Symmetric Hyperbolic Difference Schemes and Matrix Problems

E. Turkel*

Courant Institute of Mathematical Sciences
New York University
New York, New York 10012

Submitted by P. D. Lax

ABSTRACT

Sufficient conditions for the stability of multidimensional finite difference schemes are difficult to obtain. It is shown that for special families of amplification matrices $G(A,B)$ a sufficient condition for power boundedness can be obtained by replacing the matrices by appropriate scalars, and so the problem is reduced to a scalar one. As one application it is shown that the Lax-Wendroff scheme in two dimensions is stable if $|Au|^{2/3} + |Bu|^{2/3} < 1$ for all real unit vectors $u$. The Lax-Wendroff scheme with stabilizer does not always permit such large time steps. It is conjectured that the analysis for all symmetric hyperbolic schemes can be reduced to the scalar case.

1. INTRODUCTION

It is well known that many of the equations of mathematical physics can be expressed as symmetric hyperbolic systems of partial differential equations. These include Maxwell's equations, linear dynamic elasticity, the wave equation, inviscid dynamic fluid dynamics, and magnetohydrodynamics. It is thus of importance to investigate finite differences for such equations and to utilize the additional properties of these systems to assist the analysis of the approximating equations. A particular case has been analyzed by Friedrichs [4] where the coefficients of the finite difference scheme are all positive. Kreiss [6] has studied difference equations with symmetric matrices with the additional requirement that the schemes be dissipative. In this paper we shall study approximations that are in some sense the most appropriate for symmetric hyperbolic systems.

*This work was partially supported by a National Science Foundation Energy Related Fellowship and by ERDA, contract no. E (11-1)-3077.
Given a system

\[ u_t + \sum_{i=1}^{d} A_i \frac{\partial u}{\partial x_i} = 0, \]  

(1)

where \( A_i \) are real symmetric constant matrices. We consider finite difference approximations of the form

\[ u^{n+1} = P(A_i, T)u^n, \]

(2)

where \( P \) is a polynomial in the matrices \( A_i \) and in the shift operator \( T \).

Taking the Fourier transform of this equation, we have

\[ \hat{u}^{n+1} = G(A_i; \xi_i)\hat{u}^n. \]

(3)

Here \( G \) is the amplification matrix and depends on the Fourier variables \( \xi_i \) as well as the matrices \( A_i \). In general, the matrix \( G \) and be expressed as

\[ G = R + iJ, \]

(4)

where \( R \) and \( J \) are polynomials with real coefficients in the real symmetric matrices \( A_i \).

We shall call a scheme symmetric hyperbolic if the amplification of the scheme has the form (4) where the \( A_i \) are real symmetric and \( R \) and \( J \) are also real symmetric matrices. Schemes with the property preserve the "nice" properties of the analytic equations.

In general, the study of the stability properties of schemes is a difficult manner. The Kreiss matrix theorem (see Richtmyer and Morton [10]) yields several necessary and sufficient conditions, none of which can really be used for practical purposes. For specific equations the Buchanan criterion (see [10]) can be used, but again is of limited use. The numerical radius gives a sufficient condition for stability that can be used for families of partial differential equations. However, up to date the application of this technique has been limited (see [8] and [9]). However, there is one case for which the stability theory is relatively simple. This is the case when the matrices \( A_i \) all commute. Then they share a common complete set of eigenvectors, and the spectral mapping theorem is applicable. In this simple case the amplification matrix \( G \) is normal, and its power boundedness is reduced to the scalar case. Hence, the von Neumann criterion is both necessary and sufficient. Even when analytic conditions cannot be determined for the amplification factor to be within the unit circle, the question can always be decided computationally. Since we are only dealing with a scalar polynomial depending on the eigenvalues of \( A_i \), a simple computer scan will yield necessary and sufficient stability conditions for all commuting matrices \( A_i \).
This method has already been used by Eilon, Gottlieb and Zwas [3], as well as Turkel, Abarbanel and Gottlieb [12], to at least indicate what the situation is in general. Lax and Wendroff [8] also note that for their scheme the most severe restrictions arise when \( A = B \). The major objection to analyzing the case of commuting \( A_i \) is that this hypothesis is rarely met in physical situations. Thus, there seems to be little advantage in studying systems for which the analysis is relatively easy but has little relevance to practical cases. Nevertheless, in the papers quoted above the most severe cases have always been when the matrices \( A_i \) commute. Based on this, we propose the following conjecture.

**Conjecture**  Let \( G \) be the amplification matrix for a symmetric hyperbolic finite difference scheme as defined above. Assume that \( G \) is power bounded when the matrices \( A_i \) commute and their eigenvalues all lie within some set \( S \subset \mathbb{R}^d \). Then \( G \) is power bounded for all real symmetric matrices \( A_i \) whose eigenvalues lie within the set \( S \).

Before proceeding, we note that this conjecture cannot be strengthened. The example of Yamaguti [13] shows that a scheme which is stable for real symmetric matrices \( A_i \) need not be stable for arbitrary hyperbolic systems. It is also trivial to see that if the matrices \( A_i \) are symmetric but the scheme is not symmetric hyperbolic in the above sense, then the commuting case need not imply anything above the general case. For example, let \( G \) be any power bounded amplification matrix for which \( G(\pi, \pi) = I \) (e.g., rotated Richtmyer; see [14]). Let \( G_1(\xi, \eta) = G(\xi, \eta) + \alpha(\xi, \eta)(AB - BA) \). When \( A \) and \( B \) commute, \( G = G_1 \), and so \( G_1 \) is power bounded. However, for noncommuting \( A, B, \rho(G_1) > 1 \), and so the scheme is not stable even though \( A \) and \( B \) are real symmetric. We also note that the conjecture is a property of difference schemes but not of matrices. Let

\[
G = 4(ABA + BAB - AB - BA).
\]

When \( A, B \) are real symmetric, so is \( G \). Furthermore, if \( A \) and \( B \) commute and have eigenvalues of zero or one, then the spectral mapping theorem shows that \( G \) is the zero matrix and so is power bounded. Now consider

\[
A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.
\]

Here \( A \) and \( B \) are real symmetric with eigenvalues of zero or one. Nevertheless, \( G = \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} \), \( \rho(G) = \sqrt{2} \), and so \( G \) is not power bounded.

From the theorems that follow, it may be possible that one also must require some type of geometrical symmetry of the scheme in addition to the"
algebraic symmetric of the matrices involved. We also note that the conjecture yields only a sufficient condition for stability. An example given later shows that, in general, the permissible time steps for schemes with noncommuting matrices are strictly larger than those allowed in the commuting case.

We define the spectral radius of a matrix $A$ as

$$\rho(A) = \max_i |\lambda_i(A)|,$$

where $\lambda_i$ are the eigenvalues of $A$. The numerical radius of $A$ is defined as

$$r(A) = \sup |\langle Au, u \rangle|,$$

where $u$ is a complex unit vector.

2. SECOND ORDER SYSTEMS

As indicated in the introduction, we conjecture that among symmetric hyperbolic schemes the commuting case allows the smallest time steps consistent with stability. In this section we shall verify this conjecture for special cases of finite difference schemes which include many of the second order methods. We first prove a theorem about complex symmetric matrices (see Gantmacher [5]), i.e., complex matrices $C$ such that $C^t = C$. Such matrices can always be expressed as $C = R + iJ$, where $R, J$ are real symmetric matrices. So the theory of complex symmetric matrices is closely related to the properties of symmetric hyperbolic schemes. A vector is said to be real if all its components are real. We then have

**Theorem 1.** Let $C$ be a complex symmetric matrix. Then the numerical range of $C$ is given by

$$r(C) = \sup |\langle Cx, x \rangle|,$$

where the supremum is over all real unit vectors.

**Proof.** We shall show that, given a complex unit vector $x$, we can construct a real unit vector $y$, such that $|\langle Cy, y \rangle| > |\langle Cx, x \rangle|$. The result then follows.
For any vector \( x \) we let \( x = x_1 + ix_2 \) where \( x_1, x_2 \) are real. Then

\[
(Cx, x) = (Rx, x) + i(Jx, x)
\]

\[
= (Rx_1, x_1) + (Rx_1, ix_2) + i(Rx_2, x_1) + i(Rx_2, ix_2)
+ i(Jx_1, x_1) + i(Jx_1, ix_2) + i(Jx_2, x_1) + i(Jx_2, ix_2)
\]

\[
= (Rx_1, x_1) - i(Rx_1, x_2) + i(Rx_2, x_1) + (Rx_2, x_2)
+ i(Jx_1, x_1) + (Jx_1, x_2) - (Jx_2, x_1) + i(Jx_2, x_2).
\]

but \((Rx_1, x_2) = (x_1, Rx_2) = (Rx_2, x_1)\), since \( R \) is symmetric and all quantities are real. Similarly \((Jx_1, x_2) = (Jx_2, x_1)\), so

\[
(Cx, x) = (Rx_1, x_1) + (Rx_2, x_2) + i(Jx_1, x_1) + i(Jx_2, x_2) = (Cx_1, x_1) + (Cx_2, x_2) \tag{5}
\]

If \( x_2 \) is zero, then \( x \) is real, and we are finished. If \( x_1 \) is zero, choose \( y = x_2 = ix \). Then \( y \) is a unit real vector and \( |(Cy, y)| = |(Cx, x)| \). Thus, we can assume that both \( x_1, x_2 \) are nonzero.

We now define

\[
ri = \frac{(Rx_i, x_i)}{(x_i, x_i)}, \quad hi = \frac{(Jx_i, x_i)}{(x_i, x_i)}. \tag{6}
\]

\( r_i, h_i \) are real, since \( R \) and \( J \) are symmetric. Also define

\[
y = \begin{cases} 
\frac{x_1}{\|x_1\|} & \text{if } r_1^2 + h_1^2 \geq r_2^2 + h_2^2, \\
\frac{x_2}{\|x_2\|} & \text{if } r_2^2 + h_2^2 > r_1^2 + h_1^2. 
\end{cases} \tag{7}
\]

Then \( y \) is a real unit vector and

\[
(Cy, y) = (Ry, y) + i(Jy, y), \quad |(Cy, y)|^2 = (Ry, y)^2 + (Jy, y)^2
\]

\[
= \begin{cases} 
\frac{(Rx_1, x_1)}{|x_1|^2} & r_1^2 + h_1^2 \geq r_2^2 + h_2^2, \\
\frac{(Rx_2, x_2)}{|x_2|^2} & r_2^2 + h_2^2 < r_1^2 + h_1^2. 
\end{cases}
\]
We thus wish to show that for any complex unit vector $x$,

$$|(Cx, x)|^2 < \max(r_1^2 + j_1^2, r_2^2 + j_2^2).$$

From Eq. (5) together with (6) we have

$$|(Cx, x)|^2 = \left[ r_1 \|x_1\|^2 + r_2 \|x_2\|^2 + i \left[ f_1 \|x_1\|^2 + f_2 \|x_2\|^2 \right] \right]^2 + \left[ f_1 \|x_1\|^2 + f_2 \|x_2\|^2 \right]^2$$

$$= (r_1^2 + j_1^2) \|x_1\|^4 + (r_2^2 + j_2^2) \|x_2\|^4 + 2(r_1 r_2 + j_1 j_2) \|x_1\|^2 \|x_2\|^2.$$

Let $u = \|x_1\|^2$. Since $\|x\|^2 = \|x_1\|^2 + \|x_2\|^2 = 1$, we have

$$|(Cx, x)|^2 = f(u) = (r_1^2 + j_1^2)u^2 + (r_2^2 + j_2^2)(1 - u)^2 + 2(r_1 r_2 + j_1 j_2) u(1 - u).$$

We can consider $f(u)$ as a quadratic function in $u$ for all finite $u$:

$$f(u) = au^2 + bu + c,$$

where

$$a = r_1^2 + j_1^2 + r_2^2 + j_2^2 - 2(r_1 r_2 + j_1 j_2) > 0.$$ 

Hence $f$ is concave upwards and

$$\max_{0 < u < 1} f(u) < \max(f(0), f(1)) = \max(r_1^2 + j_1^2, r_2^2 + j_2^2),$$

or

$$|(Cx, x)|^2 < \max(r_1^2 + j_1^2, r_2^2 + j_2^2) = |(Cy, y)|^2.$$

Using this result, we can prove the following theorems, which are a partial verification of our conjecture.
**Theorem 2.** Let

\[ G = I + P_1(A) + P_2(B) + P_3(A,B) + i[P_4(A) + P_5(B)], \]

with \( P_3(A,B) = \sum_{k,l>0} y_{kl}(A^k B^l + B^l A^k) \). We make the following assumptions:

1. The coefficients of the polynomials \( P_i \) are all real.
2. When \( A \) and \( B \) commute, \( G \) is power bounded if the eigenvalues of \( A \) and \( B \) satisfy \((a_i, b_i) \in S\) for some set \( S\).
3. \( A \) and \( B \) are 2x2 matrices.
4. If \( P_3(A,B) \) is replaced by \(- P_3(A,B) \) (possibly accompanied by changes in \( P_4 \) or \( P_5 \)) the power-boundedness of \( G \) (for the commuting case) is not affected.

**Note:** In practice, the coefficients of \( G \) are functions of the Fourier variables \( \xi \) and \( \eta \). Changing \( \xi \) to \(-\xi \) frequently changes the sign of \( P_3 \) and stability still holds. Thus, assumption (4) is quite natural.

**Proof.** A sufficient condition for stability is that the numerical range of \( G \) be bounded by 1 (see [8, 10]). By Theorem 1, it is sufficient to consider \(|(Gx, x)|\) for real vectors \( x \). Since \( x \) is real and two dimensional, we can express it as

\[ x = \begin{pmatrix} \cos \psi \\ \sin \psi \end{pmatrix}. \]

Furthermore, since \( A \) is real and symmetric, we can diagonalize it by a real orthogonal matrix \( T \). Letting \( G = TGT^{-1} \), we see that none of the hypotheses are disturbed. Hence we can assume, without loss of generality, that \( A \) is diagonal. In addition, we introduce the matrices \( A_1 = A - qI \), \( B_1 = B - pI \), \( q, p \) real. \( G \) can be considered as a polynomial in \( A_1, B_1 \), and it will again have the same form as Eq. (1). The hypotheses of the theorem will still hold with the stability set \( S \) replaced by \( S_1 \), where \((a_1,b_1) \in S_1 \) iff \((a - q, b - p) \in S\). This follows from the spectral mapping theorem. Thus, again we can choose \( q \) and \( p \) arbitrarily. By choosing \( q \) and \( p \) to be eigenvalues of \( A \) and \( B \), we can assume, without loss of generality, that \( A \) and \( B \) both have at most one nonzero eigenvalue. If both eigenvalues are zero, then the matrix is the zero matrix (since \( A, B \) are symmetric) and the result is trivial. We thus assume that \( A \) and \( B \) both have one nonzero eigenvalue, labeled as \( a \) and \( b \) respectively.
Since $B$ is real and symmetric, it has a complete set of real orthonormal vectors. These can be represented as

$$y_1 = \begin{pmatrix} \cos \theta \\ -\sin \theta \end{pmatrix}, \quad y_2 = \begin{pmatrix} \sin \theta \\ \cos \theta \end{pmatrix},$$

(11)

where $By_1 = by_1$, $By_2 = 0$. Since $A$ is diagonal, we have

$$A = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}.$$  

(12)

Let

$$\phi = \psi + \theta.$$  

(13)

It is readily seen that

$$\langle Gx, x \rangle = 1 - p_1(a)\cos^2\psi - p_2(b)\cos^2\phi - 2p_3(a, b)\cos \psi \cos \phi \cos \theta$$

$$+ i\left[p_4(a)\cos^2\psi + p_5(b)\cos^2\phi\right]$$

(14)

and

$$|\langle Gx, x \rangle|^2 \leq \left[1 - p_1(a)\cos^2\psi - p_2(b)\cos^2\phi + 2p_3(a, b)\cos \psi \cos \phi\right]^2$$

$$+ \left[p_4(a)\cos^2\psi + p_5(b)\cos^2\phi\right].$$  

(15)

for the appropriate sign preceding $p_3$. The choice of this sign does not affect stability by hypothesis (4). When $A$ and $B$ commute, $|\langle Gx, x \rangle|$ is maximized by choosing $x$ as one of the common eigenvectors, and so $\cos \theta = \pm 1$ or $0$ and $\cos \psi = \pm 1$ or $0$. By calculus it can be shown that the right hand side of Eq. (15) has its maximum when $\cos \theta = \pm 1, 0$ and $\cos \psi = \pm 1, 0$, i.e., the commuting case. Hence, $|\langle Gx, x \rangle|^2$ is maximal when $A$ and $B$ commute, and the theorem follows.

We now consider a subfamily of the above schemes which contains many of the second order accurate methods. For this subfamily we obtain much stronger conclusions.

**Theorem 3.** Let

$$G = I - \alpha A^2 - \beta B^2 - \gamma (AB + BA) + i[\delta A + \epsilon B].$$

(16)

Assume:

1. When $A$ and $B$ commute, $G$ is power bounded if the eigenvalues of $A, B$ satisfy $(a_i, b_i) \in S$ for some set $S$. The eigenvalues $a_i, b_i$ correspond to the same common eigenvector.
(2) When $A$ and $B$ commute, the stability does not depend on the sign of either $\gamma$ or $\delta \epsilon - 2\gamma$.

Then for real symmetric $A$ and $B$ the scheme is stable if $|Au|, |Bu| \in S$ for all real unit vectors $u$.

**Proof.** We write $G$ in the form $G = R + ij$. As before, $R$ and $J$ are real symmetric, and by Theorem 1 we need only show that $|(Gu, u)| \leq 1$ for real unit vectors $u$:

$$
(Gu, u) = (Ru, u) + i(Ju, u),
$$

$$
|Gu, u|^2 = (Ru, u)^2 + (Ju, u)^2
$$

$$
\leq (Ru, u)^2 + |Ju|^2 \quad \text{by Schwarz inequality}
$$

$$
= \left[ 1 - \alpha(A^2u, u) - \beta(B^2u, u) - \gamma(ABu, u) - \gamma(BAu, u) \right]^2
$$

$$
+ (\delta Au + \epsilon Bu, \delta Au + \epsilon Bu)
$$

$$
= \left[ 1 - \alpha(Au, Au) - \beta(Bu, Bu) - 2\gamma(Au, Bu) \right]^2
$$

$$
+ \delta^2(Au, Au) + 2\delta \epsilon(Au, Bu) + \epsilon^2(Bu, Bu).
$$

Since $A, B$ are symmetric and all quantities are real,

$$
|Gu, u|^2 \leq 1 - 2\alpha |Au|^2 - 2\beta |Bu|^2 - 4\gamma(Au, Bu) + \alpha^2 |Au|^4 + \beta^2 |Bu|^4
$$

$$
+ 4\gamma^2(Au, Bu)^2 + 4\alpha \gamma |Au|^2(Au, Bu) + 4\beta \gamma |Bu|^2(Au, Bu)
$$

$$
+ 2\alpha \beta |Au|^2|Bu|^2 + \delta^2 |Au|^2 + 2\delta \epsilon (Au, Bu) + \epsilon^2|Bu|^2
$$

$$
= 1 - (2\alpha - \delta^2)|Au|^2 - (2\beta - \epsilon^2)|Bu|^2 + (2\delta \epsilon - 4\gamma)(Au, Bu) + \alpha^2 |Au|^4
$$

$$
+ \beta^2 |Bu|^4 + 2\alpha \beta |Au|^2|Bu|^2 + 4\gamma(Au, Bu)[\alpha |Au|^2 + \beta |Bu|^2]
$$

$$
\leq 1 - (2\alpha - \delta^2)|Au|^2 - (2\beta - \epsilon^2)|Bu|^2 + |2\delta \epsilon - 4\gamma||Au| |Bu| + \alpha^2 |Au|^4
$$

$$
+ \beta^2 |Bu|^4 + 2\alpha \beta |Au|^2|Bu|^2 + 4\gamma_0 |Au| |Bu|[\alpha |Au|^2 + \beta |Bu|^2].
$$

(17)
γ₀ is chosen equal to ±γ so that the last term is positive. However, when A and B commute we have

\[
|g_i|^2 = (1 - αa_i^2 - βb_i^2 - 2γa_i b_i) + (δa_i + b_i)^2
\]

\[= 1 - (2α - δ^2)a_i^2 - (2β - δ^2)b_i^2 + (2δε - 4γ)a_i b_i + α^2 a_i^4 + β^2 b_i^4
\]

\[+ 2αβa_i b_i^2 + 4γa_i b_i (αa_i^2 + βb_i^2).
\] (18)

By assumption \(|g_i|^2 < 1\). Comparing Eqs. (17) and (18), we see that they are identical except for the fact that γ may possibly be replaced by −γ and \(2δε - 4γ\) is replaced by its absolute value. Thus by hypothesis (2) we have \(|(Gu, u)|^2 < 1\) for all real symmetric A and B.

We see that while Theorem 3 treats only a subclass of the schemes considered in Theorem 2, the results are stronger. First, we are no longer restricted to \(2 \times 2\) matrices. Second, the new stability depends on \(|Au|, |Bu|\) rather than the eigenvalues \(a_i, b_i\). Even in the case of commuting matrices the largest eigenvalue of A and B need not correspond to the same eigenvector. Thus, the use of \((a_i, b_i)\) rather than \((a_i, b_i)\) can significantly lower the allowable Δt. Thus, if the stability criterion is a monotonic function of the eigenvalues, the correct inequality for the noncommuting case is to use the pair \(|Au|, |Bu|\) for all unit vectors u rather than the pair \((ρ(A), ρ(B))\), which allows for no interaction between A and B [see, for example, Eqs. (25), (26)].

**Corollary 3.1.** Let G be given by Eq. (16) of Theorem 3, and assume the same hypotheses as in Theorem 3. In addition, assume that when \((a, b) ∈ S\) and A, B commute, the scheme is dissipative in the sense of Kreiss. Then for all real symmetric matrices A, B the scheme is dissipative in the sense of Kreiss whenever \(|Au|, |Bu|) ∈ S\) for all real unit vectors u.

**Proof.** When A, B are symmetric and commute, G is normal, and so

\[r_c(G) = \rho_c(G) < 1 - δ|ξ|^2r, \quad |ξ| < π.
\]

By the proof of the previous theorem, when A and B are symmetric but noncommuting,

\[ρ(G) < r(G) < r_c(G) < 1 - δ|ξ|^2r, \quad |ξ| < π.
\]
We note that in both Theorems 2 and 3 we have shown not only that is $G$ power bounded, but that $r(G) < 1$ and hence $\|G^n\| < 2$. From the viewpoint of matrix theory the 2 is a deficiency, since we allow any constant. However, from the viewpoint of numerical analysis, the bound of 2 is a virtue, since the stability constant affects the rate of convergence: the smaller the constant, the faster the convergence can be (see, for example, Kreiss and Oliger [7]).

3. APPLICATIONS

Having proven the conjecture for special families of schemes, we wish to use these results to study the stability of several known algorithms. We consider the equation

$$u_t = A_1 u_x + B_1 u_y \tag{19}$$

and let

$$A = A_1 \frac{\Delta t}{\Delta x}, \quad B = B_1 \frac{\Delta t}{\Delta y}.$$ 

Then the amplification matrix of the Lax-Wendroff method is given by

$$G_1 (\xi, \eta) = I + i(A \sin \xi + B \sin \eta) - A^2(1 - \cos \xi) - B^2 (1 - \cos \eta)$$

$$- \frac{i}{2} (AB + BA) \sin \xi \sin \eta. \tag{20}$$

Lax and Wendroff also consider a modification of the above scheme which includes a stabilizing term. The amplification matrix of this modified scheme is given by

$$G_2 (\xi, \eta) = G_1 (\xi, \eta) - \frac{A^2 + B^2}{2} (1 - \cos \xi)(1 - \cos \eta). \tag{21}$$

We shall now consider the stability properties of these schemes.

**Lemma.** Assume the matrices $A$ and $B$ are real symmetric and commute. Then the Lax-Wendroff scheme is stable if

$$|a|^2 / 3 + |b|^2 / 3 < 1, \tag{22}$$
where \( a_i, b_i \) are eigenvalues of \( A \) and \( B \) corresponding to the same eigenvector.

**Proof.** Let 
\[
\alpha = \sin \frac{\xi}{2}, \quad \beta = \sin \frac{\eta}{2}.
\]

Then the amplification matrix (Eq. 20) can be expressed as
\[
G = I + 2i \left[ A\alpha \sqrt{1 - a^2} + B\beta \sqrt{1 - \beta^2} \right] - 2A^2\alpha^2 - 2B^2\beta^2 \\
- 2(AB + BA)\alpha\beta \sqrt{1 - a^2} \sqrt{1 - \beta^2}.
\]

Since \( A \) and \( B \) commute, they share a common set of eigenvectors, and we can apply the spectral mapping theorem. Letting \( g_i \) be an eigenvalue of \( G \), we have
\[
g_i = 1 + 2i \left[ a_i\alpha \sqrt{1 - a^2} + b_i\beta \sqrt{1 - \beta^2} \right] - 2a_i^2\alpha^2 - 2b_i^2\beta^2 \\
- 4a_i b_i \alpha\beta \sqrt{1 - a^2} \sqrt{1 - \beta^2}.
\]

Furthermore, when \( A \) and \( B \) are real symmetric, \( G \) is a normal matrix, and the von Neumann condition is both necessary and sufficient. Hence, we must show that \( \max_i |g_i|^2 < 1 \).

Since \( A \) and \( B \) are symmetric, \( a_i, b_i \) are real, and so
\[
|g_i|^2 = \left[ 1 - 2a_i^2\alpha^2 - 2b_i^2\beta^2 - 4a_i b_i \alpha\beta \sqrt{1 - a^2} \sqrt{1 - \beta^2} \right]^2 \\
+ 4 \left[ a_i\alpha \sqrt{1 - a^2} + b_i\beta \sqrt{1 - \beta^2} \right]^2 \\
= 1 - 4a_i^2\alpha^4 - 4b_i^2\beta^4 + 4a_i^4\alpha^4 + 4b_i^4\beta^4 \\
+ 16a_i^2 b_i^2 \alpha^2 \beta^2 (1 - a^2)(1 - \beta^2) + 8a_i^2 b_i^2 \alpha^2 \beta^2 \\
+ 16a_i^2 b_i^3 \alpha^3 \beta \sqrt{1 - a^2} \sqrt{1 - \beta^2} + 16a_i b_i^3 \alpha \beta^3 \sqrt{1 - a^2} \sqrt{1 - \beta^2}.
\]

We first demonstrate the sufficiency of the condition. Since \(-1 < \alpha, \beta < 1\), we have \((1 - a^2)(1 - \beta^2) < 1\), and so
\[
|g_i|^2 \leq 1 - 4a_i^2\alpha^4 - 4b_i^2\beta^4 + 4a_i^4\alpha^4 + 4b_i^4\beta^4 + 24a_i^2 b_i^2 + 16|a_i b_i \alpha \beta| + 16|a_i b_i^3 \alpha \beta^3| \\
= 1 - 4 \left[ a_i^3 \alpha^4 + b_i^3 \beta^4 - (|a_i\alpha| + |b_i\beta|)^4 \right]. \tag{23}
\]
Let $\xi = \alpha / \beta$. Then
\[
|g_i|^2 = 1 - 4\beta^4 \left[ |a_i\xi^4| + |b_i|^2 - (|a_i\xi| + |b_i|)^4 \right].
\]

Thus, for stability, it is sufficient that
\[
q(\xi) = a_i^{2\xi^4} + b_i^2 - (|a_i\xi| + |b_i|)^4 > 0 \quad \text{for} \quad -\infty < \xi < \infty.
\]

For $\xi$ large, $q(\xi) \approx (a_i^2 - a_i^4)\xi^2 > 0$, since by the Courant condition we must have $a < 1$. We must thus find the minimum of $q$ and show that all possible real minima of $q$ are still positive.

$q'(\xi^*) = 0$ implies $4a_i^2(\xi^*)^3 - 4a_i(|a_i\xi^*| + |b_i|)^3 = 0$, or
\[
\xi^* = \frac{|b_i|}{|a_i|^{1/3} - |a_i|}.
\]
Then
\[
q(\xi^*) = b_i^3 \left[ 1 - \frac{|a_i|^{2/3}b_i^2}{(|a_i|^{1/3} - |a_i|)^3} \right] = \frac{|a_i|^{1/3}b_i^2}{(|a_i|^{1/3} - |a_i|)^3} \left[ (1 - |a_i|^{2/3})^3 - (|b_i|^{2/3})^3 \right],
\]
and so $q(\xi^*) > 0$ if $1 - |a_i|^{2/3} > |b_i|^{2/3}$, or equivalently $|a_i|^{2/3} + |b_i|^{2/3} < 1$ is a sufficient condition for stability.

To show necessity, we expand the amplification matrix $G$ for small $\xi$ and $\eta$. Using the spectral mapping theorem as before, we find that
\[
|g_i|^2 = 1 - 4 \left[ a_i^2\xi^4 + h_i^2\eta^4 - (|a_i\xi| + |b_i\eta|)^4 \right] + O(\xi^6 + \eta^6).
\]

We see that this is exactly the same right hand side as in Eq. (22), and so the rest of the argument proceeds as before.

Since a necessary condition for stability is that $G$ be power bounded for small $\xi$ and $\eta$, the condition
\[
|a_i|^{2/3} + |b_i|^{2/3} < 1
\]
is both necessary and sufficient for stability.

**Theorem 4.** Let $A$ and $B$ be real and symmetric matrices. Then a sufficient condition for the stability of the Lax-Wendroff scheme is that
\[
|Au|^{2/3} + |Bu|^{2/3} < 1
\]
for all real unit vectors $u$. A weaker condition is that

\[ \left[ \rho(A) \right]^{2/3} + \left[ \rho(B) \right]^{2/3} < 1. \tag{26} \]

**Proof.** Comparing the amplification matrix (20) for the Lax-Wendroff scheme with the general family (16) of Theorem 3, we see that

\[ y = \frac{1}{2} \sin \xi \sin \eta, \quad \delta = \sin \xi, \quad \epsilon = \sin \eta, \]

\[ \delta \epsilon - 2\gamma = 0. \]

From the proof of Theorem 3 it is obvious that $\xi \to -\xi$ does not affect stability (it corresponds to reversing the $x$ axis). Thus, hypothesis (1) together with Theorems 3 and 4 yield the result. Since $A$ and $B$ are symmetric, $\rho(A) = \max_{\|u\|=1} \|Au\|$, and so $\|Au\|^{2/3} + \|Bu\|^{2/3} < [\rho(A)]^{2/3} + [\rho(B)]^{2/3}$. Hence, condition (2) is also sufficient, but will usually allow smaller time steps than condition (1). If we reintroduce the original matrices $A_1, B_1$ of the differential equation (19) and assume that $\Delta x = \Delta y$, then we can rewrite the condition (25) as

\[ \Delta t \leq \frac{\Delta x}{\max \left[ \|A_1 u\|^{2/3} + \|B_1 u\|^{2/3} \right]^{3/2}}, \tag{27} \]

the maximum being taken over all real unit vectors.

**Corollary 4.1.** The Lax-Wendroff method is stable for real symmetric matrices $A$ and $B$ if $\rho(A) < 1/\sqrt{8}$, $\rho(B) < 1/\sqrt{8}$.

**Proof.** If $\rho(A) < 1/\sqrt{8}$, $\rho(B) < 1/\sqrt{8}$, then

\[ \left[ \rho(A) \right]^{2/3} + \left[ \rho(B) \right]^{2/3} < \frac{1}{2} + \frac{1}{2} < 1, \]

and the result follows from Theorem 4.

**Corollary 4.2.** The Lax-Wendroff method is stable for real symmetric matrices $A$ and $B$ if

\[ A^2 + B^2 \leq \frac{1}{4}. \tag{28} \]
Proof. Let $a = |Au|$, $b = |Bu|$; then from Theorem 4 the scheme is stable if
\[ 1 > (|a|^{2/3} + |b|^{2/3})^3 = |a|^2 + 3 |a|^{4/3} |b|^{2/3} + 3 |a|^{2/3} |b|^{4/3} + |b|^2 \]
but
\[ 0 < (|a|^{2/3} - |b|^{2/3})^2 (|a|^{2/3} + |b|^{2/3}) = |a|^2 - |a|^{4/3} |b|^{2/3} - |a|^{2/3} |b|^{4/3} + |b|^2, \]
or equivalently
\[ |a|^{4/3} |b|^{2/3} + |a|^{2/3} |b|^{4/3} < |a|^2 + |b|^2. \]
Substituting this into Eq. (29), we have that a sufficient condition for stability is that
\[ 1 > 4(|a|^2 + |b|^2) \]
but
\[ |a|^2 + |b|^2 = |Au|^2 + |Bu|^2 = (Au, Au) + (Bu, Bu) = (A^2 u, u) + (B^2 u, u), \]
i.e.,
\[ (A^2 u, u) + (B^2 u, u) \leq \frac{1}{4}, \]
and this is equivalent to the condition (28).

Corollary 4.1 is the original stability criterion of Lax and Wendroff, while Corollary 4.2 is an improvement found by Tadmor [11]. In Fig. 1 the boundary of the domain of stability is plotted according to Theorem 4 and its corollaries, as well as the CFL condition. At $A = B$ all the criteria agree and are the best possible. The improvement of Theorem 4 is most noticeable when $A$ is in some sense much larger or much smaller than $B$. For example, when $B = 0$, the two dimensional Lax-Wendroff scheme coincides with the one dimensional version, and so the correct stability condition is $(\Delta t/\Delta x)\rho(A_1) \leq 1$, and this is predicted by theorem 4. The original condition of Lax-Wendroff only allows $(\Delta t/\Delta x)\rho(A_1) \leq 1/\sqrt{8}$, while Tadmor's improvement allows $(\Delta t/\Delta x)\rho(A_1) \leq \frac{1}{2}$. 

Fig. 1. Graph 1: original stability condition of Lax and Wendroff for their scheme; graph 2: improvement of Tadmor; graph 3: stability condition given by Theorem 5; graph 4: stability condition for Lax-Wendroff scheme with stabilizer; graph 5: CFL stability condition.

Nevertheless, even the condition given by Theorem 4 need not be the best possible for noncommuting matrices. As an example, we consider the wave equation in two dimensions:

\[ u_t = c(u_x + v_y), \]
\[ v_t = c(u_y - v_x). \]  

Here

\[ A_1 = \begin{pmatrix} c & 0 \\ 0 & -c \end{pmatrix}, \quad B_1 = \begin{pmatrix} 0 & c \\ c & 0 \end{pmatrix} \]
and \(A_1B_1 \neq B_1A_1\); however, \(A_1^2B_1^2 = B_1^2A_1^2\). In this case both Theorem 4 and its two corollaries yield a sufficient condition that \(c \Delta t / \Delta x < 1/\sqrt{8}\). However, Clifton [2] has shown that for this particular case a sufficient condition for stability is \(c \Delta t / \Delta x < 1/\sqrt{2}\).

We now pass to the modified Lax-Wendroff method given by Eq. (21).

**Theorem 5.** A necessary and sufficient condition for the Lax-Wendroff scheme with stabilizer to be stable, for symmetric \(A\) and \(B\), is \(A^2 + B^2 < \frac{1}{2}\).

**Proof.** The sufficiency is shown by Lax and Wendroff [8]. The necessity follows from the observation that

\[
G_2(\pi, \pi) = I - 4(A^2 + B^2).
\]

Comparing the result with that of Theorem 4, we see that the addition of the stabilizer does not always allow larger time steps. Considering the case \(B=0\), the original scheme allows \(\rho(A) < 1\), while the "improved" method only allows \(\rho(A) < 1/\sqrt{2}\). In Fig. 1 one can see for which domains each of the schemes allows larger time steps. Of course, when \(A\) and \(B\) don't commute, the situation is more complicated. This shows that the effect of stabilizers is not always that expected. It is well known that the introduction of an artificial viscosity, to prevent nonlinear instabilities, may require the use of smaller time steps, since we have introduced parabolic type terms into the differential equation. On the other hand, in several dimensions the introduction of an artificial viscosity occasionally increases the allowable time step.

The above Theorems 2 and 3 can frequently be applied even in cases where the amplification matrix does not seem to have the required form. By introducing new matrices \(D, E\), which are functions of the matrices \(A\) and \(B\), the matrix \(G(D, E)\) now has the required form. As an example, we present the scheme of Livne [9], which is a noncentered scheme and has a domain of dependence of seven points, but is second order accurate. The amplification matrix of this scheme is

\[
G = I - A^2(1 - \cos \xi) - B^2(1 - \cos \eta)
\]

\[
+ \frac{1}{2}(AB + BA)[(1 - \cos \xi)(1 - \cos \eta) - \sin \xi \sin \eta]
\]

\[
+ \frac{i}{2} \left\{ A[\sin(\xi + \eta) + \sin \xi - \sin \eta] + B[\sin(\xi + \eta) - \sin \xi + \sin \eta] \right\}.
\]

We see that this amplification matrix does not satisfy the conditions of either
Theorem 2 or 3, since it is not true that changing the coefficient of $AB + BA$ does not affect stability. Since the scheme is not centered, changing $A$ to $-A$ (or equivalently $x$ to $-x$) does affect the stability criterion.

Instead we introduce the new matrices

$$D = A - B, \quad E = A + B.$$ 

In terms of these matrices Eq. (31) can be written as

$$G = I - \frac{1}{4}D^2[3 - 2\cos \xi - 2\cos \eta + \cos(\xi + \eta)] - \frac{1}{4}E^2[1 - \cos(\xi + \eta)]$$

$$+ \frac{1}{4}(DE + ED)(\cos \xi - \cos \eta) + \frac{i}{2} [E \sin(\xi + \eta) + D(\sin \xi - \sin \eta)].$$

If we write this in the form of Theorem 3 (Eq. 16), we have

$$G(\xi, \eta) = I - \alpha D^2 - \beta E^2 - \gamma(DE + ED) + i(\delta D + \epsilon E),$$

$$G(\eta, \xi) = I - \alpha D^2 - \beta E^2 + \gamma(DE + ED) + i(-\delta D + \epsilon E);$$

also

$$(\delta \epsilon - 2\gamma)(\xi, \eta) = -(\delta \epsilon - 2\gamma)(\eta, \xi).$$

Since interchanging $\xi$ and $\eta$ cannot affect stability in the commuting case, the hypotheses of Theorem 3 are satisfied. Hence, it suffices to find conditions for the power boundedness of Eq. (32) when $D$ and $E$ commute. When $A$ and $B$ commute, so do $D$ and $E$, and so this scheme satisfies the basic conjecture although it does not fit directly under Theorem 2 or 3. This indicates that the conjecture applies to a wider class of schemes than those given by Theorems 2 and 3.

To find a sufficient condition for stability of this scheme, we reduce the problem to the scalar case by Theorem 3. This scalar problem can now be solved computationally by allowing the scalars $d, e$ to vary discretely between plus and minus one and computing $G$ for a series of $\xi$ and $\eta$ between $-\pi$ and $+\pi$. One then checks for which $d$ and $e$ the absolute value of $G$ is less than or equal to one for all $\xi, \eta$. The result of this computer study yields

**Theorem 6.** The scheme of Livne, with amplification matrix given by Eq. (31), is stable for real symmetric $A, B$ if

$$(A - B)^2 < \frac{1}{4} \quad \text{and} \quad (A + B)^2 < 1$$
or else

\[ A^2 + B^2 \leq 0.36, \]

whichever is weaker—i.e., if \( \Delta t \) satisfies either of these inequalities, we have stability.

This condition is considerably better than that obtained by Livne, requiring that

\[ \|A - B\|^2 + \|A + B\|^2 \leq 1, \]

\[ \|A - B\|^2 \leq \frac{1}{4}. \quad (34) \]

We note that this condition is not the best possible, even for the commuting case.

Similarly, one can easily show

**Theorem 7.** The Lax-Friedrichs scheme with amplification matrix

\[ G(\xi, \eta) = \frac{1}{2}(\cos \xi + \cos \eta) + i(\Lambda \sin \xi + B \sin \eta) \]

is stable for real symmetric \( \Lambda, B \) if

\[ A^2 + B^2 \leq \frac{1}{2}. \]

If the matrices \( A \) and \( B \) commute, this condition is both necessary and sufficient for stability.

4. **Conclusions**

For a large family of schemes, which include most of the important second order methods, we have shown that the most stringent time steps are required when \( A \) and \( B \) commute. When \( A \) and \( B \) are real symmetric (and the scheme is hyperbolic symmetric) but do not commute, the time steps allowed without causing instability are usually strictly larger than that allowed for commuting \( A, B \). Theorem 5 shows that the strict inequality is not always true.

For the schemes considered, we have also shown the stronger condition that when there is stability for the symmetric hyperbolic case, we in fact have \( \|G^n\| \leq 2 \). It is interesting to note that in many special cases one can in
fact show that the scheme is strongly stable, \( \|G\| \leq 1 \), and so \( \|G^n\| \leq 1 \). Thus, for example, the rotated Richtmyer two step method can be shown to be strongly stable if \( A^2 + B^2 < 1 \) (see, for example, Zwas [14]). It is also trivial to show that Lax-Friedrichs scheme is strongly stable when \( A^2 < \frac{1}{4}, B^2 < \frac{1}{4} \), which is stronger than the condition of Theorem 7. Furthermore, Abarbanel and Gottlieb [1] have shown that the Lax-Wendroff scheme with stabilizer is strongly stable if \( A^2 < \frac{1}{4}, B^2 < \frac{1}{4} \). Again, Theorem 5 permits a larger time step than this. It is interesting to speculate whether one can strengthen the basic conjecture by even claiming that when a symmetric hyperbolic scheme is stable, it is powerbounded with constant 2 and for sufficiently small \( \Delta t/\Delta x \) even strongly stable.

The improved stability condition (27) for the Lax-Wendroff method goes part way to explaining the phenomenon that the allowable time step used in practice is considerably larger than that allowed by the original Lax-Wendroff criterion. However, even this condition is probably an unnecessarily strong restriction on \( \Delta t \) for the fluid dynamic equations. In this case one must analyze more carefully the affect of the noncommutation of the matrices.

The main theorems (Theorems 2 and 3) suffer from two disadvantages. First, they allow only simple combinations of the matrices \( A \) and \( B \)—e.g., \( A^4B^4 \) but not \( A^4B^4A^nB^n \cdots \). The use of the numerical radius is much more complicated in this latter case. The second restriction is that we have confined our attention to two dimensional schemes rather than multidimensional ones.

Thus, there is still much work left in verifying (or finding counterexamples to) the major conjecture about symmetric hyperbolic schemes.

REFERENCES


Received 5 January 1976