On the implementation of boundary conditions for the method of lines

Bertil Gustafsson
Department of Scientific Computing
Uppsala University
Box 120
S-75104 Uppsala
Sweden

April 3, 1997

Contents

Abstract

The method of lines for difference approximations of hyperbolic first order systems of partial differential equations is analyzed. The approximations are based on strictly semibounded difference operators including high order ones. The formulation of the ODE-system requires that the implementation of the boundary conditions is done carefully. We shall illustrate how different ways of implementation give rise to different stability properties. In particular, we derive a way of implementation that leads to an approximation that is strongly stable in the generalized sense.

It has been an open problem, whether for semidiscrete approximations with this strong stability property, the timestep for the ODE-solver is governed by the Cauchy problem. We present a counterexample showing that it is not.

The analysis presented in this paper also serves as an illustration of the significant difference between different stability concepts.

1 Introduction

In this paper we shall discuss the method of lines for difference approximations of hyperbolic first order systems of partial differential equations. For wave propagation problems, such as those occuring in acoustics, electromagnetics and fluid dynamics, high order methods are usually more efficient, see [2], and these methods will get particular attention.

The versatility of a general code intended for use on real world problems is significantly increased if it is built up as a modular system. The method of lines is usually interpreted as an application of a standard ODE solver to a PDE system obtained by a discretization in space. This procedure is very convenient, since it allows the space discretization to be handled separately. The direct discretization of the PDE system is most naturally obtained in the form

$$\begin{aligned} \frac{d\underline{v}}{dt} &= \tilde{Q}\underline{v} + \underline{\tilde{F}}, \\ B\underline{v}(t) &= \underline{g}(t), \\ \underline{v}(0) &= \underline{\tilde{f}}, \end{aligned} \tag{1.1}$$

where $\underline{v}$ is the semidiscrete solution vector, and B represents the boundary conditions in space. However, a direct application of the ODE solver requires the standard form

$$\frac{d\underline{u}}{dt} = Q\underline{u} + \underline{F}, \tag{1.2}$$

$$\underline{u}(0) = \underline{f}, \tag{1.3}$$

for the semidiscrete solution vector $\underline{u} = \underline{u}(t)$. This pure initial-value problem is also the most appropriate form for applying the stability theory when deriving stability conditions for the fully discrete approximation. However, even if the basic discretization in space is given, the building of the operator Q is not necessarily straightforward. Each equation in (1.2) is a differential equation. In general, the boundary conditions for the PDE are not expressed as differential equations in time, and therefore we need some reformulation procedure. In order to secure stability for the final fully discrete approximation, this formulation has to be done carefully.

The most general and systematic way of constructing stable difference methods for first order systems has been devised by Olsson [6] and [7]. The procedure is based on *strictly semibounded* approximations of $\partial/\partial x$ as introduced by Kreiss and Scherer in [3], and later realized by Strand [9]. The method is described in [2]. For first order systems we define difference approximations of the differential operators $\partial/\partial x$, $\partial/\partial y$, $\ldots$ on the whole grid. By applying these operators premultiplied by the coefficient matrices at all grid-points, we have a new gridfunction defined on the whole grid. The application of the physical boundary conditions is then implemented as a separate projection step. In this paper we shall analyze this approach in more detail than has been done before. In particular, we shall show that the resulting procedure has some peculiar stability properties, that in some cases may lead to less robust bahaviour. As a result we propose a slightly modified procedure, that eliminates this difficulty.

In the last section, we shall discuss the discretization in time. Except for simple first and second order time-discretizations, very few stability results exist for initial-boundary value problems. Levy and Tadmor [5] derived conditions for Runge-Kutta schemes based on a so called coercivity condition, which is in general not satisfied for non-dissipative approximations. The most general stability result for first order systems is due to Kreiss and Wu [4], and it applies to Runge-Kutta and linear multi-step approximations. It turns out that the sufficient stability condition given there leads to a restriction on the time-step, which is severe for some of the high order operators. Practical experience shows that longer time-steps can be used. Hence, the question is whether it is possible to derive a weaker sufficient stability condition. In particular one would like to know if there is a sufficient stability condition that is no more severe than the time-step constraint for the Cauchy problem. By studying a counterexample, we shall show that this is impossible for general approximations, even if they are stable for the semi-discrete problem.

The analysis presented in this paper also serves as an illustration of the significant difference between different stability concepts.

2 Preliminaries

We shall discuss approximations of first order hyperbolic systems, and consider first semidiscrete difference approximations. Let u , F , g , f be the solution, the forcing function, the boundary data and the initial data respectively. Furthermore, let Ω_h be the collection of inner points x_j in space where the difference operator Q is applied, and let $\partial\Omega_h$ denote the boundary points.

The approximation is

$$\begin{array}{rcll} \frac{du}{dt} & = & Qu + F \ , & x_j \in \Omega_h \\ Bu & = & g \ , & x_j \in \partial\Omega_h \\ u(0) & = & f & \end{array} \tag{2.1}$$

Unfortunately, it is not possible to use one simple definition of stability for approximations of PDE. If one chooses a strict version, many useful approximations fall outside. If one uses a weak definition, many essential properties of the approximation cannot be singled out. These matters are discussed in detail in [2]. In Appendix A the relevant stability definitions are given, and the relations between them as they are currently known.

We shall now discuss one-dimensional difference operators, that are the building blocks for the general multi-dimensional case. The difference approximations are constructed starting from strictly semibounded difference operators, and we shall first define this concept. Consider functions $u(x)$ with compact support defined on the positive x-axis and the differential operator $\partial/\partial x$. For the usual L_2 scalar product we have

$$(u, \partial u/\partial x)_{L_2} = -\frac{1}{2}|u(0)|^2 \tag{2.2}$$

With gridfunctions defined on the grid $\{x_j = jh\}_{j=0}^{\infty}$, let D be a difference approximation of $\partial/\partial x$, and let

$$(u, v)_h = \sum_{i,j=0}^{r-1} h_{ij} u_i v_j h + \sum_{j=r}^{\infty} u_i v_j h \tag{2.3}$$

be the scalar product with the corresponding norm

$$||u||_h^2 = (u, u)_h \,. \tag{2.4}$$

The coefficients h_{ij} are elements of a symmetric positive definite $r \times r$ matrix H_0.

Definition 2.1 *The difference operator D is called* strictly semibounded *if for every gridfunction u with $||u||_h < \infty$*

$$(u, Du)_h = -\frac{1}{2}|u_0|^2 \,. \tag{2.5}$$

Obviously, this is a nice property, since it resembles exactly the differential operator. Furthermore, and perhaps more importantly, it leads to stability also for approximations of systems of PDEs by a simple projection based only on the physical boundary conditions.

Assume now that we use the standard centered difference approximation of order $2s$ at all inner points, i.e. the stencil is $2s + 1$ points wide. In order to obtain an overall accuracy of order $2s$, we can sacrifice one order near the boundary, but not more.

Theorem 2.1 *For every order of accuracy $2s$ in the interior, there are boundary approximations accurate of order $\tau = 2s - 1$ such that D is strictly semibounded.*

There are s points near the boundary, where the centered formula cannot be applied. Accordingly, the minimal number of points to be modified is s. However, it turns out that in general we need different formulas at more than s points, the only exception is $s = 1$. For a given order of accuracy, D is not unique. In [9], examples of D up to order 6 is found, where D contains one or more free parameters.

For the scalar outflow equation

$$u_t = u_x$$

these difference operators give the exact energy dissipation at the boundary:

$$\frac{d}{dt}||u||_h^2 = -|u_0|^2 .$$

In the next few sections, we shall discuss other situations, where the stability properties are less obvious.

3 A scalar example

In order to illustrate various types of boundary condition implementation, we consider the problem

$$
\begin{aligned}
u_t + u_x &= 0 \ , \ 0 \le x < \infty\,, \ 0 \le t\,, \\
u(x,0) &= f(x)\,, \\
u(0,t) &= g(t)\,,
\end{aligned} \tag{3.1}
$$

where we shall discuss the treatment of the inflow boundary at $x = 0$. We assume that $f(x)$ has compact support, and as a consequence for the numerical experiments, we substitute zeros for u near the right boundary. Let D be a strictly semibounded difference operator as defined above. We first consider the straightforward procedure, where the boundary condition is enforced for all t, i.e. we work with the unknowns $u_1, u_2, \ldots$ at inner points, and assume that $u_0 = g$. The semidiscrete approximation is

$$
\begin{aligned}
\frac{du_j}{dt} + D(j)u_j &= 0 \ , \ j = 1,2,\ldots\,, \ 0 \le t\,, \\
u_j(0) &= f_j\,, \\
u_0(t) &= g(t)\,.
\end{aligned} \tag{3.2}
$$

Here we use the notation $D(j)$ to indicate that the approximation varies from one point to the next near the boundaries. Furthermore, it is not applied at the boundary point x_0. This seems to be a very natural formulation, and one may think that it is a trivial case giving an energy estimate by using the norm $(u,u)_h$ defined in (2.3). But it is not. In fact, for the $6th$ order operator D_6 defined in [9], Strand [10] showed that there are exponentially growing solutions to this problem even if $g(t) = 0$. This may be somewhat surprising, and we will show what goes wrong in the derivation of the energy estimate.

Let $\underline{u} = (u_0, u_1, \ldots)^T$ be the *full* vector including the boundary element. With a slight abuse of notation, we let D denote both the difference operator and the corresponding matrix. Let H be the matrix corresponding to the scalar product in (2.3), i.e. H is the unity matrix with the upper left $r \times r$ corner substituted by H_0. Let $\langle \cdot, \cdot \rangle$ denote the usual scalar product for vectors. Strict stability means that for real gridfunctions we have

$$
-\langle \underline{u}, HD\underline{u} \rangle = \frac{1}{2} u_0^2\,.
$$

For $u_0 = 0$ we can also write this as

$$
-\langle \tilde{u}, \widetilde{HD}\tilde{u} \rangle = 0\,,
$$

where $\widetilde{HD}$ is the matrix HD with the first row and column eliminated.

Consider next the vector $\tilde{u} = (u_1, u_2, \dots)^T$, where the boundary value u_0 has been eliminated, and let $\tilde{D}$ be the matrix corresponding to the difference operator $D(j)$, $j = 1, 2, \dots$, i.e. the first row and column have been eliminated. Furthermore, let $\tilde{H}$ be the corresponding reduced matrix occurring in the the scalar product. Since $d\tilde{u}/dt = \tilde{D}\tilde{u}$, we have

$$\frac{d}{dt}\langle \tilde{u}, \tilde{H}\tilde{u}\rangle = 2\langle \tilde{u}, \tilde{H}\tilde{u}_t\rangle = -2\langle \tilde{u}, \tilde{H}\tilde{D}\tilde{u}\rangle \,. \tag{3.3}$$

This is where the derivation stops. In order to get a non-increasing norm, we would need $\tilde{H}\tilde{D} = \widetilde{HD}$. In general this is not true.

In the formulation above we have eliminated u_0 from the system of differential equations. Let us next consider a different formulation, where u_0 is kept in the system. For the fully discrete approximation, where also time has been discretized, it is natural to advance the scheme at *all* points, since D is defined everywhere. We then modify the boundary value by imposing the physical boundary condition. For the semidiscrete case, this corresponds to introducing a projection operator P. Let the vector $\underline{u}$, and the matrix D be defined as above, and let P be defined by

$$P = diag(0\ 1\ 1\ 1\dots) \tag{3.4}$$

For simplicity we assume $g(t) = 0$, and write the problem (3.2) in the form

$$\begin{aligned} \tfrac{du}{dt} + PD\underline{u} &= 0\,, \\ \underline{u}(0) &= P\underline{f} \end{aligned} \tag{3.5}$$

Since $f(x)$ is restricted to satisfy the boundary condition, the first equation guarantees that $u_0 = 0$, and obviously the two formulations are equivalent. We can also formulate the problem in a third way:

$$\begin{aligned} \tfrac{du}{dt} + PDP\underline{u} &= 0\,, \\ \underline{u}(0) &= P\underline{f} \end{aligned} \tag{3.6}$$

The spectrum of $-PDP$ with $D = D_6$ is shown in Figure 1 (cf. Strand [10]). Here we have assumed zero boundary values to the right such that the matrix can be made finite of order $N \times N$.

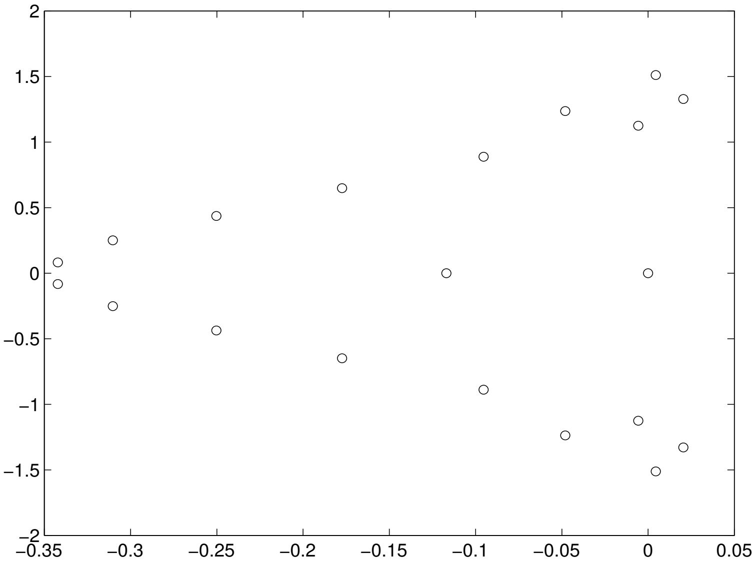

Figure 1: Spectrum of $-PD_6P$, $\quad N = 20$

The operator has eigenvalues λ_j in the right halfplane, implying that there are exponentially growing solutions. For large N we have

$$\max_{1\leq j\leq N} \; Re(\lambda_j) \geq \delta > 0\,,$$

where δ is independent of N. This means that there is no possibility in *any* norm to have an energy estimate of the form

$$d|\underline{u}|^2/dt \leq 0\,.$$

Remark:
There is a theorem by Goldberg and Tadmor [1] stating that specifying data at the s missing points for the centered symmetric difference operator of order $2s$ gives a strongly stable scheme. For our operators, this theorem applies only for the case $s = 1$.

The result above shows that we need another implementation of the boundary conditions. Olsson [6] constructed such a new implementation procedure, that we shall describe next. Consider for a moment the general case including systems of differential equations, and let $\underline{u}$ denote the discrete approximation in vector form including the boundary values. Formulate the physical boundary conditions as

$$L\underline{u} = 0\,, \tag{3.7}$$

where L is a rectangular matrix with as many rows as there are physical boundary conditions. The approximation is formulated as in (3.5), where the projection operator is defined by:

$$P = I - H^{-1}L^T(LH^{-1}L^T)^{-1}L \tag{3.8}$$

In our case L has the form

$$L = (1\ 0\ \ 0\ 0 \ldots\ 0)\,. \tag{3.9}$$

The problem (3.5) now satisfies an energy estimate, but note that the initial data must satisfy the boundary condition.

The upper left corner of the matrix representation of P is in this case

$$P = \left[\begin{array}{cccccccc} 0.0000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0.0676 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \\ 0.0806 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \\ -0.0048 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \\ -0.1137 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \\ 0.15167 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \\ \vdots & & & & & & & & \ddots \end{array}\right. \tag{3.10}$$

Compared to the previously used projection, the solution is now modified also at the points $x_1 \ldots x_5$. The spectrum of $-PD_6P$ is contained in the left halfplane as shown in Figure 2.

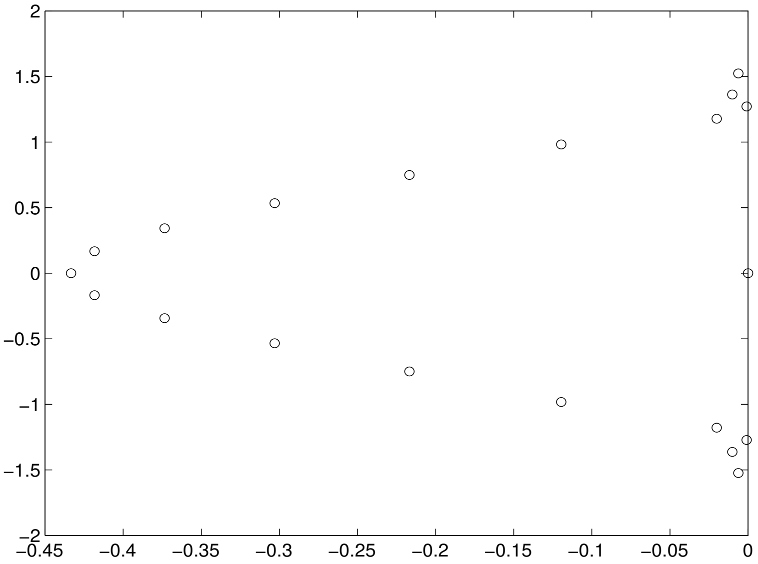

Figure 2: Spectrum of $-PD_6P$, $N = 20$, new projection P

Remark:

If all elements of the matrix H_0 are zero outside the diagonal in the first row and column, i.e. H as the form

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & & & \\ 0 & \times & \times & \times & \times & & & & \\ 0 & \times & \times & \times & \times & & & & \\ 0 & \times & \times & \times & \times & & & & \\ 0 & \times & \times & \times & \times & & & & \\ \vdots & & & & & 1 & & & \\ & & & & & & 1 & & \\ & & & & & & & 1 & \\ & & & & & & & & \ddots \end{bmatrix}, \tag{3.11}$$

there is no need to use the more complicated projection (3.8), see [2], the projection (3.4) leads to stability. The more general condition for obtaining a boundary projection operator that is independent of H is discussed in [6].

In the next section we shall show that for the general case, there is reason to consider still another method for implementing the boundary conditions.

4 A 2×2 system

Let

$$w = \begin{bmatrix} u \\ v \end{bmatrix}, \quad A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix},$$

and consider the system

$$\begin{aligned} w_t &= A w_x\,, \quad 0 \le x < \infty\,,\ 0 \le t\,, \\ w(x,0) &= f(x)\,, \\ u(0,t) &= v(0,t)\,. \end{aligned} \tag{4.1}$$

We want to use the standard centered second order difference operator together with a projection operator for the physical boundary condition. In this case, the approximation of $\partial/\partial x$ is modified only at the boundary point itself, where a first order one-sided approximation is used. By using vector-matrix representation and the vector $\underline{w}$ defined by

$$\underline{w} = (u_0\ v_0\ u_1\ v_1\ \dots)^T\,,$$

the difference operator is

$$Q = \frac{1}{2h}\begin{bmatrix} 2 & 0 & -2 & 0 & 0 & 0 & & \\ 0 & -2 & 0 & 2 & 0 & 0 & & \\ 1 & 0 & 0 & 0 & -1 & 0 & & \\ 0 & -1 & 0 & 0 & 0 & 1 & & \\ & & \ddots & & \ddots & & \ddots & \\ & & & \ddots & & \ddots & & \ddots \end{bmatrix}$$

Correspondingly, the matrix H in the norm is

$$H = diag\,[\ 1/2 \quad 1/2 \quad 1 \quad 1 \quad \dots\ ]$$

The boundary conditions are written in the form

$$Lw = 0\,, \quad L = [\ 1 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots\ ]\,, \tag{4.2}$$

and the projection operator P is then given by (3.8).

The resulting semi-discrete scheme is

$$\frac{d\underline{w}}{dt} = PQ\underline{w}\,, \tag{4.3}$$

where the explicit forms of P and PQ are

$$P = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 & & \\ 1 & 1 & 0 & 0 & & \\ 0 & 0 & 2 & 0 & & \\ 0 & 0 & 0 & 2 & & \\ & & & & \ddots & \\ & & & & & \ddots \end{bmatrix}$$

$$PQ = \frac{1}{2h}\begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 & & \\ 1 & -1 & -1 & 1 & 0 & 0 & & \\ 1 & 0 & 0 & 0 & -1 & 0 & & \\ 0 & -1 & 0 & 0 & 0 & 1 & & \\ & & \ddots & & \ddots & & \ddots & \\ & & & \ddots & & \ddots & & \ddots \end{bmatrix}$$

The method fulfills all conditions that are required for an energy estimate, provided that the initial data fulfill the boundary conditions, see [6]. However, we shall now prove that the approximation does not fulfill any of the stability definitions for general initial data as stated in Appendix A.

We use the normal mode analysis, and prove that there is a generalized eigenvalue. The theory is based on a Laplace transformation of (4.3), assuming $f = 0$. This transformation gives

$$\begin{aligned} s\underline{\hat{w}} &= PQ\underline{\hat{w}}\,, \\ \sum_{j=0}^{\infty} |\hat{w}_j|^2 h &< \infty\,. \end{aligned}$$

The solution $\hat{w}_j = (\hat{u}_j, \hat{v}_j)^T$ has the form

$$\begin{aligned} \hat{u}_j &= \hat{u}_0 \kappa_1^j \\ \hat{v}_j &= \hat{v}_0 \kappa_2^j\,, \quad j = 0, 1, \dots\,, \end{aligned}$$

where $\kappa_1\,,\kappa_2$ are the roots of the characteristic equation

$$2\tilde{s}\kappa = \mp(\kappa^2 - 1)\ ,\quad \tilde{s} = sh\,, \tag{4.4}$$

respectively, with

$$|\kappa_1|, |\kappa_2| < 1 \quad \text{for } Res > 0\,.$$

There are two equations left to determine the parameters $\hat{u}_0, \hat{v}_0$. However, for $\tilde{s} = 0$, the first two rows of PQ are linearly dependent giving identical equations for $\hat{u}_0, \hat{v}_0$. Accordingly, there is always a nontrivial solution, and as a consequence, $\tilde{s} = 0$ is a generalized eigenvalue. Both κ_1 and κ_2 are simple roots of (4.4) and located on the unit circle for $\tilde{s} = 0$. Therefore, we have an instability of the kind that is spreading over the whole domain, see [2]. None of the stability definitions in Appendix A is satisfied.

This is no contradiction. The energy estimate is obtained for special initial data, where the unstable mode has been removed. All the stability definitions require that the estimates hold for *general* initial data $\underline{f}$. This may also be of practical significance. For real problems in several space dimensions, it is sometimes difficult to ensure that the initial function satisfies the boundary condition, and the unstable mode may be triggered. Furthermore, the scheme will be less robust if it does not fulfill the general stability condition.

It can be shown that this type of instability occurs also for higher order accurate operators. We have run a numerical experiment with the 6th order operator described in the previous section, but now applied to our 2×2 system. Figure 3 shows the l_2-norm of $\underline{w}$ (normalized to 1 at $t = 0$) with the initial data

$$w_j(0) = \begin{bmatrix} (-1)^j \\ 1 \end{bmatrix} \quad , \; j = 0, 1, \dots\,,$$

which satisfy the boundary condition at $x = 0$. At the right boundary we simply assume that the solution vanishes, and prescribe

$$w_j = 0 \quad , \; j = N, N+1, N+2\,,$$

which is a stable procedure. The norm is bounded for any time.

We have also run the same scheme with the oscillation in $u_j(0)$ shifted one step, such that the initial data do not satisfy the boundary condition:

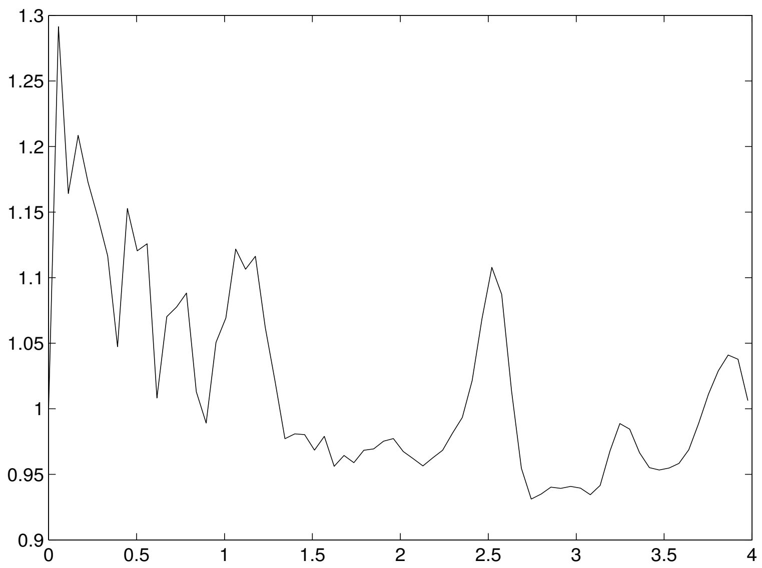

Figure 3: $||w(t)||_h$ for compatible initial data

$$w_j(0) = \begin{bmatrix} (-1)^{j-1} \\ 1 \end{bmatrix} \quad , \; j = 0, 1, \ldots \,.$$

The result is shown in Figure 4.

There is now a linear growth in the norm, just like the theory predicts.

In [8] methods leading to energy estimates for nonlinear problems are derived. Here we shall analyze what happens with nonlinear problems for the more direct and simple implementation derived from the linear theory.

We consider an example, by modifying (4.1) slightly:

$$\begin{aligned} w_t &= |u| A w_x \,, \quad 0 \leq x < \infty \,, \; 0 \leq t \,, \\ w(x,0) &= f(x) \,, \\ u(0,t) &= v(0,t) \,, \end{aligned} \tag{4.5}$$

where A is defined in (4.1). We first run the scheme with no boundary projection formulation, but rather use direct enforcement of the boundary condition in analogy with the scalar problem formulation (3.2). The approximation is unstable, and there is a nonlinear explosive effect as shown in Figure 5.

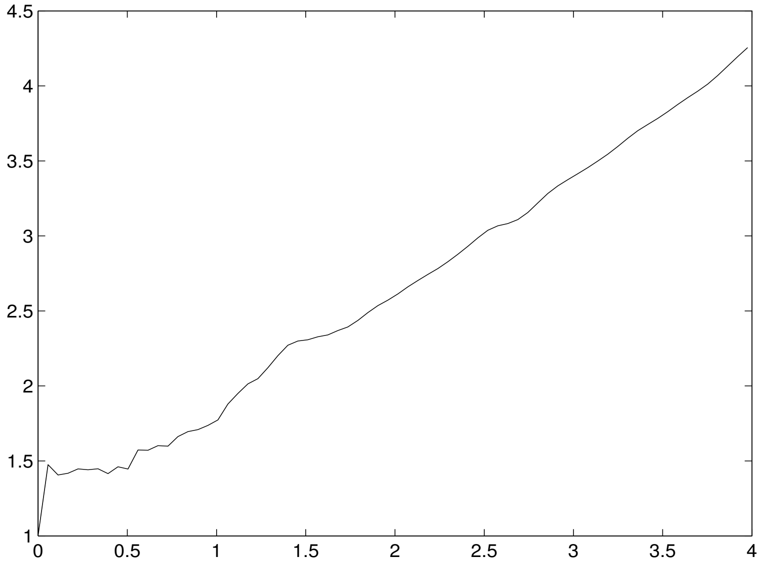

Figure 4: $||w(t)||_h$ for incompatible initial data

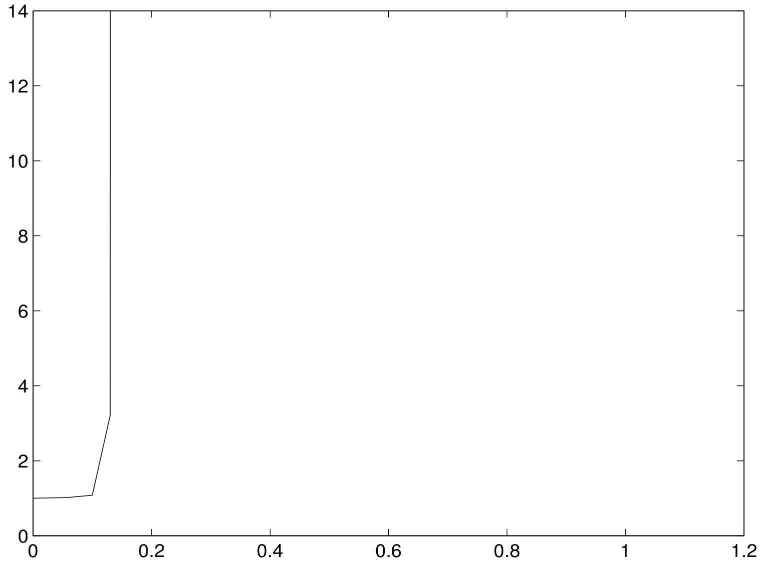

Figure 5: $||w(t)||_h$ for boundary condition enforced without projection operator. Nonlinear system.

We next turn to the projection formulation. The boundary operator (4.2) is the same as for the linear case, and the projection operator has the form (3.8). We use the initial data

$$\begin{aligned} u_0 &= 1\,, \\ u_j &= 0 \quad , j = 1, 2, \dots\,, \\ v_j &= 1 \quad , j = 0, 1, \dots\,, \end{aligned}$$

i.e. the boundary condition is fulfilled. The solution behaves very well as shown in Figure 6.

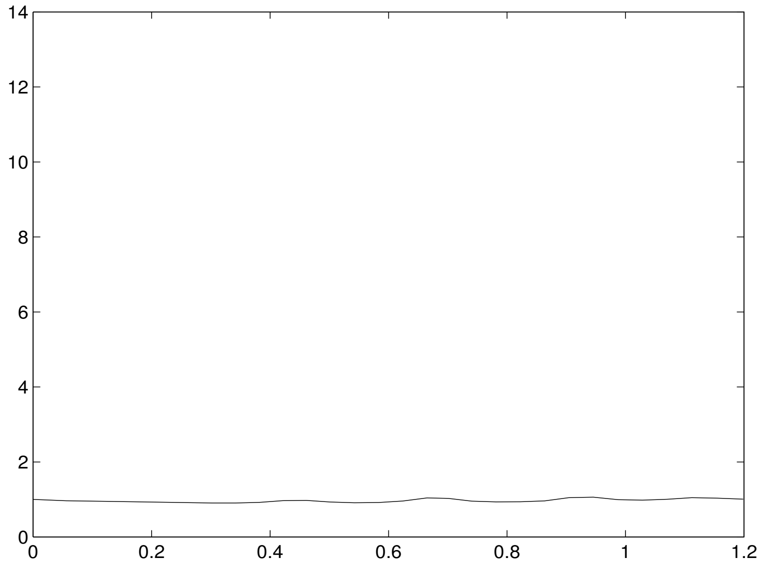

Figure 6: $||w(t)||_h$ for PQ, no perturbation. Nonlinear system.

Next, we introduce a perturbation ϵ at the boundary in each time-step:

$$v_0 \rightarrow v_0 + \epsilon\,.$$

The result for $\epsilon = 0.1$ is shown in Figure 7.

There is now a nonlinear effect, such that the weak linear instability is converted to an explosive growth in the solution. The reason is the generalized eigenvalue for the linear problem, as described above.

The weak instability can be removed. The cure is a modification such that the physical boundary condition is enforced, but *after* the projection operator

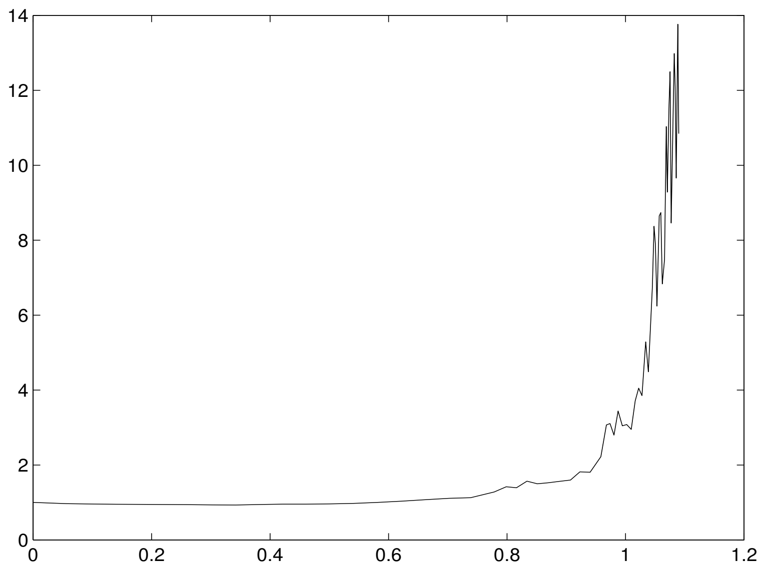

Figure 7: $||w(t)||_h$ for PQ and perturbation $\epsilon = 0.1$. Nonlinear system.

has been applied. For our example it simply means that u_0 is eliminated such that we get a shorter vector

$$\underline{\tilde{w}} = [v_0 \; u_1 \; v_1 \; \ldots] \,,$$

The new operator $\widetilde{PQ}$ is obtained by adding the first and second column of PQ $(u_0 = v_0)$ and eliminating the first row. The system is

$$\begin{aligned} \frac{d\underline{\tilde{w}}}{dt} &= \widetilde{PQ}\underline{\tilde{w}}\,, \\ \underline{\tilde{w}}(0) &= \underline{\tilde{f}}\,. \end{aligned} \tag{4.6}$$

We apply the same type of analysis as for the second order approximation above. There is a complication now, since there are 11 points near the boundary that are modified instead of 6 points as is required by the standard theory. The complete analysis is given Appendix B.

The numerical result for our example with the same perturbation as above, is shown in Figure 8.

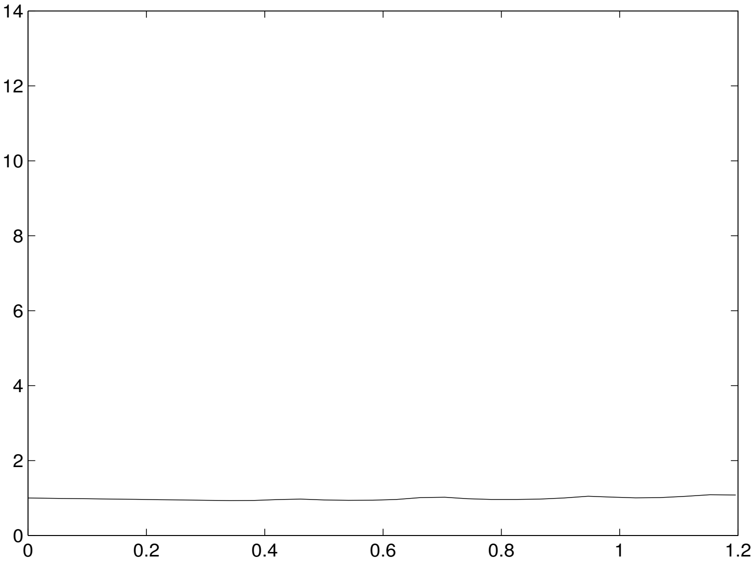

Figure 8: $||w(t)||_h$ for $\widetilde{PQ}$ and perturbation $\epsilon = 0.1$. Nonlinear system.

The norm is now bounded. Note the significant difference between the case presented in Figure 5 and this one. In the latter well behaving case, the enforcement of the boundary condition is preceded by the application of the projection operator.

5 Time-discretization

The experiments in the previous sections are carried out using a standard fourth order Runge-Kutta method for time-discretization. In this section we shall discuss the theoretical basis for methods of this type.

We assume that the boundary variables have been eliminated, as demonstrated above. The semi-discrete approximation is

$$\begin{aligned} du_j/dt &= Su_j + F_j \ , \\ u_j^0 &= f_j \ , \quad j = 0, 1, \dots \ , \end{aligned} \tag{5.1}$$

where S represents any of the difference operators above with or without a projection operator included. For convenience we consider only one-step schemes

$$
\begin{aligned}
u_j^{n+1} &= Tu_j^n + kG_j^n \;, \\
u_j^0 &= f_j \;, \quad j = 0, 1, \dots \;,
\end{aligned}
\tag{5.2}
$$

where k is the time-step.

Methods for discretizing systems of ODE are usually analyzed by considering its properties for the scalar model equation $du/dt = \lambda u$, where λ is a complex number. The stability domain, i.e. those λk giving nongrowing solutions, is well known for most methods. The analysis can then be easily generalized to systems with a fixed number of equations, if they can be diagonalized. However, our systems become larger when the step-size h decreases, and we need uniform estimates as $h \to 0$. Furthermore, we want to obtain estimates also for the case with variable coefficients.

The most general stability result for initial-boundary value problems for first order systems is due to Kreiss and Wu [4]. Assume that the time-discretization method has part of its stability domain Ω in the left halfplane, and let $\mathcal{C}$ be the largest semi-circle that can be inscribed in Ω:

$$
\mathcal{C} = \{z | z \in \Omega \,, \; |z| < R_1 \,, \; Re \; z < 0\}
$$

The theorem requires a few technical assumptions, which are fulfilled by most standard methods, and we omit them here. The result is stated in terms of generalized stability, which for fully discrete approximations is defined by

Definition 5.1 *The approximation (5.2) is* stable in the generalized sense *if for $f = 0$ the solutions satisfy*

$$
\sum_{n=1}^{\infty} e^{-2\eta t_n} ||u^n||_h^2 k \leq K(\eta) \sum_{n=1}^{\infty} e^{-2\eta t_n} ||G^{n-1}||_h^2 k \;, \quad \lim_{\eta \to \infty} K(\eta) = 0 \,.
$$

Theorem 5.1 *Assume that the semi-discrete approximation (5.1) is stable in the generalized sense with*

$$
K(\eta) \leq \frac{const.}{\eta} \,.
$$

Then (5.2) is stable in the generalized sense if

$$
k \leq R/||S||_h \;, \quad R < R_1 \,. \tag{5.3}
$$

The condition (5.3) is restrictive. As an example, consider the 2×2 system (4.1) and the 6th order approximation D_6 of $\partial/\partial x$. At the right boundary, we prescribe zero values, which is a strongly stable condition. The l_2-norm of the corresponding matrix $S = PQ$ in (5.1) is $||S||_h = 16.86/h$ for large N. The radius R_1 for the fourth order Runge-Kutta method can be taken as $R_1 = \sqrt{8}$. This means that the time-step is restricted by the condition

$$k < \frac{h\sqrt{8}}{16.86}\,. \tag{5.4}$$

For comparison, consider the Cauchy problem for the same system. We use the centered high order difference approximations at all gridpoints. Since this leads to an anti-Hermitian operator S, the standard l_2-norm equals the spectral radius of S, which by Parseval's relation equals the spectral radius of the symbol $\hat{S}$. The eigenvalues of the coefficient matrix A are 1 and -1, hence the symbol $\hat{S} = A\hat{D}_6$ has the same spectral radius as $\hat{D}_6$, and we have

$$\rho(\hat{S}) = \rho(\hat{D}_6) = 1.58/h\,.$$

Hence, we get the much less restrictive condition for the Cauchy problem

$$k < \frac{h\sqrt{8}}{1.58}\,. \tag{5.5}$$

The condition (5.4) is sufficient, but not necessary, and there is a good chance that much larger time-steps can be used. First of all one can construct other norms providing lower bounds. A first candidate would be the H-norm $||u||_h = \langle \underline{u}, H\underline{u}\rangle^{1/2}$ used in the theory above. For our example we have $||S||_h = 8.97/h$. Since this norm is equivalent with the l_2 norm, we have stability under a less restrictive condition. Furthermore, there are other norms that gives still lower bounds on $||S||_h$.

Ideally, we would like to have a situation where the the time-step is governed by the Cauchy problem. However, there is no possibility to get such a general stability result even if the semi-discrete problem is stable in the generalized sense. Consider the system (4.6) with the same 6th order approximation, where the boundary projection operator is applied, but with the variable u_0 eliminated afterwards. As explained in the previous section, this system is strongly stable. A numerical computation shows that

$$\rho(\tilde{PQ}) = 2.26/h$$

for large N. If the time-step for the fourth order Runge-Kutta method is chosen close to the limit as given by the condition (5.5) for the Cauchy problem, we will obviously get a spectral radius of the corresponding matrix T in (5.2) larger than one, and we have

$$\rho(T) > 1 + \delta$$

where δ is independent of h. This gives an unstable scheme.

6 Conclusions

We have analyzed the procedure of deriving the standard form for the method of lines as a pure initial value problem. The approximation is based on high order difference operators of strictly semi-bounded type. The results illustrate the variety of stability properties that may be obtained by different boundary condition implementations. A direct implementation of the physical boundary conditions may lead to exponential instabilities. There is a systematic procedure for constructing approximations leading to energy estimates for compatible initial data. We have shown that this leads to a weak instability for general initial data. Sometimes it may be desirable to remove this effect, and we have devised a method for doing that. Finally we have given a counterexample for the method of lines, showing that even if the semidiscrete approximation is strongly stable in the generalized sense, the time-step restriction obtained from the Cauchy problem is not enough to ensure stability for the fully discrete approximation.

A Stability definitions

Consider the semidiscrete approximation (2.1). Let $||u||_h$ denote a discrete Hilbert norm defined over the inner points in space, and let $|u|_B$ be the boundary norm defined by

$$|u|_B^2 = \sum_{x_j \in \partial\Omega} |u_j|^2 \tag{A.1}$$

We consider a finite time-interval $0 \leq t \leq T$ and make

Definition A.1 *The approximation is called* stable *if for $g = 0$ in (2.1) there is a constant $K = K(T)$ such that an estimate*

$$||u(t)||_h^2 \leq K(||f||_h^2 + \int_0^t ||F(\tau)||_h^2 d\tau) \tag{A.2}$$

holds.

Definition A.2 *The approximation is called* strongly stable *if there is a constant $K = K(T)$ such that an estimate*

$$||u(t)||_h^2 + \int_0^t |u(\tau)|_B^2 d\tau \leq K(||f||_h^2 + \int_0^t (||F(\tau)||_h^2 + |g(\tau)|_B^2) d\tau) \tag{A.3}$$

holds.

Definition A.3 *The approximation is called* stable in the generalized sense *if for $f = 0$, $g = 0$ in (2.1) there is a constant $K = K(\eta)$ such that an estimate*

$$\int_0^\infty e^{-2\eta t} ||u(t)||_h^2 dt \leq K(\eta) \int_0^\infty e^{-2\eta t} ||F(t)||_h^2 dt \tag{A.4}$$

holds for $\eta \geq \eta_0$, where $\lim_{\eta \to \infty} K(\eta) = 0$.

Definition A.4 *The approximation is called* strongly stable in the generalized sense *if for $f = 0$ in (2.1) there is a constant $K = K(\eta)$ such that an estimate*

$$\int_0^\infty e^{-2\eta t} (||u(t)||_h^2 + |u(t)|_B^2) dt \leq K(\eta) \int_0^\infty e^{-2\eta t} (||F(t)||_h^2 + |g(t)|_B^2) dt \tag{A.5}$$

holds for $\eta \geq \eta_0$, where $\lim_{\eta \to \infty} K(\eta) = 0$.

We also use

Definition A.5 *A difference operator* Q *is* semibounded *if, for all gridfunctions* u *that satisfy the homogeneous boundary conditions* $Bu = 0$*, there is a scalar product, a norm and a constant* α *such that the inequality*

$$Re(u, Qu)_h \leq \alpha ||u||_h^2 \tag{A.6}$$

holds.

Next follows a few relations between the various stability concepts.

- *Strongly stable* $\Rightarrow$ *stable*
- *Stable* for $F = 0$ $\Rightarrow$ *stable* for $F \neq 0$ (Duhamel's principle)
- *Strongly stable in the generalized sense* $\Rightarrow$ *stable in the generalized sense*
- Q *semibounded* $\Rightarrow$ *stable*
- *Stable* $\Rightarrow$ *stable in the generalized sense*
- *Strongly stable* $\Rightarrow$ *strongly stable in the generalized sense*

B Stability proof

We shall prove that the approximation (4.6) is strongly stable in the generalized sense. For that purpose we assume $\underline{\tilde{f}} = 0$ and Laplace transform the differential equation with respect to t:

$$s\hat{w} = \widetilde{PQ}\hat{w} \,. \tag{B.1}$$

This is not the standard form required for the stability theory. The complication is that the constant difference operator at inner points requires 3 boundary conditions for each component $\hat{u}\,, \hat{v}\,$, i.e. 6 conditions altogether. But the operator $\widetilde{PQ}$ differs from the basic centered operator at 11 points (recall that the variable v_0 has been eliminated). Therefore, in order to apply the theory, we first have to eliminate five unknowns from the system. We need to find 5 rows where the first 5 columns form a submatrix that is non-singular for all $\tilde{s}$, $Re\tilde{s} \geq 0$. It turns out that row 2 to 6 contain such a matrix. We then write the new system as

$$\begin{aligned} \tilde{s}\hat{\phi}_j &= R\hat{\phi}_j \ , \quad j = 3, 4, \dots \ , \ \tilde{s} = sh \, , \\ E\hat{\phi}_0 &= 0 \ , \\ ||\hat{\phi}||_h &< \infty \, . \end{aligned} \tag{B.2}$$

Here

$$\hat{\phi}_j = \begin{bmatrix} \hat{\phi}_{pj} \\ \hat{\phi}_{mj} \end{bmatrix} = \begin{bmatrix} \hat{u}_{j+3} \\ \hat{v}_{j+3} \end{bmatrix}$$

has two components, and R is the difference operator (premultiplied by h) at inner points representing the 6th order approximation represented by the coefficients

$$\pm(\ -\tfrac{1}{60} \quad \tfrac{3}{20} \quad -\tfrac{3}{4} \quad 0 \quad \tfrac{3}{4} \quad -\tfrac{3}{20} \quad \tfrac{1}{60} \ ) \, .$$

$E = E(\tilde{s})$ is the new boundary operator representing 6 different conditions. It remains to show that there is no eigenvalue or generalized eigenvalue $\tilde{s}$ to the problem (B.2). The solution $\hat{\phi}$ can be expressed in terms of the roots of the characteristic equation

$$\tilde{s} \pm (\frac{1}{60}\kappa^3 - \frac{3}{20}\kappa^2 + \frac{3}{4}\kappa - \frac{3}{4}\kappa^{-1} + \frac{3}{20}\kappa^{-2} - \frac{1}{60}\kappa^{-3}) = 0 \, . \tag{B.3}$$

There are 3 roots with $|\kappa| < 1$ for $Re\tilde{s} > 0$. We denote these by $\kappa_{p1}, \kappa_{p2}, \kappa_{p3}$ corresponding to the plus-sign in (B.3), and by $\kappa_{m1}, \kappa_{m2}, \kappa_{m3}$ corresponding to the minus-sign. In order to get a convenient representation covering also all cases with multiple roots, we write

$$\begin{aligned} \phi_{pj} &= c_{p1} f_j^{(1)}(\kappa_{p1}) + c_{p2} f_j^{(2)}(\kappa_{p2}, \kappa_{p1}) + c_{p3} f_j^{(3)}(\kappa_{p3}, \kappa_{p2}, \kappa_{p1}) \, , \\ \phi_{mj} &= c_{m1} f_j^{(1)}(\kappa_{m1}) + c_{m2} f_j^{(2)}(\kappa_{m2}, \kappa_{m1}) + c_{m3} f_j^{(3)}(\kappa_{m3}, \kappa_{m2}, \kappa_{m1}) \, . \end{aligned}$$

Here the gridfunctions $f_j^{(\nu)}$ are defined by

$$f_j^{(1)}(\kappa_1) = \kappa^j \, ,$$

$$f_j^{(2)}(\kappa_2,\kappa_1) = \begin{cases} j\kappa_1^{j-1} & \text{if} \kappa_1 = \kappa_2\,, \\ (\kappa_2^j - \kappa_1^j)/(\kappa_2 - \kappa_1) & \text{else}\,, \end{cases}$$

$$f_j^{(3)}(\kappa_3,\kappa_2,\kappa_1) = \begin{cases} \frac{j(j-1)}{2}\kappa_1^{j-2} & \text{if } \kappa_1 = \kappa_2 = \kappa_3\,, \\ \frac{1}{\kappa_1-\kappa_2}(j - \frac{\kappa_1}{\kappa_1-\kappa_2})\kappa_1^{j-1} + \frac{\kappa_2^j}{(\kappa_1-\kappa_2)^2} & \text{if } \kappa_1 = \kappa_3 \neq \kappa_2\,, \\ (f_j^{(2)}(\kappa_3,\kappa_2) - f_j^{(2)}(\kappa_2,\kappa_1))/(\kappa_3 - \kappa_1) & \text{else}\,. \end{cases}$$

Let the vectors $\underline{c}$ and $\underline{\phi}$ be defined by

$$\underline{c} = \begin{bmatrix} c_{p1} \\ c_{p2} \\ c_{p3} \\ c_{m1} \\ c_{m2} \\ c_{m3} \end{bmatrix} \quad , \quad \underline{\phi} = \begin{bmatrix} \phi_{p0} \\ \phi_{m0} \\ \phi_{p1} \\ \phi_{m1} \\ \phi_{p2} \\ \phi_{m2} \end{bmatrix} \quad ,$$

and the matrix Z by

$$Z = \begin{bmatrix} f_{p0}^{(1)} & f_{p0}^{(2)} & f_{p0}^{(3)} & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{m0}^{(1)} & f_{m0}^{(2)} & f_{m0}^{(3)} \\ f_{p1}^{(1)} & f_{p1}^{(2)} & f_{p1}^{(3)} & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{m1}^{(1)} & f_{m1}^{(2)} & f_{m1}^{(3)} \\ f_{p2}^{(1)} & f_{p2}^{(2)} & f_{p2}^{(3)} & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{m2}^{(1)} & f_{m2}^{(2)} & f_{m2}^{(3)} \end{bmatrix} .$$

Here we have used the notation $f_{pj}^{(1)} = f_j^{(1)}(\kappa_{p1})$ etc. Then $\underline{\phi}$ can be expressed as

$$\underline{\phi} = Z\underline{c}\,. \tag{B.4}$$

There is an eigenvalue or generalized eigenvalue $\tilde{s}$ if and only if $E\underline{\phi} = 0$ has a nontrivial solution. By (B.4) we get the equivalent condition

$$\det(EZ) = 0\,. \tag{B.5}$$

Assume next that there is an eigenvalue $\tilde{s}$ with $Re\tilde{s} > 0$, and a corresponding eigengridfunction $\hat{\phi}_j$. By going back to the original variable $\hat{w}$ in

(B.1), we have a nontrivial solution satisfying the physical boundary condition that has an exponential growth in time. But then we can add the boundary variable $\hat{v}_0 = \hat{u}_0$ to the gridvectorfunction, and we have an exponentially growing solution to the full original system. This contradicts the assumption of an energy estimate. The conclusion is that it is enough to check (B.5) for $\tilde{s}$ on the imaginary axis.

We have done so by using a numerical algorithm. The final result is that

$$\min_{Re\tilde{s}=0} |\det(EZ)| = 0.0470 \quad \text{for } \tilde{s} = 1.273i \ ,$$

and we conclude that the approximation is strongly stable in the generalized sense.

References

[1] M. Goldberg and E. Tadmor. Scheme-independent stability criteria for difference approximations of hyperbolic initial-boundary value problems. ii. *Mathematics of Computation*, 36:605–626, 1981.

[2] B. Gustafsson, H.-O. Kreiss, and J. Oliger. *Time Dependent problems and Difference Methods*. Wiley and Sons, 1995.

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

[4] H.-O. Kreiss and L. Wu. On the stability definition of difference approximations for the initial boundary value problem. *Appl. Num. Math.*, 12:213–227, 1993.

[5] D. Levy and E. Tadmor. From semi-discrete to fully-discrete: Stability of runge-kutta schemes by the energy method. Technical Report 97-3, CAM, UCLA, 1997.

[6] P. Olsson. Summation by parts, projections, and stability I. *Mathematics of Computation*, 64(211):1035–1065, 1995.

[7] P. Olsson. Summation by parts, projections, and stability II. *Mathematics of Computation*, 64(212):1473–1493, 1995.

[8] P. Olsson. Summation by parts, projections, and stability. III. Technical Report 95.06, RIACS, 1995.

[9] B. Strand. Summation by parts for finite difference approximations for d/dx. *Journal of Computational Physics*, 110(1):47–67, 1994.

[10] B. Strand. Numerical studies of hyperbolic IBVP with high-order finite difference operators satisfying a summation by parts rule. Technical Report 176, Uppsala University, Dept of Scientific Computing, 1996. Part of PhD-thesis.