Nonlinear self-modulation: An exactly solvable model

E. R. Tracy
Department of Physics, The College of William and Mary, Williamsburg, Virginia 23185

H. H. Chen
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
(Received 22 December 1986; revised manuscript received 10 August 1987)

The cubic Schrödinger equation (CSE) ($iu_t + u_{xx} \pm 2 | u |^2 u = 0$) is a generic model equation used in the study of modulational problems in one spatial dimension. The CSE is exactly solvable using inverse-scattering techniques. Periodic solutions of the focusing CSE ("+" sign in the above equation) are also well known to be subject to modulational instabilities. This unique mixture of solvability and instability allows the development of a complete and explicit analytical theory for the long-time behavior of the instabilities. Among the results to be discussed are (i) a method for calculating the growth rates of instabilities around (spatially nonuniform) initial states, (ii) a discussion of recurrence phenomena for systems with finite spatial period, and (iii) a method for calculating the recurrence time.

I. INTRODUCTION

An important class of problems in nonlinear physics concerns the propagation of disturbances in a nonlinear medium. Examples include Langmuir waves in strongly turbulent plasmas, nonlinear optics, and the theory of ocean waves. Of particular physical importance in such cases is the question of stability: Is a given wave train stable under perturbations, or will it evolve into something very different from its initial form? A common process in nonlinear wave propagation is self-modulation. This occurs when the wave is large enough to "distort" the medium it is propagating through. Self-modulation can, in some cases, lead to instability so an understanding of its effects is important.

Unfortunately, in almost all the situations of interest in physics, the mathematical equations which describe the nonlinear evolution of the wave are intractable analytically, and not generally amenable to numerical study because they are fully three dimensional and involve a wide range of space and time scales.

One common method of attack in such situations, which is the approach we follow here, is to study model systems which are simplified idealizations but which still correctly capture the behavior of the physical system of interest. The model we use to study self-modulation is the focusing cubic Schrödinger equation (CSE),

$$iu_t + u_{xx} + 2 | u |^2 u = 0 . \tag{1}$$

This nonlinear PDE arises in a wide variety of fields, such as nonlinear optics,[1,2] the theory of deep water waves,[3] and plasma physics.[4]

In these contexts, as in most of the others where the CSE has arisen, the solution of Eq. (1), $u(x,t)$, describes the slow space-time evolution of the envelope of a fast oscillation. For example, suppose $A(x,t)$ is the amplitude of a surface wave on deep water and also that $A(x,t)$ is nearly a plane wave with some slow modulation. We expand $A(x,t)$ in an asymptotic series which has as its leading term

$$A(x,t) = u(x,t)e^{ikx - i\omega t} + \text{c.c.}$$

(where c.c. means complex conjugate). Here we are restricting attention to one spatial dimension. This expression is inserted back into the evolution equation for $A(x,t)$ (in this case the Stokes formulation of the water-wave problem) and this evolution equation is averaged over the fast space-time oscillations. What is left is the evolution equation for the wave envelope.

Equation (1) and its relevance to water waves was examined in a series of numerical *and* laboratory experiments by Yuen and Lake.[5] They found excellent agreement between the numerical predictions of Eq. (1) and wave-tank data in the case where the wave propagation was essentially one dimensional (i.e., a long thin wave tank). The higher-dimensional generalization of Eq. (1) has proved to be far less useful for describing higher-dimensional processes.

The CSE appeared also in the work of Zakharov[4] on Langmuir turbulence in plasmas. In this case, because of the complex array of physical processes taking place in plasmas it is not clear that the CSE has direct application, but it has been used to gain qualitative insight into the energetics of Langmuir turbulence.[6,7]

The CSE also appears in nonlinear optics[1,8] where, for example, it describes the envelope of an electromagnetic wave propagating in an optical fiber. This is essentially a one-dimensional problem and here again the one-dimensional CSE [Eq. (1)] gives excellent quantitative predictions when compared with laboratory experiments.[9]

By now it should be clear that the CSE has general applicability to problems in self-modulation. In fact, using heuristic arguments, it can be shown that the CSE will describe the nonlinear self-modulation of wave

trains under very general assumptions about the nature of the nonlinear and dispersive properties of the medium.[10,11]

Without a doubt, the two most important physical predictions of Eq. (1) are (i) the existence of solitons when the wave profile is spatially localized ($|u| \to 0$ as $x \to \pm\infty$) and (ii) modulational instability for periodic boundary conditions [$u(x+d)=u(x)$ for some d].

The existence of soliton solutions of the CSE was demonstrated by Zakharov and Shabat.[12] They showed that the CSE could be solved exactly by using a modification of a method developed by Gardner, Greene, Kruskal, and Miura[13] to solve the Korteweg–de Vries equation

$$u_t+6uu_x+u_{xxx}=0 .$$

(This method, now called the inverse-scattering method, is described in the Appendix.)

The most important physical property of solitons is that they are localized wave packets which survive collisions with one another. This has been observed experimentally in water waves.[5] In optical fibers "breathers" have been observed. (A breather is a bound state of several solitons traveling at the same speed. They have a very distinctive envelope oscillation. See Ref. 9 for the comparison of numerical predictions and experimental data.)

Solitons, however, are modulationally stable. When *periodic* boundary conditions are enforced the possibility of instability arises. By instability we mean the following. Suppose we start with a solution of Eq. (1). Call it $u(x,t)$. At $t=0$ add a small perturbation to it,

$$\tilde{u}(x,0)=u(x,0)+\epsilon\phi(x) .$$

Now we ask, do $\tilde{u}$ and u stay "close" to one another in some appropriate sense, or does $\tilde{u}$ diverge from u as time progresses? If $\tilde{u}$ and u initially grow apart *exponentially* fast in time, the perturbation is unstable.

The standard way to analyze this problem is to linearize Eq. (1) about the solution $u(x,t)$ and solve a linear eigenvalue problem. This gives us the various unstable perturbations $\phi(x)$ and the growth rates for these perturbations. Let us look at a concrete example. Suppose our initial solution to the CSE is the envelope of an exact plane wave in one dimension. Such an envelope solution would look constant in x. It is easy to show that $u(x,t)$ will then have the form

$$u(x,t)=ae^{2ia^2t} ,$$

where the parameter a is real. Let us perturb this,

$$\tilde{u}(x,t)=u[1+\epsilon(A_1\phi+A_2\phi^*)] ,$$

where ϵ, A_1, and A_2 are real numbers ($\epsilon \ll 1$), and $\phi=\exp(ikx-i\Omega t)$. Here k is real, but Ω may be complex. This ansatz is plugged into Eq. (1) which is then linearized. This leads to

$$[(2a^2+\Omega-k^2)\phi+2a^2\phi^*]A_1 + [(2a^2-\Omega^*-k^2)\phi^*+2a^2\phi]A_2=0 .$$

The complex conjugate of this must also hold,

$$[(2a^2+\Omega^*-k^2)\phi^*+2a^2\phi]A_1 + [(2a^2-\Omega-k^2)\phi+2a^2\phi^*]A_2=0 .$$

For there to be nontrivial solutions of this pair of linear equations for A_1 and A_2 the determinant of the coefficients must be zero. This gives our dispersion relation

$$\Omega=\pm k(k^2-4a^2)^{1/2} . \tag{2}$$

For *long* wavelengths, $|k|<2|a|$, the modulation is unstable; u and $\tilde{u}$ diverge exponentially fast in time.

For more general initial conditions [e.g., if $u(x,0)$ is *not* constant in x] the general procedure is the same, however this now results in a nontrivial eigenvalue problem which is difficult to solve analytically. In fact this calculation has up to now only been carried out for the traveling wave (or "cnoidal" wave) solution of the CSE. These traveling wave solutions are constructed by assuming $u(x,t)=A(x-vt)e^{i\psi(x,t)}$. Here A is a real amplitude function and ψ is a phase. This ansatz is plugged into Eq. (1) and ordinary differential equations generated for A and ψ. Such solutions have elliptic functions as envelopes (hence the term cnoidal). The stability of such solutions was studied by Walstead.[14]

While linearizing to study stability is an enormously useful technique, the drawbacks of such an approach are sizeable. (i) As the complexity of the initial state increases the linearized problem rapidly becomes difficult to analyze (without resorting to numerics), and this prevents any generic understanding of the stability properties of nontrivial solutions. (ii) Even if problem (i) could be overcome by diligence and hard work the solution gives us information about the instability only for short times (i.e., only as long as the linearization is approximately correct). It tells us nothing about the long-time behavior, which is of course what we really want to understand; the linear analysis is merely a first step in the right direction.

These points, while obvious to most readers, are brought up to emphasize that for Eq. (1) we *can* do much better. Using the inverse-scattering technique to solve the CSE with *periodic* boundary conditions has proven to be more difficult than the soliton problem but it is now possible to generate a very large class of spatially periodic solutions to the CSE.[11,15,16] While there is still no claim to be able to solve the initial-value problem in general there is excellent evidence (which we discuss in the body of the paper) to support the argument that the class of solutions that *can* be constructed contain all of the periodic solutions which are *physically* important.

The techniques described below allow the construction of approximate solutions which are uniformly valid in t. This allows one to follow the instability until it saturates and well beyond. The errors associated with these approximations are unimportant in a structural sense. To clarify this, Ercolani, Forest, and McLaughlin—in studies of the periodic sine-Gordon equation[17]—have shown that linear instabilities of soliton systems have a pro-

found geometric interpretation. Solutions which are linearly unstable to perturbations of initial conditions are structurally unstable geometrically. A small change in initial conditions produces a finite change in the underlying topology of the function space associated with these solutions.

The solutions of the CSE developed below include these new degrees of freedom. If all possible instabilities have been taken into account this new solution will be structurally stable to any further perturbations.

The solutions we deal with here are called N-phase wave trains since they can be represented by hyperelliptic functions which have N different phases. The theory of hyperelliptic functions is mostly unknown to physicists, and we will not delve more deeply into it than is needed to explain our results.

The existence of these exact solutions gives us a powerful new tool for the analysis of the modulational instability. This combination of dynamical instability with exact solvability is rare in mathematical physics. Such a combination occurs for those soliton systems which are related to scattering problems which are not self-adjoint. The sine-Gordon equation[17] is another important example. Such opportunities for complete analysis should be exploited to the fullest extent possible. With this new tool we can construct *exact* solutions which initially start as close as we please to a plane wave (for example). By examining the limiting ($t \to 0$) of the exact solution it is possible to recover all of the results of the linear analysis: the existence of instabilities and their initial growth rates.[18] The new approach, however, gives more. Since the exact solution is known we can follow the instability into the nonlinear stage with ease and explicitly follow the long-time behavior.

The same approach can be used to study perturbations around arbitrary initial states. As in the linear analysis the mathematics rapidly becomes difficult, but for comparable amounts of numerical work the new approach gives much more information about the *long-time* behavior of the system.

For example, the present method helps to explain the recurrence phenomena seen in numerical studies of the CSE.[5] By showing that the perturbed solution can be well approximated by a finite-dimensional hyperelliptic function (see Secs. II and III) this immediately implies recurrence, since such functions are quasiperiodic. But it also gives us a method for *computing the recurrence time.*

The emphasis in this paper is on the physical aspects of the theory. Many of the basic mathematical details have been treated elsewhere[11,15,16] so they will simply be summarized in the Appendix. The application of these mathematical techniques to the study of the modulational instability of the CSE has been reported only briefly elsewhere[18] so the details are given here.

Conclusions similar to those reported here were arrived at independently by Ercolani, Forest, and McLaughlin[17] in their studies of the periodic sine-Gordon equation. The approach reported here, however, is different from that in Ref. 17; here we examine the limiting behavior of an exact solution while in Ref. 17 the emphasis is on the use of "squared eigenfunctions" as a basis for the linearized stability analysis. Ercolani, Forest, and McLaughlin also paid strict attention to the geometric aspects of the problem.

The approach used here leads not only to the characterization of which parts of the degenerate spectra are unstable to perturbation, but also leads to a method for predicting the recurrence time.

II. MODULATION INSTABILITIES: PLANE WAVES

As promised in the Introduction, we are now going to apply the results stated in the Appendix to the study of modulational instabilities. The reader who is unfamiliar with the inverse-scattering method may wish to glance at the Appendix before proceeding.

The simplest case to study is that of a perturbed plane-wave solution of the CSE. By "plane-wave solution" we mean that the CSE solution is the envelope of a plane wave. The standard linear analysis of this case was discussed in Sec. I where we derived Eq. (2), the linear dispersion relation for the perturbing sinusoidal disturbance, [Eq. (2)]

$$\Omega = \pm k(k^2 - 4a^2)^{1/2} .$$

Let us set $a=1$ without loss of generality. Now our initial plane-wave solution of the CSE is

$$u(x,t) = e^{2it} .$$

We imagine this placed in a box of length d ($d < \infty$).

Our approach to the study of the stability of solutions to the CSE is as follows. We are starting with an exact N-band solution to the CSE (we shall see that the plane wave is what is called an $N=0$ solution). We can find the main spectrum associated with this solution. Since it is an N-band wave only a finite number $(2N+2)$ of the main spectrum eigenvalues will be nondegenerate. We consider all of the degenerate eigenvalue pairs to be potential degrees of freedom which have been "frozen out" by the special initial conditions. Under perturbation of the initial conditions the degeneracies will be broken and new degrees of freedom will appear. We study the stability of these new degrees of freedom by constructing the *new exact solution* of the CSE which includes many of the new modes (in general, of course, there would be an infinite number but, as we shall see, it is possible to neglect the majority of them).

A. The main spectrum (unperturbed)

Our first step, therefore, is to find all of the main spectrum associated with $u(x,t)=e^{2it}$. We do this as follows. Plug $u(x,0)$ into $L(u)$ (see the Appendix) and solve $L\Phi = \lambda\Phi$ subject to the initial conditions $\Phi(0;\lambda) = 1$ (where 1 is the 2×2 identity matrix),

$$\begin{bmatrix} i\partial_x & 1 \\ -1 & -i\partial_x \end{bmatrix} \Phi = \lambda\Phi, \quad \Phi(0,\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} .$$

It is straightforward to check that

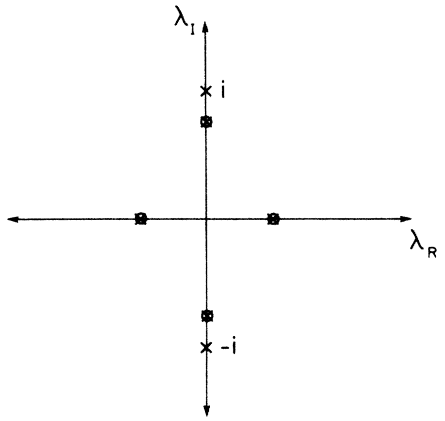

FIG. 1. Unperturbed spectrum of the plane wave $u(x,t)=\exp(2it)$. The nondegenerate main spectrum is shown at $\lambda=\pm i$. The degenerate main spectra are the other crosses, while the auxiliary variables are shown as circles.

$$\Phi(x;\lambda)=\begin{bmatrix}\cos kx & 0\\ 0 & \cos kx\end{bmatrix}+i\frac{\sin kx}{k}\begin{bmatrix}-\lambda & 1\\ 1 & \lambda\end{bmatrix}$$

is the fundamental solution matrix. Here $k^2=1+\lambda^2$. The monodromy matrix is

$$M(\lambda)=\Phi(d;\lambda)$$

and the discriminant $\Delta(\lambda)$ is given by the trace of the monodromy matrix,

$$\Delta(\lambda)=2\cos[d(1+\lambda^2)^{1/2}]\ .$$

Recall that the spectrum splits into regions of stability or instability depending on the behavior of $\Delta(\lambda)$. In this case $\Delta(\lambda)$ is real when $(1+\lambda^2)^{1/2}$ is either real or purely imaginary. This occurs for λ either real or purely imaginary. Along these two lines we have $-2\le\Delta(\lambda)\le 2$ for all real λ and for $\lambda=i\alpha(-1\le\alpha\le 1)$. These are the bands of stable eigenvalues. The main spectrum consists of those λ where $\Delta(\lambda)=\pm 2$. This happens when

$$\lambda_n^0=\pm\left[\frac{n^2\pi^2}{d^2}-1\right]^{1/2}\ . \tag{3}$$

From this formula we can glean some important information. For all $n\neq 0$ $(n=\pm 1,\pm 2,\ldots)$ the eigenvalues are doubly degenerate (see the Appendix). So long as d is finite it is clear from Eq. (3) that only a finite number of degenerate pairs will appear on the imaginary axis. This will turn out to be very important. At $n=0$ we have $\lambda_0=\pm i$. These two eigenvalues are nondegenerate. One interesting way to see this is to show that only one Bloch eigenfunction exists at $\lambda=\pm i$. Consider the fundamental solution matrix,

$$\Phi(x;\lambda=\pm i)=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}+x\begin{bmatrix}\pm 1 & i\\ i & \mp 1\end{bmatrix}\ .$$

Notice the linear behavior in x. The Bloch eigenfunction must be constructed by a linear combination of the column vectors of Φ. Only one such linear combination is periodic; If we add column 1 to $(\pm i)$ times column 2 we get

$$\psi(x;\lambda=\pm i)=\begin{bmatrix}1\\ \pm i\end{bmatrix}\ .$$

Any other independent solution at $\lambda=\pm i$ must contain the x dependence, which is aperiodic. Therefore the eigenvalues $\lambda=\pm i$ are nondegenerate.

This means that what we shall call the zeroth-order spectrum—namely, the spectrum of the initial solution without any perturbations—is like that shown in Fig. 1. If we now perturb the initial conditions we will, in general, break all of the degeneracies.

B. The main spectrum (perturbed)

We assume that under the perturbation

$$u(x,0)\rightarrow u(x,0)+\epsilon\phi(x)$$

all of the spectrum moves $O(\epsilon)$. We now wish to construct a typical solution to the CSE which has a spectrum of the type shown in Fig. 2. The numbering is chosen for ease in the later computation. For imaginary-axis modes ϵ_j is assumed real [Fig. 2(b)] while for real-axis modes ϵ_j is assumed to be imaginary [Fig. 2(c)].

The first assumption concerning the imaginary-axis modes is made in order to simplify the analysis. The only symmetry required of the main spectra is that they

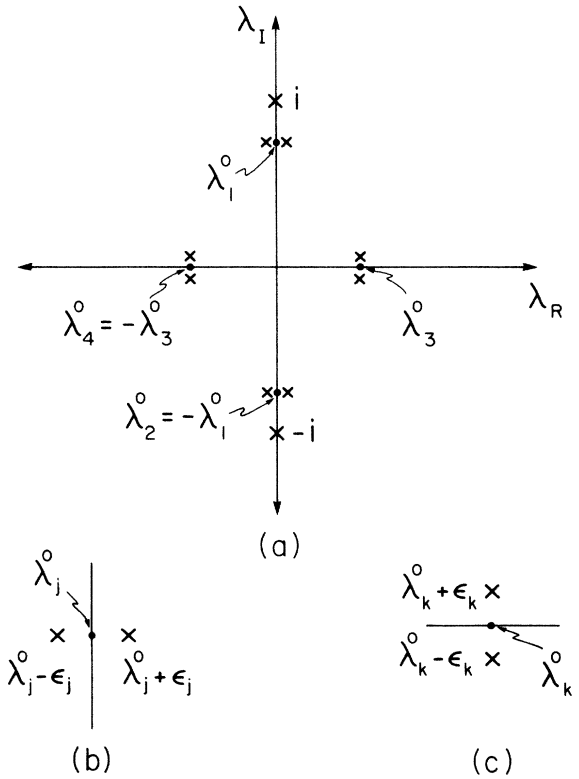

FIG. 2. (a) Typical spectrum after perturbation. No auxiliary variables are shown. The dots represent the numbers λ_k^0. (b) A typical imaginary-axis mode. (c) A typical real-axis mode.

appear in complex-conjugate pairs. Extending the analysis to the more general case where ϵ_j is not purely real is straightforward. The only change occurs in the phases of the theta function [see Eq. (A32) in the Appendix and the discussion which follows it].

When the degeneracy of real-axis modes is broken, however, they become a complex-conjugate pair of spectra, so the assumption that ϵ_j is imaginary for real-axis modes is always valid. Notice also that the numbering is chosen so that the imaginary-axis modes appear first, and then the real-axis modes.

Before constructing the desired solution we must close back up all but a finite number of these nearly degenerate double points. We can include as many as we want in the analysis, as long as we keep only a finite number. We shall see at the end of the computation that this is acceptable on physical grounds since we shall show that the degrees of freedom corresponding to double points at asymptotically large $|\lambda|$ (for real λ) will always be stable and therefore neglectable. In this particular case (plane waves) *all* of the real axis modes are neglectable if we are concerned only with questions of stability.

C. Construction of exact solutions which start "near" a given plane wave

To construct such a solution of the CSE we must proceed with care since the degenerate limit (where we "pinch down" a pair of eigenvalues) is a singular limit for the theta functions. We need to take the limit in such a way that we still get usable results at the end.

In what follows we construct these solutions in a step-by-step manner. Given the general structure of the main spectrum, do the following.

(i) Choose appropriate holomorphic differentials and loop cycles.

(ii) Compute the matrix of a periods.

(iii) Compute the matrix of b periods.

(iv) Compute the $\underline{\tau}$ matrix. In the plane-wave case this is trivial since, using the choice of basis described below, the $\underline{B}$ matrix becomes the $\underline{\tau}$ matrix as $\epsilon \to 0$.

(v) Construct the Abel map and from it find the wave numbers and frequencies.

(vi) Construct the theta function, and thereby the solution of the CSE.

The key is to choose the correct basis for the holomorphic differentials and the correct a and b cycles. The appropriate bases simplify the calculation tremendously. We will call the new basis of holomorphic differentials the "modulation basis" since it is also central to the general modulation problem. The modulation basis is a linear combination of the old "standard" basis,

$$dU_j = \frac{1}{2\pi i}[1+(\lambda_j^0)^2]^{1/2} \frac{\prod_{m\,(\neq j)} (\lambda-\lambda_m^0)d\lambda}{R(\lambda)} ,$$

$$j = 1,2,\ldots,N$$

$$R(\lambda) = (1+\lambda^2)^{1/2} \prod_{k=1}^{N} [(\lambda-\lambda_k^0-\epsilon_k)(\lambda-\lambda_k^0+\epsilon_k)]^{1/2}$$

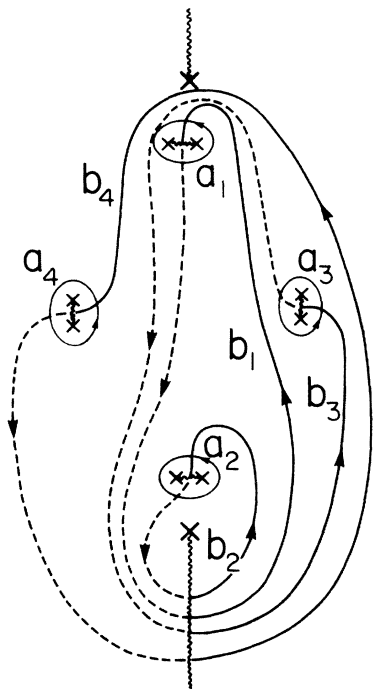

FIG. 3. Positions of the branch cuts and choices for the a and b cycles. The reader should imagine superimposing Fig. 2(a) on this diagram. Cycles a_1 and a_2 lie near the imaginary λ axis, a_3 and a_4 near the real λ axis. Note: A dashed curve implies the path is on the lower sheet of the Riemann surface.

The reason for this choice of basis becomes obvious when we take the limit as $\epsilon \to 0$ (here and in the rest of this paper the symbol $\lim_{\epsilon\to 0}$ is used to symbolize the limit as all of the ϵ_k's are taken to zero simultaneously),

$$\lim_{\epsilon\to 0} dU_j = \frac{1}{2\pi i}[1+(\lambda_j^0)^2]^{1/2} \frac{d\lambda}{(\lambda-\lambda_j^0)(1+\lambda^2)^{1/2}} .$$

Thus, in this limit, each differential "sees" only a pole at one double point and the branch points at $\pm i$.

The appropriate a and b cycles are shown in Fig. 3. This figure looks complicated, but it will simplify when we take the small-ϵ limit.

Since the loop a_j surrounds λ_j^0 we have

$$\lim_{\epsilon\to 0} A_{kj} = \lim_{\epsilon\to 0} \int_{a_j} dU_k = \int_{a_j} \lim_{\epsilon\to 0} dU_k = \delta_{kj} .$$

(This is true because if a_j does not enclose a pole we can shrink the contour down to zero.)

Now the advantage of using these bases for the differentials and loop cycles becomes apparent. As $\epsilon \to 0$, matrix $\underline{A}$ automatically goes over to the identity matrix [to $O(\epsilon)$] so the differentials approach the *normalized* differentials [to $O(\epsilon)$]. This also means that

$$\lim_{\epsilon\to 0} B_{jk} = \tau_{jk} + O(\epsilon) .$$

However, this limit is singular so we must be careful. The off-diagonal terms of B_{jk} are well behaved and can be integrated analytically,

$$\lim_{\epsilon\to 0} B_{kj} = \int_{b_j} \lim_{\epsilon\to 0} dU_k, \quad j\neq k$$

$$\tau_{kj} = \frac{1}{2\pi i}[1+(\lambda_k^0)^2]^{1/2} \int_{b_j} \frac{d\lambda}{(\lambda-\lambda_k^0)(1+\lambda^2)^{1/2}} + O(\epsilon) .$$

As mentioned earlier, the integrand does not contain any information about the poles anywhere else but at λ_k^0. The situation is essentially that of Fig. 4. We have indicated two generic possibilities. In Fig. 4(a) the b cycle does not encounter the pole (i.e., if we deform the contour so that the paths on the top and bottom sheets lie over one another we do not "hit" the pole). In this case the b integral immediately reduces to an elementary integral.

In Fig. 4(b) the b cycle encounters the pole if we attempt to deform it as in 4(a). This is fixed by deforming the contour as shown so that it passes through the branch cut above $+i$ instead of below $-i$. Now this can be reduced to an elementary integral. For double points and poles on the real axis the approach can be generalized in an obvious way.

An important point to note is that the off-diagonal elements depend only on the positions of the original double points [to $O(\epsilon)$]. These in turn are determined only by d, the length of the box. Thus the off-diagonal terms of the τ matrix contain no information about the initial perturbation. As we will now see, only the diagonal terms are affected by the initial perturbation.

The diagonal terms of the τ matrix are singular when $\epsilon=0$. To find how they behave near $\epsilon=0$ consider

$$\tau_{kk} = \frac{1}{2\pi i}[1+(\lambda_k^0)^2]^{1/2} \times \int_{b_k} \frac{d\lambda}{\{[(\lambda-\lambda_k^0)^2-\epsilon_k^2](1+\lambda^2)\}^{1/2}} + O(\epsilon) .$$

Suppose λ_k^0 is on the imaginary axis. By considering Fig. 5 we can see

$$\tau_{kk} = \frac{[1+(\lambda_k^0)^2]^{1/2}}{2\pi i}\left[2\int_{-i}^{\lambda_k^0} + \tfrac{1}{2}\int_{a_k}\right] \times \frac{d\lambda}{\{[(\lambda-\lambda_k^0)^2-\epsilon_k^2](1+\lambda^2)\}^{1/2}} + O(\epsilon)$$

$$= \frac{1}{2} + \frac{[1+(\lambda_k^0)^2]^{1/2}}{\pi i} \times \int_{-i}^{\lambda_k^0} \frac{d\lambda}{\{[(\lambda-\lambda_k^0)^2-\epsilon_k^2](1+\lambda^2)\}^{1/2}} + O(\epsilon) .$$

The integral in the second term is elliptic, but we can make use of the asymptotic form near $\epsilon_k=0$,

$$\tau_{kk} = \frac{1}{2} - \frac{i}{\pi}\ln|\epsilon_k| + O(1) + O(\epsilon) ,$$

where $O(1)$ is the contribution from the nonsingular part of the integral. For singularities on the imaginary axis this $O(1)$ contribution is purely imaginary. For λ_k^0 on the real axis we can write the integral as

$$\int_{-i}^{\lambda_k^0} = \int_{-i}^{0} + \int_{0}^{\lambda_k^0}$$

(where the second integral is along the real axis). In the first integral we can take $\epsilon_k=0$ (unless $\lambda_k^0=0$, which is not true in general; if $\lambda_k^0=0$ the analysis needs to be modified, but in fairly obvious ways).

The first integral can be done easily. The second integral is purely real (recall that for real λ_k^0 we have ϵ_k purely imaginary). Putting this all together we have

$$\tau_{kk} = \frac{1}{2} - \frac{i}{\pi}\ln|\epsilon_k| + iO(1) + O(\epsilon_k)$$

(for λ_k^0 imaginary) ,

$$\tau_{kk} = -\frac{i}{\pi}\ln|\epsilon_k| + iO(1) + O(\epsilon_k)$$

(for λ_k^0 real) ,

$$\tau_{jk} = iO(1), \quad j\neq k .$$

As $\epsilon\to 0$ the τ matrix becomes logarithmically singular along the diagonal.

What about the arguments of the theta function W_j?

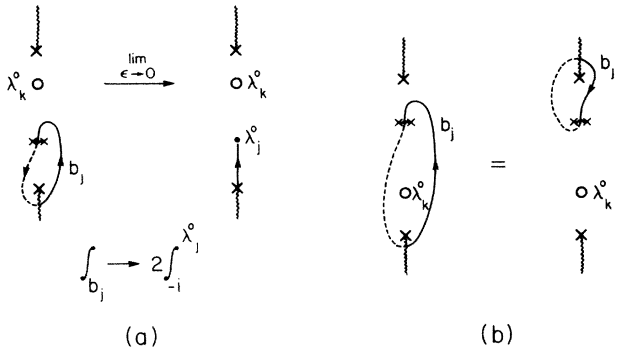

FIG. 4. Evaluation of a typical off-diagonal element of τ_{kj}.

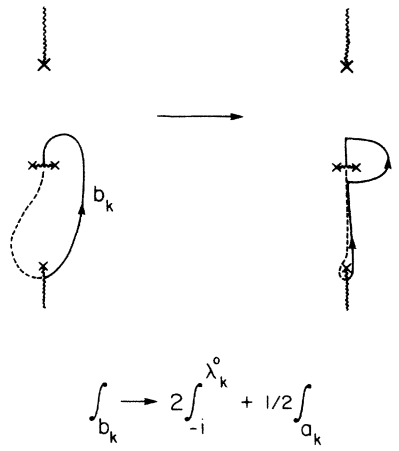

FIG. 5. Evaluation of a typical diagonal element τ_{kk}.

We have

$$W_j(x,t)=\sum_{k=1}^{N}\int_{p_0}^{\mu_k(x,t)} dU_j$$
$$=\frac{[1+(\lambda_j^0)^2]^{1/2}}{2\pi i}\sum_{k=1}^{N}\int_{p_0}^{\mu_k(x,t)}\frac{\prod_{m\,(\neq j)}(\lambda-\lambda_m^0)d\lambda}{R(\lambda)} .$$

A calculation similar to the one leading to Eq. (A18) in the Appendix shows

$$\frac{dW_j}{dx}=\frac{-1}{\pi}[1+(\lambda_j^0)^2]^{1/2} ,$$
$$k_j=-2[1+(\lambda_j^0)^2]^{1/2} , \tag{4}$$
$$\frac{dW_j}{dt}=\frac{-2[1+(\lambda_j^0)^2]^{1/2}}{\pi}\left[\tfrac{1}{2}\sum_{k=1}^{2N}\lambda_k-\sum_{m\,(\neq j)}\lambda_j^0\right] ,$$

for $j=1,\ldots,N$. (The first sum does *not* involve the branch points at $\pm i$ which have canceled out.) As $\epsilon\to 0$ almost all of the terms in the two summations cancel one another and we are left with

$$\frac{dW_j}{dt}=\frac{-2}{\pi}[1+(\lambda_j^0)^2]^{1/2}\lambda_j^0+O(\epsilon) ,$$
$$\Omega_j=-4\lambda_j^0[1+(\lambda_j^0)^2]^{1/2}+O(\epsilon) . \tag{5}$$

This leads to

$$\Omega_j=\pm k_j(k_j^2-4)^{1/2}+O(\epsilon),\quad j=1,\ldots,N . \tag{6}$$

Thus, reassuringly, we get that (as $\epsilon\to 0$) the wave number and frequency arguments of the theta function are related to one another *through the linear dispersion relation [Eq. (2)]*.

The perturbations (there are N of them) separate into two distinct types. Using Eq. (5) we see that the real-axis degeneracies have Ω_j real, while the imaginary-axis degeneracies have Ω_j imaginary. Notice these correspond precisely to the linearly stable and linearly unstable wave numbers, respectively. As we shall see in Sec. III, this is a general result. The degeneracies on the real axis are linearly stable, while those off the real axis are linearly unstable.

With all of the necessary calculations done, it is now possible to write down a solution of the CSE which, at $t=0$, looks like a slightly perturbed plane wave. Using Eq. (A31) from the Appendix we see

$$u(x,t)=e^{2it}\frac{\theta(W^+\,|\,\tau)}{\theta(W^-\,|\,\tau)} ,$$
$$W_j^{\pm}=(1/2\pi)(k_jx+\Omega_jt+\delta_j^{\pm}) . \tag{7}$$

Here (see Appendix) the external phase $\exp(ik_0x-i\omega_0t)$ has simplified to give us our zeroth-order plane wave $\exp(2it)$ and

$$\delta_j^{\pm}=-\pi\tau_{jj}+\pi+2\pi\int_{\mu_j(0,0)}^{\infty\pm} dU_j+O(\epsilon) .$$

We now wish to show that $\delta_j^{\pm}$ is real. As mentioned earlier and in the Appendix the initial points $\mu_j(0,0)$ must satisfy a constraint. The general form of the constraint is given in Refs. 11 and 19. In the present situation we know that for $u(x,0)$ to be nearly a plane wave we must have

$$\mu_j(x,0)\sim\lambda_j^0 .$$

This is because at $\epsilon_j=0$ this degree of freedom is frozen out. The position of $\mu_j(0,0)$ within the band can be found as follows. The initial conditions $\mu_j(0,0)$ and $|u(0,0)|$ must be chosen so that the function

$$f(\lambda)\equiv\sqrt{P+gh}$$

is a finite polynomial in λ (see the Appendix for the definitions of P, g, and h and Refs. 11 and 19 for proof that this constraint is necessary and sufficient). Consider the case where two conjugate imaginary-axis modulation bands are present and they are completely degenerate. The generalization to real-axis bands and more degrees of freedom is straightforward. In this situation the polynomials P, g, and h take the form

$$P(\lambda)=(1+\lambda^2)(\lambda-\lambda_1^0)^2(\lambda+\lambda_1^0)^2 ,$$
$$g=i(\lambda-\lambda_1^0)(\lambda+\lambda_1^0) ,$$
$$h=i(\lambda-\lambda_1^0)(\lambda+\lambda_1^0) .$$

Here we have put $|u(0,0)|=1$ and $\mu_1(0,0)=\lambda_1^0$, $\mu_2(0,0)=\lambda_2^0=-\lambda_1^0$; therefore,

$$f(\lambda)=\sqrt{P+gh}=\sigma\lambda(\lambda-\lambda_1^0)(\lambda+\lambda_1^0)\quad(\sigma=\pm 1) .$$

Such initial conditions and spectra have $\mu_j(x,t)=\lambda_j^0$ and $|u(x,t)|=1$ for all x and t. By breaking the degeneracy in the roots of $P(\lambda)$ while still keeping the roots of $f(\lambda)$ fixed we force the $\mu_j(0,0)$ to move away from λ_j^0. By computing their new position as a perturbation series in ϵ_j we find

$$\mu_j(0,0)=\lambda_j^0\pm[1+(\lambda_j^0)^2]^{1/2}\epsilon_j+O(\epsilon_j^2) .$$

A similar result is obtained for real-axis bands. The modifications due to more bands being present are $O(\epsilon^2)$. For small but finite ϵ_j the spatial evolution of such μ_j is a simple oscillation between the nearby branch points with a wave number determined by λ_j^0 (see Ref. 11). Because $\mu_j(0,0)$ lies near λ_j^0 we can write

$$\int_{\mu_j(0,0)}^{\infty\pm} dU_j=\int_{\mu_j(0,0)}^{\lambda_j^0} dU_j+\int_{\lambda_j^0}^{\pm\infty} dU_j .$$

For $\mu_j(0,0)$ lying on the line between λ_j^0 and $\lambda_j^0+\epsilon_j$ or $\lambda_j^0-\epsilon_j$ the first integral on the right is real. In fact, the first integral on the right is simply a fraction of the cycle a_j.

The second integral is, up to a real constant of $O(1)$, equal to $\frac{1}{2}\tau_{jj}$. This can be seen by taking the path of integration to be the real (imaginary) axis for λ_j^0 real (imaginary). The choice of sheet is fixed so that the singular part of the integral cancels the similar term in $\delta_j^{\pm}$. Therefore $\delta_j^{\pm}$ are $2N$ real constants of $O(1)$. In what follows we do not need the explicit form for $\delta_j^{\pm}$ but shall only need the property of reality.

D. Examination of the behavior of the exact solution

Let us look more closely at Eq. (7). Expanding the theta functions using the representation given in Eq. (A19) in the Appendix we see

$$\theta(W^{\pm}\mid\tau) = \sum_{m_1=-\infty}^{\infty}\cdots\sum_{m_N=-\infty}^{\infty}\exp(i\bar{m}\cdot\bar{k}x + i\bar{m}\cdot\bar{\Omega}t + i\bar{m}\cdot\bar{\delta}^{\pm} + \pi i\bar{m}\cdot\underline{\tau}\cdot\bar{m}) . \quad (8)$$

Where do the dominant terms come from in the theta series? If we take the magnitude of any individual term in the summation we find

$$|\exp(2\pi i\bar{m}\cdot\bar{W}^{\pm} + \pi i\bar{m}\cdot\underline{\tau}\cdot\bar{m})| = \exp[-\bar{m}\cdot(\mathrm{Im}\bar{\Omega})t - \pi\bar{m}\cdot(\mathrm{Im}\underline{\tau})\cdot\bar{m}] , \quad (9)$$

where $(\mathrm{Im}\underline{\tau})$ is the imaginary part of the τ matrix (which is positive definite). At $t=0$ we have

$$|\exp(2\pi i\bar{m}\cdot\bar{W}^{\pm} + \pi i\bar{m}\cdot\underline{\tau}\cdot\bar{m})| = \exp[-\pi\bar{m}\cdot(\mathrm{Im}\underline{\tau})\cdot\bar{m}] .$$

For small ϵ we get

$$\pi\bar{m}\cdot(\mathrm{Im}\underline{\tau})\cdot\bar{m} = \pi\sum_{j,k=1}^{N} m_j m_k \mathrm{Im}\underline{\tau}_{jk} \sim -\sum_{j=1}^{N} m_j^2 \ln|\epsilon_j| .$$

Therefore (at $t=0$)

$$|\exp(2\pi i\bar{m}\cdot\bar{W}^{\pm} + \pi i\bar{m}\cdot\underline{\tau}\cdot\bar{m})| \sim \sum_{j=1}^{N} e^{m_j^2\ln|\epsilon_j|} ,$$

which is $O(\epsilon)$ or smaller unless all of the m_j's are zero. Thus

$$\theta(W^{\pm}\mid\tau) = 1 + O(\epsilon)$$

at $t=0$. We can find the first-order (in each ϵ_j) contribution to the theta series easily. It comes from those terms in the expansion [Eq. (8)] which have one of the m_j's equal to ± 1 while all of the others are zero. Collecting this together we have

$$\theta(W^{\pm}\mid\tau) = 1 + 2\sum_{j=1}^{N} |\epsilon_j|\cos(k_j x + d_j^{\pm}) + O(\epsilon_j^2) ,$$

where $d_j^{\pm}$ is a set of phases. This gives us (still at $t=0$)

$$u(x,0) = \frac{\theta(W^{-}\mid\tau)}{\theta(W^{+}\mid\tau)} = 1 + 2\sum_{j=1}^{N} |\epsilon_j|\cos(k_j x + a_j) + O(\epsilon^2) ,$$

where a_j are constants. This shows that, at $t=0$, the exact solution we have constructed looks like a uniform solution ($u=1$) plus a collection of sinusoidal perturbations of $O(\epsilon)$.

As time moves forward, this situation changes. If there are imaginary-axis modes present, then the vector $\mathrm{Im}\bar{\Omega}$ will have the following form:

$$\mathrm{Im}\bar{\Omega} = (\gamma_1, -\gamma_1, \gamma_2, -\gamma_2, \ldots, \gamma_n, -\gamma_n, 0, 0, \ldots, 0) + O(\epsilon) ,$$

where the γ's represent imaginary-axis modes and the zeros represent real-axis modes. We have purposely numbered the basis differentials so the imaginary-axis modes would appear first. Let us go back to Eq. (9) but now allow $t\neq 0$. The magnitudes of the various terms in the theta series are determined by

$$\phi(\bar{m}) = \bar{m}\cdot(\mathrm{Im}\bar{\Omega})t + \pi\bar{m}\cdot(\mathrm{Im}\underline{\tau})\cdot\bar{m} .$$

The larger this term is (in a positive sense) the smaller is the contribution to the theta sum. Where is it a minimum? Let us treat $\bar{m}$ as a vector of continuous variables and find $\min\phi(\bar{m})$,

$$\frac{\partial}{\partial m_k}\phi(\bar{m}) = 0 \quad (\text{for each } m_k) .$$

This implies

$$(\mathrm{Im}\Omega_j)t + 2\pi\sum_{j=1}^{N}\mathrm{Im}\tau_{kj}m_j = 0 .$$

Therefore $\phi(\bar{m})$ has a minimum at

$$\bar{m}_0(t) = -\frac{1}{2\pi}(\mathrm{Im}\underline{\tau})^{-1}(\mathrm{Im}\bar{\Omega})t . \quad (10)$$

Since $\phi(\bar{m})$ is a quadratic form its level sets (in $\bar{m}$ space) are $(N-1)$-dimensional ellipsoids whose principle axes and radii are determined by the eigenvalues and eigenvectors of the matrix $(\mathrm{Im}\tau)$. For small enough ϵ this matrix is nearly diagonal with diagonal elements,

$$(\mathrm{Im}\tau)_{kk} \simeq -\frac{1}{\pi}\ln|\epsilon_k| + O(1) .$$

Therefore

$$(\mathrm{Im}\tau)^{-1}_{jk} \simeq -\pi(1/\ln|\epsilon_k|)\delta_{jk}$$

and

$$m_{0j}(t) \simeq \tfrac{1}{2}[(\mathrm{Im}\Omega_j)/\ln|\epsilon_j|]t .$$

This shows that in the stable subspace ($\mathrm{Im}\Omega_j = 0$) the dominant term always comes from $m_j = 0$. This means that the contributions to the theta series from the stable subspace are always $O(\epsilon)$ or smaller.

With this comment in mind we can now perform a further truncation in the theta-function representation of $u(x,t)$. Since the real-axis modes do not make a contribution greater than $O(\epsilon)$ we will neglect any contribution they make to the theta series.

To examine the recurrence properties of such solutions it is best at this point to examine a simple case in some detail. Let us assume there is only one pair of unstable modes present. The theta function in this case looks like

$$\theta(W^{\pm}\mid\tau) = \sum_{m_1=-\infty}^{\infty}\sum_{m_2=-\infty}^{\infty}\exp(i\bar{m}\cdot\bar{k}x + i\bar{m}\cdot\bar{\Omega}t + i\bar{m}\cdot\bar{\delta}^{\pm} + \pi i\bar{m}\cdot\underline{\tau}\cdot\bar{m}) . \quad (11)$$

The vectors $\bar{k}$ and $\bar{\Omega}$ look like

$$\bar{k}=k\begin{bmatrix}1\\1\end{bmatrix},\quad \bar{\Omega}=i\gamma\begin{bmatrix}1\\-1\end{bmatrix},$$

with

$$\gamma=|k(k^2-4)^{1/2}| \text{ for } k<2 .$$

The (2×2) matrix $\underline{\tau}$ takes the form

$$\underline{\tau}=\begin{bmatrix}\frac{1}{2}+\frac{i}{\pi}|\ln\epsilon| & 0\\ 0 & \frac{1}{2}+\frac{i}{\pi}|\ln\epsilon|\end{bmatrix}+iO(1),$$

where here $O(1)$ represents a real positive definite symmetric matrix with entries of order unity. The function $\phi(\bar{m})$, which determines the magnitude of the various terms in the theta series, is given by

$$\phi(\bar{m})\simeq\bar{m}\cdot(\mathrm{Im}\underline{\tau})\cdot\bar{m},$$

$$\phi(\bar{m})\simeq\gamma t(m_1-m_2)+|\ln\epsilon|(m_1^2+m_2^2).$$

This has a minimum at

$$\bar{m}_0(t)\simeq-\frac{1}{2}\frac{\gamma t}{|\ln\epsilon|}\begin{bmatrix}1\\-1\end{bmatrix}.$$

The level sets of $\phi(\bar{m})$ in this case are approximately circles centered at $\bar{m}_0(t)$. This point moves along the diagonal line indicated in Fig. 6. As shown earlier, at $t=0$ the point $\bar{m}_0(t)$ is at the origin and the only important term in the theta series comes from $\bar{m}=0$.

There are two sets of important times as $\bar{m}_0$ traces out its path: (1) when $\bar{m}_0(t)$ is at a lattice point ($t=2n|\ln\epsilon|/\gamma$) and (2) when $\bar{m}_0(t)$ is midway between two lattice points [$t=(2n+1)|\ln\epsilon|/\gamma$].

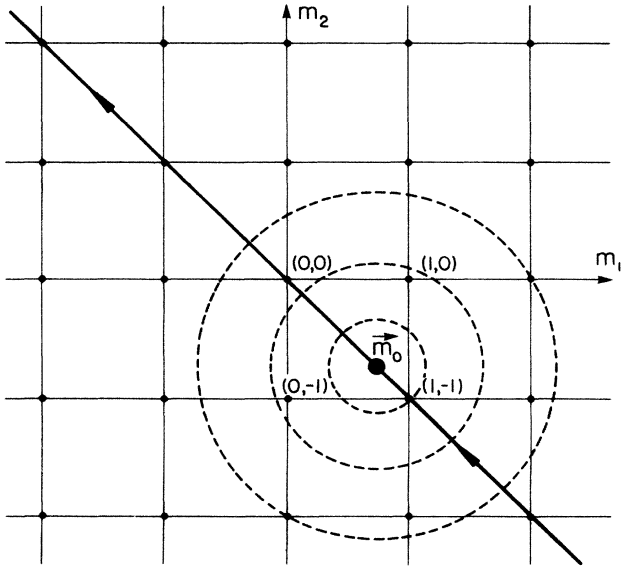

FIG. 6. Diagram of the path followed by $\bar{m}_0(t)$ in lattice space. The terms in the theta series come from the integer lattice. $\bar{m}_0(t)$ [the minimum point of $\phi(\bar{m})$] is the large dot. It moves along the line shown as time varies. The approximate level sets of $\phi(\bar{m})$ are shown as dotted circles.

When $\bar{m}_0(t)$ reaches the first lattice point we have

$$\bar{m}_0(t_1)=\begin{bmatrix}-1\\1\end{bmatrix}.$$

Going back to Eq. (10) we see that at $t=t_1$,

$$\bar{m}_0(t_1)=\frac{-1}{2\pi}(\mathrm{Im}\underline{\tau})^{-1}(\mathrm{Im}\bar{\Omega})t_1,$$

which implies

$$\mathrm{Im}\bar{\Omega}t_1=-2\pi(\mathrm{Im}\underline{\tau})\bar{m}_0=-2\pi\underline{\tau}\bar{m}_0+2\pi(\mathrm{Re}\underline{\tau})\bar{m}_0 .$$

Using

$$\mathrm{Re}\underline{\tau}=\frac{1}{2}\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

we see that $\mathrm{Im}\bar{\Omega}t_1=2\pi\tau\bar{m}_0-\pi\bar{m}_0$. Therefore

$$\bar{W}^{\pm}(x,t_1)=\bar{W}^{\pm}(x,0)+\underline{\tau}\bar{m}_0-\tfrac{1}{2}\bar{m}_0$$

and

$$\theta(W^{\pm}(x,t_1)|\tau)=\theta(W^{\pm}(x,0)+\tau m_0-\tfrac{1}{2}m_0|\tau),$$

and using a slight generalization of Eq. (A21) we find

$$\theta(W^{\pm}(x,t_1)|\tau)=\phi^{\pm}(x)\theta(W^{\pm}(x,0)-\tfrac{1}{2}m_0|\tau),$$

where

$$\phi^{\pm}(x)=\exp[-i\bar{m}_0\cdot(\bar{k}x+\bar{\delta}^{\pm})+\pi i\bar{m}_0\cdot\underline{\tau}\cdot\bar{m}_0].$$

The solution of the CSE at $t=t_1$ is [using Eq. (7)]

$$u(x,t_1)=e^{2it_1}e^{-i\bar{m}_0\cdot(\bar{\delta}^- -\bar{\delta}^+)}\times\frac{\theta(W^-(x,0)-\frac{1}{2}m_0|\tau)}{\theta(W^+(x,0)-\frac{1}{2}m_0|\tau)}.$$

Since $\bar{\delta}^{\pm}$ has real entries both terms out front are simply phases. The arguments of the theta functions are shifted by a real vector. If we ask what the dominant terms are we find, once again, that only the term $\bar{m}'=0$ [see discussion preceding Eq. (A21)] is $O(1)$, all others are $O(\epsilon)$ or smaller. The system has returned (with a phase shift) to the initial plane-wave state. Thus even though individually the theta functions have exponential growth in time, because of their special quasiperiod properties and the fact that the solution $u(x,t)$ is a *ratio* of theta functions, the initial state will recur.

At $t=2t_1$ the arguments of the theta functions will be shifted by an integer in each argument so, using Eq. (A20), the theta functions have *exactly* returned to their values at $t=0$. The external phase of $u(x,t)$ has changed, but the modulus, $|u(x,t)|$, has returned to its initial value,

$$|u(x,2t_1)|=|u(x,0)| .$$

Thus the modulus is periodic with period,

$$T=2t_1=(4\pi/\gamma)\mathrm{Im}(\tau_{11}-\tau_{12})\simeq4|\ln\epsilon|/\gamma .$$

Now consider the case when $\bar{m}_0(t)$ is midway between two lattice points. For instance, suppose $\bar{m}_0(t)$ has moved from $\bar{m}=0$ to the middle of the first adjacent square. Then, since the level sets of $\phi(\bar{m})$ are nearly circles, and since $\phi(0)=0$ we have that each corner of this first square must have $\phi(\bar{m})\simeq 0$. Thus all four terms from these lattice points in $\bar{m}$ space have unit magnitude in the theta series. This is the time when the instability "saturates." This time is

$$t_s=\mathrm{Im}(\tau_{11}-\tau_{12})/\gamma \simeq |\ln\epsilon|/\gamma ,$$

which is just $t_1/2$. This agrees with an estimate of the saturation time of an instability given by purely linear arguments. For example, suppose $a(t)$ is the amplitude of an exponentially unstable mode. Assume it starts with amplitude ϵ and has a growth rate of γ. Now ask, how much time will it take for $a(t)$ to reach order unity? We have

$$a(t)=\epsilon e^{\gamma t} ,$$

$$a(t_s)\sim 1 \rightarrow t_s \sim |\ln\epsilon|/\gamma ,$$

which agrees with the result given above.

Since we can always shift our origin in $\bar{m}$ space we choose to do it in such a way that the point $\bar{m}_0(t)$ is always inside the square formed by the origin and the three points,

$$\begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} .$$

This we call the "unit cell" and, from the arguments we have already presented, we can see that only those terms in the theta series which come from the corners of the unit cell need be evaluated. All of the other contributions are always of $O(\epsilon)$ or smaller. A plot of this solution, using the theta-function representation, is shown in Fig. 7.

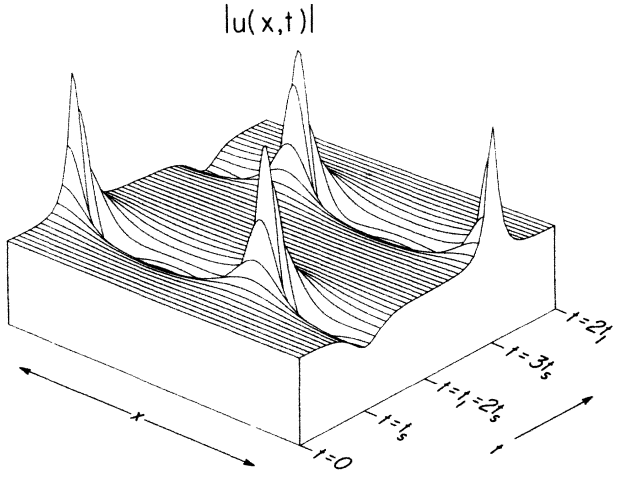

FIG. 7. Typical solution of the CSE with two unstable modes present. The characteristic times, t_s and t_1, are defined in the text.

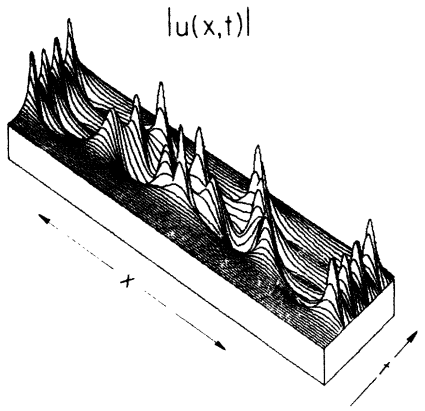

FIG. 8. Typical solution of the CSE with ten unstable modes present. Notice the recurrence property which is generic, as discussed in the text.

It should be fairly clear by now how to generalize the approach to more general cases, for example, where many unstable modes are present. It is always possible to shift the origin in lattice space so that $\bar{m}_0(t)$ lies inside the higher-dimensional unit cell. The important terms in the theta series come from the corners of this hypercube. The point $\bar{m}_0(t)$ draws out a line in lattice space but now not along a simple diagonal since its "speeds" along the various axes are, in general, irrationally related. Thus it will not strike another lattice point once it leaves the origin at $t=0$. This means that the solution $u(x,t)$ will no longer have a modulus which is periodic in time, but it will be quasiperiodic.

It is possible to estimate the recurrence time by assuming, for ϵ small enough, that the diagonal terms in the $\underline{\tau}$ matrix dominate the off-diagonal terms. The periods of the individual modes are then given by their periods without the other modes present,

$$T_j=4\frac{|\ln\epsilon_j|}{|\gamma_j|} .$$

The recurrence time is the smallest T which is (approximately) an integer multiple of each of the individual periods,

$$T=n_jT_j, \quad n_j\text{integer}, \quad j=1,\ldots,N .$$

A more accurate calculation of the recurrence time can be gotten by using Eq. (10) and asking when $\bar{m}_0(t)$ passes close to a lattice point.

As long as there are only a finite number of modes involved $u(x,t)$ will exhibit this recurrence property. From Eq. (3), for a box of finite size there will only be a finite number of unstable modes present (recall these originate from degeneracies on the imaginary axis). Since the real-axis modes remain of $O(\epsilon)$ the system recurs, and the recurrence time can be calculated.

A solution with ten unstable modes present is shown in Fig. 8. We will now discuss the general analysis of the modulational stability.

III. MODULATIONAL INSTABILITIES: THE GENERAL CASE

The analysis of the general case proceeds along the same path as that of the simpler plane wave, but the formulas are, necessarily, more complicated. An added difficulty here is the fact that finding the degenerate part of the spectrum once we know the nondegenerate part is not trivial, as in the plane-wave case. In that simpler case we could write explicit formulas for the degenerate main spectrum [see Eq. (3)] while in the more general case the degenerate spectrum is found by forcing conditions on a complicated integral expression (called Hochstadt's formula, see the following discussion). An outline of the logic of the approach may help here.

(1) Start with an arbitrary N-band solution [called $u^0(x,t)$] which corresponds to an arbitrary main spectrum with a finite number of nondegenerate spectra.
(2) Find the degenerate spectrum.
(3) Perturb the initial conditions [$u(x,0)=u^0(x,0)+\epsilon\phi(x)$]. This breaks the degeneracies.
(4) Study the behavior of the new degrees of freedom which have been introduced.

Step (1) has been discussed in the Appendix. In this section we perform steps (2) and (3) *at the same time.* This is done by examining what occurs if we open a slightly nondegenerate spectral pair at an arbitrary point on the complex plane, and showing that in general this would not represent a small perturbation. This is true because in general the k value corresponding to this degree of freedom would be complex and therefore not represent a simple oscillation. By requiring the k value associated with the new degree of freedom to be (1) real and (2) commensurate with the other k values of the system, so it is a *periodic* perturbation, we arrive at the following result: it is possible to find the degenerate spectra starting only with knowledge of the nondegenerate spectra. This is a well-known result in the theory of Hill's equation due to Hochstadt.[20,21] The *stability* of the degrees of freedom associated with the degenerate spectra is studied by then examining the time behavior associated with them [step (4)].

A. The zeroth-order N-band solution

Here, as in the simpler plane-wave case, the key is to pick the correct bases for the holomorphic differentials and period cycles. Let us suppose we start with an arbitrary N-band solution of the CSE. Call this the zeroth-order solution. Since it is an N-band solution it can be represented by

$$u^0(x,t)=e^{ik_0x-i\omega_0t}\frac{\theta^0(W^{0-}\mid\tau^0)}{\theta^0(W^{0+}\mid\tau^0)}\ .$$

The spectrum associated with $u^0(x,t)$ is the "zeroth-order spectrum" and generically will look something like Fig. 9. Degeneracies are shown as crossed circles. It should be kept in mind that $u^0(x,t)$ is completely determined by the *nondegenerate* spectrum. There are $N+1$ conjugate pairs ($2N+2$ λ_j's in all).

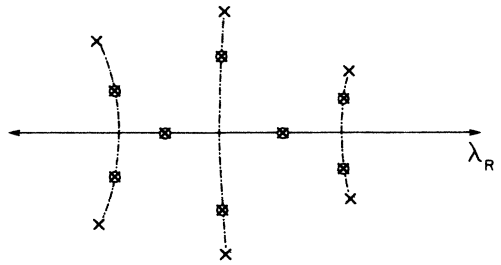

FIG. 9. Typical zeroth-order spectrum. The crosses mark nondegenerate main spectra, the dotted lines indicate spines (see discussion in the Appendix). Degenerate main spectra must appear either on the real axis or a spine. The degenerate main spectra and trapped auxiliary variables are indicated by crossed circles.

As in the plane-wave case when we perturb the initial conditions

$$u^0(x,0)\rightarrow u^0(x,0)+\epsilon\phi(x)$$

the spectrum will shift by $O(\epsilon)$. The degeneracies will be broken and new degrees of freedom injected into the system. Once again we will keep all of the complex degeneracies (which lie off the real axis) and many of the real degeneracies and show that the degrees of freedom associated with the complex degeneracies are unstable while the real ones are stable.

For the zeroth-order spectrum branch cuts are made between conjugate pairs. The a and b cycles are as shown in Fig. 10. The zeroth-order basis for the holomorphic differentials we choose to be the standard one,

$$dU_j^0\equiv\frac{\lambda^{j-1}d\lambda}{R^0(\lambda)},\quad j=1,2,\ldots,N$$

$$R^0(\lambda)\equiv\prod_{k=1}^{2N+2}(\lambda-\lambda_k)^{1/2}\ .$$

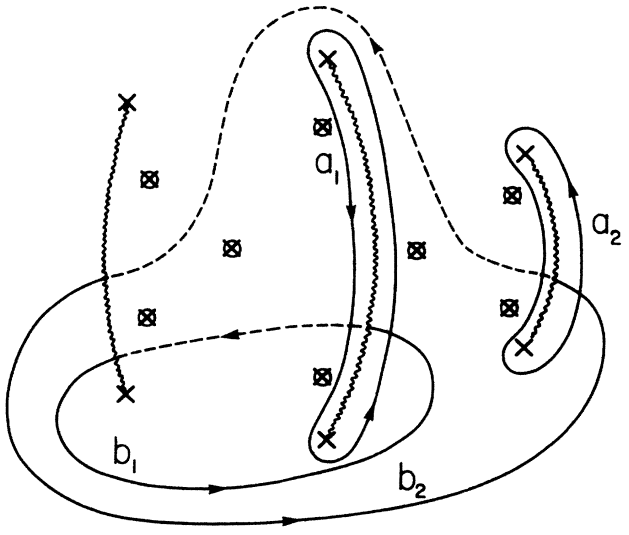

FIG. 10. Zeroth-order branch cuts and basis curves. The zeroth-order branch cuts avoid any of the degenerate points. The a cycles are chosen so as not to encircle any of the degeneracies. The b cycles are required to avoid degeneracies, but need not pass around them in any particular manner.

The zeroth-order A and B matrices are

$$A^0_{jk}=\int_{a_k} dU^0_j,\quad B^0_{jk}=\int_{b_k} dU^0_j\ ,$$

leading to

$$\underline{C}^0=(\underline{A}^0)^{-1},\quad \underline{\tau}^0=\underline{C}^0\underline{B}^0\ .$$

Keep in mind that these are $N\times N$ matrices. From matrix $\underline{C}$ it is possible to find the wave numbers and frequencies of this N-band solution of the CSE [see Eqs. (A16)–(A18)].

B. Adding new degrees of freedom

When the modulational bands are opened the a and b cycles become complicated. We have indicated in Fig. 11 what they should be. (Branch cuts are placed between nearly degenerate band pairs.) The important requirement is that a_k orbit the kth pair which is slightly nondegenerate and that b_k pass through the kth branch cut.

Now assume that there are M degenerate band pairs that are going to open up. Number these so that the *complex* degeneracies appear first and the *real* degeneracies last (assume there are M_1 and M_2 of these, respectively, so $M=M_1+M_2$). Each of these nearly degenerate pairs is centered about a point λ^0_j, $j=1,2,\ldots,M$. Each of the slightly nondegenerate band edges is measured from these points,

$$\lambda_{2j-1}=\lambda^0_j-\epsilon_j,\lambda_{2j}=\lambda^0_j+\epsilon_j\ .$$

The basis for the holomorphic differentials is

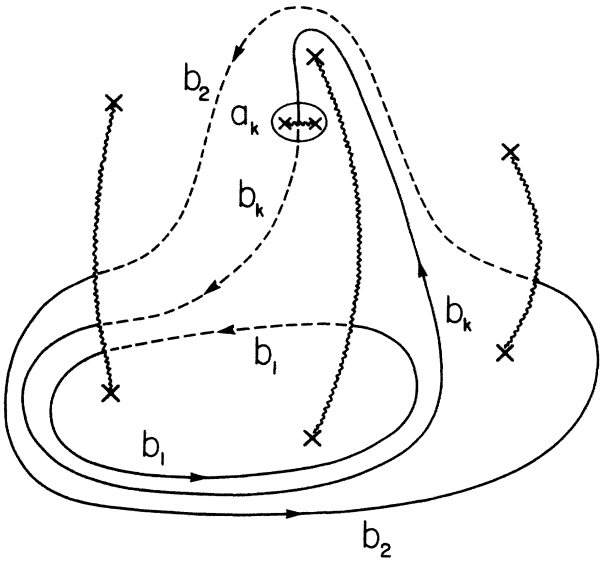

FIG. 11. Typical a and b cycles for a slightly nondegenerate spectral pair. The new branch cut is placed between the pair. The cycle a_k orbits nearby. The b_k cycle must pass through this branch cut without intersecting any other a or b cycles.

$$dU_j=\lambda^{j-1}\frac{\prod_{k=1}^{M}(\lambda-\lambda^0_k)d\lambda}{R(\lambda)},\quad j=1,2,\ldots,N$$

$$dU_j=\frac{R^0(\lambda^0_j)}{2\pi i}\frac{\prod_{k=1\ (k\neq j)}^{M}(\lambda-\lambda^0_k)d\lambda}{R(\lambda)}\ ,$$

$$j=N+1,N+2,\ldots,N+M$$

$$R(\lambda)=\prod_{k=1}^{2(M+N+1)}(\lambda-\lambda_k)^{1/2}$$

$$=R^0(\lambda)\prod_{k=2N+3}^{2(M+N+1)}(\lambda-\lambda_k)^{1/2}\ .$$

As $\epsilon\to 0$ these become

$$\lim_{\epsilon\to 0} dU_j=\frac{\lambda^{j-1}d\lambda}{R^0(\lambda)}=dU^0_j,\quad j=1,2,\ldots,N$$

$$\lim_{\epsilon\to 0} dU_j=\frac{R^0(\lambda^0_j)}{2\pi i}\frac{d\lambda}{(\lambda-\lambda^0_j)R^0(\lambda)}\ ,$$

$$j=N+1,N+2,\ldots,N+M\ .$$

Therefore [recall $\underline{A}$ and $\underline{B}$ are now $(N+M)\times(N+M)$ matrices]

$$\lim_{\epsilon\to 0} A_{jk}=\begin{cases} A^0_{jk}, & 1\le j,k\le N\\ 0, & 1\le j\le N,\ N+1\le k\le N+M\\ S_{jk}, & N+1\le j\le N+M,\ 1\le k\le N\\ \delta_{jk}, & N+1\le j,k\le N+M\ .\end{cases}$$

$\underline{S}$ is an $M\times N$ matrix of "singular periods" since these are the a periods of the singular differentials dU_j $(N+1\le j\le N+M)$. The matrix $\underline{A}$ schematically has the form

$$\underline{A}^{((N+M)\times(N+M))}=\begin{bmatrix}\underline{A}^0 & \underline{0}_{N\times M}\\ \underline{S} & \underline{1}_{M\times M}\end{bmatrix}\ .$$

[All of these results are of $O(\epsilon)$ which is *not* indicated for the sake of brevity.] Since matrix $\underline{A}$ has the form given, matrix $\underline{C}$ $(\underline{A}^{-1})$ must take the form

$$\underline{C}^{((N+M)\times(N+M))}=\begin{bmatrix}\underline{C}^0 & \underline{0}_{N\times M}\\ \underline{\Gamma} & \underline{1}_{M\times M}\end{bmatrix}\ ,$$

where $\underline{\Gamma}$ is an $M\times N$ matrix which is equal to $\underline{\Gamma}=-\underline{S}\,\underline{C}^0$. Using this full $\underline{C}$ matrix we can construct the $N+M$ normalized differentials and then W_j, which will be the argument in the new theta function,

$$W_j=\sum_{k=1}^{N+M}\sum_{m=1}^{N+M}C_{jk}\int_{p_0}^{\mu_m(x,t)}dU_k,\quad j=1,2,\ldots,N+M\ .$$

First we find $k_j = 2\pi\, dW_j/dx$,

$$\frac{dW_j}{dx} = -2i\left[\sum_{k=1}^{N} C_{jk}S_1(k) + \frac{1}{2\pi i}\sum_{k=N+1}^{N+M} R^0(\lambda_k^0)C_{jk}S_2(k)\right],$$

where

$$S_1(k) = \frac{\sum_{n=1}^{N+M}\mu_n^{k-1}\prod_{m=1}^{M}(\mu_n-\lambda_m^0)}{\prod_{r(\neq n)}(\mu_n-\mu_r)}, \quad k=1,\ldots,N$$

$$S_2(k) = \sum_{n=1}^{N+M}\frac{\prod_{m=1}^{M}(\mu_n-\lambda_m^0)}{\prod_{r(\neq n)}^{(m\neq k)}(\mu_n-\mu_r)}, \quad k=N+1,\ldots,N+M .$$

Using the same trick as earlier, convert both sums to contour integrals,

$$S_1(k) = \frac{1}{2\pi i}\int_C \lambda^{k-1}\frac{\prod_{m=1}^{M}(\lambda-\lambda_m^0)d\lambda}{\prod_{r=1}^{N+M}(\lambda-\mu_r)}$$

$$= \delta_{k,N}, \quad k=1,2,\ldots,N$$

$$S_2(k) = \frac{1}{2\pi i}\int_C \frac{\prod_{m=1}^{M}(\lambda-\lambda_m^0)d\lambda}{\prod_{r=1}^{N+M}(\lambda-\mu_r)}$$

$$= \begin{cases} 0, & N\geq 1 \\ 1, & N=0 \text{ (plane wave)} . \end{cases}$$

Therefore we get

$$\frac{dW_j}{dx} = 2i\left[C_{jN} + \frac{1}{2\pi i}\sum_{k=N+1}^{N+M} R^0(\lambda_k^0)C_{jk}\delta_{N,0}\right].$$

Using the results for the $\underline{C}$ matrix we can write this as

$$\frac{dW_j}{dx} = \begin{cases} -2iC_{jN}^0, & j=1,\ldots,N \\ -2i\Gamma_{j-N,N} - \dfrac{1}{\pi}R^0(\lambda_j^0)\delta_{N,0}, & \\ & j=N+1,\ldots,N+M \end{cases}$$

(with the understanding that when $N=0$ both $\underline{C}^0$ and $\underline{\Gamma}$ vanish). All of this leads to

$$k_j = \begin{cases} -4\pi i C_{jN}^0 (= k_j^0), & j=1,\ldots,N \\ -4\pi i\,\Gamma_{j-N,N} - 2R^0(\lambda_j^0)\delta_{N,0}, & \\ & j=N+1,\ldots,N+M . \end{cases} \tag{12}$$

Notice that when $N=0$ we recover our plane-wave result since, in that case, $R^0(\lambda_j^0)=[1+(\lambda_j^0)^2]^{1/2}$. Let us assume $N\neq 0$ for now and examine k_j, for $j>N$, in more detail. The jth normalized holomorphic differential is given by

$$d\psi_j = \sum_{k=1}^{N+M} C_{jk}\,dU_k .$$

As $\epsilon\to 0$ this becomes (for $j>N$)

$$\lim_{\epsilon\to 0} d\psi_j = \sum_{k=1}^{N+M}\lim_{\epsilon\to 0} C_{jk}\,dU_k = \sum_{k=1}^{N}\Gamma_{j-N,k}\,dU_k^0 + \frac{R^0(\lambda_j^0)}{2\pi i}\frac{d\lambda}{(\lambda-\lambda_j^0)R^0(\lambda)} . \tag{13}$$

This differential has a pole at λ_j^0 with unit residue ($+1$ on the top sheet, -1 on the bottom sheet). The differentials dU_k^0 are the standard holomorphic differentials and involve only the zeroth-order spectrum. The complex numbers $\Gamma_{j-N,k}$ are chosen to "normalize" the singular differential $d\psi_j$ so that it has zero a periods when integrated around the zeroth-order a cycles. The differential $d\psi_j$ is called a normalized differential of the third kind.[22,23] The wave number of the modulation, k_j, is determined by $\Gamma_{j-N,N}$ which in turn is determined only by the zeroth-order spectrum (λ_k, $k=1,\ldots,2N$) and the position of the pole λ_j^0. None of the other modulational modes appear, so at this stage in the calculation we can proceed as though only *one* modulational mode had been opened up.

C. Conditions for real k and periodic perturbations

If we, for the moment, consider λ_j^0 to be a free parameter then we can ask, what value do we want to choose for λ_j^0 to generate the type of solution to the CSE that we are interested in (namely, one which starts close to our zeroth-order N-band solution)? It is obvious that the first requirement is that the k_j generated by λ_j^0 is *real*. Otherwise the new theta functions would not be *spatially* stable and the modulation would not be small everywhere in x. The other requirement is that the wavelength of the modulation be some rational multiple of the original period of the zeroth-order wave so that we still have a periodic solution. This implies k_j is an integral multiple of $2\pi/d$, where d is the size of our box (the original N-band wave could, in principle, have many oscillations inside the box). This result—that given the nondegenerate spectrum we can construct the degenerate spectrum—is called Hochstadt's theorem.

A few general statements about the $\Gamma_{j-N,k}(\lambda_j^0)$ are possible. The most important is that if λ_j^0 is real $\Gamma_{j-N,k}(\lambda_j^0)$ is pure imaginary, which in turn implies that

k_j is real (as we expect). This can be shown by noticing that $R^0(\lambda)$ contains conjugate pairs of roots and that the zeroth-order a cycles have an "anti-" conjugate symmetry (for a more detailed argument see Ref. 24). This means that any λ_j^0 on the real axis would generate a modulation with real k_j. Only a discrete subset of this line will generate modulations with spatial periods which are commensurate with the original period.

There is also the possibility that k_j is real on "spines" in the complex plane. In fact, as mentioned briefly in the Appendix, each of the original points in the main spectrum which generates the zeroth-order N-band wave must be the end point of a spine. In the plane-wave case, discussed earlier, the imaginary axis between $\pm i$ is a spine. We saw in that case that modulational modes could exist on the spine. In this more general case the spines will not be so simple, but the basic idea is still the same. On a spine, k_j is real, and at discrete points on the spine it will take values which are commensurate with the period of the zeroth-order solution.

D. Discussion of the asymptotic properties of k

Some asymptotic estimates can be made of the behavior of k_j with respect to λ_j^0. These estimates will allow us to identify where the degenerate main spectra are located which correspond to very short wavelength and very long wavelength modulations.

For example, as λ_j^0 gets very large on the real axis ($|\lambda_j^0| \gg |\lambda_m|$ where λ_m is any member of the original zeroth-order nondegenerate spectrum) it is possible to get estimates of the entries of Γ as follows. Consider the normalized differential

$$d\psi_j = \sum_{m=1}^{N} \Gamma_{j-N,m} dU_m^0 + \frac{R^0(\lambda_j^0)}{2\pi i} \frac{d\lambda}{(\lambda-\lambda_j^0)R^0(\lambda)} \quad (j>N) .$$

The entries of Γ are determined by integration around the zeroth-order a cycles which, by assumption, are close to the origin in comparison with λ_j^0 (see Fig. 12). Thus we can expand the singular part of the differential assuming $|\lambda_j^0| \gg |\lambda|$ and balance term by term with the first summation,

$$\lim_{|\lambda_j^0|\to\infty} d\psi_j = \sum_{m=1}^{N} \Gamma_{j-N,m} dU_m^0 - \frac{(\lambda_j^0)^N}{2\pi i} \left[\sum_{n=0}^{\infty} \left[\frac{\lambda}{\lambda_j^0} \right]^n \right] \frac{d\lambda}{R^0(\lambda)}$$

$$= \sum_{m=1}^{N} \Gamma_{j-N,m} dU_m^0 - \frac{1}{2\pi i} \sum_{n=0}^{\infty} (\lambda_j^0)^{N-n} \frac{\lambda^n d\lambda}{R^0(\lambda)} \quad (j>N) .$$

FIG. 12. Diagram of the situation when $|\lambda_j^0| \gg |\lambda_k|$, $k=1,2,\ldots,2N+2$, to justify the expansion used in the examination of the asymptotic properties of k.

Here we have used

$$R^0(\lambda_j^0) \simeq (\lambda_j^0)^{N+1}$$

and

$$(\lambda-\lambda_j^0)^{-1} = -\frac{1}{\lambda_j^0} \sum_{n=0}^{\infty} \left[\frac{\lambda}{\lambda_j^0} \right]^n .$$

To insure $d\psi_j$ has zero a cycles we need to balance all of the large terms (those with λ_j^0 to a positive power) which means

$$\Gamma_{j-N,k} \simeq (1/2\pi i)(\lambda_j^0)^{N+1-k} + O(1) .$$

Therefore

$$k_j = -4\pi i \Gamma_{j-N,N} = -2\lambda_j^0 + O(1) ,$$

which agrees with the plane-wave result [Eq. (4)] for large λ_j^0. This is to be expected since, at large eigenvalues, the spectral problem does not "see" the potential.

Where are the long-wavelength modulations? These occur when $\Gamma_{j-N,N} \to 0$. Take the pole at λ_j^0 and move it close to one of the zeroth-order band edges. At first it looks as though we will have to pinch the a cycle between the pole and branch point but it is possible to slip the contour past the pole by adding its residue (see Fig. 13). The differential $d\psi_j$ is now nowhere singular on the contour. In the last term of $d\psi_j$, however, $R^0(\lambda_j^0) \to 0$ when λ_j^0 approaches one of the band edges. Therefore

$$\int_{a_k} d\psi_j \to 0$$

if $\Gamma_{j-N,m} \to 0$. The long-wavelength modulations arise from poles which appear close to the zeroth-order band edges. This will turn out to be an important result since it will help us prove that *any* N-band solution of the CSE is unstable *if* we go to long enough wavelengths.

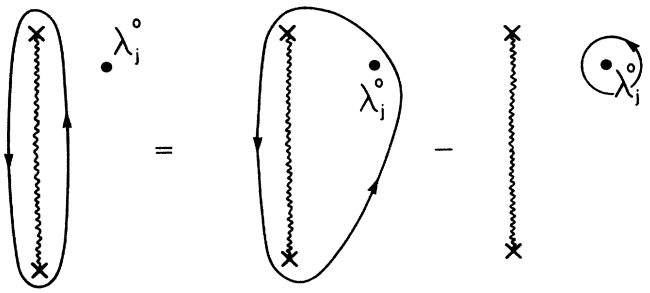

FIG. 13. Diagram of the situation when λ_j^0 approaches a zeroth-order branch point. To avoid getting a singularity on the contour, the contour is slipped past the singularity, but appropriate account must be taken of the residue, as shown.

E. Examination of the temporal behavior

Now consider the time derivative of $W_j(x,t)$,

$$\frac{dW_j}{dt}=4i\sum_{m=1}^{N}C_{jm}S_1'(m)+\frac{1}{2\pi i}\sum_{m=N+1}^{N+M}R^0(\lambda_m^0)C_{jm}S_2'(m)\ ,$$

where

$$S_1'(m)=\sum_{k=1}^{N+M}\frac{\mu_k^{m-1}\prod_{n=1}^{M}(\mu_k-\lambda_n^0)}{\prod_{r\neq k}(\mu_k-\mu_r)}\left[\frac{1}{2i}\partial_x\ln u-\mu_k\right],$$

$$S_2'(m)=\sum_{k=1}^{N+M}\prod_{\substack{n=1\\ n\neq m}}^{M}(\mu_k-\lambda_n^0)\frac{\left[\frac{1}{2i}\partial_x\ln u-\mu_k\right]}{\prod_{r\neq k}(\mu_k-\mu_r)}\ .$$

[We have used Eq. (A13) to simplify the equations slightly.]

Once again these two sums can be written as contour integrals with the result

$$S_1'(m)=\left[-\tfrac{1}{2}\sum_{j=1}^{2N+2}\lambda_j\right]\delta_{m,N}-\delta_{m,N-1}\ ,$$

$$S_2'(m)=\left[-\tfrac{1}{2}\sum_{j=1}^{2N+2}\lambda_j-\lambda_m^0\right]\delta_{N,0}-\delta_{N,1}\ .$$

Notice that the summations include only the original zeroth-order spectrum. The contributions from the modulation spectra have canceled except where explicitly indicated in $S_2'(m)$. Putting this all together leads us to

$$\frac{dW_j}{dt}=\begin{cases}-4i\left[\left[\tfrac{1}{2}\sum_{k=1}^{2N+2}\lambda_k\right]C_{j,N}+C_{j,N-1}\right], & 1\le j\le N\\ -4i\left[\left[\tfrac{1}{2}\sum_{k=1}^{2N+2}\lambda_k\right]\Gamma_{j-N,N}+\Gamma_{j-N,N-1}\right]\\ \quad+\frac{2}{\pi}R^0(\lambda_j^0)\left[\left[-\tfrac{1}{2}\sum_{k=1}^{2N+2}\lambda_k-\lambda_m^0\right]\delta_{N,0}-\delta_{N,1}\right], & N+1\le j\le N+M\ .\end{cases}$$

Notice that as $\epsilon\to 0$ the first N frequencies go over to the original frequencies of the unperturbed N-band solution [see Eq. (A18) and the immediately preceding calculation]. In the "modulation subspace" $(N+1\le j\le N+M)$ there are two special cases.

(1) When $N=0$ the plane-wave results are recovered.

(2) When $N=1$, this is the cnoidal (or traveling-wave) solution of the CSE. For this case we have

$$\frac{dW_j}{dt}=-4i\left[\tfrac{1}{2}\sum_{k=1}^{4}\lambda_k\right]\Gamma_{j-1,1}-\frac{2}{\pi}R^0(\lambda_j^0),\quad 2\le j\ .$$

Using Eq. (12) we get

$$\frac{dW_j}{dt}=\frac{1}{\pi}\left[\left[\tfrac{1}{2}\sum_{k=1}^{4}\lambda_k\right]k_j-2R^0(\lambda_j^0)\right]\ .$$

Since k_j is real and the λ_k's come in conjugate pairs the first term is real. Any instability comes from an imaginary part in the second term, since

$$R^0(\lambda_j^0)=\left[\prod_{k=1}^{4}(\lambda_j^0-\lambda_k)\right]^{1/2}\ .$$

If λ_j^0 is real $R^0(\lambda_j^0)$ is real also. Once again, all of the real-axis modulations are linearly stable. When λ_j^0 is complex $R^0(\lambda_j^0)$ will be complex, except for rare special cases. Thus the growth rate of a modulated cnoidal wave [modulated at a wave number $k_j(\lambda_j^0)$] is

$$\gamma=4\,|\,\mathrm{Im}R^0(\lambda_j^0)\,|\ .$$

Leaving aside the discussion of special cases when $N\ge 2$ we have

$$\frac{dW_j}{dt}=\frac{1}{\pi}\left[\tfrac{1}{2}\sum_{k=1}^{2N+2}\lambda_k\right]k_j-4i\Gamma_{j-N,N-1}\ ,\qquad N+1\le j\le N+M\ .$$

The first term is purely real, so any instability must come from the second term. This gives us an explicit expression for the growth rates of any stability,

$$\gamma_j=8\pi\,|\,\mathrm{Re}\Gamma_{j-N,N-1}\,|\ .$$

From our earlier discussion we know that as $\epsilon\to 0$, $\Gamma_{j-N,N-1}(\lambda_j^0)$ becomes purely imaginary [to $O(\epsilon)$] for all λ_j^0 which appear on the real axis. This means that *any double point which appears on the real λ-axis corresponds to a linearly stable modulation.*

Also, since $\Gamma_{j-N,N-1}$ is complex—in general—if λ_j^0 is complex we have that *any double point which appears off the real λ axis corresponds (except possibly for rare special cases) to a linearly unstable modulation.*

It was shown earlier that the long-wavelength modulations occur when λ_j^0 is near one of the original branch points (λ_k). It has now been shown that for λ_j^0 complex we have linear instability. This shows that *all* solutions of the CSE are linearly unstable if we allow perturbations of arbitrarily long wavelength.

On the other hand, if we fix the maximum allowed wavelength at a finite value (though large enough to allow instabilities) then it is possible to see that only a *finite* number of complex double points will appear. This is because $k_j(\lambda_j^0)$ is an analytic function of λ_j^0 (except possibly for branch-cut singularities and a pole at

$\lambda_j^0 \to \infty$). Therefore along any given spine [lines in the complex λ plane where $k_j(\lambda_j^0)$ is real] it will be well behaved and take a given value only a finite number of times as long as the spine has finite length. We shall see in a moment that this result in turn implies that unstable solutions to the CSE will always recur.

F. Examination of the behavior of the perturbed solution

Consider matrix $\underline{\tau}$. By a generalization of the arguments presented for the plane wave and our knowledge of the general structure of matrix $\underline{C}$ we can see that matrix $\underline{\tau}$ for a modulated N-band solution takes the following form:

$$\underline{\tau} = \begin{bmatrix} \underline{\tau}^0 & O(1)\underline{1}_{N\times M} \\ O(1) & [O(1)+(i/\pi)\,|\ln\epsilon_j|\,\delta_{jk}]\underline{1}_{M\times M} \end{bmatrix} + O(\epsilon).$$

Because of our choice of bases for the differentials and period cycles matrix $\underline{\tau}$ takes on a fairly simple structure in the $\epsilon \to 0$ limit. The original $N\times N$ matrix $\underline{\tau}$ associated with the unperturbed solution of the CSE appears in the upper left hand corner. The modulation subspace appears in the lower right hand corner (notice the familiar logarithmic divergence along the diagonal).

The analysis to ferret out the dominant terms in the theta series proceeds in the same manner as for the plane-wave case. It is more complicated because of the nonsingular nature of the $N\times N$ $\underline{\tau}^0$ subspace. For example, at $t=0$ we have

$$\begin{aligned}\phi(\overline{m}) &= \pi \sum_{j,k=1}^{N+M} \mathrm{Im}(\tau_{jk}) m_j m_k \\ &= \pi \sum_{j,k=1}^{N} \mathrm{Im}(\tau_{jk}^0) m_j m_k \\ &\quad + \sum_{k=N+1}^{N+M} |\ln\epsilon_k|\, m_k^2 + \zeta(\overline{m}) \ .\end{aligned}$$

We have grouped the terms as follows.

(1) All of those terms in the sum which involve only contributions from the original matrix $\underline{\tau}^0$ (first sum).

(2) The dominant (diagonal) terms from the modulational subspace (second sum).

(3) Everything else [definition of $\zeta(\overline{m})$].

The function $\phi(\overline{m})$ is large and positive (and hence corresponds to a negligible term in the theta series) if any m_k with $N+1 \le k \le N+M$ is not zero. With all such m_k set to zero, $\zeta(\overline{m})$ is also zero. This leaves us only with the first summation, which is identical to the sum for the unperturbed N-band solution of the CSE. This means

$$\begin{aligned}\theta(x,0) &= \sum_{m_{N+M}=-\infty}^{\infty} \cdots \sum_{m_N=-\infty}^{\infty} \cdots \sum_{m_1=-\infty}^{\infty} \exp(2\pi i\overline{m}\cdot\overline{W}^{\pm} + \pi i\overline{m}\cdot\underline{\tau}\cdot\overline{m}) \\ &= \sum_{m_N=-\infty}^{\infty} \cdots \sum_{m_1=-\infty}^{\infty} \exp(2\pi i\overline{m}\cdot\overline{W}^{\pm 0} + \pi i\overline{m}\cdot\underline{\tau}^0\cdot\overline{m}) + O(\epsilon) \\ &= \theta^0(x,0) + O(\epsilon) \ .\end{aligned}$$

Once again as time progresses new terms in the theta series will become important. However, it is still true that, because the modulations corresponding to real λ_j^0 have $\gamma_j = 0$, all of the terms which the real-axis modes contribute to the theta series are always $O(\epsilon)$ or smaller. This means, if we are interested solely in the instability, we can neglect those real-axis modes entirely, just as in the plane-wave case. Since C_{jk}^0 is pure imaginary we know that

$$\mathrm{Im}\overline{\Omega} = (0,0,\ldots,0;\ \gamma_1,\gamma_2,\ldots,\gamma_{M_1};\ 0,0,\ldots,0) \ ,$$

where the first zeros are the N-dimensional zeroth-order subspace, the γ's are M_1-dimensional unstable subspace, and the last zeros are the M_2-dimensional stable subspace. Thus the minimum value of $\phi(\overline{m})$,

$$\phi(\overline{m}) = \overline{m}\cdot(\mathrm{Im}\overline{\Omega})t + \pi\overline{m}\cdot(\mathrm{Im}\underline{\tau})\cdot\overline{m}$$

comes at $\overline{m}_0(t)$ which is given by

$$\overline{m}_0(t) = -\frac{1}{2\pi}(\mathrm{Im}\underline{\tau})^{-1}(\mathrm{Im}\overline{\Omega})t \ .$$

This moves linearly in time but only through the unstable subspace. Thus the instability will saturate and demodulate when $\overline{m}_0(t)$ has traveled from the origin [$\overline{m}_0(0)=0$] to the vicinity of another lattice point in $\overline{m}$ space. The reason for this is the same as for the earlier plane-wave case. When $\overline{m}(t_D)$ is near a lattice point we can expand the theta function about this new origin and, using the period properties of the theta function show that the new theta function looks like an N-band theta function plus some $O(\epsilon)$ contributions from the modulations. It is *not* the same as the N-band theta function at $t=0$ because the original solution to the CSE had N temporal periods itself, which must be taken into account when discussing recurrence.

Suppose the "demodulation" time is t_D, and that the original unperturbed N-band solution of the CSE had a recurrence time of t_R^0. Then the full system, zeroth-order and with the modulations added, will recur when $n_1 t_D + n_2 t_R^0 = 0$ for some integers n_1 and n_2. As long as the theta functions are finite dimensional recurrence shall occur. Since we have already shown that only the complex modulational modes are unstable and that there will only be a finite number of those (for a system with finite spatial period) this implies recurrence in general.

ACKNOWLEDGMENTS

One of the authors (E.R.T.) wishes to acknowledge the support of the physics department of the College of William and Mary. This work was also supported, in part, by the National Science Foundation and the U.S. Office of Naval Research.

APPENDIX

The inverse-scattering transformation (IST) first appeared in Ref. 13 where it was developed to solve the Korteweg–de Vries equation. The necessary modifications for application to the CSE were reported in Ref. 12.

The IST method of solution starts by introducing a pair of linear operators, called the Lax pair.[25,26] In this case,

$$L \equiv \begin{bmatrix} i\partial_x & u(x,t) \\ -u^*(x,t) & -i\partial_x \end{bmatrix},$$

$$A \equiv \begin{bmatrix} i\,|u|^2-2i\lambda^2 & -u_x+2i\lambda u \\ u_x^*+2i\lambda u^* & -i\,|u|^2+2i\lambda^2 \end{bmatrix}.$$

These two operators act on a two-component wave function ϕ such that

$$L\phi \equiv \lambda\phi \quad \text{and} \quad \phi_t=(\partial/\partial_t)\phi \equiv A\phi, \quad \phi \equiv \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}.$$

In order for there to be a solution to these two operator equations we must have $\phi_{xt}=\phi_{tx}$. A little algebra shows that this is true if and only if $u(x,t)$ satisfies the CSE.

The next step is to use the initial conditions for $u(x,t)$ in the operator L and solve the eigenvalue problem. In other words, treat the operator equation

$$L\phi=\lambda\phi \tag{A1}$$

as a time-independent scattering problem and find the spectrum associated with $u(x,0)$. In the present case u is periodic so $u(x+d,0)=u(x,0)$. Pick a base point $x=x_0$ and introduce the two independent solutions of Eq. (A1) which take the following values at $x=x_0$:

$$\phi(x_0)=\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \tilde{\phi}(x_0)=\begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The solution matrix $\Phi(x,x_0;\lambda)$ is given by

$$\Phi(x,x_0;\lambda) \equiv \begin{bmatrix} \phi_1(x,x_0;\lambda) & \tilde{\phi}_1(x,x_0;\lambda) \\ \phi_2(x,x_0;\lambda) & \tilde{\phi}_2(x,x_0;\lambda) \end{bmatrix}.$$

It satisfies

$$L\Phi=\lambda\Phi, \quad \Phi(x_0,x_0;\lambda)=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{A2}$$

The Wronskian of any two solutions is defined as $W(f,g) \equiv f_1g_2-g_1f_2$. Therefore

$$W(\phi,\tilde{\phi})=\det(\Phi).$$

Using Eq. (A1) it is easy to show $\partial_x W=0$ which means $\det(\Phi(x))=\det(\Phi(x_0))=1$. Using Eq. (A2) we can find $\Phi(x)$ for any x. In particular, we can find $\Phi(x_0+d,x_0;\lambda)$. This matrix

$$\Phi(x_0+d,x_0;\lambda)=\begin{bmatrix} \phi_1(x_0+d,x_0;\lambda) & \tilde{\phi}_1(x_0+d,x_0;\lambda) \\ \phi_2(x_0+d,x_0;\lambda) & \tilde{\phi}_2(x_0+d,x_0,\lambda) \end{bmatrix} \tag{A3}$$

is called the monodromy, or transfer, matrix and we will represent it by

$$M(x_0;\lambda) \equiv \Phi(x_0+d,x_0;\lambda).$$

If we change our base point to x_0' the new functions

$$\phi(x,x_0';\lambda) \quad \text{and} \quad \tilde{\phi}(x,x_0';\lambda)$$

are simple linear combinations of the original ϕ and $\tilde{\phi}$. This means $x_0 \to x_0'$ results in a change of basis and that $M(x_0',\lambda)$ is related to $M(x_0,\lambda)$ by a similarity transformation,

$$M(x_0',\lambda)=SM(x_0,\lambda)S^{-1}.$$

The trace and determinant are preserved under similarity transformations so

$$[\mathrm{Tr}M](x_0,\lambda)=[\mathrm{Tr}M](\lambda) \equiv \Delta(\lambda),$$

$$\det M=1.$$

The function $\Delta(\lambda)$, which is independent of x_0, is called the discriminant and is central to understanding the spectral properties of L. We can see this by constructing the Bloch (or Floquet) solutions of Eq. (A1). The Bloch functions have the property

$$\psi(x+d;\lambda)=e^{ip(\lambda)}\psi(x;\lambda),$$

where $p(\lambda)$ is called the Floquet exponent or the "quasi-momentum." One important result of Floquet theory[27] is that for every λ (we now think of λ as a complex parameter) there exists at least one such solution. Since such a solution can be expressed as a linear combination of our standard solutions (ϕ and $\tilde{\phi}$) we have

$$\psi(x;\lambda)=A\phi(x;\lambda)+B\tilde{\phi}(x;\lambda).$$

From Eq. (A2) we know

$$\psi(x_0;\lambda)=A\phi(x_0;\lambda)+B\tilde{\phi}(x_0;\lambda)=\begin{bmatrix} A \\ B \end{bmatrix}. \tag{A4}$$

We want

$$\psi(x_0+d;\lambda)=m(\lambda)\psi(x_0;\lambda). \tag{A5}$$

Using Eqs. (A4) and (A5) we find

$$m(\lambda)\begin{bmatrix} A \\ B \end{bmatrix}=M\begin{bmatrix} A \\ B \end{bmatrix},$$

where M is the monodromy matrix. This implies that we can construct the Bloch eigenfunctions and find their related Floquet exponents by finding the eigenvectors

and eigenvalues of the monodromy matrix. The eigenvalues are given by

$$\det(M-m)=m^2-(\mathrm{Tr}M)m+\det M$$
$$=m^2-\Delta(\lambda)m+1 .$$

Therefore,

$$m^{\pm}(\lambda)=\frac{\Delta(\lambda)\pm(\Delta^2-4)^{1/2}}{2} . \tag{A6}$$

Since the scattering problem [Eq. (A1)] is not self-adjoint we shall have to study the properties of $\Delta(\lambda)$ [and hence of $m^{\pm}(\lambda)$] on the complex λ plane. For details and proofs of the following assertations see Ref. 27 for an introduction to Floquet theory, Ref. 16 for its application to the focusing and defocusing CSE, and Ref. 24 for a related discussion of the scattering problem for the sine-Gordon equation (which is also non-self-adjoint).

The discriminant, $\Delta(\lambda)$, considered as a function of the complex variable λ is analytic in the finite complex plane. This means $\Delta(\lambda)$ will be real (i.e., $\mathrm{Im}[\Delta(\lambda)]=0$) along one-dimensional curves in the complex λ plane. The real λ axis is one such curve. Along such curves, there will be three distinct regions: (I) $\Delta^2(\lambda)<4$, (II) $\Delta^2(\lambda)=4$, (III) $\Delta^2(\lambda)>4$.

Region (I) is termed the region or band of stability because when $\Delta^2<4$ the Floquet multiplier $m(\lambda)$ [see Eq. (A6)] is a complex number whose magnitude is unity. This means the Bloch eigenfunctions are stable under translation. The entire real λ axis is a band of stability. Other stable bands on the complex plane are called spines.

Region (II) consists of discrete points where $\Delta(\lambda)=\pm 2$. At these points $m(\lambda)=\pm 1$ and the Bloch eigenfunctions are periodic or antiperiodic. This set of discrete values of λ is called the *main spectrum* and is symbolized by $\{\lambda_j\}$.

Region (III) has $m(\lambda)$ real, but $|m(\lambda)|\neq 1$, which implies that the Bloch eigenfunctions are unstable to translations either to the right or the left along the x axis.

The rest of the complex λ plane (where $\mathrm{Im}[\Delta(\lambda)]\neq 0$) is unstable. Since $\Delta(\lambda)$ is real for real λ we know

$$[\Delta(\lambda)]^*=\Delta(\lambda^*)$$

(where the asterisk means complex conjugation). Therefore, if λ lies in a band of stability or is a point in the main spectrum, then λ^* will be also. Thus the spectrum is symmetric under complex conjugation.

For all of the regions of the complex λ plane *except* where $\Delta=\pm 2$ the two eigenvalues of the monodromy matrix M are distinct. This means the eigenvectors will be independent and there exist two independent Bloch eigenfunctions

$$\psi^+(x+d)=m^+\psi^+(x),\psi^-(x+d)=m^-\psi^-(x) .$$

However, when $\Delta(\lambda)=\pm 2$ the two eigenvalues are no longer distinct and it is not true in general that *two* independent Bloch eigenfunctions exist.

Definition. If λ_j is a point in the main spectrum which has *two* independent Bloch eigenfunctions, then λ_j is said to be a *degenerate* eigenvalue. If λ_j is a point in the main spectrum which has only one Bloch eigenfunction then λ_j is said to be a *nondegenerate* eigenvalue.

This distinction is important for understanding the origin of the instability for the CSE. It is shown in the main text that the existence and location of degeneracies in the main spectrum determine whether or not a given solution is unstable.

The spines of stable λ can end only at nondegenerate eigenvalues. A potential $u(x,0)$ which has no degenerate eigenvalues we will call "generic" (following Ref. 28). A spectrum with degeneracies is "nongeneric" in the sense that if we change it by adding a small perturbation $[u(x,0)\rightarrow u(x,0)+\epsilon\phi(x)]$ then the degeneracy will be broken in general.

An important consequence of the fact that the CSE is the compatibility condition for the operators L and A is that if we now let $u(x,0)$ be an initial condition for the CSE and evolve it forward in time,

$$u(x,0)\xrightarrow{\text{CSE}}u(x,t) .$$

It can be shown[16] that the discriminant is invariant. Thus all of the spectral structure deduced from $\Delta(\lambda)$, the main spectrum and the stable bands, for example, are invariant. The eigenvalues $\{\lambda_j\}$ are constants of the motion for the CSE.

An important class of initial conditions $u(x,0)$ consists of those which generate only a *finite* number of nondegenerate eigenvalues λ_j. These are called finite band potentials and they are central to our study.

The simplest method for generating solutions starts by first constructing the "squared" eigenfunctions. As we shall see the squared eigenfunctions depend only on the nondegenerate part of the spectrum, which simplifies many of the computations. Later we will study how these special solutions behave under perturbations and find that if there are any complex degeneracies in the spectrum the solution will be unstable to perturbations. Degeneracies on the real axis will play *no* role in the instability.

To construct the squared eigenfunctions we start with our two Bloch eigenfunctions, ψ^+ and ψ^-. If there is only one Bloch eigenfunction we square that. Define

$$f(x,t;\lambda)\equiv-\frac{i}{2}(\psi_1^+\psi_2^-+\psi_2^+\psi_1^-) ,$$

$$g(x,t;\lambda)\equiv\psi_1^+\psi_1^- ,$$

$$h(x,t;\lambda)\equiv-\psi_2^+\psi_2^- .$$

From this definition we see

$$f(x+d,t;\lambda)=f(x,t;\lambda)$$

for *all* λ. The same is true for g and h. From Eq. (A1) and the definition of $L[u]$ and A, it is easy to show

$$
\begin{aligned}
f_x &= u^* g - u h \ , \\
g_x &= -2i\lambda g - 2u f \ , \qquad \text{(A7)} \\
h_x &= 2i\lambda h + 2u^* f \ ,
\end{aligned}
$$

$$
\begin{aligned}
f_t &= -i(u_x^* + 2i\lambda u^*) g + i(-u_x + 2i\lambda u) h \ , \\
g_t &= 2i(|u|^2 - 2\lambda^2) g + 2i(-u_x + 2i\lambda u) f \ , \qquad \text{(A8)} \\
h_t &= -2i(|u|^2 - 2\lambda^2) h - 2i(u_x^* + 2i\lambda u^*) f \ .
\end{aligned}
$$

The condition on $u(x,0)$ for there to be only a finite number of nondegenerate spectra is equivalent to the requirement that f, g, and h be finite-order polynomials in λ,[11]

$$
\begin{aligned}
f(x,t;\lambda) &= \sum_{j=0}^{N+1} f_j(x,t)\lambda^j \ , \\
g(x,t;\lambda) &= \sum_{j=0}^{N+1} g_j(x,t)\lambda^j \ , \qquad \text{(A9)} \\
h(x,t;\lambda) &= \sum_{j=0}^{N+1} h_j(x,t)\lambda^j \ .
\end{aligned}
$$

The exact form of these conditions are given in Ref. 11. They also are given in Ref. 16. We will not need the conditions here.

Equations (A7) and (A8) have one immediate integral,

$$\partial_x(f^2 - gh) = \partial_t(f^2 - gh) = 0 \ .$$

By using the expansions given in Eq. (A9) in Eq. (A7) it can be shown[11] that both g and h must be of order N in λ. Since g and h are both of order N in λ the coefficient of λ^{2N+2} in $f^2 - gh$ is simply f_{N+1}^2. This means f_{N+1} is a constant. Without any loss of generality we can set $f_{N+1} = 1$ since changing f_{N+1} simply results in multiplying f, g, and h by a constant. The coefficient of λ^N in g (using $f_{N+1} = 1$) is $iu(x,t)$. This means if we write g as a finite product it takes the form

$$g(x,t;\lambda) = iu(x,t) \prod_{j=1}^{N} [\lambda - \mu_j(x,t)] \ .$$

By similar arguments we find that the coefficient of λ^N in h is $iu^*(x,t)$. The zeros of h are μ_j^*. If we look back to the original definitions of f, g, and h we find that $f^2 - gh$ is simply related to the Wronskian of ψ^+ and ψ^-,

$$f^2 - gh = -\tfrac{1}{4}[W(\psi^+, \psi^-)]^2 \ .$$

The function $f^2 - gh$ is a $(2N+2)$th-order polynomial in λ with constant coefficients,

$$f^2 - gh \equiv P(\lambda) = \sum_{k=0}^{2N+2} P_k \lambda^k = \prod_{j=1}^{2N+2} (\lambda - \lambda_j) \ . \qquad \text{(A10)}$$

We know from our earlier discussion that ψ^+ and ψ^- are linearly independent *except* at nondegenerate eigenvalues. This implies that, in fact, the zeros of $P(\lambda)$ are the nondegenerate eigenvalues. These must appear in complex conjugate pairs so $P(\lambda)$ has real coefficients.

If we evaluate Eq. (A10) at one of the zeros of g (μ_m, for example) then we arrive at the identity

$$f(x,t;\mu_m) = \sigma_m \sqrt{P(\mu_m)} \ , \qquad \text{(A11)}$$

where σ_m is a sheet index that indicates which sheet of the Riemann surface associated with $\sqrt{P(\lambda)}$ the complex μ_m lies on. We shall have more to say about this in a moment.

Using Eq. (A7) we can evaluate g_x at one of the zeros of g to find

$$\mu_{mx} = \frac{-2i\sigma_m \sqrt{P(\mu_m)}}{\prod_{n \neq m} (\mu_m - \mu_n)}, \quad m = 1,2,\ldots,N \ . \qquad \text{(A12)}$$

If we look at the λ^N term for g_x in Eq. (A7) we find

$$iu_x = -2ig_{N-1} - 2uf_N \ .$$

Using the product expansion of g leads to

$$\partial_x \ln u = 2i \left[\sum_{j=1}^{N} \mu_j + f_N \right] .$$

Now use Eq. (A10) to find f_N,

$$f_N = -\tfrac{1}{2} \sum_{k=1}^{2N+2} \lambda_k \ .$$

Putting this all together we have

$$\partial_x \ln u = 2i \left[\sum_{j=1}^{N} \mu_j - \tfrac{1}{2} \sum_{k=1}^{2N+2} \lambda_k \right] . \qquad \text{(A13)}$$

Applying the same approach to g_t in Eq. (A8) we can show

$$\mu_{jt} = -2 \left[\sum_{m \neq j} \mu_m - \tfrac{1}{2} \sum_{k=1}^{2N+2} \lambda_k \right] \mu_{jx} \ , \qquad \text{(A14)}$$

$$
\begin{aligned}
\partial_t \ln u = 2i &\left[\sum_{j>k} \lambda_j \lambda_k - \tfrac{3}{4} \left[\sum_{k=1}^{2N+2} \lambda_k \right]^2 \right] \\
&- 4i \left[\left[-\tfrac{1}{2} \sum_{k=1}^{2N+2} \lambda_k \right] \left[\sum_{j=1}^{N} \mu_j \right] + \sum_{j>k} \mu_j \mu_k \right] .
\end{aligned} \qquad \text{(A15)}
$$

If we knew the initial conditions for $\mu_j(0,0)$ and $|u(0,0)|$ we could proceed by integrating Eqs. (A12) and (A14) to find $\mu_j(x,t)$ and then integrate Eqs. (A13) and (A15) to find $u(x,t)$. In fact we do not know the initial conditions for a simple reason: Eqs. (A12)–(A15) are valid *only* for N-band potentials. How can we construct such potentials? *By solving Eqs. (A12)–(A15) with appropriate initial conditions.* The argument is circular. We can avoid this by shifting our point of view. First, take the parameters $\{\lambda_j \mid j = 1,2,\ldots,2N+2\}$ as known. Second, choose initial conditions $\mu_j(0,0)$ and $|u(0,0)|$. Third, solve Eqs. (A12)–(A15). We shall see that this can be done analytically and results in a closed-form expression for $u(x,t)$.

How do we accomplish the second step? It is at this point that a technical difficulty arises. We cannot

choose these initial conditions freely; they must satisfy a constraint. This is due to the fact that the system has $N+1$ real degrees of freedom (see Ref. 11, Chap. 4) while the N complex $\mu_j(0,0)$'s and $|u(0,0)|$ constitute $2N+1$ real parameters. N of these parameters must be constrained. A solution of this problem is presented in Refs. 11 and 19. An independent solution of the constraint problem using a very different approach was developed by Previato.[29]

A similar constraint problem also appears in the periodic sine-Gordon equation where it was solved in different ways by Ercolani and Forest,[28] and Ting, Tracy, and Chen.[30] The constraint problems for the Liouville and sinh-Poisson equations were solved in Refs. 31 and 32 respectively. The same constraint appears in the work of Bogolyubov and Prikarpatskii on a discrete form of the nonlinear Schrödinger equation[33] and the modified nonlinear Schrödinger equation.[34]

Consider Eq. (A10). Suppose we choose arbitrary initial conditions $\mu_j(0,0)$ and $|u(0,0)|$ and from these construct the polynomials g and h. Now construct

$$Y(\lambda)\equiv\sqrt{P+gh}\ .$$

A moment's reflection shows that, for most choices of zeros for g and h, the function $Y(\lambda)$ will not be a finite-order polynomial in λ. This is the constraint needed. The initial conditions $\mu_j(0,0)$ and $|u(0,0)|$ must be chosen in such a way to insure that $Y(\lambda)$ is a finite-order polynomial in λ.[11,19]

Therefore, using the approach outlined above, choose an appropriate set of initial conditions for the $\mu_j(0,0)$'s and $|u(0,0)|$. Now solve Eqs. (A12)–(A15). The crucial observation, first made by Its, Matveev, and Novikov (see Ref. 35 and papers cited therein) is that the μ_j's should be treated as points lying on the two-sheeted Riemann surface associated with

$$\sqrt{P(\lambda)}=\left[\prod_{k=1}^{2N+2}(\lambda-\lambda_k)\right]^{1/2}\equiv R(\lambda)\ ,$$

which has branch points at each of the nondegenerate band edges, λ_k. For $R(\lambda)$ to be a well-defined function we must also specify which sheet of the Riemann surface λ is on. This is taken care of by the sheet index σ. (For more details see Ref. 11. For an introduction to Riemann surfaces and hyperelliptic-function theory see Refs. 22 and 23.)

On this two-sheeted Riemann surface, which we will denote by M, it is possible to find N linearly independent holomorphic (regular) differentials. For the present, we will choose as our basis the following set:

$$dU_1\equiv\frac{d\lambda}{R(\lambda)}\ ,$$

$$dU_2\equiv\frac{\lambda d\lambda}{R(\lambda)}\ ,$$

$$\vdots$$

$$dU_j\equiv\frac{\lambda^{j-1}d\lambda}{R(\lambda)}\ ,$$

$$\vdots$$

$$dU_N\equiv\frac{\lambda^{N-1}d\lambda}{R(\lambda)}\ .$$

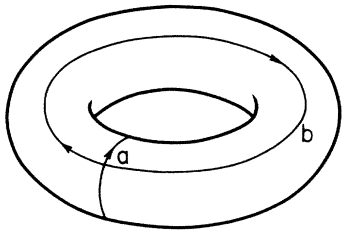

FIG. 14. Riemann surface M for $N=1$ is topologically equivalent to a torus. The curves a and b form a basis for closed curves on M.

These differentials are regular everywhere on M. At each branch point of $R(\lambda)$ they have an integrable singularity (as long as no two branch points coincide). Near $\lambda\to\infty$,

$$\lim_{\lambda\to\infty}dU_j=\lim_{\lambda\to\infty}\frac{\lambda^{j-1}}{R(\lambda)}d\lambda=\lambda^{j-2-N}d\lambda\ ,$$

which is integrable for $j\le N$, as in the present case. On the surface M there are $2N$ topologically distinct closed curves (or cycles). For example, when $N=1$ the surface M is topologically equivalent to a torus. A torus has two distinct closed curves which cannot be continuously deformed into each other or shrunk to zero, see Fig. 14. For more general N the surface M can be shown to be topologically equivalent to a sphere with N handles sewn onto it. The number N is called the genus of the surface M. Notice that it equals the number of independent holomorphic differentials. On this more general surface the $2N$ distinct closed curves are split into two classes, denoted a_j cycles and b_j cycles ($j=2,\ldots,N$). Each of these cycles has a specified direction in which it is to be traversed (e.g., in Fig. 14 we have attached arrows to the cycles to indicate that these curves also carry a specified orientation). The rules for constructing a and b cycles are as follows.

(1) a_j cycles do not cross any other a_j cycles; b_j cycles do not cross any other b_j cycles.

(2) The cycle a_k intersects b_k only once and intersects no other b cycle. The intersection is such that at the point where they meet, if we imagine deforming them so they meet in a right angle, the vector tangent to a_k maps into the vector tangent to b_k after rotation by $\pi/2$ counterclockwise (see Fig. 15).

This is symbolized as

$$a\circ a=0,\quad b\circ b=0\ ,$$

$$a_j\circ b_k=\delta_{jk}\ .$$

Using the a and b cycles we can construct the A and B period matrices,

$$A_{kj}\equiv\int_{a_j}dU_k,\quad B_{kj}\equiv\int_{b_j}dU_k\ ,$$

where dU_k are the holomorphic differentials given earlier. It is possible to show[22,23] that the matrices $\underline{A}$ and $\underline{B}$ are invertible. We shall only need $\underline{A}^{-1}$. Now introduce

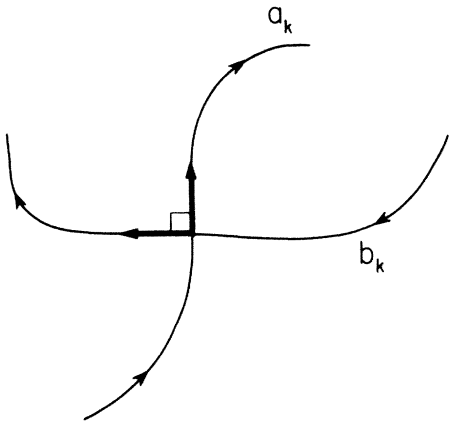

FIG. 15. "Canonical" intersection of an a and b cycle.

the $\underline{C}$ and $\underline{\tau}$ matrices defined by

$$\underline{C}=\underline{A}^{-1},\quad \underline{\tau}=\underline{A}^{-1}\underline{B}\ .$$

The $\underline{\tau}$ matrix can be shown to be symmetric ($\tau_{jk}=\tau_{kj}$) and it has positive definite imaginary part ($\mathrm{Im}\underline{\tau}>0$).

If we change our holomorphic basis of differentials to the new basis,

$$d\psi_j\equiv\sum_{k=1}^{N}C_{jk}dU_k\ .$$

We find

$$\int_{a_n}d\psi_j=\sum_{k=1}^{N}C_{jk}\int_{a_n}dU_k=\sum_{k=1}^{N}C_{jk}A_{kn}=\delta_{jn}\ .$$

Also

$$\int_{b_n}d\psi_j=\tau_{jn}\ .$$

These new differentials and the τ matrix will prove crucial to finding the solution of the CSE in closed form.

Using these differentials construct a change of variables, called an Abel map, from the Riemann surface M onto C^N (the space of complex n-tuples of dimension N). This is done as follows. Choose a base point on M and call it p_0. Define the variables $W_j(x,t)$ as follows:

$$W_j(x,t)=\sum_{k=1}^{N}\int_{p_0}^{\mu_k(x,t)}d\psi_j$$
$$=\sum_{k=1}^{N}\sum_{m=1}^{N}C_{jm}\int_{p_0}^{\mu_k}\frac{\lambda^{m-1}d\lambda}{R(\lambda)}\ . \qquad \text{(A16)}$$

These variables have the remarkable property that their x and t dependence is trivial,

$$\frac{d}{dx}W_j(x,t)=\sum_{k=1}^{N}\sum_{m=1}^{N}C_{jm}\frac{\mu_k^{m-1}\mu_{kx}}{\sigma_k R(\mu_k)}\ .$$

Using

$$\mu_{kx}=\frac{-2i\sigma_k R(\mu_k)}{\prod_{n\neq k}(\mu_k-\mu_n)}$$

we find

$$\frac{d}{dx}W_j=\sum_{m=1}^{N}C_{jm}\sum_{k=1}^{N}\frac{-2i\mu_k^{m-1}}{\prod_{n\neq k}(\mu_k-\mu_n)}\ .$$

The summation

$$\sum_{k=1}^{N}\frac{\mu_k^{m-1}}{\prod_{n\neq k}(\mu_k-\mu_n)}$$

can be performed with the help of the following contour integral:

$$I_m\equiv 1/2\pi i\int_C\frac{\lambda^{m-1}d\lambda}{\prod_{n=1}^{N}(\lambda-\mu_n)}\ , \qquad \text{(A17)}$$

where the contour C encloses all of the poles μ_n in a counterclockwise manner. It is easily seen by the residue theorem that

$$\sum_{k=1}^{N}\frac{\mu_k^{m-1}}{\prod_{n\neq k}(\mu_k-\mu_n)}=I_m\ .$$

The integral [Eq. (A17)] can also be evaluated by a different route. Consider Fig. 16. If we consider the complex λ plane as a sphere then we see there are two ways of evaluating I_m. One is to shrink the contour C around each of the poles at μ_k. The residue theorem then leads to the summation of interest. On the other hand, we can try to shrink the contour around the other side of the sphere. This leads to an evaluation of the residue at $z=\infty$. The contour orbits the pole in a clockwise manner, but that is taken care of by the minus sign in the variable transformation,

$$t\equiv\frac{1}{\lambda}\rightarrow\frac{dt}{t}=-\frac{d\lambda}{\lambda}\ .$$

Evaluating the residue at infinity leaves us with

$$\sum_{k=1}^{N}\frac{\mu_k^{m-1}}{\prod_{n\neq k}(\mu_k-\mu_n)}=\delta_{m,N}\ .$$

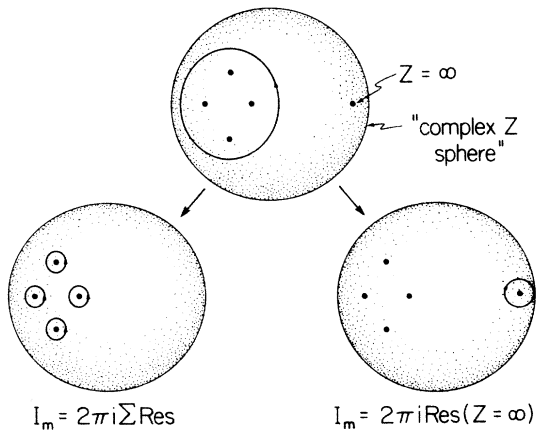

FIG. 16. Two ways to evaluate I_m.

So

$$\frac{d}{dx}W_j(x,t) = -2iC_{j,N} \equiv \frac{1}{2\pi}k_j \ .$$

A similar calculation gives

$$\frac{d}{dt}W_j(x,t) = -4i\left[C_{j,N-1} + \left[\frac{1}{2}\sum_{k=1}^{2N+2}\lambda_k\right]C_{j,N}\right]$$
$$= \frac{1}{2\pi}\Omega_j \ .$$

Therefore

$$W_j(x,t) = \frac{1}{2\pi}(k_j x + \Omega_j t + d_j) \ , \tag{A18}$$

where d_j is a phase which is determined by the initial μ_k's. We will discuss this more in a moment. The Abel map has simplified the μ motions. Now, in order to find $u(x,t)$ (the solution of the CSE) we need to invert the Abel map. This involves the solution of what is called Jacobi's inversion problem. We will not go into details here but will sketch the broad outlines of the solution. The interested reader is referred to Refs. 35, 22, 23, 11 and 15. The derivation was first done in Ref. 15.

The inversion is carried out by first introducing a new function, the Riemann theta function. This function has as its argument a vector of N complex numbers: $\bar{z}$ ($\bar{z} \in C^N$). Keep in mind that here, and in what follows, the symbol $\bar{z}$ represents an N-dimensional array of complex number *not* the complex conjugate of z. Recall that the $\underline{\tau}$ matrix is symmetric ($\tau_{jk} = \tau_{kj}$) and has positive definite imaginary part ($\mathrm{Im}\underline{\tau} > 0$). The significance of these two properties will soon become apparent.

The Riemann theta function is defined as follows:

$$\theta(z \mid \tau) \equiv \sum_{m_1=-\infty}^{\infty} \cdots \sum_{m_N=-\infty}^{\infty} \exp(2\pi i\bar{m}\cdot\bar{z} + \pi i\bar{m}\cdot\underline{\tau}\cdot\bar{m}) \ . \tag{A19}$$

Here

$$\bar{m}\cdot\bar{z} = \sum_{k=1}^{N} m_k z_k, \quad \bar{m}\cdot\underline{\tau}\cdot\bar{m} = \sum_{k,j=1}^{N} \tau_{jk} m_j m_k \ .$$

Because the $\underline{\tau}$ matrix has positive definite imaginary part it can be proven that this series representation of $\theta(z \mid \tau)$ converges absolutely for all finite $\bar{z}$.[23]

The θ function has some useful period properties. For example, suppose we displace $\bar{z}$ by $\bar{e}_k$ where

$$\bar{e}_k = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

where the 1 is in the kth place. Then from the definition it is obvious that

$$\theta(z + e_k \mid \tau) = \theta(z \mid \tau) \ . \tag{A20}$$

Therefore the θ function has N real periods. It also has N "quasiperiods." Let us add the vector $\bar{\tau}_k \equiv \underline{\tau}\bar{e}_k$ to $\bar{z}$. (Notice that $\bar{\tau}_k$ is the kth column of the $\underline{\tau}$ matrix),

$$\theta(z + \tau_k \mid \tau) \equiv \sum_m \exp(2\pi i\bar{m}\cdot\bar{z} + 2\pi i\bar{m}\cdot\bar{\tau}_k + \pi i\bar{m}\cdot\underline{\tau}\cdot\bar{m}) \ .$$

Since the summation goes over all possible entries in $\bar{m}$, we can shift to a new dummy index $\bar{m}' = \bar{m} + \bar{e}_k$,

$$\theta(z + \tau_k \mid \tau) = \sum_{m'} \exp[2\pi i(\bar{m}' - \bar{e}_k)\cdot\bar{z} + 2\pi i(\bar{m}' - \bar{e}_k)\cdot\tau_k$$
$$+ \pi i(\bar{m}' - \bar{e}_k)\cdot\underline{\tau}\cdot(\bar{m}' - \bar{e}_k)] \ .$$

Using the fact that $\underline{\tau}$ is a symmetric matrix we can write

$$(\bar{m}' - \bar{e}_k)\cdot\underline{\tau}\cdot(\bar{m}' - \bar{e}_k) = \bar{m}'\cdot\underline{\tau}\cdot\bar{m}' - 2\bar{m}'\cdot\bar{\tau}_k + \tau_{kk} \ ,$$

where τ_{kk} is the kth diagonal component of the $\underline{\tau}$ matrix. Putting this all together we get

$$\theta(z + \tau_k \mid \tau) = \exp(-2\pi i z_k - \pi i\tau_{kk})\theta(z \mid \tau) \ . \tag{A21}$$

Now consider the N functions

$$\psi_j(p) = \int_{p_0}^{p} d\psi_j, \quad j = 1,2,\ldots,N \ . \tag{A22}$$

Here $d\psi_j$ is the jth normalized holomorphic differential. The base point p_0 is a fixed point on the surface M. The integral is over a contour on the Riemann surface N connecting the points p_0 and p (the *same* contour for each j). These functions on M are multiple valued because we can add a closed loop to the path without changing the end points. Since any closed loop can be decomposed into a combination of a and b cycles the functions $\psi_j(p)$ are defined modulo the group of a and b periods.

Now consider the following function, $F(p)$, defined on the surface M:

$$F(p) \equiv \theta(\psi(p) - K \mid \tau) \ , \tag{A23}$$

where θ is the N-dimensional Riemann theta function associated with the surface M, $\psi(p)$ is the N-dimensional array of complex functions $\psi_j(p)$, and K is an N-dimensional array of complex numbers which is independent of the point p.

$F(p)$ is an entire function but it is multivalued on M. If we move the point p continuously around one of the b cycles and bring it back to the original point the ψ_j's change by adding elements of the τ matrix and this means that the function $F(p)$ will change by an overall factor given by Eq. (A21).

To make the function single valued the surface M is cut (dissected) in the canonical fashion.[11,21,22] This leads to the surface M^* which is simply connected. Since $F(p)$ (now defined on M^*) is entire the following integral is well defined:

$$I_0 \equiv \frac{1}{2\pi i}\int_{\partial M^*} d\ \ln F(p) \ .$$

The integral is around the boundary of M^*. Since this region is simply connected we can apply Cauchy's

theorem. The integral I_0 is the number of zeros of $F(p)$ on the surface M^*. Riemann proved that the number of zeros is N.

By an appropriate choice of the N complex numbers K_j $(j=1,2,\ldots,N)$ in Eq. (A23) the n zeros of $F(p)$ can be chosen to coincide with the N points $(\mu_j(x,t),\sigma_j)$. Recall that the auxiliary variables are points on the surface M^* which have coordinates specified by a complex number (μ_j) and a sheet index (σ_j). The appropriate K_j's are[11,23,15]

$$K_j \equiv \tfrac{1}{2}\tau_{jj} - \sum_{m=1}^{N} \int_{a_m} \psi_j d\psi_m + \sum_{k=1}^{N} \psi_j(\mu_k) \tag{A24}$$

[notice that the last summation in Eq. (A24) is $W_j(x,t)$ from Eq. (A16)]. Since the zeros of $F(p)$ now correspond to the auxiliary variables it is possible to evaluate the summations

$$\sum_{m=1}^{N} \mu_m(x,t) \quad \text{and} \quad \sum_{m=1}^{N} \mu_m^2(x,t) .$$

Both of these are needed in Eqs. (A13) and (A15). This is true because

$$\sum_{j>k} \mu_j \mu_k = \tfrac{1}{2}\left[\left[\sum_{m=1}^{N} \mu_m \right]^2 - \sum_{m-1}^{N} \mu_m^2 \right] .$$

Consider the two integrals

$$I_1 = \frac{1}{2\pi i} \int_{\partial M^*} \lambda(p) d\ln[F(p)] ,$$

$$I_2 \equiv \frac{1}{2\pi i} \int_{\partial M^*} \lambda^2(p) d\ln[F(p)] .$$

Here $\lambda(p)$ is the coordinate (complex number) associated with the point p. It is possible to evaluate these integrals in two independent ways.

(1) Direct evaluation using the properties of $F(p)$. This leads to[11,23,15]

$$I_1 = \sum_{m=1}^{N} \int_{a_m} \lambda d\psi_m \equiv A_1 , \tag{A25}$$

$$I_2 = \sum_{m=1}^{N} \int_{a_m} \lambda^2 d\psi_m \equiv A_2 . \tag{A26}$$

Using the standard definitions for the a cycles (see Fig. 10) and the fact that the main spectra come in conjugate pairs it is possible to show that A_1 and A_2 are real. They are independent of x and t since they depend only on the constants $\{\lambda_j\}$.

(2) I_1 and I_2 can also be evaluated via the residue theorem. This leads to

$$I_1 = \sum_{m=1}^{N} \mu_m + \operatorname*{Res}_{\lambda\to\infty^+} [\lambda(p) d \ln F(p)] + \operatorname*{Res}_{\lambda\to\infty^-} [\lambda(p) d \ln F(p)] ,$$

$$I_2 = \sum_{m=1}^{N} \mu_m^2 + \operatorname*{Res}_{\lambda\to\infty^+} [\lambda^2(p) d \ln F(p)] + \operatorname*{Res}_{\lambda\to\infty^-} [\lambda^2(p) d \ln F(p)] .$$

There are two "residues at infinity" because $\lambda=\infty$ is not a branch point on M. Therefore there are two points on M corresponding to $\lambda=\infty$, one on the top sheet and one on the bottom. Let us examine

$$\operatorname*{Res}_{\lambda\to\infty^+} [\lambda(p) d \ln F(p)] .$$

The other three residues are evaluated in a similar way. Start by writing

$$\lambda(p) d \ln F(p) = \lambda \sum_{j=1}^{N} \frac{d\psi_j}{dp} D_j \ln F(p) d\lambda = \lambda \sum_{j=1}^{N} \sum_{m=1}^{N} C_{jm} \frac{\lambda^{m-1}}{\sqrt{P(\lambda)}} D_j d \ln F(p) d\lambda ,$$

where D_j signifies a derivative with respect to the jth argument of $F(p)$. Expanding the above expression as $\lambda\to\infty^+$ we find

$$\lim_{\lambda\to\infty^+} \lambda(p) d \ln F(p) = \sum_{j=1}^{N} \sum_{m=1}^{N} C_{jm} D_j \ln F(p) \{ [\lambda^{m-N-1} + O(\lambda^{m-N-2})] d\lambda \} .$$

The only residue comes from the first term when $m=N$. In that case change variables to

$$t = \frac{1}{\lambda} ,$$

$$\frac{dt}{t} = -\frac{d\lambda}{\lambda} ,$$

to get

$$\operatorname*{Res}_{\lambda\to\infty^+} [\lambda d \ln F] = -\sum_{j=1}^{N} C_{jN} D_j \ln F(r^+ - K) .$$

Here r^+ is the value of $\psi(p)$ when p is the point on M^* where $\lambda\to\infty^+$. Going back to the definition of K_j [Eq. (A24)] and using Eq. (A16) it is possible to show that

$$\partial_x \ln F(r^+ - K) = -\sum_{j=1}^{N} K_{jx} D_j \ln F(r^+ - K) = 2i \sum_{j=1}^{N} C_{jN} D_j \ln F(r^+ - K) ,$$

and therefore

$$\operatorname*{Res}_{\lambda\to\infty^+} (\lambda d \ln F) = \frac{-i}{2} \partial_x \ln F(r^+ - K) .$$

The analysis for $\lambda\to\infty^-$ is the same except

$$[P(\lambda^-)]^{1/2} = -[P(\lambda^+)]^{1/2} .$$

The final result is

$$I_1 = \sum_{j=1}^{N} \mu_j + \frac{i}{2} \partial_x \ln \left[\frac{F(r^- - K)}{F(r^+ - K)} \right] ,$$

which can be written more succinctly as

$$I_1=\sum_{j=1}^{N}\mu_j+\frac{i}{2}\partial_x\ln\left[\frac{\theta^-}{\theta^+}\right] \quad \text{with } \theta^{\pm}\equiv F(r^{\pm}-K)\ . \tag{A27}$$

Using the same techniques we find

$$I_2=\sum_{j=1}^{N}\mu_j^2-\tfrac{1}{4}\partial_x\ln(\theta^+\theta^-)+\frac{i}{4}\partial_t\ln\left[\frac{\theta^-}{\theta^+}\right]\ . \tag{A28}$$

Using Eqs. (A25)–(A28) we can rewrite Eqs. (A13) and (A15),

$$\partial_x\ln u=\partial_x\ln J_1+2iA_1-i\sum_{m=1}^{2N+2}\lambda_m\ , \tag{A29}$$

$$\partial_t\ln u=\partial_t\ln J_1+iN(x,t)\ , \tag{A30}$$

where

$$\begin{aligned}N(x,t)=&-\frac{i}{2}\partial_t\ln J_1-2A_2-\tfrac{1}{2}\partial_x^2\ln J_2\\&+\left[-\frac{i}{2}\partial_x\ln J_1+A_1\right]^2\\&-2\left[\sum_{j>k}\lambda_j\lambda_k-\tfrac{3}{4}\left[\sum_{j=1}^{2N+2}\lambda_j\right]^2\right]\\&+i\left[\sum_{j=1}^{2N+2}\lambda_j\right]\partial_x\ln J_1-2A_1\sum_{j=1}^{2N+2}\lambda_j\end{aligned}$$

and $J_1\equiv\theta^+/\theta^-$, $J_2\equiv\theta^+\theta^-$. Because Eqs. (A29) and (A30) are compatible $N(x,t)$ can only be a function of time. In fact, Kotljarov and Its show that N is a real constant.[15]

The solution of Eqs. (A29) and (A30) can now be written as

$$u(x,t)=u(0,0)e^{ik_0x-i\omega_0t}\frac{\theta(W^-\mid\tau)}{\theta(W^+\mid\tau)}\ , \tag{A31}$$

where $W_j^{\pm}=1/2\pi(k_jx+\Omega_jt+\delta_j^{\pm})$. The $2N$ quantities $W_j^{\pm}$ $(j=1,\ldots,N)$ are not the same as the N W_j's of Eq. (A18). They are different because the $2N$ phases, $\delta_j^{\pm}$, are determined not just by the initial values of $\mu_j(0,0)$, as the d_j's of Eq. (A18) are. The general form of these constants is given in Eq. (A32) below. The $\delta_j^{\pm}$ are simply the part of $r_j^{\pm}-K_j$ which is independent of x and t.

The external wave number and frequency, k_0 and ω_0, are given by[11,15]

$$k_0=2A_1-\sum_{j=1}^{2N+2}\lambda_j\ ,$$

$$\omega_0=N\ ,$$

where N was given above.

Note: This expression for ω_0 differs from that given in Ref. 11 by a few constant factors. The expression above is the correct one.

For our purposes we never need to evaluate this complicated expression for ω_0 since the external phase is irrelevant when studying instabilities. There is one special case, however, when ω_0 is evaluated easily: the plane-wave case. Suppose there are only two branch points: $\lambda_j=\pm i$. This case has genus zero so there are no a cycles: $A_1=A_2=0$. The theta functions $\theta(W^+)$ and $\theta(W^-)$ are completely degenerate (see body of text) so $\theta(W^+)=\theta(W^-)=1$. This leads to

$$k_0=0,\quad \omega_0=-2\ ,$$

as required.

The $2N$ phases $\delta_j^{\pm}$ are given by

$$\frac{1}{2\pi}\delta_j^{\pm}=\int_{p_0}^{\infty\pm}d\psi_j-\sum_{k=1}^{N}\left[\int_{p_0}^{\mu_j(0,0)}d\psi_j-\int_{a_k}\psi_jd\psi_k\right]-\tfrac{1}{2}\tau_{jj}\ . \tag{A32}$$

The term $k=j$ in the summation is special so separate it out,

$$\frac{1}{2\pi}\delta_j^{\pm}=\int_{\mu_j(0,0)}^{\infty\pm}d\psi_j+\tfrac{1}{2}\int_{a_j}d(\psi_j^2)-\tfrac{1}{2}\tau_{jj}-\sum_{k\,(\neq j)}\left[\int_{p_0}^{\mu_j(0,0)}d\psi_j-\int_{a_k}\psi_jd\psi_k\right].$$

The second integral on the right can be evaluated since the function ψ_j changes by unity around the a_j cycle,

$$\begin{aligned}\tfrac{1}{2}\int_{a_j}d(\psi_j^2)&=\tfrac{1}{2}[\psi_j(p_0)+1]^2-\tfrac{1}{2}\psi_j^2(p_0)\\&=\psi_j(p_0)+\tfrac{1}{2}\\&=\tfrac{1}{2}\ .\end{aligned}$$

where we have used $\psi_j(p_0)=0$ since p_0 is the base point of the integral defining $\psi_j(p)$.

Thus

$$\frac{1}{2\pi}\delta_j^{\pm}=\int_{\mu_j(0,0)}^{\infty\pm}d\psi_j-\tfrac{1}{2}\tau_{jj}+\tfrac{1}{2}-\sum_{k\,(\neq j)}\left[\int_{p_0}^{\mu_k(0,0)}d\psi_j-\int_{a_k}\psi_jd\psi_k\right].$$

In general this is as far as we can go without numerically evaluating the remaining integrals. However, we can go a bit further when dealing with modulational problems. In this case the last integral can be evaluated by residues when k is in the modulation subspace [see Eq. (13)],

$$\begin{aligned}\int_{a_k}\psi_jd\psi_k&=\psi_j(\lambda_k^0)+O(\epsilon_k)\\&=\int_{p_0}^{\lambda_k^0}d\psi_j+O(\epsilon_k),\quad N+1\le k\le N+M\ .\end{aligned}$$

Therefore

$$\int_{p_0}^{\mu_k(0,0)}d\psi_j-\int_{a_k}\psi_jd\psi_k=\int_{\lambda_k^0}^{\mu_k(0,0)}d\psi_j+O(\epsilon_k),\quad N+1\le k\le N+M$$

for k in the modulation subspace. For the initial modulation to be small $\mu_k(0,0)$ must start near λ_k^0. Since $d\psi_j$ has no singularity there (for $j\neq k$) the integral will be $O(\epsilon_k)$. Therefore there is no contribution to the last

summation in $\delta_j^{\pm}$ from the modulation subspace [so long as $\mu_k(0,0) \sim \lambda_k^0$]. For the plane-wave case we have simply

$$\frac{1}{2\pi}\delta_j^{\pm} = \int_{\mu_j(0,0)}^{\infty \pm} d\psi_j - \tfrac{1}{2}\tau_{jj} + \tfrac{1}{2} + O(\epsilon) .$$

By performing the above integral on the lower sheet the singular behavior at $\mu_j(0,0) \sim \lambda_j^0$ will cancel with that of $\frac{1}{2}\tau_{jj}$ (consider Fig. 3). This will leave an $O(1)$ contribution. These constants [the $2(N)\delta^{\pm}$] are completely determined by the main spectrum and the initial conditions of the $\mu_j(0,0)$'s. We have seen that these initial conditions are not arbitrary but must satisfy a constraint. This translates into the $\delta_j^{\pm}$ also not being arbitrary. Since we have satisfied the constraint at the level of the μ_j's and the Abel map is locally one to one, it is also solved at the level of the $\delta^{\pm}$. Alternatively, one could solve the reality constraint at the level of the $\delta^{\pm}$ without reference to the μ_j's. This is the approach taken in Ref. 30. The preferred method depends upon the application one has in mind.

[1]P. L. Kelley, Phys. Rev. Lett. **15**, 1005 (1965).

[2]V. I. Talanov, Pis'ma Zh. Eksp. Teor. Fiz. **2**, 223 (1965) [JETP Lett. **2**, 141 (1965)].

[3]D. J. Benney and A. C. Newell, J. Math. Phys. (Cambridge, Mass.) **46**, 133 (1967).

[4]V. E. Zakharov, Zh. Eksp. Teor. Fiz. **62**, 1745 (1972) [Sov. Phys.—JETP **35**, 908 (1972)].

[5]H. C. Yuen and B. M. Lake, Adv. Appl. Mech. **22**, 67 (1982).

[6]M. V. Goldman, Rev. Mod. Phys. **55**, 709 (1984).

[7]H. T. Moon and M. V. Goldman, Phys. Rev. Lett. **53**, 1821 (1984).

[8]A. Hasegawa and F. Tappert, Appl. Phys. Lett. **23**, 142 (1973).

[9]L. F. Mollenauer and R. H. Stolen, Laser Focus **18**, 193 (1982).

[10]G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).

[11]E. R. Tracy, Ph.D. thesis, University of Maryland, 1984 (unpublished).

[12]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys.—JETP **34**, 62 (1972)].

[13]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).

[14]A. E. Walstead, Ph.D. thesis, University of California, 1980 (unpublished).

[15]V. P. Kotljarov and A. R. Its, Dopov. Akad. Nauk Ukr RSR., Ser. A **11**, 965 (1976) (in Ukrainian).

[16]Y. C. Ma and M. J. Ablowitz, Stud. Appl. Math. **65**, 113 (1981).

[17]N. Ercolani, M. G. Forest, and D. W. McLaughlin, Physica **18D**, 472 (1986).

[18]E. R. Tracy, H. H. Chen and Y. C. Lee, Physica **18D**, 481 (1986).

[19]E. R. Tracy, H. H. Chen, and Y. C. Lee, Phys. Rev. Lett. **53**, 218 (1984).

[20]H. P. McKean and P. van Moerbeke, Inventiones Math. **30**, 217 (1975).

[21]H. Hochstadt, Math. Z. **82**, 237 (1963).

[22]G. Springer, *Introduction to Riemann Surfaces* (Chelsea, New York, 1981).

[23]C. L. Siegel, *Topics in Complex Function Theory* (Wiley-Interscience, New York, 1971), Vol. 2.

[24]M. G. Forest and D. W. McLaughlin, J. Math. Phys. **23**, 1248 (1982).

[25]M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).

[26]R. K. Dodd, J. C. Eilbeck, J. D. Gibbon, and H. C. Morris, *Solitons and Nonlinear Wave Equations* (Academic, New York, 1982).

[27]W. Magnus and S. Winkler, *Hill's Equation* (Dover, New York, 1979).

[28]M. G. Forest and N. Ercolani, Commun. Math. Phys. **99**, 1 (1985).

[29]E. Previato, Duke Math. J. **52**, 329 (1985).

[30]A. C. Ting, E. R. Tracy, H. H. Chen, and Y. C. Lee, Phys. Rev. A **30**, 6, 3355 (1984).

[31]E. R. Tracy, C. H. Chin, and H. H. Chen, Physica **23D**, 91 (1986).

[32]A. C. Ting, H. H. Chen, and Y. C. Lee, Phys. Rev. Lett. **53**, 1348 (1984).

[33]N. N. Bogolyubov, Jr. and A. K. Prikarpatskii, Sov. Phys. Dokl. **27**, 113 (1982).

[34]A. K. Prikarpatskii, Theor. Mat. Phys. **47**, 487 (1981).

[35]B. A. Dubrovin, V. B. Mateev, and S. P. Novikov, Russian Math Surveys **31**, 59 (1976).