Provably Stable Discontinuous Spectral-Element Methods with the Summation-by-Parts Property: Unified Matrix Analysis and Efficient Tensor-Product Formulations on Curved Simplices

by

Tristan Montoya

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Institute for Aerospace Studies
University of Toronto

Provably Stable Discontinuous Spectral-Element Methods with the Summation-by-Parts Property: Unified Matrix Analysis and Efficient Tensor-Product Formulations on Curved Simplices

Tristan Montoya
Doctor of Philosophy
Institute for Aerospace Studies, University of Toronto, 2024

Abstract

Discontinuous spectral-element methods (DSEMs) provide a flexible high-order spatial discretization approach for time-dependent conservation laws, but are traditionally regarded as lacking robustness relative to conventional second-order numerical methods for such partial differential equations. This thesis describes a unifying matrix analysis framework based on the summation-by-parts (SBP) property for the construction and analysis of DSEMs which are provably stable for linear or nonlinear hyperbolic problems. Within such a framework, we obtain algebraic proofs of conservation and energy stability as well as of the discrete equivalence between strong and weak formulations for discontinuous Galerkin and flux reconstruction schemes on general element types. The proposed methodology is also applied to the development of a new class of efficient and robust DSEMs on triangles and tetrahedra through the construction of sparse tensor-product operators in collapsed coordinates which satisfy the SBP property and are amenable to efficient sum-factorization algorithms. These SBP operators are used to construct split-form and flux-differencing DSEMs on curved triangular and tetrahedral unstructured grids which are conservative, free-stream preserving, and energy-stable for the linear advection equation or entropy stable for nonlinear hyperbolic systems. By exploiting the structure of the proposed operators as well as that of the Proriol–Koornwinder–Dubiner polynomial basis and using a weight-adjusted approximation to avoid the inversion of the curvilinear modal mass matrix, we obtain fully explicit algorithms of reduced complexity which facilitate the efficient extension of provably stable DSEMs on triangles and tetrahedra to arbitrarily high-order accuracy. The proposed schemes are compared to non-tensorial energy- and entropy-stable formulations on triangles and tetrahedra in a series of numerical experiments involving the solution of the linear advection and compressible Euler equations.

Acknowledgements

I would first like to thank my supervisor, Prof. David Zingg, for providing me with the freedom to pursue my research interests as well as the guidance I needed to keep my goals in sight. Prof. Zingg's commitment to the pursuit of a deeper understanding and his insight in asking the key questions which drive a research project forward set an example to which I will forever aspire. I am also thankful for the opportunities Prof. Zingg has given me to present my work at conferences around the world.

Next, I would like to thank Prof. Gianmarco Mengaldo for the enlightening conversations on various research topics and for being such a great host during my month in Singapore. I'd also like to thank my Doctoral Examination Committee members, Prof. Masayuki Yano, Prof. Prasanth Nair, and Prof. Christina Christara, for their feedback on this work as well as for the excellent courses I had the opportunity to take with each of them. I am grateful as well for the feedback and encouragement provided by my external examiner, Prof. Jesse Chan, and my internal examiner, Prof. Kirill Serkh.

I extend my gratitude to all my UTIAS colleagues for their advice, mentorship, encouragement, and friendship over the years. In particular, I'd like to thank Zelalem, Alex, André, and Aiden for sharing this journey with me. The work of the administrative and building staff to keep things running smoothly is also appreciated, as is the financial support from the University of Toronto, the Ontario Graduate Scholarship, and the Natural Sciences and Engineering Research Council of Canada. The computing resources provided by the SciNet High-Performance Computing Consortium are gratefully acknowledged as well.

Many thanks go out to my Ottawa friends, including Rafi, Russell, Anton, Silvia, Justin, and Nasir. The bike rides, jam sessions, ski days, board games, coffee, walks, and all-you-can-eat sushi dinners with you made my life so full during the past few years in "the city that fun forgot."

Finally, I would like to dedicate this thesis to my family, to whom I owe everything. To my mother, Heather, my father, Frédéric, my sister, Tamara, and to my extended family and in-laws, there is no way I can thank you all enough for your unwavering support. To Archie, Leo, and Willow, thank you for always making me smile and for teaching me to find joy in the little things. To Paige, thank you for your love and companionship, and for the confidence you have in me even when I lack it in myself; I couldn't have done it without you.

Statement of contributions

This thesis is based on content which has appeared in the following co-authored publications, with portions reproduced in accordance with the appropriate copyright policies.

(I) T. Montoya and D. W. Zingg. "A unifying algebraic framework for discontinuous Galerkin and flux reconstruction methods based on the summation-by-parts property." *Journal of Scientific Computing*, vol. 92, no. 3, article no. 87, 2022.

(II) T. Montoya and D. W. Zingg. "Stable and conservative high-order methods on triangular elements using tensor-product summation-by-parts operators." *Eleventh International Conference on Computational Fluid Dynamics*, 2022.

(III) T. Montoya and D. W. Zingg. "Efficient tensor-product spectral-element operators with the summation-by-parts property on curved triangles and tetrahedra." *SIAM Journal on Scientific Computing*, vol. 46, no. 4, pp. A2270–A2297, 2024.

(IV) T. Montoya and D. W. Zingg. "Efficient entropy-stable discontinuous spectral-element methods using tensor-product summation-by-parts operators on triangles and tetrahedra." In revision for *Journal of Computational Physics*, 2024.

In all of the above, the present author was responsible for the mathematical analysis, software implementation, and numerical verification, as well as writing complete drafts of the manuscripts and producing all tables and figures. The project conceptualization, methodological development, as well as the review and revision of manuscripts are the work of the present author in collaboration with David W. Zingg, who supervised this research.

Contents

List of figures

List of abbreviations

DG	Discontinuous Galerkin
DSEM	Discontinuous spectral-element method
FR	Flux reconstruction
JG	Jacobi–Gauss
JGL	Jacobi–Gauss–Lobatto
JGR	Jacobi–Gauss–Radau
KHI	Kelvin–Helmholtz instability
LG	Legendre–Gauss
LGL	Legendre–Gauss–Lobatto
LGR	Legendre–Gauss–Radau
ODE	Ordinary differential equation
PDE	Partial differential equation
PKD	Proriol–Koornwinder–Dubiner
SAT	Simultaneous approximation term
SBP	Summation-by-parts
SEM	Spectral-element method
SPD	Symmetric positive definite
SPSD	Symmetric positive semidefinite
TGV	Taylor–Green vortex
VCJH	Vincent–Castonguay–Jameson–Huynh

Chapter 1

Introduction

Hyperbolic or advection-dominated systems of time-dependent conservation laws constitute a class of partial differential equations (PDEs) of considerable importance in numerous scientific and engineering disciplines. For example, electromagnetic, geophysical, and acoustic wave propagation can be modelled by linear hyperbolic systems, whereas atmospheric, aerodynamic, and magnetohydrodynamic fluid flow problems are typically modelled by nonlinear hyperbolic and mixed hyperbolic-parabolic systems. These PDEs present significant challenges in the design of efficient, automated, and robust numerical methods, especially when a wide range of spatial and temporal scales are present. Although such behaviour is particularly characteristic of nonlinear problems, for example, in the context of scale-resolving simulations of turbulent flow (see, for example, Wang *et al.* [1]), similar challenges arise when simulating linear wave propagation over distances much larger than the wavelength (see, for example, Kreiss and Oliger [2] as well as Zingg [3]). *Discontinuous spectral-element methods* (DSEMs) have emerged as an attractive numerical approach for these problems, where we use such terminology in a general sense to refer to any spatial discretization with the following characteristics.

- The spatial domain is subdivided into elements, each containing one or more degrees of freedom (e.g. nodal values or modal expansion coefficients). Local refinement can be performed either by decreasing the element size (h-refinement) or increasing the order of the approximation within an element (p-refinement).
- The numerical solution is not required to be continuous at element interfaces. Boundary conditions and inter-element coupling are imposed weakly using numerical fluxes or penalty terms, aside from which all operations are entirely local to a given element.

These methods have received considerable attention in recent years due to their performance on modern hardware (see, for example, Klöckner *et al.* [4], Abdi *et al.* [5], and Vermiere *et al.* [6]) resulting from their relatively high arithmetic intensity (i.e. the ratio of floating-point operations to memory accesses) and data locality. Moreover, DSEMs are highly flexible in their support for general unstructured grids as well as their amenability to adaptation,

facilitating the efficient and automated solution of physically and geometrically complex problems, as exemplified in recent work by Parsani *et al.* [7] and Mossier *et al.* [8].

Although they offer the potential for improved efficiency relative to the low-order finite-difference, finite-volume, and finite-element methods (i.e. those of at most second-order accuracy) which continue to drive the majority of computational physics simulations, DSEMs, like most high-order methods, are regarded to be more prone to numerical instability when applied to nonlinear or variable-coefficient problems as well as when using curvilinear meshes. While the use of an upwind numerical flux at element interfaces often provides sufficient dissipation to obtain a stable simulation, practitioners applying DSEMs to problems involving under-resolved scales often employ additional regularization techniques such as modal filtering or de-aliasing approaches based on over-integration to address such robustness issues in practice (see, for example, Gassner and Beck [9] and Mengaldo *et al.* [10]). However, regularization approaches generally require problem-specific parameter tuning to ensure that the enhanced robustness is obtained without significant detriment to the accuracy of the simulation,[1] whereas over-integration incurs significant computational expense while still not ensuring stability for nonlinear systems (see, for example, Winters *et al.* [12]).

Modern formulations based on the summation-by-parts (SBP) property produce mathematical guarantees that discretizations will respect certain auxiliary integral balances or invariants satisfied by the linear or nonlinear PDEs they approximate, such as those governing the conservation or dissipation of energy or entropy. Satisfying such auxiliary balances ensures that important physical constraints such as the second law of thermodynamics cannot be violated by a numerical method, and there is strong numerical evidence suggesting that such schemes are more robust than conventional high-order methods, for which such balances are not guaranteed to hold (see, for example, the review by Gassner and Winters [13]). Energy and entropy balances are also important in a mathematical sense as they can, under certain conditions, enable one to bound the numerical solution *a priori* in terms of the initial and boundary data to obtain a *provably stable* discretization in a rigorous sense. Moreover, the theoretical underpinnings of the SBP approach are applicable to a wide variety of numerical methods, relying on the algebraic properties of the matrix operators which constitute a given discretization, rather than the specifics of how such a discretization was constructed. The SBP property can therefore be viewed as providing a *unifying* algebraic framework for the construction and analysis of numerical methods for conservation laws. This perspective is the focus of the present work, in which we aim to develop a unifying matrix analysis framework for energy-stable and entropy-stable DSEMs and to apply such a framework to the characterization of existing methods as well as the construction of novel discretizations.

[1] Highlighting the need for a careful balance between accuracy and robustness, Hesthaven and Warburton recommend in [11, Section 5.3] that one "filter as little as possible [...] but as much as is needed."

1.1 Background

Before outlining the main contributions of this thesis, it is helpful to review some of the historical development of finite-element and spectral methods, discontinuous spectral-element methods, sum-factorization algorithms, summation-by-parts operators, and entropy-stable schemes, in order to situate such contributions within the context of the broader literature.

1.1.1 Finite-element and spectral methods

Modern spectral-element methods are the result of a confluence of advances in spectral and finite-element methods. These methods share a common origin in the work of Bubnov [14], who in 1913 recognized that the *Ritz method* [15], which seeks an approximate solution to a boundary-value problem based on a principle of energy minimization, could be rewritten as a statement of orthogonality between the residual (i.e. the amount by which the numerical solution fails to satisfy the exact differential equation) and the basis functions used to approximate the solution. Galerkin then used Bubnov's formulation to extend the Ritz method to a much broader class of problems for which no corresponding minimization principle exists [16], and the resulting approximation technique would come to be known as a *Bubnov–Galerkin method* or simply a *Galerkin method*. The reader is referred to Gander and Wanner [17] for an overview of these early contributions.

The first numerical schemes resembling modern finite-element methods were introduced in the early 1940s, with Courant's description of an approximation using piecewise linear functions with compact support on a mesh of triangles often cited as the originating work [18]. Similar ideas, however, were proposed independently by structural engineers (see, for example, Hrennikoff [19]), and the 1950s and 1960s would see significant progress in the development of finite-element methods for structural analysis, enabled by the increasing availability and processing power of digital computers. We note important contributions during this period by Turner *et al.* [20], Argyris [21], and Zienkiewicz and Cheung [22].

Beginning in the late 1960s with the influential work of Orszag [23], spectral methods based on high-order algebraic or trigonometric polynomials with global support emerged as a popular numerical approach for time-dependent conservation laws, especially for applications in fluid mechanics, differing substantially from the aforementioned finite-element methods which had become commonplace in structural applications. By the 1980s, however, advances in finite-element methods began to trend from low-order piecewise polynomial approximations towards those of higher order, known as p-version and hp-version finite elements (see, for example, the review by Babuška and Suri [24]). Meanwhile, multidomain approaches were introduced within the context of spectral methods to facilitate the treatment of complex geometries and to enable local refinement, with an early example being the work of Orszag

[25]. These developments can be interpreted as the convergence of spectral and finite-element methods to a new class of schemes combining the accuracy of spectral methods with the flexibility of finite-element methods. In this thesis, we use the term *spectral-element method* (SEM), originally coined by Patera [26], to refer to this entire family of methods, noting that some authors adopt a narrower usage of such terminology, specifically referring to collocated nodal formulations employing tensor-product Gaussian quadrature rules. We refer to Young [27] for a historical perspective and some clarification regarding the terminology employed in the literature relating to collocation-type spectral and spectral-element approximations, and we note that the first multidomain formulations of such schemes were proposed for one-dimensional problems by Wheeler [28] and Díaz [29].

1.1.2 Discontinuous spectral-element methods

Characterized by a Galerkin formulation (i.e. requiring the residual to be orthogonal to test functions belonging to the same space as the numerical solution) permitting discontinuities between elements, the first *discontinuous Galerkin* (DG) methods for hyperbolic PDEs were introduced by Reed and Hill in 1973 to solve the steady neutron transport equation [30]. However, it was not until the work of Cockburn, Shu, and collaborators in the late 1980s and 1990s [31–34] combining a DG spatial discretization with explicit Runge-Kutta time integration and slope limiting that such schemes would begin to gain traction for time-dependent nonlinear systems. A parallel approach to the construction of DSEMs in differential form would also materialize in the 1990s and 2000s as a multidomain evolution of the penalty-type boundary treatment for spectral methods introduced by Funaro and Gottlieb [35], with popular schemes of such lineage including staggered-grid collocation methods [36] and spectral-difference methods [37, 38].

Introduced by Huynh in 2007 [39], the *flux reconstruction* (FR) approach provides a general methodology for constructing a variety of DSEMs in differential form through the use of *correction functions* to reconstruct a continuous flux from the discontinuous numerical solution. Initially proposed in one spatial dimension, FR methods were subsequently extended to triangular elements by Wang and Gao [40] through the so-called lifting collocation penalty formulation and further developed by several other authors, including Vincent *et al.* [41], Castonguay *et al.* [42], and Williams and Jameson [43], who identified one-parameter families of FR schemes on one-dimensional, triangular, and tetrahedral elements which are energy stable with respect to particular broken norms of Sobolev type. These energy-stable FR methods are collectively termed *Vincent–Castonguay–Jameson–Huynh* (VCJH) schemes.

Although DG methods are derived from a weak (i.e. variational) formulation whereas FR methods are derived from a strong (i.e. differential) formulation, the two approaches are

closely related [44–47], with the most popular choice of FR correction function being that which recovers a strong-form nodal DG method. Furthermore, like all DSEM formulations, DG and FR methods are both highly amenable to efficient implementation on modern massively parallel computer architectures, particularly in conjunction with explicit time integration, with all operations aside from the numerical flux evaluation consisting entirely of local computations employing highly structured memory access patterns with minimal indirection. Such arithmetically intense local operations enable the hiding of communication latency when used within a distributed-memory environment, resulting in parallel algorithms which scale efficiently to hundreds of thousands of cores. Examples of high-performance DSEM implementations which have been successfully applied to large-scale problems include the open-source `Nektar++` [48, 49], `PyFR` [50], and `FLEXI` [51] solvers.

1.1.3 Sum-factorization algorithms

Spectral and spectral-element methods on domains which are diffeomorphic to the square or cube offer the natural advantage of supporting *sum-factorization* algorithms for the efficient evaluation of tensor-product operators, an approach first described in the context of spectral methods by Orszag [25]. In such algorithms, multidimensional approximation procedures (e.g. those involving differentiation or interpolation) using tensor-product bases are decomposed so as to consist of individual one-dimensional operations, similarly to the application of one-dimensional stencils in a tensor-product finite-difference scheme. As such algorithms do not require explicitly forming multidimensional operator matrices, the sum factorization is often described as a *matrix-free* approach (see, for example, Kronbichler and Kormann [52]). Considering a polynomial spectral-element approximation in d dimensions, local operations such as spatial differentiation are typically of $\mathcal{O}(p^{2d})$ time complexity with respect to the polynomial degree p when no such tensor-product structure is exploited, whereas sum factorization generally results in algorithms of $\mathcal{O}(p^{d+1})$ complexity.[2] Sum factorization is therefore considered to be essential to maximizing the efficiency of spectral-element methods at high polynomial degrees, although, as discussed by Świrydowicz *et al.* [53], achieving optimal performance requires some degree of hardware-specific optimization.

The restriction of the classical sum-factorization technique to quadrilaterals in two dimensions and hexahedra in three dimensions limits the geometric flexibility of the resulting schemes, whereas the $\mathcal{O}(p^{2d})$ complexity of non-tensor-product discretizations restricts the range of polynomial degrees which can be efficiently utilized, for example, on triangles and tetrahedra when using inherently multidimensional formulations. This tradeoff can

[2] Unless otherwise specified (e.g. when discussing memory usage), the term *complexity* is taken in this thesis to denote the asymptotic time complexity of an algorithm on the basis of operation count.

be reconciled through the use of a *collapsed coordinate transformation*, sometimes referred to as a *Duffy transformation* [54], which allows for the geometric flexibility of simplicial (e.g. triangular and tetrahedral) elements as well as other element types such as prisms and pyramids to be combined with the aforementioned computational benefits of a tensor-product operator structure. Beginning with theoretical development by Dubiner [55] and the subsequent application of such ideas to continuous Galerkin [56, 57] as well as discontinuous Galerkin methods [58, 59], collapsed-coordinate formulations have enabled the construction of efficient discretizations of arbitrary order on general element types (see, for example, Vos *et al.* [60] and Cantwell *et al.* [61]). Furthermore, such schemes have been shown to be amenable to efficient implementation on modern hardware (see, for example, Moxey *et al.* [62]).

1.1.4 Summation-by-parts operators

A central component of the framework described in this thesis is the SBP property, which was introduced in 1974 by Kreiss and Scherer [63] in order to obtain energy-stable high-order finite-difference methods (i.e. those for which a discrete L^2 norm is bounded in terms of the problem data) for linear hyperbolic problems. By mimicking the integration-by-parts property of the derivative operator using discrete inner products, the continuous energy estimates satisfied by standard Galerkin finite-element methods could be extended in a straightforward manner to finite-difference schemes. Matrix difference operators satisfying such an analogue of integration by parts are commonly referred to as *summation-by-parts operators*, for which boundary conditions and inter-block coupling are typically imposed weakly using *simultaneous approximation terms* (SATs), which were developed by Carpenter *et al.* [64] based on the penalty approach introduced in [35]. We refer to the review papers by Del Rey Fernández *et al.* [65] and Svärd and Nordström [66] for more information regarding the development, analysis, and application of SBP-SAT finite-difference methods.

Although introduced within a finite-difference setting, the SBP property provides a general methodology for the construction and analysis of a very broad class of discretizations. Building on the work of Carpenter and Gottlieb [67] and Kopriva and Gassner [68], the connection between the SBP property and DSEMs, which forms the crux of this thesis, was first made explicit in a 2013 paper by Gassner [69], who recognized that the matrix operators employed within DG methods based on collocated Legendre–Gauss–Lobatto quadrature are, in fact, SBP operators. In the same work, Gassner exploited this equivalence in order to construct discretely conservative and nonlinearly stable DSEMs for Burgers' equation by way of a skew-symmetric *split formulation* resulting from the discretization of a judiciously chosen linear combination of analytically equivalent but numerically distinct forms of the PDE (see, for example, Pirozolli [70] or Fisher [71]). Again inspired by developments in SBP

finite-difference methods [72, 73], Kopriva and Gassner used a similar splitting to obtain energy-stable tensor-product DSEMs for variable-coefficient advection problems on curvilinear meshes [74]. Gassner would also use this approach to construct a kinetic-energy-preserving DSEM for the Euler equations [75] based on a split formulation introduced for finite-difference methods by Morinishi [76]. An important aspect of Gassner and Kopriva's approach is the fact that the proofs of stability are valid despite the inexactness of the collocated quadrature rules, which, while often advantageous from an efficiency perspective, are otherwise prone to inducing aliasing-driven instabilities (see, for example, Kirby and Karniadakis [77]).

A generalized theoretical framework for nodal SBP operators in one dimension (encompassing spectral-element as well as finite-difference approximations) was presented in 2014 by Del Rey Fernández *et al.* [78] based on the connection between SBP operators and quadrature rules established in a classical finite-difference setting by Hicken and Zingg [79]. Notably, such a framework extends the SBP approach to nodal sets which do not include one or both endpoints of the interval, including collocated spectral-element operators on Legendre–Gauss quadrature nodes, which are often used for the construction of efficient tensor-product DG methods (see, for example, Black [80] and Hindenlang *et al.* [81]). A similar methodology was adopted by Ranocha *et al.* in [82], who used such a generalized notion of the SBP property for the analysis of one-dimensional FR methods. In 2016, Hicken *et al.* [83] presented a further generalization of the SBP approach to multidimensional operators on general element types including triangles and tetrahedra, which was used by Del Rey Fernández *et al.* in [84] along with suitable discontinuous coupling procedures to obtain energy-stable and conservative discretizations of variable-coefficient advection problems in split form on curved simplicial meshes. While advantageous in terms of geometric flexibility relative to tensor-product operators on quadrilaterals and hexahedra, the multidimensional SBP operators constructed in [83, 84] and subsequent work are based on non-tensorial formulations of $\mathcal{O}(p^{2d})$ complexity, restricting their practical use to modest polynomial degrees.

1.1.5 Entropy-stable discretizations

Although stable discretizations of certain nonlinear or variable-coefficient PDEs can be obtained by way of a split formulation using SBP operators, the extension of such an approach to more complex systems of conservation laws requires the use of *entropy-conservative two-point flux functions*, which were proposed by Tadmor [85] in the context of first-order and second-order finite-volume methods. These flux functions guarantee (under certain physical admissibility criteria) that a strictly convex *entropy function* remains bounded from above for all time, mimicking the well-known entropy analysis for nonlinear hyperbolic PDEs (see, for example, Lax [86]). Tadmor's approach was later extended to high-order

accuracy on one-dimensional periodic domains by LeFloch *et al.* [87]. The modern era of entropy-stable discretizations, however, began in 2012 when Fisher [88] combined affordable entropy-conservative flux functions proposed by Ismail and Roe [89] with SBP operators in order to obtain entropy-stable high-order finite-difference methods for the compressible Euler and Navier–Stokes equations on curvilinear block-structured grids (see also Fisher and Carpenter [90] and Fisher *et al.* [71]). Such a combination of SBP operators with two-point flux functions came to be known in the entropy stability community as *flux differencing*, a term not to be confused with the similarly named *flux-difference splitting* technique introduced by Roe [91] decades earlier for approximate Riemann solvers.

Building on Fisher's work in the context of finite-difference methods, entropy-stable DSEMs for systems of conservation laws on tensor-product quadrilateral and hexahedral elements were introduced by Carpenter *et al.* [92] and Gassner *et al.* [93], with the latter demonstrating that a broad class of split-form and entropy-stable DSEMs (including those in [75]) could be recovered by choosing different two-point flux functions. These schemes are distinguished from the approach taken by Barth [94] as well as Hiltebrand and Mishra [95] based on the work of Hughes *et al.* [96], wherein space-time DG schemes are formulated in terms of the entropy variables. Unlike flux-differencing DSEMs, the latter methodology results in discretizations which are only entropy stable under the assumption that all integrals in the corresponding variational formulation are evaluated exactly (which is impractical, if not impossible, for many PDEs of interest to practitioners, including the compressible Euler and Navier–Stokes equations) and, furthermore, cannot be formulated explicitly in time due to the dependence of the mass matrix on the solution state.

Entropy-stable DSEMs on simplicial elements using multidimensional SBP operators were first introduced by Crean *et al.* [97], Chen and Shu [98], and Chan [99], and are systematically reviewed by Chen and Shu [100]. In the context of an entropy-stable scheme, multidimensional (i.e. non-tensor-product) SBP operators on triangles and tetrahedra require the evaluation of entropy-conservative two-point flux functions between all pairs of quadrature nodes, rather than simply along lines of nodes as in a tensor-product formulation. Since in the case of the Euler and Navier–Stokes equations, the evaluation of an entropy-conservative flux is a relatively expensive operation involving the logarithmic mean, this represents a significant increase in computational cost relative to discretizations on quadrilaterals and hexahedra, particularly at high polynomial degrees. Due in part to these limitations, as well as the difficulty in constructing suitable quadrature rules, entropy-stable discretizations based on multidimensional SBP operators rarely employ polynomial degrees greater than four or five.

1.2 Summary of contributions

Having reviewed the relevant literature and current state of the art, we are now equipped to summarize the main contributions of this thesis, which can be organized into efforts towards meeting three main objectives, which are listed and discussed below.

Development of a unifying matrix analysis framework for discontinuous spectral-element methods based on the summation-by-parts property (Chapter 3)

As discussed in [69] and [82], the stability properties of standard DG and FR methods can be understood as a consequence of the SBP property, at least in the one-dimensional, collocated case. However, the role of the multidimensional SBP property in the analysis of more general DG formulations and the energy-stable FR methods introduced for simplicial meshes in [41–43] had not been fully explored prior to the present work. We address this in the present work by reformulating the standard DG and FR methods for curvilinear unstructured grids in terms of a common set of matrix operators on the reference element and demonstrating that, under certain conditions, they can be unified as SBP schemes within a more general algebraic framework. As a result of this unification, the existing equivalences between the DG and FR methods established in [44–47] are reinterpreted from the perspective of the SBP property. Furthermore, we demonstrate algebraically that the resulting schemes are locally (i.e. element-wise) and globally conservative as well as energy stable with respect to suitable quadrature rules and discrete norms, recovering the analysis in [41–43] within the context of a more general and arguably simpler theoretical framework. Besides providing a new perspective to the analysis of standard DG and FR methods, the proposed framework constitutes a general methodology for the formulation and analysis of DSEMs based on the SBP property and serves as a unifying theoretical backbone for the construction of the novel schemes which are the focus of the balance of this thesis.

Construction of tensor-product spectral-element operators with the summation-by-parts property on the reference triangle and tetrahedron (Chapter 4)

To the author's knowledge, all existing high-order SBP operators on triangles and tetrahedra proposed prior to the present work are based on inherently multidimensional formulations which are dense and lack a tensor-product structure, which results in matrix operations of $\mathcal{O}(p^{2d})$ complexity, comparing unfavourably to the $\mathcal{O}(p^{d+1})$ complexity of tensor-product formulations on quadrilaterals and hexahedra resulting from sum factorization. In the context of an entropy-stable scheme, the number of required entropy-conservative flux evaluations scales similarly, further disadvantaging the multidimensional approach as a result of the

dense coupling between each pair of quadrature nodes. In this thesis, we describe a critical step towards addressing such limitations through the construction of SBP operators of arbitrary order on the reference triangle and tetrahedron which are sparse and possess a tensor-product structure amenable to sum factorization. These operators are obtained through a collocation approach based on the use of tensor-product Gaussian quadrature rules in collapsed coordinates, corresponding to a rational approximation which remains exact for polynomials on the triangle or tetrahedron. Our approach differs from existing work such as [56] or [57] in that the Jacobian determinant of the collapsed coordinate transformation is *not* entirely subsumed by a Jacobi-type quadrature weight, allowing for a cancellation of the singular factors arising from the use of the chain rule to differentiate on the reference element. This cancellation, along with the careful construction of facet quadrature rules and interpolation/extrapolation operators, results in sparse tensor-product operators of any order which satisfy the SBP property on the reference triangle or tetrahedron, and are therefore suitable for the construction of efficient energy-stable and entropy-stable schemes on curvilinear unstructured simplicial meshes.

Development of efficient energy- and entropy-stable tensor-product discontinuous spectral-element methods on curved triangles and tetrahedra (Chapters 5 to 7)

Using the proposed tensor-product SBP operators within split-form and flux-differencing DSEM formulations built upon the generalized framework described above, we construct discretizations on curvilinear triangular and tetrahedral unstructured grids which are free-stream preserving, conservative, as well as energy stable for the linear advection equation or entropy stable for nonlinear hyperbolic systems of conservation laws. Although a collocated nodal formulation can be constructed using the proposed tensor-product operators, our main focus is on a modal approach based on a projection onto a particular orthonormal polynomial basis on the reference triangle or tetrahedron which is separable with respect to the collapsed coordinate system. This projection onto a standard total-degree polynomial space on the reference element alleviates the time step restriction associated with the singularity of the collapsed coordinate transformation while retaining the advantages of the proposed tensor-product operators for nodal operations such as differentiation and interpolation/extrapolation. While the use of curved elements within a modal formulation results in a dense local mass matrix for each element, which is unfavourable in the context of explicit time integration, the inversion of such a matrix is avoided in this work through the use of a weight-adjusted approximation from Chan *et al.* [101], which results in the overall cost of evaluating the time derivative on each element scaling as $\mathcal{O}(p^{d+1})$, with optimal $\mathcal{O}(p^d)$ storage requirements.

The proposed energy-stable and entropy-stable schemes are implemented within a flexible open-source PDE solver written in the Julia programming language [102], for which the

software design closely parallels the unified framework described in this thesis. We present numerical experiments involving the solution of the linear advection and compressible Euler equations on curved triangular and tetrahedral meshes using the proposed energy-stable and entropy-stable DSEM formulations using tensor-product SBP operators as well as those using multidimensional SBP operators based on symmetric quadrature rules. These experiments indicate that for a given mesh and polynomial degree, the proposed tensor-product approach is very similar in accuracy to a comparable multidimensional formulation, and that the spectral radius of the semi-discrete advection operator, which dictates the maximum stable time step for explicit schemes, is similar for both approaches when a weight-adjusted modal formulation is used. Furthermore, the proposed approach is shown to result in highly robust schemes for challenging nonlinear problems characteristic of under-resolved turbulence. We also provide estimates of the computational expense of such formulations in terms of operation count as well as the required number of entropy-conservative two-point flux evaluations, which suggest that the proposed schemes have the potential for significantly improved efficiency at high polynomial degrees relative to existing provably stable formulations on simplicial elements.

Chapter 2

Mathematical preliminaries

The purpose of this chapter is to present some important mathematical notation and definitions used throughout the remainder of this thesis and to review the essential concepts which are required for one to grasp the contributions therein. Portions of this section are adapted from Papers I and IV.

2.1 Notation

In this thesis, mathematical symbols are generally defined as they appear, where we employ the following conventions throughout.

- Single underlines are used to denote vectors (treated as column matrices), whereas double underlines denote matrices. Symbols in bold are used specifically to denote Cartesian (i.e. spatial) vectors, for which we employ the usual dot product $\boldsymbol{x} \cdot \boldsymbol{y} \coloneqq x_1 y_1 + \cdots + x_d y_d$ and Euclidean norm $\|\boldsymbol{x}\|^2 \coloneqq \boldsymbol{x} \cdot \boldsymbol{x}$.
- The symbol ∇ is used to denote componentwise differentiation with respect to either type of vector, for example, as $\nabla_{\boldsymbol{x}} \coloneqq [\partial/\partial_{x_1}, \ldots, \partial/\partial_{x_d}]^{\mathrm{T}}$ or $\nabla_{\underline{U}} \coloneqq [\partial/\partial_{U_1}, \ldots, \partial/\partial_{U_d}]^{\mathrm{T}}$.
- The symbols $\mathbb{R}$, $\mathbb{R}^+$, $\mathbb{R}_0^+$, $\mathbb{N}$, $\mathbb{N}_0$, and $\mathbb{S}^{d-1}$ are used, respectively, to denote the real numbers, the positive real numbers, the non-negative real numbers, the natural numbers (excluding zero), the natural numbers including zero, and the unit $(d-1)$-sphere given by $\mathbb{S}^{d-1} \coloneqq \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = 1\}$.
- The symbols $\underline{0}^{(N)}$ and $\underline{1}^{(N)}$ are reserved for vectors of length N containing all zeros and all ones, respectively, and $\underline{\underline{I}}^{(N)}$ denotes the N by N identity matrix. The symbol $\underline{\underline{0}}^{(M \times N)}$ likewise denotes an M by N matrix of all zeros.
- A matrix $\underline{\underline{A}} \in \mathbb{R}^{N \times N}$ is *symmetric positive semidefinite* (SPSD) if it is symmetric (i.e. $\underline{\underline{A}}^{\mathrm{T}} = \underline{\underline{A}}$) and satisfies $\underline{v}^{\mathrm{T}} \underline{\underline{A}}\, \underline{v} \geq 0$ for all $\underline{v} \in \mathbb{R}^N$. Such a matrix is *symmetric positive definite* (SPD) if it is symmetric and satisfies $\underline{v}^{\mathrm{T}} \underline{\underline{A}}\, \underline{v} > 0$ for all $\underline{v} \in \mathbb{R}^N \setminus \{\underline{0}^{(N)}\}$.

- The notation $\{1:N\}$ is used as shorthand for the index set $\{1,2,\dots,N\}$. Given a set $\mathcal{X}$, we abuse the notation $\{x^{(i)}\}_{i\in\{1:N\}} \subset \mathcal{X}$ to denote a sequence of N elements satisfying $x^{(i)} \in \mathcal{X}$, where such elements are ordered and not necessarily distinct.
- Given a bounded domain $\Omega \subset \mathbb{R}^d$, we use $\partial\Omega$ to denote its boundary and $\bar{\Omega} \coloneqq \Omega \cup \partial\Omega$ to denote its closure; the interior of a closed domain Ω is then given by $\mathring{\Omega} \coloneqq \Omega \setminus \partial\Omega$.

2.2 Hyperbolic conservation laws

We are interested in systems of conservation laws governing the evolution of N_c *conservative variables* given by $\underline{U}(\boldsymbol{x},t) \in \Upsilon \subset \mathbb{R}^{N_c}$ on the spatial domain $\Omega \subset \mathbb{R}^d$ over the time interval $(0,T) \subset \mathbb{R}_0^+$, where Υ denotes the set of admissible solution states. Such PDEs can be expressed in the general form

$$\frac{\partial \underline{U}(\boldsymbol{x},t)}{\partial t} + \sum_{m=1}^{d} \frac{\partial \underline{F}_m(\underline{U}(\boldsymbol{x},t))}{\partial x_m} = \underline{0}^{(N_c)}, \qquad \forall\,(\boldsymbol{x},t) \in \Omega \times (0,T), \tag{2.1a}$$

$$\underline{U}(\boldsymbol{x},0) = \underline{U}^0(\boldsymbol{x}), \qquad \forall\,\boldsymbol{x} \in \Omega, \tag{2.1b}$$

subject to appropriate boundary conditions, where $\underline{F}_m(\underline{U}(\boldsymbol{x},t)) \in \mathbb{R}^{N_c}$ denotes the Cartesian flux component in the x_m direction, and $\underline{U}^0(\boldsymbol{x}) \in \Upsilon$ denotes the initial data. For convenience, we define the flux in any direction $\boldsymbol{n} \in \mathbb{S}^{d-1}$ for an arbitrary solution state $\underline{U} \in \Upsilon$ as

$$\underline{F}(\underline{U},\boldsymbol{n}) \coloneqq \sum_{m=1}^{d} n_m \underline{F}_m(\underline{U}). \tag{2.2}$$

A system of conservation laws in the form of (2.1) is then called *hyperbolic* if the flux Jacobian

$$\nabla_{\underline{U}} \underline{F}(\underline{U},\boldsymbol{n}) \coloneqq \Big[\nabla_{\underline{U}} F_1(\underline{U},\boldsymbol{n}), \dots, \nabla_{\underline{U}} F_{N_c}(\underline{U},\boldsymbol{n})\Big]^{\mathrm{T}} \tag{2.3}$$

is diagonalizable with all real eigenvalues for all states $\underline{U} \in \Upsilon$ and all directions $\boldsymbol{n} \in \mathbb{S}^{d-1}$. We are specifically interested in hyperbolic systems of conservation laws endowed with an *entropy function* and corresponding *entropy flux*, which are defined as follows.

Definition 2.1. The functions $\mathcal{S} : \Upsilon \to \mathbb{R}$ and $\boldsymbol{\mathcal{F}} : \Upsilon \to \mathbb{R}^d$ are, respectively, an *entropy function* and *entropy flux* if $\mathcal{S}$ is a strictly convex function (i.e. its Hessian is SPD) satisfying

$$\Big(\nabla_{\underline{U}} \mathcal{F}_m(\underline{U})\Big)^{\mathrm{T}} = \Big(\nabla_{\underline{U}} \mathcal{S}(\underline{U})\Big)^{\mathrm{T}} \Big(\nabla_{\underline{U}} \underline{F}_m(\underline{U})\Big), \quad \forall\,\underline{U} \in \Upsilon, \quad \forall\, m \in \{1:d\}, \tag{2.4}$$

where $\nabla_{\underline{U}} \mathcal{S}(\underline{U}), \nabla_{\underline{U}} \mathcal{F}_m(\underline{U}) \in \mathbb{R}^{N_c}$ denote the gradients of the entropy function and entropy flux components with respect to the conservative variables, and $\nabla_{\underline{U}} \underline{F}_m(\underline{U}) \in \mathbb{R}^{N_c \times N_c}$ denotes

the Jacobian of the m^{th} Cartesian flux component, which is defined similarly to (2.3).

The entries of the vector $\underline{\mathcal{W}}(\underline{U}) \coloneqq \nabla_{\underline{U}}\mathcal{S}(\underline{U})$ are referred to as the *entropy variables*, where the mapping $\underline{\mathcal{W}}$ has an inverse given by $\underline{\mathcal{U}}$ due to the strict convexity of $\mathcal{S}$ over the admissible set Υ. As described by Friedrichs and Lax [103], the existence of an entropy–entropy flux pair in the sense of Definition 2.1 implies that any classical (i.e. continuously differentiable) solution to (2.1) satisfies an auxiliary conservation law of the form

$$\frac{\partial \mathcal{S}(\underline{U}(\boldsymbol{x},t))}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\mathcal{F}}(\underline{U}(\boldsymbol{x},t)) = 0, \quad \forall\, (\boldsymbol{x},t) \in \Omega \times (0,T). \tag{2.5}$$

Integrating (2.5) over the spatial domain and using the divergence theorem then results in

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega} \mathcal{S}(\underline{U}(\boldsymbol{x},t))\,\mathrm{d}\boldsymbol{x} &= -\int_{\partial\Omega} \boldsymbol{\mathcal{F}}(\underline{U}(\boldsymbol{x},t)) \cdot \boldsymbol{n}(\boldsymbol{x})\,\mathrm{d}s \\ &= \int_{\partial\Omega} \Big(\boldsymbol{\Psi}\Big(\underline{\mathcal{W}}(\underline{U}(\boldsymbol{x},t))\Big) \cdot \boldsymbol{n}(\boldsymbol{x}) - \Big(\underline{\mathcal{W}}(\underline{U}(\boldsymbol{x},t))\Big)^{\mathrm{T}} \underline{F}(\underline{U}(\boldsymbol{x},t), \boldsymbol{n}(\boldsymbol{x})) \Big)\,\mathrm{d}s, \end{aligned} \tag{2.6}$$

where $\boldsymbol{n}(\boldsymbol{x}) \in \mathbb{S}^{d-1}$ denotes the outward unit normal to $\partial\Omega$, and the components of the *flux potential* $\boldsymbol{\Psi}(\underline{W}) \in \mathbb{R}^d$ are given by

$$\Psi_m(\underline{W}) \coloneqq \underline{W}^{\mathrm{T}} \underline{F}_m(\underline{\mathcal{U}}(\underline{W})) - \mathcal{F}_m(\underline{\mathcal{U}}(\underline{W})). \tag{2.7}$$

We are, however, often interested in *weak solutions*, which satisfy (2.1) in the sense of distributions and can therefore be discontinuous, describing phenomena such as shock waves. Replacing (2.6) with an *entropy inequality* of the form

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega} \mathcal{S}(\underline{U}(\boldsymbol{x},t))\,\mathrm{d}\boldsymbol{x} \leq \int_{\partial\Omega} \Big(&\boldsymbol{\Psi}\Big(\underline{\mathcal{W}}(\underline{U}(\boldsymbol{x},t))\Big) \cdot \boldsymbol{n}(\boldsymbol{x}) \\ &- \Big(\underline{\mathcal{W}}(\underline{U}(\boldsymbol{x},t))\Big)^{\mathrm{T}} \underline{F}(\underline{U}(\boldsymbol{x},t), \boldsymbol{n}(\boldsymbol{x})) \Big)\,\mathrm{d}s \end{aligned} \tag{2.8}$$

then provides an admissibility criterion for physically relevant weak solutions (see, for example, Kružkov [104] or Lax [105]). Provided that $\underline{U}(\boldsymbol{x},t)$ remains within Υ and that the boundary conditions are imposed correctly, it can be shown (see, for example, Dafermos [106]) that (2.8) implies a bound on the solution itself due to the strict convexity of the entropy function.

Remark 2.1. The requirement for a strictly convex entropy function in Definition 2.1 is consistent with the mathematical literature. However, such a convention is opposite that used in physics, wherein the Clausius–Duhem statement of the second law of thermodynamics can be expressed in the form of (2.8) with the direction of the inequality reversed.

2.3 Discontinuous spectral-element methods

In this section, we will review the essential components of discontinuous spectral-element formulations on general curvilinear unstructured grids, with the classical DG and FR methods serving as representative examples within such a class of methods.

2.3.1 Mesh and coordinate transformation

The first step in constructing a continuous or discontinuous spectral-element discretization of the conservation law in (2.1) is to subdivide the spatial domain Ω into a *mesh* or *grid* of characteristic element size $h > 0$, which consists of a collection $\{\Omega^{(\kappa)}\}_{\kappa\in\{1:N_e\}}$ of N_e closed, bounded, and connected elements $\Omega^{(\kappa)} \subset \Omega$ with nonempty interiors, satisfying

$$\bigcup_{\kappa=1}^{N_e} \Omega^{(\kappa)} = \bar{\Omega} \quad \text{and} \quad \mathring{\Omega}^{(\kappa)} \cap \mathring{\Omega}^{(\nu)} = \emptyset, \quad \forall\, \kappa \neq \nu. \tag{2.9}$$

In this work, we assume that each element is the image of a polytopal (i.e. polygonal in two dimensions or polyhedral in three dimensions) *reference element* denoted by $\hat{\Omega} \subset \mathbb{R}^d$ under a smooth, time-invariant mapping $\boldsymbol{X}^{(\kappa)} : \hat{\Omega} \to \Omega^{(\kappa)}$. Denoting the Jacobian of the mapping as

$$\boldsymbol{J}^{(\kappa)}(\boldsymbol{\xi}) \coloneqq \left[\nabla_{\boldsymbol{\xi}} X_1(\boldsymbol{\xi}), \ldots, \nabla_{\boldsymbol{\xi}} X_d(\boldsymbol{\xi})\right]^{\mathrm{T}} \tag{2.10}$$

and its determinant as $J^{(\kappa)}(\boldsymbol{\xi}) \coloneqq \det(\boldsymbol{J}^{(\kappa)}(\boldsymbol{\xi}))$, we assume that the transformation from reference to physical coordinates is bijective and orientation preserving, satisfying

$$J^{(\kappa)}(\boldsymbol{\xi}) > 0, \quad \forall\, \boldsymbol{\xi} \in \hat{\Omega}. \tag{2.11}$$

The adjugate of the Jacobian is then given by $\boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi}) \coloneqq J^{(\kappa)}(\boldsymbol{\xi})(\boldsymbol{J}^{(\kappa)}(\boldsymbol{\xi}))^{-1}$, with entries (often referred to as the *metric terms* in the literature) satisfying the *metric identities*

$$\sum_{l=1}^{d} \frac{\partial G_{lm}^{(\kappa)}(\boldsymbol{\xi})}{\partial \xi_l} = 0, \quad \forall\, \boldsymbol{\xi} \in \hat{\Omega}, \quad \forall\, m \in \{1:d\}, \tag{2.12}$$

which can be used to express (2.1a) in conservation form on the reference element (see, for example, Pulliam and Zingg [107, Section 4.2] or Kopriva [108, Section 6.2]) as

$$J^{(\kappa)}(\boldsymbol{\xi}) \frac{\partial \underline{U}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}), t)}{\partial t} + \sum_{l=1}^{d} \frac{\partial}{\partial \xi_l} \left(\sum_{m=1}^{d} G_{lm}^{(\kappa)}(\boldsymbol{\xi}) \underline{F}_m(\underline{U}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}), t)) \right) = \underline{0}^{(N_c)}. \tag{2.13}$$

Since the reference element is a polytope, we can partition its boundary into a N_f flat facets $\hat{\Gamma}^{(\zeta)} \subset \partial\hat{\Omega}$ with disjoint interiors and (constant) outward unit normal vectors denoted by $\hat{\boldsymbol{n}}^{(\zeta)} \in \mathbb{S}^{d-1}$. Defining $J^{(\kappa,\zeta)}(\boldsymbol{\xi}) \coloneqq \|\boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi})^{\mathrm{T}}\hat{\boldsymbol{n}}^{(\zeta)}\|$, the outward unit normal $\boldsymbol{n}^{(\kappa,\zeta)}(\boldsymbol{x}) \in \mathbb{S}^{d-1}$ to the facet $\Gamma^{(\kappa,\zeta)} \subset \partial\Omega^{(\kappa)}$ which is the image of $\hat{\Gamma}^{(\zeta)}$ under the mapping then satisfies a relation known in the continuum mechanics literature as *Nanson's formula*,

$$J^{(\kappa,\zeta)}(\boldsymbol{\xi})\boldsymbol{n}^{(\kappa,\zeta)}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi})) = \boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi})^{\mathrm{T}}\hat{\boldsymbol{n}}^{(\zeta)}, \quad \forall\, \boldsymbol{\xi} \in \hat{\Omega}, \tag{2.14}$$

for which derivations can be found in standard texts such as Gurtin *et al.* [109, Section 8.1].

2.3.2 Polynomial approximation spaces

Standard DG and FR methods are based on the approximation of each component of the transformed numerical solution in (2.13) within a polynomial space of finite dimension, for which the natural choice depends on the geometry of the reference element. For example, on the triangle or tetrahedron, one typically considers the *total-degree polynomial space*

$$\mathbb{P}_p(\hat{\Omega}) \coloneqq \operatorname{span}\Big\{\hat{\Omega} \ni \boldsymbol{\xi} \mapsto \xi_1^{\alpha_1}\cdots\xi_d^{\alpha_d} : \boldsymbol{\alpha} \in \mathcal{P}(p)\Big\}, \tag{2.15}$$

where we define the multi-index set $\mathcal{P}(p) \coloneqq \{\boldsymbol{\alpha} \in \mathbb{N}_0^d : |\boldsymbol{\alpha}| \leq p\}$ with $|\boldsymbol{\alpha}| \coloneqq \alpha_1 + \cdots + \alpha_d$. On the quadrilateral or hexahedron, the natural choice is the *tensor-product polynomial space*

$$\mathbb{Q}_p(\hat{\Omega}) \coloneqq \operatorname{span}\Big\{\hat{\Omega} \ni \boldsymbol{\xi} \mapsto \xi_1^{\alpha_1}\cdots\xi_d^{\alpha_d} : 0 \leq \alpha_l \leq p,\ \forall\, l \in \{1:d\}\Big\}. \tag{2.16}$$

For generality, we use $\mathbb{V}_p(\hat{\Omega})$ to denote a generic polynomial space on the reference element, and we define the corresponding *trace space* on a given facet $\hat{\Gamma}^{(\zeta)} \subset \partial\hat{\Omega}$ as

$$\mathbb{V}_p(\hat{\Gamma}^{(\zeta)}) \coloneqq \Big\{V|_{\hat{\Gamma}^{(\zeta)}} \colon V \in \mathbb{V}_p(\hat{\Omega})\Big\}, \tag{2.17}$$

where the specific cases of $\mathbb{P}_p(\hat{\Gamma}^{(\zeta)})$ and $\mathbb{Q}_p(\hat{\Gamma}^{(\zeta)})$ are defined in an analogous manner. The dimension of the approximation space is given by $N_p \coloneqq \dim(\mathbb{V}_p(\hat{\Omega}))$, where we have

$$\dim\Big(\mathbb{P}_p(\hat{\Omega})\Big) = \binom{p+d}{d}, \quad \dim\Big(\mathbb{Q}_p(\hat{\Omega})\Big) = \Big(p+1\Big)^d, \tag{2.18}$$

for the total-degree and tensor-product polynomial spaces in (2.15) and (2.16), respectively.

2.3.3 Discontinuous Galerkin formulation

A weak formulation of (2.1) can be derived on the reference element either by integrating by parts against a smooth test function on the physical element and performing a change of variables within each of the resulting integrals, or by integrating the transformed conservation law in (2.13) by parts against a smooth test function on the reference element and using (2.14) on the boundary. In either case, we obtain a weak formulation given by

$$\begin{aligned}\int_{\hat{\Omega}} V(\boldsymbol{\xi}) J^{(\kappa)}(\boldsymbol{\xi}) \frac{\partial \underline{U}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}), t)}{\partial t} \,\mathrm{d}\boldsymbol{\xi} &= \sum_{l=1}^{d} \int_{\hat{\Omega}} \frac{\partial \phi^{(i)}(\boldsymbol{\xi})}{\partial \xi_l} \sum_{m=1}^{d} G_{lm}^{(\kappa)}(\boldsymbol{\xi}) \underline{F}_m(\underline{U}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}), t)) \,\mathrm{d}\boldsymbol{\xi} \\ &\quad - \sum_{\zeta=1}^{N_f} \int_{\hat{\Gamma}^{(\zeta)}} V(\boldsymbol{\xi}) J^{(\kappa,\zeta)}(\boldsymbol{\xi}) \sum_{m=1}^{d} \underline{F}_m(\underline{U}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}), t)) n_m^{(\kappa,\zeta)}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi})) \,\mathrm{d}\hat{s}. \end{aligned} \tag{2.19}$$

To obtain a DG approximation of (2.19), we then approximate each solution variable in space in terms of a basis $\{\phi^{(i)}\}_{i\in\{1:N_p\}}$ for the generic polynomial space $\mathbb{V}_p(\hat{\Omega})$, resulting in an expansion of the form

$$U_e(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}), t) \approx U_e^{(h,\kappa)}(\boldsymbol{\xi}, t) := \sum_{i=1}^{N_p} \tilde{u}_i^{(h,\kappa,e)}(t) \phi^{(i)}(\boldsymbol{\xi}). \tag{2.20}$$

Considering test functions belonging to the same space as the solution variables and noting that it is sufficient to test with each basis function due to the linearity of each term in (2.19) with respect to the test function, we obtain a DG formulation given by

$$\begin{aligned}\int_{\hat{\Omega}} \phi^{(i)}(\boldsymbol{\xi}) J^{(\kappa)}(\boldsymbol{\xi}) \frac{\partial \underline{U}^{(h,\kappa)}(\boldsymbol{\xi}, t)}{\partial t} \,\mathrm{d}\boldsymbol{\xi} &= \sum_{l=1}^{d} \int_{\hat{\Omega}} \frac{\partial \phi^{(i)}(\boldsymbol{\xi})}{\partial \xi_l} \sum_{m=1}^{d} G_{lm}^{(\kappa)}(\boldsymbol{\xi}) \underline{F}_m(\underline{U}^{(h,\kappa)}(\boldsymbol{\xi}, t)) \,\mathrm{d}\boldsymbol{\xi} \\ &\quad - \sum_{\zeta=1}^{N_f} \int_{\hat{\Gamma}^{(\zeta)}} \phi^{(i)}(\boldsymbol{\xi}) J^{(\kappa,\zeta)}(\boldsymbol{\xi}) \underline{F}^{(*,\kappa,\zeta)}(\boldsymbol{\xi}, t) \,\mathrm{d}\hat{s}, \quad \forall\, i \in \{1 : N_p\}, \end{aligned} \tag{2.21}$$

where the numerical flux function $\underline{F}^* : \Upsilon \times \Upsilon \times \mathbb{S}^{d-1} \to \mathbb{R}^{N_c}$ has been used to resolve the discontinuity in the global numerical solution at each facet, on which we define

$$\underline{F}^{(*,\kappa,\zeta)}(\boldsymbol{\xi}, t) := \underline{F}^*(\underline{U}^{(h,\kappa)}(\boldsymbol{\xi}, t), \underline{U}^{(h,\kappa,+)}(\boldsymbol{\xi}, t), \boldsymbol{n}^{(\kappa,\zeta)}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}))), \tag{2.22}$$

with the exterior solution state given by $\underline{U}^{(h,\kappa,+)}(\boldsymbol{\xi}, t) \in \Upsilon$. The numerical flux is typically assumed to satisfy the *conservation* and *consistency* properties given, respectively, by

$$\underline{F}^*(\underline{U}^-, \underline{U}^+, \boldsymbol{n}) = -\underline{F}^*(\underline{U}^+, \underline{U}^-, -\boldsymbol{n}), \quad \forall\, \underline{U}^-, \underline{U}^+ \in \Upsilon, \quad \forall\, \boldsymbol{n} \in \mathbb{S}^{d-1}, \tag{2.23a}$$

$$\underline{F}^*(\underline{U}, \underline{U}, \boldsymbol{n}) = \underline{F}(\underline{U}, \boldsymbol{n}), \quad \forall\, \underline{U} \in \Upsilon, \quad \forall\, \boldsymbol{n} \in \mathbb{S}^{d-1}. \tag{2.23b}$$

The initial condition for a DG scheme is typically imposed through a Galerkin projection with respect to the L^2 inner product on the physical element; expressing such a projection in reference coordinates, we therefore obtain $\underline{U}^{(h,\kappa)}(\cdot, 0)$ by solving

$$\int_{\hat{\Omega}} \phi^{(i)}(\boldsymbol{\xi}) J^{(\kappa)}(\boldsymbol{\xi}) \underline{U}^{(h,\kappa)}(\boldsymbol{\xi}, 0) \, \mathrm{d}\boldsymbol{\xi} = \int_{\hat{\Omega}} \phi^{(i)} J^{(\kappa)}(\boldsymbol{\xi}) \underline{U}^{0}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi})) \, \mathrm{d}\boldsymbol{\xi}, \quad \forall\, i \in \{1 : N_p\}. \tag{2.24}$$

Inserting the expansion (2.20) into (2.21) and (2.24) then results in a system of N_p equations which must be solved in order to obtain the expansion coefficients for each component of the time derivative and initial condition, producing a system of ordinary differential equations (ODEs) to which a standard implicit or explicit time-marching method can be applied.

2.3.4 Flux reconstruction formulation

Unlike the DG method, the FR approach involves the discretization of the strong form of (2.13) rather than the weak form, and traditionally employs a nodal collocation-based approximation rather than a Galerkin approach. Expanding each component of the numerical solution using a Lagrange basis $\{\ell^{(i)}\}_{i \in \{1:N_p\}}$ for the polynomial space $\mathbb{V}_p(\hat{\Omega})$ as

$$U_e(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}), t) \approx U_e^{(h,\kappa)}(\boldsymbol{\xi}, t) \coloneqq \sum_{i=1}^{N_p} u_i^{(h,\kappa,e)}(t) \ell^{(i)}(\boldsymbol{\xi}), \tag{2.25}$$

where such a basis satisfies the cardinal property $\ell^{(i)}(\boldsymbol{\xi}^{(j)}) = \delta_{ij}$ on the nodes $\{\boldsymbol{\xi}^{(i)}\}_{i \in \{1:N_p\}} \subset \hat{\Omega}$ such that $u_i^{(h,\kappa,e)}(t) = U_e^{(h,\kappa)}(\boldsymbol{\xi}^{(i)}, t)$, we can define a *discontinuous flux* approximation as

$$\underline{\hat{F}}_D^{(\kappa,l)}(\boldsymbol{\xi}, t) \coloneqq \sum_{i=1}^{N_p} \sum_{m=1}^{d} G_{lm}^{(\kappa)}(\boldsymbol{\xi}^{(i)}) \underline{F}_m(\underline{U}^{(h,\kappa)}(\boldsymbol{\xi}^{(i)}, t)) \ell^{(i)}(\boldsymbol{\xi}). \tag{2.26}$$

The central idea of the FR method as proposed by Huynh [39] is to reconstruct a continuous flux through the addition of a *correction flux* to the discontinuous flux in (2.26), which is defined based on the difference between the scaled numerical flux and the normal component of the discontinuous flux at the nodes $\{\boldsymbol{\xi}^{(\zeta,i)}\}_{i \in \{1:N_{q_f}^{(\zeta)}\}} \subset \hat{\Gamma}^{(\zeta)}$ on each facet as

$$\underline{\hat{F}}_C^{(\kappa,l)}(\boldsymbol{\xi}, t) \coloneqq \sum_{\zeta=1}^{N_f} \sum_{i=1}^{N_{q_f}^{(\zeta)}} \left(J^{(\kappa,\zeta)}(\boldsymbol{\xi}^{(\zeta,i)}) \underline{F}^{(*,\kappa,\zeta)}(\boldsymbol{\xi}^{(\zeta,i)}, t) - \sum_{m=1}^{d} \hat{n}_m^{(\zeta)} \underline{\hat{F}}_D^{(\kappa,m)}(\boldsymbol{\xi}^{(\zeta,i)}, t) \right) G_l^{(\zeta,i)}(\boldsymbol{\xi}), \tag{2.27}$$

where the numerical flux is evaluated as in (2.22) and $\boldsymbol{G}^{(\zeta,i)} : \hat{\Omega} \to \mathbb{R}^d$ denote vector-valued correction functions satisfying

$$\nabla_{\boldsymbol{\xi}} \cdot \boldsymbol{G}^{(\zeta,i)} \in \mathbb{V}_p(\hat{\Omega}), \quad \text{and} \quad \boldsymbol{G}^{(\zeta,i)}(\boldsymbol{\xi}^{(\eta,j)}) \cdot \hat{\boldsymbol{n}}^{(\eta)} = \delta_{\zeta\eta}\delta_{ij}. \tag{2.28}$$

Such conditions are typically supplemented with additional constraints in order to ensure energy stability in a particular norm, a topic which we will consider in Chapter 3 within the context of the SBP property. Substituting the approximation in (2.25) into (2.13) and using the above correction procedure to approximate the transformed flux components, we obtain a semi-discrete formulation given for $i \in \{1 : N_q\}$ by

$$\frac{\mathrm{d}\underline{U}^{(h,\kappa)}(\boldsymbol{\xi}^{(i)}, t)}{\mathrm{d}t} = -\frac{1}{J^{(\kappa)}(\boldsymbol{\xi}^{(i)})} \sum_{l=1}^{d} \frac{\partial(\underline{\hat{F}}_D^{(\kappa,l)} + \underline{\hat{F}}_C^{(\kappa,l)})}{\partial \xi_l}(\boldsymbol{\xi}^{(i)}, t), \tag{2.29}$$

where the initial condition is obtained as $\underline{U}^{(h,\kappa)}(\boldsymbol{\xi}^{(i)}, 0) = \underline{U}^0(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}^{(i)}))$. The nodal values can then be advanced in time using a standard time-marching method for systems of ODEs.

2.4 Orthogonal polynomials and Gaussian quadrature rules

The fundamental building blocks used for constructing the novel DSEMs described in this work are Jacobi and Legendre polynomials as well as their associated Gaussian quadrature rules and interpolants, the basic properties of which we will review here. The normalized *Jacobi polynomials* are denoted by $P_i^{(a,b)} \in \mathbb{P}_i([-1, 1])$ and satisfy

$$\int_{-1}^{1} P_i^{(a,b)}(\xi) P_j^{(a,b)}(\xi)(1-\xi)^a(1+\xi)^b \,\mathrm{d}\xi = \delta_{ij}, \quad \forall\, a, b > -1, \tag{2.30}$$

where the case of $(a, b) = (0, 0)$ corresponds to the normalized *Legendre polynomials*. The Jacobi polynomials can be constructed through recurrence relations, as shown, for example, by Hesthaven and Warburton [11, Appendix A]. For a given non-negative integer q, the *Gaussian quadrature rules* corresponding to a Jacobi weight with exponents a and b have nodes $\{\xi^{(i)}\}_{i\in\{0:q\}} \subset [-1, 1]$ given by the $q + 1$ solutions to a polynomial equation. Such equations are given by

$$\text{Gauss:} \qquad P_{q+1}^{(a,b)}(\xi) = 0, \tag{2.31a}$$

$$\text{Gauss–Radau:} \qquad (1+\xi)P_q^{(a,b+1)}(\xi) = 0, \tag{2.31b}$$

$$\text{Gauss–Lobatto:} \qquad (1-\xi^2)P_{q-1}^{(a+1,b+1)}(\xi) = 0, \tag{2.31c}$$

where the Gauss, Gauss–Radau, and Gauss–Lobatto families of quadrature rules include zero, one, and two endpoints of the interval, respectively, and we note that Gauss–Lobatto rules require $q \geq 1$ due to the requirement for at least two nodes.

Remark 2.2. Here, we define Gauss–Radau quadrature rules including a node at the left endpoint, with those including the right endpoint obtained by flipping the sign of ξ in (2.31b).

The *Lagrange polynomials* $\{\ell^{(i)}\}_{i\in\{0:q\}}$ associated with the quadrature nodes $\{\xi^{(i)}\}_{i\in\{0:q\}}$ constitute a basis for $\mathbb{P}_q([-1,1])$ satisfying the *cardinal property* $\ell^{(i)}(\xi^{(j)}) = \delta_{ij}$ and are given explicitly as

$$\ell^{(i)}(\xi) \coloneqq \prod_{j\in\{0:q\}\setminus\{i\}} \frac{\xi - \xi^{(j)}}{\xi^{(i)} - \xi^{(j)}}. \tag{2.32}$$

The corresponding Gaussian quadrature weights $\{\omega^{(i)}\}_{i\in\{0:q\}}$ can then be expressed as

$$\omega^{(i)} \coloneqq \int_{-1}^{1} \ell^{(i)}(\xi)(1-\xi)^a(1+\xi)^b \,\mathrm{d}\xi, \tag{2.33}$$

where alternative expressions for such weights can be found, for example, in references such as Karniadakis and Sherwin [110, Appendix B]. The resulting quadrature rule satisfies

$$\sum_{i=0}^{q} V(\xi^{(i)})\,\omega^{(i)} = \int_{-1}^{1} V(\xi)(1-\xi)^a(1+\xi)^b \,\mathrm{d}\xi, \quad \forall\, V \in \mathbb{P}_{\tau^{(a,b)}}([-1,1]), \tag{2.34}$$

where $\tau^{(a,b)} = 2q+1$ for Gauss nodes, $\tau^{(a,b)} = 2q$ for Gauss–Radau nodes, and $\tau^{(a,b)} = 2q-1$ for Lobatto nodes. In the case of $(a,b) = (0,0)$, we recover the familiar *Legendre–Gauss* (LG), *Legendre–Gauss–Radau* (LGR), *Legendre–Gauss–Lobatto* (LGL) quadrature rules for integration with respect to the unit weight. Gauss, Gauss–Radau, and Gauss–Lobatto quadrature rules for which $(a,b) \neq (0,0)$ will be referred to in this thesis as those of *Jacobi–Gauss* (JG), *Jacobi–Gauss–Radau* (JGR), and *Jacobi–Gauss–Lobatto* (JGR) type, respectively.

2.5 Summation-by-parts operators

In this section, we will review the generalized definitions presented in [78] and [111] for one-dimensional and multidimensional SBP operators, respectively. We will use the term *SBP operator* to denote operators adhering to either of such definitions, which are sometimes denoted as *generalized SBP operators* to distinguish them from the classical finite-difference SBP operators introduced in [63]. Due to the focus of this thesis on DSEMs rather than finite-difference methods, this is not expected to cause confusion.

2.5.1 One-dimensional summation-by-parts operators

The following generalized definition of a nodal SBP operator in one dimension was proposed in [78], extending earlier notions of SBP operators used in the finite-difference community to encompass element-type operators used in DSEMs, including those based on quadrature

rules not including one or both endpoints of the interval.

Definition 2.2. Let $\{x^{(i)}\}_{i\in\{1:N_q\}} \subset [x_L, x_R]$ denote a set of N_q distinct nodes, on which the functions $U, V : [x_L, x_R] \to \mathbb{R}$ can be evaluated as

$$\underline{u} \coloneqq \left[U(x^{(1)}), \ldots, U(x^{(N_q)})\right]^{\mathrm{T}} \quad \text{and} \quad \underline{v} \coloneqq \left[V(x^{(1)}), \ldots, V(x^{(N_q)})\right]^{\mathrm{T}}. \tag{2.35}$$

A matrix $\underline{\underline{D}} \in \mathbb{R}^{N_q \times N_q}$ approximating $\mathrm{d}/\mathrm{d}x$ is an *SBP operator* of degree q if it satisfies

$$\underline{\underline{D}}\,\underline{v} = \left[(\mathrm{d}V/\mathrm{d}x)(x^{(1)}), \ldots, (\mathrm{d}V/\mathrm{d}x)(x^{(N_q)})\right]^{\mathrm{T}}, \quad \forall\, V \in \mathbb{P}_q([x_L, x_R]), \tag{2.36}$$

and may be decomposed as $\underline{\underline{D}} = \underline{\underline{W}}^{-1}\underline{\underline{Q}}$ such that $\underline{\underline{W}} \in \mathbb{R}^{N_q \times N_q}$ is SPD and $\underline{\underline{Q}} \in \mathbb{R}^{N_q \times N_q}$ satisfies the *SBP property* $\underline{\underline{Q}} + \underline{\underline{Q}}^{\mathrm{T}} = \underline{\underline{E}}$, where $\underline{\underline{E}} \in \mathbb{R}^{N_q \times N_q}$ satisfies

$$\underline{u}^{\mathrm{T}}\underline{\underline{E}}\,\underline{v} = U(x_R)V(x_R) - U(x_L)V(x_L), \quad \forall\, U, V \in \mathbb{P}_q([x_L, x_R]). \tag{2.37}$$

Remark 2.3. Polynomial exactness conditions such as (2.36) imply that any SBP operator of degree q is, by definition, also an SBP operator of any non-negative integer degree $p < q$. Where such ambiguity arises, the degree q of an SBP operator is taken to refer uniquely to the *maximum* integer value for which such a condition holds.

The SBP property mimics the integration-by-parts property of the first derivative $\mathrm{d}/\mathrm{d}x$, which is given for continuously differentiable functions U and V as

$$\begin{array}{ccccc} \displaystyle\int_{x_L}^{x_R} U(x)\frac{\mathrm{d}V(x)}{\mathrm{d}x}\,\mathrm{d}x & + & \displaystyle\int_{x_L}^{x_R} \frac{\mathrm{d}U(x)}{\mathrm{d}x}V(x)\,\mathrm{d}x & = & U(x_R)V(x_R) - U(x_L)V(x_L). \\ \wr\!\wr & & \wr\!\wr & & \wr\!\wr \\ \underline{u}^{\mathrm{T}}\underline{\underline{Q}}\,\underline{v} & + & \underline{u}^{\mathrm{T}}\underline{\underline{Q}}^{\mathrm{T}}\underline{v} & = & \underline{u}^{\mathrm{T}}\underline{\underline{E}}\,\underline{v} \end{array} \tag{2.38}$$

We refer to any SBP operator for which the associated matrix $\underline{\underline{W}}$ is diagonal as a *diagonal-norm* SBP operator, where $\underline{\underline{W}}$ is denoted as the *norm matrix* in the finite-difference literature and as the *mass matrix* in the context of spectral and finite-element methods. It was shown in [78, Section 4.1] that for any diagonal-norm SBP operator of degree q, the corresponding quadrature rule with positive weights $\{\omega^{(i)}\}_{i\in\{1:N_q\}} \subset \mathbb{R}^+$ given by $\omega^{(i)} \coloneqq W_{ii}$ satisfies

$$\sum_{i=1}^{N_q} V(x^{(i)})\,\omega^{(i)} = \int_{x_L}^{x_R} V(x)\,\mathrm{d}x, \quad \forall\, V \in \mathbb{P}_{2q-1}([x_L, x_R]). \tag{2.39}$$

Remark 2.4. We devote particular attention to diagonal-norm SBP operators in this thesis, since the matrix $\underline{\underline{W}}$ must be diagonal for the construction of provably stable schemes on curvilinear meshes and for nonlinear PDEs, which are the focus of Chapters 5 and 6.

2.5.2 Multidimensional summation-by-parts operators

We now present the following definition of a nodal SBP operator from [83], which extends Definition 2.2 to the multidimensional setting.

Definition 2.3. Let $\hat{\Omega} \subset \mathbb{R}^d$ denote a closed, bounded, and connected reference domain on which we define a set of N_q distinct nodes $\{\boldsymbol{\xi}^{(i)}\}_{i\in\{1:N_q\}} \subset \hat{\Omega}$, and let

$$\underline{u} \coloneqq \left[U(\boldsymbol{\xi}^{(1)}), \dots, U(\boldsymbol{\xi}^{(N_q)})\right]^{\mathrm{T}} \quad \text{and} \quad \underline{v} \coloneqq \left[V(\boldsymbol{\xi}^{(1)}), \dots, V(\boldsymbol{\xi}^{(N_q)})\right]^{\mathrm{T}} \tag{2.40}$$

contain the nodal values of functions $U, V : \hat{\Omega} \to \mathbb{R}$. A matrix $\underline{\underline{D}}^{(l)} \in \mathbb{R}^{N_q\times N_q}$ approximating $\partial/\partial\xi_l$ is then a *multidimensional SBP operator* of degree q if it satisfies

$$\underline{\underline{D}}^{(l)}\underline{v} = \left[(\partial V/\partial\xi_l)(\boldsymbol{\xi}^{(1)}), \dots, (\partial V/\partial\xi_l)(\boldsymbol{\xi}^{(N_q)})\right]^{\mathrm{T}}, \quad \forall\, V \in \mathbb{P}_q(\hat{\Omega}), \tag{2.41}$$

and may be decomposed as $\underline{\underline{D}}^{(l)} = \underline{\underline{W}}^{-1}\underline{\underline{Q}}^{(l)}$ such that $\underline{\underline{W}} \in \mathbb{R}^{N_q\times N_q}$ is SPD and $\underline{\underline{Q}}^{(l)} \in \mathbb{R}^{N_q\times N_q}$ satisfies the *SBP property* $\underline{\underline{Q}}^{(l)} + (\underline{\underline{Q}}^{(l)})^{\mathrm{T}} = \underline{\underline{E}}^{(l)}$, where $\underline{\underline{E}}^{(l)} \in \mathbb{R}^{N_q\times N_q}$ satisfies

$$\underline{u}^{\mathrm{T}}\underline{\underline{E}}^{(l)}\underline{v} = \int_{\partial\hat{\Omega}} U(\boldsymbol{\xi})V(\boldsymbol{\xi})\hat{n}_l(\boldsymbol{\xi})\,\mathrm{d}\hat{s}, \quad \forall\, U, V \in \mathbb{P}_q(\hat{\Omega}), \tag{2.42}$$

where $\hat{\boldsymbol{n}} : \partial\hat{\Omega} \to \mathbb{S}^{d-1}$ denotes the outward unit normal vector to $\hat{\Omega}$.

As in the one-dimensional case, the multidimensional SBP property mimics the integration-by-parts relation given for continuously differentiable functions U and V as

$$\begin{array}{ccccc} \displaystyle\int_{\hat{\Omega}} U(\boldsymbol{\xi})\frac{\partial V(\boldsymbol{\xi})}{\partial\xi_l}\,\mathrm{d}\boldsymbol{\xi} & + & \displaystyle\int_{\hat{\Omega}} \frac{\partial U(\boldsymbol{\xi})}{\partial\xi_l}V(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} & = & \displaystyle\int_{\partial\hat{\Omega}} U(\boldsymbol{\xi})V(\boldsymbol{\xi})\hat{n}_l(\boldsymbol{\xi})\,\mathrm{d}\hat{s}, \\ \wr\wr & & \wr\wr & & \wr\wr \\ \underline{u}^{\mathrm{T}}\underline{\underline{Q}}^{(l)}\underline{v} & + & \underline{u}^{\mathrm{T}}\left(\underline{\underline{Q}}^{(l)}\right)^{\mathrm{T}}\underline{v} & = & \underline{u}^{\mathrm{T}}\underline{\underline{E}}^{(l)}\underline{v} \end{array} \tag{2.43}$$

and is equivalent to the decomposition $\underline{\underline{Q}}^{(l)} = \underline{\underline{S}}^{(l)} + \frac{1}{2}\underline{\underline{E}}^{(l)}$, where $\underline{\underline{S}}^{(l)}$ is skew-symmetric and $\underline{\underline{E}}^{(l)}$ is a symmetric matrix satisfying (2.42). In the case of a diagonal-norm multidimensional SBP operator of degree q, the corresponding quadrature rule satisfies

$$\sum_{i=1}^{N_q} V(\boldsymbol{\xi}^{(i)})\,\omega^{(i)} = \int_{\hat{\Omega}} V(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}, \quad \forall\, V \in \mathbb{P}_\tau(\hat{\Omega}), \tag{2.44}$$

where, similarly to (2.39), it was shown in [83, Theorem 3.2] that $\tau \geq 2q-1$.

2.5.3 Decomposition of the boundary operators

Assuming as in Section 2.3.1 that the reference element is a polytope, we introduce $N_{q_f}^{(\zeta)}$ distinct nodes $\{\boldsymbol{\xi}^{(\zeta,i)}\}_{i\in\{1:N_{q_f}^{(\zeta)}\}} \subset \hat{\Gamma}^{(\zeta)}$ on each facet $\hat{\Gamma}^{(\zeta)} \subset \partial\hat{\Omega}$. As in [84, Section 3], we then restrict our attention to the class of SBP operators for which the boundary matrices in (2.42) are constructed as

$$\underline{\underline{E}}^{(l)} \coloneqq \sum_{\zeta=1}^{N_f} \hat{n}_l^{(\zeta)} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{R}}^{(\zeta)}, \tag{2.45}$$

in terms of the SPSD facet mass matrices $\underline{\underline{B}}^{(\zeta)} \in \mathbb{R}^{N_q \times N_q}$ as well as interpolation/extrapolation operators $\underline{\underline{R}}^{(\zeta)} \in \mathbb{R}^{N_{q_f}^{(\zeta)} \times N_q}$ satisfying

$$\underline{\underline{R}}^{(\zeta)} \underline{v} = \left[V(\boldsymbol{\xi}^{(\zeta,1)}), \ldots, V(\boldsymbol{\xi}^{(\zeta,N_{q_f}^{(\zeta)})})\right]^{\mathrm{T}}, \quad \forall\, V \in \mathbb{P}_q(\hat{\Omega}). \tag{2.46}$$

As a special case of (2.45), we note that $\underline{\underline{E}}^{(l)}$ can be made diagonal by constructing SBP operators for which the facet quadrature nodes all form subsets of the volume quadrature nodes, wherein $\underline{\underline{R}}^{(\zeta)}$ simply selects boundary values from nodal vectors. Such SBP operators, denoted here as *diagonal-E* operators, are analogous to those on one-dimensional nodal sets including both endpoints, and are constructed on simplicial elements, for example, by Chen and Shu [98], Del Rey Fernández *et al.* [112], and Worku and Zingg [113].

Chapter 3

Unifying framework for discontinuous spectral-element methods based on the summation-by-parts property

In this chapter, which is based on the content of Paper I, we introduce the fundamental components of the proposed discontinuous spectral-element framework based on the SBP property within the context of standard DG and FR methods. In doing so, we aim to provide new insights regarding the existing schemes through their interpretation as SBP discretizations, as well as to introduce the principles and notation which will be used later in this thesis to construct provably stable schemes for nonlinear problems and curved meshes.

3.1 Reference operator matrices

In order to obtain an algebraic DSEM formulation amenable to analysis based on the SBP property, we must first define the matrix operators which constitute the discretization on the reference element. Letting $\{\boldsymbol{\xi}^{(i)}\}_{i\in\{1:N_q\}} \subset \hat{\Omega}$ and $\{\omega^{(i)}\}_{i\in\{1:N_q\}} \subset \mathbb{R}^+$ denote the nodes and weights, respectively, for a volume quadrature rule on the reference element of the form

$$\sum_{i=1}^{N_q} V(\boldsymbol{\xi}^{(i)})\,\omega^{(i)} \approx \int_{\hat{\Omega}} V(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}, \tag{3.1}$$

and letting $\{\boldsymbol{\xi}^{(\zeta,i)}\}_{i\in\{1:N_{q_f}^{(\zeta)}\}} \subset \hat{\Gamma}^{(\zeta)}$ and $\{\omega^{(\zeta,i)}\}_{i\in\{1:N_{q_f}^{(\zeta)}\}} \subset \mathbb{R}_0^+$ define facet quadrature rules as

$$\sum_{i=1}^{N_{q_f}^{(\zeta)}} V(\boldsymbol{\xi}^{(\zeta,i)})\,\omega^{(\zeta,i)} \approx \int_{\hat{\Gamma}^{(\zeta)}} V(\boldsymbol{\xi})\,\mathrm{d}\hat{s}, \tag{3.2}$$

we construct the diagonal matrices

$$\underline{\underline{W}} \coloneqq \begin{bmatrix} \omega^{(1)} & & \\ & \ddots & \\ & & \omega^{(N_q)} \end{bmatrix}, \quad \underline{\underline{B}}^{(\zeta)} \coloneqq \begin{bmatrix} \omega^{(\zeta,1)} & & \\ & \ddots & \\ & & \omega^{(\zeta,N_{q_f}^{(\zeta)})} \end{bmatrix}. \tag{3.3}$$

Defining the *generalized Vandermonde matrix* corresponding to an arbitrary (i.e. nodal or modal) basis $\{\phi^{(i)}\}_{i\in\{1:N_p\}}$ for the generic polynomial space $\mathbb{V}_p(\hat{\Omega})$ as

$$\underline{\underline{V}} \coloneqq \begin{bmatrix} \phi^{(1)}(\boldsymbol{\xi}^{(1)}) & \cdots & \phi^{(N_p)}(\boldsymbol{\xi}^{(1)}) \\ \vdots & \ddots & \vdots \\ \phi^{(1)}(\boldsymbol{\xi}^{(N_q)}) & \cdots & \phi^{(N_p)}(\boldsymbol{\xi}^{(N_q)}) \end{bmatrix}, \tag{3.4}$$

we require that differentiation and interpolation/extrapolation operators exist on the chosen nodal sets which satisfy the following assumption.

Assumption 3.1. *The basis $\{\phi^{(i)}\}_{i\in\{1:N_p\}}$ spans a polynomial space $\mathbb{V}_p(\hat{\Omega})$ which is closed under partial differentiation, and the rank of the corresponding Vandermonde matrix $\underline{\underline{V}}$ in (3.4) is equal to the dimension N_p of such a space (i.e. it is of full column rank). Moreover, there exist matrices $\underline{\underline{D}}^{(l)} \in \mathbb{R}^{N_q\times N_q}$ and $\underline{\underline{R}}^{(\zeta)} \in \mathbb{R}^{N_{q_f}^{(\zeta)}\times N_q}$ satisfying*

$$\underline{\underline{D}}^{(l)}\underline{\underline{V}} = \begin{bmatrix} (\partial\phi^{(1)}/\partial\xi_l)(\boldsymbol{\xi}^{(1)}) & \cdots & (\partial\phi^{(N_p)}/\partial\xi_l)(\boldsymbol{\xi}^{(1)}) \\ \vdots & \ddots & \vdots \\ (\partial\phi^{(1)}/\partial\xi_l)(\boldsymbol{\xi}^{(N_q)}) & \cdots & (\partial\phi^{(N_{q_f}^{(\zeta)})}/\partial\xi_l)(\boldsymbol{\xi}^{(N_q)}) \end{bmatrix}, \quad \forall\, l \in \{1:d\}, \tag{3.5a}$$

$$\underline{\underline{R}}^{(\zeta)}\underline{\underline{V}} = \begin{bmatrix} \phi^{(1)}(\boldsymbol{\xi}^{(\zeta,1)}) & \cdots & \phi^{(N_p)}(\boldsymbol{\xi}^{(\zeta,1)}) \\ \vdots & \ddots & \vdots \\ \phi^{(1)}(\boldsymbol{\xi}^{(\zeta,N_{q_f}^{(\zeta)})}) & \cdots & \phi^{(N_p)}(\boldsymbol{\xi}^{(\zeta,N_{q_f}^{(\zeta)})}) \end{bmatrix}, \quad \forall\, \zeta \in \{1:N_f\}, \tag{3.5b}$$

as well as the SBP property given for boundary operators in the form of (2.45) by

$$\underline{\underline{W}}\,\underline{\underline{D}}^{(l)} + \left(\underline{\underline{D}}^{(l)}\right)^{\mathrm{T}}\underline{\underline{W}} = \sum_{\zeta=1}^{N_f} \hat{n}_l^{(\zeta)} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)}\underline{\underline{R}}^{(\zeta)}, \quad \forall\, l \in \{1:d\}, \tag{3.6}$$

where the matrices $\underline{\underline{W}} \in \mathbb{R}^{N_q\times N_q}$ and $\underline{\underline{B}}^{(\zeta)} \in \mathbb{R}^{N_{q_f}^{(\zeta)}\times N_{q_f}^{(\zeta)}}$ are SPD and SPSD, respectively.

Remark 3.1. Polynomial spaces which are closed under partial differentiation are called *downward closed* (see, for example, Cohen and Migliorati [114]) and include both $\mathbb{P}_p(\hat{\Omega})$ and $\mathbb{Q}_p(\hat{\Omega})$. Since one can differentiate any monomial of the form $\pi^{\boldsymbol{\alpha}}(\boldsymbol{\xi}) \coloneqq \xi_1^{\alpha_1}\cdots\xi_d^{\alpha_d}$ belonging to a downward-closed polynomial space α_l times with respect to each coordinate ξ_l while

remaining within such a space, it follows that such spaces must always include constants.

3.1.1 Collocation-based operators

As discussed, for example, by Bos [115], the distinct nodes $\{\boldsymbol{\xi}^{(i)}\}_{i\in\{1:N_q\}}$ constitute a *unisolvent nodal set* for the space $\mathbb{V}_p(\hat{\Omega})$ when $\underline{\underline{V}}$ is invertible, where $N_q = N_p$ provides a necessary condition, but not a sufficient one when $d \geq 2$. When the nodes are unisolvent, the matrices in (3.5a) and (3.5b) are given uniquely by multiplying from the right by $\underline{\underline{V}}^{-1}$ to obtain

$$\underline{\underline{D}}^{(l)} = \begin{bmatrix} (\partial\ell^{(1)}/\partial\xi_l)(\boldsymbol{\xi}^{(1)}) & \cdots & (\partial\ell^{(N_q)}/\partial\xi_l)(\boldsymbol{\xi}^{(1)}) \\ \vdots & \ddots & \vdots \\ (\partial\ell^{(1)}/\partial\xi_l)(\boldsymbol{\xi}^{(N_q)}) & \cdots & (\partial\ell^{(N_q)}/\partial\xi_l)(\boldsymbol{\xi}^{(N_q)}) \end{bmatrix}, \tag{3.7a}$$

$$\underline{\underline{R}}^{(\zeta)} = \begin{bmatrix} \ell^{(1)}(\boldsymbol{\xi}^{(\zeta,1)}) & \cdots & \ell^{(N_q)}(\boldsymbol{\xi}^{(\zeta,1)}) \\ \vdots & \ddots & \vdots \\ \ell^{(1)}(\boldsymbol{\xi}^{(\zeta,N_{q_f}^{(\zeta)})}) & \cdots & \ell^{(N_q)}(\boldsymbol{\xi}^{(\zeta,N_{q_f}^{(\zeta)})}) \end{bmatrix}, \tag{3.7b}$$

where $\{\ell^{(i)}\}_{i\in\{1:N_q\}}$ is a nodal (i.e. Lagrange) basis given by

$$\begin{bmatrix} \ell^{(1)}(\boldsymbol{\xi}) \\ \vdots \\ \ell^{(N_q)}(\boldsymbol{\xi}) \end{bmatrix} \coloneqq \underline{\underline{V}}^{-\mathrm{T}} \begin{bmatrix} \phi^{(1)}(\boldsymbol{\xi}) \\ \vdots \\ \phi^{(N_p)}(\boldsymbol{\xi}) \end{bmatrix}, \tag{3.8}$$

which satisfies the cardinal property $\ell^{(i)}(\boldsymbol{\xi}^{(j)}) = \delta_{ij}$ and recovers (2.32) in the one-dimensional case. The differentiation matrix in (3.7a) is then recognized as a *collocation derivative* or *interpolation derivative* operator familiar in the context of spectral methods (see, for example, [108, Section 3.5.2] or [110, Section 2.4.2]). Defining the *exact* nodal mass matrix as

$$\underline{\underline{W}} \coloneqq \begin{bmatrix} \int_{\hat{\Omega}} \ell^{(1)}(\boldsymbol{\xi})\ell^{(1)}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} & \cdots & \int_{\hat{\Omega}} \ell^{(1)}(\boldsymbol{\xi})\ell^{(N_q)}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} \\ \vdots & \ddots & \vdots \\ \int_{\hat{\Omega}} \ell^{(N_q)}(\boldsymbol{\xi})\ell^{(1)}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} & \cdots & \int_{\hat{\Omega}} \ell^{(N_q)}(\boldsymbol{\xi})\ell^{(N_q)}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} \end{bmatrix}, \tag{3.9}$$

we then obtain the SBP property by applying integration by parts to the entries of the matrix

$$\underline{\underline{W}}\,\underline{\underline{D}}^{(l)} = \begin{bmatrix} \int_{\hat{\Omega}} \ell^{(1)}(\boldsymbol{\xi})(\partial\ell^{(1)}/\partial\xi_l)(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} & \cdots & \int_{\hat{\Omega}} \ell^{(1)}(\boldsymbol{\xi})(\partial\ell^{(N_q)}/\partial\xi_l)(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} \\ \vdots & \ddots & \vdots \\ \int_{\hat{\Omega}} \ell^{(N_q)}(\boldsymbol{\xi})(\partial\ell^{(1)}/\partial\xi_l)(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} & \cdots & \int_{\hat{\Omega}} \ell^{(N_q)}(\boldsymbol{\xi})(\partial\ell^{(N_q)}/\partial\xi_l)(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} \end{bmatrix}, \tag{3.10}$$

which, as shown in the one-dimensional case by Carpenter and Gottlieb [67], results in

$$\underline{\underline{W}}\,\underline{\underline{D}}^{(l)} + \left(\underline{\underline{D}}^{(l)}\right)^{\mathrm{T}} \underline{\underline{W}} = \begin{bmatrix} \int_{\partial\hat{\Omega}} \ell^{(1)}(\boldsymbol{\xi})\ell^{(1)}(\boldsymbol{\xi})\hat{n}_l(\boldsymbol{\xi})\,\mathrm{d}\hat{s} & \cdots & \int_{\partial\hat{\Omega}} \ell^{(1)}(\boldsymbol{\xi})\ell^{(N_q)}(\boldsymbol{\xi})\hat{n}_l(\boldsymbol{\xi})\,\mathrm{d}\hat{s} \\ \vdots & \ddots & \vdots \\ \int_{\partial\hat{\Omega}} \ell^{(N_q)}(\boldsymbol{\xi})\ell^{(1)}(\boldsymbol{\xi})\hat{n}_l(\boldsymbol{\xi})\,\mathrm{d}\hat{s} & \cdots & \int_{\partial\hat{\Omega}} \ell^{(N_q)}(\boldsymbol{\xi})\ell^{(N_q)}(\boldsymbol{\xi})\hat{n}_l(\boldsymbol{\xi})\,\mathrm{d}\hat{s} \end{bmatrix}. \tag{3.11}$$

We therefore obtain boundary operators $\underline{\underline{E}}^{(l)}$ in the form of (2.45) and, hence, an SBP property in the form of (3.6) when

$$\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{R}}^{(\zeta)} = \begin{bmatrix} \int_{\hat{\Gamma}^{(\zeta)}} \ell^{(1)}(\boldsymbol{\xi})\ell^{(1)}(\boldsymbol{\xi})\,\mathrm{d}\hat{s} & \cdots & \int_{\hat{\Gamma}^{(\zeta)}} \ell^{(1)}(\boldsymbol{\xi})\ell^{(N_q)}(\boldsymbol{\xi})\,\mathrm{d}\hat{s} \\ \vdots & \ddots & \vdots \\ \int_{\hat{\Gamma}^{(\zeta)}} \ell^{(N_q)}(\boldsymbol{\xi})\ell^{(1)}(\boldsymbol{\xi})\,\mathrm{d}\hat{s} & \cdots & \int_{\hat{\Gamma}^{(\zeta)}} \ell^{(N_q)}(\boldsymbol{\xi})\ell^{(N_q)}(\boldsymbol{\xi})\,\mathrm{d}\hat{s} \end{bmatrix} \tag{3.12}$$

holds for each facet of the reference element. Such is the case when $\underline{\underline{B}}^{(\zeta)}$ is defined as in (3.3) in terms of a facet quadrature rule which exactly integrates the product of two basis functions, or when $\underline{\underline{B}}^{(\zeta)}$ is defined analogously to (3.9) as the exact facet mass matrix, which is generally dense and SPD. The above construction therefore establishes the existence of a dense-norm multidimensional SBP operator on any unisolvent nodal set.

Remark 3.2. We refer the reader to Marchildon and Zingg [116] for the derivation of necessary conditions for obtaining symmetrical nodal sets which are unisolvent for total-degree polynomial spaces on triangles and tetrahedra, as well as to Chen and Babuška [117, 118], Hesthaven [119], Hesthaven and Teng [120], Taylor *et al.* [121], Warburton [122], and Chan and Warburton [123], who describe algorithms for constructing such nodal sets and bases.

3.1.2 Quadrature-based operators

While Section 3.1.1 provides a straightforward procedure for constructing SBP operators on unisolvent nodal sets, a diagonal matrix $\underline{\underline{W}}$ and, thus, a diagonal-norm SBP operator, is not obtained unless the corresponding quadrature rule with weights given similarly to (2.33) by

$$\omega^{(i)} \coloneqq \int_{\hat{\Omega}} \ell^{(i)}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} \tag{3.13}$$

is exact for the product of any two of the basis functions in (3.8), in which case the cardinal property results in the definitions given for $\underline{\underline{W}}$ in (3.3) and (3.9) being equivalent. For the line segment, quadrilateral, and hexahedron, this can be achieved, for example, by constructing spectral collocation operators on (tensor-product) LG quadrature rules. Aside from such special cases, however, obtaining a diagonal nodal mass matrix requires *mass lumping*, in which the collocation-based operators in (3.7a) and (3.7b) are used with the diagonal mass

matrix in (3.3) containing the interpolatory quadrature weights in (3.13). In a landmark paper, Gassner [69] recognized that if the integration-by-parts relation in (2.43) holds under quadrature for all $U, V \in \mathbb{V}_p(\hat{\Omega})$, which was shown to be the case for tensor-product LGL quadrature rules on quadrilaterals and hexahedra [68], the mass-lumped formulation satisfies a diagonal-norm SBP property. For other element types such as triangles and tetrahedra, however, one often requires more than N_p volume quadrature nodes for integration by parts to hold for polynomials in $\mathbb{V}_p(\hat{\Omega})$, and hence constructing a diagonal-norm SBP operator requires consideration of the general notion of a multidimensional SBP operator in Definition 2.3, which allows for the case of $N_q > N_p$. Unlike the matrices in (3.7a) and (3.7b), such operators are not uniquely defined on a given set of nodes, and there are many approaches to constructing such operators [83, 84]. To provide a concrete example of such an approach, we consider the methodology proposed by Chan [99], which employs the quadrature-based projection matrix

$$\underline{\underline{P}} \coloneqq \underline{\underline{M}}^{-1} \underline{\underline{V}}^{\mathrm{T}} \underline{\underline{W}}, \tag{3.14}$$

where $\underline{\underline{W}}$ is defined as in (3.3) using positive quadrature weights and $\underline{\underline{V}}$ is of full column rank, which ensures positive-definiteness of the modal mass matrix $\underline{\underline{M}} \coloneqq \underline{\underline{V}}^{\mathrm{T}} \underline{\underline{W}} \underline{\underline{V}}$. The operators

$$\underline{\underline{D}}^{(l)} \coloneqq \begin{bmatrix} (\partial \phi^{(1)}/\partial \xi_l)(\boldsymbol{\xi}^{(1)}) & \cdots & (\partial \phi^{(N_p)}/\partial \xi_l)(\boldsymbol{\xi}^{(1)}) \\ \vdots & \ddots & \vdots \\ (\partial \phi^{(1)}/\partial \xi_l)(\boldsymbol{\xi}^{(N_q)}) & \cdots & (\partial \phi^{(N_p^{(\zeta)}_{q_f})}/\partial \xi_l)(\boldsymbol{\xi}^{(N_q)}) \end{bmatrix} \underline{\underline{P}}, \tag{3.15a}$$

$$\underline{\underline{R}}^{(\zeta)} \coloneqq \begin{bmatrix} \phi^{(1)}(\boldsymbol{\xi}^{(\zeta,1)}) & \cdots & \phi^{(N_p)}(\boldsymbol{\xi}^{(\zeta,1)}) \\ \vdots & \ddots & \vdots \\ \phi^{(1)}(\boldsymbol{\xi}^{(\zeta,N_{q_f}^{(\zeta)})}) & \cdots & \phi^{(N_p)}(\boldsymbol{\xi}^{(\zeta,N_{q_f}^{(\zeta)})}) \end{bmatrix} \underline{\underline{P}} \tag{3.15b}$$

then satisfy (3.5a) and (3.5b), respectively, since $\underline{\underline{P}}\underline{\underline{V}} = \underline{\underline{I}}^{(N_p)}$, and can be shown to satisfy the SBP property in (3.6) when the quadrature is exact for each term in (2.43) when $U, V \in \mathbb{V}_p(\hat{\Omega})$. We therefore have a procedure for constructing multidimensional diagonal-norm SBP operators for general (e.g. non-collocated) choices of quadrature and basis. Such a topic will be revisited in Chapter 4 with the aim of obtaining more efficient operators on triangles and tetrahedra.

3.2 Discontinuous spectral-element formulations

The operators constructed in Section 3.1 will now be used to obtain algebraic formulations for the DG and FR methods described in Section 2.3. Specifically, we aim to obtain the time derivative of the vector $\underline{u}^{(h,\kappa,e)}(t)$ or $\underline{\tilde{u}}^{(h,\kappa,e)}(t)$, which contains the degrees of freedom for the local expansion in (2.20) or (2.25), in terms of the solution degrees of freedom for given

element and (through the numerical flux) those with which such an element shares a common interface. Such formulations also require the diagonal matrices $\underline{\underline{J}}^{(\kappa)}, \underline{\underline{G}}^{(\kappa,l,m)} \in \mathbb{R}^{N_q \times N_q}$ and $\underline{\underline{J}}^{(\kappa,\zeta)}, \underline{\underline{N}}^{(\kappa,\zeta,m)} \in \mathbb{R}^{N_{q_f}^{(\zeta)} \times N_{q_f}^{(\zeta)}}$ with entries given by

$$\begin{aligned} J_{ij}^{(\kappa)} &:= J^{(\kappa)}(\boldsymbol{\xi}^{(i)})\delta_{ij}, & G_{ij}^{(\kappa,l,m)} &:= G_{lm}^{(\kappa)}(\boldsymbol{\xi}^{(i)})\delta_{ij}, \\ J_{ij}^{(\kappa,\zeta)} &:= J^{(\kappa,\zeta)}(\boldsymbol{\xi}^{(\zeta,i)})\delta_{ij}, & N_{ij}^{(\kappa,\zeta,m)} &:= n_m^{(\kappa,\zeta)}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}^{(\zeta,i)}))\delta_{ij}, \end{aligned} \tag{3.16}$$

which describe the geometry of each physical element, and we use the relation

$$\underline{u}^{(h,\kappa,e)}(t) = \underline{\underline{V}}\,\underline{\tilde{u}}^{(h,\kappa,e)}(t), \tag{3.17}$$

to obtain the nodal values from the modal expansion coefficients. We note that although the matrix $\underline{\underline{P}}$ defined in (3.14) can be used to obtain a modal expansion from a vector of nodal values, such an expansion is not guaranteed to be unique unless $\underline{\underline{V}}$ is invertible, which results in $\underline{\underline{P}} = \underline{\underline{V}}^{-1}$. In such a case, the operators in (3.7) are then equal to those in (3.15).

3.2.1 Discontinuous Galerkin method

Using matrices $\underline{\underline{D}}^{(l)}$ and $\underline{\underline{R}}^{(\zeta)}$ satisfying (3.5a) and (3.5b), respectively, to exactly differentiate and interpolate/extrapolate each basis function, an algebraic formulation of the DG method is obtained by approximating the integrals in (2.21) using $\underline{\underline{W}}$ and $\underline{\underline{B}}^{(\zeta)}$, resulting in

$$\begin{aligned} \underline{\underline{V}}^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{\underline{V}}\frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa,e)}(t)}{\mathrm{d}t} &= \sum_{l=1}^{d}\left(\underline{\underline{D}}^{(l)}\underline{\underline{V}}\right)^{\mathrm{T}}\underline{\underline{W}}\sum_{m=1}^{d}\underline{\underline{G}}^{(\kappa,l,m)}\underline{f}^{(\kappa,m,e)}(t) \\ &\quad - \sum_{\zeta=1}^{N_f}\left(\underline{\underline{R}}^{(\zeta)}\underline{\underline{V}}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t). \end{aligned} \tag{3.18}$$

In the above, we gather the conservative variables at each node as

$$\underline{u}_i^{(h,\kappa)}(t) := \begin{bmatrix} u_i^{(h,\kappa,1)}(t) \\ \vdots \\ u_i^{(h,\kappa,N_c)}(t) \end{bmatrix}, \quad \underline{u}_i^{(h,\kappa,\zeta)}(t) := \begin{bmatrix} [\underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa,1)}(t)]_i \\ \vdots \\ [\underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa,N_c)}(t)]_i \end{bmatrix}, \tag{3.19}$$

and evaluate the physical flux and numerical flux components, respectively, as

$$\underline{f}^{(\kappa,m,e)}(t) := \begin{bmatrix} F_{me}(\underline{u}_1^{(h,\kappa)}(t)) \\ \vdots \\ F_{me}(\underline{u}_{N_q}^{(h,\kappa)}(t)) \end{bmatrix} \tag{3.20}$$

and

$$\underline{f}^{(*,\kappa,\zeta,e)}(t) \coloneqq \begin{bmatrix} F_e^*(\underline{u}_1^{(h,\kappa,\zeta)}(t),\, \underline{u}_1^{(+,\kappa,\zeta,e)}(t),\, \boldsymbol{n}^{(\kappa,\zeta)}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}^{(\zeta,1)}))) \\ \vdots \\ F_e^*(\underline{u}_{N_{q_f}^{(\zeta)}}^{(h,\kappa,\zeta)}(t),\, \underline{u}_{N_{q_f}^{(\zeta)}}^{(+,\kappa,\zeta,e)}(t),\, \boldsymbol{n}^{(\kappa,\zeta)}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}^{(\zeta,N_{q_f}^{(\zeta)})}))) \end{bmatrix}, \tag{3.21}$$

in terms of the above nodal values and the exterior solution state $\underline{u}_i^{(+,\kappa,\zeta,e)}(t) \in \Upsilon$. The projection of the initial condition in (2.24) can similarly be formulated algebraically as

$$\underline{\underline{V}}^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{\underline{V}}\,\underline{\tilde{u}}^{(h,\kappa,e)}(0) = \underline{\underline{V}}^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)} \begin{bmatrix} U^0(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}^{(1)})) \\ \vdots \\ U^0(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}^{(N_q)})) \end{bmatrix}, \tag{3.22}$$

where we note that both (3.18) and (3.22) require the solution of a linear system involving the elemental mass matrix $\underline{\underline{V}}^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{\underline{V}}$. Finally, in order to provide a unified analysis of nodal as well as modal formulations, we note that the DG scheme in (3.18) can be expressed as

$$\underline{\underline{V}}^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{\underline{V}}\,\frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = \underline{\underline{V}}^{\mathrm{T}}\underline{r}^{(h,\kappa,e)}(t), \tag{3.23}$$

where the nodal right-hand side is given by

$$\underline{r}^{(h,\kappa,e)}(t) \coloneqq \sum_{l=1}^{d}\left(\underline{\underline{D}}^{(l)}\right)^{\mathrm{T}}\underline{\underline{W}}\sum_{m=1}^{d}\underline{\underline{G}}^{(\kappa,l,m)}\underline{f}^{(\kappa,m,e)}(t) - \sum_{\zeta=1}^{N_f}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t). \tag{3.24}$$

Remark 3.3. The formulation in (3.23) can be interpreted as a discrete projection of the time derivative for a nodal discretization onto the chosen polynomial basis. However, for a given choice of approximation space and quadrature, the solution for such a scheme is in fact independent of the particular matrices $\underline{\underline{D}}^{(l)}$ and $\underline{\underline{R}}^{(\zeta)}$ used in (3.24) when (3.5a) and (3.5b) are satisfied, since, in a standard DG method, such operators are only applied to polynomial test functions, for which discrete differentiation and interpolation/extrapolation are exact. Nevertheless, the form of the operator and the choice of polynomial basis (which also has no influence on the accuracy of the numerical solution in infinite precision) can have a significant impact on the computational expense of the resulting algorithm.

3.2.2 Flux reconstruction method

Following Castonguay *et al.* [42] and Williams *et al.* [43], we define the scalar-valued *correction field* associated with each vector-valued correction function as

$$H^{(\zeta,i)}(\boldsymbol{\xi}) \coloneqq \nabla_{\boldsymbol{\xi}} \cdot \boldsymbol{G}^{(\zeta,i)}(\boldsymbol{\xi}), \tag{3.25}$$

and separate the divergence of the discontinuous flux in (2.26) from that of the correction flux in (2.27) in order to rewrite the FR method in (2.29) as

$$\begin{aligned}\frac{\mathrm{d}\underline{U}^{(h,\kappa)}(\boldsymbol{\xi}^{(i)},t)}{\mathrm{d}t} = -\frac{1}{J^{(\kappa)}(\boldsymbol{\xi}^{(i)})}\Bigg(&\sum_{l=1}^{d}\frac{\partial \underline{\hat{F}}_D^{(\kappa,l)}}{\partial \xi_l}(\boldsymbol{\xi}^{(i)},t)\\ &+\sum_{\zeta=1}^{N_f}\sum_{j=1}^{N_{qf}^{(\zeta)}} H^{(\zeta,j)}(\boldsymbol{\xi}^{(i)})\Big(J^{(\kappa,\zeta)}(\boldsymbol{\xi}^{(\zeta,j)})\underline{F}^{(*,\kappa,\zeta)}(\boldsymbol{\xi}^{(\zeta,j)},t)-\sum_{m=1}^{d}\hat{n}_m^{(\zeta)}\underline{\hat{F}}_D^{(\kappa,m)}(\boldsymbol{\xi}^{(\zeta,j)},t)\Big)\Bigg).\end{aligned} \tag{3.26}$$

Using the collocation-based operators $\underline{\underline{D}}^{(l)}$ and $\underline{\underline{R}}^{(\zeta)}$ given in (3.7a) and (3.7b), respectively, such a formulation can then be expressed algebraically as

$$\begin{aligned}\frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = -\left(\underline{\underline{J}}^{(\kappa)}\right)^{-1}\Bigg(&\sum_{l=1}^{d}\underline{\underline{D}}^{(l)}\underline{\hat{f}}^{(\kappa,l,e)}(t)\\ &+\sum_{\zeta=1}^{N_f}\underline{\underline{L}}^{(\zeta)}\Big(\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t)-\underline{\underline{R}}^{(\zeta)}\sum_{l=1}^{d}\hat{n}_l^{(\zeta)}\underline{\hat{f}}^{(\kappa,l,e)}(t)\Big)\Bigg),\end{aligned} \tag{3.27}$$

where the nodal values of the discontinuous flux components in (2.26) are given by

$$\underline{\hat{f}}^{(\kappa,l,e)}(t) \coloneqq \sum_{m=1}^{d}\underline{\underline{G}}^{(\kappa,l,m)}\underline{f}^{(\kappa,m,e)}(t), \tag{3.28}$$

and we define the *lifting matrix* in terms of the correction fields in (3.25) as

$$\underline{\underline{L}}^{(\zeta)} \coloneqq \begin{bmatrix} H^{(\zeta,1)}(\boldsymbol{\xi}^{(1)}) & \cdots & H^{(\zeta,N_{qf}^{(\zeta)})}(\boldsymbol{\xi}^{(1)}) \\ \vdots & \ddots & \vdots \\ H^{(\zeta,1)}(\boldsymbol{\xi}^{(1)}) & \cdots & H^{(\zeta,N_{qf}^{(\zeta)})}(\boldsymbol{\xi}^{(N_q)}) \end{bmatrix}. \tag{3.29}$$

Remark 3.4. Our analysis is based on properties of the scalar correction fields encoded within the lifting matrices, rather than on properties of the vector-valued correction functions, where we note that the latter do not appear directly in (3.26) nor (3.27). While energy-stable correction fields for triangles and tetrahedra are constructed in [42, 43], the corresponding correction functions are never explicitly constructed and thus serve only a conceptual role.

We consider a particular class of correction fields, satisfying the following assumption.

Assumption 3.2. *The correction fields correspond to a lifting matrix of the form*

$$\underline{\underline{L}}^{(\zeta)} \coloneqq \left(\underline{\underline{W}}+\underline{\underline{K}}\right)^{-1}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}, \tag{3.30}$$

where $\underline{\underline{W}} + \underline{\underline{K}}$ *is SPD and* $\underline{\underline{K}} \in \mathbb{R}^{N_q \times N_q}$ *is a symmetric matrix satisfying*

$$\underline{\underline{K}}\,\underline{\underline{D}}^{(l)} = \underline{0}^{(N_q \times N_q)}, \quad \forall\, l \in \{1 : d\}, \tag{3.31a}$$

$$\underline{\underline{K}}\,\underline{1}^{(N_q)} = \underline{0}^{(N_q)}. \tag{3.31b}$$

The precise form of $\underline{\underline{K}}$ generally depends on the element type and approximation space, where we note that the case of a tensor-product polynomial space on the quadrilateral or hexahedron was treated by Cicchino and Nadarajah [124] and is not detailed in this thesis. For total-degree polynomial spaces $\mathbb{P}_p(\hat{\Omega})$ on simplices, we present the following lemma.

Lemma 3.1. *Defining the derivative operator with respect to a multi-index* $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ *as*

$$\underline{\underline{D}}^{\boldsymbol{\alpha}} \coloneqq \left(\underline{\underline{D}}^{(1)}\right)^{\alpha_1} \cdots \left(\underline{\underline{D}}^{(d)}\right)^{\alpha_d}, \tag{3.32}$$

in terms of the collocation derivative matrices $\underline{\underline{D}}^{(l)}$ *given as* (3.7a) *in terms of a nodal basis for the space* $\mathbb{P}_p(\hat{\Omega})$ *with* $p \geq 1$*, the correction fields given in terms of* (3.30) *with*

$$\underline{\underline{K}} \coloneqq \begin{cases} \dfrac{c}{|\hat{\Omega}|}\left(\underline{\underline{D}}^{p}\right)^{\mathrm{T}} \underline{\underline{W}}\,\underline{\underline{D}}^{p}, & d = 1, \\ \dfrac{c}{|\hat{\Omega}|}\displaystyle\sum_{i=0}^{p} \binom{p}{i} \left(\underline{\underline{D}}^{(p-i,i)}\right)^{\mathrm{T}} \underline{\underline{W}}\,\underline{\underline{D}}^{(p-i,i)}, & d = 2, \\ \dfrac{c}{|\hat{\Omega}|}\displaystyle\sum_{i=0}^{p}\sum_{j=0}^{i} \binom{p}{i}\binom{i}{j} \left(\underline{\underline{D}}^{(p-i,i-j,j)}\right)^{\mathrm{T}} \underline{\underline{W}}\,\underline{\underline{D}}^{(p-i,i-j,j)}, & d = 3, \end{cases} \tag{3.33}$$

satisfy the conditions of Assumption 3.2, where it is sufficient to take $c \geq 0$ *when* $\underline{\underline{W}}$ *is SPD.*

Proof. Since it is clear from the definitions in (3.33) that $c \geq 0$ results in an SPSD matrix $\underline{\underline{K}}$, such a condition is sufficient[1] for $\underline{\underline{W}} + \underline{\underline{K}}$ to be SPD, as the sum of a positive-definite matrix and a positive-semidefinite matrix is positive definite. The property in (3.31a) then follows from the fact that for $|\boldsymbol{\alpha}| = q$, applying the operator $\underline{\underline{D}}^{\boldsymbol{\alpha}}\underline{\underline{D}}^{(l)}$ to any vector of nodal values corresponds to the exact differentiation of a total-degree q polynomial interpolant a total of $p+1$ times, which always yields the zero vector. Similarly, (3.31b) results from the fact that $\underline{\underline{D}}^{\boldsymbol{\alpha}}\underline{1}^{(N_q)} = \underline{0}^{(N_q)}$, since the p^{th} derivative of a constant is zero for any $p \geq 1$. □

Remark 3.5. The definitions given for $\underline{\underline{K}}$ in (3.33) are such that the original VCJH correction fields described in [41–43] are recovered when the lifting operator in (3.30) is defined using the exact reference mass matrix in (3.9) and facet mass matrices satisfying (3.12).

[1] As discussed in [41, Section 3.3], [44, Section 3.2], and [82, Section 3.6], the condition $c \geq 0$ is sufficient but not necessary for positive-definiteness of $\underline{\underline{W}} + \underline{\underline{K}}$ in the one-dimensional case.

3.3 Analysis

With the DG and FR methods recast in terms of matrix operators on the reference element, we are now able to analyze such schemes based on the algebraic properties of such constituent operators. In this section, such an approach will allow us to demonstrate the discrete equivalence between strong and weak formulations of the conservation law (and, hence, the relation between the DG and FR methods) and will facilitate unified proofs of local and global conservation as well as energy stability for the constant-coefficient linear advection equation on meshes for which the mapping from reference coordinates is affine.

3.3.1 Equivalence and unification

Just as integration by parts allows for the weak form of a PDE to be obtained from the strong form, the SBP property may be used analogously to transform a strong-form discretization into a weak-form discretization, and vice versa (as discussed, for example, in [65, Section 8.1] and [69, Section 2.1]). The following theorem establishes such an equivalence within the context of the present framework.

Theorem 3.1. *Let Assumption 3.1 hold and define* $\underline{\hat{f}}^{(\kappa,l,e)}(t)$ *as in (3.28). The weak-form right-hand side in (3.24) is then equivalent to a strong formulation, as given by*

$$\begin{aligned}\underline{r}^{(h,\kappa,e)}(t) = &-\sum_{l=1}^{d} \underline{\underline{W}}\,\underline{\underline{D}}^{(l)}\,\underline{\hat{f}}^{(\kappa,l,e)}(t) \\ &- \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \left(\underline{\underline{J}}^{(\kappa,\zeta)}\,\underline{f}^{(*,\kappa,\zeta,e)}(t) - \underline{\underline{R}}^{(\zeta)} \sum_{l=1}^{d} \hat{n}_l^{(\zeta)}\,\underline{\hat{f}}^{(\kappa,l,e)}(t) \right).\end{aligned} \tag{3.34}$$

Proof. Applying (3.6) to the first term on the right-hand side of (3.24), we obtain

$$\begin{aligned}\underline{r}^{(h,\kappa,e)}(t) = &\sum_{l=1}^{d} \left(\sum_{\zeta=1}^{N_f} \hat{n}_l^{(\zeta)} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{R}}^{(\zeta)} - \underline{\underline{W}}\,\underline{\underline{D}}^{(l)} \right) \sum_{m=1}^{d} \underline{\underline{G}}^{(\kappa,l,m)}\,\underline{f}^{(\kappa,m,e)}(t) \\ &- \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)}\,\underline{\underline{J}}^{(\kappa,\zeta)}\,\underline{f}^{(*,\kappa,\zeta,e)}(t).\end{aligned} \tag{3.35}$$

Combining the facet contributions and using (3.28) then results in (3.34). □

Remark 3.6. As discussed by Kopriva and Gassner [68] in the context of DG methods using tensor-product LG and LGL quadrature rules, strong formulations such as the right-hand side of (3.34) require the extrapolation of the normal component of the transformed flux to the facet quadrature nodes, resulting in a potential increase in computational expense relative

to the weak formulation, particularly for volume quadrature rules not including boundary nodes, in which case the operator $\underline{\underline{R}}^{(\zeta)}$ must be applied for each facet and coordinate index.

The equivalence in Theorem 3.1 can be used to express FR methods with VCJH correction fields as nodal DG methods with modified mass matrices, which we demonstrate as follows.

Theorem 3.2. *Under Assumptions 3.1 and 3.2, the scheme in (3.27) can be expressed as*

$$\left(\underline{\underline{W}} + \underline{\underline{K}}\right)\underline{\underline{J}}^{(\kappa)} \frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = \underline{r}^{(h,\kappa,e)}(t), \tag{3.36}$$

where $\underline{r}^{(h,\kappa,e)}(t)$ is the nodal right-hand side for a weak-form DG scheme given as in (3.24).

Proof. Substituting (3.30) into (3.27) and left-multiplying by $(\underline{\underline{W}} + \underline{\underline{K}})\underline{\underline{J}}^{(\kappa)}$ results in

$$\begin{aligned}\left(\underline{\underline{W}} + \underline{\underline{K}}\right)\underline{\underline{J}}^{(\kappa)} \frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t} &= -\sum_{l=1}^{d} \left(\underline{\underline{W}} + \underline{\underline{K}}\right)\underline{\underline{D}}^{(l)} \underline{\hat{f}}^{(\kappa,l,e)}(t) \\ &\quad - \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \left(\underline{\underline{J}}^{(\kappa,\zeta)} \underline{f}^{(*,\kappa,\zeta,e)}(t) - \underline{\underline{R}}^{(\zeta)} \sum_{l=1}^{d} \hat{n}_l^{(\zeta)} \underline{\hat{f}}^{(\kappa,l,e)}(t)\right).\end{aligned} \tag{3.37}$$

The contribution of the matrix $\underline{\underline{K}}$ to the first term on the right-hand side of (3.37) then vanishes as a consequence of (3.31a), which, as noted in [44, Section 3.2] and [43, Section 5], results in the following strong-form nodal DG method with a modified mass matrix:

$$\begin{aligned}\left(\underline{\underline{W}} + \underline{\underline{K}}\right)\underline{\underline{J}}^{(\kappa)} \frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t} &= -\sum_{l=1}^{d} \underline{\underline{W}}\,\underline{\underline{D}}^{(l)} \underline{\hat{f}}^{(\kappa,l,e)}(t) \\ &\quad - \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \left(\underline{\underline{J}}^{(\kappa,\zeta)} \underline{f}^{(*,\kappa,\zeta,e)}(t) - \underline{\underline{R}}^{(\zeta)} \sum_{l=1}^{d} \hat{n}_l^{(\zeta)} \underline{\hat{f}}^{(\kappa,l,e)}(t)\right).\end{aligned} \tag{3.38}$$

Finally, noting that the right-hand side of (3.38) is identical to the right-hand side of (3.34), which holds as a consequence of the SBP property in (3.6), we therefore obtain (3.36). □

Remark 3.7. As noted in Remark 3.6, the weak formulation eliminates the flux extrapolation step. Implementing the FR method in weak form as in (3.36) as opposed to the conventional strong formulation in (3.27) therefore yields a potential improvement in efficiency.

Remark 3.8. Defining the *filter matrix* as $\underline{\underline{F}} \coloneqq (\underline{\underline{I}}^{(N_q)} + \underline{\underline{W}}^{-1}\underline{\underline{K}})^{-1}$ and letting the corresponding nodal DG scheme be given by (3.36) with $\underline{\underline{K}} = \underline{\underline{0}}^{(N_q \times N_q)}$, the FR and nodal DG formulations are related as

$$\left(\underline{\underline{J}}^{(\kappa)} \frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t}\right)_{\mathrm{FR}} = \underline{\underline{F}} \left(\underline{\underline{J}}^{(\kappa)} \frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t}\right)_{\mathrm{DG}}, \tag{3.39}$$

where, as shown in [44, Section 3.2] and [43, Appendix B], the filter acts only on modes for which $|\boldsymbol{\alpha}| = q$ when $\underline{\underline{K}}$ is defined as in (3.33) using the exact mass matrix in (3.9). Since

Theorem 3.1 establishes that the strong and weak formulations are equivalent when the SBP property is satisfied, we see that VCJH schemes can be obtained through filtering of the Jacobian-weighted time derivative for a nodal DG scheme *in strong or weak form.*

Representing the solution as in (2.20) in terms of an arbitrary (i.e. nodal or modal) basis $\{\phi^{(i)}\}_{i\in\{1:N_p\}}$ and defining the modified FR mass matrix for such a basis as

$$\underline{\underline{\tilde{M}}}^{(\kappa)} := \underline{\underline{V}}^{\mathrm{T}}\Big(\underline{\underline{W}} + \underline{\underline{K}}\Big)\underline{\underline{J}}^{(\kappa)}\underline{\underline{V}}, \tag{3.40}$$

we obtain a generalized formulation given in terms of the nodal right-hand side in (3.24) as

$$\underline{\underline{\tilde{M}}}^{(\kappa)}\frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = \underline{\underline{V}}^{\mathrm{T}}\underline{r}^{(h,\kappa,e)}(t). \tag{3.41}$$

Such a formulation recovers the DG scheme in (3.18) when $\underline{\underline{K}} = \underline{\underline{0}}^{(N_q\times N_q)}$, where any set of SBP operators satisfying Assumption 3.1 may be used. We likewise obtain the standard FR scheme in (3.27) under Assumptions 3.1 and 3.2 when $\underline{\underline{V}}$ is the identity matrix and $\underline{\underline{D}}^{(l)}$ and $\underline{\underline{R}}^{(\zeta)}$ are defined as in (3.7a) and (3.7b), respectively, where we note that such a simplification results from the choice of a nodal basis satisfying $\phi^{(i)}(\boldsymbol{\xi}^{(j)}) = \delta_{ij}$, for which $N_q = N_p$ is a necessary condition. If we make both of the above simplifications, we recover a nodal DG scheme given by

$$\begin{aligned}\frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = \Big(\underline{\underline{J}}^{(\kappa)}\Big)^{-1}\Bigg(&\sum_{l=1}^{d}\underline{\underline{W}}^{-1}\Big(\underline{\underline{D}}^{(l)}\Big)^{\mathrm{T}}\underline{\underline{W}}\sum_{m=1}^{d}\underline{\underline{G}}^{(\kappa,l,m)}\underline{f}^{(\kappa,m,e)}(t)\\ &-\sum_{\zeta=1}^{N_f}\underline{\underline{W}}^{-1}\Big(\underline{\underline{R}}^{(\zeta)}\Big)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t)\Bigg).\end{aligned} \tag{3.42}$$

The present framework can then be said to *unify* several existing discretization approaches, and the remainder of the analysis in this section will therefore employ the unified formulation in (3.41), where we note the above simplifications occurring for $\underline{\underline{K}} = \underline{\underline{0}}^{(N_q\times N_q)}$ and $\underline{\underline{V}} = \underline{\underline{I}}^{(N_q)}$.

Remark 3.9. We are not restricted to the use of discretizations employing polynomial expansions, as the analysis in this section immediately extends to methods in the form of (3.42) using *any* SBP approximation $\underline{\underline{D}}^{(l)}$ of each partial derivative $\partial/\partial\xi_l$ satisfying the conditions of Assumption 3.1, which may not correspond to any analytical (i.e. polynomial or non-polynomial) basis. For such schemes, the solution to (2.1) is approximated as

$$U_e(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}^{(i)}), t) \approx u_i^{(h,\kappa,e)}(t), \quad \forall\, i \in \{1 : N_q\}, \tag{3.43}$$

where, when no explicit basis is present, the numerical solution is only defined at the nodes, similarly to the finite-difference methods from which the SBP concept originates.

3.3.2 Local and global conservation

The following theorem establishes that the unified DSEM formulation in (3.41) is locally conservative in an element-wise sense, a property which, by way of the well-known Lax–Wendroff theorem (see, for example, the original result in [125] and its modern generalization by Shi and Shu [126]) ensures that a convergent discretization converges to a weak solution under mesh refinement, and is thus suitable for the treatment of problems with discontinuities.

Theorem 3.3. *Let Assumptions 3.1 and 3.2 hold and define* $\underline{r}^{(h,\kappa,e)}(t)$ *and* $\underline{\underline{\tilde{M}}}^{(\kappa)}$ *as in (3.24) and (3.40), respectively. The resulting scheme in (3.41) is then locally conservative, satisfying*

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{u}^{(h,\kappa,e)}(t) = -\sum_{\zeta=1}^{N_f}\left(\underline{1}^{(N_{q_f}^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\,\underline{\underline{J}}^{(\kappa,\zeta)}\,\underline{f}^{(*,\kappa,\zeta,e)}(t). \tag{3.44}$$

Proof. Since constant functions lie within the span of $\{\phi^{(i)}\}_{i\in\{1:N_q\}}$ under Assumption 3.1, there exists a vector $\underline{1} \in \mathbb{R}^{N_p}$ satisfying $\underline{\underline{V}}\,\underline{1} = \underline{1}^{(N_q)}$. Multiplying both sides of (3.41) from the left by the transpose of such a vector (i.e. taking a constant test function in a discrete sense), we can use the fact that (3.31b) holds under Assumption 3.2 to obtain

$$\begin{aligned}\underline{1}^{\mathrm{T}}\underline{\underline{\tilde{M}}}^{(\kappa)}\frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa,e)}(t)}{\mathrm{d}t} &= \left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\left(\underline{\underline{W}}+\underline{\underline{K}}\right)\underline{\underline{J}}^{(\kappa)}\frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t}\\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{u}^{(h,\kappa,e)}(t).\end{aligned} \tag{3.45}$$

Noting that the volume contributions in (3.24) vanish when left-multiplied by $(\underline{1}^{(N_q)})^{\mathrm{T}}$ as a consequence of the fact that constants are differentiated exactly under Assumption 3.1 as $\underline{\underline{D}}^{(l)}\underline{1}^{(N_q)} = \underline{0}^{(N_q)}$ and similarly using $\underline{\underline{R}}^{(\zeta)}\underline{1}^{(N_q)} = \underline{1}^{(N_{q_f}^{(\zeta)})}$ to simplify the facet contributions then gives

$$\begin{aligned}\underline{1}^{\mathrm{T}}\underline{\underline{V}}^{\mathrm{T}}\underline{r}^{(h,\kappa,e)}(t) &= \left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{r}^{(h,\kappa,e)}(t)\\ &= -\sum_{\zeta=1}^{N_f}\left(\underline{1}^{(N_{q_f}^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\,\underline{\underline{J}}^{(\kappa,\zeta)}\,\underline{f}^{(*,\kappa,\zeta,e)}(t).\end{aligned} \tag{3.46}$$

Using (3.45) on the left-hand side and (3.46) on the right-hand side therefore results in the statement of local conservation in (3.44). □

To prove that the scheme is globally conservative, meaning that no spurious generation or destruction of the integrated conservative variables occurs over the entire domain, we make the following assumption regarding the conformity of the mesh.

Assumption 3.3. *For each pair of element indices* $\kappa, \nu \in \{1 : N_e\}$ *with* $\kappa \neq \nu$ *such that* $\partial\Omega^{(\kappa)} \cap \partial\Omega^{(\nu)} \neq \emptyset$*, there exist a unique pair of indices* $\zeta, \eta \in \{1 : N_f\}$ *such that* $\Gamma^{(\kappa,\zeta)} = \Gamma^{(\nu,\eta)}$ *and the following two conditions are satisfied:*

- *For every* $i \in \{1 : N_{q_f}^{(\zeta)}\}$ *there exists a unique* $j \in \{1 : N_{q_f}^{(\eta)}\}$ *for which we have*

$$\boldsymbol{n}^{(\kappa,\zeta)}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}^{(\zeta,i)})) = -\boldsymbol{n}^{(\nu,\eta)}(\boldsymbol{X}^{(\nu)}(\boldsymbol{\xi}^{(\eta,j)})). \tag{3.47}$$

- *The facet matrices* $\underline{\underline{B}}^{(\eta)}\underline{\underline{J}}^{(\nu,\eta)}$ *and* $\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}$ *are related through the permutation matrix* $\underline{\underline{T}}^{(\kappa,\zeta)} \in \mathbb{R}^{N_{q_f}^{(\zeta)} \times N_{q_f}^{(\zeta)}}$ *corresponding to the bijective mapping* $i \mapsto j$ *in (3.47) as*

$$\underline{\underline{B}}^{(\eta)}\underline{\underline{J}}^{(\nu,\eta)} = \left(\underline{\underline{T}}^{(\kappa,\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{\underline{T}}^{(\kappa,\zeta)}. \tag{3.48}$$

Both of such conditions also hold for facets $\Gamma^{(\kappa,\zeta)}$ *and* $\Gamma^{(\nu,\eta)}$ *which are connected through periodic boundary conditions imposed using the numerical flux function.*

We now have the following theorem establishing global conservation.

Theorem 3.4. *Suppose that a discretization is locally conservative, satisfying (3.44) for all* $\kappa \in \{1 : N_e\}$*, that the numerical flux is conservative in the sense of (2.23b), and that Assumption 3.3 holds. The resulting scheme is then globally conservative, satisfying*

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{\kappa=1}^{N_e}\left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{u}^{(h,\kappa,e)}(t) = -\sum_{\Gamma^{(\kappa,\zeta)}\subset\partial\Omega}\left(\underline{1}^{(N_{q_f}^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t), \tag{3.49}$$

where the right-hand side vanishes when the entire boundary $\partial\Omega$ *is taken to be periodic.*

Proof. To demonstrate that the scheme is globally conservative, we sum (3.44) over all elements in order to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{\kappa=1}^{N_e}\left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{u}^{(h,\kappa,e)}(t) = \sum_{\kappa=1}^{N_e}\sum_{\zeta=1}^{N_f}\left(\underline{1}^{(N_{q_f}^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t). \tag{3.50}$$

Considering two adjacent elements $\Omega^{(\kappa)}$ and $\Omega^{(\nu)}$ which share an interior facet $\Gamma^{(\kappa,\zeta)} = \Gamma^{(\nu,\eta)}$, the net contribution to (3.50) arising from such an interface is given by

$$\Phi_e^{(\kappa,\zeta)}(t) := \left(\underline{1}^{(N_{q_f}^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t) + \left(\underline{1}^{(N_q^{(\eta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\eta)}\underline{\underline{J}}^{(\nu,\eta)}\underline{f}^{(*,\nu,\eta,e)}(t). \tag{3.51}$$

Using (3.48) to combine the two terms in (3.51), we then obtain

$$\begin{aligned}\Phi_e^{(\kappa,\zeta)}(t) &= \left(\underline{1}^{(N_{q_f}^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t) + \left(\underline{1}^{(N_q^{(\eta)})}\right)^{\mathrm{T}}\left(\underline{\underline{T}}^{(\kappa,\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{\underline{T}}^{(\kappa,\zeta)}\underline{f}^{(*,\nu,\eta,e)}(t) \\ &= \left(\underline{1}^{(N_{q_f}^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\left(\underline{f}^{(*,\kappa,\zeta,e)}(t) + \underline{\underline{T}}^{(\kappa,\zeta)}\underline{f}^{(*,\nu,\eta,e)}(t)\right),\end{aligned} \tag{3.52}$$

which is zero due to the fact that $\underline{f}^{(*,\kappa,\zeta,e)}(t) = -\underline{\underline{T}}^{(\kappa,\zeta)}\underline{f}^{(*,\nu,\eta,e)}(t)$ when (2.23a) and (3.47) are satisfied. Since only the facets lying on the domain boundary remain in the sum on

the right-hand side of (3.50), we therefore obtain the statement of global conservation in (3.49). Since periodic interfaces also satisfy (3.47) and (3.48) under Assumption 3.3, their contribution to the right-hand side of (3.50) thus vanishes as well. □

Remark 3.10. Although we invoke Assumption 3.3 in the above proof and throughout this thesis to simplify our analysis, the extension of the theory to nonconforming meshes arising from h-adaptivity or p-adaptivity is, in principle, straightforward, provided that suitable interface interpolation operators are employed (see, for example, Kozdon and Wilcox [127], Del Rey Fernández *et al.* [128], Shadpey and Zingg [129], and Chan *et al.* [130]).

3.3.3 Energy stability

While certain modifications, which we will discuss in Chapters 5 and 6, are required to obtain provably stable schemes for curvilinear meshes and nonlinear problems, the SBP property allows one to prove that the standard DG and FR methods described in Section 3.2 are energy stable for the constant-coefficient linear advection equation on affine meshes. To demonstrate this, we first require the following lemma.

Lemma 3.2. *Let $\underline{r}^{(h,\kappa)}(t)$ denote the nodal right-hand side in (3.24) for a discretization of the constant-coefficient linear advection equation, which is given for $\boldsymbol{a} \in \mathbb{R}^d$ by*

$$\frac{\partial U(\boldsymbol{x},t)}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \Big(\boldsymbol{a} U(\boldsymbol{x},t)\Big) = 0, \quad \forall\, (\boldsymbol{x},t) \in \Omega \times (0,T), \tag{3.53}$$

where Assumption 3.1 holds and the mapping is affine, satisfying $\boldsymbol{X}^{(\kappa)} \in [\mathbb{P}_1(\hat{\Omega})]^d$. Any nodal solution vector $\underline{u}^{(h,\kappa)}(t) \in \mathbb{R}^{N_q}$ then satisfies

$$\begin{aligned}\Big(\underline{u}^{(h,\kappa)}(t)\Big)^{\mathrm{T}} \underline{r}^{(h,\kappa)}(t) = \sum_{\zeta=1}^{N_f} \Bigg(\sum_{m=1}^{d} \frac{a_m}{2} \Big(\underline{u}^{(h,\kappa)}(t)\Big)^{\mathrm{T}} \Big(\underline{\underline{R}}^{(\zeta)}\Big)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{J}}^{(\kappa,\zeta)} \underline{\underline{N}}^{(\kappa,\zeta,m)} \underline{\underline{R}}^{(\zeta)} \underline{u}^{(h,\kappa)}(t) \\ - \Big(\underline{u}^{(h,\kappa)}(t)\Big)^{\mathrm{T}} \Big(\underline{\underline{R}}^{(\zeta)}\Big)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{J}}^{(\kappa,\zeta)} \underline{f}^{(*,\kappa,\zeta)}(t) \Bigg),\end{aligned} \tag{3.54}$$

where the equation index has been suppressed since we are considering a scalar problem.

Proof. Multiplying (3.24) from the left by $(\underline{u}^{(h,\kappa)}(t))^{\mathrm{T}}$, which corresponds to taking the solution as a test function, and inserting the linear flux $\underline{f}^{(\kappa,m,e)}(t) = a_m \underline{u}^{(h,\kappa)}(t)$, we obtain

$$\begin{aligned}\Big(\underline{u}^{(h,\kappa)}(t)\Big)^{\mathrm{T}} \underline{r}^{(h,\kappa)}(t) = \sum_{m=1}^{d} a_m \sum_{l=1}^{d} \Big(\underline{u}^{(h,\kappa)}(t)\Big)^{\mathrm{T}} \Big(\underline{\underline{D}}^{(l)}\Big)^{\mathrm{T}} \underline{\underline{W}}\,\underline{\underline{G}}^{(\kappa,l,m)} \underline{u}^{(h,\kappa)}(t) \\ - \sum_{\zeta=1}^{N_f} \Big(\underline{u}^{(h,\kappa)}(t)\Big)^{\mathrm{T}} \Big(\underline{\underline{R}}^{(\zeta)}\Big)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{J}}^{(\kappa,\zeta)} \underline{f}^{(*,\kappa,\zeta)}(t).\end{aligned} \tag{3.55}$$

Using the SBP property in (3.6) to discretely integrate by parts in reference coordinates as

$$\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{D}}^{(l)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{u}^{(h,\kappa)}(t) = \frac{1}{2}\sum_{\zeta=1}^{N_f}\hat{n}_l^{(\zeta)}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa)}(t) \tag{3.56}$$

and noting that the metric terms are constant such that $\underline{\underline{G}}^{(\kappa,l,m)} = G_{lm}^{(\kappa)}\underline{\underline{I}}^{(N_q)}$ when the mapping is affine, we can express the first term on the right-hand side of (3.55) as

$$\begin{aligned}&\sum_{m=1}^{d} a_m \sum_{l=1}^{d}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{D}}^{(l)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{G}}^{(\kappa,l,m)}\underline{u}^{(h,\kappa)}(t) = \\ &\quad\sum_{m=1}^{d}\frac{a_m}{2}\left(\sum_{l=1}^{d}G_{lm}^{(\kappa)}\hat{n}_l^{(\zeta)}\right)\sum_{\zeta=1}^{N_f}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa)}(t).\end{aligned} \tag{3.57}$$

Using (2.14) on the right-hand side of (3.57) to obtain $\underline{\underline{J}}^{(\kappa,\zeta)}\underline{\underline{N}}^{(\kappa,\zeta,m)}$ from the metric terms and substituting the result into (3.55) then results in (3.54). □

The result in (3.54) allows for the contributions to the semi-discrete energy balance arising from the volume terms of (3.24) to be expressed as interface terms, which we use to prove the following theorem.

Theorem 3.5. *Let all of Assumptions 3.1 to 3.3 hold and assume that the numerical flux for the constant-coefficient linear advection equation takes the form*

$$F^*(U^-,U^+,\boldsymbol{n}) \coloneqq \frac{1}{2}(\boldsymbol{a}\cdot\boldsymbol{n})(U^+ + U^-) - \frac{\lambda}{2}|\boldsymbol{a}\cdot\boldsymbol{n}|(U^+ - U^-), \tag{3.58}$$

where the parameter $\lambda \in \mathbb{R}_0^+$ takes the same value on each side of an interior or periodic interface, and that the mapping from the reference element onto each physical element is affine and satisfies (2.11). Defining the solution energy for the formulation in (3.41) as

$$\mathcal{E}^h(t) \coloneqq \frac{1}{2}\sum_{\kappa=1}^{N_e}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{W}} + \underline{\underline{K}}\right)\underline{\underline{J}}^{(\kappa)}\underline{u}^{(h,\kappa)}(t), \tag{3.59}$$

such a scheme then satisfies the following semi-discrete energy estimate when applied to the constant-coefficient linear advection equation:

$$\begin{aligned}\frac{\mathrm{d}\mathcal{E}^h(t)}{\mathrm{d}t} \leq \sum_{\Gamma^{(\kappa,\zeta)}\subset\partial\Omega}\Bigg(&\sum_{m=1}^{d}\frac{a_m}{2}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{\underline{N}}^{(\kappa,\zeta,m)}\underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa)}(t) \\ &- \left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta)}(t)\Bigg).\end{aligned} \tag{3.60}$$

Proof. Multiplying (3.41) from the left by $(\underline{\tilde{u}}^{(h,\kappa)}(t))^{\mathrm{T}}$ and using (3.17) as well as the fact

that $(\underline{\underline{W}} + \underline{\underline{K}})\underline{\underline{J}}^{(\kappa)}$ is symmetric when the mapping is affine, the left-hand side of the energy balance becomes

$$\begin{aligned}\left(\underline{\tilde{u}}^{(h,\kappa)}(t)\right)^{\mathrm{T}} \underline{\underline{\tilde{M}}}^{(\kappa)} \frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa)}(t)}{\mathrm{d}t} &= \left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}} \left(\underline{\underline{W}} + \underline{\underline{K}}\right) \underline{\underline{J}}^{(\kappa)} \frac{\mathrm{d}\underline{u}^{(h,\kappa)}(t)}{\mathrm{d}t} \\ &= \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}} \left(\underline{\underline{W}} + \underline{\underline{K}}\right) \underline{\underline{J}}^{(\kappa)} \underline{u}^{(h,\kappa)}(t).\end{aligned} \tag{3.61}$$

Using (3.17) and (3.54) on the right-hand side and summing over all elements, the rate of change in the solution energy given by (3.59) can be expressed as

$$\begin{aligned}\frac{\mathrm{d}\mathcal{E}^h(t)}{\mathrm{d}t} = \sum_{\kappa=1}^{N_e}\sum_{\zeta=1}^{N_f}\Bigg(&\sum_{m=1}^{d}\frac{a_m}{2}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{\underline{N}}^{(\kappa,\zeta,m)}\underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa)}(t) \\ &- \left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta)}(t)\Bigg).\end{aligned} \tag{3.62}$$

As in the proof of Theorem 3.4, we can bound the right-hand side of (3.62) by considering two adjacent elements $\Omega^{(\kappa)}$ and $\Omega^{(\nu)}$ which share an interior facet $\Gamma^{(\kappa,\zeta)} = \Gamma^{(\nu,\eta)}$. Defining the diagonal matrix of wave speeds in the normal direction for convenience, as given by

$$\underline{\underline{A}}^{(\kappa,\zeta)} \coloneqq \sum_{m=1}^{d} a_m \underline{\underline{N}}^{(\kappa,\zeta,m)}, \tag{3.63}$$

with $|\underline{\underline{A}}^{(\kappa,\zeta)}|$ containing the absolute values of the entries of $\underline{\underline{A}}^{(\kappa,\zeta)}$, we can use Assumption 3.3 to express the contributions from the numerical flux in (3.58) as

$$\begin{aligned}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta)}(t) = &\frac{1}{2}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\kappa,\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{\underline{A}}^{(\kappa,\zeta)}\underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa)}(t) \\ &+ \frac{1}{2}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\kappa,\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{\underline{A}}^{(\kappa,\zeta)}\underline{\underline{T}}^{(\kappa,\zeta)}\underline{\underline{R}}^{(\eta)}\underline{u}^{(h,\nu)}(t) \\ &- \frac{\lambda}{2}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\kappa,\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}|\underline{\underline{A}}^{(\kappa,\zeta)}|\underline{\underline{T}}^{(\kappa,\zeta)}\underline{\underline{R}}^{(\eta)}\underline{u}^{(h,\nu)}(t) \\ &+ \frac{\lambda}{2}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\kappa,\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}|\underline{\underline{A}}^{(\kappa,\zeta)}|\underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa)}(t)\end{aligned} \tag{3.64}$$

and

$$\begin{aligned}\left(\underline{u}^{(h,\nu)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\eta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\eta)}\underline{\underline{J}}^{(\nu,\eta)}\underline{f}^{(*,\kappa,\zeta)}(t) = &-\frac{1}{2}\left(\underline{u}^{(h,\nu)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\eta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\eta)}\underline{\underline{J}}^{(\nu,\eta)}\underline{\underline{A}}^{(\nu,\eta)}\underline{\underline{R}}^{(\eta)}\underline{u}^{(h,\nu)}(t) \\ &- \frac{1}{2}\left(\underline{u}^{(h,\nu)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\eta)}\right)^{\mathrm{T}}\left(\underline{\underline{T}}^{(\kappa,\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{\underline{A}}^{(\kappa,\zeta)}\underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa)}(t) \\ &- \frac{\lambda}{2}\left(\underline{u}^{(h,\nu)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\eta)}\right)^{\mathrm{T}}\left(\underline{\underline{T}}^{(\kappa,\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\left|\underline{\underline{A}}^{(\kappa,\zeta)}\right|\underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa)}(t) \\ &+ \frac{\lambda}{2}\left(\underline{u}^{(h,\nu)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\left(\underline{\underline{T}}^{(\kappa,\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\left|\underline{\underline{A}}^{(\kappa,\zeta)}\right|\underline{\underline{T}}^{(\kappa,\zeta)}\underline{\underline{R}}^{(\eta)}\underline{u}^{(h,\nu)}(t).\end{aligned} \tag{3.65}$$

Using the above expressions, the net contribution to the right-hand side of (3.62) arising at such an interface is then given by

$$-\frac{\lambda}{2}\left(\underline{d}^{(\kappa,\zeta)}(t)\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\left|\underline{\underline{A}}^{(\kappa,\zeta)}\right|\underline{d}^{(\kappa,\zeta)}(t), \tag{3.66}$$

where the jump in the solution at such an interface is given by

$$\underline{d}^{(\kappa,\zeta)}(t) \coloneqq \underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa)}(t) - \underline{\underline{T}}^{(\kappa,\zeta)}\underline{\underline{R}}^{(\eta)}\underline{u}^{(h,\nu)}(t), \tag{3.67}$$

and we note that the dissipation rate in (3.66) is zero for $\lambda = 0$ and non-positive for $\lambda > 0$ due to the positive semidefiniteness of $\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}|\underline{\underline{A}}^{(\kappa,\zeta)}|$. Applying the above procedure to all interior interfaces, we therefore obtain (3.60). □

Remark 3.11. Treating periodic interfaces in the same fashion as interior interfaces, a similar argument to that employed in the proof of Theorem 3.5 can be used to obtain

$$\frac{\mathrm{d}\mathcal{E}^h(t)}{\mathrm{d}t} \leq 0 \tag{3.68}$$

when all boundary conditions are periodic and $\lambda \geq 0$, where such a relation becomes an equality in the case of $\lambda = 0$, corresponding to semi-discrete *energy conservation*.

3.4 Chapter summary

This chapter presents the essential components of our proposed methodology for the matrix-based formulation and analysis of DSEMs for conservation laws. We describe the construction of collocation-based and quadrature-based discretization operators satisfying the SBP property on the reference element, which we use to obtain a unified matrix formulation for standard DG and FR methods in strong or weak form. Algebraic proofs of conservation and energy stability are then presented for such a unified formulation, establishing a common set of tools which will be employed throughout the remainder of this thesis for the analysis of the novel schemes described therein.

Chapter 4

Tensor-product summation-by-parts operators on the reference triangle and tetrahedron

In this chapter, which is based on the content of Papers II and III, a methodology is presented for constructing tensor-product spectral-element operators of any order with the SBP property on the reference triangle and tetrahedron. These operators are sparse and support sum factorization, and will be used in Chapters 5 and 6 to construct efficient energy-stable and entropy-stable discretizations, respectively, on curved triangles and tetrahedra.

4.1 Tensor-product SBP operators on the reference triangle

The operators which we will construct in this section are defined on the reference triangle with vertices at $[-1,-1]^{\mathrm{T}}$, $[1,-1]^{\mathrm{T}}$, and $[-1,1]^{\mathrm{T}}$, which is given by

$$\hat{\Omega} \coloneqq \Big\{ \boldsymbol{\xi} \in [-1,1]^2 : \xi_1 + \xi_2 \leq 0 \Big\}, \tag{4.1}$$

where, as a convention, we number the facets (i.e. edges of the triangle) as

$$\hat{\Gamma}^{(1)} \coloneqq \Big\{ \boldsymbol{\xi} \in \hat{\Omega} : \xi_2 = -1 \Big\}, \tag{4.2a}$$
$$\hat{\Gamma}^{(2)} \coloneqq \Big\{ \boldsymbol{\xi} \in \hat{\Omega} : \xi_1 + \xi_2 = 0 \Big\}, \tag{4.2b}$$
$$\hat{\Gamma}^{(3)} \coloneqq \Big\{ \boldsymbol{\xi} \in \hat{\Omega} : \xi_1 = -1 \Big\}. \tag{4.2c}$$

The outward unit normal vectors to each facet are then given by

$$\hat{\boldsymbol{n}}^{(1)} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \quad \hat{\boldsymbol{n}}^{(2)} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \quad \hat{\boldsymbol{n}}^{(3)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}. \tag{4.3}$$

The proposed schemes are constructed based on a *collapsed coordinate system* $\boldsymbol{\eta} \in [-1,1]^2$, from which any point can be mapped onto the *reference coordinate system* $\boldsymbol{\xi} \in \hat{\Omega}$ through

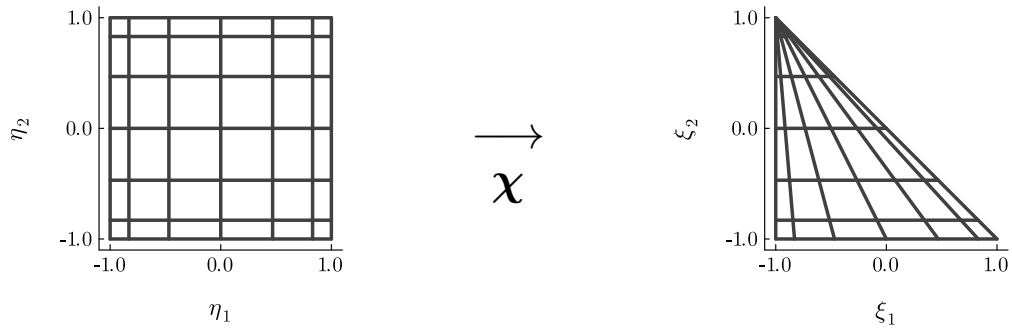


Figure 4.1: Illustration of the mapping $\boldsymbol{\xi} = \boldsymbol{\chi}(\boldsymbol{\eta})$ from the square to the reference triangle

the mapping $\boldsymbol{\chi} : [-1, 1]^2 \to \hat{\Omega}$ depicted in Figure 4.1, which is given by

$$\boldsymbol{\chi}(\boldsymbol{\eta}) \coloneqq \begin{bmatrix} \frac{1}{2}(1+\eta_1)(1-\eta_2) - 1 \\ \eta_2 \end{bmatrix}, \tag{4.4}$$

where we adopt a similar notation to [110, Section 3.2]. Such a mapping has an inverse given away from the singularity by

$$\boldsymbol{\chi}^{-1}(\boldsymbol{\xi}) \coloneqq \begin{bmatrix} 2(1+\xi_1)/(1-\xi_2) - 1 \\ \xi_2 \end{bmatrix}, \quad \forall\, \boldsymbol{\xi} \in \hat{\Omega} \setminus \left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}. \tag{4.5}$$

It is then straightforward to show that integrals on the reference element can be expressed in terms of the collapsed coordinate system as

$$\int_{\hat{\Omega}} V(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} = \int_{-1}^{1} \int_{-1}^{1} V(\boldsymbol{\chi}(\boldsymbol{\eta})) \frac{1-\eta_2}{2} \, \mathrm{d}\eta_1 \mathrm{d}\eta_2. \tag{4.6}$$

Similarly, partial derivatives with respect to each reference coordinate may be computed in terms of the collapsed coordinate system via the chain rule as

$$\frac{\partial V}{\partial \xi_1}(\boldsymbol{\chi}(\boldsymbol{\eta})) = \frac{2}{1-\eta_2} \frac{\partial}{\partial \eta_1} V(\boldsymbol{\chi}(\boldsymbol{\eta})), \quad \frac{\partial V}{\partial \xi_2}(\boldsymbol{\chi}(\boldsymbol{\eta})) = \left(\frac{1+\eta_1}{1-\eta_2} \frac{\partial}{\partial \eta_1} + \frac{\partial}{\partial \eta_2} \right) V(\boldsymbol{\chi}(\boldsymbol{\eta})), \tag{4.7}$$

where we note that such expressions are undefined at the top edge (i.e. $\eta_2 = 1$), which collapses onto the singular vertex $\boldsymbol{\xi} = [-1, 1]^{\mathrm{T}}$ under the transformation in (4.4).

4.1.1 Nodal sets and interpolants

Let q_1 and q_2 denote the degrees of the polynomial approximations with respect to η_1 and η_2, respectively, and define $q \coloneqq \min(q_1, q_2)$, where we allow for the use of different polynomial degrees in each direction and note that this flexibility could be exploited, for example, within

the context of anisotropic p-adaptivity. We then construct Gaussian quadrature rules as in Section 2.4 with nodes and weights given by

$$\begin{aligned} \{\eta_1^{(i)}\}_{i\in\{0:q_1\}} &\subset [-1,1], \quad \{\eta_2^{(i)}\}_{i\in\{0:q_2\}} \subset [-1,1), \\ \{\omega_1^{(i)}\}_{i\in\{0:q_1\}} &\subset \mathbb{R}^+, \qquad \{\omega_2^{(i)}\}_{i\in\{0:q_2\}} \subset \mathbb{R}^+, \end{aligned} \tag{4.8}$$

where we note that a half-open interval in the η_2 coordinate is used to avoid the singularity of the mapping in (4.4). Letting $a_m, b_m > -1$ for $m \in \{1,2\}$, such quadrature rules satisfy

$$\sum_{i=0}^{q_m} V(\eta_m^{(i)})\,\omega_m^{(i)} = \int_{-1}^{1} V(\eta_m)(1-\eta_m)^{a_m}(1+\eta_m)^{b_m}\,\mathrm{d}\eta_m, \quad \forall\, V \in \mathbb{P}_{\tau_m^{(a_m,b_m)}}([-1,1]), \tag{4.9}$$

where we recall that $\tau_m^{(a_m,b_m)} = 2q_m + 1$ for Gauss quadrature, $\tau_m^{(a_m,b_m)} = 2q_m$ for Gauss–Radau quadrature, and $\tau_m^{(a_m,b_m)} = 2q_m - 1$ for Gauss–Lobatto quadrature rules. Defining the corresponding Lagrange polynomials as in (2.32), we can interpolate a function $V : \hat{\Omega} \to \mathbb{R}$ on the reference triangle by way of the mapping in (4.4) as

$$(\mathcal{I}_q V)(\boldsymbol{\chi}(\boldsymbol{\eta})) \coloneqq \sum_{\alpha_1=0}^{q_1} \sum_{\alpha_2=0}^{q_2} V(\boldsymbol{\chi}(\eta_1^{(\alpha_1)}, \eta_2^{(\alpha_2)}))\ell_1^{(\alpha_1)}(\eta_1)\ell_2^{(\alpha_2)}(\eta_2). \tag{4.10}$$

The above interpolant is, in general, a rational function with a singularity at $\boldsymbol{\xi} = [-1,1]^{\mathrm{T}}$ when expressed in reference coordinates under the inverse mapping in (4.5). However, despite being based on a rational function space (the approximation properties of which are analyzed by Shen *et al.* [131] and Li and Wang [132]), the interpolation is exact for all polynomials of up to degree q, as characterized by the following lemma.

Lemma 4.1. *The tensor-product interpolation operator defined in (4.10) is exact for any polynomial $V \in \mathbb{P}_q(\hat{\Omega})$, satisfying $(\mathcal{I}_q V)(\boldsymbol{\xi}) = V(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \hat{\Omega} \setminus \{[-1,1]^{\mathrm{T}}\}$.*

Proof. First, we note that any monomial of the form $\pi^{(\alpha_1,\alpha_2)}(\boldsymbol{\xi}) \coloneqq \xi_1^{\alpha_1}\xi_2^{\alpha_2}$ with $\boldsymbol{\alpha} \in \mathcal{P}(q)$ can be expressed in terms of the collapsed coordinate system as

$$\pi^{(\alpha_1,\alpha_2)}(\boldsymbol{\chi}(\boldsymbol{\eta})) = \left(\tfrac{1}{2}(1+\eta_1)(1-\eta_2) - 1\right)^{\alpha_1}\eta_2^{\alpha_2}, \tag{4.11}$$

which is of maximum degree α_1 in η_1 and of maximum degree $\alpha_1 + \alpha_2$ in η_2. Since $\alpha_1 \le q_1$ and $\alpha_1 + \alpha_2 \le q_2$ for all $\boldsymbol{\alpha} \in \mathcal{P}(q)$, it follows from expressing any polynomial $V \in \mathbb{P}_q(\hat{\Omega})$ as a linear combination of monomials in the form of (4.11) that the composite function $V \circ \boldsymbol{\chi}$ lies within the span of the tensor-product basis, and is therefore interpolated exactly. □

4.1.2 Volume quadrature

Defining the multi-index set $\mathcal{Q}(q_1, q_2) \coloneqq \{0 : q_1\} \times \{0 : q_2\}$, we let $\sigma : \mathcal{Q}(q_1, q_2) \to \{1 : N_q\}$ denote a bijective mapping which defines an ordering of the $N_q \coloneqq (q_1 + 1)(q_2 + 1)$ tensor-product quadrature nodes on the square. Using the mapping in (4.4), we can then define a quadrature rule in the form of (3.1) approximating integrals on the reference triangle under the change of variables in (4.6), where we let the nodes $\{\boldsymbol{\xi}^{(i)}\}_{i\in\{1:N_q\}} \subset \hat{\Omega}$ be given for a multi-index $\boldsymbol{\alpha} \in \mathcal{Q}(q_1, q_2)$ by

$$\boldsymbol{\xi}^{(\sigma(\boldsymbol{\alpha}))} \coloneqq \boldsymbol{\chi}(\eta_1^{(\alpha_1)}, \eta_2^{(\alpha_2)}). \tag{4.12}$$

The corresponding weights $\{\omega^{(i)}\}_{i\in\{1:N_q\}}$ are given for $(a_1, b_1) = (0, 0)$ and $(a_2, b_2) = (0, 0)$ by

$$\omega^{(\sigma(\boldsymbol{\alpha}))} \coloneqq \frac{1 - \eta_2^{(\alpha_2)}}{2} \omega_1^{(\alpha_1)} \omega_2^{(\alpha_2)} \tag{4.13}$$

or, alternatively, by

$$\omega^{(\sigma(\boldsymbol{\alpha}))} \coloneqq \frac{1}{2} \omega_1^{(\alpha_1)} \omega_2^{(\alpha_2)} \tag{4.14}$$

if a Jacobi-type quadrature rule with $(a_2, b_2) = (1, 0)$ is instead used as in [56, Section 2.2.1] to subsume the factor of $1 - \eta_2$ arising from the change of variables in (4.6).

4.1.3 Facet quadrature

In order to integrate over the facets of the reference triangle, we introduce an additional one-dimensional quadrature rule with $q_f + 1$ nodes and weights given by

$$\{\eta_f^{(i)}\}_{i\in\{0:q_f\}} \subset [-1, 1], \quad \{\omega_f^{(i)}\}_{i\in\{0:q_f\}} \subset \mathbb{R}_0^+, \tag{4.15}$$

where we assume that such a rule is of degree τ_f with respect to the Legendre weight, satisfying

$$\sum_{i=0}^{q_f} V(\eta_f^{(i)})\, \omega_f^{(i)} = \int_{-1}^{1} V(\eta_f)\, \mathrm{d}\eta_f, \quad \forall\, V \in \mathbb{P}_{\tau_f}([-1, 1]). \tag{4.16}$$

Transforming integrals on each facet in (4.2) as

$$\int_{\hat{\Gamma}^{(1)}} V(\boldsymbol{\xi})\, \mathrm{d}\hat{s} = \int_{-1}^{1} V(\boldsymbol{\chi}(\eta_f, -1))\, \mathrm{d}\eta_f, \tag{4.17a}$$

$$\int_{\hat{\Gamma}^{(2)}} V(\boldsymbol{\xi})\, \mathrm{d}\hat{s} = \sqrt{2} \int_{-1}^{1} V(\boldsymbol{\chi}(1, \eta_f))\, \mathrm{d}\eta_f, \tag{4.17b}$$

$$\int_{\hat{\Gamma}^{(3)}} V(\boldsymbol{\xi})\, \mathrm{d}\hat{s} = \int_{-1}^{1} V(\boldsymbol{\chi}(-1, \eta_f))\, \mathrm{d}\eta_f, \tag{4.17c}$$

we can then define quadrature rules using $N_{q_f}^{(\zeta)} \coloneqq q_f + 1$ nodes on each facet as

$$\begin{aligned} \boldsymbol{\xi}^{(1,i)} &\coloneqq \boldsymbol{\chi}(\eta_f^{(i-1)}, -1), \quad & \boldsymbol{\xi}^{(2,i)} &\coloneqq \boldsymbol{\chi}(1, \eta_f^{(i-1)}), \quad & \boldsymbol{\xi}^{(3,i)} &\coloneqq \boldsymbol{\chi}(-1, \eta_f^{(i-1)}), \\ \omega^{(1,i)} &\coloneqq \omega_f^{(i-1)}, & \omega^{(2,i)} &\coloneqq \sqrt{2}\omega_f^{(i-1)}, & \omega^{(3,i)} &\coloneqq \omega_f^{(i-1)}, \end{aligned} \tag{4.18}$$

where the corresponding approximation in the form of (3.2) is exact for $V \in \mathbb{P}_{\tau_f}(\hat{\Gamma}^{(\zeta)})$ since each facet is an affine mapping of the interval over which the integral in (4.15) is evaluated.

4.1.4 Summation-by-parts operators

The interpolation in (4.10) can be expressed in reference coodinates for $\boldsymbol{\xi} \in \hat{\Omega} \setminus \{[-1,1]^{\mathrm{T}}\}$ as

$$(\mathcal{I}_q V)(\boldsymbol{\xi}) = \sum_{i=1}^{N_q} V(\boldsymbol{\xi}^{(i)}) \ell^{(i)}(\boldsymbol{\xi}), \tag{4.19}$$

where the nodal basis functions $\{\ell^{(i)}\}_{i \in \{1:N_q\}}$ satisfy $\ell^{(i)}(\boldsymbol{\xi}^{(j)}) = \delta_{ij}$ for the volume quadrature nodes constructed as in (4.12), and are defined in terms of the one-dimensional Lagrange polynomials as

$$\ell^{(\sigma(\boldsymbol{\alpha}))}(\boldsymbol{\chi}(\boldsymbol{\eta})) \coloneqq \ell_1^{(\alpha_1)}(\eta_1) \ell_2^{(\alpha_2)}(\eta_2). \tag{4.20}$$

Differentiating each basis function in (4.20) with respect to ξ_1 and ξ_2 at each volume quadrature node using the chain-rule expressions in (4.7), we then obtain collocation derivative operators $\underline{\underline{D}}^{(1)}, \underline{\underline{D}}^{(2)} \in \mathbb{R}^{N_q \times N_q}$ in the form of (3.7a) with entries given by

$$D^{(1)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} \coloneqq \frac{2}{1 - \eta_2^{(\alpha_2)}} \frac{\mathrm{d}\ell_1^{(\beta_1)}}{\mathrm{d}\eta_1}(\eta_1^{(\alpha_1)}) \delta_{\alpha_2\beta_2}, \tag{4.21a}$$

$$D^{(2)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} \coloneqq \frac{1 + \eta_1^{(\alpha_1)}}{1 - \eta_2^{(\alpha_2)}} \frac{\mathrm{d}\ell_1^{(\beta_1)}}{\mathrm{d}\eta_1}(\eta_1^{(\alpha_1)}) \delta_{\alpha_2\beta_2} + \delta_{\alpha_1\beta_1} \frac{\mathrm{d}\ell_2^{(\beta_2)}}{\mathrm{d}\eta_2}(\eta_2^{(\alpha_2)}). \tag{4.21b}$$

Similarly, we can evaluate (4.20) at the facet quadrature nodes in order to obtain interpolation/extrapolation operators $\underline{\underline{R}}^{(1)}, \underline{\underline{R}}^{(2)}, \underline{\underline{R}}^{(3)} \in \mathbb{R}^{N_{q_f}^{(\zeta)} \times N_q}$ in the form of (3.7b) as

$$R^{(1)}_{i,\sigma(\boldsymbol{\beta})} \coloneqq \ell_1^{(\beta_1)}(\eta_f^{(i-1)}) \ell_2^{(\beta_2)}(-1), \tag{4.22a}$$

$$R^{(2)}_{i,\sigma(\boldsymbol{\beta})} \coloneqq \ell_1^{(\beta_1)}(1) \ell_2^{(\beta_2)}(\eta_f^{(i-1)}), \tag{4.22b}$$

$$R^{(3)}_{i,\sigma(\boldsymbol{\beta})} \coloneqq \ell_1^{(\beta_1)}(-1) \ell_2^{(\beta_2)}(\eta_f^{(i-1)}). \tag{4.22c}$$

Using the above expressions and defining $\underline{\underline{W}}$ and $\underline{\underline{B}}^{(\zeta)}$ as in (3.3) using the quadrature rules introduced in Sections 4.1.2 and 4.1.3, the following theorem provides sufficient conditions under which the proposed operators satisfy the requirements of Definition 2.3.

Theorem 4.1. *Suppose that (4.9) and (4.16) hold for the quadrature rules in (4.8) and (4.15), respectively, such that* $\tau_1^{(0,0)} \geq 2q_1$, $\tau_2^{(0,0)} \geq 2q_2$, *and* $\tau_f \geq 2\max(q_1, q_2)$. *The matrices* $\underline{\underline{D}}^{(1)}$ *and* $\underline{\underline{D}}^{(2)}$ *given in (4.21a) and (4.21b) are then diagonal-norm SBP operators of degree* q *approximating* $\partial/\partial\xi_1$ *and* $\partial/\partial\xi_2$, *respectively, with boundary operators in the form of (2.45).*

Proof. The accuracy conditions for $\underline{\underline{D}}^{(l)}$ in (2.41) and those for $\underline{\underline{R}}^{(\zeta)}$ in (2.46) follow from Lemma 4.1 as well as the fact that the action of such operators can be expressed as

$$\underline{\underline{D}}^{(l)} \begin{bmatrix} V(\boldsymbol{\xi}^{(1)}) \\ \vdots \\ V(\boldsymbol{\xi}^{(N_q)}) \end{bmatrix} = \begin{bmatrix} (\partial \mathcal{I}_q V/\partial \xi_l)(\boldsymbol{\xi}^{(1)}) \\ \vdots \\ (\partial \mathcal{I}_q V/\partial \xi_l)(\boldsymbol{\xi}^{(N_q)}) \end{bmatrix}, \tag{4.23a}$$

$$\underline{\underline{R}}^{(\zeta)} \begin{bmatrix} V(\boldsymbol{\xi}^{(1)}) \\ \vdots \\ V(\boldsymbol{\xi}^{(N_q)}) \end{bmatrix} = \begin{bmatrix} (\mathcal{I}_q V)(\boldsymbol{\xi}^{(\zeta,1)}) \\ \vdots \\ (\mathcal{I}_q V)(\boldsymbol{\xi}^{(\zeta,N_{q_f}^{(\zeta)})}) \end{bmatrix}. \tag{4.23b}$$

Defining $\underline{\underline{E}}^{(l)}$ as in (2.45) and using the polynomial exactness of $\underline{\underline{R}}^{(\zeta)}$ as well as the fact that $\tau_f \geq 2q$ implies that each facet quadrature rule is exact for the corresponding trace space $\mathbb{P}_{2q}(\hat{\Gamma}^{(\zeta)})$ defined as in (2.17), we therefore obtain (2.42). Since $\underline{\underline{W}}$ is SPD due to the fact that the weights in (4.13) are positive when there is no node at $\eta_2 = 1$, as is the case for the quadrature rules in (4.8), all that remains is to show that the SBP property is satisfied.

Beginning with the ξ_1 direction, we define $\underline{\underline{Q}}^{(1)} \coloneqq \underline{\underline{W}}\underline{\underline{D}}^{(1)}$ and note that the denominator in (4.21a) is cancelled by the factor $1 - \eta_2^{(\alpha_2)}$ appearing in (4.13). Taking $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{Q}(q_1, q_2)$ and using the cardinal property of the Lagrange bases as well as the polynomial exactness of the quadrature, we then obtain

$$Q^{(1)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = \underbrace{\int_{-1}^{1} \ell_1^{(\alpha_1)}(\eta_1) \frac{\mathrm{d}\ell_1^{(\beta_1)}}{\mathrm{d}\eta_1}(\eta_1)\,\mathrm{d}\eta_1}_{\tau_1^{(0,0)} \geq 2q_1 - 1} \underbrace{\int_{-1}^{1} \ell_2^{(\alpha_2)}(\eta_2)\ell_2^{(\beta_2)}(\eta_2)\,\mathrm{d}\eta_2}_{\tau_2^{(0,0)} \geq 2q_2} \tag{4.24}$$

and

$$E^{(1)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = \ell_1^{(\alpha_1)}\ell_1^{(\beta_1)}\Big|_{-1}^{1} \underbrace{\int_{-1}^{1} \ell_2^{(\alpha_2)}(\eta_f)\ell_2^{(\beta_2)}(\eta_f)\,\mathrm{d}\eta_f}_{\tau_f \geq 2q_2}, \tag{4.25}$$

where the latter expression follows from substitution of (4.22) and (4.3) into (2.45), and the SBP property results from a straightforward application of integration by parts to obtain

$$Q^{(1)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = E^{(1)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} - Q^{(1)}_{\sigma(\boldsymbol{\beta})\sigma(\boldsymbol{\alpha})}. \tag{4.26}$$

Noting that a similar cancellation of $1-\eta_2^{(\alpha_2)}$ occurs for $\underline{\underline{Q}}^{(2)} \coloneqq \underline{\underline{W}}\,\underline{\underline{D}}^{(2)}$, we obtain

$$\begin{aligned} Q^{(2)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = &\underbrace{\int_{-1}^{1} \frac{1+\eta_1}{2}\ell_1^{(\alpha_1)}(\eta_1)\frac{\mathrm{d}\ell_1^{(\beta_1)}}{\mathrm{d}\eta_1}(\eta_1)\,\mathrm{d}\eta_1}_{\tau_1^{(0,0)} \geq 2q_1}\underbrace{\int_{-1}^{1} \ell_2^{(\alpha_2)}(\eta_2)\ell_2^{(\beta_2)}(\eta_2)\,\mathrm{d}\eta_2}_{\tau_2^{(0,0)} \geq 2q_2} \\ &+ \underbrace{\int_{-1}^{1} \ell_1^{(\alpha_1)}(\eta_1)\ell_1^{(\beta_1)}(\eta_1)\,\mathrm{d}\eta_1}_{\tau_1^{(0,0)} \geq 2q_1}\underbrace{\int_{-1}^{1} \frac{1-\eta_2}{2}\ell_2^{(\alpha_2)}(\eta_2)\frac{\partial \ell_2^{(\beta_2)}}{\partial \eta_2}(\eta_2)\,\mathrm{d}\eta_2}_{\tau_2^{(0,0)} \geq 2q_2} \end{aligned} \tag{4.27}$$

and

$$\begin{aligned} E^{(2)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = &\ell_1^{(\alpha_1)}(1)\ell_1^{(\beta_1)}(1)\underbrace{\int_{-1}^{1} \ell_2^{(\alpha_2)}(\eta_f)\ell_2^{(\beta_2)}(\eta_f)\,\mathrm{d}\eta_f}_{\tau_f \geq 2q_2} \\ &- \ell_2^{(\alpha_2)}(-1)\ell_2^{(\beta_2)}(-1)\underbrace{\int_{-1}^{1} \ell_1^{(\alpha_1)}(\eta_f)\ell_1^{(\beta_1)}(\eta_f)\,\mathrm{d}\eta_f}_{\tau_f \geq 2q_1}. \end{aligned} \tag{4.28}$$

Similarly to (4.26), applying integration by parts and the product rule to (4.27) results in

$$Q^{(2)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = E^{(2)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} - Q^{(2)}_{\sigma(\boldsymbol{\beta})\sigma(\boldsymbol{\alpha})}. \tag{4.29}$$

Since σ maps every multi-index in $\mathcal{Q}(q_1, q_2)$ to a unique scalar index in $\{1 : N_q\}$, and since $\underline{\underline{W}}$ is SPD and therefore invertible under the present assumptions, the fact that (4.26) and (4.29) hold for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{Q}(q_1, q_2)$ implies that the operators in (4.21) can be expressed in the form $\underline{\underline{D}}^{(l)} = \underline{\underline{W}}^{-1}\underline{\underline{Q}}^{(l)}$, where the SBP property $\underline{\underline{Q}}^{(l)} + (\underline{\underline{Q}}^{(l)})^{\mathrm{T}} = \underline{\underline{E}}^{(l)}$ is satisfied for $l = 1$ as well as $l = 2$, thus fulfilling the requirements of Definition 2.3. □

Remark 4.1. The discrete derivative operators in (4.21a) and (4.21b) are of the same form as those proposed in [56, Section 2.2.3]. However, while the quadrature rules employed in their work are constructed as in (4.14) in order to subsume the Jacobian determinant of the coordinate transformation appearing in integrals such as (4.6), such choices do not, in general, result in nodal SBP operators on the reference triangle in the sense of Definition 2.3. This is because the factor of $1-\eta_2^{(\alpha_2)}$ in (4.13), which is subsumed through the use of a Jacobi weight with $(a_2, b_2) = (1, 0)$ in (4.14), is precisely what leads to the cancellation of the denominators in (4.21a) and (4.21b) when $\underline{\underline{D}}^{(1)}$ and $\underline{\underline{D}}^{(2)}$ are left-multiplied by $\underline{\underline{W}}$, enabling the exact evaluation of the integrals in (4.24) and (4.27), respectively. While the nodal operators for $(a_2, b_2) = (1, 0)$ remain exact for polynomials of degree q as a consequence of Lemma 4.1, they do not, in general, satisfy the SBP property. Theorem 4.1 therefore requires the use of quadrature rules based on the Legendre weight for both η_1 and η_2.

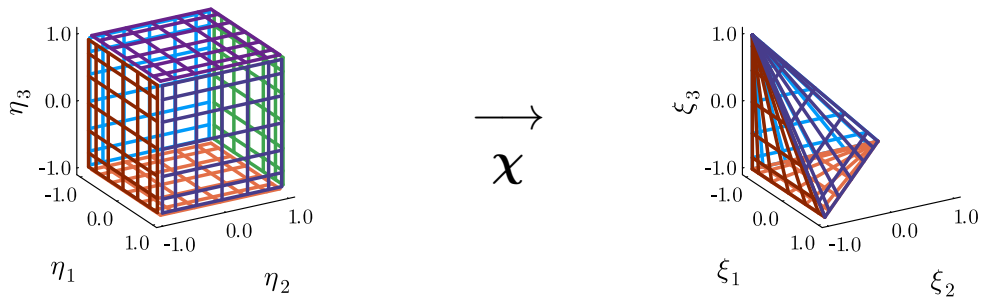


Figure 4.2: Illustration of the mapping $\boldsymbol{\xi} = \boldsymbol{\chi}(\boldsymbol{\eta})$ from the cube to the reference tetrahedron

4.2 Tensor-product SBP operators on the reference tetrahedron

We now extend the methodology presented in Section 4.1 to the three-dimensional case, wherein approximations are constructed on the reference tetrahedron given by

$$\hat{\Omega} \coloneqq \left\{ \boldsymbol{\xi} \in [-1,1]^3 : \xi_1 + \xi_2 + \xi_3 \leq -1 \right\}, \tag{4.30}$$

which has vertices at $[-1,-1,1]^{\mathrm{T}}$, $[1,-1,-1]^{\mathrm{T}}$, $[-1,1,-1]^{\mathrm{T}}$, and $[-1,-1,1]^{\mathrm{T}}$ as well as facets given by

$$\begin{aligned} \hat{\Gamma}^{(1)} &\coloneqq \left\{ \boldsymbol{\xi} \in \hat{\Omega} : \xi_2 = -1 \right\}, \quad \hat{\Gamma}^{(2)} \coloneqq \left\{ \boldsymbol{\xi} \in \hat{\Omega} : \xi_1 + \xi_2 + \xi_3 = -1 \right\}, \\ \hat{\Gamma}^{(3)} &\coloneqq \left\{ \boldsymbol{\xi} \in \hat{\Omega} : \xi_1 = -1 \right\}, \quad \hat{\Gamma}^{(4)} \coloneqq \left\{ \boldsymbol{\xi} \in \hat{\Omega} : \xi_3 = -1 \right\}. \end{aligned} \tag{4.31}$$

The corresponding outward unit normal vectors are then given as

$$\hat{\boldsymbol{n}}^{(1)} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \quad \hat{\boldsymbol{n}}^{(2)} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \quad \hat{\boldsymbol{n}}^{(3)} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \quad \hat{\boldsymbol{n}}^{(4)} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}. \tag{4.32}$$

As described, for example, in [110, Section 3.2], the mapping $\boldsymbol{\chi} : [-1,1]^3 \to \hat{\Omega}$ from the cube to the tetrahedron is constructed from three successive applications of (4.4) in order to obtain

$$\boldsymbol{\chi}(\boldsymbol{\eta}) \coloneqq \begin{bmatrix} \frac{1}{4}(1+\eta_1)(1-\eta_2)(1-\eta_2) - 1 \\ \frac{1}{2}(1+\eta_2)(1-\eta_3) - 1 \\ \eta_3 \end{bmatrix}, \tag{4.33}$$

which is shown in Figure 4.2 and has an inverse given away from the singularities by

$$\boldsymbol{\chi}^{-1}(\boldsymbol{\xi}) = \begin{bmatrix} 2(1+\xi_1)/(-\xi_2-\xi_3) - 1 \\ 2(1+\xi_2)/(1-\xi_3) - 1 \\ \xi_3 \end{bmatrix}, \quad \forall \boldsymbol{\xi} \in \hat{\Omega} \setminus \left\{ \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \right\}. \tag{4.34}$$

Using such a transformation, integrals can be evaluated in collapsed coordinates as

$$\int_{\hat{\Omega}} V(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} = \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} V(\boldsymbol{\chi}(\boldsymbol{\eta}))\frac{(1-\eta_2)(1-\eta_3)^2}{8}\,\mathrm{d}\eta_1\mathrm{d}\eta_2\mathrm{d}\eta_3, \tag{4.35}$$

whereas partial derivatives can be evaluated for $\boldsymbol{\eta} \in [-1,1] \times [-1,1) \times [-1,1)$ as

$$\frac{\partial V}{\partial \xi_1}(\boldsymbol{\chi}(\boldsymbol{\eta})) = \frac{4}{(1-\eta_2)(1-\eta_3)}\frac{\partial}{\partial \eta_1}V(\boldsymbol{\chi}(\boldsymbol{\eta})), \tag{4.36a}$$

$$\frac{\partial V}{\partial \xi_2}(\boldsymbol{\chi}(\boldsymbol{\eta})) = \left(\frac{2(1+\eta_1)}{(1-\eta_2)(1-\eta_3)}\frac{\partial}{\partial \eta_1} + \frac{2}{1-\eta_3}\frac{\partial}{\partial \eta_2}\right)V(\boldsymbol{\chi}(\boldsymbol{\eta})), \tag{4.36b}$$

$$\frac{\partial V}{\partial \xi_3}(\boldsymbol{\chi}(\boldsymbol{\eta})) = \left(\frac{2(1+\eta_1)}{(1-\eta_2)(1-\eta_3)}\frac{\partial}{\partial \eta_1} + \frac{1+\eta_2}{1-\eta_3}\frac{\partial}{\partial \eta_2} + \frac{\partial}{\partial \eta_3}\right)V(\boldsymbol{\chi}(\boldsymbol{\eta})). \tag{4.36c}$$

4.2.1 Nodal sets and interpolants

Similarly to the triangular case, we let q_1, q_2, and q_3 denote the degrees of the approximations with respect to the η_1, η_2, and η_3 coordinates, respectively, with $q \coloneqq \min(q_1, q_2, q_3)$, and define Gaussian quadrature rules with nodes and weights given by

$$\begin{aligned} &\{\eta_1^{(i)}\}_{i\in\{0:q_1\}} \subset [-1,1], \quad \{\eta_2^{(i)}\}_{i\in\{0:q_2\}} \subset [-1,1), \quad \{\eta_3^{(i)}\}_{i\in\{0:q_3\}} \subset [-1,1), \\ &\{\omega_1^{(i)}\}_{i\in\{0:q_1\}} \subset \mathbb{R}^+, \qquad \{\omega_2^{(i)}\}_{i\in\{0:q_2\}} \subset \mathbb{R}^+, \qquad \{\omega_3^{(i)}\}_{i\in\{0:q_3\}} \subset \mathbb{R}^+, \end{aligned} \tag{4.37}$$

satisfying accuracy conditions analogous to (4.9) for $m \in \{1:3\}$. Defining Lagrange bases $\{\ell_m^{(i)}\}_{i\in\{0:q_m\}}$ as in (2.32) and using the mapping in (4.33), we obtain the interpolant

$$(\mathcal{I}_q V)(\boldsymbol{\chi}(\boldsymbol{\eta})) \coloneqq \sum_{\alpha_1=0}^{q_1}\sum_{\alpha_2=0}^{q_2}\sum_{\alpha_3=0}^{q_3} V(\boldsymbol{\chi}(\eta_1^{(\alpha_1)}, \eta_2^{(\alpha_2)}, \eta_3^{(\alpha_3)}))\ell_1^{(\alpha_1)}(\eta_1)\ell_2^{(\alpha_2)}(\eta_2)\ell_3^{(\alpha_3)}(\eta_3), \tag{4.38}$$

whose exactness for polynomials of up to degree q on the reference tetrahedron in (4.30) is characterized with the following lemma.

Lemma 4.2. *The tensor-product interpolation operator in (4.38) is exact for any polynomial $V \in \mathbb{P}_q(\hat{\Omega})$, satisfying $(\mathcal{I}_q V)(\boldsymbol{\xi}) = V(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \hat{\Omega} \setminus \{[-1,1,-1]^{\mathrm{T}}, [-1,-1,1]^{\mathrm{T}}\}$.*

Proof. The proof is similar to that of Lemma 4.1, wherein the monomial $\pi^{(\alpha_1,\alpha_2,\alpha_3)}(\boldsymbol{\xi}) \coloneqq \xi_1^{\alpha_1}\xi_2^{\alpha_2}\xi_3^{\alpha_3}$ with $\boldsymbol{\alpha} \in \mathcal{P}(q)$ is evaluated under the mapping in (4.33) to obtain

$$\pi^{(\alpha_1,\alpha_2,\alpha_3)}(\boldsymbol{\chi}(\boldsymbol{\eta})) = \left(\tfrac{1}{4}(1+\eta_1)(1-\eta_2)(1-\eta_3) - 1\right)^{\alpha_1}\left(\tfrac{1}{2}(1+\eta_2)(1-\eta_3) - 1\right)^{\alpha_2}\eta_3^{\alpha_3}, \tag{4.39}$$

which is of maximum degree α_1 in η_1, $\alpha_1+\alpha_2$ in η_2, and $\alpha_1+\alpha_2+\alpha_3$ in η_3. The accuracy of the interpolation then follows from the fact that $\alpha_1 \le q_1$, $\alpha_1+\alpha_2 \le q_2$, and $\alpha_1+\alpha_2+\alpha_3 \le q_3$

for all $\boldsymbol{\alpha} \in \mathcal{P}(q)$, placing any function $V \circ \boldsymbol{\chi}$ which can be expressed as a linear combination of such monomials within the span of the tensor-product basis used in (4.38). □

4.2.2 Volume quadrature

Defining the multi-index set $\mathcal{Q}(q_1, q_2, q_3) \coloneqq \{0 : q_1\} \times \{0 : q_2\} \times \{0 : q_3\}$, volume quadrature rules in the form of (3.1) can be constructed by ordering the $N_q \coloneqq (q_1+1)(q_2+1)(q_3+1)$ tensor-product volume quadrature nodes using the bijective mapping $\sigma : \mathcal{Q}(q_1, q_2, q_3) \to \{1 : N_q\}$ and applying the mapping in (4.33) to obtain

$$\boldsymbol{\xi}^{(\sigma(\boldsymbol{\alpha}))} \coloneqq \boldsymbol{\chi}(\eta_1^{(\alpha_1)}, \eta_2^{(\alpha_2)}, \eta_3^{(\alpha_3)}). \tag{4.40}$$

Recalling that integrals transform under the mapping in (4.33) as in (4.35), we can define quadrature rules on the nodes in (4.40) for $(a_1, b_1) = (a_2, b_2) = (a_3, b_3) = (0, 0)$ with weights given by

$$\omega^{(\sigma(\boldsymbol{\alpha}))} \coloneqq \frac{(1-\eta_2^{(\alpha_2)})(1-\eta_3^{(\alpha_3)})^2}{8}\omega_1^{(\alpha_1)}\omega_2^{(\alpha_2)}\omega_3^{(\alpha_3)}, \tag{4.41}$$

whereas for $(a_1, b_1) = (a_2, b_2) = (0, 0)$ and $(a_3, b_3) = (1, 0)$ we can subsume one factor of $(1 - \eta_3)$ appearing in (4.35) as

$$\omega^{(\sigma(\boldsymbol{\alpha}))} \coloneqq \frac{(1-\eta_2^{(\alpha_2)})(1-\eta_3^{(\alpha_3)})}{8}\omega_1^{(\alpha_1)}\omega_2^{(\alpha_2)}\omega_3^{(\alpha_3)}. \tag{4.42}$$

By a similar argument to that in Remark 4.1, it can be shown that, unlike the quadrature rules specified above, the choices of $(a_1, b_1) = (0, 0)$, $(a_2, b_2) = (1, 0)$, and $(a_3, b_3) = (2, 0)$ employed in [57, Section 3.1] do not, in general, result in SBP operators on the volume quadrature nodes, and are thus not considered in this work.

4.2.3 Facet quadrature

Defining a collapsed coordinate system (η_{f1}, η_{f2}) on each facet of the reference tetrahedron, we let q_{f1} and q_{f2} denote the polynomial degree in each of such coordinates and introduce the bijective mapping $\sigma_f : \mathcal{Q}(q_{f1}, q_{f2}) \to \{1 : N_{q_f}^{(\zeta)}\}$, where $N_{q_f}^{(\zeta)} \coloneqq (q_{f1} + 1)(q_{f2} + 1)$. Next, we define one-dimensional quadrature rules for integration with respect to each component of the facet coordinate system as

$$\begin{aligned} &\{\eta_{f1}^{(i)}\}_{i\in\{0:q_{f1}\}} \subset [-1, 1], \quad &&\{\eta_{f2}^{(i)}\}_{i\in\{0:q_{f2}\}} \subset [-1, 1], \\ &\{\omega_{f1}^{(i)}\}_{i\in\{0:q_{f1}\}} \subset \mathbb{R}_0^+, &&\{\omega_{f2}^{(i)}\}_{i\in\{0:q_{f2}\}} \subset \mathbb{R}_0^+, \end{aligned} \tag{4.43}$$

which satisfy accuracy conditions given by

$$\sum_{i=0}^{q_{f1}} V(\eta_{f1}^{(i)})\,\omega_{f1}^{(i)} = \int_{-1}^{1} V(\eta)(1-\eta)^{a_{f1}}(1+\eta)^{b_{f1}}\,\mathrm{d}\eta, \quad \forall\, V \in \mathbb{P}_{\tau_{f1}^{(a_{f1},b_{f1})}}([-1,1]), \tag{4.44a}$$

$$\sum_{i=0}^{q_{f2}} V(\eta_{f2}^{(i)})\,\omega_{f2}^{(i)} = \int_{-1}^{1} V(\eta)(1-\eta)^{a_{f2}}(1+\eta)^{b_{f2}}\,\mathrm{d}\eta, \quad \forall\, V \in \mathbb{P}_{\tau_{f2}^{(a_{f2},b_{f2})}}([-1,1]). \tag{4.44b}$$

Integrals on each facet of the reference element then transform as

$$\int_{\hat{\Gamma}^{(1)}} V(\boldsymbol{\xi})\,\mathrm{d}\hat{s} = \int_{-1}^{1}\int_{-1}^{1} V(\boldsymbol{\chi}(\eta_{f1},-1,\eta_{f2}))\frac{1-\eta_{f2}}{2}\,\mathrm{d}\eta_{f1}\mathrm{d}\eta_{f2}, \tag{4.45a}$$

$$\int_{\hat{\Gamma}^{(1)}} V(\boldsymbol{\xi})\,\mathrm{d}\hat{s} = \int_{-1}^{1}\int_{-1}^{1} V(\boldsymbol{\chi}(1,\eta_{f1},\eta_{f2}))\frac{1-\eta_{f2}}{2}\,\mathrm{d}\eta_{f1}\mathrm{d}\eta_{f2}, \tag{4.45b}$$

$$\int_{\hat{\Gamma}^{(3)}} V(\boldsymbol{\xi})\,\mathrm{d}\hat{s} = \int_{-1}^{1}\int_{-1}^{1} V(\boldsymbol{\chi}(-1,\eta_{f1},\eta_{f2}))\frac{\sqrt{3}(1-\eta_{f2})}{2}\,\mathrm{d}\eta_{f1}\mathrm{d}\eta_{f2}, \tag{4.45c}$$

$$\int_{\hat{\Gamma}^{(4)}} V(\boldsymbol{\xi})\,\mathrm{d}\hat{s} = \int_{-1}^{1}\int_{-1}^{1} V(\boldsymbol{\chi}(\eta_{f1},\eta_{f2},-1))\frac{1-\eta_{f2}}{2}\,\mathrm{d}\eta_{f1}\mathrm{d}\eta_{f2}, \tag{4.45d}$$

and hence the one-dimensional quadrature nodes map onto each facet as

$$\begin{aligned}
\boldsymbol{\xi}^{(1,\sigma_f(\boldsymbol{\alpha}))} &\coloneqq \boldsymbol{\chi}(\eta_{f1}^{(\alpha_1)},-1,\eta_{f2}^{(\alpha_2)}), \quad \boldsymbol{\xi}^{(2,\sigma_f(\boldsymbol{\alpha}))} \coloneqq \boldsymbol{\chi}(1,\eta_{f1}^{(\alpha_1)},\eta_{f2}^{(\alpha_2)}),\\
\boldsymbol{\xi}^{(3,\sigma_f(\boldsymbol{\alpha}))} &\coloneqq \boldsymbol{\chi}(-1,\eta_{f1}^{(\alpha_1)},\eta_{f2}^{(\alpha_2)}), \quad \boldsymbol{\xi}^{(4,\sigma_f(\boldsymbol{\alpha}))} \coloneqq \boldsymbol{\chi}(\eta_{f1}^{(\alpha_1)},\eta_{f2}^{(\alpha_2)},-1).
\end{aligned} \tag{4.46}$$

For $(a_{f1},b_{f1}) = (a_{f2},b_{f2}) = (0,0)$, the corresponding weights are given by

$$\begin{aligned}
\omega^{(1,\sigma_f(\boldsymbol{\alpha}))} &\coloneqq \frac{1}{2}\omega_{f1}^{(\alpha_1)}\omega_{f2}^{(\alpha_2)}, \quad \omega^{(2,\sigma_f(\boldsymbol{\alpha}))} \coloneqq \frac{\sqrt{3}}{2}\omega_{f1}^{(\alpha_1)}\omega_{f2}^{(\alpha_2)},\\
\omega^{(3,\sigma_f(\boldsymbol{\alpha}))} &\coloneqq \frac{1}{2}\omega_{f1}^{(\alpha_1)}\omega_{f2}^{(\alpha_2)}, \quad \omega^{(4,\sigma_f(\boldsymbol{\alpha}))} \coloneqq \frac{1}{2}\omega_{f1}^{(\alpha_1)}\omega_{f2}^{(\alpha_2)},
\end{aligned} \tag{4.47}$$

whereas for $(a_{f1},b_{f1}) = (0,0)$ and $(a_{f2},b_{f2}) = (1,0)$, we have

$$\begin{aligned}
\omega^{(1,\sigma_f(\boldsymbol{\alpha}))} &\coloneqq \frac{1-\eta_{f2}^{(\alpha_2)}}{2}\omega_{f1}^{(\alpha_1)}\omega_{f2}^{(\alpha_2)}, \quad \omega^{(2,\sigma_f(\boldsymbol{\alpha}))} \coloneqq \frac{\sqrt{3}(1-\eta_{f2}^{(\alpha_2)})}{2}\omega_{f1}^{(\alpha_1)}\omega_{f2}^{(\alpha_2)},\\
\omega^{(3,\sigma_f(\boldsymbol{\alpha}))} &\coloneqq \frac{1-\eta_{f2}^{(\alpha_2)}}{2}\omega_{f1}^{(\alpha_1)}\omega_{f2}^{(\alpha_2)}, \quad \omega^{(4,\sigma_f(\boldsymbol{\alpha}))} \coloneqq \frac{1-\eta_{f2}^{(\alpha_2)}}{2}\omega_{f1}^{(\alpha_1)}\omega_{f2}^{(\alpha_2)}.
\end{aligned} \tag{4.48}$$

4.2.4 Summation-by-parts operators

Noting that the interpolation in (4.38) can be expressed in the form of (4.19) for $\boldsymbol{\xi} \in \hat{\Omega} \setminus \{[-1,1,-1]^{\mathrm{T}},[-1,-1,1]^{\mathrm{T}}\}$ in terms of the nodal basis functions $\{\ell^{(i)}\}_{i\in\{1:N_q\}}$ given by

$$\ell^{(\sigma(\boldsymbol{\alpha}))}(\boldsymbol{\chi}(\boldsymbol{\eta})) \coloneqq \ell_1^{(\alpha_1)}(\eta_1)\ell_2^{(\alpha_2)}(\eta_2)\ell_3^{(\alpha_3)}(\eta_3), \tag{4.49}$$

the entries of the matrices $\underline{\underline{D}}^{(l)} \in \mathbb{R}^{N_q \times N_q}$ and $\underline{\underline{R}}^{(\zeta)} \in \mathbb{R}^{N_{q_f}^{(\zeta)} \times N_q}$ can be obtained analogously to (4.21) and (4.22) for each coordinate index and facet index, respectively, as

$$D^{(1)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} := \frac{4}{(1-\eta_2^{(\alpha_2)})(1-\eta^{(1,0)}_{q_3,\alpha_3})} \frac{\mathrm{d}\ell_1^{(\beta_1)}}{\mathrm{d}\eta_1}(\eta_1^{(\alpha_1)})\delta_{\alpha_2\beta_2}\delta_{\alpha_3\beta_3}, \tag{4.50a}$$

$$\begin{aligned} D^{(2)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} :=& \frac{2(1+\eta_1^{(\alpha_1)})}{(1-\eta_2^{(\alpha_2)})(1-\eta^{(1,0)}_{q_3,\alpha_3})} \frac{\mathrm{d}\ell_1^{(\beta_1)}}{\mathrm{d}\eta_1}(\eta_1^{(\alpha_1)})\delta_{\alpha_2\beta_2}\delta_{\alpha_3\alpha_3} \\ &+ \frac{2}{1-\eta^{(1,0)}_{q_3,\alpha_3}}\delta_{\alpha_1\beta_1}\frac{\mathrm{d}\ell_2^{(\beta_2)}}{\mathrm{d}\eta_2}(\eta_2^{(\alpha_2)})\delta_{\alpha_3\beta_3}, \end{aligned} \tag{4.50b}$$

$$\begin{aligned} D^{(3)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} :=& \frac{2(1+\eta_1^{(\alpha_1)})}{(1-\eta_2^{(\alpha_2)})(1-\eta^{(1,0)}_{q_3,\alpha_3})} \frac{\mathrm{d}\ell_1^{(\beta_1)}}{\mathrm{d}\eta_1}(\eta_1^{(\alpha_1)})\delta_{\alpha_2\beta_2}\delta_{\alpha_3\alpha_3} \\ &+ \frac{1+\eta_2^{(\alpha_2)}}{1-\eta^{(1,0)}_{q_3,\alpha_3}}\delta_{\alpha_1\beta_1}\frac{\mathrm{d}\ell_2^{(\beta_2)}}{\mathrm{d}\eta_2}(\eta_2^{(\alpha_2)})\delta_{\alpha_3\beta_3} + \delta_{\alpha_1\beta_1}\delta_{\alpha_2\beta_2}\frac{\mathrm{d}\ell_3^{(\beta_3)}}{\mathrm{d}\eta_3}(\eta^{(1,0)}_{q_3,\alpha_3}), \end{aligned} \tag{4.50c}$$

and

$$R^{(1)}_{\sigma_f(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} := \ell_1^{(\beta_1)}(\eta_{f1}^{(\alpha_1)})\ell_2^{(\beta_2)}(-1)\ell_3^{(\beta_3)}(\eta_{f2}^{(\alpha_2)}), \tag{4.51a}$$

$$R^{(2)}_{\sigma_f(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} := \ell_1^{(\beta_1)}(1)\ell_2^{(\beta_2)}(\eta_{f1}^{(\alpha_1)})\ell_3^{(\beta_3)}(\eta_{f2}^{(\alpha_2)}), \tag{4.51b}$$

$$R^{(3)}_{\sigma_f(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} := \ell_1^{(\beta_1)}(-1)\ell_2^{(\beta_2)}(\eta_{f1}^{(\alpha_1)})\ell_3^{(\beta_3)}(\eta_{f2}^{(\alpha_2)}), \tag{4.51c}$$

$$R^{(4)}_{\sigma_f(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} := \ell_1^{(\beta_1)}(\eta_{f1}^{(\alpha_1)})\ell_2^{(\beta_2)}(\eta_{f2}^{(\alpha_2)})\ell_3^{(\beta_3)}(-1). \tag{4.51d}$$

Defining the diagonal matrices $\underline{\underline{W}}$ and $\underline{\underline{B}}^{(\zeta)}$ as in (3.3) using the quadrature weights in Sections 4.2.2 and 4.2.3, we then have the following three-dimensional analogue of Theorem 4.1.

Theorem 4.2. *Assuming that the one-dimensional quadrature rules in (4.37) satisfy (2.34) with* $\tau_1^{(0,0)} \geq 2q_1$, $\tau_2^{(0,0)} \geq 2q_2+1$, *and either* $\tau_3^{(0,0)} \geq 2q_3+1$ *or* $\tau_3^{(1,0)} \geq 2q_3$, *and that the facet quadrature rules in (4.43) satisfy (4.44) with* $\tau_{f1}^{(0,0)} \geq 2\max(q_1,q_2)$ *and either* $\tau_{f2}^{(0,0)} \geq 2\max(q_2,q_3)+1$ *or* $\tau_{f2}^{(1,0)} \geq 2\max(q_2,q_3)$, *the matrices* $\underline{\underline{D}}^{(1)}$, $\underline{\underline{D}}^{(2)}$, *and* $\underline{\underline{D}}^{(3)}$ *given as in (4.50a) to (4.50c) are diagonal-norm SBP operators of degree* q *approximating* $\partial/\partial\xi_1$, $\partial/\partial\xi_2$, *and* $\partial/\partial\xi_3$, *respectively, with boundary operators given as in (2.45).*

Proof. The proof is similar to that of Theorem 4.1, with the accuracy conditions in (2.41) resulting from expressing the action of the operators in (4.50) and (4.51) as in (4.23) and using Lemma 4.2. The conditions in (2.42) then follow from the fact that polynomials of degree q are extrapolated exactly to the boundaries as in (2.46) as a consequence of Lemma 4.2 and the fact that facet quadrature rules satisfying the stated assumptions are exact for all polynomials within the corresponding trace space $\mathbb{P}_{2q}(\hat{\Gamma}^{(\zeta)})$, which can be shown by expressing such polynomials in terms of the (η_{f1}, η_{f2}) coordinate system in a similar manner to (4.11).

To demonstrate that the SBP property holds in the ξ_1 direction, we define $\underline{\underline{Q}}^{(1)} \coloneqq \underline{\underline{W}}\,\underline{\underline{D}}^{(1)}$, which, noting the cancellation of the factor $(1-\eta_2^{(\alpha_2)})(1-\eta_3^{(\alpha_3)})$ in the denominator of (4.50a), results in

$$Q^{(1)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = \underbrace{\int_{-1}^{1} \ell_1^{(\alpha_1)} \frac{\mathrm{d}\ell_1^{(\beta_1)}}{\mathrm{d}\eta_1}\,\mathrm{d}\eta_1}_{\tau_1^{(0,0)} \geq 2q_1-1} \underbrace{\int_{-1}^{1} \ell_2^{(\alpha_2)}\ell_2^{(\beta_2)}\,\mathrm{d}\eta_2}_{\tau_2^{(0,0)} \geq 2q_2} \underbrace{\int_{-1}^{1} \frac{1-\eta_3}{2}\ell_3^{(\alpha_3)}\ell_3^{(\beta_3)}\,\mathrm{d}\eta_3}_{\tau_3^{(0,0)} \geq 2q_3+1 \text{ or } \tau_3^{(1,0)} \geq 2q_3}, \tag{4.52}$$

where the dependence of the Lagrange polynomials on the variable of integration has been suppressed within such one-dimensional integral factors for a clearer presentation, and we have used the cardinal property of the Lagrange basis and the polynomial exactness of the quadrature rules to obtain such an expression. Expressing the boundary operator $\underline{\underline{E}}^{(1)}$ given in (2.45) in terms of the interpolation/extrapolation operators in (4.51) as well as the outward unit normal vectors in (4.32) and using the exactness of the facet quadrature rules, we obtain

$$E^{(1)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = \ell_1^{(\alpha_1)}\ell_1^{(\beta_1)}\Big|_{-1}^{1} \underbrace{\int_{-1}^{1} \ell_2^{(\alpha_2)}\ell_2^{(\beta_2)}\,\mathrm{d}\eta_{f1}}_{\tau_{f1}^{(0,0)} \geq 2q_2} \underbrace{\int_{-1}^{1} \frac{1-\eta_{f2}}{2}\ell_3^{(\alpha_3)}\ell_3^{(\beta_3)}\,\mathrm{d}\eta_{f2}}_{\tau_{f2}^{(0,0)} \geq 2q_3+1 \text{ or } \tau_{f2}^{(1,0)} \geq 2q_3}. \tag{4.53}$$

The SBP property in the ξ_1 direction then follows from applying integration by parts to the first factor in (4.52). Similarly, defining $\underline{\underline{Q}}^{(2)} \coloneqq \underline{\underline{W}}\,\underline{\underline{D}}^{(2)}$ results in

$$\begin{aligned} Q^{(2)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = {} & \underbrace{\int_{-1}^{1} \frac{1+\eta_1}{2}\ell_1^{(\alpha_1)} \frac{\mathrm{d}\ell_1^{(\beta_1)}}{\mathrm{d}\eta_1}\,\mathrm{d}\eta_1}_{\tau_1^{(0,0)} \geq 2q_1} \underbrace{\int_{-1}^{1} \ell_2^{(\alpha_2)}\ell_2^{(\beta_2)}\,\mathrm{d}\eta_2}_{\tau_2 \geq 2q_2} \underbrace{\int_{-1}^{1} \frac{1-\eta_3}{2}\ell_3^{(\alpha_3)}\ell_3^{(\beta_3)}\,\mathrm{d}\eta_3}_{\tau_3^{(0,0)} \geq 2q_3+1 \text{ or } \tau_3^{(1,0)} \geq 2q_3} \\ & + \underbrace{\int_{-1}^{1} \ell_1^{(\alpha_1)}\ell_1^{(\beta_1)}\,\mathrm{d}\eta_1}_{\tau_1^{(0,0)} \geq 2q_1} \underbrace{\int_{-1}^{1} \frac{1-\eta_2}{2}\ell_2^{(\alpha_2)} \frac{\mathrm{d}\ell_2^{(\beta_2)}}{\mathrm{d}\eta_2}\,\mathrm{d}\eta_2}_{\tau_2^{(0,0)} \geq 2q_2} \underbrace{\int_{-1}^{1} \frac{1-\eta_3}{2}\ell_3^{(\alpha_3)}\ell_3^{(\beta_3)}\,\mathrm{d}\eta_3}_{\tau_3^{(0,0)} \geq 2q_3+1 \text{ or } \tau_3^{(1,0)} \geq 2q_3} \end{aligned} \tag{4.54}$$

and

$$\begin{aligned} E^{(2)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = {} & \ell_1^{(\alpha_1)}(1)\ell_1^{(\beta_1)}(1) \underbrace{\int_{-1}^{1} \ell_2^{(\alpha_2)}\ell_2^{(\beta_2)}\,\mathrm{d}\eta_{f1}}_{\tau_{f1}^{(0,0)} \geq 2q_2} \underbrace{\int_{-1}^{1} \frac{1-\eta_{f2}}{2}\ell_3^{(\alpha_3)}\ell_3^{(\beta_3)}\,\mathrm{d}\eta_{f2}}_{\tau_{f2}^{(0,0)} \geq 2q_3+1 \text{ or } \tau_{f2}^{(1,0)} \geq 2q_3} \\ & - \ell_2^{(\alpha_2)}(-1)\ell_2^{(\beta_2)}(-1) \underbrace{\int_{-1}^{1} \ell_1^{(\alpha_1)}\ell_1^{(\beta_1)}\,\mathrm{d}\eta_{f1}}_{\tau_{f1}^{0,0} \geq 2q_1} \underbrace{\int_{-1}^{1} \frac{1-\eta_{f2}}{2}\ell_3^{(\alpha_3)}\ell_3^{(\beta_3)}\,\mathrm{d}\eta_{f2}}_{\tau_{f2}^{(0,0)} \geq 2q_3+1 \text{ or } \tau_{f2}^{(1,0)} \geq 2q_3}, \end{aligned} \tag{4.55}$$

with the SBP property in the ξ_2 direction resulting from the application of integration by parts and the product rule to (4.54). Finally, one can similarly show that the SBP property

in the ξ_3 direction follows from expressing the entries of $\underline{\underline{Q}}^{(3)} \coloneqq \underline{\underline{W}}\,\underline{\underline{D}}^{(3)}$ as

$$\begin{aligned} Q^{(3)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = &\underbrace{\int_{-1}^{1} \frac{1+\eta_1}{2}\ell_1^{(\alpha_1)}\frac{\mathrm{d}\ell_1^{(\beta_1)}}{\mathrm{d}\eta_1}\,\mathrm{d}\eta_1}_{\tau_1^{(0,0)} \geq 2q_1} \underbrace{\int_{-1}^{1} \ell_2^{(\alpha_2)}\ell_2^{(\beta_2)}\,\mathrm{d}\eta_2}_{\tau_2 \geq 2q_2} \underbrace{\int_{-1}^{1} \frac{1-\eta_3}{2}\ell_3^{(\alpha_3)}\ell_3^{(\beta_3)}\,\mathrm{d}\eta_3}_{\tau_3^{(0,0)} \geq 2q_3+1 \text{ or } \tau_3^{(1,0)} \geq 2q_3} \\ &+ \underbrace{\int_{-1}^{1} \ell_1^{(\alpha_1)}\ell_1^{(\beta_1)}\,\mathrm{d}\eta_1}_{\tau_1^{(0,0)} \geq 2q_1} \underbrace{\int_{-1}^{1} \frac{(1-\eta_2)(1+\eta_2)}{4}\ell_2^{(\alpha_2)}\frac{\mathrm{d}\ell_2^{(\beta_2)}}{\mathrm{d}\eta_2}\,\mathrm{d}\eta_2}_{\tau_2^{(0,0)} \geq 2q_2+1} \underbrace{\int_{-1}^{1} \frac{1-\eta_3}{2}\ell_3^{(\alpha_3)}\ell_3^{(\beta_3)}\,\mathrm{d}\eta_3}_{\tau_3^{(0,0)} \geq 2q_3+1 \text{ or } \tau_3^{(1,0)} \geq 2q_3} \\ &+ \underbrace{\int_{-1}^{1} \ell_1^{(\alpha_1)}\ell_1^{(\beta_1)}\,\mathrm{d}\eta_1}_{\tau_1^{(0,0)} \geq 2q_1} \underbrace{\int_{-1}^{1} \frac{1-\eta_2}{2}\ell_2^{(\alpha_2)}\ell_2^{(\beta_2)}\,\mathrm{d}\eta_2}_{\tau_2^{(0,0)} \geq 2q_2+1} \underbrace{\int_{-1}^{1} \frac{(1-\eta_3)^2}{4}\ell_3^{(\alpha_3)}\frac{\mathrm{d}\ell_3^{(\beta_3)}}{\mathrm{d}\eta_3}\,\mathrm{d}\eta_3}_{\tau_3^{(0,0)} \geq 2q_3+1 \text{ or } \tau_3^{(1,0)} \geq 2q_3} \end{aligned} \tag{4.56}$$

and those of the corresponding boundary operator as

$$\begin{aligned} E^{(3)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = &\ \ell_1^{(\alpha_1)}(1)\ell_1^{(\beta_1)}(1) \underbrace{\int_{-1}^{1} \ell_2^{(\alpha_2)}\ell_2^{(\beta_2)}\,\mathrm{d}\eta_{f1}}_{\tau_{f1}^{(0,0)} \geq 2q_2} \underbrace{\int_{-1}^{1} \frac{1-\eta_{f2}}{2}\ell_3^{(\alpha_3)}\ell_3^{(\beta_3)}\,\mathrm{d}\eta_{f2}}_{\tau_{f2}^{(0,0)} \geq 2q_3+1 \text{ or } \tau_{f2}^{(1,0)} \geq 2q_3} \\ &- \ell_3^{(\alpha_3)}(-1)\ell_3^{(\beta_3)}(-1) \underbrace{\int_{-1}^{1} \ell_1^{(\alpha_1)}\ell_1^{(\beta_1)}\,\mathrm{d}\eta_{f1}}_{\tau_{f1}^{(0,0)} \geq 2q_1} \underbrace{\int_{-1}^{1} \frac{1-\eta_{f2}}{2}\ell_2^{(\alpha_2)}\ell_2^{(\beta_2)}\,\mathrm{d}\eta_{f2}}_{\tau_{f2}^{(0,0)} \geq 2q_2+1 \text{ or } \tau_{f2}^{(1,0)} \geq 2q_2}. \end{aligned} \tag{4.57}$$

Since, for $l = 1$, $l = 2$, and $l = 3$, the entries of the matrices $\underline{\underline{Q}}^{(l)} \coloneqq \underline{\underline{W}}\,\underline{\underline{D}}^{(l)}$ satisfy

$$Q^{(l)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} = E^{(l)}_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} - Q^{(l)}_{\sigma(\boldsymbol{\beta})\sigma(\boldsymbol{\alpha})} \tag{4.58}$$

and $\underline{\underline{W}}$ is SPD as the one-dimensional volume quadrature weights are positive and there are no nodes at $\eta_2 = 1$ nor at $\eta_3 = 1$, the SBP property is then satisfied in the ξ_1, ξ_2, and ξ_3 directions, and, hence, the operators in (4.50) fulfil the requirements of Definition 2.3. □

4.3 Chapter summary

This chapter describes a methodology for the construction of spectral-element operators of arbitrary order on the reference triangle and tetrahedron which are sparse, satisfy the SBP property, and possess a tensor-product structure supporting the use of sum-factorization algorithms for matrix-free operator evaluation. Employing the polynomial exactness of tensor-product interpolation in collapsed coordinates and careful choices of Gaussian quadrature rules for volume and facet integration, we establish conditions under which the proposed operators satisfy the conditions of Definition 2.3 and are therefore suitable for use within

spatial discretizations based on SBP operators for simplicial elements. The following two chapters will discuss the formulation, analysis, and implementation of such discretizations, in which we will exploit the operators' tensor-product structure and sparsity as well as the summation-by-parts property to obtain efficient algorithms which are provably stable for linear and nonlinear problems on curved triangular and tetrahedral unstructured grids.

Chapter 5

Energy-stable tensor-product discontinuous spectral-element methods on curved triangles and tetrahedra

In this chapter, which is based on the content of Papers II and III as well as portions of Paper IV, we will employ the SBP operators introduced in Chapter 4 within a skew-symmetric formulation in order to construct efficient DSEMs curved triangular and tetrahedral elements which are discretely conservative, free-stream preserving, as well as energy stable for the linear advection equation. The fundamental aspects of such discretizations are introduced in Section 5.1, followed by an analysis in Section 5.2 based on the general framework introduced in Chapter 3. We describe the efficient implementation of the proposed schemes in Section 5.3.

5.1 Energy-stable discontinuous spectral-element formulation

This section describes several important numerical techniques which are employed in synergy to construct the proposed schemes, including the approximation of the metric terms in conservative curl form, the use of tensor-product polynomial expansions in collapsed coordinates, the discretization in skew-symmetric form using tensor-product SBP operators on the triangle or tetrahedron, and the weight-adjusted approximation of the inverse mass matrix.

5.1.1 Approximation of the metric terms in conservative curl form

Considering a Lagrange basis $\{\ell_{p_g}^{(i)}\}_{i\in\{1:N_{p_g}\}}$ for $\mathbb{P}_{p_g}(\hat{\Omega})$ satisfying $\ell_{p_g}^{(i)}(\boldsymbol{\xi}_{p_g}^{(j)}) = \delta_{ij}$ on the nodes $\{\boldsymbol{\xi}_{p_g}^{(i)}\}_{i\in\{1:N_{p_g}\}} \subset \hat{\Omega}$, we can construct a polynomial mapping $\boldsymbol{X}^{(\kappa)} \in [\mathbb{P}_{p_g}(\hat{\Omega})]^d$ from the reference triangle or tetrahedron to each physical element as

$$\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}) \coloneqq \sum_{i=1}^{N_{p_g}} \boldsymbol{x}_{p_g}^{(\kappa,i)} \ell_{p_g}^{(i)}(\boldsymbol{\xi}), \tag{5.1}$$

where $\{\boldsymbol{x}_{p_g}^{(\kappa,i)}\}_{i\in\{1:N_{p_g}\}}$ are the prescribed physical positions of the mapping nodes. To ensure a watertight mesh, we assume that the mapping nodes contain a subset of nodes on each facet $\hat{\Gamma}^{(\zeta)} \subset \partial\hat{\Omega}$ which are unisolvent for the corresponding trace space $\mathbb{P}_{p_g}(\hat{\Gamma}^{(\zeta)})$, thus resulting in continuity at element interfaces. Considering a mapping using polynomials of degree p_g as in (5.1), the metric terms are polynomials of degree $p_g - 1$ in two dimensions and degree $2p_g - 2$ in three dimensions. Since operations such as differentiation using SBP operators are exact for polynomials of at most degree q, we cannot expect that a discrete analogue of the metric identities in (2.12) will hold unless $p_g \leq q + 1$ in two dimensions or $p_g \leq \lfloor q/2 \rfloor + 1$ in three dimensions. To circumvent this requirement for a subparametric mapping in three-dimensional case, we use an adaptation by Chan and Wilcox [133, Sect. 5] of Kopriva's approximation of the metric terms in *conservative curl form* [134, Eq. (36)], which is itself based on techniques introduced by Thomas and Lombard [135]. To obtain such an approximation, we introduce a Lagrange basis $\{\ell_{q+1}^{(i)}\}_{i\in\{1:N_{q+1}\}}$ for $\mathbb{P}_{q+1}(\hat{\Omega})$ associated with a set of nodes $\{\boldsymbol{\xi}_{q+1}^{(i)}\}_{i\in\{1:N_{q+1}\}} \subset \hat{\Omega}$ and construct the polynomial interpolants

$$\boldsymbol{r}_{q+1}^{(\kappa,1)}(\boldsymbol{\xi}) \coloneqq \sum_{i=1}^{N_{q+1}} X_3^{(\kappa)}(\boldsymbol{\xi}_{q+1}^{(i)}) \nabla_{\boldsymbol{\xi}} X_2^{(\kappa)}(\boldsymbol{\xi}_{q+1}^{(i)}) \ell_{q+1}^{(i)}(\boldsymbol{\xi}), \tag{5.2a}$$

$$\boldsymbol{r}_{q+1}^{(\kappa,2)}(\boldsymbol{\xi}) \coloneqq \sum_{i=1}^{N_{q+1}} X_3^{(\kappa)}(\boldsymbol{\xi}_{q+1}^{(i)}) \nabla_{\boldsymbol{\xi}} X_1^{(\kappa)}(\boldsymbol{\xi}_{q+1}^{(i)}) \ell_{q+1}^{(i)}(\boldsymbol{\xi}), \tag{5.2b}$$

$$\boldsymbol{r}_{q+1}^{(\kappa,3)}(\boldsymbol{\xi}) \coloneqq \sum_{i=1}^{N_{q+1}} X_1^{(\kappa)}(\boldsymbol{\xi}_{q+1}^{(i)}) \nabla_{\boldsymbol{\xi}} X_2^{(\kappa)}(\boldsymbol{\xi}_{q+1}^{(i)}) \ell_{q+1}^{(i)}(\boldsymbol{\xi}). \tag{5.2c}$$

The functions $\boldsymbol{r}_{q+1}^{(\kappa,m)} \in [\mathbb{P}_{q+1}(\hat{\Omega})]^d$ are used to define the matrix of *approximate* metric terms

$$\boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi}) \coloneqq \left[-\nabla_{\boldsymbol{\xi}} \times \boldsymbol{r}_{q+1}^{(\kappa,1)}(\boldsymbol{\xi}),\ \nabla_{\boldsymbol{\xi}} \times \boldsymbol{r}_{q+1}^{(\kappa,2)}(\boldsymbol{\xi}),\ \nabla_{\boldsymbol{\xi}} \times \boldsymbol{r}_{q+1}^{(\kappa,3)}(\boldsymbol{\xi})\right]^{\mathrm{T}}, \tag{5.3}$$

which has entries of degree q and satisfies (2.12) by construction. Since using (5.3) with the exact normal vector violates (2.14), we then use the matrix of approximate metric terms in (5.3) to compute the normal vector and surface element *approximately* as

$$\boldsymbol{n}^{(\kappa,\zeta)}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi})) \coloneqq \frac{\boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi})^{\mathrm{T}} \hat{\boldsymbol{n}}^{(\zeta)}}{\|\boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi})^{\mathrm{T}} \hat{\boldsymbol{n}}^{(\zeta)}\|}, \quad J^{(\kappa)}(\boldsymbol{\xi}) \coloneqq \|\boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi})^{\mathrm{T}} \hat{\boldsymbol{n}}^{(\zeta)}\|. \tag{5.4}$$

If the analytically defined mesh is watertight and the nodes used for the interpolants in (5.2) define a continuous approximation space, the approximate normals computed as in (5.4) then remain equal and opposite at element interfaces (see, for example, [133, Theorem 5]).

Remark 5.1. We will demonstrate numerically in Chapter 7 that computing the outward unit normal vector as in (5.4) using the design-order metric approximation in (5.3) results in

optimal rates of convergence for smooth periodic problems. However, further investigation is required regarding the effectiveness of such an approach in the context of the imposition of wall boundary conditions, which have been shown in recent work by Navah and Nadarajah [136] as well as Craig Penner and Zingg [137] to be sensitive to the method used to compute the wall normal.

5.1.2 Nodal and modal tensor-product expansions

In order to discretize systems of conservation laws in the form of (2.13) on the reference element, we can approximate the solution and flux components in collapsed coordinates using the tensor-product expansions in (4.10) and (4.38) as

$$\underline{U}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\chi}(\boldsymbol{\eta})),t) \approx \sum_{\alpha_1=0}^{q_1} \cdots \sum_{\alpha_d=0}^{q_d} \underline{u}_{\sigma(\boldsymbol{\alpha})}^{(h,\kappa)}(t)\ell_1^{(\alpha_1)}(\eta_1)\cdots\ell_d^{(\alpha_d)}(\eta_d), \tag{5.5a}$$

$$\underline{F}_m(U(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\chi}(\boldsymbol{\eta})),t)) \approx \sum_{\alpha_1=0}^{q_1} \cdots \sum_{\alpha_d=0}^{q_d} \underline{F}_m(\underline{u}_{\sigma(\boldsymbol{\alpha})}^{(h,\kappa)}(t))\ell_1^{(\alpha_1)}(\eta_1)\cdots\ell_d^{(\alpha_d)}(\eta_d). \tag{5.5b}$$

In the above, the vector $\underline{u}_{\sigma(\boldsymbol{\alpha})}^{(h,\kappa)}(t) \in \mathbb{R}^{N_c}$ contains the solution variables evaluated as in (3.19) in terms of the nodal solution vector $\underline{u}^{(h,\kappa,e)}(t) \in \mathbb{R}^{N_q}$ at each of the tensor-product volume quadrature nodes, which are given generically (i.e. for either triangles or tetrahedra) by

$$\boldsymbol{\xi}^{(\sigma(\boldsymbol{\alpha}))} \coloneqq \boldsymbol{\chi}(\eta_1^{(\alpha_d)},\ldots,\eta_d^{(\alpha_d)}). \tag{5.6}$$

Although it is often advantageous from an efficiency standpoint to collocate the solution degrees of freedom at the volume quadrature points by directly evolving $\underline{u}^{(h,\kappa,e)}(t)$ as in (3.42), such an approach presents two disadvantages in the present context, which we list below.

- The number of tensor-product quadrature nodes N_q required for a collapsed-coordinate formulation using SBP operators of degree q can be much larger than the cardinality of a total-degree polynomial basis for the space $\mathbb{P}_q(\hat{\Omega})$, thus requiring more degrees of freedom than would otherwise be needed to achieve a given order of accuracy.
- The collapsed coordinate system and the resulting rational approximation space in reference coordinates introduces a clustering of resolution near the singularity, and therefore limits the maximum stable time step size for explicit schemes.[1]

We therefore propose the use of a modal approach in which a basis $\{\phi^{(i)}\}_{i\in\{1:N_p\}}$ for $\mathbb{P}_p(\hat{\Omega})$ with $p \leq q$ is used to represent each solution variable on the reference element in terms of the

[1] Quoting Dubiner [55] concerning the time step restriction for tensor-product approximations in collapsed coordinates, "[resolution] is a good thing, but this is a case of too much of a good thing."

modal expansion coefficients $\underline{\tilde{u}}^{(h,\kappa,e)}(t) \in \mathbb{R}^{N_p}$ as in (2.20). Such an expansion is evaluated at each node using the generalized Vandermonde matrix in (3.4) to obtain the nodal expansion coefficients $\underline{u}^{(h,\kappa)}(t) = \underline{\underline{V}}\,\underline{\tilde{u}}^{(h,\kappa,e)}(t)$ used to approximate the solution and flux components as in (5.5a). Since any polynomial $V \in \mathbb{P}_p(\hat{\Omega})$ with $p \leq q$ admits a unique expansion in terms of the tensor-product Lagrange basis functions in (4.20) and (4.49), the mapping in (3.17) is injective and hence $\underline{\underline{V}}$ is of rank N_p (i.e. full column rank) as required for Assumption 3.1, regardless of the chosen basis for $\mathbb{P}_p(\hat{\Omega})$. We must, however, carefully choose such a basis to ensure that the cost of applying $\underline{\underline{V}}$ is minimized and the tensor-product structure of the discretization is preserved.

5.1.3 Proriol–Koornwinder–Dubiner basis functions

In order to construct polynomial bases on the triangle and tetrahedron for which operations such as (3.17) are amenable to sum factorization, we follow [110, Section 3.2] and define

$$\psi_1^{(\alpha_1)}(\eta_1) \coloneqq \sqrt{2} P_{\alpha_1}^{(0,0)}(\eta_1) \tag{5.7a}$$

$$\psi_2^{(\alpha_1,\alpha_2)}(\eta_2) \coloneqq (1-\eta_2)^{\alpha_1} P_{\alpha_2}^{(2\alpha_1+1,0)}(\eta_2), \tag{5.7b}$$

$$\psi_3^{(\alpha_1,\alpha_2,\alpha_3)}(\eta_3) \coloneqq 2(1-\eta_3)^{\alpha_1+\alpha_2} P_{\alpha_3}^{(2\alpha_1+2\alpha_2+2,0)}(\eta_3), \tag{5.7c}$$

in terms of the normalized Jacobi polynomials, which are defined in Section 2.4. The Proriol–Koornwinder–Dubiner (PKD) polynomials [55, 138, 139] are then given in collapsed coordinates on the reference triangle as

$$\phi^{(\pi(\boldsymbol{\alpha}))}(\boldsymbol{\chi}(\boldsymbol{\eta})) \coloneqq \psi_1^{(\alpha_1)}(\eta_1)\psi_2^{(\alpha_1,\alpha_2)}(\eta_2), \tag{5.8}$$

and on the reference tetrahedron as

$$\phi^{(\pi(\boldsymbol{\alpha}))}(\boldsymbol{\chi}(\boldsymbol{\eta})) \coloneqq \psi_1^{(\alpha_1)}(\eta_1)\psi_2^{(\alpha_1,\alpha_2)}(\eta_2)\psi_3^{(\alpha_1,\alpha_2,\alpha_3)}(\eta_3), \tag{5.9}$$

where we order the multi-indices $\boldsymbol{\alpha} \in \mathcal{P}(p)$ using the bijection $\pi : \mathcal{P}(p) \to \{1 : N_p\}$. Although the generalized Vandermonde matrix $\underline{\underline{V}}$ resulting from the evaluation of such basis functions at the nodes in (5.6) cannot be expressed as a standard Kronecker product, the "warped" tensor-product structure of the PKD bases nevertheless allows for such an operator (as well as its transpose) to be applied through sum factorization, as described, for example, in [110, Sections 4.1.6.1 and 4.1.6.2]. Moreover, such bases are orthonormal with respect to the L^2 inner product on the reference element, satisfying

$$\int_{\hat{\Omega}} \phi^{(i)}(\boldsymbol{\xi})\phi^{(j)}(\boldsymbol{\xi}) \,\mathrm{d}\boldsymbol{\xi} = \delta_{ij}, \tag{5.10}$$

resulting in the reference mass matrix $\underline{\underline{M}} \coloneqq \underline{\underline{V}}^{\mathrm{T}}\underline{\underline{W}}\underline{\underline{V}}$ being the identity matrix if the quadrature rule in (2.44) is of degree $\tau \geq 2p$. As will be discussed in Section 5.3, the proposed algorithms exploit both the tensor-product structure and the orthonormality of the basis.

5.1.4 Skew-symmetric discretization using summation-by-parts operators

Using the nodal basis functions defined in (4.20) and (4.49) on the reference triangle and tetrahedron, respectively, the right-hand side of (5.5b) can be expressed in terms of $\boldsymbol{\xi} = \boldsymbol{\chi}(\boldsymbol{\eta})$ as

$$\underline{F}_m^{(h,\kappa)}(\boldsymbol{\xi}, t) \coloneqq \sum_{i=1}^{N_q} \underline{F}_m(\underline{u}_i^{(h,\kappa)}(t))\ell^{(i)}(\boldsymbol{\xi}). \tag{5.11}$$

Inserting such an approximation into (2.19) and applying integration by parts, the product rule, and the metric identities in (2.12) to half of the volume term on the right-hand side, we obtain a skew-symmetric variational formulation, which is given by

$$\begin{aligned}
&\int_{\hat{\Omega}} V(\boldsymbol{\xi}) J^{(\kappa)}(\boldsymbol{\xi}) \frac{\partial \underline{U}^{(h,\kappa)}(\boldsymbol{\xi}, t)}{\partial t} \,\mathrm{d}\boldsymbol{\xi} = \\
&\frac{1}{2}\sum_{l=1}^{d} \int_{\hat{\Omega}} \left(\frac{\partial V(\boldsymbol{\xi})}{\partial \xi_l} \sum_{m=1}^{d} G_{lm}^{(\kappa)}(\boldsymbol{\xi}) \underline{F}_m^{(h,\kappa)}(\boldsymbol{\xi}, t) - V(\boldsymbol{\xi}) \sum_{m=1}^{d} G_{lm}^{(\kappa)}(\boldsymbol{\xi}) \frac{\partial \underline{F}_m^{(h,\kappa)}(\boldsymbol{\xi}, t)}{\partial \xi_l} \right) \mathrm{d}\boldsymbol{\xi} \\
&- \sum_{\zeta=1}^{N_f} \int_{\hat{\Gamma}^{(\zeta)}} V(\boldsymbol{\xi}) J^{(\kappa,\zeta)}(\boldsymbol{\xi}) \left(\underline{F}^{(*,\kappa,\zeta)}(\boldsymbol{\xi}, t) - \frac{1}{2} \sum_{m=1}^{d} n_m^{(\kappa,\zeta)}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi})) \underline{F}_m^{(h,\kappa)}(\boldsymbol{\xi}, t) \right) \mathrm{d}\hat{s}.
\end{aligned} \tag{5.12}$$

Using the tensor-product quadrature rules introduced in Chapter 4 to approximate the integrals in the above formulation, a nodal approximation within a rational function space is then obtained by expanding the solution and test function in terms of the tensor-product Lagrange basis $\{\ell^{(i)}\}_{i\in\{1:N_q\}}$, while a modal approximation within a total-degree polynomial approximation space is recovered by expanding the solution and test function in terms of the PKD basis $\{\phi^{(i)}\}_{i\in\{1:N_p\}}$. Such formulations are then given, respectively, by

$$\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)} \frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = \underline{r}^{(h,\kappa,e)}(t), \tag{5.13a}$$

$$\underline{\underline{V}}^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{\underline{V}} \frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = \underline{\underline{V}}^{\mathrm{T}}\underline{r}^{(h,\kappa,e)}(t), \tag{5.13b}$$

where we define

$$\begin{aligned}
\underline{r}^{(h,\kappa,e)}(t) \coloneqq &\frac{1}{2}\sum_{l=1}^{d} \left(\left(\underline{\underline{D}}^{(l)}\right)^{\mathrm{T}} \underline{\underline{W}} \sum_{m=1}^{d} \underline{\underline{G}}^{(\kappa,l,m)} \underline{f}^{(\kappa,m,e)}(t) - \sum_{m=1}^{d} \underline{\underline{G}}^{(\kappa,l,m)} \underline{\underline{W}}\,\underline{\underline{D}}^{(l)} \underline{f}^{(\kappa,m,e)}(t) \right) \\
&- \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{J}}^{(\kappa,\zeta)} \left(\underline{f}^{(*,\kappa,\zeta,e)}(t) - \frac{1}{2} \sum_{m=1}^{d} \underline{\underline{N}}^{(\kappa,\zeta,m)} \underline{\underline{R}}^{(\zeta)} \underline{f}^{(\kappa,m,e)}(t) \right)
\end{aligned} \tag{5.14}$$

in terms of the SBP operators introduced in Chapter 4 and recall the definitions of $\underline{f}^{(\kappa,m,e)}(t) \in \mathbb{R}^{N_q}$ and $\underline{f}^{(*,\kappa,\zeta,e)}(t) \in \mathbb{R}^{N_f^{(\zeta)}}$ in (3.20) and (3.21), respectively.

5.1.5 Summation-by-parts operators on the physical element

To provide an alternative perspective which is useful for our analysis, we can group together the terms which left-multiply the vector $\underline{f}^{(\kappa,m,e)}(t)$ in (5.14) in order to obtain

$$\underline{r}^{(h,\kappa,e)}(t) = \sum_{m=1}^{d} \left(\underline{\underline{Q}}^{(\kappa,m)}\right)^{\mathrm{T}} \underline{f}^{(\kappa,m,e)}(t) - \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{J}}^{(\kappa,\zeta)} \underline{f}^{(*,\kappa,\zeta,e)}(t), \tag{5.15}$$

which is of a similar form to (3.24), but involves the discretization of the volume terms using the operator

$$\underline{\underline{Q}}^{(\kappa,m)} \coloneqq \frac{1}{2} \sum_{l=1}^{d} \left(\underline{\underline{G}}^{(\kappa,l,m)} \underline{\underline{W}}\, \underline{\underline{D}}^{(l)} - \left(\underline{\underline{D}}^{(l)}\right)^{\mathrm{T}} \underline{\underline{W}}\, \underline{\underline{G}}^{(\kappa,l,m)} \right) + \frac{1}{2} \underline{\underline{E}}^{(\kappa,m)}, \tag{5.16}$$

where we define

$$\underline{\underline{E}}^{(\kappa,m)} \coloneqq \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{J}}^{(\kappa,\zeta)} \underline{\underline{N}}^{(\kappa,\zeta,m)} \underline{\underline{R}}^{(\zeta)}. \tag{5.17}$$

Although obtained from the variational formulation in (5.12), the matrices in (5.16) and (5.17) are of the same form as those constructed in [97, Section 5]. Since the proposed tensor-product SBP operators on the reference triangle or tetrahedron satisfy the assumptions of [97, Theorem 9], it then follows that the differentiation matrix on the physical element given by

$$\underline{\underline{D}}^{(\kappa,m)} \coloneqq \left(\underline{\underline{W}}\, \underline{\underline{J}}^{(\kappa)}\right)^{-1} \underline{\underline{Q}}^{(\kappa,m)} \tag{5.18}$$

provides an $\mathcal{O}(h^q)$ approximation to the partial derivative $\partial/\partial x_m$ for sufficiently regular functions and mesh sequences. Moreover, it is easy to show that the SBP property is satisfied in physical space as

$$\underline{\underline{Q}}^{(\kappa,m)} + \left(\underline{\underline{Q}}^{(\kappa,m)}\right)^{\mathrm{T}} = \underline{\underline{E}}^{(\kappa,m)}, \quad \forall\, m \in \{1:d\}. \tag{5.19}$$

Applying the SBP property on the reference element to (5.16) using (3.6), we then obtain

$$\begin{aligned} \underline{\underline{Q}}^{(\kappa,m)} = {} & \frac{1}{2} \sum_{l=1}^{d} \underline{\underline{W}} \left(\underline{\underline{G}}^{(\kappa,l,m)} \underline{\underline{D}}^{(l)} + \underline{\underline{D}}^{(l)} \underline{\underline{G}}^{(\kappa,l,m)} \right) \\ & + \frac{1}{2} \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \left(\underline{\underline{N}}^{(\kappa,\zeta,m)} \underline{\underline{R}}^{(\zeta)} - \sum_{l=1}^{d} \hat{n}_l^{(\zeta)} \underline{\underline{R}}^{(\zeta)} \underline{\underline{G}}^{(\kappa,l,m)} \right), \end{aligned} \tag{5.20}$$

where the first line corresponds to the more familiar splitting of the derivative operator (see, for example, Kopriva and Gassner [74] or Nordström [72]) as

$$J^{(\kappa)}(\boldsymbol{\xi})\frac{\partial V}{\partial x_m}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi})) = \frac{1}{2}\sum_{l=1}^{d}\left(G_{lm}^{(\kappa)}(\boldsymbol{\xi})\frac{\partial V(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}))}{\partial \xi_l} + \frac{\partial}{\partial \xi_l}\left(G_{lm}^{(\kappa)}(\boldsymbol{\xi})V(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}))\right)\right), \tag{5.21}$$

which holds at the continuous level for smooth functions when $\boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi})$ contains the exact metric terms, and the second line of (5.20) can be interpreted as a correction term which vanishes for diagonal-E SBP operators.

5.1.6 Weight-adjusted approximation of the inverse mass matrix

Even when using orthonormal bases on the reference element, the mass matrix appearing on the left-hand side of (5.13b) is dense when the mapping from the reference element to the physical element is not affine, and its inverse lacks a tensor-product structure amenable to sum factorization. Obtaining the time derivative for such a scheme in the context of explicit temporal integration thus requires either the storage and application of a non-tensorial factorization or inverse, or, otherwise, the solution of a dense N_p by N_p linear system, for each curved element in the mesh. To obtain a fully explicit formulation for the time derivative, we use a *weight-adjusted* approximation given by

$$\left(\underline{\underline{V}}^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{\underline{V}}\right)^{-1} \approx \underline{\underline{M}}^{-1}\underline{\underline{V}}^{\mathrm{T}}\underline{\underline{W}}\left(\underline{\underline{J}}^{(\kappa)}\right)^{-1}\underline{\underline{V}}\,\underline{\underline{M}}^{-1} =: \left(\tilde{\underline{\underline{M}}}^{(\kappa)}\right)^{-1}. \tag{5.22}$$

The above approximation was initially proposed by Chan *et al.* [101] for the purpose of reducing storage requirements for curved elements, and an *a priori* analysis reveals that the corresponding weight-adjusted projection differs from the standard L^2 projection by $\mathcal{O}(h^{p+2})$ for sufficiently regular functions and mesh sequences, thereby introducing an error which is at least one order higher than that of the discretization itself and is typically negligible in practice [133, Sections 4.3 and 6.1]. Besides reducing storage, however, the weight-adjusted approximation also preserves the tensor-product operator structure which would otherwise be lost by taking the inverse of the mass matrix. The time derivative for the resulting scheme can then be obtained explicitly as

$$\frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = \left(\tilde{\underline{\underline{M}}}^{(\kappa)}\right)^{-1}\underline{\underline{V}}^{\mathrm{T}}\underline{r}^{(h,\kappa,e)}(t), \tag{5.23}$$

where we can exploit sum factorization in the application of the operators $\underline{\underline{V}}$ and $\underline{\underline{V}}^{\mathrm{T}}$ in (5.22). While the formulation in (5.23) is not, in general, discretely conservative with respect to the quadrature rule defined by the diagonal entries of $\underline{\underline{W}}$ as in (2.44), we restore conservation using

a technique proposed by Chan and Wilcox [133, Lemma 2]. In the context of a mapping in the form of (5.1), such a modification involves approximating the determinant of the mapping Jacobian, which is of degree $2p_g - 2$ in two dimensions and $3p_g - 3$ in three dimensions, by an interpolant of degree p_g given in terms of the nodal basis in (5.1) as

$$J^{(\kappa)}(\boldsymbol{\xi}) \coloneqq \sum_{i=1}^{N_{p_g}} \det(\boldsymbol{J}^{(\kappa)}(\boldsymbol{\xi}_{p_g}^{(i)}))\ell_{p_g}^{(i)}(\boldsymbol{\xi}), \tag{5.24}$$

and using such an approximation to define $\underline{\underline{J}}^{(\kappa)}$ in (5.22), noting that such a modification does not affect the (approximate) metric terms $\boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi})$ used to compute $\underline{r}^{(h,\kappa,e)}(t)$ in (5.23).

5.2 Analysis

We now begin our theoretical analysis of the schemes constructed in the previous section, which will make use of several techniques introduced in Chapter 3 within the context of standard DG and FR methods. Although the focus of this chapter is on discretizations using the operators constructed in Chapter 4, the analysis applies to nodal and modal DSEMs using any set of multidimensional SBP operators satisfying the following assumption.

Assumption 5.1. *The matrices $\underline{\underline{D}}^{(l)}$ are multidimensional diagonal-norm SBP operators of degree $q \geq p$ on the reference element, satisfying the conditions of Definition 2.3 with boundary operators given as in (2.45) in terms of diagonal positive-semidefinite matrices $\underline{\underline{B}}^{(\zeta)}$ and interpolation/extrapolation operators $\underline{\underline{R}}^{(\zeta)}$ satisfying the accuracy conditions in (2.46).*

We will also invoke Assumption 3.3 in our analysis, which could, at first sight, appear to be somewhat restrictive in the tetrahedral case due to the asymmetry of the tensor-product facet quadrature rules described in Section 4.2.3. However, through the use of simple preprocessing algorithms (see, for example, Sherwin and Karniadakis [57, Section 2.3] or Warburton *et al.* [140]), a suitable orientation can be defined for the local coordinate system on each element so as to obtain matching facet quadrature node positions in physical space at a cost scaling linearly with the number of elements in the mesh. Using such techniques, the proposed operators can be used with standard mesh generation tools without the need for nonconforming interface procedures. Otherwise, mortar-based techniques similar to those introduced in [130] could be employed in order to enable the use of symmetric nodal sets at element interfaces, thereby eliminating the need for a preprocessing step, although such an approach would increase the cost of applying the interpolation/extrapolation operators.

5.2.1 Discrete metric identities and free-stream preservation

Although polynomials are not necessarily differentiated exactly in physical space by the split-form derivative operators in (5.18) when the mapping $\boldsymbol{X}^{(\kappa)}$ is not affine, we will require that constants lie within the nullspace of the physical SBP operator, a property which can be ensured with the following lemma.

Lemma 5.1. *Let Assumption 5.1 hold and assume that the metric terms, whether computed exactly or approximately as in (5.3), satisfy $G_{lm}^{(\kappa)} \in \mathbb{P}_q(\hat{\Omega})$ as well as the metric identities in (2.12) and that the normals are computed using such metric terms as in (5.4). Then, the operators in (5.16) satisfy the following discrete metric identities:*

$$\underline{\underline{Q}}^{(\kappa,m)}\underline{1}^{(N_q)} = \underline{0}^{(N_q)}, \quad \forall\, m \in \{1:d\}, \quad \forall\, \kappa \in \{1:N_e\}. \tag{5.25}$$

Proof. The proof follows similarly to that of [97, Theorem 6], where we use the SBP property in (3.6) to obtain (5.20), where applying such an operator to a vector of ones results in

$$\begin{aligned}\underline{\underline{Q}}^{(\kappa,m)}\underline{1}^{(N_q)} = &\frac{1}{2}\underline{\underline{W}}\left(\sum_{l=1}^{d}\underline{\underline{G}}^{(\kappa,l,m)}\underline{\underline{D}}^{(l)}\underline{1}^{(N_q)} + \sum_{l=1}^{d}\underline{\underline{D}}^{(l)}\underline{\underline{G}}^{(\kappa,l,m)}\underline{1}^{(N_q)}\right)\\ &+\frac{1}{2}\sum_{\zeta=1}^{N_f}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\left(\underline{\underline{N}}^{(\kappa,\zeta,m)}\underline{\underline{R}}^{(\zeta)}\underline{1}^{(N_q)} - \sum_{l=1}^{d}\hat{n}_l^{(\zeta)}\underline{\underline{R}}^{(\zeta)}\underline{\underline{G}}^{(\kappa,l,m)}\underline{1}^{(N_q)}\right).\end{aligned} \tag{5.26}$$

The first term within the parentheses on the first line then vanishes due to the fact that $\underline{\underline{D}}^{(l)}\underline{1}^{(N_q)} = \underline{0}^{(N_q)}$ holds under Assumption 3.1, while for the second term we can use exactness of the derivative operators for $G_{lm}^{(\kappa)} \in \mathbb{P}_q(\hat{\Omega})$ as well as the metric identities in (2.12) to obtain

$$\begin{aligned}\sum_{l=1}^{d}\underline{\underline{D}}^{(l)}\underline{\underline{G}}^{(\kappa,l,m)}\underline{1}^{(N_q)} &= \sum_{l=1}^{d}\underline{\underline{D}}^{(l)}\left[G_{lm}^{(\kappa)}(\boldsymbol{\xi}^{(1)}),\ldots,G_{lm}^{(\kappa)}(\boldsymbol{\xi}^{(N_q)})\right]^{\mathrm{T}}\\ &= \sum_{l=1}^{d}\left[\frac{\partial G_{lm}^{(\kappa)}}{\partial \xi_l}(\boldsymbol{\xi}^{(1)}),\ldots,\frac{\partial G_{lm}^{(\kappa)}}{\partial \xi_l}(\boldsymbol{\xi}^{(N_q)})\right]^{\mathrm{T}} = \underline{0}^{(N_q)}.\end{aligned} \tag{5.27}$$

Similarly, we can use the exactness of the interpolation/extrapolation operators for $G_{lm}^{(\kappa)} \in \mathbb{P}_q(\hat{\Omega})$ as well as the fact that the exact or approximate normals satisfy (2.14) to obtain

$$\begin{aligned}\sum_{l=1}^{d}\hat{n}_l^{(\zeta)}\underline{\underline{R}}^{(\zeta)}\underline{\underline{G}}^{(\kappa,l,m)}\underline{1}^{(N_q)} &= \sum_{l=1}^{d}\hat{n}_l^{(\zeta)}\underline{\underline{R}}^{(\zeta)}\left[G_{lm}^{(\kappa)}(\boldsymbol{\xi}^{(1)}),\ldots,G_{lm}^{(\kappa)}(\boldsymbol{\xi}^{(N_q)})\right]^{\mathrm{T}}\\ &= \begin{bmatrix} J^{(\kappa,\zeta)}(\boldsymbol{\xi}^{(\zeta,1)})\boldsymbol{n}^{(\kappa,\zeta)}(\boldsymbol{\xi}^{(\zeta,1)}) \\ \vdots \\ J^{(\kappa,\zeta)}(\boldsymbol{\xi}^{(\zeta,N_{q_f}^{(\zeta)})})\boldsymbol{n}^{(\kappa,\zeta)}(\boldsymbol{\xi}^{(\zeta,N_{q_f}^{(\zeta)})}) \end{bmatrix}\\ &= \underline{\underline{N}}^{(\kappa,\zeta,m)}\underline{\underline{R}}^{(\zeta)}\underline{1}^{(N_q)},\end{aligned} \tag{5.28}$$

where we have used $\underline{\underline{R}}^{(\zeta)}\underline{1}^{(N_q)} = \underline{1}^{(N_{q_f}^{(\zeta)})}$ to obtain the last line, resulting in a cancellation on the second line of (5.26). Using the fact that the volume and facet contributions to the right-hand side of (5.26) are each zero, we therefore obtain (5.25). □

Remark 5.2. While we obtain the discrete metric identities by using the fact that the volume and facet contributions to (5.26) individually vanish when the metric terms and normals are approximated using the conservative curl form described in Section 5.1.1, Crean *et al.* take a different approach in [97], solving quadratic optimization problems to obtain approximate metric terms and normals for which the volume and facet contributions to the metric identities cancel, but do not necessarily vanish individually.

Invoking the assumptions of a conforming mesh and a consistent numerical flux, we can show that the discrete metric identities imply that a uniform solution state remains constant in time. Such a property, which is known as *free-stream preservation* in the context of fluid dynamics, is established with the following theorem.

Theorem 5.1. *Let Assumption 3.3 hold and assume that the discrete metric identities in (5.25) are satisfied and that the numerical flux satisfies the consistency condition in (2.23b). For any uniform solution state satisfying the boundary conditions, the right-hand side given by (5.14) then vanishes as*

$$\underline{r}^{(h,\kappa,e)}(t) = \underline{0}^{(N_q)}, \quad \forall\, e \in \{1 : N_c\}, \quad \forall\, \kappa \in \{1 : N_e\}, \tag{5.29}$$

such that the solutions to (5.13a), (5.13b), and (5.23) remain constant in time.

Proof. Using the equivalent formulation in (5.15), we can use the SBP property in physical space given by (5.19) to obtain an equivalent strong formulation as

$$\begin{aligned}\underline{r}^{(h,\kappa,e)}(t) = &- \sum_{m=1}^{d} \underline{\underline{Q}}^{(\kappa,m)} \underline{f}^{(\kappa,m,e)}(t) \\ &- \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \left(\underline{\underline{J}}^{(\kappa,\zeta)} \underline{f}^{(*,\kappa,\zeta,e)}(t) - \sum_{m=1}^{d} \underline{\underline{N}}^{(\kappa,\zeta,m)} \underline{\underline{R}}^{(\zeta)} \underline{f}^{(\kappa,m,e)}(t) \right).\end{aligned} \tag{5.30}$$

Since the physical flux components in (2.1a) depend only on the solution variables and not, for example, on a spatially varying coefficient, the first term on the right-hand side of (5.30) vanishes for any constant solution when the metric identities in (5.25) are satisfied. The second term then vanishes due to the consistency of the numerical flux and the exactness of the interpolation/extrapolation operators for constant functions. The entire right-hand side of (5.30) then vanishes for all elements and all solution variables. □

5.2.2 Conservation

We demonstrate that the skew-symmetric nodal and modal formulations in (5.13a) are locally and globally conservative with the following theorem, where, unlike in the proofs of conservation for the DG and FR methods in Chapter 3, we must explicitly invoke the discrete metric identities to ensure that the volume terms vanish.

Theorem 5.2. *Under Assumptions 3.3 and 5.1, the skew-symmetric nodal and modal DSEM formulations given by (5.13a) and (5.13b), respectively, with $\underline{r}^{(h,\kappa,e)}(t)$ defined as in (5.14) are locally and globally conservative, satisfying*

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{u}^{(h,\kappa,e)}(t) = -\sum_{\zeta=1}^{N_f}\left(\underline{1}^{(N_q^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t), \tag{5.31a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{\kappa=1}^{N_e}\left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{u}^{(h,\kappa,e)}(t) = -\sum_{\Gamma^{(\kappa,\zeta)}\subset\partial\Omega}\left(\underline{1}^{(N_q^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t), \tag{5.31b}$$

when the discrete metric identities in (5.25) are satisfied and the numerical flux satisfies the conservation property in (2.23a).

Proof. Multiplying both sides of the nodal formulation in (5.13a) from the left by $(\underline{1}^{(N_q)})^{\mathrm{T}}$ or multiplying both sides of the modal formulation in (5.13b) from the left by $\underline{1}^{\mathrm{T}}$, where, as in Theorem 3.3, we define $\underline{1} \in \mathbb{R}^{N_p}$ such that $\underline{\underline{V}}\,\underline{1} = \underline{1}^{(N_q)}$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{u}^{(h,\kappa,e)}(t) = \left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{r}^{(h,\kappa,e)}(t). \tag{5.32}$$

Expressing $\underline{r}^{(h,\kappa,e)}(t)$ as in (5.15), the right-hand side then becomes

$$\begin{aligned}\left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{r}^{(h,\kappa,e)}(t) &= \sum_{m=1}^{d}\left(\underline{\underline{Q}}^{(\kappa,m)}\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{f}^{(\kappa,m,e)}(t) - \sum_{\zeta=1}^{N_f}\left(\underline{\underline{R}}^{(\zeta)}\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t)\\ &= -\sum_{\zeta=1}^{N_f}\left(\underline{1}^{(N_q^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t),\end{aligned} \tag{5.33}$$

where $\underline{\underline{Q}}^{(\kappa,m)}\underline{1}^{(N_q)} = \underline{0}^{(N_q)}$ and $\underline{\underline{R}}^{(\zeta)}\underline{1}^{(N_q)} = \underline{1}^{(N_q^{(\zeta)})}$ are used to obtain the final equality. We then obtain the statement of local conservation in (3.44), and the proof of global conservation follows identically to that of Theorem 3.4. □

The following theorem extends the above proof of conservation to the weight-adjusted formulation in (5.23).

Theorem 5.3. *Let Assumptions 3.3 and 5.1 hold, and assume that the quadrature rule in (2.44) is of degree $\tau \geq p + p_g$, where we use (5.24) to obtain $J^{(\kappa)} \in \mathbb{P}_{p_g}(\hat{\Omega})$. The skew-*

symmetric weight-adjusted modal DSEM given by (5.23) with $\underline{r}^{(h,\kappa,e)}(t)$ defined as in (5.14) is then locally and globally conservative, satisfying (5.31a) and (5.31b) when the discrete metric identities in (5.25) hold and the numerical flux satisfies the conservation property in (2.23a).

Proof. As with the standard (i.e. not weight-adjusted) modal formulation in (5.13b), we multiply from the left by $\underline{1}^{\mathrm{T}}$ and use $\underline{\underline{V}}\,\underline{1} = \underline{1}^{(N_q)}$ as well as (5.33) to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{1}^{\mathrm{T}}\underline{\underline{\tilde{M}}}^{(\kappa)}\underline{\tilde{u}}^{(h,\kappa,e)}(t) = -\sum_{\zeta=1}^{N_f}\left(\underline{1}^{(N_{q_f}^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa)}\underline{f}^{(*,\kappa,\zeta,e)}(t), \tag{5.34}$$

which is a statement of local conservation with respect to the weight-adjusted mass matrix. Following [133, Lemma 2], we use the exactness of the quadrature for $J^{(\kappa)}U_e^{(h,\kappa)}(\cdot,t) \in \mathbb{P}_{p+p_g}(\hat{\Omega})$ to obtain

$$\begin{aligned}\underline{1}^{\mathrm{T}}\underline{\underline{\tilde{M}}}^{(\kappa)}\underline{\tilde{u}}^{(h,\kappa,e)}(t) &= \int_{\hat{\Omega}} J^{(\kappa)}(\boldsymbol{\xi})U_e^{(h,\kappa)}(\boldsymbol{\xi},t)\,\mathrm{d}\boldsymbol{\xi} \\ &= \left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{u}^{(h,\kappa,e)}(t).\end{aligned} \tag{5.35}$$

The remainder of the proof is then identical to those of Theorems 3.4 and 5.2. □

5.2.3 Energy stability

While the standard DG and FR methods were shown in Chapter 3 to be energy stable for the linear advection equation when the SBP property is satisfied and the mapping from reference to physical coordinates is affine, the proposed skew-symmetric discretizations are energy stable on curvilinear meshes, even when the integrals in (5.12) are not computed exactly under the chosen quadrature rules. To prove this, we require the following lemma.

Lemma 5.2. *Let $\underline{r}^{(h,\kappa)}(t)$ denote the nodal right-hand side for the skew-symmetric formulation in (5.14) applied to the constant-coefficient linear advection equation and assume that $\underline{\underline{W}}$ is diagonal. Any nodal solution vector $\underline{u}^{(h,\kappa)}(t) \in \mathbb{R}^{N_q}$ then satisfies (3.54).*

Proof. Substituting $\underline{f}^{(\kappa,m)}(t) = a_m\underline{u}^{(h,\kappa)}(t)$ into (5.14) and multiplying from the left by $(\underline{u}^{(h,\kappa)}(t))^{\mathrm{T}}$, we obtain

$$\begin{aligned}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\underline{r}^{(h,\kappa)}(t) = &-\sum_{m=1}^{d} a_m\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\underline{\underline{S}}^{(\kappa,m)}\underline{u}^{(h,\kappa)}(t) \\ &+\sum_{\zeta=1}^{N_f}\sum_{m=1}^{d}\frac{a_m}{2}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{N}}^{(\kappa,\zeta,m)}\underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa)}(t) \\ &-\sum_{\zeta=1}^{N_f}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta)}(t),\end{aligned} \tag{5.36}$$

where we define

$$\underline{\underline{S}}^{(\kappa,m)} := \frac{1}{2}\sum_{l=1}^{d}\left(\underline{\underline{G}}^{(\kappa,l,m)}\underline{\underline{W}}\,\underline{\underline{D}}^{(l)} - \left(\underline{\underline{D}}^{(l)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{G}}^{(\kappa,l,m)}\right). \tag{5.37}$$

Since the matrix $\underline{\underline{S}}^{(\kappa,m)}$ defined above is skew-symmetric when $\underline{\underline{W}}$ and $\underline{\underline{G}}^{(\kappa,l,m)}$ are both diagonal, the first term on the right-hand side of (5.36) therefore vanishes, resulting in (3.54). □

Energy stability then follows from the following theorem, where in the case of the weight-adjusted formulation, the solution energy is defined in terms of the weight-adjusted mass matrix, which is the inverse of the matrix defined in (5.22).

Theorem 5.4. *Let Assumptions 3.3 and 5.1 hold and assume that the numerical flux for the constant-coefficient linear advection equation takes the form given in (3.58), where the parameter $\lambda \in \mathbb{R}_0^+$ takes the same value on each side of an interior or periodic interface, and that the (curvilinear) mapping from the reference element onto each physical element satisfies (2.11). Defining the solution energy for the formulations in (5.13a) or (5.13b) as*

$$\mathcal{E}^h(t) := \frac{1}{2}\sum_{\kappa=1}^{N_e}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{u}^{(h,\kappa)}(t), \tag{5.38}$$

and for the weight-adjusted formulation in (5.23) as

$$\mathcal{E}^h(t) := \frac{1}{2}\sum_{\kappa=1}^{N_e}\left(\underline{\tilde{u}}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\underline{\underline{\tilde{M}}}^{(\kappa)}\underline{\tilde{u}}^{(h,\kappa)}(t), \tag{5.39}$$

such a scheme then satisfies a semi-discrete energy estimate in the form of (3.59) when applied to the constant-coefficient linear advection equation.

Proof. For the formulation in (5.13a), we multiply from the left by $(\underline{u}^{(h,\kappa)}(t))^{\mathrm{T}}$ and use the fact that $\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}$ is symmetric for a general curvilinear mapping when $\underline{\underline{W}}$ is diagonal to obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{u}^{(h,\kappa)}(t) = \left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\underline{r}^{(h,\kappa)}(t). \tag{5.40}$$

The above relation is also obtained for the modal formulation in (5.13b) through multiplication from the left by $(\underline{\tilde{u}}^{(h,\kappa)}(t))^{\mathrm{T}}$ and using $\underline{u}^{(h,\kappa)}(t) = \underline{\underline{V}}\,\underline{\tilde{u}}^{(h,\kappa)}(t)$, whereas for the weight-adjusted formulation, we similarly have

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\underline{\tilde{u}}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\underline{\underline{\tilde{M}}}^{(\kappa)}\underline{\tilde{u}}^{(h,\kappa)}(t) = \left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}}\underline{r}^{(h,\kappa)}(t). \tag{5.41}$$

Using Lemma 5.2 on the right-hand side of (5.40) or (5.41), summing over all elements, and

using the respective definitions of $\mathcal{E}^h(t)$ in (5.38) and (5.39), we then obtain the energy balance in (3.62). The remainder of the proof therefore follows identically to that of Theorem 3.5. □

5.3 Efficient implementation

In this section, we discuss the efficient implementation of the proposed skew-symmetric tensor-product DSEMs on curved triangles and tetrahedra, particularly in conjunction with explicit time integration. With the exception of the numerical flux evaluation, all operations are entirely local to a given element, and can therefore be executed in parallel in a straightforward manner. Several strategies for computing such local operations are described below.

5.3.1 Reference-operator algorithms

Expressing (5.14) in terms of operators on the reference element and separating the volume and facet contributions, we can compute the nodal right-hand side as

$$\begin{aligned}\underline{r}^{(h,\kappa,e)}(t) = &\sum_{l=1}^{d}\sum_{m=1}^{d}\left(\left(\underline{\underline{D}}^{(l)}\right)^{\mathrm{T}}\left[\tfrac{1}{2}\underline{\underline{W}}\,\underline{\underline{G}}^{(\kappa,l,m)}\right]\underline{f}^{(\kappa,m,e)}(t) - \left[\tfrac{1}{2}\underline{\underline{W}}\,\underline{\underline{G}}^{(\kappa,l,m)}\right]\underline{\underline{D}}^{(l)}\,\underline{f}^{(\kappa,m,e)}(t)\right)\\ &- \sum_{\zeta=1}^{N_f}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\left[\underline{\underline{B}}^{(\zeta)}\,\underline{\underline{J}}^{(\kappa,\zeta)}\right]\left(\underline{f}^{(*,\kappa,\zeta,e)}(t) - \sum_{m=1}^{d}\left[\tfrac{1}{2}\underline{\underline{N}}^{(\kappa,\zeta,m)}\right]\underline{\underline{R}}^{(\zeta)}\,\underline{f}^{(\kappa,m,e)}(t)\right),\end{aligned} \tag{5.42}$$

where square brackets are used to denote operators which must be precomputed and stored for each element, which in (5.42) are all diagonal for the proposed operators. For discretizations on triangles and tetrahedra using non-tensorial operators, the reference operators $\underline{\underline{D}}^{(l)}$ and $\underline{\underline{R}}^{(\zeta)}$ are typically stored as dense matrices and applied, for example, using standard BLAS operations. In the context of the proposed schemes, however, we have the additional option of exploiting the tensor-product structure of such operators through sum factorization. To obtain a further optimization in such cases, we redefine the following operators:

$$\begin{aligned}\left[\tfrac{1}{2}\underline{\underline{W}}\,\underline{\underline{G}}^{(\kappa,l,m)}\right]_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} &\leftarrow \frac{1}{2}\sum_{k=1}^{d}\left[\left(\nabla_{\boldsymbol{\eta}}\boldsymbol{\chi}(\eta_1^{(\alpha_1)},\ldots,\eta_d^{(\alpha_d)})\right)^{-1}\right]_{lk} G_{km}^{(\kappa)}(\boldsymbol{\xi}^{(\sigma(\boldsymbol{\alpha}))})\omega^{(\sigma(\boldsymbol{\alpha}))}\delta_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})},\\ \left[\underline{\underline{D}}^{(l)}\right]_{\sigma(\boldsymbol{\alpha})\sigma(\boldsymbol{\beta})} &\leftarrow \frac{\mathrm{d}\ell_l^{(\beta_l)}}{\mathrm{d}\eta_l}(\eta_l^{(\alpha_l)})\prod_{m\in\{1:d\}\setminus\{l\}}\delta_{\alpha_m\beta_m}.\end{aligned} \tag{5.43}$$

These modifications combine the geometric factors arising from the transformations $\boldsymbol{\chi} : [-1,1]^d \to \hat{\Omega}$ and $\boldsymbol{X}^{(\kappa)} : \hat{\Omega} \to \Omega^{(\kappa)}$, allowing for the volume contributions in (5.42) to be evaluated in collapsed coordinates with an equivalent cost per element to that of a comparable tensor-product discretization on curved quadrilaterals or hexahedra. To evaluate the time

derivative for the nodal formulation, we simply left-multiply the right-hand-side vector computed as in (5.42) by the inverse of the diagonal nodal mass matrix, resulting in

$$\frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = \left[\left(\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\right)^{-1}\right]\underline{r}^{(h,\kappa,e)}(t). \tag{5.44}$$

For the weight-adjusted modal formulation in (5.23), the time derivative can be computed explicitly as

$$\frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = \underline{\underline{M}}^{-1}\underline{\underline{V}}^{\mathrm{T}}\left[\underline{\underline{W}}\left(\underline{\underline{J}}^{(\kappa)}\right)^{-1}\right]\underline{\underline{V}}\,\underline{\underline{M}}^{-1}\underline{\underline{V}}^{\mathrm{T}}\underline{r}^{(h,\kappa,e)}(t), \tag{5.45}$$

where we recall that the application of $\underline{\underline{M}}^{-1}$ can be avoided by choosing an orthonormal basis and using a volume quadrature rule of degree $2p$ or higher. Since the use of the PKD basis allows for $\underline{\underline{V}}$ and $\underline{\underline{V}}^{\mathrm{T}}$ to be applied using sum factorization, and all other operators are either diagonal or possess a standard Kronecker-product structure, the number of operations required for evaluating the time derivative in either (5.44) or (5.23) scales as $\mathcal{O}(p^{d+1})$, assuming in the triangular case that q_1, q_2, and q_f scale as $\mathcal{O}(p)$, and in the tetrahedral case that q_1, q_2, q_3, q_{f1}, and q_{f2} scale as $\mathcal{O}(p)$. Asymptotically, this compares favourably to the $\mathcal{O}(p^{2d})$ complexity of a standard (i.e. non-tensor-product) multidimensional scheme similarly employing $\mathcal{O}(p^d)$ volume quadrature nodes and $\mathcal{O}(p^{d-1})$ facet quadrature nodes. Furthermore, since the only operators which must be stored for each element are diagonal, the storage requirements for reference-operator algorithms scale with the number of quadrature nodes, that is, as $\mathcal{O}(p^d)$.

5.3.2 Physical-operator algorithms

Whether or not sum factorization is employed, and whether a nodal or modal formulation is chosen, the algorithms described in Section 5.3.1 all share the feature of avoiding the precomputation and storage of dense operator matrices for each physical element. However, provided that sufficient memory is available and that the latency associated with accessing local operators from different memory locations for each element is not prohibitive, one has the option of instead precomputing physical operator matrices, an approach which can be competitive with sum factorization at lower polynomial degrees despite scaling asymptotically as $\mathcal{O}(p^{2d})$. Through such an approach, the time derivative can then be obtained for the nodal formulation as

$$\begin{aligned}\frac{\mathrm{d}\underline{u}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = &\sum_{m=1}^{d}\left[\left(\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\right)^{-1}\left(\underline{\underline{Q}}^{(\kappa,m)}\right)^{\mathrm{T}}\right]\underline{f}^{(\kappa,m,e)}(t)\\ &-\sum_{\zeta=1}^{N_f}\left[\left(\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\right)^{-1}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\,\underline{\underline{J}}^{(\kappa,\zeta)}\right]\underline{f}^{(*,\kappa,\zeta,e)}(t),\end{aligned} \tag{5.46}$$

and for the modal formulation as

$$\begin{aligned}\frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa,e)}(t)}{\mathrm{d}t} &= \sum_{m=1}^{d}\left[\left(\underline{\underline{\tilde{M}}}^{(\kappa)}\right)^{-1}\underline{\underline{V}}^{\mathrm{T}}\left(\underline{\underline{Q}}^{(\kappa,m)}\right)^{\mathrm{T}}\right]\underline{f}^{(\kappa,m,e)}(t)\\ &\quad - \sum_{\zeta=1}^{N_f}\left[\left(\underline{\underline{\tilde{M}}}^{(\kappa)}\right)^{-1}\underline{\underline{V}}^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\right]\underline{f}^{(*,\kappa,\zeta,e)}(t),\end{aligned} \tag{5.47}$$

where, although we have used $\underline{\underline{\tilde{M}}}^{(\kappa)}$ instead of $\underline{\underline{M}}^{(\kappa)}$ for consistency with (5.23), there is no advantage in operation count (aside from that of precomputing the inverse) nor in storage to using such an approximation for physical-operator algorithms.

5.4 Chapter summary

This chapter describes the application of the novel tensor-product SBP operators on the reference triangle and tetrahedron introduced in Chapter 4 to the construction of efficient skew-symmetric energy-stable discretizations on curvilinear unstructured grids. Nodal and modal variants of the proposed approach are presented, the latter making use of a projection onto an orthonormal basis for a total-degree polynomial space in order to circumvent the explicit time step restriction resulting from the singular nature of the collapsed coordinate transformation. We approximate the metric terms in conservative curl form in order to ensure that the discrete metric identities are satisfied in three dimensions, and, in the case of the modal approach, a weight-adjusted approximation is used to obtain fully explicit low-storage algorithms of $\mathcal{O}(p^{d+1})$ time complexity, avoiding the inversion of a dense local mass matrix. Proofs of free-stream preservation, conservation, as well as energy stability for the linear advection equation are presented, and techniques for the efficient implementation of the proposed schemes are described.

Chapter 6

Entropy-stable tensor-product discontinuous spectral-element methods on curved triangles and tetrahedra

In this chapter, which is based on the content of Paper IV, we detail how the SBP operators described in Chapter 4 can be used to construct efficient entropy-stable DSEMs of any order for nonlinear systems of conservation laws on curvilinear unstructured grids. Such formulations extend the approach devised in Chapter 5 to entropy-stable schemes using techniques introduced by Chan [99] as well as Chan and Wilcox [133]. We will focus on the modal formulation due to the fact that the use of the PKD basis described in Section 5.1.3 together with the weight-adjusted approximation described in Section 5.1.6 allows for all operators to be evaluated using efficient sum-factorization techniques of $\mathcal{O}(p^{d+1})$ complexity without negatively affecting the explicit time step restriction.[1] The proposed entropy-stable schemes will be constructed so as to enable the use of the sum-factorization techniques discussed in the previous chapter while additionally exploiting the sparsity of the SBP operators described in Chapter 4 to reduce the number of required entropy-conservative two-point flux evaluations. We introduce the essential components of such schemes in Section 6.1 and analyze the resulting formulations in Section 6.2. Finally, Section 6.3 describes how the tensor-product structure and sparsity of the proposed tensor-product operators on triangles and tetrahedra can be used to obtain efficient entropy-stable algorithms.

6.1 Entropy-stable discontinuous spectral-element formulation

Similarly to the energy-stable modal DSEMs described in Chapter 5, the formulations described can be expressed in the form of (5.23) using the weight-adjusted inverse in (5.22), where the metric terms are approximated using the conservative curl formulation in (5.3)

[1] This will be verified numerically in Chapter 7 through examination of the spectral radius of the semi-discrete operator arising from the spatial discretization of the linear advection equation.

and the Jacobian determinant of the mapping is approximated as in (5.24). We will combine such ingredients with an entropy projection, an entropy-conservative volume flux, an entropy-stable interface flux, and a flux-differencing formulation to obtain efficient entropy-stable discretizations on curved triangles and tetrahedra.

6.1.1 Entropy projection

We begin by considering a fundamental issue in the development of entropy-stable modal DSEMs, which is the fact that the entropy variables may not lie within the approximation space in which the solution is sought. As such, they cannot be taken as test functions in the discrete variational formulation, a critical step in establishing entropy stability for such schemes. To resolve this, entropy-stable modal formulations employ the *entropy projection* procedure introduced in [99], referring to the approximation of the entropy variables as

$$\mathcal{W}_e(\underline{U}^{(h,\kappa)}(\boldsymbol{\xi},t)) \approx \sum_{i=1}^{N_p} \tilde{w}_i^{(h,\kappa,e)}(t)\phi^{(i)}(\boldsymbol{\xi}), \tag{6.1}$$

where, as in [133, Eq. (31)], we obtain the coefficients through a weight-adjusted projection as

$$\underline{\tilde{w}}^{(h,\kappa,e)}(t) \coloneqq \left(\underline{\underline{\tilde{M}}}^{(\kappa)}\right)^{-1} \underline{\underline{V}}^{\mathrm{T}} \underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)} \begin{bmatrix} \mathcal{W}_e(\underline{u}_1^{(h,\kappa)}(t)) \\ \vdots \\ \mathcal{W}_e(\underline{u}_{N_q}^{(h,\kappa)}(t)) \end{bmatrix}. \tag{6.2}$$

Using the generalized Vandermonde matrix to evaluate the projected entropy variables at the volume quadrature nodes as

$$\underline{w}^{(h,\kappa,e)}(t) \coloneqq \underline{\underline{V}}\,\underline{\tilde{w}}^{(h,\kappa,e)}(t), \tag{6.3}$$

we obtain

$$\underline{w}_i^{(h,\kappa)}(t) \coloneqq \begin{bmatrix} w_i^{(h,\kappa,1)}(t) \\ \vdots \\ w_i^{(h,\kappa,N_c)}(t) \end{bmatrix}, \quad \underline{w}_i^{(h,\kappa,\zeta)}(t) \coloneqq \begin{bmatrix} [\underline{\underline{R}}^{(\zeta)}\underline{w}^{(h,\kappa,1)}(t)]_i \\ \vdots \\ [\underline{\underline{R}}^{(\zeta)}\underline{w}^{(h,\kappa,N_c)}(t)]_i \end{bmatrix}. \tag{6.4}$$

The conservative variables are then evaluated in terms of the projected entropy variables in order to obtain the *entropy-projected conservative variables* as $\underline{\mathcal{U}}(\underline{w}_i^{(h,\kappa)}(t))$ and $\underline{\mathcal{U}}(\underline{w}_i^{(h,\kappa,\zeta)}(t))$.

6.1.2 Entropy-conservative and entropy-stable flux functions

Next, we present the following definition of an *entropy-conservative two-point flux*, which is an essential component of the entropy-stable methods described in this work.

Definition 6.1. A continuously differentiable function $\underline{F}_m^{\#} : \Upsilon \times \Upsilon \to \mathbb{R}^{N_c}$ is an *entropy-conservative two-point flux* if it is symmetric and consistent with (2.1a), satisfying

$$\underline{F}_m^{\#}(\underline{U}^-, \underline{U}^+) = \underline{F}_m^{\#}(\underline{U}^+, \underline{U}^-), \quad \forall \underline{U}^-, \underline{U}^+ \in \Upsilon, \tag{6.5a}$$

$$\underline{F}_m^{\#}(\underline{U}, \underline{U}) = \underline{F}_m(\underline{U}), \qquad \forall \underline{U} \in \Upsilon, \tag{6.5b}$$

and, recalling the definition of the flux potential in (2.7), the following condition holds:

$$\Big(\underline{\mathcal{W}}(\underline{U}^+) - \underline{\mathcal{W}}(\underline{U}^-)\Big)^{\mathrm{T}} \underline{F}_m^{\#}(\underline{U}^-, \underline{U}^+) = \Psi_m(\underline{\mathcal{W}}(\underline{U}^+)) - \Psi_m(\underline{\mathcal{W}}(\underline{U}^-)), \quad \forall \underline{U}^-, \underline{U}^+ \in \Upsilon. \tag{6.6}$$

First proposed in [85], the property in (6.6) is referred to in the literature as *Tadmor's condition* or the *shuffle condition*, and enables the chain rule to be circumvented when deriving semi-discrete forms of bounds such as (2.8) for an entropy-stable discretization. At element interfaces, we use *entropy-stable* or *entropy-conservative* directional numerical fluxes, for which the following definition is introduced (see, for example, [98, Definitions 3.1 and 3.2]).

Definition 6.2. A conservative and consistent numerical flux $\underline{F}^* : \Upsilon \times \Upsilon \times \mathbb{S}^{d-1} \to \mathbb{R}^{N_e}$ is *entropy stable* if, for any direction vector $\boldsymbol{n} \in \mathbb{S}^{d-1}$, the following condition is satisfied:

$$\Big(\underline{\mathcal{W}}(\underline{U}^+) - \underline{\mathcal{W}}(\underline{U}^-)\Big)^{\mathrm{T}} \underline{F}^*(\underline{U}^-, \underline{U}^+, \boldsymbol{n}) \leq \Big(\boldsymbol{\Psi}(\underline{\mathcal{W}}(\underline{U}^+)) - \boldsymbol{\Psi}(\underline{\mathcal{W}}(\underline{U}^-))\Big) \cdot \boldsymbol{n}, \quad \forall \underline{U}^-, \underline{U}^+ \in \Upsilon. \tag{6.7}$$

Such a numerical flux is *entropy conservative* if (6.7) holds as an equality for all arguments.

In this work, we employ a numerical interface flux consisting of an entropy-conservative two-point flux in the normal direction augmented by a local Lax–Friedrichs dissipative term (see, for example, Ranocha [141, Section 6.1]). Such a flux takes the form

$$\underline{F}^*(\underline{U}^-, \underline{U}^+, \boldsymbol{n}) \coloneqq \underline{F}^{\#}(\underline{U}^-, \underline{U}^+, \boldsymbol{n}) - \frac{1}{2} \Lambda(\underline{U}^-, \underline{U}^+, \boldsymbol{n})(\underline{U}^+ - \underline{U}^-), \tag{6.8}$$

where the entropy-conservative directional flux is given by

$$\underline{F}^{\#}(\underline{U}^-, \underline{U}^+, \boldsymbol{n}) \coloneqq \sum_{m=1}^{d} n_m \underline{F}_m^{\#}(\underline{U}^-, \underline{U}^+), \tag{6.9}$$

and $\Lambda(\underline{U}^-, \underline{U}^+, \boldsymbol{n}) \in \mathbb{R}^+$ is an estimate of the maximum normal wave speed in the normal direction to the interface. As in [99, 133], the arguments $\underline{U}^-$ and $\underline{U}^+$ are taken in this work to be the entropy-projected conservative variables, although we note that alternative interface dissipation approaches have been proposed based on penalization of the jump in the entropy variables, either through scalar dissipation (which is used, for example, in [93] and [97]) or using matrix dissipation operators such as those proposed by Winters *et al.* [142].

6.1.3 Flux-differencing formulation

In order to construct efficient entropy-stable formulations on curvilinear meshes, we take a similar approach to Fisher [88, Section 4.5] and Gassner *et al.* [93, Appendix B] and evaluate the averaged metric terms between pairs of volume quadrature nodes and pairs of volume and facet quadrature nodes as

$$\{\{G_{lm}^{(\kappa)}\}\}_{ij} \coloneqq \frac{1}{2}[\boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi}^{(i)}) + \boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi}^{(j)})]_{lm}, \tag{6.10a}$$

$$\{\{J^{(\kappa,\zeta)}n_m^{(\kappa,\zeta)}\}\}_{ij} \coloneqq \frac{1}{2}[\boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi}^{(i)})^{\mathrm{T}}\hat{\boldsymbol{n}}^{(\zeta)} + \boldsymbol{G}^{(\kappa)}(\boldsymbol{\xi}^{(\zeta,j)})^{\mathrm{T}}\hat{\boldsymbol{n}}^{(\zeta)}]_m, \tag{6.10b}$$

which define the matrices $\{\{G_{lm}^{(\kappa)}\}\} \in \mathbb{R}^{N_q\times N_q}$ and $\{\{J^{(\kappa,\zeta)}n_m^{(\kappa,\zeta)}\}\} \in \mathbb{R}^{N_q\times N_{q_f}^{(\zeta)}}$, respectively. The entropy-conservative two-point fluxes are similarly computed between pairs of quadrature nodes using the entropy-projected conservative variables as

$$F_{ij}^{(\kappa,m,e)}(t) \coloneqq F_{me}^{\#}(\underline{\mathcal{U}}(\underline{w}_i^{(h,\kappa)}(t)), \underline{\mathcal{U}}(\underline{w}_j^{(h,\kappa)}(t))), \tag{6.11a}$$

$$F_{ij}^{(\kappa,\zeta,m,e)}(t) \coloneqq F_{me}^{\#}(\underline{\mathcal{U}}(\underline{w}_i^{(h,\kappa)}(t)), \underline{\mathcal{U}}(\underline{w}_j^{(h,\kappa,\zeta)}(t))), \tag{6.11b}$$

defining the matrices $\underline{\underline{F}}^{(\kappa,m,e)}(t) \in \mathbb{R}^{N_q\times N_q}$ and $\underline{\underline{F}}^{(\kappa,\zeta,m,e)}(t) \in \mathbb{R}^{N_q\times N_{q_f}^{(\zeta)}}$. Denoting the exterior values of the entropy variables as $\underline{w}_i^{(h,\kappa,\zeta,+)}(t) \in \mathbb{R}^{N_c}$, we evaluate $\underline{f}^{(*,\kappa,\zeta,e)}(t) \in \mathbb{R}^{N_{q_f}^{(\zeta)}}$ as

$$f_i^{(*,\kappa,\zeta,e)}(t) \coloneqq F_e^*\Big(\underline{\mathcal{U}}(\underline{w}_i^{(h,\kappa,\zeta)}(t)),\, \underline{\mathcal{U}}(\underline{w}_i^{(h,\kappa,\zeta,+)}(t)),\, \boldsymbol{n}^{(\kappa,\zeta)}(\boldsymbol{X}^{(\kappa)}(\boldsymbol{\xi}^{(\zeta,i)}))\Big). \tag{6.12}$$

Having introduced the essential components of the scheme, a flux-differencing weight-adjusted modal DSEM is now obtained by computing the time derivative as in (5.23), with the nodal right-hand side computed as

$$\begin{aligned}\underline{r}^{(h,\kappa,e)}(t) \coloneqq &-\sum_{l=1}^{d}\left(2\underline{\underline{S}}^{(l)} \odot \sum_{m=1}^{d}\{\{G_{lm}^{(\kappa)}\}\} \odot \underline{\underline{F}}^{(\kappa,m,e)}(t)\right)\underline{1}^{(N_q)} \\ &-\sum_{\zeta=1}^{N_f}\left(\underline{\underline{C}}^{(\kappa,\zeta,e)}(t)\underline{1}^{(N_{q_f}^{(\zeta)})} + \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\left(\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t) - \left(\underline{\underline{C}}^{(\kappa,\zeta,e)}(t)\right)^{\mathrm{T}}\underline{1}^{(N_q)}\right)\right),\end{aligned} \tag{6.13}$$

where $\odot$ denotes the Hadamard product given by $\left[\underline{\underline{A}} \odot \underline{\underline{B}}\right]_{ij} \coloneqq A_{ij}B_{ij}$, and we define

$$\underline{\underline{C}}^{(\kappa,\zeta,e)}(t) \coloneqq \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)} \odot \sum_{m=1}^{d}\{\{J^{(\kappa,\zeta)}n_m^{(\kappa,\zeta)}\}\} \odot \underline{\underline{F}}^{(\kappa,\zeta,m,e)}(t). \tag{6.14}$$

The above formulation can be used with any set of SBP operators for which the boundary matrices can be decomposed as in (2.45), and it will be shown in Section 6.2 that the formulation is, in fact, mathematically equivalent to that in [133, Eq. (35)] when the same

SBP operators are used in both schemes, a fact which we will exploit in our analysis of conservation, free-stream preservation, and entropy stability in the following section.

Remark 6.1. When diagonal-E operators are used, the terms in (6.13) involving the correction operator $\underline{\underline{C}}^{(\kappa,\zeta,e)}(t)$ in (6.14) cancel, and hence the scheme simplifies to

$$\begin{aligned}\underline{r}^{(h,\kappa,e)}(t) = &- \sum_{l=1}^{d} \left(2\underline{\underline{S}}^{(l)} \odot \sum_{m=1}^{d} \{\{G_{lm}^{(\kappa)}\}\} \odot \underline{\underline{F}}^{(\kappa,m,e)}(t) \right) \underline{1}^{(N_q)} \\ &- \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)} \right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{J}}^{(\kappa,\zeta)} \underline{f}^{(*,\kappa,\zeta,e)}(t),\end{aligned} \tag{6.15}$$

a formulation which Ranocha *et al.* employ in [143, Section 2.1] to obtain highly efficient algorithms using tensor-product LGL quadrature rules on curved quadrilateral and hexahedral elements. However, since the tensor-product SBP operators on the reference triangle and tetrahedron introduced in Chapter 4 are not diagonal-E operators, the proposed schemes require the evaluation of the full right-hand side in (6.13) including the facet correction, the efficient implementation of which will be discussed in Section 6.3.

6.2 Analysis

We will now demonstrate that the schemes introduced in the previous section are conservative, free-stream preserving, and entropy stable due to their equivalence to formulations based on *hybridized summation-by-parts operators*,[2] which were proposed by Chan in [99].

6.2.1 Equivalent hybridized summation-by-parts formulation

The construction of such hybridized operators from SBP operators on the reference element satisfying the conditions of Definition 2.3 is demonstrated with the following lemma.

Lemma 6.1. *Given any nodal SBP operator* $\underline{\underline{D}}^{(l)} \coloneqq \underline{\underline{W}}^{-1}(\underline{\underline{S}}^{(l)} + \frac{1}{2}\underline{\underline{E}}^{(l)})$ *on the reference element in the sense of Definition 2.3 for which* $\underline{\underline{E}}^{(l)}$ *takes the form of (2.45), the block matrix*

$$\underline{\underline{\bar{Q}}}^{(l)} \coloneqq \begin{bmatrix} \underline{\underline{S}}^{(l)} & \frac{1}{2}\hat{n}_l^{(1)} \left(\underline{\underline{R}}^{(1)} \right)^{\mathrm{T}} \underline{\underline{B}}^{(1)} & \cdots & \frac{1}{2}\hat{n}_l^{(N_f)} \left(\underline{\underline{R}}^{(N_f)} \right)^{\mathrm{T}} \underline{\underline{B}}^{(N_f)} \\ -\frac{1}{2}\hat{n}_l^{(1)} \underline{\underline{B}}^{(1)} \underline{\underline{R}}^{(1)} & \frac{1}{2}\hat{n}_l^{(1)} \underline{\underline{B}}^{(1)} & & \\ \vdots & & \ddots & \\ -\frac{1}{2}\hat{n}_l^{(N_f)} \underline{\underline{B}}^{(N_f)} \underline{\underline{R}}^{(N_f)} & & & \frac{1}{2}\hat{n}_l^{(N_f)} \underline{\underline{B}}^{(N_f)} \end{bmatrix}, \tag{6.16}$$

[2] Such operators were originally introduced in [99] as *decoupled SBP operators*, with the term *hybridized SBP operator* popularized following the review paper by Chen and Shu [100]. We note, however, that this notion of "hybridization" differs from that used to describe finite-element methods which exploit static condensation to reduce the number of coupled degrees of freedom (see, for example, [144] and [145]).

satisfies the following "SBP-like" property from [99, Eq. (32)]:

$$\underline{\underline{\bar{Q}}}^{(l)} + \left(\underline{\underline{\bar{Q}}}^{(l)}\right)^{\mathrm{T}} = \underline{\underline{\bar{E}}}^{(l)}, \quad \textit{where} \quad \underline{\underline{\bar{E}}}^{(l)} \coloneqq \begin{bmatrix} \underline{\underline{0}}^{(N_q \times N_q)} & & & \\ & \hat{n}_l^{(1)} \underline{\underline{B}}^{(1)} & & \\ & & \ddots & \\ & & & \hat{n}_l^{(N_f)} \underline{\underline{B}}^{(N_f)} \end{bmatrix}. \tag{6.17}$$

Proof. Aside from the fact that our notation separates the operators on individual facets, the result is identical to [146, Theorem 1]. The proof relies on the fact that when adding (6.16) to its transpose, the top-left block vanishes by the skew-symmetry of $\underline{\underline{S}}^{(l)}$ and the off-diagonal blocks vanish due to the blocks of the first row being the negative transposes of those in the first column, and hence only the diagonal matrix $\underline{\underline{\bar{E}}}^{(l)}$ remains. □

The hybridized operators in (6.16) are of dimension $\bar{N}_q$ by $\bar{N}_q$, where we define $\bar{N}_q \coloneqq N_q + N_{q_f}^{(1)} + \cdots + N_{q_f}^{(N_f)}$. Within a flux-differencing formulation, such operators act on block matrices of the form

$$\underline{\underline{\bar{F}}}^{(\kappa,m,e)}(t) \coloneqq \begin{bmatrix} \underline{\underline{F}}^{(\kappa,m,e)}(t) & \underline{\underline{F}}^{(\kappa,1,m,e)}(t) & \cdots & \underline{\underline{F}}^{(h,N_f,m,e)}(t) \\ \underline{\underline{F}}^{(\kappa,1,m,e)}(t) & \underline{\underline{F}}^{(\kappa,1,1,m,e)}(t) & \cdots & \underline{\underline{F}}^{(\kappa,1,N_f,m,e)}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\underline{F}}^{(\kappa,N_f,m,e)}(t) & \underline{\underline{F}}^{(\kappa,N_f,1,m,e)}(t) & \cdots & \underline{\underline{F}}^{(\kappa,N_f,N_f,m,e)}(t) \end{bmatrix}, \tag{6.18}$$

where $\underline{\underline{F}}^{(\kappa,\zeta,\eta,m,e)}(t) \in \mathbb{R}^{N_{q_f}^{(\zeta)} \times N_{q_f}^{(\eta)}}$ couples quadrature nodes on the facets $\hat{\Gamma}^{(\zeta)}, \hat{\Gamma}^{(\eta)} \subset \partial\hat{\Omega}$ as

$$F_{ij}^{(\kappa,\zeta,\eta,m,e)}(t) \coloneqq F_{me}^{\#}(\underline{\mathcal{U}}(\underline{w}_i^{(h,\kappa,\zeta)}(t)), \underline{\mathcal{U}}(\underline{w}_j^{(h,\kappa,\eta)}(t))). \tag{6.19}$$

As in [133], hybridized SBP operators on the physical element are constructed in split form as

$$\underline{\underline{\bar{Q}}}^{(\kappa,m)} \coloneqq \frac{1}{2} \sum_{l=1}^{d} \left(\underline{\underline{\bar{Q}}}^{(l)} \underline{\underline{\bar{G}}}^{(\kappa,l,m)} + \underline{\underline{\bar{G}}}^{(\kappa,l,m)} \underline{\underline{\bar{Q}}}^{(l)} \right), \tag{6.20}$$

using the diagonal matrices of concatenated volume and facet metric terms given by

$$\underline{\underline{\bar{G}}}^{(\kappa,l,m)} \coloneqq \begin{bmatrix} \underline{\underline{G}}^{(\kappa,l,m)} & & & \\ & \underline{\underline{G}}^{(\kappa,1,l,m)} & & \\ & & \ddots & \\ & & & \underline{\underline{G}}^{(\kappa,N_f,l,m)} \end{bmatrix}, \tag{6.21}$$

where the entries of $\underline{\underline{G}}^{(\kappa,\zeta,l,m)} \in \mathbb{R}^{N_{q_f}^{(\zeta)} \times N_{q_f}^{(\zeta)}}$ are given by $G_{ij}^{(\kappa,\zeta,l,m)} \coloneqq G_{lm}^{(\kappa)}(\boldsymbol{\xi}^{(\zeta,i)})\delta_{ij}$. The following lemma demonstrates the relation of (6.20) to the approach adopted in Section 6.1.3

based on averaging of the metric terms between pairs of quadrature nodes.

Lemma 6.2. *The split-form operator in (6.20) can be rewritten as*

$$\underline{\underline{\bar{Q}}}^{(\kappa,m)} = \sum_{l=1}^{d} \underline{\underline{\bar{Q}}}^{(l)} \odot \{\{\bar{G}_{lm}^{(\kappa)}\}\}, \tag{6.22}$$

where the entries of the matrix $\{\{\bar{G}_{lm}^{(\kappa)}\}\} \in \mathbb{R}^{\bar{N}_q \times \bar{N}_q}$ *are given by*

$$\{\{\bar{G}_{lm}^{(\kappa)}\}\}_{ij} \coloneqq \frac{1}{2}\left(\bar{G}_{ii}^{(\kappa,l,m)} + \bar{G}_{jj}^{(\kappa,l,m)}\right). \tag{6.23}$$

Additionally, a property analogous to (6.17) is satisfied in physical space as

$$\underline{\underline{\bar{Q}}}^{(\kappa,m)} + \left(\underline{\underline{\bar{Q}}}^{(\kappa,m)}\right)^{\mathrm{T}} = \begin{bmatrix} \underline{\underline{0}}^{(N_q \times N_q)} & & \\ & \ddots & \\ & & \underline{\underline{B}}^{(N_f)} \underline{\underline{J}}^{(\kappa,N_f)} \underline{\underline{N}}^{(\kappa,N_f,m)} \end{bmatrix}, \tag{6.24}$$

provided that $\underline{\underline{J}}^{(\kappa,\zeta)}$ *and* $\underline{\underline{N}}^{(\kappa,\zeta,m)}$ *are computed using (5.4) based on the volume metric terms.*

Proof. Expressing (6.20) in indicial form and factoring out the scalar $\bar{Q}_{ij}^{(l)}$, we obtain

$$\begin{aligned} \bar{Q}_{ij}^{(\kappa,m,n)} &= \frac{1}{2}\sum_{l=1}^{d}\left(\bar{Q}_{ij}^{(l)}\bar{G}_{jj}^{(\kappa,l,m)} + \bar{G}_{ii}^{(\kappa,l,m)}\bar{Q}_{ij}^{(l)}\right) \\ &= \frac{1}{2}\sum_{l=1}^{d}\bar{Q}_{ij}^{(l)}\left(\bar{G}_{ii}^{(\kappa,l,m)} + \bar{G}_{jj}^{(\kappa,l,m)}\right), \end{aligned} \tag{6.25}$$

and hence the result in (6.22) follows directly from the definition of the Hadamard product. The SBP-like property in (6.24) follows from (6.17) and (6.20), where we obtain the right-hand side using (5.4) and the fact that the hybridized boundary operators $\underline{\underline{\bar{E}}}^{(l)}$ are diagonal. □

We now have the following theorem relating the formulation proposed in Section 6.1.3 to one which is readily analyzed using the properties of hybridized SBP operators.

Theorem 6.1. *Under Assumption 5.1, the DSEM given by (5.23) with* $\underline{r}^{(h,\kappa,e)}(t)$ *defined as in (6.13) is equivalent to the following hybridized SBP formulation proposed in [133, Eq. (35)]:*

$$\begin{aligned} \underline{\underline{\tilde{M}}}^{(\kappa)} \frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = &- \begin{bmatrix} \underline{\underline{V}} \\ \underline{\underline{V}}^{f} \end{bmatrix}^{\mathrm{T}} \sum_{m=1}^{d} \left(2\underline{\underline{\bar{Q}}}^{(\kappa,m)} \odot \underline{\underline{\bar{F}}}^{(\kappa,m,e)}(t)\right)\underline{1}^{(\bar{N}_q)} \\ &- \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{J}}^{(\kappa,\zeta)} \left(\underline{f}^{(*,\kappa,\zeta,e)}(t) - \sum_{m=1}^{d} \underline{\underline{N}}^{(\kappa,\zeta,m)} \underline{f}^{(\kappa,\zeta,m,e)}(t)\right), \end{aligned} \tag{6.26}$$

where we define

$$\underline{\underline{V}}^f \coloneqq \begin{bmatrix} \underline{\underline{R}}^{(1)}\underline{\underline{V}} \\ \vdots \\ \underline{\underline{R}}^{(N_f)}\underline{\underline{V}} \end{bmatrix}, \quad \underline{f}^{(\kappa,\zeta,m,e)}(t) \coloneqq \begin{bmatrix} F_{me}\Big(\underline{\mathcal{U}}(\underline{w}_1^{(h,\kappa,\zeta)}(t))\Big) \\ \vdots \\ F_{me}\Big(\underline{\mathcal{U}}(\underline{w}_{N_{q_f}^{(\zeta)}}^{(h,\kappa,\zeta)}(t))\Big) \end{bmatrix}. \tag{6.27}$$

Proof. Substituting (6.13) into (5.23), multiplying from the left by $\underline{\underline{\tilde{M}}}^{(\kappa)}$, and grouping Hadamard products into block matrices, we obtain

$$\begin{aligned} \underline{\underline{\tilde{M}}}^{(\kappa)} \frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = &- \begin{bmatrix} \underline{\underline{V}} \\ \underline{\underline{V}}^f \end{bmatrix}^{\mathrm{T}} \sum_{l=1}^{d} \left(2\underline{\underline{\bar{S}}}^{(l)} \odot \sum_{m=1}^{d} \{\{\bar{G}_{lm}^{(\kappa)}\}\} \odot \underline{\underline{\bar{F}}}^{(\kappa,m,e)}(t) \right) \underline{1}^{(\bar{N}_q)} \\ &- \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{J}}^{(\kappa,\zeta)} \underline{f}^{(*,\kappa,\zeta,e)}(t), \end{aligned} \tag{6.28}$$

where the facet correction terms in (6.14) have been incorporated into the hybridized operator

$$\underline{\underline{\bar{S}}}^{(l)} \coloneqq \begin{bmatrix} \underline{\underline{S}}^{(l)} & \frac{1}{2}\hat{n}_l^{(1)}\left(\underline{\underline{R}}^{(1)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(1)} & \cdots & \frac{1}{2}\hat{n}_l^{(N_f)}\left(\underline{\underline{R}}^{(N_f)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(N_f)} \\ -\frac{1}{2}\hat{n}_l^{(1)}\underline{\underline{B}}^{(1)}\underline{\underline{R}}^{(1)} & & & \\ \vdots & & & \\ -\frac{1}{2}\hat{n}_l^{(N_f)}\underline{\underline{B}}^{(N_f)}\underline{\underline{R}}^{(N_f)} & & & \end{bmatrix}. \tag{6.29}$$

Recognizing such a matrix as the skew-symmetric part of $\underline{\underline{\bar{Q}}}^{(l)}$, we invoke the SBP-like property from Lemma 6.1 to obtain $2\underline{\underline{\bar{S}}}^{(l)} = 2\underline{\underline{\bar{Q}}}^{(l)} - \underline{\underline{\bar{E}}}^{(l)}$. Substituting such a relation into (6.28) and using the consistency of the two-point flux as well as the fact that $\underline{\underline{\bar{E}}}^{(l)}$ is diagonal, the scheme becomes

$$\begin{aligned} \underline{\underline{\tilde{M}}}^{(\kappa)} \frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa,e)}(t)}{\mathrm{d}t} = &- \begin{bmatrix} \underline{\underline{V}} \\ \underline{\underline{V}}^f \end{bmatrix}^{\mathrm{T}} \sum_{l=1}^{d} \left(2\underline{\underline{\bar{Q}}}^{(l)} \odot \sum_{m=1}^{d} \{\{\bar{G}_{lm}^{(\kappa)}\}\} \odot \underline{\underline{\bar{F}}}^{(\kappa,m,e)}(t) \right) \underline{1}^{(\bar{N}_q)} \\ &- \sum_{\zeta=1}^{N_f} \left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}} \underline{\underline{B}}^{(\zeta)} \underline{\underline{J}}^{(\kappa,\zeta)} \left(\underline{f}^{(*,\kappa,\zeta,e)}(t) - \sum_{m=1}^{d} \underline{\underline{N}}^{(\kappa,\zeta,m)} \underline{f}^{(\kappa,\zeta,m,e)}(t) \right), \end{aligned} \tag{6.30}$$

Finally, considering the first term on the right-hand side of (6.30), we move the sum over the index m outside of the sum over l and invoke Lemma 6.2 to obtain the split-form operator in (6.20). The resulting scheme is then given by (6.26). □

6.2.2 Conservation

As a consequence of the above equivalence, the conservation, free-stream preservation, and entropy stability of the proposed discretizations follow directly from the analysis in [99] and

[133]. Referring to [99] and [133] for further details, we summarize the critical steps of the analysis in the remainder of this section, beginning with the following theorem establishing discrete conservation.

Theorem 6.2. *Let Assumptions 3.3 and 5.1 hold and assume that the quadrature rule in (2.44) is of degree $\tau \geq p + p_g$, where we use (5.24) to obtain $J^{(\kappa)} \in \mathbb{P}_{p_g}(\hat{\Omega})$. The entropy-stable weight-adjusted modal DSEM given by (5.23) with $\underline{r}^{(h,\kappa,e)}(t)$ defined as in (5.14) is then locally and globally conservative, satisfying (5.31a) and (5.31b) when the two-point volume flux satisfies (6.5a) and the directional interface flux satisfies (2.23a).*

Proof. Using the equivalent formulation in (6.28), we proceed as in the proofs of Theorems 3.3 and 5.3, wherein both sides of the formulation are multiplied from the left by $\underline{1}^{\mathrm{T}}$. Noting that the skew-symmetry of $\underline{\underline{\bar{S}}}^{(\kappa,m)}$ and the symmetry of $\{\{\bar{G}_{lm}^{(\kappa)}\}\}$ as well as $\underline{\underline{\bar{F}}}^{(\kappa,m,e)}(t)$ imply that the matrix $2\underline{\underline{\bar{S}}}^{(l)} \odot \{\{\bar{G}_{lm}^{(\kappa)}\}\} \odot \underline{\underline{\bar{F}}}^{(\kappa,m,e)}(t)$ is skew-symmetric, we then use

$$\left(\underline{1}^{(\bar{N}_q)}\right)^{\mathrm{T}} \left(2\underline{\underline{\bar{S}}}^{(l)} \odot \{\{\bar{G}_{lm}^{(\kappa)}\}\} \odot \underline{\underline{\bar{F}}}^{(\kappa,m,e)}(t)\right) \underline{1}^{(\bar{N}_q)} = 0 \tag{6.31}$$

to show that the volume terms of (6.28) vanish, resulting in (5.34). The remainder of the proof is then identical to those of Theorem 5.3. □

6.2.3 Discrete metric identities and free-stream preservation

We present the following lemma from [133, Theorem 5], which establishes conditions under which discrete metric identities analogous to (5.25) are satisfied.

Lemma 6.3. *Assume that the metric terms, whether computed exactly or approximately, satisfy $G_{lm}^{(\kappa)} \in \mathbb{P}_q(\hat{\Omega})$ as well as (2.12), and that the normals are computed as in (5.4). Then, the hybridized SBP operators in (6.20) satisfy the following discrete metric identities:*

$$\underline{\underline{\bar{Q}}}^{(\kappa,m)} \underline{1}^{(\bar{N}_q)} = \underline{0}^{(\bar{N}_q)}, \quad \forall\, m \in \{1:d\}, \quad \forall\, \kappa \in \{1:N_e\}. \tag{6.32}$$

We now use the above lemma to demonstrate that the scheme is free-stream preserving.

Theorem 6.3. *Let Assumptions 3.3 and 5.1 hold, and also assume that the conditions of Lemma 6.3 are satisfied and that the two-point volume flux and the directional interface flux satisfy (2.23b) and (6.5b), respectively. The scheme in (6.26), or, equivalently, in (5.23) with $\underline{r}^{(h,\kappa,e)}(t)$ defined as in (6.13), is then free-stream preserving, such that the right-hand side of (5.23) vanishes for any uniform solution state satisfying the boundary conditions.*

Proof. Considering the formulation in (6.26), the facet penalty on the second line vanishes when the solution is identical on both sides of the interface due to the consistency property

of the numerical interface flux. Invoking the consistency of the two-point flux as well, we then see that the entire right-hand side of (6.26) vanishes when (6.32) holds, which follows from Lemma 6.3. Since the weight-adjusted mass matrix is invertible by construction, the time derivative is zero, and the scheme is therefore free-stream preserving. □

6.2.4 Entropy stability

Invoking the entropy conditions in (6.6) and (6.7), we now present the following theorem, which establishes that the proposed discretizations satisfy a discrete version of (2.8).

Theorem 6.4. *Let Assumptions 3.3 and 5.1 hold, and also assume that the conditions of Lemma 6.3 are satisfied, that the two-point volume flux is entropy conservative in the sense of Definition 6.1, that the numerical interface flux is entropy stable in the sense of Definition 6.2. The discretization given by (6.26), or, equivalently, by (5.23) with $\underline{r}^{(h,\kappa,e)}(t)$ defined as in (6.13), is then discretely entropy stable, satisfying the entropy balance*

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\sum_{\kappa=1}^{N_e}\left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{s}^{(h,\kappa)}(t) \leq \sum_{\Gamma^{(\kappa,\zeta)}\subset\partial\Omega}\Bigg(&\sum_{m=1}^{d}\left(\underline{1}^{(N_{qf}^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa)}\underline{\underline{N}}^{(\kappa,\zeta,m)}\underline{\psi}^{(\kappa,\zeta,m)}(t)\\ &-\sum_{e=1}^{N_c}\left(\underline{w}^{(h,\kappa,e)}(t)\right)^{\mathrm{T}}\left(\underline{\underline{R}}^{(\zeta)}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t)\Bigg),\end{aligned} \tag{6.33}$$

where we define

$$\underline{s}^{(h,\kappa)}(t) \coloneqq \begin{bmatrix}\mathcal{S}(\underline{u}_1^{(h,\kappa)}(t))\\ \vdots\\ \mathcal{S}(\underline{u}_{N_q}^{(h,\kappa)}(t))\end{bmatrix}, \quad \underline{\psi}^{(\kappa,\zeta,m)}(t) \coloneqq \begin{bmatrix}\Psi_m(\underline{w}_1^{(h,\kappa,\zeta)}(t))\\ \vdots\\ \Psi_m(\underline{w}_{N_{qf}^{(\zeta)}}^{(h,\kappa,\zeta)}(t))\end{bmatrix}. \tag{6.34}$$

Moreover, the entropy balance in (6.33) holds as an equality when the numerical interface flux is entropy conservative.

Proof. The proof of entropy conservation for a non-dissipative interface flux is identical to that of [133, Theorem 2], wherein we left-multiply (6.26) by $(\underline{w}^{(h,\kappa,e)}(t))^{\mathrm{T}}$, sum over $e \in \{1 : N_c\}$, and use (6.32) as well as (6.24), which follow from Lemmas 6.2 and 6.3, respectively, to obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{s}^{(h,\kappa)}(t) = &\sum_{m=1}^{d}\left(\underline{1}^{(N_{qf}^{(\zeta)})}\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa)}\underline{\underline{N}}^{(\kappa,\zeta,m)}\underline{\psi}^{(\kappa,\zeta,m)}(t)\\ &-\sum_{e=1}^{N_c}\sum_{\zeta=1}^{N_f}\left(\underline{w}^{(h,\kappa,\zeta,e)}(t)\right)^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}\underline{f}^{(*,\kappa,\zeta,e)}(t).\end{aligned} \tag{6.35}$$

For an entropy-conservative interface flux, summing (6.35) over all elements and splitting

the contributions arising from each interior interface between the two elements sharing such an interface results in a global statement of entropy conservation, corresponding to (6.33) being satisfied as an equality. The entropy inequality for an entropy-stable interface flux then follows, for example, from the analysis in [98, Theorems 3.4 and 4.3]. □

Remark 6.2. When periodic boundary conditions are imposed in all directions, the boundary contributions on the right-hand side of (6.33) vanish similarly to the interior interface contributions for an entropy-conservative interface flux. For an entropy-stable interface flux, we then obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_{\kappa=1}^{N_e}\left(\underline{1}^{(N_q)}\right)^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\underline{s}^{(h,\kappa)}(t) \leq 0, \tag{6.36}$$

where, recalling the differing sign conventions, a suitable choice of mathematical entropy results in a non-decreasing physical entropy, consistent with the second law of thermodynamics.

6.3 Efficient implementation

In this section, we discuss and analyze several important algorithmic considerations pertaining to the implementation of the proposed schemes, particularly regarding techniques for exploiting the sparsity and tensor-product structure of the SBP operators described in Chapter 4 within the context of an entropy-stable flux-differencing DSEM.

6.3.1 Exploiting operator sparsity in the flux-differencing volume terms

As discussed, for example, by Ranocha *et al.* [143, Figure 3], the cost of an entropy-stable scheme is dominated by the flux-differencing terms, for which the primary expense is the evaluation of two-point entropy-conservative flux functions between pairs of quadrature nodes. By rewriting the volume contributions appearing in the first term on the right-hand side of (6.13) or (6.15) as

$$\begin{aligned}&\left[\left(2\underline{\underline{S}}^{(l)}\odot\sum_{m=1}^{d}\{\{G_{lm}^{(\kappa)}\}\}\odot\underline{\underline{F}}^{(\kappa,m,e)}(t)\right)\underline{1}^{(N_q)}\right]_i\\ &\quad=\sum_{j=1}^{N_q}S_{ij}^{(l)}F_e^{\#}\left(\underline{\mathcal{U}}(\underline{w}_i^{(h,\kappa)}(t)),\,\underline{\mathcal{U}}(\underline{w}_j^{(h,\kappa)}(t)),\,\{\{2\boldsymbol{g}^{(\kappa,l)}\}\}_{ij}\right),\end{aligned} \tag{6.37}$$

we observe, as noted in [143, Section 2.2], that due to the symmetry of $\{\{G_{lm}^{(\kappa)}\}\}\odot\underline{\underline{F}}^{(\kappa,m,e)}(t)$ and the skew-symmetry of $\underline{\underline{S}}^{(l)}$, it is only necessary to iterate over the indices i and j corresponding to the strictly upper-triangular parts of such matrices. Moreover, the sum need only be taken over the indices for which $S_{ij}^{(l)}\neq 0$, and the corresponding values of the

vector

$$\{\{2\boldsymbol{g}^{(\kappa,l)}\}\}_{ij} \coloneqq \left[G_{l1}^{(\kappa)}(\boldsymbol{\xi}^{(i)}) + G_{l1}^{(\kappa)}(\boldsymbol{\xi}^{(j)}), \ldots, G_{ld}^{(\kappa)}(\boldsymbol{\xi}^{(i)}) + G_{ld}^{(\kappa)}(\boldsymbol{\xi}^{(j)})\right]^{\mathrm{T}} \tag{6.38}$$

within the directional two-point flux can be computed on the fly in order to avoid storing the dense matrices $\{\{G_{lm}^{(\kappa)}\}\}$ in memory. Computing the sum on the right-hand side of (6.37) for all $i \in \{1 : N_q\}$ therefore requires the evaluation of the two-point flux function, which for entropy-stable schemes is typically a relatively expensive operation involving the logarithmic mean, once per nonzero entry in the strictly upper-triangular part of $\underline{\underline{S}}^{(l)}$. Recalling from [83] and [84] that the minimum number of volume quadrature nodes for an SBP operator of degree q is the dimension of the associated total-degree polynomial space, which scales as $\mathcal{O}(q^d)$, the number of two-point fluxes, and hence the work required to evaluate the flux-differencing volume terms, is therefore expected to scale as $\mathcal{O}(q^{2d})$ when $\underline{\underline{S}}^{(l)}$ is dense, which, to the author's knowledge, is the case for all high-order SBP operators on triangles or tetrahedra proposed prior to this work. By contrast, such matrices are sparse for the operators described in Chapter 4, with their one-dimensional coupling along lines of nodes resulting in the same $\mathcal{O}(q^{d+1})$ complexity as for tensor-product DSEMs on quadrilaterals or hexahedra.

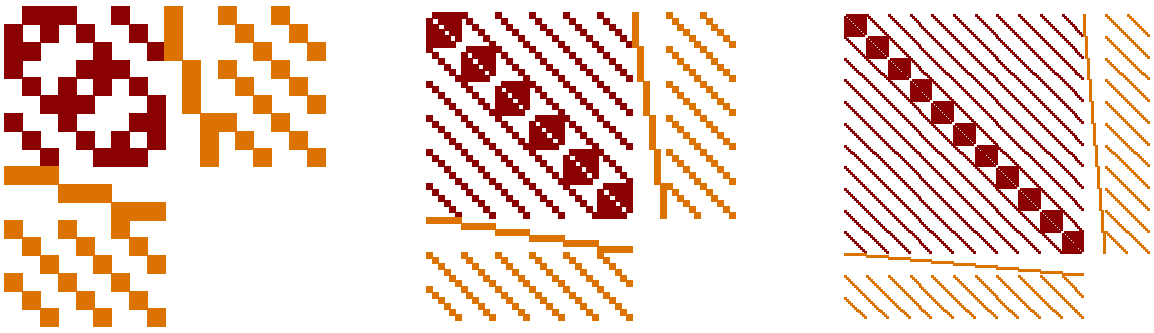
6.3.2 Exploiting operator sparsity in the flux-differencing facet correction

The second line of (6.13) requires the evaluation of correction terms of the form $\underline{\underline{C}}^{(\kappa,\zeta,e)}\underline{1}^{(N_{q_f}^{(\zeta)})}$ and $(\underline{\underline{C}}^{(\kappa,\zeta,e)})^{\mathrm{T}}\underline{1}^{(N_q)}$ for each facet. While the simplified formulation for diagonal-E operators in (6.15) does not require such corrections, diagonal-E operators on triangles and tetrahedra require quadrature rules using a much larger number of nodes for a given degree than would otherwise be needed, and are currently only available for modest polynomial degrees.[3] Similarly to how the algorithms described in the previous subsection exploit the sparsity of $\underline{\underline{S}}^{(l)}$, we can compute such terms in a manner that exploits the sparsity of the operators $(\underline{\underline{R}}^{(\zeta)})^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}$. To do this, we express the averaged metric terms in (6.10b) as

$$\{\{J^{(\kappa,\zeta)}\boldsymbol{n}^{(\kappa,\zeta)}\}\}_{ij} \coloneqq \left[\{\{J^{(\kappa,\zeta)}n_1^{(\kappa,\zeta)}\}\}_{ij}, \ldots, \{\{J^{(\kappa,\zeta)}n_d^{(\kappa,\zeta)}\}\}_{ij}\right]^{\mathrm{T}}, \tag{6.39}$$

and evaluate the contributions from $\underline{\underline{C}}^{(\kappa,\zeta,e)}(t)\underline{1}^{(N_{q_f}^{(\zeta)})}$ and $(\underline{\underline{C}}^{(\kappa,\zeta,e)}(t))^{\mathrm{T}}\underline{1}^{(N_q)}$ on the second line of (6.13) simultaneously by initializing both such vectors to zero and then iterating over values of i and j such that $[(\underline{\underline{R}}^{(\zeta)})^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}]_{ij} \neq 0$. Within each iteration, we compute the corresponding two-point flux and multiply each component by the corresponding nonzero

[3] To the author's knowledge, the highest-order diagonal-E SBP operators on tetrahedra are those of Worku *et al.* [113], who provide quadrature rules of up to degree 10 suitable for SBP operators of up to degree 5.



(a) $q = 2$ **(b)** $q = 5$ **(c)** $q = 10$

Figure 6.1: Sparsity patterns for skew-symmetric hybridized tensor-product operators on the triangle; upper-left blocks indicate coupling between pairs of volume quadrature nodes, while off-diagonal blocks indicate coupling between volume and facet quadrature nodes

matrix entry to obtain

$$C_{ij}^{(\kappa,\zeta,e)}(t) = \left[(\underline{\underline{R}}^{(\zeta)})^{\mathrm{T}}\underline{\underline{B}}^{(\zeta)}\right]_{ij} F_e^{\#}\Big(\underline{\mathcal{U}}(\underline{w}_i^{(h,\kappa)}(t)), \underline{\mathcal{U}}(\underline{w}_j^{(h,\kappa,\zeta)}(t)), \{\{J^{(\kappa,\zeta)}\boldsymbol{n}^{(\kappa,\zeta)}\}\}_{ij}\Big), \tag{6.40}$$

which we accumulate within $[\underline{\underline{C}}^{(\kappa,\zeta,e)}(t)\underline{1}^{(N_{q_f}^{(\zeta)})}]_i$ and $[(\underline{\underline{C}}^{(\kappa,\zeta,e)}(t))^{\mathrm{T}}\underline{1}^{(N_q)}]_j$ as

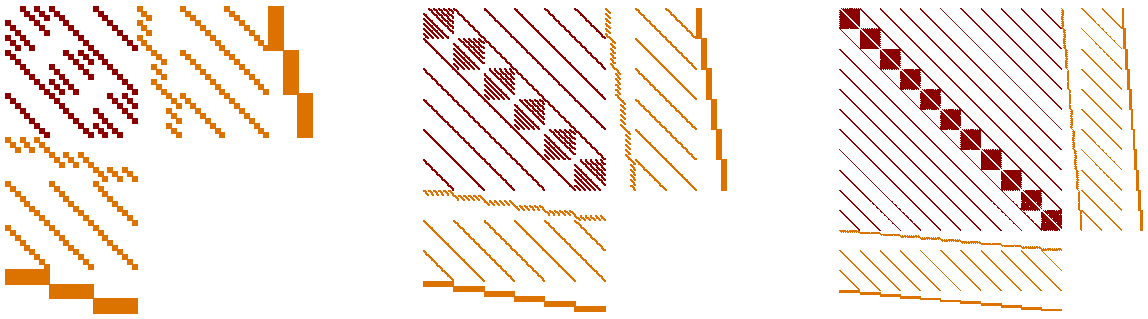
$$\left[\underline{\underline{C}}^{(\kappa,\zeta,e)}(t)\underline{1}^{(N_{q_f}^{(\zeta)})}\right]_i \leftarrow \left[\underline{\underline{C}}^{(\kappa,\zeta,e)}(t)\underline{1}^{(N_{q_f}^{(\zeta)})}\right]_i + C_{ij}^{(\kappa,\zeta,e)}(t), \tag{6.41a}$$

$$\left[\left(\underline{\underline{C}}^{(\kappa,\zeta,e)}(t)\right)^{\mathrm{T}}\underline{1}^{(N_q)}\right]_j \leftarrow \left[\left(\underline{\underline{C}}^{(\kappa,\zeta,e)}(t)\right)^{\mathrm{T}}\underline{1}^{(N_q)}\right]_j + C_{ij}^{(\kappa,\zeta,e)}(t). \tag{6.41b}$$

Remark 6.3. The sparsity patterns for the hybridized operators described in Section 6.2 can be used to illustrate which pairs of quadrature nodes must be coupled via two-point flux evaluations within an entropy-stable flux-differencing formulation. In Figures 6.1 and 6.2, we visualize the sparsity patterns for skew-symmetric matrices constructed as in (6.29) with $l = d$ using the proposed tensor-product SBP operators on the reference triangle and tetrahedron, respectively, where we note that alternative choices of nodal ordering would lead to different patterns with the same number of nonzero elements. Quantitative comparisons of the required number of two-point flux evaluations for the proposed operators relative to standard multidimensional SBP operators employing symmetric quadrature rules are deferred to Section 7.3.4, although we note here as a general trend for tensor-product operators that sparsity increases with the polynomial degree and the number of spatial dimensions.

6.3.3 Exploiting sum factorization for tensor-product operators

In addition to the reduction in computational complexity of the flux-differencing terms, the algorithmic benefits of tensor-product operators discussed in Section 5.3 with respect to



(a) $q = 2$ **(b)** $q = 5$ **(c)** $q = 10$

Figure 6.2: Sparsity patterns for skew-symmetric hybridized tensor-product operators on the tetrahedron; upper-left blocks indicate coupling between pairs of volume quadrature nodes, while off-diagonal blocks indicate coupling between volume and facet quadrature nodes

the skew-symmetric scheme in (5.14) extend directly to the matrix-vector products in the proposed entropy-stable formulations, which involve $\underline{\underline{V}}$ and $\underline{\underline{R}}^{(\zeta)}$ as well as their transposes. As such, we are able to exploit the structure of such matrices through sum factorization in the evaluation of the conservative variables at the volume and facet quadrature nodes as

$$\underline{u}^{(h,\kappa,e)}(t) = \underline{\underline{V}}\,\underline{\tilde{u}}^{(h,\kappa,e)}(t), \tag{6.42a}$$

$$\underline{u}^{(h,\kappa,\zeta,e)}(t) = \underline{\underline{R}}^{(\zeta)}\underline{u}^{(h,\kappa,e)}(t), \quad \forall\, \zeta \in \{1 : N_f\}. \tag{6.42b}$$

Similarly, we use sum factorization to compute the weight-adjusted entropy projection as

$$\underline{\tilde{w}}^{(h,\kappa,e)}(t) = \underline{\underline{M}}^{-1}\underline{\underline{V}}^{\mathrm{T}}\underline{\underline{W}}\left(\underline{\underline{J}}^{(\kappa)}\right)^{-1}\underline{\underline{V}}\,\underline{\underline{M}}^{-1}\underline{\underline{V}}^{\mathrm{T}}\underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)}\begin{bmatrix}\mathcal{W}_e(\underline{u}_1^{(h,\kappa)}(t))\\ \vdots \\ \mathcal{W}_e(\underline{u}_{N_q}^{(h,\kappa)}(t))\end{bmatrix}, \tag{6.43}$$

as well as to evaluate such a projection at the volume and facet quadrature nodes as

$$\underline{w}^{(h,\kappa,e)}(t) = \underline{\underline{V}}\,\underline{\tilde{w}}^{(h,\kappa,e)}(t), \tag{6.44a}$$

$$\underline{w}^{(h,\kappa,\zeta,e)}(t) = \underline{\underline{R}}^{(\zeta)}\underline{w}^{(h,\kappa,e)}(t), \quad \forall\, \zeta \in \{1 : N_f\}, \tag{6.44b}$$

where we also exploit the fact that $\underline{\underline{M}}^{-1}$ is the identity matrix due to the PKD basis remaining orthonormal under all quadrature rules considered in this work. Furthermore, we employ sum factorization when applying $(\underline{\underline{R}}^{(\zeta)})^{\mathrm{T}}$ to obtain $\underline{r}^{(h,\kappa,e)}(t)$ in (6.13), and, finally, in the evaluation of the time derivative in (5.23) using the weight-adjusted inverse as in (5.45). Such an approach results in the entire algorithm for computing the local time derivative requiring $\mathcal{O}(p^{d+1})$ floating-point operations under the standard assumption that q scales as $\mathcal{O}(p)$ with the polynomial degree p of the modal expansion. To the author's knowledge, this is not

achieved by any prior entropy-stable method on triangles or tetrahedra, for which the required dense matrix operations are of $\mathcal{O}(p^{2d})$ complexity. Moreover, by avoiding the construction of physical operator matrices and averaging the metric terms on the fly, memory usage is minimized, with per-element memory requirements scaling as $\mathcal{O}(p^d)$ due to the only necessary storage being geometric information at the quadrature nodes, and, optionally, precomputed diagonal entries of the matrices $\underline{\underline{W}}(\underline{\underline{J}}^{(\kappa)})^{-1}$, $\underline{\underline{W}}\underline{\underline{J}}^{(\kappa)}$, and $\underline{\underline{B}}^{(\zeta)}\underline{\underline{J}}^{(\kappa,\zeta)}$.

6.4 Chapter summary

This chapter extends the energy-stable methods on curved triangular and tetrahedral unstructured grids described in Chapter 5 to entropy-stable formulations for nonlinear hyperbolic systems of conservation laws. The schemes described in this chapter employ the tensor-product SBP operators in collapsed coordinates introduced in Chapter 4 within a weight-adjusted modal flux-differencing formulation, which is shown to be free-stream preserving, locally and globally conservative, and entropy conservative or entropy stable for suitable choices of numerical interface flux. We also describe techniques for the efficient implementation of the proposed schemes through the exploitation of operator sparsity to reduce the number of required entropy-conservative two-point flux evaluations as well as the use of sum factorization for efficient tensor-product operator evaluation, resulting in algorithms which are equivalent in time and memory complexity to comparable schemes on quadrilaterals and hexahedra.

Chapter 7

Numerical experiments

In this chapter, which is based on portions of Papers III and IV, the energy-stable schemes presented in Chapter 5 as well as the entropy-stable schemes presented in Chapter 6 are applied to linear and nonlinear model problems, respectively, on curved simplicial meshes.

7.1 Simulation setup

The proposed schemes are implemented within `StableSpectralElements.jl`, an open-source solver for conservation laws developed by the author of this thesis.[1] Written in Julia, the implementation in `StableSpectralElements.jl` closely parallels the present mathematical framework for semi-discrete DSEM formulations and interfaces with `OrdinaryDiffEq.jl`, which implements a wide variety of time-marching methods for ODEs [147]. A unified matrix notation is used to represent linear operators, with the actual strategy for evaluating the action of such operators on vectors (e.g. using matrix-free sum-factorization techniques to apply tensor-product operators or applying explicitly formed matrix operators on the reference or physical element) dispatched based on the chosen operator and algorithm subtypes. Such a software design lends itself naturally to Julia's "just ahead of time" compilation and multiple-dispatch paradigms, which provide a high level of flexibility without substantial performance compromises relative to traditional compiled languages [102].

7.1.1 Tensor-product and multidimensional SBP operators

Letting p denote the SBP operator degree, which we take to be equal to the degree of the polynomial expansion employed in the case of a modal formulation, the tensor-product SBP operators which we construct on the triangle employ LG quadrature rules with $p+1$ nodes for integration with respect to η_1, η_2, and η_f, corresponding to $q_1 = q_2 = q_f = p$, whereas those

[1] `StableSpectralElements.jl` is a registered Julia package and is available under the GNU General Public License at https://github.com/tristanmontoya/StableSpectralElements.jl.

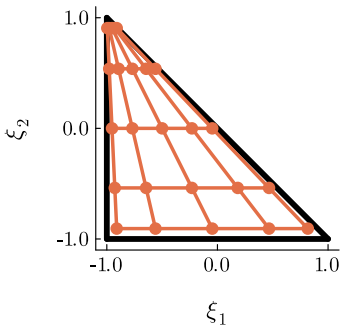
on the tetrahedron employ LG quadrature rules for η_1 and η_2 alongside a JG quadrature rule with $(a_3, b_3) = (1, 0)$ for η_3, with $p+1$ nodes in each direction (i.e. taking $q_1 = q_2 = q_3 = p$). The facet quadrature on the tetrahedron consists of an LG rule in the η_{f1} direction and a JG rule with $(a_{f2}, b_{f2}) = (1, 0)$ in the η_{f2} direction, where we use $p+1$ nodes in each direction, corresponding to $q_{f1} = q_{f2} = p$. As such quadrature rules satisfy the conditions of Theorems 4.1 and 4.2, valid SBP operators are obtained for all polynomial degrees.

To provide a point of comparison for the proposed tensor-product discretizations, we construct multidimensional SBP operators using symmetric quadrature rules with positive weights on the reference element following the approach proposed in [99, Lemma 1], which is also detailed in Section 3.1.2. Specifically, for SBP operators of degree p on the triangle, we use degree $2p$ Xiao–Gimbutas quadrature rules [148] for volume integration and degree $2p+1$ LG quadrature rules for facet integration. On the tetrahedron, we use degree $2p$ Jaśkowiec–Sukumar quadrature rules [149] for volume integration and degree $2p$ triangular quadrature rules from [148] for facet integration. The resulting operators are henceforth denoted as *multidimensional* to distinguish them from the *tensor-product* operators introduced in this work, and are available for degrees $p \leq 25$ on the triangle and $p \leq 10$ on the tetrahedron. These are, to the author's knowledge, the highest-order SBP operators on triangles and tetrahedra constructed to date using symmetric quadrature rules, and therefore facilitate comparisons with the proposed tensor-product approach over a wide range of polynomial degrees. Examples of volume quadrature nodes used to construct SBP operators employed for the simulations in this chapter are pictured in Figure 7.1.

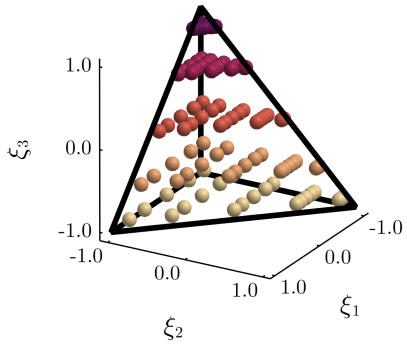
7.1.2 Curvilinear mesh generation

The problems considered in this work are defined on the spatial domain $\Omega := (0, L)^d$, where $L \in \mathbb{R}^+$ and $d \in \{2, 3\}$. The meshes are generated by beginning with a regular Cartesian grid with M edges in each direction and splitting each quadrilateral into two triangles or each hexahedron into six tetrahedra, resulting in $N_e = 2M^2$ in two dimensions and $N_e = 6M^3$ in three dimensions. We use the second algorithm described in [57, Section 2.3], which originally appeared in [140], to orient the local coordinate systems such that Assumption 3.3 is satisfied for tetrahedral meshes. The mapping nodes on the reference triangle and tetrahedron are obtained using the interpolatory warp-and-blend procedure from [123], and an affine transformation is used to obtain the positions of the mapping nodes on each element of the split Cartesian mesh. Following Chan *et al.* [150, Section 5], the mapping nodes are then perturbed as

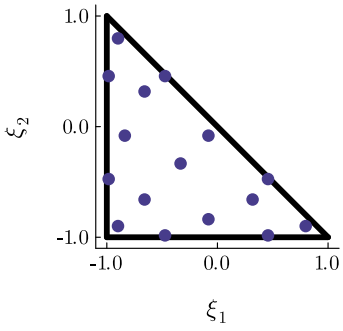
$$\begin{aligned} \tilde{x}_1 &\leftarrow x_1 + \varepsilon L \cos\Big(\tfrac{\pi}{L}\Big(x_1 - \tfrac{1}{2}\Big)\Big) \cos\Big(\tfrac{3\pi}{L}\Big(x_2 - \tfrac{1}{2}\Big)\Big), \\ \tilde{x}_2 &\leftarrow x_2 + \varepsilon L \sin\Big(\tfrac{4\pi}{L}\Big(\tilde{x}_1 - \tfrac{1}{2}\Big)\Big) \cos\Big(\tfrac{\pi}{L}\Big(x_2 - \tfrac{1}{2}\Big)\Big), \end{aligned} \tag{7.1}$$



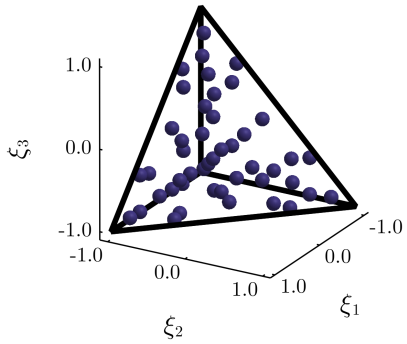
(a) Tensor-product quadrature nodes on the triangle



(b) Tensor-product quadrature nodes on the tetrahedron



(c) Multidimensional quadrature nodes on the triangle



(d) Multidimensional quadrature nodes on the tetrahedron

Figure 7.1: Volume quadrature nodes for SBP operators on the triangle and tetrahedron with $p = 4$

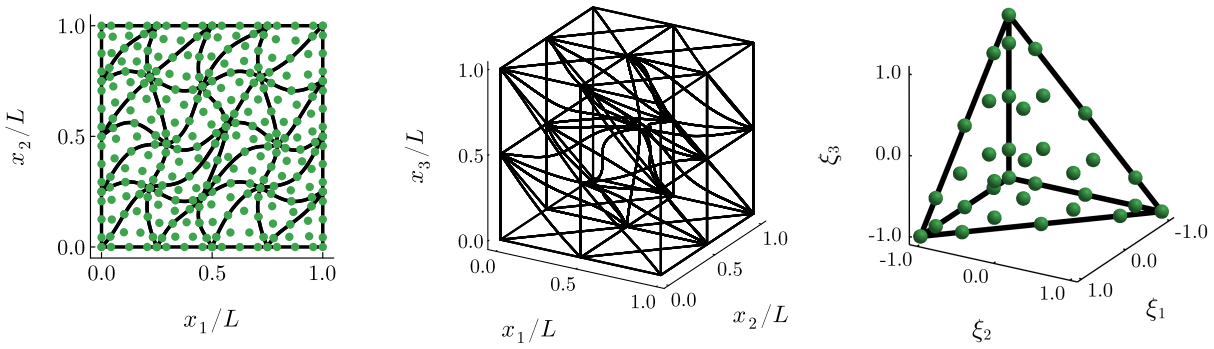
in two dimensions, and as

$$\begin{aligned}
\tilde{x}_2 &\leftarrow x_2 + \varepsilon L \cos\left(\tfrac{3\pi}{L}\left(x_1 - \tfrac{1}{2}\right)\right) \cos\left(\tfrac{\pi}{L}\left(x_2 - \tfrac{1}{2}\right)\right) \cos\left(\tfrac{\pi}{L}\left(x_3 - \tfrac{1}{2}\right)\right), \\
\tilde{x}_1 &\leftarrow x_1 + \varepsilon L \cos\left(\tfrac{\pi}{L}\left(x_1 - \tfrac{1}{2}\right)\right) \sin\left(\tfrac{4\pi}{L}\left(\tilde{x}_2 - \tfrac{1}{2}\right)\right) \cos\left(\tfrac{\pi}{L}\left(x_3 - \tfrac{1}{2}\right)\right), \\
\tilde{x}_3 &\leftarrow x_3 + \varepsilon L \cos\left(\tfrac{\pi}{L}\left(\tilde{x}_1 - \tfrac{1}{2}\right)\right) \cos\left(\tfrac{2\pi}{L}\left(\tilde{x}_2 - \tfrac{1}{2}\right)\right) \cos\left(\tfrac{\pi}{L}\left(x_3 - \tfrac{1}{2}\right)\right),
\end{aligned} \tag{7.2}$$

in three dimensions, where we take $\varepsilon = 1/16$ in both cases, and we note that such an asymmetric warping is used to ensure that the results are not systematically biased by the regularity of the mesh. The new node positions $\tilde{\boldsymbol{x}}$ are then used to define the curvilinear mapping in (5.1). Finally, the metric terms are computed using the approach described in Section 5.1.1, where in the three-dimensional case we use the conservative curl formulation, with the nodes used for the interpolation in (5.2) again obtained as in [123]. Examples of curvilinear meshes and the mapping nodes used to obtain such meshes are shown in Figure 7.2.

7.2 Linear advection equation

In this section, we solve the linear advection equation given by (3.53) on the domain $\Omega := (0, 1)^d$, with an advection velocity of $\boldsymbol{a} := [1, 1]^{\mathrm{T}}$ in two dimensions and $\boldsymbol{a} := [1, 1, 1]^{\mathrm{T}}$ in



(a) Triangular mesh **(b)** Tetrahedral mesh **(c)** Tetrahedral mapping nodes

Figure 7.2: Examples of warped meshes and mapping nodes for $p_g = 4$

three dimensions. Periodic boundary conditions are imposed in all directions, and the initial condition is given by

$$U^0(\boldsymbol{x}) \coloneqq \prod_{m=1}^{d} \sin(2\pi x_m). \tag{7.3}$$

Considering skew-symmetric nodal formulations in the form of (5.13a) as well as weight-adjusted modal formulations in the form of (5.23) using the tensor-product and multidimensional SBP operators on triangles and tetrahedra described in Section 7.1.1, we construct the mesh as in Section 7.1.2 using an isoparametric mapping, corresponding to $p_g = p$. The systems of ODEs resulting from the proposed spatial discretizations are then integrated in time until $T = 1$ using the five-stage, fourth-order explicit low-storage Runge-Kutta method of Carpenter and Kennedy [151], with the time step taken to be sufficiently small for the error due to the temporal discretization to be dominated by that due to the spatial discretization.

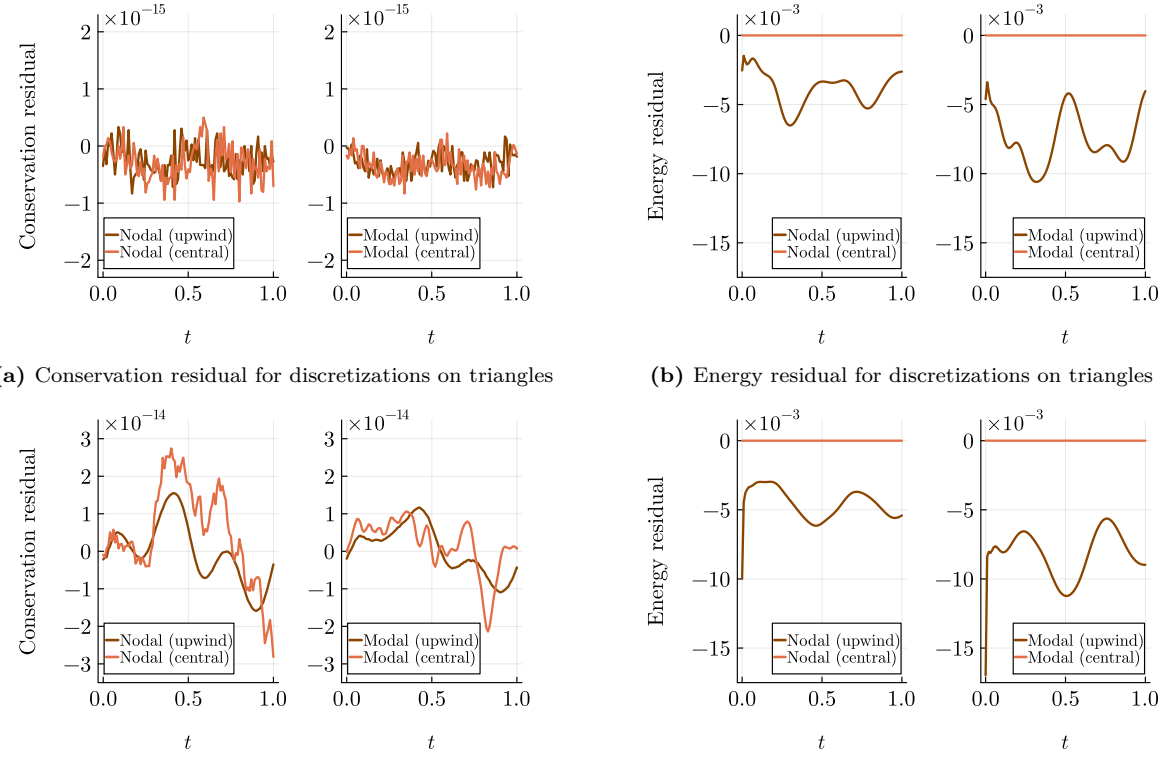
7.2.1 Conservation and energy stability

We plot the time evolution of the conservation and energy residuals in Figure 7.3 for the proposed tensor-product formulations on triangles and tetrahedra, where we present results for $p = 4$ and $M = 2$ as an illustrative example. Such quantities are given by

$$\text{Conservation residual} \coloneqq \sum_{\kappa=1}^{N_e} \left(\underline{1}^{(N_q)}\right)^{\mathrm{T}} \underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)} \frac{\mathrm{d}\underline{u}^{(h,\kappa)}(t)}{\mathrm{d}t} \tag{7.4}$$

and

$$\text{Energy residual} \coloneqq \begin{cases} \displaystyle\sum_{\kappa=1}^{N_e} \left(\underline{u}^{(h,\kappa)}(t)\right)^{\mathrm{T}} \underline{\underline{W}}\,\underline{\underline{J}}^{(\kappa)} \frac{\mathrm{d}\underline{u}^{(h,\kappa)}(t)}{\mathrm{d}t} & \text{(nodal)} \\ \displaystyle\sum_{\kappa=1}^{N_e} \left(\underline{\tilde{u}}^{(h,\kappa)}(t)\right)^{\mathrm{T}} \underline{\underline{\tilde{M}}}^{(\kappa)} \frac{\mathrm{d}\underline{\tilde{u}}^{(h,\kappa)}(t)}{\mathrm{d}t} & \text{(modal)} \end{cases}, \tag{7.5}$$



(a) Conservation residual for discretizations on triangles

(b) Energy residual for discretizations on triangles

(c) Conservation residual for discretizations on tetrahedra

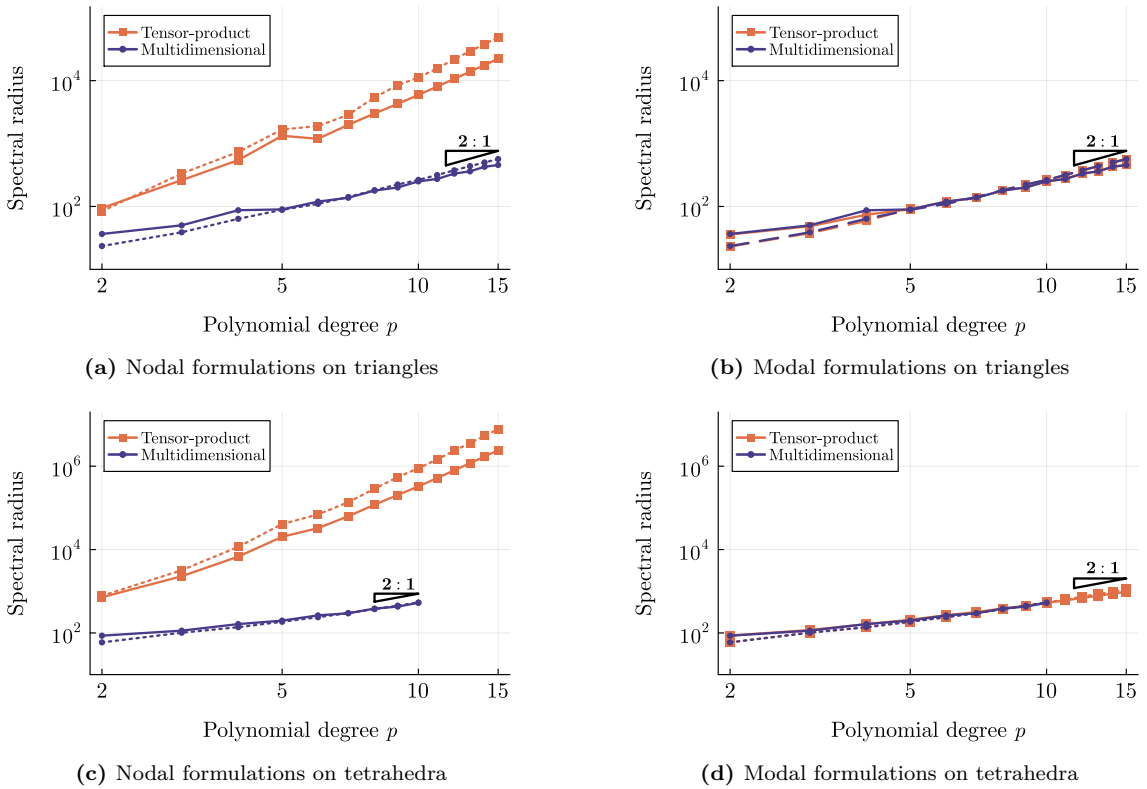
(d) Energy residual for discretizations on tetrahedra

Figure 7.3: Time evolution of the conservation and energy residuals for skew-symmetric tensor-product discretizations of the linear advection equation on triangles and tetrahedra with $p = 4$ and $M = 2$

corresponding to the time derivative of the discretely integrated numerical solution and that of the discrete solution energy, respectively. As expected for conservative and energy-stable SBP discretizations, the conservation residual remains close to machine precision for both the upwind and central variants of the nodal and modal tensor-product schemes on triangles as well as tetrahedra, while the energy residual is near machine precision for a central numerical flux and negative for an upwind numerical flux. Note that we have deliberately used coarse meshes for such tests in order to demonstrate that the conservation and energy stability properties established in Theorems 5.2 to 5.4 are satisfied discretely (up to roundoff error) at finite resolution, rather than only in the limit of mesh refinement.

7.2.2 Spectral radius

While the proposed spatial discretizations of the linear advection equation are provably energy stable in a semi-discrete sense, the stability of the fully discrete problem requires the spectrum of the semi-discrete operator to lie within the stability region of the chosen time-marching method. Thus, for explicit time integration, the maximum stable time step size is dictated by the spectral radius of the global semi-discrete operator. Figure 7.4 illustrates the effect of



(a) Nodal formulations on triangles

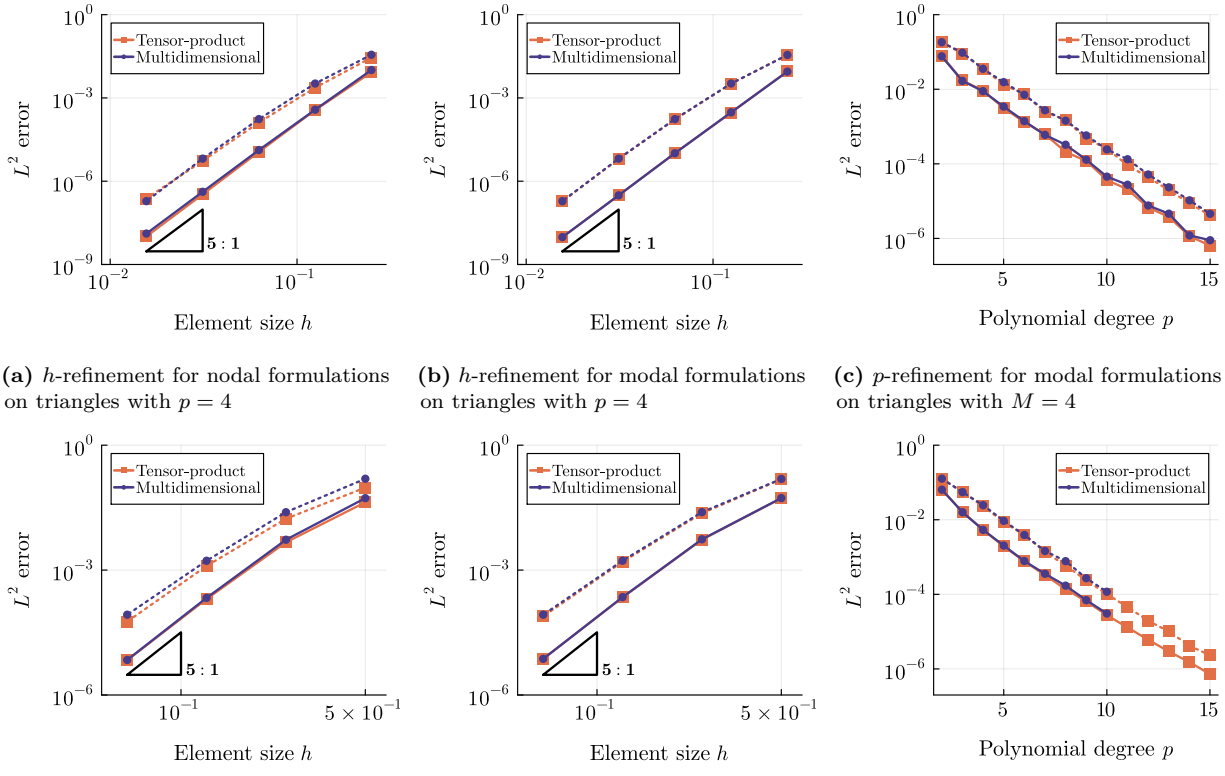
(b) Modal formulations on triangles

(c) Nodal formulations on tetrahedra

(d) Modal formulations on tetrahedra

Figure 7.4: Variation in spectral radius of the semi-discrete advection operator with polynomial degree for skew-symmetric discretizations on triangles and tetrahedra with $M = 2$; solid and dashed lines denote upwind and central numerical fluxes, respectively

varying the polynomial degree on the spectral radius of the semi-discrete advection operator for nodal and modal formulations using tensor-product as well as multidimensional SBP operators on triangles and tetrahedra, with the number of edges in each direction held fixed at $M = 2$. Considering the nodal and modal multidimensional schemes, the spectral radius grows roughly quadratically with the polynomial degree in the case of a central flux, with slightly slower growth observed for an upwind flux. The spectral radii as well as their growth rates with respect to the polynomial degree are much larger for the nodal tensor-product schemes than for all other methods, which results in a severe restriction on the time step when such schemes are used with explicit temporal integration. This limitation resulting from the concentration of resolution near the singularity of the collapsed coordinate transformation is remedied through the use of the modal formulation, in which the nodal time derivative is projected onto a standard total-degree polynomial space, resulting in spectral radii which are similar to those of the multidimensional schemes. Such behaviour is consistent with the literature (see, for example, [110, Section 6.3]) and favours the use of the modal formulation at higher polynomial degrees, at least for explicit schemes.



(a) h-refinement for nodal formulations on triangles with $p = 4$

(b) h-refinement for modal formulations on triangles with $p = 4$

(c) p-refinement for modal formulations on triangles with $M = 4$

(d) h-refinement for nodal formulations on tetrahedra with $p = 4$

(e) h-refinement for modal formulations on tetrahedra with $p = 4$

(f) p-refinement for modal formulations on tetrahedra with $M = 4$

Figure 7.5: Convergence with respect to h and p for skew-symmetric discretizations of the linear advection equation; solid and dashed lines denote upwind and central numerical fluxes, respectively

7.2.3 Accuracy

Keeping the polynomial degree fixed at $p = 4$ and successively doubling the number of edges in each direction, the discrete L^2 error is evaluated for each scheme using its associated volume quadrature rule and plotted with respect to the nominal element size, which is given by $h = L/M$. We see from Figures 7.5a, 7.5b, 7.5d, and 7.5e that all methods considered converge approximately as $\mathcal{O}(h^{p+1})$ in the case of an upwind numerical flux, with similar accuracy levels on a given mesh observed for all schemes. Recalling from Section 7.2.2 that the spectral radius for the nodal tensor-product formulation becomes extremely large at high polynomial degrees, we examine the schemes' convergence under p-refinement with $M = 4$, considering modal formulations only. As shown in Figures 7.5c and 7.5f, the modal tensor-product schemes on triangles and tetrahedra are similar in accuracy to comparable multidimensional formulations, with all schemes exhibiting exponential convergence. The reduction in accuracy observed when using a central numerical flux is well known in the context of DSEMs for hyperbolic PDEs, and, as discussed, for example, by Hu *et al.* [152] and Asthana *et al.* [153] and justified through linear eigensolution analysis, can be attributed to the failure of the central scheme to dissipate spurious solution modes contributing to excessive

dispersive error, which would otherwise be damped out by the upwind flux. Furthermore, although not proven rigorously for the particular schemes proposed in this work, *a priori* error analyses such as that presented by Brezzi *et al.* [154] for a steady linear hyperbolic problem demonstrate that, unlike its central counterpart, the standard upwind DG method admits a bound on a norm which controls the magnitudes of the inter-element jumps, resulting in an improved error estimate (we refer to the textbook by Di Pietro and Ern [155] for a detailed treatment of such estimates for DG schemes). The results of our h-refinement and p-refinement studies indicate that for this smooth model problem, the proposed skew-symmetric tensor-product discretizations on triangles and tetrahedra offer similar accuracy to comparable multidimensional schemes for a given mesh and polynomial degree, but, unlike their multidimensional counterparts, support the use of efficient sum-factorization algorithms as described in Section 5.3.

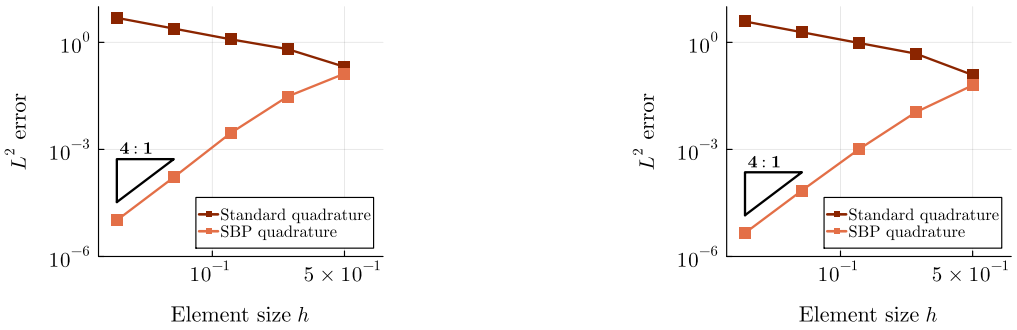
Remark 7.1. A similar numerical experiment to that described above can be used to verify the accuracy of the split-form derivative approximation obtained using (5.18). As an example, we consider a smooth vector field $\boldsymbol{F} : \Omega \to \mathbb{R}^d$ with components given by

$$F_m(\boldsymbol{x}) := -\frac{\cos(2\pi x_m)}{2\pi d} \prod_{l \in \{1:d\} \setminus m} \sin(2\pi x_l), \tag{7.6}$$

which we differentiate in split form using tensor-product operators of degree $p = 4$ on a sequence of successively refined curvilinear meshes constructed as in Section 7.1.2. At each grid level, the resulting approximation to $\nabla_{\boldsymbol{x}} \cdot \boldsymbol{F}(\boldsymbol{x})$ is compared to the exact divergence, which is equal to the initial condition given in (7.3), and the L^2 error is evaluated using the scheme's corresponding volume quadrature rule. Plotting the results of such an experiment in Figure 7.6, we observe $\mathcal{O}(h^p)$ convergence for the proposed tensor-product SBP operators on the reference triangle and tetrahedron, as expected from [97, Theorem 9]. However, a loss of convergence can be seen when differentiating in split form using tensor-product operators based on "standard" choices of Jacobi-Gauss quadrature with exponents taken as $(a_1, b_1) = (0, 0)$, $(a_2, b_2) = (1, 0)$, and $(a_3, b_3) = (2, 0)$, following [56, 57]. This loss of convergence can likely be attributed to the fact that, as discussed in Remark 4.1 and Section 4.2.2, the resulting operators do not satisfy the SBP property on the reference simplex, which is assumed to hold when deriving the truncation error estimate in [97, Theorem 9].

7.2.4 Estimated computational cost

Figure 7.7 displays the number of floating-point operations incurred in evaluating the time derivative on a single element for each scheme at varying polynomial degrees, including the evaluation of the physical flux at all volume quadrature nodes and the evaluation of the



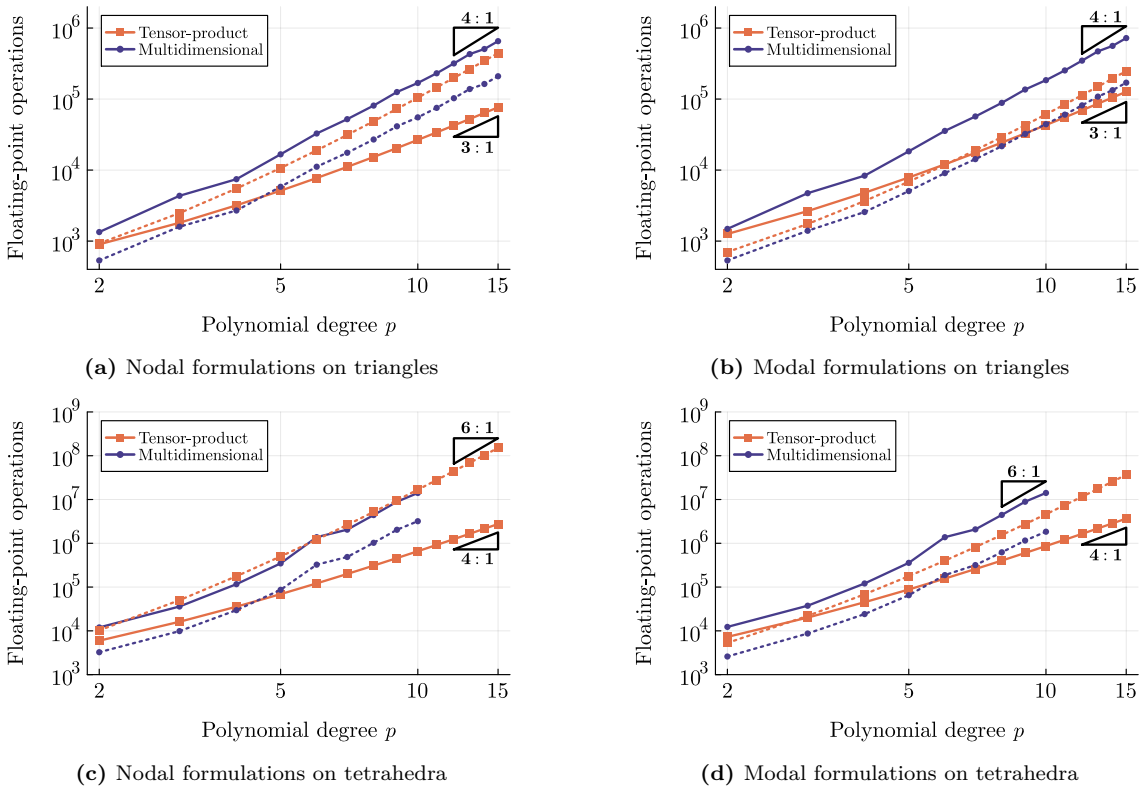
(a) Triangles **(b)** Tetrahedra

Figure 7.6: Accuracy of the split-form divergence approximation using tensor-product operators with $p = 4$

numerical flux at all facet quadrature nodes.[2] Since the results in Section 7.2.3 indicate that for a given mesh and polynomial degree, the accuracy of the proposed tensor-product approach is comparable to that of a multidimensional scheme employing a symmetric quadrature rule, such an analysis is expected to provide a fair efficiency comparison, assuming that the implementations of all methods are similarly optimized and that the resulting algorithms are compute-bound rather than memory-bound. If only the reference-operator algorithms described in Section 5.3.1 are considered (i.e. ignoring the dashed lines in Figure 7.7), we see that the sum-factorization algorithms used for the proposed tensor-product operators yield a significant reduction in operation count relative to the multidimensional operators for all orders of accuracy, scaling approximately as $\mathcal{O}(p^{d+1})$ for both the nodal and modal tensor-product formulations, in contrast to the $\mathcal{O}(p^{2d})$ scaling of the nodal and modal multidimensional formulations. If, however, we also consider the physical-operator algorithms described in Section 5.3.2, which require the precomputation and storage of operator matrices for each element, and assume that neither memory size nor latency are limiting factors, the multidimensional operators can be competitive with the proposed tensor-product operators at lower polynomial degrees. Examining the dashed lines in Figure 7.7, we observe that the physical-operator implementation using multidimensional SBP operators, while asymptotically requiring $\mathcal{O}(p^{2d})$ operations, nevertheless requires fewer floating-point operations than the $\mathcal{O}(p^{d+1})$ sum-factorization implementation of the tensor-product approach when $p \leq 4$ or $p \leq 9$ for nodal or modal formulations, respectively, on triangles, and when $p \leq 4$ or $p \leq 5$ for nodal or modal formulations, respectively, on tetrahedra.

Remark 7.2. Comparisons on the basis of floating-point operation count, while more objective than implementation-specific and hardware-specific timing comparisons, are not necessarily representative of efficiency at lower polynomial degrees, for which performance is often substantially limited by memory bandwidth. However, due to the arithmetic intensity of

[2]The operation count for each algorithm (implemented in native Julia without calls to BLAS) was evaluated using `GFlops.jl`, which is available at https://github.com/triscale-innov/GFlops.jl.



(a) Nodal formulations on triangles

(b) Modal formulations on triangles

(c) Nodal formulations on tetrahedra

(d) Modal formulations on tetrahedra

Figure 7.7: Number of floating-point operations in local time derivative evaluation for skew-symmetric discretizations of the linear advection equation; solid and dashed lines denote reference-operator and physical-operator algorithms, respectively

an SEM increasing with the polynomial degree of the discretization (see, for example, the roofline analysis in [62]), floating-point operation count indeed becomes a relevant measure at higher polynomial degrees, which is precisely where the proposed tensor-product approach shows significant cost savings. The point at which such benefits become substantial is dependent on the specifics of the implementation and hardware (e.g. considering the memory access pattern, cache size, as well as the use of single-instruction-multiple-data vectorization or multithreading) as well as the PDE and the specifics of the numerical method, and is therefore an important topic of future investigation within the context of the high-performance implementation and evaluation of the algorithms proposed in this thesis.

7.3 Euler equations

The Euler equations constitute a system of coupled nonlinear PDEs governing the conservation of mass, momentum, and energy for a compressible, inviscid, and adiabatic fluid. Such a

system takes the form of (2.1), where we define

$$\underline{U}(\boldsymbol{x},t) \coloneqq \begin{bmatrix} \rho(\boldsymbol{x},t) \\ \rho(\boldsymbol{x},t)V_1(\boldsymbol{x}) \\ \vdots \\ \rho(\boldsymbol{x},t)V_d(\boldsymbol{x}) \\ E(\boldsymbol{x},t) \end{bmatrix}, \quad \underline{F}_m(\underline{U}(\boldsymbol{x},t)) \coloneqq \begin{bmatrix} \rho(\boldsymbol{x},t)V_m(\boldsymbol{x},t) \\ \rho(\boldsymbol{x},t)V_1(\boldsymbol{x},t)V_m(\boldsymbol{x},t) + P(\boldsymbol{x},t)\delta_{1m} \\ \vdots \\ \rho(\boldsymbol{x},t)V_d(\boldsymbol{x},t)V_m(\boldsymbol{x},t) + P(\boldsymbol{x},t)\delta_{dm} \\ V_m(\boldsymbol{x},t)(E(\boldsymbol{x},t) + P(\boldsymbol{x},t)) \end{bmatrix}, \tag{7.7}$$

in terms of the density $\rho(\boldsymbol{x},t) \in \mathbb{R}$, velocity $\boldsymbol{V}(\boldsymbol{x},t) \in \mathbb{R}^d$, total energy per unit volume $E(\boldsymbol{x},t) \in \mathbb{R}$, and pressure $P(\boldsymbol{x},t) \in \mathbb{R}$. We compute the pressure in terms of the other variables using the equation of state

$$P(\boldsymbol{x},t) = (\gamma - 1)\Big(E(\boldsymbol{x},t) - \frac{1}{2}\rho(\boldsymbol{x},t)\big\|\boldsymbol{V}(\boldsymbol{x},t)\big\|^2\Big), \tag{7.8}$$

where $\gamma > 1$ is the specific heat ratio, and we have assumed that the fluid is an ideal gas with constant specific heat. In all tests considered here, we take $\gamma = 7/5$ for air. The Euler equations are hyperbolic for solutions belonging to the admissible set

$$\Upsilon \coloneqq \Big\{\underline{U}(\boldsymbol{x},t) \in \mathbb{R}^{d+2} : P(\boldsymbol{x},t), \rho(\boldsymbol{x},t) > 0\Big\}. \tag{7.9}$$

While there exist many possible entropy–entropy flux pairs for the Euler equations satisfying the conditions of Definition 2.1, we restrict our attention to the pair given by

$$\mathcal{S}(\underline{U}(\boldsymbol{x},t)) \coloneqq -\frac{\rho(\boldsymbol{x},t)}{\gamma - 1}\ln\left(\frac{P(\boldsymbol{x},t)}{\rho(\boldsymbol{x},t)^\gamma}\right), \tag{7.10a}$$

$$\boldsymbol{\mathcal{F}}(\underline{U}(\boldsymbol{x},t)) \coloneqq -\frac{\rho(\boldsymbol{x},t)\boldsymbol{V}(\boldsymbol{x},t)}{\gamma - 1}\ln\left(\frac{P(\boldsymbol{x},t)}{\rho(\boldsymbol{x},t)^\gamma}\right), \tag{7.10b}$$

which also symmetrizes the viscous terms of the compressible Navier–Stokes equations with heat conduction, and was shown by Hughes *et al.* [96] to be the unique member (up to an affine transformation) of Harten's family of entropy–entropy flux pairs [156] to do so.

7.3.1 Entropy-conservative and entropy-stable flux functions

To define an entropy-conservative two-point flux function in the sense of Definition 6.1 with respect to the entropy–entropy flux pair in (7.10), we make use of the notation

$$\{\{a\}\} \coloneqq \frac{1}{2}\big(a^- + a^+\big) \tag{7.11}$$

and

$$\{\{a\}\}_{\ln} \coloneqq \begin{cases} \frac{a^+ - a^-}{\ln(a^+) - \ln(a^-)}, & a^- \neq a^+, \\ a^-, & a^- = a^+, \end{cases} \tag{7.12}$$

for the arithmetic mean and logarithmic mean, respectively, where Taylor-series approximations from [143, Algorithms 2 and 3] are used to compute the logarithmic mean and its reciprocal when a^- and a^+ are nearly equal. The particular two-point flux employed in this work was proposed by Ranocha [157, 158], and is given by

$$\underline{F}_m^{\#}(\underline{U}^-, \underline{U}^+) \coloneqq \begin{bmatrix} \{\{\rho\}\}_{\ln}\{\{V_m\}\} \\ \{\{\rho\}\}_{\ln}\{\{V_m\}\}\{\{V_1\}\} + \{\{P\}\}\delta_{1m} \\ \vdots \\ \{\{\rho\}\}_{\ln}\{\{V_m\}\}\{\{V_d\}\} + \{\{P\}\}\delta_{dm} \\ \frac{1}{2}\{\{\rho\}\}_{\ln}\{\{V_m\}\}(\boldsymbol{V}^- \cdot \boldsymbol{V}^+ + \frac{1}{\gamma-1}\{\{\rho/P\}\}_{\ln}^{-1}) + \frac{1}{2}(P^- V_m^+ + P^+ V_m^-) \end{bmatrix}. \tag{7.13}$$

In addition to satisfying the entropy conservation property in (6.6), Ranocha's flux is kinetic energy preserving as well as pressure equilibrium preserving (see, for example, Ranocha and Gassner [159]). The interface flux takes the form of (6.8), where we use Davis's wave speed estimate [160],

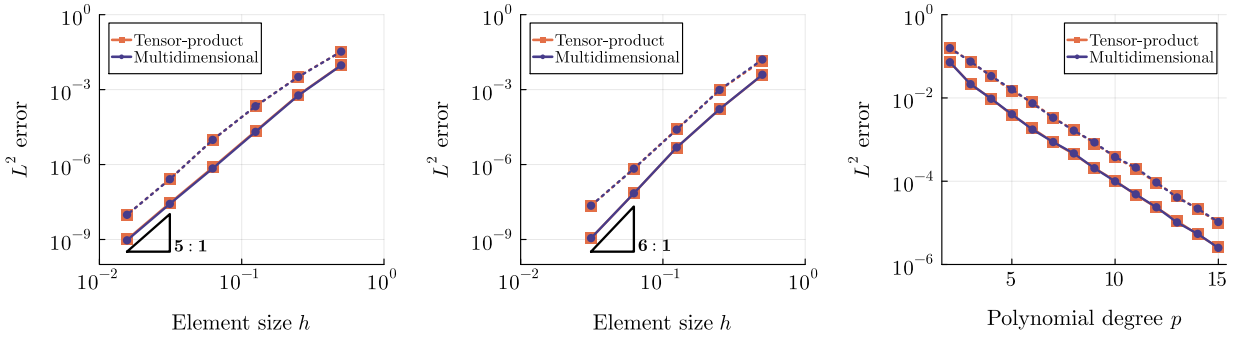
$$\Lambda(\underline{U}^-, \underline{U}^+, \boldsymbol{n}) \coloneqq \max\left(\left|\boldsymbol{V}^- \cdot \boldsymbol{n}\right|, \left|\boldsymbol{V}^+ \cdot \boldsymbol{n}\right|\right) + \max\left(\sqrt{\gamma P^-/\rho^-}, \sqrt{\gamma P^+/\rho^+}\right). \tag{7.14}$$

In all the results which follow, we employ weight-adjusted modal flux-differencing formulations, which take the form of (5.23) with $\underline{r}^{(h,\kappa,e)}(t)$ computed as in (6.13). We refer to the schemes employing local Lax–Friedrichs dissipation based on the entropy-projected conservative variables as *entropy-stable* methods. We also implement a variant without dissipation, in which the second term on the right-hand side of (6.8) is absent; such schemes are denoted as *entropy-conservative* methods.

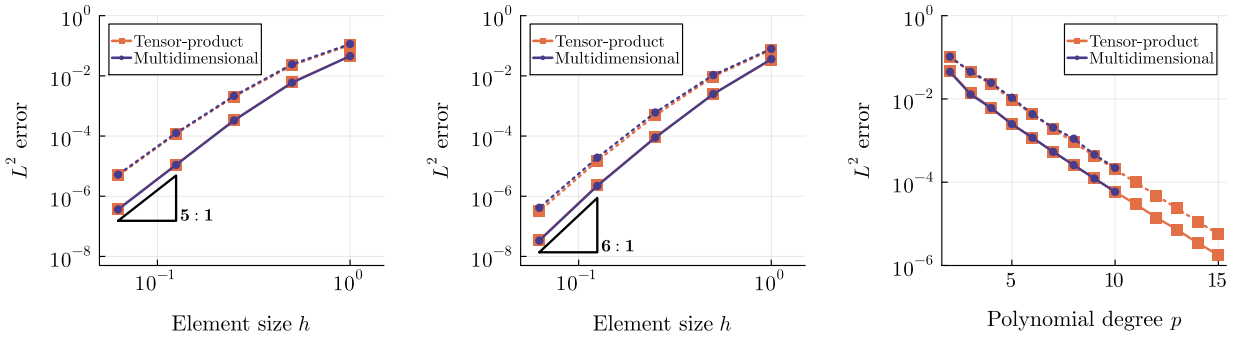
7.3.2 Accuracy

We assess the accuracy of the proposed entropy-conservative and entropy-stable discretizations of the Euler equations under refinement with respect to the nominal element size h as well as the polynomial degree p in the context of smooth problems with known analytical solutions. The initial condition in (2.1b) is prescribed as

$$\underline{U}^0(\boldsymbol{x}) \coloneqq \begin{bmatrix} \rho_0(\boldsymbol{x}) \\ \rho_0(\boldsymbol{x})\boldsymbol{V}_0(\boldsymbol{x}) \\ \frac{1}{\gamma-1}P_0(\boldsymbol{x}) + \frac{1}{2}\rho_0(\boldsymbol{x})\|\boldsymbol{V}_0(\boldsymbol{x})\|^2 \end{bmatrix} \tag{7.15}$$



(a) h-refinement on triangles with $p = 4$ **(b)** h-refinement on triangles with $p = 5$ **(c)** p-refinement on triangles with $M = 4$



(d) h-refinement on tetrahedra with $p = 4$ **(e)** h-refinement on tetrahedra with $p = 5$ **(f)** p-refinement on tetrahedra with $M = 4$

Figure 7.8: Convergence with respect to h and p for discretizations of the Euler equations; dashed and solid lines denote density error for entropy-conservative and entropy-stable schemes, respectively

in terms of the primitive variables $\rho_0(\boldsymbol{x})$, $\boldsymbol{V}_0(\boldsymbol{x})$, and $P_0(\boldsymbol{x})$. We consider the smooth density wave problem from Jiang and Shu [161, Section 7], for which the initial values of the primitive variables are given by

$$\rho_0(\boldsymbol{x}) \coloneqq 1 + \frac{1}{5}\sin\left(\frac{2\pi}{L}\sum_{m=1}^{d} x_m\right), \quad \boldsymbol{V}_0(\boldsymbol{x}) \coloneqq \begin{bmatrix}1, \dots, 1\end{bmatrix}^{\mathrm{T}}, \quad P_0(\boldsymbol{x}) \coloneqq 1, \tag{7.16}$$

on the domain $\Omega \coloneqq (0, 2)^d$, with periodic boundary conditions applied in all directions. The solution is advanced in time for one period of wave propagation (i.e. until a final time of $T = 2$) using the eighth-order Dormand–Prince algorithm described in [162, Section II.5], where, as in the previous section, the time step is taken to be sufficiently small for the temporal discretization error to be negligible in comparison to that due to the spatial discretization. The meshes for such simulations, as for all numerical experiments which we will present for in this section, are constructed as in Section 7.1.2, where we use an isoparametric mapping (i.e. taking $p_g = p$) and employ the conservative curl formulation described in Section 5.1.1 to satisfy the metric identities in the three-dimensional case.

Convergence is examined with respect to the nominal element size as well as the polynomial degree for entropy-conservative and entropy-stable DSEMs using the tensor-product and

multidimensional SBP operators on triangles and tetrahedra considered in Section 6.3. The degree p of the solution expansion in (2.20) is taken to be equal to the degree q of the SBP operators, and we report the L^2 norm of the density error, which is computed numerically using a quadrature rule of degree 35. Similar convergence behaviour was observed for the other solution variables as well as when computing the error norm as in Section 7.2.3 using each scheme's respective volume quadrature rule. Figure 7.8 demonstrates optimal $\mathcal{O}(h^{p+1})$ algebraic convergence under h-refinement for the entropy-stable discretizations (i.e. those including interface dissipation) as well as exponential convergence under p-refinement, where we recall from Section 7.1.1 that the multidimensional SBP operators on tetrahedra are only available for degrees up to ten. For a given mesh and polynomial degree, the error norms obtained for the proposed tensor-product discretizations are found to be very close to those obtained for their multidimensional counterparts, which suggests that their algorithmic advantages discussed in Section 6.3 with respect to the ability to exploit operator sparsity and sum factorization do not come at the expense of accuracy.

Remark 7.3. The numerical results are consistent with the conservation property established in Theorem 6.2, with the time derivative on the left-hand side of (5.31b) remaining close to machine precision for all solution variables at approximately 100 equispaced snapshots taken during each test. Furthermore, the rates of entropy dissipation given by the left-hand side of (6.33) are verified for the entropy-conservative and entropy-stable schemes to be zero and non-positive, respectively, at all snapshots (up to roundoff error levels close to machine precision), as is consistent with Theorem 6.4.

7.3.3 Robustness

High-order methods such as DSEMs are prone to numerical stability issues, particularly in the context of under-resolved nonlinear problems, in which high-frequency numerical modes are produced and potentially amplified by the discretization, resulting in non-physical blow-up or negative values of thermodynamic quantities such as pressure or density (i.e. corresponding to solutions outside the admissible set Υ). Such under-resolution commonly arises in simulations of turbulent fluid flow problems, which are characterized by the cascade of energy from larger eddies to those progressively smaller in size, eventually dissipating as heat due to the viscosity of the fluid (see, for example, Pope [163, Chapter 6]). As the Euler equations do not model physical viscosity, the only dissipation present is that inherent in the numerical scheme, which is typically small for a high-order method, thus exposing the potentially destabilizing effects of under-resolved eddies. We therefore use simulations of inviscid vortical flows to mimic worst-case scenarios with respect to under-resolved turbulence, where we are interested in assessing whether the simulations run to completion and whether the entropy bounds

established in Section 6.2 hold, rather than in evaluating the accuracy of such simulations.

In the two-dimensional case, we consider the Kelvin–Helmholtz instability (KHI) problem described by Rueda-Ramírez and Gassner [164] and used for robustness tests by Chan *et al.* [165], for which we define the smoothed step function

$$B(\boldsymbol{x}) \coloneqq \tanh\Big(15(x_2 - \tfrac{1}{2})\Big) - \tanh\Big(15(x_2 - \tfrac{3}{2})\Big) \tag{7.17}$$

in order to obtain the initial condition, which is given in terms of the primitive variables by

$$\rho_0(\boldsymbol{x}) \coloneqq \frac{1}{2} + \frac{3}{4}B(\boldsymbol{x}), \quad \boldsymbol{V}_0(\boldsymbol{x}) \coloneqq \begin{bmatrix} \frac{1}{2}(B(\boldsymbol{x}) - 1) \\ \frac{1}{10}\sin(2\pi x_1) \end{bmatrix}, \quad P_0(\boldsymbol{x}) \coloneqq 1, \tag{7.18}$$

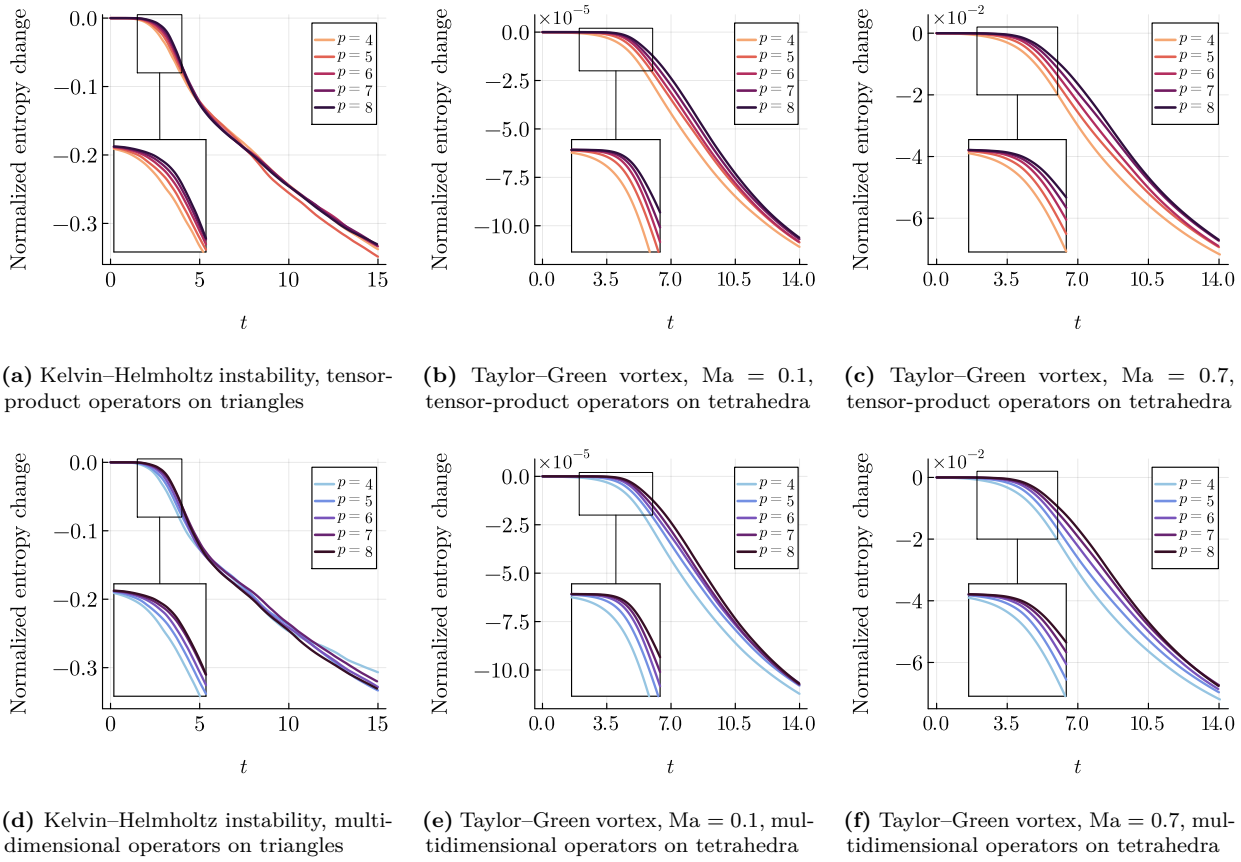
on the domain $\Omega \coloneqq (0, 2)^2$, with periodic boundary conditions in both directions. As in [165], we integrate until a final time of $T = 15$. In three dimensions, we consider an inviscid Taylor–Green vortex (TGV) problem, for which the initial condition is given on the periodic domain $\Omega \coloneqq (0, 2\pi)^3$ as

$$\begin{aligned} &\rho_0(\boldsymbol{x}) \coloneqq 1, \quad \boldsymbol{V}_0(\boldsymbol{x}) \coloneqq \Big[\sin(x_1)\cos(x_2)\cos(x_3),\ -\cos(x_1)\sin(x_2)\cos(x_3),\ 0\Big]^{\mathrm{T}}, \\ &P_0(\boldsymbol{x}) \coloneqq \frac{1}{\gamma \mathrm{Ma}^2} + \frac{1}{16}\Big(\cos(2x_1) + 2\cos(2x_2) + \cos(2x_1)\cos(2x_3) + \cos(2x_2)\cos(2x_3)\Big), \end{aligned} \tag{7.19}$$

where $\mathrm{Ma} \in \mathbb{R}^+$ is the nominal Mach number. We run the TGV simulations until a final time of $T = 14$, and, as in [166], we consider the nearly incompressible case of $\mathrm{Ma} = 0.1$ as well as the case of $\mathrm{Ma} = 0.7$, where the latter is expected to pose a greater challenge to the robustness of the proposed schemes, specifically regarding positivity preservation.

The Euler equations are solved for the above initial conditions using the proposed entropy-conservative and entropy-stable DSEMs for polynomial degrees 4 to 8, taking $M = 16$ for the KHI problem and $M = 4$ for the TGV problem. We integrate in time using the same explicit eighth-order Dormand–Prince method used for the accuracy tests. The time step is taken to be sufficiently small to ensure that the reported instances of instability result purely from the spatial discretization, and we did not find any of the reported instabilities to be remedied by decreasing the time step size. Without any additional stabilization beyond the interface dissipation provided by the numerical flux in (6.8), all simulations ran to completion for the entropy-stable schemes using tensor-product as well as multidimensional SBP operators on triangles and tetrahedra. Figure 7.9 demonstrates that the entropy is nonincreasing for all time in each of such cases, as expected from Theorem 6.4 for periodic boundary conditions.

While the entropy-conservative simulations ran to completion for the TGV with $\mathrm{Ma} = 0.1$, incurring a change in entropy close to machine precision, such computations crashed for the TGV with $\mathrm{Ma} = 0.7$ as well as for the KHI due to negative densities or pressures. This



(a) Kelvin–Helmholtz instability, tensor-product operators on triangles

(b) Taylor–Green vortex, Ma = 0.1, tensor-product operators on tetrahedra

(c) Taylor–Green vortex, Ma = 0.7, tensor-product operators on tetrahedra

(d) Kelvin–Helmholtz instability, multidimensional operators on triangles

(e) Taylor–Green vortex, Ma = 0.1, multidimensional operators on tetrahedra

(f) Taylor–Green vortex, Ma = 0.7, multidimensional operators on tetrahedra

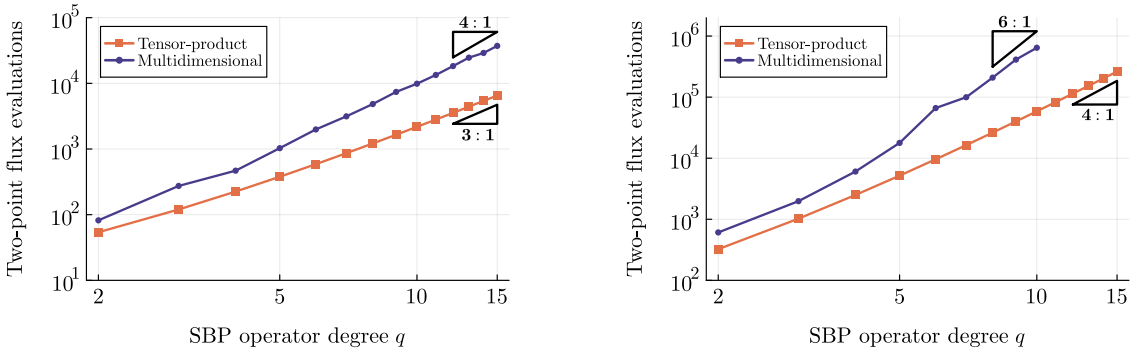
Figure 7.9: Normalized entropy change for entropy-stable discretizations of the Euler equations

reflects a well-known limitation of the entropy analysis, namely that entropy stability does not guarantee positivity of thermodynamic quantities. Although not provably positivity preserving, the entropy-stable schemes using the dissipative interface flux in (6.8) did not incur negative densities or pressures for any of the tests considered in this work, likely as a consequence of the dissipative term within the numerical flux serving to dampen any oscillations which would otherwise eventually lead to a violation of positivity. As a result, such schemes are highly robust even in the presence of substantially under-resolved solution features. These results are consistent with the observations in [165], where the authors demonstrated numerically that entropy-stable schemes which incorporate an entropy projection are often able to avoid negative densities or pressures for challenging under-resolved problems without the need for positivity-preserving limiters. We recognize, however, that such an approach may not be sufficient to preserve positivity for problems with discontinuities, and the extension of subcell limiting techniques such as those developed by Rueda-Ramírez *et al.* [167], Yamaleev and Upperman [168], and Lin *et al.* [169] to the proposed tensor-product discretizations on triangles and tetrahedra is an important topic of future research.

7.3.4 Estimated computational cost

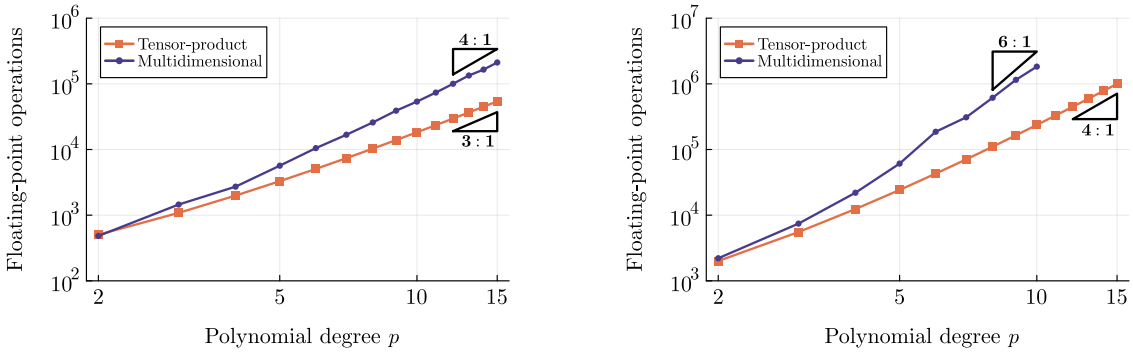
As discussed in Section 6.3 and in [143], the overall expense of an entropy-stable DSEM is dominated by that of the local flux-differencing terms in (6.13), the cost of which is proportional to the number of two-point entropy-conservative flux functions between pairs of quadrature nodes. Letting q denote the SBP operator degree, which may differ in general from the degree p of the polynomial expansion, we plot the number of required two-point flux evaluations for each class of operator in Figure 7.10. Since dense matrix storage is used in our implementation of the multidimensional operators, the only zero entries considered for such operators are those along the main diagonal of the skew-symmetric matrix $\underline{\underline{S}}^{(l)}$, although we observe numerically that a small fraction of the off-diagonal entries are, in fact, on the order of machine precision. The scaling is observed to be slightly better than the asymptotic estimates for both classes of operators, which we recall from Section 6.3 to be $\mathcal{O}(q^{d+1})$ and $\mathcal{O}(q^{2d})$ for the tensor-product and multidimensional operators, respectively, and the number of two-point flux evaluations required for the proposed tensor-product approach is smaller than that required when using multidimensional SBP operators for all polynomial degrees considered. As expected, the benefit of the tensor-product approach increases with the polynomial degree; for example, the number of two-point flux evaluations is reduced by factors of 1.56 at $q = 2$, 2.78 at $q = 5$, and 4.57 at $q = 10$ on triangles, and by factors of 1.88 at $q = 2$, 3.44 at $q = 5$, and 10.99 at $q = 10$ on tetrahedra. We also consider the total number of floating-point operations per variable incurred in evaluating the matrix-vector products in (6.13), (6.42), (6.43), and (5.45) on a given element, where, as in the numerical experiments presented earlier in this chapter, we take $p = q$ in all cases. The results of such an analysis, which are displayed in Figure 7.11, are qualitatively similar to those in Figure 7.10 for higher polynomial degrees, although the benefit of the tensor-product operators for low polynomial degrees is less substantial in this regard, requiring roughly the same number of floating-point operations as the multidimensional operators, for example, at $p = 2$.

Remark 7.4. The discussion in Remark 7.2 regarding comparisons based on operation count applies equally to the entropy-stable discretizations considered in this section. Additionally, due to their use of a larger number of volume and facet quadrature nodes than typical multidimensional operators of the same degree, the tensor-product operators require a somewhat greater number of conversions between conservative and entropy variables as well as a somewhat larger number of numerical interface flux evaluations. These operations do not, however, typically constitute the most significant contribution to the overall expense of an entropy-stable scheme and incur a cost which grows more slowly with the polynomial degree than that of the flux-differencing terms.



(a) Triangles **(b)** Tetrahedra

Figure 7.10: Number of two-point flux evaluations in local flux-differencing terms



(a) Triangles **(b)** Tetrahedra

Figure 7.11: Number of floating-point operations per variable in local operator evaluation

7.4 Chapter summary

This chapter presents numerical experiments involving the solution of the linear advection and compressible Euler equations using provably stable DSEMs on curved triangular and tetrahedral meshes employing the tensor-product SBP operators constructed in Chapter 4. The discrete conservation properties and energy or entropy estimates established theoretically in Chapters 5 and 6 are confirmed numerically, and the schemes are compared to those using standard multidimensional SBP operators based on symmetric quadrature rules in terms of their accuracy under h- and p-refinement, robustness for under-resolved nonlinear problems, spectral radius of the semi-discrete linear advection operator, and estimated computational cost. We conclude from such experiments that the proposed energy-stable and entropy-stable DSEMs using tensor-product SBP operators on triangles and tetrahedra enable a reduction in computational complexity (and, hence, a significantly reduced cost for high polynomial degrees) to be achieved on triangular and tetrahedral grids without compromising accuracy, robustness, or spectral radius relative to existing multidimensional formulations.

Chapter 8

Conclusions

This thesis presents a comprehensive approach to the formulation and analysis of provably stable DSEMs for conservation laws based on the SBP property. Such a framework enables the unified analysis of a broad class of existing DG and FR methods by recasting them as SBP schemes, through which conservation and energy stability can be proven through matrix analysis. Furthermore, the present framework facilitates the construction and analysis of a new class of DSEMs on triangles and tetrahedra which combine the geometric flexibility afforded through the use of simplicial elements with the efficiency of a tensor-product formulation and the robustness of an energy-stable or entropy-stable discretization with the SBP property. Such schemes are constructed through a novel synergy of the following technologies.

- The discretization on the reference triangle or tetrahedron uses tensor-product spectral-element operators with the SBP property in collapsed coordinates, which enable the use of sum-factorization algorithms for efficient matrix-free operator evaluation.
- A split formulation is used to obtain energy-stable schemes for the linear advection equation on curved simplicial meshes. The method is extended to entropy-stable discretizations of nonlinear hyperbolic systems through a flux-differencing approach.
- The use of a modal formulation employing a projection onto a total-degree polynomial space alleviates the explicit time step restriction resulting from the singular nature of the collapsed coordinate transformation.
- By exploiting the "warped" tensor-product structure of the modal basis functions alongside a weight-adjusted approximation to the inverse of the curvilinear mass matrix, we obtain an explicit algorithm for computing the time derivative on a given element in $\mathcal{O}(p^{d+1})$ floating-point operations, with optimal $\mathcal{O}(p^d)$ storage requirements.
- In addition to the use of sum factorization for efficient tensor-product operator evaluation, the proposed entropy-stable schemes exploit the sparsity of the proposed SBP operators to compute the local flux-differencing terms, reducing the number of required

entropy-conservative two-point flux evaluations by a factor ranging from 1.56 at $p = 2$ to 4.57 at $p = 10$ for triangles, and from 1.88 at $p = 2$ to 10.99 at $p = 10$ for tetrahedra.

While possessing a particular structure which lends itself to an efficient implementation, the proposed tensor-product operators satisfy the same algebraic properties as existing multidimensional SBP operators and are therefore amenable to the same analysis of free-stream preservation, conservation, and energy or entropy stability within the proposed theoretical framework, thus illustrating the utility of such a unified approach. Through numerical experiments, the proposed methods are shown to be at least as accurate on a given mesh as otherwise identical discretizations using multidimensional SBP operators based on symmetric quadrature rules. Furthermore, the two operator families both result in entropy-stable schemes which exhibit excellent robustness properties for challenging nonlinear problems containing under-resolved scales. The proposed tensor-product triangular and tetrahedral DSEMs thus possess very similar approximation properties to existing approaches while offering the potential for significant efficiency benefits at higher polynomial degrees. Considering this thesis in its entirety, the analysis of existing DG and FR methods as well as the development of novel energy-stable and entropy-stable tensor-product DSEMs on triangles and tetrahedra therefore serve as demonstrative examples which illustrate the utility of the SBP property as a general guiding principle for the construction and analysis of numerical methods for conservation laws.

8.1 Recommendations

There remain numerous avenues for further development of the proposed framework as well as certain limitations which should be addressed in future work. Specific recommendations of topics which merit further investigation are provided below.

- Further accuracy, efficiency, and robustness comparisons between the schemes encompassed within the present framework are required. In particular, fair efficiency comparisons should be performed between the proposed tensor-product DSEMs on triangles and tetrahedra and those using different families of multidimensional SBP operators, including diagonal-E operators such as those recently proposed in [113].
- Numerical studies should be undertaken in order to assess the influence of the techniques used to approximate the metric terms and normals on the accuracy of the resulting schemes. In particular, the conservative curl formulation from [133] which we employ in this thesis to should be compared to the optimization-based approach proposed in [97]. Additionally, the effect of approximating the outward unit normal as in (5.4) should be examined within the context of wall-bounded problems.

- Although we address only hyperbolic conservation laws in this thesis, the proposed tensor-product DSEMs on simplices could likely be applied in a straightforward manner to problems with parabolic or nonconservative terms. This could be done, for example, by adapting existing techniques suitable for entropy-stable modal formulations such as those proposed by Chan *et al.* [170] and Waruzewski *et al.* [171] in the context of the compressible Navier–Stokes equations and the Euler equations with gravity, respectively.
- The novel SBP operators developed for triangles could be extended through a tensor-product construction to triangular prisms, which are useful for boundary layer meshing as well as for three-dimensional atmospheric simulations on geodesic (e.g. icosahedral) grids. It would also be useful to develop SBP operators with a tensor-product structure on pyramids, as such elements are often used to transition between hexahedral and tetrahedral elements within a hybrid mesh.
- While the spatial discretizations developed and analyzed in this thesis are suitable for use with either explicit or implicit temporal integration, the weight-adjusted approximation of the inverse mass matrix and the comparisons of the schemes' spectral radii pertain primarily to the explicit case. In order to facilitate the use of the proposed tensor-product discretizations on simplices in conjunction with implicit time-marching methods, it would be helpful to compare the conditioning of the linear systems resulting from the nodal and modal formulations in collapsed coordinates and to develop preconditioners which exploit the tensor-product structure and sparsity of the proposed operators.
- Other approaches to the construction of tensor-product DSEMs on triangles and tetrahedra could be explored. For example, Li *et al.* [172] and Zhou *et al.* [173, 174] construct tensor-product formulations on triangular elements based on an alternative mapping from the square to the reference triangle which leads to a more uniform nodal distribution, and they describe a similar mapping from the cube to the reference tetrahedron. While it is not clear whether SBP operators can be constructed using such a mapping, the approach should be studied in further detail in order to determine its feasibility as an alternative to the proposed formulations based on collapsed coordinates.
- As discussed in Section 7.3.3, the proposed entropy-stable schemes on do not guarantee positivity of thermodynamic variables such as pressure or density. In order to apply such schemes to problems with discontinuous solutions, positivity-preserving subcell limiting techniques should be developed, potentially taking advantage of the fact that the tensor-product quadrature nodes on the triangle or tetrahedron define a grid on which a low-order scheme could be constructed and blended with the high-order DSEM in a similar manner to that employed in [167] or [168] for quadrilaterals and hexahedra.
- The convergence properties of entropy-stable high-order methods for nonlinear hyper-

bolic systems are not fully understood, and the schemes described in this work are not guaranteed to converge to weak solutions which satisfy the entropy inequality in (2.8). Moreover, as discussed by Gassner *et al.* [175], split-form and entropy-stable discretizations of nonlinear problems are not necessarily locally energy stable when linearized about an arbitrary base flow, which can result in non-physical behaviour, particularly at coarser resolutions. Further investigation is therefore required in order to better understand and potentially overcome these limitations.

References

[1] Z. J. Wang, K. Fidkowski, R. Abgrall, F. Bassi, D. Caraeni, A. Cary, H. Deconinck, R. Hartmann, K. Hillewaert, H. T. Huynh, N. Kroll, G. May, P.-O. Persson, B. van Leer, and M. Visbal, "High-order CFD methods: Current status and perspective," *International Journal for Numerical Methods in Fluids*, vol. 72, no. 8, pp. 811–845, 2013.

[2] H.-O. Kreiss and J. Oliger, "Comparison of accurate methods for the integration of hyperbolic equations," *Tellus*, vol. 24, no. 3, pp. 199–215, 1972.

[3] D. W. Zingg, "Comparison of high-accuracy finite-difference methods for linear wave propagation," *SIAM Journal on Scientific Computing*, vol. 22, no. 2, pp. 476–502, 2000.

[4] A. Klöckner, T. Warburton, J. Bridge, and J. S. Hesthaven, "Nodal discontinuous Galerkin methods on graphics processors," *Journal of Computational Physics*, vol. 228, no. 21, pp. 7863–7882, 2009.

[5] D. S. Abdi, L. C. Wilcox, T. Warburton, and F. X. Giraldo, "A GPU-accelerated continuous and discontinuous Galerkin non-hydrostatic atmospheric model," *The International Journal of High Performance Computing Applications*, vol. 33, no. 1, pp. 81–109, 2017.

[6] B. C. Vermeire, F. D. Witherden, and P. E. Vincent, "On the utility of GPU accelerated high-order methods for unsteady flow simulations: A comparison with industry-standard tools," *Journal of Computational Physics*, vol. 334, pp. 497–521, 2017.

[7] M. Parsani, R. Boukharfane, I. R. Nolasco, D. C. Del Rey Fernández, S. Zampini, B. Hadri, and L. Dalcin, "High-order accurate entropy-stable discontinuous collocated Galerkin methods with the summation-by-parts property for compressible CFD frameworks: Scalable SSDC algorithms and flow solver," *Journal of Computational Physics*, vol. 424, article no. 109844, 2021.

[8] P. Mossier, A. Beck, and C.-D. Munz, "A p-adaptive discontinuous Galerkin method with hp-shock capturing," *Journal of Scientific Computing*, vol. 91, no. 1, article no. 4, 2022.

[9] G. J. Gassner and A. D. Beck, "On the accuracy of high-order discretizations for underresolved turbulence simulations," *Theoretical and Computational Fluid Dynamics*, vol. 27, no. 3–4, pp. 221–237, 2012.

[10] G. Mengaldo, D. De Grazia, D. Moxey, P. E. Vincent, and S. J. Sherwin, "Dealiasing techniques for high-order spectral element methods on regular and irregular grids," *Journal of Computational Physics*, vol. 299, pp. 56–81, 2015.

[11] J. S. Hesthaven and T. Warburton, *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*. Springer, 2008.

[12] A. R. Winters, R. C. Moura, G. Mengaldo, G. J. Gassner, S. Walch, J. Peiro, and S. J. Sherwin, "A comparative study on polynomial dealiasing and split form discontinuous Galerkin schemes for under-resolved turbulence computations," *Journal of Computational Physics*, vol. 372, pp. 1–21, 2018.

[13] G. J. Gassner and A. R. Winters, "A novel robust strategy for discontinuous Galerkin methods in computational fluid mechanics: Why? when? what? where?" *Frontiers in Physics*, vol. 8, 2021.

[14] I. G. Bubnov, "Report on the works of Professor Timoshenko which were awarded the Zhuranskyi Prize," in *Symposium of the Institute of Communication Engineers*, vol. 81, 1913, in Russian.

[15] W. Ritz, "Über eine neue methode zur lösung gewisser variationsprobleme der mathematischen physik," *Journal für die Reine und Angewandte Mathematik*, vol. 135, no. 1, pp. 1–61, 1909.

[16] B. G. Galerkin, "Rods and plates: Series occurring in various questions concerning the elastic equilibrium of rods and plates," *Vestnik Inzhenerov i Tekhnikov*, pp. 897–908, 1915, in Russian.

[17] M. J. Gander and G. Wanner, "From Euler, Ritz, and Galerkin to modern computing," *SIAM Review*, vol. 54, no. 4, pp. 627–666, 2012.

[18] R. Courant, "Variational methods for the solution of problems of equilibrium and vibrations," *Bulletin of the American Mathematical Society*, vol. 49, no. 1, pp. 1–23, 1943.

[19] A. Hrennikoff, "Solution of problems of elasticity by the framework method," *Journal of Applied Mechanics*, vol. 8, no. 4, A169–A175, 1941.

[20] M. J. Turner, R. W. Clough, H. C. Martin, and L. J. Topp, "Stiffness and deflection analysis of complex structures," *Journal of the Aeronautical Sciences*, vol. 23, no. 9, pp. 805–823, 1956.

[21] J. H. Argyris, *Energy Theorems and Structural Analysis*. Butterworth & Co., 1960.

[22] O. C. Zienkiewicz and Y. K. Cheung, "The finite element method for analysis of elastic isotropic and orthotropic slabs," *Proceedings of the Institution of Civil Engineers*, vol. 28, no. 4, pp. 471–488, 1964.

[23] S. A. Orszag, "Numerical methods for the simulation of turbulence," *Physics of Fluids*, vol. 12, no. 12, pp. II-250–II-257, 1969.

[24] I. Babuška and M. Suri, "The *p*- and *h*-*p* versions of the finite element method: An overview," *Computer Methods in Applied Mechanics and Engineering*, vol. 80, no. 1, pp. 5–26, 1990.

[25] S. A. Orszag, "Spectral methods for problems in complex geometries," *Journal of Computational Physics*, vol. 37, no. 1, pp. 70–92, 1980.

[26] A. T. Patera, "A spectral element method for fluid dynamics: Laminar flow in a channel expansion," *Journal of Computational Physics*, vol. 54, no. 3, pp. 468–488, 1984.

[27] L. C. Young, "Orthogonal collocation revisited," *Computer Methods in Applied Mechanics and Engineering*, vol. 345, pp. 1033–1076, 2019.

[28] M. F. Wheeler, "A C^0-collocation-finite element method for two-point boundary value problems and one space dimensional parabolic problems," *SIAM Journal on Numerical Analysis*, vol. 14, no. 1, pp. 71–90, 1977.

[29] J. C. Díaz, "A collocation-Galerkin method for the two point boundary value problem using continuous piecewise polynomial spaces," *SIAM Journal on Numerical Analysis*, vol. 14, no. 5, pp. 844–858, 1977.

[30] W. H. Reed and T. R. Hill, "Triangular mesh methods for the neutron transport equation," Los Alamos Scientific Laboratory, Tech. Rep. LA-UR-73-479, 1973.

[31] B. Cockburn and C.-W. Shu, "TVB Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws II: General framework," *Mathematics of Computation*, vol. 52, no. 186, pp. 411–435, 1989.

[32] B. Cockburn, S.-Y. Lin, and C.-W. Shu, "TVB Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws III: One-dimensional systems," *Journal of Computational Physics*, vol. 84, no. 1, pp. 90–113, 1989.

[33] B. Cockburn, S. Hou, and C.-W. Shu, "The Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws. IV: The multidimensional case," *Mathematics of Computation*, vol. 54, no. 190, pp. 545–581, 1990.

[34] B. Cockburn and C.-W. Shu, "The Runge–Kutta discontinuous Galerkin method for conservation laws V: Multidimensional systems," *Journal of Computational Physics*, vol. 141, no. 2, pp. 199–224, 1998.

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

[36] D. A. Kopriva and J. H. Kolias, "A conservative staggered-grid Chebyshev multidomain method for compressible flows," *Journal of Computational Physics*, vol. 125, no. 1, pp. 244–261, 1996.

[37] Y. Liu, M. Vinokur, and Z. J. Wang, "Spectral difference method for unstructured grids I: Basic formulation," *Journal of Computational Physics*, vol. 216, no. 2, pp. 780–801, 2006.

[38] Z. J. Wang, Y. Liu, G. May, and A. Jameson, "Spectral difference method for unstructured grids II: Extension to the Euler equations," *Journal of Scientific Computing*, vol. 32, no. 1, pp. 45–71, 2006.

[39] H. T. Huynh, "A flux reconstruction approach to high-order schemes including discontinuous Galerkin methods," in *18th AIAA Computational Fluid Dynamics Conference*, American Institute of Aeronautics and Astronautics, 2007.

[40] Z. J. Wang and H. Gao, "A unifying lifting collocation penalty formulation including the discontinuous Galerkin, spectral volume/difference methods for conservation laws on mixed grids," *Journal of Computational Physics*, vol. 228, no. 21, pp. 8161–8186, 2009.

[41] P. E. Vincent, P. Castonguay, and A. Jameson, "A new class of high-order energy stable flux reconstruction schemes," *Journal of Scientific Computing*, vol. 47, no. 1, pp. 50–72, 2010.

[42] P. Castonguay, P. E. Vincent, and A. Jameson, "A new class of high-order energy stable flux reconstruction schemes for triangular elements," *Journal of Scientific Computing*, vol. 51, no. 1, pp. 224–256, 2011.

[43] D. M. Williams and A. Jameson, "Energy stable flux reconstruction schemes for advection–diffusion problems on tetrahedra," *Journal of Scientific Computing*, vol. 59, no. 3, pp. 721–759, 2013.

[44] Y. Allaneau and A. Jameson, "Connections between the filtered discontinuous Galerkin method and the flux reconstruction approach to high order discretizations," *Computer Methods in Applied Mechanics and Engineering*, vol. 75, pp. 3628–3636, 2011.

[45] D. De Grazia, G. Mengaldo, D. Moxey, P. E. Vincent, and S. J. Sherwin, "Connections between the discontinuous Galerkin method and high-order flux reconstruction schemes," *International Journal for Numerical Methods in Fluids*, vol. 75, pp. 860–877, 2014.

[46] G. Mengaldo, D. De Grazia, P. E. Vincent, and S. J. Sherwin, "On the connections between discontinuous Galerkin and flux reconstruction schemes: Extension to curvilinear meshes," *Journal of Scientific Computing*, vol. 67, no. 3, pp. 1272–1292, 2015.

[47] P. Zwanenburg and S. Nadarajah, "Equivalence between the energy stable flux reconstruction and filtered discontinuous Galerkin schemes," *Journal of Computational Physics*, vol. 306, pp. 343–369, 2016.

[48] C. D. Cantwell, D. Moxey, A. Comerford, A. Bolis, G. Rocco, G. Mengaldo, D. De Grazia, S. Yakovlev, J.-E. Lombard, D. Ekelschot, B. Jordi, H. Xu, Y. Mohamied, C. Eskilsson, B. Nelson, P. E. J. Vos, C. Biotto, R. M. Kirby, and S. J. Sherwin, "*Nektar*++: An open-source spectral/*hp* element framework," *Computer Physics Communications*, vol. 192, pp. 205–219, 2015.

[49] D. Moxey, C. D. Cantwell, Y. Bao, A. Cassinelli, G. Castiglioni, S. Chun, E. Juda, E. Kazemi, K. Lackhove, J. Marcon, G. Mengaldo, D. Serson, M. Turner, H. Xu, J. Peiró, R. M. Kirby, and S. J. Sherwin, "*Nektar*++: Enhancing the capability and application of high-fidelity spectral/*hp* element methods," *Computer Physics Communications*, vol. 249, article no. 107110, 2020.

[50] F. D. Witherden, A. M. Farrington, and P. E. Vincent, "PyFR: An open source framework for solving advection–diffusion type problems on streaming architectures using the flux reconstruction approach," *Computer Physics Communications*, vol. 185, no. 11, pp. 3028–3040, 2014.

[51] N. Krais, A. D. Beck, T. Bolemann, H. Frank, D. Flad, G. J. Gassner, F. Hindenlang, M. Hoffmann, T. Kuhn, M. Sonntag, and C.-D. Munz, "FLEXI: A high order discontinuous Galerkin framework for hyperbolic-parabolic conservation laws," *Computers & Mathematics with Applications*, vol. 81, pp. 186–219, 2021.

[52] M. Kronbichler and K. Kormann, "Fast matrix-free evaluation of discontinuous Galerkin finite element operators," *ACM Transactions on Mathematical Software*, vol. 45, no. 3, pp. 1–40, 2019.

[53] K. Świrydowicz, N. Chalmers, A. Karakus, and T. Warburton, "Acceleration of tensor-product operations for high-order finite element methods," *The International Journal of High Performance Computing Applications*, vol. 33, no. 4, 735–757, 2019.

[54] M. G. Duffy, "Quadrature over a pyramid or cube of integrands with a singularity at a vertex," *SIAM Journal on Numerical Analysis*, vol. 19, no. 6, pp. 1260–1262, 1982.

[55] M. Dubiner, "Spectral methods on triangles and other domains," *Journal of Scientific Computing*, vol. 6, no. 4, pp. 345–390, 1991.

[56] S. J. Sherwin and G. E. Karniadakis, "A triangular spectral element method: Applications to the incompressible Navier–Stokes equations," *Computer Methods in Applied Mechanics and Engineering*, vol. 123, no. 1-4, pp. 189–229, 1995.

[57] S. J. Sherwin and G. E. Karniadakis, "Tetrahedral *hp* finite elements: Algorithms and flow simulations," *Journal of Computational Physics*, vol. 124, no. 1, pp. 14–45, 1996.

[58] I. Lomtev and G. E. Karniadakis, "A discontinuous Galerkin method for the Navier–Stokes equations," *International Journal for Numerical Methods in Fluids*, vol. 29, no. 5, pp. 587–603, 1999.

[59] R. M. Kirby, T. C. Warburton, I. Lomtev, and G. E. Karniadakis, "A discontinuous Galerkin spectral/*hp* method on hybrid grids," *Applied Numerical Mathematics*, vol. 33, no. 1-4, pp. 393–405, 2000.

[60] P. E. J. Vos, S. J. Sherwin, and R. M. Kirby, "From h to p efficiently: Implementing finite and spectral/*hp* element methods to achieve optimal performance for low- and high-order discretisations," *Journal of Computational Physics*, vol. 229, no. 13, pp. 5161–5181, 2010.

[61] C. D. Cantwell, S. J. Sherwin, R. M. Kirby, and P. H. J. Kelly, "From h to p efficiently: Strategy selection for operator evaluation on hexahedral and tetrahedral elements," *Computers & Fluids*, vol. 43, no. 1, pp. 23–28, 2011.

[62] D. Moxey, R. Amici, and R. M. Kirby, "Efficient matrix-free high-order finite element evaluation for simplicial elements," *SIAM Journal on Scientific Computing*, vol. 42, no. 3, pp. C97–C123, 2020.

[63] H.-O. Kreiss and G. Scherer, "Finite element and finite difference methods for hyperbolic partial differential equations," in *Mathematical Aspects of Finite Elements in Partial Differential Equations*, C. de Boor, Ed., Academic Press, 1974, pp. 195–212.

[64] M. H. Carpenter, D. Gottlieb, and S. Abarbanel, "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, pp. 220–236, 1994.

[65] D. C. Del Rey Fernández, J. E. Hicken, and D. W. Zingg, "Review of summation-by-parts operators with simultaneous approximation terms for the numerical solution of partial differential equations," *Computers & Fluids*, vol. 95, pp. 171–196, 2014.

[66] M. Svärd and J. Nordström, "Review of summation-by-parts schemes for initial-boundary-value problems," *Journal of Computational Physics*, vol. 268, pp. 17–38, 2014.

[67] M. H. Carpenter and D. Gottlieb, "Spectral methods on arbitrary grids," *Journal of Computational Physics*, vol. 129, pp. 74–86, 1996.

[68] D. A. Kopriva and G. J. Gassner, "On the quadrature and weak form choices in collocation type discontinuous Galerkin spectral element methods," *Journal of Scientific Computing*, vol. 44, pp. 136–155, 2010.

[69] G. J. Gassner, "A skew-symmetric discontinuous Galerkin spectral element discretization and its relation to SBP-SAT finite difference methods," *SIAM Journal on Scientific Computing*, vol. 35, no. 3, pp. A1233–A1253, 2013.

[70] S. Pirozzoli, "Generalized conservative approximations of split convective derivative operators," *Journal of Computational Physics*, vol. 229, no. 19, 7180–7190, 2010.

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

[72] J. Nordström, "Conservative finite difference formulations, variable coefficients, energy estimates and artificial dissipation," *Journal of Scientific Computing*, vol. 29, no. 3, pp. 375–404, 2006.

[73] J. E. Kozdon, E. M. Dunham, and J. Nordström, "Simulation of dynamic earthquake ruptures in complex geometries using high-order finite difference methods," *Journal of Scientific Computing*, vol. 55, no. 1, pp. 92–124, 2012.

[74] D. A. Kopriva and G. J. Gassner, "An energy stable discontinuous Galerkin spectral element discretization for variable coefficient advection problems," *SIAM Journal on Scientific Computing*, vol. 36, no. 4, pp. A2076–A2099, 2014.

[75] G. J. Gassner, "A kinetic energy preserving nodal discontinuous Galerkin spectral element method," *International Journal for Numerical Methods in Fluids*, vol. 76, no. 1, pp. 28–50, 2014.

[76] Y. Morinishi, "Skew-symmetric form of convective terms and fully conservative finite difference schemes for variable density low-Mach number flows," *Journal of Computational Physics*, vol. 229, no. 2, pp. 276–300, 2010.

[77] R. M. Kirby and G. E. Karniadakis, "De-aliasing on non-uniform grids: Algorithms and applications," *Journal of Computational Physics*, vol. 191, no. 1, pp. 249–264, 2003.

[78] D. C. Del Rey Fernández, P. D. Boom, and D. W. Zingg, "A generalized framework for nodal first derivative summation-by-parts operators," *Journal of Computational Physics*, vol. 266, pp. 214–239, 2014.

[79] J. Hicken and D. Zingg, "Summation-by-parts operators and high-order quadrature," *Journal of Computational and Applied Mathematics*, vol. 237, no. 1, pp. 111–125, 2013.

[80] K. Black, "A conservative spectral element method for the approximation of compressible fluid flow," eng, *Kybernetika*, vol. 35, no. 1, pp. 133–146, 1999.

[81] F. Hindenlang, G. J. Gassner, C. Altmann, A. D. Beck, M. Staudenmaier, and C.-D. Munz, "Explicit discontinuous Galerkin methods for unsteady problems," *Computers & Fluids*, vol. 61, pp. 86–93,

[82] H. Ranocha, P. Öffner, and T. Sonar, "Summation-by-parts operators for correction procedure via reconstruction," *Journal of Computational Physics*, vol. 311, pp. 299–328, 2016.

[83] J. E. Hicken, D. C. Del Rey Fernández, and D. W. Zingg, "Multidimensional summation-by-parts operators: General theory and application to simplex elements," *SIAM Journal on Scientific Computing*, vol. 38, no. 4, pp. A1935–A1958, 2016.

[84] D. C. Del Rey Fernández, J. E. Hicken, and D. W. Zingg, "Simultaneous approximation terms for multi-dimensional summation-by-parts operators," *Journal of Scientific Computing*, vol. 75, no. 1, pp. 83–110, 2018.

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

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

[87] P. G. LeFloch, J. M. Mercier, and C. Rohde, "Fully discrete, entropy conservative schemes of arbitrary order," *SIAM Journal on Numerical Analysis*, vol. 40, no. 5, pp. 1968–1992, 2002.

[88] T. C. Fisher, "High-order L^2 stable multi-domain finite difference method for compressible flows," Ph.D. dissertation, Purdue University, 2012.

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

[90] T. C. Fisher and M. H. Carpenter, "High-order entropy stable finite difference schemes for nonlinear conservation laws: Finite domains," *Journal of Computational Physics*, vol. 252, pp. 518–557, 2013.

[91] P. L. Roe, "Approximate Riemann solvers, parameter vectors, and difference schemes," *Journal of Computational Physics*, vol. 43, pp. 357–372, 1981.

[92] M. H. Carpenter, T. C. Fisher, E. J. Nielsen, and S. H. Frankel, "Entropy stable spectral collocation schemes for the Navier–Stokes equations: Discontinuous interfaces," *SIAM Journal on Scientific Computing*, vol. 36, no. 5, pp. B835–B867, 2014.

[93] G. J. Gassner, A. R. Winters, and D. A. Kopriva, "Split form nodal discontinuous Galerkin schemes with summation-by-parts property for the compressible Euler equations," *Journal of Computational Physics*, vol. 327, pp. 39–66, 2016.

[94] T. J. Barth, "Numerical methods for gasdynamic systems on unstructured meshes," in *An Introduction to Recent Developments in Theory and Numerics for Conservation Laws*, D. Kröner, M. Ohlberger, and C. Rohde, Eds., Springer, 1999, pp. 195–285.

[95] A. Hiltebrand and S. Mishra, "Entropy stable shock capturing space-time discontinuous Galerkin schemes for systems of conservation laws," *Numerische Mathematik*, vol. 126, no. 1, pp. 103–151, 2014.

[96] T. J. R. Hughes, L. P. Franca, and M. Mallet, "A new finite element formulation for computational fluid dynamics: I. symmetric forms of the compressible Euler and Navier–Stokes equations and the second law of thermodynamics," *Computer Methods in Applied Mechanics and Engineering*, vol. 54, no. 2, pp. 223–234, 1986.

[97] J. Crean, J. E. Hicken, D. C. Del Rey Fernández, D. W. Zingg, and M. H. Carpenter, "Entropy-stable summation-by-parts discretization of the Euler equations on general curved elements," *Journal of Computational Physics*, vol. 356, pp. 410–438, 2018.

[98] T. Chen and C.-W. Shu, "Entropy stable high order discontinuous Galerkin methods with suitable quadrature rules for hyperbolic conservation laws," *Journal of Computational Physics*, vol. 345, pp. 427–461, 2017.

[99] J. Chan, "On discretely entropy conservative and entropy stable discontinuous Galerkin methods," *Journal of Computational Physics*, vol. 362, pp. 346–374, 2018.

[100] T. Chen and C.-W. Shu, "Review of entropy stable discontinuous Galerkin methods for systems of conservation laws on unstructured simplex meshes," *CSIAM Transactions on Applied Mathematics*, vol. 1, no. 1, pp. 1–52, 2020.

[101] J. Chan, R. J. Hewett, and T. Warburton, "Weight-adjusted discontinuous Galerkin methods: Curvilinear meshes," *SIAM Journal on Scientific Computing*, vol. 39, no. 6, pp. A2395–A2421, 2017.

[102] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017.

[103] K. O. Friedrichs and P. D. Lax, "Systems of conservation equations with a convex extension," *Proceedings of the National Academy of Sciences*, vol. 68, no. 8, pp. 1686–1688, 1971.

[104] S. N. Kružkov, "First order quasilinear equations in several independent variables," *Mathematics of the USSR-Sbornik*, vol. 10, no. 2, pp. 217–243, 1970.

[105] P. D. Lax, "Shock waves and entropy," in *Contributions to Nonlinear Functional Analysis*, Academic Press, 1971, pp. 603–634.

[106] C. M. Dafermos, *Hyperbolic Conservation Laws in Continuum Physics*. Springer, 2016.

[107] T. H. Pulliam and D. W. Zingg, *Fundamentals of Computational Fluid Dynamics*. Springer, 2014.

[108] D. A. Kopriva, *Implementing Spectral Methods for Partial Differential Equations: Algorithms for Scientists and Engineers*. Springer, 2009.

[109] M. E. Gurtin, E. Fried, and L. Anand, *The Mechanics and Thermodynamics of Continua*. Cambridge University Press, 2010.

[110] G. E. Karniadakis and S. J. Sherwin, *Spectral/hp Element Methods for Computational Fluid Dynamics*, 2nd ed. Oxford University Press, 2005.

[111] J. E. Hicken and D. W. Zingg, "Dual consistency and functional accuracy: A finite-difference perspective," *Journal of Computational Physics*, vol. 256, pp. 161–182, 2014.

[112] D. C. Del Rey Fernández, J. Crean, M. H. Carpenter, and J. E. Hicken, "Staggered-grid entropy-stable multidimensional summation-by-parts discretizations on curvilinear coordinates," *Journal of Computational Physics*, vol. 392, pp. 161–186, 2019.

[113] Z. A. Worku, J. E. Hicken, and D. W. Zingg, "Quadrature rules on triangles and tetrahedra for multidimensional summation-by-parts operators," *Preprint, arXiv:2311.15576 [math.NA]*, 2023.

[114] A. Cohen and G. Migliorati, "Multivariate approximation in downward closed polynomial spaces," in *Contemporary Computational Mathematics–A Celebration of the 80th Birthday of Ian Sloan*, J. Dick, F. Y. Kuo, and H. Woźniakowski, Eds., Springer, 2018, pp. 233–282.

[115] L. Bos, "On certain configurations of points in $\mathbb{R}^n$ which are unisolvent for polynomial interpolation," *Journal of Approximation Theory*, vol. 64, no. 3, pp. 271–280, 1991.

[116] A. L. Marchildon and D. W. Zingg, "Unisolvency for polynomial interpolation in simplices with symmetrical nodal distributions," *Journal of Scientific Computing*, vol. 92, no. 2, article no. 50, 2022.

[117] Q. Chen and I. Babuška, "Approximate optimal points for polynomial interpolation of real functions in an interval and in a triangle," *Computer Methods in Applied Mechanics and Engineering*, vol. 128, no. 3, pp. 405–417, 1995.

[118] ——, "The optimal symmetrical points for polynomial interpolation of real functions in the tetrahedron," *Computer Methods in Applied Mechanics and Engineering*, vol. 137, no. 1, pp. 89–94, 1996.

[119] J. S. Hesthaven, "From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex," *SIAM Journal on Numerical Analysis*, vol. 35, no. 2, pp. 655–676, 1998.

[120] J. S. Hesthaven and C. H. Teng, "Stable spectral methods on tetrahedral elements," *SIAM Journal on Scientific Computing*, vol. 21, no. 6, pp. 2352–2380, 2000.

[121] M. A. Taylor, B. A. Wingate, and R. E. Vincent, "An algorithm for computing Fekete points in the triangle," *SIAM Journal on Numerical Analysis*, vol. 38, no. 5, pp. 1707–1720, 2001.

[122] T. Warburton, "An explicit construction of interpolation nodes on the simplex," *Journal of Engineering Mathematics*, vol. 56, no. 3, pp. 247–262, 2006.

[123] J. Chan and T. Warburton, "A comparison of high order interpolation nodes for the pyramid," *SIAM Journal on Scientific Computing*, vol. 37, no. 5, pp. A2151–A2170, 2015.

[124] A. Cicchino and S. Nadarajah, "A new norm and stability condition for tensor product flux reconstruction schemes," *Journal of Computational Physics*, vol. 429, article no. 110025, 2021.

[125] P. D. Lax and B. Wendroff, "Systems of conservation laws," *Communications on Pure and Applied Mathematics*, vol. 13, no. 2, pp. 217–237, 1960.

[126] C. Shi and C.-W. Shu, "On local conservation of numerical methods for conservation laws," *Computers & Fluids*, vol. 169, pp. 3–9, 2018.

[127] J. E. Kozdon and L. C. Wilcox, "An energy stable approach for discretizing hyperbolic equations with nonconforming discontinuous Galerkin methods," *Journal of Scientific Computing*, vol. 76, no. 3, pp. 1742–1784, 2018.

[128] D. C. Del Rey Fernández, M. H. Carpenter, L. Dalcin, S. Zampini, and M. Parsani, "Entropy stable h/p-nonconforming discretization with the summation-by-parts property for the compressible Euler and Navier–Stokes equations," *SN Partial Differential Equations and Applications*, vol. 1, no. 2, article no. 9, 2020.

[129] S. Shadpey and D. W. Zingg, "Entropy-stable multidimensional summation-by-parts discretizations on *hp*-adaptive curvilinear grids for hyperbolic conservation laws," *Journal of Scientific Computing*, vol. 82, no. 3, article no. 70, 2020.

[130] J. Chan, M. J. Bencomo, and D. C. Del Rey Fernández, "Mortar-based entropy-stable discontinuous Galerkin methods on non-conforming quadrilateral and hexahedral meshes," *Journal of Scientific Computing*, vol. 89, no. 2, article no. 51, 2021.

[131] J. Shen, L.-L. Wang, and H. Li, "A triangular spectral element method using fully tensorial rational basis functions," *SIAM Journal on Numerical Analysis*, vol. 47, no. 3, pp. 1619–1650, 2009.

[132] H. Li and L.-L. Wang, "A spectral method on tetrahedra using rational basis functions," *International Journal of Numerical Analysis and Modeling*, vol. 7, no. 2, pp. 330–355, 2010.

[133] J. Chan and L. C. Wilcox, "On discretely entropy stable weight-adjusted discontinuous Galerkin methods: Curvilinear meshes," *Journal of Computational Physics*, vol. 378, pp. 366–393, 2019.

[134] D. A. Kopriva, "Metric identities and the discontinuous spectral element method on curvilinear meshes," *Journal of Scientific Computing*, vol. 26, no. 3, pp. 301–327, 2006.

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

[136] F. Navah and S. Nadarajah, "On the verification of CFD solvers of all orders of accuracy on curved wall-bounded domains and for realistic RANS flows," *Computers & Fluids*, vol. 205, article no. 104504, 2020.

[137] D. A. Craig Penner and D. W. Zingg, "Accurate high-order tensor-product generalized summation-by-parts discretizations of hyperbolic conservation laws: General curved domains and functional superconvergence," *Journal of Scientific Computing*, vol. 93, no. 2, article no. 36, 2022.

[138] J. Proriol, "Sur une famille de polynomes à deux variables orthogonaux dans un triangle," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, vol. 245, pp. 2459–2461, 1957.

[139] T. Koornwinder, "Two-variable analogues of the classical orthogonal polynomials," in *Theory and Application of Special Functions*, R. Askey, Ed., Academic Press, 1975, pp. 435–495.

[140] T. Warburton, S. J. Sherwin, and G. E. Karniadakis, "Unstructured *hp*/spectral elements: Connectivity and optimal ordering," in *Computational Mechanics '95: Theory and Applications*, S. N. Atluri, G. Yagawa, and T. Cruse, Eds., Springer, 1995, pp. 433–444.

[141] H. Ranocha, "Comparison of some entropy conservative numerical fluxes for the Euler equations," *Journal of Scientific Computing*, vol. 76, no. 1, pp. 216–242, 2018.

[142] A. R. Winters, D. Derigs, G. J. Gassner, and S. Walch, "A uniquely defined entropy stable matrix dissipation operator for high Mach number ideal MHD and compressible Euler simulations," *Journal of Computational Physics*, vol. 332, pp. 274–289, 2017.

[143] H. Ranocha, M. Schlottke-Lakemper, J. Chan, A. M. Rueda-Ramírez, A. R. Winters, F. Hindenlang, and G. J. Gassner, "Efficient implementation of modern entropy stable and kinetic energy preserving discontinuous Galerkin methods for conservation laws," *ACM Transactions on Mathematical Software*, vol. 49, no. 4, pp. 1–30, 2023.

[144] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*. Springer New York, 1991.

[145] B. Cockburn, J. Gopalakrishnan, and R. Lazarov, "Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems," *SIAM Journal on Numerical Analysis*, vol. 47, no. 2, pp. 1319–1365, 2009.

[146] J. Chan, "Skew-symmetric entropy stable modal discontinuous Galerkin formulations," *Journal of Scientific Computing*, vol. 81, no. 1, pp. 459–485, 2019.

[147] C. Rackauckas and Q. Nie, "DifferentialEquations.jl–a performant and feature-rich ecosystem for solving differential equations in Julia," *Journal of Open Research Software*, vol. 5, no. 1, article no. 15, 2017.

[148] H. Xiao and Z. Gimbutas, "A numerical algorithm for the construction of efficient quadrature rules in two and higher dimensions," *Computers & Mathematics with Applications*, vol. 59, no. 2, pp. 663–676, 2010.

[149] J. Jaśkowiec and N. Sukumar, "High-order symmetric cubature rules for tetrahedra and pyramids," *International Journal for Numerical Methods in Engineering*, vol. 122, no. 1, pp. 148–171, 2021.

[150] J. Chan, D. C. Del Rey Fernández, and M. H. Carpenter, "Efficient entropy stable Gauss collocation methods," *SIAM Journal on Scientific Computing*, vol. 41, no. 5, pp. A2938–A2966, 2019.

[151] M. H. Carpenter and C. A. Kennedy, "Fourth-order $2N$-storage Runge–Kutta schemes," NASA Technical Memorandum 109112, Tech. Rep., 1994.

[152] F. Q. Hu, M. Y. Hussaini, and P. Rasetinera, "An analysis of the discontinuous Galerkin method for wave propagation problems," *Journal of Computational Physics*, vol. 151, pp. 921–946, 1999.

[153] K. Asthana, J. Watkins, and A. Jameson, "On consistency and rate of convergence of flux reconstruction for time-dependent problems," *Journal of Computational Physics*, vol. 334, pp. 367–391, 2017.

[154] F. Brezzi, L. D. Marini, and E. Süli, "Discontinuous Galerkin methods for first-order hyperbolic problems," *Mathematical Models and Methods in Applied Sciences*, vol. 14, no. 12, pp. 1893–1903, 2004.

[155] D. A. Di Pietro and A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*. Springer, 2012.

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

[157] H. Ranocha, "Generalised summation-by-parts operators and entropy stability of numerical methods for hyperbolic balance laws," Ph.D. dissertation, Technische Universität Braunschweig, 2018.

[158] ——, "Entropy conserving and kinetic energy preserving numerical methods for the Euler equations using summation-by-parts operators," in *Spectral and High Order Methods for Partial Differential Equations ICOSAHOM 2018*, S. J. Sherwin, D. Moxey, J. Peiró, P. E. Vincent, and C. Schwab, Eds., Springer, 2020, pp. 525–535.

[159] H. Ranocha and G. J. Gassner, "Preventing pressure oscillations does not fix local linear stability issues of entropy-based split-form high-order schemes," *Communications on Applied Mathematics and Computation*, vol. 4, no. 3, pp. 880–903, 2022.

[160] S. F. Davis, "Simplified second-order Godunov-type methods," *SIAM Journal on Scientific and Statistical Computing*, vol. 9, no. 3, pp. 445–473, 1988.

[161] G.-S. Jiang and C.-W. Shu, "Efficient implementation of weighted ENO schemes," *Journal of Computational Physics*, vol. 126, no. 1, pp. 202–228, 1996.

[162] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I. Nonstiff Problems*, 2nd revised ed. Springer, 1993.

[163] S. B. Pope, *Turbulent Flows*. Cambridge University Press, 2000.

[164] A. M. Rueda-Ramírez and G. J. Gassner, "A subcell finite volume positivity-preserving limiter for DGSEM discretizations of the Euler equations," in *WCCM–ECCOMAS Congress*, 2021.

[165] J. Chan, H. Ranocha, A. M. Rueda-Ramírez, G. J. Gassner, and T. Warburton, "On the entropy projection and the robustness of high order entropy stable discontinuous Galerkin schemes for under-resolved flows," *Frontiers in Physics*, vol. 10, 2022.

[166] W. Pazner and P.-O. Persson, "Analysis and entropy stability of the line-based discontinuous Galerkin method," *Journal of Scientific Computing*, vol. 80, no. 1, pp. 376–402, 2019.

[167] A. M. Rueda-Ramírez, W. Pazner, and G. J. Gassner, "Subcell limiting strategies for discontinuous Galerkin spectral element methods," *Computers & Fluids*, vol. 247, article no. 105627, 2022.

[168] N. K. Yamaleev and J. Upperman, "High-order positivity-preserving entropy stable schemes for the 3-D compressible Navier-–Stokes equations," *Journal of Scientific Computing*, vol. 95, no. 1, article no. 11, 2023.

[169] Y. Lin, J. Chan, and I. Tomas, "A positivity preserving strategy for entropy stable discontinuous Galerkin discretizations of the compressible Euler and Navier–Stokes equations," *Journal of Computational Physics*, vol. 475, article no. 111850, 2023.

[170] J. Chan, Y. Lin, and T. Warburton, "Entropy stable modal discontinuous Galerkin schemes and wall boundary conditions for the compressible Navier–Stokes equations," *Journal of Computational Physics*, vol. 448, article no. 110723, 2022.

[171] M. Waruszewski, J. E. Kozdon, L. C. Wilcox, T. H. Gibson, and F. X. Giraldo, "Entropy stable discontinuous Galerkin methods for balance laws in non-conservative form: Applications to the Euler equations with gravity," *Journal of Computational Physics*, vol. 468, article no. 111507, 2022.

[172] Y. Li, L.-L. Wang, H. Li, and H. Ma, "A new spectral method on triangles," in *Spectral and High Order Methods for Partial Differential Equations ICOSAHOM 2009*, J. S. Hesthaven and E. M. Rønquist, Eds., Springer, 2011, pp. 237–246.

[173] B. Zhou, B. Wang, L.-L. Wang, and Z. Xie, "A new triangular spectral element method II: Mixed formulation and *hp*-error estimates," *Numerical Mathematics: Theory, Methods and Applications*, vol. 12, no. 1, pp. 72–97, 2019.

[174] ——, "A hybridizable discontinuous triangular spectral element method on unstructured meshes and its *hp*-error estimates," *Numerical Algorithms*, vol. 91, no. 3, pp. 1231–1260, 2022.

[175] G. J. Gassner, M. Svärd, and F. Hindenlang, "Stability issues of entropy-stable and/or split-form high-order schemes," *Journal of Scientific Computing*, vol. 90, no. 3, article no. 79, 2022.