Numerical computation of a finite amplitude sound beam

by

Jarle Berntsen and Erlend Vefring

Report no. 81  November 1986
Numerical computation of a finite amplitude sound beam

by

Jarle Berntsen and Erlend Vefring

Report no. 81 November 1986

Abstract. Numerical techniques for computing the harmonic content of a finite amplitude sound beam are analysed and tested. A test problem is introduced, and we find that the analysis of our numerical methods with the help of this test problem also explains the behaviour of the numerical methods on the problem from acoustics. Three different routines for solving the acoustical problem are presented, and test results that may guide us in choosing the most efficient routine and a reasonable step size, are given.
CONTENTS

List of symbols ............................................................ 3

1. Introduction ............................................................. 4

2. Analysis of two numerical techniques on a test problem ....... 6

3. Choice of difference operator and step size in x-direction ...... 14

4. Choice of step size in o-direction .................................. 18

5. Conclusions ............................................................. 21

Acknowledgments .......................................................... 21

References ................................................................. 22
LIST OF SYMBOLS

A = See eq. (2.2)
A_1, B_1 = See eq. (2.7)
a = Radius of the source
a_n, b_n = Fourier coefficients (See eq. (1.2))
c_0 = Isentropic speed of sound at ambient values of pressure and density
h = Discretization in x-direction
IMAX = Number of mesh points across the axis
k = Step size in t-direction (Test problem)
L = Step size in x-direction (Problem from acoustics)
l_0 = Shock formation distance for a plane wave
NUMLI = Number of harmonics in calculation
p = Acoustic pressure normalized to peak amplitude on the source
P_{abs} = Parameter of absorption = \omega r_0
P_{non} = Parameter of nonlinearity = r_0/l_0
r = k/h^2
r_0 = Rayleigh distance
s = Abbreviation for (1+\sigma)
T = (1+\sigma)p
= See eq. (2.8)
t = Time
x = Dimensionless coordinate across axis = (x_1^2 + x_2^2)^{1/2}/(as)
x_1, x_2, z = Dimensional cartesian coordinates, z along the direction of propagation
\alpha = Linear absorption coefficient
\omega = Frequency of source
= See eq. (2.11)
\sigma = Dimensionless coordinate along acoustic axis z/r_0
\tau = Dimensionless retarded time, \omega(t-z/c_0) - x^2/s
1. Introduction.

The propagation of a nonlinear sound beam generated by a circular piston is considered. An equation obtained in 1971 by Kutznetsov [14] provides an excellent model for the combined effects of nonlinearity, diffraction and absorption. Throughout the 1970's there appeared in the Soviet literature a large volume of work based on Kutznetsov's equation. Bakhvalov and coworkers [3,4,5] presented numerical solutions, but not in the practical case of a circular piston. A numerical procedure for solving Kutznetsov's equation was given by Zhileikin [17].

In Norway the mathematicians Naze Tjøtta and Tjøtta and coworkers [6,9,10,11] studied analytical solutions in the case of weak nonlinearity. Aanonsen [1] developed a numerical procedure for calculating the harmonic content of the sound beam in the case of moderate nonlinearity. This procedure was used by Aanonsen et al. [2] to consider the nearfield of a finite amplitude sound beam. Hamilton et al. [12] introduced a transformation which facilitated numerical evaluation of the sound beam through the transition zone and into the farfield region.

In our notation Kutznetsov's equation is written

\[
\frac{\partial ^2 \tau}{\partial \sigma \partial \tau} = p_{\text{abs}} \frac{\partial ^3 \tau}{\partial \tau^3} + \frac{1}{4s} \nabla^2 \tau + \frac{p_{\text{non}}}{2s} \frac{\partial ^2 \tau}{\partial \tau^2} \tag{1.1}
\]

where

\[
\nabla^2 \tau = \left( \frac{\partial ^2}{\partial x^2} + \frac{1}{x} \frac{\partial}{\partial x} \right)
\]

and \( \sigma \) is the coordinate along the acoustic axis, \( x \) the coordinate across the axis, \( \tau \) retarded time, \( s=1+o \), \( T=(1+o)p \), where \( p \) is dimensionless pressure, \( p_{\text{abs}} \) parameter of absorption, and \( p_{\text{non}} \) parameter of nonlinearity. The first term on the right hand side is due to absorption, the second to diffraction and the third to nonlinearity. If a solution of equation (1.1) is sought in the form
\[ T = \sum_{n=1}^{\infty} \left( a_n \cos(n\tau) + b_n \sin(n\tau) \right) \]

the following set of coupled equations are obtained.

\[ \frac{\partial a_n}{\partial \sigma} = -p_{\text{abs}}^n a_n - \frac{1}{4n \sigma} \sum_{\pi=1}^{\infty} Y_{\pi \pi} b_n \]
\[ + p_{\text{non}} \sum_{p=1}^{n-1} \left[ \sum_{p=1}^{\infty} \left( b_n - a_n \right) - a_{n-p} - a_{n-p} \right] \]
\[ \frac{\partial b_n}{\partial \sigma} = -p_{\text{abs}}^n b_n + \frac{1}{4n \sigma} \sum_{\pi=1}^{\infty} Y_{\pi \pi} a_n \]
\[ + p_{\text{non}} \sum_{p=1}^{n-1} \left[ \sum_{p=1}^{\infty} \left( b_n + a_n \right) - a_{n-p} + a_{n-p} \right] \]

(1.3)

In order to solve system (1.3) both Zhileikhin and Aanonsen used finite difference techniques. Zhileikhin applied Crank-Nicolsons method while Aanonsen integrated the absorption term and the diffraction term by a fully implicit method and the nonlinear term by an explicit method. In sec. 2 we introduce a test problem which is a simplification of (1.3). The simplified system only accounts for diffraction, but this term have to be integrated with care in the case of a piston source. The numerical techniques applied to the test problem are analysed, and this analysis also gives insight in why the techniques behave as they do on problem (1.3).

In a user documentation by Berntsen and Vefring [7] three subroutines were presented for solving the problem

\[ \frac{\partial a_n}{\partial \sigma} = -c(n,\sigma)a_n + k(n,\sigma)\sum_{\pi=1}^{\infty} Y_{\pi \pi} b_n + \sum_{\pi=1}^{\infty} l_1(n,\sigma, a, b) \]
\[ \frac{\partial b_n}{\partial \sigma} = -c(n,\sigma)b_n - k(n,\sigma)\sum_{\pi=1}^{\infty} Y_{\pi \pi} a_n + \sum_{\pi=1}^{\infty} l_2(n,\sigma, a, b) \]

(1.4)

This system is a convenient generalisation of (1.3). The three subroutines differ in the way the Laplacian is approximated. In sec. 3 the three subroutines are compared. In sec. 4 we study the choice of step size in \( \sigma \)-direction.
2. Analysis of two numerical techniques on a test problem.


\[ \begin{align*}
\frac{\partial v}{\partial t} &= -\frac{\partial^2 w}{\partial x^2} & & 0 < x < 1, \quad t > 0 \\
\frac{\partial w}{\partial t} &= -\frac{\partial^2 v}{\partial x^2}
\end{align*} \]  

subject to initial conditions \( v(x,0) = v_0(x), \quad w(x,0) = w_0(x), \) and boundary conditions \( v(0,t) = f_0(t), \quad v(1,t) = f_1(t), \quad w(0,t) = g_0(t), \quad w(1,t) = g_1(t), \) is studied. We hoped that the methods developed for solving (2.1) would be applicable to our system (1.3), since (2.1) may be considered as an simplification of (1.3), and we will therefore describe a solution technique due to Richtmeyer [15].

We introduce some notation used by Fairweather and Gourlay[8]

\[
\Omega = \begin{pmatrix} v \\ w \end{pmatrix} \quad A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}
\]  

(2.2)

(2.1) may then be rewritten

\[ \frac{\partial \Omega}{\partial t} = A \frac{\partial^2 \Omega}{\partial x^2} \]  

(2.3)

A rectangular network of points with mesh sizes h and k in the x and t directions respectively, where Nh=1 is superimposed on the region \( 0 < x < 1, \quad t > 0. \) The values of the functions \( v(x,t), w(x,t) \) and \( \Omega(x,t) \) at the mesh points \( x = ih, t = nk; i = 0, 1, \ldots, N; n = 0, 1, \ldots \) are given by \( v_i,n, w_i,n \) and \( \Omega_i,n \) respectively.

When we apply Crank-Nicolsons method[16] to (2.3), we obtain the algorithm used by Richtmeyer in the form
\[
(I - \frac{1}{2} rA\delta^2_x)\Omega_{i,n+1} = (I + \frac{1}{2} rA\delta^2_x)\Omega_{i,n} \quad i=1,\ldots,N-1
\]  

(2.4)

where \( r = k/h^2 \), \( I \) is the 2x2 unit matrix and \( \delta^2_x \) is the usual central difference operator in the \( x \)-direction.

In each time step we have to solve a system of \((N-1)\) linear equations for the \((N-1)\) unknowns

\[
\Omega_{i,n} = \begin{pmatrix} v_{i,n} \\ w_{i,n} \end{pmatrix} \quad i=1,\ldots,N-1
\]

(2.5)

which may be written in the form

\[
A_1 W_{n+1} = B_1 W_n + c
\]

(2.6)

where \( W_m = [\Omega_{1,m}, \ldots, \Omega_{N-1,m}]^T \), \( m = n, n+1 \), and \( c \) is a constant vector arising from the boundary conditions. \( A_1 \) and \( B_1 \) are given by

\[
A_1 = I + \frac{1}{2} rT \quad B_1 = I - \frac{1}{2} rT
\]

(2.7)

where \( T \) is a block tridiagonal matrix given by

\[
T = \begin{pmatrix}
2A & -A \\
-A & 2A & -A \\
-A & 2A & -A \\
-A & 2A & -A
\end{pmatrix}
\]

(2.8)

The block tridiagonal system (2.6) may be solved by the well known algorithm described in Richtmeyer[15] where also the stability properties of the algorithm are proved to be very satisfactory.

If we integrate the nonlinear terms in (1.3) by an explicit difference formula, the described method for solving (2.1) may also be applied to (1.3). If we compare this method with the method described by Aanonsen [1], the two methods differ in two ways:
1. Richtmeyers method uses an implicit 2nd order approximation in $\sigma$-direction, whereas Aanonsen uses a fully implicit first order approximation.

2. The numbering of the linear equations is different. If we use the notation (2.2), we turn the wide band matrix in [1] into a narrow band matrix which can be solved by a direct equation solver. In [1] an iterative solution technique is used for solving the linear system of equations appearing in each step in $\sigma$-direction. In order to make the iterations converge we have to impose constraints on the step size in $\sigma$-direction ($r < 1/2$). When we apply Richtmeyers method, we have no such constraints on the step size, and moreover the solution of (2.6) is very stable. Tests show that we get $W_{N+1}$ to almost full machine precision.

We have tried to use Richtmeyers method to solve equation (1.3). The initial conditions in (1.3) are typically discontinuous, and the numerical solution is highly oscillatory. In [8] (2.1) is used as a test problem together with the initial and boundary conditions

$$Q(x,0) = \begin{cases} 0 & \text{if } x < 1 \\ x^2 & \text{if } x \geq 1 \end{cases}$$

(2.9)

$$Q(0,t) = Q(1,t) = \begin{cases} 0 & \text{if } t > 0 \end{cases}$$

This problem has a theoretical solution, see [8]. However, if we replace (2.9) with
\[ \mathbf{Q}(x,0) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad 0 \leq x < 0.1 \text{ and } 0.9 < x < 0.1 \]

\[ \mathbf{Q}(x,0) = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad 0.1 \leq x < 0.9 \]

\[ \mathbf{Q}(0,t) = \mathbf{Q}(1,t) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad t > 0 \]

and solve this problem with Richtmeyer's method, here also unphysical oscillations appeared in the solution. We replaced the 2nd order method with the fully implicit method and solved the same problem. The oscillations then disappeared completely. In fig 2.1 we plot the numerical approximations to \( v(x,0.2) \) given by the two methods, \( h=0.05 \) and \( k=0.00125 \). The plots for \( w(x,0.2) \) are similar.

Fig 2.1. A comparison of the 2nd order method and the fully implicit method on problem (2.10).
The explanation of this phenomena must lie in the stability properties of the two methods. In [15] von Neumanns technique for studying stability properties is applied on the 2nd order method. They define

\[ w = 4r \sin^2 \left( \frac{\beta h}{2} \right) \]  \hspace{1cm} (2.11)

The amplification matrix is then

\[ G_1 = \frac{1}{1 + w^2/4} \begin{pmatrix} 1 - w^2/4 & w \\ -w & 1 - w^2/4 \end{pmatrix} \]  \hspace{1cm} (2.12)

The eigenvalues of \( G_1 \) lie on the unit circle, and the numerical solution is always stable.

We have used the same technique to study the stability properties of the fully implicit method applied to (2.1), and the amplification matrix is

\[ G_2 = \frac{1}{1 + w^2} \begin{pmatrix} 1 & w \\ -w & 1 \end{pmatrix} \]  \hspace{1cm} (2.13)

The eigenvalues of this matrix are

\[ \lambda_{1,2} = \frac{1 \pm iw}{1 + w^2} \]  \hspace{1cm} (2.14)

and thus, for \( \beta h \) close to odd multiples of \( \pi \), we should have a dramatic (and unphysical) numerical damping when we apply the fully implicit method.

When we apply this method to (2.1) with the initial and boundary conditions given in (2.10), the damping effect is almost negligible compared to what could be expected from (2.14). We have therefore applied the matrix method, see [16], in order to study in more detail the stability properties of the two methods. The eigenvectors of the matrices \( A_1 \) and \( B_1 \) in (2.7) are the same as the eigenvectors of the matrix \( T \). Using the theory in [16] we find the eigenvalues of \( T \) to be
\[ \lambda_{\pm s} = \pm \sin^2(s\pi/N) \quad s=1,2,\ldots,N-1 \]  
\hspace{1cm} (2.15)

and the corresponding eigenvectors to be

\[
V_{\pm s} = \begin{pmatrix} 
\sin(s\pi/N) \\
\pm\sin(s\pi/N) \\
\sin(2s\pi/N) \\
\pm\sin(2s\pi/N) \\
& \ddots \\
& & \sin((N-2)s\pi/N) \\
& & \pm\sin((N-2)s\pi/N) \\
& & \sin((N-1)s\pi/N) \\
& & \pm\sin((N-1)s\pi/N) \\
\end{pmatrix} \quad s=1,2,\ldots,N-1 \]  
\hspace{1cm} (2.16)

The eigenvalues of \( A_1^{-1}B_1 \) will then be

\[
\lambda_{\pm s} = \frac{1 \pm 2\sin^2(s\pi/2N)}{1 \pm 2\sin^2(s\pi/2N)} \quad s=1,2,\ldots,N-1 \]  
\hspace{1cm} (2.17)

which lie on the unit circle.

When we apply the fully implicit method on (2.1), the corresponding system of linear equations may be written

\[
A_2 W_{n+1} = B_2 W_n + c \]  
\hspace{1cm} (2.18)

where \( A_2 = I + rT \) and \( B_2 = I \). The eigenvalues of \( A_2^{-1}B_2 \) will now be

\[
\lambda_{\pm s} = \frac{1 \pm 4\sin^2(s\pi/2N)}{1 + 16r^2\sin^2(s\pi/2N)} \quad s=1,2,\ldots,N-1 \]  
\hspace{1cm} (2.19)

This means that we get the damping predicted from (2.14), but the damping will be dependent of \( s \), and we will take a closer look at what this means.
We may use the eigenvectors (2.16) to expand the initial values of $Q$. If the initial values are represented as a slowly varying function, the function may be represented well by the first few terms (small values of $s$). The highly oscillating eigenvectors will almost be missing, and the 2nd order method will prohibit growth of these terms. The use of the fully implicit method will on the other hand cause no dramatic damping for reasonable $r$'s since only the smallest values of $s$ are interesting. For instance when we apply the fully implicit method on (2.1) with initial conditions (2.9) ($k=0.00125$ and $h=0.05$), we would expect the theoretical damping of the first eigenvector ($s=1$) to be $0.9999242$ per step and the corresponding damping of the last eigenvector ($s=N-1$) to be $0.5765$. If we integrate to $t=0.02$, the average damping is $0.9999152$ which is very close to what we expected for the first eigenvector. Here we measured the solutions in the rms norm

$$\|q\|_{\text{rms}} = \left[ \frac{\int |q|^2}{M} \right]^{1/2}$$  \hspace{1cm} (2.20)

where $q=(v^2+w^2)^{1/2}$ and $M$ is the number of meshpoints. On the other hand if the initial values are represented by a step function, as is the case in our application, the higher oscillating eigenvectors are used in a much greater extent to represent the initial values. If we use the 2nd order method, these Gibbs oscillations will be maintained in the numerical solution. If we use the fully implicit method, these unwanted oscillations will be damped. The fully implicit method acts as a filter on the higher harmonics and this explains why Aanonsen found through his numerical experiments the fully implicit method to be the best choice. When we apply the fully implicit method on (2.1) with initial conditions (2.10) ($k=0.00125$ and $h=0.05$) and integrate to $t=0.02$, the average damping is $0.99804$. Here the damping is clearly greater than in the previous case. However by the time we get to $t=0.02$ the higher harmonics have lost much of their strength, and the damping per step will be reduced as we proceed. The damping per step from $t=0.02$ to $t=0.04$ is $0.9997859$ and from $t=0.04$ to $t=1.0$ it is $0.9998153$. 
To conclude we would like to remark that the insight gained by studying our methods on the test problem (2.1) has given us a deeper understanding of why the methods behave as they do on our problem (1.3). The oscillations appearing when we solve (1.3) with Richtmeyers method, stem from the initial conditions and not from the extra terms in (1.3). The fully implicit method is for reasonable values of r an appropriate method for solving (1.3). This is in agreement with Aanonsens experience. On the other hand we want to keep Richtmeyers numbering of the equations to get a narrow band solution matrix in each step in o-direction. We will then be relieved from the constraints on the step size which Aanonsens method suffers from.
3. Choice of difference operator and step size in $x$-direction.

In a user documentation by Berntsen and Vefring [7] three subroutines for solving (1.4) are given. The three subroutines differ in the way the Laplacian is approximated. Table 1 shows the algebraic formulas used

<table>
<thead>
<tr>
<th>Finite difference formula</th>
<th>$u_x$</th>
<th>$u_{xx}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three point</td>
<td>$(-u_{i-1}+u_{i+1})/2h$</td>
<td>$(u_{i-1}-2u_i+u_{i+1})/h^2$</td>
</tr>
<tr>
<td>Five point</td>
<td>$\left(\frac{u_{i-2} - 8u_{i-1} + 8u_{i+1} - u_{i+2}}{12h}\right)$</td>
<td>$\left(\frac{-u_{i-2} + 16u_{i-1} - 30u_i + 16u_{i+1} - 4u_{i+2}}{12h^2}\right)$</td>
</tr>
<tr>
<td>Seven point</td>
<td>$\left(\frac{-u_{i-3} + 9u_{i-2} - 45u_{i-1} + 45u_{i+1} - 9u_{i+2} + u_{i+3}}{60h}\right)$</td>
<td>$\left(\frac{2u_{i-3} - 27u_{i-2} + 270u_{i-1} - 490u_i + 270u_{i+1} - 27u_{i+2} + 2u_{i+3}}{180h^2}\right)$</td>
</tr>
</tbody>
</table>

To compare accuracy and computational efficiency we state a test problem.

$$a_1(x, \sigma = 0) = \begin{cases} \sin(x^2) & \text{when } 0 < x < 1 \\ 0 & \text{when } 1 < x < 8 \end{cases}$$

$$b_1(x, \sigma = 0) = \begin{cases} \cos(x^2) & \text{when } 0 < x < 1 \\ 0 & \text{when } 1 < x < 8 \end{cases}$$  \hspace{1cm} (3.1)
\[ a_2(x, \sigma = 0) = 0 \quad 0 \leq x \leq 8 \]
\[ b_2(x, \sigma = 0) = 0 \]

cont. (3.1)

\[ a_n(x=8, \sigma) = 0 \quad \sigma > 0, n=1,2 \]
\[ b_n(x=8, \sigma) = 0 \]

The terms \( c(n, \sigma), k(n, \sigma), i_{11}(n, \sigma, a, b) \) and \( i_{12}(n, \sigma, a, b) \) are defined according to (1.3) with \( p_{\text{abs}} = 4.4 \times 10^{-6} \) and \( p_{\text{non}} = 3.4 \times 10^{-2} \).

The parameters of absorption and nonlinearity are chosen relatively small because we want diffraction to be the major effect. The number of harmonics retained in the calculations are two. Solutions to this problem are computed using different subroutines and different discretization in the \( x \)-direction. In all cases \( k \) is \( 3.5 \times 10^{-4} \). In the experiments described in this and the next section, we adjust \( k \) by multiplying by \((1+\sigma)^2\) as we proceed in order to keep \( k/(h^2(1+\sigma)^2) \) constant, see (1.3).

The numerical solutions are computed by marching step by step in the \( \sigma \)-direction. For each step and each harmonic nonlinear terms have to be calculated and a system of linear equations have to be solved. Table 2 shows the computational cost of the different subroutines.

**TABLE 2**

<table>
<thead>
<tr>
<th>Linear system</th>
<th>Nonlinear terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three point</td>
<td>( 26^*\text{IMAX}\times\text{NUMLI} ) +</td>
</tr>
<tr>
<td>Five point</td>
<td>( 60^*\text{IMAX}\times\text{NUMLI} ) +</td>
</tr>
<tr>
<td>Seven point</td>
<td>( 111^*\text{IMAX}\times\text{NUMLI} ) +</td>
</tr>
</tbody>
</table>
IMAX defines the number of mesh points in x-direction and NUMLI the number of harmonics.

We define

\[ p_1 = \left( a_1^2 + b_1^2 \right)^{1/2} / (1+\sigma) \]
\[ p_2 = \left( a_2^2 + b_2^2 \right)^{1/2} / (1+\sigma) \]  

(3.2)

where \( p_1, p_2 \) are the normalized peak pressure amplitudes for the fundamental and second harmonic respectively. The step size across the axis was halved until the difference between to consecutive solutions were nonsignificant. The results from computations with the seven point approximation and \( h = 0.1875 \times 10^{-2} \) were taken as an estimate of the exact solution and defines \( p_{1e} \) and \( p_{2e} \).

Amplitudes at \( \sigma = 0.5, 1.0 \) and 1.5 were compared. Only results at \( \sigma = 1.0 \) are shown. The results for \( \sigma = 0.5 \) and \( \sigma = 1.5 \) are similar. Table 3 shows the results for the primary.

**TABLE 3**

<table>
<thead>
<tr>
<th>[|p_1 - p_{1e}|_{\text{rms}}]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h )</td>
</tr>
</tbody>
</table>

| | \( 0.85 \times 10^{-2} \) | \( 0.84 \times 10^{-2} \) | \( 0.83 \times 10^{-3} \) |
| Three point | \( 0.42 \times 10^{-2} \) | \( 0.26 \times 10^{-2} \) | \( 0.84 \times 10^{-3} \) |
| Five point | \( 0.42 \times 10^{-2} \) | \( 0.25 \times 10^{-2} \) | \( 0.84 \times 10^{-3} \) |

Results for the second harmonic are shown in Table 4.
TABLE 4

\[ \|p_2 - p_2^e\|_{\text{rms}} \]

<table>
<thead>
<tr>
<th>h</th>
<th>(0.60 \times 10^{-1})</th>
<th>(0.30 \times 10^{-1})</th>
<th>(0.15 \times 10^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three point</td>
<td>(0.70 \times 10^{-4})</td>
<td>(0.86 \times 10^{-4})</td>
<td>(0.18 \times 10^{-4})</td>
</tr>
<tr>
<td>Five point</td>
<td>(0.10 \times 10^{-3})</td>
<td>(0.66 \times 10^{-4})</td>
<td>(0.21 \times 10^{-4})</td>
</tr>
<tr>
<td>Seven point</td>
<td>(0.10 \times 10^{-3})</td>
<td>(0.66 \times 10^{-4})</td>
<td>(0.21 \times 10^{-4})</td>
</tr>
</tbody>
</table>

We are aware that this is a very limited experiment. The results presented in tables 3 and 4 are, however, representative for what we have observed for other choices of parameters.

The most obvious conclusion is that there is nothing to gain by using the seven point approximation. In Hamilton et al. [12] they remark that it is most efficient to use the seven point approx. In [12], however, they use an iterative solution technique for the systems of linear equations, and the extra cost in going from a three point approximation to a seven point approximation is negligible.

To choose between the three point approximation and the five point approximation is more difficult. The cost of using the five point approximation with \(h=0.06(0.03)\) is almost equal to the cost of using the three point approximation with \(h=0.03(0.015)\) when only a few harmonics are retained in the solution. From tables 3 and 4 we see that in 3 out of 4 cases the three point approximation is more reliable than the five point approximation we get to the same cost. This indicates that the best choice is the three point approximation. However, as the number of harmonics increases we see from Table 2 that the work with computing the nonlinear terms becomes dominant and if we use the five point approximation, we can hope to achieve the same accuracy with a smaller \(\text{IMAX}\). (As we see from Table 4 this is not always the case.)
4. Choice of step size in $\sigma$-direction.

Before we discuss how to choose step sizes on our problem (1.3), we will study the effect of varying the discretisation on the test problem (2.1). When we apply the fully implicit method, the eigenvectors are given by (2.16) and the eigenvalues by (2.19). The parameter defining the damping is

$$w(s) = 4rsin^2(s\pi/2N) \quad s=1,2,\ldots,N-1$$

(4.1)

(4.1) gives the impression that the damping is a parameter of $r=k/h^2$. This is however only to some extent true. The first eigenvalue is a function of

$$w(1) = 4rsin^2(\pi/2N) = 4rsin^2(\pi h/2) \approx km^2$$

(4.2)

and we see that the damping of the first eigenvector is a function of $k$ (and not of $h$). This means that if we reduce $h$, we do not have to reduce $k$ to keep $r$ unchanged if we are concerned about the damping of the first eigenvector only.

As we have seen in sec.2 only the lower harmonics are maintained in the numerical solution when we apply the fully implicit method, and the above comment on the first harmonic also to some extent applies to the other lower harmonics. This means that if we want the same damping when we reduce $h$, we can almost keep $k$ unchanged. For small values of $h$ we can thus expect relatively large values of $r$ to give a reasonable damping.

We have run our problem (1.3) with $p_{abs} = p_{non} = 0$ for two values of $h$ and different values of $k$ to illustrate this effect. Since $p_{non} = 0$ only the fundamental is present in the sound beam. The problem is integrated to $\sigma=1$. We define the damping to be the rms norm, see (2.20), of the numerical solution measured at $\sigma=1$ divided to the rms norm of the initial values. The damping is affected both by the diffraction and the numerical damping caused by the use of the fully
implicit method. In order to study the effect of increasing $k$ or $r$, relative damping is also defined. We define the relative damping to be the rms norm of the numerical solution for a specific $k$ measured at $x=1$ divided to the rms norm of the numerical solution for $k=3.5 \times 10^{-5}$ measured at $x=1$. Tables 5 and 6 show the damping for $h=0.03$ and $h=0.096$ respectively.

**TABLE 5**

Damping for $h=0.03$

<table>
<thead>
<tr>
<th>$r$</th>
<th>$3.8 \times 10^{-2}$</th>
<th>$3.8 \times 10^{-1}$</th>
<th>$3.8 \times 10^{0}$</th>
<th>$3.8 \times 10^{1}$</th>
<th>$3.8 \times 10^{2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>$3.5 \times 10^{-5}$</td>
<td>$3.5 \times 10^{-4}$</td>
<td>$3.5 \times 10^{-3}$</td>
<td>$3.5 \times 10^{-2}$</td>
<td>$3.5 \times 10^{-1}$</td>
</tr>
<tr>
<td>Damp.</td>
<td>0.6406</td>
<td>0.6390</td>
<td>0.6369</td>
<td>0.6201</td>
<td>0.5408</td>
</tr>
<tr>
<td>Rel. Damp.</td>
<td>1.0000</td>
<td>0.9989</td>
<td>0.9943</td>
<td>0.9681</td>
<td>0.8443</td>
</tr>
</tbody>
</table>

**TABLE 6**

Damping for $h=0.096$

<table>
<thead>
<tr>
<th>$r$</th>
<th>$3.8 \times 10^{-3}$</th>
<th>$3.8 \times 10^{-2}$</th>
<th>$3.8 \times 10^{-1}$</th>
<th>$3.8 \times 10^{0}$</th>
<th>$3.8 \times 10^{1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>$3.5 \times 10^{-5}$</td>
<td>$3.5 \times 10^{-4}$</td>
<td>$3.5 \times 10^{-3}$</td>
<td>$3.5 \times 10^{-2}$</td>
<td>$3.5 \times 10^{-1}$</td>
</tr>
<tr>
<td>Damp.</td>
<td>0.6776</td>
<td>0.6691</td>
<td>0.6606</td>
<td>0.6420</td>
<td>0.5495</td>
</tr>
<tr>
<td>Rel. Damp.</td>
<td>1.0000</td>
<td>0.9873</td>
<td>0.9749</td>
<td>0.9474</td>
<td>0.8109</td>
</tr>
</tbody>
</table>

From tables 5 and 6 we observe that for $h$ constant the damping increase with increasing $r$. This is in agreement with the results from the analysis of the test problem in sec. 2.

When $r$ is increased the high frequency components are killed and we expect the damping to be characterized mainly by $k$, and we see that for $k=3.5 \times 10^{-1}$ the damping for $h=0.03$ is almost equal to the damping
for $h=0.096$. When we reduce $k$, the damping for $h=0.03$ tends to be greater than the corresponding damping for $h=0.096$.

From Table 6 we see that the relative damping for $h=0.096$ and $r=3.8\times10^{-1}$ is 0.9749. From Table 5 we find that the relative damping for $h=0.03$ and $r=3.8\times10^{-1}$ is 0.9989 and for $r=3.8$ it is 0.9943. This illustrates that for small values of $h$ the damping will be acceptable for greater values of $r$. The step sizes used in [12] are $h=0.03$ and $k=3.5\times10^{-4}$ ($r=3.8\times10^{-1}$). Table 5 shows that the extra damping introduced by multiplying this $k$ by 10 is quite small. This suggests that computations can be performed with reasonable accuracy when we choose $k$ to be $3.5\times10^{-3}$ thus reducing the computation time by 90%.
5. Conclusions.

The main achievement in this report we consider to be the analysis of two numerical techniques with help of a test problem. This analysis has given us a much better understanding of why the numerical techniques perform as they do also on the problem (1.3) from acoustics, and we believe that this insight may be used to improve the numerical techniques even further.

From a practical point of view the main achievement is that the restriction on the step size in \( \sigma \)-direction appearing in [1] is removed. This means that for reasonable discretizations in \( x \)-direction we may use approximately 10 times larger step sizes and thereby save the same amount of computer time.

Acknowledgments.

This report has been supported by the Norwegian Council of Humanities and Sciences and STATOIL, Norway.
References.


