Stability and Functional Superconvergence of Narrow-Stencil Second-Derivative Generalized Summation-By-Parts Discretizations

Zelalem Arega Worku[*] · **David W. Zingg**[*]

Abstract We analyze the stability and functional superconvergence of discretizations of diffusion problems with the narrow-stencil second-derivative generalized summation-by-parts (SBP) operators coupled with simultaneous approximation terms (SATs). Provided that the primal and adjoint solutions are sufficiently smooth and the SBP-SAT discretization is primal and adjoint consistent, we show that linear functionals associated with the steady diffusion problem superconverge at a rate of $2p$ when a degree $p+1$ narrow-stencil or a degree p wide-stencil generalized SBP operator is used for the spatial discretization. Sufficient conditions for stability of adjoint consistent discretizations with the narrow-stencil generalized SBP operators are presented. The stability analysis assumes nullspace consistency of the second-derivative operator and the invertibility of the matrix approximating the first derivative at the element boundaries. The theoretical results are verified by numerical experiments with the one-dimensional Poisson problem.

Keywords Summation-by-parts · Adjoint consistency · Simultaneous approximation term · Narrow-stencil · Functional superconvergence

Mathematics Subject Classification (2010) 65M06 · 65M12 · 65N06 · 65N12

1 Introduction

It has been observed in many cases that compared to wide-stencil[1] summation-by-parts (SBP) operators, explicitly formed narrow-stencil[2] second-derivative SBP operators provide smaller solution error, superior solution convergence rates, compact stencil width, and better damping of high frequency modes [30,31,27,12,13]. As with the wide-stencil operators, narrow-stencil second-derivative operators are coupled by simultaneous approximation terms (SATs) [6]. However, the SAT coefficients derived for wide-stencil SBP operators must be modified for implementations with narrow-stencil SBP operators to achieve stability and adjoint consistency simultaneously. Unfortunately, the analysis required to find such SAT coefficients for narrow-stencil SBP operators is more involved, e.g., see [13].

Hicken and Zingg [22] showed that adjoint consistent SBP-SAT discretizations of linear elliptic partial differential equations (PDEs) lead to functional superconvergence (see also [5,24,20]). In their study, they analyzed discretizations with wide-stencil second-derivative classical SBP (CSBP) operators by posing second-order linear PDEs as a system of first-order equations and determined

Zelalem Arega Worku
E-mail: zelalem.worku@mail.utoronto.ca

David W. Zingg
E-mail: dwz@oddjob.utias.utoronto.ca

[*]Institute for Aerospace Studies, University of Toronto, Toronto, Ontario, M3H 5T6, Canada

[1] Second-derivative operators formed by applying first-derivative operators twice.

[2] Also known as compact-stencil second-derivative operators.

the conditions that the SATs must satisfy for adjoint consistency and functional superconvergence. A similar analysis is conducted in [37] for multidimensional SBP operators, but without posing the second-order linear PDEs as a system of first-order equations. The latter approach enables analysis of functional accuracy of adjoint consistent discretizations of diffusion problems with narrow-stencil SBP operators. While the stability of discretizations arising from narrow-stencil second-derivative operators is well-studied (e.g., see [7,31,27,16,29,28]), it is only recently (see, e.g., [13,14]) that conditions for which such discretizations satisfy both stability and adjoint consistency requirements are presented. Eriksson [13], used the eigendecomposition technique to find the conditions on the SAT coefficients that enable construction of stable and adjoint consistent discretizations of diffusion problems. In a subsequent paper [14], Eriksson and Nordström used a variant of the approach in [13] to find a more general set of SAT coefficients. Although these SAT coefficients lead to adjoint consistent and stable discretizations, the analysis in [14] assumes positive definiteness of a component matrix of the discrete second derivative operator (i.e., M_k in (2.4)). However, this condition is not satisfied by many narrow-stencil second-derivative operators in the literature, including those in [30,29,27,12,28]. Furthermore, it is not straightforward how the theory extends to narrow-stencil generalized SBP operators which have one or more of the following characteristics: exclusion of one or both boundary nodes, non-repeating interior point operators, and non-uniform nodal distribution [12]. For a discussion on generalized SBP operators and a comparison of some of their properties with corresponding properties of classical SBP operators, we refer the reader to [10,12].

The first objective of this paper is to establish the conditions required for the stability of adjoint consistent SBP-SAT discretizations of diffusion problems with the generalized narrow-stencil second-derivative SBP operators. We use the "borrowing trick"[7] in the energy stability analysis which directly applies to the diagonal- and block-norm[3] narrow-stencil second-derivative SBP operators in [30,29,27,28] and to the generalized SBP operators of Del Rey Fernández and Zingg [12] upon minor modifications of the derivative operators at element boundaries. The second objective is to show that primal and adjoint consistent discretizations lead to functional convergence rates of $2p$ when a degree $p + 1$ narrow-stencil or a degree p wide-stencil diagonal-norm second-derivative generalized SBP operator is used to discretize steady diffusion problems for which the primal and adjoint solutions are sufficiently smooth. We also show that the functional converges at a rate of $2p$ irrespective of whether or not the scheme is adjoint consistent when a degree $2p - 1$ dense-norm wide- or narrow-stencil second-derivative SBP operator is used to discretize the spatial derivatives. Finally, we specialize the generalized form of the SATs given in [38,37] for one-dimensional implementation and provide penalty coefficients corresponding to a few known types of SAT such that they lead to consistent, adjoint consistent, conservative, and stable discretizations when coupled with the narrow-stencil second-derivative generalized SBP operators.

The paper is organized as follows. Section 2 presents the notation and some important definitions. In Section 3, we state the model problem and its SBP-SAT discretization. The main theoretical results that establish the functional superconvergence and energy stability of the SBP-SAT discretizations are presented in Section 4. The theoretical results are verified using the steady version of the model problem, the Poisson equation, in Section 5 and concluding remarks are presented in Section 6.

2 Preliminaries

We closely follow the notation used in [12,10,38,37]. A one-dimensional compact domain is considered, and it is tessellated into n_e non-overlapping elements, $\mathcal{T}_h := \{\{\Omega_k\}_{k=1}^{n_e} : \Omega = \cup_{k=1}^{n_e} \Omega_k\}$. The boundaries of each element will be referred to as interfaces, and we denote their union by $\Gamma_k := \partial\Omega_k$. The set of all interior interfaces is denoted by $\Gamma^I := \{\Gamma_k \cap \Gamma_v : k, v = 1, \dots, n_e, k \neq v\}$, while the element interfaces for which Dirichlet and Neumann boundary conditions are enforced are in the sets Γ^D and Γ^N, respectively, and $\Gamma := \Gamma^I \cup \Gamma^D \cup \Gamma^N$. Operators associated with the left and right interfaces of Ω_k bear the subscripts ℓ and r, respectively, and the left and right most elements are indicated by the subscripts L and R, respectively, e.g., $\mathsf{D}_{\ell L}$ is a derivative operator at the left interface of the left most element. The set of n_p volume nodes in element Ω_k is represented by $\boldsymbol{x}_k = \{x_i\}_{i=1}^{n_p}$. Uppercase script type, e.g., $\mathcal{U}_k \in \mathcal{C}^\infty(\Omega_k)$, is used for continuous functions, and $\mathbb{P}^p(\hat{\Omega})$ denotes the space of polynomials up to total degree p, which has a cardinality of $n_p^* = p + 1$. Bold letters, e.g., $\boldsymbol{u}_k \in \mathbb{R}^{n_p}$, delineate the restriction of $\mathcal{U}_k$ to grid points $\boldsymbol{x}_k$, while solution vectors to the discrete systems of equations have subscript h, e.g., $\boldsymbol{u}_{h,k} \in \mathbb{R}^{n_p}$. For the purpose of the functional convergence analysis in Section 4.3, we define $h := \max_{a,b\in\boldsymbol{x}_k} |a-b|$ as the size of an element. Matrices are denoted by sans-serif uppercase letters, e.g., $\mathsf{V} \in \mathbb{R}^{n_p \times n_p}$; $\mathbf{1}$ denotes a vector consisting of all ones, $\mathbf{0}$ denotes a vector or matrix consisting of all zeros. The sizes of $\mathbf{1}$ and $\mathbf{0}$ should be clear from context.

---

[3] Also referred to as full-norm matrix.

Definitions of the first- and second-derivative SBP operators presented in [12] are stated below. For the construction of narrow-stencil second-derivative SBP operators, we refer the reader to [10, 12, 27, 28]. $D_k \boldsymbol{p} = \frac{\partial}{\partial \mathcal{P}} x$ $D_k \boldsymbol{p} = \frac{\partial \mathcal{P}}{\partial x}$

Definition 2.1 (Generalized first-derivative SBP operator) The matrix $D_k \in \mathbb{R}^{n_p \times n_p}$ is a degree p SBP operator approximating the first derivative $\frac{\partial}{\partial x}$ on the set of nodes $\boldsymbol{x}_k$, which need neither be uniform nor include nodes on the boundaries and may have nodes outside the domain of element Ω_k, if [12]

1. $D_k \boldsymbol{p} = \frac{\partial \mathcal{P}}{\partial x}$ for all $\mathcal{P} \in \mathbb{P}^p(\Omega_k)$
2. $D_k = H_k^{-1} Q_k$, where H_k is a symmetric positive definite (SPD) matrix, and
3. $Q_k = S_k + \frac{1}{2} E_k$, where $S_k = -S_k^T$, $E_k = E_k^T$, and E_k satisfies $\boldsymbol{p}^T E_k \boldsymbol{q} = \sum_{\gamma \in \Gamma_k} [\mathcal{P}]_\gamma [\mathcal{Q}]_\gamma n_{\gamma k}$ for all $\mathcal{P}, \mathcal{Q} \in \mathbb{P}^\tau(\Omega_k)$, where $\tau \geq p$, and $n_{\gamma k} = 1$ if γ is the right interface of Ω_k, otherwise $n_{\gamma k} = -1$.

The norm matrix, H_k, may be diagonal or dense. A dense-norm matrix refers to any norm matrix that is not diagonal, which includes the block-norm matrix. The block-norm matrix has diagonal entries at the interior points (containing h) and dense blocks at the top-left and bottom-right corners corresponding to the boundary nodes. The L^2 inner product of two functions $\mathcal{P}$ and $\mathcal{Q}$ is approximated by [23, 10, 21, 11]

$$\boldsymbol{p}^T H_k \boldsymbol{q} = \int_{\Omega_k} \mathcal{P}\mathcal{Q} \, \mathrm{d}\Omega + \mathcal{O}\left(h^{2p}\right),$$

and H_k defines the norm

$$\boldsymbol{u}^T H_k \boldsymbol{u} = \|\boldsymbol{u}\|_H^2 = \int_{\Omega_k} \mathcal{U}^2 \, \mathrm{d}\Omega + \mathcal{O}\left(h^{2p}\right).$$

The E_k matrix is constructed as [10, 11]

$$E_k = \sum_{\gamma \subset \Gamma_k} n_{\gamma k} R_{\gamma k}^T R_{\gamma k} = R_{rk}^T R_{rk} - R_{\ell k}^T R_{\ell k}, \tag{2.1}$$

where $R_{\gamma k}$ is an extrapolation row vector of at least order $h^{\tau+1}$ accuracy, i.e., $R_{\gamma k} \boldsymbol{u}_k = [\mathcal{U}_k]_\gamma + \mathcal{O}(h^{\geq \tau+1})$. Furthermore, we define an operator that extrapolates the product of the diffusion coefficient and the derivative of the solution from volume nodes to an interface as

$$D_{\gamma k} := n_{\gamma k} R_{\gamma k} \Lambda_k D_{b,k}. \tag{2.2}$$

Definition 2.2 (Order-matched narrow-stencil second-derivative generalized SBP operator) The narrow-stencil second-derivative operator $D_k^{(2)}(\Lambda)$ of degree $p+1$, approximating $\frac{\partial}{\partial x}(\lambda_k \frac{\partial \mathcal{U}_k}{\partial x})$, is order-matched with the first-derivative operator $D_k = H_k^{-1} Q_k$ of degree p on the nodal set $\boldsymbol{x}_k$ if [12]

$$D_k^{(2)}(\Lambda) \boldsymbol{p}_k = \frac{\partial}{\partial x}\left(\lambda_k \frac{\partial \mathcal{P}_k}{\partial x}\right), \quad \forall \, (\lambda_k \mathcal{P}_k) \in \mathbb{P}^{p+1}(\Omega_k), \tag{2.3}$$

and $D_k^{(2)}(\Lambda)$ is of the form

$$D_k^{(2)}(\Lambda) = H_k^{-1}[-M_k + E_k \Lambda_k D_{b,k}], \tag{2.4}$$

where $M_k = \sum_{i=1}^{n_p} \Lambda_k(i,i) \bar{M}_i$, $\bar{M}_i$ are symmetric positive semidefinite matrices,

$$\Lambda_k = \mathrm{diag}(\lambda_k(x_1), \lambda_k(x_2), ..., \lambda_k(x_{n_p})),$$

and $D_{b,k}$ is an approximation to the first derivative of degree and order $\geq p+1$.

The order-matched SBP operators in Definition 2.2 are assumed to have a diagonal-norm matrix. Note that for the mth derivative, the degree and order are related by $order = degree - m + 1$; consequently, both the diagonal-norm narrow-stencil $D_k^{(2)}(\Lambda)$ and D_k operators are order p accurate, while a diagonal-norm wide-stencil second-derivative operator, which has the decomposition

$$D_k \Lambda_k D_k = H_k^{-1}[-D_k^T H_k \Lambda_k D_k + E_k \Lambda_k D_k], \tag{2.5}$$

is order $p-1$ accurate [12]. Similar to the diagonal-norm SBP operators, block-norm SBP operators have an order $2p$ centered-difference interior operator. At the boundaries, however, the block-norm wide- and narrow-stencil second-derivative operators are closed with order $2p-2$ one-sided stencils, unlike the order $p-1$ and p one-sided stencils used with the diagonal-norm wide- and narrow-stencil SBP operators, respectively. Furthermore, the $D_{b,k}$ matrix of a block-norm operator contains order $2p-1$ approximations of the first derivative at rows corresponding to the boundary nodes (see, e.g., [28, 30] for definition and discussion regarding the block-norm SBP operators).

Remark 2.1 In this work, we do not assume that M_k is necessarily symmetric positive semidefinite; rather we assume that $\mathsf{M}_k + \mathsf{M}_k^T$ is symmetric positive semidefinite, which allows the analysis to be extended to a more general class of explicitly formed second-derivative operators, including the block-norm SBP operators in [28,30], which do not have symmetric M_k matrix.

Another decomposition of second-derivative SBP operators, which is instrumental for the adjoint consistency and functional superconvergence analyses in Section 4, is presented below.

Proposition 2.1 *A second-derivative operator of the form* (2.4)*, for which* M_k *is not necessarily symmetric, can be written as*

$$
\begin{aligned}
\mathsf{D}_k^{(2)}(\Lambda) = \mathsf{H}_k^{-1}\left(\mathsf{D}_k^{(2)}(\Lambda)\right)^T \mathsf{H}_k - \mathsf{H}_k^{-1}\mathsf{D}_{rk}^T\mathsf{R}_{rk} - \mathsf{H}_k^{-1}\mathsf{D}_{\ell k}^T\mathsf{R}_{\ell k} \\
+ \mathsf{H}_k^{-1}\mathsf{R}_{rk}^T\mathsf{D}_{rk} + \mathsf{H}_k^{-1}\mathsf{R}_{\ell k}^T\mathsf{D}_{\ell k} - \mathsf{H}_k^{-1}\left(\mathsf{M}_k - \mathsf{M}_k^T\right).
\end{aligned}
\tag{2.6}
$$

Proof The proof is given in Appendix A.

3 Model Problem and SBP-SAT Discretization

We consider the one-dimensional diffusion problem

$$
\begin{aligned}
\frac{\partial \mathcal{U}}{\partial t} - \frac{\partial}{\partial x}\left(\lambda\frac{\partial \mathcal{U}}{\partial x}\right) = \mathcal{F} \quad \forall x \in \Omega, \quad \mathcal{U} = \mathcal{U}_0 \text{ at } t = 0, \\
\mathcal{U}|_{\Gamma^D} = \mathcal{U}_D, \\
n_\gamma\left(\lambda\frac{\partial \mathcal{U}}{\partial x}\right)\Big|_{\Gamma^N} = \mathcal{U}_N,
\end{aligned}
\tag{3.1}
$$

where $\mathcal{F} \in L^2(\Omega)$, $\lambda = \lambda(x)$ is a positive diffusivity coefficient, and Γ^D is not empty. For functional error analysis and numerical experiment purposes, we consider the steady version of (3.1), the Poisson problem. We also consider a compatible linear functional of the form

$$
\mathcal{I}(\mathcal{U}) = \int_\Omega \mathcal{G}\mathcal{U}\,\mathrm{d}\Omega - \psi_D\left[\lambda\frac{\partial \mathcal{U}}{\partial x}n_\gamma\right]_{\Gamma^D} + \psi_N\mathcal{U}|_{\Gamma^N},
\tag{3.2}
$$

where $\mathcal{G} \in L^2(\Omega)$, $\psi_N = n_\gamma(\lambda\frac{\partial \psi}{\partial x}) \in L^2(\Gamma^N)$, and $\psi_D \in L^2(\Gamma^D)$. A linear functional is compatible with the steady version of (3.1) if [18]

$$
\begin{aligned}
\int_\Omega \psi\frac{\partial}{\partial x}\left(\lambda\frac{\partial \mathcal{U}}{\partial x}\right)\mathrm{d}\Omega + \mathcal{U}_D\left[\lambda\frac{\partial \psi}{\partial x}n_\gamma\right]_{\Gamma^D} - \mathcal{U}_N\psi|_{\Gamma^N} \\
= \int_\Omega \mathcal{U}\frac{\partial}{\partial x}\left(\lambda\frac{\partial \psi}{\partial x}\right)\mathrm{d}\Omega + \psi_D\left[\lambda\frac{\partial \mathcal{U}}{\partial x}n_\gamma\right]_{\Gamma^D} - \psi_N\mathcal{U}|_{\Gamma^N},
\end{aligned}
\tag{3.3}
$$

i.e.,

$$
\mathcal{I}(\mathcal{U}) = \mathcal{I}(\psi) = \int_\Omega \psi\mathcal{F}\mathrm{d}\Omega - \mathcal{U}_D\left[\lambda\frac{\partial \psi}{\partial x}n_\gamma\right]_{\Gamma^D} + \mathcal{U}_N\psi|_{\Gamma^N}.
\tag{3.4}
$$

Under the compatibility condition on the functional, the adjoint, ψ, satisfies the PDE (see, e.g., [22, 38,18])

$$
-\frac{\partial}{\partial x}\left(\lambda\frac{\partial \psi}{\partial x}\right) = \mathcal{G} \quad \forall x \in \Omega, \quad \psi|_{\Gamma^D} = \psi_D, \quad n_\gamma\left(\lambda\frac{\partial \psi}{\partial x}\right)\Big|_{\Gamma^N} = \psi_N.
\tag{3.5}
$$

The SBP-SAT semi-discretization of the diffusion problem, (3.1), is given by

$$
\frac{\mathrm{d}\boldsymbol{u}_{h,k}}{\mathrm{d}t} = \mathsf{D}_k^{(2)}(\Lambda)\boldsymbol{u}_{h,k} + \boldsymbol{f}_k - \mathsf{H}_k^{-1}\boldsymbol{s}_k^I(\boldsymbol{u}_{h,k}) - \mathsf{H}_k^{-1}\boldsymbol{s}_k^B(\boldsymbol{u}_{h,k}, u_D, u_N) =: R_{h,u},
\tag{3.6}
$$

where $\boldsymbol{f}_k$ is the restriction of $\mathcal{F}$ to the volume nodes in Ω_k and the interface SATs, $\boldsymbol{s}_k^I$, and boundary SATs, $\boldsymbol{s}_k^B$, given in [38,37] are specialized for one-dimensional implementation as

$$
\boldsymbol{s}_k^I(\boldsymbol{u}_{h,k}) = \sum_{\gamma \subset \Gamma_k^I}\begin{bmatrix}\mathsf{R}_{\gamma k}^T & \mathsf{D}_{\gamma k}^T\end{bmatrix}\begin{bmatrix}\mathsf{T}_{\gamma k}^{(1)} & \mathsf{T}_{\gamma k}^{(3)} \\ \mathsf{T}_{\gamma k}^{(2)} & \mathsf{T}_{\gamma k}^{(4)}\end{bmatrix}\begin{bmatrix}\mathsf{R}_{\gamma k}\boldsymbol{u}_{h,k} - \mathsf{R}_{\gamma v}\boldsymbol{u}_{h,v} \\ \mathsf{D}_{\gamma k}\boldsymbol{u}_{h,k} + \mathsf{D}_{\gamma v}\boldsymbol{u}_{h,v}\end{bmatrix}
\tag{3.7}
$$

and

$$s_k^B(\boldsymbol{u}_{h,k}, \boldsymbol{u}_D, \boldsymbol{u}_N) = \left\{ \begin{bmatrix} \mathsf{R}_{\gamma k}^T & \mathsf{D}_{\gamma k}^T \end{bmatrix} \begin{bmatrix} \mathsf{T}_{\gamma k}^{(D)} \\ -1 \end{bmatrix} (\mathsf{R}_{\gamma k} \boldsymbol{u}_{h,k} - u_D) \right\}_{\gamma \subset \Gamma^D}$$
$$+ \left\{ \mathsf{R}_{\gamma k}^T (\mathsf{D}_{\gamma k} \boldsymbol{u}_{h,k} - u_N) \right\}_{\gamma \subset \Gamma^N} . \tag{3.8}$$

The SAT coefficients $\mathsf{T}_{\gamma k}^{(1)}, \mathsf{T}_{\gamma k}^{(2)}, \mathsf{T}_{\gamma k}^{(3)}, \mathsf{T}_{\gamma k}^{(4)}, \mathsf{T}_{\gamma k}^{(D)} \in \mathbb{R}$ are determined such that the scheme satisfies desired properties such as conservation, adjoint consistency, and energy stability. For implementations with wide-stencil operators, we replace $\mathsf{D}_k^{(2)}(\Lambda)$ by $\mathsf{D}_k \Lambda_k \mathsf{D}_k$ in (3.6) and $\mathsf{D}_{b,k}$ by D_k in (2.2).

Substituting the restriction of a sufficiently smooth solution to grid points, $\boldsymbol{u}_k$, into (3.6) to (3.8), we see that the right-hand side (RHS) of (3.6) yields a discretization error of $\mathcal{O}(h^p)$ when an order-matched narrow-stencil second-derivative SBP operator is used; hence, the discretization of the primal problem is consistent. In contrast, for diagonal-norm wide-stencil SBP operators, the discretization error is $\mathcal{O}(h^{p-1})$ while for block-norm second-derivative SBP operators, it is $\mathcal{O}(h^{2p-2})$.

4 Theoretical Results

In this section, we present the two main results of this paper. After establishing the conditions required for adjoint consistency and conservation, we show that primal and adjoint consistent SBP-SAT discretizations of the Poisson problem with the diagonal-norm narrow-stencil second-derivative operators lead to functional superconvergence. To achieve this goal, we closely follow the technique used to show functional superconvergence in [37]. Then, we use the energy method to find sufficient conditions that the SATs must satisfy for the stability of discretizations with narrow-stencil generalized SBP operators before stating a few concrete examples of such SATs.

4.1 Adjoint Consistency

Adjoint consistency requires that the discrete adjoint problem,

$$\sum_{\Omega_k \in \mathcal{T}_h} \left(L_{h,k}^*(\boldsymbol{\psi}_h) - \boldsymbol{g}_k \right) = \boldsymbol{0}, \tag{4.1}$$

where $L_{h,k}^*$ is the discrete adjoint operator, corresponding to the steady version of the primal problem (3.6) satisfies

$$\lim_{h \to 0} \sum_{\Omega_k \in \mathcal{T}_h} \left\| L_{h,k}^*(\psi_k) - \boldsymbol{g}_k \right\|_{\mathsf{H}_k} = 0. \tag{4.2}$$

To find the discrete adjoint operator, we begin by discretizing the two forms of the functional, (3.2) and (3.4), as

$$I_h(\boldsymbol{u}_h) = \sum_{\Omega_k \in \mathcal{T}_h} \boldsymbol{g}_k^T \mathsf{H}_k \boldsymbol{u}_{h,k} - \psi_D \mathsf{D}_{\ell L} \boldsymbol{u}_{h,L} + \psi_N \mathsf{R}_{rR} \boldsymbol{u}_{h,R}$$
$$+ \psi_D \mathsf{T}_{\ell L}^{(D)} (\mathsf{R}_{\ell L} \boldsymbol{u}_{h,L} - u_D), \tag{4.3}$$
$$I_h(\boldsymbol{\psi}_h) = \sum_{\Omega_k \in \mathcal{T}_h} \boldsymbol{f}_k^T \mathsf{H}_k \boldsymbol{\psi}_{h,k} - u_D \mathsf{D}_{\ell L} \boldsymbol{\psi}_{h,L} + u_N \mathsf{R}_{rR} \boldsymbol{\psi}_{h,R}$$
$$+ u_D \mathsf{T}_{\ell L}^{(D)} (\mathsf{R}_{\ell L} \boldsymbol{\psi}_{h,L} - \psi_D), \tag{4.4}$$

where we have assumed that the Dirichlet and Neumann boundary conditions are enforced on the left and right boundaries, respectively. The last terms in (4.3) and (4.4) arise from consistent modifications of the functional, see [18,22,20,38,37]. Note that in cases where a Dirichlet boundary condition is enforced on both boundaries, we apply the Dirichlet SATs given in (3.8) on both boundaries and modify the discrete functionals, (4.3) and (4.4), by replacing the Neumann boundary terms by Dirichlet right boundary terms similar to those given for the left boundary. The theory developed holds for such cases without significant modification.

To derive the conditions required for adjoint consistency, we set $I_h(\boldsymbol{u}_h) - I_h(\boldsymbol{\psi}_h) = 0$, which is a discrete analogue of the relation $\mathcal{I}(\mathcal{U}) - \mathcal{I}(\psi) = 0$. Adding $\sum_{\Omega_k \subset \mathcal{T}_h} \boldsymbol{\psi}_{h,k}^T \mathsf{H}_k R_{h,k} + I_h(\boldsymbol{\psi}_h) - I_h(\boldsymbol{\psi}_h) = 0$

to the RHS of (4.3) and rearranging we find

$$
\begin{aligned}
I_h\left(\boldsymbol{u}_h\right) = {} & \sum_{\Omega_k \subset \mathcal{T}_h} \boldsymbol{g}_k^T \mathsf{H}_k \boldsymbol{u}_{h,k} - \psi_D \mathsf{D}_{\ell L} \boldsymbol{u}_{h,L} + \psi_N \mathsf{R}_{rR} \boldsymbol{u}_{h,R} \\
& + \psi_D \mathsf{T}_{\ell L}^{(D)} \left(\mathsf{R}_{\ell L} \boldsymbol{u}_{h,L} - u_D\right) + u_D \mathsf{D}_{\ell L} \boldsymbol{\psi}_{h,L} - u_N \mathsf{R}_{rR} \boldsymbol{\psi}_{h,R} \\
& + \sum_{\Omega_k \subset \mathcal{T}_h} \left[\boldsymbol{\psi}_{h,k}^T \mathsf{H}_k \mathsf{D}_k^{(2)}(\Lambda) \boldsymbol{u}_{h,k} - \boldsymbol{\psi}_{h,k}^T \boldsymbol{s}_k^I(\boldsymbol{u}_{h,k}) - \boldsymbol{\psi}_{h,k}^T \boldsymbol{s}_k^B(\boldsymbol{u}_{h,k}, \boldsymbol{u}_D, \boldsymbol{u}_N) \right] \\
& - u_D \mathsf{T}_{\ell L}^{(D)} \left(\mathsf{R}_{\ell L} \boldsymbol{\psi}_{h,L} - \psi_D\right) + I_h\left(\boldsymbol{\psi}_h\right).
\end{aligned}
\tag{4.5}
$$

Transposing (4.5), enforcing $I_h(\boldsymbol{u}_h) - I_h(\boldsymbol{\psi}_h) = 0$, applying identity (2.6), and simplifying, we obtain

$$
\begin{aligned}
& \sum_{\Omega_k \subset \mathcal{T}_h} \left\{ \boldsymbol{u}_{h,k}^T \mathsf{H}_k \left(\mathsf{D}_k^{(2)}(\Lambda) \boldsymbol{\psi}_{h,k} + \boldsymbol{g}_k \right) + \boldsymbol{u}_{h,k}^T \left(\mathsf{M}_k - \mathsf{M}_k^T \right) \boldsymbol{\psi}_{h,k} \right\} \\
& - \sum_{\gamma \subset \Gamma^I} \begin{bmatrix} \mathsf{R}_{\gamma k} \boldsymbol{u}_{h,k} \\ \mathsf{R}_{\gamma v} \boldsymbol{u}_{h,v} \\ \mathsf{D}_{\gamma k} \boldsymbol{u}_{h,k} \\ \mathsf{D}_{\gamma v} \boldsymbol{u}_{h,v} \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{\gamma k}^{(1)} & -\mathsf{T}_{\gamma v}^{(1)} & \mathsf{T}_{\gamma k}^{(2)}+1 & -\mathsf{T}_{\gamma v}^{(2)} \\ -\mathsf{T}_{\gamma k}^{(1)} & \mathsf{T}_{\gamma v}^{(1)} & -\mathsf{T}_{\gamma k}^{(2)} & \mathsf{T}_{\gamma v}^{(2)}+1 \\ \mathsf{T}_{\gamma k}^{(3)}-1 & \mathsf{T}_{\gamma v}^{(3)} & \mathsf{T}_{\gamma k}^{(4)} & \mathsf{T}_{\gamma v}^{(4)} \\ \mathsf{T}_{\gamma k}^{(3)} & \mathsf{T}_{\gamma v}^{(3)}-1 & \mathsf{T}_{\gamma k}^{(4)} & \mathsf{T}_{\gamma v}^{(4)} \end{bmatrix} \begin{bmatrix} \mathsf{R}_{\gamma k} \boldsymbol{\psi}_{h,k} \\ \mathsf{R}_{\gamma v} \boldsymbol{\psi}_{h,v} \\ \mathsf{D}_{\gamma k} \boldsymbol{\psi}_{h,k} \\ \mathsf{D}_{\gamma v} \boldsymbol{\psi}_{h,v} \end{bmatrix} \\
& - \boldsymbol{u}_{h,L}^T \mathsf{R}_{\ell L}^T \mathsf{T}_{\ell L}^{(D)} \left(\mathsf{R}_{\ell L} \boldsymbol{\psi}_{h,L} - \psi_D\right) + \boldsymbol{u}_{h,L}^T \mathsf{D}_{\ell L}^T \left(\mathsf{R}_{\ell L} \boldsymbol{\psi}_{h,L} - \psi_D\right) \\
& - \boldsymbol{u}_{h,R}^T \mathsf{R}_{rR}^T \left(\mathsf{D}_{rR} \boldsymbol{\psi}_{h,k} - \psi_N\right) = 0,
\end{aligned}
\tag{4.6}
$$

from which we extract the discrete adjoint operator on element Ω_k as

$$
\begin{aligned}
L_{h,k}^*(\psi_h) = {} & -\mathsf{D}_k^{(2)}(\Lambda) \boldsymbol{\psi}_{h,k} - \mathsf{H}_k^{-1}(\mathsf{M}_k - \mathsf{M}_k^T) \boldsymbol{\psi}_{h,k} \\
& + \mathsf{H}_k^{-1}(\boldsymbol{s}_k^I)^*(\boldsymbol{\psi}_{h,k}) + \mathsf{H}_k^{-1}(\boldsymbol{s}_k^B)^*(\boldsymbol{\psi}_{h,k}, \psi_D, \psi_N),
\end{aligned}
\tag{4.7}
$$

where the interface and boundary SATs for the adjoint problem are given, respectively, by

$$
\left(\boldsymbol{s}_k^I\right)^* = \sum_{\gamma \subset \Gamma_k^I} \begin{bmatrix} \mathsf{R}_{\gamma k}^T & \mathsf{D}_{\gamma k}^T \end{bmatrix} \begin{bmatrix} \mathsf{T}_{\gamma k}^{(1)} & -\mathsf{T}_{\gamma v}^{(1)} & \mathsf{T}_{\gamma k}^{(2)}+1 & -\mathsf{T}_{\gamma v}^{(2)} \\ \mathsf{T}_{\gamma k}^{(3)}-1 & \mathsf{T}_{\gamma v}^{(3)} & \mathsf{T}_{\gamma k}^{(4)} & \mathsf{T}_{\gamma v}^{(4)} \end{bmatrix} \begin{bmatrix} \mathsf{R}_{\gamma k} \boldsymbol{\psi}_{h,k} \\ \mathsf{R}_{\gamma v} \boldsymbol{\psi}_{h,v} \\ \mathsf{D}_{\gamma k} \boldsymbol{\psi}_{h,k} \\ \mathsf{D}_{\gamma v} \boldsymbol{\psi}_{h,v} \end{bmatrix},
\tag{4.8}
$$

$$
\begin{aligned}
\left(\boldsymbol{s}_k^B\right)^* = {} & \left\{ \begin{bmatrix} \mathsf{R}_{\gamma k}^T & \mathsf{D}_{\gamma k}^T \end{bmatrix} \begin{bmatrix} \mathsf{T}_{\gamma k}^{(D)} \\ -1 \end{bmatrix} \begin{bmatrix} \mathsf{R}_{\gamma k} \boldsymbol{\psi}_{h,k} - \psi_D \end{bmatrix} \right\}_{\gamma \subset \Gamma^D} \\
& + \left\{ \mathsf{R}_{\gamma k}^T \left(\mathsf{D}_{\gamma k} \boldsymbol{\psi}_{h,k} - \psi_N\right) \right\}_{\gamma \subset \Gamma^N}.
\end{aligned}
\tag{4.9}
$$

Furthermore, we define the residual of the SBP-SAT discretization of the adjoint problem as

$$
\begin{aligned}
R_{h,\psi} := {} & \mathsf{D}_k^{(2)} \boldsymbol{\psi}_{h,k} + \boldsymbol{g}_k + \mathsf{H}_k^{-1}(\mathsf{M}_k - \mathsf{M}_k^T)\psi_{h,k} \\
& - \mathsf{H}_k^{-1}(\boldsymbol{s}_k^I)^*(\boldsymbol{\psi}_{h,k}) - \mathsf{H}_k^{-1}(\boldsymbol{s}_k^B)^*(\boldsymbol{\psi}_{h,k}, \psi_D, \psi_N) = \boldsymbol{0},
\end{aligned}
\tag{4.10}
$$

Substituting the exact adjoint solution, $\boldsymbol{\psi}$, into (4.10) and provided that $\boldsymbol{\psi}$ is sufficiently smooth, we observe that $R_{h,\psi}$ is $\mathcal{O}(h^{\geq p-1})$, i.e., the discretization of the adjoint problem is consistent, if and only if

$$
\mathsf{T}_{\gamma k}^{(1)} = \mathsf{T}_{\gamma v}^{(1)}, \qquad \mathsf{T}_{\gamma k}^{(2)} + 1 = -\mathsf{T}_{\gamma v}^{(2)}, \qquad \mathsf{T}_{\gamma k}^{(3)} - 1 = -\mathsf{T}_{\gamma v}^{(3)}, \qquad \mathsf{T}_{\gamma k}^{(4)} = \mathsf{T}_{\gamma v}^{(4)}, \qquad \mathsf{M}_k = \mathsf{M}_k^T.
\tag{4.11}
$$

For discretizations with wide-stencil second-derivative operators, the last condition, $\mathsf{M}_k = \mathsf{M}_k^T$, is satisfied by default.

4.2 Conservation

For conservation, the homogeneous diffusion problem (3.1), i.e., $\mathcal{F} = 0$, should satisfy Gauss's theorem discretely, i.e., $\sum_{\Omega_k \subset \mathcal{T}_h} \mathbf{1}^T \mathsf{H}_k \mathrm{d}\boldsymbol{u}_k / \mathrm{d}t$ must depend only on the boundary terms. Premultiplying $R_{h,u}$ defined in (3.6) by $\mathbf{1}^T \mathsf{H}_k$, setting $\boldsymbol{f}_k = 0$, summing over all elements, and applying the decomposition of $\mathsf{D}_k^{(2)}(\Lambda)$ given in (2.4) yields

$$
\sum_{\Omega_k \subset \mathcal{T}_h} \mathbf{1}^T \mathsf{H}_k R_{h,u} = -\sum_{\gamma \subset \Gamma^I} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{\gamma k}^{(1)} & -\mathsf{T}_{\gamma k}^{(1)} & \mathsf{T}_{\gamma k}^{(3)} - 1 & \mathsf{T}_{\gamma k}^{(3)} \\ -\mathsf{T}_{\gamma v}^{(1)} & \mathsf{T}_{\gamma v}^{(1)} & \mathsf{T}_{\gamma v}^{(3)} & \mathsf{T}_{\gamma v}^{(3)} - 1 \\ \mathsf{T}_{\gamma k}^{(2)} & -\mathsf{T}_{\gamma k}^{(2)} & \mathsf{T}_{\gamma k}^{(4)} & \mathsf{T}_{\gamma k}^{(4)} \\ -\mathsf{T}_{\gamma v}^{(2)} & \mathsf{T}_{\gamma v}^{(2)} & \mathsf{T}_{\gamma v}^{(4)} & \mathsf{T}_{\gamma v}^{(4)} \end{bmatrix} \begin{bmatrix} \mathsf{R}_{\gamma k} \boldsymbol{u}_{h,k} \\ \mathsf{R}_{\gamma v} \boldsymbol{u}_{h,v} \\ \mathsf{D}_{\gamma k} \boldsymbol{u}_{h,k} \\ \mathsf{D}_{\gamma v} \boldsymbol{u}_{h,v} \end{bmatrix}
$$
$$
- \sum_{\Omega_k \subset \mathcal{T}_h} \mathbf{1}^T \mathsf{M}_k \boldsymbol{u}_{h,k} - \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_\gamma^D & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \mathsf{R}_{\gamma k} \boldsymbol{u}_k - u_D \\ \mathsf{D}_{\gamma k} \boldsymbol{u}_k \end{bmatrix} \right\}_{\gamma \subset \Gamma^D} + u_N,
\tag{4.12}
$$

which reduces to a sum of boundary terms only,

$$
\sum_{\Omega_k \subset \mathcal{T}_h} \mathbf{1}^T \mathsf{H}_k R_{h,u} = \left\{ \mathsf{D}_{\gamma k} \boldsymbol{u}_k - \mathsf{T}_\gamma^D \left(\mathsf{R}_{\gamma k} \boldsymbol{u}_k - u_D \right) \right\}_{\gamma \subset \Gamma^D} + u_N,
\tag{4.13}
$$

as required for conservation of the discretization if

$$
\mathsf{T}_{\gamma k}^{(1)} = \mathsf{T}_{\gamma v}^{(1)}, \qquad\qquad \mathsf{T}_{\gamma k}^{(3)} - 1 = -\mathsf{T}_{\gamma v}^{(3)}, \qquad\qquad \mathbf{1}^T \mathsf{M}_k = \mathbf{0}.
\tag{4.14}
$$

Comparing (4.14) and (4.11) and noting that $\mathsf{M}_k \mathbf{1} = \mathbf{0}$, we see that adjoint consistency implies conservation, as noted in [2, 19, 37].

4.3 Functional Superconvergence

Without loss of generality, we assume that the domain is tessellated using two elements, Ω_L and Ω_R. In the subsequent analysis, we will use the vectors $\boldsymbol{u}, \boldsymbol{u}_h, \boldsymbol{\psi}, \boldsymbol{\psi}_h, \boldsymbol{f}, \boldsymbol{g}, \mathbb{E}(\boldsymbol{u}_h), \mathbb{F}(\boldsymbol{\psi}_h) \in \mathbb{R}^{2n_p}$ given by

$$
\boldsymbol{u}_h = \begin{bmatrix} \boldsymbol{u}_{h,L} \\ \boldsymbol{u}_{h,R} \end{bmatrix}, \quad \boldsymbol{\psi}_h = \begin{bmatrix} \boldsymbol{\psi}_{h,L} \\ \boldsymbol{\psi}_{h,R} \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}_L \\ \boldsymbol{u}_R \end{bmatrix},
$$
$$
\boldsymbol{\psi} = \begin{bmatrix} \boldsymbol{\psi}_L \\ \boldsymbol{\psi}_R \end{bmatrix}, \qquad \boldsymbol{f} = \begin{bmatrix} \boldsymbol{f}_L \\ \boldsymbol{f}_R \end{bmatrix}, \qquad \boldsymbol{g} = \begin{bmatrix} \boldsymbol{g}_L \\ \boldsymbol{g}_R \end{bmatrix},
\tag{4.15}
$$

$$
\mathbb{E}(\boldsymbol{u}_h) = \begin{bmatrix} \left(\mathsf{R}_{\ell L}^T \mathsf{T}_{\ell L}^{(D)} - \mathsf{D}_{\ell L}^T \right) \left(\mathsf{R}_{\ell L} \boldsymbol{u}_{h,L} - u_D \right) \\ \mathsf{R}_{rR}^T \left(\mathsf{D}_{rR} \boldsymbol{u}_{h,R} - u_N \right) \end{bmatrix},
$$
$$
\mathbb{F}(\boldsymbol{\psi}_h) = \begin{bmatrix} \left(\mathsf{R}_{\ell L}^T \mathsf{T}_{\ell L}^{(D)} - \mathsf{D}_{\ell L}^T \right) \left(\mathsf{R}_{\ell L} \boldsymbol{\psi}_{h,L} - \psi_D \right) \\ \mathsf{R}_{rR}^T \left(\mathsf{D}_{rR} \boldsymbol{\psi}_{h,R} - \psi_N \right) \end{bmatrix},
\tag{4.16}
$$

and the matrices $\mathbb{A}, \mathbb{B}, \mathbb{H}, \mathbb{D}^{(2)}, \mathbb{M} \in \mathbb{R}^{2n_p \times 2n_p}$ with block entries

$$
\mathbb{A}_{11} = \begin{bmatrix} \mathsf{R}_{rL} \\ \mathsf{D}_{rL} \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{rL}^{(1)} & \mathsf{T}_{rL}^{(3)} \\ \mathsf{T}_{rL}^{(2)} & \mathsf{T}_{rL}^{(4)} \end{bmatrix} \begin{bmatrix} \mathsf{R}_{rL} \\ \mathsf{D}_{rL} \end{bmatrix}, \qquad \mathbb{B}_{11} = \begin{bmatrix} \mathsf{R}_{rL} \\ \mathsf{D}_{rL} \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{rL}^{(1)} & \mathsf{T}_{rL}^{(2)} + 1 \\ \mathsf{T}_{rL}^{(3)} - 1 & \mathsf{T}_{rL}^{(4)} \end{bmatrix} \begin{bmatrix} \mathsf{R}_{rL} \\ \mathsf{D}_{rL} \end{bmatrix},
$$

$$
\mathbb{A}_{12} = \begin{bmatrix} \mathsf{R}_{rL} \\ \mathsf{D}_{rL} \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{rL}^{(1)} & \mathsf{T}_{rL}^{(3)} \\ \mathsf{T}_{rL}^{(2)} & \mathsf{T}_{rL}^{(4)} \end{bmatrix} \begin{bmatrix} -\mathsf{R}_{\ell R} \\ \mathsf{D}_{\ell R} \end{bmatrix}, \qquad \mathbb{B}_{12} = \begin{bmatrix} \mathsf{R}_{rL} \\ \mathsf{D}_{rL} \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{rL}^{(1)} & \mathsf{T}_{rL}^{(2)} + 1 \\ \mathsf{T}_{rL}^{(3)} - 1 & \mathsf{T}_{rL}^{(4)} \end{bmatrix} \begin{bmatrix} -\mathsf{R}_{\ell R} \\ \mathsf{D}_{\ell R} \end{bmatrix},
$$

$$
\mathbb{A}_{21} = \begin{bmatrix} \mathsf{R}_{\ell R} \\ \mathsf{D}_{\ell R} \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{\ell R}^{(1)} & \mathsf{T}_{\ell R}^{(3)} \\ \mathsf{T}_{\ell R}^{(2)} & \mathsf{T}_{\ell R}^{(4)} \end{bmatrix} \begin{bmatrix} -\mathsf{R}_{rL} \\ \mathsf{D}_{rL} \end{bmatrix}, \qquad \mathbb{B}_{21} = \begin{bmatrix} \mathsf{R}_{\ell R} \\ \mathsf{D}_{\ell R} \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{\ell R}^{(1)} & \mathsf{T}_{\ell R}^{(2)} + 1 \\ \mathsf{T}_{\ell R}^{(3)} - 1 & \mathsf{T}_{\ell R}^{(4)} \end{bmatrix} \begin{bmatrix} -\mathsf{R}_{rL} \\ \mathsf{D}_{rL} \end{bmatrix},
$$

$$
\mathbb{A}_{22} = \begin{bmatrix} \mathsf{R}_{\ell R} \\ \mathsf{D}_{\ell R} \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{\ell R}^{(1)} & \mathsf{T}_{\ell R}^{(3)} \\ \mathsf{T}_{\ell R}^{(2)} & \mathsf{T}_{\ell R}^{(4)} \end{bmatrix} \begin{bmatrix} \mathsf{R}_{\ell R} \\ \mathsf{D}_{\ell R} \end{bmatrix}, \qquad \mathbb{B}_{22} = \begin{bmatrix} \mathsf{R}_{\ell R} \\ \mathsf{D}_{\ell R} \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{\ell R}^{(1)} & \mathsf{T}_{\ell R}^{(2)} + 1 \\ \mathsf{T}_{\ell R}^{(3)} - 1 & \mathsf{T}_{\ell R}^{(4)} \end{bmatrix} \begin{bmatrix} \mathsf{R}_{\ell R} \\ \mathsf{D}_{\ell R} \end{bmatrix},
$$

$$
\mathbb{H} = \begin{bmatrix} \mathsf{H}_L & \\ & \mathsf{H}_R \end{bmatrix}, \qquad\qquad \mathbb{D}^{(2)} = \begin{bmatrix} \mathsf{D}_L^{(2)}(\Lambda) & \\ & \mathsf{D}_R^{(2)}(\Lambda) \end{bmatrix},
\tag{4.17}
$$

$$
\mathbb{M} = \begin{bmatrix} \mathsf{M}_L - \mathsf{M}_L^T & \\ & \mathsf{M}_R - \mathsf{M}_R^T \end{bmatrix}.
$$

Note that for adjoint consistent schemes, it can be shown, using (4.11), that

$$\mathbb{A}_{12}^T = \mathbb{B}_{21}, \quad \text{and} \quad \mathbb{A}_{21}^T = \mathbb{B}_{12}. \tag{4.18}$$

The discrete residuals corresponding to the steady version of (3.1) and the adjoint problem (3.5) can now be written, respectively, as

$$R_{h,u}(\boldsymbol{u}_h) = -\mathbb{D}^{(2)}\boldsymbol{u}_h - \boldsymbol{f} + \mathbb{H}^{-1}\mathbb{A}\boldsymbol{u}_h + \mathbb{H}^{-1}\mathbb{E}(\boldsymbol{u}_h) = \mathbf{0}, \tag{4.19}$$

$$R_{h,\psi}(\boldsymbol{\psi}_h) = -\mathbb{D}^{(2)}\boldsymbol{\psi}_h - \boldsymbol{g} + \mathbb{H}^{-1}\mathbb{B}\boldsymbol{\psi}_h + \mathbb{H}^{-1}\mathbb{F}(\boldsymbol{\psi}_h) - \mathbb{H}^{-1}\mathbb{M}\boldsymbol{\psi}_h = \mathbf{0}. \tag{4.20}$$

Before stating the main result, we present an assumption regarding the primal and adjoint solution accuracy.

Assumption 1 *We assume that unique numerical solutions for the steady version of the discrete primal equation* (3.6) *and the discrete adjoint problem* (4.10) *exist, and these solutions are at least order* h^{p+1} *accurate in the maximum norm, i.e.,* $\|\boldsymbol{u} - \boldsymbol{u}_h\|_\infty = \mathcal{O}(h^{\geq p+1})$ *and* $\|\boldsymbol{\psi} - \boldsymbol{\psi}_h\|_\infty = \mathcal{O}(h^{\geq p+1})$.

Assumption 1 is not necessary if pointwise stability of the SBP-SAT discretization for the diffusion problem can be demonstrated, see [17,35,36,20,22,9]. Numerical experiments with adjoint consistent discretizations show primal and adjoint solution convergence rates of $p+1$ when a degree p diagonal-norm wide-stencil second-derivative SBP operator is used. In contrast, primal and adjoint solution convergence rates of $p+2$ are observed when a degree $p+1$ order-matched narrow-stencil second-derivative SBP operator is used with adjoint consistent SATs. The block-norm wide- and narrow-stencil second-derivative operators, on the other hand, exhibit primal solution convergence rates of $2p$.

We present the order of accuracy of the discrete functional approximating $\mathcal{I}(\mathcal{U}) = \mathcal{I}(\psi)$ in the following theorem.

Theorem 4.1 *Let the primal solution of the steady version of* (3.1) *and the adjoint solution of* (3.5) *be* $\mathcal{U}, \psi \in \mathcal{C}^{2p+2}(\Omega)$, *respectively, the variable coefficient in* (3.1) *and* (3.5) *be* $\lambda \in \mathcal{C}^{2p+1}(\Omega)$, *and the source terms in* (3.1) *and* (3.5) *be* $\mathcal{F}, \mathcal{G} \in \mathcal{C}^{2p}(\Omega)$, *respectively. If* $\boldsymbol{u}_h, \boldsymbol{\psi}_h \in \mathbb{R}^{n_e n_p}$ *are solutions to consistent discretizations of the steady version of* (3.1) *and* (3.5), *respectively, and Assumption 1 holds, then the discrete functionals* (4.3) *and* (4.4) *are order* h^{2p} *accurate approximations to the compatible linear functional* $\mathcal{I}(\mathcal{U}) = \mathcal{I}(\psi)$ *given by* (3.2) *and* (3.4), *i.e.,*

$$\mathcal{I}(\mathcal{U}) - I_h(\boldsymbol{u}_h) = \mathcal{O}\left(h^{2p}\right), \tag{4.21}$$

$$\mathcal{I}(\psi) - I_h(\boldsymbol{\psi}_h) = \mathcal{O}\left(h^{2p}\right). \tag{4.22}$$

Proof The proof can be found in Appendix B.

Remark 4.1 For implementations with the block-norm wide- or narrow-stencil second-derivative SBP operators of the type presented in [28], the estimate in (4.21) is attained even for adjoint inconsistent schemes. Note that for these types of operator, we have $\|\boldsymbol{u}_{h,k} - \boldsymbol{u}_k\|_\infty = \mathcal{O}(h^{2p})$ in (B.6). The block-norm narrow-stencil operators have $\mathsf{M}_k \neq \mathsf{M}_k^T$; hence, they lead to adjoint inconsistent schemes even when coupled with adjoint consistent SATs.

4.4 Stability Analysis

We use the energy method to analyze the stability of the SBP-SAT discretization of (3.1). The residual of the discretization for the homogeneous version of the problem, i.e., $\mathcal{F} = 0$, $\mathcal{U}_D = 0$, and $\mathcal{U}_N = 0$, summed over all elements can be written as

$$
R_h(\boldsymbol{u}_h, \boldsymbol{v}) = -\sum_{\Omega_k \in \mathcal{T}_h} \boldsymbol{v}_k^T \mathsf{M}_k \boldsymbol{u}_{h,k} - \sum_{\gamma \subset \Gamma^D} \begin{bmatrix} \mathsf{R}_{\gamma k} \boldsymbol{v}_k \\ \mathsf{D}_{\gamma k} \boldsymbol{v}_k \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{\gamma k}^{(D)} & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \mathsf{R}_{\gamma k} \boldsymbol{u}_{h,k} \\ \mathsf{D}_{\gamma k} \boldsymbol{u}_{h,k} \end{bmatrix}
$$
$$
-\sum_{\gamma \subset \Gamma^I} \begin{bmatrix} \mathsf{R}_{\gamma k} \boldsymbol{v}_k \\ \mathsf{R}_{\gamma v} \boldsymbol{v}_v \\ \mathsf{D}_{\gamma k} \boldsymbol{v}_k \\ \mathsf{D}_{\gamma v} \boldsymbol{v}_v \end{bmatrix}^T \begin{bmatrix} \mathsf{T}_{\gamma k}^{(1)} & -\mathsf{T}_{\gamma k}^{(1)} & \mathsf{T}_{\gamma k}^{(3)}-1 & \mathsf{T}_{\gamma k}^{(3)} \\ -\mathsf{T}_{\gamma v}^{(1)} & \mathsf{T}_{\gamma v}^{(1)} & \mathsf{T}_{\gamma v}^{(3)} & \mathsf{T}_{\gamma v}^{(3)}-1 \\ \mathsf{T}_{\gamma k}^{(2)} & -\mathsf{T}_{\gamma k}^{(2)} & \mathsf{T}_{\gamma k}^{(4)} & \mathsf{T}_{\gamma k}^{(4)} \\ -\mathsf{T}_{\gamma v}^{(2)} & \mathsf{T}_{\gamma v}^{(2)} & \mathsf{T}_{\gamma v}^{(4)} & \mathsf{T}_{\gamma v}^{(4)} \end{bmatrix} \begin{bmatrix} \mathsf{R}_{\gamma k} \boldsymbol{u}_{h,k} \\ \mathsf{R}_{\gamma v} \boldsymbol{u}_{h,v} \\ \mathsf{D}_{\gamma k} \boldsymbol{u}_{h,k} \\ \mathsf{D}_{\gamma v} \boldsymbol{u}_{h,v} \end{bmatrix} \tag{4.23}
$$

for $\boldsymbol{v} \in \mathbb{R}^{n_e n_p}$. In [38,37], a factorization of M_k for wide-stencil operators allowed the use of the borrowing trick and enabled $R_h(\boldsymbol{u}_h, \boldsymbol{v})$ to be written in terms of interface contributions only. However, the same factorization cannot be applied for narrow-stencil operators because $\mathsf{D}_{\gamma k}$ is constructed using a modified derivative operator at the element boundaries, $\mathsf{D}_{b,k}$, instead of D_k. To circumvent this, we make the following assumption.

Assumption 2 *The $\mathsf{D}_{b,k}$ matrix is invertible or can be modified such that it is invertible.*

The invertibility requirement on $\mathsf{D}_{b,k}$ is not too restrictive. In fact, all the narrow-stencil second-derivative operators in [30, 29, 27, 28] either have invertible $\mathsf{D}_{b,k}$ matrix or their $\mathsf{D}_{b,k}$ matrix can be modified such that it is invertible. For operators that include nodes at element boundaries, the only requirement for $\mathsf{D}_{b,k}$ to be invertible is that its interior diagonal entries are nonzero, e.g., $\mathsf{D}_{b,k}$ can be constructed from the identity matrix by modifying the first and last rows such that these rows approximate the first derivative to degree $\geq p + 1$. The invertibility of $\mathsf{D}_{b,k}$ matrix constructed in this manner can be verified using Gershgorin's theorem. Note that all the eigenvalues of $\mathsf{D}_{b,k}$ are nonzero as all other entries of either the row or column corresponding to each diagonal element is zero. For generalized narrow-stencil second-derivative operators with nodes at element boundaries, e.g., the hybrid Gauss-trapezoidal-Lobatto (HGTL) operators in [12], a similar modification can be applied to obtain an invertible $\mathsf{D}_{b,k}$ matrix. In contrast, all except the degree two hybrid Gauss-trapezoidal (HGT) operators in [12], which do not include boundary nodes, do not yield an invertible $\mathsf{D}_{b,k}$ matrix even after applying the modification discussed. However, it is likely possible to construct HGT operators such that $\mathsf{D}_{b,k}$ is invertible by enforcing a condition on the free variables during the construction of the operators. The structures of the invertible $\mathsf{D}_{b,k}$ matrices of the degree two CSBP and HGT operators are,

$$
\begin{bmatrix}
\times & \times & \times & \times & & & & & \\
& 1 & & & & & & & \\
& & 1 & & & & & & \\
& & & & & & & & \\
& & & & \ddots & & & & \\
& & & & & & & & \\
& & & & & & 1 & & \\
& & & & & & & 1 & \\
& & & & & & \times & \times & \times & \times
\end{bmatrix},
\qquad
\begin{bmatrix}
\times & \times & \times & \times & \times & & & & \\
\times & \times & \times & \times & \times & & & & \\
\times & \times & \times & \times & \times & & & & \\
\times & \times & \times & \times & \times & & & & \\
& & & & 1 & & & & \\
& & & & & \ddots & & & \\
& & & & & & 1 & & \\
& & & & & & \times & \times & \times & \times & \times \\
& & & & & & \times & \times & \times & \times & \times \\
& & & & & & \times & \times & \times & \times & \times \\
& & & & & & \times & \times & \times & \times & \times
\end{bmatrix},
$$

respectively, where each row containing $\times$ in its entries approximate the first derivative.

Another important assumption that is required in the subsequent energy stability analysis for adjoint consistent discretizations with narrow-stencil second-derivative operators is presented below.

Assumption 3 *The first and second-derivative SBP operators, D_k and $\mathsf{D}_k^{(2)}$, are nullspace consistent, i.e., the nonzero vectors in the nullspace of D_k and $\mathsf{D}_k^{(2)}$ are $\mathcal{N}(\mathsf{D}_k) = \mathrm{span}\{\mathbf{1}\} =: \boldsymbol{v}_c$ and $\mathcal{N}(\mathsf{D}_k^{(2)}) = \mathrm{span}\{\mathbf{1}, \boldsymbol{x}_k\}$, respectively.*

It should be noted that consistency of an SBP operator does not necessarily imply nullspace consistency and vice versa. The operators defined in Definitions 2.1 and 2.2 are consistent because they satisfy the accuracy conditions, i.e., they differentiate polynomials up to a required degree exactly [36]. In contrast, nullspace consistency requires that the nullspaces of D_k and $\mathsf{D}_k^{(2)}(\Lambda)$ exclusively contain vectors in $\mathrm{span}\{\mathbf{1}\}$ and $\mathrm{span}\{\mathbf{1}, \boldsymbol{x}_k\}$, respectively. SBP derivative operators are consistent by construction, and most of them are nullspace consistent as well [36].

Using Assumption 2 and enforcing the conditions necessary for conservation, (4.14), we can now write the sum of the residual and its transpose as

$$
\begin{aligned}
2R_h(\boldsymbol{u}_h, \boldsymbol{u}_h) = R_h(\boldsymbol{u}_h, \boldsymbol{u}_h) + R_h^T(\boldsymbol{u}_h, \boldsymbol{u}_h) = \\
- \sum_{\gamma \subset \Gamma^I}
\begin{bmatrix} \mathsf{R}_{\gamma k}\boldsymbol{u}_{h,k} \\ \mathsf{R}_{\gamma v}\boldsymbol{u}_{h,v} \\ \mathsf{D}_{b,k}\boldsymbol{u}_{h,k} \\ \mathsf{D}_{b,v}\boldsymbol{u}_{h,v} \end{bmatrix}^T
\begin{bmatrix}
2\mathsf{T}_{\gamma k}^{(1)} & -2\mathsf{T}_{\gamma k}^{(1)} & \sigma_k \mathsf{C}_{\gamma k} & -\sigma_v \mathsf{C}_{\gamma v} \\
-2\mathsf{T}_{\gamma k}^{(1)} & 2\mathsf{T}_{\gamma v}^{(1)} & -\sigma_k \mathsf{C}_{\gamma k} & \sigma_v \mathsf{C}_{\gamma v} \\
\sigma_k \mathsf{C}_{\gamma k}^T & -\sigma_k \mathsf{C}_{\gamma k}^T & \alpha_{\gamma k}\mathsf{V}_k & 0 \\
-\sigma_v \mathsf{C}_{\gamma v}^T & \sigma_v \mathsf{C}_{\gamma v}^T & 0 & \alpha_{\gamma v}\mathsf{V}_v
\end{bmatrix}
\begin{bmatrix} \mathsf{R}_{\gamma k}\boldsymbol{u}_{h,k} \\ \mathsf{R}_{\gamma v}\boldsymbol{u}_{h,v} \\ \mathsf{D}_{b,k}\boldsymbol{u}_{h,k} \\ \mathsf{D}_{b,v}\boldsymbol{u}_{h,v} \end{bmatrix} \\
- \sum_{\gamma \subset \Gamma^I}
\begin{bmatrix} \mathsf{D}_{\gamma k}\boldsymbol{u}_{h,k} \\ \mathsf{D}_{\gamma v}\boldsymbol{u}_{h,v} \end{bmatrix}^T
\begin{bmatrix} 2\mathsf{T}_{\gamma k}^{(4)} & 2\mathsf{T}_{\gamma k}^{(4)} \\ 2\mathsf{T}_{\gamma k}^{(4)} & 2\mathsf{T}_{\gamma}^{(4)} \end{bmatrix}
\begin{bmatrix} \mathsf{D}_{\gamma k}\boldsymbol{u}_{h,k} \\ \mathsf{D}_{\gamma v}\boldsymbol{u}_{h,v} \end{bmatrix} \\
- \sum_{\gamma \subset \Gamma^D}
\begin{bmatrix} \mathsf{R}_{\gamma k}\boldsymbol{u}_{h,k} \\ \mathsf{D}_{b,k}\boldsymbol{u}_{h,k} \end{bmatrix}^T
\begin{bmatrix} 2\mathsf{T}_{\gamma k}^{(D)} & -2\mathsf{C}_{\gamma k} \\ -2\mathsf{C}_{\gamma k}^T & \alpha_{\gamma k}\mathsf{V}_k \end{bmatrix}
\begin{bmatrix} \mathsf{R}_{\gamma k}\boldsymbol{u}_{h,k} \\ \mathsf{D}_{b,k}\boldsymbol{u}_{h,k} \end{bmatrix},
\end{aligned}
\tag{4.24}
$$

where $\mathsf{C}_{\gamma k} = n_{\gamma k}\mathsf{R}_{\gamma k}\Lambda_k$, $\mathsf{C}_{\gamma v} = n_{\gamma v}\mathsf{R}_{\gamma v}\Lambda_v$, $\sigma_k = \mathsf{T}_{\gamma k}^{(2)} + \mathsf{T}_{\gamma k}^{(3)} - 1$, $\sigma_v = \mathsf{T}_{\gamma v}^{(2)} + \mathsf{T}_{\gamma v}^{(3)} - 1$, $\alpha_{\gamma k}$ is a positive interface weight factor satisfying the relation $\sum_{\gamma \in \Gamma_k} \alpha_{\gamma k} = 1$, and

$$
\mathsf{V}_k = \mathsf{D}_{b,k}^{-T}(\mathsf{M}_k + \mathsf{M}_k^T)\mathsf{D}_{b,k}^{-1}, \qquad \mathsf{V}_v = \mathsf{D}_{b,v}^{-T}(\mathsf{M}_v + \mathsf{M}_v^T)\mathsf{D}_{b,v}^{-1}.
\tag{4.25}
$$

We note that V_k is positive semidefinite since $\boldsymbol{v}^T(\mathsf{M}_k + \mathsf{M}_k^T)\boldsymbol{v} \geq 0$ for all $\boldsymbol{v} \in \mathbb{R}^{n_p}$ implies

$$(\mathsf{D}_{b,k}^{-1}\boldsymbol{v})^T(\mathsf{M}_k + \mathsf{M}_k^T)(\mathsf{D}_{b,k}^{-1}\boldsymbol{v}) \geq 0. \tag{4.26}$$

Moreover, we have

$$\mathsf{D}_{b,k}\boldsymbol{v}_c = \boldsymbol{v}_0, \quad \text{or} \quad \mathsf{D}_{b,k}^{-1}\boldsymbol{v}_0 = \boldsymbol{v}_c, \tag{4.27}$$

where $\boldsymbol{v}_0$ represents vectors containing zero at the entries corresponding to the rows for which $\mathsf{D}_{b,k}$ contains consistent approximations of the first derivative and the values of $\boldsymbol{v}_c$ at all other entries.

For diagonal-norm narrow-stencil SBP operators that are constructed as in [30,29,27,12,28], we can determine the vectors in the nullspace of V_k using Assumptions 2 and 3.

Lemma 4.1 *Consider a consistent diagonal-norm narrow-stencil second-derivative SBP operator of the form* (2.4) *for which* $\mathsf{M}_k = \mathsf{M}_k^T$, *the* E_k *matrix is constructed such that it has nonzero rows only at row indices where the* $\mathsf{D}_{b,k}$ *matrix contains consistent approximations of the first derivative, the* $\mathsf{D}_k^{(2)}(\Lambda)$ *matrix has larger dense blocks at the top left and bottom right corners than the* E_k *matrix, and Assumptions 2 and 3 hold. Then, we have* $\mathcal{N}(\mathsf{M}_k) = \boldsymbol{v}_c$ *and* $\mathcal{N}(\mathsf{V}_k) = \mathcal{N}(\mathsf{V}_v) = \boldsymbol{v}_0$.

Proof The proof is given in Appendix C.

In [14], the stability conditions that the SATs must satisfy were derived for diagonal-norm narrow-stencil SBP operators assuming that M_k is SPD; however, most operators in the literature, e.g., [30, 29,27,12,28], do not satisfy this requirement. For dense-norm narrow-stencil second-derivative SBP operators, we make the following assumption regarding the nullspaces of M_k and M_k^T:

Assumption 4 *For dense-norm narrow-stencil SBP operators,* $\boldsymbol{v}_c$ *is the only nontrivial vector in the nullspaces of* M_k *and* M_k^T, *i.e.,* $\mathcal{N}(\mathsf{M}_k) = \mathcal{N}(\mathsf{M}_k^T) = \boldsymbol{v}_c$.

Under Assumption 4, (C.5) gives $\mathcal{N}(\mathsf{V}_k) = \mathcal{N}(\mathsf{V}_v) = \boldsymbol{v}_0$ for dense-norm narrow-stencil SBP operators. Before proceeding with the energy analysis of the SBP-SAT discretization, we state an essential theorem, which is proved in [1,15].

Theorem 4.2 *A symmetric matrix of the form* $\mathsf{Y} = \begin{bmatrix} \mathsf{Y}_{11} & \mathsf{Y}_{12} \\ \mathsf{Y}_{12}^T & \mathsf{Y}_{22} \end{bmatrix}$ *is positive semidefinite if and only if*

$$\mathsf{Y}_{22} \succeq 0, \quad (\mathsf{I} - \mathsf{Y}_{22}\mathsf{Y}_{22}^+)\mathsf{Y}_{12}^T = \mathbf{0}, \quad and \quad \mathsf{Y}_{11} - \mathsf{Y}_{12}\mathsf{Y}_{22}^+\mathsf{Y}_{12}^T \succeq 0, \tag{4.28}$$

where Y^+ *denotes the Moore-Penrose pseudoinverse of* Y *and* $\mathsf{Y} \succeq 0$ *indicates that* Y *is positive semidefinite.*

An SBP-SAT discretization is energy stable if

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|\boldsymbol{u}_h\|_{\mathsf{H}}^2\right) = \boldsymbol{u}_h^T\mathsf{H}\frac{\mathrm{d}\boldsymbol{u}_h}{\mathrm{d}t} + \frac{\mathrm{d}\boldsymbol{u}_h^T}{\mathrm{d}t}\mathsf{H}\boldsymbol{u}_h = 2R_h\left(\boldsymbol{u}_h, \boldsymbol{u}_h\right) \leq 0. \tag{4.29}$$

The sum of the residual and its transpose for conservative schemes, $2R_h(\boldsymbol{u}_h, \boldsymbol{u}_h)$, given in (4.24) satisfies the energy stability condition if

$$\mathsf{A} := \begin{bmatrix} \mathsf{A}_{11} & \mathsf{A}_{12} \\ \mathsf{A}_{21} & \mathsf{A}_{22} \end{bmatrix} = \begin{bmatrix} 2\mathsf{T}_{\gamma k}^{(1)} & -2\mathsf{T}_{\gamma k}^{(1)} & \sigma_k\mathsf{C}_{\gamma k} & -\sigma_v\mathsf{C}_{\gamma v} \\ -2\mathsf{T}_{\gamma k}^{(1)} & 2\mathsf{T}_{\gamma v}^{(1)} & -\sigma_k\mathsf{C}_{\gamma k} & \sigma_v\mathsf{C}_{\gamma v} \\ \sigma_k\mathsf{C}_{\gamma k}^T & -\sigma_k\mathsf{C}_{\gamma k}^T & \alpha_{\gamma k}\mathsf{V}_k & \mathbf{0} \\ -\sigma_v\mathsf{C}_{\gamma v}^T & \sigma_v\mathsf{C}_{\gamma v}^T & \mathbf{0} & \alpha_{\gamma v}\mathsf{V}_v \end{bmatrix},$$

$$\begin{bmatrix} 2\mathsf{T}_{\gamma k}^{(4)} & 2\mathsf{T}_{\gamma k}^{(4)} \\ 2\mathsf{T}_{\gamma k}^{(4)} & 2\mathsf{T}_{\gamma}^{(4)} \end{bmatrix}, \quad and \quad \begin{bmatrix} 2\mathsf{T}_{\gamma k}^{(D)} & -2\mathsf{C}_{\gamma k} \\ -2\mathsf{C}_{\gamma k}^T & \alpha_{\gamma k}\mathsf{V}_k \end{bmatrix} \tag{4.30}$$

are positive semidefinite. We partition the matrix $\mathsf{A} \in \mathbb{R}^{(2+2n_p)\times(2+2n_p)}$ using four blocks, namely $\mathsf{A}_{11} \in \mathbb{R}^{2\times2}$, $\mathsf{A}_{12} \in \mathbb{R}^{2\times2n_p}$, $\mathsf{A}_{21} \in \mathbb{R}^{2n_p\times2}$, and $\mathsf{A}_{22} \in \mathbb{R}^{2n_p\times2n_p}$.

For adjoint consistent SATs, we enforce all the conditions in (4.11). Furthermore, we require that $\mathsf{T}_{\gamma k}^{(3)} - \mathsf{T}_{\gamma k}^{(2)} = 1$, as in [38], which is a condition satisfied by the SATs corresponding to some of the popular discontinuous Galerkin fluxes for elliptic PDEs [37], e.g., the modified method of Bassi and Rebay (BR2) [3], local discontinuous Galerkin (LDG) [34], and compact discontinuous Galerkin (CDG) [32] methods. With this condition in place, the components of the A matrix for adjoint consistent schemes become

$$\mathsf{A}_{11} = \begin{bmatrix} 2\mathsf{T}_{\gamma k}^{(1)} & -2\mathsf{T}_{\gamma k}^{(1)} \\ -2\mathsf{T}_{\gamma k}^{(1)} & 2\mathsf{T}_{\gamma v}^{(1)} \end{bmatrix}, \qquad \mathsf{A}_{12} = \begin{bmatrix} 2\mathsf{T}_{\gamma k}^{(2)}\mathsf{C}_{\gamma k} & -2\mathsf{T}_{\gamma v}^{(2)}\mathsf{C}_{\gamma v} \\ -2\mathsf{T}_{\gamma k}^{(2)}\mathsf{C}_{\gamma k} & 2\mathsf{T}_{\gamma v}^{(2)}\mathsf{C}_{\gamma v} \end{bmatrix},$$

$$\mathsf{A}_{21} = \begin{bmatrix} 2\mathsf{T}_{\gamma k}^{(2)}\mathsf{C}_{\gamma k}^T & -2\mathsf{T}_{\gamma k}^{(2)}\mathsf{C}_{\gamma k}^T \\ -2\mathsf{T}_{\gamma v}^{(2)}\mathsf{C}_{\gamma v}^T & 2\mathsf{T}_{\gamma v}^{(2)}\mathsf{C}_{\gamma v}^T \end{bmatrix}, \quad \mathsf{A}_{22} = \begin{bmatrix} \alpha_{\gamma k}\mathsf{V}_k & \mathbf{0} \\ \mathbf{0} & \alpha_{\gamma v}\mathsf{V}_v \end{bmatrix}. \tag{4.31}$$

Theorem 4.3 *Let Assumptions 2 to 4 hold, then an adjoint consistent SBP-SAT discretization of the diffusion problem with coefficients satisfying $\mathsf{T}^{(3)}_{\gamma k} - \mathsf{T}^{(2)}_{\gamma k} = 1$, and uses a narrow-stencil second-derivative operator of the form (2.4) for the spatial discretization is energy stable if*

$$\mathsf{T}^{(1)}_{\gamma k} \geq \frac{2}{\alpha_{\gamma k}} \mathsf{T}^{(2)}_{\gamma k} \mathsf{R}_{\gamma k} \varLambda_k \mathsf{V}^+_k \varLambda_k \mathsf{R}^T_{\gamma k} \mathsf{T}^{(2)}_{\gamma k} + \frac{2}{\alpha_{\gamma v}} \mathsf{T}^{(2)}_{\gamma v} \mathsf{R}_{\gamma v} \varLambda_v \mathsf{V}^+_v \varLambda_v \mathsf{R}^T_{\gamma v} \mathsf{T}^{(2)}_{\gamma v}, \tag{4.32}$$

$$\mathsf{T}^{(D)}_{\gamma k} \geq \frac{2}{\alpha_{\gamma k}} \mathsf{R}_{\gamma k} \varLambda_k \mathsf{V}^+_k \varLambda_k \mathsf{R}^T_{\gamma k}, \tag{4.33}$$

$$\mathsf{T}^{(4)}_{\gamma k} \geq 0. \tag{4.34}$$

Proof The proof is given in Appendix D.

Remark 4.2 In practice, the pseudoinverse V^+_k for a given SBP operator is computed once and for all on the reference element. It is then scaled by the inverse of the metric Jacobian when the operator is mapped to the physical elements. In Table 1 of [13], equivalent values of $2\mathsf{R}_{\gamma k} \mathsf{V}^+_k \mathsf{R}^T_{\gamma k}$, denoted by q and scaled by the mesh spacing, are tabulated for the constant-coefficient diagonal-norm narrow-stencil SBP operators presented in [30]. The scaling used can be written as $h = 1/[(n_e n_p - 1) - (n_e - 1)]$, and $2\mathsf{V}^+_k = (\mathsf{D}^{-T}_{b,k} \mathsf{M}_k \mathsf{D}^{-1}_{b,k})^+$ for operators with $\mathsf{M}_k = \mathsf{M}^T_k$ is computed as $\mathsf{D}_{b,k}(\widetilde{\mathsf{M}}_k)^{-1}\mathsf{D}^T_{b,k}$, where $\widetilde{\mathsf{M}}_k$ is obtained by perturbing M_k such that the corner values of $\mathsf{D}_{b,k}(\widetilde{\mathsf{M}}_k)^{-1}\mathsf{D}^T_{b,k}$ are independent of the perturbation. The difference in the values of qh and $2h\mathsf{R}_{\gamma k}\mathsf{V}^+_k\mathsf{R}^T_{\gamma k}$ lies in the approaches pursued to evaluate $2\mathsf{V}^+_k$.

Remark 4.3 For stability of discretizations with wide-stencil second-derivative operators, the terms V^+_k and V^+_v in Theorem 4.3 are replaced by $(\mathsf{H}_k\varLambda_k + \varLambda_k\mathsf{H}_k)^{-1}$ and $(\mathsf{H}_v\varLambda_v + \varLambda_v\mathsf{H}_v)^{-1}$, respectively.

4.5 Interface SATs

In this section, we present a few concrete examples of SATs for diffusion problems. The type of SAT used in the discretization affects several numerical properties such as accuracy, stability, conditioning, symmetry, and sparsity. However, we do not analyze many of these properties; rather we limit our focus to aspects of solution and functional convergence. With this in mind, we introduce four SATs, of which two are stable and adjoint consistent while the other two are stable but adjoint inconsistent when implemented with narrow-stencil SBP operators. A more comprehensive analysis of SATs for diffusion problems is presented in [37], and additional types of SAT that are not studied in this work can be found therein.

4.5.1 BR2 SAT: The modified method of Bassi and Rebay

A stabilized version of the BR2 method [3] for implementation with the narrow-stencil second-derivative SBP operators can be obtained by choosing

$$\begin{aligned}
\mathsf{T}^{(1)}_{\gamma k} = \mathsf{T}^{(1)}_{\gamma v} &= \frac{1}{2\alpha_{\gamma k}} \mathsf{R}_{\gamma k} \varLambda_k \mathsf{V}^+_k \varLambda_k \mathsf{R}^T_{\gamma k} + \frac{1}{2\alpha_{\gamma v}} \mathsf{R}_{\gamma v} \varLambda_v \mathsf{V}^+_v \varLambda_v \mathsf{R}^T_{\gamma v}, \\
-\mathsf{T}^{(2)}_{\gamma k} = -\mathsf{T}^{(2)}_{\gamma v} = \mathsf{T}^{(3)}_{\gamma k} = \mathsf{T}^{(3)}_{\gamma k} &= \frac{1}{2}, \\
\mathsf{T}^{(D)}_{\gamma k} &= \frac{2}{\alpha_{\gamma k}} \mathsf{R}_{\gamma k} \varLambda_k \mathsf{V}^+_k \varLambda_k \mathsf{R}^T_{\gamma k}, \\
\mathsf{T}^{(4)}_{\gamma k} = \mathsf{T}^{(4)}_{\gamma v} &= 0.
\end{aligned} \tag{4.35}$$

The general form of the BR2 SAT was first proposed in [38] by discretizing the primal formulation of the DG method using SBP operators. It is straightforward to show that the BR2 SAT coefficients satisfy the adjoint consistency conditions in (4.11) and the stability requirements in Theorem 4.3.

4.5.2 LDG SAT: The local discontinuous Galerkin method

We determine the SAT coefficients corresponding to the LDG method [34] by discretizing the primal LDG formulation of the diffusion problem (see, e.g., [2,32] for the primal LDG formulation). Similar analysis with the wide-stencil second-derivative SBP operators can be found in [8,16,5,37]. Stable

LDG SAT coefficients (with no mesh dependent parameter for stabilization) for discretizations with the narrow-stencil second-derivative SBP operators are given by

$$\mathsf{T}_{\gamma k}^{(1)} = \mathsf{T}_{\gamma v}^{(1)} = \mathsf{T}_{\gamma k}^{(D)} = \frac{2}{\alpha_{\gamma k}} \mathsf{R}_{\gamma k} \Lambda_k \mathsf{V}_k^{+} \Lambda_k \mathsf{R}_{\gamma k}^{T}, \quad -\mathsf{T}_{\gamma k}^{(2)} = \mathsf{T}_{\gamma v}^{(3)} = 1,$$
$$\mathsf{T}_{\gamma k}^{(3)} = \mathsf{T}_{\gamma v}^{(2)} = \mathsf{T}_{\gamma k}^{(4)} = \mathsf{T}_{\gamma v}^{(4)} = 0. \tag{4.36}$$

Clearly, the LDG SAT coefficients in (4.36) satisfy both the adjoint consistency conditions in (4.11) and the stability demands in Theorem 4.3.

Remark 4.4 The LDG SAT coefficients presented in (4.36) are obtained by using a switch function value of $1/2$ and a global vector pointing to the positive x-axis. We refer the reader to [2,32,37] for details regarding the switch function and to [33] for a discussion on the need to use a global vector.

Remark 4.5 In one space dimension the LDG and CDG fluxes are identical [32]; therefore, the SAT coefficients presented in (4.36) define the CDG SAT as well.

4.5.3 BO SAT: The Baumann-Oden method

The SAT coefficients corresponding to the BO method [4] do not satisfy the adjoint consistency conditions in (4.11); hence, the energy stability requirements in Theorem 4.3 do not apply. The BO SAT coefficients for implementation with narrow-stencil second-derivative operators are given by

$$\mathsf{T}_{\gamma k}^{(2)} = \mathsf{T}_{\gamma v}^{(2)} = \mathsf{T}_{\gamma k}^{(3)} = \mathsf{T}_{\gamma v}^{(3)} = \frac{1}{2}, \quad \mathsf{T}_{\gamma k}^{(D)} = \frac{2}{\alpha_{\gamma k}} \mathsf{R}_{\gamma k} \Lambda_k \mathsf{V}_k^{+} \Lambda_k \mathsf{R}_{\gamma k}^{T},$$
$$\mathsf{T}_{\gamma k}^{(1)} = \mathsf{T}_{\gamma v}^{(1)} = \mathsf{T}_{\gamma k}^{(4)} = \mathsf{T}_{\gamma v}^{(4)} = 0. \tag{4.37}$$

Except for the conditions on $\mathsf{T}_{\gamma k}^{(D)}$, the stability conditions for the narrow- and wide-stencil SBP operators are the same when the BO SAT is used. Stability analysis for discretizations with wide-stencil SBP operators and the BO SAT can be found in [8,16,37]. Note that for the BO SAT coefficients in (4.37), we have $\mathsf{T}_{\gamma k}^{(1)} = \mathsf{T}_{\gamma v}^{(1)} = \sigma_k = \sigma_v = 0$ in the matrix A given in (4.30), and thus A is positive semidefinite. The stability analyses for the second and third matrices in (4.30) remain the same as those presented in the proof of Theorem 4.3. The BO SAT satisfies the conditions for conservation (4.14); hence, it leads to a conservative and stable but not adjoint consistent scheme.

4.5.4 CNG SAT: The Carpenter-Nordström-Gottlieb method

A version of the CNG SAT [7] that leads to a stable discretization when implemented with narrow-stencil second-derivative SBP operators has the SAT coefficients

$$\mathsf{T}_{\gamma k}^{(1)} = \mathsf{T}_{\gamma v}^{(1)} = \frac{1}{8\alpha_{\gamma k}} \mathsf{R}_{\gamma k} \Lambda_k \mathsf{V}_k^{+} \Lambda_k \mathsf{R}_{\gamma k}^{T} + \frac{1}{8\alpha_{\gamma v}} \mathsf{R}_{\gamma v} \Lambda_v \mathsf{V}_v^{+} \Lambda_v \mathsf{R}_{\gamma v}^{T},$$
$$\mathsf{T}_{\gamma k}^{(2)} = \mathsf{T}_{\gamma v}^{(2)} = \mathsf{T}_{\gamma k}^{(4)} = \mathsf{T}_{\gamma v}^{(4)} = 0,$$
$$\mathsf{T}_{\gamma k}^{(3)} = \mathsf{T}_{\gamma v}^{(3)} = \frac{1}{2}, \tag{4.38}$$
$$\mathsf{T}_{\gamma k}^{(D)} = \frac{2}{\alpha_{\gamma k}} \mathsf{R}_{\gamma k} \Lambda_k \mathsf{V}_k^{+} \Lambda_k \mathsf{R}_{\gamma k}^{T}.$$

Clearly, the coefficients in (4.38) satisfy all the conditions in (4.11) except the second one. Therefore, the CNG SAT leads to conservative but adjoint inconsistent schemes. The stability analyses of the second and third matrices in (4.30) are the same as those presented in the proof of Theorem 4.3. The positive semidefiniteness of the matrix A in (4.30) requires all the conditions in (4.28) to be satisfied. Substituting the CNG SAT coefficients in A, we see that $\mathsf{A}_{22} \succeq 0$, and it can be shown that $\left(\mathsf{I} - \mathsf{A}_{22}\mathsf{A}_{22}^{+}\right)\mathsf{A}_{21} = 0$. Hence, it only remains to find conditions such that $\mathsf{A}_{11} - \mathsf{A}_{12}\mathsf{A}_{22}^{+}\mathsf{A}_{21} \succeq 0$, which, after simplification, yields

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes \left[2\mathsf{T}_{\gamma k}^{(1)} - \left(\frac{1}{4\alpha_{\gamma k}} \mathsf{C}_{\gamma k}\mathsf{V}_k^{+}\mathsf{C}_{\gamma k}^{T} + \frac{1}{4\alpha_{\gamma k}} \mathsf{C}_{\gamma v}\mathsf{V}_v^{+}\mathsf{C}_{\gamma v}^{T} \right) \right] \succeq 0. \tag{4.39}$$

The inequality in (4.39) is satisfied by the $\mathsf{T}_{\gamma k}^{(1)} = \mathsf{T}_{\gamma v}^{(1)}$ coefficient given in (4.38).

5 Numerical Results

We consider the one-dimensional Poisson problem with Dirichlet and Neumann boundary conditions,

$$-\frac{\partial^2 \mathcal{U}}{\partial x^2} = \mathcal{F} \quad \text{in } \Omega = [0,1], \qquad \mathcal{U}\big|_{x=0} = \mathcal{U}_D, \qquad \frac{\partial \mathcal{U}}{\partial x}\bigg|_{x=1} = \mathcal{U}_N. \qquad (5.1)$$

We use the method of manufactured solution and let $\mathcal{U} = \cos(30x)$ as in [13]; thus, $\mathcal{F} = 30^2 \cos(30x)$. We also consider a compatible linear functional given by

$$\mathcal{I}(\mathcal{U}) = \int_0^1 \cos^2(30x)\, \mathrm{d}\Omega + \frac{1}{30^2}(1 - 30\sin(30) - \cos(30))\cos(30). \qquad (5.2)$$

Some of the operators mentioned in this paper are not designed to discretize variable-coefficient problems. The choice of constant-coefficient problem (with $\lambda = 1$) in (5.1) allows to compare such operators with those designed to handle variable coefficients. We note that the construction of variable-coefficient narrow-stencil SBP operators is more challenging than the construction of constant-coefficient narrow-stencil operators (e.g., see [27,12]). Their implementation after construction, however, is not significantly more complicated. A discussion on efficient implementation of variable-coefficient, narrow-stencil SBP operators can be found in [12].

We are interested in the convergence of the solution and functional errors under mesh refinement. Figure 5.1 presents the solution convergence for discretizations with the diagonal-norm narrow-stencil CSBP operators in [30] and the generalized SBP operators in [12] that have an invertible $\mathsf{D}_{b,k}$ matrix and satisfy the accuracy conditions, i.e., the $p = \{2,3\}$ HGTL and $p = 2$ HGT operators. Note that the degree four HGTL operator in [12] meets the accuracy requirements given in Definitions 2.1 and 2.2 to order h^5 only, while the degree three and four HGT operators have $\mathsf{D}_{b,k}$ matrices that cannot be modified as described in Section 4 to ensure their invertibility. The interface weight parameters in the SAT coefficients are set as $\alpha_{\gamma k} = \alpha_{\gamma v} = 1/2$ in all cases.

The solution error is computed as

$$\sqrt{\sum_{\Omega_k \in \mathcal{T}_h} (\boldsymbol{u}_{h,k} - \boldsymbol{u}_k)^T \mathsf{H}_k (\boldsymbol{u}_{h,k} - \boldsymbol{u}_k)}.$$

The convergence rates in Fig. 5.1 through Fig. 5.8 are calculated by fitting a line through the error values on the mesh resolutions indicated by the short, thin lines, and "dof" stands for the number of degrees of freedom in the spatial discretization. Figure 5.1 shows that a solution convergence rate of $p+2$ is attained when order-matched narrow-stencil operators are coupled with the adjoint consistent SATs, except with the degree one SBP operators. The adjoint inconsistent SATs, BO and CNG, exhibit solution convergence rates of $p+2$ with all the order-matched narrow-stencil SBP operators, except the degree one and three CSBP operators which yield convergence rates of $p+1$. This is consistent with the results[4] presented in [12] but somewhat surprising since the well-known even-odd convergence phenomenon (see e.g., [8,34,26]) that the BO method displays is not observed with the narrow-stencil SBP operators. Numerical experiments in [37] show that the BO and CNG SATs converge at rates of $p+1$ and p when implemented with odd and even degree multidimensional SBP operators, respectively. Indeed, this even-odd convergence phenomenon is also observed when the BO and CNG SATs are implemented with the Legendre-Gauss-Lobatto (LGL) and Legendre-Gauss (LG) wide-stencil SBP operators, as shown in Fig. 5.2. However, this trend does not hold consistently with the wide-stencil CSBP and HGT operators, as convergence rates of $p+1$ are achieved with the $p = 4$ operators, as depicted in Fig. 5.2d. Therefore, it appears that the even-odd convergence property of the BO and CNG SATs is dependent on the type of SBP operator used, and it is not observed with the diagonal-norm narrow-stencil SBP operators consistently. Implementations of the diagonal-norm wide-stencil SBP operators with the BR2 and LDG SATs lead to solution convergence rates of $p+1$, as depicted in Fig. 5.3. Finally, Fig. 5.4 shows that the block-norm wide- and narrow-stencil SBP operators presented in [28], denoted by CSBP2, achieve a solution convergence rate of $2p$ regardless of the type of SAT used.

The functional error is calculated as $|I_h(\boldsymbol{u}_h) - \mathcal{I}(\mathcal{U})|$. Figure 5.5 shows the functional convergence rates resulting from discretizations with the order-matched narrow-stencil SBP operators. As established in Theorem 4.1, the figure shows that the functional superconverges at a rate of $2p$ when adjoint consistent SATs are used. The adjoint inconsistent SATs yield larger functional error values and lower functional convergence rates when coupled with the degree three and four diagonal-norm

---

[4] Although not presented here, we observe a solution convergence rate of $p+2$ with all the diagonal-norm narrow-stencil CSBP operators presented in [12].

(a) $p = 1$

(b) $p = 2$

(c) $p = 3$

(d) $p = 4$
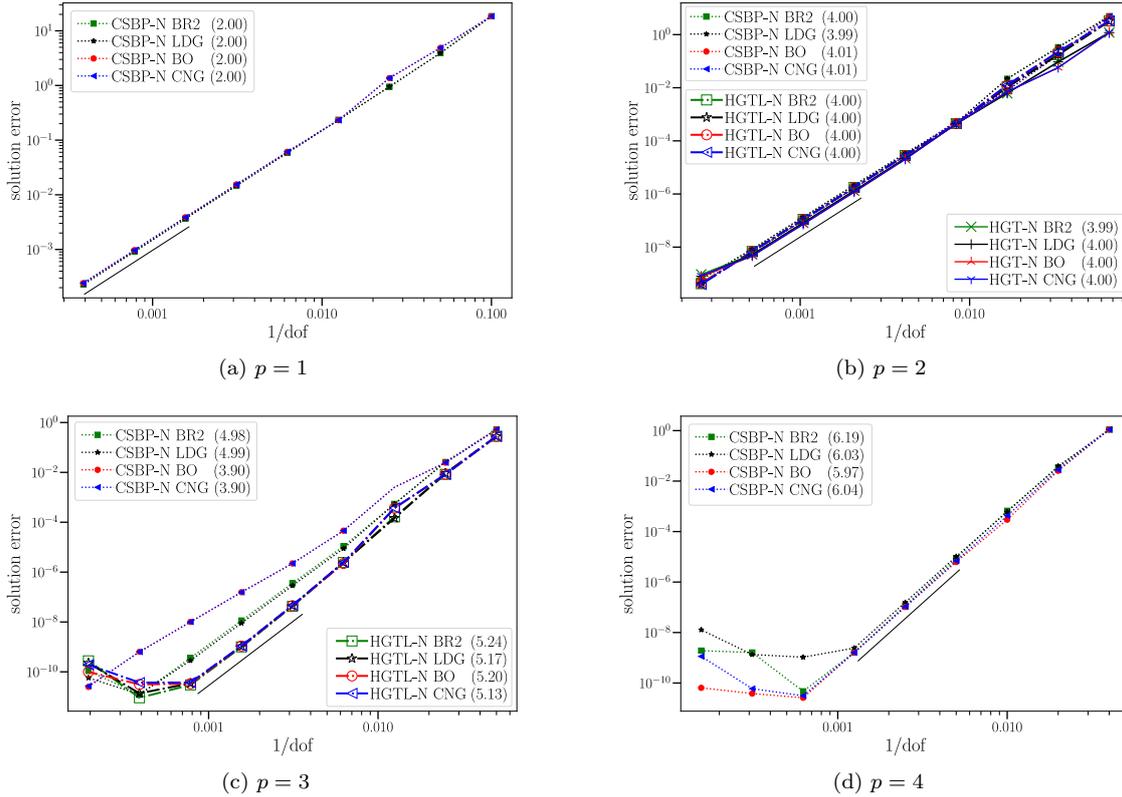
Fig. 5.1 Solution convergence under mesh refinement. The values in parentheses are the convergence rates, and "N" stands for narrow-stencil SBP operator. The short, thin lines indicate the mesh refinement levels used to compute the convergence rates.

narrow-stencil SBP operators, but they attain lower error values and a convergence rate of $2p$ for degree one and two operators. As can be seen from Fig. 5.6, when the adjoint inconsistent SATs are used with the diagonal-norm wide-stencil SBP operators, convergence rates of $2p$ are not attained, except for the $p = 1$ case, which, however, does not necessarily indicate superconvergence, as $p + 1$ and $2p$ are indistinguishable for this case. It is also evident from Figs. 5.5 and 5.6 that in most cases the diagonal-norm narrow-stencil operators result in lower functional error and larger functional convergence rates than the diagonal-norm wide-stencil operators when used with the adjoint inconsistent SATs. Figure 5.7 shows that the adjoint consistent SATs lead to functional convergence rates of $2p$ when used with the diagonal-norm wide-stencil SBP operators. Comparing the results depicted in Figs. 5.5 and 5.7, we can conclude that diagonal-norm narrow-stencil SBP operators do not offer better functional convergence rates than diagonal-norm wide-stencil SBP operators when coupled with the adjoint consistent SATs, which agrees with the theory. Similarly, the functional convergence rates attained with the block-norm wide- and narrow-stencil SBP operators are comparable, as depicted in Fig. 5.8. Furthermore, the functional convergence rate with the block-norm SBP operators is $2p$ for adjoint consistent as well as adjoint inconsistent schemes, except for the degree three block-norm wide-stencil SBP operator, which exhibits a $2p - 1$ convergence rate when implemented with the BO and CNG SATs. In most cases, the BO and CNG SATs yield lower functional error values than the BR2 and LDG SATs when implemented with the block-norm narrow-stencil SBP operators. For the $p = 1$ case, in particular, the BO and CNG SATs converge at a rate of ≈ 2.5. The superior functional convergence rates observed with the BO and CNG SATs are unexpected, especially given the fact that the solution errors for all the SATs implemented with the block-norm narrow-stencil SBP operators coincide, as depicted in Fig. 5.4. The lower functional error values of the BO and CNG SATs seem to be caused by error cancellation due to the oscillation around zero of the error in the integrand of the volume integral term of the functional, although it remains unclear to the authors as to why the BO and CNG SATs exhibit such a behavior when implemented with the block-norm narrow-stencil SBP operators.

(a) $p = 1$

(b) $p = 2$

(c) $p = 3$

(d) $p = 4$

Fig. 5.2 Solution convergence under mesh refinement with adjoint inconsistent SATs. The values in parentheses are the convergence rates, and "W" stands for wide-stencil SBP operator. The short, thin lines indicate the mesh refinement levels used to compute the convergence rates.

6 Conclusion

In this paper, we have shown that primal and adjoint consistent SBP-SAT discretizations of diffusion problems with diagonal-norm second-derivative generalized SBP operators lead to functional superconvergence if the primal and adjoint solutions are sufficiently smooth. For block-norm second-derivative operators, however, the analysis and the numerical experiments show that adjoint inconsistency does not degrade the functional convergence rate provided that the boundary closure for the second derivative operator and the matrix approximating the first derivative at the element boundaries are at least order $2p - 2$ and $2p - 1$ accurate, respectively. We have also derived the conditions required for the stability of adjoint consistent SBP-SAT discretizations with the narrow-stencil second-derivative generalized SBP operators under the assumptions that the operators are consistent and nullspace consistent. The stability analysis also requires that the derivative operator at the element boundaries, $\mathsf{D}_{b,k}$, be invertible. For most operators, $\mathsf{D}_{b,k}$ is invertible or can easily be modified to be invertible. For some operators, however, this is not the case, and it might be necessary to enforce the invertibility of this matrix during the construction of the SBP operators to ensure that SBP-SAT discretizations with these operators are stable and adjoint consistent in addition to the other attractive numerical properties that narrow-stencil operators offer.

Four different types of stable SATs for narrow-stencil SBP operators, among which two are adjoint consistent, are proposed and implemented in the numerical experiments. As predicted by the theory, the numerical experiments show that functionals superconverge at a rate of $2p$ when a diagonal-norm degree $p + 1$ narrow-stencil or degree p wide-stencil generalized SBP operator is used along with adjoint consistent SATs. It is also observed that the adjoint consistent BR2 and LDG SATs yield solution convergence rates of $p+1$ and $p+2$ when implemented with diagonal-norm wide- and narrow-stencil SBP operators, respectively. Implementations with the block-norm wide- and narrow-stencil SBP operators show a solution and functional convergence rates of $2p$ regardless of the type of SAT used. The even-odd convergence properties of the adjoint inconsistent BO and CNG SATs are not observed when these SATs are implemented with the diagonal- and block-norm narrow-stencil SBP operators.

(a) $p = 1$

(b) $p = 2$
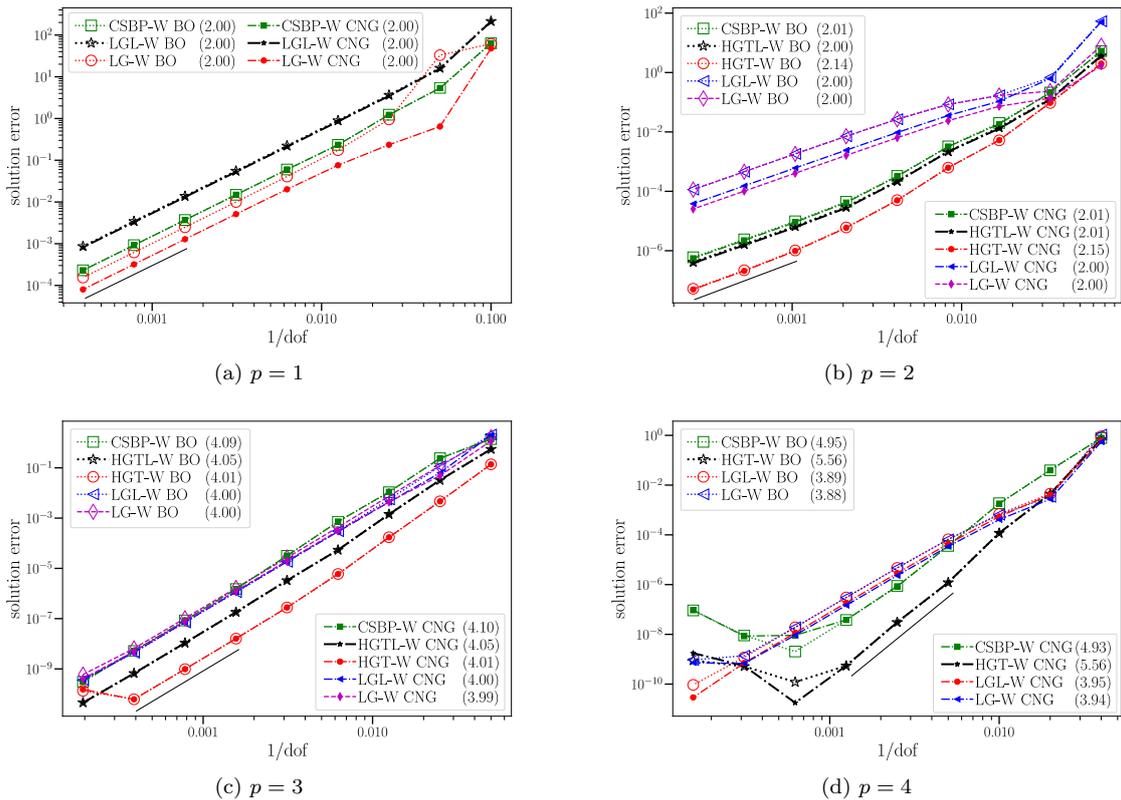

(c) $p = 3$

(d) $p = 4$
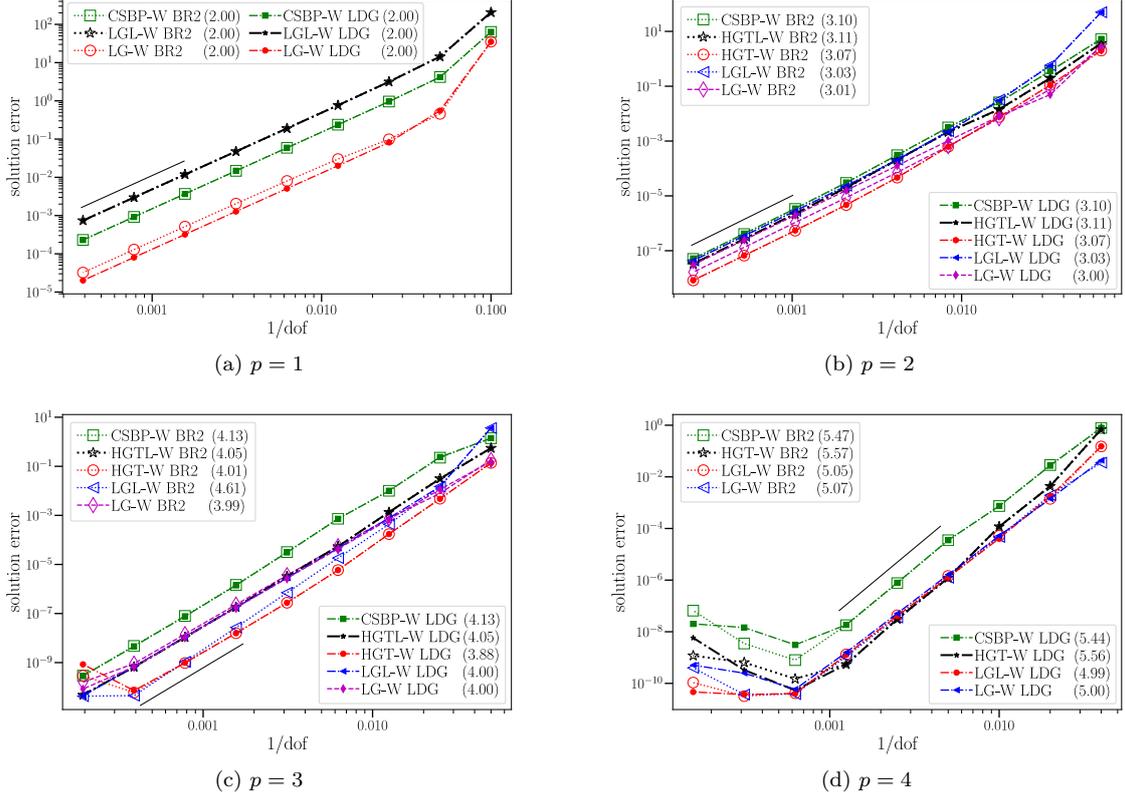
Fig. 5.3 Solution convergence under mesh refinement with adjoint consistent SATs. The values in parentheses are the convergence rates, and "W" stands for wide-stencil SBP operator. The short, thin lines indicate the mesh refinement levels used to compute the convergence rates.

While the SATs presented in this work ensure the consistency, conservation, adjoint consistency, energy stability, and functional superconvergence of SBP-SAT discretizations with narrow-stencil generalized SBP operators, optimization of the SAT coefficients to achieve improved numerical properties, e.g., better conditioning and spectral radius, may be pursued in the future. Extension of the theoretical results to the Navier-Stokes equations is also left for future work.

A Proof of Proposition 2.1

The form of the second derivative we wish to prove, (2.6), corresponds to application of integration by parts twice. Substituting (2.1) and (2.2) into (2.4), we have

$$\mathsf{D}_k^{(2)}(\Lambda) = \mathsf{H}_k^{-1}\left[-\mathsf{M}_k + \mathsf{E}_k\Lambda_k\mathsf{D}_{b,k}\right] = -\mathsf{H}_k^{-1}\mathsf{M}_k + \mathsf{H}_k^{-1}\left(\mathsf{R}_{rk}^T\mathsf{R}_{rk} - \mathsf{R}_{\ell k}^T\mathsf{R}_{\ell k}\right)\Lambda_k\mathsf{D}_{b,k}$$
$$= -\mathsf{H}_k^{-1}\mathsf{M}_k + \mathsf{H}_k^{-1}\mathsf{R}_{rk}^T\mathsf{D}_{rk} + \mathsf{H}_k^{-1}\mathsf{R}_{\ell k}^T\mathsf{D}_{\ell k}$$

Adding and subtracting $\mathsf{H}_k^{-1}\left(\mathsf{D}_k^{(2)}(\Lambda)\right)^T\mathsf{H}_k$, we get

$$\mathsf{D}_k^{(2)}(\Lambda) = -\mathsf{H}_k^{-1}\mathsf{M}_k + \mathsf{H}_k^{-1}\left(\mathsf{D}_k^{(2)}(\Lambda)\right)^T\mathsf{H}_k - \mathsf{H}_k^{-1}\left(\mathsf{D}_k^{(2)}(\Lambda)\right)^T\mathsf{H}_k$$
$$+ \mathsf{H}_k^{-1}\mathsf{R}_{rk}^T\mathsf{D}_{rk} + \mathsf{H}_k^{-1}\mathsf{R}_{\ell k}^T\mathsf{D}_{\ell k}$$
$$= \mathsf{H}_k^{-1}\left(\mathsf{D}_k^{(2)}(\Lambda)\right)^T\mathsf{H}_k - \mathsf{H}_k^{-1}\left[-\mathsf{M}_k^T + \mathsf{D}_{b,k}^T\Lambda_k^T\mathsf{E}_k^T\right]$$
$$+ \mathsf{H}_k^{-1}\mathsf{R}_{rk}^T\mathsf{D}_{rk} + \mathsf{H}_k^{-1}\mathsf{R}_{\ell k}^T\mathsf{D}_{\ell k} - \mathsf{H}_k^{-1}\mathsf{M}_k$$
$$= \mathsf{H}_k^{-1}\left(\mathsf{D}_k^{(2)}(\Lambda)\right)^T\mathsf{H}_k - \mathsf{H}_k^{-1}\mathsf{D}_{rk}^T\mathsf{R}_{rk} - \mathsf{H}_k^{-1}\mathsf{D}_{\ell k}^T\mathsf{R}_{\ell k}$$
$$+ \mathsf{H}_k^{-1}\mathsf{R}_{rk}^T\mathsf{D}_{rk} + \mathsf{H}_k^{-1}\mathsf{R}_{\ell k}^T\mathsf{D}_{\ell k} - \mathsf{H}_k^{-1}\left(\mathsf{M}_k - \mathsf{M}_k^T\right),$$

which is the desired result.

Fig. 5.4 Solution convergence under mesh refinement with block-norm wide-stencil ("W") and narrow-stencil ("N") CSBP operators. The values in parentheses are the convergence rates. The short, thin lines indicate the mesh refinement levels used to compute the convergence rates.

B Proof of Theorem 4.1

It is sufficient to show that the result holds for a domain tessellated by two elements, Ω_L and Ω_R, as the interface SATs considered couple immediate neighboring elements only. We let the Dirichlet and Neumann boundary conditions be implemented at the left and right boundaries of the domain. The boundary terms in both forms of the functional, (3.2) and (3.4), involve the products $\psi\lambda\partial\mathcal{U}/\partial x$ and $\mathcal{U}\lambda\partial\psi/\partial x$. Using the continuity of $\psi, \mathcal{U}$, and λ, we can approximate $(\psi\lambda\partial\mathcal{U}/\partial x) \in \mathcal{C}^{2p+1}(\Omega)$ and $(\mathcal{U}\lambda\partial\psi/\partial x) \in \mathcal{C}^{2p+1}(\Omega)$ at the boundary nodes by degree $\leq 2p$ polynomials. The integrands in the volume integrals of (3.2) and (3.4) are $2p$ times differentiable, i.e., $(\mathcal{GU}), (\psi\mathcal{F}) \in \mathcal{C}^{2p}(\Omega)$. Since integrals are approximated by quadratures of order h^{2p}, replacing $(\mathcal{GU}), (\psi\mathcal{F}) \in \mathcal{C}^{2p}(\Omega)$ by $(\widetilde{\mathcal{GU}}), (\widetilde{\psi\mathcal{F}}) \in \mathbb{P}^{2p-1}(\Omega)$ in the functionals introduces an error of order h^{2p}. Therefore, we consider $\widetilde{\mathcal{U}}, (\widetilde{\lambda\partial\mathcal{U}/\partial x}) \in \mathbb{P}^p(\Omega)$ to be at least order h^{p+1} approximations of $\mathcal{U}$ and $(\lambda\partial\mathcal{U}/\partial x)$, respectively, and thus $\widetilde{\mathcal{F}} \in \mathbb{P}^{p-1}(\Omega)$ due to the steady version of the primal PDE, (3.1). Similarly, considering $\widetilde{\psi}, (\widetilde{\lambda\partial\psi/\partial x}) \in \mathbb{P}^p(\Omega)$ to be at least order h^{p+1} approximations of ψ and $(\lambda\partial\psi/\partial x)$, respectively, gives $\widetilde{\mathcal{G}} \in \mathbb{P}^{p-1}(\Omega)$ due to the adjoint PDE, (3.5). For primal and adjoint consistent discretizations, the numerical primal and adjoint solutions are order $h^{\geq p+1}$ accurate despite the polynomial approximations; hence, it is sufficient to show that either (4.21) or (4.22) hold for the polynomial integrands instead of the general continuous functions. Note that compatible functionals satisfy $\mathcal{I}(\mathcal{U}) = \mathcal{I}(\psi)$, and we enforced the condition $I_h(\boldsymbol{u}_h) = I_h(\boldsymbol{\psi}_h)$ to find the discrete adjoint problem; hence, $\mathcal{I}(\mathcal{U}) - I_h(\boldsymbol{u}_h) = \mathcal{I}(\psi) - I_h(\boldsymbol{\psi}_h)$. For the rest of the proof, we drop the tilde sign used to distinguish polynomials from the general continuous functions.

If $\mathcal{U} \in \mathbb{P}^p(\Omega)$ and $(\lambda\partial\mathcal{U}/\partial x) \in \mathbb{P}^p(\Omega)$, then we discretize (3.2) to find

$$\mathcal{I}(\mathcal{U}) = \boldsymbol{u}_L^T\mathsf{H}_L\boldsymbol{g}_L + \boldsymbol{g}_R^T\mathsf{H}_R\boldsymbol{u}_R - \psi_D w_{\ell L} + \psi_N u_{rR} + \mathcal{O}\left(h^{2p}\right), \tag{B.1}$$

where $w_{\ell L} = [\lambda\frac{\partial\mathcal{U}}{\partial x}n_\ell]_{\Gamma^D}$ and $u_{rR} = \mathcal{U}|_{\Gamma^N}$. Subtracting (4.3) from (B.1) and rearranging, we have

$$\begin{aligned}
\mathcal{I}\left(\mathcal{U}\right) = {} & I_h\boldsymbol{u}_h - \boldsymbol{g}_L^T\mathsf{H}_L\left(\boldsymbol{u}_{h,L} - \boldsymbol{u}_L\right) + \psi_D\left(\mathsf{D}_{\ell L}\boldsymbol{u}_{h,L} - w_{\ell L}\right) \\
& - \psi_D\mathsf{T}_{\ell L}^{(D)}\left(\mathsf{R}_{\ell L}\boldsymbol{u}_{h,L} - u_{\ell L}\right) - \boldsymbol{g}_R^T\mathsf{H}_R\left(\boldsymbol{u}_{h,R} - \boldsymbol{u}_R\right) \\
& - \psi_N\left(\mathsf{R}_{rR}\boldsymbol{u}_{h,R} - u_{rR}\right) + \mathcal{O}\left(h^{2p}\right).
\end{aligned} \tag{B.2}$$

Since $\mathcal{U} \in \mathbb{P}^p(\Omega)$, the $\mathsf{R}_{\gamma k}$ and $\mathsf{D}_{\gamma k}$ matrices are exact when applied to the restriction of $\mathcal{U}$ to the grid points, e.g., $\mathsf{R}_{rR}\boldsymbol{u}_R = u_{rR}$ and $\mathsf{D}_{\ell L}\boldsymbol{u}_L = w_{\ell L}$. Applying this property in (B.2) and simplifying we obtain

$$\mathcal{I}\left(\mathcal{U}\right) = I_h\left(\boldsymbol{u}_h\right) - \boldsymbol{g}^T\mathbb{H}\left(\boldsymbol{u}_h - \boldsymbol{u}\right) - \begin{bmatrix}\psi_D\mathsf{T}_{\ell L}^{(D)}\mathsf{R}_{\ell L} - \psi_D\mathsf{D}_{\ell L} \\ \psi_N\mathsf{R}_{rR}\end{bmatrix}\left(\boldsymbol{u}_h - \boldsymbol{u}\right) + \mathcal{O}\left(h^{2p}\right). \tag{B.3}$$

(a) $p = 1$

(b) $p = 2$

(c) $p = 3$

(d) $p = 4$

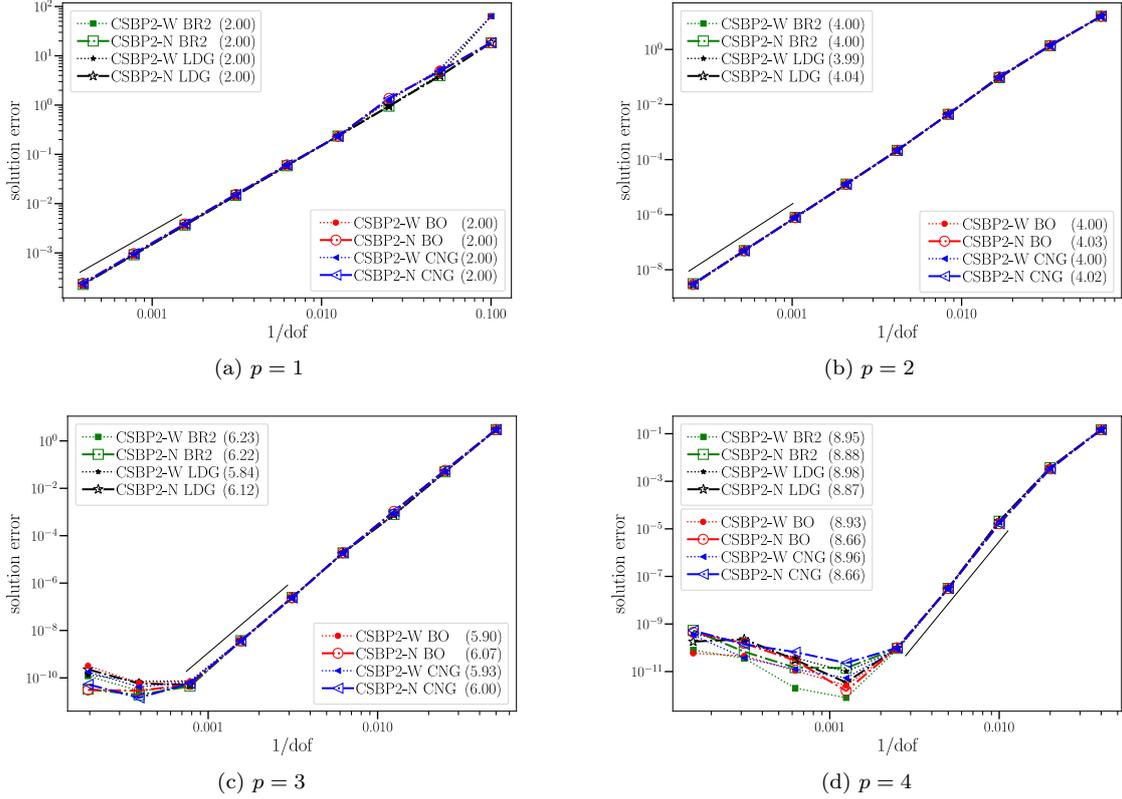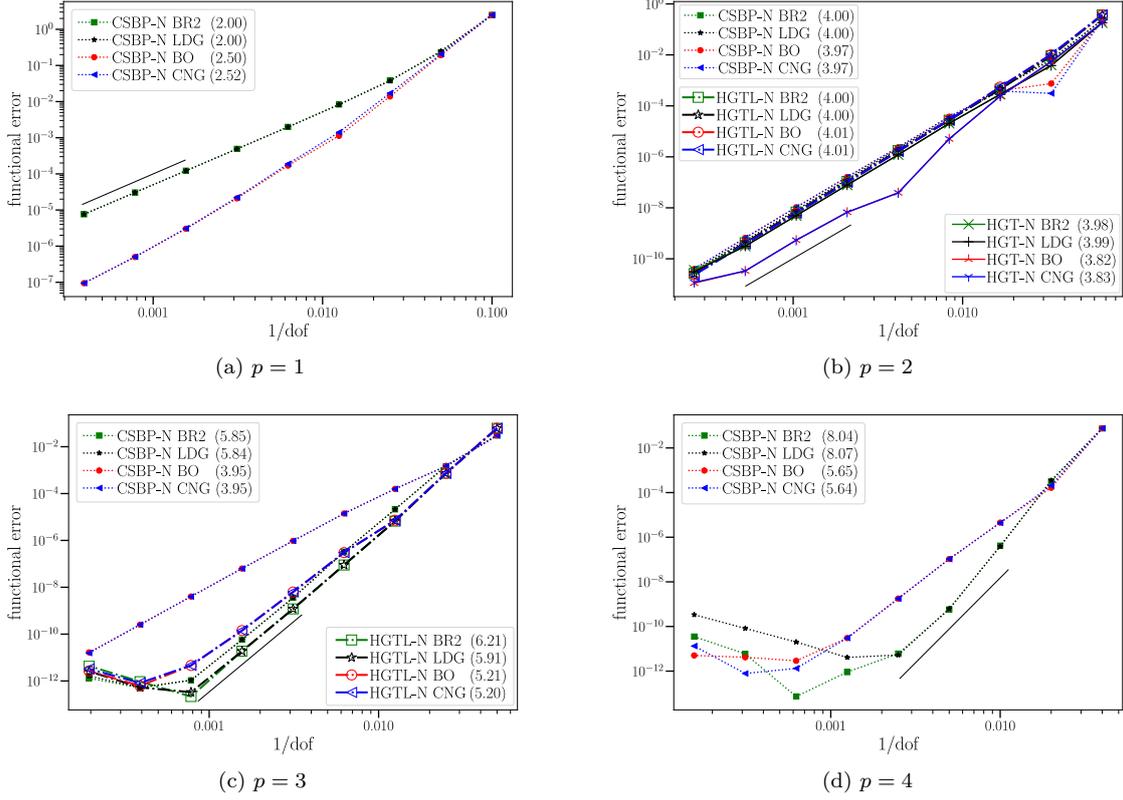**Fig. 5.5** Functional convergence under mesh refinement. The values in parentheses are the convergence rates, and "N" stands for narrow-stencil SBP operator. The short, thin lines indicate the mesh refinement levels used to compute the convergence rates.

Adding $\boldsymbol{\psi}^T \mathbb{H} R_{h,u}\left(\boldsymbol{u}_h\right) = 0$ to the RHS of (B.3) and rearranging terms, we have

$$
\begin{aligned}
\mathcal{I}\left(\mathcal{U}\right) = I_h\left(\boldsymbol{u}_h\right) - \boldsymbol{\psi}^T \mathbb{H} \mathbb{D}^{(2)} \boldsymbol{u} - \boldsymbol{\psi}^T \mathbb{H} \boldsymbol{f} + \bigg\{ &- \boldsymbol{g}^T - \boldsymbol{\psi}^T \mathbb{H} \mathbb{D}^{(2)} \mathbb{H}^{-1} \\
&+ \boldsymbol{\psi}^T \mathbb{A} \mathbb{H}^{-1} - \begin{bmatrix} \psi_D \mathsf{T}_{\ell L}^{(D)} \mathsf{R}_{\ell L} - \psi_D \mathsf{D}_{\ell L} \\ \psi_N \mathsf{R}_{rR} \end{bmatrix} \mathbb{H}^{-1} \\
&+ \boldsymbol{\psi}^T \begin{bmatrix} \mathsf{R}_{\ell L}^T \mathsf{T}_{\ell L}^{(D)} \mathsf{R}_{\ell L} - \mathsf{D}_{\ell L}^T \mathsf{R}_{\ell L} \\ \mathsf{R}_{rR}^T \mathsf{D}_{rR} \end{bmatrix} \mathbb{H}^{-1} \bigg\} \mathbb{H}\left(\boldsymbol{u}_h - \boldsymbol{u}\right) + \mathcal{O}\left(h^{2p}\right).
\end{aligned}
\tag{B.4}
$$

Using the identity in (2.6) we can write

$$
\begin{aligned}
-\mathbb{H} \mathbb{D}^{(2)} \mathbb{H}^{-1} = &- \left(\mathbb{D}^{(2)}\right)^T + \begin{bmatrix} \mathsf{D}_{\ell L}^T \mathsf{R}_{\ell L} - \mathsf{R}_{\ell L}^T \mathsf{D}_{\ell L} & \mathbf{0} \\ \mathbf{0} & \mathsf{D}_{rR}^T \mathsf{R}_{rR} - \mathsf{R}_{rR}^T \mathsf{D}_{rR} \end{bmatrix} \mathbb{H}^{-1} \\
&+ \begin{bmatrix} \mathsf{D}_{rL}^T \mathsf{R}_{rL} - \mathsf{R}_{rL}^T \mathsf{D}_{rL} & \mathbf{0} \\ \mathbf{0} & \mathsf{D}_{\ell R}^T \mathsf{R}_{\ell R} - \mathsf{R}_{\ell R}^T \mathsf{D}_{\ell R} \end{bmatrix} \mathbb{H}^{-1} + \mathbb{M} \mathbb{H}^{-1},
\end{aligned}
\tag{B.5}
$$

which, after substituting into (B.4) and simplifying, gives

$$
\begin{aligned}
\mathcal{I}\left(\mathcal{U}\right) = I_h\left(\boldsymbol{u}_h\right) - \boldsymbol{\psi}^T \mathbb{H} \left[\mathbb{D}^{(2)} \boldsymbol{u} + \boldsymbol{f}\right] + \bigg\{ &- \boldsymbol{g}^T - \boldsymbol{\psi}^T \left(\mathbb{D}^{(2)}\right)^T \\
&+ \boldsymbol{\psi}^T \mathbb{B}^T \mathbb{H}^{-1} + \left[\mathbb{F}\left(\boldsymbol{\psi}\right)\right]^T \mathbb{H}^{-1} - \boldsymbol{\psi}^T \mathbb{M}^T \mathbb{H}^{-1} \bigg\} \mathbb{H}\left(\boldsymbol{u}_h - \boldsymbol{u}\right) + \mathcal{O}\left(h^{2p}\right).
\end{aligned}
\tag{B.6}
$$

Since $\mathcal{U} \in \mathbb{P}^p(\Omega)$, the second term on the RHS vanishes due to the primal PDE. The third term is $\mathcal{O}(h^{\geq 2p+1})$ due to the consistency of the adjoint discretization, the fact that $\mathbb{H}$ is $\mathcal{O}(h)$, and Assumption 1. Therefore, $\mathcal{I}\left(\mathcal{U}\right) = I_h\left(\boldsymbol{u}_h\right) + \mathcal{O}\left(h^{2p}\right)$.

Alternatively, if we consider $\psi \in \mathbb{P}^p(\Omega)$ and $(\lambda \partial \psi / \partial x) \in \mathbb{P}^p(\Omega)$, we start by discretizing the second form of the functional, (3.4),

$$
\mathcal{I}(\psi) = \boldsymbol{\psi}_L^T \mathsf{H}_L \boldsymbol{f}_L + \boldsymbol{\psi}_R^T \mathsf{H}_R \boldsymbol{f}_R - u_D z_{\ell L} + u_N \psi_{rR} + \mathcal{O}\left(h^{2p}\right),
\tag{B.7}
$$

where $z_{\ell L} = \left[\lambda \frac{\partial \psi}{\partial x} n_\ell\right]_{\Gamma^D}$ and $\psi_{rR} = \psi|_{\Gamma^N}$. Subtracting (4.4) from (B.7) and rearranging, we obtain

$$
\begin{aligned}
\mathcal{I}(\psi) = I_h(\boldsymbol{\psi}_h) &- \boldsymbol{f}_L^T \mathsf{H}_L \left(\boldsymbol{\psi}_{h,L} - \boldsymbol{\psi}_L\right) + u_D \left(\mathsf{D}_{\ell L} \boldsymbol{\psi}_{h,L} - z_{\ell L}\right) \\
&- u_D \mathsf{T}_{\ell L}^{(D)} \left(\mathsf{R}_{\ell L} \boldsymbol{\psi}_{h,L} - \psi_{\ell L}\right) - \boldsymbol{f}_R^T \mathsf{H}_R \left(\boldsymbol{\psi}_{h,R} - \boldsymbol{\psi}_R\right) \\
&- u_N \left(\mathsf{R}_{rR} \boldsymbol{\psi}_{h,R} - \psi_{rR}\right) + \mathcal{O}\left(h^{2p}\right).
\end{aligned}
\tag{B.8}
$$

Fig. 5.6 Functional convergence under mesh refinement with adjoint inconsistent SATs. The values in parentheses are the convergence rates, and "W" stands for wide-stencil SBP operator. The short, thin lines indicate the mesh refinement levels used to compute the convergence rates. The convergence rates for the $p = 3$ HGTL and $p = 4$ HGT operators are calculated using the error values at the mesh levels indicated by the short, thin lines with a star marker.

Using the accuracies of $R_{\gamma k}$ and $D_{\gamma k}$ to approximate the boundary terms, adding $\boldsymbol{u}^T \mathbb{H} R_{h,\psi}(\boldsymbol{\psi}_h) = 0$, and simplifying leads to

$$
\begin{aligned}
\mathcal{I}(\psi) = I_h(\boldsymbol{\psi}_h) &- \left[\boldsymbol{u}^T \mathbb{H} \mathbb{D}^{(2)} \boldsymbol{\psi} + \boldsymbol{u}^T \mathbb{H} \boldsymbol{g} \right] + \left\{ -\boldsymbol{f}^T - \boldsymbol{u}^T \mathbb{H} \mathbb{D}^{(2)} \mathbb{H}^{-1} \right. \\
&+ \boldsymbol{u}^T \mathbb{B} \mathbb{H}^{-1} - \left[\begin{matrix} u_D \mathsf{T}_{\ell L}^{(D)} \mathsf{R}_{\ell L} - u_D \mathsf{D}_{\ell L} \\ u_N \mathsf{R}_{rR} \end{matrix} \right] \mathbb{H}^{-1} \\
&+ \boldsymbol{u}^T \left. \left[\begin{matrix} \mathsf{R}_{\ell L}^T \mathsf{T}_{\ell L}^{(D)} \mathsf{R}_{\ell L} - \mathsf{D}_{\ell L}^T \mathsf{R}_{\ell L} \\ \mathsf{R}_{rR}^T \mathsf{D}_{rR} \end{matrix} \right] \mathbb{H}^{-1} \right\} \mathbb{H}(\boldsymbol{\psi}_h - \boldsymbol{\psi}) - \boldsymbol{u}^T \mathbb{M} \boldsymbol{\psi}_h + \mathcal{O}\left(h^{2p}\right).
\end{aligned}
\tag{B.9}
$$

Using the identity (B.5) and simplifying, we find

$$
\begin{aligned}
\mathcal{I}(\psi) = I_h(\boldsymbol{\psi}_h) - \boldsymbol{u}^T \mathbb{H} \left[\mathbb{D}^{(2)} \boldsymbol{\psi} + \boldsymbol{g} \right] &+ \left\{ -\boldsymbol{f}^T - \boldsymbol{u}^T \left(\mathbb{D}^{(2)} \right)^T + \boldsymbol{u}^T \mathbb{A}^T \mathbb{H}^{-1} \right. \\
&+ \left. \left[\mathbb{E}(\boldsymbol{u}) \right]^T \mathbb{H}^{-1} - \boldsymbol{u}^T \mathbb{M}^T \mathbb{H}^{-1} \right\} \mathbb{H}(\boldsymbol{\psi}_h - \boldsymbol{\psi}) - \boldsymbol{u}^T \mathbb{M} \boldsymbol{\psi}_h + \mathcal{O}\left(h^{2p}\right).
\end{aligned}
\tag{B.10}
$$

Noting that $\mathbb{M} + \mathbb{M}^T = \boldsymbol{0}$, we have

$$
\begin{aligned}
\mathcal{I}(\psi) = I_h(\boldsymbol{\psi}_h) - \boldsymbol{u}^T \mathbb{H} \left[\mathbb{D}^{(2)} \boldsymbol{\psi} + \boldsymbol{g} \right] &+ \left\{ -\boldsymbol{f}^T - \boldsymbol{u}^T \left(\mathbb{D}^{(2)} \right)^T + \boldsymbol{u}^T \mathbb{A}^T \mathbb{H}^{-1} \right. \\
&+ \left. \left[\mathbb{E}(\boldsymbol{u}) \right]^T \mathbb{H}^{-1} \right\} \mathbb{H}(\boldsymbol{\psi}_h - \boldsymbol{\psi}) - \boldsymbol{u}^T \mathbb{M} \boldsymbol{\psi} + \mathcal{O}\left(h^{2p}\right).
\end{aligned}
\tag{B.11}
$$

The second and fourth terms on the RHS of (B.11) vanish due to the adjoint PDE and the adjoint consistency requirement that $\mathbb{M} = \boldsymbol{0}$, respectively. The third term is $\mathcal{O}(h^{\geq 2p+1})$ due to the consistency of the primal discretization, the scaling of the norm matrix, and Assumption 1. Therefore, the estimates in (4.21) and (4.22) hold.
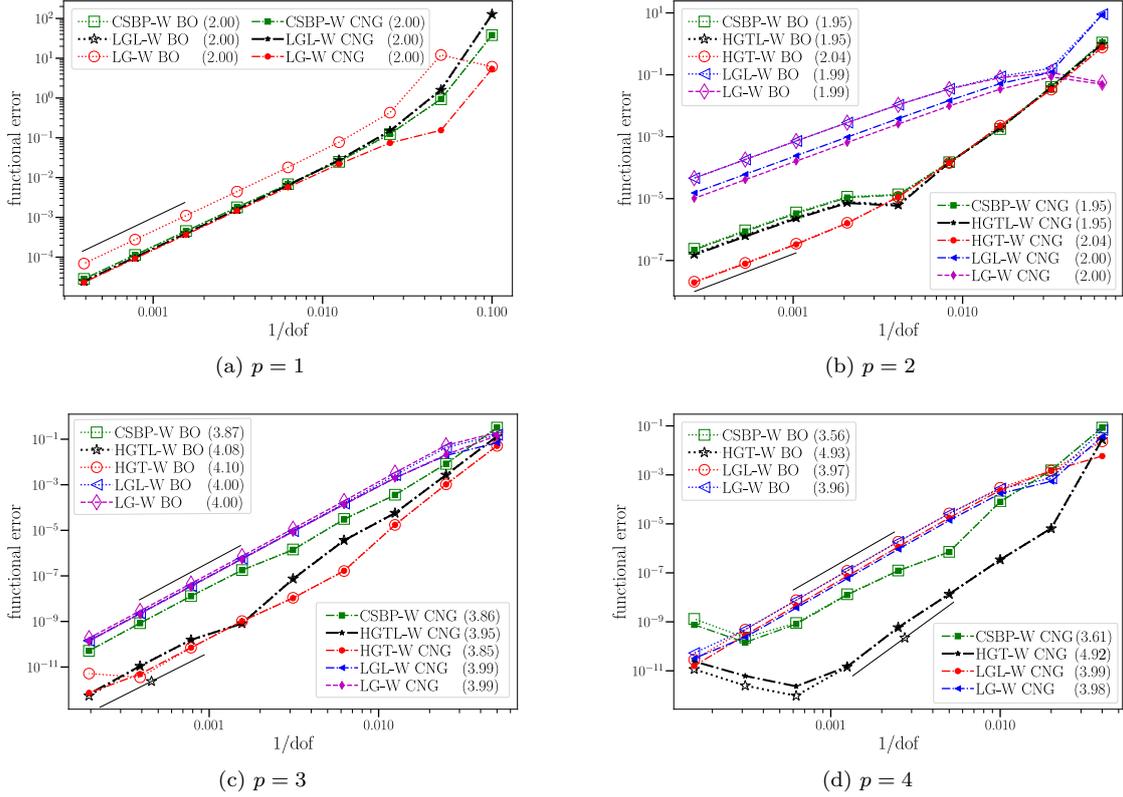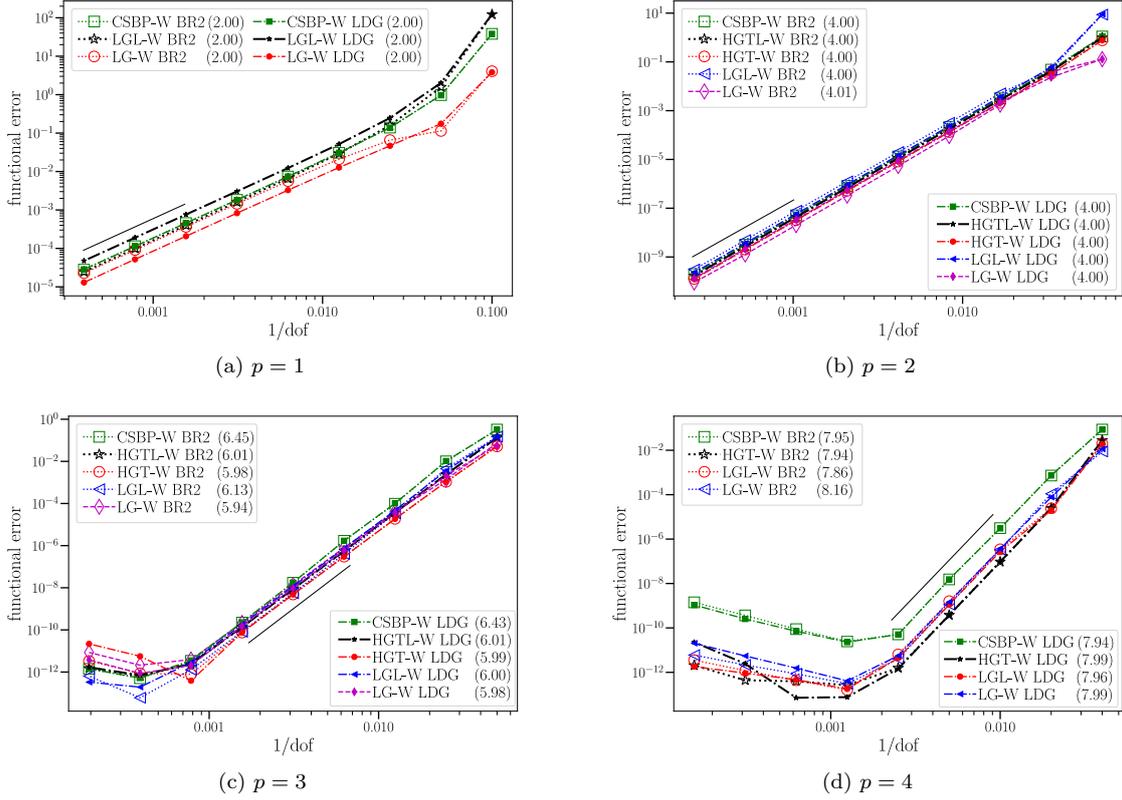
Fig. 5.7 Functional convergence under mesh refinement with adjoint consistent SATs. The values in parentheses are the convergence rates, and "W" stands for wide-stencil SBP operator. The short, thin lines indicate the mesh refinement levels used to compute the convergence rates.

C Proof of Lemma 4.1

The nullspace consistency of the $D_k^{(2)}(\Lambda)$ in Assumption 3 implies the following:

$$H_k D_k^{(2)}(\Lambda)\boldsymbol{v}_c = \left(-M_k + E_k \Lambda_k D_{b,k}\right)\boldsymbol{v}_c = -M_k\boldsymbol{v}_c + E_k\Lambda_k D_{b,k}\boldsymbol{v}_c = \boldsymbol{0}, \tag{C.1}$$

$$H_k D_k^{(2)}(\Lambda)\boldsymbol{x}_k = \left(-M_k + E_k \Lambda_k D_{b,k}\right)\boldsymbol{x}_k = -M_k\boldsymbol{x}_k + E_k\Lambda_k D_{b,k}\boldsymbol{x}_k = \boldsymbol{0}. \tag{C.2}$$

The second term in the last equality in (C.1) is zero due to the structure of the E_k matrix, i.e., $E_k\Lambda_k D_{b,k}\boldsymbol{v}_c = \boldsymbol{0}$; thus, $M_k\boldsymbol{v}_c = \boldsymbol{0}$. Furthermore, $M_k\boldsymbol{x}_k \neq \boldsymbol{0}$ in (C.2) since otherwise we would obtain

$$H_k D_k^{(2)}(\Lambda)\boldsymbol{x}_k = E_k\Lambda_k D_{b,k}\boldsymbol{x}_k = R_{rk}^T D_{rk}\boldsymbol{x}_k + R_{\ell k}^T D_{\ell k}\boldsymbol{x}_k = R_{rk}^T \lambda_r - R_{\ell k}^T \lambda_\ell = \boldsymbol{0}, \tag{C.3}$$

which is not possible as R_{rk} and $R_{\ell k}$ do not have nonzero values at the same entries, and $\lambda > 0$. We have used the accuracy of D_{rk} and $D_{\ell k}$ in the penultimate equality in (C.3), i.e., $D_{rk}\boldsymbol{x}_k = \lambda_r$ and $D_{\ell k}\boldsymbol{x}_k = -\lambda_\ell$, where λ_ℓ and λ_r are at least order h^{p+1} approximations of λ at the left and right boundaries of Ω_k, respectively. Hence, there is no vector spanned by $\{\boldsymbol{1}, \boldsymbol{x}_k\}$ other than $\boldsymbol{v}_c$ that is in the nullspace of M_k. If there exists a nontrivial vector $\boldsymbol{v}$ such that $\boldsymbol{v} \notin \mathrm{span}\{\boldsymbol{1}, \boldsymbol{x}_k\}$ and $M_k\boldsymbol{v} = \boldsymbol{0}$, then

$$H_k D_k^{(2)}(\Lambda)\boldsymbol{v} = E_k\Lambda_k D_{b,k}\boldsymbol{v} \neq \boldsymbol{0}, \tag{C.4}$$

because $D_k^{(2)}(\Lambda)$ is nullspace consistent and H_k is SPD. The vector $E_k\Lambda_k D_{b,k}\boldsymbol{v}$ has zero entries at rows corresponding to the zero rows of the E_k matrix. By construction, $D_k^{(2)}(\Lambda)$ has larger dense blocks at the top left and bottom right corners (consisting of more rows and columns) than the E_k matrix; therefore, it follows from the nullspace consistency of the $D_k^{(2)}(\Lambda)$ matrix that $[H_k D_k^{(2)}(\Lambda)\boldsymbol{v}]_i \neq 0$ and $[E_k\Lambda_k D_{b,k}\boldsymbol{v}]_i = 0$, at least for one entry, the i-th entry, near the boundaries. This implies that the equality in (C.4) cannot hold for any vector $\boldsymbol{v} \notin \mathrm{span}\{\boldsymbol{1}, \boldsymbol{x}_k\}$; hence, we have $\mathcal{N}(M_k) = \boldsymbol{v}_c$. Since $M_k = M_k^T$, it follows that $\mathcal{N}(M_k^T) = \boldsymbol{v}_c$. Using the result in (4.27) with the fact that $\mathcal{N}(M_k) = \mathcal{N}(M_k^T) = \boldsymbol{v}_c$, we obtain

$$V_k\boldsymbol{v}_0 = D_{b,k}^{-T}(M_k + M_k^T)D_{b,k}^{-1}\boldsymbol{v}_0 = D_{b,k}^{-1}(M_k + M_k^T)\boldsymbol{v}_c = \boldsymbol{0}. \tag{C.5}$$

Thus, $\boldsymbol{v}_0$ is the only nontrivial vector in the nullspace of V_k. Analogous results hold for V_v.

(a) $p = 1$

(b) $p = 2$

(c) $p = 3$

(d) $p = 4$

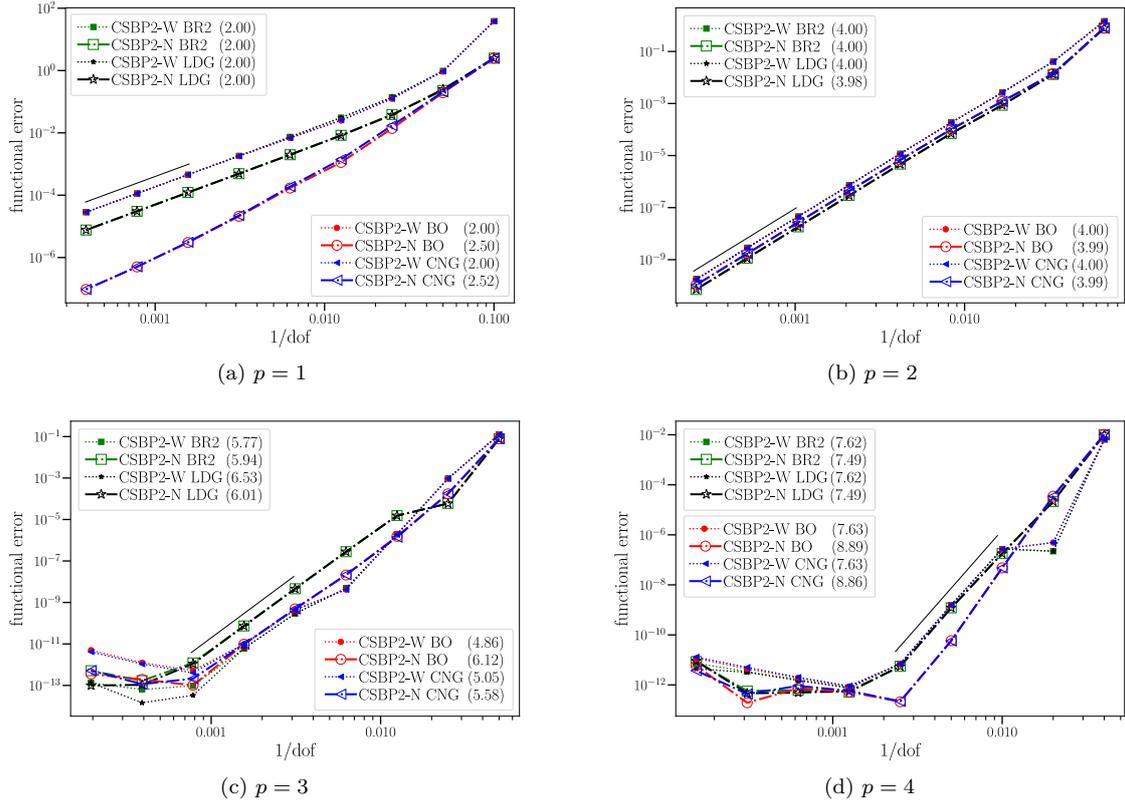**Fig. 5.8** Functional convergence under mesh refinement with block-norm wide-stencil ("W") and narrow-stencil ("N") CSBP operators. The values in parentheses are the convergence rates. The short, thin lines indicate the mesh refinement levels used to compute the convergence rates.

D Proof of Theorem 4.3

We wish to show that the matrices in (4.30) are positive semidefinite. The matrix A, whose components are given in (4.31), is symmetric; thus, we can use Theorem 4.2 to determine the conditions required for it to be positive semidefinite. We have $\mathsf{A}_{22} \succeq 0$ because V_k and V_v are positive semidefinite, $\alpha_{\gamma k} > 0$, and $\alpha_{\gamma v} > 0$. Therefore, the first condition in Theorem 4.2 is satisfied. The second condition in Theorem 4.2 requires that

$$\left(\mathsf{I}_{2n_p} - \mathsf{A}_{22}\mathsf{A}_{22}^+\right)\mathsf{A}_{21} = 2\begin{bmatrix} \mathsf{I}_{n_p} - \mathsf{V}_k\mathsf{V}_k^+ & 0 \\ 0 & \mathsf{I}_{n_p} - \mathsf{V}_v\mathsf{V}_v^+ \end{bmatrix}\begin{bmatrix} \mathsf{C}_{\gamma k}^T\mathsf{T}_{\gamma k}^{(2)} & -\mathsf{C}_{\gamma k}^T\mathsf{T}_{\gamma k}^{(2)} \\ -\mathsf{C}_{\gamma v}^T\mathsf{T}_{\gamma v}^{(2)} & \mathsf{C}_{\gamma v}^T\mathsf{T}_{\gamma v}^{(2)} \end{bmatrix} = 0, \quad (\text{D.1})$$

where I_{2n_p} and I_{n_p} are identity matrices of size $2n_p \times 2n_p$ and $n_p \times n_p$, respectively. To show that (D.1) holds, we consider the singular value decomposition of V_k,

$$\mathsf{V}_k = \mathsf{X}\Sigma\mathsf{Y}^T, \quad (\text{D.2})$$

where the columns of X and Y contain orthonormal basis vectors of the column and row spaces, respectively, and Σ is a diagonal matrix containing the singular values of V_k along its diagonal. Lemma 4.1 and Assumption 4 ensure that the matrix $\mathsf{V}_k \in \mathbb{R}^{n_p \times n_p}$ has only one nontrivial vector in its nullspace; hence, the first $n_p - 1$ columns of X contain orthonormal basis vectors that span the column space of V_k and the last column contains the vector in the nullspace of V_k, which is $\boldsymbol{v}_0$. We also note that

$$\mathsf{V}_k\mathsf{V}_k^+ = \mathsf{X}\Sigma\mathsf{Y}^T\mathsf{Y}\Sigma^+\mathsf{X}^T = \mathsf{X}\Sigma\Sigma^+\mathsf{X}^T = \mathsf{X}\mathsf{I}_m\mathsf{X}^T, \quad (\text{D.3})$$

where we have used the orthonormality of Y in the second equality, i.e., $\mathsf{Y}^T\mathsf{Y} = \mathsf{I}_{n_p}$, and I_m denotes an identity matrix of size $n_p \times n_p$ with the last diagonal entry set to zero (since $\Sigma_{ii} = 0$ for $i > m$, where $m = n_p - 1$ is the rank of V_k). Therefore, for operators that include the boundary nodes, we have $\mathsf{I}_{n_p} - \mathsf{V}_k\mathsf{V}_k^+ = \mathsf{I}_{n_p} - \mathsf{X}\mathsf{I}_m\mathsf{X}^T$, which gives

$$\mathsf{I}_{n_p} - \mathsf{X}\mathsf{I}_m\mathsf{X}^T = \mathsf{I}_{n_p} - \begin{bmatrix} \times & \ldots & \times & 0 \\ \times & \ldots & \times & \times \\ \vdots & & \vdots & \vdots \\ \times & \ldots & \times & \times \\ \times & \ldots & \times & 0 \end{bmatrix}\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & & 0 \end{bmatrix}\begin{bmatrix} \times & \times & \ldots & \times & \times \\ \vdots & \vdots & & \vdots & \vdots \\ \times & \times & \ldots & \times & \times \\ 0 & \times & \ldots & \times & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ 0 & \times & \ldots & \times & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & \times & \ldots & \times & 0 \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix}, \quad (\text{D.4})$$

where $\times$ denotes an entry that we do not need to specify for this analysis. Similarly, for operators that do not include boundary nodes, we obtain

$$
\mathsf{I}_{n_p} - \mathsf{V}_k \mathsf{V}_k^+ = \mathsf{I}_{n_p} - \mathsf{X} \mathsf{I}_m \mathsf{X}^T = \begin{bmatrix} 0 & & & & & & & \\ & \ddots & & & & & & \\ & & 0 & & & & & \\ & & & \times & \dots & \times & & \\ & & & \vdots & & \vdots & & \\ & & & \times & \dots & \times & & \\ & & & & & & 0 & \\ & & & & & & & \ddots \\ & & & & & & & & 0 \end{bmatrix}, \tag{D.5}
$$

i.e., the first and last s rows are zero, where s is half of the number of zero entries in $\boldsymbol{v}_0$. For operators that have nodes at the boundaries, the LHS of (D.1) can be evaluated using (D.4) as

$$
\left(\mathsf{I}_{2n_p} - \mathsf{A}_{22} \mathsf{A}_{22}^+ \right) \mathsf{A}_{21} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & & & & & \\ 0 & \times & \dots & \times & 0 & & & & & \\ \vdots & \vdots & & \vdots & \vdots & & & & & \\ 0 & \times & \dots & \times & 0 & & & & & \\ 0 & 0 & \dots & 0 & 0 & & & & & \\ & & & & & 0 & 0 & \dots & 0 & 0 \\ & & & & & 0 & \times & \dots & \times & 0 \\ & & & & & \vdots & \vdots & & \vdots & \vdots \\ & & & & & 0 & \times & \dots & \times & 0 \\ & & & & & 0 & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \mathsf{T}_{\gamma k}^{(2)} \lambda_{\gamma k} & -\mathsf{T}_{\gamma k}^{(2)} \lambda_{\gamma k} \\ -\mathsf{T}_{\gamma v}^{(2)} \lambda_{\gamma v} & \mathsf{T}_{\gamma v}^{(2)} \lambda_{\gamma v} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = 0. \tag{D.6}
$$

Similarly, substituting (D.5) into (D.1), it is straightforward to show that the condition $\left(\mathsf{I}_{2n_p} - \mathsf{A}_{22} \mathsf{A}_{22}^+ \right) \mathsf{A}_{21} = 0$ also holds for operators that do not include boundary nodes. For A to be positive semidefinite, it remains to find sufficient conditions to satisfy the last requirement in Theorem 4.3, i.e., $\mathsf{A}_{11} - \mathsf{A}_{12} \mathsf{A}_{22}^+ \mathsf{A}_{21} \succeq 0$, which, after some algebra, gives the condition

$$
\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes \left[\mathsf{T}_1 - \left(\frac{2}{\alpha_{\gamma k}} \mathsf{T}_{\gamma k}^{(2)} \mathsf{C}_{\gamma k} \mathsf{V}_k^+ \mathsf{C}_{\gamma k}^T \mathsf{T}_{\gamma k}^{(2)} + \frac{2}{\alpha_{\gamma v}} \mathsf{T}_{\gamma v}^{(2)} \mathsf{C}_{\gamma v} \mathsf{V}_v^+ \mathsf{C}_{\gamma v}^T \mathsf{T}_{\gamma v}^{(2)} \right) \right] \succeq 0, \tag{D.7}
$$

where $\otimes$ denotes the Kronecker product. Since $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \succeq 0$, the inequality in (D.7) is satisfied if

$$
\mathsf{T}_\gamma^{(1)} \geq \frac{2}{\alpha_{\gamma k}} \mathsf{T}_{\gamma k}^{(2)} \mathsf{C}_{\gamma k} \mathsf{V}_k^+ \mathsf{C}_{\gamma k}^T \mathsf{T}_{\gamma k}^{(2)} + \frac{2}{\alpha_{\gamma v}} \mathsf{T}_{\gamma v}^{(2)} \mathsf{C}_{\gamma v} \mathsf{V}_v^+ \mathsf{C}_{\gamma v}^T \mathsf{T}_{\gamma v}^{(2)}, \tag{D.8}
$$

which is the same as the condition given by (4.32). Note that $\mathsf{C}_{\gamma k} \mathsf{V}_k^+ \mathsf{C}_{\gamma k}^T = \mathsf{R}_{\gamma k} \Lambda_k \mathsf{V}_k^+ \Lambda_k \mathsf{R}_{\gamma k}^T$ since $n_{\gamma k}^2 = 1$.

The second matrix in (4.30) can be written as

$$
\begin{bmatrix} 2\mathsf{T}_{\gamma k}^{(4)} & 2\mathsf{T}_{\gamma k}^{(4)} \\ 2\mathsf{T}_{\gamma k}^{(4)} & 2\mathsf{T}_{\gamma k}^{(4)} \end{bmatrix} = 2\mathsf{T}_{\gamma k}^{(4)} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \tag{D.9}
$$

which is positive semidefinite provided $\mathsf{T}_{\gamma k}^{(4)} = \mathsf{T}_{\gamma v}^{(4)} \geq 0$, since $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \succeq 0$. Finally, we note that the last matrix in (4.30) is symmetric, and $\alpha_{\gamma k} \mathsf{V}_k \succeq 0$. Furthermore, using the same approach used to obtain (D.6) it can be shown that

$$
(\mathsf{I}_{n_p} - \mathsf{V}_k \mathsf{V}_k^+)(-2\mathsf{C}_{\gamma k}^T) = 0 \tag{D.10}
$$

irrespective of whether or not the operator includes boundary nodes. The last condition required for positive semidefiniteness of the last matrix in (4.30) is

$$
2\mathsf{T}_{\gamma k}^{(D)} - 4\mathsf{C}_{\gamma k} \left(\alpha_{\gamma k} \mathsf{V}_k \right)^+ \mathsf{C}_{\gamma k}^T \geq 0, \tag{D.11}
$$

which is satisfied if (4.33) holds. Therefore, the conditions in Theorem 4.3 are indeed sufficient for energy stability.

Declarations

Conflicts of Interest

The authors have no conflicts of interest to declare.

Data Availability

Not applicable.

Acknowledgments

References

1. Albert, A.: Conditions for positive and nonnegative definiteness in terms of pseudoinverses. SIAM Journal on Applied Mathematics **17**(2), 434–440 (1969)
2. Arnold, D.N., Brezzi, F., Cockburn, B., Marini, L.D.: Unified analysis of discontinuous Galerkin methods for elliptic problems. SIAM Journal on Numerical Analysis **39**(5), 1749–1779 (2002)
3. Bassi, F., Rebay, S., Mariotti, G., Pedinotti, S., Savini, M.: A high-order accurate discontinuous finite element method for inviscid and viscous turbomachinery flows. In: R. Decuypere, G. Dibelius (eds.) Proceedings of the 2nd European Conference on Turbomachinery, Fluid Dynamics and Thermodynamics, pp. 99–109. Technologisch Instituut, Antwerpen, Belgium (1997)
4. Baumann, C.E., Oden, J.T.: A discontinuous hp finite element method for convection–diffusion problems. Computer Methods in Applied Mechanics and Engineering **175**(3-4), 311–341 (1999)
5. Berg, J., Nordström, J.: Superconvergent functional output for time-dependent problems using finite differences on Summation-By-Parts form. Journal of Computational Physics **231**(20), 6846–6860 (2012)
6. Carpenter, M.H., Gottlieb, D., Abarbanel, S.: Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: methodology and application to high-order compact schemes. Journal of Computational Physics **111**(2), 220–236 (1994)
7. Carpenter, M.H., Nordström, J., Gottlieb, D.: A stable and conservative interface treatment of arbitrary spatial accuracy. Journal of Computational Physics **148**(2), 341–365 (1999)
8. Carpenter, M.H., Nordström, J., Gottlieb, D.: Revisiting and extending interface penalties for multi-domain summation-by-parts operators. Journal of Scientific Computing **45**(1-3), 118–150 (2010)
9. Craig Penner, D.A., Zingg, D.W.: Superconvergent functional estimates from tensor-product generalized summation-by-parts discretizations in curvilinear coordinates. Journal of Scientific Computing **82**(2), 41 (2020)
10. Del Rey Fernández, D.C., Boom, P.D., Zingg, D.W.: A generalized framework for nodal first derivative summation-by-parts operators. Journal of Computational Physics **266**, 214–239 (2014)
11. Del Rey Fernández, D.C., Hicken, J.E., Zingg, D.W.: Simultaneous approximation terms for multi-dimensional summation-by-parts operators. Journal of Scientific Computing **75**(1), 83–110 (2018)
12. Del Rey Fernández, D.C., Zingg, D.W.: Generalized summation-by-parts operators for the second derivative. SIAM Journal on Scientific Computing **37**(6), A2840–A2864 (2015)
13. Eriksson, S.: A dual consistent finite difference method with narrow stencil second derivative operators. Journal of Scientific Computing **75**(2), 906–940 (2018)
14. Eriksson, S., Nordström, J.: Finite difference schemes with transferable interfaces for parabolic problems. Journal of Computational Physics **375**, 935–949 (2018)
15. Gallier, J.H.: Notes on the Schur complement. Penn Engineering pp. 1–12 (2010)
16. Gong, J., Nordström, J.: Interface procedures for finite difference approximations of the advection–diffusion equation. Journal of Computational and Applied Mathematics **236**(5), 602–620 (2011)
17. Gustafsson, B.: The convergence rate for difference approximations to general mixed initial-boundary value problems. SIAM Journal on Numerical Analysis **18**(2), 179–190 (1981)
18. Hartmann, R.: Adjoint consistency analysis of discontinuous Galerkin discretizations. SIAM Journal on Numerical Analysis **45**(6), 2671–2696 (2007)
19. Hartmann, R., Leicht, T.: Higher order and adaptive DG methods for compressible flows. In: Deconinck H, ed. VKI LS 2014-03: 37th Advanced VKI CFD Lecture Series: Recent Developments in Higher Order Methods and Industrial Application in Aeronautics, Dec. 9-12, 2013. Rhode-Saint-Genese, Belgium: Von Karman Institute for Fluid Dynamics (2014)
20. Hicken, J.E.: Output error estimation for summation-by-parts finite-difference schemes. Journal of Computational Physics **231**(9), 3828–3848 (2012)
21. Hicken, J.E., Del Rey Fernández, D.C., Zingg, D.W.: Multidimensional summation-by-parts operators: General theory and application to simplex elements. SIAM Journal on Scientific Computing **38**(4), A1935–A1958 (2016)
22. Hicken, J.E., Zingg, D.W.: Superconvergent functional estimates from summation-by-parts finite-difference discretizations. SIAM Journal on Scientific Computing **33**(2), 893–922 (2011)
23. Hicken, J.E., Zingg, D.W.: Summation-by-parts operators and high-order quadrature. Journal of Computational and Applied Mathematics **237**(1), 111–125 (2013)
24. Hicken, J.E., Zingg, D.W.: Dual consistency and functional accuracy: a finite-difference perspective. Journal of Computational Physics **256**, 161–182 (2014)
25. Hunter, J.D.: Matplotlib: A 2d graphics environment. IEEE Annals of the History of Computing **9**(03), 90–95 (2007)
26. Kirby, R.M., Karniadakis, G.E.: Selecting the numerical flux in discontinuous Galerkin methods for diffusion problems. Journal of Scientific Computing **22**(1-3), 385–411 (2005)
27. Mattsson, K.: Summation by parts operators for finite difference approximations of second-derivatives with variable coefficients. Journal of Scientific Computing **51**(3), 650–682 (2012)
28. Mattsson, K., Almquist, M.: A solution to the stability issues with block norm summation by parts operators. Journal of Computational Physics **253**, 418–442 (2013)
29. Mattsson, K., Ham, F., Iaccarino, G.: Stable and accurate wave-propagation in discontinuous media. Journal of Computational Physics **227**(19), 8753–8767 (2008)
30. Mattsson, K., Nordström, J.: Summation by parts operators for finite difference approximations of second derivatives. Journal of Computational Physics **199**(2), 503–540 (2004)
31. Mattsson, K., Svärd, M., Shoeybi, M.: Stable and accurate schemes for the compressible Navier–Stokes equations. Journal of Computational Physics **227**(4), 2293–2316 (2008)

32. Peraire, J., Persson, P.O.: The compact discontinuous Galerkin (CDG) method for elliptic problems. SIAM Journal on Scientific Computing **30**(4), 1806–1824 (2008)
33. Sherwin, S., Kirby, R., Peiró, J., Taylor, R., Zienkiewicz, O.: On 2D elliptic discontinuous Galerkin methods. International Journal for Numerical Methods in Engineering **65**(5), 752–784 (2006)
34. Shu, C.W.: Different formulations of the discontinuous Galerkin method for the viscous terms. Z.-C. Shu, M. Mu, W. Xue, J. Zou (Eds.), Advances in Scientific Computing pp. 144–155 (2001)
35. Svärd, M., Nordström, J.: On the order of accuracy for difference approximations of initial-boundary value problems. Journal of Computational Physics **218**(1), 333–352 (2006)
36. Svärd, M., Nordström, J.: On the convergence rates of energy-stable finite-difference schemes. Journal of Computational Physics **397**, 108819 (2019)
37. Worku, Z.A., Zingg, D.W.: Simultaneous approximation terms and functional accuracy for diffusion problems discretized with multidimensional summation-by-parts operators. Journal of Computational Physics **445**, 110634 (2021)
38. Yan, J., Crean, J., Hicken, J.E.: Interior penalties for summation-by-parts discretizations of linear second-order differential equations. Journal of Scientific Computing **75**(3), 1385–1414 (2018)