Entropy-Stable Summation-By-Parts Discretization of the Euler Equations on General Curved Elements[☆]

Jared Crean[a,3,*], Jason E. Hicken[a,1], David C. Del Rey Fernández[d,4], David W. Zingg[b,2], Mark H. Carpenter[d,5]

[a] *Department of Mechanical, Aerospace, and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, New York, United States*
[b] *University of Toronto Institute for Aerospace Studies, Toronto, Canada*
[c] *National Institute of Aerospace, Hampton, Virgina, United States*
[d] *Computational AeroSciences Branch, NASA Langley Research Center, Hampton, Virgina, United States*

Abstract

We present and analyze an entropy-stable semi-discretization of the Euler equations based on high-order summation-by-parts (SBP) operators. In particular, we consider general multidimensional SBP elements, building on and generalizing previous work with tensor-product discretizations. In the absence of dissipation, we prove that the semi-discrete scheme conserves entropy; significantly, this proof of nonlinear L^2 stability does not rely on integral exactness. Furthermore, interior penalties can be incorporated into the discretization to ensure that the total (mathematical) entropy decreases monotonically, producing an entropy-stable scheme. SBP discretizations with curved elements remain accurate, conservative, and entropy stable provided the mapping Jacobian satisfies the discrete metric invariants; polynomial mappings at most one degree higher than the SBP operators automatically satisfy the metric invariants in two dimensions. In three-dimensions, we describe an elementwise optimization that leads to suitable Jacobians in the case of polynomial mappings. The properties of the semi-discrete scheme are verified and investigated using numerical experiments.

Keywords: nonlinear entropy stability, summation-by-parts, simultaneous approximation terms, high-order discretizations, general elements, curved elements, unstructured grid

[☆] J. Crean was supported by the National Science Foundation under Grant No. 1554253, and J. Hicken was partially funded by the Air Force Office of Scientific Research Award FA9550-15-1-0242 under Dr. Jean-Luc Cambier. The authors gratefully acknowledge this support.

[*] corresponding author

Email addresses: jcrean01@gmail.com (Jared Crean), hickej2@rpi.edu (Jason E. Hicken), dcdelrey@gmail.com (David C. Del Rey Fernández), dwz@oddjob.utias.utoronto.ca (David W. Zingg), mark.h.carpenter@nasa.gov (Mark H. Carpenter)

[1] Assistant Professor
[2] Professor
[3] Graduate Student
[4] Postdoctoral Fellow
[5] Senior Research Scientist

1. Introduction

Scientists and engineers increasingly rely on computational fluid dynamics (CFD) for both academic and industrial applications. To be useful to these practitioners, a CFD algorithm must be robust, geometrically flexible, and accurate. These three requirements are met by first- and second-order accurate discretizations that use unstructured grids, provided a sufficiently fine mesh is adopted. This explains, in part, why most commercial CFD software is based on low-order discretizations.

However, the caveat of "a sufficiently fine mesh" can be a significant burden in practice. For a given computational budget, the user wants the most accurate solution possible; that is, they want efficient algorithms. There are two common approaches to improving the efficiency of CFD workflows: mesh adaptation and high-order discretizations. The present work focuses on the latter, although these methods are not mutually exclusive and are often complimentary.

Constructing a robust high-order CFD discretization that can accommodate complex geometries is not straightforward. The most obvious candidate is the finite-element (FE) method, which can be implemented on unstructured grids using either continuous or discontinuous high-order polynomial basis functions. In addition, for linear partial-differential equations (PDEs), many FE discretizations are provably stable, and, therefore, robust.

A potential drawback of FE methods is that proving stability can be difficult for nonlinear PDEs, like the Euler or Navier-Stokes equations. Nonlinear FE stability proofs typically rely on exact integration of the rational functions that appear in the compressible Euler and Navier-Stokes equations. For example, Hughes, Franca, and Mallet [1] presented a FE discretization of the Navier-Stokes equations that satisfies the second law of thermodynamics and is, therefore, stable, provided the integration is exact; see, also, the work of Barth [2]. This integration is impractical for low polynomial order $p < 2$, and impossible for arbitrary orders using standard cubature rules. In practice, inexact cubatures are used in the FE method, which leads to semi-discrete schemes that are not provably stable, in general.

To guarantee robustness, the use of inexact cubature must be accounted for in the stability analysis. One way to do this is to build the cubature rule into the discretization from the beginning, as is done in collocation FE methods like the Legendre-Gauss spectral-element method. More generally, the discretization can use summation-by-parts (SBP) operators [3, 4], which are high-order finite-difference (FD) operators that mimic integration by parts discretely. SBP methods can be regarded as collocation FE methods that do not necessarily have an underlying basis.

Recently, Fisher and Carpenter [5] combined SBP FD operators with the numerical flux of Ismail and Roe [6] to produce a high-order, entropy-conservative, semi-discrete scheme. By introducing appropriate dissipation, they were able to produce an entropy-stable scheme. Their approach was subsequently extended to tensor-product spectral-element methods that obey the SBP property [7]. In both cases, the essential ingredients necessary for nonlinear stability were the use of a diagonal mass matrix, the SBP property, and an entropy-conservative flux.

A potential criticism of traditional finite-difference discretizations, including SBP methods, is that they are limited to tensor-product elements. To address this, Hicken, Del Rey Fernández, and Zingg [8] proposed a multidimensional SBP definition for more general domains such as triangles and tetrahedra. The primary objective of this work is to synthesize this multidimensional SBP methodology with the entropy-stable framework of [5]. We believe this offers an attractive path toward an efficient, robust, and geometrically flexible CFD solver. In addition to the primary objective, this work makes the following contributions:

- We show that any high-order FD operator, not just SBP, can be combined with a symmetric, dyadic flux function to approximate the derivative of the flux; this generalizes a previous result in [7] that only applied to Tadmor's flux [9].
- We clarify how to construct entropy-conservative interior penalties, i.e. simultaneous approximation terms [10–12], for multidimensional SBP operators, including those operators that do not have nodes on the interface.
- We present a method for (implicitly) constructing SBP operators on high-order curvilinear elements while maintaining the properties necessary for accuracy, conservation and stability.

The paper is organized as follows. After introducing notation and definitions, Section 2 reviews the properties of the continuous equations the discretization will mimic. Section 3 presents our generic multidimensional SBP semi-discretization. In Section 4, we prove that the discretization is accurate, conservative (with respect to the conservative variables), and entropy conservative. Our approach to handling curvilinear elements is described in Section 5. Numerical experiments are provided in Section 6 to verify the theory. Section 7 presents our conclusions and future work.

2. Preliminaries

This section introduces our notation, and it reviews the continuous equations and their entropy properties. In developing the discrete equations, we will refer to these properties extensively.

2.1. Notation and definitions

Let $\Omega \subset \mathbb{R}^2$ denote a compact, connected set in the plane. The boundary of Ω will be denoted by $\partial\Omega$, and we assume $\partial\Omega$ is piecewise smooth. Consider a partition of Ω into a set of K non-overlapping elements. The domain of the element with index κ is denoted by Ω_κ and its boundary is denoted by $\partial\Omega_\kappa$. Thus, $\Omega = \bigcup_{\kappa=1}^{K} \Omega_\kappa$. We will also assume that $\partial\Omega_\kappa$ is piecewise smooth for all elements κ.

The set of interfaces between elements is denoted Γ^I, and this set is formally defined by

$$\Gamma^I = \{\partial\Omega_\kappa \cap \partial\Omega_\nu \mid \kappa, \nu = 1, \dots, K,\ \kappa \neq \nu\}\,.$$

In order to perform numerical integration over Γ^I and $\partial\Omega_\kappa$, it is helpful to partition these sets into smooth subsets. For example, the notations

$$\sum_{\gamma \subset \Gamma^I} (\cdot) \qquad \text{and} \qquad \sum_{\gamma \subset \partial\Omega_\kappa} (\cdot)$$

are used to indicate a sum over all smooth subsets of Γ^I and $\partial\Omega_\kappa$, respectively. In practice, γ is an individual (smooth) face in a mesh for which a cubature rule can be defined.

Functions on a particular domain are represented using a script type, and bold font is used for vector-valued functions. For example, $\mathcal{U} \in L^2(\Omega)$ is a square integrable function on Ω, while $\boldsymbol{\mathcal{F}} \in [H^1(\Omega_\kappa)]^4$ is a 4-vector-valued function whose components have square integrable derivatives on the element κ. We will reserve $\mathcal{P}, \mathcal{Q} \in \mathbb{P}_d(\Omega)$ for polynomials, where $\mathbb{P}_d(\Omega)$ is the space of polynomials of total degree d or less on Ω. These polynomials will be used to define the properties of the SBP operators and prove the properties of the discretization.

We discretize functions using their nodal values at points within the elements. Suppose element κ has n_κ nodes given by $S_\kappa = \{(x_i, y_i)\}_{i=1}^{n_\kappa}$. Then the function $\mathcal{U} \in L^2(\Omega_\kappa)$ evaluated at the nodes is represented using the column vector

$$\boldsymbol{u} = \begin{bmatrix} \mathcal{U}(x_1, y_1) & \mathcal{U}(x_2, y_2) & \cdots & \mathcal{U}(x_n, y_n) \end{bmatrix}^T \in \mathbb{R}^{n_\kappa}.$$

We will reserve $\boldsymbol{p}$ and $\boldsymbol{q}$ for generic polynomials $\mathcal{P}$ and $\mathcal{Q}$ evaluated at the points in S_κ. The symbols $\mathbf{1}$ and $\mathbf{0}$ denote vectors consisting of all ones and all zeros, respectively. The number of entries in $\mathbf{1}$ and $\mathbf{0}$ can always be inferred from the context. When discussing the accuracy of the discretization, it will be convenient to define the nominal element size $h = \max_{\boldsymbol{x}_i, \boldsymbol{x}_j \in S_\kappa} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$. To facilitate operating on the faces of the elements, we also define the face nodes $S_\gamma = \{(x_j, y_j)\}_{j=1}^{n_\gamma} \subset \gamma$.

Matrices are represented with an uppercase sans-serif type, for example $\mathsf{A} \in \mathbb{R}^{n \times m}$. The $n \times n$ identity matrix is represented by I_n. The Kronecker product of two matrices $\mathsf{A} \in \mathbb{R}^{n \times m}$ and $\mathsf{B} \in \mathbb{R}^{p \times q}$ is the matrix $\mathsf{A} \otimes \mathsf{B} \in \mathbb{R}^{np \times mq}$ defined by

$$\mathsf{A} \otimes \mathsf{B} = \begin{bmatrix} a_{11}\mathsf{B} & a_{12}\mathsf{B} & \cdots & a_{1m}\mathsf{B} \\ a_{21}\mathsf{B} & a_{22}\mathsf{B} & \cdots & a_{2m}\mathsf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}\mathsf{B} & a_{n2}\mathsf{B} & \cdots & a_{nm}\mathsf{B} \end{bmatrix}.$$

Two properties of Kronecker products that we will use are that $(\mathsf{A} \otimes \mathsf{B})(\mathsf{C} \otimes \mathsf{D}) = (\mathsf{AC}) \otimes (\mathsf{BD})$ and $(\mathsf{A} \otimes \mathsf{B})^T = \mathsf{A}^T \otimes \mathsf{B}^T$.

The Hadamard, or entrywise, product of two conforming matrices $\mathsf{A}, \mathsf{B} \in \mathbb{R}^{n \times m}$ is the matrix $\mathsf{A} \circ \mathsf{B} \in \mathbb{R}^{n \times m}$ whose entries are defined by $(\mathsf{A} \circ \mathsf{B})_{ij} = (\mathsf{A})_{ij} (\mathsf{B})_{ij}$. The Hadamard product is commutative, associative, and distributive over addition. Furthermore, we will need the following Lemma for the subsequent analysis; we omit the proof, which follows from the definitions of matrix multiplication and the Hadamard product.

Lemma 1. *If $\mathsf{A} \in \mathbb{R}^{n \times n}$ is a diagonal matrix, and $\mathsf{B}, \mathsf{C} \in \mathbb{R}^{n \times m}$, then*

$$\mathsf{A}\,(\mathsf{B} \circ \mathsf{C}) = (\mathsf{AB}) \circ \mathsf{C} = \mathsf{B} \circ (\mathsf{AC})\,.$$

2.2. The Euler equations and the entropy equation

The differential form of the two-dimensional Euler equations is

$$\frac{\partial \boldsymbol{\mathcal{U}}}{\partial t} + \frac{\partial \boldsymbol{\mathcal{F}}_x}{\partial x} + \frac{\partial \boldsymbol{\mathcal{F}}_y}{\partial y} = \mathbf{0}, \qquad \forall\, \boldsymbol{x} \in \Omega, \tag{1}$$

where $\boldsymbol{\mathcal{U}} = [\rho, \rho u, \rho v, e]^T$ denotes the conservative variables and the flux vectors are

$$\boldsymbol{\mathcal{F}}_x = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ \rho u v \\ (e + p)u \end{bmatrix}, \qquad \boldsymbol{\mathcal{F}}_y = \begin{bmatrix} \rho v \\ \rho v u \\ \rho v^2 + p \\ (e + p)v \end{bmatrix}.$$

The pressure, which appears in the flux vectors above, is defined by the callorically perfect ideal gas law as $p = (\gamma - 1)[e - \frac{\rho}{2}(u^2 + v^2)]$, where the heat capacity ratio is assumed to be $\gamma = 1.4$.

Remark 1. The Euler equations must be supplemented with suitable boundary conditions, in general. For this work we consider only periodic boundary conditions. Determining nonlinearly stable boundary conditions for more general types of boundaries remains an active area of research.

In addition to conserving mass, momentum, and energy, the Euler equations also conserve entropy in the absence of shocks [13, 14]. More generally, the entropy should be monotonically decreasing[6] in accordance with the second-law of thermodynamics. Our objective is to design a high-order discretization that respects these properties because, as described in Section 2.2, this can be used to prove nonlinear stability.

We briefly review the proof that (1) conserves entropy, following [15], since it helps elucidate the semi-discrete analysis presented later. For concreteness, we define the (mathematical) entropy to be $\mathcal{S} \equiv -\rho s/(\gamma-1)$, where $s = \ln(p/\rho^\gamma)$ is the physical entropy; other choices of entropy function are possible, but this particular choice is the only one that can be extended to the Navier-Stokes equations [1]. Whenever we refer to the entropy in the following text, we are referring to mathematical entropy, $\mathcal{S}$, as defined above.

Based on the above definition of entropy, we obtain the entropy variables

$$\boldsymbol{\mathcal{W}} \equiv \frac{\partial \mathcal{S}}{\partial \boldsymbol{\mathcal{U}}} = \begin{bmatrix} \frac{\gamma-s}{\gamma-1} - \frac{1}{2}\frac{\rho}{p}(u^2+v^2), & \frac{\rho u}{p}, & \frac{\rho v}{p}, & -\frac{\rho}{p} \end{bmatrix}^T.$$

Next, we define the entropy fluxes, $\mathcal{G}_x$ and $\mathcal{G}_y$, using the following differential relations.

$$\boldsymbol{\mathcal{W}}^T \frac{\partial \boldsymbol{\mathcal{F}}_x}{\partial \boldsymbol{\mathcal{U}}} \equiv \frac{\partial \mathcal{G}_x}{\partial \boldsymbol{\mathcal{U}}}, \qquad \text{and} \qquad \boldsymbol{\mathcal{W}}^T \frac{\partial \boldsymbol{\mathcal{F}}_y}{\partial \boldsymbol{\mathcal{U}}} \equiv \frac{\partial \mathcal{G}_y}{\partial \boldsymbol{\mathcal{U}}}.$$

Entropy conservation now follows for smooth solutions by contracting the entropy variables with the Euler equations and integrating over the domain Ω:

$$\begin{aligned}
& \int_\Omega \boldsymbol{\mathcal{W}}^T \left[\frac{\partial \boldsymbol{\mathcal{U}}}{\partial t} + \frac{\partial \boldsymbol{\mathcal{F}}_x}{\partial x} + \frac{\partial \boldsymbol{\mathcal{F}}_y}{\partial y} \right] d\Omega = 0, \\
\Rightarrow \quad & \int_\Omega \frac{\partial \mathcal{S}}{\partial \boldsymbol{\mathcal{U}}} \left[\frac{\partial \boldsymbol{\mathcal{U}}}{\partial t} + \frac{\partial \boldsymbol{\mathcal{F}}_x}{\partial \boldsymbol{\mathcal{U}}}\frac{\partial \boldsymbol{\mathcal{U}}}{\partial x} + \frac{\partial \boldsymbol{\mathcal{F}}_y}{\partial \boldsymbol{\mathcal{U}}}\frac{\partial \boldsymbol{\mathcal{U}}}{\partial y} \right] d\Omega = 0, \\
\Rightarrow \quad & \int_\Omega \frac{\partial \mathcal{S}}{\partial t} + \frac{\partial \mathcal{G}_x}{\partial x} + \frac{\partial \mathcal{G}_y}{\partial y} \, d\Omega = 0, \\
\Rightarrow \quad & \frac{d}{dt}\int_\Omega \mathcal{S} \, d\Omega + \int_{\partial\Omega} (\mathcal{G}_x n_x + \mathcal{G}_y n_y) \, d\Gamma = 0.
\end{aligned} \tag{2}$$

We reiterate that (2) holds only for smooth solutions. If the solutions are discontinuous, then the left hand side should be less than zero [13]. Before proceeding, we list a few important definitions and a property we will use later. The entropy flux in the x direction is $\mathcal{G}_x = -\rho u \mathcal{S}$, and, similarly, the y direction flux is given by substituting v for u [15]. The Hessian $\frac{\partial^2 \mathcal{S}}{\partial \boldsymbol{\mathcal{U}}^2}$ is positive definite and symmetric, therefore its eigenvalues are real [1]. In addition the potential flux will play an important role in the semi-discrete analysis; the potential flux in the x and y directions are $\psi_x = \boldsymbol{\mathcal{W}}^T \boldsymbol{\mathcal{F}}_x - \mathcal{G}_x$ and $\psi_y = \boldsymbol{\mathcal{W}}^T \boldsymbol{\mathcal{F}}_y - \mathcal{G}_y$, respectively[16].

[6] Here we adopt the mathematical sign convention for entropy

Remark 2. A primary motivation for seeking entropy-conservative and entropy-stable schemes is based on the work in [17], where Dafermos proved that bounding the entropy function $\mathcal{S}$ can be used to bound the norm of the solution itself, provided density and pressure remain positive. For completeness, we quote the result from Svärd [18], who extended the proof to systems that contain mass diffusion:

$$\int \boldsymbol{\mathcal{U}}^T \boldsymbol{\mathcal{U}} d\Omega \leq 2\frac{\mathcal{C}}{\mathcal{S}''_{min}} + \int \boldsymbol{\mathcal{U}}_0^T \boldsymbol{\mathcal{U}}_0 d\Omega, \tag{3}$$

where $\mathcal{S}''_{min}$ is the minimum eigenvalue of the Hessian of the entropy function, $\boldsymbol{\mathcal{U}}_0$ is the initial state, and $\mathcal{C}$ is a constant that depends on $\boldsymbol{\mathcal{U}}_0$ and the boundary conditions.

3. SBP-SAT discretization of the Euler equations

In this section and the next we restrict our focus to two-dimensional discretizations to keep the presentation concise; however, the results generalize to three-dimensional discretizations as the results in Section 6 attest. There are some differences that arise on curvilinear elements, and these are discussed in Section 5.

3.1. Multidimensional SBP operators

We rely on the abstract definition of multidimensional SBP operators proposed in [8]. The definition for an SBP approximation of $\partial/\partial x$ is given below for completeness. The operator in the y direction is defined similarly.

Definition 1. Two-dimensional summation-by-parts operator: The matrix D_x is a degree p SBP approximation to the first derivative $\frac{\partial}{\partial x}$ on the nodes $S_\kappa = \{(x_i, y_i)\}_{i=1}^{n_\kappa}$ if

1. $\mathsf{D}_x \boldsymbol{p}$ is equal to $\partial \mathcal{P}/\partial x$ at the nodes S_κ, for all polynomials $\mathcal{P} \in \mathbb{P}_p(\Omega_\kappa)$;
2. $\mathsf{D}_x = \mathsf{H}^{-1}\mathsf{Q}_x$, where H is symmetric positive-definite, and;
3. $\mathsf{Q}_x = \mathsf{S}_x + \frac{1}{2}\mathsf{E}_x$, where $\mathsf{S}_x^T = -\mathsf{S}_x$, $\mathsf{E}_x^T = \mathsf{E}_x$, and E_x satisfies

$$\boldsymbol{p}^T \mathsf{E}_x \boldsymbol{q} = \int_{\partial\Omega_\kappa} \mathcal{P}\, \mathcal{Q}\, n_x \mathrm{d}\Gamma,$$

for all polynomials $\mathcal{P}, \mathcal{Q} \in \mathbb{P}_r(\Omega_\kappa)$, where $r \geq p$. In the above integral, n_x is the x component of $\boldsymbol{n} = [n_x, n_y]^T$, the outward pointing unit normal on $\partial\Omega_\kappa$.

Remark 3. Despite the appearance of polynomials in Definition 1, the solution of an SBP-based discretization is not defined in terms of a polynomial basis, in general. Polynomials are used only to define the accuracy conditions, and there are usually more nodes than basis functions for a given total degree basis.

Remark 4. This work is concerned exclusively with diagonal-norm SBP operators, for which H is a diagonal matrix with strictly positive entries. In this case, the nodes S_κ and entries in H define a cubature that is exact for polynomials of degree $2p-1$, at least [8].

The matrices that appear in the SBP definition are closely related to integral bilinear forms; see [8] for the case of multidimensional SBP operators. We list the relationships here for completeness and for the benefit of readers more familiar with finite-element discretizations. The diagonal matrix H_κ can be interpreted as a (lumped) mass matrix and can be used to approximate an inner product of two functions:

$$\boldsymbol{p}^T \mathsf{H} \boldsymbol{q} = \int_{\Omega_\kappa} \mathcal{P}\mathcal{Q} \mathrm{d}\Omega + \mathrm{O}(h^{2p}).$$

Furthermore, for the matrix Q_x we have that

$$\boldsymbol{p}^T \mathsf{Q}_x \boldsymbol{q} = \int_{\Omega_\kappa} \mathcal{P} \frac{\partial \mathcal{Q}}{\partial x} \mathrm{d}\Omega + O(h^{\min(r+1,2p)})$$

and

$$\boldsymbol{p}^T \mathsf{Q}_x^T \boldsymbol{q} = \int_{\Omega_\kappa} \frac{\partial \mathcal{P}}{\partial x} \mathcal{Q} \mathrm{d}\Omega + O(h^{\min(r+1,2p)})$$

can be used to discretize first derivative terms of the variational form of the PDE. We can use these operators to mimic integration-by-parts; hence the name summation-by-parts. To begin, notice that $\mathsf{Q}_x + \mathsf{Q}_x^T = (\mathsf{S}_x + \frac{1}{2}\mathsf{E}_x) + (\mathsf{S}_x + \frac{1}{2}\mathsf{E}_x)^T$. Using the skew-symmetry of S_x and solving for Q_x, we have $\mathsf{Q}_x = -\mathsf{Q}_x^T + \mathsf{E}_x$. Comparing with the integral forms above, we can see this is integration-by-parts:

$$\underbrace{\boldsymbol{p}^T \mathsf{Q}_x \boldsymbol{q}}_{\int_{\Omega_\kappa} \mathcal{P} \frac{\partial \mathcal{Q}}{\partial x} \mathrm{d}\Omega} = \underbrace{-\boldsymbol{p}^T \mathsf{Q}_x^T \boldsymbol{q}}_{-\int_{\Omega_\kappa} \frac{\partial \mathcal{P}}{\partial x} \mathcal{Q} \mathrm{d}\Omega} + \underbrace{\boldsymbol{p}^T \mathsf{E}_x \boldsymbol{q}}_{\int_{\partial\Omega_\kappa} \mathcal{P}\, \mathcal{Q}\, n_x \mathrm{d}\Gamma} .$$

In order to discretize systems of partial differential equations, like the Euler equations, it will be convenient to define matrices that operate on all variables simultaneously. This is denoted using an overbar to implicitly insert a Kronecker product, for example, $\overline{\mathsf{D}}_x \equiv \mathsf{D}_x \otimes \mathsf{I}_4$. This requires the variables in the solution vector to be grouped by node, e.g. the density, momenta, and energy at the first node occupy the first four entries, then the variables for the second node occupy the next four, etc.

3.2. Matrices for face-based operations

We couple adjacent elements using interior penalty terms. The penalties are usually called simultaneous-approximation terms (SATs) in the SBP literature [10–12, 19]. In the classical SBP setting, that is, using tensor product operators with nodes on the boundary, SATs are applied pointwise between coincident nodes of adjacent elements/blocks.

A pointwise implementation of SATs is not (directly) possible for generalized one-dimensional SBP operators [20] and multidimensional SBP operators, since these operators may not have nodes on $\partial\Omega_\kappa$. To address this issue, we follow the approach presented in [21, 22], and use interpolation/extrapolation operators to reconstruct the solution from the SBP nodes to nodes on the element's boundary.

Remark 5. It is possible to construct multidimensional SBP operators where the face nodes are collocated with the volume nodes, making the interpolation operators, and thus the boundary operator E_x, diagonal. Herein, we consider the general case where E_x is a dense matrix.

Under the moderate assumptions described in [21], we can construct a matrix $\mathsf{R}_{\gamma\kappa} \in \mathbb{R}^{n_\gamma \times n_\kappa}$ that is a degree r interpolation/extrapolation operator from the volume nodes $S_\kappa = \{(x_i, y_i)\}_{i=1}^{n_\kappa}$ to the face nodes $S_\gamma = \{(x_j, y_j)\}_{j=1}^{n_\gamma}$, that is

$$\left(\mathsf{R}_{\gamma\kappa}\boldsymbol{p}\right)_j = \sum_{i=1}^{n_\kappa} (\mathsf{R}_{\gamma\kappa})_{ji} \mathcal{P}(x_i, y_i) = \mathcal{P}(x_j, y_j), \qquad \forall\, j = 1, \ldots, n_\gamma,$$

and for all polynomials $\mathcal{P} \in \mathbb{P}_r(\Omega_\kappa)$, where S_γ supports a cubature rule with positive weights.

In [21] it was shown that the E_x operator for an element κ can be decomposed into the sum of $\mathsf{E}_x^{\gamma\kappa}$ operators, one for each face γ, using the interpolation/extrapolation operators. Specifically,

$$\mathsf{E}_x = \sum_{\gamma \subset \partial\Omega_\kappa} \mathsf{E}_x^{\gamma\kappa}, \qquad \text{where} \qquad \mathsf{E}_x^{\gamma\kappa} \equiv \mathsf{R}_{\gamma\kappa}^T \mathsf{N}_{x,\gamma} \mathsf{B}_\gamma \mathsf{R}_{\gamma\kappa}, \tag{4}$$

$\mathsf{B}_\gamma = \operatorname{diag}(b_1, b_2, \ldots, b_{n_\gamma})$ is a diagonal matrix whose entries are the (positive) cubature weights for face γ, and $\mathsf{N}_{x,\gamma} = \operatorname{diag}\left[(n_x)_1, (n_x)_2, \ldots, (n_x)_{n_\gamma}\right]$ is a diagonal matrix whose entries are the x component of the unit outward normal to $\partial\Omega_\kappa$ at the cubature points of face γ. If the face cubature rule is degree $2p$ or greater, there exists at least one SBP operator whose E_x has the decomposition (4) for a given S_κ and H [21], and we assume that such an SBP operator is used throughout this work.

To keep the subsequent presentation compact, we also use the interpolation/extrapolation operators to define the following matrices, which play a role similar to $\mathsf{E}_x^{\gamma\kappa}$ and $\mathsf{E}_y^{\gamma\kappa}$, but are skew-symmetric with respect to adjacent elements κ and ν:

$$\begin{aligned}
\mathsf{E}_x^{\kappa\nu} &\equiv \mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \mathsf{N}_{x,\gamma} \mathsf{R}_{\gamma\nu} = -(\mathsf{E}_x^{\nu\kappa})^T, \\
\mathsf{E}_y^{\kappa\nu} &\equiv \mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \mathsf{N}_{y,\gamma} \mathsf{R}_{\gamma\nu} = -(\mathsf{E}_y^{\nu\kappa})^T.
\end{aligned}$$

These operators apply the discrete test function corresponding to one element's nodes to the adjacent element's trial function interpolated to their common interface. In Section 4, the skew-symmetry property will be essential in proving entropy conservation.

Remark 6. We will assume that the normal is outward pointing with respect to κ in the adjacent pair of generic elements κ and ν.

3.3. Strong-form discretization

In order to describe our entropy-conservative discretization of (1), we must introduce dyadic flux functions with specific properties. Let $\boldsymbol{\mathcal{F}}_x^\star(\cdot,\cdot) : \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}^4$ denote a numerical flux function that is symmetric in its two arguments, $\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{U}}') = \boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{\mathcal{U}}',\boldsymbol{\mathcal{U}})$, and consistent with the continuous flux, $\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{U}}) = \boldsymbol{\mathcal{F}}_x(\boldsymbol{\mathcal{U}})$. The function $\boldsymbol{\mathcal{F}}_y^\star$ is required to satisfy analogous properties.

In addition, we require the numerical fluxes $\boldsymbol{\mathcal{F}}_x^\star$ and $\boldsymbol{\mathcal{F}}_y^\star$ to satisfy the following entropy consistency conditions: if $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{U}}'$ are two states, and $\boldsymbol{\mathcal{W}}$ and $\boldsymbol{\mathcal{W}}'$ are their corresponding entropy variables, then

$$\begin{aligned}
\left(\boldsymbol{\mathcal{W}} - \boldsymbol{\mathcal{W}}'\right)^T \boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{U}}') &= \psi_x - \psi_x', \\
\left(\boldsymbol{\mathcal{W}} - \boldsymbol{\mathcal{W}}'\right)^T \boldsymbol{\mathcal{F}}_y^\star(\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{U}}') &= \psi_y - \psi_y',
\end{aligned} \tag{5}$$

where ψ_x and ψ_y are the potential fluxes reviewed in Section 2.2. .

Remark 7. The definition of the potential flux was first given by Tadmor [16]; in the same paper he also defined the first numerical flux function to satisfy the entropy consistency condition. More recently, Ismail and Roe [6] developed a less expensive flux that satisfies the entropy consistency condition. Subsequently, Chandrashekar [23] developed a numerical flux that satisfies the entropy consistency condition and preserves kinetic energy. Many commonly used numerical flux functions, such as Roe's [24], do not satisfy the entropy consistency condition, and, hence, are not suitable for constructing entropy conservative schemes.

Let $\boldsymbol{u}_\kappa \in \mathbb{R}^{4n_\kappa}$ denote the solution of the discretized Euler equations on element κ. We assume that the conservative variables are ordered in $\boldsymbol{u}_\kappa$ such that the value at the ith node is given by

$$\boldsymbol{u}_{\kappa,i} \equiv [\boldsymbol{u}_\kappa]_{4(i-1)+1:4} = [(\rho)_i, (\rho u)_i, (\rho v)_i, (\rho e)_i]^T .$$

When necessary, we will use $\boldsymbol{u}_h \in \mathbb{R}^{4\sum n_\kappa}$ to represent the global discrete solution, which is the concatenation of all the $\boldsymbol{u}_\kappa$ vectors. Each element has a distinct set of unknowns, even for those nodes that may be coincident. Hence, this spectral collocation scheme allows the solution to be discontinuous at element interfaces.

Based on the ordering of unknowns in $\boldsymbol{u}_\kappa$, the strong form of the SBP-SAT discretization of the Euler equations on element κ is given by

$$\frac{\mathrm{d}\boldsymbol{u}_\kappa}{\mathrm{d}t} + \left[\overline{\mathsf{D}}_x \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1} + \left[\overline{\mathsf{D}}_y \circ \mathsf{F}_y(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1} = \overline{\mathsf{H}}_\kappa^{-1} \boldsymbol{r}_\kappa(\boldsymbol{u}_h). \tag{6}$$

recalling the overbar notation for matrices $\overline{\mathsf{D}}_x \equiv \mathsf{D}_x \otimes \mathsf{I}_4$ introduced previously. The remainder of this section is devoted to describing the operators F_x, F_y and $\boldsymbol{r}_\kappa$.

In general, $\mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)$ is the $4n_\kappa \times 4n_\nu$ matrix defined by

$$\mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) \equiv 2 \begin{bmatrix} \mathrm{diag}\left[\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,1}, \boldsymbol{u}_{\nu,1})\right] & \dots & \mathrm{diag}\left[\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,1}, \boldsymbol{u}_{\nu,n_\nu})\right] \\ \vdots & \ddots & \vdots \\ \mathrm{diag}\left[\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,n_\kappa}, \boldsymbol{u}_{\nu,1})\right] & \dots & \mathrm{diag}\left[\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,n_\kappa}, \boldsymbol{u}_{\nu,n_\nu})\right] \end{bmatrix}.$$

The matrix F_y is defined similarly. In words, F_x and F_y are block matrices that evaluate (twice) the numerical flux function at every nodal combination of the first and second input. It follows from the symmetry of $\boldsymbol{\mathcal{F}}_x^\star$ that F_x has the properties

$$\mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) = \mathsf{F}_x(\boldsymbol{u}_\nu, \boldsymbol{u}_\kappa)^T \qquad \text{and} \qquad \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa) = \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)^T.$$

Similar properties hold in the y direction. These properties will be used extensively in the analysis. Note that the volume terms in (6) use $\mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)$, however the face terms (described below) use $\mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)$ to couple the elements together.

Remark 8. The operation $\left[\overline{\mathsf{D}}_x \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1}$ in (6), while unusual in appearance, simply amounts to applying the difference operator $\overline{\mathsf{D}}_x$ row-wise to $\mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)$. Furthermore, for each row of F_x one of the two states in the flux function is fixed, a fact that we will use when proving the accuracy of the discretization.

The vector $\boldsymbol{r}_\kappa$, which appears on the right-hand side of (6), constitutes the SAT penalty for element κ. The penalty is defined by

$$\begin{aligned}\boldsymbol{r}_\kappa(\boldsymbol{u}_h) = &\frac{1}{2}\left[\bar{\mathsf{E}}_x \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right]\mathbf{1} + \frac{1}{2}\left[\bar{\mathsf{E}}_y \circ \mathsf{F}_y(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right]\mathbf{1} \\ &- \frac{1}{2}\sum_{\gamma \subset \partial\Omega_\kappa}\left[\bar{\mathsf{E}}_x^{\kappa\nu} \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) + \bar{\mathsf{E}}_y^{\kappa\nu} \circ \mathsf{F}_y(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\right]\mathbf{1},\end{aligned} \tag{7}$$

where $\boldsymbol{u}_\nu \in \mathbb{R}^{4n_\nu}$ denotes the discrete solution for the element adjacent to κ on face γ. The form of the SAT is motivated by the desire to remove the $\bar{\mathsf{E}}_x$ and $\bar{\mathsf{E}}_y$ boundary operators from the weak form after applying integration-by-parts, leaving only the $\bar{\mathsf{E}}_x^{\kappa\nu}$ and $\bar{\mathsf{E}}_y^{\kappa\nu}$ operators, which have the desired skew-symmetry property. We will show that because of the symmetry of $\boldsymbol{\mathcal{F}}_x^\star$ and $\boldsymbol{\mathcal{F}}_y^\star$, (7) is a non-dissipative SAT.

3.4. Weak-form discretization

The equivalent weak form of the SBP-SAT discretization is obtained by left multiplying (6) by $\boldsymbol{v}_\kappa^T \bar{\mathsf{H}}_\kappa$, which is the discrete analog of multiplying the PDE by a test function and integrating over the element. Using part 3 of Definition 1 to decompose $\bar{\mathsf{H}}_\kappa \bar{\mathsf{D}}_x = \bar{\mathsf{Q}}_x$ into its component matrices and rearranging the result with the help of Lemma 1, we arrive at the weak formulation: find $\boldsymbol{u}_h \in \mathbb{R}^{4\sum n_\kappa}$ such that, for all elements $\kappa = 1, 2, \ldots, K$ and all $\boldsymbol{v}_\kappa \in \mathbb{R}^{4n_\kappa}$,

$$\begin{aligned}&\boldsymbol{v}_\kappa^T \bar{\mathsf{H}}_\kappa \frac{\mathrm{d}\boldsymbol{u}_\kappa}{\mathrm{d}t} + \boldsymbol{v}_\kappa^T\left[\bar{\mathsf{S}}_x \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right]\mathbf{1} + \boldsymbol{v}_\kappa^T\left[\bar{\mathsf{S}}_y \circ \mathsf{F}_y(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right]\mathbf{1} \\ &\qquad = -\frac{1}{2}\sum_{\gamma \subset \partial\Omega_\kappa} \boldsymbol{v}_\kappa^T\left[\bar{\mathsf{E}}_x^{\kappa\nu} \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) + \bar{\mathsf{E}}_y^{\kappa\nu} \circ \mathsf{F}_y(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\right]\mathbf{1}.\end{aligned} \tag{8}$$

4. Analysis of accuracy, conservation, and stability

4.1. Accuracy of the SBP-SAT discretization

We divide the accuracy analysis into two theorems. The first theorem, immediately below, concerns the spatial discretization on the left-hand side of (6). The second theorem addresses the SAT penalties on the right-hand side of (6).

Theorem 1. *Let D_x and D_y be any degree p finite-difference approximations of the first derivative operators $\partial/\partial x$ and $\partial/\partial y$, respectively, defined on the node set $S_\kappa = \{(x_i, y_i)\}_{i=1}^{n_\kappa}$. Consider a conservation law whose fluxes in the x and y directions are the continuously differentiable functions $\boldsymbol{\mathcal{F}}_x, \boldsymbol{\mathcal{F}}_y : \mathbb{R}^m \to \mathbb{R}^m$. If $\boldsymbol{\mathcal{F}}_x^\star, \boldsymbol{\mathcal{F}}_y^\star : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ are dyadic functions that are continuously differentiable, symmetric in their arguments, and satisfy $\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{U}}) = \boldsymbol{\mathcal{F}}_x(\boldsymbol{\mathcal{U}})$ and $\boldsymbol{\mathcal{F}}_y^\star(\boldsymbol{\mathcal{U}},\boldsymbol{\mathcal{U}}) = \boldsymbol{\mathcal{F}}_y(\boldsymbol{\mathcal{U}})$, then for sufficiently smooth solutions $\boldsymbol{\mathcal{U}}$*

$$\left\{\left[\bar{\mathsf{D}}_x \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right]\mathbf{1}\right\}_i = \frac{\partial \boldsymbol{\mathcal{F}}_x}{\partial x}(\boldsymbol{u}_{\kappa,i}) + O(h^p), \tag{9}$$

and

$$\left\{\left[\bar{\mathsf{D}}_y \circ \mathsf{F}_y(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right]\mathbf{1}\right\}_i = \frac{\partial \boldsymbol{\mathcal{F}}_y}{\partial y}(\boldsymbol{u}_{\kappa,i}) + O(h^p). \tag{10}$$

PROOF. We will prove (9), since the proof of (10) is conceptually the same.

For row i of F_x appearing in (9), the first argument of $\boldsymbol{\mathcal{F}}_x^\star$ is fixed at $\boldsymbol{\mathcal{U}}(x_i, y_i) = \boldsymbol{u}_{\kappa,i}$, so $\boldsymbol{\mathcal{F}}_x^\star$ is a function of (x, y) via only its second argument. Therefore, in order to prove that the operation defined by (9) is an order p approximation, it is sufficient to show that it is exact for polynomial fluxes $\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i}, \boldsymbol{\mathcal{U}}) \in \mathbb{P}_p(\Omega)$. In this case we have

$$\begin{aligned}
\left\{\left[\overline{\mathsf{D}}_x \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1}\right\}_i &= \sum_{j=1}^{n_\kappa} 2[\mathsf{D}_x]_{ij} \boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i}, \boldsymbol{u}_{\kappa,j}) && \text{(Recall } \boldsymbol{u}_{\kappa,i} \text{ is fixed for each row } i) \\
&= 2\left[\frac{\partial}{\partial x} \boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i}, \boldsymbol{\mathcal{U}}(x,y))\right]_{\boldsymbol{\mathcal{U}}=\boldsymbol{u}_{\kappa,i}} && \text{(since } \boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i}, \boldsymbol{\mathcal{U}}) \in \mathbb{P}_p(\Omega)) \\
&= 2\frac{\partial \boldsymbol{\mathcal{F}}_x^\star}{\partial \boldsymbol{\mathcal{U}}_2}(\boldsymbol{u}_{\kappa,i}, \boldsymbol{u}_{\kappa,i}) \frac{\partial \boldsymbol{\mathcal{U}}}{\partial x}, && \text{(by chain rule)} \qquad (11)
\end{aligned}$$

where $\partial \boldsymbol{\mathcal{F}}_x^\star / \partial \boldsymbol{\mathcal{U}}_2$ is the Jacobian of $\boldsymbol{\mathcal{F}}_x^\star$ with respect to its second argument.

Next, using the symmetry of $\boldsymbol{\mathcal{F}}_x^\star$, it is easy to show that $\partial \boldsymbol{\mathcal{F}}_x^\star / \partial \boldsymbol{\mathcal{U}}_1 = \partial \boldsymbol{\mathcal{F}}_x^\star / \partial \boldsymbol{\mathcal{U}}_2$ whenever its first and second arguments are equal. Furthermore, since $\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{U}}) = \boldsymbol{\mathcal{F}}_x(\boldsymbol{\mathcal{U}})$, we have

$$\begin{aligned}
\frac{\partial \boldsymbol{\mathcal{F}}_x}{\partial \boldsymbol{\mathcal{U}}}(\boldsymbol{u}_{\kappa,i}) = \left[\frac{\partial}{\partial \boldsymbol{\mathcal{U}}} \boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{U}})\right]_{\boldsymbol{\mathcal{U}}=\boldsymbol{u}_{\kappa,i}} &= \left[\frac{\partial \boldsymbol{\mathcal{F}}_x^\star}{\partial \boldsymbol{\mathcal{U}}_1}(\boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{U}}) + \frac{\partial \boldsymbol{\mathcal{F}}_x^\star}{\partial \boldsymbol{\mathcal{U}}_2}(\boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{U}})\right]_{\boldsymbol{\mathcal{U}}=\boldsymbol{u}_{\kappa,i}} \\
&= 2\frac{\partial \boldsymbol{\mathcal{F}}_x^\star}{\partial \boldsymbol{\mathcal{U}}_2}(\boldsymbol{u}_{\kappa,i}, \boldsymbol{u}_{\kappa,i}).
\end{aligned}$$

Substituting this into the third line of (11) gives $(\partial \boldsymbol{\mathcal{F}}_x / \partial \boldsymbol{\mathcal{U}}) \partial \boldsymbol{\mathcal{U}} / \partial x = \partial \boldsymbol{\mathcal{F}}_x / \partial x$, which is the desired result for the degree p polynomial flux. □

Remark 9. A previous result regarding the accuracy of the approximation (9) in [7] was restricted to Tadmor's flux [9], which is generally too expensive to be used in practice. In contrast, Theorem 1 applies to any dyadic flux that is symmetric, differentiable, and consistent. Furthermore, we note that Theorem 1 does not require D_x and D_y to be SBP operators.

We now show that the SAT penalty contributes an error that is no worse, asymptotically, than the left-hand side of the discretization (6).

Theorem 2. *If the fluxes $\boldsymbol{\mathcal{F}}_x^\star, \boldsymbol{\mathcal{F}}_y^\star : \mathbb{R}^m \to \mathbb{R}^m$ and the exact solution $\boldsymbol{\mathcal{U}}$ are sufficiently smooth, then the SAT penalty for element κ, defined by (7), satisfies*

$$\boldsymbol{r}_\kappa(\boldsymbol{u}_h) = O(h_{\max}^r),$$

where $h_{\max}$ is the largest value of $\max_{\boldsymbol{x}_i, \boldsymbol{x}_j \in S_\kappa} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$ over κ and its nearest neighbors and r is the constant from part 3 of Definition 1

PROOF. As in the proof of Theorem 1, it is sufficient to show that $\boldsymbol{r}_\kappa(\boldsymbol{u}_h)$ vanishes for all polynomial fluxes of total degree r, where $r \geq p$. In addition, we will consider only the x component of the SAT on one surface, since the proof for the y component and the remaining surfaces is similar. Thus, we must show that

$$\left[\bar{\mathsf{E}}_x^{\gamma\kappa} \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1} - \left[\bar{\mathsf{E}}_x^{\kappa\nu} \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\right] \mathbf{1} = \mathbf{0}, \qquad (12)$$

for all fluxes $\boldsymbol{\mathcal{F}}_x^\star \in \mathbb{P}_r(\Omega_\kappa \cup \Omega_\nu)$, where we dropped the factor of $1/2$ from $\boldsymbol{r}_\kappa$. Note that the matrix $\bar{\mathsf{E}}_x^{\gamma\kappa} = \mathsf{E}_x^{\gamma\kappa} \otimes \mathsf{I}_4$, where $\mathsf{E}_x^{\gamma\kappa} = \mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \mathsf{N}_{x,\gamma} \mathsf{R}_{\gamma\kappa}$, arises from the decomposition of E_x; see (4).

Now, since the interpolation operators are exact for polynomials of degree r, and $\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i}, \boldsymbol{\mathcal{U}}) \in \mathbb{P}_r(\Omega_\kappa \cup \Omega_\nu)$, we have

$$\begin{aligned}
\left\{ \left[\bar{\mathsf{E}}_x^{\gamma\kappa} \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa) \right] \mathbf{1} \right\}_i &= \sum_{j=1}^{n_\kappa} (\mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \mathsf{N}_{x,\gamma} \mathsf{R}_{\gamma\kappa})_{ij} \boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i}, \boldsymbol{u}_{\kappa,j}) \\
&= \sum_{k=1}^{n_\gamma} (\mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \mathsf{N}_{x,\gamma})_{ik} \sum_{j=1}^{n_\kappa} (\mathsf{R}_{\gamma\kappa})_{kj} \boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i}, \boldsymbol{u}_{\kappa,j}) \\
&= \sum_{k=1}^{n_\gamma} (\mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \mathsf{N}_{x,\gamma})_{ik} \boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i}, \boldsymbol{u}_{\gamma,k}),
\end{aligned} \tag{13}$$

where $i = 1, 2, \ldots, n_\kappa$, and $\boldsymbol{u}_{\gamma,k}$ is the kth node on the interface γ. Similarly, we have

$$\left\{ \left[\bar{\mathsf{E}}_x^{\kappa\nu} \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) \right] \mathbf{1} \right\}_i = \sum_{k=1}^{n_\gamma} (\mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \mathsf{N}_{x,\gamma})_{ik} \boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i}, \boldsymbol{u}_{\gamma,k}), \tag{14}$$

for all $i = 1, 2, \ldots, n_\kappa$. The right-hand sides of (13) and (14) are equal under the assumptions; therefore, (12) is satisfied. □

4.2. Elementwise conservation analysis

Volume integration of the spatial derivatives in the Euler equations results in a surface integral, a well-known consequence of the Gauss-Divergence theorem. This is a property that the SBP-SAT discretization must mimic in order for the scheme to be conservative elementwise. Elementwise conservation, in turn, implies that the scheme approximates the weak-form of the PDE for discontinuous solutions, at least to the order of the element size. This is a necessary ingredient if we eventually want to use the scheme to capture shocks.

Consider a subset of the domain, $\Omega' = \bigcup_{\kappa \in \zeta} \Omega_\kappa \subseteq \Omega$, where $\zeta \subseteq \{1, 2, \ldots, K\}$ is an index set. Let $\partial\Omega'$ denote the boundary of Ω', and let

$$\Gamma' = \{ \partial\Omega_\kappa \cap \partial\Omega_\nu \mid \kappa, \nu \in \zeta, \kappa \neq \nu \}$$

be the interfaces internal to Ω'.

Theorem 3. *For any subset Ω' as defined above, the SBP-SAT discretization (6) (or (8)) is conservative in an elementwise sense. For example, in the case of conservation of mass we have*

$$\frac{\mathrm{d}}{\mathrm{d}t} \sum_{\kappa \in \zeta} \mathbf{1}^T \mathsf{H}_\kappa \boldsymbol{\rho}_\kappa = -\frac{1}{2} \sum_{\gamma \subset \partial\Omega'} \mathbf{1}^T \left[\mathsf{E}_x^{\kappa\nu} \circ \mathsf{F}_x^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) + \mathsf{E}_y^{\kappa\nu} \circ \mathsf{F}_y^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) \right] \mathbf{1},$$

where $\boldsymbol{\rho}_\kappa$ is the density variable at the nodes of element κ, and F_x^ρ (resp. F_y^ρ) are the rows and columns of F_x (resp. F_y) corresponding to the x-direction (resp. y-direction) mass flux.

PROOF. We will prove elementwise conservation of mass, since the proofs for the other conservation laws are similar. The weak-form of the SBP-SAT discretization of conservation of mass is (refer to (8))

$$\boldsymbol{v}_\kappa^T \mathsf{H}_\kappa \frac{\mathrm{d}\boldsymbol{\rho}_\kappa}{\mathrm{d}t} + \boldsymbol{v}_\kappa^T \left[\mathsf{S}_x \circ \mathsf{F}_x^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1} + \boldsymbol{v}_\kappa^T \left[\mathsf{S}_y \circ \mathsf{F}_y^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1} \\ = -\frac{1}{2} \sum_{\gamma \subset \partial\Omega_\kappa} \boldsymbol{v}_\kappa^T \left[\mathsf{E}_x^{\kappa\nu} \circ \mathsf{F}_x^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) + \mathsf{E}_y^{\kappa\nu} \circ \mathsf{F}_y^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\right] \mathbf{1}, \qquad \forall\, \boldsymbol{v}_\kappa \in \mathbb{R}^{n_\kappa}.$$

For conservation of mass, we replace $\boldsymbol{v}_\kappa$ with $\mathbf{1}$ in the above equation. In doing so, the spatial derivative terms on the left vanish, because $\mathsf{S}_x \circ \mathsf{F}_x^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)$ and $\mathsf{S}_y \circ \mathsf{F}_y^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)$ are skew symmetric matrices[7]. Therefore, we are left with

$$\mathbf{1}^T \mathsf{H}_\kappa \frac{\mathrm{d}\boldsymbol{\rho}_\kappa}{\mathrm{d}t} = -\frac{1}{2} \sum_{\gamma \subset \partial\Omega_\kappa} \mathbf{1}^T \left[\mathsf{E}_x^{\kappa\nu} \circ \mathsf{F}_x^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) + \mathsf{E}_y^{\kappa\nu} \circ \mathsf{F}_y^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\right] \mathbf{1}.$$

Next, we sum these element-based equations over all $\kappa \in \zeta$, and we collect the boundary and interface terms in two separate sums on the right-hand side as follows:

$$\sum_{\kappa \in \zeta} \mathbf{1}^T \mathsf{H}_\kappa \frac{\mathrm{d}\boldsymbol{\rho}_\kappa}{\mathrm{d}t} = -\frac{1}{2} \sum_{\gamma \subset \partial\Omega'} \mathbf{1}^T \left[\mathsf{E}_x^{\kappa\nu} \circ \mathsf{F}_x^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) + \mathsf{E}_y^{\kappa\nu} \circ \mathsf{F}_y^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\right] \mathbf{1} \\ - \frac{1}{2} \sum_{\gamma \subset \Gamma'} \mathbf{1}^T \left[\mathsf{E}_x^{\kappa\nu} \circ \mathsf{F}_x^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) + \mathsf{E}_x^{\nu\kappa} \circ \mathsf{F}_x^\rho(\boldsymbol{u}_\nu, \boldsymbol{u}_\kappa)\right] \mathbf{1} \\ - \frac{1}{2} \sum_{\gamma \subset \Gamma'} \mathbf{1}^T \left[\mathsf{E}_y^{\kappa\nu} \circ \mathsf{F}_y^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) + \mathsf{E}_y^{\nu\kappa} \circ \mathsf{F}_y^\rho(\boldsymbol{u}_\nu, \boldsymbol{u}_\kappa)\right] \mathbf{1}. \quad (15)$$

Recall that $\mathsf{E}_x^{\kappa\nu} = -(\mathsf{E}_x^{\nu\kappa})^T$, $\mathsf{E}_y^{\kappa\nu} = -(\mathsf{E}_y^{\nu\kappa})^T$, $\mathsf{F}_x^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) = \mathsf{F}_x^\rho(\boldsymbol{u}_\nu, \boldsymbol{u}_\kappa)^T$ and $\mathsf{F}_y^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) = \mathsf{F}_y^\rho(\boldsymbol{u}_\nu, \boldsymbol{u}_\kappa)^T$. Therefore, using the identity $(\mathsf{A} \circ \mathsf{B})^T = (\mathsf{A}^T \circ \mathsf{B}^T)$, we have that

$$\mathbf{1}^T \left[\mathsf{E}_x^{\kappa\nu} \circ \mathsf{F}_x^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\right] \mathbf{1} = -\mathbf{1}^T \left[\mathsf{E}_x^{\nu\kappa} \circ \mathsf{F}_x^\rho(\boldsymbol{u}_\nu, \boldsymbol{u}_\kappa)\right] \mathbf{1}, \\ \text{and} \qquad \mathbf{1}^T \left[\mathsf{E}_y^{\kappa\nu} \circ \mathsf{F}_y^\rho(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\right] \mathbf{1} = -\mathbf{1}^T \left[\mathsf{E}_y^{\nu\kappa} \circ \mathsf{F}_y^\rho(\boldsymbol{u}_\nu, \boldsymbol{u}_\kappa)\right] \mathbf{1}.$$

Using the above two equations in the right-side of (15), and taking the time derivative outside the sum on the left, we arrive at the desired result. □

Remark 10. Tensor-product SBP schemes can show sub-cell conservation in addition to elementwise conservation [25]. It is not clear if general multi-dimensional SBP schemes can be cast into the the flux-difference form required to show sub-cell conservation.

4.3. Entropy-conservation analysis

We now prove that the SBP-SAT discretization (8) conserves entropy. The main entropy-conservation result depends on the following two lemmas. Lemma 2 concerns the volume terms on the left of (8), and Lemma 3 concerns the boundary terms on the right of (8). Lemma 2 was previously proven in [7], but is included to keep the presentation self-contained.

[7] Recall that a skew symmetric matrix A satisfies $\boldsymbol{x}^T \mathsf{A} \boldsymbol{x} = \boldsymbol{x}^T \mathsf{A}^T \boldsymbol{x} = -\boldsymbol{x}^T \mathsf{A} \boldsymbol{x} = 0$.

Lemma 2. *Let $\boldsymbol{w}_\kappa \in \mathbb{R}^{4n_\kappa}$ denote the entropy variables corresponding to the conservative variables $\boldsymbol{u}_\kappa$, evaluated at the SBP nodes of element κ. If the numerical flux functions $\boldsymbol{\mathcal{F}}_x^\star(\cdot,\cdot)$ and $\boldsymbol{\mathcal{F}}_y^\star(\cdot,\cdot)$ are symmetric and satisfy the entropy-consistency condition (5), then*

$$\boldsymbol{w}_\kappa^T \left[\bar{\mathsf{S}}_x \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1} = -\mathbf{1}^T \mathsf{E}_x \boldsymbol{\psi}_{x,\kappa} \tag{16}$$

and

$$\boldsymbol{w}_\kappa^T \left[\bar{\mathsf{S}}_y \circ \mathsf{F}_y(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1} = -\mathbf{1}^T \mathsf{E}_y \boldsymbol{\psi}_{y,\kappa}, \tag{17}$$

where $\boldsymbol{\psi}_{x,\kappa} \in \mathbb{R}^{n_\kappa}$ and $\boldsymbol{\psi}_{y,\kappa} \in \mathbb{R}^{n_\kappa}$ are the potential fluxes, $\psi_x = \rho u$ and $\psi_y = \rho v$, evaluated at the nodes of element κ.

Proof. $\bar{\mathsf{S}}_x$ is the skew symmetric part of $\bar{\mathsf{Q}}_x$, so $\bar{\mathsf{S}}_x = 1/2(\bar{\mathsf{Q}}_x - \bar{\mathsf{Q}}_x^T)$. Using this, Lemma 1, and the identities $\boldsymbol{x}^T \mathsf{A} \boldsymbol{y} = \boldsymbol{y}^T \mathsf{A}^T \boldsymbol{x}$ and $\boldsymbol{w}_\kappa = \operatorname{diag}(\boldsymbol{w}_\kappa)\mathbf{1}$, we find

$$\begin{aligned}
&\boldsymbol{w}_\kappa^T \left[\bar{\mathsf{S}}_x \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1} \\
&\quad = \frac{1}{2} \mathbf{1}^T \operatorname{diag}(\boldsymbol{w}_\kappa)^T \left[\left(\bar{\mathsf{Q}}_x - \bar{\mathsf{Q}}_x^T\right) \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1} \\
&\quad = \frac{1}{2} \mathbf{1}^T \left[\bar{\mathsf{Q}}_x \circ (\operatorname{diag}(\boldsymbol{w}_\kappa) \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa) - \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa) \operatorname{diag}(\boldsymbol{w}_\kappa)\right] \mathbf{1}.
\end{aligned}$$

To proceed, we make use of the entropy-consistency condition and the definition of F_x (In the following expressions, $\boldsymbol{w}_{\kappa,i}$ and $\psi_{x,i}$ denote the entropy variables and x potential flux, respectively, at node i of element κ):

$$\begin{aligned}
&\boldsymbol{w}_\kappa^T \left[\bar{\mathsf{S}}_x \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa)\right] \mathbf{1} \\
&\quad = \frac{1}{2} \sum_{i=1}^{n_\kappa} \sum_{j=1}^{n_\kappa} 2[\mathsf{Q}_x]_{ij} \mathbf{1}^T \left[\operatorname{diag}(\boldsymbol{w}_{\kappa,i}) - \operatorname{diag}(\boldsymbol{w}_{\kappa,j})\right] \operatorname{diag}\left[\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i}, \boldsymbol{u}_{\kappa,j})\right] \mathbf{1} \\
&\quad = \sum_{i=1}^{n_\kappa} \sum_{j=1}^{n_\kappa} [\mathsf{Q}_x]_{ij} \left(\boldsymbol{w}_{\kappa,i} - \boldsymbol{w}_{\kappa,j}\right)^T \boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i}, \boldsymbol{u}_{\kappa,j}) \\
&\quad = \sum_{i=1}^{n_\kappa} \sum_{j=1}^{n_\kappa} [\mathsf{Q}_x]_{ij} \left(\psi_{x,i} - \psi_{x,j}\right) \\
&\quad = \boldsymbol{\psi}_{x,\kappa}^T \mathsf{Q}_x \mathbf{1} - \mathbf{1}^T \mathsf{Q}_x \boldsymbol{\psi}_{x,\kappa} = -\mathbf{1}^T \left(\mathsf{E}_x - \mathsf{Q}_x^T\right) \boldsymbol{\psi}_{x,\kappa} = -\mathbf{1}^T \mathsf{E}_x \boldsymbol{\psi}_{x,\kappa},
\end{aligned}$$

where, in the last line, we used $\mathsf{Q}_x \mathbf{1} = \mathsf{H}\mathsf{D}_x \mathbf{1} = \mathbf{0}$. The proof of (17) is analogous. □

Lemma 3. *Let κ and ν be two adjacent elements whose common interface is γ. If $\boldsymbol{w}_\kappa \in \mathbb{R}^{4n_\kappa}$ and $\boldsymbol{w}_\nu \in \mathbb{R}^{4n_\nu}$ are the entropy variables on κ and ν, respectively, then*

$$\frac{1}{2} \boldsymbol{w}_\kappa^T \left[\bar{\mathsf{E}}_x^{\kappa\nu} \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\right] \mathbf{1} + \frac{1}{2} \boldsymbol{w}_\nu^T \left[\bar{\mathsf{E}}_x^{\nu\kappa} \circ \mathsf{F}_x(\boldsymbol{u}_\nu, \boldsymbol{u}_\kappa)\right] \mathbf{1} = \mathbf{1}^T \mathsf{E}_x^{\gamma\kappa} \boldsymbol{\psi}_{x,\kappa} + \mathbf{1}^T \mathsf{E}_x^{\gamma\nu} \boldsymbol{\psi}_{x,\nu} \tag{18}$$

and

$$\frac{1}{2} \boldsymbol{w}_\kappa^T \left[\bar{\mathsf{E}}_y^{\kappa\nu} \circ \mathsf{F}_y(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\right] \mathbf{1} + \frac{1}{2} \boldsymbol{w}_\nu^T \left[\bar{\mathsf{E}}_y^{\nu\kappa} \circ \mathsf{F}_y(\boldsymbol{u}_\nu, \boldsymbol{u}_\kappa)\right] \mathbf{1} = \mathbf{1}^T \mathsf{E}_y^{\gamma\kappa} \boldsymbol{\psi}_{y,\kappa} + \mathbf{1}^T \mathsf{E}_y^{\gamma\nu} \boldsymbol{\psi}_{y,\nu}. \tag{19}$$

PROOF. Again, we will only prove (18), since the proof of (19) is similar. Expressing (18) as an explicit sum we have

$$\begin{aligned}
&\frac{1}{2}\boldsymbol{w}_\kappa^T\left[\bar{\mathsf{E}}_x^{\kappa\nu}\circ\mathsf{F}_x(\boldsymbol{u}_\kappa,\boldsymbol{u}_\nu)\right]\mathbf{1}+\frac{1}{2}\boldsymbol{w}_\nu^T\left[(\bar{\mathsf{E}}_x^{\nu\kappa}\circ\mathsf{F}_x(\boldsymbol{u}_\nu,\boldsymbol{u}_\kappa)\right]\mathbf{1}\\
&\quad=\sum_{i=1}^{n_\kappa}\sum_{j=1}^{n_\nu}\boldsymbol{w}_{\kappa,i}^T\left(\mathsf{E}_x^{\kappa\nu}\right)_{ij}\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i},\boldsymbol{u}_{\nu,j})+\sum_{i=1}^{n_\nu}\sum_{j=1}^{n_\kappa}\boldsymbol{w}_{\nu,i}^T\left(\mathsf{E}_x^{\nu\kappa}\right)_{ij}\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\nu,i},\boldsymbol{u}_{\kappa,j})\\
&\quad=\sum_{i=1}^{n_\kappa}\sum_{j=1}^{n_\nu}\left(\mathsf{E}_x^{\kappa\nu}\right)_{ij}\left(\boldsymbol{w}_{\kappa,i}-\boldsymbol{w}_{\nu,j}\right)^T\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{u}_{\kappa,i},\boldsymbol{u}_{\nu,j}).
\end{aligned}$$

To get the last line above, we used $\mathsf{E}_x^{\kappa\nu}=-\left(\mathsf{E}_x^{\nu\kappa}\right)^T$ and the symmetry of the numerical flux function. Continuing, we use the entropy consistency condition to find that

$$\begin{aligned}
&\frac{1}{2}\boldsymbol{w}_\kappa^T\left[\bar{\mathsf{E}}_x^{\kappa\nu}\circ\mathsf{F}_x(\boldsymbol{u}_\kappa,\boldsymbol{u}_\nu)\right]\mathbf{1}+\frac{1}{2}\boldsymbol{w}_\nu^T\left[\bar{\mathsf{E}}_x^{\nu\kappa}\circ\mathsf{F}_x(\boldsymbol{u}_\nu,\boldsymbol{u}_\kappa)\right]\mathbf{1}\\
&\quad=\sum_{i=1}^{n_\kappa}\sum_{j=1}^{n_\nu}\left(\mathsf{E}_x^{\kappa\nu}\right)_{ij}\left(\psi_{x,\kappa,i}-\psi_{x,\nu,j}\right)\\
&\quad=\mathbf{1}^T\mathsf{R}_{\gamma\nu}^T\mathsf{B}_\gamma\mathsf{N}_{x,\gamma}\mathsf{R}_{\gamma\kappa}\boldsymbol{\psi}_{x,\kappa}-\mathbf{1}^T\mathsf{R}_{\gamma\kappa}^T\mathsf{B}_\gamma\mathsf{N}_{x,\gamma}\mathsf{R}_{\gamma\nu}\boldsymbol{\psi}_{x,\nu}\\
&\quad=\mathbf{1}^T\mathsf{R}_{\gamma\kappa}^T\mathsf{B}_\gamma\mathsf{N}_{x,\gamma}\mathsf{R}_{\gamma\kappa}\boldsymbol{\psi}_{x,\kappa}-\mathbf{1}^T\mathsf{R}_{\gamma\nu}^T\mathsf{B}_\gamma\mathsf{N}_{x,\gamma}\mathsf{R}_{\gamma\nu}\boldsymbol{\psi}_{x,\nu}\\
&\quad=\mathbf{1}^T\mathsf{E}_x^{\gamma\kappa}\boldsymbol{\psi}_{x,\kappa}+\mathbf{1}^T\mathsf{E}_x^{\gamma\nu}\boldsymbol{\psi}_{x,\nu},
\end{aligned}$$

where we have used $\mathsf{R}_{\gamma\kappa}\mathbf{1}=\mathsf{R}_{\gamma\nu}\mathbf{1}$ in the penultimate step. □

We can now state and prove the main entropy-conservation result.

Theorem 4. *The SBP-SAT discretization of the Euler equations, given by (6) or (8), is entropy conservative provided the numerical flux functions $\boldsymbol{\mathcal{F}}_x^\star(\cdot,\cdot)$ and $\boldsymbol{\mathcal{F}}_y^\star(\cdot,\cdot)$ are symmetric in their arguments and satisfy the entropy consistency condition (5).*

PROOF. As in Lemma 3, let $\boldsymbol{w}_\kappa\in\mathbb{R}^{4n_\kappa}$ and $\boldsymbol{w}_\nu\in\mathbb{R}^{4n_\nu}$ denote the entropy variables on generic elements κ and ν, respectively. Replacing $\boldsymbol{v}_\kappa$ with $\boldsymbol{w}_\kappa$ in (8) we have

$$\begin{aligned}
\boldsymbol{w}_\kappa^T\bar{\mathsf{H}}_\kappa\frac{\mathrm{d}\boldsymbol{u}_\kappa}{\mathrm{d}t}=&-\boldsymbol{w}_\kappa^T\left[\bar{\mathsf{S}}_x\circ\mathsf{F}_x(\boldsymbol{u}_\kappa,\boldsymbol{u}_\kappa)\right]\mathbf{1}-\boldsymbol{w}_\kappa^T\left[\bar{\mathsf{S}}_y\circ\mathsf{F}_y(\boldsymbol{u}_\kappa,\boldsymbol{u}_\kappa)\right]\mathbf{1}\\
&-\frac{1}{2}\sum_{\gamma\subset\partial\Omega_\kappa}\boldsymbol{w}_\kappa^T\left[\bar{\mathsf{E}}_x^{\kappa\nu}\circ\mathsf{F}_x(\boldsymbol{u}_\kappa,\boldsymbol{u}_\nu)+\bar{\mathsf{E}}_y^{\kappa\nu}\circ\mathsf{F}_y(\boldsymbol{u}_\kappa,\boldsymbol{u}_\nu)\right]\mathbf{1},
\end{aligned}\tag{20}$$

where we have moved all the spatial terms to the right hand side.

To simplify the right-hand side we use Lemma 2 and the decomposition (4). We also note that, due to the diagonal mass matrix, $\boldsymbol{w}_\kappa^T\bar{\mathsf{H}}_\kappa\mathrm{d}\boldsymbol{u}_\kappa/\mathrm{d}t=\mathbf{1}^T\mathsf{H}_\kappa\mathrm{d}\boldsymbol{s}_\kappa/\mathrm{d}t$, where $\boldsymbol{s}_\kappa\in\mathbb{R}^{n_\kappa}$ is the entropy evaluated at the SBP nodes of element κ. Thus,

$$\begin{aligned}
\mathbf{1}^T\mathsf{H}_\kappa\frac{\mathrm{d}\boldsymbol{s}_\kappa}{\mathrm{d}t}=&\sum_{\gamma\subset\Gamma_\kappa}\mathbf{1}^T\left[\mathsf{E}_x^{\gamma\kappa}\boldsymbol{\psi}_{x,\kappa}+\mathsf{E}_y^{\gamma\kappa}\boldsymbol{\psi}_{y,\kappa}\right]\\
&-\frac{1}{2}\sum_{\gamma\subset\partial\Omega_\kappa}\boldsymbol{w}_\kappa^T\left[\bar{\mathsf{E}}_x^{\kappa\nu}\circ\mathsf{F}_x(\boldsymbol{u}_\kappa,\boldsymbol{u}_\nu)+\bar{\mathsf{E}}_y^{\kappa\nu}\circ\mathsf{F}_y(\boldsymbol{u}_\kappa,\boldsymbol{u}_\nu)\right]\mathbf{1}.
\end{aligned}$$

Next, we sum the above equation over all K elements to get

$$\begin{aligned}\sum_{\kappa=1}^{K} \mathbf{1}^T \mathsf{H}_\kappa \frac{\mathrm{d}\boldsymbol{s}_\kappa}{\mathrm{d}t} = &\sum_{\kappa=1}^{K} \sum_{\gamma \subset \Gamma_\kappa} \mathbf{1}^T \left[\mathsf{E}_x^{\gamma\kappa} \boldsymbol{\psi}_{x,\kappa} + \mathsf{E}_y^{\gamma\kappa} \boldsymbol{\psi}_{y,\kappa} \right] \\ &- \frac{1}{2} \sum_{\gamma \subset \Gamma^I} \left\{ \boldsymbol{w}_\kappa^T \left[\bar{\mathsf{E}}_x^{\kappa\nu} \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) \right] \mathbf{1} + \boldsymbol{w}_\nu^T \left[\bar{\mathsf{E}}_x^{\nu\kappa} \circ \mathsf{F}_x(\boldsymbol{u}_\nu, \boldsymbol{u}_\kappa) \right] \mathbf{1} \right\} \\ &- \frac{1}{2} \sum_{\gamma \subset \Gamma^I} \left\{ \boldsymbol{w}_\kappa^T \left[\bar{\mathsf{E}}_y^{\kappa\nu} \circ \mathsf{F}_y(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) \right] \mathbf{1} + \boldsymbol{w}_\nu^T \left[\bar{\mathsf{E}}_y^{\nu\kappa} \circ \mathsf{F}_y(\boldsymbol{u}_\nu, \boldsymbol{u}_\kappa) \right] \mathbf{1} \right\}.\end{aligned}$$

We can now simplify the interface terms on the right-hand side by invoking identities (18) and (19) from Lemma 3. We arrive at

$$\begin{aligned}\sum_{\kappa=1}^{K} \mathbf{1}^T \mathsf{H}_\kappa \frac{\mathrm{d}\boldsymbol{s}_\kappa}{\mathrm{d}t} = &\sum_{\kappa=1}^{K} \sum_{\gamma \subset \Gamma_\kappa} \mathbf{1}^T \left[\mathsf{E}_x^{\gamma\kappa} \boldsymbol{\psi}_{x,\kappa} + \mathsf{E}_y^{\gamma\kappa} \boldsymbol{\psi}_{y,\kappa} \right] \\ &- \sum_{\gamma \subset \Gamma^I} \left[\mathbf{1}^T \mathsf{E}_x^{\gamma\kappa} \boldsymbol{\psi}_{x,\kappa} + \mathbf{1}^T \mathsf{E}_x^{\gamma\nu} \boldsymbol{\psi}_{x,\nu} + \mathbf{1}^T \mathsf{E}_y^{\gamma\kappa} \boldsymbol{\psi}_{y,\kappa} + \mathbf{1}^T \mathsf{E}_y^{\gamma\nu} \boldsymbol{\psi}_{y,\nu} \right].\end{aligned} \tag{21}$$

The interface terms cancel if we rewrite the single sum on the right-hand side as a double sum over elements and their faces; thus, we are left with

$$\sum_{\kappa=1}^{K} \mathbf{1}^T \mathsf{H}_\kappa \frac{\mathrm{d}\boldsymbol{s}_\kappa}{\mathrm{d}t} = 0.$$

Therefore, the volume integrated entropy is constant in time on the periodic domain. □

Remark 11. While Theorem 4 considers global entropy conservation, it is straightforward to generalize this result to elementwise conservation of entropy using Lemmas 2 and 3.

4.4. Entropy stability via dissipative interior penalties

Because the entropy-conservative SBP-SAT discretization (6) does not introduce dissipation through an upwinded flux, it has no mechanism to control aliasing errors and will generally produce non-physical "saw-tooth" solutions. To control such errors, we introduce dissipative penalties at the interfaces. These penalties must not destroy the accuracy and conservation properties of the scheme, and they must be strictly dissipative of entropy, i.e. entropy-stable.

Unlike the tensor-product SBP-SAT discretizations in [5] and [26], we cannot obtain an entropy-stable scheme simply by modifying the numerical flux function used in the SAT penalties. The underlying reason is that the E_x, E_y, and E_z matrices are not diagonal for general multidimensional SBP operators. Instead, an entropy-stable discretization can be obtained by adding a dissipative term to the right-hand side of (6). For a generic element κ the discretization becomes

$$\frac{\mathrm{d}\boldsymbol{u}_\kappa}{\mathrm{d}t} + \left[\bar{\mathsf{D}}_x \circ \mathsf{F}_x(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa) \right] \mathbf{1} + \left[\bar{\mathsf{D}}_y \circ \mathsf{F}_y(\boldsymbol{u}_\kappa, \boldsymbol{u}_\kappa) \right] \mathbf{1} = \bar{\mathsf{H}}_\kappa^{-1} \boldsymbol{r}_\kappa(\boldsymbol{u}_h) - \bar{\mathsf{H}}_\kappa^{-1} \boldsymbol{d}_\kappa(\boldsymbol{u}_h), \tag{22}$$

where the dissipation is defined by

$$\boldsymbol{d}_\kappa(\boldsymbol{u}_h) = \sum_{\gamma \subset \partial\Omega_\kappa} (\mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \otimes \mathsf{I}_4) \Lambda_\gamma(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) \left[(\mathsf{R}_{\gamma\kappa} \otimes \mathsf{I}_4) \boldsymbol{w}_\kappa - (\mathsf{R}_{\gamma\nu} \otimes \mathsf{I}_4) \boldsymbol{w}_\nu \right].$$

Theorem 5. *Let $\Lambda_\gamma(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu) \in \mathbb{R}^{4n_\gamma \times 4n_\gamma}$ be a block-diagonal symmetric semi-definite matrix with a block size of 4. Then the SBP-SAT semi-discretization of the Euler equations given by (22) is*

1. *an order h^p approximation to the differential form of the Euler equations;*
2. *discretely conservative of ρ, ρu, ρv, and e, and;*
3. *entropy stable, in that the entropy is non-increasing in time.*

Proof. The preceding theory has established properties 1, 2, and 3 in the absence of $\boldsymbol{d}_\kappa(\boldsymbol{u}_h)$. Therefore it is sufficient to show that the dissipative penalty does not interfere with these properties.

First, we address property 1: accuracy. Suppose the entropy variables, $\mathcal{W}$, are degree p polynomials. Then the interpolation operations $\mathsf{R}_{\gamma\kappa}$ and $\mathsf{R}_{\gamma\nu}$ are exact for $\boldsymbol{w}_\kappa$ and $\boldsymbol{w}_\nu$, and we have $(\mathsf{R}_{\gamma\kappa} \otimes \mathsf{I}_4)\boldsymbol{w}_\kappa - (\mathsf{R}_{\gamma\nu} \otimes \mathsf{I}_4)\boldsymbol{w}_\nu = \mathbf{0}$, i.e. the penalty $\boldsymbol{d}_\kappa(\boldsymbol{u}_h)$ vanishes for polynomials of degree p.

To show conservation we consider the density equation over a subset of elements, Ω' (see Section 4.2 for the notation), since the proof for the remaining variables is analogous. Left multiplying the dissipative penalty by $\mathbf{1}^T\mathsf{H}_\kappa$ we find

$$\mathbf{1}^T\boldsymbol{d}_{\rho,\kappa}(\boldsymbol{u}_h) = \sum_{\gamma\subset\partial\Omega_\kappa} \mathbf{1}^T\mathsf{B}_\gamma\mathsf{L}_{\rho,\gamma}(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\left[(\mathsf{R}_{\gamma\kappa} \otimes \mathsf{I}_4)\boldsymbol{w}_\kappa - (\mathsf{R}_{\gamma\nu} \otimes \mathsf{I}_4)\boldsymbol{w}_\nu\right],$$

where $\boldsymbol{d}_{\rho,\kappa}$ and $\mathsf{L}_{\rho,\gamma}$ are the vector and matrix obtained by retaining those rows in $\boldsymbol{d}_\kappa$ and Λ_γ, respectively, corresponding to the density variable. To arrive at the above, we used $\mathsf{R}_{\gamma\kappa}\mathbf{1} = \mathbf{1}$. Summing over all elements in Ω' we obtain

$$\sum_{\kappa\in\zeta} \mathbf{1}^T\boldsymbol{d}_{\rho,\kappa} = \sum_{\gamma\subset\partial\Omega'} \mathbf{1}^T\mathsf{B}_\gamma\mathsf{L}_{\rho,\gamma}(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\left[(\mathsf{R}_{\gamma\kappa} \otimes \mathsf{I}_4)\boldsymbol{w}_\kappa - (\mathsf{R}_{\gamma\nu} \otimes \mathsf{I}_4)\boldsymbol{w}_\nu\right]$$

where we used the symmetry of $\Lambda_\gamma(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)$ to eliminate the terms on internal interfaces $\gamma \subset \Gamma'$. Thus, the discrete integral depends only on terms on the boundary $\partial\Omega'$, as required for conservation.

Finally, to show entropy stability, we start by left multiplying the penalty by $\boldsymbol{w}_\kappa^T\overline{\mathsf{H}}_\kappa$ to get

$$\boldsymbol{w}_\kappa^T\boldsymbol{d}_\kappa(\boldsymbol{u}_h) = \sum_{\gamma\subset\partial\Omega_\kappa} \boldsymbol{w}_{\gamma\kappa}^T\Lambda_\gamma(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\left[\boldsymbol{w}_{\gamma\kappa} - \boldsymbol{w}_{\gamma\nu}\right],$$

where $\boldsymbol{w}_{\gamma\kappa} = (\mathsf{R}_{\gamma\kappa} \otimes \mathsf{I}_4)\boldsymbol{w}_\kappa$ and $\boldsymbol{w}_{\gamma\nu} = (\mathsf{R}_{\gamma\nu} \otimes \mathsf{I}_4)\boldsymbol{w}_\nu$. Summing the above terms over all elements we get

$$\sum_{\kappa=1}^{K} \boldsymbol{w}_\kappa^T\boldsymbol{d}_\kappa(\boldsymbol{u}_h) = \sum_{\gamma\subset\Gamma^I} \left[\boldsymbol{w}_{\gamma\kappa} - \boldsymbol{w}_{\gamma\nu}\right]^T\Lambda_\gamma(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\left[\boldsymbol{w}_{\gamma\kappa} - \boldsymbol{w}_{\gamma\nu}\right].$$

The above is clearly non-negative provided $\Lambda_\gamma(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)$ is positive semi-definite, which it is by assumption □

In general, the matrix Λ_γ can be any symmetric semi-definite matrix; however, it is natural to adopt a matrix whose scaling is appropriate to the wave speeds at the interface. The matrix we adopt here is based on a Lax-Friedrichs approach where Λ_γ is a block diagonal matrix whose ith 4×4 block is given by

$$\left[\Lambda_\gamma(\boldsymbol{u}_\kappa, \boldsymbol{u}_\nu)\right]_{ii} = \left[|\lambda_{\max}|\frac{\partial\mathcal{U}}{\partial\mathcal{W}}\right]_{ii},$$

where $|\lambda_{\max}|$ an approximation to the magnitude of the fastest wave speed of the flux in the direction $[n_x, n_y]_i^T$. The above matrix and wave speed are evaluated at an average state based on the interpolated/extrapolated states at the ith node of the face.

Remark 12. In general, $\boldsymbol{d}_\kappa$ is an entropy-stable dissipation provided $(\mathsf{B}_\gamma \otimes \mathsf{I}_4)\mathsf{\Lambda}_\gamma$ is symmetric semi-definite, even if $\mathsf{\Lambda}_\gamma$ is not block diagonal.

5. Curvilinear elements

In practice, high-order elements must be curved to accommodate complex geometries and, potentially, flow features. The introduction of curvilinear elements requires careful treatment of the mapping Jacobian if we are to maintain an accurate, conservative, and entropy-conservative scheme. This section describes one possible approach for handling curvilinear elements.

The treatment of curvilinear tensor-product elements for entropy-stable SBP discretizations has been previously addressed by Fisher [27]; see also [26]. In order to generalize the analysis to tensor-product elements, Fisher showed that the mapping Jacobian must satisfy the discrete metric invariants and be averaged before they are combined with the numerical flux. As we show below, the situation is similar, but not identical, for the multidimensional SBP discretization (6).

5.1. Conditions on the SBP operators in physical space

To handle curvilinear elements, we (indirectly) construct SBP operators on each element for the physical coordinates x and y (and z). Strictly speaking, we do not construct SBP operators that adhere to Definition 1 exactly. Specifically, we do not require that the operators be exact for polynomials of a certain degree in x and y; nevertheless, we will show that the operators are accurate to the order of the scheme.

Before describing our approach, it is helpful to review the specific properties of the matrix operators that were used in the analysis[8] presented in Section 4:

- For elementwise conservation, we required that S_x be skew symmetric and that $\mathsf{R}_{\gamma\kappa}$ and $\mathsf{R}_{\gamma\nu}$ exactly interpolate constants, i.e. $\mathsf{R}_{\gamma\kappa}\mathbf{1} = \mathsf{R}_{\gamma\nu}\mathbf{1} = \mathbf{1}$.
- For Lemma 2, we again relied on the skew symmetry of S_x. We also used the exactness of the SBP operator when applied to a constant, namely $\mathsf{H}^{-1}\mathsf{Q}_x\mathbf{1} = \mathbf{0}$, and we applied the SBP property $\mathsf{Q}_x + \mathsf{Q}_x^T = \mathsf{E}_x$.
- Like conservation, Lemma 3 required that $\mathsf{R}_{\gamma\kappa}$ and $\mathsf{R}_{\gamma\nu}$ exactly interpolate the constant.
- In addition to the above requirements, Theorem 4 used the fact that H is diagonal.

Obviously, the same properties must hold for E_y and, for three-dimensional operators, E_z. Note that the accuracy of E_x found in Property 3 of Definition 1 was not used for any of the analysis, although it will prove useful below.

5.2. Construction of E_x, E_y, *and* E_z

We will make the following assumption regarding the mesh elements.

Assumption 1. For each element domain $\Omega_\kappa \subset \Omega$, there exists a bijective polynomial mapping of degree at most $q \leq p+1$ that maps a reference domain $\hat{\Omega}_\kappa$ to Ω_κ. Moreover, there exist degree p SBP operators, $\mathsf{D}_\xi = \hat{\mathsf{H}}^{-1}\mathsf{Q}_\xi$ and $\mathsf{D}_\eta = \hat{\mathsf{H}}^{-1}\mathsf{Q}_\eta$ (and $\mathsf{D}_\zeta = \hat{\mathsf{H}}^{-1}\mathsf{Q}_\zeta$), satisfying Definition 1 on $\hat{\Omega}_\kappa$, where (ξ, η) are Cartesian coordinates in the reference space. Additionally, the boundary of $\hat{\Omega}_\kappa$ is piecewise smooth and each of its smooth faces admits a strong cubature rule of degree $2r \geq 2p$.

[8] With the exception of accuracy, which we will address separately

Remark 13. A bijective polynomial mapping has a Jacobian with either a strictly positive or strictly negative determinant. Without loss of generality, we will assume the determinant of the mapping Jacobian is positive.

Our motivation for introducing Assumption 1 is so we can use a decomposition of E_x, E_y, and E_z that satisfies the divergence theorem in physical space (see Proposition 1 below). In particular, we use the now familiar decomposition

$$\mathsf{E}_x = \sum_{\gamma \subset \Gamma_\kappa} \mathsf{E}_x^{\gamma\kappa} = \sum_{\gamma \subset \Gamma_\kappa} \mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \mathsf{N}_{x,\gamma} \mathsf{R}_{\gamma\kappa}, \tag{23}$$

where B_γ hold the cubature weights for face γ in reference space (see Section 3.1), and the $\mathsf{N}_{x,\gamma}$ now hold scaled normal vectors, which are defined below. The interpolation/extrapolation operators $\mathsf{R}_{\gamma\kappa}$ are exact for polynomials of total degree $r \geq p$ *in reference space*, not physical space; however, constant functions are an important exception. In particular, $\mathsf{R}_{\gamma\kappa}\mathbf{1} = \mathbf{1}$ is required by the analysis in Section 4.

Remark 14. One could construct the $\mathsf{R}_{\gamma\kappa}$ such that they are exact for polynomials of degree $r \geq p$ in physical space, but this would require storing a unique $\mathsf{R}_{\gamma\kappa}$ for each face of each element.

In two-dimensions, the scaled face-normal components along the diagonal of $\mathsf{N}_{x,\gamma}$ are computed as

$$\begin{aligned}
\left(\mathsf{N}_{x,\gamma}\right)_{jj} &= \left(\mathcal{J}\frac{\partial \xi}{\partial x} n_\xi + \mathcal{J}\frac{\partial \eta}{\partial x} n_\eta\right)_{(\xi_j,\eta_j)} \\
&= \left(\frac{\partial y}{\partial \eta} n_\xi - \frac{\partial y}{\partial \xi} n_\eta\right)_{(\xi_j,\eta_j)}, \qquad \forall\, j = 1, 2, \ldots, n_\gamma,
\end{aligned}$$

where $\mathcal{J}$ is the determinant of the mapping Jacobian, $[n_\xi, n_\eta]^T$ is the outward unit normal on γ in reference space, and $(\xi_j, \eta_j) \in \gamma$ denotes the location of the jth cubature node on the face. In three dimensions the scaled face-normal components are

$$\begin{aligned}
\left(\mathsf{N}_{x,\gamma}\right)_{jj} &= \left(\mathcal{J}\frac{\partial \xi}{\partial x} n_\xi + \mathcal{J}\frac{\partial \eta}{\partial x} n_\eta + \mathcal{J}\frac{\partial \zeta}{\partial x} n_\zeta\right)_{(\xi_j,\eta_j,\zeta_j)} \\
&= \left[\left(\frac{\partial y}{\partial \eta}\frac{\partial z}{\partial \zeta} - \frac{\partial z}{\partial \eta}\frac{\partial y}{\partial \zeta}\right) n_\xi + \left(\frac{\partial y}{\partial \zeta}\frac{\partial z}{\partial \xi} - \frac{\partial z}{\partial \zeta}\frac{\partial y}{\partial \xi} n_\eta\right) n_\eta + \left(\frac{\partial y}{\partial \xi}\frac{\partial z}{\partial \eta} - \frac{\partial z}{\partial \xi}\frac{\partial y}{\partial \eta}\right) n_\zeta\right]_{(\xi_j,\eta_j,\zeta_j)},
\end{aligned}$$

for all $j = 1, 2 \ldots, n_\gamma$. Analogous expressions hold for $\mathsf{N}_{y,\gamma}$ and $\mathsf{N}_{z,\gamma}$.

We now come to the special case of the divergence theorem satisfied by E_x, which will be useful later. A similar result holds for E_y and E_z.

Proposition 1. *If the conditions of Assumption 1 hold, then the matrix* E_x *defined by* (23) *satisfies the discrete geometric divergence condition*

$$\mathbf{1}^T \mathsf{E}_x \mathbf{1} = \int_{\partial\Omega_\kappa} n_x \mathrm{d}\Gamma = 0.$$

PROOF. In two dimensions, the scaled normal vectors, $\mathsf{N}_{x,\gamma}$ and $\mathsf{N}_{y,\gamma}$, consist of derivatives of the coordinate mapping (see the definitions above); therefore, by Assumption 1, the scaled normal vectors are degree $q-1 \leq p < 2r$ polynomials in two dimensions. In three dimensions, the scaled normal vectors are products of derivatives of the coordinate mapping, so they are degree $2(q-1) \leq 2p \leq 2r$ polynomials under Assumption 1. Consequently, the cubature rules on the faces are exact for the components of the normal, and we have

$$\mathbf{1}^T \mathsf{E}_x \mathbf{1} = \sum_{\gamma \subset \partial\Omega_\kappa} \mathbf{1}^T \mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \mathsf{N}_{x,\gamma} \mathsf{R}_{\gamma\kappa} \mathbf{1} = \sum_{\gamma \subset \partial\Omega_\kappa} \mathbf{1}^T \mathsf{B}_\gamma \mathsf{N}_{x,\gamma} \mathbf{1} = \sum_{\gamma \subset \partial\Omega_\kappa} \sum_{j=1}^{n_\gamma} b_j (n_x)_j = \int_{\partial\Omega_\kappa} n_x \mathrm{d}\Gamma = 0.$$

□

5.3. Two-dimensional curvilinear elements

We now describe the construction of the SBP operator $\mathsf{D}_x = \mathsf{H}^{-1}(\mathsf{S}_x + \frac{1}{2}\mathsf{E}_x)$ on a two-dimensional curvilinear element. Consider an arbitrary element κ and suppose the conditions of Assumption 1 hold. Then E_x is defined by (23), and

$$\mathsf{H} \equiv \operatorname{diag}\left[\mathcal{J}(\xi_1,\eta_1), \mathcal{J}(\xi_2,\eta_2), \ldots, \mathcal{J}(\xi_{n_\kappa},\eta_{n_\kappa})\right] \hat{\mathsf{H}}, , \tag{24}$$

$$\text{and} \qquad \mathsf{S}_x \equiv \frac{1}{2}\left(\Lambda_{\xi,x}\mathsf{Q}_\xi + \Lambda_{\eta,x}\mathsf{Q}_\eta\right) - \frac{1}{2}\left(\mathsf{Q}_\xi^T \Lambda_{\xi,x} + \mathsf{Q}_\eta^T \Lambda_{\eta,x}\right), \tag{25}$$

where $\hat{\mathsf{H}}$ is the SBP norm matrix on the reference element. The matrices $\Lambda_{\xi,x}$ and $\Lambda_{\eta,x}$ are $n_\kappa \times n_\kappa$ diagonal matrices that hold the scaled metric terms. That is,

$$(\Lambda_{\xi,x})_{ii} = \mathcal{J}(\xi_i,\eta_i)\frac{\partial \xi}{\partial x}(\xi_i,\eta_i), \qquad \forall\, i = 1,2,\ldots,n_\kappa,$$

$$\text{and} \qquad (\Lambda_{\eta,x})_{ii} = \mathcal{J}(\xi_i,\eta_i)\frac{\partial \eta}{\partial x}(\xi_i,\eta_i), \qquad \forall\, i = 1,2,\ldots,n_\kappa.$$

With the exception of polynomial exactness, the operator $\mathsf{D}_x = \mathsf{H}^{-1}(\mathsf{S}_x + \frac{1}{2}\mathsf{E}_x)$ based on (23), (24), and (25) satisfies the requirements for a multidimensional SBP operator (Definition 1): E_x is symmetric; S_x is skew-symmetric, and; the norm matrix is positive definite (and diagonal), since $\hat{\mathsf{H}}$ is diagonal and positive definite and $\mathcal{J}_i = \mathcal{J}(\xi_i,\eta_i) > 0$ for all $i = 1,2,\ldots,n_\kappa$ (see the remark below Assumption 1).

The only remaining condition we require for the conservation and entropy analysis is that $\mathsf{D}_x\mathbf{1} = \mathbf{0}$, which is the topic of the following theorem. We will address accuracy more generally in Section 5.5.

Theorem 6. *Let Assumption 1 hold. Let $\mathsf{D}_x = \mathsf{H}^{-1}(\mathsf{S}_x + \frac{1}{2}\mathsf{E}_x)$ be defined by (23), (24), and (25) on a given element κ. Let $\mathsf{D}_y = \mathsf{H}^{-1}(\mathsf{S}_y + \frac{1}{2}\mathsf{E}_y)$ be defined similarly. Then $\mathsf{D}_x\mathbf{1} = \mathbf{0}$ provided the matrices $\Lambda_{\xi,x}$ and $\Lambda_{\eta,x}$ satisfy the discrete divergence-free condition*

$$\left(\mathsf{D}_\xi \Lambda_{\xi,x} + \mathsf{D}_\eta \Lambda_{\eta,x}\right)\mathbf{1} = \hat{\mathsf{H}}^{-1}\left(\mathsf{E}_\xi \Lambda_{\xi,x} + \mathsf{E}_\eta \Lambda_{\eta,x} - \mathsf{E}_x\right)\mathbf{1}, \tag{26}$$

and $\mathsf{D}_y\mathbf{1} = \mathbf{0}$ provided $\Lambda_{\xi,y}$ and $\Lambda_{\eta,y}$ satisfy

$$\left(\mathsf{D}_\xi \Lambda_{\xi,y} + \mathsf{D}_\eta \Lambda_{\eta,y}\right)\mathbf{1} = \hat{\mathsf{H}}^{-1}\left(\mathsf{E}_\xi \Lambda_{\xi,y} + \mathsf{E}_\eta \Lambda_{\eta,y} - \mathsf{E}_y\right)\mathbf{1}. \tag{27}$$

PROOF. Under the assumptions, it is equivalent to show that, $\mathsf{Q}_x\mathbf{1} = \mathbf{0}$, where $\mathsf{Q}_x = \mathsf{S}_x + \frac{1}{2}\mathsf{E}_x$.

$$\begin{aligned}\mathsf{Q}_x\mathbf{1} &= \left[\frac{1}{2}\left(\Lambda_{\xi,x}\mathsf{Q}_\xi + \Lambda_{\eta,x}\mathsf{Q}_\eta\right) - \frac{1}{2}\left(\mathsf{Q}_\xi^T\Lambda_{\xi,x} + \mathsf{Q}_\eta^T\Lambda_{\eta,x}\right) + \frac{1}{2}\mathsf{E}_x\right]\mathbf{1} \\ &= \frac{1}{2}\left(\Lambda_{\xi,x}\mathsf{Q}_\xi + \Lambda_{\eta,x}\mathsf{Q}_\eta + \mathsf{Q}_\xi\Lambda_{\xi,x} + \mathsf{Q}_\eta\Lambda_{\eta,x} - \mathsf{E}_\xi\Lambda_{\xi,x} - \mathsf{E}_\eta\Lambda_{\eta,x} + \mathsf{E}_x\right)\mathbf{1} \\ &= \frac{1}{2}\left(\mathsf{Q}_\xi\Lambda_{\xi,x} + \mathsf{Q}_\eta\Lambda_{\eta,x} - \mathsf{E}_\xi\Lambda_{\xi,x} - \mathsf{E}_\eta\Lambda_{\eta,x} + \mathsf{E}_x\right)\mathbf{1}\end{aligned}$$

where we have used $-\mathsf{Q}_\xi^T = \mathsf{Q}_\xi - \mathsf{E}_\xi$, $-\mathsf{Q}_\eta^T = \mathsf{Q}_\eta - \mathsf{E}_\eta$, and $\mathsf{Q}_\xi\mathbf{1} = \mathsf{Q}_\eta\mathbf{1} = \mathbf{0}$. The last line above vanishes if (26) is satisfied. The proof for $\mathsf{D}_y\mathbf{1} = \mathbf{0}$ is analogous. □

Remark 15. The conditions (26) and (27) are discretizations of the metric invariants; see, for example, [28]. The left-hand sides are a direct discretization of the invariants using the SBP operators, while the right-hand sides are essentially differences between the face-normal vectors and the interpolated metrics.

Theorem 7. *Suppose that the diagonal entries of $\Lambda_{\xi,x}$, $\Lambda_{\eta,x}$, $\Lambda_{\xi,y}$, and $\Lambda_{\eta,y}$ are the exact scaled metrics. Then the discrete metric invariants, (26) and (27), are satisfied in two dimensions if Assumption 1 holds.*

PROOF. By Assumption 1, the scaled metrics are degree $q-1 \leq p$ polynomials. Therefore, the SBP operators D_ξ and D_η are exact for these terms, and the left-hand sides of (26) and (27) vanish by virtue of the analytical metric invariants. In addition, the interpolation operators are exact for degree $r \geq q$ polynomials in ξ and η; consequently we have, for example, the vanishing residual

$$(\boldsymbol{z})_j \equiv \sum_{i=1}^{n_\kappa}(\mathsf{R}_{\gamma\kappa})_{ji}\left[(n_\xi)_j\left(\mathcal{J}\frac{\partial\xi}{\partial x}\right)_i + (n_\eta)_j\left(\mathcal{J}\frac{\partial\eta}{\partial x}\right)_i\right] - (\mathsf{N}_{x,\gamma})_{jj} = 0,$$

for all $j = 1, 2, \ldots, n_\gamma$, and so

$$\left(\mathsf{E}_\xi\Lambda_{\xi,x} + \mathsf{E}_\eta\Lambda_{\eta,x} - \mathsf{E}_x\right)\mathbf{1} = \sum_{\gamma\subset\Gamma_\kappa}\mathsf{R}_{\gamma\kappa}^T\mathsf{B}_\gamma\boldsymbol{z} = 0.$$

Therefore, both the left-hand and right-hand sides of (26) and (27) vanish under the assumptions. □

5.4. Three-dimensional curvilinear elements

In three-dimensions, the definitions of D_x, D_y, and D_z are similar to the two-dimensional operators. The difference operator in the x direction has the form $\mathsf{D}_x = \mathsf{H}^{-1}(\mathsf{S}_x + \frac{1}{2}\mathsf{E}_x)$, with

$$\mathsf{H} \equiv \operatorname{diag}\left[\mathcal{J}(\xi_1,\eta_1,\zeta_1), \mathcal{J}(\xi_2,\eta_2,\zeta_2), \ldots, \mathcal{J}(\xi_{n_\kappa},\eta_{n_\kappa},\zeta_{n_\kappa})\right]\ \hat{\mathsf{H}}, \tag{28}$$

$$\mathsf{S}_x \equiv \frac{1}{2}\left(\Lambda_{\xi,x}\mathsf{Q}_\xi + \Lambda_{\eta,x}\mathsf{Q}_\eta + \Lambda_{\zeta,x}\mathsf{Q}_\zeta\right) - \frac{1}{2}\left(\mathsf{Q}_\xi^T\Lambda_{\xi,x} + \mathsf{Q}_\eta^T\Lambda_{\eta,x} + \mathsf{Q}_\zeta^T\Lambda_{\zeta,x}\right), \tag{29}$$

and, as before, E_x is defined by (23). The generalization to D_y and D_z is straightforward.

The diagonal matrices $\Lambda_{\xi,x}$, $\Lambda_{\eta,x}$, and $\Lambda_{\zeta,x}$ are similar to their two-dimensional counterparts; however, we will not necessarily use the exact metric terms along the diagonals. That is,

$$\begin{aligned}
(\Lambda_{\xi,x})_{ii} &\approx \mathcal{J}(\xi_i,\eta_i,\zeta_i)\frac{\partial \xi}{\partial x}(\xi_i,\eta_i,\zeta_i), \qquad \forall\, i = 1,2,\ldots,n_\kappa,\\
(\Lambda_{\eta,x})_{ii} &\approx \mathcal{J}(\xi_i,\eta_i,\zeta_i)\frac{\partial \eta}{\partial x}(\xi_i,\eta_i,\zeta_i), \qquad \forall\, i = 1,2,\ldots,n_\kappa,\\
(\Lambda_{\zeta,x})_{ii} &\approx \mathcal{J}(\xi_i,\eta_i,\zeta_i)\frac{\partial \zeta}{\partial x}(\xi_i,\eta_i,\zeta_i), \qquad \forall\, i = 1,2,\ldots,n_\kappa.
\end{aligned}$$

We will define $\Lambda_{\xi,x}$, $\Lambda_{\eta,x}$, and $\Lambda_{\zeta,x}$ more precisely below, at which point the need for permitting flexibility in their definition will become clearer.

As was the case in two dimensions, the only property that is not automatically satisfied by the SBP operators is that $\mathsf{D}_x\mathbf{1} = \mathsf{D}_y\mathbf{1} = \mathsf{D}_z\mathbf{1} = \mathbf{0}$. The following theorem is the three-dimensional version of Theorem 6; its proof is essentially the same and is omitted.

Theorem 8. *Let Assumption 1 hold. Let* $\mathsf{D}_x = \mathsf{H}^{-1}(\mathsf{S}_x + \frac{1}{2}\mathsf{E}_x)$ *be defined by* (23), (28), *and* (29) *on a given element* κ. *Let* $\mathsf{D}_y = \mathsf{H}^{-1}(\mathsf{S}_y + \frac{1}{2}\mathsf{E}_y)$ *and* $\mathsf{D}_z = \mathsf{H}^{-1}(\mathsf{S}_z + \frac{1}{2}\mathsf{E}_z)$ *be defined similarly. Then* $\mathsf{D}_x\mathbf{1} = \mathsf{D}_y\mathbf{1} = \mathsf{D}_z\mathbf{1} = \mathbf{0}$ *provided the following discretized metric invariants are satisfied:*

$$\left(\mathsf{D}_\xi\Lambda_{\xi,x} + \mathsf{D}_\eta\Lambda_{\eta,x} + \mathsf{D}_\zeta\Lambda_{\zeta,x}\right)\mathbf{1} = \hat{\mathsf{H}}^{-1}\left(\mathsf{E}_\xi\Lambda_{\xi,x} + \mathsf{E}_\eta\Lambda_{\eta,x} + \mathsf{E}_\zeta\Lambda_{\zeta,x} - \mathsf{E}_x\right)\mathbf{1}, \tag{30}$$

$$\left(\mathsf{D}_\xi\Lambda_{\xi,y} + \mathsf{D}_\eta\Lambda_{\eta,y} + \mathsf{D}_\zeta\Lambda_{\zeta,y}\right)\mathbf{1} = \hat{\mathsf{H}}^{-1}\left(\mathsf{E}_\xi\Lambda_{\xi,y} + \mathsf{E}_\eta\Lambda_{\eta,y} + \mathsf{E}_\zeta\Lambda_{\zeta,y} - \mathsf{E}_y\right)\mathbf{1}, \tag{31}$$

$$\left(\mathsf{D}_\xi\Lambda_{\xi,z} + \mathsf{D}_\eta\Lambda_{\eta,z} + \mathsf{D}_\zeta\Lambda_{\zeta,z}\right)\mathbf{1} = \hat{\mathsf{H}}^{-1}\left(\mathsf{E}_\xi\Lambda_{\xi,z} + \mathsf{E}_\eta\Lambda_{\eta,z} + \mathsf{E}_\zeta\Lambda_{\zeta,z} - \mathsf{E}_z\right)\mathbf{1}. \tag{32}$$

Unlike the two-dimensional situation, the discrete metric invariants (30)–(32) are not satisfied if the conditions of Assumption 1 are met, in general. Thus, our challenge is to find $\Lambda_{\xi,x}$, $\Lambda_{\eta,x}$, $\Lambda_{\zeta,x}$, etc., that satisfy the discrete metric invariants.

Equations (30)–(32) take the form of the discretized divergence-free condition found in [21]. In that work, the exact face-normal and volume velocity fields were projected onto "nearby" fields that satisfy the discrete divergence-free conditions. In the present context projection would be performed by solving a pair of quadratic optimization problems: 1) one global optimization problem for the face normals in $\mathsf{N}_{x,\gamma}$, $\mathsf{N}_{y,\gamma}$, and $\mathsf{N}_{z,\gamma}$, and; 2) one local-to-the-element optimization problem for the matrices $\Lambda_{\xi,x}$, $\Lambda_{\eta,x}$, $\Lambda_{\zeta,z}$, etc.

The global optimization problem for the $\mathsf{N}_{x,\gamma}$, $\mathsf{N}_{y,\gamma}$, and $\mathsf{N}_{z,\gamma}$ matrices is motivated by multiplying the conditions (30)–(32) by H and summing the result. For example, in the case of (30) this produces

$$\begin{aligned}
\mathbf{1}^T\left(\mathsf{Q}_\xi\Lambda_{\xi,x} + \mathsf{Q}_\eta\Lambda_{\eta,x} + \mathsf{Q}_\zeta\Lambda_{\zeta,x}\right)\mathbf{1} &= \mathbf{1}^T\left(\mathsf{E}_\xi\Lambda_{\xi,x} + \mathsf{E}_\eta\Lambda_{\eta,x} + \mathsf{E}_\zeta\Lambda_{\zeta,x} - \mathsf{E}_x\right)\mathbf{1}\\
\Rightarrow \qquad \mathbf{1}^T\mathsf{E}_x\mathbf{1} &= \mathbf{0}.
\end{aligned}$$

However, this is precisely the divergence-free condition guaranteed by Proposition 1. Thus, *it is unnecessary to solve a global optimization problem for the face normals* if Assumption 1 holds.

To satisfy (30)–(32), we now have a decoupled problem involving the discrete metrics that we solve using the approach in [21]. To illustrate, the entries in the diagonal matrices $\Lambda_{\xi,x}$, $\Lambda_{\eta,x}$, and $\Lambda_{\zeta,x}$ are determined by solving the following strictly convex quadratic optimization problem:

$$\min_{\boldsymbol{m}} \frac{1}{2}(\boldsymbol{m} - \boldsymbol{m}_{\text{targ}})^T(\boldsymbol{m} - \boldsymbol{m}_{\text{targ}}), \qquad \text{subject to} \qquad \mathsf{Q}^T\boldsymbol{m} = \boldsymbol{c}, \tag{33}$$

where

$$\boldsymbol{m}^T \equiv \mathbf{1}^T \begin{bmatrix} \Lambda_{\xi,x} & \Lambda_{\eta,x} & \Lambda_{\zeta,x} \end{bmatrix}, \qquad \mathsf{Q}^T = \begin{bmatrix} \mathsf{Q}_\xi^T & \mathsf{Q}_\eta^T & \mathsf{Q}_\zeta^T \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{c} = \mathsf{E}_x \mathbf{1}.$$

The optimal solution is the vector $\boldsymbol{m}$ that is the closest to $\boldsymbol{m}_{\text{targ}}$ in the Cartesian 2-norm while satisfying the constraint $\mathsf{Q}^T\boldsymbol{m} = \boldsymbol{c}$, which is equivalent to (30) up to a minus sign. The target variables, $\boldsymbol{m}_{\text{targ}}$, are the exact metrics, or $\mathrm{O}(h^{p+2})$ accurate approximations to the metrics.

Proposition 2. *The solution to (33) is given by*

$$\boldsymbol{m} = \boldsymbol{m}_{targ} - (\mathsf{Q}^T)^\dagger \left(\mathsf{Q}^T\boldsymbol{m}_{targ} - \boldsymbol{c}\right), \tag{34}$$

where $(\mathsf{Q}^T)^\dagger$ is the Moore-Penrose pseudoinverse of Q^T. Furthermore,

$$\|\boldsymbol{m} - \boldsymbol{m}_{targ}\|_\infty \leq Mh^{p+2}, \tag{35}$$

where $h > 0$ is the element size and $M > 0$ is independent of h.

Proof. The solution to the optimization problem (33) is equivalent to the minimum 2-norm solution of the linear system

$$\mathsf{Q}^T\Delta\boldsymbol{m} = -\mathsf{Q}^T\boldsymbol{m}_{\text{targ}} + \boldsymbol{c},$$

where $\Delta\boldsymbol{m} = \boldsymbol{m} - \boldsymbol{m}_{\text{targ}}$. The solution to this problem follows from standard linear algebra arguments using a singular value decomposition of Q^T; see, for example, [29]. The solution is (34).

Rearranging (34) to isolate $\boldsymbol{m} - \boldsymbol{m}_{\text{targ}}$ and taking the infinity norm we find

$$\|\boldsymbol{m} - \boldsymbol{m}_{\text{targ}}\|_\infty \leq \|(\mathsf{Q}^T)^\dagger\|_\infty \|\mathsf{Q}^T\boldsymbol{m}_{\text{targ}} - \boldsymbol{c}\|_\infty.$$

The matrix $(\mathsf{Q}^T)^\dagger$ depends only on the SBP operator in reference space, so it has a bounded infinity norm that is independent of h. The term $\mathsf{Q}^T\boldsymbol{m}_{\text{targ}} - \boldsymbol{c}$ is the residual form of Equation (30) with the discrete metrics replaced with the target values, which, by assumption, are exact or $\mathrm{O}(h^{p+2})$ approximations. Furthermore, (30) is an $\mathrm{O}(h^{p+2})$ discretization of one of the metric invariants, since the SBP operators D_ξ, D_η, and D_ζ are exact for polynomials of degree p, as is $\mathsf{R}_{\gamma\kappa}$: note that the scaled metrics are order h^2 in three dimensions. Thus,

$$\|\mathsf{Q}^T\boldsymbol{m}_{\text{targ}} + \boldsymbol{c}\|_\infty \leq M'h^{p+2},$$

for h sufficiently small and some M' independent of h. Combining this bound with the bound on $(\mathsf{Q}^T)^\dagger$ produces the desired result (35). □

Remark 16. The astute reader will notice that Q^T is rank deficient, since the constant vector is in the null space of the difference operators Q_ξ, Q_η, and Q_ζ. In general, such rank deficiency would lead to an infeasible optimization problem. Fortunately, the constraint is made consistent by Proposition 1, which ensures that $\mathbf{1}^T\boldsymbol{c} = \mathbf{1}^T\mathsf{E}_x\mathbf{1} = 0$.

5.5. Accuracy of the curvilinear operators

We conclude this section with an analysis of the accuracy of the two-dimensional curvilinear operators. The accuracy analysis of the three-dimensional operators is similar.

Theorem 9. *Let Assumption 1 hold; furthermore, assume the diagonal entries of $\Lambda_{\xi,x}$ and $\Lambda_{\eta,x}$ are the exact scaled metrics. Let $\mathsf{D}_x = \mathsf{H}^{-1}(\mathsf{S}_x + \frac{1}{2}\mathsf{E}_x)$ be defined by (23), (24), and (25) on element κ. If $\mathcal{U} \in C^{p+1}(\hat{\Omega}_\kappa)$ and $\boldsymbol{u}_\kappa \in \mathbb{R}^n_\kappa$ holds its values at the nodes S_κ, then*

$$[\mathsf{D}_x \boldsymbol{u}_\kappa]_i = \frac{\partial \mathcal{U}}{\partial x}(\xi_i, \eta_i) + O(h^p).$$

Proof. Let J be the diagonal matrix that holds the determinant of the mapping Jacobian:

$$\mathsf{J} \equiv \operatorname{diag}\left[\mathcal{J}(\xi_1, \eta_1), \mathcal{J}(\xi_2, \eta_2), \ldots, \mathcal{J}(\xi_{n_\kappa}, \eta_{n_\kappa})\right].$$

Then, using the definition of D_x, the properties of $\mathsf{D}_\xi = \hat{\mathsf{H}}^{-1}\mathsf{Q}_\xi$ and $\mathsf{D}_\eta = \hat{\mathsf{H}}^{-1}\mathsf{Q}_\eta$, and the fact that diagonal matrices commute, we have

$$\begin{aligned}
\mathsf{D}_x \boldsymbol{u}_\kappa &= \mathsf{J}^{-1}\hat{\mathsf{H}}^{-1}\left[\frac{1}{2}\left(\Lambda_{\xi,x}\mathsf{Q}_\xi + \Lambda_{\eta,x}\mathsf{Q}_\eta\right) - \frac{1}{2}\left(\mathsf{Q}_\xi^T\Lambda_{\xi,x} + \mathsf{Q}_\eta^T\Lambda_{\eta,x}\right) + \frac{1}{2}\mathsf{E}_x\right]\boldsymbol{u}_\kappa \\
&= \frac{1}{2}\mathsf{J}^{-1}\left(\Lambda_{\xi,x}\mathsf{D}_\xi + \Lambda_{\eta,x}\mathsf{D}_\eta\right)\boldsymbol{u}_\kappa && \text{(a)} \\
&\quad + \frac{1}{2}\mathsf{J}^{-1}\left(\mathsf{D}_\xi\Lambda_{\xi,x} + \mathsf{D}_\eta\Lambda_{\eta,x}\right)\boldsymbol{u}_\kappa && \text{(b)} \\
&\quad + \frac{1}{2}\mathsf{J}^{-1}\hat{\mathsf{H}}^{-1}\left(\mathsf{E}_x - \mathsf{E}_\xi\Lambda_{\xi,x} - \mathsf{E}_\eta\Lambda_{\eta,x}\right)\boldsymbol{u}_\kappa. && \text{(c)}
\end{aligned}$$

We will consider each of the three terms (a), (b), and (c) individually.

For (a), recall that $\Lambda_{\xi,x}$ and $\Lambda_{\eta,x}$ hold the exact scaled metric terms. Therefore, since D_ξ and D_η are exact for polynomials of degree p, and $\mathsf{J}^{-1}\Lambda_{\xi,x}$ and $\mathsf{J}^{-1}\Lambda_{\eta,x}$ are $\mathrm{O}(h^{-1})$,

$$\begin{aligned}
\left[\frac{1}{2}\mathsf{J}^{-1}\left(\Lambda_{\xi,x}\mathsf{D}_\xi + \Lambda_{\eta,x}\mathsf{D}_\eta\right)\boldsymbol{u}_\kappa\right]_i &= \frac{1}{2\mathcal{J}_i}\left(\mathcal{J}\frac{\partial\xi}{\partial x}\right)_i (\mathsf{D}_\xi \boldsymbol{u}_\kappa)_i + \frac{1}{2\mathcal{J}_i}\left(\mathcal{J}\frac{\partial\eta}{\partial x}\right)_i (\mathsf{D}_\eta \boldsymbol{u}_\kappa)_i \\
&= \frac{1}{2}\left(\frac{\partial\xi}{\partial x}\frac{\partial\mathcal{U}}{\partial\xi}\right)_i + \mathrm{O}(h^{-1})\mathrm{O}(h^{p+1}) + \frac{1}{2}\left(\frac{\partial\eta}{\partial x}\frac{\partial\mathcal{U}}{\partial\eta}\right)_i + \mathrm{O}(h^{-1})\mathrm{O}(h^{p+1}) \\
&= \frac{1}{2}\left(\frac{\partial\mathcal{U}}{\partial x}\right)_i + \mathrm{O}(h^p). \qquad (36)
\end{aligned}$$

Next, we rewrite (b) using the accuracy of D_ξ and D_η and the product rule:

$$\frac{1}{2}\mathsf{J}^{-1}\left(\mathsf{D}_\xi\Lambda_{\xi,x} + \mathsf{D}_\eta\Lambda_{\eta,x}\right)\boldsymbol{u}_\kappa = \frac{1}{2}\mathsf{J}^{-1}\left[\left(\Lambda_{\xi,x}\mathsf{D}_\xi + \Lambda_{\eta,x}\mathsf{D}_\eta\right)\boldsymbol{u}_\kappa + \mathsf{U}\left(\mathsf{D}_\xi\Lambda_{\xi,x} + \mathsf{D}_\eta\Lambda_{\eta,x}\right)\mathbf{1}\right] + \mathrm{O}(h^p),$$

where $\mathsf{U} = \operatorname{diag}(\boldsymbol{u}_\kappa)$. The first term on the right is identical to (a). The second term on the right vanishes because $\Lambda_{\xi,x}$ and $\Lambda_{\eta,x}$, which hold the exact scaled metrics by assumption, satisfy the metric invariants. Thus,

$$\left[\frac{1}{2}\mathsf{J}^{-1}\left(\mathsf{D}_\xi\Lambda_{\xi,x} + \mathsf{D}_\eta\Lambda_{\eta,x}\right)\boldsymbol{u}_\kappa\right]_i = \frac{1}{2}\left(\frac{\partial\mathcal{U}}{\partial x}\right)_i + \mathrm{O}(h^p). \qquad (37)$$

Finally, (c) amounts to the difference between interpolating a product versus the product of an interpolation. Under the assumption that $\mathcal{U}$ is C^{p+1} continuous on Ω_κ and the scaled metrics are exact, we have

$$\begin{aligned}(\boldsymbol{z})_j &\equiv \sum_{i=1}^{n_\kappa} (\mathsf{R}_{\gamma\kappa})_{ji} \left[(n_\xi)_j \left(\mathcal{J} \frac{\partial \xi}{\partial x} \mathcal{U} \right)_i + (n_\eta)_j \left(\mathcal{J} \frac{\partial \eta}{\partial x} \mathcal{U} \right)_i \right] - (\mathsf{N}_{x,\gamma})_{jj} (\mathsf{R}_{\gamma\kappa})_{ji} \mathcal{U}_i \\ &= \left[(n_\xi)_j \left(\mathcal{J} \frac{\partial \xi}{\partial x} \mathcal{U} \right)_j + (n_\eta)_j \left(\mathcal{J} \frac{\partial \eta}{\partial x} \mathcal{U} \right)_j \right] - (\mathsf{N}_{x,\gamma})_{jj} \mathcal{U}_j + \mathrm{O}(h)\mathrm{O}(h^{p+1}) \\ &= \mathrm{O}(h^{p+2}),\end{aligned}$$

where we have used the fact that $\mathsf{N}_{x,\gamma}$, $\mathsf{\Lambda}_{\xi,x}$, and $\mathsf{\Lambda}_{\eta,x}$ are order h, and the interpolation operator is exact for degree p polynomials. Consequently, noting $\mathsf{J}^{-1}\hat{\mathsf{H}}^{-1} = \mathrm{O}(h^{-2})$, we find that

$$\frac{1}{2}\mathsf{J}^{-1}\hat{\mathsf{H}}^{-1} \left(\mathsf{E}_x - \mathsf{E}_\xi \mathsf{\Lambda}_{\xi,x} - \mathsf{E}_\eta \mathsf{\Lambda}_{\eta,x} \right) \boldsymbol{u}_\kappa = \frac{1}{2}\mathsf{J}^{-1}\hat{\mathsf{H}}^{-1} \sum_{\gamma \subset \Gamma_\kappa} \mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \boldsymbol{z} = \mathrm{O}(h^p). \tag{38}$$

Substituting (36), (37), and (38) for Terms (a), (b), and (c), respectively, we have

$$[\mathsf{D}_x \boldsymbol{u}_\kappa]_i = \frac{1}{2}\left(\frac{\partial \mathcal{U}}{\partial x} \right)_i + \frac{1}{2}\left(\frac{\partial \mathcal{U}}{\partial x} \right)_i + \mathrm{O}(h^p) = \left(\frac{\partial \mathcal{U}}{\partial x} \right)_i + \mathrm{O}(h^p),$$

as desired. □

Remark 17. The proof that establishes the accuracy of the three-dimensional curvilinear operators is essentially the same as the above proof, except that the exact scaled metrics are replaced with $\mathrm{O}(h^{p+2})$ approximations.

6. Numerical experiments

In this section we present several numerical experiments to verify the theorems in Sections 4 and 5. The experiments share several implementation details in common, and we list these here to avoid repetition.

The spatial discretizations use the so-called SBP-Γ elements described in [8][9]. These elements are similar to Lagrange finite elements, but have more nodes for $p \geq 2$ to obtain a diagonal mass matrix. An important feature is that the R matrices are sparse, operating only on the nodes of a particular face, substantially reducing the computational cost of the face terms. Preliminary results using both SBP-Γ and other SBP operators can be found in [30].

All cases discretize the time term using the five-stage, fourth-order low-storage explicit Runge-Kutta method [31]. CFL numbers are computed using the minimum distance between nodes of an element. We also use the numerical flux of Ismail and Roe [6] for all the results presented, unless indicated otherwise[10]. To compute the logarithmic average, we modify the procedure described in [6] reducing the cutoff parameter ϵ to 10^{-3}, originally proposed in [32], and by extending the Taylor series to 5 terms. With these changes, the agreement between the Taylor series and the standard computation is $O(10^{-16})$ at the cutoff.

[9] Source code can be found at: https://github.com/OptimalDesignLab/SummationByParts.jl
[10] Source code can be found at: https://github.com/OptimalDesignLab/PDESolver.jl

Remark 18. In the appendix, we prove that the Ismail-Roe flux is continuously differentiable, as required by Theorem 1.

When plotting solution error versus mesh size, the nominal mesh size h is calculated as $1/N$, where N is the number of elements in each direction. Integrals are approximated using the SBP matrix H. In particular, if the error at the nodes of element κ is $\boldsymbol{e}_\kappa$, then the L^2 error is estimated as $\|\boldsymbol{e}_h\| = \sqrt{\sum_\kappa \boldsymbol{e}_\kappa^T \mathsf{H}_\kappa \boldsymbol{e}_\kappa}$.

The solver is implemented in the Julia language [33] and the mesh data structure is managed by the Parallel Unstructured Mesh Interface [34]. Parallel communication uses the Message Passing Interface.

6.1. Inviscid isentropic vortex

In order to verify the two dimensional code, we solve the unsteady isentropic vortex problem; see, for example, [35]. The analytical solution is known to be [36]

$$
\begin{aligned}
u =& 1 - \frac{\epsilon y}{2\pi} \exp\left(\frac{f(x,y,t)}{2}\right), \qquad v = \frac{\epsilon((x-x_0)-t)}{2\pi} \exp\left(\frac{f(x,y,t)}{2}\right), \\
\rho =& \left(1 - \frac{\epsilon^2(\gamma-1)M^2}{8\pi^2} \exp\left(f(x,y,t)\right)\right)^{\frac{1}{1-\gamma}}, \qquad p = \frac{\rho^\gamma}{\gamma M^2},
\end{aligned}
$$

where $f(x,y,t) = 1 - (((x-x_0)-t)^2 + y^2)$ and the Mach number is $M = 0.5$. The vortex strength ϵ is set to 1, and x_0, the x coordinate of the center of the vortex at $t = 0$, is 5. The y coordinate of the vortex's center is zero.

We solve the isentropic vortex problem on a rectangular domain $\Omega = \{(x,y) \mid x \in [0,20], y \in [-5,5]\}$, with periodic boundary conditions applied to all edges of the domain in order to satisfy the assumptions found in Section 4. The analytical solution is imposed for the initial condition, and the simulation is run for 5 time units. A CFL of 0.01 is used to ensure that the temporal error remains smaller than the spatial error.

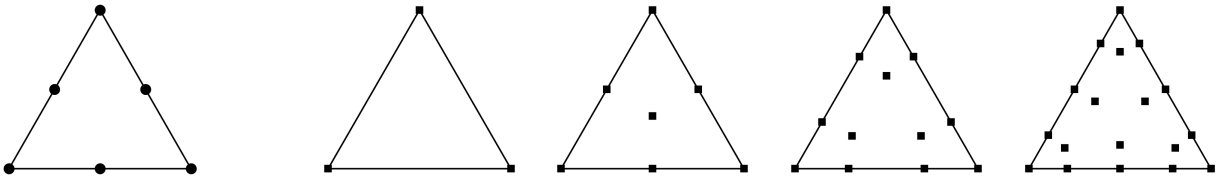
Remark 19. The imposition of periodic boundary conditions introduces a small error in the solution of the isentropic vortex; however, the exact solution tends rapidly to a free-stream flow away from the vortex, and the maximum absolute error on the boundary was found to be 9.78×10^{-6}. The convergence studies (see below) suggest this error is significantly smaller than the discretization error for the grids considered in this work.

Each curvilinear mesh is defined by a second-order Lagrangian coordinate field, which defines the coordinates of each mesh vertex and the midpoint of each edge. In order to curve the elements, the vertices and mid-edge nodes of the coordinate field are remapped according to

$$
x_i \leftarrow a h_i \sin\left(\frac{10\pi x_i}{L_i}\right), \tag{39}
$$

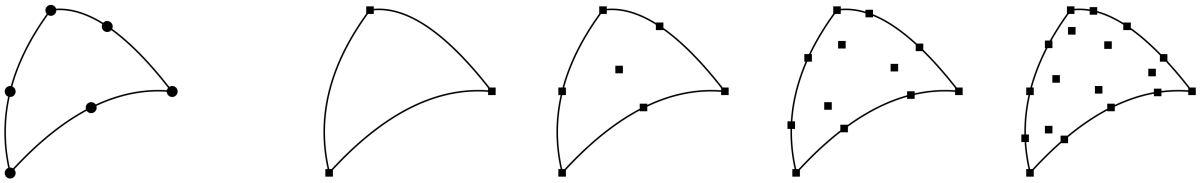
where L_i is the domain length in direction i and the amplitude is $a = 0.2$. The effects of the curvilinear transformation applied to individual elements is illustrated in Figures 1 and 2, and an example mesh is plotted in Figure 3.

The SBP L^2 solution errors are plotted versus h in Figure 4. The results use meshes with $N =$ 20 to 80 elements in each direction in increments of 5. The slopes shown are the least squares



(a) Coordinate nodes (b) $p = 1$ (c) $p = 2$ (d) $p = 3$ (e) $p = 4$

Figure 1: Straight-sided elements



(a) Coordinate nodes (b) $p = 1$ (c) $p = 2$ (d) $p = 3$ (e) $p = 4$

Figure 2: Curvilinear elements

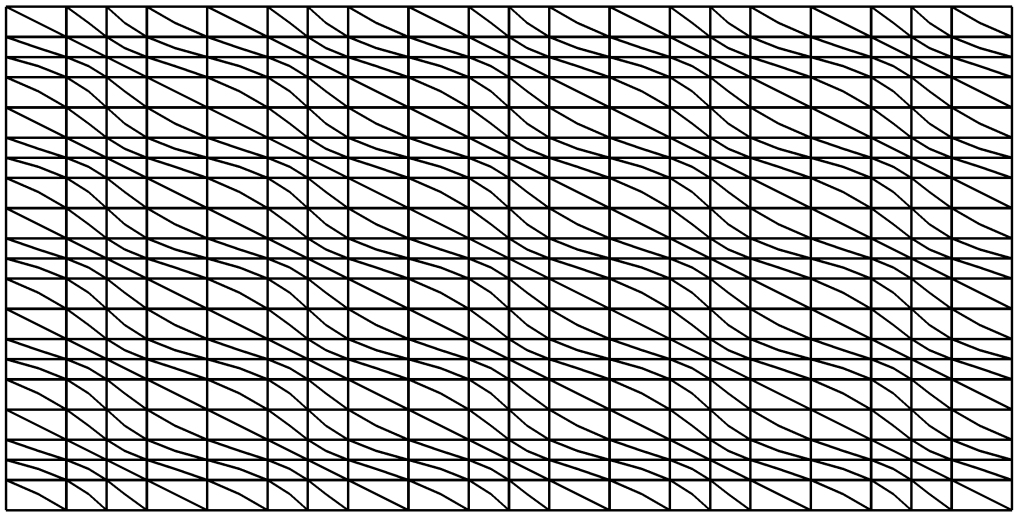
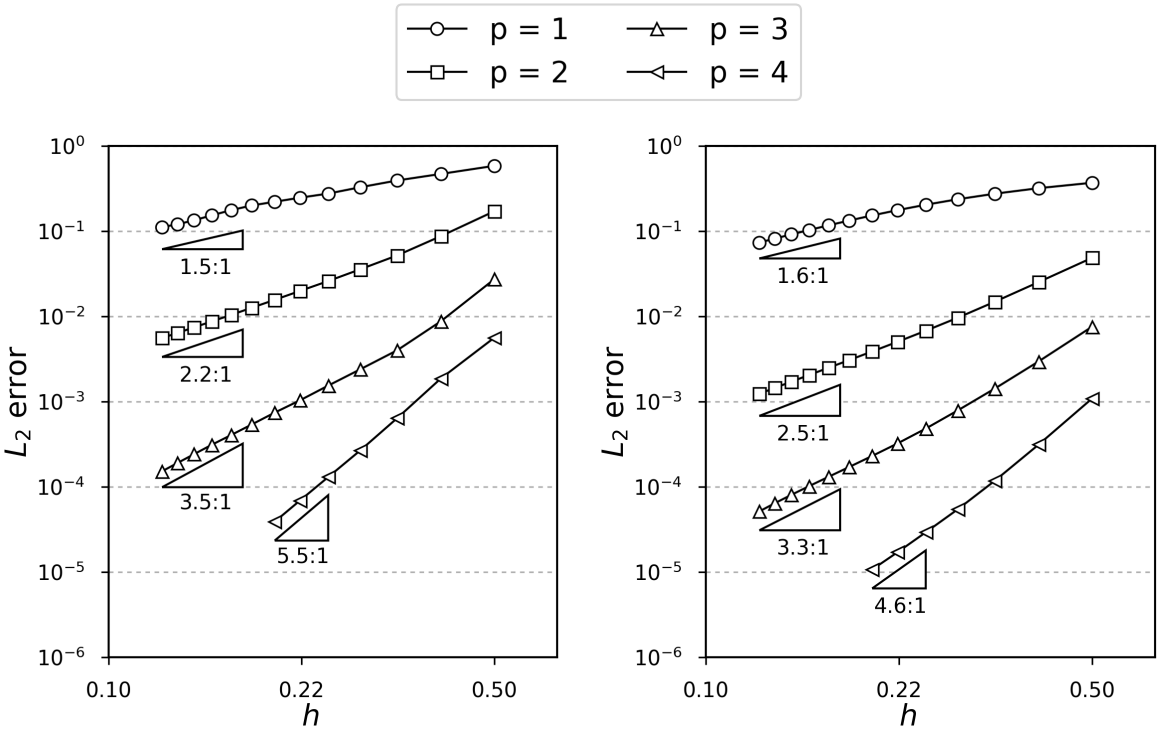


Figure 3: 2D curvilinear mesh



(a) Entropy-conservative

(b) Entropy-stable

Figure 4: Unsteady vortex convergence studies

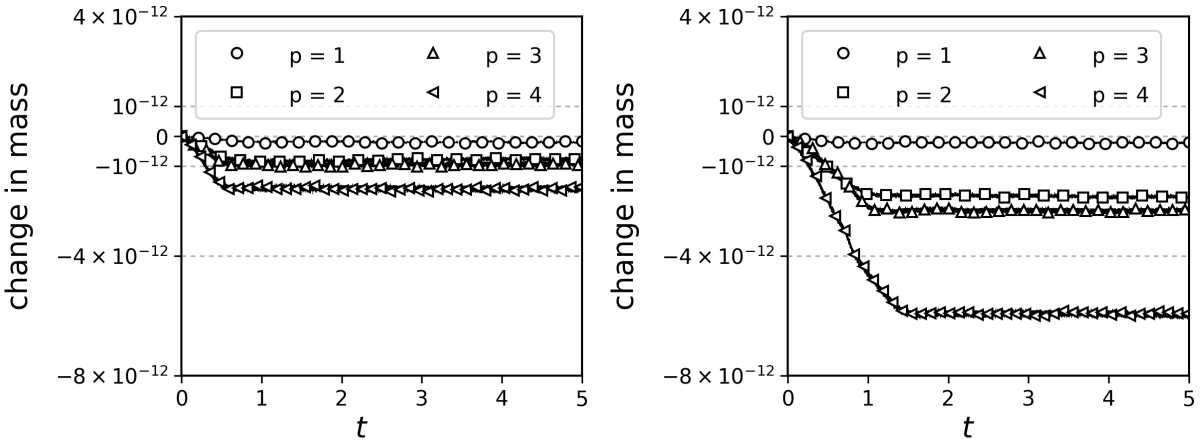
slope of the last 4 meshes. Even without dissipation, the high-order operators show convergence rates between p and $p+1$, with the exception of the $p=4$ entropy conservative discretization, which shows a convergence rate of $p+1.5$. For the entropy-stable scheme, the convergence rates are approximately $p+\frac{1}{2}$. Including the the dissipation term improves the convergence rate of the $p=1$ and $p=2$ operators and improves the error constant for all operators.

Figure 5 shows the change in mass over the course of the simulation. Theorem 3 proves the spatial discretization is conservative, and, after an initial transient, the results show the mass is constant. . The plots of the other conserved quantities (not shown) are similar.

The entropy conservation and stability properties are verified by Figures 6a and 6b. Specifically, Figure 6a shows that the SBP approximation to the integral of the entropy function, $\mathbf{1}^T \mathsf{H} \boldsymbol{s}_h$, remains constant after a small initial transient for the entropy-conservative discretization (6). The results in Figure 6b show that, when the entropy-stable discretization (22) is used, entropy monotonically decreases in all cases, with less dissipation in the higher-order schemes.

6.2. Discontinuous Solution

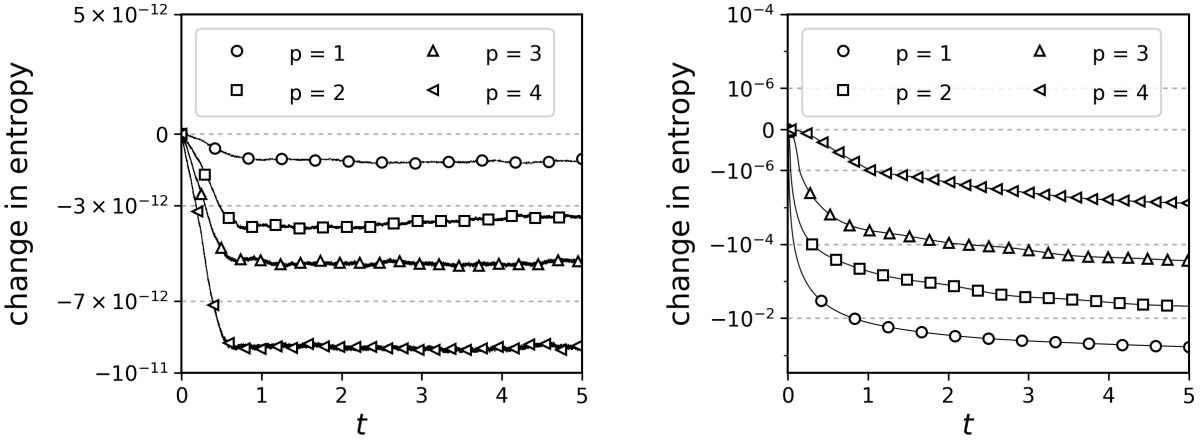
For smooth solutions like the isentropic vortex, even non-entropy-conservative high-order discretizations may approximately conserve entropy. Therefore, to better test the robustness of the entropy-conservative discretization, we consider a discontinuous initial condition. Physically, a solution with shocks will lead to a decrease in the integral of (mathematical) entropy, however a scheme like (8) has no mechanism to dissipate entropy (in the spatial discretization), and the entropy should therefore remain constant for periodic boundary conditions. To test this claim, we use the 20 x 20 mesh from the previous section and the initial condition



(a) Entropy-conservative

(b) Entropy-stable

Figure 5: Unsteady vortex mass conservation. Data is shown with solid lines, symbols are plotted at intervals



(a) Entropy-conservative

(b) Entropy-stable

Figure 6: Unsteady vortex entropy. Data is shown with solid lines, symbols are plotted at intervals

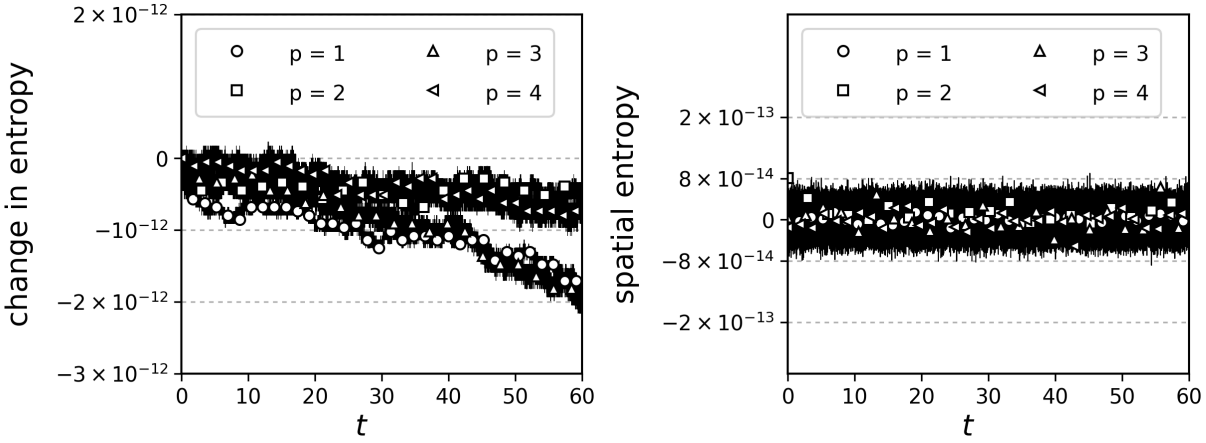


Figure 7: Discontinuous solution results. Data is shown with solid lines, symbols are plotted at intervals

$$\boldsymbol{\mathcal{U}}(x,y) = \left[\rho, \rho u, \rho v, E\right]^T = \begin{cases} \left[1.1, 0.4, 0.4, 5.1\right]^T & \forall \quad 7.5 < x < 12.5, -1.25 < y < 1.25 \\ \left[1.0, 0.3, 0.3, 5.0\right]^T & \text{otherwise} \end{cases}$$

The case is run time a maximum time of 60 time units, at a CFL of 0.01. During this time the numerical solution shows the rectangle in the middle of the domain traversing the top and right boundaries and re-entering the domain in bottom left corner due to the periodic boundary conditions.

The entropy is shown in Figure 7a. There is a trend of slightly decreasing entropy, more pronounced for the $p = 1$ and $p = 4$ operators than for the $p = 2$ and $p = 3$. We note that Theorem 4 addresses only the the spatial discretization and not the temporal discretization. To verify the spatial discretization is entropy conservative, the right-hand side of Equation (20) is plotted in Figure 7b. This figure clearly shows that the spatial contribution to the entropy is small and the entropy change can be attributed to the time term.

6.3. Three-dimensional manufactured solution

The accuracy of the 3 dimensional scheme was investigated using the periodic manufactured solution from [26]. For completeness, the solution is reproduced below:

$$\begin{aligned} \rho &= 2 + \frac{1}{10}\sin(\pi(x + y + z - 2t)), \\ u &= v = w = 1, \\ \rho e &= \left(2 + \frac{1}{10}\sin(\pi(x + y + z - 2t))\right)^2, \end{aligned}$$

with source term

$$
\begin{aligned}
S_\rho &= c_1 \cos(\pi(x+y+z-2t)),\\
S_{\rho u} &= c_2 \cos(\pi(x+y+z-2t)) + c_3 \cos(2\pi(x+y+z-2t)),\\
S_{\rho v} &= c_2 \cos(\pi(x+y+z-2t)) + c_3 \cos(2\pi(x+y+z-2t)),\\
S_{\rho w} &= c_3 \cos(\pi(x+y+z-2t)) + c_3 \cos(2\pi(x+y+z-2t)),\\
S_{\rho e} &= c_4 \cos(\pi(x+y+z-2t)) + c_5 \cos(2\pi(x+y+z-2t)).
\end{aligned}
$$

where $c_1 = \frac{\pi}{10}$, $c_2 = \frac{\pi}{5} + \frac{\pi}{20}(1+5\gamma)$, $c_3 = \frac{\pi}{100}(\gamma - 1)$, $c_4 = \frac{\pi}{20}(-7+15\gamma)$, and $c_5 = \frac{\pi}{100}(3\gamma - 2)$. The domain is a periodic cube, $[-1,1]^3$.

In order to produce a tetrahedral mesh, the domain is divided into cubes and each cube is subdivided into 6 tetrahedra according to the Marching Tetrahedra algorithm [37]. The curvilinear mesh coordinates are produced according to (39) with the amplitude reduced to 0.05 to prevent invalid elements. The convergence studies were run at CFL numbers of 0.35, 0.275, 0.25, and 0.12 for the degree 1 through 4 operators, respectively, using the entropy-stable discretization.

Convergence rates are shown in Figure 8 for linear and curvilinear meshes with 10 to 55 edges in each direction for $p = 1$ and $p = 2$ operators and 10 to 25 edges in each direction for the two higher-order operators. The slopes shown are the least squares slopes of the four finest meshes. The results reveal an even-odd effect for both linear and curvilinear grids, where the degree 2 and 4 operators appear to have an asymptotic convergence rate of p while the odd-order operators appear to converge at a rate of $p+1$. The $p = 1$ discretization requires a much finer mesh to reach the asymptotic region than the higher-order schemes. The $p = 2$ operator was tested on the same set of meshes as the $p = 1$ operator, yet shows the same convergence rate for the last two point and the least squares slope. This suggests the high-order error terms have become negligible and the scheme has reached its asymptotic convergence rate. Tensor-product entropy-stable schemes have previously reported the opposite pattern, where the odd-order operators lose one order of accuracy [7]. Despite the sub-optimal rate of the even-order operators, the error constant is substantially improved by increasing the degree from 1 to 2 and 3 to 4.

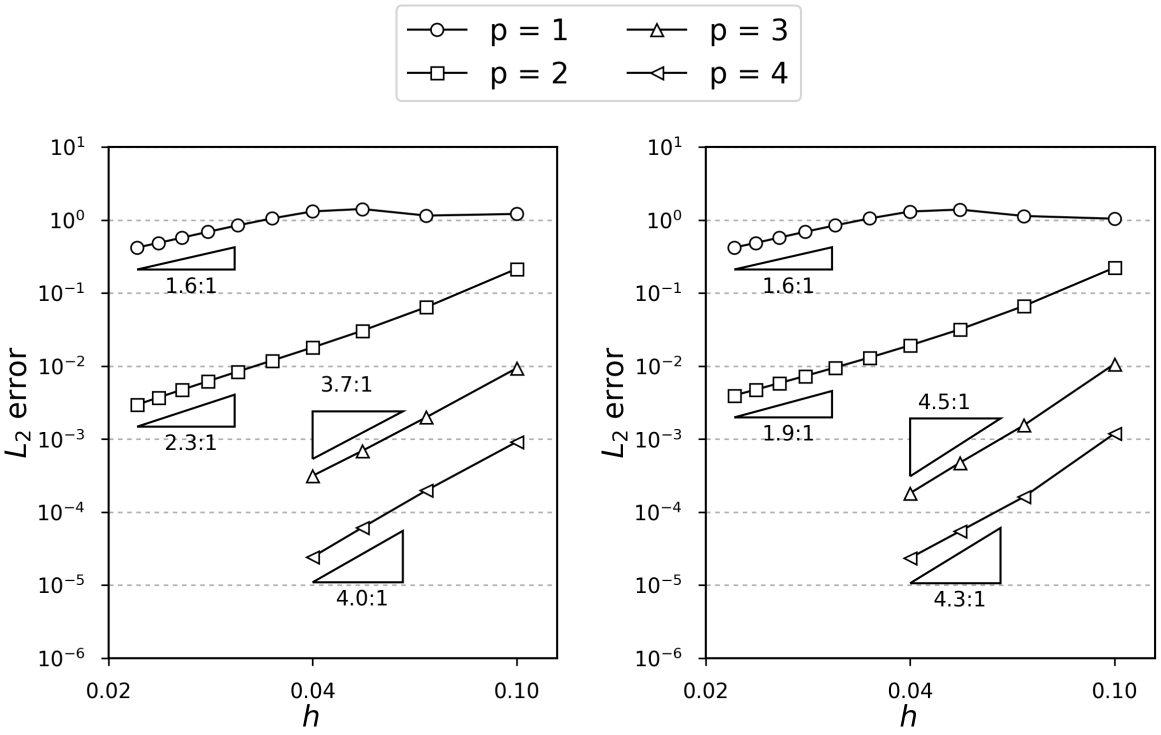
6.4. Taylor-Green vortex

Similar to the discontinuous solution considered in 6.2, we use the Taylor-Green vortex case in order to investigate the entropy properties and robustness of the scheme. The Taylor-Green vortex was originally developed to study turbulence [38], and it offers a challenging test for an inviscid simulation: its initial condition decays into smaller eddies as time progresses, and, in the inviscid limit, there is no limiting scale at which the eddies vanish. This makes the problem a good test case for the stability of a scheme in under-resolved flows.

The domain is a periodic cube $[-\pi, \pi]^3$, and the initial condition is given by

$$
\begin{aligned}
&\rho = 1, \quad u = \sin(x)\cos(y)\cos(z), \quad v = -\cos(x)\sin(y)\cos(z), \quad w = 0\\
&p = \frac{100}{\gamma} + \frac{1}{16}(\cos(2x)\cos(2z) + 2\cos(2y) + 2\cos(2y)\cos(2z)).
\end{aligned}
$$

The meshes have approximately 256,000 nodal degrees of freedom for each degree operator. The precise values are shown in Table 1. Curvilinear meshes were generated using the same perturbation as the 3D manufactured solution case, and all cases were run at a CFL of 0.05.



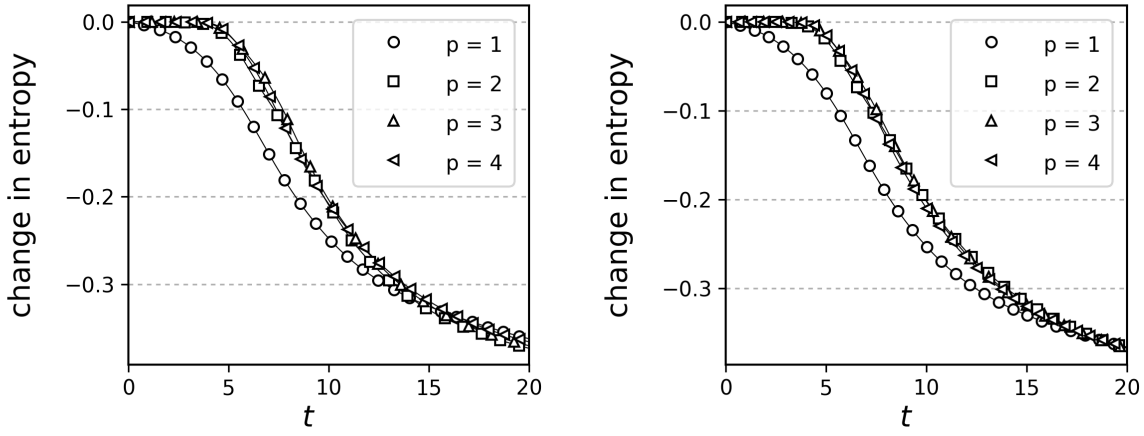
(a) Linear mesh

(b) Curvilinear mesh

Figure 8: 3D solution-error convergence studies

Operator	Edges per direction	Total Elements	Nodal DOFs
$p = 1$	22	63,888	255,552
$p = 2$	16	24,576	270,336
$p = 3$	12	10,368	248,832
$p = 4$	10	6,000	270,000

Table 1: Taylor-Green mesh statistics



(a) Linear mesh

(b) Curvilinear mesh

Figure 9: Taylor-Green entropy-stable results

The results of simulating the Taylor-Green problem using the entropy-stable discretizations are plotted in Figure 9. As expected, entropy is smoothly dissipated for all orders of discretization on both linear and curvilinear meshes.

The entropy-conservative results are shown in Figure 10. Unexpectedly, they show some decrease in entropy. The $p = 1$ discretization has a small loss of entropy over the duration of the simulation, while the higher-order discretizations show progressively larger losses for both linear and curvilinear meshes.

Similar to the discontinuous solution case, we have plotted the right-hand side of Equation (20) at each time step in Figure 11. These results verify that the entropy contribution from the spatial terms is small for both linear and curvilinear meshes.

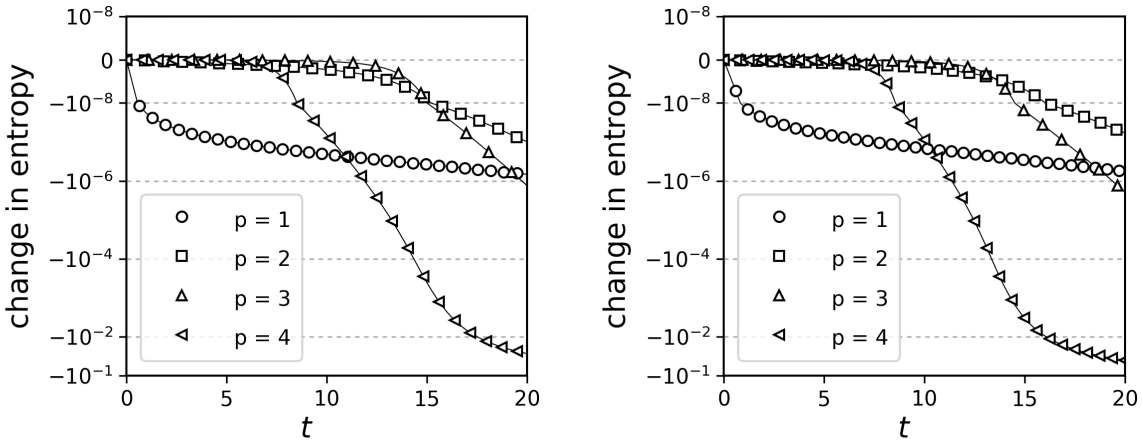
With the spatial terms verified, the temporal discretization remains as the likely source of the discrete entropy error. In order to verify this, a time step refinement study was performed on the linear meshes, examining the entropy change as a function of CFL number. Figure 12 demonstrates that the magnitude of the entropy change reduces at fifth order for all spatial discretization orders. We can conclude that the time-marching scheme is indeed responsible for the entropy dissipation observed in Figure 10. It is not yet clear why the temporal entropy error converges at a fifth-order rate rather than fourth-order, as would be expected from the low-storage Runge-Kutta scheme.

6.5. Comparison to Roe scheme

We conclude the numerical experiments with a demonstration of a standard (non-entropy stable) scheme applied to the Taylor-Green vortex. Specifically, we consider an SBP-SAT discretization of (1) given by by

$$\overline{\mathsf{H}}_\kappa \frac{\mathrm{d}\boldsymbol{u}_\kappa}{\mathrm{d}t} = \overline{\mathsf{Q}}_x^T \boldsymbol{g}_x(\boldsymbol{u}_\kappa) + \overline{\mathsf{Q}}_y^T \boldsymbol{g}_y(\boldsymbol{u}_\kappa) - \sum_{\gamma \subset \partial\Omega_\kappa} (\mathsf{R}_{\gamma\kappa}^T \mathsf{B}_\gamma \otimes \mathsf{I}_4) \boldsymbol{g}_x^\star(\mathsf{R}_{\gamma\kappa}\boldsymbol{u}_\kappa, \mathsf{R}_{\gamma\nu}\boldsymbol{u}_\nu, \mathsf{N}_{x,\gamma}, \mathsf{N}_{y,\gamma}), \tag{40}$$

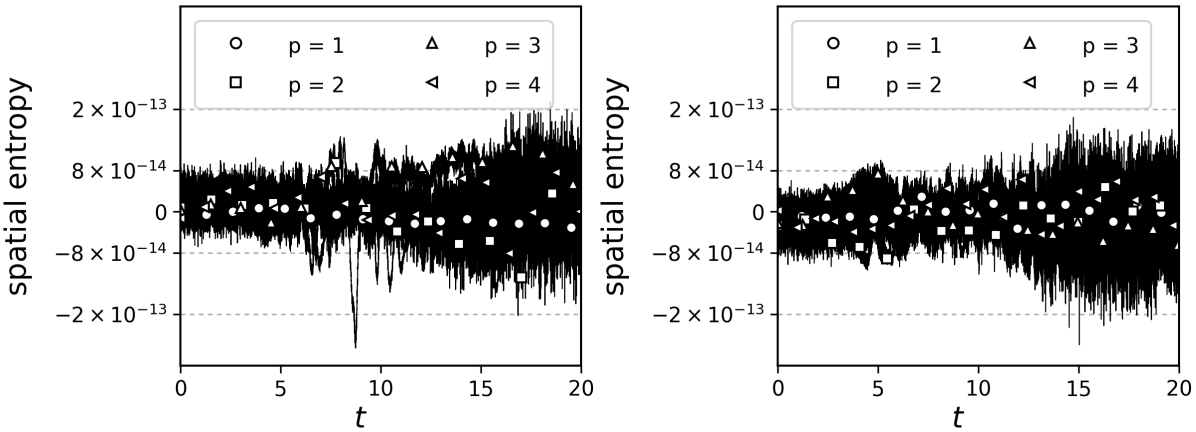
where $\boldsymbol{g}_x$ and $\boldsymbol{g}_y$ are the Euler fluxes $\boldsymbol{\mathcal{F}}_x$ and $\boldsymbol{\mathcal{F}}_y$, respectively, evaluated at the nodes of the element κ. The vector $\boldsymbol{g}_x^\star \in \mathbb{R}^{n_\gamma}$ denotes the Roe numerical flux [24] evaluated at the nodes of the face γ. Unlike (8), this discretization does not utilize Hadamard products nor the newly-introduced $\mathsf{E}_x^{\kappa\nu}$



(a) Linear mesh

(b) Curvilinear mesh

Figure 10: Taylor-Green entropy-conservative results



(a) Linear mesh

(b) Curvilinear mesh

Figure 11: Taylor-Green entropy contribution of the spatial terms. Data is shown with solid lines, symbols are plotted at intervals

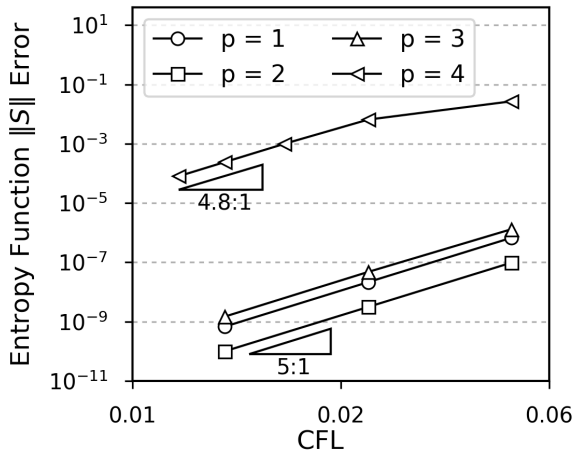


Figure 12: Time-step convergence study

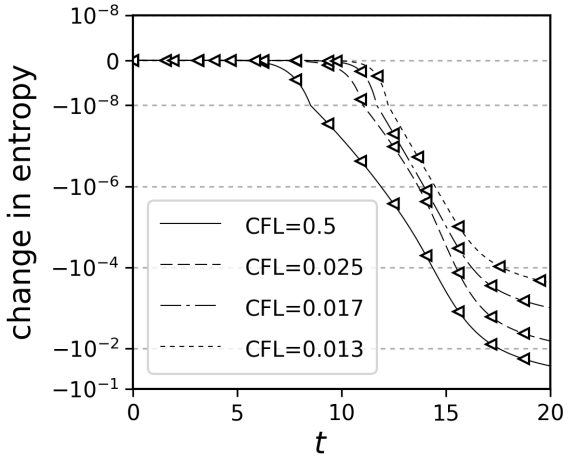


Figure 13: $p = 4$ Entropy history for convergence study

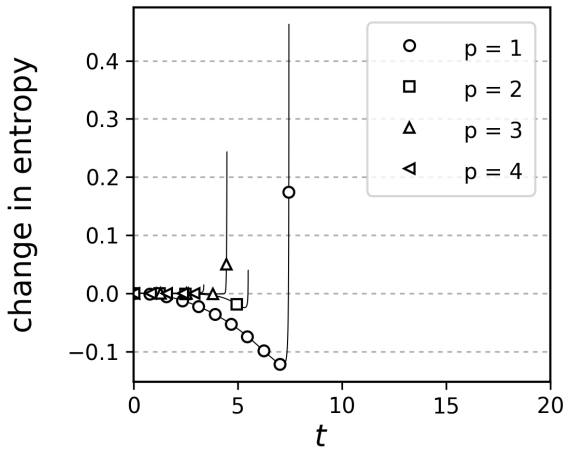


Figure 14: Taylor-Green Roe-scheme results

boundary operator. Instead, it uses the SBP operators to discretize the weak form directly and uses the Roe numerical flux function at the interfaces.

The above discretization introduces some numerical dissipation through the Roe solver, but it is not provably entropy stable. The results of applying discretization (40) to the Taylor-Green vortex are shown in Figure 14. All degree discretizations eventually become unstable and introduce non-physical entropy growth. The $p = 1$ scheme is the most dissipative and consequently remains stable for longest time, but at $t = 7.5$ the under-resolved flow features cause the pressure to become negative. This example demonstrates the utility of provable stability for both low and high order schemes.

7. Conclusions

We have presented a new multidimensional SBP-SAT discretization of the Euler equations that discretely conserves entropy in addition to mass, momentum, and energy. The discretization can be

made entropy stable by introducing dissipation at the interfaces; we considered a Lax-Friedrichs-type dissipation, but other choices are possible. This extends previous work on entropy-stable SBP discretizations to non-tensor product elements, including those with dense boundary operators, i.e. dense E_x, E_y, E_z. By extending the theory beyond tensor-product elements, we hope to facilitate robust simulations on general unstructured grids.

We also addressed how to preserve the properties of the SBP discretization in the presence of curved elements. In two dimensions, it is sufficient to ensure the mapping is polynomial and at most one degree higher than the SBP operator. In three dimensions, one can satisfy the necessary metric invariants by solving a small, convex quadratic optimization problem on each element.

The discretizations were tested on two- and three-dimensional problems to investigate and verify their properties. The two-dimensional discretization was found to have nearly $p+1$ accuracy based on the isentropic-vortex problem. Mass and entropy conservation were also verified for the dissipation-free discretization using the isentropic vortex. When the Lax-Friedrichs dissipation was introduced, mass conservation remained, and the total entropy was found to monotonically decrease for all orders of discretization.

The three-dimensional discretization exhibited an even-odd effect on a manufactured solution: odd p operators were $(p+1)$-order accurate while even p operators were p-order accurate. The source of this three-dimensional even-odd effect remains unclear. In addition, the entropy-conservative scheme did not conserve entropy to numerical precision on the inviscid Taylor-Green vortex, and instead the scheme dissipated entropy; however, we provided evidence that this dissipation was a result of the time-marching scheme. When the entropy-stable discretization was applied to the Taylor-Green problem, total entropy was again monotonically decreasing for all orders.

Finally, we contrasted the robustness of the entropy-stable scheme with that of a standard discretization based on Roe-numerical-flux functions. When applied to the Taylor-Green vortex problem, the Roe scheme eventually produced nonphysical changes in entropy leading to negative pressure, for all orders of discretization.

Significant progress has been made developing entropy-stable SBP-SAT discretizations, but much work remains. For example, the present study considered only periodic boundary conditions, for which the interface SATs can be applied directly. There has been work in the tensor-product SBP literature on entropy-stable boundary conditions [39], and we believe this can be adapted to more general elements *in the case where the boundary operators are diagonal*, i.e. when E_x, E_y, and E_z are diagonal; however, it is not immediately clear if these boundary conditions can be applied to non-diagonal boundary operators to produce entropy-stable discretizations on finite domains. Additionally, as the results from the Taylor-Green case illustrate, time stepping methods that discretely conserve entropy will be required to obtain a fully discrete entropy-conserving scheme.

Acknowledgments

The authors thank the anonymous referees for their valuable feedback improving the manuscript. The authors are grateful to RPI's Scientific Computation Research Center for the use of computer facilities. Finally, the authors would like to acknowledge the use of Matplotlib [40] in producing the figures.

AppendixA. Differentiability of the Ismail-Roe Flux

In this appendix, we prove that the Ismail-Roe (IR) flux [6] is continuously differentiable, which is a requirement for accuracy if the IR flux is used in the discretization (6); see Theorem 1.

Let the arithmetic and logarithmic means be defined by

$$\langle x, y\rangle = \frac{1}{2}(x+y), \qquad \text{and} \qquad \langle x, y\rangle^{\ln} = \frac{x-y}{\ln(x)-\ln(y)},$$

respectively. We begin by proving

Lemma 4. *The logarithmic mean is continuously differentiable on its domain, $x, y > 0$.*

Proof. We will consider the partial derivative of the logarithmic mean with respect to x only, since the derivative with respect to y follows by symmetry. We have

$$\frac{\partial}{\partial x}\langle x, y\rangle^{\ln} = \frac{-x\ln(y) - x + x\ln(x) + y}{x\left[\ln(x)-\ln(y)\right]^2}.$$

The denominator of the partial derivative vanishes only if $x = 0$ or $x = y$. Points for which $x = 0$ are outside the given domain, but we do need to consider $x = y$. Taking the limit as $y \to x$, and applying L'Hôpital's Rule twice, we find

$$\begin{aligned}
\lim_{y\to x} \frac{\partial}{\partial x}\langle x, y\rangle^{\ln} &= \lim_{y\to x} \frac{-x/y + 1}{-2(x/y)\left[\ln(x)-\ln(y)\right]} && \text{(L'Hôpital's Rule)}\\
&= \lim_{y\to x} \frac{-x + y}{-2x\left[\ln(x)-\ln(y)\right]}\\
&= \lim_{y\to x} \frac{1}{2x/y} && \text{(L'Hôpital's Rule)}\\
&= \frac{1}{2}.
\end{aligned}$$

Thus, the limit exists along $x = y$ ($x, y > 0$), and the logarithmic mean is continuously differentiable on its domain. □

The IR flux is defined by[11]

$$\boldsymbol{\mathcal{F}}_x^\star(\boldsymbol{\mathcal{U}}_1, \boldsymbol{\mathcal{U}}_2) = \begin{bmatrix} \hat{\rho}\hat{u} \\ \hat{\rho}\hat{u}^2 + \hat{p} \\ \hat{\rho}\hat{u}\hat{v} \\ \hat{\rho}\hat{u}\hat{H} \end{bmatrix},$$

where the averaged states are defined by

$$\hat{\rho} \equiv \langle \alpha_1, \alpha_2\rangle \langle \beta_1, \beta_2\rangle^{\ln}, \qquad \hat{u} \equiv \frac{\langle u_1\alpha_1, u_2\alpha_2\rangle}{\langle \alpha_1, \alpha_2\rangle}, \qquad \hat{v} \equiv \frac{\langle v_1\alpha_1, v_2\alpha_2\rangle}{\langle \alpha_1, \alpha_2\rangle}$$

$$\hat{p} \equiv \frac{\langle \beta_1, \beta_2\rangle}{\langle \alpha_1, \alpha_2\rangle}, \qquad \hat{H} \equiv \frac{\hat{a}^2}{\gamma - 1} + \frac{1}{2}(\hat{u}^2 + \hat{v}^2), \qquad \hat{a}^2 \equiv \frac{\gamma}{\hat{\rho}}\left[\frac{\gamma+1}{2\gamma}\frac{\langle \beta_1, \beta_2\rangle^{\ln}}{\langle \alpha_1, \alpha_2\rangle^{\ln}} + \frac{\gamma-1}{2\gamma}\frac{\langle \beta_1, \beta_2\rangle}{\langle \alpha_1, \alpha_2\rangle}\right]$$

with

$$\alpha_1 \equiv \sqrt{\frac{\rho_1}{p_1}}, \qquad \alpha_2 \equiv \sqrt{\frac{\rho_2}{p_2}}, \qquad \beta_1 \equiv \sqrt{\rho_1 p_1}, \qquad \beta_2 \equiv \sqrt{\rho_2 p_2}.$$

[11] we consider the x direction flux without loss of generality

Theorem 10. *The Ismail-Roe (IR) flux is continuously differentiable provided ρ_1, ρ_2, p_1, and p_2 are strictly positive.*

Proof. Differentiating the IR flux with respect to $\boldsymbol{\mathcal{U}}_1$ we have

$$\frac{\partial \boldsymbol{\mathcal{F}}_x^\star}{\partial \boldsymbol{\mathcal{U}}_1} = \frac{\partial \boldsymbol{\mathcal{F}}_x^\star}{\partial \hat{\boldsymbol{\mathcal{U}}}} \frac{\partial \hat{\boldsymbol{\mathcal{U}}}}{\partial \boldsymbol{\mathcal{U}}_1},$$

where $\hat{\boldsymbol{\mathcal{U}}} \equiv \begin{bmatrix} \hat{\rho} & \hat{u} & \hat{v} & \hat{p} & \hat{a}^2 \end{bmatrix}^T$ denotes the vector of averaged states. The first Jacobian on the right, $\partial \boldsymbol{\mathcal{F}}_x^\star / \partial \hat{\boldsymbol{\mathcal{U}}}$, is continuously differentiable, since $\boldsymbol{\mathcal{F}}_x^\star$ is polynomial in the elements of $\hat{\boldsymbol{\mathcal{U}}}$. Using the chain rule, the second Jacobian on the right can be expressed as

$$\frac{\partial \hat{\boldsymbol{\mathcal{U}}}}{\partial \boldsymbol{\mathcal{U}}_1} = \frac{\partial \hat{\boldsymbol{\mathcal{U}}}}{\partial \alpha_1} \frac{\partial \alpha_1}{\partial \boldsymbol{\mathcal{U}}_1} + \frac{\partial \hat{\boldsymbol{\mathcal{U}}}}{\partial \beta_1} \frac{\partial \beta_1}{\partial \boldsymbol{\mathcal{U}}_1} + \frac{\partial \hat{\boldsymbol{\mathcal{U}}}}{\partial u_1} \frac{\partial u_1}{\partial \boldsymbol{\mathcal{U}}_1} + \frac{\partial \hat{\boldsymbol{\mathcal{U}}}}{\partial v_1} \frac{\partial v_1}{\partial \boldsymbol{\mathcal{U}}_1},$$

where the intermediate variables α_1 and β_1 are defined above, and u_1 and v_1 are the x and y components of the velocity based on $\boldsymbol{\mathcal{U}}_1$ ($\hat{u}$ and $\hat{v}$ depend on the velocity components). The gradients $\partial \alpha_1 / \partial \boldsymbol{\mathcal{U}}_1$, $\partial \beta_1 / \partial \boldsymbol{\mathcal{U}}_1$, $\partial u_1 / \partial \boldsymbol{\mathcal{U}}_1$, and $\partial v_1 / \partial \boldsymbol{\mathcal{U}}_1$ are continuously differentiable provided that ρ_1, ρ_2, p_1, and p_2 are strictly positive. Furthermore, it follows from Lemma 4 that the derivatives $\partial \hat{\boldsymbol{\mathcal{U}}} / \partial \alpha_1$, $\partial \hat{\boldsymbol{\mathcal{U}}} / \partial \beta_1$, $\partial \hat{\boldsymbol{\mathcal{U}}} / \partial u_1$, and $\partial \hat{\boldsymbol{\mathcal{U}}} / \partial v_1$ are also continuously differentiable if α_1, α_2, β_1, and β_2 are strictly positive. These requirements are also satisfied provided ρ_1, ρ_2, p_1, and p_2 are strictly positive. This completes the proof that $\partial \boldsymbol{\mathcal{F}}_x^\star / \partial \boldsymbol{\mathcal{U}}_1$ is continuously differentiable. The proof for $\partial \boldsymbol{\mathcal{F}}_x^\star / \partial \boldsymbol{\mathcal{U}}_2$ is similar. □

References

[1] Hughes, T. J. R., Franca, L. P., and Mallet, M., "A new finite element formulation for computational fluid dynamics: I. symmetric forms of the compressible Navier-Stokes equations and the second law of thermodymaics," *Computer Methods in Applie Mechanics and Engineering*, Vol. 54, No. 2, April 1986, pp. 223–234.

[2] Barth, T. J., "Numerical methods for gasdynamic systems on unstructured meshes," *An introduction to recent developments in theory and numerics for conservation laws*, Springer, 1999, pp. 195–285.

[3] Del Rey Fernández, D. C., Hicken, J. E., and Zingg, D. W., "Review of summation-by-parts operators with simultaneous approximation terms for the numerical solution of partial differential equations," *Computers & Fluids*, Vol. 95, No. 22, 2014, pp. 171–196.

[4] Svärd, M. and Nordström, J., "Review of summation-by-parts schemes for initial-boundary-value-problems," *Journal of Computational Physics*, Vol. 268, No. 1, 2014, pp. 17–38.

[5] Fisher, T. C. and Carpenter, M. H., "High-order entropy stable finite difference schemes for nonlinear conservation laws: Finite domains," *Journal of Computational Physics*, Vol. 252, No. 1, 2013, pp. 518–557.

[6] Ismail, F. and Roe, P. L., "Affordable, entropy-consistent Euler flux functions II: entropy production at shocks," *Journal of Computational Physics*, Vol. 228, No. 15, 2009, pp. 5410–5436.

[7] Carpenter, M. H., Parsani, M., Nielsen, E. J., and Fisher, T. C., "Towards an Entropy Stable Spectral Element Framework for Computational Fluid Dynamics," *54th AIAA Aerospace Sciences Meeting*, American Institute of Aeronautics and Astronautics, Jan. 2016.

[8] Hicken, J. E., Del Rey Fernández, D. C., and Zingg, D. W., "Multi-dimensional Summation-By-Parts Operators: General Theory and Application to Simplex Elements," *SIAM Journal on Scientific Computing*, Vol. 38, No. 4, 2016, pp. A1935–A1958.

[9] Tadmor, E., "Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems," *Acta Numerica*, Vol. 12, 5 2003, pp. 451–512.

[10] Funaro, D. and Gottlieb, D., "A new method of imposing boundary conditions in pseudospectral approximations of hyperbolic equations," *Mathematics of Computation*, Vol. 51, No. 184, Oct. 1988, pp. 599–613.

[11] Carpenter, M. H., Gottlieb, D., and Abarbanel, S., "Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: Methodology and application to high-order compact schemes," *Journal of Computational Physics*, Vol. 111, No. 2, 1994, pp. 220–236.

[12] Carpenter, M. H., Nordström, J., and Gottlieb, D., "Revisiting and extending interface penalties for multi-domain summation-by-parts operators," *Journal of Scientific Computing*, Vol. 45, No. 1, June 2010, pp. 118–150.

[13] Lax, P., *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, Society for Industrial and Applied Mathematics, 1973.

[14] Tadmor, E., "Skew-selfadjoint form for systems of conservation laws," *Journal of Mathematical Analysis and Applications*, Vol. 103, No. 2, Oct. 1984, pp. 428–442.

[15] Harten, A., "On the symmetric form of systems of conservation laws with entropy," *Journal of Computational Physics*, Vol. 49, No. 1, 1983, pp. 151–164.

[16] Tadmor, E., "The numerical viscosity of entropy stable schemes for systems of conservation laws I," *Mathematics of Computation*, Vol. 49, No. 179, July 1987, pp. 91–103.

[17] Dafermos, C. M., *Hyperbolic Conservation Laws in Continuum Physics*, Vol. 325, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[18] Svärd, M., "Weak solutions and convergent numerical schemes of modified compressible NavierStokes equations," *Journal of Computational Physics*, Vol. 288, May 2015, pp. 19–51.

[19] Carpenter, M. H., Nordström, J., and Gottlieb, D., "A stable and conservative interface treatment of arbitrary spatial accuracy," *Journal of Computational Physics*, Vol. 148, No. 2, 1999, pp. 341–365.

[20] Del Rey Fernández, D. C., Boom, P. D., and Zingg, D. W., "A Generalized Framework for Nodal First Derivative Summation-By-Parts Operators," *Journal of Computational Physics*, Vol. 266, No. 1, 2014, pp. 214–239.

[21] Del Rey Fernández, D. C., Hicken, J. E., and Zingg, D. W., "Simultaneous approximation terms for multi-dimensional summation-by-parts operators," *Journal of Scientific Computing*, 2016, submitted.

[22] Hicken, J. E., Del Rey Fernández, D. C., and Zingg, D. W., "Simultaneous approximation terms for multi-dimensional summation-by-parts operators," *46th AIAA Fluid Dynamics Conference*, Washington, DC, June 2016, AIAA–2016–3971.

[23] Chandrashekar, P., "Kinetic Energy Preserving and Entropy Stable Finite Volume Schemes for Compressible Euler and Navier-Stokes Equations," *Communications in Computational Physics*, Vol. 14, No. 5, 2013, pp. 12521286.

[24] Roe, P., "Approximate Riemann solvers, parameter vectors, and difference schemes," *Journal of Computational Physics*, Vol. 43, No. 2, 1981, pp. 357 – 372.

[25] Fisher, T. C., Carpenter, M. H., Nordstrm, J., Yamaleev, N. K., and Swanson, C., "Discretely conservative finite-difference formulations for nonlinear conservation laws in split form: Theory and boundary conditions," *Journal of Computational Physics*, Vol. 234, 2013, pp. 353 – 375.

[26] Gassner, G. J., Winters, A. R., and Kopriva, D. A., "Split Form Nodal Discontinuous Galerkin Schemes with Summation-by-parts Property for the Compressible Euler Equations," *Journal Computational Physics*, Vol. 327, No. C, Dec. 2016, pp. 39–66.

[27] Fisher, T. C., *High-order L2 stable multi-domain finite difference method for compressible flows*, Ph.D. thesis, Purdue University, 2012.

[28] Thomas, P. D. and Lombard, C. K., "Geometric conservation law and its application to flow computations on moving grids," *AIAA Journal*, Vol. 17, No. 10, Oct. 1979, pp. 1030–1037.

[29] Golub, G. H. and Van Loan, C. F., *Matrix Computations*, The John Hopkins University Press, 3rd ed., 1996.

[30] Crean, J., Hicken, J. E., Del Rey Fernández, D. C., Zingg, D. W., and Carpenter, M., chap. High-Order, Entropy-Conservative Discretizations of the Euler Equations for Complex Geometries, 23rd AIAA CFD Conference, American Institute of Aeronautics and Astronautics, Jun 2017.

[31] Carpenter, M. H. and Kennedy, C. A., "Fourth-Order 2N-Storage Runge-Kutta Schemes," Tech. rep., NASA, Nasa Langley Research Center, Hampton, Virginia, June 1994.

[32] Chan, J., "On Discretely Entropy conservative and Entropy Stable Discontinuous Galerkin Methods," *arXiv*, 2017, https://arxiv.org/abs/1708.01243.

[33] Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B., "Julia: A Fresh Approach to Numerical Computing," *SIAM Review*, Vol. 59, No. 1, 2017, pp. 65–98.

[34] Seol, S., Smith, C. W., Ibanez, D. A., and Shephard, M. S., "A Parallel Unstructured Mesh Infrastructure," *High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:*, Nov 2012, pp. 1124–1132.

[35] Mattsson, K., Svärd, M., Carpenter, M., and Nordström, J., "High Order Accurate Computations for Unsteady Aerodynamics," *Computers and Fluids*, Vol. 36, No. 3, 2007, pp. 636–649.

[36] Erlebacher, G., Hussaini, M. Y., and Shu, C.-W., "Interaction of a shock with a longitudinal vortex," *Journal of Fluid Mechanics*, Vol. 337, Apr 1997, pp. 129–153.

[37] Doi, A. and Koide, A., "An Efficient Method of Triangulation Equi-Valued Surfaces by Using Tetrahedral Cells," *IEICE Transactions on Information Systems*, Vol. E74-D, No. 1, January 1991, pp. 214–224.

[38] Taylor, G. and Green, A., "Mechanism of the Production of Small Eddies from Large Ones," *Proceedings of hte Royal Society of London*, Vol. 158, February 1937, pp. 499–521.

[39] Parsani, M., Carpenter, M. H., and Nielsen, E. J., "Entropy stable wall boundary conditions for the three-dimensional compressible Navier-Stokes equations," *Journal of Computational Physics*, Vol. 292, No. C, 2015, pp. 88–113.

[40] Hunter, J. D., "Matplotlib: A 2D graphics environment," *Computing In Science & Engineering*, Vol. 9, No. 3, 2007, pp. 90–95.