Interface Procedures for Finite Difference Approximations of the Advection-Diffusion Equation

Jing Gong\textsuperscript{a,}\textsuperscript{*}, Jan Nordström\textsuperscript{b}

\textsuperscript{a}Department of Information Technology, Scientific Computing Division, Uppsala University, Box 337, SE-751 05 Uppsala, Sweden
\textsuperscript{b}Department of Mathematics, Linköping University, SE-581 83 Linköping, Sweden

Abstract

We investigate several existing interface procedures for finite difference methods applied to advection-diffusion problems. The accuracy, stiffness and reflecting properties of the various interface procedures are investigated.

The analysis and numerical experiments show that there are only minor differences between the various methods once a proper parameter choice has been made.

Keywords: high order finite difference methods, numerical stability, accuracy, interface conditions, summation-by-parts, weak boundary conditions

1. Introduction

The conventional multi-block methodology for structured meshes is often, for efficiency and ease of mesh generation, used in computational physics (see [1],[2],[3],[4],[5],[6],[7]). A stable and accurate coupling at the block interfaces is therefore of utmost importance. However, there are many potential traps

\textsuperscript{*}Corresponding author.

Email address: jing.gong@gmail.com (Jing Gong)
and possibilities for failure. Instabilities introduced at the block boundaries or interfaces are often handled by adding artificial dissipation. When advection is the dominant transport process, excessive amounts can easily reduce the accuracy. The artificial interfaces will also inevitably introduce numerical reflections, and care must be taken to minimize them. Another third important aspect when constructing interface procedures is to minimize the potential additional stiffness due to a large spectral radius.

The development of numerical schemes that overcome the problems mentioned above is an ongoing challenge, especially for high order finite difference methods. Strictly stable and accurate high order finite difference methods for both hyperbolic, parabolic and incompletely parabolic problems were derived in [8], [9], [10], [11], [12], [13], [14], [15]. These methods employ so called Summation-by-Parts (SBP) operators and the Simultaneous Approximation Term (SAT) procedure for imposing boundary conditions, see [16], [8], [11], [17], [15], [18]. With well-posed boundary conditions for the continuous problem, SBP operators and the SAT procedure, it is straightforward to prove stability using the energy-method. The methods discussed above have been implemented and tested in realistic flow calculations, see [19], [20], [21], [22].

In [8], [12] various versions of the SAT method in multiple domains were presented. That work was continued in [23] where the theoretical properties of interface procedures were investigated in detail. The main focus in [23] was on the stability and formal accuracy properties of the various schemes. We continue this investigation and focus on the stiffness and reflecting properties of the different interface treatments. For clarity, we follow the path
in [23], and consider one-dimensional problems in this paper. However, the
SAT formulation can easily be extended to several space dimensions and to
complicated boundary conditions (see [12], [13], [24], [14], [19], [20], [21]).
Examples of other types of hybrid methods and approaches can be found in
[25], [26], [27], [28], [29], [30], [31].

In Section 2 we derive conditions for well-posedness of the continuous
advection-diffusion problem. Section 3 deals with the various semi-discrete
multiple domain problems. We present the formulations and give a short
theoretical overview of the existing stability theory. The size and location of
the eigenvalues for both the continuous and discrete problems are considered
in Section 4. In Section 5 we perform numerical experiments and compare
the different interface procedures. We present both one- and two-dimensional
calculations. Conclusions are drawn in Section 6.

2. The continuous problem

Consider the advection-diffusion problem in one space dimension,
\[ u_t + au_x = \varepsilon u_{xx} + F, \quad 0 \leq x \leq 1, \quad t > 0, \quad (1a) \]
\[ \alpha u(0, t) + \beta u_x(0, t) = g^L(t), \quad t \geq 0, \quad (1b) \]
\[ \gamma u(1, t) + \delta u_x(1, t) = g^R(t), \quad t \geq 0, \quad (1c) \]
\[ u(x, 0) = f(x), \quad 0 \leq x \leq 1, \quad (1d) \]
where \( a, \varepsilon > 0 \) and \( \varepsilon \ll a \). In most cases we use \( F = 0 \) and we limit ourself
to Robin boundary conditions with \( \beta, \delta \neq 0 \). The functions \( F, g^L, g^R \) and \( f \)
are the data of the problem.
Remark: When the solution can be estimated in terms of all types of data, the problem (1) is called strongly well posed, see [32] for more details.

Let the inner product for real valued functions $a, b \in L^2[0, 1]$ be defined by 
\[(a, b) = \int_0^1 ab \, dx\] and the corresponding norm by $\|a\|^2 = (a, a)$. The energy method applied to (1) with $F = 0$ yields
\[
\frac{d}{dt}(\|u\|^2) + 2\varepsilon \|u_x\|^2 = \left( a + \frac{2\alpha}{\beta} \varepsilon \right) \left( u(0, t) - \frac{\varepsilon}{\beta} a + \frac{2\alpha}{\beta} \varepsilon g^L(t) \right)^2 \\
- \left( a + \frac{2\gamma}{\delta} \varepsilon \right) \left( u(1, t) - \frac{\varepsilon}{\delta} a + \frac{2\gamma}{\delta} \varepsilon g^R(t) \right)^2 \tag{2}
\]
Hence an energy estimate is obtained if
\[
a + \frac{2\alpha}{\beta} \varepsilon < 0 \quad \text{and} \quad a + \frac{2\gamma}{\delta} \varepsilon > 0. \tag{3}
\]

Remark: With the choice (3), the last two terms in (2) are positive but bounded since they contain only boundary data.

We have proved the following proposition.

**Proposition 2.1** With condition (3) satisfied, the problem (1) is strongly well posed.

### 3. The semi-discrete problem

In this Section we give a short theoretical overview of the existing stability theory for interface procedures. Most of the material, in scattered form, can be found in [8], [12], [23], [21], [33], [34], [35] but is summarized here for completeness. Section 3.1 deals with the single domain problem and the
general SBP-SAT theory while Section 3.2 deals with the specifics related to the multiple domain problem.

3.1. Single domain in one-dimension

Consider the problem (1) discretized on the single domain \([0, 1]\) with a uniform mesh of \((N+1)\) points. The vector \(u = [u_0, u_1, \ldots, u_N]\) is the discrete approximation of \(u\). The discrete approximation of \(u\) at the grid point \(i\) is denoted \(u_i\). \(u_x\) and \(u_{xx}\) are the approximations of \(u_x\) and \(u_{xx}\), respectively. By using the SBP operators constructed in [9] and [15] we have

\[
\begin{align*}
    u_x &= D_1 u = P^{-1} Q u, \\
    u_{xx} &= D_{2n} u = D_1 (D_1 u) = (P^{-1} Q)^2 u, \quad \text{or} \quad (4) \\
    u_{xx} &= D_{2c} u = P^{-1} (-A + B S) u,
\end{align*}
\]

where \(A\) is a matrix with that satisfies \(A + A^T \geq 0\). \(P\) is a symmetric positive definite matrix. \(Q\) is an almost skew-symmetric matrix that satisfies

\[
Q + Q^T = B = \text{diag}([-1, 0, \ldots, 0, 1]). \quad (5)
\]

Both operators \(D_{2n}\) and \(D_{2c}\) satisfy the second derivative SBP property (8) below. Moreover, \(D_{2c}\) is a difference operator with minimum band-width. The operator \(S\) has the form (see [11]),

\[
S = \begin{bmatrix}
    -s_1 & \cdots & -s_r & 0 & \cdots \\
    1 \\
    \vdots \\
    1 \\
    \cdots & 0 & s_r & \cdots & s_1
\end{bmatrix}. \quad (6)
\]
The first and last row of $S$ approximates the first derivative at the two boundaries, respectively. For simplicity, we denote

$$
(D_{1B} u)_0 = \begin{cases} 
(D_1 u)_0, & \text{if } D_{2n} \text{ used,} \\
(S u)_0, & \text{if } D_{2c} \text{ used,}
\end{cases}
$$

$$
(D_{1B} u)_N = \begin{cases} 
(D_1 u)_N, & \text{if } D_{2n} \text{ used,} \\
(S u)_N, & \text{if } D_{2c} \text{ used.}
\end{cases}
$$

As a result, the semi-discrete approximation of (1) can be written

$$
\begin{align*}
\dot{u} + a D_1 u &= \varepsilon D_2 u + \tau L P^{-1} e_0 \left[ \alpha u_0 + \beta (D_{1B} u)_0 - g^L \right] \\
&\quad + \tau R P^{-1} e_N \left[ \gamma u_N + \delta (D_{1B} u)_N - g^R \right],
\end{align*}
$$

$$
(u(t = 0) = f.
$$

In (7), $D_2$ is either $D_{2n}$ or $D_{2c}$. The SAT treatment (see [16], [8], [11], [17], [15], [18]) is used to implement the boundary condition and the coefficients $\tau L$ and $\tau R$ are chosen to give a stable scheme. The vectors $e_0 = [1, 0, \ldots, 0]^T$ and $e_N = [0, 0, \ldots, 1]^T$ are used to place the penalty terms at the boundary points.

**Remark:** When the solution can be estimated in terms of all types of data, the problem is called strongly stable, see [32] for more details.

We define a discrete inner product and norm for the grid functions by

$$
(u, v)_P = u^T P v, \quad \|u\|_P^2 = (u, u)_P = u^T P u
$$

and

$$
(u, v)_{A+A^T} = u^T (A + A^T) v, \quad \|u\|_{A+A^T}^2 = (u, u)_{A+A^T} = u^T (A + A^T) u.
$$

If $P$ is a diagonal matrix with positive elements, it is referred to (with a slight abuse of notation) as a diagonal norm [10]. The relations (4)-(6) together
with the definitions of the norms above lead to the SBP relations

\[ u^T [PD_1 + (PD_1)^T] u = -u_0^2 + u_N^2, \]
\[ u^T [PD_{2n} + (PD_{2n})^T] u = \|D_1 u\|^2_P - 2u_0(D_1B u)_0 + 2u_N(D_1B u)_N, \]  \hspace{1cm} (8)
\[ u^T [PD_{2c} + (PD_{2c})^T] u = (u, u)_{A+A^T} - 2u_0(D_1B u)_0 + 2u_N(D_1B u)_N. \]

For more details on SBP approximations of second derivatives, see [11].

We apply the energy method by multiplying (7a) by \( u^T P \), and adding the transpose. This yields,

\[ \frac{d}{dt} \left( \|u\|^2_P \right) + 2\varepsilon \text{Diss} = au_0^2 - au_N^2 - 2\varepsilon u_0(D_1B u)_0 + 2\varepsilon u_N(D_1B u)_N \]
\[ + 2\tau^L \left[ \alpha u_0^2 + \beta u_0(D_1B u)_0 - u_0 g^L \right] + 2\tau^R \left[ \gamma u_N^2 + \delta u_N(D_1B u)_N - u_N g^R \right] \]
\[ = (a + 2\tau^L \alpha) \left( u_0 - \frac{\tau^L}{a + 2\tau^L \alpha} g^L \right)^2 - (a - 2\tau^R \gamma) \left( u_N + \frac{\tau^R}{a - 2\tau^R \gamma} g^R \right)^2 \]
\[ - \frac{\tau^L^2}{a + 2\tau^L \alpha} g^L^2 + \frac{\tau^R^2}{a - 2\tau^R \gamma} g^R^2 - 2(\varepsilon - \tau^L \beta) u_0(D_1B u)_0 \]
\[ + 2(\varepsilon + \tau^R \delta) u_N(D_1B u)_N. \]

(9)

where Diss represents \( \|D_1 u\|^2_P \) if \( D_{2n} \) is used and \( (u, u)_{A+A^T} \) if \( D_{2c} \) is used.

To cancel the indefinite terms (I) and (II) in equation (9), we choose

\[ \tau^L = \frac{\varepsilon}{\beta} \quad \text{and} \quad \tau^R = -\frac{\varepsilon}{\delta}. \]

(10)
Substituting (10) into (9) we have

\[ \frac{d}{dt}(\|u\|^2) + 2\varepsilon \text{Diss} = \left( a + \frac{2\alpha}{\beta} \varepsilon \right) \left( u_0 - \frac{\varepsilon}{\beta} \frac{1}{a + \frac{2\alpha}{\beta} \varepsilon} g^L \right)^2 \]

\[ - \left( a + \frac{2\gamma}{\delta} \varepsilon \right) \left( u_N - \frac{\varepsilon}{\delta} \frac{1}{a + \frac{2\gamma}{\delta} \varepsilon} g^R \right)^2 \]

\[ - \left( \frac{\varepsilon}{\beta} \right)^2 \left( \frac{1}{a + \frac{2\alpha}{\beta} \varepsilon} \right) g^{L2} + \left( \frac{\varepsilon}{\delta} \right)^2 \left( \frac{1}{a + \frac{2\gamma}{\delta} \varepsilon} \right) g^{R2}. \]

(11)

We have obtained the following result.

**Proposition 3.1** If condition (3) for well-posedness and (10) are satisfied, the approximation (7) is strongly stable.

**Remark:** The estimate (11) is completely similar to the continuous estimate (2), see also the remark above Proposition 2.1.

3.2. Multiple domains and interface conditions

Without loss of generality, we consider a computational domain which consists of two sub-domains. The unknown on the left sub-domain is denoted by \( u \) and on the right sub-domain by \( v \), respectively. The same technique described in the previous section is used here to discretize both domains. The corresponding notations are also modified by adding superscripts \( L \) and \( R \) in order to identify the left and right sub-domains.

Since the outer boundary treatment has already been discussed, we will only focus on the interface treatment. The coupling of \( u \) and \( v \) as well as the first derivatives \( D^L_i u \) and \( D^R_i v \) at the interface will be done by using various forms of the SAT technique. The content in this Section summarize some of the results in [23] but we specifically identify the difference between the compact and non-compact form of the second derivative.
3.2.1. The Baumann-Oden (BO) method

In this method (first proposed in [36]), the semi-discrete approximation of (1) is given by

\[ u_t + a D_L^1 u = \varepsilon D_L^2 u + \sigma_L^1 (P_L)^{-1} e_{N}^L (u_N - v_0) + \]
\[ \sigma_L^2 (P_L)^{-1} e_{N}^L [(D_{1B}^L u)_N - (D_{1B}^R v)_0] + \]
\[ \sigma_L^3 (P_L)^{-1} (D_{1B}^L)^T e_{N}^L (u_N - v_0) + B T_L, \]

\[ (12a) \]

\[ v_t + a D_R^1 v = \varepsilon D_R^2 v + \sigma_R^1 (P_R)^{-1} e_{0}^R (v_0 - u_N) + \]
\[ \sigma_R^2 (P_R)^{-1} e_{0}^R [(D_{1B}^R v)_0 - (D_{1B}^L u)_N] + \]
\[ \sigma_R^3 (P_R)^{-1} (D_{1B}^R)^T e_{0}^R (v_0 - u_N) + B T_R, \]

\[ (12b) \]

on the left and right subdomain respectively. The coefficients \( \sigma_L^1, \sigma_L^2, \sigma_L^3, \sigma_R^1, \sigma_R^2, \sigma_R^3 \) will be determined by stability considerations. \( D_L^2 \) represents \( D_{2n}^L \) or \( D_{2c}^L \) and \( D_R^2 \) represents \( D_{2n}^R \) or \( D_{2c}^R \). \( B T_L \) and \( B T_R \) are introduced to represent the stable left and right boundary terms respectively.

We apply the energy method by multiplying (12a) and (12b) with \( u^T P_L \) and \( v^T P_R \) respectively, adding the transposes, using the relations (8) and
summing up. That leads to

\[
\frac{d}{dt} (\|u\|_{PL}^2 + \|v\|_{PR}^2) + 2\varepsilon \text{Diss}^L + 2\varepsilon \text{Diss}^R = \\
= au_0^2 - av_N^2 - 2\varepsilon u_0(D_{1B}^L u)_0 \\
+ 2\varepsilon u_N(D_{1B}^L u)_N + 2\sigma_1^L u_N(u_N - v_0) \\
+ 2\sigma_2^L u_N[(D_{1B}^L u)_N - (D_{1B}^R v)_0] \\
+ 2\sigma_3^L (D_{1B}^L u)_N(u_N - v_0) + 2u^T BT^L \\
+ av_0^2 - av_N^2 - 2\varepsilon v_0(D_{1B}^R v)_0 \\
+ 2\varepsilon v_N(D_{1B}^R v)_N + 2\sigma_1^R v_0(v_0 - u_N) \\
+ 2\sigma_2^R v_0[(D_{1B}^R v)_0 - (D_{1B}^L u)_N] \\
+ 2\sigma_3^R (D_{1B}^R v)_0(v_0 - u_N) + 2v^T BT^R, 
\]

(13)

where we have used

\[
\text{Diss}^{L,R}(w) = \begin{cases} \\
\|D_1 w\|_{PL,R}^2, & \text{if } D_{2L,R}^2 = D_{2n}^{L,R} \\
(w, w)_{A_{L,R} + (A_{L,R})^T}, & \text{if } D_{2L,R}^2 = D_{2c}^{L,R} 
\end{cases} 
\]

(14)

and

\[
u^T e_0^L = u_0, \quad u^T e_N^L = u_N, \quad u^T [Q^L + (Q^L)^T] u = u_N^2 - u_0^2, \\
v^T e_0^R = v_0, \quad v^T e_N^R = v_N, \quad v^T [Q^R + (Q^R)^T] v = v_N^2 - v_0^2.
\]

Equation (13) can be written in matrix form as

\[
\frac{d}{dt} (\|u\|_{PL}^2 + \|v\|_{PR}^2) + 2\varepsilon \text{Diss}^L + 2\varepsilon \text{Diss}^R = u^T M_I u_I + BT. 
\]

(15)

In (15) BT collects all terms on the outer boundaries, that is,

\[
BT = 2u \cdot BT^L + au_0^2 - 2\varepsilon u_0(D_{1B}^L u)_0 + 2v \cdot BT^R - av_N^2 + 2\varepsilon v_N(D_{1B}^R v)_N,
\]

where

\[
BT^L = 2u \cdot BT^L + au_0^2 - 2\varepsilon u_0(D_{1B}^L u)_0, \\
BT^R = 2v \cdot BT^R - av_N^2 + 2\varepsilon v_N(D_{1B}^R v)_N.
\]
and
\[ u_I = \begin{bmatrix} u_N (D^L_{1B}u)_N & v_0 (D^R_{1B}v)_0 \end{bmatrix}^T.\]

\[ M_I = \begin{bmatrix} -a + 2\sigma^L_1 & \varepsilon + \sigma^L_2 & -\sigma^L_1 - \sigma^R_1 & -\sigma^L_2 - \sigma^R_3 \\ \varepsilon + \sigma^L_2 + \sigma^L_3 & 0 & -\sigma^L_3 - \sigma^R_2 & 0 \\ -\sigma^L_1 - \sigma^R_1 & -\sigma^L_3 - \sigma^R_2 & a + 2\sigma^R_1 & -\varepsilon + \sigma^R_2 + \sigma^R_3 \\ -\sigma^L_2 - \sigma^R_3 & 0 & -\varepsilon + \sigma^R_2 + \sigma^R_3 & 0 \end{bmatrix}. \]

Note that we already shown in Section 3.1 that BT is bounded and causes no stability problems.

We need a negative semi-definite \( M_I \) for stability. The three \( 2 \times 2 \) submatrices along the diagonal must be negative semi-definite, which yields the necessary conditions,

\[ -a + 2\sigma^L_1 \leq 0, \quad a + 2\sigma^R_1 \leq 0, \tag{16} \]

and

\[ \varepsilon + \sigma^L_2 + \sigma^L_3 = 0, \quad -\varepsilon + \sigma^R_2 + \sigma^R_3 = 0, \quad -\sigma^L_3 - \sigma^R_2 = 0. \tag{17} \]

The conditions (17) inserted into matrix \( M_I \) yields,

\[ M_I = \begin{bmatrix} -a + 2\sigma^L_1 & 0 & -\sigma^L_1 - \sigma^R_1 & 0 \\ 0 & 0 & 0 & 0 \\ -\sigma^L_1 - \sigma^R_1 & 0 & a + 2\sigma^R_1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \]

We can verify that \( M_I \) is negative semi-definite if

\[ \sigma^L_1 \leq \frac{a}{2}, \quad \sigma^R_1 = \sigma^L_1 - a, \tag{18} \]

which also satisfies (16).

The following proposition has been proved.
Proposition 3.2  Consider the semi-discrete scheme (12) for the well-posed problem (1). If
\[
\sigma_1^L \leq \frac{a}{2}, \quad \sigma_1^R = \sigma_1^L - a, \\
\sigma_2^R = \varepsilon + \sigma_2^L, \quad \sigma_3^L = -\varepsilon - \sigma_2^L, \quad \sigma_3^R = -\sigma_2^L. 
\]
and Proposition 3.1 holds, then (12) is stable.

This proof was also given in [23].

3.2.2. The Carpenter-Nordström-Gottlieb (CNG) method

In [8] the authors used $\sigma_3^L = \sigma_3^R = 0$ in the interface treatment. The semi-discrete approximation of (1) is given by
\[
u_t + aD_1^R
\]
\[
\begin{align*}
\mu_t + aD_1^L \mu &= \varepsilon D_2^L \mu + \sigma_1^L (P_L)^{-1} e_N^L (u_N - v_0) + \\
&\quad \sigma_2^L (P_L)^{-1} e_N^L [(D_{1B}^L \mu)_N - (D_{1B}^R v)_0] + B T^L, \\
\nu_t + aD_1^R \nu &= \varepsilon D_2^R \nu + \sigma_1^R (P_R)^{-1} e_0^R (v_0 - u_N) + \\
&\quad \sigma_2^R (P_R)^{-1} e_0^R [(D_{1B}^R \nu)_0 - (D_{1B}^L \mu)_N] + B T^R,
\end{align*}
\]
on the left and right subdomain respectively. The same notation as in (12a),(12b) is used.

By applying the energy method introduced in Section 3.2.1, we obtain the corresponding interface matrix $M_I$,
\[
M_I = \begin{bmatrix}
-a + 2\sigma_1^L & \varepsilon + \sigma_2^L & -\sigma_1^L - \sigma_1^R & -\sigma_2^L \\
\varepsilon + \sigma_2^L & 0 & -\sigma_2^R & 0 \\
-\sigma_1^L - \sigma_1^R & -\sigma_2^R & a + 2\sigma_1^R & -\varepsilon + \sigma_2^R \\
-\sigma_2^L & 0 & -\varepsilon + \sigma_2^R & 0
\end{bmatrix}.
\]
To show stability we need the following relations

\[
\text{Diss}^L = \kappa^L (D^L_{1B}u)^2_N + \text{Diss}^L, \quad \text{Diss}^R = \kappa^R (D^R_{1B}v)^2_0 + \text{Diss}^R
\]

where \(\kappa^L, \kappa^R > 0\) and \(\text{Diss}^L, \text{Diss}^R > 0\). \(\text{Diss}^L\) and \(\text{Diss}^R\) are defined in (14).

Note that \(\kappa^L, \kappa^R\) are proportional to the mesh sizes in the left and right domain respectively.

With this substitution equation (15) becomes

\[
\frac{d}{dt}(\|u\|^2_{PL} + \|v\|^2_{PR}) + 2\varepsilon \text{Diss}^L + 2\varepsilon \text{Diss}^R = u^T M'_{I} u_I + BT
\]

with

\[
M'_{I} = \begin{bmatrix}
-a + 2\sigma_1^L & \varepsilon + \sigma_2^L & -\sigma_1^L - \sigma_1^R & -\sigma_2^L \\
\varepsilon + \sigma_2^L & -2\varepsilon \kappa^L & -\sigma_2^R & 0 \\
-\sigma_1^L - \sigma_1^R & -\sigma_2^R & a + 2\sigma_1^R & -\varepsilon + \sigma_2^R \\
-\sigma_2^L & 0 & -\varepsilon + \sigma_2^R & -2\varepsilon \kappa^R
\end{bmatrix}
\]

and

\[
BT = 2u \cdot BT^L + a u_0^2 - 2\varepsilon u_0(D^L_{1B}u)_0 + 2v \cdot BT^R - av_N^2 + 2\varepsilon v_N(D^R_{1B}v)_N.
\]

BT from the outer boundaries is bounded as shown in Section 3.1.

**Remark:** The terms \(-2\varepsilon \kappa^L\) and \(-2\varepsilon \kappa^R\) inserted in \(M_{I}\) to form \(M'_{I}\) are necessary. Without them \(M'_{I}\) would not be negative semi-definite.

A sufficient condition for semi-negative definiteness of \(M'_{I}\) derived in [8] is written,

\[
\sigma_1^R = \sigma_1^L - a, \quad \sigma_2^R = \varepsilon + \sigma_2^L, \quad \sigma_1^L \leq \frac{a}{2} - \frac{\varepsilon}{4} \left[ \frac{(\sigma_2^L)^2}{\kappa^L} + \frac{(\sigma_2^L)^2}{\kappa^R} \right].
\]

The following proposition was proved in [8].
**Proposition 3.3** Consider the semi-discrete scheme (20) for the well-posed problem (1). If the conditions in (22) and Proposition 3.1 are satisfied, then (20) is stable.

**Remark:** The coefficients in (22) depend on $\kappa^L$ and $\kappa^R$, which in turn depend on the mesh size. As a result, the coefficients must be modified, when the grid is refined.

**Remark:** The number of unknown parameters in (12) and (20) are reduced once stability has been shown, see (19) and (22).

### 3.2.3. The Local Discontinuous Galerkin (LDG) method

In the LDG method (first introduced in [37], see also [38] and [39]), equation (1a) is written in first order form as

$$
\begin{align*}
    u_t + a u_x - \varepsilon p_x &= \sigma^L_1 (u - v) \delta(x_i) + \sigma^L_2 (p - q) \delta(x_i) \\
    -\varepsilon u_x + \varepsilon p &= \sigma^L_3 (u - v) \delta(x_i) \\
    &\text{for } x \in [0, x_i],
\end{align*}
$$

(23a)

and

$$
\begin{align*}
    v_t + a v_x - \varepsilon q &= \sigma^R_1 (v - u) \delta(x_i) + \sigma^R_2 (q - p) \delta(x_i) \\
    -\varepsilon v_x + \varepsilon q &= \sigma^R_3 (v - u) \delta(x_i) \\
    &\text{for } x \in [x_i, 1].
\end{align*}
$$

(23b)

In (23) $x_i$ denotes the location of the interface and $p \approx u_x$ and $q \approx v_x$ are intermediate variables. $\delta(x_i)$ is the delta function.

The semi-discrete approximation of (23) is

$$
\begin{align*}
    \mathbf{u}_t + a D^L_1 \mathbf{u} - \varepsilon D^L_1 \mathbf{p} &= \sigma^L_1 (P^L)^{-1} e^L_N (u_N - v_0) + \\
    &\quad \sigma^L_2 (P^L)^{-1} e^L_N (p_N - q_0) + B^L \\
    -\varepsilon D^L_1 \mathbf{u} + \varepsilon \mathbf{p} &= \sigma^L_3 (P^L)^{-1} e^L_N (u_N - v_0)
\end{align*}
$$

(24a)
on the left subdomain, and
\[ v_t + aD_1^R v - \varepsilon D_1^R q = \sigma_1^R (P^R)^{-1} e_0^R (v_0 - u_N) + \]
\[ \sigma_2^R (P^R)^{-1} e_0^R (q_0 - p_N) + B^T R \] (24b)

on the right subdomain. In (24) we have \( p \approx D_1^L u \) and \( q \approx D_1^R v \).

**Proposition 3.4** The conditions (19) in Proposition 3.2 lead to stability of the LDG method (24).

**Proof**: Applying the energy method introduced in Section 3.2.1 to (24) yields
\[ \frac{d}{dt} \left( \| u \|_{PL}^2 + \| v \|_{PR}^2 \right) + 2\varepsilon \| p \|_{PL}^2 + 2\varepsilon \| q \|_{PR}^2 = v^T I M I v + B^T, \] (25)

where \( M_I \) and \( B^T \) are given in (15) and \( v_I = [u_N \ p_N \ v_0 \ q_0]^T \).

Consequently, the conditions (19) also lead to an energy estimate for the LDG method. □

The relation between Proposition 3.2 and 3.4 was originally given in [23].

### 4. Spectral analysis

In this section we investigate the spectral properties of the various schemes. There are two main reasons for this investigation. Firstly, we need an accurate prediction of the eigenvalue with the largest real part. A positive real part leads to exponential growth and instability while a negative real part determines the convergence rate to steady-state, see [40]. An accurate prediction of the largest eigenvalue is also a requirement for an accurate prediction of the time development of the numerical solution. Secondly, to reduce stiffness and increase efficiency we want a spectrum with a limited size of the spectral radius, see [41].
4.1. The spectrum of the continuous problem

The Laplace transform of (1) with zero initial data gives

\[ s \hat{u} + a \hat{u}_x = \varepsilon \hat{u}_{xx}, \quad 0 \leq x \leq 1 \]

\[ \hat{u}(0) + \beta \hat{u}_x(0) = 0, \quad \hat{u}_x(1) = 0. \]

(26)

where \( \hat{u} \) is the Laplace transform of \( u \) and we have chosen \( \alpha = 1, \gamma = 0 \) and \( \delta = 1 \) as an example.

The general solution of (26) is \( \hat{u} = \sigma_1 e^{\kappa_1 x} + \sigma_2 e^{\kappa_2 x} \) with

\[ \kappa_{1,2} = \frac{a}{2\varepsilon} \left( 1 \pm \sqrt{1 + \frac{4\varepsilon}{a^2 s}} \right). \]

By applying the boundary conditions and demanding a unique solution we obtain that the Kreiss condition (see [32]) for stability of (26) is

\[ \det C(s) = \begin{vmatrix} 1 + \beta \kappa_1 & 1 + \beta \kappa_2 \\ \kappa_1 e^{\kappa_1} & \kappa_2 e^{\kappa_2} \end{vmatrix} \neq 0 \quad \text{for} \ \Re(s) \geq 0. \]

The spectrum of the continuous problem consist of \( s \)-values making

\[ \det C(s) = (1 + \beta \kappa_2) e^{\kappa_1} - (1 + \beta \kappa_1) e^{\kappa_2} = 0, \]

(27)

(for more details, see [10]).

We have the following lemma.

**Lemma 4.1** Equation (27) has a solution if \( s \leq -a^2/(4\varepsilon), \ s \in \Re. \)

The proof is presented in Appendix.

By choosing \( a = 1, \ \varepsilon = 0.1 \) and \( \beta = -2\varepsilon/a = -0.2 \) in (26), we obtain the two maximum eigenvalues \( s_c^1 = -2.67261695452793 \) and \( s_c^2 = -4.12696691192682 \) for the continuous system in (26). The least negative
part of the spectrum for continuous system is presented in Figure 1. Note that all the eigenvalues are real.

Remark: The purely real spectrum of the continuous advection-diffusion problem has also been observed in [10]. The spectrum of the advection problem ($\epsilon = 0$) with only one boundary condition at $x = 0$ has no continuous spectrum ($\operatorname{det} C(s) \neq 0$ for all $s$). The existence of the second derivative, albeit with a small $\epsilon$, changes the mathematical character of the problem completely; introduces one more boundary condition, produces a spectrum and make the advection-diffusion problem behave spectrally as the diffusion problem.

4.2. The spectrum of the semi-discrete problem

It is convenient to introduce notations for the methods introduced in the previous sections. If the approximation for the second derivative SBP operator is of the form $D_2 = D_1 \cdot D_1$ we denote it the non-compact form. The
formulation $D_2 = P^{-1}(-A + BS)$ is denoted the compact form. Moreover, we denote

$SIN$: the scheme (7) with non-compact form on single domain;

$SIC$: the scheme (7) with compact form on single domain;

$BON$: the Baumann-Oden scheme (12) with non-compact form;

$BOC$: the Baumann-Oden scheme(12) with compact form;

$CNGN$: the borrowing scheme (20) with non-compact form;

$CNGC$: the borrowing scheme (20) with compact form;

$LDG$: the local discontinuous Galerkin method (24) (No compact form);

$CON$: the continuous system (1).

All the semi-discrete schemes can be written in the form,

$$u_t = Au + \tilde{F},$$

(28)

where $A$ is a matrix and $\tilde{F}$ is a function of $F$, $g^L$ and $g^R$ given in (1). Note that the matrix $A$ may not be symmetric since we introduce boundary and interface terms. This means that parts of the spectra can be complex. This is contrary to spectrum for the continuous problem, which is purely real. If the number of grid points $N \to \infty$ the spectra of the semi-discrete problems converge to the spectra of the continuous problems since our approximations are stable and accurate. For a finite number of grid points, part of the spectra of semi-discrete problem corresponds to the spectra of the continuous problem.
Remark: The most important eigenvalue of $A$ in (28) is the one with the largest real part. A positive real part leads to exponential growth and instability while a negative real part determines the convergence rate to steady-state. By computing that eigenvalue and comparing it to the corresponding eigenvalue for the continuous problem we can determine whether the discrete and continuous problem have the same convergence rate, see for example [40]. It is of course also necessary for an accurate prediction of the time evolution. Moreover, it is a good test of the numerical scheme to investigate if it can capture that important quantity.

4.2.1. The eigenvalue with the largest real part

We compare the spectra of these schemes in the two sub-domains with uniform mesh of 161 grid points. The computation on $[0, 1]$ is divided into two sub-domains $[0, 1/2]$ and $[1/2, 1]$. For the BAN, BOC, and LDG schemes, we choose the coefficients $\sigma_{L_1} = a/2 = 1/2$ and $\sigma_{L_2} = \varepsilon$. The other coefficients are decided by the stability conditions in (19). In the CNGN and CNGC schemes, the coefficient $\sigma_{L_1}$ depends on $\kappa_L$ and $\kappa_R$, which increase with grid refinement. In all tests, $\sigma_{L_1}$ is determined by the maximum value under the stability condition (22), that is,

$$\sigma_{L_1} = \frac{a}{2} - \frac{1}{4\varepsilon} \left[ \left( \frac{\sigma_{R_2}^2}{\kappa_L} \right)^2 + \left( \frac{\sigma_{L_2}^2}{\kappa_R} \right)^2 \right].$$

Table 1 presents that the maximum eigenvalues for different $\varepsilon$ ($a = 1$). Note that LDG cannot employ the compact form to discretize the second derivative. Due to the compact form, the maximum eigenvalues of SIC, BOC, and CNGC agree well with the continuous system (see Table 1). But for the non-compact form, if $\varepsilon$ is small ($\varepsilon \leq 0.06$ for the second order as well
as $\varepsilon \leq 0.04$ for the sixth order), the maximum eigenvalues of the semi-discrete schemes do not correspond to that of the continuous system.

(a) Second order accuracy

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>0.005</th>
<th>0.01</th>
<th>0.02</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIN</td>
<td>-3.3187</td>
<td>-3.9873</td>
<td>-4.6158</td>
<td>-4.6724</td>
<td>-2.6732</td>
<td>-1.5110</td>
</tr>
<tr>
<td>BON</td>
<td>-3.8587</td>
<td>-4.8050</td>
<td>-5.6747</td>
<td>-5.0748</td>
<td>-2.6732</td>
<td>-1.5110</td>
</tr>
<tr>
<td>CNGN</td>
<td>-3.9489</td>
<td>-5.4306</td>
<td>-6.7694</td>
<td>-5.1153</td>
<td>-2.6732</td>
<td>-1.5110</td>
</tr>
<tr>
<td>LDG</td>
<td>-3.4369</td>
<td>-4.0825</td>
<td>-4.7007</td>
<td>-4.7111</td>
<td>-2.6728</td>
<td>-1.5110</td>
</tr>
<tr>
<td>SIC</td>
<td>-28.8214</td>
<td>-23.5382</td>
<td>-12.6223</td>
<td>-5.1068</td>
<td>-2.6732</td>
<td>-1.5110</td>
</tr>
<tr>
<td>BOC</td>
<td>-29.6753</td>
<td>-22.5115</td>
<td>-12.6223</td>
<td>-5.1068</td>
<td>-2.6732</td>
<td>-1.5110</td>
</tr>
<tr>
<td>CNGC</td>
<td>-29.6825</td>
<td>-22.5296</td>
<td>-12.6223</td>
<td>-5.1068</td>
<td>-2.6732</td>
<td>-1.5110</td>
</tr>
<tr>
<td>CON</td>
<td>-49.3632</td>
<td>-25.2135</td>
<td>-12.4380</td>
<td>-5.1021</td>
<td>-2.6726</td>
<td>-1.5109</td>
</tr>
</tbody>
</table>

(b) Sixth order accuracy

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>0.005</th>
<th>0.01</th>
<th>0.02</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIN</td>
<td>-5.1959</td>
<td>-6.4197</td>
<td>-7.7066</td>
<td>-5.1021</td>
<td>-2.6726</td>
<td>-1.5109</td>
</tr>
<tr>
<td>BON</td>
<td>-6.1718</td>
<td>-7.7586</td>
<td>-9.1441</td>
<td>-5.1021</td>
<td>-2.6726</td>
<td>-1.5109</td>
</tr>
<tr>
<td>CNGN</td>
<td>-6.7816</td>
<td>-9.2849</td>
<td>-10.8014</td>
<td>-5.1021</td>
<td>-2.6726</td>
<td>-1.5109</td>
</tr>
<tr>
<td>LDG</td>
<td>-5.3114</td>
<td>-6.5127</td>
<td>-8.3412</td>
<td>-5.1021</td>
<td>-2.6726</td>
<td>-1.5109</td>
</tr>
<tr>
<td>SIC</td>
<td>-29.3339</td>
<td>-22.9183</td>
<td>-12.5456</td>
<td>-5.1021</td>
<td>-2.6726</td>
<td>-1.5109</td>
</tr>
<tr>
<td>BOC</td>
<td>-28.9662</td>
<td>-22.5006</td>
<td>-12.5456</td>
<td>-5.1021</td>
<td>-2.6726</td>
<td>-1.5109</td>
</tr>
<tr>
<td>CNGC</td>
<td>-27.9956</td>
<td>-23.7056</td>
<td>-12.5456</td>
<td>-5.1021</td>
<td>-2.6726</td>
<td>-1.5109</td>
</tr>
<tr>
<td>CON</td>
<td>-49.3632</td>
<td>-25.2135</td>
<td>-12.4380</td>
<td>-5.1021</td>
<td>-2.6726</td>
<td>-1.5109</td>
</tr>
</tbody>
</table>

Table 1: The maximum eigenvalues with 161 grid points. $a = 1$.

Denote the convergence rate $q_e$ for the maximum eigenvalue by

$$ q_e = \frac{\log_{10} \left( \left| s_c - s_d^{(1)} \right| / \left| s_c - s_d^{(2)} \right| \right)}{\log_{10} \left( N^{(1)} / N^{(2)} \right)} $$

(29)
where $s_c$ and $s_d$ are the maximum eigenvalues of the continuous system and semi-discrete schemes. $s_d^{(1)}$ and $s_d^{(2)}$ are the maximum eigenvalues on the meshes of $N^{(1)}$ and $N^{(2)}$ grid points (including boundary points), respectively. The convergence rate $q_e$ with $a = 1$ and $\varepsilon = 0.1$ are shown in Tables 2 and 3. The eigenvalues of the semi-discrete system converges to the eigenvalues of continuous system with the grid refinement. With non-compact form, the SIN and LDG schemes have higher precision. For example, by using the sixth order accuracy SBP operator and 81 grid points, the SIN has eight digits precision while both BON and CNGN have 6 digits precision on multiple domain. The convergence rates of the non-compact forms (SIN, BON and CNGN) are almost similar to those of the compact forms (SIC, BOC and CNGC). However there are significant differences in error levels between the non-compact form and compact form (see Tables 2 and 3).

**Remark** All the semi-discrete approximations of (1) have spectra located in the left half of the complex plane, which means that the long-time behavior of the solution is correct.

4.2.2. The spectral radius

To speed up convergence to steady state and increase computational efficiency it is essential that the spectral radius of the numerical scheme is minimal, see [41]. In (19) there are six unknown variables and four equations. Let $\sigma_1^L$ and $\sigma_2^L$ be the free parameters. Table 4 shows the spectral radius of these schemes on a uniform mesh of 161 grid points on two sub-domains with different $\sigma_2^L$. $\sigma_1^L$ is fixed to $a/2$. We do not present more results with different $\sigma_1^L$ since $\sigma_1^L$ only affect the results marginally when $\sigma_1^L \in [-a, a/2]$. The
(a) Second order accuracy

```
| Points | $|s_d - s_c|$ | $|s_d - s_c|$ | $|s_d - s_c|$ | $|s_d - s_c|$ |
|--------|-------------|-------------|-------------|-------------|
| 41     | 2.2e-02     | 3.3e-03     | 5.4e-03     | 1.8e-02     |
| 81     | 1.8e-03     | 3.69        | 7.2e-04     | 2.23        |
| 121    | 4.9e-04     | 3.26        | 3.0e-04     | 2.19        |
| 161    | 2.1e-04     | 2.86        | 1.6e-04     | 2.16        |
| 201    | 1.2e-04     | 2.62        | 1.0e-04     | 2.13        |
```

(b) Sixth order accuracy

```
| Points | $|s_d - s_c|$ | $|s_d - s_c|$ | $|s_d - s_c|$ | $|s_d - s_c|$ |
|--------|-------------|-------------|-------------|-------------|
| 41     | 1.1e-05     | 6.5e-05     | 1.3e-04     | 9.9e-06     |
| 81     | 6.4e-08     | 7.38        | 2.5e-06     | 4.85        |
| 121    | 2.9e-09     | 7.60        | 3.5e-07     | 4.79        |
| 161    | 3.1e-10     | 7.81        | 8.6e-08     | 4.91        |
| 201    | 5.3e-11     | 7.88        | 2.9e-08     | 4.93        |
```

Table 2: The convergence rate of the maximum eigenvalue for non-compact form $D_2 = D_1$. $s_d$ and $s_c$ are the maximum eigenvalues for the semi-discrete schemes and continuous problem, respectively. $a = 1$ and $\varepsilon = 0.1$.  

22
spectral radius are almost same for all these schemes if reasonable coefficients are chosen. From these tables we find that \( \sigma_2^L = \mathcal{O}(\varepsilon) \) minimize the spectral radius. Note that the LDG scheme always has a minimal spectral radius when \( \sigma_1^L = a/2 \) and \( \sigma_2^L = -\varepsilon/2 \), that is, with the centered fluxes. When \( \sigma_1^L = 0 \) and \( \sigma_2^L = -\varepsilon \) the one-sided fluxes (see [39]) have been obtained. In this case, the LDG scheme has a rather small spectral radius (see Table 4).

(a) Second order accuracy

| Points | SIC \( |s_d - s_c| \) | \( q_e \) | BOC \( |s_d - s_c| \) | \( q_e \) | CNGC \( |s_d - s_c| \) | \( q_e \) |
|--------|------------------|-----|------------------|-----|------------------|-----|
| 41     | 8.9e-03          | 8.9e-03 | 8.8e-03          |     |                  |     |
| 81     | 2.2e-03          | 2.2e-03 | 2.2e-03          | 2.05| 2.2e-03          | 2.05|
| 121    | 9.8e-04          | 9.8e-04 | 9.8e-04          | 2.01| 9.8e-04          | 2.01|
| 161    | 5.5e-04          | 5.5e-04 | 5.5e-04          | 2.01| 5.5e-04          | 2.00|
| 201    | 3.5e-04          | 3.5e-04 | 3.5e-04          | 2.00| 3.5e-04          | 2.00|

(b) Sixth order accuracy

| Points | SIC \( |s_d - s_c| \) | \( q_e \) | BOC \( |s_d - s_c| \) | \( q_e \) | CNGC \( |s_d - s_c| \) | \( q_e \) |
|--------|------------------|-----|------------------|-----|------------------|-----|
| 41     | 1.7e-08          | 5.9e-08 | 2.2e-07          |     |                  |     |
| 81     | 1.5e-10          | 6.99 | 4.6e-09          | 3.74| 1.6e-08          | 3.89|
| 121    | 9.8e-12          | 6.74 | 8.8e-10          | 4.09| 3.0e-09          | 4.11|
| 161    | 7.1e-13          | 9.23 | 2.5e-10          | 4.47| 8.4e-10          | 4.51|
| 201    | 6.9e-12          |    | 8.0e-11          | 5.09| 2.9e-10          | 4.75|

Table 3: The convergence rate of the maximum eigenvalue for compact form \( D_2 = P^{-1}(-A + BS) \). \( s_d \) and \( s_c \) are the maximum eigenvalues for the semi-discrete schemes and continuous problem, respectively. \( a = 1 \) and \( \varepsilon = 0.1 \).
(a) Second order accuracy. SIN: $5.1e + 3$; SIC: $1.0e + 4$.

<table>
<thead>
<tr>
<th>$\sigma_L^2$</th>
<th>$-10\varepsilon$</th>
<th>$-\varepsilon$</th>
<th>$-\varepsilon/2$</th>
<th>$0$</th>
<th>$\varepsilon/2$</th>
<th>$\varepsilon$</th>
<th>$10\varepsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BON</td>
<td>$1.1e + 5$</td>
<td>$7.1e + 3$</td>
<td>$5.2e + 3$</td>
<td>$6.6e + 3$</td>
<td>$1.3e + 4$</td>
<td>$2.0e + 4$</td>
<td>$1.2e + 5$</td>
</tr>
<tr>
<td>CNGN</td>
<td>$9.4e + 5$</td>
<td>$5.2e + 3$</td>
<td>$5.2e + 3$</td>
<td>$5.6e + 3$</td>
<td>$1.3e + 4$</td>
<td>$2.5e + 4$</td>
<td>$1.0e + 6$</td>
</tr>
<tr>
<td>LDG</td>
<td>$3.7e + 6$</td>
<td>$1.6e + 4$</td>
<td>$5.2e + 3$</td>
<td>$1.8e + 4$</td>
<td>$4.7e + 4$</td>
<td>$9.9e + 4$</td>
<td>$4.5e + 5$</td>
</tr>
<tr>
<td>BOC</td>
<td>$2.0e + 5$</td>
<td>$1.2e + 4$</td>
<td>$1.0e + 4$</td>
<td>$1.2e + 4$</td>
<td>$2.2e + 4$</td>
<td>$3.3e + 4$</td>
<td>$2.4e + 5$</td>
</tr>
<tr>
<td>CNGC</td>
<td>$9.4e + 5$</td>
<td>$1.0e + 4$</td>
<td>$1.0e + 4$</td>
<td>$1.0e + 4$</td>
<td>$1.3e + 4$</td>
<td>$2.3e + 4$</td>
<td>$1.1e + 6$</td>
</tr>
</tbody>
</table>

(b) Sixth order accuracy. SIN: $1.5e + 4$; SIC: $3.6e + 4$.

<table>
<thead>
<tr>
<th>$\sigma_L^2$</th>
<th>$-10\varepsilon$</th>
<th>$-\varepsilon$</th>
<th>$-\varepsilon/2$</th>
<th>$0$</th>
<th>$\varepsilon/2$</th>
<th>$\varepsilon$</th>
<th>$10\varepsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BON</td>
<td>$3.1e + 5$</td>
<td>$1.7e + 4$</td>
<td>$1.6e + 4$</td>
<td>$1.6e + 4$</td>
<td>$3.1e + 4$</td>
<td>$4.8e + 4$</td>
<td>$3.3e + 5$</td>
</tr>
<tr>
<td>CNGN</td>
<td>$2.4e + 6$</td>
<td>$1.6e + 4$</td>
<td>$1.5e + 4$</td>
<td>$1.4e + 4$</td>
<td>$3.0e + 4$</td>
<td>$6.0e + 4$</td>
<td>$2.9e + 6$</td>
</tr>
<tr>
<td>LDG</td>
<td>$9.3e + 6$</td>
<td>$4.4e + 4$</td>
<td>$1.5e + 4$</td>
<td>$4.4e + 4$</td>
<td>$1.1e + 5$</td>
<td>$2.4e + 5$</td>
<td>$1.0e + 7$</td>
</tr>
<tr>
<td>BOC</td>
<td>$5.6e + 5$</td>
<td>$3.4e + 4$</td>
<td>$3.4e + 4$</td>
<td>$3.7e + 4$</td>
<td>$3.7e + 4$</td>
<td>$6.1e + 4$</td>
<td>$6.1e + 5$</td>
</tr>
<tr>
<td>CNGC</td>
<td>$2.3e + 6$</td>
<td>$3.6e + 4$</td>
<td>$3.4e + 4$</td>
<td>$3.6e + 4$</td>
<td>$3.6e + 4$</td>
<td>$5.3e + 4$</td>
<td>$2.8e + 6$</td>
</tr>
</tbody>
</table>

Table 4: The spectral radius. $a = 1$, $\varepsilon = 0.1$ and $\sigma_L^2 = a/2$. 

24
Remark By comparing with the schemes SIN and SIC (without interface), it is clear that the coupling schemes (with interfaces) do not significantly increase the spectral radius.

5. Numerical experiments

Denote the convergence rate $q$ in the computational domain by

$$q = \frac{\log_{10} \left( \frac{||u - v^{(1)}||_2}{||u - v^{(2)}||_2} \right)}{\log_{10} \left( \frac{N^{(1)}}{N^{(2)}} \right)}$$

(30)

where $u$ is an exact solution. $v^{(1)}$ and $v^{(2)}$ are the corresponding numerical solutions on the meshes of $N^{(1)}$ and $N^{(2)}$ grid points (including boundary points), respectively.

With a diagonal norm, the first derivative SBP operator was constructed with $2p$-th order internal accuracy and $p$-th order at the boundary (see [9], [11] and [15]). According to [42], $(p + 1)$-th order accuracy is achieved in a hyperbolic equation which only includes the first derivative. For example, an SBP operator with sixth order internal accuracy and third order accurate boundary closures will lead to a fourth order accurate scheme.

In the advection-diffusion equation, as described previously, there are two options to construct the SBP operator for the second derivative. The non-compact form is obtained by using the first derivative operator $D_1 = P^{-1}Q$ twice, that is, $D_2 = D_1 \cdot D_1$. With a diagonal norm, we obtain a boundary closure of order $(p - 1)$-th. In the compact form we use the minimal width operator $D_2 = P^{-1}(-A + BS)$, and the second derivative SBP operators have $p$-th order accuracy at the boundaries, see [11] for details. It was proved in [43] that if the solution is point-wise bounded, the accuracy
of advection-diffusion equation is two orders higher than the accuracy of the second derivative approximation at the boundaries. For clarity, the theoretical convergence rate is shown in Table 5.

<table>
<thead>
<tr>
<th>Hyperbolic</th>
<th>Viscous</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>internal</td>
<td>boundary</td>
<td>internal</td>
</tr>
<tr>
<td>$q$ with non-compact form</td>
<td>$2p$</td>
<td>$p$</td>
</tr>
<tr>
<td>$q$ with compact form</td>
<td>$2p$</td>
<td>$p$</td>
</tr>
</tbody>
</table>

Table 5: The theoretical convergence rate by using different SBP operators with diagonal norm. $p = 1, 2, \text{ and } 3$. (*) For the compact form and $p = 1$ we get $q = 2$.

One exact solution to the advection-diffusion equation (1) is

$$u = \sin(w(x - ct))e^{-bx}, \quad c > 0, \quad w = \sqrt{\frac{c^2 - a^2}{2\varepsilon}}, \quad b = \frac{c - a}{2\varepsilon}, \quad |c| > |a|.$$ 

In the following analysis we have chosen $a = 1$, $\varepsilon = 0.01$, $c = 1.01$ and $\alpha = 1$, $\beta = -0.01$, $\gamma = 0$, $\delta = 1$. We use the classical fourth-order Runge-Kutta method for the time integration. A small time-step is used to minimize the temporal errors.

5.1. One dimension

5.1.1. Single domain

We begin by studying the accuracy of the SBP operators on a single domain. The convergence rate for both options of the second derivatives are
shown in Table 6. The results are in line with the theoretical prediction in Table 5.

\[
D_2 = D_1 \cdot D_1 = D_2 = P^{-1}(-A + BS)
\]

<table>
<thead>
<tr>
<th>Points</th>
<th>(L_2)-Err</th>
<th>(q)</th>
<th>(L_2)-Err</th>
<th>(q)</th>
<th>(L_2)-Err</th>
<th>(q)</th>
<th>(L_2)-Err</th>
<th>(q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>-1.10</td>
<td>-</td>
<td>-2.38</td>
<td>-</td>
<td>-1.11</td>
<td>-</td>
<td>-3.02</td>
<td>-</td>
</tr>
<tr>
<td>81</td>
<td>-1.71</td>
<td>2.05</td>
<td>-3.55</td>
<td>3.95</td>
<td>-1.71</td>
<td>2.05</td>
<td>-4.29</td>
<td>4.28</td>
</tr>
<tr>
<td>161</td>
<td>-2.31</td>
<td>2.02</td>
<td>-4.75</td>
<td>3.95</td>
<td>-2.32</td>
<td>2.02</td>
<td>-5.82</td>
<td>5.14</td>
</tr>
<tr>
<td>321</td>
<td>-2.91</td>
<td>2.00</td>
<td>-5.94</td>
<td>4.02</td>
<td>-2.92</td>
<td>2.01</td>
<td>-7.43</td>
<td>5.35</td>
</tr>
<tr>
<td>641</td>
<td>-3.51</td>
<td>2.00</td>
<td>-7.15</td>
<td>3.99</td>
<td>-3.52</td>
<td>2.01</td>
<td>-9.06</td>
<td>5.35</td>
</tr>
</tbody>
</table>

Table 6: Grid convergence of \(u_t + u_x = 0.01u_{xx}\). A single domain \([-1, 1]\).

5.1.2. Two sub-domains with an interface

Recall that the LDG method prohibits the use of the compact form for the second derivative. We apply the non-compact form for a fair comparison between the different methods.

The convergence rates are calculated on two sub-domains of uniform mesh with an interface at \(x = 0\) (see Table 7). The uniform mesh is refined from 42 to 1282 grid points. As in the single domain case, the convergence rates for the non-compact formulation agree with the theory in [42], [44], [11], and [43]. Note that the convergence rate \(q_e\) of maximum eigenvalue with the non-compact form (see Table 2) is one order higher than the convergence rate \(q\) for the approximations (see Table 7).
(a) second order accuracy

<table>
<thead>
<tr>
<th>Points</th>
<th>( L_2 )-Err</th>
<th>( q )</th>
<th>( L_2 )-Err</th>
<th>( q )</th>
<th>( L_2 )-Err</th>
<th>( q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>-1.07</td>
<td>-</td>
<td>-1.07</td>
<td>-</td>
<td>-1.07</td>
<td>-</td>
</tr>
<tr>
<td>81</td>
<td>-1.69</td>
<td>2.13</td>
<td>-1.69</td>
<td>2.13</td>
<td>-1.70</td>
<td>2.14</td>
</tr>
<tr>
<td>161</td>
<td>-2.30</td>
<td>2.06</td>
<td>-2.30</td>
<td>2.06</td>
<td>-2.30</td>
<td>2.06</td>
</tr>
<tr>
<td>321</td>
<td>-2.90</td>
<td>2.02</td>
<td>-2.90</td>
<td>2.03</td>
<td>-2.91</td>
<td>2.03</td>
</tr>
<tr>
<td>641</td>
<td>-3.51</td>
<td>2.01</td>
<td>-3.51</td>
<td>2.01</td>
<td>-3.51</td>
<td>2.01</td>
</tr>
</tbody>
</table>

(b) sixth order accuracy

<table>
<thead>
<tr>
<th>Points</th>
<th>( L_2 )-Err</th>
<th>( q )</th>
<th>( L_2 )-Err</th>
<th>( q )</th>
<th>( L_2 )-Err</th>
<th>( q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>-2.02</td>
<td>-</td>
<td>-2.02</td>
<td>-</td>
<td>-1.93</td>
<td>-</td>
</tr>
<tr>
<td>321</td>
<td>-5.79</td>
<td>4.15</td>
<td>-5.80</td>
<td>4.09</td>
<td>-5.93</td>
<td>4.39</td>
</tr>
<tr>
<td>641</td>
<td>-7.00</td>
<td>4.00</td>
<td>-7.00</td>
<td>4.00</td>
<td>-7.17</td>
<td>4.14</td>
</tr>
</tbody>
</table>

Table 7: Grid convergence of \( u_t + u_x = 0.01u_{xx} \). Two sub-domains, uniform mesh in \([-1,1]\).
Now the convergence rate $q$ is tested on two sub-domains with non-uniform grid. We start with 41 grid points in the left subdomain and 11 grid points in the right subdomain. For each refinement the grid points are doubled in both sub-domains. Table 8 presents the results using a non-compact second derivative. The convergence rate exactly coincide with the theoretical values.

(a) second order accuracy

<table>
<thead>
<tr>
<th>Points</th>
<th>BON $L_2$-Err</th>
<th>$q$</th>
<th>CNGN $L_2$-Err</th>
<th>$q$</th>
<th>LDG $L_2$-Err</th>
<th>$q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>41 + 11</td>
<td>-0.99</td>
<td>-</td>
<td>-0.93</td>
<td>-</td>
<td>-0.96</td>
<td>-</td>
</tr>
<tr>
<td>81 + 21</td>
<td>-1.64</td>
<td>2.19</td>
<td>-1.63</td>
<td>2.29</td>
<td>-1.60</td>
<td>2.14</td>
</tr>
<tr>
<td>161 + 41</td>
<td>-2.26</td>
<td>2.12</td>
<td>-2.27</td>
<td>2.13</td>
<td>-2.22</td>
<td>2.05</td>
</tr>
<tr>
<td>321 + 81</td>
<td>-2.88</td>
<td>2.05</td>
<td>-2.89</td>
<td>2.03</td>
<td>-2.83</td>
<td>2.01</td>
</tr>
<tr>
<td>641 + 161</td>
<td>-3.48</td>
<td>2.01</td>
<td>-3.48</td>
<td>2.01</td>
<td>-3.43</td>
<td>2.00</td>
</tr>
</tbody>
</table>

(b) sixth order accuracy

<table>
<thead>
<tr>
<th>Points</th>
<th>BON $L_2$-Err</th>
<th>$q$</th>
<th>CNGN $L_2$-Err</th>
<th>$q$</th>
<th>LDG $L_2$-Err</th>
<th>$q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>41 + 11</td>
<td>-1.45</td>
<td>-</td>
<td>-1.34</td>
<td>-</td>
<td>-1.33</td>
<td>-</td>
</tr>
<tr>
<td>81 + 21</td>
<td>-2.62</td>
<td>3.97</td>
<td>-2.65</td>
<td>4.37</td>
<td>-2.18</td>
<td>2.83</td>
</tr>
<tr>
<td>161 + 41</td>
<td>-3.81</td>
<td>4.04</td>
<td>-3.86</td>
<td>4.03</td>
<td>-3.27</td>
<td>3.62</td>
</tr>
<tr>
<td>321 + 81</td>
<td>-5.05</td>
<td>4.13</td>
<td>-5.06</td>
<td>3.97</td>
<td>-4.47</td>
<td>3.98</td>
</tr>
<tr>
<td>641 + 161</td>
<td>-6.26</td>
<td>4.04</td>
<td>-6.26</td>
<td>3.99</td>
<td>-5.68</td>
<td>4.03</td>
</tr>
</tbody>
</table>

Table 8: Grid convergence for non-compact form. Two sub-domains, non-uniform mesh.

So far we have used the non-compact form. Table 9 shows the conver-
gence $q$ for the BON and CNGN schemes on compact form. In Table 9 the convergence rate for the second and sixth order accurate schemes are in line with the theoretical conclusion in Table 5. Note that the convergence rates using the sixth order scheme in Table 9 attain almost 6 while the theoretical value is 5.

<table>
<thead>
<tr>
<th>Points</th>
<th>2nd (BOC)</th>
<th>6th (BOC)</th>
<th>2nd (CNGC)</th>
<th>6th (CNGC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$L_2$-Err</td>
<td>$q$</td>
<td>$L_2$-Err</td>
<td>$q$</td>
</tr>
<tr>
<td>41</td>
<td>-1.07</td>
<td>-2.35</td>
<td>-1.07</td>
<td>-2.37</td>
</tr>
<tr>
<td>81</td>
<td>-1.68</td>
<td>2.02</td>
<td>-3.77</td>
<td>4.89</td>
</tr>
<tr>
<td>161</td>
<td>-2.29</td>
<td>2.02</td>
<td>-5.34</td>
<td>5.30</td>
</tr>
<tr>
<td>321</td>
<td>-2.89</td>
<td>2.00</td>
<td>-7.04</td>
<td>5.70</td>
</tr>
<tr>
<td>641</td>
<td>-3.50</td>
<td>2.00</td>
<td>-8.83</td>
<td>5.70</td>
</tr>
</tbody>
</table>

Table 9: Grid convergence for compact form. Two sub-domains, non-uniform mesh.

5.1.3. The reflecting properties

To test the reflecting and oscillation properties of these schemes, a “wave” like analytic solution of (1) is chosen

$$u = \kappa \exp(-\theta(x - ct + b)^2).$$  \hspace{1cm} (31)

The exact solutions at $T = 0.3$, $T = 0.8$ and $T = 1.3$ are shown in Figure
2. With increasing time, the solution propagate from left to right without changing form. The calculation in this section is done on an equidistant grid for both domains.

![Figure 2: Exact solution.](image)

The error of the schemes are presented at $T = 0.3$, $T = 0.8$ and $T = 1.3$ in Figures 3 and 4. We notice that for the SIN scheme (without an interface), the error propagate from left to right. In both the second order accurate cases (see Figures 3) and the sixth order accurate case (see Figures 4), the error propagate from left to right via the interface at $x = 0$ without reflection for all compact schemes. However the schemes BON, CNGN and LDG (with non-compact form) lead to an oscillatory error caused by the interface. The error attains a maximum when the “wave” pass the interface around $T = 0.8$ (see Figures 4). But it is also clearly seen propagating backwards at $T = 0.8$. The schemes BOC and CNGC (with compact form) also introduces an oscillation at interface, however the magnitude of the error is very small compared with the non-compact schemes.

Sixth order accurate dissipation operators [45] are introduced in the non-compact schemes BON, CNGN and LDG. The calculations are shown in
Figure 3: Error. Second order accuracy. 81 points used. $\sigma_L^2 = -\varepsilon/2$. 
Figure 4: Error. Sixth order accuracy. 81 points used. $\sigma_2^L = -\varepsilon/2$. 
Figure 5. By comparing with the calculations in Figure 4 we find that the artificial dissipation operators kill the non-physical numerical oscillations efficiently for schemes BON and CNGN. However the artificial dissipation only reduce the magnitude of the oscillation of LDG to 30%.

Figure 5: Error. Sixth order accuracy, artificial dissipation. 81 points used. $\sigma_2^L = -\varepsilon/2$.

5.2. Multi-domains in two-dimensions

The SAT formulation can easily be generalized to several space dimensions. We demonstrate that by using the Baumann-Orden scheme described
in Section 3.2.1 with unequally spaced sub-domains. The domain $-1 \leq x, y \leq 1$ is divided into four sub-domains, each with different number of points and a uniform distribution. The domain interfaces are located on $x = 0$ and $y = 0$ (see Figure 6). Note that the mesh is discontinuous at the interfaces.

![Figure 6: A mesh with four sub-domains. Subdomain 1: 91 $\times$ 61 grid points; Subdomain 2: 91 $\times$ 23 grid points; Sub-domain 3: 15 $\times$ 61 grid points; Sub-domain 4: 15 $\times$ 23 grid points.](image)

The model problem in two dimensions can be written

$$u_t + au_x + bu_y = \varepsilon(u_{xx} + u_{yy}) + F, \quad -1 \leq x, y \leq 1, \quad t > 0, \quad (33)$$

with suitable initial data and boundary data. In the test below we used $a = 1$, $b = 1$, and $\varepsilon = 0.1$. In order to estimate the accuracy of the scheme,
an exact solution \( u = \sin(2\pi(x + y - 2t)) \) has been chosen. The initial data, boundary data and the forcing function \( F \) are adjusted to correspond to the exact solution. Table 10 shows a grid-refinement study for three different orders of accuracy. Note that the convergence rate approaches the theoretical rates studied previously in the one-dimensional cases.

\[
\begin{array}{cccccc}
\text{Points} & \text{2nd} & \text{4th} & \text{6th} \\
& L_2\text{-Err} & q & L_2\text{-Err} & q & L_2\text{-Err} & q \\
4144 & -1.57 & - & -2.17 & - & -2.08 & - \\
8904 & -1.94 & 2.20 & -2.63 & 2.77 & -2.58 & 3.07 \\
15904 & -2.23 & 2.30 & -3.00 & 2.93 & -3.03 & 3.57 \\
24603 & -2.43 & 2.04 & -3.28 & 2.92 & -3.38 & 3.73 \\
35404 & -2.58 & 2.02 & -3.51 & 2.94 & -3.67 & 3.70 \\
62604 & -2.84 & 2.03 & -3.88 & 2.97 & -4.14 & 3.78 \\
97304 & -2.84 & 2.03 & -4.17 & 3.00 & -4.57 & 3.85 \\
\end{array}
\]

Table 10: The convergence rate of \( u_t + u_x + u_y = 0.1(u_{xx} + u_{yy}) + F \) with non-compact form in four subdomains of nonuniform mesh in two dimensions.

We also consider the reflection properties from the interfaces in two dimensions. The analytic solution

\[
u(x, y, t) = \kappa \exp(-\theta((x - c_1 t + b_1)^2 + (y - c_2 t + b_2)^2)), \quad (34)\]

is used as boundary and initial data.

In this test, \( \kappa = 0.5, \theta = 50, c_1 = 1, b_1 = 0.5, c_2 = 1, \) and \( c_2 = 0.5. \) Figure 7 shows the numerical results at \( t = 0.1, 0.3, 0.5, 0.7, 0.9, \)
and 1.5 with the scheme. Between $t = 0.3$ and $t = 0.7$ the vortex propagates close to the interfaces $y = 0$ and $x = 0$. No problems could be detected at the interfaces and the reflexion is very small indeed, see Figure 8. For even more complex geometries, we can use our technique with hybrid methods, see [14], [24] and [46] or the recently developed method with non-matching grid lines [47].

### 6. Conclusions

Stable and accurate interface treatments for the linear advection-diffusion equation have been studied. The treatment is based on SBP operators and the SAT technique, which lead to an energy estimate and stability. Accurate high order calculations are performed in both single domain and multiple domains with an interface.

Three stable interface procedures: the Baumann-Oden method, the Carpenter-Nordström-Gottlieb method and the local discontinuous Galerkin method have been investigated. The compact form and non-compact form of the second derivative SBP operators have also been compared.

The spectral radius for the schemes depend on the chosen coefficients. The interface procedures do not increase the spectral radius if suitable penalty parameters chosen. In particular, when the centered fluxes were used in the LDG, the minimal spectral radius was been obtained.

By using the compact form we can obtain one order higher accuracy than for the non-compact form. Moreover, the compact form introduces less reflection and oscillation than the non-compact form. Artificial dissipation can reduce the non-physical oscillation from the interface for the non-compact
Figure 7: The calculation on a mesh with four subdomains and sixth order accuracy. Subdomain 1: 81 × 61 points; subdomain 2: 81 × 41 points; subdomain 3: 31 × 61 points; subdomain 4: 31 × 41 points.
In short, this analysis show that only minor differences separates the different interface procedures. However the local discontinuous Galerkin method is more difficult to implement since the scheme requires one to rewrite the original viscous problem as a first order system of equations.

Appendix

Here we prove Lemma 4.1.

When \( s \leq -a^2/(4\varepsilon) \), \( s \in \mathbb{R} \), the term \( \sqrt{1 + 4s\varepsilon/a^2} \) is a pure imaginary number and \( \kappa_{1,2} \) becomes

\[
\kappa_{1,2} = \frac{a}{2\varepsilon} \left( 1 \pm \sqrt{1 + \frac{4s\varepsilon}{a^2}} \right) = c(1 \pm b(s)i)
\]

where \( c = a/(2\varepsilon) > 0 \in \mathbb{R} \) is a constant and \( b(s) = \left| \sqrt{-(1 + 4s\varepsilon/a^2)} \right| \geq 0 \in \mathbb{R} \).
Substituting \( \kappa_{1,2} \) into (27) yields
\[
\det C(s) = \left[ 1 + \beta c(1 - b(s)i) \right] c(1 + b(s)i)e^{c(1+b(s))i}
- \left[ 1 + \beta c(1 + b(s)i) \right] c(1 - b(s)i)e^{c(1-b(s))i}
= 2ce^{c\left[ b(s) \cos(cb(s)) + (1 + \beta c + \beta cb^2(s)) \sin(cb(s)) \right]i}
= F(b(s))i
\]
which only contains imaginary parts and the continuous real-function
\[
F(b(s)) = 2ce^{c\left[ b(s) \cos(cb(s)) + (1 + \beta c + \beta cb^2(s)) \sin(cb(s)) \right]}, \quad F(b(s)) \in \mathbb{R}.
\]
We have
\[
F\left(b(s) = \frac{2n\pi}{c}\right) = 2ce^{-\frac{2n\pi}{c}} = 4ne^c \pi > 0, \quad n = 1, 2, \ldots
F\left(b(s) = \frac{(2n+1)\pi}{c}\right) = -2ce^{-\frac{2n\pi}{c}} = -4ne^c \pi < 0, \quad n = 1, 2, \ldots
\]
As a result, there exists \( b(s) \in [2n\pi/c, (2n+1)\pi/c] \) or
\[
s = -\frac{a^2(b^2+1)}{4\varepsilon} \in \left[ -\frac{a^2((2n+1)\pi)^2 + c^2}{4\varepsilon c^2}, -\frac{a^2(2n\pi)^2 + c^2}{4\varepsilon c^2} \right], \quad n = 1, 2, \ldots
\]
such that \( F(b(s)) = 0 \). Therefore, (27) has a solution for \( s \leq -a^2/(4\varepsilon) \), \( s \in \mathbb{R} \).

However, we note that when \( s > -a^2/(4\varepsilon) \), \( s \in \mathbb{R} \), the equation (27) has no solution. To verify this, we rewrite the term \( \kappa_{1,2} \),
\[
\kappa_{1,2} = \frac{a}{2\varepsilon} \left( 1 \pm \sqrt{1 + \frac{4s\varepsilon}{a^2}} \right) = c(1 \pm b(s))
\]
where \( b(s) = |\sqrt{1 + 4s\varepsilon/a^2}| > 0, b(s) \in \mathbb{R} \).

Applying \( \kappa_{1,2} \) to (27) leads to
\[
\det C(s) = \left[ 1 + \beta c(1 - b(s)) \right] c(1 + b(s))e^{c(1+b(s))}
- \left[ 1 + \beta c(1 + b(s)) \right] c(1 - b(s))e^{c(1-b(s))}
= ce^{c\left(1-b(s)\right)}\left[ e^{2c\beta b(s)}(1 + b(s) + \beta c - \beta cb^2(s)) - (1 - b(s) + \beta c - \beta cb^2(s)) \right]
\]
The inequality derived from the well-posedness condition (3),

$$-1 \leq \beta c \leq 0,$$

implies that the term $1 + b(s) + \beta c - \beta cb^2(s)$ is positive. Consequently we have

$$\det C(s) > ce^{c(1-b(s))}[(1 + b(s) + \beta c - \beta cb^2(s)) - (1 - b(s) + \beta c - \beta cb^2(s))]$$

$$= 2ce^{c(1-b(s))}b(s) > 0$$

(37)

Therefore, $\det C(s)$ is always positive for $s > -a^2/(4\varepsilon)$, $s \in \mathbb{R}$. ■

References


[26] F. Edelvik and G. Ledfelt. Explicit hybrid time domain solver for the


