

SUPERCONVERGENT FUNCTIONAL ESTIMATES FROM SUMMATION-BY-PARTS FINITE-DIFFERENCE DISCRETIZATIONS*

JASON E. HICKEN[†] AND DAVID W. ZINGG[†]

Abstract. Diagonal-norm summation-by-parts (SBP) operators can be used to construct time-stable high-order accurate finite-difference schemes. However, to achieve both stability and accuracy, these operators must use s -order accurate boundary closures when the interior scheme is $2s$ -order accurate. The boundary closure limits the solution to $(s + 1)$ -order global accuracy. Despite this bound on solution accuracy, we show that functional estimates can be constructed that are $2s$ -order accurate. This superconvergence requires dual-consistency, which depends on the SBP operators, the boundary condition implementation, and the discretized functional. The theory is developed for scalar hyperbolic and elliptic partial differential equations in one dimension. In higher dimensions, we show that superconvergent functional estimates remain viable in the presence of curvilinear multiblock grids with interfaces. The generality of the theoretical results is demonstrated using a two-dimensional Poisson problem and a nonlinear hyperbolic system—the Euler equations of fluid mechanics.

Key words. finite-difference schemes, superconvergence, functional estimates, summation-by-parts operators, dual consistency, simultaneous approximation terms

AMS subject classifications. 65N06, 65N12

DOI. 10.1137/100790987

1. Introduction. The finite-difference method is an efficient discretization for a wide range of partial differential equations (PDEs). In addition, high-order finite-difference methods are straightforward to derive and implement, which increases their appeal. Unfortunately, constructing high-order schemes that are provably stable is more difficult, and practitioners often resort to numerical experiments to test the stability of a finite-difference method; this is never a satisfying or rigorous argument.

In an effort to construct stable high-order finite-difference discretizations, Kreiss and Scherer [15] designed summation-by-parts (SBP) operators to mimic the stability properties of Galerkin finite-element methods. SBP operators remained relatively obscure until Carpenter, Gottlieb, and Abarbanel [4] combined them with a boundary-condition penalty called a simultaneous approximation term (SAT) [9]. Subsequently, the SBP–SAT method was extended to handle block interfaces [5], curvilinear domains [24], and nonlinear problems requiring dissipation [22]. The SBP–SAT combination has proven to be a powerful approach, with applications including the Euler [22, 12], Navier–Stokes [23, 25, 26], and Einstein equations [17, 27].

Svärd [32] showed that diagonal-norm SBP operators are necessary to guarantee time stability when coordinate transformations are used. Unfortunately, to achieve both stability and high-order accuracy, diagonal-norm SBP operators have interior stencils that are twice the accuracy of their boundary stencils. For example, a sixth-

*Submitted to the journal’s Methods and Algorithms for Scientific Computing section April 2, 2010; accepted for publication (in revised form) January 26, 2011; published electronically April 7, 2011. This work was supported by the Natural Sciences and Engineering Research Council (NSERC), the Canada Research Chairs program, Bombardier Aerospace, Mathematics of Information Technology and Complex Systems (MITACS), and the University of Toronto.

<http://www.siam.org/journals/sisc/33-2/79098.html>

[†]Institute for Aerospace Studies, University of Toronto, Toronto, Ontario, M3H 5T6, Canada (jason.hicken@gmail.com, dwz@oddjob.utias.utoronto.ca).

order interior scheme must be paired with a third-order boundary scheme, and this limits the global accuracy of the solution to fourth order. Thus, the price paid for time stability is a decrease in solution accuracy for a given interior stencil size.

In many applications, solving the discretized PDEs is not the ultimate goal. Often, the objective is an accurate estimate of some functionals that depend on the solution. For example, a numerical solution to the Navier–Stokes equations may be sought because it can be used to approximate the lift and drag on an aerodynamic shape. In this context a discretization is efficient if it uses fewer computational resources, relative to other discretizations, to obtain the same accuracy in a functional estimate.

The goal of this paper is to show that discretizations based on diagonal-norm SBP operators produce superconvergent functional estimates. Specifically, if the interior scheme is $2s$ -order accurate, then a properly discretized functional will also be $2s$ -order accurate despite the reduced accuracy of the boundary closure. This superconvergence follows from the dual-consistency of SBP–SAT discretizations. Examples of superconvergence are documented in the literature, e.g., [7, 35, 3, 2, 28], but this is the first example, to our knowledge, of superconvergent functional estimates produced by a finite-difference method.

The paper is organized as follows. Section 2 introduces the notation and definitions used in the subsequent sections. We also use section 2 to review SBP–SAT discretizations. We present the main theoretical results in section 3. In particular, we prove that SBP–SAT discretizations of one-dimensional scalar hyperbolic and elliptic PDEs yield superconvergent functional estimates. In section 4, we discuss issues that arise in practical implementations—curvilinear grids and subdomain interfaces—and derive the necessary conditions for functional estimates to remain superconvergent in these situations. We verify the theory in section 5 using a two-dimensional Poisson PDE and the Euler equations of fluid dynamics. Conclusions and a summary are presented in section 6.

2. Notation and definitions. This section introduces the notation and definitions used throughout the paper. The notation is consistent with [13], and readers familiar with that reference may wish to proceed to section 3.

The SBP difference operators are defined on a computational grid consisting of $n + 1$ uniformly spaced points $x_k = kh$, $k = 0, 1, \dots, n$, with mesh spacing $h = 1/n$. This implies that the computational domain is the interval $[0, 1]$; if this is not the case, it is assumed that an invertible and sufficiently differentiable transformation can be introduced to map the domain to $[0, 1]$.

We use capital letters with a script type to denote functions on a specified domain Ω . For example, $\mathcal{U}(x) \in C^p[0, 1]$ is a function in the set of p -times differentiable functions on the interval $[0, 1]$. Small roman letters in a serif type are used to indicate a function restricted to the grid. This is illustrated with $\mathcal{U}(x)$ and $u \in \mathbb{R}^{n+1}$ as follows:

$$u = [\mathcal{U}(x_0) \quad \mathcal{U}(x_1) \quad \cdots \quad \mathcal{U}(x_n)]^T.$$

If a subscript h appears on a vector, for example, $u_h \in \mathbb{R}^{n+1}$, this indicates that the vector is the solution of a difference equation.

The “big O” notation is used extensively in the proofs of section 3. We write $F(h) = O(h^p)$ if and only if $\exists M > 0$ and $h_\star > 0$ such that

$$|F(h)| \leq Mh^p \quad \forall h < h_\star.$$

To manipulate the SAT penalties, we make use of the unit vectors $e_0, e_n \in \mathbb{R}^{n+1}$, which are the first and last columns of the $(n+1) \times (n+1)$ identity matrix, respectively:

$$e_0 = [1 \ 0 \ 0 \ \cdots \ 0 \ 0]^T,$$

$$e_n = [0 \ 0 \ 0 \ \cdots \ 0 \ 1]^T.$$

We will frequently use these vectors to construct the rank-one matrices

$$e_0 e_0^T = \text{diag}(1, 0, 0, \dots, 0) \quad \text{and} \quad e_n e_n^T = \text{diag}(0, 0, \dots, 0, 1).$$

The Kronecker and Hadamard products are nonstandard matrix products that will prove useful in the development of the theory, particularly for problems involving multiple space dimensions. For matrices $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{q \times r}$ the Kronecker product is defined by

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

Thus, $A \otimes B$ is an $(nq) \times (mr)$ matrix. The Kronecker product is associative and distributive, but it is not commutative. In addition, we highlight the following two properties of the Kronecker product:

$$(A \otimes B)(C \otimes D) = AC \otimes BD,$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

The Hadamard product, also called the entrywise or Schur product, is defined by

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

where $A, B, (A \circ B) \in \mathbb{R}^{n \times m}$. The Hadamard product is the entrywise multiplication of each set of elements; in this sense it is analogous to matrix addition.

2.1. Summation-by-parts operators. We now formally define SBP finite-difference operators, which were first introduced by Kreiss and Scherer [15] and later refined by Strand [31].

DEFINITION 1 (summation-by-parts operator). *The matrix $D \in \mathbb{R}^{(n+1) \times (n+1)}$ is a summation-by-parts (SBP) operator for the first derivative if it has the form*

$$D = H^{-1}Q,$$

where $H \in \mathbb{R}^{(n+1) \times (n+1)}$ is a symmetric-positive-definite weight matrix with entries $H_{ij} = O(h)$, and $Q \in \mathbb{R}^{(n+1) \times (n+1)}$ satisfies

$$Q + Q^T = \text{diag}(-1, 0, 0, \dots, 0, 1)$$

$$= e_n e_n^T - e_0 e_0^T.$$

Furthermore, D is a $2s$ -order accurate approximation to d/dx at the interior nodes, $\{x_k\}_{k=r}^{n-r}$, and a τ -order accurate approximation at the boundary nodes, $\{x_k\}_{k=0}^{r-1}$ and $\{x_k\}_{k=n-r+1}^n$.

To be clear, there are r boundary points at each end of the domain. The even order of accuracy $2s$ for the internal nodes is a consequence of using centered-difference schemes consisting of s points on either side of the node in question. This provides the lowest error for a given stencil size.

This work is concerned with SBP operators that have diagonal weight matrices [15, 31] of the form

$$H = h \begin{pmatrix} H_L & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & H_R \end{pmatrix},$$

where

$$\begin{aligned} H_L &= \text{diag}(\rho_0, \rho_1, \dots, \rho_{r-1}), \\ H_R &= \text{diag}(\rho_{r-1}, \dots, \rho_1, \rho_0) \end{aligned}$$

with $\rho_i > 0$. The matrix H is symmetric and positive-definite, so we can use it to define an inner product and corresponding norm for vectors. Let $u, z \in \mathbb{R}^{n+1}$ be two functions restricted to the grid. Then

$$(u, z)_H \equiv u^T H z \quad \text{and} \quad \|u\|_H^2 \equiv (u, u)_H$$

define the H inner product and H norm, respectively.

SBP operators with diagonal weight matrices are limited to $\tau \leq s$; the truncation error of the difference approximation at the boundary is at most s -order accurate when the approximation is $2s$ -order accurate in the interior. Therefore, when these operators are used in a time-stable discretization of a hyperbolic PDE, the solution error is $O(h^{s+1})$, i.e., one order higher than the boundary truncation error [10].

Thus, the interior-node stencils of “diagonal norm” SBP operators seem unnecessarily large for the achievable solution accuracy. However, the intrinsic properties of SBP operators produce two surprising consequences related to accuracy. The first is that the weight matrix defines a $2s$ -order accurate quadrature [13]. The second, which is the focus of this paper, is that functionals based on SBP-discretized PDEs are also $2s$ -order accurate.

2.2. SBP–SAT discretizations of PDEs. We will use a simple constant-coefficient linear advection equation to introduce SATs and illustrate an SBP–SAT discretization of an initial-boundary value problem. Consider the PDE

$$\begin{aligned} (2.1) \quad \frac{\partial \mathcal{U}}{\partial t} + a \frac{\partial \mathcal{U}}{\partial x} &= 0 \quad \forall x \in \Omega = [0, 1], \\ \mathcal{U}(x, 0) &= \mathcal{U}_0(x), \\ \mathcal{U}(0, t) &= \mathcal{U}_L(t), \end{aligned}$$

where $a > 0$ is the advection velocity. If we discretize the spatial derivative in (2.1) using an SBP operator D , we obtain the semidiscrete equation

$$(2.2) \quad \frac{du_h}{dt} + a D u_h = 0.$$

Equation (2.2) is not useful in practice, because the boundary data is not supplied to the system of ODEs. To impose the boundary conditions in a strong sense, we could use the injection method, which is a common approach in finite-difference discretizations. In the present example, injection would delete the first ODE in (2.2) and replace it with the boundary condition

$$(2.3) \quad [I - e_0 e_0^T] \left[\frac{du_h}{dt} + aDu_h \right] + e_0 [e_0^T u_h - \mathcal{U}_L(t)] = 0.$$

While simple to implement, injection can destroy the stability properties of high-order SBP operators [20].

SATs offer a time-stable alternative to the injection method.¹ An SAT is a penalty term used to impose the boundary conditions in a weak sense [9, 4]. The SBP–SAT semidiscretization of (2.1) would take the form

$$(2.4) \quad \frac{du_h}{dt} + aDu_h = -\sigma aH^{-1}e_0 (e_0^T u_h - \mathcal{U}_L(t)),$$

where σ is a penalty parameter. The SAT penalty on the right-hand side of (2.4) is zero for all equations except the first.

To find acceptable values for σ , we apply the energy method to (2.4) and seek a time-stable solution with $\mathcal{U}_L = 0$. Multiplying (2.4) from the left by $u_h^T H$ and adding the transpose of the result, we find

$$\begin{aligned} \frac{d}{dt} \|u_h\|_H^2 &= -au_h^T(Q + Q^T)u_h - 2\sigma au_0^2 \\ &= -au_n^2 + a(1 - 2\sigma)u_0^2, \end{aligned}$$

where we have used $Q + Q^T = e_n e_n^T - e_0 e_0^T$. To obtain a nongrowing solution, it follows that $\sigma \geq \frac{1}{2}$. A similar condition for σ arises when considering time-stability at interfaces between grids (see, for example, [6]). These inequalities provide considerable flexibility in choosing σ from the perspective of stability; however, we shall see that only certain values lead to superconvergent functionals.

2.3. The dual problem and dual-consistency. Duality plays a central role in the theory of superconvergent functionals; therefore, in this section we review the continuous and discrete dual problems, and we introduce the concept of dual or adjoint consistency [19] (dual and adjoint will be used interchangeably throughout the paper).

Let L be a linear differential operator, and consider the PDE

$$(2.5) \quad L\mathcal{U} - \mathcal{F} = 0 \quad \forall x \in \Omega,$$

where $\mathcal{F} \in L_2(\Omega)$. For simplicity, we will assume that \mathcal{U} satisfies homogeneous boundary conditions; see [16] for a more general discussion. Consider a linear functional of the solution to (2.5),

$$(2.6) \quad \mathcal{I}(\mathcal{U}) = (\mathcal{G}, \mathcal{U})_\Omega,$$

where $(\cdot, \cdot)_\Omega$ is an inner product on Ω , and $\mathcal{G} \in L_2(\Omega)$. Taking the inner product of the PDE with the generic function \mathcal{V} , the functional can be written as

$$(2.7) \quad \begin{aligned} \mathcal{I}(\mathcal{U}) &= (\mathcal{G}, \mathcal{U})_\Omega - (\mathcal{V}, L\mathcal{U} - \mathcal{F})_\Omega \\ &= (\mathcal{V}, \mathcal{F})_\Omega - (L^*\mathcal{V} - \mathcal{G}, \mathcal{U})_\Omega, \end{aligned}$$

¹A semidiscretization of (2.1) is time stable if, for $\mathcal{U}_L = 0$, the discrete solution satisfies $\|u_h(t)\| \leq K\|u_0\|$ for some constant $K \in \mathbb{R}$ (see, e.g., [11]).

where the formal adjoint operator L^* is defined by Green's identity [16] $(\mathcal{V}, L\mathcal{U})_\Omega = (L^*\mathcal{V}, \mathcal{U})_\Omega$. The dual problem associated with (2.5) and the functional \mathcal{I} can be extracted from the second term in (2.7):

$$(2.8) \quad L^*\mathcal{V} - \mathcal{G} = 0 \quad \forall x \in \Omega.$$

If we assume that the dual problem is well posed, with solution \mathcal{V} , then the second term in (2.7) vanishes, and the functional becomes

$$\mathcal{I}(\mathcal{U}) = \mathcal{I}(\mathcal{V}) = (\mathcal{V}, \mathcal{F})_\Omega.$$

Note that the roles of \mathcal{F} and \mathcal{G} in the primal formulation (i.e., the PDE (2.5) and $\mathcal{I}(\mathcal{U})$) are reversed in the adjoint formulation (i.e., the PDE (2.8) and $\mathcal{I}(\mathcal{V})$).

The preceding analysis can be adapted to the discrete problem. Let

$$L_h u_h - f = 0$$

be a discretization of (2.5), and let

$$\mathcal{I}_h(u_h) = (g, u_h)_h$$

be a discrete approximation of the functional \mathcal{I} . In general, the linear operator L_h incorporates both the discretization of the PDE and the boundary conditions. Similarly, the discrete inner product $(\cdot)_h$ may include internal and boundary terms. Using this inner product, we have

$$\begin{aligned} \mathcal{I}_h(u_h) &= (g, u_h)_h - (v_h, L_h u_h - f)_h \\ &= (v_h, f)_h - (L_h^T v_h - g, u_h)_h, \end{aligned}$$

where $(v_h, L_h u_h)_h = (L_h^T v_h, u_h)$ defines the discrete adjoint operator L_h^T , assuming a real-valued problem. In analogy with (2.8), we define the discrete dual equation

$$(2.9) \quad L_h^T v_h - g = 0.$$

DEFINITION 2 (dual (adjoint) consistency). *A discrete operator L_h and functional \mathcal{I}_h are dual (adjoint) consistent of order $q \geq 1$ with respect to a corresponding continuous PDE and functional if*

$$L_h^T v - g = O(h^q),$$

where v is the solution to the continuous dual problem projected onto the discrete solution space.

In other words, a discretization is dual-consistent if it leads to a discrete dual problem that is a consistent discretization of the continuous dual PDE. It is important to recognize that dual-consistency does not follow from consistency of the primal PDE discretization in general.

3. Theory. The main results of this section assume a linear functional and a linear scalar PDE in one space dimension. These assumptions help simplify the presentation, and they can be relaxed without significant difficulties. For example, the results can be generalized to nonlinear PDEs and functionals by using Fréchet derivatives to define the linear adjoint problem [28], and multidimensional PDEs can be

discretized using tensor products to construct the necessary SBP derivative operators. Indeed, the results will be illustrated in section 5 using the two-dimensional Euler equations.

We will require the following lemma to prove the superconvergence of SBP–SAT functional estimates. The proof can be found in [13].

LEMMA 3. *Let $D = H^{-1}Q$ be an SBP first derivative operator. Then*

$$(z, Du)_H = z^T Qu$$

is a $2s$ -order accurate approximation to the integral

$$\int_0^1 \mathcal{Z} \frac{d\mathcal{U}}{dx} dx,$$

where $\mathcal{Z} \frac{d\mathcal{U}}{dx} \in C^{2s-1}[0, 1]$.

3.1. Linear hyperbolic PDEs. Consider a scalar hyperbolic PDE of the form

$$\begin{aligned} \frac{\partial \mathcal{U}}{\partial t} + \frac{\partial}{\partial x} (\lambda \mathcal{U}) &= \mathcal{F} \quad \forall x \in \Omega = [0, 1], \quad t > 0, \\ \mathcal{U}(x, 0) &= \mathcal{U}_0(x), \\ \mathcal{U}(0, t) &= \mathcal{U}_L(t), \end{aligned}$$

where $\lambda(x) > 0$ is the spatially varying wave speed. We will restrict our attention to stationary problems, in which case the above PDE simplifies to

$$(3.1) \quad \begin{aligned} \frac{\partial}{\partial x} (\lambda \mathcal{U}) &= \mathcal{F} \quad \forall x \in \Omega = [0, 1], \\ \mathcal{U}(0) &= \mathcal{U}_L. \end{aligned}$$

The SBP–SAT discretization of this PDE is given by

$$(3.2) \quad D(\Lambda u_h) = f - H^{-1}e_0\lambda_0(e_0^T u_h - \mathcal{U}_L),$$

where

$$\begin{aligned} \lambda_i &= \lambda(x_i), \quad i = 0, \dots, n, \\ \Lambda &= \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_N), \\ f &= [\mathcal{F}(x_0) \quad \mathcal{F}(x_1) \quad \dots \quad \mathcal{F}(x_n)]^T. \end{aligned}$$

The vector $u_h \in \mathbb{R}^{n+1}$ denotes the unique solution to the linear system (3.2). For diagonal-norm SBP operators with an s -order accurate boundary closure, Gustafsson’s theory [10] and numerical experiments suggest that the solution error $\|u_h - u\|_p$ is at best $O(h^{s+1})$. Despite this bound on the accuracy of the discrete solution, the following theorem concludes that a functional estimate based on u_h can approximate the actual functional to order $2s$.

THEOREM 4. *Let $\mathcal{I} : L^2[0, 1] \rightarrow \mathbb{R}$ be a linear functional defined by*

$$(3.3) \quad \mathcal{I}(\mathcal{U}) = \int_0^1 \mathcal{G} \mathcal{U} dx + \alpha (\lambda \mathcal{U})|_{x=1},$$

where $\mathcal{G} \in C^{2s}[0, 1]$, and let $H^{-1}A \equiv H^{-1}(Q + e_0 e_0^T)\Lambda$ denote the matrix appearing in the linear system (3.2). Assume the PDE (3.1) is well posed and admits the solution $\mathcal{U} \in C^{2s}[0, 1]$. Furthermore, assume the linear system (3.2) is nonsingular with solution $u_h \in \mathbb{R}^{n+1}$, and $\|A^{-T}H\|_\infty \leq C$ for some constant $C \in \mathbb{R}$, independent of n . Then the functional estimate

$$(3.4) \quad \mathcal{I}_h(u_h) = (g, u_h)_H + \alpha \lambda_n (e_n^T u_h),$$

where

$$g = [\mathcal{G}(x_0) \quad \mathcal{G}(x_1) \quad \cdots \quad \mathcal{G}(x_n)]^T,$$

is a $2s$ -order accurate approximation to $\mathcal{I}(\mathcal{U})$.

Proof. Using the SBP norm to integrate the functional $\mathcal{I}(\mathcal{U})$, we have [13]

$$\begin{aligned} \mathcal{I}(\mathcal{U}) &= (g, u)_H + \alpha \lambda_n (e_n^T u) + O(h^{2s}) \\ &= g^T H u + \alpha \lambda_n e_n^T u + O(h^{2s}) \\ &= g^T H u_h - g^T H (u_h - u) + \alpha \lambda_n e_n^T u_h - \alpha \lambda_n e_n^T (u_h - u) + O(h^{2s}) \\ &= \mathcal{I}_h(u_h) - (g^T + \alpha \lambda_n e_n^T H^{-1}) H (u_h - u) + O(h^{2s}). \end{aligned}$$

We must show that the second term in the last line is order h^{2s} . Our strategy will be to rearrange this expression to find a suitable equation for the discrete adjoint variables. The result will then follow by showing that the discrete adjoint variables are a sufficiently accurate approximation to the continuous adjoint variables.

We begin by left-multiplying (3.2) by H and grouping terms involving u_h on the left-hand side,

$$(3.5) \quad Au_h = Hf + \lambda_0 e_0 \mathcal{U}_L,$$

and we observe that

$$(3.6) \quad \begin{aligned} A &= (Q + e_0 e_0^T) \Lambda \\ &= (\lambda_n e_n e_n^T - Q^T) \Lambda \end{aligned}$$

by the properties of SBP operators. Equation (3.6) expresses A in a way that facilitates the definition of the discrete adjoint variables below. Note that when A is applied to the exact solution u , rather than u_h , we have

$$(3.7) \quad Au = HD(\Lambda u) + \lambda_0 e_0 \mathcal{U}_L.$$

By assumption, the linear system (3.2) is nonsingular, which implies that A^{-1} exists. Inserting $I = HA^{-1}AH^{-1}$ into the equation for $\mathcal{I}(\mathcal{U})$, and using (3.5) and (3.7), we find

$$\begin{aligned} \mathcal{I}(\mathcal{U}) &= \mathcal{I}_h(u_h) - (g^T + \alpha \lambda_n e_n^T H^{-1}) HA^{-1}AH^{-1}H(u_h - u) + O(h^{2s}) \\ &= \mathcal{I}_h(u_h) - (g^T H + \alpha \lambda_n e_n^T) A^{-1}H(f - D\Lambda u) + O(h^{2s}) \\ &= \mathcal{I}_h(u_h) - v_h^T H(f - D\Lambda u) + O(h^{2s}). \end{aligned}$$

We have introduced the discrete adjoint variables, v_h , which are defined by

$$v_h \equiv A^{-T} (Hg + e_n \lambda_n \alpha).$$

Left-multiplying this definition by $H^{-1}A^T$ and inserting (3.6), we have

$$(3.8) \quad -\Lambda Dv_h = g - H^{-1}e_n\lambda_n(e_n^T v_h - \alpha).$$

Inspecting (3.8), we see that it is the SBP–SAT discretization of the continuous dual problem corresponding to (3.1) and (3.3):

$$(3.9) \quad \begin{aligned} -\lambda \frac{\partial \mathcal{V}}{\partial x} &= \mathcal{G} & \forall x \in \Omega = [0, 1], \\ \mathcal{V}(1) &= \alpha. \end{aligned}$$

The wave speed for the dual problem is $-\lambda(x)$, and the boundary condition is imposed at $x = 1$ rather than $x = 0$.

Let $v \in \mathbb{R}^{n+1}$ be the projection of the continuous adjoint variable \mathcal{V} onto the grid; then, the truncation error of the discretization (3.8) is given by

$$T_h \equiv g + H^{-1}e_n\lambda_n\alpha - H^{-1}A^T v,$$

and the SBP–SAT discretization implies that $\|T_h\|_\infty = O(h^s)$. Consequently, the error between the discrete adjoint and continuous adjoint variables satisfies

$$H^{-1}A^T(v_h - v) = T_h,$$

from which we can obtain the following bound using the assumption on $\|A^{-T}H\|_\infty$:

$$\|v_h - v\|_\infty \leq C\|T_h\|_\infty = O(h^s).$$

The above bound on the discrete adjoint error is suboptimal, but it is adequate for the present proof.

Finally, we have

$$\begin{aligned} \mathcal{I}(U) &= \mathcal{I}_h(u_h) - v_h^T H(f - D\Lambda u) + O(h^{2s}) \\ &= \mathcal{I}_h(u_h) - v^T H(f - D\Lambda u) + (v - v_h)^T H(f - D\Lambda u) + O(h^{2s}). \end{aligned}$$

The first error term, $v^T H(f - D\Lambda u)$, is order h^{2s} in light of Lemma 3 and the PDE definition (3.1). The second error term is the SBP inner product between the adjoint error $(v - v_h)$, which we have already shown is $O(h^s)$, and the truncation error $f - D\Lambda u$, which is also order h^s . Consequently, the second error term is $O(h^{2s})$, and we are left with

$$(3.10) \quad \mathcal{I}(U) = \mathcal{I}_h(u_h) + O(h^{2s}),$$

completing the proof. \square

Remark. The assumption $\|A^{-T}H\|_\infty \leq C$ is a form of stability requirement on the discrete adjoint system. Such conditions, together with consistency, are used to establish the convergence of a discrete solution to its continuous counterpart in steady boundary value problems; see, for example, [18]. As with similar stability conditions, this assumption is difficult to prove in general. Nevertheless, numerical experiments and experience suggest that it does hold for diagonal-norm SBP operators for a large range of problems. For example, Figure 3.1 plots $\|A^{-T}H\|_\infty$ versus $n \in [12, 200]$ for a constant wave speed $\lambda = 1$ and three diagonal-norm SBP schemes (second-

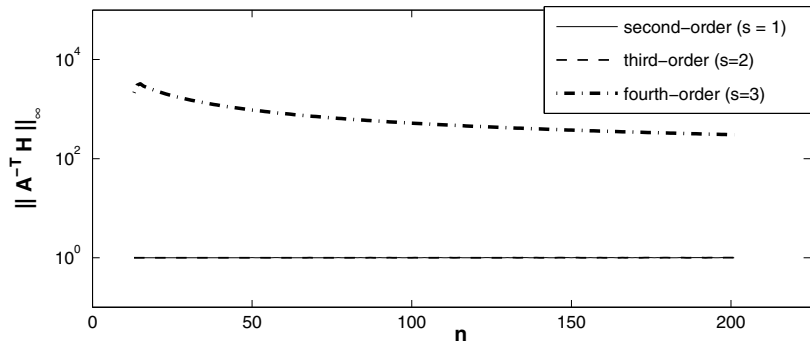


FIG. 3.1. Variation of $\|A^{-T}H\|_{\infty}$ versus n for second-, third-, and fourth-order diagonal-norm SBP operators ($\lambda = 1$).

order, third-order, and fourth-order).² The second- and third-order schemes yield $\|A^{-T}H\|_{\infty} = 1$, to machine precision, while the norm for the fourth-order scheme is monotonically decreasing for $n > 15$.

The proof of Theorem 4 relies on the dual-consistency of the SBP–SAT discretization. That is, the SBP–SAT discretization of the PDE and the discrete functional estimate lead to a discrete dual problem that is a consistent, and sufficiently accurate, discretization of the dual PDE. If the difference operator did not satisfy the SBP property, or if there were a mismatch between the SAT penalties and the discrete functional definition, the functional estimate would not be superconvergent.

3.2. Linear elliptic PDEs. Next, we consider a model linear elliptic PDE that includes both Dirichlet and Neumann boundary conditions. To simplify the analysis, we cast the second-order scalar PDE in the form of a first-order system:

$$\begin{aligned}
 (3.11) \quad & -\frac{\partial}{\partial x}(\gamma \mathcal{W}) = \mathcal{F} \quad \forall x \in \Omega = [0, 1], \\
 & \mathcal{W} = \frac{\partial \mathcal{U}}{\partial x} \quad \forall x \in \Omega = [0, 1], \\
 & \mathcal{U}(0) = \mathcal{U}_L, \quad \mathcal{W}(1) = \mathcal{W}_R.
 \end{aligned}$$

In general, the diffusion coefficient, $\gamma(x) > 0$, is permitted to be a function of the space coordinate. An SBP–SAT discretization of the PDE (3.11) is given by

$$\begin{aligned}
 (3.12) \quad & -D(\Gamma w_h) = f - H^{-1}e_0(e_0^T u_h - \mathcal{U}_L) - H^{-1}e_n \gamma_n(e_n^T w_h - \mathcal{W}_R), \\
 & w_h = Du_h + H^{-1}e_0(e_0^T u_h - \mathcal{U}_L),
 \end{aligned}$$

where

$$\begin{aligned}
 & \gamma_i = \gamma(x_i), \quad i = 0, \dots, n, \\
 & \Gamma = \text{diag}(\gamma_0, \gamma_1, \dots, \gamma_N), \\
 & f = [\mathcal{F}(x_0) \quad \mathcal{F}(x_1) \quad \dots \quad \mathcal{F}(x_n)]^T.
 \end{aligned}$$

²Fourth-order diagonal-norm SBP operators are not unique; here we use the fourth-order minimum bandwidth operator from [8].

To be clear, w_h depends explicitly on u_h in the discrete formulation; i.e., this is not a compact- or Padé-type scheme.

For constant γ , the discretization in (3.12) would approximate the second derivative using two applications of the first derivative operator. This approach may not be optimal from the perspective of stencil size. Narrow-stencil SBP operators for the second derivative have been proposed [21] but only for constant diffusion coefficients. Research into narrow-stencil operators that permit spatially varying diffusion coefficients is ongoing [14].

The order of accuracy of u_h in (3.12) is not obvious. The truncation errors of the individual equations in (3.12) correspond with the accuracy of the first-derivative SBP operator, $O(h^s)$ at the boundary and $O(h^{2s})$ in the interior. However, if we eliminate w_h , we obtain a single equation for u_h that has a boundary truncation error of $O(h^{s-1})$. Svård and Nordström [33] have shown that the boundary closure can be reduced two orders for strongly pointwise-stable discretizations of parabolic equations. This suggests that the truncation error present in (3.12) leads to a solution error that is $O(h^{s+1})$, since SBP–SAT discretizations of parabolic equations, with suitable penalty parameters, are strongly pointwise-stable [33].

The following theorem extends Theorem 4 to the elliptic case. Here, the functional includes contributions from the function and its derivative in the interior and on the boundary. The boundary contributions to the functional are at opposite ends of the domain with respect to the given boundary conditions; if the opposite were true, we would simply use the given boundary data.

THEOREM 5. *Let $\mathcal{I} : L^2[0, 1] \rightarrow \mathbb{R}$ be a linear functional defined by*

$$(3.13) \quad \mathcal{I}(U) = \int_0^1 \mathcal{G}_1 U \, dx + \int_0^1 \mathcal{G}_2 \frac{\partial U}{\partial x} \, dx + \alpha U|_{x=1} + \beta \left(\gamma \frac{\partial U}{\partial x} \right) \Big|_{x=0}.$$

Assume that $\mathcal{G}_1, \mathcal{G}_2 \in C^{2s}[0, 1]$. If $U \in C^{2s}[0, 1]$ is the solution to (3.11), and $u_h, w_h \in \mathbb{R}^{n+1}$ are the solutions to (3.12), then the functional estimate

$$(3.14) \quad \begin{aligned} \mathcal{I}_h(u_h) &= (g_1, u_h)_H + (g_2, w_h)_H \\ &+ \alpha(e_n^T u_h) + \beta\gamma_0(e_0^T w_h) + \beta(e_0^T u_h - \mathcal{U}_L), \end{aligned}$$

where

$$\begin{aligned} g_1 &= [\mathcal{G}_1(x_0) \quad \mathcal{G}_1(x_1) \quad \cdots \quad \mathcal{G}_1(x_n)]^T, \\ g_2 &= [\mathcal{G}_2(x_0) \quad \mathcal{G}_2(x_1) \quad \cdots \quad \mathcal{G}_2(x_n)]^T, \end{aligned}$$

is a $2s$ -order accurate approximation to $\mathcal{I}(U)$.

Proof. As in the hyperbolic PDE case, we begin by integrating the appropriate terms in the functional $\mathcal{I}(U)$ using the SBP norm:

$$\mathcal{I}(U) = (g_1, u)_H + (g_2, w)_H + \alpha(e_n^T u) + \beta\gamma_0(e_0^T w) + O(h^{2s}).$$

Subsequently, we add and subtract terms to relate the functional to the estimate $\mathcal{I}_h(u_h)$:

$$\begin{aligned} \mathcal{I}(\mathcal{U}) &= g_1^T H u_h + g_2^T H w_h + \alpha(e_n^T u_h) + \beta\gamma_0(e_0^T w_h) + \beta e_0^T (u_h - u) \\ &\quad - g_1^T H(u_h - u) - g_2^T H(w_h - w) \\ &\quad - \alpha e_n^T (u_h - u) - \beta\gamma_0 e_0^T (w_h - w) - \beta e_0^T (u_h - u) + O(h^{2s}) \\ &= \mathcal{I}_h(u_h) + O(h^{2s}) \\ &\quad - (g_1^T + \alpha e_n^T H^{-1} + \beta e_0^T H^{-1}) H(u_h - u) \\ &\quad - (g_2^T + \beta\gamma_0 e_0^T H^{-1}) H(w_h - w). \end{aligned}$$

To prove the result, we must show the error terms in the last two lines are $O(h^{2s})$. In the hyperbolic case, we used the system matrix and its inverse to rearrange the error term and find expressions for the discrete adjoint variables. While this is possible in the elliptic case, it becomes considerably more tedious due to nested operators. Instead, we proceed by introducing the discrete adjoint variables as unspecified Lagrange multipliers and then determine suitable equations for the discrete dual problems.

Define the primal-equation residuals as follows:

$$\begin{aligned} r_u &\equiv D(\Gamma w_h) + f - H^{-1} e_0 (e_0^T u_h - \mathcal{U}_L) - H^{-1} e_n \gamma_n (e_n^T w_h - \mathcal{W}_R), \\ r_w &\equiv D u_h - w_h + H^{-1} e_0 (e_0^T u_h - \mathcal{U}_L). \end{aligned}$$

The residuals are equal to zero, because u_h and w_h satisfy the discrete equations. By exploiting the properties of SBP operators, and noting that $\mathcal{U}_L = e_0^T u$ and $\mathcal{W}_R = e_n^T w$, we can rewrite the residuals as

$$\begin{aligned} r_u &= f + D(\Gamma w) - H^{-1} [e_0 e_0^T H^{-1}] H(u_h - u) \\ &\quad - H^{-1} [D^T \Gamma^T + e_0 e_0^T \gamma_0 H^{-1}] H(w_h - w) \end{aligned}$$

and

$$r_w = D u - w - H^{-1} [D^T - e_n e_n^T H^{-1}] H(u_h - u) - (w_h - w).$$

This reformulation of the residuals, in which $(u_h - u)$ and $(w_h - w)$ terms are grouped, will simplify the subsequent algebra.

Let $v_h \in \mathbb{R}^{n+1}$ and $z_h \in \mathbb{R}^{n+1}$ be unspecified vectors—the Lagrange multipliers—associated with r_u and r_w , respectively. We take the SBP inner product of the multipliers v_h and z_h with their corresponding primal-equation residuals and subtract the result from the functional. We then group terms that contain $(u_h - u)$ and $(w_h - w)$ in order to define the discrete adjoint equations:

$$\begin{aligned} \mathcal{I}(\mathcal{U}) &= \mathcal{I}_h(u_h) + O(h^{2s}) \\ &\quad - (g_1^T + \alpha e_n^T H^{-1} + \beta e_0^T H^{-1}) H(u_h - u) \\ &\quad - (g_2^T + \beta\gamma_0 e_0^T H^{-1}) H(w_h - w) \\ &\quad - v_h^T H r_u - z_h^T H r_w \\ &= \mathcal{I}_h(u_h) - v_h^T H (f + D\Gamma w) - z_h^T H (D u - w) + O(h^{2s}) \\ &\quad + [z_h^T D^T - g_1^T - (z_h^T e_n + \alpha) e_n^T H^{-1} + (v_h^T e_0 - \beta) e_0^T H^{-1}] H(u_h - u) \\ &\quad + [v_h^T D^T \Gamma^T - g_2^T + z_h^T + (v_h^T e_0 - \beta) e_0^T \gamma_0 H^{-1}] H(w_h - w). \end{aligned}$$

We have not yet specified v_h and z_h , so we use this freedom to eliminate the terms multiplying $(u_h - u)$ and $(w_h - w)$. Hence, the discrete adjoint variables satisfy the following equations:

$$(3.15) \quad \begin{aligned} Dz_h &= g_1 + H^{-1}e_n (e_n^T z_h + \alpha) - H^{-1}e_0 (e_0^T v_h - \beta), \\ -z_h &= \Gamma Dv_h - g_2 + H^{-1}\gamma_0 e_0 (e_0^T v_h - \beta). \end{aligned}$$

These linear equations are an SBP–SAT discretization of the continuous dual problem, which is given here for reference:

$$(3.16) \quad \begin{aligned} \frac{\partial \mathcal{Z}}{\partial x} &= \mathcal{G}_1 & \forall x \in \Omega = [0, 1], \\ -\mathcal{Z} &= \gamma \frac{\partial \mathcal{V}}{\partial x} - \mathcal{G}_2 & \forall x \in \Omega = [0, 1], \\ \mathcal{V}(0) &= \beta, & \mathcal{Z}(1) = -\alpha. \end{aligned}$$

The SBP–SAT discretization (3.15) of the dual problem is pointwise-stable, so it follows from the truncation error that v_h is $(s + 1)$ -order accurate [33] and z_h is at least s -order accurate.

Returning to the expression for the functional, we have

$$\begin{aligned} \mathcal{I}(\mathcal{U}) &= \mathcal{I}_h(u_h) - v_h^T H (f + D\Gamma w) - z_h^T H (Du - w) + O(h^{2s}) \\ &= \mathcal{I}_h(u_h) - v^T H (f + D\Gamma w) - z^T H (Du - w) \\ &\quad - (v_h - v)^T H (f + D\Gamma w) - (z_h - z)^T H (Du - w) + O(h^{2s}). \end{aligned}$$

The first and second error terms, $v^T H (f + D\Gamma w)$ and $z^T H (Du - w)$, can be shown to be order h^{2s} using Lemma 3. The third and fourth error terms are also order $2s$, since $(v_h - v)$ and $(z_h - z)$ are order h^{s+1} and order h^s , respectively, and the discretization errors in the primal PDE are order h^s , e.g., $\|f + D\Gamma w\| = O(h^s)$. Therefore, we are left with

$$\mathcal{I}(\mathcal{U}) = \mathcal{I}_h(u_h) + O(h^{2s}),$$

which is the desired result. \square

4. Practical issues. In developing the present theory we have considered very simple model problems. In this section, we address two issues that arise when finite-difference methods are used in practice—curvilinear and semistructured grids.

4.1. Curvilinear grids. SBP operators can be used in higher-space dimensions by applying one-dimensional operators to each spatial direction independently. For example, suppose the cubic domain $\Omega = [0, 1]^3$ is discretized using a Cartesian product of $n + 1$ uniformly spaced nodes. Thus, the grid nodes will have the spatial coordinates

$$\mathbf{x}_{jkm} = (x_j, y_k, z_m) = \frac{1}{n}(j, k, m) \quad \forall (j, k, m) \in \{0 \leq j, k, m \leq n\}.$$

Discrete functions on this grid are represented as vectors in $\mathbb{R}^{(n+1)^3}$, and we assume that the entries in these vectors are ordered by j first, then k , and finally m . Under these assumptions, the SBP operator for $\partial/\partial x$ is given by

$$(4.1) \quad D_x = (I \otimes I \otimes D),$$

where D is a one-dimensional SBP operator and I is the $(n + 1) \times (n + 1)$ identity matrix.

The above definition of D_x is suitable only for regular grids on a cubic domain. More general geometries and grids (with the same topology) must be handled by using curvilinear grids. However, curvilinear transformations introduce geometric quantities into the discretization, and the influence of these terms on superconvergence is not obvious.

To extend the results of section 3 to curvilinear grids, we need to show that

1. discrete integration on a curvilinear domain remains $2s$ -order accurate, and
2. the discrete primal and dual equations have truncation errors of $O(h^s)$.

For simplicity, we will restrict our focus to domains in \mathbb{R}^2 . Extension to higher dimensions should be straightforward.

First, consider integration of $\mathcal{U}(x, y)$ over a domain Ω_x that is a closed compact set of \mathbb{R}^2 . Moreover, assume that Ω_x is such that there exists an invertible transformation $\mathcal{T}(x, y) = (\xi(x, y), \eta(x, y))$ of class C^{2s} that maps Ω_x to the square $\Omega_\xi = [0, 1]^2$. The domain Ω_ξ will be associated with computational space. Using the change of variable theorem, we have

$$\iint_{\Omega_x} \mathcal{U} dx dy = \iint_{\Omega_\xi} \mathcal{U} \mathcal{J} d\xi d\eta,$$

where

$$\mathcal{J} \equiv \frac{\partial(x, y)}{\partial(\xi, \eta)} = \frac{\partial x}{\partial \xi} \frac{\partial y}{\partial \eta} - \frac{\partial y}{\partial \xi} \frac{\partial x}{\partial \eta}$$

is the Jacobian of the transformation \mathcal{T}^{-1} (since the transformation is invertible, we will assume that \mathcal{J} is positive, without loss of generality).

In practice the transformation \mathcal{T} is not explicitly available, so the Jacobian at the grid points is approximated by

$$(4.2) \quad J = (D_\xi x) \circ (D_\eta y) - (D_\xi y) \circ (D_\eta x),$$

where we have abused notation to let x and y denote vectors of nodal coordinates. To make use of Kronecker products, we assume that the unknowns are ordered first in the ξ direction and then in the η direction. Thus,

$$D_\xi = (I \otimes D) \quad \text{and} \quad D_\eta = (D \otimes I).$$

We have tacitly assumed that both directions are discretized using the same number of points, since this allows us to use the same one-dimensional operator D to define both D_ξ and D_η ; however, this assumption is not necessary. In general, the coordinate directions can have different numbers of nodes.

When SBP operators are used to calculate J , the error in approximating \mathcal{J} at the boundary is $O(h^s)$. Despite this reduced order of accuracy, the following theorem indicates that superconvergence of discrete integration is not affected.

THEOREM 6. *Let Ω_x and Ω_ξ be connected compact subsets of \mathbb{R}^2 . Furthermore, let $\mathcal{T} : \Omega_x \rightarrow \Omega_\xi$ be a C^{2s} invertible mapping. For grids based on the uniform discretization of Ω_ξ and mapped to Ω_x using \mathcal{T}^{-1} , the quadrature*

$$(4.3) \quad \mathcal{I}_h(u) = u (H \otimes H) J$$

is a $2s$ -order accurate estimate of the integral

$$(4.4) \quad \mathcal{I}(\mathcal{U}) = \iint_{\Omega_x} \mathcal{U} \, dx dy$$

for $\mathcal{U} \in C^{2s}(\Omega_x)$.

The proof is based on an iterated application of Lemma 3 and is given in Appendix A.

Next, we need to show that an SBP–SAT discretization of a PDE on a curvilinear domain has a truncation error of $O(h^s)$. Moreover, this discretization must lead to a set of discrete dual equations that have a truncation error of $O(h^s)$ with respect to the adjoint PDE.

Let Ω_x and Ω_ξ be the physical and computational spaces, respectively, defined previously. Consider the hyperbolic PDE

$$\frac{\partial}{\partial x} (\lambda_x \mathcal{U}) + \frac{\partial}{\partial y} (\lambda_y \mathcal{U}) = \mathcal{F} \quad \forall \mathbf{x} \in \Omega_x,$$

where (λ_x, λ_y) is a spatially varying velocity. Transforming this PDE to computational space, we find

$$(4.5) \quad \frac{\partial}{\partial \xi} (\lambda_\xi \mathcal{U}) + \frac{\partial}{\partial \eta} (\lambda_\eta \mathcal{U}) = \mathcal{F} \mathcal{J} \quad \forall \boldsymbol{\xi} \in \Omega_\xi.$$

Using the metric relations [29, 34], the velocity components in computational space can be expressed as

$$(4.6) \quad \begin{aligned} \lambda_\xi &= \mathcal{J} \left(\frac{\partial \xi}{\partial x} \lambda_x + \frac{\partial \xi}{\partial y} \lambda_y \right) = \frac{\partial y}{\partial \eta} \lambda_x - \frac{\partial x}{\partial \eta} \lambda_y, \\ \lambda_\eta &= \mathcal{J} \left(\frac{\partial \eta}{\partial x} \lambda_x + \frac{\partial \eta}{\partial y} \lambda_y \right) = -\frac{\partial y}{\partial \xi} \lambda_x + \frac{\partial x}{\partial \xi} \lambda_y. \end{aligned}$$

For simplicity, we will assume that $\lambda_\xi > 0$ and $\lambda_\eta > 0$ and impose suitable boundary conditions for the inlet boundary:

$$(4.7) \quad \mathcal{U}(0, \eta) = \mathcal{U}_W(\eta) \quad \text{and} \quad \mathcal{U}(\xi, 0) = \mathcal{U}_S(\xi).$$

The PDE in computational space contains geometric quantities that must be approximated using a finite-difference operator. In particular, the Jacobian is approximated using (4.2), and the components of the velocity field (4.6) are approximated using

$$\begin{aligned} \Lambda_\xi &= \text{diag} (A_x D_\eta y - A_y D_\eta x), \\ \Lambda_\eta &= \text{diag} (-A_x D_\xi y + A_y D_\xi x), \end{aligned}$$

where the diagonal matrices A_x and A_y hold the velocity field components (λ_x, λ_y) evaluated at the grid nodes (x_{jk}, y_{jk}) :

$$\begin{aligned} A_x &= \text{diag} (\lambda_x(x_{jk}, y_{jk})), \\ A_y &= \text{diag} (\lambda_y(x_{jk}, y_{jk})). \end{aligned}$$

With the geometric quantities approximated, the SBP–SAT discretization of the PDE (4.5), with boundary conditions (4.7), can be written as

$$(4.8) \quad \begin{aligned} D_\xi(\Lambda_\xi u_h) + D_\eta(\Lambda_\eta u_h) &= (J \circ f) - [I \otimes (H^{-1} e_0 e_0^T)] \Lambda_\xi(u_h - u) \\ &\quad - [(H^{-1} e_0 e_0^T) \otimes I] \Lambda_\eta(u_h - u), \end{aligned}$$

where $u_h \in \mathbb{R}^{(n+1)^2}$ is the discrete solution, and $u \in \mathbb{R}^{(n+1)^2}$ is the exact solution— u is used only to supply boundary conditions here. If we use the same SBP operator for (4.8) and the geometric terms $(J, \Lambda_x, \Lambda_y)$, then the truncation error of the discretized primal PDE is $O(h^s)$ at the boundary and $O(h^{2s})$ in the interior, as required for superconvergence of functional estimates.³

Finally, we turn to the accuracy of the discrete adjoint equation. The algebra leading to this equation is analogous to the algebra involved in the one-dimensional problems of section 3, so we will simply state the final result:

$$(4.9) \quad \begin{aligned} -\Lambda_\xi D_\xi v_h - \Lambda_\eta D_\eta v_h &= (J \circ g) - \Lambda_\xi [I \otimes (H^{-1} e_n e_n^T)] (v_h - v) \\ &\quad - \Lambda_\eta [(H^{-1} e_n e_n^T) \otimes I] (v_h - v), \end{aligned}$$

where $v_h \in \mathbb{R}^{(n+1)^2}$ are the discrete adjoint variables, and $v \in \mathbb{R}^{(n+1)^2}$ is the restriction to the grid of the continuous adjoint variables. Equation (4.9) is a discretization of the continuous dual problem and, like the primal discretization, has a truncation error of $O(h^s)$ at the boundary and $O(h^{2s})$ in the interior.

We have shown that a two-dimensional hyperbolic PDE can be discretized with SBP operators and SATs such that superconvergent functional estimates are possible. While this is only one particular example, the generalization to higher dimensions and diffusion-type PDEs is straightforward.

4.2. Interface SATs for semistructured grids. Semistructured grids are necessary when the domain is not topologically equivalent to a hyperrectangle. These grids may also be employed to improve mesh quality, to avoid singularities, and to partition work for parallel computations [12].

A semistructured grid partitions a domain into a set of nonoverlapping curvilinear subdomains, such that each subdomain can be mapped to a hyperrectangle. Thus, the individual subdomains can be treated as in the previous case.

The subdomains introduce artificial boundaries that must be addressed by the discretization. One strategy is to use SATs at the interfaces between subdomains. Indeed, the SAT methodology offers several advantages in this context [12]: SATs require only C^0 grid continuity, they help reduce communication in parallel solution algorithms, and they lead to a (linearly) time-stable scheme.

In this section, we will show that using SATs to couple subdomains does not affect the superconvergence of functional estimates for the PDEs of interest. Thus, we consider the advection-diffusion PDE

$$(4.10) \quad \begin{aligned} \frac{\partial}{\partial x}(\lambda \mathcal{U}) - \frac{\partial}{\partial x}(\gamma \mathcal{W}) &= 0 & \forall x \in \Omega = [0, 2], \\ \mathcal{W} &= \frac{\partial \mathcal{U}}{\partial x} & \forall x \in \Omega = [0, 2]. \end{aligned}$$

³The SBP operator appearing in Λ_ξ and Λ_η is D_η and D_ξ , respectively, so the SBP operator is not applied twice in the same direction in (4.8).

Note that the domain has changed to the interval $[0, 2]$; this is merely for convenience. We do not specify the boundary conditions, since the boundary SATs can be ignored in the analysis of the interface SATs.

Partition the domain $\Omega = [0, 2]$ into the two subdomains $\Omega_L = [0, 1]$ and $\Omega_R = [1, 2]$, with an interface at $x = 1$. Discretize Ω_L using $n + 1$ uniformly spaced points, and Ω_R using $m + 1$ uniformly spaced points. The last grid point of Ω_L lies on the interface and coincides with the first grid point of Ω_R .

To simplify the presentation, we need to introduce some additional notation. Let $D_L = H_L^{-1}Q_L$ be a first-derivative SBP operator for Ω_L . Similarly, let $D_R = H_R^{-1}Q_R$ be an SBP operator on Ω_R . We define the following $\mathbb{R}^{(n+m+2) \times (n+m+2)}$ matrices:

$$\hat{D} = \begin{bmatrix} D_L & 0 \\ 0 & D_R \end{bmatrix}, \quad \hat{H} = \begin{bmatrix} H_L & 0 \\ 0 & H_R \end{bmatrix}, \quad \text{and} \quad \hat{Q} = \begin{bmatrix} Q_L & 0 \\ 0 & Q_R \end{bmatrix}.$$

Using the properties of Q_L and Q_R , we note that

$$\hat{Q} + \hat{Q}^T = e_n e_n^T - e_0 e_0^T - e_{n+1} e_{n+1}^T + e_{n+m+1} e_{n+m+1}^T,$$

where e_0, e_n, e_{n+1} , and e_{n+m+1} are the appropriate columns of the $(n + m + 2) \times (n + m + 2)$ identity matrix. Ignoring boundary SATs corresponding to e_0 and e_{n+m+1} , the SBP–SAT discretization of (4.10) is

$$\begin{aligned} \hat{H}\hat{D}(\Lambda u_h) - \hat{H}\hat{D}(\Gamma w_h) &= \sigma_L e_n (e_n^T - e_{n+1}^T) \Lambda u_h \\ &\quad + \rho_L e_n (e_n^T - e_{n+1}^T) \Gamma w_h \\ &\quad + \sigma_R e_{n+1} (e_{n+1}^T - e_n^T) \Lambda u_h \\ &\quad + \rho_R e_{n+1} (e_{n+1}^T - e_n^T) \Gamma w_h, \\ \hat{H}w_h - \hat{H}\hat{D}u_h &= \epsilon_L e_n (e_n^T - e_{n+1}^T) u_h \\ &\quad + \epsilon_R e_{n+1} (e_{n+1}^T - e_n^T) u_h, \end{aligned} \tag{4.11}$$

where $\sigma_L, \sigma_R, \rho_L, \rho_R, \epsilon_L, \epsilon_R \in \mathbb{R}$ are SAT penalty parameters. In contrast with previous SBP–SAT discretizations in this paper, we have left-multiplied by \hat{H} to help “clean up” the right-hand side. As before, $u_h, w_h \in \mathbb{R}^{(n+m+2)}$ are the discrete solution vectors, and Λ and Γ are diagonal $(n + m + 2) \times (n + m + 2)$ matrices representing λ and γ on the grid. Due to the coincident grid points at the interface, we have $\lambda_n = \lambda_{n+1}$ and $\gamma_n = \gamma_{n+1}$.

To find conditions on the penalty parameters that lead to dual-consistency, we take the inner product of the equations (4.11) with their corresponding discrete adjoint variables v_h and z_h . After rearranging the result, we identify the adjoint equations with the terms multiplying u_h and w_h . Omitting the algebra, we find the discrete dual equations

$$\begin{aligned} -\hat{H}\Lambda\hat{D}v_h + \hat{H}\hat{D}z_h &= \Lambda [(\sigma_L - 1)e_n e_n^T - \sigma_R e_n e_{n+1}^T] v_h \\ &\quad + [(\epsilon_L + 1)e_n e_n^T - \epsilon_R e_n e_{n+1}^T] z_h \\ &\quad + \Lambda [(\sigma_R + 1)e_{n+1} e_{n+1}^T - \sigma_L e_{n+1} e_n^T] v_h \\ &\quad + [(\epsilon_R - 1)e_{n+1} e_{n+1}^T - \epsilon_L e_{n+1} e_n^T] z_h, \\ \hat{H}z_h + \hat{H}\Gamma\hat{D}v_h &= \Gamma [(\rho_L + 1)e_n e_n^T - \rho_R e_n e_{n+1}^T] v_h \\ &\quad + \Gamma [(\rho_R - 1)e_{n+1} e_{n+1}^T - \rho_L e_{n+1} e_n^T] v_h, \end{aligned} \tag{4.12}$$

where boundary terms involving e_0 and e_{n+m+1} have been dropped. To obtain consistent interface SAT terms in the adjoint equations (4.12), we require

$$(4.13) \quad \sigma_L = \sigma_R + 1,$$

$$(4.14) \quad \rho_L = \rho_R - 1,$$

$$(4.15) \quad \epsilon_L = \epsilon_R - 1.$$

The first two constraints, (4.13) and (4.14), are also required for discrete conservation [6]. The third equation appears to be a unique requirement for dual-consistency.

The penalty parameters in (4.11) must satisfy additional constraints to guarantee time-stability. We do not discuss these constraints here and direct the interested reader to [6] for a review. For discretizations involving SAT interfaces, we will use the following SAT parameter values, which satisfy both the dual-consistency conditions (4.13)–(4.14) and the stability conditions:

$$\begin{aligned} \sigma_L &= -\frac{1}{2} [\operatorname{sgn}(\lambda_n) - 1], & \rho_L &= -\frac{1}{2}, & \epsilon_L &= -\frac{1}{2}, \\ \sigma_R &= -\frac{1}{2} [\operatorname{sgn}(\lambda_{n+1}) + 1], & \rho_R &= \frac{1}{2}, & \epsilon_R &= \frac{1}{2}. \end{aligned}$$

5. Numerical examples. We have chosen two examples to illustrate the superconvergence of SBP–SAT functional estimates. The examples were selected to verify the generality of the theory, and for each case we consider second-, third-, and fourth-order SBP operators ($s = 1$, $s = 2$, and $s = 3$, respectively). Diener et al. [8] describe the specific diagonal-norm operators that we use here. For the fourth-order scheme we use their minimum bandwidth operator.

5.1. Two-dimensional Poisson problem. Consider the following Poisson-type PDE on $\Omega = [0, 1] \times [0, 1]$:

$$(5.1) \quad \begin{aligned} -\nabla \cdot (\gamma \nabla \mathcal{U}) &= \mathcal{F} & \forall \mathbf{x} \in \Omega, \\ \mathcal{U}(\mathbf{x}) &= \mathcal{B}(\mathbf{x}) & \forall \mathbf{x} \in \partial\Omega. \end{aligned}$$

To demonstrate the theory, it is sufficient to specify a solution and diffusion coefficient for (5.1) and then determine the appropriate boundary conditions and source term, i.e., the method of manufactured solutions. Hence, we choose

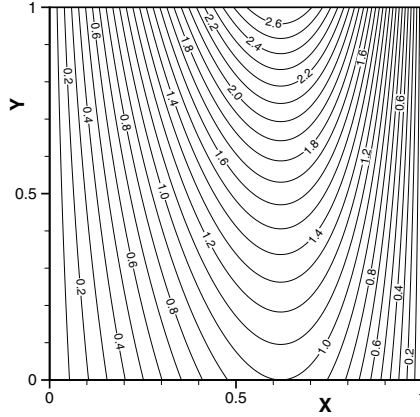
$$\mathcal{U}(\mathbf{x}) = e^y \sin\left(\frac{\pi(e^x - 1)}{e - 1}\right) \quad \text{and} \quad \gamma(\mathbf{x}) = \frac{\pi e^x}{e - 1}.$$

The contours of the function \mathcal{U} are plotted in Figure 5.1 for reference.

We are also free to specify the functional in this example and will consider functionals of the form

$$(5.2) \quad \mathcal{I}(\mathcal{U}) = \int_{\partial\Omega_1} \beta \gamma (\hat{\mathbf{n}} \cdot \nabla \mathcal{U}) \, dx,$$

where $\partial\Omega_1 = \{(x, y) \mid x \in [0, 1], y = 0\}$, i.e., $\partial\Omega_1$ is the bottom edge of the square domain. As we shall see, some care must be exercised in the choice of the function $\beta(x)$, because the functional defines the adjoint variables (see (3.16), for example). If the adjoint variables are not sufficiently differentiable, the functional estimate will not be superconvergent.


 FIG. 5.1. Contours of the exact solution U .

The PDE (5.1) is discretized using SBP operators and SATs. Geometric quantities, such as $\partial_x \xi$ and \mathcal{J} , are approximated using the methods described in section 4.1. Thus, the discrete equations are given by

$$\begin{aligned}
 (5.3) \quad & -D_\xi \left(\Gamma_{\xi\xi} w_h^{(\xi)} + \Gamma_{\xi\eta} w_h^{(\eta)} \right) - D_\eta \left(\Gamma_{\eta\xi} w_h^{(\xi)} + \Gamma_{\eta\eta} w_h^{(\eta)} \right) = f \circ J \\
 & - [I \otimes (H^{-1} e_0 e_0^T + H^{-1} e_n e_n^T)] \Gamma_{\xi\xi} (u_h - u) \\
 & - [(H^{-1} e_0 e_0^T + H^{-1} e_n e_n^T) \otimes I] \Gamma_{\eta\eta} (u_h - u), \\
 & w_h^{(\xi)} = D_\xi u_h + [I \otimes (H^{-1} e_0 e_0^T - H^{-1} e_n e_n^T)] (u_h - u), \\
 & w_h^{(\eta)} = D_\eta u_h + [(H^{-1} e_0 e_0^T - H^{-1} e_n e_n^T) \otimes I] (u_h - u).
 \end{aligned}$$

The vectors $w_h^{(\xi)}$ and $w_h^{(\eta)}$ approximate the components of the gradient $(\partial_\xi \mathcal{U}, \partial_\eta \mathcal{U})$. The diffusion coefficient and geometric quantities are gathered into the diagonal matrices

$$\begin{aligned}
 \Gamma_{\xi\xi} &= \text{diag}(\gamma) \text{diag}[(D_\eta y) \circ (D_\eta y) + (D_\eta x) \circ (D_\eta x)] \text{diag}(J)^{-1}, \\
 \Gamma_{\xi\eta} = \Gamma_{\eta\xi} &= -\text{diag}(\gamma) \text{diag}[(D_\eta y) \circ (D_\xi y) + (D_\eta x) \circ (D_\xi x)] \text{diag}(J)^{-1}, \\
 \Gamma_{\eta\eta} &= \text{diag}(\gamma) \text{diag}[(D_\xi y) \circ (D_\xi y) + (D_\xi x) \circ (D_\xi x)] \text{diag}(J)^{-1}.
 \end{aligned}$$

We emphasize that the geometric quantities are approximated using SBP operators, even when an analytical mapping is available.

Including the matrices $\Gamma_{\xi\xi}$ and $\Gamma_{\eta\eta}$ in the SAT penalties of (5.3) is not strictly necessary; these terms simply alter the scaling applied to the difference $(u_h - u)$. Thus, there is flexibility in the choice of scaling for these SATs, and the functional estimate will be superconvergent provided it uses the same scaling. Of course, the choice of scaling will impact the condition number of the system matrix.

5.1.1. Dual-consistency of the discrete functional. To obtain a superconvergent functional estimate, the theory requires that the functional be discretized such that the discrete adjoint equations are a consistent and accurate approximation to the continuous adjoint problem. We will illustrate this requirement using two alternative discretizations of the functional (5.2).

TABLE 5.1

Comparison of the convergence rates of dual-consistent and dual-inconsistent functional estimates.

n	Dual consistent			Dual inconsistent		
	2nd	3rd	4th	2nd	3rd	4th
24	2.00597	4.16802	6.29756	1.89587	5.12368	6.52419
48	2.00064	4.13503	5.90746	1.95935	1.83905	8.17144
96	1.99998	4.09146	5.80347	1.98600	2.08327	1.88536
192	1.99997	4.06162	6.35784	1.99533	2.70180	4.17460

We fix the functional (5.2) by choosing

$$(5.4) \quad \beta(x) = -\frac{\pi^2(e^x - 1)(e - e^x)}{(e - 1)^2},$$

which yields $\mathcal{I}(\mathcal{U}) = 4$ for the exact solution. This $\beta(x)$ vanishes at $x = 0$ and $x = 1$, which ensures that the continuous adjoint variables are sufficiently smooth; the role of β in superconvergence will be explored further in the next section.

If we define $b = [\beta(x_0) \ \beta(x_1) \ \cdots \ \beta(x_n)]^T$, then a dual-consistent discretization of (5.2) is given by

$$(5.5) \quad \begin{aligned} \mathcal{I}_h(u_h) = (e_0 \otimes b)^T (I \otimes H) & \left\{ \Gamma_{\eta\xi} w_h^{(\xi)} + \Gamma_{\eta\eta} w_h^{(\eta)} \right. \\ & \left. + [(H^{-1} e_0 e_0^T) \otimes I] \Gamma_{\eta\eta} (u_h - u) \right\}. \end{aligned}$$

A functional estimate that is dual-inconsistent is given by

$$(5.6) \quad \tilde{\mathcal{I}}_h(u_h) = (e_0 \otimes b)^T (I \otimes H) [\Gamma_{\eta\xi} D_\xi u_h + \Gamma_{\eta\eta} D_\eta u_h].$$

The only difference between (5.5) and (5.6) is the SAT-like penalties in the former. These penalties provide consistent boundary conditions to the adjoint PDE. Therefore, although the penalties contribute terms that are only $O(h^s)$, these small corrections are necessary for superconvergence.

The accuracy and convergence rate of the two functional estimates are determined using a sequence of grids. The grids consist of $(n+1) \times (n+1)$ nodes in each coordinate direction, with $n \in \{n_i\} = \{12, 18, 24, 36, 48, 72, 96, 144, 192\}$. The discrete solution u_h is obtained by solving (5.3) using a direct (banded) solver from LAPACK [1].

For mesh level i , the functional convergence rate is given by

$$q_i = \frac{1}{\ln(n_i/n_{i-1})} \ln \left(\frac{|E_{i-1}|}{|E_i|} \right),$$

where E_i is the functional error using $n_i + 1$ nodes in each direction. Table 5.1 lists the convergence rates for every other value of n_i starting at $n_i = 24$. The convergence rates are consistent with the theory. The second-order discretization produces a second-order functional estimate, independent of the two formulations considered. In contrast, the third- and fourth-order discretizations yield superconvergent functional estimates, but only for the dual-consistent functional.

Figure 5.2 plots the functional errors E_i versus mesh spacing $h = 1/n_i$. Initially, the third- and fourth-order dual-inconsistent functionals converge at higher-than-expected rates; however, for smaller values of h they behave as predicted. Salas and Atkins [30] observed similar behavior in functional estimates—rapid convergence followed by asymptotic behavior—and attributed it to interactions between the errors

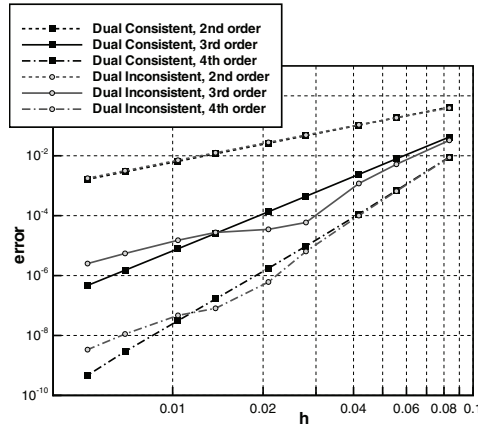


FIG. 5.2. Dual consistent and dual inconsistent functional errors.

in u_h and the numerical quadrature. The present results suggest that dual consistency may eliminate this behavior, since the corresponding dual-consistent functionals appear to converge at the same rate throughout the range of n considered. Therefore, the dual-consistent functionals may be more amenable to Richardson extrapolation.

5.1.2. Adjoint smoothness and functional accuracy. We have seen that the superconvergence of SBP–SAT functionals is destroyed if the discretization is not dual-consistent. Here we show that, even if the discretization is dual-consistent, superconvergence will not be possible if the continuous adjoint variable is not sufficiently smooth.

For an arbitrary $\beta(x)$, the PDE (5.1) and functional (5.2) have the associated adjoint PDE

$$\begin{aligned}
 (5.7) \quad & \nabla \cdot (\gamma \nabla \mathcal{V}) = 0 && \forall \mathbf{x} \in \Omega, \\
 & \mathcal{V}(\mathbf{x}) = \beta(\mathbf{x}) && \forall \mathbf{x} \in \partial\Omega_1, \\
 & \mathcal{V}(\mathbf{x}) = 0 && \forall \mathbf{x} \in \partial\Omega \setminus \partial\Omega_1.
 \end{aligned}$$

The functional is an integral over $\partial\Omega_1$, so $\beta(x)$ defines the adjoint variable on this subset of the boundary only. On the remaining boundaries, the absence of the functional implicitly defines $\mathcal{V} = 0$; therefore, $\beta(x)$ must equal zero at $x = 0$ and $x = 1$ for the boundary data to be continuous. This requirement was satisfied by (5.4), so the resulting adjoint variable was smooth, and the dual-consistent functional was superconvergent.

Now, consider adopting $\beta(x) = 1$ in the definition of the functional (5.2). This apparently innocuous choice leads to discontinuous boundary conditions in (5.7) and, consequently, the adjoint variable plotted in Figure 5.3. Singularities in \mathcal{V} are visible at $(0, 0)$ and $(1, 0)$.

Table 5.2 lists the convergence rates of dual-consistent discretizations of (5.2) with $\beta(x) = 1$. We use the same methodology and sequence of grids as in the previous study. As predicted by the theory, the singularities in the adjoint variable prevent superconvergence of the functional estimate.

5.2. Impact of a curvilinear grid on functional accuracy. Next, we verify the analysis presented in section 4.1, namely, that the functional estimates remain

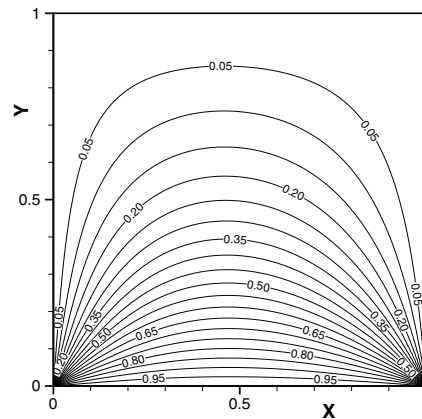


FIG. 5.3. Contours of the adjoint variable when $\beta(x) = 1$ in (5.2).

TABLE 5.2

Convergence rates of functional estimates corresponding to a nonsmooth adjoint variable.

n	2nd	3rd	4th
24	1.83336	1.97355	3.00296
48	1.88125	2.59152	3.74879
96	1.90415	2.80494	3.91383
192	1.91653	2.89948	3.96385

superconvergent on curvilinear grids. We introduce a smooth mapping from computational to physical space:

$$\begin{aligned} x(\xi, \eta) &= \xi + \frac{1}{40} \sin(2\pi\xi) \sin(2\pi\eta), \\ y(\xi, \eta) &= \eta + \frac{1}{40} \sin(2\pi\eta) \sin(2\pi\xi). \end{aligned}$$

Such a curvilinear grid, while unnecessary for the domain Ω , eliminates certain error cancellations that can arise on uniform Cartesian grids; such cancellations would not be representative of the grids encountered in practice. Figure 5.4 shows an example grid with $n = 48$ in each computational direction.

We consider a sequence of grids with $(n_i + 1)$ nodes in the ξ - and η -coordinate directions, where, as before,

$$n_i \in \{12, 18, 24, 36, 48, 72, 96, 144, 192\}.$$

The functional is (5.2) with β given by (5.4). Table 5.3 lists the convergence rate of the functional estimate (5.5) on both uniform and curvilinear grids, for every other grid in the sequence beginning with $n_i = 24$. The third-order functional estimate is clearly superconvergent on the curvilinear grid. The fourth-order functional does not exhibit its expected convergence rate (sixth-order) for the range of h considered.

Figure 5.5 plots the error in the functional estimates calculated on the uniform and curvilinear grids. The second- and third-order estimates are nearly identical on both grid families. The fourth-order estimate varies slightly between the two grid families; the fourth-order method appears to be more sensitive to the curvilinear grid.

TABLE 5.3
Convergence rates of the functional estimate on uniform and curvilinear grids.

n	Uniform grid			Curvilinear grid		
	2nd	3rd	4th	2nd	3rd	4th
24	2.00597	4.16802	6.29756	2.02270	4.17574	12.32711
48	2.00064	4.13503	5.90746	2.00454	4.19971	6.07181
96	1.99998	4.09146	5.80347	2.00087	4.12854	9.04578
192	1.99997	4.06162	6.35784	2.00018	4.07716	4.88753

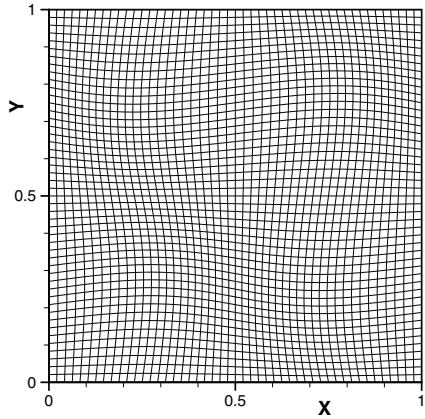


FIG. 5.4. Example curvilinear grid with $n = 48$.

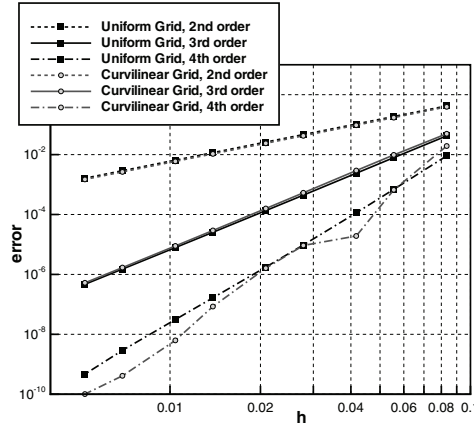


FIG. 5.5. Functional estimate errors on the uniform and curvilinear grids.

5.3. Nonlinear problem: Two-dimensional Euler equations. The Euler equations model the flow of a compressible inviscid fluid. They provide a useful verification of the present theory, since they are nonlinear. In two dimensions, the steady Euler equations take the form

$$\frac{\partial}{\partial x} \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho v u \\ (e + p)u \end{pmatrix} + \frac{\partial}{\partial y} \begin{pmatrix} \rho v \\ \rho u v \\ \rho v^2 + p \\ (e + p)v \end{pmatrix} = 0,$$

where ρ is the density, $\mathbf{u} = (u \ v)^T$ is the velocity, e is the energy, and p is the pressure. The Euler equations are closed using the perfect-gas equation of state and are subject to appropriate boundary conditions. On wall boundaries, a tangential flow is imposed by setting $\hat{\mathbf{n}}^T \mathbf{u} = 0$, where $\hat{\mathbf{n}}$ is a unit normal to the boundary. On the remaining boundaries, incoming characteristics are set equal to the exact solution.

The Euler equations are a system of nonlinear PDEs, so some additional remarks on their discretization are warranted. Numerical dissipation must be added to the discrete equations to damp high-frequency modes, because the SBP operators are nearly skew-symmetric. To ensure that the discretization is dual-consistent and stable, we use dissipation operators of the form [22, 8]

$$A_s = H^{-1} D_s^T \Sigma D_s,$$

where D_s denotes an s -order undivided difference operator, Σ is a diagonal matrix of flux Jacobian spectral radii (see Appendix B), and H is the weight matrix from the

first-derivative SBP operator. The operator A_s is applied to the conservative variables $(\rho, \rho u, \rho v, e)$ individually. The matrix HA_s is symmetric semidefinite (Σ contains only positive entries), so the dissipation operator is both stable and dual consistent with respect to the H inner product. We pair A_s with the first-derivative SBP operator that is s -order accurate at the boundary, thereby maintaining $(s + 1)$ -order accuracy of the solution.⁴

Until now, we have discussed SAT penalties only for scalar PDEs. For hyperbolic systems, such as the Euler equations, the SAT penalties can be applied to the incoming characteristics by analyzing the eigenvalues of the flux Jacobian matrix; see [12] for details of our implementation. However, this approach must be modified for the tangential-flow boundary condition to ensure dual consistency. To avoid a lengthy digression, the appropriate boundary treatment is analyzed in Appendix B.

We will use the two-dimensional supersonic vortex to illustrate SBP–SAT functional convergence for the Euler equations. This isentropic flow has a smooth analytical solution that can be used to verify the solution; indeed, this is one of the few flows for which an analytical solution of the two-dimensional Euler equations is known. The streamlines of the vortex are concentric circles about the origin, and the density is defined by (in polar coordinates)

$$\rho(r) = \rho_{\text{in}} \left[1 + \frac{\gamma - 1}{2} M_{\text{in}}^2 \left(1 - \frac{r_{\text{in}}^2}{r^2} \right) \right]^{\frac{1}{\gamma - 1}}.$$

The remaining variables can be obtained using isentropic relations. The subscript “in” indicates values along the inner radius r_{in} . Assuming nondimensionalized variables, we have chosen $r_{\text{in}} = 1$, $\rho_{\text{in}} = 2$, $M_{\text{in}} = 2$, and $p_{\text{in}} = 1/\gamma$, where γ is the ratio of heat capacities.

The geometry and grid topology for this example are shown in Figure 5.6. The grid consists of four curvilinear blocks that conform to the radii $r = 1$ and $r = 3$. Angular perturbations have been introduced in the radial grid lines to prevent error cancellations that may appear with orthogonal grid lines.

An inviscid-wall boundary condition is applied along the inner radius, $\partial\Omega_{\text{in}} = \{r = 1, \theta \in [0, \pi/2]\}$, and characteristic-based SATs supply the exact solution along the remaining boundaries. The functional is the force in the x direction on the boundary $\partial\Omega_{\text{in}}$. The functional’s exact value is $F = -1/\gamma$.

We use a set of six grids for the grid refinement study. The finest grid consists of 321×321 nodes on each block, or 412 164 nodes in total. The five remaining grid levels are obtained by successively removing every other node from the finest grid. The grid in Figure 5.6 is the second coarsest grid (21×21 nodes per block). The fourth-order SBP stencil is not compatible with the coarsest grid in the set, so only five-grid levels are used with the fourth-order operator.

We begin by examining the L^2 error in the solution. Figure 5.7 plots the error in density versus the normalized mesh spacing h_i/h_0 , where $h_i = 1/n$ is the mesh spacing on a block of grid level i . Results from second-, third-, and fourth-order accurate SBP operators are included, and both dual-consistent and dual-inconsistent boundary treatments are shown; the boundary treatments are described in Appendix B. The figure demonstrates that the discretizations obtain their designed order of accuracy. Moreover, dual-consistency has very little impact on the solution’s order of accuracy.

⁴In the notation of reference [22], the present dissipation operators do not use an operator B whose entries decrease to $O(h^{s-1})$ toward the boundaries, so A_s must use a wider stencil than the dissipation operators of [22] to obtain the necessary accuracy.

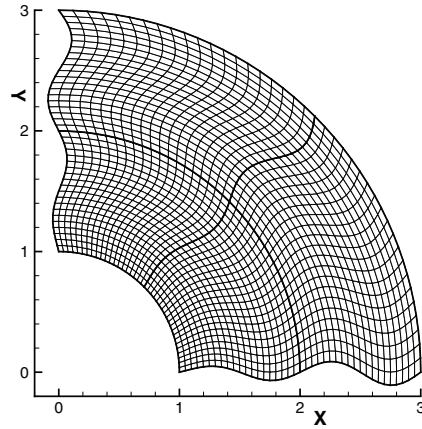


FIG. 5.6. Geometry and grid topology for the inviscid-vortex study.

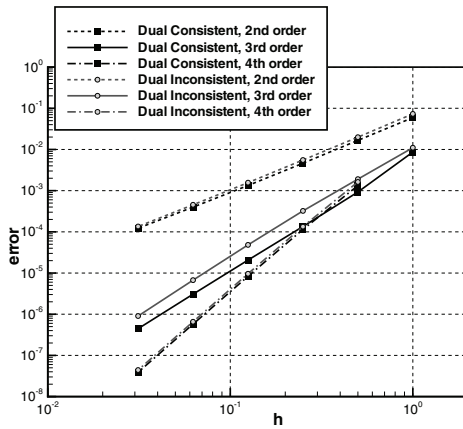


FIG. 5.7. Vortex flow, L^2 error in ρ .

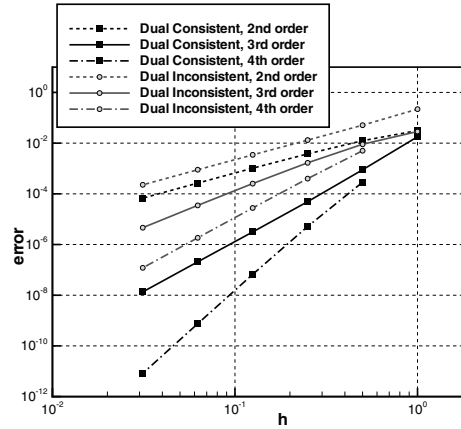


FIG. 5.8. Vortex flow functional error.

TABLE 5.4

Convergence rates of the dual consistent and dual inconsistent functional estimates for the vortex-flow example.

n	Dual consistent			Dual inconsistent		
	2nd	3rd	4th	2nd	3rd	4th
20	1.32696	4.32777	—	2.10279	1.67020	—
40	1.72834	4.17770	5.79869	1.95301	2.42847	3.66070
80	1.87561	3.96340	6.27190	1.93302	2.72474	3.84329
160	1.95240	3.91546	6.46379	1.95034	2.85689	3.90725
320	1.97473	3.93968	6.50283	1.97506	2.92272	3.93769

Next, examine the functional error, plotted in Figure 5.8. Here we see a significant difference in error between the dual-consistent and dual-inconsistent discretizations (note the different y -axis range from Figure 5.7). Table 5.4 lists the order of accuracy of the functional estimates. As predicted by the theory, the third- and fourth-order dual-consistent SBP discretizations produce superconvergent functionals.

6. Conclusions. Diagonal-norm SBP operators have interior stencils that are $2s$ -order accurate and boundary closures that are s -order accurate. This reduction in accuracy at the boundary is required to guarantee time stability. The boundary-stencil accuracy limits the global solution accuracy to $O(h^{s+1})$ at best.

Nevertheless, we have shown that SBP–SAT discretizations lead to functional estimates that are $O(h^{2s})$. To achieve this superconvergence, the functional and SATs must be carefully discretized to ensure that the SBP–SAT method is dual-consistent.

The superconvergence of SBP–SAT functional estimates was illustrated using a two-dimensional Poisson PDE and the Euler equations. The theory was verified in the presence of curvilinear grids and multiblock grids with interfaces. Dual-inconsistent boundary conditions and functional estimates were shown to destroy superconvergence. In addition, superconvergence requires the primal and adjoint solutions to be sufficiently smooth; this was illustrated by considering a particular functional in the Poisson PDE example.

Appendix A. Proof of Theorem 6.

THEOREM 6. *Let Ω_x and Ω_ξ be connected compact subsets of \mathbb{R}^2 . Furthermore, let $\mathcal{T} : \Omega_x \rightarrow \Omega_\xi$ be a C^{2s} invertible mapping. For grids based on the uniform discretization of Ω_ξ and mapped to Ω_x using \mathcal{T}^{-1} , the quadrature*

$$(4.3) \quad \mathcal{I}_h(u) = u(H \otimes H) J$$

is a $2s$ -order accurate estimate to the integral

$$(4.4) \quad \mathcal{I}(U) = \iint_{\Omega_x} U \, dx dy$$

for $U \in C^{2s}(\Omega_x)$.

Proof. Expanding the Jacobian in the change of variable theorem, we find

$$(A.1) \quad \begin{aligned} \iint_{\Omega_x} U \, dx dy &= \iint_{\Omega_\xi} U \mathcal{J} \, d\xi d\eta \\ &= \iint_{\Omega_\xi} U \frac{\partial x}{\partial \xi} \frac{\partial y}{\partial \eta} \, d\xi d\eta - \iint_{\Omega_\xi} U \frac{\partial y}{\partial \xi} \frac{\partial x}{\partial \eta} \, d\xi d\eta. \end{aligned}$$

Expressing the first integral as an iterated integral, we have

$$(A.2) \quad \iint_{\Omega_\xi} U \frac{\partial x}{\partial \xi} \frac{\partial y}{\partial \eta} \, d\xi d\eta = \int_{\eta=0}^1 \left[\int_{\xi=0}^1 \left(U \frac{\partial y}{\partial \eta} \right) \frac{\partial x}{\partial \xi} \, d\xi \right] \, d\eta.$$

If we associate the term $U(\partial y/\partial \eta)$ with \mathcal{Z} in Lemma 3, we can express the inner integral as

$$\begin{aligned} \int_{\xi=0}^1 \left(U \frac{\partial y}{\partial \eta} \right) \frac{\partial x}{\partial \xi} \, d\xi &= (\bar{u} \circ \bar{y}')^T H D \bar{x} + O(h^{2s}) \\ &= (\bar{u} \circ D \bar{x})^T H \bar{y}' + O(h^{2s}), \end{aligned}$$

where we have introduced the following vector functions of length $n + 1$:

$$\begin{aligned} \bar{u}(\eta) &= [U(\xi_0, \eta) \quad U(\xi_1, \eta) \quad \cdots \quad U(\xi_n, \eta)]^T, \\ \bar{y}'(\eta) &= \left[\frac{\partial y}{\partial \eta}(\xi_0, \eta) \quad \frac{\partial y}{\partial \eta}(\xi_1, \eta) \quad \cdots \quad \frac{\partial y}{\partial \eta}(\xi_n, \eta) \right]^T, \\ \bar{x}(\eta) &= [x(\xi_0, \eta) \quad x(\xi_1, \eta) \quad \cdots \quad x(\xi_n, \eta)]^T. \end{aligned}$$

For notational convenience, we define $\bar{w} \equiv H(\bar{u} \circ D\bar{x})$. Substituting the above expression for the inner integral into (A.2), we have

$$\begin{aligned} \int_{\eta=0}^1 \left[\int_{\xi=0}^1 \left(\mathcal{U} \frac{\partial y}{\partial \eta} \right) \frac{\partial x}{\partial \xi} d\xi \right] d\eta &= \int_{\eta=0}^1 \bar{w}^T \bar{y}' d\eta + O(h^{2s}) \\ &= \sum_{i=0}^n \int_{\eta=0}^1 \bar{w}_i(\eta) \frac{\partial y}{\partial \eta}(\xi_i, \eta) d\eta + O(h^{2s}). \end{aligned}$$

We can invoke Lemma 3 on each integral in the above sum. Hence,

$$\begin{aligned} \iint_{\Omega_\xi} \mathcal{U} \frac{\partial x}{\partial \xi} \frac{\partial y}{\partial \eta} d\xi d\eta &= w^T (H \otimes I)(D \otimes I)y + O(h^{2s}) \\ &= [u \circ (I \otimes D)x]^T (I \otimes H)(H \otimes I)(D \otimes I)y + O(h^{2s}) \\ \text{(A.3)} \quad &= u^T (H \otimes H) [(D_\xi x) \circ (D_\eta y)] + O(h^{2s}), \end{aligned}$$

where w, u, x , and y are vectors on the whole grid corresponding to the η discretization of $\bar{w}, \bar{u}, \bar{x}$, and \bar{y} , respectively.

The analysis above can be repeated to show

$$\text{(A.4)} \quad \iint_{\Omega_\xi} \mathcal{U} \frac{\partial y}{\partial \xi} \frac{\partial x}{\partial \eta} d\xi d\eta = u^T (H \otimes H) [(D_\xi y) \circ (D_\eta x)] + O(h^{2s}).$$

Using (A.3) and (A.4) in (A.1), we find

$$\begin{aligned} \iint_{\Omega_\xi} \mathcal{U} \mathcal{J} d\xi d\eta &= u(H \otimes H) [(D_\xi x) \circ (D_\eta y) - (D_\xi y) \circ (D_\eta x)] + O(h^{2s}) \\ &= u(H \otimes H)J + O(h^{2s}), \end{aligned}$$

which is the desired result \square

Appendix B. Dual-consistent tangential-flow boundary condition and force functional for the Euler equations. We will illustrate the appropriate dual-consistent inviscid-wall SAT boundary treatment using the linearized one-dimensional Euler equations

$$\text{(B.1)} \quad \frac{\partial}{\partial x} (A\mathcal{Q}) = 0 \quad \forall x \in [0, 1],$$

where $\mathcal{Q} = [\rho, \rho u, e]^T$ and A is the 3×3 (constant) flux Jacobian matrix evaluated at a fixed state:

$$A = \frac{\partial}{\partial \mathcal{Q}} \begin{pmatrix} \rho u \\ \rho u^2 + p \\ (e + p)u \end{pmatrix}.$$

We impose a boundary condition at $x = 0$ that mimics the tangential-flow boundary condition in two or three dimensions:

$$P\mathcal{Q} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \hat{n}\hat{n}^T & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathcal{Q} = 0,$$

where $\hat{n} = 1$. Note that P is an orthogonal projection matrix, and its definition remains valid in multiple dimensions provided \hat{n} is a unit normal vector to the solid wall.

The SBP–SAT discretization of (B.1) on a grid with $n + 1$ nodes is (ignoring the SAT at $x = 1$ for simplicity)

$$(D \otimes I_3) \Lambda q_h = - [(H^{-1} e_0 e_0^T) \otimes (I_3 - P)] \Lambda q_h - [(H^{-1} e_0 e_0^T) \otimes P] \Sigma q_h,$$

where $q_h \in \mathbb{R}^{3(n+1)}$ is the discrete solution vector and I_3 is the 3×3 identity matrix. The matrix Λ is block diagonal and holds the flux Jacobian matrix at each node, and Σ is a diagonal matrix containing the spectral radii, σ , of the flux Jacobian matrices. Specifically, we have that

$$\begin{aligned} \Lambda &= \text{diag}(A_0, A_1, \dots, A_n), \\ \Sigma &= \text{diag}(I_3 \sigma_0, I_3 \sigma_1, \dots, I_3 \sigma_n). \end{aligned}$$

The SAT penalty at $x = 0$ is unusual, insofar as it consists of a penalty on the flux and a penalty on the variable.

The above discretization will be dual-consistent for the pressure force at the left boundary if this functional is evaluated as

$$\mathcal{I}_h = -g^T P (A_0 - I_3 \sigma_0) q_0,$$

where $g^T = [0 \quad \hat{t}^T \quad 0]$. The unit normal vector \hat{t} gives the direction of the desired force; here it can take on only the scalar values 1 or -1 , but in multiple dimensions it will define the direction of lift or drag.

If we now follow the algebra in Theorem 4, but replace the scalar hyperbolic PDE with the linearized Euler equations, we arrive at the following discrete adjoint equation:

$$-\Lambda^T (D \otimes I_3) v_h = (\Lambda^T - \Sigma) [(H^{-1} e_0 e_0^T) \otimes P] (v_h - g).$$

The difference $\Lambda^T - \Sigma$ ensures that the adjoint SAT penalty is applied to waves that, in the primal equation, are exiting the domain or stagnant (one of the wave speeds is zero at the wall). This is a stable boundary treatment for the adjoint variables. The boundary condition enforced is $P(v_h - g) = 0$; i.e., the normal component of the adjoint variables corresponding to momentum is equal to the normal component of g . This is the correct boundary condition for the dual equations.

Acknowledgments. We sincerely thank the anonymous referees whose input and suggestions helped improve the clarity and quality of the manuscript. We also acknowledge and thank the associate editor for his time and effort during the review process.

REFERENCES

- [1] E. ANDERSON, Z. BAI, C. BISCHOF, S. BLACKFORD, J. DEMMEL, J. DONGARRA, J. DU CROZ, A. GREENBAUM, S. HAMMARLING, A. MCKENNEY, AND D. SORENSEN, *LAPACK Users' Guide*, 3rd ed., Software Environ. Tools 9, SIAM, Philadelphia, 1999.
- [2] J. W. BARRETT AND C. M. ELLIOTT, *Total flux estimates for a finite-element approximation of elliptic equations*, IMA J. Numer. Anal., 7 (1987), pp. 129–148.
- [3] J. H. BRAMBLE AND A. H. SCHATZ, *Higher order local accuracy by averaging in the finite element method*, Math. Comp., 31 (1977), pp. 94–111.

- [4] 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*, J. Comput. Phys., 111 (1994), pp. 220–236.
- [5] M. H. CARPENTER, J. NORDSTRÖM, AND D. GOTTLIEB, *A stable and conservative interface treatment of arbitrary spatial accuracy*, J. Comput. Phys., 148 (1999), pp. 341–365.
- [6] M. H. CARPENTER, J. NORDSTRÖM, AND D. GOTTLIEB, *Revisiting and extending interface penalties for multi-domain summation-by-parts operators*, J. Sci. Comput., 45 (2010), pp. 118–150.
- [7] C. DE BOOR AND B. SWARTZ, *Collocation at Gaussian points*, SIAM J. Numer. Anal., 10 (1973), pp. 582–606.
- [8] P. DIENER, E. N. DORBAND, E. SCHNETTER, AND M. TIGLIO, *Optimized high-order derivative and dissipation operators satisfying summation by parts, and applications in three-dimensional multi-block evolutions*, J. Sci. Comput., 32 (2007), pp. 109–145.
- [9] D. FUNARO AND D. GOTTLIEB, *A new method of imposing boundary conditions in pseudospectral approximations of hyperbolic equations*, Math. Comp., 51 (1988), pp. 599–613.
- [10] B. GUSTAFSSON, *The convergence rate for difference approximations to general mixed initial boundary value problems*, SIAM J. Numer. Anal., 18 (1981), pp. 179–190.
- [11] B. GUSTAFSSON, *High Order Difference Methods for Time Dependent PDE*, Springer-Verlag, Berlin, 2008.
- [12] J. E. HICKEN AND D. W. ZINGG, *A parallel Newton-Krylov solver for the Euler equations discretized using simultaneous approximation terms*, AIAA J., 46 (2008), pp. 2773–2786.
- [13] J. E. HICKEN AND D. W. ZINGG, *Summation-by-parts operators and high-order quadrature*, <http://arxiv.org/abs/1103.5182>.
- [14] R. KAMAKOTI AND C. PANTANO, *High-order narrow stencil finite-difference approximations of second-order derivatives involving variable coefficients*, SIAM J. Sci. Comput., 31 (2009), pp. 4222–4243.
- [15] 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, New York, 1974, pp. 195–212.
- [16] C. LANCZOS, *Linear Differential Operators*, D. Van Nostrand, London, 1961.
- [17] L. LEHNER, O. REULA, AND M. TIGLIO, *Multi-block simulations in general relativity: High-order discretizations, numerical stability and applications*, Classical Quantum Gravity, 22 (2005), pp. 5283–5321.
- [18] R. J. LEVEQUE, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*, SIAM, Philadelphia, 2007.
- [19] J. C.-C. LU, *An A Posteriori Error Control Framework for Adaptive Precision Optimization Using Discontinuous Galerkin Finite Element Method*, Massachusetts Institute of Technology, Cambridge, MA, 2005.
- [20] K. MATTSSON, *Boundary procedures for summation-by-parts operators*, J. Sci. Comput., 18 (2003), pp. 133–153.
- [21] K. MATTSSON AND J. NORDSTRÖM, *Summation by parts operators for finite difference approximations of second derivatives*, J. Comput. Phys., 199 (2004), pp. 503–540.
- [22] K. MATTSSON, M. SVÄRD, AND J. NORDSTRÖM, *Stable and accurate artificial dissipation*, J. Sci. Comput., 21 (2004), pp. 57–79.
- [23] K. MATTSSON, M. SVÄRD, AND M. SHOEBYBI, *Stable and accurate schemes for the compressible Navier-Stokes equations*, J. Comput. Phys., 227 (2008), pp. 2293–2316.
- [24] J. NORDSTRÖM AND M. H. CARPENTER, *High-order finite difference methods, multidimensional linear problems, and curvilinear coordinates*, J. Comput. Phys., 173 (2001), pp. 149–174.
- [25] J. NORDSTRÖM, J. GONG, E. VAN DER WEIDE, AND M. SVÄRD, *A stable and conservative high order multi-block method for the compressible Navier-Stokes equations*, J. Comput. Phys., 228 (2009), pp. 9020–9035.
- [26] M. OSUSKY, J. E. HICKEN, AND D. W. ZINGG, *A parallel Newton-Krylov-Schur flow solver for the Navier-Stokes equations using the SBP-SAT approach*, in 48th AIAA Aerospace Sciences Meeting, Orlando, FL, 2010, AIAA–2010–0116.
- [27] E. PAZOS, M. TIGLIO, M. D. DUEZ, L. E. KIDDER, AND S. A. TEUKOLSKY, *Orbiting binary black hole evolutions with a multipatch high order finite-difference approach*, Phys. Rev. D, 80 (2009), 024027.
- [28] N. A. PIERCE AND M. B. GILES, *Adjoint recovery of superconvergent functionals from PDE approximations*, SIAM Rev., 42 (2000), pp. 247–264.
- [29] T. H. PULLIAM AND J. L. STEGER, *Implicit finite-difference simulations of three-dimensional compressible flow*, AIAA J., 18 (1980), pp. 159–167.

- [30] M. D. SALAS AND H. L. ATKINS, *On problems associated with grid convergence of functionals*, *Comput. & Fluids*, 38 (2009), pp. 1445–1454.
- [31] B. STRAND, *Summation by parts for finite difference approximations for d/dx* , *J. Comput. Phys.*, 110 (1994), pp. 47–67.
- [32] M. SVÄRD, *On coordinate transformations for summation-by-parts operators*, *J. Sci. Comput.*, 20 (2004), pp. 29–42.
- [33] M. SVÄRD AND J. NORDSTRÖM, *On the order of accuracy for difference approximations of initial-boundary value problems*, *J. Comput. Phys.*, 218 (2006), pp. 333–352.
- [34] P. D. THOMAS AND C. K. LOMBARD, *Geometric conservation law and its application to flow computations on moving grids*, *AIAA J.*, 17 (1979), pp. 1030–1037.
- [35] V. THOMÉE AND B. WENDROFF, *Convergence estimates for Galerkin methods for variable coefficient initial value problems*, *SIAM J. Numer. Anal.*, 11 (1974), pp. 1059–1068.