JOURNAL OF COMPUTATIONAL PHYSICS 51, 83–106 (1983)

Focusing: A Mechanism for Instability of Nonlinear Finite Difference Equations

WILLIAM L. BRIGGS,* ALAN C. NEWELL,† AND TALIB SARIE*

*Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam, New York 13676 and †Department of Mathematics, University of Arizona, Tucson, Arizona 85721

Received January 12, 1982; revised December 7, 1982

A mechanism for the destabilization of numerical algorithms for partial differential equations is suggested. The novelty of the work is that it attempts to explain the dynamical process by which noise can localize on a spatial grid and cause finite amplitude instability thresholds to be exceeded at distinct locations.

I. INTRODUCTION AND GENERAL DISCUSSION

The stability of partial difference equations which arise in the discretization of time dependent differential equations is well understood for linear problems with constant coefficients. Progress has also been made in studying linear, variable coefficient problems. However, once nonlinear terms are introduced into difference equations, there are few general statements which can be made and global results are available only for isolated cases [2, 4, 9]. Except for the pioneering work of Phillips [9], and Arakawa and his colleagues [1, 6], very little work has gone into analyzing the nature of instabilities in the way that fluid mechanicists investigate instabilities as they occur in the transition to turbulence. It is the goal of this paper to make such an investigation for a class of nonlinear finite difference equations which are typified by the leapfrog (second order) method applied to the quasi-linear equation

$$u_t + uu_x = 0.$$

What we find is a totally new and very subtle mechanism for the triggering of nonlinear instabilities. It is insidious and at first very slow to develop. As a certain threshold is reached, however, sudden outbursts of unbounded noise occur at various *local* positions in the spatial grid. The mechanism is dynamic in character and does not necessarily rely on large initial perturbations or on a large flow of energy into the high wave numbers. It makes its appearance in schemes which are energy conserving and neutrally stable over short time scales. It is a mechanism which is universal in character and closely related to the mechanisms responsible for the breakdown of monochromatic gravity waves on the sea surface, Langmuir turbulence in plasmas, and the intense laser beams seen in nonlinear dielectrics [10]. Our goal is to

0021-9991/83 $3.00

581/51/1-7

understand the nature of this mechanism and to develop from this understanding plausible criteria for the surgical application of various remedies which are necessary to suppress instability and sustain computations over long times. In particular we will discuss, in the context of the example used in this paper, ways in which one might judiciously choose the frequency at which one must apply these remedies.

In order to develop a feeling for how this instability arises, we first recount the ideas on which nonlinear stability theory is usually based. It is natural to decompose the field $u(m, n)$ (where $t = m\Delta t$ and $x = n\Delta x$) into components $U(m, n)$ and $u'(m, n)$, where $U(m, n)$, the approximation to the exact solution, changes slowly with respect to the grid length Δx and $u'(m, n)$, the noise, consists of a small number N_1 of low period (high wavenumber) modes which are small integer multiples of the grid length. This sort of decomposition is chosen because (i) it is known that the potentially most unstable modes have wavelengths on the scale of the grid (often linear stability analysis can suggest which modes to include) and (ii) it is desirable to reduce the dimensionality of the problem from N, the number of grid points which is generally large, to N_1 which is much smaller. Use of this ansatz in the partial difference equations leads to a set of N_1 coupled, nonlinear, ordinary difference equations for the amplitudes of the N_1 modes which constitute $u'(m, n)$. The background field $U(m, n)$ appears as a coefficient which, because it is slowly varying, can be taken to be locally constant. Because the original equation is nonlinear, these N_1 amplitude equations do not close automatically, but are often derived through perturbation procedures as asymptotic approximations. In the case we shall examine, the amplitude equations do, in fact, close exactly because of the aliasing phenomenon. The nonlinear terms in the equations are quadratic and are due to both direct interactions of the form

$$\exp\left(i2\pi l_1 \cdot \frac{n}{N}\right) \exp\left(i2\pi l_2 \cdot \frac{n}{N}\right) \to \exp\left(i2\pi\left(l_1 + l_2\right)\frac{n}{N}\right),$$

where $l_1 + l_2 < N/2$, and to "indirect" interactions which involve aliasing error [6, 9] or (what crystal physicists would call) Umklapp processes in which a wavenumber $l = l_1 + l_2 > N/2$ is misrepresented by the wavenumber $l = N - l_1 - l_2$ due to the inability of the grid to resolve wavelengths smaller than $2\Delta x$. It is evident that the wavenumber sets $\{\pm N/4, \pm N/2\}$ and $\{\pm N/6, \pm N/3, \pm N/2\}$ are closed under quadratic interactions (e.g. $N/4 + N/2 = 3N/4 = N - N/4$).

One can now solve the initial value problem for the ordinary difference equations and determine stability curves such as those given in Fig. 1a. Roughly speaking, the stability curve divides into two regions: the plane coordinatized by E, a measure of the initial energy in the noise, and α, a nondimensional stability parameter (e.g. $U\Delta t/\Delta x$). In one region, solutions grow without bound (overflowing in 10–100 time steps), whereas, in the other region, solutions simply oscillate neutrally.

This curve provides all of the information usually associated with nonlinear stability theory. If the stability curve intersects the α axis ($E = 0$) at a finite point $\alpha_c(0)$, then the scheme is unstable to infinitesimal disturbances. We note that $\alpha_c(0)$

can be infinite, in which case the scheme is unconditionally stable in the linear sense. For $\alpha < \alpha_c(0)$, a finite E can push the computation into the unstable region. We call this value for $E(\alpha)$ the critical threshold at α. This is the instability discovered, in the numerical context, by Phillips. However, in a carefully designed numerical scheme which inhibits the flow of energy from small to large wavenumbers, there is neither the source of large spontaneous or driven perturbations nor a process analogous to the role that imperfections play in destabilizing elastic shells, through which the critical threshold can be reached. The size and growth rate of roundoff error in numerical schemes is simply too small. Our aim is to show that there is indeed a mechanism, dynamic in character, by which the critical threshold can be attained locally without the benefit of large initial perturbations.

This instability can be described as follows. The solutions of the ordinary difference equations which correspond to values of (α, E) in the neutrally stable region of Fig. 1a and which are exact solutions of the original partial difference equations *are unstable*. They are unstable to modes which are their immediate neighbors in wavenumber space. The instability, which results from the nonlinear interaction between the original modes and their sidebands, manifests itself as a distortion of the envelope of the noise. The exact solution will have a spatial period of the order of the grid length ($4\Delta x$ or $6\Delta x$ in our examples) depending on which set of N_1 modes is used. The envelope of the exact solution is constant in space and oscillates in time. In our experiment, the initial noise in the sideband modes triggering the instability is due to roundoff error. In real calculations, there would generally be some energy already in these modes. The instability mechanism itself is a noise amplifier. Its character (initial growth rate and wavelength) is independent of the size

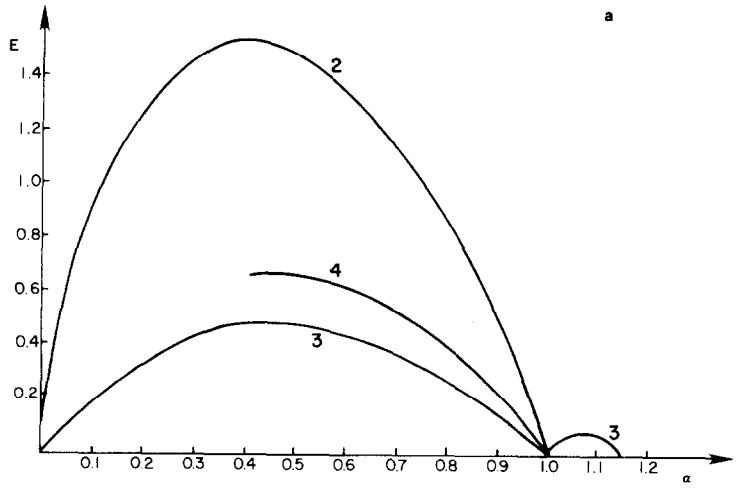

FIG. 1a. Stability curves in (α, E) plane as determined from amplitude equations: (2) two mode solution $(\pi/2, \pi)$, (3) three mode solution $(\pi/3, 2\pi/3, \pi)$, (4) four mode solution $(\pi/4, \pi/2, 3\pi/4, \pi)$.

of the grid and the degree of precision used in the calculations. Its wavelength is chosen dynamically as being the one of optimal growth. We understand the initial stages of this process. For the later stages, we have developed an envelope equation which appears to describe the subsequent growth reasonably well. The envelope begins to distort and slowly develops sharp peaks (focuses) at isolated points along the grid. When the local amplitude reaches the critical threshold given by Fig. 1a, the noise level accelerates dramatically and becomes unbounded within relatively few time steps.

Thus the process by which the *partial* difference equation destabilizes is a twofold one. At first the noise level in the potentially unstable modes (introduced in real computations by nonlinear cascade) is not large. The noise begins to focus and, if the spatial grid is large enough, can eventually reach the critical threshold locally. Depending on the initial noise amplitude, this focusing process can take a long time (often on the order of 10^3–10^4 time steps) to develop. At this point, the conditions for nonlinear finite amplitude instability are satisfied and the noise grows without bound. The critical parameter in determining whether a partial difference equation is unstable is a combination of both noise level and grid size. In this sense, for large enough grids, the leapfrog method is always unstable!

In summary then, we have provided an explanation for spatially local instabilities in locally neutrally stable schemes. To our knowledge, all other theories of nonlinear instability are global in that breakdown occurs uniformly throughout the spatial grid. One can, of course, inhibit the instability by attacking its source of life, namely, the energy in the small scales. Indeed, such remedies as (i) filtering the high wavenumbers, (ii) using a finite difference scheme that impedes the energy cascade to the point where the stability curve is almost vertical so that no finite amplitude instability is present, (iii) averaging the solution at successive time intervals, and (iv) inserting a forward time step at prescribed intervals, will suppress or delay the appearance of the instability. However, these remedies may also have undesirable side effects. Although we have no precise algorithm, we will discuss ways in which these techniques (particularly (iv)) might be applied in order to suppress the instability and at the same time minimize extraneous side effects.

The contents of the paper are as follows: In Section 2, we present the stability diagrams on which the discussion of nonlinear instability is usually based. Along the way we see, in the context of our example, the nonlinear instabilities described by Phillips and Kreiss and Oliger. In Section 3, we illustrate the instability of the solutions used in Section 2 and display, through numerical experiments, the focusing property. We include careful experiments showing that this behavior is not simply due to lack of computational precision, but rather is a genuine instability whose initial growth rate is independent of the precision and the size of the grid. In Section 4, we introduce an envelope equation as an attempt to give a universal equation which will describe the focusing process for a larger class of schemes. In the last section, we discuss some ideas about how to apply various remedies to inhibit instability and we advance some conjectures concerning the parameters on which the focusing property might depend.

II. Difference Scheme and Amplitude Equations

We will presently consider the stability of a particular finite difference scheme applied to the nonlinear advection equation

$$u_t + uu_x = 0,$$

subject to periodic boundary conditions $u(t, 0) = u(t, 1)$ and initial conditions $u(0, x) = f(x)$. The stability of the constant solution $u = U$ $(U > 0)$ will be handled first. Perturbations u' about the constant solution satisfy

$$u'_t + (u' + U)\, u'_x = 0.$$

We discretize the perturbation equation over a grid with time step k and space step $h = 1/N$ and let $u(m, n)$ be the discrete approximation to the exact solution $u'(mk, nh)$. Using second order finite differences in x and t gives the set of difference equations

$$\begin{aligned} u(m+1, n) - u(m-1, n) + \frac{\theta\gamma}{2}\left[u^2(m, n+1) - u^2(m, n-1)\right] \\ + \left[(1-\theta)\,\gamma u(m, n) + \alpha\right]\left[u(m, n+1) - u(m, n-1)\right] = 0 \\ \text{for } 0 \leqslant n \leqslant N-1, \quad m \geqslant 1; \quad (2.1) \\ u(m, 0) = u(m, N), \end{aligned}$$

where $\theta \in \mathbb{R}$, $\gamma = k/h$, $\alpha = kU/h$. The nonlinear term has been discretized in two different ways. It is not difficult to show that with $\theta = \frac{2}{3}$ the scheme satisfies the conservation properties that

$$M = \sum_{n=1}^{N} u(m, n) \qquad \text{and} \qquad E = \sum_{n=1}^{N} u(m+1, n)\, u(m, n) \tag{2.2}$$

are independent of m. In the calculations that follow, the choice $\theta = \frac{2}{3}$ will be used. In addition, we assume $\gamma = 1$ $(k = h)$ to eliminate one degree of freedom in parameter space.

A brief look at the associated linear problem will be useful. The linear difference equations

$$u(m+1, n) - u(m-1, n) + \alpha\left[u(m, n+1) - u(m, n-1)\right] = 0, \qquad 0 \leqslant n \leqslant N-1,$$

have normal mode solutions of the form

$$u_p(m, n) \sim e^{(in\, 2p\pi/N) - im\,\phi_i} \qquad \text{for } 0 \leqslant p \leqslant n-1, \qquad i = 1, 2.$$

The frequencies ϕ_1 and ϕ_2 are real and given by

$$\phi_{1,2} = \arctan\left\{\frac{\alpha \sin(2p\pi/N)}{\pm\sqrt{1-\alpha^2\sin^2(2p\pi/N)}}\right\},$$

provided that α is less than the critical value

$$\alpha_p^* = (\sin(2p\pi/N))^{-1}.$$

The frequency ϕ_1 is associated with the physical mode and converges to the exact solution, while ϕ_2 belongs to a spurious or computational mode [6]. Note that $\alpha_p^* \geqslant 1$ and that if $\alpha \leqslant \alpha_p^*$, then mode p is neutral. However, if $\alpha > \alpha_p^*$, then mode p grows exponentially in m (or time). The smallest critical value of α occurs for $p = N/4$, when $\alpha_p^* = 1$. This corresponds to the spatial mode $e^{in\pi/2}$, which has a wavelength of $4h$. This is identified as the most unstable mode. With $p = 0$, the corresponding spatial mode $e^{in\pi}$ has wavelength $2h$ and is neutrally stable for all α. Finally with $\alpha = 1$, $\phi_{1,2} = 2p\pi/N$, $\pi - (2p\pi/N)$, one choice of which gives the dispersion relation of the continuous problem.

We now turn to the full nonlinear difference equation (2.1) and look for exact solutions consisting of a superposition of linear modes. These sets of modes can be chosen by noting that each mode in the set must include its subharmonic which appears through the quadratic, nonlinear term. The resulting equations for the mode amplitudes are closed and therefore their solutions provide exact solutions of the original partial difference equation. The various sets of modes we consider are as follows:

A. *One Mode Solution*

A solution of the form

$$u(m, n) = A(m)\, e^{(i2\pi/3)n} + (\text{complex conjugate}) \tag{2.3}$$

has a wavelength of $3h$ and an amplitude which depends only on time. Substitution of this solution into Eq. (2.1) gives an ordinary difference equation for the amplitude

$$A(m+1) - A(m-1) + i\alpha\sqrt{3}\, A(m) = i\sqrt{3}/2\, \gamma(2-3\theta)\, A^{*2}(m).$$

This equation, for $\alpha = 0$, $\theta \neq \frac{2}{3}$, contains the result of Fornberg [2] who noted that, in the continuous limit, iA behaves in time like $(t_0 - t)^{-1}$. It also includes the observation of Kreiss and Oliger [4] that a spatial pattern $u(m, 0) = u(m, 3) = 0$ with $u(m, 1)$, $u(m, 2)$ of opposite sign (that is, an $e^{(i2\pi/3)n}$ solution) is unstable. To see this, simply take $A(m)$ to be pure imaginary ($A(m) = ia(m)$). Then (2.3) gives $u(m, 0) = u(m, 3) = 0$, $u(m, 1) = -u(m, 2) = -\sqrt{3}\, a(m)$, where $a(m+1) = a(m-1) - a^2(m)$. This pattern leads to unbounded growth and the choice $\theta = \frac{2}{3}$ is again advisable in order to suppress this fast acting instability.

B. *Two Mode Solution*

In order to investigate the nonlinear behavior of the most unstable linear mode, we assume a solution of the form

$$u(m, n) = A(m)\, e^{i(\pi/2)n} + (*) + B(m)\, e^{i\pi n}, \qquad B \in R, \tag{2.4}$$

and obtain the *exact* amplitude equations

$$A(m+1) - A(m-1) = -2i\alpha A(m) - \frac{2i\gamma}{3} A^*(m)\, B(m), \tag{2.5a}$$

$$B(m+1) - B(m-1) = -\tfrac{2}{3} i\gamma (A^2(m) - A^{*2}(m)). \tag{2.5b}$$

In order to recover the linear stability result, it is useful to include (thinking of α as close to one) the linear (fast) time response in the exponential by setting

$$A(m) = a(m)\, e^{-i(\pi/2)m}, \qquad B(m) = b(m)\, e^{-i\pi m},$$

whence (2.5a), (2.5b) become

$$a(m+1) - 2a(m) + a(m-1) = 2(\alpha - 1)\, a(m) + \frac{2\gamma}{3} a^*(m)\, b(m), \tag{2.6a}$$

$$b(m+1) - b(m-1) = \tfrac{2}{3} i\gamma (a^2(m) - a^{*2}(m)). \tag{2.6b}$$

Note that the necessary linear stability criterion $\alpha \leqslant 1$ for the scheme is contained in (2.6). However, when the nonlinear terms are included, Eqs. (2.6) can exhibit unbounded growth even when $\alpha < 1$ provided the initial disturbance is sufficiently large.

The amplitude equations (2.5a), (2.5b) are very revealing and deserve careful analysis. First notice that the A^*B term which appears in the $\pi/2$ mode equation (2.5a) represents an interaction between the π mode and the $\pi/2$ mode. This is precisely the nonlinear interaction due to aliasing error identified by Phillips (see also the discussion in Mesinger and Arakawa [6]). The result of this interaction is the production of a $3\pi/2$ mode which is resolved by the system as contributing to the change in A^*, the amplitude connected with the $-\pi/2$ mode. Indeed, it is precisely because of aliasing error that an exact closure of the amplitude equations is achieved. In Phillips' example, the equations equivalent to (2.5) would have $\alpha = 0$ in (2.5a) and no right-hand side in (2.5b). In his case, he would allow a solution in which $B(m)\, e^{-i\pi m}$ has the same sign at successive time steps. This leads to immediate exponential growth. On the other hand, if this quantity has opposite signs at successive m, a certain amplitude threshold is required in order to initiate the instability. It is the latter case which our situation parallels.

Equations (2.5a), (2.5b) also show clearly the role which the computational mode plays in the development of finite amplitude instability. Assume that initially the

amplitudes $A(m)$ and $B(m)$ are small, in which case the linear portions of (2.5a), (2.5b) will determine their growth. We then have

$$A(m) = A_p e^{-im\phi_1} + A_c e^{-im\phi_2} = A_p e^{-im\phi_1} + A_c(-1)^m e^{+im\phi_1},$$
$$B(m) = B_p(-1)^m + B_c,$$

where ϕ_1 and $\phi_2 = \pi - \phi_1$ are given by the linear dispersion relation, A_p and B_p are coefficients of the physical mode, and A_c and B_c are coefficients of the computational mode. The linear solution given above will begin to contribute to the right-hand side of the B equation (2.5b) in the following way:

$$B(m+1) - B(m-1) = -\tfrac{2}{3} i\gamma(A_c^2 - A_p^{*2})\, e^{2im\phi} - \tfrac{2}{3} i\gamma(A_p^2 - A_c^{*2})\, e^{-2im\phi} + \tfrac{8}{3}\gamma(-1)^m \operatorname{Im}(A_p A_c).$$

The third term of the right-hand side is a homogeneous solution and hence gives rise to a resonant solution. We find that

$$B(m) = \{\text{homogeneous solutions}\} + ae^{2im\phi} + be^{-2im\phi} - \tfrac{2}{3}\gamma \operatorname{Im}(A_p A_c)(-1)^m m,$$

where a, b are constant independent of m. The nonlinear term of the right-hand side of the A equation (2.5a) will now reflect this growth in $B(m)$,

$$\begin{aligned} A^*(m)\, B(m) = {} & e^{im\phi}\{-\tfrac{2}{3}\gamma \operatorname{Im}(A_p A_c)\, A_c^* m + bA_p^* + A_c^* B_p + A_c^* B_c(-1)^m\} \\ & + e^{im\phi}\{-\tfrac{2}{3}\gamma \operatorname{Im}(A_p A_c)\, A_p^* m + aA_c^* + A_p^* B_c + A_p^* B_p(-1)^m\} \\ & + \text{higher harmonics.} \end{aligned}$$

We see that $A(m)$ is driven by terms which grow linearly in m and which involve the computational mode of $A(m)$ itself. This interaction triggers the finite amplitude instability. When the $A^*(m)\, B(m)$ term overcomes the linear (restoring) term, rapid growth of the solution sets in. Analogous arguments could also be carried out in the three and four mode amplitude equations.

We compute the nonlinear stability threshold as follows: Let

$$u(0, n) = u(1, n) = \sigma\{(1+i)\, e^{i(\pi/2)n} + (*) + e^{i\pi n}\},$$

and with the initial conditions $A(0) = A(1) = \sigma(1+i)$, $B(0) = B(1) = \sigma$ compute solutions for (2.5). Note that the total amplitude is given by

$$E = \max_{0 \leqslant n \leqslant N,\, m = 0,1} |u(m, n)| = 3\sigma.$$

In Figs. 1a, b we show the regions of the (α, E) plane which correspond to bounded (for 2×10^4 time steps, the solution oscillates) and unbounded (usually overflow occurs in less than 10^2 time steps) solutions. The transition in the (α, E) plane from

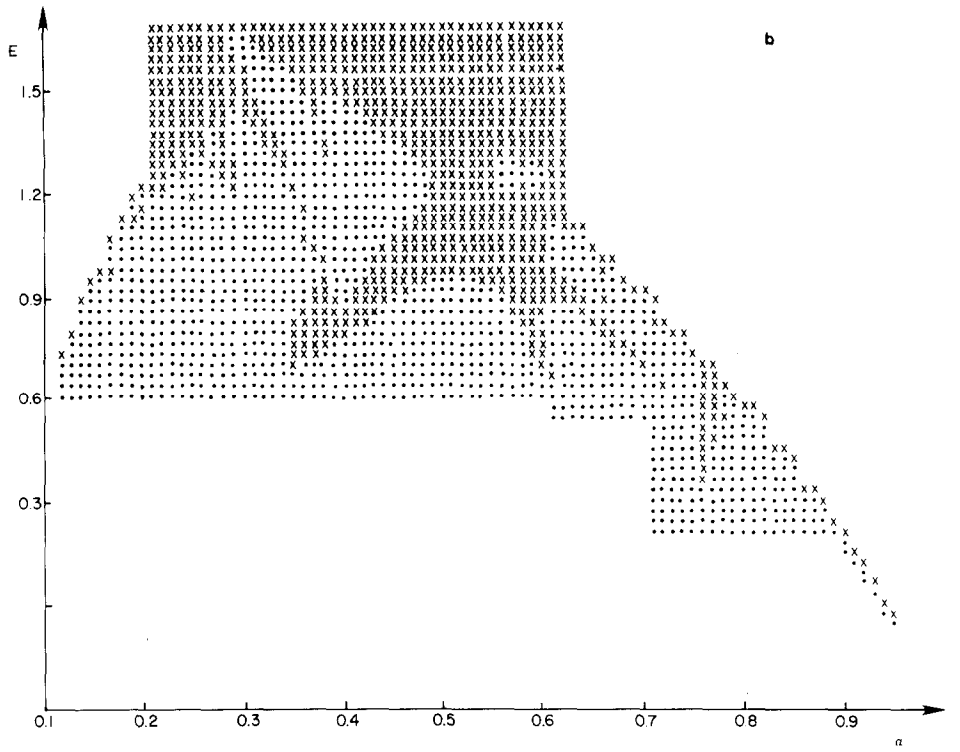

FIG. 1b. Enlargement of (α, E) parameter plane for two mode solution: $(\cdot)$, stable solution to 2×10^4 time steps; $(\times)$, unstable solution.

bounded to unbounded solutions is not smooth. Given recent experience with mappings, it is not surprising that the boundary is irregular and that the domains of attraction of the bounded and unbounded solutions are interspersed. We stress, however, that when we examine the stability of solutions in Section 3, we begin with initial conditions which belong to the stable region of the amplitude equations.

In Fig. 1a, we draw smooth curves to indicate roughly where the boundary lies. In Fig. 1b, we give a more detailed picture of the (α, E) plane in the case of two modes (Eqs. (2.5)). The dots correspond to initial values of α and E for which the solution remains stable for 2×10^4 time steps; the crosses indicate values for which the solutions rapidly (in less than 10^2 steps) blow up.

Part of this behavior is due to the fact that the initial phase (which we have chosen to be fixed) is also important in determining the final disposition of the solution. Our choice of weighting the three modes $e^{\pm i\pi n/2}$, $e^{i\pi n}$ equally does not significantly affect the average position of the stability boundary nor the qualitative features of Fig. 1b. It does change quantitatively the complicated patterns seen near the boundary, and it

does change the actual number of steps needed to reach instability. We confirmed this by choosing different weightings while keeping the "energy"

$$\sum_{n=1}^{N} u(0,n)\,u(1,n) = N(A(0)A^*(1) + A^*(0)A(1) + B(0)B(1))$$

fixed.

The curves in Fig. 1a can be considered to be representative. It is interesting to note that the stability boundary reaches a maximum at approximately $\alpha = 0.5$ and then returns to $E = 0$ at $\alpha = 0$. Recall that since we keep γ fixed (equal to one) in these experiments, α getting smaller means that the size of the solution U about which we perturb is getting smaller. One might argue from (2.6a) that the smaller α is, the larger is the linear restoring force which the nonlinearity must overcome. However, this thinking is really only of value when α is close to one and we can take the continuous time limit of (2.6). It is better to consider Eqs. (2.5). If we write $A(m) = x_m + iy_m$, $B(m) = b_m$, then (2.5a) reads for $\alpha = 0$, $\gamma = 1$,

$$x_{m+1} = x_{m-1} - \tfrac{2}{3} b_m y_m,$$
$$y_{m+1} = y_{m-1} - \tfrac{2}{3} b_m x_m,$$

which if $x_0 = y_0, x_1 = y_1$ allows the solution $x_m = y_m$ for all m; thus

$$b_{m+1} = b_{m-1} + \tfrac{8}{3} x_m^2,$$

and therefore always increases. This result is not significantly affected by a change of initial conditions. For example, if the energy is redistributed in a different manner among $x_0, x_1, y_0, y_1, b_0, b_1$ the stability threshold at $\alpha = 0$ can increase to as much as $+0.05$. Thus the role of α, for α small, is to dephase x_m and y_m which inhibits the monotonic growth of b_m.

Similar comments apply to the other stability curves of Fig. 1a which are calculated by solving the initial value problem for the ordinary difference equations (2.7), (2.9), describing three and four mode behavior respectively. The reason that the stability boundary for the three mode solution touches $E = 0$ at $\alpha = 1$ is that, at this value of α, the solution $A(m) = \exp(-2\pi i m)/3$, $B(m) = \exp(-\pi i m)$, corresponding to the undistorted travelling wave, exactly cancels the linear terms in the equation. Thus the nonlinearity has no linear restoring force to overcome.

C. *Three Mode Solution*

The $2\pi/3$ (period 3) mode can also appear as a solution with the π (period 2) and $\pi/3$ (period 6) modes. A solution of the form

$$u(m,n) = A(m)\,e^{i(\pi/3)n} + B(m)\,e^{i(2\pi/3)n} + C(m)\,e^{i\pi n} + (*) \qquad (2.7)$$

is an exact solution of the full partial difference equations provided the amplitudes satisfy

$$A(m+1) - A(m-1) + \frac{2i\gamma}{\sqrt{3}}(A^*B + B^*(C + C^*))_m + i\alpha\sqrt{3}\,A(m) = 0,$$

$$B(m+1) - B(m-1) + \frac{i\gamma}{\sqrt{3}}(2A^2 + A^*(C + C^*))_m + i\alpha\sqrt{3}\,B(m) = 0,$$

$$C(m+1) - C(m-1) + \frac{i2\gamma}{\sqrt{3}}A(m)\,B(m) = 0. \tag{2.8}$$

Once again a stability curve relating α to the critical value of the initial amplitude has been determined experimentally. This curve is also shown in Fig. 1a, for the case $A(0) = A(1) = B(0) = B(1) = \sigma(1 + i)$, $C(0) = C(1) = \sigma$. Now $E = \max_{0 \leqslant n \leqslant N,\, m = 0,1} |u(m, n)| = 5\sigma$.

D. *Four Mode Solution*

An exact solution to the full partial difference equations consisting of four linear modes takes the form

$$u(m, n) = A(m)\,e^{i(\pi/4)n} + B(m)\,e^{i(\pi/2)n} + C(m)\,e^{i(3\pi/4)n} + D(m)\,e^{i\pi n} + (*). \tag{2.9}$$

The amplitudes must satisfy the ordinary difference equations

$$A(m+1) - A(m-1) + i\gamma\left[\frac{1}{3}(2 + \sqrt{2})A^*B + \left(\frac{2}{3} - \sqrt{2}\right)B^*C + \frac{\sqrt{2}}{3}C^*D\right]_m$$
$$+ i\alpha\sqrt{2}\,A(m) = 0,$$

$$B(m+1) - B(m-1) + i\gamma\left[\tfrac{1}{3}(2 + \sqrt{2})A^2 + \tfrac{4}{3}A^*C + \tfrac{2}{3}B^*D + \tfrac{1}{3}(2 - \sqrt{2})C^{*2}\right]_m$$
$$+ 2i\alpha B(m) = 0,$$

$$C(m+1) - C(m-1) + i\gamma\left[\frac{\sqrt{2}}{3}A^*D + \left(\frac{2}{3} + \sqrt{2}\right)AB + \frac{1}{3}(\sqrt{2} - 2)B^*C^*\right]_m$$
$$+ i\alpha\sqrt{2}\,C(m) = 0,$$

$$D(m+1) - D(m-1) + i\gamma\tfrac{2}{3}(\sqrt{2}\,AC + B^2)_m = 0. \tag{2.10}$$

The stability curve determined from these amplitude equations is also shown in Fig. 1a for the case $A(0) = A(1) = B(0) = B(1) = C(0) = C(1) = \sigma(1 + i)$, $D(0) = D(1) = \sigma$. Now $E = \max_{0 \leqslant n \leqslant N,\, m = 0,1} |u(m, n)| = 7\sigma$.

III. Focusing in the Partial Difference Equations

The calculations of the previous section provide the regions of stability for exact solutions to the full partial difference equations. However, these stability curves were determined, not from the full partial difference equations, but rather from a set of ordinary difference equations that govern the amplitudes of various Fourier modes. We now return to the partial difference equations for a numerical experiment that can be thought of as a verification of the stability results of the previous section. In all cases, we will begin from initial conditions which give rise to stable solutions of the ordinary difference equations (2.5) and (2.8).

Consider the specific case of the exact three mode solution

$$u(m, n) = A(m)\, e^{i(\pi/3)n} + B(m)\, e^{i(2\pi/3)n} + C(m)\, e^{i\pi n} + (*).$$

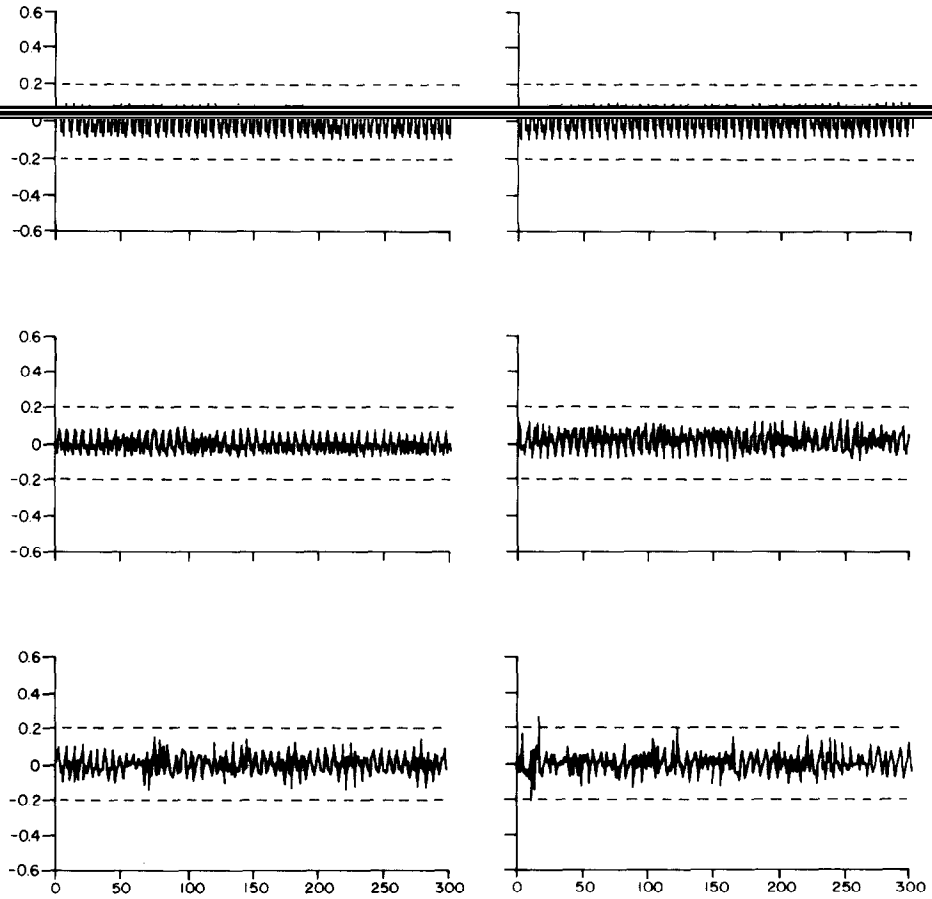

FIG. 2. Solution to the partial difference equations: 3 modes, $N = 300$, $\alpha = 0.9$, $E = 0.1$, $m = 400$, 1000, 2000, 2200, 2400, 2680.

According to the stability curve of Fig. 1a, a value of $\alpha = 0.9$ and an initial amplitude of $E = 0.1$ should produce a stable solution of the amplitude equations. Yet when the partial difference equations are solved with $\alpha = 0.9$ and $E = 0.1$, something unexpected happens. The results of this calculation are shown in Fig. 2. With $N = 300$ grid points on the interval $0 \leqslant x \leqslant 1$, the solution is plotted at time steps $m = 400$, 1000, 2000, 2200, 2400, 2680. The dashed lines indicate the critical amplitude at which finite amplitude instability sets in according to the stability curve of Figs 1a, b. (In this case, the critical amplitude is about 0.2.) Clearly, the initial amplitude in this case is subcritical. At $m = 400$ and $m = 1000$ (Figs. 2a, b), the solution still retains the periodic structure of the initial conditions; its amplitude is well below critical everywhere. By $m = 2000$ (Fig. 2c), the constant envelope of the initial profile begins to vary slowly in x. The solution remains well contained through $m = 2400$ (Fig. 2e), although local amplitudes have exceeded their initial value. At $m = 2680$ (Fig. 2f) the solution exceeds the threshold value at a single grid point. This completes the first stage of the development of the instability. It is characterized by the slow gathering or focusing of the solution locally. Once the solution reaches the critical threshold at even a single grid point, the second stage of the development takes place swiftly. By $m = 2700$, finite amplitude instability, as predicted by the amplitude equations, has taken over and the solution grows without bound. During the integration, the two quantities $\sum_n u(m, n)$ and $\sum_n u(m+1), n)\, u(m, n)$ are conserved exactly.

Some understanding of this process may be gained by looking at the Fourier spectrum of the solution at the same time steps shown in Fig. 2. On a grid of $N = 300$ points, there are 150 distinct modes with mode j of the form $e^{i(\pi j/150)}$ having wavelength of $(300/j)\, h$. After $m = 1000$ time steps (Fig. 3a) the energy is still in the three modes of the initial conditions. By $m = 2000$ time steps (Fig. 3b) the energy has

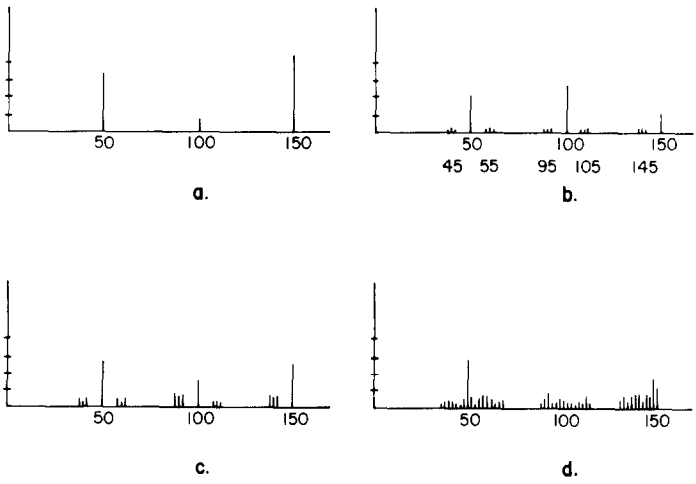

FIG. 3. Spectral resolution of (a) $u(1000, n)$, (b) $u(2000, n)$, (c) $u(2200, n)$, (d) $u(2400, n)$, of Fig. 2.

spread to the sidebands with wavenumbers $\mu = 45, 55$, $\mu = 95, 105$, and $\mu = 145$ ($\mu = 50$ is the period 6 or $\pi/3$ mode). This corresponds to an envelope modulation of wavelength $300/5 = 60$, and the constant envelope of the initial profile begins to vary slowly in x. *In short, the exact solutions of the amplitude equations are unstable solutions of the full partial difference equations.* This is the beginning of the focusing process. The slow modulation of the envelope is triggered only by the presence of errors, either in the initial conditions or in computation. Figure 3c at $m = 2200$ time steps shows a further spreading of the energy in wave number space corresponding to a continued enhancement of the modulation in the envelope. During time steps $m \geqslant 2400$ (Fig. 3d), the energy is distributed through all wavenumbers, approaching a uniform distribution. In physical space, this corresponds to the envelope of the

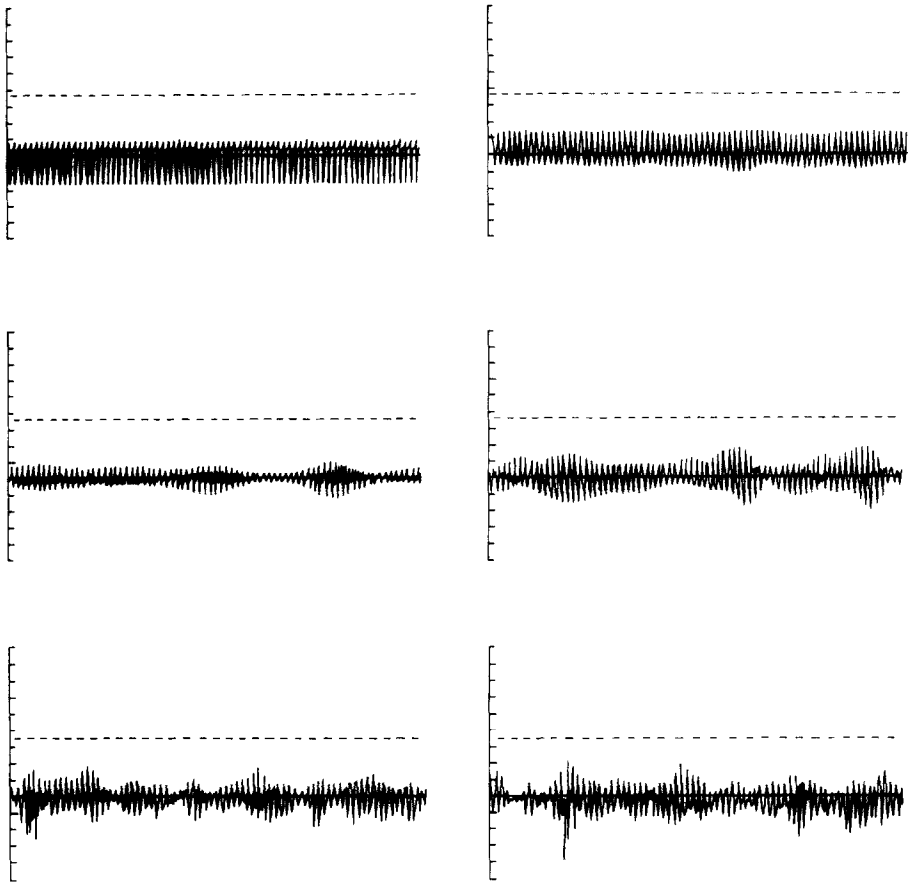

FIG. 4. Solution to partial difference equations: 2 modes, $N = 300$, $\alpha = 0.9$, $E = 0.15$, $m = 100, 400, 450, 500, 800, 850$.

solution having focused into a local peak with subcritical amplitude. Once focusing elevates the maximum amplitude above the threshold, finite amplitude instability sets in, leading to a rapid deterioration of the solution.

We point out again that the initial conditions for this experiment correspond to an exact and bounded solution of the full partial difference equations. The focusing mechanism feeds on errors in the calculations and magnifies them at a level which is subcritical even for finite amplitude (nonlinear) instability. The effect of focusing can be accelerated by adding small perturbations to the initial conditions. It can be delayed by doing the calculation in higher precision.

Figure 4 shows another sequence of experimental results. We choose initial conditions consisting of the $\pi/2$ and π modes and N, the number of grid points, is 300. With $\alpha = 0.9$, the critical amplitude (by Figs. a, b) is $E_c = 0.36$ and is marked by a dashed line in the figures. For these calculations, an initial amplitude of $E = 0.15$ was chosen. For early times, the envelope oscillates in a manner almost

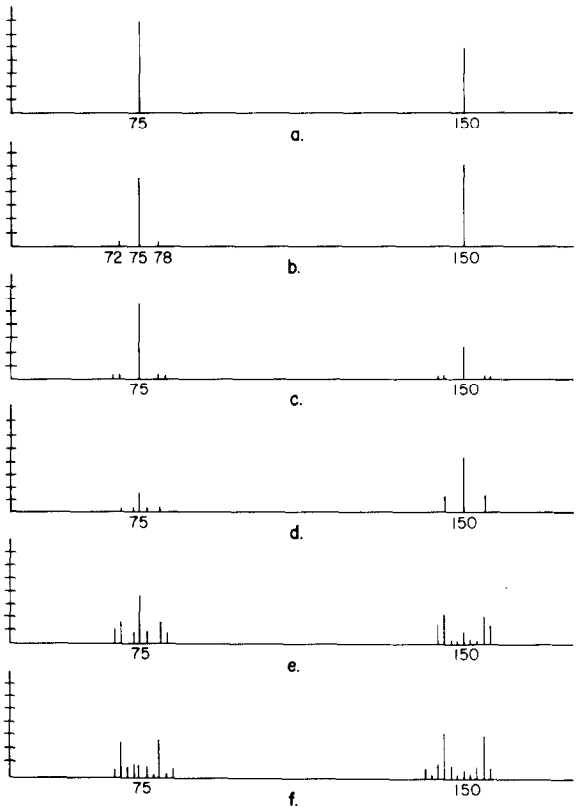

FIG. 5. Spectral resolution of solutions of Fig. 4 at $m = 0$, 100, 400, 450, 500, 550, respectively, (a)–(f).

independent of x and in precise agreement with the motion predicted by Eqs. (2.5a), (2.5b). One can think of the system as consisting of a chain of coupled oscillators in a nonlinear potential. For early times, their orbits are almost synchronized. However, a careful analysis of the spectrum reveals that the envelope has begun to deform and one can already see the long wave modulation at times $m = 100, 400$ (Figs. 4a, b). The spectral decompositions of $u(100, n)$ (Fig. 5b) shows that the sidebands $k = 72$, 78 ($k = 75$ is the period 4 or $\pi/2$ mode) are excited. This corresponds to an envelope modulation of wavelength $300/3 = 100$. This wavelength is chosen dynamically, and is a function of the initial amplitude but not of N, the number of grid points. This fact was verified by taking values of N ranging from 60 to 600. The fact that there is a most unstable sideband, and that the wavelength of the most unstable mode is inversely proportional to the initial amplitude, is consistent with parallel theories of modulation instabilities [8, 11]. Returning to the experiment shown in Fig. 4, we note that by $m = 400$, the deformation of the envelope into a wave of wavelength 100 is clear to the eye, although by this step some energy has also been transferred to the sidebands $k = 71$ and 79 (Fig. 5c). By $m = 800$ (Fig. 4e), the envelope has deformed so that in several locations it is about to exceed the critical threshold. Within fifty more time steps (Fig. 4f), the solution becomes rapidly unbounded. Note that the maximum negative peak (Figs. 4e, f) travels with a speed of almost one consistent with the envelope description discussed in the next section.

An important question is whether the behavior observed in these two experiments is inherent in the difference equations themselves or whether it can be attributed to finite precision arithmetic. To address this question, the growth rate of the instability was measured for various cases in both single and double precision. One measure of growth rate was obtained by monitoring the quantity

$$e_1(m) = \max_{1 \leqslant n \leqslant N} |u(m, n) - \hat{u}(m, n)|,$$

where u is the solution of the full partial difference equation and $\hat{u}$ is the solution of the partial difference equation as reconstructed from the solution of the amplitude equations. The quantity e_1 measures the deviation of the exact solution ($\hat{u}$) from the destabilized solution (u) (assuming the same initial conditions) and thus gives an indication of the rate at which the instability is developing. Figures 6a, b show plots of m vs $e_1(m)$ for a single and a double precision calculation on a grid of $N = 240$ with two mode $(\pi/2, \pi)$ initial conditions. The average growth rates, as determined from each curve's interval of uniform growth, are essentially identical. A similar run with $N = 300$ in single and double precision also yields the same growth rate. A second quantity

$$e_2(m) = \max \left\{ \max_n u(m, n) \right\} - \min \left\{ \max_n u(m, n) \right\}$$

measures the rate at which the amplitude of the envelope modulation grows. When this quantity is monitored, a growth rate is obtained which not only agrees in single

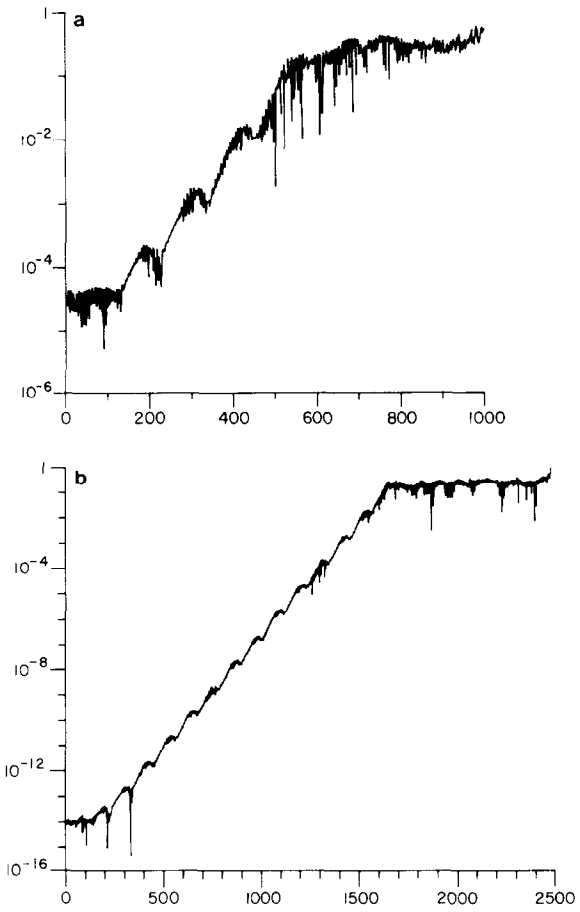

FIG. 6. Growth rate curves. Single (a) and double (b) precision, $N = 240$.

and double precision, but also agrees with the growth rate obtained from e_1. It seems reasonable to conclude that the mechanism which is responsible for focusing resides in the difference equations and is not an artifact of finite precision arithmetic.

IV. Envelope Description

Since the instability which leads to focusing involves wave numbers in the immediate neighborhood of the primary modes, it is natural to seek an envelope description of the process. We carry out this analysis for the situation in which the energy is initially in the $\pi/2$ and π modes. In (2.6) let the amplitudes $a(m, n)$, $b(m, n)$ be slowly varying functions of both the time and space variable. When substituted

into the full partial difference equations (2.1) the following envelope equations for the amplitudes are obtained:

$$|a(m+1,n)-2a(m,n)+a(m-1),n)|-|a(m,n+1)-2a(m,n)+a(m,n-1)|$$
$$=2(\alpha-1)\,a(m,n)+\frac{2i\gamma}{3}\,a^*(m,n)\,b(m,n), \tag{4.1a}$$

$$b(m+1,n)-b(m-1,n)+b(m,n+1)-b(m,n-1)$$
$$=+\frac{2i\gamma}{3}\,(a^2(m,n)-a^{*2}(m,n)). \tag{4.1b}$$

The approximations used in obtaining these equations are valid provided that the spatial gradient $a(m,n+1)-a(m,n-1)$ of $a(m,n)$ is small with respect to the

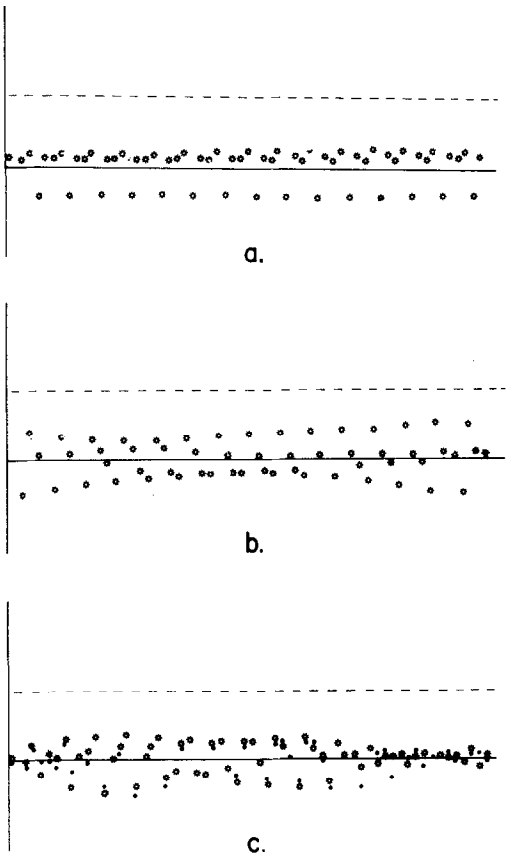

FIG. 7. Solution of partial difference equations (·) and envelope equations (*): 2 modes, $N=60$, $\alpha=0.9$, $E=0.15$, $m=100$, 200, 300, respectively, (a)–(c).

amplitude $a(m, n)$ itself. The advantage of Eqs. (4.1) is that they are universal and will apply to a broad class of partial difference equations. In addition, the envelope equations (4.1) are a better representation of the full partial difference equations (2.1) than the amplitude equations (2.6). Certainly they contain the amplitude equations. They are also a valid approximation to the full partial difference equations at least for early times as, in the initial steps of growth of the envelope instability, the criterion that $a(m, n+1) - a(m, n-1)$ is small with respect to $a(m, n)$ is well satisfied.

In Fig. 7, we show the result of comparing $u(m, n)$ as calculated from (2.1) and as constructed from a solution of the envelope equations (4.1). The initial conditions consist of the $\pi/2$ and π modes only with an amplitude $E = 0.15$ modulated by a long wave perturbation with an amplitude of $0.05E$. The parameter values $\alpha = 0.9$ and $N = 60$ grid points are used. For $m < 200$ (Figs. 7a, b) the two computations produce identical results. When $m > 300$ (Fig. 7c) the approximations used to derive the envelope equations cease to be valid. For example, a universal term, such as $2a^*(m, n)\, b(m, n)$, no longer represents $a^*(m, n+1)\, b(m, n+1) + a^*(m, n-1)\, b(m, n-1)$, a term which is peculiar to the particular partial difference equation under study. Nevertheless, the envelope equations do exhibit the focusing property and display qualitatively similar behavior to the full difference equations even though, in this computation, the critical threshold is reached much sooner (at $m = 650$) by the envelope equations. The full difference equations exhibit focusing behavior which reaches the critical threshold at about $m = 1800$ time steps.

V. Conclusions and Conjectures

The results of the previous sections were obtained in a purely experimental way. These experiments provide evidence for the presence of a universal mechanism for instability in certain nonlinear difference schemes. We have considered the leapfrog scheme which has

(i) potentially unstable modes which are neutral by linear stability analysis,

(ii) a subcritical amplitude threshold governing the onset of finite amplitude instability, and

(iii) the focusing property.

We believe that any difference scheme possessing these properties will be susceptible to instability through this mechanism. Of these three properties, the focusing mechanism is the most difficult to predict. One necessary criterion for focusing is that the envelope equations (4.1) possess n-independent solutions which are always unstable. Although certainly accessible, an analytic result to this effect has not been proved, but, to date, experimental evidence strongly suggests that this instability is always present. A second, more difficult, question is whether the focusing envelope always attains the critical threshold. In order to gain some insight into this question, we plot in Figs. 8 and 9 the number of time steps needed for the critical threshold to

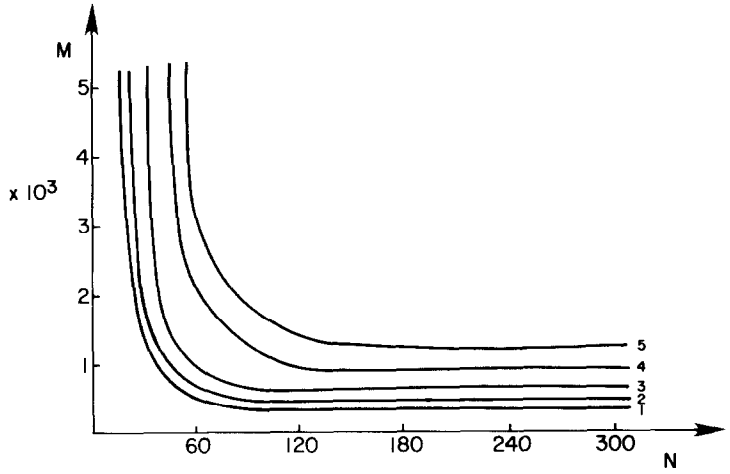

FIG. 8. Number of grid points (N) vs number of time steps to critical threshold (M): 2 modes, $\alpha = 0.9$, various $E = (1)\ 0.3,\ (2)\ 0.27,\ (3)\ 0.21,\ (4)\ 0.18,\ (5)\ 0.15$.

be reached (M) as a function of the number of spatial grid points (N). The different curves are parameterized by E, the amplitude of the solution from which the envelope starts to deform. Figure 8 refers to the case of two mode $(\pi/2, \pi)$ initial conditions in which, with $\alpha = 0.9$, the critical amplitude is $E_c = 0.36$. Figure 9 refers to the case of three mode $(\pi/3,\ 2\pi/3,\ \pi)$ initial conditions in which, with $\alpha = 0.9$, the critical amplitude is $E_c = 0.15$. These results also exhibit some interesting features.

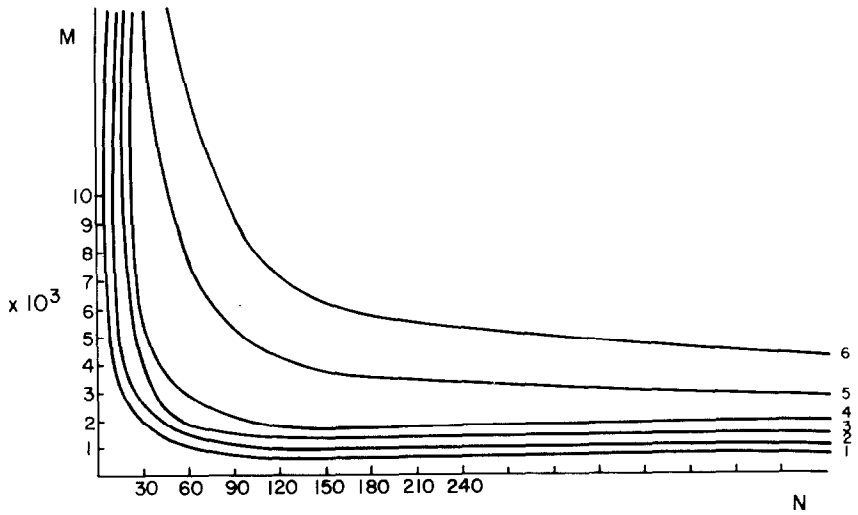

FIG. 9. Number of grid points (N) vs number of time steps to critical threshold (M): 3 modes, $\alpha = 0.9$, various $E = (1)\ 0.14,\ (2)\ 0.12,\ (3)\ 0.1,\ (4)\ 0.09,\ (5)\ 0.106,\ (6)\ 0.05$.

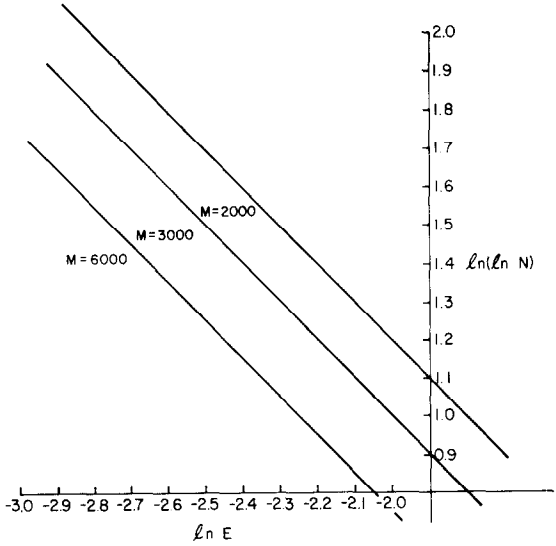

FIG. 10. ln (ln N) vs ln(E), $\alpha = 0.9$, 3 modes for $M = 2000, 3000, 6000$.

(i) The closer E is to its critical value E_c, the larger is the range for which M is independent of N. The fact that these curves are asymptotic at nonzero values of M reflects the fact that the time for the perturbation to reach an amplitude of $E_c - E$ depends on the size of the initial fluctuations and the growth rate of the envelope instability.

(ii) The smaller E is, the larger N must be in order for the envelope to attain the critical threshold. From the data on Fig. 9, we plot in Fig. 10 ln E vs ln (ln N) for fixed M, the number of time steps needed for the envelope to attain the critical threshold. The straight lines indicate that $E \ln N$ is constant for fixed M and furthermore we note that even then $E \ln N$ is only weakly dependent on M. Thus it is not simply the amplitude which determines the ultimate fate of the solution. Rather, the critical parameter appears to be a global quantity which measures a weighted average of the original perturbations. It should be pointed out once again that these results are sensitive to the choice of initial conditions. Figures 8–10 show the situation in which all components of all modes are given equal weight in the initial conditions. It is expected that a different weighting would give qualitatively similar, but quantitatively different, pictures. Since in a typical calculation initial errors are distributed fairly randomly, it would be difficult to use the curves of Figs. 8–10 to predict the number of steps needed to reach the threshold.

This raises the question of how instability due to focusing can be avoided. In Section 2 an argument was given to show the role of the computational mode in the onset of finite amplitude instability. It appears that the presence of spurious, neutral modes also contributes to the focusing mechanism. A number of strategies have been

developed to eliminate the computational mode from calculations that use nondissipative schemes. Among these strategies we tested the following and reached some conclusions.

(i) *Averaging* the solution on two consecutive time levels at regular intervals effectively eliminates the development of envelope instability and keeps the solution intact for any number of time steps. However, averaging amounts to a nonphysical "time" step and, not surprisingly, the conservation of quantities such as M and E is badly violated.

(ii) Periodic *restarts* or insertion of a step with a two-level scheme also appears to suppress the focusing mechanism, but has a negligible effect on the conserved quantities. To show the effectiveness of this strategy the case of Fig. 4 was run again, this time with a two-level Matsuno step inserted every 200 time steps. After 420 time steps, when the original (nonrestarted) solution was showing noticeable modulation of the envelope, the restarted solution still shows a uniform envelope over a perfectly periodic wave. After 880 time steps, when the original solution has become unbounded, the restarted solution still has a uniform envelope.

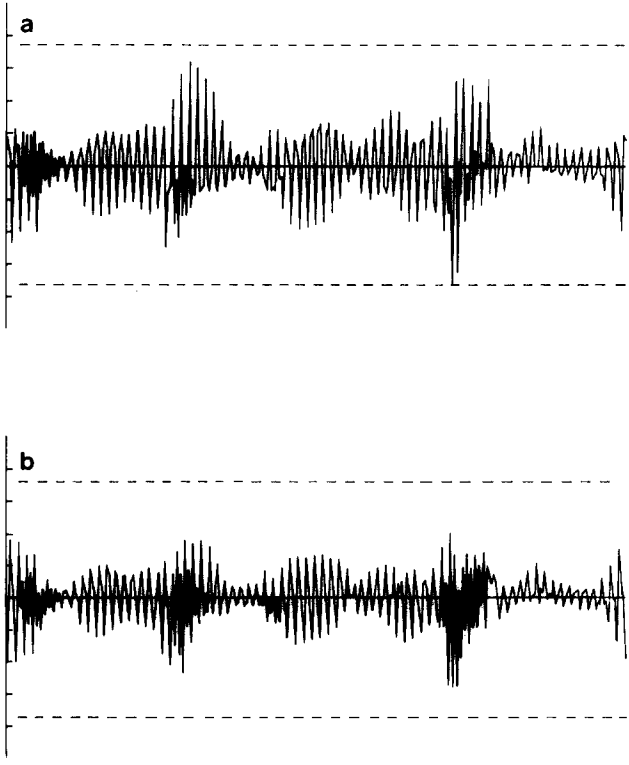

FIG. 11. The effect of one Matsuno step. $N = 300$, $\alpha = 0.9$, $m = 853$, 855, respectively, (a), (b).

In this case with a forward time step taken every 200 time steps, any growth that has begun in the envelope is small enough that it can be eliminated by the damping in one Matsuno step. On the other hand, if a forward step is taken less frequently, then the buckling and growth of the envelope has enough time to develop and one Matsuno step will not restore the uniform envelope. This latter situation is illustrated in Fig. 11: (a) shows a well-developed envelope wave over a grid of $N = 300$ after $m = 853$ time steps; (b) shows the solution one step after a Matsuno step. The effect of the forward step is to reduce the amplitude of the solution uniformly over the grid. One local peak which has reached the critical threshold has been reduced to about 75% of its value.

It is difficult to determine how much of that reduction is due to the damping of the Matsuno step. Some portion of it is due to the inherent oscillation of the envelope. A careful look at the spectrum shows that the energy has been reduced fairly uniformly across all the modes in contrast to the linear case in which the damping is strongest for the high wavenumbers. This particular integration which would have terminated a few steps after $m = 855$ without the forward step continues for several hundred additional steps.

It would be useful to derive a rough prescription for the frequency with which forward steps should be inserted. There are some assumptions in such a calculation which may mitigate its usefulness as a general result, but it does show that damping and envelope growth can be made to compensate each other in an effective way. The growth rate of the focusing instability can be estimated either from Fig. 6 or from the flat portions ($N > 200$) of the curves in Fig. 8. At the same time a linear analysis gives the amount of damping associated with one Matsuno step. For example, the most highly damped mode ($\pi/2$ mode) is damped by a factor of 0.92 when $\alpha = 0.9$. Assuming a constant exponential growth of the envelope, it is possible to determine how often a forward step should be inserted to exactly cancel the growth of the envelope. With $N = 300$, such a calculation yields a frequency of 400 time steps which agrees well with empirically determined strategies. This calculation is certainly oversimplified. The growth rate is not uniform and is somewhat amplitude dependent. Furthermore, the linear estimate of the damping factor is not exactly correct, especially in the later stages of the computation. Nevertheless, the argument does explain qualitatively the success of a forward step in inhibiting the instability.

(iii) For smooth solutions, periodic *filtering* on high wavenumbers has been used successfully to suppress instability. This has not been tried in the present one-dimensional runs (in which the solutions are far from smooth). We would expect filtering to be effective in suppressing focusing which feeds preferentially on the high wavenumbers. At the same time filtering could have an undesirable effect on the budget of conserved quantities.

The qualitative similarity between all of the features reported above and the properties of focusing envelopes of partial differential equations leads us to conjecture that the strength of the focusing mechanism increases with the dimension of the problem. Indeed, preliminary calculations with analogous two-dimensional equations

have borne out this conclusion. We expect that this two step instability process will be potentially present in all large scale computations. Our goal, in this and future work, is to understand the nature of the breakdown of numerical algorithms to the point that we can (a) appreciate why certain ad hoc instability inhibitors (such as filtering and the introduction of artificial viscosity) work and (b) to devise new and more enlightened ways to control instabilities without sacrificing accuracy.

ACKNOWLEDGMENTS

This research was supported in part by National Science Foundation Grants PHY77-27084 and MCS-07548 A01. One of us (A.C.N) is very appreciative of the hospitality and scientific atmosphere he enjoyed while visiting the Institute for Theoretical Physics at the University of California, Santa Barbara. The work was also supported in part by U.S. Army Contract DAAG29-82-C-0068 and ONR Contract N00014-76-C-0867. The authors are also particularly grateful to David Rand whose healthy scepticism and insightful comments encouraged us to examine Fig. 1 in greater detail.

REFERENCES

1. A. ARAKAWA AND V. R. LAMB, *in* "Methods in Computational Physics," p. 173, Academic Press, New York, 1977.
2. B. FORNBERG, *Math. Comput.* **27** (1973), 45.
3. R. H. G. HELLEMAN, *in* "Fundamental Problems in Statistical Mechanics," Vol. 5 (E. G. D. Cohen, Ed.), p. 165, North-Holland, Amsterdam, 1980.
4. H. O. KREISS AND J. OLIGER, *Tellus* **24** (1972), 199.
5. H. O. KREISS AND J. OLIGER, "Methods for the Approximate Solution of Time Dependent Problems," GARP Publication No. 10, 1973.
6. F. MESINGER AND A. ARAKAWA, "Numerical Methods Used in Atmospheric Models," Vol. 1, GARP Publication Series 17.
7. A. C. NEWELL, "Bifurcation and Nonlinear Focusing, Pattern Formation and Pattern Recognition" (H. Haken, Ed.), p. 244, Springer-Verlag, New York/Berlin, 1979.
8. A. C. NEWELL, *SIAM J. Appl. Math.* **33** (1977), 133.
9. N. A. PHILLIPS, *in* "The Atmosphere and the Sea in Motion" (B. Bolin, Ed.), p. 501, Rockefeller Institute, New York, 1959.
10. Y. S. SIGOV AND V. E. ZAKHAROV, *J. Physique* **40** (1979).
11. G. B. WHITHAM, "Linear and Nonlinear Waves," Wiley–Interscience, New York, 1974.