Constructing non-reflecting boundary conditions using summation-by-parts in time

Hannes Frenander and Jan Nordström

Journal Article

Tweet

Postprint available at: Linköping University Electronic Press

http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-133853

LiU LINKÖPINGS UNIVERSITET

Constructing non–reflecting boundary conditions using summation–by–parts in time

Hannes Frenander, Jan Nordström

Division of Computational Mathematics, Department of Mathematics, Linköping University, SE-58183 Linköping. Sweden

Abstract

In this paper we provide a new approach for constructing non–reflecting boundary conditions. The boundary conditions are based on summation–by–parts operators and derived without Laplace transformation in time. We prove that the new non–reflecting boundary conditions yield a well-posed problem and that the corresponding numerical approximation is unconditionally stable. The analysis is demonstrated on a hyperbolic system in two space dimensions, and the theoretical results are confirmed by numerical experiments.

Keywords: Non-reflecting boundary conditions, summation-by-parts, simultaneous approximation terms, finite differences, stability, accuracy

1. Introduction

Many applications in physics and engineering involve unbounded physical domains, which must be limited by Artificial Boundary Conditions (ABC's). These boundary conditions will, if not chosen appropriately, produce non-physical reflections that pollute the solution in the interior of the domain.

Non–Reflecting Boundary Conditions (NRBC's), i.e. ABC's that do not produce reflections, are in most cases obtained in a transformed dual space, see [4, 5, 15]. The NRBC's are exact in the transformed space, but hard to implement since they are expressed in the dual variables. To circumvent this issue, the exact boundary conditions are often approximated using various types of expansions in combination with suitable size assumptions on the

frequencies involved [4, 15]. The resulting approximate boundary conditions are local in both space and time, and relatively easy to implement.

Unfortunately, this approach has some drawbacks. First and foremost, although the exact NRBC's result in a well–posed problem, an approximation of these often leads to an ill–posed problem [4], and consequently an unstable scheme. Secondly, even if the approximate NRBC's lead to a well-posed problem and a stable scheme, the accuracy is ruined since the amplitude of the reflections is independent of the mesh size. The reflections will therefore not vanish during mesh–refinement [5].

Another quite different approach for constructing approximate NRBC's is to introduce buffer zones as artificial boundaries, where incoming waves are damped. When the interface between the computational domain and the buffer zone is exactly non-reflecting, it is called a Perfectly Matched Layer (PML) [1]. Buffer zone techniques will not be discussed further in this paper. For comprehensive reviews of NRBC's including PML's as well as other techniques, the reader is referred to [9, 6, 18].

The use of Summation–By–Parts (SBP) operators [7, 11, 16] for the discrete formulation of partial differential equations has proven to be very successful. The boundary conditions for SBP approximations are implemented weakly by using Simultaneous Approximation Terms (SAT) [2, 3]. In this paper, we will make use of the SBP-SAT technique in time [10, 12, 13] when constructing the NRBC's. By using this technique, the complete analysis can be performed in real space, and hence, the NRBC's can be implemented relatively easily. The resulting boundary conditions are global in space and time, and the whole procedure bypasses the transformation and accuracy problems mentioned above.

The remainder of this paper will proceed as follows. We describe the SBP-SAT technique on a one-dimensional problem in Section 2. In Section 3, NRBC's are derived for a general two-dimensional hyperbolic problem. It is shown that the boundary conditions result in a well-posed problem. By considering the discrete problem, we introduce an alternative way of deriving the NRBC's in Section 4. The theoretical results are verified by numerical experiments in Section 5. Finally, in Section 6, we summarize the results and draw conclusions.

2. The SBP-SAT technique

To introduce the SBP-SAT technique, we consider the advection equation in one space dimension,

$$\begin{aligned}
u_t + u_x &= 0, \quad x \in [0,1], \quad t > 0, \\
u(0,t) &= g(t), \\
u(x,0) &= f(x).
\end{aligned} \tag{1}$$

To examine well-posedness of (1), we multiply with u and integrate in space and time to get

$$||u||^2 + \int_0^t u^2(1,\tau)d\tau = ||f||^2 + \int_0^t g^2(\tau)d\tau, \tag{2}$$

where the boundary and initial conditions in (1) and the notation $||u||^2 = \int_0^1 u^2(x,t)dx$ have been used. Obviously, u is bounded by the data f and g which implies that the problem is strongly well-posed. For a detailed discussion on well-posedness of initial boundary value problems, see [8, 14].

2.1. The fully discrete problem

Equation (1) is discretized using the SBP-SAT technique in space and time,

$$\begin{aligned}
(P_t^{-1}Q_t \otimes I_x)v + (I_t \otimes P_x^{-1}Q_x)v = \\
\alpha_t(P_t^{-1}E_{0t} \otimes I_x)(v - \bar{f}) + \\
\alpha_x(I_t \otimes P_x^{-1}E_{0x})(v - \bar{g}),
\end{aligned} \tag{3}$$

where $P_{t,x}$ and $Q_{t,x}$ are the SBP-operators satisfying $P_{t,x} = P_{t,x}^T > 0$ and $Q_{t,x} + Q_{t,x}^T = diag(-1,0,...,0,1)$. Moreover, P_t is proportional to the time step Δt and P_x is proportional to the grid spacing Δx. In (3), $E_{0t,x} = diag(1,0,...,0)$ are of the same sizes as $P_{t,x}$, respectively, and $\bar{f}$, $\bar{g}$ are grid functions with the values of f, g injected at appropriate grid points. The symbol $\otimes$ denotes the Kronecker product which, for two arbitrary matrices M and N, is defined as

$$M \otimes N = \begin{bmatrix} M_{11}N & \cdots & \cdots & M_{1n}N \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ M_{m1}N & \cdots & \cdots & M_{mn}N \end{bmatrix}.$$

The penalty coefficients $\alpha_{t,x}$ will be determined such that the scheme becomes stable.

We now multiply (3) with $v^T(P_t \otimes P_x)$ from the left, choose $\alpha_t = \alpha_x = -1$ and add the transpose of the outcome to itself to find

$$v^T(E_{Nt} \otimes P_x)v + v^T(P_t \otimes E_{Nx})v = \bar{f}^T(E_{0t} \otimes P_x)\bar{f} + \bar{g}^T(P_t \otimes E_{0x})\bar{g}$$
$$-(v - \bar{f})^T(E_{0t} \otimes P_x)(v - \bar{f}) - (v - \bar{g})^T(P_t \otimes E_{0x})(v - \bar{g}). \qquad (4)$$

Since the matrices $E_{0x,t}$ and $E_{Nx,t}$ in (4) are positive semi–definite, v is bounded by the data $\bar{f}, \bar{g}$, just as in the continuous case. This implies that the numerical scheme is strongly stable. The estimate (4) is valid for any choice of timestep and hence (3) is unconditionally stable.

Furthermore, the terms on the left-hand side of (4) mimics the terms on the left–hand side of (2). Also, if $v = \bar{g}$ at $x = 0$ and $v = \bar{f}$ when $t = 0$, the right–hand side of (4) mimic the right–hand side of (2). In summary: the discrete energy estimate (4) mimics the continuous energy estimate (2) with additional damping terms that vanish with mesh–refinement. For more details on the SBP-SAT technique, the reader is referred to [17].

3. NRBC for the continuous problem

Consider a linear, hyperbolic system of equations in two space dimensions,

$$u_t + Au_x + Bu_y = 0, \quad x, y \in [0, 1],\ t \geq 0, \qquad (5)$$

where A and B are symmetric and constant matrices. Equations of the form (5) govern many problems in science and engineering, such as the linearized Euler and shallow water equations, as well as the Maxwell and elastic wave equations. We will henceforth consider B to be diagonal and non-singular. If B is not diagonal, one can always diagonalize it by an appropriate transformation.

We focus exclusively on non-reflecting properties of the boundaries at $y \in \{0, 1\}$, and therefore assume that (5) is periodic in x. With this assumption, we can apply a Fourier transform in the horizontal space direction and a Laplace transform in time (with zero initial data), such that (5) becomes

$$\hat{u}_y + C(s, \omega)\hat{u} = 0, \qquad (6)$$

where

$$C(s, \omega) = sB^{-1} + i\omega B^{-1}A, \qquad (7)$$

while s and $i\omega$ are the dual variables to t and x, respectively. To simplify the derivations below, we assume that $C(s,\omega)$ has distinct eigenvalues, and can therefore be diagonalized.

By inserting the ansatz $\hat{u} = e^{-ky}\psi(s,\omega)$ in (6), we get

$$(C(s,\omega) - kI)\psi(s,\omega) = 0. \tag{8}$$

Equation (8) has non-trivial solutions if $\det(C(s,\omega) - kI) = 0$. Hence, k is an eigenvalue of $C(s,\omega)$ and $\psi_j(s,\omega) \neq 0$ is the corresponding eigenvector. By solving the eigenvalue problem (8), the solution to (6) can be written

$$\hat{u}(y) = \sum_{j=1}^{n} \sigma_j e^{-\lambda_j(s,\omega)y}\psi_j(s,\omega) = \Psi(s,\omega)e^{-\Lambda_C(s,\omega)y}\bar{\sigma}, \tag{9}$$

where $e^{-\Lambda_C(s,\omega)y} = diag(e^{-\lambda_1(s,\omega)y}, e^{-\lambda_2(s,\omega)y}, ..., e^{-\lambda_n(s,\omega)y})$. Here, λ_j are the eigenvalues of $C(s,\omega)$; $\Psi = [\psi_1, ..., \psi_n]$ is the matrix of eigenvectors, $\bar{\sigma} = [\sigma_1, ..., \sigma_n]^T$ is the vector of coefficients and $\Lambda_C = diag(\lambda_1, \lambda_2..., \lambda_n)$. The coefficients σ_j, for $j = 1, ..., n$, will be determined by the boundary conditions at $y = \{0, 1\}$. Note that both the eigenvectors ψ_j and the eigenvalues λ_j depend on s and ω in general.

To construct the NRBC at $y = 0$, all decaying modes with $Re(\lambda_j) > 0$, must be removed, which means that the corresponding σ_j must be zero. Similarly, all increasing modes must vanish at $y = 1$. We need,

$$\begin{aligned} \sigma_j &= 0, \quad Re(\lambda_j) > 0; \quad y = 0 \\ \sigma_j &= 0, \quad Re(\lambda_j) < 0; \quad y = 1. \end{aligned} \tag{10}$$

The requirements (10) can be enforced by the boundary conditions

$$C^+(s,\omega)\hat{u}(0) = 0, \quad C^-(s,\omega)\hat{u}(1) = 0, \tag{11}$$

where $C^\pm(s,\omega) = \Psi\Lambda_C^\pm\Psi^{-1}$ in which $\Lambda_C^\pm$ denotes the part of Λ_C with positive and negative real parts, respectively, such that $\Lambda_C = \Lambda_C^+ + \Lambda_C^-$. The equivalence between (11) and (10) can be realized by inserting $\hat{u}(0)$ and $\hat{u}(1)$ on the form (9) in (11).

For the upcoming well-posedness analysis, the following lemma is needed.

Lemma 1. *Let $Re(s) = \eta > 0$. If B in (5) has l_+ positive and $l_- = n - l_+$ negative eigenvalues, then $C(s,\omega)$ in (6) has l_+ eigenvalues with positive real part and l_- eigenvalues with negative real part.*

Proof. Consider the eigenvalue problem

$$C(s,\omega)\psi_j = B^{-1}(sI + i\omega A)\psi_j = \lambda_j \psi_j. \tag{12}$$

By multiplying (12) with $\psi_j^* B$ from the left and adding the complex conjugate, one obtains

$$\eta|\psi_j|^2 = Re(\lambda_j)\psi_j^* B\psi_j,$$

where ψ_j^* is the hermitian conjugate of ψ_j and $s = \eta + i\xi$. Since $\eta > 0$ and $\psi_j \neq 0$, both $Re(\lambda_j) \neq 0$ and $\psi_j^* B\psi_j \neq 0$, which implies that the number of eigenvalues with positive and negative real parts is constant, and independent of ω. Phrased in another way: since $Re(\lambda_j) \neq 0$, for all ω, no eigenvalue can cross the imaginary axis.

The number of eigenvalues with positive and negative real part can now be determined by considering the simplified case $C(s, \omega = 0) = sB^{-1}$, which has l_+ eigenvalues with positive real part and l_- eigenvalues with negative real part for $\eta > 0$. Hence, $C(s, \omega)$ has l_+ eigenvalues with positive real part and l_- eigenvalues with negative real part for all ω and $\eta > 0$. $\qquad\square$

3.1. Well-posedness and energy estimates

To ensure well-posedness of (5) with the boundary conditions (11), the solution must must exist, be unique and the growth must be limited. We start by showing that the solution of (5) with boundary conditions (11) cannot grow. There is no growth if there are no solutions to (6) with the real part of the dual variable $s = \eta + i\xi$ being positive (see [5] for a similar analysis).

By multiplying (6) with $\hat{u}^* B$ from the left, integrating in y and adding the complex conjugate of the result, one obtains

$$2\eta \int_0^1 |\hat{u}|^2 dy = \hat{u}^*(0)B\hat{u}(0) - \hat{u}^*(1)B\hat{u}(1), \tag{13}$$

where we have used that $s + s^* = 2\eta$, $A = A^T$ and $|\hat{u}|^2 = \hat{u}^*\hat{u}$.

The matrix $C(s,\omega)$ can be diagonalized as $C(s,\omega) = \Psi\Lambda_C\Psi^{-1}$, and we partition $\Psi = [\tilde{\Psi}^+, \tilde{\Psi}^-]$, where $\tilde{\Psi}^\pm$ contain the eigenvectors corresponding to eigenvalues with positive and negative real part, respectively. According to Lemma 1, $\tilde{\Psi}^+$ is a $n \times l_+$ matrix and $\tilde{\Psi}^-$ is a $n \times l_-$ matrix, in which $l_\pm$ denotes the number of positive and negative eigenvalues of B. In the remainder of this paper, the square matrices

$$\Psi^+ = [\tilde{\Psi}^+, 0], \quad \Psi^- = [0, \tilde{\Psi}^-], \tag{14}$$

6

such that $\Psi = \Psi^+ + \Psi^-$, will be used. The vector $\bar{\sigma} = [\sigma_+, \sigma_-]^T$ is also partitioned as

$$\bar{\sigma}_+ = [\sigma_+, 0]^T, \quad \bar{\sigma}_- = [0, \sigma_-]^T,$$

where $\sigma_\pm$ are vectors of size l_+ and l_- corresponding to the growing and decaying modes, respectively.

By using (9), the solution $\hat{u}$ can be written as

$$\hat{u}(y) = \Psi e^{-\Lambda c y} \bar{\sigma} = \Psi^+ e^{-\Lambda c y} \bar{\sigma}_+ + \Psi^- e^{-\Lambda c y} \bar{\sigma}_-. \tag{15}$$

Next, the NRBC's in (11) leading to (10) are imposed. By letting $\bar{\sigma}_+ = 0$ at $y = 0$ and $\bar{\sigma}_- = 0$ at $y = 1$, (13) becomes

$$2\eta \int_0^1 |\hat{u}|^2 dy = \bar{\sigma}_-^* (\Psi^{-,*} B \Psi^-) \bar{\sigma}_- - (e^{-\Lambda c} \bar{\sigma}_+)^* (\Psi^{+,*} B \Psi^+)(e^{-\Lambda c} \bar{\sigma}_+). \tag{16}$$

To proceed, the following lemma is needed.

Lemma 2. *For $\eta > 0$, the matrices $\Psi^\pm$ in (14) have the properties*

$$rank(\Psi^{-,*} B \Psi^-) = l_-, \quad rank(\Psi^{+,*} B \Psi^+) = l_+,$$

independently of ω.

Proof. See Appendix A. $\square$

We are now ready to state the main result of this section.

Proposition 1. *The problem (5) with the boundary conditions (11) is well-posed.*

Proof. Assume that $\eta > 0$. For $\omega = 0$, the matrix $C(s, 0) = sB^{-1}$. The standard basis vectors are then the eigenvectors of C since B is diagonal. This implies that

$$\Psi^+(s, 0) = \begin{bmatrix} I_+ & 0 \\ 0 & 0 \end{bmatrix}, \quad \Psi^-(s, 0) = \begin{bmatrix} 0 & 0 \\ 0 & I_- \end{bmatrix},$$

where $I_\pm$ are identity matrices of size $l_\pm$, respectively. It then follows that

$$\Psi^{-,*}(s, 0) B \Psi^-(s, 0) = \begin{bmatrix} 0 & 0 \\ 0 & I_- \end{bmatrix} \begin{bmatrix} B^+ & 0 \\ 0 & B^- \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & I_- \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & B^- \end{bmatrix} \leq 0 \tag{17}$$

where $B^\pm$ is the positive and negative part of B, respectively. In (17), we have used that B^- and I_- have the same dimensions.

Since the rank of $\Psi^{-,*}B\Psi^-$ is constant for all $\eta > 0$ and all ω (according to Lemma 2), the number of eigenvalues on each side of the imaginary axis is constant. No eigenvalue can cross the origin ($\Psi^{-,*}B\Psi^-$ is hermitian, and its eigenvalues are therefore real) without lowering the rank of $\Psi^{-,*}B\Psi^-$, and thus violating Lemma 2. Consequently, if $\Psi^{-,*}B\Psi^- \leq 0$ for $\omega = 0$, then $\Psi^{-,*}B\Psi^- \leq 0$ for all ω, Hence, we must have

$$\Psi^{-,*}B\Psi^- \leq 0, \quad \forall \ \omega, \eta > 0. \tag{18}$$

By repeating the procedure above for $\Psi^{+,*}B\Psi^+$, one finds that

$$\Psi^{+,*}B\Psi^+ \geq 0, \quad \forall \ \omega, \eta > 0. \tag{19}$$

The relations (18) and (19) inserted into (16) now yields

$$2\eta \int_0^1 |\hat{u}|^2 dy = \bar{\sigma}_-^* (\Psi^{-,*}B\Psi^-)\bar{\sigma}_- - (e^{-\Lambda_c}\bar{\sigma}_+)^*(\Psi^{+,*}B\Psi^+)(e^{-\Lambda_c}\bar{\sigma}_+) \leq 0,$$

which means that there are no non-trivial solutions for $\eta > 0$. In summary, the boundary conditions (11) applied to (6) yields $\hat{u} = 0$ for any $\eta > 0$, which means that there are no growing solutions to (5).

Uniqueness follow directly by looking at the difference problem for (5), i.e. the same problem with another solution and the same data. Existence is given by the fact that we use the correct number of boundary conditions. Consequently, (5) with the boundary conditions (11) is well-posed. $\qquad \square$

Note that Proposition 1 is consistent with (2); the solution must be zero if the data is zero. Moreover, Proposition 1 holds for all systems of the form (5) with the NRBC's given by (11). This means that the exact NRBC to all two-dimensional hyperbolic systems of the form (5), that are periodic in x, results in a well-posed problem.

4. The semi-discrete problem

Consider (5) with zero initial data and a periodic solution in x. The SBP-SAT technique described in Section 2.1 is used to discretize (5) in the t and x-direction,

$$(\tilde{D}_t \otimes I_x \otimes I)v + (I_t \otimes D_x \otimes A)v + (I_t \otimes I_x \otimes B)v_y = 0 \tag{20}$$

where $D_x = -D_x^T$ approximates the derivative in x with periodic boundary conditions. Also, $\tilde{D}_t = P_t^{-1}(Q_t + E_{0t})$, where the added $E_{0,t} = diag(1, 0, ..., 0)$ comes from the penalty term in time with $\alpha_t = -1$ and $f = 0$, as in (4). Since D_x is skew-symmetric, it has purely imaginary eigenvalues and can be diagonalized with an orthonormal matrix X_x; that is, $D_x = iX_x\hat{\Omega}X_x^*$ where $\hat{\Omega} = diag(\hat{\omega}_1, \hat{\omega}_2, ..., \hat{\omega}_{N_x})$ is a diagonal matrix.

We make the following assumption, based on the results in [12]:

Assumption 1. *The matrix $\tilde{D}_t$ is diagonalizable and has eigenvalues with strictly positive real part.*

By Assumption 1, the matrix $\tilde{D}_t$ can be written as $\tilde{D}_t = X_t \hat{S} X_t^{-1}$ where $\hat{S} = diag(\hat{s}_1, \hat{s}_2, ..., \hat{s}_{N_t})$ with $Re(\hat{s}_j) > 0$. Multiplying (20) with $X_t^{-1} \otimes X_x^* \otimes B^{-1}$ from the left results in,

$$w_y + \hat{C}w = 0 \tag{21}$$

where $w = (X_t^{-1} \otimes X_x^* \otimes I)v$ and $\hat{C} = (\hat{S} \otimes I_x \otimes B^{-1}) + i(I_t \otimes \hat{\Omega} \otimes B^{-1}A)$.

Remark 1. *Note that (21) is similar to the continuous relation (6). This indicates that we might be able to re-use part of the analysis in Section 3.*

As in Section 3, the ansatz $w = e^{-\hat{k}y}\hat{\psi}$ is applied to (21), and results in

$$(\hat{C} - \hat{k}I)\hat{\psi} = 0,$$

which has non-trivial solutions $\hat{\psi}$ if $\det(\hat{C} - \hat{k}I) = 0$. Hence, $\hat{k}$ are the eigenvalues of $\hat{C}$ and $\hat{\psi}_j \neq 0$ the corresponding eigenvectors.

The solution to (21) can be written

$$w = \sum_{j=1}^{N} \sigma_j e^{-\hat{\lambda}_j y}\hat{\psi}_j = \hat{\Psi}e^{-\hat{\Lambda}_{\hat{C}}y}\hat{\sigma} \tag{22}$$

where $\hat{\Lambda}_{\hat{C}} = diag(\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_N)$ in which $\hat{\lambda}_j$ are the eigenvalues of $\hat{C}$. The matrix $\hat{\Psi} = [\hat{\psi}_1, \hat{\psi}_2..., \hat{\psi}_N]$ is the matrix of eigenvectors and $\hat{\sigma} = [\sigma_1, \sigma_2, ..., \sigma_N]^T$ are coefficients to be determined by the boundary conditions.

The boundaries at $y = \{0, 1\}$ produce no reflections if

$$\begin{aligned} \sigma_j &= 0, \quad Re(\hat{\lambda}_j) > 0, \quad y = 0, \\ \sigma_j &= 0, \quad Re(\hat{\lambda}_j) < 0, \quad y = 1. \end{aligned} \tag{23}$$

Just as in the continuous case, the requirements (23) is enforced by the boundary conditions

$$\hat{C}^+ w(0) = 0, \quad \hat{C}^- w(1) = 0, \tag{24}$$

where $\hat{C}^\pm = \hat{\Psi}\hat{\Lambda}_{\hat{C}}^\pm \hat{\Psi}^{-1}$ in which $\hat{\Lambda}_{\hat{C}}^\pm$ denotes the part of $\hat{\Lambda}_{\hat{C}}$ with positive and negative real part, respectively. Note the similarity between (24) and (11).

Before moving on to the stability analysis, we make some observations. The matrix $\hat{C}$ is block diagonal:

$$\hat{C} = \begin{bmatrix} C(\hat{s}_1, \hat{\omega}_1) & 0 & \dots & 0 \\ 0 & C(\hat{s}_1, \hat{\omega}_2) & \dots & \vdots \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & C(\hat{s}_{N_t}, \hat{\omega}_{N_x}) \end{bmatrix}, \tag{25}$$

where $C(\cdot, \cdot)$ is given by (7). Consequently, the matrix of eigenvectors has the form

$$\hat{\Psi} = diag(\Psi(\hat{s}_1, \hat{\omega}_1), \Psi(\hat{s}_2, \hat{\omega}_2), ..., \Psi(\hat{s}_{N_t}, \hat{\omega}_{N_x})) \tag{26}$$

where $\Psi(\cdot, \cdot)$ has the functional form of matrix of eigenvectors in (9).

We are now ready to state the following Lemma.

Lemma 3. *Let $\hat{\Psi}^\pm = diag(\Psi^\pm(\hat{s}_1, \hat{\omega}_1), \Psi^\pm(\hat{s}_1, \hat{\omega}_2), ..., \Psi^\pm(\hat{s}_{N_t}, \hat{\omega}_{N_x}))$ contain the eigenvectors of $\hat{C}$ corresponding to eigenvalues with positive and negative real part, respectively. We then have*

$$\hat{\Psi}^{+,*}\hat{B}\hat{\Psi}^+ \geq 0, \quad \hat{\Psi}^{-,*}\hat{B}\hat{\Psi}^- \leq 0$$

where $\hat{B} = I_t \otimes I_x \otimes B$.

Proof. The matrices $\hat{\Psi}^{\pm,*}\hat{B}\hat{\Psi}^\pm$ are block diagonal with the matrices $\Psi^{\pm,*}(\hat{s}_k, \hat{\omega}_l)B\Psi^\pm(\hat{s}_k, \hat{\omega}_l)$ on the diagonal. Hence, the claim follows directly from (18) and (19). □

4.1. Stability and energy estimates

Consider the semi-discrete formulation (21). The boundary conditions (24) are imposed by using lifting operators and penalty terms, such that

$$w_y + \hat{C}w = L_0 \left(\hat{B}^{-1}\Sigma_0 \hat{C}^+ w(0) \right) + L_1 \left(\hat{B}^{-1}\Sigma_1 \hat{C}^- w(1) \right) \tag{27}$$

10

approximates (5) with the boundary conditions (11). The matrices $\Sigma_{0,1}$ are penalty matrices to be determined such that (27) is stable. The lifting operators L_z in (27) are defined such that for two arbitrary functions ϕ, θ, we have $\int_0^1 \phi L_z(\theta) dy = \phi\theta|_{y=z}$.

Analogously to the technique in the continuous setting, (27) is multiplied with $w^*\hat{B}$ from the left, integrated in y and added to the complex conjugate of the result. One obtains

$$
\begin{aligned}
2\int_0^1 w^*(\hat{\eta} \otimes I_x \otimes I)w\,dy = {} & w^*(0)\left(\Sigma_0\hat{C}^+ + (\Sigma_0\hat{C}^+)^* + \hat{B}\right)w(0) \\
& + w^*(1)\left(\Sigma_1\hat{C}^- + (\Sigma_1\hat{C}^-)^* - \hat{B}\right)w(1),
\end{aligned}
\tag{28}
$$

where the eigenvalue matrix $\hat{S}$ has been divided into a real and imaginary part, $\hat{S} = \hat{\eta} + i\hat{\xi}$ in which $\hat{\eta}$ and $\hat{\xi}$ are real diagonal matrices. Recall that $\hat{\eta} > 0$ by Assumption 1, so that the left hand side of (28) is positive for $w \neq 0$.

By specific choices of the penalty matrices $\Sigma_{0,1}$, we will now show that the right–hand side of (28) is non–positive, such that there are no non-trivial solutions. To proceed we write $w(0)$ and $w(1)$ on the form (22), such that

$$
w(0) = \hat{\Psi}\hat{\sigma}, \quad w(1) = \hat{\Psi}e^{-\hat{\Lambda}\hat{c}}\hat{\sigma}
\tag{29}
$$

and make the ansatz

$$
\begin{aligned}
\Sigma_0 &= \hat{\Psi}^{-1,*}\left(\alpha\hat{\Psi}^{+,*} - \hat{\Psi}^{-,*}\right)\hat{B}\hat{C}^{-1}, \\
\Sigma_1 &= \hat{\Psi}^{-1,*}\left(\beta\hat{\Psi}^{-,*} + \hat{\Psi}^{+,*}\right)\hat{B}\hat{C}^{-1},
\end{aligned}
\tag{30}
$$

where α, β are real constants to be determined.

Inserting (29) and (30) into (28) yields

$$
\begin{aligned}
2\int_0^1 w^*(\hat{\eta} \otimes I_x \otimes I)w\,dy = {} & \hat{\sigma}\left(\hat{\Psi}^{-,*}\hat{B}\hat{\Psi}^- + (1 + 2\alpha)\hat{\Psi}^{+,*}\hat{B}\hat{\Psi}^+\right)\hat{\sigma} \\
& - (e^{-\hat{\Lambda}\hat{c}}\hat{\sigma})^*\left(\hat{\Psi}^{+,*}\hat{B}\hat{\Psi}^+ + (1 - 2\beta)\hat{\Psi}^{-,*}\hat{B}\hat{\Psi}^-\right)e^{-\hat{\Lambda}\hat{c}}\hat{\sigma}.
\end{aligned}
\tag{31}
$$

The right-hand side of (31) is non-positive if $\alpha \leq -1/2$ and $\beta \geq 1/2$, by the use of Lemma 3. In particular, if $\alpha = -\beta = -1/2$, we get

$$
2\int_0^1 w^*(\hat{\eta} \otimes I_x \otimes I)w\,dy = \hat{\sigma}\left(\hat{\Psi}^{-,*}\hat{B}\hat{\Psi}^-\right)\hat{\sigma} - (e^{-\hat{\Lambda}\hat{c}}\hat{\sigma})^*\left(\hat{\Psi}^{+,*}\hat{B}\hat{\Psi}^+\right)e^{-\hat{\Lambda}\hat{c}}\hat{\sigma},
\tag{32}
$$

11

which is analogous to the continuous energy estimate (16). Equation (32) implies that no non-trivial solutions exist, which proves stability. Note the clear connection to the continous problem in Section 3.

We summarize the results of this section in the following proposition.

Proposition 2. *The numerical scheme* (27) *is stable with* $\Sigma_{0,1}$ *chosen according to* (30) *with* $\alpha \leq -1/2$ *and* $\beta \geq 1/2$.

We also state the following corollary.

Corollary 1. *The numerical scheme* (27) *with the penalty matrices* (30) *is unconditionally stable.*

Proof. Proposition 2 holds for any time-discretization D_t, i.e. for any Δt. $\square$

The fully discrete SBP-SAT scheme with an initial data $\bar{f}$ and boundary operators (24) expressed in the original variable v can now be written,

$$
\begin{aligned}
(\tilde{D}_t \otimes I_x \otimes I_y \otimes I)v + (I_t \otimes D_x \otimes I_y \otimes A)v + (I_t \otimes I_x \otimes D_y \otimes B)v = \\
(I_t \otimes I_x \otimes P_y^{-1} E_{0y} \otimes I)\tilde{X}\tilde{\Sigma}_0\tilde{C}^+\tilde{X}^{-1}v_{y=0}+ \\
(I_t \otimes I_x \otimes P_y^{-1} E_{Ny} \otimes I)\tilde{X}\tilde{\Sigma}_1\tilde{C}^-\tilde{X}^{-1}v_{y=1}+ \\
(P_t^{-1} E_{0t} \otimes I_x \otimes I_y \otimes I)\bar{f}
\end{aligned}
\tag{33}
$$

where $D_y = P_y^{-1} Q_y$ is the SBP finite difference operator in the y-direction. In (33), $\tilde{X} = X_t \otimes X_x \otimes I_y \otimes I$ and $\tilde{C}^{\pm}$, $\tilde{\Sigma}_{0,1}$ are the boundary operators $\hat{C}^{\pm}$ and the penalty matrices $\Sigma_{0,1}$ injected at the appropriate grid points. Also, we have used the notation

$$
v_{y=0,1} = (I_t \otimes I_x \otimes E_{0,Ny} \otimes I)v, \quad v_{t=0} = (E_{0t} \otimes I_x \otimes I_y \otimes I)v.
$$

The matrices $\tilde{X}$, $\tilde{C}^{\pm}$ and $\tilde{\Sigma}_{0,1}$ are obtained numerically, as part of the numerical procedure.

5. Numerical experiments

To test the new NRBC's, we consider the problem (5) in the domain $\Omega = \{0 \leq x \leq 2, 0 \leq y \leq 1\}$ with

$$
A = \begin{bmatrix} \bar{u} & \bar{c}/\sqrt{2} & -\bar{c}/\sqrt{2} \\ \bar{c}/\sqrt{2} & \bar{u} & 0 \\ -\bar{c}/\sqrt{2} & 0 & \bar{u} \end{bmatrix}, \quad B = \begin{bmatrix} \bar{v} & 0 & 0 \\ 0 & \bar{v} - \bar{c} & 0 \\ 0 & 0 & \bar{v} + \bar{c} \end{bmatrix},
$$

which constitutes the linearized and symmetrized shallow water equations, without the Coriolis term. In the following numerical calculations, the reference state velocities are $\bar{u} = \bar{v} = 1/2$ and the gravity wave speed is $\bar{c} = 1$. As the initial condition, a Gaussian pulse centered around $(x_0, y_0) = (1, 0.7)$,

$$u_{1,2,3}(x, y, 0) = e^{-80\left((x-1)^2 + (y-0.7)^2\right)}$$

is used. A reference solution is created by performing the calculation on the domain $\Omega_{ref} = \{0 \leq x \leq 2, -0.5 \leq y \leq 1.5\}$, such that the physical reflections from the far field boundaries are not present at $y = \{0, 1\}$ for $t \leq 0.4$.

Let us define $B^{\pm}$ as the negative and positive parts of B, i.e.

$$B^+ = \begin{bmatrix} \bar{v} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \bar{v} + \bar{c} \end{bmatrix}, \quad B^- = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \bar{v} - \bar{c} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The first order classical approximate NRBC's [4] are

$$B^+ u(x, 0, t) = 0, \quad B^- u(x, 1, t) = 0. \tag{34}$$

Equation (5) together with the boundary conditions (34) result in a well–posed problem, and produce relatively small reflections. The performance of the new exact NRBC's will be compared to results obtained using (34).

In Figure 1, the reference solution is displayed at different time levels. We only show the results for the first component u_1; the results for $u_{2,3}$ are similar. In Figure 2, the error, i.e. the deviation from the reference solution, at different time levels is displayed when using the first order classical NRBC's (34). As one can see, the error increases significantly at $t = 0.4$, due to reflections at the boundary $y = 1$ when using (34); these reflections are not present when using the new exact NRBC's (24), as can be seen in Figure 3.

The calculations above are repeated for different mesh sizes using SBP schemes of second, third and fourth order overall accuracy, and the P-norm of the error, defined as $||e||_P^2 = e^T(P_x \otimes P_y)e$, at time $t = 0.4$ is computed. The results are summarized in Table 1 and 2. In Table 1, the approximate NRBC's (34) has been used. Note that the solution does not converge during mesh refinement. Applying the new exact NRBC (24) yields significantly smaller errors and the solution converges, as illustrated in Table 2. Moreover, the rate of convergence approach the overall accuracy of the scheme.
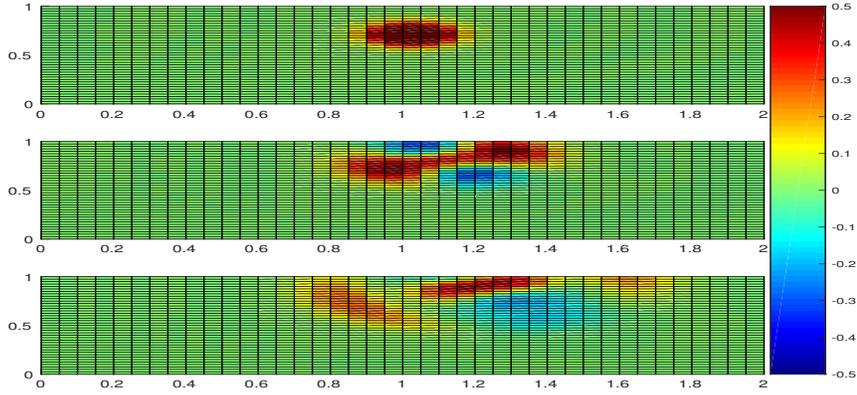
13

Figure 1: The reference solution u_1 at different time levels using the grid $\Delta x = 2/40$, $\Delta y = 1/40$. Upper: $t = 0$, middle: $t = 0.2$ and bottom: $t = 0.4$. A third order SBP scheme has been used.
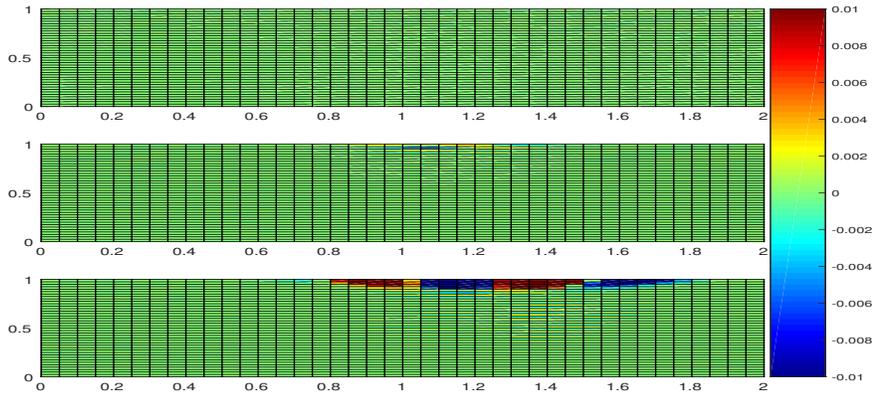

Figure 2: The error of the component u_1 at different time levels using the grid $\Delta x = 2/40$, $\Delta y = 1/40$ and the boundary conditions (34). Upper: $t = 0$, middle: $t = 0.2$ and bottom: $t = 0.4$. A third order SBP scheme has been used.

Figure 3: The error of the component u_1 at different time levels using the grid $\Delta x = 2/40$, $\Delta y = 1/40$ and the boundary conditions (24). Upper: $t = 0$, middle: $t = 0.2$ and bottom: $t = 0.4$. A third order SBP scheme has been used.
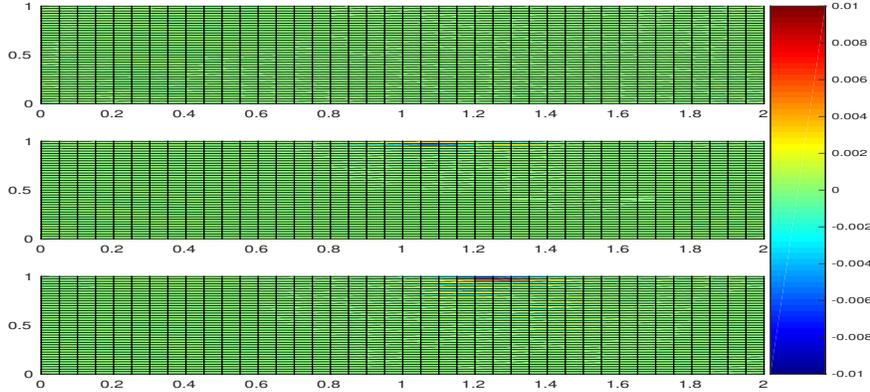
6. Summary and conclusions

A new approach for constructing NRBC's, based on SBP-SAT technique in time, has been investigated. The derived NRBC's results in an unconditionally stable scheme for linear hyperbolic problems.

Since the boundary operators are constructed in the discrete setting, they can be implemented directly. This new technique bypasses the complicated inverse Fourier-Laplace transform normally required when implementing standard NRBC's.

The new technique is exact, in the sense that it does not rely on approximations based on the angle of incidence or the size of the frequencies involved.

The method was applied to the linearized shallow water equations. Numerical experiments show that the numerical method is stable and accurate. Moreover, the derived boundary conditions produce very small reflections compared to approximate NRBC's, and converge with the correct rate.

N	SBP(2,1)	Rate	SBP(4,2)	Rate	SBP(6,3)	Rate
12	$4.34 \cdot 10^{-2}$	–	$7.26 \cdot 10^{-2}$	–	$1.32 \cdot 10^{-1}$	–
20	$4.92 \cdot 10^{-2}$	-0.2	$5.67 \cdot 10^{-2}$	0.5	$6.55 \cdot 10^{-2}$	1.4
30	$4.73 \cdot 10^{-2}$	0.1	$4.68 \cdot 10^{-2}$	0.5	$4.80 \cdot 10^{-2}$	0.8
40	$4.53 \cdot 10^{-2}$	0.2	$4.48 \cdot 10^{-2}$	0.2	$4.49 \cdot 10^{-2}$	0.2
50	$4.46 \cdot 10^{-2}$	0.1	$4.43 \cdot 10^{-2}$	0.1	$4.42 \cdot 10^{-2}$	0.1

Table 1: The P-norm of the error at $t = 0.4$ for different mesh-sizes when using a second (SBP(2,1)), third (SBP(4,2)) and fourth (SBP(6,3)) order SBP scheme and the approximate NRBC's (34).

N	SBP(2,1)	Rate	SBP(4,2)	Rate	SBP(6,3)	Rate
12	$2.79 \cdot 10^{-2}$	–	$5.57 \cdot 10^{-2}$	–	$1.41 \cdot 10^{-1}$	–
20	$1.55 \cdot 10^{-2}$	1.2	$1.95 \cdot 10^{-2}$	2.1	$4.02 \cdot 10^{-2}$	2.5
30	$8.00 \cdot 10^{-3}$	1.6	$6.64 \cdot 10^{-3}$	2.7	$1.13 \cdot 10^{-2}$	3.1
40	$4.56 \cdot 10^{-3}$	2.0	$3.13 \cdot 10^{-3}$	2.6	$3.56 \cdot 10^{-3}$	4.0
50	$2.92 \cdot 10^{-3}$	2.0	$1.39 \cdot 10^{-3}$	3.6	$1.28 \cdot 10^{-3}$	4.6

Table 2: The P-norm of the error at $t = 0.4$ for different mesh-sizes when using a second (SBP(2,1)), third (SBP(4,2)) and fourth (SBP(6,3)) order SBP scheme and the exact NRBC's (24).

Appendix A. Proof of Lemma 2

We will only prove the statement for the term $\Psi^{-,*}B\Psi^-$, since the proof for the term $\Psi^{+,*}B\Psi^+$ is analogous. First, we note that $\Psi^- = \Psi I^-$ where

$$I^- = \begin{bmatrix} 0 & 0 \\ 0 & I_{l_-} \end{bmatrix}$$

in which I_{l_-} is an identity matrix of size l_-. Note that, due to Lemma 1, I_{l_-} have the same dimensions as I_- in (17), i.e. $I_{l_-} = I_-$. Also, recall that Ψ diagonalizes $C = sB^{-1} + i\omega B^{-1}A$; that is, $C\Psi = \Psi\Lambda_C$. We have,

$$rank(\Psi^{-,*}B\Psi^-) = rank(\Psi^{-,*}B\Psi I^-\Lambda_C) =$$
$$rank(\Psi^{-,*}BC\Psi I^-) = rank(\Psi^{-,*}(sI + i\omega A)\Psi^-).$$

In the first step, we have multiplied the expression with the non-singular matrix Λ_C from the right. This can be done without altering the rank, as Λ_C is non-singular due to Lemma 1.

Next, let the vector $x \neq 0$ satisfy $\Psi^{-,*}(sI + i\omega A)\Psi^- x = 0$, and let $s = \eta + i\xi$ where $\eta > 0$. We then have

$$0 = x^*\Psi^{-,*}(sI + i\omega A)\Psi^- x = \eta||\Psi^- x||_2^2 + i(\xi||\Psi^- x||_2^2 + \omega(\Psi^- x)^*A(\Psi^- x)).$$
$$(A.1)$$

Note that $||\Psi^- x||_2^2$ and $(\Psi^- x)^*A(\Psi^- x)$ are real (since A is symmetric). Accordingly, (A.1) can only be satisfied if $||\Psi^- x||_2 = 0$ (the real part of (A.1) will be non-zero otherwise). This means that $\Psi^{-,*}(sI + i\omega A)\Psi^- x = 0$ if, and only if, $\Psi^- x = 0$. In other words,

$$ker(\Psi^{-,*}(sI + i\omega A)\Psi^-) = ker(\Psi^-)$$

where $ker(\cdot)$ denotes the null-space. This means that

$$rank(\Psi^{-,*}(sI + i\omega A)\Psi^-) = rank(\Psi^-) = l_-.$$

Hence, we have

$$rank(\Psi^{-,*}B\Psi^-) = rank(\Psi^{-,*}(sI + i\omega A)\Psi^-) = rank(\Psi^-) = l_-,$$

which completes the proof.

References

[1] J.P Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200, 1994.

[2] M. Carpenter, D. Gottlieb, and S. Aberbanel. Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: methodology and application to high-order compact schemes. *Journal of Computational Physics*, 111:220–236, 1994.

[3] M. Carpenter, J. Nordström, and D. Gottlieb. A stable and conservative interface treatment of arbitrary spatial accuracy. *Journal of Computational Physics*, 148:341–365, 1999.

[4] B. Engquist and A. Majda. Absorbing boundary conditions for the numerical simulation of waves. *Mathematics of computation*, 31:629–651, 1977.

[5] S. Eriksson and J. Nordström. Exact non-reflecting boundary conditions revisted: well-posedness and stability. *Foundations of Computational Mathematics*, February 2016. doi: 10.1007/s10208-016-9310-3.

[6] D. Givoli. High-order local non-reflecting boundary conditions: a review. *Wave Motion*, 39:319–326, 2004.

[7] B. Gustafsson, H. Kreiss, and A. Sundström. Stability theory of difference approximations for mixed initial boundary value problems II. *Mathematics of Computation*, 26:649–686, 1972.

[8] B. Gustafsson, H.O Kreiss, and J. Oliger. *Time dependent problems and difference methods.* John Wiley & Sons, 1995.

[9] T. Hagstrom. Radiation boundary conditions for the numerical simulation of waves. *Acta Numerica*, 8:47–106, 1999.

[10] T. Lundquist and J. Nordström. The SBP-SAT technique for initial value problems. *Journal of Computational Physics*, 270:86–104, 2014.

[11] K. Mattsson and J. Nordström. Summation by parts for finite difference approximations of second derivatives. *Journal of Computational Physics*, 199:503–540, 2004.

[12] J. Nordström and T. Lundquist. Summation-by-parts in time. *Journal of Computational Physics*, 251:487–499, 2013.

[13] J. Nordström and T. Lundquist. Summation-by-parts in time: the second derivative. *SIAM Journal of Scientific Computing*, 38:A1561–A1586, 2016.

[14] J. Nordström and M. Svärd. Well-posed boundary conditions for the Navier-Stokes equations. *SIAM Journal of Numerical Analysis*, 43:1231–1255, 2005.

[15] C.W. Rowley and T. Colonius. Discretely non-reflecting boundary conditions for linear hyperbolic systems. *Journal of Computational Physics*, 157:500–538, 2000.

[16] B. Strand. Summation by parts for finite difference approximations for $\frac{d}{dx}$. *Journal of Computational Physics*, 110:47–67, 1994.

[17] M. Svärd and J. Nordström. Review of summation-by-parts schemes for initial-boundary-value problems. *Journal of Computational Physics*, 268:17–38, 2014.

[18] S. Tsynkov. Numerical solution of problems on unbounded domains. A review. *Applied Numerical Mathematics*, 27:465–532, 1998.