Macroscopic lattice dynamics

M.H. Hays [a,1], C.D. Levermore [b,2], P.D. Miller [a,3]

[a] *Program in Applied Mathematics, University of Arizona, Tucson, Arizona 85721, USA*

[b] *Department of Mathematics and Program in Applied Mathematics, University of Arizona, Tucson, Arizona 85721, USA*

Abstract

Fully nonlinear modulation equations are presented that govern the slow evolution of single-phase harmonic wavetrains under a large family of spatially discrete nonintegrable flows, among which is the discrete nonlinear Schrödinger equation (DNLS). These modulation equations are a pair of partial differential equations in conservation form that are hyperbolic for some data and elliptic for other data. In some cases these equations are capable of dynamically changing type from hyperbolic to elliptic, a phenomenon that has been associated with the modulational instability of the underlying wavetrain. By putting the modulation equations in Riemann invariant form, one can select initial data that avoid this dynamic change of type. Numerical experiments demonstrating the theoretical results are presented.

1. Introduction

In this paper, we begin an investigation of the behavior and interactions of modulated wavetrains in discrete lattices described by an infinite number of coupled ordinary differential equations. Specifically, we study the evolution of a slowly modulated single-phase wavetrain under a large family of (generally nonintegrable) flows, by formally deriving the corresponding modulation equations.

In a more general setting, the goal is to identify regions of the lattice phase space in which the dynamics can be described by a macroscopic model that is simpler (in the sense of being more intuitive or amenable to analysis) than the full microscopic lattice system. The macroscopic model is obtained by representing the behavior locally in space and time by one of the known exact solutions.

Our primary interest in constructing modulation equations for discrete systems is to extend families of known exact solutions using analytical methods. Although the discrete systems considered here have many physical applications, typically all that is known analytically about the system is the existence of such a solution family. As a concrete representative example of such a discrete lattice, we use the discrete nonlinear Schrödinger equation (DNLS),

$$i\dot{A}_j + \varepsilon(A_{j-1} - 2A_j + A_{j+1}) + \gamma|A_j|^2 A_j = 0, \quad (1)$$

where j is an integer index, and $A_j(t)$ is a complex function of time. This lattice model has been used as a numerical scheme for integrating the continuous nonlinear Schrödinger equation [8], as a model for phonon-exciton interactions in alpha-helix proteins [5,6], and as an envelope equation for harmonic os-

---

[1] hays@math.arizona.edu
[2] lvrmr@math.arizona.edu
[3] millerpd@math.arizona.edu

cillations in the linear frequency band of weakly non-linear, strongly dispersive lattice equations [10]. It is well known that the DNLS equation has a two-parameter family of harmonic plane wave solutions of the form

$$A_j(t) = Ae^{i(kj-\omega t)} , \tag{2}$$

where the complex amplitude A, unit cell twist angle (or wavenumber) k, and frequency ω are linked by the dispersion relation

$$\omega = 2\varepsilon(1 - \cos k) - \gamma|A|^2 . \tag{3}$$

This solution is spatially periodic only in the case when k is a rational multiple of 2π. We are interested in an explicit analytical characterization of any solutions of the underlying DNLS lattice that may be considered to be wavetrains in which the parameters are taken to be slowly varying in space and time. For example, it is known that the continuous nonlinear Schrödinger equation (NLS) has an analogous family of plane wave solutions, and in the defocusing case ($\varepsilon\gamma < 0$) there exist stable, slowly modulated plane waves. One of our interests here is the degree to which the same story holds for the harmonic plane wave solutions to the DNLS lattice.

However, our interests run deeper as well. We believe that in several ways the modulation theory for lattices plays a no less fundamental role in applied mathematics than the corresponding theory for PDE's. On the one hand, this is because the lattice modulation theory presented in this paper is capable of reproducing certain formal results from standard PDE modulation theory. For example, it will be shown later that, in the case where the lattice has a partial differential evolution equation as its continuum limit, the lattice modulation equations contain the corresponding continuum modulation equations in an appropriate limit; in this sense, the modulation theory for the lattice is richer than that of the corresponding PDE. On the other hand, numerical analysis favors the study of the appearance of macroscopic behavior in lattices to that in PDEs. This is because the full microscopic dynamics can be integrated numerically without spatial discretization errors. Such ease of computation should

be contrasted with the case of macroscopic behavior in PDEs, where accurate numerical integration of the microscopic system involves keeping track of high wavenumber modes present in the microstructure as well as the low wavenumber modes associated with the modulations; the problem is inherently spatially stiff.

Let us illustrate the phenomenology. Denote by ρ the quantity $|A|^2$. Take ρ and k to be the two independent real parameters of the DNLS harmonic plane wave family, and consider preparing a lattice of 400 sites arranged in a periodic ring in a state described by a wave train with spatially constant k and an envelope given by an expression of the form

$$|A_j|^2 = a + b\,\text{sech}(cj) , \tag{4}$$

with the constant c chosen so that the envelope has stabilized sufficiently in the wings to approximately satisfy the periodic boundary conditions. We integrated the DNLS equation with this data using a variable order Adams-Bashforth ODE solver, and, at several subsequent times, the macroscopic quantities ρ and k were extracted from the data. Snapshots of these fields are reproduced in Fig. 1, with the initial condition fields emphasized using thicker curves. It is clear that the ansatz of a modulated harmonic wavetrain is in fact valid for a time long compared to the local period of the wavetrain determined by (3). In fact, the ratio of the length of time between snapshots in the figure to the microscopic period is about 5. However, the regularity is destroyed locally by the eventual formation of a cusp-like singularity, as well as a steepening suggestive of shock formation.

This numerical evidence suggests that the macroscopic fields ρ and k should have a dynamical description that is independent of the details of the microscopic evolution. In Section 2 it will be shown that this is indeed the case, and we will develop equations for the macroscopic fields. Our procedure is based on the averaging of discrete versions of local conservation laws, followed by a passage to the continuum limit in the macroscopic field variables. This method is very much in the spirit of the work of Whitham in the mid 1960's [16], and an extension of the method to nonintegrable discrete problems is utilized in the re-
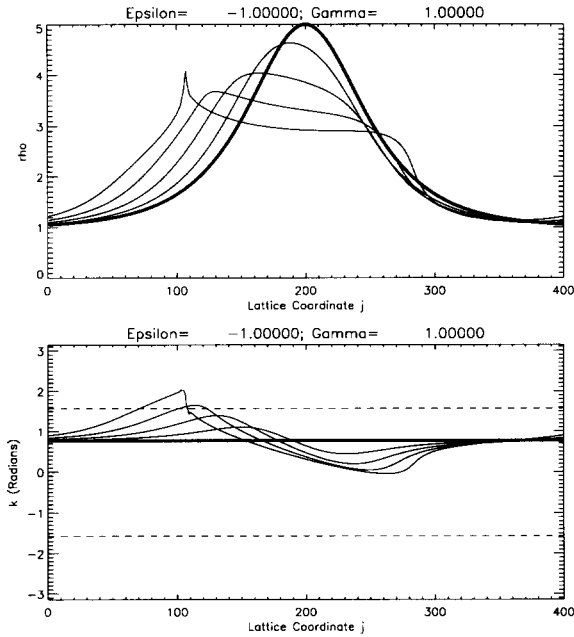
Fig. 1. Above: The macroscopic quantity ρ. Below: The macroscopic quantity k.

cent work of Levermore and Liu [11,12]. In deriving modulation equations, we proceed in some generality by embedding the DNLS equation in a large family of lattice models that also includes the integrable discretization of NLS due to Ablowitz and Ladik [2].

Section 3 contains an analysis of the modulation equations, with particular emphasis on the special case of the DNLS equation. The characteristic velocities and the corresponding domains of hyperbolicity will be written down. It will then be clear that the formation of the cusp in Fig. 1 is preceded locally by a change of type of the equations from hyperbolic to elliptic, which echos Whitham's condition for modulational instability and local breakdown of the plane wave ansatz. Section 3 also includes a discussion of the corresponding linear stability analysis, which, while less useful prior to change of type, provides the explicit local stable and unstable manifolds after change of type.

Section 4 introduces the Riemann-invariant form of the modulation equations, which permits the identification of a class of macroscopic data that never experiences dynamic change of type. For data in this class,

the only type of singularity that can form in the modulational system is an infinite derivative in finite time. If the modulation equations were not representing a certain microscopic system in a macroscopic limit, one might interpret such a steepening as the onset of a hyperbolic shock; in the present case however, the loss of regularity corresponds to the local reemergence of the microscopic dynamics. In some cases, this class of permanently hyperbolic macroscopic data can be described explicitly in terms of the "physical" macroscopic variables ρ and k, although it is much more naturally described in terms of the Riemann invariants. The utility of our classification scheme will be demonstrated with a numerical experiment in which initial data is chosen to avoid change of type, and it is shown that our prediction is borne out and that the evolution is regular until a shock forms.

Finally, Section 5 explores some questions left unanswered by our description of modulated harmonic plane waves. Also the possible extension of our methodology to discrete analogs of multiphase wavetrains will be discussed, along with comments on the universality of and interactions among such wavetrains.

2. Modulation equations for plane waves

Below we will develop a fully nonlinear modulation theory for harmonic plane wave solutions of a general lattice model that contains the DNLS (1) and several other important models as special cases and takes the form

$$i\dot{A}_j + f(|A_j|^2)(A_{j+1} + A_{j-1}) + F(|A_j|^2)A_j = 0.$$
$$(5)$$

Our numerical simulations employ periodic boundary conditions on the integer index j; however, the analysis will be purely local, so that the results will hold for arbitrary boundary conditions. The functions f and F will be assumed to be polynomial functions of their arguments with real coefficients. In addition, it is useful to take $f(\rho) > 0$ for all $\rho \geq 0$ (or alternatively to consider modulations of wavetrains with amplitude ρ such that $f(\rho) > 0$).

There are several choices one can make for the functions f and F such that the system (5) has the NLS equation as its continuum limit:

- $f(\rho) = \epsilon$ and $F(\rho) = -2\epsilon + \gamma\rho$. This corresponds to the DNLS equation.
- $f(\rho) = \epsilon + \gamma\rho/3$ and $F(\rho) = -2\epsilon + \gamma\rho/3$. This is a new discretization of the NLS which was proposed recently by Salerno [14].
- $f(\rho) = \epsilon + \gamma\rho/2$ and $F(\rho) = -2\epsilon$. This is the Ablowitz–Ladik equation [2]. It is the only nontrivial case of (5) (here, nontrivial means F and f nonconstant) that is known to be completely integrable.

In order to obtain the continuum limit, note that it is necessary to scale ϵ as $1/h^2$, where h is the vanishing lattice spacing.

The system (5) is Hamiltonian and possesses a *nonstandard sympletic structure*; the reader is referred to Appendix A for details on this formulation. Generally, there are two known constants of motion:

$$H = \sum_j \left\{ 2|A_j|^2 - |A_{j+1} - A_{j-1}|^2 \right.$$
$$\left. + \int_0^{|A_j|^2} \frac{F(y)\,dy}{f(y)} \right\},$$

$$N = \sum_j \int_0^{|A_j|^2} \frac{dy}{f(y)}. \qquad (6)$$

Here, H is the Hamiltonian for (5) and N is called the norm or number.

We seek a modulational description of harmonic plane waves of (5). It is easy to see that a solution of the form

$$A_j = \sqrt{\rho}\,e^{i(kj - \omega t)} \qquad (7)$$

satisfies (5) if ω obeys the dispersion relation

$$\omega + 2f(\rho)\cos k + F(\rho) = 0. \qquad (8)$$

Taking ρ and k as the independent parameters of this solution family results in a 2-manifold of microscopic states for the lattice system (5).

In order to obtain modulation equations it is necessary in our scheme to have a set of conservation laws

in *local form*. Given the form of the global conserved quantities H and N, it is reasonable to seek local conservation laws of the form

$$\frac{d}{dt}N_j + (Q_j - Q_{j-1}) = 0,$$
$$\frac{d}{dt}E_j + (R_j - R_{j-1}) = 0, \qquad (9)$$

where the *densities* are given by the summands in (6)

$$N_j = \int_0^{|A_j|^2} \frac{dy}{f(y)},$$

$$E_j = 2|A_j|^2 - |A_{j+1} - A_j|^2 + \int_0^{|A_j|^2} \frac{F(y)}{f(y)}\,dy. \qquad (10)$$

It turns out that such a local representation does indeed exist and that the *fluxes* are given by

$$Q_j = 2\operatorname{Im}(A_j^* A_{j+1}),$$
$$R_j = 2\operatorname{Im}(A_j^* A_{j+2} + A_{j+2}^* A_{j+1} + A_j^* A_{j+1})$$
$$\times f(|A_{j+1}|^2) + 2\operatorname{Im}(A_j^* A_{j+1})F(|A_{j+1}|^2). \qquad (11)$$

We will obtain modulation equations for the plane waves by inserting the plane wave solution ansatz into the density and flux terms and passing to the continuum limit in the parameters of the ansatz.

To introduce the macroscopic scale, allow ρ and k to depend on the continuous variable $x = jh$ and let $h \downarrow 0$. Our ansatz becomes (suppressing the time dependence)

$$A_j = \sqrt{\rho(x)}\,e^{i\theta(x)/h}, \qquad (12)$$

where k is defined as θ_x. Evaluating the densities N_i, E_i, and the fluxes Q_i and R_i on this ansatz and passing to the continuum limit $h \downarrow 0$ yields the limiting expressions

$$N = \int_0^\rho \frac{dy}{f(y)},$$

$$Q = 2\rho\sin k,$$

$$E = 2\rho\cos k + \int_0^\rho \frac{F(y)}{f(y)}\,dy,$$

$$R = 2\rho(f(\rho) \cos k + F(\rho)) \sin k. \qquad (13)$$

Finally, in order to make the flux difference converge to a derivative in x, we must divide (9) by the vanishing lattice spacing, h. This means that, formally, we must pass to the limit $h \downarrow 0$ on a stretched time scale of $T = ht$.

This procedure yields a system of PDEs in conservation form:

$$\frac{\partial N}{\partial T} + \frac{\partial Q}{\partial x} = 0,$$

$$\frac{\partial E}{\partial T} + \frac{\partial R}{\partial x} = 0. \qquad (14)$$

From these, one easily obtains equations for the parameters ρ and k:

$$\frac{\partial \rho}{\partial T} - f(\rho) \frac{\partial}{\partial x}[2\rho \sin k] = 0,$$

$$\frac{\partial k}{\partial T} + \frac{\partial}{\partial x}[2f(\rho) \cos k + F(\rho)] = 0. \qquad (15)$$

It is evident from these equations that the "total twist" $\int k(x)dx$ is a conserved quantity. Note that, while this discrete analog of "conservation of waves" follows naturally from the conservation law limiting procedure, it is built-in *a priori* by the alternative approach outlined in Appendix A based on averaging of a variational principle.

In the special case of the DNLS equation, the above modulation equations (15) take the form

$$\frac{\partial \rho}{\partial T} + \frac{\partial}{\partial x}(2\varepsilon\rho \sin k) = 0,$$

$$\frac{\partial k}{\partial T} + \frac{\partial}{\partial x}(-\gamma\rho - 2\epsilon \cos k) = 0. \qquad (16)$$

So, for the DNLS equation, the "total norm" $\int \rho(x)dx$ is also a conserved quantity.

3. Structure of the modulation equations

The behavior of a system of equations like (15) is determined by the existence or nonexistence of a sufficiently large family of real characteristic curves along which information propagates. In this section, the nature of the characteristic directions for (15) will

be interpreted in the context of the stability of the underlying plane wave ansatz, and connections will be drawn to the numerical experiment discussed in Section 1.

The characteristic velocities for the modulation equations (15) are given by

$$\lambda_\pm = -2f(\rho) \sin k$$
$$\pm \sqrt{-2\rho f(\rho) \left[2f'(\rho) \cos k + F'(\rho)\right] \cos k}. \qquad (17)$$

In the specific case of the DNLS this expression reduces to

$$\lambda_\pm = 2\epsilon \sin k \pm \sqrt{-2\epsilon\gamma\rho \cos k}. \qquad (18)$$

Strict hyperbolicity of the equations, which leads to stable modulations, requires real and distinct velocities. For the DNLS modulation equations there are two cases:

- In the focusing case, $\epsilon\gamma > 0$, the equations are hyperbolic when $\cos k < 0$, which corresponds to *short waves*.
- In the defocusing case, $\epsilon\gamma < 0$, the equations are hyperbolic when $\cos k > 0$, which corresponds to *long waves*.

At this point, we once again draw the reader's attention to Fig. 1. The initial data lies entirely in the hyperbolic region. Note that the location of the cusp singularity corresponds to a region where $\cos k$ has changed sign (the boundaries of the region of hyperbolicity are shown as dashed lines in the lower plot), and hence the modulation equations have locally become elliptic. Our numerical experiments suggest that the length of time (as measured on the T scale) between the change of sign of $\cos k$ and the onset of less regular behavior vanishes as the lattice is refined. This dynamic change of type has not been observed in the modulation equations corresponding to integrable models, where the only singularities that are observed are hyperbolic shocks.

Let us compare these stability results to those obtained from straightforward linear analysis (carried out in detail in Appendix B). The result of this standard analysis is that a uniform wavetrain is unstable to a perturbation of relative wavenumber Δk if

- $\gamma\rho/\epsilon > (1 - \cos \Delta k) \cos k$ when $\cos k > 0$ (long waves),
- $\gamma\rho/\epsilon < (1 - \cos \Delta k) \cos k$ when $\cos k < 0$ (short waves).

Thus when there is a Δk satisfying these conditions, the wavenumber k experiences instability. Existence of such a Δk corresponds exactly to the conditions derived above for ellipticity of the corresponding modulation equations. Hence, the modulation equation approach to modulational instability and the linear stability analysis approach agree on this crucial point. However, we also want to point out the complementarity of these two approaches. The modulation equation approach gives useful dynamical information (a hyperbolic system) prior to instability, but results in ill-posed elliptic equations at the onset of instability. The linear stability analysis, on the other hand, while saying virtually nothing about well-behaved global modulations, provides an explicit description of the local unstable manifolds following the onset of the instability. In the particular case of the DNLS, for example, the instability begins with an infinite-dimensional manifold composed of perturbative wavenumbers in a narrow sideband of the local wavenumber k.

The stability criterion for the modulation equations associated with the Ablowitz–Ladik equation is very different from that given above in the DNLS case. In the Ablowitz–Ladik case, the velocities are *never* real in the focusing version $\epsilon\gamma > 0$. Thus all plane waves are unstable in this case. The defocusing version of the system is somewhat more delicate; there is a threshold value of ρ below which the velocities are always real and above which the velocities are always complex. Moreover the threshold value is a barrier which cannot be dynamically crossed. The consequence is that there can be no dynamical change of type in the modulation equations for the Ablowitz–Ladik system.

Furthermore, the behavior of modulated harmonic waves in the DNLS equation should be strongly contrasted with the corresponding behavior in the continuous NLS equation[4]. In the focusing NLS equation, all such modulated waves are *unstable* (the modula-

tion equations are elliptic), while for the defocusing case, all such modulated waves are *stable*. There is no dynamic change of type, and moreover, the stability does not even depend on the data. In the DNLS equation, however, hyperbolic and elliptic data coexist in *both* focusing and defocusing cases. Also, as in the example in Section 1, it is possible for an initial condition that is globally hyperbolic to evolve under the modulational system into a state that is locally elliptic.

In spite of the differences, the modulation equations for the continuous NLS equation can be recovered from (16) through the following formal limit process. Replacing ϵ by ϵ/h^2 and k by hk in (16), and passing to the limit $h \downarrow 0$ yields the familiar plane wave modulation system for the NLS equation

$$\frac{\partial \rho}{\partial T} + \frac{\partial}{\partial \tilde{x}} (2\epsilon \rho k) = 0 \,,$$

$$\frac{\partial k}{\partial T} + \frac{\partial}{\partial \tilde{x}} (-\gamma \rho + \epsilon k^2) = 0 \,, \tag{19}$$

where the modulation equations then contain the spatial scale $\tilde{x} = hx = h^2 j$, and the time scale $T = ht$. In this way the modulation theory for the DNLS is more general than that of the continuous NLS.

4. Singularities and Riemann invariants

In general, only two types of singularities may evolve from initially smooth hyperbolic data. One is a hyperbolic singularity associated with a folding-over of the manifold of real characteristics. The other is a dynamic change of type in which the characteristic speeds become complex. In this section we present tools which allow us to identify hyperbolic data that will *not* undergo such a change of type.

In order to proceed, the system of modulation equations (15) is written in Riemann invariant form

$$\frac{\partial r_\pm}{\partial T} + \lambda_\pm (r_+, r_-) \frac{\partial r_\pm}{\partial x} = 0 \,, \tag{20}$$

where $\lambda_\pm$ are the characteristic velocities of the original modulation equations expressed in terms of the Riemann invariants, $r_\pm$. Expressing hyperbolic initial data parametrically as functions of x in terms of $r_\pm$, one obtains a curve in the (r_+, r_-) plane, which

---

[4] Considered in the whole-line case to allow for a continuum of perturbative wavenumbers.

we call the *Riemann signature* of the data. Note that the map taking the data to its signature has no inverse, since all the dependence on x has been removed; that is, $r_\pm(x)$ and $r_\pm(h(x))$ have the same signature whenever $h(x)$ is a homeomorphism of the line. The extreme values of $r_\pm$ in a particular signature define a rectangle in the (r_+, r_-) plane which we call the *Riemann signature box*. Under the hyperbolic evolution of (20) the signature will generally change, but the signature box remains invariant.

In general the Riemann invariants for hyperbolic data, $r_\pm(\rho, k)$, only take values in a subset of the (r_+, r_-) plane, called the *admissible region*. Although the initial signature of such data must lie in this region, part of the box may lie outside of the admissible region. Thus, if the box lies entirely within the admissible region, one may conclude that the modulation equations will not change type, as the data will be confined to the admissible region, where the modulation equations are hyperbolic. On the other hand, if some portion of the box lies outside the admissible region, the hyperbolic evolution of (20) may lead the signature to the edge of the admissible region, at which time a change of type occurs.

Note that any system of two conservation laws can, in principle, be put into the form (20), while larger systems cannot generally be written this way. In practice, however, it is necessary even for the two-equation case to integrate simultaneous differentials for which there is no obvious integrating factor. We therefore focus on cases where we can identify an integrating factor.

Finding the Riemann invariants for our system involves diagonalizing the Jacobian of the spatial derivative terms in the modulation equations. The Jacobian is

$$J = \begin{bmatrix} -2f(\rho)\sin k & -2\rho f(\rho)\cos k \\ 2f'(\rho)\cos k + F'(\rho) & -2f(\rho)\sin k \end{bmatrix}. \quad (21)$$

In each of the specific examples we have mentioned, the functions $f'(\rho)$ and $F'(\rho)$ have been proportional to the same positive function, say $\varphi(\rho)$. In such cases, there is an obvious integrating factor; hence, the Riemann invariants can be expressed as

$$r_\pm = -\frac{1}{2} \int_{k_0}^{k} \sqrt{\frac{\cos u}{-2(2\alpha \cos u + \beta)}} \, du$$

$$\pm \frac{1}{2} \int_0^\rho \sqrt{\frac{\varphi(\eta)}{\eta f(\eta)}} \, d\eta, \quad (22)$$

where $f'(\rho) = \alpha\varphi(\rho)$ and $F'(\rho) = \beta\varphi(\rho)$. The base point k_0 in the first integral is chosen to be in the region of hyperbolicity (which corresponds to reality of the integrand). These Riemann invariants are of the general form

$$r_\pm = A(k) \pm B(\rho) ; \quad (23)$$

we will say more about the consequences of this below. The admissible region for these Riemann invariants is a tilted rectangle defined by

$$A_{min} < \frac{r_+ + r_-}{2} < A_{max} ,$$

$$B_{min} < \frac{r_+ - r_-}{2} < B_{max} , \quad (24)$$

where A_{max}, A_{min}, B_{max}, and B_{min} are the extreme values taken on by the functions A and B for hyperbolic data, some of which may not be finite.

In the case of the DNLS equation, the result is particularly simple. Let $A(k)$ be the primitive of $(1/\gamma)\sqrt{-2\epsilon\gamma\cos k}$ that has mean value zero over the bounded interval of hyperbolicity in k. Note that the domain of $A(k)$ depends upon the sign of $\epsilon\gamma$, whereas the range does not. Then the Riemann invariants are

$$r_\pm(\rho, k) = A(k) \mp 2\sqrt{\rho}. \quad (25)$$

The admissible region is a half strip given by

$$\left| \frac{r_+ + r_-}{2} \right| < 1.69 \sqrt{\left| \frac{\epsilon}{\gamma} \right|} ,$$

$$r_+ < r_- . \quad (26)$$

The factor 1.69 is the maximum of $|A(k)|\sqrt{|\gamma/\epsilon|}$.

Fig. 2 is a graph of the signatures of the data snapshots shown in Fig. 1, along with the admissible region and signature box. Once again the initial signature is emphasized using larger points. Note that whereas the initial data lies within the admissible region, the upper right-hand corner of the box does not. The hyper-
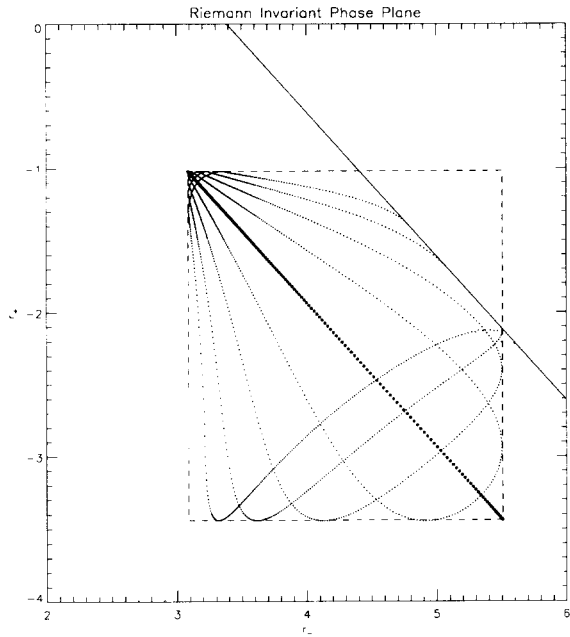
Fig. 2. The evolution of the Riemann invariant signature for Fig. 1.

Fig. 3. Above: The macroscopic quantity ρ. Below: The macroscopic quantity k.

bolic evolution leads the signature to the edge of the admissible region, at which point a change of type occurs corresponding to the formation of the cusp-like singularity seen in Fig. 1.

Next we select initial data whose signature box lies entirely within the admissible region. The evolution of this data is displayed in Fig. 3. Note that the entire evolution of the k field lies between the dashed lines, and hence the data is hyperbolic forever.

The corresponding Riemann invariant signature is shown in Fig. 4. Note that in this case a change of type is prevented since the signature box is completely contained within the admissible region, and the only singularity to appear is a hyperbolic shock. Hence, the Riemann invariant signature is a useful tool for identifying data which will *not* experience a change of type.

The most convenient representation of a modulated plane wave is through the quantities ρ and k, which have the interpretations of amplitude and wavenumber (or unit cell twist) respectively. Ideally then, one would like to be able to prepare a lattice in a hyperbolic macroscopic state described in terms of these "convenient" variables such that the evolution will not
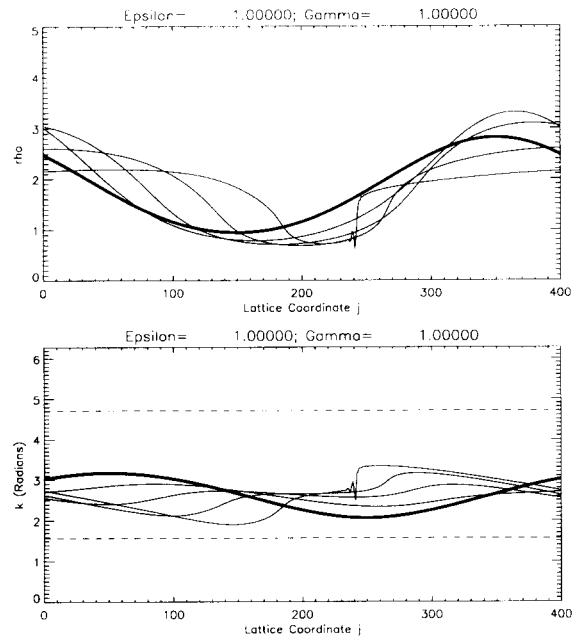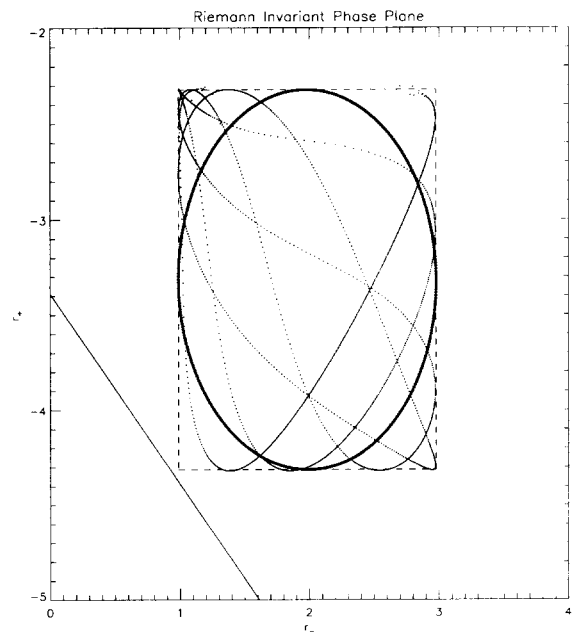

Fig. 4. The evolution of the Riemann invariant signature for Fig. 3.

change type. This requires a direct translation of the condition that the signature box be contained in the admissible region in terms of the given data $\rho(x)$ and $k(x)$. There are two interesting types of data for which such a translation is possible given Riemann invariants of the separated form $r_\pm = A(k) \pm B(\rho)$. We use the following names for these two types of data:

- Amplitude modulated (AM) data. In this case, the twist k is constant in x.
- Frequency modulated (FM) data. In this case, the amplitude ρ is constant in x.

These names are meant to be suggestive of the waveforms; they are accurate if the data in x is visualized as a time series. In each of these two cases, the Riemann invariant signature of the data is a straight diagonal line segment, the endpoints of which define the signature box. In order that this signature box lies within the tilted rectangle that is the admissible region, we must only check that the two remaining corners of the square-shaped box are inside the admissible region. The length of the signature segment is the variation of the data measured using the A function in the FM case and using the B function in the AM case. The conditions for the permanence of hyperbolicity are:

- in the AM case,

$$A(k) + \tfrac{1}{2}[\sup_x B(\rho(x)) - \inf_x B(\rho(x))] < A_{max},$$

$$A(k) - \tfrac{1}{2}[\sup_x B(\rho(x)) - \inf_x B(\rho(x))] > A_{min};$$

(27)

- and in the FM case,

$$B(\rho) + \tfrac{1}{2}[\sup_x A(k(x)) - \inf_x A(k(x))] < B_{max},$$

$$B(\rho) - \tfrac{1}{2}[\sup_x A(k(x)) - \inf_x A(k(x))] > B_{min}.$$

(28)

As an illustration of this result, consider once again the DNLS case with FM data depicted in Fig. 5. We see that the admissible region is a semi-infinite strip, so that B_{max} is not finite. It appears that it is very easy to find FM data that is permanently hyperbolic, since there is only one effective constraint which takes the form

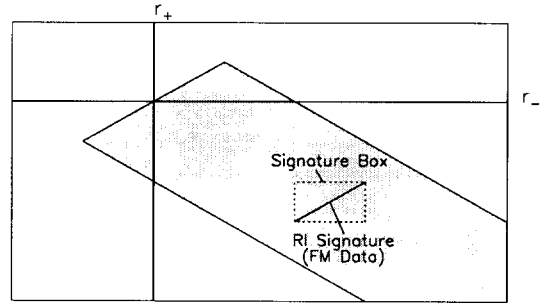$$B(\rho) > \tfrac{1}{2}[\sup_x A(k(x)) - \inf_x A(k(x))],$$

(29)

Fig. 5. The Riemann signature for the DNLS with FM data. The admissable region is shaded.

since $B_{min} = 0$. Furthermore, it is possible to use the bounds on the function A to obtain a sufficient condition for permanence of hyperbolicity for FM data which takes the form

$$B(\rho) > \tfrac{1}{2}(A_{max} - A_{min}),$$

(30)

and is independent of the twist data $k(x)$. Thus, sufficiently large amplitude FM data will always be permanently hyperbolic in the DNLS equation.

We want to point out that the results for AM and FM hyperbolic data can be extended to cover the case of data which is only nearly of AM or FM type. This follows from the observation that the signature box is defined by the endpoints of a line segment, so that small changes in the middle of the signature segment will not change the box or the conditions for permanence of hyperbolicity, which depend only on the endpoints. So, for example, in the quasi AM case, if the previously spatially constant k field is perturbed slightly in the neighborhood of x values where $B(\rho(x))$ does not take on extreme values corresponding to the endpoints of the signature segment, permanence of hyperbolicity will be guaranteed by the same conditions on the endpoints of the segment as given above. A similar argument holds for quasi FM data. In fact, the reader will observe that the quasi AM and quasi FM classes are invariant under the modulational system (15) while the pure AM and FM classes are not.

5. Conclusions and future work

In this paper we have presented a nonlinear modulation theory for plane wave solutions of a large class of discrete dynamical systems. This theory allows one to predict the hyperbolic evolution of such modulated solutions and to identify the types of singularities which may form. By passing to the Riemann invariant form of the modulation equations, one can identify conditions under which the initial data are immune to dynamic change of type. In certain instances it is easy to translate these conditions into constraints on the original field variables.

The development of modulation theory for single phase wavetrains in linear dispersive PDEs was motivated in part by the result of stationary phase analysis that all initial disturbances can be described by a modulated single phase wavetrain in an appropriate long time limit. Similar harmonic waves are observed in the asymptotic aftermath of an initial shock in the Toda lattice, as observed and studied by Holian, Flaschka, and McLaughlin [9], Venakides, Deift, and Oba [15], and most recently by Bloch and Kodama [4]. However, the spontaneous appearance of modulated wavetrains is not restricted to integrable systems; for example, the numerical results of Levermore and Liu [11,12] have suggested that the binary oscillations whose behavior they have studied in a discrete Hopf equation can arise naturally from the local breakdown of an initially smooth profile. In contrast to these well known examples of modulated wavetrains, the single phase wavetrains discussed in this paper have *not* been observed to appear spontaneously. We want to emphasize that we make *no* such claim of this kind of universality of our plane wave solutions.

We have said nothing about the resulting local behavior of the lattice when the modulated harmonic plane wave picture breaks down. It is possible that when this happens, a more complicated multiperiodic microstructure appears, which may yield to analysis similar to that we have presented above. The complete macroscopic picture of the lattice then corresponds to the interactions of several spatially separated phases (in the sense of thermodynamics). Each such phase has a nonequilibrium thermodynamical description given by the appropriate set of modulation equations, and as the modulation equations become locally singular, new microscopic behavior takes over. It is believed that such a picture of a cascade of interacting phases is valid in microscopic systems that are integrable, but in nonintegrable problems, it is possible for local dynamics to appear that are not sufficiently constrained by constants of motion to be called quasiperiodic, and thus are best referred to as a chaotic phase. We point out that Levermore and Liu [11,12] observed the appearance of such a chaotic phase in their numerical experiments with a nonintegrable discrete Hopf equation. Our preliminary numerical experiments indicate that, in the DNLS equation, there may be at least one genuine multiperiodic phase between the breakdown of the harmonic plane waves and the onset of a local chaotic phase. We plan to catalog the phases we observe and obtain a modulational description for them.

How would the procedure presented in Section 2 be adapted to the analysis of a more complicated local ansatz? Let us review in some generality the procedure used to obtain modulation equations for the plane waves. The algorithm followed above was a special case of the averaging of local conservation laws. The idea is to identify a family of quasiperiodic solutions to the lattice model of the form $A_j(t) = \phi(\theta_1, \theta_2, \dots)$ where $\theta_i = k_i j - \omega_i t$ and ϕ is a periodic function of each of its arguments independently. In the plane wave case above, the family depended on only a single phase variable θ. The solution family is inserted into the local conservation laws and the family's parameters (which are an independent subset of the k_i, ω_i, and possibly other parameters describing, say, amplitudes) are allowed to vary slowly across the lattice on the scale $x = jh$. In general, the density and flux terms will still contain rapidly oscillating degrees of freedom that must be eliminated on the macroscopic scale. Averaging over any such remaining dependence on the fast phase variables θ_i results in PDE's which govern the evolution of the slowly varying parameters[5]. Finally, comparing the magnitudes of the den-

---

[5] Note that in the plane wave case, the limiting density and flux expressions no longer involved any fast oscillations; this is

sities and fluxes in the limit $h \downarrow 0$ gives the time scale that corresponds in the macroscopic limit to the spatial scale of $x = jh$.

In principle, the procedure outlined above could be carried out for general multiphase solutions of a general member of the flow family (5), although the practicality of this is governed by two constraints: the kind of information one is able to obtain about families of exact multiperiodic solutions, and the availability of a sufficient number of local conservation laws to control the motion [1]. It is not necessary to obtain an analytical expression for the solution family; since the modulation equations in general only involve certain averages of functionals of the solution family evaluated in terms of the family's parameters, one only needs to be convinced that a solution family exists and be able to construct the appropriate averages. Even this procedure appears to be quite difficult in the general (nonintegrable) context. This is because understanding properties of multiphase waves of the form $\phi(\theta_1, \ldots, \theta_n)$ in discrete nonintegrable problems amounts to classifying and analyzing the solutions of a nonlinear complex (partial) differential equation with delay and advance terms. We believe that this analysis is a problem best attacked perturbatively (e.g. in the case of weak nonlinearity or extreme wavenumbers).

On the other hand, in the integrable case of the Ablowitz–Ladik equation, much more is possible on a purely formal level without resorting to the kind of analysis mentioned above. The integrable structure provides an inexhaustible number of conservation laws, as well as a hierarchy of exact quasiperiodic solution families constructed from Riemann theta functions. A further benefit of working with the modulational behavior of integrable systems such as the Ablowitz–Ladik equation is that it is easy to place the modulation equations in Riemann invariant form regardless of the number of modulational degrees of freedom; thus, the characteristic velocities are easily recovered, and hyperbolic structure of the equations can then be discerned. The recent work of Miller,

Ercolani, Krichever, and Levermore [13] describes the class of microscopic oscillations in the Ablowitz–Ladik system as well as the formalism of averaging the Ablowitz–Ladik conservation laws about multiphase wavetrain solutions, following a procedure first used by Flaschka, Forest, and McLaughlin [7] on the Korteweg–deVries equation, and recently applied to the Toda lattice [15,4]. The analysis of the hyperbolic structure of the corresponding modulation equations will be developed in a future work along with the asymptotic modulation theory of lattice models that can be considered to be close to the Ablowitz–Ladik model.

Finally, we plan to investigate the possibility of studying the regularization of hyperbolic shocks and perhaps even change of type singularities through a small dissipation limit. We would replace the underlying Hamiltonian lattice system with a nearby system of discrete complex Ginzburg-Landau type, and derive small dissipative corrections to the modulation equations, leading us to an analysis of the shock structure, including the shock speed. In this way, we can understand the motion of hyperbolic shock boundaries between stable periodic states of the underlying microscopic lattice.

Acknowledgements

---

because the oscillatory motion in the plane wave ansatz is a local gauge transformation related to the global gauge symmetry of the equation described by $A_j(t) \rightarrow e^{i\alpha} A_j(t)$. Hence, the explicit averaging over fast phases was unnecessary for this ansatz.

Appendix A. Legendre transform and averaging of Lagrangians

In this appendix we will set up the Hamiltonian formulation of (5) and proceed to derive modulation equations for a uniform harmonic wavetrain using Whitham's averaged Lagrangian principle. The reader will recall that we are considering the system

$$i\dot{A}_j + f(|A_j|^2)(A_{j+1} + A_{j-1}) + F(|A_j|^2)A_j = 0. \tag{A.1}$$

This system admits a Hamiltonian structure, as well as a second independent conserved quantity. Indeed, define the functions $g(\rho)$ and $G(\rho)$ by

$$g(\rho) = \int_0^\rho \frac{dy}{f(y)} \quad \text{and} \quad G(\rho) = \int_0^\rho \frac{F(y)\,dy}{f(y)}. \tag{A.2}$$

It is easily verified that (A.1) may be expressed as

$$i\dot{A}_j = f(|A_j|^2)\frac{\partial H}{\partial A_j^*}, \tag{A.3}$$

with the Hamiltonian function

$$H = \sum_j A_j^*(A_{j+1} + A_{j-1}) + \sum_j G(|A_j|^2). \tag{A.4}$$

This is the same function as the Hamiltonian presented in the main text, written here with a different summand to clarify the dependence on A_j and A_j^* as independent dynamical variables. The gradient is derived from a Poisson bracket, and the nonzero fundamental brackets are given by

$$\{A_j, A_j^*\} = -if(|A_j|^2), \tag{A.5}$$

or, in general,

$$\{B, C\} = -i\sum_j \left(\frac{\partial B}{\partial A_j}\frac{\partial C}{\partial A_j^*} - \frac{\partial B}{\partial A_j^*}\frac{\partial C}{\partial A_j}\right)f(|A_j|^2), \tag{A.6}$$

for any functions B and C of the field variables. Given the gauge symmetry $A_j \to e^{i\alpha}A_j$ of (A.1), we may obtain a second conserved quantity via Noether's theorem as

$$N = \sum_j g(|A_j|^2). \tag{A.7}$$

There is a family of plane wave solutions for (A.1), and we would like to apply Whitham's averaged variational principle [16] to this family of solutions and obtain the PDE's which govern the evolution of the plane wave parameters. In order to apply this method, we must pass from a Hamiltonian formulation of (A.1) to a Lagrangian formulation. Since our Hamiltonian is expressed in terms of noncanonical variables, we must find canonical variables before carrying out the Legendre transform from which we obtain the Lagrangian formulation [3]. In what follows we obtain such a Lagrangian from a specific choice of canonical variables.

The simplest set of canonical variables to work with are those that preserve the form of our plane wave ansatz (2). Accordingly, we propose

$$\psi_j = A_j P(|A_j|^2) \quad \text{and} \quad \psi_j^* = A_j^* P(|A_j|^2), \tag{A.8}$$

where

$$P(x) = \sqrt{\frac{g(x)}{x}} . \tag{A.9}$$

$P(x)$ exists and is nonsingular due to the definition of $g(x)$ and the restrictions placed on $f(x)$. Let $\bar{g}(x)$ be the compositional inverse of $g(x)$. Again, this exists due the definition of $g(x)$ and the restrictions placed on $f(x)$. The inverse transformation to (A.8) is given by

$$A_j = \psi_j M(|\psi_j|^2) \quad \text{and} \quad A_j^* = \psi_j^* M(|\psi_j|^2) , \tag{A.10}$$

where

$$M(x) = \sqrt{\frac{\bar{g}(x)}{x}} . \tag{A.11}$$

In the ψ variables, the plane wave solution becomes

$$\psi_j = AP(|A|^2)e^{i(kj-\omega t)} . \tag{A.12}$$

We also have

$$\{\psi_j, \psi_j^*\} = -i , \tag{A.13}$$

with all other Poisson brackets vanishing – thus we have constructed canonical variables for (A.1). The Legendre transform of H is then

$$L = \sum_j \mathcal{L}_j = \frac{i}{2} \sum_j \left(\dot{\psi}_j \psi_j^* - \psi_j \dot{\psi}_j^* \right) - H , \tag{A.14}$$

with H written in terms of the ψ variables. Recalling the definition $\rho = |A|^2$ and evaluating the Lagrangian density $\mathcal{L}_j$ on the ansatz (A.12) gives

$$\mathcal{L}_j = \omega g(\rho) - 2\rho \cos k - G(\rho) . \tag{A.15}$$

We are now in a position to apply Whitham's averaging method. Introduce the usual phase variable θ by the expressions

$$k = \theta_x, \quad \omega = -\theta_t. \tag{A.16}$$

In addition, we allow the parameters θ and ρ to vary slowly in such a way that the sum (A.14) may be approximated by the integral

$$\bar{L} = - \int \left[g(\rho)\theta_t + 2\rho \cos \theta_x + G(\rho) \right] dx . \tag{A.17}$$

Whitham proposes the following variational principle: the evolution of the modulated plane wave solution extremizes the action functional

$$\int_{t_0}^{t_1} \bar{L} \, dt , \tag{A.18}$$

with respect to independent variations in θ and ρ.

The variation with respect to ρ results in an algebraic equation which we supplement with the consistency relation

$$k_t + \omega_x = 0$$

to obtain a PDE. This relation imposes "conservation of waves." We obtain the second equation by varying with respect to θ. The resulting system is given by

$$\frac{\partial}{\partial t} g(\rho) - \frac{\partial}{\partial x} \left[2\rho \sin k \right] = 0 \,,$$
$$\frac{\partial}{\partial t} k + \frac{\partial}{\partial x} \left[2 f(\rho) \cos k + F(\rho) \right] = 0 \,. \tag{A.19}$$

We can also rewrite the first of these equations in the equivalent form

$$\frac{\partial}{\partial t} \rho - f(\rho) \frac{\partial}{\partial x} \left[2\rho \sin k \right] = 0 \,. \tag{A.20}$$

These expressions are the same as those derived by averaging local conservation laws.

Appendix B. Linear stability analysis

Here, we compute the conditions for linear stability of the uniform harmonic wavetrain solutions to the DNLS equation. We begin with the system in the form of (1). Under the transformations

$$\Phi_j(t) = A_j(t/\varepsilon) e^{2it} \quad \text{and} \quad \Gamma = \gamma/\varepsilon \,, \tag{B.1}$$

the system takes the form

$$i\dot{\Phi}_j + \Phi_{j+1} + \Phi_{j-1} + \Gamma |\Phi_j|^2 \Phi_j = 0 \,. \tag{B.2}$$

This system possesses uniform wavetrain solutions given by

$$\Phi_j(t) = a e^{i(kj - \omega(k,a)t)} \,, \tag{B.3}$$

where ω is given by the amplitude dependent dispersion relation

$$\omega(k,a) = -2\cos k - \Gamma|a|^2 \,, \tag{B.4}$$

and $k \in \mathcal{R}$ while $a \in \mathcal{C}$. Consider choosing particular values for k and a and adding a "rotating" complex-valued perturbation $p_j(t)$ to the wavetrain solution:

$$\Phi_j(t) = (a + p_j(t)) e^{i(kj - \omega(k,a)t)} \,. \tag{B.5}$$

The reduced DNLS equation (B.2) then becomes

$$i\dot{p}_j + (2\Gamma|a|^2 + \omega(k,a) + \Gamma a^* p_j + 2\Gamma a p_j^* + \Gamma|p_j|^2) p_j$$
$$+ \Gamma a^2 p_j^* + e^{ik} p_{j+1} + e^{-ik} p_{j-1} = 0 \,, \tag{B.6}$$

which, in the case of infinitesimal perturbations $p_j(t)$ becomes

$$i\dot{p}_j + (2\Gamma|a|^2 + \omega(k,a)) p_j + \Gamma a^2 p_j^* + e^{ik} p_{j+1} + e^{-ik} p_{j-1} = 0 \,. \tag{B.7}$$

If we let

$$a = \sqrt{\rho}e^{i\alpha} ,$$

$$p_j(t) = [U(t)e^{i\Delta k} + \text{c.c.}] + i[V(t)e^{i\Delta k} + \text{c.c.}] , \tag{B.8}$$

Eq. (B.7) takes the form

$$\frac{d}{dt}\begin{bmatrix} U \\ V \end{bmatrix} = B\begin{bmatrix} U \\ V \end{bmatrix} , \tag{B.9}$$

where the matrix B is given by

$$B = \begin{bmatrix} -\Gamma\rho\sin 2\alpha - 2i\sin\Delta k\sin k & (\cos 2\alpha - 1)\Gamma\rho + 2(1 - \cos\Delta k)\cos k \\ (\cos 2\alpha + 1)\Gamma\rho - 2(1 - \cos\Delta k)\cos k & \Gamma\rho\sin 2\alpha - 2i\sin\Delta k\sin k \end{bmatrix} . \tag{B.10}$$

Because we have selected a rotating frame, Δk represents a *relative* wavenumber. The perturbation $p_j(t)$ will be linearly unstable if there are any eigenvalues of B in the right half-plane. The matrix B either has both eigenvalues pure imaginary or one eigenvalue on either side of the imaginary axis. The second case occurs if

$$\Gamma\rho > (1 - \cos\Delta k)\cos k , \tag{B.11}$$

for $\cos k > 0$ or

$$\Gamma\rho < (1 - \cos\Delta k)\cos k , \tag{B.12}$$

for $\cos k < 0$ which is a sufficient condition for the growth of the perturbation of relative wavenumber Δk.

References

[1] M.J. Ablowitz and D.J. Benney, The evolution of multiphase modes for nonlinear dispersive waves, Stud. Appl. Math. 42(3) (1970) 225–238.

[2] M.J. Ablowitz and J.F. Ladik, Nonlinear differential-difference equations and Fourier analysis, J. Math. Phys. 17 (1976) 1011–1018.

[3] V.I. Arnol'd, Mathematical Methods of Classical Mechanics, 2nd edition (Springer, Berlin, 1989).

[4] A.M. Bloch and Y. Kodama, Dispersive Regularization of the Whitham Equation for the Toda Lattice, SIAM J. Appl. Math. 52(4) (1992) 909–928.

[5] A.S. Davydov, Solitons in Molecular Systems (Kluwer, Dordrecht, 1985).

[6] J.C. Eilbeck, P.S. Lomdahl and A.C. Scott, The discrete self-trapping equation, Physica D 16 (1985) 318–338.

[7] H. Flaschka, M.G. Forest and D.W. McLaughlin, Multiphase averaging and the inverse spectral solution of the Korteweg-de Vries equation, CPAM 33 (1980) 739–784.

[8] B.M. Herbst and M.J. Ablowitz, Numerically induced chaos in the nonlinear Schrödinger equation, Phys. Rev. Lett. 62(18) (1989) 2065–2068.

[9] B.L. Holian, H. Flaschka and D.W. McLaughlin, Shock waves in the Toda lattice: Analysis, Phys. Rev. A 24(5) (1981) 2595–2623.

[10] Y.S. Kivshar and M. Peyrard, Modulational Instabilities in Discrete Lattices, Phys. Rev. A 46 (1992) 3198–3205.

[11] C.D. Levermore and J.G. Liu, Oscillations arising in numerical experiments, in: Singular Limits of Dispersive Waves, N. Ercolani, I. Gabitov, D. Levermore and D. Serre, eds., NATO ARW series (Plenum, New York, 1994).

[12] C.D. Levermore and J.G. Liu, Large oscillations arising in a dispersive numerical scheme, University of Arizona Preprint (1994).

[13] P.D. Miller, N.M. Ercolani, I.M. Krichever and C.D. Levermore, Finite genus solutions to the Ablowitz–Ladik equations, University of Arizona Preprint (1994).

[14] M. Salerno, Quantum deformations of the discrete nonlinear Schrödinger equation, Phys. Rev. A 46(11) (1992) 6856–6859.

[15] S. Venakides, P. Deift and R. Oba, The Toda shock problem, CPAM 44 (1991) 1171–1242.

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

Erratum

Erratum to "Macroscopic lattice dynamics" [Physica D 79 (1994) 1–15] *

M.H. Hays [a,1], C.D. Levermore [b,2], P.D. Miller [a,3]

[a] *Program in Applied Mathematics, University of Arizona, Tucson, AZ 85721, USA*
[b] *Department of Mathematics and Program in Applied Mathematics, University of Arizona, Tucson, AZ 85721, USA*

Some physically inconsequential sign errors appeared in this paper. The first of these occurs in Eq. (15), which should read

$$\frac{\partial \rho}{\partial T} + f(\rho) \frac{\partial}{\partial x} \left[2\rho \sin k \right] = 0 \,,$$

$$\frac{\partial k}{\partial T} - \frac{\partial}{\partial x} \left[2f(\rho) \cos k + F(\rho) \right] = 0 \,.$$

This error (in the sign of the time variable, really) also occurs in Appendix A – first in Eq. (A.3) which should read

$$i\dot{A}_j + f(|A_j|^2) \frac{\partial H}{\partial A_j^*} = 0 \,,$$

and then propagating into Eqs. (A.14), (A.15), (A.17), (A.19), and (A.20). The corresponding sign of the characteristic velocity $\lambda_\pm$ appearing in Eq. (17) should also be changed (although Eq. (18) is correct), as well as the sign of the matrix J in Eq. (21).

The reader should also be advised that there was a typographical error in Eq. (13), the final line of which should read

$$R = 2\rho \Big(2f(\rho) \cos k + F(\rho) \Big) \sin k \,,$$

and that the two symbols ε and ϵ that appear throughout the text should be interpreted identically.