 The spatial jump operator $[\![\cdot]\!]^-_+$,

$$\left[\!\left[\boldsymbol{v}^h\right]\!\right]^-_+ \equiv \boldsymbol{v}^h_- - \boldsymbol{v}^h_+. \tag{5.1}$$

Definition 5.2 The spatial average operator $\{\{\cdot\}\}_{+}^{-}$,

$$\left\{\!\!\left\{ \boldsymbol{v}^h \right\}\!\!\right\}_{+}^{-} \equiv \frac{1}{2}\left(\boldsymbol{v}_{-}^{h} + \boldsymbol{v}_{+}^{h}\right). \tag{5.2}$$

Note that the average operator does not explicitly appear in the statement of the following theorem, however it is closely related to the jump operator, and it is utilized in subsequent sections.

Definition 5.3 The Mean-Value (or Symmetric Mean-Value) flux $\boldsymbol{h}_{MV}^{f}\left(\boldsymbol{v}_{-}^{h}, \boldsymbol{v}_{+}^{h}; \widehat{\boldsymbol{n}}\right) : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^m$, [5] p. 216 and [4] p. 10

$$\begin{aligned} &\boldsymbol{h}_{MV}^{f}\left(\boldsymbol{v}_{-}^{h}, \boldsymbol{v}_{+}^{h}; \widehat{\boldsymbol{n}}\right) \\ &\quad \equiv \int_0^1 (1-\theta)\left(\left|\widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}^h(\theta)\right)\widehat{\boldsymbol{n}}^i\right|_{\widetilde{A}_0} + \left|\widetilde{A}_i\left(\overline{\boldsymbol{v}}^h(\theta)\right)\widehat{\boldsymbol{n}}^i\right|_{\widetilde{A}_0}\right)\left[\!\left[\boldsymbol{v}^h\right]\!\right]_{+}^{-} d\theta, \end{aligned} \tag{5.3}$$

where

$$\overline{\boldsymbol{v}}^h(\theta) = \boldsymbol{v}_{+}^{h} - \theta\left[\!\left[\boldsymbol{v}^h\right]\!\right]_{-}^{+}, \qquad \overline{\overline{\boldsymbol{v}}}^h(\theta) = \boldsymbol{v}_{-}^{h} + \theta\left[\!\left[\boldsymbol{v}^h\right]\!\right]_{-}^{+},$$

$$\left|\widetilde{A}_i(\cdot)\widehat{\boldsymbol{n}}^i\right|_{\widetilde{A}_0} = \left|A_i(\cdot)\widehat{\boldsymbol{n}}^i\right|\widetilde{A}_0(\cdot).$$

In the last expression, the matrix absolute value of $\widetilde{A}_i(\cdot)\widehat{\boldsymbol{n}}^i$ is symmetric positive semi-definite (SPSD).

Definition 5.4 The non-negative function $|||\cdot|||_{|\underline{\widetilde{A}}(\overline{\boldsymbol{v}})|, F_\ell} : \mathbb{R}^m \rightarrow \mathbb{R}$, [5] p. 223 and [4] pp. 10–12

$$\begin{aligned} |||\cdot|||^2_{|\underline{\widetilde{A}}(\overline{\boldsymbol{v}})|, F_\ell} \equiv \int_{F_\ell}\int_0^1 (1-\theta)(\cdot)^T &\left(\left[\widetilde{A}_i^{+}\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right]_{\widetilde{A}_0}\right. \\ &\left. - \left[\widetilde{A}_i^{-}\left(\overline{\boldsymbol{v}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right]_{\widetilde{A}_0}\right)(\cdot)\, d\theta\, d\hat{x}, \end{aligned} \tag{5.4}$$

where

$$\left[\widetilde{A}_i^{+}\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right]_{\widetilde{A}_0} = \left(A_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right)^{+}\widetilde{A}_0\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right),$$

$$\left[\widetilde{A}_i^{-}\left(\overline{\boldsymbol{v}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right]_{\widetilde{A}_0} = \left(A_i\left(\overline{\boldsymbol{v}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right)^{-}\widetilde{A}_0\left(\overline{\boldsymbol{v}}(\theta)\right),$$

and where the matrices with + and − superscripts are defined such that

$$\left|A_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right| = \left(A_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right)^{+} - \left(A_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right)^{-},$$

$$\left|A_i\left(\overline{\boldsymbol{v}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right| = \left(A_i\left(\overline{\boldsymbol{v}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right)^{+} - \left(A_i\left(\overline{\boldsymbol{v}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right)^{-}.$$

The principal result of this section can now be summarized in the following theorem.

Theorem 5.5 *Consider the DG scheme in Eq.* (4.1) *with* $p \geq 0$ *and the term* $\boldsymbol{h}^f\left(\boldsymbol{v}_-^h, \boldsymbol{v}_+; \widehat{\boldsymbol{n}}\right)$ *chosen to be the Mean-Value flux* [Eq. (5.3)]. *The entropy stability of this scheme is governed by the following equation*

$$\sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} U_{,t}\left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)\right) dx \right] + \sum_{F_\ell \in \mathcal{E}_h^0} \left[\left\| \left[\!\left[\boldsymbol{v}^h \right]\!\right]_+^- \right\|^2_{\left|\widetilde{\underline{A}}(\overline{\boldsymbol{v}})\right|, F_\ell} \right]$$
$$= \Lambda_{bc,inv}\left(\boldsymbol{v}^h, \boldsymbol{v}^h\right) - \sum_{F_\ell \in \mathcal{E}_h^\partial} \left[\int_{F_\ell} F^i\left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)\right) \widehat{\boldsymbol{n}}^i \, d\hat{x} \right], \tag{5.5}$$

where the term $\Lambda_{bc,inv}\left(\boldsymbol{v}^h, \boldsymbol{v}^h\right)$ [*which is defined in Eq.* (6.2) *and by* [69]] *quantifies the effects of boundary conditions.*

Proof The proof of Theorem (5.5) follows directly from the work in [5] and [69]. In particular, the same result is obtained in [5] with the caveat that [5] integrates each term over space and time, whereas here, only integration in space has been assumed. This result also appears in [69], with the caveat that a hybrid variable $\widehat{\boldsymbol{v}}^h$ is introduced to provide additional stabilization, and integration is once again performed in both space and time instead of just in space. Hence, the result presented here requires only minor modifications to previous work, and does not require an additional detailed proof. □

Remark 5.6 It immediately follows from Eq. (5.5), that the DG scheme in Eq. (4.1) is entropy stable for the compressible Euler equations when the boundary conditions are chosen such that the terms

$$\Lambda_{bc,inv}\left(\boldsymbol{v}^h, \boldsymbol{v}^h\right) - \sum_{F_\ell \in \mathcal{E}_h^\partial} \left[\int_{F_\ell} F^i\left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)\right) \widehat{\boldsymbol{n}}^i \, d\hat{x} \right],$$

vanish or are non-positive, $p \geq 0$, and $\boldsymbol{h}^f\left(\boldsymbol{v}_-^h, \boldsymbol{v}_+; \widehat{\boldsymbol{n}}\right)$ is chosen to be the Mean-Value flux [Eq. (5.3)]. For example, this holds true when the boundary conditions are periodic, the solution has compact support, or the boundary conditions are chosen in conjunction with the approaches of [26,63]. Under these circumstances, the time-rate of change of the solution is governed by the following equation

$$\sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} U_{,t}\left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)\right) dx \right] \leq 0, \tag{5.6}$$

or equivalently

$$\sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} U\left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)(t)\right) dx \right] \leq \sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} U\left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)(t_0)\right) dx \right], \quad \forall t \geq t_0.$$

6 Entropy stability for the DG methods with conservative variables

In accordance with the techniques outlined in the previous section, one can obtain results governing the stability of the DG methods in Eqs. (4.3) and (4.4). In preparation for these results, one begins by reformulating Eq. (4.3) by performing integration by parts and replacing summations over individual element faces with summations over faces in the mesh as follows

$$\Lambda_{sol}(\boldsymbol{w}^h, \boldsymbol{u}^h) + \Lambda_{inv}(\boldsymbol{w}^h, \boldsymbol{u}^h) - \Lambda_{bc,inv}(\boldsymbol{w}^h, \boldsymbol{u}^h) = 0, \tag{6.1}$$

where

$$\begin{aligned}
\Lambda_{sol}\left(\boldsymbol{w}^h, \boldsymbol{u}^h\right) &= \sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} \left(\boldsymbol{w}^h\right)^T \boldsymbol{u}^h_{,t}\, dx \right], \\
\Lambda_{inv}\left(\boldsymbol{w}^h, \boldsymbol{u}^h\right) &= \sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} \left(\boldsymbol{w}^h\right)^T \boldsymbol{f}^i_{,x_i}\left(\boldsymbol{u}^h\right) dx \right] \\
&+ \sum_{F_\ell \in \mathcal{E}_h^0} \left[\int_{F_\ell} \left(\left[\!\left[\boldsymbol{w}^h \right]\!\right]_+^- \right)^T \boldsymbol{f}^\star\left(\boldsymbol{v}\left(\boldsymbol{u}^h_-\right), \boldsymbol{v}\left(\boldsymbol{u}^h_+\right); \widehat{\boldsymbol{n}}\right) d\hat{x} \right] \\
&+ \sum_{F_\ell \in \mathcal{E}_h^0} \left[\int_{F_\ell} \left(-\left(\boldsymbol{w}^h_-\right)^T \boldsymbol{f}^i\left(\boldsymbol{u}^h_-\right)\widehat{\boldsymbol{n}}^i + \left(\boldsymbol{w}^h_+\right)^T \boldsymbol{f}^i\left(\boldsymbol{u}^h_+\right)\widehat{\boldsymbol{n}}^i \right) d\hat{x} \right], \\
\Lambda_{bc,inv}\left(\boldsymbol{w}^h, \boldsymbol{u}^h\right) &= -\sum_{F_\ell \in \mathcal{E}_h^\partial} \left[\int_{F_\ell} \left(\boldsymbol{w}^h\right)^T \left(-\boldsymbol{f}^i\left(\boldsymbol{u}^h\right)\widehat{\boldsymbol{n}}^i + \boldsymbol{f}^\star\left(\boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{v}\left(\boldsymbol{u}^\partial\right); \widehat{\boldsymbol{n}}\right)\right) d\hat{x} \right],
\end{aligned} \tag{6.2}$$

and where it has been assumed that a Dirichlet boundary condition $\boldsymbol{u}^\partial$ has been specified. It is also convenient for the purpose of analysis to introduce the following quantity

$$\Lambda_{proj}\left(\,\cdot\,, \boldsymbol{u}^h\right) = -\Lambda_{inv}\left(\,\cdot\,, \boldsymbol{u}^h\right) + \Lambda_{bc,inv}\left(\,\cdot\,, \boldsymbol{u}^h\right).$$

Finally, one may introduce the following definition.

Definition 6.1 The Mean-Value flux $\boldsymbol{h}^f_{MV}\left(\boldsymbol{v}\left(\boldsymbol{u}^h_-\right), \boldsymbol{v}\left(\boldsymbol{u}^h_+\right); \widehat{\boldsymbol{n}}\right) : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^m$, [5] p. 216 and [4] p. 10

$$\begin{aligned}
&\boldsymbol{h}^f_{MV}\left(\boldsymbol{v}\left(\boldsymbol{u}^h_-\right), \boldsymbol{v}\left(\boldsymbol{u}^h_+\right); \widehat{\boldsymbol{n}}\right) \\
&\quad\equiv \int_0^1 (1-\theta)\left(\left| \widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)\widehat{\boldsymbol{n}}^i \right|_{\widetilde{A}_0} + \left| \widetilde{A}_i\left(\overline{\boldsymbol{v}}(\theta)\right)\widehat{\boldsymbol{n}}^i \right|_{\widetilde{A}_0} \right) \left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right) \right]\!\right]_+^- d\theta,
\end{aligned} \tag{6.3}$$

where

$$\overline{\boldsymbol{v}}(\theta) = \boldsymbol{v}\left(\boldsymbol{u}^h_+\right) - \theta \left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right) \right]\!\right]_-^+, \qquad \overline{\overline{\boldsymbol{v}}}(\theta) = \boldsymbol{v}\left(\boldsymbol{u}^h_-\right) + \theta \left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right) \right]\!\right]_-^+.$$

The first principal result of this section can now be summarized in the following theorem.

Theorem 6.2 *Consider the DG scheme in Eq.* (4.3) *with* $p \geq 0$ *and the term* $\boldsymbol{h}^f\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right), \boldsymbol{v}\left(\boldsymbol{u}_+^h\right); \widehat{\boldsymbol{n}}\right)$ *chosen to be the Mean-Value flux* [*Eq.* (6.3)]. *The entropy stability of this scheme is governed by the following equation*

$$\sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} U_{,t}\left(\boldsymbol{u}^h\right) dx \right] + \sum_{F_\ell \in \mathcal{E}_h^0} \left[\left\| \left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right) \right]\!\right]_+^- \right\|^2_{|\widetilde{\underline{A}}(\overline{\boldsymbol{v}})|, F_\ell} \right]$$
$$= \Lambda_{proj}\left(\boldsymbol{\varepsilon}_\Pi, \boldsymbol{u}^h\right) + \Lambda_{bc,inv}\left(\boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{u}^h\right) - \sum_{F_\ell \in \mathcal{E}_h^\partial} \left[\int_{F_\ell} F^i\left(\boldsymbol{u}^h\right) \widehat{\boldsymbol{n}}^i d\hat{x} \right], \tag{6.4}$$

where the 'entropy projection error' is defined such that

$$\boldsymbol{\varepsilon}_\Pi = \Pi^h \boldsymbol{v}\left(\boldsymbol{u}^h\right) - \boldsymbol{v}\left(\boldsymbol{u}^h\right), \tag{6.5}$$

and Π^h *is a* L_2 *projection operator from* $\boldsymbol{L}_2$ *onto the space* $\boldsymbol{\mathcal{W}}^h$.

Proof One may begin by replacing $\boldsymbol{w}^h$ in Eq. (6.1) with the test function $\boldsymbol{v}\left(\boldsymbol{u}^h\right) + \boldsymbol{\varepsilon}_\Pi$ in order to obtain

$$\Lambda_{sol}\left(\boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{u}^h\right) + \Lambda_{inv}\left(\boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{u}^h\right) - \Lambda_{bc,inv}\left(\boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{u}^h\right)$$
$$+ \Lambda_{sol}\left(\boldsymbol{\varepsilon}_\Pi, \boldsymbol{u}^h\right) - \Lambda_{proj}\left(\boldsymbol{\varepsilon}_\Pi, \boldsymbol{u}^h\right) = 0. \tag{6.6}$$

The quantity $\boldsymbol{\varepsilon}_\Pi$ is effectively a measure of the error between exactly evaluating the entropy function $\boldsymbol{v}\left(\boldsymbol{u}^h\right)$ and evaluating its L_2 projection. Of course, by construction, the sum of $\boldsymbol{\varepsilon}_\Pi$ and $\boldsymbol{v}\left(\boldsymbol{u}^h\right)$ lies in $\boldsymbol{\mathcal{W}}^h$, and is thus suitable for replacing $\boldsymbol{w}^h$.

In what follows, several of the terms in Eq. (6.6) will be analyzed in more detail. The solution term on the first line of Eq. (6.6) can be expanded as follows

$$\Lambda_{sol}\left(\boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{u}^h\right) = \sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} \boldsymbol{v}\left(\boldsymbol{u}^h\right)^T \boldsymbol{u}^h_{,t} \, dx \right]. \tag{6.7}$$

In accordance with Eq. (3.4), one observes that $\boldsymbol{v}\left(\boldsymbol{u}^h\right)^T \boldsymbol{u}^h_{,t} = U\left(\boldsymbol{u}^h\right)_{,\boldsymbol{u}} \boldsymbol{u}^h_{,t} = U_{,t}\left(\boldsymbol{u}^h\right)$. Therefore, upon substituting this identity into Eq. (6.7), one obtains

$$\Lambda_{sol}\left(\boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{u}^h\right) = \sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} U_{,t}\left(\boldsymbol{u}^h\right) dx \right]. \tag{6.8}$$

In a similar fashion, the solution term on the second line of Eq. (6.6) can be expanded as follows

$$\Lambda_{sol}\left(\boldsymbol{\varepsilon}_{\Pi},\boldsymbol{u}^{h}\right)=\sum_{T_k\in\mathcal{T}_h}\left[\int_{T_k}\left(\Pi^{h}\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)-\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)\right)^{T}\boldsymbol{u}^{h}_{,t}\,dx\right]=0. \tag{6.9}$$

This term vanishes because the L_2 projection error is orthogonal to all polynomials of degree $\leq p$.

Next, the inviscid flux term in Eq. (6.6) can be expanded as follows

$$\begin{aligned}\Lambda_{inv}\left(\boldsymbol{v}\left(\boldsymbol{u}^{h}\right),\boldsymbol{u}^{h}\right)&=\sum_{T_k\in\mathcal{T}_h}\left[\int_{T_k}\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)^{T}\boldsymbol{f}^{i}_{,\boldsymbol{x}_i}\left(\boldsymbol{u}^{h}\right)dx\right]\\&+\sum_{F_\ell\in\mathcal{E}_h^0}\left[\int_{F_\ell}\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{+}^{-}\right)^{T}\boldsymbol{f}^{\star}\left(\boldsymbol{v}\left(\boldsymbol{u}^{h}_{-}\right),\boldsymbol{v}\left(\boldsymbol{u}^{h}_{+}\right);\widehat{\boldsymbol{n}}\right)d\hat{x}\right]\\&+\sum_{F_\ell\in\mathcal{E}_h^0}\left[\int_{F_\ell}\left(-\boldsymbol{v}\left(\boldsymbol{u}^{h}_{-}\right)^{T}\boldsymbol{f}^{i}\left(\boldsymbol{u}^{h}_{-}\right)\widehat{\boldsymbol{n}}^{i}+\boldsymbol{v}\left(\boldsymbol{u}^{h}_{+}\right)^{T}\boldsymbol{f}^{i}\left(\boldsymbol{u}^{h}_{+}\right)\widehat{\boldsymbol{n}}^{i}\right)d\hat{x}\right].\end{aligned} \tag{6.10}$$

Consider rewriting the first term on the RHS of Eq. (6.10) as follows

$$\begin{aligned}&\sum_{T_k\in\mathcal{T}_h}\left[\int_{T_k}\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)^{T}\boldsymbol{f}^{i}_{,\boldsymbol{x}_i}\left(\boldsymbol{u}^{h}\right)dx\right]\\&\quad=\sum_{F_\ell\in\mathcal{E}_h^0}\left[\int_{F_\ell}\left[\!\left[F^{i}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{+}^{-}\widehat{\boldsymbol{n}}^{i}d\hat{x}\right]+\sum_{F_\ell\in\mathcal{E}_h^\partial}\left[\int_{F_\ell}F^{i}\left(\boldsymbol{u}^{h}\right)\widehat{\boldsymbol{n}}^{i}d\hat{x}\right],\end{aligned} \tag{6.11}$$

where the identity $\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)^{T}\boldsymbol{f}^{i}_{,\boldsymbol{x}_i}\left(\boldsymbol{u}^{h}\right)=F^{i}_{,\boldsymbol{x}_i}\left(\boldsymbol{u}^{h}\right)$ [which follows from Eqs. (3.3) and (3.5)] has been used, in conjunction with the divergence theorem.

Upon substituting Eq. (6.11) and the definition of the numerical flux [Eq. (4.2)] into Eq. (6.10), one obtains

$$\begin{aligned}&\Lambda_{inv}\left(\boldsymbol{v}\left(\boldsymbol{u}^{h}\right),\boldsymbol{u}^{h}\right)\\&\quad=\sum_{F_\ell\in\mathcal{E}_h^0}\left[\int_{F_\ell}\left[\!\left[F^{i}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{+}^{-}\widehat{\boldsymbol{n}}^{i}d\hat{x}\right]+\sum_{F_\ell\in\mathcal{E}_h^\partial}\left[\int_{F_\ell}F^{i}\left(\boldsymbol{u}^{h}\right)\widehat{\boldsymbol{n}}^{i}d\hat{x}\right]\\&\qquad+\sum_{F_\ell\in\mathcal{E}_h^0}\left[\int_{F_\ell}\left(\left\{\!\left\{\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)\right\}\!\right\}_{+}^{-}\right)^{T}\left[\!\left[\boldsymbol{f}^{i}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{-}^{+}\widehat{\boldsymbol{n}}^{i}\,d\hat{x}\right]\\&\qquad+\frac{1}{2}\sum_{F_\ell\in\mathcal{E}_h^0}\left[\int_{F_\ell}\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{+}^{-}\right)^{T}\boldsymbol{h}^{f}\left(\boldsymbol{v}\left(\boldsymbol{u}^{h}_{-}\right),\boldsymbol{v}\left(\boldsymbol{u}^{h}_{+}\right);\widehat{\boldsymbol{n}}\right)d\hat{x}\right].\end{aligned} \tag{6.12}$$

Eq. (6.12) can be rewritten by invoking Eq. (B.4) from Lemma B.2. The equation from the lemma is restated here for convenience.

$$\begin{aligned}&\left[\!\left[F^i\left(\boldsymbol{u}^h\right)\right]\!\right]_+^- + \left(\left\{\!\!\left\{\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right\}\!\!\right\}_+^-\right)^T \left[\!\left[\boldsymbol{f}^i\left(\boldsymbol{u}^h\right)\right]\!\right]_-^+ \\ &\quad = \frac{1}{2}\int_0^1 (1-\theta)\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_-^+\right)^T\left(\widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)-\widetilde{A}_i\left(\overline{\boldsymbol{v}}(\theta)\right)\right)\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_-^+ d\theta.\end{aligned} \tag{6.13}$$

On substituting Eq. (6.13) into Eq. (6.12), one obtains

$$\begin{aligned}\Lambda_{inv}\left(\boldsymbol{v}\left(\boldsymbol{u}^h\right),\boldsymbol{u}^h\right) &= \sum_{F_\ell\in\mathcal{E}_h^\partial}\left[\int_{F_\ell} F^i\left(\boldsymbol{u}^h\right)\widehat{\boldsymbol{n}}^i\, d\hat{x}\right] \\ &+\frac{1}{2}\sum_{F_\ell\in\mathcal{E}_h^0}\left[\int_{F_\ell}\int_0^1 (1-\theta)\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_-^+\right)^T\left(\widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)-\widetilde{A}_i\left(\overline{\boldsymbol{v}}(\theta)\right)\right)\widehat{\boldsymbol{n}}^i\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_-^+ d\theta\, d\hat{x}\right. \\ &+\left.\frac{1}{2}\int_{F_\ell}\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right)^T\boldsymbol{h}^f\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right),\boldsymbol{v}\left(\boldsymbol{u}_+^h\right);\widehat{\boldsymbol{n}}\right)d\hat{x}\right].\end{aligned} \tag{6.14}$$

In Eq. (6.14), the function $\boldsymbol{h}^f\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right),\boldsymbol{v}\left(\boldsymbol{u}_+^h\right);\widehat{\boldsymbol{n}}\right)$ can be replaced by the Mean-Value flux $\boldsymbol{h}_{MV}^f\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right),\boldsymbol{v}\left(\boldsymbol{u}_+^h\right);\widehat{\boldsymbol{n}}\right)$ (cf. Definition 6.1) as follows

$$\begin{aligned}\Lambda_{inv}\left(\boldsymbol{v}\left(\boldsymbol{u}^h\right),\boldsymbol{u}^h\right) &= \sum_{F_\ell\in\mathcal{E}_h^\partial}\left[\int_{F_\ell} F^i\left(\boldsymbol{u}^h\right)\widehat{\boldsymbol{n}}^i\, d\hat{x}\right] \\ &+\frac{1}{2}\sum_{F_\ell\in\mathcal{E}_h^0}\left[\int_{F_\ell}\int_0^1 (1-\theta)\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right)^T\left(\widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)-\widetilde{A}_i\left(\overline{\boldsymbol{v}}(\theta)\right)\right)\widehat{\boldsymbol{n}}^i\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^- d\theta\, d\hat{x}\right. \\ &+\frac{1}{2}\int_{F_\ell}\int_0^1 (1-\theta)\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right)^T\left(\left|\widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right|_{\widetilde{A}_0}\right. \\ &+\left.\left.\left|\widetilde{A}_i\left(\overline{\boldsymbol{v}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right|_{\widetilde{A}_0}\right)\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^- d\theta\, d\hat{x}\right].\end{aligned} \tag{6.15}$$

On manipulating Eq. (6.15), one obtains

$$\Lambda_{inv}\left(\boldsymbol{v}\left(\boldsymbol{u}^h\right),\boldsymbol{u}^h\right) = \sum_{F_\ell\in\mathcal{E}_h^\partial}\left[\int_{F_\ell} F^i\left(\boldsymbol{u}^h\right)\widehat{\boldsymbol{n}}^i\, d\hat{x}\right] + \sum_{F_\ell\in\mathcal{E}_h^0}\left[\left|\!\left|\!\left|\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right|\!\right|\!\right|^2_{|\underline{\widetilde{A}}(\overline{\boldsymbol{v}})|,F_\ell}\right]. \tag{6.16}$$

Here, the non-negative function from Definition 5.4 has been utilized. Upon substituting Eqs. (6.16), (6.8), and (6.9) into Eq. (6.6), one obtains Eq. (6.4). □

Remark 6.3 It immediately follows from Eq. (6.4), that the DG scheme in Eq. (4.3) is entropy stable for the compressible Euler equations when the boundary conditions are chosen appropriately (in line with Remark 5.6), $p \geq 0$, $\boldsymbol{h}^f \left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right), \boldsymbol{v}\left(\boldsymbol{u}_+^h\right); \widehat{\boldsymbol{n}}\right)$ is chosen to be the Mean-Value flux [Eq. (6.3)], and the projection error ($\boldsymbol{\varepsilon}_\Pi$) vanishes pointwise, or the projection error terms ($\Lambda_{proj}\left(\boldsymbol{\varepsilon}_\Pi, \boldsymbol{u}^h\right)$) are non-positive. Under these circumstances, the time-rate of change of the solution is governed by the following equation

$$\sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} U_{,t}\left(\boldsymbol{u}^h\right) dx\right] \leq 0, \tag{6.17}$$

or equivalently

$$\sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} U\left(\boldsymbol{u}^h(t)\right) dx\right] \leq \sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} U\left(\boldsymbol{u}^h(t_0)\right) dx\right], \quad \forall t \geq t_0.$$

It is worth discussing the conditions under which the projection error ($\boldsymbol{\varepsilon}_\Pi$) vanishes pointwise. Suppose that $\boldsymbol{v}\left(\boldsymbol{u}^h\right)$ is sufficiently smooth such that

$$\left\| D^{p+1}\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right\|_{L_\infty, T_k} < M_p, \quad \forall h \text{ and some fixed } p, \tag{6.18}$$

or

$$\left\| D^{p+1}\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right\|_{L_\infty, T_k} < M_h, \quad \forall p \text{ and some fixed } h, \tag{6.19}$$

where $0 < M_p < \infty$ and $0 < M_h < \infty$ are positive constants, $D^{p+1}: \mathbb{R} \rightarrow \mathbb{R}$ is a spatial derivative operator of order $p+1$, and

$$\left\| D^{p+1}\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right\|_{L_\infty, T_k} \equiv \max_{x \in T_k}\left[\max\left(D^{p+1}\boldsymbol{v}_1\left(\boldsymbol{u}^h\right), \ldots, D^{p+1}\boldsymbol{v}_m\left(\boldsymbol{u}^h\right)\right)\right].$$

Under these circumstances, the following vector-valued generalization of Ciarlet and Raviart's classic interpolation result (cf. [12], p. 20) holds

$$\left\|\boldsymbol{v}\left(\boldsymbol{u}^h\right) - \Pi^h \boldsymbol{v}\left(\boldsymbol{u}^h\right)\right\|_{L_\infty, T_k} \leq C_{T_k}\left\| D^{p+1}\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right\|_{L_\infty, T_k} h^{p+1},$$

or equivalently,

$$\|\boldsymbol{\varepsilon}_\Pi\|_{L_\infty, T_k} \leq C_{T_k}\left\| D^{p+1}\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right\|_{L_\infty, T_k} h^{p+1}, \tag{6.20}$$

where C_{T_k} is a constant that is element dependent, and

$$\begin{aligned}&\left\|\boldsymbol{v}\left(\boldsymbol{u}^h\right) - \Pi^h \boldsymbol{v}\left(\boldsymbol{u}^h\right)\right\|_{L_\infty, T_k} \\ &\quad \equiv \max_{x \in T_k}\left[\max\left(\boldsymbol{v}_1\left(\boldsymbol{u}^h\right) - \Pi^h \boldsymbol{v}_1\left(\boldsymbol{u}^h\right), \ldots, \boldsymbol{v}_m\left(\boldsymbol{u}^h\right) - \Pi^h \boldsymbol{v}_m\left(\boldsymbol{u}^h\right)\right)\right].\end{aligned}$$

By Eq. (6.20), it immediately follows that the projection error vanishes pointwise within each element when Eq. (6.18) holds and $h \to 0$, or Eq. (6.19) holds and $p \to \infty$. As a result, the DG scheme in Eq. (4.3) is 'asymptotically entropy stable', although it was necessary to require that the very strong assumptions in Eqs. (6.18) or (6.19) hold in order to obtain this result.

Now, from the author's perspective, this asymptotic result is somewhat weak, as there is no clear indication that the required assumptions will hold in practice. It appears that a better approach is to control the entropy projection errors by adding more dissipation, as suggested in Remark 6.4.

Remark 6.4 One may modify the DG scheme in Eq. (4.3) by adding the following 'stabilization term' in order to control the entropy projection error terms ($\Lambda_{proj}\left(\boldsymbol{\varepsilon}_{\Pi}, \boldsymbol{u}^h\right)$),

$$\Lambda_{stable}\left(\boldsymbol{w}^h, \boldsymbol{u}^h\right) = \sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} \mathcal{S}\left(\boldsymbol{w}^h, \boldsymbol{u}^h\right) dx\right],$$

where $\mathcal{S}(\boldsymbol{w}^h, \boldsymbol{u}^h)$ is a stabilization function that will be subsequently defined. In order to ensure that the stabilization term actually improves the scheme's robustness, one requires that the stabilization function $\mathcal{S}$ is dissipative (semi-coercive) when $\boldsymbol{w}^h = \Pi^h \boldsymbol{v}\left(\boldsymbol{u}^h\right)$, i.e. one requires $\mathcal{S}\left(\Pi^h \boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{u}^h\right) \geq 0$. In addition, one requires that the stabilization function is 'primally consistent', i.e. $\mathcal{S}\left(\boldsymbol{w}^h, \boldsymbol{u}\right) = 0$ for the exact solution $\boldsymbol{u}$, as well as 'dimensionally consistent', i.e. $\mathcal{S}\left(\Pi^h \boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{u}^h\right)$ has the same units as $U_{,t}\left(\boldsymbol{u}^h\right)$. It is difficult to construct a stabilization function that satisfies all of these requirements. However, it is theoretically possible if one chooses $\mathcal{S}\left(\boldsymbol{w}^h, \boldsymbol{u}^h\right)$ as follows

$$\mathcal{S}\left(\boldsymbol{w}^h, \boldsymbol{u}^h\right) = \left(\left(\boldsymbol{w}^h - \boldsymbol{v}\left(\boldsymbol{u}^h\right)\right)^T \boldsymbol{S}\mathcal{R}\left(\boldsymbol{u}^h\right)\right)\left(\left(\Pi^h \boldsymbol{v}\left(\boldsymbol{u}^h\right) - \boldsymbol{v}\left(\boldsymbol{u}^h\right)\right)^T \boldsymbol{S}\mathcal{R}\left(\boldsymbol{u}^h\right)\right), \tag{6.21}$$

or equivalently

$$\mathcal{S}\left(\boldsymbol{w}^h, \boldsymbol{u}^h\right) = \left(\left(\boldsymbol{w}^h - \boldsymbol{v}\left(\boldsymbol{u}^h\right)\right)^T \boldsymbol{S}\mathcal{R}\left(\boldsymbol{u}^h\right)\right)\left(\left(\boldsymbol{\varepsilon}_{\Pi}\right)^T \boldsymbol{S}\mathcal{R}\left(\boldsymbol{u}^h\right)\right), \tag{6.22}$$

where $\boldsymbol{S} \in \mathbb{R}^{m \times m}$ is an SPD stabilization matrix, and $\mathcal{R} : \mathbb{R}^m \to \mathbb{R}^m$ is the strong form of the residual operator

$$\mathcal{R}\left(\boldsymbol{u}^h\right) = \boldsymbol{u}^h_{,t} + \boldsymbol{f}^i_{,x_i}\left(\boldsymbol{u}^h\right),$$

where $\mathcal{R}(\boldsymbol{u}) = \mathbf{0}$ for the exact solution $\boldsymbol{u}$.

In order to preserve dimensional consistency of the stabilization function $\mathcal{S}$, the stabilization matrix $\boldsymbol{S}$ in Eq. (6.22) must have units of $\sqrt{t/\rho}$. A simple choice that meets this requirement is $\boldsymbol{S} = \sqrt{c}\,\boldsymbol{I}$, where

$$c = \frac{\eta \Delta t}{\rho_\infty},$$

where $\eta > 0$ is a user specified constant, and Δt is the characteristic time-step for a time marching scheme.

It remains to show that $\mathcal{S}\left(\Pi^h \boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{u}^h\right) \geq 0$. However, this immediately follows from substituting $\boldsymbol{w}^h = \Pi^h \boldsymbol{v}\left(\boldsymbol{u}^h\right)$ into Eq. (6.22) in order to obtain

$$\mathcal{S}\left(\Pi^h \boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{u}^h\right) = \left(\left(\boldsymbol{\varepsilon}_\Pi\right)^T \boldsymbol{S}\, \mathcal{R}\left(\boldsymbol{u}^h\right)\right)^2 \geq 0.$$

In conclusion, the proposed stabilization function [Eq. (6.22)] is of theoretical interest, as it will guarantee entropy stability for sufficiently large values of η. However, in this work we will not investigate this term further because, from a practical standpoint, it makes more sense to just utilize entropy variables, thereby exactly vanishing the entropy projection terms. The stabilization term is presented here only to inspire the development of similar terms, which may be necessary in instances where conservative variables cannot be avoided (for instance, a piece of software is hard-coded with conservative variables).

Remark 6.5 A modification of Eq. (6.4) holds for a more general class of numerical fluxes $\boldsymbol{h}^f$. In particular, for any numerical flux $\boldsymbol{h}^f$ that is more dissipative than the Mean-Value flux [Eq. (6.3)], i.e.

$$\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right)^T \boldsymbol{h}_{MV}^f \leq \left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right)^T \boldsymbol{h}^f \tag{6.23}$$

then the following modification of Eq. (6.4) holds

$$\begin{aligned} &\sum_{T_k \in \mathcal{T}_h} \left[\int_{T_k} U_{,t}\left(\boldsymbol{u}^h\right) dx\right] + C \sum_{F_\ell \in \mathcal{E}_h^0} \left[\left|\!\left|\!\left| \left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^- \right|\!\right|\!\right|^2_{|\underline{\widetilde{A}}(\overline{\boldsymbol{v}})|, F_\ell}\right] \\ &= \Lambda_{proj}\left(\boldsymbol{\varepsilon}_\Pi, \boldsymbol{u}^h\right) + \Lambda_{bc,inv}\left(\boldsymbol{v}\left(\boldsymbol{u}^h\right), \boldsymbol{u}^h\right) - \sum_{F_\ell \in \mathcal{E}_h^\partial} \left[\int_{F_\ell} F^i\left(\boldsymbol{u}^h\right) \widehat{\boldsymbol{n}}^i d\hat{x}\right], \end{aligned} \tag{6.24}$$

where $C \geq 1$. An extensive list of numerical fluxes that satisfy Eq. (6.23) is given in [4]. This list includes the Lax-Friedrichs flux (which is presented in, for example, [38]) and a modified form of the Harten-Lax-Van-Leer-Einfeldt (HLLE) flux (which is presented in, for example [66]). These fluxes are more dissipative than the Mean-Value flux, however, they are less expensive to compute. In particular, costly numerical quadrature procedures are required in order to compute the Mean-Value flux. The cost is controlled by the number of quadrature points that are required in order to accurately approximate the trajectory integral in Eq. (6.3), which is difficult to determine a priori. However, a general heuristic for finite element methods is that a quadrature rule of at least degree $3p$ is required in order to integrate inner products that arise in nonlinear convection problems [47]. Barth et al. [4] recommended using a Gauss-Lobatto quadrature rule for this purpose, which exactly integrates a polynomial of degree $2q-3$ with q points [47,54]. As a result, the computation of the Mean-Value flux will likely require at least $q = \lceil (3p+3)/2 \rceil$ quadrature points. Conversely, the cost of computing the Lax-Friedrichs or modified HLLE fluxes is close to the cost of integrating with a single quadrature point. Therefore, for large values of p, the Mean-Value flux will not be competitive from a cost standpoint with the simpler alternative fluxes.

It is useful to introduce the following definitions prior to presenting the second principal result of this section.

Definition 6.6 The 'Volpert matrix' $\widetilde{G} = \widetilde{G}\left(\boldsymbol{v}\left(\boldsymbol{u}_{-}^{h}\right), \boldsymbol{v}\left(\boldsymbol{u}_{+}^{h}\right)\right) : \mathbb{R}^{m} \times \mathbb{R}^{m} \rightarrow \mathbb{R}^{m\times m}$, [5] p. 225. The Volpert matrix is required to be SPD and satisfy the following equation

$$\left[\!\left[\boldsymbol{u}^{h}\right]\!\right]_{+}^{-} = \widetilde{G}\left(\boldsymbol{v}\left(\boldsymbol{u}_{-}^{h}\right), \boldsymbol{v}\left(\boldsymbol{u}_{+}^{h}\right)\right)\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{+}^{-}. \tag{6.25}$$

It turns out that $\widetilde{G}$ is guaranteed to exist due to Volpert trajectory integration theory (cf. [5]). In fact, a precise formulation for $\widetilde{G}$ can be obtained by invoking a particular Mean-Value Theorem for vector-valued functions (cf. [53] p. 278). This theorem yields the following identity

$$\begin{aligned}\left[\!\left[\boldsymbol{u}^{h}\right]\!\right]_{+}^{-} &= \left[\int_{0}^{1} \boldsymbol{u}_{,\boldsymbol{v}}\left(\overline{\boldsymbol{v}}\left(\theta\right)\right) d\theta\right]\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{+}^{-} \\ &= \left[\int_{0}^{1} \widetilde{A}_{0}\left(\overline{\boldsymbol{v}}\left(\theta\right)\right) d\theta\right]\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{+}^{-},\end{aligned} \tag{6.26}$$

where the last line follows from the identities in Eq. (3.7). Upon comparing Eq. (6.26) with Eq. (6.25), it is immediately obvious that $\widetilde{G}\left(\boldsymbol{v}\left(\boldsymbol{u}_{-}^{h}\right), \boldsymbol{v}\left(\boldsymbol{u}_{+}^{h}\right)\right)$ takes the following form

$$\widetilde{G}\left(\boldsymbol{v}\left(\boldsymbol{u}_{-}^{h}\right), \boldsymbol{v}\left(\boldsymbol{u}_{+}^{h}\right)\right) = \int_{0}^{1} \widetilde{A}_{0}\left(\overline{\boldsymbol{v}}\left(\theta\right)\right) d\theta. \tag{6.27}$$

Based on Eq. (6.27), it is clear that $\widetilde{G}\left(\boldsymbol{v}\left(\boldsymbol{u}_{-}^{h}\right), \boldsymbol{v}\left(\boldsymbol{u}_{+}^{h}\right)\right)$ is SPD because $\widetilde{A}_{0}$ is SPD (under the assumptions of Lemma B.1).

Definition 6.7 The 'Volpert flux' $\boldsymbol{h}_{VP}^{f}\left(\boldsymbol{u}_{-}^{h}, \boldsymbol{u}_{+}^{h}, \widehat{\boldsymbol{n}}\right) : \mathbb{R}^{m} \times \mathbb{R}^{m} \times \mathbb{R}^{d} \rightarrow \mathbb{R}^{m}$

$$\boldsymbol{h}_{VP}^{f}\left(\boldsymbol{u}_{-}^{h}, \boldsymbol{u}_{+}^{h}, \widehat{\boldsymbol{n}}\right) \equiv B\left(\boldsymbol{u}_{-}^{h}, \boldsymbol{u}_{+}^{h}, \widehat{\boldsymbol{n}}\right)\left[\!\left[\boldsymbol{u}^{h}\right]\!\right]_{+}^{-}, \tag{6.28}$$

where $B\left(\boldsymbol{u}_{-}^{h}, \boldsymbol{u}_{+}^{h}, \widehat{\boldsymbol{n}}\right) \in \mathbb{R}^{m \times m}$ is a matrix that is defined as follows

$$\begin{aligned} & B\left(\boldsymbol{u}_{-}^{h}, \boldsymbol{u}_{+}^{h}, \widehat{\boldsymbol{n}}\right) \\ & \quad \equiv \left[\int_{0}^{1}(1-\theta)\left(\left|\widetilde{A}_{i}\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right) \widehat{\boldsymbol{n}}^{i}\right|_{\widetilde{A}_{0}}+\left|\widetilde{A}_{i}\left(\overline{\boldsymbol{v}}(\theta)\right) \widehat{\boldsymbol{n}}^{i}\right|_{\widetilde{A}_{0}}\right) d \theta\right]\left(\widetilde{G}\left(\boldsymbol{v}\left(\boldsymbol{u}_{-}^{h}\right), \boldsymbol{v}\left(\boldsymbol{u}_{+}^{h}\right)\right)\right)^{-1}. \end{aligned}$$

The second principal result of this section can now be summarized in the following theorem.

Theorem 6.8 *Consider the DG scheme in Eq.* (4.4) *with* $p \geq 0$ *and the term* $\boldsymbol{h}^{f}\left(\boldsymbol{u}_{-}^{h}, \boldsymbol{u}_{+}^{h} ; \widehat{\boldsymbol{n}}\right)$ *chosen to be the Volpert flux* [*Eq.* (6.28)]. *The entropy stability of this scheme is governed by the following equation*

$$\begin{aligned} & \sum_{T_{k} \in \mathcal{T}_{h}}\left[\int_{T_{k}} U_{, t}\left(\boldsymbol{u}^{h}\right) d x\right]+\sum_{F_{\ell} \in \mathcal{E}_{h}^{0}}\left[\left|\left|\left|\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{+}^{-}\right|\right|\right|_{|\underline{\widetilde{A}}(\overline{\boldsymbol{v}})|, F_{\ell}}^{2}\right] \\ & \quad=\Lambda_{proj}\left(\boldsymbol{\varepsilon}_{\Pi}, \boldsymbol{u}^{h}\right)+\Lambda_{bc, inv}\left(\boldsymbol{v}\left(\boldsymbol{u}^{h}\right), \boldsymbol{u}^{h}\right)-\sum_{F_{\ell} \in \mathcal{E}_{h}^{\partial}}\left[\int_{F_{\ell}} F^{i}\left(\boldsymbol{u}^{h}\right) \widehat{\boldsymbol{n}}^{i} d \hat{x}\right]. \end{aligned} \tag{6.29}$$

Note that Eqs. (6.29) *and* (6.4) *are identical.*

Proof The proofs of Theorems 6.8 and 6.2 are virtually identical. The only apparent difference arises when the numerical flux is evaluated. The numerical flux term

$$\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{+}^{-}\right)^{T} \boldsymbol{h}^{f}$$

takes the following form in the proof of Theorem 6.2

$$\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{+}^{-}\right)^{T} \boldsymbol{h}_{MV}^{f}\left(\boldsymbol{v}\left(\boldsymbol{u}_{-}^{h}\right), \boldsymbol{v}\left(\boldsymbol{u}_{+}^{h}\right), \widehat{\boldsymbol{n}}\right), \tag{6.30}$$

and the following alternative form in the proof of Theorem 6.8

$$\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^{h}\right)\right]\!\right]_{+}^{-}\right)^{T} \boldsymbol{h}_{VP}^{f}\left(\boldsymbol{u}_{-}^{h}, \boldsymbol{u}_{+}^{h}, \widehat{\boldsymbol{n}}\right). \tag{6.31}$$

However, upon substituting Definitions 6.1, 6.6, and 6.7, into Eqs. (6.30) and (6.31), it is clear that the numerical flux terms are equivalent. In particular

$$\begin{aligned}
&\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right)^T \boldsymbol{h}_{VP}^f\left(\boldsymbol{u}_-^h, \boldsymbol{u}_+^h, \widehat{\boldsymbol{n}}\right)\\
&= \left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right)^T \left[\int_0^1 (1-\theta)\left(\left|\widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right|_{\widetilde{A}_0} + \left|\widetilde{A}_i\left(\overline{\boldsymbol{v}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right|_{\widetilde{A}_0}\right) d\theta\right]\\
&\quad\times\left(\widetilde{G}\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right), \boldsymbol{v}\left(\boldsymbol{u}_+^h\right)\right)\right)^{-1}\left[\!\left[\boldsymbol{u}^h\right]\!\right]_+^-\\
&= \left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right)^T \left[\int_0^1 (1-\theta)\left(\left|\widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right|_{\widetilde{A}_0} + \left|\widetilde{A}_i\left(\overline{\boldsymbol{v}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right|_{\widetilde{A}_0}\right) d\theta\right]\\
&\quad\times\left(\widetilde{G}\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right), \boldsymbol{v}\left(\boldsymbol{u}_+^h\right)\right)\right)^{-1}\widetilde{G}\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right), \boldsymbol{v}\left(\boldsymbol{u}_+^h\right)\right)\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right)\\
&= \left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right)^T \int_0^1 (1-\theta)\left(\left|\widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right|_{\widetilde{A}_0} + \left|\widetilde{A}_i\left(\overline{\boldsymbol{v}}(\theta)\right)\widehat{\boldsymbol{n}}^i\right|_{\widetilde{A}_0}\right)\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right) d\theta\\
&= \left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_+^-\right)^T \boldsymbol{h}_{MV}^f\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right), \boldsymbol{v}\left(\boldsymbol{u}_+^h\right), \widehat{\boldsymbol{n}}\right).
\end{aligned}$$

This completes the proof. □

Remark 6.9 Remarks 6.3–6.5 for Theorem 6.2 also apply to Theorem 6.8 with the Volpert flux [Eq. (6.28)] in place of the Mean-Value flux [Eq. (6.3)], and the scheme in Eq. (4.4) in place of the scheme in Eq. (4.3).

7 L_2 stability for the DG method with entropy variables

If the entropy stability condition [Eq. (5.6)] holds for the DG scheme in Eq. (4.1), it can be shown that certain conditions on the 'L_2 stability' of the scheme also hold. These conditions are summarized in the following theorems. The first theorem is prefaced with several necessary definitions in order to facilitate its presentation.

Definition 7.1 The L_2 norm $\|\cdot\|_{L_2,\mathcal{T}_h}^2$ on the entire domain

$$\|\cdot\|_{L_2,\mathcal{T}_h}^2 \equiv \sum_{T_k\in\mathcal{T}_h}\int_{T_k}(\cdot)^T(\cdot)\,dx. \tag{7.1}$$

Definition 7.2 The domain-averaged solution $\boldsymbol{u}^*$, [6,21,39]

$$\boldsymbol{u}^* \equiv \frac{1}{\text{meas}\,(\Omega)}\left[\sum_{T_k\in\mathcal{T}_h}\int_{T_k}\boldsymbol{u}\left(\boldsymbol{v}^h(t_0)\right)dx\right], \tag{7.2}$$

where t_0 is an arbitrary, but fixed initial time.

Definition 7.3 The function $H\left(\boldsymbol{u}, \boldsymbol{u}^*\right) : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ of the solution and the domain-averaged solution

$$H\left(\boldsymbol{u}, \boldsymbol{u}^*\right) \equiv U\left(\boldsymbol{u}\right) - U\left(\boldsymbol{u}^*\right) - \left(U_{,\boldsymbol{u}}\left(\boldsymbol{u}^*\right)\right)^T\left(\boldsymbol{u} - \boldsymbol{u}^*\right). \tag{7.3}$$

In Eq. (7.3) it is implicitly assumed that $\boldsymbol{u} = \boldsymbol{u}\left(\boldsymbol{v}^h\right)$.

The first result of this section can now be summarized in the following theorem.

Theorem 7.4 *Suppose the DG scheme in Eq.* (4.1) *is conservative and satisfies Eq.* (5.6)*. Furthermore, suppose that* $\boldsymbol{u}\left(t_0\right) \in \boldsymbol{L}_2\left(\Omega\right)$*, and in accordance with Lemma B.3,* $H\left(\boldsymbol{u}, \boldsymbol{u}^*\right)$ *is bounded at time* t_0 *in the following weak sense*

$$\sum_{T_k \in \mathcal{T}_h} \int_{T_k} H\left(\boldsymbol{u}\left(t_0\right), \boldsymbol{u}^*\right) dx \leq \mathfrak{C} \|\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\|^2_{L_2, \mathcal{T}_h}, \tag{7.4}$$

where $\mathfrak{C} > 0$ *is a constant that depends on the initial data. Then, in accordance with the work of Dafermos* [21]*, the scheme in Eq.* (4.1) *is* L_2 *stable for all* $t \geq t_0$ *in the following sense*

$$\|\boldsymbol{u}\left(t\right) - \boldsymbol{u}^*\|_{L_2, \mathcal{T}_h} \leq c \|\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\|_{L_2, \mathcal{T}_h}, \tag{7.5}$$

or more precisely

$$\|\boldsymbol{u}\left(\boldsymbol{v}^h\left(t\right)\right) - \boldsymbol{u}^*\left(\boldsymbol{v}^h\right)\|_{L_2, \mathcal{T}_h} \leq c \|\boldsymbol{u}\left(\boldsymbol{v}^h\left(t_0\right)\right) - \boldsymbol{u}^*\left(\boldsymbol{v}^h\right)\|_{L_2, \mathcal{T}_h}, \tag{7.6}$$

where $c \geq 1$ *is a constant that depends on* $\mathfrak{C}$.

Proof A somewhat complicated proof appears in [6,21,62]. However, the result in Theorem 7.4 can also be obtained with a simpler proof, a detailed formulation of which is presented in what follows.

In order to obtain the L_2 stability condition in question, one may begin by introducing the concept of a domain-averaged solution $\boldsymbol{u}^*$ [as was done in Eq. (7.2)]. Here, one should not confuse the asterisk superscript $*$ with the star superscript $\star$ that has been previously used to denote the (essentially) unrelated numerical fluxes. In addition, one should observe that $\boldsymbol{u}^*$ is constant in time, as it is assumed that the scheme (and the associated boundary conditions) are designed such that mass, momentum, and energy are conserved within the spatial domain, and therefore the total domain-averaged measure of these conservative quantities does not change in time. This assumption is valid for all of the DG methods proposed in our paper, provided the boundary conditions are enforced via numerical fluxes. Under these circumstances, the schemes are 'locally conservative' in accordance with the definition in, for example [13,19]. Therefore, $\boldsymbol{u}^*$ remains fixed for these schemes, and one may use it as a convenient reference quantity.

An L_2 stability condition that relates the reference quantity $\boldsymbol{u}^*$ and the time-dependent solution $\boldsymbol{u}$ may now be obtained in a relatively straightforward manner by analyzing the function $H\left(\boldsymbol{u}, \boldsymbol{u}^*\right)$. Towards this end, one may observe that the quantity $U_{,\boldsymbol{u}}$ in the definition of $H(\boldsymbol{u},\boldsymbol{u}^*)$ [Eq. (7.3)] is equivalent to $\boldsymbol{v}^T$ by Eq. (3.4). Upon utilizing this fact in Eq. (7.3), one obtains

$$H\left(\boldsymbol{u},\boldsymbol{u}^*\right)=U\left(\boldsymbol{u}\right)-U\left(\boldsymbol{u}^*\right)-\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)\right)^T\left(\boldsymbol{u}-\boldsymbol{u}^*\right). \tag{7.7}$$

In its current form, Eq. (7.7) is defined pointwise within every element of the domain. One may obtain a weak form of Eq. (7.7) by integrating over the entire domain as follows

$$\begin{aligned}\sum_{T_k\in\mathcal{T}_h}\int_{T_k}H\left(\boldsymbol{u},\boldsymbol{u}^*\right)dx&=\sum_{T_k\in\mathcal{T}_h}\left(\int_{T_k}\left(U\left(\boldsymbol{u}\right)-U\left(\boldsymbol{u}^*\right)\right)dx\right)\\&\quad-\sum_{T_k\in\mathcal{T}_h}\left(\int_{T_k}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)\right)^T\left(\boldsymbol{u}-\boldsymbol{u}^*\right)dx\right).\end{aligned} \tag{7.8}$$

The second term on the RHS of Eq. (7.8) vanishes by the definition of $\boldsymbol{u}^*$ in Eq. (7.2), and by the fact that $\boldsymbol{v}(\boldsymbol{u}^*)$ is independent of the spatial coordinates. With these simplifications, Eq. (7.8) takes the following form

$$\sum_{T_k\in\mathcal{T}_h}\int_{T_k}H\left(\boldsymbol{u},\boldsymbol{u}^*\right)dx=\sum_{T_k\in\mathcal{T}_h}\int_{T_k}\left(U\left(\boldsymbol{u}\right)-U\left(\boldsymbol{u}^*\right)\right)dx. \tag{7.9}$$

Upon carefully examining the RHS of Eq. (7.9), one realizes that $U(\boldsymbol{u}^*)$ is independent of time because $\boldsymbol{u}^*$ is independent of time. Therefore, this part of the equation can be eliminated by differentiating both sides with respect to time in order to obtain

$$\begin{aligned}\frac{d}{dt}\left[\sum_{T_k\in\mathcal{T}_h}\int_{T_k}H\left(\boldsymbol{u},\boldsymbol{u}^*\right)dx\right]&=\frac{d}{dt}\left[\sum_{T_k\in\mathcal{T}_h}\int_{T_k}\left(U\left(\boldsymbol{u}\right)-U\left(\boldsymbol{u}^*\right)\right)dx\right]\\&=\frac{d}{dt}\left[\sum_{T_k\in\mathcal{T}_h}\int_{T_k}U\left(\boldsymbol{u}\right)dx\right]\\&=\sum_{T_k\in\mathcal{T}_h}\left[\int_{T_k}U_{,t}\left(\boldsymbol{u}\right)dx\right]\\&=\sum_{T_k\in\mathcal{T}_h}\left[\int_{T_k}U_{,t}\left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)\right)dx\right].\end{aligned} \tag{7.10}$$

The RHS of Eq. (7.10) is guaranteed to be non-positive by the entropy stability condition [Eq. (5.6)]. Utilizing this fact in conjunction with Eq. (7.10), it immediately follows that

$$\frac{d}{dt}\left[\sum_{T_k\in\mathcal{T}_h}\int_{T_k}H\left(\boldsymbol{u},\boldsymbol{u}^*\right)dx\right]\leq 0.$$

The above inequality ensures that a particular measure of $H(\boldsymbol{u}, \boldsymbol{u}^*)$ monotonically decreases in time. More precisely, it implies that

$$\sum_{T_k \in \mathcal{T}_h} \int_{T_k} H\left(\boldsymbol{u}(t), \boldsymbol{u}^*\right) dx \leq \sum_{T_k \in \mathcal{T}_h} \int_{T_k} H\left(\boldsymbol{u}(t_0), \boldsymbol{u}^*\right) dx, \quad \forall t \geq t_0, \tag{7.11}$$

where the fact that $\boldsymbol{u}^*$ is constant in time has been used. One may now introduce an expression for the L_2 norm into the inequality in Eq. (7.11). This can be done by noting that $U(\boldsymbol{u})$ is a strongly convex function (under the assumptions of Lemma B.5). As a result, the following inequality holds in accordance with [7,52]

$$U\left(\boldsymbol{u}^*\right) + \left(U_{,\boldsymbol{u}}\left(\boldsymbol{u}^*\right)\right)^T \left(\boldsymbol{u} - \boldsymbol{u}^*\right) + C\|\boldsymbol{u} - \boldsymbol{u}^*\|_{L_2}^2 \leq U(\boldsymbol{u}), \tag{7.12}$$

where $C > 0$ is a constant independent of $\boldsymbol{u}$ and $\boldsymbol{u}^*$, and where the L_2 norm is defined pointwise such that

$$\|\cdot\|_{L_2}^2 = (\cdot)^T (\cdot).$$

Upon rearranging the inequality in Eq. (7.12), one obtains

$$\begin{aligned} C\|\boldsymbol{u} - \boldsymbol{u}^*\|_{L_2}^2 &\leq U(\boldsymbol{u}) - U\left(\boldsymbol{u}^*\right) - \left(U_{,\boldsymbol{u}}\left(\boldsymbol{u}^*\right)\right)^T \left(\boldsymbol{u} - \boldsymbol{u}^*\right), \\ C\|\boldsymbol{u} - \boldsymbol{u}^*\|_{L_2}^2 &\leq H\left(\boldsymbol{u}, \boldsymbol{u}^*\right), \end{aligned} \tag{7.13}$$

where the definition of $H(\boldsymbol{u}, \boldsymbol{u}^*)$ [Eq. (7.3)] has been used. Upon combining the result in Eq. (7.13) with the expression on the LHS of Eq. (7.11), one obtains

$$C\|\boldsymbol{u}(t) - \boldsymbol{u}^*\|_{L_2, \mathcal{T}_h}^2 \leq \sum_{T_k \in \mathcal{T}_h} \int_{T_k} H\left(\boldsymbol{u}(t), \boldsymbol{u}^*\right) dx. \tag{7.14}$$

Similarly, for the expression on the RHS of Eq. (7.11), one obtains

$$C\|\boldsymbol{u}(t_0) - \boldsymbol{u}^*\|_{L_2, \mathcal{T}_h}^2 \leq \sum_{T_k \in \mathcal{T}_h} \int_{T_k} H\left(\boldsymbol{u}(t_0), \boldsymbol{u}^*\right) dx. \tag{7.15}$$

Upon comparing Eq. (7.15) with Eq. (7.4), it is clear that $\mathfrak{C} \geq C$.

The final L_2 stability condition is obtained by taking the positive square roots of Eqs. (7.11), (7.14), and (7.4) and thereafter combining the results in order to obtain

$$\|\boldsymbol{u}(t) - \boldsymbol{u}^*\|_{L_2, \mathcal{T}_h} \leq \left(\frac{\mathfrak{C}}{C}\right)^{1/2} \|\boldsymbol{u}(t_0) - \boldsymbol{u}^*\|_{L_2, \mathcal{T}_h}, \quad \forall t \geq t_0. \tag{7.16}$$

One immediately obtains Eq. (7.5) upon setting $c = \left(\frac{\mathfrak{C}}{C}\right)^{1/2} \geq 1$ in Eq. (7.16). □

Remark 7.5 Theorem 7.4 establishes an L_2 bound on $\boldsymbol{u}\,(t) = \boldsymbol{u}\left(\boldsymbol{v}^h\,(t)\right)$ as it evolves in time. This is a useful result, however, it is somewhat unnatural in the following sense: a function of the entropy variables $\boldsymbol{u}\left(\boldsymbol{v}^h\,(t)\right)$ is bounded in time, and yet the entropy variables themselves $\boldsymbol{v}^h\,(t)$ are not guaranteed to be bounded in a similar fashion. It would be more natural and more convenient from the standpoint of analysis to establish an L_2 bound directly on the entropy variables. This 'more natural' bound is derived in the second theorem of this section.

The following definition helps facilitate the presentation of the second theorem.

Definition 7.6 The function $\mathcal{H}\left(\boldsymbol{v}\,(\boldsymbol{u}^*)\,,\boldsymbol{v}^h\right) : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$

$$\mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right),\boldsymbol{v}^h\right) \equiv \mathcal{U}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)\right) - \mathcal{U}\left(\boldsymbol{v}^h\right) - \left(\mathcal{U}_{,\boldsymbol{v}}\left(\boldsymbol{v}^h\right)\right)^T\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right) - \boldsymbol{v}^h\right). \quad (7.17)$$

The theorem itself is as follows.

Theorem 7.7 *Suppose the DG scheme in Eq.* (4.1) *is conservative and satisfies Eq.* (5.6)*. Furthermore, suppose that* $\boldsymbol{v}^h\,(t_0) \in \boldsymbol{L}_2\,(\Omega)$*,* $\boldsymbol{u}\,(t_0) \in \boldsymbol{L}_1\,(\Omega)$*, and each component of* $\boldsymbol{v}\,(\boldsymbol{u}^*)$ *is bounded. Finally, suppose that in accordance with Lemma* B.4*,* $\mathcal{H}\left(\boldsymbol{v}\,(\boldsymbol{u}^*)\,,\boldsymbol{v}^h\right)$ *is bounded at time* t_0 *in the following weak sense*

$$\sum_{T_k \in \mathcal{T}_h} \int_{T_k} \mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right),\boldsymbol{v}^h\,(t_0)\right) dx \leq \mathfrak{C} \|\boldsymbol{v}^h\,(t_0) - \boldsymbol{v}\left(\boldsymbol{u}^*\right)\|^2_{L_2,\mathcal{T}_h}, \quad (7.18)$$

where $\mathfrak{C} > 0$ *is a constant that depends on the initial data. Then, the scheme in Eq.* (4.1) *is* L_2 *stable for all* $t \geq t_0$ *in the following sense*

$$\|\boldsymbol{v}^h\,(t) - \boldsymbol{v}\left(\boldsymbol{u}^*\right)\|_{L_2,\mathcal{T}_h} \leq c\|\boldsymbol{v}^h\,(t_0) - \boldsymbol{v}\left(\boldsymbol{u}^*\right)\|_{L_2,\mathcal{T}_h}, \quad (7.19)$$

where $c \geq 1$ *is a constant that depends on* $\mathfrak{C}$*.*

Proof The proof of this theorem is very similar to the proof of Theorem 7.4. One may begin by substituting the identity $\mathcal{U}_{,\boldsymbol{v}} = \boldsymbol{u}^T$ from Eq. (3.2) into the definition of $\mathcal{H}\left(\boldsymbol{v}\,(\boldsymbol{u}^*)\,,\boldsymbol{v}^h\right)$ in Eq. (7.17), in order to obtain

$$\mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right),\boldsymbol{v}^h\right) = \mathcal{U}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)\right) - \mathcal{U}\left(\boldsymbol{v}^h\right) - \left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)\right)^T\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right) - \boldsymbol{v}^h\right). \quad (7.20)$$

Next, one may utilize the leftmost expression in Eq. (3.5) in order to obtain the following identities

$$\mathcal{U}\left(\boldsymbol{v}^h\right) = \left(\boldsymbol{v}^h\right)^T \boldsymbol{u}\left(\boldsymbol{v}^h\right) - U\left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)\right), \quad (7.21)$$

$$\mathcal{U}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)\right) = \left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)\right)^T \boldsymbol{u}^* - U\left(\boldsymbol{u}^*\right). \quad (7.22)$$

On substituting the identities in Eqs. (7.21) and (7.22) into Eq. (7.20), one obtains

$$\mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right), \boldsymbol{v}^h\right) = U\left(\boldsymbol{u}\left(\boldsymbol{v}^h\right)\right) - U\left(\boldsymbol{u}^*\right) - \left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)\right)^T \left(\boldsymbol{u}\left(\boldsymbol{v}^h\right) - \boldsymbol{u}^*\right),$$

or equivalently

$$\mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right), \boldsymbol{v}^h\right) = U\left(\boldsymbol{u}\right) - U\left(\boldsymbol{u}^*\right) - \left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)\right)^T \left(\boldsymbol{u} - \boldsymbol{u}^*\right), \tag{7.23}$$

where in the latter expression, the fact that $\boldsymbol{u} = \boldsymbol{u}\left(\boldsymbol{v}^h\right)$ has been implicitly assumed.

Upon examining Eq. (7.23), one may immediately observe that the RHS is identical to the RHS of Eq. (7.7) from Theorem 7.4. As a result, one may follow the steps that appear in the proof of Theorem 7.4, cf. Eqs. (7.8)–(7.10), in conjunction with the entropy stability condition in Eq. (5.6), in order to prove that the time derivative of $\mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right), \boldsymbol{v}^h\right)$ is bounded above as follows

$$\frac{d}{dt}\left[\sum_{T_k \in \mathcal{T}_h} \int_{T_k} \mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right), \boldsymbol{v}^h\right) dx\right] \leq 0,$$

and

$$\sum_{T_k \in \mathcal{T}_h} \int_{T_k} \mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right), \boldsymbol{v}^h(t)\right) dx \leq \sum_{T_k \in \mathcal{T}_h} \int_{T_k} \mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right), \boldsymbol{v}^h(t_0)\right) dx, \quad \forall t \geq t_0. \tag{7.24}$$

Next, one should note that in accordance with Lemma B.6, the function $\mathcal{U}(\boldsymbol{v})$ is strongly convex, and therefore the following holds

$$\begin{aligned} C\|\boldsymbol{v}^h - \boldsymbol{v}\left(\boldsymbol{u}^*\right)\|_{L_2}^2 &\leq \mathcal{U}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)\right) - \mathcal{U}\left(\boldsymbol{v}^h\right) - \left(\mathcal{U}_{,\boldsymbol{v}}\left(\boldsymbol{v}^h\right)\right)^T \left(\boldsymbol{v}\left(\boldsymbol{u}^*\right) - \boldsymbol{v}^h\right), \\ C\|\boldsymbol{v}^h - \boldsymbol{v}\left(\boldsymbol{u}^*\right)\|_{L_2}^2 &\leq \mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right), \boldsymbol{v}^h\right), \end{aligned}$$

and

$$C\|\boldsymbol{v}^h(t) - \boldsymbol{v}\left(\boldsymbol{u}^*\right)\|_{L_2,\mathcal{T}_h}^2 \leq \sum_{T_k \in \mathcal{T}_h} \int_{T_k} \mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right), \boldsymbol{v}^h(t)\right) dx. \tag{7.25}$$

Lastly, upon combining Eqs. (7.18), (7.24), and (7.25) and taking positive square roots, one obtains the principal result [Eq. (7.19)]. □

8 L_2 stability for the DG methods with conservative variables

If the entropy stability condition [Eq. (6.17)] holds for either of the DG schemes in Eqs. (4.3) or (4.4), it can be shown that certain 'L_2 stability' conditions also hold. These conditions are summarized in the following theorems. The first theorem is prefaced with a necessary definition in order to facilitate its presentation.

Definition 8.1 The domain-averaged solution $\boldsymbol{u}^*$,

$$\boldsymbol{u}^* \equiv \frac{1}{\text{meas}\,(\Omega)} \left[\sum_{T_k \in \mathcal{T}_h} \int_{T_k} \boldsymbol{u}^h\,(t_0)\,dx \right], \tag{8.1}$$

where t_0 is an arbitrary, but fixed initial time.

The first result of this section can now be summarized in the following theorem.

Theorem 8.2 *Suppose the DG schemes in either Eqs.* (4.3) *or* (4.4) *are conservative and satisfy Eq.* (6.17). *Furthermore, suppose that* $\boldsymbol{u}^h\,(t_0) \in \boldsymbol{L}_2\,(\Omega)$, *and that* $H\left(\boldsymbol{u}^h, \boldsymbol{u}^*\right)$ *is bounded at time* t_0 *in the following weak sense*

$$\sum_{T_k \in \mathcal{T}_h} \int_{T_k} H\left(\boldsymbol{u}^h\,(t_0)\,, \boldsymbol{u}^*\right) dx \leq \mathfrak{C} \|\boldsymbol{u}^h\,(t_0) - \boldsymbol{u}^*\|^2_{L_2,\mathcal{T}_h}, \tag{8.2}$$

where $\mathfrak{C} > 0$ *is a constant that depends on the initial data. Then, the schemes in either Eqs.* (4.3) *or* (4.4) *are* L_2 *stable for all* $t \geq t_0$ *in the following sense*

$$\|\boldsymbol{u}^h\,(t) - \boldsymbol{u}^*\|_{L_2,\mathcal{T}_h} \leq c\|\boldsymbol{u}^h\,(t_0) - \boldsymbol{u}^*\|_{L_2,\mathcal{T}_h}, \tag{8.3}$$

where $c \geq 1$ *is a constant that depends on* $\mathfrak{C}$.

Proof The proof of Theorem 8.2 is essentially identical to the proof of Theorem 7.4. □

The second theorem of this section takes the following form.

Theorem 8.3 *Suppose the DG schemes in either Eqs.* (4.3) *or* (4.4) *are conservative and satisfy Eq.* (6.17). *Furthermore, suppose that* $\boldsymbol{u}^h\,(t_0) \in \boldsymbol{L}_1\,(\Omega)$, $\boldsymbol{v}\left(\boldsymbol{u}^h\,(t_0)\right) \in \boldsymbol{L}_2\,(\Omega)$, *and each component of* $\boldsymbol{v}\,(\boldsymbol{u}^*)$ *is bounded. Finally, suppose that* $\mathcal{H}\left(\boldsymbol{v}\,(\boldsymbol{u}^*)\,, \boldsymbol{v}\left(\boldsymbol{u}^h\right)\right)$ *is bounded at time* t_0 *in the following weak sense*

$$\sum_{T_k \in \mathcal{T}_h} \int_{T_k} \mathcal{H}\left(\boldsymbol{v}\,(\boldsymbol{u}^*)\,, \boldsymbol{v}\left(\boldsymbol{u}^h\,(t_0)\right)\right) dx \leq \mathfrak{C}\|\boldsymbol{v}\left(\boldsymbol{u}^h\,(t_0)\right) - \boldsymbol{v}\,(\boldsymbol{u}^*)\,\|^2_{L_2,\mathcal{T}_h}, \tag{8.4}$$

where $\mathfrak{C} > 0$ *is a constant that depends on the initial data. Then, the schemes in either Eqs.* (4.3) *or* (4.4) *are* L_2 *stable for all* $t \geq t_0$ *in the following sense*

$$\|\boldsymbol{v}\left(\boldsymbol{u}^h\,(t)\right) - \boldsymbol{v}\,(\boldsymbol{u}^*)\,\|_{L_2,\mathcal{T}_h} \leq c\|\boldsymbol{v}\left(\boldsymbol{u}^h\,(t_0)\right) - \boldsymbol{v}\,(\boldsymbol{u}^*)\,\|_{L_2,\mathcal{T}_h}, \tag{8.5}$$

where $c \geq 1$ *is a constant that depends on* $\mathfrak{C}$.

Proof The proof of Theorem 8.3 is essentially identical to the proof of Theorem 7.7. □

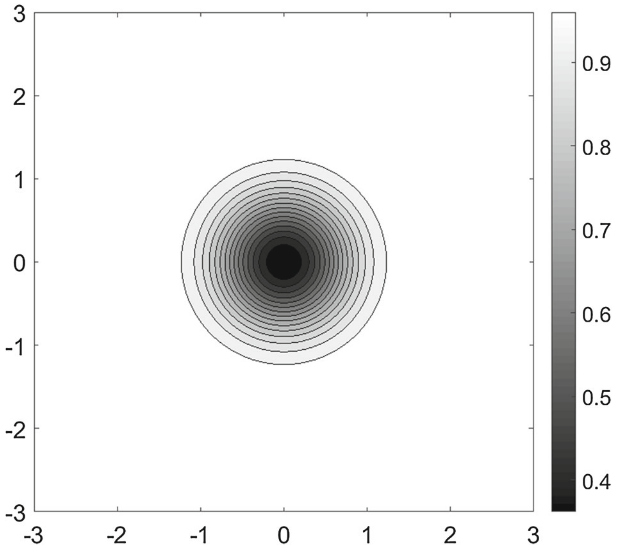

Fig. 1 Density contours for the vortex at time t = 0

9 Numerical experiments

Thus far, this article has focused on the semi-discrete formulations of DG methods. However, in this section, we consider the behavior of fully-discrete formulations that are obtained by combining DG methods for spatial discretization with RK methods for temporal discretization.

In order to evaluate the fully-discrete formulations, numerical experiments were performed on a well-known vortex propagation problem [9,57,60,68]. The problem consists of a 2D vortex which is initially centered at the point $(0, 0)$ and travels horizontally in the x-direction with a unit velocity. It has the following exact solution

$$\begin{aligned} u &= 1 - \varphi \exp\left(1 - r^2\right) \frac{y}{2\pi}, \\ v &= \varphi \exp\left(1 - r^2\right) \frac{x}{2\pi}, \\ \rho &= \left(1 - \left(\frac{\gamma - 1}{16\gamma\pi^2}\right) \varphi^2 \exp\left(2\left(1 - r^2\right)\right)\right)^{\frac{1}{\gamma-1}}, \\ p &= \rho^\gamma, \end{aligned}$$

where

$$r = \sqrt{(x-t)^2 + y^2},$$

and where φ is the vortex strength. Throughout the experiments, φ was set to 5 and γ was set to 1.4. An illustration of the vortex at time $t = 0$ is shown in Fig. 1.

The vortex was simulated on a 10×10 domain ($-5 \leq x, y \leq 5$) with periodic boundary conditions imposed on all boundaries. The domain was covered with a structured mesh of triangular elements, and the mesh was formed by creating a 10×10 Cartesian grid of square elements, and then splitting the squares along the diagonals (from top-left to bottom-right) to form triangles. On this mesh, two representative DG methods were evaluated: the DG method with entropy variables [Eq. (4.1)] and the DG method with conservative variables [Eq. (4.4)]. The DG method in Eq. (4.3) was omitted for brevity's sake. Throughout the experiments, the spatial polynomial order p for each method was set to 4, unless indicated otherwise. Finally, the numerical flux for each method was chosen to be the Lax-Friedrichs flux (as described in [38]). As mentioned previously in Remark 6.5, this flux is known to be entropy stable if it is utilized in conjunction with entropy variables [4,5].

In each experiment, the vortex was propagated until a final time of $t = 500$. The time discretization was performed with an algebraically stable, 4th-order, 4-stage, Singly Diagonally Implicit Runge–Kutta (SDIRK) method due to [1,8] (see "Appendix C"). The non-linear system at each stage of the SDIRK method was linearized using Newton's method, and the resulting linear system was solved using the Generalized Minimal Residual Method (GMRES) [55]. The step-size for Newton's method was controlled using a bisecting, line-search procedure. The iterative solution process was implemented with the aid of PETSc, the Portable, Extensible Toolkit for Scientific Computations, cf. [2,31] for details.

Before proceeding further, we note that an SDIRK method was used because, unlike explicit RK schemes, its time-step is not constrained by a CFL limit, which gives it broader applicability to high-Reynolds number flows that require high aspect-ratio mesh elements. While the current problem is not a problem of this type, these flows will be an important application area in future work.

Time-steps of $t = 0.1$ and $t = 0.05$ were chosen. At each time-step, the norm $\left\| \boldsymbol{v}^h - \boldsymbol{v}\left(\boldsymbol{u}^*\right) \right\|_{L_2, \mathcal{T}_h}$ in Theorem 7.7 was evaluated for the DG method with entropy variables [Eq. (4.1)], and the norm $\left\| \boldsymbol{u}^h - \boldsymbol{u}^* \right\|_{L_2, \mathcal{T}_h}$ in Theorem 8.2 was evaluated for the DG method with conservative variables [Eq. (4.4)]. In order to virtually eliminate integration errors, the norms and the inner products in the finite element methods were evaluated with numerical quadrature rules which exactly integrated polynomials of degree ≤ 30.

In accordance with Theorems 7.7 and 8.2, it was optimistically expected that the norms, $\left\| \boldsymbol{v}^h - \boldsymbol{v}\left(\boldsymbol{u}^*\right) \right\|_{L_2, \mathcal{T}_h}$ and $\left\| \boldsymbol{u}^h - \boldsymbol{u}^* \right\|_{L_2, \mathcal{T}_h}$, would remain bounded in time. However, temporal stability was not guaranteed, as the assumptions of Theorems 7.7 and 8.2 were not exactly satisfied. In particular, both theorems govern the semi-discrete formulations of the DG methods and not the fully-discrete formulations. In addition, Theorem 8.2 assumes that the entropy projection errors are small, or that the entropy projection terms are non-positive; neither of these assumptions is guaranteed to hold true for this test case, as the mesh was very coarse. As a result, it was optimistically expected that Theorems 7.7 and 8.2 would capture the general tendencies of the methods, but not the exact behavior.

The norms for both methods are plotted in Figs. 2 and 3. Based on Fig. 2, it is clear that the DG method with entropy variables experiences some significant oscillations,

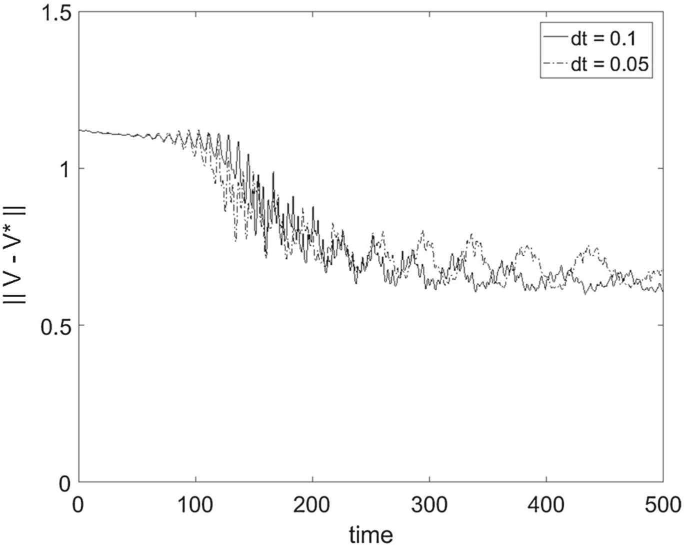

Fig. 2 The evolution of $\left\| \boldsymbol{v}^h - \boldsymbol{v}\left(\boldsymbol{u}^*\right) \right\|_{L_2, \mathcal{T}_h}$ for the vortex test case. This result was obtained using the DG method with entropy variables in Eq. (4.1)

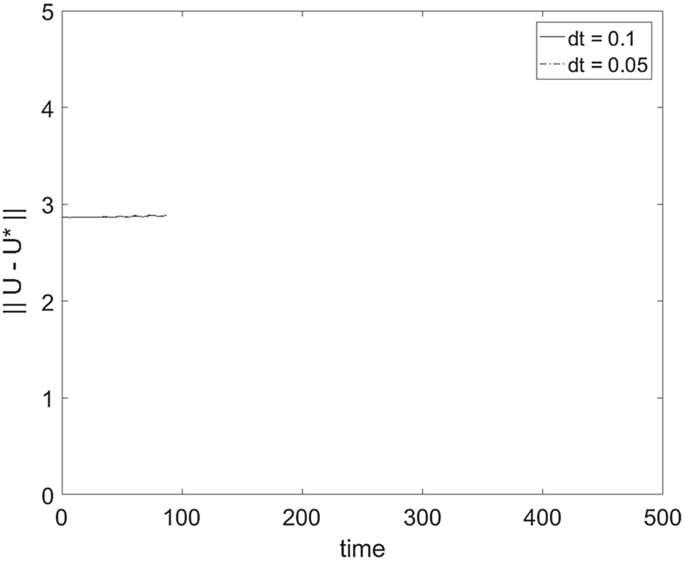

Fig. 3 The evolution of $\left\| \boldsymbol{u}^h - \boldsymbol{u}^* \right\|_{L_2, \mathcal{T}_h}$ for the vortex test case. This result was obtained using the DG method with conservative variables in Eq. (4.4)

but does not diverge. In addition, the norm for this scheme decrease monotonically (on average). However, based on Fig. 3, it is clear that the DG method with conservative variables stops running at roughly $t = 90$ seconds. At this point, NAN's were detected in the solution.

The experiments were repeated with $p = 5$ and $dt = 0.1$. The results are similar to those of the $p = 4$ case, and are not shown for the sake of brevity. As before, the DG method with conservative variables diverges, and the DG method with entropy variables remains stable.

Overall, the DG method with entropy variables behaves in accordance with Theorem 7.7 for long times, whereas, the DG method with conservative variables does not behave in accordance with Theorem 8.2.

10 Conclusion

Several 'typical' DG methods were utilized to spatially discretize the compressible Euler equations, and thereafter, the stability characteristics of the resulting semi-discrete formulations were evaluated. Entropy and L_2 stability were shown for the semi-discrete formulation of a DG method with entropy variables as its unknowns (cf. Theorems 5.5, 7.4, and 7.7). In addition, entropy and L_2 stability were shown for the semi-discrete formulations of two DG methods with conservative variables as their unknowns, under the assumption that entropy projection errors vanish, or the entropy projection terms are non-positive (cf. Theorems 6.2, 6.8, 8.2, and 8.3).

In the latter part of this work, the semi-discrete formulations associated with the DG methods were fully discretized using algebraically stable RK methods. Thereafter, the temporal stability properties of the resulting fully-discrete formulations were evaluated with numerical experiments on a vortex propagation problem. Based on the experiments, it was clear that the DG method with entropy variables remained stable and (on average) demonstrated monotonic behavior for long times, in accordance with an optimistic interpretation of Theorem 7.7. Conversely, a DG method with conservative variables diverged, failing to behave in accordance with an optimistic interpretation of Theorem 8.2. In this case, it is theorized that the entropy projection errors were too large, or the entropy projection terms had the wrong sign, preventing Theorem 8.2 from holding. In future work, it may be beneficial to construct approaches that specifically minimize or control these entropy projection errors. Also, it is the author's belief that additional numerical experiments are required for a more complete comparison of DG methods which utilize entropy and conservative variables.

Acknowledgements The author would like to thank Dr. Siddhartha Mishra (ETH Zürich), Dr. Sigal Gottlieb (University of Massachusetts, Dartmouth), Dr. Dmitry Kamenetskiy (Boeing Research & Technology), Dr. Krzysztof Fidkowski (University of Michigan), Dr. Jesse Chan (Rice University), and Dr. Jason Hicken (Rensselaer Polytechnic Institute) for their participation in conversations that helped shape this work. The author would also like to thank Dr. David Darmofal, Dr. Marshall Galbraith, and Dr. Steven Allmaras (Massachusetts Institute of Technology) for developing the Solution Adaptive Numerical Simulator (SANS) code [30,31] that was utilized for the numerical experiments.

Appendix A: Notational conventions

The principal results in this paper are expressed via a mixture of index notation and vector notation. This combination of notation is best explained with an example. Consider the term $\left(\boldsymbol{w}^h_{,\boldsymbol{x}_i}\right)^T \boldsymbol{f}^i\left(\boldsymbol{u}^h\right)$. This can be expanded as follows when $d = 2$

$$\left(\boldsymbol{w}^h_{,\boldsymbol{x}_i}\right)^T \boldsymbol{f}^i\left(\boldsymbol{u}^h\right) = \left(\frac{\partial \boldsymbol{w}^h}{\partial \boldsymbol{x}_1}\right)^T \boldsymbol{f}^1\left(\boldsymbol{u}^h\right) + \left(\frac{\partial \boldsymbol{w}^h}{\partial \boldsymbol{x}_2}\right)^T \boldsymbol{f}^2\left(\boldsymbol{u}^h\right),$$

where the repeated index i facilitates the standard Einstein summation over d dimensions, and the transpose facilitates the standard dot product between a pair of m-vectors [for instance $\boldsymbol{w}^h_{,\boldsymbol{x}_1}$ and $\boldsymbol{f}^1\left(\boldsymbol{u}^h\right)$].

Appendix B: Supporting lemmas

Lemma B.1 *Suppose that the pressure p and density ρ are bounded in the following fashion*

$$0 < p \leq M, \quad 0 < \rho \leq N, \tag{B.1}$$

where M and N are positive numbers. Then, the symmetric Jacobian matrices $\widetilde{A}_0$ and $\widetilde{A}_0^{-1}$ (defined in Eq. (3.7)) are guaranteed to be positive definite (PD).

Proof In what follows, we present the proof for the 2D case. The proof for the 3D case and for higher dimensions, is very similar. In 2D, one may begin by examining the matrix $\widetilde{A}_0$, which can be expressed as follows

$$\widetilde{A}_0 = \begin{bmatrix} \rho & \rho u & \rho v & \rho H - p \\ \rho u & \rho u^2 + p & \rho u v & \rho u H \\ \rho v & \rho u v & \rho v^2 + p & \rho v H \\ \rho H - p & \rho u H & \rho v H & \rho H^2 - \gamma e p \end{bmatrix},$$

where

$$H = \gamma e + \frac{1}{2}\left(u^2 + v^2\right).$$

The characteristic polynomial of $\widetilde{A}_0$ takes the following form

$$\frac{1}{4}(\lambda - p)\left(a\lambda^3 + b\lambda^2 + c\lambda + d\right) = 0, \tag{B.2}$$

where

$$\begin{aligned}
a &= 4,\\
b &= -\rho\left(\left(2+u^2+v^2\right)^2+4e\left(\gamma\left(1+u^2+v^2+e\right)-1\right)\right),\\
c &= \rho p\left(\left(2+u^2+v^2\right)^2+4e\left(\gamma e+1\right)\right),\\
d &= -4\rho e p^2.
\end{aligned} \tag{B.3}$$

It immediately follows from Eq. (B.2), that the pressure p is an eigenvalue. Evidently, this eigenvalue is positive since Eq. (B.1) holds.

One may determine the signs of the remaining eigenvalues of $\widetilde{A}_0$ by first observing that the matrix is symmetric and will have all real entries if Eq. (B.1) is satisfied. As a result, the eigenvalues of $\widetilde{A}_0$ will be real, i.e. the roots of its characteristic polynomial will be real. In accordance with Decartes' rule of signs, a cubic polynomial with real roots, and with coefficients $a > 0, b < 0, c > 0$, and $d < 0$ is guaranteed to have three positive roots. By inspection of Eq. (B.3), it immediately follows that the coefficients a, b, c, and d will have the required signs if Eq. (B.1) holds, if $\gamma > 1$ (which is true for virtually all gases), and if one notes that $e = p/(\rho(\gamma-1)) > 0$. This completes the proof that $\widetilde{A}_0$ has positive eigenvalues.

The eigenvalues of $\widetilde{A}_0^{-1}$ are obtained by taking the reciprocals of the eigenvalues of $\widetilde{A}_0$. The eigenvalues are positive as long as Eq. (B.1) holds, as under these circumstances, the eigenvalues of $\widetilde{A}_0$ remain positive, and the reciprocals of positive numbers are positive. This completes the proof that $\widetilde{A}_0^{-1}$ has positive eigenvalues. □

Lemma B.2 *The jump in the entropy functional $F^i(\boldsymbol{v})$ across an interface is governed by the following*

$$\begin{aligned}
&\left[\!\left[F^i\left(\boldsymbol{u}^h\right)\right]\!\right]_+^- + \left(\left\{\!\left\{\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right\}\!\right\}_+^-\right)^T \left[\!\left[\boldsymbol{f}^i\left(\boldsymbol{u}^h\right)\right]\!\right]_-^+\\
&\quad= \frac{1}{2}\int_0^1 (1-\theta)\left(\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_-^+\right)^T \left(\widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}(\theta)\right)-\widetilde{A}_i\left(\overline{\boldsymbol{v}}(\theta)\right)\right)\left[\!\left[\boldsymbol{v}\left(\boldsymbol{u}^h\right)\right]\!\right]_-^+ d\theta.
\end{aligned} \tag{B.4}$$

Proof The proof is virtually identical to the proofs of similar statements in [69] and [5]. It is repeated here for the sake of completeness.

Recall that $\mathcal{F}^i = \mathcal{F}^i(\boldsymbol{v}(\boldsymbol{u}))$. On utilizing this fact in conjunction with Taylor's theorem, one can obtain the following

$$\begin{aligned}&\mathcal{F}^i\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)\right)-\mathcal{F}^i\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)-\mathcal{F}^i_{,\boldsymbol{v}}\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)\right)\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)\\&\quad+\int_0^1(1-\theta)\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)^T\mathcal{F}^i_{,\boldsymbol{v},\boldsymbol{v}}\left(\overline{\boldsymbol{v}}\left(\theta\right)\right)\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)d\theta=0,\end{aligned}\tag{B.5}$$

$$\begin{aligned}&\mathcal{F}^i\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)\right)-\mathcal{F}^i\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)-\mathcal{F}^i_{,\boldsymbol{v}}\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)\\&\quad-\int_0^1(1-\theta)\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)^T\mathcal{F}^i_{,\boldsymbol{v},\boldsymbol{v}}\left(\overline{\overline{\boldsymbol{v}}}\left(\theta\right)\right)\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)d\theta=0.\end{aligned}\tag{B.6}$$

Upon multiplying Eqs. (B.5) and (B.6) by (1/2) and summing the results, one obtains

$$\begin{aligned}&\mathcal{F}^i\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)\right)-\mathcal{F}^i\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)-\frac{1}{2}\left(\boldsymbol{f}^i\left(\boldsymbol{u}_+^h\right)+\boldsymbol{f}^i\left(\boldsymbol{u}_-^h\right)\right)^T\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)\\&\quad=\frac{1}{2}\int_0^1(1-\theta)\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)^T\left(\widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}\left(\theta\right)\right)-\widetilde{A}_i\left(\overline{\boldsymbol{v}}\left(\theta\right)\right)\right)\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)d\theta,\end{aligned}\tag{B.7}$$

where the fact that $\left(\boldsymbol{f}^i\right)^T=\mathcal{F}^i_{,\boldsymbol{v}}$ and $\widetilde{A}_i=\mathcal{F}^i_{,\boldsymbol{v},\boldsymbol{v}}$ has been used (cf. Eqs. (3.3) and (3.7)). Setting Eq. (B.7) aside for the moment, consider the following jump identity that derives from Eq. (3.5)

$$\begin{aligned}&F^i\left(\boldsymbol{u}_+^h\right)-F^i\left(\boldsymbol{u}_-^h\right)+\mathcal{F}^i\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)\right)-\mathcal{F}^i\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)\\&\quad=\frac{1}{2}\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)+\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)^T\left(\boldsymbol{f}^i\left(\boldsymbol{u}_+^h\right)-\boldsymbol{f}^i\left(\boldsymbol{u}_-^h\right)\right)\\&\qquad+\frac{1}{2}\left(\boldsymbol{f}^i\left(\boldsymbol{u}_+^h\right)+\boldsymbol{f}^i\left(\boldsymbol{u}_-^h\right)\right)^T\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right).\end{aligned}\tag{B.8}$$

Substituting Eq. (B.7) into Eq. (B.8) yields

$$\begin{aligned}&F^i\left(\boldsymbol{u}_-^h\right)-F^i\left(\boldsymbol{u}_+^h\right)+\frac{1}{2}\left(\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)+\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)\right)^T\left(\boldsymbol{f}^i\left(\boldsymbol{u}_+^h\right)-\boldsymbol{f}^i\left(\boldsymbol{u}_-^h\right)\right)\\&\quad=\frac{1}{2}\int_0^1(1-\theta)\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)^T\left(\widetilde{A}_i\left(\overline{\overline{\boldsymbol{v}}}\left(\theta\right)\right)-\widetilde{A}_i\left(\overline{\boldsymbol{v}}\left(\theta\right)\right)\right)\left(\boldsymbol{v}\left(\boldsymbol{u}_+^h\right)-\boldsymbol{v}\left(\boldsymbol{u}_-^h\right)\right)d\theta.\end{aligned}\tag{B.9}$$

Upon substituting the spatial jump [Eq. (5.1)] and average [Eq. (5.2)] operators into Eq. (B.9) one obtains Eq. (B.4). □

Lemma B.3 *Suppose that the initial condition $\boldsymbol{u}\left(t_0\right)\in\boldsymbol{L}_2\left(\Omega\right)$, and the pressure p and density ρ are positive and bounded for all convex combinations of states $\boldsymbol{u}\left(t_0\right)$ and $\boldsymbol{u}^*$ defined by $\widehat{\widehat{\boldsymbol{u}}}\left(\theta\right)=\boldsymbol{u}^*+\theta\left(\boldsymbol{u}\left(t_0\right)-\boldsymbol{u}^*\right)$, where $0\leq\theta\leq 1$. Under these circumstances, the following statements hold: (i) the L_2 norm of $\boldsymbol{u}\left(t_0\right)-\boldsymbol{u}^*$ is well-defined; (ii) the eigenvalues of $\widetilde{A}_0^{-1}\left(\widehat{\widehat{\boldsymbol{u}}}\left(\theta\right)\right)$ are real and positive; and (iii)*

the broken integral of $H\left(\boldsymbol{u}\left(t_0\right), \boldsymbol{u}^*\right)$ over the domain $\mathcal{T}_h$ is bounded in the following fashion

$$\sum_{T_k \in \mathcal{T}_h} \int_{T_k} H\left(\boldsymbol{u}\left(t_0\right), \boldsymbol{u}^*\right) dx \leq \frac{\lambda_{1,\mathcal{T}_h}}{2} \|\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\|^2_{L_2,\mathcal{T}_h}, \tag{B.10}$$

where $\lambda_{1,\mathcal{T}_h}$ is the maximum eigenvalue of $\widetilde{A}_0^{-1}\left(\widehat{\widehat{\boldsymbol{u}}}\left(\theta\right)\right)$ over all elements in the domain.

Note: throughout this lemma we implicitly assume that $\boldsymbol{u}\left(t_0\right) = \boldsymbol{u}\left(\boldsymbol{v}^h\left(t_0\right)\right)$.

Proof The proof of part (i) follows from inspection, and the proof of part (ii) is given in Lemma B.1. Therefore, it remains to prove part (iii). One may begin the proof by utilizing Taylor's Theorem in order to express $U\left(\boldsymbol{u}\left(t_0\right)\right)$ in terms of $U\left(\boldsymbol{u}^*\right)$ as follows

$$\begin{aligned} U\left(\boldsymbol{u}\left(t_0\right)\right) &= U\left(\boldsymbol{u}^*\right) + \left(U_{,\boldsymbol{u}}\left(\boldsymbol{u}^*\right)\right)^T \left(\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right) \\ &\quad + \int_0^1 \left(1-\theta\right) \left(\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right)^T U_{,\boldsymbol{u},\boldsymbol{u}}\left(\widehat{\widehat{\boldsymbol{u}}}\left(\theta\right)\right) \left(\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right) d\theta. \end{aligned} \tag{B.11}$$

Upon substituting the expression for $H\left(\boldsymbol{u}\left(t_0\right), \boldsymbol{u}^*\right)$ (as given by definition 7.3) into Eq. (B.11), one obtains

$$\begin{aligned} H\left(\boldsymbol{u}\left(t_0\right), \boldsymbol{u}^*\right) &= \int_0^1 \left(1-\theta\right) \left(\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right)^T U_{,\boldsymbol{u},\boldsymbol{u}}\left(\widehat{\widehat{\boldsymbol{u}}}\left(\theta\right)\right) \left(\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right) d\theta \\ &= \int_0^1 \left(1-\theta\right) \left(\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right)^T \widetilde{A}_0^{-1}\left(\widehat{\widehat{\boldsymbol{u}}}\left(\theta\right)\right) \left(\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right) d\theta, \end{aligned} \tag{B.12}$$

where the definition of the inverse Jacobian matrix $\widetilde{A}_0^{-1} = U_{,\boldsymbol{u},\boldsymbol{u}}$ has been used in the last line (see Eqs. (3.4) and (3.7)). The matrix $\widetilde{A}_0^{-1}\left(\widehat{\widehat{\boldsymbol{u}}}\left(\theta\right)\right)$ has a maximum eigenvalue denoted by $\lambda_1\left(\theta\right) = \lambda_1\left(\widetilde{A}_0^{-1}\left(\widehat{\widehat{\boldsymbol{u}}}\left(\theta\right)\right)\right)$. Therefore, the last three terms that appear under the integral sign in Eq. (B.12) can be bounded from above as follows

$$\begin{aligned} &\left(\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right)^T \widetilde{A}_0^{-1}\left(\widehat{\widehat{\boldsymbol{u}}}\left(\theta\right)\right) \left(\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right) \\ &\quad \leq \lambda_1\left(\theta\right) \left(\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right)^T \left(\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right), \\ &\quad \leq \lambda_1\left(\theta\right) \left\|\boldsymbol{u}\left(t_0\right) - \boldsymbol{u}^*\right\|^2_{L_2}, \end{aligned} \tag{B.13}$$

where several standard results from Linear Algebra (cf. for instance [61]) have been used. On substituting Eq. (B.13) into Eq. (B.12) and integrating both sides over the entire domain, one obtains

$$\begin{aligned}\sum_{T_k\in\mathcal{T}_h}\int_{T_k} H\left(\boldsymbol{u}\left(t_0\right),\boldsymbol{u}^*\right)dx &\leq \sum_{T_k\in\mathcal{T}_h}\int_{T_k}\int_0^1\left(\left(1-\theta\right)\lambda_1\left(\theta\right)\left\|\boldsymbol{u}\left(t_0\right)-\boldsymbol{u}^*\right\|_{L_2}^2\right)d\theta\, dx,\\ &\leq \sum_{T_k\in\mathcal{T}_h}\left(\int_{T_k}\lambda_1\left\|\boldsymbol{u}\left(t_0\right)-\boldsymbol{u}^*\right\|_{L_2}^2 dx\right)\int_0^1\left(1-\theta\right)d\theta,\\ &\leq \frac{1}{2}\lambda_{1,\mathcal{T}_h}\left\|\boldsymbol{u}\left(t_0\right)-\boldsymbol{u}^*\right\|_{L_2,\mathcal{T}_h}^2, \end{aligned}\qquad\text{(B.14)}$$

where we have defined λ_1 to be the maximum of $\lambda_1(\theta)$ over all $0 \leq \theta \leq 1$, and $\lambda_{1,\mathcal{T}_h}$ to be the maximum of λ_1 over all elements in the domain. Finally, the proof is completed by comparing Eq. (B.14) with Eq. (B.10). □

Lemma B.4 *Suppose that the initial condition* $\boldsymbol{v}^h(t_0) \in \boldsymbol{L}_2(\Omega)$*, the initial condition* $\boldsymbol{u}(t_0) \in \boldsymbol{L}_1(\Omega)$*, each component of* $\boldsymbol{v}(\boldsymbol{u}^*)$ *is bounded, and the pressure* p *and density* ρ *are positive and bounded for all convex combinations of states* $\boldsymbol{v}^h(t_0)$ *and* $\boldsymbol{v}(\boldsymbol{u}^*)$ *defined by* $\widehat{\widehat{\boldsymbol{v}}}(\theta) = \boldsymbol{v}^h(t_0) + \theta\left(\boldsymbol{v}(\boldsymbol{u}^*) - \boldsymbol{v}^h(t_0)\right)$*, where* $0 \leq \theta \leq 1$*. Under these circumstances, the following statements hold: (i) the* L_2 *norm of* $\boldsymbol{v}^h(t_0) - \boldsymbol{v}(\boldsymbol{u}^*)$ *is well-defined; (ii) the eigenvalues of* $\widetilde{A}_0\left(\widehat{\widehat{\boldsymbol{v}}}(\theta)\right)$ *are real and positive; and (iii) the broken integral of* $\mathcal{H}\left(\boldsymbol{v}(\boldsymbol{u}^*), \boldsymbol{v}^h(t_0)\right)$ *over the domain* $\mathcal{T}_h$ *is bounded in the following fashion*

$$\sum_{T_k\in\mathcal{T}_h}\int_{T_k}\mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right),\boldsymbol{v}^h\left(t_0\right)\right)dx \leq \frac{\lambda_{1,\mathcal{T}_h}}{2}\|\boldsymbol{v}^h\left(t_0\right)-\boldsymbol{v}\left(\boldsymbol{u}^*\right)\|_{L_2,\mathcal{T}_h}^2, \qquad\text{(B.15)}$$

where $\lambda_{1,\mathcal{T}_h}$ *is the maximum eigenvalue of* $\widetilde{A}_0\left(\widehat{\widehat{\boldsymbol{v}}}(\theta)\right)$*, over all elements in the domain.*

Note: throughout this lemma we implicitly assume that $\boldsymbol{u}(t_0) = \boldsymbol{u}\left(\boldsymbol{v}^h(t_0)\right)$*.*

Proof The proof of this lemma is very similar to the proof of Lemma B.3. In particular, only part (iii) requires proof. One may begin the proof by using Taylor's Theorem in order to express $\mathcal{U}(\boldsymbol{v}(\boldsymbol{u}^*))$ in terms of $\mathcal{U}\left(\boldsymbol{v}^h(t_0)\right)$ as follows

$$\begin{aligned}\mathcal{U}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)\right) =& \mathcal{U}\left(\boldsymbol{v}^h\left(t_0\right)\right) + \left(\mathcal{U}_{,\boldsymbol{v}}\left(\boldsymbol{v}^h\left(t_0\right)\right)\right)^T\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)-\boldsymbol{v}^h\left(t_0\right)\right)\\ &+\int_0^1\left(1-\theta\right)\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)-\boldsymbol{v}^h\left(t_0\right)\right)^T\mathcal{U}_{,\boldsymbol{v},\boldsymbol{v}}\left(\widehat{\widehat{\boldsymbol{v}}}\left(\theta\right)\right)\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)-\boldsymbol{v}^h\left(t_0\right)\right)d\theta.\end{aligned}\qquad\text{(B.16)}$$

On substituting the expression for $\mathcal{H}\left(\boldsymbol{v}(\boldsymbol{u}^*), \boldsymbol{v}^h(t_0)\right)$ (as given by Definition 7.6) into Eq. (B.16), one obtains

$$\begin{aligned}&\mathcal{H}\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right),\boldsymbol{v}^h\left(t_0\right)\right)\\ &\quad=\int_0^1\left(1-\theta\right)\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)-\boldsymbol{v}^h\left(t_0\right)\right)^T\mathcal{U}_{,\boldsymbol{v},\boldsymbol{v}}\left(\widehat{\widehat{\boldsymbol{v}}}\left(\theta\right)\right)\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)-\boldsymbol{v}^h\left(t_0\right)\right)d\theta\\ &\quad=\int_0^1\left(1-\theta\right)\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)-\boldsymbol{v}^h\left(t_0\right)\right)^T\widetilde{A}_0\left(\widehat{\widehat{\boldsymbol{v}}}\left(\theta\right)\right)\left(\boldsymbol{v}\left(\boldsymbol{u}^*\right)-\boldsymbol{v}^h\left(t_0\right)\right)d\theta,\end{aligned}\qquad\text{(B.17)}$$

where the definition of the Jacobian matrix $\widetilde{A}_0 = \mathcal{U}_{,\boldsymbol{v},\boldsymbol{v}}$ has been used in the last line [see Eqs. (3.2) and (3.7)]. The remainder of the proof requires a simple eigenvalue analysis in order to find the maximum eigenvalue $\lambda_1(\theta) = \lambda_1\left(\widetilde{A}_0\left(\widehat{\widehat{\boldsymbol{v}}}(\theta)\right)\right)$, and thereafter, the construction of an upper bound with the maximum eigenvalue and the L_2 norm

$$\|\boldsymbol{v}^h(t_0) - \boldsymbol{v}\left(\boldsymbol{u}^*\right)\|_{L_2},$$

in accordance with the proof of Lemma B.3. □

Lemma B.5 *The functional $U(\boldsymbol{u})$ is strongly convex on the set $\mathscr{C} \subset$ **dom** $U(\boldsymbol{u})$, under the assumptions that $\mathscr{C}$ is a convex set, and that the eigenvalues of the matrix $\widetilde{A}_0^{-1}(\boldsymbol{u})$ are bounded away from zero for all $\boldsymbol{u} \in \mathscr{C}$, where $\widetilde{A}_0^{-1}$ is defined in Eq. (3.7).*

Proof In order to prove the strong convexity of U, one must obtain a particular inequality governing the Hessian, $U_{,\boldsymbol{u},\boldsymbol{u}}$, for all $\boldsymbol{u} \in \mathscr{C}$. Towards this end, one may first note that the following identity holds in accordance with Eqs. (3.4) and (3.7)

$$U_{,\boldsymbol{u},\boldsymbol{u}} = \boldsymbol{v}_{,\boldsymbol{u}} = \widetilde{A}_0^{-1}.$$

Here, the inverse Jacobian ($\widetilde{A}_0^{-1}$) is SPD because the Jacobian itself ($\widetilde{A}_0$) is SPD under the assumptions of Lemma B.1. As a result, the eigenvalues of $U_{,\boldsymbol{u},\boldsymbol{u}} = \widetilde{A}_0^{-1}$ are greater than zero, and are guaranteed to satisfy the following inequality

$$\lambda_1 \geq \cdots \geq \lambda_m > 0.$$

Furthermore, it is common practice (cf. [6,62]) to assume that the minimum eigenvalue of $\widetilde{A}_0^{-1}$, for all $\boldsymbol{u} \in \mathscr{C}$, is bounded away from zero

$$\lambda_m \geq 2C,$$

where $C > 0$ is a constant independent of $\boldsymbol{u}$. Setting this result aside for the moment, one may utilize the Hessian $U_{,\boldsymbol{u},\boldsymbol{u}}$ in order to construct a quadratic form

$$\boldsymbol{u}_b^T\left(U_{,\boldsymbol{u},\boldsymbol{u}}(\boldsymbol{u}_a)\right)\boldsymbol{u}_b, \tag{B.18}$$

where $\boldsymbol{u}_a$ and $\boldsymbol{u}_b$ are arbitrary values of $\boldsymbol{u} \in \mathscr{C}$. In accordance with several standard results from Linear Algebra (cf. for instance [61]), the quadratic form in Eq. (B.18) satisfies the following inequalities

$$\begin{aligned} \boldsymbol{u}_b^T\left(U_{,\boldsymbol{u},\boldsymbol{u}}(\boldsymbol{u}_a)\right)\boldsymbol{u}_b &\geq \lambda_m \boldsymbol{u}_b^T \boldsymbol{u}_b, \\ &= \lambda_m \|\boldsymbol{u}_b\|_{L_2}^2, \\ &\geq 2C\|\boldsymbol{u}_b\|_{L_2}^2. \end{aligned} \tag{B.19}$$

From Eq. (B.19) and the strong convexity criterion in [7], p. 459, it immediately follows that $U(\boldsymbol{u})$ is strongly convex. □

Lemma B.6 *The functional $\mathcal{U}(\boldsymbol{v})$ is strongly convex on the set $\mathscr{C} \subset$ **dom** $\mathcal{U}(\boldsymbol{v})$, under the assumptions that $\mathscr{C}$ is a convex set, and that the eigenvalues of the matrix $\widetilde{A}_0(\boldsymbol{v})$ are bounded away from zero for all $\boldsymbol{v} \in \mathscr{C}$, where $\widetilde{A}_0$ is defined in Eq.* (3.7)*.*

Proof The proof of this lemma is very similar to the proof of Lemma B.5. The proof begins with the observation that $\widetilde{A}_0$ is SPD under the assumptions of Lemma B.1, and that

$$\mathcal{U}_{,\boldsymbol{v},\boldsymbol{v}} = \boldsymbol{u}_{,\boldsymbol{v}} = \widetilde{A}_0,$$

in accordance with Eqs. (3.2) and (3.7). The rest of the proof requires a simple eigenvalue analysis, and the construction of inequalities involving the quadratic form,

$$\boldsymbol{v}_b^T \left(\mathcal{U}_{,\boldsymbol{v},\boldsymbol{v}}(\boldsymbol{v}_a)\right) \boldsymbol{v}_b,$$

in accordance with the proof of Lemma B.5. □

Appendix C: Algebraically stable SDIRK methods

In this section, the Butcher tables for some well-known algebraically stable SDIRK methods are presented.

The 1-stage 1st-order 'Backward Euler' method has the following Butcher table

$$\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}$$

The 2-stage 2nd-order method due to [1] and [8] has the following Butcher table

$$\begin{array}{c|cc} \zeta & \zeta & 0 \\ c_2 & a_{21} & \zeta \\ \hline & b_1 & b_2 \end{array}$$

where

$$\zeta = \frac{\left(2 \pm \sqrt{2}\right)}{2},$$
$$a_{21} = 1 - 2\zeta, \qquad b_1 = \frac{1}{2}, \qquad b_2 = \frac{1}{2}, \qquad c_2 = 1 - \zeta.$$

The 3-stage 3rd-order method due to [1] and [8] has the following Butcher table

$$\begin{array}{c|ccc} \zeta & \zeta & 0 & 0 \\ c_2 & a_{21} & \zeta & 0 \\ c_3 & a_{31} & a_{32} & \zeta \\ \hline & b_1 & b_2 & b_3 \end{array}$$

where

$$
\begin{aligned}
&\zeta = 0.43586652150845899941601945,\\
&|d_1| \geq 1.7742942470727850730963332, \qquad d_2 = \left(\frac{1}{2} - \zeta - d_1\right)\frac{b_2}{b_3},\\
&a_{21} = \frac{1}{2} - \zeta + d_1, \qquad a_{31} = 1 - 2\zeta - d_2, \qquad a_{32} = d_2,\\
&b_1 = \frac{d_1\,(1-2\zeta) + \frac{1}{6}}{(2\zeta - 1)\,(2\zeta - 1 - 2d_1)}, \qquad b_2 = \frac{\zeta^2 - \zeta + \frac{1}{6}}{\left(\zeta - \frac{1}{2}\right)^2 - d_1^2},\\
&b_3 = \frac{d_1\,(2\zeta - 1) + \frac{1}{6}}{(2\zeta - 1)\,(2\zeta - 1 + 2d_1)},\\
&c_2 = \frac{1}{2} + d_1, \qquad c_3 = 1 - \zeta.
\end{aligned}
$$

Note that the expression for d_2 given here is different than the corresponding expression in [8]. The definition given here is the correct definition, as the original definition contains a typographical error that prevents the method from being algebraically stable.

Finally, the 4-stage 4th-order method due to [1,8] has the following Butcher table

$$
\begin{array}{c|cccc}
\zeta & \zeta & 0 & 0 & 0\\
c_2 & a_{21} & \zeta & 0 & 0\\
c_3 & a_{31} & a_{32} & \zeta & 0\\
c_4 & a_{41} & a_{42} & a_{43} & \zeta\\
\hline
 & b_1 & b_2 & b_3 & b_4
\end{array}
$$

where

$$
\begin{aligned}
&\zeta = 0.5728160624821348554080014,\\
&d_1 = \zeta^2 - \frac{\zeta}{2} + \frac{1}{12}, \qquad d_2 = -\frac{\zeta\left(\zeta - \frac{1}{3}\right)}{\left(2\zeta^2 - \zeta + \frac{1}{6}\right)}, \qquad d_3 = \zeta^3 - \frac{3\zeta^2}{2} + \frac{\zeta}{2} - \frac{1}{24},\\
&a_{21} = \frac{\zeta\left(\frac{1}{3} - \zeta\right)}{2d_1}, \qquad a_{31} = \frac{1}{2} - \zeta - \frac{d_3}{d_1} + \frac{8d_3\left(\zeta - \frac{1}{4}\right)}{\zeta\left(\zeta - \frac{1}{3}\right)}, \qquad a_{32} = -\frac{8d_3\left(\zeta - \frac{1}{4}\right)}{\zeta\left(\zeta - \frac{1}{3}\right)},\\
&a_{41} = 1 - 2\zeta - a_{42} - \frac{d_1}{(4\zeta - 1)\left(b_2 - \frac{1}{2}\right)}, \qquad a_{42} = \frac{\left(\zeta^2 - \zeta + \frac{1}{6}\right)/2 + d_3/d_2}{d_2\left(\frac{1}{2} - b_2\right)},\\
&a_{43} = \frac{d_1}{(4\zeta - 1)\left(b_2 - \frac{1}{2}\right)}, \qquad b_1 = \frac{1}{2} - b_2, \qquad b_2 = \frac{d_1^2}{2\zeta\left(\zeta - \frac{1}{3}\right)\left(\zeta - \frac{1}{4}\right)}, \qquad b_3 = b_2,\\
&b_4 = \frac{1}{2} - b_2, \qquad c_2 = \frac{\zeta\left(\zeta - \frac{1}{2}\right)^2}{d_1}, \qquad c_3 = \frac{1}{2} - \frac{d_3}{d_1}, \qquad c_4 = 1 - \zeta.
\end{aligned}
$$

Note that the expression for a_{31} given here is different than the corresponding expression in [8]. The definition given here is the correct definition, as the original definition contains a typographical error that prevents the method from being algebraically stable.

References

1. Alexander, R.: Diagonally implicit Runge-Kutta methods for stiff ODE's. SIAM J. Numer. Anal. **14**(6), 1006–1021 (1977)
2. Balay, S., Abhyankar, S., Adams, M., Brown, J., Brune, P., Buschelman, K., Dalcin, L.D., Eijkhout, V., Gropp, W., Kaushik, D.: PETSc users manual revision 3.8. Tech. rep., Argonne National Lab.(ANL), Argonne, IL (United States) (2017)
3. Balsara, D.S., Shu, C.W.: Monotonicity preserving weighted essentially non-oscillatory schemes with increasingly high order of accuracy. J. Comput. Phys. **160**(2), 405–452 (2000)
4. Barth, T., Charrier, P., Mansour, N.N.: Energy stable flux formulas for the discontinuous Galerkin discretization of first order nonlinear conservation laws. Tech. rep, NASA (2001)
5. Barth, T.J.: Numerical methods for gasdynamic systems on unstructured meshes. In: An Introduction to Recent Developments in Theory and Numerics for Conservation Laws, pp. 195–285. Springer, Berlin Heidelberg (1999)
6. Barth, T.J.: On discontinuous Galerkin approximations of Boltzmann moment systems with Levermore closure. Comput. Methods Appl. Mech. Eng. **195**(25), 3311–3330 (2006)
7. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
8. Burrage, K.: Efficiently implementable algebraically stable Runge-Kutta methods. SIAM J. Numer. Anal. **19**(2), 245–258 (1982)
9. Castonguay, P., Vincent, P., Jameson, A.: Application of high-order energy stable flux reconstruction schemes to the Euler equations. In: 49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition (2011)
10. Chan, J., Demkowicz, L., Moser, R.: A DPG method for steady viscous compressible flow. Comput. Fluids **98**, 69–90 (2014)
11. Chan, J., Demkowicz, L., Moser, R., Roberts, N.: A class of discontinuous Petrov–Galerkin methods. Part V: Solution of 1D Burgers and Navier–Stokes equations. ICES Report **29** (2010)
12. Ciarlet, P.G., Raviart, P.A.: General Lagrange and Hermite interpolation in Rn with applications to finite element methods. Arch. Ration. Mech. Anal. **46**(3), 177–199 (1972)
13. Cockburn, B.: Discontinuous Galerkin methods. ZAMM—J. Appl. Math. Mech. **83**(11), 731–754 (2003)
14. Cockburn, B., Hou, S., Shu, C.W.: The Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws. IV. The multidimensional case. Math. Comput. **54**(190), 545–581 (1990)
15. Cockburn, B., Karniadakis, G.E., Shu, C.W.: The development of discontinuous Galerkin methods. In: Discontinuous Galerkin Methods, pp. 3–50. Springer (2000)
16. Cockburn, B., Lin, S.Y., Shu, C.W.: TVB Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws III: one-dimensional systems. J. Comput. Phys. **84**(1), 90–113 (1989)
17. Cockburn, B., Shu, C.W.: TVB Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws II. General framework. Math. Comput. **52**(186), 411–435 (1989)
18. Cockburn, B., Shu, C.W.: The Runge–Kutta local projection p^1-discontinuous Galerkin finite element method for scalar conservation laws. RAIRO-Modélisation mathématique et analyse numérique **25**(3), 337–361 (1991)
19. Cockburn, B., Shu, C.W.: The local discontinuous Galerkin method for time-dependent convection-diffusion systems. SIAM J. Numer. Anal. **35**(6), 2440–2463 (1998)
20. Cockburn, B., Shu, C.W.: The Runge–Kutta discontinuous Galerkin method for conservation laws V: multidimensional systems. J. Comput. Phys. **141**(2), 199–224 (1998)
21. Dafermos, C.: Hyperbolic Conservation Laws in Continuum Physics. Springer, Berlin (2005)
22. Demkowicz, L., Gopalakrishnan, J.: A class of discontinuous Petrov–Galerkin methods. Part I: the transport equation. Comput. Methods Appl. Mech. Eng. **199**(23), 1558–1572 (2010)
23. Demkowicz, L., Gopalakrishnan, J.: A class of discontinuous Petrov–Galerkin methods. Part II: optimal test functions. Numer. Methods Partial Differ. Equ. **27**(1), 70–105 (2011)
24. Demkowicz, L., Gopalakrishnan, J., Niemi, A.H.: A class of discontinuous Petrov–Galerkin methods. Part III: adaptivity. Appl. Numer. Math. **62**(4), 396–427 (2012)
25. Di Pietro, D.A., Ern, A.: Mathematical Aspects of Discontinuous Galerkin Methods, vol. 69. Springer, Heidelberg (2011)

26. Dutt, P.: Stable boundary conditions and difference schemes for Navier–Stokes equations. SIAM J. Numer. Anal. **25**(2), 245–267 (1988)
27. Ellis, T., Demkowicz, L., Chan, J.: Locally conservative discontinuous Petrov–Galerkin finite elements for fluid problems. Comput. Math. Appl. **68**(11), 1530–1549 (2014)
28. Fjordholm, U.S., Mishra, S., Tadmor, E.: Arbitrarily high-order accurate entropy stable essentially nonoscillatory schemes for systems of conservation laws. SIAM J. Numer. Anal. **50**(2), 544–573 (2012)
29. Fjordholm, U.S., Mishra, S., Tadmor, E.: ENO reconstruction and ENO interpolation are stable. Found. Comput. Math. **13**(2), 139–159 (2013)
30. Galbraith, M.C., Allmaras, S., Darmofal, D.L.: A verification driven process for rapid development of CFD software. In: 53rd AIAA Aerospace Sciences Meeting (2015)
31. Galbraith, M.C., Allmaras, S.R., Darmofal, D.L.: SANS RANS solutions for 3D benchmark configurations. In: 2018 AIAA Aerospace Sciences Meeting, Kissimmee, Florida (2018)
32. Godunov, S.K.: An interesting class of quasilinear systems. In: Dokl. Akad. Nauk SSSR, pp. 521–523 (1961)
33. Harten, A.: On the symmetric form of systems of conservation laws with entropy. J. Comput. Phys. **49**(1), 151–164 (1983)
34. Harten, A.: ENO schemes with subcell resolution. J. Comput. Phys. **83**(1), 148–184 (1989)
35. Harten, A., Engquist, B., Osher, S., Chakravarthy, S.R.: Uniformly high order accurate essentially nonoscillatory schemes: III. In: Upwind and High-Resolution Schemes, pp. 218–290. Springer (1987)
36. Harten, A., Osher, S.: Uniformly high-order accurate nonoscillatory schemes: I. SIAM J. Numer. Anal. **24**(2), 279–309 (1987)
37. Harten, A., Osher, S., Engquist, B., Chakravarthy, S.R.: Some results on uniformly high-order accurate essentially nonoscillatory schemes. Appl. Numer. Math. **2**(3–5), 347–377 (1986)
38. Hesthaven, J.S., Warburton, T.: Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications. Springer, New York (2007)
39. Hiltebrand, A., Mishra, S.: Entropy stable shock capturing space–time discontinuous Galerkin schemes for systems of conservation laws. Numerische Mathematik **126**(1), 103–151 (2014)
40. Hou, S., Liu, X.D.: Solutions of multi-dimensional hyperbolic systems of conservation laws by square entropy condition satisfying discontinuous Galerkin method. J. Sci. Comput. **31**(1–2), 127–151 (2007)
41. Hu, C., Shu, C.W.: Weighted essentially non-oscillatory schemes on triangular meshes. J. Comput. Phys. **150**(1), 97–127 (1999)
42. Hughes, T.J.R., Franca, L.P., Mallet, M.: A new finite element formulation for computational fluid dynamics: I. Symmetric forms of the compressible Euler and Navier–Stokes equations and the second law of thermodynamics. Comput. Methods Appl. Mech. Eng. **54**(2), 223–234 (1986)
43. Hughes, T.J.R., Mallet, M.: A new finite element formulation for computational fluid dynamics: III. The generalized streamline operator for multidimensional advective-diffusive systems. Comput. Methods Appl. Mech. Eng. **58**(3), 305–328 (1986)
44. Hughes, T.J.R., Mallet, M.: A new finite element formulation for computational fluid dynamics: IV. A discontinuity-capturing operator for multidimensional advective-diffusive systems. Comput. Methods Appl. Mech. Eng. **58**(3), 329–336 (1986)
45. Hughes, T.J.R., Mallet, M., Akira, M.: A new finite element formulation for computational fluid dynamics: II. Beyond SUPG. Comput. Methods Appl. Mech. Eng. **54**(3), 341–355 (1986)
46. Jiang, G.S., Shu, C.W.: On a cell entropy inequality for discontinuous Galerkin methods. Math. Comput. **62**(206), 531–538 (1994)
47. Kirby, R.M., Karniadakis, G.E.: De-aliasing on non-uniform grids: algorithms and applications. J. Comput. Phys. **191**(1), 249–264 (2003)
48. Lax, P.D.: Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves, vol. 11. SIAM, Philadelphia (1973)
49. Lefloch, P.G., Mercier, J.M., Rohde, C.: Fully discrete, entropy conservative schemes of arbitrary order. SIAM J. Numer. Anal. **40**(5), 1968–1992 (2002)
50. Liu, X.D., Osher, S., Chan, T.: Weighted essentially non-oscillatory schemes. J. Comput. Phys. **115**(1), 200–212 (1994)
51. Mock, M.S.: Systems of conservation laws of mixed type. J. Differ. Equ. **37**(1), 70–88 (1980)
52. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Springer, New York (2004)
53. Pugh, C.C.: Real Mathematical Analysis. Springer, New York (2002)

54. Quarteroni, A., Sacco, R., Saleri, F.: Numerical Mathematics, vol. 37. Springer, New York (2010)
55. Saad, Y., Schultz, M.H.: GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Stat. Comput. **7**(3), 856–869 (1986)
56. Shakib, F., Hughes, T.J.R., Johan, Z.: A new finite element formulation for computational fluid dynamics: X. The compressible Euler and Navier–Stokes equations. In: Computer Methods in Applied Mechanics and Engineering, pp. 141–219 (1991)
57. Shu, C.W.: Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In: Advanced Numerical Approximation of Nonlinear Hyperbolic Equations, pp. 325–432. Springer (1998)
58. Shu, C.W., Osher, S.: Efficient implementation of essentially non-oscillatory shock-capturing schemes. J. Comput. Phys. **77**(2), 439–471 (1988)
59. Shu, C.W., Osher, S.: Efficient implementation of essentially non-oscillatory shock-capturing schemes: II. In: Upwind and High-Resolution Schemes, pp. 328–374. Springer (1989)
60. Spiegel, S.C., Huynh, H.T., DeBonis, J.R.: A survey of the isentropic Euler vortex problem using high-order methods. In: 22nd AIAA Computational Fluid Dynamics Conference (2015)
61. Strang, G.: Introduction to Linear Algebra. Wellesley-Cambridge Press Wellesley, Wellesley (2016)
62. Svärd, M.: Weak solutions and convergent numerical schemes of modified compressible Navier-Stokes equations. J. Comput. Phys. **288**, 19–51 (2015)
63. Svärd, M., Özcan, H.: Entropy-stable schemes for the Euler equations with far-field and wall boundary conditions. J. Sci. Comput. **58**(1), 61–89 (2014)
64. Tadmor, E.: The numerical viscosity of entropy stable schemes for systems of conservation laws. I. Math. Comput. **49**(179), 91–103 (1987)
65. Tadmor, E.: Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems. Acta Numerica **12**, 451–512 (2003)
66. Toro, E.F.: Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction. Springer, Berlin (2013)
67. Wang, Z.J.: Adaptive High-Order Methods in Computational Fluid Dynamics, vol. 2. World Scientific, Singapore (2011)
68. Wang, Z.J., Liu, Y., May, G., Jameson, A.: Spectral difference method for unstructured grids II: extension to the Euler equations. J. Sci. Comput. **32**(1), 45–71 (2007)
69. Williams, D.: An entropy stable, hybridizable discontinuous Galerkin method for the compressible Navier-Stokes equations. Math. Comput. **87**(309), 95–121 (2018)
70. Zitelli, J., Muga, I., Demkowicz, L., Gopalakrishnan, J., Pardo, D., Calo, V.M.: A class of discontinuous Petrov–Galerkin methods. Part IV: the optimal test norm and time-harmonic wave propagation in 1D. J. Comput. Phys. **230**(7), 2406–2432 (2011)