A stable and accurate hybrid FE-FD method

Tuan Anh Dao
Abstract

A stable and accurate hybrid FE-FD method

Tuan Anh Dao

We develop a hybrid method to couple finite difference methods and finite element methods in a nonconforming multiblock fashion. The aim is to optimize computational efficiency when complex geometries present. The proposed coupling technique requires minimal changes in the existing schemes while maintaining strict stability, accuracy, and conservation. Analysis and computational results are shown for a linear problem (to the advection-diffusion equation) and a nonlinear problem (to the viscous Burger’s equation) in two spatial dimensions.
## Contents

1 Introduction .................................................. 3

2 Summation-by-parts operators ............................... 4
   2.1 The continuous problem to the linear advection-diffusion equation .......... 4
   2.2 Discrete advection-diffusion operators ........................................... 5

3 The two numerical schemes in summation-by-parts form .......... 6
   3.1 The finite difference (FD) scheme ................................................. 6
   3.2 The finite element (FE) scheme .................................................... 8

4 Simultaneous-approximation-term technique for interface coupling ...... 11
   4.1 Continuous analysis ................................................................. 11
   4.2 Necessary interface conditions for stability and conservation ............... 11
   4.3 FD–FD coupling ........................................................................... 12
   4.4 FD–FE coupling ........................................................................... 13

5 Numerical results .................................................. 15
   5.1 Eigenvalue analysis for validation of the interface treatment ................. 16
   5.2 Convergence study ........................................................................ 17

6 A nonlinear case – 2D Burger’s Equation ......................... 22
   6.1 The continuous problem ............................................................... 22
   6.2 FD schemes .................................................................................. 23
   6.3 FE schemes .................................................................................. 23
   6.4 The interface coupling .................................................................... 24
   6.5 Numerical results ........................................................................... 28

7 Conclusion ................................................................... 30

Appendices ................................................................. 31

A The choice of parameters for outer boundary SAT terms ............. 31

B The assembly of FE operators .......................................... 32
1 Introduction

Both finite element (FE) methods and finite difference (FD) methods are largely employed to obtain numerical solutions of partial differential equations (PDEs), which serve many applications in engineering and especially physics (e.g., fluid dynamics, solid mechanics). The FD methods are often seen as the most efficient scheme for high-order approximations despite being restricted to structured grids. The more flexible FE methods can handle complex geometries. However, the obtained matrices can sometimes be storage demanding and the method itself is considered to be more computationally expensive. The greatest strengths of both methods can be combined by developing a hybrid scheme that allows different methods and refinement levels to be involved simultaneously. Thus, computational efficiency can be optimized by using the most suitable treatment for each portion of the domain. The major complication is known to consist in designing a method-to-method interface that results in an accurate and stable approximation. An innovative method of this type should also accommodate the demands for simplicity, freedom of node distribution, and minimal interference in the existing schemes to obtain optimal performance while preserving the native properties of the involving methods.

A well-designed numerical framework for energy-stable approximation of time-dependent problems is the combination of the summation-by-parts (SBP) operators [7] (to approximate the governing equation) and the simultaneous-approximation-term (SAT) technique [3] (to impose boundary conditions). Some examples are [4, 6, 10, 13, 17, 20]. A crucial benefit of numerical schemes in this form is that proper formulation strictly prevents nonphysical energy growth, a property often referred to as “strict stability”. The vast majority of previous papers on the SBP-SAT framework were committed to FD methods due to apparent advantages of higher-order operators. Nontrivial geometries are then usually handled with curvilinear grids [17] and multiblock coupling [4]. To carry out more difficult geometries, successful attempts have been done to couple FD methods with unstructured finite volume (FV) methods (e.g., [18, 19]). Stability is achieved by modifying the FV scheme at the interface. It was not until the work of [12] when nonconforming grids using mixed schemes were made possible in an energy-stable manner. Nonmatching interfaces are resolved by the newly introduced SBP-preserving interpolation operators again maintaining fundamental properties of block-to-block coupling: strict stability, accuracy and conservation. This result paves the way for many possibilities to couple different methods that can be written in SBP form. So far hybrid methods of this type are available for the coupling of FD–FD [12], FD–discontinuous Galerkin (dG) method [6] and FD–FV [10].

Attempts to combine FD methods and the FE methods have rarely been successful despite it being of great interest. Several techniques developed for the seismic wave propagation problems involving surface topography (e.g., [11, 5]) use the concept of an overlapping region as the interface. With the same study objective in seismology, [9] utilizes the SBP-SAT framework for an added advantage of energy-conservation; however the coupling technique can only handle matching interface, and quadrilateral elements are required for the discretization of FE domain. Although there exists evidence that the standard Galerkin FE also has SBP properties [20], a big obstacle lies within its nondiagonal mass matrix (can be understood as the norm matrix in SBP manner). The general procedure [10] for coupling arbitrary SBP schemes is thus not directly applicable because diagonal norms are required. Some pivotal restrictions in [11, 5, 9] are also made in return for a diagonal mass matrix.

In this thesis, we propose a hybrid method to couple FD methods and FE methods in a nonconforming multiblock fashion. The approximation operators yielding from the formulation of the two methods are proved to have SBP properties. The interface continuity condition is weakly imposed by the SAT technique using the SBP-preserving interpolation operators [12] for nonconforming node distribution. The SAT treatment on the FD side is the same as for FD–FD coupling [12].
(hence FD operators of any order can be applied) while the treatment on the FE side is written in a general expression in the sense that it does not depend on the other method, the mesh type or the degree of polynomial of the solution space. We show that in both a linear case and a nonlinear case, nondiagonal norm matrices (i.e., the standard FE mass matrices) are applicable without further justification (e.g., mass lumping [20]) to the classical Galerkin formulation. We show further that strict stability, accuracy, and conservation are preserved.

An important part that could have been included in this thesis is the implementation of residual-based viscosity (RV) method [14, 15, 16] to handle nonsmooth solutions. RV method is a state-of-the-art stabilization technique greatly suited for the present study since an explicit expression can be derived and directly applied on both FD and FE in SBP form. However, within the limit of time, the author could not complete the correct formulation that allows the nonsmooth solution part to traverse the interface.

This thesis is organized as follows. In Section 2, we formulate a general SBP framework for the linear advection-diffusion problem. The numerical schemes of both FE and FD are derived and proved to have SBP properties in Section 3. The proposed coupling technique is described in Section 4. We show a numerical verification of the coupling treatment and some computational results in Section 5. Section 6 demonstrates the coupling technique for a nonlinear problem (to the viscous Burger’s equation); some numerical results are also shown. Section 7 is the conclusion.

2 Summation-by-parts operators

2.1 The continuous problem to the linear advection-diffusion equation

We employ the following 2D advection-diffusion problem for the main analysis

\[
\begin{align*}
  u_t + a^T \nabla u &= \varepsilon \nabla^T \nabla u, & (x, y) \in \Omega \subset \mathbb{R}^2 \quad (1) \\
  \frac{1}{2} (a^T n - |a^T n|) u - \varepsilon n^T \nabla u &= g(x, y, t), & (x, y) \in \partial \Omega, t > 0 \quad (2) \\
  u &= u_0(x, y), & (x, y) \in \Omega, t = 0
\end{align*}
\]

where \( a = (a_1, a_2)^T \) is a constant vector, \( \varepsilon > 0 \) is a small real number, \( \partial \Omega \) denotes the domain boundary, and \( n \) is the normal vector pointing outward at the boundary. The notation \( \nabla \) refers to the gradient operator \( \nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)^T \). For the two-dimensional analysis of finite difference schemes, it is more convenient to use the following alternative form of (1),

\[
  u_t + a_1 u_x + a_2 u_y = \varepsilon (u_{xx} + u_{yy}), \quad (x, y) \in \Omega \subset \mathbb{R}^2 \quad (3)
\]

where subscript notations \( x, y \) indicate the partial derivatives with respect to the corresponding variables.

Definition 2.1. Given two real-valued vector functions \( u \) and \( v \), we define the following continuous inner product and its corresponding norm

\[
  (u, v)_\Omega = \int_{\Omega} u^T v \, dx; \quad (u, v)_{\partial \Omega} = \int_{\partial \Omega} u^T v \, ds; \quad ||u||_\Omega^2 = (u, u)_\Omega = \int_{\Omega} u^T u \, dx.
\]

Where no confusion may arise, subscript domain notations are skipped for simplicity \( (\cdot, \cdot) \equiv (\cdot, \cdot)_\Omega \) and \( \| \cdot \| \equiv \| \cdot \|_\Omega \). Well-posedness can be shown for (1) with the Robin boundary condition (2) using the energy method. For continuous problems that can be written in the form \( u_t = F(u) \), a basic idea of the energy method is to investigate the energy growth utilizing the equality \( \frac{d}{dt} ||u||_\Omega^2 = (u, u_t)_\Omega + (u_t, u)_\Omega \). A well-posed problem ensures that \( \frac{d}{dt} ||u||_\Omega^2 \) is nonpositive given zero boundary data.
(energy damping, if applicable, only through the boundary). Applying the following continuous integration rules
\[ (v, a^T \nabla u)_\Omega = (v, a^T \mathbf{n} u)_{\partial \Omega} - (a^T \nabla v, u)_\Omega \]
\[ (v, \nabla^T \nabla u)_\Omega = (v, \mathbf{n}^T \nabla u)_{\partial \Omega} - (\nabla v, \nabla u)_\Omega, \]  
(4)
we have
\[ (u, u)_\Omega = -(u, a^T \mathbf{n} u)_{\partial \Omega} + (a^T \nabla u, u)_\Omega + \varepsilon(u, \mathbf{n}^T \nabla u)_{\partial \Omega} - \varepsilon(\nabla u, \nabla u)_\Omega \]
\[ (u, u)_\Omega = -(a^T \nabla u, u)_\Omega + \varepsilon(\mathbf{n}^T \nabla u, u)_{\partial \Omega} - \varepsilon(\nabla u, \nabla u)_\Omega. \]
Adding up the two equalities above yields
\[ \frac{d}{dt} \left( \|u\|_{\Omega}^2 \right) + 2\varepsilon \|\nabla u\|_{\Omega}^2 = -(u, a^T \mathbf{n} u)_{\partial \Omega} + 2\varepsilon(u, \mathbf{n}^T \nabla u)_{\partial \Omega}. \]
It is now easy to see that the boundary condition (2) imposes well-posedness to (1) since bounded energy is assured alongside. Indeed, inserting (2) with zero boundary data \( g = 0 \) gives
\[ \frac{d}{dt} \left( \|u\|_{\Omega}^2 \right) + 2\varepsilon \|\nabla u\|_{\Omega}^2 = -(u, |a^T \mathbf{n}| u)_{\partial \Omega}. \]  
(5)

2.2 Discrete advection-diffusion operators

In this section, we formulate a general framework that can later unify the two numerical methods. A set of matrix operators \((\mathcal{H}, \mathcal{B}, \mathcal{L})\) is normally required for any numerical scheme in SBP form, where
\[ \diamond \mathcal{H} \text{ is an integration operator over the whole domain;} \]
\[ \diamond \mathcal{B} \text{ is an integration operator over the boundary;} \]
\[ \diamond \mathcal{L} \text{ is a boundary selection operator.} \]
For the current study, we restrict \((\mathcal{H}, \mathcal{B}, \mathcal{L})\) into the assumptions: (i) \(\mathcal{H}, \mathcal{B}\) are symmetric, positive definite; \(\mathcal{H}\) is of size \(N \times N\) where \(N\) is the number of nodes; and \(\mathcal{B}\) is of size \(N_{\partial \Omega} \times N_{\partial \Omega}\) where \(N_{\partial \Omega}\) is the number of nodes on the boundary; (ii) \(\mathcal{L}\) comprises a nonsquare zero-one matrix of which the only “one” element in each row one-to-one corresponds to a boundary node. Notice that the restriction (i) is weaker than the corresponding requirement in [10] in which, in addition, \(\mathcal{H}\) and \(\mathcal{B}\) need to be diagonal. This allows the use of nondiagonal norm matrices yielding from standard Galerkin FE approximations.

**Definition 2.2.** Given two discrete solutions \(u, v \in \mathbb{R}^N\), we define the following *discrete inner product* and its corresponding *discrete norm*
\[ (u, v)_{\mathcal{H}} = u^T \mathcal{H} v; (\mathcal{L} u, \mathcal{L} v)_{\mathcal{B}} = (\mathcal{L} u)^T \mathcal{B} (\mathcal{L} v); \|u\|_{\mathcal{H}}^2 = (u, u)_{\mathcal{H}} = u^T \mathcal{H} u. \]
Advantages of an SBP scheme come from the fact that its discrete operators, by design, mimic the continuous integration-by-parts. Below we define a discrete advection operator and a discrete diffusion operator that hold properties analogous to the continuous integration rules (4).

**Definition 2.3.** A matrix operator \(\mathcal{H}^{-1} \mathcal{Q}\) approximating \(a^T \nabla\) is said to be an *SBP advection operator* if
\[ \mathcal{Q} + \mathcal{Q}^T = \mathcal{L}^T \mathcal{B} \text{Diag}(a^T \mathbf{n}) \mathcal{L}. \]  
(6)

**Definition 2.4.** A matrix operators \(\mathcal{H}^{-1} \mathcal{R}\) approximating \(\nabla^T \nabla\) is said to be an *SBP diffusion operator* if
\[ \mathcal{R} = \mathcal{L}^T \mathcal{B} \mathcal{S} - \mathcal{A}. \]  
(7)
where \( A = A^T \geq 0 \) and \( S \) is a consistent approximation of \( n^T \nabla \) at the boundary. Moreover, if there exist consistent approximations \( D_{1x}, D_{1y} \) of \( \partial / \partial x, \partial / \partial y \) respectively, such that \( A = D_1^T (I_2 \otimes \mathcal{H}) D_1 \), where \( I_2 \) is the identity matrix of size \( 2 \times 2 \), \( D_1 = [D_{1x}^T, D_{1y}^T]^T \), then \( \mathcal{H}^{-1} \mathcal{R} \) is called a complete SBP diffusion operator (see [13]).

The energy method can be analogously applied to SBP schemes utilizing the properties (6), (7), and the relation \( \frac{\partial}{\partial t} \| u \|_H^2 = u^T \mathcal{H} u + u^T \mathcal{H} u_t \). A strictly stable approximation ensures that \( \frac{\partial}{\partial t} \| u \|_H^2 \) is nonpositive given zero boundary data.

### 3 The two numerical schemes in summation-by-parts form

#### 3.1 The finite difference (FD) scheme

Consider a one-dimensional domain \( [x_l, x_r] \) discretized by \( n \) equidistant grid points \( \{x_i \equiv x_1, x_2, \ldots, x_n \equiv x_r\} \). We first define some general one-dimensional SBP operators for FD.

**Definition 3.1.** [13] A finite difference operator \( D_1 = H^{-1} Q \) approximating \( \partial / \partial x \) using \( q \)-th order interior stencils is called a \( q \)-th order accurate first derivative SBP operator if \( H = H^T > 0 \) and \( Q + Q^T = \text{Diag}(-1, 0, \ldots, 0, 1) \).

Let \( e_1 = (1, 0, \ldots, 0)^T, e_n = (0, \ldots, 0, 1)^T \) be column vectors of length \( n \), and \( d_1, d_n \) be row vectors approximating first derivatives at \( x_1, x_n \).

**Definition 3.2.** [13] A finite difference operator \( D_2 = H^{-1} (-M - e_1 d_1 + e_n d_n) \) approximating \( \partial^2 / \partial x^2 \) using \( q \)-th order interior stencils is called a \( q \)-th order accurate second derivative SBP operator if \( H = H^T > 0, M = M^T \geq 0 \).

**Definition 3.3.** Given matrices \( A = (a_{ij}) \) of size \( p \times q \) and \( B \) of arbitrary size, the Kronecker product of \( A \) and \( B \) is defined as

\[
A \otimes B = \begin{bmatrix}
a_{11}B & a_{12}B & \cdots & a_{1q}B \\
a_{21}B & a_{22}B & \cdots & a_{2q}B \\
\vdots & \vdots & \ddots & \vdots \\
a_{p1}B & a_{p2}B & \cdots & a_{pq}B
\end{bmatrix}
\]

The following properties of the Kronecker product are frequently used.

**Property 3.4.** Assuming that matrices \( A, B, C, D \) are of suitable sizes, we have

(i) \((A \otimes B)(C \otimes D) = (AC \otimes BD)\);

(ii) \((A \otimes B)^T = (A^T \otimes B^T)\).

Given a column vector \( \bar{a} \), for any matrices \( B, D \) of compatible dimensions, a direct consequence of Property 3.4 (i) is \((\bar{a} \otimes BD) = (\bar{a} \otimes B)D \) and \((\bar{a}^T \otimes BD) = B(\bar{a}^T \otimes D)\).

FD SBP schemes for higher number of spatial dimensions are naturally extended from the one-dimensional operators. Consider a bounded rectangle domain \( \Omega = [x_l, x_r] \times [y_l, y_r] \subset \mathbb{R}^2 \) illustrated in Figure 1. The discrete solution is approximated on \( n \times m \) equidistant grid points \( \{(x_i, y_j), i = 1, 2, \ldots, n, j = 1, 2, \ldots, m\} \), where

\[
x_i = x_l + (i - 1)h_x, \quad h_x = \frac{x_r - x_l}{n - 1}, \\
y_j = y_l + (j - 1)h_y, \quad h_y = \frac{y_r - y_l}{m - 1}
\]
Figure 1: A rectangle domain discretized by an equidistant grid

A solution vector $u$ approximating a continuous function $\psi$ is then presented by

$$u = \begin{bmatrix} u_{1,1}, u_{1,2}, \ldots, u_{1,m}, & u_{2,1}, u_{2,2}, \ldots, u_{2,m}, & \ldots & u_{n,1}, u_{n,2}, \ldots, u_{n,m} \end{bmatrix}^T$$

where $u_{i,j}$ approximates $\psi(x_i, y_j)$, which can be interpreted as a “vector of vectors”.

Some necessary operators for the two-dimensional approximation are listed below. Denote the identity matrix of size $k \times k$ by $I_k$. The following finite difference operators will be frequently used in the sequel,

$$D_x = D_1 \otimes I_m, \quad D_y = I_n \otimes D_1,$$
$$D_{2x} = D_2 \otimes I_m, \quad D_{2y} = I_n \otimes D_2,$$
$$H_x = H_x \otimes I_m, \quad H_y = I_n \otimes H_y,$$

where $H_x, H_y$ are the one-dimensional norm matrices by $x-$ and $y-$ direction. For ease of reading, it is worth noting that the first components in all of the above Kronecker products are of size $n \times n$, and the second components are of size $m \times m$. The corresponding two-dimensional norm is $H = H_x \cdot H_y = H_x \otimes H_y$.

A finite difference semi-discretization of (3) using diagonal norm SBP operators is given by

$$u_t = \begin{bmatrix} -a_1 D_x - a_2 D_y + \varepsilon D_{2x} + \varepsilon D_{2y} \end{bmatrix} u + \text{SAT}. \quad (9)$$

The following proposition is obvious from the way the above operators are defined, but useful for the connection with finite element schemes in the next section.

**Proposition 3.1.** The approximation operators used in (9) satisfy the summation-by-parts rules (6) and (7), where

$$\mathcal{H} = H;$$
$$\mathcal{H}^{-1} Q = a_1 D_x + a_2 D_y;$$
$$\mathcal{H}^{-1} R = D_{2x} + D_{2y}.$$

Since this property is intended by definition. No proof is given.

The normal vectors at the north, east, south, west boundaries are $n_N = (0, 1)^T, n_E = (1, 0)^T, n_S = (0, -1)^T$ and $n_W = (-1, 0)^T$, respectively. The boundary condition (2) can be written more
precisely at the boundaries

\[
\begin{align*}
\text{North} : & \quad \frac{a_2 - |a_2|}{2} u - \epsilon u_y = g_N(t), \\
\text{East} : & \quad \frac{a_1 - |a_1|}{2} u - \epsilon u_x = g_E(t), \\
\text{South} : & \quad \frac{-a_2 - |a_2|}{2} u + \epsilon u_y = g_S(t), \\
\text{West} : & \quad \frac{-a_1 - |a_1|}{2} u + \epsilon u_x = g_W(t).
\end{align*}
\]

We show that the following penalty terms weakly forcing (2) yield stability to (9),

\[
\begin{align*}
\text{SAT} & = \text{SAT}_N + \text{SAT}_E + \text{SAT}_S + \text{SAT}_W, \\
\text{SAT}_N & = \tau_N \left( \frac{a_2 - |a_2|}{2} u_N - \epsilon (D_y u)_N - g_N \right) \otimes (H^{-1}_y e_m), \\
\text{SAT}_E & = \tau_E (H^{-1}_x e_n) \otimes \left( \frac{a_1 - |a_1|}{2} u_E - \epsilon (D_x u)_E - g_E \right), \\
\text{SAT}_S & = \tau_S \left( -\frac{a_2 - |a_2|}{2} u_S + \epsilon (D_y u)_S - g_S \right) \otimes (H^{-1}_y e_1), \\
\text{SAT}_W & = \tau_W (H^{-1}_x e_1) \otimes \left( -\frac{a_1 - |a_1|}{2} u_W + \epsilon (D_x u)_W - g_W \right),
\end{align*}
\]  

(10)

where

\[
\begin{align*}
& u_N = (I_n \otimes e_m^T) u, & & u_E = (e_n^T \otimes I_m) u, & & u_S = (I_n \otimes e_1^T) u, & & u_W = (e_1^T \otimes I_m) u, \\
& (D_y u)_N = (I_n \otimes d_m) u, & & (D_x u)_E = (d_n \otimes I_m) u, & & (D_y u)_S = (I_n \otimes d_1) u, & & (D_x u)_W = (d_1 \otimes I_m) u.
\end{align*}
\]

(11)

The following theorem completes the formulation of the finite difference approximation to (1), (2).

**Theorem 3.5.** By choosing \( \tau_N = \tau_E = \tau_S = \tau_W = 1 \), the approximation (9) with the outer boundary SAT terms given in (10) is stable.

**Proof.** See Appendix A. \( \Box \)

### 3.2 The finite element (FE) scheme

**Weak formulation**

Multiplying both sides of (1) by a test function \( v \in V := \{v(\cdot, t) : \|v\|_{L^2(\Omega)} + \|\nabla v\|_{L^2(\Omega)} < \infty \}\) and integrating by parts, a weak formulation can be written as: find \( u \in V \) such that

\[
(\partial_t u, v) + (a^T \nabla u, v) = (\epsilon \nabla^T \nabla u, v) \quad \forall v \in V.
\]

(12)

Partition the boundary \( \partial \Omega \) into inflow part and outflow part, \( \partial \Omega = \Gamma_{in} \cup \Gamma_{out} \), where

\[
\begin{align*}
\Gamma_{in} & = \{(x, y) \in \partial \Omega \mid a^T n < 0\} \\
\text{and} \quad \Gamma_{out} & = \{(x, y) \in \partial \Omega \mid a^T n \geq 0\}.
\end{align*}
\]

For example, if \( a \geq 0 \), it is easy to see that \( \Gamma_{in} = \Gamma_S \cap \Gamma_W \) consists of the South and West boundaries and \( \Gamma_{out} = \Gamma_N \cap \Gamma_E \) consists of the North and East boundaries, denoted respectively. The boundary condition (2) becomes

\[
\begin{align*}
\left\{ \begin{array}{ll}
a^T n u - \epsilon n^T \nabla u = g & \quad \text{on} \ \Gamma_{in}, \\
-\epsilon n^T \nabla u = g & \quad \text{on} \ \Gamma_{out}.
\end{array} \right.
\end{align*}
\]

(13)
Applying the Green’s formula on (12), we obtain
\[ (\partial_t u, v) + (a^T \nabla u, v) + \varepsilon(\nabla u, \nabla v) - \int_{\partial \Omega} (\varepsilon n^T \nabla u) v ds = 0, \forall v \in V. \]
Inserting (13) gives
\[ (\partial_t u, v) = -(a^T \nabla u, v) - \varepsilon(\nabla u, \nabla v) + \int_{\Gamma_{in}} (a^T n u - g) v ds - \int_{\Gamma_{out}} g v ds, \forall v \in V. \]

**Galerkin finite element approximation**

A finite element approximation can then be formulated as: find \( u_h \in V_h := \{ v(., t) \in C^0(\Omega) \mid v|_K \in P^1(K), \forall K \in T_h \} \) such that
\[
(\partial_t u_h, v) = -(a^T \nabla u_h, v) - \varepsilon(\nabla u_h, \nabla v) + \int_{\Gamma_{in}} (a^T n u_h - g) v ds - \int_{\Gamma_{out}} g v ds, \forall v \in V_h. \tag{14}
\]
where \( T_h = \{ K_i \}_{i=1}^N \approx \Omega \), \( N \) is the number of nodes the mesh and \( P^1(K_i) \) is the space of linear interpolations on \( K_i \). We can write \( V_h = \text{span}\{ \varphi_i \}_{i=1}^N \), where \( \varphi_i, i = 1, \ldots, N \) are piecewise linear basis functions, \( \varphi_i = 1 \) at the \( i \)-th vertex, zero elsewhere. Since \( u_h \in V_h \), there exist real numbers \( \{ \xi_i \}_{i=1}^N \) such that \( u_h = \sum_{j=1}^N \xi_j(t) \varphi_j(x, y) \). Inserting this into (14) and letting \( v = \varphi_i, i = 1, \ldots, N \) gives
\[
M \dot{\xi}_i(t) = -C \xi_i(t) - \varepsilon A \xi_i(t) + R_{M}^{\text{in}} \xi_i(t) - \tau_i, \tag{15}
\]
where the over dot denotes the time derivative; \( M, C, A, R_{M}^{\text{in}}, \tau \) are respectively called the mass matrix, the convection matrix, the stiffness matrix, the Robin boundary mass matrix, and the Robin boundary vector; to be more precise, for \( i, j = 1, \ldots, N \),
\[
M_{ij} = (\varphi_j, \varphi_i), \\
C_{ij} = (a^T \nabla \varphi_j, \varphi_i), \\
A_{ij} = (\nabla \varphi_j, \nabla \varphi_i), \\
R_{M}^{\text{in}}_{ij} = \int_{\Gamma_{in}} a^T n \varphi_j \varphi_i ds, \\
\tau_i = \int_{\partial \Omega} g \varphi_i ds.
\]
The assembly of the above operators is described in [8]. Interestingly, if we treat (15) in an SBP manner by considering \( u(t) = \xi(t) \) as an approximation vector of the solution at mesh points and applying the energy method with zero boundary data (i.e., writing it in the form \( u_t = F(u) \), multiplying both side by \( a^T M \), integrating by parts and adding the transpose \( a^T M v \)), it deduces that
\[
\frac{d}{dt} \| u \|^2_M + 2\varepsilon u^T A u = -u^T (C + C^T) u + 2u^T R_{M}^{\text{in}} u.
\]
To achieve the above equality, the symmetry and positive definiteness of \( M \) is necessary \( (M^{-1})^T = M^{-1} \). We now apply the continuous integration rule (4) on \( C + C^T \) to obtain
\[
\frac{d}{dt} \| u \|^2_M + 2\varepsilon u^T A u = -u^T R_{M}^{\Omega} u + 2u^T R_{M}^{\text{in}} u = -u^T [R_M] u,
\]
where \( R_{M}^{\Omega}_{ij} = \int_{\partial \Omega} a^T n \varphi_j \varphi_i ds \). In the interior, the sum \( C + C^T \) vanishes because the local contributions of adjoining triangles give opposite signs of the same value at both ends of their shared edge. Since the stiffness matrix \( A \) is positive definite, the above equality shows stability in the same way regularly seen in SBP finite difference approximations (see Appendix A). It is also analogous to the continuous estimate (5). This observation reveals some SBP properties of (15).
Proposition 3.2. The approximation operators used in (15) satisfy the summation-by-parts rules (6) and (7), where

\[ H = M; \]
\[ H^{-1} Q = M^{-1} C; \]
\[ H^{-1} R = -M^{-1} A + M^{-1} R^{\partial \Omega}_C, \]

where \( R^{\partial \Omega}_{C_{ij}} = \int_{\partial \Omega} n^T \nabla \phi_j \phi_i ds, i, j = 1, 2, \ldots, N. \)

Proof. We prove that the FE advection operator \( Q = C \) satisfies (6). From the above analysis, we already have \( C + C^T = R^{\partial \Omega}_M \). We need to show that \( R^{\partial \Omega}_M = L^T B \text{Diag}(a^T n) L \) with the FE formulation of \( B \) and \( L \). For a general boundary selection operator \( L \), each row of \( L \) must be a one-to-one mapping from its “one” element with a node on the boundary. The combination of a left multiplier \( L^T \) and a right multiplier \( L \) thus simply expands the local boundary operators into the full domain size (with zeros on the extended rows and columns). The relation (6) becomes obvious with the local boundary integration operator inserted \( B_{ij} = \int_{\partial \Omega} \phi_j \phi_i ds \), for \( i, j = 1, 2, \ldots, N_{\partial \Omega} \), where \( N_{\partial \Omega} \) is the number of boundary nodes.

For the diffusion operator \( R = -A + R^{\partial \Omega}_M \), we need to show that \( -A + R^{\partial \Omega}_M = L^T B S - A \) with the FE formulation of \( B, L, D_1, S, \) and \( H \). Firstly, we formulate \( S \) approximating \( n^T \nabla u \) (i.e., given \( u_h \in V_h, S(u_h) = n^T \nabla u_h \)). We seek this approximation in the space \( V_h \). Since \( n^T \nabla u \) does not belong to \( V_h \), the projection \( pS(u_h) \) of \( S(u_h) \) onto \( V_h \) must be used instead. Namely, we find \( pS(u_h) \in V_h \) such that

\[ (pS(u_h) - S(u_h), v)_{\partial \Omega} = 0, \forall v \in V_h \]

\[ (pS(u_h), v)_{\partial \Omega} = (n^T \nabla u_h, v)_{\partial \Omega}, \forall v \in V_h. \]

Since \( pS(u_h) \in V_h \), there exists some \( \bar{n} = (\eta_1, \eta_2, \ldots, \eta_N) \) such that \( pS(u_h) = \sum_{j=1}^N \eta_j \phi_j \). Inserting this and \( u_h = \sum_{j=1}^N \xi_j \phi_j \), the above equation can be written in matrix form,

\[ \bar{B}_{\eta} = R^{\partial \Omega}_C \bar{\xi}, \]

(16)

where \( \bar{B}_{ij} = \int_{\partial \Omega} \phi_j \phi_i ds, i, j = 1, 2, \ldots, N \) and \( R^{\partial \Omega}_C \) is defined above. Since obviously \( \int_{\partial \Omega} \phi_j \phi_i ds = 0 \) if either \( i \) or \( j \) corresponds to a nonboundary node,

\[ L^T B L \eta = \bar{B}_{\eta} \]

(17)

The use of \( L \) is simply because \( B \) is defined in the SBP framework of size \( N_{\partial \Omega} \times N_{\partial \Omega} \). By definition, \( L \eta = S \bar{\xi} \). \( (S \) is in general not square because of (7).) From (16) and (17), we obtain

\[ R^{\partial \Omega}_C = L^T BS. \]

(18)

What is left is to prove the semi-positive definiteness of \( A \). Starting from a symmetric bilinear form in the weak formulation \( \Psi(u, v) := (\nabla u, \nabla v)_{\Omega} \), for \( u, v \in V \), applying the Poincaré’s inequality yields

\[ \Psi(v, v) = ||\nabla v||^2_{L^2(\Omega)} \geq c ||v||^2_{L^2(\Omega)}, \forall v \in V. \]

Since \( V_h \subset V \), for all \( v_h \in V_h \), the property still holds \( \Psi(v_h, v_h) \geq c ||v_h||^2_{L^2(\Omega)} \). Given \( \bar{\xi} = (\xi_1, \xi_2, \ldots, \xi_N)^T \) and \( v_h = \sum_{j=1}^N \xi_j \phi_j \), we have

\[ \bar{\xi}^T A \bar{\xi} = \sum_{i=1}^N \sum_{j=1}^N \xi_i \xi_j = \sum_{i=1}^N \sum_{j=1}^N \Psi(\xi_i \phi_i, \xi_j \phi_j) = \Psi \left( \sum_{i=1}^N \xi_i \phi_i, \sum_{j=1}^N \xi_j \phi_j \right) = \Psi(v_h, v_h) \geq 0. \]

(19)

From (18) and (19), the desired property of \( R \) is achieved. Note also that at some boundary points where \( n \) is not well-defined (e.g., corners of a rectangle domain), a boundary splitting [10] is necessary for the above proof to be valid. However, this proof routine does not change. \( \square \)
4 Simultaneous-approximation-term technique for interface coupling

4.1 Continuous analysis

Consider the coupling of two domains $\Omega_L$, $\Omega_R$ illustrated in Figure 2, where $u$ and $v$ present the parts of solution on the left and right domains, respectively,

\begin{align*}
  u_t + a^T \nabla u &= \varepsilon \nabla^T \nabla u, \quad (x, y) \in \Omega_L \\
  v_t + a^T \nabla v &= \varepsilon \nabla^T \nabla v, \quad (x, y) \in \Omega_R.
\end{align*}

\hspace{1cm} \text{(20)}

\begin{figure}[h]
  \centering
  \includegraphics[width=0.5\textwidth]{figure2.png}
  \caption{Coupling of two adjoining blocks}
  \end{figure}

The North, South, West boundaries of $\Omega_L$ and the North, East, South boundaries of $\Omega_R$ will be referred to as the “outer boundary” and the shared edge of the two adjoining blocks is called the “interface”, denoted $\Gamma$. Applying the energy method on (20), a total energy estimate of the whole solution on $\Omega_L \cup \Omega_R$ reads

\[ \frac{d}{dt} \left( \|u\|^2_{\Omega_L} + \|v\|^2_{\Omega_R} \right) + 2\varepsilon \|\nabla u\|^2_{\Omega_L} + 2\varepsilon \|\nabla v\|^2_{\Omega_R} = BT + IT, \]

where $BT$ contains all the outer boundary terms (being bounded with the boundary condition (2)) and $IT$ contains the interface terms given by

\[ IT = -(u, a^T n_E u)_{\Omega_I} + 2\varepsilon (u, n^T_E \nabla u)_{\Omega_I} - (v, a^T n_W v)_{\Omega_I} + 2\varepsilon (v, n^T_W \nabla v)_{\Omega_I} \]

\[ = -a_1 \|\Omega_I\|^2 + a_1 \|v\|^2_{\Omega_I} + 2\varepsilon (u, u_x)_{\Omega_I} - 2\varepsilon (v, v_x)_{\Omega_I}. \]

Therefore, the necessary continuity condition for an accurate coupling (also for conservation $IT \equiv 0$) is

\[ \begin{cases} 
  u = v \\
  u_x = v_x 
\end{cases} \quad \text{at } x \in \Gamma. \quad \text{(21)} \]

4.2 Necessary interface conditions for stability and conservation

An SBP-SAT semi-discretization of (20) is given by

\begin{align*}
  u_t + \mathcal{H}^{-1}_L Q_L u &= \varepsilon \mathcal{H}^{-1}_L R_L u + \text{SAT}_L + \text{SAT}_{IL} \\
  v_t + \mathcal{H}^{-1}_R Q_R v &= \varepsilon \mathcal{H}^{-1}_R R_R v + \text{SAT}_R + \text{SAT}_{IR}, \quad \text{(22)}
\end{align*}

where SAT$_{R,L}$ present the weak boundary treatment, and SAT$_{I,R,IL}$ present the weak interface treatment. Another advantage of this weakly forcing technique is that SAT$_{R,L,IR,IL}$ each can be
treated separately by the energy method. Therefore, in the following subsections, we assume the stable outer boundary treatment as described in Section 3 and only analyze the estimate terms regarding the additional involvement of SAT_{IR,IL}.

The nonmatching nodes on \( \Omega_L \) are resolved by the interpolation operators \( I_{L2R} \) and \( I_{R2L} \), where \( I_{L2R} \) maps the interface values from the left domain to the right domain, and \( I_{R2L} \) does the opposite direction (illustrated in Figure 3). For proven stability and energy conservation, the use of the following class of interpolation operators is necessary (first introduced in [12]).

**Definition 4.1.** Let \( H_y^L, H_y^R \) be some integration operators over the interface of the left and the right domains, respectively. \( \{ I_{L2R}, I_{R2L} \} \) is called a pair of **SBP-preserving interpolation operators** if

\[
H_y^R I_{L2R} = I_{R2L}^T H_y^L. \tag{23}
\]

For example, for FD, \( H_y \) is the \( H_y \)-norm that we defined earlier. For FV, this operator is a diagonal matrix containing the Euclid distances between two neighboring interface nodes [10]. For dG, this operator is a diagonal matrix presenting integration weights for interface edges [6]. However, it is quite intuitive that this matrix operator is not diagonal for FE. We will investigate this in more detail in Section 4.4.

### 4.3 FD–FD coupling

We consider the case where FD schemes are used on both \( \Omega_L \) and \( \Omega_R \). The coupling SAT treatment is given by

\[
\begin{align*}
\text{SAT}_{IL} &= \alpha_L(H_x^{L-1} e_n) \otimes (u_I - I_{R2L}v_I) \\
&\quad + \delta_L(H_x^{L-1} e_n) \otimes ((D_x u)_I - I_{R2L}(D_x v)_I) + \sigma_L(H_x^{L-1} d_n^T) \otimes (u_I - I_{R2L}v_I), \\
\text{SAT}_{IR} &= \alpha_R(H_x^{R-1} e_1) \otimes (v_I - I_{L2R}u_I) \\
&\quad + \delta_R(H_x^{R-1} e_1) \otimes ((D_x v)_I - I_{L2R}(D_x u)_I) + \sigma_R(H_x^{R-1} d_1^T) \otimes (v_I - I_{L2R}u_I), \tag{24}
\end{align*}
\]

where \( u_I \equiv u_E, v_I \equiv v_W, (D_x u)_I \equiv (D_x u)_E, (D_x v)_I \equiv (D_x v)_W \) are defined in (11).
Note. Though \(e_1, e_n, e_m, d_1, d_n, d_m\) all differ from the \(x\)- to \(y\)- dimensions, or from the left to right domains, it would be much less readable if we introduced separate notation for each case. Fortunately, there is unlikely serious confusion that may arise even when different meanings involve. We take serious care of situations like that.

**Theorem 4.2.** By choosing \(\alpha_L = \frac{1}{2} \alpha_1, \alpha_R = -\frac{1}{2} \alpha_1, \delta_L = -\sigma_R, \sigma_L = -\delta_R \) and \(\delta_L + \sigma_L = -\varepsilon\), assuming proper outer boundary treatment, the approximation (22) with the interface SAT terms (24) is stable.

**Proof.** Applying the energy method on each equation of (22) and adding them up, a total energy estimate can be written as

\[
\frac{d}{dt} \|u\|^2_H + 2\varepsilon u^T (M \otimes H^L_y) u + 2\varepsilon v^T (H^L_y \otimes M) u \\
+ \frac{d}{dt} \|v\|^2_H + 2\varepsilon v^T (M \otimes H^R_y) v + 2\varepsilon v^T (H^R_y \otimes M) v = BT + IT,
\]

where the outer boundary terms all contained in \(BT\) are assumed to be finely tuned (as given in (10) and Theorem 3.5); and the interface terms consist of

\[
IT = \varepsilon u^T (e_n d_n + (e_n d_n)^T \otimes H^L_y) u + \varepsilon v^T (-e_1 d_1 - (e_1 d_1)^T \otimes H^R_y) v \\
- a_1 u^T (e_n e_n^T \otimes H^L_y) u + a_1 v^T (e_1 e_1^T \otimes H^R_y) v \\
+ \alpha_L [2u^T (e_n e_n^T \otimes H^L_y) u - u^T (e_n e_n^T \otimes H^L_y I_{R2L}) u] \phi v^T (e_1 e_1^T \otimes I_{R2L} H^L_y) u] \\
+ \alpha_R [2v^T (e_1 e_1^T \otimes H^R_y) v - v^T (e_1 e_1^T \otimes H^R_y I_{L2R}) u] \phi u^T (e_n e_n^T \otimes I_{L2R} H^R_y) v] \\
+ \delta_L [u^T (e_n d_n + (e_n d_n)^T \otimes H^L_y) u - u^T (e_n d_n + (e_n e_n)^T \otimes H^L_y I_{R2L}) u] \phi v^T (e_1 d_1 + (e_1 d_1)^T \otimes H^R_y I_{L2R}) u] \\
+ \delta_R [v^T (e_1 d_1 + (e_1 d_1)^T \otimes H^R_y) v - v^T (e_1 d_1 + (e_1 d_1)^T \otimes H^R_y I_{L2R}) u] \phi u^T (e_n e_n^T \otimes I_{R2L} H^R_y) v]
\]

It can be verified that with the conservation property (23) and the above choice of parameters, all the interface terms vanish (i.e., \(IT \equiv 0\)). The interface treatment is thus stable. \(\square\)

### 4.4 FD–FE coupling

This section contains the main contribution of this thesis. We consider the case when FD is used on \(\Omega_L\) and FE is used on \(\Omega_R\). The following introduction of some appropriate boundary selection operators is essential to the SAT technique for FE schemes.

**Definition 4.3.** Given a rectangle domain, the zero-one matrices \(L_N, L_E, L_S, L_W\) of size \(m \times N\), where \(m\) is the number of nodes on the concerned boundary and \(N\) is the number of total nodes, are respectively called the North, East, South and West boundary FD–FE compatible selection operators if each satisfies the following properties

(i) each row of it contains exactly one “1” element uniquely corresponding to a node on the concerned boundary;

(ii) it rearranges the boundary nodes to the same spatial order of the adjoining FD boundary.

The following properties of the above boundary selection operators are useful for the proof of stability in this section.

**Property 4.4.** Assuming \(A, B\) are square matrices of compatible dimensions, for \(L \in \{L_N, L_E, L_S, L_W\}\), the following statements hold

(i) \(AL = B \implies A = BL^T\);
(ii) $BL^T = A \iff B = AL$ if $(I_N - L^T L)B^T = 0$.

Proof. We prove (i). From $AL = B$, right multiplying both sides by $L^T$ gives $ALL^T = BL^T$. Since each row of $L$ contains exactly one “1” element corresponding to the same position on each column of $L^T$, it is clear that $LL^T = I_m$. We have the first statement proven.

Similarly, we prove (ii) by right multiplying both sides by $L$ to obtain $BL^T L = AL$. However, since $L^T L \neq I_N$, the addition of $(I_N - L^T L)B^T = 0$ is needed to obtain (ii) using basic matrix operations. Assuming that $B$ is a square matrix, this additional constraint can be understood as, all the rows and columns of $B$ of which indices corresponding the nonboundary nodes are zero. □

For the FD–FE coupling, notations $H_x, H_y$ are used instead of $H^L_x, H^L_y$ since there is no $H^R_x, H^R_y$. The following definition is central to the FD–FE coupling.

**Definition 4.5.** \{IL_{22}, IR_{22}\} is called a pair of SBP-preserving FD–FE interpolation operators if

$$M_{IL_{22}} = I_{IR_{22}}^T H_y,$$

where $M_I = L_W R_M^W L_W^T, R_{M ij} = \int_{\Gamma_W} \varphi_j \varphi_i ds, i, j = 1, 2, \ldots, N$.

Notice that $M_I$ is analogous to the norm $H^R_y$ in the FD–FD case. Despite it not being diagonal, $M_I$ has nice properties including being symmetric and positive definite. For example, if the node distribution on $\Gamma_W$ is equidistant (with step length $h_y$), and $P^1$ is used, then $M_I$ is a banded matrix,

$$M_I = h_y \begin{bmatrix}
1/3 & 1/6 & 0 & 0 & \ldots & 0 \\
1/6 & 2/3 & 1/6 & 0 & \ldots & 0 \\
0 & 1/6 & 2/3 & 1/6 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \ldots & 0 & 1/6 & 2/3 & 1/6 \\
0 & \ldots & 0 & 0 & 1/6 & 1/3
\end{bmatrix}.$$

We will prove that the following FD–FE coupling SAT terms impose stability to (22),

$$\begin{align*}
SAT_{IL} &= \alpha_L (H^{-1}_x e_n) \otimes (u_I - IR_{22} L_W v) \\
&\quad + \delta_L (H^{-1}_x e_n) \otimes (D_x u_I - IR_{22} D^W_x v) + \sigma_L (H^{-1}_x d_n^T) \otimes (u_I - IR_{22} L_W v), \\
SAT_{IR} &= \alpha_R M^{-1} R_M^W L_W^T (L_W v - IL_{22} u_I) \\
&\quad + \delta_R M^{-1} R_M^W L_W^T (D_x v - IL_{22} (D_x u_I)) + \sigma_R M^{-1} (-R^W_C) L_W^T (L_W v - IL_{22} u_I).
\end{align*}$$

where $R^W_{C ij} = \int_{\Gamma_W} n^T \nabla \varphi_j \varphi_i ds, D^W_x = -M_I^{-1} L_W R^W_C$.

The assembly of the above matrices is described in [8], except the Robin convection matrix $R^W_C$ which is hence described in Appendix B. Note that the minus signs coming with $R^W_C$ in $D^W_x$ and the third component of SAT$_{IR}$ are in order to make $R^W_C$ analogous to $(d_1 \otimes I_m)$ since $n_W$ is a negative vector.

Strict stability for the FD–FE coupling is shown in Theorem 4.6.

**Theorem 4.6.** By choosing $\alpha_L = \frac{1}{2} a_1, \alpha_R = -\frac{1}{2} a_1, \delta_L = -\sigma_R, \sigma_L = -\delta_R$ and $\delta_L + \sigma_L = -\varepsilon$, assuming proper outer boundary treatment, the approximation (22) with the interface SAT terms (26) is stable.
Proof. A total energy estimate can be obtained by applying the energy method to both schemes
\[
\frac{d}{dt} \|u\|_H^2 + 2\varepsilon u^T (M \otimes H_y) u + 2\varepsilon u^T (H_x \otimes M) u + \frac{d}{dt} \|v\|^2_M + 2\varepsilon v^T A v = BT + IT,
\]
where the energy estimate at the interface is
\[
IT = \varepsilon u^T ((e_n d_n + (e_n d_n)^T \otimes H_y) u - a_1 u^T (e_n e_n^T \otimes H_y) u + a_1 v^T (R_M^W) + (R_M^W)^T) v
+ a_L [2u^T (e_n e_n^T \otimes H_y) u - u^T (e_n \otimes H_y I_{R2L} L_W) v - v^T (e_n^T \otimes (H_y I_{R2L} L_W)^T) u]
+ 2\alpha v^T R_M^W v - \alpha_R v^T R_M^W L_W I_{L2R} (e_n^T \otimes I_m) u - \alpha_R u^T (e_n \otimes I_m) I_{L2R}^T L_W R_M^W v
+ \delta_L [u^T (e_n d_n + (e_n d_n)^T \otimes H_y) u - u^T (e_n \otimes H_y I_{R2L} D_x^W) v - v^T (e_n^T \otimes (H_y I_{R2L} D_x^W)^T) u]
+ \alpha_R [u^T (e_n d_n + (e_n d_n)^T \otimes H_y) u - u^T (d_n^T \otimes H_y I_{R2L} D_x^W) v - v^T (d_n^T \otimes I_m I_{L2R} D_x^W) u]
+ \delta_R [v^T (R_M^W + (R_M^W)^T) v - v^T R_M^W L_W I_{L2R} (d_n \otimes I_m) u - u^T (d_n^T \otimes I_m I_{L2R} D_x^W) v
+ \sigma_R v^T (R_M^W)^T L_W I_{L2R} (e_n^T \otimes I_m) u
+ \sigma_R u^T (e_n \otimes I_m) I_{L2R}^T L_W R_M^W v].
\]
The nice property \( R_M^W L_W L_W^T v = R_M^W v \) no longer holds true with \( R_M^W \) replaced by \( R_M^C \) since the construction of \( R_M^C \) consists of nonzero local contributions from nodes not lying on \( \Gamma_W \). Though, from Property 4.4, it is useful that
\[
L_W^T L_W R_C^W v = R_C^W v \quad \text{and thus,} \quad v^T (R_C^W)^T L_W^T L_W = v^T (R_C^W)^T.
\]
It can also be shown that
\[
H_y I_{R2L} D_x^W = -I_{L2R}^T L_W R_C^W.
\]
Indeed, from the conservation property (25), we have
\[
H_y I_{R2L} = I_{L2R}^T M_I
\Rightarrow H_y I_{R2L} (-M_I^{-1} L_W R_C^W) = -I_{L2R}^T L_W R_C^W.
\]
The equalities (25), (27), and (28) can be applied on \( IT \) to verify that the interface terms vanish (i.e., \( IT \equiv 0 \)) with the above choice of parameters. In the case \( a_1 = 0 \), the interface terms weakly forcing the requirement \( u = v \) all become zero. We also obtain \( IT \equiv 0 \) in that case with the same parameters. The interface treatment is thus stable. \( \square \)

Remark 4.1. Similarly to the FD–FD coupling [12], the use of SBP-preserving interpolation operators used for FD–FE can also be shown to preserve the full order of accuracy. It is known that when narrow stencil FD operators of \( q \)-th-order are used, the accuracy at boundary closures is of \( q/2 \)-th-order resulting in \((q/2 + 1)\)-th-order of overall accuracy. However, since the solution space \( \mathbb{P}^1 \) is currently used for FE, we expect that with roughly equal grid resolutions and roughly equal amount of simulation by each method, the total of accuracy is of second-order although if higher order FD operators are used. This statement will be verified numerically in the next section.

5 Numerical results

Note: Within the limit of the time, the author could not complete a general procedure (which solves an optimization problem) to construct the interpolation operators satisfying the SBP-preserving property (23). The interpolation operators used for the following numerical simulations are: identity matrices for the interface matching case, and the second-order FD–FD ratio 1:2 SBP-preserving operators for the nonmatching case. However, the obtained results are still desirable since the violation of (23) does not seem to result in significant issues for the convergence rate.
5.1 Eigenvalue analysis for validation of the interface treatment

In this subsection, we show a numerical validation of Theorem 4.6. Consider the model problem (20) subject to the boundary condition (2). Second-order FD operators are used on $\Omega_L = [-1, 1] \times [-2, 0]$, and a FE scheme using piecewise linear basis functions ($P^1$) is used on $\Omega_R = [-1, 1] \times [0, 2]$. The FD domain is discretized with $41^2$ grid points ($m = n = 41$). The discretization of FE scheme uses 2796 vertices for an unstructured mesh using Delaunay triangulation with maximum cell diameter $h_{\text{max}} = 0.025$ and interface nodes matched to the FD grid (intended for an equally distributed image of eigenvalues). The problem settings are $a = (1, 0)^T$, $\varepsilon = 0.01$.

The semi-discrete problem can be written in the form

\[
\begin{bmatrix}
  u \\
  v
\end{bmatrix}_t = F \times \begin{bmatrix}
  u \\
  v
\end{bmatrix},
\]

where $F$ is some square real-valued matrix. In Figure 4, the eigenvalues of $F$ are plotted for several choices of the penalty parameters imposing the interface conditions in Theorem 4.6. Figure 4(a) and 4(c) are with two different sets of penalty parameters both satisfying Theorem 4.6 while the parameters used in Figure 4(b) and 4(d) violate the theorem. It can be clearly seen that in the former cases, all the yielding eigenvalues have nonpositive real part. Therefore, with some explicit method to integrate in time (e.g., Runge-Kutta 4) and a sufficiently small time step, the fully discrete scheme is strictly stable.
(a) With correct parameters $\alpha_L = 0.5, \alpha_R = 0.5, \delta_L = -0.5, \delta_R = 0.5, \sigma_L = -\varepsilon, \sigma_R = -\varepsilon, \sigma_L = \delta_L = 0$

(b) Without coupling terms $\alpha_{L,R} = \delta_{L,R} = \sigma_{L,R} = 0$. There exist positive eigenvalues.

(c) With correct parameters $\alpha_L = 0.5, \alpha_R = 0.5, \delta_L = \sigma_L = -\varepsilon/2, \delta_R = \sigma_R = \varepsilon/2$

(d) With incorrect parameters $\alpha_L = -0.05, \delta_L = 0.001, \alpha_R = -0.05, \sigma_R = 0.001, \sigma_L = \delta_R = 0$

Figure 4: Eigenvalues of the system (22) coupling second-order FD – $\mathbb{P}^1$ FE with parameters satisfying Theorem 4.6 (on the left column) and violating the theorem (on the right column). All are with proper outer boundary treatment.

5.2 Convergence study

We use the following two dimensional Gaussian as the reference solution

$$u(x, y) = \frac{r}{4\pi(t - t_0)\varepsilon} \exp \left[ \frac{-1}{t - t_0} \left( \frac{(x - x_0 - a_1 t)^2 + (y - y_0 - a_2 t)^2}{4\varepsilon} \right) \right].$$

The solution satisfies the boundary condition (2) with zero boundary data $g = 0$ if the initial solution is small enough at the inflow boundary. For convergence results in this section, we choose $r = 0.005$. The problem settings are: $\varepsilon = 0.01$; the initial position and the advection direction is

$$\begin{cases} (x_0, y_0)^T = (-1, 0)^T \\ a = (1, 0)^T \end{cases}$$

when transporting the solution from the FD domain to the FE domain, and

$$\begin{cases} (x_0, y_0)^T = (1, 0)^T \\ a = (-1, 0)^T \end{cases}$$
when transporting it in the opposite direction. With these settings, the solution is smaller than $10^{-30}$ at $\Gamma_{in}$.

Figure 5 shows an example when using $m_L = n_L = 21$ for the FD domain, and an unstructured mesh using Delaunay triangulation $h_{max} = 0.05$ for the FE domain (nonmatching interface). The solution is initialized at $(x_0, y_0)^T = (0, 0)^T$ and transported from left to right $a = (1, 0)^T$. The diffusion parameter is $\varepsilon = 0.01$. Fourth-order FD operators are used. The solution is examined at $t = 0$ and $t = 0.3$.

![Figure 5](image)

Figure 5: An example of FD–FE nonconforming multiblock coupling for the advection-diffusion problem.

For the convergence results, we use a more uniform mesh type for the FE discretization (regular mesh), illustrated in Figure 6. This is for two reasons: (i) for shorter computation time (fewer nodes than the Delaunay triangulation); (ii) for a ‘roughly’ equal refinement level between the two methods when matching interface is used.
The error is calculated by $L_2$-norm,

$$l_2(u) = \sqrt{(u - u_e)^T H (u - u_e)}, \quad l_2(v) = \sqrt{(v - v_e)^T M (v - v_e)},$$  \hspace{0.5cm} (30)

where $u_e, v_e$ are the exact solutions at evaluation time point. The convergence rate is calculated as

$$Q = \log_{10} \left( \frac{\log_{10}(l_2(u) + l_2(v))^{(m_1)}}{\log_{10}(l_2(u) + l_2(v))^{(m_2)}} \right) / \log_{10} \left( \frac{m_2}{m_1} \right).$$  \hspace{0.5cm} (31)

We use the FD diagonal norm SBP operators formulated in [13].

Tables 1, 2, 3, 4 show the $L_2$-error and convergence rates for the problem (1) subject to the boundary condition (2) with different numbers of grid points. (39) is integrated in time using the standard Runge-Kutta 4 with time step $\Delta t = \beta h^2, \beta = 0.1$. The solution is examined at $t = 2$. The convergence rate is denoted by $Q$ and calculated by (31).

In Tables 1, 2, $m$ is equal to both grid size in the FD domain ($m = n_L = m_L$), and the number of discretization points along each axis for the regular mesh (matching interface). In both tables, we can observe the expected convergence rate of 2. The use of higher-order FD operators considerably lowers the error and improves the convergence rate.

![Figure 6: The computational mesh for convergence study](image)

<table>
<thead>
<tr>
<th>$m$</th>
<th>$2^{nd}$ order FD to $P^1$ FE</th>
<th></th>
<th>$4^{th}$ order FD to $P^1$ FE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\log_{10}(l_2(u))$</td>
<td>$\log_{10}(l_2(v))$</td>
<td>$\log_{10}(l_2(u) + l_2(v))$</td>
</tr>
<tr>
<td>31</td>
<td>-4.16</td>
<td>-2.29</td>
<td>-2.29</td>
</tr>
<tr>
<td>41</td>
<td>-5.46</td>
<td>-2.56</td>
<td>-2.56</td>
</tr>
<tr>
<td>51</td>
<td>-5.27</td>
<td>-2.76</td>
<td>-2.76</td>
</tr>
<tr>
<td>61</td>
<td>-5.32</td>
<td>-2.92</td>
<td>-2.92</td>
</tr>
</tbody>
</table>

Table 1: Convergence rates when transporting the solution from the FD domain to the FE domain (matching interface)

<table>
<thead>
<tr>
<th>$m$</th>
<th>$P^1$ FE to $2^{nd}$ order FD</th>
<th></th>
<th>$P^1$ FE to $4^{th}$ order FD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\log_{10}(l_2(u))$</td>
<td>$\log_{10}(l_2(v))$</td>
<td>$\log_{10}(l_2(u) + l_2(v))$</td>
</tr>
<tr>
<td>31</td>
<td>-2.32</td>
<td>-5.92</td>
<td>-2.32</td>
</tr>
<tr>
<td>41</td>
<td>-2.56</td>
<td>-6.08</td>
<td>-2.56</td>
</tr>
<tr>
<td>51</td>
<td>-2.75</td>
<td>-6.23</td>
<td>-2.75</td>
</tr>
</tbody>
</table>

Table 2: Convergence rates when transporting the solution from the FE domain to the FD domain (matching interface)
In Tables 3, 4, $m_L$ is equal to grid size in FD domain ($m_L = n_L$), and $m_R$ is equal to the number of discretization points along each axis for the regular mesh ($m_L \neq m_R$, non-matching interface).

<table>
<thead>
<tr>
<th>$m_L$</th>
<th>$m_R$</th>
<th>$2^{nd}$ order FD to $P^1$ FE</th>
<th>$4^{th}$ order FD to $P^1$ FE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\log_{10}(l_2(u))$</td>
<td>$\log_{10}(l_2(v))$</td>
</tr>
<tr>
<td>21</td>
<td>41</td>
<td>-2.69</td>
<td>-1.92</td>
</tr>
<tr>
<td>31</td>
<td>61</td>
<td>-4.14</td>
<td>-2.29</td>
</tr>
<tr>
<td>41</td>
<td>81</td>
<td>-5.45</td>
<td>-2.56</td>
</tr>
</tbody>
</table>

Table 3: Convergence rates when transporting the solution from the FD domain to the FE domain (nonmatching interface)

<table>
<thead>
<tr>
<th>$m_L$</th>
<th>$m_R$</th>
<th>$P^1$ FE to $2^{nd}$ order FD</th>
<th>$P^1$ FE to $4^{th}$ order FD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\log_{10}(l_2(u))$</td>
<td>$\log_{10}(l_2(v))$</td>
</tr>
<tr>
<td>21</td>
<td>41</td>
<td>-2.00</td>
<td>-4.43</td>
</tr>
<tr>
<td>31</td>
<td>61</td>
<td>-2.32</td>
<td>-6.23</td>
</tr>
<tr>
<td>41</td>
<td>81</td>
<td>-2.56</td>
<td>-6.49</td>
</tr>
</tbody>
</table>

Table 4: Convergence rates when transporting the solution from the FE domain to the FD domain (nonmatching interface)

The obtained rates when using more coarse grid and higher order operators for FD domain are much higher than 2. It can be explained that since the FE mesh is much finer, the error contribution from the FE scheme does not significantly influence the convergence rate. Thus, for a sufficiently high level of refinement (keeping the same ratio 1:2), the rates will eventually become 2.

We also show some experiments for the inviscid problem when the diffusion term is set to zero ($\varepsilon = 0$) in Tables 5, 6. In that case, the following 2D Gaussian is used as the reference solution

$$u(x,y) = r_1 \exp \left( \frac{(x-x_0-a_1 t)^2 + (y-y_0-a_2 t)^2}{r_2^2} \right).$$

For the following experiments, we choose $r_1 = 1$, $r_2 = 0.2$. $m$ in Tables 5, 6 holds the same meaning as above (matching interface). We explain the obtained convergence results in Remark 5.1.

<table>
<thead>
<tr>
<th>$m$</th>
<th>$2^{nd}$ order FD to $P^1$ FE</th>
<th>$4^{th}$ order FD to $P^1$ FE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\log_{10}(l_2(u))$</td>
<td>$\log_{10}(l_2(v))$</td>
</tr>
<tr>
<td>31</td>
<td>-1.72</td>
<td>-1.10</td>
</tr>
<tr>
<td>41</td>
<td>-2.05</td>
<td>-1.30</td>
</tr>
<tr>
<td>51</td>
<td>-2.26</td>
<td>-1.48</td>
</tr>
</tbody>
</table>

Table 5: Convergence rates when transporting the solution from the FD domain to the FE domain (inviscid case $\varepsilon = 0$)
Table 6: Convergence rates when transporting the solution from the FE domain to the FD domain (inviscid case $\varepsilon = 0$)

A three-dimensional illustration is shown in Figure 7.

Remark 5.1. The convergence rates in fourth order FD cases in Tables 5 and 6 are noticeably close to a fourth-order–fourth-order coupling (cf., FD–FD fourth order case [12]). Here we incidentally
observe a phenomenon in finite element approximations so-called superconvergence [2] where the
order of convergence at some special points is higher than the convergence rate of the total error
if calculated analytically. Those points are called superconvergence points and they happen to
be the nodal points of the regular mesh (Figure 6) with the advection operator $a^T \nabla v$. For the
viscous case, there also exist other certain points which give rise to this behaviour. It turns out
that (30) is not the most correct way to assess the discrete error in the sense of mimicking the
continuous measurement $\|e\|_\Omega = \sqrt{\int_\Omega e^2 dx dy}$. A straightforward fix to avoid this problem is
to calculate the error with many more sample points added to the computation mesh. From a
different angle, it cannot be denied that (30) is desirable once our FE approximation fits in the
SBP framework. (Simply the defined discrete norm was applied.) To that extent, interpreting the
advection operator as fourth order and the diffusion operator as second order may explain all the
rates obtained. (The author has done separate numerical experiments on the FE approximation to
verify that the current advection operator is fourth order on the regular mesh regarding the error
assessment (30).) This is a strong evidence that the proposed coupling technique preserves native
properties of the existing schemes.

6 A nonlinear case – 2D Burger’s Equation

We consider the following nonlinear problem in two spatial dimensions

$$u_t + a^T \nabla \left( \frac{u^2}{2} \right) = \varepsilon \nabla^T \nabla u, \quad (x, y) \in \Omega \subset \mathbb{R}^2 \tag{32}$$

subject to boundary condition and initial condition

$$\frac{1}{3} u(a^T n u - |a^T n u|) - \varepsilon n^T \nabla u = g(x, y, t), \quad (x, y) \in \partial \Omega, t > 0 \tag{33}$$

$$u = u_0(x, y), \quad (x, y) \in \Omega, t = 0.$$ 

where $a > 0$ is a constant vector and $\varepsilon$ is a real constant. (32) can be written in an alternative
form

$$u_t + a_1 uu_x + a_2 uu_y = \varepsilon(u_{xx} + u_{yy}), \quad (x, y) \in \Omega \subset \mathbb{R}^2. \tag{34}$$

6.1 The continuous problem

Since $\nabla \left( \frac{u^2}{2} \right) = u a^T \nabla u$, the following skew-symmetric presentation is equivalent to (32),

$$u_t + a a^T \nabla \left( \frac{u^2}{2} \right) + (1 - \alpha) u a^T \nabla u = \varepsilon \nabla^T \nabla u.$$

Applying the energy method with the continuous integration rules (4) yields

$$(u, u_t) = \frac{\alpha}{2} (a^T \nabla u, u^2)_{\Omega} - \frac{\alpha}{2} \int_{\partial \Omega} a^T n u^3 ds + (1 - \alpha)(a^T \nabla u^2, u)_{\Omega} - (1 - \alpha) \int_{\partial \Omega} a^T n u^3 ds$$

$$+ \varepsilon (u, n^T \nabla u)_{\partial \Omega} - \varepsilon (\nabla u, \nabla u)_{\Omega}$$

$$(u_t, u) = -\frac{\alpha}{2} (a^T \nabla u^2, u)_{\Omega} - (1 - \alpha)(a^T \nabla u, u^2)_{\Omega} + \varepsilon (n^T \nabla u, u)_{\partial \Omega} - \varepsilon (\nabla u, \nabla u)_{\Omega}.$$ 

Choosing $\alpha = \frac{2}{3}$ eliminates all the interior terms

$$\frac{d}{dt} (\|u\|_\Omega^2) + 2 \varepsilon \|\nabla u\|_{\partial \Omega}^2 = \frac{2}{3} \int_{\partial \Omega} a^T n u^3 ds + 2 \varepsilon (u, n^T \nabla u)_{\partial \Omega}. \tag{35}$$
The boundary condition (33) imposes well-posedness to (32) since bounded energy is assured alongside. Indeed, inserting (33) with zero boundary data \( g = 0 \) gives

\[
\frac{d}{dt} \left( \|u\|_{\Omega}^2 \right) + 2\varepsilon \|\nabla u\|_{\Omega}^2 = -\frac{2}{3} \int_{\partial\Omega} u^2 |a^T n u| \, ds. \tag{36}
\]

### 6.2 FD schemes

A finite difference SBP-SAT semi-discretization of (32) is given by

\[
u_t = -\frac{1}{3} a_1 \nabla_x u - \frac{1}{3} a_1 \nabla_x u - \frac{1}{3} a_2 \nabla_y u - \frac{1}{3} a_2 \nabla_y u + \varepsilon \nabla x u + \varepsilon \nabla y u + \text{SAT}, \tag{37}
\]

where \( \overline{\pi} = \text{Diag}(u) \). Denote \( u^+ = \frac{u + |u|}{2} \), \( u^- = \frac{u - |u|}{2} \). It comes with several properties,

\[
u^+ + u^- = u, u + |u| = 2u^+, u - |u| = 2u^- \text{ and } u^+ - u^- = |u|. \tag{38}
\]

The boundary condition (33) is equivalent to

\[
\begin{align*}
\text{North} : & \quad \frac{2}{3} a_2 uu^- - \varepsilon u_y = g_N(t), \\
\text{East} : & \quad \frac{2}{3} a_1 uu^- - \varepsilon u_x = g_E(t), \\
\text{South} : & \quad -\frac{2}{3} a_2 uu^+ + \varepsilon u_y = g_S(t), \\
\text{West} : & \quad -\frac{2}{3} a_1 uu^+ + \varepsilon u_x = g_W(t).
\end{align*}
\]

The corresponding boundary treatment terms imposing the above boundary conditions are

\[
\begin{align*}
\text{SAT} &= \text{SAT}_N + \text{SAT}_E + \text{SAT}_S + \text{SAT}_W \\
\text{SAT}_N &= \tau_N \left( \frac{2a_2 u_N u_N^-}{3} - \varepsilon (\nabla_y u)_N - g_N \right) \otimes (H_y^{-1} e_m), \\
\text{SAT}_E &= \tau_E (H_x^{-1} e_n) \otimes \left( \frac{2a_1 u_E u_E^-}{3} - \varepsilon (\nabla_x u)_E - g_E \right), \\
\text{SAT}_S &= \tau_S \left( -\frac{2a_2 u_S u_S^+}{3} + \varepsilon (\nabla_y u)_S - g_S \right) \otimes (H_y^{-1} e_1), \\
\text{SAT}_W &= \tau_W (H_x^{-1} e_1) \otimes \left( -\frac{2a_1 u_W u_W^+}{3} + \varepsilon (\nabla_x u)_W - g_W \right).
\end{align*}
\tag{39}
\]

**Theorem 6.1.** By choosing \( \tau_N = \tau_E = \tau_S = \tau_W = 1 \), the approximation (37) with the outer boundary SAT terms given in (39) is stable.

**Proof.** See Appendix A. \( \square \)

### 6.3 FE schemes

Let \( f(u) = \left[ \frac{a_1 u^2}{2}, \frac{a_2 u^2}{2} \right]^T \). We have \( \nabla f(u) = f'(u) \cdot \nabla u \), where \( f'(u) = [a_1 u, a_2 u]^T \). With skew-symmetric splitting, (32) is equivalent to

\[
u_t + \frac{1}{3} f'(u) \cdot \nabla u + \frac{1}{3} a^T \nabla u^2 = \varepsilon \nabla^T \nabla u. \tag{40}
\]
Weak formulation

Multiplying both sides of (40) by a test function \( v \in V := \{ v : \| v \|_{L^2(\Omega)}^2 + \| \nabla v \|_{L^2(\Omega)}^2 < \infty \} \) and integrating by parts, a weak formulation can be written as: find \( u \in V \) such that

\[
(\partial_t u, v) + \frac{1}{3} (f'(u) \cdot \nabla u, v) + \frac{1}{3} (a^T \nabla u^2, v) = (\varepsilon \nabla^T \nabla v, v) \quad \forall v \in V. \tag{41}
\]

Applying the Green’s formula on (41), we obtain

\[
(\partial_t u, v) + \frac{1}{3} (f'(u) \cdot \nabla u, v) + \frac{1}{3} (a^T \nabla u^2, v) + \varepsilon (\nabla u, \nabla v) = \int_{\partial \Omega} (\varepsilon \mathbf{n}^T \nabla u) v ds, \quad \forall v \in V.
\]

Inserting the boundary condition (33) gives

\[
(\partial_t u, v) = -\frac{1}{3} (f'(u) \cdot \nabla u, v) - \frac{1}{3} (a^T \nabla u^2, v) - \varepsilon (\nabla u, \nabla v) + \frac{1}{3} \int_{\partial \Omega} u(a^T \mathbf{n} u - |a^T \mathbf{n} u|) v ds, \quad \forall v \in V.
\]

Galerkin finite element approximation

A finite element approximation can then be formulated as: find \( u_h \in V_h := \{ v \in C^0(\Omega) \mid v|_K \in P^1(K), \forall K \in T_h \} \) such that

\[
(\partial_t u_h, v) = \frac{1}{3} (f'(u_h) \cdot \nabla u_h, v) - \frac{1}{3} (a^T \nabla u_h^2, v) - \varepsilon (\nabla u_h, \nabla v) + \frac{1}{3} \int_{\partial \Omega} u_h(a^T \mathbf{n} u_h - |a^T \mathbf{n} u_h|) v ds, \quad \forall v \in V
\]

where \( T_h = \{ K_i \}_{i=1}^N \approx \Omega \), and \( P^1(K_i) \) is the space of linear interpolations on \( K_i \). Let \( V_h = \text{span}\{ \varphi_i \}_{i=1}^N \). Since \( u_h \in V_h \), there exist real numbers \( \{ \xi_i \}_{i=1}^N \) such that \( u_h = \sum_{j=1}^N \xi_j \varphi_j \). Inserting this into (42) and letting \( v = \varphi_i, i = 1, \ldots, N \) gives

\[
\mathbf{M}\dot{\xi}(t) = -\frac{1}{3} \mathbf{C}(\xi)\dot{\xi}(t) - \frac{1}{3} \mathbf{G}(\xi)\dot{\xi}(t) - \varepsilon \mathcal{A}\xi(t) + \frac{1}{3} \mathcal{R}(a^T \mathbf{n} \xi(t)) - \frac{1}{3} \mathcal{R}(|a^T \mathbf{n} \xi|)\dot{\xi}(t). \tag{43}
\]

Matrices \( \mathbf{M}, \mathbf{A} \) are the same as formulated in Section 3.2. The new matrices \( \mathbf{C}(\xi), \mathbf{G}(\xi), \mathcal{R}(\phi) \) are respectively the two nonlinear convection matrices, and the nonlinear Robin mass matrix. For \( i, j = 1, \ldots, N \),

\[
\mathbf{C}_{ij} = (f'(\xi) \cdot \nabla \varphi_j, \varphi_i),
\mathbf{G}_{ij} = (a^T \nabla (\varphi_j \xi), \varphi_i),
\mathcal{R}_{ij}(\phi) = \int_{\partial \Omega} \phi \varphi_j \varphi_i ds.
\]

\[
\mathbf{C}_{ij} = (f'(\xi) \cdot \nabla \varphi_j, \varphi_i), \mathbf{G}_{ij} = (a^T \nabla (\varphi_j \xi), \varphi_i), \mathcal{R}_{ij}(\phi) = \int_{\partial \Omega} \phi \varphi_j \varphi_i ds.
\]

The assembly of \( \mathcal{C}(\xi), \mathcal{G}(\xi), \) and \( \mathcal{R}(\phi) \) is described in Appendix B.

6.4 The interface coupling

Consider the coupling of two domains \( \Omega_L, \Omega_R \) illustrated in Figure 2, where \( u \) and \( v \) present the parts of solution on the left and right domains, respectively,

\[
u_t + \frac{2}{3} a^T \nabla \left( \frac{u^2}{2} \right) + \frac{1}{3} ua^T \nabla u = \varepsilon \nabla^T \nabla u, \quad (x, y) \in \Omega_L
\]

\[
v_t + \frac{2}{3} a^T \nabla \left( \frac{v^2}{2} \right) + \frac{1}{3} va^T \nabla v = \varepsilon \nabla^T \nabla v, \quad (x, y) \in \Omega_R \tag{44}
\]
Applying the energy method to (44) gives
\[
\frac{d}{dt} \left( \left\| u \right\|_{\Omega_L}^2 + \left\| v \right\|_{\Omega_R}^2 \right) + 2\varepsilon \left\| \nabla u \right\|_{\Omega_L}^2 + 2\varepsilon \left\| \nabla v \right\|_{\Omega_R}^2 = BT + IT,
\]
where the interface energy estimate reads
\[
IT = -\frac{2}{3} \int_{\Omega_I} a^T n_E w^3 ds + 2\varepsilon (u, n_E^T \nabla u)_{\Omega_I} - \frac{2}{3} \int_{\Omega_I} a^T n_W v^3 ds + 2\varepsilon (v, n_W^T \nabla v)_{\Omega_I}
\]
\[
= \frac{2}{3} \int_{\Omega_I} a_1 (v^3 - u^3) ds + 2\varepsilon (u, u_x)_{\Omega_I} - 2\varepsilon (v, v_x)_{\Omega_I}.
\]
Therefore, the necessary continuity condition for accurate coupling (so that IT ≡ 0) is
\[
\begin{cases}
  u = v & \text{at } x \in \Omega_I.
  \\
  u_x = v_x
\end{cases}
\] (45)

**FD–FD**

A FD–FD semi-discretization of (44) is given by
\[
uu_i = -\frac{1}{3} a_1 \pi D_x u - \frac{1}{3} u_1 D_x \pi u - \frac{1}{3} a_2 \pi D_y u - \frac{1}{3} a_2 D_y \pi u + \varepsilon D_{2x} u + \varepsilon D_{2y} u + \text{SAT}_L + \text{SAT}_L,
\] (46)
\[
v_i = -\frac{1}{3} a_1 \pi D_x v - \frac{1}{3} u_1 D_x \pi v - \frac{1}{3} a_2 D_y v - \frac{1}{3} a_2 D_y \pi v + \varepsilon D_{2x} v + \varepsilon D_{2y} v + \text{SAT}_R + \text{SAT}_R.
\]
The interface SAT terms can be written as
\[
\text{SAT}_L = \alpha_L (H_x^{L-1} e_n) \otimes ((\pi u)_I - I_{R2L} (\pi v)_I) + \beta_L (H_x^{L-1} e_n) \otimes [\pi_I (u_I - I_{R2L} v_I)]
\]
\[
+ \delta_L (H_x^{L-1} e_n) \otimes ((D_x u)_I - I_{R2L} (D_x v)_I) + \sigma_L (H_x^{L-1} d_n^T) \otimes (u_I - I_{R2L} v_I),
\] (47)
\[
\text{SAT}_R = \alpha_R (H_x^{R-1} e_1) \otimes ((\pi v)_I - I_{L2R} (\pi u)_I) + \beta_R (H_x^{R-1} e_1) \otimes [\pi_I (v_I - I_{L2R} u_I)]
\]
\[
+ \delta_R (H_x^{R-1} e_1) \otimes ((D_x v)_I - I_{L2R} (D_x u)_I) + \sigma_R (H_x^{R-1} d_n^T) \otimes (v_I - I_{L2R} u_I),
\]
where \((\pi u)_I = (e_n^T \otimes I_n) \pi u, \pi_I = \text{Diag}(u_I), \pi_I = \text{Diag}(v_I). We have a direct consequence,
\[
(\pi u)_I \equiv \pi_I u_I
\] (48)
which is a diagonal matrix of which diagonal is the elementwise self-product of \(u_I\); and from properties of block matrix multiplication in linear algebra, one can show that
\[
(e_n^T \otimes E) \pi = (e_n^T \otimes E) (I_n \otimes \pi_I),
\]
and \((e_n^T \otimes E) \pi = (e_n^T \otimes E) (I_n \otimes \pi_I),
\] (49)
where \(E\) is an arbitrary matrix of compatible dimensions.

**Theorem 6.2.** By choosing \(\delta_L = -\sigma_R, \sigma_L = -\delta_R, \delta_L + \sigma_L = -\varepsilon, \alpha_L = -\beta_R, \beta_L = -\alpha_R\) and \(\alpha_L + \beta_L = \frac{a}{3}\) assuming proper outer boundary treatment, the approximation (46) with the interface SAT terms (47) is stable.

**Proof.** Applying the energy method on each equation of (46) and adding them up, a total energy estimate of \(u, v\) can be written as
\[
\frac{d}{dt} \left\| u \right\|_{\Omega}^2 + 2\varepsilon u^T (M \otimes H_y^L) u + 2\varepsilon u^T (H_y^L \otimes M) u
\]
\[
+ \frac{d}{dt} \left\| v \right\|_{\Omega}^2 + 2\varepsilon v^T (M \otimes H_y^R) v + 2\varepsilon v^T (H_y^R \otimes M) v = BT + IT,
\]
where the interface terms are
\[
IT = -\frac{2}{3}a_1 \nu^T (e_n e_n^T \otimes H_y^L) \nu + \frac{2}{3}a_1 \nu^T (-e_1 e_1^T \otimes H_y^R) \nu
\]
\[
+ \varepsilon^T (e_n d_n + (e_n d_n)^T \otimes H_y^L) u + \varepsilon v^T (-e_1 d_1 - (e_1 d_1)^T \otimes H_y^R) v
\]
\[
+ 2a_L u^T (e_n e_n^T \otimes H_y^L) \nu - 2a_L u^T (e_n e_n^T \otimes H_y^L I_{L2R}) \nu
\]
\[
+ 2\alpha R u^T (e_1 e_1^T \otimes H_y^R) \nu - 2\alpha R u^T (e_1 e_1^T \otimes H_y^R I_{L2R}) \nu
\]
\[
+ 2\beta_L u^T (e_n e_n^T \otimes H_y^L) \nu - 2\beta_L v^T (e_n e_n^T \otimes I_{L2R} H_y^L) \nu
\]
\[
+ 2\beta_R u^T (e_1 e_1^T \otimes H_y^R) \nu - 2\beta_R v^T (e_1 e_1^T \otimes I_{L2R} H_y^R) \nu
\]
\[
+ \delta_L [u^T (e_n d_n + (e_n d_n)^T \otimes H_y^L) u - u^T (e_1 d_1 + (e_1 d_1)^T \otimes H_y^R I_{L2R}) u]
\]
\[
+ \sigma_L [u^T (e_n d_n + (e_n d_n)^T \otimes H_y^L) u - u^T (e_1 d_1 + (e_1 d_1)^T \otimes H_y^R I_{L2R}) u]
\]
\[
+ \delta_R [v^T (e_1 d_1 + (e_1 d_1)^T \otimes H_y^R) v - v^T (e_1 d_1 + (e_1 d_1)^T \otimes H_y^R I_{L2R}) u]
\]
\[
+ \sigma_R [v^T (e_1 d_1 + (e_1 d_1)^T \otimes H_y^R) v - v^T (e_1 d_1 + (e_1 d_1)^T \otimes H_y^R I_{L2R}) u].
\]

The relations (48), (49) and following observations are necessary for the deduction of the above equality
\[
u^T \Pi^T [(H_x^{L^{-1}} e_n) \otimes (\nu u)_I] = [(H_x^{L^{-1}} e_n) \otimes (\nu u)_I]^T \Pi^L u;
\]
\[
u^T \Pi^R [(H_x^{R^{-1}} e_1) \otimes (\nu^T I_{L2R} u)_I] = [(H_x^{R^{-1}} e_1) \otimes (\nu^T I_{L2R} u)_I]^T \Pi^R u = u^T (e_n e_n^T \otimes I_{L2R} H_y^R) \nu.
\]

For the proof of (50), the right hand side is equal to \((\nu u)^T (e_n e_n^T \otimes H_y^L) u = u^T \Pi (e_n e_n^T \otimes H_y^L) u\).

Since the whole product is only a scalar which is equal to its transpose, taking transpose leads to the desired conclusion. A similar reasoning is sufficient for the first equality of (51). Because \([H_x^{L^{-1}} e_1] \otimes (\nu^T I_{L2R} u)_I]^T \Pi^R v = u^T (e_n e_n^T \otimes I_{L2R} \nu^T H_y^R) v\) and \(\nu^T H_y^R u = H_y^R \nu^T \Pi u\) hold true for diagonal norm FD operators, applying (49) yields the second equality of (51). It can be verified that with the conservation property \(H_y^{R T} I_{L2R} = I_{L2R}^T H_y^L\) and the above choice of parameters, the interface terms vanish (i.e., \(IT = 0\)). The interface treatment is thus stable. \(\square\)

**FD–FE**

A FD–FE semi-discretization of (44) is given by
\[
u_t = -\frac{1}{3} a_1 \Pi D_x \nu - \frac{1}{3} a_1 D_x \nu - \frac{1}{3} a_2 \Pi D_y \nu - \frac{1}{3} a_2 D_y \nu + \varepsilon D_x \nu + \varepsilon D_y \nu + SAT_L + SAT_{IL}
\]
\[
u_t = -\frac{1}{3} M^{-1} \mathcal{C}(\nu) v - \frac{1}{3} M^{-1} \mathcal{G}(\nu) v - \varepsilon M^{-1} A(\nu) v + \frac{\alpha_n}{3} M^{-1} [\mathcal{R}^N(\nu)] v
\]
\[
+ \frac{\alpha_1}{3} M^{-1} [\mathcal{R}^E(\nu) - \mathcal{R}^E(|\nu|)] v + \frac{\alpha_2}{3} M^{-1} [-\mathcal{R}^S(\nu) - \mathcal{R}^S(|\nu|)] v + SAT_{IR},
\]

where \(\mathcal{R}^N, \mathcal{R}^E, \mathcal{R}^S\) are parts of the nonlinear Robin mass matrix \(\mathcal{R}\) formulated in Section 6.3 regarding the contributions along \(\Gamma_N, \Gamma_E, \Gamma_S\), respectively.

The penalty terms imposing the interface continuity condition (45) are given by
\[
SAT_{IL} = \alpha_L (H_x^{L^{-1}} e_n) \otimes ((\nu u)_I - I_{R2L} L_W \nu) + \beta_L (H_x^{L^{-1}} e_n) \otimes [\nu^T (u_I - I_{R2L} L_W v)]
\]
\[
+ \delta_L (H_x^{L^{-1}} e_n) \otimes ((D_x u)_I - I_{R2L} D_y^T v) + \sigma_L (H_x^{L^{-1}} d_1) \otimes (u_I - I_{R2L} L_W v),
\]
\[
SAT_{IR} = \alpha_R M^{-1} R^W_M (\nu - L_W^T I_{L2R} \nu) + \beta_R M^{-1} (\nu - L_W^T I_{L2R} \nu) + \delta_R M^{-1} R^W_M^T (D_x^T v - I_{L2R} (D_x u)_I) + \sigma_R M^{-1} (-R^W_M^T) L_W^T (L_W v - I_{L2R} u_I).
\]

27
From (43), a semi-discrete energy estimate for the solution part on the FE domain with proper outer boundary treatment and without interface treatment \((\alpha_L = \beta_L = \delta_L = \sigma_L = 0)\) reads
\[
\frac{d}{dt} \|v\|^2_M + 2\varepsilon v^T A v = BT_{N.E.S} + \frac{2\alpha_1}{3} v^T B^W(v)v + \varepsilon v^T (R_C^W + (R_C^W)^T)v,
\]
where \(B^W_{ij}(\cdot) = \int_{\Gamma_w} \phi \varphi_j \varphi_i ds\).

The following property of \(B^W\) negates the requirement of being diagonal for \(M_i\).

**Lemma 6.3.** Let \(\phi \in V\) be some smooth function and \(\Phi\) its projection on \(V_h\). We have
\[
B^W(\phi) = \frac{1}{2}(\Phi R_M^W + R_M^W \Phi),
\]
where \(\Phi = \text{Diag}(\Phi)\).

**Proof.** We have \(B^W_{ij}(\cdot) = \int_{\Gamma_w} \phi \varphi_j \varphi_i ds\). A local edge contribution to \(B^W(\cdot)\) is
\[
\int_{E \in \Gamma_w} \phi \varphi_j \varphi_i ds \approx \Phi \left( \frac{x_i + x_j}{2} \right) \int_E \varphi_j \varphi_i ds = \frac{1}{2} \Phi(x_i) \int_E \varphi_j \varphi_i ds + \frac{1}{2} \Phi(x_j) \int_E \varphi_j \varphi_i ds
\]
since \(\Phi \in \mathbb{V}_h\). Thus, \(B^W(\phi) = \frac{1}{2}(\Phi R_M^W + (\Phi R_M^W)^T)\). This leads to the desired conclusion because \(\Phi\) and \(M_i^W\) are both symmetric. 

The following theorem completes the FD–FE coupling for the concerning nonlinear problem.

**Theorem 6.4.** By choosing \(\delta_L = -\sigma, \sigma_L = -\delta_R, \sigma_L = -\varepsilon, \alpha_L = -\beta_R, \beta_L = -\alpha_R\) and \(\alpha_L + \beta_L = \frac{a_1}{3}\) assuming proper outer boundary treatment, the approximation (52) with the interface SAT terms (53) is stable.

**Proof.** A total energy estimate can be obtained by applying the energy method on both schemes in (52)
\[
\frac{d}{dt} \|u\|^2_M + 2\varepsilon u^T (M \otimes H_y) u + 2\varepsilon u^T (H_x \otimes M) u + \frac{d}{dt} \|v\|^2_M + 2\varepsilon v^T A v = BT + IT,
\]
where the interface energy is
\[
IT = -\frac{2}{3} \alpha_1 u^T (e_n^T \otimes H_y) u + \varepsilon u^T (e_n^T + (e_n d_n)^T \otimes H_y) u + \frac{2\alpha_1}{3} v^T B^W(v)v + \varepsilon v^T (R_C^W + (R_C^W)^T)v
\]

\[
+ 2\alpha_L u^T (e_n^T \otimes H_y) u - 2\alpha_L u^T (e_n \otimes H_y I_{R2L} L_W) v + 2\alpha_L u^T (e_n^T \otimes H_y) u - 2\alpha_L u^T (e_n \otimes H_y I_{R2L} L_W) v
\]

\[
+ \alpha_R u^T (\tau R_M^W + R_M^W) v - 2\alpha_R u^T (\tau R_M^W + R_M^W) v - 2\beta_L u^T (e_n^T \otimes H_y) u + 2\beta_L u^T (e_n^T \otimes H_y) u
\]

\[
+ \beta_R u^T (\tau R_M^W + R_M^W) v - 2\beta_R u^T (\tau R_M^W + R_M^W) v - 2\delta_L [u^T (e_n d_n) (e_n d_n) \otimes H_y) u - u^T (e_n \otimes H_y I_{R2L} D_x^W) v - v^T (e_n^T \otimes (H_y I_{R2L} D_x^W)^T) u]
\]

\[
+ \sigma_L [u^T (e_n d_n) (e_n d_n) \otimes H_y) u - u^T (d_n \otimes H_y I_{R2L} L_W) v - v^T (d_n \otimes H_y I_{R2L} L_W) v]
\]

\[
+ \delta_R [u^T (R_C^W + (R_C^W)^T) v - v^T R_M^W L_W I_{R2L}(d_n \otimes I_m) u - u^T (d_n \otimes I_m) I_{L2R}(d_n \otimes I_m) v]
\]

\[
- \sigma_R u^T ((R_C^W)^T L_W L_W + L_W^T L_W R_M^W) v + \sigma_R u^T (R_C^W)^T L_W I_{L2R}(e_n \otimes I_m) u + \sigma_R u^T (\tau e_n \otimes I_m) I_{L2R}(e_n \otimes I_m) u
\]

We skip the proof of \(\delta_L\) and \(\sigma_R\) since it is the same as in Theorem 4.6. Under Lemma 6.3 and the conservation property given in Definition 4.5, one can verify that the interface terms vanish (i.e., \(IT \equiv 0\)) with the above choice of parameters. The interface treatment is thus stable. 

\[\square\]
6.5 Numerical results

We do some convergence study in this section. It can be verified that an analytic solution of (32) subject to the boundary condition (33) is

\[ u(x, y, t) = c - b \tanh \left( b \frac{x - x_0 - ct}{2\varepsilon} \right), \quad t > 0 \]

where \( b \) and \( c \) are arbitrary constants. In the following experiments, we choose \( b = 1, c = 2 \). The semi-discrete scheme (52) with the outer boundary parameters in Theorem 6.1 and the coupling parameters in Theorem 6.4 is integrated in time using the standard Runge-Kutta 4. The time step size is chosen to be \( \Delta t = \beta h^2, \beta = 0.1 \). The problem settings are: \( a = (1, 0)^T, \varepsilon = 0.1, x_0 = -1 \). The solution is examined at \( t = 1 \).

A simulation is shown in Figure 8. A coarse grid \( m_L = n_L = 21 \) is used for the FD domain, and a finer regular mesh with \( h_{\text{max}} = 0.05 \) is used for the FE domain (nonmatching interface). Fourth-order FD operators are used.
A convergence study is shown in Tables 7 and 8. The notations hold the same meaning as defined in Section 5.2. In both tables, we can observe the expected convergence rate (second order). It can also be seen that the use of higher-order FD operators (fourth-order) significantly lowers the error and improves the convergence rate.

<table>
<thead>
<tr>
<th>$m$</th>
<th>Second-order FD</th>
<th>Fourth-order FD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\log_{10}(l_2(u))$</td>
<td>$\log_{10}(l_2(v))$</td>
</tr>
<tr>
<td>21</td>
<td>−4.39</td>
<td>−1.91</td>
</tr>
<tr>
<td>31</td>
<td>−4.55</td>
<td>−2.25</td>
</tr>
<tr>
<td>41</td>
<td>−4.75</td>
<td>−2.49</td>
</tr>
<tr>
<td>51</td>
<td>−4.92</td>
<td>−2.68</td>
</tr>
</tbody>
</table>

Table 7: Convergence rates for the case of matching interface
<table>
<thead>
<tr>
<th>$m_L$</th>
<th>$m_R$</th>
<th>Second-order FD</th>
<th>Fourth-order FD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\log_{10}(l_2(u))$</td>
<td>$\log_{10}(l_2(v))$</td>
</tr>
<tr>
<td>11</td>
<td>21</td>
<td>$-1.96$</td>
<td>$-1.38$</td>
</tr>
<tr>
<td>21</td>
<td>41</td>
<td>$-4.41$</td>
<td>$-2.15$</td>
</tr>
<tr>
<td>31</td>
<td>61</td>
<td>$-4.57$</td>
<td>$-2.48$</td>
</tr>
</tbody>
</table>

Table 8: Convergence rates for the case of nonmatching interface

7 Conclusion

The multiblock coupling of FD schemes and FE schemes in this thesis is done within the SBP framework using the SAT technique to weakly impose the continuity conditions, and the SBP-preserving interpolation operators to resolve the nonconforming node distribution. Accuracy, stability and conservation are proven for a linear hyperbolic/parabolic problem and a nonlinear problem.

Future works may concern extensions for curvilinear grids, optimal time-stepping, an option for upwind operators, or a solution to the case when block corners are unmatching. The RV-method with a high potential is a great remedy for nonsmooth solutions. It could also be possible to write a general procedure to couple two arbitrary SBP schemes without the diagonal-norm constraint.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Murtazo Nazarov and my reviewer Ken Mattsson who helped me tremendously and gave me fundamental suggestions throughout this thesis. I would like to thank Lukas Lundgren and Vidar Stiernström for the valuable discussions. I also want to thank Ludvig Lindeberg and Nam Nguyen for carefully proofreading this thesis.

This thesis is conducted during a Master’s programme fully funded by the Swedish Institute Study Scholarship (SISS) 2017–2019. Without the generous support, all of these would not have been possible.
Appendices

A The choice of parameters for outer boundary SAT terms

For the advection-diffusion equation. To shorten the proof, we consider \( a \geq 0 \). Other cases can be proved similarly. We first apply the energy method to the approximation (9) only considering the SBP terms, (some operators in (8) are concerned)

\[
\begin{align*}
\dot{u}^T \mathbf{H} u_t &= - a_1 u^T (Q \otimes H_y) u - a_2 u^T (H_x \otimes Q) u \\
&\quad + \varepsilon u^T (-M - e_1 d_1 + e_n d_n \otimes H_y) u + \varepsilon u^T (H_x \otimes -M - e_1 d_1 + e_m d_m) u \\
u^T \mathbf{H} u &= - a_1 u^T (Q^T \otimes H_y) u - a_2 u^T (H_x \otimes Q^T) u \\
&\quad + \varepsilon u^T (-M^T - (e_1 d_1)^T + (e_n d_n)^T \otimes H_y) u + \varepsilon u^T (H_x \otimes -M^T - (e_1 d_1)^T + (e_m d_m)^T) u
\end{align*}
\]

Recall a property for the discrete norm \( \frac{d}{dt} \|u\|^2_H = u^T \mathbf{H} u_t + u^T \dot{u}^T \mathbf{H} u \). For presentation, an energy estimate with only regards to the SBP terms (i.e., \( \tau_N = \tau_E = \tau_S = \tau_W = 0 \)) can be written as

\[
\begin{align*}
\frac{d}{dt} \|u\|^2_H &= 2u^T (M \otimes H_y) u + 2\varepsilon u^T (H_x \otimes M) u \\
&= - a_1 u^T (-e_1 c_1^T + e_n c_n^T \otimes H_y) u - a_2 u^T (H_x \otimes -e_1 c_1^T + e_m c_m^T) u \\
&\quad + \varepsilon u^T (-e_1 d_1 + e_n d_n - (e_1 d_1)^T + (e_n d_n)^T \otimes H_y) u \\
&\quad + \varepsilon u^T (H_x \otimes -e_1 d_1 + e_m d_m - (e_1 d_1)^T + (e_m d_m)^T) u = RT.
\end{align*}
\]

In order to obtain strict stability, the addition of the SAT terms (10) with some specific choice of penalty parameters should yield bounded energy for the above discrete estimate. With zero boundary data \( g_N = g_E = g_S = g_W = 0 \), those additional energy terms are

\[
\begin{align*}
\begin{cases}
BT_N = -\tau_N \varepsilon u^T [H_x (I_n \otimes d_m) u \otimes e_m] - \tau_N \varepsilon [u^T (I_n \otimes d_m^T) H_x \otimes e_m^T] u, \\
BT_E = -\tau_E \varepsilon u^T [e_n \otimes H_y (d_n \otimes I_m) u] - \tau_E \varepsilon [c_n^T \otimes u^T (d_n^T \otimes I_m) H_y] u, \\
BT_S = 2\tau_S u^T [H_x (-a_2 I_n \otimes e_1^T) \otimes e_1] u \\
\quad + \tau_S u^T [H_y (\varepsilon I_n \otimes d_1) u \otimes e_1] + \tau_S [u^T (\varepsilon I_n \otimes d_1^T) H_x \otimes e_1^T] u, \\
BT_W = 2\tau_W u^T [e_1 \otimes H_y (-a_1 c_1^T \otimes I_m)] u \\
\quad + \tau_W u^T [e_1 \otimes H_y (\varepsilon d_1 \otimes I_m) u] + \tau_W [e_1^T \otimes u^T (\varepsilon d_1^T \otimes I_m) H_y] u.
\end{cases}
\end{align*}
\]

Indeed, by choosing \( \tau_N = \tau_E = \tau_S = \tau_W = 1 \), the total energy estimate becomes

\[
RT + BT_N + BT_E + BT_S + BT_W = -a_1 u^T (e_1 c_1^T + e_n c_n^T \otimes H_y) u - a_2 u^T (H_x \otimes e_1 c_1^T + e_m c_m^T) u,
\]

which is less than or equal to zero in our assumption \( a_1, a_2 \geq 0 \). A strict stability is thus guaranteed. Also notice that the above estimate is analogous to the continuous energy estimate (5).

For the Burger’s equation. A same routine is here performed. Assuming \( \pi H_y = H_y \pi \) (holding true for diagonal norm FD operators), we apply the energy method to the approximation (37) only considering the SBP terms,

\[
\begin{align*}
\dot{u}^T \mathbf{H} u_t &= - \frac{1}{3} a_1 u^T \pi (Q \otimes H_y) u - \frac{1}{3} a_1 u^T (Q \otimes H_y) \pi u - \frac{1}{3} a_2 u^T \pi (H_x \otimes Q) u - \frac{1}{3} a_2 u^T (H_x \otimes Q) \pi u \\
&\quad + \varepsilon u^T (-M - e_1 d_1 + e_n d_n \otimes H_y) u + \varepsilon u^T (H_x \otimes -M - e_1 d_1 + e_m d_m) u \\
\dot{u}_i^T \mathbf{H} u &= - \frac{1}{3} a_1 u^T (Q^T \otimes H_y) \pi u - \frac{1}{3} a_1 u^T (Q^T \otimes H_y) \pi u - \frac{1}{3} a_2 u^T (H_x \otimes Q^T) \pi u - \frac{1}{3} a_2 u^T (H_x \otimes Q^T) \pi u \\
&\quad + \varepsilon u^T (-M^T - (e_1 d_1)^T + (e_n d_n)^T \otimes H_y) u + \varepsilon u^T (H_x \otimes -M^T - (e_1 d_1)^T + (e_m d_m)^T) u
\end{align*}
\]
A total estimate without boundary treatment is followed

\[
\frac{d}{dt} \|u\|^2_T + 2\varepsilon u^T(M \otimes H_y)u + 2\varepsilon u^T(H_x \otimes M)u = -\frac{1}{3} a_1 u^T(-e_1e_1^T + e_n e_n^T \otimes H_y)u - \frac{1}{3} a_1 u^T\pi(-e_1e_1^T + e_n e_n^T \otimes H_y)u
\]

\[
-\frac{1}{3} a_2 u^T(H_x \otimes -e_1e_1^T + e_m e_m^T)u - \frac{1}{3} a_2 u^T\pi(H_x \otimes -e_1e_1^T + e_m e_m^T)u + \varepsilon u^T(-e_1d_1 + e_n d_n - (e_1d_1)^T + (e_n d_n)^T \otimes H_y)u
\]

\[
+ \varepsilon u^T(H_x \otimes -e_1d_1 + e_m d_m - (e_1d_1)^T + (e_m d_m)^T)u = RT.
\]

Additional energy terms from SAT technique on outer boundaries are

\[
\begin{aligned}
BT_N &= 2\tau_N a_2 u^T(H_x \otimes e_m e_m^T)\pi u - \frac{2}{3} a_2\tau_N u^{-T}\pi(H_x \otimes e_m e_m^T)u - \tau_N \varepsilon u^T[H_x(I_n \otimes d_m)u \otimes e_n - \tau_N \varepsilon [u^T(I_n \otimes d_m^T)H_x \otimes e_n^T]u],

BT_E &= -\frac{2}{3} a_1\tau_N [u^T(e_n e_n^T \otimes H_y)\pi u] - \frac{2}{3} a_1\tau_N u^{-T}\pi(e_n e_n^T \otimes H_y)u - \tau_N \varepsilon u^T(e_n \otimes H_y(d_n \otimes I_m)u - \tau_N \varepsilon [e_n^T \otimes u^T(d_m^T \otimes I_m)H_y]u),

BT_S &= -\frac{2}{3} a_2\tau_N u^T(H_x \otimes e_1 e_1^T)\pi u - \frac{2}{3} a_2\tau_N u^{-T}\pi(H_x \otimes e_1 e_1^T)u + \tau_N [u^T(e_n \otimes d_n^T)H_x \otimes e_1^T]u,

BT_W &= -\frac{2}{3} \tau_N a_1 u^T(e_1 e_1^T \otimes H_y)\pi u - \frac{2}{3} \tau_N u^{-T}\pi(e_1 e_1^T \otimes H_y)u.
\end{aligned}
\]

By choosing \(\tau_N = \tau_E = \tau_S = \tau_W = 1\), using properties in (38), the total energy estimate becomes

\[
RT + BT_N + BT_E + BT_S + BT_W = -\frac{1}{3} a_2 u^T(H_x \otimes e_1 e_1^T + e_m e_m^T)\pi |u| - \frac{1}{3} a_2 |u|^T\pi(H_x \otimes e_1 e_1^T + e_m e_m^T)u - \frac{1}{3} a_1 u^T(e_1 e_1^T + e_n e_n^T \otimes H_y)\pi |u| - \frac{1}{3} a_1 |u|^T\pi(e_1 e_1^T + e_n e_n^T \otimes H_y)u,
\]

which is less than or equal to zero. A strict stability is thus guaranteed. Notice also that the above estimate is analogous to the continuous energy estimate (36).

**B The assembly of FE operators**

The Robin mass matrix \(R_W^W\) and the nonlinear Robin mass matrix \(\mathcal{R}^W\) can be assembled using the procedure in [8], Section 4.6.2, looping over nodes on the West boundary, setting \(\kappa = 1\) and \(\kappa = \Phi\), respectively. Other FE operators not included here can also be looked up in [8].

The Robin convection matrix. We calculate the Robin convection matrix \(R_W^W\) at the West boundary in which

\[
R_W^W_{i,j} = \int_{\Gamma_W} n^T \nabla \varphi_j \varphi_i ds, \text{for } i, j = 1, 2, \ldots, N.
\]

Inserting \(n_W = (-1, 0)^T\), a local contribution on edge \(e \in \Gamma_W\) is

\[
R_W^W_{e} = \int_e n^T \nabla \varphi_j \varphi_i ds = -\int_e \frac{\partial}{\partial x} \varphi_j \varphi_i ds.
\]
In order to make the explanation intuitive, we consider a specific case illustrated in Figure 9. Different from the assembly of the Robin mass matrices, $R^W_C$ contains nonzero local contributions from some nodes not lying on the boundary (e.g., $\int_{\Gamma_W} \frac{\partial}{\partial x} \varphi_7 \varphi_1 ds \neq 0$).

The local contributions are thus $2 \times 3$ blocks. In this example, $R^W_C$ is of the form

$$R^W_C = \begin{bmatrix} c_{11} & c_{12} & \cdots & \cdots & \cdots & \cdots \\ c_{21} & c_{22} & c_{23} & \cdots & c_{27} & \cdots & \cdots \\ \cdots & \cdots & c_{32} & c_{33} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix},$$

where the local contributions of triangles $\{1, 2, 7\}$ and $\{2, 3, 8\}$ are respectively highlighted in red and blue. Those contributions can be calculated quite simply, for example, considering $\{1, 2, 7\}$

$$c_{11} = - \int_{e_{12}} \frac{\partial}{\partial x} \varphi_1 \varphi_1 ds = -b_1 \int_{e_{12}} \frac{h_{12} - s}{h_{12}} ds = -b_1 \frac{h_{12}}{2},$$

where $e_{12}$ is the segment connecting nodes 1 and 2, $h_{12} = |y_1 - y_2|$, $b_i = \frac{\partial}{\partial x} \varphi_i$, $i = 1, 2$. Similarly, we obtain

$$c_{12} = -b_2 \frac{h_{12}}{2}; c_{17} = -b_7 \frac{h_{12}}{2}; c_{21} = -b_1 \frac{h_{12}}{2}; c_{22} = -b_2 \frac{h_{12}}{2}; c_{27} = -b_7 \frac{h_{12}}{2}.$$ 

The local matrices can now be written explicitly as

$$\begin{bmatrix} c_{11} & c_{12} & c_{17} \\ c_{21} & c_{22} & c_{27} \end{bmatrix} = -\begin{bmatrix} b_1 & b_2 & b_7 \\ b_1 & b_2 & b_7 \end{bmatrix} \frac{h}{2}.$$

We wrap up everything in the following Matlab routine.

```matlab
function RC = RobinConvectionAssembler(p,t,x_l)
    np=size(p,2);
    m=size(t,2);
    RC=zeros(np,np);
    for i=1:m
        globalIndices=t(1:3,i);
        x1=p(1,globalIndices);
        x2=p(2,globalIndices);
        if sum(x1==x_l)==2
```
\[
S = \text{polyarea}(x1,x2); \\
b = [x2(2)-x2(3);x2(3)-x2(1);x2(1)-x2(2)]/2; \\
RCK = -(x1==x_l)'*b'*\sqrt((x1(1)-x1(2))^2+(x2(1)-x2(2))^2)/2; \\
R(C(\text{globalIndices},\text{globalIndices}))=R(C(\text{globalIndices},\text{globalIndices}))+RCK; \\
\]

The input variables \( p \) and \( t \) are from the mesh generator \texttt{initmesh} in Matlab. \( x_{-1} \) is the left limit by \( x \)-direction of the FE domain (which is the \( x \)-coordinate of all the nodes on the West boundary).

The \textbf{nonlinear convection matrix} \( C \). One can assemble \( C \) from the routine described in [8], Section 10.1. However, the equivalent operator \( \mathcal{C} \) is less frequently seen. We want to calculate it explicitly where

\[
\mathcal{C}_{ij} = (a^T \nabla(\Phi_i \varphi_j), \varphi_i), \Phi \in V_h, a = (a_1,a_2)^T.
\]

At a local element \( K = \{i,j,k\} \), let us consider the product

\[
\varphi_j \Phi = \varphi_j (\Phi_i \varphi_i + \Phi_j \varphi_j + \Phi_k \varphi_k),
\]

where \( \Phi_{i,j,k} \) are the values of \( \Phi \) at nodes \( i,j,k \), respectively. The above equality holds because \( \Phi \in V_h \). Since \( \varphi_{i,j,k} \in \mathbb{P}^1(K) \), they can be presented as a plane in 2D (i.e., \( \varphi_l = a_l + b_l x + c_l y, l = i,j,k \)). The above product can then be written more precisely,

\[
\varphi_j \Phi = x[2\Phi_j a_j b_j + \Phi_i (a_j b_i + a_i b_j) + \Phi_k (a_j b_k + a_k b_j)] \\
+ y[2\Phi_j a_j b_j + \Phi_i (a_j c_i + a_i c_j) + \Phi_k (a_j c_k + a_k c_j)] \\
+ x^2[\Phi_j b_j^2 + \Phi_i b_j b_i + \Phi_k b_j b_k] \\
+ y^2[\Phi_j c_j^2 + \Phi_i c_j c_i + \Phi_k c_j c_k] \\
+ xy[2\Phi_j b_j c_j + \Phi_i (b_j c_i + b_i c_j) + \Phi_k (b_j c_k + b_k c_j)] \\
+ b_j a_j^2 + \Phi_i a_j a_i + \Phi_k a_j a_k.
\]

We obtain the components of \( \nabla(\varphi_j \Phi) \)

\[
\frac{\partial}{\partial x}(\varphi_j \Phi) = [2\Phi_j a_j b_j + \Phi_i (a_j b_i + a_i b_j) + \Phi_k (a_j b_k + a_k b_j)] \\
+ 2x[\Phi_j b_j^2 + \Phi_i b_j b_i + \Phi_k b_j b_k] \\
+ y[2\Phi_j b_j c_j + \Phi_i (b_j c_i + b_i c_j) + \Phi_k (b_j c_k + b_k c_j)],
\]

\[
\frac{\partial}{\partial y}(\varphi_j \Phi) = [2\Phi_j a_j b_j + \Phi_i (a_j c_i + a_i c_j) + \Phi_k (a_j c_k + a_k c_j)] \\
+ 2y[\Phi_j c_j^2 + \Phi_i c_j c_i + \Phi_k c_j c_k] \\
+ x[2\Phi_j b_j c_j + \Phi_i (b_j c_i + b_i c_j) + \Phi_k (b_j c_k + b_k c_j)].
\]

Grouping the terms with regards to \( \Phi_{i,j,k} \) gives

\[
\frac{\partial}{\partial x}(\varphi_j \Phi) = 2b_j \varphi_j \Phi_j + (b_i \varphi_j + b_j \varphi_i) \Phi_i + (b_k \varphi_j + b_j \varphi_k) \Phi_k.
\]

For simplicity, we consider \( a = (1,0)^T \) so that \( (a^T \nabla(\Phi \varphi_j), \varphi_l) = (\frac{\partial}{\partial x}(\varphi_j \Phi), \varphi_l), l = i,j,k. \) We have

\[
\left( \frac{\partial}{\partial x}(\varphi_j \Phi), \varphi_l \right) = (2\Phi_j b_j + \Phi_i b_i + \Phi_k b_k)(\varphi_j, \varphi_l) + \Phi_i b_j (\varphi_i, \varphi_l) + \Phi_k b_j (\varphi_k, \varphi_l), l = i,j,k.
\]

From this point, \( \mathcal{C} \) can be assembled easily using the well-known result \( (\varphi_j, \varphi_l), l = i,j,k \) from the assembly of the mass matrix. The details are described in the following Matlab routine.
function C = NonlinearConvectionAssembler(p,t,Phi)
    n=size(p,2);
    m=size(t,2);
    C=zeros(n,n);
    for i=1:m
        globalIndices=t(1:3,i);
        x1=p(1,globalIndices);
        x2=p(2,globalIndices);
        S=polyarea(x1,x2);
        PhiLocal=Phi(globalIndices);
        b=[x2(2)-x2(3); x2(3)-x2(1); x2(1)-x2(2)]/2/S;
        CK_i= [2 1 1]*(2*PhiLocal(1)*b(1)+PhiLocal(2)*b(2)+PhiLocal(3)*b(3))... 
            + [1 2 1]*PhiLocal(2)*b(1)... 
            + [1 1 2]*PhiLocal(3)*b(1);
        CK_j= [1 2 1]*(PhiLocal(1)*b(1)+2*PhiLocal(2)*b(2)+PhiLocal(3)*b(3))... 
            + [1 1 2]*PhiLocal(3)*b(2)... 
            + [2 1 1]*PhiLocal(1)*b(2);
        CK_k= [1 1 2]*(PhiLocal(1)*b(1)+PhiLocal(2)*b(2)+2*PhiLocal(3)*b(3))... 
            + [2 1 1]*PhiLocal(1)*b(3)... 
            + [1 2 1]*PhiLocal(2)*b(3);
        CK=[CK_i;CK_j;CK_k]'/12*S;
        C(globalIndices,globalIndices)=C(globalIndices,globalIndices)+CK;
    end
end

Note also that the assembly of $C$ and $C^T$ can be combined for efficiency because they are always calculated together.

References


