On Difference Schemes for Hyperbolic Equations with Discontinuous Initial Values

By Mats Y. T. Apelkrans

1. Introduction. In this paper we consider the hyperbolic equation

$$\begin{aligned} &\frac{\partial u}{\partial t} = \rho(x,t)\frac{\partial u}{\partial x} + \phi(x,t)u \quad \text{with initial values} \\ &u(x,0) = f(x) \end{aligned} \tag{1.1}$$

in the upper halfplane. It is well known that if $f(x)$ is a sufficiently smooth function (1.1) can be closely approximated by stable difference schemes, and very realistic error bounds are given in a number of papers (see, e.g., Lax [6]). But in applications there often arise initial functions, with simple discontinuities or discontinuities in the higher derivatives. A discontinuity in $f(x)$ propagates in the solution to (1.1) along a characteristic, but this propagation is disturbed in the solution of the corresponding difference equation. It is the aim of this paper to give rather sharp bounds for the error and to show how these bounds depend on the order of accuracy and dissipation of the difference scheme at hand.

We show that the error between the solutions of (1.1) and the corresponding difference approximation has two components E_1 and E_2, where $E_1 = O(h^p)$, p is the order of accuracy. The other component $E_2 = O(e^{-a})$, where $a = N^q d(x,t)$ with $0 < q < 1$. Here $d(x,t)$ is the distance (cf. Section 2) from the characteristic to (1.1) through the jump-point. $h = N^{-1}$ is the mesh-size and q depends on p and the order of dissipation. Furthermore, for constant coefficients we can take $E_1 \equiv 0$. In Section 2 we list a number of results. The first one shows that if p is odd, then a dissipative scheme is necessarily dissipative of order $p + 1$. But for even p, the dissipative schemes with largest q are those with dissipation of order $p + 2$. In Section 5, where we list the results of a number of numerical experiments, Experiment 9 shows that a nondissipative scheme can behave very badly; e.g., if $\rho \equiv 1$, $\phi \equiv 0$ and $N = 200$, then the error for $t = 1$, is bigger than 0.01 in an interval of length 2.

The main idea of this paper, due to H.-O. Kreiss, is to extend the real variable ξ to the complex variable $\xi + i\alpha$ in the symbol $\hat{Q}(\xi)$ of the difference operator at hand. This technique is also used in a paper of H.-O. Kreiss and E. Lundqvist [5], where they consider a mixed problem for a hyperbolic equation.

The main part of our proofs, collected in Section 3, consists of algebraic manipulations of the transformed symbol $\hat{Q}(\xi + i\alpha)$, which lead to the definition of contractivity (Definition 2.4). Contraction can be considered as an algebraic condition on the difference scheme. Now, $\hat{Q}(\xi + i\alpha)$ is the symbol of a difference operator that originates from an equation obtained by a simple transformation of the original difference equation. This fact is used in Section 4, where the rest of the proofs are found. For the variable coefficients case, we also use a theorem of Lax-Nirenberg [7].

Received December 11, 1967.

In a forthcoming paper we intend to generalize our results to symmetric hyperbolic systems with variable coefficients.

There are some earlier results which ought to be mentioned. First, H.-O. Kreiss [4] has shown that the same relation between p and the order of dissipation $2s = p + 1$ or $2s = p + 2$ guarantees L.R.-stability for difference approximations to hyperbolic systems in any number of independent variables. (L.R.-stable means stable in the sense of Lax and Richtmyer.) B. Parlett has also discussed stability of dissipative schemes in [8], with the same relations between p and $2s$. We should also mention papers of V. Thomée [11], J. Peetré and V. Thomée [9], G. Hedström [3] and R. P. Fedorenko [2], which consider problems with discontinuous initial functions.

In [11] V. Thomée has proved a theorem like ours, but in his proof he uses a quite different technique containing Fourier-Stieltjes transforms of functions of bounded variation. His results are restricted to difference operators, which are stable in maximum-norm. But many of the difference schemes used in practice do not belong to this class; e.g., every scheme with even order of accuracy. Furthermore, nothing is said about equations with variable coefficients.

In [9], J. Peetré and V. Thomée give estimates for the rate of convergence of a very general class of difference schemes. They use for their proof the theory of interpolation of Banach spaces. In our special case they give an L_2-estimate of the error of order h^a, where $a = \frac{1}{2}p/(p + 1)$. These estimates are better than the corresponding L_2-estimates using our results. However, an L_2-estimate in the problem of this paper gives quite misleading information, because the local error outside an interval of length $O(h^q|\log h|)$ is only of $O(h^p)$.

In [3] G. W. Hedström gives an estimate for the constant coefficient case. His results give $E_1 = O(h^p)$ and $E_2 = O(e^{-a})$, where $a = \text{const}\, N \cdot d(x, t)^{(p+1)/p}$ for $d(x, t) \geqq \text{const}\, N^{-p/(p+1)}$. These results are therefore more precise* in this part of the x-axis, but are less general. In his proof he uses estimates of the norms of powers of absolutely convergent Fourier series. Generalizations to variable coefficients can hardly be treated with this method.

Finally, R. P. Fedorenko [2] describes a method that gives very good numerical results in a number of tested examples. The paper is of experimental character and contains no proofs. He starts with a high order method, say $p = 3$, but when he comes to the discontinuity he changes the method to $p = 1$ in order to avoid the parasitic waves, "the Gibbs phenomenon," and afterwards, when the danger is over, he again takes the first method with $p = 3$. In order that the computer should recognize the discontinuities, he lets it compare the second and first differences in every mesh-point. This works well for simple examples like $\partial u/\partial t = \partial u/\partial x$, but to generalize to more complicated equations without this a priori knowledge about the characteristics seems to be hard.

Acknowledgements. I should like to thank my teacher H.-O. Kreiss who proposed this problem to me and who guided me in a very unselfish way. I am also very grateful to O. Widlund, who has read the manuscript and given me many valuable comments.

I am very indebted to James M. Varah, who made my English understandable and suggested some improvements.

---

* Compare Experiment 7 in Section 5.

2. Preliminaries and Statement of Results. Consider the hyperbolic equation

$$\frac{\partial u}{\partial t}(x, t) = \rho(x, t)\frac{\partial u}{\partial x}(x, t) + \phi(x, t)u(x, t)\,, \tag{2.1}$$

where ρ and ϕ are real valued functions, $-\infty < x < \infty$ and $t \geqq 0$. Its (generalized) solution is uniquely determined by initial values

$$u(x, 0) = f(x)\,. \tag{2.2}$$

We want to study the behavior of difference approximations to this problem. We therefore introduce a time-step $k > 0$, a mesh-width h, and grid-points $x_\nu = \nu h$, $\nu = 0, \pm 1, \pm 2, \cdots$. Furthermore let $k/h = \lambda$, $\lambda > 0$ constant. Denoting by $v_\nu(t) = v(x_\nu, t)$ a function defined for all $t = t_n = nk$, $n = 0, 1, 2, \cdots$ we approximate (2.1), (2.2) either by

$$Q_1 v_\nu(t + k) = Q_2 v_\nu(t)\,, \tag{2.3}$$

or by

$$v_\nu(t + k) = Q_0 v_\nu(t)\,, \tag{2.3'}$$

with initial conditions in either case

$$v_\nu(0) = f(x_\nu)\,. \tag{2.4}$$

Q_j, $j = 0, 1, 2$ are difference operators of the form

$$Q_1(x, t, h) = \sum_{j=-l}^{m} b_j(x, t, h)E^j\,,$$

$$Q_2(x, t, h) = \sum_{j=-l}^{m} a_j(x, t, h)E^j$$

and

$$Q_0(x, t, h) = \sum_{j=-\infty}^{\infty} d_j(x, t, h)E^j$$

respectively, where E is the translation operator defined by

$$Eg_\nu = g_{\nu+1}\,.$$

We always assume that a_j, b_j and d_j are smooth functions of h. To get an algebraic description of the behavior of the solution of (2.3) or (2.3') we introduce the symbols

$$\hat{Q}_j(x, t, \xi, h) = e^{-i\omega x}Q_j e^{i\omega x}\,, \qquad j = 0, 1, 2\,.$$

We always assume that for the implicit scheme (2.3)

$$|\hat{Q}_1(x, t, \xi, h)| \geqq \text{const}$$

for all x, t, ξ and for sufficiently small h.

To be complete, we have assumed h-dependent coefficients in (2.3) and (2.3'), but the essential features of a difference scheme come from the principal part of the difference operator, i.e. when $h = 0$. From the assumption above, we see that $Q_1^{-1}(x, t, 0)$ is a bounded operator in L_2 and hence

$$\hat{Q}_{-1}(x, t, \xi, 0) = \hat{Q}_1^{-1}(x, t, \xi, 0)\hat{Q}_2(x, t, \xi, 0)$$

is well defined.

Furthermore, we define the solution operator S of (2.3)–(2.4) or (2.3′)–(2.4) by

$$v_\nu(t) = S(t, h)f(x_\nu) .$$

We also introduce a discrete L_2-norm $\|\cdot\|_h$ by the scalar-product

$$(f, g)_h = \sum_{\nu=-\infty}^{\infty} hf_\nu g_\nu , \qquad \|f\|_h^2 = (f, f)_h ,$$

where f and g are mesh-functions.

Some definitions are now in order.

Definition 2.1. The difference scheme (2.3) or (2.3′) is stable if the operator $S(t, h)$ is uniformly bounded in L_2 for $0 \leqq t \leqq T$, independent of h.

Definition 2.2. The difference scheme (2.3) or (2.3′) is accurate of order p, (consistent if $p \geqq 1$), if p is the largest integer such that for all sufficiently differentiable solutions $u(x, t)$ of (2.1)

$$u(x, t + k) = Q_i(x, t, h)u(x, t) + O(h^{p+1}) , \qquad h \to 0 , i = 0, -1 .$$

It is well known that a scheme (2.3) or (2.3′) which is accurate of order p satisfies

$$\hat{Q}_j(x, t, \xi, 0) = e^{ip(x,t)\lambda\xi} + c_0(x, t)\xi^{p+1}(1 + o(1)) , \qquad \xi \to 0 , \tag{2.5}$$

(cf. Lax [6]), $j = 0, -1$.

Definition 2.3. A difference scheme defined by the operator $Q(x, t, h)$ is dissipative of order $2s$ if s is the least integer such that for all x and t, and for all ξ with $|\xi| \leqq \pi$,

$$|\hat{Q}(x, t, \xi, 0)| \leqq 1 - \delta|\xi|^{2s} ,$$

$\delta > 0$ some constant.

We now introduce another algebraic condition in terms of which we will state our main theorems.

Definition 2.4. Let α be a real constant. A difference scheme defined by the operator $Q(x, t, h)$ is contractive of order $\tau = (\tau_-, \tau_+)$ if there is a uniformly bounded function $\sigma(x, t)$, independent of ξ, such that, for α sufficiently small,

$$\hat{Q}(x, t, \xi + i\alpha, 0) = \exp(-\alpha\lambda\rho(x, t) + \sigma(x, t)|\alpha|^\tau)\cdot R(x, t, \xi) , \qquad |\xi| \leqq \pi ,$$

with $\tau = \tau_-$ for $\alpha\rho < 0$ and $\tau = \tau_+$ for $\alpha\rho > 0$ and where $R(x, t, \xi)$ is such that $|R| \leqq 1$ for all x and t, and for all ξ with $|\xi| \leqq \pi$.

Let $r = 2s - p$; then it is easy to see that there are no dissipative difference schemes with $r \leqq 0$, and the following theorem says that for odd p we can only have $r = 1$.

THEOREM 2.1. *A dissipative difference scheme of odd order of accuracy p is dissipative of order* $p + 1$.

But for even order of accuracy p, we can construct schemes with $r = 2j$, $j = 1, 2, \cdots$. (Section 5, Experiment 11 gives an example.)

There are connections between τ, p and s for a given scheme, which we are going to describe in the next theorem.

THEOREM 2.2. *Suppose that the difference scheme* (2.3) *or* (2.3′) *is*

(i) *accurate of order p,*

(ii) *dissipative of order* $2s$, *and that*

(iii) $\rho(x, t)$ *is uniformly bounded and does not change sign; then, for* $p = 2j - 1$, $j = 1, 2, \cdots$, *it is contractive of order* $\tau = (p + 1, p + 1)$. *Moreover, for* $p = 2j$, $j = 1, 2, \cdots$, *if*

(iv) $ic_0(x, t)\rho(x, t) > 0$

is satisfied, then it is contractive of order $\tau = (6s/(r + 2), 2s/r)$. ($c_0(x, t)$ *was defined in* (2.5).)

We believe condition (iv) is satisfied for all schemes used in practice. It is satisfied for explicit schemes (2.3) of maximum accuracy. (Cf. Strang [10] and our report [1].)

THEOREM 2.3. *Consider the difference scheme* (2.3) *with* $Q_1 = I$ (*identity operator*) *and* $Q_2 = \sum_{j=-N}^{N} a_j E^j$ *of maximum accuracy, i.e.,* $p = 2N$. *For these difference schemes condition* (iv) *of Theorem* 2.2 *is satisfied, if* $|\rho\lambda| \leqq 1$.

The proofs of these three theorems will be given in Section 3. They are purely algebraic in nature. The next three theorems have analytical proofs, which are presented in Section 4.

In this paper we only consider initial functions $f(x)$, which are step-functions

$$f(x) = 0 \qquad x < 0$$
$$= 1 \qquad x \geqq 0\,. \tag{2.6}$$

More general discontinuous initial functions could be divided into a sum of a smooth function and step-functions, which are handled in our original report [1]. That this works is essentially due to the linear character of our problem. For smooth initial functions we can refer to [6].

The error estimates in Theorems 2.4–2.5 show that the influence to the error from the discontinuity-jump is exponentially decreasing with the distance $d(x, t)$ from the characteristic through the origin.

Definition 2.5. The distance $d(x, t)$ from the characteristic through the origin is given by $d(x, t) = |g(x, t)|$, where $g(x, t)$ satisfies $\partial g/\partial t = \rho(x, t)\partial g/\partial x$ with $g(x, 0) = x$.

We first study the constant coefficient case. Then $d(x, t) = |\rho t + x|$. It is no restriction to assume that $\phi = 0$ in (2.1). Furthermore, let T_{+1} denote the region in the (x, t)-plane, where $0 \leqq t \leqq T$ and where (x, t) lies to the right of the characteristic of (2.1) through the origin. T_{-1} is the corresponding region to the left of the same characteristic.

THEOREM 2.4. *Let* $\rho \neq 0$ *be a constant and let* $\phi \equiv 0$. *Consider the difference scheme* (2.3) *with coefficients independent of* x, t *and* h, *and let* $f(x)$ *be defined by* (2.6). *Suppose that*

(i) (2.3) *is consistent with* (2.1),

(ii) (2.3) *is contractive of order* $\tau = (\tau_-, \tau_+)$. *Then for* h *sufficiently small,*

$$|u(x_\nu, t) - v_\nu(t)| \leqq ch^{(q-1)/2} \exp\,(-h^{-q}|\rho t + x_\nu|)\,, \tag{2.7}$$

where c *is a constant. Here* $q = (\tau_\mp - 1)/\tau_\mp$, *with minus sign in* $T_{-\mathrm{sign}(\rho)}$ *and plus sign in* $T_{+\mathrm{sign}(\rho)}$.

We now turn to the variable coefficient case, and in order to be able to use a stability theorem of Lax and Nirenberg (see [7] and Section 4) we choose the scheme (2.3′). This theorem also needs some smoothness properties on the coefficients d_j in

(2.3′). To simplify our proof, we assume that $\rho(x, t)$ and $\phi(x, t)$ have compact support. Here $\|\cdot\|$ means the ordinary L_2-norm.

THEOREM 2.5. *Consider the difference scheme* (2.3′) *with initial values defined by* (2.6). *Suppose that*

(i) (2.3′) *is accurate of order* p,

(ii) (2.3′) *is contractive of order* $\tau = (\tau_-, \tau_+)$,

(iii) $\rho(x, t)$ *and* $\phi(x, t) \in c_0^{\infty}$,** *and* $\rho(x, t)$ *does not change sign.*

(iv)
$$\sum_j \max\left\{||d_j(\cdot, t, 0)||\,,\ \left\|\frac{\partial}{\partial x} d_j(\cdot, t, 0)\right\|,\ \left\|\frac{\partial^2}{\partial x^2} d_j(\cdot, t, 0)\right\|\right\} \leqq \text{const}$$
$$\sum_j ||d_j(\cdot, t, 0)||(1 + j^2) \leqq \text{const} \quad \textit{for all} \quad t\,.$$

Then, for h *sufficiently small,*

$$\begin{aligned} |u(x_\nu, t) - v_\nu(t)| &\leqq c_1 h^p + c_2 D(x_\nu, t) \quad \textit{in} \quad T_1\,, \\ &\leqq c_3 D(x_\nu, t) \qquad\quad \textit{in} \quad T_{-1}\,, \end{aligned} \tag{2.8}$$

where

$$D(x_\nu, t) = h^{(q-1)/2} \exp\left(-h^{-q} d(x_\nu, t)\right)$$

and where c_i *are constants independent of* x, t *and* h. *Here* $q = (\tau_\mp - 1)/\tau_\mp$, *with minus sign in* $T_{-\text{sign}(\rho)}$ *and plus sign in* $T_{+\text{sign}(\rho)}$.

Remarks. The assumption, that $\rho \in c_0^{\infty}$ simplifies the proof a lot, but it will of course work even for more general coefficients, which are sufficiently smooth. Furthermore, by modifying the proof of a stability theorem of Kreiss [4], it certainly will be possible to prove an analogous theorem for the scheme (2.3), with the following assumptions: (1) the coefficients in the scheme are Lipschitz continuous, and (2) $2s = p + 2 - l$, $l = 0, 1$.

Note that the q's of Theorem 2.4–2.5 satisfy $0 < q < 1$. Furthermore the best q's for dissipative schemes with even order of accuracy are obtained for schemes with $r = 2$. Then

$$\begin{aligned} q &= p/(p + 2) \qquad\quad \text{in} \quad T_{+\text{sign}(\rho)}\,, \\ &= (3p + 2)/(3p + 6) \quad \text{in} \quad T_{-\text{sign}(\rho)}\,. \end{aligned}$$

Example. The Lax-Wendroff scheme $Q_1 \equiv I$, $Q_2 = I + \rho k D_0 + \sigma\rho^2(k^2/2)D_+D_-$, where $2hD_0 = E - E^{-1}$, $hD_+ = E - I$ and $hD_- = I - E^{-1}$, is accurate of order 2 and dissipative of order 4 if $\sigma = 1$.

Hence, $q = 1/2$ in $T_{+\text{sign}(\rho)}$ and $q = 2/3$ in $T_{-\text{sign}(\rho)}$. If $\sigma \neq 1$, $p = 1$ and $2s = 2$, i.e., $q = 1/2$ on both sides of the characteristic. But the estimates (2.8) show that the local error outside an interval of length $O(h^q|\log h|)$ is only $O(h^2)$ for $\sigma = 1$. Therefore the solution to the Lax-Wendroff scheme with $\sigma = 1$, behaves fairly well even for discontinuous initial functions.

3. Proofs of Algebraic Theorems. We are going to prove Theorems 2.1–2.3 and begin by rewriting (2.5) in the form

** $h(x, t) \in c_0^{\infty}$ if h has partial derivatives of all order and $h \equiv 0$ outside a bounded region in the (x, t)-plane.

$$\hat{Q}(x, t, \xi, 0) = e^{i\rho(x,t)\lambda\xi + U(x,t,\xi)},$$

where

$$U(x, t, \xi) = \sum_{\nu=0}^{\infty} c_\nu(x, t)\xi^{p+\nu+1},$$

which is analytic in ξ. In this section $\hat{Q}$ stands for either $\hat{Q}_{-1}$ or $\hat{Q}_0$.

Proof of Theorem 2.1. Since (2.3) or (2.3′) is dissipative of some order $2s$, we have

$$e^{\mathrm{Re}\{U(x,t,\xi)\}} = |\hat{Q}(x, t, \xi, 0)| \leqq 1 - \delta|\xi|^{2s}, \qquad \delta > 0, \tag{3.1}$$

and thus $2s \geqq p + 1$. Write $G(i\xi) = \hat{Q}(x, t, \xi, 0)$, then

$$1 = |G(i\xi)/G(-i\xi)| = e^{2\mathrm{Im}(c_0)\xi^{p+1}+\cdots},$$

i.e., Im $(c_0(x, t)) = 0$ for odd p.

Moreover $c_0(x, t) \leqq 0$, since we otherwise could find a ξ, such that $|\hat{Q}| > 1$. But p is the order of accuracy and hence $c_0(x, t) \not\equiv 0$ and $c_0(x, t) < 0$. Thus we finally conclude that $2s = p + 1$.

Proof of Theorem 2.2. We first start with constant coefficients ρ and ϕ, where $\rho \neq 0$.

We are going to estimate $U(\xi + i\alpha) = \sum_{\nu=0}^{\infty} c_\nu(\xi + i\alpha)^{p+\nu+1}$ and therefore we state the following variant of Hölder's inequality.

Hölder's inequality. Let $A \geqq 0$, $B \geqq 0$ and $\epsilon > 0$ be given. Then

$$AB \leqq \eta A^r + \epsilon B^q,$$

where $r = q/(q - 1)$, $q > 1$ and $\eta = \eta(\epsilon)$ is uniformly bounded for $\epsilon \geqq \epsilon_0 > 0$.

From (3.1) and assumption (ii) it follows that

$$|\hat{Q}(\xi + i\alpha, 0)| = \exp\left(-\alpha\rho\lambda - \delta'\xi^{2s} + H(\alpha, \xi, p)\right),$$

where $\delta' > 0$.

Here

$$\begin{aligned} H(\alpha, \xi, p) &= \mathrm{Re}\left\{i\alpha U'(\xi) + \frac{(i\alpha)^2}{2!}U''(\xi) + \cdots\right\} \\ &= \mathrm{Re}\left\{c_0(d_1(i\alpha)\xi^p + d_2(i\alpha)^2\xi^{p-1} + \cdots + d_p(i\alpha)^p\xi + (i\alpha)^{p+1}\right. \\ &\qquad \left. + O\left(\sum_{j+k=p+2} |\alpha|^j\xi^k\right)\right\}, \end{aligned} \tag{3.2}$$

where

$$d_k = \binom{p+1}{k}, \qquad k = 1, 2, \cdots, p.$$

Now we have from Hölder's inequality the estimates

$$|\alpha|^{j+1}|\xi|^{p-j} \leqq \eta_j|\alpha|^{\tau_{r,j}} + \epsilon\xi^{2s},$$

where for $r \geqq 1$, $\tau_{r,j} = 2s(j + 1)/(r + j)$, $(r = 2s - p)$, $j = 0, 1, \cdots, (p - 1)$.

Thus, we can find a $\tau_r \geqq 1$ such that

$$H(\alpha, \xi, p) \leqq \nu_1|\alpha|^{\tau_r}(1 + o(1)) + \nu_2\xi^{2s}(1 + O(\alpha) + O(\xi))$$

when $\alpha, \xi \to 0$ and where $\nu_2 = O(\epsilon)$, $\epsilon \to 0$.

It is easy to see that we have to take $\tau_r = \min_j \tau_{r,j}$ over the terms in (3.2) which are positive. Thus for odd p we conclude that

$$\tau_r^{(1)} = \tau_{r,1} = \tau_{1,1},$$

since $r = 1$. (Theorem 2.1.) From this it follows that $\tau = (p + 1, p + 1)$ for odd p.

When $p = 2l$, $l = 1, 2, \cdots$, we find that

$$\tau_r^{(0)} = \tau_{r,0} \tag{3.3}$$

when sign (ic_0) sign $(\alpha) = 1$, and

$$\tau_r^{(2)} = \tau_{r,2} \tag{3.3'}$$

when sign (ic_0) sign $(\alpha) = -1$.

Note that Im $(ic_0) = 0$. Summarizing, we have

$$|\hat{Q}(\xi + i\alpha, 0)| = \exp(-\alpha\rho\lambda - (\delta' - \nu_2)\xi^{2s} + \nu_1|\alpha|^{\tau_r^{(m)}} + \cdots), \quad m = 0, 1, 2. \tag{3.4}$$

Moreover, from assumption (iv) we know that

$$\text{sign } (ic_0) = \text{sign } \rho, \tag{3.5}$$

and together with (3.3) this gives that (3.3) holds when

$$\text{sign } (\alpha) = \text{sign } (\rho), \quad \text{i.e.} \quad \alpha\rho > 0.$$

Analogously, (3.3') holds for $\alpha\rho < 0$, i.e., for even p, τ is given by

$$\tau = (\tau_r^{(2)}, \tau_r^{(0)}) = (6s/(r + 2), 2s/r).$$

Now it is easy to see that we can choose ϵ so small that (3.4) guarantees contractivity of order τ. Indeed, there exists a ξ_0, such that for all ξ, $|\xi| < \xi_0$,

$$|\hat{Q}(\xi + i\alpha, 0)| = \exp(-\alpha\rho\lambda + \sigma|\alpha|^{\tau_r^{(m)}})R(\xi), \quad m = 0, 1, 2, \tag{3.6}$$

where $|R(\xi)| \leqq 1$ for $|\xi| < \xi_0$. But for $|\xi| \geqq \xi_0$ there exists an $\alpha_0 > 0$ such that for all $|\alpha| < \alpha_0$

$$|\hat{Q}(\xi + i\alpha, 0)| \leqq \exp(-\alpha\rho\lambda + \sigma|\alpha|^{\tau_r^{(m)}} - \delta'/2 \cdot \xi^{2s}), \quad m = 0, 1, 2.$$

Thus, Theorem 2.2 is proved for constant coefficients.

From the assumptions of Theorem 2.2 we have local contractivity of order τ for every x and t, i.e.,

$$\hat{Q}(x, t, \xi + i\alpha, 0) = \exp(-\alpha\lambda\rho(x, t) + \sigma(x, t)|\alpha|^{\tau}) \times R(x, t, \xi), \quad |\xi| \leqq \pi.$$

What remains to show is that $\sigma(x, t)$ is uniformly bounded. From Definition 2.2, we see that

$$\hat{Q}(x, t, \xi, 0) = \exp(i\rho(x, t)\lambda\xi + c_0(x, t)\xi^{p+1} + \cdots).$$

This relation and assumptions (ii) and (iii) of Theorem 2.2 then show that $c_0(x, t)$ is uniformly bounded in the upper halfplane. Now $\nu_2 = \nu_2(\epsilon)$ in (3.4) depends

essentially on $c_0(x, t)$ and thus $\nu_2 = \epsilon\nu_3(x, t)$, where $\nu_3(x, t)$ is uniformly bounded for $t \geqq 0$.

Therefore it is sufficient to choose $\epsilon(x, t) \geqq \epsilon_0 > 0$ to obtain the relation (3.6) and from Hölder's inequality, we then see that $\nu_1(x, t, \epsilon)$ is uniformly bounded. Hence we can conclude that $\sigma(x, t)$ is uniformly bounded for $t \geqq 0$, and the proof of Theorem 2.2 is completed.

Finally we give the proof of Theorem 2.3. (Cf. Strang [10] and [1].)

Proof of Theorem 2.3. Put $e^{i\rho(x,t)\lambda\xi} = \hat{Q}_2(x, t, \xi, 0) + R(\xi)$, and take $y = \rho\lambda$; then $\hat{Q}_2(\xi, y)$ is the Lagrangian interpolation polynomial of degree $2N$ through the points $\{(\nu, e^{i\xi\nu})\}_{\nu=-N}^{N}$, and hence the remainder $R(\xi, y)$ is given by

$$R(\xi, y) = \xi^{p+1}i^{p+1}e^{i\xi\eta}\cdot y\prod_{j=1}^{N}(y^2 - j^2)/(2N + 1)!\,, \qquad |\eta| < N\,.$$

Thus

$$c_0(x, t) = -i\rho(x, t)\lambda\prod_{j=1}^{N}(j^2 - \rho^2\lambda^2)/(2N + 1)!$$

and since $|\rho(x, t)\lambda| \leqq 1$ we see that condition (iv) of Theorem 2.2 holds.

4. Proofs of Analytic Theorems. *Proof of Theorem* 2.4 (*constant coefficients*). The generalized solution of (2.1) with $\phi = 0$ and with initial function $f(x)$ defined in (2.6) is

$$u(x, t) = f(x + \rho t) = 0 \qquad x + \rho t < 0\,, \tag{4.1}$$
$$= 1 \qquad x + \rho t \geqq 0\,.$$

Now make the transformation

$$w_\nu(t) = e^{\alpha(x_\nu+\rho t)/h}v_\nu(t)\,.$$

If $v_\nu(t)$ satisfies (2.3) then $w_\nu(t)$ satisfies

$$Q_1'w_\nu(t + k) = Q_2'w_\nu(t)\,, \tag{4.2}$$

where

$$Q_1'w_\nu = e^{-\alpha\rho\lambda}e^{\alpha\nu}Q_1e^{-\alpha\nu}w_\nu$$

and

$$Q_2'w_\nu = e^{\alpha\nu}Q_2e^{-\alpha\nu}w_\nu\,.$$

Since $|\hat{Q}_1(\xi, 0)| \geqq \text{const} > 0$ it follows that for α sufficiently small

$$|\hat{Q}_1'(\xi, 0)| = e^{-\alpha\rho\lambda}|\hat{Q}_1(\xi + i\alpha, 0)| \geqq \text{const} > 0\,.$$

Thus

$$\hat{Q}'_{-1}(\xi + i\alpha, 0) = e^{\alpha\rho\lambda}\hat{Q}_1(\xi + i\alpha, 0)^{-1}Q_2(\xi + i\alpha, 0)$$

is well defined.

From assumption (ii) it now follows that

$$\hat{Q}'_{-1}(\xi + i\alpha, 0) = \exp(\sigma|\alpha|^{r\mp}) \times R(\xi)\,,$$

where $|R(\xi)| \leqq 1$ for $|\xi| \leqq \pi$. Then it is an easy matter to see that

$$|Q'_{-1}(\xi + i\alpha, h)| \leqq 1 + \nu_0 h \quad \text{for} \, 0 \leqq h \leqq h_1 , \tag{4.3}$$

where ν_0 is independent of h, if α is chosen such that

$$\alpha = \alpha_{\mp} = \text{sign } (\rho t + x_\nu) h^{1/\tau_{\mp}} .$$

It is well known (cf. Richtmyer [12]) that (4.3) implies stability of the scheme (4.2), i.e.,

$$\|w_\nu(t)\|_h \leqq e^{\beta t} \|w_\nu(0)\|_h = e^{\beta t} \|e^{\alpha(x_\nu + \rho t)/h} f(x_\nu)\|_h ,$$

for some constant β.

If $\alpha < 0$, then

$$\begin{aligned} \|w_\nu(0)\|_h &\leqq \max_\nu f(x_\nu) \left(\sum_{\nu=0}^{\infty} h e^{2\alpha\nu} \right)^{1/2} = h^{1/2}/(1 - e^{2\alpha})^{1/2} \\ &= O(h^{1/2} |\alpha|^{-1/2}) . \end{aligned} \tag{4.4}$$

Since $\|w_\nu(t)\|_h \geqq h^{1/2} |w_\nu(t)|$, we get

$$|v_\nu(t)| \leqq \nu_1 e^{\beta t} h^{(q-1)/2} \exp (-h^{-q} |x_\nu + \rho t|) ,$$

where $q = q_{\mp} = (\tau_{\mp} - 1)/\tau_{\mp}$, and ν_1 is a constant.

Note that this estimate only holds to the left of the characteristic $\rho t + x = 0$.

Let

$$\begin{aligned} f_1(x) = f(-x) &= 1 , \qquad x \leqq 0 , \\ &= 0 , \qquad x > 0 . \end{aligned}$$

With $f_1(x)$ as initial function we get a corresponding estimate to the right of the characteristic $\rho t + x = 0$, when $\alpha > 0$.

Now $f_2(x) = f_1(x) + f_0(x) \equiv 1$, $-\infty < x < \infty$, $(f_0 = f)$ as initial function gives rise to solutions of (2.1) and (2.3) that are identically equal to one. This follows from assumption (i). An elementary application of the triangle inequality then gives

$$|v_\nu(t) - 1| \leqq \nu_2 e^{\beta t} h^{(q-1)/2} \exp (-h^{-q} |x_\nu + \rho t|) ,$$

where ν_2 is a constant.

This estimate is used to the right of the characteristic $\rho t + x = 0$. Thus the proof of Theorem 2.4 is completed.

Before we are able to prove Theorem 2.5 we are going to state the following stability theorem due to Lax-Nirenberg [7].

THEOREM 4.1. *Consider the difference scheme*

$$y_\nu(t + k) = S y_\nu(t) ,$$

where S is a difference operator of the form

$$S = \sum d_j(x_\nu, t, 0) E^j .$$

Suppose that for S we have

$$|\hat{S}(x, t, \xi, 0)| \leqq 1$$

for all t, x *and* ξ.

Furthermore assume that condition (iv) *of Theorem* 2.5 *holds. Then the difference scheme is stable in the sense that*

$$\|y_\nu(t)\|_h \leqq e^{\beta t}\|y_\nu(0)\|_h ,$$

where β *is a constant.*

Proof of Theorem 2.5 (*variable coefficients*). In this case we have the scheme (2.3′) and we make the transformation

$$w_\nu(t) = e^{\alpha g(x_\nu, t)/h} v_\nu(t) ,$$

where $g(x, t)$ was defined in Section 2.

The new mesh-function $w_\nu(t)$ satisfies

$$w_\nu(t + k) = Q_0' w_\nu(t) = e^{\alpha g(x_\nu, t+k)/h} Q_0 e^{-\alpha g(x_\nu, t)/h} w_\nu(t) . \tag{4.5}$$

From Definition 2.5 it follows that $\partial g/\partial t = \rho \partial g/\partial x$, and moreover, since $\rho(x, t) \in c_0^\infty$, that $\partial^2 g/\partial t^2$ and $\partial^2 g/\partial x^2$ is uniformly bounded for $t \geqq 0$.

Thus it is easy to conclude that

$$\hat{Q}_0'(x, t, \xi, \alpha, h) = e^{\alpha \lambda \rho(x,t) g_x(x,t)} \hat{Q}_0(x, t, \xi + i\alpha g_x(x, t), h) + |\alpha| h \hat{Q}(x, t, \xi) , \qquad x = x_\nu ,$$

where $\hat{Q}$ is the symbol of a uniformly bounded operator Q.

Using assumption (ii) we have for α sufficiently small

$$\hat{Q}_0'(x, t, \xi, \alpha, 0) = \exp\left(\sigma(x, t)|\alpha g_x(x, t)|^{\tau \mp}\right) \times R(x, t, \xi, \alpha) ,$$

where $|R| \leqq 1$ for $|\xi| \leqq \pi$.

Now

$$R(x, t, \xi, \alpha) = R(x, t, \xi, 0) + |\alpha|^{\tau \mp} \hat{Q}_3(x, t, \xi) ,$$

and hence, since $R(x, t, \xi, 0) = \hat{Q}_0(x, t, \xi, 0)$,

$$\hat{Q}_0'(x, t, \xi, \alpha, h) = \hat{Q}_0(x, t, \xi, 0) + (|\alpha|^{\tau \mp} + |\alpha| h + h)\hat{Q}_4(x, t, \xi, h) .$$

Here, $\hat{Q}_j$, $j = 3, 4$ correspond to uniformly bounded operators Q_j, $j = 3, 4$.

Now the difference scheme $y_\nu(t + k) = Q_0 y_\nu(t)$ satisfies the conditions of Theorem 4.1, and therefore it is stable. If we choose

$$\alpha = \operatorname{sign}(g(x, t)) h^{1/\tau \mp}$$

then the scheme (4.5) is also stable, since $Q_0' = Q_0 + hQ_5$, where Q_5 is a uniformly bounded operator. Summarizing, we can find a constant β_1 such that

$$\|w_\nu(t)\|_h \leqq e^{\beta_1 t}\|w_\nu(0)\|_h = e^{\beta_1 t}\|e^{\alpha g(\cdot, 0)/h} f(\cdot)\|_h .$$

But now $g(x, 0) = x$, and therefore estimates corresponding to (4.4) hold even in the variable coefficient case. Since $u(x, t) = 0$ for $g(x, t) < 0$, the estimate (2.8) in T_{-1} follows. Now call the solutions of (2.1) and (2.3′) with initial function $f_2(x) \equiv 1$ $u^{(2)}(x, t)$ and $v_\nu^{(2)}(t)$ respectively. Then from assumption (i) we have

TABLE. *Eleven numerical experiments for testing* τ.

Experiment	ρ	Q_2 $(Q_1 \equiv 1)$	p	$2s$	σ	ϵ	$n = 1/k$	D_h^{-1}	D_h^{+1}	D_h	E_h^{-1}	E_h^{+1}	E_h	*Estimate for* τ	*Theoretical* τ
1	0	$I + \sigma h^2 D_+ D_-$	1	2	0.5	0.01	400			96					
							500			107					
							800			131			1.36	(2.2,2.2)	(2,2)
							1000			147			1.37	(2.2,2.2)	(2,2)
			1	2	0.5	0.001	400			120					
							500			136					
							800			176			1.47	(1.8,1.8)	(2,2)
							1000			194			1.44	(1.9,1.9)	(2,2)
2	0	$I + 0.1h^3 D_0 D_+ D_- + \sigma h^4 D_+^2 D_-^2$	2	4	−0.025	0.001	300	14	60						
							330	14	62						
							600	18	85		1.29	1.42		(2.7,2.0)	(3,2)
							660	19	92		1.36	1.48		(2.3,1.8)	(3,2)
3	0	$I + \sigma h^4 D_+^2 D_-^2$	3	4	−0.1	0.01	350			27					
							400			27					
							700			31			1.15	(5.0,5.0)	(4,4)
							800			33			1.22	(3.5,3.5)	(4,4)
4	1	$I + kD_0 + \sigma \frac{k^2}{2} D_+ D_-$	1	2	1.2	0.01	350			25					
							400			27					
							700			34			1.36	(2.3,2.3)	(2,2)
							800			36			1.33	(2.4,2.4)	(2,2)
		with $\lambda = 1/2$	1	2	1.6	0.01	450			49					
							500			52					
							900			68			1.39	(2.1,2.1)	(2,2)
							1000			71			1.37	(2.2,2.2)	(2,2)
5	1	$I + kD_0 + \sigma \frac{k^3}{2} D_+ D_-$	2	4	1	0.01	75	4	17						
							100	5	20						
		with $\lambda = 1/2$					150	5	23		1.25	1.36		(3.1,2.2)	(3,2)
							200	6	28		1.20	1.40		(3.8,2.1)	(3,2)
6	1	$I + kD_0 + \frac{k^2}{2} D_+ D_- + (k^3 - kh^2)/6 D_0 D_+ D_-$	3	4	0.75	0.001	250			17					
							300			17					
							500			19			1.12	(6.1,6.1)	(4,4)
		$+ \sigma(k^4 - k^2h^2)/24 D_+^2 D_-^2$ with $\lambda = 0.9$					600			20			1.18	(4.2,4.2)	(4,4)
					1.2		250			13					
							300			14					
							500			16			1.23	(3.3,3.3)	(4,4)
							600			16			1.14	(5.3,5.3)	(4,4)

7	1	the same as in exp. 6	4	6	1	0.001	280 300 560 600	7 7 8 8	16 17 21 22		1.14 1.14	1.31 1.29		(5.3,2.6) (5.3,2.7)	(4.5,3) (4.5,3)
		Implicit methods $Q = Q_1^{-1}Q_2$													
8	1	$(I - k/2D_0)^{-1}(9E + 32I - E^{-1})/40$ with $\lambda = 1/2$	1	2	–	0.01	100 150 200 300			20 24 28 34			1.40 1.42	(2.1,2.1) (2.0,2.0)	(2,2) (2,2)
9	1	$(I - k/2D_0)^{-1}\left(I + \frac{k}{2} D_0\right)$ with $\lambda = 1/2$	2	0	–	0.01	50 100 150 200	5 6 7 8	54 95 155 206	59 104 162 214				(1,1) (1,1)	
10	1	$(I - k/2D_0)^{-1}\left(I + \frac{k}{2} D_0 + \sigma h^4 D_+^2 D_-^2\right)$ with $\lambda = 1/2$	2	4	−0.01	0.01	170 180 340 360	8 8 10 10	48 49 69 70		1.25 1.25	1.44 1.45		(3.1,1.9) (3.1,1.9)	(3,2) (3,2)
11	1	$(I - k/2D_0)^{-1}\left(I + \frac{k}{2} D_0 + \sigma h^6 D_+^3 D_-^3\right)$ with $\lambda = 1/2$	2	6	0.01	0.01	170 180 340 360	8 8 10 10	46 49 73 77		1.25 1.25	1.59 1.57		(3.1,1.5) (3.1,1.5)	(3,1.5) (3,1.5)

$$|u^{(2)}(x_\nu, k) - v_\nu^{(2)}(k)| = |u^{(2)}(x_\nu, k) - Q_0 u^{(2)}(x_\nu, 0)| = (h^{p+1}) .$$

A standard calculation then shows that

$$|u^{(2)}(x_\nu, t) - v_\nu^{(2)}(t)| = O(h^p) ,$$

and hence the estimate (2.8) in T_{+1} follows as in the proof of Theorem 2.4.

5. Numerical Experiments. In this section our theoretical results of Sections 2–4 will be tested by a series of numerical experiments performed on the CDC 3600 computer at the University of Uppsala.

In the main theorem of Section 2, we have estimates of the form

$$|u(x_\nu, t) - v_\nu(t)| \leqq c(t) \exp(-h^{-q}|\rho t + x_\nu|) . \tag{5.1}$$

We are now going to study the interval $I_\epsilon(t)$ outside which this error is smaller than ϵ, $\epsilon > 0$. Therefore, we let our computer program count the number of points in this interval. We call the number to the left of the characteristic line, $\rho t + x = 0$, D_h^{-1} and the number to the right D_h^{+1}, and the total number of points within $I_\epsilon(t)$, $D_h = D_h^{-1} + D_h^{+1}$. The width of $I_\epsilon(t) = m(I_\epsilon(t))$, is therefore approximately $h \cdot D_h$. We can also find $m(I_\epsilon(t))$ by solving

$$c(t) \exp(-h^{-q}|\rho t + x|) = \epsilon . \tag{5.2}$$

This last equation has two roots x_1, x_2, $x_1 < x_2$, and hence $m(I_\epsilon(t)) = x_2 - x_1$.

If we also introduce the double step-size $2h$ and count D_{2h}^{-1}, D_{2h}^{+1} and D_{2h} respectively, we can define the following quantities

$$E_h^{\,i} = D_h^{\,i}/D_{2h}^{i} \doteqdot 2^{-q+1} = \sqrt[\tau]{2} , \qquad i = -1, +1$$

$$E_h = D_h/D_{2h} \doteqdot \sqrt[\tau]{2} , \qquad \tau = \tau_\mp \text{ defined in Section 2} .$$

Therefore we could use these quantities to control our theoretical value of $\tau_\mp$.

In the table on the following pages we describe 11 numerical experiments. In column 8 we have the number of iterations, called n, which we can relate to h and k in the following way. Set $nk = 1$ and $h = k/\lambda$. We have thus normalised our results to the time $t = 1$. $\phi \equiv 0$ in all of the experiments. Observe that for odd ρ, we have $\tau_- = \tau_+$, and therefore we are only considering E_h for these experiments.

We are now going to comment on the results of the experiments. A general remark must be that $E_h^{\,i}$, $i = -1, +1$ and E_h are rational step functions of h, and when $D_h^{\,i}$, $i = -1, +1$, and D_h respectively are rather small natural numbers, the risk of obtaining crude estimates of $\tau_\mp$ increases seriously.

In Experiment 5 we have the Lax-Wendroff scheme mentioned in Section 2, and hence $\tau = (3, 2)$.

The difference operator in Experiment 7 is the most accurate one, when $N = 2$ in Theorem 2.3. Hence $\tau = (4.5, 3)$. The more precise results of Hedström [3] gives for this special case $\tau = (5, 3)$, which agrees better with the experiment.

The Crank-Nicolson scheme of Experiment 9 was introduced to see that some sort of dissipation is necessary to get a meaningful result. (Cf. Section 1.)

The results of Experiment 10 verify the theory closely, and finally in the last experiment we see the effect of increasing r from 2 to 4, i.e., to $2s = 6$.

Department of Computer Sciences
Uppsala University
Uppsala, Sweden

1. M. APELKRANS, *On Difference Schemes for Hyperbolic Equations with Discontinuous Initial Values*, Report No. 5, Department of Computer Sciences, Uppsala, 1967.

2. R. P. FEDORENKO, "The application of high-accuracy difference schemes to the numerical solution of hyperbolic equations," *Ž. Vyčisl. Mat. i Mat. Fiz.*, v. 2, 1962, pp. 1122–1128. (Russian) MR **26** #5739.

3. G. W. HEDSTRÖM, "The rate of convergence of some difference schemes." (To appear.)

4. H.-O. KREISS, "On difference approximations of the dissipative type for hyperbolic differential equations," *Comm. Pure Appl. Math.*, v. 17, 1964, pp. 335–353. MR **29** #4210.

5. H.-O. KREISS & E. LUNDQVIST, "On difference approximations with wrong boundary values," *Math. Comp.*, v. 22, 1968, pp. 1–12.

6. P. D. LAX, "On the stability of difference approximations to solutions of hyperbolic equations with variable coefficients," *Comm. Pure Appl. Math.*, v. 14, 1961, pp. 497–520. MR **26** #3215.

7. P. D. LAX & L. NIRENBERG, "On stability for difference schemes: A sharp form of Gårding's Inequality," *Comm. Pure Appl. Math.*, v. 19, 1966, pp. 473–492. MR **34** #6352.

8. B. PARLETT, "Accuracy and dissipation in difference schemes," *Comm. Pure Appl. Math.*, v. 19, 1966, pp. 111–123. MR **33** #5141.

9. J. PEETRÉ & V. THOMÉE, "On the rate of convergence for discrete initial-value problems." (To appear.)

10. G. STRANG, "Trigonometric polynomials and difference methods of maximum accuracy," *J. Math. and Phys.*, v. 41, 1962, pp. 147–154.

11. V. THOMÉE, "On maximum-norm stable difference operators" in *Numerical Solution of Partial Differential Equations*, edited by, J. H. Bramble, Academic Press, New York, 1966.

12. R. D. RICHTMYER, *Difference Methods for Initial-Value Problems*, Interscience Tracts in Pure and Applied Mathematics, Interscience, New York, 1957. MR **20** #438.