Compact difference methods applied to initial-boundary value problems for mixed systems

Author(s):
Bodenmann, Rudolf Paul; Schroll, Hans Joachim

Publication Date:
1995

Permanent Link:
https://doi.org/10.3929/ethz-a-004284307

Compact difference methods applied to initial-boundary value problems for mixed systems

R. Bodenmann and H.J. Schroll [1]

Research Report No. 95-04
February 1995

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

[1]Numerische Mathematik, RWTH-Aachen, D-52056 Aachen, Germany

Compact difference methods applied to initial–boundary value problems for mixed systems

R. Bodenmann and H.J. Schroll [1]

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Research Report No. 95-04 February 1995

ABSTRACT. An initial-boundary value problem to a system of nonlinear partial differential equations, which consists of a hyperbolic and a parabolic part, is taken into consideration. The problem is discretised by a compact finite difference method. An approximation of the numerical solution is constructed, at which the difference scheme is linearised. Nonlinear convergence is proved using the stability of the linearised scheme. Finally, a computational experiment for a noncompact scheme is presented.

Keywords: Compact difference method, nonlinear hyperbolic-parabolic system, initial-boundary value problem, local stability, convergence.

Subject Classification: 35L65, 65M10

[1]Numerische Mathematik, RWTH-Aachen, D-52056 Aachen, Germany

1. Introduction

The question of existence and uniqueness of solutions of nonlinear hyperbolic and mixed systems is still an open problem. Strongly related to this question is the theory of numerical approximations.

In the present paper a class of implicit finite difference methods applied to initial-boundary value problems for the following mixed type systems is analysed:

$$
\begin{pmatrix} v \\ w \end{pmatrix}_t + \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix}_x = \begin{pmatrix} B_{11} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix}_{xx} + \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}. \tag{1}
$$

The coefficients are smooth matrix functions depending on the unknown $u = (v, w)^T \in \mathcal{R}^n$ and $(t, x) \in \Omega \subset \mathcal{R}^+ \times \mathcal{R}$. Neglecting the coupling terms, we assume the v-equation to be strongly parabolic and the system for w is supposed to be strictly hyperbolic. Initial-boundary value problems for mixed systems have been studied by several authors. We refer to the book by Kreiss and Lorenz ([2] chapter 7) and references therein, especially Strikwerda [9].

Examples of mixed systems are the compressible Navier Stokes equations and the viscous shallow water equations; cf. [1].

Implicit finite difference methods applied to pure Cauchy problems for mixed systems have been analysed in [7]. In order to compensate for the lack of stability of the nonlinear operators, a highly consistent and attracting approximation to the solution, the so-called pilot function, was applied. The ansatz for that pilot function goes back to Strang [8]. In [5], Michelson extended Strang's analysis for Cauchy problems for hyperbolic systems to initial-boundary value problems. In order to handle numerical boundary conditions, Michelson had to extend Strang's ansatz for the pilot function.

The present paper demonstrates that in the case of compact difference methods, which do not use numerical boundary conditions, the original pilot function leads to convergence results for initial-boundary value problems for nonlinear hyperbolic, parabolic, and mixed systems.

It should be mentioned here that we do not aim for an existence proof. In fact, we assume smooth initial data and compatible boundary conditions, such that a smooth solution exists on some finite time interval. This smooth solution is used to define the pilot function.

In the next section, the initial-boundary value problem, the numerical scheme, and the result are stated precisely. Furthermore, we give an outline of the argument and review the convergence theory. The pilot function will be constructed in the third section, while the fourth section provides linearised stability theory. Finally, in Section 5, we give an outlook concerning noncompact schemes.

2. Statement of the result and outline of the argument

To specify the boundary conditions for system (1), we make the following assumptions. The coefficients are smooth matrix functions:

$$A, B \in C^\infty((\mathcal{R}^n \times \Omega), \mathcal{R}^{n\times n}) \tag{2}$$

and C is a smooth vector function:

$$C \in C^\infty((\mathcal{R}^n \times \Omega), \mathcal{R}^n). \tag{3}$$

The system for $v \in \mathcal{R}^m$, $m \leq n$ is strongly parabolic, i.e.

$$B_{11}(u,t,x) \in C^\infty((\mathcal{R}^n \times \Omega), \mathcal{R}^{m\times m}), \quad B_{11} + B_{11}^T \geq 2\beta I > 0.$$

For symmetric matrices, the relation $A \leq B$ is defined by $u^T A u \leq u^T B u$ for all possible u.

Furthermore, the system for $w \in \mathcal{R}^{n-m}$ is supposed to be strictly hyperbolic in the sense that all eigenvalues of A_{22} are real and distinct. We want to point out that both the purely parabolic and the purely hyperbolic case are included in this setting.

We are especially interested in smooth solutions to the initial-boundary value problem for system (1). Therefore, the initial data is a smooth function

$$u(0,x) = z(x), \quad x \in [0,1], \quad z \in C^\infty([0,1], \mathcal{R}^n). \tag{4}$$

For the parabolic component, we assume to have Dirichlet boundary conditions at $x = 0$ and $x = 1$

$$v(t,0) = f_0(t), \quad v(t,1) = f_1(t), \quad t \geq 0, \quad f_i \in C^\infty(\mathcal{R}^+, \mathcal{R}^m). \tag{5}$$

Concerning the hyperbolic component, we can only prescribe the ingoing characteristic variables at each boundary. To formalise this, we assume that the eigenvalues of A_{22} are bounded away from zero by some constant γ. If we denote by R the matrix of right eigenvectors of A_{22}, then

$$\Lambda := R^{-1} A_{22} R = \Lambda^+ + \Lambda^-$$

can be split into a positive and a negative part, where

$$\Lambda^+ := \begin{pmatrix} \mathrm{diag}(\lambda_i \geq \gamma) & 0 \\ 0 & 0 \end{pmatrix}, \quad \Lambda^- := \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{diag}(\lambda_i \leq -\gamma) \end{pmatrix}.$$

The corresponding characteristic variables are

$$\begin{pmatrix} w^+ \\ w^- \end{pmatrix} := R^{-1} w.$$

If there are q positive eigenvalues, then

$$w^+(t,x) \in \mathcal{R}^q, \quad w^-(t,x) \in \mathcal{R}^{n-m-q}.$$

Ingoing characteristic variables at the left and right boundary are w^+ and w^- respectively. Proper boundary conditions are

$$\begin{aligned} &w^+(t,0) = g_0(t), \quad w^-(t,1) = g_1(t), \quad t \geq 0, \\ &g_0 \in C^\infty(\mathcal{R}^+, \mathcal{R}^q), \quad g_1 \in C^\infty(\mathcal{R}^+, \mathcal{R}^{n-m-q}). \end{aligned} \tag{6}$$

Furthermore, all the data must be compatible. If we assume that f_i, g_i, $(i = 0, 1)$, and z vanish in a neighborhood of the corners $(0, 0)$ and $(0, 1)$, then the existence of a smooth solution can be shown. The linear case is discussed in [2] chapter 7 and nonlinear problems can locally be solved by linearisation. Therefore, we make the following assumption.

Assumption 2.1. *Given the conditions above, there is a finite time $T > 0$ such that the initial-boundary value problem (1), (2), (3), (4), (5) and (6) has a unique smooth solution in $\Omega := [0, T] \times [0, 1]$.*

The particular finite difference scheme that will be analysed is implicit in time and uses central differences in space for the parabolic component. For the convective terms, an upwind technique is applied such that the overall scheme is compact, i.e. does not use so-called numerical boundary conditions. At the left boundary, v and w^+ are known and w^- has to be determined from the data and inner grid points. Therefore, we want to apply only forward differences in space for the w^- components. Similarly, at the right boundary w^+ will be computed by backward differences.

To be more concrete, let us introduce some more notation. Choosing step sizes $\Delta x = \frac{1}{J}$, $J \in \mathcal{N}$ and $\Delta t = \frac{T}{K}$, $K \in \mathcal{N}$, we define a grid

$$\begin{aligned} \Omega_h &:= \Omega_{\Delta t} \times \Omega_{\Delta x} \subset \Omega, \\ \Omega_{\Delta t} &:= \{t_k = k\Delta t,\ k = 0, 1, \ldots K\}, \\ \Omega_{\Delta x} &:= \{x_j = j\Delta x,\ j = 0, 1, \ldots J\}. \end{aligned}$$

For grid functions u on $\Omega_{\Delta x}$, we have the shift operator

$$(Eu)_j = u_{j+1} = u(x_{j+1}),$$

as well as forward, backward

$$D_+ = \frac{E - I}{\Delta x}, \quad D_- = \frac{I - E^{-1}}{\Delta x},$$

and central difference operators

$$D_0 = \frac{D_+ + D_-}{2}, \quad D_+ D_- = \frac{E - 2I + E^{-1}}{\Delta x^2}.$$

Next, we transform the mixed system (1) to characteristic form

$$\begin{pmatrix} v_t \\ R^{-1} w_t \end{pmatrix} + \begin{pmatrix} A_{11} & A_{12} R \\ R^{-1} A_{21} & \Lambda \end{pmatrix} \begin{pmatrix} v_x \\ R^{-1} w_x \end{pmatrix} = \begin{pmatrix} B_{11} v_{xx} \\ 0 \end{pmatrix} + \begin{pmatrix} C_1 \\ R^{-1} C_2 \end{pmatrix}$$

and split up $R^{-1}A_{21}$ according to the decomposition of Λ into

$$R^{-1}A_{21} = (R^{-1}A_{21})^{+} + (R^{-1}A_{21})^{-},$$

where the positive part consists of the first q rows of $R^{-1}A_{21}$ and $n-m-q$ zero rows. Furthermore, we denote by $(R^{-1})^{+}$ the matrix that consists of the first q rows of R^{-1} and $n-m-q$ zero rows and define $(R^{-1})^{-}$ such that $R^{-1} = (R^{-1})^{+} + (R^{-1})^{-}$. Using the notation

$$A_{22}^{\pm} = R\Lambda^{\pm}R^{-1},$$

the equation for w reads

$$[(R^{-1})^{+} + (R^{-1})^{-}]w_t + [(R^{-1}A_{21})^{+} + (R^{-1}A_{21})^{-}]v_x + R^{-1}(A_{22}^{+} + A_{22}^{-})w_x = R^{-1}C_2.$$

Now, it is possible to write down a compact upwind scheme. Initially, $u_{0,j} = z(x_j)$ is given. The step from t_{k-1} to t_k can be described as follows:

At the left boundary,

$$v_{k,0} = f_0(t_k) \text{ and } w_{k,0}^{+} = g_0(t_k) \tag{7}$$

are given. The unknown $w_{k,0}^{-}$ can be computed from

$$\begin{aligned} (R^{-1})_{k,0}^{-}\frac{w_{k,0} - w_{k-1,0}}{\Delta t} + (R^{-1}A_{21})_{k,0}^{-}(D_+v_k)_0 + R_{k,0}^{-1}(A_{22}^{-})_{k,0}(D_+w_k)_0 \\ = \quad (R^{-1})_{k,0}^{-}(C_2)_{k,0}. \end{aligned} \tag{8}$$

Similarly, at the right boundary, we solve

$$\begin{aligned} (R^{-1})_{k,J}^{+}\frac{w_{k,J} - w_{k-1,J}}{\Delta t} + (R^{-1}A_{21})_{k,J}^{+}(D_-v_k)_J + R_{k,J}^{-1}(A_{22}^{+})_{k,J}(D_-w_k)_J \\ = \quad (R^{-1})_{k,J}^{+}(C_2)_{k,J} \end{aligned} \tag{9}$$

under the conditions

$$v_{k,J} = f_1(t_k) \text{ and } w_{k,J}^{-} = g_1(t_k). \tag{10}$$

At inner grid points, we compose both boundary methods and solve

$$\begin{aligned} \frac{u_{k,j} - u_{k-1,j}}{\Delta t} + (A_1)_{k,j}(D_*u_k)_j + (A_2^{+})_{k,j}(D_-u_k)_j + (A_2^{-})_{k,j}(D_+u_k)_j \\ = \quad B_{k,j}(D_+D_-u_k)_j + C_{k,j}, \quad j = 1, 2, \dots J-1, \end{aligned} \tag{11}$$

where

$$A_1 = \begin{pmatrix} A_{11} & A_{12} \\ 0 & 0 \end{pmatrix}, \quad A_2^{\pm} = \begin{pmatrix} 0 & 0 \\ R(R^{-1}A_{21})^{\pm} & A_{22}^{\pm} \end{pmatrix},$$

and D_* may be one of the operators D_+, D_- or D_0.

Since the argument will be based on energy estimates for properly linearised systems, convergence will be measured in a discrete L^2-norm in space but uniformly in time. Throughout the paper, the following scalar products and norms are used: For real vectors, the Euklidian norm $|u|^2 = u^T u$ is induced by the product $< u, v >= u^T v$.

For grid functions on $\Omega_{\Delta x}$, we have the discrete L^2-norm $\|u\|_2^2 = \Delta x \sum_{x\in\Omega_{\Delta x}} |u(x)|^2$ with product $(u,v) = \Delta x \sum_{x\in\Omega_{\Delta x}} < u(x), v(x) >$, the discrete L^1-norm $\|u\|_1 = \Delta x \sum_{x\in\Omega_{\Delta x}} |u(x)|$, and the maximum norm $\|u\|_\infty = \max_{x\in\Omega_{\Delta x}} |u(x)|$. Grid functions on Ω_h will be measured by the combined norm $\|u\|_{\infty,2} = \max_{t\in\Omega_{\Delta t}} \|u(t,\cdot)\|_2$.

Now, we are in the position to state the convergence result:

Theorem 2.1. *Consider a mixed system (1), (2), (3), where B_{11} is strongly parabolic and A_{22} is strictly hyperbolic and regular. Let $\tilde{u}$ be the unique smooth solution of the initial-boundary value problem (1), (2), (3), (4), (5) and (6).*

For sufficiently small step sizes $\Delta t = \mu\Delta x$, where μ is arbitrarily but fixed, the implicit scheme (7), (8), (9), (10), (11) has a unique solution U defined on Ω_h. Furthermore, there is a smooth pilot function $u^{pi} \in C^\infty(\Omega, \mathcal{R}^n)$ such that

$$\|U - u^{pi}|_h\|_{\infty,2} = \mathcal{O}(\Delta x^3), \quad \Delta x \to 0.$$

By construction, the pilot function converges at first order to the desired solution. This implies

$$\|U - \tilde{u}|_h\|_{\infty,2} = \mathcal{O}(\Delta x), \quad \Delta x \to 0. \quad \Box$$

Before going into detail, let us outline the argument. For notational convenience, we assume that the hyperbolic component is already diagonal, $A_{22} = \Lambda$. Then, the vector of unknowns is partitioned as $u = (v, w^+, w^-)^T$. Furthermore, we denote by a tilde the non zero blocks of the matrices

$$A_2 = A_2^+ + A_2^- = \begin{pmatrix} 0 & 0 & 0 \\ \tilde{A}_{21}^+ & \tilde{\Lambda}^+ & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \tilde{A}_{21}^- & 0 & \tilde{\Lambda}^- \end{pmatrix}$$

and of the vectors

$$C_2 = C_2^+ + C_2^- = \begin{pmatrix} \tilde{C}_2^+ \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \tilde{C}_2^- \end{pmatrix}.$$

It will be convenient to write the scheme as a root equation

$$\Phi_h(u) = 0$$

for the discretisation operator

$$\Phi_h(u) = \begin{pmatrix} u_0 - z|_{\Delta x} \\ \{\Phi_h(u)_k\}_{k=1,2,\ldots K} \end{pmatrix}, \tag{12}$$

$$
(13) \qquad \Phi_h(u)_k = \left(\begin{array}{c} v_{k,0} - f_0(t_k) \\ w^+_{k,0} - g_0(t_k) \\ \frac{1}{\Delta t} \left(\begin{array}{c} w^-_{k,0} - w^-_{k-1,0} \\ \{u_{k,j} - u_{k-1,j}\}_{j=1,2,\dots J-1} \\ w^+_{k,J} - w^+_{k-1,J} \end{array} \right) + P_{\Delta x}(u_k) \\ w^-_{k,J} - g_1(t_k) \\ v_{k,J} - f_1(t_k) \end{array} \right)_{k=1,2,\dots K}.
$$

Here, $h = (\Delta t, \Delta x)^T$ and $P_{\Delta x}$ is the spatial discretisation operator which is defined by

$$
(14) \qquad P_{\Delta x}(u_k) = \left(\begin{array}{c} (\tilde{A}^-_{21})_{k,0}(D_+ v_k)_0 + \tilde{\Lambda}^-_{k,0}(D_+ w^-_k)_0 - (\tilde{C}^-_2)_{k,0} \\ \left(\begin{array}{r} (A_1)_{k,j}(D_* u)_{k,j} + (A^+_2)_{k,j}(D_- u_k)_j + (A^-_2)_{k,j}(D_+ u_k)_j \\ -B_{k,j}(D_+ D_- u_k)_j - C_{k,j} \end{array} \right) \\ (\tilde{A}^+_{21})_{k,J}(D_+ v_k)_J + \tilde{\Lambda}^+_{k,J}(D_- w^+_k)_J - (\tilde{C}^+_2)_{k,J} \end{array} \right).
$$

Again, the middle line has to be repeated for $j = 1, 2, \dots J-1$.

In this situation, the stability theory by López-Marcos and Sanz-Serna applies. We denote by $\mathcal{B}(u, R)$ the open ball of radius R centered at u. The concept can be summarised as follows.

Theorem 2.2. *Assume there is a smooth function $u^{pi} = u^{pi}(t, x, \Delta x)$ such that*

i) *$D\Phi_h$ is Lipschitz continuous near u^{pi}*

$$
\|D\Phi_h(u^{pi}|_h) - D\Phi_h(u)\|_{\infty,2} \leq Lip_h \|u^{pi}|_h - u\|_{\infty,2}, \quad u \in \mathcal{B}(u^{pi}|_h, C_h).
$$

ii) *The scheme linearised at u^{pi} is stable, i.e. $D\Phi_h(u^{pi}|_h)$ is regular and*

$$
\|D\Phi_h^{-1}(u^{pi}|_h)\|_{\infty,2} \leq L.
$$

Then, there is a constant $S > L$ such that Φ_h is locally stable

$$
\|u - v\|_{\infty,2} \leq S\|\Phi_h(u) - \Phi_h(v)\|_{\infty,2}, \quad u, v \in \mathcal{B}(u^{pi}|_h, R_h)
$$

with stability radius $R_h = \min(C_h, (L^{-1} - S^{-1})/Lip_h)$.

If, furthermore, Φ_h is consistent with u^{pi}

$$
(15) \qquad \|\Phi_h(u^{pi}|_h)\|_{\infty,2} = o(R_h), \quad h \to 0,
$$

then, there is (for sufficiently small h) a unique solution U of $\Phi_h(u) = 0$ in $\mathcal{B}(u^{pi}|_h, R_h)$ and

$$
\|U - u^{pi}|_h\|_{\infty,2} \leq S\|\Phi_h(u^{pi}|_h)\|_{\infty,2}. \quad \square
$$

For the proof and discussion of this result, we refer to [3], [4] and [6].

In order to prove convergence towards $\tilde{u}$, one would like to apply Theorem 2.2 directly to $u^{pi} = \tilde{u}$, but this is not possible because our scheme is only first order consistent with $\tilde{u}$ and the stability radius is of order $r > 2$ in Δx.

The reason for this is that the Lipschitz constant Lip_h depends on negative powers of Δx. Obviously, $\Phi_h(u)$ involves terms like Δt^{-1}, Δx^{-1} and Δx^{-2}, which are introduced by finite difference expressions. Since time differences appear linearly with respect to u and the Jacobian $D\Phi_h(u)$ is block three diagonal, it is possible (cf. [7]) to show an estimate

$$Lip_h \leq \frac{C}{\Delta x^{2.5}}\,. \tag{16}$$

Consequently, the stability radius is of order $R_h = \mathcal{O}(\Delta x^{2.5})$ and condition (15) requires a pilot function u^{pi} that is third order consistent with Φ_h. This pilot function is constructed in the next section.

3. The pilot function

In [8] Strang constructed a high order approximate solution to a difference scheme for a hyperbolic Cauchy problem by the ansatz

$$u^{pi} = \tilde{u} + \Delta x u^{(1)} + \Delta x^2 u^{(2)} + \dots \tag{17}$$

In [7] the same ansatz was applied to Cauchy problems for mixed systems.

The procedure to define the error terms is as follows:

- Substitute the ansatz (17) into Φ_h,
- expand with respect to the step size,
- set the coefficients of Δx and Δx^2 to zero.

The result are linear initial-boundary value problems that define $u^{(1)}$ and $u^{(2)}$. The point is to make sure that the conditions obtained by the interior scheme are compatible with the conditions arising at the boundary. This is the case here since the present scheme applies the interior difference formula on the boundary as well, using the given data.

From the initial conditions it follows

$$u^{(1)}(0,x) = u^{(2)}(0,x) = 0. \tag{18}$$

At inner grid points x_j, $j = 1, 2, \dots J-1$ the difference equation is

$$\begin{aligned}
&\frac{u^{pi}_{k,j} - u^{pi}_{k-1,j}}{\Delta t} + A_1(u^{pi}_{k,j})(D_* u^{pi}_k)_j + A_2^+(u^{pi}_{k,j})(D_- u^{pi}_k)_j + A_2^-(u^{pi}_{k,j})(D_+ u^{pi}_k)_j \\
&\quad = \; B(u^{pi}_{k,j})(D_+ D_- u^{pi}_k)_j + C(u^{pi}_{k,j}).
\end{aligned}$$

If D_* represents D_+ or D_- respectively, the resulting systems for $u^{(1)}$ and $u^{(2)}$ are:

$$
\begin{aligned}
(19)\qquad & u^{(1)}{}_t + A(\tilde{u})u^{(1)}{}_x + DA(\tilde{u})u^{(1)}\tilde{u}_x + 1/2\,(A_2^-(\tilde{u}) - A_2^+(\tilde{u}) \pm A_1(\tilde{u}))\tilde{u}_{xx} \\
&\quad = \; B(\tilde{u})u^{(1)}{}_{xx} + DB(\tilde{u})u^{(1)}\tilde{u}_{xx} + DC(\tilde{u})u^{(1)} + \mu/2\,\tilde{u}_{tt},
\end{aligned}
$$

$$
\begin{aligned}
(20)\qquad & u^{(2)}{}_t + A(\tilde{u})(u^{(2)}{}_x + 1/6\,\tilde{u}_{xxx}) \\
&\quad + DA(\tilde{u})u^{(1)}u^{(1)}{}_x + DA(\tilde{u})u^{(2)}\tilde{u}_x + 1/2\,D^2A(\tilde{u})u^{(1)}u^{(1)}\tilde{u}_x \\
&\quad + 1/2\,(A_2^-(\tilde{u}) - A_2^+(\tilde{u}) \pm A_1(\tilde{u}))u^{(1)}{}_{xx} \\
&\quad + 1/2\,(DA_2^-(\tilde{u}) - DA_2^+(\tilde{u}) \pm DA_1(\tilde{u}))u^{(1)}\tilde{u}_{xx} \\
&= \; B(\tilde{u})(u^{(2)}{}_{xx} + 1/12\,\tilde{u}_{xxxx}) \\
&\quad + DB(\tilde{u})u^{(1)}u^{(1)}{}_{xx} + DB(\tilde{u})u^{(2)}\tilde{u}_{xx} + 1/2\,D^2B(\tilde{u})u^{(1)}u^{(1)}\tilde{u}_{xx} \\
&\quad + DC(\tilde{u})u^{(2)} + 1/2\,D^2C(\tilde{u})u^{(1)}u^{(1)} + \mu/2\,u^{(1)}{}_{tt} - \mu^2/6\,\tilde{u}_{ttt}.
\end{aligned}
$$

Here, we have used $\Delta t = \mu\Delta x$. If $D_* = D_0$ the $\pm$-terms are zero.

Next, we consider the left boundary. Using the notation $u^{pi} = (v^{pi}, w^{pi+}, w^{pi-})^T$, $u^{(l)} = (v^{(l)}, w^{(l)+}, w^{(l)-})^T$, $l = 1, 2$ and the boundary conditions it follows

$$
(21)\qquad
\begin{aligned}
v^{(1)}(t_k, 0) &= v^{(2)}(t_k, 0) = 0, \\
w^{(1)+}(t_k, 0) &= w^{(2)+}(t_k, 0) = 0.
\end{aligned}
$$

The solution on the outgoing characteristics is determined by the equation

$$
\frac{w^{pi-}_{k,0} - w^{pi-}_{k-1,0}}{\Delta t} + \tilde{A}^-_{21}(u^{pi}_{k,0})(D_+v^{pi}_k)_0 + \tilde{\Lambda}^-(u^{pi}_{k,0})(D_+w^{pi-}_k)_0 = \tilde{C}^-_2(u^{pi}_{k,0}).
$$

The expansion leads to:

$$
\begin{aligned}
& w^{(1)-}_t + \tilde{A}^-_{21}(\tilde{u})(v^{(1)}_x + 1/2\,\tilde{v}_{xx}) + D\tilde{A}^-_{21}(\tilde{u})u^{(1)}\tilde{v}_x \\
&\quad + \tilde{\Lambda}^-(\tilde{u})(w^{(1)-}_x + 1/2\,\tilde{w}^-_{xx}) + D\tilde{\Lambda}^-(\tilde{u})u^{(1)}\tilde{w}^-_x \\
&= \; D\tilde{C}^-_2(\tilde{u})u^{(1)} + \mu/2\,\tilde{w}^-_{tt},
\end{aligned}
$$

$$
\begin{aligned}
& w^{(2)-}_t + \tilde{A}^-_{21}(\tilde{u})(v^{(2)}_x + 1/2\,v^{(1)}_{xx} + 1/6\,\tilde{v}_{xxx}) \\
&\quad + D\tilde{A}^-_{21}(\tilde{u})u^{(1)}(v^{(1)}_x + 1/2\,\tilde{v}_{xx}) + D\tilde{A}^-_{21}(\tilde{u})u^{(2)}\tilde{v}_x + 1/2\,D^2\tilde{A}^-_{21}(\tilde{u})u^{(1)}u^{(1)}\tilde{v}_x \\
&\quad + \tilde{\Lambda}^-(\tilde{u})(w^{(2)-}_x + 1/2\,w^{(1)-}_{xx} + 1/6\,\tilde{w}^-_{xxx}) \\
&\quad + D\tilde{\Lambda}^-(\tilde{u})u^{(1)}(w^{(1)-}_x + 1/2\,\tilde{w}^-_{xx}) + D\tilde{\Lambda}^-(\tilde{u})u^{(2)}\tilde{w}^-_x + 1/2\,D^2\tilde{\Lambda}^-(\tilde{u})u^{(1)}u^{(1)}\tilde{w}^-_x \\
&= \; D\tilde{C}^-_2(\tilde{u})u^{(2)} + 1/2\,D^2\tilde{C}^-_2(\tilde{u})u^{(1)}u^{(1)} + \mu/2\,w^{(1)-}_{tt} - \mu^2/6\,\tilde{w}^-_{ttt}.
\end{aligned}
$$

Finally, we find that these systems are exactly the $w^{(1)-}$- and $w^{(2)-}$-blocks of (19) and (20) respectively.

A similar argument for the right boundary results in

$$v^{(1)}(t_k,1) = v^{(2)}(t_k,1) = 0, \quad w^{(1)-}(t_k,1) = w^{(2)-}(t_k,1) = 0, \tag{22}$$

and the appropriate equations, corresponding to (19) and (20).

Since (19) and (20) are linear systems, the corresponding homogeneous initial-boundary value problem is well posed (cf. [2]). Let us summarise this section:

Lemma 3.1. *If $u^{(1)}$ solves the initial-boundary value problem (18), (21), (22) and (19) and $u^{(2)}$ is defined by (18), (21), (22) and (20), then the pilot function*

$$u^{pi} = \tilde{u} + \Delta x u^{(1)} + \Delta x^2 u^{(2)}$$

is third order consistent with (12), (13) and (14). □

In order to apply the convergence result in Theorem 2.2, the stability of the scheme linearised at u^{pi} has to be verified.

4. Linearised stability

The nonlinear scheme $\Phi_h(u) = 0$ linearised at u^{pi} reads

$$\Phi_h(u^{pi}) + D\Phi_h(u^{pi})(u - u^{pi}) = 0.$$

The coefficients of this scheme depend on x and t as well as on the mesh parameter Δx. It is well known, that the linearised scheme is stable if and only if there exist constants h_0 and L such that $D\Phi_h(u^{pi})$ is regular and the inverse is uniformly bounded $\|D\Phi_h^{-1}(u^{pi})\|_{\infty,2} \leq L$ for all $\Delta t \leq \Delta t_0$ and $\Delta x \leq \Delta x_0$. Consequently, the linear scheme is stable, if and only if the corresponding homogeneous scheme $D\Phi_h(u^{pi})u = 0$ is stable. In this section, we therefore have to treat homogeneous problems with zero initial- and boundary data.

Since often the boundaries must be treated separately, we need some more notation:

$$(u,v)^{(l,r)} := \Delta x \sum_{j=l}^{r} < u_j, v_j > \quad \text{and} \quad (\|u\|^{(l,r)})^2 := (u,u)^{(l,r)}.$$

For any matrix function A and any vector function f on $\Omega_{\Delta x}$, we have the following discrete analogues to Leibnitz' rule:

$$D_+(Af) = AD_+f + (D_+A)Ef \quad \text{and} \quad D_-(Af) = AD_-f + (D_-A)E^{-1}f. \tag{23}$$

For any two vector functions f and g on $\Omega_{\Delta x}$, one can easily prove the summation by parts formulae

$$(f, D_+g)^{(0,J-1)} = -(D_-f,g)^{(1,J)} + < f,g > |_0^J \quad \text{and} \quad (f, D_-g)^{(1,J)} = -(D_+f,g)^{(0,J-1)} + < f,g > |_0^J \tag{24}$$

respectively

$$(25) \qquad \begin{aligned} (f, D_+ g)^{(1,J-1)} &= -(D_- f, g)^{(1,J-1)} + < f_{J-1}, g_J > - < f_0, g_1 > \quad \text{and} \\ (f, D_- g)^{(1,J-1)} &= -(D_+ f, g)^{(1,J-1)} + < f_J, g_{J-1} > - < f_1, g_0 > . \end{aligned}$$

Furthermore, we will use a simple version of Young's inequality:

$$(26) \qquad ab \le \frac{\sigma^2 a^2}{2} + \frac{b^2}{2\sigma^2}, \quad \sigma, a, b \in \mathcal{R}.$$

4.1. Strongly parabolic problems. Let us consider the following discretisation of a strongly parabolic problem:

$$\frac{v_{k,j} - v_{k-1,j}}{\Delta t} = B(t_k, x_j, \Delta x)(D_+ D_- v_k)_j + C(t_k, x_j, \Delta x) v_{k,j}, \quad B + B^T \ge 2\beta,$$
$$k = 1, 2, \dots K, \; j = 1, 2, \dots J - 1.$$

As mentioned above, it is sufficient to investigate homogeneous initial- and boundary data

$$(27) \qquad \begin{aligned} v_{0,j} &= 0, & j &= 0, 1, \dots J, \\ v_{k,0} &= v_{k,J} = 0, & k &= 0, 1, \dots K. \end{aligned}$$

With $(P_{\Delta x}(t_k) v_k)_j := -B(t_k, x_j, \Delta x)(D_+ D_- v_k)_j - C(t_k, x_j, \Delta x) v_{k,j}$, we get

$$(28) \qquad \begin{aligned} \|v_{k-1}\|_2^2 &= (\|v_k + \Delta t P_{\Delta x}(t_k) v_k\|^{(1,J-1)})^2 \\ &\ge \|v_k\|_2^2 + 2\Delta t (v_k, P_{\Delta x}(t_k) v_k)^{(1,J-1)}. \end{aligned}$$

Furthermore, by omitting the time variable, we have

$$-2(v, P_{\Delta x} v)^{(1,J-1)} = (v, B D_+ D_- v)^{(1,J-1)} + (v, B D_- D_+ v)^{(1,J-1)} + 2(v, Cv)^{(1,J-1)}.$$

Applying the discrete product rule (23) and the summation by parts formula (25), and by making use of the boundary conditions, we get

$$(29) \qquad \begin{aligned} &(v, B D_+ D_- v)^{(1,J-1)} \\ &= (v, D_+(B D_- v))^{(1,J-1)} - (v, (D_+ B) E D_- v)^{(1,J-1)} \\ &= -(D_- v, B D_- v)^{(1,J-1)} - (v, (D_+ B) E D_- v)^{(1,J-1)} \\ &\quad + < v_{J-1}, B_J D_- v_J > - < v_0, B_1 D_- v_1 > \\ &= -(D_- v, B D_- v)^{(1,J-1)} - (v, (D_+ B) E D_- v)^{(1,J-1)} \\ &\quad - \Delta x < D_- v_J, B_J D_- v_J > \\ &\le -\beta(\|D_- v\|^{(1,J)})^2 + c_1 \|v\|^{(1,J-1)} \|D_+ v\|^{(1,J-1)}. \end{aligned}$$

Similarly, it follows

$$(30) \qquad \begin{aligned} &(v, B D_- D_+ v)^{(1,J-1)} \\ &\le -\beta(\|D_+ v\|^{(0,J-1)})^2 + c_2 \|v\|^{(1,J-1)} \|D_- v\|^{(1,J-1)}. \end{aligned}$$

Since C is a smooth matrix function, it is obvious that

$$|(v, Cv)| \leq c_3 \|v\|^2. \tag{31}$$

(29), (30) and (31) are used in (28) and Young's inequality leads to

$$\|v_k\|_2^2 \leq (1 + \mathcal{O}(\Delta t))\|v_{k-1}\|_2^2.$$

4.2. Strictly hyperbolic problems. In the next step, we want to develop analogous estimates for a strictly hyperbolic linear problem that is discretised by a compact finite difference method

$$\begin{aligned} &\frac{w_{k,j} - w_{k-1,j}}{\Delta t} + \Lambda^+(t_k, x_j, \Delta x)(D_- w_k)_j + \Lambda^-(t_k, x_j, \Delta x)(D_+ w_k)_j \\ &= \ C(t_k, x_j, \Delta x) w_{k,j}, \quad k = 1, 2, \ldots K, \ j = 1, 2, \ldots J-1, \end{aligned}$$

with zero initial and boundary data:

$$\begin{aligned} &w_{0,j} = 0, \qquad j = 0, 1, \ldots J, \\ &\left.\begin{aligned} w_{k,0}^+ = 0, \\ w_{k,J}^- = 0, \end{aligned}\right\} \quad k = 0, 1, \ldots K. \end{aligned} \tag{32}$$

At internal points, the method reads

$$w_{k-1,j} = w_{k,j} + \Delta t \cdot (P_{\Delta x}(t_k) w_k)_j, \quad j = 1, 2, \ldots J-1,$$

where

$$(P_{\Delta x}(t_k) w_k)_j = \Lambda^+(t_k, x_j, \Delta x)(D_- w_k)_j + \Lambda^-(t_k, x_j, \Delta x)(D_+ w_k)_j - C(t_k, x_j, \Delta x) w_{k,j}.$$

On the outflow boundaries the equations are:

$$\begin{aligned} w_{k-1,0}^- &= w_{k,0}^- + \Delta t \{\tilde{\Lambda}^-(t_k, x_0, \Delta x)(D_+ w_k^-)_0 - \tilde{C}^-(t_k, x_0, \Delta x) w_{k,0}\} \text{ and} \\ w_{k-1,J}^+ &= w_{k,J}^+ + \Delta t \{\tilde{\Lambda}^+(t_k, x_J, \Delta x)(D_- w_k^+)_J - \tilde{C}^+(t_k, x_J, \Delta x) w_{k,J}\}. \end{aligned}$$

Now we have

$$\begin{aligned} \|w_{k-1}\|_2^2 &= \left(\|w_k + \Delta t P_{\Delta x}(t_k) w_k\|^{(1,J-1)}\right)^2 + \Delta x \left(|w_{k-1,0}|^2 + |w_{k-1,J}|^2\right) \\ &= \|w_k\|_2^2 + 2\Delta t\, (w_k, P_{\Delta x}(t_k) w_k)^{(1,J-1)} + \Delta t^2 (\|P_{\Delta x}(t_k) w_k\|^{(1,J-1)})^2 \\ &\quad + \Delta x (|w_{k-1,0}^-|^2 - |w_{k,0}^-|^2 + |w_{k-1,J}^+|^2 - |w_{k,J}^+|^2) \\ &\geq \|w_k\|_2^2 + 2\Delta t \Big\{(w_k, P_{\Delta x}(t_k) w_k)^{(1,J-1)} \\ &\quad + \Delta x < w_{k,0}^-, \tilde{\Lambda}_{k,0}^-(D_+ w_k^-)_0 - \tilde{C}_{k,0}^- w_{k,0} > \\ &\quad + \Delta x < w_{k,J}^+, \tilde{\Lambda}_{k,J}^+(D_- w_k^+)_J - \tilde{C}_{k,J}^+ w_{k,J} > \Big\} \\ &= \|w_k\|_2^2 + 2\Delta t \{(w_k^-, \tilde{\Lambda}_k^- D_+ w_k^-)^{(0,J-1)} + (w_k^+, \tilde{\Lambda}_k^+ D_- w_k^+)^{(1,J)} \\ &\quad - (w_k, C_k w_k)\} \end{aligned} \tag{33}$$

By using summation by parts (24) and the discrete product rule (23) and by omitting the time index k, we continue:

$$
\begin{aligned}
&-2(w^-, \tilde{\Lambda}^- D_+ w^-)^{(0,J-1)} \\
&\quad = -(w^-, \tilde{\Lambda}^- D_+ w^-)^{(0,J-1)} + (D_- w^-, \tilde{\Lambda}^- w^-)^{(1,J)} \\
&\qquad + (w^-, (D_+ \tilde{\Lambda}^-) E w^-)^{(0,J-1)} - < w^-, \tilde{\Lambda}^- w^- > |_0^J \\
&\quad = (w^-, -\tilde{\Lambda}^- (D_+ - D_-) w^-)^{(1,J-1)} + (w^-, (D_+ \tilde{\Lambda}^-) E w^-)^{(0,J-1)} \\
&\qquad + \Delta x (< w_J^-, \tilde{\Lambda}_J^- D_- w_J^- > - < w_0^-, \tilde{\Lambda}_0^- D_+ w_0^- >) \\
&\qquad - < w^-, \tilde{\Lambda}^- w^- > |_0^J .
\end{aligned}
$$

Similarly we have

$$
\begin{aligned}
-2(w^+, \tilde{\Lambda}^+ D_- w^+)^{(1,J)} &= (w^+, \tilde{\Lambda}^+ (D_+ - D_-) w^+)^{(1,J-1)} + (w^+, (D_- \tilde{\Lambda}^+) E^{-1} w^+)^{(1,J)} \\
&\quad - \Delta x (< w_J^+, \tilde{\Lambda}_J^+ D_- w_J^+ > - < w_0^+, \tilde{\Lambda}_0^+ D_+ w_0^+ >) \\
&\quad - < w^+, \tilde{\Lambda}^+ w^+ > |_0^J .
\end{aligned}
$$

Since $D_+ - D_- = \Delta x D_+ D_- = \Delta x D_- D_+$, $\tilde{\Lambda}^+ = (\tilde{\Lambda}^+)^T > \gamma I$ and $-\tilde{\Lambda}^- = -(\tilde{\Lambda}^-)^T > \gamma I$, the new terms can be treated as in the parabolic case and it follows

$$
\begin{aligned}
&-2(w^-, \tilde{\Lambda}^- D_+ w^-)^{(0,J-1)} \\
&\quad = \frac{\Delta x}{2} \Big\{ (D_- w^-, \tilde{\Lambda}^- D_- w^-)^{(1,J)} + (w^-, (D_+ \tilde{\Lambda}^-)(D_+ w^-))^{(0,J-1)} \\
&\qquad + (D_+ w^-, \tilde{\Lambda}^- D_+ w^-)^{(0,J-1)} + (w^-, (D_- \tilde{\Lambda}^-)(D_- w^-))^{(1,J)} \Big\} \\
&\qquad + (w^-, (D_+ \tilde{\Lambda}^-) E w^-)^{(0,J-1)} - < w^-, \tilde{\Lambda}^- w^- > |_0^J \\
&\quad \le -\frac{\gamma}{2} \Big\{ (\|D_- w^-\|^{(1,J)})^2 + (\|D_+ w^-\|^{(0,J-1)})^2 \Big\} \\
&\qquad + c_4 \left(\|w^-\|^{(0,J-1)} \|D_+ w^-\|^{(0,J-1)} + \|w^-\|^{(1,J)} \|D_- w^-\|^{(1,J)} \right) \\
&\qquad + c_5 \|w^-\|^{(0,J-1)} \|w^-\|^{(1,J)} - < w^-, \tilde{\Lambda}^- w^- > |_0^J ,
\end{aligned}
$$

where we have used that $\Delta x \in (0, 1)$. Again Young's inequality (26) and the boundary conditions imply

$$
\begin{aligned}
-2(w^-, \tilde{\Lambda}^- D_+ w^-)^{(0,J-1)} &\le c_6 \Big\{ (\|w^-\|^{(0,J-1)})^2 + (\|w^-\|^{(1,J)})^2 \Big\} - < w^-, \tilde{\Lambda}^- w^- > |_0^J \\
&\le c_6 \Big\{ (\|w^-\|^{(0,J-1)})^2 + (\|w^-\|^{(1,J)})^2 \Big\} \qquad (34)
\end{aligned}
$$

and similarly

$$
-2(w^+, \tilde{\Lambda}^+ D_- w^+)^{(1,J)} \le c_7 \Big\{ (\|w^+\|^{(0,J-1)})^2 + (\|w^+\|^{(1,J)})^2 \Big\} . \qquad (35)
$$

Finally, (31), (33), (34) and (35) result in

$$\|w_k\|_2^2 \le (1+\mathcal{O}(\Delta t))\|w_{k-1}\|_2^2.$$

4.3. Mixed systems. We are now ready to treat the linearised version of mixed type systems from Section 2. For $j = 1, 2, \dots, J-1$, let $P_{\Delta x}$ be defined as

$$\begin{aligned}(P_{\Delta x}(t_k)u_k)_j &= (A_1)_{k,j}(D_* u_k)_j + (A_2^+)_{k,j}(D_- u_k)_j + (A_2^-)_{k,j}(D_+ u_k)_j \\ &\quad - B_{k,j}(D_+ D_- u_k)_j - C_{k,j} u_{k,j}.\end{aligned}$$

Here, the matrices depend on t, x and Δx but not on u. For the parabolic part, we assume to have boundary conditions (27) and for the hyperbolic part (32). Then we have

$$\begin{aligned}\|u_{k-1}\|_2^2 &= (\|u_k + \Delta t P_{\Delta t}(t_k)u_k\|^{(1,J-1)})^2 + \Delta x(|u_{k-1,0}|^2 + |u_{k-1,J}|^2) \\ &= \|u_k\|_2^2 + 2\Delta t(u_k, P_{\Delta x}(t_k)u_k)^{(1,J-1)} + \Delta t^2(\|P_{\Delta x}(t_k)u_k\|^{(1,J-1)})^2 \\ &\quad + \Delta x(|u_{k-1,0}|^2 - |u_{k,0}|^2 + |u_{k-1,J}|^2 - |u_{k,J}|^2) \\ &= \|u_k\|_2^2 + 2\Delta t(u_k, P_{\Delta x}(t_k)u_k)^{(1,J-1)} + \Delta t^2(\|P_{\Delta x}(t_k)u_k\|^{(1,J-1)})^2 \\ &\quad + \Delta x(|w_{k-1,0}^-|^2 - |w_{k,0}^-|^2 + |w_{k-1,J}^+|^2 - |w_{k,J}^+|^2) \\ &\ge \|u_k\|_2^2 + 2\Delta t\Big\{(u_k, P_{\Delta x}(t_k)u_k)^{(1,J-1)} \\ &\quad + \ \Delta x < w_{k,0}^-, (\tilde{A}_{21}^-)_{k,0}D_+ v_{k,0} + \tilde{\Lambda}_{k,0}^- D_+ w_{k,0}^- - (\tilde{C}_2^-)_{k,0}u_{k,0} > \\ &\quad + \ \Delta x < w_{k,J}^+, (\tilde{A}_{21}^+)_{k,J}D_- v_{k,J} + \tilde{\Lambda}_{k,J}^+ D_- w_{k,J}^+ - (\tilde{C}_2^+)_{k,J}u_{k,J} >\Big\}.\end{aligned}$$

We need to estimate the following terms:

$$\begin{aligned}&-(u, P_{\Delta x}u)^{(1,J-1)} \\ &\qquad -\Delta x < w_0^-, (\tilde{A}_{21}^-)_0 D_+ v_0 + \tilde{\Lambda}_0^- D_+ w_0^- - (\tilde{C}_2^-)_0 u_0 > \\ &\qquad - \Delta x < w_J^+, (\tilde{A}_{21}^+)_J D_- v_J + \tilde{\Lambda}_J^+ D_- w_J^+ - (\tilde{C}_2^+)_J u_J > \\ &= \underbrace{\frac{1}{2}(v, B_{11}D_+D_-v)^{(1,J-1)}}_{I} \underbrace{+\frac{1}{2}(v, B_{11}D_-D_+v)^{(1,J-1)}}_{II} \\ &\qquad \underbrace{-(v, A_{11}D_*v)^{(1,J-1)}}_{III} \underbrace{-\ (v, A_{12}D_*w)^{(1,J-1)}}_{IV} \\ &\qquad \underbrace{-(w^-, \tilde{A}_{21}^- D_+ v)^{(0,J-1)}}_{V} \underbrace{-(w^+, \tilde{A}_{21}^+ D_- v)^{(1,J)}}_{VI} \\ &\qquad \underbrace{-(w^-, \tilde{\Lambda}^- D_+ w^-)^{(0,J-1)}}_{VII} \underbrace{-\ (w^+, \tilde{\Lambda}^+ D_- w^+)^{(1,J)}}_{VIII} \underbrace{+(u, Cu)}_{IX}.\end{aligned}$$

For I and II we already have the estimates in (29) and (30), for VII and $VIII$ we have (34) and (35) respectively, and to IX we apply (31). The remaining terms are estimated as follows:

$$\begin{aligned} III &\leq c_8\|v\|^{(1,J-1)}(\|D_+v\|^{(1,J-1)} + \|D_-v\|^{(1,J-1)}) \\ V &\leq c_9\|w^-\|^{(0,J-1)}\|D_+v\|^{(0,J-1)} \\ VI &\leq c_{10}\|w^+\|^{(1,J)}\|D_-v\|^{(1,J)} \end{aligned}$$

To get a useful estimate for term IV, we apply again (23) and (25). In this way, we manage to differentiate v instead of w:

$$\begin{aligned} -(v, A_{12}D_+w)^{(1,J-1)} &= (D_-v, A_{12}w)^{(1,J)} + (v, (D_+A_{12})Ew)^{(1,J-1)} \\ -(v, A_{12}D_-w)^{(1,J-1)} &= (D_+v, A_{12}w)^{(0,J-1)} + (v, (D_-A_{12})E^{-1}w)^{(1,J-1)}. \end{aligned}$$

Since $-(v, A_{12}D_0w)^{(1,J-1)} = -1/2\{(v, A_{12}D_+w)^{(1,J-1)} + (v, A_{12}D_-w)^{(1,J-1)}\}$, it follows

$$IV \leq c_{11}\|w\|^{(0,J)}\left(\|D_-v\|^{(1,J)} + \|D_+v\|^{(0,J-1)} + \|v\|^{(1,J-1)}\right)$$

and by Young's inequality (26), we get

$$\|u_k\|_2^2 \leq (1 + \mathcal{O}(\Delta t))\|u_{k-1}\|_2^2.$$

From this, it can easily be seen that $D\Phi_h(u^{pi})^{-1}$ exists and is uniformly bounded. Therefore, the following lemma is proved:

Lemma 4.1. *The difference scheme (12), (13) and (14) linearised at the pilot function*

$$\Phi_h(u^{pi}) + D\Phi_h(u^{pi})(u - u^{pi}) = 0$$

is stable with respect to $\|\cdot\|_{\infty,2}$. □

The proof of the main result, Theorem 2.1, follows from Lemma 3.1, the estimate (16), Lemma 4.1 and Theorem 2.2.

5. A noncompact scheme

In the previous sections, we treated compact schemes, which is the reason why we did not need to introduce artificial boundary conditions. But, as soon as we want to use higher order methods in space direction, we are obliged to investigate noncompact schemes and therefore have to introduce artificial boundary conditions. It is well known that artificial boundary conditions lead to boundary layers, cf. [5]. If we want to construct the pilot function to such a numerical solution, we have to make sure that these layers are approximated as well. Since the phenomenon of boundary layer occurrence can already be observed in linear problems, we focus on the linear advection equation:

$$
\begin{aligned}
(36)\qquad & u_t + a u_x &=& \ 0, \quad a > 0, \\
(37)\qquad & u(t,0) &=& \ 0, \\
& u(0,x) &=& \ \sin^3(\pi x), \quad x \in [0,1].
\end{aligned}
$$

Clearly, this initial boundary value problem has the solution $\tilde{u}(t,x) = \sin^3(\pi(x-at))$. We discretise the problem by an implicit scheme that is consistent of third order in space direction:

$$
\frac{u_{k,j} - u_{k-1,j}}{\Delta t} + a \frac{2u_{k,j+1} + 3u_{k,j} - 6u_{k,j-1} + u_{k,j-2}}{6\Delta x} = 0. \tag{38}
$$

These equations are well defined at grid points away from the boundary, i.e. for $j = 2, 3, \ldots J-1$ and $k = 1, 2, \ldots K$. From (37) we have the initial and boundary conditions

$$
\begin{aligned}
u_{0,j} = \sin^3(\pi x_j), \quad j = 0, 1, \ldots J, \\
u_{k,0} = 0, \quad k = 0, 1, \ldots K,
\end{aligned} \tag{39}
$$

but these are not enough. So we have to choose more boundary conditions both on the left and on the right hand side.

5.1. Boundary conditions without boundary layers. Suppose we want to construct a pilot function that is second order consistent with the scheme defined by (38), (39) and (40). First, we choose artificial boundary conditions with high enough order of consistency:

$$
\begin{aligned}
\frac{u_{k,1} - u_{k-1,1}}{\Delta t} + a \frac{u_{k,2} - u_{k,0}}{2\Delta x} &= 0 \quad \text{and} \\
\frac{u_{k,J} - u_{k-1,J}}{\Delta t} + a \frac{3u_{k,J} - 4u_{k,J-1} + u_{k,J-2}}{2\Delta x} &= 0.
\end{aligned} \tag{40}
$$

Using the ansatz

$$
u^{pi} = \tilde{u} + \Delta x u^{(1)},
$$

the defining equations, which we get from (38) and (40), are identical

$$
u_t^{(1)} + a u_x^{(1)} = \frac{\mu}{2} \tilde{u}_{tt} \tag{41}
$$

and the pilot function is defined uniquely.

5.2. Boundary conditions with boundary layers. We are now going to change the artificial boundary conditions. Instead of (40) we choose:

$$\begin{aligned} \frac{u_{k,1}-u_{k-1,1}}{\Delta t} + a\frac{u_{k,1}-u_{k,0}}{\Delta x} &= 0 \quad \text{and} \\ \frac{u_{k,J}-u_{k-1,J}}{\Delta t} + a\frac{u_{k,J}-u_{k,J-1}}{\Delta x} &= 0. \end{aligned} \tag{42}$$

Using the same ansatz as above, we get again equation (41) from the expansion of (38), but from the expansion of (42), we find:

$$u_t^{(1)} + au_x^{(1)} = \frac{\mu}{2}\tilde{u}_{tt} + \frac{1}{2}\tilde{u}_{xx}.$$

The reason is that the boundary conditions (42) are only first order approximations to the true solution. Therefore, the ansatz for the pilot function must be modified. According to Michelson [5] we set

$$u^{pi}(t,x,\frac{x}{\Delta x},\frac{1-x}{\Delta x},\Delta x) = \tilde{u}(t,x) + \Delta x u^{(1)}(t,x) + \Delta x^2 l^{(1)}(t,\frac{x}{\Delta x}) + \Delta x^2 r^{(1)}(t,\frac{1-x}{\Delta x}).$$

Here, $l^{(1)}$ and $r^{(1)}$ describe the boundary layers on the left and on the right hand side respectively.

We substitute this ansatz into (38) and (42) and expand the terms except for the boundary layer terms, which are not expanded in the x-direction. In this way, we find again equation (41) to define $u^{(1)}$. The following set of difference equations defines $l^{(1)}$:

$$\begin{aligned} 2l_{k,j+1}^{(1)} + 3l_{k,j}^{(1)} - 6l_{k,j-1}^{(1)} + l_{k,j-2}^{(1)} &= 0, \quad j = 2,3,\dots J-1, \\ l_{k,1}^{(1)} - l_{k,0}^{(1)} &= \frac{1}{2}(\tilde{u}_{xx})_{k,0}, \\ l_{k,J-1}^{(1)} = l_{k,J}^{(1)} &= 0, \end{aligned}$$

Here, we have used $(\tilde{u}_{xx})_{k,1} = (\tilde{u}_{xx})_{k,0} + \mathcal{O}(\Delta x)$. Because of $(\tilde{u}_{xx})_{k,0} = 0$ the linear set of equations has the trivial solution $l^{(1)}(t,\frac{x}{\Delta x}) \equiv 0$. Finally, the conditions for $r^{(1)}$ are:

$$\begin{aligned} 2r_{k,j+1}^{(1)} + 3r_{k,j}^{(1)} - 6r_{k,j-1}^{(1)} + r_{k,j-2}^{(1)} &= 0, \quad j = 2,3,\dots J-1, \\ r_{k,J}^{(1)} - r_{k,J-1}^{(1)} &= \frac{1}{2}(\tilde{u}_{xx})_{k,J}, \\ r_{k,0}^{(1)} = r_{k,1}^{(1)} &= 0. \end{aligned}$$

These can be solved by elementary computation:

$$r^{(1)}(t,y) = \frac{1}{2}\tilde{u}_{xx}(t,1)\left(\alpha + \beta\exp\left(y\log\lambda_2\right) + \gamma\exp\left(y\log|\lambda_3|\right)\cdot\cos\left(\pi y\right)\right), \tag{43}$$

$$\text{where} \quad y = \frac{1-x}{\Delta x}, \quad \lambda_2 = \frac{5+\sqrt{33}}{2}, \quad \lambda_3 = \frac{5-\sqrt{33}}{2},$$

$$\alpha = \frac{(\lambda_2 - \lambda_3)\lambda_2^{J-1}\lambda_3^{J-1}}{(\lambda_2 - 1)(\lambda_3 - 1)(\lambda_2^{J-1} - \lambda_3^{J-1})}, \quad \beta = \frac{\lambda_3^{J-1}}{(\lambda_2 - 1)(\lambda_2^{J-1} - \lambda_3^{J-1})},$$

$$\text{and} \quad \gamma = \frac{\lambda_2^{J-1}}{(\lambda_3 - 1)(\lambda_3^{J-1} - \lambda_2^{J-1})}.$$

On the other hand, it is also possible to compute the boundary layer numerically. Let U_1 denote the solution of the scheme (38), (39) with the artificial boundary condition (40), i.e. the method without layer. By U_2 we denote the solution of (38), (39) and (42). The layer, which is caused by the numerical condition (42), can be approximated by the difference $R = \Delta x^2(U_2 - U_1)$.

In Figures 1 and 2, we compare the analytical expression for $r^{(1)}$ given by (43) and the numerical approximation R at time $t = 0.1$ and $t = 0.5$ respectively. All the computations are done with $a = 1$, $\mu = \frac{\Delta t}{\Delta x} = 1$ and $\Delta x = 0.01$. We observe that the numerical approximation—printed as discrete values—represents the analytical expression quite well.

References

1. B. Gustafsson and A. Sundström, *Incompletely parabolic problems in fluid dynamics*, SIAM J. Appl. Math., 35 (1978), pp. 343–357.
2. H.-O. Kreiss and J. Lorenz, *Initial–Boundary Value Problems and the Navier–Stokes Equations*, Academic Press, London, 1989.
3. J. C. López-Marcos and J. M. Sanz-Serna, *A definition of stability for nonlinear problems*, in Numerical Treatment of Differential Equations, K. Strehmel, ed., Teubner, Leipzig, 1988, pp. 216–226.
4. J. C. López-Marcos and J. M. Sanz-Serna, *Stability and convergence in numerical analysis III: Linear investigation of nonlinear stability*, IMA J. Numer. Anal., 8 (1988), pp. 71–84.
5. D. Michelson, *Convergence theorem for difference approximations of hyperbolic quasi-linear initial-boundary value problems*, Math. Comput., 49 (1987), pp. 445–459.
6. J. M. Sanz-Serna, *Two topics in nonlinear stability*, in Advances in Numerical Analysis, Vol. 1, W. Light, ed., Clarendon Press, Oxford, 1991, pp. 147–174.
7. H. J. Schroll, *Convergence of implicit finite difference methods applied to nonlinear mixed systems.* Preprint, Dep. of Informatics, Univ. of Oslo, Norway 1993; to appear in SIAM J. Numer. Anal.
8. G. Strang, *Accurate partial difference methods II. Non–linear problems*, Numer. Math., 6 (1964), pp. 37–46.
9. J. Strikwerda, *Initial boundary value problems for incomplete parabolic systems*, Comm. Pure Appl. Math., 30 (1977), pp. 797–822.

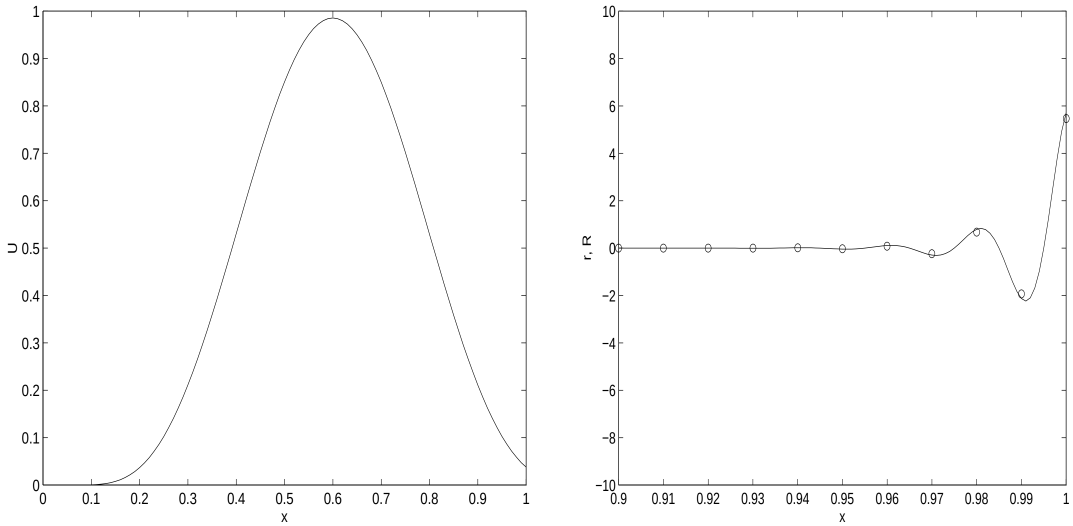

FIGURE 1. The solution U_2 and the boundary layer $r^{(1)}$ respectively R at time $t = 0.1$.

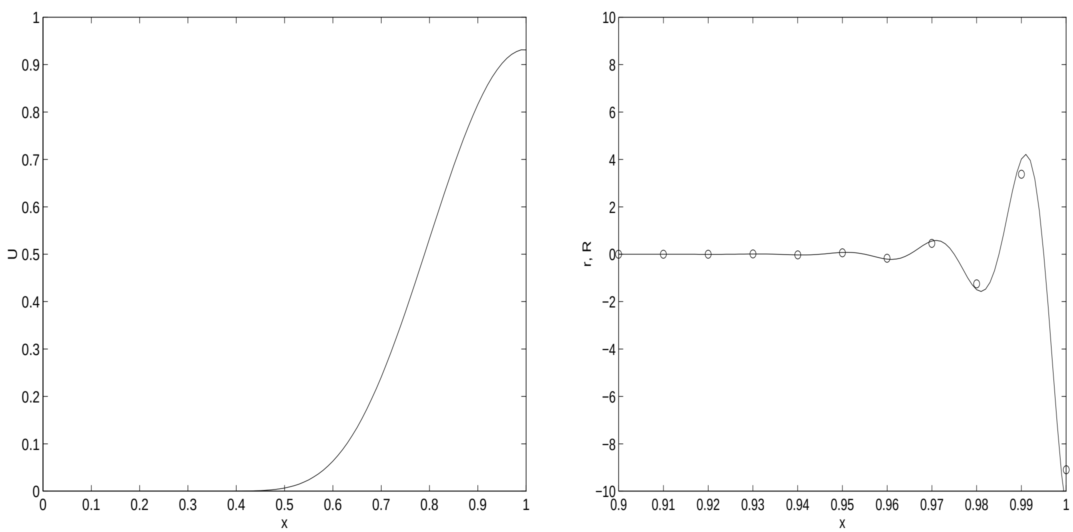

FIGURE 2. The solution U_2 and the boundary layer $r^{(1)}$ respectively R at time $t = 0.5$.