On the Instability of Leap-Frog and Crank-Nicolson Approximations of a Nonlinear Partial Differential Equation

By B. Fornberg

Abstract. It is well known that nonlinear instabilities may occur when the partial differential equations, describing, for example, hydrodynamic flows, are approximated by finitedifference schemes, even if the corresponding linearized equations are stable. A scalar model equation is studied, and it is proved that methods of leap-frog and Crank-Nicolson type are unstable, unless the differential equation is rewritten to make the approximations quasi-conservative. The local structure of the instabilities is discussed.

1. Introduction. There are many theorems, based for example on Fourier or energy methods, which can be used to find precise stability conditions for difference approximations of linear partial differential equations with constant coefficients. By stability, we mean that the L_2 -norm of the difference approximation does not increase in time faster than a fixed exponential function even if the mesh is refined. Many of the results for equations with constant coefficients can be carried over to the case of variable coefficients. It is often sufficient to freeze the coefficients and consider only the local stability properties to get an estimate of the over-all stability. However, as yet very little has been proved about the stability for approximations of nonlinear equations. It turns out that the properties of the linearized equations are not at all sufficient for determining stability.

The first example of pure nonlinear instability was given by N. A. Phillips [4] for a difference approximation of the barotropic vorticity equation for two-dimensional flow. Richtmyer [5] gives another example, which can also be found in Richtmyer-Morton [6]. It is very similar to that of Phillips, but here a model equation is studied:

(1)
$$\frac{\partial u}{\partial t} + \frac{1}{2} \frac{\partial u^2}{\partial x} = 0$$

A solution for the leap-frog approximation of (1) is found, which increases exponentially with the number of time steps. In this paper, we shall also study approximations of (1), which, more generally, can be rewritten as

(2)
$$\frac{\partial u}{\partial t} + \frac{\theta}{2} \frac{\partial u^2}{\partial x} + (1 - \theta)u \frac{\partial u}{\partial x} = 0,$$

for any value of the (real) parameter θ . This equation is approximated by finitedifference equations of leap-frog or Crank-Nicolson type. We use the difference

Copyright © 1973, American Mathematical Society

Received June 2, 1971, revised February 22, 1972.

AMS (MOS) subject classifications (1970). Primary 65M10.

Key words and phrases. Nonlinear instability, leap-frog scheme, Crank-Nicolson scheme.

approximations for $\partial/\partial x$, which make the corresponding approximation for the linear equation

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0$$

energy-conserving. In Lemma 2, we prove that this means that the difference operator for space derivatives is antisymmetric. The main purpose of this paper is to add to the understanding of nonlinear instability by showing that, under a slight restriction, a necessary and, for Crank-Nicolson also a sufficient, condition for stability in the nonlinear case is that $\theta = 2/3$. We shall also discuss the effect of adding smoothing operators to improve stability. The discussion in Section 4 indicates that instabilities occur only when the solution alternates around zero.

2. Notations and Preliminaries. The mesh sizes in the x- and t-directions are denoted by h and k, respectively, and their ratio by $\lambda = k/h$. The translation operator E is defined by Eg(x) = g(x + h), and by Q we denote a difference operator of the form $Q = \sum_{i=-p}^{q} a_i E^i$ which approximates $\partial/\partial x$. Since we are only concerned with real solutions to Eq. (2), we restrict ourselves in the following to operators where the a_i 's are real numbers. The centered, forward and backward difference approximations, D_0 , D_+ and D_- of $\partial/\partial x$, respectively, are defined by $2hD_0 = E - E^{-1}$, $hD_+ = E - I$ and $hD_- = I - E^{-1}$.

The symbol or the Fourier transform of Q is the analytic function $\hat{Q}(\omega h) = \sum_{i=-p}^{q} a_i e^{i i \omega h}$.

The solution of the difference approximation in the point x, t is denoted by v(x, t). By $\{s(x)\}$, we denote the doubly infinite sequence $\cdots s(x - h) s(x)$ $s(x + h) \cdots$. If T is an operator (or a constant), $T\{s(x)\}$ means $\{Ts(x)\}$.

LEMMA 1. If Re $\hat{Q} = 0$ for all real values of ωh , then Q can be written as $Q = D_0 P$, where

$$P = \sum_{\nu=0}^{N} \alpha_{\nu} h^{2\nu} (D_{+} D_{-})^{\nu}.$$

Proof of Lemma 1. Let $z = e^{i\omega \hbar}$, ω real. Then $\hat{Q}(\omega \hbar) = p(z) = a_N z^N + a_{N-1} z^{N-1}$ + \cdots + $a_{-N} z^{-N}$. By assumption, all coefficients a_ν are real and $\operatorname{Re}(p(z)) = 0$. Hence, p(1) = p(-1) = 0, which shows that (z - 1/z) can be factored out, i.e., $p(z) = (z - 1/z) \cdot q(z)$. Let $q(z) = b_{N-1} z^{N-1} + \cdots + b_{1-N} z^{1-N}$, b_ν real. We want to show that $b_\nu = b_{-\nu}$, $\nu = 1, 2, \cdots, N-1$. Observing that $\operatorname{Re}(z - 1/z) = 0$ for |z| = 1, it follows that $\operatorname{Im}(q(z)) = 0$. We can write q(z) in the form q(z) = r(z) + s(z), where

$$r(z) = c_{N-1}z^{N-1} + \cdots + c_1z + c_0$$

is a polynomial with real coefficients and s(z) is a sum of cosines,

$$s(z) = 2b_{-1} \cos \omega h + \cdots + 2b_{1-N} \cos(N-1)\omega h.$$

Hence, Im s(z) = 0 and, thus, Im r(z) = 0 for z with |z| = 1. Im r(z) is a harmonic function, and since it is bounded inside the unit circle it must be identically zero. Therefore, the equation r(z) = i is not solvable, which contradicts the fundamental theorem of algebra unless r(z) is a constant. This proves the lemma.

LEMMA 2. Let Q be some difference approximation of $\partial/\partial x$ such that the Crank-Nicolson scheme

(4)
$$(I + kQ/2)v(x, t + k) = (I - kQ/2)v(x, t)$$

or the leap-frog approximation

(5)
$$v(x, t+k) - v(x, t-k) = 2kQv(x, t),$$

is L_2 -energy-conserving for the linear equation (3). Then, Q can be written in the form $Q = D_0 P$, where

$$P = \sum_{\nu=0}^{N} \alpha_{\nu} h^{2\nu} (D_{+} D_{-})^{\nu}$$

Proof of Lemma 2. The Crank-Nicolson equation (4) is energy-conserving if and only if the absolute value of the Fourier transform of the implicit operator $(I + kQ/2)^{-1}(I - kQ/2)$ is equal to one for all real ωh , i.e.,

$$\left|\frac{1-k\hat{Q}/2}{1+k\hat{Q}/2}\right| = 1$$

If we write $k\hat{Q}/2$ as a + ib (a and b real) and square, we get $(1 + a)^2 + b^2 = (1 - a)^2 + b^2$, i.e., a = 0 and \hat{Q} is pure imaginary. Lemma 1 can now be used.

Let $v(x, t) = \xi^{t/k} e^{i\omega x}$ in the leap-frog approximation (5). We get

$$\xi^2 - 2k\hat{Q}(\omega h)\xi - 1 = 0.$$

In order that the method be energy-conserving (or even stable, since the roots cannot depend on k or h individually), it is necessary that no root of this equation has an absolute value greater than one. By the relations between roots and coefficients, the product of the roots is -1, and, hence, they are of the form $e^{i(\pi/2\pm\eta)}$, where η is real. Their sum is then $2i \cos \eta$, i.e., \hat{Q} is pure imaginary. The result now follows from Lemma 1.

3. Main Theorem.

THEOREM. Let Q be of the same form as in Lemma 2. It is now used to approximate $\partial/\partial x$ in the Crank-Nicolson scheme

(6)
$$v(x, t + k) + \frac{\theta k}{4} Qv(x, t + k)^{2} + \frac{k(1 - \theta)}{2} v(x, t + k) Qv(x, t + k)$$
$$= v(x, t) - \frac{\theta k}{4} Qv(x, t)^{2} - \frac{k(1 - \theta)}{2} v(x, t) Qv(x, t)$$

and the leap-frog approximation

(7)
$$v(x, t + k) - v(x, t - k) + k \theta Q v(x, t)^2 + 2k(1 - \theta)v(x, t)Q v(x, t) = 0$$

of Eq. (2). If P satisfies $P\{r(x)\} \neq \{0\}$, where $\{r(x)\}$ is the sequence

(8)
$$\{r(x)\} = \cdots = 0 - \epsilon \epsilon = 0 - \epsilon \epsilon = 0 \cdots , \quad \epsilon \neq 0,$$

then $\theta = 2/3$ is a necessary condition for stability. For the Crank-Nicolson scheme (6), $\theta = 2/3$ is also a sufficient condition for stability.

Proof of the Theorem. Assuming that v(x, t) in Eq. (6) has the form

$$v(x, t) = r(x) \cdot a(t)$$

and using $Q = D_0 P$, we obtain, after some rearrangement of terms,

$$r(x) \cdot (a(t+k) - a(t)) + (a(t+k)^2 + a(t)^2) \\ \cdot \left[\frac{\lambda\theta}{8} \cdot P(r(x+h)^2 - r(x-h)^2) + \frac{\lambda(1-\theta)}{4} \cdot r(x) \cdot P(r(x+h) - r(x-h))\right] = 0.$$

This equation is satisfied, for example, if a(t) and r(x) satisfy the following recursion relations

(9)
$$a(t+k) - a(t) = C \frac{\lambda}{8} (a(t+k)^2 + a(t)^2),$$

(10)
$$Tr(x) = -C \cdot r(x),$$

where

(11)
$$Tr(x) = 2(1 - \theta)r(x)P(r(x + h) - r(x - h)) + \theta P(r(x + h)^2 - r(x - h)^2)$$

and $C \neq 0$ is some constant.

The approximation is certainly not stable if we can find a bounded sequence which satisfies (10). By starting the recursion (9) with plus or minus one, the sequence a(t) and, hence, the solution v(x, t) will eventually explode. A rough estimate of the rate of divergence can be obtained by observing that Eq. (9) can be regarded as a difference approximation of the ordinary differential equation

 $a'(t) = (C/4h)a(t)^2,$

 $a(0) = \pm 1$, the sign to be chosen such that $C \cdot a(0) > 0$.

The solution of this equation diverges to infinity already at the finite time t = 4h/C. A suitable choice for the solution of (10) turns out to be the sequence in Eq. (8):

$$\{r(x)\} = \cdots = 0 - \epsilon \epsilon = 0 - \epsilon \epsilon = 0 \cdots, \quad \epsilon \neq 0.$$

We find that $P\{r(x)\} = K_P\{r(x)\}$, where K_P is a constant factor depending only on *P*. To see this, we observe that $\{r(x)\}$ has period 3 and that *P* is symmetric. The values of $P\{r(x)\}$ at r(x) = 0 and $\pm \epsilon$ have to be calculated. The value at r(x) = 0is zero, while the other two values differ only by sign. Thus we find

$$\{r(x+h)^2-r(x-h)^2\} = \cdots 0 \epsilon^2 - \epsilon^2 0 \epsilon^2 - \epsilon^2 0 \cdots,$$

i.e.,

$$\{P(r(x + h)^{2} - r(x - h)^{2})\} = -\epsilon K_{P}\{r(x)\}$$

and

$$\{P(r(x + h) - r(x - h))\} = \{Pr(x + h)\} - \{Pr(x - h)\}$$
$$= \cdots -2\epsilon K_P \epsilon K_P \epsilon K_P - 2\epsilon K_P \epsilon K_P \epsilon K_P - 2\epsilon K_P \cdots$$

Hence,

$$\{r(x) \cdot P(r(x+h) - r(x-h))\} = \cdots 0 - \epsilon^2 K_P \epsilon^2 K_P 0 - \epsilon^2 K_P \epsilon^2 K_P 0 \cdots$$
$$= \epsilon \cdot K_P \{r(x)\}.$$

Therefore,

$$T\{r(x)\} = 2(1 - \theta) \cdot \epsilon \cdot K_P\{r(x)\} - \theta \cdot \epsilon \cdot K_P\{r(x)\}$$
$$= (2 - 3\theta) \cdot \epsilon \cdot K_P\{r(x)\}$$

which satisfies (10) if we choose the constant C to be

(12)
$$C = -(2 - 3\theta) \cdot \epsilon \cdot K_P$$

According to the assumptions, $K_P \neq 0$ and $\epsilon \neq 0$. Thus, $C \neq 0$ if $\theta \neq 2/3$, and the proof of this part is complete.

We shall now prove that $\theta = 2/3$ is sufficient for stability. With $\theta = 2/3$, Eq. (6) can be written as

(13)
$$v(x, t+k) + \frac{k}{6}Rv(x, t+k) = v(x, t) - \frac{k}{6}Rv(x, t),$$

where R denotes the operator

 $Rv(x, t) = D_0 Pv(x, t)^2 + v(x, t) D_0 Pv(x, t).$

Squaring and summing both sides gives

$$||v(\cdot, t+k)||^{2} + \frac{k^{2}}{36} ||Rv(\cdot, t+k)||^{2} - \left[||v(\cdot, t)||^{2} + \frac{k^{2}}{36} ||Rv(\cdot, t)||^{2} \right]$$
$$= -\frac{k}{3} (v(\cdot, t+k), Rv(\cdot, t+k)) - \frac{k}{3} (v(\cdot, t), Rv(\cdot, t)).$$

Hence, $(v(\cdot, t), Rv(\cdot, t)) = 0$ is sufficient for L_2 -stability, i.e.,

$$\left(v(\cdot, t), D_0 \sum_{\nu=0}^N \alpha_{\nu} h^{2\nu} (D_+ D_-)^{\nu} v(\cdot, t)^2 + v(\cdot, t) D_0 \sum_{\nu=0}^N \alpha_{\nu} h^{2\nu} (D_+ D_-)^{\nu} v(\cdot, t) \right) = 0.$$

This follows immediately from repeated partial summation.

We now turn to the leap-frog case. Assuming again that

$$v(x, t) = r(x) \cdot a(t),$$

Eq. (7) gives

$$r(x) \cdot (a(t+k) - a(t-k)) + a(t)^{2}$$
$$\cdot \left[\frac{\lambda \theta}{2} P(r(x+h)^{2} - r(x-h)^{2}) + \lambda(1-\theta)r(x)P(r(x+h) - r(x-h))\right] = 0$$

which is satisfied if

(14)
$$a(t+k) - a(t-k) = C \frac{\lambda}{2} a(t)^2$$

(15)
$$Tr(x) = -C \cdot r(x),$$

where T is to be understood as defined in Eq. (11). The proof now proceeds exactly as in the Crank-Nicolson case, since Eqs. (9) and (14) are equivalent and (10) and (15) are equal.

Finally, we observe that this main theorem may be generalized to be valid for the corresponding approximations of $\partial u/\partial t + (\theta/n)\partial u^n/\partial x + (1 - \theta)u^{n-1}(\partial u/\partial x) = 0$, where *n* is an even positive integer. All parts of the proof follow through exactly in the same way. $\theta = 2/3$ generalizes to $\theta = n/(n + 1)$.



– – First time level L L L DIAGRAM 1. Crank-Nicolson with $\theta = 1, \lambda = 1/4, P = I$. two elements which are -0.1 and 0.1, respectively.

v(x,t) 0.6 -





DIAGRAM Z. Ferioaic Crank-Nicolson with v = 1, $\lambda = 1/4$, F = 1, ----3 in the diagram, are random numbers between -0.1 and 0.1.



v(x,t) 0.5 .

0.4

0.3



LEAP-FROG AND CRANK-NICOLSON APPROXIMATIONS

53

v(x,t) 1.2

٠

4. Numerical Results, Conclusions. In Fornberg [2], the behavior of the leapfrog scheme for the case P = I = identity operator and $\theta = 1$ was studied. The choice of $\theta = 1$ is natural because it corresponds to a conservation form. It gives rise to comparatively few arithmetic operations and conserves the sum of the solution over the mesh points for successive time levels. This holds even for more general equations than (1). The conservation of the sum is important, for example, when, in a practical computation, a correct shock speed is wanted. Although the pattern (8) seemed to be the one which gave the strongest divergence, we found in the paper [2] that explosions more often were characterized by locally raising and wandering spikes in schemes with values around zero. When a certain local combination of values happens to appear in the solution, this process starts, and waves of the new and bigger amplitude spread over the mesh. This may repeat on a larger scale when the critical combination of values again happens to occur. (See Diagrams 1 and 2. Here the Crank-Nicolson scheme is used instead of the leap-frog scheme, but the diagrams are almost identical to those for the leap-frog scheme in [2].) Soon even the linear stability limits are passed. If all values of the solution have the same sign and are not too close to zero, the stability seems perfect, as is indicated in Diagram 3. Similar test runs for $\theta = 0$ and P = I and for $\theta = 0$ or 1 and $P = I - (1/6)h^2 D_+ D_-$ (corresponding to fourth-order accuracy in space) show the same stability. Probably, this property of stability of solutions bounded away from zero is quite general. Of course, the solution must not be so far away from zero that the linearized stability condition is violated. In [2], $(\theta = 1, P = I)$ we also studied the influence of a boundary condition equal to zero, and proved that it was sufficient for making the scheme divergent. Theoretically, there exist bounded solutions even in this case, but a small number, in practice only one, of small perturbations made the solution divergent.

From Eq. (2), we can form a differential-difference equation

(16)
$$\frac{\partial v(x, t)}{\partial t} + \frac{\theta}{2} Q v(x, t)^2 + (1 - \theta) v(x, t) Q v(x, t) = 0,$$

where only the x-direction is made discrete. For $\theta = 2/3$, it is quasi-conservative in the sense that

(17)
$$(\partial/\partial t) ||v(x, t)||^2 = 0,$$

which is arrived at by taking the scalar product of (16) with v(x, t) and using partial summation $(Q = D_0 P, P \text{ symmetric})$. In spite of (17), the approximation (7) with $\theta = 2/3$ diverges, as is proved by Kreiss and Oliger [3] and is also seen from Diagram 4, in both cases for P = I. However, this instability is not as serious as the one we obtained for $\theta \neq 2/3$ and is also of a somewhat different nature. In the same way as (17) was derived from (16), we find that the scalar product between two successive time levels for the scheme (7) is invariant in time. This shows that (7) cannot diverge, unless v(x, t) changes considerably between successive time levels. This was not necessary for divergence when $\theta \neq 2/3$.

In order to increase the chance for stability and especially to avoid divergence for $\theta \neq 2/3$ from the pattern $\{r(x)\}$ defined in Eq. (8), it is natural to apply smoothing to the difference scheme. Since $\{r(x)\}$ is just multiplied by a constant K_s if a smoothing operator, which replaces the values v(x, t) at some time level by

$$Sv(x, t) = \sum_{\nu=0}^{M} \beta_{\nu} h^{2\nu} (D_{+} D_{-})^{\nu} v(x, t)$$

is applied, it can easily be estimated how much smoothing is necessary for keeping this special pattern from increasing. A smoothing of this type can be applied, for example, to the new values obtained when a time step is completed, or we can apply the smoothing operator to the earliest time level involved when the solution on a new time level shall be evaluated. For S to be a reasonable smoothing operator, K_s should be slightly less than one. It follows from Eqs. (9) and (14) that, apart from terms which are $O(\epsilon^2)$, the amplitude of $\{r(x)\}$ (or $-\{r(x)\}$) in a single time step is increased by the factor

$$1 + |C| \cdot \frac{\lambda}{4} = 1 + \frac{1}{4} \cdot K_P \cdot \lambda \cdot |2 - 3\theta| \cdot |\epsilon|.$$

At the same time, smoothing reduces it by the factor K_s . To prevent our special solution from diverging, we choose S such that K_s satisfies

(18)
$$K_{s} \leq 1 - \frac{1}{4} \cdot K_{P} \cdot \lambda \cdot |2 - 3\theta| \cdot |\epsilon|.$$

T.

If we apply smoothing to the earliest involved time level at each step, (18) can be used for two-level schemes, but the factor 1/4 has to be changed to 1/2 for three-level schemes, for example for the leap-frog scheme.

The estimate in Eq. (18) on smoothing, which is necessary to keep our special solution bounded, has been tested for both Crank-Nicolson and leap-frog equations. Choosing $\theta = 1$, $\epsilon = 0.1$, $\lambda = 1/4$ and the operators P = I, $S = I + \alpha \cdot (D_+D_-)$, i.e., $K_P = 1$, $K_S = 1 - 3\alpha$ and applying S to the earliest involved time level in every step, Eq. (18) suggests $\alpha = 1/480$ for Crank-Nicolson and $\alpha = 1/240$ for leap-frog. The amplitudes without smoothing are also calculated in the table below.

Method	α	0	50	100	140	150	159	163	170	300
C-N	1/480 0	0.100 0.100	0.100 0.144	0.100 0.262	0.099 0.763	0.099 1.460	0.099 9.486	0.099	0.099	0,097
l-f	1/240 0	0.100 0.100	0.100 0.143	0.100 0.260	0.100 0.743	0.100 1.381	0.100 5.684	0.100 158.2	0.100 >10 ¹⁶⁷	0.100 >10 ^{10*1}

Number of Time Steps

The use of the Crank-Nicolson scheme involves, in every time step, the solving of a system of nonlinear equations. In the example above, this could not be done beyond the 159th time step.

It is not quite clear whether the dissipation as a function of the original amplitude, as given by Eq. (18), is sufficient for stability in the case of more general initial values. In the experiment reported in Diagram 5, random initial values between -0.1 and 0.1 on the first two time levels were used to start a leap-frog scheme with $\theta = 1$, $\lambda = 1/4$ and $\alpha = 1/240$. The norm of the solution remained almost unchanged during the last 2000 time steps.



v(x,t) 0.3 -

It is probable that the best way to proceed in a practical case is to use a quasiconservative scheme and add only a small dissipative term. For the real hydrodynamic equations, quasi-conservative schemes are well known (A. Arakawa [1]). To minimize the number of arithmetic operations, and increase the chance for a correct shock speed, if shocks are present, the choice of $\theta = 1$ might be preferable, especially if the solution is not close to zero. There seems to be no great danger in choosing $\theta \neq 2/3$ provided we add a rather small dissipative term, for example, of a size suggested by (18), in critical regions. Kreiss and Oliger [3] have shown that in the case $\theta = 0$, $P = I - (h^2/6)D_+D_-$, it is sufficient to add a small dissipative term, such as $(h^4/6)D_+D_- |D_0v| D_+D_-v$, to satisfy the condition $(\partial/\partial t) ||v(x, t)||^2 \leq 0$, and test runs carried out by them with leap-frog type time differencing indicate stability around zero.

Finally, we observe that the transformation of the independent variables

$$x' = x + at, \qquad t' = t,$$

where a is a constant, transforms Eq. (2) into

(20)
$$\frac{\partial(a+u)}{\partial t'} + \frac{\theta}{2} \frac{\partial(a+u)^2}{\partial x'} + (1-\theta)(a+u) \frac{\partial(a+u)}{\partial x'} = 0.$$

This equation has the same form as (2), but the solution has been translated by the constant *a*. In this way, the solution can be bounded away from zero, if it originally was around zero, and stability problems can be avoided. Actually, however, this type of transformation is used in meteorological applications in the opposite direction, i.e., to eliminate a constant flow and then calculate only the remaining oscillations around zero. As we have seen above, this makes the introduction of a dissipative mechanism necessary.

University of Uppsala Uppsala, Sweden

1. A. ARAKAWA, "Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. I," J. Computational Phys., v. 1, 1966, pp. 119–143.

2. B. FORNBERG, A Study of the Instability of the Leap-Frog Approximation of a Non-Linear Differential Equation, Report NR 22, June 1969, Department of Computer Sciences, Uppsala University.

3. H.-O. KREISS & J. OLIGER, Comparison of Accurate Methods For the Integration of Hyperbolic Equations, Report NR 36, October 1971, Department of Computer Sciences, Uppsala University.

4. N. A. PHILLIPS, "An example of non-linear computational instability," *The Atmosphere* and the Sea in Motion, Edited by B. Bolin, 1959, Rockefeller Institute, New York, pp. 501-504.

5. R. D. RICHTMYER, A Survey of Difference Methods for Non-Steady Fluid Dynamics, NCAR Technical Note 63-2, National Center for Atmospheric Research, Boulder, Colorado, 1962, pp. 16–19.

6. R. D. RICHTMYER & K. W. MORTON, Difference Methods for Initial-Value Problems, 2nd ed., Interscience Tracts in Pure and Appl. Math., no. 4, Interscience, New York, 1967, pp. 128–130. MR 36 #3515.