HIROSHIMA MATH. J.
4 (1974), 357-374

On the Stability of Finite-difference Schemes of Lax-Wendroff Type

Hisayoshi SHINTANI

(Received December 25, 1973)

1. Introduction

Let us consider the initial value problem for a linear hyperbolic system

$$(1.1)\qquad \frac{\partial u}{\partial t}=\sum_{j=1}^{n}A_j\frac{\partial u}{\partial x_j}\qquad(-\infty<x_j<\infty,\ 0\leqq t\leqq T),$$

$$(1.2)\qquad u(x,0)=u_0(x),$$

where u is an N-vector function of the real variables $x=(x_1, x_2,\dots, x_n)$ and t, $A_j(j=1, 2,\dots, n)$ are real constant $N\times N$ matrices, and $u_0(x)$ is a vector function belonging to L_2. It is assumed that the solution to this initial value problem exists and is unique.

For the numerical solution of this problem we use the finite-difference schemes of Lax-Wendroff type. Several sufficient conditions for their stability in the sense of Lax-Richtmyer [4][1)] are obtained when (1.1) is a symmetric hyperbolic system [4, 3, 2] and when it is a strictly hyperbolic system [5]. The object of this paper is to obtain some sufficient conditions for stability when (1.1) is a strongly hyperbolic system.

2. Notations and preliminaries

We denote by $|y|$ the Euclidean norm of the vector $y=(y_1, y_2,\dots, y_n)$, also denote by $|A|$ the spectral norm of the matrix A and put

$$(2.1)\qquad A(y)=\sum_{j=1}^{n}A_jy_j,\qquad A_0(y)=A\left(\frac{y}{|y|}\right)\qquad(y\neq 0).$$

In the sequel we assume that the eigenvalues of $A_0(y)$ are all real for any real $y\neq 0$ and that there exist a non-singular matrix $T(y)$ and a constant C_1 independent of y such that

$$(2.2)\qquad T(y)A_0(y)T(y)^{-1}=D_0(y),$$

1) Numbers in square brackets refer to the references listed at the end of this paper.

(2.3) $$|T(y)|\leqq C_1, \qquad |T(y)^{-1}|\leqq C_1,$$

where $D_0(y)$ is a diagonal matrix. Such a system (1.1) is called a strongly hyperbolic system. The system (1.1) is called strictly hyperbolic if the eigenvalues of $A_0(y)$ are all real and distinct for any real $y\neq 0$.

We consider a mesh imposed on the (x, t)-space with a spacing of $h>0$ in each x_j-direction $(j=1, 2,\dots, n)$ and a spacing of $k>0$ in the t-direction. The ratio $\lambda=k/h$ is to be kept constant as h varies. We wish to approximate (1.1) and (1.2) by the finite-difference scheme of the form

(2.4) $$v(x, t+k)=S_h v(x, t),$$

(2.5) $$v(x, 0)=u_0(x),$$

where

(2.6) $$S_h=\sum_\alpha C_\alpha T_1^{\alpha_1}T_2^{\alpha_2}\cdots T_n^{\alpha_n}, \qquad \alpha=(\alpha_1, \alpha_2,\dots, \alpha_n),$$

T_j is a translation operator defined by

(2.7) $$T_j^{\pm 1}v(x_1, x_2,\dots, x_n)=v(x_1,\dots, x_{j-1}, x_j\pm h, x_{j+1},\dots, x_n),$$

C_α's are constant $N\times N$ matrices and the summation extends over a finite number of terms.

To study the stability of the finite-difference scheme (2.4), we consider the amplification matrix

(2.8) $$C(\omega)=\sum_\alpha C_\alpha e^{i(\alpha, \omega)},$$

where

(2.9) $$(\alpha, \omega)=\sum_{j=1}^n \alpha_j\omega_j, \qquad \omega=h\xi,$$

$\xi=(\xi_1, \xi_2,\dots, \xi_n)$ is the variable vector dual to x in the Fourier transform. Let $\triangle_j=\sum_l b_l T_j^l$ be a finite-difference operator that approximates the differential operator $h\partial/\partial x_j$ and put $\sum_l b_l \exp(il\omega_j)=is_j(\omega)$. Then we assume that $s_j(\omega)$ is a sufficiently smooth real-valued periodic function of ω_j with period 2π and that for some positive integer r it can be written as follows:

(2.10) $$s_j(\omega)=\omega_j+O(|\omega_j|^{r+1}) \qquad (|\omega_j|\leqq\pi;\, j=1, 2,\dots, n).$$

Put

(2.11) $$s(\omega)=(s_1(\omega),\ s_2(\omega),\dots, s_n(\omega)).$$

Then the amplification matrix corresponding to the operator

(2.12) $$P_h=\lambda\sum_{j=1}^n A_j\triangle_j$$

can be expressed as $i\lambda A(s(\omega))$.

We denote by A^* the conjugate transpose of the matrix A and denote by $\lambda_j(A)$ $(j=1, 2,\dots, N)$ the eigenvalues of A. For hermitian matrices A and B we use the notation $A\geqq B$ when $A-B$ is positive semidefinite.

We shall make use of the following

LEMMA 1. *Let X and Y be $N\times N$ matrices and assume that all linear combinations with real coefficients of X and Y have only real eigenvalues. Let $\sigma=\sigma_1+i\sigma_2$ be any eigenvalue of the matrix $X+iY$, where σ_1 and σ_2 are real numbers. Then*

$$\lambda_1(X)\geqq\sigma_1\geqq\lambda_N(X), \qquad \lambda_1(Y)\geqq\sigma_2\geqq\lambda_N(Y),$$

where $\lambda_1(X)$ and $\lambda_N(X)$ are the largest and the smallest eigenvalues of X respectively.

This lemma follows from Lax's theorem on hyperbolic matrices [1, 6].

3. Schemes of Lax-Wendroff type

We are concerned with the case where the amplification matrix $C(\omega)$ can be written as follows:

$$C(\omega)=I+\sum_{j=1}^{r}\frac{1}{j!}[i\lambda A(s(\omega))]^j-\lambda^{2m}R(\omega, \lambda), \tag{3.1}$$

where

$$R(\omega, \lambda)=Q(t(\omega))+O(\lambda|t(\omega)|), \tag{3.2}$$

$$r\geqq 2m \qquad (m\geqq 1), \tag{3.3}$$

$$Q(y)=\sum_{j=1}^{n}Q_jy_j, \tag{3.4}$$

$R(\omega, \lambda)$ is continuous in ω and λ, Q_j $(j=1, 2,\dots, n)$ are real constant $N\times N$ matrices, $t(\omega)=(t_1(\omega), t_2(\omega),\dots, t_n(\omega))$, and $t_j(\omega)$ is a sufficiently smooth real-valued periodic function of ω_j with period 2π. For ω such that $t(\omega)\neq 0$ put

$$Q_0(\omega)=Q(t(\omega)/|t(\omega)|). \tag{3.5}$$

Let S be the set of all points ω such that $|\omega_j|\leqq\pi$ $(j=1, 2,\dots, n)$ and decompose S into the following three subsets:

$$S_1=\{\omega\in S: s(\omega)\neq 0\}, \quad S_2=\{\omega\in S: s(\omega)=0,\ t(\omega)\neq 0\},$$

$$S_3=\{\omega\in S: s(\omega)=0,\ t(\omega)=0\}.$$

In the sequel we assume that $s(\omega)$ does not vanish in S except for a finite number of points and that there exists a constant C_2 such that

$$|s(\omega)|^{r+l} \leqq C_2 |t(\omega)|, \tag{3.6}$$

where

$$l = \begin{cases} 1 & \text{if } r \text{ is odd}, \\ 2 & \text{if } r \text{ is even}. \end{cases} \tag{3.7}$$

Since S_2 and S_3 are finite sets, we can write them as follows:

$$S_2 = \{\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(s)}\}, \qquad S_3 = \{\omega^{(s+1)}, \ldots, \omega^{(t)}\}. \tag{3.8}$$

Put

$$\rho = \lambda |s(\omega)|, \qquad \sigma = \lambda^{2m} |t(\omega)|, \tag{3.9}$$

$$e(\omega; \lambda) = 1 - \max_j |\lambda_j(C(\omega))|^2. \tag{3.10}$$

For $\omega \in S_1$ put

$$T(s(\omega)) = T(\omega), \quad D_0(s(\omega)) = D_0(\omega), \quad |s(\omega)| D_0(\omega) = D(\omega), \tag{3.11}$$

$$D_0(\omega) = \operatorname{diag}(d_1(\omega), d_2(\omega), \ldots, d_N(\omega)), \tag{3.12}$$

$$T(\omega) Q_0(\omega) T(\omega)^{-1} = \tilde{Q}_0(\omega), \tag{3.13}$$

$$T(\omega) C(\omega) T(\omega)^{-1} = \tilde{C}(\omega). \tag{3.14}$$

Then $\tilde{C}(\omega)$ can be written as follows:

$$\tilde{C}(\omega) = I + \textstyle\sum_{j=1}^{r} \frac{1}{j!} [i\lambda D(\omega)]^j - \sigma[\tilde{Q}_0(\omega) + O(\lambda)]. \tag{3.15}$$

Now we shall show the following

THEOREM 1. *Suppose that there exist positive numbers δ and λ_0 such that*

$$|\lambda_j(C(\omega))| \leqq 1 - \delta\sigma \quad \text{for} \quad \lambda \leqq \lambda_0 \qquad (j = 1, 2, \ldots, N). \tag{3.16}$$

Then the scheme (2.4) is stable for $\lambda \leqq \lambda_0$.

PROOF. We consider first the case where $\omega \in S_1$. When r is odd, since by (3.6)

$$\rho^{r+1} = \lambda^{r+1} |s(\omega)|^{r+1} \leqq C_2 \lambda^{r+1} |t(\omega)| = C_2 \lambda^{r+1-2m} \sigma$$

and $r+1-2m \geqq 1$ by (3.3), $\tilde{C}(\omega)$ can be written as follows:

(3.17) $$\tilde{C}(\omega)=\exp(i\rho D_0(\omega))-\sigma[\tilde{Q}_0(\omega)+O(\lambda)].$$

When r is even, since

$$\rho^{r+2}=\lambda^{r+2}|s(\omega)|^{r+2}\leqq C_2\lambda^{r+2}|t(\omega)|=C_2\lambda^{r+2-2m}\sigma$$

and $r+2-2m\geqq 2$, we can write $\tilde{C}(\omega)$ as follows:

(3.18) $$\tilde{C}(\omega)=\exp\left(i\rho D_0(\omega)-\frac{1}{(r+1)!}(i\rho D_0(\omega))^{r+1}\right)-\sigma[\tilde{Q}_0(\omega)+O(\lambda)].$$

In both cases we have

(3.19) $$\tilde{C}(\omega)^*\tilde{C}(\omega)=I-\sigma[\tilde{Q}_0(\omega)^*+\tilde{Q}_0(\omega)+O(\lambda)].$$

There exists a unitary matrix $U(\omega)$ by which $\tilde{C}(\omega)$ is transformed into an upper triangular matrix, namely,

$$C'(\omega)=U\tilde{C}(\omega)U^*=K+R,$$

where

$$K=\mathrm{diag}(\lambda_1, \lambda_2,\dots, \lambda_N),\quad \lambda_j=\lambda_j(C(\omega))\qquad (j=1, 2,\dots, N),$$

$$R=(r_{ij}),\quad r_{ij}=0\qquad (i\geqq j).$$

Since by (3.16) and (3.19)

$$C'(\omega)^*C'(\omega)=K^*K+K^*R+R^*K+R^*R,$$

$$K^*K=I+O(\sigma),\quad C'(\omega)^*C'(\omega)=U\tilde{C}(\omega)^*\tilde{C}(\omega)U^*=I+O(\sigma),$$

it follows that

$$K^*R+R^*K+R^*R=O(\sigma).$$

From this it can be shown that $r_{ij}=O(\sigma)$ $(i<j)$. Hence $|R|\leqq\beta\sigma$ for some constant β. Put

$$\delta\sigma=y,\qquad \gamma=\max(1, (\beta/\delta)^{N-1}).$$

Then since

$$|(K+R)^p|\leqq\sum_{j=1}^{q}\binom{p}{j}|K|^{p-j}|R|^j,\qquad q=\min(p, N-1),$$

we have

$$|(K+R)^p|\leqq\sum_{j=1}^{q}\binom{p}{j}(1-y)^{p-j}(\beta y/\delta)^j\leqq\gamma\sum_{j=1}^{q}\binom{p}{j}(1-y)^{p-j}y^j\leqq\gamma.$$

Next we consider the case where $\omega\in S_2$. Since

$$C(\omega^{(j)})=I+O(\sigma_j),\quad \sigma_j=\lambda^{2m}|t(\omega^{(j)})| \qquad (j=1, 2,\dots, s),$$

there exist unitary matrices U_j and constants β_j $(j=1, 2,\dots, s)$ such that

$$C'(\omega^{(j)})=U_jC(\omega^{(j)})U_j^*=K_j+R_j,\qquad |R_j|\leqq\beta_j\sigma_j \quad (j=1, 2,\dots, s),$$

where K_j and R_j $(j=1, 2,\dots, s)$ are diagonal and strictly upper triangular matrices respectively. Put

$$\gamma_j=\max(1, (\beta_j/\delta)^{N-1}) \qquad (j=1, 2,\dots, s).$$

Then it can be shown as before that

$$|(K_j+R_j)^p|\leqq\gamma_j \qquad (j=1, 2,\dots, s).$$

In the case where $\omega\in S_3$, since $C(\omega)=I$, we put $C'(\omega)=I$.

Now put

$$T_0(\omega)=\begin{cases} U(\omega)T(\omega) & \text{if } \omega\in S_1, \\ U_j & \text{if } \omega=\omega^{(j)} \quad (j=1, 2,\dots, s), \\ I & \text{if } \omega\in S_3. \end{cases}$$

Then we can choose a constant C_0 such that

$$|T_0(\omega)|\leqq C_0, \qquad |T_0(\omega)^{-1}|\leqq C_0,$$

and it follows that

$$|C(\omega)^p|=|T_0(\omega)^{-1}C'(\omega)^pT_0(\omega)|\leqq C_0^2\gamma_0$$

for all p such that $pk\leqq T$, where $\gamma_0=\max(1, \gamma, \gamma_1, \gamma_2,\dots, \gamma_s)$. This implies the stability of the scheme (2.4).

In the following we shall give some sufficient conditions under which (3.16) is valid.

We consider the following two conditions.

CONDITION (I) : *There is a positive number p such that*

$$\lambda_j(Q_0(\omega))\geqq p \qquad \text{for all} \quad \omega\in S_2 \qquad (j=1, 2,\dots, N).$$

CONDITION (II): *There is a positive number p such that*

$$Q_0(\omega)^*+Q_0(\omega)\geqq 2pI \qquad \text{for all} \quad \omega\in S_2.$$

Then we have the following

LEMMA 2. *Suppose that the condition* (I) *or* (II) *is satisfied. Then there exists a positive number μ_1 such that*

(3.20) $$e(\omega;\lambda)\geqq p\sigma \quad \text{for} \quad \lambda\leqq\mu_1 \quad \text{and for all} \quad \omega\in S_2.$$

PROOF. We put for simplicity $\omega^{(k)}=\omega_0$ $(1\leqq k\leqq s)$ and $\lambda^{2m}|t(\omega_0)|=\sigma_0$. Then

$$C(\omega_0)=I-\sigma_0[Q_0(\omega_0)+O(\lambda)].$$

In the case where the condition (II) is satisfied, since

$$C(\omega_0)^*C(\omega_0)=I-\sigma_0[Q_0(\omega_0)^*+Q_0(\omega_0)+O(\lambda)],$$

there is a positive number μ_1' such that

$$|C(\omega_0)|^2\leqq 1-p\sigma_0 \qquad \text{for} \quad \lambda\leqq\mu_1',$$

and it follows that

$$e(\omega_0;\lambda)\geqq 1-|C(\omega_0)|^2\geqq p\sigma_0 \qquad \text{for} \quad \lambda\leqq\mu_1'.$$

Next we consider the case where the condition (II) is satisfied. There is a unitary matrix U such that $UQ_0(\omega_0)U^*=K+R$, where K is a diagonal matrix and

$$R=(r_{ij}), \quad r_{ij}=0 \qquad (i\geqq j).$$

Let g be a positive number and put

$$G=\mathrm{diag}(g, g^2,\dots, g^N), \qquad V=GU.$$

Then we have

$$VQ_0(\omega_0)V^{-1}=K+\tilde{R}, \qquad \tilde{R}=GRG^{-1}=(\tilde{r}_{ij}),$$

where

$$\tilde{r}_{ij}=r_{ij}g^{i-j} \quad (i<j), \qquad \tilde{r}_{ij}=0 \quad (i\geqq j).$$

Hence we can choose g so that

$$|\tilde{r}_{ij}|\leqq p/(2N) \qquad (i<j).$$

Then since $K\geqq pI$, by Gerschgorin's theorem

$$2K+\tilde{R}^*+\tilde{R}\geqq(3p/2)I.$$

Put $C'(\omega_0)=VC(\omega_0)V^{-1}$. Then since

$$C'(\omega_0)^*C'(\omega_0)=I-\sigma_0(2K+\tilde{R}^*+\tilde{R})+O(\lambda\sigma_0),$$

for some constant $\mu_1'>0$

$$|\lambda_j(C'(\omega_0))|^2 \leqq 1 - p\sigma_0 \quad \text{for} \quad \lambda \leqq \mu_1' \quad (j=1, 2, \dots, N).$$

From this it follows that

$$e(\omega_0; \lambda) \geqq p\sigma_0 \quad \text{for} \quad \lambda \leqq \mu_1'.$$

Since S_2 is a finite set, we can choose a positive number μ_1 so that (3.20) is valid. This completes the proof of lemma 2.

By continuity of eigenvalues, we have the following

COROLLARY. *Suppose that the condition* (I) *or* (II) *is satisfied. Then, for each* $\omega^{(k)} \in S_2$ $(1 \leqq k \leqq s)$, *there exist a neighborhood* $N(\omega^{(k)})$ *of* $\omega^{(k)}$ *and a positive number* μ_2 *independent of* k *such that*

$$e(\omega; \lambda) \geqq p\sigma/2 \quad \text{for} \quad \lambda \leqq \mu_2 \quad \text{and} \quad \omega \in N(\omega^{(k)}). \tag{3.21}$$

We have the following stability criterion in terms of the symmetric part of $\tilde{Q}_0(\omega)$.

THEOREM 2. *Assume that there exists a positive number* q *such that*

$$\tilde{Q}_0(\omega)^* + \tilde{Q}_0(\omega) \geqq 2qI \tag{3.22}$$

and that the condition (I) *or* (II) *is satisfied. Then the scheme* (2.4) *is stable for sufficiently small* λ.

PROOF. By (3.22) and (3.19) we can choose a constant $\mu > 0$ such that

$$e(\omega; \lambda) \geqq q\sigma \quad \text{for} \quad \lambda \leqq \mu \quad \text{and} \quad \omega \in S_1.$$

By lemma 2 we have a constant μ_1 such that (3.20) is valid for $\omega \in S_2$. When $\omega \in S_3$, it is clear that $\rho = 0$ and $\lambda_j(C(\omega)) = 1$ $(j=1, 2, \dots, N)$. Hence there exist positive numbers δ and λ_0 such that

$$e(\omega; \lambda) \geqq 2\delta\sigma \quad \text{for} \quad \lambda \leqq \lambda_0.$$

From this it follows that

$$|\lambda_j(C(\omega))| \leqq 1 - \delta\sigma \quad \text{for} \quad \lambda \leqq \lambda_0 \quad (j=1, 2, \dots, N)$$

and the scheme (2.4) is stable for $\lambda \leqq \lambda_0$ by theorem 1.

We now introduce the following two assumptions.

ASSUMPTION (A): *For each* $\omega^{(k)} \in S_3$ $(s+1 \leqq k \leqq t)$, *there exists a neighborhood* $V(\omega^{(k)})$ *of* $\omega^{(k)}$ *satisfying the following conditions*:

(i) $s(\omega) \neq 0$ *in* $V(\omega^{(k)})$ *except for* $\omega = \omega^{(k)}$;

(ii) *there exists a constant* C_3 *such that*

$$|t(\omega)| \leqq C_3 |s(\omega)| \qquad \text{for} \quad \omega \in V(\omega^{(k)}); \tag{3.23}$$

(iii) $y = s(\omega)$ *has the inverse function* $\omega = f(y)$ *in* $V(\omega^{(k)})$.

ASSUMPTION (B): *For each* $\omega^{(k)} \in S_3$ $(s+1 \leqq k \leqq t)$, *there exists a neighborhood* $V(\omega^{(k)})$ *of* $\omega^{(k)}$ *satisfying the conditions* (i) *and* (ii).
Then we have the following stability criterion in terms of $\tilde{Q}_0(\omega)$.

THEOREM 3. *Under the assumption* (A), *suppose that there exists a positive number* q *such that all the eigenvalues of any principal submatrix of* $\tilde{Q}_0(\omega)$ *are not less than* q. *Suppose also that the condition* (I) *or* (II) *is satisfied. Then the scheme* (2.4) *is stable for sufficiently small* λ.

PROOF. Put for simplicity $\omega^{(k)} = \omega_0$. By the assumption there is a positive number γ_0 such that

$$f(y) \in V(\omega_0) \qquad \text{for} \quad |y| < \gamma_0 .$$

Let S^{n-1} be the unit spherical surface in the real n-space and define $N(\omega_0)$ by

$$N(\omega_0) = \{\omega \colon \omega = f(\gamma l),\ 0 \leqq \gamma < \gamma_0,\ l \in S^{n-1}\} .$$

Then $N(\omega_0)$ is a neighborhood of ω_0.

For any fixed $l \in S^{n-1}$, put $\hat{\omega} = f(\gamma l)$ $(0 < \gamma < \gamma_0)$. Then since $s(\hat{\omega}) = \gamma l$ and $|s(\hat{\omega})| = \gamma$, $D_0(\hat{\omega})$ does not depend on γ. Let e_j $(j = 1, 2, \ldots, p)$ be all the distinct eigenvalues of $D_0(\hat{\omega})$ and let m_j $(j = 1, 2, \ldots, p)$ be their multiplicities respectively. Without loss of generality we may assume that $D_0(\hat{\omega})$ is of the form

$$D_0(\hat{\omega}) = \begin{pmatrix} e_1 I_1 & & O \\ & e_2 I_2 & \\ & & \ddots \\ O & & e_p I_p \end{pmatrix},$$

where I_k is the unit matrix of order m_k. Corresponding to this form, we partition $\tilde{Q}_0(\hat{\omega})$ as follows:

$$\tilde{Q}_0(\hat{\omega}) = \begin{pmatrix} Q_{11} & Q_{12} & \cdots & Q_{1p} \\ \cdots & \cdots & & \cdots \\ Q_{p1} & Q_{p2} & \cdots & Q_{pp} \end{pmatrix},$$

where $Q_{jk}(\hat{\omega})$ is an $m_j \times m_k$ matrix.

There is a unitary matrix $U_j(\hat{\omega})$ $(1 \leqq j \leqq p)$ such that

$$U_j(\hat{\omega})^* Q_{jj}(\hat{\omega}) U_j(\hat{\omega}) = K_j(\hat{\omega}) + R_j(\hat{\omega}),$$

where the matrices $K_j(\hat{\omega})$ and $R_j(\hat{\omega})$ are diagonal and strictly upper triangular respectively. Making use of these, we construct the following matrices:

$$U = \mathrm{diag}(U_1, U_2, \dots, U_p),$$

$$\mathrm{E} = \mathrm{diag}(K_1 + R_1, K_2 + R_2, \dots, K_p + R_p), \quad F = (F_{jk}),$$

where

$$F_{jk}(\hat{\omega}) = (e_k - e_j)^{-1} Q_{jk}(\hat{\omega}) U_k(\hat{\omega}) \qquad (j \neq k),$$

$$F_{jj}(\hat{\omega}) = 0 \qquad (j, k = 1, 2, \dots, p).$$

Put

$$\hat{\rho} R = \hat{\rho} U + i \hat{\sigma} F,$$

where

$$\hat{\rho} = \lambda \gamma, \qquad \hat{\sigma} = \lambda^{2m} |t(\hat{\omega})|.$$

Then it follows that

$$(i\hat{\rho} D_0 - \hat{\sigma} \tilde{Q}_0) \hat{\rho} R = \hat{\rho} R (i \hat{\rho} D_0 - \hat{\sigma} E) + O(\hat{\sigma}^2).$$

$|F(\hat{\omega})|$ is bounded because $\tilde{Q}_0(\hat{\omega})$ is bounded in norm. Since by (3.23) $|t(\hat{\omega})| \leqq C_3 \gamma$, for some constant $\mu_3 > 0$

$$|\hat{\rho}^{-1} \hat{\sigma} U^* F| < 1 \qquad \text{for} \quad \lambda \leqq \mu_3.$$

For such λ, R^{-1} exists and we have

$$R^{-1}(i\hat{\rho} D_0 - \hat{\sigma} \tilde{Q}_0) R = i \hat{\rho} D_0 - \hat{\sigma} E + O(\hat{\rho}^{-1} \hat{\sigma}^2).$$

Since $R_j(\hat{\omega})$ $(1 \leqq j \leqq p)$ is bounded in norm, there is a positive number g_j such that

$$|\tilde{r}_{kl}^{(j)}| \leqq q/(2m_j) \qquad (k < l),$$

where

$$\tilde{R}_j(\hat{\omega}) = G_j R_j(\hat{\omega}) G_j^{-1} = (\tilde{r}_{kl}^{(j)}), \qquad G_j = \mathrm{diag}\,(g_j, g_j^2 \dots, g_j^{m_j}).$$

Put

$$G = \mathrm{diag}(G_1, G_2, \dots, G_p), \quad G R^{-1} = V, \quad C'(\hat{\omega}) = V \tilde{C}(\hat{\omega}) V^{-1},$$

$$\tilde{E}=\mathrm{diag}(K_1+\tilde{R}_1,\ K_2+\tilde{R}_2,\dots,\ K_p+\tilde{R}_p)\,.$$

Then we have

$$C'(\hat{\omega})=I+\textstyle\sum_{j=1}^{r}\frac{1}{j!}(i\lambda D(\hat{\omega}))^j-\hat{\sigma}\tilde{E}+O(\lambda\hat{\sigma})\,,$$

and so

$$C'(\hat{\omega})^*C'(\hat{\omega})=I-\hat{\sigma}(\tilde{E}^*+\tilde{E})+O(\lambda\hat{\sigma})\,.$$

Since $K_j\geqq qI_j$ $(j=1, 2,\dots, p)$ by the assumption, it follows that

$$\tilde{E}^*+\tilde{E}\geqq(3q/2)I\,,$$

and for some constant $\mu'_3>0$

$$e(\hat{\omega};\lambda)\geqq q\hat{\sigma}\qquad\text{for}\quad \lambda\leqq\mu'_3\,.$$

By continuity of $e(\omega;\lambda)$, there exist a positive number $\tilde{\mu}_3$ and a neighborhood $U(l)$ of l on S^{n-1} such that

$$e(\omega;\lambda)\geqq q\sigma/2\quad\text{for}\quad \omega=f(\gamma\tilde{l})\quad\text{and}\quad \lambda\leqq\tilde{\mu}_3\,,$$

where $\tilde{l}\in U(l)$ and $0<\gamma<\gamma_0$. Then by the Heine-Borel theorem we can cover S^{n-1} by a finite number of such neighborhoods. Hence we can choose a positive number μ such that for $\omega\in N(\omega_0)$ $(\omega\neq\omega_0)$

$$e(\omega;\lambda)\geqq q\sigma/2\qquad\text{for}\quad\lambda\leqq\mu\,. \tag{3.24}$$

By continuity of eigenvalues, (3.24) holds for all $\omega\in N(\omega_0)$.

Since S_3 is a finite set, there exist a positive number μ_3 and neighborhoods $N(\omega^{(k)})$ of $\omega^{(k)}$ $(k=s+1,\ s+2,\dots,\ t)$ such that

$$e(\omega;\lambda)\geqq q\sigma/2\quad\text{for}\quad\lambda\leqq\mu_3\quad\text{and}\quad\omega\in N(\omega^{(k)})\qquad(k=s+1,\dots,t)\,.$$

Put

$$\Omega=S-\cup_{j=1}^{t}N(\omega^{(j)}),\quad \varepsilon=\inf_{\omega\in\Omega}|s(\omega)|,\quad \alpha=\sup_{\omega\in\Omega}|t(\omega)|\,.$$

Let ω_0 be any point belonging to Ω, e_j $(j=1, 2,\dots, p)$ be all the distinct eigenvalues of $D_0(\omega_0)$ and m_j $(j=1, 2,\dots, p)$ be their multiplicities respectively. Replacing $\hat{\omega}$, $\hat{\rho}$ and $\hat{\sigma}$ by ω_0, $\rho_0=\lambda|s(\omega_0)|$ and $\sigma_0=\lambda^{2m}|t(\omega_0)|$ respectively, we define the matrices U, E, F and R analogously. Since $\rho_0^{-1}\sigma_0\leqq\lambda^{2m-1}\alpha/\varepsilon$, we can find a constant $\mu'_4>0$ such that

$$|\rho_0^{-1}\sigma_0U^*F|<1\qquad\text{for}\quad\lambda\leqq\mu'_4\,.$$

Then R^{-1} exists for such λ and there holds

$$e(\omega_0;\lambda)\geqq q\sigma_0 \qquad \text{for} \quad \lambda\leqq\mu_4'.$$

By continuity of eigenvalues there exist a positive number μ_4'' and a neighborhood $N(\omega_0)$ of ω_0 such that

$$e(\omega;\lambda)\geqq q\sigma/2 \quad \text{for} \quad \lambda\leqq\mu_4'' \quad \text{and} \quad \omega\in N(\omega_0).$$

By the Heine-Borel theorem we can cover Ω by a finite number of such neighborhoods, and so for some constant $\mu_4>0$

$$e(\omega;\lambda)\geqq q\sigma/2 \quad \text{for} \quad \lambda\leqq\mu_4 \quad \text{and} \quad \omega\in\Omega.$$

If we put

$$\lambda_0=\min(\mu_2,\mu_3,\mu_4), \qquad 4\delta=\min(p,q),$$

then (3.16) is satisfied and the theorem has been proved.

We have the following stability criterion for a strictly hyperbolic system in terms of the diagonal elements of $\tilde{Q}_0(\omega)$.

THEOREM 4. *For a strictly hyperbolic system* (1.1), *under the assumption* (B), *suppose that there exists a positive number q such that the diagonal elements of $\tilde{Q}_0(\omega)$ are all not less than q. Suppose also that the condition* (I) *or* (II) *is satisfied. Then the scheme* (2.4) *is stable for sufficiently small λ.*

PROOF. By the assumption there is a constant β such that

$$|d_j(\omega)-d_k(\omega)|\geqq\beta>0 \qquad (j\neq k;\ j,k=1,2,\dots,N). \tag{3.25}$$

Put

$$E(\omega)=\operatorname{diag}(q_{11}(\omega),q_{22}(\omega),\dots,q_{NN}(\omega)),$$

$$\rho R=\rho I+i\sigma P, \qquad \Omega_1=S-\cup_{j=1}^{s}N(\omega^{(j)}),$$

where

$$\tilde{Q}_0(\omega)=(q_{jk}(\omega)), \qquad P=(p_{jk}),$$

$$p_{jk}=q_{jk}/(d_k-d_j) \quad (j\neq k), \qquad p_{jj}=0 \quad (j,k=1,2,\dots,N).$$

Then by (3.25) we have

$$(i\lambda D-\sigma\tilde{Q}_0)\rho R=\rho R(i\lambda D-\sigma E)+O(\sigma^2),$$

because $|P|$ is bounded. Since $|t(\omega)|/|s(\omega)|$ is bounded in $\Omega_1\cap S_1$, R^{-1} exists for sufficiently small λ and

$$R^{-1}(i\lambda D-\sigma\tilde{Q}_0)R=i\lambda D-\sigma E+O(\rho^{-1}\sigma^2).$$

If we put $C'(\omega)=R^{-1}\tilde{C}(\omega)R$, then

$$C'(\omega)=I+\sum_{j=1}^{r}\frac{1}{j!}(i\lambda D)^{j}-\sigma E+O(\lambda\sigma)\,,$$

so that

$$C'(\omega)^{*}C'(\omega)=I-2\sigma E+O(\lambda\sigma)\,.$$

Since $E\geqq qI$ by the assumption, there is a positive number μ_5 such that

$$e(\omega;\lambda)\geqq q\sigma \quad \text{for} \quad \lambda\leqq\mu_5 \quad \text{and} \quad \omega\in\Omega_1\cap S_1\,.$$

By continuity of $e(\omega;\lambda)$ this result is valid also for $\omega\in S_3$. Thus if we choose

$$\lambda_0=\min(\mu_2,\mu_5), \qquad 2\delta=\min(p/2,q)\,,$$

then (3.16) is satisfied and the theorem has been proved.

Now we shall show the following

THEOREM 5. *Suppose that all linear combinations with real coefficients of $A(s(\omega))$ and $Q(t(\omega))$ have only real eigenvalues and that there exists a positive number q such that the eigenvalues of $Q_0(\omega)$ are all not less than q. Then the scheme* (2.4) *is stable for sufficiently small λ.*

PROOF. Put

$$M(\omega)=i\rho D_0(\omega)-\sigma\tilde{Q}_0(\omega)$$

and let $-\sigma_j+i\rho_j$ $(j=1,2,\dots,N)$ be the eigenvalues of $M(\omega)$. Then since

$$T(\omega)^{-1}M(\omega)T(\omega)=i\lambda A(s(\omega))-\lambda^{2m}Q(t(\omega))\,,$$

by lemma 1 we have

$$\sigma_j\geqq q\sigma \qquad (j=1,2,\dots,N)\,.$$

By Gerschgorin's theorem we can find a suffix $k(j)$ such that

$$\rho_j=\rho d_{k(j)}+O(\sigma), \qquad \sigma_j=O(\sigma)\,.$$

There exists a unitary matrix $U(\omega)$ such that $UMU^{*}=K+R$, where

$$K=\mathrm{diag}(-\sigma_1+i\rho_1,\dots,-\sigma_N+i\rho_N), \qquad R=(r_{ij}), \quad r_{ij}=0 \qquad (i\geqq j)\,.$$

Put

$$U\tilde{Q}_0U^{*}=L_1+E_1+R_1\,,$$

$$\rho UD_0U^{*}=\rho E+\sigma E_2+L_2+L_2^{*}\,,$$

where the matrices L_1 and L_2 are strictly lower triangular, R_1 is strictly upper triangular, E_1, E_2 and E are diagonal matrices and they are all bounded in norm. Then it follows that $iL_2=\sigma L_1$. Hence

$$(3.26)\qquad i\rho UD_0U^*=i\rho E+\sigma(L_1+iE_2-L_1^*),$$
$$K=i\rho E+i\sigma E_2-\sigma E_1,\ R=\sigma S,\quad S=-L_1^*-R_1.$$

There are positive numbers g and C_4 such that

$$VMV^{-1}=K+\sigma\tilde{S},\quad \tilde{S}=GSG^{-1}=(\tilde{s}_{ij}),\quad |\tilde{s}_{ij}|\leqq q/(4N)\quad (i<j),$$
$$|V|\leqq C_4,\qquad |V^{-1}|\leqq C_4,$$

where

$$V=GU,\qquad G=\mathrm{diag}(g, g^2,\dots, g^N).$$

We consider first the case where r is odd. By (3.17) $\tilde{C}(\omega)$ can be written as follows:

$$\tilde{C}(\omega)=\exp(M(\omega))+O(\lambda\sigma).$$

Since

$$C'(\omega)=V\tilde{C}(\omega)V^{-1}=\exp(K+\sigma\tilde{S})+O(\lambda\sigma),$$

it follows that

$$C'(\omega)^*C'(\omega)=\exp(K^*+K)+\sigma(\tilde{S}^*+\tilde{S})+O(\lambda\sigma).$$

By Gerschgorin's theorem the eigenvalues of $\exp(K^*+K)+\sigma(\tilde{S}^*+\tilde{S})$ are not greater than

$$\max_j \exp(-2\sigma_j)+q\sigma/4.$$

Since

$$\exp(-2\sigma_j)+q\sigma/4=1-(2\sigma_j-q\sigma/4)+O(\sigma^2),\qquad 2\sigma_j-q\sigma/4\geqq 7q\sigma/4,$$

we have $e(\omega;\lambda)\geqq q\sigma$ for sufficiently small λ. The condition (I) is satisfied by the assumption and $e(\omega;\lambda)=\sigma=0$ for $\omega\in S_3$. Hence there exist constants λ_0 and δ such that (3.16) is satisfied and the scheme (2.4) is stable for $\lambda\leqq\lambda_0$.

Next we consider the case where r is even. Put

$$M_1(\omega)=M(\omega)-\frac{1}{(r+1)!}(i\rho D_0(\omega))^{r+1}.$$

Then by (3.26) we have

$$U(i\rho D_0)^{r+1}U^* = (i\rho E)^{r+1} + \lambda^r \sigma W,$$

where $|W|$ is bounded. Hence

$$VM_1V^{-1} = K - \frac{1}{(r+1)!}(i\rho E)^{r+1} - \sigma\tilde{S} + \lambda^r\sigma\tilde{W},$$

$$\tilde{W} = GWG^{-1} = (\tilde{w}_{ij}).$$

Put

$$\frac{1}{(r+1)!} i^r E^{r+1} = \mathrm{diag}\,(e_1, e_2 \dots, e_N)$$

and let $-\alpha + i\beta$ be any eigenvalue of $M_1(\omega)$. Then by Gerschgorin's theorem we can find a suffix k such that

$$|-\sigma_k + i(\rho_k - \rho^{r+1}e_k) + \alpha - i\beta| \leqq \sigma[\textstyle\sum_{j=k+1}^N |\tilde{s}_{kj}| + \lambda^r \sum_{j=1}^N |\tilde{w}_{kj}|].$$

Since

$$\textstyle\sum_{j=k+1}^N |\tilde{s}_{kj}| \leqq q/4$$

and $\sigma_k \geqq q$, for sufficiently small λ we have $|\alpha - \sigma_k| \leqq q\sigma/2$ and $\alpha \geqq q\sigma/2$. Hence there is a positive number μ_5 such that

$$\alpha_j \geqq q\sigma/2 \quad \text{for} \quad \lambda \leqq \mu_5 \qquad (j = 1, 2, \dots, N),$$

where $-\alpha_j + i\beta_j$ $(j = 1, 2, \dots, N)$ are the eigenvalues of $M_1(\omega)$. By (3.18) $\tilde{C}(\omega)$ can be written as follows:

$$\tilde{C}(\omega) = \exp(M_1(\omega)) + O(\lambda\sigma).$$

The stability of the scheme (2.4) can be shown as in the previous case.

EXAMPLE. Consider the Lax-Wendroff scheme for the system (1.1) with $n = 2$, $N = 3$ and

$$A_1 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 2 & 1 & 4 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Then $r = 2$, $m = 1$ and

$$s_j(\omega) = \sin\omega_j, \quad t_j(\omega) = \sin^4(\omega_j/2) \qquad (j = 1, 2),$$

$$C(\omega) = I + i\lambda A(s(\omega)) - \frac{1}{2}\lambda^2 A(s(\omega))^2 - \lambda^2 Q(t(\omega)),$$

where

$$A(y)=\begin{pmatrix} 3y_1+2y_2 & y_2 & 4y_2 \\ y_2 & y_1+2y_2 & 0 \\ 0 & 0 & y_1+2y_2 \end{pmatrix},$$

$$Q(y)=2(A_1^2y_1+A_2^2y_2)=\begin{pmatrix} 18y_1+10y_2 & 8y_2 & 32y_2 \\ 8y_2 & 2y_1+10y_2 & 8y_2 \\ 0 & 0 & 2y_1+8y_2 \end{pmatrix}.$$

If we choose

$$T(\omega)=\begin{pmatrix} 1 & -p & 0 \\ p & 1 & -4 \\ 0 & 0 & 1 \end{pmatrix},$$

then

$$T(\omega)^{-1}=\begin{pmatrix} q & pq & 4pq \\ -pq & q & 4q \\ 0 & 0 & 1 \end{pmatrix},$$

$$|T(\omega)|\leqq 5, \qquad |T(\omega)^{-1}|\leqq 5,$$

$$d_1(\omega)=2(s_1'+s_2')+\operatorname{sgn}(s_1'), \qquad d_2(\omega)=2(s_1'+s_2')-\operatorname{sgn}(s_1'),$$

$$d_3(\omega)=s_1'+2s_2',$$

where

$$s_j'=s_j(\omega)/|s(\omega)| \quad (j=1, 2), \qquad p=\operatorname{sgn}(s_1')s_2'/(1+|s_1'|),$$

$$q=1/(1+p^2), \qquad \operatorname{sgn}(x)=\begin{cases} 1 & \text{if } x\geqq 0 \\ -1 & \text{if } x<0 \end{cases}.$$

Hence this system is strongly hyperbolic but not strictly hyperbolic. The condition (3.6) is satisfied because

$$|s(\omega)|^4\leqq 32\sqrt{2}\,|t(\omega)|.$$

Since $Q(y)$ has only real eigenvalues for any real y and

$$\lambda_j(A_1^2)\geqq 1, \quad \lambda_j(A_2^2)\geqq 1 \qquad (j=1, 2, 3),$$

by Lax's concavity theorem for hyperbolic matrices [1]

$$\lambda_j(Q_0(\omega)) \geqq 2(t_1(\omega)+t_2(\omega))/|t(\omega)| \geqq 2 \qquad (j=1, 2, 3),$$

and the condition (I) is satisfied. It is easily verified that the conditions of theorems 2, 3 and 5 are all satisfied. It can be shown that, when $\omega_1=0$ and $\omega_2=\pi$, $|C(\omega)|>1$ for sufficiently small λ.

4. Examples of the schemes

We shall present examples of the schemes that satisfy the conditions (3.2), (3.3) and (3.6). For this end we introduce the following finite-difference operators:

$$P_1=\sum_{j=1}^n A_j \triangle_j, \quad P_2=\sum_{j=1}^n A_j \triangle_j^{(2)},$$

$$Q_1=\sum_{j=1}^n A_j^2 D_{2j}+\sum_{j\neq k} A_j A_k \triangle_j \triangle_k,$$

$$Q_2=\sum_{j=1}^n A_j^2 D_{2j}^{(2)}+\sum_{j\neq k} A_j A_k \triangle_j^{(2)} \triangle_k^{(2)},$$

$$Q_3=\sum_{j=1}^n A_j^2 D_{2j}^{(3)}+\sum_{j\neq k} A_j A_k \triangle_j^{(2)} \triangle_k^{(2)},$$

where

$$\triangle_j=\frac{1}{2}(T_j-T_j^{-1}), \qquad D_{2j}=T_j-2I+T_j^{-1} \qquad (j=1, 2,\dots, n),$$

$$\triangle_j^{(2)}=\triangle_j\Big(I-\frac{1}{6}D_{2j}\Big), \qquad D_{2j}^{(2)}=\frac{1}{3}(4D_{2j}-\triangle_j^2),$$

$$D_{2j}^{(3)}=\frac{1}{9}(16D_{2j}-7\triangle_j^2).$$

Put

$$\alpha_j=\sin\omega_j, \quad X_j=\sin^2(\omega_j/2) \qquad (j=1, 2,\dots, n),$$

$$p_1=\sum_{j=1}^n A_j\alpha_j, \qquad p_2=\sum_{j=1}^n A_j\alpha_j\Big(1+\frac{2}{3}X_j\Big), \qquad r_1=\sum_{j=1}^n A_j\alpha_j X_j,$$

$$q_1=\sum_{j=1}^n A_j^2 X_j(3-8X_j-4X_j^2)+\sum_{j\neq k} A_j A_k\Big[\frac{3}{2}(X_j+X_k)+X_jX_k\Big],$$

$$q_2=\sum_{j=1}^n A_j^2 X_j^3(2+X_j), \qquad q_3=\sum_{j=1}^n A_j^2 X_j^2(1+X_j)^2,$$

$$r_2=4\sum_{j=1}^n A_j^2 X_j\Big(1+\frac{1}{3}X_j\Big)+\sum_{j\neq k} A_j A_k\alpha_j\alpha_k\Big(1+\frac{2}{3}X_j\Big)\Big(1+\frac{2}{3}X_k\Big).$$

Then we obtain the following scheme with accuracy of order 3:

$$S_h = I + \lambda P_2 + \frac{1}{2}\lambda^2 Q_3 + \frac{1}{6}\lambda^3 P_1 Q_1 ,$$

$$C(\omega) = I + \sum_{j=1}^{3} \frac{1}{j!}(i\lambda p_2)^j - \frac{8}{9}\lambda^2 q_3 + \frac{1}{27}\lambda^3(3r_1 p_2^2 + 2p_1 q_1) .$$

We have also the following scheme with accuracy of order 4:

$$S_h = I + \lambda P_2 + \frac{1}{2}\lambda^2 Q_2\left(I + \frac{1}{3}\lambda P_2 + \frac{1}{12}\lambda^2 Q_2\right),$$

$$C(\omega) = I + \sum_{j=1}^{4} \frac{1}{j!}(i\lambda p_2)^j - \frac{8}{9}\lambda^2 q_2 - \frac{8}{27} i\lambda^3 q_2 p_2 + \frac{2}{27}\lambda^4(p_2^2 q_2 + q_2 r_2) .$$

References

[1] Lax, P. D., *Differential equations, difference equations and matrix theory,* Comm. Pure Appl. Math., **11** (1958), 175–194.

[2] Lax, P. D. and Wendroff, B., *Difference schemes for hyperbolic equations with high order of accuracy,* Comm. Pure Appl. Math., **17** (1964), 381–398.

[3] Parlett, B., *Accuracy and dissipation in difference schemes,* Comm. Pure Appl. Math., **19** (1966), 111–123.

[4] Richtmyer, R. D. and Morton, K. W., *Difference methods for initial-value problems.* Interscience Publishers, New York, 1967.

[5] Yamaguti, M., *Some remarks on the Lax-Wendroff finite-difference scheme for nonsymmetric hyperbolic systems,* Math. Comp., **21** (1967), 611–619.

[6] Yamaguti, M. and Nogi, T., *Basis of numerical analysis* (Japanese). Kyoritsu Syuppan, Tokyo, 1969.

Department of Mathematics,
Faculty of Science,
Hiroshima University