Theorem 6.2 If the conditions (1)—(5) hold, then the process of computing $\mu_0^{(m)}$, $\mu^{(m)}$ of (5.14)′ for $m=1, 2, \cdots, M$ successively, and the process of computing U_m for $m=M-1, M-2, \cdots, 0$ successively by using (5.13) are stable.

In order to prove this theorem, we need to notice only the following fact. The relation (5.14)′

$$\tilde{\mu}_0^{(m+1)} U_{m+1} = \tilde{\mu}^{(m+1)}$$

can be obtained from the relation (5.4)′, the equations (5.1b) with $m=0$ and (5.11), i.e., from

$$\begin{cases} \mu_0^{(m+1)} U_{m+1} + \mu_1^{(m+1)} U_0 = \mu^{(m+1)}, \\ \nu_0^{(m+1)} U_{m+1} + \nu_1^{(m+1)} U_0 = \nu^{(m+1)}, \end{cases} \tag{5.4}'$$

$$\bar{G}_2^{(0)} U_0 = \bar{F}_2^{(0)}, \tag{5.1b}$$

$$\tilde{\mu}_0^{(0)} U_0 = \tilde{\mu}^{(0)}. \tag{5.11}$$

Hence, there is an invertible matrix α_{m+1} such that

$$\mu_0^{(m+1)} - \mu_1^{(m+1)} \begin{pmatrix} \tilde{\mu}_0^{(0)} \\ \bar{G}_2^{(0)} \\ \nu_1^{(m+1)} \end{pmatrix}^{-1} \begin{pmatrix} O \\ O \\ \nu_0^{(m+1)} \end{pmatrix} = \alpha_{m+1} \tilde{\mu}_0^{(m+1)},$$

$$\mu^{(m+1)} - \mu_1^{(m+1)} \begin{pmatrix} \tilde{\mu}_0^{(0)} \\ \bar{G}_2^{(0)} \\ \nu_1^{(m+1)} \end{pmatrix}^{-1} \begin{pmatrix} \tilde{\mu}^{(0)} \\ \bar{F}_2^{(0)} \\ \nu^{(m+1)} \end{pmatrix} = \alpha_{m+1} \tilde{\mu}^{(m+1)}.$$

In order to distinguish $\mu_0^{(m)}$, $\mu^{(m)}$ in the two procedures, we shall change $\mu_0^{(m)}$, $\mu^{(m)}$ in (5.11) and (5.14)′ into $\tilde{\mu}_0^{(m)}$, $\tilde{\mu}^{(m)}$.

By using the above fact, this theorem can be proved, but to save space we shall not give the proof here.

Appendix 1

Stability of Difference Schemes for Pure-Initial-Value Problems with Variable Coefficients[1)]

Introduction

Lax et al.[1]–[4], and Kreiss[5] have discussed the stability of difference schemes for pure-initial-value problems with variable coefficients, and have developed some theorems. However, concerning this subject, there still remain problems to be solved.

This paper discusses the stability of difference schemes for hyperbolic

1) This paper is an English translation of the paper in "Mathematicae Numericae Sinica, 1978, No. 1, 33—43".

systems with two independent variables. For any explicit and implicit horizontal–three–point schemes and for several horizontal–multi–point explicit schemes, in which the Rusanov[6] third–order scheme, and the Burstein–Mirin[7] third–order scheme are included (i.e., for almost all schemes[6]–[15] being used in practical work), we present sufficient conditions of stability of schemes with variable coefficients. These are: (i) the von Neumann condition (4); (ii) condition (5), which guarantees the difference equations to be well–conditioned, and which is automatically fulfilled for explicit schemes; (iii) the condition that the coefficients of the schemes are smooth functions of x and t. Conditions (4) and (5) are also necessary for schemes with constant coefficients. We do not require that the schemes be dissipative, nor do we require that the operators be symmetric. These conditions can be applied conveniently to both explicit and implicit schemes.

1. *General Results*

We consider the following initial–value problem of hyperbolic systems:

$$\begin{cases} \dfrac{\partial U}{\partial t} + A(x, t)\dfrac{\partial U}{\partial x} = 0, \\ U(x, 0) = f(x), \end{cases} \tag{1}$$

where $U(x, t)$ and $f(x)$ are N–dimensional vectors, and $A(x, t)$ is an $N \times N$–matrix. $A(x, t)$ has N real eigenvalues and N linearly independent eigenvectors, i.e., there is a matrix G such that

$$A = G^{-1}\Lambda G, \tag{2}$$

where $\Lambda(x, t)$ is a real diagonal matrix. We discuss the following horizontal $(H+1)$–point difference scheme:

$$\sum_{h=0}^{H} R_{h,m}^{k}(\Delta)\, U_{m+h}^{k+1} = \sum_{h=0}^{H} S_{h,m}^{k}(\Delta)\, U_{m+h}^{k}, \tag{3}$$

where we adopt the following notation:

$$U_m^k = U(m\Delta x, k\Delta t);$$
$$R_{h,m}^{k}(\Delta) = R_h(m\Delta x, k\Delta t, \Delta x, \Delta t);$$
$$S_{h,m}^{k}(\Delta) = S_h(m\Delta x, k\Delta t, \Delta x, \Delta t);$$
$R_{h,m}^{k}(\Delta)$ and $S_{h,m}^{k}(\Delta)$ are $N \times N$–matrices.

We also introduce other notation as follows: $R_{h,m}^{k}$, $S_{h,m}^{k}$ denote $R_{h,m}^{k}(\Delta)\,|_{\Delta\omega=\Delta t=0}$, $S_{h,m}^{k}(\Delta)\,|_{\Delta\omega=\Delta t=0}$ respectively; $\overline{F}$ denotes $(F_0, F_1, \cdots, F_H)$, where F_h is an $N \times N$–matrix, $h = 0, 1, \cdots, H$; $\overline{F}^*$ is the conjugate transposition of $\overline{F}$; and $F(\theta) \equiv \sum_{h=0}^{H} F_h e^{ih\theta}$. Moreover, the superscript k and subscript m will be omitted, if no confusion results. We will discuss problems in L_2 space and suppose $\Delta t/\Delta x$ is bounded.

First, we give a lemma.

Lemma 1 If $\sum_l \Big(\sum_{h=0}^{H} D_{h,l}e^{ih\theta}\Big)^* E_l \Big(\sum_{h=0}^{H} D_{h,l}e^{ih\theta}\Big)=0$, then every sum[1] of all "N-matrix-elements" located on a diagonal line of $Q=\sum_l \bar{D}_l^* E_l \bar{D}_l$ is a null-matrix, where E_l and $D_{h,l}$ are $N\times N$-matrices, and where the matrix Q is an $N(H+1)\times N(H+1)$-matrix.

Proof From

$$\sum_l \Big(\sum_{h=0}^{H} D_{h,l}e^{ih\theta}\Big)^* E_l \Big(\sum_{h=0}^{H} D_{h,l}e^{ih\theta}\Big)$$

$$=\sum_l \sum_{h=-H}^{H} \sum_j D_{j,l}^* E_l D_{j+h,l}e^{ih\theta}=\sum_{h=-H}^{H} \sum_l \sum_j D_{j,l}^* E_l D_{j+h,l}e^{ih\theta}=0,$$

we obtain

$$\sum_l \sum_j D_{j,l}^* E_l D_{j+h,l}=0, \quad h=0, \pm 1, \pm 2, \cdots, \pm H,$$

where j satisfies $0\leqslant j\leqslant H$ and $0\leqslant j+h\leqslant H$. However, $\sum_l \sum_j D_{j,l}^* E_l D_{j+h,l}$ is the sum of all "N-matrix-elements" located on "the h-th diagonal line" of $\sum_l \bar{D}_l^* E_l \bar{D}_l$. Therefore, we obtain the desired result.

Then, we give two theorems.

Theorem 1 If for any m, k, and θ, the following conditions are fulfilled:

(i) there exist two invertible matrices N_m^k and G_m^k such that

$$N_m^k R_m^k(\theta) G_m^{k-1}=\Lambda_{1,m}^k(\theta), \quad N_m^k S_m^k(\theta) G_m^{k-1}=\Lambda_{2,m}^k(\theta),$$

and

$$\Lambda_{1,m}^{k*}(\theta)\Lambda_{1,m}^k(\theta)-\Lambda_{2,m}^{k*}(\theta)\Lambda_{2,m}^k(\theta)\geqslant 0, \tag{4}$$

where $\Lambda_{1,m}^k(\theta)$ and $\Lambda_{2,m}^k(\theta)$ are diagonal matrices, and $\geqslant 0$ denotes that the matrix on its left side is nonnegative definite;

(ii)
$$\Lambda_{1,m}^{k*}(\theta)\Lambda_{1,m}^k(\theta)-c_1 I\geqslant 0, \tag{5}$$

where c_1 is a positive constant and I is an $N\times N$-unit matrix;

(iii) R_h, S_h, N, N^{-1}, G and G^{-1} satisfy the Lipschitz condition with respect to x and t, i.e., for every element f of these matrices, the following relations

$$|f_{m+1}^k-f_m^k|\leqslant c_2\Delta x, \quad |f_m^{k+1}-f_m^k|\leqslant c_2\Delta t$$

---

1) Let $A_{i,j}$ be an $N\times N$-matrix. We call

$$\sum_{0<\{i,\,i+h\}<H} A_{i,i+h}$$

the sum of all "N-matrix-elements" located on the h-th diagonal line of

$$\begin{pmatrix} A_{00}, A_{01}, \cdots, A_{0H} \\ A_{10}, A_{11}, \cdots, A_{1H} \\ \cdots\cdots\cdots\cdots \\ A_{H0}, A_{H1}, \cdots, A_{HH} \end{pmatrix},$$

where $0\leqslant\{i, i+h\}\leqslant H$ denotes the fact that $0\leqslant i\leqslant H$ and $0\leqslant i+h\leqslant H$.

are fulfilled, and these elements are bounded;

(iv) among R_h, S_h and $R_h(\Delta)$, $S_h(\Delta)$, there are the relations

$$\|R_h(\Delta)-R_h\|\leqslant c_2\Delta x,$$
$$\|S_h(\Delta)-S_h\|\leqslant c_2\Delta x;$$

(v)
$$V(\theta)\equiv R^*(\theta)N^*NR(\theta)-S^*(\theta)N^*NS(\theta)$$
$$=G^*\Lambda_1^*(\theta)\Lambda_1(\theta)G-G^*\Lambda_2^*(\theta)\Lambda_2(\theta)G$$

can be rewritten as

$$V(\theta)=\sum_l D_l^*(\theta)M_lD_l(\theta), \tag{6}$$

where M_l and $D_{h,l}$ are $N\times N$-matrices, satisfying the Lipschitz condition, and M_l is a nonnegative definite matrix,

then the scheme (3) is stable in the space L_2, i.e., there exists a constant c_3 such that

$$\|U^k\|^2\leqslant c_3\|U^0\|^2,\quad 0\leqslant k\Delta t\leqslant T_1,$$

where T_1 is a bounded constant and

$$\|U^k\|^2=\sum_m U^*(m\Delta x,\ k\Delta t)\,U(m\Delta x,\ k\Delta t)\,\Delta x.$$

Proof We use the energy method. We take

$$T^k=\sum_m\Big(N_m^k\sum_{h=0}^H R_{h,m}^{k-1}(\Delta)\,U_{m+h}^k\Big)^*\Big(N_m^k\sum_{h=0}^H R_{h,m}^{k-1}(\Delta)\,U_{m+h}^k\Big)\Delta x$$
$$=\sum_m(\overline{U},\ \overline{R}^*N^*N\overline{R}\overline{U})_m^k\Delta x+O(\Delta x)\|U^k\|^2,$$

as the energy sum, where $\overline{U}_m^*=(U_m^*,\ U_{m+1}^*,\ \cdots,\ U_{m+H}^*)$.

First, we prove that from conditions (i), (iii)—(v) we can obtain the inequality

$$T^{k+1}-T^k\leqslant c_4\Delta x\|U^k\|^2, \tag{7}$$

where c_4 is a constant. In fact, by using (3) and conditions (iii), (iv), we have

$$T^{k+1}-T^k=\sum_m\Big[\Big(N_m^k\sum_{h=0}^H S_{h,m}^kU_{m+h}^k\Big)^*\Big(N_m^k\sum_{h=0}^H S_{h,m}^kU_{m+h}^k\Big)$$
$$-\Big(N_m^k\sum_{h=0}^H R_{h,m}^kU_{m+h}^k\Big)^*\Big(N_m^k\sum_{h=0}^H R_{h,m}^kU_{m+h}^k\Big)\Big]\Delta x$$
$$+O(\Delta x)\|U^k\|^2$$
$$=\sum_m(\overline{U},\ (\overline{S}^*N^*N\overline{S}-\overline{R}^*N^*N\overline{R})\overline{U})_m^k\Delta x+O(\Delta x)\|U^k\|^2.$$

According to conditions (iii) and (v) and by using Lemma 1, we know that every sum of all "N-matrix-elements" located on a "diagonal line" of the matrix

$$Q_1=\overline{S}^*N^*N\overline{S}-\overline{R}^*N^*N\overline{R}+\sum_l\overline{D}_l^*M_l\overline{D}_l$$

is a null-matrix of order N, and that every element of Q_1 satisfies the Lipschitz condition with respect to x. Thus, from the fact that $\sum_t \bar{D}_t^* M_t \bar{D}_t$ is nonnegative definite, we can obtain inequality (7) immediately.

Then, from conditions (ii) and (iii), we derive

$$c_5 \| U^k \|^2 \leqslant T^k \leqslant c_5^{-1} \| U^k \|^2, \tag{8}$$

where c_5 is a positive constant. It is easy to obtain the right half of the inequality. In the following we derive the left half. By using condition (ii) and the Fejér-Reisz theorem[17], we know that there exists a diagonal matrix-polynomial $J(\theta) = \sum_{h=0}^{H} J_h e^{ih\theta}$ such that

$$R^*(\theta) N^* N R(\theta) - G^* \frac{c_1}{2} G$$

$$= G^* \Lambda_1^*(\theta) \Lambda_1(\theta) G - G^* \frac{c_1}{2} G = G^* J^*(\theta) J(\theta) G.$$

Moreover, because $J^*(\theta) J(\theta) \geqslant (c_1/2) I > 0$, from the fact that the elements of $\Lambda_1(\theta)$ satisfy the Lipschitz condition, we can see that the elements of $J(\theta)$ also satisfy the Lipschitz condition [2]. Thus, according to Lemma 1, every sum of all "N-matrix-elements" located on a "diagonal line" of the matrix

$$Q_2 = \bar{R}^* N^* N \bar{R} - \bar{G}^* \frac{c_1}{2} \bar{G} - \bar{J}_G^* \bar{J}_G$$

is a null-matrix of order N, where

$$\bar{G} = (G, \underbrace{0, \cdots, 0}_{H}), \quad 0 \text{ being an } N \times N\text{-null-matrix},$$

$$\bar{J}_G = (J_0 G, J_1 G, \cdots, J_H G).$$

In addition,

$$\left(\bar{U}, \bar{G}^* \frac{c_1}{2} \bar{G} \bar{U} \right)_m = \left(U, G^* \frac{c_1}{2} G U \right)_m \geqslant \frac{c_1}{2 \| G_m^{-1} \|^2} (U, U)_m.$$

Therefore, from the fact that $\bar{J}_G^* \bar{J}_G$ is nonnegative definite, we can derive the left half of (8) immediately. Obviously, the inequality

$$T^{k+1} - T^k \leqslant c_4 \Delta x \| U^k \|^2 \leqslant \frac{c_4}{c_5} \Delta x T^k$$

follows from (7) and (8) immediately. Furthermore, we can obtain the following inequalities:

$$T^{k+1} \leqslant \left(1 + \frac{c_4}{c_5} \Delta x\right) T^k \leqslant \left(1 + \frac{c_4}{c_5} \Delta x\right)^{k+1} T^0,$$

and

$$\| U^{k+1} \|^2 \leqslant c_5^{-2} \left(1 + \frac{c_4}{c_5} \Delta x\right)^{k+1} \| U^0 \|^2.$$

Therefore, when $\Delta x/\Delta t$ is bounded, there is a positive constant c_3 such that

$$\|U^k\|^2 \leqslant c_3 \|U^0\|^2,$$

i.e., the conclusion of the theorem is proved.

Definition Let $\tilde{U}^k = (\cdots, U_0^{k^*}, U_1^{k^*}, \cdots, U_m^{k^*}, \cdots)^*$ be the solution of the system of equations

$$B\tilde{U}^k = C.$$

If there exists a positive constant such that the inequality

$$\|U^k\|^2 \leqslant c_6 \|C\|^2$$

is fulfilled for any solution $\tilde{U}^k$, then we say that the system is well-conditioned in L_2, where $\|C\|^2 = \sum_j |c_j|^2 \Delta x$ and c_j is an element of C.

Obviously, if the left half of (8) is fulfilled and $\|N_m^k\|$ is bounded, then the difference equations in (3) are well-conditioned. Therefore, we can obtain the following theorem:

Theorem 2 If for any m and θ, there are two invertible matrices N and G such that $NR(\theta)G^{-1} = \Lambda_1(\theta)$ and

$$\Lambda_1^*(\theta)\Lambda_1(\theta) - c_1 I \geqslant 0, \tag{5}$$

and if conditions (iii) and (iv) of Theorem 2 are fulfilled, then the difference equations in (3) are well-conditioned. In the case with constant coefficients, if there exist two invertible matrices N and G such that all NR_hG^{-1} are diagonal matrices, then condition (5) is also necessary.

Proof From the proof of Theorem 1, we know that (8) also is fulfilled under the conditions of this theorem. Therefore the first part of the conclusion is proved.

If (5) is not fulfilled, then there exists a number θ^* such that the j-th element of $\Lambda_1(\theta) = NR(\theta)G^{-1}$ is equal to zero when $\theta = \theta^*$. Therefore, in the case of constant coefficients, the equations

$$N\sum_{h=0}^{H} R_h U_{m+h}^k = \sum_{h=0}^{H} NR_h U_{m+h}^k = 0, \quad m = \cdots, 0, 1, 2, \cdots,$$

have a nontrivial solution:

$$U_{m+h}^k = G^{-1}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ e^{i(m+h)\theta^*} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \longleftarrow \text{the } j\text{-th element.}$$

Therefore, the equations are not well conditioned, i.e., the second part of the theorem is proved.

From Theorem 2, we know that the Fourier method can be used for studying the properties of difference equations.

2. *Applications*

Before we apply the results in Section 1 to some concrete schemes, we point out the following fact: for quite a few difference schemes which can be written in the form (3), there exists an invertible matrix $\tilde{H}_m^k$ such that $R_{h,m}^k$ and $S_{h,m}^k$ can be rewitten as

$$\begin{cases} R_{h,m}^k = [\tilde{H}\sum\limits_j c_{j,h}A^j]_m^k = [\tilde{H}G^{-1}(\sum\limits_j c_{j,h}\Lambda^j)G]_m^k, \\ S_{h,m}^k = [\tilde{H}\sum\limits_j b_{j,h}A^j]_m^k = [\tilde{H}G^{-1}(\sum\limits_j b_{j,h}\Lambda^j)G]_m^k, \end{cases} \tag{9}$$

where $c_{j,h}$ and $b_{j,h}$ are scalars. The C-I-R scheme[8], the Lax scheme[9], the Richtmyer two step scheme[14], the Wendroff scheme[10], and the Thomée scheme[11], ···, have this feature, and $\tilde{H}_m^k = I$ or G_m^k. For these schemes, there are the following results:

Lemma 2 If the coefficients of scheme (3) satisfy (9), then there exists N such that

$$R^*(\theta)N^*NR(\theta) = G^*\Big(\sum_{h=0}^{H} d_h \sin^{2h}\frac{\theta}{2}\Big)G, \tag{10}$$

$$R^*(\theta)N^*NR(\theta) - S^*(\theta)N^*NS(\theta) = G^*\Big(\sum_{h=0}^{H} a_h \sin^{2h}\frac{\theta}{2}\Big)G, \tag{11}$$

where a_h and d_h are real diagonal matrices (for their concrete expressions see (14) and (13)). If condition (iii) of Theorem 1 is also fulfilled, then every element of a_h satisfies the Lipschitz condition with respect to x and t.

Proof Let $N = G\tilde{H}^{-1}$. From (9), we have

$$G^{-1*}R^*(\theta)N^*NR(\theta)G^{-1} = \Big(\sum_{h=0}^{H}\sum_j c_{j,h}\Lambda^j e^{ih\theta}\Big)^*\Big(\sum_{h=0}^{H}\sum_j c_{j,h}\Lambda^j e^{ih\theta}\Big)$$
$$= \sum_{h=-H}^{H} \tilde{d}_h e^{ih\theta},$$

where $\tilde{d}_h = \sum\limits_{0<(i,i+h)\leq H} (\sum\limits_j c_{j,i}\Lambda^j)(\sum\limits_j c_{j,i+h}\Lambda^j)$. Obviously, $\tilde{d}_h$ is a real diagonal matrix, and $\tilde{d}_h = \tilde{d}_{-h}$. Moreover, for any integer k, there is the expression

$$\cos k\theta = \sum_{j=0}^{k} e_{k,j}\sin^{2j}\frac{\theta}{2},$$

where any $e_{k,j}$ is a real constant. (For $k=1, 2, 3, 4$, the concrete expressions are as follows:

$$\begin{cases} \cos\theta = 1 - 2\sin^2\dfrac{\theta}{2}, \\ \cos 2\theta = 1 - 8\sin^2\dfrac{\theta}{2} + 8\sin^4\dfrac{\theta}{2}, \\ \cos 3\theta = 1 - 18\sin^2\dfrac{\theta}{2} + 48\sin^4\dfrac{\theta}{2} - 32\sin^6\dfrac{\theta}{2}, \\ \cos 4\theta = 1 - 32\sin^2\dfrac{\theta}{2} + 160\sin^4\dfrac{\theta}{2} - 256\sin^6\dfrac{\theta}{2} + 128\sin^8\dfrac{\theta}{2}). \end{cases} \tag{12}$$

Therefore, we have (10):

$$R^*(\theta)N^*NR(\theta)=G^*\Big(\tilde{d}_0+2\sum_{h=1}^{H}\tilde{d}_h\cos h\theta\Big)G=G^*\Big(\sum_{h=0}^{H}d_h\sin^{2h}\frac{\theta}{2}\Big)G,$$

where

$$\begin{cases} d_0=\tilde{d}_0+2\sum\limits_{k=1}^{H}e_{k,0}\tilde{d}_k=\Big(\sum\limits_{h=0}^{H}\sum\limits_{j}c_{j,h}\Lambda^j\Big)^2, \quad h=1,2,\cdots H, \\ d_h=2\sum\limits_{k=h}^{H}e_{k,h}\tilde{d}_k=2\sum\limits_{k=h}^{H}e_{k,h}\sum\limits_{0\leqslant i\leqslant i+k\leqslant H}(\sum\limits_{j}c_{j,i}\Lambda^j)(\sum\limits_{j}c_{j,i+k}\Lambda^j). \end{cases} \tag{13}$$

We can also obtain a similar expression for $S^*(\theta)N^*NS(\theta)$ in which the first term is $\Big(\sum\limits_{h=0}^{H}\sum\limits_{j}b_{j,h}\Lambda^j\Big)^2$. Moreover,

$$\sum_{h=0}^{H}\sum_{j}c_{j,h}\Lambda^j=\sum_{h=0}^{H}\sum_{j}b_{j,h}\Lambda^j.$$

Therefore, we obtain the equality (11), and a_h has the following expression:

$$\begin{aligned} a_h=2\sum_{k=h}^{H}e_{k,h}\sum_{0\leqslant i\leqslant i+k\leqslant H}&[(\sum_{j}c_{j,i}\Lambda^j)(\sum_{j}c_{j,i+k}\Lambda^j) \\ &-(\sum_{j}b_{j,i}\Lambda^j)(\sum_{j}b_{j,i+k}\Lambda^j)]. \end{aligned} \tag{14}$$

Furthermore, it is obvious that every element of a_h satisfies the Lipschitz condition if condition (iii) is fulfilled. Thus, we have obtained all conclusions of this theorem.

Lemma 3 Let $f_0+f_1\sin^2\frac{\theta}{2}+f_2\sin^4\frac{\theta}{2}$ be a scalar quadratic polynomial in $\sin^2\frac{\theta}{2}$. Necessary and sufficient conditions for $f_0+f_1\sin^2\frac{\theta}{2}+f_2\sin^4\frac{\theta}{2}\geqslant 0$ are

$$f_0\geqslant 0, \quad 2f_0+f_1+2(f_0(f_0+f_1+f_2))^{1/2}\geqslant 0, \quad f_0+f_1+f_2\geqslant 0. \tag{15}$$

Proof According to the equality

$$\begin{aligned} &f_0+f_1\sin^2\frac{\theta}{2}+f_2\sin^4\frac{\theta}{2} \\ &=f_0\cos^4\frac{\theta}{2}+(2f_0+f_1)\sin^2\frac{\theta}{2}\cos^2\frac{\theta}{2}+(f_0+f_1+f_2)\sin^4\frac{\theta}{2} \\ &=\Big[f_0^{1/2}\cos^2\frac{\theta}{2}-(f_0+f_1+f_2)^{1/2}\sin^2\frac{\theta}{2}\Big]^2 \\ &\quad+[2f_0+f_1+2f_0^{1/2}(f_0+f_1+f_2)^{1/2}]\sin^2\frac{\theta}{2}\cos^2\frac{\theta}{2}, \end{aligned}$$

we know that if

$$f_0\geqslant 0, \quad 2f_0+f_1+2f_0^{1/2}(f_0+f_1+f_2)^{1/2}\geqslant 0, \quad \text{and} \quad f_0+f_1+f_2\geqslant 0,$$

then

$$f_0+f_1\sin^2\frac{\theta}{2}+f_2\sin^4\frac{\theta}{2}\geqslant 0,$$

i.e., (15) is a sufficient condition. Furthermore, if we let $\sin^2\frac{\theta}{2}\approx 0$, $\cos^2\frac{\theta}{2}\approx 0$ and $f_0^{1/2}\cos^2\frac{\theta}{2}-(f_0+f_1+f_2)^{1/2}\sin^2\frac{\theta}{2}=0$, and observe the right-hand side of the equality, then we can find that (15) is also a necessary condition.

From Lemma 2 we know that if the coefficients of a scheme satisfy equality (9), then conditions (i) and (ii) are reduced to

$$\sum_{h=1}^{H}a_h\sin^{2h}\frac{\theta}{2}\geqslant 0 \quad\text{and}\quad \sum_{h=0}^{H}d_h\sin^{2h}\frac{\theta}{2}-c_1I\geqslant 0$$

respectively. Furthermore, because a_h and d_h are real diagonal matrices, these inequalities can be reduced to some inequalities on scalar polynomials in $\sin^2\frac{\theta}{2}$. From Lemma 3 we know that if H is not too large, these conditions are further reduced to some inequalities on the coefficients of schemes $c_{j,i}$ and $b_{j,i}$, and the eigenvalues λ_m of A. In the following, we shall prove that the other conditions of Theorem 1 guarantee that condition (v) is fulfilled in a series of cases. Therefore, from Theorem 1, we shall obtain some stability criteria which are convenient in applications. Moreover, for constant coefficients, these conditions are necessary.

Theorem 3 For a horizontal three-point scheme with (9), if

(i)
$$a_1\geqslant 0,\quad a_1+a_2\geqslant 0, \tag{16}$$

(ii) there exists $c_1>0$ such that

$$\begin{cases} d_0-c_1I\geqslant 0,\\ 2(d_0-c_1I)+d_1+2(d_0-c_1I)^{1/2}(d_0-c_1I+d_1+d_2)^{1/2}\geqslant 0,\\ d_0-c_1I+d_1+d_2\geqslant 0, \end{cases} \tag{17}$$

and if conditions (iii) and (iv) of Theorem 1 are fulfilled, then the scheme is stable.

Proof According to Lemmas 2 and 3, conditions (i) and (ii) here guarantee that conditions (i) and (ii) of Theorem 1 are fulfilled. Therefore, the proof of this theorem is reduced to proving that condition (v) of Theorem 1 is fulfilled. In fact, we have the equality

$$\begin{aligned} V(\theta)&=R^*(\theta)N^*NR(\theta)-S^*(\theta)N^*NS(\theta)=G^*\Big(\sum_{h=1}^{2}a_h\sin^{2h}\frac{\theta}{2}\Big)G\\ &=G^*\Big(a_1\sin^2\frac{\theta}{2}\cos^2\frac{\theta}{2}+(a_1+a_2)\sin^4\frac{\theta}{2}\Big)G\\ &=[(1-e^{i2\theta})G]^*\frac{a_1}{16}[(1-e^{i2\theta})G]\\ &\quad+[(1-2e^{i\theta}+e^{i2\theta})G]^*\frac{a_1+a_2}{16}[(1-2e^{i\theta}+e^{i2\theta})G]. \end{aligned}$$

Moreover, condition (iii) guarantees that a_1 and a_2 satisfy the Lipschitz condition with respect to x and t, Therefore, condition (v) is fulfilled, i.e., the scheme is stable.

These schemes in [8]—[16] are horizontal three-point schemes, and satisfy equalities (9). Therefore, we can discuss their stability by using Theorem 3. We give the following corollaries.

Corollary 1 When the C-I-R scheme[8], the Lax scheme[9] and the Lax-Wendroff second-order scheme[13] (or the Richtmyer[14] and MacCormack[15] two-step L-W scheme) are applied to (1), if

(1) $$\|A\|\frac{\Delta t}{\Delta x}\leqslant 1, \tag{18}$$

i.e., for every element λ_m of A,

$$|\lambda_m|\frac{\Delta t}{\Delta x}\leqslant 1;$$

(2) $A(x, t)$ and $G(x, t)$ satisfy the Lipschitz condition, i.e., for every element f of A and G,

$$|f(x+\Delta x, t+\Delta t)-f(x, t)|\leqslant c_2(|\Delta x|+|\Delta t|), \tag{19}$$

and

$$|G|^2\geqslant\varepsilon\geqslant 0, \tag{20}$$

where ε is a positive constant.

then the schemes are stable.

Corollary 2 When the Wendroff[10] and Thomée[11] scheme is applied to (1), if

$$\|A\|\geqslant\varepsilon>0, \tag{21}$$

and (19) and (20) are fulfilled, then the scheme is stable.

Corollary 3 When the Keller-Thomée[12] scheme and the scheme

$$U_m^{k+1}-U_m^k+\frac{\Delta t}{\Delta x}A_m^{k+1/2}\cdot\frac{1}{4}(U_{m+1}^{k+1}-U_{m-1}^{k+1}+U_{m+1}^k-U_{m-1}^k)=0,$$

which is similar to the Crank-Nicolson[16] scheme, are applied to (1), if (19) and (20) are fulfilled, then the schemes are stable.

In order to prove these corollaries, we only need to prove that (16) and (17) are fulfilled. This is easy. In fact, from the coefficients $c_{j,k}$ and $b_{j,k}$ of these schemes, we may easily obtain a_1, a_2, d_0, d_1 and d_2 by using (12), (13), and (14), and prove immediately that (16) and (17) are fulfilled. We shall not give the proof for all schemes. In the following, taking the L-W scheme and the scheme similar to the Crank-Nicolson scheme as examples, we shall explain the procedure of the proof. For the L-W scheme, we have

$$\sum_j c_{j,0}A^j=0,\quad \sum_j c_{j,1}A^j=I,\quad \sum_j c_{j,2}A^j=0,$$

$$\sum_j b_{j,0}A^j=\frac{1}{2}\frac{\Delta t}{\Delta x}A+\frac{1}{2}\left(\frac{\Delta t}{\Delta x}A\right)^2,\quad \sum_j b_{j,1}A^j=I-\left(\frac{\Delta t}{\Delta x}A\right)^2,$$

$$\sum_j b_{j,2}A^j=-\frac{1}{2}\frac{\Delta t}{\Delta x}A+\frac{1}{2}\left(\frac{\Delta t}{\Delta x}A\right)^2.$$

According to (12),

$$e_{1,0}=1, \quad e_{1,1}=-2,$$
$$e_{2,0}=1, \quad e_{2,1}=-8, \quad e_{2,2}=8.$$

Therefore, using (13) and (14), we obtain

$$d_0=I, \quad d_1=0, \quad d_2=0,$$
$$a_1=0, \quad a_2=4\left[I-\left(\frac{\Delta t}{\Delta x}A\right)^2\right]\left(\frac{\Delta t}{\Delta x}A\right)^2,$$

i.e., (16) is reduced to (18), and (17) is always fulfilled. Thus, it is easy to prove that the conditions of Theorem 3 are fulfilled if (18)—(20) are fulfilled. This is the desired conclusion. For the scheme similar to the Crank–Nicolson scheme[16],

$$\sum_j c_{j,0}A^j=-\frac{1}{4}\frac{\Delta t}{\Delta x}A, \quad \sum_j c_{j,1}A^j=I, \quad \sum_j c_{i,2}A^j=\frac{1}{4}\frac{\Delta t}{\Delta x}A,$$
$$\sum_j b_{j,0}A^j=\frac{1}{4}\frac{\Delta t}{\Delta x}A, \quad \sum_j b_{j,1}A^j=I, \quad \sum_j b_{j,2}A^j=-\frac{1}{4}\frac{\Delta t}{\Delta x}A.$$

Moreover, by using (13) and (14), we obtain

$$d_0=I, \quad d_1=\left(\frac{\Delta t}{\Delta x}A\right)^2, \quad d_2=-\left(\frac{\Delta t}{\Delta x}A\right)^2,$$
$$a_1=0, \quad a_2=0.$$

Thus, (16) and (17) are always fulfilled, and it is easy to obtain the conclusion we want.

Lemma 4 For a horizontal five-point explicit scheme with (9), if the order of accuracy of the scheme is greater than 2, then $a_1=0$.

Proof The horizontal five-point explicit scheme can be rewritten in the following form:

$$U_{m+2+j}^{k+1}=S_0(A)U_m^k+S_1(A)U_{m+1}^k+S_2(A)U_{m+2}^k+S_3(A)U_{m+3}^k+S_4(A)U_{m+4}^k,$$

where j is equal to any one among -2, -1, 0, 1, and 2. Because the scheme is at least of the second order accuracy, we have the following relations:

$$\begin{cases} S_0+S_1+S_2+S_3+S_4=I, \\ -2S_0-S_1+S_3+2S_4=-\left(\dfrac{\Delta t}{\Delta x}A-jI\right), \\ 4S_0+S_1+S_3+4S_4=\left(\dfrac{\Delta t}{\Delta x}A-jI\right)^2. \end{cases}$$

These expressions can be rewritten as

$$\begin{cases} S_1=\dfrac{1}{2}\left[\left(\dfrac{\Delta t}{\Delta x}A-jI\right)^2+\left(\dfrac{\Delta t}{\Delta x}A-jI\right)\right]-3S_0-S_4, \\ S_2=I-\left(\dfrac{\Delta t}{\Delta x}A-jI\right)^2+3S_0+3S_4, \\ S_3=\dfrac{1}{2}\left[\left(\dfrac{\Delta t}{\Delta x}A-jI\right)^2-\left(\dfrac{\Delta t}{\Delta x}A-jI\right)\right]-S_0-3S_4. \end{cases} \tag{22}$$

In addition, according to (14) and (12), in the case here,

$$a_1=-2[-2(\tilde{S}_0\tilde{S}_1+\tilde{S}_1\tilde{S}_2+\tilde{S}_2\tilde{S}_3+\tilde{S}_3\tilde{S}_4)-8(\tilde{S}_0\tilde{S}_2+\tilde{S}_1\tilde{S}_3+\tilde{S}_2\tilde{S}_4)$$
$$-18(\tilde{S}_0\tilde{S}_3+\tilde{S}_1\tilde{S}_4)-32\tilde{S}_0\tilde{S}_4],\quad \tilde{S}_k=G S_k G^{-1}.$$

By using these expressions, it is easy to prove the conclusion of this lemma. In fact, putting (22) into the expression for a_1, we immediately know $a_1=0$.

From this lemma and the other results mentioned above, we obtain the following result:

Theorem 4 When the Rusanov[6] and the Burstein–Mirin[7] third–order schemes are applied to (1), if (18)—(20) and

$$4\left(\lambda_n\frac{\Delta t}{\Delta x}\right)^2-\left(\lambda_n\frac{\Delta t}{\Delta x}\right)^4\leqslant\omega\leqslant 3 \tag{23}$$

are fulfilled, then the schemes are stable. (ω is a parameter of the Rusanov[6] scheme.)

Proof The schemes are third order, horizontal five–point explicit schemes, and (9) is fulfilled, so $a_1=0$. Therefore, according to Lemma 2,

$$V(\theta)=G^*\left(a_2\sin^4\frac{\theta}{2}+a_3\sin^6\frac{\theta}{2}+a_4\sin^8\frac{\theta}{2}\right)G$$
$$=G^*\left\{\frac{1}{256}|1-2e^{i\theta}+e^{i2\theta}|^2[a_2|1+2e^{i\theta}+e^{i2\theta}|^2\right.$$
$$\left.+(2a_2+a_3)|1-e^{i2\theta}|^2+(a_2+a_3+a_4)|1-2e^{i\theta}+e^{i2\theta}|^2]\right\}G.$$

Moreover, it can be verified that for the schemes, only when

$$a_2\geqslant 0,\quad 2a_2+a_3\geqslant 0,\quad a_2+a_3+a_4\geqslant 0,$$

condition (4) can be fulfilled. Therefore, when conditions (i), (iii) and (iv) of Theorem 1 are fulfilled, condition (v) of Theorem 1 is also fulfilled, i.e., in order to apply Theorem 1, we only need to prove that (i)—(iv) are fulfilled. Obviously, if (19) and (20) are fulfilled, then (iii) and (iv) are fulfilled. For explicit schemes, (ii) is always fulfilled. Moreover (18) and (23) guarantee that (i) is fulfilled. Therefore, Theorem 1 can be applied, i.e., the schemes are stable.

Theorem 5 For a second–order or third–order explicit scheme with (9) and $H=3$, if

(i) $$a_2\geqslant 0,\quad a_2+a_3\geqslant 0, \tag{24}$$

and if conditions (iii) and (iv) of Theorem 1 are fulfilled, then the scheme is stable.

Proof By using Lemma 4 and the method for proving Theorem 3, the result can be obtained. The concrete proof is omitted.

3. *Conclusions*

From the above results, we see that in a series of cases, conditions (i)—(iv) of Theorem 1 guarantee that condition (v) of Theorem 1 is

fulfilled, i.e., if (i)—(iv) of Theorem 1 are fulfilled, then the schemes for pure-initial-value problems with variable coefficients are stable. Moreover, we also see that in a series of cases, if the von Neumann condition (4), condition (5) guaranteeing that the difference equations are well conditioned, and condition (iv) are fulfilled, and if G and A are smooth functions, then the schemes are stable. Condition (5) is always fulfilled for explicit schemes. In general, if A and G are smooth, then condition (iv) is fulfilled. Thus for a series of explicit schemes, if the von Neumann condition is fulfilled, and if G and A are smooth, then the schemes with variable coefficients are stable; for a series of implicit schemes, if these two conditions and condition (5) are fulfilled, then the schemes are also stable. Therefore, for a series of explicit schemes, the conditions of the papers [2] and [5] can be weakened. In [4], those results of [2] and [5] have been improved. The paper [4] also has discussed the stability of implicit schemes. However, the stability criterion in this paper seems to be more convenient. Moreover, this paper gives certain conditions which do not contain functions of θ, for example, conditions (16), (17) and (24). Therefore the results of this paper are more useful in practical applications.

References

[1] P. D. Lax, The scope of the energy method, *Bull. of the Amer. Math. Soc.*, **66** (1960), 1, 32—35.

[2] P. D. Lax, On the stability of difference approximations to solutions of hyperbolic equations with variable coefficients, *Comm. Pure Appl. Math.*, **14** (1961), 3, 497—520.

[3] P. D. Lax and B. Wendroff, On the stability of difference schemes, *Comm. Pure Appl. Math.*, **15** (1962), 4, 363—371.

[4] P. D. Lax and L. Nirenberg, On stability for difference schemes: A sharp form of Gårding's inequality, *Comm. Pure Appl. Math.*, **19** (1966), 4, 473—492.

[5] H.-O. Kreiss, On difference approximations of the dissipative type for hyperbolic differential equations, *Comm. Pure Appl. Math.*, **17** (1964), 3, 335—353.

[6] V. V. Rusanov, Third-order difference schemes for "through"-calculation of discontinuous solutions, D. A. N. USSR, **180** (1968), 6, 1303—1305 (in Russian).

[7] S. Z. Burstein and A. A. Mirin, Third order difference methods for hyperbolic equations, *J. Comp. Phys.*, **5** (1970), 3, 547—571.

[8] R. Courant, E. Isaacson and M. Rees, On the solution of nonlinear hyperbolic differential equations by finite differences, *Comm. Pure Appl. Math.*, **5** (1952), 3, 243—255.

[9] P. D. Lax, Weak solution of nonlinear hyperbolic equations and their numerical computation, *Comm. Pure Appl. Math.*, **7** (1954), 1, 159—193.

[10] B. Wendroff, On centered difference equations for hyperbolic systems, *J. Soc. Indust. Appl. Math.*, **8** (1960), 3, 549—555.

[11] V. Thomée, A stable difference scheme for the mixed boundary problem for a hyperbolic first order system in two dimensions. *J. Soc. Indust. Appl. Math.*, **10** (1962), 2, 229—245.

[12] H. B. Keller and V. Thomée, Unconditionally stable difference methods for mixed problems for quasi-linear hyperbolic systems in two dimensions, *Comm. Pure Appl. Math.*, **15** (1962), 1, 63—73.

[13] P. D. Lax and B. Wendroff, Systems of conservation laws, *Comm. Pure Appl. Math.*, **13**

(1960), 2, 217—237.

[14] R. D. Richtmyer, A survey of difference methods for nonsteady fluid dynamics, NCAR TN63-2, 1962.

[15] R. W. MacCormack, The effects of viscosity in hypervelocity impact cratering, AIAA Paper No. 69-354, 1969.

[16] J. Crank and P. Nicolson, A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type, Proc. Cambridge Philos. Soc., 43 (1947), Part I, 50—67.

[17] N. I. Ahiezer, Lectures on theory of approximation, Gostehizdat, Moscow, 1947 (in Russian).

Appendix 2

A Block-Double-Sweep Method for "Incomplete" Linear Algebraic Systems and Its Stability[1)]

Abstract

The block-double-sweep methods for linear algebraic systems have been discussed by many authors. However, sometimes we may meet a system which consists of many linear and a few nonlinear equations, and in which the number of unknowns in the linear equations is greater than the number of linear equations. In this situation, these linear equations constitute an "incomplete" system. This paper presents a block-double-sweep method which can be used to solve the "incomplete" systems of linear equations and which is stable under very weak conditions. In addition, this paper points out that for linear tridiagonal systems we can obtain the stability conditions which are weaker than the conditions obtained previously.

Introduction

Rusanov's paper [1] discusses a block-double-sweep method——a direct method for solving a system in the following form:

$$b_i x_i + a_{i+1} x_{i+1} = c_{i+1}, \quad i = 0, 1, \cdots, M-1, \tag{1}$$

$$\begin{cases} g_0 x_0 = d_0, \\ h_M x_M = d_M, \end{cases} \tag{2}$$

where b_i and a_{i+1} are $n \times n$-matrices; x_i and c_i are n-dimensional vectors; g_0 and h_M are $s \times n$- and $(n-s) \times n$-matrices respectively; d_0 and d_M are s- and $(n-s)$-dimensional vectors respectively; and $0 \leqslant s \leqslant n$. We may meet this type of system when solving differential equations numerically. In this case, equations (1) are difference equations approximating differential equations; and every equation of (2) is either a boundary condition for differential equations or a difference equation. Of course, this type of system can also appear in other problems. [2] discusses

1) This paper is an English translation of the paper in "Mathematicae Numericae Sinica, 1978, No. 3, 1—27".