Numerical Stabilizers and Computing Time for Second-Order Accurate Schemes*

B. EILON, D. GOTTLIEB, AND G. ZWAS

Department of Mathematical Sciences, Tel-Aviv University, Israel

Received May 17, 1971

This paper deals with explicit second-order accurate schemes for solving quasilinear hyperbolic equations with two spatial dimensions. The effect of certain stabilizing terms that allow a larger time step is studied and numerical examples are given, taking into account the simplicity of the schemes so as to shorten the actual computing time.

INTRODUCTION¹

The numerical solution of nonlinear initial value problems in more than one spatial dimension can cause severe problems of computing time, especially if second-order accuracy is desired.

We will first deal with hyperbolic systems of the form

$$W_t = A \cdot W_x + B \cdot W_y, \qquad (1)$$

where A and B depend on the components of W so that

$$A(W) W_x \equiv F_x, \qquad B(W) W_y \equiv G_y; \qquad (2)$$

in other words (1) is a system of conservation laws. The equations of compressible fluid dynamics are an example for such a system. When examining linear stability A and B will be taken as constant matrices. We will assume that A and B can be symmetrized by the same similarity transformation so that our system (1) is guaranteed to be hyperbolic.

• The basic explicit second-order accuracy scheme for solving (1) with given

* This research has been sponsored in part by the Air Force Office of Scientific Research (NAM) through the European Office of Aerospace Research AFSC, U. S. Aire Force, under contract F611052-69-C-0041.

¹ The computations reported herein were carried out on the CDC-6600 computer at the Tel-Aviv University Computation Center.

© 1972 by Academic Press, Inc.

initial data is the Lax-Wendroff method [6]. This method uses the fact that (1) and (2) lead to

$$W_{tt} = [A(F_x + G_y)]_x + [B(F_x + G_y)]_y.$$
(3)

The Lax–Wendroff scheme will be denoted by S_1 , namely,

$$W_{j,m}^{n+1} = S_1 W_{j,m}^n \,, \tag{4}$$

where $W_{j,m}^n = W(t_n, x_j, y_m)$. Using the notations

 $\sigma = \text{largest eigenvalue of } A \text{ or } B,$ $\lambda = \Delta t/h,$ and taking $h = \Delta x = \Delta y,$

the stability requirement for S_1 is given [6] by,

$$\lambda \leqslant 1/\sigma \sqrt{8}.$$
 (5)

In the special case where A and B commute, σ_1 replaces σ in (5) where

$$\sigma_1 = \max_i \left[(a_i^2 + b_i^2)/2 \right]^{1/2}, \tag{6}$$

and where a_i and b_i are the corresponding eigenvalues of A and B. We will refer in some cases to the situation where AB = BA knowing fully well that we lose generality but on the other hand we gain easier insight into the problems since we are able to apply the spectral mapping theorem.

We will also use the notation

$$\sigma_2 = \left[\frac{1}{2}\rho(A^2 + B^2)\right]^{1/2},\tag{6'}$$

where $\rho(A)$ and $\rho(B)$ are the spectral radii of A and B, and where, obviously, $\sigma_1 \leqslant \sigma_2 \leqslant \sigma$.

In addition to the maximal stable λ we will be interested in a simplicity number r, denoting the number of times that the functions F and G are computed for each net point. This number r is indeed some measure of the simplicity of a scheme, as pointed out by Strang [10], since F and G can be complicated computationally and since the quantity r/λ turns out to be significant as far as computing time is concerned. For the Lax-Wendroff scheme, S_1 , the number r is at least 15 (scalar case) and (9 + 6p) for a system of p equations. This is so because six matrix evaluations are needed in S_1 , and for a system of order p, such a matrix computation is roughly equivalent to p evaluations of the vector functions F or G.

EVEN STABILIZERS

In their paper [6], Lax and Wendroff suggested adding to their scheme a stabilizing term which is the centered finite-difference representation of

$$-(\lambda^2/8) h^4(A^2 + B^2) W_{xxyy}.$$
 (7)

This term allows a larger time step so that only

$$\lambda \leqslant [1/2\sigma_2] \tag{8}$$

is required for linear stability, namely a gain of at least $\sqrt{2}$ in Δt is achieved. Here again σ_1 can replace σ_2 if A and B commute. We will refer to terms like (7) as "even stabilizers" since they use derivatives of even orders, hence contributing only to the real part of the corresponding amplification matrix G. The scheme S_1 combined with (7) will be denoted by S_2 . The corresponding amplification matrices are (see [6]),

$$G_{1} = I + i\lambda(A\sin\xi + B\sin\eta) + \lambda^{2}[A^{2}(\cos\xi - 1) + B^{2}(\cos\eta - 1) - 1/2(AB + BA)\sin\xi\sin\eta],$$
(9)

$$G_{2} = G_{1} - 1/2(A^{2} + B^{2})(1 - \cos\xi)(1 - \cos\eta)\lambda^{4}.$$

Here A and B were taken to be locally constant and ξ and η are the dual variables after the usual Fourier transform. It should be noted that the stabilizer (7) uses only the same nine points that were already used by S_1 .

Let us now take the special case where AB = BA. By the spectral mapping theorem the eigenvalues of G_1 and G_2 , i.e., g_1 and g_2 , are given by

$$g_{1} = \left\{ 1 - \lambda^{2} \left[(1 - \cos \xi) a^{2} + (1 - \cos \eta) b^{2} + \frac{ab + ba}{2} \sin \xi \sin \eta \right] \right\}$$

+ $i\lambda \{a \sin \xi + b \sin \eta\},$ (10)
$$g_{2} = g_{1} - 1/2(a^{2} + b^{2})(1 - \cos \xi)(1 - \cos \eta) \lambda^{4},$$

where a and b are corresponding eigenvalues of A and B. In order to meet the well-known Von-Neumann condition we must impose $|g_1|^2 \leq 1$ and $|g_2|^2 \leq 1$ for all $|\xi|$; $|\eta| \leq \pi$. Now we observe that inequality

$$\{I_{M}[g_{1}]\}^{2} = \lambda^{2} \{a \sin \xi + b \sin \eta\}^{2} \leqslant 1 \qquad (|\xi|; |\eta| \leqslant \pi)$$

leads already to the stability requirement (8) with σ_1 replacing σ_2 . Consequently, since even stabilizers contribute only to the real part of the amplification matrix, we see that in a sense (7) is an optimal even stabilizer for S_1 . In order to further stabilize the Lax-Wendroff scheme, "odd stabilizers" are needed.

ODD STABILIZERS

If we take only the real part of the amplification matrix G_1 and impose the Von-Neumann condition, we find for the case where A and B commute,

$$\lambda \leqslant 1/\sqrt{2} \,\sigma_1 \,. \tag{11}$$

This means that an odd stabilizer that will yield the stability condition (11) when added to S_1 , is an optimal odd stabilizer in the above mentioned sense.

We now claim that the centered finite-difference representation of

$$(\lambda^{3}/4) h^{3}[(AB^{2} + B^{2}A) W_{xyy} + (BA^{2} + A^{2}B) W_{yxx}]$$
(12)

is such an optimal odd stabilizer.

We start the proof by observing that (12) does not damage the second-order accuracy and that the scheme S_2 with (12) as a stabilizer is again a regular nine-point scheme.

Now since (12) adds the term

$$-(i/2)\,\lambda^{3}[(AB^{2}+B^{2}A)\sin\xi(1-\cos\eta)+(BA^{2}+A^{2}B)\sin\eta(1-\cos\xi)] \qquad (13)$$

to G_1 , we obtain for the Von-Neumann requirement the inequality,

$$\begin{aligned} 4c^{2} \left\{ \alpha \sqrt{\left[\beta(1-\beta)\right]} + c\beta \sqrt{\left[\alpha(1-\alpha)\right]} \right\}^{2} \mu^{2} \\ + \left[4c^{2}\alpha^{2}\beta^{2} + (\alpha - c^{2}\beta)^{2}\right] \mu - \left[\alpha^{2} + c^{2}\beta^{2}\right] \leqslant 0, \end{aligned}$$
(14)

where

$$lpha = \sin^2(\xi/2);$$
 $eta = \sin^2(\eta/2);$
 $c = b/a;$ $\mu = \lambda^2 a^2.$

Again, we have assumed AB = BA so that the spectral-mapping theorem could be applied.

Inequality (14) is a convex parabola in μ having real roots of opposite signs and we want to show that for $\mu_0 = 1/(1 + c^2)$ the inequality still holds for all α ; $\beta \in [0; 1]$.

Substitution of μ_0 in (14) leads to

$$[(\alpha + \beta)/2]^{2}(1 + c^{2}) \ge \alpha\beta\{\alpha + c^{2}\beta + 2c[\alpha\beta(1 - \alpha)(1 - \beta)]^{1/2}\}$$
(15)

and of course it is enough to show that (15) holds for all α ; $\beta \in [0; 1]$ if $(\alpha\beta)$ replaces $[(\alpha + \beta)/2]^2$ on the left-hand side.

We are now left with the inequality

$$(1-\beta) c^{2} - 2[\alpha\beta(1-\alpha)(1-\beta)]^{1/2} c + (1-\alpha) \ge 0$$
 (16)

which is easily seen to be true for every real c and all α ; $\beta \in [0; 1]$, leading finally to the stability criteria (11) as claimed. The hyperbolicity of our system of equations guarantees that c is real.

We do not suggest the stabilizer (13) for practical computations but we feel it clarifies the role of stabilizing terms at least for second-order accurate schemes.

TWO-STEP SCHEMES

Since our main interest is shortening the real computing time we next mention the two step schemes, first suggested by Richtmeyer [7]. These schemes do not perform any matrix calculations for the case of conservation laws and are therefore considerably faster. In [7] Richtmyer gives the following scheme

$$W_{j,m}^{n+1} = \tilde{W}_{j,m}^{n} + (\lambda/2)(F_{j+1,m}^{n} - F_{j-1,m}^{n}) + (\lambda/2)(G_{j,m+1}^{n} - G_{j,m-1}^{n}),$$
(17)
$$W_{j,m}^{n+2} = W_{j,m}^{n} + \lambda(F_{j+1,m}^{n+1} - F_{j-1,m}^{n+1}) + \lambda(G_{j,m+1}^{n+1} - G_{j,m-1}^{n+1}),$$

where

$$\tilde{W}_{j,m}^n = 1/4(W_{j+1,m}^n + W_{j-1,m}^n + W_{j,m+1}^n + W_{j,m-1}^n)$$

the stability condition (for the case AB = BA) is

$$\lambda \leqslant 1/2\sigma_1 \,. \tag{18}$$

Scheme (17), which will be denoted by S_3 , advances the time by $2\Delta t$ while computing F and G only 8 times (i.e., $r = 4 \text{ per } \Delta t$). However, as will be seen later, in order to get results as accurate as with S_1 or S_2 it is necessary to take a finer mesh. The reason for this is probably the "diamond-shaped" numerical domain of dependence and the fact that S_3 does not use the nearest data, namely $W_{j\pm 1,m}$ and $W_{j,m\pm 1}^n$. In this sense S_3 is not a regular nine-point scheme.

There is also a possibility to consider a regular nine-point version of the original two-step method, namely the scheme S_4 (see [12])

$$W_{j+1/2,m+1/2}^{n+1/2} = \tilde{W}_{j+1/2,m+1/2}^{n} + (\lambda/2)(\hat{F}_{j+1,m+1/2}^{n} - \hat{F}_{j,m+1/2}^{n}) + (\lambda/2)(\hat{G}_{j+1/2,m+1}^{n} - \hat{G}_{j+1/2,m}^{n}),$$
(19)
$$W_{j,m}^{n+1} = W_{j,m}^{n} + \lambda(\hat{F}_{j+1/2,m}^{n+1/2} - \hat{F}_{j-1/2,m}^{n+1/2}) + \lambda(\hat{G}_{j,m+1/2}^{n+1/2} - \hat{G}_{j,m-1/2}^{n+1/2}),$$

where

$$ilde{W}^n_{j+1/2,m+1/2} = 1/4(W^n_{j+1,m+1} + W^n_{j+1,m} + W^n_{j,m+1} + W^n_{j,m}), \\ \hat{F}^n_{j+1,m+1/2} = F((W^n_{j+1,m+1} + W^n_{j+1,m})/2), \quad \text{etc.}$$

It turns out that the stability condition for this scheme, S_4 , is given (with AB = BA) by

$$\lambda \leqslant 1/\sqrt{2} \,\sigma_1 \tag{20}$$

and r = 8 per net point. It will be demonstrated later that S_4 produces errors much smaller than S_3 .

For the hydrodynamic equations where $AB \neq BA$ the allowed time steps are given by

$$\lambda \leqslant 1/[\sqrt{2} (|\mathbf{V}| + c)] \quad \text{for} \quad S_3$$

$$\lambda \leqslant 1/(|\mathbf{V}| + c) \quad \text{for} \quad S_4 ,$$
(21)

and

where c is the sound speed and V the fluid velocity (see [7] and [12]).

OPTIMAL-STABILITY SCHEMES

Second-order accurate schemes with optimal stability for solving twodimensional systems were proposed by Strang [9, 10], and multistep formulations for these schemes were constructed by Gourlay and Morris [4, 5]. These schemes are

$$S_5 = 1/2(L_x L_y + L_y L_x) \tag{22}$$

and

$$S_6 = L_{x/2} L_y L_{x/2}$$
 or $S_6' = L_{y/2} L_x L_{y/2}$, (23)

where L_x is a one-dimensional Lax-Wendroff operator in the x direction and L_y in the y direction. The stability condition is $\lambda \sigma \leq 1$ which is optimal for explicit schemes. S_6 even allows a time step according to $\lambda \leq \min[2/\rho(A); 1/\rho(B)]$ and this is of course an advantage especially when $\rho(A)$ is considerably larger than $\rho(B)$, however, it should be remembered that S_6 uses also $W_{j\pm 2,m}^n$. Similar results hold for S_6' .

As pointed out by Burstein [1], the Strang schemes have built-in "mixed stabilizers" which allow for the optimal stable λ . S_5 , for example, has a built-in mixed stabilizer containing (12) together with the even term

$$((\Delta t)^4/8) (A^2 B^2 + B^2 A^2) W_{xxyy}$$
.

The mixed stabilizer built-in S_6 is more complicated and includes terms with $(\Delta t)^5$ and $(\Delta t)^6$.

The simplicity of these schemes, in the sense of Strang, is represented by the fact that r = 16 for S_5 and r = 12 for S_6 , with the Gourlay-Morris formulation.

392

It is interesting to note that Gourlay and Morris [5], found a ratio of nearly 4/3 for the real computing times of S_5 and S_6 and that this is also the ratio of the corresponding r's. The real time ratio is influenced, of course, by the programming technique and by the specific computer used, but in general we see that r/λ is indeed a rough first estimate for a typical computing time.

The last scheme we would like to include in our comparison is influenced by the fact that the original Richtmyer scheme, S_3 , produces results of second-order accuracy only every second time step, i.e., after 8 evaluations of F and G. This scheme is also generated as a result of the disappointing results that the compounded S_6 scheme yielded in some cases (see [5]). This last scheme, S_7 , is a Strang-type scheme with optimal stability producing second-order results only every second time step, namely, we operate with $L_x L_y$ and then with $L_y L_x$ and so on, alternatively² (see [3]). The scheme S_7 computes F and G 8 times per Δt and has $r/\lambda = 8\sigma$, where again the Gourlay-Morris formulation should be used.

NUMERICAL RESULTS

In order to test and demonstrate the effects of the stabilizers and the simplicity of the schemes on the real computing time, we first examined the nonlinear equation

$$U_t + (U^3/3)_x + (U^5/5)_y = 0$$
with initial data $U(0, x, y) = (x + y)^{1/2}$
(24)

in the region $0 \le t \le t^*$, $1 \le x, y \le 2$. We did not use the square $0 \le x, y \le 1$, because we preferred the relative and the absolute errors to be approximately of the same magnitude. The analytic solution of (24) is

$$U = \{(1/2t)[((1+t)^2 + 4t(x+y))^{1/2} - (1+t)]\}^{1/2}$$
(25)

enabling us to compute a table of the errors for each time step for each scheme. The boundary values are taken from the exact solution. In a scalar case like (24) we have $\sigma_1 = \sigma_2 < \sigma$. As a yardstick we used the Lax-Wendroff scheme and fixed t^* so that S_1 will run for a thousand seconds producing errors as small as 10^{-6} . It turned out that h = 1/40 was the right grid for S_1 in order to have the above mentioned errors; these same t^* and h were used for all the other schemes except for S_3 where a finer grid (d = 1/60) had to be used since the errors were very much larger than Lax-Wendroff's. The programming took into account that there is no memory space for storing F and G for all net points at each cycle.

 $^{2}S_{7}$ is similar to the method of fractional time steps given by Marchuk in the SYNSPADE 1970 Proceedings, page 478.

°z	Scheme	Type of scheme	Real time in sec	r/À	Most of the errors	Maximal error	Remarks
-	S1	Lax-Wendroff	1000	$30 \sqrt{2} \sigma_1$	~10-₅	7 - 10 ⁻⁶	σ_1 replaces σ since $AB = BA$
7	S.	LW with even stabilizer	640	$30\sigma_1$	~10-"	9 · 10-°	σ_1 replaces σ_2 since $AB = BA$
3	S_3	Richtmyer, two step	102	8a1	$10^{-5} - 10^{-4}$	120 · 10 ⁻⁶	Greater errors near boundaries
4	S_3	Richtmyer, two step	340	801	$\sim 10^{-5}$	72 · 10-6	h = 1/60
5	S_4	Modified two step	158	$8 \sqrt{2} \sigma_1$	~10-6	$9 \cdot 10^{-6}$	
9	\mathcal{S}_{s}	Strang, $1/2(L_xL_y+L_yL_x)$	260	160	∼10 ^{-"}	6 · 10- ⁶	
7	S_6	Strang, $L_{x/2}L_yL_{x/2}$	201	12σ	~10-"	$5 \cdot 10^{-6}$	Uses $u_{j\pm 2,m}^{n}$
œ	S_6	Strang, $L_{y/2}L_xL_{y/2}$	206	120	$10^{-6} - 10^{-7}$	3 · 10-6	Uses $u_{j,m\pm 2}^{n}$
6	S_7	$L_x L_y$ and $L_y L_x$ alternatively	140	βσ	∼10-°	5 · 10 ⁻⁶	Second order at even cycles
10	S_7	$L_x L_y$ and $L_y L_x$ alternatively	128	80	$\sim 10^{-6}$	$6\cdot 10^{-6}$	<i>dr</i> computed at odd cycles only

TABLE I $(\sigma_1 = \sigma_2 < \sigma)$

EILON, GOTTLIEB, AND ZWAS

The results of the numerical solutions of problem (24) are given in Table I below, where the real computing times are for advancing the solution up to t^* . The grid size is taken as h = 1/40 except for the fourth run where h = 1/60 is used. The Δt 's are taken to be the maximal stable time steps at each cycle for each scheme.

Similar results were obtained for other problems. As a second example we took the equation $U_t + (U^2/2)_x + (U^2/2)_y = 0$ where $\sigma_1 = \sigma_2 = \sigma$ and where the solution for the initial data $(x + y)^{1/2}$ is $U = (t^2 + x + y)^{1/2} - t$. The results showed again that S_3 has by far larger errors than all the other schemes which for h = 1/40 had errors of $\sim 10^{-6}$.

The reason for this lies, probably, in the boundary treatment as well as in the "diamond-shaped" numerical domain of dependence. When using S_3 it seems that we do not have second-order accuracy near the boundaries. This is not the case in all the other schemes tested here, i.e., S_3 is much more sensitive to boundary effects. It is true that more appropriate treatment of the boundaries is possible pointed out by Gourlay and Morris [5], on the other hand, however, this sensitivity of S_3 is a serious disadvantage in practical calculations.

In hydrodynamical computations, for example, when calculating flows around bodies, boundary values are usually imposed on the body surface.

The advantage of S_4 over S_3 in this sense is interesting when observing that S_4 is a regular nine-point scheme which does not need boundary values at intermediate stages. When taking S_3 with the exact solution on a double boundary without using boundaries in the intermediate stages (usually, this cannot be achieved), the accuracy turned out to be as good as with S_4 .

The results of the second example are given in Table II.

TABLE II ($\sigma = \sigma_1 = \sigma_2$

				·			
Scheme	S_1	S_2	S_3	S_4	S_5	S_6	S_7
Real time in sec	1000	752	139	200	259	207	140
Typical errors	∼10-6	~10⁻6	$\sim 10^{-5}$	~10-6	~10-6	~10-6	~10⁻6

As Richtmyer pointed out [7], it is enough to use only half the mesh points when using S_3 , but as seen from Table II additional refinement is needed and the time is considerably longer than the times for S_7 and even S_4 . Our real time ratio between S_5 and S_6 is near the ratio of the corresponding r's namely 4/3, as was also found by Gourlay and Morris [5]. In general schemes S_7 and then S_4 and S_6 seem to be fast and accurate at least for problems with smooth solutions; they are also easy to program.

For the hydrodynamic equations in two spatial variables the time steps for S_7 and S_4 are given by

$$\lambda_n \leqslant 1/\max_{j,m} \left[|v_i| + c \right] \quad \text{for} \quad S_7 \tag{26}$$

and

$$\lambda_n \leqslant 1/\max_{j,m} \left[|\mathbf{v}| + c \right] \quad \text{for} \quad S_4 \,, \tag{27}$$

where c is the speed of sound and $|v_i|$ the largest component of the particle velocity v. If for some time t_n the denominator of (27) is maximal at a point where the velocity is in the x (or y) direction then (26) and (27) yield the same Δt_n .

THREE SPATIAL DIMENSIONS

The Lax–Wendroff scheme with and without even stabilizers can be easily extended to the three-dimensional case having stability conditions $\lambda \leq 1/(3\sqrt{3}\sigma)$ and $\lambda \leq 1/(3\sigma_2)$, respectively (see [11]).

Richtmyer's two-step method (with $\lambda \leq 1/3\sigma_1$) and its more compact version (with $\lambda \leq 1/\sqrt{3}\sigma_1$) can also be given for three dimensions (see [8, 2, 12]). In the class of optimal-stability methods (see [5, 3]) the extended S_7 scheme seems to be the fastest as pointed out by Gottlieb (scheme L_7 in [3]); i.e., use for example $L_x L_y L_z$ at odd time cycles and $L_z L_y L_x$ at even ones. This last scheme has $r/\lambda = 12\sigma$ and it is practical and efficient to use, at least when smooth solutions are involved.

A great deal of work is still to be done in handling multidimensional shockwave like discontinuities. For treating shocks it is desirable to have a simple smoothing operator that will act *automatically only near discontinuities* without being enormously complicated and time consuming. For schemes like S_5 , S_6 and S_7 such smoothing operators can be introduced one-dimensionally into L_x , L_y and L_z , as pointed out by Strang [10]. We hope to report on results in this direction in the near future.

References

- 1. S. Z. BURSTEIN, Numerical methods in multidimensional shocked flows, AIAA J. 2 (1964), 2111-2117.
- 2. B. EILON, A note concerning the two step Lax-Wendroff method in three dimensions, to appear in *Math. of Comp.*, also TAU Math. report MS-7014.
- 3. D. GOTTLIEB, Strang type difference schemes for multidimensional problems, to appear in SIAM Journal on Num. Anal.

- 4. A. R. GOURLAY AND J. L. MORRIS, A multistep formulation of the optimized Lax-Wendroff method for nonlinear hyperbolic systems in two space variables, *Math. Comp.* 22 (1968), 715-719.
- 5. A. R. GOURLAY AND J. L. MORRIS, On the comparison of multistep formulations of the optimized Lax-Wendroff method for nonlinear hyperbolic systems in two space variables, *Comp. Phys.* 5 (1970), 229-243.
- 6. P. D. LAX AND B. WENDROFF, Difference schemes for hyperbolic equations with high order of accuracy, *Comm. Pure Appl. Math.* 17 (1964), 381–398.
- 7. R. D. RICHTMYER, "A survey of difference methods for nonsteady fluid dynamics," N.C.A.R. Tech. Notes 63-2, 1963.
- E. L. RUBIN AND S. PREISER, Three dimensional second order difference schemes for discontinuous flows, *Math. Comp.* 24 (1970), 57–63.
- 9. G. STRANG, Accurate partial difference methods—nonlinear problems, *Numer. Math.* 13 (1964), 37–46.
- 10. G. STRANG, On the construction and comparison of difference schemes, SIAM J. Numer. Anal. 5 (1968), 506-517.
- 11. G. Zwas, Stability conditions for three dimensional Lax-Wendroff schemes, to appear, TAU Math. Report.
- 12. G. ZWAS, On two step Lax-Wendroff methods in several dimensions, to appear.