On the Order of Accuracy for Difference Approximations of Initial–Boundary Value Problems

Magnus Svärd and Jan Nordström

23rd September 2004

Abstract

Finite difference approximations of the second derivative in space appearing in, parabolic, incompletely parabolic systems of, and second order hyperbolic, partial differential equations are considered. If the solution is pointwise bounded, we prove that finite difference approximations of those classes of equations can be closed with two orders less accuracy at the boundary without reducing the global order of accuracy.

This result is generalised to initial-boundary value problems with an mth order principal part. Then, the boundary accuracy can be lowered m orders.

Further, it is shown that summation-by-parts operators with approximating second derivatives are pointwise bounded. Linear and nonlinear computations corroborates the theoretical results.

1 Introduction

For computations of numerical solutions to an initial-boundary value problems, it is commonly known that one order less accuracy at the boundary is allowed. This stems from two articles by Gustafsson, [1, 2], and refers to the order of accuracy of the numerical boundary conditions. The physical boundary conditions have to be approximated to the global order of accuracy. Also, in [2] it was shown that 2 orders is recovered at the boundary if Dirichlet boundary conditions are used and a number of algebraic conditions are satisfied.

Abarbanel et al. showed in [3] that 1.5 orders of accuracy can be recovered theoretically at the boundary for parabolic problems with general

boundary conditions. They present computations where two orders of accuracy is recovered, indicating that their theoretical estimate is not sharp.

In [4] Mattsson and Nordström suggested that for parabolic problems as well as incompletely-parabolic problems, the numerical boundary conditions (or numerical closure) can be approximated with two orders less accaracy for parabolic terms. Further, the physical boundary conditions are allowed to be approximated with one order less accuracy when the boundary conditions are weakly implemented. These conclusions are supported with extensive numerical experiments and an analysis giving conditions for the hypothesis to be true. However, the conditions derived are algebraically difficult to evaluate for the actual numerical scheme.

In this article we consider parabolic, as well as incompletely parabolic systems of partial differential equations (PDE:s) with general boundary conditions. We prove that two orders less accuracy is allowed for the approximation of second derivatives at the boundary, if the scheme yields a pointwise bounded solution. It is also proven that the results carry over to discretisations of second order hyperbolic equations, such as the wave equation.

The theory is also taken one step further by considering equations with an mth order principal part. Then the order of accuracy for numerical boundary conditions can be lowered m orders if the scheme is pointwise stable.

The article is organised as follows: in Section 2, accuracy theorems are proven under specific stability assumptions; Section 3 proves that the theorems are applicable to Summation-by-Parts operators (SBP operators) with the Simultaneous Approximation Term technique (SAT) approximating the boundary conditions; in Section 4 computations that corroborate the theoretical results are presented.

2   Analysis

The focus in this paper will be discretisations near the boundary. To simplify the notation we consider, without restriction, semi-infinite problems in space.

2.1   The Advection-Diffusion Equation

Consider the parabolic equation,

$$
\begin{aligned}
u_t + a u_x &= \epsilon u_{xx} + F(x,t), \quad 0 \leq x \leq \infty, \quad t \geq t_0, \\
u(0,t) + \alpha u_x(0,t) &= g(t), \\
|u| &\to 0, \qquad x \to \infty \\
u(x,t_0) &= f(x),
\end{aligned}
\tag{1}
$$

where $\epsilon > 0$ and $\| \cdot \|$ denotes some norm; f is the intital data; g is the boundary data and F is the forcing function. With $a > 0$, the energy method applied to (1) leads to an energy estimate if,

$$-\frac{2\epsilon}{a} \leq \alpha \leq 0. \tag{2}$$

With the above condition satisfied, equation (1) is well posed, i.e. it has a unique and bounded solution.

A general semidiscretisation of (1), with grid spacing h, would be,

$$\begin{aligned} v_t &= M_h v + B_h, \\ v(0) &= f \end{aligned} \tag{3}$$

where M_h is the discretisation operator and B_h includes the boundary data and the forcing function. Further, v is the vector function approximating the solution of (1) and f is the vector function identical to $f(x)$ at the grid points. Note that, the general formulation (3) covers both weak and strong implementation of boundary conditions.

Next, we define and discuss a few notions that frequently will be used. Let $\| \cdot \|_h$ denote the l^2-norm, i.e. $\|v\|_h^2 = hv^T v$. In [5] the following definition is given.

Definition 2.1 *The approximation, v, is* strongly stable *if, for all $h \leq h_0$, the estimate*

$$\|v(t)\|_h^2 \leq K(t)(\|f\|_h^2 + \max_{0 \leq \tau \leq t} \|F(\tau)\|_h^2 + \max_{0 \leq \tau \leq t} g(\tau)^2) \tag{4}$$

holds. Here $K(t)$ is a bounded function in any finite time interval and does not depend on the data.

With the norm $\|v\|_\infty = sup_i \, v_i$ we modify the previous definition.

Definition 2.2 *The approximation, v, is* strongly pointwise stable *if, for all $h \leq h_0$, the estimate*

$$\|v(t)\|_\infty^2 \leq K(t)(\|f\|^2 + \max_{0 \leq \tau \leq t} \|F(\tau)\|^2 + \max_{0 \leq \tau \leq t} g(\tau)^2) \tag{5}$$

holds. Here $K(t)$ is a bounded function in any finite time interval and does not depend on the data. ($\| \cdot \|$ denotes some norm.)

We also define the space l^∞ as the space of all grid functions f with the property that $\|f\|_\infty$ is bounded.

Remark We call a scheme stable or pointwise stable, if the problem with $g(t) = 0$ fulfils the estimate (4) or (5).

Lemma 2.3 *Assume that F, f and g are smooth such that the solution u of (1) is smooth. Let v denote the solution to the consistent discretisation (3) of (1) with grid spacing h. Let u_h denote the projection of the exact solution onto the grid. If v is pointwise stable, for all $h \leq h_0$, v converges to u_h uniformly.*

Proof Insert u_h into (3) to obtain,

$$
\begin{aligned}
(u_h)_t &= M_h u_h + B_h + T_h, \\
u_h(0) &= f,
\end{aligned}
$$

where T_h denotes the truncation error vector. Using (3) we obtain,

$$
\begin{aligned}
(u_h - v)_t &= M_h(u_h - v) + T_h, \\
(u_h - v)(0) &= 0.
\end{aligned}
$$

Since the scheme is pointwise stable we have the estimate,

$$
\|u_h - v\|_\infty \leq K(t)(\sup_{0 \leq \tau \leq t} \|T_h(\tau)\|_\infty^2). \tag{6}
$$

By consistency and smoothness of u, $\|T_h(\tau)\|_\infty^2 \to 0$ as $h \to 0$. Thus, we have uniform convergence. ∎

Remark The Lemma is true for any discretisation of a well posed partial differential equation, if it is pointwise stable and the solution is smooth. Not just for parabolic partial differential equations.

Lemma 2.4 *Assume that v is stable in some norm, i.e. the estimate (4) holds when $g = 0$ for a specific norm. Then v is uniquely defined in that norm.*

Proof Assume that there exist two solutions w and v to equation (3). By linearity we have the error equation, $(v-w)_t = M_h(v-w)$, with $(v-w)(0) = 0$ and the bound $\|v - w\| < 0$ for $h \leq h_0$ follows. ∎

Remark Note that if the norm is weak, i.e. the continuous norm can be zero if the function is nonzero on a set of Lebesgue measure zero, then Lemma 2.4 does not imply convergence pointwise. Still, there can only be one solution to the numerical problem in the sense of that norm. Further, for $h > 0$ the numerical solution is pointwise unique, since there can be no set of measure 0.

4

Lemma 2.5 *If v is bounded in l^∞, then v converges uniformly and uniquely to u, in the sense of $\|u_h - v\|_\infty \to 0$ as $h \to 0$.*

Proof Lemma 2.3 and Lemma 2.4.

To analyse the order of accuracy we shift our focus to consider the error equation by subtracting the true solution, $u(x,t)$ from v, i.e. $e = v - u_h$. Using either a strong or weak approximation of the boundary conditions we would arrive at,

$$
\begin{aligned}
e_t &= M_h e + T_h, \\
e(0) &= 0.
\end{aligned}
\tag{7}
$$

As before, T_h denotes the truncation error and generally we have,

$$
T = (\mathcal{O}(h^r), ..., \mathcal{O}(h^r), \mathcal{O}(h^{2p}), ..., \mathcal{O}(h^{2p}), \mathcal{O}(h^r), ..., \mathcal{O}(h^r))^T,
\tag{8}
$$

where h denotes the grid spacing. To describe the size and structure of T_h, we will use $T_h = \mathcal{O}(h^r, h^{2p})$ for boundary and interior points respectively. If (3) is stable and $r = 2p$ we immediately obtain the desired order of accuracy $2p$ of the scheme by applying the energy method (See proof of Lemma 2.3, where the norm may be different from the supremum norm.). However, we will consider $r < 2p$. The first theorem below states that two orders less accuracy is allowed on the boundary in the purely parabolic case, $a = 0$.

Theorem 2.6 *If v is a pointwise stable discretisation of (3) for $h \le h_0$ and $a = 0$, then with $r = 2p - 2$, the global order of accuary of the approximation (3) is $2p$.*

Proof We follow a technique presented in [5] and also used in [4], and split the truncation error into a boundary and internal part,

$$
T_i = [0, ..., 0, \mathcal{O}(h^{2p}), ..., \mathcal{O}(h^{2p}), 0, ..., 0]^T = \mathcal{O}(0, h^{2p}),
\tag{9}
$$
$$
T_b = [\mathcal{O}(h^r), ..., \mathcal{O}(h^r), 0, ..., 0, \mathcal{O}(h^r), ..., \mathcal{O}(h^r)]^T = \mathcal{O}(h^r, 0),
\tag{10}
$$

such that $T = T_i + T_b$. Similarly, the error is split into $e = e_i + e_b$. Note that e_i and e_b are both nonzero everywhere since there is in general a strong coupling between the boundaries and the interior. By the boundedness in l^∞ of v, and since e_i is discretised with the same scheme as v, e_i satisafies the same estimate, such that,

$$
\|e_i(t)\|_\infty \le K(t)\|T_i(t)\|_\infty \le \mathcal{O}(h^{2p}).
\tag{11}
$$

5

Next, we turn to the boundary part. Laplace transform (7) and consider only errors coming from the discretisation at the boundary,

$$s\hat{e}_b = M_h\hat{e}_b + \hat{T}_b, \quad Re\, s \geq 0. \tag{12}$$

In the purely parabolic case all the entries of M are proportional to $1/h^2$. Thus, we multiply by h^2 such that $\tilde{M} = h^2 M_h$ to make every nonzero entry of $\tilde{M}$ of order $\mathcal{O}(1)$. With $\tilde{s} = sh^2$ we obtain,

$$\tilde{s}\hat{e}_b = \tilde{M}\hat{e}_b + h^2\hat{T}_b, \quad Re\, \tilde{s} \geq 0. \tag{13}$$

Note that, the scheme is the same at every point except at points near the boundary. We consider (13) to be a homogeneous difference equation where $h^2\hat{T}_b$ is its initial data. We write the solution to (13) as,

$$(\hat{e}_b)_j = \sum_{l=1}^{2p} \sigma_l \kappa_l^j. \tag{14}$$

Assume without loss of generality that the interior scheme is $2p = k + q + 1$ points wide. Since $(\hat{T}_b)_j = 0$ at an interior point j we have,

$$\tilde{s}(\hat{e}_b)_j = \sum_{i=-k}^{q} \alpha_i (\hat{e}_b)_{i+j}, \tag{15}$$

where α_i are constants. Inserting the ansatz (14) into (15) yields the characteristic equation,

$$\tilde{s}\kappa^j = \sum_{i=-k}^{q} \alpha_i \kappa^{i+j}, \tag{16}$$

which has solutions $\kappa_l(\tilde{s})$ for $l = 1..2p$. Denote by $\kappa_1, .., \kappa_m$ the roots with $|\kappa_i| \leq 1$ for $i = 1..m$. The remaining roots are discarded due to boundedness of the solution. That is $\sigma_{m+1} = ... = \sigma_{2p} = 0$. Hence, the solution reads,

$$(\hat{e}_b)_j = \sum_{l=1}^{m} \sigma_l \kappa_l^j. \tag{17}$$

We have yet not used the constants σ_l, $l = 1..m$. Those constants are determined by the scheme near the boundary.

Define $\sigma = (\sigma_1, ..., \sigma_m)^T$ such that,

$$\bar{\kappa}\sigma = \hat{e}_{rb}, \tag{18}$$

where $\hat{e}_{rb}$ now denotes the restriction of $\hat{e}_b$ to the $\nu + 1$ boundary points and,

$$\bar{\kappa} = \begin{pmatrix} \kappa_1^0 & \ldots & \kappa_m^0 \\ \vdots & \ldots & \vdots \\ \kappa_1^\nu & \ldots & \kappa_m^\nu \end{pmatrix} \tag{19}$$

We will use $\bar{\kappa}$ to determine σ which is why we exclude the interior points since (13) is already fulfilled at the interior points by the κ_l:s, independent of σ. Let I_r denote the $(\nu + 1) \times (\nu + 1)$ identity matrix. At the $\nu + 1$ boundary points where the interior scheme is altered we obtain,

$$(\tilde{s} I_r - \tilde{M}_r) \bar{\kappa} \sigma = h^2 \hat{T}_{rb}, \tag{20}$$

where M_r and $\hat{T}_{rb}$ denotes the restrictions to the $(\nu + 1)$ boundary points. To estimate σ we note again that $\tilde{M}_r$ is a matrix with coefficients independent of h and $\tilde{s}$. We have,

$$(\tilde{s} \bar{\kappa} - \tilde{M}_r \bar{\kappa}) \sigma = h^2 \hat{T}_{rb}, \tag{21}$$

where the coeffiecients of $R = (\tilde{s} \bar{\kappa} - \tilde{M} \bar{\kappa})$ are independent of h. Thus if a unique solution to (21) exists, σ would be of order $h^2 \hat{T}_{rb}$, i.e. we would gain two orders of accuracy at the boundary. Then by Parseval's relation we can transform back to e to conclude that the desired order of accuracy is obtained. However, R is in general non-square and the system seems overdetermined. (If it is underdetermined, more numerical boundary conditions need to be supplied.) We need to prove that (21) has a solution for all $Re\,\tilde{s} \geq 0$, i.e. that the overdetermined system in fact has a number of linearly dependent equations and that the remaining system is nonsingular for $Re\,\tilde{s} \geq 0$.

By well posedness, the exact continuous solution is unique. From pointwise stability of the numerical scheme and Lemma 2.5, v converges uniquely and pointwise to u. The same properties carries over to e and e_i and they will converge uniquely and pointwise to 0. Hence, $e_b = e - e_i$ is unique.

Suppose σ is not uniquely determined by (21) then $\hat{e}_b$ would not be unique. However, since e and e_i are bounded, $\hat{e}_b$ has to be bounded and the inverse Laplace transform could be performed and yield a nonunique e_b. A contradiction. ∎

Remark Note the necessity to split the errors into two parts referring to boundary points and interior points respectively. If this is not done it is impossible to determine the κ_i:s from the interior scheme since the right hand side is nonzero. We would end up with, $(sI - M_h)e = \hat{T}$ for the whole scheme. Although, the uniqueness of e means that $(sI - M)$ is nonsingular even when multiplied by h^2 it does not provide information about the size of e since the number of entries of M increases as $h \to 0$.

This was the purely parabolic case, i.e. $a = 0$ in (1). Next, we want to add a low order term, that is $a \neq 0$ in (1), and still recover the same accuracy result. We need the following Lemma.

Lemma 2.7 *If A is an invertible matrix and E a matrix, then $A + E$ will be invertible if $\rho(A^{-1}E) < 1$, where $\rho(\cdot)$ denotes the spectral radius, and*

$$(A + E)^{-1} = A^{-1} - \sum_{k=1}^{\infty} (-1)^{k+1} (A^{-1}E)^k A^{-1}. \tag{22}$$

Proof See [6]. ∎

The main difference compared to the purely parabolic case is that $\tilde{M}$ will not be a constant matrix but rather, $\tilde{M} = A + Bh$, where A, B are constant matrices. A results from the discretisation of the second derivatives and B from first derivatives. These perturbations follows through the whole proof such that the elements of $M_r \sim \mathcal{O}(1 + h)$ and hence $\kappa_l \sim \mathcal{O}(1 + h)$, and we end up with equation (21) where,

$$(\tilde{s}\bar{\kappa} - \tilde{M}_r \kappa) \sim \mathcal{O}(1 + h). \tag{23}$$

The same reasoning applies and we conclude that also in this case (23) can be reduced to a square nonsingular matrix. By Lemma 2.7 the inverse would be of order $1 + h$ and the desired size of σ is obtained. This result is stated in the following theorem.

Theorem 2.8 *If (3) is a pointwise stable discretisation of (1) for $h \leq h_0$, then with the order of accuracy $r = 2p - 2$ at the boundary, the global order of accuracy of the approximation (3) is $2p$.*

Remark The truncation errors, T_b, include errors from all terms in Theorem 2.8. That means that it is allowed for the hyperbolic terms to be 2 orders less accurate at the boundary as long as parabolic terms are present.

Remark In the case using a weak implementation of the boundary conditions such as the Simultaneous Approximation Term technique (SAT, [7]) to approximate the boundary conditions a penalty term scaled by h^{-1} is used. Thus, when multiplied by h^2 we can allow the physical boundary conditions to be approximated with one order of accuracy less than global accuracy.

Remark If a system of parabolic equations are discretised with a pointwise stable scheme the same theorem applies also. We omit the proof here. It is precisely the same though with a more involved notation.

As a concluding observation, note that equation (13) can be written as,

$$(\tilde{s}I - \tilde{M})\hat{e}_b = h^2 \hat{T}_b, \tag{24}$$

and that Theorem 2.8 implies that $(\tilde{s}I - \tilde{M})^{-1}$ exists and is of order 1. In [5] the following definition is introduced which we will need below.

Definition 2.9 *If* $det(\tilde{s}I - \tilde{M}) \neq 0$ *for* $Re\,\tilde{s} \geq 0$ *in, equation (24). Then* $(\tilde{s}I - \tilde{M})^{-1}$ *exists and we say that the* determinant condition *is satisfied.*

2.2 Incompletely Parabolic Systems

Consider the discretisation,

$$
\begin{aligned}
v_t + a_{11}D_{11}v &= B, \quad a_{11} > 0, \\
v(0) &= f,
\end{aligned}
\tag{25}
$$

of,

$$
\begin{aligned}
u_t + a_{11}u_x &= 0, \quad 0 \leq x < \infty \\
u(0,t) &= g(t), \\
u(x,0) &= f(x),
\end{aligned}
\tag{26}
$$

where B holds the boundary data and forcing function. Suppose that the determinant condition for (25) holds, such that, for some constant $\delta > 0$,

$$|(\tilde{s}I + a_{11}\tilde{D}_{11})^{-1}| > \delta, \;\; Re\,\tilde{s} \geq 0. \tag{27}$$

The tilde denotes the undivided difference such that $hD_{11} = \tilde{D}_{11}$.

Consider the following incompletely parabolic system,

$$
\begin{pmatrix} u^{(1)} \\ u^{(2)} \end{pmatrix}_t + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} u^{(1)} \\ u^{(2)} \end{pmatrix}_x = \begin{pmatrix} 0 \\ \epsilon u^{(2)} \end{pmatrix}_{xx}, x \geq 0, t \geq 0,
$$
$$
\begin{aligned}
L_0(t)\mathbf{u} &= g_0(t), \\
\mathbf{u}(x,0) &= f(x),
\end{aligned}
\tag{28}
$$

where $\mathbf{u} = (u^{(1)}, u^{(2)})^T$. Let equation (28) be discretised by,

$$
\begin{pmatrix} v^{(1)} \\ v^{(2)} \end{pmatrix}_t + \begin{pmatrix} a_{11}D_{11} & a_{12}D_{12} \\ a_{21}D_{21} & a_{22}D_{22} - \epsilon D_2 \end{pmatrix} \begin{pmatrix} v^{(1)} \\ v^{(2)} \end{pmatrix} = \begin{pmatrix} B^{(1)} \\ B^{(2)} \end{pmatrix}, \tag{29}
$$

9

where $B^{(1)}$ and $B^{(2)}$ are vectors that introduce the boundary data. Further, $v^{(1)}$ and $v^{(2)}$ are the discrete solution vectors. With the splitting of the error $e = e_i + e_b$ and the truncation error $T = T_i + T_b$ we obtain for e_b,

$$
\begin{pmatrix} e_b^{(1)} \\ e_b^{(2)} \end{pmatrix}_t + \begin{pmatrix} a_{11}D_{11} & a_{12}D_{12} \\ a_{21}D_{21} & a_{22}D_{22} - \epsilon D_2 \end{pmatrix} \begin{pmatrix} e_b^{(1)} \\ e_b^{(2)} \end{pmatrix} = \begin{pmatrix} T_b^{(1)} \\ T_b^{(2)} \end{pmatrix}, \qquad (30)
$$

where $T_b^{(1)} = \mathcal{O}(h^r, 0)$ and $T_b^{(2)} = \mathcal{O}(h^q, 0)$.

Remark Note that D_{11}, D_{12}, D_{21} and D_{22} are not necessarily pure first derivative approximations but can include terms from the boundary treatment. The same is true for D_2 wich is mainly an approximation of the second derivative.

Below, we state and prove a theorem based on the following conditions.

Condition 2.10 *Assume that the discretisation (29) of (28) is pointwise stable.*

Condition 2.11 *Assume that the discretisation (29) of (28) is stable and, with $a_{11} = a_{12} = a_{21} = a_{22} = 0$, fulfills Theorem 2.8.*

Theorem 2.12 *Assume that the discretisation (25) of (26) satisfies the determinant condition (27). If either Condition 2.10 or Condition 2.11 is satisafied and, D_{11} and D_{12} are closed with $r = 2p - 1$ whereas D_{21}, D_{22} and D_2 are closed with $q = 2p - 2$, then (29) is of order $2p$.*

Proof Laplace transform (30) to obtain,

$$
\begin{pmatrix} sI + a_{11}D_{11} & a_{12}D_{12} \\ a_{21}D_{21} & sI + a_{22}D_{22} - \epsilon D_{22} \end{pmatrix} \begin{pmatrix} \hat{e}_b^{(1)} \\ \hat{e}_b^{(2)} \end{pmatrix} = \begin{pmatrix} \hat{T}_b^{(1)} \\ \hat{T}_b^{(2)} \end{pmatrix}, \qquad (31)
$$

or,

$$
A\hat{e} = \hat{T}_b, \quad Re\,\tilde{s} \geq 0. \qquad (32)
$$

Rotate equation (32) using,

$$
R = \begin{pmatrix} I & \alpha \\ 0 & I \end{pmatrix}, \quad B = \begin{pmatrix} I & 0 \\ \beta & I \end{pmatrix} \qquad (33)
$$

such that,

$$
BARR^{-1}\hat{e} = B\hat{T}_b. \qquad (34)
$$

10

To make BAR block diagonal we choose $\alpha = -h(\tilde{s}I + a_{11}\tilde{D}_{11})^{-1}a_{12}D_{12} = -(\tilde{s}I + a_{11}\tilde{D}_{11})^{-1}a_{12}\tilde{D}_{12}$ and $\beta = -(\tilde{s}I + a_{11}\tilde{D}_{11})^{-1}a_{21}\tilde{D}_{21}$. By assumption, $(\tilde{s}I + a_{11}\tilde{D}_{11})^{-1}$ exists. Thus, α and β are of order 1. The matrices R and B are both non-singular justifying the transformation. Further,

$$R^{-1} = \begin{pmatrix} I & -\alpha \\ 0 & I \end{pmatrix}, BT_b = \begin{pmatrix} \hat{T}_b^{(1)} \\ \hat{T}_b^{(2)\prime} \end{pmatrix} \tag{35}$$

where $\hat{T}_b^{(2)\prime} = \hat{T}_b^{(2)} + \beta\hat{T}_b^{(1)}$. Note that $B\hat{T}_b$ is of the same size as T_b. Furthermore,

$$R^{-1}\hat{e}_b = \begin{pmatrix} \hat{e}_b^{(1)} - \alpha\hat{e}_b^{(2)} \\ \hat{e}_b^{(2)} \end{pmatrix} = \begin{pmatrix} \hat{e}_b^{(1)\prime} \\ \hat{e}_b^{(2)} \end{pmatrix} = \hat{e}_b' \tag{36}$$

Finally,

$$BAR = \begin{pmatrix} \frac{1}{h}(\tilde{s}I + a_{11}\tilde{D}_{11}) & 0 \\ 0 & D_{21}\alpha + sI + a_{22}D_{22} - \epsilon D_2 \end{pmatrix}, \tag{37}$$

Multiply equation (34) by $diag(hI, h^2I)$ such that,

$$\begin{pmatrix} (\tilde{s}I + a_{11}\tilde{D}_{11}) & 0 \\ 0 & ha_{21}\tilde{D}_{21}\alpha + \tilde{s}I + ha_{22}\tilde{D}_{22} - \epsilon\tilde{D}_2 \end{pmatrix} \hat{e}_b' = \begin{pmatrix} h\hat{T}_b^{(1)} \\ h^2\hat{T}_b^{(2)\prime} \end{pmatrix}. \tag{38}$$

The upper left block is invertible by assumption (27), yielding that $\hat{e}_b^{(1)\prime}$ is order $r + 1$.

The lower left is $\tilde{s}I - \epsilon\tilde{D}_2 + \mathcal{O}(h)$. Two different approaches may be considered for this term. We can use Assumption 2.11 that the purely parabolic equation is uniquely determined such that the inverse of $(\tilde{s}I - \epsilon\tilde{D}_2)$ exists. Then $det((\tilde{s}I - \epsilon\tilde{D}_2)) \geq const > 0$ for $\tilde{s} \geq 0$. By Lemma 2.7, if h is small enough the perturbation does not make the matrix singular. Or, we use Assumption 2.10 that the incomplete parabolic system is pointwise stable in which case the inverse must exist by uniqueness of the numerical as well as the mathematical solution. Either of the two assumptions leads to a solution $\hat{e}_b^{(2)}$ of order $q + 2$. Solving for $\hat{e}_b$ yields,

$$\hat{e}_b = R\hat{e}_b' = \begin{pmatrix} \hat{e}_b^{(1)\prime} + \alpha\hat{e}_b^{(2)} \\ \hat{e}_b^{(2)} \end{pmatrix} \tag{39}$$

We conclude that $\hat{e}_b^{(1)} \sim max(\mathcal{O}(h^{r+1}), \mathcal{O}(h^{q+2}))$ and $\hat{e}_b^{(2)} \sim \mathcal{O}(h^{q+2})$. Inverting the Laplace transform yield the same order of magnitude to $e_b^{(1)}$ and $e_b^{(2)}$, respectivley.

11

Finally, we consider e_i. With Condition 2.10, (29) is pointwise stable and with Condition 2.11, (29) is stable, such that an estimate analogous to (11) is obtained in both cases. ∎

2.3 The Wave Equation

Consider a second order hyperbolic partial differential equation such as the wave equation.

$$
\begin{aligned}
u_{tt} &= u_{xx}, \quad 0 \le x \le \infty, \, 0 \le t \le T \\
L_0(t)u &= g_1(t), \text{ at } x = 0, \\
u(x,0) &= f(x)
\end{aligned}
\tag{40}
$$

We assume that (40) is supplied with boundary conditions such that it is well posed. A semidiscretisation of (40) can be written,

$$
\begin{aligned}
v_{tt} &= Mv + B \\
v(0) &= f,
\end{aligned}
\tag{41}
$$

where B includes the boundary data. We assume that the order of accuracy is $2p$ for the interior scheme and r at a finite number of boundary points (as $h \to 0$). Laplace transforming the error equation yields,

$$
(s')^2 e = Me + T
\tag{42}
$$

where T is the truncation error. We state the following theorem.

Theorem 2.13 *If v is a pointwise stable discrete solution to (41), then with $r = 2p - 2$ the global order of accuracy is $2p$.*

Proof With the transformation $(s')^2 = s$ the proof of Theorem 2.6 applies. ∎

Remark A necessary requirement for stability is that M is symmetric and negative semidefinite.

2.4 A General Statement

Consider the advection-diffusion equation, $u_t + au_x = \epsilon u_{xx}$. The above theory shows that pointwise stability of a scheme approximating the equation is sufficient to gain two orders of accuracy at the boundary. A key part in the proof is the multiplication of the truncation error by h^2 in (13).

On the other hand, with $\epsilon = 0$ and the assumption of a pointwise stable scheme, we could use the same proof but this time only multiplying the boundary error by h in (13). Then, the components of $\tilde{M}$ are $\mathcal{O}(1)$ and $\tilde{s} = sh$. That is just proving that we can lower the accuracy at the boundary by one order for hyperbolic equations, i.e. what is proven in [1, 2]. We have also shown above, that lower order terms will not affect the resulting accuracy.

The above reasoning justifies the study of the following equation,

$$
\begin{aligned}
u_t &= \frac{\partial^m u}{\partial x^m}, \quad 0 \le x < \infty \\
L_0 u &= g(t), \\
u(x,0) &= f(x)
\end{aligned}
\tag{43}
$$

since lower order terms will not affect the sufficient order at the boundary. We assume that (43) is well-posed. A semi-discretisation would be,

$$
\begin{aligned}
v_t &= M_h v + B, \\
v(0) &= f.
\end{aligned}
\tag{44}
$$

As before, we assume that the discrete scheme is pointwise stable. Again we study the error equation,

$$
\begin{aligned}
e_t &= M_h e + T_h, \\
e(0) &= 0.
\end{aligned}
\tag{45}
$$

and split the error into two parts, internal and boundary. (e_i, e_b, T_i, T_b) The internal error directly yields the correct order. The boundary part is Laplace transformed and viewed as a homogeneous difference equation with initial data. In order to obtain $\mathcal{O}(1)$ coefficients in M we need to multiply the error equation (corresponding to (13)) by h^m. With these observation the proof of the following theorem is identical to the previous proof of Theorem 2.6.

Theorem 2.14 *Assume that (43) is well-posed and its semi-discretisation (44) is pointwise stable. Then with the order of accuracy p in the interior and order $p - m$ at the boundary closure, the global order of accuracy is p.*

Remark In the proof it follows trivially that the weak SAT implementation of boundary conditions allow the physical boundary conditions to be approximated to order $p - m + 1$ without reducing the global order p.

3   Analysis of SBP Schemes

The conditions in Theorems 2.8, 2.12 and 2.13 are quite general and we will devote this chapter to prove that SBP-schemes with SAT implementation of boundary conditions, do fulfil these requirements. (For theory of SBP-schemes see [8, 9, 10, 7, 11, 12, 13, 14, 15, 16, 4])

3.1   The Heat Equation

3.1.1   The Continuous Problem

Again, we begin by analysing the simplest time dependent equation including a second derivative in space,

$$
\begin{aligned}
u_t &= u_{xx}, \quad 0 \leq x \leq 1, 0 \leq t \leq T \\
L_1 u &= g_1(t), \\
L_0 u &= 0,
\end{aligned}
\tag{46}
$$

where $L_1 u = u(1,t) + \alpha u_x(1,t)$, $\alpha > 0$ and $L_0 u = u_x(0,t)$ and (46) is assumed to have bounded initial data. The energy method applied to (46) leads to,

$$
\frac{1}{2}\|u\|_t^2 + \int u_x^2 \, dx = [uu_x]_0^1 = \leq -(1-\eta)\frac{|u(1,t)|^2}{\alpha} + \frac{1}{\eta}\frac{|g_1(t)|^2}{\alpha}.
\tag{47}
$$

where $\|u\|^2 = \int_0^1 u^2 dx$ and $0 < \eta \leq 1$. Integrating in time yields,

$$
\frac{1}{2}\|u(\cdot,T)\|^2 + \int_0^T (\|u_x(\cdot,t)\| + (1-\eta)\frac{|u(1,t)|^2}{\alpha})dt \leq
\tag{48}
$$
$$
\frac{1}{2}\|u(\cdot,0)\|^2 + \int_0^T \frac{1}{\eta}\frac{|g_1(t)|^2}{\alpha}dt,
$$

and well-posedness follows. Note also that $\|u_x(\cdot,t)\|$ is bounded. Then u can be pointwise estimated by a Sobolev inequality. For any point $x_1 \in [0,1]$ and every $\epsilon > 0$ we have,

$$
|u(x_1)| \leq \epsilon\|u_x\|^2 + (\epsilon^{-1} + 1)\|u\|^2.
\tag{49}
$$

3.1.2   The Semidiscrete Problem

In order to discretise (46), an approximation of the second derivative is needed. Such approximations in the SBP-framework are derived in [4] for different orders of accuracy, see also [11]. For any order, those can be expressed as,

$$
D_2 = P^{-1}(-A + BS).
\tag{50}
$$

In (50), P is an l^2-equivalent norm, that is P is symmetric and positive definite and $v^T P v = \|v\|_P^2$. Further, $A + A^T \geq 0$; $B = diag(-1, 0..., 0, 1)$ and S is a matrix approximating u_x at the boundaries. We will also need the following,

$$e_0 = (1, 0, ..., 0)^T, \; E_0 = diag(1, 0, ..., 0), \tag{51}$$
$$e_N = (0, ..., 0, 1)^T, \; E_N = diag(0, ..., 0, 1).$$

Further, we will frequently use the notation $(w)_i$ to denote the ith component of some vector w. Discretise equation (46) with $N+1$ grid points and denote the solution vector v. The operator (50) together with an SAT treatment for the boundary conditions lead to,

$$v_t = P^{-1}(-A + BS)v + \sigma_1 P^{-1} L_1^D(v, g_1) + \sigma_0 P^{-1} L_0^D v. \tag{52}$$

where $L_1^D(v, g_1) = (E_N(I + \alpha BS)v - e_N g_1(t))$, $L_0^D v = E_0 DSv$ and I denotes the identity matrix. The initial data is the vector f, i.e. the function $f(x)$ projected onto the grid. Next, we multiply (52) by $v^T P$ and add the result to its transpose. We obtain, with $\sigma_0 = 1$ and $\sigma_1 = -1/\alpha$,

$$(\|v\|_P^2)_t + v^T(A + A^T)v = \frac{-2}{\alpha} v_N(v_N - g_1(t)), \tag{53}$$

i.e. the discrete counterpart of (47). We omit the integration in time and conclude directly that the term $v^T(A + A^T)v$ will be bounded and is the discrete analogue of $\|u_x\|^2$ in (48).

The following properties of the SBP operators can be shown to hold and we state those without a proof in an assumption.

Assumption 3.1 *The matrix A, in the diagonal norm schemes we consider, is symmetric and the row sums are zero. Further, if A is an $n \times n$-matrix then $rank(A) = n - 1$.*

Remark The rank of A in Assumption 3.1 can be checked for some n. Then A is extended in the interior by the difference stencil which is linearly independent to the rest of the matrix. Hence, the rank does not change as n increases.

Lemma 3.2 *Let A be defined above and satisfy Assumption 3.1, c a positive constant and C a function depending only on data (f, g and F, denoting initial data , boundary data and forcing function respectively). Then, any scheme with an estimate $\|v\|_P^2 + cv^T(A + A^T)v < C(f, g, F)$, is pointwise stable.*

Proof In [5] the following discrete Sobolev inequality is proved,

$$|v_i| \leq \|v\|^2 + \epsilon \|D_+ v\|_2^2, \quad i = 1..N$$

where $\|v\|_2^2 = h \sum_1^N |v_i|^2$, $(D_+ v)_i = (v_i - v_{i-1})/h$ and $\epsilon > 0$ is a constant.

The first observation is the following,

$$0 \leq v^T \hat{A} v \leq ch. \tag{54}$$

where $\hat{A} = hA$ and all $\hat{A}_{ij}$ are of order 1. We will need a few properties of $\hat{A}$. For the diagonal norm case $\hat{A}$ is symmetric and the row sums are zero. Then,

$$ch \geq \sum_{i=1}^n \sum_{j=1}^n v_i \hat{A}_{ij} v_j = \sum_{i=1}^n v_i (\hat{A}_{ii} v_i + \sum_{j \neq i} \hat{A}_{ij} v_j) =$$

$$\sum_{i=1}^n v_i ((-\sum_{j \neq i} \hat{A}_{ij}) v_i + \sum_{j \neq i} \hat{A}_{ij} v_j) = \sum_{i=1}^n v_i (\sum_{j \neq i} \hat{A}_{ij} (v_j - v_i)) \geq 0.$$

Since $\hat{A}$ is symmetric this can be rewritten as,

$$\sum_{i=1}^n v_i (\sum_{j \neq i} \hat{A}_{ij} (v_j - v_i)) = \sum_{i=2}^n \sum_{i<j} (v_i - v_j)^2 (-\hat{A}_{ij}).$$

Next, consider,

$$(v_i - v_j)^2 = ((v_i - v_{i-1}) + (v_{i-1} - v_{i-2}) + ... + (v_{j+1} - v_j))^2. \tag{55}$$

From this observation we conclude that,

$$v^T \hat{A} v = v^T D^T B D v, \tag{56}$$

where B is an $(n - 1 \times n - 1)$-matrix and

$$D = \begin{pmatrix} -1 & 1 & 0 & \ldots & \\ 0 & -1 & 1 & 0 & \ldots \\ & & \ddots & \ddots & \\ & 0 & -1 & 1 \end{pmatrix}, \tag{57}$$

is an $(n - 1) \times n$ matrix. The crucial part is to prove that B is positive definite. Extend B by a top row and left column of zeros such that it becomes

an $n \times n$-matrix denoted by $\tilde{B}$. Further, let

$$\tilde{D} = \begin{pmatrix} 1 & 0 & \ldots & & & \\ -1 & 1 & 0 & \ldots & & \\ 0 & -1 & 1 & 0 & \ldots & \\ & & \ddots & \ddots & & \\ & & & 0 & -1 & 1 \end{pmatrix} \tag{58}$$

such that $\tilde{D}$ becomes a non-singular $n \times n$-matrix. Then $v^T A v = v^T \tilde{D}^T \tilde{B} \tilde{D} v$. Since $\tilde{D}$ is non-singular $\tilde{B}$ and A have the same rank, i.e. $rank(\tilde{B}) = rank(A) = n - 1$. Then, since B was extended by zeros, B itself must be non-singular, i.e. positive definite. Then $0 < Dv \leq c'h$. Hence, the discrete Sobolev inequality applies and we conclude that v is pointwise bounded. ∎

Remark In the example above the estimate is bounded by, $C(f, g_1, F)$, i.e. we have a bound for nonhomogeneous boundary data. Hence, the proof shows *strong pointwise stability*. In general, it might be easier to prove an energy estimate with $g_1 = 0$ in which case the above proof concerns *pointwise stability*.

Proposition 3.3 *With $\sigma_0 = 1$ and $\sigma_1 = -1/\alpha$, the discretisation (52) of (46) leads to strong pointwise stability and two orders of accuracy are gained at the boundary.*

Proof Equation (52) with $\sigma_0 = 1$ and $\sigma_1 = -1/\alpha$ leads to boundedness of $v^T(A + A^T)v$. Then by Lemma 3.2, (52) is (strongly) pointwise stable and two orders are gained by Theorem 2.8. ∎

Remark Note that, applying an SBP first derivative twice yields a noncompact second derivative in the interior. However, this does not affect the proof since the resulting A matrix has the same properties as those derived in [4] and stated in Assumption 3.1

3.1.3 Relations between Different Stability Definitions

To conclude this subsection we will discuss the differences of the various stability notions. To prove pointwise stability of an SBP-SAT scheme with compact second derivatives we have shown that it suffice to derive an energy estimate. In Definition 2.1 strong stability is defined. The energy estimate, (53), implies strong stability since we can obtain an estimate with nonhomogeneous boundary data. If we would need to assume that the boundary data is zero to obtain an energy estimate, the resulting scheme would have

17

been described as stable. However, consider the following transformation of a nonhomogeneous continuous problem,

$$
\begin{aligned}
u_t &= u_{xx} + F, \quad 0 \le x \le 1, \\
u(0,t) &= g_0(t), \\
u(1,t) &= g_1(t), \\
u(x,0) &= f(x).
\end{aligned}
\tag{59}
$$

Let $h(x,t)$ be a smooth function such that $h(0,t) = g_0(t)$, $h(1,t) = g_1(t)$ and $h(x,0) = f(x)$ and let $v = u - h$, then,

$$
\begin{aligned}
v_t &= v_{xx} + F - h_t + h_{xx} = v_{xx} + \tilde{F}, \\
v(0,t) &= v(1,t) = v(x,0) = 0.
\end{aligned}
\tag{60}
$$

Hence, with the assumption of sufficiently smooth boundary data an energy estimate for v also gives a bound on u. In the same manner we can define a grid function $h_i = h(x_i, t)$. The same transformation can be applied for the discrete problem with nonhomogeneous boundary data and a forcing function analogous to (59) and (60) using h_i. Hence, if the data is smooth, stability suffice to get an energy estimate and pointwise stability.

3.2 The Advection-Diffusion Equation

Consider,

$$
\begin{aligned}
u_t + a u_x &= \epsilon u_{xx} + F(x,t), \quad 0 \le x \le 1, \quad t \ge t_0, \\
L_0 u &= g_0(t), \\
L_1 u &= g_1(t), \\
u(x,t_0) &= f(x),
\end{aligned}
\tag{61}
$$

where $L_0 u = u(0,t) + \alpha u_x(0,t)$ and $L_1 u = u(1,t) + \beta u_x(1,t)$. Assume that $a > 0$, then equation (61) can be proven well-posed with the energy method if,

$$
-\frac{2\epsilon}{a} \le \alpha \le 0, \ \beta \le -\frac{2\epsilon}{a}, \beta > 0.
\tag{62}
$$

From the previous subsection, we have the tools to prove pointwise stability by deriving an energy estimate. Equation (61) is discretised as,

$$
\begin{aligned}
v_t + a P^{-1} Q v &= \epsilon P^{-1}(-A + BS)v - P^{-1}\sigma_0 L_0^D v - P^{-1}\sigma_1 L_1^D \\
v(0) &= f.
\end{aligned}
\tag{63}
$$

where $L_0^D v = (E_0(I - \alpha BS)v - e_0 g_0(t))$ and $L_1^D v = (E_N(I + \beta BS)v - e_N g_1(t))$ and I denotes the identity matrix.

The first derivative approximation operator $P^{-1}Q$ satisfies the following relation, $Q + Q^T = B$, where $B = diag(-1, 0, ..., 0, 1)$. Next, the energy method is applied by multiplying equation (63) by $v^T P$ and adding the result to its transpose. An energy estimate is obtained if $\sigma_0 = \epsilon/\alpha$, $\sigma_1 = -\epsilon/\beta$ and (62) hold. We have,

$$\frac{d}{dt}(v^T P v) + a v^T B v = -\epsilon v^T (A + A^T) v \tag{64}$$
$$+2\frac{\epsilon}{\alpha} v_0 (v_0 - g_0(t)) - 2\frac{\epsilon}{\beta} v_N (v_N - g_1(t)).$$

Condition (62) ensures that the boundary terms are bounded such that the desired estimate of the semidiscrete initial-boundary value problem is obtained. Denoting the boundary terms by BT, (64) becomes,

$$\frac{d}{dt}(v^T P v) + BT = -\epsilon v^T (A + A^T) v.$$

Omitting the integration in time we conclude, using Lemma 3.2, that v is strongly pointwise bounded. Since the requirement of Theorem 2.8 is fulfilled, we have proved the following theorem.

Proposition 3.4 *With $\sigma_0 = -\epsilon/\beta_0$ and $\sigma_1 = \epsilon/\beta_1$ and (62), the discretisation (63) of (61) with internal order of accuracy $2p$ and boundary accuracy r has global accuracy $min(2p, r)$.*

We have justified that for these SBP schemes 2 orders less accuracy at the boundary does not reduce the global accuracy of the scheme Note also that in the case with parabolic terms we can also reduce the accuracy of the hyperbolic terms two orders at the boundary.

3.3 An Incompletely Parabolic System

3.3.1 The Continuous Problem

We proceed by considering one example of an incompletely parabolic system of equations and test the conditions in Theorem 2.12.

$$\begin{pmatrix} \tilde{u}^{(1)} \\ \tilde{u}^{(2)} \end{pmatrix}_t + \mathbf{A} \begin{pmatrix} \tilde{u}^{(1)} \\ \tilde{u}^{(2)} \end{pmatrix}_x = \epsilon \mathbf{C} \begin{pmatrix} \tilde{u}^{(1)} \\ \tilde{u}^{(2)} \end{pmatrix}_{xx} \tag{65}$$
$$\tilde{u}^{(1)}(0) = g^{(1)}(t),$$
$$\tilde{u}^{(2)}(0) = g^{(2)}(t),$$
$$\tilde{u}_x^{(2)}(1) = g^{(3)}(t).$$

where $\mathbf{A}$ is a symmetric positive definite (2×2)-matrix such that $[\mathbf{A}]_{ij} = a_{ij}$, $\mathbf{C} = diag(0, 1)$ and $\epsilon > 0$. We define the norm $\|\tilde{u}^{(i)}\|^2 = \int_0^1 (\tilde{u}^{(i)})^2 \, dx$ and apply the energy method,

$$\frac{1}{2} \left(\|\tilde{u}^{(1)}\|_t^2 + \|\tilde{u}^{(2)}\|_t^2 \right) + \frac{1}{2}(\tilde{u}^{(1)} \, \tilde{u}^{(2)}) \mathbf{A} (\tilde{u}^{(1)} \, \tilde{u}^{(2)})^T|_0^1 - \tilde{u}^{(2)} \tilde{u}_x^{(2)}|_0^1 =$$

$$-\epsilon \int_0^1 (\tilde{u}_x^{(2)})^2 \, dx.$$

Imposing the boundary conditions, and for simplicity assuming that $g^{(2)} = g^{(3)} = 0$, yields,

$$\frac{1}{2} \left(\|\tilde{u}^{(1)}\|_t^2 + \|\tilde{u}^{(2)}\|_t^2 \right) + \frac{1}{2}(\tilde{u}^{(1)}(1) \, \tilde{u}^{(2)}(1)) \mathbf{A} (\tilde{u}^{(1)}(1) \, \tilde{u}^{(2)}(1))^T +$$

$$\epsilon \int_0^1 (\tilde{u}_x^{(2)})^2 \, dx = \frac{1}{2} g^{(1)} a_{11} g^{(1)},$$

Thus the problem (65) is well posed.

3.3.2 The Semidiscrete Problem

To analyse systems of partial differential equations it is convenient to introduce the *Kronecker product*,

$$A \otimes B = \begin{pmatrix} a_{0,0} B & \dots & a_{0,q-1} B \\ \vdots & & \vdots \\ a_{p-1,0} B & \dots & a_{p-1,q-1} B \end{pmatrix}, \tag{66}$$

where A is a $(p \times q)$ matrix and B an $m \times n$ matrix. The Kronecker product satisfies the following rules: $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ and $(A \otimes B)^T = A^T \otimes B^T$.

We proceed by constructing a semidiscretisation of (65) with $N + 1$ grid points. Let $v_i^{(1)}$ and $v_i^{(2)}$ denote the approximation of $u^{(1)}(x_i)$ and $u^{(2)}(x_i)$. Further, let $v_i = (v_i^{(1)}, v_i^{(2)})^T$ and $\mathbf{v} = (v_0, v_1, ..., v_N)^T$. Finally, we will need, $\mathbf{v}_0^{(1)} = (v_0^{(1)}, 0, ...)^T, \mathbf{v}_0^{(2)} = (0, v_0^{(2)}, 0, ...)^T$, $\mathbf{v}_N^{(2)} = (..., 0, v_N^{(2)})^T$ and $E_N^{(2)}$ such that $E_N^{(2)} \mathbf{v} = \mathbf{v}_N^{(2)}$.

The basic scheme approximating (65), without boundary conditions, is,

$$\mathbf{v}_t + (P^{-1} Q \otimes \mathbf{A}) \mathbf{v} = (P^{-1}(-A + BS) \otimes \epsilon \mathbf{C}) \mathbf{v}. \tag{67}$$

To determine the structure of the penalty terms the energy method is applied to (67) by multiplying $v^T (P \otimes I)$, where I is the (2×2) identity matrix, and adding the transpose.

The resulting boundary terms determines the form of the penalties and the full SBP-SAT scheme approximating (65) becomes,

$$\mathbf{v}_t + (P^{-1}Q \otimes \mathbf{A})\mathbf{v} = (P^{-1}(-A + BS) \otimes \epsilon\mathbf{C})\mathbf{v} +$$
$$\sigma_0(P^{-1} \otimes \mathbf{A})(\mathbf{v}_0^{(1)} - G_1) + \sigma_1(P^{-1}(BS)^T \otimes \epsilon\mathbf{C})(\mathbf{v}_0^{(2)} - G_2) + \quad (68)$$
$$\sigma_2(P^{-1} \otimes \epsilon\mathbf{C})(E_N^{(2)}(BS \otimes I)\mathbf{v} - G_3),$$

where $G_i = (g^{(i)}, 0, ...)^T$, $i = 1, 1$ and $G_3 = (0, ..., 0, g^{(3)})$. For simplicity, we assume that $g^{(2)} = g^{(3)} = 0$.

We have used $\|\mathbf{v}\|_M^2 = \mathbf{v}^T(P \otimes I)\mathbf{v}$ to denote the norm. With $\sigma_0 \leq -1/2$ and $\sigma_1 = \sigma_2 = -1/2$ we have,

$$(\|\mathbf{v}\|_M^2)_t + v_N^T \mathbf{A} v_N + \mathbf{v}^T((A + A^T) \otimes \epsilon\mathbf{C})\mathbf{v} = \quad (69)$$
$$(1 + 2\sigma_0)a_{11}v_0^2 - 2\sigma_0 v_0^{(1)} a_{11} g^{(1)}$$

If $\sigma_0 = -1/2$ in (69) we obtain exactly the same estimate as in the continuous case.

In the previous subsection we proved the heat equation to be pointwise stable, which is a requirement for Theorem 2.12 to apply. It remains to show that the hyperbolic part satisfies the determinant condition. The hyperbolic part of the scheme is in general of the form,

$$v_t + P^{-1}Qv = \sigma_0 P^{-1} E_0(v_0 - g(t)). \quad (70)$$

(In the specific example above $g(t) = 0$, but that is not necessary.) For a hyperbolic equation it is not sufficient that the scheme is strongly stable for it to be pointwise stable (which is equivalent to the determinant condition).

We begin by considering dissipative schemes and restrict ourselves to schemes where P is diagonal. Then $P^{-1}Q$ is replaced by,

$$P^{-1}(Q + R). \quad (71)$$

In [14] dissipation operators that do not destroy the SBP-property are derived, such that $R = \gamma h \tilde{D}_p^T B \tilde{D}_p$ where $\tilde{D}_p/h^p$ is a first order accurate approximation of the pth space derivative, $\gamma > 0$ is a parameter and B a positive definite matrix. If p is chosen such that $2p$ is the order of the scheme this dissipation operator will keep the order of accuracy without widening the stencil. In order to prove pointwise stability we must choose $\gamma \sim 1/h$. Then the accuracy is lowered one order or we must choose a larger p, i.e. widening the stencil. We can prove the following proposition.

Proposition 3.5 *The scheme (70), discretised with a dissipative SBP operator (71), satisfies the determinant condition (27), i.e. it is pointwise stable.*

21

Proof See Appendix I.

For a central difference scheme we can not use the energy method to prove pointwise stability. Hence, we have to turn to the Laplace transform technique. (See [5] for a thorough presentation of the theory.) For this reason we have to prove that the determinant condition is satisfied for each particular type of scheme and order of accuracy. However, the Laplace transform technique becomes increasingly difficult to apply for higher order schemes. We state the following conjecture and give some justification.

Conjecture 3.6 *The scheme (70), discretised using a central difference SBP scheme, satisfies the determinant condition, i.e. it is pointwise stable.*

A proof that the Conjecture is true in the second order case is included in Appendix II. For an internally fourth order scheme we give in Appendix III some analysis indicating the truth of the Conjecture although it is strictly not a proof. For higher order methods than four, we refer to computations where the measured global order of accuracy can be explained if the Conjecture is true.

We conclude that the requirements of Theorem 2.12 may be fulfilled and summarise the results in a proposition.

Proposition 3.7 *If either Proposition 3.5 or Conjecture 3.6 holds, then with $\sigma_0 = -1$ and $\sigma_1 = \sigma_2 = -1/2$ the discretisation (68) of (65) with internal order of accuracy $2p$ and boundary accuracy r for the parabolic and $r+1$ for the hyperbolic equation has global order of accuracy $min(r+2, 2p)$.*

This is just one example of an incompletely parabolic system that we use to show the techniques to prove the conditions of Theorem 2.12. However, for any well posed incompletely parabolic system, discretised with an SBP and SAT scheme that satisfy a discrete energy estimate, those conditions will be fulfilled.

3.4 The Wave Equation

Consider (40) with homogeneous Dirichlet boundary conditions. In the SBP-setting we discretise by,

$$v_{tt} = P^{-1}(-A + DS)v + \sigma_0 P^{-1}E_0 S(v - 0) + \sigma_1 P^{-1}E_N S(v - 0). \quad (72)$$

In this case A has to be symmetric which the energy method will reveal below. Applying the energy method to (72) yields,

$$(\|v_t\|_P^2 + v^T A v)_t = -2(1 - \sigma_0)(v_t)_0(Sv)_0 + 2(1 + \sigma_1)(v_t)_N(Sv)_N. \quad (73)$$

Note that without symmetry of A it would not be possible to obtain the total derivative $(v^T A v)_t$. With $\sigma_0 = 1$ and $\sigma_1 = -1$ stability follows. In this case we do not directly have a bound on $\|v\|$ and $v^T A v$. However, with $\|f\| \leq \infty$ we can solve the ordinary differential equation (73) to bound $\|v_t\|$ and $v^T A v$. Since the norm of $v(0)$ and $\|v_t\|$ is bounded it follows that $\|v\|$ has to be bounded. Then we can estimate the solution pointwise using Lemma 3.2 such that the requirement of Theorem 2.13 is fulfilled. We summarise the results in a proposition.

Proposition 3.8 *With $\sigma_0 = 1, \sigma_1 = -1$, the discretisation (72) of (40) has global order of accuracy $min(2p, r + 2)$ where r is the boundary and $2p$ the internal order of accuracy.*

4 Computations

In [4] extensive computations on the advection-diffusion and incompletely parabolic equations were performed with SBP-schemes. We will not redo the calculations for the advection-diffusion but only give their results. We will omit computations for the heat equation, since it is a special case of the advection–diffusion equation. We present novel results for the wave equation. We will also test the validity of the linear theory for the nonlinear viscous Burgers' equation. Finally, we compute numerical solutions to a simple 4th order equation.

Throughout this section we will consider approximations of the second derivatives derived in [4]. Also, first derivative approximations are used. Those were originally derived in [8, 9] and given as exact expressions in [10]. We distinguish between two types of operators. Those with a diagonal norm, i.e. P is diagonal, and those with a block norm where P is diagonal except at the upper-left and lower-right corners where blocks are situated. In [8, 9] it was proven that diagonal norm schemes can only be closed at the boundary with half the internal accuracy.

4.1 Equations with First Derivative in Time

The contents of this subsection was originally presented in [4] and we briefly quote some of their computational results.

N	$log(l_2 - error)$	q	$log(l_v^2 - error)$	q^v
40	-4.25		-2.59	
60	-5.02	4.30	-3.13	3.01
100	-5.98	4.25	-3.81	3.01
200	-7.24	4.17	-4.72	3.01
300	-7.97	4.11	-5.25	3.00

Table 1: SBP-scheme with 4th order internal accuracy and 2nd order boundary closure. The two right columns are results for scheme with stability estimates violated.

4.1.1 The Advection-Diffusion Equation

Consider equation (1) discretised by (63). The Cauchy problem have the solution,

$$u = sin(\omega(x - ct))e^{-bx}, \, c > 0, \, \omega = \frac{\sqrt{c^2 - a^2}}{2\epsilon}, b = \frac{c - a}{2\epsilon}, |c| > |a|. \quad (74)$$

The computional domain is $0 < x < 1$ and (74) is used both as initial and boundary conditions. Further, $a = 1$, $c = 2$ and $\epsilon = 0.1$ have been used. The convergence rate is calculated as,

$$q = log\left(\frac{\|u - v^{h_1}\|_h}{\|u - v^{h_2}\|_h}\right) / log\left(\frac{h_1}{h_2}\right), \quad (75)$$

where u is the analytical solution and v^{h_1} is the corresponding numerical solution with grid size h_1. Further, $\|u - v^{h_1}\|_h$ is the l_2-error.

In [4] results are presented for schemes with both 4th and 6th order internal acuracy. The results agree with the theory and we choose only to quote the results for a 4th order diagonal norm scheme, Table 1. Note that with a diagonal norm an internally 4th order accurate scheme can be closed to maximally 2nd order at the boundary. However, two orders are gained at the boundary and the scheme is globally 4th order.

Two different cases are shown:

1. Theoretically strongly stable scheme. Hence, also pointwise stable.

2. The theoretical estimates are violated by altering the penalty parameter. Hence, the scheme is not pointwise stable. However, the computations are stable in the sense that the eigenvalues are located in the left half plane. This case is marked with superscript v.

24

Notable is that if the penalty parameter is chosen such that the scheme is not energy stable (though the computations are not unstable), the global order of accuracy is reduced by 1 indicating that the conditions in Theorem 2.8 are not only necessary but also sufficient. From the present article this is justified since the scheme is not pointwise stable.

Finally, if $\epsilon = 0$ in the above computations, i.e. we have a hyperbolic equation, the accuracy drops to 3rd order in full agreement with the results in [1, 2].

4.1.2   An Incompletely Parabolic System

The system (65) was considered in [4] with,

$$u = \left(\begin{array}{c} u^{(1)} \\ u^{(2)} \end{array} \right), \ C = \left(\begin{array}{cc} 1 & 1 \\ 1 & -1 \end{array} \right), \ D = \left(\begin{array}{cc} 0 & 0 \\ 0 & \epsilon \end{array} \right). \tag{76}$$

The system is transformed such that the hyperbolic part is diagonal and provided with well posed boundary conditions. The system is discretised using SBP and SAT technique such that the scheme is strongly stable.

We will discuss the results from two test cases:

1. An internally fourth order block norm scheme. The second derivatives are closed to second order accuracy and the first derivatives to third order accuracy.

2. An internally fourth order accurate diagonal norm scheme. Both the first and second derivatives are closed to second order accuracy.

In Table 2 the results of Test Case 1 are displayed. Also in this case, orders of accuracy to the problem with a non-energy stable choice of the penalty parameter are presented. This reduces the global order of accuracy by one. This indicates that the conditions of Theorem (2.12) are both necessary and sufficient.

Next, we turn to Test Case 2. The results are shown in Table 3. As expected the scheme is only third order accurate. All the hyperbolic terms are discretised with second order boundary closure but Theorem 2.12 requires the hyperbolic equation in the system to have a boundary closure of only one order less than the internal scheme. Hence, the violation of the energy estimates does not affect the accuracy either, as long as the scheme remains stable in the numerical computations.

Note that, since fourth order accuracy is recovered in Table 2, the Conjecture 3.6 seems to be true. The hyperbolic part need to be pointwise stable for Theorem 2.12 to be true.

N	$log(l_2 - error)$	q	$log(l_v^2 - error)$	q^v
30	-3.31		-3.26	
60	-4.52	3.91	-4.25	3.24
90	-5.23	4.00	-4.81	3.10
120	-5.74	4.03	-5.19	3.05
150	-6.13	4.03	-5.48	3.03

Table 2: SBP-scheme with 4th order internal accuracy and 2nd order boundary closure for the second derivative and third order for the first derivative. The two right columns are results for scheme with stability estimates violated.

N	$log(l_2 - error)$	q	$log(l_v^2 - error)$	q^v
30	-2.59		-2.60	
60	-3.61	3.33	-3.55	3.10
90	-4.18	3.19	-4.10	3.05
120	-4.58	3.13	-4.48	3.05
150	-4.88	3.11	-4.78	3.04

Table 3: SBP-scheme with 4th order internal accuracy and 2nd order boundary closure for both first and second derivative. The two right columns are results for scheme with stability estimates violated.

N	$log(l_2 - error)$	q
10	-1.09	
20	-5.23	5.97
40	-8.59	4.84
80	-11.27	3.87
160	-14.07	4.04

Table 4: SBP-scheme with 4th order internal accuracy and 2nd order boundary closure.

4.2 The Wave Equation

Using the scheme (72) we have computed convergence rates to corroborate Theorem 2.13. (In (72) it is assumed that the boundary data is zero. This is sufficient for Theorem 2.13 to hold. However, in the case below, it is possible to show strong stability which allows less smoothness in the data.)

We have considered the following wave equation,

$$
\begin{aligned}
u_{tt} &= c^2 u_{xx}, \ 0 \le x \le \pi, \ 0 \le t \le 0.5 \\
u_x(0, t) &= cos(ct) \\
u_x(\pi, t) &= -cos(ct) \\
u(x, 0) &= sin(x), \\
u_t(x, 0) &= 0,
\end{aligned}
$$
(77)

(78)

where $c = 2$. The exact solution is $u(x, t) = \frac{1}{2}(sin(x - ct) + sin(x + ct))$. The l_2-error and convergence rate are computed at $t = 0.5$. The results are shown in Table 4. In Table 4 there are no data for the scheme with the energy estimate violated. This is due to the scheme being unstable for $\sigma_0 \ne 1$ or $\sigma_1 \ne -1$. Further, we note that 4th order accuracy is achieved in accordance with the theory.

4.3 The viscous Burgers' equation

Consider,

$$
\begin{aligned}
u_t + u u_x &= \epsilon u_{xx}, \quad 0 \le x \le L, \quad t \ge t_0, \\
L_0 u &= g_0(t), \\
L_1 u &= g_1(t), \\
u(x, t_0) &= f(x),
\end{aligned}
$$
(79)

where $L_0 u = u(0, t) + \alpha u_x(0, t)$ and $L_1 u = u(L, t) + \beta u_x(L, t)$. If equation (79) is linearised we obtain (61) and from the linear theory we derive a numerical

scheme that is similar to (63). For linear well-posedness we have,

$$-\frac{2\epsilon}{u(0)} \leq \alpha \leq 0, \quad \beta > 0, \ \beta \leq -\frac{2\epsilon}{u(L)}. \tag{80}$$

Equation (79) is discretised as,

$$
\begin{aligned}
v_t + P^{-1}Q(\frac{v^2}{2}) &= \epsilon D_2 v - P^{-1}\sigma_0 L_0^D v - P^{-1}\sigma_1 L_1^D \\
v(0) &= f.
\end{aligned}
\tag{81}
$$

where $L_0^D v = (E_0(I - \alpha BS)v - e_0 g_0(t))$ and $L_1^D v = (E_N(I - \beta BS)v - e_N g_1(t))$ and I denotes the identity matrix. We use a standard fourth order Runge-Kutta scheme to discretise (81) in time. The computations are done with a constant small time step and 100 iterations. In (79) we choose $t_0 = 0.16$ and $L = 0.5$. The exact solution to the viscous Burgers' equation is,

$$u(x,t) = -a \cdot tanh(a\frac{x-ct}{2\epsilon}) + c \quad -\infty < x < \infty, \ t \geq 0. \tag{82}$$

The solution (82) is used as initial and boundary data with $a = 1$, $c = 2$ and $\epsilon = 0.02$.

Remark Due to the specific form of the solution the accuracy limit of the computer is reached very quickly in a large part of the domain destroying the measurents of the order of accuracy. Hence, the parameters of the problem have to be chosen with some care to obtain the correct order of accuracy within the finite precision. With a more complex solution it would be easier to measure the order of accuracy but then no analytical solution is available.

We test two different cases,

1. $D_2 = P^{-1}(-A + BS)$; internally 4th order accurate; 2nd order boundary scheme; Su is 3rd order discretisation of u_x at the boundary points.

2. $D_2 = P^{-1}QP^{-1}Q$; internally 4th order accurate; 2nd order boundary scheme; $Su = P^{-1}Qu$, i.e. 2nd order accurate.

Table 1 displays fourth order convergence for Test Case 1 just as in the linear case. Indicating that the linear theory is applicable in this nonlinear case also. Table 2 displays third order accuracy. This is due to the second order accuracy of the discretisation of the boundary condition. However, this also indicates that the linear theory applies. If two orders of accuracy were not gained at the boundary Test Case 2 would result in globally 2nd order accuracy.

N	$log(l_2 - error)$	q
210	-21.0	
230	-21.4	3.98
250	-21.7	3.97
270	-22.0	3.97
290	-22.3	3.97

Table 5: SBP-scheme with second derivative approximation according to Test Case 1.

N	$log(l_2 - error)$	q
210	-20.8	
230	-21.1	3.47
250	-21.4	3.32
270	-21.6	3.17
290	-21.8	3.02

Table 6: SBP-scheme with second derivative approximation according to Test Case 2.

4.4 The biharmonic operator

Consider,

$$
\begin{aligned}
u_t &= -u_{xxxx}, \quad 0 \le x \le L, \ t \ge 0 \\
u_{xx}(0) &= g_1(t), \\
u_{xx}(L) &= g_2(t). \\
u(0) &= g_3(t), \\
u(L) &= g_4(t),
\end{aligned}
\tag{83}
$$

With the energy method it is easily shown that (83) is well-posed. $u = sin(x)e^{-t}$ is a solution to the Cauchy problem and by choosing $g_{1,2,3,4}$ accordingly we have an exact solution to (83). The equation is discretised by,

$$
u_t = -D_4 u + penalty,
\tag{84}
$$

where $D_4 = D_1 \cdot D_1 \cdot D_1 \cdot D_1$ and $D_1 = P^{-1}Q$ is an SBP operator with 6th order internal accuracy and 3rd order boundary accuracy. Hence, D_4 is 0th order at the boundary and 6th order in the interior. Further,

$$
\begin{aligned}
penalty = P^{-1}(\sigma_1 D_1^T E_0(D_2 u - g_1) + \sigma_2 D_1^T E_N(D_2 u - g_2) + \\
\sigma_3 D_3^T E_0(u - g_3) + \\
\sigma_4 D_3^T E_N(u - g_4)),
\end{aligned}
$$

N	$log(l_2 - error)$	q
20	-8.93	
30	-10.63	4.0295
40	-11.83	4.0391
50	-12.75	4.0362
60	-13.50	4.0321
70	-14.13	4.0285
80	-14.67	4.0255
90	-15.15	4.0230

Table 7: SBP-scheme with second derivative approximation according to Test Case 2.

where $\sigma_1 = 1, \sigma_2 = -1, \sigma_3 = -1$ and $\sigma_4 = 1$ lead to stability. The first two penalty terms are 1st order implementation of the boundary condition multiplied by P^{-1} which leads to 0th order truncation error at the boundary. The second two terms does not have a truncation error. Altogether, we have a globally 4th order accurate scheme when Theorem 2.14 has been applied. (We omit the proof of pointwise stability since it is similar to all the previous.)

The results of computations with the scheme above is shown in Table 7. As before we use a 4th order Runge-Kutta in time. We choose $L = 3\pi/4$ to obtain non-zero boundary data and the final time is $t = 0.01$ in order not to introduce a large temporal error. We see that the convergence is 4th order as predicted by theory.

5   Summary and Conclusions

The results of this article can be divided into three parts. In the first part we consider partial differential equations including spatial second derivatives. We show that finite difference discretisations of such equations can be closed with 2 orders less accuracy at the boundary without reducing the global accuracy, if the scheme is pointwise stable. In particular it should be noted that this result also applies to second order hyperbolic equations such as the wave equation.

An immediate consequence of this theory is a generalisation to PDE:s with mth order derivatives. With the same stability assumption on the scheme it is possible to lower the order of the boundary closure m orders of accuracy.

In the second part, it is shown that summation-by-parts operators with either compact second derivatives or, with the first derivative applied twice, fulfil these requirements. For summation-by-parts operators the task of prov-

ing pointwise boundedness is reduced to derive an energy estimate for the scheme which is considerably simpler. (See [8, 9, 10, 7, 11, 12, 13, 14, 15, 16])

The third part concerns numerical results. In [4] and [3] the newly developed theory is verified for different schemes with a first derivative in time. In [4], stable computations with the energy estimate (and hence, the pointwise stability), violated, were performed showing that 2 orders of accuracy are not gained at the boundary. That is in full agreement with the theory developed in this article and indicates that pointwise stability is a necessary condition.

Further, numerical computations with the wave equation supports the theoretical results showing that the scheme can be closed with 2 orders of accuracy less at the boundary. We also test the validity of the linear theory on the nonlinear viscous Burgers' equation. Computations show that the linear theory is applicable. At last, we perform computations for a time-dependent 4th order equation and show that 4 orders of accuracy are gained at the boundary.

As a final observation, consider a first derivative approximation with reduced order at the boundary. The truncation error at the boundary is increased by one order for each new application of the first derivative operator to approximate a higher derivative. However, the theory of this article shows that the decreasing order of accuracy at the boundary is precisely cancelled resulting in the same global accuracy.

APPENDIX

I   Proof of Proposition 3.5

Throughout this proof, the tilde sign indicates that it is an undivided property, i.e. the components have no dependence on h. Furthermore, C always denotes a constant, not necessary the same in every expression.

Here we will prove that,

$$
\begin{aligned}
v_t + P^{-1}(Q + R)v &= \sigma_0 P^{-1}(v_0 - g(t)), \\
v(0) &= f,
\end{aligned}
\tag{85}
$$

is pointwise stable. In [14] a numerical dissipation of the form $\tilde{P}^{-1}R = \gamma \tilde{P}^{-1}\tilde{D}_p^T B \tilde{D}_p$ was derived. B is an $\mathcal{O}(1)$ positive definite matrix. With $\gamma \sim 1/h$ we can prove the theorem, which corresponds to an upwind scheme, i.e. the order of accuracy drops by one order.

Apply the energy method (85),

$$
(\|v\|_P^2)_t + v^T B v + v^T(R + R^T)v = 2\sigma_0 v_0(v_0 - g(t)),
$$

Using that $v_0 g(t) \leq \eta v_0^2 + \frac{1}{\eta}(g(t))^2$, $\eta > 0$, we obtain with $\eta < 1$ an estimate of $\|v\|_P^2$ in $g(t)$. Thus, the scheme (70) is strongly stable. Furthermore, $v^T(R + R^T)v < C$.

First, we consider boundedness of $\|v\|_P^2$. The norm $\|\cdot\|_P$ is l_2 equivalent. Hence,

$$
\sum_{i=1}^{N} h|v_i|^2 < C, \text{ or } \sum_{i=1}^{N} |v_i|^2 < \frac{C}{h}.
\tag{86}
$$

We see that $|v_i|$ may become infinite as the total number of points $N = 1/h \to \infty$. However, the total number of unbounded points n satisfies, $n/N \to 0$ as $N \to \infty$.

Next, since $\tilde{D}_p$ is an high order undivided difference, $v^T(R + R^T)v < C$ implies,

$$
v^T(R + R^T)v \sim \sum_{p}^{N-r} \gamma h(\tilde{D}_p v)_i^2 \Rightarrow \sum_{p}^{N-r} (\tilde{D}_p v)_i^2 < C,
\tag{87}
$$

The sum goes between the points closest to the boundary such that the difference do not pass over the boundary. (Every point will be 'touched' by the sum.) Equation (87) yields directly,

$$
(\tilde{D}_p v)_i^2 < C.
\tag{88}
$$

Assume that some $|v_j|$ tend to infinity without violating (86). Since only a decreasing fraction of grid points may become unbounded we can choose v_j such that its p closest neighbours are bounded. Consider, $(\tilde{D}_p v)_j^2 = (\alpha_0 v_j + \alpha_1 v_{j+1} + ... \alpha_p v_{j+p})^2 \to \infty$ as $h \to 0$ which is a violation of (88). Hence, $|v_i| < \infty$ for all i as $N \to \infty$.

Remark If the direction of convection would be the opposite we would consider the scheme,

$$v_t - P^{-1}(Q - R)v = \sigma_0 P^{-1}(v_0 - g(t)),$$

and be able to prove pointwise stability. This corresponds to upwinding and downwinding.

II Proof of Conjecture 3.6 for second order central differences

The first part of this Appendix applies to all central difference SBP operators.

To have the most general case we do not assume homogeneous boundary data. We restate (70) here for convenience,

$$\begin{aligned} v_t + P^{-1}Qv &= \sigma_0 P^{-1}(v_0 - g(t)), &(89)\\ v(0) &= f. \end{aligned}$$

First of all we note that $\sigma_0 = -1$ derived in Subsection 3.3 results in a strongly stable scheme. Apply the energy method to (89),

$$(\|v\|_P^2)_t + v^T B v = 2\sigma_0 v_0 (v_0 - g(t)),$$

or, with $\sigma_0 = -1$,

$$(\|v\|_P^2)_t + v_N^2 + v_0^2 = v_0 g(t).$$

Using that $v_0 g(t) \leq \eta v_0^2 + \frac{1}{\eta}(g(t))^2$, $\eta > 0$, we obtain with $\eta < 1$ an estimate of $\|v\|_P^2$ in $g(t)$. Thus, the scheme (70) is strongly stable. For $\sigma_0 = -1$ we will prove that the scheme also satisfies the determinant condition.

In [5] it is shown that it suffice to study the Laplace transformed quarter space problem,

$$\tilde{s}\varphi + hP^{-1}Q\varphi = h\sigma_0 P^{-1}(\varphi_0 - \hat{g}), \quad \|\varphi\|_P < \infty, \tag{90}$$

where $\tilde{s} = sh$ and initial data is zero.

We restrict ourselves to the second order case. The internal scheme in (95) is,

$$\tilde{s}\varphi_i - \frac{1}{2}\varphi_{i-1} + \frac{1}{2}\varphi_{i+1} = 0, \tag{91}$$

Assume solutions of the form $\varphi_i = \sum_{j=1}^{2} \tau_j \kappa_j^i$. From (91) we obtain the roots,

$$\kappa_1 = -\tilde{s} + \sqrt{\tilde{s}^2 + 1}, \tag{92}$$
$$\kappa_2 = -\tilde{s} - \sqrt{\tilde{s}^2 + 1}. \tag{93}$$

For $Re\,\tilde{s} > 0$, $|\kappa_1| < 1$ and $|\kappa_2| > 0$ (see [5]). The second root is discarded due to boundedness of $\|\varphi\|_P$. Hence, the solution is of the form $\varphi_i = \tau_1 \kappa_1^i$ and we need to show that τ_1 is bounded.

By inserting $\varphi_i = \tau_1 \kappa_1^i$ into the boundary scheme we obtain from (90) in the second order case,

$$\tilde{s}\varphi_0 + (-\varphi_0 + \varphi_1) + 2\varphi_0 = (\tilde{s} + 1 + \kappa_1)\tau_1 = 2\hat{g} \tag{94}$$

By observing that $\tilde{s} + 1 + \kappa_1 = 1 + \sqrt{\tilde{s}^2 + 1} \neq 0$ for $Re\tilde{s} \geq 0$, we conclude that,

$$det(\tilde{s}\varphi + hP^{-1}Q\varphi - h\sigma_0 P^{-1}\varphi_0) > \delta > 0, \tag{95}$$

for $Re\,\tilde{s} \geq 0$.

The last part of the proof could be simplified in this case. We know from strong stability (shown above) that the boundary point is bounded. That is, $\varphi_0 = \tau_1 \kappa_1^0$ is bounded, therefore τ_1 is bounded. Thus, φ_i is bounded for all i.

III Aspects of Conjecture 3.6 in the fourth order case

As in Appendix II, we begin with the scheme (89) and conclude that it is strongly stable by the same derivation.

The internal scheme of (90) is,

$$\tilde{s}\varphi_i + \frac{1}{12}\varphi_{i-2} - \frac{2}{3}\varphi_{i-1} + \frac{2}{3}\varphi_{i+1} - \frac{1}{12}\varphi_{i+2} = 0 \tag{96}$$

where $\tilde{s} = sh$. Assume solutions of the form $\varphi_i = \sum_{j=1}^{4} \tau_j \kappa_j^i$. In [5] it is shown that there are 2 positive and 2 negative roots for $Re\tilde{s} > 0$. The

positive roots can be discarded due to $\|\varphi\|_P \leq \infty$. We have solutions of the form,

$$\varphi_j = \tau_1 \kappa_1^j + \tau_2 \frac{\kappa_2^j - \kappa_1^j}{\kappa_2 - \kappa_1}, \tag{97}$$

for $\kappa_1 \neq \kappa_2$. If $\kappa_1 = \kappa_2$ the solution becomes,

$$\varphi_j = \tau_1 \kappa_1^j + \tau_2 j \kappa_1^{j-1}. \tag{98}$$

In this case it does not suffice that we know that the solution is bounded at the boundary point since two parameters, τ_1 and τ_2, need to be determined. Instead, we substitute the solution (98) into the boundary scheme, i.e the rows near the boundary that are altered from the the internal scheme. We will obtain a system of equations,

$$\mathbf{A}\bar{\tau} = \bar{0}, \tag{99}$$

where $\bar{\tau} = (\tau_1, \tau_2)^T$ and $\bar{0} = (0, 0)^T$. $\mathbf{A}$ is deduced from the scheme such that,

$$
\begin{aligned}
A_{11} &= 24/17 + \tilde{s} + 59/34\kappa_0 - 4/17\kappa_0^2 - 3/34\kappa_0^3, \\
A_{12} &= 24/17 + \tilde{s} + 59/34\kappa_1 - 4/17\kappa_1^2 - 3/34\kappa_1^3, \\
A_{21} &= -1/2 + \tilde{s}\kappa_0 + 1/2\kappa_0^2, \\
A_{22} &= -1/2 + s\kappa_1 + 1/2\kappa_1^2, \\
A_{31} &= 4/43 - 59/86\kappa_0 + \tilde{s}\kappa_0^2 + 59/86\kappa_0^3 - 4/43\kappa_0^4, \\
A_{32} &= 4/43 - 59/86\kappa_1 + \tilde{s}\kappa_1^2 + 59/86\kappa_1^3 - 4/43\kappa_1^4, \\
A_{41} &= 3/98 - 59/98\kappa_0^2 + \tilde{s}\kappa_0^3 + 32/49\kappa_0^4 - 4/49\kappa_0^5, \\
A_{42} &= 3/98 - 59/98\kappa_1^2 + \tilde{s}\kappa_1^3 + 32/49\kappa_1^4 - 4/49\kappa_1^5.
\end{aligned}
$$

The (4×2) matrix $\mathbf{A}$ has rank 2 so 2 rows may be deleted, since they are linear dependent to the others. We keep the first and third rows since they are linearly independent, and denote the resulting submatrix,

$$\mathbf{B} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}. \tag{100}$$

If the determinant of $\mathbf{B}$ is nonzero for all $Re\,\tilde{s} \geq 0$ (in this case this the determinant condition), (90) has no eigenvalues or generalised eigenvalues. In [5] the roots are derived for $\tilde{s}$ close to 0. The negative roots are,

$$\kappa_1 = -1 + \tilde{s} + \mathcal{O}(|\tilde{s}|^2), \quad \kappa_2 = 4 - \sqrt{15} + \tilde{s} + \mathcal{O}(|\tilde{s}|) \tag{101}$$

We have with (101) inserted,

$$det(\mathbf{B}) = -\frac{648}{833}\sqrt{15} + \frac{2280}{833} + \mathcal{O}(|\tilde{s}|) \tag{102}$$

Hence, for small $Re\,\tilde{s} \geq$ the determinant condition is satisfied. With numerical approximations to the fractions we have the series,

$$det(\mathbf{B}) \approx -0.27 - 4.48\tilde{s}..., \tag{103}$$

which clearly is nonzero for small $|\tilde{s}|$.

This was the case for small $|\tilde{s}|$. In [5] it is shown that for $|\tilde{s}|$ sufficiently large, there are no eigenvalues to (90). As mentioned before this is not a proof since there still might be eigenvalues for intermediate $|\tilde{s}|$. However, all experience from computations indicate that the scheme is pointwise stable. This also applies to schemes of higher order than four, where the analysis becomes even more algebraically complicated.

References

[1] B. Gustafsson. The convergence rate for difference approximations to mixed initial boundary value problems. *Math. Comp.*, 29(130):396–406, Apr. 1975.

[2] B. Gustafsson. The convergence rate for difference approximations to general mixed initial boundary value problems. *SIAM J. Numer. Anal.*, 18(2):179–190, Apr. 1981.

[3] Saul Abarbanel, Adi Ditkowski, and Bertil Gustafsson. On Error Bounds of Finite Difference Approximations to Partial Differential Equations-Temporal Behavior and Rate of Convergence. *Journal of Scientific Computing*, 2000.

[4] Ken Mattsson and Jan Nordström. Summation by parts operators for finite difference approximations of second derivatives. *J. Comput. Phys.*, 199(2), 2004.

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

[6] R.A. Horn and C.R. Johnson. *Matrix analysis.* Cambridge University Press, 1990.

[7] M. H. Carpenter, D. Gottlieb, and S. Abarbanel. Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: Methodology and application to high-order compact schemes. *J. Comput. Phys.*, 111(2), 1994.

[8] H.-O. Kreiss and G. Scherer. Finite element and finite difference methods for hyperbolic partial differential equations. *Mathematical Aspects of Finite Elements in Partial Differential Equations., Academic Press, Inc.*, 1974.

[9] H.-O. Kreiss and G. Scherer. On the existence of energy estimates for difference approximations for hyperbolic systems. Technical report, Dept. of Scientific Computing, Uppsala University, 1977.

[10] Bo Strand. Summation by Parts for Finite Difference Approximations for d/d x. *J. Comput. Phys.*, 110, 1994.

[11] M. H. Carpenter, J. Nordström, and D. Gottlieb. A Stable and Conservative Interface Treatment of Arbitrary Spatial Accuracy. *J. Comput. Phys.*, 148, 1999.

[12] J. Nordström and M. H. Carpenter. Boundary and interface conditions for high-order finite-difference methods applied to the Euler and Navier-Stokes equations. *J. Comput. Phys.*, 148, 1999.

[13] J. Nordström and M. H. Carpenter. High-order finite difference methods, multidimensional linear problems, and curvilinear coordinates. *J. Comput. Phys.*, 173, 2001.

[14] Ken Mattsson, Magnus Svärd, and Jan Nordström. Stable and Accurate Artificial Dissipation. *Journal of Scientific Computing*, 21, Aug. 2004.

[15] K. Mattsson, M. Svärd, J. Nordström, and M.H. Carpenter. Accuracy Requirements for Transient Aerodynamics. *AIAA paper 2003-3689*, 2003.

[16] Magnus Svärd. On coordinate transformation for summation-by-parts operators. *Journal of Scientific Computing*, 20, 2004.