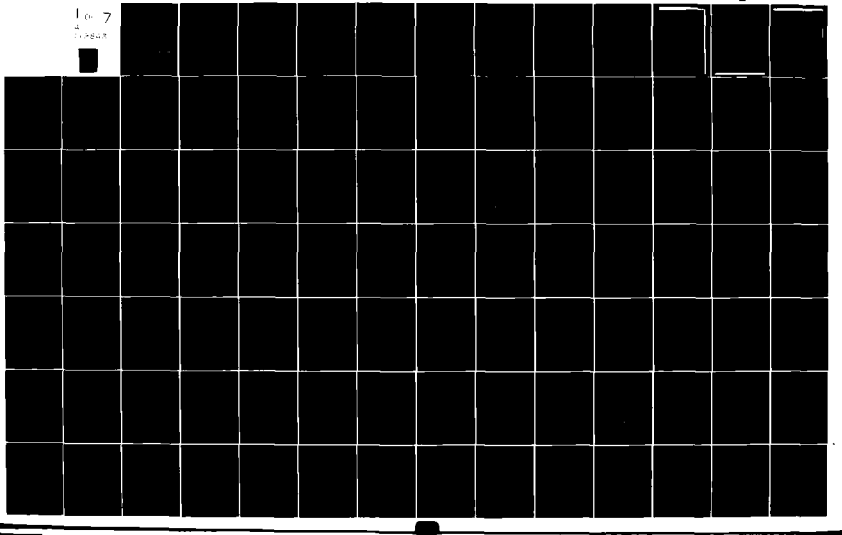
AD-A119 843 CLARKSON COLL OF TECHNOLOGY POTSDAM NY DEPT OF MATHE--ETC F/G 12/1
NONLINEAR WAVES.(U)
SEP 82 M J ABLOWITZ, D J KAUP, A C NEWELL N00014-76-C-0867
UNCLASSIFIED NL



AD A119843

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD A119843 (13) / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) NONLINEAR WAVES	5. TYPE OF REPORT & PERIOD COVERED Final Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) M.J. Ablowitz, D.J. Kaup and A.C. Newell	8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0867
9. PERFORMING ORGANIZATION NAME AND ADDRESS Clarkson College of Technology Department of Mathematics & Computer Science Potsdam, NY 13676	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Department of the Navy Office of Naval Research Arlington, Virginia 22217	12. REPORT DATE September 23, 1982
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE OCT 4 1982 S D H

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The areas of study during the period of this grant include: A class of physically significant singular nonlinear integro-differential equations; the Riemann-Hilbert boundary value problem and the Inverse Scattering Transform method; Direct linearizing methods for certain nonlinear wave equations and associated nonlinear ordinary differential equations; Focussing instabilities and the propagation of water waves; Ordinary differential equations of Painelvé type and numerical schemes for nonlinear wave equations solvable by the Inverse Scattering Transform method. Research has also been carried out into the following nonlinear

DD FORM 1 JAN 73 1473

DTIC FILE COPY

82 09 29 026

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

systems: Stimualted Raman Scattering and two-photon absorption; Nonlinear resonances and spherical solitons; Nonlinear interactions of Whistler Waves; Exact three-dimensional solutions of the three-wave resonant interaction; Discrete versions of the nonlinear Schrödinger Equation and Thermal effects on soliton propagation.

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A

DTIC COPY INSPECTED 8

NONLINEAR WAVES

Final Technical Report of ONR Contract #N00014-76-C-0867

Investigators: M. J. Ablowitz
D. J. Kaup
A. C. Newell

M.J. Ablowitz

During the period of this grant the grant funds have partially supported myself and the following individuals: Assistant Professor A.S. Fokas, Post-doctoral Associate Y. Kodama, A. Nakamura, J. Satsuma; Research Assistant T. Taha. The areas of study include:

a) A class of physically significant singular nonlinear integro-differential equations. Applications include the propagation of long internal waves in stratified fluids (with and without shear).

b) The Riemann-Hilbert boundary value problem and the Inverse Scattering Transform method. Applications include the Intermediate Long Wave Equation and the Benjamin-Ono Equation.

c) Direct linearizing methods for certain nonlinear wave equations and associated nonlinear ordinary differential equations.

d) Focussing instabilities and the propagation of water waves.

e) Ordinary differential equations of Painlevé type.

f) Numerical schemes for nonlinear wave equations solvable by the Inverse Scattering Transform method.

Publications during this period were as follows:

1. Nonlinear Wave Propagation, M.J. Ablowitz, published in Encyclopedia of Physics, pp. 663-664, 1981.

2. Remarks on Nonlinear Evolution Equations and Ordinary Differential Equations of Painlevé Type, by M.J. Ablowitz, Proceedings of Conference on Solitons held in Kiev, U.S.S.R., September, 1979, as part of a Joint U.S. - U.S.S.R. Academy of Sciences agreement, Physica 3D, 1 & 2, pp. 129-141, 1981.

3. Finite Perturbations and Solutions of the Korteweg-deVries Equation, H. Airault and M.J. Ablowitz, C.R. Academy of Science, Paris, 6292, pp. 279-281, 1981.

4. The Nonlinear Intermediate Long Wave Equation: Analysis and Method of Solution, Y. Kodama, J. Satsuma and M.J. Ablowitz, Phys. Rev. Lett., Vol. 46, 11, pp. 687-690, 1981.

5. Linearization of the Kdv and Painlevé II equations, A.S. Fokas and M.J. Ablowitz, Phys. Rev. Lett., Vol. 47, No. 16, pp. 1096-1100, October, 1981.

6. Perturbations of Solitons and Solitary Waves, Y. Kodama and M.J. Ablowitz, Stud. in Appl. Math., 64, pp. 225-245, 1981.

7. The Periodic Cubic Schrödinger Equation, Y.C. Ma and M.J. Ablowitz, Stud. in Appl. Math., 65, pp. 113-158, 1981.

8. Remarks on Nonlinear Evolution Equations and the Inverse Scattering Transform, M.J. Ablowitz, Nonlinear Phenomena in Physics and Biology, pp. 83-94 (Edited by R.H. Enns, B.L. Jones, R.M. Miura and S.S. Rangnekar, Plenum Publishing Corporation 1981).

9. Direct Linearization of the Korteweg-deVries Equation, A.S. Fokas and M.J. Ablowitz, American Institute of Physics Conference Proceedings, No. 88, Mathematical Methods in Hydrodynamics and Integrability in Dynamical Systems, (La Jolla Institute 1981) pp. 237-241 (Edited by M. Tabor and Y.M. Treve).

10. A Direct Linearization Associated with the Benjamin-Ono Equation, M.J. Ablowitz and A.S. Fokas, American Institute of Physics Conference Proceedings, No. 88, Mathematical Methods in Hydrodynamics and Integrability in Related Dynamical Systems, (La Jolla Institute 1981) pp. 229-236 (Edited by M. Tabor and Y.M. Treve).

11. Note on Asymptotic Solutions of the Korteweg-deVries Equation with Solitons, M.J. Ablowitz and Yuji Kodama, Studies in Applied Mathematics, 66, pp. 159-170 (1982).

12. Direct and Inverse Scattering Problems of Nonlinear Intermediate Long Wave Equations, Y. Kodama, M.J. Ablowitz and J. Satsuma, J. Math. Phys., Vol. 23, No. 4, pp. 564-576, 1982.

13. On the Periodic Intermediate Long Wave Equation, M.J. Ablowitz, A.S. Fokas, J. Satsuma and H. Segur, J. Phys. A: Math. Gen. 15, pp. 781-786, 1982.

14. On a Unified Approach to Transformation and Elementary Solutions of Painlevé Equations, A.S. Fokas and M.J. Ablowitz, to be published in J. Math. Phys., 1982.

15. On the Inverse Scattering and Direct Linearizing Transforms for the Kadomtsev-Petviashvili Equation, A.S. Fokas and M.J. Ablowitz, I.F.N.S. #9, submitted for publication, 1982.

16. The Direct Linearizing Transform and the Benjamin-Ono Equation, M.J. Ablowitz, A.S. Fokas and R.L. Anderson, I.F.N.S. #10, submitted for publication, 1982.

17. The Inverse Scattering Transform for the Benjamin-Ono Equation - A Pivot to Multidimensional Problems, A.S. Fokas and M.J. Ablowitz, I.F.N.S. #11, accepted for publication in Stud. Appl. Math., 1982.

18. On a Bäcklund Transformation and Scattering Problem for the Modified Intermediate Long Wave Equation, J. Satsuma, T.Taha and M.J. Ablowitz, I.F.N.S. #12, submitted for publication, 1982.

19. On Analytical & Numerical Aspects of Certain Nonlinear Evolution Equations, Part I: Analytical, I.F.N.S. #14, submitted for publication, 1982.

20. On Analytical and Numerical Aspects of Certain Nonlinear Evolution Equations, Part II: Numerical, Nonlinear Schrödinger Equations, T. Taha and M.J. Ablowitz, I.F.N.S. #15, submitted for publication, 1982.

21. On Analytical & Numerical Aspects of Certain Nonlinear Evolution Equations, Part III: Numerical, Korteweg-deVries Equation, T. Taha and M.J. Ablowitz, I.F.N.S. #16, submitted for publication, 1982.

D.J. Kaup

During the past year, I have carried out research and investigations into the following nonlinear systems and areas.

i) Stimulated Raman Scattering and two-photon absorption (with H. Steudel),

ii) Nonlinear resonances and spherical solitons,

iii) Nonlinear interactions of Whistler Waves (with S. Antani),

iv) Exact three-dimensional solutions of the three-wave resonant interaction,

v) Discrete versions of the nonlinear Schrödinger Equation (with S. Trullinger, USC), and

vi) Thermal effects on soliton propagation.

Publications during the last year consisted of:

1. Nonlinear Resonance and Colliding Spherical Ion-Acoustic Solitons (Physica 2D, 389-394 (1981).

2. Multi-Shock Solutions of Random Phase Three-Wave Interactions, A. Reiman and D.J. Kaup, Phys. Fluids 24, 228-32 (1981).

3. The Solutions of the General Initial Value Problem for the Full Three Dimensional Three-Wave Resonant Interaction, D.J. Kaup, Physica 3D, 374-95 (1981).

4. The Lump Solutions and the Bäcklund Transformation for the Three-Dimensional Three-Wave Resonant Interaction, J. Math. Phys. 22, 1176-1181 (1981).

5. Whistler Wave Self-Modulation in a Tokamak Plasma, D.J. Kaup and S.N. Antani, Phys. Fluids 24, pp. 1391-3 July (1981).

6. The Linearity of Nonlinear Soliton Equations and the Three-Wave Resonance Interaction, article in Nonlinear Phenomena in Physics and Biology, pp. 95-123, Ed. by Enns, Jones, Miura, and Rangnekar (Plenum, 1981).

7. The Method of Estabrook and Wahlquist, article in Mathematical Methods in Hydrodynamics and Integrability in Dynamical Systems, pp. 193-210, edited by M. Tabor and Y.M. Treve (Am. Inst. Physics, New York, 1982).

A.C. Newell

In the past two years this investigator has been involved in a number of studies falling under the general heading of Nonlinear Processes.

1. Soliton Mathematics: (With H. Flaschka)

In these studies our goal has been to understand the interconnections between soliton mathematics and solvable models in statistical mechanics and quantum field theory. Out of this study has come a number of papers, reprints of which are enclosed or listed.

1.1 Monodromy - and Spectrum-Preserving Deformations I, H. Flaschka and A.C. Newell, Commun.math.Phys. 76,65-116 (1980).

1.2 The Inverse Monodromy Transform Is A Canonical Transformation, H. Flaschka and A.C. Newell.

1.3 Multiphase Similarity Solutions of Integrable Evolution Equations, H. Flaschka and A.C. Newell, Physica 3D 1 & 2, pp. 203-221 (1981).

1.4 A series of 6 papers - Kac Moody algebras and soliton equations.

2. Propagation of solitary surface and internal waves in slowly varying depths: (With C.J. Knickerbocker)

This turned out to be an exciting study in which we answered many open questions.

2.1 Shelves and the Korteweg-deVries Equation, C.J. Knickerbocker and A.C. Newell, J. Fluid Mech. Vol. 98, pp. 803-818 (1980).

2.2 Internal Solitary Waves Near A Turning Point, C.J. Knickerbocker and A.C. Newell, Physics Letters, Vol. 75A, No. #5, PP. 326-330 (1980).

2.3 Reflections from Solitary Waves in Nonuniform Media, C.J. Knickerbocker and A.C. Newell.

82 10 04 139

3. The Phenomenon of Self-Focusing as a Mechanism for Initiating Subcritical Instabilities.

The basic idea here is that focusing is a dynamic mechanism by which a system can reach its finite amplitude instability threshold locally without benefit of large initial perturbations or the effects of imperfections.

3.1 Bifurcation and Nonlinear, A.C. Newell, from Pattern Formation and Recognition, Ed. H. Haken, Springer (1979).

3.2 The Mechanism by Which Many Partial Difference Equations Destabilize, W. Briggs, A.C. Newell and T. Sarie, from Chaos and Order in Nature, Ed. H. Haken, Springer (1981).

3.3 Focusing: A Mechanism for Instability on Nonlinear Finite Difference Equations, W.L. Briggs, A.C. Newell and T. Sarie (1982).

4. Chaos in Partial Differential Equations

We seek to determine what significant new qualitative features appear when one forces mechanical systems described by PDE's into the chaotic regime.

4.1 Chaos in the Inhomogeneously Driven Sine-Gordon Equation, J.C. Eilbeck, P.L. Lomdahl and A.C. Newell, Physics Letters, Vol. #87A, No. 1,2, pp. 1-4 (1981).

and above 8 kHz.) More elaborate instru-
ies portable, are widely used for analysis of
ve, fractional octave, or narrow frequency
re detailed information is usually required for
r engineering control of noise at the source.
ly for regulatory purposes.
escribe the varying noise levels that constitute
te at a given location, instruments have been
ample the noise level (usually *A*-weighted) at
of time. In this way the statistical descriptors
L, indicating the noise level *L* decibels that is
rcent of the time: L_1 or L_5 are often taken to
ximum noise levels occasionally attained, and
mbient noise level.
table device is the noise dose meter, worn by
ng in noisy locations. This instrument sums
sure of the wearer, taking into account both
ime of exposure and the noise level.
oise control procedure is to enclose the noise
completely or partially, taking certain precau-
ng the absorption of sound, though this is often
solution for a variety of reasons. Noise re-
lesign of the source usually requires expert
what mechanisms are producing the noise,
ated, how this can be reduced (many noise
ques are specific to certain processes), and
onacoustical constraints such as weight, cost,
equirements, etc. Once noise escapes the vi-
ource the intensity and spectrum shape of that
the receiver depends on the possible propa-
Inside a building noise may be either airborne
orne, and the noise level depends on numerous
s the shape and size of rooms, whether they
nt or absorbing, how the absorption is distrib-
iency of vibration isolation in solid structures,
the noise level at the receiver depends on the
edance of the ground surface, whether or not
height of the source and receiver above the
their distance apart. Wind velocity and tem-
their variation with height, and also molecular
re all important at distances of the order of 100
ater.
of noise on human activity are extensive. De-
s frequency content, intensity, time duration,
nces under which it is heard, noise can be ben-
masking sounds of high message content that
vise be distracting—annoying, or hazardous.
noying effects are interference with speech and
nce of mental tasks, and degradation of the en-
the listener. In most circumstances psycho-
on to noise increases with increasing amounts
there is considerable variability among individ-
ing on past experience, the nature of the noise,
ion of quietening it. Long or repeated exposure
of noise presents a risk of hearing damage: the
nage and its effect on hearing depend in a com-
duration, intermittency, frequency content, and
d also on the physiological susceptibility of the
Noise disturbs sleep, either by awakening or by
lepth of sleep, to an extent that depends on the
intensity and duration of the noise—whether this is merely annoying or has more serious consequences is not yet established except in extreme cases.

See also ACOUSTICS; ACOUSTICS, ARCHITECTURAL.

BIBLIOGRAPHY

L. L. Beranek (ed.), *Noise and Vibration Control*, McGraw-Hill, New York, 1971.
C. M. Harris (ed.), *Handbook of Noise Control*, 2nd ed. McGraw-Hill, New York, 1979. (I)
L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics*, 2nd ed. Wiley, New York, 1962. (E)
P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, McGraw-Hill, New York, 1968. (A)

Nonlinear Wave Propagation

Mark J. Ablowitz

Wave propagation is essentially the study of how information, or some type of signal, is transmitted. The generality in the notion of what actually constitutes a wave, or wave phenomena, necessarily makes the study rather broad and of interest to scientists and mathematicians alike. Applications abound: for example, water waves, acoustic waves, laser beams, traffic flow, and electromagnetic waves. Here we discuss some of the mathematical theories and techniques that have been developed for nonlinear waves. By nonlinear wave propagation we refer to those phenomena whose underlying equations (and hence the basic physical mechanism) is nonlinear.

The discipline of nonlinear wave propagation began in the nineteenth century with the examination of surface waves by Stokes (1847). For deep-water gravity waves, he found that the phase speed of the wave was amplitude dependent. The larger the amplitude, the faster the wave train moves. Thus, if we call c the phase velocity, k the wave number, and a the amplitude, Stokes found that $c = c(k,a)$. In linear problems, $c = c(k)$ only. This concept has been used quite recently and will be discussed later in this article.

At approximately the same time, Riemann was also studying nonlinear wave phenomena. In fact, in his classic paper (1858) on gas dynamics, he discovered a method to solve a certain system of nonlinear partial differential equations (PDE's). The ideas have been generalized, and new methods developed for a class of equations called hyperbolic systems. These are PDE's that have distinct real characteristics. A discussion of these ideas can be found in Courant and Friedrichs (1948) and Courant and Hilbert (1961). Perhaps the most outstanding new phenomenon of this nonlinear theory is the appearance of a shock wave. A shock is the abrupt change in the dependent variable(s) of the underlying equation(s). In gas dynamics these variables are the pressure and the velocity.

In recent years there have been significant advances in the solution of nonlinear dispersive wave problems. By this we mean those equations that, when linearized, have the phase speed depending nontrivially on k; i.e., $c = c(k)$. The first important step was the understanding of the effects of nonlinearity on wave interactions. These ideas are often called

weak interaction theory, because here the nonlinearity is assumed to be small. Very generally, we consider an equation of the form

$$L(u) = \epsilon N(u) \tag{1}$$

where ϵ is very small, and L and N are linear and nonlinear operators, respectively, that depend on the particular physical problem. Equation (1) is assumed to have elementary dispersive wave solutions $u_i = A_i \cos \theta_i$, $\theta_i = k_i x - \omega(k_i)t$ (assuming one spatial dimension and time) in the linear limit $\epsilon = 0$. $\omega(k_i)$ is called the dispersion relation and is related to phase speed by $c_i = \omega_i/k_i$. These solutions do not obey the full equation (1). However, approximate solutions can be constructed by using multiple-scale perturbation methods (see, e.g., Cole, 1968). It turns out that, in general, the nonlinearity creates resonant interactions between the solutions u_i. This in turn causes the amplitudes A_i to depend weakly on time (typically $A_i = A_i(\epsilon t)$). The effect can be substantial. A mode that initially is absent can "feed" off the others, and eventually become equally important. References regarding this work may be found in Phillips (1974) and Bloembergen (1965).

In 1965 Whitham discovered how to develop certain types of approximate solutions to fully nonlinear partial differential equations. The starting point is that for certain equations a single periodic uniform nonlinear "mode" is an exact solution, e.g., the Stokes water-wave solution. Whitham showed how to generate simplified approximate equations that govern gradually modulated nonuniform waves (e.g., a wave in a slowly varying medium). The methods can be viewed as a generalization of the WKB method to nonlinear partial differential equations, and can be put into an elegant variational formulation. The theory provides information as to how energy propagates, and gives a concept of group velocity in a nonlinear problem. In the linear theory, the group velocity is the velocity at which energy is transferred. The nonlinearity has the effect of splitting the Whitham's velocity into more than one velocity. Although the linear theory was applied only to a single-phase ($\theta = kx - \omega t$) mode, the ideas can be extended to multiphase modes, as noted by Ablowitz and Benney (1970).

The foregoing theories have led to important new discoveries. One of the most famous is the work of Benjamin (1967). He showed, by using weak interaction theory (the result can also be deduced using Whitham's technique), that the Stokes water wave was unstable. This surprising fact (and the very nature of the instability) has stimulated a great deal of research.

At approximately the same time, Kruskal and Zabusky were studying special aperiodic solitary-wave solutions of the physically interesting Korteweg–de Vries (KdV) equation (1895),

$$u_t + 6uu_x + u_{xxx} = 0. \tag{2}$$

A solitary-wave solution to (2) is given by

$$u = A \operatorname{sech}^2(\sqrt{A/2}(x - 2At)) \tag{3}$$

They discovered the remarkable fact that two such waves with different amplitudes interact elastically. Specifically, the initial velocity, before interaction, and the asymptotic final velocity, after interaction, are equal! They named this solitary wave a *soliton*, in conceptual analogy to elementary-particle interactions. Indeed, in 1975, some physicists suggested that an elementary particle is actually a soliton.

This result shortly preceded the pioneering work of Gardner, Greene, Kruskal, and Miura (GGKM) (1967, 1974) in which they developed a method to solve the initial-value problem for (2), for given $u(x,0)$ decaying sufficiently rapidly as $|x| \to \infty$. The key step is to associate with the KdV equation the linear Schrödinger eigenvalue problem

$$V_{xx} + (\lambda + u(x,t))V = 0. \tag{4}$$

Remarkably, Eq. (4) had been studied intensively by theoretical physicists in the 1950s. By using concepts and methods of direct and inverse scattering (see, e.g., Faddeev, 1963) GGKM found that the solution of (2) obeyed the linear Gelfand–Levitan (1951) integral equation.

Previous to KdV, Hopf (1950) and Cole (1951) showed the Burgers (1948) equation

$$u_t + uu_x = u_{xx}$$

could be reduced to the solution of the linear heat equa by using the transformation $V_x = uV$. Although (4) is, in s sense, a generalization of this, the implicit nature and "inverse" concepts are quite novel.

The work of Zakharov and Shabat (1971) showed that the methods of KdV were applicable to still another physically relevant evolution equation. Subsequently, Ablowitz, Kaup, Newell, and Segur (1974) developed a technique by which a class of nonlinear evolution equations can be isolated and solved.

It is significant that many of the equations solvable by these techniques are applicable to a wide variety of physical problems and can be derived in a rather general setting (see, e.g., Benney and Newell, 1967). The general ideas have stimulated work in a great many new directions. Numerous papers on these subjects can be found in the current literature.

See also OPTICS, NONLINEAR; WAVES.

REFERENCES

M. J. Ablowitz and D. J. Benney, *Stud. in Appl. Math.* **49**, 225 (1970). (A)

M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, *Stud. in Appl. Math.* **53**, 255 (1974). (A)

T. B. Benjamin, *Proc. Roy. Soc. London* **A299**, 59 (1967). (A)

D. J. Benney and A. C. Newell, *J. Math. Phys. (N.Y.)* **46**, 133 (1967). (A)

N. Bloembergen, *Nonlinear Optics*, Benjamin, New York, 1965. (I)

J. M. Burgers, *Adv. Appl. Mech.* **1**, 171 (1948). (A)

J. D. Cole, *Q. Appl. Math.* **9**, 225 (1951). (A)

J. D. Cole, *Perturbation Methods in Applied Mathematics*, Blaisdell, Boston, 1968. (I)

R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves*, Wiley (Interscience), New York, 1948. (I)

R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. II, Wiley (Interscience), New York, 1961. (A)

L. Faddeev, *J. Math. Phys.* **4**, 72 (1963). (A)

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

C. Gardner, J. Greene, M. D. Kruskal, and R. Miura, *Commun. Pure Appl. Math.* 27, 97 (1974). (A)
I. M. Gel'fand and B. M. Levitan, *Am. Math. Trans.* (2), 1, 253 (1951). (A)
E. Hopf, *Comm. Pure Appl. Math.* 3, 201 (1950). (A)
D. J. Korteweg and G. deVries, *Phil. Mag.* 39, 422 (1895). (A)
O. M. Phillips, *Ann. Rev. Fluid Mech.* 6, 93 (1974). (I)
B. Riemann, *Göttingen Abhandl.* 7, 43 (1858). (A)
G. G. Stokes, *Camb. Trans.* 8, 441 (1847). (A)
G. B. Whitham, *Proc. Roy. Soc. London* A283, 238 (1965a). (A)
G. B. Whitham, *J. Fluid Mech.* 22, 213 (1965b). (A)
G. B. Whitham, *Nonlinear Waves*. Wiley (Interscience), New York, 1974. (I)
N. J. Zabusky and M. D. Kruskal, *Phys. Rev. Lett.* 15, 240 (1965). (I)
V. E. Zakharov and A. B. Shabat, *Zh. Eksp. Teor. Fiz.* 61, 118 (1971) [*Soviet Phys. JETP* 34, 62 (1972)]. (A)

Nuclear Emissions *see* Angular Correlation of Nuclear Radiation

Nuclear Fission

*Harold W. Schmitt**

I. INTRODUCTION

The subject of nuclear fission, as a subfield of nuclear physics, is perhaps unique in that its importance is founded in both practical and research considerations. On the practical side, the large energy release from fission, coupled with the emission of neutrons and the consequent possibility of sustaining a controlled chain reaction, as in nuclear fission reactors, provides a source of energy applicable to the world's energy needs. In research, the phenomenon of nuclear fission is a fascinating physical process worthy of study and understanding in its own right. In addition, the increased capabilities of existing and new accelerators have opened the previously inaccessible field of heavy-ion physics; acceleration of heavy ions to energies above the Coulomb barrier permits qualitatively new studies of reactions between heavy nuclei. Importantly, heavy-ion reactions cannot be regarded as "fission in reverse"; nevertheless, many theoretical considerations are common to both subjects, and many of the experimental methods and techniques used in nuclear fission research are directly applicable in heavy-ion research.

In this short treatment, we can only describe the qualitative aspects of fission and the nature of present-day understanding of the fission process. The interested reader is referred to the appended list of selected readings for further information and study.

II. ENERGY RELEASE IN FISSION—DIVISION OF A HEAVY NUCLEUS

The division of a heavy nucleus, e.g., uranium, plutonium, etc., into two fragment nuclei is accompanied by large energy release [about 200 million electron volts (MeV)]. At one time, this large energy release, together with the emission of neutrons and gamma rays, was thought to indicate that nuclear fission was a complex, high-energy process not permitting of nuclear-structure considerations and analysis. Evidence accumulated through many years of research, however, indicates that the principal features of the fission process can reasonably be understood in phenomenological terms and that it is perhaps surprisingly amenable to analyses born in more conventional theoretical nuclear studies, e.g., shell-model and collective-model studies. Having said this, however, we must immediately point out that the nuclear many-body problem presented by the phenomenon of fission is enormously complex and remains far from being solved: even the analyses based on more conventional nuclear models are often highly complex indeed.

Consider first the nuclear fission reaction induced when a thermal neutron (0.025 eV) is incident on a ^{235}U nucleus, i.e., $^{235}U(n_{th},f)$. (This reaction has perhaps been the subject of more study than any other fission reaction.) The compound nucleus ^{236}U is formed in an excited state with excitation energy $E^* \cong 6.5$ MeV. As the compound nucleus begins to fission, this excitation energy transforms into deformation, and a sequence of shapes such as those shown in Fig. 1 may ensue. At the "scission point," i.e. the point of separation, the product nuclei (often called "fission fragments" or "fission products") are free to accelerate by mutual Coulomb repulsion. It is this Coulomb potential energy, and its subsequent transformation into fragment kinetic energy, which accounts for most (80 to 90%) of the energy released in fission. The slowing down and ionization of these high-energy fragment nuclei by atomic processes in the surrounding medium produces the heat which can then be used as a useful energy source in nuclear reactors.

The two fragment nuclei are produced in deformed excited states and decay by neutron and gamma-ray emission. An average of 2.42 neutrons is produced in the $^{235}U(n_{th},f)$ reaction; in a reactor using ^{235}U as fuel it is these neutrons

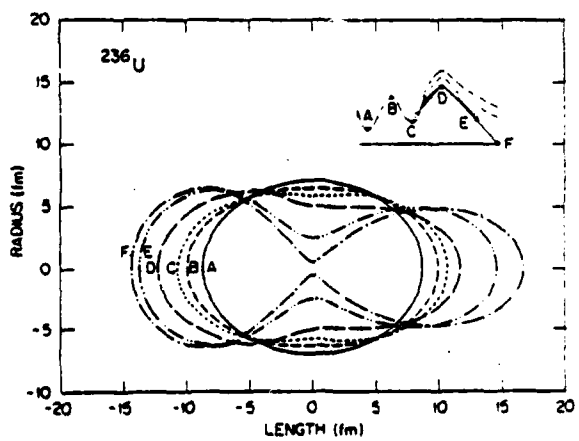


FIG. 1. A possible sequence of shapes of a fissioning nucleus. In the inset, points corresponding to these shapes are shown schematically on the minimum potential-energy curve (Section IV). [Reproduced, with permission, from the second item in Part D of the Bibliography.]

* *Editor's note:* This article was prepared by the author and accepted for publication in 1977.

Physica 3D (1981) 1 & 2, 129-141
© North-Holland Publishing Company

REMARKS ON NONLINEAR EVOLUTION EQUATIONS AND ORDINARY DIFFERENTIAL EQUATIONS OF PAINLEVÉ TYPE

Mark J. Ablowitz

Mathematics and Computer Sciences
Clarkson College
Potsdam, New York
U.S.A.

Recent work on (i) Painlevé equations and their relationship to nonlinear evolution equations solvable by inverse scattering and (ii) new results on an evolution equation describing long waves in a stratified fluid are discussed.

In these proceedings I will review some of the recent work done by Ramani, Segur and myself on the deep and important relationship between nonlinear evolution equations solvable by Inverse Scattering and ordinary differential equations (O.D.E.'s) of Painlevé type. In addition I will briefly discuss some recent work done by Satsuma, Kodama and myself on certain equations arising in the propagation of long waves in stratified fluids.

I begin with a brief discussion of O.D.E.'s and their singularities. I will classify the singularities of an O.D.E. into two main types (i) fixed singular points and (ii) movable singular points. In class (i) the location of the singularities (in the complex plane) of the equation are fixed by the equation. Specifically the location of the fixed singularities depend on the location of the singularities of the coefficients in the equation, and locations where the coefficients of the highest derivatives vanish. For example

$$z^2 w' + w = o \tag{1a}$$

has the solution

$$w = C \exp(1/z) \tag{1b}$$

where C is an arbitrary constant. Here $z = o$ is a fixed singular point (essential). Another such example is the linear, second order O.D.E.

$$w'' + p(z)\, w' + q(z) w = o \tag{2a}$$

The location of the singularities of (2a) depend only on $p(z)$, $q(z)$. Specificially the solution

$$w = C_1 w_1(z) + C_2 w_2(z) \tag{2b}$$

in such that the singularities in $w(z)$ do not depend on the constants of integration C_1, C_2. This is an example of the general statement: the location of the singularities of any linear O.D.E. are fixed. Singularities in class (ii) do not have this property. Namely the location of their singularities depend on the constants of integration. This may occur for nonlinear O.D.E.'s, and it is generic to them. So, for example the equation

$$w' + w^2 = o, \tag{3a}$$

has the solution

$$w = \frac{1}{z-z_o}, \tag{3b}$$

where z_o is a constant of integration. This singularity, a pole, is movable because its location depends on the constant of integration (i.e. initial values) z_o.

In what follows, we shall refer to a critical point as either a branch point or an essential singularity.

I shall now turn to the classical results on the question of which O.D.E.'s have no movable critical points.

The first equation (a review of this and the further work done by Painlevé and his coworkers which I will discuss below appears in Ince [1]) to be considered is the first order nonlinear O.D.E.

$$w' = F(w,z) \tag{4}$$

where F is rational in w, and locally analytic in z. It was found (Fuchs (1884) - see [1]) that the only such equation without movable critical points is the generalized Riccati Equation.

$$w' = P_o(z) + P_1(z)w + P_2(z)w^2 \tag{5}$$

Undoubtedly familiar with this result, Kovalevskya (1888) made the next significant advance. Indeed she was awarded the Bordin Prize for her major contribution to the theory of the motion of a rigid body about a fixed point. Her main idea was to carry out the apparently nonphysical calculation of determining the choices of parameters for which the equations of motion admitted no movable critical points. In all such cases she then solved the equations explicitly. In all other cases the solution is still unknown (for a discussion of this work the reader is recommended to consult [2]).

Shortly thereafter Painlevé and his coworkers examined the second order equation

$$w'' = F(w',w,z) \tag{6}$$

where F is rational in w',w and locally analytic in z. They showed that out of all the possible equations of the form (6) there are only 50 canonical equations with the property of having no movable critical points. Hereafter, we shall refer to any equation possessing this property as being of P-type (P for Painlevé). Of these equations all may be reduced to either an equation already solved or to six new transcendental functions; i.e. the six Painlevé transcendents. The first two are

$$w'' = 6w^2 + z, \tag{P_I}$$

$$w'' = zw + 2w^3 + \alpha. \tag{P_{II}}$$

The other four are listed in [1].

The question of whether an O.D.E. is of Painlevé type can be asked of an equation of any order. However, extensive results are known only for O.D.E.'s of first and second order. At higher orders it is straightforward to examine the equation for movable algebraic singularities. Essential singularities are more difficult to handle. Such a method was used by Kovalevskya [2], and Painlevé (his so-called α method [1]). For example

$$w'' = z^m w = 2w^3 \tag{7}$$

By a simple balance we see that the dominant term of an algebraic

singularity to (7) has structure $w \sim \pm 1/(z-z_o)$. Take the + root, call $\xi=z-z_o$ and look for a solution $w = 1/\xi+\rho(\xi)$. ρ Then satisfies

$$\rho''-\frac{6}{\xi^2}\rho=(z_o^m+m\, z_o^{m-1}\xi + \frac{m(m-1)}{2}\, z_o^{m-2}\xi^2+\ldots)(\frac{1}{\xi}+\rho)+\frac{6}{\xi}\rho^2 + 2\rho^3 \qquad (8)$$

Assuming $\rho(\xi)=a_o + a_1\,\xi + a_2\xi^2 + a_3\xi^3 +\ldots$, we find $a_o=0$, $a_1=-z_o^m/6$, $a_2=m\, z_o^{m-1}/4$ and at order ξ^3: $0=0\cdot a_3=\frac{1}{12}m(m-1)z_o^{m-2}$. If $m\neq 0,1$ then log terms must be inserted into the above expansion i.e.

$$\rho(\xi) = a_o + a_1\,\xi + a_2\xi^2 + (b_3\xi^3\log\xi + a_3\xi^3) +\ldots,$$

and the equation is not of P type. When $m = o,1$ we have satisfied the necessary conditions and indeed Painlevé has proven that the equations have no movable critical points. In this way we can, in principle, test an O.D.E. to see whether it satisfies necessary conditions to be of Pairlevé type. The above procedure is more extensively considered in [3].

Next we mention that there is a connection between O.D.E.'s of P-type and nonlinear evolution equations solvable by inverse scattering (I.S.T.). Namely, we have found that similarity reductions of such evolution equations are of P-type. For example the modified Korteweg deVries (mKdV) equation

$$v_t - 6v^2v_x + v_{xxx} = 0 \qquad (9)$$

has a self-similar solution of the form

$$v = w(z)/(3t)^{1/3},\quad z = x/(3t)^{1/3}.$$

Substituting this into (9) shows that $w(z)$ satisfies P_{II} (see also ref. [4]).

For our purposes we consider an evolution equation to be solvable by I.S.T. if the following holds. Consider a linear integral equation of Gel'fand-Levitan'Marchenko type:

$$K(x,y;t) = F(x,y;t) + \lambda \int_x^\infty K(x,z;t)\, N(x,z,y;t)dz. \qquad (10)$$

The cases of interest have $N(x,z,y;t)$ related to F in some definite way, e.g.

$$N = F(z,y;t)$$

$$N = \int_x^\infty F(x,s;t)\, F(s,y;t)ds$$

etc.

By requiring $F(x,y;t)$ to satisfy certain linear equations, say $L_iF=0$, $i=1,2$ we may establish equations which $K(x,y,t)$ must satisfy (see, for example [5]). Using these results $K(x,y,t)$ may be shown in every concrete example to satisfy an equation. $L_2K(x,y;t) = \hat{L}\, K(x,y;t)$ where $\hat{L}$ depends on $K(x,x;t)$. Along the line $y=x$ we have a nonlinear evolution equation. The example of mKdV is detailed in ref. [6].

With the above we have a notion of when an evolution is in the I.S.T. class. We have considered many examples of equations in the I.S.T. class and similarity reductions of such equations. By applying the expansion procedure described above we have found that each O.D.E.

satisfies necessary conditions to be of P-type. On the basis of our results we have formulated the following Conjecture [7].

Consider a nonlinear P.D.E. in the I.S.T. class. Then every O.D.E. obtained by an exact reduction of the nonlinear P.D.E. is of P-type.

Apart from mKdV, we have considered similarity reductions of KdV, certain higher order KdV equations, Boussinesq (relates to P_I), derivative nonlinear Schödinger (relates to P_{IV}), Sine Gordon (related to P_{III}- and we caution that a transformation is necessary here) etc. All of these equations have similarity reductions which are of P-type.

If an evolution equation reduces to an O.D.E. that is not of P-type then we do not expect it to be solved by I.S.T. Using the singularity analysis described above we have examined the similarity solutions of some evolution equations for which numerical or other evidence shows that solitary waves do not interact as solitons:

(a) "Cubic" nonlinear KdV: $u_t + 6u^3u_x + u_{xxx} = 0$,

where $u = w(z)/(3t)^{2/9}$, $z = x/(3t)^{1/3}$,

(b) double Sine-Gordon: $u_{xt} = \sin u + \lambda \sin 2u$,

$\lambda \neq 0$, $w(z) = \exp(iu(z))$, $z = xt$,

(c) Equation of spherical Self-focussing

$i\Psi_t + \Psi_{rr} + \frac{1}{r}\Psi_r \pm \Psi^2\Psi^* = 0$ where $\Psi = e^{i\lambda t}w(r)$

(d) Fisher's equation $u_t = u(1-u) + u_{xx}$

where $u = w(z)$, $z = x-ct$.

In all cases except (d) when $c=c_0 = \pm 5/\sqrt{6}$, or $c = 0$ we find that the resulting O.D.E. is not of P-type. We therefore conjectured that none of these evolution equations are solvable by I.S.T. In the special cases $c=c_0$, $c=0$ of (d) the equation is of Painlevé Type. Moreover the solutions can be found in closed form (apparently these are the only such solutions known [8]).

It should also be noted that in those cases where the kernel $N(x,z,y;t)$ in (10) decays rapidly enough such that $|N(x,z,y)| \leq M(x,y)| < \infty$ for all z, $\int_x^\infty M(x,y)dy = T(x) < \infty$ and $\int_x^\infty F(x,z)M(x,z)dz < \infty$

then (10) has a unique solution (Fredholm Theory) which is given by

$$K(x,z;t) = F(x,y;t) + \lambda \int_x^\infty F(x,z;t)\mathcal{R}(x,z,y;t)dx \qquad (11)$$

$$\mathcal{R}(x,z,y;t) = \frac{C(x,z,y;t)}{D(x;t)} \qquad (12)$$

Hence C and D are nontrivial entire functions of λ which are given as absolutely convergent series. In important subclasses where $F(x,y;t]$ is analytic in its arguments and rapidly decreasing then C,D are analytic in each of their arguments. Hence the only singularities of K come from the zeros of $D(x;t)$ and as a function of x can only be poles. With certain modifications, this argument can be extended to cases in which F has fixed singularities. This then shows that any solution arising from such a linear integral equation (11) cannot have movable critical points (i.e. this solution branch

has the Painlevé property - see ref [3]).

Not only does the argument following (11) work for evolution equations, it also holds directly for those O.D.E.'s obtained from an evolution equation in the I.S.T. class. For example if we take the kernel in (11): N to be:

$$N(x,z,y) = \int_x^\infty F(x,z)\, F(z,y)dz$$

and require $L_i F(x,y)=0, i=1,2$ where

$$L_1 = \partial_x - \partial_y \tag{13a}$$

$$L_2 = (\partial_x+\partial_y)^2 - \left(\frac{x+y}{2}\right) \tag{13b}$$

then a solution $F(x,y)=F(z)$, $z=x+y$ satisfies,

$$F_{zz} - \frac{z}{8} F = 0 \tag{13c}$$

with a decaying solution

$$F(z) = r_o \mathrm{Ai}\left(\frac{z}{2}\right). \tag{13d}$$

In [6], [7] it is shown by direct application of these L_i, $i=1,2$ on (11) with this kernel, that we may solve for a one parameter family of solutions of P_{II} with $\alpha=0$. Namely, solve for $K(x,y)$, $y>x$:

$$K(x,y) - r_o\, A_i\left(\frac{x+y}{2}\right) - \left(\frac{r_o}{2}\right)^2 \int_x^\infty \int_x^\infty K(x,z)\, A_i\left(\frac{z+s}{2}\right) A_i\left(\frac{s+y}{2}\right)dzds=0 \tag{14a}$$

then $w(x) = K(x,x)$ satisfies

$$w'' = xw + 2w^3 \tag{14b}$$

with $w \sim r_o\, A_i(x)$ as $x \to +\infty$.

If $|r_o| < 1$ it may be proved [6] that a solution exists for all x, whereas for $|r_o| > 1$ we have a pole for finite x. $|r_o|=1$ is the "critical" branch. In these proceedings Segur will discuss how we may use the asymptotics ($t\to\infty$) of the mKdV equation to determine the so called connection formula in the case $|r_o| < 1$ i.e.

As $x\to\infty$, for $|r_o| < 1$

$$w \sim \frac{d_o}{(-x)^{1/4}} \sin\left(\frac{2}{3}(-x)^{3/2} - \frac{3}{2} d_o^2 \log(-x) + \theta_o\right) \tag{15a}$$

where $d_o = -\frac{1}{\pi}\log(1-r_o^2)$ and

$$\theta_o = \frac{\pi}{4} - \arg\Gamma\left(1-i\frac{d_o^2}{2}\right) - \frac{3}{2} d_o^2 \log 2 - \arg r_o$$

(r_o real). If $|r_o| = 1$, then as $x \to -\infty$

$$w \sim \mathrm{sign}(r_o)\left(\frac{-x}{2}\right)^{1/2} \tag{15b}$$

and if $|r_o| > 1$

$$w \sim \mathrm{sign}\,(r_o)\,(x - x_o(r_o))^{-1} \tag{15c}$$

as $x \to x_o(r_o)$.

It should be stressed, however, that this theory goes much deeper and far beyond the case of P_{II}. It allows us to linearize and obtain global information for many nonlinear O.D.E.'s resulting from evolution equations in the I.S.T. class. Whereas certain of these O.D.E.'s are the classical Painlevé transcendents (e.g. special cases of P_{II}, P_{III}, P_{IV} etc.), others are higher order nonlinear O.D.E.'s (e.g. similarity forms of the higher KdV, mKdV flows). Moreover the deep connection between evolution equations in the I.S.T. class and O.D.E.'s of P-type give us an apriori test as to whether an evolution equation can, in fact, be solvable by I.S.T.

As a final comment about these nonlinear O.D.E.'s it should be noted that there are methods by which we can give exact representations to "perturbations" from solutions. Namely if $w(z)$ satisfies some nonlinear O.D.E. solvable in the above manner, and $w_o(z)$ represents a special solution, then $v(z)$ via

$$w(z) = w_o(z) + v(z):$$

can also be represented by a linear integral equation. For convenience let's consider the KdV equation

$$u_t - 6uu_x + u_{xxx} = o. \tag{16}$$

A similarity solution is of the form

$$u = w(z)/(3t)^{2/3}, \quad z = x/(3t)^{1/3}$$

Substituting this into (16) we have

$$w''' - 6ww' - (2w + zw') = o \tag{17}$$

(17) has a class of special solutions given by

$$w_o(z) = -2 \frac{d^2}{dz^2} \log \theta_n(z) \tag{18}$$

$$\theta_o = 1, \ \theta_1 = z, \ \theta_2 = z^3 + 4, \ \ldots$$

(see for example [9]). There is a recursion relation for the higher θ_n's:

$$\theta_{n+1}\,\theta_{n-1} - \theta_{n+1}\,\theta_{n-1} = (2n-1)\theta_n^2.$$

With a method similar to that used, (for example) in ref. [10] we may establish linear integral equations for the perturbation $v(z)$ (above) from $w_o(z)$.

In the case $w_o(x) = 0$ ($\theta_1 = 1$) we have the standard Gel'fand Levitan Marchenko equation

$$K(x,y) + r_o \, Ai\left(\frac{x+y}{2}\right) + r_o \int_x^\infty K(x,z) \, Ai\left(\frac{z+y}{2}\right) dz = o$$

$$w(x) = -2 \frac{d}{dx} K(x,x) \underset{x \to a}{\sim} -2 \, r_o \, A_i'(x), \tag{19}$$

where $w(x)$ satisfies (17) directly. For $\theta_1 = x$ we have

$$w(x) = \frac{2}{x^2} + v(x) \tag{20}$$

where $v(x)$ is obtained via:

$$K(x,y) + r_o\, G(x,y) + r_o \int_x^\infty K(x,z)\, G(z,y)\,dz = o,$$

$$G(x,y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left(1+\frac{2i}{kx}\right)\left(1+\frac{2i}{ky}\right) e^{ik(x+y)/2+ik^3/3}dk$$

$$v(x) = -2\frac{d}{dx} K(x,x) \qquad (21)$$

Hence G(x,y) is a generalization of the Airy function in (19)

$$Ai(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx+ik^3/3}dk.$$

Higher order $N(\theta_n(x))$ may be done analogously [10].

At this point I shall leave the subject of nonlinear O.D.E.'s and consider a rather different class of nonlinear evolution equations of physical interest. The particular class of equations that I have in mind arise in the context of long internal gravity waves in a stratified fluid with finite depth.

The diagram in figure 1 is useful: ρ is the density, δ = total depth of the fluid, h is the scale upon which the stratification is felt. The wavelength λ is such that $\lambda > > h$, and the waves are weakly nonlinear with amplitude a, such that $a << h$. The references [11, 12] discuss the derivation in detail.

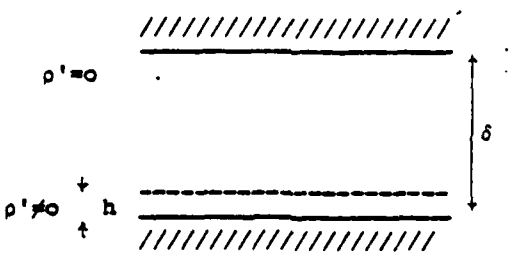


Figure 1.

The equation may be written in the form

$$u_t + 2uu_x + \frac{\partial}{\partial x}\int_{-\infty}^{\infty} K(x-\xi)\, u(\xi)d\xi = o \qquad (22a)$$

where

$$K(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} C(k)\, e^{ikx}dk, \qquad (22b)$$

$$C(k) = -k \coth k\delta + 1/\delta, \qquad (22c)$$

or in the alternative form

$$u_t + 2\, uu_x + T(u_{xx}) = \qquad (23a)$$

where

$$T[u_x] = \int_{-\infty}^{\infty} \left[\frac{1}{2\delta} \coth \frac{(x-\xi)}{2\delta} + \frac{1}{2\delta}\, \mathrm{sgn}\,(x-\xi)\right] u_\xi d\xi, \qquad (23b)$$

by taking the Fourier Transform in (22b, c). As $\delta \to o$ we have the KdV

equation

$$u_t + 2uu_x + \frac{\delta}{3} u_{xxx} = 0 \tag{24}$$

whereas if $\delta \to \infty$ we have the so called Benjamin-Ono equation:

$$u_t + 2uu_x + H(u_{xx}) = 0; \tag{25}$$

where $H(u) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(\xi)}{\xi - x} d\xi$, is the Hilbert Transform.

The interest here is that even though the equation (22) or (23) is an integro-differential equation, it still has solitons, an infinite number of conserved quantities, a Bäcklund transformation, and is related to a linear problem (i.e. a linear scattering-like equation)

Its been pointed out in ref. [13] that the equation (22) has an N soliton solution (see also [13a]). The idea is to put (22) into a differential-difference equation by formally replacing $k \to i \frac{\partial}{\partial x}$ in (22). Doing this yields (there are further conditions necessary in order to do this [16]).

$$u_t + \frac{1}{\delta} u_x + 2uu_x - i \frac{L^+(\delta)}{L^-(\delta)} u_{xx} = 0, \tag{26a}$$

where

$$L^{\pm}(\delta) = e^{i\delta\frac{\partial}{\partial x}} \pm e^{-i\delta\frac{\partial}{\partial x}} \tag{26b}$$

Using the dependent variable transformation

$$u = -i L^-(\delta) \frac{\partial}{\partial x} \log f = -i \frac{\partial}{\partial x} \log \frac{f^+}{f^-} \tag{27a}$$

where

$$f^{\pm} \equiv f(x \pm i\delta) \tag{27b}$$

we find

$$(i D_t + \frac{1}{\delta} i D_x + D_x^2) f^+ f^- = 0 \tag{28a}$$

with

$$D_x^n \, a \cdot b = (\frac{\partial}{\partial x} - \frac{\partial}{\partial x'})^n \, a(x) \, b(x') \, | \, x' = x. \tag{28b}$$

(28) is the same form as that in ref. [14]. (27) via (28) has an N soliton solution

$$f = \sum_{\mu=0,1} \exp(\sum_{i=1}^{N} \mu_i \eta_i + \sum_{1 \le i < j}^{N} \mu_i \mu_j A_{ij}) \tag{29a}$$

where

$$\exp(A_{ij}) = \frac{(k_i - k_j)^2 + (k_i \cot k_i\delta - k_j \cot k_j\delta)^2}{(k_i + k_j)^2 + (k_i \cot k_i\delta - k_j \cot k_j\delta)^2} \tag{29b}$$

A one soliton solution is given by

$$u = \frac{k \tan k\delta}{(1 + \sec(k\delta) \cosh(kx - (\frac{k}{\delta} - k^2 \cot k\delta)t]} \tag{30}$$

As $\delta \to 0$

$$u \sim \tfrac{1}{2}\delta^2 k^2 \operatorname{sech}^2 \tfrac{1}{2}(kx - \tfrac{1}{3}\delta k^3 t) \tag{31a}$$

i.e. the KdV soliton. Whereas if we take $\delta \to \infty$ with $k_1\delta = \pi - k_1/c$ ($k_1 \to 0$ such that $k_1\delta$ is finite, c finite)

$$u \sim \frac{2c}{1+c^2(x-ct)^2}, \tag{32b}$$

we find the Benjamin Ono Rational Soliton (see Case [15a] or Chen and Lee [15b], for the N soliton solution). These arguments extend to the N soliton case as well and show explicitly that the soliton (30) contains as limits both the KdV and Benjamin Ono solitons (in contradiction to what is stated in [13]). Moreover it should be noted that there is a dynamical system which describes the motion of an arbitrary number of poles for the intermediate equation (22)-(23), which reduces to the KdV and Benjamin - Ono dynamical systems (see [16]). However I will not go into that in this lecture.

Rather I wish to note that one can derive the following Bäcklund transformation (B-T) [17] with $w_x = u$:

$$(w + w')_x = \lambda + i\,T(w'-w)_x - i\delta^{-1}(w'-w) + \mu e^{i(w'-w)}, \tag{32a}$$

$$(w'-w)_t = -(\delta^{-1}+\lambda)(w'-w)_x + i(w'+w)_{xx}$$

$$-i(w'-w)_x\,T(w'-w)_x + i\delta^{-1}(w'-w)(w'-w)_x, \tag{32b}$$

for this equation. Indeed, substituting $(w+w')_x$ from (32a) into the right hand side of (32b) we have

$$\{w_t + (w_x)^2 + T(w_{xx})\} - \{w'_t + (w'_x)^2 + T(w'_{xx})\} = 0 \tag{33}$$

so that if w' satisfies (23) (with $w'_x = u'$) then so does w. These equations may be used to generate solitons from a vaccuum. Moreover the limiting forms are as follows. Calling $\lambda = -k\cot k\delta$, $\mu = k \operatorname{cosec} k\delta$, $\hat{w} = \delta^{-1}w$, $\hat{w}' = \delta^{-1}w'$ (32) reduces to

$$(\hat{w}+\hat{w}')_x = \tfrac{1}{2}k^2 - \tfrac{1}{2}(\hat{w}'-\hat{w})^2 + o(\delta) \tag{34a}$$

$$\delta^{-1}(\hat{w}'-\hat{w})_t = -\tfrac{1}{2}k^2(\hat{w}'-\hat{w})_x - \tfrac{1}{3}(\hat{w}'-\hat{w})_{xxx} + \tfrac{1}{6}((\hat{w}'-\hat{w})^3)_x + o(\delta) \tag{34b}$$

as $\delta \to 0$; i.e. the KdV B-T. Whereas if $\delta \to \infty$ we find the Benjamin Ono B-T [17, 18] and a special case of the result in [19]:

$$(w+w')_x = \lambda + iH(w'-w)_x + \mu e^{i(w'-w)} \tag{35a}$$

$$(w'-w)_t = -\lambda(w'-w)_x + i(w'+w)_{xx} - i(w'-w)_x H(w'-w)_x, \tag{35b}$$

(calling $w'-w = v$, we may rewrite (33) to obtain a generalized Miura transformation:

$$V_x + 2u = \lambda + iT(V_x) - i\delta^{-1}V + \mu e^{iV} \tag{36a}$$

$$V_t + (\delta^{-1}+\lambda)V_x - iV_{xx} - 2iu_x + iV_xT(V_x) - i\delta^{-1}VV_x = o \qquad (36b)$$

Substituting for u from (36a) into (36b) gives the following modified intermediate equation:

$$V_t + [\lambda V + T(V_x) - \mu e^{iV} - \frac{i}{2\delta}V^2]_x + iV_xT(V_x) = o \qquad (37a)$$

Since $\int_{-\infty}^{\infty} uT(u)dx = o$ (if $u \to o$ appropriately as $|x| \to \infty$) we have:

$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} V\,dx = o$ ($V \to o$ appropriately as $|x| \to \infty$). So calling

$\chi = iV + \log(-\lambda/\mu)$ gives from (36a).

$$e^{\chi} - 1 = \lambda^{-1}(-i\chi_x - T(\chi_x) + \delta^{-1}\chi + 2u). \qquad (37b)$$

For $\lambda \to \infty$, expanding $\chi = \sum_{n=1}^{\infty} \lambda^{-n}\chi_n$ and using the above argument yields an infinite number of conserved quantities χ_n. The first few are

$$\chi_1 = u,\ \chi_2 = u^2,\ \chi_3 = u^3 + \tfrac{3}{2}uTu_x,\ \chi_4 = u^4 + 3u^2Tu_x + \tfrac{1}{2}u_x^2 + \tfrac{3}{2}(Tu_x)^2 + \tfrac{3}{2}\delta^{-1}uTu_x, \ldots$$

In order to derive a linear problem we define

$$\log \psi^+/\psi^- = iV, \qquad (37c)$$

$$(\log \psi^+\psi^-)_x = -T(V_x) + \delta^{-1}V.$$

In the limit $\delta \to \infty$, for appropriate V, this amounts to splitting the function V into functions analytically extendable into the upper (-), lower (+) half plane i.e. $(\log \psi^{\pm})_x = \pm(i+H)V_x$ (note sign convention). Substituting these equations into (36) yields the linear problems:

$$\psi_x^- - i(u - \tfrac{\lambda}{2})\psi^- = -\tfrac{i}{2}\mu\psi^+ \qquad (38a)$$

$$i\psi_t^{\pm} + i(\delta^{-1}+\lambda)\psi_x^{\pm} + \psi_{xx}^{\pm} + \{\pm iu_x - T(u_x) + \delta^{-1}u + \nu\}\psi^{\pm} = o \qquad (38b)$$

(μ, λ, ν are constants). In deriving (38b) we actually use a transformed but equivalent form of (36b):

$$V_t + (\delta^{-1}+\lambda)V_x - i(V_{xx}+2u_x) + iV_x\{T(V_x) - \delta^{-1}V\}$$

$$\pm i[T(V_t) - \delta^{-1}\int_{-\infty}^{x} V_t dx + (\delta^{-1}+\lambda)\{T(V_x) - \delta^{-1}V\}$$

$$-i\{T(V_{xx} + 2u_x) - \delta^{-1}(V_x + 2u)\} - iV_x^2/2$$

$$+i\{T(V_x) - \delta^{-1}V\}^2/2 + 2i\nu] = o. \qquad (38c)$$

For small δ, introduce $V = 2\delta(\log\phi)_x$ choose the parameters λ, μ as we did above (34) and $\nu = -k^2/4$. Then taking $u = o(\delta)$,

$$\psi^{\pm} = \phi[1 \pm i\delta(\log\phi)_x - \tfrac{1}{3}\delta^2(\log\phi)_{xxx} + \ldots]$$

(from (37)), (38a), (38c) yields

$$\phi_{xx} - (k^2 4-\delta^{-1}u)\phi=o, \tag{39a}$$

$$\phi_t + \delta(k^2/4+\delta^{-1}u)\phi_x + \phi_{xxx/3} = o \tag{39b}$$

On the other hand for $\delta\to\infty$ we have

$$\Psi_x^- - i(u-\tfrac{\lambda}{2})\Psi^- = -\tfrac{1}{2} i\mu\Psi^+ \quad , \tag{40a}$$

$$i\Psi_t^\pm + i\lambda\Psi_x^\pm + \{\pm iu_x - H(u_x) + \nu\}\Psi^+ = o. \tag{40b}$$

(40a) is a differential Riemann-Hilbert Problem.

As a final remark (about a different topic), we have recently discovered evidence [20] which suggests that the "higher nonlinear" KdV equations:

$$U_t + 6U^P U_x + U_{xxx} = 0 \tag{41}$$

have a focusing singularity for $P\geq 4$, with $P=4$ being the critical value. At this time we do not have a rigorous proof, but rather have discovered certain unusual behavior associated with the evolution of the solitary wave under weak dissipative perturbations when $P\geq 4$.

Acknowledgements This work was partially supported by the office of Naval Research and A.F.O.S.R. - Mathematics divisions.

References

[1] E.L. Ince: Ordinary differential equations (New York, N.Y. 1947).

[2] V.V. Golubev, Lectures on the integration of the equations of motion of a rigid body about a fixed point, (State Publication House, Moscow, 1953).

[3] M.J. Ablowitz, A. Ramani, H. Segur, A connection between nonlinear evolution equation and ordinary differential equations of P-type I. To be published in Journal of Mathematical and Physics.

[4] M.J. Ablowitz and H. Segur, Physical Review Lett., 38, 103 (1977).

[5] V.E. Zakharov and A.B. Shabat, Funkt. Anal. Pulozh., 8, 43 (1974) (Funct. Anal. Appl., 8, 226 (1974)).

[6] M.J. Ablowitz, A. Ramani, H. Segur, A connection between nonlinear evolution equations and ordinary differential equations of P-type II. To be published in Journal of Mathematical Physics.

[7] M.J. Ablowitz, A. Ramani, and H. Segur, Lett. Nuovo Cimento 23, 333 (1978).

[8] M.J. Ablowitz and T. Zeppetella, Explicit solutions of Fisher's equation for a special wave speed, To be published in the Bulletin of Mathematics and Biology.

[9] M.J. Ablowitz and J. Satsuma, Journal of Mathematics and Physics. 19, p. 2180, (1978).

[10] M. Airault and M.J. Ablowitz, Finite perturbations and solutions of the Korteweg deVries equation; Preprint.

[11] R.I. Joseph, J. Phys. A; Math. Gen 10 (1977) L225.

[12] T. Kubota, D.R.S. Ko and L. Dobbs, Report of AIAA 16th Aerospace sciences meeting, Huntsville, Alabama, USA (1978).

[13] H.H. Chen and Y.C. Lee Phys. Rev. Lett. 43, p. 264, (1979)

[13a] R.I. Joseph and R. Egri, Journal of Phys. A: Math 11, No. 5, L97, (1978).

[14] J. Satsuma and Y. Ishimori, J. Phys. Soc. Japan 46 (1979) 681.

[15] K.M. Case, Proc. Nat'l Acad. Sci. USA, 75, p. 3562, (1978).

[15b] J. Satsuma, H.H. Chen and Y.C. Lee, Phys. Fluids 22, 187 (1979).

[16] J. Satsuma and M.J. Ablowitz, Solutions of an internal wave equation describing a stratified fluid with finite depth. To be publ. Proc. Conf. on Nonlinear Partial Differential Equations, Univ. of Rhode Island, June, 1979, Preprint.

[17] J. Satsuma, M.J. Ablowitz, and Y. Kodama, Phys. Lett., 73A,

References (2)

No. 4, 283, (1979).

[18] A. Nakamura, Bäcklund Transform and Conservation lwas of the Benjamin-Ono Equation, Journal of the Physical Society of Japan, Vol. 47, No. 4, p. 1335-1340, October, 1979.

[19] T.L. Boch and M.D. Kruskal, A two parameter Miura Transformation of the Benjamin-Ono Equation, Physics Letters A, Vol. 74A Number 3, 4, November 12, 1979, p. 173-176.

[20] Y. Kodama and M.J. Ablowitz, Perturbations of solitons and solitary waves, to be publ. Stud. Appl. Math.

FONCTIONS SPÉCIALES. – *Perturbations finies et forme particulière de certaines solutions de l'équation de Korteweg de Vries.* Note (*) de **Mark J. Ablowitz** et **Hélène Airault**, transmise par Paul Malliavin.

On construit une famille de solutions de l'équation de Korteweg de Vries. Ces solutions sont obtenues en perturbant les potentiels $d(d+1)/x^2$, (d est un entier positif). Cette sorte de solutions a un analogue dans le cas périodique.

We construct a family of solutions of the Korteweg de Vries equation. These solutions are obtained with a perturbation of the potentials $d(d+1)/x^2$ (d is a positive integer). This kind of solution has an analogue in the periodic case.

I. On vérifie facilement que le potentiel $u(x, t)=2/x^2$ est une solution de l'équation (de Korteweg de Vries) :

$$u_t = 6uu_x - u_{xxx}. \tag{1}$$

Dans [1], on a construit une famille de solutions de l'équation (1). Ces solutions décroissent comme $2/x^2$ quand $|x|$ tend vers l'infini. Pour construire ces solutions, on a utilisé la méthode de perturbation de Schabat [2], dont on a donné une dérivation directe.

Dans ce qui suit, on construit pour chaque entier d, une famille de solutions de l'équation (1). Ces solutions décroissent comme $d(d+1)/x^2$ quand $|x|$ tend vers l'infini. Lorsque $d=1$, on retrouve les potentiels de [1]. Si d n'est pas égal à 1, le potentiel $d(d+1)/x^2$ n'est pas une solution de (1) et une perturbation directe ne permet pas d'engendrer de nouvelles solutions de (1). Pour obtenir celles-ci, on perturbe [2] la solution rationnelle connue ([3], [4], [5]) de l'équation (1) :

$$q_d(x, t) = -2\frac{d^2}{dx^2}\log \theta_d(x+\tau_1, \tau_2, \tau_3, \ldots, \tau_d), \tag{2}$$

où $\tau_2 = 12t$; $\tau_i (i=1, \ldots, d,\ i\neq 2)$ sont des constantes arbitraires et où :

$$\theta_d(x+\tau_1, \tau_2, \ldots, \tau_d)$$

est un polynôme de degré $n_d = d(d+1)/2$ en la variable x.

On rappelle que les polynomes θ_d sont définis par la formule de récurrence [4] :

$$\theta'_{d+1}\theta_{d-1} - \theta_{d+1}\theta'_{d-1} = (2d+1)\theta_d^2, \tag{3}$$

avec $\theta_0 = 1$, $\theta_1 = x$ (la prime désigne la dérivée par rapport à la variable x).

Ici, on fixe la constante d'intégration à l'étape $(d+1)$ par la condition que le coefficient de x^{n_d-1} dans θ_{d+1} soit égal à :

$$1^2.3^2.5^2\ldots(2d-1)^2(-1)^{d+1}\tau_{d+1}.$$

Par exemple, on a :

$$\theta_2 = x^3 + \tau_2,$$

$$\theta_3 = x^6 + 5\tau_2 x^3 - 9\tau_3 x - 5\tau_2^2,$$

$$\theta_4 = x^{10} + 15\tau_2 x^7 - 63\tau_3 x^5 + 315\tau_2\tau_3 x^2 + 175\tau_2^3 x - 189\tau_3^2 + 225\tau_4(x^3+\tau_2).$$

En particulier, pour chaque entier d, on a obtenu les solutions :

$$u(x, \tau_2) = -2\frac{d^2}{dx^2}\log(\theta_d(x+\tau_1, \tau_2, \ldots, \tau_d)\det A), \tag{4}$$

où A est la matrice :

$$(5)\quad A_{ij}=\delta_{ij}-C_iC_j.\frac{e^{-4(k_i^3+k_j^3)t}}{k_i+k_j}\times e^{(k_i+k_j)x}$$
$$\times\frac{\theta_d(x+\tau_1-(1/k_i)-(1/k_j),\ \tau_2+(1/k_i^3)+(1/k_j^3),\ \ldots,\ \tau_d+(1/k_i^{2d-1})+(1/k_j^{2d-1}))}{\theta_d(x+\tau_1,\ \tau_2,\ \ldots,\ \tau_d)}.$$

On remarque que ces quasi-solitons solutions (4)-(5) sont des dégénerescences de combinaisons de solutions périodiques et exponentielles.

On s'attend à ce que les formules (4)-(5) aient un analogue quand on perturbe la solution [6] :

$$q=-2\frac{d^2}{dx^2}\log\Theta$$

de l'équation de Korteweg de Vries, où Θ désigne la fonction thêta de Riemann. Le cas où $q=2\mathcal{P}(x)$, où $\mathcal{P}$ est la fonction elliptique de Weierstrass a été étudié dans [7].

On a une démonstration de la représentation (5) pour les solutions dans les cas $d=1, 2, 3, 4$ seulement, mais on s'attend à ce que la représentation (5) pour les solutions, existe pour tout entier d.

II. Les solutions (4)-(5) sont obtenues de la manière suivante : Soit $\psi(x, k_i)(i=1, \ldots, n)$ une solution de :

$$(6)\qquad \frac{d^2}{dx^2}\psi=(k_i^2+q_d)\psi$$

telle que :

$$(7)\qquad \frac{d}{dt}\psi=(-4k_i^2+2q_d)\psi'-q'\psi+\lambda\psi$$

(λ est une constante indépendante de x, mais peut dépendre de t). Soit A la matrice :

$$(8)\qquad A_{ij}=\delta_{ij}+C_iC_j\int_x^\infty\psi(s, k_i)\psi(s, k_j)\,ds,$$

alors :

$$(9)\qquad u_d(x, t)=q_d(x, t)-2\frac{d^2}{dx^2}\log(\det A)$$

est solution de l'équation (1) (*voir* [2]).

Dans [8], on a démontré que dans les cas $d=1, 2, 3, 4$, les fonctions :

$$(10)\qquad \psi_d(k, x)=\frac{\theta_d(x-(1/k),\ \tau_2+(1/k^3),\ \ldots,\ \tau_d+(1/k^{2d-1}))}{\theta_d(x,\ \tau_2,\ \ldots,\ \tau_d)}e^{kx}$$

et $\psi_d(-k, x)$ forment une base de solutions de (6).

Une autre représentation des solutions de (6) a été étudiée dans [4]. De plus, on a :

$$(11)\quad \int\psi_d(k_i, x)\psi_d(k_j, x)\,dx=e^{(k_i+k_j)x}$$
$$\times\frac{\theta_d(x-(1/k_i)-(1/k_j),\ \tau_2+(1/k_i^3)+(1/k_j^3),\ \ldots,\ \tau_d+(1/k_i^{2d-1})+(1/k_j^{2d-1}))}{(k_i+k_j)\theta_d(x,\ \tau_2,\ \tau_3,\ \ldots,\ \tau_d)},$$

$$(12)\qquad \int\psi_d^2(k, x)\,dx=e^{2kx}\times\frac{\theta_d(x-(2/k),\ \tau_2+(2/k^3),\ \ldots,\ \tau_d+(2/k^{2d-1}))}{2k\,\theta_d(x,\ \tau_2,\ \tau_3,\ \ldots,\ \tau_d)}.$$

La vérification de (10), (11), (12) dans les cas $d=1$, $d=2$ est triviale.

Dans le cas de potentiels périodiques, des formules analogues à (11) et (12) existent, *voir*, par exemple [7] pour le cas où le potentiel est $q=2\mathscr{P}(x)$.

III. Il est bien connu que les formules (8)-(9) peuvent être obtenues soit par la méthode de perturbation de Schabat, soit en effectuant sur le potentiel q_d des transformations de Bäcklund-Darboux successives (*voir* par exemple [9]). La méthode de la transformation de Bäcklund-Darboux [10] est la suivante : Supposons que ψ soit une solution de (6) et (7) et que le potentiel q satisfasse (1), alors le nouveau potentiel :

$Q=q-2(\psi'/\psi)'$ est aussi une solution de (1).

Dans le cas de potentiels considérés ci-dessus, on peut obtenir les formules (4)-(5) avec un nombre réduit de transformations de Bäcklund-Darboux. En fait, certaines des transformations se reduisent à des translations. Ceci est dû aux propriétés d'addition que possèdent les polynômes θ_d (*voir* [8]).

Remarquons qu'en perturbant le potentiel q_d [formule (2)] par la méthode de Schabat, on construit une classe plus large de potentiels que celle donnée par les équations (4)-(5). Tous les potentiels ainsi construits décroissent comme $d(d+1)/x^2$, lorsque $|x|$ tend vers l'infini (*voir* [8]).

(*) Remise le 15 septembre 1980, acceptée le 13 octobre 1980, revenue à l'imprimerie le 26 janvier 1981.

[1] M. J. Ablowitz et H. Cornille, *Physics Letters*, 72 A, n° 2, juin 1979.

[2] A. B. Schabat, *Soviet Math. Dokl.*, 14, n° 4, 1973.

[3] H. Airault, H. P. McKean et J. Moser, *Comm Pure and Appl. Math.*, XXX, 1977, p. 95-148.

[4] M. Adler et J. Moser, *Comm. in Math. Physics*, 61, n° 1, 1978.

[5] M. J. Ablowitz et J. Satsuma, *J. Math. Phys.*, 19, (10), octobre 1978.

[6] M. P. McKean et E. Trubowitz, *Bull. Amer. Math. Soc.*, 84, n° 6, novembre 1978.

[7] E. A. Kuznetsov et A. V. Mikhailov, *Zh. Eksp. Teor Fiz.*, 67, novembre 1974, p. 1717-1727.

[8] H. Airault et M. J. Ablowitz, Preprint, Clarkson, 1979 (à paraitre).

[9] H. Flaschka et D. W. McLaughlin, *Lecture Notes in Math.*, Springer, 515, 1974, R. M. Miura, éd., Springer, N.Y., 1976.

[10] G. Darboux, *Leçons sur la théorie générale des surfaces*, II, livre IV, chap. IX, p. 210 ou *Comptes rendus*, 94, 1882, p. 1456.

M. J. A. : *Clarkson College of Technology, Mathematics Department, Potsdam, N.Y, 13676, U.S.A.*

H. A. : *Mathématiques, Paris-VI*, Tour 45, 4, *place Jussieu* 75 *Paris* 5e.

PHYSICAL REVIEW
LETTERS

VOLUME 46 | 16 MARCH 1981 | NUMBER 11

Nonlinear Intermediate Long-Wave Equation: Analysis and Method of Solution

Yuji Kodama,[a] Junkichi Satsuma,[b] and Mark J. Ablowitz
Department of Mathematics, Clarkson College of Technology, Potsdam, New York 13676
(Received 23 June 1980)

A physically interesting nonlinear singular integro-differential equation which is an intermediary between the Korteweg–deVries and Benjamin-Ono equations is considered via the inverse-scattering transform. Novel aspects of the theory and limits to the Benjamin-Ono equation are discussed.

PACS numbers: 93.40.Kf, 02.30.+g

Recent studies have shown that the equation

$$u_t + \delta^{-1}u_x + 2uu_x + T(u_{xx}) = 0, \tag{1}$$

where $(Tu)(x) = \int_{-\infty}^{\infty} dy\, u(y) \coth[\pi(y-x)/2\delta]/2\delta$ ($\int_{-\infty}^{\infty}$ represents the principal-value integral), is of mathematical and physical interest. Physically it represents long waves in a stratified fluid of finite depth characterized by the parameter δ.[1,2] Depending on δ we get the Korteweg–de Vries (KdV) equation as $\delta \to 0$ (shallow-water limit),

$$u_t + 2uu_x + (\delta/3)u_{xxx} = 0, \tag{2}$$

and the Benjamin-Ono (BO) equation as $\delta \to \infty$ (deep-water limit),

$$u_t + 2uu_x + H(u_{xx}) = 0, \tag{3}$$

where $(Hu)(x) = \pi^{-1}\int_{-\infty}^{\infty} dy\, u(y)/(y-x)$ (Hilbert transform). Hence Eq. (1) is an intermediary equation between these two very interesting nonlinear evolution equations. Hereafter we shall refer to (1) as the intermediate long-wave (ILW) equation. Mathematically speaking, Eq. (1) has soliton solution,[3,4] a Bäcklund transformation, and a novel type of linear scattering problem.[5,6] In this Letter we shall do the following:

(a) Relate Eq. (1) directly to a linear Gel'fand-Levitan integral equation which has N-soliton solutions.

(b) Discuss how to deal with this new scattering problem. We show in what sense the above Gel'fand-Levitan equation can be derived from analytical considerations of suitable scattering data.

(c) We shall also briefly discuss the limiting case of the BO equation for which there has also been considerable study (see, for example, Refs. 7–12) regarding solitons, Bäcklund transformations, and linear scattering problems.

We begin with point (a). The operator T defined below Eq. (1) immediately suggests a splitting of the function $u(x)$ into appropriate analytic functions. Namely, if we call $U(z) = (Tu)(z)$, $\mathrm{Im}z \neq 0$, then the boundary values on $z = x$, x real, satisfy $U^{\pm}(x) = (Tu)(x) \pm iu(x)$. Here $U^{\pm}(x)$ are the boundary values of functions analytic in the horizontal strip between $\mathrm{Im}z = 0$ and $\mathrm{Im}z = \pm 2\delta$, and are periodically extended vertically. Moreover, periodicity requires that $U^{-}(x) = U^{+}(x + 2i\delta)$. It is convenient to define $g(x) = -U^{+}(x + i\delta)/2$ [here $g(z)$ is analytic in the strip $-\delta < \mathrm{Im}z < \delta$], whereupon the splitting takes the form $u(x) = i[g(x - i\delta) - g(x + i\delta)]$, $(Tu)(x) = -[g(x - i\delta) + g(x + i\delta)]$. Hence

© 1981 The American Physical Society

(1) takes the form

$$i(g^+ - g^-)_t + i\delta^{-1}(g^+ - g^-)_x - 2(g^+ - g^-)(g^+ - g^-)_x - (g_{xx}{}^+ + g_{xx}{}^-) = 0, \tag{4}$$

where $g^\pm(x) \equiv g(x \mp i\delta) = -U^\pm(x)/2$.

Consider the following linear Gel'fand-Levitan integral equation:

$$K(x,y) + F(x,y) + \int_x^\infty K(x,s)F(s,y)\,ds = 0, \quad \text{for } y > x. \tag{5}$$

Following the basic idea of Zakharov and Shabat,[13] we introduce linear operators on F, such that

$$L_1 F = (i\partial_x + \tfrac{1}{2}\delta^{-1})F^+(x,y) + (i\partial_y - \tfrac{1}{2}\delta^{-1})F^-(x,y) = 0, \tag{6a}$$

$$L_2 F = (i\partial_t + \partial_x{}^2 - \partial_y{}^2)F(x,y) = 0, \tag{6b}$$

where $F^\pm = F(x \mp i\delta, y \mp i\delta)$. Then, direct calculation shows that $K(x,y)$ must satisfy

$$[i\partial_x + \tfrac{1}{2}\delta^{-1} + iK^+(x,x) - iK^-(x,x)]K^+(x,y) + (i\partial_y - \tfrac{1}{2}\delta^{-1})K^-(x,y) = 0, \tag{7a}$$

$$[i\partial_t + \partial_x{}^2 - \partial_y{}^2 + 2\partial_x K(x,x)]K(x,y) = 0. \tag{7b}$$

Compatibility between Eqs. (7a) and (7b) gives us Eq. (4) with $g^\pm(x) = K^\pm(x,x) \equiv K(x \mp i\delta, x \mp i\delta)$. The N-soliton solutions to Eq. (1) can now be readily constructed (in the usual manner) by assuming exponential solutions for F; i.e., $F(x,y) = \sum_{i=1}^N C_i(t)\exp(i\zeta_{-i}x + i\zeta_{+i}y)$, where $\zeta_{\pm i} = i\kappa_i \pm [\kappa_i \cot(2\kappa_i\delta) - \frac{1}{2}\delta^{-1}]$, $\kappa_i > 0$ and $C_i(t) = C_i(0)\exp[-4\kappa_i(\kappa_i \cot 2\kappa_i\delta - \frac{1}{2}\delta^{-1})t]$. A one-soliton solution is given by $u = 2\kappa_1 \sin(2\kappa_1\delta)/\{\cos(2\kappa_1\delta) + \cosh[2\kappa_1(x - x_0(t))]\}$, where $x_0(t) = (2\kappa_1)^{-1}\ln[C_1(t)/2\kappa_1]$.

We now pass to point (b). As discussed in Ref. 6, the linear scattering problem obeys (with some changes in notation)

$$i\phi_x{}^+ + (\mu - \lambda)\phi^+ = \mu\phi^-, \tag{8a}$$

$$i\phi_t{}^\pm + 2i(\lambda + \tfrac{1}{2}\delta^{-1})\phi_x{}^\pm + \phi_{xx}{}^\pm + [\mp iu_x - T(u_x) + \nu]\phi^\pm = 0, \tag{8b}$$

where $\lambda = k\coth(2k\delta)$, $\mu = k\,\mathrm{csch}(2k\delta)$, $\nu = k^2 - 2k(\lambda + \frac{1}{2}\delta^{-1})$, and here $\phi^\pm(x)$ represent the boundary values of functions analytic in the horizontal strips between $\mathrm{Im}\,z = 0$ and $\mathrm{Im}\,z = \pm 2\delta$, and periodically extended. As mentioned earlier this implies $\phi^-(x) = \phi^+(x + 2i\delta)$. We note that this condition immediately leads to $T(\phi^+ - \phi^-) = i(\phi^+ + \phi^-)$ which is required in Ref. 6, and that compatibility of Eqs. (8a) and (8b) yields Eq. (1). In order to analyze the scattering problem, it is convenient to define a new function, $W^+(x,k) = \phi^+(x,k)\exp[ik(x - i\delta)]$, whereupon the scattering problem becomes

$$\mathfrak{L}W \equiv iW_x{}^+ + (\zeta_+ + \tfrac{1}{2}\delta^{-1})(W^+ - W^-) = -uW^+ \tag{9}$$

with $W^-(x) = W^+(x + 2i\delta)$, $\zeta_+(k) = k + [k\coth(2k\delta) - \frac{1}{2}\delta^{-1}]$ (We shall need the definition of ζ_- subsequently.) Now we define specific Jost functions for real k: $M^+(x;k) \to 1$, as $x \to -\infty$, and $N^+(x;k) \to \exp[2ik(x - i\delta)]$, $\bar{N}^+(x;k) \to 1$, as $x \to +\infty$. Each of these functions can be shown to satisfy an integral equation. For this purpose, we introduce the notion of a Green function satisfying $\mathfrak{L}G(x,y;k) = -\delta(x - y)$ [$\mathfrak{L}$ defined by Eq. (9).] Then

$$M^+(x;k) = 1 + \int_{-\infty}^\infty G_1{}^+(x,y;k)u(y)M^+(y;k)\,dy, \tag{10a}$$

$$N^+(x;k) = \exp[2ik(x - i\delta)] + \int_{-\infty}^\infty G_2{}^+(x,y;k)u(y)N^+(y;k)\,dy, \tag{10b}$$

$$\bar{N}^+(x;k) = 1 + \int_{-\infty}^\infty G_2{}^+(x,y;k)u(y)\bar{N}^+(y;k)\,dy, \tag{10c}$$

where

$$G_{1,2}{}^+(x,y;k) = \tfrac{1}{2}\pi\int_{C_{1,2}} \hat{G}^+(p;k)\exp[ip(x - y)]\,dp, \tag{11}$$

and $\hat{G}^+(p;k) = \{p - (\zeta_+ + \frac{1}{2}\delta^{-1})[1 - \exp(-2i\delta p)]\}^{-1}$. $\hat{G}^+(p;k)$ has poles at $p_0 = 0$, $p = 2\zeta_+{}^{-1}(\zeta_+(k))$. We note $\zeta_+{}^{-1}(\cdots)$ is a multivalued function and we have an infinite number of poles p for which we shall define $p_{-1} = 2k$ and $p_n, \bar{p}_n$ ($n \geq 1$) such that $(2n - 1)\pi/2\delta < \mathrm{Im}\,p_n < (2n + 3)\pi/2\delta$ and similarly for $-\mathrm{Im}\,\bar{p}_n$. Moreover double poles occur at special values of $\zeta_+(k)$ satisfying $k = 0$ and $p = \zeta_+(k)$. We call these values $\{\zeta_+^{(i)}, \bar{\zeta}_+^{(i)}\}_{i=1}^m$ [$\mathrm{Im}\,\zeta_+^{(i)} > 0$, $\mathrm{Im}\,\bar{\zeta}_+^{(i)} < 0$]. The contours C_1, C_2 are taken to be the lines $\mathrm{Re}\,p - i0$ and $\mathrm{Re}\,p + i0$, respectively (this is necessary in order to preserve the bound-

ary conditions). It is important to remark that Eqs. (10a)–(10c) are Fredholm-type integral equations, unlike the usual case of the Schrödinger equation where the Jost functions satisfy Volterra equations. In addition we note that, by using residue calculus, Eqs. (10a)–(10c) can be represented in an explicit manner useful for the proof of existence and analyticity of the solution (convergence of Neumann series). From Eqs. (10a)–(10c) one can establish the following [assuming $u(x)$ decays rapidly as $|x| \to \infty$]:

(i) M^+, N^+, and $\bar{N}^+$ have convergent Neumann series in certain regions of ζ_+ plane for given δ and $\max|u|$ chosen small enough.

(ii) In the ζ_+ plane, p_n $(n \geq 1)$ has a logarithmic branch point at $\zeta_+ = -1/2\delta$ and square-root branch points at $\zeta_+^{(n)}$ and $\zeta_+^{(n+1)}$.

(iii) Despite (ii), $M^+(x;k)$ and $\bar{N}^+(x;k)$ are analytic in the upper and lower half ζ_+ plane, respectively, whenever the Neumann series converges in this region. Moreover, as $|\zeta_+| \to \infty$, $M^+, \bar{N}^+ \to 1 + O(1/\zeta_+)$. Here we note that k is a multivalued function of ζ_+, and we are required to define an appropriate branch in k plane. For the functions M^+ and $\bar{N}^+$, our principal branch is that one containing the real k axis, and which has $\operatorname{Im} k \gtrless 0$ corresponding to $\operatorname{Im} \zeta_+ \gtrless 0$. There is a branch point at $\zeta_+ = -1/2\delta$ and a branch cut from $\zeta_+ = -1/2\delta$ to $\zeta_+ = -\infty$.

By virtue of the fact $G_1^+(x, y; k) - G_2^+(x, y; k) = (2i\delta\zeta_+)^{-1} - \exp[2ik(x-y)]/2i\delta\zeta_+$, we have a relation among $M^+(x;k)$, $N^+(x;k)$, and $\bar{N}^+(x;k)$ for real k [i.e., $\zeta_+ > -1/(2\delta)$],

$$M^+(x;k) = a(k)\bar{N}^+(x;k) + b(k)N^+(x;k), \tag{12}$$

where $a(k) = 1 + [\int_{-\infty}^{\infty} u(y)M^+(y;k)dy]/2i\delta\zeta_+$, $b(k) = -\{\int_{-\infty}^{\infty} u(y)M^+(y;k)\exp[-2ik(y-i\delta)]dy\}/2i\delta\zeta_+$. Hence $a(k)$ takes on the same analyticity as $M^+(x;k)$, and as $|\zeta_+| \to \infty$, $a(k) \to 1$. On the other hand, for $\zeta_+ + i0$ with $\zeta_+ < -1/2\delta$ and real (i.e., k is in the upper half plane at the edge of the principal branch), we have a relation $G_1^+(x, y; k) - G_2^+(x, y; k^*) = 1/2i\delta\zeta_+$ [note $\zeta_+(k+i0) = \zeta_+(k^*-i0)$, k^* complex conjugate of k], which yields

$$M^+(x;k) = a(k)\bar{N}^+(x;k^*). \tag{13}$$

The bound states [as $x \to \infty$, $M^+(x;k) \to 0$] are defined by $a(k_l) = 0$, $M^+(x;k_l) = b_l N^+(x;k_l)$, for $\operatorname{Im} k_l > 0$ $(l = 1, 2, \ldots, N)$. The scattering data are now given by $S = \{a(k), b(k), \{k_l, b_l\}_{l=1}^N\}$. We have found that $a(k)$ has only simple zeros and they lie on the imaginary k axis, i.e., $k_l = i\kappa_l$. From Eqs. (10b–10c) and consistent with our analyticity requirements, we assume, for $\bar{N}^+$ and N^+, the triangular representations

$$\bar{N}^+(x;k) = 1 + \int_x^{\infty} ds\, K^+(x,s)\exp[i\zeta_+(x-s)], \quad \text{for } \operatorname{Im}\zeta_+ < 0, \tag{14a}$$

$$N^+(x;k) = \exp[2ik(x-i\delta)] + \int_x^{\infty} ds K^+(x,s)\exp[i\zeta_+(x-s) + 2ik(s-i\delta)], \tag{14b}$$

where $K^+(x,s)$ satisfies Eq. (7) and $K^+(x,s) \to 0$ as $s \to \infty$. Inverse scattering formulas are obtained as follows: Divide Eq. (12) and Eq. (13) by $a(k)$ and operate with $(1/2\pi)\int_{-\infty}^{\infty} d\zeta_+ \exp[i\zeta_+(y-x)]$ (i.e. Fourier transform) for $y > x$. Then using Eq. (14), we obtain the linear Gel'fand-Levitan integral Eq. (5) with

$$F(x,y) = \tfrac{1}{2}\pi \int_{-1/2\delta}^{\infty} d\zeta_+ \frac{b(k)}{a(k)} \exp(i\zeta_- x + i\zeta_+ y) + \sum_{l=1}^{N} C_l \exp(i\zeta_{-l}x + i\zeta_{+l}y), \tag{15}$$

where $C_l = -ib_l/\dot{a}_l$ and $\dot{a}_l = [\partial a/\partial\zeta_+]_{\zeta_+ = \zeta_{+l}}$. From Eq. (8b), the time dependence of the scattering data is given by $a(k,t) = a(k,0)$, $b(k,t) = b(k,0)\exp[-4 \times ik(\lambda + \frac{1}{2}\delta^{-1})t]$, for real k, $b_l(t) = b_l(0)\exp[4\kappa_l(\lambda_l + \frac{1}{2}\delta^{-1})t]$. We expect the Gel'fand-Levitan equation is valid when the Neumann series expansions of Eq. (10) converge. For fixed $\max|u(x,0)|$, when $\delta \to \infty$ (the BO limit), this will not hold and new singularities due to the Fredholm nature of Eq. (10) may have to be taken into account. We briefly mention this later.

We now pass on to point (c). Our basic philosophy regarding the BO equation is to obtain information by taking the limit process $\delta \to \infty$. First of all if we simply take $\delta \to \infty$, then for real k, $\zeta_+ = 2k\theta(k)$ and $\hat{G}^+(p;k) = (p - 2k)^{-1}\theta(p)$, where $\hat{G}^+(p;k)$ is defined below Eq. (11) and $\theta(\cdots)$ is the usual Heaviside step function. Similarly from the fact that $G^-(x,y;k) = G^+(x+2i\delta, y; k)$ we have $\hat{G}^-(p;k) = \hat{G}^+(p;k)\exp(-2\delta p) - \theta(-p)/2k$. These formulas suggest a natural splitting of the δ function and the Green function. Hence from these results, we may deduce the split equations for the eigenfunctions,

$$iM_x^+(x;k) + 2k[M^+(x;k) - 1] = \hat{P}^+(uM^+)(x;k), \tag{16a}$$

$$M^-(x;k) = 1 + (2k)^{-1}\hat{P}^-(uM^+)(x;k), \tag{16b}$$

where $\hat{P}^{\pm} = \frac{1}{2}(1 \mp i\hat{H})$ are the usual projection operators. It is also worthwhile noting that the eigenvalue problem with $k < 0$ $(\delta \to \infty)$ in Eq. (16) corresponds to what happens to $\zeta_- < -1/2\delta$ for finite

δ. Moreover we have found a solution of the *homogeneous* equation for $N^+(x;k)$ with some $k<0$ (this seems to be related to BO solitons). In this regard, we note that one can actually compute certain eigenfunctions of the scattering problem for the BO equation. We shall use the scattering problem [from Eq. (9) with $\delta\to\infty$]:

$$iW_x^+ + (u+2k)W^+ = 2kW^-. \tag{17}$$

Let us consider $u(x) = 2\nu/(x^2+1) = i\nu[1/(x+i) - 1/(x-i)]$ as an example. There is a natural way to split the eigenvalue problem Eq. (17). Namely multiply by $(x-i)$ and require both sides to be an entire function. For the case of bound states (solitons) we take $2k(x-i)W^- = 1$. Then the solution for W^+ can be found to be

$$W^+ = -i\left(\frac{x-i}{x+i}\right)^{\nu}\int_{-\infty}^{x}\left(\frac{y+i}{y-i}\right)^{\nu}\frac{\exp[2ik(x-y)]}{y-i}\,dy, \tag{18}$$

requiring $W^+\to 0$ as $|x|\to\infty$. This implies immediately that

$$D_\nu(k) = \int_{-\infty}^{\infty}\left(\frac{y+i}{y-i}\right)^{\nu}\frac{\exp(-2iky)}{y-i}\,dy = 0 \tag{19}$$

is the condition which determines the discrete eigenvalues. For $\nu = n$ = integer, Eq. (19) is the Laguerre polynomial of degree n, i.e., $D_n(k) = L_n(-4k) = 0$, for $k<0$. Hence, for $\nu = n$, there are n-real distinct eigenvalues, e.g., for $n=1$, $k_1 = -1/4$; for $n=2$, $k_{1,2} = -(2\pm\sqrt{2})/4$, etc. Moreover, this condition corresponds to the requirement that W^+ is, in fact, analytic in the upper half plane. Thus we expect to find n solitons when $\nu = n$ (in agreement with Ref. 14). The situation with $\nu \neq$ integer is more difficult. Nevertheless we found that Eq. (18) has n eigenvalues for ν in the range $n-1<\nu\leq n$ ($n=1, 2, \ldots$). We also remark that when $\nu = 1$ the eigenfunction $W^+(x;k)$ satisfies the homogeneous equation of Eq. (10b) with G_2^+ given by $G_2^+(x,y;k) = (2\pi)^{-1}\int_0^\infty dp(p-2k)^{-1}\exp[ip(x-y)]$ for $k = k_1 < 0$.

This work was supported by the U. S. Office of Naval Research, Mathematics Division under Grant No. 0014-76-C-0867, and the Air Force Systems Command, U. S. Air Force under Grant No. AFOSR-78-3674.

---

[a]Present address: Bell Laboratories, 1E-347, 600 Mountain Ave., Murray Hill, N. J. 07974.

[b]Permanent address: Department of Applied Mathematics and Physics, Faculty of Engineering, Kyoto University, Kyoto 606, Japan.

[1]R. I. Joseph, J. Phys. A 10, L225 (1977).

[2]T. Kubota, D. R. S. Ko, and D. Dobbs, J. Hydronaut. 12, 157 (1978).

[3]R. I. Joseph and R. Egri, J. Phys. A 11, L97 (1978).

[4]H. H. Chen and Y. C. Lee, Phys. Rev. Lett. 43, 264 (1979).

[5]J. Satsuma and M. J. Ablowitz, in *Nonlinear Partial Differential Equations in Engineering and Applied Science*, edited by R. L. Sternberg, A. J. Kalinowki, and J. S. Papadakis (Marcel Dekker, New York, 1980), p. 397.

[6]J. Satsuma, M. J. Ablowitz, and Y. Kodama, Phys. Lett. 73A, 283 (1979).

[7]K. M. Case, Proc. Nat. Acad. Sci. U.S.A. 75, 3562 (1978), and 76, 1 (1979).

[8]H. H. Chen, Y. C. Lee, and N. R. Pereira, Phys. Fluids 22, 187 (1979).

[9]Y. Matsuno, J. Phys. A 12, 619 (1979).

[10]J. Satsuma and Y. Ishimori, J. Phys. Soc. Jpn. 46, 681 (1979).

[11]A. Nakamura, J. Phys. Soc. Jpn. 47, 1701 (1979).

[12]T. L. Bock and M. D. Kruskal, Phys. Lett. 74A, 173 (1979).

[13]V. E. Zakharov and A. B. Shabat, Functional Anal. Appl. 8, 226 (1974).

[14]J. D. Meiss and N. R. Pereira, Phys. Fluids 21, 700 (1978).

Linearization of the Korteweg–de Vries and Painlevé II Equations

A. S. Fokas and M. J. Ablowitz
Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam, New York 13676
(Received 22 May 1981)

A new integral equation which linearizes the Korteweg-de Vries and Painlevé II equations, and is related to the potentials of the Schrödinger eigenvalue problem, is presented. This equation allows one to capture a far larger class of solutions than the Gel'fand-Levitan equation, which may be recovered as a special case. As an application this equation, with the aid of the classical theory of singular integral equations, yields a three-parameter family of solutions to the self-similar reduction of Korteweg-de Vries which is related to Painlevé II.

PACS numbers: 02.30.+g

Since the work of Gardner *et al.* in 1967,[1] there has been wide interest in the analysis of nonlinear evolution equations solvable by the so-called inverse-scattering transform (IST). The prototype example is the Korteweg–de Vries (KdV) equation

$$u_t + 6uu_x + u_{xxx} = 0. \tag{1}$$

In this note we shall present a new linear integral equation which, in principle, allows one to capture a far larger class of solutions than does the Gel'fand-Levitan equation. Specifically we claim that if $\varphi(k; x, t)$ solves

$$\varphi(k; x, t) + i \exp[i(kx + k^3 t)] \int_L \frac{\varphi(l; x, t)}{l + k} d\lambda(l) = \exp[i(kx + k^3 t)], \tag{2}$$

where $d\lambda(k)$ and L are an appropriate measure and contour, respectively, then

$$u = -\frac{\partial}{\partial x} \int_L \varphi(k; x, t) d\lambda(k) \tag{3}$$

solves the KdV equation. The well-documented physical significance of the KdV equation, of its self-

 © 1981 The American Physical Society

similar analogue, and of the associated Schrödinger scattering problem require us to attempt to characterize the form of the most general solution/potential possible.

We now enumerate the basic results given in this note. (i) We give a direct proof that (2) and (3) solve (1); (ii) we show how the well known Gel'fand-Levitan equation can be obtained from (2) as a special case; and (iii) we characterize by a matrix Fredholm equation a three-parameter family of solutions to the similarity ordinary differential equation of (1) which is directly related to the classical second equation of Painlevé (P II). We end with some remarks regarding the role of Bäcklund transformations and relevant generalizations.

We now consider (i). The point of view we take here is, in spirit, similar to that of Zakharov and Shabat.[2] Specifically, by direct calculation we show that solutions of (2) substituted in (3) satisfy (1). We make two assumptions: (a) $d\lambda$ and L are such that differentiation by x, t may be interchanged with $\int_L$; (b) the homogeneous integral equation has only the zero solution. Defining $\tilde{L} = L_0 + 3u\partial_x$, where $L_0 = \partial_t + \partial_x^3$, after some manipulation we find

$$\tilde{L}\varphi(k; x, t) + i \exp[i(kx + k^3t)] \int_L \frac{\tilde{L}\varphi(l; x, t)}{l + k} d\lambda(l) = 3k[k\varphi_x + i\varphi_{xx} + iu\varphi]. \tag{4}$$

Similar calculations show that the quantity in brackets in the right-hand side of (4) satisfies the homogeneous integral equation. Hence $\tilde{M}\varphi = k\varphi_x + i\varphi_{xx} + iu\varphi = 0$ which implies $\tilde{L}\varphi = 0$, whereupon $\partial_x \int_L (\tilde{L}\varphi) d\lambda = 0$ is (1). Moreover the equation $\tilde{M}\varphi = 0$ is directly related to the Schrödinger eigenvalue problem. If we define

$$\varphi(k; x, t) = \psi(k; x, t) \exp[i(kx + k^3t)/2],$$

then $\tilde{M}\varphi = 0$ gives

$$\psi_{xx} + (\tfrac{1}{2}k)^2\psi + u\psi = 0. \tag{5}$$

Next we pass on to (ii). The classical theory of inverse scattering and appropriately decaying solutions of KdV may be most easily obtained as follows. Let the measure $d\lambda(k) = r_0(\frac{1}{2}k)dk/2\pi$, where $r_0(k)$ is the usual reflection coefficient of $u(x, 0)$ and the contour L goes over all the poles of $r_0(k)$. [Here we have assumed, for convenience, that $u(x, 0) \to 0$ rapidly as $|x| \to \infty$.] Then substituting the expression for φ into (2), defining

$$K(x, y, t) = -(\tfrac{1}{2}) \int_L \psi(k; x, t) \exp[i(ky + k^3t)/2] d\lambda(k),$$

and using

$$\exp[i(k + l)x/2]/(l + k) = -i\{\int_x^\infty \exp[i(k + l)\xi/2]/2 d\xi\}$$

(k, l satisfy $\mathrm{Im}k, \mathrm{Im}l > 0$), we obtain

$$K(x, y; t) + F(x + y; t) + \int_x^\infty K(x, \xi; t) F(\xi + y; t) d\xi = 0, \tag{6}$$

where

$$F(x, t) = (\tfrac{1}{2}) \int_L \exp[i(kx/2 + k^3t)] d\lambda(k),$$

and $u(x, t) = 2\partial_x K(x, x; t)$. Hence by choosing the above measure $d\lambda$ and contour L, the Gel'fand-Levitan equation (6) may now be completely bypassed.

Soliton solutions of (1) may be calculated in a particularly easy manner from (2). Locations of the poles on the imaginary k axis in $r(k, 0)$ correspond to soliton amplitudes, and the residues of $r(k, 0)$ at these locations play the role of the normalization coefficients. Pure solitons may also be obtained by taking the measure as

$$d\lambda(k)_s = \sum_{j=1}^{N} c_j \delta(k - i\kappa_j) dk$$

(L passes through the $k = i\kappa_j$). Then (2) reduces to a linear *algebraic* system from which the well known N-soliton solution is immediately obtained.

We now discuss (iii). The KdV equation admits the similarity transformation $u(x, t) = U(x')/(3t)^{2/3}$, where $x' = x/(3t)^{1/3}$. The equation for U is given by (dropping the primes)

$$K_1(U) = U''' + 6UU' - (2U + xU') = 0. \tag{7}$$

We note that (7) is directly related to P II:

$$P_2(V) = V'' - xV - 2V^3 = \alpha. \tag{8}$$

Specifically we note that the transformations $U = -V^2 - V'$, $V = (U' + \alpha)/(2U - x)$ relate (8) to the

equation

$$K_2(U) = U'' + 2U^3 - xU + [\nu + U' - (U')^2]/(2U - x) = 0$$

with $\nu = \alpha(\alpha+1)$. However, by direct calculation $[(2U-x)K_2(U)]' = (2U-x)K_1(U)$, hence $K_2(U)$ is an integral of (7), and thus there is a direct transformation between (7) and (8).[3] One may make use of these transformations fo find all the known (see, for example, Lukashevich[4] and Erugin[4]) elementary solutions of P II. Ablowitz and Segur[5] had established a connection between P II and IST and had characterized a one-parameter family of solutions via the Gel'fand-Levitan equation. Recently Flaschka and Newell[6] considered P II via monodromy theory. In the latter work the authors derived a formal system of linear singular integral equations for the general solution of P II. However, the highly nontrivial question of existence of solutions was left open.

An application of the result presented above in (i) is that a three-parameter family of solutions of (7) may be obtained from the linear singular integral equation

$$\varphi(t) + \frac{b(t)}{i\pi}\int_L \frac{\varphi(\tau)}{\tau + t}\,d\tau = f(t), \quad t \text{ on } L, \tag{9}$$

where $b(t) = f(t) = \exp[i(tx + t^3/3)]$ and $\int_L = \sum_{j=1}^5 \hat{\rho}_j \int_{L_j}$ (see Fig. 1), $\hat{\rho}_1 = \hat{\rho}_2 = \rho_1$, $-\hat{\rho}_3 = \hat{\rho}_4 = \rho_2$, $\hat{\rho}_5 = \rho_3$. (Hereafter j always stands for $j = 1, \ldots, 5$). The solution to (7) is then obtained from

$$U = \frac{1}{\pi}\frac{\partial}{\partial x}\int_L \varphi(\tau)\,d\tau$$

(φ depends parametrically on x). We note that both (9) and U are obtained from (2) and (3) by a self-similar reduction. Moreover the contours L_j are obtained by finding the solution to the linear problem $(U = w_x)\ w''' - (w + xw') = 0$ in terms of integral representations and then deforming these contours so that they all pass through the origin. For example, note that $L_1 + L_2$ may be deformed to the usual Airy-function contour. If we restrict ourselves to this Airy contour, the result in Ref. 5 is obtained in the same manner as that in (ii) above.

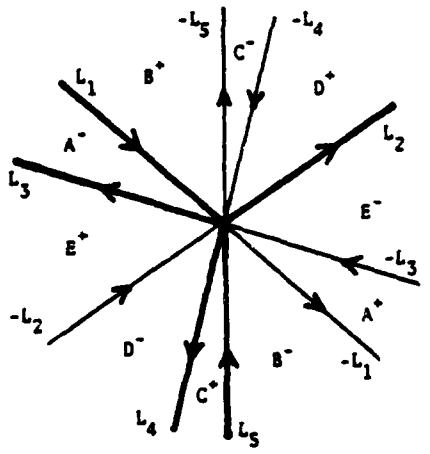


FIG. 1. Contours associated with Eqs. (9) and (12).

We shall proceed to demonstrate that (9) may be reduced to a system of Riemann-Hilbert problems which are solvable using Fredholm theory. For this we need the full power of the classical theory of singular integral equations.[7-9]

Consider the sectionally holomorphic function

$$\Phi(z) = \frac{1}{2\pi i}\int_L \frac{\varphi(\tau)}{\tau - z}\,d\tau. \tag{10}$$

The lines of discontinuity of $\Phi(z)$ are L_j; thus using the Plemelj formulas, we have

$$\Phi^+(t) = \Phi^-(t) = \frac{1}{2\pi i}\int_L \frac{\varphi(\tau)}{\tau - t}\,d\tau, \quad t \text{ on } -L_j,$$
$$\Phi^\pm(t) = \pm\tfrac{1}{2}\hat{\rho}_j\varphi(t) + \frac{1}{2i\pi}\int_L \frac{\varphi(\tau)}{\tau - t}\,d\tau, \quad t \text{ on } L_j, \tag{11}$$

where $\Phi^\pm(t)$ for t on L_j has the standard definitions[7-9] of limits of $\Phi(z)$ as $z \to t$ from the "left-hand side" (+) and "right-hand side" (−) of L_j, and where principal-value integrals are implied when needed. With use of (11), and Eq. (9) for t on L_j and $-t$ on $-L_j$, we obtain a system which we choose to write in the form

$$\underline{\Phi}^+(t) = G(t)\underline{\Phi}^-(t) + \underline{F}(t), \quad t \text{ on } \overline{L}_j, \tag{12}$$

where $\overline{L}_j = L_j + (-L_j)$, $\underline{\Phi}^+(t) = [\Phi^+(t), \Phi^-(-t)]^T$, $\underline{\Phi}^-(t) = \underline{\Phi}^+(-t)$, $\underline{F}(t) = [f(t)H(t), -f(-t)H(-t)]^T$, $H(t) = \{\hat{\beta}_j$ if t on L_j, 0 if t on $-L_j\}$ and the components of the 2×2 matrix $G(t)$ are $G_{11}(t) = -2b(t) \times H(t) = -G_{22}(-t)$, $G_{12} = G_{21} = 1$.

One can prove the following statements.

(a) $\Phi^+(-t)$, $\Phi^-(-t)$ are "minus" and "plus" functions, respectively. (b) Necessary conditions for solvability of (12) are the symmetry conditions $G(t) = [G(-t)]^{-1}$, $\underline{F}(t) + G(t)\underline{F}(-t) = 0$, which are satisfied by the above G, $\underline{F}$. (c) Thus (12) defines a system of discontinuous Riemann-Hilbert problems with the additional restriction that $\underline{\Phi}^-(t) = \underline{\Phi}^+(-t)$. However this condition can be relaxed since one can show that (12) always admits a solution with this restriction, and moreover, in our case the solution is unique.

In order to solve (12) we first consider the

homogeneous problem. The standard procedure is to transform the discontinuous homogeneous problem to a continuous one, and then obtain the fundamental set of solutions.

Associated with a given contour Σ_j, define the following auxilliary functions

$$\omega_{jk}^{+}(t)=\left(\frac{t}{t+z_j}\right)^{\lambda_{jk}},\quad \omega_{jk}^{-}(t)=\left(\frac{t}{t-z_j}\right)^{\lambda_{jk}},$$
$$\omega_{jk}=\frac{\omega_{jk}^{+}}{\omega_{jk}^{-}},\quad k=1,2, \tag{13}$$

where z_j is some j-dependent fixed point off Σ_j. The branches of the above functions are chosen such that ω_{jk}^{+} and ω_{jk}^{-} are plus and minus functions, respectively (e.g., the branch cut for ω_{jk}^{+} is taken between 0, $-z_j$ and hence lies to the right of Σ_j). The properties $\omega_{jk}(0+)=\exp[-i\pi\lambda_{jk}]$, $\omega_{jk}(0-)=\exp[i\pi\lambda_{jk}]$, $\omega_{jk}^{+}(t)=\omega_{jk}^{-}(-t)$ allow us to map the homogeneous system $\underline{\Phi}^{+}(t)=G(t)\underline{\Phi}^{-}(t)$ which has a discontinuity at $t=0$ to the following Riemann-Hilbert system which is continuous at the origin:

$$\underline{\Psi}^{+}(t)=g(t)\underline{\Psi}^{-}(t),\quad t \text{ on } \Sigma_j, \tag{14}$$

where we have used the transformation $\underline{\Phi}^{+}(t)=A\Omega^{+}(t)\underline{\Psi}^{+}(t)$, $\underline{\Phi}^{-}(t)=A\Omega^{-}(t)\underline{\Psi}^{-}(t)$ and hence $g(t)=[\Omega^{+}(t)]^{-1}A^{-1}G(t)A\Omega^{-}(t)$, with A, $\Omega^{+}(t)$ defined by

$$A_j=\begin{pmatrix}\dfrac{1-\Lambda_{j1}}{2\hat{\rho}_j}\exp[i\pi\lambda_{j1}/2] & \dfrac{1-\Lambda_{j1}}{2\hat{\rho}_j}\exp[i\pi\lambda_{j2}/2]\\ \exp[i\pi\lambda_{j1}/2] & \exp[i\pi\lambda_{j2}/2]\end{pmatrix},\quad \Omega_j^{+}=\begin{pmatrix}\omega_{j\alpha}^{+} & 0\\ 0 & \omega_{j\beta}^{+}\end{pmatrix}, \tag{15}$$

where for $j=2,3,4$ we have $\alpha=1$, $\beta=2$ and for $j=1,5$ we have $\alpha=2$, $\beta=1$; the λ_{jk} and Λ_{jk} are defined by

$$\exp[i\pi\lambda_{j1}]=\hat{\rho}_j+(1+\hat{\rho}_j^2)^{1/2},\quad \exp[i\pi\lambda_{j2}]=-\hat{\rho}_j+(1+\hat{\rho}_j^2)^{1/2},\quad \Lambda_{j1}=\exp[2i\pi\lambda_{j1}],\quad \Lambda_{j2}=\exp[2i\pi\lambda_{j2}].$$

The matrix $g(t)$ has the properties $g(t)=[g(-t)]^{-1}$ and $\det g=-1$.

One may characterize a solution of the system (14) by imposing the condition $\underline{\Psi}(z)-\underline{\Psi}_A^{-}=\underline{\chi}$ as $|z|\to\infty$ in A^{-}. This leads to a Fredholm equation for, say, $\underline{\Psi}^{+}(t)$, which however must be interpreted in a suitable principal-value sense as it does not converge in the normal sense at infinity. Alternatively, one may obtain a regular Fredholm equation of the second kind by imposing conditions at a finite point off all contours, say $z=1$. This leads to the following Fredholm equation for $\underline{\Psi}^{+}(t)$:

$$\underline{\Psi}^{+}(t)+\frac{1}{2\pi i}\int_{\Sigma}\left[\frac{1}{\tau-t}-\frac{1}{\tau-1}\right][g_j(t)g(-\tau)-I]\underline{\Psi}^{+}(\tau)d\tau=g_j(t)\underline{\beta},\quad t \text{ on } \Sigma_j, \tag{16}$$

where $\underline{\beta}=\underline{\Psi}(1)$, $\int_{\Sigma}=\sum_{j=1}^{5}(\int_{\Sigma_j}+\int_{-\Sigma_j})$, and I is the unit matrix. Any two linearly independent $\underline{\beta}$ vectors, say $\underline{\beta}_{1,2}$, lead to a fundamental matrix $Y^{+}(t)=[\underline{\Psi}_1^{+}(t),\underline{\Psi}_2^{+}(t)]$ for the system (14).

With use of the above results the fundamental matrix of the discontinuous problem (12) is given by

$$X^{+}(t)=A\Omega^{+}(t)[\underline{\Psi}_1^{+}(t),\underline{\Psi}_2^{+}(t)]. \tag{17}$$

Hence the solution of (12) is given by

$$\underline{\Phi}^{+}(t)=\frac{\underline{F}(t)}{2}+\frac{1}{2\pi i}X^{+}(t)\int_{\Sigma}\frac{[X^{+}(\tau)]^{-1}\underline{F}(\tau)}{\tau-t}d\tau. \tag{18}$$

Having obtained $\underline{\Phi}^{+}(t)$ and using (11) to obtain $\varphi(t)$, we have characterized a three-parameter family of solutions of U. With use of the results of Fredholm's theory the nonmovable critical-point property of U is easily verified.

Finally, we make some remarks. First, we only expect from (2) to obtain solutions to P II in the range $-\frac{1}{2}<\alpha<\frac{1}{2}$. To obtain the solution for all ranges of α, we believe, the Bäcklund transformations (following Rosales[10]) and "finite perturbations" (see, for example, Ablowitz and Cornille[11]) of suitable elementary solutions must be employed. Similarly, wider classes of solutions to KdV should be obtainable this way (we shall remark on this more completely in the future). Second, straightforward generalizations to the higher-order KdV equations, as well as to many other nonlinear evolution equations, are possible. Third, motivation for some of the ideas in this note originate from the concept of summing perturbation series. Relevant perturbation series can be readily developed (see, for example, Refs. 11 and 12).

We appreciate useful discussions with M. Kruskal, R. Coifman, and E. Barouch. We are also sincerely grateful to C. Cosgrove for many important comments on a first version of this paper and, in particular, for his suggestion of Eq. (16).

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

[2]V. E. Zakharov and A. B. Shabat, Funktsional. Anal. Prilozhen. 8, 43 (1974) [Functional Anal. Appl. 8, 226 (1974)].

[3]A. Fokas and M. J. Ablowitz, to be published.

[4]N. A. Lukashevich, Differ. Uravn. 7, 1124 (1971) [Differ. Equat. 7, 853 (1971)]; N. P. Erugin, Differ. Uravn. 12, 579 (1976) [Differ. Equat. 12, 405 (1976)].

[5]M. J. Ablowitz and H. Segur, Phys. Rev. Lett. 38, 1103 (1977).

[6]H. Flaschka and A. C. Newell, Commun. Math. Phys. 76, 65 (1980).

[7]N. I. Muskhelishvili, in *Singular Integral Equations*, edited by J. Radok (Noordhoff, Groningen, 1953).

[8]F. D. Gakhov, in *Boundary Value Problems*, edited by I. N. Sneddon (Pergamon, New York, 1966), Vol. 85.

[9]N. P. Vekua, in *Systems of Singular Integral Equations*, edited by J. H. Ferziger (Gordon and Breach, New York, 1967).

[10]R. Rosales, Stud. Appl. Math. 59, 117 (1978).

[11]M. J. Ablowitz and H. Cornille, Phys. Lett. 72A, 277 (1979).

[12]S. Oishi, J. Phys. Soc. Jpn. 47, 1037 (1979).

Perturbations of Solitons and Solitary Waves

By Yuji Kodama and Mark J. Ablowitz

A direct perturbation method is developed to investigate the evolution of solitary waves in the presence of small perturbations. A uniformly valid first order solution is constructed. The method is applied to several nonlinear evolution equations which support solitons or solitary waves. Finally, the method is compared with other approaches in the literature.

1. Introduction

In recent years there has been a great deal of interest in nonlinear dispersive wave problems admitting, as special solutions, localized, nondecaying waves, called solitary waves or solitons. In this paper we shall mean by solitons those waves which satisfy nonlinear evolution equations solvable by the inverse scattering transform (I.S.T.—see for example [1], [2]). Given the physical significance of these solutions, a natural question to ask is how they develop under the influence of weak perturbations. For those problems giving rise to solitons, there are three methods which have been developed that use the techniques of I.S.T. [3–5]. The methods in [3], [4] develop perturbed equations in the scattering data via an associated linear eigenvalue problem. The field variable is recovered via the inverse equations (e.g. the linear Gel'fand-Levitan equation). From a somewhat different point of view, Ref. [5] develops a perturbation theory using a Green's function to solve the associated linearized equation in the higher order problems. In order to calculate the Green's function, information from I.S.T. is needed.

On the other hand, it is well known that there exist very general perturbation techniques that are applicable to nonlinear problems where the leading order problem has a well-defined solution (e.g. solitons, breathers, solitary waves, periodic solutions, etc.). These ideas have been applied to a wide variety of

Address for correspondence: M. J. Ablowitz and Y. Kodama, Department of Mathematics, Clarkson College, Potsdam, NY 13676.

STUDIES IN APPLIED MATHEMATICS 64:225–245 (1981)

Copyright © 1981 by the Massachusetts Institute of Technology
Published by Elsevier North Holland, Inc.
0022-2526/81/030225+21$02.50

problems (for example [6]–[11]). Our main purpose in this paper is to illustrate how these simple ideas can be applied to problems with solitons and solitary waves. In the soliton case, we compare our results with those already developed in the literature. In some cases we have found discrepancies between the existing results. We shall discuss this and make suggested modifications to some of the presently used techniques. Specifically, we examine perturbations associated with the Korteweg-deVries (KdV), modified KdV (MKdV), nonlinear Schrödinger (NLS), and sine-Gordon (SG) equations. In addition, we discuss perturbations of certain other "higher" nonlinear KdV and NLS equations (presumably not solvable by I.S.T.). In some of these cases we find what appears to be a self-focusing singularity.

The basic idea of the perturbation approach is as follows: We study the solution of a perturbed nonlinear dispersive wave equation of the general form

$$K(q, q_t, q_x, \ldots) = F(q, q_x, \ldots), \qquad 0 < \varepsilon \ll 1, \tag{1.1}$$

where K and F are nonlinear functions of $q, q_x, \ldots$. The unperturbed equation for $\varepsilon = 0$ is

$$K(q_0, q_{0t}, q_{0x}, \ldots) = 0, \tag{1.2}$$

and its solution q_0 is to be taken as a solitary wave or soliton solution (or perhaps a breather). We write this solution in terms of certain natural fast and slow variables:

$$q_0 = \hat{q}_0(\theta_1, \theta_2, \ldots, \theta_m, T; P_1, P_2, \ldots, P_N). \tag{1.3}$$

In (1.3), θ_i $(i = 1, \ldots, m)$ are so-called "fast" variables, $T = \varepsilon t$ is a "slow" variable, and the P_l $(l = 1, \ldots, N)$ are parameters which depend on the slow variable (in some problems, one might need to also introduce a slow variable $X = \varepsilon x$). In many problems we need only one fast variable, such as $\theta = x - P_1 t$ in the unperturbed problem. We generalize θ to satisfy $\partial\theta/\partial x = 1$, and $\partial\theta/\partial t = -P_1$ and use $P_1 = P_1(T)$ to remove secular terms. With this, we call such a solution (1.3) a quasistationary solution and write $q = \hat{q}(\theta, T, \varepsilon)$. It is necessary that we develop equations for the $P_1, \ldots, P_N$ by using appropriate conditions, such as secularity conditions (there must be N such independent conditions). Some of these conditions are formed from Green's identity as follows. We assume an expression for $\hat{q}$ of the form

$$\hat{q} = \hat{q}_0 + \varepsilon \hat{q}_1 + \cdots$$

(after introducing appropriate variables θ_i, T, etc.). Then (1.2) is the leading order problem, and

$$L(\partial_\theta, \hat{q}_0)\hat{q}_1 = F(\hat{q}_0) - \frac{\partial K}{\partial q_t} \cdot q_T \Big|_{q = \hat{q}_0} \equiv \hat{F} \tag{1.4}$$

is the first order equation. Here $L(\partial_{\theta_i}, \hat{q}_0)u=0$ is a linearized equation of $K(q, q_t, q_x, \ldots)=0$ after transforming (x, t) to the appropriate coordinate θ_i. Denoting by v_i $(i=1,\ldots,M)$ the M solutions of the homogeneous adjoint problem satisfying the necessary boundary conditions (e.g. $v_i \to 0$ as $|\theta| \to \infty$),

$$L^A v_i = 0, \qquad i = 1,\ldots,M, \quad M \leqslant N,$$

where L^A is the adjoint operator to L, we form

$$(L\hat{q}_1)\cdot v_i - (L^A v_i)\cdot \hat{q}_1 = \hat{F} v_i. \tag{1.5}$$

Equation (1.5) is always a divergence. It may be integrated to give the secularity conditions. These secularity conditions allow us to be able to compute a solution $\hat{q}_1$ to (1.4) which satisfies suitable boundary conditions (e.g. $\hat{q}_1$ is bounded as $|\theta| \to \infty$). However, as is standard in perturbation problems, there is still freedom in the solution. This is due to the fact that some terms in the solution $\hat{q}_1$ can be absorbed in the leading order solution $\hat{q}_0$ by shifting the other parameters. The solution $\hat{q}_1$ can be made unique by imposing additional conditions which reflect specific initial conditions or other normalizations.

Some properties of this method are the following:

(1) A simple *uniform* quasistationary expansion on $|x| < \infty$ is generally not obtained (see also [6]).

(2) In its region of validity we find a quasistationary solution, i.e., the solution depends on the θ_i and T only.

(3) In order to develop a valid expansion, we match the solution obtained via the method above to a nonstationary solution for large $|\theta_i|$ [e.g. $|\theta| \sim 1/\varepsilon$].

In the following sections, we apply the general scheme outlined above to a number of nonlinear dispersive systems supporting a soliton or solitary wave solution. In a separate communication we discuss the analysis for the more complicated examples of bound solitons, i.e. breathers [18].

2. Perturbed KdV and MKdV equations

As examples of the general scheme in Sec. 1, we study the KdV and MKdV equations with a small dissipationlike perturbation. (For small dispersive perturbations, see Ref. [12].) Physically, these equations correspond to the evolution of a soliton in a slowly varying medium [10, 11]. An interesting feature of these equations is the appearance of a shelf behind the perturbed soliton due to the dissipative perturbation [3, 4, 10, 11]. We shall find that in this case the perturbation expansion is not uniform, and there are three regions to the solution. We discuss this problem in Sec. 4, and construct a uniform solution.

2.1. Perturbed KdV equation

Let the perturbed KdV equation be of the form

$$q_t + 6qq_x + q_{xxx} = -\varepsilon\gamma q, \tag{2.1}$$

in which the soliton solution to the unperturbed equation ($\varepsilon=0$) can be written

$$q_0 = 2\eta^2 \operatorname{sech}^2 \eta(\theta-\theta_0), \qquad \frac{\partial\theta}{\partial x}=1, \quad \frac{\partial\theta}{\partial t}=-4\eta^2. \tag{2.2}$$

[Although we shall present results for (2.1), one should consider (2.1) only a prototype equation. The analysis applies in much wider generality.] Here η and θ_0 are arbitrary parameters which may depend on the long time scale $T=\varepsilon t$. Under the assumption of quasistationarity, (2.1) becomes

$$-4\eta^2\hat{q}+6\hat{q}\hat{q}_\theta+\hat{q}_{\theta\theta\theta} = -\varepsilon\gamma\hat{q}-\varepsilon\hat{q}_T. \tag{2.3}$$

Expanding $\hat{q}$ in terms of ε, at leading order, we have

$$-4\eta^2\hat{q}_{0\theta}+6\hat{q}_0\hat{q}_{0\theta}+\hat{q}_{0\theta\theta\theta}=0, \tag{2.4}$$

and the solution

$$\hat{q}_0 = 2\eta^2 \operatorname{sech}^2 \eta(\theta-\theta_0). \tag{2.5}$$

At order ε, we have

$$\begin{aligned}\hat{L}\hat{q}_1 &\equiv -4\eta^2\hat{q}_{1\theta}+6(\hat{q}_0\hat{q}_1)_\theta+\hat{q}_{1\theta\theta\theta}=\hat{F}_1,\\ \hat{F}_1 &\equiv -\gamma\hat{q}_0-\hat{q}_{0T}\\ &= -\gamma\hat{q}_0-\frac{1}{\eta}\eta_T\{2\hat{q}_0+(\theta-\theta_0)\hat{q}_{0\theta}\}+\theta_{0T}\hat{q}_{0\theta}.\end{aligned} \tag{2.6}$$

From (2.4) we find that $\hat{q}_0$ is a proper solution of the adjoint problem of $\hat{L}u=0$, i.e.

$$\hat{L}^\dagger\hat{q}_0=0, \qquad \hat{L}^\dagger=4\eta^2\partial_\theta-6\hat{q}_0\partial_\theta-\partial_\theta^3, \tag{2.7}$$

decaying rapidly as $|\theta|\to\infty$. Then the compatibility condition

$$\int_{-\infty}^{\infty}\hat{q}_0\hat{F}_1\,d\theta=0 \tag{2.8}$$

leads to

$$\frac{1}{\eta}\frac{\partial\eta}{\partial T}=-\tfrac{2}{3}\gamma, \qquad \text{or} \quad \eta(T)=\eta(0)\exp\left(-\frac{2}{3}\int_0^T\gamma\,dT'\right). \tag{2.9}$$

This implies that the amplitude and speed of the soliton are decreasing ($\gamma > 0$) adiabatically according to the dissipation (e.g. a climbing soliton on a beach). Taking (2.9) into account, we solve (2.6) and obtain the solution (see Appendix for the details of the calculations).

$$\hat{q}_1 = \frac{\gamma}{6\eta}\left[-1+\tanh\phi+3\left(1+\frac{\eta}{\gamma}\theta_{0T}\right)(1-\phi\tanh\phi)\operatorname{sech}^2\phi\right.$$

$$\left.+\phi(2-\phi\tanh\phi)\operatorname{sech}^2\phi\right] \quad (|\phi| \ll \varepsilon^{-1/2}), \tag{2.10}$$

where $\phi = \eta(\theta - \theta_0)$. Higher order calculations indicate that the expansion breaks down when $|\phi| \sim \varepsilon^{-1/2}$. (It should be noted that this order is related to the breakdown of the expansion in [3, 4] for time $t \sim \varepsilon^{-1/2}$.) From (2.10), one can see that there is a shelf introduced by the dissipation, i.e. asymptotically

$$\hat{q}_1 \sim -\frac{\gamma}{3\eta}\{1-2\phi^2 e^{2\phi}\} \qquad \text{for} \quad 1 \ll -\phi \ll \varepsilon^{-1/2},$$

$$\sim -\frac{2\gamma}{3\eta}\phi^2 e^{-2\phi} \qquad \text{for} \quad 1 \ll \phi \ll \varepsilon^{-1/2}, \tag{2.11}$$

which agrees with the results in Ref. [4] (via the inverse method). We also notice at this point that the parameter θ_0 can be taken arbitrarily, since the term $\theta_{0T}(1-\phi\tanh\phi)\operatorname{sech}^2\phi$ can be absorbed into the leading order solution $\hat{q}_0$ by shifting η to $\eta - \varepsilon\theta_{0T}/(8\eta)$. However, for η to be given by certain initial data, one can determine the evolution equation of θ_0 by the following. Let us consider an initial value problem with the initial value in the form of an unperturbed solitary wave, i.e.,

$$q(x,0) = 2\eta^2 \operatorname{sech}^2 \eta x. \tag{2.12}$$

From Eq. (2.1), we have the following global relation (rate of change of energy):

$$\frac{d}{dt}\int_{-\infty}^{\infty} q^2\,dx = -2\varepsilon\gamma\int_{-\infty}^{\infty} q^2\,dx. \tag{2.13}$$

Moreover, let us assume that q takes the form $q_s + \delta q$ where q_s expresses soliton part, i.e. (2.5), and δq the correction to the soliton. Taking (2.9) into account [i.e., at leading order (2.13) is (2.9)], we have

$$\frac{d}{dt}\Delta(t) = -2\varepsilon\gamma\Delta(t) \tag{2.14}$$

where $\Delta(t) = \int_{-\infty}^{\infty}\{q_s\,\delta q + (\delta q)^2/2\}\,dx$. From $\delta q(x,0) = 0$, we obtain

$$\Delta(t) = 0. \tag{2.15}$$

It turns our that (see Sec. 4) the length of the shelf is $\sim\varepsilon^{-1}$ for $t\sim\varepsilon^{-1}$. Hence for these times the order of the second term in $\Delta(t)$ is the same as the first one. We argue that even though for short time the problem is not stationary, the nonstationary portion of the soliton quickly moves to the tail of the soliton. For times $t\sim\varepsilon^{-1}$ the region near the soliton, $|\theta|\ll\varepsilon^{-1/2}$, is quasistationary, and in this region $\delta q=\delta q(\theta, T)$. Hence, in order to determine the evolution equation of the parameter θ_0, we require the following relation, as an additional condition:

$$\int_{-\infty}^{\infty}\hat{q}_0(\theta)\hat{q}_1(\theta)\,d\theta+\frac{1}{2}\int_{-\infty}^{\infty}(\delta q)^2dx=0. \tag{2.16}$$

The condition (2.16) gives

$$\frac{\partial\theta_0}{\partial T}=-\frac{\gamma}{3\eta}-\frac{1}{2\eta}\int_{-\infty}^{\infty}(\delta q)^2\,dx. \tag{2.17}$$

Here, we notice that for the range of time $1\ll t\ll\varepsilon^{-1}$ the second term in (2.17) [i.e. $\int_{-\infty}^{\infty}(\delta q)^2\,dx$ in $\Delta(t)$] can be ignored. For this range of time, Eq. (2.17) gives the same results as Ref. [4]. Also in (2.16) and (2.17), we must use the results of Sec. 4 where we compute δq.

2.2. *Perturbed MKdV equation*

In this subsection, we consider a prototype perturbed MKdV equation which takes the form

$$q_t+6q^2q_x+q_{xxx}=-\varepsilon\gamma q. \tag{2.18}$$

The soliton solution of the unperturbed MKdV equation is given by

$$q_0=\eta\,\mathrm{sech}\,\eta(\theta-\theta_0),\qquad \frac{\partial\theta}{\partial x}=1,\quad \frac{\partial\theta}{\partial t}=-\eta^2. \tag{2.19}$$

The analysis is similar to that of the KdV equation; hence, we shall only outline the method. Letting $\hat{q}(\theta, T;\varepsilon)=\hat{q}_0+\varepsilon\hat{q}_1+\cdots$, as in the KdV case, we find

$$-\eta^2\hat{q}_\theta+6\hat{q}^2\hat{q}_\theta+\hat{q}_{\theta\theta\theta}=-\varepsilon\gamma\hat{q}-\varepsilon\hat{q}_T, \tag{2.20}$$

from which, at order ε, we have

$$\hat{L}\hat{q}_1\equiv-\eta^2\hat{q}_{1\theta}+6(\hat{q}_0^2\hat{q}_1)_\theta+\hat{q}_{1\theta\theta\theta}=\hat{F}_1,$$

$$\hat{F}_1\equiv-\gamma\hat{q}_0-\frac{1}{\eta}\eta_T\{\hat{q}_0+(\theta-\theta_0)\hat{q}_{0\theta}\}+\theta_{0T}\hat{q}_{0\theta}. \tag{2.21}$$

By using $\hat{L}^{\wedge}\hat{q}_0 = 0$, the compatibility condition is given by

$$\int_{-\infty}^{\infty} \hat{q}_0 \hat{F}_1 d\theta = 0. \tag{2.22}$$

which leads to

$$\frac{1}{\eta}\frac{\partial \eta}{\partial T} = -2\gamma. \tag{2.23}$$

From (2.23), we calculate the solution $\hat{q}_1$ of (2.21) to be

$$\begin{aligned}\hat{q}_1 = \frac{\gamma}{\eta^2}\Big[&(1-2\operatorname{sech}^2\phi)(2\tan^{-1}e^{\phi}-\pi)\\ &+\left\{\phi+\left(2\log\operatorname{sech}\phi-\tfrac{1}{2}\phi^2\right)\tanh\phi\right\}\operatorname{sech}\phi\Big]\\ &+\frac{1}{2\eta}\theta_{0T}(1-\phi\tanh\phi)\operatorname{sech}\phi \qquad (|\phi| \ll \varepsilon^{-1/2}),\end{aligned} \tag{2.24}$$

where $\phi = \eta(\theta-\theta_0)$. Again, the shelf appears, and asymptotically,

$$\begin{aligned}\hat{q}_1 &\sim -\frac{\gamma}{\eta^2}(\pi+\phi^2 e^{\phi}) \qquad \text{for} \quad 1 \ll -\phi \ll \varepsilon^{-1/2},\\ &\sim -\frac{\gamma}{\eta^2}\phi^2 e^{-\phi} \qquad \text{for} \quad 1 \ll \phi \ll \varepsilon^{-1/2}.\end{aligned} \tag{2.25}$$

As explained before in (2.24), we have the arbitrary parameter θ_0 which can be absorbed in $\hat{q}_0$ by shifting η to $\eta - \varepsilon\theta_{0T}/(2\eta)$. For the initial value problem, we take the relation to be the same as the KdV case, i.e. (2.16), and we obtain

$$\frac{\partial \theta_0}{\partial T} = -\frac{1}{\eta}\int_{-\infty}^{\infty} (\delta q)^2 dx, \tag{2.26}$$

where δq expresses the shelf part, which can be derived in the same way as for KdV (see Sec. 4).

3. Perturbed NLS and SG equations

In this section, we discuss the perturbed NLS and SG equations. Most of the procedure to be used in Sec. 2 can be given natural extensions to these systems. We find that the first order solution $\hat{q}_1$ does not have a shelf, but rather a "dress"

which is located around the soliton. We also find that due to the dress the perturbation expansion is not uniform.

3.1. *Perturbed NLS equation*

Consider a prototype perturbed NLS equation in the form

$$iq_t + q_{xx} + 2q^2q^* = -i\varepsilon\gamma q. \tag{3.1}$$

The unperturbed soliton solution is given by

$$q_0 = \eta \operatorname{sech}\eta(\theta-\theta_0)\exp\left[i\xi(\theta-\theta_0)+i(\sigma-\sigma_0)\right], \tag{3.2}$$

where

$$\frac{\partial\theta}{\partial t} = -2\xi, \qquad \frac{\partial\theta}{\partial x} = 1,$$

$$\frac{\partial\sigma}{\partial t} = \eta^2+\xi^2, \qquad \frac{\partial\sigma}{\partial x} = 0. \tag{3.3}$$

Here, ξ, η, θ_0, and σ_0 are arbitrary functions of the long time scale $T=\varepsilon t$. We assume that the quasistationary solution of (3.1) takes the form

$$q = \hat{q}(\theta, T, \varepsilon)\exp\left[i\xi(\theta-\theta_0)+i(\sigma-\sigma_0)\right]. \tag{3.4}$$

Substituting this into (3.1), we have

$$-\eta^2\hat{q} + \hat{q}_{\theta\theta} + 2\hat{q}^2\hat{q}^* = \varepsilon\hat{F}(q),$$

$$\hat{F}(q) \equiv -i\gamma\hat{q} - i\hat{q}_T + (\theta-\theta_0)\xi_T\hat{q} - (\xi\theta_{0T}+\sigma_{0T})\hat{q}. \tag{3.5}$$

We also assume that $\hat{q}$ can be expanded by

$$\hat{q}(\theta, T, \varepsilon) = \hat{q}_0(\theta, T) + \varepsilon\hat{q}_1(\theta, T) + \ldots, \tag{3.6}$$

where $\hat{q}_0$ is the leading order solution of (3.5), taken from (3.2) to be

$$\hat{q}_0 = \eta \operatorname{sech}\eta(\theta-\theta_0). \tag{3.7}$$

From (3.5) and (3.6), at order ε, we have

$$-\eta^2\hat{q}_1 + \hat{q}_{1\theta\theta} + 4\hat{q}_0^2\hat{q}_1 + 2\hat{q}_0^2\hat{q}_1^* = \hat{F}_1, \tag{3.8}$$

where $\hat{F}_1 = \hat{F}(\hat{q}_0)$. Setting $\hat{q}_1 = \hat{\phi}_1 + i\hat{\psi}_1$, where $\hat{\phi}_1$ and $\hat{\psi}_1$ are real valued functions,

we obtain a system of equations,

$$\hat{L}\hat{\phi}_1 \equiv -\eta^2\hat{\phi}_1 + \hat{\phi}_{1\theta\theta} + 6\hat{q}_0^2\hat{\phi}_1 = \mathrm{Re}\,\hat{F}_1, \tag{3.9a}$$

$$\hat{M}\hat{\psi}_1 \equiv -\eta^2\hat{\psi}_1 + \hat{\psi}_{1\theta\theta} + 2\hat{q}_0^2\hat{\psi}_1 = \mathrm{Im}\,\hat{F}_1. \tag{3.9b}$$

It is easy to see that $\hat{\phi}_1$ and $\hat{\psi}_1$ have localized solutions around the soliton. By noting that the operators $\hat{L}$ and $\hat{M}$ are self-adjoint and $\hat{L}\hat{q}_{0\theta}=0$, $\hat{M}\hat{q}_0=0$, the conditions for the solvability of (3.9) are given by the secularity conditions,

$$\int_{-\infty}^{\infty}\hat{q}_{0\theta}\mathrm{Re}\,\hat{F}_1 d\theta = 0, \tag{3.10a}$$

$$\int_{-\infty}^{\infty}\hat{q}_0\,\mathrm{Im}\,\hat{F}_1 d\theta = 0. \tag{3.10b}$$

From (3.10a, b) we obtain the evolution equations,

$$\frac{\partial\xi}{\partial T}=0 \quad \text{and} \quad \frac{1}{\eta}\frac{\partial\eta}{\partial T}=-2\gamma, \tag{3.11}$$

which show that the amplitude of soliton is decreasing ($\gamma>0$) but the velocity is constant. With (3.11) we have the solutions

$$\hat{\phi}_1 = -\frac{1}{2\eta}(\xi\theta_{0T}+\sigma_{0T})\{1-(\theta-\theta_0)\tanh\eta(\theta-\theta_0)\}\,\mathrm{sech}\,\eta(\theta-\theta_0), \tag{3.12a}$$

$$\hat{\psi}_1 = \frac{\eta}{2}(\theta-\theta_0)\{\theta_{0T}+\gamma(\theta-\theta_0)\}\,\mathrm{sech}\,\eta(\theta-\theta_0), \tag{3.12b}$$

which are valid for $|\theta-\theta_0|\ll\varepsilon^{-1/2}$. The expansions must be modified for sufficiently large $|\theta-\theta_0|$ in a manner similar to KdV (see Sec. 4). Here we have two arbitrary parameters θ_0 and σ_0 which provide the shift of the location and phase of the soliton. We notice that the terms $\{1-(\theta-\theta_0)\tanh\eta(\theta-\theta_0)\}\,\mathrm{sech}\,\eta(\theta-\theta_0)$ and $(\theta-\theta_0)\,\mathrm{sech}\,\eta(\theta-\theta_0)$ can be absorbed in the unperturbed soliton by shifting η and ξ, respectively. For the initial value problem, we take the orthogonality conditions

$$\int_{-\infty}^{\infty}\hat{q}_{0\theta}\hat{\psi}_1 d\theta = 0, \qquad \int_{-\infty}^{\infty}\hat{q}_0\hat{\phi}_1 d\theta = 0, \tag{3.13}$$

which give

$$\frac{\partial\theta_0}{\partial T}=0, \qquad \frac{\partial\sigma_0}{\partial T}=0. \tag{3.14}$$

These conditions may be derived from the following relations (modified conserved quantities), in a similar way as in the KdV case:

$$\frac{d}{dt}\int_{-\infty}^{\infty} qq^* dx = -2\varepsilon\gamma\int_{-\infty}^{\infty} qq^* dx, \tag{3.15}$$

$$\frac{d}{dt}\int_{-\infty}^{\infty} (q_x q^* - q_x^* q)\, dx = -2\varepsilon\gamma\int_{-\infty}^{\infty} (q_x q^* - q_x^* q)\, dx. \tag{3.16}$$

We note that since there is no shelf, the equations (3.13) have especially simple forms.

3.2. Perturbed SG equation

Next consider a prototype perturbed SG equation, in laboratory coordinates, which takes the form

$$q_{tt} - q_{xx} + \sin q = -\varepsilon\gamma q_t. \tag{3.17}$$

The soliton of the unperturbed system can be written

$$q_0 = 4\tan^{-1}\exp\frac{1}{\sqrt{1-C^2}}(\theta-\theta_0), \qquad \frac{\partial\theta}{\partial x} = 1, \quad \frac{\partial\theta}{\partial t} = -C. \tag{3.18}$$

Under the condition of quasistationarity, we have

$$-(1-C^2)\hat{q}_{\theta\theta} + \sin\hat{q} = \varepsilon\hat{F}(\hat{q}) + O(\varepsilon^2),$$

$$\hat{F}(\hat{q}) = C\gamma\hat{q}_\theta + 2(C\hat{q}_\theta)_T, \tag{3.19}$$

from which, at order ε, we obtain

$$\hat{L}\hat{q}_1 \equiv -(1-C^2)\hat{q}_{1\theta\theta} + \cos\hat{q}_0\cdot\hat{q}_1 = \hat{F}(\hat{q}_0). \tag{3.20}$$

By virtue of the fact that $\hat{L}$ is self-adjoint and $\hat{L}q_{0\theta}=0$, the compatibility condition can be given by

$$\int_{-\infty}^{\infty} q_{0\theta}\hat{F}_1(\hat{q}_0)\, d\theta = 0. \tag{3.21}$$

From this, we obtain the evolution of $C(T)$:

$$\frac{dC}{dT} = -\gamma C\frac{1-C^2}{2-C^2}.$$

or

$$\frac{C(T)^4}{1-C(T)^2}=\frac{C(0)^4}{1-C(0)^2}\exp\left(-2\int_0^T\gamma\,dT\right). \tag{3.22}$$

Thus, owing to dissipation ($\gamma>0$), the speed of the soliton, C, decreases to zero. From (3.22), we calculate the solution of (3.20),

$$q_1=\frac{2C}{(1-C^2)^{3/2}}(\theta-\theta_0)\left\{\frac{\partial\theta_0}{\partial T}+\frac{\gamma C^2}{2(2-C^2)}(\theta-\theta_0)\right\}\operatorname{sech}\frac{\theta-\theta_0}{\sqrt{1-C^2}}. \tag{3.23}$$

which is valid for $|\theta-\theta_0|\ll\varepsilon^{-1/2}$. The expansions must be modified for sufficiently large $|\theta-\theta_0|$ (see Sec. 4). Since the term $(\theta-\theta_0)\{(\theta-\theta_0)/\sqrt{1-C^2}\}\sim\hat{q}_{0C}$ can be absorbed into $\hat{q}_0$ by shifting C, θ_0 can be taken arbitrary. For the initial value problem, we take the orthogonality relation,

$$\int_{-\infty}^{\infty}\hat{q}_{0\theta}\hat{q}_{1\theta}\,d\theta=0, \tag{3.24}$$

which leads to

$$\frac{\partial\theta_0}{\partial T}=0. \tag{3.25}$$

[Again the simple form of (3.24) is due to the fact that there is no shelf part.] The condition (3.24) is derived from the modified conservation law,

$$\frac{d}{dt}\int_{-\infty}^{\infty}q_xq_t\,dx=-\varepsilon\gamma\int_{-\infty}^{\infty}q_xq_t\,dx. \tag{3.26}$$

For the NLS and SG equations, we have shown that the expansion is not uniform but only of algebraic type (i.e., it only has nonuniformities of the form $\varepsilon x^p e^{-x}$), unlike the KdV and MKdV cases. In the next section, we discuss this problem and construct the uniform solution.

4. Uniform solution

In the previous sections, we have developed a direct perturbation method by assuming quasistationarity. Then we have shown that there is more than one region to the solution. In this section, using the perturbed KdV equation (2.1) as an example and using the method of matched asymptotic expansions, for $t\sim\varepsilon^{-1}$ we construct a solution which is uniformly valid for all x. In the first region, $|\theta|\ll\varepsilon^{-1/2}$, the solution is quasistationary; however, in the other region, the solution depends on x and t strongly.

As shown in Sec. 2 for the perturbed KdV and MKdV equations, there are two regions which give rise to nonuniformities. In the region $-\infty<x\ll\varepsilon^{-1}$, that is, behind the soliton, the expansion is nonuniform due to the shelf.

We require that the solution to (4.5) satisfies the matching conditions, i.e. boundary conditions (2.11),

$$q \to \left(8\eta^2 - \frac{2\gamma\eta}{3\varepsilon} Y^2\right)\exp\left(-\frac{2\eta}{\varepsilon} Y\right) \qquad \text{for} \quad 1 \gg Y \sim \varepsilon^{1/2},$$
$$\to 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{as} \quad Y \to \infty. \tag{4.6}$$

Expanding $q = q_0 + \varepsilon q_1 + \ldots$, and $\phi = \phi_0 + \varepsilon\phi_1 + \ldots$, at leading order we have

$$\phi_{0Y}^3 q_{0\theta\theta\theta} + (\phi_{0T} - 4\eta^2\phi_{0Y}) q_{0\theta} = 0. \tag{4.7}$$

Taking (4.6) into account, we obtain the solution,

$$q_0 = 8\eta^2 e^{-K\theta}, \tag{4.8}$$

where $K = \{(4\eta^2\phi_{0Y} - \phi_{0T})/\phi_{0Y}^3\}^{1/2}$. Here ϕ_0 is determined by the secularity condition in the next order equation. Substituting (4.8) into the first order equation, we find that the secularity conditions lead to

$$K_T - 4\eta^2 K_Y + 3\phi_{0Y}^2 K^2 K_Y = 0, \tag{4.9a}$$

$$(K\phi_1)_T + (3\phi_{0Y}^2 K^2 - 4\eta^2)(K\phi_1)_Y = -\frac{\gamma}{3} + 3\phi_{0Y}\phi_{0YY}K^2 + 6\phi_{0Y}^2 K K_x. \tag{4.9b}$$

In order to find the solution of (4.9), it is convenient to take $K = 1$; then we have

$$\phi_{0T} - 4\eta^2\phi_{0Y} + \phi_{0Y}^3 = 0,$$
$$\phi_{1T} - 4\eta^2\phi_{1Y} + 3\phi_{0Y}^2\phi_{1Y} = -\frac{\gamma}{3} + 3\phi_{0Y}\phi_{0YY}. \tag{4.10}$$

Equations (4.10) are similar to Johnson's results [10]. We follow Johnson's work to find approximate solutions of (4.10) for $Y \approx 0$. For this purpose, we assume that the solutions take of the forms

$$\phi_0 = \alpha_1(T)\, Y + \alpha_2(T)\, Y^2 + \cdots, \tag{4.11a}$$

$$\phi_1 = \beta_1(T)\, Y + \beta_2(T)\, Y^2 + \cdots. \tag{4.11b}$$

Under the matching conditions (4.6), we have

$$\alpha_1(T) = 2\eta, \qquad \alpha_2(T) = \frac{\gamma}{12\eta}, \qquad \beta_1(T) = \frac{\gamma}{12\eta^2}, \tag{4.12}$$

and so on.

For $Y \to \infty$, we have a similarity solution. Noting that (4.10a) can be written in the form

$$\psi_\sigma + 3\psi\psi_\nu = 0, \tag{4.13}$$

where $\psi = \phi_0^2 \gamma$, $\sigma = T$, and $\nu = Y + \int_0^T 4\eta^2(T')\,dT'$ [note that (4.13) may be solved exactly], we have the similarity solution to (4.13) or (4.10):

$$\phi_0 = \frac{2}{3\sqrt{3}} \left(\frac{\nu}{\sigma^{1/3}} \right)^{3/2}. \tag{4.14}$$

Thus a description for the uniform solution is

$$q(x,t) \simeq \begin{cases} \hat{q}(x,t), & \text{for} \quad -\theta \gtrsim \varepsilon^{-1/2}, \\ \hat{q}_0(\theta, T) + \varepsilon \hat{q}_1(\theta, T), & \text{for} \quad |\theta| \lesssim \varepsilon^{-1/2}, \\ 8\eta^2 \exp\{\phi_0(Y,T)/\varepsilon\}, & \text{for} \quad \theta \gtrsim \varepsilon^{-1/2}. \end{cases} \tag{4.15}$$

where $\hat{q}_0$, $\hat{q}_1$ are quasistationary solutions given in (2.5), (2.10) respectively, and $\hat{q}$ is given by (4.3). Note that δq in (2.14) (i.e., $q = q_s + \delta q$) is given by $\delta g(x,t) = q(x,t) - \hat{q}_0(\theta, T)$ in (4.15).

For the NLS and SG equations, there is no shelf. Therefore, we can use the latter method in this section and construct the uniform solution.

5. Conservation laws

It is instructive to study the effects of the perturbation from the point of view of the conservation relations. In this section, as an example, we study the perturbed KdV equation and verify the results derived in the previous sections. It should also be noted that the conservation laws were used effectively in [3, 4, 16, 17] to determine the structure of the shelf.

From Eq. (2.1), we have the following conservation relations:

(1) Mass conservation,

$$\frac{d}{dt} \int_{-\infty}^{\infty} q\,dx = -\varepsilon\gamma \int_{-\infty}^{\infty} q\,dx. \tag{5.1}$$

(2) Energy conservation,

$$\frac{d}{dt} \int_{-\infty}^{\infty} q^2\,dx = -2\varepsilon\gamma \int_{-\infty}^{\infty} q^2\,dx. \tag{5.2}$$

(3) Motion of center of mass (momentum conservation),

$$\frac{d}{dt} \int_{-\infty}^{\infty} xq\,dx = 3\int_{-\infty}^{\infty} q^2\,dx - \varepsilon\gamma \int_{-\infty}^{\infty} xq\,dx. \tag{5.3}$$

From the solution, we have the following facts:

$$\int_{-\infty}^{\infty} q_s\, dX = 4\eta, \tag{5.4a}$$

$$\int_{-\infty}^{\infty} \delta q\, dx = 4\eta\left(e^{-\frac{1}{3}\int_0^T \gamma\, dT'} - 1\right) + O(\varepsilon), \tag{5.4b}$$

$$\int_{-\infty}^{\infty} q^2\, dx = \tfrac{16}{3}\eta^3 + O(\varepsilon^2), \tag{5.4c}$$

$$\int_{-\infty}^{\infty} xq\, dx = \frac{16\eta^{3/2}}{\varepsilon}\int_0^T \eta^{3/2}\, dT' + O(1), \tag{5.4d}$$

where $q_s = \hat{q}_0(\theta, T)$ is given by (2.2) and $q = q_s + \delta q$ is given by (4.15). On the other hand, the relation (5.1) gives

$$M = \int_{-\infty}^{\infty} q\, dx = M_0 e^{-\int_0^T \gamma\, dT'}, \tag{5.5}$$

where M expresses the total mass and $M_0 = 4\eta(0)$. Equation (5.5) is equivalent to the results (5.4a) and (5.4b). The relation (5.2) gives the same result as the secularity condition: $\eta_t = -2\varepsilon\gamma\eta/3$, or

$$\frac{d}{dt}E_s = -2\varepsilon\gamma E_s,$$

where $E_s = \frac{16}{3}\eta^3$. Therefore, the changing energy is provided by the soliton part only, or in other words, the interaction energy is canceled by the energy of the shelf part, i.e. (2.15). From (5.4d) we find that the relation (5.3) yields $\theta_t = -4\eta^2$. In order to calculate the next order correction, one should have the solution at order ε^2, say q_2 (which we do not obtain here). Since the solution q_2 may also have a shelf part, the integral $\int_{-\infty}^{\infty} xq_2\, dx$ is order 1. Thus our results are consistent with the conservation relations at leading order.

6. Perturbed higher nonlinear MKdV and NLS equations

In this section, we discuss certain perturbed "higher" nonlinear MKdV and NLS equations as examples of (presumably) nonintegrable systems. We find that if the order of nonlinearity is ≥ 5, the perturbation method suggests that the perturbed solitary wave is undergoing focusing in an analogous manner to the higher nonlinear NLS equation [13].

Let us consider the following "higher" nonlinear MKdV equation:

$$q_t + Aq^n q_x + q_{xxx} = -\varepsilon\gamma q \qquad (n \geq 3), \tag{6.1}$$

in which the unperturbed solitary wave may be written

$$q_0 = \alpha \operatorname{sech}^{2/n} \eta(\theta-\theta_0), \qquad \frac{\partial\theta}{\partial x}=1, \quad \frac{\partial\theta}{\partial t}=-4\frac{\eta^2}{n^2}. \tag{6.2}$$

where α is given by $A\alpha^n = 2(n+1)(n+2)\eta^2/n^2$. By assuming a quasistationary solution, we have

$$-4\frac{\eta^2}{n^2}\hat{q}_\theta + A\hat{q}^n\hat{q}_\theta + \hat{q}_{\theta\theta\theta} = \epsilon\hat{F}(\hat{q}),$$

$$F(\hat{q}) = -\gamma\hat{q} - \hat{q}_T, \tag{6.3}$$

from which, at order ϵ, we obtain

$$\hat{L}\hat{q}_1 \equiv -4\frac{\eta^2}{n^2}\hat{q}_{1\theta} + A(\hat{q}_0^n\hat{q}_1)_\theta + \hat{q}_{1\theta\theta\theta} = \hat{F}_1(\hat{q}_0), \tag{6.4}$$

$$\hat{F}_1(\hat{q}_0) = -\gamma\hat{q}_0 - \frac{1}{\eta}\frac{\partial\eta}{\partial T}\left\{\frac{2}{n}\hat{q}_0 + (\theta-\theta_0)\hat{q}_{0\theta}\right\} + \frac{\partial\theta_0}{\partial T}\hat{q}_{0\theta}.$$

Using the fact that $\hat{L}^A\hat{q}_0 = 0$, the compatibility condition is given by

$$\int_{-\infty}^{\infty} \hat{q}_0\hat{F}_1(\hat{q}_0)\,d\theta = 0, \tag{6.5}$$

which leads to

$$\frac{1}{\eta}\frac{\partial\eta}{\partial T} = -\frac{2n}{4-n}\gamma. \tag{6.6}$$

From (6.4) and (6.6), we find that there is a shelf which is given by

$$\hat{q}_1 \to -\frac{n^3\gamma}{4\eta^2(4-n)}\int_{-\infty}^{\infty}\hat{q}_0\,d\theta \qquad \text{as} \quad \theta\to-\infty.$$

This result [found by reduction of order of (6.4)] is consistent with KdV and MKdV. Uniform results can be obtained following the ideas in Sec. 4.

From (6.6), one can see that the perturbation scheme breaks down at $n=4$. This implies that the assumption of quasistationarity cannot be applied to this problem; that is, the effects of perturbation are not adiabatic. For the case $n \geq 4$, this result suggests that the equation admits of a self-focusing singularity. Whereas we have not proven the existence of the singularity, we can show that the similar situation occurs for "higher" nonlinear NLS equations (where the existence of a

singularity is provable). In this regard, consider a "higher" NLS equation,

$$iq_t + q_{xx} + A|q|^{2n}q = -i\varepsilon\gamma q, \qquad n \geqslant 2, \tag{6.7}$$

which has an unperturbed solitary wave of the form

$$q_0 = \alpha \operatorname{sech}^{1/n} \eta(\theta-\theta_0) \exp i(\sigma-\sigma_0), \tag{6.8}$$

where $A\alpha^{2n} = (n+1)\eta^2/n^2$. Here, for simplicity, we have taken the solitary wave in the rest frame, i.e.

$$\frac{\partial\theta}{\partial x} = 1, \quad \frac{\partial\theta}{\partial t} = 0, \quad \frac{\partial\sigma}{\partial x} = 0, \quad \frac{\partial\sigma}{\partial t} = \frac{\eta^2}{n^2}. \tag{6.9}$$

Under the assumption of quasistationarity for the solution, $q = \hat{q}(\theta, T, \varepsilon) \exp i(\sigma-\sigma_0)$, we obtain

$$-\frac{\eta^2}{n^2}\hat{q} + \hat{q}_{\theta\theta} + A|\hat{q}|^{2n}\hat{q} = \varepsilon\hat{F}(\hat{q}), \tag{6.10}$$

$$\hat{F}(\hat{q}) = -i\gamma\hat{q} - i\hat{q}_T - \sigma_{0T}\hat{q}.$$

At order ε, we have

$$-\frac{\eta^2}{n^2}\hat{q}_1 + \hat{q}_{1\theta\theta} + (n+1)A\hat{q}_0^{2n}\hat{q}_1 + nA\hat{q}_0^{2n}\hat{q}_1^* = \hat{F}_1(\hat{q}_0), \tag{6.11}$$

in which, setting $\hat{q}_1 = \hat{\phi}_1 + i\hat{\psi}_1$, we obtain

$$-\frac{\eta^2}{n^2}\hat{\phi}_1 + \hat{\phi}_{1\theta\theta} + A(2n+1)\hat{q}_0^{2n}\hat{\phi}_1 = \operatorname{Re}\hat{F}(\hat{q}_0), \tag{6.12a}$$

$$-\frac{\eta^2}{n^2}\hat{\psi}_1 + \hat{\psi}_{1\theta\theta} + A\hat{q}_0^{2n}\hat{\psi}_1 = \operatorname{Im}\hat{F}(\hat{q}_0). \tag{6.12b}$$

The compatibility condition

$$\int_{-\infty}^{\infty} \hat{q}_0 \operatorname{Im}\hat{F}(\hat{q}_0)\, d\theta = 0 \tag{6.13}$$

gives

$$\frac{1}{\eta}\frac{\partial\eta}{\partial T} = -\frac{2n}{2-n}\gamma. \tag{6.14}$$

At $n=2$, the perturbation scheme breaks down. Thus if the order of nonlinearity is greater than 5, the perturbation changes the solitary wave drastically. However, this effect is not really due to the perturbation. Rather it is inherent in the equation itself. By using the conservation laws, one can show that the equation (6.7) has a focusing singularity.

Consider the evolution of the following quantity [13] (moment of inertia):

$$J = \int_{-\infty}^{\infty} x^2 |q|^2 \, dx. \tag{6.15}$$

Then, from Eq. (6.7), we obtain

$$\frac{d^2J}{dt^2} + 4\varepsilon\gamma \frac{dJ}{dt} + 4\varepsilon^2\gamma^2 J = 8 \int \left\{ |q_x|^2 - \frac{nA}{2(n+1)} |q|^{2n+2} \right\} dx. \tag{6.16}$$

For $n=2$, this becomes

$$\frac{d^2J}{dt^2} = 8I_3 + O(\varepsilon), \tag{6.17}$$

where I_3 is one of the conserved quantities (if $\gamma=0$),

$$I_3 = \int \left\{ |q_x|^2 - \frac{A}{3} |q|^6 \right\} dx. \tag{6.18}$$

This implies that if $I_3 < 0$, J goes to zero at finite time, and the equation has a focusing singularity. No such argument has yet been given for (6.1).

7. Discussion

In this section we compare the method considered here with the other approaches in the literature. As a concrete example we discuss the results for the perturbed KdV equation (2.1).

In the existing literature a direct perturbation method has been applied to the evolution of solitons in slowly varying media, e.g. [10], [11]. However, in neither of these papers is a specific initial value problem considered. Indeed, [11] makes the claim that [10] is mistaken due to neglect of a specific phase. Our results demonstrate that the phase term θ_0 is fixed only by specific initial values. Any other choices are *ad hoc*, since addition of the term θ_0 to the phase *does not* lead to a secularity in the real field.

An alternative approach for problems admitting solitons (*not* solitary waves) is to employ methods of inverse scattering. In [3], [4] the time dependent o.d.e.'s governing the scattering data are used. (In [5] a Green's function formulation is developed.) In the presence of a small perturbation $F(q)$, where the equation takes the form

$$q_t = K(q) + \varepsilon F(q), \tag{7.1}$$

crudely speaking, the scattering data evolve according to

$$\frac{dS}{dt}=\int_{-\infty}^{\infty}\frac{\delta S}{\delta q}\left[K(q)+\varepsilon F(q)\right]dx, \tag{7.2}$$

where S represents some portion of the scattering data (e.g. the reflection coefficient) and $\delta S/\delta q$ is the functional derivative of S with respect to q. In (7.2), one can apply straightforward perturbation methods to remove time secularities. The perturbed soliton is found via the inverse scattering equations. It is interesting to note that this method, while removing all time secularities, does not remove the space secularities (see Sec. 4). Apparently, for certain regions (large x) the Gel'fand-Levitan integral equation produces convergent but nonasymptotic expansion using the simple iteration of the integral equation. It should be noted that in [3] and [4], they considered the perturbed equations for the range of time $1\ll t\ll\varepsilon^{-1}$. For this range of time, our result, which becomes $\theta_{0T}=-\gamma/3\eta$, agrees with the result in [4]. This is in disagreement with the result in [3]. The error in [3] stems from the fact that they assume a certain form for the scattering function $a(\zeta)$; namely, $a(\zeta)=(\zeta-i\eta_1)/(\zeta+i\eta_1)$, where η_1 is the location of the eigenvalue at the initial instant. It turns out that after the initial time this formula must be modified [for long time $a(\zeta)$ may be quite complicated]. In any event, the perturbation method using inverse scattering is fairly sophisticated and requires detailed knowledge of I.S.T.

We also considered the method of Keener and McLaughlin [5]. Unfortunately, a naive application of their ideas is not completely satisfactory for the KdV equation. If we simply follow in a completely analogous manner their work on the NLS equation (for which our results are consistent with theirs), the appropriate secularity condition is

$$\int_{-\infty}^{\infty}v\cdot\hat{F}\,dx=0, \tag{7.3}$$

where $\hat{F}$ is given by (2.6) and v satisfies the adjoint problem of the linearized KdV equation,

$$v_t-4\eta^2v_\theta+6q_0v_\theta+v_{\theta\theta\theta}=0, \tag{7.4}$$

with $q_0=2\eta^2\operatorname{sech}^2\eta(\theta-\theta_0)$. We have two solutions of (7.4):

$$\begin{aligned}v_1&=q_0,\\ v_2&=-1+\tanh\phi+\left(\tfrac{1}{2}+\phi-8\eta^2t\right)\operatorname{sech}^2\phi,\end{aligned} \tag{7.5}$$

where $\phi=\eta(\theta-\theta_0)$. We note that $v_2\to$constant $(\neq 0)$ as $\theta\to-\infty$; hence, it is not a proper solution in the sense of (2.7). Equation (7.3) using $v=v_1$ is the same as (2.8), and gives the evolution equation of η, i.e. (2.9). However, the second equation of (7.3), using $v=v_2$, is not satisfied. We can see this by writing

$$\int_0^t dt'\int_{-\infty}^{\infty}v\cdot\hat{F}\,dx=\int_{-\infty}^{\infty}v\cdot q_t\big|_{t'=t}dx, \tag{7.6}$$

where q_1 is the perturbation of the soliton and we've required $q_1(x,0)=0$. From (7.6), (7.3), we must have

$$\int_{-\infty}^{\infty} v_2 q_1 \, dx = 0. \tag{7.7}$$

Since q_1 generates a constant shelf whose area increases with time t, the left hand side of (7.7) is proportional to t. This implies that (7.7) cannot be satisfied. We conclude, of course, that the method of Keener and McLaughlin cannot be applied by directly substituting into formulas analogous to the NLS problem. For the KdV perturbation problem, the Green's function is, in a sense, improper, and the condition (7.3) must be modified due to contributions of the continuous spectrum. McLaughlin has also noted similar behavior [14].

Finally, we mention that the direct perturbation method can also be applied to more complicated solutions, such as breathers. We shall report on this work in a future communication.

Appendix

Here we give the calculation to solve for $\hat{q}_1$ of (2.6). Setting $y=\tanh\eta(\theta-\theta_0)$, Eq. (2.6) can be written

$$\tilde{L}\hat{q}_1 = \frac{d}{dy}(1-y^2)\frac{d}{dy}\hat{q}_1 + \left(12-\frac{4}{1-y^2}\right)\hat{q}_1 = \tilde{F}_1. \tag{A.1}$$

where $\tilde{F}$ is given by

$$\tilde{F} = \frac{2\gamma}{3\eta}\frac{1}{1+y} + \frac{2\gamma}{3\eta}\log\frac{1+y}{1-y} + 2\frac{\partial\theta_0}{\partial T}.$$

Taking into account that $\tilde{L}v=0$ is a Legendre equation, we have $v=P_3^2(y)=15y(1-y^2)$ as a proper solution. By using the variation of constant method, i.e.

$$\hat{q}_1(y) = A(y)P_3^2(y), \tag{A.2}$$

we obtain the equation for $B(y)=dA/dy$,

$$\frac{dB}{dy} + \frac{2(1-4y^2)}{y(1-y^2)}B = G, \tag{A.3}$$

$$G = \frac{2\gamma}{45\eta}\frac{1-y}{y(1-y^2)^3} + \frac{2\gamma}{45\eta}\frac{1}{y(1-y^2)^2}\log\frac{1+y}{1-y}$$

$$+\frac{2}{15}\frac{\partial\theta_0}{\partial T}\frac{1}{y(1-y^2)^2}.$$

Since this is just a first order ordinary differential equation, we easily obtain the solution

$$\hat{q}_1 = \frac{\gamma}{6\eta}\left[-1+y+\frac{3}{2}\left(1+\frac{\eta}{\gamma}\frac{\partial\theta_0}{\partial T}\right)(1-y^2)\left(2-y\log\frac{1+y}{1-y}\right)\right.$$
$$\left.+(1-y^2)\left(1-\tfrac{1}{2}y\log\frac{1+y}{1-y}\right)\log\frac{1+y}{1-y}\right]. \qquad (A.4)$$

Acknowledgments

This research was partially supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR 78-3674, and the National Science Foundation under Grant No. MCS-7903998.

References

1. M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, *Studies in Appl. Math.* 53:249 (1974).
2. M. J. Ablowitz, *Studies in Appl. Math.* 58:17 (1978).
3. D. J. Kaup and A. C. Newell, *Proc. Roy. Soc. Ser. A* 361:413 (1978).
4. V. I. Karpman and E. M. Maslov, *Ž. Ėkser. Theoret. Fiz.* 75:504 (1978).
5. J. P. Keener and D. W. McLaughlin, *Phys. Rev. A* 16:777 (1977).
6. M. J. Ablowitz, *Studies in Appl. Math.* 50:329 (1971).
7. G. B. Whitham, *Linear and Nonlinear Waves*, Wiley-Interscience, 1974.
8. V. E. Zakharov and A. M. Rubenchik, *Soviet Physics JETP* 38:494 (1974).
9. B. Kadomtsev and V. Petviashvili, *Soviet Physics Dokl.* 15:539 (1970).
10. R. S. Johnson, *J. Fluid Mech.* 60:813 (1973).
11. K. Ko and H. H. Kuehl, *Phys. Rev. Lett.* 40:223 (1978).
12. Y. Kodama and T. Taniuti, *J. Phys. Soc. Japan* 45:298 (1978).
13. V. E. Zakharov and V. S. Synakh, *Soviet Physics JETP* 41:465 (1976).
14. D. W. McLaughlin, private communication.
15. S. Leibovich and G. D. Randall, *J. Fluid Mech.* 53:481 (1973).
16. A. C. Newell, *Solitons and Condensed Matter Physics* (Bishop and Schneider, Eds.), Springer, 1978, p. 52.
17. C. J. Knickerbocker and A. C. Newell, preprint Shelves and Korteweg-deVries equation, *J. Fluid Mech.* to appear.
18. Y. Kodama and M. J. Ablowitz, *J. Math. Phys.* 21:928 (1980).

Clarkson College of Technology

(Received August 20, 1979)

The Periodic Cubic Schrödinger Equation

By Yan-Chow Ma and Mark J. Ablowitz

Solutions to the cubic nonlinear Schrödinger equation, periodic in space, are developed.

Introduction

The development of the inverse scattering transform (I.S.T.) has made it possible to solve certain physically significant nonlinear evolution equations [1]. The original ideas were applied to the Korteweg-deVries equation on the infinite interval with rapidly decaying boundary conditions [2]. The solution is obtained via a linear integral equation. A second equation of physical interest solvable by I.S.T. was discovered by Zakharov and Shabat [3]. They studied the cubic nonlinear Schrödinger equation on the infinite interval. This work suggested that indeed many equations were indeed of I.S.T. type [1].

Recently there have been significant developments regarding solutions to these equations with periodic boundary conditions. For the Korteweg-deVries equation References [4–8] develop many of the main ideas. A review of some of this work appears in Reference [9]. In this paper we examine the cubic nonlinear Schrödinger equation

$$iq_t + q_{xx} \mp 2|q|^2 q = 0, \qquad (\mathrm{I}\pm)$$

where $q(x, t)$ is periodic in x with period T for all time $q(x, t) = q(x+T, t)$. We follow closely the analysis in [4, 5]. In the defocusing problem [$-$ sign in (I$\pm$)] we find the analysis is close to that of KdV, save for the need of an additional set of scattering data. In the focusing problem [$+$ sign in (I$\pm$)] we find significant differences from the KdV theory. These differences are due to the fact that the

Address for correspondence: Professor Mark J. Ablowitz, Department of Mathematics, Clarkson College of Technology, Potsdam, NY 13676.

STUDIES IN APPLIED MATHEMATICS 65:113–158 (1981)
Copyright © 1981 by the Massachusetts Institute of Technology
Published by Elsevier North Holland, Inc.

0022-2526/81/050113+46$02.50

"nice" spectral properties for the defocusing case no longer hold. In the latter case we also discuss the question of stability. Although it is clear that the analysis presented for the periodic defocusing cubic Schrödinger equation can be readily extended to C^∞ initial values, this question (to our knowledge) still remains open in the focusing case.

Our approach here will be to review many of the ideas in [4–8] and to use elementary methods and analysis. We shall attempt to spell out the basic ideas in the clearest possible way. One of our aims is simply to isolate the key differences between the defocusing and focusing problems. These differences are the reasons why the focusing problem has still not been completely solved. The paper is divided into two parts: Part I on the defocusing case, and Part II on the focusing case. Much of this work is contained in Reference [14].

I. The periodic defocusing cubic Schrödinger equation

1. The direct scattering problem

The associated scattering equation for the defocusing cubic Schrödinger equation is

$$\frac{\partial v_1}{\partial x} + i\zeta v_1 = q v_2, \tag{1.1.a}$$

$$\frac{\partial v_2}{\partial x} - i\zeta v_2 = q^* v_1. \tag{1.1.b}$$

It is shown in [3] that if

$$v(x,\zeta) = \begin{pmatrix} v_1(x,\zeta) \\ v_2(x,\zeta) \end{pmatrix}$$

is a solution of (1.1), so is

$$\bar{v}(x,\zeta) = \begin{pmatrix} v_2^*(x,\zeta^*) \\ v_1^*(x,\zeta^*) \end{pmatrix}.$$

The two solutions are independent if their Wronskian is nonzero. We therefore fix two independent eigenfunction bases:

$$\phi(x,\zeta) = \begin{pmatrix} \phi_1(x,\zeta) \\ \phi_2(x,\zeta) \end{pmatrix}, \qquad \bar{\phi}(x,\zeta) = \begin{pmatrix} \phi_2^*(x,\zeta^*) \\ \phi_1^*(x,\zeta^*) \end{pmatrix}, \tag{1.2.a}$$

where

$$\phi(x_0,\zeta) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{1.2.b}$$

$\phi(x+T,\zeta)$ is also a solution of (1.1) due to the periodicity of the potential q, and can be expanded in terms of $\phi(x,\zeta)$ and $\bar{\phi}(x,\zeta)$:

$$\phi(x+T,\zeta) = a\phi(x,\zeta) + b\bar{\phi}(x,\zeta). \tag{1.3}$$

Evaluating x at x_0, we obtain

$$a(\zeta,x_0) = \phi_1(x_0+T,\zeta), \tag{1.4.a}$$

$$b(\zeta,x_0) = \phi_2(x_0+T,\zeta). \tag{1.4.b}$$

Using (1.1) (the Wronskian relation) and (1.4), we obtain the following relation between a and b:

$$a(\zeta,x_0)a^*(\zeta^*,x_0) - b(\zeta,x_0)b^*(\zeta^*,x_0) = 1. \tag{1.5}$$

The Bloch eigenfunction

$$\psi(x,\zeta) = \begin{pmatrix} \psi_1(x,\zeta) \\ \psi_2(x,\zeta) \end{pmatrix}$$

is defined as

$$\psi(x+T,\zeta) = \lambda\psi(x,\zeta). \tag{1.6.a}$$

Expanding ψ in terms of $\phi, \bar{\phi}$:

$$\psi = c\phi + d\bar{\phi}, \tag{1.6.b}$$

and using (1.6.a) at $x=x_0$, we find

$$c(a-\lambda) + db^* = 0 \tag{1.6.c}$$

and

$$cb + d(a^*-\lambda) = 0. \tag{1.6.d}$$

The condition for nontrivial solutions c, d is [using (1.5)]

$$\lambda^2 - 2a_R\lambda + 1 = 0. \tag{1.6.e}$$

For real ζ we have the following cases:

(1) $a_R^2 > 1$: λ is real and the two Bloch eigenfunctions are unstable.
(2) $a_R^2 < 1$: λ is on the unit circle ($\lambda = e^{i\theta}$) and the two Bloch eigenfunctions are stable.

(3) $a_R^2=1$: $\lambda=\pm 1$ and at least one of the Bloch eigenfunctions is periodic or antiperiodic.

For convenience, in the analysis that follows we separate a, b, v_1, and v_2 into real and imaginary parts:

$$a = a_R + ia_I, \qquad b = b_R + ib_I, \tag{1.7.a}$$

$$v_i = v_{iR} + iv_{iI}, \qquad i = 1,2. \tag{1.7.b}$$

a_R, a_I, b_R, v_{iR}, and v_{iI} are real when ζ is real.

Next, we look at the spectra of (1.1). We choose the following three spectra, which will simplify the reconstruction of q.

(i) *The main spectrum*: The main spectrum is composed of the eigenvalues ζ_i' at least one of whose eigenfunctions is periodic or antiperiodic. From the above we see that the ζ_i are the roots of $1-a_R^2=0$:

$$1 - a_R^2(\zeta_i') = 0 \tag{1.8.a}$$

We define the stable band to be the line segment between any two adjacent ζ_i' such that $1-a_R^2 \geqslant 0$ for real ζ. Likewise, we define the unstable band to be the line segment between any two adjacent ζ_i' such that $1-a_R^2 \leqslant 0$ for real ζ. ζ_i' is named the band edge. A typical function $a_R=a_R(\zeta)$ is given in Figure 1.

(ii) *auxiliary spectrum* (I): This is composed of the eigenvalues γ_i' such that

$$\phi_{1I}(x_0+T,\gamma_i') + \phi_{2I}(x_0+T,\gamma_i') = 0.$$

From the definitions of ϕ, $\bar{\phi}$, a, and b we see that the γ_i' correspond to the roots of $a_I+b_I=0$:

$$a_I(\gamma_i') + b_I(\gamma_i') = 0. \tag{1.8.b}$$

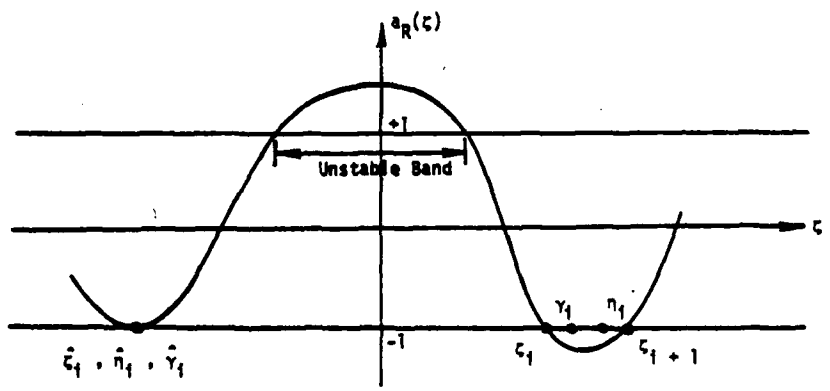


Figure 1.

Note that from (1.5) this implies that when $\zeta=\gamma_i'$, $1-a_R^2=-b_R^2\leqslant 0$. Hence the γ_i' lie inside the unstable band.

(iii) *Auxiliary spectrum* (II): This is composed of the eigenvalues η_i' such that

$$\phi_{1I}(x_0+T,\eta_i')+\phi_{2R}(x_0+T,\eta_i')=0.$$

η_i' are the roots of $a_I+b_R=0$:

$$a_I(\eta_i')+b_R(\eta_i')=0. \tag{1.8.c}$$

At $\zeta=\eta_i'$, $1-a_R^2=-b_I^2\leqslant 0$ and the η_i' lie within the unstable band.

There are several properties associated with the above three spectra. The proofs can be found in the appendix.

(1) The main spectrum ζ_i' ($i=1,2,\ldots.\infty$) is real and can be divided into what we shall call nondegenerate band edges (ζ_i) and degenerate band edges ($\hat{\zeta}_i$). The property

$$\frac{\partial a_R}{\partial \zeta}(\zeta_i)\neq 0$$

holds at the nondegenerate band edges, whereas

$$\frac{\partial a_R}{\partial \zeta}(\hat{\zeta}_i)=0$$

at the degenerate band edges. ζ_i' will be either a simple or a double root of $1-a_R^2=0$, but not of higher order.

(2) The auxiliary spectra γ_i' and η_i' ($i=1,2,\ldots,\infty$) are real and must lie inside the unstable bands or on the band edges. All the γ_i' and η_i' are simple roots. At any $\hat{\zeta}_i$ we have $a_I=b_R=b_I=0$, and we can therefore split γ_i' and η_i' into γ_i, $\hat{\gamma}_i$ and η_i, $\hat{\eta}_i$. The values of $\hat{\gamma}_i$ and $\hat{\eta}_i$ coincide with $\hat{\zeta}_i$, whereas γ_i and η_i lie in the nondegenerate unstable band. There will be only one γ_i and one η_i in each unstable band.

(3) $a_R'(\zeta)\neq 0$ for $|a_R(\zeta)|<1$. $a(\zeta)$ and $b(\zeta)$ are entire functions of ζ. [If $q(x,0)$ is real, we can show that $a(\zeta)=a^*(-\zeta)$. Thus in this case the main spectrum will appear in positive and negative pairs.] The above properties are summarized by Figure 1.

2. *N-band potential*

An arbitrary periodic potential has an infinite number of ζ_i. Here we restrict our initial conditions in such a way that we will have only a finite number ($2N$) of nondegenerate band edges ζ_i with all the other band edges being degenerate ($\hat{\zeta}_i$). This type of potential is called an N-band potential. Following [4], we now prove that any periodic solution of $\sum_{m=0}^{N}C_m\,\delta I_m/\delta q^*=0$ will be at most an N-band potential (C_m are real) of the equations (1.1), where I_m ($m=0,1,2\ldots,N$) are the

conserved quantities of the cubic Schrödinger equation:

$$I_0 = -\int_{x_0}^{x_0+T} |q|^2\,dx, \tag{2.1.a}$$

$$I_1 = \frac{i}{2}\int_{x_0}^{x_0+T} (q^* q_x - q q_x^*)\,dx, \tag{2.1.b}$$

$$I_2 = \int_{x_0}^{x_0+T} \left(|q_x|^2 + |q|^4\right) dx, \tag{2.1.c}$$

etc.

First we note that the evolution equations which can be solved exactly by the inverse scattering technique and have (1.1) as their scattering equations are

$$iq_t = \frac{\delta I_m}{\delta q^*}, \qquad m = 0,1,2,\ldots, \tag{2.2}$$

where $m=2$ corresponds to the cubic Schrödinger equation.

Next, choose the time dependence of the base eigenfunctions to be

$$\frac{\partial \phi_1}{\partial t} = A\phi_1 + B\phi_2 + \lambda\phi_1 + \mu\phi_2^*(\zeta^*), \tag{2.3.a}$$

$$\frac{\partial \phi_2}{\partial t} = C\phi_1 + D\phi_2 + \lambda\phi_1 + \mu\phi_1^*(\zeta^*). \tag{2.3.b}$$

To keep $\phi_1(x_0, t, \zeta)=1$ and $\phi_2(x_0, t, \zeta)=0$ for all time we take

$$\lambda = -A(x_0, t, \zeta), \tag{2.4.a}$$

$$\mu = -C(x_0, t, \zeta), \tag{2.4.b}$$

where λ and μ are independent of x.

The consistency of (2.3) and (1.1) gives us the evolution equation as well as $D(\zeta)=A^*(\zeta^*)=-A(\zeta)$, $C(\zeta)=B^*(\zeta^*)$. Thus, it is obvious that $A(\zeta)$ and $\lambda(\zeta)$ are imaginary whenever ζ is real. For each m we have the corresponding A_m and B_m, where A_m is a polynomial in ζ of order m and B_m is a polynomial in ζ of order $m-1$ [1]. Both have coefficients which depend on q and its derivatives. We list the first few A_m and B_m:

m	Evolution equation	A_m	B_m
0	$iq_t = -q$	$-i/2$	0
1	$q_t = q_x$	$-i\zeta$	q
2	$iq_t + q_{xx} - 2\lvert q\rvert^2 q = 0$	$-2i^2 - i\lvert q\rvert^2$	$iq_x + 2\zeta q$

Evaluating the equations (2.3) at $x = x_0 + T$, we get

$$\frac{\partial a}{\partial t} = \mu(\zeta) b^*(\zeta^*) - \mu^*(\zeta^*) b(\zeta), \tag{2.5.a}$$

$$\frac{\partial b}{\partial t} = -\mu(\zeta) a(\zeta) + \mu(\zeta) a^*(\zeta^*) + 2\lambda(\zeta) b(\zeta). \tag{2.5.b}$$

These results can be written

$$\frac{\partial \hat{T}}{\partial t} = [\Lambda, \hat{T}], \tag{2.6.a}$$

where

$$\hat{T} = \begin{pmatrix} a(\zeta) & b(\zeta) \\ b^*(\zeta^*) & a^*(\zeta^*) \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} \lambda(\zeta) & \mu(\zeta) \\ \mu^*(\zeta^*) & \lambda^*(\zeta^*) \end{pmatrix}. \tag{2.6.b}$$

Trace $\Lambda = 0$, since λ is purely imaginary when ζ is real.

When we superpose those equations in (2.2), we obtain the following equation, which also has (1.1) as its scattering equations:

$$iq_t = \sum_{m=0}^{N} C_m \frac{\delta I_m}{\delta q^*}. \tag{2.7}$$

The corresponding A, B, λ, and μ for (2.7) are

$$A = \sum_{m=0}^{N} C_m A_m, \qquad B = \sum_{m=0}^{N} C_m B_m, \tag{2.8.a}$$

$$\lambda = -A(x_0, t, \zeta), \qquad \mu = -B^*(x_0, t, \zeta^*). \tag{2.8.b}$$

If $q(x,0)$ is periodic in x and satisfies $\Sigma_{m=0}^{N} C_m \delta I_m / \delta q^* = 0$; then q will be stationary if it evolves according to (2.7). Hence, $q(x,t) = q(x,0)$, $a(\zeta,t) = a(\zeta,0)$, $b(\zeta,t) = b(\zeta,0)$, and from (2.6)

$$[\Lambda, \hat{T}] = [\Lambda, \hat{T}]|_{t=0} = 0. \tag{2.9}$$

Note here, the $a(\zeta,t)$ and $b(\zeta,t)$ are for (2.7) and *not* for the cubic Schröedinger equation. In general $a(\zeta,t)$ and $b(\zeta,t)$ will be different for different evolution equations because of the dispersion relation [1]. Still, $a(\zeta,0)$ and $b(\zeta,0)$ will be the same for all equations in (2.2) as well as for (2.7). The reason is that we obtain $a(\zeta,0)$ and $b(\zeta,0)$ from (1.1), which makes them independent of the evolution equation and only dependent on $q(x,0)$. Thus, $[\Lambda, \hat{T}]|_{t=0} = 0$ is also true for the cubic Schrödinger equation.

From $[\Lambda, \hat{T}]|_{t=0}=0$ we get

$$-4a_I^2(\zeta)\mu(\zeta)\mu^*(\zeta^*)+4\lambda(\zeta)\lambda^*(\zeta^*)b(\zeta)b^*(\zeta^*)=0. \tag{2.10}$$

It follows that the roots which correspond to $a_I^2(\zeta)=b(\zeta)b^*(\zeta^*)\neq 0$ are the roots of $\mu(\zeta)\mu^*(\zeta^*)=\lambda(\zeta)\lambda^*(\zeta^*)$. But the roots of $a_I^2(\zeta)=b(\zeta)b^*(\zeta^*)\neq 0$ are also the roots of $1-a_R^2=0$ (with $\partial a_R/\partial\zeta\neq 0$ at the roots), which are the band edges. Thus, the band edges are the roots of $\lambda(\zeta)\lambda^*(\zeta^*)=\mu(\zeta)\mu^*(\zeta^*)$. From (3.8.b) we see that λ is a polynomial in ζ of order N, and μ is a polynomial in ζ of order $N-1$. $\lambda(\zeta)\lambda^*(\zeta^*)-\mu(\zeta)\mu^*(\zeta^*)=0$ is a polynomial equation in ζ of order $2N$, and hence there are at most $2N$ simple roots of $1-a_R^2=0$. This proves that the potential is at most an N-band potential. Using (2.9), we may also show that γ_i are the roots of the polynomial equation $-i\lambda+\mu_I=0$ and η_i are the roots of $-i\lambda+\mu_R=0$ where $\mu=\mu_R+i\mu_I$. Dubrovin [13] showed that any N-band potential of the Hill equation satisfies $\sum_{m=0}^{N}C_m\,\delta I_m/\delta q=0$ (I_m are the conserved quantities of the KdV equation) with some C_m. The proof for the cubic Schrödinger equation is similar and hence omitted. We conclude that $\sum_{m=0}^{N}C_m\,\delta I_m/\delta q^*=0$ gives us the most general N-band potential. Hereafter we concentrate on the N-band potentials.

3. *The inverse scattering formulas*

Next we find the scattering data as $|\zeta|\to\infty$ using a WKB approach on $\phi(x,\zeta)$ as $|\zeta|\to\infty$, and applying (1.4) we find

$$a_R=\cos\zeta T+\frac{\sin\zeta T}{2\zeta}\int_{x_0}^{x_0+T}|q|^2\,dx+O\left(\frac{1}{\zeta^2}\right), \tag{3.1.a}$$

$$a_I=-\sin\zeta T+\frac{\cos T}{2\zeta}\int_{x_0}^{x_0+T}|q|^2\,dx+O\left(\frac{1}{\zeta^2}\right), \tag{3.1.b}$$

$$b_R=\frac{q_R(x_0)}{\zeta}\sin\zeta T+O\left(\frac{1}{\zeta^2}\right), \tag{3.1.c}$$

$$b_I=\frac{-q_I(x_0)}{\zeta}\sin\zeta T+O\left(\frac{1}{\zeta^2}\right); \tag{3.1.d}$$

hence,

$$\frac{(a_I+b_I)^2}{1-a_R^2}\sim 1. \tag{3.1.e}$$

Since a_R, a_I, b_R, and b_I are entire functions, the ratio $(a_I+b_I)^2/(1-a_R^2)$ can be

expressed as the ratio of their roots; i.e.,

$$\frac{(a_I+b_I)^2}{1-a_R^2}=\frac{\prod_{i=1}^{N}(\zeta-\gamma_i)^2}{\prod_{i=1}^{2N}(\zeta-\zeta_i)}g(\zeta), \tag{3.2.a}$$

where $g(\zeta)$ is entire with no roots. Using (3.1.e), $g(\zeta)=1$. Similarly,

$$\frac{(a_I+b_R)^2}{1-a_R^2}=\frac{\prod_{i=1}^{N}(\zeta-\eta_i)^2}{\prod_{i=1}^{2N}(\zeta-\zeta_i)}. \tag{3.2.b}$$

The inverse scattering formula allows us to reconstruct the potential from the scattering data. We expand Equation (3.2) for $|\zeta|\to\infty$ and make use of Equation (3.1) in order to compare the coefficients of the $O(1/\zeta)$ term. We get

$$q_I=\frac{1}{2}\sum_{i=1}^{2N}\zeta_i-\sum_{i=1}^{N}\gamma_i, \tag{3.3.a}$$

$$q_R=-\frac{1}{2}\sum_{i=1}^{2N}\zeta_i+\sum_{i=1}^{N}\eta_i. \tag{3.3.b}$$

The above equations can be used to reconstruct the potential at some point x_0. Using the idea of Dubrovin and Novikov [5], we note that the point x_0 is arbitrary and hence (3.3) can be used to reconstruct the potential at any point. For this reason we develop the equations of ζ_i, γ_i, and η_i with respect to x_0.

Changing the point x_0 to x_0+dx_0, we have $\phi(x, x_0+dx_0, \zeta)$; since $\phi(x, x_0+dx_0, \zeta)$ is a solution of (1.1), it can be expanded in terms of $\phi(x, x_0, \zeta)$ and $\bar{\phi}(x, x_0, \zeta)$ (with the coefficients of expansion independent of x). We write this (using Taylor's theorem) as

$$\begin{pmatrix}\phi_1(x, x_0+dx_0, \zeta) & \phi_2(x, x_0+dx_0, \zeta)\\ \phi_2^*(x, x+dx_0, \zeta^*) & \phi_1^*(x, x_0+dx_0, \zeta^*)\end{pmatrix}$$

$$=[1+Q(x_0)dx_0]\begin{pmatrix}\phi_1(x, x_0, \zeta) & \phi_2(x, x_0, \zeta)\\ \phi_2^*(x, x_0, \zeta^*) & \phi_1^*(x, x_0, \zeta^*)\end{pmatrix}, \tag{3.4.a}$$

where

$$Q(x_0)=\begin{pmatrix}Q_{11}(x_0) & Q_{12}(x_0)\\ Q_{21}(x_0) & Q_{22}(x_0)\end{pmatrix}, \quad Q_{12}^*=Q_{21}, \quad Q_{22}^*=Q_{11}. \tag{3.4.b}$$

Replacing x in (3.4.a) by $x+T$ and using (1.3) and the definition (2.6.b) for $\hat{T}$, we obtain

$$\frac{\partial \hat{T}}{\partial x_0} = Q\hat{T} - \hat{T}Q \tag{3.5}$$

(i.e., the scattering data are obtained from the base at $x=x_0$). Since $\phi_1(x_0, x_0, \zeta) = 1$ and $\phi_2(x_0, x_0, \zeta)=0$ for any x_0, we see that

$$\frac{\partial}{\partial x_0}\phi_1(x, x_0, \zeta)\Big|_{x=x_0} = -\frac{\partial}{\partial x}\phi_1(x, x_0, \zeta)\Big|_{x=x_0}, \tag{3.6.a}$$

$$\frac{\partial}{\partial x_0}\phi_2(x, x_0, \zeta)\Big|_{x=x_0} = -\frac{\partial}{\partial x}\phi_2(x, x_0, \zeta)\Big|_{x=x_0}. \tag{3.6.b}$$

Using (3.6) and (1.1) we obtain

$$Q(x_0) = \begin{pmatrix} i\zeta & -q^*(x_0) \\ -q(x_0) & -i\zeta \end{pmatrix}. \tag{3.7}$$

Hence Equation (3.5) can be written

$$\frac{d}{dx_0}\begin{pmatrix} a(\zeta) & b(\zeta) \\ b^*(\zeta^*) & a^*(\zeta^*) \end{pmatrix} = \begin{pmatrix} 2i(q_R b_I + q_I b_R) & -2i\zeta b(\zeta) + 2iq^* a_I \\ -2i\zeta b^*(\zeta^*) - 2iqa_I & -2i(q_R b_I + q_I b_R) \end{pmatrix}. \tag{3.8}$$

From (3.8) we get $(d/dx_0)a_R = 0$, which means that ζ_i $(i=1,2,\ldots,2N)$ are independent of x_0. This is not true for γ_i and η_i.

Taking the square root of Equation (3.2.a), we see that

$$\frac{a_I + b_I}{\sqrt{-1+a_R^2}} = \frac{\sigma_i' \prod_{i=1}^{N}(\zeta-\gamma_i)}{\sqrt{-\prod_{i=1}^{2N}(\zeta-\zeta_i)}}, \tag{3.9}$$

where the square root is defined so that the real part of it is positive and $\sigma' = \pm 1$. Using (3.8) and evaluating ζ at γ_j, we get

$$\frac{d}{dx_0}(a_I + b_I)\Big|_{\zeta=\gamma_j} = 2\sigma_j''\sqrt{a_R^2(\gamma_j)-1}\,(q_I + \gamma_j), \tag{3.10}$$

with $\sigma_j'' = \pm 1$. Combining (3.10) and (3.3), we obtain the differential equations of

γ_j with respect to x_0:

$$\frac{d\gamma_j}{dx_0} = \frac{2\sigma_j\sqrt{-\prod_{i=1}^{2N}(\gamma_j-\zeta_i)}}{\prod_{k\neq j}(\gamma_j-\gamma_k)}\left(\sum_{k\neq j}\gamma_k - \frac{1}{2}\sum_{i=1}^{2N}\zeta_i\right), \tag{3.11.a}$$

$i=1,2,\ldots,N$, $\sigma_j=\pm 1$. Similarly the above analysis yields the equations for η_j:

$$\frac{d\eta_j}{dx_0} = \frac{-2\hat{\sigma}_j\prod_{i=1}^{2N}(\eta_j-\zeta_i)}{\prod_{k\neq j}(\eta_j-\eta_k)}\left(\sum_{k\neq j}\eta_k - \frac{1}{2}\sum_{i=1}^{2N}\zeta_i\right), \tag{3.11.b}$$

$i=1,2,\ldots,N$, $\hat{\sigma}_j=\pm 1$.

The equations (3.11) give us the motion of γ_j and η_j with respect to x_0, which in turn determines q_l and q_R for all x_0 so long as γ_j and η_j are given at one x_0 and the σ_j ($\hat{\sigma}_j$) are specified at that point; σ_j and $\hat{\sigma}_j$ will change sign as γ_j and η_j reach the band edge. In (3.11) the ζ_i ($i=1,2,\ldots,2N$) are branch points. We make branch cuts in the nondegenerate unstable bands l_j ($l_j=\{\zeta:\zeta_{2j-1}<\zeta<\zeta_{2j}\}$) in the ζ plane. In this way we form the Riemann surface R of the roots

$$R(\zeta) = \prod_{i=1}^{2N}(\zeta-\zeta_i)^{1/2}. \tag{3.12}$$

A path for $\gamma_j(\eta_j)$ has two sections $[l_j,+]$ ($[\hat{l}_j,+]$) and $[l_j,-]$ ($[\hat{l}_j,-]$), the former being the upper sheet of $R(\zeta)$ with $\sigma_j(\hat{\sigma}_j)=+1$, and the latter being the lower sheet of $R(\zeta)$ with $\sigma_j(\hat{\sigma}_j)=-1$.

Define

$$Q_j = (\gamma_j,\sigma_j),$$
$$Q_j = (\eta_j,\hat{\sigma}_j). \tag{3.13}$$

By making a change of variables (Abel's transformation [9]) we can integrate the equations (3.11). Define

$$\Omega_m = \sum_{k=0}^{N-2} C_{k,m}\frac{E^k\,dE}{R^{1/2}(E)}, \qquad \hat{\Omega}_m = \sum_{k=0}^{N-2} C_{k,m}\frac{E^k\,dE}{R^{1/2}(E)}, \tag{3.14.a}$$

$$k_m = \sum_{j=1}^{N}\int_{\zeta_{2j-1}}^{Q_j}\Omega_m, \qquad \hat{k}_m = \sum_{j=1}^{N}\int_{\zeta_{2j-1}}^{\hat{Q}_j}\hat{\Omega}_m. \tag{3.14.b}$$

Substituting (3.14) into (3.11), we obtain

$$\frac{dk_m}{dx_0} = -2iC_{N-2,m}, \tag{3.15.a}$$

$$\frac{d\hat{k}_m}{dx_0} = 2iC_{N-2,m}, \tag{3.15.b}$$

$m=1,2,\ldots,N$, $N\geqslant 2$, where we have used

$$\sum_{j=1}^{N} \frac{\gamma_j^k}{\prod_{e\neq j}(\gamma_j-\gamma_e)} = \delta_{k,N-1}, \tag{3.16}$$

and δ_{nm} is the Kronecker delta function. The proof of (3.16) is given in the appendix.

The equations (3.15) yield straight line motion in this new coordinate system. Matveev [10] discusses the solution of $(I-)$ in terms of theta functions. Equations (3.14)–(3.15) are important in constructing the proper variables to use in the theta function. The N-band potential has N periods where each γ_j, η_j move within the unstable band and possess a definite period in x; the periods have a common commensurate period T.

4. The time dependence of the scattering data

In order to reconstruct the potential q_R and q_I at a later time we need to know the scattering data at a later time. From (2.6.a) we have

$$\frac{\partial a_R}{\partial t} = 0. \tag{4.1}$$

It is thus clear that the main spectrum is time independent. If q is an N-band potential initially, it will be an N-band potential for all time.

Using (2.6.a), we get

$$\frac{\partial}{\partial t}\left[a_I(\zeta,x_0)+b_I(\zeta,x_0)\right] = 2\left[a_I(\zeta,x_0)+b_I(\zeta,x_0)\right]B_R(\zeta,x_0) + 2b_R(\zeta,x_0)\left[B_I(\zeta,x_0)+2\zeta^2+|q(x_0)|^2\right],$$

where

$$B_R(\zeta,x_0) = 2\zeta q_R(x_0) - \frac{\partial q_I}{\partial x}(x_0),$$

$$B_I(\zeta,x_0) = 2\zeta q_I(x_0) - \frac{\partial q_R}{\partial x}(x_0).$$

Next, we evaluate ζ in the above equation at $\zeta=\gamma_j$ and make use of (3.3):

$$\frac{\partial \gamma_j}{\partial t}=\frac{-2\sigma_j\sqrt{-\sum_{i=1}^{2N}(\gamma_j-\zeta_i)}}{\prod_{k\neq j}(\gamma_j-\gamma_k)}$$

$$\times\Bigg[\left(\sum_{i=1}^{N}\gamma_i\right)^2+\frac{1}{4}\left(\sum_{i=1}^{2N}\zeta_i\right)^2-2\gamma_j\sum_{k\neq j}\gamma_k+\gamma_j\sum_{i=1}^{2N}\zeta_i$$

$$-\left(\sum_{i=1}^{N}\gamma_i\right)\left(\sum_{i=1}^{2N}\zeta_i\right)+\left(\sum_{i=1}^{N}\eta_i-\frac{1}{2}\sum_{i=1}^{2N}\zeta_i\right)^2$$

$$+\sum_{i=1}^{N}\frac{-2\hat{\sigma}_i\sqrt{-\prod_{k=1}^{N}(\eta_i-\zeta_k)}}{\prod_{k\neq i}(\eta_i-\eta_k)}\left(\frac{1}{2}\sum_{i=1}^{2N}\zeta_i-\sum_{k\neq i}\eta_k\right)\Bigg], \qquad (4.2.a)$$

$j=1,2,\ldots,N$. Similarly, we obtain the evolution equations for η_j:

$$\frac{\partial \eta_j}{\partial t}=\frac{-2\hat{\sigma}_j\sqrt{-\prod_{i=1}^{2N}(\eta_j-\zeta_i)}}{\prod_{k\neq j}(\eta_j-\eta_k)}$$

$$\times\Bigg[-\left(\sum_{i=1}^{N}\eta_i\right)^2-\frac{1}{4}\left(\sum_{i=1}^{2N}\zeta_i\right)^2+2\eta_j\sum_{k\neq j}\eta_k$$

$$-\eta_j\sum_{i=1}^{2N}\zeta_i+\left(\sum_{i=1}^{N}\eta_i\right)\left(\sum_{i=1}^{2N}\zeta_i\right)-\left(\sum_{i=1}^{N}\gamma_i-\frac{1}{2}\sum_{i=1}^{2N}\zeta_i\right)^2$$

$$+\sum_{i=1}^{N}\frac{-2\sigma_i\sqrt{-\prod_{k=1}^{2N}(\gamma_i-\zeta_k)}}{\prod_{k\neq i}(\gamma_i-\gamma_k)}\left(\frac{1}{2}\sum_{i=1}^{2N}\zeta_i-\sum_{k\neq i}\gamma_k\right)\Bigg], \qquad (4.2.b)$$

$j=1,2,\ldots,N$. In Equations (4.2) we see that the evolution of γ_j in time is coupled with the evolution of η_j, whereas in (3.11) they are decoupled in x_0.

For an N-band potential the periodic cubic Schrödinger equation (I$-$) can be solved by solving the scattering equations (1.1) and obtaining γ_i, η_i, and ζ_i at a certain point x_0. We then use (3.11) to reconstruct the potential for all x_0. It is actually easier to obtain ζ_i, γ_i, and η_i for all x_0 by simply solving the polynomial equations $\lambda(\zeta)\lambda^*(\zeta^*)-\mu(\zeta)\mu^*(\zeta^*)=0$, etc., provided we know the value of N beforehand. Equations (4.2) are used to obtain γ_i and η_i at a later time, and we use (3.3) to reconstruct the potential. Mathematically, the original nonlinear partial differential equation has been replaced by $2N$ nonlinear ordinary differential equation which can be integrated via Abel's transformation.

Although Equations (4.2) are complicated, the transformation (3.14) can be used to integrate these equations. We find after some algebra

$$\frac{dk_m}{dt} = 4iC_{N-3,m} - 2i\left(\sum_{i=1}^{2N}\zeta_i\right)C_{N-2,m}, \tag{4.3.a}$$

$$\frac{d\hat{k}_m}{dt} = 4iC_{N-3,m} + 2i\left(\sum_{i=1}^{2N}\zeta_i\right)C_{N-2,m} \tag{4.3.b}$$

for $m=1,2,\ldots,N$, $N>2$, where we have used (3.16) and

$$\sum_{j=1}^{N}\frac{\gamma_j^N}{\prod_{e\neq j}(\gamma_j-\gamma_e)} = \sum_{j=1}^{N}\gamma_j. \tag{4.3.c}$$

The proof of (4.3.c) is given in the appendix. Also,

$$\frac{dk_m}{dt} = 2iC_{0,m}\left(\sum_{i=1}^{4}\zeta_i\right), \tag{4.3.d}$$

$$\frac{d\hat{k}_m}{dt} = 2iC_{0,m}\left(\sum_{i=1}^{4}\zeta_i\right), \tag{4.3.e}$$

for $m=1,2$, $N=2$.

5. *Discussion of the defocusing problem*

The previous sections give a systematic way of integrating an N-band potential for (I$-$). In general, an arbitrary periodic potential will have an infinite number of unstable bands. Nevertheless it can be approximated by an N-band potential provided that the potential is C^∞. To see this, we consider the Bloch eigenfunction (1.6.a),

$$\psi = \begin{pmatrix}\psi_1(x,\zeta)\\ \psi_2(x,\zeta)\end{pmatrix}.$$

where from (1.6.e)

$$\lambda = a_R(\zeta) + i\sqrt{1-a_R^2(\zeta)}\,. \tag{5.1}$$

Thus, $\psi_1(x+T,\zeta)e^{i\zeta(x+T)} = \Psi_1(x)e^{i\zeta x}$ as $|\zeta|\to\infty$, and we take $e^{i\zeta x_0}\psi_1(\zeta,x_0)=1$ without loss of generality.

Defining

$$e^{\hat{\phi}} = \psi_1(x,\zeta)e^{i\zeta x}, \tag{5.2}$$

and substituting it into the scattering equations (1.1), we can generate an equation for $\hat{\phi}'$:

$$2i\zeta\hat{\phi}' = -|q|^2 + \hat{\phi}'^2 + q\frac{d}{dx}\left(\frac{1}{q}\hat{\phi}'\right). \tag{5.3}$$

Since $\hat{\phi}\to 0$, $\hat{\phi}'\to 0$ as $\zeta\to\pm\infty$, we expand $\hat{\phi}$ and $\hat{\phi}'$ in power series in $1/\zeta$:

$$\hat{\phi}\sim\sum_{m=0}^{\infty}\frac{g_m}{(2i\zeta)^{m+1}}, \qquad \hat{\phi}'\sim\sum_{m=0}^{\infty}\frac{f_m}{(2i\zeta)^{m+1}}.$$

Substituting $\hat{\phi}'$ into (5.3), we get a recurrence formula for the f_m:

$$f_{m+1} = q\frac{d}{dx}\left(\frac{1}{q}f_m\right) + \sum_{j+k=m} f_j f_k, \qquad f_0 = -|q|^2. \tag{5.4}$$

Since $\psi_1(x_0,\zeta)=e^{-i\zeta x_0}$, we have $\hat{\phi}(x_0)=0$ and $g_m(x_0)=0$ for all m; hence

$$\hat{\phi}\sim\sum_{m=0}^{\infty}\frac{\int_{x_0}^{x} f_m\,dx}{(2i\zeta)^{m+1}} \qquad \text{as} \quad |\zeta|\to\infty.$$

Using the definition of the Bloch eigenfunction, as well as (5.1) and (5.4), we can show that

$$a_R(\zeta)\sim\cos\left(\zeta T - \sum_{m=0}^{\infty}\frac{I_m}{(2i\zeta)^{m+1}}\right) \qquad \text{as} \quad |\zeta|\to\infty. \tag{5.5}$$

From (5.5) we see that

$$\zeta = \frac{1}{T}\left(N\pi + \sum_{m=0}^{\infty}\frac{I_m}{(2i\zeta)^{m+1}}\right) \qquad (\text{as} \quad |\zeta|\to\infty,\ N\to\pm\infty)$$

is a double zero of $1-a_N^2=0$ to all orders of $1/\zeta$. This means that as $|\zeta|\to\infty$ the higher band size is asymptotically zero to all orders of $1/\zeta$. The contribution to the potential due to each band is at most $\zeta_N-\zeta_{N-1}$, and hence from (3.3) it is also zero to all orders as $|\zeta|\to\infty$. The above arguments hold only for a C^∞ potential, since otherwise the conserved quantities I_m may be unbounded, and the asymptotic expansion (5.5) is not valid.

II. The periodic focusing cubic Schrödinger equation

The periodic focusing cubic Schrödinger equation (I+) differs from the periodic defocusing cubic Schrödinger equation in some characteristic ways. One of the major differences is that the spectra we are looking for in the focusing case are complex, while in the defocusing case they are real. Most of the spectral properties given in Part I of this paper are no longer necessarily true here. In order to use a similar analysis we must make more assumptions which we state explicitly herein.

1. The direct scattering problem

The associated scattering equations for this case are [3]

$$\frac{\partial V_1}{\partial x}+i\zeta V_1=qV_2, \tag{6.1.a}$$

$$\frac{\partial V_2}{\partial x}-i\zeta V_2=-q^*V_1. \tag{6.1.b}$$

As before, we choose

$$\phi(x,\zeta)=\begin{pmatrix}\phi_1(x,\zeta)\\ \phi_2(x,\zeta)\end{pmatrix} \quad \text{with} \quad \phi(x_0,\zeta)=\begin{pmatrix}1\\0\end{pmatrix}$$

to be one of the eigenfunctions of (6.1). The second independent eigenfunction is

$$\bar{\phi}(x,\zeta)=\begin{pmatrix}\phi_2^*(x,\zeta^*)\\ -\phi_1^*(x,\zeta^*)\end{pmatrix}.$$

Expanding $\phi(x+T,\zeta)$ in terms of $\phi(x,\zeta)$ and $\bar{\phi}(x,\zeta)$, we have

$$\phi(x+T,\zeta)=a\phi(x,\zeta)+b\bar{\phi}(x,\zeta). \tag{6.2}$$

Evaluating x in (1.2) at x_0, we obtain

$$a(\zeta,x_0)=\phi_1(x_0+T,\zeta), \tag{6.3.a}$$

$$b(\zeta,x_0)=-\phi_2(x_0+T,\zeta), \tag{6.3.b}$$

and from the Wronskian relation,

$$a(\zeta, x_0)a^*(\zeta^*, x_0) + b(\zeta, x_0)b^*(\zeta^*, x_0) = 1. \tag{6.3.c}$$

We denote by a_R, b_R, and V_{iR} the real parts of a, b, and V_i when ζ is real ($a_R = [a(\zeta) + a^*(\zeta^*)]/2$ etc.). a_I, b_I, and V_{iI} are the imaginary parts of a, b, and V_i.

We choose the following three spectra (which are slightly different from the spectra in the defocusing case):

(i) *The main spectrum*: The main spectrum is composed of the eigenvalues ζ_i' ($i = 1, 2, \ldots, \infty$) at least one of whose eigenfunctions is periodic or antiperiodic. The main spectrum is composed of the roots of $1 - a_R^2 = 0$. There are two special properties of ζ_i': (a) ζ_i' ($i = 1, 2, \ldots$) either are real or appear as complex conjugate pairs (or both). (b) When ζ_i' is real, it will be at least a double zero of $1 - a_R^2 = 0$. To prove (a) we use the fact that ζ_i' are the roots of $1 - a_R^2 = 0$ and hence $\{[a(\zeta_i') + a^*(\zeta_i'^*)]/2\}^2 = 1$. Taking the complex conjugate of this equation, we have $\{[a^*(\zeta_i') + a(\zeta_i'^*)]/2\}^2 = 1$. Thus, $\zeta_i'^*$ is also a root of $1 - a_R^2 = 0$. We therefore conclude that the ζ_i' either are real or appear as complex conjugate pairs (or both). To prove (b) we use the fact that when ζ_i' is real, $a_R(\zeta_i')$, $b_R(\zeta_i')$, and $b_I(\zeta_i')$ are all real. From (1.3.c) we get $a_R^2(\zeta_i') + a_I^2(\zeta_i') + b_R^2(\zeta_i') + b_I^2(\zeta_i') = 1$. Since $a_R^2(\zeta_i') = 1$, we have $a_I(\zeta_i') = b_R(\zeta_i') = b_I(\zeta_i') = 0$, and upon differentiation of this equation with respect to ζ,

$$\frac{\partial}{\partial \zeta} a_R \Big|_{\zeta = \zeta_i'} = 0.$$

Therefore ζ_i' is at least a double zero of $1 - a_R^2 = 0$. For real ζ, $|a_R(\zeta)| \leq 1$ and hence no unstable band can exist for real ζ. In this respect, the focusing problem is different from the defocusing problem. In what follows we divide ζ_i' into $\hat{\zeta}_i$ and ζ_i, where $\hat{\zeta}_i$ is real and is a (at least) double zero of $1 - a_R^2 = 0$. The quantity ζ_i is complex, and therefore it appears in complex conjugate pairs, i.e., as (ζ_i, ζ_i^*). Moreover, as in the defocusing case, $a(\zeta)$ and $b(\zeta)$ are entire functions of ζ, and if the initial condition is real, we can show that $a_R^*(-\zeta^*) = a_R(\zeta)$. This implies that if ζ is an eigenvalue of $1 - a_R^2 = 0$, then so is $-\zeta$. The eigenvalues then will appear in double pairs $(\zeta, \zeta^*, -\zeta, -\zeta^*)$.

(ii) *Auxiliary spectrum (I)*: This is composed of the eigenvalues γ_i' ($i = 1, 2, \ldots, \infty$) for which $\phi_{1I}(x_0 + T, \gamma_i') - i\phi_{2I}(x_0 + T, \gamma_i') = 0$. From the definition of ϕ we see that this implies that the γ_i are the roots of $a_I + ib_I = 0$. We divide γ_i' into γ_i and $\hat{\gamma}_i$. $\hat{\gamma}_i$ is real and coincides with $\hat{\zeta}_i$, whereas γ_i is complex.

(iii) *Auxiliary spectrum (II)*: This is composed of the eigenvalues η_i ($i = 1, 2, \ldots$) for which $\phi_{1I}(x_0 + T, \eta_i') - i\phi_{2R}(x_0 + T, \eta_i') = 0$, with η_i' the roots of $a_I + ib_R = 0$. We divide η_i' into η_i and $\hat{\eta}_i$. $\hat{\eta}_i$ is real and coincides with $\hat{\zeta}_i$, whereas η_i is complex.

Next, we state explicitly our assumptions for the scattering data of (6.1).

(a) All the $\hat{\zeta}_i$ are double zeros of $1 - a_R^2 = 0$, i.e., there are no higher order zeroes of $1 - a_R^2 = 0$. Exceptional cases are discussed in the appendix.

(b) All $\hat{\gamma}_i, \hat{\eta}_i$ are simple zeros of $a_I + ib_I = 0$ and $a_I + ib_R = 0$, respectively. They are real and coincide with $\hat{\xi}_i$.

(c) All (ζ_i, ζ_i^*) are simple zeros of $1 - a_R^2 = 0$ and are complex.

(d) All γ_i, η_i are simple zeros of $a_I + ib_I = 0$ and $a_I + ib_R = 0$, respectively, and are complex.

2. *N-band potential*

As in the defocusing problem, by an N-band potential we mean one which has $2N$ simple roots of $1 - a_R^2 = 0$. An N-band potential satisfies the equation $\sum_{m=0}^{N} C_m \delta I_m / \delta q^* = 0$, where I_m are the conserved quantities:

$$I_0 = \int_{x_0}^{x_0+T} |q|^2 \, dx, \tag{7.1.a}$$

$$I_1 = \frac{i}{2} \int_{x_0}^{x_0+T} (qq_x^* - q^* q_x) \, dx, \tag{7.1.b}$$

$$I_2 = -\int_{x_0}^{x_0+T} \left(|q|^4 - |q_x|^2\right) dx. \tag{7.1.c}$$

As before, we take the time dependence of the eigenfunction to be

$$\frac{\partial \phi_1}{\partial t} = A\phi_1 + B\phi_2 + \lambda\phi_1 + \mu\phi_2^*, \tag{7.2.a}$$

$$\frac{\partial \phi_2}{\partial t} = C\phi_1 + D\phi_2 + \lambda\phi_2 - \mu\phi_1^*. \tag{7.2.b}$$

To make $\phi_1(x_0, t, \zeta) = 1$ and $\phi_2(x_0, t, \zeta) = 0$ for all time, we require

$$\lambda = -A(x_0, t, \zeta), \qquad \mu = C(x_0, t, \zeta). \tag{7.3}$$

The consistency of (7.1) and (6.1) leads to $D(\zeta) = A^*(\zeta^*) = -A(\zeta)$, $C(\zeta) = -B^*(\zeta^*)$, and the evolution equation. $A(\zeta)$ and $\lambda(\zeta)$ are purely imaginary when ζ is real.

The evolution equations that can be solved by the inverse scattering technique and possess (1.1) as their scattering equations are

$$iq_t = \frac{\delta I_m}{\delta q^*}, \qquad m = 0, 1, 2, \ldots. \tag{7.4}$$

For each m, we have the corresponding A_m and B_m. A_m is a polynomial in ζ of order m, and B_m is a polynomial in ζ of order $m-1$. We list the first few A_m and B_m:

M	Evolution equation	A_m	B_m
0	$iq_t = q$	$-i/2$	0
1	$q_t = -q_x$	$i\zeta$	$-q$
2	$iq_t + q_{xx} + 2\lvert q\rvert^2 q = 0$	$-2i\zeta^2 + i\lvert q\rvert^2$	$iq_x + 2\zeta q$

The proof that q is an N-band potential is the same as given in Part I and hence omitted. We point out that the $2N$ simple roots are also the roots of the polynomial equation

$$\lambda(\zeta)\lambda^*(\zeta^*) + \mu(\zeta)\mu^*(\zeta^*) = 0. \tag{7.5}$$

The N complex γ_i are the roots of

$$\mu_I(\zeta) - \lambda(\zeta) = 0, \tag{7.6.a}$$

and the N complex η_i are the roots of

$$\mu_R(\zeta) - \lambda(\zeta) = 0, \tag{7.6.b}$$

where

$$\lambda(\zeta) = -\sum_{m=0}^{N} c_m A_m(x_0, t, \zeta), \tag{7.7.a}$$

$$\mu(\zeta) = -\sum_{m=0}^{N} c_m B_m^*(x_0, t, \zeta^*). \tag{7.7.b}$$

3. The inverse scattering formula and the time dependence of the scattering data

Since a and b are entire functions of ζ, we have the following relations:

$$\frac{(a_I + ib_I)^2}{1 - a_R^2} = \frac{\prod_{i=1}^{N} (\zeta - \gamma_i)^2}{\prod_{i=1}^{2N} (\zeta - \zeta_i)}, \tag{8.1.a}$$

$$\frac{(a_I + ib_R)^2}{1 - a_R^2} = \frac{\prod_{i=1}^{N} (\zeta - \eta_i)^2}{\prod_{i=1}^{2N} (\zeta - \zeta_i)}. \tag{8.1.b}$$

Expanding a_R, a_I, b_R, and b_I in terms of power series in $1/\zeta$ as $|\zeta| \to \infty$, and comparing with the coefficients of the terms $1/\zeta$ in (8.1), we obtain

$$q_I = -i\left(\frac{1}{2}\sum_{i=1}^{2N} \zeta_i - \sum_{i=1}^{N} \gamma_i\right), \tag{8.2.a}$$

$$q_R = i\left(\frac{1}{2}\sum_{i=1}^{2N} \zeta_i - \sum_{i=1}^{N} \eta_i\right). \tag{8.2.b}$$

Following Part I, we can obtain the x_0 dependence of the scattering data by shifting the reference point x_0 to x_0+dx_0 and expanding the basis eigenfunctions at x_0+dx_0 in terms of the basis eigenfunctions at x_0. We find

$$\frac{\partial}{\partial x_0}\zeta_i = 0, \qquad i = 1,2,\ldots,2N, \tag{8.3}$$

$$\frac{\partial \gamma_j}{\partial x_0} = \frac{\sigma_j \sqrt{\prod_{i=1}^{2N}(\gamma_j-\zeta_i)}\left(-2i\sum_{i\neq j}\gamma_i + i\sum_{i=1}^{2N}\zeta_i\right)}{\prod_{i\neq j}(\gamma_j-\gamma_i)}, \tag{8.4.a}$$

$$\frac{\partial \eta_j}{\partial x_0} = \frac{\delta_j \sqrt{\prod_{i=1}^{2N}(\eta_j-\zeta_i)}\left(-2i\sum_{i\neq j}\eta_i + i\sum_{i=1}^{2N}\zeta_i\right)}{\prod_{i\neq j}(\eta_j-\eta_i)}, \tag{8.4.b}$$

$$j = 1,2,\ldots,N, \qquad \sigma_j = \pm 1, \quad \delta_j = \pm 1.$$

The square roots in the above equations correspond to the positive real parts.

To obtain the time dependence of the scattering data we make use of (7.2). Hence, $\partial a_R/\partial t = 0$, or

$$\frac{\partial \zeta_i}{\partial t} = 0, \qquad i = 1,2,\ldots,2N, \tag{8.5}$$

and

$$\frac{\partial \gamma_j}{\partial t} = \frac{2i\sigma_j \sqrt{\prod_{i=1}^{2N}(\gamma_j-\zeta_i)}}{\prod_{i\neq j}(\gamma_j-\gamma_i)}\Bigg[-2\gamma_j\sum_{i\neq j}\gamma_i + \gamma_j\sum_{i=1}^{2N}\zeta_i$$

$$+\left(\frac{1}{2}\sum_{i=1}^{2N}\zeta_i - \sum_{i=1}^{N}\eta_i\right)^2 + \left(\frac{1}{2}\sum_{i=1}^{2N}\zeta_i - \sum_{i=1}^{N}\gamma_i\right)^2$$

$$+\sum_{i=1}^{N}\frac{\delta_i\sqrt{\prod_{k=1}^{2N}(\eta_i-\zeta_k)}}{\prod_{k\neq i}(\eta_i-\eta_k)}\left(i\sum_{k=1}^{2N}\zeta_k - 2i\sum_{k\neq i}\eta_k\right)\Bigg], \tag{8.6.a}$$

$$\frac{\partial \eta_j}{\partial t} = \frac{2i\hat{\sigma}_j \sqrt{\prod_{i=1}^{2N}(\eta_j - \zeta_i)}}{\prod_{i\neq j}(\eta_j - \eta_i)} \Bigg[-2\eta_j \sum_{i\neq j} \eta_i + \eta_j \sum_{i=1}^{2N} \zeta_i$$

$$+ \left(\frac{1}{2}\sum_{i=1}^{2N}\zeta_i - \sum_{i=1}^{N}\eta_i\right)^2 + \left(\frac{1}{2}\sum_{i=1}^{2N}\zeta_i - \sum_{i=1}^{2N}\gamma_i\right)^2$$

$$- \sum_{i=1}^{N} \frac{\sigma_i \sqrt{\prod_{k=1}^{2N}(\gamma_i - \zeta_k)}}{\prod_{k\neq i}(\gamma_i - \gamma_k)} \left(-2i \sum_{k\neq i} \gamma_k + i \sum_{k=1}^{2N} \zeta_k \right) \Bigg], \quad (8.6.b)$$

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

In order to integrate these equations we first determine $\sigma_j, \hat{\sigma}_j$ at $t=0$ with respect to various x_0. σ_j or $\hat{\sigma}_j$ changes sign as γ_j or η_j reaches the band edge ζ_i, respectively.

We summarize the above results as follows. For an N-band potential we solve the scattering equations (6.1) to obtain the main spectrum as well as the auxiliary spectra at $t=0$ at an arbitrary base point x_0. Then we use (8.6) to obtain the auxiliary spectra at a later time. Finally we use (8.4) to obtain the auxiliary spectra at various x_0 and make use of (8.2) to reconstruct the potential q. The Abel transformations introduced in Part I of this paper can also be used to integrate (8.4) and (8.6). In general, an N-band potential has N noncommensurate periods, and hence q is almost periodic in time. This agrees with the numerical solution of the periodic cubic Schrödinger equation [12].

4. 1-band and 2-band Potential

If we compare Equations (8.2), (8.4), and (8.6) with the corresponding equations in the defocusing problem, we find little difference between these equations. However, the geometries of the loci of the motion of these spectra with respect to x_0 or t in the two problems are completely different. For simplicity we first consider the 1-band potential. In the defocusing problem we have [from $\Sigma_{n=0}^{1} c_m \delta I_m / \delta q^* = 0$; alternatively we could solve (8.4) with $j=1$]

$$-c_0 q + i c_1 q_x = 0. \quad (9.1)$$

Hence,

$$q(x) = A \exp\left(-i \frac{c_0}{c_1} x\right).$$

From $|\lambda(\zeta)|^2-|\mu(\zeta)|^2=0$ we have

$$\zeta_1=\frac{c_0}{2c_1}-A, \qquad \zeta_2=\frac{c_0}{2c_1}+A, \tag{9.2.a}$$

and from $-i\lambda(\zeta)+\mu_I(\zeta)=0$, $-i\lambda(\zeta)+\mu_R(\zeta)=0$ we have

$$\gamma_1=\frac{c_0}{2c_1}-A\sin\left(\frac{c_0}{c_1}x_0\right), \tag{9.2.b}$$

$$\eta_1=\frac{c_0}{2c_1}+A\cos\left(\frac{c_0}{c_1}x_0\right), \tag{9.2.c}$$

respectively. Using (9.2) as initial values, we can solve the evolution equations of γ_1 and η_1 with respect to time and verify that the solution for all time is

$$q(x,t)=A\exp\left[-i\frac{c_0}{c_1}-i\left(\frac{c_0}{c_1}\right)^2t-2iA^2t\right]. \tag{9.3}$$

Hence,

$$\gamma_1=\frac{c_0}{2c_1}-A\sin\left[\frac{c_0}{c_1}x+\left(\frac{c_0}{c_1}\right)^2t+2A^2t\right], \tag{9.4.a}$$

$$\eta_1=\frac{c_0}{2c_1}-A\cos\left[\frac{c_0}{c_1}x+\left(\frac{c_0}{c_1}\right)^2t+2A^2t\right]. \tag{9.4.b}$$

The loci of γ_1 and η_1 with respect to time or x_0 are shown in Figure 2. We see that γ_1 and η_1 move back and forth along a horizontal straight line segment $(-A-c_0/2c_1, A-c_0/2c_1)$.

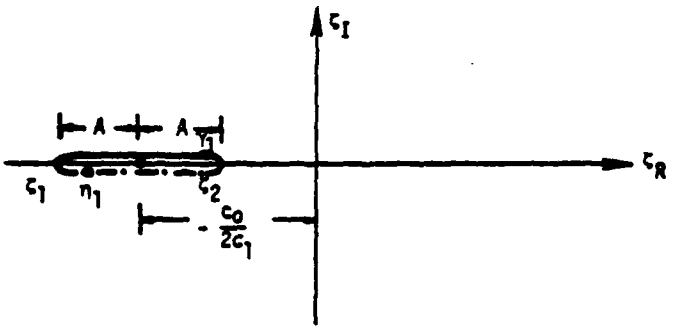


Figure 2.

In the focusing problem the differential equation for the 1-band potential is

$$c_0 q + i c_1 q_x = 0. \tag{9.5}$$

Hence

$$q(x) = A \exp\left(i \frac{c_0}{c_1} x \right). \tag{9.6}$$

It is straightforward to verify that

$$q(x,t) = A \exp\left[i \frac{c_0}{c_1} - i \left(\frac{c_0}{c_1} \right)^2 t + 2iA^2 t \right] \tag{9.7}$$

is the most general one-band potential.

The various spectra are

$$\zeta_1 = \frac{c_0}{2c_1} - iA, \qquad \zeta_2 = \frac{c_0}{2c_1} + iA, \tag{9.8.a}$$

$$\gamma_1 = \frac{c_0}{2c_1} - iA \sin\left[\frac{c_0}{c_1} x_0 - \left(\frac{c_0}{c_1} \right)^2 t + 2A^2 t \right], \tag{9.8.b}$$

$$\eta_1 = \frac{c_0}{2c_1} + iA \cos\left[\frac{c_0}{c_1} x_0 - \left(\frac{c_0}{c_1} \right)^2 t + 2A^2 t \right]. \tag{9.8.c}$$

The loci of γ_1 and η_1 are shown in Figure 3. Here γ_1 and η_1 move back and forth along a vertical straight line segment $(-iA - c_0/2c_1, iA - c_0/2c_1)$.

The above example suggests that the horizontal band in the defocusing problem become the vertical band in the focusing problem.

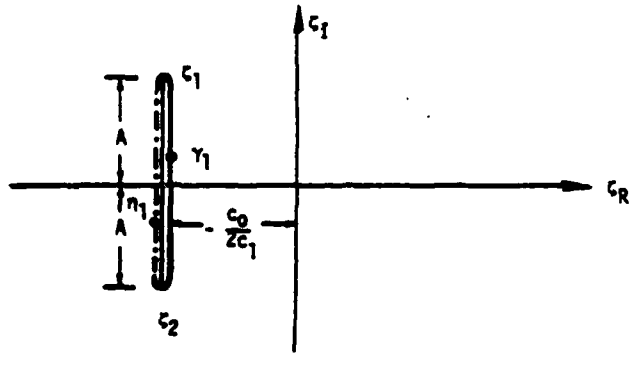


Figure 3.

Next, we consider the 2-band problem. In the defocusing case it is qualitatively the same as for the 1-band problem. We have γ_1 and η_1 moving along the horizontal straight line segment (ζ_1, ζ_2), whereas γ_2 and η_2 move along (ζ_3, ζ_4). In the focusing problem, we show that the loci can be totally different from those suggested by the 1-band problem. The differential equation for $q(x)$ is

$$c_0 q + 2ic_1 q_x - c_2 q_{xx} - 2c_2 |q|^2 q = 0. \tag{9.9}$$

We consider here the special case that $q(x)$ is real, $q(x)=v(x)$. From (9.9) we require $c_1 = 0$, and we take $c_2 = 1$ without loss of generality. Integrating (9.9) once, we get

$$[v'(x)]^2 = c_0 V^2 - V^4 + \hat{A}_0, \tag{9.10}$$

where $\hat{A}_0 + c_0 v^2 - v^4 \geqslant 0$ to guarantee a real solution.

The solution $v(x)$ of (9.10) can be expressed in terms of an elliptic function $\mathrm{cn}(x|m)$ (m is the modulus): $v(x) = \hat{A}\,\mathrm{cn}(\hat{B}x|m)$. The relations between $\hat{A}$, $\hat{B}$, m, and $\hat{A}_0$ are

$$\hat{A}^2\hat{B}^2(1-m) = \hat{A}_0, \tag{9.11.a}$$

$$c_0 = \hat{B}^2 = (2m-1), \tag{9.11.b}$$

$$\hat{A}^2 = m\hat{B}^2. \tag{9.11.c}$$

We now look for the four complex simple roots of $1 - a_R^2 = 0$. From Equation (7.5) we have

$$\zeta^4 + \frac{c_0}{2}\zeta + \frac{c_0^2}{16} + \frac{\hat{A}_0}{4} = 0. \tag{9.12}$$

We get the four roots $(\zeta_1, \zeta_1^*, -\zeta_1, -\zeta_1^*)$, which occur in double pairs (as expected):

$$\zeta_{1R} = d, \qquad \zeta_{1I}^2 = \frac{c_0}{4} + d^2, \tag{9.13.a}$$

where

$$d^2 = \frac{-c_0 + \sqrt{c_0^2 + 4\hat{A}_0}}{8}. \tag{9.13.b}$$

The roots γ_1 and γ_2 are the roots of $\lambda - \mu_f = 0$, i.e.,

$$\frac{c_0}{2} + 2\gamma_i^2 - \hat{A}^2\,\mathrm{cn}^2(Bx|m) - i\hat{A}\hat{B}\,\mathrm{sn}(Bx|m)\,\mathrm{dn}(\hat{B}x|m) = 0, \qquad i = 1,2. \tag{9.14}$$

From (9.14), γ_1 and γ_2 are related by $\gamma_2 = -\gamma_1$. Taking $\gamma_1 = \gamma_R + i\gamma_I$ and substituting it into (9.14), we get

$$2(\gamma_R^2 - \gamma_I^2) = -\frac{c_0}{2} + \hat{A}^2 \mathrm{cn}^2(\hat{B}x|m),$$

$$4\gamma_R\gamma_I = \hat{A}\hat{B}\,\mathrm{sn}(\hat{B}x|m)\,\mathrm{dn}(\hat{B}x|m).$$

Hence $\gamma_R^2 + \gamma_I^2 = \frac{1}{2}\sqrt{\hat{A}_0 + c_0^2/4}$ and we conclude that γ_1 moves along a circular arc of radius $\sqrt[4]{\hat{A}_0/4 + c_0^2/16}$ and γ_2 moves along a different circular arc of the same radius.

The roots η_1 and η_2 are the roots of $\lambda - \mu_R = 0$, or

$$2\eta_i^2 - \hat{A}^2 \mathrm{cn}^2(\hat{B}x|m) + \frac{c_0}{2} - 2i\hat{A}\eta_i \mathrm{cn}(\hat{B}x|m) = 0, \qquad i = 1,2. \tag{9.15}$$

From the above equation η_1 and η_2 are related by $\eta_2 = -\eta_1^*$. Taking $\eta_1 = \eta_R + i\eta_I$ and substituting it into (9.15), we get

$$2(\eta_R^2 - \eta_I^2) = \hat{A}^2 \mathrm{cn}^2(\hat{B}x|m) - \frac{c_0}{2} - 2\hat{A}\eta_I \mathrm{cn}(\hat{B}x|m),$$

$$4\eta_R\eta_I = 2\hat{A}\eta_R \mathrm{cn}(\hat{B}x|m).$$

Hence $\eta_R^2 - \eta_I^2 = -c_0/2$, i.e., we have hyperbolas.

Depending on the value of c_0, we have three different cases:

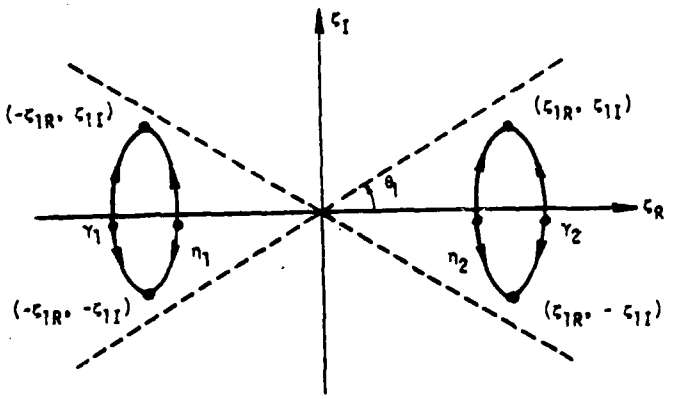
Case I: $c_0 < 0$, $m < \frac{1}{2}$, Figure 4(a);
Case II: $c_0 > 0$, $m > \frac{1}{2}$, Figure 4(b);
Case III: $c_0 = 0$, $m = \frac{1}{2}$, Figure 4(c).

In case III, $\eta_1 = \eta_2 = 0$ occurs when $\mathrm{cn}(\hat{B}x'|m = \frac{1}{2}) = 0$ for some x'. In case II, $\eta_1 = \eta_2 \neq 0$ occurs for some x''. At these points the differential equations (8.4) and (8.6) are singular and the previous analysis must be modified, as we have assumed only simple roots in each band. This demonstrates a significant difference from the defocusing problem. In that case each γ_i and η_i moves within its own unstable band and hence $\eta_1 = \eta_2$ can not happen for any x and t. In general, the auxiliary spectra move on curves between ζ_i in the focusing problem. This makes it difficult to approximate an arbitrary periodic potential by an N-band potential.

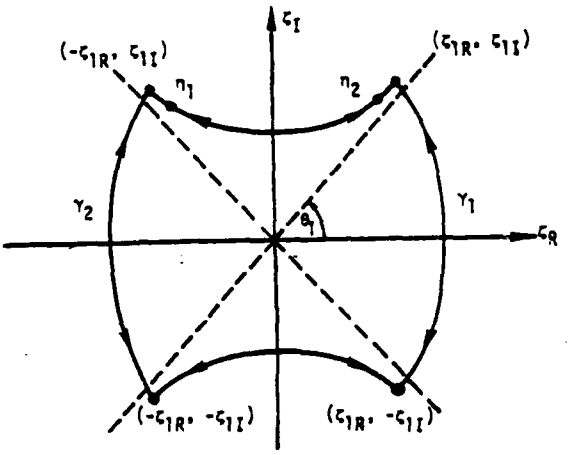
Next, we examine the stability of a 1-band potential with respect to a second band of perturbations, where the second band is small ($|\zeta_3 - \zeta_4| \ll |\zeta_1 - \zeta_2|$).

The general equation for a 2-band potential $q(x)$ is

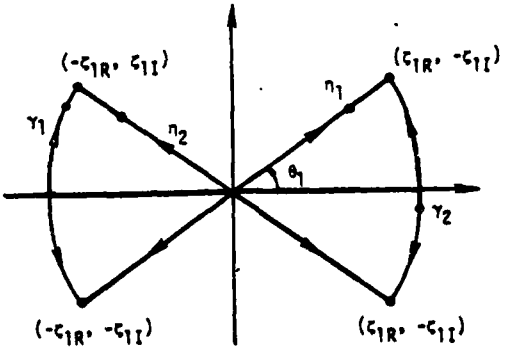
$$c_0\frac{\delta I_0}{\delta q^*} + \frac{\delta I_1}{\delta q^*} + \frac{\delta I_2}{\delta q^*} = 0. \tag{9.14.a}$$



(a)



(b)



(c)

Figure 4. (a) $\theta_1 < \pi/4$. (b) $\theta_1 > \pi/4$. (c) $\theta_1 = \pi/4$.

From (7.1) the above equation can be written

$$q_{xx} - 2ic_1 q_x - c_0 q + 2|q|^2 q = 0. \tag{9.14.b}$$

To remove the term $-2ic_1 q_x$ we take

$$q = \hat{q} e^{ic_1 x}, \tag{9.15.a}$$

and Equation (9.14b) becomes

$$\hat{q}_{xx} + (c_1^2 - c_0)\hat{q} + 2|\hat{q}|^2 \hat{q} = 0. \tag{9.16}$$

Let $\hat{q} = re^{i\theta}$, where r is real and positive. Substitute this expression into (9.16):

$$r_{xx} - r\theta_x^2 + (c_1^2 - c_0)r + 2r^3 = 0, \tag{9.17.a}$$

$$2r_x\theta_x + r\theta_{xx} = 0. \tag{9.17.b}$$

Integrating Equation (9.17.b), we obtain

$$\theta_x = ar^{-2}. \tag{9.18}$$

Equation (9.17.a) becomes

$$r_{xx} - a^2 r^{-3} + (c_1^2 - c_0)r + 2r^3 = 0, \tag{9.19.a}$$

or

$$r_x^2 + a^2 r^{-2} + (c_1^2 - c_0)r^2 + r^4 = 2\hat{b}. \tag{9.19.b}$$

For simplicity we define

$$\hat{r} = r^2, \tag{9.20}$$

and Equations (9.18) and (9.19b) can be written

$$(\hat{r}_x)^2 = 8b\hat{r} - 4\hat{r}^3 - 4(c_1^2 - c_0)\hat{r}^2 - 4a^2, \tag{9.21.a}$$

$$\theta = \int a\hat{r}^{-1} dx. \tag{9.21.b}$$

To calculate the band edges of the above potential, we note that there are two quantities associated with (9.14.b) that are independent of x:

$$2ic_1|q|^2 - (q^* q_x - q q_x^*) = iA', \tag{9.22.a}$$

$$c_0|q|^2 - |q_x|^2 - |q|^4 = B'. \tag{9.22.b}$$

For the potential (9.17) we readily obtain A' and B':

$$A' = -2\hat{a}, \tag{9.23.a}$$

$$B' = -2\hat{b} - 2c_1\hat{a}. \tag{9.23.b}$$

The band edges are the roots of $\lambda(\zeta)\lambda^*(\zeta^*)+u(\zeta)u^*(\zeta^*)=0$. From the definition of λ and μ (7.7), we get a fourth order polynomial equation for ζ:

$$\zeta^4 + 2c_1\zeta^3 + \left(\frac{c_0}{2} + c_1^2\right)\zeta^2 + \left(-\hat{a} + \frac{c_0c_1}{2}\right)\zeta + \frac{1}{4}\left(\frac{c_0^2}{4} - 2c_1\hat{a} + 2\hat{b}\right) = 0. \tag{9.24}$$

Since we want the 2-band potential to be a small perturbation of the 1-band potential, we take the four roots to be of the forms

$$\zeta_1 = iA, \tag{9.25.a}$$

$$\zeta_2 = -iA, \tag{9.25.b}$$

$$\zeta_3 = d + i\epsilon, \tag{9.25.c}$$

$$\zeta_4 = d - i\epsilon, \tag{9.25.d}$$

with $\epsilon \ll 1$. Comparing (9.24) with (9.25), we can relate A, d with c_0, c_1, $\hat{a}$, and $\hat{b}$:

$$c_1 = -d, \tag{9.26.a}$$

$$c_0 = 2(A^2 + \epsilon^2), \tag{9.26.b}$$

$$\hat{a} = (A^2 - \epsilon^2)d, \tag{9.26.c}$$

$$\hat{b} = A^2d^2 + A^2\epsilon^2 + d^2\epsilon^2 - \tfrac{1}{2}(A^4 + \epsilon^4). \tag{9.26.d}$$

Substituting the above expressions into Equation (9.21.a), we get

$$(\hat{\rho}_x)^2 = 8\left(A^2d^2 + d^2\epsilon^2 + A^2\epsilon^2 - \tfrac{1}{2}A^4 - \tfrac{1}{2}\epsilon^4\right)\hat{\rho} - 4\hat{\rho}^3$$
$$- 4(d^2 - 2A^2 - 2\epsilon^2)\hat{\rho}^2 - 4d^2(A^2 - \epsilon^2)^2. \tag{9.27}$$

To obtain the periodic solution of (9.27) we first examine the roots of the right hand side of (9.27),

$$(\hat{\rho}_x)^2 = -4(\hat{\rho} - \hat{\rho}_1)(\hat{\rho} - \hat{\rho}_2)(\hat{\rho} - \hat{\rho}_3). \tag{9.28}$$

To order ϵ, we find

$$\hat{r}_1 = -d^2, \tag{9.29.a}$$

$$\hat{r}_2 = A^2 - 2A\epsilon, \tag{9.29.b}$$

$$\hat{r}_3 = A^2 + 2A\epsilon. \tag{9.29.c}$$

Then we can write $\hat{r}$ in explicit form:

$$\hat{r} = A^2 - 2A\epsilon + 4A\epsilon \operatorname{cn}^2\left(\sqrt{d^2 + A^2 + 2A\epsilon}\, x \mid m\right). \tag{9.30.a}$$

Since $m = O(\epsilon)$ in (9.30.a), we can approximate $\operatorname{cn}(x|m)$ by $\cos x$:

$$\hat{r} \sim A^2 + 2A\epsilon \cos 2\sqrt{d^2 + A^2}\, x. \tag{9.30.b}$$

Hence

$$r \sim A + \epsilon \cos 2\sqrt{d^2 + A^2}\, x. \tag{9.31}$$

From (9.21.b),

$$\theta = A^2 d\left(\frac{x}{A^2} - \frac{\epsilon}{A^3\sqrt{d^2 + A^2}} \sin 2\sqrt{d^2 + A^2}\, x\right). \tag{9.32}$$

Finally, we obtain an approximation to the 2-band potential of (I+):

$$q(x) = A + \epsilon \cos 2\sqrt{d^2 + A^2}\, x - \frac{i\epsilon d}{\sqrt{d^2 + A^2}} \sin 2\sqrt{d^2 + A^2 x}\,. \tag{9.33}$$

Next we note that the usual linear stability analysis of the one band potential would be as follows. Take $q(x,t) = Ae^{2iA^2t}(1 + a_+ e^{i(kx+\Omega t)} + a_- e^{-i(kx+\Omega t)})$, and substitute it into the cubic Schrödinger equation (I+). We get

$$(2A^2 - \Omega - k^2)a_+ + 2A^2 a_-^* = 0, \tag{9.34.a}$$

$$2A^2 a_+ + (2A^2 - k^2 + \Omega)a_-^* = 0. \tag{9.34.b}$$

The above equations give us

$$\Omega^2 = k^2(k^2 - 4A^2). \tag{9.34.c}$$

This means that the plane wave solution is stable with respect to perturbations of wave number $k>2A$ and unstable for $k<2A$.

It is significant that the solution (9.33) has a wave number $2\sqrt{d^2+A^2}$ which is *always bigger than* $2A$. In the limit $d\to 0$, the 2-band potential (9.33) corresponds to the wave number $2A$. Hence the potential (9.33) should be stable according to linear stability analysis.

We now study the stability of the potential with initial condition (9.33) via the analysis given in this paper. Let us first determine γ_1, γ_2, η_1, η_2, σ_1, σ_2, $\hat{\sigma}_1$, and $\hat{\sigma}_2$ at $t=0$. Solving the equation $\lambda-\mu_I=0$, we obtain γ_1 and γ_2:

$$\gamma_1 = -\frac{A\epsilon}{d}\cos 2\sqrt{d^2+A^2}\,x - \frac{i\epsilon A^2}{d\sqrt{d^2+A^2}}\sin 2\sqrt{d^2+A^2}\,x, \tag{9.35.a}$$

$$\gamma_2 = d + \frac{id\epsilon}{\sqrt{d^2+A^2}}\sin 2\sqrt{d^2+A^2}\,x + \frac{A\epsilon}{d}\cos 2\sqrt{d^2+A^2}\,x$$

$$+ \frac{i\epsilon A^2}{d\sqrt{d^2+A^2}}\sin 2\sqrt{d^2+A^2}\,x. \tag{9.35.b}$$

To determine σ_2, we make use of (8.4.a) and find at $t=0$

$$\sigma_2 = -1 \quad \text{when} \quad \frac{3\pi}{2} \leqslant 2\sqrt{a^2+A^2}\,x < 2\pi \text{ or } 0 \leqslant 2\sqrt{a^2+A^2}\,x < \frac{\pi}{2}, \tag{9.35.c}$$

$$\sigma_2 = 1 \quad \text{when} \quad \frac{\pi}{2} \leqslant 2\sqrt{a^2+A^2}\,x < \frac{3\pi}{2}. \tag{9.35.d}$$

Similarly, we solve the equation $\lambda-\mu_R=0$ to obtain η_1 and η_2:

$$\eta_1 = iA, \tag{9.36.a}$$

$$\eta_2 = d + i\epsilon\cos 2\sqrt{d^2+A^2}\,. \tag{9.36.b}$$

To determine $\hat{\sigma}_2$ we make use of Equation (8.4.b) and find

$$\hat{\sigma}_2 = -1\,. \quad \text{when} \quad 0 \leqslant 2\sqrt{d^2+A^2}\,x < \pi, \tag{9.37.a}$$

$$\hat{\sigma}_2 = 1 \quad \text{when} \quad \pi \leqslant 2\sqrt{d^2+A^2}\,x < 2\pi. \tag{9.37.b}$$

Next, consider the time dependence of the auxiliary spectra. We write the auxiliary spectra in the following form:

$$\gamma_1 = -A\sin(2A^2t) + \epsilon\gamma_{11}, \tag{9.38.a}$$

$$\gamma_2 = d + \epsilon\gamma_{21}, \tag{9.38.b}$$

$$\eta_1 = iA\cos(2A^2t) + \epsilon\eta_{11}, \tag{9.38.c}$$

$$\eta_2 = d + \epsilon\eta_{21}, \tag{9.38.d}$$

where $\epsilon=0$ corresponds to the 1-band solution. γ_{11}, γ_{21}, η_{11}, and η_{21} are functions of x and t, and their values at $t=0$ and at various x are given by (9.35). The stability analysis corresponds to determining when γ_{11}, γ_{21}, η_{11}, and η_{21} grow without bound. We note that γ_{11} and γ_{21} are of order $1/d$ whereas $\eta_{11}=0$ and $\eta_{21}=i\cos 2\sqrt{d^2+A^2}$ at $t=0$.

From the equation (8.6) for $d\gamma_2/dt$ and $d\eta_2/dt$ we obtain to order ϵ

$$\frac{d\gamma_{21}}{dt} = 4i\sigma_2 d\sqrt{(1+\gamma_{21}^2)(d^2+A^2)} - \frac{2i\sigma_2 A^2\sqrt{(1+\gamma_{21}^2)(d^2+A^2)}}{d+iA\sin(2A^2t)}, \tag{9.39.a}$$

$$\frac{d\eta_{21}}{dt} = 4i\hat{\sigma}_2 d\sqrt{(1+\eta_{21}^2)(d^2+A^2)} - \frac{2i\hat{\sigma}_2 A^2\sqrt{(1+\eta_{21}^2)(d^2+A^2)}}{d-iA\cos(2A^2t)}. \tag{9.39.b}$$

From the equations for $d\gamma_1/dt$ and $d\eta_1/dt$ we obtain σ_1 and $\hat{\sigma}_1$:

$$\sigma_1 = -1 \quad \text{when} \quad 0 \leq 2A^2t < \frac{\pi}{2} \text{ or } \frac{3\pi}{2} \leq 2A^2t < 2\pi, \tag{9.40.a}$$

$$\sigma_1 = 1 \quad \text{when} \quad \frac{\pi}{2} \leq 2A^2t < \frac{3\pi}{2}, \tag{9.40.b}$$

$$\hat{\sigma}_1 = -1 \quad \text{when} \quad 0 \leq 2A^2t < \pi, \tag{9.41.a}$$

$$\hat{\sigma}_1 = 1 \quad \text{when} \quad \pi \leq 2A^2t < 2\pi. \tag{9.41.b}$$

We can integrate Equations (9.39.a) and (9.39.b) directly, and we find

$$\frac{\gamma_{21}(t)+\sqrt{1+\gamma_{21}^2(t)}}{\gamma_{21}(0)+\sqrt{1+\gamma_{21}^2(0)}} = \exp\left(4i\sigma_2 d\sqrt{d^2+A^2}\right) \times \exp\left(-i\sigma_2 \tan^{-1}\frac{d}{\sqrt{d^2+A^2}}\tan(2A^2t)\right) \times \left(\frac{\sqrt{d^2+A^2}+A\cos(2A^2t)}{\sqrt{d^2+A^2}-A\cos(2A^2t)}\,\frac{\sqrt{d^2+A^2}-A}{\sqrt{d^2+A^2}+A}\right)^{\sigma_2/2}, \tag{9.42.a}$$

$$\frac{\eta_{21}(t)+\sqrt{1+\eta_{21}^2(t)}}{\eta_{21}(0)+\sqrt{1+\eta_{21}^2(0)}} = \left(\frac{\sqrt{d^2+A^2}+A\sin(2A^2t)}{\sqrt{d^2+A^2}-A\sin(2A^2t)}\right)^{\hat{\sigma}_2/2} \times \exp\left[4i\hat{\sigma}_2 d\sqrt{d^2+A^2}\,t - i\hat{\sigma}_2 \times \tan^{-1}\left(\frac{d}{\sqrt{d^2+A^2}}\tan(2A^2t)\right)\right]. \tag{9.42.b}$$

From Equation (9.42) we find η_{21} and γ_{21} are bounded for all time for finite d. The only instability that could possibly occur is the case where $d\to 0$ and $2A^2t\to 3\pi/2$. Equations (9.39) and (9.42) are valid when $\epsilon \ll d^2$ as $d\to 0$. In this case we find

$$\gamma_{21}(t=0)+\sqrt{1+\gamma_{21}^2(t=0)} \sim \frac{2A}{d}\exp\left(2i\sqrt{d^2+A^2}\right)$$

$$\text{when} \quad 0 \leqslant 2\sqrt{d^2+A^2}\,x < \frac{\pi}{2} \quad \text{or} \quad \frac{3\pi}{2} \leqslant 2\sqrt{d^2+A^2}\,x < 2\pi, \tag{9.43.a}$$

and

$$\gamma_{21}(t=0)+\sqrt{1+\gamma_{21}^2(t=0)} \sim \frac{-d}{2A}\left(\cos 2\sqrt{d^2+A^2}\,x - i\sin 2\sqrt{d^2+A^2}\,x\right)$$

$$\text{when} \quad \frac{\pi}{2} \leqslant 2\sqrt{d^2+A^2} < \frac{3\pi}{2}. \tag{9.43.b}$$

It is straightforward to verify that $\gamma_{21}(t)\to\pm i$ as $2A^2t\to\pi/2$ and hence $\sqrt{1+\gamma_{21}^2(t)}\to 0$ as $2A^2t\to\pi/2$. σ_2 will thus change sign as $2A^2t\to\pi/2$. We conclude that at time $2A^2t=3\pi/2$, σ_2 behaves as

$$\sigma_2=-1 \qquad \text{when} \quad \frac{3\pi}{2}\leqslant 2\sqrt{d^2+A^2}<2\pi \text{ or } 0\leqslant 2\sqrt{a^2+A^2}<\frac{\pi}{2}. \tag{9.44.a}$$

$$\sigma_2=1 \qquad \text{when} \quad \pi\leqslant 2\sqrt{d^2+A^2}<\frac{3\pi}{2}. \tag{9.44.b}$$

We note that since $\sigma_2(t)$ changes sign at $t=\pi/4A^2$, the relevant equation for γ_{21} in the neighborhood of $2A^2t\sim 3\pi/2$ is

$$\frac{\gamma_{21}(t)+\sqrt{1+\gamma_{21}^2(t)}}{\gamma_{21}(t=\pi/4A^2)+\sqrt{1+\gamma_{21}^2(t=\pi/4A^2)}}$$

$$=\left(\frac{\sqrt{d^2+A^2}+A\cos(2A^2t)}{\sqrt{d^2+A^2}-A\cos(2A^2t)}\right)^{\sigma_2/2}$$

$$\times\exp\left\{4i\sigma_2 d\sqrt{d^2+A^2}\left(t-\frac{\pi}{4A^2}\right)\right.$$

$$\left.-i\sigma_2\left[\tan^{-1}\left(\frac{d}{\sqrt{d^2+A^2}}\tan 2A^2t\right)-\frac{\pi}{2}\right]\right\}, \tag{9.45}$$

where σ_2 is given in (9.44). Since $\gamma_{21}(t=\pi/4A^2)\sim i$, we find that $\gamma_{21}(t\sim 3\pi/4A^2)\sim 1/d$ and thus $\gamma_2=d+O(\epsilon/d)$. Similar arguments hold for η_{21}, and there is no instability associated with them. We conclude that a 1-band potential is stable with respect to any small 2-band perturbation. This agrees with the linear stability analysis (9.34.c), since a 2-band perturbation has a large enough wave number.

5. Discussion of the focusing case

In Part II of this paper we have given a method to integrate the N-band potential (i.e., the periodic potential that has $2N$ simple roots of $1-a_R^2=0$) under suitable conditions on the scattering data. The method is the same as given in Part I; i.e., the original nonlinear partial differential equation is replaced by $2N$ coupled nonlinear ordinary differential equations provided the spectra have certain well-behaved properties. Although we have shown explicitly that a 1-band potential is

stable with respect to any 2-band perturbation, we have not proven stability in general.

Appendix

(a) $\zeta_i (i=1,2,\ldots,\infty)$ *are real.*

The proof is based on the fact that the operator in (1.1) is Hermitian:

$$v_1^*(\zeta,x)\cdot\left[\frac{\partial v_1}{\partial x}+i\zeta v_1-qv_2\right]-v_2^*(\zeta,x)\cdot\left[\frac{\partial v_2}{\partial x}-i\zeta v_2-q^*v_1\right]+\text{c.c.}$$

$$=\frac{\partial}{\partial x}\left(|v_1(x,\zeta)|^2-|v_2(x,\zeta)|^2\right)$$

$$+i(\zeta-\zeta^*)\left(|v_1(x,\zeta)|^2+|v_2(x,\zeta)|^2\right)$$

$$=0,$$

or

$$\left(|v_1(x,\zeta)|^2-|v_2(x,\zeta)|^2\right)\Big|_{x_0}^{x_0+T}$$

$$+i(\zeta-\zeta^*)\int_{x_0}^{x_0+T}\left(|v_1(x,\zeta)|^2+|v_2(x,\zeta)|^2\right)dx=0.$$

Evaluate ζ at ζ_i'; the first term vanishes due to the periodicity or antiperiodicity of the eigenfunctions. Since

$$\int_{x_0}^{x_0+T}\left(|v_1(x,\zeta_i')|^2+|v_2(x,\zeta_i')|^2\right)dx>0,$$

we conclude $\zeta_i'=\zeta_i'^*$, or ζ_i' $(i=1,2,\ldots,\infty)$ are real.

(b) γ_i' *(or* η_i'*)* $(i=1,2,\ldots,\infty)$ *are real.*

Define

$$Y(\zeta,x)=v_{1R}-v_{2R}, \tag{A.1.a}$$

$$Z(\zeta,x)=v_{1I}+v_{2I}. \tag{A.1.b}$$

Equations (1.1) can be written

$$\frac{\partial Z}{\partial x}+\zeta Y=q_RZ-q_IY, \tag{A.2.a}$$

$$\frac{\partial Y}{\partial x}-\zeta Z=-q_RY-q_IZ, \tag{A.2.b}$$

where $Z(x_0,\gamma_i')=Z(x_0+T,\gamma_i')=0$ from the definition of γ_i'.

Using (A.2), it is straightforward to show

$$\frac{\partial}{\partial x}(YZ^* - Y^*Z) + (\zeta^* - \zeta)(|Y|^2 + |Z|^2) = 0;$$

hence,

$$(\zeta^* - \zeta)\int_{x_0}^{x_0+T}(|Y|^2 + |Z|^2)\,dx + (YZ^* - Y^*Z)\Big|_{x_0}^{x_0+T} = 0.$$

Evaluate ζ at γ_i', the second term in the above equation vanishes. Since $\int_{x_0}^{x_0+T}(|Y|^2+|Z|^2)\,dx>0$ ($Y=1$ at $x=x_0$), we conclude that $\gamma_i' = \gamma_i'^*$, i.e., γ_i' ($i=1,2,\ldots$) are real.

(c) *At any* $\hat{\zeta}_i$ ($i=1,2,\ldots$), $a_I = b_R = b_I = 0$.

From (A.2) we choose two sets of solutions with the initial conditions

$$\begin{Bmatrix} Z_1(x_0)=1 \\ Y_1(x_0)=0 \end{Bmatrix} \quad \text{and} \quad \begin{Bmatrix} Z_2(x_0)=0 \\ Y_2(x_0)=1 \end{Bmatrix}.$$

From $(\partial/\partial x)(Z_1Y_2 - Z_2Y_1)=0$ we obtain

$$Z_1(x)Y_2(x) - Z_2(x)Y_1(x) = Z_1(x_0)Y_2(x_0) - Z_2(x_0)Y_1(x_0) = 1.$$

Hence,

$$\begin{pmatrix} Z_1 \\ Y_1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} Z_2 \\ Y_2 \end{pmatrix}$$

are two sets of linearly independent solutions.

We expand the Bloch eigenfunction in terms of $\begin{pmatrix} Z_1 \\ Y_1 \end{pmatrix}$ and $\begin{pmatrix} Z_2 \\ Y_2 \end{pmatrix}$:

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = c_1\begin{pmatrix} Z_1 \\ Y_1 \end{pmatrix} + c_2\begin{pmatrix} Z_2 \\ Y_2 \end{pmatrix}.$$

We can also expand $\begin{pmatrix} Z_1(x+T) \\ Y_1(x+T) \end{pmatrix}$ in terms of $\begin{pmatrix} Z_i(x) \\ Y_i(x) \end{pmatrix}$ ($i=1,2$), due to the periodicity of the potential. The coefficients of expansion can be evaluated by setting $x=x_0$. We then have

$$\begin{pmatrix} Z_1(x+T) \\ Y_1(x+T) \end{pmatrix} = Z_1(x_0+T)\begin{pmatrix} Z_1(x) \\ Y_1(x) \end{pmatrix} + Y_1(x_0+T)\begin{pmatrix} Z_2(x) \\ Y_2(x) \end{pmatrix},$$

and similarly,

$$\begin{pmatrix} Z_2(x+T) \\ Y_2(x+T) \end{pmatrix} = Z_2(x_0+T)\begin{pmatrix} Z_1(x) \\ Y_1(x) \end{pmatrix} + Y_2(x_0+T)\begin{pmatrix} Z_2(x) \\ Y_2(x) \end{pmatrix}.$$

Thus

$$\begin{aligned} \begin{pmatrix} \psi_1(x+T) \\ \psi_2(x+T) \end{pmatrix} = \lambda\begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix} &= \lambda c_1\begin{pmatrix} Z_1(x) \\ Y_1(x) \end{pmatrix} + \lambda c_2\begin{pmatrix} Z_2(x) \\ Y_2(x) \end{pmatrix} \\ &= c_1\begin{pmatrix} Z_1(x+T) \\ Y_1(x+T) \end{pmatrix} + c_2\begin{pmatrix} Z_2(x+T) \\ Y_2(x+T) \end{pmatrix} \\ &= c_1\begin{pmatrix} Z_1(x_0+T)Z_1(x)+Y_1(x_0+T)Z_2(x) \\ Z_1(x_0+T)Y_1(x)+Y_1(x_0+T)Y_2(x) \end{pmatrix} \\ &\quad + c_2\begin{pmatrix} Z_2(x_0+T)Z_1(x)+Y_2(x_0+T)Z_2(x) \\ Z_2(x_0+T)Y_1(x)+Y_2(x_0+T)Y_2(x) \end{pmatrix}. \end{aligned}$$

Hence,

$$\begin{aligned} c_1(Z_1(x_0+T)-\lambda)+c_2Z_2(x_0+T) &= 0, \\ c_1Y_1(x_0+T)+c_2(Y_2(x_0+T)-\lambda) &= 0. \end{aligned}$$

For nontrivial c_1 and c_2 we require

$$\begin{aligned} &\lambda^2-[Z_1(x_0+T)+Y_2(x_0+T)]\lambda \\ &\quad + Z_1(x_0+T)Y_2(x_0+T)-Z_2(x_0+T)Y_1(x_0+T)=0. \end{aligned}$$

Using the fact that $Z_1(x_0+T)Y_2(x_0+T)-Z_2(x_0+T)Y_1(x_0+T)=1$, we have

$$\lambda^2-[Z_1(x_0+T)+Y_2(x_0+T)]\lambda+1=0.$$

Hence from (1.6.e)

$$a_R=\frac{Z_1(x_0+T,\xi)+Y_2(x_0+T,\xi)}{2}.$$

Defining

$$y_1=\frac{\partial}{\partial\xi}Y_1,\quad z_1=\frac{\partial}{\partial\xi}Z_1,\quad y_2=\frac{\partial}{\partial\xi}Y_2,\quad z_2=\frac{\partial}{\partial\xi}Z_2$$

and differentiating (A.2) with respect to ζ, we get

$$\frac{\partial z}{\partial x} + \zeta y - q_R z + q_I y = -Y, \tag{A.3.a}$$

$$\frac{\partial y}{\partial x} - \zeta z + q_R y + q_I = Z. \tag{A.3.b}$$

$\begin{pmatrix} z \\ y \end{pmatrix}$ satisfies the same equation as does $\begin{pmatrix} Z \\ Y \end{pmatrix}$ except that there is a forcing term on the right hand side. Take $z=z_1$, $y=y_1$, $Y=Y_1$, $Z=Z_1$ in (A.3) and expand $\begin{pmatrix} z_1 \\ y_1 \end{pmatrix}$ in terms of $\begin{pmatrix} Z_1 \\ Y_1 \end{pmatrix}$ and $\begin{pmatrix} Z_2 \\ Y_2 \end{pmatrix}$. We obtain

$$z_1 = c_1 Z_1 + c_2 Z_2,$$

$$y_1 = c_1 Y_1 + c_2 Y_2,$$

where c_1 and c_2 are functions of x. From the variation of parameters in the differential equation we have

$$c_1 = -\int_{x_0}^{x} (Y_1 Y_2 + Z_1 Z_2)\, dx + c_{10},$$

$$c_2 = \int_{x_0}^{x} (Y_1^2 + Z_1^2)\, dx + c_{20}.$$

Evaluating x at x_0, we have $c_{10} = z_1(x_0) = 0$, $c_{20} = y_1(x_0) = 0$. Then

$$z_1(x) = \left(-\int_{x_0}^{x} (Y_1 Y_2 + Z_1 Z_2)\, dx \right) Z_1(x) + \left(\int_{x_0}^{x} (Z_1^2 + Y_1^2)\, dx \right) Z_2(x), \tag{A.4.a}$$

$$y_1(x) = \left(-\int_{x_0}^{x} (Y_1 Y_2 + Z_1 Z_2)\, dx \right) Y_1(x) + \left(\int_{x_0}^{x} (Z_1^2 + Y_1^2)\, dx \right) Y_2(x). \tag{A.4.b}$$

By the same reasoning,

$$z_2(x) = \left(-\int_{x_0}^{x} (Y_2^2 + Z_2^2)\, dx \right) Z_1(x) + \left(\int_{x_0}^{x} (Z_1 Z_2 + Y_1 Y_2)\, dx \right) Z_2(x), \tag{A.5.a}$$

$$y_2(x) = \left(-\int_{x_0}^{x} (Y_2^2 + Z_2^2)\, dx \right) Y_1(x) + \left(\int_{x_0}^{x} (Z_1 Z_2 + Y_1 Y_2)\, dx \right) Y_2(x). \tag{A.5.b}$$

Differentiating a_R with respect to ζ,

$$2\frac{da_R}{d\zeta} = z_1(x_0+T,\zeta)+y_1(x_0+T,\zeta)$$

$$= \left(-\int_{x_0}^{x_0+T}(Y_1Y_2+Z_1Z_2)\,dx\right)Z_1(x_0+T,\zeta)$$

$$+Z_2(x_0+T,\zeta)\int_{x_0}^{x_0+T}(Z_1^2+Y_1^2)\,dx$$

$$-Y_1(x_0+T,\zeta)\int_{x_0}^{x_0+T}(Y_2^2+Z_2^2)\,dx$$

$$+Y_2(x_0+T,\zeta)\int_{x_0}^{x_0+T}(Z_1Z_2+Y_1Y_2)\,dx.$$

We assume for the present that $Z_2(x_0+T,\zeta)\neq 0$, and obtain

$$Y_1(x_0+T,\zeta) = \operatorname{sgn} Z_2(x_0+T,\zeta)\cdot\frac{4(a_R^2-1)-[Z_1(x_0+T,\zeta)-Y_2(x_0+T,\zeta)]^2}{4|Z_2(x_0+T,\zeta)|},$$

where

$$\operatorname{sgn} Z_0(x_0+T,\zeta) = \begin{cases} 1 & \text{if } Z_0(x_0+T,\zeta)>0, \\ -1 & \text{if } Z_0(x_0+T,\zeta)<0. \end{cases}$$

To evaluate $da_R/d\zeta$ we write

$$-Y_1(x_0+T,\zeta)\int_{x_0}^{x_0+T}Z_2^2\,dx = -\operatorname{sgn} Z_2(x_0+T,\zeta)\frac{(a_R^2-1)}{|Z_2(x_0+T,\zeta)|}\int_{x_0}^{x_0+T}Z_2^2\,dx$$

$$+\operatorname{sgn} Z_2(x_0+T,\zeta)$$

$$\times\frac{[Z_1(x_0+T,\zeta)-Y_2(x_0+T,\zeta)]^2}{4|Z_2(x_0+T,\zeta)|}\int_{x_0}^{x_0+T}Z_2^2\,dx$$

and

$$Z_2(x_0+T,\zeta)\int_{x_0}^{x_0+T}Z_1^2\,dx = \operatorname{sgn} Z_2(x_0+T,\zeta)\cdot|Z_2(x_0+T,\zeta)|$$

$$\times\int_{x_0}^{x_0+T}Z_1^2\,dx\,[\operatorname{sgn} Z_2(x_0+T,\zeta)]^2.$$

After some straightforward calculation we obtain

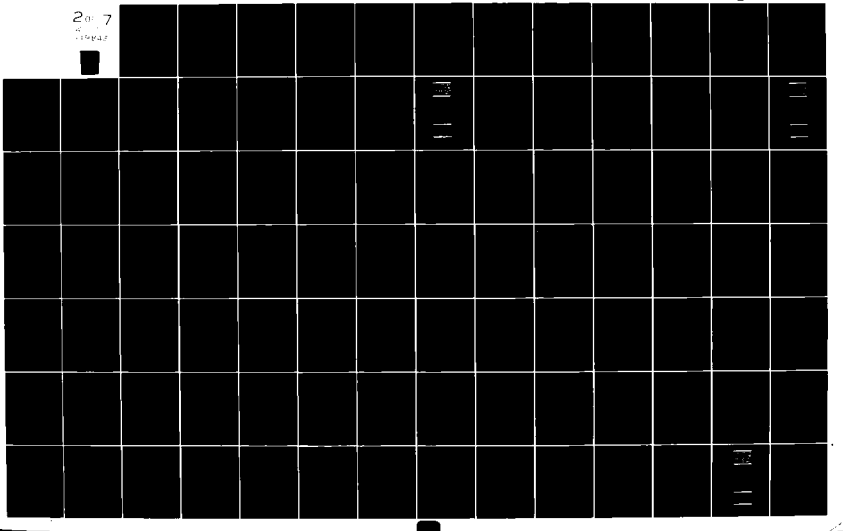
$$2\frac{da_R}{d\zeta} = \operatorname{sgn} Z_2(x_0+T,\zeta)\Bigg[\int_{x_0}^{x_0+T}\Bigg(\frac{Z_1(x_0+T,\zeta)-Y_2(x_0+T,\zeta)}{2\sqrt{|Z_2(x_0+T,\zeta)|}}Z_2 - \operatorname{sgn} Z_2(x_0+T,\zeta)\cdot\sqrt{|Z_2(x_0+T,\zeta)|}\,Z_1\Bigg)^2 dx + \int_{x_0}^{x_0+T}\Bigg(\frac{Z_1(x_0+T,\zeta)-Y_2(x_0+T,\zeta)}{2\sqrt{|Z_2(x_0+T,\zeta)|}}Y_2 - \operatorname{sgn} Z_2(x_0+T,\zeta)\cdot\sqrt{|Z_2(x_0+T,\zeta)|}\,Y_1\Bigg)^2 dx - \frac{a_R^2-1}{|Z_2(x_0+T,\zeta)|}\int_{x_0}^{x_0+T}(Y_2^2+Z_2^2)\,dx\Bigg]. \tag{A.6}$$

We now prove $Z_2(x_0+T,\hat{\zeta}_i)=0$. At $\zeta=\hat{\zeta}_i$, we have $1-a_R^2(\hat{\zeta}_i)=0$ and $da_R/d\zeta|_{\hat{\zeta}_i}=0$. Assuming $Z_2(x_0+T,\hat{\zeta}_i)\neq 0$, we can use (A.6) to obtain $da_R/d\zeta|_{\hat{\zeta}_i}$:

$$\frac{da_R}{d\zeta}\Big|_{\hat{\zeta}_i} = \tfrac{1}{2}\operatorname{sgn} Z_2(x_0+T,\hat{\zeta}_i) \times\Bigg[\int_{x_0}^{x_0+T}\Bigg(\frac{Z_1(x_0+T,\hat{\zeta}_i)-Y_2(x_0+T,\hat{\zeta}_i)}{2\sqrt{|Z_2(x_0+T,\hat{\zeta}_i)|}}Z_2 - \operatorname{sgn} Z_2(x_0+T),\hat{\zeta}_i)\sqrt{|Z_1(x_0+T,\hat{\zeta}_i)|}\,Z_1\Bigg)^2 dx + \int_{x_0}^{x_0+T}\Bigg(\frac{Z_1(x_0+T,\hat{\zeta}_i)-Y_2(x_0+T,\hat{\zeta}_i)}{2\sqrt{|Z_2(x_0+T,\hat{\zeta}_i)|}}Y_2 - \operatorname{sgn} Z_2(x_0+T,\hat{\zeta}_i)\sqrt{|Z_2(x_0+T,\hat{\zeta}_i)|}\,Y_1\Bigg)^2 dx\Bigg].$$

AD-A119 843 CLARKSON COLL OF TECHNOLOGY POTSDAM NY DEPT OF MATHE--ETC F/G 12/1
NONLINEAR WAVES.(U)
SEP 82 M J ABLOWITZ, D J KAUP, A C NEWELL N00014-76-C-0867

UNCLASSIFIED NL



The factor in the bracket is never less than zero. It is zero only when

$$\begin{pmatrix} Y_2 \\ Z_2 \end{pmatrix} = \frac{2Z_2(x_0+T)}{Z_1(x_0+T)-Y_2(x_0+T)} \begin{pmatrix} Y_1 \\ Z_1 \end{pmatrix},$$

but this cannot happen, since $\begin{pmatrix} Y_1 \\ Z_1 \end{pmatrix}$ and $\begin{pmatrix} Y_2 \\ Z_2 \end{pmatrix}$ are linearly independent. We then get $da_R/d\zeta|_{\hat{\xi}_i} \neq 0$, which contradicts the fact that $da_R/d\zeta|_{\hat{\xi}_i}=0$. We have thus proved that $Z_2(x_0+T, \hat{\xi}_i)=0$.

Therefore $\hat{\xi}_i$ is also a root of the family γ_i', and $\hat{\gamma}_i$ coincides with $\hat{\xi}_i$. Similarly, $\hat{\xi}_i$ is also a root of η_i', and $\hat{\eta}_i$ coincides with $\hat{\xi}_i$. At $\hat{\xi}_i$, $a_I+b_I=0$, $a_I+b_R=0$, $a_R^2=1$. Hence we conclude that at $\hat{\xi}_i$, $a_I=b_I=b_I=b_R=0$.

(d) $\gamma_i'(\eta_i')$ $(i=1,2,\ldots,\infty)$ *are simple roots.*

From (A.5), $z_2(x_0+T, \gamma_i')=[-\int_{x_0}^{x_0+T}(Y_2^2+Z_2^2)\,dx]Z_1(x_0+T, \gamma_i')$. Since γ_i is a root of Z_2, we have $Z_1(x_0+T, \gamma_i')Y_2(x_0+T, \gamma_i')=1$, which implies that $Z_1(x_0+T, \gamma_i')\neq 0$. Since $\int_{x_0}^{x_0+T}(Y_2^2+Z_2^2)\,dx>0$, we have $z_2(x_0+T, \gamma_i')\neq 0$. We thus conclude γ_i' is a simple root of $Z_2(x_0+T)=0$.

(e) ζ_i'' *is either a simple or a double root of* $1-a_R^2=0$.

To prove the above statement we need to prove that at $\hat{\xi}_i$, $d^2a_R/d\zeta^2|_{\hat{\xi}_i}\neq 0$. We only prove it for the periodic band edges $(\hat{\xi}_i)$; the proof for the antiperiodic band edges follows similarly.

First we claim that $Y_1(x_0+T, \hat{\xi}_i)=0$. The proof is exactly the same as given in [11]. At $\hat{\xi}_i$,

$$Z_1(x_0+T, \hat{\xi}_i)+Y_2(x_0+T, \hat{\xi}_i)=2, \tag{A.7.a}$$

$$Z_2(x_0+T, \hat{\xi}_i)=0. \tag{A.7.b}$$

Hence,

$$Z_1(x_0+T, \hat{\xi}_i)Y_2(x_0+T, \hat{\xi}_i)-Z_2(x_0+T, \hat{\xi}_i)Y_1(x_0+T, \hat{\xi}_i)=1. \tag{A.8}$$

From (A.7) and (A.8) we obtain

$$Z_1(x_0+T, \hat{\xi}_i)=Y_2(x_0+T, \hat{\xi}_i)=1.$$

Hence,

$$z_1(x_0+T,\xi_i) = -\int_{x_0}^{x_0+T}(Z_1Z_2+Y_1Y_2)\,dx,$$

$$z_2(x_0+T,\xi_i) = -\int_{x_0}^{x_0+T}(Y_2^2+Z_2^2)\,dx,$$

$$y_1(x_0+T,\xi_i) = \int_{x_0}^{x_0+T}(Y_1^2+Z_1^2)\,dx,$$

$$y_2(x_0+T,\xi_i) = \int_{x_0}^{x_0+T}(Z_1Z_2+Y_1Y_2)\,dx,$$

$$2\frac{d^2a_R}{d\zeta^2}\bigg|_{\xi_i} = 2\left\{\left[\int_{x_0}^{x_0+T}(Z_1Z_2+Y_1Y_2)\,dx\right]^2 - \int_{x_0}^{x_0+T}(Z_2^2+Y_2^2)\,dx\cdot\int_{x_0}^{x_0+T}(Z_1^2+Y_1^2)\,dx\right\} \leqslant 0.$$

The last inequality is the Cauchy-Schwarz inequality. It is clear that the equality sign occurs when

$$\begin{pmatrix} Z_1 \\ Y_1 \end{pmatrix} = c\begin{pmatrix} Z_2 \\ Y_2 \end{pmatrix}$$

with c a constant. This can never happen, since $\begin{pmatrix} Z_2 \\ Y_2 \end{pmatrix}$ and $\begin{pmatrix} Z \\ Y_1 \end{pmatrix}$ are linearly independent. Hence we have $d^2a_R/d\zeta^2|_{\xi_i} < 0$, and thus the roots of $1-a_R^2=0$ are either simple or double.

(f) *There is only one γ_i (or η_i) in each unstable band.*

First we prove that there will be at least one γ_i in each unstable band. At two adjacent band edges ξ_i and ξ_{i+1} that are both periodic or antiperiodic, $da_R/d\zeta|_{\xi=\xi_i}$ and $da_R/d\zeta|_{\xi=\xi_{i+1}}$ have *different signs*. From (A.6) this means that $Z_2(x_0+T,\xi_i)$ and $Z_2(x_0+T,\xi_{i+1})$ have different signs. From continuity there must be a point γ_i in between ξ_i and ξ_{i+1} such that $Z_2(x_0+T,\gamma_i)=0$. We thus proved that there will be at least one γ_i in each unstable band.

Next we prove that there will be at most one γ_i in each unstable band. At γ_i, $Z_2(x_0+T,\gamma_i)=0$. From $Y_2(x_0+T,\gamma_i)Z_1(x_0+T,\gamma_i)-Y_1(x_0+T,\gamma_i)Z_2(x_0+T,\gamma_i)=1$, we have

$$Z_1(x_0+T,\gamma_i) = \frac{1}{Y_2(x_0+T,\gamma_i)}$$

and

$$2a_R(\gamma_i) = Y_2(x_0+T,\gamma_i) + Z_1(x_0+T,\gamma_i)$$
$$= Y_2(x_0+T,\gamma_i) + \frac{1}{Y_2(x_0+T,\gamma_i)}.$$

Hence,

$$Y_2(x_0+T,\gamma_i) = a_R(\gamma_i) + \sqrt{a_R^2(\gamma_i)-1}\,.$$

It is clear that $Y_2(x_0+T,\gamma_i)$ can have only a definite sign in each unstable band. $Y_2(x_0+T,\gamma_i)>0$ if $a_R(\gamma_i)>1$, and $Y_2(x_0+T,\gamma_i)<0$ if $a_R(\gamma_i)<-1$.

From (A.5),

$$z_2(x_0+T,\gamma_i) = -Z_1(x_0+T,\gamma_i)\int_{x_0}^{x_0+T}(Y_2^2+Z_2^2)\,dx$$
$$= -\frac{1}{Y_2(x_0+T,\gamma_i)}\int_{x_0}^{x_0+T}(Y_2^2+Z_2^2)\,dx.$$

Since the integral is positive, this implies that $z_2(x_0+T,\gamma_i)$ can only have a definite sign in each unstable band, so that $Z_2(x_0+T,\gamma_i)$ is monotonic in one band. This proves that there will be at most one γ_i in each unstable band.

(g) *$da_R/d\zeta \neq 0$ for $|a_R|<1$.*

Consider a periodic band edge ζ_i'. If at ζ_i', $a_R(\zeta_i')=1$ and $da_R/d\zeta|_{\zeta_i'}<0$, then $a_R(\zeta)$ is a decreasing function of ζ for $\zeta_i'<\zeta<\zeta_i'+\sigma$ where σ is a small positive number. If at ζ_i', $a_R(\zeta_i')=1$ and $da_R/d\zeta|_{\zeta_i'}=0$, then $a_R(\zeta)$ is still a decreasing function of ζ for $\zeta_i'<\zeta<\zeta_i'+\sigma$ where σ is a small positive number, since $a_R''(\zeta_i')<0$. Now there exists a μ^* ($\mu^*<\zeta_i'$) such that $1>a_R(\mu^*)>-1$. We show that $da_R/d\zeta|_{\zeta=\mu^*}\neq 0$:

$$4\left[a_R^2(\mu^*)-1\right] = \left[Z_1(x_0+T,\mu^*)-Y_2(x_0+T,\mu^*)\right]^2$$
$$+4Z_2(x_0+T,\mu^*)Y_1(x_0+T,\mu^*)$$
$$<0 \qquad [\text{since} \quad -1<a_R(\mu^*)<1].$$

Hence, $Z_2(x_0+T,\mu^*)Y_1(x_0+T,\mu^*)<0$, so that $Z_2(x_0+T,\mu^*)\neq 0$. From (A.6), $da_R/d\zeta|_{\mu^*}\neq 0$. Thus $da_R/d\zeta|_{\mu^*}<0$ in any open interval $\zeta_i'<\zeta<\mu^*$ where $1>a_R(\mu^*)>-1$; this completes the proof.

(h) *$a(\zeta)$ and $b(\zeta)$ are entire functions of ζ.*

It has been shown in [1] that $a_1(\zeta)$ and $b_1(\zeta)$ are entire functions of ζ for the cubic Schrödinger equation on the infinite interval when $q=0$ as $|x|>M$ (M is some positive constant). The Jost functions (ϕ', $\bar{\phi}'$, ψ', $\bar{\psi}'$) and $a_1(\zeta)$ and $b_1(\zeta)$ are

defined as follows:

$$\phi'(x,\zeta) \sim e^{-i\zeta x}\begin{pmatrix}1\\0\end{pmatrix} \qquad \text{as} \quad x \to -\infty,$$

$$\bar{\phi}'(x,\zeta) \sim e^{i\zeta x}\begin{pmatrix}0\\-1\end{pmatrix} \qquad \text{as} \quad x \to -\infty,$$

$$\psi'(x,\zeta) \sim e^{i\zeta x}\begin{pmatrix}0\\1\end{pmatrix} \qquad \text{as} \quad x \to -\infty,$$

$$\bar{\psi}'(x,\zeta) \sim e^{-i\zeta x}\begin{pmatrix}1\\0\end{pmatrix} \qquad \text{as} \quad x \to \infty,$$

$$\phi'(x,\zeta) = a_1(\zeta)\bar{\psi}'(x,\zeta) + b_1(\zeta)\psi'(x,\zeta).$$

To obtain $a(\zeta)$ and $b(\zeta)$ in the periodic problem it would be sufficient to solve (1.1) with the following potential:

$$q(x) = 0 \qquad \text{for} \quad x > x_0 + T \text{ and } x < x_0,$$

$$q(x) = \text{given periodic potential}$$

$$\text{for} \quad x_0 \leq x \leq x_0 + T. \tag{A.9}$$

The reason is that $\phi(x)$ and $\bar{\phi}(x)$ for the potential (A.9) will be the same as the $\phi(x)$ and $\bar{\phi}(x)$ for the periodic potential in the interval $x_0 \leq x \leq x_0 + T$. Hence, we shall obtain the same $a(\zeta)$ and $b(\zeta)$ for potential (A.9) as we do for the given periodic potential. It is straightforward to show that $a(\zeta)$ and $b(\zeta)$ are related to $a_1(\zeta)$ and $b_1(\zeta)$ as follows:

$$a(\zeta) = e^{-i\zeta T}a_1(\zeta), \tag{A.10.a}$$

$$b(\zeta) = e^{i\zeta(x_0+T)}b_1(\zeta). \tag{A.10.b}$$

Since $a_1(\zeta)$ and $b_1(\zeta)$ are entire functions, we conclude that $a(\zeta)$ and $b(\zeta)$ are also entire functions.

It has been shown in [1] that if q is real, $a_1(\zeta) = a_1^*(-\zeta)$. From (A.10.a) we find that $a(\zeta) = a^*(-\zeta^*)$. Thus, the main spectrum appears in positive and negative pairs.

(i) $\sum_{j=1}^{N} \frac{\gamma_j^k}{\prod_{l\neq j}(\gamma_j - \gamma_l)} = \delta_{k,N-1}$, $0 \leq k \leq N-1$, $N \geq 2$.

We can prove this identity by using the residue theorem:

$$\sum_{j=1}^{N} \frac{\gamma_j^k}{\prod_{l\neq j}(\gamma_j - \gamma_l)} = \frac{1}{2\pi i}\oint_C \frac{\gamma^k\, d\gamma}{\prod_{l=1}^{N}(\gamma - \gamma_l)} = \frac{1}{2\pi i}\oint_O \frac{\gamma^k\, d\gamma}{\prod_{l=1}^{N}(\gamma - \gamma_l)}.$$

where C is a closed contour including all γ_j, and O is a closed circle with center at the origin of the γ plane and with infinite radius ($R\to\infty$). Then

$$\frac{1}{2\pi i}\int \frac{\gamma^k d\gamma}{\prod_{l=1}^{N}(\gamma-\gamma_l)} = \frac{1}{2\pi i}\int_0^{2\pi} \frac{R^{k+1}e^{i(k+1)\theta} i d\theta}{R^N e^{iN\theta}}$$

$$\to \begin{cases} 0 & \text{if } k<N-1, \\ 1 & \text{if } k=N-1. \end{cases}$$

Hence

$$\sum_{j=1}^{N} \frac{\gamma_j^k}{\prod_{l\neq j}(\gamma_j-\gamma_l)} = \delta_{k,N-1}, \qquad 0\leqslant k\leqslant N-1, \quad N\geqslant 2.$$

(j) $$\sum_{j=1}^{N} \frac{\gamma_j^N}{\prod_{l\neq j}(\gamma_j-\gamma_l)} = \sum_{j=1}^{N}\gamma_j.$$

The above identity is true for $N=1$ and $N=2$. Assuming it is true for N, we prove it also true for $N+1$:

$$\sum_{j=1}^{N+1} \frac{\gamma_j^{N+1}}{\prod_{l\neq j}(\gamma_j-\gamma_l)} = \sum_{j=1}^{N} \frac{\dfrac{\gamma_j^{N+1}}{\gamma_j-\gamma_{N+1}}}{\prod_{\substack{l\neq j \\ l\neq N+1}}(\gamma_j-\gamma_l)} + \frac{(\gamma_{N+1})^{N+1}}{\prod_{l\neq N+1}(\gamma_{N+1}-\gamma_l)}$$

$$= \sum_{j=1}^{N} \frac{\gamma_j^N}{\prod_{\substack{l\neq j \\ l\neq N+1}}(\gamma_j-\gamma_l)} \sum_{j=1}^{N} \frac{\gamma_j^N\gamma_{N+1}}{\prod_{l\neq j}(\gamma_j-\gamma_l)}$$

$$+ \frac{(\gamma_{N+1})^{N+1}}{\prod_{l\neq N+1}(\gamma_{N+1}-\gamma_l)}$$

$$= \sum_{j=1}^{N}\gamma_j + \gamma_{N+1}\sum_{j=1}^{N+1} \frac{\gamma_j^N}{\prod_{l\neq j}(\gamma_j-\gamma_l)}$$

$$= \sum_{j=1}^{N}\gamma_j + \gamma_{N+1} = \sum_{j=1}^{N+1}\gamma_j.$$

We have thus proved the identity by induction.

(k) *In the focusing case not all ξ_l are double zeros and not all ζ_l are simple zeros of $1-a_R^2=0$.*

We work out one example to show that ζ_i can be a higher order zero. Consider on 2-band potential satisfying the differential equation

$$c_0 q + 2ic_1 q_x - c_2 q_{xx} - 2c_2 q^2 q^* = 0 \tag{A.11}$$

(according to $\sum_{m=0}^{2} c_m \delta I_m / \delta q^* = 0$). λ and μ can be calculated by using (7.3):

$$\lambda(\zeta) = \frac{i}{2} c_0 + 2i\zeta c_1 + c_2(2i\zeta^2 - i|q|^2),$$

$$\mu(\zeta) = c_1 q^* + ic_2 \frac{\partial q^*}{\partial x}(x_0) - 2\zeta c_2 q^*(x_0).$$

From (A.11) there are two quantities independent of x which are related to the roots of $1 - a_R^2 = 0$:

$$2ic_1|q|^2 - c_2(q^* q_x - q q_x^*) = iA', \tag{A.12.a}$$

$$c_0|q|^2 - c_2|q_x|^2 - c_2|q|^4 = B'. \tag{A.12.b}$$

Then the complex roots of $1 - a_R^2 = 0$ can be calculated by using (7.5):

$$\zeta^4 + \frac{2c_1}{c_2}\zeta^3 + \left(\frac{c_0}{2c_2} + \frac{c_1^2}{c_2^2}\right)\zeta^2 + \zeta\left(\frac{A'}{2c_2} + \frac{c_0 c_1}{2c_2^2}\right)$$

$$+ \frac{1}{4c_2^2}\left(\frac{c_0^2}{4} + 2c_1 A' - c_2 B'\right) = 0. \tag{A.13}$$

Therefore, if there is a pair of complex double roots (ζ_1, ζ_1^*) where $\zeta_1 = \zeta_R + i\zeta_I$, then (A.13) can be written as $(\zeta - \zeta_1)^2(\zeta - \zeta_1^*)^2 = 0$, or as

$$\zeta^4 + 4\zeta_R^2\zeta^2 + (\zeta_R^2 + \zeta_I^2) - 4\zeta^3\zeta_R + 2\zeta^2(\zeta_R^2 + \zeta_I^2) - 4\zeta\zeta_R(\zeta_R^2 + \zeta_I^2) = 0.$$

Comparing the above equation with (A.13), we find

$$4\zeta_R = -2c_1/c_2,$$

$$4\zeta_R^2 + 2(\zeta_R^2 + \zeta_I^2) = \frac{c_0}{2c_2} + \frac{c_1^2}{c_2^2},$$

$$-4\zeta_R(\zeta_R^2 + \zeta_I^2) = \frac{A'}{2c_2} + \frac{c_0 c_1}{2c_2^2},$$

$$(\zeta_R^2 + \zeta_I^2)^2 = \frac{1}{4c_2^2}\left(\frac{c_0^2}{4} + 2c_1 A' - c_2 B'\right).$$

Solving the four equations, we have

$$\zeta_{1R} = \frac{-c_1}{2c_2}, \quad \zeta_{1I}^2 = \frac{c_0c_2 - c_1^2}{4c_2^2}, \quad A' = 0, \quad B' = 0.$$

We therefore conclude that a complex double zero is possible and will occur in the 2-band case if $A'=0$, $B'=0$, and $c_0c_2 - c_1^2 > 0$. Also, if $A'=0$, $B'=0$, and $c_0c_2 = c_1^2$, we will have a real quadruple zero of $\lambda - a_R^2 = 0$.

References

1. M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, *Studies in Appl. Math.* 53:249 (1974).
2. C. S. Gardner, J. M. Green, M. D. Kruskal, and R. M. Miura, *Phys. Rev. Lett.* 19:1095 (1967); *Comm. Pure Appl. Math.* 27:97 (1974).
3. V. E. Zakharov and A. B. Shabat, *Functional Anal. Appl.* 8:226 (1974).
4. S. P. Novikov, *Functional Anal. Appl.* 8:236 (1974).
5. B. A. Dubrovin and S. P. Novikov, *Soviet Physics JETP* 40:1058 (1975).
6. H. P. McKean and P. VanMoerbeke, *Invent. Math.* 30:217 (1975).
7. H. P. McKean and E. Trubowitz, *Comm. Pure Appl. Math.* 29:143 (1976).
8. A. R. Its and V. B. Matveev, *Functional Anal. Appl.* 9:65 (1976).
9. B. A. Dubrovin, V. B. Matveev, and S. P. Novikov, *Russian Math. Surveys* 31:59 (1976).
10. V. B. Matveev, Abelian functions and solitons, to appear.
11. W. Magnus and W. Winkler, *Hill's Equation*, Wiley-Interscience, New York, 1966.
12. B. M. Lake, H. C. Yuen, H. Rungaldier, and W. E. Ferguson, *J. Fluid Mech.* 83:49 (1977).
13. B. A. Dubrovin, *Functional Anal. Appl.* 8, No. 3 (1975).
14. Y. C. Ma, Doctoral Thesis, Princeton Univ., 1977.

TRW Inc.
Clarkson College of Technology

(Received February 4, 1980)

From: NONLINEAR PHENOMENA IN PHYSICS AND BIOLOGY
Edited by Richard H. Enns, Billy L. Jones,
Robert M. Miura, and Sadanand S. Rangnekar
(Plenum Publishing Corporation, 1981)

REMARKS ON NONLINEAR EVOLUTION EQUATIONS AND THE INVERSE SCATTERING TRANSFORM*

Mark J. Ablowitz

Department of Mathematics and Computer Science
Clarkson College of Technology
Potsdam, N.Y. 13676

* This work was partially supported by the Air Force Office of Scientific Research, and the office of Naval Research.

TABLE OF CONTENTS

I. INTRODUCTION 85

II. EXAMPLES USING THE NONLINEAR SCHRÖDINGER EQUATION 86

III. THE INTERMEDIATE LONG-WAVE EQUATION 89

REFERENCES 93

REMARKS ON NONLINEAR EVOLUTION EQUATIONS AND THE INVERSE SCATTERING TRANSFORM

Mark J. Ablowitz

Department of Mathematics and Computer Science
Clarkson College of Technology
Potsdam, N.Y. 13676

I. INTRODUCTION

In recent years there has been considerable attention devoted to a new and rapidly developing area of mathematical physics, namely the Inverse Scattering Transform (I.S.T. for short). This method has allowed us to solve certain physically interesting nonlinear evolution equations. By now there are a number of review articles [for example, see Scott et al, 1973; Miura, 1976; Ablowitz, 1978] on this subject as well as some new books [for example, see Zakharov et al, 1980; Ablowitz and Segur, to appear], all of which contain numerous references.

At this time there are quite a few areas of current interest. Some of these are the following:

"Classical" results and methods: Namely, solitons; direct and inverse scattering; direct methods for finding solutions via Hirota's bilinear forms or direct operations on Gel'fand-Levitan-Marchenko integral equations; asymptotic solutions as $|t| \to \infty$; prolongation structures; periodic and other types of boundary conditions (not the

doubly infinite line); perturbations; discretizations; physical applications; numerical simulations; etc.

Alternative formulations of inverse scattering via Riemann-Hilbert Problems [see Zakharov et al, 1980; Zakharov and Shabat, 1979; Zakharov and Mikhailov, 1978].

Multidimensional problems: Scattering and inverse scattering problems; some multidimensional nonlinear evolution equations solvable by I.S.T. are the Kadomstev-Petviashvilli equation, Davey-Stewartson equation, three wave equations, self-dual Yang-Mills equation and others [for example, see Zakharov et al, 1980; Ablowitz and Segur, to appear; Zakharov and Manakov, 1979; Kaup, 1980; Proc. of Joint U.S.-U.S.S.R. Conf. on Soliton Theory, 1979; Zakharov and Manakov (Crete Conference)].

Ordinary differential equations of Painlevé type; monodromy preserving deformations [for example, see Ablowitz et al, 1978; Sato et al, 1977; Flashka and Newell, to appear].

Direct and inverse scattering problems associated with certain nonlinear singular integro-differential equations, namely the Intermediate Long-Wave equation and Benjamin-Ono equation, [Satsuma et al, 1979; Satsuma and Ablowitz, 1980; Kodama et al, to be published; Nakamura, 1979; Bock and Kruskal, 1979].

In this lecture I intend to review some of the results associated with certain continuous and discrete (differential-difference, partial difference) nonlinear evolution equations solvable by inverse scattering.

II. EXAMPLES USING THE NONLINEAR SCHRÖDINGER EQUATION

The prototype equations we will study are for the nonlinear Schrödinger (NLS) equation.

(a) Partial differential equation,

$$iU_t = U_{xx} \pm 2U^2U^* \tag{2.1}$$

(U^* is the complex conjugate of U).

(b) Differential-difference equations,

$$iU_{nt} = \frac{1}{(\Delta x)^2}(U_{n+1}+U_{n-1}-2U_n) \pm U_nU_n^*(U_{n+1}+U_{n-1}) \ . \tag{2.2}$$

(c) Partial-difference equations,

$$\frac{i\Delta^m U_n^m}{\Delta t} = \frac{1}{2(\Delta x)^2}\left[\left(U_{n+1}^m + U_{n-1}^m \prod_{-\infty}^{n-1} \Lambda_k^m - 2U_n^m\right) + \left(U_{n-1}^{m+1} \prod_{-\infty}^{n} \Lambda_k^m - 2U_n^{m+1} + U_{n-1}^{m+1}\right)\right]$$

$$\pm \frac{1}{4}\left[2U_n^m U_n^{m*} U_{n+1}^{m+1} \prod_{-\infty}^{n} \Lambda_k^m + 2U_{n-1}^m U_n^{m+1} U_n^{m+1*} \prod_{-\infty}^{n} \Lambda_k^m\right.$$

$$+ U_n^m (U_n^{m*} U_{n+1}^m + U_n^{m+1*} U_{n+1}^{m+1}) + U_n^{m+1} (U_{n-1}^m U_n^{m*} + U_{n-1}^{m+1} U_n^{m+1*})$$

$$\left. - U_n^m \sum_{-\infty}^{n+1} \Delta^m S_k^m - U_n^{m+1} \sum_{-\infty}^{n} \Delta^m S_k^{m*}\right], \qquad (2.3)$$

where

$$\Lambda_k^m = \frac{1 \pm (\Delta x)^2 U_k^{m+1} U_k^{m+1*}}{1 \pm (\Delta x)^2 U_k^m U_k^{m*}},$$

$$S_k = U_{k-1}^m U_{k-2}^{m*} + U_k^m U_{k-1}^{m*}.$$

Each of these equations has solitons and an infinite number of conserved quantities, and is solvable by linear integral (for continuous cases) or summation (for discrete cases) equations. Each of the results for the discrete equation relax properly to the continuum limit as $\Delta x \to 0$ and $\Delta t \to 0$ (in the partial difference case). It should be stressed that these are only prototype results. In principle one may get analogous results associated with any continuous equation (e.g., Korteweg-deVries (KdV), modified Korteweg-deVries, etc.). We also remark that recently we have carried out some numerical simulations using Eq. (2.3) as a difference approximation of Eq. (2.1) and have found it to be very accurate and efficient [Taha and Ablowitz, 1980].

The basic point of view is to work with the associated linear scattering problem and its respective discretizations. Namely, for some partial differential equations, such as (2.1), the associated scattering problem is:

$$V_{1x} = -i\zeta V_1 + qV_2$$
$$V_{2x} = i\zeta V_2 + rV_1 \quad , \qquad (2.4)$$

and for differential-difference and partial-difference equations, such as (2.2), (2.3):

$$\begin{aligned} V_{1n+1} &= zV_{1n}+Q_nV_{2n} \\ V_{2n+1} &= \frac{1}{z}V_{2n}+R_nV_{1n} \end{aligned} \qquad (2.5)$$

$(Q_n = \Delta x q_n,\ R_n = \Delta x r_n)$.

Following the methods outlined in Ablowitz [1978] and Ablowitz and Segur [to appear] one can find classes of nonlinear evolution equations (both continuous associated with (2.4) and discrete associated with (2.5) where each equation has solitons, an infinite number of conserved quantities, and is solvable by I.S.T.

Briefly, the main results may be summarized as follows:

(a) For the continuous problem, from initial conditions we may calculate:

$$F(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} b(k,0)e^{-i(kx-\omega(2k)t)}dk - i\sum_{1}^{N} C_j(0)e^{-i(\zeta_j x-\omega(2\zeta_j)t)}, \qquad (2.6a)$$

where $\omega(k)$ is the linearized dispersion relation, obtained by seeking solutions of the form

$$u \sim e^{i(kx-\omega t)}.$$

For example, in the NLS equation $\omega(k) = -k^2$, and from (2.4),

$$\{b(k,0),\ \{C_j(0),\ \zeta_j\}_{j=1}^{N}\}$$

are the required scattering data [for example, see Ablowitz, 1978; Ablowitz and Segur, to appear]. If we can solve the linear integral equation $(y > x)$

$$K(x,y)-F(x+y) \pm \int_x^{\infty}\int_x^{\infty} K(x,s)F^*(s+z)\,F(z+y)dzds = 0, \qquad (2.6b)$$

then $U \equiv r = \pm q^*$ is given by

$$U(x,t) = \pm 2K^*(x,x). \qquad (2.6c)$$

(b) For the discrete problem, the results may be summarized as follows. From the initial data we must calculate:

$$F(n) = \frac{1}{2\pi i}\oint b(z;t/m)z^{n-1}dz - \sum C_j(t/m)z_j^{n-1} \ , \tag{2.7a}$$

where

$$b(z;t/m) = b(z;0) \cdot \begin{Bmatrix} e^{-i\omega(z^2)t} & \text{differential-difference} \\ (\omega(z^2))^m & \text{partial-difference} \end{Bmatrix} ,$$

$$C_j(t/m) = C_j(0) \cdot \begin{Bmatrix} e^{-i\omega(z_j^2)t} \\ (\omega(z_j^2))^m \end{Bmatrix} .$$

From (2.5) the functions $\{b(z,0), \{C_j(0), z_j\}_{j=1}^N\}$ are the required scattering data. Then we must solve the linear summation equation ($\ell > n$)

$$\kappa(n,\ell) - F(n+\ell) \pm \sum_{n',n''=n+1}\sum \kappa(n,n'')F^*(n''+n')F(n'+\ell) = 0 \tag{2.7b}$$

and find the solution ($U_n \equiv R_n/\Delta x = \mp Q_n^*$)

$$U_n = \pm \kappa^*(n,n+1)/\Delta x \ . \tag{2.7c}$$

Since the details of soliton calculations, conserved quantities, direct and inverse scattering are worked out in Ablowitz [1978], Ablowitz and Segur [to appear] and Zakharov et al [1980], I will not pursue these matters further here.

III. THE INTERMEDIATE LONG-WAVE EQUATION

In the remaining portion of these lectures I wish to make some comments on a problem I mentioned earlier. This concerns very recent work we have been doing on the intermediate long wave (I.L.W.) equation, namely,

$$U_t + 2UU_x + U_x/\delta + \hat{T}(U_{xx}) = 0 \ , \tag{3.1}$$

where

$$\hat{T}U = \frac{1}{2\delta}\int_{-\infty}^{\infty} \coth\frac{\pi}{2\delta}(\xi - x)U(\xi)d\xi \tag{3.1a}$$

and $\int_{-\infty}^{\infty}$ represents the Cauchy principal value integral. Equation (3.1) is a nonlinear singular integro-differential equation. As $\delta \to 0$ we have the KdV equation,

$$U_t + 2UU_x + \frac{\delta}{3} U_{xxx} = 0 \tag{3.2}$$

and as $\delta \to \infty$ we have the Benjamin-Ono equation

$$U_t + 2UU_x + H(U_{xx}) = 0 , \tag{3.3}$$

where

$$H(U) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{U(\xi)}{\xi - x}\, d\xi$$

is the usual Hilbert transform. Thus it is intermediate between two important equations. Physically speaking, it has been derived in the context of long internal waves in a stratified fluid [see Joseph, 1977; Kubota et al, 1978].

Recently it has been shown that there are solitons, an infinite number of conserved quantities, a Bäcklund transform, an associated novel type of linear scattering problem, and solutions via linear Gel'fand-Levitan-Marchenko type integral equations (see Satsuma et al, 1979; Satsuma and Ablowitz, 1980; Kodama et al, to be published, Nakamura, 1979; Bock and Kruskal, 1979; and associated references). I shall only mention the basic results here.

To appreciate the novelty of the scattering problem, we should recall the following Plemelj formulae associated with the operator $\hat{T}$. Namely let $\psi^{\pm}(x)$ be the boundary values of certain functions $\psi\pm(z)$ analytic in horizontal strips of width 2δ ($\psi^{+}(z)$ analytic for $0 < \mathrm{Im} z < 2\delta$, $\psi^{-}(z)$ analytic for $-2\delta < \mathrm{Im} z < 0$) and periodically extended, namely,

$$\psi^{+}(x) = \lim_{\mathrm{Im} z \to 0} \frac{1}{2\delta} \int_{-\infty}^{\infty} \coth \frac{\pi}{2\delta} (\xi - z)\, u(\xi) d\xi = (i + \hat{T})\, u(x) , \tag{3.4a}$$

$$\psi^{-}(x) = \lim_{\mathrm{Im} z \to 0} \frac{1}{2\delta} \int_{-\infty}^{\infty} \coth \frac{\pi}{2\delta} (\xi - z)\, u(\xi) d\xi = (-i + \hat{T})\, u(x) . \tag{3.4b}$$

By periodicity $\psi^{-}(x) = \psi^{+}(x + 2i\delta)$. The scattering problem and associated time dependence is given by

$$i\psi_x^+ + (u-\lambda)\psi^+ = \mu\psi^- , \tag{3.5}$$

$$i\psi_t + 2i(\lambda + \frac{1}{2\delta})\psi_x^\pm + \psi_{xx}^\pm + [(\mp i - \hat{T})U_x + \nu]\psi^\pm = 0 \tag{3.6}$$

where λ,μ,ν are constants appropriately defined by the Jost functions.

Some of the main points to note are the following:

(a) Equation (3.5) is interpreted as a differential Riemann-Hilbert problem. For finite δ it also is a differential-difference (for complex x) equation.

(b) Compatibility of (3.5) and (3.6) yields the I.L.W. equation (3.1).

(c) As $\delta \to 0$ all results associated with KdV and the Schrödinger scattering problem are recovered.

(d) The following Gel'fand-Levitan equation produces the solution of the I.L.W. equation

$$K(x,y)+F(x,y) + \int_x^\infty K(x,s)F(s,y)ds = 0, \quad (y > x)$$

where,

$$u(x) = i(K^+(x,x)-K^-(x,x)) ,$$

$$\hat{T}u(x) = -(K^+(x,x)+K^-(x,x)) ,$$

$$K^-(x,y) = K^+(x+2i\delta,y+2i\delta) , \tag{3.7}$$

and $F(x,y)$ satisfies,

$$L_1F = (i\partial_x + \frac{1}{2\delta})F^+(x,y) + (i\partial_y - \frac{1}{2\delta})F^-(x,y) = 0 , \tag{3.8a}$$

$$L_2F = (i\partial_t + \partial_x^2 - \partial_y^2)F(x,y) = 0 . \tag{3.8b}$$

N-soliton solutions can be constructed by assuming exponential solutions for F, i.e.,

$$F(x,y) = \sum_{\ell=1}^{N} C_\ell(t)\exp(i\zeta_{-\ell}x+i\zeta_{+\ell}y) , \tag{3.9}$$

where

$$\zeta_{\pm\ell} = i\kappa_\ell \pm (\kappa_\ell \cot 2\kappa_\ell\delta - \frac{1}{2\delta}) ,$$

$$C_\ell(t) = C_\ell(0)\exp(-4\kappa_\ell(\kappa_\ell \cot 2\kappa_\ell\delta - \frac{1}{2\delta})t) .$$

A one-soliton solution is given by

$$u = \frac{2\kappa_1 \sin 2\kappa_1\delta}{\cosh 2\kappa_1(x-x_o(t)) + \cos 2\kappa_1\delta} , \tag{3.10}$$

where

$$x_o(t) = (2\kappa_1)^{-1}\ln(C_1(t)/2\kappa_1) .$$

It should be remarked that the analytical scattering and inverse scattering analysis yields an explicit representation for $F(x,y)$ [see Kodama et al, to be published];

$$F(x,y) = \frac{1}{2\pi}\int_{-1/2\delta}^{\infty} \frac{b(k,t)}{a(k,t)} e^{i\zeta_- x + i\zeta_+ y} + \sum_{\ell=1}^{N} C_\ell(t)\exp(i\zeta_{-\ell}x + i\zeta_{+\ell}y) \tag{3.11}$$

where

$$b(k,t) = b(k,0)\exp(-4ik(k\coth 2k\delta + \frac{1}{2\delta})t) ,$$

$$a(k,t) = a(k,0) ,$$

$$C_\ell(t) = -ib(k_\ell = i\kappa_\ell, t)/a'(k = i\kappa_\ell) .$$

These results are valid for given δ and $\max|u(x,o)|$ chosen small enough. When $\delta \to \infty$ (the Benjamin-Ono limit) new singularities may appear. We are presently investigating this situation.

REFERENCES

Ablowitz, M.J., 1978, Lectures on the Inverse Scattering Transform, Stud. Appl. Math., 58:17.

Ablowitz, M.J., Ramani, A., and Segur, H., 1978, Nonlinear evolution equations and ordinary differential equations of Painlevé type, Lett. Nuovo Cimento, 23:333.

Ablowitz, M.J., and Segur, H., to appear, Solitons and the Inverse Scattering Transform, SIAM Monograph, Stud. in Appl. Math.

Bock, T.L., and Kruskal, M.D., 1979, A two-parameter Miura transformation of the Benjamin-Ono equation, Phys. Lett., 74A:173.

Flaschka, H., and Newell, A.C., to appear, Monodromy and spectrum preserving deformations, I. Comm. Math Phys.

Joseph, R.I., 1977, Solitary waves in a finite depth fluid, J. Phys., 10A:L225.

Kaup, D.J., 1980, The inverse scattering solution for the full three-dimensional three-wave resonant interaction, Physica D, 1:45.

Kodama, Y., Satsuma, J., and Ablowitz, M.J., to be published, The nonlinear intermediate long wave equation: Analysis and method of solution.

Kubota, T., Ko, D.R.S., and Dobbs, L., 1978, Propagation of weakly nonlinear internal waves in a stratified fluid of finite depth, AIAA J. of Hydronautics, 12:157.

Miura, R.M., 1976, The Korteweg-deVries equation: A survey of results, SIAM Rev., 18:412.

Nakamura, A., 1979, Bäcklund transform and conservation laws of the Benjamin-Ono equation, J. Phys. Soc. Jap., 47:1335.

Proceedings of Joint U.S.-U.S.S.R. Conference on Soliton Theory, Kiev, U.S.S.R., 1979, to be published.

Sato, M., Miwa, T., and Jimbo, M., 1977, A series of papers entitled "Holonomic quantum fields, I-V", Pub. RIMS, I is in 14:223; Also a series of papers entitled "Studies on holonomic quantum fields, I-SVI", Proc. Japan Acad., I is in 53A:6.

Satsuma, J., Ablowitz, M.J., and Kodama, Y., 1979, On an internal wave equation describing a stratified fluid with finite depth, Phys. Lett., 73A:283.

Satsuma, J., and Ablowitz, M.J., 1980, Solutions of an internal wave equation describing a stratified fluid with finite depth, in: "Nonlinear Partial Differential Equations in Engineering and Applied Science," R.L. Sternberg, A.J. Kalinowki and J.S. Papadakis, eds., Marcel Dekker, New York.

Scott, A.C., Chu, F.Y.F., and McLaughlin, D.W., 1973, The soliton, a new concept in applied science, Proc. IEEE, 61:1443.

Taha, T., and Ablowitz, M.J., 1980, Numerical calculations, Clarkson College.

Zakharov, V.E., and Mikhailov, A.V., 1978, Relativistically invariant two-dimensional models of field theory which are integrable by menas of the inverse scattering problem method, Sov. Phys. J.E.T.P., 47:1017.

Zakharov, V.E., and Shabat, A.B., 1979, Integration of the nonlinear equations of mathematical physics by the method of the inverse scattering problem, II, Funkts. anal. i ego prilozh, 13:13.

Zakharov, V.E., and Manakov, S.V., 1979, Soliton theory, (Sov. Sci. Rev. Sec. A) Physics Reviews, 1:133.

Zakharov, V.E., Manakov, S.V., Novikov, S.P., and Pitayevsky, L.P., to appear 1980, Theory of solitons, the method of the inverse problem.

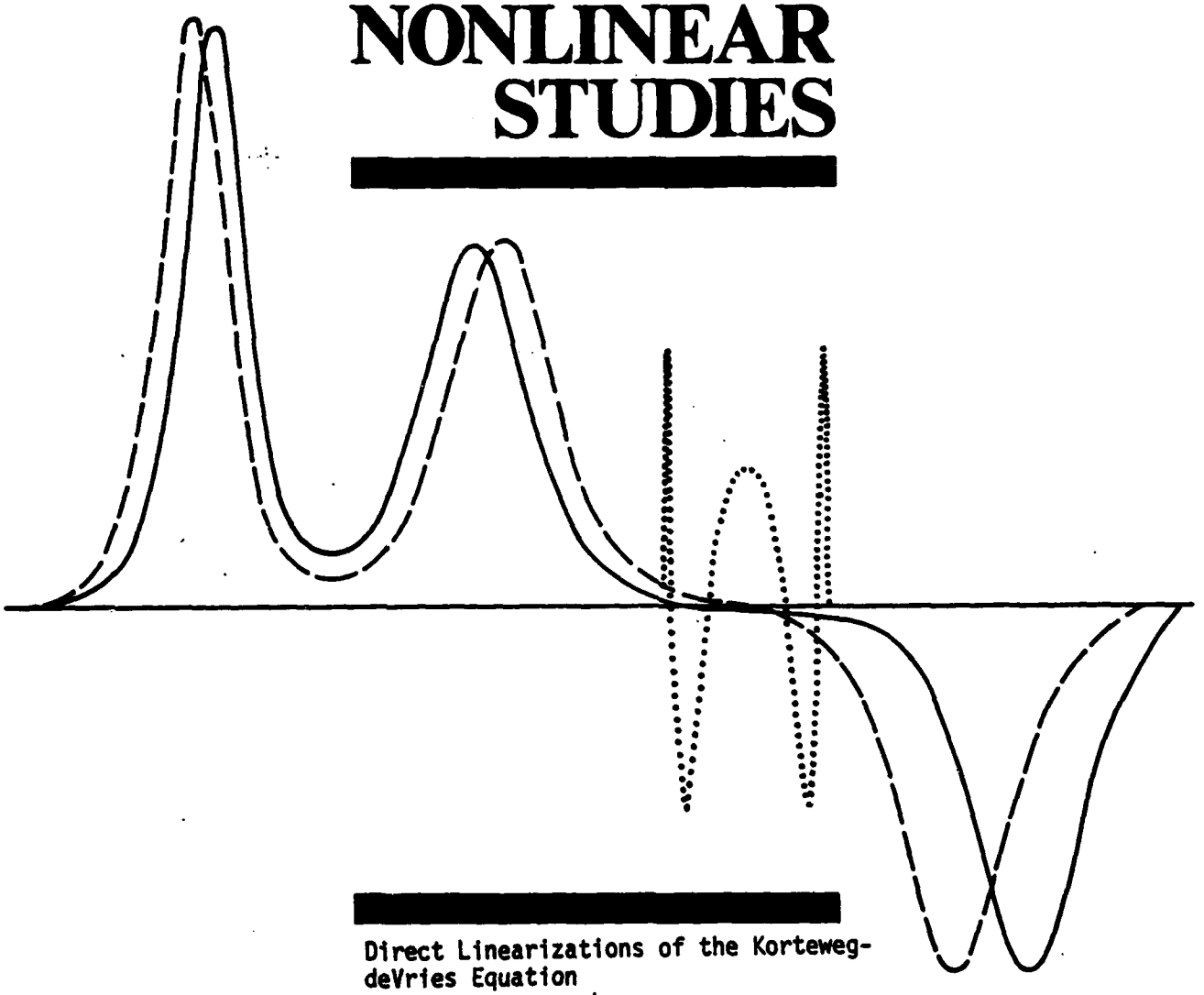
Zakharov, V.E., and Manakov, S., 1980, Lecture by S. Manakov "On the

self-dual Yang-Mills equation in 2n + 1 dimensions." Conference on Solitons, Crete, Greece.

PRINTED IN U.S.A.

I.F.N.S. - #2

INSTITUTE FOR NONLINEAR STUDIES



Direct Linearizations of the Korteweg-deVries Equation

by

A.S. Fokas and M.J. Ablowitz

Presented at the
Workshop on Mathematical Methods in Hydrodynamics and Intregability in Related Dynamical Systems

La Jolla, CA
December 7-9, 1981

Clarkson College of Technology
Potsdam, New York 13676

Direct Linearizations of the Korteweg-deVries Equation

by

A.S. Fokas and M.J. Ablowitz
Department of Mathematics and Computer Science
Clarkson College of Technology
Potsdam, New York 13676

1. INTRODUCTION

The use of the celebrated Gel'fand-Levitan-Marchenko (GLM) equation [1] for both obtaining potentials of the Schrödinger scattering problem as well as for solving the initial value problem of the Korteweg-deVries (KdV) equation [2] has been well established. However, in spite of its wide applicability the GLM equation has certain limitations. Namely it characterizes only those potentials-solutions which decay fast enough as $|x|\to\infty$ [3]. The following two well known examples readily illustrate the above point:

i) Consider the problem of finding solutions of the KdV $u_t+6uu_x+u_{xxx}=0$ which decay like $0(\frac{1}{x^2})$ as $|x|\to\infty$. These solutions are outside the range of applicability of the GLM equation. Ablowitz and Cornille [4] analyzed such solutions by "perturbing" the GLM equation around $u_o=-2/x^2$. u_o is a solution of KdV (the degenerate solutions of this "perturbed" GLM equation are the so-called quasi-solitons). Similar "perturbed" GLM equations can be found by perturbing around any "natural state" of the KdV [5].

ii) Consider the problem of finding the self-similar solutions of the KdV (i.e. the solutions of KdV invariant under a scaling transformation). These solutions satisfy a third order ODE; this ODE is of the Painlevé type and its first integral is related, through a one to one map, to Painlevé II [6]. Using the ideas of Ablowitz and Segur [7] one may obtain, via the GLM equation (properly scaled), a one parameter family of solutions of this third order ODE. How can one obtain a three parameter family of solutions? It is quite clear that one needs a more general equation than the GLM equation.

Recently [8] we have proposed such a generalization to the GLM equation. Furthermore, we have used this generalized equation to characterize a three parameter family of solutions of the ODE mentioned in ii) above, through a system

of Fredholm equations [8]. The sense in which our equation provides a generalization to the GLM equation can be best understood by recalling the "direct approach" method of Zakharov and Shabat [9]: These authors, bypassing the connection with inverse scattering and using certain linear operators, proved directly that solutions of the GLM equation (i.e. solutions of a linear integral equation of the Fredholm type) are also solutions of the KdV equation. Similarly, we have proved directly that solutions of a rather general linear integral equation (which in some cases is a singular integral equation) are also solutions of the KdV equation. The generality of our equation results from the fact that it involves an arbitrary measure-contour. Actually for a specific choice of the measure-contour it reduces to the so-called k-space equation recently introduced by Newton [10], and shown to be equivalent to the GLM equation.

It should be noted that in the linear limit our approach yields the general solution of the underlying linear equation $u_t+u_{xxx}=0$ (this is a consequence of the so-called Ehrenpreis principle). This should be contrasted with the linear limit of the GLM equation which yield only those solutions of $u_t+u_{xxx}=0$ which are obtained through the Fourier transform. In this sense, our equation can be thought of as the analogue of a generalized transform for solving a nonlinear PDE, in the same way that the GLM equation corresponds to the Fourier transform.

In this note we

a) present a direct proof of our new linearization and comment on its linear limit;

b) present the analogous generalization to the "perturbed" GLM equation discussed in i) above.

2. THE MAIN RESULTS

Theorem 1

Let $\phi(k,x,t)$ be a solution of

$$(1)\quad M\phi \doteq \phi(k;x,t)+ie^{i(kx+k^3t)}\int_L \frac{\phi(\ell;x,t)}{\ell+k}\,d\zeta(\ell)=e^{i(kx+k^3t)}$$

Then

$$(2)\quad u = -\frac{\partial}{\partial x}\int_L \phi(k;x,t)\,d\zeta(k)$$

solves the KdV equation

$$(3)\quad u_t+6uu_x+u_{xxx} = 0.$$

Before proving this result we make the following remarks.

REMARKS

1. The measure $d\zeta(\ell)$ and contour L above are quite arbitrary. The only assumption made is that equation (1) is well defined in the sense that one can interchange differentiations w.r.t. x and t and the integral along L.

2. If the nonlinearity is absent then equation (1) yields $\bar{\phi} = \exp(i(kx+k^3t))$. Hence equation (2) implies

$$(4)\quad \bar{u} = -\frac{\partial}{\partial x}\int_L e^{i(kx+k^3t)}d\zeta(k).$$

This is the general solution (Ehrenpreis principle) of the linear equation:

$$(5)\quad \bar{u}_t+\bar{u}_{xxx} = 0.$$

3. $\phi(k;x,t)$ is directly related to the Schrödinger eigenvalue problem

$$(6)\quad \phi_{xx}+u\phi-ik\phi_x = 0.$$

Let $\phi = \psi\exp(\frac{1}{2}(kx+k^3t))$ and equation (6) reduces to the usual Schrödinger equation

$$(7)\quad \psi_{xx}+((\tfrac{k}{2})^2+u)\psi = 0.$$

4. For the proof, we also assume that the homogeneous equation corresponding to (1) has only the trivial solution.

PROOF

Proving that u as defined by (2) solves the KdV is equivalent to proving that ϕ satisfies

$$(8)\quad \Omega(\phi) \doteq \phi_t+\phi_{xxx}-3(\int_L \phi_x(\ell)d\zeta(\ell))\phi_x = 0.$$

However, applying the operator $\partial_t+\partial_x^3$ to (1) one easily establishes that

$$(9)\quad M\Omega = 3k[i\phi_{xx}+k\phi_x-i(\int_L \phi_x(\ell)d\zeta(\ell))\phi] \doteq 3kw,$$

where the linear operator M is defined in (1). Similarly, applying the operator $i\partial_x^2+k\partial_x$ to (1) one obtains

$$(10)\quad Mw = 0.$$

Hence, because of the assumption 4 above, w = 0 and hence equation (9) implies $\Omega = 0$.

Particular choices of the measure-contour in (1) yield [8] i) a three parameter family of solutions of the ODE mentioned in the introduction; ii) the class of solutions obtainable through the usual inverse scattering transform method.

We now present an example of another linearization of KdV corresponding to the "perturbed" GLM equation discussed in the introduction.

Theorem 2

Let $\psi(k;x,t)$ be a solution of

$$(11)\quad \psi(k;x,t)+ie^{i(kx+k^3t)}\int_L \frac{(1+\frac{2i}{\ell x}+\frac{2i}{kx})}{k+\ell}\psi(\ell;x,t)d\zeta(\ell)$$

$$= e^{i(kx+k^3t)}(1+\frac{2i}{kx}).$$

Then

$$(12)\quad u = -\frac{2}{x^2} - \frac{\partial}{\partial x}\int_L \psi(k;x,t)(1+\frac{2i}{kx})d\zeta(k)$$

solves the KdV equation (3).

Acknowledgements:

This research was partially supported by the Air Force of Scientific Research, under Grant No. 78-3674-C and the Office of Naval Research under Grant No. N00014-76-C0867.

REFERENCES

[1] M.J. Ablowitz and H. Segur, "Solitons and the Inverse Scattering Transform", SIAM Monograph, December, 1981.

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

[3] P. Deift and E. Trubowitz, Comm. Pure Appl. Math., 32, 121 (1979).

[4] M.J. Ablowitz and H. Cornille, Phys. Lett. 72A, 277 (1979).

[5] M.J. Ablowitz and H. Airault, C.R. Acad. Sc. Paris, 292 (1981).

[6] A.S. Fokas and M.J. Ablowitz, "On a unified approach to transformations and elementary solutions of Painlevé equations", to appear in J. Math. Phys., December, 1981.

[7] M.J. Ablowitz and H. Segur, Phys. Rev. Lett. 38, 1103 (1977).

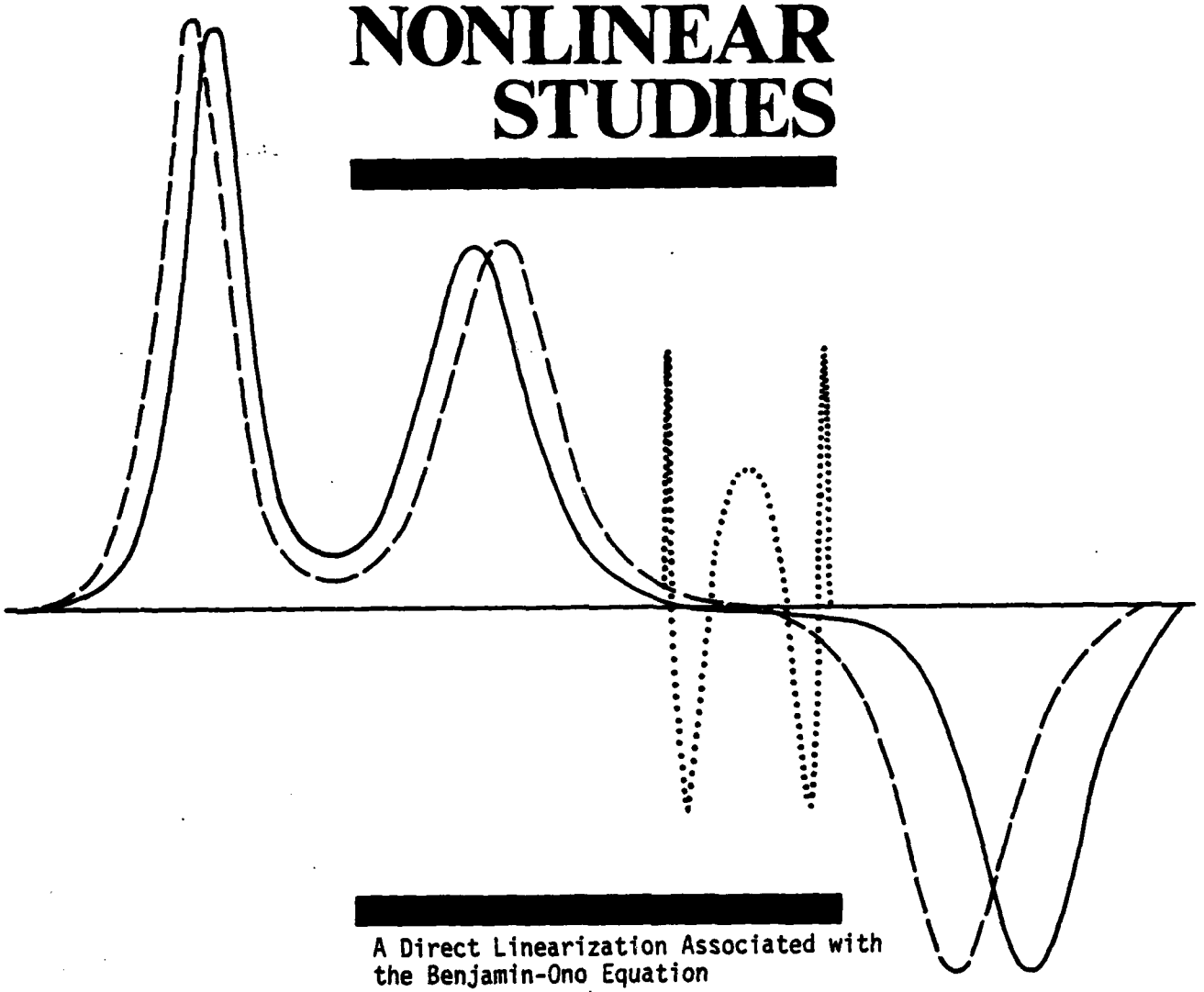
[8] A.S. Fokas and M.J. Ablowitz, Phys. Rev. Lett. 47, 1096 (1981).

[9] V.E. Zakharov and A.B. Shabat, Funkts. Anal. Pril. 8, 43 (1974), [Funct. Anal. Appl. 8, 226 (1974].

[10] R.G. Newton, J. Math. Phys. 21, 493 (1980); Geophys. J.R. Astr. Soc. 65, 191 (1981).

I.F.N.S. - #4

INSTITUTE FOR NONLINEAR STUDIES



A Direct Linearization Associated with the Benjamin-Ono Equation

by

M.J. Ablowitz and A.S. Fokas

Presented at the
Workshop on Mathematical Methods in Hydrodynamics and Intregability in Related Dynamical Systems

La Jolla, CA
December 7-9, 1981

Clarkson College of Technology
Potsdam, New York 13676

A Direct Linearization Associated with the Benjamin-Ono Equation

by

M.J. Ablowitz and A.S. Fokas
Department of Mathematics and Computer Science
Clarkson College of Technology
Potsdam, New York 13676

1. INTRODUCTION

The Benjamin-Ono (BO) equation [1] is a nonlinear singular-integro-differential equation which describes long internal gravity waves in stratified fluid. It has been established that the BO equation possesses: a) N soliton solutions [2], [3]; b) Bäcklund transformations, conservation laws, and a novel Lax pair [4], [5]; c) two nonlocal linear operators which generate its infinitely many commuting symmetries and constants of motion in involution[6].

In this note we outline a method of solution for the initial value problem of the BO equation, which we take in the form

$$(1)\quad u_t+2uu_x+Hu_{xx} = 0;\ Hv(x) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{v(\xi)}{\xi-x}d\xi,$$

where H denotes the Hilbert transform and principal value integrals are assumed if needed. We make use of the Lax pair

$$(2)\quad i\phi^+_x+\lambda(\phi^+-\phi^-) = -u\phi^+$$

$$(3)^{\pm}\quad i\phi^{\pm}_t - 2i\lambda\phi^{\pm}_x+\phi^{\pm}_{xx}-2i[u]^{\pm}_x\phi^{\pm} = -\nu\phi^{\pm},$$

where $\phi^+(x,t;\lambda)$ (ϕ^-) is the limit of a function analytic in the upper (lower) half z-plane as $z\to x$ (z is the complex extension of x); similarly $[u]^+$, $[u]^-$ are the (+) and (-) parts of $u(x,t)$ respectively, i.e. $u = [u]^+-[u]^-$, where $[u]^+$, $[u]^-$ are analytic in the upper and lower half z-plane respectively (λ is constant and is interpreted as a spectral parameter, ν is an arbitrary constant).

Our method differes substantially from the inverse scattering transform method as applied for example to the Korteweg-deVries equation. We think that the "inverse problem" associated with (2) is not solvable (in the usual sense) and, as a result of this, equations $(3)^{\pm}$ (i.e. the time-part of the Lax pair) play now a fundamental role.

The method we propose for linearizing (1) consists essentially of the following steps (we only state the results for the (+) functions, since the results for the (-) functions are similar):

a) Use equation (2) to express $[u]^+$ in terms of ϕ^+ and the "reflection coefficient".
b) Use equation (3) to find how the "reflection coefficient" evolves in time and then substitute the expression obtained in a) in equation (3), to obtain a nonlinear equation for ϕ^+.
c) Linearize the above equation for ϕ^+ to its linear part, either by an explicit transformation (in the case of solitons only) or by a linear integral equation (in the general case).

2. THE x-EIGENVALUE PROBLEM

Equation (2) should be interpreted as a differential Riemann-Hilbert problem for the analytic functions $\phi^{\pm}(z,t,\lambda)$. Equation (2) describes the jump condition across the real axis x; it yields unique solutions for $\phi^{\pm}$ provided one imposes some boundary conditions as $z\to\infty$, say, in the upper half-plane. Here we assume that either $\phi^+(z,t,\lambda)\to 0$, or 1, as $z\to\infty$, $IM(z)>0$.

i) Left and right eigenfunctions

Consideration of the equation (2) with the above boundary conditions yields that: a) there exist continuous eigenfunctions $\phi^{\pm}(x,t,\lambda)$, where λ is real and positive, and these eigenfunctions satisfy Fredholm equations of the second type; b) there exist discrete eigenfunctions $\phi^{\pm}(x,t,\lambda_j)$, where λ_j are real and negative, and these eigenfunctions satisfy homogeneous Fredholm equations. More specifically, let us consider only the (+) functions and let $M,\bar{M}$ denote "left" eigenfunctions, while $\bar{N},N$ denote "right" eigenfunctions. These eigenfunctions are specified by the following asymptotic behavior.

$$(4)\quad M\to 1 \qquad\qquad (5)\quad \bar{N}\to 1$$
$$\bar{M}\to e^{i\lambda x};\ x\to-\infty \qquad\qquad N\to e^{i\lambda x};\ x\to+\infty.$$

Furthermore, let ϕ_j denote the descrete (+) eigenfunctions. Then

$$(6)\quad \begin{pmatrix} M(x,t,\lambda) \\ \bar{M}(x,t,\lambda) \end{pmatrix} = \begin{pmatrix} 1 \\ e^{i\lambda x} \end{pmatrix} + \int_{-\infty}^{\infty} G^+(x,y,\lambda)u(y,t) \begin{pmatrix} M(y,t,\lambda) \\ \bar{M}(y,t,\lambda) \end{pmatrix} dy,$$

$$(7)\quad \begin{pmatrix} \bar{N}(x,t,\lambda) \\ N(x,t,\lambda) \end{pmatrix} = \begin{pmatrix} 1 \\ e^{i\lambda x} \end{pmatrix} + \int_{-\infty}^{\infty} G^-(x,y,\lambda)u(y,t) \begin{pmatrix} \bar{N}(y,t,\lambda) \\ N(y,t,\lambda) \end{pmatrix} dy,$$

$$(8)\quad \phi_j(x,t) = \int_{-\infty}^{\infty} G(x,y,\lambda_j)u(y,t)\phi_j(y,t)dy.$$

In the above expressions G^+, G^- are the (+) and (-) parts of the sectionally holomorphic function

$$(9)\quad G(x,y,\zeta) = \frac{1}{2\pi}\int_0^{\infty} \frac{e^{i(x-y)p}}{p-\zeta}dp,$$

where ζ denotes the complex extension of λ, i.e.

$$(10)\quad G^{\pm}(x,y,\lambda) = \lim_{\varepsilon\to 0} \frac{1}{2\pi}\int_0^{\infty} \frac{e^{i(x-y)p}\,dp}{p-(\lambda\pm i\varepsilon)}$$

Equations (6) and (7) can also be obtained from the corresponding equations associated with the intermediate long wave equation [7] in the appropriate limit.

The line of discontinuity of G is given by the positive λ axis, hence using Plemelj's formulae [8]

$$(11)\quad G^+ - G^- = \begin{cases} ie^{i(x-y)\lambda} & \lambda>0 \\ 0 & \lambda<0 \end{cases}$$

In particular for the discrete eigenfunctions λ_j, $G^+=G^-=G(x,y,\lambda_j)$.

ii) The relationship between left and right eigenfunctions

There exists the following relationship between left and right eigenfunctions

$$(12)\quad M = \bar{N} + \beta(\lambda,t)N;\ \lambda>0$$

$$M = \bar{N};\ \lambda<0$$

where $$\beta(\lambda,t) = i\int_{-\infty}^{\infty} u(y,t)M(y,t,\lambda)e^{-i\lambda y}dy$$

PROOF

Let $\Delta \doteq M-\bar{N}$. Then using (6), (7)

$$\Delta = 1 + \int_{-\infty}^{\infty} G^{+}uMdy - 1 - \int_{-\infty}^{\infty} G^{-}u\bar{N}dy =$$

$$= \int_{-\infty}^{\infty} G^{+}uMdy - \int_{-\infty}^{\infty} G^{-}u(M-\Delta)dy.$$

Hence

$$(13)\quad \Delta - \int_{\infty}^{\infty} G^{-}u\Delta dy = ie^{i\lambda x}\int_{\infty}^{\infty} e^{-i\lambda y}uMdy,$$

where we have used (11). Equation (13) implies that

$$\Delta = \beta(\lambda,t)N.$$

iii) Some analyticity properties

The results of Fredholm theory imply that M is a (+) function in the ζ-plane, except for possible poles. These poles correspond to homogeneous solutions of equation (6). Hence

$$(14)\quad M = 1+\sum_{1}^{P}\frac{c_j\phi_j}{\lambda-\lambda_j} + \mathcal{M};\quad \mathcal{M} \text{ is a (+) function in } \lambda.$$

Similarly

$$(15)\quad \bar{N} = 1+\sum_{1}^{P}\frac{c_j\phi_j}{\lambda-\lambda_j} + \bar{\mathcal{N}};\quad \bar{\mathcal{N}} \text{ is a (-) function in } \lambda.$$

The above representations, together with knowledge of β, are inadequate for solving the "inverse problem" associated with (12). However, using equations (12), (14), (15) one obtains the following important relationship.

$$(16)\quad [u]^{+} = \frac{1}{2\pi i}\int_{0}^{\infty}\beta(\lambda,t)N(x,t,\lambda)d\lambda - \sum_{1}^{P}c_j(t)\phi_j(x,t)$$

iv) The time evolution

Using equation (12) in $(3)^{+}$ one easily establishes that

$$(17)\quad c_j(t) = \text{constants},\quad \beta(\lambda,t) = \beta_o(\lambda)e^{i\lambda^2 t},$$

$$\lambda_j(t) = \text{constants}.$$

3. THE t-PART OF THE LAX PAIR

i) Solitons

Let $\beta_o = 0$. Then substituting (16) into (3) one obtains the following system of coupled nonlinear PDE's:

$$\Phi_{jt} - 2\lambda_j \Phi_{jx} - i\Phi_{jxx} + 2\left(\sum_1^P c_\ell \Phi_\ell\right)_x \Phi_j = 0. \tag{18}$$

If one assumes that

$$\Phi_j(x,t) = \sum_1^P \frac{c_{kj}(t)}{x - x_k(t)},$$

one is led to the Calogero-Moser system [9]

$$\ddot{x}_k = 8 \sum_{\ell=1}^P \frac{1}{(x_k - x_\ell)^3}, \quad k = 1, \ldots, P.$$

Hence, using (16) (and its complex conjugate) one recovers the well known P-soliton solution of the BO equation. Thus, the P-solitons correspond to the discrete spectrum of the eigenvalue equation (2).

Recently [10] we have been able to linearize equations (18): the transformations

$$\Phi_j(x,t) = V_j(x,t) - i\sum_1^P c_\ell \Phi_\ell(x,t) W_{j\ell}(x,t), \tag{19}$$

where

$$W_{j\ell} = e^{i(\lambda_j - \lambda_\ell)x} \int_a^x V_j(\xi,t) e^{-i(\lambda_j - \lambda_\ell)\xi} d\xi \tag{20}$$

$$+ B(t) e^{i(\lambda_j - \lambda_\ell)(x-a)},$$

$$B_t + i(\lambda_\ell^2 - \lambda_j^2) B = (\lambda_\ell + \lambda_j) V(a,t,\lambda) \tag{21}$$

$$+ iV_x(a,t,\lambda), \text{ a arbitrary},$$

relate equations (18) to the linear equations

$$V_{jt} - 2\lambda_j V_{jx} - iV_{jxx} = 0. \tag{22}$$

ii) The general case

Substituting (16) into (3) one obtains

$$(23)\quad \hat{N}_t - 2\lambda\hat{N}_x - i\hat{N}_{xx} - 2\Big[\frac{1}{2\pi i}\int_0^\infty \beta_0(\tau)\hat{N}d\tau - \sum_1^P c_\ell \Phi_\ell\Big]_x \hat{N} = 0$$

$$(24)\quad \Phi_{jt} - 2\lambda\Phi_{jx} - i\Phi_{jxx} - 2\Big[\frac{1}{2\pi i}\int_0^\infty \beta_0(\tau)\hat{N}d\tau - \sum_1^P c_\ell \Phi_\ell\Big]_x \Phi_j = 0,$$

where

$$\hat{N} = Ne^{i\lambda^2 t}$$

The above equations can also be linearized to their linear parts with the aid of an integral equation [10]. This linearization can be used for solving the initial value problem of the BO equation. Here, we only prove this linearization (we actually prove a slightly stronger result than the one needed here). A complete discussion of the initial value problem is given in [10].

Theorem

Let $\psi(x,t,\lambda)$ be a solution of

$$(25)\quad \psi(x,t,\lambda) = V(x,t,\lambda) - i\int_L \psi(x,t,\ell)W(\lambda,\ell,x,t)d\rho(\ell),$$

where $V(x,t,\lambda)$ and $W(\lambda,\ell,x,t)$ are defined by

$$(26)\quad \mathcal{L}(\lambda)V \doteq V_t - 2\lambda V_x - iV_{xx} = 0,$$

$$(27)\quad W(\lambda,\ell,x,t) = e^{i(\lambda-\ell)x}\int_a^x V(\lambda,\xi,t)e^{-i(\lambda-\ell)\xi}d\xi + B(t,\lambda,\ell)e^{i(\lambda-\ell)(x-a)},$$

$$(28)\quad B_t + i(\ell^2-\lambda^2)B = (\lambda+\ell)V(a,t,\lambda) + iV_x(a,t,\lambda).$$

Assume that the homogeneous equation corresponding to (25) has only the trivial solution and that one may interchange differentiations w.r.t. x,t and the integral along L. Then ψ also solves the nonlinear equation

$$(29)\quad \Omega(\psi)=\psi_t-2\lambda\psi_x-i\psi_{xx}+2\left(\int_L \psi_x(\ell)\,d\rho(\ell)\right)\psi = 0.$$

PROOF

Let ψ satisfy (25). Then

$$\Omega(\psi)(\lambda)+i\int_L \Omega(\psi)(\ell)W(\lambda,\ell)\,d\rho(\ell) = \mathcal{L}(\lambda)V$$

$$-i\int_L \psi(\ell)\mathcal{L}(\lambda)W(\lambda,\ell)\,d\rho(\ell)$$

$$-2\int_L \psi_x(\ell)[W_x(\lambda,\ell)+i(\ell-\lambda)W(\lambda,\ell)-V(\lambda)]\,d\rho(\ell) = 0.$$

Hence, $\Omega(\psi) = 0$ provided that

$$(30)\quad \mathcal{L}(\lambda)V = 0,\ \mathcal{L}(\lambda)W(\lambda,\ell) = 0,\ W_x+i(\ell-\lambda)W-V = 0.$$

Equation (30.a) is equation (26) and equation (30.c) is equivalent to equation (27). To satisfy equation (30.b) use equation (27) and integration by parts, this readily yields equation (28).

It is clear that equations (23), (24) are a special case of (29) where $d\rho(\ell) = [\sum_1^p c_j\delta(\ell-\lambda_j) - \frac{1}{2\pi i}\beta_o(\ell)]\,d\ell$.

REFERENCES

[1] i) T.B. Benjamin, J. Fluid Mech. 29, 559 (1967).
ii) H. Ono, J. Phys. Soc. Japan, 39, 1082 (1975).

[2] K.M. Case, Proc. Nat. Acad. Sc. U.S.A., 75, 3563 (1978); 76, 1 (1979).

[3] H.H. Chen, Y.C. Lee, N.R. Pereira, Phys. Fluids 22, 187 (1979).

[4] A. Nakamura, J. Phys. Soc. Jap. 47, 1701 (1979).

[5] T.L. Boch, M.D. Kruskal, Phys. Lett. 74A, 173 (1979).

[6] A.S. Fokas, B. Fuchssteiner, "The hierarchy of the Benjamin-Ono equation", to appear in Phys. Lett.

[7] Y. Kodama, M.J. Ablowitz, J. Satsuma, Phys. Rev. Lett. 46, 687 (1981).

[8] N.I. Muskhelishvili, "Singular Integral Equations", ed. by J. Radok (Noordhoff, Grouingen, 1953).

[9] i) F. Calogero, J. Math. Phys. 12, 419 (1971).
ii) J. Moser, Adv. in Math. 16, 197 (1975).

[10] M.J. Ablowitz, A.S. Fokas, R.L. Anderson, "A method of Solution of the Benjamin-Ono equation" (preprint).

Note on Asymptotic Solutions of the Korteweg–de Vries Equation with Solitons

*By Mark J. Ablowitz and Yuji Kodama**

The long time asymptotic solution of the Korteweg–de Vries equation containing both solitons and the dispersive wavetrain is described. It is shown that a soliton interacts elastically both with the dispersive wavetrain and with other solitons. An explicit formula for the phase shift produced upon interaction is given. These ideas also apply to other nonlinear evolution equations solvable by the inverse scattering transform.

I. Introduction

It is by now well known that the Korteweg–de Vries (KdV) equation has a solution which consists of solitons and a dispersive wavetrain. The solitons and dispersive wavetrain are associated with the discrete and continuous spectra, respectively, in the related Schrödinger scattering problem [1,2]. Aspects of the asymptotic behavior of the solution have been discussed by numerous authors. For example, pure N-soliton solutions and their phase shifts have been discussed in [1–5], and the long time asymptotic solution of the dispersive wavetrain without solitons has also been considered in the literature (see for example [6, 7]). A review of the latter work as well as the connection formulas for the second Painlevé equation was presented in [8]. The work in [8] clears up, once and for all, the various errors, misstatements, and ambiguities regarding the $0(1)$ asymptotic solution of the dispersive wavetrain. However, the problem of long time asymptotic states evolving from initial data containing both solitons and dispersive wavetrain remains unresolved. A difficulty that must be overcome is that the solitons and dispersive wavetrain are of differing, exponentially small asymptotic

Address for correspondence: Professor Mark J. Ablowitz, Department of Mathematics, Clarkson College of Technology, Potsdam, NY 13676.
*Present Address: Bell Laboratories, Murray Hill, NJ 07974.

STUDIES IN APPLIED MATHEMATICS 66:159–170 (1982)

Copyright © 1982 by the Massachusetts Institute of Technology
Published by Elsevier Science Publishing Co., Inc.
0022-2526/82/020159–12$02.75

orders in certain regions of space. Their relative magnitudes change drastically depending on spatial location. Our aim is to examine this question and show how to obtain the long time asymptotic solution.

A related question is the following. What are the phase shifts of the solitons as they interact both with other solitons and with the dispersive wavetrain? This question has been considered in the literature, the only useful result so far being given in [6] for the case of one soliton plus dispersive waves (see also [5] and [9]). Here we show that a soliton interacts elastically with the dispersive waves as well as with the other solitons. We give a concrete formula for the phase shift. A corollary to this result is the definition of a "perfect soliton" of an evolution equation, i.e., one which in the asymptotic limit interacts elastically with any sufficiently localized disturbance. Solitary waves of an evolution equation without this property presumably would be associated with evolution equations not solvable by the inverse scattering transform (IST) (if this is indeed true, then equations which have two solitary waves interacting nonelastically will not be solvable by the IST). Finally, we remark that the ideas and methods described herein apply, in principle, to other nonlinear evolution equations solvable by the IST.

II. Finite perturbation method

We consider the KdV equation,

$$u_t + 6uu_x + u_{xxx} = 0, \tag{2.1}$$

with arbitrary initial data, decaying sufficiently rapidly as $|x| \to \infty$, $u(x,0) = \phi(x)$. The method of finite perturbations (see for example References [10–12]) allows us to seek a solution to (2.1) of the form

$$u(x,t) = u_0(x,t) + v(x,t), \tag{2.2}$$

where $u_0(x,t)$ is any special solution of (2.1) and $v(x,t)$ given by

$$v_t + 6(u_0 v)_x + 6vv_x + v_{xxx} = 0. \tag{2.3}$$

The solution of v is obtained via the following Gel'fand-Levitan-Marchenko integral equation (see Shabat [10], and further applications of this idea [11, 12]):

$$K(x,y;t) + F(x,y;t) + \int_x^\infty K(x,s;t)F(s,y;t)\,ds = 0, \qquad y > x, \tag{2.4a}$$

$$v(x,t) = 2\frac{d}{dx}K(x,x;t). \tag{2.4b}$$

The functions $K(x,y;t)$ and $F(x,y;t)$ satisfy the following partial differential

equations:

$$\left[\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial y^2}+u_0(x,t)-u_0(y,t)+2\frac{d}{dx}K(x,x;t)\right]K(x,y;t)=0, \tag{2.5a}$$

$$\left[\frac{\partial}{\partial t}+4\left(\frac{\partial^3}{\partial x^3}+\frac{\partial^3}{\partial y^3}\right)+6\left\{u_0(x,t)\frac{\partial}{\partial x}+u_0(y,t)\frac{\partial}{\partial y}\right\}\right.$$
$$\left.+3\{u_{0x}(x,t)+u_{0y}(y,t)\}\right]F(x,y;t)=0. \tag{2.5b}$$

Generally speaking the above method is very difficult to carry out with regard to initial values. Whereas the transformation [11]

$$K(x,x;0)\overset{(2.5a)}{\rightarrow}K(x,y;0)\overset{(2.4a)}{\rightarrow}F(x,y;0)$$
$$\overset{(2.5b)}{\rightarrow}F(x,y;t)\overset{(2.4a)}{\rightarrow}K(x,y;t)\overset{(2.4b)}{\rightarrow}v(x,t)$$

gives a method of solution that is conceptually clear, the concrete application of this to an initial value problem has not yet been successfully carried out. However, in the case where u, u_0, v all decay rapidly as $|x|\to\infty$, an explicit formula for $F(x,y;t)$ in terms of the scattering data at the initial time can be obtained, thereby giving us a workable method of solution. In the appendix, we derive the formula (see also [13])

$$F(x,y;t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\{r^+(k,t)-r_0^+(k,t)\}g_0(x,t;k)g_0(y,t;k)\,dk$$
$$+\sum_{l=1}^{N}C_l^+(t)g_0(x,t;i\kappa_l)g_0(y,t;i\kappa_l)$$
$$-\sum_{l=1}^{N_0}C_{0l}^+(t)g_0(x,t;i\kappa_{0l})g_0(y,t;i\kappa_{0l}), \tag{2.6}$$

where $S^+(t)\equiv[r^+(k,t),\{C_l^+(t),\kappa_l\}_{l=1}^{N}]$ and $S_0^+(t)\equiv[r_0^+(k,t),\{C_{0l}^+(t),\kappa_{0l}\}_{l=1}^{N_0}]$ are the "right" scattering data associated with the potentials $u(x,t)$ and $u_0(x,t)$ respectively; $g_0(x,t;k)$ is the Jost function of the associated scattering problem with $u_0(x,t)$ defined in the appendix. We will show that the above formulas are particularly useful for the asymptotic case $t\to\infty$. The following formulas are

also important (see appendix for derivations):

$$M(x,y;t)+G(x,y;t)+\int_{-\infty}^{x}M(x,s;t)G(s,y;t)\,ds=0,\qquad y<x, \tag{2.7a}$$

$$v(x,t)=-2\frac{d}{dx}M(x,x;t), \tag{2.7b}$$

where $G(x,y;t)$ is given by

$$\begin{aligned}G(x,y;t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\{r^{-}(k,t)-r_0^{-}(k,t)\}f_0(x,t;k)f_0(y,t;k)\,dk\\ +\sum_{l=1}^{N}C_l^{-}(t)f_0(x,t;i\kappa_l)f_0(y,t;i\kappa_l)\\ -\sum_{l=1}^{N_0}(C_{0l}^{-}(t)f_0(x,t;i\kappa_{0l})f_0(y,t;i\kappa_{0l}).\end{aligned} \tag{2.8}$$

Here, $S^{-}(t)\equiv[r^{-}(k,t),\{C_l^{-}(t),\kappa_l\}_{l=1}^{N}]$ and $S_0^{-}(t)\equiv[r_0^{-}(k,t),\{C_{0l}^{-}(t),\kappa_{0l}\}_{l=1}^{N_0}]$ are the "left" scattering data associated with $u(x,t)$ and $u_0(x,t)$, respectively; $f_0(x,t;k)$ is the Jost function associated with $u_0(x,t)$ defined in the appendix.

We now use the above method to describe the asymptotic solution of the KdV equation. For given initial data,

$$S^{\pm}(0)=\left[r^{\pm}(k,0),\{C_l^{\pm}(0),\kappa_l\}_{l=1}^{N}\right], \tag{2.9}$$

we choose two reference potentials $u_0^{\pm}$ whose scattering data have, respectively,

$$S_0^{+}(0)=\left[0,\{C_l^{+}(0),\kappa_l\}_{l=1}^{N}\right], \tag{2.10a}$$

$$S_0^{-}(0)=\left[0,\{C_l^{-}(0),\kappa_l\}_{l=1}^{N}\right], \tag{2.10b}$$

i.e., pure N-soliton solutions:

$$u_0^{+}(x,t)=4\sum_{l=1}^{N}\kappa_l C_l^{+}(t)[g_0^{+}(x,t;i\kappa_l)]^2, \tag{2.11a}$$

$$u_0^{-}(x,t)=4\sum_{l=1}^{N}\kappa_l C_l^{-}(t)[f_0^{-}(x,t;i\kappa_l)]^2. \tag{2.11b}$$

Note that the solutions (2.11) are in terms of the squared eigenfunctions (see [2]).

It should be noted that $u_0^+(s,t)$ and $u_0^-(s,t)$ are *different* unless $u(x,t)$ has discrete spectrum only. With (2.9) the kernels of the integral equations (2.4) and (2.7) simplify to

$$F(x,y;t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} r^+(k,t)g_0^+(x,t;k)g_0^+(y,t;k)\,dk, \tag{2.12a}$$

$$G(x,y;t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} r^-(k,t)f_0^-(x,t;k)f_0^-(y,t;k)\,dk. \tag{2.12b}$$

Moreover, the Jost functions g_0 and f_0 for the case of pure N-soliton solutions, can be calculated in terms of determinants (see [2] and the appendix for the general form of g_0 and f_0). From this we can state that

$$F(x,y;t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} r^+(k,0)a_0^2(k)e^{i k(x+y)+8ik^3t}\,dk$$

$$+T^+(x,y;t)\int_{-\infty}^{\infty} R^+(k)e^{i k(x+y)+8ik^3t}\,dk, \tag{2.13a}$$

$$G(x,y;t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} r^-(k,0)a_0^2(k)e^{-i k(x+y)-8ik^3t}\,dk$$

$$+T^-(x,y;t)\int_{-\infty}^{\infty} R^-(k)e^{-i k(x+y)-8ik^3t}\,dk, \tag{2.13b}$$

where $a_0(k)\equiv\prod_{l=1}^{N}(k-i\kappa_l)/(k+i\kappa_l)$, $R^{\pm}(k)$ are "nice" functions of k, and the $T^{\pm}(x,y;t)$ are bounded functions which decay rapidly behind the solitons.

Let us order the solitons via

$$\kappa_1>\kappa_2>\cdots>\kappa_N>0; \tag{2.14}$$

i.e., the soliton $u_1=2\kappa_1^2\operatorname{sech}^2\kappa_1(x-4\kappa_1^2t-x_{10})$ with parameter κ_1 moves to the right with the largest velocity. Hence, for $|t|\to\infty$, in the soliton region, $|x|\gtrsim\kappa_N|t|$, we see that F and G decay exponentially, via the integral equations (2.4) and (2.7); $v(x,t)$ decays exponentially in this region; and

$$u(x,t)\sim u_0^+(x,t) \qquad \text{for} \quad t\to\infty, \quad x\gtrsim\kappa_N t, \tag{2.15a}$$

$$u(x,t)\sim u_0^-(x,t) \qquad \text{for} \quad t\to-\infty, \quad x\lesssim\kappa_N t. \tag{2.15b}$$

But $u_0^{\pm}(x,t)$ are the well-known soliton solutions. All results associated with $u_0^{\pm}(x,t)$ are well known. We shall return to this shortly to derive the formula of the phase shifts of solitons. First, however, we remark upon the simplicity of describing the remaining portion of the solution.

Behind the solitons $u_0^{\pm}(x,t)$, $T^{\pm}(x,y;t)$ decays exponentially fast. The only terms which contribute significantly are where we approach the regions $x\lesssim t^{1/3}$

$(t \to +\infty)$, $x \gtrsim -|t|^{1/3}$ $(t \to -\infty)$. As the *boundaries* of these regions are approached, from (2.3) v asymptotically satisfies

$$v_t + 6vv_x + v_{xxx} \simeq 0 \tag{2.16}$$

(exponentially small corrections), and

$$F \sim \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, r^+(k,0)a_0^2(k)e^{i[k(x+y)+8ik^3t} \quad \text{as} \quad t \to \infty, \quad x \to t^{1/3}, \tag{2.17a}$$

$$G \sim \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, r^-(k,0)a_0^2(k)e^{-i[k(x+y)-8ik^3t} \quad \text{as} \quad t \to -\infty, \quad x \to -|t|^{1/3}. \tag{2.17b}$$

From (2.4) and (2.7), as we approach these regions,

$$u \sim v \sim -2\frac{d}{dx}F(x,x;t) \quad \text{as} \quad t \to +\infty, \quad x \to t^{1/3}, \tag{2.18a}$$

$$u \sim v \sim -2\frac{d}{dx}G(x,x;t) \quad \text{as} \quad t \to -\infty, \quad x \to -|t|^{1/3}. \tag{2.18b}$$

This can now be taken as the beginning point of the dispersive wave analysis of Ablowitz and Segur [6,8]. Hence we could call upon all of these results with the *modified* reflection coefficients

$$\rho^+(k,0) = r^+(k,0)a_0^2(k), \quad t \to +\infty, \tag{2.19a}$$

$$\rho^-(k,0) = r^-(k,0)a_0^2(k), \quad t \to -\infty, \tag{2.19b}$$

[note that $\rho^{\pm}(k,0)$ have no poles in the upper half plane].

Let us turn to the question of phase shifts between solitons and the dispersive wavetrain. To do this, we need only call upon well-known results. As $t \to +\infty$, from [2, p. 121] we have

$$u \underset{\substack{t\to+\infty \\ x-4\kappa_l^2 t}}{\to} 2\kappa_l^2 \operatorname{sech}^2 \kappa_l\left(x - 4\kappa_l^2 t - x_{0l}^+\right), \tag{2.20a}$$

$$x_{0l}^+ = \frac{1}{2\kappa_l}\log\left[\frac{C_l^+(0)}{2\kappa_l}\prod_{m=1}^{l-1}\left(\frac{\kappa_m - \kappa_l}{\kappa_m + \kappa_l}\right)^2\right]. \tag{2.20b}$$

As $t \to -\infty$, from [3, p. 1408] we have

$$u \underset{\substack{t\to-\infty \\ x-4\kappa_l^2 t}}{\to} 2\kappa_l^2 \operatorname{sech}^2 \kappa_l\left(x-4\kappa_l^2-x_{0l}^-\right), \tag{2.20c}$$

$$x_{0l}^- = -\frac{1}{2\kappa_l}\log\left[\frac{C_l^-(0)}{2\kappa_l}\prod_{m=1}^{l-1}\left(\frac{\kappa_m-\kappa_l}{\kappa_m+\kappa_l}\right)^2\right]. \tag{2.20d}$$

(In (2.20), we note some differences in notation between [1], [2] and ourselves: (a) *our norming* coefficients are the squares of theirs; (b) in [2] they order $\kappa_N > \kappa_{N-1} > \cdots > \kappa_1 > 0$, which is opposite to (2.14). So (2.20) reflects the appropriate modifications.) Thus, we immediately have the phase shift formula:

$$x_{0l}^+ - x_{0l}^- = \frac{1}{2\kappa_l}\log\left[\frac{C_l^+C_l^-}{(2\kappa_l)^2}\prod_{m=1}^{l-1}\left(\frac{\kappa_m-\kappa_l}{\kappa_m+\kappa_l}\right)^4\right]. \tag{2.21}$$

Special cases of this result are:

(1) When $N=1$,

$$x_{01}^+ - x_{01}^- = \frac{1}{2\kappa_1}\log\frac{C_1^+C_1^-}{(2\kappa_1)^2}. \tag{2.22}$$

This agrees with the result given in [5] (note difference in the definition of the norming constant).

(2) Pure N-soliton solutions: Here

$$C_l^+C_l^- = -\frac{1}{\left[a_0'(i\kappa_l)\right]^2} = (2\kappa_l)^2\prod_{m\neq l}^{N}\left(\frac{\kappa_l+\kappa_m}{\kappa_l-\kappa_m}\right)^2, \tag{2.23}$$

so that (2.21) yields

$$x_{0l}^+ - x_{0l}^- = \frac{1}{\kappa_l}\left[\sum_{m=1}^{l-1}\log\frac{\kappa_m-\kappa_l}{\kappa_m+\kappa_l} - \sum_{m=l+1}^{N}\log\frac{\kappa_l-\kappa_m}{\kappa_l+\kappa_m}\right]. \tag{2.24}$$

III. Asymptotic solution of the KdV equation with one soliton

As a consequence of the previous section, and as a concrete example, we discuss the asymptotic solution which contains one soliton and a dispersive wavetrain.

Let $u(x,0)$ be an initial function whose scattering data are given by

$$S^{\pm}(0) = \left[r^{\pm}(k,0), C_1^{\pm}(0), \kappa_1\right]. \tag{3.1}$$

Choose $u_0^{\pm}(x,0)$ such that the scattering datas are given, respectively, by

$$S_0^+(0) = [0, C_1^+(0), \kappa_1], \tag{3.2a}$$

$$S_0^-(0) = [0, C_1^-(0), \kappa_1]; \tag{3.2b}$$

i.e., $u_0^{\pm}(x,0)$ are pure one-soliton solutions,

$$u_0^+(x,t) = 2\kappa_1^2 \operatorname{sech}^2 \kappa_1(x - 4\kappa_1^2 t - x_{01}^+), \tag{3.3a}$$

$$u_0^-(x,t) = 2\kappa_1^2 \operatorname{sech}^2 \kappa_1(x - 4\kappa_1^2 t - x_{01}^-), \tag{3.3b}$$

where $x_0^{\pm} = (2\kappa_1)^{-1} \log\{C_1^{\pm}(0)/(2\kappa_1)\}$. [Note that if $u(x,0)$ is a *pure* one soliton solution, then $C_1^+ C_1^- = (2\kappa_1)^2$, i.e. $x_{01}^+ = x_{01}^-$.] The relevant Jost functions for $u_0^{\pm}$, $g_0^+(x,t;k)$, and $f_0^-(x,t;k)$ are given by

$$g_0^+(x,t;k) = \frac{e^{ikx}}{k + i\kappa_1}\left[k + i\kappa_1 \tanh \kappa_1(x - 4\kappa_1^2 t - x_{01}^+)\right], \tag{3.4a}$$

$$f_0^-(x,t;k) = \frac{e^{-ikx}}{k + i\kappa_1}\left[k - i\kappa_1 \tanh \kappa_1(x - 4\kappa_1^2 t - x_{01}^-)\right]. \tag{3.4b}$$

Note that $g_0^+(x,t;k)$ may be rewritten in the form

$$g_0^+(x,t;k) = \left[a_0(k) + \frac{i\kappa_1}{k + i\kappa_1}\{1 + \tanh \kappa_1 \theta(x)\}\right] e^{ikx},$$

$$\theta(x) = x - 4K_1^2 t - x_{01}^+;$$

hence $F(x, y; t)$ is given by

$$\begin{aligned} F(x,y;t) = {} & \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, r^+(k,0) a_0^2(k) e^{ik(x+y)+8ik^3t} \\ & + \frac{1}{2\pi} T_1(x,y;t) \int_{-\infty}^{\infty} dk\, r^+(k,0) \frac{a_0(k) i\kappa_1}{k + i\kappa_1} e^{ik(x+y)+8ik^3t} \\ & + \frac{1}{2\pi} T_2(x,y;t) \int_{-\infty}^{\infty} dk\, r^+(k,0) \left(\frac{i\kappa_1}{k + i\kappa_1}\right)^2 e^{ik(x+y)+8ik^3t}, \end{aligned} \tag{3.5}$$

where

$$T_1(x,y;t) = 2 + \tanh \kappa_1 \theta(x) + \tanh \kappa_1 \theta(y),$$

$$T_2(x,y;t) = [1 + \tanh \kappa_1 \theta(x)][1 + \tanh \kappa_1 \theta(y)]$$

(note that T_1, T_2 are bounded and $T_1, T_2 \to 0$ as $\theta(x), \theta(y) \to -\infty$). From (3.5), for $t \to \infty$, behind the soliton $F(x, y; t)$ can be approximated by

$$F(x, y; t) \simeq \frac{1}{2\pi} \int_{-\infty}^{\infty} r^+(k,0) a_0^2(k) e^{ik(x+y)+8ik^3t} \qquad \text{for} \quad x, y \ll 4\kappa_1^2 t, \tag{3.6}$$

and for all other regions, $F(x, y; t)$ decays exponentially. For $t \to \infty$, the solution of (2.3) in the region behind the soliton can be given by

$$v_t + 6vv_x + v_{xxx} = 0 \tag{3.7}$$

(i.e., the term $u_0 v$ can be ignored). Since $F(x, y; t)$ is given by (3.6) for $t \to \infty$, one can find the asymptotic solution by using the method developed in [5] with modified reflection coefficient $\rho^+(k,0) = r^+(k,0) a_0^2(k)$.

Appendix A. Scattering problem

The scattering problem for the KdV equation (2.1) is given by

$$\psi_{xx} + (u + k^2)\psi = 0, \tag{A.1a}$$

$$\psi_t + 4\psi_{xxx} + 6u\psi_x + 3u_x\psi = 4ik^3\psi. \tag{A.1b}$$

The Jost functions are the solutions of (A.1a) defined by

$$\begin{aligned} f(x, t; k) &\sim e^{-ikx} \qquad \text{as} \quad x \to -\infty, \\ g(x, t; k) &\sim e^{ikx} \qquad \text{as} \quad x \to \infty. \end{aligned} \tag{A.2}$$

For real k, there are relations between f and g,

$$\begin{aligned} f(x, t; k) &= a(k, t) g(x, t; -k) + b(k, t) g(x, t; k), \\ g(x, t; k) &= a(k, t) f(x, t; -k) - b(-k, t) f(x, t; k), \end{aligned} \tag{A.3}$$

where $|a(k, t)|^2 - |b(k, t)|^2 = 1$.

The bound states of (A.1a) are given by

$$\begin{aligned} a(i\kappa_l, t) &= 0, \\ f(x, t; i\kappa_l) &= b_l g(x, t; i\kappa_l), \qquad l = 1, \ldots, N, \end{aligned} \tag{A.4}$$

where κ_l $(l = 1, \ldots, N)$ are real and positive. Then the right scattering data S^+ are

defined by

$$S^+(t) = \left[r^+(k,t) = \frac{b(k,t)}{a(k,t)}, \{C_l^+(t), \kappa_l\}_{l=1}^N \right], \tag{A.5a}$$

and the left scattering data S^- by

$$S^-(t) = \left[r^-(k,t) = \frac{-b(-k,t)}{a(k,t)}, \{C_l^-(t), \kappa_l\}_{l=1}^N \right], \tag{A.5b}$$

where $C_l^+ = b_l/(ia_l')$ and $C_l^+ C_l^- = -1/(a_l')^2$ $[a_l' = \partial a(k)/\partial k|_{k=i\kappa_l}]$.

From (A.1b), the time evolution of the scattering data is given by

$$\begin{aligned} a(k,t) &= a(k,0) \equiv a(k), \\ b(k,t) &= b(k,0)\exp(8ik^3t), \\ b_l(t) &= b_l(0)\exp(8\kappa_l^3 t). \end{aligned} \tag{A.6}$$

For an N-soliton solution, $g(x,t;k)$ and $f(x,t;k)$ are given by (see for example [2])

$$g(x,t;k) = \left\{ 1 - i \sum_{l=1}^{N} \frac{C_l^+ e^{\kappa_l x}}{k + i\kappa_l} g(x,t;i\kappa_l) \right\} e^{ikx}, \tag{A.7a}$$

$$f(x,t;k) = \left\{ 1 - i \sum_{l=1}^{N} \frac{C_l^- e^{\kappa_l x}}{k + i\kappa_l} f(x,t;i\kappa_l) \right\} e^{-ikx}. \tag{A.7b}$$

The bound states $g(x,t;i\kappa_l)$ and $f(x,t;i\kappa_l)$ tend to one soliton forms asymptotically:

$$g(x,t;i\kappa_l) \sim \frac{1}{2}\left(\frac{2\kappa_l}{C_l^+}\right)^{1/2} \operatorname{sech} \kappa_l [x - x_l^+(t)] \quad \text{for} \quad t \to \infty, \tag{A.8a}$$

$$f(x,t,i\kappa_l) \sim \frac{1}{2}\left(\frac{2\kappa_l}{C_l^-}\right)^{1/2} \operatorname{sech} \kappa_l [x - x_l^-(t)] \quad \text{for} \quad t \to -\infty \tag{A.8b}$$

where $x_l^{\pm}(t)$ are the centers of the solitons, given by

$$x_l^+(t) = \frac{1}{2\kappa_l} \log\left[\frac{C_l^+(t)}{2\kappa_l} \prod_{m=1}^{l-1} \left(\frac{\kappa_m - \kappa_l}{\kappa_m + \kappa_l} \right)^2 \right], \tag{A.9a}$$

$$x_l^-(t) = -\frac{1}{2\kappa_l} \log\left[\frac{C_l^-(t)}{2\kappa_l} \prod_{m=1}^{l-1} \left(\frac{\kappa_m - \kappa_l}{\kappa_m + \kappa_l} \right)^2 \right]. \tag{A.9b}$$

Appendix B. Derivations of (2.4)–(2.7)

Here, we derive Equations (2.4)–(2.7), used in the finite perturbation method.

Let $u(x,t)$ and $u_0(x,t)$ be the two potential functions in (A.1), with scattering data corresponding to these potentials: $S^{\pm}$ and $S_0^{\pm}$. Consider the following identity:

$$\frac{f(x,k)g_0(y,k)}{a(k)} = g_0(y,k)\{g(x,-k)+r^+(k)g(x,k)\}, \tag{B.1}$$

where time t has been suppressed. We make an *Ansatz*

$$g(x,k) = g_0(x,k) + \int_x^\infty K(x,s)g_0(s,k)\,ds. \tag{B.2}$$

Substituting (B.2) into (B.1) and integrating (B.1) with respect to k from $-\infty$ to ∞, one can find the relation, for $y > x$,

$$\begin{aligned} i\sum_{l=1}^{N}\frac{b_l}{a_l'}g(x,i\kappa_l)g_0(y,i\kappa_l) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}g_0(x,-k)g_0(y,k)\,dk + \frac{1}{2\pi}\int_{-\infty}^{\infty}r^+(k)g_0(x,k)g_0(y,k)\,dk \\ &+ \frac{1}{2\pi}\int_x^\infty K(x,s)\int_{-\infty}^{\infty}g_0(s,-k)g_0(y,k)\,dk\,ds \\ &+ \frac{1}{2\pi}\int_x^\infty K(x,s)\int_{-\infty}^{\infty}r^+(k)g_0(s,k)g_0(y,k)\,dk\,ds \qquad \text{for} \quad y > x. \end{aligned} \tag{B.3}$$

Here we have used the fact that $f(x,k)$, $g_0(x,k)$, and $a(k)$ are analytically continuable to the upper half plane of k and, asymptotically,

$$\left.\begin{aligned} f(x,k) &\sim e^{-ikx} \\ g_0(x,k) &\sim e^{ikx} \\ a(k) &\to 1 \end{aligned}\right\} \qquad \text{as} \quad |k| \to \infty, \quad \operatorname{Im} k \geqslant 0. \tag{B.4}$$

On the other hand, there is a completeness relation,

$$\begin{aligned} \delta(x-y) = -i\sum_{l=1}^{N}\frac{b_{0l}}{a_{0l}'}g_0(x,i\kappa_l)g_0(y,i\kappa_l) \\ + \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{b_0(k)}{a_0(k)}f_0(x,k)g_0(y,k)\,dk. \end{aligned} \tag{B.5}$$

Taking (B.5) into account, (B.3) gives the Gel'fand-Levitan-Marchenko equation (2.4a) with $F(x, y)$ given by (2.7a).

From (B.2), we have (2.5a) and (2.6a), i.e.,

$$\left[\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} + u(x,t) - u_0(y,t)\right] K(x, y; t) = 0,$$

$$u(x,t) - u_0(x,t) = 2\frac{d}{dx} K(x, x; t).$$

It is obvious that $F(x, y; t)$ satisfies (2.5b). To derive the other expression (2.4b), consider the identity

$$\frac{g(x,k) f_0(y,k)}{a(k)} = f_0(y,k)\{f(x,-k) + r^-(k) f(x,k)\}. \tag{B.6}$$

Then, in a similar way as before, one can derive (2.4b)–(2.7b).

Acknowledgments

This work was partially supported by Mathematics Division of the Air Force Office of Scientific Research and the Office of Naval Research.

References

1. C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, *Phys. Rev. Lett.* 19:1095 (1967).
2. C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, *Comm. Pure Appl. Math.* 23:97 (1974).
3. M. Wadati and M. Toda, *J. Phys. Soc. Japan* 32:1403 (1972).
4. R. Hirota, *Phys. Rev. Lett.* 27:1192 (1971).
5. V. E. Zakharov, *Soviet Phys. JETP* 33:538 (1971).
6. M. J. Ablowitz and H. Segur, *Stud. Appl. Math.* 51:13 (1977).
7. V. E. Zakharov and S. V. Manakov, *Soviet Phys. JETP* 44:106 (1976).
8. H. Segur and M. J. Ablowitz, Asymptotic solutions of nonlinear evolution equations and Painlévé transcendents, Symposium on Solitons, Kiev, U.S.S.R., Sept. 1979; to appear.
9. H. Segur, *J. Fluid Mech.* 59:721 (1973).
10. A. B. Shabat, *Soviet Math. Dokl.* 14:1266 (1973).
11. E. A. Kuznetsov and A. V. Mikhailov, *Soviet Phys. JETP* 40:855 (1975).
12. H. Airault and M. J. Ablowitz, Finite perturbations and solutions of the Korteweg–de Vries equation, to appear.
13. V. E. Zakharov and L. D. Faddeev, *Functional Anal. Appl.* 5:280 (1971).

Clarkson College of Technology

(Received April 3, 1980)

Direct and inverse scattering problems of the nonlinear intermediate long wave equation

Y. Kodama
Bell Laboratories, Murray Hill, New Jersey 07974

M. J. Ablowitz
Department of Mathematics, Clarkson College, Potsdam, New York 13676

J. Satsuma[a)]
Department of Applied Mathematics and Physics, Faculty of Engineering, Kyoto University, Kyoto, Japan

(Received 17 March 1981; accepted for publication 2 October 1981)

The inverse scattering transformation method associated with a nonlinear singular integrodifferential equation is discussed. The equation describes long internal gravity waves in a stratified fluid of finite depth, and reduces to the Korteweg–de Vries equation as shallow water limit and the Benjamin–Ono equation as deep water limit. Both limits of the method and novel aspects of the theory are also discussed.

PACS numbers: 03.40.Kf, 02.30. + g

1. INTRODUCTION

In recent years, there has been considerable physical and mathematical interest in a certain nonlinear singular integrodifferential equation,[1–6]

$$u_t + (1/\delta)u_x + 2uu_x + Tu_{xx} = 0, \tag{1.1}$$

where $T(\cdot)$ is the singular integral operator given by

$$(Tf)(x) = \frac{1}{2\delta} P\int_{-\infty}^{\infty} \coth\frac{\pi(y-x)}{2\delta} f(y)\,dy \tag{1.2}$$

($P\int_{-\infty}^{\infty}$ represents the principal value integral). Physically, Eq. (1.1) describes the long internal gravity waves in a stratified fluid with finite depth (characterized by the parameter δ). Depending on the parameter δ, (1.1) reduces to the Korteweg–de Vries (KdV) equation as $\delta\to 0$ (shallow-water limit),

$$u_t + 2uu_x + (\delta/3)u_{xxx} = 0, \tag{1.3}$$

and the Benjamin–Ono (BO) equation as $\delta\to\infty$ (deep-water limit),

$$u_t + 2uu_x + Hu_{xx} = 0. \tag{1.4}$$

Here $H(\cdot)$ is the Hilbert transform given by

$$(Hf)(x) = \frac{1}{\pi} P\int_{-\infty}^{\infty} \frac{1}{y-x} f(y)\,dy. \tag{1.5}$$

Hence Eq. (1.1) is an intermediary equation between those two very interesting nonlinear evolution equations, (describing certain long wave motion). It is now known that (1.1) has an N-soliton solution,[3,4] infinite number of conservation laws, a Bäcklund transformation and a novel type of inverse scattering transform (IST) to solve the initial value problem.[6,7]

In this paper, we discuss in detail the direct and inverse problems of this new scattering problem. This paper serves to amplify and extend the results of our previous note.[7] In Sec. 2, we reformulate the IST scheme, originated in Ref. 6, with a specific analytical requirement in the complex x plane, say z plane. Specifically, the scattering problem may be viewed as a differential Riemann–Hilbert problem. In Sec. 3, we discuss the direct scattering problem, and define specific Jost functions in terms of a Green function. Then we show that the Jost function satisfies a Fredholm type integral equation. This is unlike the usual situations where we have local scattering problems (e.g., Schrödinger equation) and where the Jost function satisfies the Volterra integral equation. Several remarkable features of the Green function corresponding to the limits $\delta\to 0$, $\delta\to\infty$ are discussed in Appendix A. Using the results obtained in the previous sections, Sec. 4 is devoted to solving the inverse problem within a certain class of initial conditions. For this class of initial conditions we construct the linear integral equation (i.e., a Gel'fand–Levitan type equation) and hence give the direct connection between the solution of (1.1) and the scattering data defined in Sec. 3. In Sec. 5, using the Gel'fand–Levitan equation, we obtain an explicit form of the N-soliton solution, and in Sec. 6, taking the analyticity of the scattering data into account, we give the trace formula for the scattering function, and we express the conserved quanitities in terms of the scattering data. In Secs. 4–6 we keep δ finite in order for use to be sure of the appropriate analyticity of our Jost functions. We discuss the case $\delta = \infty$ (the BO limit) in Sec. 7, and we illustrate several remarkable properties of the scattering problem in this case. Our basic philosophy regarding the BO equation is to obtain the information by taking the limit process $\delta\to\infty$. However, we are careful to point out that we do not present here the solution to the initial value problem of the BO equation. Nevertheless we feel that the analysis presented here should be a basis for extension to the BO equation.

2. IST SCHEME

The IST Scheme[6] for (1.1) is given by

$$i\psi_x^+ + (u-\lambda)\psi^+ = \mu\psi^-, \tag{2.1}$$

[a)] Present address: General Education, Miyazaki Medical College, Miyazaki 899-16, Japan.

 0022-2488/82/040564-13$02.50 © 1982 American Institute of Physics

$$i\psi_t^{\pm} + 2i(\lambda + 1/2\delta)\psi_x^{\pm} + \psi_{xx}^{\pm} + [\mp iu_x - Tu_x + \nu]\psi^{\pm} = 0, \tag{2.2}$$

where λ and μ are constants given by $\lambda = -k \coth 2k\delta$, $\mu = k \operatorname{cosech} 2k\delta$, and ν is a constant determined by fixing the Jost functions of (2.1) (see Sec. 3). Here $\psi^{\pm}(x)$ represent the boundary values of functions

$\left[\text{i.e., } \psi^{\pm}(x) = \lim_{\operatorname{Im} z \to 0} \psi^{\pm}(z)\right]$ analytic in the horizontal strips between $\operatorname{Im} z = 0$ and $\operatorname{Im} z = \pm 2\delta$, and periodically extended vertically. By using the operator $T(\cdot)$, $\psi^{\pm}(x)$ may be written in the form

$$\psi^{+}(x) = \lim_{\operatorname{Im} z \downarrow 0} \psi(z) = \tfrac{1}{2}(1 - iT)\Psi(x), \tag{2.3a}$$

$$\psi^{-}(x) = \lim_{\operatorname{Im} z \uparrow 0} \psi(z) = -\tfrac{1}{2}(1 + iT)\Psi(x), \tag{2.3b}$$

where $\Psi(x)$ defined on the real axis is a proper fuction for the operator $T(\cdot)$ (i.e., $|\int_{-\infty}^{\infty} \Psi(x)\,dx| < \infty$, and $\Psi(x)$ satisfies the Hölder condition on the real axis, i.e., there exists constants C and h such that $|\Psi(x) - \Psi(y)| < C|x - y|^h$, $0 < h \leq 1$), and for $\operatorname{Im} z \neq 0 (\operatorname{mod} 2\delta)$, $\psi(z)$ is given by

$$\psi(z) = \frac{1}{4\delta i} \int_{-\infty}^{\infty} \coth \frac{\pi(y - z)}{2\delta} \Psi(y)\,dy. \tag{2.4}$$

From (2.3) and (2.4), we have a relation

$$T(\psi^{+} - \psi^{-}) = i(\psi^{+} + \psi^{-}). \tag{2.5}$$

It should be noted that the relation (2.5) due to the analyticity requirement is weaker than condition (27) in Ref. 6 in a sense (i.e., if $\ln \psi^{+}$ is analytic, then so is ψ^{+}). Furthermore, from the periodicity (period 4δ) of (2.4), we have the relation between $\psi^{\pm}(x)$ in the form of difference

$$\psi^{-}(x) = \psi^{+}(x + 2i\delta). \tag{2.6}$$

We also note that the same constant can be added to $\psi^{\pm}(x)$ without violating the analytically requirement. With regard to the limits $\delta \to 0$ and $\delta \to \infty$, we make some remarks.

Remark 1: As mentioned in Ref. 6, for $\delta \to 0$, the scattering problem (2.1) tends to the Schrödinger equation with

$\phi(x) = \lim_{\delta \to 0} \psi^{\pm}(x)$, i.e.,

$$\delta(\phi_{xx} + k^2\phi) + u\phi = 0, \tag{2.7}$$

and for $\delta \to \infty$,

$$i\psi_x^{+} + (u + k)\psi^{+} = 2k\psi^{-}, \quad \text{for } k > 0, \tag{2.8}$$

where $\psi^{\pm}(z)$ are the functions analytic in the upper and lower half z plane.

Remark 2: The formula (2.4) with the limit $\delta \to \infty$ is just the Plemelj formula, that is, as $\delta \to \infty$, $\frac{1}{2}(1 \mp iT)(\cdot)$ in (2.3) tend to the usual projection operators $P^{\pm}(\cdot) = \frac{1}{2}(1 \mp iH)(\cdot)$ that decompose a function into two functions analytic in the upper and lower half z plane.

3. DIRECT SCATTERING PROBLEM

Here and in Secs. 4–6 we assume for convenience that the initial condition $u(x,0)$ decays sufficiently rapidly as $|x| \to \infty$. In order to analyze the direct scattering problem of (2.1), it is convenient to define a new function,

$$W(x;k) \equiv \psi(x;k) \exp(ikx), \tag{3.1}$$

where we have defined $\psi(x;k) \equiv \psi^{\pm}(x \pm i\delta;k)$ and $W(x;k) \equiv W^{\pm}(x \pm i\delta;k)$ by taking (2.6) into account. Hereafter, for functions $f^{\pm}(x)$, we shall frequently use the notion $f(x)$ in which $f^{\pm}(x) = f(x \mp i\delta)$.

From (3.1) Eqs. (2.1) and (2.2) become

$$iW_x^{+} + (\zeta_{+} + 1/2\delta)(W^{+} - W^{-}) = -uW^{+}, \tag{3.2}$$

$$iW_t^{\pm} - 2i\zeta_{+}W_x^{\pm} + W_{xx}^{\pm} + [\mp iu_x - Tu_x + \rho]W^{\pm} = 0, \tag{3.3}$$

where $\rho = -2k\zeta_{+} + k^2 + \nu$, $\zeta_{\pm} = \zeta_{\pm}(k) = k \pm (k \coth 2k\delta - 1/2\delta)$ (we shall need the definition of ζ_{-} subsequently). The solution to (3.2) can be given by an integral equation,

$$W(x;k) = W_0(x;k) + \int_{-\infty}^{\infty} G(x, y;k)u(y)W^{+}(y;k)\,dy, \tag{3.4}$$

where $W_0(x;k)$ is the solution of the homogeneous equation of (3.2) [i.e., $u(x) = 0$], and $G(x,y;k)$ is a Green function satisfying

$$i\frac{\partial}{\partial x} G^{+}(x, y;k) + (\zeta_{+} + 1/2\delta)[G^{+}(x, y;k) - G^{-}(x, y;k)] = -\delta(x - y). \tag{3.5}$$

Here $G^{\pm}(x, y;k) = G(x \mp i\delta\sigma, y;k)$. From (3.5), we have the Fourier representation of the Green function $G(x, y;k)$,

$$G(x, y;k) = \frac{1}{2\pi} \int_C dp\, \hat{G}(p;k) e^{ip(x - y)} \tag{3.6}$$

in which $\hat{G}(p;k)$ are given by

$$\hat{G}(p;k) = \left[pe^{p\delta} \cdots 2k \frac{\sinh p\delta}{\sinh 2k\delta} e^{2k\delta}\right]^{-1} = \frac{p}{2} \operatorname{cosech}(p\delta) \left\{ p\left[\zeta_{+}\left(\frac{p}{2}\right) - \zeta_{+}(k)\right]\right\}^{-1}, \tag{3.7}$$

where the contour C is taken to be a contour (from $-\infty$ to ∞) determined by choosing the specific solution to (3.2) (see below). In Appendix A, we discuss the properties of the Green function in both limits $\delta \to 0$ and $\delta \to \infty$. From (3.7), we see that $\hat{G}(p;k)$ have poles at $p = 0$, $p = 2\zeta_{+}^{-1}[\zeta_{+}(k)]$. Since $\zeta_{+}^{-1}(\cdot)$ is the multivalued function, we have an infinite number of poles for which we shall define $p_{-1} = 0$, $p_0 = 2k$ and $p_n, \bar{p}_n$ $(n \geq 1)$ such that, for $n \geq 1$, $(2n - 1)\pi/(2\delta) < \operatorname{Im} p_n < (2n + 3)\pi/(2\delta)$ and similarly for $-\operatorname{Im} \bar{p}_n$. Moreover double zero poles occur at special values of $\zeta_{+}(k)$ satisfying $\zeta_{+}(k) = 0$ and $p_n = \zeta_{+}(k)$, $\bar{p}_n = \zeta_{+}(k)$ $(n \geq 1)$ [i.e., $p_n = p_{n-1}$, $\bar{p}_n = \bar{p}_{n-1}$ $(n \geq 0)$]. We call these values $\zeta_{+}^{(0)} = 0$ and $\{\zeta_{+}^{(l)}, \bar{\zeta}_{+}^{(l)}\}_{l=1}^{m}$ $(\operatorname{Im} \zeta_{+}^{(l)} > 0, \operatorname{Im} \bar{\zeta}_{+}^{(l)} < 0)$. Considering the equations of p_n $(n > 0)$,

$$\frac{dp_n}{d\zeta_{+}} = \frac{p_n}{2\delta(p_n - \zeta_{+})(\zeta_{+} + 1/2\delta)} \tag{3.8}$$

($\bar{p}_n$ satisfy same equation), one can see that there are logarithmic branch points at $\zeta_{+} = -1/2\delta$ for p_0, p_n and $\bar{p}_n$ $(n \geq 1)$, and square root branch points at $\{\zeta_{+}^{(l)}, \bar{\zeta}_{+}^{(l)}\}$ for p_0, at

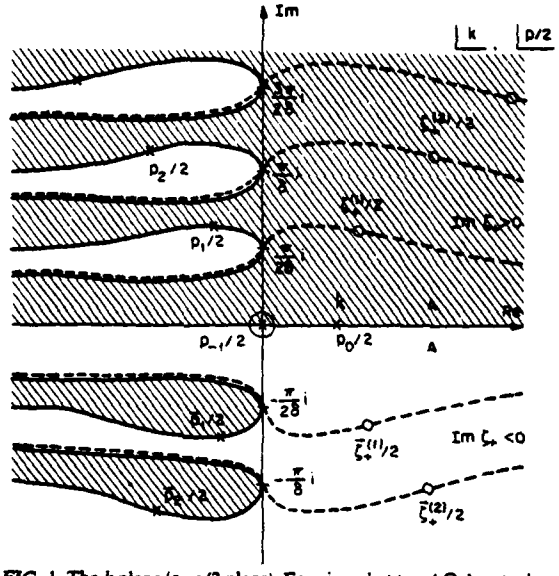


FIG. 1. The k plane (or $p/2$ plane). For given κ, $\times$ and $\circ$ denote the poles $p_{-1}=0, p_n\ (\pm p_i, \bar{p}_i)_{-1}^{n}$, and the double poles $\zeta_+^{(0)}=0, (\zeta_+^{(i)}, \bar{\zeta}_+^{(i)})_{-1}^{n}$ of Eq. (3.7), respectively. Each branch is surrounded by the dotted line, and branch A is the principal branch (PB). The shaded regions correspond to the upper half ζ_+ plane (multisheeted).

$\{\zeta_+^{(n)}, \zeta_+^{(n+1)}\}$ for p_n, and at $\{\bar{\zeta}_+^{(n)}, \bar{\zeta}_+^{(n+1)}\}$ for $\bar{p}_n$, respectively. It should be noted that from the multiplicity of k $(=\zeta_+^{-1}(\cdot))$, we are required to define an appropriate branch in k plane. We show in Fig. 1 the several of the branches in k plane, and in Fig. 2 the fundamental sheet (abbreviated hereafter as FS) corresponding to the principal branch (as PB), which is the portion A including the real k axis in Fig. 1.

From (3.2), one can see that if $W(x;k)$ is a solution, then $W(x;-k)\exp(2ikx)$ is also a solution. Taking this into account we now define specific Jost functions of (3.2). For real k, the Jost functions are defined as the solution to (3.2) with the boundary conditions

$$\left.\begin{aligned} M(x;k) &\to 1, \\ \bar{M}(x;k) &\to e^{2ikx}, \end{aligned}\right\} \quad \text{as } x \to -\infty, \tag{3.9a,b}$$

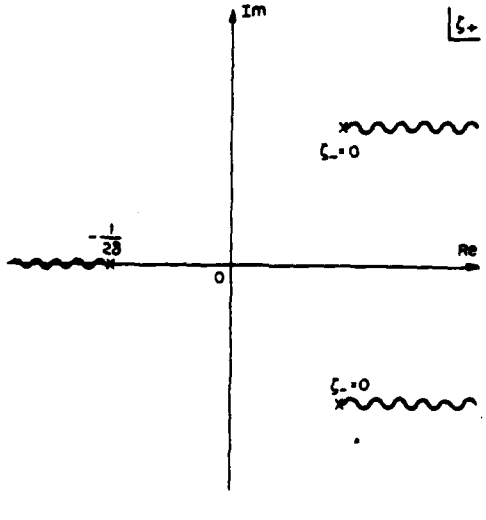


FIG. 2. The fundamental sheet (FS) corresponding to the principal branch. The wave lines show the branch cuts corresponding to the edges of PB.

$$\left.\begin{aligned} N(x;k) &\to e^{2ikx}, \\ \bar{N}(x;k) &\to 1, \end{aligned}\right\} \quad \text{as } x \to \infty. \tag{3.10a,b}$$

Here note that $\bar{M}(x;k)=M(x;-k)\exp(2ikx)$ and $\bar{N}(x;k)=N(x;-k)\exp(2ikx)$. Then in terms of the Green function (3.6), these Jost functions are given by

$$\begin{pmatrix} M(x;k) \\ \bar{M}(x;k) \end{pmatrix} = \begin{pmatrix} 1 \\ e^{2ikx} \end{pmatrix} = \int_{-\infty}^{\infty} G_1(x,y;k)u(y) \begin{pmatrix} M^+(y;k) \\ \bar{M}^+(y;k) \end{pmatrix} dy, \tag{3.11a,b}$$

$$\begin{pmatrix} N(x;k) \\ \bar{N}(x;k) \end{pmatrix} = \begin{pmatrix} e^{2ikx} \\ 1 \end{pmatrix} + \int_{-\infty}^{\infty} G_2(x,y;k)u(y) \begin{pmatrix} N^+(y;k) \\ \bar{N}^+(y;k) \end{pmatrix} dy, \tag{3.12a,b}$$

where the contours $C_{1,2}$ for $G_{1,2}(x,y;k)$ are taken to be the lines $\mathrm{Re}(p-i0)$, $\mathrm{Re}(p+i0)$, respectively [this is necessary in order to preserve the boundary conditions (3.9) and (3.10)]. Note that by taking these contours the Green functions $G_{1,2}(x,y;k)$ are bounded as $|x|\to\infty$. It is important to remark that (3.11) and (3.12) are Fredholm type integral equations, unlike the usual case of the local scattering problem (e.g., Schrödinger equation) where the Jost functions satisfy the Volterra type integral equations. In addition, we note that by using residue calculus (3.11) and (3.12) can be represented in an explicit manner useful for the proof of existence and analyticity of the solution (see Appendix B). As shown in Appendix B, we have the following propositions:

(i) $M(x;k)$, $\bar{N}(x;k)$ have convergent Neumann series in certain region of the FS for given δ and $\max|u|$ chosen small enough.

(ii) Despite the fact mentioned below Eq. (3.8) (i.e., the poles have square root branch points in the FS), $M(x;k)$ and $\bar{N}(x;k)$ are holomorphic in the upper and lower half plane of the FS, respectively. Moreover they are analytic in these regions, whenever the Neumann series converges in this region. In addition, we have the asymptotic form of $M(x;k)$ and $\bar{N}(x;k)$,

$$M(x;k) \to 1 + O\left(\frac{1}{\zeta_1}\right) \text{ as } |\zeta_+| \to \infty, \quad \mathrm{Im}\,\zeta_+ > 0, \tag{3.13}$$

$$\bar{N}(x;k) \to 1 + O\left(\frac{1}{\zeta_1}\right) \text{ as } |\zeta_+| \to \infty, \quad \mathrm{Im}\,\zeta_+ < 0. \tag{3.14}$$

We now define the appropriate scattering data corresponding to the Jost functions (3.11) and (3.12). For real k [i.e., $\zeta_+ > -1/(2\delta)$], by virtue of the fact that

$$G_1(x,y;k) - G_2(x,y;k) = \frac{1}{2i\delta\zeta_+} - \frac{1}{2i\delta\zeta_-}\exp[2ik(x-y) - 2k\delta], \tag{3.15}$$

we have alternative representations for $M(x;k)$ and $N(x;k)$, respectively,

$$M(x;k) = a(k) + b(k)e^{2ikx} + \int_{-\infty}^{\infty} G_2(x,y;k)u(y)M^+(y;k)\,dy, \tag{3.16}$$

$$N(x;k) = \bar{a}(k)e^{2ikx} + \bar{b}(k) + \int_{-\infty}^{\infty} G_1(x,y;k)u(y)N^+(y;k)\,dy. \tag{3.17}$$

Here $a(k)$, $b(k)$, $\bar{a}(k)$ and $\bar{b}(k)$ are given by

$$a(k) = 1 + \frac{1}{2i\delta\zeta_+}\int_{-\infty}^{\infty} dy\, u(y)M^+(y;k), \tag{3.18a}$$

$$b(k) = -\frac{1}{2i\delta\zeta_-}\int_{-\infty}^{\infty} dy\, u(y)M^+(y;k)e^{-2ik(y-i\delta)}, \tag{3.18b}$$

$$\bar{a}(k) = 1 + \frac{1}{2i\delta\zeta_-}\int_{-\infty}^{\infty} dy\, u(y)N^+(y;k)e^{-2ik(y-i\delta)}, \tag{3.19a}$$

$$\bar{b}(k) = \frac{1}{2i\delta\zeta_+}\int_{-\infty}^{\infty} dy\, u(y)N^+(y;k). \tag{3.19b}$$

When the solutions of the integral equations (3.16) and (3.17) are unique (or having a convergent Neumann series as a stronger condition), then we have

$$M(x;k) = a(k)\bar{N}(x;k) + b(k)N(x;k) \tag{3.20}$$

and

$$N(x;k) = \bar{a}(k)\bar{M}(x;k) + \bar{b}(k)M(x;k), \tag{3.21}$$

which are considered as "right" and "left" scattering equations, respectively. Moreover, as shown in Appendix C, one finds

$$\bar{a}(k) = a^*(-k), \tag{3.22a}$$

$$\bar{b}(k) = -b(-k) = -\left(\frac{d\zeta_+}{dk}\right)\left(\frac{d\zeta_-}{dk}\right)^{-1} b^*(k), \tag{3.22b}$$

$$|a(k)|^2 - \left(\frac{d\zeta_+}{dk}\right)\left(\frac{d\zeta_-}{dk}\right)^{-1}|b(k)|^2 = 1. \tag{3.23}$$

From (3.18a), we note that $a(k)$ takes on the same analyticity as $M(x;k)$, and $a(k)\to 1$ as $|\zeta_+|\to\infty$, $\mathrm{Im}\,\zeta_+ \geq 0$.

On the other hand, for $\zeta_+ + i0$ with $\zeta_+ < -1/(2\delta)$ and ζ_+ real [i.e., k is in the upper half plane at the edge of the *PB* where $\zeta_+(k) = \zeta_+ = \zeta_+(k^*)$] we have a relation

$$G_1(x,y;k) - G_2(x,y;k^*) = \frac{1}{2i\delta\zeta_+}, \tag{3.24}$$

from which we have

$$M(x;k) = a(k) + \int_{-\infty}^{\infty} G_2(x,y;k^*)u(y)M^+(y;k)\,dy, \tag{3.25}$$

$$\bar{N}(x;k^*) = \bar{a}(-k^*) + \int_{-\infty}^{\infty} G_1(x,y;k)u(y)\bar{N}^+(y;k^*)\,dy. \tag{3.26}$$

Here we have used the relation $\bar{N}(x;k) = N(x;-k)\exp(2ikx)$. From (3.25) and (3.26), we obtain

$$M(x;k) = a(k)\bar{N}(x;k^*), \tag{3.27}$$

$$\bar{N}(x;k^*) = \bar{a}(-k^*)M(x;k) \tag{3.28}$$

[i.e., $a(k)\bar{a}(-k^*) = 1$], whenever the solutions of (3.25) and (3.26) are unique.

The bound states defined as

$$M(x;k)\to 0 \quad \text{as} \quad x\to+\infty \tag{3.29}$$

and

$$N(x;k)\to 0 \quad \text{as} \quad x\to-\infty \tag{3.30}$$

are given by

$$a(k_l) = 0, \quad M(x;k_l) = b_l N(x;k_l) \tag{3.31}$$

and

$$\bar{a}(\bar{k}_l) = 0, \quad N(x;\bar{k}_l) = \bar{b}_l M(x;\bar{k}_l), \tag{3.32}$$

with $\bar{k}_l = -k_l^*$ for $l = 1,2,\ldots,N$. As shown in Appendix D, $a(k)$ has only simple zeroes and they lie on the imaginary k axis, i.e., $k_l = i\kappa_l = \bar{k}_l$, $0 < \kappa_l < \pi/2\delta$ in the PB, and $b_l\bar{b}_l = 1$. The right and left scattering datas are now given by

$$S_R = [a(k), b(k), \{\kappa_l, b_l\}_{l=1}^N], \tag{3.33}$$

$$S_L = [\bar{a}(k), \bar{b}(k), \{\kappa_l, \bar{b}_l\}_{l=1}^N]. \tag{3.34}$$

Let us now find the time evolution of the scattering data. By virtue of the boundary condition (3.9), the constant ρ in (3.3) is taken to be zero. From (3.3), (3.20), and (3.27), we obtain

$$a(k,t) = (k,0), \tag{3.35a}$$

$$b(k,t) = b(k,0)\exp[-4ik(\lambda + 1/2\delta)t], \tag{3.35b}$$

$$b_l(t) = b_l(0)\exp[4\kappa_l(\lambda_l + 1/2\delta)t], \tag{3.35c}$$

where $\lambda_l = \lambda(i\kappa_l) = -\kappa_l \cot 2\kappa_l\delta$.

Before closing this section, we offer several remarks:

Remark 1: Although there are infinitely many solutions to (3.2) corresponding to the poles of (3.7), we have chosen only a finite set of solutions as the Jost functions where the Green functions are bounded as $|x|\to\infty$ for real k, that is, (3.9) and (3.10). As shown in Appendix E, however, when we have a unique solution to the integral equation (3.4) [e.g., in the case that the Neumann series of the integral equation (3.4) converges], our set of the Jost functions (3.11) [or (3.12)] consists of a complete set of the functions in the sense of $L_2(-\infty < x < \infty)$.

Remark 2: In order to define the scattering data (3.33) [or (3.34)], we have assumed that the solution of the integral equation (3.4) is unique (and we have given sufficient conditions on δ, $\max|u|$ for this to hold), mathematically speaking, there is to be no nontrivial solution of the homogeneous equation of (3.4), i.e.,

$$W_h^+(x;k) = \int_{-\infty}^{\infty} G^+(x,y;k)u(y)W_h^+(y;k)\,dy. \tag{3.36}$$

However, in general, the solutions $W_h^+(x;k)$ may exist for some k [certainly for $\delta\to\infty$ (see Sec. 7) we refer to such values of k as "eigenvalues" of the Fredholm integral equation]. In this case, the solution to (3.4) may have a singularity, unless $W_0^+(x;k)$ satisfies

$$\int_{-\infty}^{\infty} [W_h^A(x;k)W_0(x;k)]^+\,dx = 0,$$

where $W_h^A(x;k)$ is the solution of the adjoint equation of (3.36),

$$[W_h^A(x;k)]^+ = u(x)\int_{-\infty}^{\infty} G^+(y,x;k)[W_h^A(y;k)]^+\,dy,$$

and we must add this eigenvalue into the scattering data (it may be a new kind of bound state). We shall discuss briefly such situations in Sec. 7.

4. INVERSE SCATTERING PROBLEM

We now discuss the inverse scattering problem associated with (3.2) in the case when the integral equations for the Jost functions have a unique solution and the Jost functions are analytic in certain region of the FS.

For $\bar{N}(x;k)$, from the analyticity and the asymptotic behavior (3.14), the following integral representation is suggested:

$$\bar{N}(x;k) = 1 + \int_x^\infty ds\, K(x,s) e^{i\zeta_+(x-s)} \quad \text{for } \operatorname{Im}\zeta_+ < 0. \tag{4.1}$$

Substituting (4.1) into (3.2) and (3.3), one finds that the kernel $K(x, y)$ satisfies

$$\{i\partial_x + 1/2\delta + u(x)\}K^+(x,y) + (i\partial_y - 1/2\delta)K^-(x,y) = 0, \tag{4.2}$$

$$\{i\partial_t + \partial_x^2 - \partial_y^2 + 2(\partial_x K(x,x))\}K(x,y) = 0, \tag{4.3}$$

with

$$u(x) = iK^+(x,x) - iK^-(x,x), \tag{4.4}$$

where $K^\pm(x,y) = K(x \mp i\delta, y \mp i\delta)$ and $K(x,y) \to 0$ as $y \to \infty$. It is important to note that (4.4) is a decomposition of $u(x)$ (see Sec. 2), and from (2.5) we also have the relation,

$$(Tu)(x) = -\{K^+(x,x) + K^-(x,x)\}. \tag{4.5}$$

Subject to (4.4) and (4.5), one can see that the compatibility between (4.2) and (4.3) gives Eq. (1.1).

From (4.1) and the relation $N(x;k) = \bar{N}(x; -k)\exp(2ikx)$, we also have

$$N(x;k) = e^{2ikx}\left[1 + \int_x^\infty ds\, K(x,s) e^{-i\zeta_-(x-s)}\right], \tag{4.6}$$

which is analytic in the upper half ζ_- plane [note $\zeta_+(-k) = -\zeta_-(k)$]. By virtue of the triangular representations (4.1) and (4.6) one can derive the linear integral equation (i.e., a Gel'fand–Levitan type equation) and hence solve the inverse problem as follows: dividing (3.20) and (3.27) by $a(k)$ and operating with $(1/2\pi)\int_{-\infty}^{\infty} d\zeta_+ \exp[i\zeta_+(y-x)]$. (i.e., Fourier transform) for $y > x$, we have

$$\frac{1}{2\pi}\int_{-\infty+i0}^{\infty+i0} d\zeta_+ \frac{M(x;k)}{a(k)} e^{i\zeta_+(y-x)} = \frac{1}{2\pi}\int_{-\infty-i0}^{\infty-i0} d\zeta\, \bar{N}(x;k) e^{i\zeta_+(y-x)} + \frac{1}{2\pi}\int_{-1/2\delta}^{\infty} d\zeta_+ \frac{b(k)}{a(k)} N(x;k) e^{i\zeta_+(y-x)}. \tag{4.7}$$

From the analyticity arguments for $M(x;k)$ and $a(k)$ (i.e., analytic on the upper half plane of the FS), the left-hand side of (4.7) can be written in the form

$$-\sum_{l=1}^{N} C_l N(x;i\kappa_l)\exp[i\zeta_{+l}(y-x)], \tag{4.8}$$

where $C_l = -ib_l/\dot{a}_l$, $\dot{a}_l = \partial a/\partial\zeta_+|_{\zeta_+=\zeta_{+l}}$, $\zeta_{+l} = \zeta_+(i\kappa_l)$. Then, using (4.1) and (4.6), we obtain the Gel'fand–Levitan equation,

$$K(x,y) + F(x,y) = \int_x^\infty K(x,s)F(s,y)\,ds = 0 \quad \text{for } y > x, \tag{4.9}$$

with

$$F(x,y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk \left(\frac{d\zeta_+}{dk}\right)\frac{b(k)}{a(k)} e^{i\zeta_- x + i\zeta_+ y} + \sum_{l=1}^{N} C_l \exp(i\zeta_{-l}x + i\zeta_{+l}y), \tag{4.10}$$

(note $\int_{-\infty}^{\infty} dk\,(d\zeta_+/dk)\cdot = \int_{-1/2\delta}^{\infty} d\zeta_+ \cdot$). In Sec. 5, by using (4.9), we construct the N-soliton solution as an example of an explicit solution of (1.1).

To close this section, we briefly discuss an alternative method of the inverse problem which is a direct method[8] based upon the Gel'fand–Levitan equation (4.9). From (4.10), taking $(\zeta_+ + 1/2\delta) = (-\zeta_- + 1/2\delta)\exp(4k\delta)$ into account, one finds that $F(x,y)$ satisfies

$$(i\partial_x + 1/2\delta)F^+(x,y) + (i\partial_y - 1/2\delta)F^-(x,y) = 0. \tag{4.11}$$

From (3.35), we also have

$$(i\partial_t + \partial_x^2 - \partial_y^2)F(x,y) = 0. \tag{4.12}$$

From (4.9), (4.11), and (4.12) one can derive (4.2) and (4.3) with (4.4) whose compatibility yields (1.1) with (4.5). Then the direct method based upon (4.9) is given as follows: For given $u(x,0)$,

(1) find $K^\pm(x,y,0)$ from (4.2) with the boundary condition $K^\pm(x,x,0) = \mp iu^\pm(x,0)$ [$u(x,0) = u^+(x,0) + u^-(x,0)$ from (4.4)],

(2) find $F^\pm(x,y,0)$ from (4.9),

(3) find $F^\pm(x,y,t)$ from (4.12),

(4) find $K^\pm(x,y,t)$ from (4.9),

(5) then $u(x,t) = iK^+(x,x,t) - iK^-(x,x,t)$.

5. N-SOLITON SOLUTION

In the case of an N-soliton solution, $F(x,y)$ in the Gel'fand–Levitan equation is given by

$$F(x,y) = \sum_{l=1}^{N} C_l \exp(i\zeta_{-l}x + i\zeta_{+l}y), \tag{5.1}$$

where C_l is a positive real function of t (see Appendix D). In order to solve (4.9) with (5.1), we assume $K(x,y)$ to be of the form

$$K(x\,y) = \sum_{l=1}^{N} (C_l)^{1/2}\Gamma_l(x)\exp(i\zeta_{+l}y). \tag{5.2}$$

Then (4.9) becomes

$$\Gamma_n + i\sum_{l=1}^{N} \frac{(C_nC_l)^{1/2}e^{i(\zeta_{-n}+\zeta_{+l})x}}{\zeta_{-n}+\zeta_{+l}}\Gamma_l = -C_n^{1/2}e^{i\zeta_{-n}x}. \tag{5.3}$$

Here we note that the matrix $\Delta(x)$ defined by

$$[\Delta(x)]_{nl} = i(C_nC_l)^{1/2}\frac{\exp[i(\zeta_{-n}+\zeta_{+l})x]}{\zeta_{-n}+\zeta_{+l}} \tag{5.4}$$

is positive definite. Namely, for an arbitrary column vector $\mathbf{V} = (v_1,\ldots,v_N)^T$ the inner product $(\mathbf{V},\Delta\mathbf{V}) \equiv (\mathbf{V}^T)^*\Delta\mathbf{V}$ takes positive value, i.e., noting $(\zeta_{+l})^* = -\zeta_{-l}$,

$$(\mathbf{V},\Delta\,\mathbf{V}) = i\sum_{n,l} v_n^* \frac{(C_n C_l)^{1/2} e^{i(\zeta_{-n}+\zeta_{+l})x}}{\zeta_{-n}+\zeta_{+l}} v_l$$

$$= \int_x^\infty \left| \sum_n^N C_n^{1/2} v_n \exp(i\zeta_{+n} s) \right|^2 ds > 0. \tag{5.5}$$

Therefore, for $\Gamma \equiv (\Gamma_1,\ldots,\Gamma_N)^T$, we have

$$\Gamma = (I+\Delta)^{-1}\mathbf{E} \tag{5.6}$$

where $\mathbf{E} \equiv (E_1,\ldots,E_N)^T$ with the element $E_l = -C_l^{1/2}\exp(i\zeta_{-l}x)$. From (5.6), we obtain

$$K(x,x) = \frac{\partial}{\partial x} \ln \det |I+\Delta(x)|, \tag{5.7}$$

which leads to

$$u(x,t) = i\frac{\partial}{\partial x} \ln \det \left| \frac{I+\Delta(x-i\delta,t)}{I+\Delta(x+i\delta,t)} \right|, \tag{5.8}$$

where I is the identity matrix. For example, a 1-soliton solution is given by

$$u(x,t) = \frac{2\kappa_1 \sin(2\kappa_1\delta)}{\cos(2\kappa_1\delta) + \cosh\{2\kappa_1[x-x_1(t)]\}}, \tag{5.9}$$

where κ_1 is an eigenvalue, $k = i\kappa_1$, and from (3.35) $x_l(t)$ is given by

$$x_1(t) = \frac{1}{2\kappa_1} \ln \frac{C_1(t)}{2\kappa_1}$$

$$= (2\lambda_1 + \delta^{-1})t + x_1(0). \tag{5.10}$$

For N-soliton solution, following the method developed in Ref. 9, one can calculate the phase shift formula for l th-soliton,

$$x_l^+(0) - x_l^-(0)$$

$$= \frac{1}{\kappa_l} \ln \left[\prod_{m=l+1}^{N} \left| \frac{\zeta_{+l}+\zeta_{-m}}{\zeta_{+l}-\zeta_{+m}} \right| \prod_{m=1}^{l-1} \left| \frac{\zeta_{+m}-\zeta_{+l}}{\zeta_{-m}+\zeta_{+l}} \right| \right]. \tag{5.11}$$

Here we have ordered $\kappa_l > \kappa_2 > \cdots > \kappa_N > 0$, and $x_l^{\pm}(0)$ are the phases of the soliton at $t \to \pm\infty$, respectively.

It is interesting to note that the N-soliton solution (5.8) can be written in terms of the squared eigenfunctions. From (5.3) we have

$$C_l N_l = -C_l^{1/2}\Gamma_l \exp(i\zeta_{+l}x) \tag{5.12}$$

and

$$N_n = e^{-2\kappa_n x}\left[1 - i\sum_{l=1}^{N} \frac{C_l N_l}{\zeta_{-n}+\zeta_{+l}}\right], \tag{5.13}$$

where $N_l(x) = N(x;i\kappa_l)$. Then we obtain

$$u(x) = iK^+(x,x) - K^-(x,x)$$

$$= -i\sum_{l=1}^{N} C_l[N_l^+(x) - N_l^-(x)]$$

$$= 2\sum_{l=1}^{N} C_l \sin(2\kappa_l\delta)|N_l^+|^2 e^{2\kappa_l x}. \tag{5.14}$$

On the other hand, from (2.1) or (3.2), one finds that the equation for $\Psi \equiv |\psi^+|^2 = |W^+\exp[-ik(x-i\delta)]|^2$ with k real or pure imaginary satisfies

$$\Psi_t + (1/\delta)\Psi_x + 2u\Psi_x + T\Psi_{xx} = 0, \tag{5.15}$$

which is the associated linear equation of (1.1). Note that this result is similar to the case of the KdV equation.[9] In fact, the results obtained in Ref. 9 for the KdV solitons are derived by taking $\lim \delta \to 0$ in our results.

6. CONSERVATION LAWS

As shown in Ref. 6, (1.1) has an infinite number of conservation laws. Here we give the direct connection between those conserved quantities and the scattering data defined in the Sec. 3. We first show that the function defined by

$$\sigma(x;k) = \ln \frac{M^-(x;k)}{M^+(x;k)} \tag{6.1}$$

is directly related to the scattering data $a(k)$. Noting that, for large ζ_+ in the upper half plane of the FS, $M^{\pm}(x;k)$ are not zero, we have a relation

$$(T\sigma)(x;k) = -i\ln[M^+(x;k)M^-(x;k)A], \tag{6.2}$$

for large $|\zeta_+|$, $\operatorname{Im}\zeta_+ > 0$, where we have used the fact mentioned below Eq. (2.6), and the constant A can be determined by the boundary condition (see below). It should be noted that $\sigma(x) \to 0$ as $|x| \to \infty$, since for $\operatorname{Im}\zeta_+ > 0$ in the FS, $M^{\pm}(x;k) \to 1$ as $x \to -\infty$ and $M^{\pm}(x;k) \to a(k)$ as $x \to +\infty$. From (6.1) and (6.2), (4.2) and (4.3) become

$$e^{\sigma} - 1 = \frac{1}{2\zeta_+}\left[\frac{1}{\delta}(1-e^{\sigma}) - i\sigma_x - T\sigma_x + 2u\right], \tag{6.3}$$

$$\sigma_t - 2\zeta_+\sigma_x - i\sigma_{xx} + \sigma_x T\sigma_x + 2u_x = 0. \tag{6.4}$$

Taking $\int_{-\infty}^{\infty} \sigma_x T\sigma_x\,dx = 0$ into account, from (6.4), one finds that $\sigma(x)$ is a conserved density. For large $|\zeta_+|$, (6.3) has an asymptotic expansion

$$\sigma(x;k) = \sum_{n=1}^{\infty} \frac{\phi_n(x)}{(2\zeta_+)^n}. \tag{6.5}$$

The first two $\phi_n(x)$ are

$$\phi_1 = 2u,$$

$$\phi_2 = -2u^2 - \frac{2}{\delta}u - 2iu_x - 2T(u_x). \tag{6.6}$$

On the other hand, from the definition of $T(\cdot)$, we have

$$(T\sigma)(x;k) \to \begin{cases} -\dfrac{1}{2\delta}\displaystyle\int_{-\infty}^{\infty} \sigma(y;k)\,dy & \text{as } x \to \infty, \\ \dfrac{1}{2\delta}\displaystyle\int_{-\infty}^{\infty} \sigma(y;k)\,dy & \text{as } x \to -\infty, \end{cases} \tag{6.7}$$

and from the boundary conditions of $M^{\pm}(x;k)$ for $\operatorname{Im}\zeta_+ > 0$,

$$\ln[M^+(x;k)M^-(x;k)A] \to \begin{cases} \ln[a^2(k)A] & \text{as } x \to +\infty, \\ \ln A & \text{as } x \to -\infty. \end{cases} \tag{6.8}$$

Comparing (6.7) with (6.8), we obtain

$$\ln a(k) = -\frac{i}{2\delta}\int_{-\infty}^{\infty} \sigma(y;k)\,dy. \tag{6.9}$$

Next we derive a closed form expression for $a(k)$ by using the analyticity requirements of the previous sections.

Namely, we employ the following facts:

(i) $a(k)$ is analytic for $\operatorname{Im}\zeta_+ > 0$.

(ii) $a(k)$ has only a finite number of simple zeros.

(iii) $a(k) \to 1$, as $|\zeta_+| \to \infty$ for $\operatorname{Im}\zeta_+ \geq 0$.

(iv) $a(k) = \bar{a}^*(-k^*)$ for real $\zeta_+(k)$.

In this case, $a(k)$ can be written in the form

$$a(k) = \hat{a}(k) \prod_{l=1}^{N} \frac{\zeta_+ - \zeta_{+l}}{\zeta_+ + \zeta_{-l}}, \tag{6.10}$$

where, for $\operatorname{Im}\zeta_+ \geq 0$, $\hat{a}(k)$ has no zeros and $\hat{a}(k) \to 1$ as $|\zeta_+| \to \infty$. From (6.10) we obtain

$$a(k) = \left(\prod_{l=1}^{N} \frac{\zeta_+ - \zeta_{+l}}{\zeta_+ + \zeta_{-l}} \right) \times \exp\left[\frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{d\zeta'_+}{\zeta'_+ - \zeta_+} \ln|a[k'(\zeta'_+)]|^2 \right], \quad \text{for } \operatorname{Im}\zeta_+ > 0, \tag{6.11}$$

$$a(k) = \lim_{\operatorname{Im}\zeta_+ \to 0} a(k), \quad \text{for real } \zeta_+.$$

For large $|\zeta_+|$, we have the asymptotic expansion,

$$\ln a(k) = \frac{1}{2\delta} \sum_{n=1}^{\infty} \frac{1}{(2\zeta_+)^n} \left[\frac{2^n}{\pi} \int_{-\infty}^{\infty} (\zeta'_+)^{n-1} \ln|a(k')|^2 \, d\zeta'_+ + \frac{2^{n+1}\delta i}{n} \sum_{l=1}^{N} \{\zeta_{+l}^n - \zeta_{+l}^{*n}\} \right]. \tag{6.12}$$

Comparing (6.5) with (6.12), we now obtain the trace formula

$$I_n \equiv \int_{-\infty}^{\infty} \phi_n(y)\, dy = -\frac{2^n\delta}{\pi} \int_{-\infty}^{\infty} d\zeta_+ \, \zeta_+^{n-1} \ln|a(k)|^2 - \frac{2^{n+1}}{n} i\delta \sum_{l=1}^{N} \{\zeta_{+l}^n - \zeta_{+l}^{*n}\}. \tag{6.13}$$

From the first two terms of I_n, we have

$$\int_{-\infty}^{\infty} u\, dx = 4 \sum_{l=1}^{N} \kappa_l \delta - \frac{\delta}{\pi} \int_{-\infty}^{\infty} d\zeta_+ \ln|a(k)|^2, \tag{6.14a}$$

$$\int_{-\infty}^{\infty} u^2\, dx = 8 \sum_{l=1}^{N} \kappa_l \delta \left(\frac{1}{2\delta} - \kappa_l \cot(2\kappa_l \delta_l) \right) + \frac{2\delta}{\pi} \int_{-\infty}^{\infty} d\zeta_+ \, \zeta_+ \ln|a(k)|^2. \tag{6.14b}$$

7. NOTES ON THE BENJAMIN–ONO EQUATION

In Secs. 4–6, we developed a method associated with the IST problem of (1.1) in the case where the Neumann series corresponding to the integral equation (3.4) converges. Sufficient conditions for this to hold are to have δ *finite* and $\max|u|$ chosen small enough (see Appendix B). We now briefly consider the scattering problem in the case $\delta = \infty$, and we find several results associated with the BO equation by solving the scattering problem. Our basic philosophy regarding this case is to obtain information by taking the limit process $\delta \to \infty$.

From (3.2) and (3.3) with $\delta \to \infty$, the IST scheme of the BO equation is given by

$$iw_x^+ + 2k(w^+ - w^-) = -uw^+, \tag{7.1}$$

$$iw_t^\pm - 4ikw_x^\pm + w_{xx}^\pm + [\mp 2iP^\pm(u_x) + \rho]w^\pm = 0, \tag{7.2}$$

where $P^\pm(\cdot)$ are the usual projection operators given by $P^\pm(\cdot) = \frac{1}{2}[1 \mp iH](\cdot)$, and $w^\pm(x)$ are the boundary values of the functions analytic in the upper and lower half z plane, respectively. Here we require that the *derivatives* of $w^\pm(x;k)$ satisfy

$$Hw_x^\pm = \pm iw_x^\pm \tag{7.3}$$

With (3.11) and (3.12), we define the Jost functions of (7.1) as follows. For real $k > 0$,

$$m^\pm(x;k) = \lim_{\delta \to \infty} M^\pm(x;k), \tag{7.4a}$$

$$\bar{m}^\pm(x;k) = \lim_{\delta \to \infty} \bar{M}^\pm(x;k) \exp(-2k\delta), \tag{7.4b}$$

$$n^\pm(x;k) = \lim_{\delta \to T} N^\pm(x;k) \exp(-2k\delta), \tag{7.5a}$$

$$\bar{n}^\pm(x;k) = \lim_{\delta \to \infty} \bar{N}^\pm(x;k). \tag{7.5b}$$

As mentioned in Appendix A, these Jost functions are the solutions of the split equations

$$iw_x^+ + 2k(w^+ - w_0) = -P^+(uw^+), \tag{7.6a}$$

$$2k(w^- - w_0) = P^-(uw^+), \tag{7.6b}$$

where w_0 *is a constant determined by the boundary condition.* The Green's functions associated with (7.4) and (7.5), defined by $g_{1,2}^\pm(x,y;k) = \lim_{\delta \to \infty} G_{1,2}^\pm(x,y;k)$, are given in the explicit form, for real $k > 0$,

$$g_1^+(x,y;k) = ie^{2ik(x-y)}\left\{ \theta(x-y) - \frac{i}{2\pi} E_1[2ik(x-y)] \right\}, \tag{7.7a}$$

$$g_2^+(x,y;k) = ie^{2ik(x-y)} + g_1^+(x,y;k), \tag{7.7b}$$

$$g_1^-(x,y;k) = \frac{1}{4\pi i k(x-y-i0)}, \tag{7.8a}$$

$$g_2^-(x,y;k) = g_1^-(x,y;k), \tag{7.8b}$$

where $\theta(\cdot)$ is the usual Heaviside step function, and $E_1(z)$ is the exponential integral,

$$E_1(z) = \int_z^{\infty} \frac{e^{-t}}{t}\, dt, \quad \text{for } |\arg z| < \pi, \tag{7.9}$$

and asymptotically $E_1(z) \to O(e^{-z}/z)$ as $|z| \to \infty$. The integral equations for m^+ and n^+ are found in the same way as in Sec. 3; hence, from (7.7), we have, for real $k > 0$,

$$m^+(x;k) = 1 + \beta(k)e^{2ikx} + \int_{-\infty}^{\infty} g_2^+(x,y;k)u(y)m^+(y;k)\, dy, \tag{7.10}$$

$$n^+(x;k) = \bar{a}(k)e^{2ikx} + \int_{-\infty}^{\infty} g_1^+(x,y;k)u(y)n^+(y;k)\, dy, \tag{7.11}$$

which leads to

$$m^+(x;k) = \bar{n}^+(x;k) + \beta(k)n^+(x;k), \tag{7.12}$$

$$\bar{n}(x;k) = \bar{a}(k)\bar{m}^+(x;k), \tag{7.13}$$

whenever the solutions of (7.10) *and* (7.11) *are unique*. Here $\beta(k)$ and $\bar{a}(k)$ are given by

$$\beta(k) = \lim_{\delta\to\infty} b(k)\exp(2k\delta) = i\int_{-\infty}^{\infty} u(y)m^+(y;k)e^{-2iky}\,dy, \tag{7.14}$$

$$\bar{a}(k) = \lim_{\delta\to\infty} \bar{a}(k) = 1 - i\int_{-\infty}^{\infty} u(y)n^+(y;k)e^{-2iky}\,dy \tag{7.15}$$

[note that, for real $k>0$, $a(k)\to 1$, $\bar{b}(k)\to 0$ as $\delta\to\infty$]. So far we have discussed the Jost functions for real $k>0$. A remarkable feature of the scattering problem (7.1) may arise for real $k<0$. In this case, the Green functions become

$$g_1^+(x,y;k) = g_2^+(x,y;k) = (1/2\pi)E_1[2ik(x-y)]e^{2ik(x-y)}, \tag{7.16}$$

and $g_{1,2}^-(x,y;k)$ are the same as (7.8). Then, the Jost functions are given by

$$m^+(x;k) = 1 + \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{2ik(x-y)} E_1[2ik(x-y)]u(y)m^+(y;k)\,dy, \tag{7.17}$$

$$n^+(x;k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{2ik(x-y)}E_1[2ik(x-y)]u(y)n^+(y;k)\,dy, \tag{7.18}$$

where we have omitted the term $\exp(2ikx)$ in (3.12a), since this function is not analytic on the upper half z plane ($\mathrm{Re}\,z = x$) for $k<0$ in the sense of (7.3). We note that if there is a nontrivial solution of (7.18), then, by the asymptotics of $E_1(x)$, $n^+(x;k)\to O(1/x)$ as $|x|\to\infty$ [i.e., this solution may be considered to be a bound state for (7.1)]. By the method of successive approximations,[10] one can find that if the kernel $K(x,y;k) \equiv (2\pi)^{-1}E_1[2ik(x-y)]u(y)\exp[2ik(x-y)]$ satisfies the following square integral conditions,

$$S_1(x) \equiv \int_{-\infty}^{\infty} |K(x,y;k)|^2\,dy < \infty,$$

$$S_2(y) \equiv \int_{-\infty}^{\infty} |K(x,y;k)|^2\,dx < \infty, \tag{7.19}$$

$$S_0 \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |K(x,y;k)|^2\,dx\,dy < \infty,$$

and if $S_0<1$, then there is only the trivial solution to (7.18), i.e., $n^+(x;k)=0$. For real k, the condition $S_0<1$ gives

$$\int_{-\infty}^{\infty} [u(x)]^2\,dx < 4\pi|k|. \tag{7.20}$$

This implies that, for given $u(x)$, one may expect to have a nontrivial solution for those k satisfying $|k| < (4\pi)^{-1}\int_{-\infty}^{\infty}[u(x)]^2\,dx$. In order to illustrate some of those features [e.g., the existence of nontrivial solutions of (7.18)] of the scattering problem (7.1), we discuss a simple example. Namely consider the following initial condition,

$$u(x) = \frac{2\nu\epsilon}{x^2+\epsilon^2}, \tag{7.21}$$

where ν and ϵ are constants, and without loss of generality ϵ can be taken 1 by virtue of the scaling asymmetry, $x\to\epsilon x$, $t\to\epsilon^2 t$, $u\to u/\epsilon$ (the BO equation with this initial condition has been discussed numerically in Ref. 11). The method discused below can be applied to the case in which $u(x)$ takes more general form of a rotational function. For a bound state (i.e., real $k<0$), by virtue of Cauchy's theorem, we immediately obtain $n^-(x;k)$, from (7.6b),

$$\begin{aligned} n^-(x;k) &= \frac{1}{2k}P^-(un^+)(x;k) \\ &= \frac{1}{4\pi ik}\int_{-\infty}^{\infty}\frac{1}{x-y-i0}\left(\frac{2\nu}{y^2+1}\right)n^+(y;k)\,dy \\ &= \frac{1}{2k}\,\mathrm{Res}\left[\frac{1}{x-y-i0}\left(\frac{2\nu}{y^2+1}\right)n^+(y;k)\right] \\ &= -\frac{i\nu n^+(i;k)}{2k(x-i)}, \end{aligned} \tag{7.22}$$

where we have used the requirement that $n^+(y;k)$ is analytic on the upper half plane. It should be noted that $n^-(x;k)$ can be calculated explicitly and the degree of singularity in the upper half plane is the same as the one of $u(x)$ when $u(x)$ is the rational function. Then the solution $n^+(x;k)$ can be found directly from (7.1) in the form

$$\frac{n^+(x;k)}{n^+(i;k)} = -\nu\left(\frac{x-i}{x+i}\right)^{\nu}\int_{-\infty}^{x}\left(\frac{y+i}{y-i}\right)^{\nu}\times\frac{\exp[2ik(x-y)]}{y-i}\,dy, \tag{7.23}$$

with the boundary condition ($n^+(x;k)\to 0$ as $|x|\to\infty$)

$$D_\nu(k) \equiv \int_{-\infty}^{\infty}\left(\frac{y+i}{y-i}\right)^{\nu}\frac{\exp(-2iky)}{y-i}\,dy = 0, \tag{7.24}$$

which determines the discrete eigenvalues (see below). For $\nu<0$, integrating (7.23) by parts, one may see that $n^+(x;k)$ has a singular point at $z=i$, and there is no bound state. As a special case of $\nu>0$, we first consider the case $\nu=n=$ integer. By a residue calculation, one can see that (7.24) is the Laguerre polynomial of degree n, i.e., $D_n(k) = (2\pi i)^{-1}L_n(-4k)\exp(4k) = 0$ for $k<0$. Hence there are n real distinct eigenvalues, e.g., for $n=1$, $k_1 = -\frac{1}{4}$, for $n=2$, $k_{1,2} = -(2\pm\sqrt{2})/4$, and so on. On the other hand, (7.23) can be written in the form,

$$\frac{n^+(x;k)}{n^+(i;k)} = -\nu\left(\frac{x-i}{x+i}\right)^n e^{2ikx}\left[\sum_{l=0}^{n}\binom{n}{l}(2i)^{n-l}I_l^{(n)}(x;k)\right], \tag{7.25}$$

TABLE I. The eigenvalues of the Fredholm integral equation (7.18).

ν	0.8	1.0	1.5	2.0	2.5	3.0
	0.48	1.0	0.12	0.59	0.085	0.42
$-4k$			1.97	3.41	1.28	2.29
					4.69	6.29

where $I_l^{(n)}(x;k)$ is given by

$$I_l^{(n)}(x;k) = \int_{-\infty}^{x} \frac{\exp(-2iky)}{(y-i)^{n-l+1}}\,dy$$
$$= -\frac{1}{n-l}\left[\frac{\exp(-2ikx)}{(x-i)^{n-l}} - 2ikI_{l+1}^{(n)}(x;k)\right]$$
$$= \cdots\cdots$$
$$= \sum_{m=1}^{n-l}\frac{\gamma_m e^{-2ikx}}{(x-i)^m} + \frac{(-2ik)^{n-l}}{(n-l)!}I_n^{(n)}(x;k). \tag{7.26}$$

Here γ_m is the constant to be calculated recursively, and $I_n^{(n)}(x;k)$ is the exponential integral $E_i[2ik(x-i)]$ which has a logarithmic singularity at $x=i$. However, the coefficient of $I_n^{(n)}$ in $n^+(x;k)$ is just the Laguerre polynomial, i.e.,

$$L_n(-4k) = \sum_{l=0}^{n}\binom{n}{l}\frac{(4k)^{n-l}}{(n-l)!}, \tag{7.27}$$

which is required to be zero by the condition (7.24). Namely, the condition (7.24) corresponds to the requirement that $n^-(x;k)$ is, in fact, analytic in the upper half plane. Thus we expect to find an n-soliton solution when $\nu = n$ (in agreement with Ref. 11). By direct calculations, one can show that the solution (7.25) with (7.27) = 0 satisfies (7.18) [i.e., this shows the existence of the nontrivial solution to the homogeneous integral equation (7.18)]. For the situation with $\nu \neq$ integer, we note that (7.24) can be expressed by

$$(1-e^{2\pi i\nu})\int_0^\infty \frac{(y+1)^\nu e^{4ky} - \sum_{l=0}^{[\nu]}\alpha_l y^l}{y^{l+\nu}}\,dy = 0, \tag{7.28}$$

where $[\nu]$ denotes the Gauss's symbol (i.e., maximum integer less than ν), and the series $\sum_{l=0}^{[\nu]}\alpha_l y^l$ is the first $[\nu]+1$ terms of the Taylor expansion of $(y+1)^\nu \exp(4ky)$ around $y=0$. From (7.28), we find that there are n roots of (7.28) when ν is in the range $n-1<\nu<n$. We have listed the values of k versus ν in Table I.

ACKNOWLEDGMENTS

Work at Clarkson College was supported by the U.S. Office of Naval Research, Mathematics Division under Grant No. 0014-76-C-0867, and the Air Force Systems Command, U.S. Air Force under Grant No. AFOSR-78-3674.

APPENDIX A: GREEN'S FUNCTIONS

We give here the form of the Green's function $G^{\pm}(x,y;k)$ in the limits of $\delta\to 0$ and $\delta\to\infty$ for real k.

For $\delta\to 0$, taking $\zeta_+\to k$ into account, the Green's function (3.6) with (3.7) becomes

$$G(x,y;k)\to\frac{1}{2\pi\delta}\int_C \frac{e^{ip(x-y)}}{p(p-2k)}\,dp, \tag{A1}$$

which is the Green's function of the Schrödinger equation of $V(x;k) = \lim_{\delta\to 0} W(x;k)$, i.e.,

$$\delta(V_{xx} + 2ikV_x) + uV = 0 \tag{A2}$$

[note $V = \phi\exp(-ikx)$ where ϕ is in (2.7)]. On the other hand, for $\delta\to\infty$, we have

$$\hat{G}(p;k)\to\frac{1}{p-2k}\theta(p), \tag{A3a}$$

$$\hat{G}^-(p;k)\to\frac{1}{2k}\theta(p), \tag{A3b}$$

where $\theta(\cdot)$ is the usual Heaviside step function. It should be noted that $\zeta_+(k)\to 2k\theta(k)$ for real k, and the negative eigenvalue in (7.1) corresponds to what happens to $\zeta_+ < -1/(2\delta)$ (i.e., k is on the edge of the PB) for finite δ. From (A3), we have

$$G^+(x,y;k)\to\frac{1}{2\pi}\int_{C^+}\frac{e^{ip(x-y+i0)}}{p-2k}\,dp, \tag{A4a}$$

$$G^-(x,y;k)\to\frac{1}{4\pi k}\int_{-\infty}^{0} e^{-ip(x-y-i0)}\,dp, \tag{A4b}$$

where the contour C^+ is taken to be a positive half p line avoiding the singularity $p=2k$ (see Sec. 7). It is remarkable that (A4b) is just the usual projection operator $P^-(\cdot) = \frac{1}{2}(1+iH)(\cdot)$, i.e.,

$$G^-(x,y;k)\to\frac{1}{4\pi ik(x-y-i0)}$$
$$= \frac{1}{4k}\left[\delta(x-y) - \frac{i}{\pi}p\frac{1}{x-y}\right]$$
$$= \frac{1}{2k}\hat{P}^-(x-y) \tag{A5}$$

[note that $P^-(\cdot) = \int_{-\infty}^{\infty} dy\,\hat{P}^-(x-y)(\cdot)$]. Indeed, the Green's functions $G^{\pm}(x,y;k)$ decompose the scattering problem into two equations:

$$iG_x^+(x,y;k) + (\zeta_+ + 1/2\delta)G^+(x,y;k) = -\Delta^+(x-y), \tag{A6a}$$

$$(\zeta_+ + 1/2\delta)G^-(x,y;k) = \Delta^-(x-y), \tag{A6b}$$

where the functions $\Delta^{\pm}(x-y)$ satisfying $\Delta^+(x-y) + \Delta^-(x-y) = \delta(x-y)$ (i.e., $\Delta^{\pm}$ is a decomposition of δ-function) are given by

$$\Delta^+(x-y) = \frac{1}{2\pi}\int_C dp\,\frac{p-(\zeta_+ + 1/2\delta)}{p-(\zeta_+ + 1/2\delta)(1-e^{-2\delta p})}e^{ip(x-y+i0)}, \tag{A7a}$$

$$\Delta^-(x-y) = \frac{1}{2\pi}\int_C dp\,\frac{(\zeta_+ + 1/2\delta)e^{-2\delta p}}{p-(\zeta_+ + 1/2\delta)(1-e^{-2\delta p})}e^{ip(x-y-i0)}. \tag{A7b}$$

Here we note that for $\delta\to\infty$, (A7) tend to the usual projection operators,

$$\Delta^{\pm}(xy)\to\hat{P}^{\pm}(x-y) = \frac{\pm i}{2\pi(x-y\pm i0)}. \tag{A8}$$

APPENDIX B: ANALYTICITY OF THE JOST FUNCTIONS

Here we discuss the analytical property of $M^+(x;k)$ in the fundamental sheet (FS). [The analyticity of $\bar{N}^+(x;k)$ can be discussed in a similar fashion to that considered here.] In order to do this, it is convenient to write $G_1^+(x,y;k)$ in the

explicit form calculated by the residue theorem, i.e.,

$$G_1^+(x,y;k) = \frac{i}{2\delta}\sum_{n=-1}^{\infty}\frac{1}{p_n-\zeta_+}\exp[ip_n(x-y)]\theta(x-y) - \frac{i}{2\delta}\sum_{n=1}^{\infty}\frac{1}{\bar{p}_n-\zeta_+}\exp[i\bar{p}_n(x-y)]\theta(y-x). \tag{B1}$$

Here p_n $(n\geq -1)$, $\bar{p}_n$ $(n\geq 1)$ are given below (3.7).

Firstly, we show that in the upper half plane of the FS, $G_1^+(x,y;k)$ [and therefore $M^+(x;k)$] is holomorphic at the points on the branch cut corresponding to the edge of the principal branch (PB) in $\operatorname{Im} k>0$.

Let ζ_{+0} be the point on the branch cut in the upper half plane of the FS. Then there are two points k_1, k_2 (say, $\operatorname{Re} k_1 < \operatorname{Re} k_2$) on the edge of the PB, such that $\zeta_{+0}(k_1) = \zeta_{+0}(k_2)$. Calculating the two limits of $G_1^+(x,y;k)$ in the FS (or PB),

$$\lim_{\substack{\operatorname{Im}\zeta_+>\operatorname{Im}\zeta_{+0}\\ \zeta_+\to\zeta_{+0}}} G_1^+(x,y;k) = \lim_{k\to k_1} G_1^+(x,y;k) \tag{B2a}$$

and

$$\lim_{\substack{\operatorname{Im}\zeta_+>\operatorname{Im}\zeta_{+0}\\ \zeta_+\to\zeta_{+0}}} G_1^+(x,y;k) = \lim_{k\to k_2} G_1^+(x,y;k), \tag{B2b}$$

we obtain (B2a) = (B2b), since at k_1, $p_0 = 2k_1$, $p_1 = 2k_2$, at k_2, $p_0 = 2k_2$, $p_1 = 2k_1$, and the other poles p_n, $\bar{p}_n$ $(n\geq 2)$ remain the same. Hence, $G_1^+(x,y;k)$ does not have the branch cut in the upper half plane of the FS. It should be noted that for $\operatorname{Im}\zeta_+<0$, $G_1^+(x,y;k)$ has a branch cut corresponding to the lower edge of the PB. For $G_2^+(x,y;k)$ [associated with $\bar{N}(x;k)$], similar reasoning suggests that $G_2^+(x,y;k)$ does not have the branch cut in the lower half plane of the FS.

Now let us discuss the analyticity of $M^+(x;k)$ for $\operatorname{Im}\zeta_+>0$. In the case $\zeta_+\neq p_n$ $(n\geq 0)$ (i.e., there is no double zero), one can estimate, from (B1),

$$\int_{-\infty}^{\infty}|G_1^+(x,y;k)u(y)|\,dy < \frac{1}{2\delta}\left(\frac{1}{|\zeta_+|}+\frac{1}{|\zeta_-|}\right)\int_{-\infty}^{\infty}|u(y)|\,dy + \frac{2\max|u|}{\pi}\sum_{n=1}^{\infty}\frac{1}{(2n-1)|p_n-\zeta_+|} \equiv \alpha_1 < \infty. \tag{B3}$$

(i.e., the kernel can be estimated by a constant which depends on δ, $\max|u|$ and $\int_{-\infty}^{\infty}|u|\,dx$). Then, by the method of successive approximation, we have the Neumann series for $|M^+(x;k)|$ in the form

$$|M^+(x;k)| \leq 1 + \int_{-\infty}^{\infty}|G_1(x,y;k)u(y)|\,dy + \int_{-\infty}^{\infty}dy_1\int_{-\infty}^{\infty}dy_2|G_1(x,y_1;k)u(y_1)G_1(y_1,y_2;k)u(y_2) + \cdots \leq 1+\alpha_1+\alpha_1^2+\cdots. \tag{B4}$$

If $\alpha_1(k,\delta)$ is less than 1, then (B4) converges and $|M^+(x;k)|$ is uniformly bounded. For the case, $p_{n_0}=p_{n_0+1}=\zeta_+$ $(n>0)$ for a certain n_0 (i.e., there is a double zero in the upper half plane of the FS), we have

$$\int_{-\infty}^{\infty}|G_1^+(x,y;k)u(y)|\,dy \leq \frac{1}{|2n_0-1|\pi}\int_{-\infty}^{\infty}|u(y)|\,dy + \frac{2\delta\max|u|}{\pi}\left[\frac{1}{\delta}\sum_{n=0}^{\infty}{}'\frac{1}{|2n-1||p_n-p_{n_0}|} + \frac{4}{3|2n_0-1|}+\frac{6}{(2n_0-1)^2\pi}\right] \equiv \alpha_2 < \infty, \tag{B5}$$

where $\Sigma'_{n=0}^{\infty}$ expresses the sum over $n>0$ except $n=n_0$ and n_0+1. Thus, the convergence of the Neumann series is given by $\alpha_2<1$, and $|M^+(x;k)|$ is also uniformly bounded. In the case $\zeta_+=0$ $(k=0)$, we need a different way to estimate $|M^+(x;k)|$. Here the estimation of the kernel $G_1^+(x,y;k)u(y)$ depends on x. We write the integral equation for $M^+(x;k)$ in the form

$$M^+(x;k) = 1+\int_{-\infty}^{x}V(x,y;k)u(y)M^+(y;k)\,dy + \int_{-\infty}^{\infty}F(x,y;k)u(y)M^+(y;k)\,dy. \tag{B6}$$

where

$$V(x,y;k) = \frac{1}{2i\delta\zeta_+} - \frac{1}{2i\delta\zeta_-}e^{2ik(x-y)}, \tag{B7a}$$

$$F(x,y;k) = G_1^+(x,y;k) - V(x,y;k). \tag{B7b}$$

Define $\Phi(x;k)$ to be

$$\Phi(x;k) = 1+\int_{-\infty}^{\infty}F(x,y;k)u(y)M^+(y;k)\,dy, \tag{B8}$$

from which we have

$$M^+(x;k) = \Phi(x;k)+\int_{-\infty}^{x}V(x,y;k)u(y)M^+(y;k)\,dy. \tag{B9}$$

Noting that (B9) is a Volterra integral equation, we have a resolvent kernel $\Gamma(x,y;k)$ given by

$$\Gamma(x,y;k) = \sum_{n=1}^{\infty}K^{(n)}(x,y;k), \tag{B10}$$

with

$$K^{(n)}(x,y;k) = \int_y^x ds\,V(x,s;k)K^{(n-1)}(s,y;k),$$
$$K^{(1)}(x,y;k) = V(x,y;k)u(y), \tag{B11}$$

and $M^+(x;k)$ is given by

$$M^+(x;k) = \Phi(x;k)+\int_{-\infty}^{x}\Gamma(x,y;k)\Phi(y;k)\,dy. \tag{B12}$$

From (B12), (B8) can be written as a Fredholm integral equation in the form

$$\Phi(x;k) = 1+\int_{-\infty}^{\infty}K(x,y;k)\Phi(y;k)\,dy, \tag{B13}$$

whose kernel

$$K(x,y;k) = F(x,y;k)u(y) + \int_y^\infty F(x,s;k)u(s)\Gamma(s,y;k)\,ds \tag{B14}$$

[note that $F(x,y;k)$ (and therefore $K(x,y;k)$] has a logarithmic singularity at $x=y$. By the method of successive approximations for (B13), if $K(x,y;k)$ satisfies the following conditions,[10]

$$\int_{-\infty}^{\infty} |K[x,y;k)|^2\,dy = \beta_1(x) < \infty, \quad \text{for all } x, \tag{B15a}$$

$$\int_{-\infty}^{\infty} |K(x,y;k)|^2\,dx \equiv \beta_2(y) < \infty, \quad \text{for all } y, \tag{B15b}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |K(x,y;k)|^2\,dx\,dy < 1, \tag{B15c}$$

then the Neumann series of (B13) converges. In the case $k=0$, we have the estimation $|K(x,y;0)|$, for $x \neq y$,

$$|K(x,y;0)| < |F(x,y;0)u(y)| + \frac{1}{2\pi}|u(y)|\int_{-\infty}^{\infty} ds|F(x,s;0)u(s)\Gamma(s,y;0)|\,ds \tag{B16}$$

with

$$|\Gamma(x,y;0)| \leq \frac{2}{3}\left[1 + \frac{3}{2\delta}(|y|+|x|)\right] \times \exp\left[\beta + \frac{1}{\delta}x\theta(x)\int_{-\infty}^{\infty}|u(s)|ds\right], \tag{B17}$$

$$|F(x,y;0)| \leq \frac{1}{2\delta}\sum_{n=1}^{\infty}\frac{1}{|p_n|}\exp\left[-\frac{(2n-1)\pi}{2\delta}|x-y|\right] < \frac{1}{2\pi}\ln\frac{1+\exp[-(\pi/2\delta)|x-y|]}{1-\exp[-(\pi/2\delta)|x-y|]}, \tag{B18}$$

$$\beta = \frac{2}{3}\int_{-\infty}^{\infty}\left(1 + \frac{34}{2\delta}|x|\right)|u(x)|\,dx. \tag{B19}$$

From (B16), one can see that, for finite δ, if $u(x)$ decays sufficiently rapidly as $|x|\to\infty$, then $|K(x,y;0)|$ is square integrable. Moreover the Neumann series converges for given δ and $\max|u|$ chosen small enough. Namely, the solution of (B13) exists for all x, and therefore $M^+(x,0)$ exists. It is interesting to note that, for the limit $\delta\to 0$ with $u(x)/\delta$ remaining finite, $F(x,y;k)$ tends to zero and (B6) becomes the Volterra equation of the Schrödinger equation (see Appendix A). Consequently, these results imply that $M^+(x;k)$ is analytic in the upper half plane of the FS whenever the Neumann series converges in this region.

APPENDIX C: RELATIONS BETWEEN $\{a(k),b(k)\}$ and $\{\bar{a}(k),\bar{b}(k)\}$

For real k, we derive the relations (3.22) and (3.23). In order to do this, it is convenient to consider the following form of the scattering problem,

$$i\phi_x^+ + \left(u + \frac{1}{2\delta}\right)\phi^+ = \Gamma(k)\phi^-, \tag{C1}$$

where

$$\phi(x;k) \equiv W(x;k)\exp(-i\zeta_+ x),$$

and $\Gamma(k) \equiv (\zeta_+ + 1/2\delta)\exp(-2\delta\zeta_+)$ $= (-\zeta_- + 1/2\delta)\exp(2\delta\zeta_-) = \Gamma(-k)$. Then, for real k, the Jost functions are defined by

$$f(x,k) \to e^{-i\zeta_- x}, \quad \text{as } x\to-\infty, \tag{C2a}$$

$$g(x;k) \to e^{i\zeta_- x}, \quad \text{as } x\to+\infty, \tag{C2b}$$

[note that from $\Gamma(k) = \Gamma(-k)$, if $\phi(x;k)$ is a solution, then $\phi(x;-k)$ is also a solution]. Then the relations (3.20) and (3.21) become

$$f(x;k) = a(k)g(x;-k) + b(k)g(x;k), \tag{C3}$$

$$g(x;k) = \bar{a}(k)f(x;-k) + \bar{b}(k)f(x;k), \tag{C4}$$

from which we have

$$a(k)\bar{a}(-k) + b(k)\bar{b}(k) = 1, \tag{C5a}$$

$$\bar{a}(k)a(-k) + b(k)\bar{b}(k) = 1, \tag{C5b}$$

and

$$a(k)\bar{b}(-k) + b(k)\bar{a}(k) = 0, \tag{C6a}$$

$$\bar{a}(k)b(-k) + \bar{b}(k)a(k) = 0. \tag{C6b}$$

On the other hand, from (C1) and its complex conjugate equation, we have

$$i\frac{\partial}{\partial x}[\phi^+\{\phi^+\}^*] = \Gamma[\phi^-\{\phi^+\}^* - \{\phi^-\}^*\phi^+]. \tag{C7}$$

By virtue of the relation $[\phi^\pm]^* = [\phi^*]^\mp$, we obtain

$$0 = i[\phi^+\{\phi^*\}^-]\Big|_{x=-\infty}^{x=\infty} + \Gamma(k)\lim_{l\to\infty}\left(\int_{-l-i\delta}^{-l+i\delta}dx + \int_{l-i\delta}^{l+i\delta}dx\right)|\phi|^2. \tag{C8}$$

Using $f(x;k)$ as $\phi(x;k)$ in (C8), and taking the boundary condition of $f(x;k)$ into account, we obtain (3.23). Also comparing (3.23) with (C5) and (C6), we have (3.22).

APPENDIX D: BOUND STATES

Here we show that the zeroes of $a(k)$ are pure imaginary and simple.

From (2.1), we have

$$i\frac{\partial}{\partial x}[\psi_1^+\{\psi_1^+\}^*] - (\lambda_1 - \lambda_1^*)[\psi_1^+\{\psi_1^+\}^*] = \mu_1\psi_1^-\{\psi_1^+\}^* - \mu_1^*\psi_1^+\{\psi_1^-\}^*, \tag{D1}$$

where $\psi_1 \equiv (x;k_1)$, $\lambda_1 = \lambda(k_1)$, $\mu_1 = \mu(k_1)$ and $\operatorname{Im} k_1 > 0$. By virtue of that $[\psi^\pm]^* = [\psi^*]^\mp$, we have

$$(\lambda_1 - \lambda_1^*)\int_{-\infty}^{\infty}|\psi_1^+|^2\,dx + (\mu_1 - \mu_1^*)\int_{-\infty}^{\infty}|\psi_1|^2\,dx = 0. \tag{D2}$$

From (D2), using the formula of λ and μ in terms of the k [see below (2.2)], we obtain $k_1 = i\kappa_1$.

In order to prove $a(k_1)\neq 0$ (i.e., k_1 is a simple), we show the following relation, for the bound state $k = i\kappa$,

$$C^{-1} \equiv i\dot{a}/b = \int_{-\infty}^{\infty}|g(x;i\kappa)|^2\,dx, \tag{D3}$$

where $g(x;i\kappa)$ is given by (C2b), and C is defined in (4.8). From

(C1) and its derivative with respect to ζ_+, i.e.,

$$i\dot{\phi}^+_x + (u + 1/2\delta)\dot{\phi}^+ = \Gamma\dot{\phi}^- + \dot{\Gamma}\phi^- \tag{D4}$$

[here the dot denotes the derivative with respect to ζ_+, e.g., $\dot{\Gamma} = -2\delta\zeta_+ \exp(-2\delta\zeta_+)$], we have, for $k = i\kappa$,

$$i\frac{\partial}{\partial x}[\dot{\phi}^+\{\phi^+\}^*] = \Gamma[\dot{\phi}^-\{\phi^+\}^* - \dot{\phi}^+\{\phi^-\}^*] - 2\delta\zeta_+ e^{-2\delta\zeta_+}\phi^-\{\phi^+\}^*, \tag{D5}$$

where $\dot{\phi} = \partial\phi/\partial\zeta_+|_{k=i\kappa}$, and we have used $\Gamma^*(i\kappa) = \Gamma(i\kappa)$. Integrating (D5) over $-\infty$ to ∞, we have

$$i\dot{\phi}^+\{\phi^+\}^*|^{x=\infty}_{x=-\infty} = \Gamma\lim_{l\to\infty}\left(\int_{-l+i\delta}^{-l-i\delta} dx + \int_{l-i\delta}^{l+i\delta} dx\right)\dot{\phi}\phi^* - 2\delta_+ e^{-2\delta\zeta_+}\lim_{l\to\infty}\left(\int_{-l+i\delta}^{-l} dx + \int_{-l}^{l} dx + \int_{l}^{l+i\delta} dx\right)|\phi|^2, \tag{D6}$$

where we have used $[\phi^\pm]^* = [\phi^*]^\mp$. From (D6) with $\phi = f$ and the boundary condition of f, we have

$$i\dot{a}b^* = \int_{-\infty}^{\infty} |f(x;i\kappa)|^2\,dx. \tag{D7}$$

Consequently, from $f(x;i\kappa) = bg(x;i\kappa)$, we obtain (D3).

APPENDIX E: CLOSURE OF THE JOST FUNCTIONS

Under the condition of the analyticity for the Jost functions considered in Secs. 4–6, we show that the set of the Jost functions defined in (3.11) or (3.12) consists of the closure for the eigenvalue problem (3.2). We use the scattering problem (C1) for convenience.

Let us define the adjoint problem of (C1) in the form

$$-i\{\phi^A\}^-_x + (u + 1/2\delta)\{\phi^A\}^- = \Gamma(k)\{\phi^A\}^+, \tag{E1}$$

where $\phi^A = \phi^A(x;k)$ is the adjoint solution of $\phi(x;k)$. For real k, the Jost functions of (E1) are given by

$$f^A(x;k) \to e^{-i\zeta_- x}, \quad \text{as } x\to-\infty, \tag{E2a}$$

$$g^A(x;k) \to e^{i\zeta_- x}, \quad \text{as } x\to+\infty. \tag{E2b}$$

From (C2) and (E2), one can see that

$$\phi^A(x;k) = \phi^*(x,-k) \quad \text{for real } k. \tag{E3}$$

By virtue of the analytical continuation, we have

$$\phi^A(x;k) = \phi^*(x,-k^*) \quad \text{for complex } k. \tag{E4}$$

Let $\langle k|k'\rangle$ be an inner product defined by

$$\langle k|k'\rangle \equiv \int_{-\infty}^{\infty} \phi(x;k)\phi^A(x;k')\,dx \equiv \int_{-\infty}^{\infty} \langle k|x\rangle\langle x|k'\rangle\,dx, \tag{E5}$$

where the Dirac symbols are defined by $\langle k|x\rangle \equiv \phi(x;k)$ and $\langle x|k\rangle \equiv [\langle k|x\rangle]^A = \phi^A(x;k)$. Then for real k, from (C1) and (E1), we have

$$[\Gamma(k) - \Gamma(k')]\langle k|k'\rangle = i[\phi^+(k)\{\phi^A(k')\}^-]|^{x=\infty}_{x=-\infty} - \Gamma(k)\lim_{l\to\infty}\left(\int_{-l+i\delta}^{-l} dx + \int_{l}^{l+i\delta} dx\right)\phi(k)\phi^A(k') + \Gamma(k')\lim_{l\to\infty}\left(\int_{-l-i\delta}^{-l} dx + \int_{l}^{l-i\delta} dx\right)\phi(k)\phi^A(k'). \tag{E6}$$

Defining

$$f(x,k) \equiv \langle k,1|x\rangle\}, \tag{E7a}$$

$$g(x,k) \equiv \langle k,2|x\rangle\}, \tag{E7b}$$

and by virtue of the boundary conditions of f and g, we have the following orthogonality relations,

$$\langle k,1|k',2\rangle = 2\pi a(k)\delta(\zeta_+ - \zeta'_+) \quad \text{for real } k,k', \tag{E8a}$$

$$\langle k,1|i\kappa_l,2\rangle = \langle i\kappa_l,1|k,2\rangle = 0 \quad \text{for real } k, \tag{E8b}$$

$$\langle i\kappa_l,1|i\kappa_m,2\rangle = i\dot{a}_l\delta_{lm}. \tag{E8c}$$

If the set of the functions $[\{f(x;k), -\infty < k < \infty\}, \{f(x;i\kappa_l)\}_{l=1}^N]$ is complete in the sense of $L_2(-\infty < x < \infty)$, then an arbitrary function $h(x)$ in L_2 can be expanded in the following form (that is, the expansion theorem),

$$h(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\zeta_+ \hat{h}(k)f(x;k) + \sum_{l=1}^{N} h_l f(x;i\kappa_l), \tag{E9}$$

where $\hat{h}(k)$ and h_l are determined by using (E8), i.e.,

$$\hat{h}(k) = \frac{1}{a(k)}\int_{-\infty}^{\infty} h(x)g^A(x;k)\,dx, \tag{E10a}$$

$$h_l = \frac{1}{i\dot{a}_l}\int_{-\infty}^{\infty} h(x)g^A(x;i\kappa_l)\,dx. \tag{E10b}$$

In order to verify the expansion theorem (E9) it is sufficient to show that, for this case, there is an identity operator given in the form[12]

$$I = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{d\zeta_+}{a(k)}|k,2\rangle\langle k,1| + \sum_{l=1}^{N}\frac{1}{i\dot{a}_l}|i\kappa_l,2\rangle\langle i\kappa_l,1|, \tag{E11}$$

or

$$\begin{aligned}\langle y|I|x\rangle &= \delta(x-y)\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{d\zeta_+}{a(k)}g^A(y;k)f(x;k)\\ &\quad + \sum_{l=1}^{N}\frac{1}{i\dot{a}_l}g^A(y;i\kappa_l)f(x;i\kappa_l)\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} g^A(y;k)g(x;-k)\,d\zeta_+\\ &\quad + \frac{1}{2\pi}\int_{-1/2\delta}^{\infty}\frac{b(k)}{a(k)}g^A(y;k)g(x;k)\,d\zeta_+\\ &\quad + \sum_{l=1}^{N} C_l g^A(y;i\kappa_l)g(x;i\kappa_l).\end{aligned} \tag{E12}$$

By using the triangular representation of

$$N(x;k) = g(x;k)\exp(i\zeta_+ x),$$

$$g(x;k) = e^{i\zeta_- x} + \int_x^{\infty} K(x,s)e^{i\zeta_- s}\,ds, \tag{E13}$$

(E12) becomes

$$\langle y|Ix\rangle = \delta(x-y) + K(x,y) + F(x,y) + \int_x^\infty K(x,s)F(s,y)\,ds + \int_y^\infty K^*(y,s)\Big[K(x,s) + F(x,s) + \int_x^\infty K(x,t)F(t,s)dt\Big]ds. \tag{E14}$$

By virtue of the Gel'fand–Levitan equation (4.9), we have

$$\langle y|I|x\rangle = \delta(x-y). \tag{E15}$$

[1] R. I. Joseph, J. Phys. A. 101, L225 (1977).
[2] T. Kubota, D. R. S. Ko, and D. Dobbs, J. Hydronaut. 12, 157 (1978).
[3] R. I. Joseph and R. Egri, J. Phys. A 11, L97 (1978).
[4] H. H. Chen and Y. C. Lee, Phys. Rev. Lett. 43, 264 (1979).
[5] J. Satsuma and M. J. Ablowitz, in *Nonlinear Partial Differential Equations in Engineering and Applied Science*, edited by R. L. Sternberg, A. J. Kalinowski, and J. S. Papadakis (Marcel Dekker, New York, 1980), p. 397.
[6] J. Satsuma, M. J. Ablowitz, and Y. Kodama, Phys. Lett. A 73, 283 (1979).
[7] Y. Kodama, J. Satsuma, and M. J. Ablowitz, Phys. Rev. Lett. 46, 687 (1981).
[8] V. E. Zakharov and A. B. Shabat, Funct. Anal. Appl. 8, 226 (1974).
[9] C. S. Gardner, J. M. Green, M. D. Kruskal, and R. M. Miura, Commun. Pure Appl. Math. 23, 97 (1974).
[10] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), p. 152.
[11] J. D. Meiss and N. R. Pereira, Phys. Fluids 21, 700 (1978).
[12] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. LIII, 249 (1974).

J. Phys. A: Math. Gen. 15 (1982) 781–786. Printed in Great Britain

On the periodic intermediate long wave equation

M J Ablowitz†, A S Fokas†, J Satsuma†§ and Harvey Segur‡

†Clarkson College of Technology, Department of Mathematics and Computer Science, Potsdam, New York 13676, USA

‡Aeronautical Research Associates of Princeton, Inc., 50 Washington Road, PO Box 2229, Princeton, New Jersey 08540, USA

Received 16 June 1981

Abstract. We discuss some properties of a certain physically interesting nonlinear integro-differential equation with periodic boundary conditions. It is the natural periodic analogue of the intermediate long wave equation, and it provides a periodic analogue of the Benjamin–Ono equation in the appropriate limit. Due to the speciality of the integral operator, the equation admits a Bäcklund transformation, an infinity of motion constants, etc. Two simple periodic solutions are exhibited. Finally we note that the equation may be transformed into more than one kind of bilinear equation.

The so-called intermediate long wave (ILW) equation (Joseph 1977, Kubota *et al* 1978) can be written in the form ($\delta > 0$)

$$u_t + \delta^{-1} u_x + 2uu_x + (Tu)_{xx} = 0 \tag{1}$$

on $-\infty < x < \infty$, where

$$Tu = -\frac{1}{2\delta} \int_{-\infty}^{\infty} \coth \frac{\pi}{2\delta} (x - \xi) u(\xi)\, d\xi \tag{2}$$

and the integral is evaluated in the principal-value sense. The equation can be solved on $(-\infty, \infty)$ via an inverse scattering transform (IST) (Kodama *et al* 1981), and has the analytical structure associated with such equations (Joseph and Egri 1978, Satsuma *et al* 1979).

The physical derivation of (1) and (2) as a model of the evolution of long internal waves of moderate amplitude assumes that $u(x)$ has a classical Fourier transform, and that u vanishes as $|x| \to \infty$. Even so, one may ask whether (1), (2) admit spatially periodic solutions. This was done by Joseph and Egri (1978), Chen and Lee (1979), and Nakamura and Matsuno (1980), using formal algebraic methods. Unfortunately, the solutions so obtained either contain errors or are subject to a limitation that was obscured by these formal methods.

Alternatively, one may seek an evolution equation for long internal waves of moderate amplitude that are spatially periodic. Then a derivation similar to the usual one leads to (1), but in the periodic case (2) is replaced with

$$Tu = \frac{1}{2L} \int_{-L}^{L} \hat{T}(x - \xi; \delta, L) u(\xi)\, d\xi \tag{3a}$$

§ Present address: General Education, Miyazaki Medical College, Kiyotaki, Miyazaki 889-16, Japan.

0305-4470/82/030781+06$02.00 © 1982 The Institute of Physics

where

$$\hat{T}(x;\delta,L)=-\frac{2K}{\pi}\left[Z\left(\frac{Kx}{L}\right)+dn\left(\frac{Kx}{L}\right)cs\left(\frac{Kx}{L}\right)\right]. \tag{3b}$$

(Actually, the physical derivation naturally leads to the Fourier representation of $\hat{T}$, given in (3c), which is then transformed into (3b).) This derivation also requires that $\int_{-L}^{L} u\,dx=0$, which we always may impose on (1) with (3), because any non-zero mean may be removed by a Galilean transformation: $u'=u+\alpha$, $x'=x+2\alpha t$. In (3b), K denotes the complete elliptic integral of the first kind, $Z(a)$ is Jacobi's zeta function, and $dn(a)$, $cs(a)$ are Jacobian elliptic functions. These all have modulus k, determined by the condition that $K'(k)/K(k)=\delta/L$, where $K'(k)$ is the associated elliptic integral of the first kind. (All of these functions are discussed by Byrd and Friedman (1971).) The purpose of this note is to discuss some of the mathematical structure of (1) with (3).

An alternative, but very useful, representation of $\hat{T}$ is its Fourier series,

$$\hat{T}(x;\delta,L)=\mathrm{i}\sum_{n\neq 0}\coth\left(\frac{n\pi\delta}{L}\right)\exp\left(\frac{\mathrm{i}n\pi x}{L}\right) \tag{3c}$$

so that

$$Tu=\mathrm{i}\sum_{n\neq 0}\coth\left(\frac{n\pi\delta}{L}\right)\hat{u}_n\exp\left(\frac{\mathrm{i}n\pi x}{L}\right) \tag{3d}$$

where $\{\hat{u}_n\}$ are the Fourier coefficients of u. This representation follows from the identities (cf Byrd and Freidman 1971, §§ 905.01 and 908.51)

$$Z(a)=\frac{2\pi}{K}\sum_{m=1}^{\infty}\frac{q^m}{1-q^{2m}}\sin\left(\frac{m\pi a}{K}\right)$$

$$cs(a)\,dn(a)=\frac{\pi}{2K}\cot\frac{\pi a}{2K}-\frac{2\pi}{K}\sum_{m=1}^{\infty}\frac{q^m}{1+q^m}\sin\left(\frac{m\pi a}{K}\right)$$

where $q=\exp(-\pi K'/K)=\exp(-\pi\delta/L)$, and from the formal representation (Gel'fand and Shilov 1964, p 32)

$$\cot\tfrac{1}{2}\theta=2\sum_{m=1}^{\infty}\sin m\theta.$$

The usual operator on $(-\infty,\infty)$ may be recovered simply by replacing the sum in (3d) with an integral and rescaling. Similarly, we may recover (2) from (3) by letting $L\to\infty$, δ fixed. Then $k\to 1$, $K\to\pi L/2\delta$, and one may show that

$$\frac{1}{2L}\hat{T}(x;\delta,L\to\infty)\to\frac{1}{2\delta}\coth\frac{\pi x}{2\delta} \tag{4a}$$

which reproduces (2). On the other hand, for $\delta\to\infty$, L fixed, we have $k\to 0$, $K\to\frac{1}{2}\pi$, and

$$\hat{T}(x;\delta\to\infty,L)\to-\cot\pi x/2L. \tag{4b}$$

This is the well known Hilbert kernel on $(-L,L)$. With this kernel, (1) is the natural periodic extension of the Benjamin–Ono equation (Benjamin 1967, Ono 1975). As one would expect, (1) with (3) reduces to the (periodic) KdV equation if $\delta\to 0$, L fixed.

If $u(x)$ is periodic with period $2L$ and with zero mean, then (Tu) according to (2) and to (3) are identical. That this is so may be seen by rewriting (2) in its Fourier transform representation, and recalling that the FT of a periodic function is a sum of Dirac delta

functions (Gel'fand and Shilov 1964). Thus (1) with (3) may be regarded simply as the most natural way to write the ILW equation when periodic solutions are of interest.

The operator T given in (3) is the most general periodic operator we have found which satisfies the 'T conditions':

$$T(uTv + vTu) = TuTv - uv \tag{T1}$$

$$\int_{-\infty}^{\infty} (uTv + vTu)\,dx = 0 \tag{T2}$$

where u, v have zero mean. We find that ($T1$), ($T2$) are necessary conditions on the T operator in order for (1) to have more than the standard number of conserved quantities. With these conditions the above evolution equations can be expected to be in the IST class. Condition ($T2$) follows from the fact that $T(x)$ is an odd function. To establish ($T1$) we use the representation (3c). Calling $\hat{T}_n = \mathrm{i}\coth n\pi\delta/L$, and $\hat{u}_n$, $\hat{v}_n$, the Fourier coefficients of u, v respectively, assuming $\hat{u}_0 = \hat{v}_0 = 0$ (i.e., zero mean) and using the convolution theorem, then

$$\begin{aligned} T(uTv) &= \sum_{n=-\infty}^{\infty} \hat{T}_n \sum_{m=-\infty}^{\infty} \hat{T}_m \hat{v}_{n-m}\hat{u}_{n-m} \exp\left(\frac{\mathrm{i}n\pi x}{L}\right) \\ &= \sum_m \hat{u}_m \sum_n \hat{T}_n \hat{T}_{n-m}\hat{v}_{n-m} \exp\left(\frac{\mathrm{i}n\pi x}{L}\right) \\ &= -\sum_m \hat{u}_m \sum_n \hat{v}_{n-m} \exp\left(\frac{\mathrm{i}n\pi x}{L}\right) - \sum_m \hat{u}_m \sum_n \hat{T}_m \hat{T}_n \hat{v}_{n-m} \exp\left(\frac{\mathrm{i}n\pi x}{L}\right) \\ &\quad + \sum_m \hat{u}_m \sum_n \hat{T}_m \hat{T}_{n-m}\hat{v}_{n-m} \exp\left(\frac{\mathrm{i}n\pi x}{L}\right) \\ &= -u(x)v(x) - T(vTu) + (Tu)(Tv) \end{aligned} \tag{5}$$

where we have used the identity

$$\coth A \coth B = 1 + \coth(A - B)(-\coth A + \coth B).$$

The order of summation in (5) may be interchanged if $\Sigma|\hat{u}_n|$, $\Sigma|\hat{v}_n|$ exist. This establishes ($T1$).

The need for the T conditions can be seen from the following. The usual ILW equation on $(-\infty, \infty)$ was considered by Satsuma *et al* (1979). They showed that the derivations of the constants of motion and of the Bäcklund transformation do not depend on the specific kernel of T, so long as T satisfies conditions ($T1$) and ($T2$). This implies that these formulae will remain valid for equation (1) with (3) since the T conditions are satisfied. Actually one easily verifies (by differentiation) that the constants of motion given by Satsuma *et al* (1979) are also constants of motion of (1) with (3).

One may use the generalised Miura transformation, $u = \frac{1}{2}(\hat{\lambda} - V_x + \mathrm{i}TV_x + \hat{\mu}\,\mathrm{e}^{\mathrm{i}V})$ to derive a generalisation of the so-called modified ILW equation. Namely if $\tilde{K}[u] = 0$ represents (1), and $\tilde{M}[v] = V_t + (1/\delta + \hat{\lambda})V_x + (TV)_{xx} + \hat{\mu}V_x\,\mathrm{e}^{\mathrm{i}V} + \mathrm{i}V_x(TV)_x = 0$, then using the T conditions we have

$$\tilde{K}[u] = \theta\tilde{M}[v] \tag{6}$$

where

$$\theta = (-\partial_x + \mathrm{i}T\partial_x + \mathrm{i}\hat{\mu}\,\mathrm{e}^{\mathrm{i}V}).$$

Similarly it can be proven, by using the results of Fokas and Fuchssteiner (1980) (which also apply to the above Miura-type transformation) that (1) admits an auto-Bäcklund transformation if and only if $(T1)$, $(T2)$ are satisfied.

The IST pair used by Kodama *et al* (1981) is also valid in the period problem. However, the simple compatibility argument does not indicate the importance of the T conditions. Namely, consider a linear scattering problem and a sequence of associated time evolutions of the form

$$i\psi_x^+ + (u-\lambda)\psi^+ = \mu\psi^- \tag{7a}$$

$$\psi_t^* = Q_n^*\psi^* \qquad n = 1, 2, \ldots \tag{7b}$$

$(7a)$ is to be thought of as a differential Riemann–Hilbert problem to find ψ^* (i.e., ψ^* are the boundary values of certain analytic functions) given an appropriate function $u(x)$. For each n, compatibility of $(7a)$ and $(7b)$ yields an evolution equation; namely, requiring $\psi_{xt}^* = \psi_{tx}^*$; and setting all coefficients of ψ^+, ψ^-, ψ_x^-, ψ_{xx}^-, etc, to zero after using $(7a)$ to eliminate derivatives of ψ^+, gives an algorithmic procedure to determine compatible equations. The equivalent to the first two equations of the KdV hierarchy are obtained as follows. First, $Q_1^* = \partial_x + A^*$, whereupon we find $A^+ - A^- = 0$. Taking $A^* = A_0 = \text{constant}$, the compatible evolution equation is $u_t = u_x$. Second, $Q_2^* = i\partial_x^2 + iB^*\partial_x + iA^*$, we find $B^+ - B^- = 0$, $A^+ - A^- = -2iu_x$. Taking $B^* = B_0 = i(2\lambda + 1/\delta) = \text{constant}$, $A^* = \pm iu_x - (Tu)_x$, the compatible evolution equation is (1), without need for the T conditions. The underlying reason why such conditions must be added, and whether, in fact, (3) is the most general singular integral operator satisfying $(T1)$, $(T2)$ are open questions. Matsuno (1980) has given a different algorithm to derive such a hierarchy of equations.

Next we consider some special solutions of (1) with (3). For a wave of permanent form, $\partial_t \to -c\partial_x$, and (1) may be integrated once to

$$(\delta^{-1} - c)u + u^2 + (Tu)_x + A = 0 \tag{8}$$

with A constant. It has been claimed (Joseph and Egri 1978, Chen and Lee 1979) that

$$u(x-ct) = \frac{-(n\pi/L)\sinh(n\pi\delta/L)}{\cosh(n\pi\delta/L) + \cos[n\pi(x-ct)/L]} \tag{9a}$$

is a solution of (8). Its Fourier series representation is

$$u(\xi) = -\frac{n\pi}{L}\left[1 + 2\sum_1^\infty (-1)^m \exp\left(-\frac{mn\pi\delta}{L}\right)\cos\left(\frac{mn\pi\xi}{L}\right)\right]. \tag{9b}$$

That (9) does not solve (8) may be seen by computing $\partial u/\partial\delta$ from $(9a)$ and $(9b)$, and Tu from $(3d)$, because, as the reader may verify by direct calculation: $\partial u/\partial\delta = -\partial(Tu)/\partial x$. The correct solution is

$$u(x-ct) = \frac{-(n\pi/L)\sinh(n\pi\delta/L)}{\cosh(n\pi\delta/L) + \cos[(n\pi/L)(x-ct-i\phi)]} \tag{10a}$$

where $\phi < -\delta$. Its Fourier series is

$$u(\xi) = \frac{2n\pi}{L}\sum_1^\infty (-1)^m \sinh\left(\frac{mn\pi\delta}{L}\right)\exp\left(\frac{imn\pi(\xi - i\phi)}{L}\right) \tag{10b}$$

and its validity may be verified by computing $\partial u/\partial\delta$ from $(10a)$ and $(10b)$. This solution

is complex valued. (The Fourier representations in (9*b*) and (10*b*) differ because the transformation $x \to x - i\phi$, $\phi < \delta$, moves ξ outside the radius of convergence of (9*b*).)

More generally, periodic solutions may be obtained systematically via Hirota's method, revising slightly the methods of Chen and Lee (1979), Nakamura and Matsuno (1980) and Satsuma and Ablowitz (1980). Let $f^{\pm}(x) = f(x \mp i\delta)$, let $f(x)$ be periodic $(2L)$ and let $f(z)$ be analytic in the rectangle: $-L \leq \mathrm{Re}(z) \leq L$, $-\delta \leq \mathrm{Im}(z) \leq \delta$. Then by integrating $\int \tilde{T}(x - \zeta + i\delta) f(\zeta) d\zeta$ around this rectangle, one finds that

$$\frac{1}{2L} \oint_{-L}^{L} \tilde{T}(x - z; \delta, L)[f^{+}(z) - f^{-}(z)]\, dz = i[f^{+}(x) + f^{-}(x)] + J_0 \tag{11}$$

where J_0 is an unimportant constant. It follows that if

$$u(x) = i[f^{-}(x) - f^{-}(x)] \tag{12}$$

where

$$f^{\pm}(x) = \partial_x(\ln F^{\pm}),$$

and if $f(x)$ is properly analytic, then (1) with (3) becomes a bilinear equation

$$(iD_t + i\delta^{-1}D_x - D_x^2 + A)F^{+} \cdot F^{-} = 0 \tag{13}$$

with the usual notation (e.g. $D_x\ a.b \equiv (\partial_x - \partial_{x'})a(x)b(x')|_{x'=x}$), and $A = A(t)$. We emphasise that only those solutions of (13) that are analytic in the rectangle yield periodic solutions of (1), a fact that was overlooked previously.

The simplest real-valued solution of (1) with (3) was given by Nakamura and Matsuno (1980). It may be written in the form

$$u(\xi; m) = \frac{iK(m)}{L}\left[Z\left(\frac{K(m)}{L}(\xi - i\delta); m\right) - Z\left(\frac{K(m)}{L}(\xi + i\delta); m\right)\right] \tag{14a}$$

where $\xi = x - ct + x_0$, and m is the modulus. The analyticity condition is that

$$\delta/L < K'(m)/K(m) \tag{14b}$$

or, because $\delta/L = K'(k)/K(k)$, that the modulus of T exceeds the modulus of the solution. The wave speed is a complicated function of m, δ and L, and is given implicitly by Nakamura and Matsuno (1980).

Finally, we note that (13) is not the only 'bilinear' equation that may be obtained from (1), (3). For example, let $f(z)$ have a real period $(2L)$, and be analytic in $-L \leq \mathrm{Re}(z) \leq L$, $-\delta \leq \mathrm{Im}(z) \leq \delta$ except for two poles at $z = z_0$ and z_0^* $(0 < \mathrm{Im}(z_0) < \delta)$ with residues b and b^*. Then the integral of $\int \tilde{T}(x - \zeta + i\delta) f(\zeta) d\zeta$ around the rectangle yields

$$\frac{1}{2L} \oint_{-L}^{L} \tilde{T}(x - z)[f^{+}(z) - f^{-}(z)]\, dz = i[f^{+}(x) + f^{-}(x)] + J_0$$

$$- \frac{\pi i}{L}[b\tilde{T}(x - z_0 + i\delta) + b^*\tilde{T}(x - z_0^* + i\delta)] \tag{15}$$

instead of (11). In this case, (12) changes (1) with (3) into

$$\{iD_t + (i/\delta)D_x - D_x^2 + A + (\pi/L)[b\partial_x\tilde{T}(x - z_0 + i\delta) + b^*\partial_x\tilde{T}(x - z_0^* + i\delta)]\}F^{+} \cdot F^{-} = 0. \tag{16}$$

The analyticity requirement is that $F(z)$ should be analytic in the usual rectangle except for simple branch points at z_0 and z_0^*, so that f^* has poles. We have not determined whether (15) yields any solutions of (1), (3) that are not available via (13).

This work was partially supported by the Air Force Office of Scientific Research, USAF and by the Office of Naval Research, Mathematics and Fluid Dynamics Programs.

References

Benjamin T B 1967 *J. Fluid Mech.* **29** 559
Byrd P F and Friedman M D 1971 *Handbook of Elliptic Integrals for Scientists and Engineers* (New York: Springer)
Chen H H and Lee Y C 1979 *Phys. Rev. Lett.* **43** 264
Fokas A S and Fuchssteiner B 1980 *Bäcklund Transformations for Hereditary Symmetries, Nonlinear Analysis, TMA*
Gel'fand I M and Shilov G 1964 *Generalized Functions, Properties and Operations* vol 1 (New York: Academic)
Joseph R I 1977 *J. Phys. A: Math. Gen.* **10** L225
Joseph R I and Egri R 1978 *J. Phys. A: Math. Gen.* **11** L97
Kodama Y, Satsuma J and Ablowitz M J 1981 *Phys. Rev. Lett.* **46** 687
Kubota T, Ko D R S and Dobbs D 1978 *J. Hydronaut.* **12** 157
Matsuno Y 1980 *J. Phys. Soc. Japan* **48** 1024
Nakamura A and Matsuno Y 1980 *J. Phys. Soc. Japan* **48** 653
Ono H 1975 *J. Phys. Soc. Japan* **39** 1082
Satsuma J and Ablowitz M J 1980 in *Nonlinear Partial Differential Equations in Engineering and Applied Science* ed R L Sternberg, A J Kalinowki and J S Papadakis (New York: Marcel Dekker) p 397
Satsuma J, Ablowitz M J and Kodama Y 1979 *Phys. Lett.* **73A** 283

ON A UNIFIED APPROACH TO TRANSFORMATIONS AND ELEMENTARY SOLUTIONS OF PAINLEVÉ EQUATIONS

by

A. S. Fokas and M. J. Ablowitz
Department of Mathematics and Computer Science
Clarkson College of Technology
Potsdam, New York 13676 U.S.A

Abstract

An algorithmic method is developed for investigating the transformation properties of second order equations of Painlevé type. This method, which utilizes the singularity structure of these equations, yields explicit transformations which relate solutions of the Painlevé equations II-VI, with different parameters. These transformations easily generate rational and other elementary solutions of the equations. The relationship between Painlevé equations and certain new equations quadratic in the second derivative of Painlevé type is also discussed.

§1. Introduction

We say that an equation is of Painlevé type if all its solutions possess the Painlevé property, i.e. their only singularities are poles or nonmovable critical points [1]. The most well known second order equations of Painlevé type are the so called six Painlevé equations, PI-PVI [1], discovered by Painlevé [2] and his school [3] at the turn of the century. They classified all equations of the form $w'' = F(w',w,z)$ where F is rational in w', algebraic in w, and locally analytic in z, which have the Painlevé property. They found that, within a Möbious

[4] transformation, there exist fifty such equations. Distinguished amongst these fifty equations are PI-PVI. Any other of the fifty equations can either be integrated in terms of known functions or can be reduced to one of these six equations. Although PI-PVI where first discovered from strictly mathematical considerations, they have recently appeared in several physical applications (see for example [5], [6], [7]).

Explicit transformations and relevant exact solutions admitted by the Painlevé equations first appeared in the Soviet literature [8] and are summarized in [9]; the main points are as follows:

i) For certain choices of the parameters, PII-V admit one-parameter families of solutions expressible in terms of the classical transcendental functions Airy [3], [10], Bessel [11], Weber-Hermite [12] and Whittaker [13] respectively.

ii) PII-V admit transformations (see [14] to [17]) which map solutions of a given Painlevé equation to solutions of the same equation but with different values of the parameters.

iii) Using ii) one can construct (for certain choices of the parameters) various elementary solutions of PII-V. These solutions are either rational or are functions which are related (through repeated differentiations and multiplications) to the above mentioned classical transcendental functions.

iv) For PVI it was only known that, for a certain single choice of its parameters, it admits a one-parameter family of solutions expressible in terms of hypergeometric functions [18].

However, the above results apparently were obtained by rather adhoc methods. Moreover, it spite of the extensive amount of research on Painlevé equations, the transformation properties of PVI were not found. Also, no other one-parameter family of solutions of PVI was found, save for the one mentioned above. It is

important to note that PVI is in a sense the most general Painlevé equation since it contains the other five as limiting cases [1].

In this paper:

1) We develop an algorithmic method for systematically investigating the transformation properties of second order equations of the Painlevé type. This method yields explicit transformations: α) Between a given Painlevé equation and the same Painlevé equation but with different values of its parameters. β) Between two different Painlevé equations (for example PIII and PV). γ) Between a Painlevé equation of the type investigated by Painlevé (i.e. linear in the second derivative) and an equation of the Painlevé type which is quadratic in the second derivative.

2) As an application of this method we rederive the known transformation properties of PII-V and also derive the transformation properties of PVI. The latter are used to obtain (for various choices of parameters) one-parameter families of solutions of PVI. Amongst these solutions are rational solutions as well as solutions which are related (through repeated differentiations and multiplications) to hypergeometric functions.

3) We relate PIII and PVI with certain new equations quadratic in the second derivative and of the Painlevé type.

Some of the results concerning PVI have been announced in [19].

The connection with inverse scattering and monodromy preserving transformations.

In recent years considerable interest has developed in Painlevé equations. Ablowitz (Ramani) and Segur [20], ([21]) have discovered a deep connection between equations of Painlevé type and the PDE's solvable by the inverse scattering transform [22]. For example PII and special cases of PIII and PIV may be obtained from the similarity reduction of the modified Korteweg-deVries, the sine-Gordon and the

nonlinear Schrödinger equations respectively. It is interesting that proper reductions of the Korteweg-deVries (KdV) equation lead to both PI and PII:

i) KdV and PII

Consider the KdV equation in the form

$$u_t + 6uu_x + u_{xxx} = 0. \tag{1.1}$$

Equation (1.1) is clearly invariant under the group of transformations $x' = \lambda x$, $t' = \lambda^3 t$, $u' = \lambda^{-2} u$, where λ is some arbitrary parameter. The solutions of (1.1) invariant under this group of transformations (the so called similarity solutions) are characterized by $u = (3t)^{-2/3} U(z)$, $z = x(3t)^{-1/3}$, where $U(z)$ satisfies

$$K_1(U) = U''' + 6UU' - (2U + zU') = 0. \tag{1.2}$$

Whitham [23] has noted that equation (1.2) is related to PII. Actually there is a one to one correspondence between the integrated form of (1.2) and PII. Equation (1.2) can be integrated once using the following indentity

$$[(2U-z)K_2(U)]' = (2U-z)K_1(U), \tag{1.3}$$

where

$$K_1(U) = U'' + 2U^2 - zU + \frac{\nu+U'-U'^2}{2U-z}. \tag{1.4}$$

Equation (1.4) is essentially equation XXIV of [1] and is related by a one to one map with PII (see §3).

ii) KdV and PI

Equation (1.1) is also invariant under the group of transformations

$$x' = x + 6t\lambda, \quad t' = t + \frac{\lambda}{\alpha}, \quad u' = u + \lambda, \tag{1.5}$$

where λ and α are arbitrary parameters. Regarding α fixed, one immediately [24] obtains the following characterization for the solutions of (1.1) invariant under (1.5):

$$u = \alpha t + U(z), \quad z = x - 3\alpha t^2, \tag{1.6}$$

where

$$U''' + 6UU' + \alpha = 0. \tag{1.7}$$

Equation (1.7), upon integration yields PI.

Using the ideas of Ablowitz and Segur one can characterize a non-elementary one parameter family of solutions of, say, (1.2) through a Gel'fand-Levitan linear integral equation of the Fredholm type. Recently we have proposed a new method [25] for linearizing the Painlevé equations, using singular integral equations and Riemann-Hilbert theory. In this way we have characterized a three parameter family of solutions of (1.2). This work suggests that the transformations given in this paper may be useful in obtaining the general solution of, say, PII using our method. Important is the fact that using these transformations one can find the general solution of a given Painlevé equation for arbitrary values of its parameters $\underline{\alpha}$, provided one knows the general solution of this equation for a range of $\underline{\alpha}$ only. For example, for PII one needs to know the general solution only for $-1/2 < \alpha < 1/2$.

We also note that there is a close connection between ODE's of Painlevé type and mondoromy preserving deformations. This was emphasized and used by Flaschka and Newell [26] and by Sato, Miwa, Jimbo and their coworkers [27]. In particular, Flaschka and Newell derived a formal system of linear singular integral equations from which the solutions of PII and of a special case of PIII are to be found. However, they did not investigate in general the question of existence of solutions of their integral equations.

From the above comments one sees the richness and broad mathematical content associated with the investigation of Painlevé equations. Undoubtedly, considerable research will continue in this area.

§2. A Method for Investigating the Transformation Properties of Second Order Equations of the Painlevé Type

Suppose we are given one of the fifty equations found by Painlevé and his school, which we write in the form

$$v'' = P_1 v'^2 + P_2 v + P_3, \tag{2.1}$$

where P_1, P_2, P_3 depend on v,z and a set of parameters denoted here by $\underline{\alpha}$.

The first problem is to find the discrete Lie-point symmetries of (2.1), i.e., to find transformations of the form

$$\hat{v}(z;\hat{\underline{\alpha}}) = F(v(z;\underline{\alpha}),z), \tag{2.2}$$

where the function F is such that if $v(z;\underline{\alpha})$ solves (2.1) with parameters $\underline{\alpha}$, then $\hat{v}(z;\hat{\underline{\alpha}})$ solves (2.1) with parameters $\hat{\underline{\alpha}}$. Using the singularity structure of (2.1), the procedure of finding such transformations is immensely simplified: Since the only transformation of the type (2.2) preserving the Painlevé property is the Möbious transformation, one immediately replaces (2.2) by

$$\hat{v}(z;\hat{\underline{\alpha}}) = \frac{a_1 v + a_2}{a_3 v + a_4}, \tag{2.3}$$

where $a_1,\ldots,a_4$ are functions of z only. Using (2.3) the Lie-point discrete symmetries of (2.1) are easily obtained.

Having obtained the Lie-point symmetries of (2.1) one may look for generalized discrete symmetries of (2.1) [28], i.e. for transformations of the form

$$\hat{v}(z;\hat{\underline{\alpha}}) = F(v'(z;\underline{\alpha}),\ v(z;\underline{\alpha}),\ z)\ . \tag{2.4}$$

However, since we are not only interested in finding transformations relating the same equation, but also relating two different equations of Painlevé type, we replace (2.4) by

$$u(z;\hat{\underline{\alpha}}) = F(v'(z;\underline{\alpha}),\ v(z;\underline{\alpha}),z), \tag{2.5}$$

where F is such that u satisfies some second order equation of the Painlevé type.

The only transformation of the type (2.5), linear in v' [29], preserving the Painlevé property is the one involving the Riccati equation, i.e.

$$u(z;\underline{\hat{\alpha}}) = \frac{v'+av^2+bv+c}{dv^2+ev+f}, \tag{2.6}$$

where a,b,...,f depend on z only. Equation (2.6) plays a central role in our analysis.

The algorithm

Given equation (2.1) determine a,...,f by requiring that (2.6) defines a one to one invertible map between solutions v of (2.1) and solutions u of some second order equation of the Painlevé type. In this process the latter equation is completely determined.

Let us be more specific. Introducing the notation

$$J = dv^2 + ev + f, \quad Y = av^2 + bv + c, \tag{2.7}$$

differentiating (2.6), and using (2.1) to replace v'' and (2.6) to replace v', one obtains

$$\begin{aligned} Ju' = {} & [P_1J^2 - 2dJv - eJ]u^2 + [-2P_1JY + JP_2 + 2avJ + bJ + 2dvY + eY \\ & - (d'v^2 + e'v + f')]u + [P_1Y^2 - P_2Y + P_3 - 2avY - bY + a'v^2 + b'v + c']. \end{aligned} \tag{2.8}$$

There are two cases to be distinguished:

A) Find a,...,f such that (2.8) reduces to a linear equation for v,

$$A(u',u,z)v + B(u',u,z) = 0. \tag{2.9}$$

Having determined a,...,f upon substitution of $v = -\frac{B}{A}$ in (2.6) one determines the equation for u, which is of the same type as (2.1) (i.e. it will be one of the fifty equations mentioned above).

B) Find a,...,f such that (2.8) reduces to a quadratic equation for v,

$$A(u',u,z)v^2 + B(u',u,z)V + C(u',u,z) = 0. \qquad (2.10)$$

Then (2.6) yields an equation for u, which is quadratic in the second derivative. These types of equations, having the Painlevé property have not previously been considered in the literature.

Note that i) It turns out that PII-V admit transformations of both types A) and B). However, PVI does not admit a transformation of the type A) above.
ii) Utilizing the way that the parameters $\underline{\alpha}$ enter in the equation for u one can find a transformation relating equation (2.1) with different $\underline{\alpha}$'s.
iii) When equations (2.9) and/or (2.10) break down (i.e. A = B = C = 0), they define one-parameter families of solutions of (2.1). Using these solutions and the transformation properties of (2.1) new one-parameter families of solutions can be obtained.

These points will be clarified after applying the above method to PII.

§3. Painlevé II

In this section we use Painleve II to illustrate: α) How the transformation (2.6) can be used for investigating the transformation properties of a given equation. β) How certain of these transformation properties can be used for obtaining elementary solutions.

Note that i) Here we look only for transformations of the type A), i.e. we invoke (2.9). This is only for convenience. We stress that transformations of the type B), (see (2.10)) exist for all PII-PVI. In this paper we shall consider such transformations by necessity for PVI (since transformations A) do not exist in this case) and for an aid to the reader for PIII.
ii) Painlevé and his school found that some of the fifty equations mentioned in §1 are related to PI-PVI. For example equations XXXIV, XXXV, XLV, XLVI, XLVII of [1]

are related to PII. An exhaustive investigation of transformations A), not only establishes this relationship, but also gives a one to one correspondence between PII and each of the above equations. However, here we only present the relevant result for XXXIV and also comment on XXXV. We note that if one is simply interested in finding a transformation mapping PII to PII then any of the above transformations may be used.

Theorem 3.1

Let $v(z;\alpha)$ be a solution of PII

$$v'' = 2v^3 + zv + \alpha. \tag{3.1}$$

Then $\bar{v}(z;\bar{\alpha})$ are also solutions of PII, where

$$\bar{v}(z;\bar{\alpha}) = -v(z;\alpha);\ \bar{\alpha} = -\alpha, \tag{3.2}$$

$$\bar{v}(z;\bar{\alpha}) = -v(z;\alpha) - \frac{1+2\alpha}{2v^2+2v'+z};\ \bar{\alpha} = \alpha+1, \alpha \neq -\frac{1}{2}. \tag{3.3}$$

The case $\alpha = -1/2$ is considered in Lemma 3.1.

Theorem 3.2

Let $v(z;\alpha)$ be a solution of PII and let $u(z;\nu)$ be a solution of

$$u'' + 2u^2 - zu + \frac{\nu+u'-(u')^2}{2u-z} = 0; \quad \nu = \alpha(\alpha+1). \tag{3.4}$$

Then there exists the following one to one correspondence between solutions of (3.1) and (3.4)

$$u = -v' - v^2, \quad v = \frac{u'+\alpha}{2u-z}. \tag{3.5}$$

Equation (3.4) under the transformation $w = (u - z/2)/(4\alpha+1)$ reduces to XXXIV of [1].

Lemma 3.1

PII admits a one-parameter family of solutions characterized by

$$v' + v^2 + z/2 = 0, \tag{3.6}$$

iff $\alpha = -1/2$.

Derivation of the above results

Comparing (3.1) and (2.1) one finds that $P_1 = P_2 = 0$ and $P_3 = 2v^3 + zv + \alpha$. In considering equation (2.8) one has to consider separately the two cases of $d = 0$ or $d \neq 0$. Here we only consider $d = 0$. Then (2.8) becomes

$$u' = -eu^2 + [\frac{3aev^2+(2af+2eb-e')v+(bf+ec-f')}{ev+f}]u+$$

$$[\frac{2(1-a^2)v^3+(a'-3ab)v^2+(z+b'-b^2-2ac)v+(\alpha-bc+c')}{ev+f}]. \qquad (3.7)$$

Our goal now is to choose a,b,c,e,f in such a way that (3.7) becomes a linear equation for v. It is clear that this will be the case if each of the above brackets (i.e. the coefficient of u and the term independent of u) is linear in v. Then, it is obvious that $a^2=1$ and that also ev+f must devide each numerator appearing in the above brackets. Demanding this to be the case with $e \neq 0$, one is lead to establish a one to one correspondence between PII and XXXV of [1]. However, a simpler possible case is $e = 0$. Then, it is clear from equation (2.6) that one may take, without loss of generality, $f = -1$ (the minus sign is only for relating u directly to (1.4)) and $c = 0$, since one can always "absorb" them in u by a Möbious transformation. Hence, inserting $c = d = e = 0$, $f = -1$ in (3.7), this equation reduces to

$$u' = (2av + b)u + 3abv^2 - (z + b'-b^2)v - \alpha. \qquad (3.8)$$

Thus necessarily $b = 0$. Hence, equations (2.6) and (3.8) imply

$$(u'+\alpha) = v(2au - z),\ u = -(v'+av^2);\ \ a^2=1. \qquad (3.9)$$

Taking for convenience $a = 1$ and substituting $v = (u'+\alpha)/(2u-z)$ in the above expression for u, equation (3.4) follows.

i) $\alpha = -1/2$

The transformation (3.5b) breaks down iff $u = z/2$. But then $\alpha + u'$ must be zero, or $\alpha = -1/2$. (Actually one can easily check that $u = z/2$, $\alpha = -1/2$ solve (3.4)). Hence equation (3.5a) implies Lemma 3.1.

ii) The transformation from PII to PII

Using the above results (i.e. the results of Theorem 3.2) one can easily derive (3.3). The basic idea is to exploit the fact that ν in (3.4) is quadratic in α. Therefore, there exist two values of the parameter α, namely α and $-(\alpha+1)$ which give the same value of ν and hence the same value of u, i.e. $u(z;\alpha) = u(z; -(\alpha+1))$. But then

$$\bar{v}(z; -(\alpha+1)) = \frac{u'(z;-(\alpha+1))-(\alpha+1)}{2u(z;-(\alpha+1))-z} = \frac{u'(z;\alpha)-(\alpha+1)}{2u(z;\alpha)-z} = v(z;\alpha) - \frac{(2\alpha+1)}{2u(z;\alpha)-z}.$$

Hence, replacing u by $-(v^2+v')$ and $\bar{v}(z;-(\alpha+1))$ by $-\bar{v}(z;(\alpha+1))$, equation (3.3) follows.

How to obtain elementary solutions

The transformations (3.2) and (3.3) can be used to obtain all known elementary solutions of PII. Similarly, one can use transformations (3.5) to obtain elementary solutions of equation (3.4). First note that (3.3), (3.5) imply that

$$\bar{u}(z;-\alpha) = u(z;\alpha-1) \tag{3.10}$$

$$\bar{u}(z;\alpha+1) = -u(z;\alpha) - 2\left(\frac{u'(z;\alpha) - (\alpha+1)}{2u(z;\alpha)-z}\right)^2. \tag{3.11}$$

i) Rational solutions of PII.

It is clear that $v=0$, $\alpha=0$ solve PII. Then using (3.3) one can obtain a rational solution of PII for every positive integer:

$$v(z;0) = 0,\ v(z;1) = -\frac{1}{z},\ v(z;2) = \frac{1}{z} - \frac{3z^2}{z^3+4}, \ldots . \tag{3.12}$$

Then equation (3.2) generates a rational solution for every negative integer.

ii) Rational solutions of (3.4).

$u = 0$, $\alpha = 0$ solve equation (3.4). Then (3.10), (3.11) imply similar results as i) above:

$$u(z;0) = 0,\ u(z;1) = -\frac{2}{z^2},\ u(z;2) = -\frac{6z(z^3-8)}{(z^3+z)^2},\ldots . \quad (3.13)$$

Note that the hierarchies of solutions (3.12) and (3.13) are related by the transformations (3.5).

iii) Airy type solutions of PII.

Lemma 3.1 implies that $v(z; -\frac{1}{2}) = y'/y$ is a solution of PII, where y is any solution of the Airy equation $y'' + (z/2)y = 0$. One can not use directly this solution in (3.3) to generate new solutions, because in this case (3.3) breaks down. The trick is to first use (3.2) and then (3.3). In this way one generates the following hierarchy of solutions:

$$v(z; -\tfrac{1}{2}) = y'/y,\ v(z;\tfrac{1}{2}) = -y'/y,\ v(z;\tfrac{3}{2}) = y'/y - \frac{y^2}{2y'^2+zy^2},\ldots . \quad (3.14)$$

iv) Airy type solutions of (3.4).

Similarly as above

$$u(z; -\tfrac{1}{2}) = z/2,\ u(z;\tfrac{1}{2}) = -\frac{2y'^2}{y^2} - \frac{z}{2},\ldots . \quad (3.15)$$

Remarks

1. The results of Theorem 3.1 and Lemma 3.1 were first given in [14] and [3] respectively. The result of Theorem 3.1 was rederived later in [30] and [31] by exploiting the connection with the inverse scattering of the KdV equation.

2. We emphasize that the logical steps used here for α) deriving (3.4) and (3.5),

β) exploiting the quadratic dependence of ν on α to obtain (3.3), γ) characterizing one-parameter families of solutions of PII when the transformation (3.5) breaks down and δ) generating elementary solutions, remain valid for considering all Painlevé equations PII-PVI.

§4. Painlevé III

In this section we consider PIII. Having familiarized the reader with our method we now present both types of transformations A) and B).

Theorem 4.1

Let $v(z;\alpha,\beta,\gamma,\delta)$ be a solution of PIII

$$v'' = \frac{v'^2}{v} - \frac{1}{z}v' + \frac{1}{z}(\alpha v^2+\beta) + \gamma v^3 + \frac{\delta}{v} . \tag{4.1}$$

Then $\bar{v}(z;\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta})$ are also solutions of PIII, where

$$\bar{v}(z;\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}) = -v(z;\alpha,\beta,\gamma,\delta);\ \bar{\alpha} = -\alpha,\ \bar{\beta} = -\beta,\ \bar{\gamma} = \gamma,\ \bar{\delta} = \delta, \tag{4.2}$$

$$\bar{v}(z;\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}) = [v(z;\alpha,\beta,\gamma,\delta)]^{-1};\ \bar{\alpha} = -\beta,\ \bar{\beta} = -\alpha,\ \bar{\gamma} = -\delta,\ \bar{\delta} = -\gamma, \tag{4.3}$$

$$\bar{v}(z;\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}) = \frac{\gamma^{1/2}}{\bar{\gamma}^{1/2}} v\left[1 + \frac{2 + \beta(-\delta)^{1/2} + \alpha(\gamma)^{-1/2}}{z(\frac{v'}{v} + \gamma^{1/2}v + \frac{(-\delta)^{1/2}}{v}) - 1 - \beta(-\delta)^{-1/2}}\right]; \tag{4.4a}$$

$$\bar{\alpha} = -[2+\beta(-\delta)^{-1/2}]\gamma^{-1/2}, \bar{\beta} = -[2+\alpha\gamma^{-1/2}](-\delta)^{1/2}(\frac{\gamma^{1/2}}{\bar{\gamma}^{1/2}}), (-\bar{\delta})^{1/2} = (-\delta)^{1/2}\frac{\gamma^{1/2}}{\bar{\gamma}^{1/2}} . \tag{4.4b}$$

In (4.4) we have assumed that

$$\gamma \neq 0 \text{ and } 2 + \alpha(\gamma)^{-1/2} + \beta(-\delta)^{1/2} \neq 0. \tag{4.4c}$$

If $\gamma = 0$ then (4.4) is replaced by

$$\bar{v}(z;\bar{\alpha},\bar{\beta},0,\bar{\delta}) = z\,\frac{v'}{v^2} - (1+\beta(-\delta)^{1/2})v + \frac{z(-\delta)^{1/2}}{v^2}, \tag{4.5a}$$

$$\bar{\alpha} = (-\delta)^{1/2},\ \bar{\beta} = \alpha(2-\beta(-\delta)^{1/2}),\ \bar{\delta} = -\alpha^2. \tag{4.5b}$$

The case $2 + \beta(-\delta)^{1/2} + \alpha(\gamma)^{-1/2} = 0$ is considered in Lemma 4.1.

Theorem 4.2

Let $v(z;\alpha,\beta,1,\delta)$ [32] be a solution of PIII and let $w(x;\bar{\alpha},\bar{\beta},\bar{\gamma},0)$ be a solution of PV,

$$\frac{d^2w}{dx^2} = \frac{3w-1}{2w(w-1)}\left(\frac{dw}{dx}\right)^2 - \frac{1}{x}\frac{dw}{dx} + \frac{(w-1)^2}{x^2}\left(\frac{\bar{\alpha}}{4}w + \frac{\bar{\beta}}{4w}\right) + \frac{(-\bar{\delta})^{1/2}w}{x}, \tag{4.6a}$$

where

$$x = \frac{z^2}{2},\ \bar{\alpha} = \frac{(1+\alpha-\mu)^2}{32},\ \bar{\beta} = -\frac{(1-\alpha-\mu)^2}{32},\ \bar{\gamma} = (-\delta)^{1/2},\ \mu = -[1+\beta(-\delta)^{1/2}]. \tag{4.6b}$$

Then there exists the following one to one correspondence between solutions of (4.1) and (4.6)

$$\begin{aligned} &v'(w-1) + (w+1)v^2 + [\tfrac{\mu}{2}v + (-\delta)^{1/2}](w-1) = 0 \\ &zv = -\frac{w'}{w} + (1+\alpha-\mu)w - \frac{\alpha}{2} + \frac{\mu+\alpha-1}{w}. \end{aligned} \tag{4.7}$$

Theorem 4.3

Let $v(z;\alpha,\beta,\gamma,\delta)$ be a solution of PIII and let $\phi(z;\rho,\sigma,\tau)$ be a solution of

$$(\phi'' + \rho\phi + \tfrac{\sigma}{2})^2 = \frac{\phi^2}{z^2}(\phi'^2 + \rho\phi^2 + \sigma\phi + \tau); \tag{4.8a}$$

$$\rho = 4\gamma^{1/2}(-\delta)^{1/2},\ \sigma = 4\,[\alpha(-\delta)^{1/2} - \beta\gamma^{1/2}],\ \tau = -4(\alpha+\gamma^{1/2})[\beta+(-\delta)^{1/2}]. \tag{4.8b}$$

Then there exists the following one to one correspondence between solutions of (4.1) and (4.8)

$$\frac{\phi}{z} = \frac{v'}{v} + \gamma^{1/2} v + \frac{(-\delta)^{1/2}}{v},$$

$$v = \frac{\phi' + \frac{z}{\phi}(\phi'' + \rho\phi + \frac{\sigma}{2})}{2(\gamma^{1/2}\phi + \alpha + \gamma^{1/2})}, \tag{4.9}$$

provided that neither $\beta = \delta = 0$, nor $\alpha = \gamma = 0$, nor $2 + \alpha\gamma^{-1/2} + \beta(-\delta)^{-1/2} = 0$. These exceptional cases are considered below.

Lemma 4.1

PIII admits a one-parameter family of solutions characterized by

$$\frac{(1-\alpha\gamma^{-1/2})}{z} = \frac{v'}{v} + \gamma^{1/2} v + \frac{(-\delta)^{1/2}}{v}, \tag{4.10}$$

iff

$$2 + \alpha\gamma^{-1/2} + \beta(-\delta)^{-1/2} = 0.$$

Lemma 4.2

The general solution of PIII in the case that $\beta = \delta = 0$ is given by

$$v = \frac{\phi'}{\gamma^{1/2}\phi + \alpha + \gamma^{1/2}}; \int \frac{d\phi}{\frac{\phi^2}{2} + \phi + c_1} = \ln z + c_2, \tag{4.11}$$

where c_1, c_2 are arbitrary constants.

Using (4.3) similar results are obtained for the case that $\alpha = \gamma = 0$.

Remarks

1. The results of Theorem 4.2 and the above Lemma's were first given in [15], [11], and [33] respectively.

2. One can clearly combine the transformations (4.2) to (4.4) to obtain new transformations. For example combining (4.3) and (4.4) one can derive the corresponding result of [15]. Also note that a finite number of products of (4.2) and (4.3) yields the identity.

3. Elementary solutions of PIII can be derived in the same manner as in PII.

Derivation of the above results

In this case $P_1 = \frac{1}{v}$, $P_2 = -\frac{1}{z}$, $P_3 = \frac{1}{z}(\alpha v^2+\beta) + \gamma v^3 + \frac{\delta}{v}$. Thus the coefficient of u^2 in (2.8) is $-(dv^2+f)/v$.

α) Requiring (2.8) to be linear in v.

Then the above coefficient of u^2 implies that either d = 0 or f = 0. In what follows we shall consider only the case f = 0. (The result of [15] is derivable by considering the case d = 0). With f = 0 equation (2.8) becomes

$$u' = (-dv)u^2 + [\frac{2adv^3+(bd+ea-d'-dz^{-1})v^2-(e'+ez^{-1})v-ec}{dv^2+ev}]u + [\ldots], \quad (4.12)$$

where the last bracket is independent of u. The coefficient of u in the above will be linear in v iff $dv^2 + ev$ is a root of the numerator. This implies that e = 0. Then, without loss of generality we can take d = 1 and a = 0, since one can always "absorb" them in u with the aid of a Möbious transformation. If e = a = 0 and d = 1 then the term in (4.12) independent of u, is linear in v iff $c^2 + \delta = 0$ and $b = -(\beta+c)/cz$. Thus, with the above choices of a,...,f equations (2.6) and (4.12) yield

$$u = \frac{v'}{v^2} + \frac{b}{v} + \frac{c}{v^2}\ ;\ c = (-\delta)^{1/2},\ b = -(\frac{\beta+c}{cz}) \quad (4.13)$$

$$v = \frac{u' + (z^{-1}-b)u - \alpha z^{-1}}{\gamma-u^2}\ . \quad (4.14)$$

Substituting (4.14) into (4.13) one obtains an equation for u, namely

$$u'' = \frac{uu'^2}{u^2-\gamma} + \ldots\ . \quad (4.15)$$

It is then clear that $\gamma = 0$ is of special interest.

i) $\gamma = 0$

Then (4.15) becomes

$$u'' = \frac{u'^2}{u} - \frac{u'}{z} - \frac{\alpha^2}{z^2 u} + \frac{\alpha}{z^2}(1 + \frac{c-\beta}{c}) + cu^2.$$

Let $u = \frac{\bar{v}}{z}$ to obtain

$$\bar{v}'' = \frac{\bar{v}'^2}{\bar{v}} - \frac{\bar{v}'}{z} + c\frac{\bar{v}^2}{z} + \alpha(1 + \frac{c-\beta}{c}) - \frac{\alpha^2}{\bar{v}}. \tag{4.16}$$

Equations (4.13) (wth $u = \bar{v}/z$) and (4.16) imply (4.5).

ii) $\gamma \neq 0$

Without loss of generality take $\gamma = 1$. Equation (4.15) must be, within a Möbious transformation, one of the fifty equations mentioned in the introduction. Thus let $u = (Aw+B)/(Cw+D)$ to transform (4.15) to

$$w'' = [2C + \frac{(Aw+B)(AD-CB)}{[(Aw+B)^2 - \gamma(Cw+D)^2]Cw+D}]w'^2 + \ldots . \tag{4.17}$$

Hence, if $A = B = z$, $D = -C = 1$ then $w'' = (1/2w - 1/(w-1))w'^2 + \ldots$. Therefore, under the transformation

$$u = -\frac{(w+1)}{w-1} \tag{4.18}$$

equation (4.15) becomes

$$w'' = \frac{(3w-1)}{2w(w-1)} w'^2 - \frac{1}{z}w' + \frac{(w-1)^2}{z^2}(\bar{\alpha}w + \frac{\bar{\beta}}{w}) + 2cw, \tag{4.19}$$

where $\bar{\alpha},\bar{\beta}$ are defined in (4.6b). Letting $z = (2x)^{1/2}$ in (4.19) and using (4.18) in (4.13) and (4.15) the result of Theorem 4.2 follows.

iii) $2 + \beta(-\delta)^{1/2} - \alpha\gamma^{-1/2} = 0$

The transformation (4.14) breaks down iff $u = \gamma^{1/2}$. Then $u' + (z^{-1}-b)u-\alpha z^{-1}$ must be zero when $u = \gamma^{1/2}$. This implies $2 + \beta(-\delta)^{1/2} - \alpha\gamma^{-1/2} = 0$. Hence, using

(4.13) one obtains the result of Lemma 4.1.

Using Theorem 4.2 one can obtain (4.4) in a similar way as in obtaining the corresponding result for PII. However, we choose to obtain (4.4) using Theorem 4.3, which is now derived.

β) Requiring (2.8) to be quadratic in v.

Take for example d = f = 0. (We remind the reader that our investigation is not exhaustive.) Then, without loss of generality b = 0 and e = 1. Hence, (2.8) becomes

$$u' = u^2\left[\frac{av^2 - z^{-1}v - c}{\gamma}\right] + [\ ]\,u + [\ ].$$

The two brackets [ ], in the above equation contain v quadratically iff

$$a^2 = \gamma,\ c^2 + \delta = 0. \tag{4.20}$$

Therefore, with the above choices of a,...,f equations (2.6) and (2.8) yield

$$\frac{\phi}{z} = \frac{v'}{v} + \gamma^{1/2}v + \frac{(-\delta)^{1/2}}{v} \tag{4.21}$$

$$(\gamma^{1/2}+\alpha+\gamma^{1/2}\phi)v^2 - \phi'v + (\beta+(-\delta)^{1/2}-(-\delta)^{1/2}\phi) = 0, \tag{4.22}$$

where we have used for convenience the substitution u = φ/z. Equation (4.22) yields

$$v = \frac{\phi' + \Delta^{1/2}}{2(\gamma^{1/2}\phi+\alpha+\gamma^{1/2})},\quad \Delta = \phi'^2 + \rho\phi^2 + \sigma\phi + \tau, \tag{4.23}$$

where ρ,σ,τ are defined in (4.8b). Noting that Δ' = 2φ'Ω, where

$$\Omega = \phi'' + \rho\phi + \frac{\sigma}{2}, \tag{4.24}$$

and substituting (4.22) into (4.21) one obtains an equation for φ. This equation, using the fact that

$$\rho\phi^2+\sigma\phi+\tau = 4(\gamma^{1/2}\phi+\alpha+\gamma^{1/2})[(-\delta)^{1/2}\phi-\beta-(-\delta)^{1/2}]$$

takes the form $(\Delta^{1/2} + \phi')(\Delta^{1/2}\phi/z - \Omega) = 0$, which implies

$$\frac{\Delta^{1/2}\phi}{z} = \Omega. \tag{4.25}$$

Therefore, the transformations (4.21) and (4.23) relate PIII and equation (4.25). Using (4.25) in (4.23) one obtains the result of theorem 4.3.

i) $\underline{\beta = \delta = 0}$

Then, using (4.21), (4.23) and (4.25) (where we pick the positive root of $\Delta^{1/2}$) we have

$$\frac{\phi}{z} = \frac{v'}{v} + \gamma^{1/2}v, \quad v = \frac{\phi'}{\gamma^{1/2}\phi+\alpha+\gamma^{1/2}}, \quad \phi'' - \frac{\phi\phi'}{z} = 0. \tag{4.26}$$

However, the equation for ϕ is now very simple and it can be immediately integrated to $z\phi' = \phi^2/2 + \phi + c_1$. Thus Lemma 4.2 immediately follows.

ii) $\underline{2 + \alpha\gamma^{-1/2} + \beta(-\delta)^{-1/2} = 0}$

The transformation (4.23) breaks down iff $\gamma^{1/2}\phi+\alpha+\gamma^{1/2} = 0$. But then $\phi' + \Delta^{1/2} = 0$ which implies $2 + \alpha\gamma^{-1/2} + \beta(-\delta)^{-1/2} = 0$. Then, using equation (4.21) the result of Lemma 4.1 is again derived.

iii) The transformation from PIII to PIII

This transformation is easily obtained by finding two sets of $\{\alpha,\beta,\gamma,\delta\}$ which give the same values for ρ,σ,τ. Solving the equations defining ρ,σ,τ for α,β,δ (keeping γ fixed) one finds

$$(-\bar{\delta})^{1/2} = \frac{\rho}{4}(\bar{\gamma})^{-1/2}, \quad \bar{\beta} = [\bar{\alpha}(-\bar{\delta})^{1/2} - \frac{\sigma}{4}](\bar{\gamma})^{-1/2}, \quad \bar{\alpha} = \frac{2}{\rho}[\frac{\sigma}{4} - \frac{\rho}{2} \pm \sqrt{\frac{\sigma^2}{16} - \frac{\tau\rho}{4}}\,](\bar{\gamma})^{1/2}. \tag{4.27}$$

However, using the definitions of ρ,σ,τ it follows that

$$\frac{\sigma^2}{16} - \frac{\tau\rho}{4} = [\alpha(-\delta)^{1/2} + \beta\gamma^{1/2} + 2\gamma^{1/2}(-\delta)^{1/2}]^2.$$

Using the positive root if (4.27) one finds the trivial result

$$\bar{\alpha} = \alpha(\frac{\bar{\gamma}^{1/2}}{\gamma^{1/2}}), \quad \bar{\beta} = \beta(\frac{\gamma^{1/2}}{\bar{\gamma}^{1/2}}), \quad (-\bar{\delta})^{1/2} = (-\delta)^{1/2}(\frac{\gamma^{1/2}}{\bar{\gamma}^{1/2}}).$$

However, using the negative root, one obtains the expressions for $\bar{\alpha},\bar{\beta},\bar{\gamma}$ appearing in (4.4b). Then using

$$\bar{v} = \frac{\phi' + \Delta^{1/2}}{2(\bar{\gamma}^{1/2}\phi+\bar{\alpha}+\bar{\gamma}^{1/2})} = \frac{v(\gamma^{1/2}\phi+\alpha+\gamma^{1/2})}{\bar{\gamma}^{1/2}\phi+\bar{\alpha}+\bar{\gamma}^{1/2}}$$

and replacing ϕ by (4.21), equations (4.4) follow.

§5. Painlevé IV and V

Using our method one can easily find transformations which map PIV and PV to themselves, but with different values of the parameters. These transformations were first given in [16] and [17] respectively. Here, for completeness, we give these transformations and advise the interested reader to derive them himself as a simple exercise of our method.

Theorem 5.1

Let $v(z;\alpha,\beta)$ be a solution of PIV

$$v'' = \frac{v'^2}{2v} + \frac{3}{2}v^3 + 4zv^2 + 2(z^2-\alpha)v + \beta v^{-1}. \tag{5.1}$$

Then $\bar{v}(z;\bar{\alpha},\bar{\beta})$ is also a solution of PIV, where

$$\bar{v} = \frac{v'-v^2-2zv-(-2\beta)^{1/2}}{2v}\ ; \ \bar{\alpha} = \frac{1}{4}[2-2\alpha+3(-2\beta)^{1/2}], \ \bar{\beta} = -\frac{1}{2}[1+\alpha+\frac{(-2\beta)^{1/2}}{2}]^2, \tag{5.2}$$

provided that

$$1 + \alpha + \frac{(-2\beta)^{1/2}}{2} \neq 0. \tag{5.3}$$

Theorem 5.2

Let $v(z;\alpha,\beta,\gamma,\delta)$ be a solution of PV

$$v'' = \frac{3v-1}{2v(v-1)}v'^2 - \frac{1}{z}v' + \frac{\alpha}{z^2}v(v-1)^2 + \frac{\beta}{z^2}\frac{(v-1)^2}{v} + \frac{\gamma}{z}v + \frac{\delta v(v+1)}{v-1}. \quad (5.4)$$

Then $\bar{v}(z;\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta})$ is also a solution of PV, where

$$\bar{v} = 1 - \frac{2(-2\delta)^{1/2}zv}{zv'-(2\alpha)^{1/2}v^2+[(2\alpha)^{1/2}-(-2\beta)^{1/2}+(-2\delta)^{1/2}z]v+(-2\beta)^{1/2}}, \quad (5.5a)$$

$$\bar{\alpha} = -\frac{1}{16\delta}[\gamma+(-2\delta)^{1/2}(1-(-2\beta)^{1/2}-(2\alpha)^{1/2})]^2, \quad \bar{\beta} = \frac{1}{16\delta}[\gamma-(-2\delta)^{1/2}(1-(-2\beta)^{1/2}+(-2\delta)^{1/2}z)]^2 \quad (5.5b)$$

$$\bar{\gamma} = (-2\delta)^{1/2}[(-2\beta)^{1/2}-(2\alpha)^{1/2}], \quad \bar{\delta} = \delta$$

provided that $\delta \neq 0$ and

$$(-2\delta)^{1/2}[1-(-2\beta)^{1/2}-(2\alpha)^{1/2}] \neq \gamma. \quad (5.6)$$

Remarks

1. One can easily find Lie-point discrete symmetries of the above equations. For example, if $v(z;\alpha,\beta,\gamma,\delta)$ solves PV then $\bar{v} = v^{-1}(z;-\beta,-\alpha,-\gamma,\delta)$ also solves PV.

2. When the above transformations break down, i.e. when (5.3) and/or (5.6) are violated then PIV and/or PV just as for PII and PIII admit one-parameter family of solutions.

3. Elementary solutions of the above equations can be derived in a similar manner as in PII.

§6. Painlevé VI

Theorem 6.1

Let $v(z;\alpha,\beta,\gamma,\delta)$ be a solution of PVI:

$$v'' = \frac{1}{2}\left(\frac{1}{v} + \frac{1}{v-1} + \frac{1}{v-z}\right)v'^2 - \left(\frac{1}{z} + \frac{1}{z-1} + \frac{1}{v-z}\right)v' + \frac{v(v-1)(v-z)}{z^2(z-1)^2}\left[\alpha + \frac{\beta z}{v^2} + \frac{\gamma(z-1)}{(v-1)^2} + \frac{\delta z(z-1)}{(v-z)^2}\right]. \tag{6.1}$$

Then $\bar{v}(z;\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}$ are also solutions of PVI, where

$$\bar{v}(z;\ \bar{\alpha},\ \bar{\beta},\ \bar{\gamma},\ \bar{\delta}) = zv\left(\frac{1}{z};\ \alpha,\ \beta,\ \gamma,\ \delta\right);\quad \bar{\alpha}=\alpha,\ \bar{\beta}=\beta,\ \bar{\gamma}=-\delta+\tfrac{1}{2},\ \bar{\delta}=-\gamma+\tfrac{1}{2}, \tag{6.2}$$

$$\bar{v}(z;\ \bar{\alpha},\ \bar{\beta},\ \bar{\gamma},\ \bar{\delta}) = 1-v(1-z;\ \alpha,\ \beta,\ \gamma,\ \delta);\ \bar{\alpha}=\alpha,\ \bar{\beta}=-\gamma,\ \bar{\gamma}=-\beta,\ \bar{\delta}=\delta, \tag{6.3}$$

$$\bar{v}(z;\ \bar{\alpha},\ \bar{\beta},\ \bar{\gamma},\ \bar{\delta}) = 1-(1-z)v\left(\frac{1}{1-z};\ \alpha,\ \beta,\ \gamma,\ \delta\right);\ \bar{\alpha}=\alpha,\ \bar{\beta}=\delta-\tfrac{1}{2},\ \bar{\gamma}=-\beta,\ \bar{\delta}=-\gamma+\tfrac{1}{2}, \tag{6.4}$$

$$\bar{v} = v + 2\,\frac{(z+1)v - 2z}{-\frac{2z(z-1)}{\kappa}\frac{\phi'}{\phi} + \frac{(z-1)I}{\kappa\phi} - (z+1)}; \tag{6.5a}$$

$$\bar{\alpha} = \frac{1}{2}\left[(-2\beta)^{1/2}-1\right]^2,\ \bar{\beta} = -\frac{1}{2}\left[(2\alpha)^{1/2}+1\right]^2,\ \bar{\gamma} = \gamma + \frac{\kappa\mu}{4},\ \bar{\delta} = \delta + \frac{\kappa\mu}{4}. \tag{6.5b}$$

In (6.5) ϕ, I,κ,μ are defined by

$$\phi = z\frac{v'}{v} + \frac{(\lambda-\kappa-1)}{2(z-1)}v + \frac{(\lambda+\kappa+1)z}{2(z-1)}\frac{1}{v} - \frac{\lambda}{2}\frac{(z+1)}{(z-1)} - \left(\tfrac{1}{2} + \frac{\mu}{4}\right), \tag{6.6}$$

$$I \doteq \phi^2 + \frac{\mu}{2}\phi + \nu, \tag{6.7}$$

$$\kappa = (-2\beta)^{1/2} - (2\alpha)^{1/2} - 1,\ \lambda = (-2\beta)^{1/2} + (2\alpha)^{1/2},\ \mu = \frac{4}{\kappa}\left(\tfrac{1}{2}-\gamma-\delta\right),\ \nu = 2\delta - 1 + \left(\frac{\mu}{4} + \frac{\kappa}{2}\right)^2 \tag{6.8}$$

In (6.5) we have assumed that

$$\phi \neq 0,\ \kappa \neq 0,\ \nu \neq 0. \tag{6.9}$$

Theorem 6.2

Let $v(z;\alpha,\beta,\gamma,\delta)$ be a solution of PVI

and let $\phi(z;\ \kappa^2,\ \lambda,\ \mu,\ \nu)$ be a solution of

$$(z-1)^2\Omega^2 = \frac{1}{z^2}\left(\phi'^2 + \frac{I^2-\kappa^2\phi^2}{z(z-1)^2}\right)\Psi^2, \tag{6.10}$$

where

$$\Omega \doteq \phi'' + \frac{(3z-1)\phi'}{2z(z-1)} + \frac{2\phi I + \frac{\mu I}{2} - \phi\kappa^2}{z(z-1)^2},\ \Psi \doteq (z+1)\phi + \frac{\mu}{4}(z+1) + \frac{\lambda}{2}(z-1), \tag{6.11}$$

and I,κ,λ,μ,ν are defined by (6.7) and (6.8). Then equations (6.6) and (6.12) below, express a one to one correspondence between solutions of (6.1) and (6.10)

$$v = \frac{-\frac{(z+1)}{z}\phi' + \frac{2\kappa}{z(z-1)}\phi - (z-1)^2\frac{\Omega}{\Psi}}{-\frac{2\phi'}{z} + \frac{I}{z^2} + \frac{\kappa(z+1)\phi}{z^2(z-1)}} \quad . \tag{6.12}$$

We have assumed that (6.9) is valid.

Lemma 3.1

PVI admits a one-parameter family of solutions characterized by

$$zv' + \frac{(\lambda-\kappa-1)}{2(z-1)}v^2 - [\frac{\lambda}{2}\frac{(z+1)}{(z-1)} + \tfrac{1}{2} + \frac{\mu}{4}]v + \frac{(\lambda+\kappa+1)z}{2(z-1)} = 0, \tag{6.13}$$

iff $\nu = 0$, $\kappa \neq 0$.

This result which is an immediate consequence of theorem 6.2 was first given in [18]. Note that if

$$v = -\frac{2z(z-1)}{\lambda-\kappa-1}\frac{w'}{w}, \tag{6.14}$$

then w satisfies a certain hypergeometric equation.

Derivation of the above results

In deriving the above results we follow the same logical steps as with PIII.

i) The transformation from PVI to (6.10).

In this case the coefficient of u^2 in (2.8) is

$$\frac{(3/2\ v^2 - (z+1)v + z/2)}{v(v-1)(v-z)}J^2 - 2dvJ - eJ.$$

Therefore, it is impossible to choose a,...,f in such a way that (2.8) reduces to a linear equation for v. However, by choosing $d = f = 0$, and then (without loss of generality) $e = 1$, $b = 0$, equation (2.8) reduces to a quadratic equation for v iff

$$a = \frac{(2\alpha)^{1/2}}{z(z-1)}, \quad c = \frac{(-2\beta)^{1/2}}{z-1}. \tag{6.15}$$

Then (2.6) and (2.8) become

$$v' + av^2 - uv + c = 0 \tag{6.16}$$

$$\hat{A}v^2 + \hat{B}v + \hat{C} = 0, \tag{6.17}$$

where $\hat{A}$, $\hat{B}$, $\hat{C}$ are known functions of u', u, z. Equation (6.17) may be simplified if one uses the transformation

$$u = \frac{\phi}{z} + \frac{\hat{\lambda}}{z-1} + \frac{\hat{\mu}}{z(z-1)} \; ; \; \hat{\lambda} = -\frac{(\alpha+(2\alpha)^{1/2}+\beta+\gamma+\delta)}{\kappa} \; , \; \hat{\mu} = \lambda - \hat{\lambda}, \; k\neq 0, \tag{6.18}$$

where κ,λ are defined in (6.8). Replacing u in terms of ϕ in (6.16), (6.17) one obtains

$$v' + \frac{(2\alpha)^{1/2}}{z(z-1)}v^2 - (\frac{\phi}{z} + \frac{\hat{\lambda}}{z-1} + \frac{\hat{\mu}}{z(z-1)})v + \frac{(-2\beta)^{1/2}}{z-1} = 0, \tag{6.19}$$

$$Av^2 + Bv + C = 0, \tag{6.20a}$$

where

$$A = -\frac{\phi'}{z} + \frac{1}{2z^2}(I + \frac{\kappa(z+1)}{(z-1)}\phi), \; B = \frac{(z+1)}{z}\phi' - \frac{2\kappa\phi}{z(z-1)} \; , \; C = -2\phi' + \frac{\kappa(z+1)}{z(z-1)}\phi - Az \tag{6.20b}$$

and I is defined by (6.7). Equations (6.19) and (6.20) are the analogues of equations (4.21) and (4.22). Equation (620) yields

$$v = \frac{-B + \frac{(z-1)}{z}\Delta^{1/2}}{2A} \; ; \quad \Delta = \phi'^2 + \frac{I^2-\kappa^2\phi^2}{z(z-1)^2} \; . \tag{6.21}$$

Substituting (6.21) into (6.19) one obtains

$$\frac{\Delta^{1/2}}{z(z-1)}\Psi + \Omega = 0, \tag{6.22}$$

where Ψ,Ω are defined by (6.11). (In obtaining this equation it is crucial to note that A and $I^2-\kappa^2\phi^2$ are common factors.) Therefore, the transformations (6.19), (6.21) define a one to one correspondence between PVI and (6.22). Note that the two different branches of $\Delta^{1/2}$ in (6.21) correspond to the two different branches of $\Delta^{1/2}$ in (6.22). If one wants to get rid of the square root in (6.22), one may

replace $\Delta^{1/2}$ in (6.21) by $-z(z-1)\Omega/\Psi$. Then (6.21) becomes (6.12), and using (6.12) in (6.22), equation (6.10) follows.

ii) $\nu = 0$

The transformation (6.21) breaks down iff $A = 0$. It then follows (requiring that $-B + (z-1)\Delta^{1/2}/z$ is also zero) that $\phi = 0$ and $\nu = 0$. Hence, substituting $\phi = 0$, $\nu = 0$ in (6.19), Lemma 6.1 follows.

iii) The transformation from PVI to PVI.

The trick again is to find two sets of parameters $\{\alpha,\beta,\gamma,\delta\}$ and $\{\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}\}$ which give rise to the same κ^2,λ,μ,ν. Hence it is clear that if the set $\{\alpha,\beta,\gamma,\delta\}$ corresponds to κ,λ,μ,ν, then the set $\{\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}\}$ must correspond to $-\kappa,\lambda,\mu,\nu$. Thus

$$\bar{\delta} = \frac{\nu}{2} + \frac{1}{2} - \frac{1}{2}\left(\frac{\mu}{4} - \frac{\kappa}{2}\right)^2 = \frac{\nu}{2} + \frac{1}{2} - \frac{1}{2}\left(\frac{\mu}{4} + \frac{\kappa}{2}\right)^2 + \frac{\kappa\mu}{4} = \delta + \frac{\kappa\mu}{4} . \qquad (6.23)$$

Similarly for $\bar{\gamma}$. Also solving the equations $(-2\bar{\beta})^{1/2}+(2\bar{\alpha})^{1/2}=(-2\beta)^{1/2}+(2\alpha)^{1/2}$ and $(-2\bar{\beta})^{1/2}-(2\bar{\alpha})^{1/2}-1 = -(-2\beta)^{1/2}+(2\alpha)^{1/2}+1$ one obtains $(2\bar{\alpha})^{1/2}= -1 +(-2\beta)^{1/2}$, $(-2\bar{\beta})^{1/2} = 1 +(2\alpha)^{1/2}$. Finally, using (6.12) with v replaced by $\bar{v}$ and κ by $-\kappa$ one obtains (6.5a).

How to obtain elementary solutions

Using theorem 6.1, one may obtain infinite hierarchies of elementary solutions of PVI. As with PII, it is important to notice that if one starts with the solution v characterized by Lemma 6.1, one cannot use directly (6.5a) (since in this case (6.5a) breaks down); one must first use a Lie-point discrete symmetry to obtain a new solution $\hat{v}$ and then use (6.5a). We also note that the Lie-point symmetry (6.2) cannot be used, because for this symmetry $\bar{\nu} = \nu$ (hence if $\nu = 0$, $\bar{\nu} = 0$). If one instead uses the Lie-point symmetry (6.4) one has the following result:

Lemma 6.2

Let $v(z;\alpha,\beta,\gamma,\delta)$ be the one-parameter family of solutions of (6.13), where κ,λ,μ are defined by (6.8) and $\nu = 0$. Use the transformation (6.4) to evaluate $\hat{v}(z;\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\delta})$. Then apply to this solution the transformation (6.5a) to obtain a new one-parameter family of solutions $\bar{v}$ of PVI, with parameters

$$\bar{\alpha} = \frac{1}{2}(\hat{\beta}_1-1)^2,\ \bar{\beta} = -\frac{1}{2}(\hat{\alpha}_1+1)^2,\ \bar{\gamma} = \hat{\gamma} + \frac{\hat{\kappa}\hat{\mu}}{4},\ \bar{\delta} = \hat{\delta} + \frac{\hat{\kappa}\hat{\mu}}{4}, \tag{6.24a}$$

where

$$\hat{\alpha} = \alpha,\ \hat{\beta} = \delta - \frac{1}{2},\ \hat{\gamma} = -\beta,\ \hat{\delta} = -\gamma + \frac{1}{2},\ \text{and } \hat{\alpha}_1=(2\alpha)^{1/2},\ \hat{\beta}_1=(-2\beta)^{1/2}. \tag{6.24b}$$

Example 6.1

The solution of equation (6.13) is in general expressible in terms of the hypergeometric functions. Hence, using the above Lemma one can obtain an infinite hierarchy of one-parameter family of solutions of PVI, all of which are related (through the repeated application of (6.5a)) to these hypergeometric functions. However, for some special choices of the parameters $\alpha,\beta,\gamma,\delta$ equation (6.13) becomes very simple. In this case one may for example derive infinite hierarchies of rational solutions. Let us pick such a choice in order to illustrate our results. Let $\alpha = \beta = 0$, $\gamma = 1/2$, $\delta = -3/2$. Then, using (6.8) $\kappa = -1$, $\lambda = 0$, $\mu = -6$, $\nu = 0$. Therefore, since $\nu = 0$, PVI must admit a one-parameter family of solutions characterized by (6.13). Actually in this case (6.13) reduces to $zv' + v = 0$ and hence $v = \tau/z$, τ some arbitrary constant. Now starting with

$$v = \tau/z;\ \alpha = \beta = 0,\ \gamma = 1/2,\ \delta = -3/2 \tag{6.25}$$

in (6.4), one finds

$$\hat{v} = 1 - \tau(1-z)^2;\ \hat{\alpha} = \hat{\gamma} = \hat{\delta} = 0,\ \hat{\beta} = -2. \tag{6.26}$$

Then, using (6.8), either

$$\hat{\kappa} = 1,\ \hat{\lambda} = 2,\ \hat{\mu} = 2,\ \hat{\nu} = 0 \quad \text{or} \quad \hat{\kappa} = -3,\ \hat{\lambda} = -2,\ \hat{\mu} = -2/3,\ \hat{\nu} = 16/9.$$

The first choice used in (6.13) rederives $\hat{v}$, however, the second choice (used in (6.5)) yields

$$\bar{v} = \frac{z(\tau z^2 - 2\tau z + \tau - 1)}{2\tau z^3 - 3\tau z^2 + \tau - 1}\ ;\ \bar{\alpha} = \frac{9}{2},\ \bar{\beta} = -\frac{1}{2},\ \bar{\gamma} = \frac{1}{2},\ \bar{\delta} = \frac{1}{2}\ . \tag{6.27}$$

One can verify directly that the functions v, v and $\bar{v}$, as defined by equations (6.25), (6.26) and (6.27) respectively, satisfy PVI.

Remarks

1) The transformation (6.4) is the product of the transformations (6.2) and (6.3). Similary one can obtain a transformation as the product of (6.3) and (6.2).

2) It is worth noting that one cannot use just the Lie-point discrete symmetries (i.e. equations (6.2) to (6.4)) to generate an infinite hierarchy of exact solutions. This, which is consistent with [28], follows from the fact that a finite number of products of these transformations yields the identity. For example, one obtains the identity after the repeated application of (6.4) three times.

Acknowledgements

This work was partially supported by the Mathematics Divisions of AFOSR and ONR.

References

[1] E.L. Ince, Ordinary Differential Equations (New York, N.Y. 1947).

[2] 1) P. Painlevé, Bull. de Soc. Math. de France 28, 214 (1900).

2) P. Painlevé, Acta Math. 25, 1 (1902).

[3] B. Gambier, Acta Math. 33, 1 (1909).

[4] Two functions u and w are related by a Möbious, or bilinear or projective transformation iff u=(Aw+B)/(Cw+D), where AD-BC ≠ 0.

[5] 1) E. Barouch, B.M. McCoy, T.T. Wu, Phys. Rev. Lett. 31, 1409 (1973).

2) T.T. Wu, B.M. McCoy, C.A. Tracy, E. Barouch, Phys. Rev. B 13, 316 (1976).

[6] 1) M. Jimbo, T. Miwa, Y. Mori and M. Sato, Physica D., 1D, No. 1, 80 (1980).

2) M. Jimbo, T. Miwa, Proc. of Japan Academy, 56 (A), No. 9, 405 (1980).

[7] J.W. Miles, On Cumulative Nonlinearity for a Weakly Dispersive Gravity Wave Proc. R. Soc. Lond. A. 361, 277 (1978).

[8] The work done in Soviet Union up to 1976 is reviewed by the editor of Diff. Urav. in the following two review articles (where he quotes 127 references):

1) N.P. Erugin, Diff. Urav., 12, No. 3, 387 (1976).

2) N.P. Erugin, Diff. Urav., 12, No. 4, 579 (1976).

[9] V.I. Gromak, Diff. Urav., 14, No. 12, 2131 (1978).

[10] N.P. Erugin, Dokl. Akad. Nark. BSSR, 2, No. 4 (1958).

[11] N.A. Lukashevich, Diff. Urav., 3, No. 11, 994 (1967).

[12] N.A. Lukashevich, Diff. Urav., 3, No. 5, 395 (1967).

[13] N.A. Lukashevich, Diff. Urav., 4, No. 8, 732 (1968).

[14] N.A. Lukashevich, Diff. Urav., 7, No. 1, 1124 (1971).

[15] V.I. Gromak, Diff. Urav., 11, No. 2, 373 (1975).

[16] N.A. Lukashevich, Author's Abstract of Doctoral Dissertation (in Russian), Kiev (1971).

[17] V.I. Gromak, Diff. Urav., 12, No. 4, 740 (1976).

[18] N.A. Lukashevich, A.I. Yablonskii, Diff. Urav., 3, No. 3, 264 (1967).

[19] A.S. Fokas and Y.C. Yortsos, Lett. Al Nuovo Cimento (1981).

[20] M.J. Ablowitz, H. Segur, Phys. Rev. Lett. 38, 1103 (1977).

[21] M.J. Ablowitz, A. Ramani and H. Segur, i) Lett. Al. Nuovo Cimento 23, 333 (1978). ii) J. Math. Phys. 21, 715 (1980). iii) J. Math. Phys. 21, 1006 (1980).

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

[23] G.B. Whitham, Private communication.

[24] Given a continuous group of transformations which is admitted by a given equation, there is a standard procedure for finding the solutions of the equation invariant under this group (the so called invariant solutions). See for example 1) L.V. Ovsjannikov, "Group properties of Differential Equations", (translated by G.W. Bluman), original publication in 1962. 2) G.W. Bluman and J.D. Cole, "Similarity Methods for Differential Equations," Springer, New York, 1974.

[25] A.S. Fokas and M.J. Ablowitz, On a linearization of the Korteweg-deVries and Painlevé II equations, Preprint 1981.

[26] H. Flaschka and A.C. Newell, Commun. Math. Phys. 76, 65 (1980).

[27] 1) M. Sato, T. Miwa, M. Jimbo, A series of papers entitled "Holonomic Quantum Fields": i) Publ. RIMS, Kyoto Univ., 14, 223 (1977). ii) ibid, 15, 201 (1979). iii) ibid, 15, 871 (1979).

2) K. Ueno, Monodromy preserving deformations of linear differential equations with irregular singular points, RIMS Preprint 301 (1979).

3) M. Jimbo, T. Miwa, Deformations of linear ODE, RIMS, Preprint 315 (1980).

4) M. Jimbo, T. Miwa, Monodromy preserving deformation of linear ODE's with rational coefficients, RIMS, Preprint, 327 (1980).

5) M. Jimbo, T. Miwa, Deformation of linear ODE's, RIMS, Preprint, 329 (1980).

[28] It is well known that exactly solvable PDE's possess infinitely many continuous symmetries. Actually there is strong indication that this is the defining feature of an exactly solvable equation. See for example, 1) A.S. Fokas, J. Math. Phys. 21, 1318 (1980). 2) A.S. Fokas, B. Fuchssteiner, Lett. Al Nuovo Cimento 28, 299 (1980). These symmetries in general involve derivatives, in contrast to Lie-point symmetries. It is quite interesting that the Painlevé equations posse s infinitely many discrete symmetries, involving derivatives (i.e. non Lie-point).

[29] It is well known that there exist equations of the form $(w')^n + \sum_{j=1}^{n} P_j(z,w)(w')^{n-j} = 0$, $n > 1$, n integer, which satisfy the Fuchs theorem concerning the absence of movable critical points. One may thus try to use appropriate transformations, related to these equations, instead of (2.6). We have not considered such transformations.

[30] M. Boiti and F. Pempinelli, Nuovo Cimento, 51B, 70 (1979).

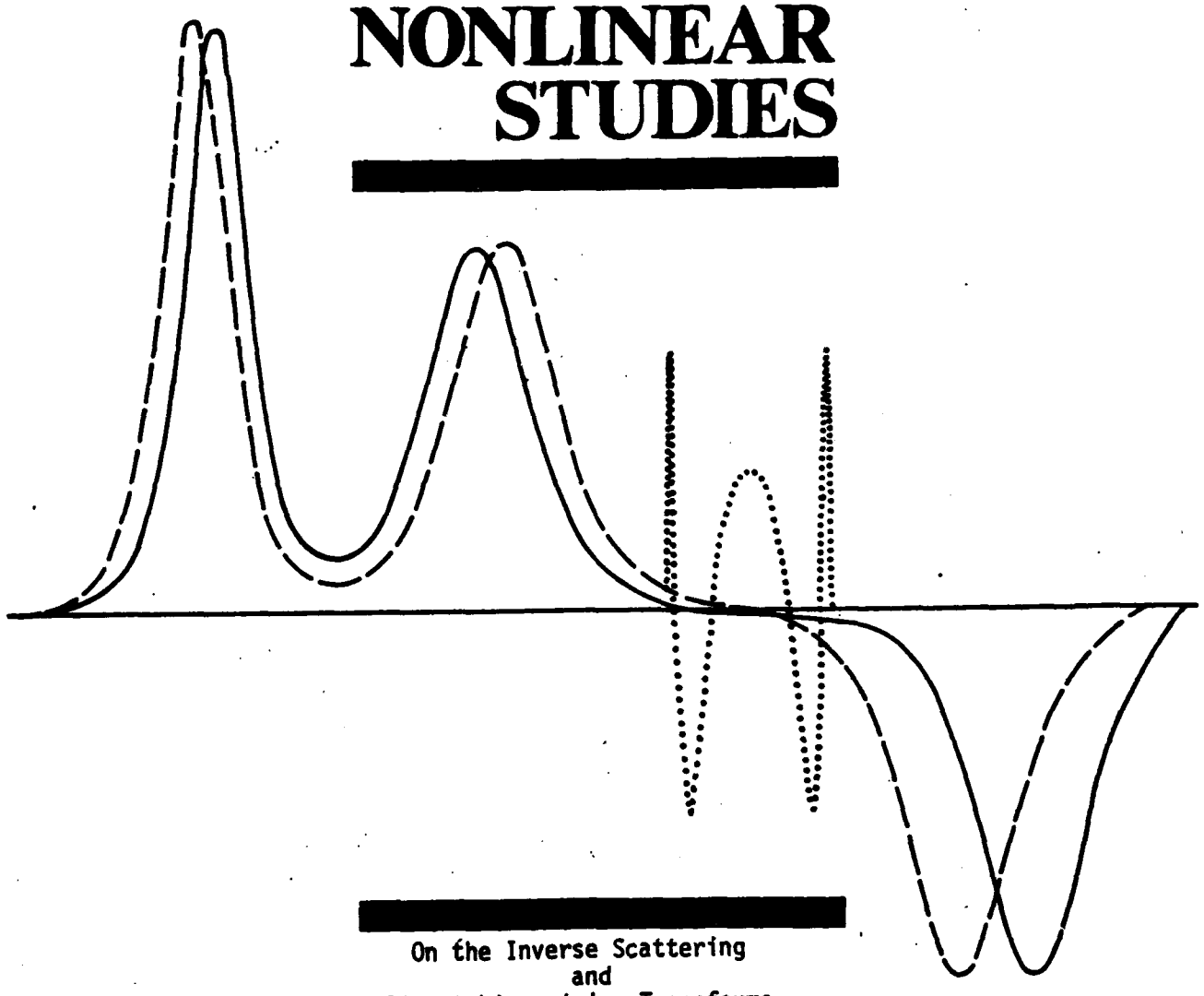
[31] H. Airault, Stud. in Appl. Math., 61, 31 (1979).

[32] If $\gamma \neq 0$, then taking $\gamma = 1$ is without loss of generality as one may always use a scaling transformation.

[33] N.A. Lukashevich, Diff. Urav., 1, No. 6, 561 (1965).

I.F.N.S. #9

INSTITUTE FOR NONLINEAR STUDIES



On the Inverse Scattering
and
Direct Linearizing Transforms
for the
Kadomtsev-Petviashvili Equation

by
A.S. Fokas and M.J. Ablowitz

March, 1982

Clarkson College of Technology
Potsdam, New York 13676

ON THE INVERSE SCATTERING AND DIRECT LINEARIZING TRANSFORMS FOR THE KADOMTSEV-PETVIASHVILI EQUATION

BY

A.S. FOKAS AND M.J. ABLOWITZ

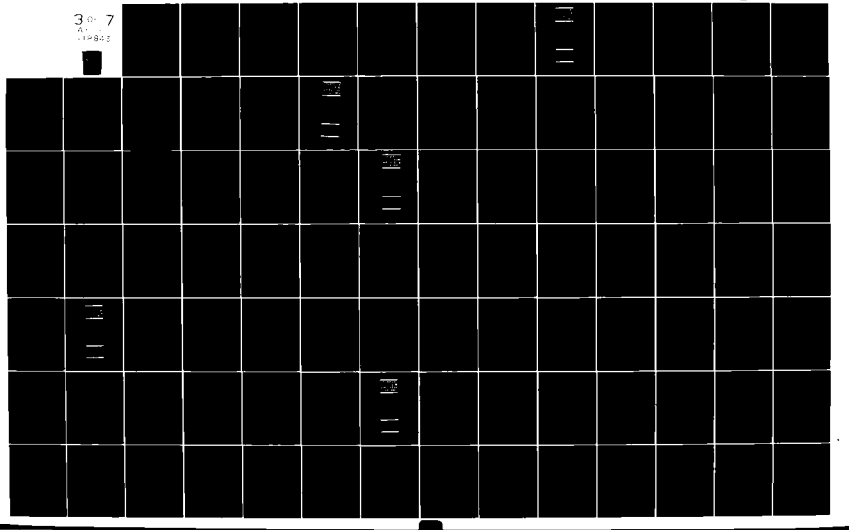
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
CLARKSON COLLEGE OF TECHNOLOGY
POTSDAM, NEW YORK 13676

Abstract

The existing formalism for solving the initial value problem associated with the Kadomtsev-Petviashvili equation, a physically significant two spatial dimensional analogue of the Korteweg-deVries equation, is both simplified and extended. The lump solutions, algebraically decaying solitons, are naturally incorporated and given a spectral characterization in the new scheme. In addition, a very general linearization of the Kadomtsev-Petviashvili equation is presented.

March, 1982

AD-A119 843 CLARKSON COLL OF TECHNOLOGY POTSDAM NY DEPT OF MATHE--ETC F/G 12/1
NONLINEAR WAVES.(U)
SEP 82 M J ABLOWITZ, D J KAUP, A C NEWELL N00014-76-C-0867
UNCLASSIFIED NL



The Kadomtsev-Petviashvili (KP) equation [1]

$$(u_t+6uu_x+u_{xxx})_x = 3u_{yy} \tag{1}$$

is a natural generalization of the Korteweg-deVries (KdV) equation to two space dimensions. The KP equation, like its descendant the celebrated KdV equation, arises generically in physical contexts. Namely, unidirectional, weak (quadratic) nonlinear disturbances perturbed from a two-spatial dimensional wave equation with weak balanced fourth order dispersion always yields KP [2]. As a result of both its physical significance as well as its mathematical importance, being a multidimensional "exactly solvable" equation, it has been investigated extensively. In particular: i) soliton solutions of KdV also solve KP but they are non decaying at infinity and are linearly unstable (for the signs chosen in (1)) [1]; ii) KP possesses a "Lax pair" [3] (see below); iii) a variety of explicit solutions of KP can be obtained (e.g. via the "direct approach" of [4]). iv) KP admits solitons, the so-called "lumps" [5] which decay algebraically (in both space dimensions) and which are not contained in the Gel'fand-Levitan-Marchenko (GLM) formulation.

Significant results associated with the initial value problem of the KP equation are contained in [6], [7] where for a restricted class of initial data, the initial value problem of KP is solved via a GLM equation. However: i) The above GLM equation is defined via a certain function $f(k,\ell)$. Relating $f(k,\ell)$ to a scattering matrix $s(k,\ell)$ is, as it is pointed out in [7], the main difficulty of the inverse problem associated with the KP. Such a relationship is established in [7] by making use of still another GLM-type equation. ii) Lumps are excluded in the above formulation (long time asymptotics were carried out in [8]).

In this letter we a) Simplify the above formalism by giving an explicit expression for $f(k,\ell)$. b) Extend the above scheme so that the lump solutions are both incorporated and given a spectral characterization. These results are relevant to the problem of solving the KP equation with initial data $u(x,y,o) = u_o(x,y)$, $u_o(x,y)$

given and going to zero rapidly enough as $(x^2+y^2)^{1/2} \to \pm\infty$ (we also assume $\int_{-\infty}^{\infty} u_0(x,y)dx = 0$[2]

The problem of characterizing solutions corresponding to initial data outside the above class is considered in part c) of this letter. In this context a very general linear integral equation is presented. This equation involves an arbitrary measure-contour; a special choice of this measure-contour yields the solutions characterized in part b). The problem of relating these far more general solutions to initial values outside the range of the inverse scattering class is open.

The Lax pair of KP can be taken in the form [3]

$$i\mu_y+\mu_{xx}+2ik\mu_x+u\mu = 0, \tag{2}$$

$$\mu_t+4\mu_{xxx}+12ik\mu_{xx}-12k^2\mu_x+6u\mu_x+3iku\mu+3iu\mu+3u_x\mu-3i\left(\int_{-\infty}^{x} u_y dx'\right)\mu = 0. \tag{3}$$

Manakov [7] introduces μ^+,μ^- which are those solutions of (2) which also satisfy the following integral equations

$$\mu^{\pm}(x,y,t,k) = 1 + \hat{G}^{\pm}\mu(\xi,\eta,t)\exp[im(x-\xi)-im(m+2k)(y-\eta)]\mu^{\pm}(\xi,\eta,t,k), \quad (4)^{\pm}$$

where $\hat{G}^+ \doteq (i/2\pi)\left(-\int_y^{\infty} d\eta \int_0^{\infty} dm \int_{-\infty}^{\infty} d\xi + \int_{-\infty}^{y} d\eta \int_{-\infty}^{0} dm \int_{-\infty}^{\infty} d\xi\right)$ and $\hat{G}^- \doteq (i/2\pi)\left(-\int_y^{\infty} d\eta \int_{-\infty}^{0} dm \int_{-\infty}^{\infty} d\xi + \int_{-\infty}^{y} d\eta \int_0^{\infty} dm \int_{-\infty}^{\infty} d\xi\right)$.

It is clear that the kernel of equation $(4)^+$ is a (+) function with respect to k, i.e. it may be analytically continued in the upper half k-plane. Similarly the kernel of equation $(4)^-$ is a (-) function. Manakov [7] then postulates the equation (hereafter for simplicity of notation we shall use $\mu^{\pm}(k)$ to replace $\mu^{\pm}(x,y,t,k)$)

$$\mu^+(k) = \mu^-(k)+\int_{-\infty}^{\infty} f(k,\ell,t)\exp[\theta(\ell,k,x,y)]\mu^-(\ell)d\ell, \tag{5}$$

where $\theta(\ell,k,x,y)\equiv i(\ell-k)x-i(\ell^2-k^2)y$, and through a series of ingenious steps he relates $f(k,\ell,t)$ to standard scattering data through certain GLM-type equations in scattering space. Finally, he assumes that μ^+,μ^- are (+) and (-) functions respectively and then solves (5) through a GLM equation in physical space. In this context Segur [9] goes somewhat further by expressing $f(k,\ell,t)$ through Volterra integral equations. Here we give an explicit expression for $f(k,\ell,t)$ in terms of initial scattering functions.

Let us now discuss (a), (b). We introduce $N(x,y,t,k,p)$ which solves (2) in k and which also satisfies the integral equation

$$N(x,y,t,k,p) = \exp[\theta(k,p,x,y)]+\hat{G}^- u(\xi,\eta,t)\exp[im(x-\xi)-im(m+2k)(y-\eta)]N(\xi,\eta,t,k,p), \qquad (6)$$

where $\hat{G}^-$, θ are defined in $(4)^-$ and (5) respectively. In general N cannot be extended off real k,p unless k=p. By manipulating equations $(4)^{\pm}$ and by using the crucial relationship

$$N(x,y,t,k,p) = \mu^+(k)\exp[\theta(k,p,x,y)] - \hat{I}\mu^-(\ell)\exp[\theta(\ell,p,x,y)]f(k,\ell,t) \qquad (7)$$

where $\hat{I} = \int_k^{\infty} d\ell + \int_{-\infty}^{p} d\ell$ for $k>p$, $\hat{I} = \int_{-\infty}^{k} d\ell + \int_p^{\infty} d\ell$ for $k<p$, it can be shown that equation (5) is valid with f explicitly given by

$$f(k,\ell,t) = -\frac{1}{2\pi}\mathrm{sgn}(\ell-k)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(\xi,\eta,t)N(\xi,\eta,t,k,\ell)d\xi d\eta \qquad (8)$$

Equations $(4)^{\pm}$ are Fredholm equations of the second type which we assume to be regular. Let $\phi_j^+(x,y,t)$ and $\phi_j^-(x,y,t)$ denote their homogeneous solutions corresponding to eigenvalues k_j^+ and k_j^- respectively (we assume that there exists a finite number of such eigenvalues and that they are all simple). Then Fredholm theory implies that μ^+, μ^- admit the representations:

$$\mu^{\pm}(x,y,t,k) = 1 + \sum_{1}^{n} \frac{c_j^{\pm}(t)\phi_j^{\pm}(x,y,t)}{k-k_j^{\pm}(t)} + \tilde{\mu}^{\pm}(x,y,t,k), \qquad (9)^{\pm}$$

where $\mathrm{Im}k_j^+>0$, $\mathrm{Im}k_j^-<0$ and $\tilde{\mu}^+$, $\tilde{\mu}^-$ are (+) and (-) functions respectively with respect to k.

By splitting equation (5) into its (+) and (-) parts, say $(5)^+$, $(5)^-$ and by using $(5)^-$ together with $(9)^{\pm}$ one obtains (we note that it is easily shown that $c_j^{\pm}(t)=i$)

$$\mu^-(k) - i\sum_{\ell=1}^{n}\left(\frac{\phi_\ell^+}{k-k_\ell^+} + \frac{\phi_\ell^-}{k-k_\ell^-}\right) - \frac{1}{2\pi i}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f(k,\ell,t)\exp[\theta(\ell,k,x,y)]\mu^-(\ell)d\ell d\nu}{\nu-k+i\varepsilon} = 1. \quad (10)$$

If one assumes that there are no homogeneous solutions of $(4)^{\pm}$ then $\phi_j^{\pm} = 0$ and equation (10) defines a Fredholm equation for μ^-. Actually in this case by using $i\int_{-\infty}^{\infty}\exp[i\xi(k-\nu-i\varepsilon)]d\xi = \exp[ix(k-\nu)]/(k-\nu-i\varepsilon)$ multiplying equation (10) by $f(k,k')\exp[i(kx-k'x')-i(k^2-k'^2)y]$ and by integrating over $dkdk'$ one directly obtains the GLM equation given in [7]. However, in general $\phi_j^{\pm}\neq 0$. Then in order to solve (10) one needs some more information about the relation of $\phi_j^{\pm}$ and $\mu^{\pm}$. This information is as follows:

$$\mu^{\pm}(j) \mp \mu^{\pm} - \frac{i\phi_j^{\pm}}{k-k_j^{\pm}} = (x-2k_j^{\pm}y+\gamma_j^{\pm}(t))\phi_j^{\pm} \qquad (11)$$

Hence (10) and the equation for $\mu^+(k)$ analogous to (10) yield (as $k\to k_j^\pm$)

$$(x-2k_j^\pm y+\gamma_j^\pm(t))\phi_j^\pm - i\sum_{\ell=1}^{n}\left(\frac{\phi_\ell^+}{k_j^\pm-k_\ell^+}+\frac{\phi_\ell^-}{k_j^\pm-k_\ell^-}\right)-\frac{1}{2\pi i}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{f(k_j^\pm,\ell,t)\exp[\theta(\ell,k_j^\pm,x,y)]\mu^-(\ell)d\ell d\nu}{\nu-k_j^\pm}=1 \quad (12)^\pm$$

where $\sum$ means summation from $\ell=1$ to n unless any of the denominators vanishes. Equations (10) and $(12)^\pm$ completely characterize the inverse problem associated with (2) in terms of the "scattering data" $k_j^\pm(t)$, $\gamma_j^\pm(t)$, $j=1,\ldots n$, $f(k,\ell,t)$.

The evolution of the scattering data is very simple: $k_j^\pm$, are constant in time, $\gamma_j^\pm(t) = 12(k_j^\pm)^2t+\gamma_j^\pm(0)$, $f(k,\ell,t) = f(k,\ell,0)\exp[4i(\ell^3-k^3)t]$ Hence given $u(x,y,0)$ the scattering data need only be evaluated at $t=0$ and equations (10), $(12)^\pm$ yield $\mu^-(x,y,t,k)$, $\phi_j^\pm(x,y,t)$. Finally $u(x,y,t)$ is obtained from

$$u(x,y,t) = \frac{\partial}{\partial x}\left[2\sum_1^n(\phi_j^++\phi_j^-)+\frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(k,\ell,t)\exp[\theta(\ell,k,x,y)]\mu^-(\ell)d\ell dk\right] \quad (13)$$

(equation (13) can be obtained from (10) asymptotically as $k\to\infty$).

The lump solutions correspond to $f(k,\ell,0) = 0$ and are then characterized through the system of linear algebraic equations $(12)^\pm$.

We now discuss c). Let $\mu(x,y,t,k)$ be a solution of the linear integral equation

$$\mu(x,y,t,k) + i\int\!\!\int_L w(x,y,t,k,\ell,\nu)\mu(x,y,t,\ell)d\zeta(\ell,\nu) = v(x,y,t,k) \quad (14)$$

where the contour, measure $L, d\zeta(\ell,\nu)$ are essentially arbitrary, v is any solution of $F(k)v_y^\pm(\partial_y-i\partial_x^2+2k\partial_x)v = 0$, $A(k)v_t^\pm(\partial_t+4\partial_x^3+12ik\partial_x^2-12k^2\partial_x)v = 0$, and w is defined in terms of v through

$w=\exp[\theta]\int_{\alpha}^{\infty} d\xi v(\xi,y,t,k)\exp[i(k-\nu)\xi]+\exp[i(\ell-k)(x-\alpha)+4i(\ell^3-\nu^3)t]b$

where $\theta=\theta(\ell,k,\nu,x,y,t)\doteq i(\ell-k)x-i(\ell^2-\nu^2)y+4i(\ell^3-\nu^3)t$ and $b=b(y,t,k,\ell,\nu)$ is defined by

$b_y+i(\ell^2-k^2)b = \exp[\theta(\ell,\nu,\alpha,y)][iv_x(\alpha)-(k+\nu)v(\alpha)]$,

$b_t+4i(k^3-\nu^3)b = -4\exp[\theta(\ell,\nu,\alpha,y)][v_{xx}(\alpha)+i(\nu+2k)v_x(\alpha)-(k^2+\nu^2+\nu k)v(\alpha)]$, $v(\alpha)\doteq v(\alpha,y,t,k)$.

Then μ also solves both (2), (3)and u solves the KP where

$$u(x,y,t) = 2i\frac{\partial}{\partial x}\int_L\int \exp[\theta(\ell,\nu,\nu,x,y,t)]\mu(x,y,t,\ell)d\zeta(\ell,\nu) \qquad (15)$$

The proof, although tedious is in the spirit to that used in both [10] and [11] and is therefore omitted.

We conclude with the following: i) By manipulation the equations (10), $(12)^{\pm}$ can be shown to be a special case of (14) where $d\zeta(\ell,\nu)$ is supported on all real ℓ,ν and $\ell=\nu=k_j^{\pm}$, $j=1,\ldots n$. ii) if $\partial_t = 0, \ell = \nu$ the linearization expressed by (14) reduces to the linearization of the Benjamin-Ono equation [10]. If $\partial_y = 0$ then (14) reduces to the linearization of the Korteweg-deVries given in [11]. iii) In this letter we have concentrated on the KP equation. However, it is evident that both equations (10), $(12)^{\pm}$ as well as equation (14) characterize potentials of the time dependent Schrödinger equation (2) as well.

We feel that the IST scheme employed here andin particular the relationship between eigenvalues associated with the scattering operator and lump-type solutions will be maintained in other multidimensional problems. Additional multidimensional examples are presently under investigation.

ACKNOWLEDGEMENTS:

This research was partially supported by the Air Force Office of Scientific Research, under Grant No. 78-3674-C and the Office of Naval Research under Grant No. N00014-76-C-0867. We appreciate useful discussions with H. Segur

References

[1] Kadomtsev, B.B. and Petviashvili, V.I., Sov. Phys. Docklady, 15, 539 (1970.

[2] Ablowitz, M.J. and Segur, H., J. Fluid Mech., 92, 691 (1979).

[3] Dryuma, V.S., Sov. Phys. JETP Lett., 19, 387 (1974).

[4] Zahkarov, V.E. and Shabat, A.B., Functional Analysis, 8 (3), 43 (1974).

[5] i) Bordag, L., Its, A., Manakov, S., Matveev, V. and Zakharov, V., Phys. Lett. 63A, 205 (1977). ii) Ablowitz, M.J. and Satsuma, J., J. Math. Phys., 19, 2180 (1978).

[6] Zakharov, V.E. and Manakov, S.V., SOV. Sci. Rev. Phys. Rev., 1 (8) 133 (1979).

[7] Manakov, S.V., Physica D, 3, 420 (1981).

[8] Manakov, S.V., Santini, P., Takhtajan, L., Phys. Lett., Vol. 75A, 451 (1980).

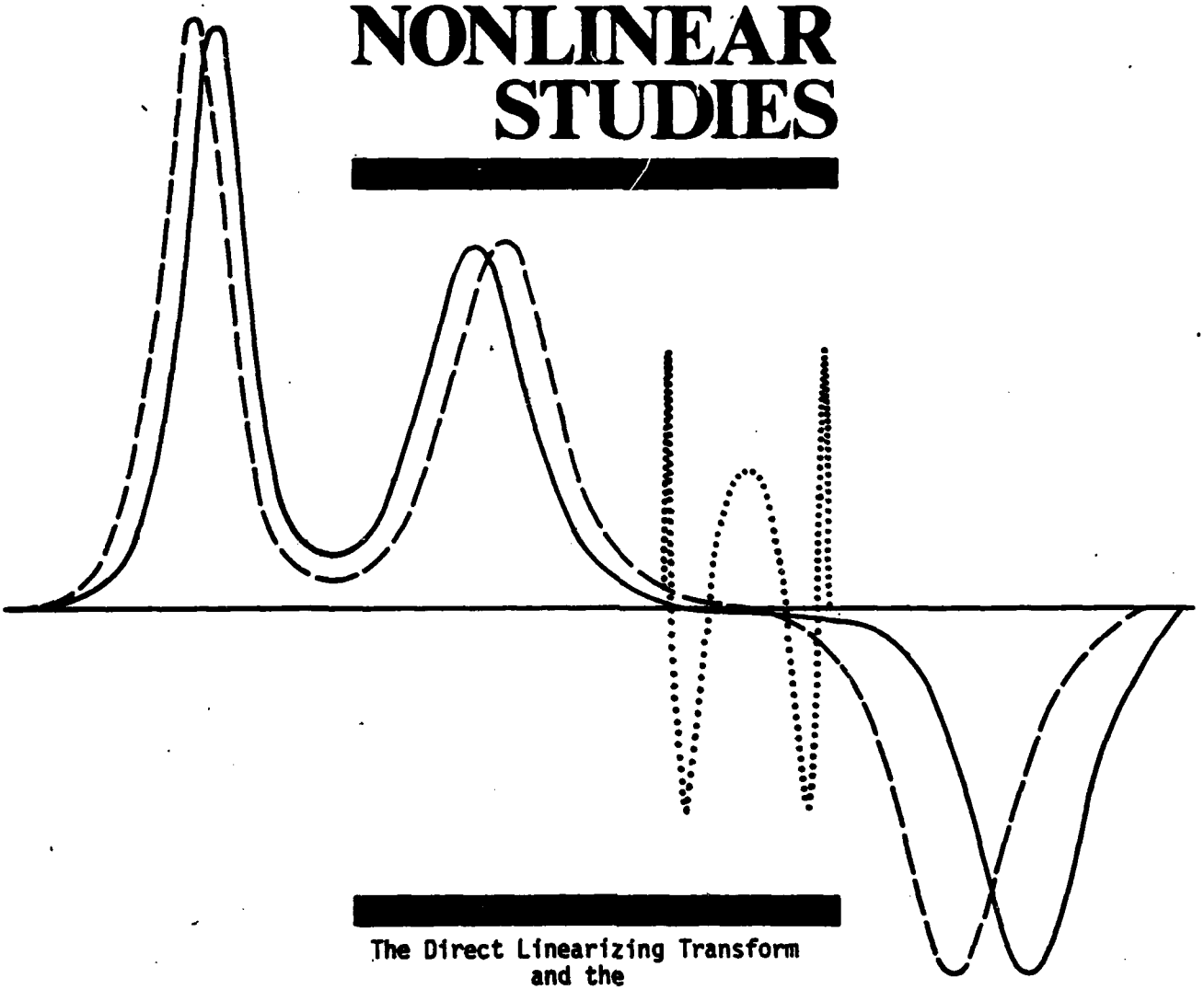
[9] Segur, H., Comments on IS for the Kadomtsev-Petviashvili equation, to appear in the proceedings of workshop, La Jolla, CA, December 7-9, 1981.

[10] Ablowitz, M.J., Fokas, A.S. and Anderson, R.L., The direct linearizing transform and the Benjamin-Ono equation (preprint, 1982).

[11] Fokas, A.S. and Ablowitz, M.J., Phys. Rev. Lett. 47, 1096 (1981).

I.F.N.S. #10

INSTITUTE FOR NONLINEAR STUDIES



The Direct Linearizing Transform
and the
Benjamin-Ono Equation

by
M.J. Ablowitz, A.S. Fokas and
R.L. Anderson†

†Permanent Address: University of Georgia, Department of Physics and Astronomy, Athens, Georgia 30602

Clarkson College of Technology
Potsdam, New York 13676

THE DIRECT LINEARIZING TRANSFORM

and the

BENJAMIN-ONO EQUATION

by

M. J. Ablowitz, A. S. Fokas and R. L. Anderson[†]

Department of Mathematics and Computer Science

Clarkson College of Technology

Potsdam, NY 13676

Abstract

The initial value problem associated with the Benjamin-Ono equation, a physically significant nonlinear singular-integro-differential equation, is linearized. The classical inverse scattering transform method is bypassed and an alternative method which we call the direct linearizing transform is introduced. This method applies to other nonlinear evolution equations, including multidimensional ones.

[†]Permanent address: University of Georgia, Department of Physics and Astronomy, Athens, Georgia 30602

January 1982

The Benjamin-Ono (BO) equation arises in the context of long internal gravity waves in a stratified fluid [1]. Recent discoveries of large amplitude internal waves in the ocean [2] anticipate the physical relevance of this equation to oceanographic applications. Moreover this equation arises in a variety of other circumstances (e.g. long waves in a stratified shear flow [3]).

Mathematically speaking it must be emphasized that the BO equation has distinctly new features, as will be brought out in the subsequent discussion. The BO equation can be taken in the form

$$u_t + 2uu_x + Hu_{xx} = 0; \quad Hv(x) \doteq \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{v(\xi)}{\xi - x} d\xi, \tag{1}$$

where Hv denotes the Hilbert transform of v and Cauchy principal value integrals are assumed if needed. It has been established that it possesses: a) n soliton solutions [4], [5]; b) Bäcklund transformations, infinitely many conservation laws and a novel "Lax pair" [6], [7]; c) two non-local operators which generate its infinitely many commuting symmetries and constants of motion in involution [8]. Its Lax pair is given by

$$i\phi_x^+ + \lambda(\phi^+ - \phi^-) = -u\phi^+ \tag{2}$$

$$i\phi_t^\pm - 2i\lambda\phi_x^\pm + \phi_{xx}^\pm - 2i[u]_x^\pm \phi^\pm = -\nu\phi^\pm; \quad [u]^\pm \doteq \pm \frac{u}{2} + \frac{1}{2i} Hu, \tag{3}^\pm$$

where $\phi^+(x,t;\lambda)$ (ϕ^-) is the limit of a function analytic in the upper (lower) half z-plane as $z \to x$ (z is the complex extension of x); similarly $[u]^+$, $[u]^-$ are analytic in the upper, lower half z-plane respectively. (λ is a constant and is interpreted as a spectral parameter, ν is an arbitrary constant).

In this letter we present a method of solution for the initial value problem of the BO equation, i.e. we solve (1) together with $u(x,o) = u_o(x) \to o$ sufficiently rapidly as $|x| \to \infty$. It has been well established that the inverse scattering transform (IST), as applied for example to the Korteweg-deVries equation [9], provides

powerful method for solving such initial value problems. The fundamental step of ıe IST consists of using the x-portion of the Lax pair to formulate and solve an inverse scattering problem". The t-portion of the Lax pair plays a far less significant ɔle. However instead of making use of the IST method here, we shall introduce an lternative method, referred to here as the direct linearizing transform method. It will e seen that in this method fundamental use of the t-portion of the Lax pair is required.

In this letter we concentrate on the (+) functions since with this knowledge esu]⁻ for the (-) functions can be obtained in a straightforward manner. In any ıse, here we assume that u is real and hence $[u]^- = -([u]^+)^*$. The method we propose ɔr linearizing (1) consists essentially of the following steps.

i) Use equation (2) to express $[u]^+$ in terms of ϕ^+, the "reflection coefficient" $\beta(\lambda,t)$, and "normalizing constants" $c_j(t)$. This step involves a detailed investigation ɔf equation (2). Namely; a) introduce left, right and discrete "Jost eigenfunctions"; ɔ) establish a "scattering relationship" between left and right eigenfunctions; c) examine the analyticity properties of ϕ^+ with respect to λ.

ii) Use equation $(3)^+$ to find how $\beta(\lambda,t)$ and $c_j(t)$ evolve in time, then substitute the expression obtained in a) in equation $(3)^+$ to obtain a nonlinear equation for ϕ^+. One then directly linearizes this equation.

iii) Use the above linear equation (which replaces the Gel'fand-Levitan-Marchenko equation) to solve the initial value problem associated with (1).

We first consider i). Equation (2) should be interpreted as a differential Riemann-Hilbert problem for the analytic functions $\phi^\pm(z,t,\lambda)$. Equation (2) describes the jump condition across the real axis x; it yields unique solutions for $\phi^\pm$ provided one imposes some boundary conditions as $z\to\infty$, say, in the upper half-plane. Here we assume that either $\phi^+(z,t,\lambda)\to 0$, or 1, as $z\to\infty$, $\mathrm{Im}(z)>0$. The solution of Riemann-Hilbert problems is usually given in terms of Fredholm integral equations [10].

This is also the case here: For arbitrary u(u→o, appropriately as $|x|\to\infty$) the solution of (2) with the above boundary conditions exists if λ is positive and real. It may be directly formulated via a Fredholm integral equation of the second type. This Fredholm equation can have homogeneous solutions for $\lambda = \lambda_j$, where λ_j are real and negative (we assume here that there exist only a finite number of λ_j and that they are simple). We say these "discrete eigenfunctions" correspond to bound states since $\phi^+ \to o$ as $|x|\to\infty$. (It turns out that these eigenfunctions are associated with solitons).

Specifically, consider only the (+) functions and let M, $\bar{M}$ denote left eigenfunctions, while $\bar{N}$, N denote right ones. They are specified by the following asymptotic behavior,

$$M\to 1,\ \bar{M}\to e^{i\lambda x} \text{ as } x\to-\infty;\ \bar{N}\to 1, N\to e^{i\lambda x} \text{ as } x\to+\infty. \tag{4}$$

and they satisfy

$$\begin{pmatrix} M(x,t,\lambda) \\ \bar{M}(x,t,\lambda) \end{pmatrix} = \begin{pmatrix} 1 \\ e^{i\lambda x} \end{pmatrix} + \int_{-\infty}^{\infty} G_+(x,y,\lambda)u(y,t) \begin{pmatrix} M(y,t,\lambda) \\ \bar{M}(y,t,\lambda) \end{pmatrix} dy, \tag{5}$$

$$\begin{pmatrix} \bar{N}(x,t,\lambda) \\ N(x,t,\lambda) \end{pmatrix} = \begin{pmatrix} 1 \\ e^{i\lambda x} \end{pmatrix} + \int_{-\infty}^{\infty} G_-(x,y,\lambda)u(y,t) \begin{pmatrix} \bar{N}(y,t,\lambda) \\ N(y,t,\lambda) \end{pmatrix} dy, \tag{6}$$

Furthermore, let ϕ_j denote the discrete (+) eigenfunctions, then

$$\phi_j(x,t) = \int_{-\infty}^{\infty} G(x,y,\lambda_j)u(y,t)\phi_j(y,t)dy;\ \lambda_j<o. \tag{7}$$

In the above expressions G_+, G_- are the (+) and (-) parts of the sectionally holomorphic function $G(x,y,\zeta) = 1/2\pi \int_0^\infty e^{i(x-y)p}/(p-\zeta)\, dp$, where ζ denotes the complex extension of λ, i.e. $G_\pm(x,y,\lambda) = \lim_{\varepsilon\to 0}(1/2\pi)\int_0^\infty e^{i(x-y)p}/(p-(\lambda\pm i\varepsilon))dp$. The line of discontinuity of G is given by the positive λ axis, hence using Plemelj's formulae [11] $G^+ - G^- = i\exp(i(x-y)\lambda)\theta(\lambda)$ where $\theta(\lambda)$ denotes the Heaviside function, i.e. $\theta(\lambda) = \{1, \lambda>0; 0, \lambda<0\}$. In particular for the discrete eigenfunctions $G_+ = G_- = G(x,y,\lambda_j)$.

By manipulating equations (5) and (6) one may establish the following "scattering relationship"

$$M = \bar{N} + B(\lambda,t)\theta(\lambda)N; \quad B(\lambda,t) = i\int_{-\infty}^{\infty} u(y,t)M(y,t,\lambda)e^{-i\lambda y}dy. \tag{8}$$

Equations (5), (6), (8) can also be obtained from the corresponding equations associated with the intermediate long wave equation [12] in the appropriate limit.

The results of Fredholm theory imply that M is a (+) function in the ζ-plane ((+) with respect to the positive real axis) except for possible poles. These poles correspond to homogeneous solutions of equation (5). Hence

$$M(x,t,\lambda) = 1 + \sum_1^n \frac{c_j(t)\phi_j(x,t)}{\lambda-\lambda_j(t)} + \mathcal{M}_+(x,t,\lambda). \tag{9}$$

Similarly

$$\bar{N}(x,t,\lambda) = 1 + \sum_1^n \frac{\bar{c}_j(t)\phi_j(x,t)}{\lambda-\lambda_j(t)} + \mathcal{N}_-(x,t,\lambda). \tag{10}$$

Then equation (8) implies that $\bar{c}_j = c_j$. Useful information about u is obtained by taking the (+) part of (8) and using (9), (10). One then obtains

$$[u]^+ = \frac{1}{2\pi i}\int_0^\infty B(\lambda,t)N(x,t,\lambda)d\lambda - \sum_1^n c_j(t)\phi_j(x,t). \tag{11}$$

At this point let us make some remarks about the inverse problem associated with (8). In order to establish that equation (8) defines a Riemann-Hilbert problem one needs to relate N to $\bar{N}$. For the Korteweg-deVries equation for example, $N(\lambda) = \bar{N}(-\lambda)$ and then the equation analogous to (8) defines a solvable Riemann-Hilbert problem (the Gel'fand-Levitan- Marchenko equation is one way to solve this Riemann-Hilbert problem). However, for the BO equation the relationship is somewhat more complicated. Specifically we find $\partial(Ne^{-i\lambda x})/\partial\lambda = F(\lambda,t)e^{-i\lambda x}\bar{N}$, where $F(\lambda,t) = -\int_{-\infty}^{\infty} uNdy/(2\pi\lambda)$. With this additional knowledge the IST can be implemented, and the results herein verified. This will be discussed in a future publication.

Let us now consider (ii). Using (8), (9) in $(3)^+$ one easily establishes that $c_j(t)$, $\lambda_j(t)$ are independent of time and that $\beta(\lambda,t) = \beta_0(\lambda)\exp(i\lambda^2 t)$. Using (11) in $(3)^+$ one obtains

$$\hat{N}_t - 2\lambda\hat{N}_x - i\hat{N}_{xx} - 2\left[\frac{1}{2\pi i}\int_0^{\infty}\beta_0(\ell)\hat{N}(\ell)d\ell - \sum_1^n c_k\Phi_k\right]_x\hat{N} = 0;\ \hat{N} = Ne^{i\lambda^2 t} \quad (12)$$

$$\Phi_{jt} - 2\lambda_j\Phi_{jx} - i\Phi_{jxx} - 2\left[\frac{1}{2\pi i}\int_0^{\infty}\beta_0(\ell)\hat{N}(\ell)d\ell - \sum_1^n c_k\Phi_k\right]_x\Phi_j = 0. \quad (13)$$

The above equations can be written in the concise form

$$\psi_t - 2\lambda\psi_x - i\psi_{xx} + 2\left[\int_{-\infty}^{\infty}\psi(\ell)d\rho(\ell)\right]_x\psi = 0 \quad (14)$$

where $\psi = \hat{N}$ for $\lambda>0$, $\psi = \Phi_j$ for $\lambda = \lambda_j$, $d\rho(\ell) = \left[\sum_1^n c_j\delta(\ell-\lambda_j) - (1/2\pi i)\beta_0(\ell)\theta(\ell)\right]d\ell$. Before linearizing equation (14) to its linear part let us first comment on the solitons: Let $\beta_0 = 0$ then (13) reduces to a system of coupled nonlinear PDE's. If one assumes

at $\Phi_j(x,t) = \sum_1^n c_{kj}(t)/(x-x_k(t))$ then the Calogero-Moser system [13]:

$\frac{x_k}{2} = 8 \sum_{\ell=1}^{n} (x_k-x)_\ell^{-3}$, k=1,...n follows. Hence, using (11) (and its complex conjugate) ie recovers the well known n-soliton solution of the BO equation. In any event e may linearize the above system of PDE's explicitly. This is a particular case of ne following result.

Let $\psi(x,t,\lambda)$ be a solution of

$$\psi(x,t,\lambda) = V(x,t,\lambda) - i \int_{-\infty}^{\infty} \psi(x,t,\ell)W(x,t,\lambda,\ell)d\rho(\ell), \tag{15}$$

whi V and W are defined by

$$L(\lambda)\, V \doteq V_t - 2\lambda V_x - iV_{xx} = 0, \tag{16}$$

$$W = e^{i(\lambda-\ell)x} \int_a^x V(\xi,t,\lambda)e^{-i(\lambda-\ell)\xi}d\xi + B(t,\lambda,\ell)e^{i(\lambda-\ell)(x-a)}, \tag{17}$$

$$B_t + i(\ell^2-\lambda^2)B = (\lambda+\ell)V(a,t,\lambda) + iV_x(a,t,\lambda). \tag{18}$$

T ψ also solves (14). The method only involves operating on (15) with $L(\lambda)$. The procedure is then similar in spirit to that presented in [14].

Let us now consider iii). At t=0 u(x,o) is given, hence $N(x,o,\lambda)$ and $\Phi_j(x,o)$ are determined from equations (6) and (7). c_j and $\beta_o(\lambda) = \beta(\lambda,o)$ are determined from equations (8) and (9). Then at t = 0 equation (15) defines a Volterra equation for $V(x,o,\lambda)$:

$$V(x,o,\lambda) - i \int_a^x V(\xi,o,\lambda)e^{i\lambda(x-\xi)}F(x,\xi)d\xi = \tilde{\psi}(x,o,\lambda), \tag{19}$$

where $F(x,\xi) = \int_{-\infty}^{\infty}\psi(x,o,\ell)e^{-i\ell(x-\xi)}d\rho(\ell)$ and

$\tilde{\psi}(x,o,\lambda) = \psi(x,o,\lambda) - i\int_{-\infty} B(o,\lambda,\ell)e^{i(\lambda-\ell)(x-a)}\psi(x,o,\ell)d\rho(\ell)$. Having obtained $v(x,o,\lambda)$, equation (16) yields $V(x,t,\lambda)$. Hence equation (15) (a Fredholm equation) defines $\psi(x,t,\lambda)$ and finally equation (11) defines $[u]^+$.

We conclude with the following remarks: a) In the case of pure solitons ($\beta=0$) equation (15) reduces to an explicit transformation for relating ϕ_j to V_j. The n soliton solutions correspond to $V_j(x,o)$ = constant. b) The generic behavior of N,N^-, β_o as $\lambda\to o$ is given by $N\sim 2\pi/(c\ell n\lambda)$ $N^-\sim -2\pi[u]^-/(c\lambda\ell n\lambda)$, $\beta_o\sim 2\pi i/\ell n\lambda$ where $c = \int_{-\infty}^{\infty} u(x,t)dx (c\neq o)$. c) It may be established that $\int_{-\infty}^{\infty} uMdx$ is constant in time which then yields the conserved quantities of the BO equation. d) The implementation of IST allows explicit representations of $V(x,t,\lambda)$, $B(t,\lambda,\ell)$ in terms of initial data.

Finally we remark that the direct linearizing transform method applies to other important nonlinear evolution equations as well, including multidimensional problems. We shall discuss this in future communications.

ACKNOWLEDGMENTS:

This research was partially supported by the Air Force Office of Scientific Research, under Grant No. 78-3674-C and the Office of Naval Research under Grant No. N00014-76-C-0867. We appreciate useful discussions with H. Segur.

REFERENCES

i) T. B. Benjamin, J. Fluid Mech. 29, 559 (1967).
ii) H. Ono, J. Phys. Soc. Japan, 39, 1082 (1975).

i) A.R. Osborn and T.L. Burch, Scientific American 208, 451 (1980).
ii) "Great Under Sea Waves May be Solitons", Physics Today, Nov. 1980.

D.J. Benney (private communication).

K.M. Case, Proc. Nat. Acad. Sc. U.S.A., 75, 3563 (1978); 76, 1 (1979).

H. Chen, Y.C. Lee, N.R. Pereira, Phys. Fluids 22, 187 (1979).

A. Nakamura, J. Phys. Soc. Jap. 47, 1701 (1979).

T.L. Boch, M.D. Kruskal, Phys. Lett. 74A, 173 (1979).

A.S. Fokas, B. Fuchssteiner, Phys. Lett. 86A, 341 (1981).

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

F.D. Gakhov, in "Boundary Value Problems", edited by I.N. Sneddon (Pergamon, New York, 1966), vol. 85.

N.I. Muskheilishvili, "Singular Integral Equations", ed. by J. Radok (Noordhoff, Grouingen, 1953).

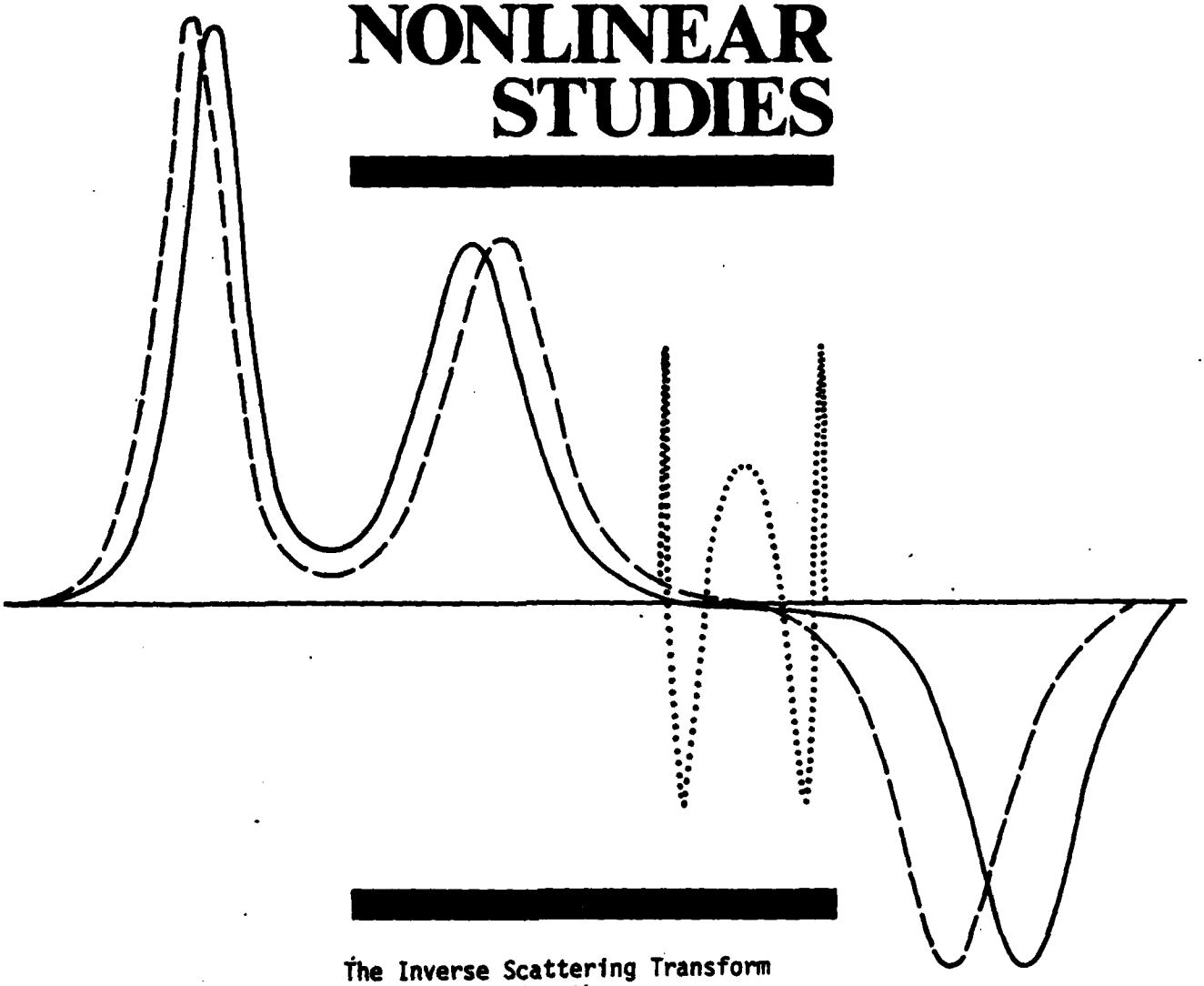
i) Y. Kodama, J. Satsuma, M.J. Ablowitz, Phys. Rev. Lett. 46, 687 (1981).
ii) Y. Kodama, M.J. Ablowitz, J. Satsuma, "Direct and Inverse Scattering Problems of the Nonlinear Intermediate Long Wave Equation", to appear in J. Math. Phys.

i) F. Calogero, J. Math. Phys. 12, 419 (1971).
ii) J. Moser, Adv. in Math. 16, 197 (1975).

A.S. Fokas and M.J. Ablowitz, Phys. Rev. Lett. 47, 1096 (1981).

I.F.N.S. #11

INSTITUTE FOR NONLINEAR STUDIES



The Inverse Scattering Transform
for the
Benjamin-Ono Equation - A Pivot To
Multidimensional Problems

by

A.S. Fokas and M.J. Ablowitz

Clarkson College of Technology
Potsdam, New York 13676

THE INVERSE SCATTERING TRANSFORM FOR THE BENJAMIN-ONO EQUATION - A PIVOT TO MULTIDIMENSIONAL PROBLEMS

by

A. S. Fokas and M. J. Ablowitz
Department of Mathematics and Computer Science
Clarkson College of Technology
Potsdam, New York 13676

Abstract

The initial value problem associated with the Benjamin-Ono equation is linearized by a suitable extension of the Inverse Scattering Transform. Essential is the formulation and solution of an associated nonlocal Riemann-Hilbert problem in terms of initial scattering data. Solitons are given a definitive spectral characterization. Pure soliton solutions are obtained by solving a linear algebraic system whose coefficients depend linearly on x,t.

April 1982

1. Introduction

The Benjamin-Ono equation arises in the context of long internal gravity waves in a stratified fluid [1] as well as in a variety of other physical circumstances, e.g. long waves in a stratified shear flow [2]. Mathematically speaking, its main importance, in our opinion, is due to the fact that it provides a conceptual bridge between the inverse scattering transforms (IST) in one and multidimensions. This startling aspect of the BO equation will be brought out in the subsequent discussion.

The BO equation can be taken in the form

$$u_t + 2uu_x + Hu_{xx} = 0;\ Hv(x) \doteq \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{v(\xi)}{\xi - x} d\xi \qquad (1)$$

where Hv denotes the Hilbert transform of v and Cauchy principal value integrals are assumed if needed. It has been established that it possesses: n soliton solutions [3], [4]; Bäcklund transformations and a novel "Lax pair" [5], [6]; infinitely many commuting symmetries and constants of motion in involution [6], [7]. Recently [8] the initial value problem associated with (1) has been linearized via what we refer to as the direct linearizing transform. This paper is a sequel to [8], though it can be read independently. Specifically in [8] u(x,t) is obtained via a Fredholm integral equation whose forcing and kernel are uniquely determined through a certain scattering function $v(x,t,\lambda)$ (λ is the spectral parameter of the associated linear scattering equation, see (2)). However, the determination of $v(x,t,\lambda)$ from initial data u(x,0) requires solving a Volterra integral equation as well as a linear partial differential equation. The above complication arises from the fact that previously we did not have enough analytic information about the Jost eigenfunctions of the associated linear eigenvalue equation in order to implement the IST.

In this paper we present an IST scheme for solving the BO equation with $u(x,0) = u_0(x) \to 0$ sufficiently rapidly as $|x| \to \infty$. We reduce (1) to a linear integral equation depending only on initial scattering data.

We note that this IST approach has two important features:
i) Not all the underlying Jost eigenfunctions may be extended off the line Re λ. Relating such eigenfunctions to those with analytic properties off the line Re λ is one of the main difficulties encountered in this IST scheme. ii) The underlying Jost eigenfunctions are characterized through Fredholm integral equations (as opposed to the standard treatment of the Korteweg-deVries equation, for example, where the corresponding eigenfunctions satisfy Volterra integral equations). These integral equations can have homogeneous solutions for certain values of λ, say λ_j, j=1,2,...n, and they give rise to solitons. Apparently both of the above features are generic to multidimensional problems [9].

The advantage of the IST approach presented here over the direct linearizing transform (DLT) method presented in [8], is that the IST essentially provides a closed form solution of the above mentioned scattering function $v(x,t,\lambda)$ explicitly in terms of initial data (see (28)). The precise relationship between IST and DLT as well various details of this IST scheme will be presented in the future.

2. The Direct and Inverse Scattering Problems

The "Lax pair" for the BO equation can be taken in the form

$$i\phi^+_x + \lambda(\phi^+ - \phi^-) = -u\phi^+, \tag{2}$$

$$i\phi^\pm_t - 2i\lambda\phi^\pm_x + \phi^\pm_{xx} - 2i[u]^\pm_x\phi^\pm = -\nu\phi^\pm, \tag{3}^\pm$$

$$[u]^\pm = \pm\frac{u}{2} + \frac{1}{2i}Hu \tag{4}^\pm$$

where λ is a constant and is interpreted as a spectral parameter, ν is an arbitrary

constant and $[u]^+$, $[u]^-$ are the boundary values of functions analytic in the upper, lower half z-planes respectively (z is the complex extension of x).

We remind the reader that any function u(x) defined on a closed contour L (assumed integrable and, say, Hölder on L [10])can be uniquely decomposed into the form $u = [u]^+ - [u]^-$, where $[u]^+$ and $[u]^-$ are analytic inside and outside the contour L respectively. The problem of finding $[u]^\pm$ (assuming that $[u]^-(z) \to 0$ as $z \to \infty$) is the simplest possible case of the so called Riemann-Hilbert problem [11]. In the case that L is the real axis the unique solution of this problem is given by equations(4).

2.1 "Left" and "Right" eigenfunctions and the scattering equation

We now analyze the linear eigenvalue problem associated with the BO equation, namely equation (2). Equation (2) should be interpreted as a differential Riemann-Hilbert problem for the functions $\phi^\pm(z,t,\lambda)$ which are analytic in the upper (+) and lower (-) half z-planes. Equation (2) describes the jump condition across the real axis x; it yields unique solutions for $\phi^\pm$ provided one imposes some boundary conditions as $z \to \infty$, say in the upper half-plane. Here we assume that either $\phi^+(z,t,\lambda) \to 0$, or 1, as $z \to \infty$, $\mathrm{Im}(z) > 0$. The solution of such Riemann-Hilbert problems is usually given in terms of Fredholm integral equations [11]. This is also the case here. For arbitrary u($u \to 0$, appropriately as $|x| \to \infty$) the solution of (2) with the above boundary conditions exists if λ is positive and real. It may be directly formulated via a Fredholm integral equation of the second type. This Fredholm equation can have homogeneous solutions for $\lambda = \lambda_j$ where λ_j are real and negative (we assume here that there exist only a finite number of λ_j and that they are simple).

Specifically, we need only consider the (+) functions We let M,$\bar{M}$ denote "left" eigenfunctions while $\bar{N}$,N denote "right"ones. They are specified by the following asymptotic behavior,

$$M\to 1,\ \bar{M}\to e^{i\lambda x} \text{ as } x\to-\infty;\ \bar{N}\to 1,\ N\to e^{i\lambda x} \text{ as } x\to+\infty. \tag{5}$$

(Note that as $|x|\to\infty$ equation (2) yields $i\phi^+_x + \lambda\phi^+ - \lambda\phi^-$. Hence $i\phi^+_x + \lambda\phi^+ - \alpha$, or $\phi^+ - \alpha/\lambda + \beta e^{i\lambda x}$ where α,β are arbitrary constants). The eigenfunctions $M,\bar{M},\bar{N},N$ satisfy the following Fredholm integral equations:

$$\begin{pmatrix} M(x,t,\lambda) \\ \bar{M}(x,t,\lambda) \end{pmatrix} = \begin{pmatrix} 1 \\ e^{i\lambda x} \end{pmatrix} + \int_{-\infty}^{\infty} G_+(x,y,\lambda)u(y,t) \begin{pmatrix} M(y,t,\lambda) \\ \bar{M}(y,t,\lambda) \end{pmatrix} dy, \tag{6}$$

$$\begin{pmatrix} \bar{N}(x,t,\lambda) \\ N(x,t,\lambda) \end{pmatrix} = \begin{pmatrix} 1 \\ e^{i\lambda x} \end{pmatrix} + \int_{-\infty}^{\infty} G_-(x,y,\lambda)u(y,t) \begin{pmatrix} \bar{N}(y,t,\lambda) \\ N(y,t,\lambda) \end{pmatrix} dy, \tag{7}$$

where G_+, G_- are the (+) and (-) parts of the sectionally holomorphic function

$$G(x,y,\zeta) = \frac{1}{2\pi}\int_0^{\infty} \frac{e^{i(x-y)p}}{p-\zeta}dp, \tag{8}$$

where ζ denotes the complex extension of λ, i.e.

$$G_\pm(x,y,\lambda) = \lim_{\epsilon\to 0} \frac{1}{2\pi}\int_0^{\infty} \frac{e^{i(x-y)p}}{p-(\lambda\pm i\epsilon)}dp, \quad \epsilon>0. \tag{9}$$

Equations (6), (7) can also be obtained from the corresponding equations associated with the intermediate long wave equation [12] in the appropriate limit. We note that in $G_\pm$, the (+), (-) are now defined with respect to the semi-real axis of the complex ζ-plane. We distinguish this from $\phi^\pm$ (where the (+), (-) are defined with respect to the real axis of the complex z-plane) by now using subscripts as opposed to the superscripts used before.

Let us indicate how the equation satisfied, say, by N can be derived: By splitting (2) into its (+) and (-) part and by using that $N \sim e^{i\lambda x}$ as $x \to \infty$ it follows that

$$N_x - i\lambda N = i[uN]^+, \tag{10}$$

or

$$(Ne^{-i\lambda x})_x = \frac{e^{-i\lambda x}}{2\pi}\int_{-\infty}^{\infty} \frac{u(y)N(y,\lambda)dy}{y-(x+i\varepsilon)} = \frac{1}{2\pi}\int_{-\infty}^{\infty} dy \int_0^{\infty} dp\, e^{ix(p-\lambda+i\varepsilon)-iyp} u(y)N(y,\lambda) \tag{11}$$

where we have used $1/(y-x-i\varepsilon) = i\int_0^{\infty}\exp[-i(y-x-i\varepsilon)p]dp$ and the fact that $(Ne^{-i\lambda x})_x \to 0$ as $x \to \infty$. Integrating (11) one finds

$$N(x,\lambda) = e^{i\lambda x} + \int_{-\infty}^{\infty} \left(\frac{1}{2\pi}\int_0^{\infty} \frac{e^{i(x-y)p}}{p-(\lambda-i\varepsilon)}dp\right) u(y)N(y,\lambda)dy,$$

i.e. equation (7b)(where we have suppressed the time dependence).

Let ϕ_j denote the discrete (+) eigenfunctions, then

$$\phi_j(x,t) = \int_{-\infty}^{\infty} G(x,y,\lambda_j)u(y,t)\phi_j(y,t)dy;\ \lambda_j<0. \tag{12}$$

Recall that the Plemelj's formulae [10] for the sectionally holomorphic function $G(x,y,\zeta)$ imply that

$$G_+(x,y,\lambda) - G_-(x,y,\lambda) = i\theta(\lambda)e^{i(x-y)\lambda},\ \theta(\lambda) = \begin{cases} 1, & \lambda>0, \\ 0, & \lambda<0. \end{cases} \tag{13}$$

In particualr for the discrete eigenfunctions $G_+ = G_- = G(x,y,\lambda_j)$.

By manipulating equations (6a) and (7) one may establish the following "scattering equation"

$$M = \bar{N} + \beta(\lambda,t)\theta(\lambda)N;\ \beta(\lambda,t) = i\int_{-\infty}^{\infty} u(y,t)M(y,t,\lambda)e^{-i\lambda y}dy \qquad (14)$$

(for the derivation of (14) see [13]). The solution of the "inverse problem" consists essentially of solving equation (14). In this respect knowledge of the analytic properties of $M,\bar{N},N$ is indispensable.

2.2 Analytic Properties of the Eigenfunctions

The kernel of equation (6a) is a (+) function with respect to λ and the forcing of (6a) is 1. Hence the results of Fredholm theory imply that M is a (+) function in the ζ- plane, except for possible poles. These poles correspond to homogeneous solutions of equation (6 a). Hence

$$M(x,t,\lambda) = 1 + \sum_{1}^{n} \frac{c_j(t)\phi_j(x,t)}{\lambda-\lambda_j(t)} + M_+(x,t,\lambda) \qquad (15)$$

where $M_+(x,t,\lambda)$ is a (+) function in λ and $M_+ \to 0$ as $\lambda \to \infty$. Similarly

$$\bar{N}(x,t,\lambda) = 1 + \sum_{1}^{n} \frac{\bar{c}_j(t)\phi_j(x,t)}{\lambda-\lambda_j(t)} + \bar{N}_-(x,t,\lambda) \qquad (16)$$

where $\bar{N}_-(x,t,\lambda)$ is a (-) function in λ and $\bar{N}_- \to 0$ as $\lambda \to \infty$. It is clear that because of the $\exp(i\lambda x)$, $\bar{M}$, N can not be analytically continued off the line Re λ.

2.3 A relationship between N and $\bar{N}$

The representations (15), (16), together with knowledge of $\beta(\lambda,t)$ are still inadequate in order to solve equation (14). To view (14) as a Riemann-Hilbert type problem one needs some further relationship between N and $\bar{N}$. This is as follows

$$\frac{\partial}{\partial\lambda}(N(x,t,\lambda)e^{-i\lambda x}) = f(\lambda,t)e^{-i\lambda x}\bar{N}(x,t,\lambda);\ f(\lambda,t)\doteq -\frac{1}{2\pi\lambda}\int_{-\infty}^{\infty} u(y,t)N(y,t,\lambda)dy \qquad (17)$$

To derive this result differentiate equation (9) with respect to λ and then integrate by parts with respect to p to obtain

$$\frac{\partial}{\partial\lambda}G_{\pm}(x,y,\lambda) = -\frac{1}{2\pi\lambda} + i(x-y)G_{\pm}(x,y,\lambda). \qquad (18)$$

Also equation (7b) implies that

$$(Ne^{-i\lambda x})_{\lambda} = \int_{-\infty}^{\infty}(G_{-}e^{-i\lambda(x-y)})_{\lambda}u(y)N(y,\lambda)e^{-i\lambda y}dy + \int_{-\infty}^{\infty} G_{-}e^{-i\lambda(x-y)}u(y)(N(y,\lambda)e^{-i\lambda y})_{\lambda}dy.$$

Hence, using (18) it follows that

$$(Ne^{-i\lambda x})_{\lambda} = e^{-i\lambda x}f(\lambda) + \int_{-\infty}^{\infty} G_{-}e^{-i\lambda(x-y)}u(y)(Ne^{-i\lambda y})_{\lambda}dy,$$

where $f(\lambda)$ is defined by (17) (and we have again supressed the time dependence). Comparing the above equation with equation (7a) (multiplied by $\exp(-i\lambda x)$) equation (17) follows.

It turns out that ϕ_j can also be related to $\bar{N}$:

$$(x+\gamma_j(t))\phi_j(x,y) = \overset{(j)}{\bar{N}}(x,y,\lambda_j); \quad \overset{(j)}{\bar{N}}(x,y,\lambda) \mp \bar{N}(x,y,\lambda) - \frac{c_j(t)\phi_j(x,y)}{\lambda-\lambda_j} \tag{19}$$

Furthermore,

$$c_j(t) = \tilde{c}_j(t) = -i, \; j = 1,2, \ldots .n, \tag{20}$$

where ϕ_j have been normalized by $x\phi_j \to 1$ as $x \to \infty$.

The above results can be proven as follows. Equation (7b), analytically continued for $\lambda<0$, implies

$$\overset{(j)}{\bar{N}}(x,\lambda) - \int_{-\infty}^{\infty} G(x,y,\lambda)u(y)\overset{(j)}{\bar{N}}(y,\lambda)dy = 1 - \frac{c_j}{\lambda-\lambda_j}\left[\phi_j(x) - \int_{-\infty}^{\infty} G(x,y,\lambda)u(y)\phi_j(y)dy\right].$$

Hence, letting $\lambda \to \lambda_j$ and using (18) (with $\varepsilon=0$ and $\lambda<0$) it follows that

$$\overset{(j)}{\bar{N}}(x,\lambda_j) - \int_{-\infty}^{\infty} G(x,y,\lambda_j)u(y)\overset{(j)}{\bar{N}}(y,\lambda_j)dy = ic_j \int_{-\infty}^{\infty} \phi_j(y)u(y)(x-y)G(x,y,\lambda_j)dy + a_j \tag{21}$$

where $a_j = 1 - c_j\int_{-\infty}^{\infty}\phi_j(y)u(y)dy/(2\pi\lambda_j)$. The asymptotics of equation (21) as $x \to \infty$ implies (assuming $x\phi_j \to 1$ as $x \to \infty$) that

$$\lambda_j = \frac{1}{2\pi i}\int_{-\infty}^{\infty} u(y,t)\phi_j(y,t)dy. \tag{22}$$

Then if $c_j = -i$, it follows that $a_j = 0$ and a particular solution of (21) is given by $\bar{N}^{(j)} = x\phi_j$. Hence, Fredholm's alternative theorem implies equation (19).

2.4 Scattering Data and their evolution

The scattering data necessary to carry out the IST associated with equation (2) is given by $\lambda_j(t)$, $\gamma_j(t)$ $\beta(\lambda,t)$, and $f(\lambda,t)$. Equations $(3)^+$, (15), (19), (12), (17) imply that

$$\{\lambda_j = \text{constant},\ \gamma_j(t) = 2\gamma_j t + \gamma_j(0)\}_{j=1}^{n}, \beta(\lambda,t) = \beta(\lambda,o)e^{i\lambda^2 t},\ f(\lambda,t) = f(\lambda,o)e^{-i\lambda^2 t} \tag{23}$$

We conclude this section by stating some important generic asymptotic relations:

$$M\text{-}N \sim -\frac{2\pi}{c\ \ell n\lambda} \text{ as } \lambda\to 0;\ \beta(\lambda,t) \sim \frac{2\pi i}{\ell n\lambda} \text{ as } \lambda\to 0, \tag{24}$$

where $c \triangleq \int_{-\infty}^{\infty} u(y,t)dy$. The above formulae are valid for u real and $c \neq 0$.

The special case $c = 0$ is nongeneric and will not be considered here. Also

$$\bar{N} \sim 1 - [\underline{u}]^+ \text{ as } \lambda\to\infty \tag{25}$$

3. The Solution of the Inverse problem

The solution of the inverse problem associated with (2) is characterized by the following linear Fredholm equations:

$$N(x,t,\lambda) - \frac{1}{2\pi}\int_0^{\infty} w(x,t,\lambda,\ell)\beta(\ell,t)N(x,t,\ell)d\ell + \sum_1^n \phi_\ell(x,t)w(x,t,\lambda,\lambda_\ell) = v(x,t,\lambda) \tag{26}$$

$$(x+\gamma_j(t))\phi_j(x,t) - \frac{1}{2\pi i}\int_0^\infty \frac{\beta(\ell,t)N(x,t,\ell)}{\ell-\lambda_j}d\ell + i\sum_{\substack{\ell=1\\ \ell\neq j}}^{n}\frac{\phi_\ell}{\lambda_j-\lambda_\ell} = 1, \qquad (27)$$

where

$$v(x,t,\lambda) = \int_0^\lambda (f(\ell,t)e^{ix(\lambda-\ell)} + f_s(\ell)e^{ix\lambda})d\ell; f_s(\lambda) = \frac{1}{\lambda\ell n\lambda} \qquad (28a)$$

$$w(x,t,\lambda,\ell) = e^{i(\lambda-\ell)x}\int_\infty^x v(\xi,\lambda)e^{-i(\lambda-\ell)\xi}d\xi, \quad \ell>0 \qquad (28b)$$

$$w(x,t,\lambda,\lambda_j) = e^{i(\lambda-\ell)x}\int_\alpha^x v(\xi,\lambda)e^{-i(\lambda-\lambda_j)\xi}d\xi + e^{i(\lambda-\lambda_j)(x-\alpha)}\int_0^\lambda \left(\frac{f(\ell,t)e^{i\alpha(\lambda-\ell)}}{\lambda_j-\ell}\right) + \frac{f_s(\ell)}{\lambda_j}d\ell. \qquad (28c)$$

By manipulation equations (26), (27) can be written in the concise form

$$\psi(x,t,\lambda) + i\int_{-\infty}^{\infty}\psi(x,t,\ell)\hat{w}(x,t,\lambda,\ell)d\rho(\ell)$$

where $\psi=N$ for $\lambda>0$, $\psi=\phi_j$ for $\lambda=\lambda_j$, $d\rho(\ell) = -[\sum_1^n(i\delta(\ell-\lambda_j) + (1/2\pi i)\beta(\ell,t)]d\ell$ and $\hat{w}$ is appropriately chosen. Hence the connection with equation (15) of [8] is readily established.

Note that equations (26), (27) define $N(x,t,\lambda)$, $\phi_j(x,t)$ in terms of λ_j, $\gamma_j(t)$, $\beta(\lambda,t)$ and $f(\lambda,t)$. Hence, because of (23) $N(x,t,\lambda)$ $\phi_j(x,t)$ are defined in terms of $u(x,0)$. Having obtained $N(x,t,\lambda)$ and $\phi_j(x,t)$, $[u]^+(x,t)$ is determined by

$$[u]^+ = \frac{1}{2\pi i}\int_0^\infty \beta(\lambda,0)e^{i\lambda^2 t}N(x,t,\lambda)d\lambda + i\sum_1^n\phi_j(x,t) \qquad (29)$$

Finally, assuming u is real $u = [u]^+ + ([u]^+)^*$.

Before sketching the proof of the above results let us make some remarks: i) $f(\lambda,t) \sim -f_s(\lambda)$ as $\lambda \to 0$ (using (17) and (24)). Hence $f(\lambda,t)$ is not integrable near $\lambda=0$. However, the singularities in the integrand of equation (28a) cancel and hence $v(x,t,\lambda)$ is well defined as $\lambda \to 0$. ii) The solitons correspond to $\beta(\lambda,0) = 0$ and hence can be obtained via the system of linear algebraic equations (27). iii) If $\phi_j = 0$, $j=1,\ldots n$ equation (26) can be formally reduced to a Gelfand-Levitan-Marchenko equation: Multiplying equation (26) by $\beta(\lambda,t)\exp[i(x'-x)\lambda]/2\pi$ and integrating over $d\lambda$ it easily follows that

$$K(x,x';t) + \int_x^\infty K(x,\xi;t)\, F(\xi,x';t)d\xi = F(x,x';t), x'<x \tag{30}$$

where

$$F(x,x';t) \doteq \frac{1}{2\pi}\int_0^\infty \beta(\lambda,0)e^{i(x'-x)\lambda+i\lambda^2 t} v(x,t,\lambda)d\lambda,$$

$$K(x,x';t) \doteq \frac{1}{2\pi}\int_0^\infty \beta(\lambda,0)e^{i(x'-x)\lambda+i\lambda^2 t} N(x,t,\lambda)d\lambda. \tag{31}$$

Also (29) yields $[u]^+ = -iK(x,x;t)$

Let us now derive equations (26) through (29). Substituting (15), (16) (with $c_j = \bar{c}_j = -i$) in (14) and then considering the (-) part of (14) it follows that

$$N(x,t,\lambda) - \frac{1}{2\pi i}\int_0^\infty \frac{\beta(\ell,t)N(x,t,\ell)d\ell}{\ell-(\lambda-i\varepsilon)} + i\sum_1^n \frac{\phi_\ell(x,t)}{\lambda-\lambda_\ell} = 1 \tag{32}$$

Equation (29) follows from the asymptotics of (32) as $\lambda \to \infty$ and from equation (25). To obtain equation (26) use equation (32), equation (32) at $\lambda = 0$, equation (17) and

the formulae

$$\frac{e^{ix(\ell-\lambda)}}{\ell-(\lambda-i\varepsilon)} = i\int_{\infty}^{x} e^{i\xi(\ell-\lambda)}d\xi, \quad \frac{e^{ix(\lambda_j-\lambda)}}{\lambda_j-\lambda} = i\int_{\alpha}^{x} e^{i\xi(\lambda_j-\lambda)}d\xi + \frac{e^{i\alpha(\lambda_j-\lambda)}}{\lambda_j-\lambda} .$$

To obtain equation (27), analytically continue (32) for $\lambda<0$, take its limit as $\lambda\to\lambda_j$ and use (19).

ACKNOWLEDGEMENTS

This research was partially supported by the Air Force Office of Scientific Research, under Grant No. 78-3674-C and the Office of Naval Research under Grant No. N00014-76-C-0867.

References

[1] i) T.B. Benjamin, J. Fluid Mech. 29, 559 (1967).

ii) H. Ono, J. Phys. Soc. Japan, 39, 1082 (1975).

[2] D.J. Benney, (private communication); S.A. Maslowe, L.G. Redekopp, J. Fl. Mech. 101, 321 (1980).

[3] K.M. Case, Proc. Nat. Acad. Sc. U.S.A., 75, 3563 (1978); 76, 1 (1979).

[4] H.H. Chen, Y.C. Lee, N.R. Pereira, Phys. Fluids 22, 187 (1979).

[5] A. Nakamura, J. Phys. Soc. Jap. 47, 1701 (1979).

[6] T.L. Boch, M.D. Kruskal, Phys. Lett. 74A, 173 (1979).

[7] A.S. Fokas, B. Fuchssteiner, Phys. Lett. 86A, 341 (1981).

[8] M.J. Ablowitz, A.S. Fokas, R.L. Anderson, "The Direct Linearizing Transform and the Benjamin-Ono Equation" (preprint), 1982.

[9] A.S. Fokas and M.J. Ablowitz, "On the Inverse Scattering and Direct Linearizing Transforms of the Kadomtsev-Petviashvili Equation" (preprint), 1982.

[10] N.I. Muskhelishvili, "Singular Integral Equations", ed. by J. Radok (Noordhoff, Groningen, 1953).

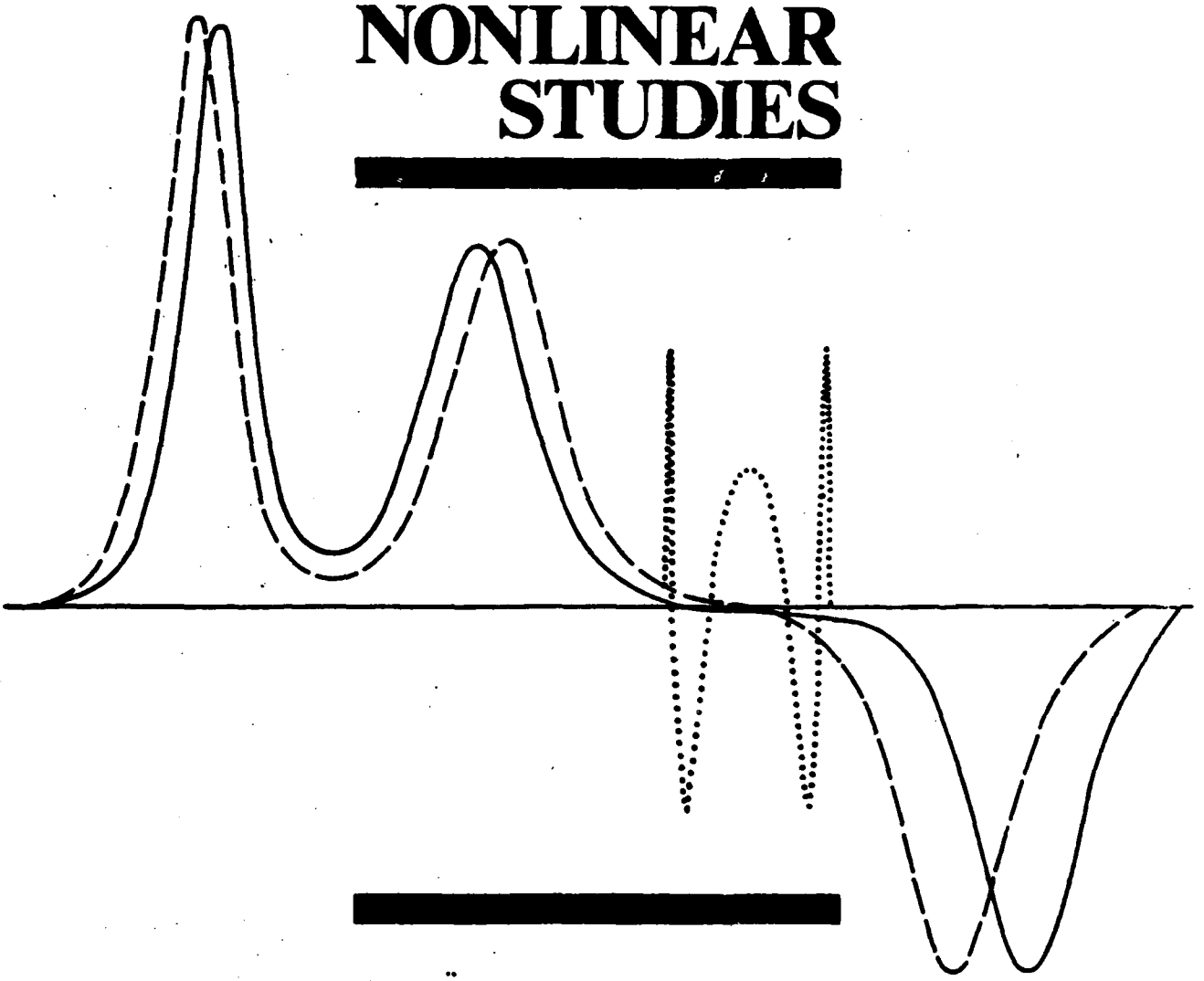
[11] F.D. Gakhov, "Boundary Value Problems", ed. by I.N. Sneddon (Pergamon, New York, 1966), Vol. 85.

[12] i) Y. Kodama, J. Satsuma, M.J. Ablowitz, Phys. Rev. Lett. 46, 687 (1981);

ii) Y. Kodama, M.J. Ablowitz, J. Satsuma, J. Math. Phys. 23(4), 564 (1982).

[13] M.J. Ablowitz and A.S. Fokas, "A Direct Linearization Associated with the Benjamin-Ono Equation", to appear in the proceedings of the Workshop of Mathematical Methods in Hydrodynamics and Integrability in Related Dynamical Systems (La Jolla, CA, December 7-9, 1981)

INSTITUTE FOR NONLINEAR STUDIES



ON A BÄCKLUND TRANSFORMATION AND SCATTERING PROBLEM FOR THE MODIFIED INTERMEDIATE LONG WAVE EQUATION

by
Junkichi Satsuma[†], Thiab R. Taha, Mark J. Ablowitz

[†]Permanent Address: General Education, Miyazaki Medical College, Kiyotake,

Clarkson College of Technology
Potsdam, New York 13676

ON A BÄCKLUND TRANSFORMATION AND SCATTERING PROBLEM FOR THE MODIFIED INTERMEDIATE LONG WAVE EQUATION

by

Junkichi Satsuma[†], Thiab R. Taha, Mark J. Ablowitz

Department of Mathematics and Computer Science

Clarkson College of Technology

Potsdam, NY 13676

Abstract

In this note we give a Bäcklund transformation, an associated linear scattering problem, and a method for finding the conservation laws for the so-called Modified Intermediate Long Wave equation. This equation reduces to the Modified Korteweg-deVries equation in the shallow water limit.

[†]Permanent Address: General Education, Miyazaki Medical College, Kiyotake, Miyazaki 889-16 Japan

May, 1982

1. Introduction

It is well known in the theory of solitons (see for example Ablowitz and Segur [1]) that multi-soliton solutions of certain nonlinear evolution equations can be obtained by more than one method. Some examples are: the inverse scattering transform (IST) ([2], [3], [4]), Bäcklund transformation (BT) [5]; and the so-called bilinear theory [6]. The IST scheme gives a method of solving the initial-value problem for a broad class of nonlinear evolution equations [4]. Nevertheless if we are given a particular evolution equation it can be difficult to set up an appropriate IST. On the other hand, if we can establish a BT associated with a given evolution equation then by suitable transformations the BT usually may be reduced to the IST. In this note we shall introduce a BT, an IST, and a method for finding the conservation laws for the so-called Modified Intermediate Long Wave (MILW) equation. In the literature the MILW equation is related to the Intermediate Long Wave (ILW) equation ([7],[8]) in the analogous way that the MKdV is to the KdV.

The ILW equation describes long internal gravity waves in a stratified fluid with finite depth. It is written in a simplified form as

$$u_t + 2uu_x + Tu_{xx} = 0, \tag{1.1}$$

where

$$T(u_x) = \hat{T}(u_x) + \frac{1}{\delta}u, \tag{1.2}$$

and

$$\hat{T}(u) = \int_{-\infty}^{\infty} [-\frac{1}{2\delta} \coth \frac{\pi(x-\xi)}{2\delta}] \; u(\xi)d\xi, \tag{1.3}$$

($\int_{-\infty}^{\infty}$ denotes the Cauchy principal value integral) and δ is a parameter representing the distance between the boundary and the internal wave layer. In the shallow water limit, $\delta \to 0$, equation (1.1) reduces to the Korteweg-deVries (KdV) equation,

$$u_t + 2uu_x + \frac{\delta}{3} u_{xxx} = 0, \tag{1.4}$$

and in the deep water limit, $\delta \to \infty$, reduces to the Benjamin-Ono (BO) equation,

$$u_t + 2uu_x + H(u_{xx}) = 0, \tag{1.5}$$

where H denotes the Hilbert transform operator

$$H(u) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(x')}{x'-x} dx' . \tag{1.6}$$

(1.1) (finite δ) and (1.5) have recently been solved by IST ([9],[10]). The scattering problem and hence solutions by IST is open for the MILW equation, see (2.4) below.

2. The Derivation of the MILW Equation

Introducing $\omega = \int^x u dx$, equation (1.1) may be written as

$$\omega_t + (\omega_x)^2 + T(\omega_{xx}) = 0. \tag{2.1}$$

The BT of equation (1.1) is expressed as [8]

$$(\omega'+\omega)_x = \lambda + iT(\omega'-\omega)_x - i\delta^{-1}(\omega'-\omega) + \mu e^{i(\omega'-\omega)}, \tag{2.2a}$$

$$(\omega'-\omega)_t = -(\delta^{-1}+\lambda)(\omega'-\omega)_x + i(\omega'+\omega)_{xx}$$

$$-i(\omega'-\omega)_x T(\omega'-\omega)_x + i\delta^{-1}(\omega'-\omega)(\omega'-\omega)_x, \tag{2.2b}$$

where λ and μ are arbitrary parameters. (See Appendix B for the BT in bilinear form for the ILW equation.) We note that if ω satisfies equation (2.1), ω' defined by equation (2.2) also satisfies equation (2.1). Substituting equation (2.2a) into (2. 2b) and introducing $V=\omega-\omega'$, we have [8]

$$V_t + T(V_{xx}) + \lambda V_x + V_x\,[\mu e^{iV} + i(T(V_x) - \tfrac{1}{\delta}V)] = 0. \tag{2.3}$$

By setting $V \to -iV$, $t \to \frac{3}{\delta}t$, $\lambda = -\frac{1}{\delta}$ and $\mu = \frac{1}{\delta}$, the above equation (2.3) takes the following form

$$\tfrac{\delta}{3} V_t + \hat{T}(V_{xx}) + V_x(\tfrac{1}{\delta}e^{V} + \hat{T}(V_x)) = 0. \tag{2.4}$$

We refer to equation (2.4) as the MILW equation, which in the limit, $\delta \to 0$, using the expansion,

$$\hat{T}(V_x) = -\tfrac{1}{\delta}V + \tfrac{\delta}{3} V_{xx} + \tfrac{\delta^3}{45}V_{xxxx} + O(\delta^5), \tag{2.5a}$$

and

$$V = 2\delta u - \tfrac{1}{6}\delta^3 u_{xx} + O(\delta^5), \tag{2.5b}$$

yields the Modified Korteweg-deVries (MKdV) equation,

$$u_t + 6u^2 u_x + u_{xxx} = 0. \tag{2.6}$$

3. The BT of the MILW Equation

Introducing the following dependent variable transformation

$$V = \log \frac{f^+ g^-}{f^- g^+}, \tag{3.1}$$

equation (2.4) can be expressed in bilinear form as

$$(\tfrac{i}{3}\delta D_t + D_x^2) f^\pm \cdot g^\pm = 0, \tag{3.2a}$$

$$2i\delta D_x f^- \cdot g^+ = -f^- g^+ + f^+ g^-, \tag{3.2b}$$

where we have used the abbreviations

$$f^\pm(x) = f(x \pm i\delta), \tag{3.3}$$

and introduced bilinear differential operators defined by [6,11] as

$$D_t^n D_x^m a \cdot b = \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right)^n \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^m a(x,t) b(x',t') \Big|_{x = x',\ t = t'} \tag{3.4}$$

(See appendix A for some properties of these operators).

An (exact) N-soliton solution of equation (3.2) has been obtained by A. Nakamura [12]. In the limit, $\delta \to 0$, these soliton solutions tend to the soliton solutions of the MKdV equation (2.6).

As discussed above the MKdV equation plays the same role with respect to the MILW equation as does the KdV equation with respect to the ILW equation [9]. (See Figure 1 below).

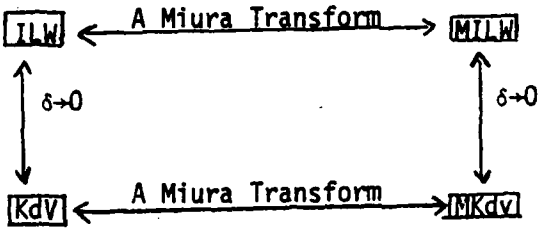


Figure 1. Describes the relations of the ILW, KdV, MILW and MKdV equations.

By analogy to the KdV-MKdV case ([13-15],[17]) we have found the following BT of equation (2.4) in bilinear form:

$$[\tfrac{i}{3}\delta D_t + i(\tfrac{1}{\delta} - c)D_x + D_x^2 - \tfrac{1}{4}k^2]\, f^{\pm}.f'^{\pm} = 0, \tag{3.5a}$$

$$[\tfrac{i}{3}\delta D_t + i(\tfrac{1}{\delta} - c)D_x + D_x^2 - \tfrac{1}{4}k^2]\, g^{\pm}.g'^{\pm} = 0, \tag{3.5b}$$

$$(2iD_x + c)f^{-}.f'^{+} = \mu f^{+}.f'^{-}, \tag{3.5c}$$

$$(2iD_x + c)g^{-}.g'^{+} = \mu g^{+}.g'^{-}, \tag{3.5d}$$

where

$$c = k \cot k\delta, \tag{3.5e}$$

and

$$\mu = k \operatorname{cosec} k\delta. \tag{3.5f}$$

Equation (3.5) relates a solution (f,g) of (3.2) with another solution (f',g') of (3.2)

(for more details see [16]).

It can be shown that the limit, $\delta \to 0$, of the bilinear form of the MILW equation (which is given in equation (3.2)) tends to the bilinear form of the MKdV equation (2.6) which is

$$(D_t + D_x^3) f \cdot f^* = 0, \tag{3.6a}$$

$$D_x^2 f \cdot f^* = 0, \tag{3.6b}$$

where $u = i(\log f^*/f)x$,
and the limit of the bilinear BT (3.5) reduces to the bilinear BT of the MKdV equation (see [17]):

$$(D_t + \tfrac{3}{4}k^2 D_x + D_x^3) f \cdot f' = 0, \tag{3.7a}$$

$$(D_t + \tfrac{3}{4}k^2 D_x + D_x^2) f^* \cdot f'^* = 0, \tag{3.7b}$$

$$D_x^2 f \cdot f' = \tfrac{1}{4}k^2 f \cdot f', \tag{3.7c}$$

$$D_x^2 f'^* \cdot f^* = \tfrac{1}{4}k^2 f^* \cdot f'^*, \tag{3.7d}$$

(where * implies a complex conjugate). Using the following relations (see [8]):

$$(\log f^{\pm})_x = \tfrac{1}{4}(i\hat{T} \pm 1)\{(i\hat{T}+1)V_x + \tfrac{i}{\delta}(e^V - 1)\}, \tag{3.8}$$

$$(\log g^{\pm})_x = \tfrac{1}{4}(i\hat{T} \pm 1)\{(i\hat{T}-1)V_x + \tfrac{i}{\delta}(e^V - 1)\}, \tag{3.9}$$

the bilinear BT (3.5) transforms to the BT in physical variables:

$$\frac{i}{3}\delta(V-V')_t + i(\frac{1}{\delta} - c)(V-V')_x + (V+V')_{xx}$$

$$- \frac{1}{2}\hat{T}(V-V')_x\{\hat{T}(V-V')_x + \frac{1}{\delta}(e^V-e^{V'})\}$$

$$- \frac{1}{2}(V-V')_x\hat{T}\{\hat{T}(V-V')_x + \frac{1}{\delta}(e^V-e^{V'})\} = 0, \qquad (3.10a)$$

$$\frac{i}{3}\delta[\hat{T}(V-V') + \frac{1}{\delta}\int_{-\infty}^{x}(e^V-e^{V'})dx]_t - (\frac{1}{\delta} - c)[\hat{T}(V-V')_x + \frac{1}{\delta}(e^V-e^{V'})]$$

$$+ i\hat{T}(V+V')_{xx} + \frac{i}{\delta}(e^V+e^{V'})_x$$

$$- \frac{i}{2}[\hat{T}\{\hat{T}(V-V')_x + \frac{1}{\delta}(e^V-e^{V'})\}]\cdot[\hat{T}(V-V')_x + \frac{1}{\delta}(e^V-e^{V'})]$$

$$+ \frac{i}{2}(V-V')_x\hat{T}(V-V')_x = 0, \qquad (3.10b)$$

$$-\hat{T}(V-V')_x - i(V+V')_x$$

$$= 2\mu \exp\frac{i}{2}[(\hat{T}(V-V') + \frac{1}{\delta}\int^{x}(e^V-e^{V'})dx] \sinh\frac{1}{2}(V-V'), \qquad (3.10c)$$

$$- i\hat{T}[\hat{T}(V-V')_x + \frac{1}{\delta}(e^V-e^{V'})] + \hat{T}(V+V')_x + \frac{1}{\delta}(e^V+e^{V'}-2)$$

$$= -2c+2\mu \exp\frac{i}{2}[\hat{T}(V-V') + \frac{1}{\delta}\int^{x}(e^V-e^{V'})dx] \quad \cosh\frac{1}{2}(V-V'). \qquad (3.10d)$$

We note that either set of equations (3.10a,c) or (3.10 b,d) are adequate to represent the BT of the MILW equation.

4. The Conserved Quantities of the MILW Equation

One may derive the conserved quantities from the BT (see for example [7], [18]). Introduce $W = V - V'$, and rewrite equation (3.10) as

$$\frac{i}{3}\delta W_t + i(\frac{1}{\delta} - c)W_x + (2V-W)_{xx} - \frac{1}{2}\hat{T}(W_x)\cdot I_x - \frac{1}{2}W_x\cdot\hat{T}(I_x) = 0 \quad (4.1a)$$

$$\frac{i}{3}\delta I_t - (\frac{1}{\delta} - c)I_x - i\{I_x - 2\hat{T}(V_x) - \frac{2}{\delta}e^V\}_x$$

$$- \frac{i}{2}I_x\hat{T}(I_x) + \frac{i}{2}W_x\hat{T}(W_x) = 0, \quad (4.1b)$$

$$- \hat{T}(W_x) + i(W-2V)_x = 2\mu(\exp\tfrac{i}{2}I)\sinh\tfrac{1}{2}W, \quad (4.1c)$$

$$- i\hat{T}[I_x] + \hat{T}(2V-W)_x + \frac{1}{\delta}(e^V + e^{V-W} - 2) = -2c + 2\mu(\exp\tfrac{i}{2}I)\cosh\tfrac{1}{2}W,$$

where

$$I = \hat{T}(W) + \frac{1}{\delta}\int_{-\infty}^{x} e^V(1-e^{-W})dx. \quad (4.2)$$

Imposing the boundary conditions $V(\pm\infty) = 0$ and using

$$\int_{-\infty}^{\infty}[u\hat{T}(v) + v\hat{T}(u)]dx = 0, \quad (4.3)$$

we have from equation (4.1a) that

$$[\int_{-\infty}^{\infty} W dx]_t = 0, \quad (4.4)$$

and from equation (4.1b) that

$$[\int_{-\infty}^{\infty} I dx]_t = 0, \quad (4.5)$$

which means that $\int_{-\infty}^{\infty} W dx$ and $\int_{-\infty}^{\infty} I dx$ are conserved quantities.

Following Miura [19], for $\mu \to \infty$. If we expand

$$W = \sum_{n=1}^{\infty} \varepsilon^n W_n , \tag{4.6}$$

where

$$\varepsilon = \frac{1}{2\mu} << 1 , \tag{4.7}$$

and equating the same power of ε in equation (4.1c), we obtain

$$W_1 = -4iV_x , \tag{4.8a}$$

$$W_2 = 8i[(\hat{T}-i)V_{xx} + V_x \hat{T}(V_x) + \frac{1}{\delta} V_x (e^V - 1)], \tag{4.8b}$$

$$W_3 = -2(\hat{T}-i)W_{2x} - \frac{1}{24} W_1^3 - W_1 \{\frac{i}{2}\hat{T}(W_2) + \frac{i}{2\delta}\int_{-\infty}^{x} e^V (W_2 - \frac{1}{2}W_1^2) dx ,$$

$$- \frac{1}{8}(\hat{T}(W_1) + \int_{-\infty}^{x} \frac{1}{\delta} e^V W_1 dx)^2\} - \frac{i}{2} W_2 \{\hat{T}(W_1) + \frac{1}{\delta}\int_{-\infty}^{x} e^{\cdot} w_1 dx\}. \tag{4.8c}$$

$$\vdots$$

Unfortunately the integrals $\int_{-\infty}^{\infty} W_1 dx$, $\int_{-\infty}^{\infty} W_2 dx$, $\int_{-\infty}^{\infty} W_3 dx$, etc. are trivial conserved quantities. But from the expansion of I:

$$I = \hat{T}(W) + \frac{1}{\delta}\int_{-\infty}^{x} e^V (1 - e^{-W}) dx = \sum_{n=1}^{\infty} \varepsilon^n I_n , \tag{4.9}$$

we obtain the following (using (4.6))

$$I_1 = -4i[\hat{T}(V_x) + \frac{1}{\delta}\int_{-\infty}^{x} e^V V_x dx]$$

$$I_2 = 8i\hat{T}[\hat{T}-i)V_{xx} + V_x\hat{T}(V_x) + \frac{1}{\delta}V_x(e^V-1)]$$

$$+ \frac{1}{\delta}\int_{-\infty}^{x} e^V[8i\{(\hat{T}-i)V_{xx} + V_x\hat{T}(V_x) + \frac{1}{\delta}V_x(e^V-1)\}$$
$$+ 8V_x^2]dx, \qquad (4.10)$$

from which are deduced the following conserved quantities

$$1)\ \frac{1}{\delta}(e^V-1) \qquad (4.11a)$$

$$2)\ \hat{T}(V_x\hat{T}(V_x)) + \frac{1}{\delta}e^V\hat{T}(V_x) + \frac{1}{2\delta^2}(e^{2V}-1) \qquad (4.11b)$$

$$\vdots$$

These may be verified directly.

In the limit as $\delta\to 0$, equations (4.11a), and (4.11b) using (2.5) give the first and the second conserved quantities of the MKdV equation (at the lowest non-trivial order of δ), which are u, and u^2 respectively [20].

5. The IST Problem for the MILW Equation

In a manner analogous to Wadati [20] and Satsuma et al [7] we obtain the IST problem of equation (2.4) from the BT given in equation (4.1). By using the following dependent variable transformation

$$(\log \psi_1^{\pm})_x = (\log \frac{f}{f}')_x = \frac{1}{4}(i\hat{T}\pm 1)\{W_x+iI_x\}, \tag{5.1a}$$

and

$$(\log \psi_2^{\pm})_x = (\log \frac{g}{g}')_x = -\frac{1}{4}(i\hat{T}\pm 1)\{-W_x+iI_x\}, \tag{5.1b}$$

(see Appendix C for useful relations which can be derived from the above equation.), and by substituting (5.1) and Appendix C relations into equation (4.1), adding and substracting the resulting equations yield

$$-2i\psi_{1x}^{+} + [c-iV_x+\hat{T}(V_x) + \frac{1}{\delta}(e^V-1)]\psi_1^{+} = \mu\psi_1^{-}, \tag{5.2a}$$

$$-2i\psi_{2x}^{+} + [c+iV_x+\hat{T}(V_x) + \frac{1}{\delta}(e^V-1)]\psi_2^{+} = \mu\psi_2^{-}, \tag{5.2b}$$

$$-\frac{i}{3\delta}\psi_{1t}^{\pm} - i(\frac{1}{\delta}-c)\psi_{1x}^{\pm} + [\frac{1}{2}(i\hat{T}\pm 1)\{V_{xx}+i\hat{T}(V_{xx})+\frac{i}{\delta}(e^V)_x\} - \frac{1}{4}k^2]\psi_1^{\pm} + \psi_{1xx}^{\pm} = 0, \tag{5.2c}$$

$$-\frac{i}{3\delta}\psi_{2t}^{\pm} - i(\frac{1}{\delta}-c)\psi_{2x}^{\pm} + [\frac{1}{2}(i\hat{T}\pm 1)\{-V_{xx}+i\hat{T}(V_{xx})+\frac{i}{\delta}(e^V)_x\} - \frac{1}{4}k^2]\psi_2^{\pm} + \psi_{2xx}^{\pm} = 0. \tag{5.2d}$$

We expect that $\psi_i^{\pm}$ have the following analytical meaning: $\psi^{\pm}$ represent the boundary values of functions (i.e. $\psi^{\pm}(x) = \lim_{\text{Im } z \to 0} \psi^{\pm}(z)$) analytic in the horizontal strips between Im z = 0 and Im z = $\mp 2\delta$, and periodically extended thereafter.

By defining

$$\psi_1 = \frac{1}{2}(v_1 - iv_2) \ , \tag{5.3a}$$

$$\psi_2 = \frac{1}{2}(v_1 + iv_2) \ , \tag{5.3b}$$

and substituting into the above equation (5.2) we get the following set of equations (by adding and substracting)

$$v_{1x}^{+} + \bar{W}v_1^{+} + Uv_2^{+} - \frac{i}{2}\mu v_1^{-} = 0, \tag{5.4a}$$

$$v_{2x}^{+} - Uv_1^{+} + \bar{W}v_2^{+} - \frac{i}{2}\mu v_2^{-} = 0, \tag{5.4b}$$

$$\frac{i}{3}\delta v_{1t}^{\pm} + i(\frac{1}{\delta} - c)v_{1x}^{\pm} - v_{1xx}^{\pm} - A^{\pm}v_1^{\pm} - B^{\pm}v_2^{\pm} = 0, \tag{5.4c}$$

$$\frac{i}{3}\delta v_{2t}^{\pm} + i(\frac{1}{\delta} - c)v_{2x}^{\pm} - v_{2xx}^{\pm} + B^{\pm}v_1^{\pm} - A^{\pm}v_2^{\pm} = 0, \tag{5.4d}$$

where

$$U = \frac{i}{2}V_x, \tag{5.5}$$

$$\bar{W} = \frac{i}{2}\{\hat{T}(V_x) + \frac{1}{\delta}(e^V - 1) + c\}, \tag{5.6}$$

$$A^{\pm} = \frac{1}{2}(i\hat{T} \pm 1)\{i\hat{T}(V_x) + \frac{i}{\delta}(e^V - 1)\}_x - \frac{1}{4}k^2, \tag{5.7}$$

$$B^{\pm} = \frac{i}{2}(i\hat{T} \pm 1)V_{xx}, \tag{5.8}$$

$$c = k \cot k\delta, \tag{5.9}$$

$$\mu = k \operatorname{cosec} k\delta . \tag{5.10}$$

It is readily shown by a cross differentiation that equations (5.4a-5.4d) constitute an IST problem associated with equation (2.4). Assuming $V = O(\delta)$, expanding $v_i^{\pm}(i=1,2)$ in terms of δ and substituting them into equation (5.4), we obtain at the lowest nontrivial order of δ,

$$v_{1xx} + (u^2 - \tfrac{1}{4}k^2)v_1 + v_2 u_x = 0 , \tag{5.11a}$$

$$v_{2xx} + (u^2 - \tfrac{1}{4}k^2)v_2 - v_1 u_x = 0 , \tag{5.11b}$$

$$v_{1t} + 2v_{2x}u_x - v_2 u_{xx} - v_1(u^2)_x + v_{1x}\{k^2+2u^2\} = 0 , \tag{5.11c}$$

$$v_{2t} - 2v_1 u_x + v_1 u_{xx} - v_2(u^2)_x + v_{2x}\{k^2+2u^2\} = 0 , \tag{5.11d}$$

which is one of the known IST problems associated with the MKdV equation (2.6). It is to be noted that equation (2.4) is a necessary compatibility condition for equations (5. 2a) and (5.2c). (See Appendix D for still another form of IST problem for the MILW equation).

Appendix A: Properties of the D operator

We have

$$D_t^n D_x^m a\cdot b = \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right)^n \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^m a(t,x)\, b(t',x')\big|_{t=t',x=x'}.$$

The following properties are easily seen from the definition.

(1) $D_x^m a\cdot 1 = \left(\frac{\partial}{\partial x}\right)^m a$,

(2) $D_x^m a\cdot b = (-1)^m D_x^m b\cdot a$,

(3) $D_x^m a\cdot a = 0$ for odd m,

(4) $D_x^m a\cdot b = D_x^{m-1}(a_x\cdot b - a\cdot b_x)$,

(5) $D_x^m a\cdot a = 2D_x^{m-1} a_x\cdot a$ for even m,

(6) $D_x D_t a\cdot a = 2D_x a_t\cdot a$

$= 2D_t a_x\cdot a$,

(7) $D_x^m \exp(k_1 x)\cdot\exp(k_2 x) = (k_1-k_2)^m \exp(k_1+k_2)x$,

Let $\phi = \log(a/b)$ and $\rho = \log ab$, we have

(8) $(D_x a\cdot b)/ab = \phi_x$,

(9) $(D_x^2 a\cdot b)/ab = \rho_{xx} + (\phi_x)^2$,

(10) $(D_x^3 a\cdot b)/ab = \phi_{xxx} + 3\phi_x\rho_{xx} + (\phi_x)^3$.

Appendix B: On the ILW Equation

The ILW equation is given by

$$u_t + 2uu_x + T(u_{xx}) = 0. \tag{B.1}$$

Introducing the following dependent variable transform

$$u = i(\log f^-/f^+)_x, \tag{B.2}$$

equation (B.1) can be expressed in bilinear form as

$$(iD_t + \frac{i}{\delta}D_x + D_x^2)f^+.f^- = 0. \tag{B.3}$$

(See Appendix A for some properties of these operators). The BT of equation (B.3) in the bilinear form is given by [8]

$$(iD_t + i(\frac{1}{\delta} + \lambda)D_x + D_x^2 - \frac{1}{4}k^2)f^{\pm}.g^{\pm} = 0, \tag{B.3a}$$

$$(2iD_x - \lambda)f^-.g^+ = \nu\, f^+g^-, \tag{B.3b}$$

where

$$\lambda = -k \cot k\delta, \tag{B.3c}$$

$$\nu = k \operatorname{cosec} k\delta. \tag{B.3d}$$

Appendix C: Useful Relations Derived from Equation (5.1)

$$(1)\quad \left(\log \frac{\psi_1^+}{\psi_1^-}\right)_x = -\frac{1}{2}\{W_x + iI_x\},$$

$$(2)\quad \left(\log \frac{\psi_2^+}{\psi_2^-}\right)_x = -\frac{1}{2}[-W_x + iI_x],$$

$$(3)\quad \left(\log \frac{\psi_2^+}{\psi_1^+}\right)_x = \frac{1}{2}(i\hat{T}+1)W_x,$$

$$(4)\quad \left(\log \frac{\psi_2^-}{\psi_1^-}\right)_x = \frac{1}{2}(i\hat{T}-1)W_x,$$

$$(5)\quad \left(\log \frac{\psi_2^+\psi_1^-}{\psi_1^+\psi_2^-}\right)_x = W_x\,,$$

$$(6)\quad \left(\log \frac{\psi_2^+\psi_2^-}{\psi_1^+\psi_1^-}\right)_x = i\hat{T}W_x\,.$$

Appendix D: Another IST Problem for the MILW Equation

In this appendix we will give an alternaive IST problem for the MILW equation (2.4). It is to be noted that we need only equations for one of the ψ_i, say ψ_1, to obtain this.

Define

$$Q^{\pm} = (\hat{T}\pm i)V + \int_{-\infty}^{x} \frac{1}{\delta}(e^{V}-1)dx. \tag{d.1}$$

Rewrite equations (5.1a) and (5.1b) as

$$(\log \psi_1^{\pm}) = -\frac{1}{4}(\hat{T}\mp i)(Q'^{-}-Q^{-}), \tag{d.2a}$$

$$(\log \psi_2^{\pm}) = -\frac{1}{4}(\hat{T}\mp i)(Q'^{+}-Q^{+}), \tag{d.2b}$$

resepctively. For convenience, we denote $\psi_1^{\pm}$ by $\xi_1^{\mp}$ and $\psi_2^{\pm}$ by $\xi_2^{\mp}$. Equations (5.2a) and (5.2c) can be rewritten as

$$-2i\xi_{1x}^{-} + (Q_x^{-}+c)\xi_1^{-} = -\mu\xi_1^{+}, \tag{d.3a}$$

$$\frac{i}{3\delta}\xi_{1t}^{\mp} + i(\frac{1}{\delta} - c)\xi_{1x}^{\mp} + \{\frac{1}{2}(\hat{T}\mp i)Q_{xx}^{-} + \frac{1}{4}k^2\}\xi_1^{\mp} - \xi_{1xxx}^{\mp} = 0, \tag{d.3b}$$

respectively, and from the analytical property note that

$$\left.\begin{aligned}(Q^{\pm})^{*} &= Q^{\mp},\\ (\xi_1^{\mp})^{*} &= \xi_2^{\pm}\end{aligned}\right\} \tag{d.4}$$

In analogy to the MKdV case if we take equations (d.3) together with their complex conjugates this will yield the IST problem of the MILW equation. Indeed, the following set of equations for ξ_1 and ξ_2 which are

$$-2i\ \xi_{1x}^{-} + (Q_x^{-}+c)\xi_1^{-} = -\mu\xi_1^{+}, \tag{d.5a}$$

$$2i\ \xi_{2x}^{+} + (Q_x^{+}+c)\xi_2^{+} = -\mu\xi_2^{-}, \tag{d.5b}$$

$$\frac{i}{3}\delta\ \xi_{1t}^{\mp} + i(\frac{1}{\delta} - c)\xi_{1x}^{\mp} + \{\frac{1}{2}(\hat{T}\mp i)Q_{xx}^{-} + \frac{1}{4}k^2\}\xi_1^{\mp} - \xi_{1xx}^{\mp} = 0, \tag{d.5c}$$

$$-\frac{i}{3}\delta\ \xi_{2t}^{\pm} - i(\frac{1}{\delta} - c)\xi_{2x}^{\pm} + \{\frac{1}{2}(\hat{T}\pm i)Q_{xx}^{+} + \frac{1}{4}k^2\}\xi_2^{\pm} - \xi_{2xx}^{\pm} = 0, \tag{d.5d}$$

yields an alternative IST problem for equation (2.4). By defining

$$\left.\begin{aligned} \xi_1^{-} &= -\tfrac{1}{2}(1+i\hat{T})A, \\ \xi_2^{+} &= \tfrac{1}{2}(1-i\hat{T})A, \\ \xi_1^{+} &= \tfrac{1}{2}(1-i\hat{T})\phi, \\ \xi_2^{-} &= -\tfrac{1}{2}(1+i\hat{T})\phi, \end{aligned}\right\} \tag{d.6}$$

equations (d.5a)-(d.5d) give the following set of equations (by adding and subtracting)

$$2i\,A_x - \bar{W}(i\hat{T}A) + iV_xA = -i\mu\hat{T}\phi\ , \tag{d.7a}$$

$$-2\hat{T}A_x - \bar{W}\cdot A + iV_x\hat{T}A = \mu\phi\ , \tag{d.7b}$$

$$\frac{i}{3\delta}A_t + i(\frac{1}{\delta} - c)A_x - U^+\cdot A - G^+(i\hat{T}A) - i\hat{T}A_{xx} = 0, \tag{d.7c}$$

$$-\frac{1}{3\delta}\hat{T}A_t - (\frac{1}{\delta} - c)\hat{T}A_x - U^+(i\hat{T}A) - G^+\cdot A - A_{xx} = 0, \tag{d.7d}$$

$$-\frac{i}{3\delta}\phi_t - i(\frac{1}{\delta} - c)\phi_x + U^-\cdot\phi - G^-(i\hat{T}\phi) - i\hat{T}\phi_{xx} = 0, \tag{d.7e}$$

$$-\frac{1}{3\delta}\hat{T}\phi_t - (\frac{1}{\delta} - c)\hat{T}\phi_x - U^-(i\hat{T}\phi) + G^-\cdot\phi + \phi_{xx} = 0, \tag{d.7f}$$

where

$$\bar{W} = c - i[i\hat{T}(V_x) + \frac{i}{\delta}(e^V-1)]\ , \tag{d.8a}$$

$$U^{\pm} = \frac{i}{2}\hat{T}(V_{xx}) \pm \frac{1}{2}[i\hat{T}V_x + \frac{i}{\delta}(e^V-1)]\ , \tag{d.8b}$$

$$G^{\pm} = \frac{i}{2}\hat{T}[i\hat{T}(V_{xx}) + \frac{i}{\delta}(e^V)_x] \pm \frac{1}{2}V_{xx} - \frac{1}{4}k^2. \tag{d.8c}$$

Acknowledgements:

This research was partially supported by the Office of Naval Research, Mathematics Division under Grant No.00014-76-C-0867, and the Air Force Office of Scientific Research, Air Froce Systems Command, USAF under Grant No. 78-3674-C. We are grateful for useful conversations with Professor A.S. Fokas and Dr. A. Nakamura.

References

[1] Ablowitz, M. and Segur, H., "Solitons and the Inverse Scattering Transform", SIAM, Philadelphia, 1981.

[2] Gardner, C., Greene, J., Kruskal, M., Miura, R., Phys. Rev. Lett. 19, 1095 (1967).

[3] Zakharov, V., and Shabat, A., Sov. Phys. - JETP 34, 62 (1972).

[4] Ablowitz, M., Kaup, D., Newell, A., and Segur, H., Stud. Appl. Math. 53, 249 (1974).

[5] Lamb, G.L., "Elements of Soliton Theory", Publ. by Wiley Interscience, 1980.

[6] Hirota, R., "Direct Method of Finding Exact Solutions of Nonlinear Evolution Equations, Bäcklund Transformations, in Lecture Notes in Mathematics, R.M. Miura, ed. Vol. 515, Springer-Verlag, pp. 40-68, (1974).

[7] Satsuma, J., Ablowitz, M.J., and Kodama, Y., "On an Internal Wave Equation Describing a Stratified Fluid With Finite Depth", Phys. Lett. 73A, pp. 283-286 (1979).

[8] Satsuma, J., and Ablowitz, M., "Solutions of an Internal Wave Equation Describing a Stratified Fluid With Finite Depth", in Nonlinear Partial Differential Equations in Engineering and Applied Science, edited by R.L. Sternberg, A.J. Kalinowki, and J.S. Papadakis (Marcel Dekker, New York, 1980), p. 397.

[9] Kodama, Y., Ablowitz, M.J., and Satsuma, J., "Direct and Inverse Scattering Problems of the Nonlinear Intermediate Long Wave Equation", J. Math. Phys. 23 (4), p. 564, April, 1982.

[10a] Ablowitz, M.J., Fokas, A.S., and Anderson, R.L., "The Direct Linearizing Transform and the Benjamin-Ono Equation", (preprint).

[10b] Fokas, A.S. and Ablowitz, M.J., "The Inverse Scattering Transform for the Benjamin-Ono Equation - A Pivot To Multidimensional Problems, (preprint).

[11] Hirota, R., and Satsuma, J., "A Variety of Nonlinear Network Equations Generated from the Bäcklund Transformations for the Toda Lattice, Prog. Theor. Phys. Suppl. 59: 64-100 (1976).

References (cont'd)

[12] Nakamura, A., J. Phys. Soc. Jpn. 47, No. 6, December (1979).

[13] Wahlquist, H.D. and Estabrook, F.B., Phys. Rev. Letters 31, 1386 (1973).

[14] Wadati, M., J. Phys. Soc. Japan 36, 1498 (1974).

[15] Lamb, G.L., Jr., J. Math. Phys. Vol. 15, No. 12, December (1974).

[16] Taha, T., PhD. Thesis, Clarkson College, Potsdam, NY, 1982.

[17] Hirota, R., and Satsuma, J., J. Phys. Soc. Jpn. 45, November 5, 1978.

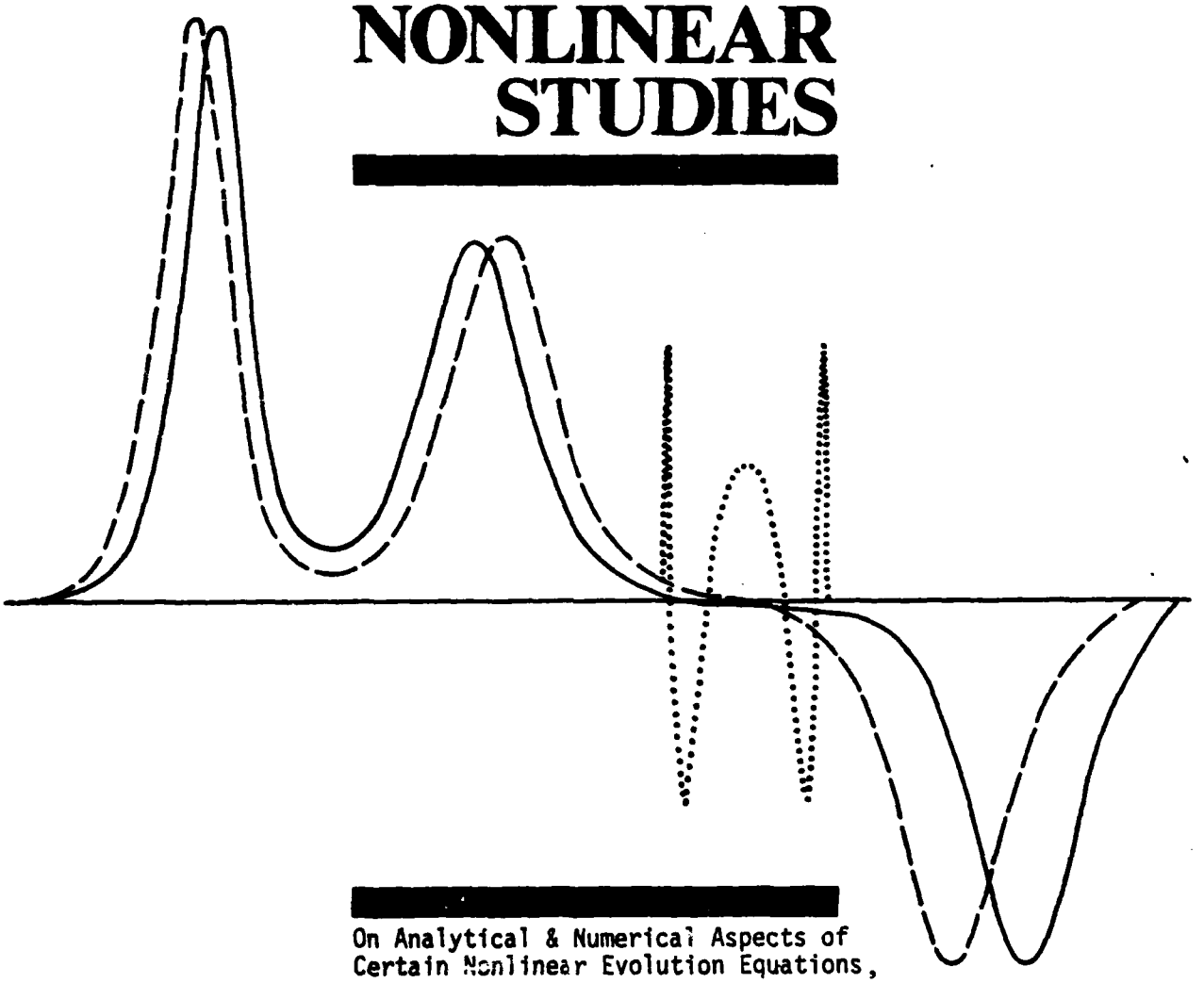
[18] Satsuma, J., Prog. Theor. Phys. 52 (1974) 1396, 53 (19750 585.

[19] Miura, R.M., J. Math. Phys. 9, 1202 (1968).
Miura, R.M. Gardner, C.S., and Kruskal, M.D., J. Math. Phys. 9, 1204 (1968).

[20] Wadati, M., Sanki, H., and Konno, K., Prog. Theor. Phys. 53, No. 2 pp. 419-436, (1975).

I.F.N.S. #14

INSTITUTE FOR NONLINEAR STUDIES



On Analytical & Numerical Aspects of Certain Nonlinear Evolution Equations,

Part I: Analytical

by

Thiab R. Taha and Mark J. Ablowitz

Clarkson College of Technology
Potsdam, New York 13676

ON ANALYTICAL AND NUMERICAL ASPECTS OF CERTAIN NONLINEAR EVOLUTION EQUATIONS, PART I: ANALYTICAL

ABSTRACT

Nonlinear partial difference equations are obtained which have as limiting forms the nonlinear Schrödinger, Korteweg-deVries and Modified Korteweg-deVries equations. These difference equations have a number of special properties. They are constructed by methods related to the inverse scattering transform. They can be used as a basis for numerical schemes to the associated nonlinear evolution equations. Experiments have shown that they compare very favorably with other known numerical methods (Parts II, III). In part II of this paper, the Ablowitz-Ladik scheme for the nonlinear Schrödinger equation is compared to other known numerical schemes, and generally proved to be faster than all utilized finite difference schemes but somewhat slower than the finite Fourier (pseudospectral) methods. In part III, a proposed scheme for the Korteweg-deVries equation proved to be faster than both the finite difference and finite Fourier methods we considered.

1. INTRODUCTION

In recent years there has been rapid advancement in the study of physically interesting nonlinear problems. The progress in this field has, in part, been due to the synergetic approach [1], which consists of the simultaneous use of conventional analysis and numerical experiments to investigate nonlinear phenomena. In this paper we will derive a numerical scheme for the Korteweg-deVries (KdV) equation and the Modified

Korteweg-deVries (MKdV) equation based on the inverse scattering transform (IST). In parts II, III of this work, we show that the schemes compare favorably with other known methods. Before proceeding, it may be helpful to review some of the recent developments in this area.

The inverse scattering transform (see for example a recent reference on this subject by Ablowitz and Segur [3]) was first discovered by Gardner, Greene, Kruskal, and Miura [4,5] in their study of the KdV equation. Subsequently, Lax [6] put the ideas in an alternative form which allows the method to be readily generalized. Zakharov and Shabat [7] found a new eigenvalue problem which led them to the solution of the nonlinear Schrödinger (NLS) equation. Ablowitz, Kaup, Newell and Segur [8] showed that a generalization of the Zakharov-Shabat eigenvalue problem allows one to find the solution to a class of interesting evolution equations which, in addition to the above, includes the sine-Gordon, MKdV, self-induced transparency equations, etc.

These ideas also apply to certain classes of nonlinear differential-difference equations. Using discrete scattering procedures developed by Case and Kac [9,10], Flaschka [11] was able to solve the Toda lattice equations. Similar results were found by Manakov [12]. Subsequently, Ablowitz and Ladik [13] presented a discretized version of the generalized Zakharov-Shabat eigenvalue problem which allowed them to isolate a class of differential-difference equations solvable by inverse scattering.

Ablowitz and Ladik [14,15] further generalized this theory to cover nonlinear partial-difference equations. They found a class of such equations and further introduced an equation which can be used as a

numerical scheme for the NLS equation. It has the following advantages [16]: (see also the following section)

(i) This scheme maintains many of the important properties of the original problem. One can associate with this scheme an infinite set of conservation laws, just as in the case of the corresponding partial difference equation. This scheme has traveling wave solutions, with special properties, these are the solitons ([1],[4],[5]).

(ii) The associated linear scheme is always neutrally stable.

(iii) This scheme maintains a certain joint x,t symmetry of the original equation.

(iv) The order of accuracy is the same for both the linear and nonlinear schemes.

(v) This scheme depends globally on the mesh points, but it does suggest others which are local.

These nice properties motivate us to look for a numerical scheme for the MKdV and the KdV equations in an analogous way.

In the next section we will review the procedure of finding the partial-difference equations together with the results for the NLS equation, which has already been discussed by Ablowitz and Ladik [16]. We will then develop and introduce a new scheme for the MKdV and the KdV equations based on the above theory.

2. Nonlinear Partial Difference Equations.

The key step in obtaining partial difference equations which can be solved by inverse scattering is to make an association between the nonlinear evolution equation and a linear eigenvalue (scattering) problem.

In this discussion all the difference equations are related to the eigenvalue problem [13]

$$V^m_{1n+1} = zV^m_{1n} + Q^m_n V^m_{2n} + S^m_n V^m_{2n+1}$$

$$V^m_{2n+1} = \frac{1}{z} V^m_{2n} + R^m_n V^m_{1n} + T^m_n V^m_{1n+1} \tag{2.1}$$

where z is the eigenvalue and the potentials R^m_n, Q^m_n, S^m_n, T^m_n are defined on the spacelike interval $|n| < \infty$ and the timelike interval $m > 0$. The various evolution equations are distinguished by the associated time (m) dependence of the eigenfunctions

$$\Delta^m V^m_{1n} = A^m_n(z) V^m_{1n} + B^m_n(z) V^m_{2n}$$

$$\Delta^m V^m_{2n} = C^m_n(z) V^m_{1n} + D^m_n(z) V^m_{2n} \tag{2.2}$$

where $\Delta^m V^m_{in} = V^{m+1}_{in} - V^m_{in}$, (i=1,2). That is to say for each partial difference equation there corresponds a set of functions A^m_n, B^m_n, C^m_n, D^m_n depending in general on the potentials. The equations for determining the sets A^m_n, ..., D^m_n, and hence the evolution equations are obtained by requiring the eigenvalue z to be invariant with respect to m and by forcing the consistency

$$\Delta^m(E_n V^m_{in}) = E_n(\Delta^m V^m_{in}), \quad i=1,2 \tag{2.3}$$

where E_n is the shift operator in the spatial coordinate defined by

$E_n V_{in}^m = V_{in+1}^m$, i=1,2. Performing the operations indicated in (2.3) results in four equations.

For the special case associated with the NLS equation we let $R_n^m = \mp Q_n^{m*}$, $S_n^m = T_n^m = 0$, and the four equations are given by

$$
\begin{aligned}
&z\Delta_n A_n^m = Q_n^{m+1} C_n^m \pm B_{n+1}^m Q_n^{m*}, \\
&\frac{1}{z} B_{n+1}^m - zB_n^m + A_{n+1}^m Q_n^m - D_n^m Q_n^{m+1} = \Delta^m Q_n^m, \\
&zC_{n+1}^m - \frac{1}{z} C_n^m \mp D_{n+1}^m Q_n^{m*} \pm A_n^m Q_n^{m+1} = \mp \Delta^m Q_n^{m*}, \\
&\frac{1}{z} \Delta_n D_n^m = \mp Q_n^{m+1*} B_n^m - Q_n^m C_{n+1}^m
\end{aligned}
\tag{2.4}
$$

where $\Delta_n A_n^m = A_{n+1}^m - A_n^m$ *etc.*

This system can be solved in a deductive way. Using the ideas in [16], expansions in powers of z and $\frac{1}{z}$ are sought. The series

$$
\begin{aligned}
&A_n^m = \sum_{k=-1}^{1} A_n^{(2k)} z^{2k}, \quad B_n^m = \sum_{\substack{k=-1 \\ k\neq 0}}^{1} B_n^{(k)} z^k \\
&C_n^m = \sum_{\substack{k=-1 \\ k\neq 0}}^{1} C_n^{(k)} z^k, \quad D_n^m = \sum_{k=-1}^{1} D_n^{(2k)} z^{2k}
\end{aligned}
\tag{2.5}
$$

are substituted into (2.4) and the various powers of z^k ($A_n^{(k)}$, ..., $D_n^{(k)}$ are assumed independent of z) are set equal to zero. One can find each of the unknowns $A_n^{(2)}$, $A_n^{(1)}$, ..., $D_n^{(-2)}$ in terms of the potentials. The

condition under which we can solve (2.4) requires that the potential evolve according to the evolution equation

$$\Delta^m Q_n^m = Q_n^m A_-^{(0)} - Q_n^{m+1} A_-^{(0)*} - A_-^{(-2)*}\left(Q_{n+1}^{m+1}\left(1\pm Q_n^m Q_n^{m*}\right)\prod_{k=-\infty}^{n}\Lambda_k^m\right)$$

$$+ A_-^{(-2)}\left(Q_{n-1}^m\left(1\pm Q_n^{m+1}\, Q_n^{m+1*}\right)\prod_{k=-\infty}^{n-1}\Lambda_k^m\right)$$

$$+ A_-^{(2)}\left[Q_{n+1}^m \pm \frac{Q_n^m}{2}\left(Q_{n+1}^{m+1}Q_n^{m+1*}+Q_{n+1}^m\, Q_n^{m*} - \sum_{k=-\infty}^{n}\Delta^m\tilde{S}_k^m\right)\right]$$

$$- A_-^{(2)*}\left[Q_{n-1}^{m+1} \pm \frac{Q_n^{m+1}}{2}\left(Q_n^{m+1*}\, Q_{n-1}^{m+1}+Q_n^{m*}Q_{n-1}^m - \sum_{k=-\infty}^{n-1}\Delta^m\tilde{S}_k^m\right)\right], \tag{2.6}$$

where $\Lambda_k^m = (1\pm Q_k^{m+1}Q_k^{m+1*})/(1\pm Q_k^m Q_k^{m*})$ and $\tilde{S}_k^m = (Q_{k+1}^m Q_k^{m*}+Q_k^m Q_{k-1}^{m*})$.

The $A_-^{(i)}$ are arbitrary constants of summation (fixed at $n\to-\infty$) and $D_-^{(-i)*} = A_-^{(i)}$, $i=2,0,-2$.

So by a suitable choice of the constants, one can obtain a partial difference equation which is consistent with the NLS equation

$$iq_t = q_{xx} \pm 2|q|^2 q. \tag{2.7}$$

In the linear limit if we want

$$i\Delta^m Q_n^m = \frac{\sigma}{2}\left(Q_{n+1}^{m+1} - 2Q_n^{m+1} + Q_{n-1}^{m+1} + Q_{n+1}^m - 2Q_n^m + Q_{n-1}^m\right)$$

where $\sigma = \dfrac{\Delta t}{(\Delta x)^2}$ and $Q_n^m = \Delta x\, q(n\Delta x, m\Delta t) = \Delta x q_n^m$ then the constants

are chosen according to $A_-^{(2)} = -i\frac{\sigma}{2}$, $A_-^{(0)} = i\sigma$, $A_-^{(-2)} = -\frac{i\sigma}{2}$. This particular choice of constants in (2.6) leads to the following evolution equation:

$$i\frac{\Delta^m q_n^m}{\Delta t} = \frac{1}{2(\Delta x)^2}(q_{n+1}^m - 2q_n^m + q_{n-1}^m P_{n-1} + q_{n+1}^{m+1} P_n - 2q_n^{m+1} + q_{n-1}^{m+1})$$

$$\pm\frac{1}{4}[q_n^m(q_n^{m*} q_{n+1}^m + q_n^{m+1*} q_{n+1}^{m+1}) + q_n^{m+1}(q_{n-1}^m q_n^{m*} + q_{n-1}^{m+1} q_n^{m+1*})$$

$$+ 2q_n^m q_n^{m*} q_{n+1}^{m+1} P_n + 2q_n^{m+1} q_n^{m+1*} q_{n-1}^m P_{n-1} - q_n^m S_n - q_n^m S_{n-1}^* \qquad (2.8)_*$$

where

$$P_n = \prod_{k=-\infty}^{n} \Lambda_k^m, \quad S_n = \sum_{k=-\infty}^{n} \Delta^m \sigma_k^m,$$

$$\sigma_k^m = q_k^m q_{k-1}^{m*} + q_{k+1}^m q_k^{m*}, \quad \Lambda_k^m = (1 \pm q_k^{m+1} q_k^{m+1*}(\Delta x)^2)/(1 \pm q_k^m q_k^{m*}(\Delta x)^2).$$

This scheme is implicit and global. However, a local scheme is suggested in which $P_n=1$ and $S_n=0$ for all n. Equation $(2.8)_*$ is consistent with the NLS equation (2.7) with the truncation error of order $O((\Delta t)^2, (\Delta x)^2)$. Similarly the local scheme also has the same truncation error.

For the special case associated with the MKdV equation

$$R_t \mp 6R^2 R_x + R_{xxx} = 0 \qquad (2.9)$$

we let $R_n^m = \pm Q_n^m$, $T_n^m = S_n^m = 0$, and the four equations are given by

$$z\Delta_n A_n^m + R_n^m B_{n+1}^m \mp R_n^{m+1} C_n^m = 0 \tag{2.10}$$

$$\mp \Delta^m R_n^m + zB_n^m \mp R_n^{m+1} D_n^m = \mp A_{n+1}^m R_n^m + \frac{1}{z} B_{n+1}^m \tag{2.11}$$

$$zC_{n+1}^m - \frac{1}{z} C_n^m + R_n^m D_{n+1}^m - R_n^{m+1} A_n^m = \Delta^m R_n^m \tag{2.12}$$

$$\frac{1}{z} \Delta_n D_n^m = R_n^{m+1} B_n^m \mp R_n^m C_{n+1}^m \tag{2.13}$$

Using the ideas in [13,15,16], the coefficients in the equations for the time dependence of the eigenfunctions are expanded as

$$\begin{aligned} A_n^m &= \sum_{k=-2}^{2} z^{2k} A_n^{(2k)} \quad , \quad B_n^m = \sum_{k=-1}^{2} z^{(2k-1)} B_n^{(2k-1)} \\ C_n^m &= \sum_{k=-1}^{2} z^{(2k-1)} B_n^{(2k-1)} \quad , \quad D_n^m = \sum_{k=-2}^{2} z^{2k} A_n^{(2k)} . \end{aligned} \tag{2.14}$$

With the expanded form of A_n^m, B_n^m, C_n^m, D_n^m, equations (2.10), (2.11), (2.12), and (2.13) yield a sequence of twenty equations in eighteen unknowns corresponding to equating powers of z^5, z^{-5}, z^4, ..., z, z^{-1}, all of which must be independently satisfied. To solve these equations it is most convenient to solve the resultant equations corresponding to z^5 and z^{-5} first, then solve the equations corresponding to z^4, $\frac{1}{z^4}$, etc. Carrying out the algebra we find the values of $A_n^{(4)}$, ..., $D_n^{(-4)}$ in terms of the potentials (see Taha [2]). The remaining two equations are consistent under the following conditions:

$$A^{(i)} = D^{(-i)}, \quad i = 4,2,0,-2,-4. \tag{2.15}$$

The following evolution equation is

$$\Delta^m R_n^m = R_{n+2}^m A_-^{(4)} - R_{n+2}^{m+1}\gamma_{n+1}D_-^{(4)} + R_{n+1}^m S_{n+1}$$

$$-R_{n+1}^{m+1}P_n - [R_{n-2}^{m+1}A_-^{(4)} - R_{n-2}^m\gamma_{n-2}D_-^{(4)} + R_{n-1}^{m+1}S_{n-2}$$

$$- R_{n-1}^m P_{n-1}] + R_n^m\Bigg\{D_-^{(0)} \pm \sum_{\ell=-\infty}^{n}\Bigg[R_\ell^{m+1}\{R_{\ell-2}^{m+1}A_-^{(4)}$$

$$- R_{\ell-2}^m\gamma_{\ell-2}D_-^{(4)} + R_{\ell-1}^{m+1}S_{\ell-2} - R_{\ell-1}^m P_{\ell-1}\}$$

$$-R_\ell^m(R_{\ell+2}^m A_-^{(4)} - R_{\ell+2}^{m+1}\gamma_{\ell+1}D_-^{(4)} + R_{\ell+1}^m S_{\ell+1} - R_{\ell+1}^{m+1}P_\ell)\Bigg]\Bigg\}$$

$$- R_n^{m+1}\Bigg\{A_-^{(0)} \pm \sum_{\ell=-\infty}^{n-1}\Bigg[R_\ell^{m+1}\{R_{\ell-2}^{m+1}A_-^{(4)} - R_{\ell-2}^m\gamma_{\ell-2}D_-^{(4)}$$

$$+ R_{\ell-1}^{m+1}S_{\ell-2} - R_{\ell-1}^m P_{\ell-1}\} - R_\ell^m[R_{\ell+2}^m A_-^{(4)} - R_{\ell+2}^{m+1}\gamma_{\ell+1}D_-^{(4)}$$

$$+ R_{\ell+1}^m S_{\ell+1} - R_{\ell+1}^{m+1}P_\ell]\Bigg]\Bigg\} \qquad (2.16)_*$$

where

$$S_n = A_-^{(2)} + A_-^{(4)}F_n + D_-^{(4)}\sum_{j=-\infty}^{n}H_j,\quad P_n = (D_-^{(2)} + \sum_{j=-\infty}^{n}[A_-^{(4)}E_j + D_-^{(4)}G_j]\eta_j)\gamma_n,$$

$$\gamma_n = \prod_{i=-\infty}^{n}(\delta_i^{m+1}/\delta_i^m),\quad \delta_i^m = 1 \mp R_i^m,$$

$$\eta_n = \gamma_n^{-1}/\delta_n^m,\quad H_n = \pm\{R_n^m R_{n+1}^{m+1}\delta_n^{m+1} - R_{n-1}^m R_n^{m+1}\delta_n^m\}\beta_{n-1},$$

$$\beta_n = \gamma_n/\delta_{n+1}^m,\quad F_n = \mp[R_{n+1}^{m+1}R_n^{m+1} - \sum_{j=-\infty}^{n}\Delta^m(R_j^m R_{j+1}^m)],$$

$$G_n = \mp(R_n^{m+1}R_{n+1}^{m+1} - R_n^m R_{n-1}^m)\gamma_{n-1}\delta_n^{m+1},$$

$$E_n = \mp(R_n^m R_{n-1}^{m+1}\delta_n^{m+1} - R_{n+1}^m R_n^{m+1}\delta_n^m). \qquad (2.17)$$

In the limit as m+1→m, equation $(2.16)_*$ becomes

$$R_{nt} = (A_-^{(4)}-D_-^{(4)})(R_{n+2}-R_{n-2}) + (A_-^{(2)}-D_-^{(2)})(R_{n+1}-R_{n-1})$$

$$\pm (A_-^{(4)}-D_-^{(4)})[R_{n-1}^2(R_n+R_{n-2}) - R_{n+1}^2(R_n+R_{n+2}) - R_n^2(R_{n+2}-R_{n-2})]$$

$$\pm (A_-^{(2)}-D_-^{(2)})\, R_n^2(R_{n-1}-R_{n+1})$$

$$+ (A_-^{(4)}-D_-^{(4)})\, R_n[R_{n+1}\, R_n(R_{n+1}R_n+R_{n+1}R_{n+2}+R_nR_{n-1})$$

$$- R_nR_{n-1}(R_nR_{n-1}+R_nR_{n+1}+R_{n-2}R_{n-1})] \qquad (2.18)$$

Let $R = \Delta x\, U$, and by a proper choice (see below) of the constants and taking limit as $\Delta x \to 0$ in equation (2.18) yields the MKdV equation

$$U_t \mp 6U^2U_x + U_{xxx} = 0 \qquad (2.19)$$

From equation $(2.16)_*$ let us consider the linear part which can be written as

$$U_n^{m+1} - U_n^m = (U_n^m-U_n^{m+1})A_-^{(0)} + (U_{n+1}^m-U_{n-1}^{m+1})A_-^{(2)}$$

$$+ (U_{n-1}^m-U_{n+1}^{m+1})D_-^{(2)} + (U_{n+2}^m-U_{n-2}^{m+1})A_-^{(4)} +$$

$$+ (U_{n-2}^m-U_{n+2}^{m+1})D_-^{(4)} \qquad (2.20)$$

To choose the constants, we require a scheme of order $0((\Delta t)^2,(\Delta x)^2)$, (expanding U^m_{n+1}, U^{m+1}_n, ... in Taylor series). With this requirement we find

$$A_-^{(2)} = -\frac{2}{3}A_-^{(0)} + \frac{1}{2}\alpha = D_-^{(-2)}$$

$$D_-^{(2)} = -\frac{2}{3}A_-^{(0)} - \frac{1}{2}\alpha = A_-^{(-2)} \qquad (2.21)$$

$$A_-^{(4)} = \frac{1}{6}A_-^{(0)} \quad \frac{1}{4}\alpha = D_-^{(-4)}$$

$$D_-^{(4)} = \frac{1}{6}A_-^{(0)} + \frac{1}{4}\alpha = A_-^{(-4)}$$

where

$$\alpha = \frac{\Delta t}{(\Delta x)^3}$$

and

$$A_-^{(0)} = \text{arbitrary constant.}$$

In order to get a local scheme of order $0((\Delta t)^2,(\Delta x)^2)$ for the MKdV equation from $(2.16)_*$, let $R^m_n = \Delta x\, U^m_n$, keep the terms through order $0((\Delta x)^3)$ and then drop the sum terms of the form

$$\sum_{j=-\infty}^{n} A^m(U^m_j U^m_{j-1}),\quad \sum_{j=-\infty}^{n}[(U^m_j)^2-(U^{m+1}_j)^2]$$

and replace γ_n by 1. Equation $(2.16)_*$ gives the following local scheme

$$U^{m+1}_n - U^m_n = \{(U^m_{n+2}-U^{m+1}_{n-2})A_-^{(4)} + (U^m_{n-2}-U^{m+1}_{n+2})D_-^{(4)}$$

$$+ (U^m_{n+1}-U^{m+1}_{n-1})A_-^{(2)} + (U^m_{n-1}-U^{m+1}_{n+1})D_-^{(2)} + (U^m_n-U^{m+1}_n)A_-^{(0)}\}$$

$$\pm (\Delta x)^2\Big[\{U_n^{m+1}[U_n^m U_{n-2}^m + U_{n-1}^m U_{n+1}^m] - U_n^m[U_n^{m+1}U_{n+2}^{m+1} + U_{n+1}^{m+1}U_{n-1}^{m+1}]$$

$$+ U_{n-1}^{m+1}[U_{n-1}^m U_{n-2}^m + U_n^m U_{n-1}^m] - U_{n+1}^m[U_n^{m+1}U_{n+1}^{m+1} + U_{n+1}^{m+1}\,U_{n+2}^{m+1}]\}A_-^{(4)}$$

$$+ \{U_{n+1}^m[U_n^m U_{n+1}^{m+1} + U_{n+1}^m U_{n+2}^{m+1}] + U_n^m[U_n^m U_{n+2}^{m+1} + U_{n+1}^{m+1}\,U_{n-1}^m]$$

$$- U_n^{m+1}[U_n^{m+1}U_{n-2}^m + U_{n+1}^{m+1}U_{n-1}^m] - U_{n-1}^{m+1}[U_{n-1}^{m+1}U_{n-2}^m + U_n^{m+1}U_{n-1}^m]\}D_-^{(4)}$$

$$+ \{U_n^m U_n^{m+1}[U_{n-1}^m - U_{n+1}^{m+1}]A_-^{(2)} + [(U_n^m)^2 U_{n+1}^{m+1} - (U_n^{m+1})^2 U_{n-1}^m]D_-^{(2)}\}\Big] \tag{2.22}_*$$

where $A_-^{(4)}, \ldots, D_-^{(-4)}$ satisfy equation (2.21). Equation $(2.22)_*$ is consistent with the MKdV equation (2.19), with the truncation error of order $0((\Delta t)^2, (\Delta x)^2)$. This truncation error holds also for the full scheme given in equation $(2.16)_*$. Since $A_-^{(0)}$ is an arbitrary constant, we have a family of schemes, each one of which satisfies the properties discussed earlier for the NLS equation scheme.

For the special case associated with the KdV equation

$$U_t + 6UU_x + U_{xxx} = 0 \tag{2.23}$$

we let $Q_n^m = R_n^m = 0$, $T_n^m = 1$. The four compatibility equations for $A_n^m, \ldots, D_n^m$ are given by

$$z\,\frac{A_{n+1}^m}{1-S_n^m} - z\,\frac{A_n^m}{1-S_n^{m+1}} + \frac{z}{1-S_n^m}\,B_{n+1}^m - \frac{S_n^m C_n^m}{z(1-S_n^{m+1})} = z\,\frac{S_n^{m+1}-S_n^m}{(1-S_n^m)(1-S_n^{m+1})} \tag{2.24}$$

$$\frac{S_n^m}{z(1-S_n^m)}\,A_{n+1}^m + \frac{B_{n+1}^m}{z(1-S_n^m)} - z\,\frac{B_n^m}{1-S_n^{m+1}} - \frac{S_n^{m+1}}{z}\,\frac{D_n^m}{(1-S_n^{m+1})} = \frac{S_n^{m+1}-S_n^m}{z(1-S_n^{m+1})(1-S_n^m)} \tag{2.25}$$

$$\frac{z}{1-S_n^m} C_{n+1}^m + \frac{z}{1-S_n^m} D_{n+1}^m - z \frac{A_n^m}{1-S_n^{m+1}} - \frac{C_n^m}{z(1-S_n^{m+1})} = z \frac{(S_n^{m+1}-S_n^m)}{(1-S_n^{m+1})(1-S_n^m)} \tag{2.26}$$

$$\frac{D_{n+1}^m}{z(1-S_n^m)} + \frac{S_n^m}{z} \frac{C_{n+1}^m}{(1-S_n^m)} - \frac{D_n^m}{z(1-S_n^{m+1})} - \frac{zB_n^m}{(1-S_n^{m+1})} = \frac{S_n^{m+1}-S_n^m}{z(1-S_n^{m+1})(1-S_n^m)} \tag{2.27}$$

Using the ideas in [13,15,16], the coefficients in the equations for the time dependence of the eigenfunctions are expanded as

$$A_n^m = \sum_{k=-2}^{2} z^{2k} A_n^{(2k)}, \quad B_n^m = \sum_{k=-2}^{2} z^{2k} B_n^{(2k)}$$
$$C_n^m = \sum_{k=-2}^{2} z^{2k} C_n^{(2k)}, \quad D_n^m = \sum_{k=-2}^{2} z^{2k} D_n^{(2k)} \tag{2.28}$$

With the expanded form of A_n^m, B_n^m, C_n^m, D_n^m, equations (2.24), (2.25), (2.26), and (2.27) yield a sequence of twenty-four equations in twenty unknowns corresponding to equating powers of z^5, z^3, z, z^{-1}, z^{-3}, z^{-5}, all of which must be independently satisfied. Twenty equations of which give the values of the twenty unknowns (see Taha [2]). The remaining four equations, two of them are trivially satisfied and the third is satisfied under the following consistency conditions

$$A_-^{(i)} = D_-^{(-i)}, \quad i = 4, 2, 0, -2, -4. \tag{2.29}$$

The fourth equations gives the following evolution equation:

$$\frac{S_n^m}{1-S_n^m}\Big\{A_-^{(0)} - \sum_{\ell=-\infty}^{n} [E_{\ell+1} + S_\ell^{m+1} W_\ell (A_-^{(2)} + C_{\ell-2})$$
$$-\{D_-^{(4)} \gamma_{\ell-1} + D_-^{(2)} + \sum_{k=-\infty}^{\ell-1} (H_k + G_k)\} S_\ell^{m+1} \gamma_\ell + (\gamma_\ell - 1)] W_\ell^{-1}\Big\} W_n$$

$$- \frac{S_n^{m+1}}{1-S_n^{m+1}} \Bigg\{ D_-^{(0)} + \sum_{\ell=-\infty}^{n-1} \Big[\frac{-S_\ell^m}{S_{\ell+1}^{m+1}} \{ \gamma_{\ell+1}^{-1} N_{\ell+1} - N_\ell + M_\ell$$

$$+ S_{\ell+1}^{m+1} Z_\ell - S_{\ell+1}^m \gamma_{\ell+1}^{-1} N_{\ell+1} \} + \gamma_\ell T_{\ell-2} + (\gamma_\ell - 1) \Big] W_\ell^{-1} \Bigg) W_{n-1}$$

$$+ \frac{1}{1-S_n^m} E_{n+1} - \frac{1}{1-S_n^{m+1}} T_{n-2} = \frac{S_n^{m+1} - S_n^m}{(1-S_n^{m+1})(1-S_n^m)} \qquad (2.30)_*$$

where

$$E_n = A_-^{(2)} S_n^m W_{n-1} - S_n^{m+1} D_-^{(2)} + H_n + G_n - S_n^{m+1} \sum_{k=-\infty}^{n} (H_k + G_k)$$

$$+ S_n^m W_{n-1} C_{n-1} - S_n^m D_-^{(4)}, \quad C_n = A_-^{(4)} + \sum_{j=-\infty}^{n} P_j W_j^{-1},$$

$$T_n = \gamma_{n+1} M_n + S_{n+1}^{m+1} \gamma_{n+1} Z_n - S_{n+1}^m N_{n+1}, \quad M_n = S_n^{m+1} W_n A_-^{(4)} - S_n^m D_-^{(4)},$$

$$Z_n = (A_-^{(2)} + \sum_{j=-\infty}^{n} Q_j W_j^{-1}) W_n, \quad N_n = D_-^{(2)} + \sum_{j=-\infty}^{n} F_j,$$

$$W_n = \prod_{i=-\infty}^{n} \gamma_i, \quad \gamma_i = \Big(\frac{1-S_i^m}{1-S_i^{m+1}} \Big),$$

$$H_k = A_-^{(4)} (S_{k+1}^m \gamma_k - S_k^m) W_{k-1}, \quad G_k = (S_k^m - S_{k+1}^{m+1}) D_-^{(4)},$$

$$F_j = A_-^{(4)} (S_j^{m+1} W_j - S_{j-1}^{m+1} W_{j-1}) + D_-^{(4)} (S_{j-1}^m - S_j^{m+1}),$$

$$P_j = A_-^{(4)} (S_j^{m+1} - S_{j+1}^m) W_j + D_-^{(4)} (S_{j+1}^{m+1} - S_j^{m+1} \gamma_j),$$

$$Q_j = (S_{j-1}^{m+1} - S_j^m) W_j A_-^{(4)} - (S_{j-1}^m \gamma_j - S_j^m) D_-^{(4)}. \qquad (2.31)$$

In the limit as m+1→m, equation $(2.30)_*$ becomes

$$(S_{n+1}-S_{n-1})\alpha+\gamma\{S_{n+2}-S_{n+1}^2-S_{n+1}S_{n+2}-S_nS_{n+1}+S_{n-1}^2+$$

$$+ S_{n-1}S_{n-2}-S_{n-2}+S_n S_{n-1}\} = \frac{S_{nt}}{1-S_n} \qquad (2.32)$$

where

$$\alpha = A_-^{(2)}-D_-^{(2)} = D_-^{(-2)}-A_-^{(-2)}$$

and

$$\gamma = A_-^{(4)}-D_-^{(4)} = D_-^{(-4)}-A_-^{(-4)}.$$

Let $S_n^m = 1 - e^{(\Delta x)^2 U_n^m}$, and by a proper choice of the constants and taking limit as $\Delta x \to 0$ in equation (2.32) yields the KdV equation (2.23).

To determine the constants in equation $(2.30)_*$ we apply the same argument as in the MKdV equation case and it turns out that the constants have the same values as given in (2.21).

In order to get a local scheme of order $0((\Delta t)^2, (\Delta x)^2)$ for the KdV equation from equation $(2.30)_*$, we follow a similar procedure to that of the MKdV equation. We can establish the following:

$$S_n^{m+1} - S_n^m = (S_n^m-S_n^{m+1})A_-^{(0)}+A_-^{(2)}(S_{n+1}^m-S_{n-1}^{m+1})+D_-^{(2)}(S_{n-1}^m-S_{n+1}^{m+1})$$

$$+ A_-^{(4)}(S_{n+2}^m-S_{n-2}^{m+1})+D_-^{(4)}(S_{n-2}^m-S_{n+2}^{m+1})$$

$$+[(S_n^m)^2-(S_n^{m+1})^2]A_-^{(0)}+D_-^{(4)}\{S_{n+1}^{m+1}(S_n^{m+1}+S_{n+1}^{m+1}+S_{n+2}^{m+1})\}$$

$$- A_-^{(4)}\{S_{n+1}^m(S_n^m+S_{n+1}^m+S_{n+2}^m)\}-D_-^{(4)}\{S_{n-1}^m(S_n^m+S_{n-1}^m+S_{n-2}^m)\}$$

$$+ A_-^{(4)}\{S_{n-1}^{m+1}(S_n^{m+1}+S_{n-1}^{m+1}+S_{n-2}^{m+1})\} \qquad (2.33)_*$$

with

$$S_n^m = 1 - e^{(\Delta x)^2 U_n^m}.$$

Equation $(2.33)_*$ is consistent with the KdV equation with truncation error of order $0((\Delta t)^2, (\Delta x)^2)$ as is the full scheme given in equation $(2.30)_*$.

As in the case of the MKdV equation, we have a family of schemes for the KdV equation and each one of them satisfies the properties given for the NLS equation scheme.

It is worth mentioning that the partial-difference equation for the KdV equation also can be deduced from the discrete Schrödinger equation

$$a_n^m V_{n+1}^m + V_{n-1}^m = \lambda V_n^m \tag{2.34}$$

with an assumed time dependence of the form

$$\Delta^m V_n^m = \tilde{A}_n^m V_{n+1}^m + \tilde{B}_n^m V_n^m \tag{2.35}$$

and expanding $\tilde{A}_n^m$, $\tilde{B}_n^m$ in powers of λ as follows

$$\tilde{A}_n^m = \tilde{A}_n^{(3)} \lambda^3 + \tilde{A}_n^{(1)} \lambda,$$

and (2.36)

$$\tilde{B}_n^m = \lambda^4 \tilde{B}_n^{(4)} + \lambda^2 \tilde{B}_n^{(2)} + \lambda^{(0)} \tilde{A}_n^{(0)}.$$

3. Conclusions.

The partial difference equations which we discussed are consistent with certain important partial differential equations (NLS, MKdV, KdV).

It can be shown that the solutions to the difference equations converge to the solutions of the corresponding partial differential equations. The partial difference equation maintains the joint x,t symmetry of the original partial differential equation. The partial difference equations suggest local schemes which still maintain the joint x,t symmetry of the original equation.

ACKNOWLEDGEMENTS

This research was partially supported by the Air Force Office of Scientific Research, under Grant No. 78-3674-C and the Office of Naval Research under Grant No. N00014-76-C-0867.

REFERENCES

[1] Zabusky, N., A synergetic approach to problems of nonlinear wave propagation and interaction in Nonlinear Partial Differential Equations (W.F. Ames, Ed.) Academic, New York, 1967, pp. 223-258.

[2] Taha, T., Ph.D. Thesis, Clarkson College, Potsdam, New York, 1982.

[3] Ablowitz, M., and Segur, H., "Solitons and the Inverse Scattering Transform", SIAM, Philadelphia, 1981.

[4] Gardner, C., Greene, J., Kruskal, M., Miura, R., Phys. Rev. Lett. 19, 1095 (1967).

[5] Gardner, C., et al., C.P.A.M., 27 (1974).

[6] Lax, P., C.P.A.M. 21, 467 (1968).

[7] Zakharov, V., and Shabat, A., Sov. Phys. - JETP 34, 62 (1972).

[8] Ablowitz, M., Kaup, D., Newell, A., and Segur, H., Stud. Appl. Math. 53, 249 (1974).

[9] Case, K., and Kac, M., J. Math. Phys. 14, 594 (1973).

[10] Case, K., J. Math. Phys. 14, 916 (1973).

[11] Flaschka, H., Prog. Theor. Phys. 51, 703 (1974).

[12] Manakov, S., Phys. JETP 40, 269 (1975).

[13] Ablowitz, M., and Ladik, J., J. Math. Phys. 16, 598 (1975).

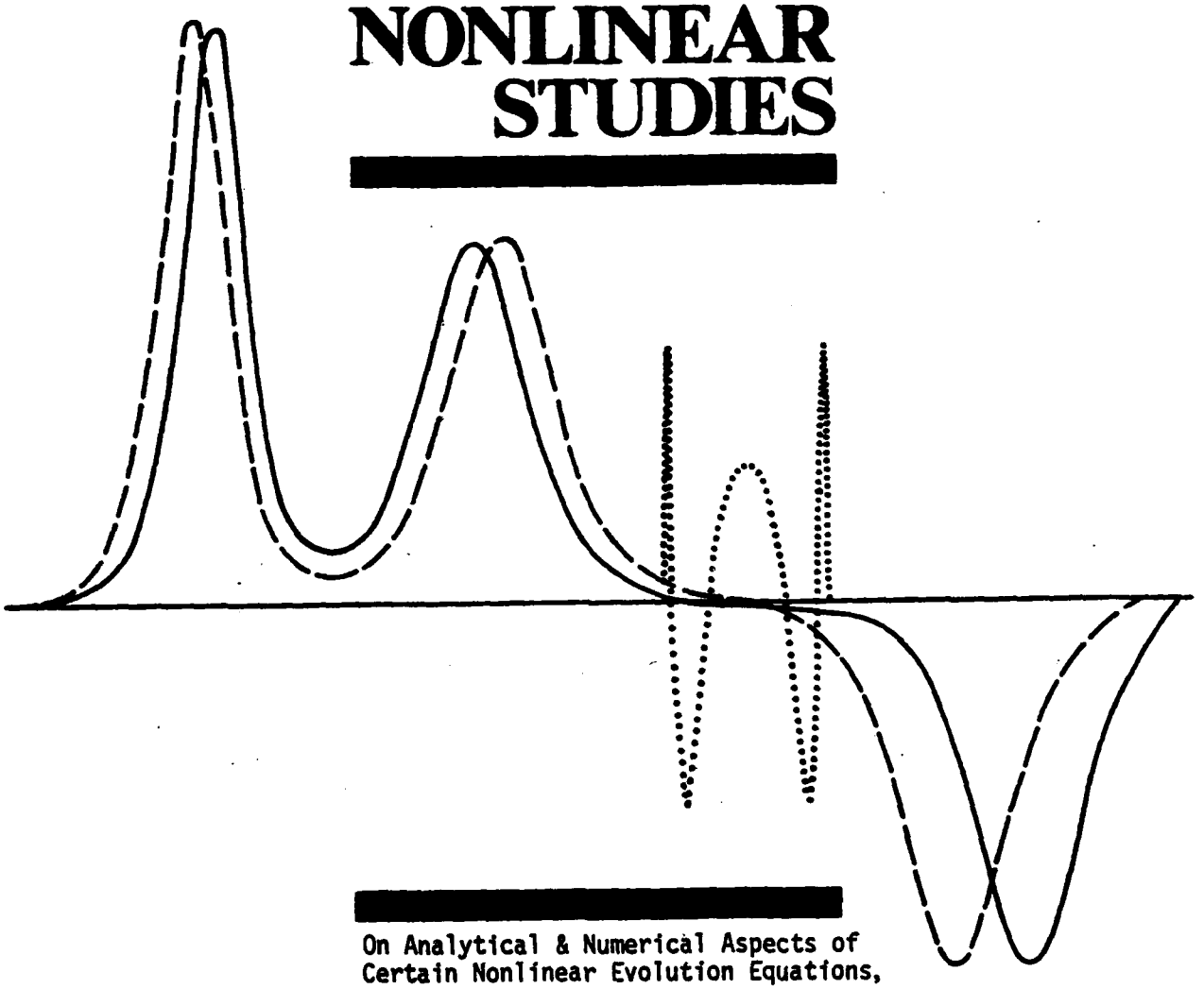
[14] Ablowitz, M., and Ladik, J., Stud. Appl. Math. 55, 213 (1976).

[15] Ablowitz, M., and Ladik, J., J. Math. Phys. 17, 1011 (1976).

[16] Ablowitz, M., and Ladik, J., Stud. Appl. Math. 57, 1-12 (1977).

INSTITUTE FOR NONLINEAR STUDIES

I.F.N.S. #15



On Analytical & Numerical Aspects of Certain Nonlinear Evolution Equations,

Part II: Numerical, Nonlinear Schrödinger Equation

by

Thiab R. Taha and Mark J. Ablowitz

Clarkson College of Technology
Potsdam, New York 13676

ON ANALYTICAL AND NUMERICAL ASPECTS OF CERTAIN NONLINEAR EVOLUTION EQUATIONS, PART II: NUMERICAL, NONLINEAR SCHRÖDINGER EQUATION

ABSTRACT

Various numerical methods were used in order to approximate the nonlinear Schrödinger equation, namely: (i) The classical explicit method, (ii) Hopscotch method, (iii) Implicit-Explicit method, (iv) Crank-Nicolson implicit scheme, (v) The Ablowitz-Ladik scheme, (vi) The split step Fourier method (F. Tappert) (vii) Pseudospectral (Fourier) method (Fornberg and Whitham). Comparisons between the Ablowitz-Ladik scheme, which was developed using notions of the Inverse Scattering Transform, and the other utilized schemes are obtained.

1. Introduction.

The nonlinear Schrödinger (NLS) equation describes a wide class of physical phenomena (e.g. modulational instability of water waves, propagation of heat pulses in anharmonic crystals, helical motion of a very thin vortex filament, nonlinear modulation of collisionless plasma waves, self-trapping of a light beam in a color-dispersive system [1]). The NLS equation was investigated numerically by Karpman and Krushkal [2], Yajima and Outi [3], and Satsuma and Yajima [4], Tappert [5] and Hardin and Tappert [6]. In the latter two works the NLS equation was integrated by the split-step Fourier method. As discussed in part I Ablowitz and Ladik [7] found nonlinear partial difference equations (based on the inverse scattering transform) which can be used as a numerical scheme for the NLS equation. This scheme has certain desirable properties [8] (see part I).

This work aims to compare the Ablowitz and Ladik scheme and other known numerical methods for the NLS equation

$$iq_t = q_{xx} + 2|q|^2 q \tag{1.1}$$

Roughly speaking numerical methods for obtaining solutions to initial value problems fall into two categories [9]: (1) finite difference methods and (2) function approximation methods. For the finite difference methods we seek approximations q_n^m to the original function $q(x,t)$ at a set of points x_n, t_m on a rectangular grid in the x,t plane, where $x_n = hn$, $t_m = km$, h is the increment in x, and k is the increment in t. By expanding function values at grid points in a Taylor series, approximations to the differential equation involving algebraic relations between grid point values can be obtained. The function approximation method approximates the exact solution $q(x,t)$ by an approximate solution defined on a finite dimensional subspace

$$q(x,t) \approx \tilde{q}(x,t) = \sum_{i=1}^{n} C_i(t)\phi_i(x) \tag{1.2}$$

The $\phi_i(x)$ are appropriately chosen basis functions. Common choices for these are the trigonometric functions, leading to a finite Fourier transform or pseudospectral method and piecewise polynomial functions with a local basis, giving the Finite Element Method.

The following numerical methods were applied to the NLS equation:

1. Finite difference methods.
   - a) Explicit methods.
     - i) The classical explicit method.
     - ii) The Hopscotch method [10].
   - b) Implicit methods.
     - i) Implicit for the linear part and explicit for the nonlinear part (Implicit-Explicit).
     - ii) Crank-Nicolson implicit scheme.
     - iii) The Ablowitz and Ladik scheme.

2. Finite Fourier transform or *pseudospectral methods.*

i) Split step Fourier method [6].

ii) Pseudospectral method by Fornberg and Whitham [11].

In order to compare schemes, our approach for comparison is to (a) fix the accuracy (L_∞) for computations beginning at t = 0 and ending at t = T; (b) leave other parameters free (e.g. Δt or Δx) and compare the computing time required to attain such accuracy for various choices of the parameters [12].

These methods are applied to the NLS equation (1.1) subject to the following conditions:

(A) The initial conditions

i) 1-soliton *solution*

The exact solution of (1.1) on the infinite interval is

$$q(x,t) = 2\eta e^{-i\{2\xi x-4(\xi^2-\eta^2)t+(\phi_0+\pi/2)\}} \operatorname{sech}(2\eta x-8\xi\eta t-x_0) \tag{1.3}$$

where x_0, η, ξ and ϕ_0 are constants.

For initial conditions, equation (1.3) is used at t = 0, and the constants are chosen to be $x_0 = 0$, $\phi_0 = 0$, $\xi = 1$, and $\eta = 0.5$, 1, 2, and 3.

ii) Collisions of two solitons [13]

The exact solution of (1.1) on the infinite interval is

$$q(x,t) = G(x,t)/F(x,t), \tag{1.4}$$

where

$$\begin{aligned} F(x,t) = {} & 1+a(1,1^*)\exp(\eta_1+\eta_1^*)+a(1,2^*)\exp(\eta_1+\eta_2^*) \\ & + a(2,1^*)\exp(\eta_2+\eta_1^*)+a(2,2^*)\exp(\eta_2+\eta_2^*) \\ & + a(1,2,1^*,2^*)\exp(\eta_1+\eta_2+\eta_1^*+\eta_2^*), \end{aligned} \tag{1.5}$$

$$G(x,t) = \exp(\eta_1)+\exp(\eta_2)+a(1,2,1^*)\exp(\eta_1+\eta_2+\eta_1^*) + a(1,2,2^*)\exp(\eta_1+\eta_2+\eta_2^*), \tag{1.6}$$

$$a(i,j^*) = (P_i+P_j^*) \tag{1.7}$$

$$a(i,j) = (P_i-P_j)^2 \tag{1.8}$$

$$a(i^*,j^*) = (P_i^*-P_j^*)^2 \tag{1.9}$$

$$a(i,j,k^*) = a(i,j)a(i,k^*)a(j,k^*), \tag{1.10}$$

$$a(i,j,k^*,\ell^*) = a(i,j)a(i,k^*)a(i,\ell^*)a(j,k^*)a(j,\ell^*)a(k^*,\ell^*), \tag{1.11}$$

where * implies a complex conjugate, and

$$\eta_j = P_j x - \Omega_j t - \eta_j^{(0)}, \quad \Omega_j = iP_j^2 \tag{1.12}$$

where P_j and $\eta_j^{(0)}$ are complex constants relating respectively to the amplitude and to the phase of the ith soliton.

For initial conditions, equation (1.4) is used at t = 0, and three different sets of parameters are studied:

i) $P_1 = 1-0.25i$, $P_2 = 0.5+0.15i$, $\eta_1^{(0)} = -2$ and $\eta_2^{(0)} = 0$

ii) $P_1 = 2-0.5i$, $P_2 = 1+0.75i$, $\eta_1^{(0)} = -2$ and $\eta_2^{(0)} = 1.0$

iii) $P_1 = 4-2i$, $P_2 = 3+i$, $\eta_1^{(0)} = -9.04$, $\eta_2^{(0)} = 2.$

(B) Boundary conditions.

Periodic boundary conditions were used. The period was chosen to be [-20,20] in the case of small amplitudes and [-10,10] in the case of relatively higher amplitudes. See Figures (1.1) and (1.2) below.

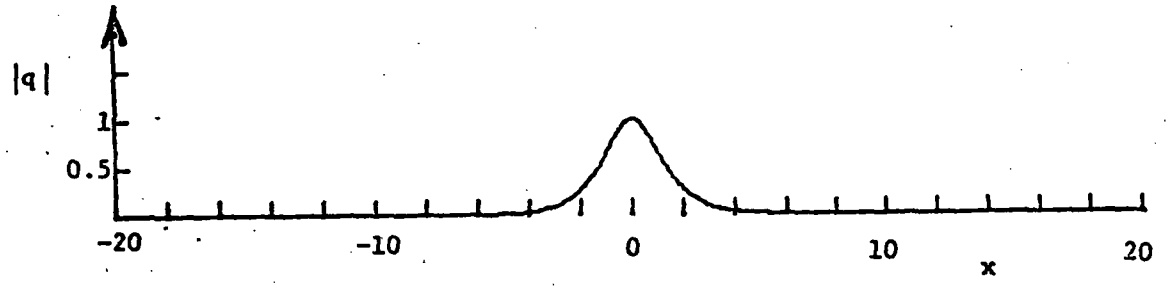


Figure (1.1). Initial condition (1.3), $\eta = 0.5$ (i.e. Amplitude = 1).

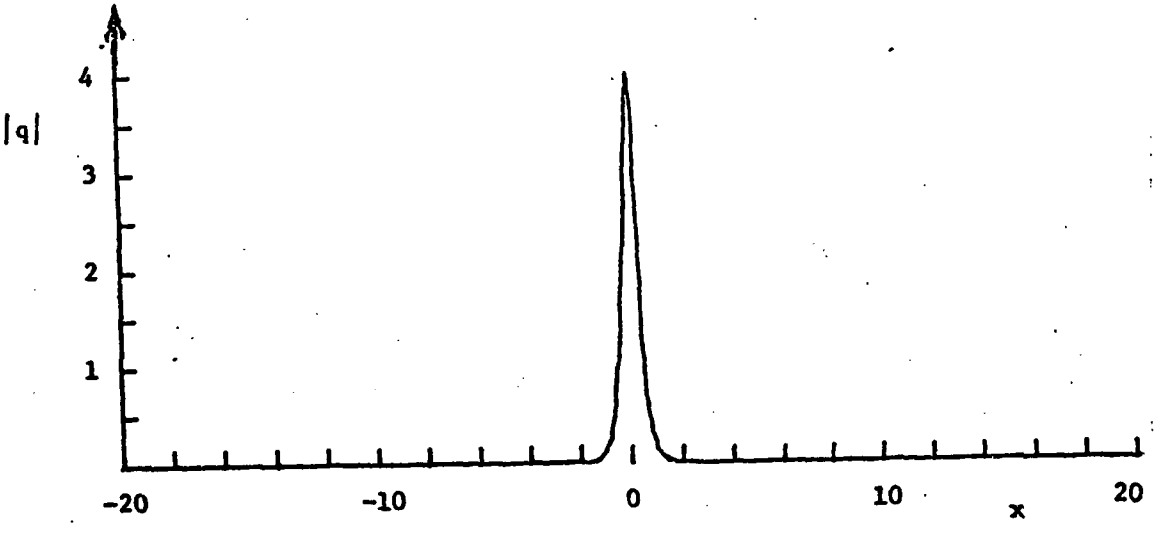


Figure (1.2). Initial condition (1.3), $\eta = 2$ (i.e. Amplitude = 4).

The numerical solution is compared with the exact solution. In addition, two of the conserved quantities were computed, namely, $\int |q|^2 dx$, and $\int (|q|^4 - |\frac{\partial q}{\partial x}|^2) dx$.

2. The Representation of the NLS Equation Using Numerical Methods:

1. Finite difference methods.

A) Explicit methods.

i) The classical explicit method.

Using the classical explicit method with central difference in time (for stability), the finite difference representation of equation (1.1) is

$$i \frac{q_n^{m+1} - q_n^{m-1}}{2\Delta t} = \frac{q_{n+1}^m - 2q_n^m + q_{n-1}^m}{(\Delta x)^2} + 2(q^2)_n^m \, q_n^{m^*} \qquad (2.1)$$

where $|n| < p$, and $m > 0$.

It is easily shown that this method is linearly (dropping the nonlinear term) stable for $\frac{\Delta t}{(\Delta x)^2} \leq \frac{1}{4}$. The truncation error of this scheme is of order $(0((\Delta t)^2) + 0((\Delta x)^2))$.

ii) A Hopscotch scheme.

The NLS equation (1.1) can be approximated by

a) an explicit scheme:

$$i \frac{q_n^{m+1} - q_n^m}{\Delta t} = \frac{q_{n-1}^m - 2q_n^m + q_{n+1}^m}{(\Delta x)^2} + [(|q|_{n-1}^m)^2 \, q_{n-1}^m + (|q|_{n+1}^m)^2 \, q_{n+1}^m] \qquad (2.2)$$

b) an implicit scheme:

$$i \frac{q_n^{m+1} - q_n^m}{\Delta t} = \frac{q_{n+1}^{m+1} - 2q_n^{m+1} + q_{n-1}^{m+1}}{(\Delta x)^2} + [(|q|_{n-1}^{m+1})^2 q_{n-1}^{m+1} + (|q|_{n+1}^{m+1})^2 q_{n+1}^{m+1}] \qquad (2.3)$$

The hopscotch scheme applies equation (2.2) at odd values of (n+m) and (2.3) at even values of (n+m). This combination makes (2.3) explicit. If we

write equation (2.3) for m = m-1, and substitute the resulting equation into (2.2), the explicit scheme (2.2) (after the first time step) may be replaced by

$$q_n^{m+1} = 2q_n^m - q_n^{m-1} \tag{2.4}$$

This scheme has truncation error of order $(O((\Delta t)^2) + O((\Delta x)^2))$, and it is unconditionally stable according to linear analysis.

B) Implicit methods.

i) Implicit-Explicit method.

The Crank-Nicolson scheme was used to approximate the linear part and an explicit average was used for the nonlinear part of the equation (1.1). The scheme is

$$i\frac{q_n^{m+1}-q_n^m}{\Delta t} = \frac{1}{2(\Delta x)^2}\left[q_{n+1}^m-2q_n^m+q_{n-1}^m+q_{n+1}^{m+1}-2q_n^{m+1}+q_{n-1}^{m+1}\right]$$

$$+ (|q|_{n-1}^m)^2 q_{n-1}^m + (|q|_{n+1}^m)^2 q_{n+1}^m \tag{2.5}$$

The scheme is unconditionally stable according to linear analysis. The truncation error is of order $(O((\Delta t)) + O((\Delta x)^2))$. To implement this scheme, a quasi tridiagonal system of equations is required to be solved at each time step. An optimization of the Gaussian elimination method is introduced to solve this system. (See appendix A).

ii) Crank-Nicolson implicit scheme.

The difference scheme for representing equation (1.1) is

$$i\frac{q_n^{m+1}-q_n^m}{\Delta t} = \frac{1}{2(\Delta x)^2}\left[q_{n+1}^m-2q_n^m+q_{n-1}^m+q_{n+1}^{m+1}-2q_n^{m+1}+q_{n-1}^{m+1}\right]$$

$$+ (|q|_n^{m+1})^2 q_n^{m+1} + (|q|_n^m)^2 q_n^m. \tag{2.6}$$

This scheme is also unconditionally stable according to linear analysis. The truncation error of this scheme is of order $(O((\Delta t)^2) + O((\Delta x)^2))$.

We briefly remark on how we solve (2.6). Rewrite equation (2.6) as

$$-\frac{\lambda}{2} q_{n-1}^{m+1} + (i+\lambda) q_n^{m+1} - \frac{\lambda}{2} q_{n+1}^{m+1} = \frac{\lambda}{2} \{q_{n+1}^m + q_{n-1}^m\} + (i-\lambda) q_n^m + \Delta t\{(|q_n^{m+1}|)^2 q_n^{m+1} + (|q_n^m|)^2 q_n^m\}, \quad n = -N, \ldots, N \tag{2.7}$$

and

$$\lambda = \frac{\Delta t}{2(\Delta x)^2}$$

where the solution is sought in the region

$(-N\,\Delta x \le x \le N\,\Delta x) \times (t > m\,\Delta t,\ m = 1,2,\ldots)$.

From the periodic boundary conditions, we have $q_{-N}^m = q_N^m$ and $q_{N+1}^m = q_{-N+1}^m$ for all m. Therefore, by applying equation (2.7) at each mesh point (i.e., $n = -N+1, \ldots, N$), we can write the totality of equations as

$$A\, Q^{m+1} = F \tag{2.8}$$

where

$$A = \begin{bmatrix} i+\lambda & -\frac{\lambda}{2} & & & -\frac{\lambda}{2} \\ -\frac{\lambda}{2} & i+\lambda & -\frac{\lambda}{2} & & \\ & & & & \\ & & -\frac{\lambda}{2} & i+\lambda & -\frac{\lambda}{2} \\ -\frac{\lambda}{2} & \cdots & & -\frac{\lambda}{2} & i+\lambda \end{bmatrix}, \quad Q^{m+1} = \begin{bmatrix} q_{-N+1} \\ q_{-N} \\ \cdot \\ \cdot \\ \cdot \\ q_N \end{bmatrix}^{m+1},$$

and

$$F_j = \frac{\lambda}{2}\{q_{j+1}^m + q_{j-1}^m\} + (i-\lambda)q_j^m + \Delta t\{(|q_j^{m+1}|)^2 q_j^{m+1} + (|q_j^m|)^2 q_j^m\},$$

$$j = -N+1, \ldots, N \qquad (2.9)$$

The right hand side of equation (2.8) is a function of known values of q at the previous time level ($t = m\Delta t$) and unknown values of q at the new time level ($t = (m+1)\Delta t$). We used an iteration technique to solve the system (2.8), and we assumed (only in the right hand side) the values of $q_n^{m+1} = q_n^m$ to start with. Therefore the right hand side becomes known, and we used the same optimization of the Gaussian elimination method used for (2.5) to solve the system (2.8). The resulting values of q at the new time level were substituted in the right hand side of equation (2.8) to start the new iteration, and we solved the new version of (2.8) by the previous method. We iterate until the condition

$$\max_{n=-N,\ldots,N} |q_n^{m+1,k} - q_n^{m+1,k+1}| < \text{tolerance}$$

(where k is the number of iterations) is satisfied. The iteration procedure is repeated at each new time level. The $q_n^{m+1,k+1}$ will be the approximated solution at the point $(n\Delta x, (m+1)\Delta t)$.

C) The Ablowitz and Ladik scheme.

The scheme is

$$i\frac{q_n^{m+1}-q_n^m}{\Delta t} = \frac{1}{2(\Delta x)^2}\left[q_{n+1}^m - 2q_n^m + q_{n-1}^m \prod_{k=-\infty}^{n-1}\Lambda_k^m + q_{n+1}^{m+1}\prod_{k=-\infty}^{n}\Lambda_k^m\right.$$

$$\left. - 2q_n^{m+1} + q_{n-1}^{m+1}\right] + \frac{1}{4}\left[q_n^m\left(q_n^{m^*}q_{n+1}^m + q_n^{m^*+1}q_{n+1}^{m+1}\right.\right.$$

$$\left. + q_n^{m+1}\left(q_{n-1}^m q_n^{m^*} + q_{n-1}^{m+1}q_n^{m^*+1}\right) + 2q_n^m q_n^{m^*}q_{n+1}^{m+1}\prod_{k=-\infty}^{n}\Lambda_k^m\right.$$

$$\left. + 2q_n^{m+1}q_n^{m^*+1}q_{n-1}^m\prod_{k=-\infty}^{n-1}\Lambda_k^m - q_n^m\sum_{k=-\infty}^{n}\Delta_m S_k^m - q_n^{m+1}\sum_{k=-\infty}^{n-1}\Delta_m S_k^{m^*}\right] \tag{2.10}$$

where

$$S_k^m = q_k^m q_{k-1}^{m^*} + q_{k+1}^m q_k^{m^*}, \quad q_n^{m^*+1} = (q_n^{m+1})^*,$$

$$\Lambda_k^m = (1+(\Delta x)^2 q_k^{m+1}q_k^{m^*+1})/(1+(\Delta x)^2 q_k^m q_k^{m^*}),$$

and

$$\Delta_m S_k^m = S_k^{m+1} - S_k^m.$$

This is a global scheme, unconditionally and nonlinearly stable, and the truncation error is of order $(O((\Delta t)^2) + O((\Delta x)^2))$. A local scheme (with the same truncation error) from (2.10) can be obtained if the sum terms are zero and the product terms are equal to one.

It is convenient to implement (2.10) as follows. Write the new time level equation as

$$q_{n+1}^{m+1} - (2+\epsilon)q_n^{m+1} + q_{n-1}^{m+1} = B_n \tag{2.11}$$

where

$$\epsilon = \frac{2i(\Delta x)^2}{\Delta t}, \quad |\epsilon| << 1$$

(Δt is supposed to be of the same order as Δx)

and

$$B_n = -q^m_{n+1} - q^m_{n-1} + (2-\epsilon)q^m_n - [q^m_{n-1}(\prod_{k=-\infty}^{n-1} \Lambda^m_k - 1) + q^{m+1}_{n+1}(\prod_{k=-\infty}^{n} \Lambda^m_k - 1)]$$

$$- \frac{(\Delta x)^2}{2} [q^m_n(q^{m^*}_n q^m_{n+1} + q^{m^*+1}_n q^{m+1}_{n+1}) + q^{m+1}_n(q^m_{n-1} q^{m^*}_n + q^{m+1}_{n-1} q^{m^*+1}_n)$$

$$+ 2q^m_n q^{m^*}_n q^{m+1}_{n+1} \prod_{k=-\infty}^{n} \Lambda^m_k + 2q^{m+1}_n q^{m^*+1}_n q^m_{n-1} \prod_{k=-\infty}^{n-1} \Lambda^m_k$$

$$- q^m_n \sum_{k=-\infty}^{n} \Delta_m S^m_k - q^{m+1}_n \sum_{k=-\infty}^{n-1} \Delta_m S^{m^*}_k], \qquad (2.12)$$

(2.11) is solved by a version of the Crank-Nicolson back and forth sweep method for the heat equation [14,15]. We seek an equation of the form (at the new time level)

$$q_{n+1} = a\, q_n + b_n \qquad (2.13)$$

suitable for computing q explicitly by sweeping to the right. For stability we require $|a| \leq 1$. Repeated substitution into (2.11) to eliminate q_{n+1} and q_n in favor of q_{n-1} gives

$$b_n + [a-(2+\epsilon)]b_{n-1} + [a^2-(2+\epsilon)a+1]q_{n-1} = B_n \qquad (2.14)$$

Requiring the q_{n-1} term to drop out determines a (uniquely since $|a| \leq 1$) as a solution of

$$a^2 - (2+\epsilon)a + 1 = 0 \qquad (2.15)$$

and leaves for b_n a first-order difference equation.

The corresponding homogeneous equation of (2.14) has a solution of the form

$$b_n = k^n \tag{2.16}$$

where the constant k satisfies

$$k + [a-(2+\varepsilon)] = 0 \tag{2.17}$$

It can be shown that the solution k of (2.17) corresponds to the second root (other than $a \leq 1$) of the quadratic equation (2.15) determining a above, and that $|k| > 1$.

It follows that b can be computed explicitly by sweeping to the left,

$$b_{n-1} = ab_n - aB_n \tag{2.18}$$

To obtain the solution q_n, first solve for b_n from (2.18) then use (2.13) to calculate q_n. In order to calculate the b's, we used an iteration procedure. We assumed that $q_n^{m+1} = q_n^m$ in equation (2.12) and $b_N = 0$ to start with, and then we applied the Gauss-Seidel technique [16] (in which the improved values are used as soon as they are computed) to calculate the rest of the b's. The calculated value of the b_N ($=b_{-N}$) was used to start the new iteration, and the iteration procedure was repeated until the condition

$$\max_{n=-N,\ldots,N} |b_{n-1}-(ab_n-aB_n)| < \text{tolerance}$$

is satisfied. Then we used the above procedure by sweeping to the right by means of (2.13) to obtain the q's. After the calculations of the q's,

we substituted their values instead of q_n^{m+1} in equation (2.12), and repeated the same procedure to calculate the b's and then the q's. This procedure was repeated until the condition

$$\max_{n=-N,\ldots,N} |q_n^{m+1,k} - q_n^{m+1,k+1}| < \text{tolerance}$$

(where k is the number of iterations)

is satisfied. The $q_n^{m+1,k+1}$ will be the approximated solution at the point $(n\Delta x, (m+1)\Delta t)$.

2. Finite Fourier transform or pseudospectral methods.

i) Split Step Fourier Method [6].

For convenience the spatial period is normalized to $[0,2\pi]$, then equation (1.1) becomes

$$iq_t = \frac{\pi^2}{P^2} q_{XX} + 2|q|^2 q \tag{2.19}$$

where P is half the length of the interval of interest, and $X = (x+P)\pi/P$. (Here we take P to be 20 or 10 depending on the calculation). This interval is discretized by N equidistant points, with spacing $\Delta X = \frac{2\pi}{N}$. The function q(X,t), numerically defined only on these points, can be transformed to the discrete Fourier space by

$$\hat{q}(k,t) = Fq = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} q(j\Delta X, t) e^{\frac{-2\pi ijk}{N}}, \tag{2.20}$$

$$k = -\frac{N}{2}, \ldots, -1, 0, 1, \ldots, \frac{N}{2} - 1$$

The inversion formula is

$$q(j\Delta X, t) = F^{-1}\hat{q} = \frac{1}{\sqrt{N}} \sum_{k} \hat{q}(k,t) e^{\frac{2\pi ijk}{N}}, \tag{2.21}$$

$$k = -\frac{N}{2}, \ldots, -1, 0, 1, \ldots, \frac{N}{2} - 1$$

These transforms can be performed, efficiently with the fast Fourier transform (FFT) algorithm [17]. Following [6] in order to apply the split step Fourier method for equation (2.19) we (a) advance the solution using only the nonlinear part:

$$i\tilde{q}_t = 2|\tilde{q}|^2\tilde{q}. \tag{2.22}$$

This can be solved exactly,

$$\tilde{q}(X,t) = e^{-2i|q(X,0)|^2 t}q(X,0) \tag{2.23}$$

where $\tilde{q}(X,t)$ is a solution of equation (2.22) and $q(X,0)$ is the solution of equation (2.19) at $t = 0$. (b) advance the solution according to

$$iq_t = \frac{\pi^2}{P^2} q_{XX} \tag{2.24}$$

by means of the discrete Fourier transform

$$q(X_j,t+\Delta t) = F^{-1}(e^{ik^2\Delta t\pi^2/P^2}F(\tilde{q}(X_j,t))) \tag{2.25}$$

This method is second order accurate in Δt and all order in Δx, and is unconditionally stable according to linear analysis.

ii) Pseudospectral method (Fornberg and Whitham) [11].

This is a Fourier (pseudospectral) method in which $q(x,t)$ is transformed into Fourier space with respect to x, and derivatives (or other operators) with respect to x are then made algebraic in the transformed variable. Again for convenience the spatial period is normalized to $[0,2\pi]$. With this scheme, q_{XX} can be evaluated as $F^{-1}\{i^2k^2F(q)\}$. Combined with a leap frog time step the NLS equation (2.19) would then be approximated by

$$q(X,t+\Delta t) = q(X,t-\Delta t) + 2i\Delta t \frac{\pi^2}{p^2} F^{-1}(k^2 F(q(X,t))$$

$$- 4i\Delta t|q|^2 q \qquad (2.26)$$

Using the ideas of Fornberg and Whitham we make a modification in approximating equation (2.19) as follows:

$$q(X,t+\Delta t) = q(X,t-\Delta t) + 2iF^{-1}(\sin(\frac{k^2\pi^2}{p^2}\Delta t)F(q(x,t)))$$

$$- 4i\Delta t|q|^2 q \qquad (2.27)$$

The difference between equations (2.26) and (2.27) is in the approximation of the linear equation (2.24). The linear part of equation (2.27) will be exactly satisfied for any solution of equation (2.24) (see [11]). Also it turns out that equation (2.26) is linearly (dropping the nonlinear term) stable for $\frac{\Delta t}{(\Delta x)^2} < \frac{1}{\pi^2}$, while equation (2.27) is unconditionally stable according to linear analysis.

3. Conclusions.

Various numerical methods are used in order to approximate the NLS equation (1.1), namely; (i) The classical explicit method (2.1), (ii) Hopscotch method (2.2), (2.3), (iii) Implicit-Explicit method (2.5), (iv) Crank-Nicolson implicit scheme (2.6), (v) The Ablowitz and Ladik scheme (2.10), (vi) The split step Fourier method (2.23), (2.25) and the (vii) Pseudo-spectral method of Fornberg and Whitham (2.27). We obtained a comparison between the Ablowitz-Ladik scheme and the other utilized schemes. Our approach for comparison is to (a) fix the accuracy (L_∞) for computations beginning at t = 0 and ending at t = T; (b) leave other parameters free

(e.g. Δt, or Δx), and compare the computing time required to attain such accuracy for various choices of the parameters. For the comparison two sets of initial conditions were studied: (A) 1-soliton solution with different values of the amplitude, (B) Collisions of two solitons with different values of the parameters. According to this approach we have made the following conclusions:

1) The schemes Explicit (i), Implicit-Explicit (iii), and the Hopscotch (ii) took more computing time than the other schemes ((iv), (v), (vi), (vii)) and the difference in the computing time increased as the amplitude increased. The Hopscotch method (ii) took less computing time than the other two methods ((i), (iii)) for the 1-soliton cases, while the explicit (i) method took less computing time than the Hopscotch (ii) and the Implicit-Explicit (iii) methods for the 2-soliton cases.

2) The previous three methods; Explicit (i), Hopscotch (ii), and Implicit-Explicit (iii) do not appear in Tables (3.1), (3.4), and (3.7) since extremely long computing time would be required.

3) The Crank-Nicolson implicit method (iv) took more computing time than the Ablowitz-Ladik (local and global), the split step Fourier method and Fornberg-Whitham method in the case of 1-soliton, and it became comparable with the Ablowitz-Ladik local scheme for high amplitudes. In the case of 2-solitons, Crank-Nicolson took less computing time than the Ablowitz-Ladik local scheme.

4) The Ablowitz-Ladik global scheme took more computing time than the local scheme in the case of small amplitudes, but for high amplitudes it was roughly five times faster than the local, and the Crank-Nicolson schemes.

5) The pseudospectral method is roughly two times faster than the Ablowitz-Ladik local scheme for small amplitudes, but it is much faster for high amplitudes. This method is roughly two times faster than the

Ablowitz-Ladik global scheme for high amplitudes. This method only proved to be faster than the split step Fourier method for small amplitude 2-soliton cases.

6) The split step Fourier method is faster than all the utilized methods for small and large amplitudes for the 1-soliton case. In the average was three times faster than the Fornberg-Whitham method. Also it proved to be faster than the Fornberg-Whitham method for high amplitude 2-soliton cases. The Tables and Figures exhibit the results.

As a conclusion we found that the experiments we conducted that the split step Fourier method was the best method for the NLS equation, followed by the pseudospectral method then the Ablowitz-Ladik global scheme. However we believe that if we were able to go to very high amplitudes (our machine capability prevented this) the Ablowitz-Ladik global scheme would improve dramatically and would prove to be better than the other methods. However it should be noted that the NLS equation is quite unusual in the sense that the nonlinearity is especially simple. This has a dramatic effect in the split step Fourier method - see equation (2.22) - which means that both steps admit to essentially exact methods. Generally this will not be true (see part III). We also note that whereas the Ablowitz-Ladik scheme is $O((\Delta t)^2,(\Delta x)^2)$ the split step Fourier and Fornberg-Whitham methods are of order $O((\Delta t)^2,(\Delta x)^p)$ for all p. (See also the calculations for the KdV equation in part III). It is also worth mentioning that we tried the sweeping technique in implementing the Implicit-Explicit and the Crank-Nicolson methods, and found that it did not affect the overall conclusions.

All the numerical calculations were inspected at every step by using the conserved quantities $\int|q|^2dx$, and $\int(|q|^4-|\frac{\partial q}{\partial x}|^2)dx$ (Table (3.1),..., Table (3.7)). The two conserved quantities were calculated by means of Simpson's Rule [18]. The Ablowitz-Ladik global scheme is the only utilized

scheme which has an infinite number of conserved quantities.

ACKNOWLEDGEMENTS

This research was partially supported by the Air Force Office of Scientific Research, under Grant No. 78-3674-C and the Office of Naval Research under Grant No. N00014-76-C-0867. We are grateful for useful conversations with Dr. H. Segur.

Table (3.1). Comparison of the computing time (E) which is required to attain an accuracy (L_∞) < 0.006 for computations beginning at t=0 and ending at t=1.0, for the numerical methods utilized in solving the NLS equation. Two conserved quantities are shown. 1-soliton as an initial condition with amplitude = 1 on the interval [-20,20].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	q_1	q_2
1.	Explicit	$\Delta x = 0.05$ $\Delta t = 0.000625$	C: 0 E: 7	23 33	0.25 0.5 1.0	0.00159 0.00308 0.00564	2.0 2.00001 2.00001	-7.29268 -7.29263 -7.29256
2.	Implicit - Explicit.	$\Delta x = 0.05$ $\Delta t = 0.001$	C: 0 E: 11	27 45	0.25 0.5 1.0	0.00146 0.00287 0.00577	1.99802 1.99605 1.99214	-7.28059 -7.26865 -7.24494
3.	Implicit (Crank-Nicolson).	$\Delta x = 0.05$ $\Delta t = 0.005$	C: 0 E: 5	32 39	0.25 0.5 1.0	0.00165 0.00320 0.00585	2.0 2.0 1.99999	-7.29262 -7.29257 -7.29247
4.	Hopscotch	$\Delta x = 0.08$ $\Delta t = 0.002$	C: 0 E: 1	23 4	0.25 0.5 1.0	0.00209 0.00321 0.00538	2.0 2.00005 2.00005	-7.22997 -7,23002 -7.23017
5.	Split step Fourier method	$\Delta x = 0.3125$ $\Delta t = 0.02$	C: 0 E: 0	21 11	0.25 0.5 1.0	0.00463 0.00502 0.00466	2.0 2.0 2.0	-5.87181 -5.86954 -5.86991
6.	The Ablowitz and Ladik local scheme	$\Delta x = 0.06$ $\Delta t = 0.0165$	C: 0 E: 2	23 2	0.25 0.5 1.0	0.00157 0.00306 0.00580	2.00002 2.00003 2.00007	-7.27475 -7.27478 -7.27486

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	q_1	q_2
7.	The Ablowitz and Ladik global scheme	$\Delta x = 0.05$ $\Delta t = 0.04$	C: 0 E: 2	29 10	0.25 0.5 1.0	0.00136 0.00272 0.00561	2.00002 2.00003 2.00006	-7.29273 -7.29280 -7.29297
8.	Pseudospectral (Fornberg-Whitham)	$\Delta x = 0.3125$ $\Delta t = 0.0026$	C: 0 E: 0	23 50	0.25 0.5 1.0	0.00513 0.00505 0.00513	2.00001 2.00000 2.00001	-5.87870 -5.87868 -5.87835

Δx = The increment in x.

Δt = The increment in t.

C = The compilation time

q_1 = A conserved quantity of the NLS equation which is $\int |q^2| dx$

q_2 = A conserved quantity of the NLS equation which is $\int (|q|^4 - |\frac{\partial q}{\partial x}|^2)$

$L_\infty = \max |\tilde{q}_n^m - q_n^m|$, $\tilde{q}_n^m$ is the numerical solution and q_n^m is the exact solution at the point $(\Delta x \cdot n, \Delta t \cdot m)$ for all n,m.

Table (3.2). Comparison of the puting time (E) which is req d to attain an accuracy (L_∞) < 0.01 for computations beginning at t=0 and ending at t=1.0, for the numerical methods utilized in solving the NLS equation. Two conserved quantities are shown. 1-soliton as an initial condition with amplitude = 2 on the interval [-20,20].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	q_1	q_2
1.	Explicit	$\Delta x = 0.02$ $\Delta t = 0.0001$	C: 0 E:116	23 50	0.25 0.5 1.0	0.00205 0.00304 0.00931	3.99124 3.99113 3.98251	-10.63542 -10.63554 -10.63641
2.	Implicit - Explicit	$\Delta x = 0.03$ $\Delta t = 0.00022$	C: 0 E: 82	27 18	0.25 0.5 1.0	0.00768 0.00938 0.00759	4.00003 4.00006 4.00010	-10.55766 -10.51480 -10.42746
3.	Implicit (Crank-Nicolson)	$\Delta x = 0.02$ $\Delta t = 0.011$	C: 0 E: 16	32 44	0.25 0.5 1.0	0.00329 0.00522 0.00971	4.0 4.0 4.00001	-10.63713 -10.63725 -10.63757
4.	Hopscotch	$\Delta x = 0.02$ $\Delta t = 0.0004$	C: 0 E: 19	23 42	0.25 0.5 1.0	0.00420 0.00744 0.00963	4.00006 4.00007 4.00006	-10.63652 -10.63615 -10.63643
5.	Split step Fourier method	$\Delta x = 0.1563$ $\Delta t = 0.0048$	C: 0 E: 1	21 26	0.25 0.5 1.0	0.00893 0.00594 0.00464	4.0 4.0 4.0	- 8.93482 - 8.95184 - 8.97181
6.	The Ablowitz and Ladik local scheme	$\Delta x = 0.06$ $\Delta t = 0.03$	C: 0 E: 2	23 28	0.25 0.5 1.0	0.00731 0.00886 0.00695	3.99997 3.99997 3.99998	-10.39668 -10.39565 -10.39722
7.	The Ablowitz and Ladik global scheme	$\Delta x = 0.07$ $\Delta t = 0.012$	C: 0 E: 7	29 33	0.25 0.5 1.0	0.00600 0.0100 0.00937	3.99987 3.99976 3.99986	-10.31049 -10.31347 -10.31216
8.	Pseudospectral (Fornberg-Whitham)	$\Delta x = 0.1563$ $\Delta t = 0.0011$	C: 0 E: 4	23 6	0.25 0.5 1.0	0.00860 0.00860 0.00840	4.00000 4.00001 4.00000	-8.95466 -8.95320 -8.95463

AD-A119 843 CLARKSON COLL OF TECHNOLOGY POTSDAM NY DEPT OF MATHE--ETC F/G 12/1
NONLINEAR WAVES.(U)
SEP 82 M J ABLOWITZ, D J KAUP, A C NEWELL N00014-76-C-0867
UNCLASSIFIED NL

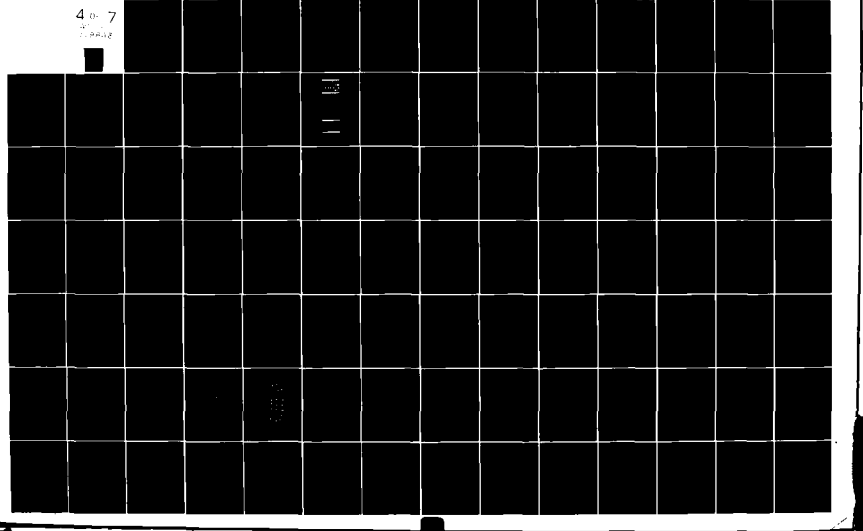


Table (3.3). Comparison of the computing time (E) which is required to attain an accuracy (L_∞) < 0.08 for computations beginning at t=0 and ending at t=1.0, for the numerical methods utilized in solving the NLS equation. Two conserved quantities are shown. 1-soliton as an initial condition with amplitude = 4 on the interval [-10,10].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	q_1	q_2
1.	Implicit (Crank-Nicolson)	$\Delta x = 0.02$ $\Delta t = 0.004$	C: 0 E: 18	32 29	0.25 0.5 1.0	0.01922 0.03751 0.07380	8.00013 8.00029 8.00061	10.95019 10.95269 10.95735
2.	Split step Fourier method	$\Delta x = 0.15625$ $\Delta t = 0.007$	C: 0 E: 0	21 30	0.25 0.5 1.0	0.06552 0.03703 0.05859	7.99743 8.01433 7.99251	22.94076 26.05543 22.94951
3.	The Ablowitz and Ladik local scheme	$\Delta x = 0.02$ $\Delta t = 0.004$	C: 0 E: 16	23 20	0.25 0.5 1.0	0.01852 0.03676 0.07270	8.00012 8.00024 8.00050	10.94663 10.94809 10.95192
4.	The Ablowitz and Ladik global scheme	$\Delta x = 0.06$ $\Delta t = 0.02$	C: 0 E: 4	29 25	0.25 0.5 1.0	0.04613 0.05490 0.07835	7.99647 7.99658 7.99704	12.92892 12.92390 12.93508
5.	Pseudospectral (Fornberg-Whitham)	$\Delta x = 0.15625$ $\Delta t = 0.0021$	C: 0 E: 1	23 2	0.25 0.5 1.0	0.05980 0.03765 0.06619	7.98569 8.02725 8.12957	24.28942 28.07237 28.16859

Table (3.4). Comparison of the computing time (E) which is required to attain an accuracy (L_∞) < 0.15 for computations beginning at t=0 and ending at t=0, for the numerical methods (listed below) utilized in solving the NLS equation. Two conserved quantities are shown. 1-soliton as an initial condition with amplitude = 6 on the interval [-10,10].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	q_1	q_2
1.	Implicit (Crank-Nicolson)	$\Delta x = 0.015$ $\Delta t = 0.00125$	C: 0 E: 56	32 59	0.25 0.5 1.0	0.03858 0.07309 0.14176	11.99968 11.99941 11.99884	96.81015 96.79710 96.77309
2.	Split step Fourier method	$\Delta x = 0.078$ $\Delta t = 0.0045$	C: 0 E: 1	21 32	0.25 0.5 1.0	0.14704 0.12475 0.13180	12.00210 11.99876 11.00092	107.97445 109.93838 108.95444
3.	The Ablowitz and Ladik local scheme	$\Delta x = 0.015$ $\Delta t = 0.00125$	C: 0 E: 57	23 39	0.25 0.5 1.0	0.03977 0.07745 0.15498	11.99938 11.99887 11.99778	96.78147 96.76401 96.71861
4.	The Ablowitz and Ladik global scheme	$\Delta x = 0.06$ $\Delta t = 0.01$	C: 0 E: 11	29 12	0.25 0.5 1.0	0.10067 0.12484 0.14704	11.97469 11.98376 11.98405	105.15448 105.92602 105.80171
5.	Pseudospectral (Fornberg-Whitham)	$\Delta x = 0.078$ $\Delta t = 0.00079$	C: 0 E: 5	23 39	0.25 0.5 1.0	0.04890 0.03489 0.12482	11.99449 11.99514 12.07188	114.73124 117.77290 114.01675

Table (3.5). Comparison of the computing time (E) which is required to attain an accuracy (L_∞) < 0.003 for computations beginning at t=0 and ending at t=2.5, for the numerical methods utilized in solving the NLS equation. Two conserved quantities are shown. Two-solitons as an initial condition with amplitudes 1 and 0.5 respectively, and they are allowed to interact, on the interval [-20,20].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	q_1	q_2
1.	Explicit	$\Delta x = 0.13$ $\Delta t = 0.0036$	C: 0 E: 1	27 1	0.5 1.0 1.6 1.8 2.5	0.00116 0.00141 0.00175 0.00189 0.00265	3.0 3.0 3.0 3.0 3.00001	0.60617 0.60541 0.60500 0.60501 0.60647
2.	Implicit - Explicit	$\Delta x = 0.05$ $\Delta t = 0.0025$	C: 0 E: 9	31 14	0.5 1.0 1.6 1.8 2.5	0.00084 0.00096 0.00124 0.00133 0.00250	3.00253 3.00451 3.00645 3.00705 3.00942	0.60474 0.60589 0.60718 0.60770 0.61114
3.	Implicit (Crank-Nicolson)	$\Delta x = 0.13$ $\Delta t = 0.04$	C: 0 E: 1	41 15	0.5 1.0 1.6 1.8 2.5	0.00116 0.00148 0.00182 0.00198 0.00272	3.00004 3.00007 3.00006 3.00004 2.99973	0.60625 0.60548 0.60505 0.60503 0.60623
4.	Hopscotch	$\Delta x = 0.05$ $\Delta t = 0.001$	C: 0 E: 5	30 58	0.5 1.0 1.6 1.8 2.5	0.00099 0.00141 0.00173 0.00188 0.00251	3.00007 3.00012 3.00014 3.00014. 3.00002	0.60255 0.60214 0.60194 0.60197 0.60288

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	q_1	q_2
5.	Split step Fourier method	$\Delta x = 0.625$ $\Delta t = 0.005$	C: 0 E: 0	26 51	0.5 1.0 1.6 1.8 2.5	0.00086 0.00097 0.00116 0.00125 0.00251	2.99917 2.99933 2.99978 2.99987 3.00212	0.66629 0.64720 0.63920 0.64037 0.69484
6.	The Ablowitz and Ladik local scheme	$\Delta x = 0.07$ $\Delta t = 0.07$	C: 0 E: 1	30 36	0.5 1.0 1.6 1.8 2.5	0.00053 0.00095 0.00142 0.00158 0.00282	2.99950 2.99917 2.99907 2.99915 3.00001	0.60272 0.60196 0.60163 0.60173 0.60344
7.	The Ablowitz and Ladik global scheme	$\Delta x = 0.08$ $\Delta t = 0.045$	C: 0 E: 2	35 13	0.5 1.0 1.6 1.8 2.5	0.00055 0.00104 0.00156 0.00170 0.00278	2.99937 2.99897 2.99878 2.99882 2.99963	0.60280 0.60182 0.60132 0.60136 0.60339
8.	Pseudospectral (Fornberg-Whitham)	$\Delta x = 0.625$ $\Delta t = 0.0071$	C: 0 E: 0	32 26	0.5 1.0 1.6 1.8 2.5	0.00136 0.00122 0.00149 0.00166 0.00266	2.99917 2.99932 2.99977 2.99985 3.00219	0.66571 0.64652 0.63815 0.63914 0.69289

Table (3.6). Comparison of the computing time (E) which is required to attain an accuracy (L_∞) < 0.0062 for computations beginning at t=0 and ending at t=1.0, for the numerical methods utilized in solving the NLS equation. Two conserved quantities are shown. Two solitons as an initial condition with amplitudes 2 and 1 respectively, and they are allowed to interact, on the interval [-20,20].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	q_1	q_2
1.	Explicit	$\Delta x = 0.032$ $\Delta t = 0.000256$	C: 0 E: 19	27 4	0.1 0.25 0.5 0.8 1.0	0.00071 0.00111 0.00192 0.00615 0.00500	6.0 6.0 6.0 6.00001 6.0	3.88835 3.88590 3.88386 3.93854 3.92292
2.	Implicit - Explicit	$\Delta x = 0.03$ $\Delta t = 0.0002$	C: 0 E: 68	31 9	0.1 0.25 0.5 0.8 1.0	0.00321 0.00576 0.00712 0.00615 0.00467	6.00162 6.00405 6.00808 6.01215 6.01629	3.89428 3.90462 3.92248 3.94080 3.95845
3.	Implicit (Crank-Nicolson)	$\Delta x = 0.03$ $\Delta t = 0.012$	C: 0 E: 6	41 58	0.1 0.25 0.5 0.8 1.0	0.00067 0.00093 0.00188 0.00601 0.00605	5.99980 5.99968 5.99927 5.99255 5.99437	3.88606 3.88329 3.87928 3.89186 3.88960
4.	Hopscotch	$\Delta x = 0.015$ $\Delta t = 0.0001$	C: 0 E: 73	30 44	0.1 0.25 0.5 0.8 1.0	0.00086 0.00149 0.00250 0.00855 0.00704	6.00002 6.00005 6.00006 5.99967 5.99976	3.87701 3.87500 3.87371 3.91727 3.90511

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	q_1	q_2
5.	Split step Fourier method	$\Delta x = 0.156$ $\Delta t = 0.001$	C: 0 E: 6	26 55	0.1 0.25 0.5 0.8 1.0	0.00157 0.00168 0.00221 0.00613 0.00598	5.99999 6.0 6.0 6.00006 5.99976	4.17099 4.09769 4.04206 5.66900 5.12597
6.	The Ablowitz and Ladik local scheme	$\Delta x = 0.02$ $\Delta t = 0.01$	C: 0 E: 13	28 24	0.1 0.25 0.5 0.8 1.0	0.00029 0.00072 0.00156 0.00617 0.00381	5.99981 5.99946 5.99929 6.00485 6.00388	3.87884 3.87555 3.87337 3.94477 3.92723
7.	The Ablowitz and Ladik global scheme	$\Delta x = 0.03$ $\Delta t = 0.03$	C: 0 E: 6	35 47	0.1 0.25 0.5 0.8 1.0	0.00057 0.00139 0.00265 0.00620 0.00632	5.99975 5.99932 5.99918 6.00712 6.00523	3.88457 3.87828 3.87377 4.01667 3.97919
8.	Pseudospectral (Fornberg-Whitham)	$\Delta x = 0.156$ $\Delta t = 0.0012$	C: 0 E: 3	32 56	0.1 0.25 0.5 0.8 1.0	0.00202 0.00228 0.00378 0.00612 0.00564	5.99999 5.99999 5.99999 6.00014 5.99981	4.17006 4.09698 4.03485 5.61584 5.16439

Table (3.7). Comparison of the puting time (E) which is req d to attain an accuracy (L_∞) < 0.013 for computations beginning at t=0 and ending at t=1.0, for the numerical methods (listed below) utilized in solving the NLS equation. Two conserved quantities are shown. Two solitons as an initial condition with amplitudes 4 and 3 respectively, on the interval [-10,10].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	q_1	q_2
1.	Implicit (Crank-Nicolson)	$\Delta x = 0.0085$ $\Delta t = 0.001$	C: 0 E: 98	41 25	0.1 0.25 0.5 0.75 1.0	0.00150 0.00276 0.01296 0.00667 0.00797	13.99988 13.99979 13.99719 13.99920 13.99890	22.72355 22.72075 22.82511 22.70690 22.70139
2.	Split step Fourier method	$\Delta x = 0.078$ $\Delta t = 0.0004$	C: 0 E: 17	26 9	0.1 0.25 0.5 0.75 1.0	0.00503 0.00416 0.01363 0.00393 0.00512	14.00002 14.0 14.00540 13.99998 14.0	27.46970 27.48880 48.66320 27.44398 27.45940
3.	The Ablowitz and Ladik local scheme	$\Delta x = 0.007$ $\Delta t = 0.0007$	C: 0 E:168	28 59	0.1 0.25 0.5 0.75 1.0	0.00083 0.00162 0.01283 0.00460 0.00637	13.99996 13.99992 14.00377 13.99979 13.99968	22.70587 22.70600 22.04042 22.70006 22.69799
4.	The Ablowitz and Ladik global scheme	$\Delta x = 0.01$ $\Delta t = 0.005$	C: 0 E: 49	35 55	0.1 0.25 0.5 0.75 1.0	0.00150 0.00193 0.01348 0.00289 0.00406	14.00016 14.00047 14.00825 14.00048 14.00091	22.75041 22.75688 23.42924 22.74976 22.75561
5.	Pseudospectral (Fornberg-Whitham)	$\Delta x = 0.078$ $\Delta t = 0.0002$	C: 0 E: 22	32 45	0.1 0.25 0.5 0.75 1.0	0.00641 0.00630 0.01301 0.00758 0.01024	14.0002 14.00001 14.00552 13.99997 14.00001	27.45347 27.50876 47.33651 27.44128 27.45570

Figure (3.1). Displays the computing time (E) which is required by each utilized method given in Table (3.1). 1-soliton, amplitude = 1.

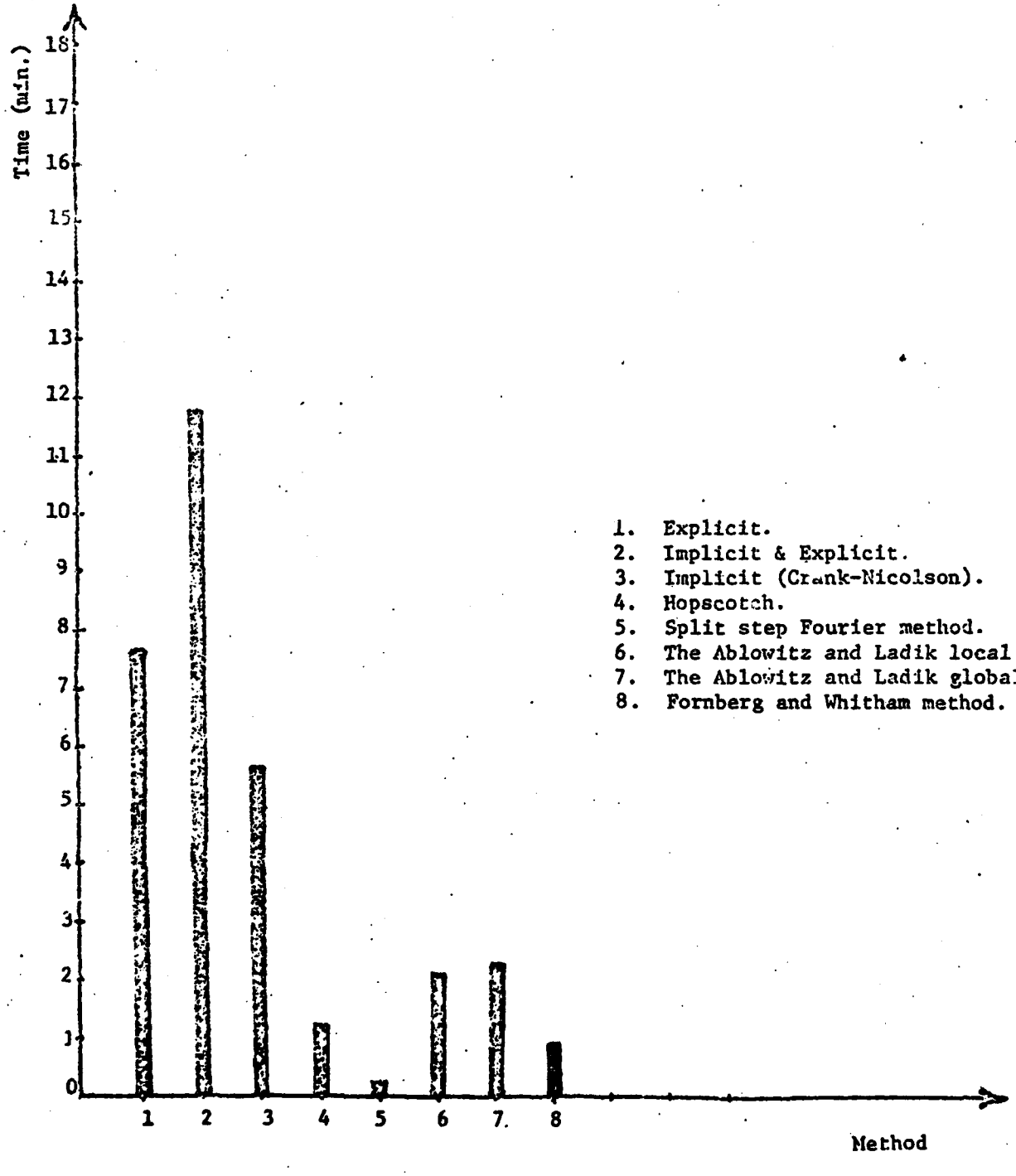


Figure (3.2). Displays the computing time (E) which is required by each utilized method given in Table (3.2). 1-soliton, amplitude = 2.

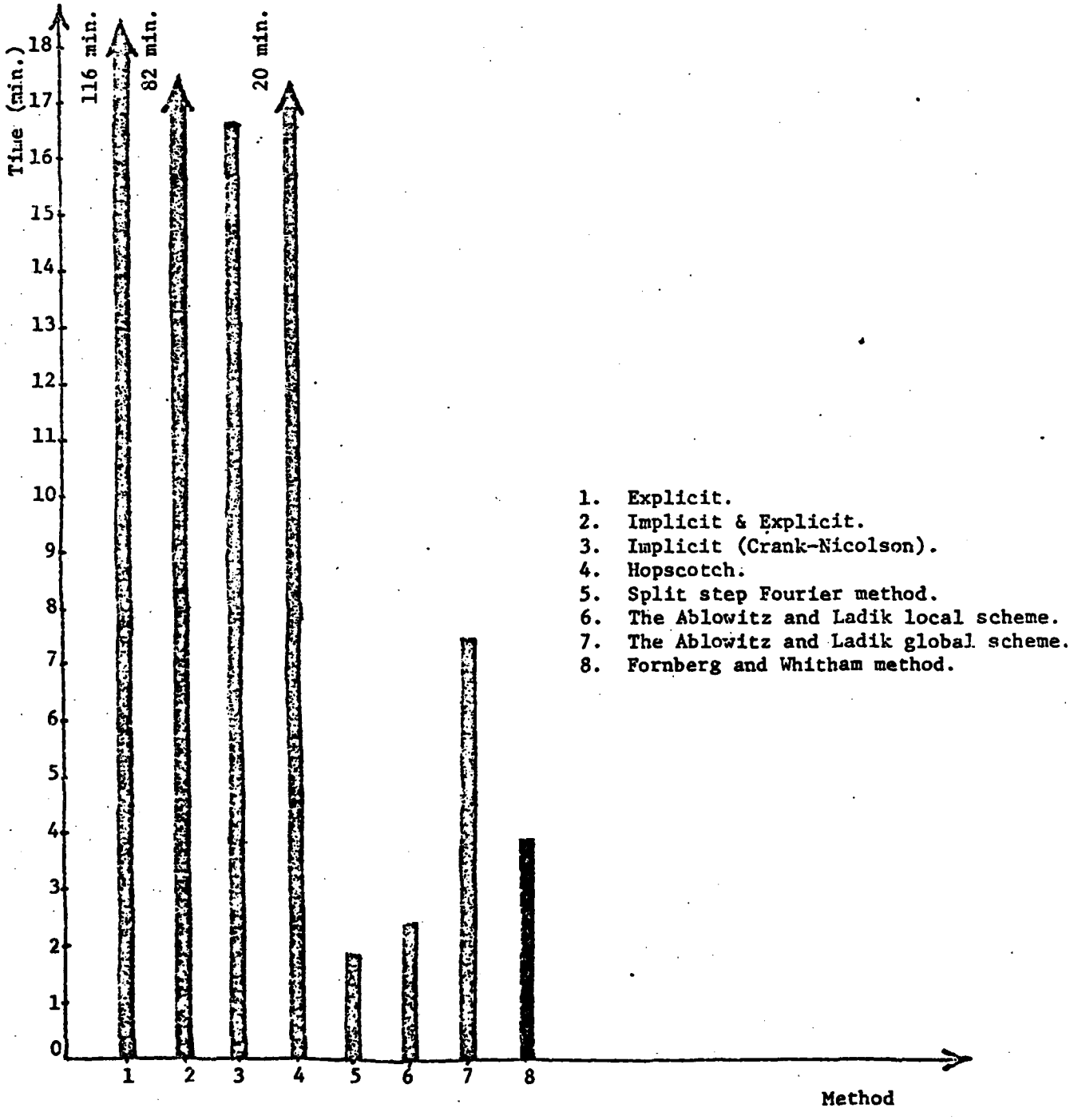


Figure (3.3). Displays the computing time (E) which is required by each utilized method given in Table (3.3). 1-soliton, amplitude = 4.

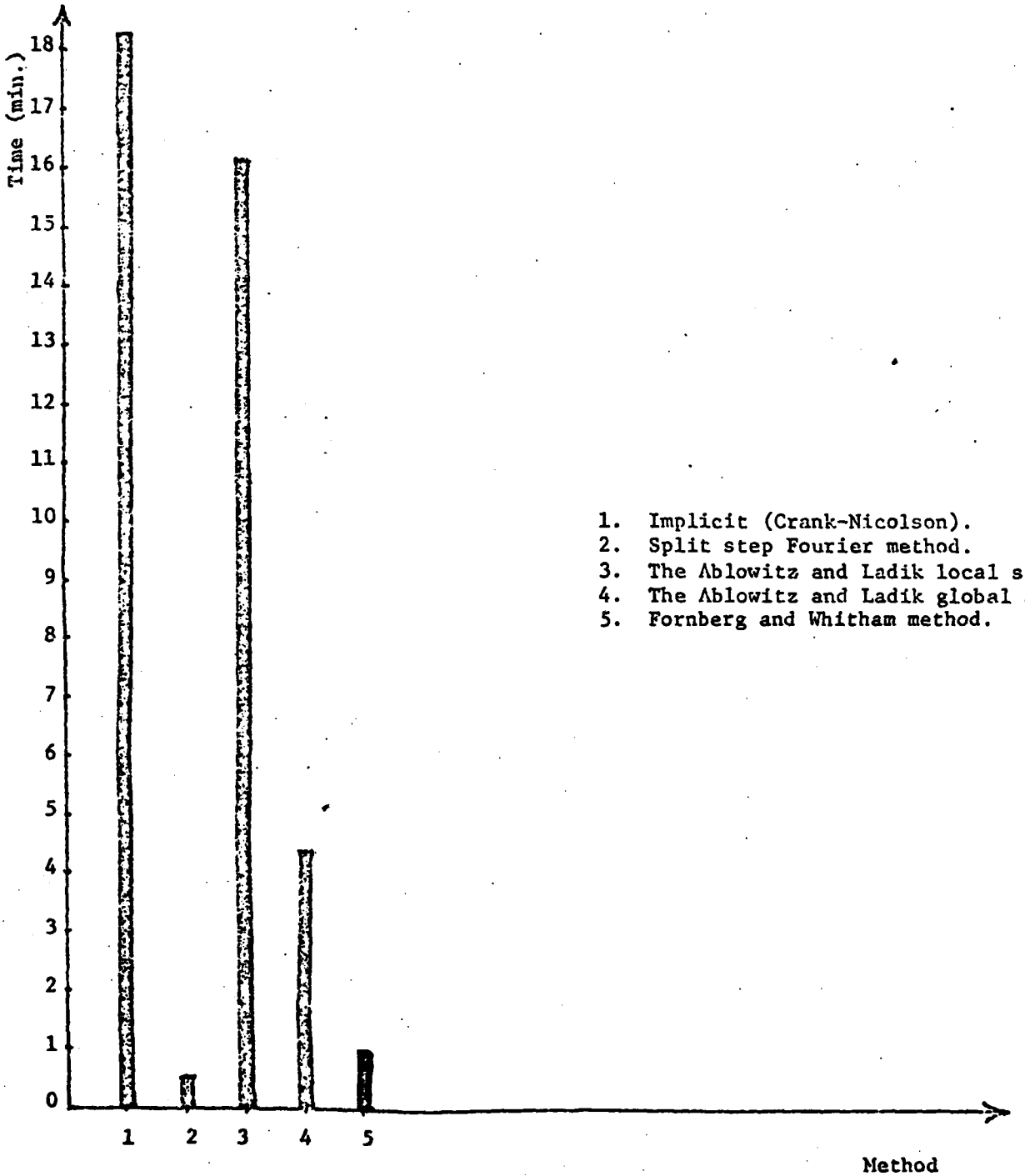
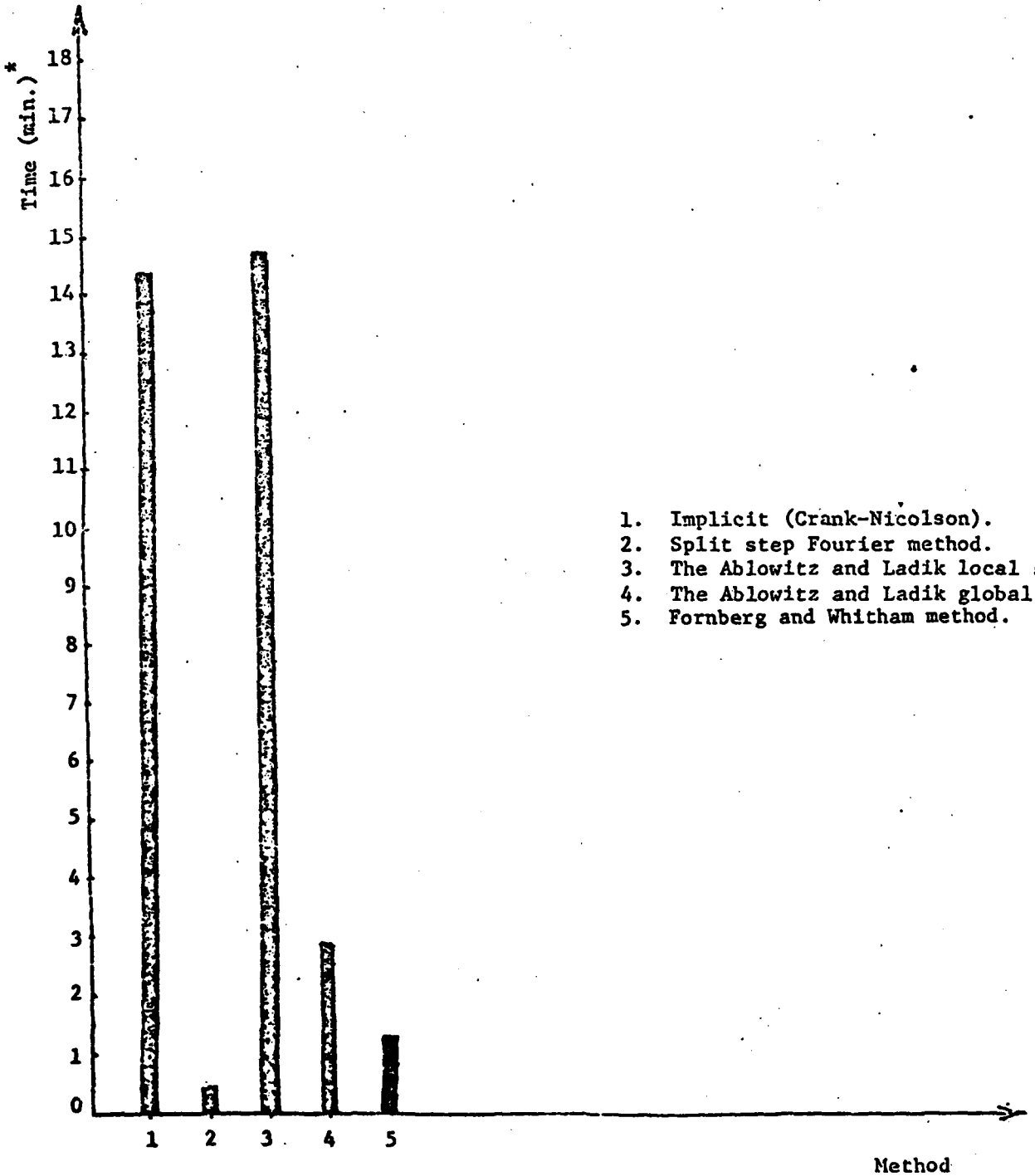


Figure (3.4). Displays the computing time (E) which is required by each utilized method given in Table (3.4). 1-soliton, amplitude = 6.



*: Each unit in time = 4 minutes.

Figure (3.5). Displays the computing time (E) which is required by each utilized method given in Table (3.5). Two solitons with amplitudes 0.5 and 1.

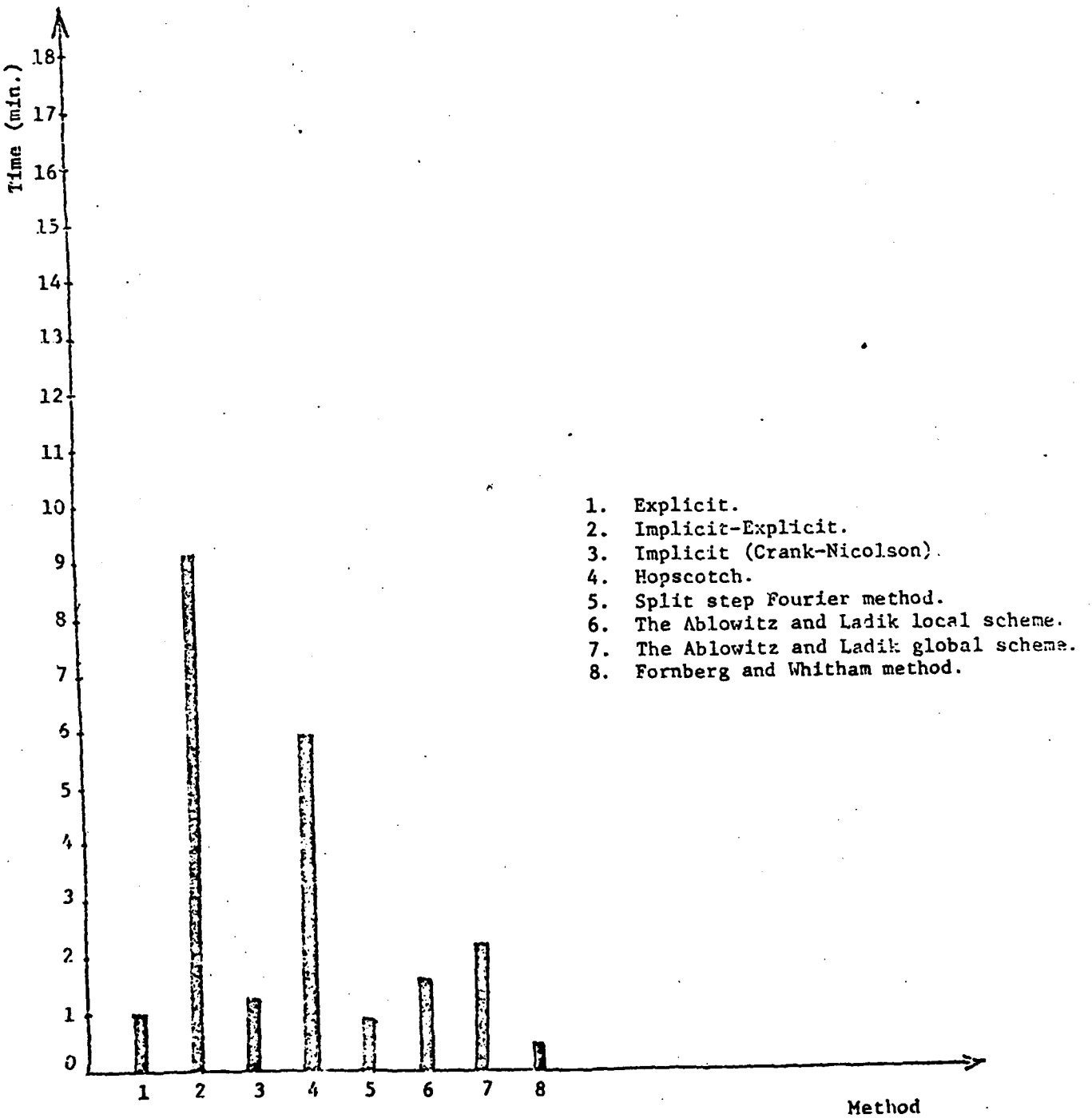
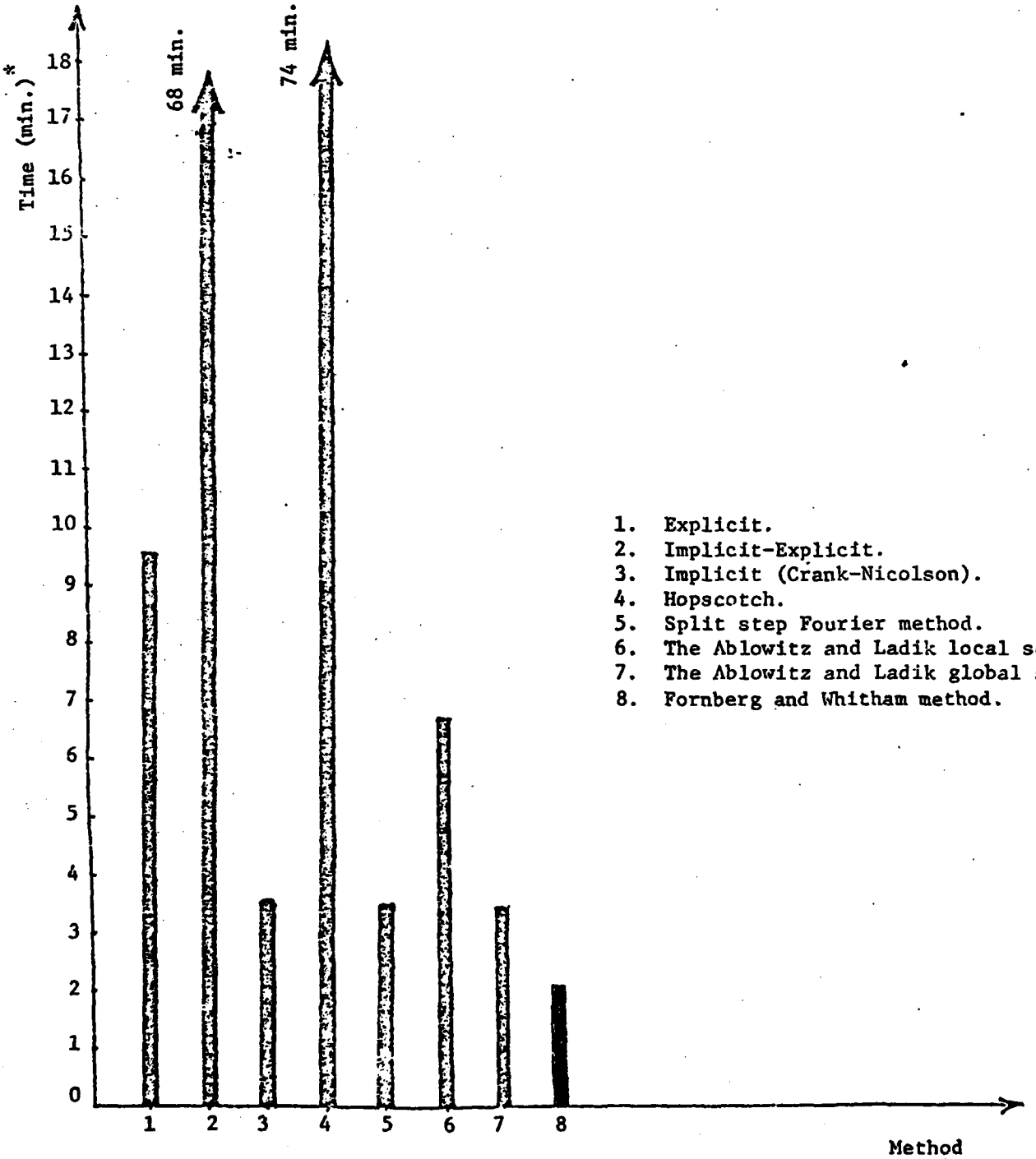
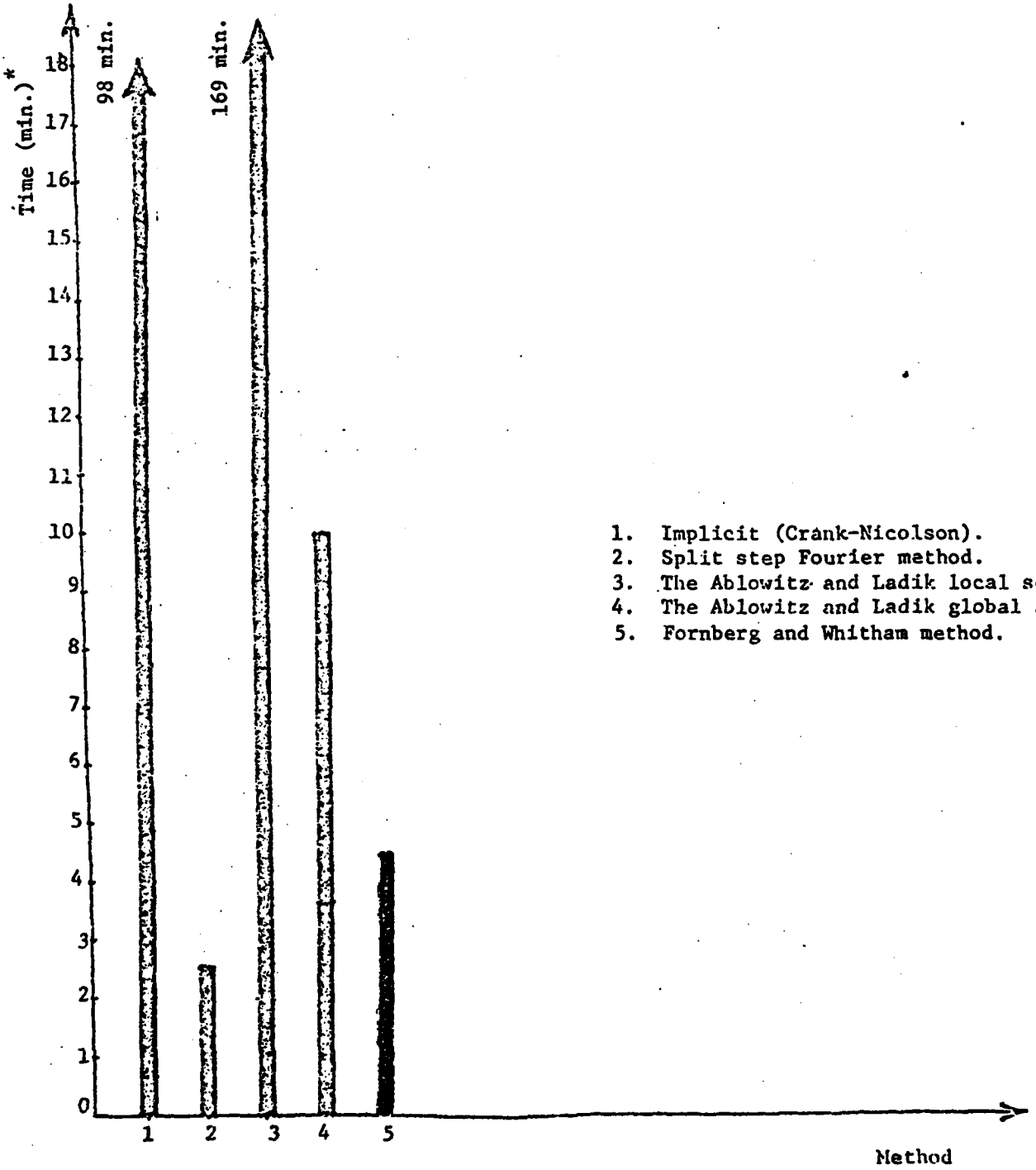


Figure (3.6). Displays the computing time (E) which is required by each utilized method given in Table (3.6). Two solitons with amplitudes 1 and 2.



*: Each unit in time ≅ 2 minutes.

Figure (3.7). Displays the computing time (E) which is required by each utilized method given in Table (3.7). Two solitons with amplitudes 3 and 4.



*: Each unit in time = 5 minutes.

APPENDIX A

AN OPTIMIZATION OF GAUSSIAN ELIMINATION

This method seeks to optimize Gaussian Elimination by eliminating unnecessary storage and multiplication by zeros. To begin we have the following quasi tridiagonal system

$$\begin{bmatrix} d_1 & u_1 & & & \alpha \\ \ell_1 & d_1 & u_2 & & \\ & & & & \\ \beta & & & \ell_{N-1} & d_N \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \\ x_N \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \\ b_N \end{bmatrix}$$

Instead of storing the NxN matrix, we store the augmented matrix in an Nx4 matrix whose elements are the tridiagonal elements and the b_i's. We then perform Gaussian elimination on the Nx4 matrix keeping in mind their original locations in the matrix. When this is done we have an upper triangular matrix and an original system is of the form

$$\begin{bmatrix} d_1 & u_1 & & & \alpha \\ & d_2^* & u_2^* & & \alpha \\ & & & & \\ & & & d_{N-1}^* & u_{N-1}^* \\ & & & & d_N^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \\ \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2^* \\ \\ \\ b_N^* \end{bmatrix}$$

where * values are the updated elements. Using back substitution we obtain the solution. The total number of operations required to obtain the solution using this method is (9N-12). The same idea can be applied to quasi pentagonal system of equations and so on.

REFERENCES

[1] Yajima, N., Oikawa, M., Satsuma, J., and Namba, C., Modulated Langmuir Waves and Nonlinear Landau Damping, Reports of Res. Ins. for Appl. Mechanics, Vol. XXII, No. 70, 1975.

[2] Karpman, V.I. and Krushkal, E.M., Modulated Waves in Nonlinear Dispersive Media, Soviet Physics JETP, Vol. 28, 1969, p. 277.

[3] Yajima, N. and Outi, A., A New Example of Stable Solitary Waves, Prog. Theor. Phys., Vol. 45, 1971, p. 1997.

[4] Satsuma, J., and Yajima, N., Initial Value Problems of One-Dimensional Self-Modulation of Nonlinear Waves in Dispersive Media, Prog. Theor. Phys. (supplement) No. 55, 1974.

[5] Tappert, F., Private communications.

[6] Hardin, R.H. and Tappert F.D., Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations, SIAM-SIGNUM Fall Meeting, Austin, TX, October 1972; SIAM Rev. (chronicle) 15 (1973), 423.

[7] Ablowitz, M., and Ladik, J., Stud. Appl. Math. 55, 213 (1976).

[8] Ablowitz, M., and Ladik, J., Stud. Appl. Math. 57, I-12 (1977).

[9] Eilbeck, J.C., "Numerical Studies of Solitons", "Solitons and Condensed Matter Physics" (Eds. A.R. Bishop and T. Schneider), Springer Series in Solid-State Sciences 8, Springer-Verlag, Berlin, 1978, pp. 28-43.

[10] Greig, I.S., and Morris, J. Ll, J. Comp. Phys. #1, 20, 60-84 (1976).

[11] Fornberg, B., and Whitham, G.B., Phil. Trans. Roy. Soc., 289, 373 (1978).

[12] We appreciate private communication with Segur, H. on this point.

[13] Hirota, R., Exact envelope-soliton solutions of a nonlinear wave equation, J. Math. Phys., Vol. 14, Nov. 7, July 1973.

[14] Richtmyer and Morton, Difference Methods for Initial Value Problems, 2nd edition, Interscience Publishers a division of John Wiley and Sons, 1967.

[15] Kruskal, M.D., Private communications.

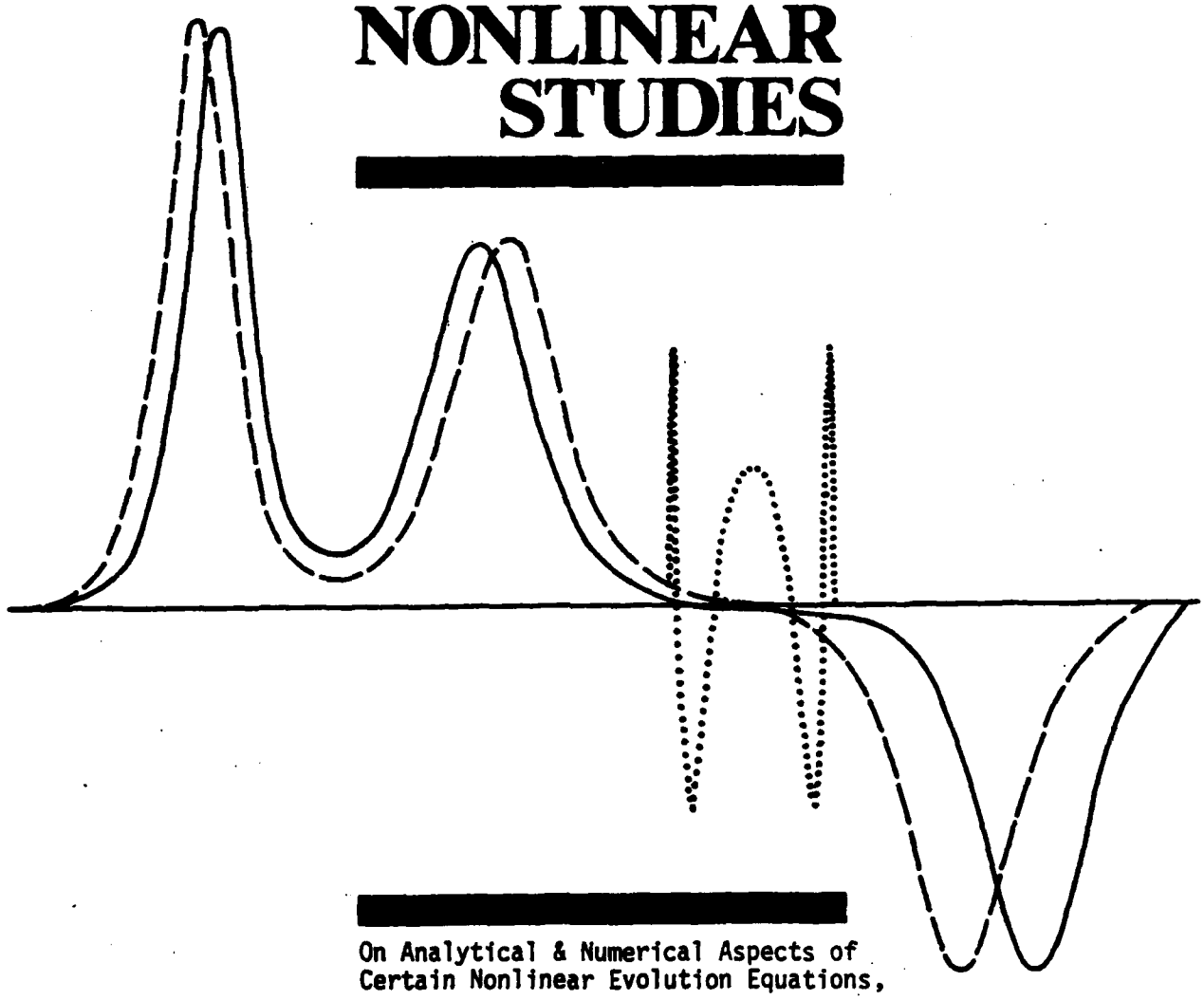
[16] Smith, G.D., Numerical Solution of Partial Differential Equations, Oxford University Press, 1965.

[17] Cooley, J.W., Lewis, P.A.W. and Welch, P.D., 1969, IEEE Trans. Education E-12, No. 1, 27-34.

[18] McCraken, Daniel D., and William, S. Dorn, "Numerical methods and FORTRAN programming with applications in engineering and science", Wiley, New York, 1964.

I.F.N.S. #16

INSTITUTE FOR NONLINEAR STUDIES



On Analytical & Numerical Aspects of Certain Nonlinear Evolution Equations,

Part III: Numerical, Korteweg-DeVries Equation

by

Thiab R. Taha and Mark J. Ablowitz

Clarkson College of Technology
Potsdam, New York 13676

ON ANALYTICAL AND NUMERICAL ASPECTS OF CERTAIN NONLINEAR EVOLUTION EQUATIONS, PART III: NUMERICAL, KORTEWEG-DE VRIES EQUATION

ABSTRACT

Various numerical methods were used in order to approximate the Korteweg-de Vries equation, namely: (i) Zabusky-Kruskal scheme, (ii) Hopscotch method, (iii) A scheme due to Goda, (iv) A proposed local scheme, (v) A proposed global scheme, (vi) A scheme suggested by Kruskal, (vii) Split step Fourier method by Tappert, (viii) An improved split step Fourier method, (ix) Pseudo-spectral method by Fornberg and Whitham. Comparisons between our proposed scheme, which was developed using notions of the Inverse Scattering Transform, and the other utilized schemes are obtained.

1. Introduction.

The Korteweg-de Vries equation (KdV) introduced in [1] was originally derived in order to describe the behavior of one-dimensional shallow water waves with small but finite amplitudes. More recently, this equation also has been found to describe wave phenomena in plasma physics [2,3], anharmonic crystals [4,5] bubble-liquid mixtures [6,7] etc. There has been great interest in this equation because of its special properties. A substantial review of this work can be found in [8,9]. Zabusky and Kruskal [10] discovered the concept of solitons localized waves with special interaction properties, while studying the results of a numerical computation (describing an anharmonic lattice) on the KdV equation. This motivated the work of Gardner etal [11] and led to the explosion of both the theoretical and numerical work which is still growing today. Many analytical results are available for equations which exhibit exact multisoliton behavior, when an

associated scattering problem can be found. Of course there are many examples inexact, or quasi-soliton behavior. For these problems little or no analytical results are known and numerical studies are essential in order to develop an understanding of the phenomena. This work aims to compare a proposed scheme which was developed in part I of this paper using notions of the inverse scattering transform (IST) and certain other known numerical methods for the KdV equation

$$u_t + 6uu_x + u_{xxx} = 0 \tag{1.1}$$

This work can be extended to cover many other related equations as well.

The following numerical methods are applied to the KdV equation:

1. Finite difference methods.
   a) Explicit methods.
      i) Zabusky and Kruskal scheme [10,12].
   b) Implicit methods.
      i) Hopscotch method [13].
      ii) A scheme due to Goda [14].
      iii) The proposed scheme [15].
      iv) A scheme suggested by M. Kruskal [16].
2. Finite Fourier transform or Pseudospectral methods.
      i) Split step Fourier method introduced by Tappert [17].
      ii) Pseudospectral method introduced by Fornberg and Whitham [18].

As in part II in order to compare schemes, our approach for comparison is to (a) fix the accuracy (L_∞) for computations beginning at $t = 0$ and ending at $t = T$; (b) leave other parameters free (e.g. Δt, or Δx), and compare the computing time required to attain such accuracy for various choices of the parameters.

Various methods were applied to the KdV equation (1.1) subject to the following conditions:

(a) The initial conditions:

i) 1-soliton solution.

The exact solution of (1.1) on the infinite interval is

$$u(x,t) = A \operatorname{sech}^2(kx-\omega t-\eta_0) \tag{1.2}$$

where

$$\omega = 4k^2,\ A = 2k^2 \text{ and } \eta_0 = \text{constant}.$$

For initial conditions, equation (1.2) was used at t = 0, and different values of A are tested and η_0 was chosen to be zero.

ii) Collisions of two solitons.

The exact solution of (1.1) on the infinite interval is

$$u(x,t) = 2(\log f)_{xx}, \tag{1.3}$$

where

$$f = 1 + e^{\eta_1} + e^{\eta_2} + e^{\eta_1+\eta_2+A_{12}},$$

$$\eta_i = k_i x - k_i^3 t + \eta_i^{(0)}$$

and

$$e^{A_{ij}} = \left(\frac{k_i-k_j}{k_i+k_j}\right)^2.$$

For initial conditions, equation (1.3) was used at t = 0, and two different sets of values of the parameters were studied, namely;

$$k_1 = 1,\ k_2 = \sqrt{2},\ \eta_1^{(0)} = 0,\ \eta_2^{(0)} = 2\sqrt{2}$$

and

$$k_1 = 1,\ k_2 = \sqrt{5},\ \eta_1^{(0)} = 0,\ \eta_2^{(0)} = 10.73$$

and the solitons were allowed to interact and return to their original shapes.

(b) The boundary conditions.

Periodic boundary conditions on the interval [-20,20] were imposed.

The numerical solution is compared with the exact solution. In addition, two of the conserved quantities were computed, namely; $\int u^2 dx$, and $\int [2u^3 - (u_x)^2] dx$.

2. The Representation of the KdV Equation Using Numerical Methods:

1. Finite difference methods.

i) Zabusky and Kruskal scheme.

In their original work, Zabusky and Kruskal [10], used the following explicit leapfrog finite difference scheme:

$$u_n^{m+1} = u_n^{m-1} - 2\frac{\Delta t}{\Delta x}(u_{n+1}^m + u_n^m + u_{n-1}^m)(u_{n+1}^m - u_{n-1}^m) - \frac{\Delta t}{(\Delta x)^3}(u_{n+2}^m - 2u_{n+1}^m + 2u_{n-1}^m - u_{n-2}^m) \qquad (2.1)$$

where $u_n^m = u(n\Delta x, m\Delta t)$; n and m are integers. This scheme is consistent with equation (1.1) and the truncation error is of order $(O((\Delta t)^2) + O((\Delta x)^2))$. The linear stability requirement for this scheme is

$$\frac{\Delta t}{\Delta x}\left|-2u_0 + \frac{1}{(\Delta x)^2}\right| \leq \frac{2}{3\sqrt{3}} \qquad (2.2)$$

(where u_0 is the maximum value of u in the range of interest).

This means that a very small time step must be used to preserve stability.

For the initial time step one may use the uncentered scheme

$$u_n^1 = u_n^0 - \frac{\Delta t}{\Delta x}(u_{n+1}^0+u_n^0+u_{n-1}^0)(u_{n+1}^0-u_{n-1}^0)$$

$$-\frac{1}{2(\Delta x)^3}(u_{n+2}^0-2u_{n+1}^0+2u_{n-1}^0-u_{n-2}^0) \qquad (2.3)$$

b) Implicit Methods.

i) Hopscotch method.

In 1976 Greig and Morris [13] proposed a Hopscotch scheme for the KdV equation (1.1).

With $f = \frac{u^2}{2}$, the scheme is

$$u_n^{m+1} = u_n^m - 3\frac{\Delta t}{\Delta x}(f_{n+1}^m-f_{n-1}^m) - \frac{\Delta t}{2(\Delta x)^3}(u_{n+2}^m-2u_{n+1}^m+2u_{n-1}^m-u_{n-2}^m) \qquad (2.4a)$$

$$u_n^{m+1} = u_n^m - 3\frac{\Delta t}{\Delta x}(f_{n+1}^{m+1}-f_{n-1}^{m+1}) - \frac{\Delta t}{2(\Delta x)^3}(u_{n+2}^{m+1}-2u_{n+1}^{m+1}+2u_{n-1}^{m+1}-u_{n-2}^{m+1}) \qquad (2.4b)$$

To implement the scheme, we employ (2.4a) for those grid points for which (n+m) is even and (2.4b) for those for which (n+m) is odd. A quasi tridiagonal system of equations must be solved at each time level. An optimization of Gaussian elimination method is used to solve this system (see part II, Appendix A). The linear stability requirement for this scheme is that [13]

$$\frac{\Delta t}{(\Delta x)^3} \leq \left|\frac{1}{(\Delta x)^2 u_0 - 2}\right| \qquad (2.5)$$

(u_0 is the maximum value of u in the range of interest).

The truncation error of this scheme is of order $(O((\Delta t)^2) + O((\Delta x)^2))$.

ii) A scheme due to Goda.

This implicit scheme for approximating the KdV equation (1.1) is given by

$$\frac{1}{\Delta t}\,(u_n^{m+1}-u_n^m) + \frac{1}{\Delta x}\,\{u_{n+1}^{m+1}(u_n^m+u_{n+1}^m)-u_{n-1}^{m+1}(u_n^m+u_{n-1}^m)\}$$

$$+\frac{1}{2(\Delta x)^3}\,\{u_{n+2}^{m+1}-2u_{n+1}^{m+1}+2u_{n-1}^{m+1}-u_{n-2}^{m+1}\} = 0 \qquad (2.6)$$

The truncation error of this scheme is $0(\Delta t) + 0((\Delta x)^2)$. This scheme is unconditionally stable according to linear analysis. In order to apply this scheme, we have to solve a quasi pentagonal system of equations at each time level.

$$\begin{bmatrix}
x & x & x & & & & & x & x \\
x & x & x & x & & & & & x \\
x & x & x & x & x & & & & \\
 & x & x & x & x & x & & & \\
 & & x & x & x & x & x & & \\
 & & & & & & & & \\
 & & & & x & x & x & x & x \\
x & & & & & x & x & x & x \\
x & x & & & & & x & x & x
\end{bmatrix}$$

An optimization of Gaussian elimination method is used to solve this system of equations.

iii) The proposed scheme which is based on the IST (see part I).

First, consider the local scheme with $A^{(0)} = \frac{3}{2}\frac{\Delta t}{(\Delta x)^3}$, which can be written as

$$\frac{u_n^{m+1}-u_n^m}{\Delta t} = \frac{1}{2(\Delta x)^3}\,[u_{n-1}^{m+1}-3u_n^{m+1}+3u_{n+1}^{m+1}-u_{n+2}^{m+1}+u_{n-2}^m-3u_{n-1}^m+3u_n^m-u_{n+1}^m]$$

$$-\frac{3}{2\Delta x}\,[(u_n^m)^2-(u_n^{m+1})^2] - \frac{1}{2\Delta x}\,\{u_{n+1}^{m+1}(u_n^{m+1}+u_{n+1}^{m+1}+u_{n+2}^{m+1})$$

$$-u_{n-1}^m(u_n^m+u_{n-1}^m+u_{n-2}^m)\} \qquad (2.7)$$

To implement this scheme, equation (2.7) for the new time level can be written as [16]

$$u_{n+2}^{m+1} - 3u_{n+1}^{m+1} + (3+\epsilon)u_n^{m+1} - u_{n-1}^{m+1} = B_n \tag{2.8}$$

where

$$\epsilon \equiv \frac{2(\Delta x)^3}{\Delta t} << 1$$

(Δt is supposed to be of the same order as Δx)

and

$$B_n = u_{n-2}^m - 3u_{n-1}^m - u_{n+1}^m + (3+\epsilon)u_n^m$$

$$- 3(\Delta x)^2[(u_n^m)^2-(u_n^{m+1})^2] - (\Delta x)^2[u_{n+1}^{m+1}(u_n^{m+1}+u_{n+1}^{m+1}+u_{n+2}^{m+1})$$

$$- u_{n-1}^m(u_n^m+u_{n-1}^m+u_{n-2}^m)] \tag{2.9}$$

This can be solved by a version of the Crank-Nicolson back and forth sweep method for the heat equation [20]. We seek an equation of the form

$$u_{n+1}^{m+1} = au_n^{m+1} + b_n^{m+1}, \tag{2.10}$$

which is suitable for computing u_n^{m+1} explicitly by sweeping to the right. For stability $|a| \leq 1$. Repeated substitution into equation (2.8) to eliminate u_{n+2}^{m+1}, u_{n+1}^{m+1}, and u_n^{m+1} in favor of u_{n-1}^{m+1} gives

$$b_{n+1}^{m+1} + (a-3)b_n^{m+1} + (a^2-3a+3+\epsilon)b_{n-1}^{m+1} + (a^3-3a^2+3a+\epsilon a-1)u_{n-1}^{m+1} = B_n \tag{2.11}$$

Requiring the u_{n-1}^{m+1} term to drop out determines a (uniquely since $|a| \leq 1$) as a solution of

$$(a-1)^3 + \epsilon a = 0 \tag{2.12}$$

and leaves for b_n (at the new time level) a second-order difference equation. The corresponding homogeneous equation of (2.11) has a solution of the form

$$b_n = k^n \tag{2.13}$$

if the constant k satisfies

$$k^2 + (a-3)k + a^2 - 3a + 3 + \epsilon = 0 \tag{2.14}$$

It can be shown, or verified, that the two roots k of this equation are the two roots with $|k| > 1$ other than that of the cubic equation determining $a < 1$ above. It follows that b can be computed explicitly by sweeping to the left

$$b_{n-1} = (3a-a^2)b_n - ab_{n+1} + aB_n \tag{2.15}$$

(equation (2.15) is obtained from (2.11) and (2.12)).
To obtain the solution u_n, first solve for b_n from (2.15) then use (2.10) to calculate u_n.

In order to implement this sweeping technique, the same iteration method used for the sweeping technique discussed in Part II was used. The only difference is that we have to assume initial values not just for b_n but also for b_{n+1}.

Second, for the global which can be written as (Part I, Equation $(2.30)_*$).

$$\frac{S_n^m}{1-S_n^m}\Big\{A_-^{(0)} - \sum_{\ell=-\infty}^{n}[E_{\ell+1} + S_\ell^{m+1}W_\ell(A_-^{(2)} + C_{\ell-2})$$

$$-\{D_-^{(4)}\gamma_{\ell-1} + D_-^{(2)} + \sum_{k=-\infty}^{\ell-1}(H_k + G_k)\}S_\ell^{m+1}\gamma_\ell + (\gamma_\ell - 1)]W_\ell^{-1}\Big\}W_n$$

$$-\frac{S_n^{m+1}}{1-S_n^{m+1}}\Big\{D_-^{(0)} + \sum_{\ell=-\infty}^{n-1}[\frac{-S_\ell^m}{S_{\ell+1}^{m+1}}\{\gamma_{\ell+1}^{-1}N_{\ell+1} - N_\ell + M_\ell$$

$$+ S_{\ell+1}^{m+1}Z_\ell - S_{\ell+1}^m\gamma_{\ell+1}^{-1}N_{\ell+1}\} + \gamma_\ell T_{\ell-2} + (\gamma_\ell - 1)]W_\ell^{-1}\Big\}W_{n-1}$$

$$+\frac{1}{1-S_n^m}E_{n+1} - \frac{1}{1-S_n^{m+1}}T_{n-2} = \frac{S_n^{m+1} - S_n^m}{(1-S_n^{m+1})(1-S_n^m)} \qquad (2.16)$$

where

$$E_n = A_-^{(2)}S_n^m W_{n-1} - S_n^{m+1}D_-^{(2)} + H_n + G_n - S_n^{m+1}\sum_{k=-\infty}^{n}(H_k + G_k)$$

$$+ S_n^m W_{n-1}C_{n-1} - S_n^m D_-^{(4)},\ C_n = A_-^{(4)} + \sum_{j=-\infty}^{n}P_j W_j^{-1},$$

$$T_n = \gamma_{n+1}M_n + S_{n+1}^{m+1}\gamma_{n+1}Z_n - S_{n+1}^m N_{n+1},\ M_n = S_n^{m+1}W_n A_-^{(4)} - S_n^m D_-^{(4)},$$

$$Z_n = (A_-^{(2)} + \sum_{j=-\infty}^{n}Q_j W_j^{-1})W_n,\ N_n = D_-^{(2)} + \sum_{j=-\infty}^{n}F_j,$$

$$W_n = \prod_{i=-\infty}^{n}\gamma_i,\ \gamma_i = (\frac{1-S_i^m}{1-S_i^{m+1}}),$$

$$H_k = A_-^{(4)}(S_{k+1}^m\gamma_k - S_k^m)W_{k-1},\ G_k = (S_k^m - S_{k+1}^{m+1})D_-^{(4)},$$

$$F_j = A_-^{(4)}(S_j^{m+1}W_j - S_{j-1}^{m+1}W_{j-1}) + D_-^{(4)}(S_{j-1}^m - S_j^{m+1}),$$

$$P_j = A_-^{(4)}(S_j^{m+1} - S_{j+1}^m)W_j + D_-^{(4)}(S_{j+1}^{m+1} - S_j^{m+1}\gamma_j),$$

$$Q_j = (S_{j-1}^{m+1} - S_j^m)W_j A_-^{(4)} - (S_{j-1}^m\gamma_j - S_j^m)D_-^{(4)},$$

$$A_-^{(2)} = -\frac{2}{3}A_-^{(0)} + \frac{1}{2}\alpha,\ D_-^{(2)} = -\frac{2}{3}A_-^{(0)} - \frac{1}{2}\alpha,$$

$$A_-^{(4)} = \frac{1}{6}A_-^{(0)} - \frac{1}{4}\alpha,\ D_-^{(4)} = \frac{1}{6}A_-^{(0)} + \frac{1}{4}\alpha,\ \alpha = \frac{\Delta t}{(\Delta x)^3},$$

$$A_-^{(0)} = \text{arbitrary constant, and } S_n^m = 1 - e^{(\Delta x)^2 u_n^m}$$

With $A_-^{(0)} = \frac{3}{2}\alpha$, the same idea is applied and the only difference is in the B_n term. This proposed scheme is unconditionally stable, and has a truncation error of order $(0((\Delta t)^2 + 0((\Delta x)^2))$.

iv) A scheme suggested by M. Kruskal.

Kruskal [16] has suggested the following numerical scheme

$$\frac{u_n^{m+1}-u_n^m}{\Delta t} + \frac{u_{n+2}^{m+1}-3u_{n+1}^{m+1}+3u_n^{m+1}-u_{n-1}^{m+1}}{2(\Delta x)^3} + \frac{u_{n+1}^m-3u_n^m+3u_{n-1}^m-u_{n-2}^m}{2(\Delta x)^3} = 0 \tag{2.17}$$

for

$$u_t + u_{xxx} = 0 \tag{2.18}$$

Kruskal did not suggest any particular numerical scheme for the nonlinear part of the KdV equation (1.1). The following scheme was used to solve the KdV equation (1.1):

$$\frac{u_n^{m+1}-u_n^m}{\Delta t} + \frac{u_{n+2}^{m+1}-3u_{n+1}^{m+1}+3u_n^{m+1}-u_{n-1}^{m+1}}{2(\Delta x)^3} + \frac{u_{n+1}^m-3u_n^m+3u_{n-1}^m-u_{n-2}^m}{2(\Delta x)^3} + 3\{\frac{\theta}{4\Delta x}[(u^2)_{n+1}^{m+1}-(u^2)_{n-1}^{m+1} + (u^2)_{n+1}^m-(u^2)_{n-1}^m] + \frac{1-\theta}{2\Delta x}[u_n^{m+1}(u_{n+1}^{m+1}-u_{n-1}^{m+1}) + u_n^m(u_{n+1}^m-u_{n-1}^m)]\} = 0 \tag{2.19}$$

To implement this scheme, the above mentioned sweeping/iteration technique was used. We tried several values of θ and experimentally we found that $\theta = \frac{2}{3}$ gave the best results.

This scheme is unconditionally stable according to linear stability, and has a truncation error of order $(O((\Delta t)^2) + O((\Delta x)^2))$.

2. Finite Fourier transform or Pseudospectral methods.

i) Split step Fourier method by F. Tappert.

For convenience the spatial period was normalized to $[0,2\pi]$, then equation (1.1) becomes

$$u_t + 6\frac{\pi}{p} uu_X + \frac{\pi^3}{p^3} u_{XXX} = 0 \tag{2.20}$$

where p is half the length of the interval of interest, and $X = (x+p)\pi/p$.

As discussed in part II of this paper the essence of the solution method is to alternate between two steps: (1) advance the solution using only the nonlinear term by means of a (implicit) finite difference approximation. (2) advance the solution using only the linear term by means of the discrete fast Fourier transform (FFT).

To implement this method for the KdV equation (2.20), as the first step, one first approximates,

$$u_t + 6\frac{\pi}{p} uu_X = 0. \tag{2.21}$$

A straightforward discretization is:

$$\tilde{u}_n^{m+1} = u_n^m - \frac{3}{4}\frac{\pi}{p}\frac{\Delta t}{\Delta X}\{(\tilde{u}^2)_{n+1}^{m+1} - (\tilde{u}^2)_{n-1}^{m+1} + (u^2)_{n+1}^m - (u^2)_{n-1}^m\} \tag{2.22}$$

where $\tilde{u}$ is a solution of equation (2.21) and u is the solution of equation (2.20). For the second step, we would take,

$$u(X_j,\ t+\Delta t) = F^{-1}(e^{ik^3\pi^3/p^3\Delta t} F(\tilde{u}(X_j,t))) \tag{2.23}$$

where F denotes discrete Fourier transform and F^{-1} its inverse.

This scheme is second order accurate in time and space (which comes from using equation (2.22) to approximate equation (2.21)), and unconditionally stable according to linear analysis. In order to find $F(\tilde{u})$ and F^{-1} the FFT technique is used. We have found however that an improved discretization of (2.21) works considerably better. Specifically the truncation error of the split step Fourier method is improved to be of order $(0((\Delta t)^2) + 0((\Delta x)^4))$ instead of order $(0((\Delta t)^2) + 0((\Delta x)^2))$, by approximating equation (2.21) according to

$$\tilde{u}_n^{m+1} = u_n^m - \frac{\Delta t}{8\Delta X} \frac{\pi}{p} \{[8(\tilde{u}^2)_{n+1}^{m+1} - 8(\tilde{u}^2)_{n-1}^{m+1} - (\tilde{u}^2)_{n+2}^{m+1} + (\tilde{u}^2)_{n-2}^{m+1}] + [8(u^2)_{n+1}^m - 8(u^2)_{n-1}^m - (u^2)_{n+2}^m + (u^2)_{n-2}^m]\} \quad (2.24)$$

Also, one may improve the truncation error to be of order $(0(\Delta t)^2 + 0(\Delta x)^p)$ for all p, see the pseudospectral method (Fornberg and Whitham) below.

ii) Pseudospectral method by Fornberg and Whitham [18].

As mentioned in part II of this paper, this is a Fourier method in which u(x,t) is transformed into Fourier space with respect to x. Again for convenience the spatial period is normalized to $[0,2\pi]$. This interval is discretized by N equidistant points, with spacing $\Delta X = \frac{2\pi}{N}$. The function u(X,t), numerically defined only on these points, can be transformed to the discrete Fourier space by

$$\hat{u}(k,t) = Fu = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} u(j\Delta X,t) e^{\frac{-2\pi ijk}{N}}, \quad (2.25)$$

$$k = -\frac{N}{2}, \ldots, -1, 0, 1, \ldots, \frac{N}{2} - 1.$$

The inversion formula is

$$u(j\Delta X,t) = F^{-1}\hat{u} = \frac{1}{\sqrt{N}} \sum_k \hat{u}(k,t) e^{2\pi ijk/N}, \quad (2.26)$$

$$k = -\frac{N}{2}, \ldots, -1, 0, 1, \ldots, \frac{N}{2} - 1$$

These transforms can be performed, efficiently with the fast Fourier transform algorithm [21]. With this scheme, u_X can be evaluated as $F^{-1}\{ikFu\}$, u_{XXX} as $F^{-1}\{-ik^3Fu\}$ and so on. Combined with a leap-frog time step the KdV equation (2.20) would then be approximated by

$$u(X,t+\Delta t) - u(X,t-\Delta t) + 2i \frac{6\pi}{p} \Delta t\, u(X,t) F^{-1}(kF(u))$$

$$- 2i\, \Delta t \frac{\pi^3}{p^3} F^{-1}(k^3 F(u)) = 0 \qquad (2.27)$$

Fornberg and Whitham make a modification in the last term, however, and take

$$u(X,t+\Delta t) - u(X,t-\Delta t) + 2i \frac{6\pi}{p} \Delta t\, u(X,t) F^{-1}(kF(u))$$

$$- 2iF^{-1}\{\sin(\frac{\pi^3 k^3}{p^3} \Delta t) F(u)\} = 0 \qquad (2.28)$$

The difference between equation (2.27) and (2.28) is in the approximation of the linear equation

$$u_t + \frac{\pi^3}{p^3} u_{XXX} = 0 \qquad (2.29)$$

The linear part of equation (2.28) is exactly satisfied for any solution of equation (2.29) [18]. Also it turns out that the linearized stability condition is less restrictive for (2.28) than for (2.27): $\frac{\Delta t}{(\Delta x)^3} < \frac{3}{2\pi^3} \simeq 0.1520$ compared to $\frac{\Delta t}{(\Delta x)^3} < \frac{1}{\pi^3} \approx 0.0323$ (for details see [18]).

Since the Fornberg and Whitham scheme is explicit, it is natural to consider Crank-Nicolson type implicit version, e.g.

$$u(X,t+\Delta t) - u(X,t) + 3i\Delta t\pi/p\{u(X,t+\Delta t)F^{-1}(kF(u(X,t+\Delta t)))$$

$$+ u(X,t)F^{-1}(kF(u(X,t)))\} - \frac{i\Delta t}{2}(\frac{\pi}{p})^3\{F^{-1}(k^3F(u(X,t+\Delta t)))$$

$$+ F^{-1}(k^3F(u(X,t)))\} = 0 \qquad (2.30)$$

This scheme (2.30) is unconditionally stable according to linear stability. We make some remarks about (2.30) in the conclusions.

3. Conclusions.

Various numerical methods are used in order to approximate the KdV equation (1.1), namely; (i) Zabusky and Kruskal scheme (2.1), (ii) Hopscotch method (2.4), (iii) A scheme due to Goda (2.6), (iv) A proposed local scheme (2.7), (v) A proposed global scheme (2.16), (vi) A scheme suggested by Kruskal (2.19), (vii) Split step Fourier method by Tappert (2.22) - (2.24), (viii, Pseudospectral method by Fornberg and Whitham (2.28).

As in part II, our approach for comparison is to (a) fix the accuracy (L_∞) for computations beginning at t = 0 and ending at t = T; (b) leave other parameters free (e.g. Δt, or Δx), and compare the computing time required to attain such accuracy for various choices of the parameters.

Two sets of initial conditions were studied: (A) 1-soliton solution with different values of the amplitude, (B) collisions of two solitons with different values of the parameters. According to this approach we have made the following conclusions:

1) The scheme of Goda required a long time compared to the other ((i),(ii),(iv),(vi), (vii), (viii)) schemes.

2) Zabusky and Kruskal's scheme was good for low amplitudes, but it needed more computing time than the other remaining methods ((ii),(iv),

(vi),(vii),(viii)) for high amplitudes.

3) The calculations for the previous two methods; Goda, and Zabusky-Kruskal, were not carried out for the 1-soliton case with amplitude = 4. They needed very long computing time.

4) The Tappert and Hopscotch schemes took less computing time than the previous two schemes, and they behave almost the same for fairly small amplitudes. For relatively high amplitudes the Tappert scheme turned out to be better.

5) The suggested scheme by Kruskal is in general faster than the previous schemes ((i),(ii),(iii),(v),(vii)).

6) The Fornberg and Whitham method is much faster than the suggested scheme by Kruskal, it is roughly three times faster for small amplitudes and six times faster for high amplitudes. Also (2.27) was tried, but (2.28) proved to be somewhat faster. In addition, the implicit version (2.30) was implemented and did not prove to be faster than (2.28).

7) The proposed local scheme is the best amongst all the utilized schemes. It was roughly eight times faster than the suggested scheme by Kruskal. (See remark belwo equation (2.18)). This certainly shows that the approximation of the nonlinear term is crucial. Also, it was roughly one and a half times faster than the Fornberg and Whitham scheme. This despite the fact that the local scheme is only $0((\Delta t)^2,(\Delta x)^2)$ whereas Fornberg and Whitham method is of order $0((\Delta t)^2,(\Delta x)^p)$ for all p. These results suggest that "IST generated" schemes will be good approximations for equations which are in fact exactly solvable by the IST. The proposed global scheme was implemented and proved to be faster than some of the utilized schemes, but much slower than its local version. Since the global scheme is so complicated and cumbersome, we are not convinced that our methods of implement were optimal. Optimizing the implementation of our

global scheme will be under further investigation in the near future. (The following Tables and Figures exhibit the results). All the numerical calculations were inspected at every step by using the conserved quantities $\int u^2 dx$, and $\int (2u^3 - (u_x)^2) dx$. (Table (3.1), ..., Table (3.5)). The two conserved quantities were calculated by means of Simpson's Rule [22]. The proposed global scheme is the only utilized scheme which has an infinite number of conserved quantities, and true soliton solutions.

ACKNOWLEDGEMENTS

This research was partially supported by the Air Force Office of Scientific Research, under Grant No. 78-3674-C and the Office of Naval Research under Grant No. N00014-76-C-0867. We are grateful for useful conversations with Professor M.D. Kruskal and Dr. H. Segur.

Table (3.1). Comparison of the computing time (E) which is required to attain an accuracy (L_∞) < 0.005 for computations beginning at t=0 and ending at t=1.0, for the numerical methods utilized in solving the KdV equation. Two conserved quantities are shown. 1-soliton as an initial condition with amplitude = 1 on the interval [-20, 20].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	u_1	u_2
1.	Explicit (Zabusky and Kruskal)	$\Delta x = 0.1739$ $\Delta t = 0.002$	C: 0 E: 0	20 28	0.25 0.5 1.0	0.00173 0.00283 0.00469	1.88562 1.88563 1.88562	2.27358 2.27359 2.27356
2.	Goda	$\Delta x = 0.1$ $\Delta t = 0.002$	C: 0 E: 4	45 4	0.25 0.5 1.0	0.00134 0.00245 0.00492	1.88411 1.88261 1.87962	2.26333 2.26032 2.25432
3.	Hopscotch	$\Delta x = 0.2$ $\Delta t = 0.003$	C: 0 E: 0	29 23	0.25 0.5 1.0	0.00178 0.00292 0.00472	1.88561 1.88562 1.88558	2.27693 2.27695 2.27675
4.	Tappert	$\Delta x = 0.3125$ $\Delta t = 0.004$	C: 0 E: 1	23 3	0.25 0.5 1.0	0.00333 0.00401 0.00494	1.88562 1.88561 1.88561	2.29630 2.29609 2.29591
5.	Kruskal	$\Delta x = 0.08$ $\Delta t = 0.04$	C: 0 E: 0	23 24	0.25 0.5 1.0	0.00202 0.00303 0.00453	1.88563 1.88564 1.88565	2.26502 2.26502 2.26504

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	u_1	u_2
6.	The proposed local scheme	$\Delta x = 0.16$ $\Delta t = 0.125$	C: 0 E: 0	23 7	0.25 0.5 1.0	0.00146 0.00162 0.00173	1.88571 1.88576 1.88582	2.27209 2.27221 2.27239
7.	The proposed global scheme	$\Delta x = 0.13$ $\Delta t = 0.12$	C: 0 E: 0	30 38	0.25 0.5 1.0	0.00138 0.00252 0.00477	1.88562 1.88562 1.88563	2.26871 2.26870 2.26869
8.	Pseudospectral by Fornberg and Whitham	$\Delta x = 0.625$ $\Delta t = 0.0096$	C: 0 E: 0	25 12	0.25 0.5 1.0	0.00157 0.00162 0.00113	1.88987 1.88344 1.88911	2.42315 2.36762 2.41751

Δx = The increment in x.

Δt = The increment in t.

C = The compilation time

u_1 = A conserved quantity of the KdV equation which is $\int u^2 dx$.

u_2 = A conserved quantity of the KdV equation which is $\int [2u^3 - (u_x)^2] dx$.

$L_\infty = \max|\tilde{u}_n^m - u_n^m|$, $\tilde{u}_n^m$ is the numerical solution and u_n^m is the exact solution at the point $(\Delta x \cdot n, \Delta t \cdot m)$ for all n,m.

Table (3.2). Comparison of the computing time (E) which is required to attain an accuracy (L_∞) < 0.01 for computations beginning at t=0 and ending at t=1.0, for the numerical methods utilized in solving the KdV equation. Two conserved quantities are shown. 1-soliton as an initial condition with amplitude = 2 on the interval [-20,20].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	u_1	u_2
1.	Explicit (Zabusky and Kruskal)	$\Delta x = 0.08$ $\Delta t = 0.00019$	C: 0 E: 9	20 51	0.25 0.5 1.0	0.00307 0.00520 0.00930	5.33333 5.33333 5.33333	12.82597 12.82595 12.82594
2.	Goda	$\Delta x = 0.04$ $\Delta t = 0.00025$	C: 0 E: 59	45 28	0.25 0.5 1.0	0.00236 0.00509 0.01282	5.32907 5.33482 5.31635	12.78946 12.77244 12.73857
3.	Hopscotch	$\Delta x = 0.1$ $\Delta t = 0.0005$	C: 0 E: 4	29 32	0.25 0.5 1.0	0.0371 0.00592 0.00994	5.33335 5.33328 5.33340	12.84042 12.84002 12.84044
4.	Tappert	$\Delta x = 0.156$ $\Delta t = 0.002$	C: 0 E: 4	23 31	0.25 0.5 1.0	0.00729 0.00820 0.00943	5.33326 5.33319 5.33304	12.89642 12.89582 12.83304
5.	Kruskal	$\Delta x = 0.04$ $\Delta t = 0.011$	C: 0 E: 2	23 43	0.25 0.5 1.0	0.00418 0.00607 0.00952	5.33337 5.33339 5.33344	12.80654 12.80665 12.80682

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	u_1	u_2
6.	The proposed local scheme	$\Delta x = 0.1$ $\Delta t = 0.1$	C: 0 E: 0	23 23	0.25 0.5 1.0	0.00237 0.00246 0.00332	5.33380 5.33384 5.33409	12.84235 12.84228 12.84313
7.	The proposed global scheme	$\Delta x = 0.05$ $\Delta t = 0.025$	C: 0 E: 6	30 7	0.25 0.5 1.0	0.00228 0.00443 0.00882	5.33347 5.33361 5.33375	12.81065 12.81119 12.81176
8.	Pseudospectral by Fornberg and Whitham	$\Delta x = 0.417$ $\Delta t = 0.0031$	C: 0 E: 0	25 57	0.25 0.5 1.0	0.00179 0.00248 0.00344	5.33087 5.34124 5.33006	13.38022 13.56912 13.36529

Table (3.3). Comparison of the computing time (E) which is required to attain an accuracy (L_∞) < 0.022 for computations beginning at t=0 and ending at t=1.0, for the numerical methods (listed below) utilized in solving the KdV equation. Two conserved quantities are shown. 1-soliton as an initial condition with amplitude = 4 on the interval [-20,20].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	u_1	u_2
1.	Hopscotch	$\Delta x = 0.05$ $\Delta t = 6.2 \times 10^{-5}$	C: 0 E: 77	29 39	0.25 0.5 1.0	0.00722 0.01240 0.02259	15.08492 15.08490 15.08500	72.52186 72.52165 72.52248
2.	Tappert	$\Delta x = 0.073$ $\Delta t = 0.0008$	C: 0 E: 26	23 30	0.25 0.5 1.0	0.016242 0.017140 0.021586	15.08493 15.08493 15.08492	72.68155 72.68204 72.68048
3.	Kruskal	$\Delta x = 0.03$ $\Delta t = 0.01$	C: 0 E: 22	23 8	0.25 0.5 1.0	0.00580 0.01069 0.02110	15.08434 15.08375 15.08257	72.44412 72.43939 72.42995
4.	The proposed local scheme	$\Delta x = 0.05$ $\Delta t = 0.0275$	C: 0 E: 2	23 20	0.25 0.5 1.0	0.00992 0.01272 0.01747	15.08469 15.08455 15.08428	72.51893 72.51760 72.51536
5.	The proposed global scheme	$\Delta x = 0.025$ $\Delta t = 0.005$	C: 0 E: 53	30 23	0.25 0.5 1.0	0.00530 0.01072 0.02163	15.08513 15.08634 15.08400	72.43785 72.44754 72.42881
6.	Pseudospectral by Fornberg and Whitham	$\Delta x = 0.3125$ $\Delta t = 0.00115$	C: 0 E: 3	25 14	0.25 0.5 1.0	0.00600 0.01095 0.01752	15.07489 15.11822 15.07162	75.99011 75.49573 75.83650

Table (3.4). Comparison of the computing time (E) which is required to attain an accuracy (L_∞) < 0.002 for computations beginning at t=0 and ending at t=3.0, for the numerical methods utilized in solving the KdV equation. Two conserved quantities are shown. Two-solitons as an initial condition with amplitudes 1/2 and 1 respectively, and they are allowed to interact, on the interval [-20,20].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	u_1	u_2
1.	Explicit (Zabusky and Kruskal)	$\Delta x = 0.12$ $\Delta t = 0.00066$	C: 0 E: 5	22 49	0.1 0.5 1.0 2.0 3.0	0.00047 0.00115 0.00159 0.00191 0.00165	2.55228 2.55228 2.55228 2.55228 2.55228	2.66732 2.66669 2.66568 2.66366 2.66362
2.	Goda	$\Delta x = 0.1$ $\Delta t = 0.0005$	C: 0 E: 46	44 4	0.1 0.5 1.0 2.0 3.0	0.00032 0.00105 0.00161 0.00200 0.00165	2.55216 2.55171 2.55123 2.55059 2.55018	2.66570 2.66439 2.66277 2.66021 2.65951
3.	Hopscotch	$\Delta x = 0.13$ $\Delta t = 0.001$	C: 0 E: 4	30 43	0.1 0.5 1.0 2.0 3.0	0.00032 0.00107 0.00147 0.00191 0.00142	2.55216 2.55171 2.55123 2.55059 2.55108	2.66570 2.66439 2.66277 2.66021 2.65951

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	u_1	u_2
4.	Tappert	$\Delta x = 0.15625$ $\Delta t = 0.005$	C: 0 E: 5	24 22	0.1 0.5 1.0 2.0 3.0	0.00171 0.00188 0.00160 0.00126 0.00186	2.55228 2.55228 2.55228 2.55229 2.55229	2.67059 2.66970 2.66845 2.66614 2.66666
5.	Kruskal scheme	$\Delta x = 0.08$ $\Delta t = 0.015$	C: 0 E: 1	24 46	0.1 0.5 1.0 2.0 3.0	0.00021 0.00088 0.00139 0.00176 0.00145	2.55228 2.55226 2.55228 2.55236 2.55234	2.66476 2.66439 2.66386 2.66293 2.66288
6.	The proposed local scheme	$\Delta x = 0.1$ $\Delta t = 0.14$	C: 0 E: 0	23 19	0.1 0.5 1.0 2.0 3.0	0.00080 0.00113 0.00135 0.00138 0.00148	2.55225 2.55215 2.55181 2.55117 2.55115	2.66590 2.66552 2.66440 2.66227 2.66213
7.	The proposed global scheme	$\Delta x = 0.07$ $\Delta t = 0.06$	C: 0 E: 4	31 21	0.1 0.5 1.0 2.0 3.0	0.00025 0.00051 0.00080 0.00115 0.00204	2.55230 2.55237 2.55246 2.55227 2.55302	2.66435 2.66433 2.66423 2.66342 2.66462
8.	Pseudospectral by Fornberg and Whitham	$\Delta x = 0.625$ $\Delta t = 0.0148$	C: 0 E: 0	29 24	0.1 0.5 1.0 2.0 3.0	0.00127 0.00101 0.00166 0.00101 0.00107	2.55528 2.55052 2.55109 2.55215 2.55273	2.80405 2.74889 2.73592 2.71420 2.71725

Table (3.5). Comparison of the computing time (E) which is required to attain an accuracy (L_∞) < 0.02 for computations beginning at t=0 and ending at t=2.4, for the numerical methods utilized in solving the KdV equation. Two conserved quantities are shown. Two solitons as an initial condition with amplitudes 1/2 and 5/2 respectively, and they are allowed to interact, on the interval [-20,20].

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	u_1	u_2
1.	Explicit (Zabusky and Kruskal)	$\Delta x = 0.06$ $\Delta t = 0.000082$	C: 0 E: 74	22 5	0.1 0.6 1.2 2.0 2.4	0.00177 0.00655 0.00842 0.01734 0.02220	8.12023 8.12022 8.12022 8.12022 8.12023	22.79202 22.78662 22.76096 22.78485 22.79111
2.	Goda	This method takes very long time to attain the required accuracy, therefore there is no need to carry out the calculations.						
3.	Hopscotch	$\Delta x = 0.0725$ $\Delta t = 0.00019$	C: 0 E: 32	30 29	0.1 0.6 1.2 2.0 2.4	0.00196 0.00649 0.00919 0.01586 0.02013	8.12027 8.12109 8.12457 8.12128 8.12037	22.80636 22.80194 22.77995 22.80020 22.80539

No.	Method	Mesh size	Time min.	sec.	t=T	L_∞	u_1	u_2
4.	Tappert	$\Delta x = 0.15625$ $\Delta t = 0.002$	C: 0 E: 10	24 48	0.1 0.6 1.2 2.0 2.4	0.01232 0.01320 0.00859 0.01802 0.02086	8.12023 8.12034 8.12061 8.11986 8.11973	22.96728 22.93491 22.79313 22.93070 22.96060
5.	Kruskal scheme	$\Delta x = 0.04$ $\Delta t = 0.0045$	C: 0 E: 11	24 56	0.1 0.6 1.2 2.0 2.4	0.00221 0.00696 0.00842 0.01534 0.02019	8.12021 8.12032 8.12141 8.12024 8.11986	22.77440 22.77159 22.76085 22.76991 22.77196
6.	The proposed local scheme	$\Delta x = 0.075$ $\Delta t = 0.055$	C: 0 E: 1	23 11	0.1 0.6 1.2 2.0 2.4	0.00229 0.00321 0.01023 0.01613 0.01502	8.12031 8.11812 8.10934 8.11656 8.11821	22.80979 22.78967 22.76395 22.77737 22.79739
7.	The proposed global scheme	$\Delta x = 0.035$ $\Delta t = 0.0125$	C: 0 E: 35	31 50	0.1 0.6 1.2 2.0 2.4	0.00092 0.00478 0.00662 0.01670 0.02002	8.12030 8.12123 8.11874 8.12127 8.11833	22.77167 22.77445 22.75541 22.77377 22.76221
8.	Pseudospectra. by Fornberg and Whitham	$\Delta x = 0.417$ $\Delta t = 0.0039$	C: 0 E: 1	29 53	0.1 0.6 1.2 2.0 2.4	0.01015 0.01297 0.00843 0.01573 0.01532	8.10244 8.10832 8.12367 8.12430 8.12919	23.71671 23.63303 23.08949 23.94873 24.21967

Figure (3.1). Displays the computing time (E) which is required by each utilized method given in Table (3.1). 1-soliton, amplitude = 1.

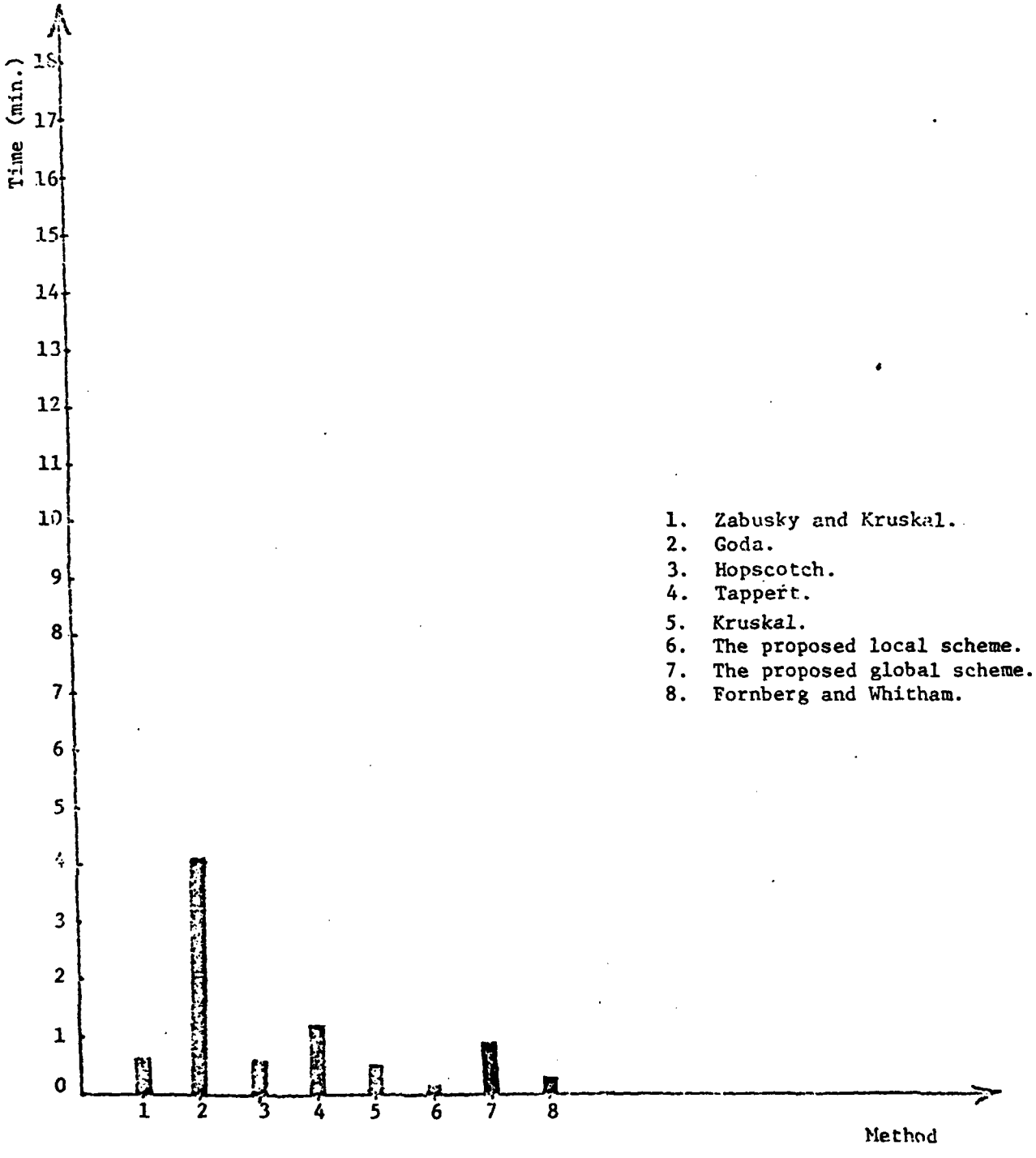


Figure (3.2). Displays the computing time (E) which is required by each utilized method given in Table (3.2). 1-soliton, amplitude = 2.

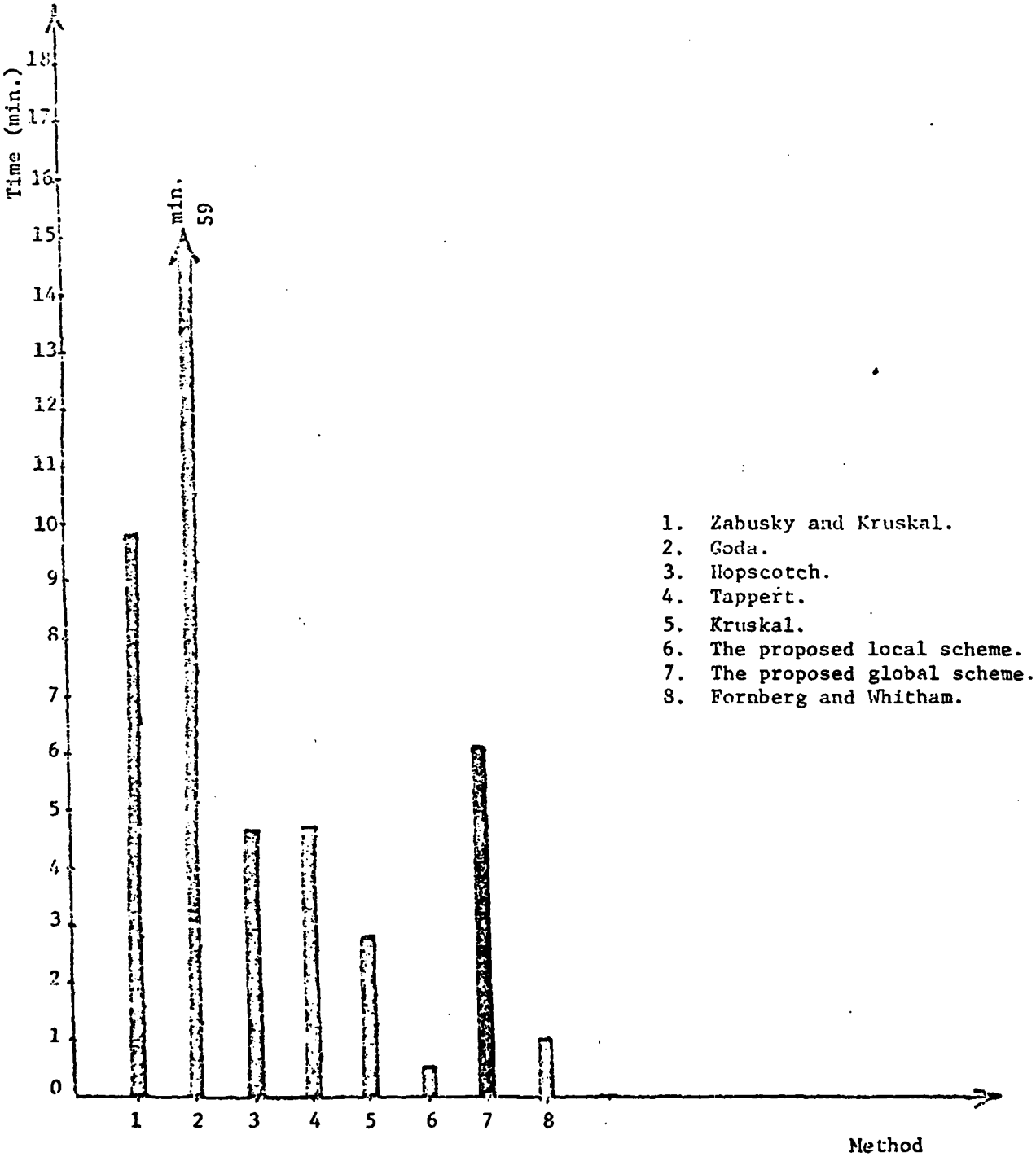
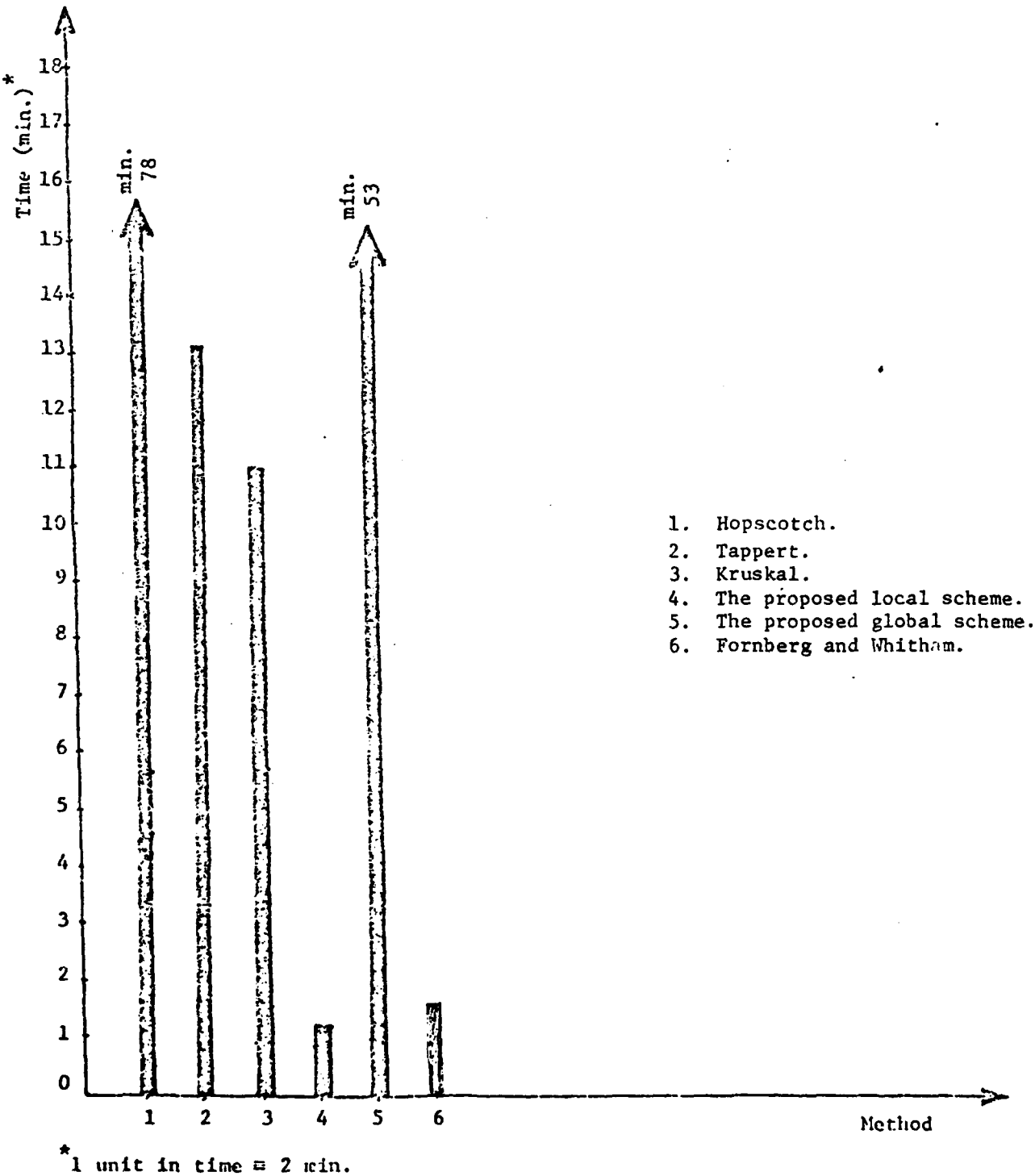


Figure (3.3). Displays the computing time (E) which is required by each utilized method given in Table (3.3). 1-soliton, amplitude = 4.



*1 unit in time = 2 min.

Figure (3.1). Displays the computing time (T) which is required by each utilized method given in Table (3.1). Two solitons with amplitudes 0.5 and 1.

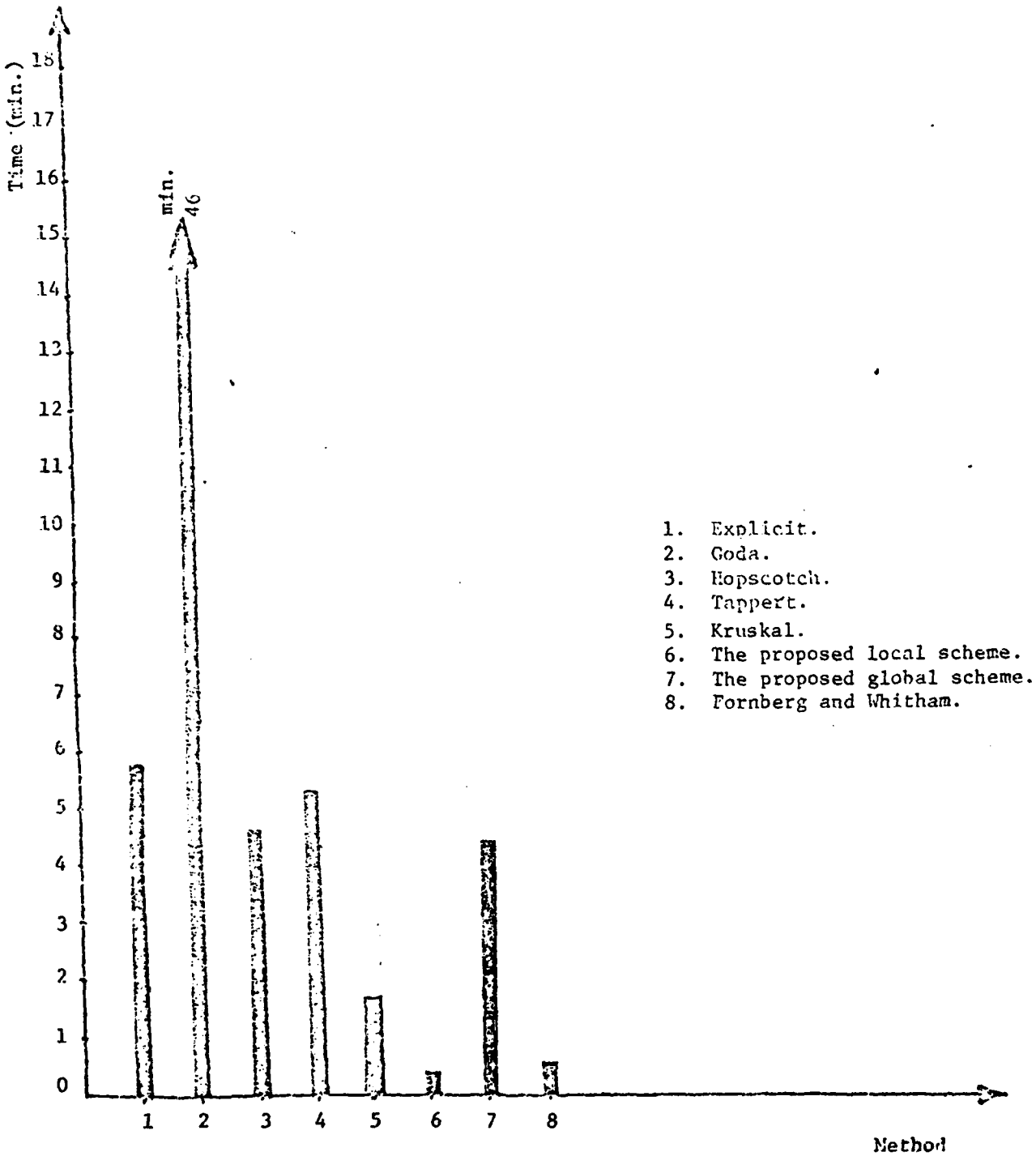
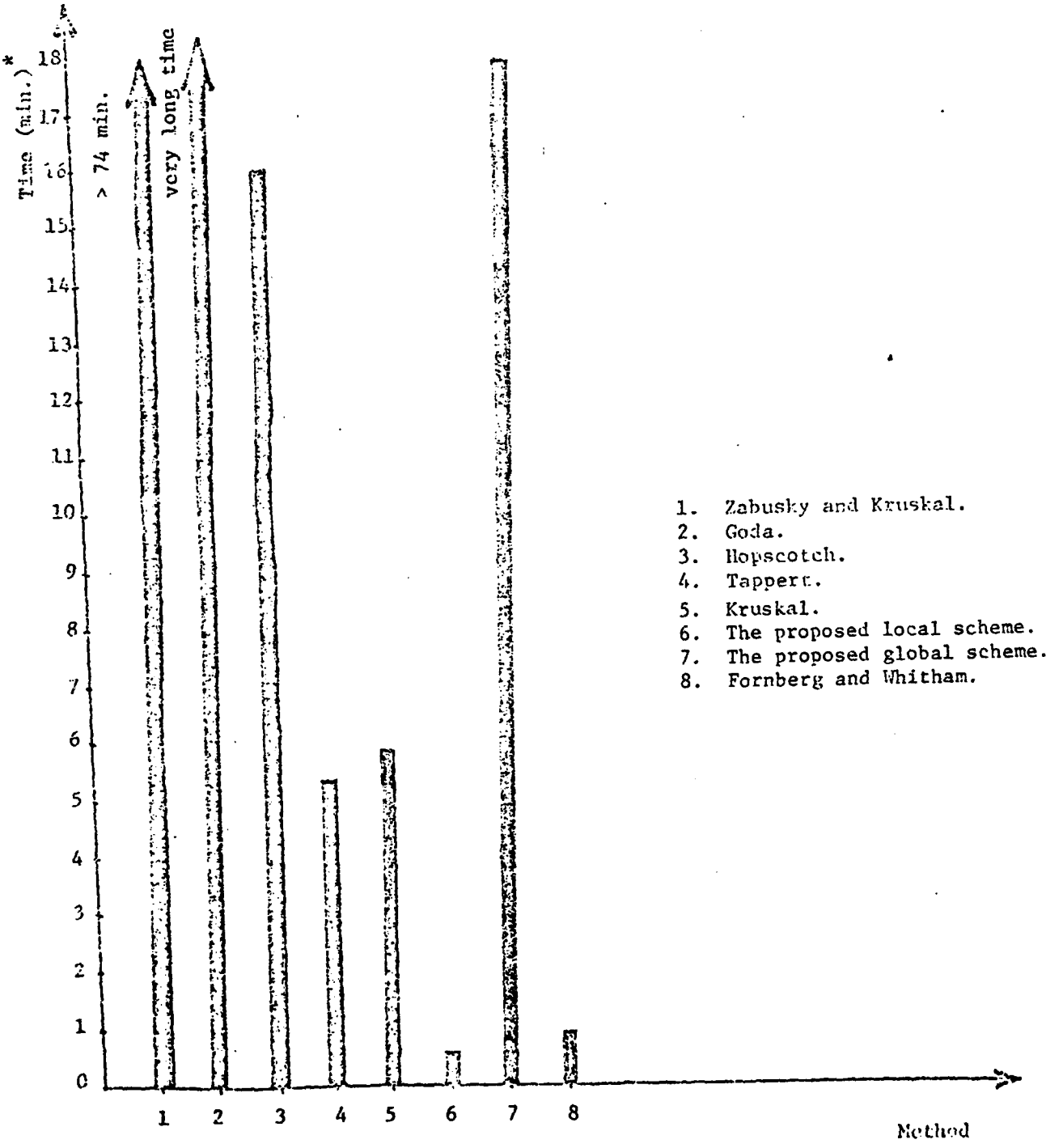


Figure (3.5). Displays the computing time (E) which is required by each utilized method given in Table (3.5). Two solitons with amplitudes 0.5 and 2.5.



*1 unit in time = 2 min.

REFERENCES

[1] Korteweg, D.J., and deVries, G., On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary wave, Philos. Mag. 39 (1895) 422-443.

[2] Gardner, C.S., and Morikawa, G.K., Similarity in the asymptotic behavior of collision-free hydromagnetic waves and water waves, New York Univ., Courant Inst. Math. Sci., Res. Report NYO-90 4, 1960.

[3] Washimi, H., and Taniuti, T., Propagation of ion-acoustic solitary waves of small amplitudes, Phys. Rev. Letters, 17 (1966) 996-998.

[4] Zabusky, N., A synergetic approach to problems of nonlinear wave propagation and interaction in Nonlinear Partial Differential Equations (W.F., Ames, Ed.) Academic, New York, 1967, pp. 223-258.

[5] Kruskal, M.D., Asymptotology in numerical computation: progress and plans on the Fermi-Pasta-Ulam problem, Proceedings of the IBM Scientific Computing Symposium on Large-Scale Problems in Physics (1965).

[6] Wijngaarden, L. van, Linear and non-linear dispersion of pressure pulses in liquid-bubble mixtures, 6th Symposium on Naval Hydrodynamics, Washington, D.C. (1966).

[7] Wijngaarden, L. van, On the equations of motion for mixtures of liquid and gas bubbles, J. Fluid Mech., 33 (1968) 465-474.

[8] Zakharov, V.E., Manakov, S.V., Novikov, S.P., and Pitayevsky, S.P., "Theory of Solitons", Moscow, Nauk, 1980.

[9] Ablowitz, M., and Segur, H., "Solitons and the Inverse Scattering Transform", SIAM, Philadelphia, 1981.

[10] Zabusky, N.J., and Kruskal, M.D., Interaction of "solitons" in a collisionless plasma and the recurrence of initial states, Phys. Rev. Letters, 15 (1965) 240-243.

[11] Gardner, C., Greene, J., Kruskal, M., Miura, R., Phys. Rev. Lett. 19, 1095 (1967).

[12] Vliegenthart, A.C., J. Engrg. Math., 151, 137 (1971).

[13] Greig, I.S., and Morris, J. Ll, J. Comp. Phys. #1, 20, 60-84 (1976).

[14] Goda, K., On instability of some finite difference schemes for the Korteweg-de Vries Equation, Journal of the Phys. Soci. of Japan 39, No. 1, July 1975.

[15] Taha, T., Ph.D. Thesis, Clarkson College, Potsdam, NY, 1982.

[16] Kruskal, M.D., Private communications.

[17] Tappert, F., Numerical solutions of the KdV equation and its Generalization by the Split-step Fourier Method, Lectures in App. Math. 15 (1974), 215.

[18] Fornberg, B., and Whitham, G.B., Phil. Trans. Roy. Soc., 289, 373 (1978).

[19] Fornberg, B., "On the Instability of Leap-Frog and Crank-Nicolson Approximations of a Nonlinear Partial Differential Equation", Math. Comp. Vol. 27, #121, Jan. 1973.

[20] Richtmyer and Morton, Difference Methods For Initial Value Problems, 2nd edition, Interscience Publishers a division of John Wiley and Sons, 1967.

[21] Cooley, J.W., Lewis, P.A.W. and Welch, P.D., 1969, IEEE Trans. Education E-12, No. 1, 27-34.

[22] Smith, G.D., Numerical Solution of Partial Differential Equations, Oxford University Press, 1965.

Physica 2D (1981) 389–394
North-Holland Publishing Company

NONLINEAR RESONANCES AND COLLIDING SPHERICAL ION-ACOUSTIC SOLITONS†

D. J. KAUP
Dept. of Physics, Clarkson College of Technology, Potsdam, NY 13676, USA

Received 15 January 1980

We find that the recent experimental results by Ze, Hershkowitz, Chan and Lonngren on colliding spherical ion-acoustic solitons can be explained in terms of a nonlinear resonance, wherein the two colliding solitons at a resonance point generate and create a third soliton of larger amplitude. However, estimates show that the resonance width is less than the required growth length, therefore full growth is not achieved, in agreement with the experimental results.

Recently, Ze, Hershkowitz, Chan and Lonngren [1] have presented experimental evidence that it is possible for colliding spherical ion-acoustic solitons to create new nonlinear objects, presumably also solitons. We shall present theoretical calculations to show that this is theoretically possible, and that there is a very good quantitative agreement between the theory and experiment. We shall call the theoretical principle involved "nonlinear resonance", and shall first quickly outline its concepts and history.

One important and very significant feature of solitons has been a continual paralleling of linear principles and concepts [2], even though the soliton is indeed a nonlinear object. Of these, we mention the "inverse scattering transform" [3] which is much like the linear Fourier transform, the infinity of conservation laws (also found in linear systems), the existence of action-angle variables [3], quantum theories for such objects [4], etc., as well as "nonlinear resonance" [5, 6], which is very analogeous to the resonance condition for three or more linear wave trains to interact. The usual linear resonance occurs for plane waves and when we have only three plane waves, it satisfies

$$k_1 + k_2 = k_3, \tag{1a}$$

$$\omega_1(k_1) + \omega_2(k_2) = \omega_3(k_3). \tag{1b}$$

In (1), k_i is the wave vector for the ith wave, and $\omega_i(k)$ is the frequency of the ith wave, as a function of k. Although the interaction arising when (1) is satisfied is indeed nonlinear (it is called the three-wave resonant interaction [7]), nevertheless we shall call such a resonance a "linear resonance", because it involves only linear parameters (i.e. -wave vectors and frequencies) of plane waves. A "nonlinear resonance" therefore will be defined to be a resonance involving nonlinear (i.e. -soliton) parameters.

There is a very strong parallel between a linear resonance and a nonlinear resonance. In order to bring out this parallel and to illustrate the analogies, we shall first state certain facts concerning solitons. The structure of any soliton is typically of the form

$$q = Af_1(\phi_r, \phi_i), \tag{2}$$

† Research supported in part by NSF Grant No. MCS78-03979 and ONR Contract No. N00014 76 C 0867.

0167-2789/81/0000-0000/$02.50© North-Holland Publishing Company

where A is an amplitude and ϕ_r and ϕ_i are the real and imaginary parts of a complexified phase [2]. This complexified phase may be defined directly from the linear phase, which is simply that phase associated with a plane wave solution of the linear limit of the nonlinear soliton equation. We shall shortly illustrate this with a couple of examples. For simplicity we shall first only consider one-dimensional problems, where the general linear phase for a plane wave is

$$\phi = kx - \omega(k)t + \phi_0. \tag{3}$$

Now, it is fairly well known [2] that if we want to know what the complexified phase required in (2) is, one simply needs to replace the real parameter k in (3) by the complex parameter 2ζ. Then (3) becomes complex in general and

$$\phi_r + i\phi_i = 2\zeta x - \omega(2\zeta)t + \phi_0, \tag{4}$$

where in general

$$\zeta = \xi + i\eta. \tag{5}$$

Of course this does not determine the functional form of f_1 in (2). It only determines how the real and imaginary parts of the complexified phase in (2) depends on x and t, and thereby the group velocity and the phase velocity. The important point is that the phases of solitons will follow upon knowing the phase of the plane wave solution of the associated linear problem.

We shall now illustrate this with a couple of examples. For the KdV equation

$$q_t + q_x q + \beta q_{xxx} = 0, \tag{6}$$

the linear problem is $q_t + \beta q_{xxx} = 0$, so we have from the linear problem that $\omega(k) = -\beta k^3$, which is the functional form required in (4). This then determines ϕ_r and ϕ_i. From the nonlinear problem, one then finds the remainder of the information, which is $f_1(\phi_r, \phi_i) = \operatorname{sech}^2(\phi_i)$, $\xi = 0$, $\eta > 0$ and $A = 12\beta\zeta^*\zeta$. For the nonlinear Schrödinger equation

$$iq_t = -q_{xx} - 2(q^*q)q, \tag{7}$$

we have $\omega(k) = k^2$, $f_1(\phi_r, \phi_i) = e^{i\phi_r}\operatorname{sech}(\phi_i)$, $\xi \neq 0$ in general, $\eta > 0$ and $A = i(\zeta - \zeta^*)$.

The point to be noted in the above, is that a soliton is like a linear wave, in that it satisfies exactly the same phase relation that a linear wave does. It differs in that the phase relation has become complexified as in (4) and (5), and in that the envelope now has a unique shape, as a function of ϕ_r and ϕ_i. These statements are also true for higher-dimensional systems, like the Kadomtsev–Petviashvili equation [8] and a two-dimensional form of the nonlinear Schrödinger equation [9], provided we replace k, x, ζ, ξ and η by appropriate vector quantities.

In addition to the one-soliton as given in (2), there exists also two-, three-, etc. soliton solutions. In particular, the two-soliton solution is typically of the form

$$q = A(\zeta_1, \zeta_1^\dagger, \zeta_2, \zeta_2^\dagger) f_2(\phi_{1r}, \phi_{1i}, \phi_{2r}, \phi_{2i}). \tag{8a}$$

Again, the exact dependence of A on the ζ's and of f_2 on the ϕ's will be determined by the specific nonlinear problem. A general feature of (8a) is that in certain regions of x and t, (8a) will approach sums of one-soliton solutions as in (2). Another feature of (8a) which seems to be even lesser known is that for certain ranges of the parameters, and in certain regions of x and t, (8a) can degenerate into the form

$$q \approx A(\zeta_1 \pm \zeta_2, \zeta_1^\dagger \pm \zeta_2^\dagger) f_1(\phi_{1r} \pm \phi_{2r}, \phi_{1i} \pm \phi_{2i}). \tag{8b}$$

This form typically occurs in the region of x and t where the two solitons are considered to be overlapping and "interacting". One should now ask himself the question of what would happen if there would exist a *third* value of ζ, ζ_3, such that we could find a ϕ_3 where $\phi_1 \pm \phi_2 = \phi_3$. Then (8b) would become (2) with ζ_3 satisfying (4). This brings us to our third fact. Although the

function f_2 is a nonlinear function, it still does mix the phases linearly, in that sums such as $\phi_{1r}+\phi_{2r}$ and $\phi_{1i}+\phi_{2i}$ do occur.

As noted by Miles [5] and by Newell and Redekopp [6], the two-soliton solution in higher dimensional systems becomes singular and breaks down whenever there exists a third value of ζ such that

$$\phi(\zeta_1)+\phi(\zeta_2)=\phi(\zeta_3), \tag{9}$$

where ζ_1 and ζ_2 are the "eigenvalues" for soliton #1 and soliton #2. By suitably choosing values for ϕ_0, one can remove this singular nature, and then the two-soliton solution takes on a three-pointed star structure [10], two arms of which are the original two solitons, while the third arm is a new (created) soliton, with a corresponding "eigenvalue" of $\zeta=\zeta_3$. This star structure occurs only when (9) is satisfied, and from (3) and (4), we have that (9) is equivalent to

$$\zeta_1+\zeta_2=\zeta_3, \tag{10a}$$

$$\omega(2\zeta_1)+(2\zeta_2)=\omega(2\zeta_3). \tag{10b}$$

Note the similarity between (1) and (10). Obviously, what has happened here is that in the region where this third arm is, (8a) has degenerated into (8b), and since $\phi(\zeta_3)$ exactly satisfies (4) for the given $\omega(k)$, the third arm can therefore freely propagate as a third soliton. It could only freely propagate if this complexified phase relation was exactly satisfied.

The above facts very strongly suggest that there can exist very strong nonlinear resonances between solitons, resulting in the creation of new solitions and/or the decay of old solitons. In fact, considering how closely that solitons do mock their linear counterparts, the existence of a soliton resonance as given by (10) should not be at all surprising. One may easily say, that it has not been recognized before, simply because up until now, there has been no outstanding examples of (10). For the one-dimensional KdV equation, eq. (6), and the nonlinear Schrödinger equation, eq. (7), one may easily show that it is impossible for these systems to satisfy (10). However, there is one one-dimensional system which, with hindsight, does exhibit the effects of (10), and that is the three-wave resonant interaction [7, 11]. For this system, we have three dispersion relations, one for each wave which are $\omega_n(k)=\pm c_n k$ for $n=1, 2$ and 3. If we let the second envelope have the highest frequency (which determines the signs via $|\omega_2|=|\omega_1|+|\omega_3|$), then (10) becomes

$$\zeta_1+\zeta_3=\zeta_2, \tag{11a}$$

$$\zeta_1 c_1+\zeta_3 c_3=\zeta_2 c_2, \tag{11b}$$

as the condition for a nonlinear resonance. For simplicity, we take the ζ's to be pure imaginary, then (11) gives

$$\eta_1(c_1-c_2)+\eta_3(c_3-c_2)=0. \tag{12}$$

We remark that for ζ to be physical [7], it must lie in the upper half complex plane, and thus all η's must be positive. This restriction causes (12) to have a solution only if $c_1<c_2<c_3$ or $c_3<c_2<c_1$. In other words, for these nonlinear resonances to occur in the one-dimension form of the three-wave resonant interaction, the wave with the middle group velocity must have the highest frequency. This is satisfied in the "Explosive Case" [7, 11] where this nonlinear resonance gives an explosive singularity. It also is satisfied in the "Soliton Decay Case" [7, 11], where the nonlinear resonance corresponds to both up-conversion and "soliton decay", but it is not satisfied in the "Stimulated Back Scattering Case" [7, 11], and thus in this latter case, we now have a "reason" as to why nothing really remarkable occurs.

Let us now turn our attention to an interpretation of the experimental data on colliding spherical ion-acoustic solitons, which to lowest

order satisfy [12]

$$\partial_r q + \frac{q}{r} + q\partial_\mu q + \beta\partial_\mu^3 q = 0, \tag{13}$$

where r is the radius of the spherical soliton, $\mu = r - ct$, with c being the ion-acoustic speed, $q = \delta n/n_0$, and $\beta = \frac{1}{2}c^2/\omega_{pi}^2$ where ω_{pi} is the ion plasma frequency. Except for the second term, (13) is just the KdV equation whose solution would be

$$q = A\,\mathrm{sech}^2\phi_i, \tag{14}$$

with $A = 12\,\beta\eta^2$ and $\phi_i = \eta(\chi r - ct)$ where $\chi = 1 - \frac{1}{3}A$. Either from energy considerations [13] or by a perturbation theory [14], one may show that the effect of the second term in (13) on the KdV soliton, (14), is in lowest order, to cause the amplitude to slowly decay as $r^{-4/3}$, and therefore η as $r^{-2/3}$. In fact, experimental results [15] indicate that an individual spherical soliton will behave like this ($A \sim r^{-4/3}$) until the nonlinear term is reduced in amplitude to about 2.5 times this second term, at which time, A drops off faster than $r^{-4/3}$, and the soliton presumably decays into linear dispersive waves. Consequently, we can consider (13) to possess approximate soliton solutions, where the soliton amplitude will decrease as r increases. And this approximation simply requires the soliton amplitudes to be sufficiently large so that one may consider this second term to be a perturbation. Thus since the experimental collision results [1] did involve large amplitude solitons, we may use the form in (14), letting A decay as $r^{-4/3}$.

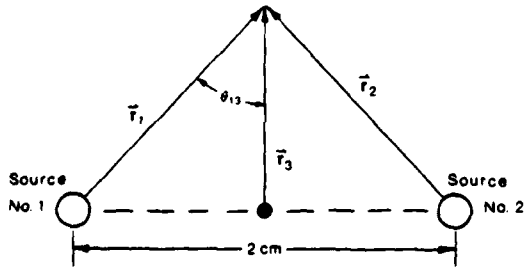


Fig. 1. The geometry for colliding spherical solitons. $r_1(r_2)$ is the vector from source #1 (#2) to the point of collision while r_3 is the vector from the apparent source of the created nonlinear object (soliton #3) to its crest. r_3 will be exactly at the intersection of r_1 and r_2 only at the resonance point.

In terms of the geometry for the colliding spherical solitons (fig. 1 of ref. 1), they will collide along the perpendicular bisecting plane between the two sources. If we let $r_1(r_2)$ be the vector from source #1 (#2) to the maximum of the first (second) spherical soliton, (see fig. 1) then if we are to have condition (9) to be satisfied, we must have

$$\eta_1 + \eta_2 = \eta_3, \tag{15a}$$

$$\frac{r_1}{r_1}\eta_1\chi_1 + \frac{r_2}{r_2}\eta_2\chi_2 = \frac{r_3}{r_3}\eta_3\chi_3, \tag{15b}$$

satisfied in order for a nonlinear resonance to occur, where η_3 will be the η value for the created third soliton and r_3 will be a vector from its (apparent) source to its maximum. From the geometry and the symmetry (fig. 1), r_3 must lie in the perpendicular bisecting plane. Thus r_1 (or r_2) is the hypotenuse of a right triangle, with r_3 and $\frac{1}{2}d$ (d is the separation distance between the two sources) being the lengths of the other two sides. We note that (15b) also corresponds to having the speed of the third soliton match the speed of the point of intersection of the two colliding solitons. In the experiment [1], $\eta_2 = \eta_1$, $A_2 = A_1$ and $r_1 = r_2$. Then (15) does have a solution when

$$\eta_3 = 2\eta_1, \quad A_3 = 4A_1, \tag{16a}$$

$$\cos\theta_{13} = \chi_3/\chi_1, \tag{16b}$$

where θ_{13} is the angle between r_1 and r_3. In the experiment at $t = 4\,\mu$s, $A_1 = 0.2$ and thus by (16), $A_3 = 0.4$ and $\theta_{13} = 38^0$. From the geometry, we would then expect to see the nonlinear resonance, and the resulting creation of a third

solition, at $r_1 = r_2 = 1.6$ cm and $r_3 = 1.3$ cm since $d = 2.0$ cm. Based on the experimental results (fig. 2 of ref. 1), we would expect this resonance to occur at or about $t = 3\,\mu$s. Their data does not show what has happened at $t = 3\,\mu$s, but we do note that the new object is well formed at $t = 4\,\mu$s. Of course, the above values of A are for $t = 4\,\mu$s, whereas we find that the collision should occur around $t = 3\,\mu$s. Thus since A at $t = 3\,\mu$s will be larger than that at $4\,\mu$s, and our values are only qualitative at this point. A more accurate result would require experimental data around $t = 3\,\mu$s. Nevertheless, it seems very likely that this nonlinear resonance is responsible for the creation of this new object, and that it indeed may be another form of a solition.

However, we also note that (16a) predicts an amplitude of 0.4, whereas their [1] fig. 2 shows that the new object as having an amplitude of only about 0.2. This suggests that the new soliton was not able to fully form while the collision passed through the resonance region. To check this, we will estimate the width of the resonance region, and compare it to the required growth length of the third soliton. To estimate the width of the resonance region, we estimate the width over which the sum of the phases of solitons #1 and #2 will match the phase of the third (created) soliton. If we let $\Delta\phi = \phi_{i1} + \phi_{i2} - \phi_{i3}$, then this will vanish at the resonance, will also have a vanishing first derivative (due to velocity matching) at the resonance, but will have a nonzero second derivative. Taking our variable to be r_3, then we would have $\Delta\phi \sim (r_3 - r_{30})^2/s_r^2$ where s_r is the resonance width and r_{30} is the value of r_3 at resonance. From (16) we then have $\Delta\phi = 2\eta_1\chi_1 r_1 - \eta_3\chi_3 r_3$ which upon taking the second derivative with respect to r_3 gives

$$s_r^2 = 2^{1/2} r_1/[2\eta_1\chi_1 \sin^2\theta_{13}], \tag{17}$$

which for the above values and $\beta = 1.74 \times 10^{-4}\,\text{cm}^2$ gives $s_r \approx 0.6$ cm. Again, since these values are based on the values at $4\,\mu$s, they are only qualitatively accurate.

Let us now estimate the required growth length. When solitons #1 and #2 collide, there will be additional nonlinearities in the plasma fluid equations, coming from their overlap. These additional nonlinearities will drive the growth of the third soliton. Treating these overlapping terms as perturbations and ignoring the depletion of the two colliding solitons, one can show that the equation of motion for the third soliton, along the perpendicular bisecting plane (where maximum growth would occur) is given by

$$\frac{\partial u_3}{\partial r_3} + \frac{u_3}{r_3} + u_3\partial_\mu u_3 + \beta\partial_\mu^3 u_3 = \Gamma, \tag{18}$$

where now $\mu = r_3 - ct$ and

$$\Gamma = -\chi_1(\cos^2\theta_{13} - \chi_1 \sin^2\theta_{13})\partial_\mu(u_1^2) + 4\eta_1^2\chi_1^2 \sin^4\theta_{13} \int_\mu^\infty d\mu(\partial_\mu u_1)^2, \tag{19}$$

with

$$u_1 = A_1 \operatorname{sech}^2[\eta_1\mu + \eta_1(\chi_1/\cos\theta_{13} - 1)r_3], \tag{20}$$

being the envelope of either colliding soliton along the bisecting plane. Assuming u_3 in (18) to be a soliton, and upon using the second conservation law, $\int_{-\infty}^{\infty} u^2 d\mu$, one obtains

$$\frac{\partial A_3}{\partial r_3} + \frac{4A_3}{3r_3} = \int_{-\infty}^{\infty} u_3\Gamma\, d\mu/(12\,\beta\eta_3), \tag{21}$$

for the rate of growth of A_3. We shall now proceed to estimate the integral on the right-hand side of (21). If we approximate u_3 by a soliton as in (14), letting $u_3 \approx A_3 \operatorname{sech}^2\{\eta_3[\mu - \frac{1}{3}A_3(v_3 - v_{30})]\}$, then at the resonance, u_3 is symmetric in μ. Meanwhile, the first term of Γ in (19) is antisymmetric with respect to μ, so we shall neglect it in (21). The second term in (19) is positive definite, can therefore give a growth, and has a maximum value of $\Gamma_{\max} = (64/15)\eta_1 A_1^2\chi_1^2$.

This term will be one half of this maximum value at the center of the collision volume, so we shall estimate Γ by taking $\Gamma \approx \frac{1}{2}\Gamma_{max}$ in (21). Then since $\int_{-\infty}^{\infty} u_3 \, d\mu = 24\beta\eta_3$, we find

$$\frac{\partial A_3}{\partial r_3} + \frac{4A_3}{3r_3} \approx (64/15)\eta_1 A_1^2 \chi_1^2. \tag{22}$$

We note that (22) is really only valid in the resonance region, because in going from (21) to (22), we have assumed the speeds of u_3 and Γ to be matched. Integrating (22) and requiring A_3 to grow from zero to its full maximum value of $4A_1$, gives us a required growth length of

$$s_g \approx 35/(16\eta_1 A_1 \chi_1^2), \tag{23}$$

which for the above values give $s_g \approx 1.3$ cm, which is about twice that of the resonance width $s_r \approx 0.6$ cm. Thus, based on these estimates, we would expect A_3 to be able to grow to only about one half of its maximum possible amplitude, and that is indeed what was observed.

In conclusion, we find that the experimental result of Ze et al [1] can be explained at least qualitatively in terms of a nonlinear resonance, wherein the colliding spherical ion-acoustic solitons do generate a third ion-acoustic (nonspherical) soliton of larger amplitude. However, since the required growth length for this resonance is about twice that of the resonance width, complete growth was not possible. Thus our current conclusions are that the phenomena of nonlinear resonance of solitons do qualitatively account for these experimental results, and that more accurate data on these collisions may also give quite good quantitative agreement.

References

[1] F. Ze, N. Hershkowitz, C. Chan and K.E. Lonngren, Phys. Rev. Lett. 42 (1979) 1747.
[2] M.J. Ablowitz, D.J. Kaup, A.C. Newell and H. Segur, Stud. Appl. Math 53 (1974) 249.
[3] H. Flaschka and A.C. Newell, Lect. Notes Phys. 38 (1975) 355.
[4] D.J. Kaup, J. Math. Phys. 16 (1975) 2036.
[5] J.W. Miles, J. Fluid Mech. 79 (1977) 157, 171.
[6] A.C. Newell and L.G. Redekopp, Phys. Rev. Lett. 38 (1977) 377.
[7] D.J. Kaup, A.H. Reiman and A. Bers, Rev. Mod. Phys. 51 (1979) 275.
[8] B.B. Kadomtsev and V.I. Petviashvili, Sov. Phys. Dok. 15 (1970) 539.
[9] M.J. Ablowitz and R. Haberman, Phys. Rev. Lett. 35 (1975) 1185.
[10] D. Anker and N.C. Freeman, On the Soliton Solutions of the Davey–Stewartson equation for long waves (preprint). Also F. Kako and N. Yajima, Interaction of Ion-Acoustic Solitons in Two-Dimensional Space (preprint).
[11] A. Bers, D.J. Kaup and A. Reiman, Phys. Rev. Lett. 37 (1976) 182.
[12] S. Maxon and J. Viecelli, Phys. Rev. Lett. 32 (1974) 4.
[13] E. Cumberbatch, Phys. Fluids 21 (1978) 374.
[14] D.J. Kaup and A.C. Newell, Proc. Roy. Soc. London A361 (1978) 413.
[15] F. Ze, N. Hershkowitz, C. Chan and L.E. Lonngren, Phys. Fluids 22 (1979) 1554.

Multi-shock solutions of random phase three-wave interactions

Allan Reiman[a] and D. Kaup

Department of Physics, Clarkson College, Potsdam, New York 13676
(Received 13 November 1979; accepted 17 October 1980)

Explicit solutions of the equations describing the space–time interaction of three random phase wave packets are presented. The solutions describe the collision and decay of previously known shock-like solutions. Stability of the solutions to a perturbation in initial conditions is also analyzed. For those solutions which are unstable, a closed form solution of the nonlinear evolution of the interaction is given.

I. INTRODUCTION

Wave-wave interactions play a central role in nonlinear plasma processes. The simplest such interaction is that between three plasma waves. Depending on whether the spectral widths of the interacting wave packets is taken to be small or large, this interaction may be described in either the coherent or the random phase approximation. Solutions of the resulting equations are obtained in standard texts by neglecting the spatial dependence of the wave packet amplitudes.[1] However, inclusion of the spatial dependence may lead to qualitatively different behavior. In recent years a rather thorough understanding of the space-time behavior of the lossless, coherent interaction in one spatial dimension has developed.[2] The space-time behavior of the random phase interaction, on the other hand, is not well understood. In this paper we describe multi-shock solutions of the lossless, random phase three-wave equations in time and one spatial dimension. The solutions are found by transformation to a bilinear form of the equations, analogous to the Hirota form of the Korteweg–deVries equation.[3] The bilinear form of the equations also greatly simplifies the stability analysis of our solutions.

Ohsawa and Taniuti discovered shock-like traveling wave solutions to the random phase three-wave interaction.[4] They have further argued that such shocks due to the interaction of two Langmuir waves with an ion-acoustic wave should be observable in laboratory plasmas.[5] Assuming a plasma frequency of 10^9/sec, an electron thermal velocity of 10^8 cm/sec, cold ions, and a Langmuir wave field strength of 1 V/cm, they describe a possible interaction having a shock width of about 9 cm. Ohsawa has presented numerical solutions showing how this shock can be formed by externally excited waves.[6]

Our solutions describe collisions and decays of the Ohsawa–Taniuti shock waves. They bear a relation to the Ohsawa–Taniuti solutions analogous to the relation between n-soliton and single soliton solutions for other equations. The analogy is deepened by our method of derivation of the solutions using a technique of Estabrook and Wahlquist[7] usually associated with n-soliton solutions, and by our bilinear form of the equations.

The nonlinear interaction of a resonant triplet of undamped, positive energy waves is described in the random phase approximation by the equations[1]

$$\left(\frac{\partial}{\partial t}+v_1\frac{\partial}{\partial x}\right)N_1=\frac{1}{|v_2-v_3|}(N_1N_2-N_1N_3+N_2N_3), \tag{1a}$$

$$\left(\frac{\partial}{\partial t}+v_2\frac{\partial}{\partial x}\right)N_2=\frac{1}{|v_1-v_3|}(N_1N_3-N_1N_2-N_2N_3), \tag{1b}$$

$$\left(\frac{\partial}{\partial t}+v_3\frac{\partial}{\partial x}\right)N_3=\frac{1}{|v_2-v_1|}(N_1N_2-N_1N_3+N_2N_3), \tag{1c}$$

where the v_j's are group velocities and the N_j's are action densities (normalized to make the coupling coefficient 1). The resonance conditions,

$$\omega_2=\omega_1+\omega_3, \tag{2a}$$

$$k_2=k_1+k_3, \tag{2b}$$

have been assumed.

If the spatial derivatives in Eqs. (1) are neglected, the resulting ordinary differential equations can be solved analytically. Note that if the N_j's are independent of x, any set of amplitudes satisfying

$$N_1N_2-N_1N_3+N_2N_3=0 \tag{3}$$

is an equilibrium solution. The solutions of the time-only equations always tend to such an equilibrium.

If the N_j's in Eqs. (1) are assumed to be functions only of $x-ut$, where u is some given constant, a set of ordinary differential equations is again obtained. The solution for each N_j is of the form

$$N_j=A_j+B_j\tanh[\nu(x-ut)]. \tag{4}$$

These are the Ohsawa – Taniuti solutions. For $|x|$ sufficiently large they approach an equilibrium solution, satisfying Eq. (3) in the limit $|x|\to\infty$. We will have more to say about these solutions in Sec. III.

In this paper we obtain a set of closed form solutions to Eqs. (1) having a nontrivial dependence on x and t. Our solutions include the Ohsawa–Taniuti solutions as a special case. In Sec. II we transform to a bilinear form of Eqs. (1). We find solutions of the resulting equation which are sums of exponentials, and we show how to transform back to obtain solutions for the N_j's. In Sec. III we describe the resulting solutions for the N_j's. Fin-

[a] Present address: Physics and Astronomy Department, University of Maryland, College Park, Md. 20742.

 0031-9171/81/020228-05$00.90 © 1981 American Institute of Physics

$$N_j = A_j + B_j \tanh(\nu\xi + \mu), \tag{13a}$$

$$\xi = x + [(\omega_2 - \omega_1)/(k_2 - k_1)]t, \tag{13b}$$

$$\nu = (k_2 - k_1)/2, \tag{13c}$$

$$A_j = \frac{v_3}{4}\left(\frac{d_{j1}}{\omega_1} + \frac{d_{j2}}{\omega_2}\right), \tag{13d}$$

$$B_j = \frac{v_3}{4}\left(\frac{d_{j2}}{\omega_2} - \frac{d_{j1}}{\omega_1}\right), \tag{13e}$$

$$\mu = \tfrac{1}{2}\ln(\omega_2 C_2/\omega_1 C_1). \tag{13f}$$

These are shock-like solutions, with the shock front centered at

$$X = -\frac{\omega_2 - \omega_1}{k_2 - k_1}t - \frac{1}{k_2 - k_1}\ln\left(\frac{\omega_1 C_1}{\omega_2 C_2}\right). \tag{14}$$

Away from the shock front, one of the two exponentials dominates, giving an equilibrium solution of the form of Eq. (12) as $|x| \to \infty$.

For $n = 3$ and $n = 4$ the solutions are new. These solutions describe collisions, decays, and coalescence of Ohsawa–Taniuti shocks. First consider the $n = 3$ solutions. We take (without loss of generality) $k_1 < k_2 < k_3$. Then, for x sufficiently large and positive we may neglect the terms containing $\exp(k_1 x + \omega_1 t)$. So, for x large and positive, the $n = 3$ solution is well approximated by a single shock, with shock front centered at

$$X_2 = \frac{\omega_3 - \omega_2}{k_3 - k_2}t + \frac{1}{k_3 - k_2}\ln\left(\frac{\omega_2 C_2}{\omega_3 C_3}\right). \tag{15}$$

Similarly, for x large and negative the solution is well approximated by an $n = 2$ solution with shock front at X_1, given by Eq. (14). If $X_2 \gg X_1$, there is an intermediate region between X_1 and X_2 where the solution is well approximated by the equilibrium solution

$$N_j = v_3 C_2 d_{j2}/(2\omega_2). \tag{16}$$

The solution then corresponds to two well separated shocks. What if $X_1 \ll X_2$? That implies that the terms containing $\exp(k_2 x + \omega_2 t)$ are everywhere dominated by one of the other two exponentials. These terms can be ignored everywhere, giving an $n = 2$ solution determined by C_1, ω_1, k_1 and C_3, ω_3, k_3.

The time evolution of the $n = 3$ solutions is determined by the sign of the expression

$$\frac{\omega_2 - \omega_1}{k_2 - k_1} - \frac{\omega_3 - \omega_2}{k_3 - k_2}. \tag{17}$$

When this expression is negative, the solution is a single shock for $t \to -\infty$ and a double shock for $t \to \infty$, corresponding to decay of an Ohsawa–Taniuti shock. When (17) is positive, the reverse is true, giving coalescence of two Ohsawa–Taniuti shocks to form a single such shock.

In Fig. 1, we show the time evolution of an $n = 3$ solution corresponding to the coalescence of two shocks. The group velocities are $v_2 = 0$, $v_3 = -v_1 = 0.3$, with the solution given by the parameters $k_1 = -1$, $\omega_1 = 1$, $k_2 = 0.147$, $\omega_2 = 0.362$, $k_3 = 0.296$, $\omega_3 = 0.326$, and $C_1 = C_2 = C_3 = 1$.

Figure 1(a) shows N_2 vs x at $t = -130$, when two well-defined shocks are still visible. Figure 1(b) shows N_2 at $t = 10$, after the two shocks have collided and formed a single Ohsawa–Taniuti shock. Note that the time reversal of this solution gives a solution of Eqs. (1) with the group velocities reversed. The time reversed solution describes the decay of an Ohsawa – Taniuti shock.

The analysis for $n = 4$ is similar. We now have six X's determining possible shock fronts. Depending on the relative values of these X's we can now have up to three shocks present simultaneously. Thus, we can have the decay of a single shock to three shocks, or the coalescence of three shocks to form a single shock. The $n = 4$ solutions can also describe the scattering of two shocks off each other to form two different shocks.

IV. STABILITY OF THE SOLUTIONS

Now we determine under what conditions our solutions are sensitive to a small change in the initial conditions. Let the perturbed solution be

$$N_j = N_{j0} + \delta N_j, \tag{18}$$

with $\delta N_j \ll N_{j0}$ initially. We want to know whether δN_j remains much smaller than N_{j3}. We again find it convenient to work with the bilinear form of the equations. Corresponding to Eq. (18), we determine a $\psi_0 + \delta\psi$ from Eqs. (5). Since $\delta N \ll N_0$ initially, $\delta\psi$ may be chosen so that $\delta\psi \ll \psi_0$ initially. From Eqs. (7) it is clear that if $\delta\psi$ remains small, then δN remains small. Conversely, from Eqs. (7) it is clear that if $\delta\psi$ is growing exponentially in time with a growth rate larger than that of ψ_0, δN does not remain small. So, we can perform our stability analysis directly on Eq. (6).

Substitute $\psi = \psi_0 + \delta\psi$ in Eq. (6), with ψ_0 of the form given by Eq. (9), and linearize in $\delta\psi$. The result is an equation of the form

$$\sum_m \sum_{j=1}^{n} P_m(\omega_j, k_j) \exp(k_j x + \omega_j t)\, Q_m\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial x}\right)\delta\psi(x, t) = 0, \tag{19}$$

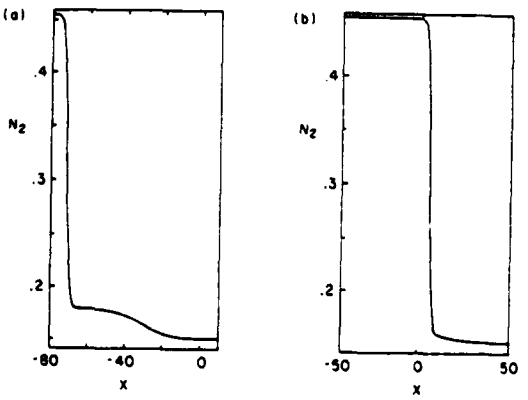


FIG. 1. An $n = 3$ solution showing coalescence of two Ohsawa–Taniuti shocks to form a single such shock. (a) The amplitude of N_2 before coalescence ($t = -130$). (b) The amplitude after coalescence ($t = 10$).

ally, in Sec. IV we consider the stability of the solutions to perturbations of the initial amplitudes. We find that in those cases where a solution is unstable, the perturbed solution is also included in our class of analytic solutions.

II. A BILINEAR FORM OF THE EQUATIONS

In order to obtain solutions of Eqs. (1) we transform to an equivalent set of bilinear equations, analogous to the Hirota form of the Korteweg–deVries equation. To find this transformation we have used a method of Estabrook and Wahlquist,[7] as further developed by Corones,[8] Kaup,[9] and Newell.[10] Much of the laborious algebra involved in this method was done on the MC computer at MIT, using MACSYMA.[11] The derivation of the transformation is outlined in the appendix.

To simplify our bookkeeping somewhat, we make use of the fact that either v_1 or v_3 must be greater than or less than the other two velocities. Since the equations are symmetric with respect to the interchange of N_1 and N_3, and are also symmetric with respect to a reflection about $x = 0$, we may without loss of generality take $v_3 > v_1$, $v_3 > v_2$.

We are still free to work in a reference frame moving at any fixed velocity. The transformation is considerably simplified if we work in a reference frame where the group velocity of the intermediate velocity wave vanishes. In the following we let α correspond to the subscript of the wave having negative group velocity (thus $\alpha = 1$ or $\alpha = 2$).

Let the quantity $\psi(x, t)$ satisfy the equations

$$\psi_{xx} = (a_1 N_1 + a_2 N_2 + a_3 N_3)e^{\theta}, \tag{5a}$$

$$\psi_t = e^{\theta}, \tag{5b}$$

with

$$\theta = -\int_{-\infty}^{x} [v_\alpha a_\alpha N_\alpha(x') + v_3 a_3 N_3(x')]dx', \tag{5c}$$

$$a_3 = 2[v_3(v_2 - v_3)(v_1 - v_3)]^{-1}, \tag{5d}$$

$$a_1 = 2(v_3 - v_2)/[v_3(v_3 - v_1)(v_1 - v_2)^2], \tag{5e}$$

and

$$a_2 = 2(v_1 - v_3)/[v_3(v_1 - v_2)^2(v_3 - v_2)]. \tag{5f}$$

Equations (5) are consistent with Eqs. (1) in the sense that $\psi_{xxt} - \psi_{txx} = 0$. From Eqs. (5) and (1) we find that ψ satisfies the equation

$$2\psi_t[\psi_{ttt} + (v_\alpha + v_3)\psi_{xtt} + v_\alpha v_3 \psi_{xxt}] - 2\psi_{tt}^2 - 2(v_\alpha + v_3)\psi_{xt}\psi_{tt} - v_3 v_\alpha(\psi_{xt}^2 + \psi_{tt}\psi_{xx}) = 0. \tag{6}$$

Given any solution $\psi(x, t)$ satisfying Eq. (6), we obtain a solution of Eqs. (1) by using Eqs. (5). From Eqs. (5) we find

$$N_1 = \frac{v_3(v_3 - v_1)}{v_3 - v_2}\,\frac{\psi_{tt} + (v_2 + v_3)\psi_{xt} + v_2 v_3 \psi_{xx}}{2\psi_t}, \tag{7a}$$

$$N_2 = \frac{v_3(v_1 - v_2)}{(v_1 - v_3)}\,\frac{\psi_{tt} + (v_1 + v_3)\psi_{xt} + v_1 v_3 \psi_{xx}}{(2\psi_t)}, \tag{7b}$$

$$N_3 = v_3[\psi_{tt} + (v_1 + v_2)\psi_{xt} + v_1 v_2 \psi_{xx}]/(2\psi_t). \tag{7c}$$

Equation (6) is bilinear. Each of its terms is quadratic in ψ, which implies that Eq. (6) has solutions of the form

$$\psi = \exp(kx + \omega t). \tag{8}$$

Substituting this in Eq. (6), we find that it is a solution for arbitrary ω and k. Pursuing the analogy with the Hirota form of the Korteweg–deVries equation, we try solutions which are sums of exponentials

$$\psi = \sum_{i=1}^{n} C_i \exp(k_i x + \omega_i t). \tag{9}$$

Substituting in Eq. (6), we find that this is a solution if for all i, j,

$$2\frac{\omega_i}{k_i}\frac{\omega_j}{k_j}\left(\frac{\omega_j - \omega_i}{k_j - k_i} + v_\alpha\right)\left(\frac{\omega_j - \omega_i}{k_j - k_i} + v_3\right) = v_\alpha v_3\left(\frac{\omega_j}{k_j} - \frac{\omega_i}{k_i}\right)^2. \tag{10}$$

For a sum of n exponentials, this gives $n(n-1)/2$ consistency conditions on the $2n$ unknowns (ω's and k's). Two of the unknowns may be fixed by the choice of the space and time normalization. We conclude that the consistency conditions can be satisfied if $n \leq 4$. The consistency condition is, in effect, a nonlinear dispersion relation.

We are interested only in those solutions for which the N_j's given by Eqs. (7) are all positive. If $v_1 = 0$, we must require $\omega_j > 0$, $k_j < 0$, and $|\omega_j/k_j| < |v_2|$, $|v_3|$ for each j. If $v_2 = 0$, we must have $\omega_j > 0$ and $|\omega_j/k_j| > v_3$ for $k_j < 0$ or $|\omega_j/k_j| > v_1$ for $k_j > 0$.

III. THE SOLUTIONS

Substituting the exponential solutions for ψ [Eq. (9)] into Eqs. (7) we obtain solutions for the N_j,

$$N_j = \frac{v_3}{2}\,\frac{\sum_{m=1}^{n} d_{jm} C_m \exp(k_m x + \omega_m t)}{\sum_{m=1}^{n} \omega_m C_m \exp(k_m x + \omega_m t)}, \tag{11a}$$

where

$$d_{1m} = \frac{v_2 - v_1}{v_3 - v_2}(\omega_m + v_2 k_m)(\omega_m + v_3 k_m), \tag{11b}$$

$$d_{2m} = \frac{v_1 - v_2}{v_1 - v_3}(\omega_m + v_1 k_m)(\omega_m + v_3 k_m), \tag{11c}$$

$$d_{3m} = (\omega_m + v_1 k_m)(\omega_m + v_2 k_m). \tag{11d}$$

Here, the ω's and k's must be chosen subject to the nonlinear dispersion relations [Eq. (10)] and subject to the condition that each of the N_j's be greater than or equal to zero. Necessary and sufficient conditons on the ω's and k's in order that the N_j's be positive have been given at the end of the previous section.

For $n = 1$, Eq. (11a) reduces to

$$N_j = v_3 C d_j/(2\omega). \tag{12}$$

These are the time-only steady-state solutions satisfying Eq. (3).

For $n = 2$ we recover the Ohsawa–Taniuti solutions. Equation (11a) then gives

$$N_j = A_j + B_j \tanh(\nu\xi + \mu), \tag{13a}$$

$$\xi = x + [(\omega_2 - \omega_1)/(k_2 - k_1)]t, \tag{13b}$$

$$\nu = (k_2 - k_1)/2, \tag{13c}$$

$$A_j = \frac{v_3}{4}\left(\frac{d_{j1}}{\omega_1} + \frac{d_{j2}}{\omega_2}\right), \tag{13d}$$

$$B_j = \frac{v_3}{4}\left(\frac{d_{j2}}{\omega_2} - \frac{d_{j1}}{\omega_1}\right), \tag{13e}$$

$$\mu = \tfrac{1}{2}\ln(\omega_2 C_2/\omega_1 C_1). \tag{13f}$$

These are shock-like solutions, with the shock front centered at

$$X = -\frac{\omega_2 - \omega_1}{k_2 - k_1}t - \frac{1}{k_2 - k_1}\ln\left(\frac{\omega_1 C_1}{\omega_2 C_2}\right). \tag{14}$$

Away from the shock front, one of the two exponentials dominates, giving an equilibrium solution of the form of Eq. (12) as $|x| \to \infty$.

For $n = 3$ and $n = 4$ the solutions are new. These solutions describe collisions, decays, and coalescence of Ohsawa–Taniuti shocks. First consider the $n = 3$ solutions. We take (without loss of generality) $k_1 < k_2 < k_3$. Then, for x sufficiently large and positive we may neglect the terms containing $\exp(k_1 x + \omega_1 t)$. So, for x large and positive, the $n = 3$ solution is well approximated by a single shock, with shock front centered at

$$X_2 = \frac{\omega_3 - \omega_2}{k_3 - k_2}t + \frac{1}{k_3 - k_2}\ln\left(\frac{\omega_2 C_2}{\omega_3 C_3}\right). \tag{15}$$

Similarly, for x large and negative the solution is well approximated by an $n = 2$ solution with shock front at X_1, given by Eq. (14). If $X_2 \gg X_1$, there is an intermediate region between X_1 and X_2 where the solution is well approximated by the equilibrium solution

$$N_j = v_3 C_2 d_{j2}/(2\omega_2). \tag{16}$$

The solution then corresponds to two well separated shocks. What if $X_1 \ll X_2$? That implies that the terms containing $\exp(k_2 x + \omega_2 t)$ are everywhere dominated by one of the other two exponentials. These terms can be ignored everywhere, giving an $n = 2$ solution determined by C_1, ω_1, k_1 and C_3, ω_3, k_3.

The time evolution of the $n = 3$ solutions is determined by the sign of the expression

$$\frac{\omega_2 - \omega_1}{k_2 - k_1} - \frac{\omega_3 - \omega_2}{k_3 - k_2}. \tag{17}$$

When this expression is negative, the solution is a single shock for $t \to -\infty$ and a double shock for $t \to \infty$, corresponding to decay of an Ohsawa–Taniuti shock. When (17) is positive, the reverse is true, giving coalescence of two Ohsawa–Taniuti shocks to form a single such shock.

In Fig. 1, we show the time evolution of an $n = 3$ solution corresponding to the coalescence of two shocks. The group velocities are $v_2 = 0$, $v_3 = -v_1 = 0.3$, with the solution given by the parameters $k_1 = -1$, $\omega_1 = 1$, $k_2 = 0.147$, $\omega_2 = 0.362$, $k_3 = 0.296$, $\omega_3 = 0.326$, and $C_1 = C_2 = C_3 = 1$.

Figure 1(a) shows N_2 vs x at $t = -130$, when two well-defined shocks are still visible. Figure 1(b) shows N_2 at $t = 10$, after the two shocks have collided and formed a single Ohsawa–Taniuti shock. Note that the time reversal of this solution gives a solution of Eqs. (1) with the group velocities reversed. The time reversed solution describes the decay of an Ohsawa – Taniuti shock.

The analysis for $n = 4$ is similar. We now have six X's determining possible shock fronts. Depending on the relative values of these X's we can now have up to three shocks present simultaneously. Thus, we can have the decay of a single shock to three shocks, or the coalescence of three shocks to form a single shock. The $n = 4$ solutions can also describe the scattering of two shocks off each other to form two different shocks.

IV. STABILITY OF THE SOLUTIONS

Now we determine under what conditions our solutions are sensitive to a small change in the initial conditions. Let the perturbed solution be

$$N_j = N_{j0} + \delta N_j, \tag{18}$$

with $\delta N_j \ll N_{j0}$ initially. We want to know whether δN_j remains much smaller than N_{j0}. We again find it convenient to work with the bilinear form of the equations. Corresponding to Eq. (18), we determine a $\psi_0 + \delta\psi$ from Eqs. (5). Since $\delta N \ll N_0$ initially, $\delta\psi$ may be chosen so that $\delta\psi \ll \psi_0$ initially. From Eqs. (7) it is clear that if $\delta\psi$ remains small, then δN remains small. Conversely, from Eqs. (7) it is clear that if $\delta\psi$ is growing exponentially in time with a growth rate larger than that of ψ_0, δN does not remain small. So, we can perform our stability analysis directly on Eq. (6).

Substitute $\psi = \psi_0 + \delta\psi$ in Eq. (6), with ψ_0 of the form given by Eq. (9), and linearize in $\delta\psi$. The result is an equation of the form

$$\sum_m \sum_{j=1}^{n} P_m(\omega_j, k_j)\exp(k_j x + \omega_j t)\, Q_m\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial x}\right)\delta\psi(x, t) = 0, \tag{19}$$

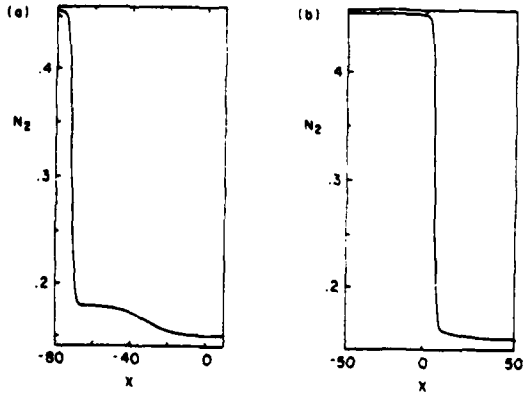


FIG. 1. An $n = 3$ solution showing coalescence of two Ohsawa–Taniuti shocks to form a single such shock. (a) The amplitude of N_2 before coalescence ($t = -130$). (b) The amplitude after coalescence ($t = 10$).

where P_m and Q_m are monomials, and the subscript m labels the different terms in the linearized equation. Laplace transform in t and x,

$$\sum_m \sum_{j=1}^{n} P_m(\omega_j, k_j) Q_m(p-\omega_j, k-k_j) \times \delta\hat{\psi}(k-k_j, p-\omega_j) = 0 . \tag{20}$$

Invoking the independence of $\delta\hat{\psi}$ at different points we may remove the sum over j. Defining $\hat{p}_j = p - \omega_j$ and $\hat{k}_j = k - k_j$, and assuming $\delta\hat{\psi}(\hat{k}_j, \hat{p}_j) \neq 0$, we finally obtain

$$\sum_m P_m(\omega_j, k_j) Q_m(\hat{p}_j, \hat{k}_j) = 0 . \tag{21}$$

This is exactly the same as the consistency condition, Eq. (10). When a closed form n-shock solution is unstable to a perturbation of the initial conditions, the nonlinear evolution of the instability is given by another closed form n-shock solution.

As an example of an instability, we recall the $n = 3$ solutions corresponding to decay of a single shock. This describes an instability of the Ohsawa–Taniuti solution.

V. CONCLUSION

The shock-like solutions of Eqs. (1) discovered by Ohsawa and Taniuti have been found to be a special case of a more general class of multi-shock solutions. These multi-shock solutions describe decay and collisions of the Ohsawa–Taniuti shocks. Any instability of such a multishock solution is itself given by a multishock solution.

The numerical work of Ohsawa, showing that the single shock solutions evolve from a variety of initial conditions, suggest that more generally the multi-shock solutions should be a common feature in those systems where instabilities saturate due to mode coupling of random phase waves. This contrasts with coherent three-wave interactions, which commonly lead to multi-soliton solutions.[2]

ACKNOWLEDGMENTS

We are grateful to A. Bers, who first suggested this problem to us. Helpful conversations with Alan Newell are also appreciated.

This work was supported by the Office of Naval Research, Contract No. N00014-76-C-0867.

APPENDIX

The bilinear form of Eqs. (1), presented in Sec. II, was obtained using a method of Estabrook and Wahlquist,[7] as further developed by Corones,[8] Kaup,[9] and Newell.[10] Originally developed in connection with equations having an inverse scattering transform, the method has been shown to also give the Cole–Hopf transformation for Burgers' equation.[6] This motivated its application to Eqs. (1).

The basic idea of the Wahlquist–Estabrook method as applied to Eqs. (1) is to embed the nonlinear Eqs. (1) as the consistency condition for a pair of linear equations,

$$\varphi_x = P(N_1, N_2, N_3)\varphi , \tag{A1a}$$

$$\phi_t = Q(N_1, N_2, N_3)\phi . \tag{A1b}$$

Here ϕ is an auxiliary vector function of x and t, and P and Q are matrix functions of the N_j's. The condition for consistency of Eqs. (A1a) and (A1b) is

$$\phi_{xt} - \phi_{tx} = 0 , \tag{A2}$$

or

$$P_t - Q_x + PQ - QP = 0 . \tag{A3}$$

If P and Q are given functions of the N_j's, (A3) is a partial differential equation (in general, nonlinear) for the N_j's. We would like to choose the matrix functions P and Q so that Eq. (A3) is the same as Eq. (1).

For those equations which are exactly soluble by an inverse scattering transform, an embedding of the form (A1) exists with a free parameter. Equations (A1) then correspond to the associated eigenvalue problem, with the free parameter corresponding to the eigenvalue. For Burgers' equation, the embedding leads to an equivalent linear equation. Neither of these is the case for Eqs. (1). Nevertheless, an embedding of Eqs. (1) does exist, and it leads to the solutions described in this paper. This suggests to us that the Estabrook–Wahlquist method may be more generally applicable than its original derivation would suggest.

With Eq. (A3) as our starting point, we apply the chain rule

$$P_t = P_{N_1} N_{1t} + P_{N_2} N_{2t} + P_{N_3} N_{3t} , \tag{A4a}$$

$$Q_x = Q_{N_1} N_{1x} + Q_{N_2} N_{2x} + Q_{N_3} N_{3x} , \tag{A4b}$$

and use Eq. (1) to substitute for N_{jt}. The resulting equation contains terms with N_{jx}. Since the initial value of the N_j's is arbitrary, we demand that the coefficient of each N_{jx} vanish. We obtain

$$v_j P_{N_j} = Q_{N_j}, \quad j = 1, 2, 3 . \tag{A5}$$

Taking cross derivatives of these equations, we find

$$v_j P_{N_j N_k} = v_k P_{N_k N_j} . \tag{A6}$$

Since the group velocities are assumed all different, Eq. (A6) implies

$$P_{N_j N_k} = 0, \tag{A7}$$

or

$$P = \sum_{j=1}^{3} g_j(N_j) + Z , \tag{A8a}$$

$$Q = -\sum_{j=1}^{3} v_j g_j(N_j) + W , \tag{A8b}$$

where g_j, W, and Z are matrices to be determined. Substituting Eqs. (A8) back into the consistency equation, we determine that $g_j''' = 0$, so that g_j must be of the form

$$g_j = N_j X_j + N_j^2 V_j , \tag{A9}$$

where we have absorbed the possible constant term into W and Z. Again substituting back into the consistency equation, we find that each V_j must commute with all other matrices. So the V_j's can be set to zero without changing the structure of the matrix algebra.

We have concluded that P and Q must be of the form

$$P = \sum_{j=1}^{3} X_j N_j + Z, \tag{A10a}$$

$$Q = -\sum_{j=1}^{3} v_j X_j N_j + W. \tag{A10b}$$

Substitute these expressions back into the consistency equation, setting the coefficient of each term in the resulting polynomial in the N_j's to zero, to obtain a set of commutation relations. To simplify the expression of these relations we define

$$Y_1 = X_1, \quad Y_2 = -X_2, \quad Y_3 = X_3. \tag{A11}$$

Let

$$[AB] \equiv AB - BA.$$

In terms of the Y_j's our commutation relations are

$$(v_i - v_j)[Y_i, Y_j] = \frac{1}{|v_2 - v_3|} Y_1 + \frac{1}{|v_3 - v_1|} Y_2 + \frac{1}{|v_2 - v_1|} Y_3, \quad i \neq j,$$

$$[Y_j, W] = v_j[Z, Y_j], \quad [W, Z] = 0. \tag{A12}$$

The commutator of Y_m with Z or W is not determined. We note that in order to recover Eqs. (1) from Eqs. (A1) and (A10)–(A12), we must also use the fact that $|v_2 - v_1|N_1 - |v_2 - v_3|N_3$ is a conserved quantity.

We now close the Lie algebra, consistent with the Jacobi identity

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0.$$

This step involves much laborious algebra. We did the algebra on the MIT MC computer, using MACSYMA.[11] Our algorithm was to define new elements of the Lie algebra equal to the unknown commutators, evaluating all Jacobi identities at each step, continuing to add new elements until the algebra collapsed back down and then closed. The result is

$$[Y_1, Z] = -a_1 W, \quad [Y_3, Z] = -a_3 W,$$

$$Y_2 = (v_1 - v_3)\left(\frac{1}{v_2 - v_3} Y_1 + \frac{1}{v_1 - v_2} Y_3\right),$$

where a_j has been defined in Eqs. (5). Note that the algorithm used gave us the largest Lie algebra consistent with Eqs. (A12). We are, therefore, assured that there can be no free parameter in the closure, of the sort which appears when an inverse scattering transform exists.

Having a closed Lie algebra, we now easily find a 3×3 matrix representation,

$$W = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix},$$

$$Z = -\frac{2}{3}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ a_1 a_3 (v_1 - v_3) & 0 & 0 \end{bmatrix},$$

$$Y_1 = \frac{v_1 a_1}{2}\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$Y_2 = -\frac{1}{2}\begin{bmatrix} v_2 a_2 & 3(v_1 - v_3)[v_1 a_1 (v_1 - v_2)]^{-1} & 0 \\ 0 & -v_2 a_2 & 0 \\ 0 & 0 & v_2 a_2 \end{bmatrix}$$

$$Y_3 = \frac{1}{2}\begin{bmatrix} v_3 a_3 & -3/(v_1 a_1) & 0 \\ 0 & -v_3 a_3 & 0 \\ 0 & 0 & v_3 a_3 \end{bmatrix}.$$

This gives us the desired matrices P and Q.

Now we have embedded Eqs. (1) as the consistency condition for a set of six linear equations in three variables. Eliminating two of the variables in terms of the third, we obtain Eqs. (5). Using Eqs. (1) to eliminate the N_j's from Eqs. (5), we obtain Eq. (6).

[1] R. Davidson, *Methods in Nonlinear Plasma Theory* (Academic, New York, 1972); V. N. Tsytovich, *Nonlinear Effects in Plasma* (Plenum, New York, 1970).

[2] A review of this work is contained in D. Kaup, A. Reiman, and A. Bers, Rev. Mod. Phys. 51, 275 (1979); and in A. Reiman, Rev. Mod. Phys. 51, 311 (1979).

[3] R. Hirota, Phys. Rev. Lett. 27, 1192 (1971); see also G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).

[4] Y. Ohsawa and T. Taniuti, J. Phys. Soc. Jpn. 35, 907 (1973).

[5] Y. Ohsawa and T. Taniuti, J. Phys. Soc. Jpn. 36, 313 (1974).

[6] Y. Ohsawa, J. Phys. Soc. Jpn. 37, 1129 (1974).

[7] H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. 16, 1 (1975); H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. 17, 1293 (1976).

[8] J. Corones, J. Math. Phys. 17, 756 (1976); J. Math. Phys. 18, 163 (1977).

[9] D. Kaup, MIT Plasma Research Report, PRR 79/6 (1979).

[10] A. Newell, Proc. R. Soc. London 365, 283 (1979).

[11] A. Reiman, in *Proceedings of the Second MACSYMA User's Conference* (Massachusetts Institute of Technology, Cambridge, Mass.) p. 385; Comput. Math. (to be published).

Physica 3D (1981) 1 & 2, 374-395
© North-Holland Publishing Company

THE SOLUTION OF THE GENERAL INITIAL VALUE PROBLEM FOR THE FULL THREE DIMENSIONAL THREE-WAVE RESONANT INTERACTION†

D.J. Kaup

Dept. of Physics
Clarkson College of Technology
Potsdam, NY 13676

The general initial value problem for the full three-dimensional three-wave resonant interaction is solved by inverse scattering. The existence and uniqueness of the solution is established when all waves have a positive energy.

I. INTRODUCTION

This is the fourth article in a series of articles on the three-dimensional three-wave resonant interaction (3D3WRI), and with it, we complete the theoretical development by presenting the method of solution for the general initial value problem. This series was initiated shortly after Cornille[1] developed an alternate set of integral equations for solving the 3D3WRI. The first set was originally proposed by Zakharov and Shabat[2]. What was significant about Cornille's results was that he had started from the scattering problem developed by Ablowitz and Haberman,[3] which could be readily cast into characteristic coordinates, which then created considerable simplifications in the required analysis. Using his results and also characteristic coordinates, we[4] were then able to develop a method for solving a special class of initial value problems, called "separable initial value problems." These separable problems only require the three initial profiles to be nonoverlapping, which is frequently the physical situation. In this first article,[4] it was not necessary to consider or use inverse scattering. We simply applied Cornill's results, in characteristic coordinates, to solving the separable initial value problem.

Now, whenever one has a set of integral (inverse scattering) equations, one can usually work backwards to the required scattering problem, and in the process determine the necessary integral dispersion relations, and thereby what analytical properties are required. Once one knows this, one can restart from the required scattering problem, and since one now knows what the analytical properties must be, one can rapidly move forward to solve both the direct and inverse scattering problem. Our initial results[5] on this were presented at the conference on "Mathematical Methods and Applications of Scattering Theory" at Catholic University, 21-25 May, 1979, wherein we were able to derive and prove almost all of the ansatzs required by Cornille. In the process, it also became necessary to solve a simpler two-dimensional scattering problem with which one could then map each separable initial profile into a set of scattering data. Then it was shown that from these three sets of initial scattering data, one could then construct the three sets of scattering data required for reconstruction of the final profiles. Thus it was now possible to go directly from initial profiles to final profiles.

In a third article,[6] these results were further expanded on and made more rigorous, in addition to several new results. Among these were the proofs of the existence and uniqueness of the solution for this simpler two-dimensional scattering

†Research supported in part by the Office of Naval Research, Contract No. N00014-76-C-0867 and NSF Grant No. MCS 7903498.

problem, of the general solution of the three-dimensional direct and inverse scattering problems for separable initial value problems, of the existence and uniqueness of the solution of the three-dimensional inverse scattering equations, closed form solutions for the scattering data when the initial profiles, as functions of the characteristic coordinates, are separable, a physical interpretation for the scattering data, and an infinity of conserved quantities, of which only the first two are local in form. Furthermore, many analogies between the 3D3WRI and the one-dimensional three-wave resonant interaction[7] (1D3WRI) were noted.

Although these results were quite comprehensive, they still could not be used to address the general initial value problem, wherein the initial profiles could be overlapping (nonseparable). And furthermore, the above results were based on the assumption that a solution existed globally, in that all profiles remained integrable and square integrable for all time. This was analogous to assuming that a solution existed, and then using that assumption to prove that the solution existed. Being a circular argument, one could thus only show consistency. However, the tools required in these previous results were exactly those required for the more general solution, and these results also allowed one to isolate precisely those items and methods to be required in the more general solution, which is finally presented in this article.

Of course, it is well recognized that the Ablowitz-Haberman scattering problem,[3] and Cornill's results[1] were not the first. Earlier, Zakharov and Shabat[2] had obtained an L and a B operator for the 3D3WRI, and based on this, Zakharov[8] and others[9] have obtained closed form solutions for the 3D3WRI. However, it just simply happened that the Ablowitz-Haberman[3] form could be most conveniently cast into characteristic coordinate form, and we cannot overemphasize the fact that this is itself allowed considerable simplification in the required analysis. Then when we move on to the general initial value problem herein, we find that we are now working directly with the Zakharov-Shabat form,[2] and we are now solving this scattering problem, both the direct and the inverse scattering parts. And, armed with the experience from the previous simpler results, the solution can be rapidly and easily obtained.

In Section II, we treat the direct scattering problem for the general initial value problem, assuming only that the initial profiles are integrable and square-integrable. Here we find that there are eight sets of fundamental solutions (three fundamental solutions per set) for a total of 24 different solutions. Any one of these eight fundamental sets is complete, in that any solution to the scattering problem can be expanded in terms of the three fundamental solutions in a given set. The analytical properties of each of these fundamental solutions are obtained, reflection and transmission coefficients are defined, and each transmission coefficient is shown to possess a unique inverse. As a consequence of this latter property, one can show that no bound states exist and therefore no soliton solutions exist for the 3D3WRI, when the profiles are integrable.[6]

In Section III, we solve the inverse scattering problem, noting first that there is a very large number of possible forms to choose from. Since our principle interest here is simply to show that the solution of the general initial value problem of the 3D3WRI does exist and is unique, we shall simply arbitrarily choose one form to work with, ignoring for now the other possible forms. We then obtain the necessarily linear dispersion relations, prove that the transformation kernels exist, obtain the inverse scattering (integral) equations, and prove that the solution of these inverse scattering equations exists and is unique in all cases, except for the explosive case. In this explosive case, at least one of the three waves must be a negative energy wave, and it is known that the singular solutions can then evolve from nonsingular solutions.[8,9] Thus, we naturally do expect a certain lack of uniqueness and existence of the solution for this explosive case. And for this reason, we do not attempt to determine just exactly what may be sufficient conditions on the initial data for an explosive case solution to exist and be unique.

In Section IV, we give some concluding remarks, comparing these inverse scattering equations against other known forms,[5,6] and noting their inequivalences. Also, we note how these fundamental solutions are related to those presented earlier,[5,6] and how one may determine the final profiles directly from initial overlapping data.

An important point not to be underemphasized here, is simply that two and three dimensional inverse scattering problems are indeed doable. Unknown to most of us, several simple two-dimensional systems had already been solved by L.P. Niznik[10] including the wave equation and the two-component Dirac equation. In fact, Niznik and Tarasov[11] have also obtained some results related to the same problem as we shall treat here. However, in their short note, they were not able to give any proofs or derivations, and we also note the bounds they require on the potentials to be more stringent than ours. However, we should also point out that these two different investigations of basically the same problem further differ in two other aspects. First is the different viewpoints involved. Niznik is interested in solving the inverse scattering problem, and formally he did so. However, for us, we must go further and apply the inverse scattering procedure to the 3D3WRI. Second, our derivations and proofs here shall differ from Niznik's, in that we shall approach (2.1) as being a physical scattering problem, and shall derive the inverse scattering equations from the analytical properties of the fundamental solutions of (2.1).

And finally, as further proof that two or three-dimensional inverse scattering is indeed doable, we simply mention that Professor Manakov[12] has just recently solved the inverse scattering for the time dependent Schrödinger equation, which is also reported in these same proceedings.

II. THE DIRECT SCATTERING PROBLEM

We start with the scattering problem in characteristic coordinates,[4] which is

$$\partial_k \psi_i = \gamma_k q_j^* \psi_k \quad , \tag{2.1a}$$

$$\partial_i \psi_k = \gamma_i q_j \psi_i \quad , \tag{2.1b}$$

where i, j, and k are cyclic in (2.1), ψ_i is the i^{th} component of the solution, γ_i carries the sign of the i^{th} coupling constant ($\gamma_i = \pm 1$), and q_i is the envelope (profiles) of the i^{th} wave. The integrability condition for the six equations contained in (2.1) are the three equations

$$\partial_i q_i = \gamma_i a_j^* q_k^* \quad , \tag{2.2}$$

with i, j, and k still cyclic. In (2.1) and (2.2), $\partial_i = \partial/\partial x_i$ where x_i is the i^{th} characteristic coordinate. These coordinates are related to ordinary space-time coordinates via

$$\partial_i = -\partial_t - \vec{v}_i \cdot \vec{\nabla} \quad , \tag{2.3a}$$

$$\partial_i x_4 = 0 \quad , \tag{2.3b}$$

where $\vec{v}_i$ is the group velocity of the i^{th} envelope, q_i, and x_4 is simply a fourth coordinate required to make the set (x_1, x_2, x_3, x_4) complete.[4] Of course, the envelopes in (2.2) may depend on x_4, but since differentiations with respect to x_4 never occur, we then simply solve (2.2) at any fixed value of x_4. Naturally,

we assume the three operators $(\partial_1,\partial_2,\partial_3)$ to be linear independent, otherwise (2.2) would degenerate into the 1D3WRI.

In References 5 and 6, the global solutions of (2.1) were analyzed, assuming the profiles, q_1, q_2, and q_3 to be suitable integrable and square integrable. Here, we shall seek to solve (2.1) at a fixed value of t, not globally. From (2.3), we have that the characteristic coordinates are related to time via

$$x_i + x_j + x_k = -t \quad . \tag{2.4}$$

Thus if we constrain t, only two of the characteristic coordinates remain independent. We shall choose the two independent coordinates to be x_i and x_k (for a given choice of i, j, k) and if we constrain t = 0, then

$$x_j(t=0) = -x_i - x_k \quad . \tag{2.5}$$

In order to constrain (2.1) to t = 0, we first must eliminate all ∂_t's from (2.1) by taking the proper linear combinations of the six equations in (2.1). From (2.1) and (2.3), one can obtain, at t = 0,

$$\partial_k\psi_i = \gamma_k q_j^*\psi_k - \gamma_j q_k\psi_j \tag{2.6a}$$

$$(\partial_k-\partial_i)\psi_j = \gamma_k q_i\psi_k - \gamma_i q_k^*\psi_i \tag{2.6b}$$

$$\partial_i\psi_k = \gamma_i q_j\psi_i - \gamma_j q_i^*\psi_j \tag{2.6c}$$

where

$$\partial_k = (\vec{v}_j-\vec{v}_k)\cdot\vec{\nabla} \quad , \tag{2.7a}$$

$$\partial_i = (\vec{v}_j-\vec{v}_i)\cdot\vec{\nabla} \quad , \tag{2.7b}$$

are simply the partial derivatives with respect to x_k and x_i when ψ is constrained to be on the surface $x_i + x_j + x_k = 0$. We remark that (2.6) and (2.7) and all further results will no longer be valid for i, j, and k cyclic. Thus we could easily replace i, j, and k by 1, 2, and 3 at this point. However, simply to emphasize that we may choose i, j, and k to be any combination of 1, 2, and 3, we shall continue to use the i, j, k notation.

A general solution of (2.6) is

$$\psi_i(x_i,x_k) = g_i(x_i) + \int_{x_k}^{\infty} (\gamma_j q_k\psi_j-\gamma_k q_j^*\psi_k)(x_i,u)du \quad , \tag{2.8a}$$

$$\psi_j(x_i,x_k) = g_j(x_i+x_k) + \int_{x_k}^{\infty} (\gamma_i q_k^*\psi_i-\gamma_k q_i\psi_k)(x_k+x_i-v,v)dv \quad , \tag{2.8b}$$

$$\psi_k(x_i,x_k) = g_k(x_k) + \int_{x_k}^{\infty} (\gamma_j q_i^*\psi_j-\gamma_i q_j\psi_i)(w,x_k)dw \quad , \tag{2.8c}$$

where g_i, g_j, and g_k are arbitrary functions. Since (2.6) is linear, we may replace each $g_n(u)$ by a plane wave or zero. There are three ways of doing this for each choice of the limits on the integrals in (2.8). We define the i^{th} fundamental solution, ψ^i, of (2.8) by

$$g_i(x_i) = e^{i\zeta x_i} \quad , \tag{2.9a}$$

$$g_j = g_k = 0 \quad , \tag{2.9b}$$

the j^{th} fundamental solution, ψ^j, by

$$g_i = g_k = 0 \quad , \tag{2.10a}$$

$$g_j(x_i+x_k) = e^{-i\zeta(x_i+x_k)} \quad , \tag{2.10b}$$

and the k^{th} fundamental solution, ψ^k, by

$$g_i = g_j = 0 \quad . \tag{2.11a}$$

$$g_k(x_k) = e^{i\zeta x_k} \quad . \tag{2.11b}$$

At this point we shall pause and describe how (2.6) is a scattering problem, by demonstrating that the solution (2.8) for the i^{th} fundamental solution, (2.9), contains an incident wave, a transmitted wave, and two scattered waves. In Figure 1, we represent the three initial separated profiles by circles, where the arrows on the circles indicate the direction in which the profiles are initially moving. Then by (2.8a) and (2.9), the ψ^i_i component propagates along the vertical characteristics in a downward direction as indicated. For values of x_k above all profiles, $\psi^i_i = e^{i\zeta x_i}$, and is simply an incident plane wave. (Note that the wave vector of ψ^i_i is in the x_i direction, while ψ^i_i propagates in the x_k direction.) As one moves down the characteristics, when they intersect the potentials, ψ^i_i is modified by the interaction, and finally emerges at the bottom as the "transmitted wave" with its profile no longer $e^{i\zeta x_i}$, except where the characteristics do not intersect the profiles. Meanwhile, due to the interaction of ψ^i_i with the profiles, (2.8b) shows that some ψ^i_j will be created, and is shown in Figure 1 as emerging from the lower right of the interaction region, and it then propagates along its own characteristics. Similarly, (2.8c) shows that ψ^i_k created to the left, propagating along the horizonal characteristics. These latter two waves are "reflected waves". Naturally, their profiles will be determined by the three potentials. And, we shall show how, by knowing certain profiles of these reflected waves, one may reconstruct the potentials.

We remark at this point that there are 8 different ways in which one could define these fundamental solutions, depending on whether one chooses $+\infty$ or $-\infty$ as one of the limits on each of the integrals in (2.8). Each possible choice will give another equivalent set of fundamental solutions. Since each set contains three fundamental solutions, we have a total of 24 possible solutions to consider. For the moment, we shall only consider the set given explicitly by the limits chosen in (2.8). First, we shall analyze the analytical properties of the j^{th} fundamental solution of this set, showing that it is analytic in the lower half ζ-plane (LHP).

We define

$$P_n(x_i,x_k) = \psi^j_n(x_i,x_k)e^{i\zeta(x_i+x_k)} \qquad (n=i,j,k) \tag{2.12a}$$

and

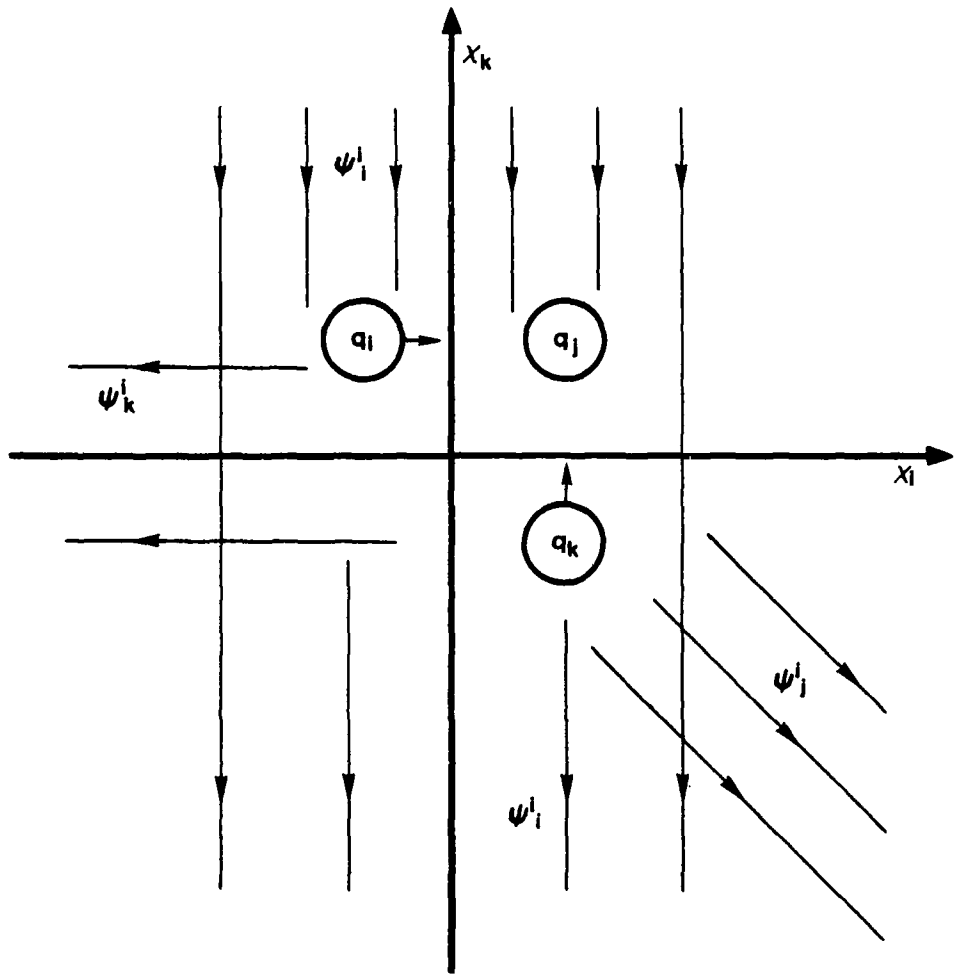


Fig. 1 - Graphics representation of the ψ^i solution of (2.8). The incident wave ($\psi^i{}_i$) is at the top, the transmitted wave ($\psi^i{}_i$) at the bottom, and the two reflected waves ($\psi^i{}_j$ and $\psi^i{}_k$) are at the left and right.

$$p_i(x_i,x_k) = \gamma_j \int_{x_k}^{\infty} e^{-i\zeta(w-x_k)} q_k P_j(x_i,w)dw$$

$$- \gamma_j\gamma_k \int_{x_k}^{\infty} dw q_j^*(x_i,w) e^{-i\zeta(w-x_k)} \int_{x_i}^{\infty} du\, e^{-i\zeta(u-x_i)} q_i^* P_j(u,w) , \qquad (2.12b)$$

$$p_k(x_i,x_k) = \gamma_j \int_{x_i}^{\infty} du e^{-i\zeta(u-x_i)} q_i^* P_j(u,x_k)$$

$$- \gamma_i\gamma_j \int_{x_j}^{\infty} du q_j(u,x_k) e^{-i\zeta(u-x_i)} \int_{x_k}^{\infty} e^{-i\zeta(w-x_k)} q_k P_j(u,w) dw \ . \qquad (2.12c)$$

Then simple substitution of (2.8b) and (2.8c) into (2.8a) gives

$$P_i = p_i + \gamma_i\gamma_k \int_{x_k}^{\infty} dw e^{-i\zeta(w-x_k)} q_j^*(x_i,w) \int_{x_i}^{\infty} du e^{-i\zeta(u-x_i)} q_j P_i(u,w) \ , \qquad (2.13a)$$

and similarly

$$P_k = p_k + \gamma_i\gamma_k \int_{x_i}^{\infty} du e^{-i\zeta(u-x_i)} q_j(u,x_k) \int_{x_k}^{\infty} dw e^{-i\zeta(w-x_k)} q_j^* P_k(u,w) \ . \qquad (2.13b)$$

Let us now assume that the profiles at t = 0 may be bounded as follows. Assume that there exists a function, V(u), such that at t = 0 and for the initial profiles,

$$|q_i(x_i,x_k)| \le V(x_i+x_k)V(x_k) \ , \qquad (2.14a)$$

$$|q_j(x_i,x_k)| \le V(x_i)V(x_k) \ , \qquad (2.14b)$$

$$|q_k(x_i,x_k)| \le V(x_i)V(x_i+x_k) \ , \qquad (2.14c)$$

with V(u) being both integrable and square integrable so that

$$\int_{-\infty}^{\infty} [V(u) + V^2(u)]du < \infty$$

Then for ζ in the LHP, it follows from the Neumann series solution of (2.13) that

$$|P_i| \le V(x_i) I_0 \int_{x_i}^{\infty} du V(u) \int_{x_k}^{\infty} dv V^2(v) \ |p_i(u,v)| \ , \qquad (2.15a)$$

$$|P_k| \le V(x_k) I_0 \int_{x_i}^{\infty} du V^2(u) \int_{x_k}^{\infty} dv V(v) \ |p_k(u,v)| \ , \qquad (2.15b)$$

where

$$I_0 = I_0(2\alpha_2) \ , \qquad (2.16a)$$

$$\alpha_2 = \int_{-\infty}^{\infty} V^2(u)du \ , \qquad (2.16b)$$

we now also define

$$\alpha_1 = \int_{-\infty}^{\infty} V(u)du \ , \qquad (2.16c)$$

and $I_0(z)$ is the modified Bessel function of order zero. Inserting (2.12) into (2.15), for ζ in the LHP, we obtain

$$|P_i| \leq U_2(x_k) I_0 V(x_i) \int_{x_i}^{\infty} du \int_{x_i}^{\infty} dw |P_j(u,w)| |V(u+w)| [V^2(u)+\alpha_2 V^2(w)] \ , \tag{2.17a}$$

$$|P_k| \leq U_2(x_i) I_0 V(x_k) \int_{x_i}^{\infty} du \int_{x_k}^{\infty} dw |P_j(u,w)| |V(u+w)| [V^2(w)+\alpha_2 V^2(u)] \ , \tag{2.17b}$$

as bounds on P_i and P_k. In (2.17a), the function U_2 is defined by

$$U_2(u) = \int_u^{\infty} V^2(v) dv \ , \tag{2.18a}$$

and we shall later also require the function

$$U_1(u) = \int_u^{\infty} V(v) dv \ . \tag{2.18b}$$

We shall now use (2.17) to find a bound on P_j. From (2.8b) and (2.12a), we have

$$|P_j - 1| \leq \alpha_2(1+\alpha_2) I_0 U_2(x_k) V(x_i+x_k) \int_{x_k}^{\infty} dw \int_{x_i+x_k-w}^{\infty} du |P_j(u,w)| |V(u+w)| \cdot [V^2(u)+V^2(w)] \ . \tag{2.19}$$

It now becomes convenient to consider $P_j(x_i, x_k)$ as a function of $x_i + x_k$ and x_k. Thus, we define

$$\tilde{P}_j(x_\ell, x_k) \equiv P_j(x_i, x_k) \ , \tag{2.20a}$$

where

$$x_\ell = x_i + x_k \ . \tag{2.20b}$$

Then

$$|\tilde{P}_j - 1| \leq \alpha_2(1+\alpha_2) I_0 U_2(x_k) V(x_\ell) \int_{x_\ell}^{\infty} du \int_{-\infty}^{\infty} dw |\tilde{P}_j(w,u)| |V(u)| [V^2(w)+V^2(u-w)] . \tag{2.21}$$

the Neumann series solution of which gives

$$|\tilde{P}_j - 1| \leq \beta V(x_\ell) U_1(x_\ell) U_2(x_k) \exp[\beta \alpha_2 U_2(x_\ell)] \ , \tag{2.22}$$

where

$$\beta = 2\alpha_2^2(1+\alpha_2) I_0 \ . \tag{2.23}$$

Now, inserting (2.20) and (2.22) into (2.17) allows us to obtain the bounds

$$|P_i| \leq \tfrac{1}{2} \alpha_1 \alpha_2 \beta^2 V(x_i) U_2(x_k) e^{\beta \alpha_2^2} \ , \tag{2.24a}$$

$$|P_k| \leq \tfrac{1}{2}\alpha_1\alpha_2\beta^2 V(x_k)U_2(x_i)e^{\beta\alpha_2^2} \tag{2.24b}$$

In a similar manner, one may show that the derivative of $\psi^j e^{i\zeta(x_i+x_k)}$ with respect to ζ is also bounded for ζ in the LHP, but not on the real axis. Thus we then have that $\psi^j e^{i\zeta(x_i+x_k)}$ is an analytic function of ζ for ζ in the LHP when (2.14) is satisfied. For $\psi^j d^{i\zeta(x_i+x_k)}$ to be analytic on the real axis, one also needs

$$\int_{-\infty}^{\infty} |u|V(u) < \infty \quad , \tag{2.25}$$

as a sufficient condition.

We shall now discuss in general all 24 solutions mentioned earlier. Each one of these solutions may be represented in the following graphic manner, which we shall illustrate with the above solution. For convenience, we shall assume at $t = 0$, that all initial profiles are on compact support, and thus there exists a circle in the $x_i x_k$-plane which circumscribes all three initial profiles. This region of the $x_i x_k$-plane shall be graphically represented by ◯. Consider now the above solution for ψ_i^j. From the bounds given by (2.24a), we observe that ψ_i^j is only nonzero for x_i within the range of this circumscribed circle, and for x_k within the circle, or below it. This part of the solution may be graphically represented by



where the circle represents where the potentials (initial profiles) are located, and the three vertical lines outline the region where ψ_i^j is in general nonzero. Similarly from (2.24b), we see that we may represent ψ_k^j by



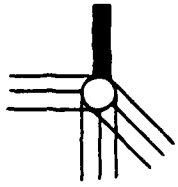
and from (2.22), we may represent ψ_j^j by



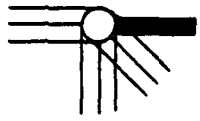
In the above, one notes that we have included a bold heavy straight line opposite the triplet of lines. This line will be used to represent the "incident wave," which for this solution, is the $e^{-i\zeta(x_i+x_k)}$ part of ψ_j^j. Now, one should further note that each component of ψ^j represented above has different characteristic lines. Thus without any confusion, we may superimpose the above graphic representations, obtaining



as a graphic representation of the above solution for ψ^j. Similarly, the solution for ψ^i, defined by (2.9), differs from the above only in the positioning of the bold line, and thus



graphically represents the solution ψ^i. (Compare with Figure 1.) Likewise,



graphically represents the solution ψ^k.

In addition to the above three fundamental solutions, one may define 21 additional fundamental solutions (not independent). They differ from the above in the positioning of the triple lines, and also occur in sets of three. Inside each set, each solution differs from the other two only in the positioning of the bold (incident) line. The graphical representation of all 24 of these fundamental solutions is given in Fig. 2. In Fig. 2, and from now on, we have expanded our notation for these fundamental solutions by adding another superscript to ψ, and thus the solution is designated by ψ^{nr} (n = 1,2,...,8 and r = i,j, or k). The first superscript simply denotes which triplet of fundamental solutions, and the second one refers to which choice of the g's, given by (2.9) - (2.11), are made. One should note that ψ^{6r} is the set of fundamental solutions given by (2.8), and ψ^{6j} is that fundamental solution whose analytical properties we explicitly worked out. Note that in the upper right corner of each graphic representation, we have placed either a U, R, or L. The U indicates that the solution is analytic in the upper half plane (UHP), the L indicates analytic in the LHP, and the R indicates that the solution in general only exists for ζ real. As we have already seen, note that ψ^{6j} is indicated to be analytic in the LHP.

We shall now proceed to define a scattering matrix for each of the eight triplets of fundamental solutions represented in Fig. 2. We define a set of ρ's by

$$\psi_i^{nr}(\zeta;x_i,x_k\to\pm\infty) = \int_{-\infty}^{\infty}\frac{d\lambda}{2\pi}\rho_{i\pm}^{nr}(\zeta,\lambda)e^{i\lambda x_i}\quad, \tag{2.26a}$$

$$\lim_{x_i-x_k\to\pm\infty}\psi_j^{nr}(\zeta;x_i,x_k) = \int_{-\infty}^{\infty}\frac{d\lambda}{2\pi}\rho_{j\pm}^{nr}(\zeta,\lambda)e^{-i\lambda(x_i+x_k)}\quad, \tag{2.26b}$$

$$\psi_k^{nr}(\zeta;x_i\to\pm\infty,x_k) = \int_{-\infty}^{\infty}\frac{d\lambda}{2\pi}\rho_{k\pm}^{nr}(\zeta,\lambda)e^{i\lambda x_k}\quad, \tag{2.26c}$$

where in (2.26), n = 1,2,...,8 and r = i, j, or k. Note that as defined, for any

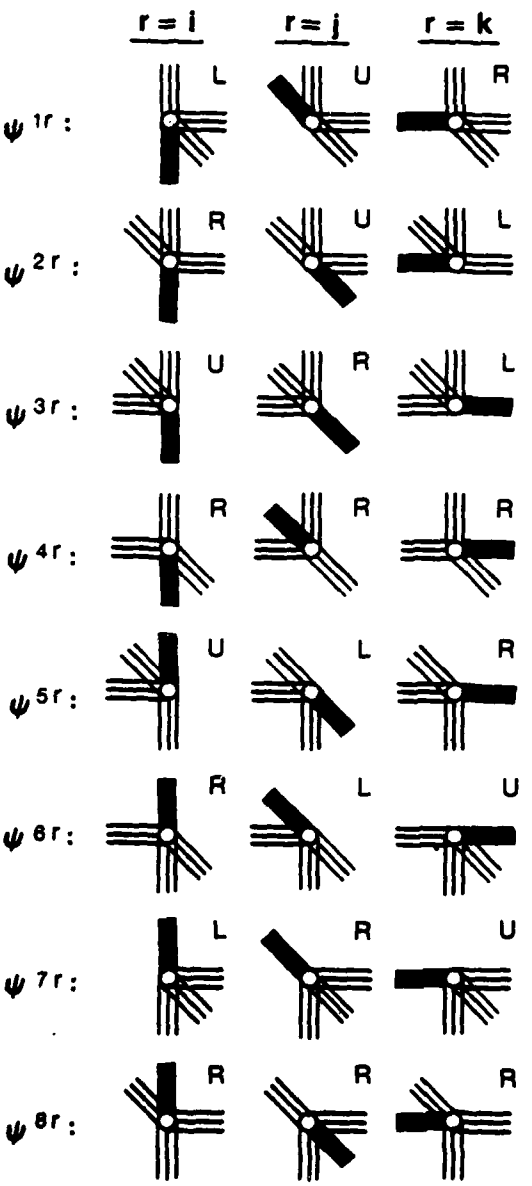


Fig. 2 - Graphic representations of all twenty-four fundamental solutions. The bold line indicates the direction of the incident wave, the circle, the region of the initial profiles. The sets of three parallel lines indicate the nonzero regions of any reflected wave or the shadow region of the transmitted wave. The symbol to the left designates the set of fundamental solutions, while the corresponding value or r is given across the top. The symbol in the upper right-hand corner of each graph indicates that the solution is analytic in the upper half plane (U), lower half plane (L), or is only bounded along the real axis (R).

fixed n and r, two of the ρ's will be zero, and one will be the identity, $I = 2\pi\delta(\zeta-\lambda)$. For example, from Fig. 2, we see that for ψ^{1i}, $\psi^{1i}_i = e^{i\zeta x_i}$ for $x_k \to -\infty$, $\psi^{1i}_j = 0$ as $x_j - x_k \to -\infty$, and $\psi^{1i}_k = 0$ as $x_i \to -\infty$. Thus we have $\rho^{1i}_{i-} = I$ and $\rho^{1i}_{j-} = 0 = \rho^{1i}_{k-}$.

From (2.6), one can derive a conservation law, which is

$$\partial_j(\gamma_k\psi_k\tilde{\psi}^*_k - \gamma_j\psi_j\tilde{\psi}^*_j) - \partial_k(\gamma_i\psi_i\tilde{\psi}^*_i - \gamma_j\psi_j\tilde{\psi}^*_j) = 0. \tag{2.27}$$

where ψ and $\tilde{\psi}$ is any two solutions of (2.6). If we integrate (2.27) over the entire x_jx_k-plane, using (2.26) to determine the contributions from the boundaries (at infinity), one obtains

$$\gamma_k(\rho^{mr}_{k+}\rho^{ns\dagger}_{k+} - \rho^{mr}_{k-}\rho^{ns\dagger}_{k-}) - \gamma_j(\rho^{mr}_{j+}\rho^{ns\dagger}_{j+} - \rho^{mr}_{j-}\rho^{ns\dagger}_{j-})$$

$$- \gamma_i(\rho^{mr}_{i+}\rho^{ns\dagger}_{i+} - \rho^{mr}_{i-}\rho^{ns\dagger}_{i-}) = 0 \quad . \tag{2.28}$$

We have introduced a convenient condensed operator notation in (2.28), which we shall use frequently hereafter. In this notation, a product such as $\sigma\rho$ simply means $\int_{-\infty}^{\infty} \frac{d\lambda}{2\pi}\sigma(\zeta,\lambda)\rho(\lambda,\zeta')$, $\rho^\dagger(\zeta,\zeta') = \rho^*(\zeta',\zeta)$, $I = 2\pi\delta(\zeta-\zeta')$, etc.

One consequence of (2.28) is that any triplet of fundamental solutions shown in Fig. 2 is a complete set of functions, by which any solution of (2.6) may be expanded. As an example, we will expand ψ^{6r} in terms of ψ^{2r}. In general we would have (in the condensed notation)

$$\psi^{6r} = a\psi^{2i} + b\psi^{2j} + c\psi^{2k} \quad . \tag{2.29}$$

To determine the matrices a, b, and c, we proceed as follows, referring to Fig. 2. As $x_k \to -\infty$ with x_i fixed, only $\psi^{2k}_k = e^{i\zeta x_i}$ is nonzero, while $\psi^{6r}_k = \rho^{6r}_{k-} \cdot e^{i\lambda x_i}$. Thus we have $c = \rho^{6r}_{k-}$. Similarly considering $x_i \to -\infty$ and $x_j - x_k \to +\infty$ gives

$$\psi^{6r} = \rho^{6r}_{i-}\psi^{2i} + \rho^{6r}_{j+}\psi^{2j} + \rho^{6r}_{k-}\psi^{2k} \quad . \tag{2.30a}$$

Likewise, we expand ψ^{2r} in terms of ψ^{6r}, obtaining

$$\psi^{2r} = \rho^{2r}_{i+}\psi^{6i} + \rho^{2r}_{j-}\psi^{6j} + \rho^{2r}_{k+}\psi^{6k} \quad . \tag{2.30b}$$

In order for these states to be complete, we must be able to insert (2.30b) into (2.30a), and obtain $\psi^{6r} = \sum_s(\delta^r_s I)\psi^{6s}$. This does indeed occur due to (2.28) and these states are therefore complete. The proof is simple straightforward algebra except for one point. And that is that (2.28) for n = 1,2,3, or 4 and for m = n + 4, gives that $\rho^{n+4,r\dagger}_{s+}$ is related to either ρ^{ns}_{r+} or ρ^{ns}_{r-}. Thus to prove the above, one needs the n = 2 and m = 6 forms of (2.28), as well as the n = 6 = m forms.

Lastly, we need to point out that the asymptotic form of ψ^{nr} as $|\zeta| \to \infty$ is exactly as one would expect, when ζ is constrained to the region where ψ^{nr} is bounded, as indicated by the letters in Figure 2. For example, directly from (2.8) and (2.10), one can verify that

$$\psi_r^{6j} e^{i\zeta(x_j+x_k)} \to \delta_r^j + o(1) \quad , \tag{2.31}$$

as $|\zeta| \to \infty$ in the LHP. Furthermore, one finds this result to be independent of n, and that in general

$$\psi_r^{nj} e^{-i\zeta x_j} \to \delta_r^j + o(1) \quad , \tag{2.32a}$$

$$\psi_r^{nj} e^{i\zeta(x_j+x_k)} \to \delta_r^j + o(1) \quad , \tag{2.32b}$$

$$\psi_r^{nk} e^{-i\zeta x_k} \to \delta_r^k + o(1) \quad , \tag{2.32c}$$

as $|\zeta| \to \infty$ in the appropriate half-plane or on the real axis.

II. THE INVERSE SCATTERING PROBLEM

As emphasized earlier, there is an almost infinite number of sets of inverse scattering equations, most of them possibly inequivalent, in that one cannot (at least easily) transform one set into another. This large number was noted first by Cornille.[1] The manner in which one set differs from another is in the set of reflection coefficients chosen to be used. As an analogy, in one dimensional scattering one may invert about either the right or the left, with each choice giving an inequivalent inverse scattering equation, since each equation uses a different reflection coefficient. The situation here is analogous, except that we now have three dimensions and a greatly expanded number of possible choices.

A general formula for deriving inverse scattering equations, and the one that I have almost always used is the following. First, one analyzes the analytical properties of the fundamental solutions of the scattering equations, define reflection coefficients, and locate all identities that they satisfy. This is the direct scattering problem and was done here in Section II for the 3D3WRI. Second, one determines what integral dispersion relations that a chosen independent set of these fundamental solutions satisfy. Once one has this, one actually at this point has a solution to the inverse scattering problem, because from these equations, one could reconstruct the chosen set of fundamental solutions and from the asymptotic form of these solutions as $|\zeta| \to \infty$, one can recover the potentials. However, existence proofs and other things are easier done using integral equations, so we then usually transform these dispersion relations into integral equations. These are actually equivalent. One is just the linear Fourier transform of the other.

So, we then have to do two additional steps. The third step is to find a set of "transformation kernals" for the fundamental solutions. Usually, consideration of the analytical properties of the fundamental solutions allows one to easily guess a proper form, and a proof of their existence directly follows. Then finally, the fourth step consists of simply of substituting the expressions for the fundamental solutions, in terms of the transformation kernals, into the linear dispersion relations, and upon taking a Fourier transform, one obtains the linear integral (inverse scattering) equations.

So, proceeding according to the above formula, we next select a set of independent fundamental solutions, and proceed to determine their linear dispersion relations. Arbitrarily, we shall select the set ψ^{6r} to work with. Even though we have selected a set to work with, here one still needs a little ingenuity to find the proper expressions, due to the large number of states present. First, let's note that to derive these dispersion relations, one basically is only using Cauchy's theorem. Thus one must be able to extend (or at least a part of each) these

fundamental solutions into the UHP and the LHP. This is best illustrated by example. ψ^{6i} exists in general only for real ζ. We note that ψ^{5i} and ψ^{3i} are both analytic in the UHP. But ψ^{5i} has the incident beam (the bold line in Fig. 2) orientated the same as ψ^{6i}. Thus if we expand ψ^{5i} in terms of the set ψ^{6r}, we find

$$\psi^{6i} = \psi^{5i} - \rho^{5i}_{j-}\psi^{6j} \quad . \tag{3.1a}$$

We want this form because ψ^{5i} can be extended into the UHP. For the LHP, we expand ψ^{7i} in terms of the set ψ^{6r}, obtaining

$$\psi^{6i} = \psi^{7i} - \rho^{7i}_{k+}\psi^{6k} \quad . \tag{3.1b}$$

For the ψ^{6j} solution, since it is already analytic in the LHP, we only need a form for extending a part of it into the UHP. Inspection of Fig. 2 shows that we want to use ψ^{1j}, which gives

$$\psi^{6j} = \psi^{1j} - \rho^{1j}_{k+}\psi^{6k} - \rho^{1j}_{i+}\psi^{6i} \quad . \tag{3.1c}$$

And to get ψ^{6k} into the LHP, we will use

$$\psi^{6k} = \psi^{3k} - \rho^{3k}_{i+}\psi^{6i} - \rho^{3k}_{j-}\psi^{6j} \quad . \tag{3.1d}$$

Now, simply applying Cauchy's theorem to each fundamental solution, using the asymptotic forms given in (2.32), we obtain the dispersion relations

$$\psi^{6i}_r e^{-i\zeta x_i} = \delta^i_r - G(x_i)\rho^{5i}_{j-}\psi^{6j}_r + \bar{G}(x_i)\rho^{7i}_{k+}\psi^{6k}_r \quad , \tag{3.2a}$$

$$\psi^{6j}_r e^{i\zeta(x_i+x_k)} = \delta^j_r - G(-x_i-x_k)[\rho^{1j}_{k+}\psi^{6k}_r + \rho^{1j}_{i+}\psi^{6i}_r] \quad , \tag{3.2c}$$

where the singular operators $G(u)$ and $\bar{G}(u)$ are defined by (for $\varepsilon \to 0^+$)

$$G(u;\zeta,\zeta') = i\,\frac{e^{-i\zeta' u}}{\zeta'-\zeta-i\varepsilon} \quad , \tag{3.3a}$$

$$\bar{G}(u;\zeta,\zeta') = i\,\frac{e^{-i\zeta' u}}{\zeta'-\zeta+i\varepsilon} \quad . \tag{3.3b}$$

This completes the second step.

For the third step, we seek to find a set of transformation kernals. Consideration of the analytical properties of ψ^{6r} simply suggests

$$\psi^{6i}_r e^{-i\zeta x_i} = \delta^i_r + \int_0^\infty e^{i\zeta s} U^{6i}_r(s;x_i,x_k)ds + \int_0^\infty e^{-i\zeta s} L^{6i}_r(s;x_i,x_k)ds \quad , \tag{3.4a}$$

$$\psi^{6j}_r e^{i\zeta(x_i+x_k)} = \delta^j_r + \int_0^\infty e^{-i\zeta s} L^{6j}_r(s;x_i,x_k)ds \quad , \tag{3.4b}$$

$$\psi^{6k}_r e^{-i\zeta x_k} = \delta^k_r + \int_0^\infty e^{i\zeta s} U^{6k}_r(s;x_i,x_k)ds \quad , \tag{3.4c}$$

where the U's and the L's are the transformation kernals. We must now prove that

the assumed U's and L's do exist. To do this, we require (3.4) to satisfy (2.6). This then gives us the series of equations

$$U_j^{6i}(s=\infty) = 0 = U_k^{6i}(s=\infty) \quad , \tag{3.5}$$

$$\partial_k U_i^{6i} = \gamma_k q_j^* U_k^{6i} - \gamma_j q_k U_j^{6i} \quad , \tag{3.6a}$$

$$(\partial_k - \partial_i + \partial_s) U_j^{6i} = \gamma_k q_i U_k^{6i} - \gamma_i q_k^* U_i^{6i} \quad , \tag{3.6b}$$

$$(\partial_i - \partial_s) U_k^{6i} = \gamma_i q_j U_i^{6i} - \gamma_j q_i^* U_j^{6i} \quad , \tag{3.6c}$$

$$L_j^{6i}(s=0) - U_j^{6i}(s=0) = \gamma_i q_k^* \quad , \tag{3.7a}$$

$$L_k^{6i}(s=0) - U_k^{6i}(s=0) = \gamma_i q_j \quad , \tag{3.7b}$$

$$L_j^{6i}(s=\infty) = 0 = L_k^{6i}(s=\infty) \quad , \tag{3.8}$$

$$\partial_k L_i^{6i} = \gamma_k q_j^* L_k^{6i} - \gamma_j q_k L_j^{6i} \quad , \tag{3.9a}$$

$$(\partial_k - \partial_i - \partial_s) L_j^{6i} = \gamma_k q_i L_k^{6i} - \gamma_i q_k^* L_i^{6i} \quad , \tag{3.9b}$$

$$(\partial_i + \partial_s) L_k^{6i} = \gamma_i q_j L_i^{6i} - \gamma_j q_i^* L_j^{6i} \quad , \tag{3.9c}$$

$$L_i^{6j}(s=\infty) = 0 = L_k^{6j}(s=\infty) \quad , \tag{3.10}$$

$$L_i^{6j}(s=0) = \gamma_j q_k \quad , \tag{3.11a}$$

$$L_k^{6j}(s=0) = \gamma_j q_i^* \quad , \tag{3.11b}$$

$$(\partial_k - \partial_s) L_i^{6j} = \gamma_k q_j^* L_k^{6j} - \gamma_j q_k L_j^{6j} \quad , \tag{3.12a}$$

$$(\partial_k - \partial_i) L_j^{6j} = \gamma_k q_i L_k^{6j} - \gamma_i q_k^* L_i^{6j} \quad , \tag{3.12b}$$

$$(\partial_i - \partial_s) L_k^{6j} = \gamma_i q_j L_i^{6j} - \gamma_j q_i^* L_j^{6j} \quad , \tag{3.12c}$$

$$U_i^{6k}(s=\infty) = 0 = U_j^{6k}(s=\infty) \quad , \tag{3.13}$$

$$U_i^{6k}(s=0) = -\gamma_k q_j^* \quad , \tag{3.14a}$$

$$U_j^{6k}(s=0) = -\gamma_k q_i \quad , \tag{3.14b}$$

$$(\partial_k - \partial_s) U_i^{6k} = \gamma_k q_j^* U_k^{6k} - \gamma_j q_k U_j^{6k} \quad , \tag{3.15a}$$

$$(\partial_k-\partial_i-\partial_s)U_j^{6k} = \gamma_k q_i U_k^{6k} - \gamma_i q_k^* U_i^{6k} \quad , \tag{3.15b}$$

$$\partial_i U_k^{6k} = \gamma_i q_j U_i^{6k} - \gamma_j q_i^* U_j^{6k} \quad . \tag{3.15c}$$

The above equations, (3.5) - (3.15) only insure that (3.4) satisfies (2.6). In order that (3.4) give *the solution* ψ^{6r}, and not another set, we must also require

$$U_i^{6r}(x_k=\infty) = 0 = L_i^{6r}(x_k=\infty) \quad , \tag{3.16a}$$

$$U_j^{6r}(x_k-x_i=\infty) = 0 = L_j^{6r}(x_k-x_i=\infty) \quad , \tag{3.16b}$$

$$U_k^{6r}(x_i=\infty) = 0 = L_k^{6r}(x_i=\infty) \quad . \tag{3.16c}$$

One may now turn (3.6), (3.9), (3.12), and (3.15) into integral equations, using (3.5), (3.7), (3.8), (3.10), (3.11), (3.13), (3.14), and (3.16) for the boundary conditions. This gives

$$U_i^{6i} = \int_0^\infty [\gamma_j q_k U_j^{6i} - \gamma_k q_j^* U_k^{6i}](s;x_i,x_k+v)dv \quad , \tag{3.17a}$$

$$U_j^{6i} = -\int_s^\infty [\gamma_k q_i U_k^{6i} - \gamma_i q_k^* U_i^{6i}](v;x_i-v+s,x_k+v-s)dv \quad , \tag{3.17b}$$

$$\begin{aligned} U_k^{6i} = {} & -\gamma_i q_j(x_i+s,x_k) - \int_0^\infty [\gamma_i q_j L_i^{6i} - \gamma_j q_i^* L_j^{6i}](v;x_i+s+v,x_k)dv \\ & - \int_0^s [\gamma_i q_j U_i^{6i} - \gamma_j q_i^* U_j^{6i}](v;x_i+s-v,x_k)dv \quad , \end{aligned} \tag{3.17c}$$

$$L_i^{6i} = \int_0^\infty [\gamma_j q_k L_j^{6i} - \gamma_k q_j^* L_k^{6i}](s;x_i,x_k+v)dv \quad , \tag{3.18a}$$

$$\begin{aligned} L_j^{6i} = {} & \gamma_i q_k^*(x_i-s,x_k+s) - \int_0^\infty [\gamma_k q_i U_k^{6i} - \gamma_i q_k^* U_i^{6i}](v;x_i-s-v,x_k+s+v)dv \\ & - \int_0^s [\gamma_k q_i L_k^{6i} - \gamma_i q_k^* L_i^{6i}](v;x_i-s+v,x_k+s-v)dv \quad , \end{aligned} \tag{3.18b}$$

$$L_k^{6i} = -\int_s^\infty [\gamma_i q_j L_i^{6i} - \gamma_j q_i^* L_j^{6i}](v;x_i+v-s,x_k)dv \quad , \tag{3.18c}$$

$$L_i^{6j} = \gamma_j q_k(x_i,x_k+s) - \int_0^s [\gamma_k q_j^* L_k^{6j} - \gamma_j q_k L_j^{6j}](v;x_i,x_k+s-v)dv \quad , \tag{3.19a}$$

$$L_j^{6j} = \int_0^\infty [\gamma_i q_k^* L_i^{6j} - \gamma_k q_i L_k^{6j}](s;x_i-v,x_k+v)dv, \tag{3.19b}$$

$$L_k^{6j} = \gamma_j q_i^*(x_i+s,x_k) + \int_0^s [\gamma_j q_i^* L_j^{6j} - \gamma_i q_j L_i^{6j}](v;x_i+s-v,x_k)dv \quad , \tag{3.19c}$$

$$U_i^{6k} = -\gamma_k q_j^*(x_i,x_k+s) + \int_0^s [\gamma_j q_k U_j^{6k} - \gamma_k q_j^* U_k^{6k}](v;x_i,x_k+s-v)dv \quad , \tag{3.20a}$$

$$U_j^{6k} = -\gamma_k q_i(x_i-s,x_k+s) + \int_0^s [\gamma_i q^*{}_k U_i^{6k} - \gamma_k q_i U_k^{6k}](v;\, x_i-s+v,x_k+s-v)dv \quad , \tag{3.20b}$$

$$U_k^{6k} = \int_0^\infty [\gamma_j q_i^* U_j^{6k} - \gamma_i q_j U_i^{6k}](s;x_i+v,x_k)dv \quad . \tag{3.20c}$$

Note that (3.17) and (3.18) is a coupled set of equations. By techniques almost identical to those used in Section II, one may show that the Neumann series solutions of the above exist, and therefore the kernals defined in (3.4) do exist.

We are now ready to obtain the inverse scattering equations. But first, we shall remark on the time dependence of the scattering data. From (2.1), (2.3), and (2.26), one obtains

$$\partial_t \rho^{nr}_{i\pm} = 0 = \partial_t \rho^{nr}_{k\pm} \quad , \tag{3.21a}$$

$$(\partial_t + i\lambda)\rho^{nr}_{j\pm}(\zeta,\lambda) = 0 \ . \tag{3.21b}$$

From (3.21b), one finds

$$\rho^{nr}_{j\pm}(t;\zeta,\lambda)e^{-i\lambda(x_i+x_k)} = \rho^{nr}_{j\pm}(0;\zeta,\lambda)e^{i\lambda x_j} \ . \tag{3.21c}$$

As a consequence of (3.21c), we may now replace the term on the left-hand side of (3.21c) wherever it occurs, with the right-hand side, provided we interpret all ρ's as those at $t = 0$. This will be a minor simplification in the following inverse scattering equations, but more important, it allows us to immediately write them in a form valid for all t. To obtain these equations, we simply substitute (3.4) into (3.2), and upon taking a Fourier transform, one obtains

$$U^{6i}_r(s) + \delta^k_r F^{7i}_{k+}(x_i+s,x_k) + \int_0^\infty F^{7i}_{k+}(x_i+s,x_k+v)U^{6k}_r(v)dv = 0 \quad , \tag{3.22a}$$

$$L^{6i}_r(s) + \delta^j_r F^{5i}_{j-}(x_i-s,x_j) + \int_0^\infty F^{5i}_{j-}(x_i-s,x_j-v)L^{6j}_r(v)dv = 0 \quad , \tag{3.22b}$$

$$\begin{aligned} L^{6j}_r(s) &+ \delta^k_r F^{1j}_{k+}(x_j-s,x_k) + \delta^i_r F^{1j}_{i+}(x_j-s,x_i) \\ &+ \int_0^\infty dv \left[\begin{array}{l} F^{1j}_{k+}(x_j-s,x_k+v)U^{6k}_r(v) \\ \quad + F^{1j}_{i+}(x_j-s,x_i+v)U^{6i}_r(v) \\ \quad\quad + F^{1j}_{i+}(x_j-s,x_i-v)L^{6i}_r(v) \end{array} \right] = 0 \ , \end{aligned} \tag{3.22c}$$

$$\begin{aligned} U^{6k}_r(s) &+ \delta^j_r F^{3k}_{i+}(x_k+s,x_i) + \delta^j_r F^{3k}_{j-}(x_k+s,x_j) \\ &+ \int_0^\infty dv \left[\begin{array}{l} F^{3k}_{i+}(x_k+s,x_i+v)U^{6i}_r(v) \\ \quad + F^{3k}_{i+}(x_k+s,x_i-v)L^{6i}_r(v) \\ \quad\quad + F^{3k}_{j-}(x_k+s,x_j-v)L^{6j}_r(v) \end{array} \right] = 0 \ , \end{aligned} \tag{3.22d}$$

where

$$F^{nr}_{p\pm}(u,v) = \int_{-\infty}^{\infty} \frac{d\zeta}{2\pi} \int_{-\infty}^{\infty} \frac{d\lambda}{2\pi} e^{-i\zeta u} \rho^{nr}_{p\pm}(0;\zeta,\lambda)e^{i\lambda v} \ . \tag{3.23}$$

In (3.22), we have six different kernals (the F's); however, only three are independent. From the $m = 7$ and $n = 3$ form of (2.28), one finds

$$\rho_{k+}^{7i} = \gamma_i\gamma_k\rho_{i+}^{3k+} \quad , \tag{3.24a}$$

from the m = 5 and n = 1 form, one finds

$$\rho_{j-}^{5i} = \gamma_i\gamma_j\rho_{i+}^{1j+} \quad , \tag{3.24b}$$

and from the m = 1 and n = 3 form we have

$$\rho_{i+}^{1j}\rho_{i+}^{3k+} - \gamma_i\gamma_j\rho_{j-}^{3k+} - \gamma_i\gamma_k\rho_{k+}^{1j} = 0 \quad . \tag{3.24c}$$

Thus from (3.23) and (3.24) we have

$$F_{k+}^{7i}(u,v) = \gamma_i\gamma_k F_{i+}^{3k*}(v,u) \quad , \tag{3.25a}$$

$$F_{j-}^{5i}(u,v) = \gamma_i\gamma_j F_{i+}^{1j*}(v,u) \quad , \tag{3.25b}$$

$$F_{j-}^{3k}(u,v) = -\gamma_j\gamma_k F_{k+}^{1j*}(v,u) + \gamma_i\gamma_j \int_{-\infty}^{\infty} ds F_{i+}^{3k}(u,s)F_{i+}^{1j*}(v,s) \quad , \tag{3.25c}$$

leaving only F_{i+}^{3k}, F_{k+}^{1j}, and F_{i+}^{1j} as independent.

A proof of the existence and uniqueness for the solution (3.22) is fairly easy to do, except in the explosive case ($\gamma_i = \gamma_k = \gamma_j = \pm 1$), which shall not handle here. For a proof, we need only to show that (3.22) has no homogeneous solutions. Using the identities in (3.25), it is fairly easy to show that the homogeneous solution of (3.22) must satisfy

$$\int_0^{\infty} ds \begin{bmatrix} \gamma_i|U_r^{6i}(s)|^2 + \gamma_i|L_r^{6i}(s)|^2 \\ -\gamma_k|U_r^{6k}(s)|^2 - \gamma_j|L_r^{6j}(s)|^2 \end{bmatrix} = 0 \quad . \tag{3.26}$$

Now, only in the explosive case will all the γ's be the same sign. In this case (3.26) tells us nothing important. The only other case is when one of the γ's differs in sign from the other two. Let this γ be designated as γ_i. Then (2.26) shows that only the trivial zero solution exists. Therefore for $\gamma_i = -\gamma_j = -\gamma_k$, the solution of (3.22) will be *unique*. (This is no real restriction, since we may initially designate any one of the three waves to be the i^{th} wave.) The reader should note that Niznik's statement in Ref. 11 that these systems are always solvable is valid in general only for the nonexplosive cases, unless further restrictions are placed on the scattering data. This is because we have a counterexample, in that for the explosive case, the profiles can initially satisfy (2.14), but can become singular as it evolves in time.[8,9]

This concludes this solution of the general initial value problem for the 3D3WRI. From the arbitrary initial data at t = 0, satisfying (2.14), one solves (2.6) for the ρ's defined in (2.26), of which one really only needs ρ_{i+}^{3k}, ρ_{k+}^{1j}, and ρ_{i+}^{1j} for this solution. Then one constructs the F's as in (3.23), solves the linear integral equations, (3.22), with the potentials being recovered from Eqs. (3.7), (3.11), and (3.14).

We remark that the F's may be obtained directly from (2.6). For example, define

$$\psi^{1j}(u;x_1,x_k) = \int_{-\infty}^{\infty} \frac{d\zeta}{2\pi} e^{-i\zeta u} \psi^{1j}(\zeta;x_1,x_k) \quad , \tag{3.27}$$

which is a solution of (2.6) satisfying the initial condition

$$\lim_{L\to\infty} \Psi_r^{lj}(u;\tfrac{1}{2}v-L,\tfrac{1}{2}v+L) = \delta_r^j\delta(u-v) \quad . \tag{3.28}$$

Then from (2.26), we have

$$F_{i+}^{lj}(u,v) = \Psi_i^{lj}(u;v,x_k \to +\infty) \quad , \tag{3.29a}$$

$$F_{k+}^{lj}(u,v) = \Psi_k^{lj}(u;x_i \to +\infty,v) \quad . \tag{3.29b}$$

Similarly, one could also obtain $F_{i+}^{3k}(u,v)$.

IV. CONCLUDING REMARKS

In an earlier work,[6] we had obtained the inverse scattering solution for separable initial data assuming global properties on the solution. The inverse scattering equations presented here, (3.22), are not equivalent to those in Ref. 6 [Eq. (45)]. The truth of this statement shall become clear shortly. First, we want to relate the fundamental solutions used here with those of Ref. 6. We do this by expanding the ψ^n fundamental solutions of Ref. 6 [defined by Eq. (7)], in terms of ψ^{6r}. Thus we have

$$\psi^n = a_i^n\psi^{6i} + a_j^n\psi^{6j} + a_k^n\psi^{6k} \quad . \tag{4.1}$$

To determine the coefficients, we proceed as follows. From Fig. 2, we have at $t = 0$, as $x_k \to +\infty$, x_i fixed [and therefore by (2.4), $x_j \to -\infty$], that only ψ^{6i} is nonzero. Thus in this limit, if we consider the i^{th} component of (4.1) we find

$$\psi_i^n(\zeta;x_i,-,+) = \int_{-\infty}^{\infty} \frac{d\lambda}{2\pi}\, a_i^n(\zeta,\lambda)e^{i\lambda x_i} \quad . \tag{4.2}$$

From Eqs. (25) and (A5) of Ref. 6, we can evaluate the left-hand side of (4.2), obtaining

$$a_i^n = \mu_{ki}^n \quad , \tag{4.3a}$$

where the μ's were defined in Ref. 6. Similarly, one also finds

$$a_j^n = \mu_{kj}^n \quad , \tag{4.3b}$$

$$a_k^n = \mu_{ik}^n \quad , \tag{4.3c}$$

giving

$$\psi^n = \mu_{ki}^n\psi^{6i} + \mu_{kj}^n\psi^{6j} + \mu_{ik}^n\psi^{6k} \quad . \tag{4.4}$$

Of course, the μ's are unknown when we do the general initial value problem, so we now seek to see how well we may determine them. To do this, we consider (4.4) in the limits opposite to those used to obtain (4.3). Specifically, we consider $\psi_i^n(x_i,+,-)$, $\psi_j^n(x_j,-,+)$, and $\psi_k^n(x_k,-,+)$. Then we have

$$\mu_{ji}^n = \mu_{ki}^n\rho_{i-}^{6i} + \mu_{kj}^n\rho_{i-}^{6j} + \mu_{ik}^n\rho_{i-}^{6k} \quad , \tag{4.5a}$$

$$\mu^{n}_{ij} = \mu^{n}_{ki}\rho^{6i}_{j+} + \mu^{n}_{kj}\rho^{6j}_{j+} + \mu^{n}_{ik}\rho^{6k}_{j+} \quad , \tag{4.5b}$$

$$\mu^{n}_{jk} = \mu^{n}_{ki}\rho^{6i}_{k-} + \mu^{n}_{kj}\rho^{6j}_{k-} + \mu^{n}_{ik}\rho^{6k}_{k-} \quad . \tag{4.5c}$$

Using the identities satisfied by the μ's, Eqs. (A9) - (A15) of Ref. 6, we may solve for the ρ's in terms of the μ's and ν's. This gives

$$\rho^{6i}_{i-} = \mu^{i\dagger}_{ki}(I - \nu^{i}_{kj}\nu^{j}_{ik}\nu^{k}_{ji})\mu^{i}_{ji} \quad , \tag{4.6a}$$

$$\rho^{6i}_{j+} = \mu^{i\dagger}_{ki}\nu^{i}_{kj}\nu^{j}_{ij} \quad , \tag{4.6b}$$

$$\rho^{6i}_{k-} = -\gamma_i\gamma_k\tilde{\nu}^{k\dagger}_{ji}\nu^{k}_{ik}\mu^{k}_{jk} \quad , \tag{4.6c}$$

$$\rho^{6j}_{i-} = \gamma_i\gamma_j\mu^{j\dagger}_{kj}\nu^{j}_{ij}\tilde{\nu}^{i\dagger}_{kj} \quad , \tag{4.6d}$$

$$\rho^{6j}_{j+} = \mu^{j\dagger}_{kj}\nu^{j}_{ij} \quad , \tag{4.6e}$$

$$\rho^{6j}_{k-} = \tilde{\nu}^{j}_{ik} \quad , \tag{4.6f}$$

$$\rho^{6k}_{i-} = -\nu^{k}_{ik}\nu^{k}_{ji}\mu^{i}_{ji} \quad , \tag{4.6g}$$

$$\rho^{6k}_{j+} = \nu^{k}_{ij} \quad , \tag{4.6h}$$

$$\rho^{6k}_{k-} = \nu^{k}_{ik}\mu^{k}_{jk} \quad . \tag{4.6i}$$

As was the convention in Ref. 6 a wavy line over any quantity indicates that it is that quantity for the final profiles ($t \to +\infty$), while those without wavy lines over them are those for the distant past ("asymptotic initial," $t \to -\infty$) profiles. Note that (4.6f) and (4.6h) gives us directly a ν and a $\tilde{\nu}$. If we now make use of some of the many identities in (2.28), from (4.6) one can then show that

$$\rho^{1i}_{j+} = \tilde{\nu}^{i}_{kj} \quad , \tag{4.7a}$$

$$\rho^{3k}_{i+} = \tilde{\nu}^{k}_{ij} \quad , \tag{4.7b}$$

$$\rho^{6j}_{k-} = \tilde{\nu}^{j}_{ik} \quad , \tag{4.7c}$$

and

$$\rho^{1j}_{i+} = \nu^{j}_{ki} \quad , \tag{4.8a}$$

$$\rho^{3i}_{k-} = \nu^{i}_{jk} \quad , \tag{4.8b}$$

$$\rho^{6k}_{j+} = \nu^{k}_{ij} \quad , \tag{4.8c}$$

Eq. (4.7) contains exactly those $\tilde{\nu}$'s required for constructing the $\tilde{E}$'s [Eq. (A23)

of Ref. 6] needed for constructing the final profiles (Eqs. (A22) and (A24) of Ref. 6). Therefore, we now have that given any general initial value problem, by evaluating the ρ's listed in (4.7), we may in general go directly to the final profiles by using the results in Ref. 6. We say in general simply because we are not necessarily always insured that the ρ's in (4.7) will satisfy those conditions sufficient for (A22) of Ref. 6 to possess a solution. (Those sufficient conditions depended on certain global properties.) One may likewise do the same with (4.8) in determining what the "asymptotic initial" envelopes were, except that we cannot use (A22) of Ref. 6 directly. This is because the permutation of i, j, and k on the ν's in (4.8) is in the wrong order for (A22) of Ref. 6. However, by finding the linear dispersion relations for the N solutions (defined in Appendix A of Ref. 6) as in (A19) of Ref. 6, one may proceed to obtain inverse scattering equations which would involve the ν's of these permutations.

Let's now determine the ρ's required by (3.22) in terms of the μ's and ν's. Again, from the identities in (2.28) and with (4.6), one finds

$$\rho_{i+}^{1j} = \nu_{ki}^{j} \quad , \tag{4.9a}$$

$$\rho_{k+}^{1j} = \gamma_j \gamma_k \nu_{kj}^{j} \mu_{ij}^{j} \nu_{ij}^{k\dagger} \quad , \tag{4.9b}$$

$$\rho_{i+}^{3k} = \nu_{ji}^{\sim k} \quad , \tag{4.9c}$$

$$\rho_{j-}^{3k} = \gamma_j \gamma_k \mu_{ik}^{k\dagger} \nu_{jk}^{k} \nu_{ik}^{\sim j\dagger} \quad , \tag{4.9d}$$

$$\rho_{j-}^{5i} = \gamma_i \gamma_j \nu_{ki}^{j\dagger} \quad , \tag{4.9e}$$

$$\rho_{k+}^{7i} = \gamma_i \gamma_k \nu_{ji}^{\sim k\dagger} \quad . \tag{4.9f}$$

In (4.9), we simply wish to point out that the scattering data required for (3.22), when viewed from the point of view of Ref. 6, are a "mixture." They contain one reflection coefficient for the asymptotic initial k^{th} envelope, ν_{ki}^{j}, one reflection coefficient for the final j^{th} envelope, $\nu_{ji}^{\sim k}$, and a conbination of the transmission coefficients for the initial i^{th} and k^{th} envelopes with the reflection coefficient of the i^{th} initial envelope. (ρ_{j-}^{3k} is not independent.)

One result of using this type of an admixture of the scattering data for inverse scattering, is that the separable solutions of (3.22) will not be the same as the separable solutions of (A22) in Ref. 6. [By separable solution, we mean the solution that results when the kernals of the inverse scattering equations are separable.] The separable solutions of (A22) of Ref. 6 are briefly discussed in Appendix B of Ref. 6. We simply note that assuming the F's to separate is equivalent to assuming that the reflection coefficients also separate. Thus the separable solutions of Ref. 6 require ν_{ik}^{j}, ν_{ji}^{k}, and ν_{kj}^{i} to separate. But as one may verify by reconstructing the fundamental solutions from this separable scattering data, the transmission coefficients and the other reflection coefficients do not in general separate. Thus the ρ's listed in (4.9) would not in general separate although ν_{ik}^{j} etc. did. Likewise, if we require the ρ's in (4.9) to separate, we can obtain separable solutions obtained from (3.22) then ν_{ik}^{j} etc. would not separate. Thus separable solutions obtained from (3.22) will not be equivalent to those separable solutions obtained from (A22) of Ref. 6. Likewise, any other inverse scattering equations which are based on a set of reflection coefficients different from the above two sets would have still another

inequivalent set of separable solutions.

What this suggests is that each different set of inverse scattering equations [Cornille calculated the number to be 3^{15} different sets] produces one subset of separable solutions from a very large class of closed form solutions of the 3D3WRI. How large this class of closed form solutions really is, remains to be seen.

REFERENCES

[1] Cornille, H., J. Math. Phys. 20, 1653 (1979).

[2] Zakharov, V.E. and Shabat, A.B., Funkts. Anal. Prilozh. 8, 43 (1974).

[3] Ablowitz, M.J. and Haberman, R., J. Math. Phys. 16, 2301 (1975).

[4] Kaup, D.J., "A Method for Solving the Separable Initial Value Problem of the Full Three Dimensional Three-Wave Interaction" (to appear in Stud. Appl. Math.).

[5] Kaup, D.J., "Determining the Final Profiles from the Initial Profiles for the Full Three Dimensional Three-Wave Resonant Interaction" (preprint).

[6] Kaup, D.J., "The Inverse Scattering Solution for the Full Three-Dimensional Three-Wave Resonant Interaction" (to appear in Physica D).

[7] Kaup, D.J., Reiman, A.H., and Bers, A., Rev. Mod. Phys. 51, 275 (1979).

[8] Zakharov, V.E., Dokl. Akad. Nauk. SSSR 228, 1314 (1976) [Sov. Phys. Dokl. 21, 322 (1976)].

[9] Craik, A.D.D., "Evolution in Space and Time of Resonant Wave Triads. Part II: A Class of Exact Solutions" (preprint).

[10] Niznik, L.P., The Nonstationary Scattering Problem, ["Naukova Dumka", Kiev U.S.S.R. (1973)], (in Russian).

[11] Niznik, L.P. and Tarasov, V.G., Dokl. Adad. Nauk SSSR 233, No. 3 (1977) [Soviet Math. Dokl. 18, 397 (1977)].

[12] Manakov, S.V., (to appear in Sov. Sci. Review).

The lump solutions and the Bäcklund transformation for the three-dimensional three-wave resonant interaction

D. J. Kaup
Department of Physics, Clarkson College of Technology, Potsdam, New York 13676

(Received 19 May 1980; accepted for publication 21 November 1980)

A Bäcklund transformation is found for the three-dimensional three-wave resonant interaction, and from it, N-lump exact solutions may be constructed. The one-lump solution is analyzed in detail, and it is shown that it describes such effects as pulse decay, upconversion, and explosive instabilities, all in three dimensions.

PACS numbers: 02.30.Jr, 03.40.Kf

I. INTRODUCTION

Since the general initial value problem of the three-dimensional three-wave resonant interaction (3D3WRI) has recently been completed,[1] one may now turn his attention to see how one may make use of this method of solution. Obviously, one would seek first to exploit the simplest classes of solutions. These are called "lump" solutions, the first ones of which were found by Zakharov[2] and later by Craik.[3] They were then rediscovered by Cornille[4] from his integral equations, and were also noted to arise from the many and varied inverse scattering transforms associated with the 3D3WRI.[1,5] Following a suggestion by Professor Corones,[6] it was soon discovered that the complicated machinery required for the general solution of this problem was not needed if one was willing to only generate new solutions. Of course, to solve an initial value problem you still needed this machinery. But if one would only seek to investigate special solutions, then an old observation made by Professor Corones,[6] would allow one to almost immediately construct the Bäcklund transformation for these equations, from which these special lump solutions could be quickly generated.

These lump solutions in three dimensions are *not* analogous to the soliton solutions from one-dimensional inverse scattering theories, although they are similar in some respects. First of all, when the envelope profiles are required to be integrable, and square integrable in three dimensions, the scattering data for the 3D3WRI has been shown to consist of only a continous spectrum, with no bound states allowed.[1,5,7] On the other hand, soliton solutions in one-dimensional inverse scattering theories always correspond to a pure bound state spectrum, with the continous spectrum being distinctly absent. Furthermore, solitons always have a relation between amplitude and width, so that only one is independent, and with a distinct shape or profile. But lump solutions have no such relation between amplitude and widths, and the profile shape can be quite arbitrary. Thus in three dimensions, we have much more freedom in shaping and forming lump solutions than was possible with one-dimensional soliton solutions.

However, there are also similarities. The one-dimensional soliton solutions always have a closed form solution, since the kernels in the one-dimensional inverse scattering equations were separable. The same is true for these lump solutions in three dimensions, in that the kernels are also separable, again allowing closed form solutions. In fact, this is how one can define a lump solution. And as one could construct N-soliton solutions, so can one also construct N-lump solutions. Furthermore, N-soliton solutions can be constructed from a Bäcklund transformation, and as we shall show here, the same can be done for N-lump solutions.

These N-lump solutions are important for another reason too. Nižnik has shown that any set of potentials (envelope profiles) can be approximated to any given accuracy by a suitably chosen N-lump solution.[8] Thus, many physical situations can therefore be reasonably well approximated by these N-lump solutions, at least for a finite time. The latter condition is necessary, because Nižnik's proof is only valid for a given instant of time, and in general one would not expect such an approximation to remain valid for all time.

What we shall do in this paper is to derive the Bäcklund transformation for the 3D3WRI, showing how one may construct an N-lump solution, and shall briefly discuss these solutions. Then we shall use the one-lump solution to analyze the collision of envelopes in the positive and negative energy cases. What we shall find here are solutions illustrating pulse decay, upconversion, and explosive singularities. These effects in three dimensions are shown to be both analogous and different from the corresponding effects in one-dimensional solutions.[9]

II. THE BÄCKLUND TRANSFORMATION

As discussed by Corones,[6] one way to find a Bäcklund transformation is to simply assume that one exists of the form

$$q' = f(q, Q, \lambda), \tag{2.1}$$

where q is a known solution of the nonlinear system, q' is to be a new solution, Q is the pseudopotential (s), and λ is the eigenvalue (if one is present). One now simply requires q' to be a solution of the nonlinear system, which then can determine what the functional form for f must be.

Well, the same basic idea also works for the 3D3WRI, as we shall now illustrate. The nonlinear equations are

$$\frac{\partial q_i}{\partial \chi_i} = \gamma_i q_j^* q_k^*, \tag{2.2}$$

where i, j, k are cyclic and equal to (1,2,3), q_i is the i^{th} envelope, χ_i is the i^{th} characteristic coordinate, defined by

$$\frac{\partial}{\partial \chi_i} = -\partial_t - \mathbf{v}_i \cdot \nabla, \tag{2.3a}$$

 0022-2488/81/061176-06$1.00 © 1981 American Institute of Physics

where $\mathbf{v}_i$ is the group velocity for the i^{th} envelope. The γ_i in (2.2) carries the sign of the coupling constants, and we assume the γ's to be real and scaled to unity in magnitude.

We use the characteristic coordinates because of the resulting simplifications in the analysis. From (2.3) one can obtain the transformation from characteristic coordinates to space–time coordinates. Of course, by (2.3a), we only have three characteristic coordinates, and to complete the transformation, we must supplement (2.3a) with

$$(\partial_t + \mathbf{v}_i\cdot\nabla)\chi_4 = 0, \tag{2.3b}$$

which defines a fourth coordinate. A solution of (2.3) for $\mathbf{x}$ and t as functions of χ_i and χ_4 is

$$x = -\sum_{i=1}^{3}\mathbf{v}_i\chi_i + \mathbf{n}\chi_4, \tag{2.4a}$$

$$t = -\sum_{i=1}^{3}\chi_i, \tag{2.4b}$$

where $\mathbf{n}$ is a unit vector parallel to $(\mathbf{v}_1 - \mathbf{v}_2)\times(\mathbf{v}_2 - \mathbf{v}_3)$, and as such, makes an equal angle with all three group velocities (i.e., $-\mathbf{v}_1\cdot\mathbf{n} = \mathbf{v}_2\cdot\mathbf{n} = \mathbf{v}_3\cdot\mathbf{n}$). Note that when we keep any two characteristic coordinates fixed, and let $t\to\pm\infty$, then the free characteristic coordinate approaches $\mp\infty$.

As far as (2.2) is concerned χ_4 is like a dummy coordinate, since no differentials with respect to it occur. Thus, in solving (2.2), we may keep χ_4 fixed, then let χ_4 take on another value, solve (2.2) again, etc. Thus, since χ_4 is a dummy coordinate, we shall ignore it from now on.

The scattering problem for (2.2) is

$$\partial_k\psi_i = \gamma_k q_j^*\psi_k, \tag{2.5a}$$

$$\partial_i\psi_k = \gamma_i q_j\psi_i, \tag{2.5b}$$

where in (2.5), i, j, k are still cyclic, giving six equations altogether. Equation (2.5) has three conservation laws, which are

$$\partial_i(\gamma_k\psi_k^*\psi_k) = \partial_k(\gamma_i\psi_i^*\psi_i), \tag{2.6}$$

and a consequence of these, is that a function D exists such that

$$\partial_i D = -\gamma_i\psi_i^*\psi_i. \tag{2.7}$$

To find a Bäcklund transformation, we now assume that we can find three functions, $\tilde{q}_1, \tilde{q}_2$, and $\tilde{q}_3$, where

$$\tilde{q}_i = \tilde{q}_i(\psi^*, \psi, q, D), \tag{2.8}$$

and such that these functions will then satisfy (2.2). Using (2.2), (2.5), and (2.7), it is rather simple to show that such can be done, and that

$$\tilde{q}_j = q_j + \psi_i^*\psi_k/D \tag{2.9}$$

is one such transformation, provided D is real. With (2.9), one may now generate new solutions of (2.2) from old solutions. Given any solution of (2.2), one constructs ψ_i from (2.5) and D from (2.7). Then these functions inserted into (2.9) will give a new solution.

In particular, we may start with $q_i = 0$, in which case we obtain a one-lump solution. The solution of (2.5) and (2.7) is then

$$\psi_i = g_i(\chi_i), \tag{2.10a}$$

$$D = \delta + \sum_{i=1}^{3}\gamma_i G_i(\chi_i), \tag{2.10b}$$

where δ is some real constant, and

$$G_i(u) = \int_u^\infty g_i^* g_i(u)\,du. \tag{2.11}$$

In (2.10), we may, without loss of generality, assume $\delta = \pm 1$, since (2.5) is linear in ψ, and only the ratio of $\psi_i^*\psi_k/D$ is significant. Furthermore, if we note that (2.2) and (2.5) are invariant under the change of the signs of all the q's and γ's, we may also choose $\delta = +1$, which we shall do.

Then from (2.9) and (2.10), we have the one-lump solution as being

$$q_j = g_i^* g_k/D, \tag{2.12a}$$

where

$$D = 1 + \sum_{i=1}^{3}\gamma_i G_i. \tag{2.12b}$$

Note that in (2.12), g_i is only a function of χ_i, and similarly for G_i.

Given the above solution, one may proceed to obtain a two-lump solution. Denoting the one-lump D in (2.12) by D_g, then the general solution of (2.5) is now

$$\psi_i = h_i(\chi_i) + g_i F/D_g, \tag{2.13}$$

where

$$F = \eta - \sum_{i=1}^{3}\gamma_i\int_{\chi_i}^\infty g_i^* H_i(u)\,du, \tag{2.14}$$

with η being some constant. The solution of (2.7) is then

$$D = D_h - F - F/D_g, \tag{2.15}$$

where

$$D_h = 1 + \sum_{i=1}^{3}\gamma_i H_i(\chi_i), \tag{2.16a}$$

$$H_i(u) = \int_u^\infty h_i^* h_i\,du. \tag{2.16b}$$

This gives the two-lump solution as being

$$q_j = \bar{q}_j/(D_g D_h - F^*F), \tag{2.17a}$$

where

$$\bar{q}_j = D_h g_i^* g_k + h_i^* g_k F + h_k g_i^* F^* + D_h h_i^* h_k. \tag{2.17b}$$

Although we do not now have a general proof, it seems fairly obvious that by this procedure one may generate an arbitrary N-lump solution.

Let us now consider the simple one-lump solution given by (2.11). The simplest form is when one of the q's is zero. If we choose $g_1 = 0$, then

$$q_1 = g_2^*(\chi_2)g_3(\chi_3)/D(\chi_2,\chi_3), \tag{2.18a}$$

$$q_2 = q_3 = 0, \tag{2.18b}$$

which is a trivial solution of (2.2). It has only one nonzero envelope, and thus corresponds to a freely propagating envelope without any interactions. And in the absence of any interactions, it travels along its characteristic, χ_1 without any change (i.e., it is independent of χ_1). It's profile and shape is determined by two independent functions g_2 and g_3, one for each free dimension. (Actually, g_2 and g_3 will also depend on χ_4, but as stated before, we shall ignore this dependence since it can be understood that all functions and parameters

can always depend on χ_4.) Note that these functions are completely arbitrary; thus there is a large amount of freedom in shaping this pulse. However, the shape is not totally arbitrary, because the functional dependence of q_1 on χ_2 and χ_3 is required to be of the form of a simple product, except for the terms in the denominator.

When all the g's are nonzero in general, we have nontrivial interactions and solutions with the general solution being given by (2.12). To obtain the asymptotic initial profiles (the profiles as $t\to-\infty$) according to (2.4b), we take the limit of the corresponding characteristic coordinate going to $+\infty$. Thus, to get the initial profile for g_1, which we shall designate by Q_1, we let $\chi_1\to+\infty$ in (2.12) for $(i,j,k)=(3,1,2)$ giving,

$$Q_1=g_3^*(\chi_3)g_2(\chi_2)/[1+\gamma_2G_2(\chi_2)+\gamma_3G_3(\chi_3)]. \quad (2.19a)$$

Similarly, the other initial profiles are

$$Q_2=g_1^*(\chi_1)g_3(\chi_3)/[1+\gamma_1G_1(\chi_1)+\gamma_3G_3(\chi_3)], \quad (2.19b)$$

$$Q_3=g_2^*(\chi_2)g_1(\chi_1)/[1+\gamma_1G_1(\chi_1)+\gamma_2G_2(\chi_2)]. \quad (2.19c)$$

Again, one should note that we have a large degree of flexibility in shaping any *one* profile in (2.19). However, after the first one is shaped, say Q_3, then g_1 and g_2 have been specified, and only g_3 is left arbitrary. Now, we are only free to specify "one-half," of another profile. So for the general one-lump solution, we may only shape "one and one-half" profiles arbitrarily, subject to the product form in (2.19). Even with this restriction, one can still obtain useful information concerning the nature of the interactions in the 3D3WRI, as we shall illustrate later.

One can similarly obtain the final profiles, designated by $\tilde{Q}_i$, by taking the limit of $t\to+\infty$ in (2.12). This gives

$$\tilde{Q}_j=g_i^*g_k/(1+\gamma_j\Gamma_j+\gamma_iG_i+\gamma_kG_k), \quad (2.20)$$

where

$$\Gamma_j=G_j(-\infty)=\int_{-\infty}^{\infty}g_j^*g_j(u)\,du. \quad (2.21)$$

Lastly, we should comment that the two-lump solution given by (2.17) has six arbitrary functions, whereas the one-lump solution has only three. Thus, with a two-lump solution one may be able to shape the asymptotic initial profiles independently, instead of only "one and one-half", as is the case for the one-lump solution. However, for now we shall only concentrate on the one-lump solution in spite of the restrictions, due to the simplicity.

III. ONE-LUMP INTERACTIONS

In this section, we shall investigate the range of interactions available in the one-lump solutions. What we shall do, in as far as is possible, is to specify the initial actions in the asymptotic envelopes, and determine via the one-lump solution what the final actions must be. In general we will not be concerned with what the exact initial shape or form is. We will only be interested in how much action is in the envelope. As we shall see, although the shapes are cross-correlated as in (2.19), the actions are not, and can be specified independently. Thus due to this simplification, it is well worthwhile to study this case.

We define the action in the j^{th} envelope A_j by

$$A_j(t)=\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dy\int_{-\infty}^{\infty}dz\,q_j^*q_j. \quad (3.1)$$

In order to evaluate the above integral, we transform to characteristic coordinates. Directly from (2.4) for t constant, one can show that

$$\begin{aligned}dx\,dy\,dz&=J\,d\chi_1\,d\chi_2\,d\chi_4\\&=J\,d\chi_2\,d\chi_3\,d\chi_4\\&=J\,d\chi_3\,d\chi_1\,d\chi_4\end{aligned} \quad (3.2)$$

where

$$J=|(\mathbf{v}_1-\mathbf{v}_2)\times(\mathbf{v}_2-\mathbf{v}_3)|. \quad (3.3)$$

Thus (3.1) becomes

$$A_j(t)=J\int_{-\infty}^{\infty}d\chi_4a_j(\chi_4,t), \quad (3.4)$$

where the "reduced action" a_j is defined by

$$a_j(\chi_4,t)=\int_{-\infty}^{\infty}d\chi_i\int_{-\infty}^{\infty}d\chi_kq_j^*q_j|_{\chi_j=-t-\chi_i-\chi_k}. \quad (3.5)$$

In practice, it is impossible to evaluate (3.5) when $\chi_j=-t-\chi_i-\chi_k$. However, if instead we define

$$\alpha_j(\chi_4,\chi_j)=\int_{-\infty}^{\infty}d\chi_i\int_{-\infty}^{\infty}d\chi_k\,q_j^*q_j, \quad (3.6)$$

and since as $t\to\pm\infty$, q_j becomes independent of χ_j and we have

$$a_j(\chi_4,t\to\pm\infty)=\alpha_j(\chi_4,\chi_j\to\mp\infty), \quad (3.7)$$

thus allowing us to evaluate the initial and final reduced actions from α_j. In fact, from (2.11) and (2.12), we have the closed form solution

$$\alpha_j=\gamma_k\gamma_k\ln\rho_j, \quad (3.8)$$

where

$$\rho_j=\frac{(1+\gamma_i\Gamma_i+\gamma_jG_j)(1+\gamma_k\Gamma_k+\gamma_jG_j)}{(1+\gamma_jG_j)(1+\gamma_i\Gamma_i+\gamma_k\Gamma_k+\gamma_jG_j)}.$$

Directly from this, we have that the asymptotic initial reduced action a_{jo} is

$$a_{jo}=\gamma_i\gamma_k\ln[(1+\gamma_i\Gamma_i)(1+\gamma_k\Gamma_k)/(1+\gamma_i+\Gamma_i+\gamma_k\Gamma_k)], \quad (3.9)$$

and the final reduced action a_{jf} is

$$a_{jf}=a_{jo}+\gamma_i\gamma_k\ln\sigma, \quad (3.10)$$

$$\sigma=\frac{(1+\gamma_1\Gamma_1+\gamma_2\Gamma_2)(1+\gamma_1\Gamma_1+\gamma_3\Gamma_3)(1+\gamma_2\Gamma_2+\gamma_3\Gamma_3)}{(1+\gamma_i\Gamma_i)(1+\gamma_2\Gamma_2)(1+\gamma_3\Gamma_3)(1+\gamma_1\Gamma_1+\gamma_2\Gamma_2+\gamma_3\Gamma_3)}. \quad (3.11)$$

Note that up to a sign carried by $\gamma_i\gamma_k$, the change in the action is the same for all three envelopes, in agreement with the three conservation laws of action, which are

$$\frac{d}{dt}\int d^3\chi(\gamma_iq_i^*q_i+\gamma_jq_j^*q_j)=0, \quad (3.12)$$

Now, according to (3.9), the initial reduced actions are a function of the three Γ's, and one can also invert this to

can always depend on χ_4.) Note that these functions are completely arbitrary; thus there is a large amount of freedom in shaping this pulse. However, the shape is not totally arbitrary, because the functional dependence of q_1 on χ_2 and χ_3 is required to be of the form of a simple product, except for the terms in the denominator.

When all the g's are nonzero in general, we have nontrivial interactions and solutions with the general solution being given by (2.12). To obtain the asymptotic initial profiles (the profiles as $t\to-\infty$) according to (2.4b), we take the limit of the corresponding characteristic coordinate going to $+\infty$. Thus, to get the initial profile for g_1, which we shall designate by Q_1, we let $\chi_1\to+\infty$ in (2.12) for $(i,j,k)=(3,1,2)$ giving,

$$Q_1 = g_3^*(\chi_3)g_2(\chi_2)/[1+\gamma_2 G_2(\chi_2)+\gamma_3 G_3(\chi_3)]. \quad (2.19a)$$

Similarly, the other initial profiles are

$$Q_2 = g_1^*(\chi_1)g_3(\chi_3)/[1+\gamma_1 G_1(\chi_1)+\gamma_3 G_3(\chi_3)], \quad (2.19b)$$

$$Q_3 = g_2^*(\chi_2)g_1(\chi_1)/[1+\gamma_1 G_1(\chi_1)+\gamma_2 G_2(\chi_2)]. \quad (2.19c)$$

Again, one should note that we have a large degree of flexibility in shaping any *one* profile in (2.19). However, after the first one is shaped, say Q_3, then g_1 and g_2 have been specified, and only g_3 is left arbitrary. Now, we are only free to specify "one-half," of another profile. So for the general one-lump solution, we may only shape "one and one-half" profiles arbitrarily, subject to the product form in (2.19). Even with this restriction, one can still obtain useful information concerning the nature of the interactions in the 3D3WRI, as we shall illustrate later.

One can similarly obtain the final profiles, designated by $\tilde{Q}_i$, by taking the limit of $t\to+\infty$ in (2.12). This gives

$$\tilde{Q}_j = g_i^* g_k/(1+\gamma_j\Gamma_j+\gamma_i G_i+\gamma_k G_k), \quad (2.20)$$

where

$$\Gamma_j = G_j(-\infty) = \int_{-\infty}^{\infty} g_j^* g_j(u)\,du. \quad (2.21)$$

Lastly, we should comment that the two-lump solution given by (2.17) has six arbitrary functions, whereas the one-lump solution has only three. Thus, with a two-lump solution one may be able to shape the asymptotic initial profiles independently, instead of only "one and one-half", as is the case for the one-lump solution. However, for now we shall only concentrate on the one-lump solution in spite of the restrictions, due to the simplicity.

III. ONE-LUMP INTERACTIONS

In this section, we shall investigate the range of interactions available in the one-lump solutions. What we shall do, in as far as is possible, is to specify the initial actions in the asymptotic envelopes, and determine via the one-lump solution what the final actions must be. In general we will not be concerned with what the exact initial shape or form is. We will only be interested in how much action is in the envelope. As we shall see, although the shapes are cross-correlated as in (2.19), the actions are not, and can be specified independently. Thus due to this simplification, it is well worthwhile to study this case.

We define the action in the j^{th} envelope A_j by

$$A_j(t) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz\, q_j^* q_j. \quad (3.1)$$

In order to evaluate the above integral, we transform to characteristic coordinates. Directly from (2.4) for t constant, one can show that

$$\begin{aligned} dx\,dy\,dz &= J\,d\chi_1\,d\chi_2\,d\chi_4 \\ &= J\,d\chi_2\,d\chi_3\,d\chi_4 \\ &= J\,d\chi_3\,d\chi_1\,d\chi_4 \end{aligned} \quad (3.2)$$

where

$$J = |(\mathbf{v}_1-\mathbf{v}_2)\times(\mathbf{v}_2-\mathbf{v}_3)|. \quad (3.3)$$

Thus (3.1) becomes

$$A_j(t) = J\int_{-\infty}^{\infty} d\chi_4 a_j(\chi_4,t), \quad (3.4)$$

where the "reduced action" a_j is defined by

$$a_j(\chi_4,t) = \int_{-\infty}^{\infty} d\chi_i \int_{-\infty}^{\infty} d\chi_k q_j^* q_j|_{\chi_j=-t-\chi_i-\chi_k}. \quad (3.5)$$

In practice, it is impossible to evaluate (3.5) when $\chi_j = -t-\chi_i-\chi_k$. However, if instead we define

$$\alpha_j(\chi_4,\chi_j) = \int_{-\infty}^{\infty} d\chi_i \int_{-\infty}^{\infty} d\chi_k\, q_j^* q_j, \quad (3.6)$$

and since as $t\to\pm\infty$, q_j becomes independent of χ_j and we have

$$a_j(\chi_4, t\to\pm\infty) = \alpha_j(\chi_4,\chi_j\to\mp\infty), \quad (3.7)$$

thus allowing us to evaluate the initial and final reduced actions from α_j. In fact, from (2.11) and (2.12), we have the closed form solution

$$\alpha_j = \gamma_k\gamma_k \ln\rho_j, \quad (3.8)$$

where

$$\rho_j = \frac{(1+\gamma_i\Gamma_i+\gamma_j G_j)(1+\gamma_k\Gamma_k+\gamma_j G_j)}{(1+\gamma_j G_j)(1+\gamma_i\Gamma_i+\gamma_k\Gamma_k+\gamma_j G_j)}.$$

Directly from this, we have that the asymptotic initial reduced action a_{jo} is

$$a_{jo} = \gamma_i\gamma_k \ln[(1+\gamma_i\Gamma_i)(1+\gamma_k\Gamma_k)/(1+\gamma_i+\Gamma_i+\gamma_k\Gamma_k)], \quad (3.9)$$

and the final reduced action a_{jf} is

$$a_{jf} = a_{jo} + \gamma_i\gamma_k \ln\sigma, \quad (3.10)$$

$$\sigma = \frac{(1+\gamma_1\Gamma_1+\gamma_2\Gamma_2)(1+\gamma_1\Gamma_1+\gamma_3\Gamma_3)(1+\gamma_2\Gamma_2+\gamma_3\Gamma_3)}{(1+\gamma_1\Gamma_1)(1+\gamma_2\Gamma_2)(1+\gamma_3\Gamma_3)(1+\gamma_1\Gamma_1+\gamma_2\Gamma_2+\gamma_3\Gamma_3)}. \quad (3.11)$$

Note that up to a sign carried by $\gamma_i\gamma_k$, the change in the action is the same for all three envelopes, in agreement with the three conservation laws of action, which are

$$\frac{d}{dt}\int d^3\chi(\gamma_i q_i^* q_i + \gamma_j q_j^* q_j) = 0, \quad (3.12)$$

Now, according to (3.9), the initial reduced actions are a function of the three Γ's, and one can also invert this to

obtain the Γ's in terms of the initial reduced actions. To do this, one first defines

$$m_j = [\gamma_i\gamma_k - \gamma_i\gamma_k \exp(-\gamma_i\gamma_k a_{j_0})]^{1/2}, \tag{3.13}$$

which is always real, and for $a_{j_0} \ll 1$,

$$m_j \to (a_{j_0})^{1/2}. \tag{3.14}$$

From (3.9) and (3.13) one can show

$$\Gamma_j/(1+\gamma_j\Gamma_j) = m_i m_k/m_j. \tag{3.15}$$

One should note that no nontrivial solution exists when any one of the reduced actions is chosen to be zero. This is a consequence of using the one-lump solution. However, we may choose any reduced action to be as small as desired, but not zero, as long as a physical solution will exist for (3.15).

Assuming that a physical solution does exist for (3.15), σ as given by (3.11) becomes

$$\sigma = \frac{(1-\gamma_2\gamma_3 m_1^2)(1-\gamma_1\gamma_3 m_2^2)(1-\gamma_1\gamma_2 m_3^2)}{1-\gamma_2\gamma_3 m_1^2-\gamma_1\gamma_3 m_2^2-\gamma_1\gamma_2 m_3^2+2\gamma_1\gamma_2\gamma_3 m_1 m_2 m_3}. \tag{3.16}$$

Then from (3.10), (3.13), and (3.16), one finally obtains

$$m_{jf}^2 = (m_j - \gamma_j m_i m_k)^2/(1-\gamma_j\gamma_k m_i^2)(1-\gamma_i\gamma_j m_k^2), \tag{3.17}$$

where m_f is that value of m corresponding to the final value of the reduced action, as given in (3.13). We shall now consider the two possible physical cases.

A. The positive energy case

When all envelopes have a positive energy one of the γ's must differ in sign from the other two,[9] and that corresponding wave then has the largest central frequency. Without loss of generality we shall choose this wave to be wave number 1, and thus we have

$$\gamma_1 = -\gamma_2 = -\gamma_3. \tag{3.18}$$

We still have a freedom in choosing the sign of γ_1, since we set the value of δ in (2.10b) equal to $+1$, and as we shall see, changing the sign of γ_1 will give a different solution.

In order to orientate ourselves, we first list out the values for m_j^2 and Γ_j. From (3.13)

$$m_1^2 = 1 - e^{-a_{10}}, \quad \text{thus } 0 < m_1 < 1, \tag{3.19a}$$

$$m_2^2 = e^{a_{20}} - 1, \quad \text{thus } 0 < m_2 < \infty, \tag{3.19b}$$

$$m_3^2 = e^{a_{30}} - 1, \quad \text{thus } 0 < m_3 < \infty, \tag{3.19c}$$

and from (3.15),

$$\Gamma_1 = m_2 m_3/(m_1 - \gamma_1 m_2 m_3), \tag{3.20a}$$

$$\Gamma_2 = m_1 m_3/(m_2 + \gamma_1 m_1 m_3), \tag{3.20b}$$

$$\Gamma_3 = m_1 m_2/(m_3 + \gamma_1 m_1 m_2). \tag{3.20c}$$

Note that by (3.19), as the initial reduced actions are increased, the values of m increase also, and that m_1 must be bounded from above by unity, while m_2 and m_3 are unbounded. From (3.20) we shall obtain the allowed range of the initial reduced actions if the collision is to be represented by one-lump solutions.

First we shall consider the $\gamma_1 = +1$ case. Then for Γ_1 to remain positive and finite we must require

$$m_1 > m_2 m_3, \tag{3.21}$$

and then (3.17) becomes

$$m_{1f}^2 = (m_1 - m_2 m_3)^2/(1+m_2^2)(1+m_3^2), \tag{3.22a}$$

$$m_{2f}^2 = (m_2 + m_1 m_3)^2/(1-m_1^2)(1+m_3^2), \tag{3.22b}$$

$$m_{3f}^2 = (m_3 + m_1 m_2)^2/(1-m_1^2)(1+m_2^2), \tag{3.22c}$$

When $m_1 \gg m_2 m_3$ this solution corresponds to a decay of the high frequency pulse q_1 into the two daughter waves q_2 and q_3. To illustrate this, we shall assume $m_2 = m_3$, and then from (3.19) and (3.22) we find that

$$e^{a_{1f}} = e^{a_{20}} e^{a_{10}}/(1 + 2m_1 m_2^2 e^{a_{10}}), \tag{3.23}$$

which for $a_{10} \to \infty$ and $a_{20} \ll 1$ gives

$$a_{1f} \to -\ln(2a_{20}), \tag{3.24}$$

so that the final action in a very intense high frequency pump, will be determined solely by the initial actions in the daughter waves. This is illustrated in Fig. 1, where we have plotted $\log a_{1f}$ vs $\log a_{1_0}$ for various values of $a_{20} = a_{30}$. The main feature to note is that as a_{10} increases (for a given value of $a_{20} = a_{30}$), a_{1f} remains equal to a_{10} (i.e., no depletion occurs) until a critical value of about $-\ln(2a_{20})$ is reached, at which point, all additional action in the pump beyond this critical value is dumped into the daughter waves. Note that this is analogous to the soliton decay case in the one-dimensional 3WRI.[9] However, there the critical value depended essentially only on the absolute *area* under the profile of the pump itself, whereas here, it depends on the amount of action in the perturbing *daughter* waves.

This phenomena of pump decay would be expected to hold true even if the three pulse profiles were not one-lump solutions because all that is basically happening is that the daughter waves are traveling through the pump, and at the

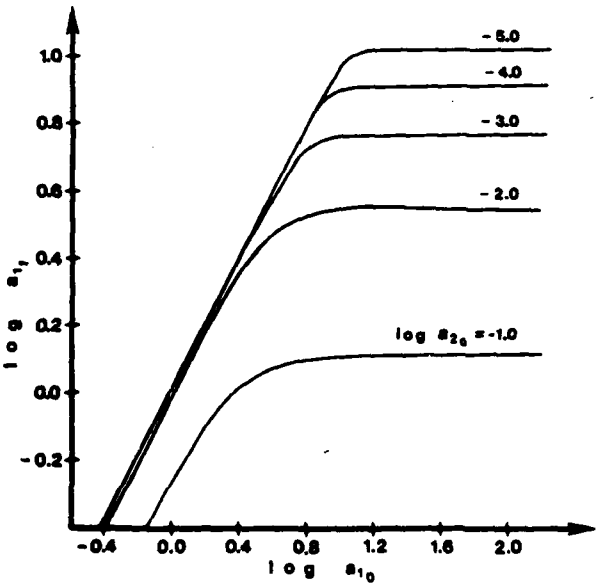


FIG. 1. The decay case. An initially intense high frequency pump of reduced action a_{1_0} will decay to a final reduced action of a_{1f}, depending on the initial reduced actions ($a_{20} = a_{30}$) in the daughter waves.

critical value, the growth rate has become equal to the relative pulse width allowing depletion to occur. Beyond this simple picture our result confirms what one would expect. Namely, the relative depletion becomes more and more as the initial action in the pump is increased.

There is another region of this solution where m_1 approaches $m_2 m_3$ from above, and from (3.22a), one can see that for this, virtually total depletion of the pump can occur. But we do not believe this feature to be a general feature. Rather, we consider it to be a special property of 1-lump solutions, and in general it should not occur for general pulses.

When $\gamma_1 = -1$ we have the opposite of decay, which is upconversion. Without loss of generality, we may assume

$$m_3 \geq m_2, \tag{3.25a}$$

in which case the allowed physical parameter space, from (3.20), is

$$m_1 < m_2/m_3, \tag{3.25b}$$

and (3.17) becomes

$$m_{f_1}^2 = (m_1 + m_2 m_3)^2/(1 + m_2^2)(1 + m_3^2), \tag{3.26a}$$

$$m_{f_2}^2 = (m_2 - m_1 m_3)^2/(1 - m_1^2)(1 + m_3^2), \tag{3.26b}$$

$$m_{f_3}^2 = (m_3 - m_2 m_1)^2/(1 - m_1^2)(1 + m_2^2). \tag{3.26c}$$

If we again consider the case where $m_2 = m_3$, then (3.25) becomes trivial and no restrictions are placed on our parameter space. Then

$$e^{a_{1f}} = e^{2a_{20}}/[(1 - m_1)(2e^{a_{20}} - 1 + m_1)], \tag{3.27}$$

and for $a_{20} \to \infty$ with $a_{10} \ll 1$, we find

$$a_{1f} \to a_{20} - \ln 2, \tag{3.28a}$$

while in the limit of $a_{10} \to 0$ and $a_{20} \to 0$ we find

$$a_{1f} \to [a_{20} + (a_{10})^{1/2}]^2. \tag{3.28b}$$

Equation (3.28) gives two important limits for the final action in the pump. In Fig. 2, we have plotted $\log a_{1f}$ vs $\log a_{20}$ for various values of $\log a_{10}$. The main features are that as $a_{20} \to 0$, $a_{1f} \simeq a_{10}$, so that no conversion has occured. Then as a_{20} increases, when it reaches the critical value of $a_{20} \simeq (a_{10})^{1/2}$, upconversion starts to occur with a_{1f} rapidly approaching the critical value given by (3.28a). Again, this feature is similar to that which occurs in the one-dimensional 3WRI,[9] except that there the critical value depended only on the absolute areas under the daughter profiles, whereas here it depends only on the amount of action in the high frequency pump. Still, for large amounts of action in the daughter waves, an almost total upconversion can occur.

B. The negative energy case

The only other case which differs from the above is when the high frequency wave is a negative energy wave, in which case

$$\gamma_1 = \gamma_2 = \gamma_3. \tag{3.29}$$

Now, the corresponding expressions for the m's and the Γ's are

$$m_j^2 = 1 - e^{-a_{j0}}, \qquad \text{thus } 0 < m_j < 1, \tag{3.30}$$

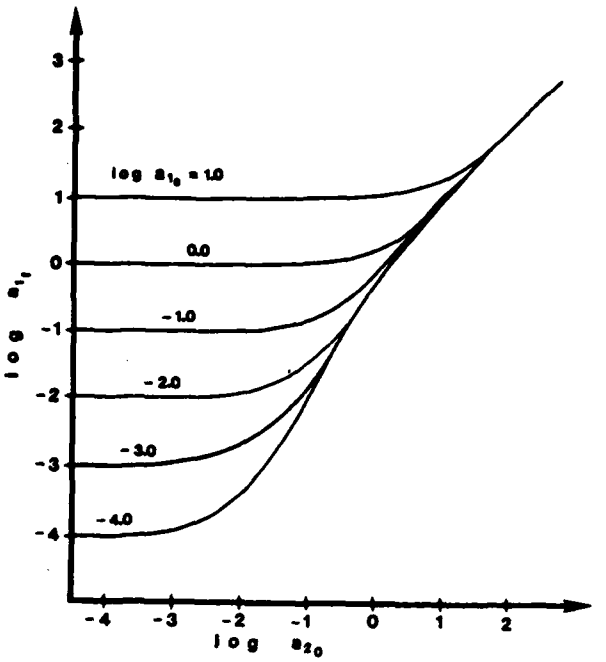


FIG. 2. The upconversion case. Two interacting daughter waves of reduced action $a_{20} = a_{30}$ interact and pump energy into the high frequency wave, creating a pump wave of final reduced action a_{1f} from a_{10}.

$$\Gamma_j = m_i m_k/(m_j - \gamma_1 m_i m_k), \tag{3.31}$$

while (3.17) becomes

$$m_{jf}^2 = (m_j - \gamma_1 m_i m_k)^2/(1 - m_i^2)(1 - m_k^2). \tag{3.32}$$

When $\gamma_1 = -1$ (3.31) gives no restrictions on the parameter space, so we shall handle this solution first. By (3.30), if the final reduced actions are to be real then $m_{jf}^2 < 1$, which by (3.32) requires

$$m_1^2 + m_2^2 + m_3^2 + 2m_1 m_2 m_3 < 1. \tag{3.33}$$

If this requirement is not satisfied by the initial reduced actions, then a singular solution will develop with all three a_j's becoming infinite at some finite time. This can be verified from (2.12). A similar result was obtained in the one-dimensional case.[9] However, here the system seems to be *linearly stable*, since if m_2 and m_3 become infinitesimal, then q_1 is only unstable for m_1 infinitesimally close to unity (i.e., $a_{10} \to \infty$).

In Fig. 3 we show this stability region for the case where $a_{20} = a_{30}$. In this case, the stability line from (3.33) is determined by

$$e^{a_{20}} = 2/[1 + (1 - e^{-a_{10}})^{1/2}]. \tag{3.34}$$

As $a_{10} \to 0$, (3.34) gives that

$$a_{20} \to \ln 2 - (a_{10})^{1/2}, \tag{3.35a}$$

while as $a_{20} \to 0$ we have

$$a_{10} \to -\ln(4a_{20}). \tag{3.35b}$$

If we consider q_1 to be the pump and q_2 and q_3 to be the daughter waves, then (3.35) and Fig. 3 show that once the action in each of the two daughter waves has exceeded ln2, the solution will be unstable for any value of the action in the

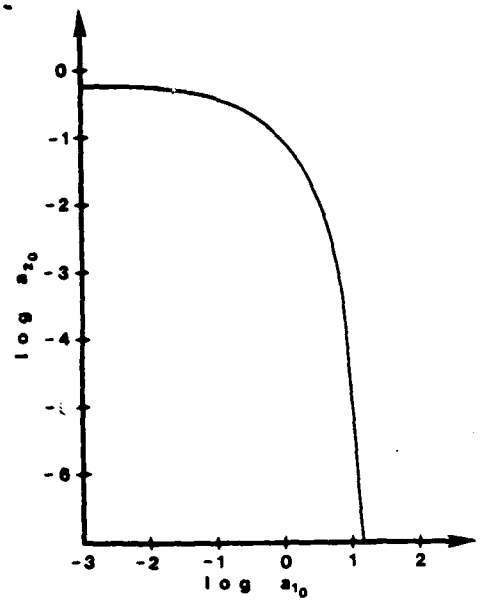


FIG. 3. The negative energy case for $a_{20} = a_{30}$. When the initial reduced actions lie outside and above the curve, the solution will always become singular in a finite time. All solutions under the curve are stable.

pump. On the other hand, as the daughter waves become infinitesimal, a large pump is stable until its action exceeds the value in (3.35b). Thus in theory, a large negative energy pump can be stable, but due to the critical value in (3.35b) being logarithmic, even the smallest finite noise level will place a practical upper limit on this stability.

The $\gamma_1 = +1$ solution is stable, although the parameter space is restricted. Without loss of generality, we can take

$$m_3 > m_2 > m_1, \tag{3.36a}$$

in which case (3.31) gives

$$m_1 > m_2 m_3. \tag{3.36b}$$

Now, we use (3.32) to demand that $m_{jf}^2 < 1$ gives

$$m_1^2 + m_2^2 + m_3^2 - 2m_1 m_2 m_3 < 1, \tag{3.37}$$

which due to (3.36) is always satisfied.

IV. CONCLUSIONS

With somewhat remarkable simplicity, we have found a quite interesting set of exact solutions for the 3D3WRI which very nicely illustrate such effects as pulse decay, upconversion, and explosive instabilities. This was all accomplished by considering the very simple one-lump solution obtained from the Bäcklund transformation. We would expect that two- and three-lump solutions would provide even more interesting results.

[1] D. J. Kaup, "The Solution of the General Initial Value Problem for the Full Three Dimensional Three-Wave Resonant Interaction" (To appear in the Proceedings of the 1970 Soviet–American Soliton Symposium, Kiev, USSR, Sept., 1979, and in Physica D.)

[2] V. E. Zakharov, Dokl. Akad. Nauk. SSSR Scr. Mat. Fiz. 228, 1314 (1976) [Sov. Phys. Dokl. 21, 322 (1976)].

[3] A. D. D. Craik, "Evolution in Space and Time of Resonant Wave Triads. Part II: A Class of Exact Solutions," Proc. R. Soc. London, Ser. A 363, 257–269 (1978).

[4] H. Cornille, J. Math. Phys. 20, 1653 (1979).

[5] D. J. Kaup, Physica D 1, 45 (1980).

[6] J. Corones, J. Math. Phys. 17, 756 (1976).

[7] D. J. Kaup, *Mathematical Methods and Applications of Scattering Theory*, Vol. 130 of "Lecture Notes in Physics", edited by J. A. DeSanto, A. W. Saenz, and W. W. Zachary (Springer, New York, 1980), pp. 247–254.

[8] L. P. Nižnik (to appear in the Transactions of the Conference in the Memory of I. G. Petrzovskii, 1976, Moscow State University Press).

[9] D. J. Kaup, A. H. Reiman, and A. Bers, Rev. Mod. Phys. 51, 275 (1979).

Whistler wave self-modulation in a tokamak plasma

S. N. Antani and D.J. Kaup

Department of Physics, Clarkson College, Potsdam, New York 13676
(Received 31 March 1980; accepted 17 March 1981)

Steady-state self-modulation of a whistler wave propagating at an arbitrary angle to the applied magnetic field, in a typical tokamak plasma, is investigated. The envelope is found to be modulationally stable for the parameter regime chosen.

Linear analysis of whistler propagation in a tokamak plasma has been reported by Theilhaber and Bers.[1] In the present work we include the effects of weak, ponderomotive nonlinearity. Several workers have reported the modulational instability of whistler wave in a magnetoplasma.[2-4] However, the parameter space of tokamak plasma has not been considered.

We work under the following basic assumptions: (i) The necessary accessibility condition for propagation into the plasma interior is fulfilled. This together with the requirement of minimum attenuation at the plasma edge limits the allowable range of the parallel refractive index, $n_z = k_z c/\omega$, to a narrow band, typically, $1.2 < n_z < 3$. (ii) The exciting structure is infinite in the y direction (thus, $d/dy = 0$). (iii) The density gradient can be treated as a weak perturbation, becoming important at the same order as the nonlinearity. (iv) Only a cold, collisionless plasma is considered. The temperature effects enter only through ponderomotive density modification. In our two-dimensional model, we take the external magnetic field along the z direction and let the plasma be weakly inhomogeneous in the x direction. We start with Maxwell's equations written in component form,[1]

$$(\partial_z^2 + K_\perp)E_x - i|K_x|E_y - \partial_x\partial_z E_z = 0\,, \tag{1a}$$

$$i|K_x|E_x + (\partial_z^2 + \partial_x^2 + K_\perp)E_y = 0\,, \tag{1b}$$

$$-\partial_z\partial_x E_x + (\partial_x^2 + K_\parallel)E_z = 0\,. \tag{1c}$$

In Eq. (1), the time dependence has been removed as $\exp(-i\omega t)$, where ω is the incident frequency such that $\Omega_i \ll \omega \ll \Omega_e$, with Ω_i, Ω_e being the electron and ion gyrofrequencies, respectively. Also, the space variables x, z and the electric field components are expressed in units of the incident wavenumber $k_0 = \omega/c$, and the thermal energy $(16\pi n_{0c} T_{0c})^{1/2}$, respectively. T_{0c}, n_{0c} are the central values of plasma temperature and density. Equation (1) is nonlinear due to the density dependence of the dielectric tensor elements

$$K_\perp = K_{\perp 0} + (K_{\perp 0} - 1)\frac{\delta n}{n_0}\,;\quad K_x = K_{x0}\left(1 + \frac{\delta n}{n_0}\right) K_\parallel = K_{\parallel 0}\left(1 + \frac{\delta n}{n_0}\right),$$

where as usual

$$K_{\perp 0} \simeq 1 - \frac{\omega_{pi0}^2(x)}{\omega^2} + \frac{\omega_{pe0}^2(x)}{\Omega_e^2}\,,\quad K_{x0} \simeq \frac{\omega_{pe0}^2(x)}{\omega\,\Omega_e}\,,$$

and

$$K_{\parallel 0} \simeq -\frac{\omega_{pe0}^2(x)}{\omega^2}\,.$$

$\omega_{pj0}^2(x)$ are the plasma frequencies of species j, corresponding to the unperturbed density $n_0(x)$. δn is the density perturbation caused by the ponderomotive force. One derives the expression for $\delta n/n_0$ in the usual way by considering the dc part of the electron-ion motion and subsequently balancing the ponderomotive and thermokinetic pressures. The result[5] can be expressed in matrix form as

$$\delta_n/n_0 = -(E^* M E)\,, \tag{2}$$

 0031-9171/81/071391-03$00.90 © 1981 American Institute of Physics

where $+$ means an adjoint, E is the column vector with elements E_x, E_y, and E_z. M is the matrix formed out of the linear part of the dielectric tensor elements $K_{\perp 0}$, K_{x0}, and $K_{\parallel 0}$. We then represent Eq. (1) in operator form as

$$LE = (\delta n/n_0)ME\,. \tag{3}$$

In L is the linear operator constructed from Eq. (1) after taking out the nonlinear part.

Next, we carry out the standard multiple scale procedure[6,7] on the system given by (3). Write the electric field as

$$E = \mathcal{E}\exp\left(i\int n_x\,dx + in_z z\right) + \text{c.c.}\,, \tag{4}$$

where $n_x = k_x c/\omega$ is the perpendicular refractive index, and expand the slowly varying envelope $\mathcal{E}$ as,

$$\mathcal{E} \simeq \epsilon\mathcal{E}^1 + \epsilon^2\mathcal{E}^2 + \cdots, \tag{5}$$

where $\epsilon \ll 1$ is a formal ordering parameter. We assume the ordering

$$|in_x| \gg |u\partial_\zeta| \gg |\partial_x|;\ |in_z| \gg |\partial_\zeta|\,,$$

where $\zeta = z - \int^x u\,dx$ and u will be defined shortly. Substituting (4) and (5) into (3) we find at first order that $\mathcal{E}^1$ must be an eigenvector of $L(in_x, in_z)$ with a zero eigenvalue, and that the dispersion relation for the fast wave, using the approximations, $|K_\parallel| \gg |K_\perp| \simeq 1$ and $K_{x0}^2 \ll -K_\parallel$ must be

$$n_x^2 \simeq K_{x0}^2/(n_z^2 - K_{\perp 0}) - (n_z^2 - K_{\perp 0})\,. \tag{6}$$

At the next order we find that u must equal the ratio of the group velocity components, or

$$u = \frac{v_{gz}}{v_{gx}} = -\frac{dn_x}{dn_z}\,. \tag{7}$$

Finally, at third order, we arrive at the following form of the nonlinear Schrödinger equation

$$i\partial_x\psi + p\partial_\zeta^2\psi + q|\psi|^2\psi - ir\partial_x(\ln n_0)\psi = 0\,, \tag{8}$$

where ψ is simply the amplitude of $\mathcal{E}^1$ [i.e., $\mathcal{E}^1 = \psi V_1$, where $L(in_x, in_z)V_1 = 0$ and $V_1^+ V_1 = 1$]. Equation (8) describes the slow, stationary evolution of the envelope $\psi(x,\zeta)$. The last term in (8) arises from the spatial nonuniformity which we have treated perturbatively. Further, the coefficients p, q, and r depend on x and represent, respectively, the effects of group dispersion, nonlinearity, and nonuniformity. We find the expressions for p, q, and r using the linear dispersion (6) and noting that $n_z^2 \ll |K_{\parallel 0}|$. These are

$$p \simeq \frac{[-(2n_z^2 - K_{\perp 0})K_{x0}^4 + 6n_z^2K_{x0}^2(n_z^2 - K_{\perp 0})^2 + (n_z^2 - K_{\perp 0})^4K_{\perp 0}]}{2n_z(n_z^2 - K_{\perp 0})^3[K_{x0}^2 - (n_z^2 - K_{\perp 0})^2]}\,, \tag{9}$$

$$q \simeq \frac{[K_{x0}^2(2n_z^2 - K_{\perp 0} - 1) + (K_{\perp 0} - 1)(n_z^2 - K_{\perp 0})^2]^2}{2n_z(n_z^2 - K_{\perp 0})^3[K_{x0}^2 + (n_z^2 - K_{\perp 0})^2]\theta(x)}\,, \tag{10}$$

$$r \simeq \frac{(3n_z^2 - K_{\perp 0} - 2)K_{x0}^2 + (K_{\perp 0} - 1)(n_z^2 - K_{\perp 0})^2}{(n_z^2 - K_{\perp 0})[K_{x0}^2 + (n_z^2 - K_{\perp 0})^2]}\,. \tag{11}$$

Equations (3) and (9)–(11) are our basic results. In Eq. (10)

$$\theta(x) = n_0(x)T(x)/n_{0c}T_{0c}\,.$$

We next choose the following typical tokamak plasma parameters: central density, $n_{0c} \simeq 2\times 10^{13}$ cm^{-3}; rf frequency, $\omega = 5\times 10^9$ rad/sec; magnetic field $B_0 \simeq 20$ kG; plasma radius $a = 10$ cm; central electron temperature $T_{ec} = 1$ keV; central ion temperature $T_{ic} = 200$ eV and $n_z = 1.5$. The density and temperature profiles are taken to be parabolic. Then, $\theta(x) = (1 - x^2/a^2)^2$.

In the central, quasi-homogeneous region, we can ignore the last term in Eq. (8) and have roughly, $|K_{\perp 0}| \simeq 1$, $|K_{x0}| \sim 10$, and $|K_{\parallel 0}| \sim 10^3$. Then, Eqs. (9)–(11) simplify. In Eq. (9) for p, the first term in the numerator is dominant showing p to be negative. From (10) we can see that q is positive definite since the numerator is squared. [This follows from the form of (2) and (3). The same matrix M occurs in (2) and (3) making the numerator of q a perfect square.] Thus $pq < 0$, which means that the wave is modulationally stable. Therefore, no solitons can form and the nonlinear propagation will qualitatively follow that of the linear wave. In addition to nonlinearity, as one goes toward the edge, inhomogeneities can also become important effects. In order to discuss these, it is convenient to rescale Eq. (8). Define $\phi = \psi v^{-1}$, $v^2 = 2q^{-1}$, and $\xi = (-p)^{-1/2}\zeta$. Then, (8) becomes

$$i\partial_x\phi - \partial_\xi^2\phi + 2|\phi|^2\phi + if(x)\phi = 0\,, \tag{12}$$

which again is modulationally stable and

$$f(x) = \partial_x[\ln(2/q)^{1/2}] - r\partial_x(\ln n_0)\,. \tag{13}$$

Now, if ϕ is sufficiently small so that the nonlinear term in (12) can be neglected, then (12) becomes linear and nothing more needs to be said since it is well known how to solve such a linear problem. Hence, we are left with only the case when both the inhomogeneity and the nonlinearity are simultaneously important. Frequently, in such cases, nothing definite can be said without much work. However, for lower hybrid waves in the electrostatic approximation, Leclert *et al.*,[8] treated the inhomogeneity as a weak perturbation. At the edge of the plasma using the approximations, $\omega_{pi}^2/\omega^2 \ll 1$ and $\omega_{pe}^2/\Omega_e^2 \ll 1$ and assuming a linear temperature profile, they found that $f(x)$ can be approximated as $f(x) \simeq 1/2x$. In this case they found that Eq. (8) was exactly solvable.

We shall note that $f(x)$ as given by (13) is profile dependent and goes almost like $1/2x$, but not exactly. For a parabolic density profile we find that $f(x)$ corresponds to the exactly solvable case within about 30% in the range, $0.2 < \bar{x} \equiv x/a < 0.5$. On the other hand, for linear density-temperature profiles, we find that $f(x)$ could be approximated over a wider $\bar{x}$ range in order to correspond to the exactly solvable case. Thus, we conclude, subject to the profile giving an $f(x)$ reasonably close to $1/2x$, that one also can simply apply the results of Leclert *et al.* to whistler wave propagation.

It seems that self-focusing (i.e., soliton formation) will not occur for the nonlinear "fast" wave in a tokamak plasma. This could prove to be advantageous in heating the bulk of the plasma for fusion applications.

This research was supported in part by Office of Naval Research contract number N00014-76-C-0867.

[1]K. Theilhaber and A. Bers, Nucl. Fusion 20, 547 (1980).
[2]V. I. Karpman and H. J. Washimi, J. Plasma Phys. 18, 173 (1977).
[3]M. Watanabe, Phys. Rev. A 18, 2689 (1978).
[4]K. H. Spatscheck, P. K. Shukla, M. Y. Yu, and V. I. Karpman, Phys. Fluids 22, 576 (1979).
[5]W. S. Wang and H. H. Kuehl, Phys. Fluids 22, 1708 (1979).
[6]D. J. Benney and A. C. Newell, J. Math. Phys. 46, 133 (1967).
[7]A. C. Newell, Lect. Appl. Math. 15, 157 (1974).
[8]G. P. Leclert, C. F. F. Karney, A. Bers, and D. J. Kaup, Phys. Fluids 22, 1545 (1979).

From: NONLINEAR PHENOMENA IN PHYSICS AND BIOLOGY
Edited by Richard H. Enns, Billy L. Jones,
Robert M. Miura, and Sadanand S. Rangnekar
(Plenum Publishing Corporation, 1981)

THE LINEARITY OF NONLINEAR SOLITON EQUATIONS AND THE THREE WAVE RESONANCE INTERACTION

D.J. Kaup

Physics Department
Clarkson College of Technology
Potsdam, N.Y. 13676

ABSTRACT

Based on a representation found by Professor Newell, it is illustrated how the general solution of nonlinear soliton equations are really simple linear sums. Then the general three-wave resonant interaction is discussed, what the resonance is, and the three-dimensional scattering problem which is used to solve it. In particular, we contrast the standard one-dimensional inverse scattering with this form of three-dimensional inverse scattering. The principle difference is that no bound states occur in this form of three-dimensional inverse scattering, and thus no solitons exist in three dimensions. However, it is shown that from a Bäcklund transformation, one can construct localized solutions, called "lump" solutions, and their properties are discussed.

TABLE OF CONTENTS

I. INTRODUCTION 97

II. THE SCATTERING PROBLEM FOR THE THREE-DIMENSIONAL THREE-WAVE RESONANT INTERACTION 105

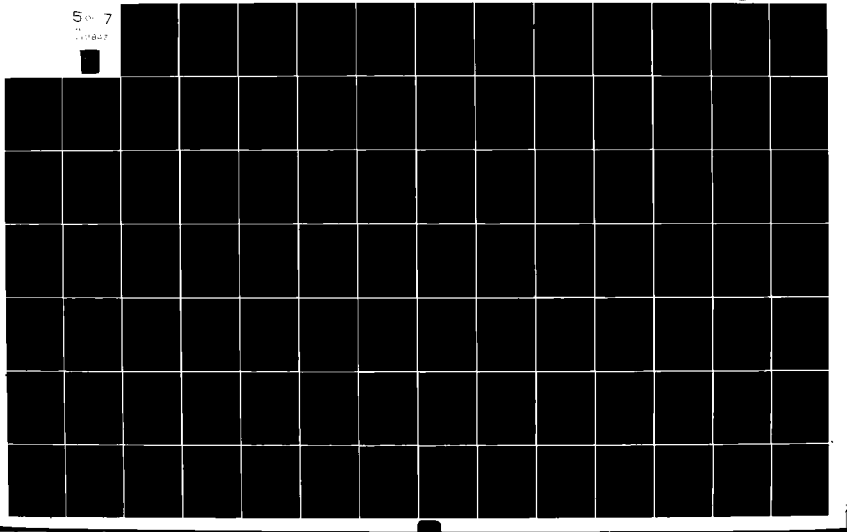
III. THE THREE-DIMENSIONAL BÄCKLUND TRANSFORMATION 115

REFERENCES 122

AD-A119 843 CLARKSON COLL OF TECHNOLOGY POTSDAM NY DEPT OF MATHE--ETC F/G 12/1
NONLINEAR WAVES.(U)
SEP 82 M J ABLOWITZ, D J KAUP, A C NEWELL N00014-76-C-0867

UNCLASSIFIED NL

5 of 7



THE LINEARITY OF NONLINEAR SOLITON EQUATIONS AND THE THREE WAVE RESONANCE INTERACTION

D.J. Kaup

Physics Department
Clarkson College of Technology
Potsdam, N.Y. 13676

I. INTRODUCTION

The purpose of a school such as this is to provide a background of information for those interested in a given particular subject. The basic fundamentals of solitons have already been well expanded on by the previous talks, and in general, the most that I could add without excessive duplication is simply additional references [Kaup, 1977; Kaup and Newell, 1978a, b; Kaup et al, 1979], which describe my viewpoint on these basics. However, there is one basic point that still has not been emphasized here, and which I have always considered to be striking and important. That is the fact that although these systems are indeed nonlinear, their behavior so closely mocks or imitates linear systems, that one is frequently ahead if he simply forgets that it is nonlinear, and looks upon the system as being essentially linear. For example, Professor Newell [Kaup and Newell, 1978b] has demonstrated a striking representation of the general solution for q in terms of the squared eigenstates. This expansion, for the Zakharov-Shabat (ZS) case with $r = -q^*$, is

$$q = -\frac{1}{\pi}\int_{-\infty}^{\infty}\left(\frac{b}{a}\psi_1^2 + \frac{b^*}{a^*}\psi_2^{*2}\right)d\zeta + 2i\sum_k\left(\gamma_k\psi_{1k}^2 - \gamma_k^*\psi_{2k}^{*2}\right), \quad (1.1)$$

with the notation as in Kaup and Newell [1978b]. Note that the solution, as given in (1.1), is an exact linear sum (integral) over a radiation (continuous spectrum) part, and a linear sum over each and every bound state (soliton). There is absolutely no difficulty in identifying what part of the solution is radiation, and what part belongs to the k^{th} soliton. It's right there in the expansion

given by (1.1). The only way that the nonlinearity enters at all in (1.1) is in the squared eigenstates. In the radiation part, the eigenstate is a plane wave only as $x \to \pm\infty$, and for x finite, undergoes phase shifts and a WKB swelling of its amplitude as it passes over the position of any bound state. But still, qualitatively these continuous eigenstates are very much like plane waves, and the radiation part of (1.1) is simply a "nonlinear Fourier transform", [Ablowitz et al, 1974], where b/a is the continuous part of the nonlinear Fourier transform. Similarly, since the soliton part of (1.1) is given in terms of the bound states, it follows that each soliton solution is localized. And again, the only way the nonlinearity enters in the soliton part of (1.1) is through the squared bound states, where the radiation will only, at most, minorly influence the shape of these bound states.

Thus these exactly integrable systems exhibit the feature that although they are nonlinear, their behavior is very remarkably linear. Mathematically, this is demonstrated by the system being exactly integrable [Ablowitz et al, 1974], by (1.1), and even may be visually seen in almost any computer plot of such solutions. As is well known, by a simple cursory visual examination of almost any such solution, one can pick out all the solitons and identify the radiation part, due to the usual striking contrast between the bound states and the continuous eigenstates. And, except for phase shifts, one can actually visually see that these states are noninteracting.

So, what I want to leave with you at this point, is the empirical observation, that if you consider these nonlinear systems to be "almost" linear, many of their "strange" features become understandable, and one can be lead to predictions [Kaup, 1977; Kaup and Newell, 1978 a,b; Kaup et al, 1979] concerning further features of such systems.

For the remainder of this lecture, I want to concentrate on some of the basics associated with three-dimensional inverse scattering, such as that which is used in solving the three dimensional three-wave resonant interaction (3WRI). But, before we even look at that, we first should describe this 3WRI, and explain what it is, because a basic understanding of it is very useful in understanding the associated 3D inverse scattering.

There are many examples that one can use to illustrate the 3WRI, one of which dealing with electromagnetic interactions, was given in our general review [Kaup et al, 1979] of the one-dimensional case. Instead of duplicating that one here, or something similar, I shall go almost to the other extreme, and instead consider only a very simple mathematical model. We could actually choose any hyperbolic system possessing a quadratic nonlinearity as its lowest order nonlinearity. But to keep things simple, we shall choose one

of the simplest of all such models, namely the wave-equation with a nonlinear term,

$$\partial_t^2\phi - \nabla^2\phi = \varepsilon\phi^2 \tag{1.2}$$

where ϕ is a real field. Again, I emphasize that (1.2) is simply a mathematical model, which contains the essential features which we require: propagating waves and a quadratic nonlinearity. Almost any such system with propagating waves and a quadratic nonlinearity will give rise to the 3WRI in the weakly nonlinear limits. We shall now illustrate this with the model (1.2). With $\varepsilon = 0$, (1.2) is exactly linear, and one can create solutions of (1.2) using linear superposition. And, since such a linear solution remains linear for all time when $\varepsilon = 0$, there is nothing more that needs to be said about such a solution. But, when we "turn-on" the interaction and allow ε to be nonzero, sooner or later something is going to happen to destroy this linear superposition. Letting ϕ be a general linear solution becomes too difficult to attempt when $\varepsilon \neq 0$. So we simplify and the easiest simplification is to let ϕ be a sum of almost planar waves, that is let

$$\phi = \sum_{j=1}^{N} e^{i\theta_j} u_j(\vec{x},t) + \text{c.c.} \tag{1.3}$$

where

$$\theta_j = \vec{k}_j\cdot\vec{x} - \omega_j t \ , \tag{1.4}$$

c.c. means the complex conjugate, and where the envelopes, u_j, are allowed to be complex and are assumed to be slowly varying with respect to the phase of the plane waves. In other words, any envelope will contain many, many oscillations or wavelengths under it. Furthermore, we assume each $\vec{k}_j$ to be well separated from the others.

Under these conditions, the zeroth-order solution (1.3) will appear to be composed of N distinct envelopes. Each one of these envelopes will contain many, many oscillations, and each of these oscillations will be described by a wave-vector, $\vec{k}_j$, and a frequency, ω_j. And as one looks at any specific wave or oscillation, inside of any envelope, and considers only its nearest neighboring waves, the solution does appear to be a plane wave. It is only when one considers the solution over many oscillations that one does then note the eventual decay of the wave, due to the envelope. The purpose of these envelopes are two-fold. First, they serve to keep the total energy and spacial extent finite, and second, there are no real physical plane-wave solutions. All physical solutions are really bounded, and so using an envelope gives us a physical solution.

When $\varepsilon = 0$, (1.3) is an exact solution, providing we restrict our choices of ω_j to

$$\omega_j^2 = \omega^2(\vec{k}_j) = k_j^2 \quad . \tag{1.5}$$

This is just the linear dispersion relation. Now, to understand what can happen to the solution when $\varepsilon \neq 0$, we shall first consider the $\varepsilon\phi^2$ term to be an inhomogeneous term, and thereby driving ϕ, in (1.3), like a source term would. So, we use the linear Green's function, and have that

$$\phi = \phi_{hom} + \varepsilon \int_{-\infty}^{t} dt' \int d^3x' \; G(\vec{x}-\vec{x}',t-t')\phi^2(\vec{x}',t') \quad , \tag{1.6}$$

where

$$G(\vec{x},t) = \frac{1}{(2\pi)^3} \int \frac{d^3k}{\omega(\vec{k})} e^{i\vec{k}\cdot\vec{x}} \sin[t\omega(\vec{k})] \quad , \tag{1.7}$$

and

$$\omega(\vec{k}) = (k^2)^{\frac{1}{2}} \quad . \tag{1.8}$$

In (1.6), ϕ_{hom} is the homogeneous solution of (1.3). Now, I want to remind you that one interpretation of (1.6) is in terms of Huygen's principle. This interpretation is as follows. At $\vec{x}'$ and t', an amount of ϕ is created by the ϕ^2 term, of amplitude and phase ϕ^2, which is then propagated to $\vec{x}$ and t by G, and its final amplitude and phase at $\vec{x}$ and t is simply $\varepsilon\phi^2$. Now, one simply integrates over all possible values of $\vec{x}'$ and t' for the final result, (1.6). Well, let's simply consider what terms might be present in ϕ^2, and take the simple case where we have only two envelopes. Then

$$\phi^2 = u_1^2 e^{2i\theta_1} + u_2^2 e^{2i\theta_2} + u_1^* u_1 + u_2^* u_2$$
$$+ 2u_1 u_2 e^{i(\theta_1-\theta_2)} + 2u_1 u_2 e^{i(\theta_1+\theta_2)} + c.c. \tag{1.9}$$

Now, take (1.6) and (1.7) and rearrange it so that we do the integral over $\vec{x}'$ first, over $\vec{k}$ second, and over t' last. The first four items in (1.9) correspond to simple harmonic generation [Kaup, 1978] and side band modulations, which are only special cases of the 3WRI. So, we shall ignore these terms here, and put our attention on the 5th and 6th terms, which represent an interaction between the two envelopes. Then in (1.6), the integral over $\vec{x}'$ will give its largest contribution when $\vec{k} = \vec{k}_1 \pm \vec{k}_2$, and also when $\vec{k} = -(\vec{k}_1 \pm \vec{k}_2)$.

And, at these values of $\vec{k}$, the integral over $\vec{x}'$ also gives very strongly peaked functions of $\vec{k}$, which are almost delta functions in the above variables. Thus, in $\vec{k}$-space, the interaction separates into distinct regions, centered at each of the above values of $\vec{k}$. The width of each region in $\vec{k}$-space is, of course, inversely proportional to the width of the associated envelope in $\vec{x}$-space. Now each of these regions in $\vec{k}$-space will have a time dependent phase associated with it. At $\vec{k} = \vec{k}_1 + \vec{k}_2$, this phase is $-(\omega_1 + \omega_2)t$, at $\vec{k} = \vec{k}_1 - \vec{k}_2$, this phase is $-(\omega_1 - \omega_2)t$, etc., all of which follows from (1.4), (1.6), and (1.9).

Next, we consider what happens when we do the d^3k integral in (1.7). And since the d^3x' integral gave essentially delta functions in $\vec{k}$-space, the d^3k integral simply gives us a sum, one term for each region in $\vec{k}$-space. But this just, in effect, gives us back what we started with, except that the amplitude and phase factors in (1.7) have been inserted. For example, about $\vec{k} = \vec{k}_1 + \vec{k}_2$, we obtain

$$2\varepsilon \frac{u_1 u_2}{\omega(\vec{k}_1+\vec{k}_2)} e^{i(\vec{k}_1\cdot\vec{x}+\vec{k}_2\cdot\vec{x}-\omega_1 t'-\omega_2 t')} \sin[(t-t')\omega(\vec{k}_1+\vec{k}_2)], \quad (1.10a)$$

and about the region $\vec{k} = \vec{k}_1 - \vec{k}_2$, we similarly have

$$2 \frac{u_1 u_2^*}{\omega(\vec{k}_1-\vec{k}_2)} e^{i(\vec{k}_1\cdot\vec{x}-\vec{k}_2\cdot\vec{x}-\omega_1 t'+\omega_2 t')} \sin[(t-t')\omega(\vec{k}_1-\vec{k}_2)], \quad (1.10b)$$

and similarly for the other terms.

The result exemplified by (1.10), interpreted in terms of Huygen's principle, is simply the net result of all wavelets produced by $\varepsilon\phi^2$ at $\vec{x}'$ and at t', upon summing over all possible values of $\vec{x}'$, chosen such that all wavelets end up at $\vec{x}$, and at the time t.

We now consider the last integral over dt'. Note the time dependencies in (1.10a) and (1.10b). We shall simply state the result. If in (1.10a), $\omega \neq \omega_1 + \omega_2$, then this term will remain bounded by the order of ε for all time. Thus the effect of this term is simply to produce a small perturbation on the linear solution. But, if $\omega = \omega_1 + \omega_2$, then this term will grow linearly in t, and cannot be considered as a small perturbation. In fact, it can grow so large that it starts to deplete the initial envelopes. The interaction has now become fully nonlinear. This term does not then produce just a small perturbation on the linear solution. This term will drastically alter these initially linear solutions, and will cause them to interact. In terms of Huygen's principle, the wavelets produced by the $\vec{k}_1 + \vec{k}_2$ term, (1.10a), at t', reach t with the phase factors $(\omega_1 + \omega_2 \pm \omega)t$, since the sine term in (1.7) contains both signs for ωt. Thus, in general, these wavelets

will reach t with various and almost random phases, such that when one adds over all t', much cancellation occurs. But, if $\omega = \omega_1 + \omega_2$, every one of these wavelets reaches t with exactly the same phase as every other wavelet. The wavelets are now all coherent and in phase, so that when one adds over all t', one obtains a linear growth in (t-t'), which can only be limited in its growth by the actual finite extent of the initial interacting envelopes, which up until now, we have considered to be absent, by using only plane waves. This is a resonance phenomena, when $\omega = \omega_1 + \omega_2$. When this resonance phenomena occurs, a small perturbation, as in (1.2), can drastically affect the linear solution, so much so, that one must now consider the nonlinear solution, and recognize that (1.6) is not valid when resonance occurs. And similarly, for those terms which do not contain resonance, (1.6) is still a valid method of treatment, since they always remain small.

Now, what happens at such a resonance? As one can see from (1.10a), a new wave, of wavevector $\vec{k}_1 + \vec{k}_2$, is being created. It has a frequency $\omega = \omega(\vec{k}_1 + \vec{k}_2)$, and thus is a linear solution of (1.2), due to (1.5). It grows linear in t (at least initially), and when the interaction turns off (by the two initial envelopes eventually separating) it propagates as a free wave. So, two waves have interacted resonantly, producing a third, wave, whence the name "three-wave resonant interaction".

Of course, this interaction is not restricted to the model (1.2), but will occur in any system which has envelope waves and a quadratic nonlinear interaction, and where the resonance conditions

$$\vec{k}_1 + \vec{k}_2 + \vec{k}_3 = 0, \tag{1.11a}$$

$$\omega_1 + \omega_2 + \omega_3 = 0, \tag{1.11b}$$

$$\omega_i = \omega(\vec{k}_i), \tag{1.11c}$$

can be satisfied. In (1.11), we are introducing a convention where the sum of all $\vec{k}$'s and ω's are zero, which can always be done for a specific set of $\vec{k}$'s and ω's, by simply redefining them with appropriate signs. This form is chosen, since the resulting 3WRI equations will be highly symmetric.

So, when the resonance condition (1.11) is satisfied, we cannot obtain a solution by simple perturbation techniques, as in (1.6), but must consider the weakly nonlinear ($\varepsilon \to 0$) case. Since two waves are interacting to produce a third, we start with three waves

$$\phi = e^{i\theta_1}u_1 + e^{i\theta_2}u_2 + e^{i\theta_3}u_3 + \text{c.c.} \tag{1.12}$$

and due to (1.11), the three phases are not independent, but satisfy

$$\theta_1 + \theta_2 + \theta_3 = 0 \quad . \tag{1.13}$$

Now let's insert (1.12) into (1.2), retaining only the lowest order terms. This gives

$$-2i \sum_{j=1}^{3} \omega_j e^{i\theta_j} (\partial_t - \frac{1}{\omega_j} \vec{k}_j \cdot \vec{\nabla}) u_j$$

$$= \varepsilon \begin{bmatrix} 2u_1^* u_2^* e^{-i(\theta_1+\theta_2)} + 2u_1^* u_2 e^{-i(\theta_1-\theta_2)} \\ + u_1^* u_1 + \ldots \end{bmatrix} , \tag{1.14}$$

and where we shall still ignore all effects of harmonic generation[1]. Upon balancing terms of like phase in (1.14), one obtains

$$(\partial_t - \vec{v}_j \cdot \vec{\nabla}) u_j = \frac{i\varepsilon}{\omega_j} u_k^* u_\ell^* \quad , \tag{1.15}$$

where $\vec{v}_j = \vec{k}_j/\omega_j$ is the group velocity (since dispersion is absent) and (j,k,ℓ) in (1.15) is cyclic in (1,2,3). There are thus three equations contained in (1.15). Upon scaling the envelope amplitudes, one can obtain (1.15) in the canonical form [Kaup et al, 1979],

$$(\partial_t - \vec{v}_j \cdot \vec{\nabla}) q_j = i\gamma_j q_k^* q_\ell^* \quad , \tag{1.16a}$$

where the q's are the scaled amplitudes, and

---

[1] Strictly speaking, for the model (1.2), one cannot ignore harmonic generation, since given any $\vec{k}$, there exists a solution for $\omega(2\vec{k}) = 2\omega(\vec{k})$. Thus due to the absence of dispersion in (1.2), harmonic generation is always resonant, and can never be ignored. In more physical models which have dispersion, solutions will exist to $\omega(2\vec{k}) = 2\omega(\vec{k})$, only for special values of $\vec{k}$. Thus, in general, harmonic generation is not resonant, and indeed may be ignored, unless one of the $\vec{k}$'s happen to be one of these special values.

$$\gamma_j = \mathrm{sgn}(\omega_j E_j) \quad , \tag{1.16b}$$

where E_j is the energy of the j^{th} envelope.

Once given (1.16), the method of solution depends first on the number of independent dimensions, or directions, appearing in (1.16a). For example, if you were given the q's as being independent of $\vec{x}$, then you have only a one-dimensional problem (only a t-dependence) regardless whether or not the group velocities are in independent directions. Another way to have a one-dimensional problem is to have all three group velocities equal. Now, we have only one characteristic direction, with the different streamlines never crossing or mixing. So, we simply solve the one-dimensional problem along each and every streamline.

In each case when the 3WRI reduces to a one-dimensional problem, one can always give the general solution in terms of elliptic functions [Armstrong et al, 1962]. When only two of the derivatives in (1.16) are independent, one has what is commonly referred to as the "one-dimensional" case, by which one means one spatial dimension (plus one time dimension, giving a total of two dimensions). Solutions for this case were given first by Zakharov and Manakov [1973], and the recent paper [Kaup et al, 1979] in Reviews of Modern Physics gives an up-to-date account of what is known about this two-dimensional case, both in terms of inverse scattering and numerical calculations, and how they both compare.

So, this only leaves the most general case, namely when (1.16) has all three derivatives as being independent. This is the full three-dimensional (3D) case, and is the case I shall discuss here. I shall not attempt to explain all the "ins-and-outs" of 3D inverse scattering here, since first there is not sufficient time, and second, this audience is not primarily interested in these ins-and-outs, but rather in the general principles involved. So, here I shall seek to emphasize how this 3D problem differs from 1D inverse scattering, and how it is still similar. As for references on the 3D-3WRI inverse scattering, Cornille [1979] and Kaup [1980a] give the first successful attempt to create a workable, albeit clumsy, inverse scattering. The papers by Kaup [1980b; 1980c] are concerned with the "global scattering problem", which could nevertheless solve a special class of initial value problems. Kaup [1979] gives the final and complete solution of the general initial value problem, and with remarkable hindsight, the paper by Kaup [1981] then allows one to completely bypass inverse scattering, and construct physical solutions with a technique very similar to, and undoubtedly related to Bäcklund transformations. What I shall do next is to discuss the principle ideas involved in Kaup [1979], and conclude by discussing the three principle types of solutions given in Kaup [1981].

II. THE SCATTERING PROBLEM FOR THE THREE-DIMENSIONAL THREE-WAVE RESONANT INTERACTION

Before discussing the scattering problem for this system, I first want to ensure that the shape and nature of the solutions of the 3D-3WRI are understood, because centering on this, we can simplify the method of solution. First, the interaction region in space-time is often finite, and is defined to be that region where two or more of the envelopes overlap. If you are in a region of space-time where there is no overlap of the envelopes, there is no interaction, since $q_j^* q_k^* = 0$, and by (1.16), each envelope then propagates as a free envelope. To know what happens in the interaction region, we must solve the full nonlinear equations. But if we block this region off, and consider only all other parts of space-time, then we are considering only the noninteracting region, where the envelopes are free. And since there is no dispersion, in this region, these free envelopes simply undergo a simple translation, as shown by the free solution of (1.16). Thus in the noninteraction region, we know exactly what the solution is. And, usually this region is much larger than the interaction region. Contrast this with other cases, where dispersion is present. Then, one only knows the solution at infinity. Almost all of space-time is now the interaction region, and you only know the solution on the boundary. But due to the absence of dispersion in the 3WRI equations, the opposite occurs here, in that one knows the solution in almost all of space-time, while it is usually unknown in only a finite region of space-time.

So, let's now consider the solution in the noninteracting region. By (1.16), each envelope here satisfies

$$(\partial_t + \vec{v}_j \cdot \vec{\nabla}) q_j = 0, \tag{2.1}$$

which therefore moves with the velocity $\vec{v}_j$, without any distortion. Now, note that in total, we only have three operators, $(\partial_t + \vec{v}_j \cdot \vec{\nabla})$, while there are four space-time coordinates. This means that there is at least one linear combination of the space-time coordinates which is independent of these three operators. We shall call this combination χ_4, and define it by

$$(\partial_t + \vec{v}_j \cdot \vec{\nabla}) \chi_4 = 0, \tag{2.2}$$

for $j = 1,2$, and 3. For example, to find χ_4, go to the rest frame of one envelope, say q_1, then $\vec{v}_1 = 0$. Now rotate your coordinate system such that $\vec{v}_2$ and $\vec{v}_3$ lie in the x-y plane. Now, it is very obvious from (2.2) that $\chi_4 = z$. As a further point, consider (1.16) in this situation. Then no derivatives with respect to z ever occur in (1.16). Thus z (or rather, χ_4 in general) appears in (1.16) only as a parameter. This means that you may solve (1.16)

by fixing χ_4, obtain a solution, then fix χ_4 to be another value, etc. In other words, solutions at different values of χ_4 never mix or affect one another.

So, with χ_4 being defined by (2.2), there remains three other independent coordinates, which we shall define by

$$\frac{\partial}{\partial \chi_j} = \partial_j = -\partial_t - \vec{v}_j \cdot \vec{\nabla} \quad , \tag{2.3}$$

for j = 1,2,3. These coordinates we shall call "characteristic coordinates". As an example, if $\vec{v}_1 = 0$, $\vec{v}_2 = \hat{i}$, and $\vec{v}_3 = \hat{j}$, then a solution of (2.2) and (2.3) is

$$t = -\chi_1 - \chi_2 - \chi_3 \quad , \tag{2.4a}$$

$$x = -\chi_2 \quad , \tag{2.4b}$$

$$y = -\chi_3 \quad , \tag{2.4c}$$

$$z = \chi_4 \quad . \tag{2.4d}$$

Of course, other different solutions will exist for other orientations of the velocities.

From (2.1) and (2.3), we have for the free solutions that

$$\partial_j q_j = 0, \tag{2.1'}$$

or

$$q_j = q_j(\chi_k, \chi_\ell, \chi_4) \quad , \tag{2.5}$$

with j,k,ℓ cyclic in 1,2,3. Explicitly in terms of the very special and model situation represented by (2.4), the three equations contained in (2.5) are

$$q_1 = q_1(-x,-y,z), \tag{2.6a}$$

$$q_2 = q_2(-y, t-x-y,z), \tag{2.6b}$$

$$q_3 = q_3(t-x-y,-x,z). \tag{2.6c}$$

As one will note, q_1 satisfies $\partial_t q_1 = 0$, and q_2 satisfies $(\partial_t+\partial_x)q_2 = 0$, and q_3 satisfies $(\partial_t+\partial_y)q_3 = 0$.

So, what the above illustrates is that these characteristic coordinates are very special, are determined by the velocities, and are natural coordinates to use in this problem. When we use these coordinates, the free solutions take on a very special form, as given by (2.5).

Since χ_4 will only enter our equations as a parameter, we shall hereafter ignore it.

Note that at constant χ_k and χ_ℓ, the free solution for q_j does not change as χ_j is changed (provided we remain in the noninteracting region, where (2.1) is valid). Thus q_j can be considered to pa- gate along the $(\pm)\chi_j$-axis.

Usually, the initial envelopes are localized and could b considered to be bounded in space. As seen in the model (2.6 if bounded in x,y, and z, the solution is also bounded in chara stic coordinates. Since characteristic coordinates are natural c ates, we shall use these coordinates instead of space-time c - ates. Thus our results shall be independent of the actual group velocities, and shall be quite general.

Consider a localized free solution for q_j as in (2.5), and as shown in Figure 1. In Figure 1, I have attempted to pictorially represent what a typical solution of the 3D-3WRI (for some value of χ_4) might look like, using characteristic coordinates, and not space-time coordinates. (I want to emphasize that the angles between the χ-axes in Fig. 1 need not be 90° as shown. What these angles would be, would depend on the three group velocities, and their directions. However, since no result depends on the angles between these characteristic coordinates, we may arbitrarily stretch these angles out to be 90°. Thus, in Fig. 1, we show the angles to be 90°.) As it is drawn, we have two envelopes q_i and q_j coming together and interacting, creating some q_k, and then everything separates as $t \to +\infty$, with different final envelopes. We note that, as in (2.4a), the relation

$$t = -\chi_i - \chi_j - \chi_k \quad , \tag{2.7}$$

is a consequence of (2.3), and does not depend on how the velocities are directed. It is from (2.7) that one can understand how time is directed in Fig. 1, since (2.7) defines a plane with the directional derivative (1,1,1). Imagine such a plane in Fig. 1. As this plane would be moved toward the right in Fig. 1, t would become more negative. At any time t, the intersection of this plane with q_i, gives the profile of q_i at that value of t. Note how if this plane would be moved toward the origin, q_i and q_j would then approach each other. This would correspond to two pulses (envelopes) approaching each other. At the origin, they overlap and interact.

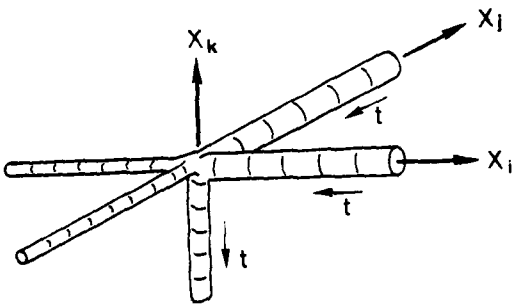


Fig. 1. A graphic representation of a solution of the 3D-3WRI in characteristic coordinates. This solution represents an initial q_j envelope and q_j envelope (each at the right) colliding, and then emerging with a reduced energy. In the process of interacting, an amount of q_k has been produced, and is seen traveling away from the collision center along its own characteristic.

Then as t becomes positive, they both come out the other side, with their original profiles changed and with q_k now being nonzero. Once t has become sufficiently large so that the envelopes no longer overlap, then the envelopes once again become free envelopes, and propagate according to (2.1).

Of course, there can be initial conditions where the envelopes would never separate, and there can be initial conditions whereby the solutions become singular. However those are special cases, and before one attempts to understand them, he had best understand this simple case as shown in Fig. 1. All it is, is two envelopes coming together, colliding and interacting, then separating. And in the process of interacting, a third envelope has been produced or created, which moves away along its own characteristic direction.

Let us now turn to the scattering problem itself. The original scattering problem was given by Zakharov and Shabat [1974]. However, it was not in a form that one could introduce characteristic coordinates into, which caused it to appear to be more complicated than it really was. Ablowitz and Haberman [1975] found another scattering problem which was in characteristic coordinate form, and is the one which we shall use. It is

$$\partial_k \psi_i = \gamma_k q_j^* \psi_k, \tag{2.8a}$$

$$\partial_i \psi_k = \gamma_i q_j \psi_i, \tag{2.8b}$$

where i,j, and k are cyclic, so that there are a total of six equations in (2.8). If one considers the integrability conditions for these six equations, one finds that the q's must satisfy

$$\partial_i q_i = \gamma_i q_j^* q_k^* , \tag{2.9}$$

where again, (i,j,k) are cyclic, so that (2.9) contains three equations, which if we interpret ∂_i as $\partial/\partial x_i$, are just the 3D-3WRI equations, (1.16).

Now in the initial value problem, one is given all three q's at some value of t, which we shall take to be at t = 0 for convenience. From (2.7), in terms of characteristic coordinates, we therefore initially know the q's only along the plane

$$x_j = -x_i - x_k . \tag{2.10}$$

So, we want to solve (2.8), only knowing this information. When one considers (2.8) relative to the plane defined by (2.10), one finds that three of the equations in (2.8) describe how the three

ψ's will evolve off of the plane (analogous to the t-evolution equations in 1D-IST) and the other three describe how the three ψ's propagate in this plane. Of course, since the plane is two-dimensional, having only three-first-order partial differential equations for three functions leaves the solution underdetermined, and it is exactly due to this underdeterminancy that this problem, even without any eigenvalue in (2.8), is still a scattering problem. It is not an eigenvalue problem as in 1D-IST, but it is still a scattering problem.

To restrict ourselves to the plane given by (2.10), we arbitrarily pick χ_i and χ_k to be our independent coordinates in this plane. Then provided we are restricted to be in this plane, the derivatives with respect to our two independent characteristic coordinates are related to the $\vec{\nabla}$ operator by

$$\partial_k = (\vec{v}_j - \vec{v}_k)\cdot\vec{\nabla} \quad , \tag{2.11a}$$

$$\partial_i = (\vec{v}_j - \vec{v}_i)\cdot\vec{\nabla} \quad , \tag{2.11b}$$

and the three equations from (2.8) which correspond to the scattering problem are

$$\partial_k\psi_i = \partial_k q_k^*\psi_k - \gamma_j q_k\psi_j \quad , \tag{2.12a}$$

$$(\partial_k - \partial_i)\psi_j = \gamma_k q_i\psi_k - \gamma_i q_k^*\psi_i \quad , \tag{2.12b}$$

$$\partial_i\psi_k = \gamma_i q_j\psi_i - \gamma_j q_i^*\psi_j \quad , \tag{2.12c}$$

and the reason this is a scattering problem is simply because $\partial_i\psi_i$, $(\partial_k + \partial_i)\psi_j$, and $\partial_k\psi_k$ are unspecified in (2.12). What (2.12a) does specify is how ψ_i changes as one moves in the χ_k direction, and nothing is said about how ψ_i must change as one moves in the χ_i direction. Similarly for (2.12b) and (2.12c), with different directions being involved. Due to this underspecification of these functions, integration of (2.12) gives

$$\psi_i = g_i(\chi_i) + \int_{\chi_k}^{\infty} \{\ \}(\chi_i,u)du \quad , \tag{2.13a}$$

$$\psi_j = g_j(\chi_i+\chi_k) + \int_{\chi_k}^{\infty} \{\ \}(\chi_i+\chi_k-v,v)dv \quad , \tag{2.13b}$$

$$\psi_k = g_k(\chi_k) + \int_{\chi_i}^{\infty} \{\ \}(w,\chi_k)dw \quad , \tag{2.13c}$$

where the brackets in (2.13) simply stand for the right-hand sides of (2.12), the exact form of which is unimportant for (2.13) and the following arguments. The three g's in (2.13) are arbitrary functions, and their specification will uniquely determine the solution. We emphasize that they are arbitrary, and they are arbitrary simply because (2.12) did underdetermine the solution for the ψ's.

Now, (2.12) are linear equations. So, the general solution may be given as a sum over some selected complete set of fundamental solutions. Furthermore, we can replace the arbitrary g's with the Fourier components, $e^{i\zeta x_i}$, etc., since from these plane waves, one can reconstruct any arbitrary function. So, for the fundamental solution we call ψ^i, we take

$$g_i(x_i) = e^{i\zeta x_i} , \tag{2.14a}$$

$$g_j = g_k = 0 , \tag{2.14b}$$

while for the fundamental solution ψ^j, we take

$$g_i = g_k = 0 , \tag{2.15a}$$

$$g_j(x_i + x_k) = e^{-i\zeta(x_i+x_k)} , \tag{2.15b}$$

and for ψ^k, we take

$$g_i = g_j = 0 , \tag{2.16a}$$

$$g_k(x_k) = e^{i\zeta x_k} . \tag{2.16b}$$

Now, let us see what the solution ψ^i, (2.14), corresponds to. In Fig. 2, we have pictorially represented this solution in the $t = 0$ plane. We assume the initial data to be on compact support, whence we can assume that the potentials (the q's) are only nonzero inside the central circle, as indicated. From (2.14) and (2.13a), we have for any x_k above the circle, that $\psi^i_i = e^{i\zeta x_i}$ exactly. Also note that if x_i lies to the left or the right of the circular region, then again $\psi^i_i = e^{i\zeta x_i}$. Thus ψ^i_i can only differ from $e^{i\zeta x_i}$ either inside the circular region, or directly underneath it. Similarly, one can argue that ψ^i_k is only nonzero in the circular region or directly to its left, and ψ^i_j is only nonzero in the circular region or to the lower right, as indicated in Fig. 2.

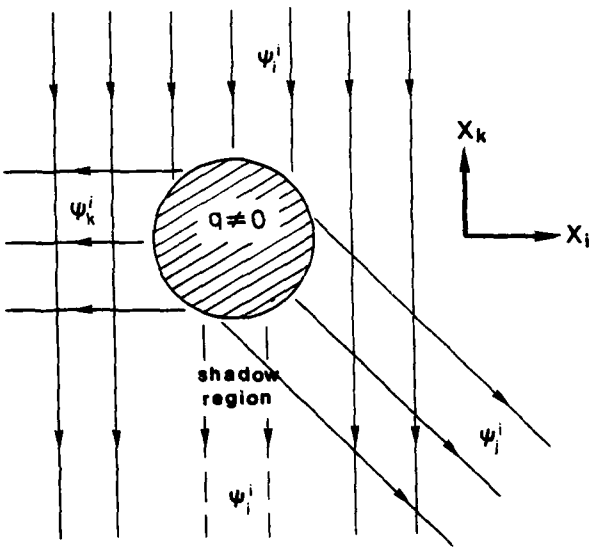


Fig. 2. The ψ^i solution in the x_i-x_k plane. The incident wave, ψ_i^i, is at the top, the transmitted wave, ψ_i^i, is the shadow region below, and ψ_k^i and ψ_j^i are the two reflected waves.

Now, note that each integral in (2.13) serves to propagate the solution of a particular component in a particular direction. These directions are indicated by the arrows in Fig. 2. We can now interpret Fig. 2 as follows. In Fig. 2, we see an incident wave, ψ_i^i entering at the top, which has transverse oscillations, $e^{i\zeta x_i}$. This wave interacts with the q's and creates out of this interaction two different waves, ψ_k^i and ψ_j^i, each of which propagates off in its own characteristic direction. So (2.12) is indeed a scattering problem, in that one wave is partially converted (scattered) into two others.

But what about the final profiles of $\psi_k^i(x_k)$ and $\psi_j^i(x_i+x_k)$? Well, as one would suspect, each of these final profiles, when Fourier transformed, defines a reflection coefficient, which depends on the exact structure of the potentials inside the circular region. Similarly, the final profile of $\psi_i^i(x_i)$ defines a transmission coefficient. So, we do have the potentials being mapped into reflection coefficients, as in 1D-IST. From the three solutions, ψ^i, ψ^j, and ψ^k, defined by (2.14)-(2.16), we therefore have two reflection coefficients per solution, for a total of six reflection coefficients. But, we have only three profiles, so that three of these six reflection coefficients are not independent, but are dependent on the other three.

Furthermore, (2.13) is not the only choice for a solution to (2.12), since one could choose different and the opposite limits for the integrals in (2.13). For example, changing the integral in (2.13a) to be from $-\infty$ to x_k, corresponds in Fig. 2 to reversing the direction of the corresponding reflected wave. In all, since there are 8 independent ways of choosing the limits on the integrals in (2.13), and since there are three independent solutions per choice, we have a total of 24 independent fundamental solutions, and therefore 48 different reflection coefficients, of which only three are independent. Contrast this with the 1D-IST case, where we have only 4 fundamental solutions [Ablowitz et al, 1974], $\psi, \overline{\psi}, \phi$, and $\overline{\phi}$, and only four different reflection coefficients [Ablowitz et al, 1974], b/a, $b/\overline{a}$, $\overline{b}/a$, and $\overline{b}/\overline{a}$, of which at most, two of the latter are independent.

So, 3D-IST has a much larger redundancy in it than does 1D-IST, with the reason for this arising from the larger number of directions which one could use for the incident beam, and the reflected beams. Furthermore, when one starts considering the number of possible forms for the inverse scattering equations themselves, we again find a similar large number, whereas in 1D-IST, there are only two possible forms: inversion about $+\infty$ or about $-\infty$. The methods required to obtain the 3D inverse scattering equations are detailed in Kaup [1979] and due to the complexity of the algebra involved, we shall not reproduce those results here. Instead we shall simply outline them, and also contrast them with 1D-IST.

First, we note that the t-dependence of the reflection coefficient can be determined in the standard manner. We use the three equations in (2.8) which corresponds to evolving the ψ's off of the $t = 0$ plane, and simply evaluate them for $\chi_j = \pm\infty$ and/or $\chi_k = \pm\infty$, whence the t-evolution of the reflection coefficients will follow, since the potentials are assumed to vanish at these limits. Second, one finds that each fundamental solution of (2.12), where the incident wave is represented as in (2.14)-(2.16), has a particular analytic property with respect to ζ. For example, the fundamental solution ψ^i, defined by (2.13) and (2.14) is in general only bounded on the real ζ-axis, whereas $\psi^j(\psi^k)$ is analytic in the lower-(upper-)half ζ-plane. Similarly, each of the 24 fundamental solutions has a definite analytical property, with respect to ζ. Furthermore, each fundamental solution can be related to almost any other set of three fundamental solutions, via the various transmission and reflection coefficients, since only three of the 24 fundamental solutions are linearly independent. From this, one can construct "linear integral dispersion relations" whereby one fundamental solution is given in terms of an inhomogeneous part and an integral in the complex ζ-plane, over products of reflection coefficients and other fundamental solutions. Once these are constructed, one has, in effect, solved the inverse scattering problem, because all that is left is simply "doing the algebra".

So far, everything which has been described above, has very closely paralleled 1D-IST, and has differed really only in the dimensionality, larger number of solutions, etc. But, it is at the point of constructing the linear integral dispersion relations, that one will note a distinct difference between 3D-IST and 1D-IST.

In 1D-IST, the poles of the transmission coefficient, $a^{-1}(\zeta)$, in the upper half ζ-plane, define the bound-state eigenvalues [Ablowitz et al, 1974], each one of which gives rise to a soliton. But in this 3D-IST, every transmission coefficient is bounded, and therefore no poles occur, and therefore no solitons exist, as we know them from 1D-IST. The inverse scattering is done completely in terms of the continuous spectrum, and only in terms of the continuous spectrum. No bound states ever occur.

But, in 1D-IST, one property of the inverse scattering equations is that when the continuous spectrum is absent, closed form solutions exist, which are the N-soliton solutions. This occurs because the kernels in the integral equations (which are the inverse scattering equations) become what is called "separable". So, in 1D-IST, one could equivalently define solitons as being the separable solutions of the IST equations. So, when one considers the 3D-IST equations, quite surprising, we find that these equations can possess separable kernels, and the corresponding solutions are

localized! And, one may also construct the analogy of the 1D N-soliton solution, except that these 3D objects are not quite solitons. And to distinguish them, we shall call these 3D objects "lumps". These 3D "lump solutions" we shall now discuss from a different point of view in the next section.

III. THE THREE DIMENSIONAL BÄCKLUND TRANSFORMATION

As one will recall, one of the more general means of constructing N-soliton solutions in one-dimension, is to use the Bäcklund transformation, whereby one starts with a trivial zero solution, and generates nontrivial solutions. Following a suggestion by Jim Corones [1976], simply applying a general technique, I have found that a Bäcklund transformation does exist for the 3D-3WRI, which is completely integrable, and is

$$\tilde{q}_j = q_j + \psi_i^* \psi_k / D \quad , \tag{3.1a}$$

where

$$\partial_i D = -\gamma_i \psi_i^* \psi_i \quad , \tag{3.1b}$$

and each above equation is cyclic in (i,j,k). In (3.1a), q_j is an initial known solution, and ψ_i is a solution of (2.8) for the known q's. By direct substitution, one may verify that if (2.8), (2.9) and (3.1b) is true, then $\tilde{q}_j$ will satisfy (2.9), being therefore a <u>new</u> solution. From (2.8), one may easily show that (3.1b) is integrable.

To see what kind of solutions we can generate with (3.1), just set $q_i = 0$, whence it follows from (2.8) that $\psi_j = g_j(x_j)$ and

$$\tilde{q}_j = \frac{1}{D} g_i^*(x_i) g_k(x_k) \quad , \tag{3.2a}$$

where

$$D = 1 + \gamma_1 G_1 + \gamma_2 G_2 + \gamma_3 G_3 \quad , \tag{3.2b}$$

$$G_i(x_i) = \int_{x_i}^{\infty} g_i^* g_i(u) du \quad . \tag{3.3}$$

The solution (3.2) is called a "1-lump" solution.

A trivial 1-lump solution is when one of the g's is chosen to be zero. We take $g_1 = 0$, then it follows that $q_2 = 0 = q_3$, while

$$q_1 = \frac{g_3^* g_2}{1+\gamma_2 G_2+\gamma_3 G_3} , \tag{3.4}$$

where in (3.4), g_i is a function only of χ_i. Now, (3.4) is not a soliton solution as we know from 1D-IST. First, we note that its shape and amplitude is quite arbitrary. Except for the denominator, it can be any product of any function of the other two characteristic coordinates. Note that it is independent of χ_1, which must be so, for a free envelope to exist. Thus in characteristic coordinate space, (3.4) describes a "tube" parallel to the χ_1-axis, when g_2 and g_3 are localized.

Now, these N-lump solutions can be constructed from a Bäcklund transformation, as we have seen here for the 1-lump solution. But, in 1D-IST, the Bäcklund transformation is a well-known and well-used procedure for generating 1D N-soliton solutions. So, one could say that these lump solutions are the extension of 1D solitons, and at the same time one could say that they are not. The choice that one takes will depend on one's point of view. But, the main point to be recognized is that these lump solutions are similar to and also are different from solitons. Thus they are indeed a different quantity, and to emphasize this, we have chosen to use the word "lump" to describe them.

Now, the general 1-lump solution, (3.2) although simple, is found to contain in it a breadth of information and examples, which is remarkable. With just this one solution, we can generate and exhibit such effects as pulse decay, upconversion, and explosive instabilities, all with one simple mathematical solution [Kaup, 1981]. The detailed analysis of (3.2) is carried out in Kaup [1981], so what I shall do here is to present the results, and hopefully enough of the mathematics to convince you that these results are true and reasonable. The first thing which must be done is to parameterize the solutions (3.2) in some manner. We choose to use the standard concept of "action" and define the initial and final action (a) of an envelope by

$$a_j = \int_{-\infty}^{\infty} d\chi_i \int_{-\infty}^{\infty} d\chi_k \, q_j^* q_j . \tag{3.5}$$

The relation that the above integral has to the corresponding integral in space-time, is discussed in Kaup [1981]. Here, we shall simply comment that for the initial and final pulses (but not necessarily for in-between times) the above integral does reduce to the usual definition of action.

Now, with the definition of initial and final actions by (3.5), one then finds that the specification of the three initial actions either i) uniquely determine all constants in (3.2), except for trivial phase factors, or ii) no solution of (3.2) exists for those initial actions. In the latter case, usually the 1-lump solution was just not sufficiently general to represent such a solution. So, there are some interaction regimes which are inaccessible to the 1-lump solution. But at the same time, we find that a very large interaction regime is accessible to this solution.

Of course, if one can determine an initial solution from the three initial actions, it follows that the final actions can be given as functions of the initial actions. This is the manner in which we shall discuss our results. We input a certain amount of action into each of the three "channels", or initial envelopes. Note that we say nothing about the shape of the profile. All we do is to specify a global quantity, the "action". Then after the interaction, the initial action will be redistributed among the three envelopes. We shall usually graph one of these final actions vs. two of the initial actions.

The results that one finds are as follows. For the positive energy case, one of the γ's must differ in sign from the other two, which we shall choose to be $\gamma_1 = -\gamma_2 = -\gamma_3$, where ω_1 is then the largest frequency. The decay case is one of the possible solutions in this case, and is pictorially represented in Fig. 3 as a function of space and time. The large bottom cylindrically shaped object corresponds to an initially intense high frequency envelope, and the lower dashed lines correspond to small amounts of the other waves, on a collision course with the high frequency envelope. These three waves collide and interact in the middle of the graph, and during the interaction, action is lost by the high frequency envelope (it decays in strength), and reappears in the low frequency waves. Thus, by this interaction, high frequency waves can decay into lower frequency waves, with the exact amount of loss of action by the high frequency envelope depending on the initial actions. In Fig. 4, I have graphed the final action of the high frequency envelope vs. the initial actions for $a_{20} = a_{30}$. One should note that, for a_{20} fixed, as a_{10} is increased from zero, at first nothing happens since $a_{1f} = a_{10}$. But soon, a critical threshold is reached after which a_{1f} remains at some constant value, with all action beyond this value being lost into the two lower frequency daughter waves.

The other positive energy solution is just the time reversal of the above decay interaction. In this case, simply reverse the direction of time in Fig. 3, and you will see two low frequency waves coming together on a collision course (also with a small amount of the high-frequency wave present), which then interact, surrendering action to the high frequency wave, which emerges

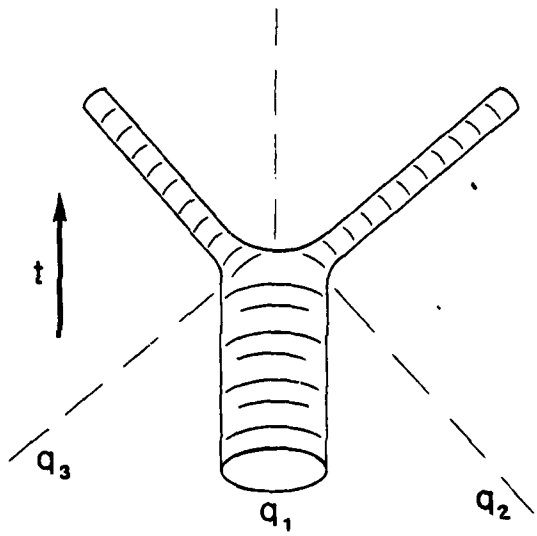


Fig. 3. The decay solution of the 3D-3WRI 1-lump solution. As shown, q_1 is the high frequency envelope, and is very intense. In the center region, a very small amount of q_2 and q_3 collides with q_1, causing a decay of q_1 into q_2 and q_3.

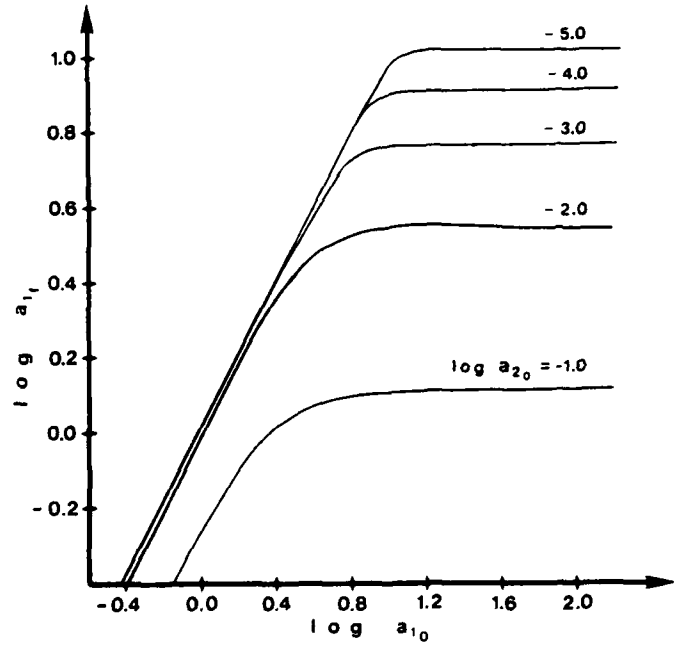


Fig. 4. The decay case. An initially intense high frequency pump of action a_{10} will decay to a final action a_{1f}, depending on the initial actions ($a_{20} = a_{30}$) in the daughter waves.

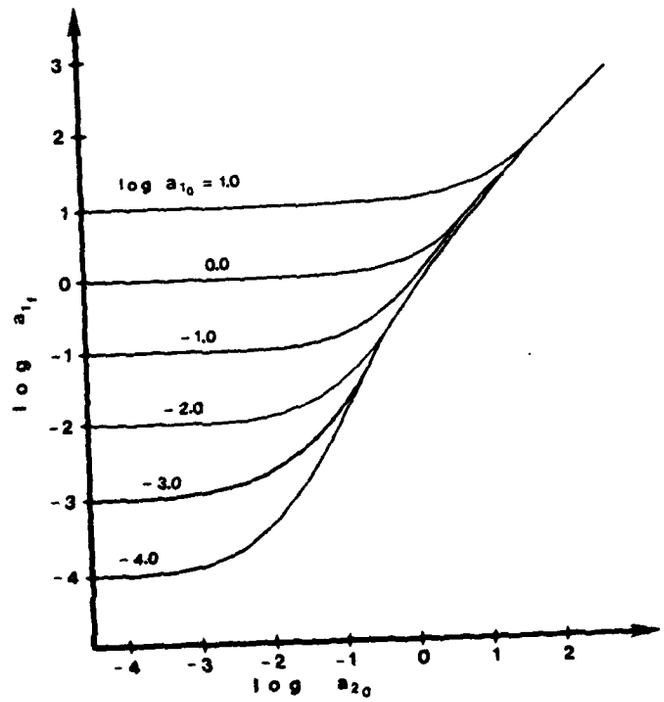


Fig. 5. The upconversion case. Two interacting daughter waves of action $a_{20} = a_{30}$ interact and pump energy into the high frequency wave, creating a pump wave of final action a_{1f}.

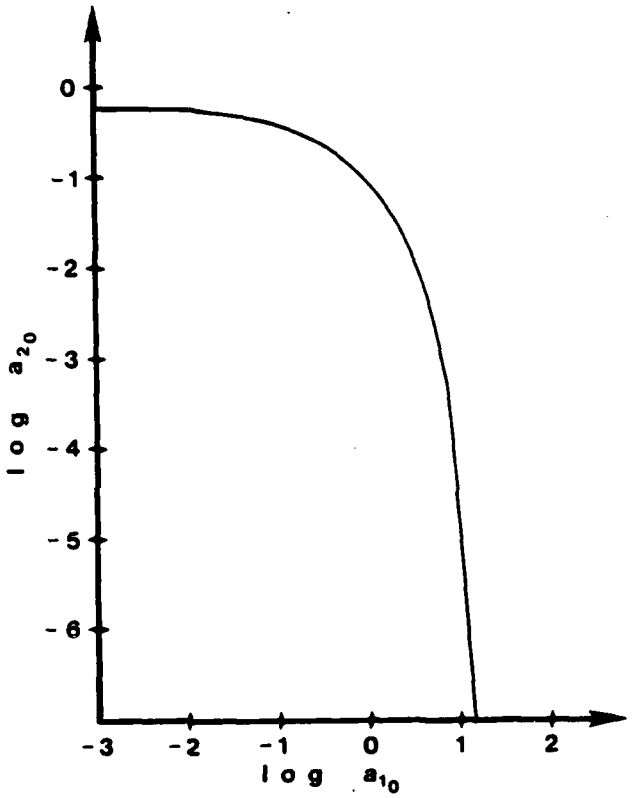


Fig. 6. The negative energy case for $a_{20} = a_{30}$. When the initial actions lie to the right and above the curve, the solution will always become singular in a finite time. All solutions under the curve are stable.

enhanced. This process is called upconversion, whereby energy passes from low frequency waves up into a higher frequency wave. The general results in this case for the 1-lump solution is demonstrated in Fig. 5. Here one should note again the threshold phenomena, whereby for fixed a_{10}, a_{1f} does not change until a_{20} passes a critical value, beyond which a_{1f} increases linearly with a_{20}.

The last case of a 1-lump solution is what is known as an explosive instability, which involves negative energy waves. In this case, $\gamma_1 = \gamma_2 = \gamma_3 = -1$, and as one can readily see, the denominator in (3.2a) may now become zero. When this happens, the solution becomes singular in a finite time, with each envelope having a singularity. The range of initial actions, for $a_{20} = a_{30}$, in which this instability occurs is shown in Fig. 6. Briefly, Fig. 6 simply shows that if a negative energy solution is to remain nonsingular, one must keep the initial actions within certain limits.

Of course, one is not assured that these one-lump solutions are typical of 3D interactions since the solution is, in a sense, quite special. However, comparing these solutions with 1D solutions, with the linear limit, and with what one expects to occur in 3D, one is lead to strongly suspect that these 1-lump solutions do indeed, at least qualitatively, represent the typical 3D-3WRI. However, at the moment, this is only a conjecture, which shall have to wait for further results for verification.

REFERENCES

Ablowitz, M.J., Kaup, D.J., Newell, A.C., and Segur, H., 1974, The inverse scattering transform - Fourier analysis for nonlinear problems, Stud. Appl. Math., 53:249.

Ablowitz, M.J., and Haberman, R., 1975, Resonantly coupled nonlinear evolution equations, J. Math. Phys., 16:2301.

Armstrong, J.A., Bloembergen, N., Ducuing, J., and Pershan, P.S., 1962, Interactions between light waves in a nonlinear dielectric, Phys. Rev., 127:1918.

Cornille, H., 1979, Solutions of the nonlinear 3-wave equations in three spatial dimensions, J. Math. Phys., 20:1653.

Corones, J., 1976, Solitons and simple pseudopotentials, J. Math. Phys., 17:756.

Kaup, D.J., 1977, Coherent pulse propagation: a comparison of the complete solution with the McCall-Hahn theory and others, Phys. Rev., A16:704.

Kaup, D.J., and Newell, A.C., 1978a, Theory of nonlinear oscillating dipolar excitations in one-dimensional condensates, Phys. Rev., B18:5162.

Kaup, D.J., and Newell, A.C., 1978b, Solitons as particles, oscillators, and in slowly changing media: a singular perturbation theory, Proc. Roy. Soc. London, A361:413.

Kaup, D.J., 1978, Simple harmonic generation: an exact method of solution, Stud. Appl. Math., 59:25.

Kaup, D.J., Rieman, A., and Bers, A., 1979, Space-time evolution of nonlinear three-wave interactions. I. Interaction in a homogeneous medium, Rev. Mod. Phys., 51:275.

Kaup, D.J., 1979, to appear in the Proceedings of the Soviet-American Soliton Symposium, Kiev, USSR, 2-15 September 1979, (Physica D, 1981).

Kaup, D.J., 1980a, A method for solving the separable initial-value problem of the full three-dimensional three-wave interaction, Stud. Appl. Math., 62:75.

Kaup, D.J., 1980b, The inverse scattering solution for the full three-dimensional three-wave resonant interaction, Physica D, 1:45.

Kaup, D.J., 1980c, Determining the final profiles from the initial profiles for the full three-dimensional three-wave resonant interaction, in: "Mathematical Methods and Applications of Scattering Theory," DeSanto, Sáenz and Zachary, eds., Lecture Notes in Physics 130, Springer-Verlag, New York.

Kaup, D.J., 1981, to appear, The Backlünd transformation and N-lump solutions of the three-dimensional three-wave resonant interaction, J. Math. Phys.

Zakharov, V.E., and Manakov, S.V., 1973, Resonant interaction of wave packets in nonlinear media, ZhETF Pis. Red., 18:413. [JETP Letts., 18:243.]

Zakharov, V.E., and Shabat, A.B., 1974, A scheme for integrating the nonlinear equations of mathematical physics by the method of inverse scattering problem. I., Funkts. Anal. Prilozh., 8:43.

Zakharov, V.E., 1976, Exact solutions to the problem of the parametric interaction of three-dimensional wave packets, Dokl. Akad. Nauk. SSSR, 228:1314. [Sov. Phys. Dokl., 21:322.]

1.1.

Commun. math. Phys. 76, 65–116 (1980)

Communications in Mathematical Physics
© by Springer-Verlag 1980

Monodromy- and Spectrum-Preserving Deformations I

Hermann Flaschka* and Alan C. Newell

Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam N. Y. 13676 USA

Abstract. A method for solving certain nonlinear ordinary and partial differential equations is developed. The central idea is to study monodromy preserving deformations of linear ordinary differential equations with regular and irregular singular points. The connections with isospectral deformations and with classical and recent work on monodromy preserving deformations are discussed. Specific new results include the reduction of the general initial value problem for the Painlevé equations of the second type and a special case of the third type to a system of linear singular integral equations. Several classes of solutions are discussed, and in particular the general expression for rational solutions for the second Painlevé equation family is shown to be $-\frac{d}{dx}\ln(\Delta_+/\Delta_-)$, where Δ_+ and Δ_- are determinants. We also demonstrate that each of these equations is an exactly integrable Hamiltonian system. The basic ideas presented here are applicable to a broad class of ordinary and partial differential equations; additional results will be presented in a sequence of future papers.

1. Introduction and Outline

This paper is the first in what is planned to be a series of studies on deformations of linear ordinary differential equations with coefficients rational on a Riemann surface. The deformations in question preserve the monodromy at singular points of the linear equation, and this requirement forces the coefficients of the linear equation to satisfy certain nonlinear ordinary or partial differential equations of considerable interest. The theory of monodromy-preserving deformations overlaps the theory of isospectral deformations (i.e., soliton theory), and indeed one of our aims will be to understand the connections between these two types of problems. Applications of the nonlinear equations governing monodromy-preserving deformations have been discovered in nonlinear waves, statistical

* Permanent address: Mathematics Department, University of Arizona, Tucson, AZ 85721 USA

0010-3616/80/0076/0065/$10.40

mechanics and quantum field theory; we shall present some new results along these lines, and we intend to develop relations between soliton ideas and the applications mentioned above.

The next section of this paper will describe in considerable detail the history of relevant work on monodromy-preserving deformations, as well as connections between different approaches and problems. The paper itself deals primarily with two representative examples. The first of these is the system

$$\begin{aligned} v_{1\zeta} &= -i(4\zeta^2 + x + 2q^2)v_1 + (4\zeta q - 2ir + \nu/\zeta)v_2 \\ v_{2\zeta} &= (4\zeta q + 2ir + \nu/\zeta)v_1 + i(4\zeta^2 + x + 2q^2)v_2 , \end{aligned} \tag{1.1}$$

with x, q, r, ν constant in ζ. The singular points of (1.1) are at $\zeta = \infty$ (irregular) and $\zeta = 0$ (regular) if $\nu \neq 0$. A deformation of (1.1) is monodromy-preserving if the Stokes multipliers associated with formal solutions about $\zeta = \infty$, the monodromy matrix about $\zeta = 0$, and the matrix connecting fundamental solutions at $\zeta = 0$ and ∞ are unchanged (when x is varied, and q, r change as functions of x). For this it is necessary and sufficient that (as function of x) $r = q_x$, and

$$q_{xx} = 2q^3 + xq - \nu . \tag{1.2}$$

(1.2) is the *second Painlevé equation* [1]. By posing an irregular-singular Riemann-Hilbert problem, we exploit the connection with (1.1) to reduce (1.2) to an equivalent system of *linear* singular integral equations. In a special case, this reproduces the solution, due to Ablowitz and Segur, of (1.2) by a Marchenko integral equation [2]. Other special cases yield the rational or Airy-function solutions of (1.2) discovered by Airault [3], but this time by procedures very familiar from the inverse-scattering derivation of multisoliton formulae.

The second and somewhat more complicated example, whose study we begin in this paper, is afforded by the system

$$\begin{aligned} v_{1\zeta} &= \left(-ix + \frac{i}{\zeta^2}\cosh u\right)v_1 + \left(-\frac{xu_x}{2\zeta} + \frac{i}{\zeta^2}\sinh u\right)v_2 , \\ v_{2\zeta} &= \left(-\frac{xu_x}{2\zeta} - \frac{i}{\zeta^2}\sinh u\right)v_1 + \left(ix - \frac{i}{\zeta^2}\cosh u\right)v_2 . \end{aligned} \tag{1.3}$$

This system has irregular singular points at $\zeta = 0, \infty$. The deformation equation is equivalent to a special case of the third Painlevé equation,

$$(xu_x)_x = -4\sinh u . \tag{1.4}$$

This equation and its linearization via the Riemann-Hilbert problem for (1.3) provide a link between the inverse scattering transform and the extensive work of Barouch–McCoy–Tracy–Wu [4–6] on the Ising model and of Sato et al. [7] on monodromy-preserving deformations and quantum field theory.

Our principal aims in this first paper are:

1. To point out relations between mondromy – and spectrum – preserving deformations.

2. To introduce a new method for solving the initial value problem for equations such as (1.2) and (1.4) which can be written as monodromy preserving deformations.

3. To point out that both (1.2) and (1.4) are exactly integrable Hamiltonian systems.

4. To discuss the connections between our method, the inverse scattering transform and the novel ideas of Krichever and Novikov for investigating multiperiodic solutions of soliton equations.

Following a general discussion in Sect. 2, in Sects. 3 and 4 we describe in detail the mapping from the coefficients of the differential equations (1.1) and (1.3) to the monodromy data and prove that this data is independent of x. In each case we also derive the inverse mapping which allows one to reconstruct the solutions to (1.1) and (1.3) (and therefore both the equations they satisfy and the coefficients in these equations, the quantities of interest) from a knowledge of the monodromy data. The result appears as a set of coupled linear singular integral equations. We examine several limiting cases and derive some useful formulae for special classes of solutions, although we have not, as yet, been able to prove the existence and uniqueness of solutions in the general case.

2. Introduction and Discussion

2A. Deformations

Nonlinear ordinary and partial differential equations do not, in general, admit explicit solutions, because the solutions of the typical nonlinear equation are so wildly irregular that they could not possibly be represented by known functions. Conversely, nonlinear equations with very well-behaved solutions should be expected to have uncommon properties. One feature shared by many such special nonlinear equations was discovered towards the end of the 19th century, was exploited for about forty years and then (apparently) forgotten; quite recently it has re-emerged in a somewhat different form. This is the observation that "solvable" nonlinear equations arise as *integrability conditions for certain kinds of deformations of linear equations.*

The most recent version of this technique centers on the idea of *isospectral deformation.* The best known and most celebrated example involves the Schrödinger operator [8] $L=-\frac{d^2}{dx^2}+q(x)$. One asks, how can one deform the coefficient $q(x)$ as a function of an additional parameter t, so that the eigenvalues of L [as operator on $L^2(R)$, say] do not change? The simplest nontrivial such deformation is already one of great physical importance: the deformation is isospectral if $q(x,t)$ satisfies the *Korteweg-de Vries equation,*

$$q_t-6qq_x+q_{xxx}=0. \tag{2.1}$$

In recent years, studies of isospectral deformations have uncovered many more nonlinear equations of physical relevance and mathematical interest; the whole subject appears to be deeply involved with Lie groups and differential and algebraic geometry.

Another kind of deformation, perhaps not as widely known, is associated with the monodromy group of Fuchsian differential equations. Consider a (matrix)

system of ordinary differential equations,

$$Y' = Y \sum_{j=1}^{n} \frac{A^j}{\zeta - a_j}, \tag{2.2}$$

where the A^j are constant $m \times m$ matrices. The fundamental solution of (2.2) is (in general) a multi-valued function of complex ζ. If ζ moves on a path $|\zeta - a_j| = \varepsilon$ encircling a_j, the solution $Y(a_j + \varepsilon)$ will change into $Y(a_j + \varepsilon e^{2\pi i}) = \tilde{Y}$ which is a matrix whose rows are linear combinations of the rows of $Y(\zeta)$,

$$\tilde{Y}(\zeta) = M_j Y(\zeta). \tag{2.3}$$

M_j is the *monodromy matrix at* a_j. The deformation problem is: how can one change the A^j, as function of the poles a_j, so that the monodromy matrices M_j remain fixed? The simplest nontrivial example is again of considerable interest. The linear Eq. (2.2) is taken to be a 2×2 system, with three poles fixed at $0, 1, \infty$, and one pole τ subject to variation. *A priori*, there are twelve adjustable entries in the coefficient matrices $A^j (j = 1, 2, 3)$, but they can all be expressed in terms of a single function $z(\tau)$ which satisfies the equation [9, 18]

$$z'' + \left[\frac{1}{\tau} + \frac{1}{\tau - 1} + \frac{1}{z - \tau}\right] z' - \frac{1}{2}\left[\frac{1}{z} + \frac{1}{z - 1} + \frac{1}{z - \tau}\right](z')^2 - 2\frac{z(z-1)(z-\tau)}{\tau^2(\tau-1)^2}$$
$$\cdot \left\{\alpha - \beta\frac{\tau}{z^2} + \gamma\frac{\tau - 1}{(z-1)^2} - \delta\frac{\tau(\tau-1)}{(z-\tau)^2}\right\} = 0. \tag{2.4}$$

The frightening Eq. (2.4) is the most general 2nd order equation

$$z'' = R(\tau, z, z')$$

with R rational in z, z' and analytic in τ, which has the property: the location of any *algebraic, logarithmic, or essential singularity of its solutions is independent of the initial conditions.*

Equations with this property were studied in exhaustive (and exhausting) detail by Painlevé and Gambier [1]. There are fifty canonical types, which include linear equations such as $z'' = z$, equations solved by elliptic functions, such as

$$z'' = 2z^3 + cz - \nu, \tag{2.5}$$

and *six* equation types whose general solutions can be proved not to be expressible in terms of the basic special functions (except for isolated cases, see Sect. 3 below). These six equations are called *Painlevé equations*, and their solutions *Painlevé transcendents*. These equations are, as the summary above indicates, distinguished among non-classical ordinary differential equations of the form (2.4) in that the nonpolar singularities of their solutions can be predicted from the equation alone. Equations (1.2) and (1.4), which we study in this paper, fall into this class [(1.4) only after the change of variables $f = e^u$]. We shall describe later some of the important applications of the Painlevé transcendents.

A third kind of deformation involves the properties of solutions of ordinary differential equations near *irregular singular points*. This, indeed, will be the main concern of the present paper, and for the moment we provide a brief description

only. Consider the system

$$Y' = A(\zeta)\, Y, \tag{2.6}$$

where the (matrix) function A is polynomial in ζ. $\zeta = \infty$ is an irregular singular point, and typically only formal solutions

$$\tilde{Y}(\zeta) = e^{P(\zeta)} \zeta^{L} \{1 + c_1 \zeta^{-1} + c_2 \zeta^{-2} + \ldots\}, \tag{2.7}$$

are available near ∞. A formal series such as (2.7) is, in fact, asymptotic to a true solution Y in certain sectors S in the complex ζ-plane, and one may have

$$Y_j \sim \tilde{Y} \quad \text{in} \quad S_j$$

$$Y_k \sim \tilde{Y} \quad \text{in} \quad S_k$$

for different true solutions Y_j, Y_k in different sectors S_j, S_k. The discontinuity of asymptotic expansions is known as the *Stokes phenomenon*, and the matrices which connect the different true solutions with fixed asymptotic expansions in the various sectors are called *Stokes multipliers*, $Y_j = Y_k M_{kj}$. The deformation problem we pose is: *to change the coefficients in* (2.6) *so that the Stokes multipliers* M_{kj} *remain constant.*

The immediate object of our paper is to study this deformation problem for a particular 2×2 system which leads to the second Painlevé equation

$$q'' = 2q^3 + xq - \nu \tag{2.8}$$

as integrability condition. (2.8) is a nonautonomous version of the elliptic function Eq. (2.5). We will reduce this nonlinear ordinary differential equation to a system of linear integral equations, and in the process we will recover some known special solutions. There are, however, deep and entirely unexplored connections amongst these various types of deformation problems, and equally interesting relations between monodromy-preserving deformations and questions in statistical mechanics, quantum field theory, and wave dynamics. We plan to address some of these topics in later papers, and want to detail in this overall introduction the ingredients of what we think will eventually become a beautiful and useful complement to current soliton theory.

2B. Applications

Painlevé transcendents are encountered in several important physical problems, of which we describe two. In one of these, the connection with solitons (more precisely, with isospectral deformations) is evident. In the other, there are certain analogies with soliton theory which first stimulated our interest. We now outline the relevant facts.

1. Self-Similar Solutions of Wave Equations. Although the following considerations apply to many soliton equations, we restrict ourselves to the modified Korteweg-de Vries (MKdV) equation, in the form

$$q_t - 6q^2 q_x + q_{xxx} = 0. \tag{2.9}$$

If $q(x,t)$ solves (2.9), then so does

$$\tilde{q}(x,t)=\beta q(\beta x,\beta^3 t)$$

for any β. A solution which is invariant under this scaling is called *self-similar*: $q(x,t)=\tilde{q}(x,t)$. It follows that for such q,

$$q(x,t)=(3t)^{-1/3}f(x(3t)^{-1/3}). \tag{2.10}$$

The function $f(\xi)$ satisfies

$$f''=2f^3+\xi f-\nu \tag{2.11}$$

(ν is an arbitrary integration constant), as can be seen by substitution of (2.10) into (2.9). (2.11) is the second Painlevé equation. Now, it is known that MKdV can be integrated by the inverse-scattering problem for

$$\begin{aligned} v_{1x}+i\zeta v_1&=qv_2\\ v_{2x}-i\zeta v_2&=qv_1 \end{aligned} \tag{2.12}$$

with the t-evolution of the eigenfunctions governed by

$$\begin{aligned} v_{1t}&=(-4i\zeta^3-2iq^2\zeta)v_1+(4\zeta^2q+2i\zeta q_x-q_{xx}+2q^3)v_2,\\ v_{2t}&=(4\zeta^2q-2i\zeta q_x-q_{xx}+2q^3)v_1+(4i\zeta^3+2iq^2\zeta)v_2. \end{aligned} \tag{2.13}$$

For the self-similar solution (2.10), the t equation is in a sense redundant, since $q(x,t)$ is known for all t once it is given for $t=1/3$, say $(q(x,1/3)=f(x))$. This observation can be put to use in two ways.

Ablowitz and Segur [10] in their analysis of the asymptotic behavior of solutions of the KdV equation were led to the following procedure (described here for MKdV). Apply the usual inverse method for MKdV up to the Marchenko equation by which $q(x,t)$ is determined from the scattering data. At that stage, assume $q(x,t)$ to be self-similar, and observe that t can be scaled out of the Marchenko equation altogether. This leads to the Fredholm equation,

$$\begin{aligned} K_1(x,y)&=\varrho\,\mathrm{Ai}\left(\frac{x+y}{2}\right)-\frac{1}{2}\varrho\int_x^\infty K_2(x,s)\,\mathrm{Ai}\left(\frac{s+y}{2}\right)ds\\ K_2(x,y)&=-\frac{1}{2}\varrho\int_x^\infty K_1(x,s)\,\mathrm{Ai}\left(\frac{s+y}{2}\right)ds \end{aligned} \tag{2.14}$$

and one recovers the Painlevé transcendent $f(x)$ by

$$f(x)=K_1(x,x). \tag{2.15}$$

It is a consequence of the scaling invariance of the Marchenko equation that the reflection coefficient $r(\zeta)$ has the special form

$$r(\zeta)=\varrho e^{8i\zeta^3/3}. \tag{2.16}$$

The Fourier transform of $R(\zeta)$ is the kernel of (2.14), which explains the occurrence of the Airy function. This method produces the unique solution of (2.11) which satisfies

$$f(x)\sim\varrho\,\mathrm{Ai}(x),\qquad x\to+\infty$$

(Hastings and McLeod [11]). More recently, Ablowitz et al. [12] have shown how to bypass the t-evolution in linearizing (2.11) and similar equations for solutions which decay at $+\infty$. This is more satisfactory, since scattering theory is not really applicable for potentials such as $f(x)$ which have poles, or at least decay slowly at $+$ or $-\infty$.

It is also possible to use the scaling invariance directly on the t-equation (2.13). If $v(x,t,\zeta)$ solves (2.12), (2.13), then so does $v(\beta x, \beta^3 t, \beta^{-1}\zeta)$ for the potential $\tilde{q}$. Define, in the case of self-similar q,

$$w(x,\zeta) = v(x, \tfrac{1}{3}, \zeta);$$

then

$$v(x,t,\zeta) = w(x(3t)^{-1/3}, \zeta(3t)^{1/3}).$$

Hence

$$\frac{\partial v}{\partial t} = -(3t)^{-4/3}\frac{\partial w}{\partial x} + (3t)^{-2/3}\zeta\frac{\partial w}{\partial \zeta},$$

and upon using the x-equation (2.12) to eliminate $\dfrac{\partial w}{\partial x}$, one can rewrite (2.13) as follows [we replace $x(3t)^{-1/3}$ by x, and $\zeta(3t)^{1/3}$ by ζ]:

$$\begin{aligned} w_{1\zeta} &= -i(4\zeta^2 + x + 2f^2)w_1 + (4\zeta f + 2if')w_2 \\ w_{2\zeta} &= (4\zeta f - 2if')w_1 + i(4\zeta^2 + x + 2f^2)w_2 . \end{aligned} \tag{2.17}$$

This is coupled to (2.12), rewritten now without use of t:

$$\begin{aligned} w_{1x} + i\zeta w_1 &= f w_2 \\ w_{2x} - i\zeta w_2 &= f w_1 . \end{aligned} \tag{2.18}$$

If one imagines $f(x)$ to decay so rapidly at $\pm\infty$ that scattering theory can be applied, one looks at a solution of (2.18) which satisfies

$$w(x,\zeta) \sim \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta x}, \qquad x \to -\infty .$$

At $+\infty$,

$$w(x,\zeta) \sim a(\zeta)\begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta x} + b(\zeta)\begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta x}$$

and it follows from (2.17) that

$$r(\zeta) = \frac{b(\zeta)}{a(\zeta)} = \varrho e^{8i\zeta^3/3}.$$

Inverse scattering then reproduces the results of Ablowitz and Segur. There is a conceptual question, which leads to the main point of our paper. In MKdV theory, the x-equation (2.12) is basic. It is deformed in t in a special way; namely, so that the transmission coefficient $a(\zeta)^{-1}$ is independent of t. The t-evolution (2.13) is one possible expression of this requirement, and MKdV is the integrability condition

for these two systems. For self-similar solutions, however, there is no t in (2.18), so what is being deformed? As we explain in this paper, it is now the ζ-equation (2.17) (alias t-evolution) which is basic. Vary x in (2.17), and change the remaining coefficients as functions of x so that the *Stokes multipliers of* (2.17) *are independent of* x. *The original* x*-equation* (2.18) *describes the appropriate evolution of the solutions of* (2.17) *in* x, *and the Painlevé equation* (2.11) *with* $\nu=0$ *is the integrability condition.* In this way an isospectral deformation leads, for self-similar solutions, to monodromy-preserving deformations associated with the irregular singular point.

We also want to point out that whereas the self-similar solutions of integrable evolution equations appear to give solvable nonautonomous ordinary differential equations, there are many members of the latter class which do not result from self-similar limits of the former. (See Example 2, Appendix I.)

II. Correlation Functions of the Ising Model. A most remarkable occurrence of Painlevé transcendents was discovered by Barouch et al. [4] and was incorporated into a powerful general framework by Sato et al. [7]. The interest here is in computing the k-point correlation functions of the rectangular Ising model in the scaling limit. First we recall some terminology about the Ising model. A spin variable $\sigma_{mn}=\pm 1$ is attached to each point of an $M\times N$ lattice. The energy $E(\sigma)$ corresponding to a configuration $\sigma=\{\sigma_{mn}\}$ is

$$E(\sigma)=-J_1\sum_{m=1}^{N}\sum_{n=1}^{N}\sigma_{mn}\sigma_{m+1,n}-J_2\sum_{m=1}^{M}\sum_{n=1}^{N}\sigma_{mn}\sigma_{m,n+1}$$

(periodic boundary conditions). The partition function is

$$Z_{MN}=\sum_{\sigma}e^{-\beta E(\sigma)},\qquad \beta=\frac{1}{kT},$$

(summed over all possible σ), and the k-point correlation functions are

$$\varrho_k((m_1,n_1),\ldots,(m_k,n_k))=Z_{MN}^{-1}\sum_{\sigma}\sigma_{m_1n_1}\ldots\sigma_{m_kn_k}e^{-\beta E(\sigma)}.$$

One is interested in obtaining explicit formulae for these quantities as $M,N\to\infty$. It is well known that the partition function is not analytic at some critical β_c, corresponding to a critical temperature T_c. The *scaling limit* is a continuum limit of the lattice as $T\to T_c^{\pm}$ [13]. The 2-point correlation functions and their scaling limits were first evaluated and studied in [4]. Of particular interest for us is the fact that the 2-point functions in the scaling limit admit *closed expressions in terms of solutions of the third Painlevé equation* (see the survey by Tracy [13]). Without invoking any deformation ideas, Wu et al. derived series expansions for the correlation functions; certain of these expansions for $k=2$ are equivalent to the Neumann series subsequently studied by Ablowitz et al. [12], and identical to iterative solutions obtained from our singular integral equations (see Sect. 4 below).

The deformation idea was introduced into this circle of problems by Sato et al. [7]. Their work presents a remarkable synthesis of apparently unrelated fields: the theory of rotations in Clifford algebras, monodromy-preserving deformations, and

quantum field theory. Certain of their discoveries show a particularly strong similarity to facts familiar from inverse-scattering theory and lead directly into the investigations of the present paper.

We present a brief sketch of the basic deformation problem of [7] (algebraic details of the connection are given in Appendix III). [7] deals with solutions of the two-dimensional Dirac equation for $w=\begin{pmatrix} w_+ \\ w_- \end{pmatrix}$,

$$\frac{\partial}{\partial z} w_- = m w_+, \quad \frac{\partial}{\partial \bar{z}} w_+ = m w_-, \quad m>0, \tag{2.19}$$

which: are multivalued, having root type branch points at $2n$ points $a_1, \ldots, a_k$, $\bar{a}_1, \ldots, \bar{a}_k$,

$$\begin{aligned} &w_\pm(a_\nu + e^{2\pi i}(z-a_\nu), \bar{a}_\nu + e^{-2\pi i}(\bar{z}-\bar{a}_\nu)) \\ &\quad = -e^{-2\pi i l_\nu} w(z,\bar{z}), \end{aligned} \tag{2.20}$$

satisfy certain growth conditions at $(a_\nu, \bar{a}_\nu)$, $\nu=1, \ldots, k$ and decay as $e^{-2m|z|}$ as $|z|\to\infty$. For $n=1$, the only solution is a modified Bessel function of the second kind. In general, the space W^k of such solutions is k dimensional [7]. Let

$$\mathbf{W}=\begin{pmatrix} w^{(1)} \\ \vdots \\ w^{(k)} \end{pmatrix}, \qquad w^{(j)}=(w_+^{(j)}, w_-^{(j)})^T \tag{2.21}$$

be a basis (appropriately normalized) of W^k. There is a differential equation satisfied by $\mathbf{W}$,

$$M_F \mathbf{W} = \left(B\frac{\partial}{\partial z} + \bar{B}\frac{\partial}{\partial \bar{z}} + E \right) \mathbf{W}, \tag{2.22}$$

where $M_F = z\frac{\partial}{\partial z} - \bar{z}\frac{\partial}{\partial z} + \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and B, $\bar{B}$, E are constant matrices which depend on l_ν, a_ν and $\bar{a}_\nu$, $\nu=1, \ldots, k$. Equation (2.22) together with (2.19) completely characterizes the space W^k. One now asks: *how do the matrices in* (2.22) *change as functions of the branch points* $a_j, \bar{a}_j$? Because these matrices characterize W^k, which in turn is the space of solutions determined by the monodromy requirement (2.20), this question is about deformations of (2.19), (2.22) preserving the root monodromy and growth conditions. The solutions of the deformation equations lead to closed expressions for the k-point functions; for $k=2$, these reduce to the formulae mentioned above involving the third Painlevé transcendent. In [7], Eq. (3.3.39), Satō et al. point out that a formal Laplace transform

$$\mathbf{W}(z,\bar{z}) = \int \frac{du}{2\pi u} \begin{pmatrix} \sqrt{u} \\ \sqrt{u^{-1}} \end{pmatrix} e^{m(zu+\bar{z}u^{-1})} \hat{w}(u)$$

converts (2.19), (2.22) into

$$\left(u\frac{d}{du} + mAu - G^{-1} m \tilde{A} G u^{-1} + F \right) \hat{w}(u) = 0. \tag{2.23}$$

This again produces a system of ordinary differential equations with polynomial coefficients, with irregular singular points at $u=0,\infty$. These equations involve the a_j, $\bar{a}_j$ and other parameters. We may ask: how should these other parameters change as functions of a_j, $\bar{a}_j$ if the Stokes multipliers of (2.23) are to remain constant? For $k=1$, the system of deformation equations is trivial, and for $k=2$, one recovers the sinh-Gordon similarity solution (i.e. a special case of Painlevé III). The exact connections are given in Appendices I and III.

2C. *Classical Work*

The work of Satō et al. was in part stimulated by, and then re-applied to, some classical problems of deformation theory. We have already mentioned the work of Fuchs [9] on Painlevé VI as a deformation equation. The general system of deformation equations for (2.2) was derived by Schlesinger [14]:

$$\frac{\partial A^j}{\partial a_i} = \frac{[A^i, A^j]}{a_j - a_i}, \tag{2.24}$$

$$\sum_{j=1}^{n} \frac{\partial A^j}{\partial a_i} = 0.$$

[7, Part II] details a constructive solution method for (2.24) based on a specialization of the deformation theory of the Dirac equation (Sect. B, above). Equation (2.24), of course, relates to monodromy groups of systems whose singular points are all regular. The question of irregular singular points was also taken up, by Garnier [15], but (as far as we can tell) on a purely formal level, by analogy with results of Fuchs on the regular-singular case. Garnier considers, in connection with Painlevé II, the 2nd order equation for $y(\zeta)$,

$$y_{\zeta\zeta} = \left[\sum_0^m a_j \zeta^j + \sum_0^n \left(\frac{3/4}{(\zeta-\lambda_k)^2} + \frac{\varrho_k}{\zeta-\lambda_k}\right)\right] y. \tag{2.25}$$

He then asks: how can one complement (2.25) by an equation

$$y_x = Ay + By_\zeta \tag{2.26}$$

with A, B rational in ζ, so that (2.25), (2.26) is an integrable system (in the sense: $y_{x\zeta\zeta} = y_{\zeta\zeta x}$)? If $n=0$ in (2.25) and $\lambda_0 = \lambda$, it turns out that one possible choice for (2.26) leads to Painlevé II as the integrability condition,

$$\lambda'' = 2\lambda^3 + x\lambda - \nu.$$

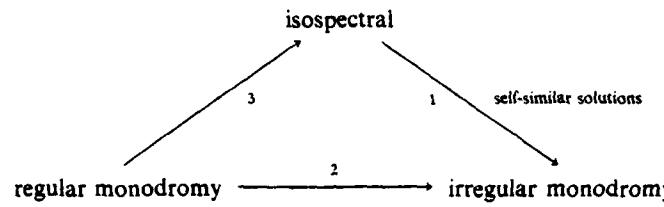
There has apparently been no discussion, so far, of the deformation theory underlying (2.25), (2.26), even though the work of Birkhoff [16, 17] presents most of the necessary ideas relating to the monodromy concept for an irregular singular point[1]. Starting from the requirement that the Stokes multipliers of (2.25) [or rather, of (1.1)] be independent of x, we will prove (in Sect. 3) that an equation like (2.26), with rational A, B, must be satisfied.

---

1 After this work was completed, we learned of work of K. Ueno on this question (see Sect. 5)

2D. Further Connections Between Isospectral and Monodromy-Preserving Deformations

Of the three deformation problems listed in Sect. 2A, only two have been shown to be related up to this point of our discussion.



It is probably not surprising that connection 2 can be established. If one coalesces two or more regular singular points of an ordinary differential equation, one expects to get an irregular one. Indeed, Garnier [15] states (without details) that an equation such as (2.25) is obtained from one with all singular points regular by confluence, with the corresponding deformation equations passing into each other. This amounts to the observation of Painlevé (see, e.g. [1]) that the 6th equation, (2.4), (which is the integrability condition for a regular monodromy-preserving deformation) yields all other Painlevé equations I–V (associated with irregular monodromy) by appropriate limiting procedures. Yet, the geometric content of this statement is still very obscure.

Much more surprising is connection 3, again due in large part to Garnier [18]. It is fairly well known (Davis [19]) that the change of variables

$$f=\sqrt{x}\,g, \quad t=\tfrac{2}{3}x^{3/2}$$

converts Painlevé II into

$$g''=2g^3-2g-\frac{1}{t}g'+\frac{1}{9}\frac{g}{t^2}-\frac{2\nu}{3t}$$

which, for $|t|\to\infty$ in certain regions of the t plane reduces to

$$g''=2g^3-2g.$$

This is solvable in terms of elliptic functions. Indeed, the solutions of Painlevé II are "asymptotically elliptic" (much as Bessel functions are asymptotically trigonometric). Garnier takes the general Schlesinger system (2.24), lets the $a_j\to\infty$ so that $\frac{a_i}{a_j}\to\frac{\alpha_i}{\alpha_j}$ (fixed) and sets $\alpha_j\log\alpha_j=\tau_j$; then (2.24) formally becomes

$$\frac{\partial A^j}{\partial\tau_i}=\frac{[A^i,A^j]}{\alpha_j-\alpha_i}, \qquad \sum_{j=1}^{n}\frac{\partial A^j}{\partial\tau_i}=0. \tag{2.27}$$

This is now an autonomous system, which Garnier integrates in terms of Abelian functions. It is remarkable that his basic lemma says that the eigenvalues of the matrix

$$\prod_{i=1}^{n}(\zeta-\alpha_i)\sum_{j=1}^{n}\frac{A^j}{\zeta-\alpha_j}$$

are unchanged when the A^j solve (2.27) – a typical isospectral flow. One special case singled out (and solved) by Garnier is the system

$$\begin{aligned}\xi_i'' &= -\lambda_i\xi_i+\left(2\sum_1^n\xi_i\eta_j\right)\xi_i\\ \eta_i'' &= -\lambda_i\eta_i+\left(2\sum_1^n\xi_j\eta_j\right)\eta_i\end{aligned}\tag{2.28}$$

which is now known [20, 21] to contain the finite-gap KdV theory as special case. It was, in fact, discovered some 60 years earlier by C. Neumann as describing uncoupled harmonic oscillators constrained to move on a sphere.

Garnier's system (2.27) covers a large class of the integrable periodic problems solvable by inverse spectral methods. There are other known periodic isospectral flows contained in (2.27). We will discuss these in the next paper, in which we will also re-interpret Garnier's method in the language of isospectral flows.

3. The General Solution of Painlevé II

3A. Outline

We have already mentioned (and full details are given in Appendix I) that $r=q_x$ and

$$q_{xx}=2q^3+xq-\nu \tag{3.1}$$

are the integrability conditions of the Eqs. (3.2) and (3.3) below

$$v_{1\zeta}=(-4i\zeta^2-i(x+2q^2))v_1+\left(4\zeta q+\frac{\nu}{\zeta}+2ir\right)v_2, \tag{3.2a}$$

$$v_{2\zeta}=\left(4\zeta q+\frac{\nu}{\zeta}-2ir\right)v_1+(4i\zeta^2+i(x+2q^2))v_2, \tag{3.2b}$$

$$v_{1x}=-i\zeta v_1+qv_2, \tag{3.3a}$$

$$v_{2x}=qv_1+i\zeta v_2. \tag{3.3b}$$

The method of isospectral deformation, or IST, concentrates its principal attention on (3.3). In order to implement the method, one must have some information on q as function of x; for example, that it decays to zero or a constant as $x\to\pm\infty$ and also that certain moments exist. If

$$\int_{-\infty}^{\infty}(1+|x|)|q|dx<\infty, \tag{3.4}$$

one can define the fundamental solution matrices $\hat{\Phi}(x,\zeta)$ and $\hat{\Psi}(x,\zeta)$ by the asymptotic properties

$$\hat{\Phi}(x,\zeta)\to\hat{E}(x,\zeta)=\begin{pmatrix} e^{-i\zeta x} & 0 \\ 0 & e^{i\zeta x}\end{pmatrix} \quad \text{as} \quad x\to-\infty \tag{3.5a}$$

and

$$\hat{\Psi}(x,\zeta)\to\hat{E}(x,\zeta) \quad \text{as} \quad x\to+\infty\,. \tag{3.5b}$$

Condition (3.4) ensures that certain analyticity properties hold, and in particular that the scattering matrix $A(\zeta)=\hat{\Phi}^{-1}\hat{\Psi}$ is defined with its diagonal entries admitting analytic extension. The ζ behavior of $A(\zeta)$ (or the (ζ,t) behavior of $A(\zeta,t)$ – see [33]) is inferred from (3.2)

For the class of equations and solutions we wish to discuss, condition (3.4) does not hold. Therefore, we propose a new method, in which one focuses central attention on Eq. (3.2) and uses (3.3) as an auxiliary equation. We note straightaway that (3.2) is much simpler: the coefficients are polynomials in the independent variable ζ. The points $\zeta=0$ and $\zeta=\infty$ are regular and irregular singular points of the equation, respectively, and the solution matrix is a meromorphic function of ζ on an appropriate Riemann surface. If ν is an integer, the solution matrix is meromorphic in the finite complex ζ plane; otherwise, one must introduce the multisheeted *Riemann surface of* ζ^{ν}.

The steps in the method exactly parallel the steps used in IST. First, at a given value of x, where q and q_x are given, one determines various properties of the solution matrix connected with the singular points $\zeta=0$ and $\zeta=\infty$ of the equation. Around the singular point $\zeta=\infty$, one has the Stokes phenomenon: the analytic continuation of a solution from one sector to another does not have as its asymptotic expansion, as $\zeta\to\infty$ in the new sector, the analytic continuation of the asymptotic expansion in the first sector. If one identifies a solution matrix in each of the sectors abutting infinity by a fixed asymptotic behavior, then these solutions will not evolve from one sector to another in a continuous fashion, but will be connected by *Stokes multiplier matrices*. The entries of these Stokes multiplier matrices are the Stokes multipliers and are part of the characteristic data of the singular point. The other data needed at $\zeta=\infty$ are the coefficients of the polynomial in the exponent of the formal asymptotic expansion of the fundamental solution matrix. *If the rank of the irregular singular point is r*, then the components of the solution vector will have asymptotic expansions of the form

$$\exp\left\{\sum_{i=1}^{r}\omega_i\zeta^i\right\}\zeta^l\left\{1+O\left(\frac{1}{\zeta}\right)\right\}. \tag{3.6}$$

In the case under discussion, $r=3$, $\omega_3=4i/3$, $\omega_2=0$, $\omega_1=ix$, $l=0$. One also needs to know how the solution changes as the point $\zeta=\infty$ is encircled; does it return to the value in the first sector? This can be determined by examining the behavior of the solution in the neighborhood of the regular singular point. Thus, one introduces the monodromy matrix at $\zeta=0$. Finally, one must specify the

connection matrix between canonical fundamental solution matrices at $\zeta=0$ and $\zeta=\infty$. The characteristic parameters at $\zeta=\infty$ (the Stokes multiplier matrices and $\{\omega_j\}$, l), the monodromy matrix at $\zeta=0$, and the connection matrix are together called the *transform data.*

The second step in the method is crucial. We ask: how do the transform data evolve as the parameters ω_i in (3.6) change? From (3.3), we will show that their evolution is trivial: they are constant.

Therefore, the third step is to reconstruct the coefficients r and q in (3.2) and hence q and q_x at any x. This is achieved by deriving linear singular integral equations for the columns of the fundamental solution matrix.

We now discuss these steps in detail in connection with the system (3.1)–(3.3).

3B. The Direct Transform

A formal asymptotic analysis of (3.2) at $\zeta=\infty$ will show that the two linearly independent solutions have the expansions

$$\psi^{(1)}(\zeta,x)\sim\tilde{\psi}^{(1)}=e^{-4i\zeta^3/3-i\zeta x}\begin{pmatrix}1-\dfrac{i}{2\zeta}(q_x^2-xq^2-q^4+2vq)+\dots\\ \dfrac{iq}{2\zeta}+\dots\end{pmatrix}, \tag{3.6a}$$

$$\psi^{(2)}(\zeta,x)\sim\tilde{\psi}^{(2)}=e^{4i\zeta^3/3+i\zeta x}\begin{pmatrix}-\dfrac{iq}{2\zeta}+\dots\\ 1+\dfrac{i}{2\zeta}(q_x^2-xq^2-q^4+2vq)+\end{pmatrix}. \tag{3.6b}$$

Since our concern is with (3.1), we will henceforth write q_x for r. If (3.1) holds, $q_x^2-xq^2-q^4+2vq=-\int_x^\infty q^2dx+\text{const}$ and if q, q_x tend to zero sufficiently rapidly as $x\to\infty$, this coefficient is simply $\int_x^\infty q^2dx$. The growth or decay of the two formal asymptotic expansions $\tilde{\psi}^{(1)}(\zeta,x)$ and $\tilde{\psi}^{(2)}(\zeta,x)$ as $\zeta\to\infty$ is determined by the exponential factor $e^{\pm 4i\zeta^3/3}$; the former (latter) series is dominant (recessive), meaning exponentially growing (decaying), as $\zeta\to\infty$ in the sectors S_1, S_3 and S_5 shown in Fig. 1 below, and recessive (dominant) in the sectors S_2, S_4, S_6,

$$S_j=\left\{\zeta\,|\,|\zeta|>\varrho,\ \text{some }\varrho,\ \frac{(j-1)\pi}{3}\leqq\arg\zeta<\frac{j\pi}{3}\right\}.$$

The initial lines of the sectors S_j are called the *anti-Stokes lines.* The lines on which the solutions are maximally dominant or recessive (in this case $\pi/6+\pi j/3$, $j=0,1,\dots,5$) are called the *Stokes lines.*

Consider the solutions $\psi_1^{(1)}(\zeta,x)$ and $\psi_1^{(2)}(\zeta,x)$ of (3.2) which in S_1 have the asymptotic expansions $\tilde{\psi}^{(1)}$ and $\tilde{\psi}^{(2)}$. These solutions will usually be defined by integral representations following the procedure suggested by Birkhoff [16], or as solutions to integral equations. Then, by standard methods (steepest descent, iterative solutions), one can find asymptotic expansions for these solutions in other

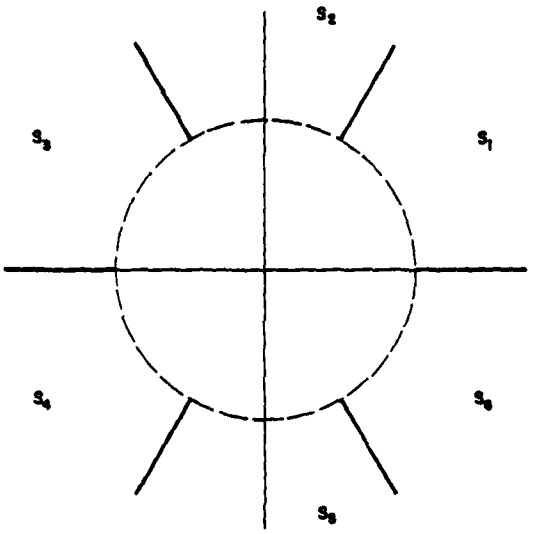


Fig. 1. The six sectors of infinity for (3.2)

sectors and, by taking appropriate linear combinations, one can determine canonical bases $\{\psi_j^{(1)}(\zeta, x), \psi_j^{(2)}(\zeta, x)\}_{j=2}^{7}$ which have the properties

$$\psi_j^{(1)}(\zeta, x) \sim \tilde{\psi}^{(1)}(\zeta, x), \quad \psi_j^{(2)}(\zeta, x) \sim \tilde{\psi}^{(2)}(\zeta, x) \quad \text{in } S_j. \tag{3.7}$$

We emphasize that the sector S_j includes its initial ray on which the asymptotic expansions are neutral (neither growing nor decaying). The fundamental matrix $\Psi_j(\zeta, x)$ will not, in general, be equal to its contiguous neighbors Ψ_{j-1} and Ψ_{j+1}, but will be related to them by the *Stokes multiplier matrices* $A_j(x)$,

$$\Psi_{j+1}(\zeta, x) = \Psi_j(\zeta, x) A_j(x). \tag{3.8}$$

Each $A_j(x)$ is triangular and has the form $\begin{pmatrix} 1 & a_j \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & 0 \\ a_j & 1 \end{pmatrix}$.

The entries a_j are called the Stokes multipliers. The reason the matrices are triangular is this: one can show by analytic continuation of the integral representation, that a solution which is recessive in S_j admits analytic continuation to S_{j+1} and has the same asymptotic expansion (which is now dominant) there [16]. On the other hand, the dominant solution in a sector S_j may need to pick up a recessive component before it can represent a recessive solution in the neighboring sector S_{j+1}.

This phenomenon was first discovered by Stokes in 1857 in his analysis of the Airy function

$$\mathrm{Ai}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx + ik^3/3} dk. \tag{3.9}$$

Stokes noticed that whereas the asymptotic expansion

$$\mathrm{Ai}(x) \sim \frac{1}{2\sqrt{\pi}x^{1/4}} \exp(-\tfrac{2}{3}x^{3/2}) \tag{3.10}$$

is valid in the Poincaré sense in the region $|\arg x| < \pi$, it is necessary that a portion of the solution whose asymptotic behavior is $\frac{1}{2\sqrt{\pi}x^{1/4}} \exp(\frac{2}{3}x^{3/2})$ be added before $\arg x$ reaches π. At which value of $\arg x$ the extra portion is first added was the subject of debate, a debate only recently resolved by Olver [22], who pointed out that uniform bounds on the error were only obtainable for $|\arg x| < \frac{2\pi}{3}$, that is, up to the last Stokes line before $\arg x = \pi$. In fact, Stokes himself knew by direct calculation that the asymptotic representation (3.10) was a poor approximation to the exact solution, which he computed by power series, once $|\arg x|$ exceeded $\frac{2\pi}{3}$. For our purposes, it is not crucial to know on which line a given dominant solution must pick up a multiple of the recessive solution; all we use is that in each sector, $\Psi_j \sim \tilde{\Psi} = (\tilde{\psi}^{(1)}, \tilde{\psi}^{(2)})$ in the *Poincaré sense* of asymptotic expansions (that is, exponentially small terms can be omitted).

The Stokes multiplier matrices A_j have certain symmetry properties which follow from the symmetry properties of (3.2). If $\Psi(\zeta, x)$ is a solution of (3.2), so is $M\Psi(-\zeta, x)$, where $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Thus,

$$\psi_4^{(1)}(\zeta, x) = M\psi_1^{(2)}(-\zeta, x),\ \psi_5^{(2)}(\zeta, x) = M\psi_2^{(1)}(-\zeta, x),\ \psi_6^{(1)}(\zeta, x) = M\psi_3^{(2)}(-\zeta, x) \tag{3.10a}$$

as each of these solutions is recessive in the sector indicated by the subscript, and recessive solutions are uniquely determined by their asymptotic expansion. The dominant solutions satisfy the same symmetry properties:

$$\psi_4^{(2)}(\zeta, x) = M\psi_1^{(1)}(-\zeta, x),\ \psi_5^{(1)}(\zeta, x) = M\psi_2^{(2)}(-\zeta, x),$$

$$\psi_6^{(2)}(\zeta, x) = M\psi_3^{(1)}(-\zeta, x). \tag{3.10b}$$

Indeed, since $M\psi_1^{(1)}(-\zeta, x)$ goes like $\begin{pmatrix} 0 \\ 1 \end{pmatrix} \exp(4i\zeta^3/3 + i\zeta x)$ on $\arg\zeta = \pi$ and is also a dominant solution in S_4, it is exactly $\psi_4^{(2)}(\zeta, x)$. The remaining relations of (3.10b) follow similarly. If $q(x)$ is real, then if $\Psi(\zeta, x)$ is a solution so is $M\Psi^*(\zeta^*, x)$, which will imply further restrictions on the Stokes multipliers [see (3.26)].

We now write down the fundamental solution matrices $\Psi_j(\zeta, x) = (\psi_j^{(1)}(\zeta, x), \psi_j^{(2)}(\zeta, x))$ which have the asymptotic behavior

$$\tilde{\Psi} = \begin{pmatrix} 1 - \frac{i}{2\zeta}(q_x^2 - xq^2 - q^4 + 2vq) + \ldots, & \frac{-iq}{2\zeta} + \ldots \\ \frac{iq}{2\zeta} + \ldots, & 1 + \frac{i}{2\zeta}(q_x^2 - xq^2 - q^4 + 2vq) + \ldots \end{pmatrix} \begin{pmatrix} e^{-\theta} & 0 \\ 0 & e^{\theta} \end{pmatrix}, \tag{3.11}$$

where $\theta = 4i\zeta^3/3 + i\zeta x$. From now on, we drop the subscript on Ψ_1.

$$0 \leqq \arg\zeta < \frac{\pi}{3}, \quad \Psi_1(\zeta, x) \equiv \Psi(\zeta, x) = (\psi^{(1)}, \psi^{(2)}), \tag{3.12a}$$

$$\frac{\pi}{3} \leqq \arg\zeta < \frac{2\pi}{3}, \quad \Psi_2(\zeta, x) = \Psi(\zeta, x)\begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}, \tag{3.12b}$$

$$\frac{2\pi}{3} \leqq \arg\zeta < \pi, \quad \Psi_3(\zeta, x) = \Psi_2(\zeta, x)\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \tag{3.12c}$$

$$\pi \leqq \arg\zeta < \frac{4\pi}{3}, \quad \Psi_4(\zeta, x) = \Psi_3(\zeta, x)\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}, \tag{3.12d}$$

$$\frac{4\pi}{3} \leqq \arg\zeta < \frac{5\pi}{3}, \quad \Psi_5(\zeta, x) = \Psi_4(\zeta, x)\begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}, \tag{3.12e}$$

$$\frac{5\pi}{3} \leqq \arg\zeta < 2\pi, \quad \Psi_6(\zeta, x) = \Psi_5(\zeta, x)\begin{pmatrix} 1 & 0 \\ e & 1 \end{pmatrix}, \tag{3.12f}$$

$$2\pi \leqq \arg\zeta < \frac{7\pi}{3}, \quad \Psi_7(\zeta, x) = \Psi_6(\zeta, x)\begin{pmatrix} 1 & f \\ 0 & 1 \end{pmatrix}. \tag{3.12g}$$

We remark now, and prove later, that each $\Psi_j(\zeta, x)$ also satisfies the x-equation (3.3). The fundamental matrix $\Psi_7(\zeta, x)$ is defined on a sector including the positive real axis $\arg\zeta = 2\pi$. If Ψ is meromorphic in the complex ζ-plane, $\Psi_7 = \Psi$. However, multivaluedness cannot be seen from the asymptotic behavior at $\zeta = \infty$, but can be inferred from the behavior of the solution about $\zeta = 0$. This we shall consider in a moment. First, however, let us use the symmetry properties (3.10) to show that $d = a$, $e = b$, $f = c$. For example, since $\psi_2^{(1)} = \psi_1^{(1)} + a\psi_1^{(2)}$, one has $M\psi_2^{(1)}(-\zeta) = M\psi_1^{(1)}(-\zeta) + aM\psi_1^{(2)}(-\zeta)$ or $\psi_5^{(2)}(\zeta) = \psi_4^{(2)}(\zeta) + a\psi_4^{(1)}(\zeta)$. Comparison with (3.12e) shows $d = a$. The remaining relations follow similarly.

It is straightforward to write down linearly independent solutions of (3.2) near $\zeta = 0$. When ν is not a half integer $\frac{2n+1}{2}$, they are of the form

$$\phi^{(1)}(\zeta, x) = \frac{1}{\sqrt{2}\zeta^{\nu}} e^{-u(x)} \left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix} - \frac{i\zeta}{1-2\nu}(2q_x + 2q^2 + x)\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \ldots \right\}, \tag{3.13a}$$

$$\phi^{(2)}(\zeta, x) = \frac{1}{\sqrt{2}}\zeta^{\nu} e^{u(x)} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \ldots \right\}, \tag{3.13b}$$

where the normalizing factors $e^{\pm u(x)}$, $u_x = q(x)$ have been introduced in order that (3.13) satisfy (3.3). The coefficient vectors in (3.13) can always be chosen to alternate between $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$; this is automatic when $\nu \neq n$, but we impose this pattern also when $\nu = n$. The reader can readily verify that even though ν be an integer, so that the difference 2ν of the indicial roots $-\nu$ and $+\nu$ is an integer, no

logarithm terms are needed. When $\nu = \frac{2n+1}{2}$, there will, in general, be logarithm terms. The two linearly independent solutions will then be $\phi^{(2)}(\zeta, x)$ and

$$\phi^{(1)}(\zeta, x) = -ij\phi^{(2)}(\zeta, x)\ln\zeta + \phi^{(1)'}(\zeta, x), \tag{3.14}$$

where $\phi^{(1)'}$ has the form of $\zeta^{-\nu}$ times a holomorphic function. In (3.14), j is proportional to the coefficient of $\zeta^{2\nu-1/2}$ in the series (3.13a). For example, when $\nu = \frac{1}{2}$,

$$j = 2\left(q_x + q^2 + \frac{x}{2}\right)e^{-2u}. \tag{3.15}$$

Note that the logarithms will disappear if $j=0$; when $\nu = \frac{1}{2}$ this implies

$$q_x + q^2 + \frac{x}{2} = 0. \tag{3.16}$$

(3.16) defines a one-parameter family of solutions of (3.1) for which the second order Painlevé equation reduces to a first order equation (see 3F).

The solutions $\phi^{(1)}$ and $\phi^{(2)}$ satisfy the symmetry condition

$$M\phi^{(1)}(\zeta e^{-i\pi}) = -e^{\nu\pi i}\phi^{(1)}(\zeta) - \pi j e^{-\nu\pi i}\phi^{(2)}(\zeta), \tag{3.17a}$$

$$M\phi^{(2)}(\zeta e^{-i\pi}) = e^{-\nu\pi i}\phi^{(2)}(\zeta). \tag{3.17b}$$

Also, from (3.14), if $\Phi(\zeta, x)$ is the fundamental matrix $(\phi^{(1)}, \phi^{(2)})$ in $0 \leqq \arg\zeta < 2\pi$, then

$$\Phi(\zeta e^{2\pi i}, x) = \Phi(\zeta, x)J \tag{3.18a}$$

is a fundamental solution matrix in the sector $(2\pi, 4\pi)$. The matrix J is

$$J = \begin{pmatrix} e^{-2\pi i\nu} & 0 \\ 2\pi j e^{2\pi i\nu} & e^{2\pi i\nu} \end{pmatrix}, \tag{3.18b}$$

where j is only nonzero when ν is a half integer, in which case $e^{-2\pi i\nu} = e^{2\pi i\nu} = -1$. We remark now, and prove later, that J is independent of x.

Finally, we specify the relation between $\Psi(\zeta, x)$ and $\Phi(\zeta, x)$ to be

$$\Psi(\zeta, x) = \Phi(\zeta, x)A, \tag{3.19a}$$

where

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tag{3.19b}$$

and $\alpha\delta - \beta\gamma = 1$ since $\det\Psi = \det\Phi = 1$.

The set of data

$$T = \{a, b, c, \alpha, \beta, \delta, \alpha\delta - \beta\gamma = 1, \nu, j, \omega_1, \omega_2, \omega_3, l\} \tag{3.20}$$

is the *transform data*; in our case, $\omega_3 = 4i/3$, $\omega_2 = l = 0$, $\omega_1 = ix$.

3C. Some Properties of the Transform Data.

With the conventions adopted here, the transform data are not uniquely specified by (3.2), (3.3) because there are some ambiguities in the definitions (3.13), (3.14) of Φ. First, we consider this situation when 2ν is not an integer. The exponent $\pm u(x)$ in (3.13) is determined only up to an arbitrary constant; a change in this constant amounts to the multiplication of $\phi^{(1)}$ by some k and of $\phi^{(2)}$ by k^{-1}. Therefore, a one-parameter family of connection matrices A is consistent with a normalization of the type (3.13). One could remedy this by imposing

$$\Phi(x_0,\zeta)\begin{pmatrix}\zeta^{\nu} & \\ & \zeta^{-\nu}\end{pmatrix}\Big|_{\zeta=0}=\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix},$$

but as the ambiguity will cause no problems, we shall not insist on such a condition. When ν is an integer, another potential indeterminacy appears because any linear combination $\phi^{(1)}+k\phi^{(2)}$ will be of the form $\zeta^{-n}x$ holomorphic function. This would, however, violate the $\begin{pmatrix}1\\ \pm 1\end{pmatrix}$ patterns imposed on the coefficients in (3.13), and hence the symmetry (3.17). When $\nu=\dfrac{2n+1}{2}$ and $j\neq 0$, there is again one free scaling parameter, and A and the *jump* j are determined up to this parameter.

The only really singular situation arises when $\nu=\dfrac{2n+1}{2}$ and $j=0$. Then the symmetry condition (3.17) does not distinguish between different linear combinations $\phi^{(1)}+k\phi^{(2)}$. This case is exceptional in many respects; to illustrate its peculiar features, we turn now to an analysis of various relations among the transform data.

Our attitude is always that x and ν *are given*, and that q and r in (3.2) are to be found. If the transform data do indeed determine those two complex numbers, then all of $a, b, c, j, \alpha, \beta, \gamma, \delta$ should depend on only two quantities in the list. Of course, $\alpha\delta-\beta\gamma=1$, and there will always be one free parameter in A and j due to the scaling freedom just discussed. Modulo this indeterminacy, we have the following results:

(A) As long as $\nu\neq\dfrac{2n+1}{2}$ or $\nu=\dfrac{2n+1}{2}$ and $j\neq 0$, any two of the three Stokes multipliers determine all transform data.

(B) If $\nu=\dfrac{2n+1}{2}$ and $j=0$, $\dfrac{\alpha}{\beta}$ or $\dfrac{\beta}{\alpha}$ determines all remaining transform data (the second determining constant *in* this case being j). In particular, $a=b=c=(-1)^{n+1}i$.

Remark. The inverse problem for the exceptional case (B) can be solved – all systems (3.2) with such transform data can be constructed explicitly.

We now prove assertions (A), (B). To this end, we derive two sets of relations among the transform data; the relations are consistent but carry slightly different information.

First, we note that from (3.19a) and (3.18a)

$$\Psi(\zeta e^{2\pi i}) = \Phi(\zeta) J A,$$

and thus

$$\Psi_\gamma(\zeta e^{2\pi i}) \left\{ \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix} \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \right\}^{-1} = \Psi(\zeta) A^{-1} J A. \quad (3.21)$$

Now choose $\zeta \in S_1$ (whence $\zeta e^{2\pi i} \in S_\gamma$) and let $|\zeta| \to \infty$; both $\Psi_\gamma(\zeta e^{2\pi i})$ and $\Psi(\zeta)$ then tend to the same asymptotic matrices. Hence,

$$\begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix} \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} = A^{-1} J^{-1} A.$$

If we write

$$G = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix} \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

then

$$AG^2 = J^{-1} A. \quad (3.22)$$

(3.22) does not use the symmetry (3.17) and so misses some information. Thus, we derive a second set of relations. In (3.19a), set $\zeta = \tilde{\zeta} e^{-i\pi}$, and apply $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Use $M\psi^{(1)}(\tilde{\zeta} e^{-i\pi}) = \psi_4^{(2)}(\tilde{\zeta})$, $M\psi^{(2)}(\tilde{\zeta} e^{-i\pi}) = \psi_4^{(1)}(\tilde{\zeta})$ on Ψ and (3.17) on Φ in the resulting expression. Finally, re-express $\psi_4^{(1)}$, $\psi_4^{(2)}$, $\phi^{(1)}$ and $\phi^{(2)}$ in terms of $\psi^{(1)}$, $\psi^{(2)}$ by (3.12) and (3.19a), and equate coefficients of $\psi^{(1)}$, $\psi^{(2)}$. The result is:

$$\begin{aligned} b &= -\alpha\delta e^{\nu\pi i} - \beta\gamma e^{-\nu\pi i} + \pi j \alpha\beta e^{-\nu\pi i} \\ 1 + ab &= 2\alpha\gamma \cos \nu\pi - \pi j \alpha^2 e^{-\nu\pi i} \\ 1 + bc &= -2\beta\delta \cos \nu\pi + \pi j \beta^2 e^{-\nu\pi i} \\ a + c + abc &= \beta\gamma e^{\nu\pi i} + \alpha\delta e^{-\nu\pi i} - \pi j \alpha\beta e^{-\nu\pi i}. \end{aligned} \quad (3.23)$$

From (3.23), we deduce immediately that,

$$a + b + c + abc = -2i \sin \nu\pi, \quad (3.24)$$

so that any two Stokes multipliers (and ν) determine the third [(3.22) almost yields (3.24); the sign of the right-hand side is not determined]. (3.23) is not convenient if one wants, as we do, to express everything in terms of these two Stokes multipliers. (3.22), on the other hand, is linear in the entries of A. A tedious but straightforward computation shows the following.

When 2ν is not an integer, or when $\nu = \dfrac{2n+1}{2}$ and $j \neq 0$, (3.22) has rank 2, so that $\alpha, \beta, \gamma, \delta, j$ are determined up to two arbitrary constants; one is fixed by $\alpha\delta - \beta\gamma = 1$, the other reflects the scaling freedom in Φ.

When $\nu = n$, (3.22) reduces to $I = I$, while (3.23) yields expressions for $\alpha, \beta, \gamma, \delta$ in terms of two Stokes multipliers (again, up to the one free constant).

When $\nu=\dfrac{2n+1}{2}$ and $j=0$, (3.22) reduces to $I=I$ again. (3.23) can be solved for a,b,c to give

$$a=b=c=(-1)^{n+1}i. \tag{3.25}$$

This has used $\alpha\delta-\beta\gamma=1$, and there are no further constraints contained in (3.23). Thus, A depends on three free parameters. Two are connected with the indeterminacy of the definition of Φ, as explained above. The third parameter, which can be taken to be $\dfrac{\alpha}{\beta}$ or $\dfrac{\beta}{\alpha}$ (α or β may be zero), labels different systems (3.2) corresponding to the one parameter family of solutions of (3.16) (this will be shown later in this section; the systems will be constructed explicitly).

Remark 1. (3.22) provides a quick proof of the absence of logarithms when $\nu=n$. Indeed, (3.22) says $\operatorname{trace} G^2=\operatorname{trace} J^{-1}=2$; computing trace G^2, one finds that this implies $a+b+c+abc=0$. The characteristic polynomial of G then turns out to be $\lambda^2-1=0$, so its eigenvalues are ± 1. G is therefore diagonalizable, and hence so is G^2. But $J=AG^2A^{-1}$ has Jordan block form unless $j=0$.

Remark 2. When can all Stokes multipliers be zero? Since then $G^2=I$, (3.22) shows that $J=I$, whence $\nu=n$. This is another case of which all systems (3.2) can be constructed explicitly.

Remark 3. If, contrary to our previous position, ν is considered unknown, another parameter from the transform data must, of course, take it place. Suppose, for example, that a,b,c are given. Then

$$a+b+c+abc=-2i\sin\nu\pi \tag{3.24}$$

determines $\nu \bmod 2$, and it is clear that (3.22) and (3.23) are not affected by a replacement $\nu\to\nu+2m$. If a,b,c are replaced by their negatives, (3.22) and (3.23) remain consistent provided $\nu\to\nu+2m+1$. In other words, sets of transform data which differ only by an integer translation in ν and sign of a,b,c are possible; this circumstance is related to Airault's [3] Bäcklund transformation, which produces a solution of (3.1) for $\nu\pm1$ from a solution of (3.1) for ν.

Remark 4. If x,q,r in (3.2) are real, the symmetry $V(\zeta,x)\to MV^*(\zeta^*,x)$ shows

$$a=-c^*. \tag{3.26}$$

3D. The x-Dependence of T

It has been mentioned repeatedly that the transform data are independent of x, provided that the Painlevé equation (3.1) is satisfied. We verify this result, and also prove a strong converse: given matrix functions $\Psi_j(x,\zeta)$, $\Phi(x,\zeta)$ with global connection properties characterized by transform data independent of x, there are unique systems (3.2), (3.3) satisfied by Ψ_j, Φ. As a corollary, one can see that there is at most one set of functions Ψ_j, Φ possessing given transform data.

Because Eq. (3.2) is determined solely by the global connection properties of its solution, while (3.3) follows from the x-independence of the connection parameters, the whole theory can be rephrased in a much more geometrical manner without reference to differential equations; this is done in Sect. 6.

In the remainder of this section, we deal with functions satisfying the following conditions:

(a) matrix functions Ψ_j holomorphic in $S_j = \left\{\zeta \,|\, |\zeta|>0,\ (j-1)\frac{\pi}{3} \leqq \arg\zeta < \frac{j\pi}{3}\right\}$, such that

$$\Psi_j \sim \tilde{\Psi} = \left(I + \frac{C_1}{\zeta} + \ldots\right)\begin{pmatrix} e^{-\theta} & 0 \\ 0 & e^{\theta}\end{pmatrix}, \quad \theta = \frac{4}{3}i\zeta^3 + i\zeta x,$$

as $|\zeta| \to \infty$ in S_j, and

$$\Psi_{j+1} = \Psi_j A_j, \quad |\zeta| > \text{some } \varrho.$$

(b) A matrix function Φ of the form

$$\Phi(\zeta) = \hat{\Phi}(\zeta)\begin{pmatrix} \zeta^{-\nu} & 0 \\ 0 & \zeta^{\nu}\end{pmatrix},$$

with $\hat{\Phi}(\zeta)$ holomorphic, such that for $\zeta \in S_1$

$$\Psi_1(\zeta) = \Phi(\zeta) A, \det A = 1.$$

[For sake of simplicity, we omit the modifications necessary when $\Phi(\zeta)$ contains logarithms.]

The formal series $\tilde{\Psi}(\zeta)$ is assumed to have the symmetry

(c) $$M\tilde{\Psi}(-\zeta)M = \tilde{\Psi}(\zeta), \quad M = \begin{pmatrix} 0 & 1 \\ 1 & 0\end{pmatrix}.$$

The functions studied earlier have all these properties.

We now prove the following facts:

Proposition 1. *Let Ψ_j, Φ satisfy* (3.2), *with Ψ_j, Φ normalized as in* (3.6) *and* (3.13). *If the Painlevé equation* (3.1) *holds (so that* (3.2), (3.3) *are compatible), then these Ψ_j, Φ are also solutions of the x-equation* (3.3).

Proposition 2. *If Ψ_j, Φ satisfy both* (3.2) *and* (3.3), *then the Stokes multipliers A_j, the connection matrix A, and the jump matrix J are independent of x.*

Remark. Propositions 1 and 2 show that the transform data introduced above are all independent of x.

Proposition 3. *Let Ψ_j, Φ have properties (a), (b), (c), and suppose that A_j and A are independent of x. Then Ψ_j, Φ satisfy differential equations of the form* (3.2), (3.3).

Proposition 4. *There can be no more than one set of functions Ψ_j, Φ, satisfying properties (a), (b), (c) above.*

Proof of Proposition 1. Write Eq. (3.2) as $\Psi_\zeta = P\Psi$ and (3.3) as $\Psi_x = Q\Psi$. Differentiate the first of these equations with respect to x and solve the resulting inhomogeneous equations by variation of parameters to obtain

$$\Psi_x = \Psi(\Psi^{-1}\Psi_x)_{\zeta_0} + \Psi\int_{\zeta_0}^{\zeta} \Psi^{-1} P_x \Psi \, d\zeta. \tag{3.27}$$

But the compatibility condition of (3.2) and (3.3) is $P_x = Q_\zeta + [Q, P]$ and, since $\Psi^{-1}(Q_\zeta + [Q, P]) = \frac{d}{d\zeta}(\Psi^{-1} Q \Psi)$, we have from (3.27) that

$$\Psi_x = Q\Psi + \Psi[\Psi^{-1}(\Psi_x - Q\Psi)]_{\zeta_0}.$$

Now let ζ_0 tend to infinity on the initial ray of that sector in which $\Psi \sim \tilde{\Psi}$ as given in (3.6). One finds $\Psi_x - Q\Psi = O\left(\frac{1}{\zeta_0}\right)$, which tends to zero. Hence, $\Psi_x = Q\Psi$. A similar proof holds for Φ.

Proof of Proposition 2. We prove that A_j is independent of x. $A_j = \Psi_j^{-1}\Psi_{j+1}$, and thus $A_{jx} = -\Psi_j^{-1}\Psi_{jx}\Psi_j^{-1}\Psi_{j+1} + \Psi_j^{-1}\Psi_{j+1\,x} = -\Psi_j^{-1}Q\Psi_{j+1} + \Psi_j^{-1}Q\Psi_{j+1} = 0$. The proof that A and J are independent of x is similar.

Remark. From the form of J when, for example, $v = \frac{1}{2}$, it is by no means obvious that $j_x = 0$. However, note that the x derivative of $j = 2\left(q_x + q^2 + \frac{x}{2}\right)e^{-2u}$ is zero by virtue of (3.1).

Proof of Proposition 3. First we note that (c) implies the following symmetry for the coefficient C_1 in the expansion $\tilde{\Psi}$:

$$-MC_1M = C_1.$$

It follows that $C_{1,12} = -C_{1,21}$ and $C_{1,11} = -C_{1,22}$; we set $C_{1,12} = -\frac{i}{2}q$, and $C_{1,11} = p$. Now differentiate $\Psi_{j+1} = \Psi_j A_j$ with respect to x, and multiply by Ψ_{j+1}^{-1}:

$$\Psi_{j+1\,x}\Psi_{j+1}^{-1} = \Psi_{jx}A_j\Psi_{j+1}^{-1} = \Psi_{jx}A_j(\Psi_j A_j)^{-1} = \Psi_{jx}\Psi_j^{-1}.$$

$\Psi_{jx}\Psi_j^{-1}$ is therefore well-defined and holomorphic in a deleted neighborhood of ∞, and its asymptotic expansion is $\tilde{\Psi}_x\tilde{\Psi}^{-1}$, uniformly for $|\zeta| >$ some ϱ.

The asymptotic expansion is therefore convergent; we set

$$\tilde{\Psi}_x\tilde{\Psi}^{-1} = Q(\zeta), \tag{3.28}$$

$Q(\zeta)$ being a Laurent series in ζ^{-1}, with a simple pole at $\zeta = \infty$. Near $\zeta = 0$, we find in a similar way that

$$\Phi_x\Phi^{-1} = \tilde{\Phi}_x\tilde{\Phi}^{-1} = \Psi_{1x}\Psi_1^{-1} \equiv Q(\zeta).$$

But $\tilde{\Phi}_x\tilde{\Phi}^{-1}$ is holomorphic at $\zeta = 0$, so that $Q(\zeta)$ contains no negative powers of ζ. The explicit form of $Q(\zeta)$ is now easy to obtain by inserting the expansion $\tilde{\Psi}$ into (3.28). The result is

$$Q(\zeta) = i\zeta\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} + i\left[C_1, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\right]$$

or

$$Q(\zeta) = \begin{pmatrix} -i\zeta & q \\ q & i\zeta \end{pmatrix}.$$

Hence, Ψ_j, Φ satisfy an equation of the form (3.3) in x. A similar argument shows that $\Psi_{j\zeta}\Psi_j^{-1}$ is well-defined and holomorphic away from 0, ∞, so that

$$\Psi_{j\zeta}\Psi_j^{-1} = \tilde{\Psi}_\zeta \tilde{\Psi}^{-1} \equiv P(\zeta).$$

$P(\zeta)$ has a double pole at $\zeta = \infty$. Near $\zeta = 0$, we find

$$P(\zeta) = \hat{\Phi}_\zeta \hat{\Phi}^{-1} + \frac{\nu}{\zeta}\hat{\Phi}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\hat{\Phi}^{-1}.$$

Hence, $P(\zeta)$ contains at most the negative power ζ^{-1}. We wish to show that in fact ζ^{-1} enters $P(\zeta)$ in the form

$$\pm\frac{\nu}{\zeta}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Put $\hat{\Phi}(0)\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\hat{\Phi}^{-1}(0) = (\varepsilon_{ij})$. First, we note that the diagonal entries ε_{11}, ε_{22} must be zero; otherwise there would be a term $\zeta^{\pm\nu}$ in the expansion $\tilde{\Psi}$ [16]. In terms of the entries $\hat{\Phi}_{ij}(0)$, this means

$$\hat{\Phi}_{11}(0)\hat{\Phi}_{22}(0) + \hat{\Phi}_{12}(0)\hat{\Phi}_{21}(0) = 0.$$

But also $\det\hat{\Phi}(0) = 1$, and a short computation shows that $\varepsilon_{12} = \varepsilon_{21}^{-1}$. Property (c) implies the symmetry

$$P(-\zeta) = -MP(\zeta)M,$$

and this forces $\varepsilon_{12} = \varepsilon_{21}$, so $\varepsilon_{12} = \varepsilon_{21} = \pm 1$, as was to be shown.

It now remains to compute $\tilde{\Psi}_\zeta \tilde{\Psi}^{-1} = P(\zeta)$ through the ζ^0 term. There are certain non-obvious cancellations, e.g. of ζ^{-2} terms, which are automatic by the analysis at $\zeta = 0$. By (c), C_2 has the form $\begin{pmatrix} y & z \\ z & y \end{pmatrix}$. One finds that $\left[\text{with } N = \begin{pmatrix} -1 & \\ & 1 \end{pmatrix}\right]$

$$P(\zeta) = (4i\zeta^2 + ix)\left\{N + \frac{1}{\zeta}[C_1, N] + \frac{1}{\zeta^2}([C_2, N] + [N, C_1]C_1)\right\}$$

or

$$P(\zeta) = (4i\zeta^2 + ix + 2iq^2)\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} + 4\zeta q\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
$$+ (4z - 2ipq)\begin{pmatrix} 0 & 2i \\ -2i & 0 \end{pmatrix} \pm \frac{\nu}{\zeta}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

If $r \equiv 4z - 2ipq$, this is precisely the form required by (3.2), up to the ambiquity in the sign of ν. The hypotheses do not allow one to distinguish between

$$+\frac{\nu}{\zeta}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{with } \Phi \text{ as defined in (3.13),}$$

and

$$+\frac{(-\nu)}{\zeta}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{with} \quad \Phi = (\phi^{(2)}, -\phi^{(1)}).$$

Had we assumed, in addition, that $\hat{\Phi}(0)$ is proportional to $\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$, the choice $+\frac{\nu}{\zeta}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ would have been the only acceptable one.

Remark. In the first case, q will satisfy Eq. (3.1). In the second case, q solves $q_{xx}=2q^3+xq+\nu$, but $-q$ again solves (3.1).

Proof of Proposition 4. Suppose that Ψ_j and Ψ_j' $(j=1,\ldots,7)$ have property (a), with the same Stokes multipliers A_j, and suppose that $\Psi_1=\Phi A$, $\Psi_1'=\Phi' A$ with Φ, Φ' satisfying (b). Then

$$\Psi_{j+1}'\Psi_{j+1}^{-1}=\Psi_j' A_j A_j^{-1}\Psi_j^{-1}=\Psi_j'\Psi_j^{-1},$$

so $S\equiv\Psi_j'\Psi_j^{-1}$ is well-defined and holomorphic about $\zeta=\infty$, and from the asymptotic expansion one sees that S is a Laurent series in ζ, $S(\zeta)=I+O\left(\frac{1}{\zeta}\right)$. Near $\zeta=0$,

$$S(\zeta)=\Psi_1'\Psi_1^{-1}=\Phi'\Phi^{-1}=\hat{\Phi}'\hat{\Phi}^{-1}=\text{const}+O(\zeta),$$

a Taylor series with no negative powers. Hence $S(\zeta)\equiv I$, as was to be shown.

3E. *The Inverse Transform*

We now turn to the existence problem; are there solutions Ψ_j, Φ giving rise to a prescribed set of transform data? Here we derive linear singular integral equations

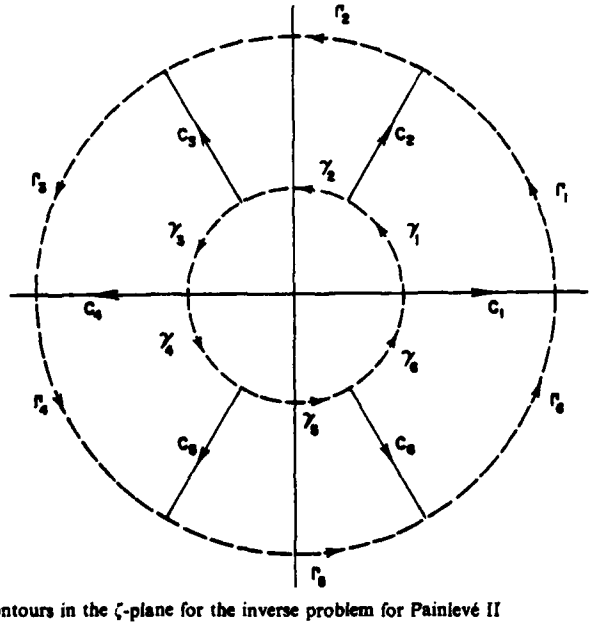


Fig. 2. The contours in the ζ-plane for the inverse problem for Painlevé II

from which the solutions Ψ_j and Φ may be constructed. The parameters in these equations are x and the Stokes multipliers. For simplicity of presentation, we take $\nu = n$, a positive integer. The results for $\nu \neq n$ are given in Appendix IV.
Consider

$$\int_{C_1} \frac{\psi^{(1)}(\xi, x) e^{+\theta}}{\xi - \zeta} d\xi, \quad \theta = 4i\zeta^3/3 + i\zeta x; \tag{3.31}$$

for convenience, we take ζ to lie in the interior of S_1. Our goal is to write a system of linear singular equations for $\psi^{(1)}$ and $\psi^{(2)}$. The strategy will be to relate, by Cauchy's theorem, the integral along C_1 to one along C_2 and to continue this process around the singular point $\zeta = \infty$.

We begin by noticing that

$$\int_{C_1} \frac{\psi^{(1)} e^{+\theta}}{\xi - \zeta} d\xi = 2\pi i \psi^{(1)}(\zeta) e^{\theta} - \frac{\pi i}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \int_{C_2} \frac{\psi_2^{(1)} - a\psi_2^{(2)}}{\xi - \zeta} e^{\theta} d\xi + \int_{\gamma_1} \frac{\psi^{(1)} e^{\theta}}{\xi - \zeta} d\xi, \tag{3.32a}$$

where we have used (3.12b). Now $\psi_2^{(2)} e^{\theta} \to \infty$ as $\zeta \to \infty$ in S_2, and therefore it is not possible to relate the term $-a \int_{C_2} \frac{\psi_2^{(2)} e^{\theta}}{\xi - \zeta} d\xi$ by contour integration to an integral along C_3. This integral is not transformed any further and appears in this form in the final equation. We can, however, continue with

$$\int_{C_2} \frac{\psi_2^{(1)} e^{\theta}}{\xi - \zeta} d\xi = -\frac{\pi i}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \int_{C_3} \frac{\psi_3^{(1)} e^{\theta}}{\xi - \zeta} d\xi + \int_{\gamma_2} \frac{\psi_2^{(1)} e^{\theta}}{\xi - \zeta} d\xi. \tag{3.32b}$$

Continuing in this manner around the ζ-plane we find,

$$\int_{C_3} \frac{\psi_3^{(1)} e^{\theta}}{\xi - \zeta} d\xi = -\frac{\pi i}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \int_{C_4} \frac{\psi_4^{(1)} - c\psi_4^{(2)}}{\xi - \zeta} e^{\theta} d\xi + \int_{\gamma_3} \frac{\psi_3^{(1)} e^{\theta}}{\xi - \zeta} d\xi, \tag{3.32c}$$

$$\int_{C_4} \frac{\psi_4^{(1)} e^{\theta}}{\xi - \zeta} d\xi = -\frac{\pi i}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \int_{C_5} \frac{\psi_5^{(1)} e^{\theta}}{\xi - \zeta} d\xi + \int_{\gamma_4} \frac{\psi_4^{(1)} e^{\theta}}{\xi - \zeta} d\xi, \tag{3.32d}$$

$$\int_{C_5} \frac{\psi_5^{(1)} e^{\theta}}{\xi - \zeta} d\xi = -\frac{\pi i}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \int_{C_6} \frac{\psi_6^{(1)} - b\psi_6^{(2)}}{\xi - \zeta} e^{\theta} d\xi + \int_{\gamma_5} \frac{\psi_5^{(1)} e^{\theta}}{\xi - \zeta} d\xi, \tag{3.32e}$$

and

$$\int_{C_6} \frac{\psi_6^{(1)} e^{\theta}}{\xi - \zeta} d\xi = -\frac{\pi i}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \int_{C_1} \frac{\psi^{(1)} e^{\theta}}{\xi - \zeta} + \int_{\gamma_6} \frac{\psi_6^{(1)} e^{\theta}}{\xi - \zeta} d\xi. \tag{3.32f}$$

Here, we have used $\psi_7^{(1)} = \psi^{(1)}$, which is a consequence of the assumption $\nu = n$. Adding the Eqs. (3.32a) through (3.32f), we find

$$\psi^{(1)}(\zeta) e^{\theta(\zeta)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \operatorname*{Res}_{\xi = 0} \frac{\psi^{(1)} e^{\theta(\xi)}}{\xi - \zeta} + \frac{a}{2\pi i} \int_{C_{42}} \frac{\psi^{(2)} e^{\theta}}{\xi - \zeta} d\xi$$
$$+ \frac{b}{2\pi i} \int_{C_{46}} \frac{\psi^{(2)} e^{\theta}}{\xi - \zeta} d\xi + \frac{bc}{2\pi i} \int_{C_{64}} \frac{\psi^{(1)} e^{\theta}}{\xi - \zeta} d\xi, \tag{3.33}$$

where we have substituted for $\psi_j^{(2)}$, $j=2,4,6$ from (3.12) *in terms* of $\psi^{(1)}$, $\psi^{(2)}$ and used identity (3.24), $a+b+c+abc=0$. The contour C_{42} is one which runs inward along C_4, then clockwise along γ_4 and γ_3 and goes out along C_2. The contours C_{64} and C_{46} (which is $-C_{64}$) are defined in a similar way (see Fig. 3).

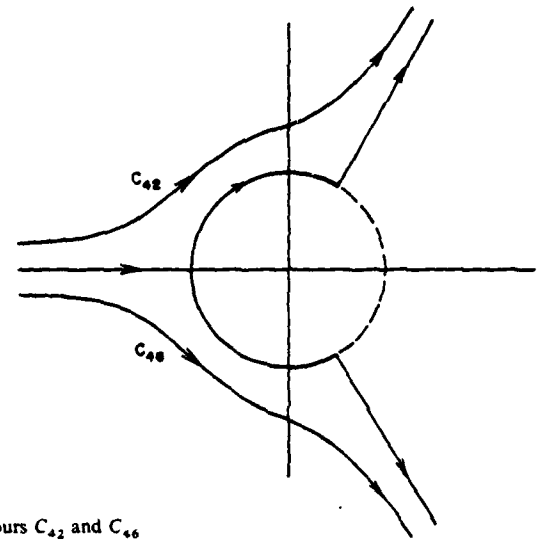


Fig. 3. The contours C_{42} and C_{46}

Remark. Note that the same contours are used in the integral representations of the Airy functions. In fact, (3.33) contains the Airy function representations as a limiting case [see example (i) which follows in 3F].

By considering

$$\int_{C_1} \frac{\psi^{(2)} e^{-\theta}}{\xi-\zeta} d\xi, \tag{3.34}$$

we find

$$\psi^{(2)} e^{-\theta} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \operatorname*{Res}_{\xi=0} \frac{\psi^{(2)} e^{-\theta}}{\xi-\zeta} + \frac{b}{2\pi i} \int_{C_{53}} \frac{\psi^{(1)} e^{-\theta}}{\xi-\zeta} d\xi + \frac{c}{2\pi i} \int_{C_{51}} \frac{\psi^{(1)} e^{-\theta}}{\xi-\zeta} d\xi$$
$$+ \frac{ab}{2\pi i} \int_{C_{53}} \frac{\psi^{(2)} e^{-\theta}}{\xi-\zeta} d\xi. \tag{3.35}$$

Equations (3.33) and (3.35) are linear integral equations whose solution determines $\psi^{(1)}$ and $\psi^{(2)}$, and hence all the coefficients in the Eq. (3.2). In particular, from the asymptotic expansions (3.6), we known that

$$q = \lim_{\zeta \to \infty} 2i\zeta \psi_{,1}^{(2)} e^{-\theta} \tag{3.36a}$$
$$= -\lim_{\zeta \to \infty} 2i\zeta \psi_{,2}^{(1)} e^{\theta} \tag{3.36b}$$

where the subscripts in (3.36) refer to the component, and not to the sector.

The integral Eqs. (3.33) and (3.35) have been derived on the assumption that functions Ψ_j with the prescribed Stokes multipliers exist. Conversely, one wants to answer the following questions:

(1) For which x, a, b, n does a solution of (3.33), (3.35) exist?

(2) Does it exhibit the prescribed Stokes jumps?

(3) What properties of $q(x)$ can one deduce from the solution Ψ so constructed?

(4) What is the dependence of $q(x)$ on the parameter a, b?

Some preliminary observations follow from the known properties of the Painlevé transcendent, $q(x)$. It is never an entire function of x, unless it is identically zero; it is, however, the ratio of two entire functions. This suggests that for given n, a, b, the solution of (3.33) and (3.35) will exist for all but a countable set of x. In particular, the inverse problem (Riemann-Hilbert problem) is not always solvable subject to the symmetry imposed on (3.2).

It is not too hard to see that any solution of the integral equations will have the required jumps; this follows from the behavior of the Cauchy integrals when ζ crosses an integration contour. The first problem, however, is existence of a solution. We have not found a proof; indeed, we have not really looked for one. The reason is this: other irregular-singular monodromy problems will lead to different, and more complicated sets of singular integral equations – see Sect. 4, for example. One needs a quite general theorem, if case-by-case existence proofs are to be avoided. It is clear that existence will depend in a subtle way on the exponent $\theta(\zeta)$, particularly when θ contains several independently varying parameters (cf. Appendix II). Local existence in x is probably easier to get, but not of much interest for applications to Painlevé equations. Thus, we restrict ourselves in this paper to the examination of various limiting cases and special examples. Other properties of Eqs. (3.33), (3.35) are under investigation.

3F. Special Solutions

(i) The linear limit

Take $v=0$ and a, b, c, small. The usual procedure in solving (3.33) and (3.35) is to form the Neumann series. Here we keep only the terms linear in the parameters a, b, and c. We find

$$\psi^{(1)}e^{\theta} = \begin{pmatrix}1\\0\end{pmatrix} + \frac{a}{2\pi i}\int_{C_{42}} \frac{\begin{pmatrix}0\\1\end{pmatrix}e^{2\theta}}{\xi-\zeta}d\xi + \frac{b}{2\pi i}\int_{C_{46}} \frac{\begin{pmatrix}0\\1\end{pmatrix}e^{2\theta}}{\xi-\zeta}d\xi.$$

Therefore from (3.36)

$$q = \frac{a}{\pi}\int_{C_{42}} e^{8i\xi^3/3+2i\xi x}dx + \frac{b}{\pi}\int_{C_{46}} e^{8i\xi^3/3+2i\xi x}d\xi$$

$$= \left(a+\frac{b}{2}\right)\mathrm{Ai}(x) - \frac{ib}{2}\mathrm{Bi}(x) \tag{3.37}$$

when $\mathrm{Ai}(x)$ and $\mathrm{Bi}(x)$ are the Airy functions. Note that when $q(x)$ is real, $c=-a^*$, $b=a^*-a$ to leading order and

$$q=\frac{a+a^*}{2}\mathrm{Ai}(x)+\frac{i(a-a^*)}{2}\mathrm{Bi}(x)$$
$$=\mathrm{Re}a\,\mathrm{Ai}(x)-\mathrm{Im}a\,\mathrm{Bi}(x). \tag{3.38}$$

(ii) Solutions which decay as $x\to+\infty$. Throughout this example, ν is taken to be zero. In this case, a one-parameter family of solutions of (3.1) has been studied by use of the Marchenko equation of the inverse scattering transform ([2, 10–12]), Eq. (2.14). Such techniques apply when $q(x)\to 0$ sufficiently rapidly as $x\to+\infty$: (3.1) then reduces to $q_{xx}=xq$ for large x. It can be proved that $q(x)$ has the expected asymptotic behavior,

$$q(x)\sim\varrho\,\mathrm{Ai}(x) \tag{3.39}$$

for some constant ϱ.

The rapid decay (3.39) ensures that the constructions of scattering theory can be used on the x-equation (3.3) at $x=+\infty$. In particular, the eigenfunction $v(x,\zeta)$ of (3.3) with asymptotic behavior

$$v(x,\zeta)\sim e^{i\zeta x}\begin{pmatrix}0\\1\end{pmatrix},\qquad x\to+\infty$$

is defined for large enough x, and it admits the triangular representation

$$v(x,\zeta)e^{-i\zeta x}=\begin{pmatrix}0\\1\end{pmatrix}+\int_x^\infty K(x,s)e^{i\zeta(s-x)}\,ds. \tag{3.40}$$

By repeated integration by parts in (3.40), one may derive the asymptotic expansion

$$v(x,\zeta)e^{-i\zeta x}\sim\begin{pmatrix}0\\1\end{pmatrix}+\frac{C_1}{\zeta}+\dots \tag{3.41}$$

valid in the upper half ζ-plane. Now, $v(x,\zeta)e^{-4/3\,i\zeta^3}$ is precisely the solution $\psi^{(2)}$ of (3.2) on $\arg\zeta=0$, and since the expansion (3.41) is valid in S_1, S_2, S_3, the Stokes multiplier b must vanish. In this case, (3.33) reduces to

$$\psi^{(1)}(\zeta)e^{\theta(\zeta)}=\begin{pmatrix}1\\0\end{pmatrix}+\frac{a}{2\pi i}\int_{C_{4_2}}\frac{\psi^{(2)}e^{\theta}}{\xi-\zeta}\,d\xi$$
$$=\begin{pmatrix}1\\0\end{pmatrix}-a\psi^{(2)}(\zeta)e^{\theta(\zeta)}+\frac{a}{2\pi i}\int_{-\infty}^{\infty}\frac{\psi^{(2)}e^{\theta}}{\xi-\zeta}\,d\xi. \tag{3.42}$$

Since $\psi^{(1)}+a\psi^{(2)}=\psi_4^{(1)}=M\psi^{(2)}(-\zeta)$, we may change $\zeta\to-\zeta$ in (3.42); then

$$\begin{pmatrix}0&1\\1&0\end{pmatrix}\psi^{(2)}(\zeta)e^{-\theta(\zeta)}=\begin{pmatrix}1\\0\end{pmatrix}+\frac{a}{2\pi i}\int_{-\infty}^{\infty}\frac{\psi^{(2)}e^{\theta}}{\xi+\zeta}\,d\xi. \tag{3.43}$$

Substitution of (3.40) into (3.43), followed by a Fourier transform [23], will recover the Marchenko equation (2.14). The constant ϱ in the asymptotic form (3.39) of $q(x)$ coincides with the nonzero Stokes multiplier a.

In this example, therefore, one can see explicitly how the Stokes multipliers a and b characterize the Painlevé transcendent $q(x)$:

$b=0$ implies $q(x)\to 0$ as $x\to+\infty$,

a describes the asymptotic behavior, $q(x)\sim a\mathrm{Ai}(x)$.

It is shown in [11] that there is exactly one $q(x)$ satisfying these conditions. A more detailed analysis [11] of the Marchenko equation (2.14) reveals that $q(x)$ has no pole on the real axis when $-1<a<1$, whereas it has at least one real pole when $|a|>1$. A proof based on (3.43) would be quite analogous to the argument in [11]: for $|a|<1$, the inhomogeneous Eq. (3.43) can be solved for any real x (remember that x enters in θ), while any a with $|a|>1$ is eigenvalue of (3.43) for some real x. It may be possible, however, to go beyond the results of [11, 12]. Since (3.43) is local in x, one can study the limits $x\to\pm\infty$ separately. This is relevant to the solution of the *connection problem* for the second Painlevé transcendent [10, 11]: what is the behavior at $x=-\infty$ of the solution which goes as (3.39) at $+\infty$? The Marchenko equation (2.14), by the time x goes to $-\infty$, contains information about $q(x)$ for all real x, and it is apparently difficult to extract the asymptotics at $-\infty$.

Connection formulas between singular points of Painlevé transcendents are important in several applications; in the Ising model, for example, the behavior near $x=0$ of the third transcendent is of physical interest [4, 13]. We hope to return to these questions in a later paper.

(iii) The rational solutions: the "solitons" of the Painlevé equations. In the inverse scattering transform, the solitons are associated with the bound states of the x-equations (3.3) which are located at the poles of the reflection coefficient in the upper half ζ-plane. The analogue to the multisoliton solution is a class of rational solutions which are associated with the poles of the fundamental solution matrix $\Psi(\zeta,x)$ at $\zeta=0$. When $\nu=n$, we find these solutions by setting $a=b=c=0$. Then from (3.35),

$$\psi^{(2)}(\zeta)e^{-\theta}=\begin{pmatrix}0\\1\end{pmatrix}-\operatorname*{Res}_{\xi=0}\frac{\psi^{(2)}e^{-\theta}}{\xi-\zeta}. \tag{3.44}$$

We mention that Eq. (3.44) will provide the rational solutions for the full class of equations of the Painlevé II family which is defined and discussed in Appendix II.

Near $\zeta=0$, the solution $\psi^{(2)}$ will be a linear combination of $\phi^{(1)}$ and $\phi^{(2)}$ as given in (3.13) and thus will have the form

$$\psi^{(2)}=\zeta^{-n}\sum_{j=0}^{2n-1}a_j\begin{pmatrix}1\\(-1)^{j+1}\end{pmatrix}\zeta^j+\sum_{2n}^{\infty}\mathbf{a}_j\zeta^j. \tag{3.45}$$

Now since $a=b=c=0$, $\psi^{(2)}e^{-\theta}$ is meromorphic and tends to $\begin{pmatrix}0\\1\end{pmatrix}$ as $\zeta\to\infty$ and therefore must have the form

$$u=\psi^{(2)}e^{-\theta}=\zeta^{-n}\left(u_0+\zeta u_1+\ldots+\zeta^{n-1}u_{n-1}+\begin{pmatrix}0\\1\end{pmatrix}\zeta^n\right),$$

which follows from (3.44). From (3.6),

$$q(x)=2iu_{n-1,1}, \qquad q^2=+2i\frac{\partial}{\partial x}u_{n-1,2} \tag{3.46}$$

where the subscripts 1 and 2 refer to the components of u_{n-1}. Set

$$e^{\theta}=\sum_{j=0}^{\infty}T_j\zeta^j, \qquad \theta=\sum_{j=0}i\omega_{2j+1}\zeta^{2j+1}/2j+1+i\zeta x. \tag{3.47}$$

Expand $\psi^{(2)}$ in a series in ζ,

$$\psi^{(2)}=ue^{\theta}=\left(\frac{u_0}{\zeta^n}+\ldots+\frac{u_{n-1}}{\zeta}+\begin{pmatrix}0\\1\end{pmatrix}\right)\sum_0^{\infty}T_j\zeta^j.$$

We compute the coefficient of ζ^k, and demand that it be orthogonal to $\begin{pmatrix}1\\(-1)^{n+k}\end{pmatrix}$, $k=-n,\ldots,n-1$, as required by (3.45). With the notation $\xi_j=u_1^j+u_2^j$, $\eta_j=u_1^j-u_2^j$, the resulting $2n$ equations can be written (if we use $T_0=1$):

$$\begin{aligned}
&\xi_0=0\\
&\eta_0T_1+\eta_1=0\\
&\xi_0T_2+\xi_1T_1+\xi_2=0\\
&\vdots\\
&\eta_0T_{2n-1}+\ldots+\eta_{n-1}T_n=(-1)^nT_{n-1}.
\end{aligned} \tag{3.48}$$

(3.48) decomposes into two separate systems for ξ, η:

$$\begin{pmatrix}1&&&\\T_2&T_1&1&\\T_4&T_3&T_2&T_1&1\\T_{2n-2}&\cdots&&T_{n-1}\end{pmatrix}\begin{pmatrix}\xi_0\\\vdots\\\xi_{n-1}\end{pmatrix}=\begin{pmatrix}0\\0\\-T_1\\-T_3\\-T_{n-2}\end{pmatrix}\text{ or }\begin{pmatrix}0\\0\\-1\\-T_2\\-T_{n-2}\end{pmatrix} \tag{3.49a}$$

depending on whether n is odd or even;

$$\begin{pmatrix}T_1&1&0&\ldots\\T_3&T_2&T_1&1&\ldots\\T_{2n-1}&\cdots&&T_n\end{pmatrix}\begin{pmatrix}\eta_0\\\vdots\\\eta_{n-1}\end{pmatrix}=\begin{pmatrix}0\\1\\T_2\\T_{n-1}\end{pmatrix}(n\text{ odd}),\begin{pmatrix}0\\T_1\\T_3\\T_{n-1}\end{pmatrix}(n\text{ even}). \tag{3.49b}$$

It is useful to observe that the matrices in (3.49) are Wronskians. Indeed, $\frac{d}{dx}e^{\theta}=i\zeta e^{\theta}$, so that $T_{jx}=iT_{j-1}$. Hence, the derivative of each column is i times the next column.

We now solve (3.49) for ξ_{n-1}, η_{n-1} by Cramer's rule. The denominator determinants are denoted by Δ_+, Δ_-. By the Wronskian property, $\frac{d}{dx}\Delta_{\pm}$ is a single

determinant in which the last column is replaced by its derivative, and it is easy to see that this derivative is in all cases proportional to the right-hand side of (3.49). Specifically,

$$\xi_{n-1} = i\frac{\Delta_{+x}}{\Delta_+}, \quad \eta_{n-1} = -i\frac{\Delta_{-x}}{\Delta_-}. \tag{3.50}$$

From (3.46) we find

$$q = -\left(\ln\frac{\Delta_+}{\Delta_-}\right)_x, \quad q^2 = -(\log\Delta_+\Delta_-)_{xx}. \tag{3.51}$$

These are the solutions determined by Airault [3] through a series of recursion relations. The first two are, for the case $\theta = \frac{4}{3}i\zeta^3 + i\zeta x$:

$$n=1. \quad \Delta_+ = 1, \quad \Delta_- = ix, \quad q = \frac{1}{x}.$$

$$n=2. \quad \Delta_+ = ix, \quad \Delta_- = -\frac{4}{3}i - \frac{ix^3}{3}, \quad q = -\frac{1}{x} + \frac{3x^2}{4+x^3}.$$

Observe that all these solutions satisfy a Riccati equation:

$$q_x + q^2 = -2(\log\Delta_+)_{xx}. \tag{3.52}$$

An interesting class of solutions of the higher-order equations of the Painlevé II family with θ given by (3.47) is obtained when $-2(\ln\Delta_+)_{xx}$ in (3.57) is

$$\frac{n(n-1)}{x^2}.$$

These solutions are related, by a Miura transformation, to certain rational solutions of the KdV hierarchy [25]. To get these solutions, let $\theta = i(\omega_{2k+1}\zeta^{2k+1} + \zeta x)$ and pick $n \leqq k+1$. In that case, all entries of Δ_+ arise from powers of $i\zeta x$ in the expansion (3.47) of e^θ, and Δ_+ is the Wronskian determinant of $1, \frac{(ix)^2}{2!}, \frac{(ix)^4}{4!}, \ldots, \frac{(ix)^{2n-2}}{(2n-2)!}$. This is easily seen to be proportional to $x \cdot x^2 \cdot x^3 \ldots x^{n-1} = x^{\frac{n(n-1)}{2}}$. Hence, $-2(\log\Delta_+)_{xx} = n(n-1)/x^2$.

For each n, only one solution of

$$q_x + q^2 = \frac{n(n-1)}{x^2}, \quad n \geqq 1$$

will also solve a Painlevé equation, and in fact it will solve all the equations of the Painlevé II family (Appendix II) in which the first nonzero power of ζ (besides $i\zeta x$) is at least ζ^{2n-1}.

(iv) Solutions with $\nu = \frac{2n+1}{2}$, $j=0$. It was noted in 3C that this is an exceptional case. We begin with a detailed description of the situation for $\nu = \frac{1}{2}$.

According to (3.15) and (3.16), $j=0$ implies

$$q_x+q^2+\frac{x}{2}=0.$$

Let $q(x)=\dfrac{d}{dx}\ln\chi$; then

$$\chi_{xx}=-\frac{x}{2}\chi,$$

whose general solution is

$$\chi=C_1\mathrm{Ai}(-2^{-1/3}x)+C_2\mathrm{Bi}(-2^{-1/3}x).$$

The solutions of (3.2), (3.3) can be verified to be

$$F(\zeta,x)=\frac{\alpha_0}{2}\zeta^{1/2}\begin{pmatrix}\left(1-\dfrac{iq}{\zeta}\right)\mathrm{Ai}(z)-2^{-1/3}\dfrac{i}{\zeta}\mathrm{Ai}'(z)\\ \left(1+\dfrac{iq}{\zeta}\right)\mathrm{Ai}(z)+2^{-1/3}\dfrac{i}{\zeta}\mathrm{Ai}'(z)\end{pmatrix}$$

and $G(\zeta,x)$, which contains Bi instead of Ai; here

$$z=e^{i\pi}(2^{2/3}\zeta^2+2^{-1/3}x),\ \alpha_0=2^{1/6}\sqrt{\pi}\,e^{i\pi/4}.$$

Using the asymptotic properties of Airy functions, one may verify that

$$\Psi_1=(-iF+G,F-iG),$$

$$\Psi_2=(-2iF,F-iG)=\Psi_1\begin{pmatrix}1&0\\-i&1\end{pmatrix},$$

$$\Psi_3=(-2iF,-F-iG)=\Psi_2\begin{pmatrix}1&-i\\0&1\end{pmatrix},$$

$$\Psi_4=(-iF-G,-F-iG)=\Psi_3\begin{pmatrix}1&0\\-i&1\end{pmatrix},$$

$$\Psi_5=(-iF-G,-2F)=\Psi_4\begin{pmatrix}1&-i\\0&1\end{pmatrix},$$

$$\Psi_6=(iF-G,-2F)=\Psi_5\begin{pmatrix}1&0\\-i&1\end{pmatrix},$$

$$\Psi_7=(iF-G,-F+iG)=\Psi_6\begin{pmatrix}1&-i\\0&1\end{pmatrix},$$

from which it is evident that $a=b=c=-i$. The solution matrix $\Phi(\zeta,x)=(\phi^{(1)}(\zeta,x),\phi^{(2)}(\zeta,x))$ defined by its behavior at the origin $\zeta=0$ is

$$\phi^{(1)}=\frac{(K_1F+K_2G)}{\beta_0(K_1C_2-K_2C_1)},\quad\text{any}\quad K_2/K_1\neq C_2/C_1$$

$$\phi^{(2)}=\frac{2}{\alpha_0}(C_1F+C_2G),$$

where $\beta_0 = i\alpha_0/2^{4/3}\pi$. The indeterminacy described in 3C is seen in the arbitrariness of K_1, K_2.

Finally, note that since Ai(z) and Bi(z) are single valued functions of ζ, the only contribution to the jump matrix J is from the square root $\zeta^{-1/2}$. Hence, $J = -I$. We next describe the general case $\nu = \frac{2n+1}{2}$ $(n \geqq 0)$.

Let $A(x,\zeta)$ be a fixed solution of $A_{xx} = (-\frac{1}{2}x - \zeta^2)A$. The solution $\phi^{(2)}$ of (3.2) is sought in the form

$$\alpha_0 \zeta^{1/2}\left\{A(x,\zeta)\left(C_0 + \frac{C_1}{\zeta} + \ldots + \frac{C_{n+1}}{\zeta^{n+1}}\right) + A_x(x,\zeta)\left(\frac{d_1}{\zeta} + \ldots + \frac{d_{n+1}}{\zeta^{n+1}}\right)\right\}, \tag{3.53}$$

where $C_0 = \begin{pmatrix} 1 \\ (-1)^n \end{pmatrix}$, $d_1 = i\begin{pmatrix} 1 \\ (-1)^{n+1} \end{pmatrix}$, and the coefficients follow the alternating $\begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}$ pattern. By definition, the leading power in $\phi^{(2)}$ is $\zeta^{\frac{2n+1}{2}}$. One must therefore equate to zero the coefficients of $\zeta^{-\frac{2n+1}{2}}, \ldots, \zeta^{\frac{2n-1}{2}}$ in (3.53); this gives $2n+1$ equations for $C_1, \ldots, C_{n+1}, d_2, \ldots, d_{n+1}$, and the coefficients of these equations will involve derivatives $\left.\frac{\partial^j}{\partial \zeta^j} A(x,\zeta)\right|_{\zeta=0}$. The coefficient of $\zeta^{\frac{2n+1}{2}}$ in (3.53) is $e^{\alpha(x)}$ [see (3.13)], and its log derivative is the required $q(x)$.

A concise expression is afforded by

$$q(x) = -\frac{d}{dx}\ln\frac{W(\gamma, \ldots, \gamma^{(n-1)})}{W(\gamma, \ldots, \gamma^{(n)})},$$

where $\gamma(x) = A(x,0)$, and W is the Wronskian determinant. (These solutions were discovered by Airault [3].)

4. Painlevé III; Solution of an Initial Value Problem

4A. Outline

In Appendix I, we show that the equation

$$(xu_x)_x = -4\sinh u, \tag{4.1}$$

a special case of the Painlevé equation of the third kind, is the integrability condition for

$$v_{1\zeta} = \left(-ix + \frac{i}{\zeta^2}\cosh u\right)v_1 + \left(\frac{-xu_x}{2\zeta} + \frac{i}{\zeta^2}\sinh u\right)v_2, \tag{4.2a}$$

$$v_{2\zeta} = \left(\frac{-xu_x}{2\zeta} - \frac{i}{\zeta^2}\sinh u\right)v_1 + \left(ix - \frac{i}{\zeta^2}\cosh u\right)v_2, \tag{4.2b}$$

and

$$v_{1x} = -i\zeta v_1 + qv_2, \quad q = -u_x/2, \tag{4.3a}$$

$$v_{2x} = qv_1 + i\zeta v_2. \tag{4.3b}$$

In Appendix III, we show how these equations relate to the work of Satō et al. [7]. Following the steps described in Sect. 3, we give the general solution of (4.1). The principal difference between (3.2) and (4.2) is that the former has one regular and one irregular singular point, while the latter has two irregular singular points, one at $\zeta=\infty$ and the other at $\zeta=0$.

4B. *Step 1: The Direct Transform and Properties of the Transform Data*

It is straightforward to write down the asymptotic forms $\tilde{\psi}^{(1)}$, $\tilde{\psi}^{(2)}$ of two linearly independent solutions of (4.2) as $\zeta\to\infty$:

$$\tilde{\psi}^{(1)}(\zeta,x)=e^{-i\zeta x-i/\zeta}\left\{\begin{pmatrix}1\\0\end{pmatrix}+\frac{1}{\zeta}\begin{pmatrix}-ixq^2/2+i(1-\cosh u)\\ iq/2\end{pmatrix}+\ldots\right\}, \tag{4.4a}$$

$$\tilde{\psi}^{(2)}(\zeta,x)=e^{i\zeta x+i/\zeta}\left\{\begin{pmatrix}0\\1\end{pmatrix}+\frac{1}{\zeta}\begin{pmatrix}-iq/2\\ ixq^2/2-i(1-\cosh u)\end{pmatrix}+\ldots\right\}. \tag{4.4b}$$

Note that, from (4.1), $xq^2/2+\cosh u-1=\frac{1}{2}\int^x q^2+\text{const}$. We have found it convenient for reasons of symmetry to include the i/ζ term in the exponent of (4.4). At $\zeta=0$, two linearly independent solutions $\phi^{(1)}$, $\phi^{(2)}$ have the asymptotic behavior:

$$\tilde{\phi}^{(1)}(\zeta,x)=e^{-i\zeta x-i/\zeta}\begin{pmatrix}\cosh\frac{u}{2}\\ -\sinh\frac{u}{2}\end{pmatrix}\quad(1+\ldots), \tag{4.5a}$$

$$\tilde{\phi}^{(2)}(\zeta,x)=e^{i\zeta x+i/\zeta}\begin{pmatrix}-\sinh\frac{u}{2}\\ \cosh\frac{u}{2}\end{pmatrix}\quad(1+\ldots). \tag{4.5b}$$

In particular, we note that, if we write $\tilde{\Psi}=(\tilde{\psi}^{(1)},\tilde{\psi}^{(2)})$ and $\tilde{\Phi}=(\tilde{\phi}^{(1)},\tilde{\phi}^{(2)})$, the following relations hold:

$$M\tilde{\Psi}(-\zeta,x)=\tilde{\Psi}(\zeta,x)M,\qquad M\tilde{\Phi}(-\zeta,x)=\tilde{\Phi}(\zeta,x)M \tag{4.6}$$

and

$$N\tilde{\Psi}\left(-\frac{1}{\zeta x},x\right)=\tilde{\Phi}(\zeta,x)\begin{pmatrix}0&1\\-1&0\end{pmatrix}, \tag{4.7}$$

where

$$M=\begin{pmatrix}0&1\\1&0\end{pmatrix}\quad\text{and}\quad N=\begin{pmatrix}\cosh\frac{u}{2}&-\sinh\frac{u}{2}\\ -\sinh\frac{u}{2}&\cosh\frac{u}{2}\end{pmatrix}\begin{pmatrix}0&1\\-1&0\end{pmatrix}. \tag{4.8}$$

Indeed, from Eqs. (4.2), (4.3) one can show that if $\Psi(\zeta, x)=(\psi^{(1)}(\zeta, x), \psi^{(2)}(\zeta, x))$ is a solution, then so is

$$M\Psi(-\zeta, x) \quad \text{and} \quad N\Psi\left(-\frac{1}{\zeta x}, x\right). \tag{4.9}$$

From the asymptotic expressions (4.4), (4.5), we observe that the lines on which the asymptotic solutions change from recessive to dominant and vice versa, the anti-Stokes lines, are $\arg\zeta = n\pi$, $n=-2,-1,0,1,2,\ldots$. Accordingly, we designate the sectors $-2\pi \leqq \arg\zeta < -\pi$, $-\pi \leqq \arg\zeta < 0$, $0 \leqq \arg\zeta < \pi$, $\pi \leqq \arg\zeta < 2\pi$, and $2\pi \leqq \arg\zeta < 3\pi$ as $S_{-2}, S_{-1}, S_1, S_2, S_3$, respectively, and define $\Psi_j(\zeta, x)$ and $\Phi_j(\zeta, x)$, $j=-2,-1,1,2,3$ to be the solutions in these sectors which have the asymptotic behaviors (4.4) and (4.5) on the initial ray of each sector. For convenience, we omit the subscript 1 when referring to the first sector.

We will now show that all the Stokes multiplier matrices for the Ψ_j are $\begin{pmatrix}1 & a\\ 0 & 1\end{pmatrix}$ or $\begin{pmatrix}1 & 0\\ a & 1\end{pmatrix}$ and for the Φ_j $\begin{pmatrix}1 & \bar{a}\\ 0 & 1\end{pmatrix}$ or $\begin{pmatrix}1 & 0\\ \bar{a} & 1\end{pmatrix}$. First, if $x<0$, $\tilde{\psi}^{(1)}$ and $\tilde{\phi}^{(1)}$ are recessive in S_{-2}, S_1, S_3, and dominant in S_{-1}, S_2, $\tilde{\psi}^{(2)}$, $\tilde{\phi}^{(2)}$ are recessive in S_{-1}, S_2 and dominant in S_{-2}, S_1, S_3. Therefore,

$$\begin{aligned}
&\Psi_{-1}=\Psi_{-2}\begin{pmatrix}1 & a_{-2}\\ 0 & 1\end{pmatrix}, \quad \Psi=\Psi_{-1}\begin{pmatrix}1 & 0\\ a_{-1} & 1\end{pmatrix},\\
&\Psi_2=\psi\begin{pmatrix}1 & a\\ 0 & 1\end{pmatrix}, \quad \psi_3=\psi_2\begin{pmatrix}1 & 0\\ a_2 & 1\end{pmatrix}.
\end{aligned} \tag{4.10}$$

From the symmetry (4.9), we have that

$$M\Psi(\zeta e^{-i\pi}, x)M^{-1}=\Psi_2(\zeta, x), \tag{4.11a}$$

$$M\Psi_2(\zeta e^{-i\pi}, x)M^{-1}=\Psi_3(\zeta, x). \tag{4.11b}$$

To see this, observe that the left-hand sides of (4.11) are solutions of (4.2) with the required asymptotic properties on the initial rays of the respective sectors S_2 and S_3. Substitute in (4.11b) from (4.10) and find

$$M\begin{pmatrix}1 & a\\ 0 & 1\end{pmatrix}=\begin{pmatrix}1 & 0\\ a_2 & 1\end{pmatrix}M$$

whence $a_2=a$. By a similar argument, $a_{-2}=a_{-1}=a$. We can also prove that $\bar{a}$ is the only Stokes multiplier connecting the matrices Φ_j. If $x>0$, the Stokes multiplier matrices are the transposes of those for the case $x<0$:

$$\Psi_2=\Psi\begin{pmatrix}1 & 0\\ a & 1\end{pmatrix}, \quad \Psi_3=\Psi_2\begin{pmatrix}1 & a\\ 0 & 1\end{pmatrix}. \tag{4.12}$$

The matrices Ψ and Φ will, in general, not be single-valued. We next calculate the monodromy matrices $J_\pm$ and $\tilde{J}$ defined by the relations

$$\Psi(\zeta e^{2\pi i})=\Psi(\zeta)J_\pm, \quad \Phi(\zeta e^{2\pi i})=\Phi(\zeta)\tilde{J} \tag{4.13}$$

in terms of the Stokes multipliers a and $\bar{a}$. The subscripts $-$, $+$ refer to the cases $x<0$ and $x>0$, respectively. From (4.10) (the case $x<0$), we have

$$\Psi_3(\zeta e^{2\pi i})\begin{pmatrix} 1 & -a \\ -a & 1+a^2 \end{pmatrix} = \Psi(\zeta) J_- . \tag{4.14}$$

Now let $\zeta \in S_1$ tend to infinity; then

$$J_- = \begin{pmatrix} 1 & -a \\ -a & 1+a^2 \end{pmatrix} \tag{4.15}$$

since both $\Psi_3(\zeta e^{2\pi i})$ and $\Psi(\zeta)$ tend to $\tilde{\Psi}(\zeta)$. Similarly, if $x>0$,

$$J_+ = \begin{pmatrix} 1+a^2 & -a \\ -a & 1 \end{pmatrix}. \tag{4.16}$$

Also,

$$\bar{J} = \begin{pmatrix} 1 & -\bar{a} \\ -\bar{a} & 1+\bar{a}^2 \end{pmatrix}. \tag{4.17}$$

Finally, we specify the connection matrix A between Ψ and Φ,

$$\Psi = \Phi A, \quad A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}. \tag{4.18}$$

From the normalizations of Ψ and Φ, $\alpha\delta - \beta\gamma = 1$. Let us now derive relations between $\bar{J}$, J_-, A, for the case $x<0$. Let $\zeta \to \zeta e^{2\pi i}$ in (4.18), and use (4.13) to find

$$AJ_- = \bar{J}A. \tag{4.19}$$

Since J_- and $\bar{J}$ are similar, their traces are equal; from (4.15) and (4.17), $\bar{a} = sa$, $s = \pm 1$. Comparison of the other entries in (4.19) gives

$$\beta = s\gamma, \quad \alpha = s(\delta + a\gamma). \tag{4.20}$$

Among the transform data, therefore, there are only two independent constants which we take to be a and $\frac{\gamma}{\delta}$. The reason for the sign parameter s is discussed later.

Similarly, if $x>0$, we find

$$AJ_+ = \bar{J}A, \tag{4.21}$$

from which we have

$$\bar{a} = sa, \quad \alpha = s\delta, \quad \beta = s(\gamma + a\delta). \tag{4.22}$$

4C. Step 2: The x-Dependence of the Transform Data

Arguments which exactly parallel those given in Sect. 3 show that:

(1) If (4.2) and (4.3) and therefore (4.1) hold, the *transform data*, which consist of the Stokes multiplier matrix $\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$, the monodromy matrices J_+, J_-, $\bar{J}$ and the connection matrix A, are independent of x.

(2) Conversely, if the transform data are independent of x, and if we specify functions Ψ_j, Φ_j to be (a) connected as in (4.10), (4.13), and (4.18), with (b) the symmetry properties stemming from (4.9), and (c) with the asymptotic expansions (4.4) and (4.5), then the functions Ψ_j, Φ_j satisfy differential equations in ζ and x which have precisely the form (4.2), (4.3). The coefficients in the equations are directly related to the coefficients in the specified asymptotic expansions.

4D. Step 3: The Inverse Transform

Here we show how to derive, at any fixed position x, a singular integral equation for the matrix function $\Psi(\zeta, x)$ from which one can, in principle, construct the function $\Psi(\zeta, x)$ and the Eq. (4.2) which it satisfies. Since the coefficients in this equation are functions of the solution $u(x)$ of (4.1), we have therefore found $u(x) (\mathrm{mod} 2\pi i)$ for any given x. We first look at the case $x<0$. Consider

$$\int_{C_1} \frac{\psi^{(1)} e^{\theta}}{\xi - \zeta} d\xi \quad \text{with} \quad \theta = i\xi x + \frac{i}{\xi}. \tag{4.23}$$

The reader should refer to Fig. 4 for the definition of the contours. The contour Γ has a large radius, the contour γ a small one. Let $\zeta \in S_1$, outside the circle designated by λ.

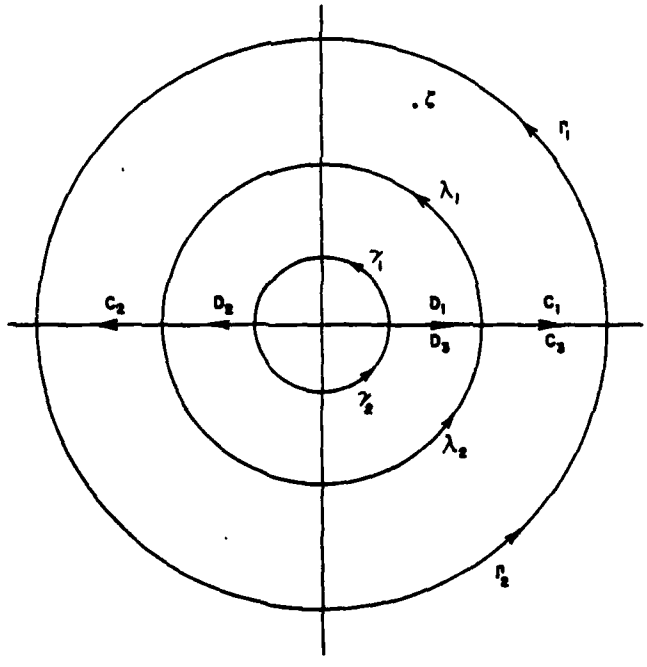


Fig. 4. The contours in the ζ-plane for the inverse problem for Painlevé III, $x<0$

Using Cauchy's theorem, we obtain

$$\int_{C_1} \frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi+\pi i\begin{pmatrix}1\\0\end{pmatrix}=2\pi i\psi^{(1)}e^{\theta}+\int_{C_2}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi+\int_{\lambda_1}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi. \tag{4.24}$$

But $\psi^{(1)}=\psi_2^{(1)}$ and

$$\int_{C_2}\frac{\psi_2^{(1)}e^{\theta}}{\xi-\zeta}d\xi+\pi i\begin{pmatrix}1\\0\end{pmatrix}=\int_{C_3}\frac{\psi_2^{(1)}e^{\theta}}{\xi-\zeta}d\xi+\int_{\lambda_2}\frac{\psi_2^{(1)}e^{\theta}}{\xi-\zeta}d\xi. \tag{4.25}$$

We add the two equations to obtain

$$\psi^{(1)}e^{\theta}=\begin{pmatrix}1\\0\end{pmatrix}+\frac{a}{2\pi i}\int_{C_1}\frac{\psi^{(2)}e^{\theta}}{\xi-\zeta}d\xi-\frac{1}{2\pi i}\int_{\lambda}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi, \tag{4.26}$$

upon making use of the relations $\psi_2^{(1)}=\psi^{(1)}$ and [from (4.13)] $\psi^{(1)}(\zeta e^{2\pi i})=\psi^{(1)}(\zeta)-a\psi^{(2)}(\zeta)$. λ is the union of the two contours λ_1 and λ_2. We will see later that, in the linear limit, the second term on the RHS of (4.26) gives rise to the solution $K_0(4\sqrt{-x})$ of the linearized (4.1), and the third term, which will be associated with the off-diagonal terms in the connection matrix A, gives rise to the solution $I_0(4\sqrt{-x})$ of the linearized (4.1).

Using (4.18), we find

$$\int_{\lambda}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi=\frac{1}{\delta}\int_{\lambda_1}\frac{\phi^{(1)}e^{\theta}}{\xi-\zeta}d\xi+\frac{\gamma}{\delta}\int_{\lambda_1}\frac{\psi^{(2)}e^{\theta}}{\xi-\zeta}d\xi+\int_{\lambda_2}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi,$$

which, because $\phi^{(1)}e^{\theta}$ is bounded on the contours γ_1 and γ_2 whose radius is arbitrarily small, is equal to

$$-\frac{1}{\delta}\int_{D_1}\frac{\phi^{(1)}e^{\theta}}{\xi-\zeta}d\xi+\frac{1}{\delta}\int_{D_2}\frac{\phi^{(1)}e^{\theta}}{\xi-\zeta}d\xi+\frac{\gamma}{\delta}\int_{\lambda_1}\frac{\psi^{(2)}e^{\theta}}{\xi-\zeta}+\int_{\lambda_2}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi.$$

But $\phi_2^{(1)}=\phi^{(1)}$, and continuing the second integral into the lower half plane and using the relations $\phi^{(1)}(\zeta e^{2\pi i})=\phi^{(1)}(\zeta)-sa\phi^{(2)}(\zeta)$ and $\psi^{(1)}-\frac{1}{\delta}\phi^{(1)}=\frac{\gamma}{\delta}\psi^{(2)}$, we find

$$\int_{\lambda}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi=-\frac{sa}{\delta}\int_{D_1}\frac{\phi^{(2)}e^{\theta}}{\xi-\zeta}d\xi+\frac{\gamma}{\delta}\int_{\lambda}\frac{\psi^{(2)}e^{\theta}}{\xi-\zeta}d\xi.$$

Finally using $\phi^{(2)}=-\beta\psi^{(1)}+\alpha\psi^{(2)}$ and replacing α and β from (4.20), we find

$$\begin{aligned}\psi^{(1)}(\zeta)e^{\theta}=\begin{pmatrix}1\\0\end{pmatrix}&+\frac{a}{2\pi i}\int_0^{\infty}\frac{\psi^{(2)}(\xi)e^{\theta}}{\xi-\zeta}-\frac{\gamma}{2\pi i\delta}\int_{\lambda}\frac{\psi^{(2)}(\xi)e^{\theta}}{\xi-\zeta}d\xi\\&+\frac{a^2\gamma}{\delta}\int_{D_1}\frac{\psi^{(2)}(\xi)e^{\theta}}{\xi-\zeta}d\xi-\frac{a\gamma}{2\pi i\delta}\int_{D_1}\frac{\psi^{(1)}(\xi)e^{\theta}}{\xi-\zeta}d\xi,\end{aligned} \tag{4.27}$$

which together with $M\psi^{(2)}(\zeta e^{i\pi})=\psi^{(1)}(\zeta)$ defines the solution matrix $\Psi(\zeta,x)$.

(4.27) is a singular integral equation in which the parameter $x<0$ only appears in the exponent $\theta(\xi)=i\xi x+\frac{i}{\xi}$. The initial conditions $u(x_0)$, $u_x(x_0)$ are represented by the parameters a and $\frac{\gamma}{\delta}$. Notice that the sign parameter s has disappeared

altogether and plays no role in the determination of the solution $u(x)$ of (4.1). Its appearance in the transform data is a consequence of the symmetry

$$u \to -u, \quad v \to \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} v \quad \text{or} \quad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} v,$$

present in (4.2), (4.3). This is reflected in (4.27): note that if $\psi^{(1)}\left(\zeta, a, \frac{\gamma}{\delta}\right)$, $\psi^{(2)}\left(\zeta, a, \frac{\gamma}{\delta}\right)$ is a solution, then so is $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\psi^{(1)}\left(\zeta, -a, -\frac{\gamma}{\delta}\right)$ and $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\psi^{(2)}\left(\zeta, -a, -\frac{\gamma}{\delta}\right)$. Observe that this transformation also satisfies the symmetry condition, since $M\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} M$.

We now apply the linear limit to (4.27) by taking a and $\frac{\gamma}{\delta}$ to be small, neglecting all quadratic and cubic terms in the parameters and approximating $\psi^{(2)}$ by $\begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{\theta}$ to leading order. Then, using

$$q(x) = \frac{-u_x}{2} = -\lim_{\zeta \to \infty} 2i\zeta \psi_{,2}^{(1)} e^{-\theta} = \lim_{\zeta \to \infty} 2i\zeta \psi_{,1}^{(2)} e^{\theta} \tag{4.28}$$

(the subscripts refer to the components), we find that

$$q(x) = \frac{a}{\pi} \int_0^{\infty} e^{2i\xi x + 2i/\xi}\, d\xi - \frac{\gamma}{\pi\delta} \int_{\lambda} e^{2i\xi x + 2i/\xi}\, d\xi.$$

Integrating with respect to x, setting $\xi = \frac{1}{\sqrt{-x}} e^{\phi}$ in the first integral, taking λ to be the circle $|\xi| = \frac{1}{\sqrt{-x}}$ in the second and setting $\xi = -\frac{1}{\sqrt{-x}} e^{i\phi}$, we find

$$\begin{aligned} u(x) &= \frac{ia}{\pi} \int_0^{\infty} \cos(4\sqrt{-x}\sinh\phi)\, d\phi + \frac{\gamma}{\pi\delta} \int_0^{2\pi} e^{4\sqrt{-x}\sin\phi}\, d\phi \\ &= \frac{ia}{\pi} K_0(4\sqrt{-x}) + \frac{2\gamma}{\delta} I_0(4\sqrt{-x}). \end{aligned} \tag{4.29}$$

We observe that if u is real, a is pure imaginary and $\frac{\gamma}{\delta}$ is real. This may be proved as follows. If u is real, $M\psi^*(\zeta^*, x)$ is a solution if $\psi(\zeta, x)$ is. From the asymptotic behaviors we have that $M\Psi_2^*(\zeta^*) = \Psi_{-1}(\zeta)M$ and $M\Psi^*(\zeta^*) = \Psi(\zeta)M$. But $\Psi_{-1} = \Psi\begin{pmatrix} 1 & 0 \\ -a & 1 \end{pmatrix}$ and $\Psi_2 = \Psi\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$ and from these relations we find $a^* = -a$. The reality of $\frac{\gamma}{\delta}$ follows from similar arguments.

The solution of (4.1) studied by Satō et al. corresponds to the case $\frac{\gamma}{\delta} = 0$; that is, the inside-outside connection matrix A is the identity ($s > 0$). From the remarks in example 2 of Sect. 3F, the reader may convince himself that this is also the case

which can be treated by inverse scattering theory. In other words, the function $\psi^{(1)}e^{i\zeta}$ is analytic in $\mathrm{Im}\,\zeta>0$ and in particular bounded as $\zeta\to 0$ only if $\psi^{(1)}$ contains no component of $\phi^{(2)}$, which from (4.18) and (4.20) means that $\beta=\gamma=0$.

When $x>0$, the derivation of the singular integral equation is simpler and can be accomplished without the contour λ. Let $\zeta\in S_2$.

Use Cauchy's theorem to express $\int_{C_1}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi$ in terms of an integral along C_2. Because this will involve an integral along γ_1 on which contour $\phi^{(2)}e^{\theta}$ is

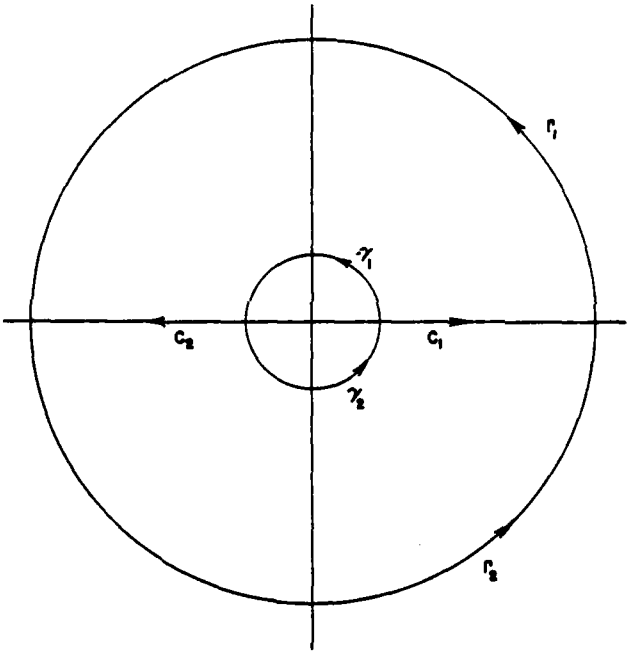


Fig. 5. The contours in the ζ-plane for the inverse problem for Painlevé III, $x>0$

exponentially large, we must subtract an appropriate amount of $\psi^{(2)}$ from $\psi^{(1)}$ in order that the $\phi^{(2)}$ component be eliminated:

$$\int_{C_1}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi=\int_{C_1}\frac{\psi^{(1)}-\frac{\gamma}{\delta}\psi^{(2)}}{\xi-\zeta}e^{\theta}\,d\xi+\frac{\gamma}{\delta}\int_{C_1}\frac{\psi^{(2)}e^{\theta}}{\xi-\zeta}d\xi$$

$$=\int_{C_2}\frac{\psi^{(1)}-\frac{\gamma}{\delta}\psi^{(2)}}{\xi-\zeta}e^{\theta}\,d\xi-\pi i\begin{pmatrix}1\\0\end{pmatrix}+\frac{\gamma}{\delta}\int_{C_1}\frac{\psi^{(2)}e^{\theta}}{\xi-\zeta}d\xi.$$

But $\psi^{(2)}=\psi_2^{(2)}$, $\psi^{(1)}=\psi_2^{(1)}-a\psi_2^{(2)}$. Thus we have

$$\int_{C_1}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi=\int_{C_2}\frac{\psi_2^{(1)}-\left(a+\frac{\gamma}{\delta}\right)\psi_2^{(2)}}{\xi-\zeta}e^{\theta}\,d\xi-\pi i\begin{pmatrix}1\\0\end{pmatrix}+\frac{\gamma}{\delta}\int_{C_1}\frac{\psi_2^{(2)}e^{\theta}}{\xi-\zeta}d\xi.$$

Now $\psi_2^{(1)}e^{\theta}$, $\phi^{(1)}e^{\theta}$, $\phi^{(2)}e^{\theta}$ are bounded on Γ_2 and γ_2 respectively, and thus $\int_{C_1}\frac{\psi_2^{(1)}e^{\theta}}{\xi-\zeta}$ can be continued to $\arg\zeta=2\pi$. On this line, $\psi_2^{(1)}(\zeta e^{2\pi i})=\psi^{(1)}(\zeta e^{2\pi i})+a\psi^{(2)}(\zeta e^{2\pi i})$ $=(1+a^2)\psi^{(1)}(\zeta)-a\psi^{(2)}(\zeta)+a\psi^{(2)}(\zeta)-a^2\psi^{(1)}(\zeta)=\psi^{(1)}(\zeta)$. Thus the integrals $\int_{C_1}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi$ on $\arg\xi=0$ and $\int_{C_1}\frac{\psi_2^{(1)}e^{\theta}}{\xi-\zeta}d\xi$ on $\arg\xi=2\pi$ cancel. We find

$$\psi_2^{(1)}(\zeta)e^{\theta}=\begin{pmatrix}1\\0\end{pmatrix}+\frac{a}{2\pi i}\int_0^{-\infty}\frac{\psi_2^{(2)}(\zeta)e^{\theta}}{\xi-\zeta}d\xi-\frac{\gamma}{2\pi i\delta}\int_{-\infty}^{\infty}\frac{\psi_2^{(2)}(\zeta)e^{\theta}}{\xi-\zeta}d\xi. \tag{4.30}$$

Taking the linear limit, we obtain (as expected)

$$u(x)=aH_0^{(2)}(4\sqrt{x})+\frac{\gamma}{\delta}J_0(4\sqrt{x})\quad\text{with}\quad x>0, \tag{4.31}$$

the solution of the linearized (4.1).

We remark, in conclusion, that just as the contours involved in the singular integral equations for finding the solution of Painlevé II are those used in the representations of Airy functions, so the contours appropriate for (4.1) are familiar from the representations of Bessel functions.

5. Hamiltonian Systems

Another property shared by monodromy – and spectrum – preserving deformations is that the deformation equations can be written as completely integrable Hamiltonian systems. The modified Korteweg-deVries equation (2.9) can be written $q_t=\partial/\partial x\frac{\delta H}{\delta q}$, with $H=\frac{1}{2}\int_{-\infty}^{\infty}(q_x^2+q^4)dx$. It can be shown [24] that the mapping to scattering data is canonical and that suitable combinations of the scattering data are action – angle variables. In contrast, Eqs. (1.2) and (1.4) can each be written as a four-dimensional Hamiltonian system with x playing the role of time.

Consider (1.2) and let

$$p_1=q_x, q_1=q, p_2=\tfrac{1}{2}\int_{x_0}^{x}q^2\,dx, q_2=x. \tag{5.1}$$

Then

$$H(p_1,q_1,p_2,q_2)=\tfrac{1}{2}p_1^2-\tfrac{1}{2}q_2q_1^2+p_2-\tfrac{1}{2}q_1^4+\nu q_1 \tag{5.2}$$

is the Hamiltonian. It can be verified directly that H is constant and that Hamilton's equations are satisfied by (1.2). But we have already shown that each

piece of monodromy data $(a, b, c, j, \alpha, \beta, \gamma, \delta)$ is also a constant of the motion and thus

$$a_x = \{H, a\} = 0, \tag{5.3}$$

where $\{H, a\}$ is the usual Poisson bracket. Therefore we have two independent constants of the motion in involution and thus the system is exactly integrable.

Similarly, for (1.4), set

$$p_1 = xu_x, q_1 = u, p_2 = \tfrac{1}{2}\int_{x_0}^{x} u_x^2\, dx, q_2 = x \tag{5.4}$$

and

$$H(p_1, q_1, p_2, q_2) = \frac{1}{2q_2} p_1^2 + p_2 + 4\cosh q_1 . \tag{5.5}$$

Again any one of the monodromy data is an independent constant of the motion which commutes with H and the system is again integrable.

The Hamiltonian property carries through for any one of the members of the Painlevé II family discussed in Appendix II.

We have not as yet examined in what sense the transformation from x, q, q_x to the monodromy data is canonical nor have we identified the appropriate angle variables in the new coordinates.

6. Further Discussion

There is a vast literature on monodromy problems, starting with the investigations of Riemann on analytic functions defined by their branching properties up to the algebraic geometry studies of recent years (for an introductory survey, see [ ]). We have not found modern mathematical work which has concerned itself with the "linearizability" properties of the nonlinear deformation equations which express the monodromy preservation property, excepting, of course, the papers by Satō et al., and the recent thesis of Ueno.

It has been mentioned repeatedly that the work of Satō et al. [7] not only provided the stimulus for the present paper, but also suggests many further problems about singular points and deformation theory. After completing a first draft of this paper, we learned that K. Ueno of the Kyoto University RIMS had carried out investigations [27] which overlap ours to some extent. He derives deformation equations for nxn systems with singular points of various ranks at $\zeta = 0, \infty$, and establishes results of the kind contained in Sect. 3C above. In other respects, his work and ours are complementary. He has not yet considered the inverse problem; on the other hand, he had found a remarkable generalization of our rational-solution example in 3F. By including *apparent singularities* in equations such as (3.2) or (4.2), i.e., by including a term

$$\sum_{1}^{N} \frac{H_j}{\zeta - \alpha_j}$$

with certain assumptions about the indicial roots at the α_j, *he can recover the N-soliton solutions of MKdV, sine-Gordon, etc.* (an appropriate limit, in which all

$\alpha_j \to 0$, should yield the rational solutions). This is the clearest evidence yet that soliton and self-similar solutions will fit into a unified framework.

As we expect that monodromy-preserving deformations will play an increasingly important role in diverse problems of physics and applied mathematics, we now give a brief description of some of the other papers we have found particularly useful. We also outline various reformulations of the inverse problem discussed in earlier sections, in order to emphasize the similarities with well-established and successful approaches to other inverse spectral problems.

The pioneering work on irregular-singular Riemann-Hilbert problems is that of Birkhoff [16, 17]. In [16], Birkhoff analyzed the behavior of certain canonical equations near an irregular singular point, counted the number of characteristic parameters of the asymptotic solutions and showed it to equal the number of adjustable coefficients of the differential equation, and hence suggested the possibility of solving the inverse problem. In our formulation of the problem, we have been guided more by inverse scattering than by Birkhoff's solution, which we have found to be inaccessible on certain points.

While the structure of solutions near an irregular singular point has been the object of many studies since Birkhoff's influential papers, the problem of finding equations with prescribed Stokes multipliers is encountered only rarely. The extensive work of Sibuya [28] should be mentioned in this connection; its relevance to deformation problems is still unexplored. The work of Satō et al. is obviously relevant to problems of irregular singular monodromy; except for a brief comment in [7, III], however, they have not developed this aspect of their theory. Their work shows that irregular singular points of ordinary differential equations can be transformed to regular singular points of partial differential equations. Indeed, the idea suggested by their approach, to use partial differential equations to find representations of and to investigate solutions of ordinary differential equations with irregular singular points, has not yet been explored and seems to be a fruitful area for study.

The irregular-singular Riemann-Hilbert problem fits quite naturally into the inverse spectral approaches developed for the solution of nonlinear evolution equations. We briefly describe the various connections.

In [29], Zakharov and Shabat solve the inverse scattering problem as follows. Define solutions $\hat{\phi}$, $\hat{\psi}$ of

$$v_{1x} + i\zeta v_1 = qv_2$$

$$v_{2x} - i\zeta v_2 = qv_1$$

by

$$\hat{\phi} \sim \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta x}, x \to -\infty, \hat{\psi} \sim \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta x}, x \to +\infty. \tag{6.1}$$

Then $\hat{\phi} = a\bar{\hat{\psi}} + b\hat{\psi}$, where $\bar{\hat{\psi}} = (\hat{\psi}_2^*(x, \zeta^*), \hat{\psi}_1^*(x, \zeta^*))^T$ (q is real). Set

$$\Phi(\zeta, x) = \begin{cases} a^{-1}(\zeta)\hat{\phi}(x, \zeta)e^{i\zeta x}, \operatorname{Im}\zeta > 0 \\ \begin{pmatrix} \hat{\psi}_2^*(x, \zeta^*) \\ \hat{\psi}_1^*(x, \zeta^*) \end{pmatrix} e^{i\zeta x}, \operatorname{Im}\zeta < 0. \end{cases} \tag{6.2}$$

From the jump of Φ across the real axis,

$$\eta(\xi, x) \equiv \Phi(\xi + i0, x) - \Phi(\xi - i0, x), \tag{6.3}$$

reconstruct Φ by a Cauchy integral; this leads to a system of singular integral equations for $\hat{\psi}$.

One can reword this approach. For $\mathrm{Im}\zeta > 0$, $\Phi(\zeta, x)$ has an asymptotic expansion $\begin{pmatrix}1\\0\end{pmatrix} + \sum \Phi_n \zeta^{-n}$. The analytic continuation of this expansion to $\mathrm{Im}\zeta < 0$ is not the expansion of an analytic continuation of Φ from $\mathrm{Im}\zeta > 0$; rather, it is the expansion of Φ as defined by (6.2) in $\mathrm{Im}\zeta < 0$. The jump (6.3), $\eta(\zeta, x)$, is given by

$$\frac{b(\zeta)}{a(\zeta)} e^{i\zeta x} \hat{\psi}(\zeta, x)\,; \tag{6.4}$$

it is analogous to, and has the form of, a product of a "Stokes multiplier" and a "recessive solution". In contrast to the cases discussed in the present paper, however, $\frac{b}{a}(\zeta)$ has a very general dependence on ζ. The reason is that $\hat{\phi}(\zeta, x)$ and $\hat{\psi}(\zeta, x)$ do not satisfy differential equations in ζ with polynomial coefficients. When they do [if q evolves in a self-similar manner, or if the t dependence is dropped in (A.4)], the reflection coefficient has the form of the product of $\frac{b}{a}(0)$ and $\exp 8i\zeta^3/3$. Appropriate scaling of the solutions $\hat{\phi}$ and $\hat{\psi}$ with the factors $\exp \pm 4i\zeta^3/3$ then shows the jump (6.4) to be the product of a Stokes multiplier and a recessive solution.

Zakharov [30] has recently propounded an extension of inverse-scattering ideas, based on Riemann-Hilbert problems. This generalizes the Zakharov-Shabat method sketched above, in that the eigenfunctions are reconstructed from prescribed jumps across arbitrary closed curves in the ζ-plane. Whereas Zakharov has formulated his new method in extreme generality, the only solutions published so far have a soliton character, in that they correspond to point spectra of certain operators. Our paper, from this perspective, provides the first other type of solution derivable by ideas related to Zakharov's and not accessible to inverse scattering. The curves across which the "jumps" are prescribed are more complicated in our examples than is envisaged in [30].

It is interesting that the irregular monodromy preserving deformations relate as naturally to periodic inverse spectral theory as they do to scattering theory. The approach developed by Krichever ([31], see also Novikov [32]) is particularly relevant. Give

i) a Riemann surface S, of genus g, with a point called ∞,

ii) a nonspecial divisor $P_1 + \ldots + P_g$,

and seek a function $\psi(x, t, y, P)$, meromorphic for $P \in S$, except at $P = \infty$, such that

iii) $\psi \sim \exp(kx + R(k)t + Q(k)y)$ near $k = \infty$ $\left(\frac{1}{k}\right.$ is the local parameter at $\left.\infty\right)$,

iv) $\psi(0, 0, 0, P) = 1$,

v) the poles of ψ are at $P_1, \ldots, P_g$, independently of x, t, y.

There is a unique such function; it is analogous to the Bloch eigenfunction of a differential operator with periodic coefficients. The coefficients in the asymptotic expansion of ψ at $k=\infty$ are functions of x, t, y, from which solutions of a certain Zakharov-Shabat equation

$$L_t - M_y = [L, M]$$

can be constructed.

These requirements determine a holomorphic line bundle over the Riemann surface S. The transition functions of this bundle are defined in the finite part of S from functions which locally realize the divisor (ii), and at ∞ by the function (iii). The bundle depends on x, t, y because of this construction at ∞, and it turns out that the variation with x, t, y is linear in the space of moduli of line bundles, i.e. the Jacobian variety of R. This leads to the well-known linearization of isospectral flows on the Jacobian, by means of θ-functions.

The Stokes multiplier problem suggests an analogous construction. In connection with (3.2), for example, we cover the complex plane by six slightly overlapping sectors. To each point, we attack the group $SL(2, \mathbb{C})$. The transition functions of an $SL(2, \mathbb{C})$ principal bundle are defined by the Stokes matrices (3.12), with some modifications to incorporate a $\nu \neq 0$ branch point or a $\nu = n$ pole. This construction is not really useful until the discontinuity of Stokes jumps at ∞ can be resolved; only a bundle over a non-contractible surface will carry nontrivial geometric information. Nevertheless, the idea is suggestive and is currently under investigation. We expect that the bundle will vary with the coefficients $x, t, \ldots$ in the exponent $\theta(\zeta)$; this geometric interpretation should make quite clear that in any deformation problem, the coefficients of $\theta(\zeta)$ are to be regarded as independent variables. In particular, our method is applicable to classes of nonlinear, non-autonomous equations in several independent variables (some examples are given in a report to appear in the Proceedings of the 1979 US-USSR Symposium on Solitons held in Kiev).

Acknowledgments. The authors are grateful for support by NSF grants MCS75-07548 A01, MPS75-07530, MCS79-03533, ONR grant N00014-76-C-0867, and US Army Contract DAAG29-78-G-0059.

H. Flaschka thanks Prof. M. Satō, Drs. T. Miwa and M. Jimbo, and Mr. K. Ueno for many stimulating discussions; these were made possible by a US-Japan Cooperative Science Program grant, and by the kind hospitality of the Kyoto University RIMS in January 1978 and April 1979.

Appendix I

Here we quote results published in reference [33]. It was shown there that the most general equation with x-dependent coefficients which can be solved by the inverse scattering transform associated with the n^{th} order system

$$V_x = (\zeta R_0 + P(x,t))V, \tag{A.1}$$

$R_0 = (\beta_j \delta_{ij})$, $\beta_i \neq \beta_j$, $P = (p_{ij})$, $p_{ii} = 0$, V an n-vector, is

$$G(D_R, t)P_t = \Omega(D_R, t)[C, P] + F(D_R, t)x[R_0, P]. \tag{A.2}$$

In (A.2), G, Ω, and F are entire functions of an integro-differential operator D_R, C is a diagonal matrix and the bracket denotes the commutator. The operator D_R,

which is only applied to off-diagonal matrices $H_F(h_{lm})$, $h_{ll}=0$, is defined as follows. First introduce the matrix $(H_F)_R=\left(\frac{h_{lm}}{i(\beta_l-\beta_m)}\right)$, the solution of $[R_0,(H_F)_R]=H_F$. The diagonal counterpart of H_F is defined to be $H_D=-\int_\infty^x [H_F,P]_D\,dy$. Secondly, define the operator D acting on H_F to be

$$DH_F=\frac{\partial}{\partial x}H_F+[H_F,P]_F+\left[\int_x^\infty [H_F,P]_D\,dy,P\right]$$
$$=\frac{\partial}{\partial x}H+[H,P],\ H=H_F+H_D. \tag{A.3}$$

The subscripts F and D in (A.3) denote the off-diagonal and diagonal parts of the designated matrices, respectively. Finally, $D_RH_F=D(H_F)_R$. We use the lower limit ∞ in the definition of H_D in a formal manner, to indicate that we simply ignore the constant of integration. The corresponding t and ζ dependence of $V(x,t,\zeta)$ is given by

$$G(\zeta,t)V_t+F(\zeta,t)V_\zeta=(xFR_0+(Q+S-T))V. \tag{A.4}$$

The quantities Q, S, and T are defined as follows. Let $\Omega(D_R,t)=\sum_0 \omega_m(t)D_R^m$, then $Q=\sum\omega_mQ^{(m)}$ where $Q^{(m)}=Q_m+Q_{m-1}\zeta+\ldots+Q_1\zeta^{m-1}+C\zeta^m$, and

$$Q_{SF}=(D_R^{S-1}[C,P])_R,\quad S=1,\ldots,m,$$

$$Q_{SD}=\int_x^\infty [Q_{SF},P]_D\,dy,$$

$$Q_S=Q_{SF}+Q_{SD}.$$

Define the sequences $\{T_k\}$, $\{S_l\}$ in a similar way;

$$T_{kF}=(D_R^{k-1}P_t)_R,\ T_{kD}=\int_x^\infty [T_{kF},P]_D\,dy,\ T_k=T_{kF}+T_{kD},\ k\geqq 1,$$

$$S_{lF}=(D_R^{l-1}x[R_0,P])_R,\ S_{lD}=\int_x^\infty [S_{lF},P]_D\,dy,\ S_l=S_{lF}+S_{lD},\ l\geqq 1.$$

Then if $G=\sum g_k(t)D_R^k$ and $F=\sum f_l(t)D_R^l$,

$$T=\sum g_k(t)T^{(k)},\quad T^{(k)}=T_k+T_{k-1}\zeta+\ldots+T_1\zeta^{k-1},$$

$$S=\sum f_l(t)S^{(l)},\quad S^{(l)}=S_l+S_{l-1}\zeta+\ldots+S_1\zeta^{l-1}.$$

Example 1. Let us look at the case where $n=2$, $G=1$, $F=f_1D_R$, $\Omega=\omega_3D_R^3$, $P=q(Y_2+Y_3)$, $R_0=C_0=-iY_1$, $Y_1=\begin{pmatrix}1&0\\0&-1\end{pmatrix}$, $Y_2=\begin{pmatrix}0&1\\0&0\end{pmatrix}$, $Y_3=\begin{pmatrix}0&0\\1&0\end{pmatrix}$. Then $T=0$; $S=f_1S^{(1)}=f_1S_1=f_1xq(Y_2+Y_3)$; $Q=\omega_3Q^{(3)}=\omega_3(Q_3+Q_2\zeta+Q_1\zeta^2-iY_1\zeta^3)$; $Q_1=q(Y_2+Y_3)$, $Q_2=\frac{i}{2}q_x(Y_2-Y_3)-\frac{i}{2}q^2Y_1$, $Q_3=\frac{-1}{4}(q_{xx}-2q^3)(Y_2+Y_3)$. Then,

the Eq. (A.2) is

$$q_t = \frac{-\omega_3}{4}(q_{xx} - 2q^3)_x + f_1(xq)_x. \tag{A.5}$$

Note that the Painlevé equation can be found by either taking $f_1 = 0$ and looking for the solution $q(x,t) = \frac{1}{(3t)^{1/3}}\hat{q}\left(\frac{x}{(3t)^{1/3}}\right)$ or by simply setting $f_1 = 1$ and ignoring the t-dependence altogether. In fact, all the self-similar solutions of evolution equations with $F = 0$ can be obtained by taking appropriate functions F and ignoring the t-dependence. Moreover, there are many equations which are solvable by the method introduced in this paper and which are not the result of a self-similar transformation on a solvable evolution equation. Ignoring the t-dependence, the ζ and x equations for V are precisely (3.2) and (3.3) respectively. Note that the term $Q_3 + S_1$ vanishes by virtue of (3.1).

Example 2. We present an equation solvable by our method which is not obtained as a self-similar limit of an evolution equation. Let $R_0 = C = -iY_1$, $P = qY_2 + rY_3$, $F = \frac{1}{4}$, $\Omega = D_R^2$, $r = -q^*$; we obtain $q_{xx} + 2q^2q^* - xq = 0$. When our method is applied to this equation, the details are similar to those introduced in Sect. 3.

Example 3. We show that (4.1) is the integrability condition for (4.1), (4.2). Take $n = 2$, $\Omega = 1$, $G = 0$, $F = -D_R^2$, $R = C = -iY_1$, $P = q(Y_2 + Y_3)$ and $q = -\frac{u_x}{2}$. Then Eq. (A.2) reads

$$(xu_x)_x = -4\sinh u. \tag{A.5}$$

We also find that $Q = -iY_1$, $S = -S_2 - \zeta S_1 = \frac{xu_x}{2}(Y_2 + Y_3) - i\sinh u(Y_2 - Y_3)$ $+ i(1 - \cosh u)Y_1$. Therefore, (A.1) is

$$\begin{aligned} v_{1x} &= -i\zeta v_1 + qv_2 \\ v_{2x} &= qv_1 + i\zeta v_1 \end{aligned} \tag{A.6}$$

and (A.4) is

$$\begin{aligned} v_{1\zeta} &= \left(-ix + \frac{i}{\zeta^2}\cosh u\right)v_1 + \left(\frac{i}{\zeta^2}\sinh u - \frac{1}{2\zeta}xu_x\right)v_2 \\ v_{2\zeta} &= \left(-\frac{i}{\zeta^2}\sinh u - \frac{1}{2\zeta}xu_x\right)v_1 + \left(ix - \frac{i}{\zeta^2}\cosh u\right)v_2. \end{aligned} \tag{A.7}$$

Appendix II: The Painlevé II Family

It is well known [23] that the MKdV equation is only one of an infinite family of equations, all of which are solvable by the scattering problem (2.12). The equations of this family can be derived from Hamiltonians [24],

$$q_t = \frac{\partial}{\partial x}\frac{\delta H_{2j+1}}{\delta q}. \tag{A.8}$$

It is easily shown that a self-similar solution of the form

$$t^{-\frac{1}{2j+1}} f\left(xt^{-\frac{1}{2j+1}}\right) \tag{A.9}$$

exists. The resulting ordinary differential equation for f is again the condition that a deformation of a system like (3.2) be monodromy preserving; this time, the asymptotic expansion of Ψ will, of course, involve $e^{\pm\theta}$ with $\theta = i\left(\omega_{2j+1}\frac{\zeta^{2j+1}}{2j+1} + \zeta x\right)$. The general autonomous equation of the MKdV family,

$$q_t = \frac{\partial}{\partial x}\frac{\delta H}{\delta q}, \tag{A.10}$$

where

$$H = \sum_0^N c_j H_{2j+1} \tag{A.11}$$

with constant c_j, does not have self-similar solutions, since the H_{2j+1} scale differently. If the c_j are time-dependent, however,

$$c_j = \hat{c}_j t^{\frac{2j+1}{2N+1}},$$

then a self-similar solution exists, and is associated with the monodromy for a system $\Psi_\zeta = Q\Psi$ which gives rise to expansions involving $e^{\pm\theta}$,

$$\theta = i\left(\sum_0^N \omega_j \frac{\zeta^{2j+1}}{2j+1} + \zeta x\right). \tag{A.12}$$

This is the *Painlevé II family* of nonlinear ordinary differential equations. It is apparent that all these equations can be reduced to a system of linear singular integral equations according to the pattern described earlier. Although the solution method parallels that of Sect. 3, the details are too cumbersome to reproduce here. It is possible, however, to give compact formulae for the rational solutions of all these equations (Stokes multipliers zero, and an n-th order pole at $\zeta = 0$). This was done in Sect. 3F.

We note that the equations of the Painlevé II family, which we designate as (PII, n) can be generated very quickly by the formalism described in Appendix I. We take $R_0 = C = -iY_1$, $P = q(Y_2 + Y_3)$. Then the first few equations are:

$$(\text{PII}, 0): \omega_1 D_R[C, P] + D_R x[R_0, P] = 0\,; \tag{A.13a}$$

this integrates to

$$(\omega_1 + x)q = v \tag{A.13b}$$

and

$$\theta(\zeta) = i\omega_1 + i\zeta x. \tag{A.13c}$$

$$(\text{PII}, 1): \omega_3 D_R^3[C, P] + \omega_1 D_R[C, P] + D_R x[R_0, P] = 0, \tag{A.14a}$$

which integrates to

$$-\frac{\omega_3}{4}(q_{xx} - 2q^3) + (\omega_1 + x)q = v, \tag{A.14b}$$

with

$$\theta(\zeta)=i\omega_3\frac{\zeta^3}{3}+i\omega_1\zeta+i\zeta x. \tag{A.14c}$$

$$(\mathrm{PII},2):\omega_3 D_R^5[C,P]+\omega_3 D_R^3[C,P]+\omega_1 D_R[C,P]+D_R x[R_0,P]=0, \tag{A.15a}$$

which integrates to

$$\frac{-5\omega_5}{16}[(q_{xx}-2q^3)_{xx}-2q(2qq_{xx}-q_x^2-3q^4)]$$

$$-\frac{\omega_3}{4}(q_{xx}-2q^3)+(\omega_1+x)q=\nu, \tag{A.15b}$$

with

$$\theta(\zeta)=\frac{i\omega_5\zeta^5}{5}+i\omega_3\frac{\zeta^3}{3}+i\omega_1\zeta+i\zeta x. \tag{A.15c}$$

The general equation

$$(\mathrm{PII},n):\sum_{r=0}^{N}\omega_{2r+1}D_R^{2r+1}[C,P]+D_R x[R_0,P]=0 \tag{A.16}$$

has $\theta(\zeta)$ given by (A.12).

Appendix III

Here we give the explicit connection between the formulae of Satō et al. and Eqs. (4.2), (4.3). In [7, III] it is shown that the basis W, to which we have already referred in the introduction, satisfies a holonomic system of differential equations. Applying a formal Laplace transform (transform variable u), Satō et al. obtain a system of ordinary differential equations in u,

$$\left(u\frac{d}{du}+mAu-G^{-1}m\bar{A}Gu^{-1}+F\right)\hat{w}(u)=0. \tag{A.17}$$

The dependence of $\hat{w}$ on the parameters $(a_j,\bar{a}_j)$, $j=1,\ldots,k$, at which points the solutions of the underlying Dirac equation have multivalued behavior, is provided by an auxiliary set of equations. We now write down this system when $n=2$ and $a_1-a_2=\frac{te^{i\theta}}{2m}$. Then using G, $\bar{G}$, F as given in [7, III] and taking $l_1=l_2=l=0$, $\frac{f_+}{\varepsilon\kappa^{-1}}=\frac{f_-}{\bar{\varepsilon}\kappa}=\frac{f}{2}=\frac{t\psi_t}{2}$, we find

$$\hat{w}_u=\left(-\frac{t}{4}\begin{pmatrix}1&0\\0&-1\end{pmatrix}+\frac{t}{4u^2}\begin{pmatrix}C&\varepsilon\kappa^{-1}S\\-\bar{\varepsilon}\kappa S&-C\end{pmatrix}-\frac{1}{u}\begin{pmatrix}0&f_+\\f_-&0\end{pmatrix}\right)\hat{w}, \tag{A.18}$$

$$\hat{w}_t=\left(-\frac{u}{4}\begin{pmatrix}1&0\\0&-1\end{pmatrix}-\frac{1}{4u}\begin{pmatrix}C&\varepsilon\kappa^{-1}S\\-\bar{\varepsilon}\kappa S&-C\end{pmatrix}-\frac{1}{t}\begin{pmatrix}0&f_+\\f_-&0\end{pmatrix}\right)\hat{w}, \tag{A.19}$$

where $C=\cosh 2\psi$, $S=\sinh 2\psi$. By scaling the components $\hat{w}_1$ and $\hat{w}_2$ by $\sqrt{\bar{\varepsilon}\kappa}$ and $\sqrt{\varepsilon\kappa^{-1}}$, respectively, and using $\varepsilon\bar{\varepsilon}=1$, we can eliminate the factors $\varepsilon\kappa^{-1}$ and $\bar{\varepsilon}\kappa$. Then, introducing the transformation $x=-\frac{1}{16}t^2$, $u=\frac{-i}{4}\zeta t$, in which case $f=t\psi_t=xv_x$ (where $v=2\psi$), we find

$$\hat{w}_x=\left(-i\zeta\begin{pmatrix}1&0\\0&-1\end{pmatrix}+\begin{pmatrix}0&q\\q&0\end{pmatrix}\right)\hat{w},$$

$$\hat{w}_\zeta=\left(-ix\begin{pmatrix}1&0\\0&-1\end{pmatrix}-\frac{xv_x}{2\zeta}\begin{pmatrix}0&1\\1&0\end{pmatrix}+\frac{i}{\zeta^2}\begin{pmatrix}\cosh v&\sinh v\\-\sinh v&-\cosh v\end{pmatrix}\right)\hat{w}$$

with $q=\frac{-v_x}{2}$. These are our Eqs. (A.6), (A.7), and (4.2) and (4.3).

Appendix IV. The Inversion Equations for PII When $\nu\neq n$

Following the ideas outlined in 3E, we can find the equations analogous to (3.33), (3.35) when $\nu\neq n$. For $\zeta\in S_1$, they are:

$$\psi^{(2)}(\zeta)e^{-\theta(\zeta)}=\begin{pmatrix}0\\1\end{pmatrix}-\frac{1}{2\pi i}\int_C\frac{\psi^{(2)}e^{-\theta}}{\xi-\zeta}d\xi+\frac{b}{2\pi i}\int_{C_{53}}\frac{\psi^{(1)}e^{-\theta}}{\xi-\zeta}d\xi+\frac{c}{2\pi i}\int_{C_{57}}\frac{\psi^{(1)}e^{-\theta}}{\xi-\zeta}d\xi$$
$$+\frac{ab}{2\pi i}\int_{C_{53}}\frac{\psi^{(2)}e^{-\theta}}{\xi-\zeta}d\xi+\frac{s}{2\pi i}\int_{C_{57}}\frac{\psi_2^{(1)}e^{-\theta}}{\xi-\zeta}d\xi, \tag{A.20}$$

$$\psi^{(1)}(\zeta)e^{\theta(\zeta)}=\begin{pmatrix}1\\0\end{pmatrix}-\frac{1}{2\pi i}\int_C\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi+\frac{a}{2\pi i}\int_{C_{42}}\frac{\psi^{(2)}e^{\theta}}{\xi-\zeta}d\xi+\frac{b+s}{2\pi i}\int_{C_{46}}\frac{\psi^{(2)}e^{\theta}}{\xi-\zeta}d\xi$$
$$-\frac{bc}{2\pi i}\int_{C_{46}}\frac{\psi^{(1)}e^{\theta}}{\xi-\zeta}d\xi+\frac{s}{2\pi i}\int_{C_{67}}\frac{\psi_3^{(2)}e^{\theta}}{\xi-\zeta}d\xi. \tag{A.21}$$

In (A.20) and (A.21), C is a contour originating at $\xi=\infty$, travelling on top of the branch cut along the positive real ξ axis, circling the origin and returning to $\xi=\infty\exp 2\pi i$ along the lower edge of the branch cut. The parameters $s=2i\sin\nu\pi$ and $\psi_2^{(1)}$, $\psi_3^{(2)}$ are defined in terms of $\psi^{(1)}$, $\psi^{(2)}$ by (3.12). Note that the terms involving the contour C simply become $-\operatorname{Res}_{\xi=0}\frac{\psi^{(j)}e^{\pm\theta}}{\xi-\zeta}$, $j=1,2$, when $\nu=n$ as in this case $\psi^{(j)}$ is single valued.

References

1. Ince, E.L.: Ordinary differential equations. New York: Dover Publications 1947
2. Ablowitz, M.J., Segur, H.: Phys. Rev. Lett. 38, 1103–1106 (1977)
3. Airault, H.: Rational solutions of Painlevé equations (to appear)
4. Wu, T.T., McCoy, B.M., Tracy, C.A., Barouch, E.: Phys. Rev. B 13, 316–371 (1976)
5. Barouch, E., McCoy, B.M., Wu, T.T.: Phys. Rev. Lett. 31, 1409–1411 (1973)
6. McCoy, B.M., Tracy, C.A., Wu, T.T.: J. Math. Phys. 18, 1058–1092 (1977)

7. Satō, M., Miwa, T., Jimbo, M.: A series of papers entitled Holonomic Quantum Fields: I. Publ. RIMS, Kyoto Univ. **14**, 223–267 (1977); II. Publ. RIMS, Kyoto Univ. **15**, 201–278 (1979); III. Publ. RIMS Kyoto Univ. **15**, 577–629 (1979). IV., V. RIMS Preprints 263 (1978), and 267 (1978). The paper we refer to most often is III. See also a series of short notes: Studies on holonomic quantum fields, I–XV
8. Gardner, C.S., Greene, J.M., Kruskal, M.D., Miura, R.M.: Commun. Pure Appl. Math. **27**, 97–133 (1976)
9. Fuchs, R.: Math. Ann. **63**, 301–321 (1906)
10. Ablowitz, M.J., Segur, H.: Stud. Appl. Math. **57**, 13–44 (1977)
11. Hastings, S.P., McLeod, J.B.: Univ. of Wisconsin, MRC Report No. 1861 (1978)
12. Ablowitz, M.J., Ramani, A., Segur, H.: (i) Lett. Nuovo Cimento **23**, 333 (1978). (ii) Two preprints: A connection between nonlinear evolution equations and ordinary differential equations of *P*-type, I, II
13. Tracy, C.A.: Proc. NATO Advanced Study Institute on: Nonlinear equations in physics and mathematics, 1978, (ed. A. Barut). Dordrecht, Holland: Reidel 1978
14. Schlesinger, L.: J. Reine Angewandte Math. **141**, 96–145 (1912)
15. Garnier, R.: Ann. Ec. Norm. Sup. **29**, 1–126 (1912)
16. Birkhoff, G.D.: Trans. AMS **10**, 436–470 (1909)
17. Birkhoff, G.D.: Proc. Am. Acad. Arts Sci. **49**, 521–568 (1913)
18. Garnier, R.: Rend. Circ. Mat. Palermo, **43**, 155–191 (1919)
19. Davis, H.T.: Introduction to nonlinear differential and integral equations. New York: Dover Publications 1962
20. Choodnovsky, D.V., Choodnovsky, G.V.: Completely integrable class of mechanical systems connected with Korteweg-deVries and multicomponent Schrödinger equations. I. Preprint, École Polytechnique, 1978
21. Moser, J., Trubowitz, E.: (to appear)
22. Olver, F.W.J.: Asymptotics and special functions. New York: Academic Press 1974
23. Ablowitz, M.J., Kaup, D.J., Newell, A.C., Segur, H.: Stud. Appl. Math. **53**, 249–315 (1974)
24. Flaschka, H., Newell, A.C.: Springer Lecture Notes in Physics **38**, 355–440 (1975)
25. Airault, H., McKean, Jr., H.P., Moser, J.: Comm. Pure Appl. Math. **30**, 95–148 (1977)
26. Brieskorn, E.: Jber. Dt. Math.-Verein. **78**, 93–112 (1976)
27. Ueno, K.: Kyoto, RIMS master's thesis, Dec. 1978. RIMS Preprints 301, 302 (1979)
28. Sibuya, Y.: (i) Proc. Int. Conf. Diff. Eq. pp. 709–738. (ed. H. A. Antosiewicz). New York: Academic Press 1975; (ii) Bull. AMS **83**, 1075–1077 (1977)
29. Zakharov, V.E., Shabat, A.B.: Sov. Phys. JETP **34**, 62–69 (1972)
30. Zakharov, V.E.: Paper at I. G. Petrovskii Memorial Converence, Moscow State Univ., Jan. 1976 (this paper has been referred to in many subsequent publications, but has apparently never been published)
31. Krichever, I.M.: Funkts. Anal. Prilozen **11**, 15–31 (1977)
32. Novikov, S.P.: Rocky Mt. J. Math. **8**, 83–94 (1978)
33. Newell, A.C.: Proc. Roy. Soc. London A **365**, 283–311 (1979)

Communicated by A. Jaffe

Received July 29, 1979; in revised form March 19, 1980

1.2

THE INVERSE MONODROMY TRANSFORM IS A CANONICAL TRANSFORMATION

H. Flaschka† and A.C. Newell*

*Clarkson College of Technology
Potsdam, NY 13676
U.S.A.

†University of Arizona
Tucson, AZ 85721
U.S.A.

I. INTRODUCTION AND GENERAL DISCUSSION

In earlier papers [1], [2], we used deformation theory to study the Painlevé equations which govern the self-similar solutions of the modified Korteweg deVries (MKdV) and sine-Gordon (SG) equations. In doing so we introduced the Inverse Monodromy Transform (IMT) which parallels the Inverse Scattering Transform (IST); whereas the latter is used to linearize the initial value problem for certain classes of nonlinear evolution equations, the former allows us to find, by linear methods, an important class of solutions, the multiphase similarity solutions to these equations. They are to the standard (one-phase) similarity solutions what multisoliton or finite gap periodic solutions are to solitons and cnoidal waves.

We begin in section 2 by introducing the family of integrable evolution equations

$$P_{t_j} = P_j(P, P_x, \cdots P_{jx}), \tag{1.1}$$

(P_{jx} means the j^{th} derivative of P with respect to x) which are integrability conditions for a certain eigenvalue problem

$$V_x = (\zeta R + P(x,t_j))V, \quad -\infty<x<\infty, \tag{1.2}$$

and associated family

$$V_{t_j} = Q^{(j)}(\zeta; P, \cdots P_{(j-1)x})V. \tag{1.3}$$

Applying a further constraint

$$V_\zeta = S(\zeta; P, \cdots P_{(n-2)x})V \tag{1.4}$$

with S rational in ζ, defines a finite dimensional manifold of solutions common to a subset j=0,1,--n-1 of the equation set (1.1). The manifold is defined by

a nonautonomous, nonlinear ordinary differential equation in x, the coefficients depending on x and t_j, which is the analogue of the Lax-Novikov (LN) [3,4] equation that defines the finite-gap and multisoliton solutions of the members of (1.1). There are many parallels between the LN equation and its nonautonomous (NLN) counterpart.

In this paper, we

(1) Introduce the NLN equation and the hierarchy of time flows with which it is associated.
(2) Show that, with respect to a given symplectic form and Poisson bracket, the NLN equation (the x-flow) and the companion time flows are generated by a sequence of commuting Hamiltonians, closely related to those which generate the multiperiodic flows;
(3) Define and illustrate how the IMT provides a mapping from the old coordinates $P,P_x,\cdots$ to new coordinates, the monodromy data M associated with the solution of (1.4), which admit trivial integration. In the case considered, S is polynomial in ζ of degree (n-1) and so M consists of the Stokes multipliers $\{s_j\}_{j=1}^{2n}$ defined in connection with the rank n irregular singular point at $\zeta=\infty$.
(4) Prove that the map is canonical and obtain expressions for the Poisson brackets of the Stokes multipliers.

Along the way, we

(5) Develop expressions for the infinitesimal changes in the new coordinates in terms of a natural inner product between the old coordinates and a set of (2n-2) vectors V_k which act as a basis in the space of dependent variables and which are formed from the "squared eigenfunctions". These expressions may be used to calculate the effects of perturbations on the NLN equations (for example, add a perturbation to the Painlevé equation; does it still retain its special properties? (see [1]); and provides a starting point for an attempt to prove a KAM (Kolmogoroff, Arnold, Moser) theorem - the flow does not, in general, take place on a compact manifold.
(6) Develop expressions for the new coordinates themselves as inner products between the old coordinates and V_k.
(7) Write expansions for the infinitesimal changes in the old coordinates and for the old coordinates themselves in terms of the basis vectors V_k. The latter provide expressions for solutions of the Painlevé and related equations in terms of contour integral representations.
(8) The relation with analogous expressions which arise in connection with the inverse scattering transform are explored.

A more general discussion of the ideas presented here, together with additional material on multiperiodic systems, particle systems and perturbed systems will be given in a series of forthcoming papers [5]. Our work on these questions was stimulated by several works by Sato, Miwa, Jimbo, Mori and Ueno [6] relating to the correlation functions of exactly solvable models in statistical mechanics and quantum field theory. In [6], they discuss the Hamiltonian structure of deformation equations.

2. DEFINING THE EQUATIONS

The inverse scattering transform focuses principal attention on (1.2) where ζ is the eigenvalue, R a constant matrix and $P(x,\vec{t})$ a matrix of potentials which evolve in time $\vec{t}=(t_0,t_1,\cdots t_{n-1})$ according to (1.1), j=0,--n-1. From (1.2) and (1.3), we have

$$P_{t_j} - Q^{(j)}_x + [\zeta R + P, Q^{(j)}] = 0, \quad j = 0,1,2,\cdots \tag{2.1a}$$

$$Q^{(j)}_{t_k} - Q^{(k)}_{t_j} + [Q^{(j)}, Q^{(k)}] = 0, \quad j,k = 0,1,2,\cdots \tag{2.1b}$$

where $[\ ,\ ]$ is the commutator. We seek solutions of (2.1a) for $Q^{(j)}$ polynomial in ζ of degree j

$$Q^{(j)} = R\zeta^j + \sum_{k=0}^{j-1} Q^{(j)}_{j+1-k}\,\zeta^k, \quad \text{(the } Q^{(j)}_{j+1-k} \text{ are independent of } j) \tag{2.2}$$

and the $Q^{(j)}_{j+1-k}$ are solved from the coefficients of ζ^k (the part of $Q^{(j)}_{j+1-k}$ which does not commute with R) and of ζ^{k-1} (the part that does) in (2.1a) (see [7]). The last equation in the sequence gives the evolution equation (1.1). The part of $Q^{(j)}_{j+1-k}$ which commutes with R is determined up to a constant in x (it can depend on $t_0, t_1, \cdots t_{j-1}$) and we take this constant to be zero without loss of generality. If nonzero, we can make it zero by a transformation in the time coordinates

Example: Define

$$Y_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad Y_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad Y_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{2.3}$$

$$R = -iY_1, \quad P = q(Y_2+Y_3), \quad Q^{(3)} = (-i\zeta^3 - \tfrac{1}{2}q^2\zeta)Y_1$$

$$+ (\zeta^2 q - \tfrac{1}{4}q_{xx} + \tfrac{1}{2}q^3)(Y_2+Y_3) + \frac{iq_x}{2}\zeta(Y_2-Y_3),$$

and (1.1) is

$$q_{t_3} = -\tfrac{1}{4}(q_{xxx} - 6q^2 q_x). \tag{2.4}$$

Finite-gap solutions of (1.1) are obtained by subjecting the vector V in (1.2) and (1.3) to a further (algebraic) constraint

$$\sum_1^{n-1} u_j V_{t_j} + u_0 V_x = \lambda V \tag{2.5a}$$

or

$$(Q - \lambda I)V = 0, \quad Q = \sum_1^{n-1} u_j\, Q^{(j)} + u_0(\zeta R + P). \tag{2.5b}$$

To motivate this choice, consider solutions of (2.6)

$$q = q(x - ct_3), \quad X = x - ct_3, \tag{2.6}$$

whence q(X) satisfies

$$-\tfrac{1}{4}q_{XXX} + 3/2\,q^2 q_X = cq_X. \tag{2.7}$$

In this case $Q^{(3)}$ is a function of ζ and X. Introduce the change of variables $X = x - ct_3$, $T = t_3$ and (1.2), (1.3) with j=3 become

$$V_X = (\zeta R + P(X))V, \quad V_T = (Q^{(3)} + c(\zeta R + P))V \tag{2.8}$$

respectively. Now set $V \to e^{\lambda T} V$ and obtain (2.5) with $u_3=1$, $u_0=c$, $u_j=0$ $j\neq 3$. Indeed the integrability condition of (1.2) and (2.5b) for the values chosen in the example is (2.7). In general, the integrability condition is

$$\sum_{j=1}^{n} u_j P_{t_j} + u_0 P_x = 0. \tag{2.9}$$

From (1.1), we see (2.9) is a nonlinear ordinary differential equation in x for the matrix P. The finite dimensional solution manifold defined by (2.12) is left invariant by the flows (1.1). By this we mean that finite-gap solutions considered as function of x satisfy (2.9) for all values the time parameters. It is well known (and details have been worked out in several cases [8]) that as functions of x, it is a completely integrable autonomous Hamiltonian system.

Solutions of (2.9) (as well as their time evolution under (1.1)) can be constructed explicitly; they are abelian functions. They are closely connected with the Riemann surface R defined by the polynomial relation $\Gamma(\zeta,\lambda) = 0$ between ζ and λ which comes about by the vanishing of the determinant of $Q-\lambda I$. The points of the Riemann surface parametrize $V(x,t)$ in (1.2), (1.3) and (2.5b). By proper normalization of V as eigenvector of (2.5b), one can arrange that it satisfies (1.2) and (1.3) and at the point(s) at infinity on R,

$$V \sim [1+O(\tfrac{1}{k})]\exp\,(kx+\sum_{j=1}^{n}\Omega_j(k)t_j) \tag{2.10}$$

k being a local parameter. The $\Omega_j(k)$ are polynomials, usually the dispersion relations of the linearized equations (1.1). In addition, V as function of (ζ,λ) on R has poles μ_s which are independent of x and t_j. Specification of the surface R, the poles μ_s, the asymptotics (2.10) and of a certain normalization condition determines V uniquely. Knowing V one can, by differentiation with respect to x, recover P from (1.2).

Now let us turn to a new class of solutions defined by adding another kind of constraint on the function $V(x,t,\zeta)$. This time we will ask it to satisfy an ordinary differential equation in ζ,

$$\zeta V_\zeta = \Sigma j t_j V_{t_j} + xV_x, \tag{2.11a}$$

$$= (\sum_{1}^{n} j t_j Q^{(j)} + x(\zeta R+P))V. \tag{2.11b}$$

Equation (2.11) has coefficients which are rational functions of ζ; all dependence on x,t_j is only parametric as far as ζ is concerned. The integrability condition on (2.11) and (1.2) gives us that (we omit the t_0 flow)

$$\sum_{j=1}^{n} j t_j P_{t_j} + (xP)_x = 0 \tag{2.12}$$

which shows that P is a function of the phases $\dfrac{x}{(nt_n)^{1/n}}$, $\dfrac{t_j}{(nt_n)^{1/n}}$, j=1,---n-1.

Equation (2.12), the analogue of (2.9), is an ordinary differential equation in x of order n with coefficients which depend on t_j. If we begin at time $\vec{t}^{(o)}$ with a solution of (2.12) and then let it evolve for a time $\vec{t}-\vec{t}^{(o)}$ in the flows (1.1), then at time $\vec{t}$, it will again satisfy (2.12) with the coefficients $t_1,---t_n$ evaluated at the new time

The choice of (2.11) can be motivated with the aid of the example (2.4) which has the self similar solution

$$q = \frac{1}{(3t_3)^{1/3}} f\left(\frac{x}{(3t_3)^{1/3}}\right) \tag{2.13a}$$

and $f(X)$, $X = x/(3t_3)^{1/3}$, satisfies the Painlevé equation of the second kind

$$\frac{1}{4}(f_{XX}-2f^3) = Xf-\nu. \tag{2.13b}$$

If we choose q to have the form (2.13a) in (1.2) and (1.3) with j=3, then V is a function of x,t_3 and ζ only through the combinations $X = x/(3t_3)^{1/3}$, $\zeta = \zeta(3t_3)^{1/3}$. Then (1.3), with j=3, becomes (2.11b) with $t_j=0$ $j\neq 3$. In the context of the example, the integrability condition of (2.11b) and (1.2) is (2.13b).

Another class of solutions to (1.1) can be found by adding the constraint

$$V_\zeta = (\sum_1^n \beta_j Q^{(j-1)}+xR)V \tag{2.14}$$

which leads to the integrability condition

$$\sum_1'' \beta_j P_{t_{j-1}} + xP_{t_0} = 0. \tag{2.15}$$

Recall $P_{t_0} = [R,P]$ is the scaling flow and $P_{t_1} = P_x$ is translation. In order that (2.14) and (1.3) are compatible (examine the asymptotic behavior of V as $\zeta\to\infty$), $\beta_j = jt_j$. Equation (2.15) then implies that $P(t_{n-1},t_{n-2},\cdots t_2,x,t_0)$ is a function only of the n phases

$$P = P(\tau_{n-2},\cdots\tau_1,\tau_0) \tag{2.10}$$

where $\{\tau_j\}_0^{n-2}$ are the integration constants defined by solving the equations

$$\frac{dt_{n-1}}{nt} = \frac{dt_{n-2}}{(n-1)t_{n-1}} = \cdots = \frac{dx}{2t_2} = \frac{d_{to}}{x} \tag{2.17}$$

where without loss of generality we can take $nt_n=1$. It is this class of solutions which we discuss when (1.2) is the Zakharov-Shabat system $\zeta R+P = \begin{pmatrix} -i\zeta & q \\ r & i\zeta \end{pmatrix}$.

By taking the coefficient matrix in the constraint equation to be a more complicated combination of rational functions of ζ, we can of course build more elaborate classes of solutions. In [2], we briefly indicated how to include a combination of solitons and self-similar solutions. In [5], we describe how to include the x dependence of the NLN system in different ways.

3. THE ZAKHAROV-SHABAT SYSTEM

3a. The equations: Here (1.2), (1.3) and the constraint equation (2.14) are

$$V_x = \begin{pmatrix} -i\zeta & q \\ r & i\zeta \end{pmatrix} V, \tag{3.1a}$$

$$V_{t_j} = \begin{pmatrix} -a^{(j)} & b^{(j)} \\ c^{(j)} & a^{(j)} \end{pmatrix} V, \tag{3.1b}$$

$$\zeta V_\zeta = \begin{pmatrix} -a & b \\ c & a \end{pmatrix} V, \tag{3.1c}$$

the integrability conditions for which are

$$a_x = rb-qc+i\zeta,\ b_x+2i\zeta b = 2qa,\ c_x-2i\zeta c = -2ra, \tag{3.2}$$

$$a^{(j)}_x = rb^{(j)}-qc^{(j)},\ b^{(j)}_x+2i\zeta b^{(j)} = q_{t_j}+2qa^{(j)},\ c^{(j)}_x-2i\zeta c^{(j)} = r_{t_j}-2ra^{(j)}, \tag{3.3}$$

$$a_{t_j} = c^{(j)}b-b^{(j)}c+\zeta a^{(j)}_\zeta,\ b_{t_j}+2a^{(j)}b = 2b^{(j)}a+\zeta b^{(j)}_\zeta,\ c_{t_j}-2a^{(j)}c = -2c^{(j)}a+\zeta c^{(j)}_\zeta . \tag{3.4}$$

3b. The Solutions To (3.2): We seek solutions to (3.2) polynomial in ζ. Set

$$a = \sum_{k=-\infty}^{n} a_{n+1-k}\zeta^k,\ b = \sum_{k=-\infty}^{n-1} b_{n+1-k}\zeta^k,\ c = \sum_{k=-\infty}^{n-1} c_{n+1-k}\zeta^k. \tag{3.5}$$

We find $a_1=\alpha_1$ constant, a_2 can be omitted. The ordinary differential equations (2.12) or (2.15) are formed by terminating the series at ζ^0 and ζ^1 respectively. We follow the latter course and then the equations are

$$b_{n+1} = c_{n+1} = 0 \tag{3.6}$$

where b_j, c_j, $j=2,\text{---}n+1$ are defined

$$b_{j+1} = \tfrac{1}{2}b_{jx}-ia_jq+xq\delta_{j,n},c_{j+1}= -\tfrac{1}{2}c_{jx}-ia_jr+r\delta_{j,n},j=1,\text{---}n. \tag{3.7}$$

The coefficients a_j, $j=3,\text{--}n$, are found by noting that the function

$$\Omega = a^2+bc-2i\zeta\textstyle\int^x a\,dx \tag{3.8}$$

is independent of x. We set this equal to

$$\Omega = \left(\alpha_1\zeta^n+\alpha_3\zeta^{n-2}+\text{---}+\alpha_n\zeta+\alpha_{n+1}+\frac{\alpha_{n+2}}{\zeta}+\text{--}\right)^2, \tag{3.9}$$

and equate the powers of $\zeta^{n+\ell}$ in (3.8). For $\ell=n-2$ to $\ell=-1$, this equation defines $a_{n+1-\ell}$ in terms of the b_j,c_j already determined to this level from (3.7) and $\alpha_{n+1-\ell}$ is effectively the additive constant associated with integrating a_x in (3.2). For $b=0$ to $b=-n+2$, this equation determines $\alpha_{n+1-\ell}$ ($\alpha_{n+1},\text{--}\alpha_{2n-1}$) in terms of the already known a_j,b_j,c_j, $j\le n$ and $(x, \alpha_3,\alpha_4,\text{--}\alpha_n)$. For $\ell<-n+2$, we simply choose $\alpha_{n+1-\ell}$ such that the coefficients of $\zeta^{n+\ell}$, $\ell<-n+2$ are zero. These choices play no role in the later analysis. We now list the first six coefficients.

	a	b	c
1.	1	0	0
2.	0	$-i\alpha_1 q$	$-i\alpha_1$
3.	$\alpha_3+\alpha_1 qr/2$	$\frac{\alpha_1}{2}q_x$	$\frac{-\alpha_1}{2}r_x$
4.	$\alpha_4+\frac{i\alpha_1}{4}(rq_x-r_xq)$	$\frac{i\alpha_1}{4}(q_{xx}-2q^2r)-i\alpha_3 q$	$\frac{i\alpha_1}{4}(r_{xx}-2qr^2)-i\alpha_3 r$
5.	$\alpha_5+\alpha_3\frac{qr}{2}$	$\frac{-\alpha_1}{8}(q_{xxx}-6qrq_x)+\frac{\alpha_3}{2}q_x$	$\frac{\alpha_1}{8}(r_{xxx}-6qrr_x)-\frac{\alpha_3}{2}r_x$
	$\frac{-\alpha_1}{8}(rq_{xx}+qr_{xx}-r_xq_x-3q^2r^2)$	$-i\alpha_4 q$	$-i\alpha_4 r$ (3.10)
6.	$\alpha_6+\alpha_4\frac{qr}{2}+\frac{i\alpha_3}{4}(rq_x-v_xq)$	$-\frac{i}{16}\alpha_1(q_{xxx}-6qrq_x)_x$	$-\frac{i}{16}\alpha_1(r_{xxx}-6qrr_x)_x$
	$\frac{-i}{16}\alpha_1(rq_{xxx}-qr_{xxx}-r_xq_{xx}$	$+\frac{i\alpha_1}{8}(rq_{xx}+qr_{xx}-q_xr_x-3q^2r^2)q$	$+\frac{i\alpha_1}{8}(rq_{xx}+qr_{xx}-q_xr_x-3q^2r^2)r$
	$+q_xr_{xx}-6qr(rq_x-r_xq))$	$+\frac{i\alpha_3}{4}(q_{xx}-2q^2r)+\frac{\alpha_4}{2}q_x-i\alpha_3 q$	$+\frac{i\alpha_3}{4}(r_{xx}-2qr^2)-\frac{\alpha_4}{2}r_x-i\alpha_3 r$

3c. The Solutions of (3.3): We seek polynomial solutions to (3.3) in the form

$$a^{(j)}=i\zeta^j+\sum_0^{j-1}a^{(j)}_{j+1-k}\zeta^k,\quad b^{(j)}=\sum_0^{j-1}b^{(j)}_{j+1-k}\zeta^k,\quad c^{(j)}=\sum_0^{j-1}c^{(j)}_{j+1-k}\zeta^k. \tag{3.11}$$

It is simple to verify that $a^{(j)}_{n+1-k}$, $b^{(j)}_{n+1-k}$, $c^{(j)}_{n+1-k}$ are simply a_{n+1-k}, b_{n+1-k}, c_{n+1-k} when $n=j$, $\alpha_1=i$, $\alpha_k=0$, $k=3,\cdots j$. For example, $a^{(3)}=i\zeta^3+\frac{iqr}{2}\zeta-\frac{1}{4}(rq_x-r_xq)$. The ζ^0 balance in (3.3) gives the evolution equations for q and r as functions of t_j. Because of the normalization of $a^{(j)}$, the compatibility of (3.2 and (3.3) requires that

$$\alpha_{n+1-j}=ijt_j\quad j=2,\cdots n. \tag{3.12}$$

We take $t_{n-1}=0$ and omit the t_{n-1} flow. We will show very shortly how this dependence can be reincorporated. In essence, the t_{n-1} flow is simply a restatement of the NLN equations (3.6) once the particular choice of scaling in $t_0,x,t_2,\cdots t_{n-1}$ is made. We can take ntn=1. Also we will take $t_1=0$ as this flow simply mimics the x-flow. We will often for convenience call x by t_1. We are left with the x-flow given by (3.6) and the time flows q_{t_j}, r_{t_j}, j=0,2,--n-2. We list and discuss the cases n=3,4.

n=3: The x-flow is

$$b_{3x}=2(a_3+ix)q,\quad c_{3x}=-2(a_3+ix)r$$

or

$$\frac{i}{2}(q_{xx}-2q^2r)=2ixq,\quad \frac{i}{2}(r_{xx}-2qr^2)=2ixr. \tag{3.13}$$

The only time flow is

$$q_{t_o} = -2iq,\ r_{t_o} = 2ir. \tag{3.14}$$

Now let us comment on what this solution has to do with solutions of

$$q_{t_2} = \tfrac{1}{2}(q_{xx}-2q^2r),\ r_{t_2} = -\tfrac{1}{2}(r_{xx}-2qr^2). \tag{3.15}$$

If we had left $\alpha_2 \neq 0$ and indeed chosen it to be $2it_2$, then (3.13) would read with $\alpha_1=1$,

$$q_{t_2}+2t_2q_{t_1}+xq_{t_o} = 0,\ r_{t_2}+2t_2r_{t_1}+xr_{t_o} = 0 \tag{3.16}$$

which means that q and r are functions of

$$\tau_1 = x-t_2^2,\ \tau_o = t_o+\tfrac{2}{3}t_2^3-xt_2 \tag{3.17}$$

Now impose (3.17), and since (3.14) holds, we have

$$q(t_o,x,t_2) = e^{-2i(t_o+\frac{2t_2^3}{3}-xt_2)}\ f(x-t_2^2)$$

$$r(t_o,x,t_2) = e^{2i(t_o+\frac{2t_2^3}{3}-xt_2)}\ g(x-t_2^2) \tag{3.18}$$

which when substituted into (2.15) has f and g satisfying (3.13) with $\alpha_1=i$ and x replaced by $x-t_2^2$.

Thus in what follows we simply ignore the equation for $q_{t_{n-1}}$, $r_{t_{n-1}}$ and find q,r as functions of $t_o,x,t_2,\cdots t_{n-2}$. Then to incorporate t_{n-1} simply replace $t_o,x,t_2,\cdots t_{n-2}$ by $\tau_o,\tau_1\cdots\tau_{n-2}$ which are found by integrating

$$dt_{n-1} = \frac{dt_{n-2}}{(n-1)t_{n-1}} = \frac{dx}{2t_2} = \frac{dto}{x}. \tag{3.19}$$

n=4 The x-flows are

$$b_{4x} = 2(a_4+ix)q,\ c_{4x} = -2(a_4+ix)r$$

or

$$-\tfrac{1}{4}(q_{xxx}-6qrq_x)-i\alpha_3q_x-2ixq = 0$$

$$-\tfrac{1}{4}(r_{xxx}-6qrq_x)-i\alpha_3r_x+2ixr = 0. \tag{3.20}$$

The associated time flows are

$$q_{t_o} = -2iq,\ r_{t_o} = 2ir \tag{3.21}$$

$$q_{t_2} = \tfrac{i}{2}(q_{xx}-2q^2r),\ r_{t_2} = -\tfrac{i}{2}(r_{xx}-2qr^2). \tag{3.22}$$

Again we may also include solutions to

$$q_{t_3} = -\frac{1}{4}(q_{xxx}-6qrq_x),\ r_{t_3} = -\frac{1}{4}(r_{xxx}-6qrr_x) \tag{3.23}$$

by replacing the solutions

$$q(t_o,x,t_2),\ r(t_o,x,t_2)$$

by

$$q(t_o-\frac{3t_3^4}{4}+\frac{t_2t_3^2}{2}-xt_3,\ x+t_3^3-t_2t_3,\ t_2-\frac{3t_3^2}{2})$$

and

$$r(t_o-\frac{3t_3^4}{4}+\frac{t_2t_3^2}{2}-xt_3,\ x+t_3^3-t_2t_3,\ t_2-\frac{3t_3^2}{2}).$$

In the corresponding periodic problem, ζV_ζ in (3.1c) is replaced by λV. This removes the $i\zeta x$ term in a, the function Ω is a^2+bc and the identification (3.12) is no longer necessary; the α_j are constants. The Lax-Novikov equations are

$$b_{nx} = 2a_nq,\ c_{nx} = -2a_nr. \tag{3.24}$$

3d. The Hamiltonian Structure: The function Ω generates the Hamiltonian for each of the flows. In particular

$$H_j = 4\alpha_{n+1+j},\ j = 0,---n-2. \tag{3.25}$$

The conjugate variables are

$$\beta_j,\ \bar{c}_{n+2-j}\quad j=2,--n \tag{3.26}$$

where $\bar{b}_j$, $\bar{c}$, are the coefficients of ζ^{-j+2}, j=2,--n in the asymptotic expansions for

$$\zeta^{n-1}\sqrt{\frac{b+\alpha}{\alpha_1\zeta^n}}\ ,\ \zeta^{n-1}\sqrt{\frac{c+\alpha}{\alpha_1\zeta^n}}\ ,\quad \alpha = \alpha_1\zeta^n+\alpha_3\zeta^{n-2}+--- \tag{3.27}$$

as $\zeta\to\infty$. We can identify the correct choice of conjugate variables from three sources:

(i) they are the same as those of the periodic problem (the detailed analysis leading to these choices is given in [6];
(ii) the Hamiltonian H_ρ which generates the scaling flow must decompose into a sum of products of conjugate variables;
(iii) the choices are necessary in order to define certain inner products correctly.

For now, we will simply take (3.27) as given. The Hamiltonians contain the independent variables $x,t_2,---t_{n-2}$. We make the system autonomous by including these as dependent variables and adding conjugates $T_1=X,T_2,--T_{n-2}$ which are defined from the term $\int^x a_j dx$ in Ω j=3,--n in such a way that

$$H_j = T_j+\hat{H}_j(\beta_k,\bar{c}_k,x,t_2--t_{n-2}). \tag{3.28}$$

From this choice we see that

$$t_j = \frac{\partial H_j}{\partial T_j} = 1, \quad t_k = \frac{\partial H_j}{\partial T_k} = 0, \quad k \neq j (t_1 = x),$$

and so t_j is the "time" variable for the flow generated by H_j.

Let us look in detail at some examples.

n=3: $\alpha_1 = i$, $\alpha_3 = 0$.

$$H_0 = -2i(b_2c_3 + b_3c_2)$$

$$H_1 = -2i\left(b_3c_3 - \frac{b_2^2c_2^2}{4} - xb_2c_2 + \int^x b_2c_2 dx\right).$$

The conjugate variables are

$$b_2, b_3, x \text{ and } c_3, c_2, X = -2i\int^x b_2c_2 dx.$$

It is easy to verify that H_0, H_1 generate the flows (3.14) and (3.13), respectively. Also, both H_0 and H_1 are constants with respect to t_0 and x.

n=4: $\alpha_1 = i$, $\alpha_3 = 2it_2$, $\alpha_4 = 0$.

$$H_0 = -2i(\tilde{b}_2\bar{c}_4 + \tilde{b}_3\bar{c}_3 + \tilde{b}_4\bar{c}_2)$$

where $\tilde{b}_2 = b_2$, $\tilde{b}_3 = b_3$, $\tilde{b}_4 = b_4 - \frac{a_3+\alpha_3}{4i}b_2 = b_4 - \frac{b_2^2c_2}{8} - t_2b_2$

and the c's are defined analogously.

$$H_1 = -2i\{\tilde{b}_3(\bar{c}_4 + \frac{\tilde{b}_2\bar{c}_2^2}{8} + t_2\bar{c}_2) + \bar{c}_3(\tilde{b}_4 + \frac{\tilde{b}_2^2\bar{c}_2}{8} + t_2\tilde{b}_2)$$

$$+ \frac{1}{2}(2it_2 + \frac{1}{2}\tilde{b}_2\bar{c}_2)(\tilde{b}_2\bar{c}_3 + \tilde{b}_3\bar{c}_2) - x\tilde{b}_2\bar{c}_2 + \int^x \tilde{b}_2\bar{c}_2 dx\}$$

$$H_2 = -2i\{(\tilde{b}_4 + \frac{\tilde{b}_2^2\bar{c}_2}{8} + t_2\tilde{b}_2)(\bar{c}_4 + \frac{\tilde{b}_2\bar{c}_2^2}{8} + t_2\bar{c}_2) - x(\tilde{b}_2\bar{c}_3 + \tilde{b}_3\bar{c}_2)$$

$$+ \int^x (\tilde{b}_2\bar{c}_3 + \tilde{b}_3\bar{c}_2) - 2t_2(\tilde{b}_2\bar{c}_4 + \tilde{b}_3\bar{c}_3 + \tilde{b}_4\bar{c}_2)\}.$$

Note that the last term in H_2 comes from $-4i\alpha_3\alpha_5$ appearing in the coefficient of ζ^2 in (3.8). The conjugate pairs are $(\tilde{b}_2, \bar{c}_4)$, $(\tilde{b}_3, \bar{c}_3)$, $(\tilde{b}_4, \bar{c}_2)$, $(x, X = -2i\int^x \tilde{b}_2\bar{c}_2 dx)$,

$(t_2, T_2 = -2i\int^x (\tilde{b}_2\bar{c}_3 + \tilde{b}_3\bar{c}_2)dx)$.

In particular the form of T_2 may be obtained by noting that, since H_1 generates the x flow,

$$\frac{dT_2}{dx} = \frac{\partial H_1}{\partial t_2} = -2i(\tilde{b}_2\bar{c}_3 + \tilde{b}_3\bar{c}_2).$$

In this way, the form of T_k, k=2,--n-2 can always be found from the corresponding H_1.

Before we continue, a convention or notation. $\frac{df}{dt_j}$ means differentiate all variables in f which depend on t_j, including t_j itself, keeping $t_k \neq t_j$ (t_1=x) constant. $\frac{\partial f}{\partial t_j}$ means we differentiate f only with respect to explicit t_j dependence.

For the multiperiodic case, we define H_0, H_1, --H_{n-2} by leaving out the contributions from $2i\zeta\int^x a dx$. We also leave the α_j^0 as constants. Then the conjugate variables are simply (3.26), the equations are the same with ix replaced by constant α_n.

3e. Poisson Brackets And The Commutability of $\{H_j\}_{j=0}^{n-2}$.

It is natural to define the Poisson bracket (t_1=x, T_1=X)

$$[F,G] = \sum_{n=2}^{n} \left(\frac{\partial F}{\partial \bar{b}_j} \frac{\partial G}{\partial \bar{c}_{n+2-j}} - \frac{\partial F}{\partial \bar{c}_{n+2-j}} \frac{\partial G}{\partial \bar{b}_j} \right)$$

$$+ \sum_{j=0}^{n-2} \left(\frac{\partial F}{\partial t_j} \frac{\partial G}{\partial T_j} - \frac{\partial F}{\partial T_j} \frac{\partial G}{\partial t_j} \right). \tag{3.29}$$

The first term is the Poisson bracket for the periodic problem and we write this $[F,G]_p$. Note that if

$$\nabla = \left(\frac{\partial}{\partial \bar{c}_2}, \frac{\partial}{\partial \bar{b}_2}, \frac{\partial}{\partial \bar{c}_3}, \frac{\partial}{\partial \bar{b}_3}, \cdots \frac{\partial}{\partial \bar{c}_n}, \frac{\partial}{\partial \bar{b}_n}\right)^T, \tag{3.30}$$

$$J = \begin{pmatrix} 0 & 0 \cdots 0 & 1 \\ 0 & \cdots & -1 & 0 \\ -1 & 0 \cdots \cdots \cdots & 0 \end{pmatrix} \tag{3.31}$$

then

$$[F,G]_p = \langle \nabla F, J \nabla G \rangle, \tag{3.32}$$

where $\langle u,v \rangle$ is $\sum_{j=2}^{n} u_j v_j$, the usual inner product. Also, if

$$F = F(\bar{b}_j, \bar{c}_j, t_j, T_j),$$

then

$$\frac{dF}{dt_k} = [F, H_k]. \tag{3.33}$$

In [5], we show that for each j,k = 0,--n-2,

$$[H_j, H_k] = 0. \tag{3.34}$$

Thus each H_j is a constant of the motion for each of the flows generated by the

other H_k. We therefore have (n-1) independent constants of the motion in involution. The periodic system, which has (2n-2) dependent variables, is then completely integrable. The multiphase self-similar solutions on the other hand are 4n-4 dimensional and thus we need another (n-1) independent constants of the motion. These new constants are introduced in the next section.

4. The Inverse Monodromy Transform (IMT):

Consider (3.1c),

$$\zeta V_\zeta = \begin{pmatrix} -a & b \\ c & a \end{pmatrix} V \tag{4.1}$$

with a,b,c, given by (3.5) with $a_{n+1}=b_{n+1}=c_{n+1} = 0$. Then $\zeta=\infty$ is an irregular singular point of rank n and the fundamental solution matrix

$$\Phi(x,t_j;\zeta) = (\psi,\bar{\psi}) \tag{4.2}$$

admits the formal asymptotic expansion (see Appendix I) as $\zeta\to\infty$,

$$\Phi\sim\tilde{\Phi} = (I+\sum_{s=1}^{\infty}\frac{C_s}{\zeta^s})e^{-\phi}\begin{pmatrix} e^{-\theta} & 0 \\ 0 & e^{\theta} \end{pmatrix} \tag{4.3}$$

where

$$\theta = \frac{i\zeta^n}{n}+it_{n-2}\zeta^{n-2}+\cdots+it_j\zeta^j+i\zeta x+it_0+\frac{iH_0}{4}\ln\zeta+o(1),$$

$$\phi = o(1),\ H_0 = +4\alpha_{n+1} \tag{4.4}$$

$\tilde{\Phi}$ has been normalized so as to satisfy (3.1a) and (3.1b). Now, it is known that in general Φ will not have the asymptotic expansion $\tilde{\Phi}$ in every neighborhood of $\zeta=\infty$. This neighborhood naturally divides into 2n equal sectors separated by rays, called anti-Stokes lines, on which $Re\theta\sim Re i\zeta^n = 0$. Define S_k as $\{\zeta;|\zeta|>\rho,\ \frac{\pi}{n}(k-1)\le Arg\zeta\le\frac{\pi}{n}k\}$ and R_k as $Arg\zeta=\frac{\pi}{n}(k-1)$. Let $\Phi=(\psi,\bar{\psi})$ be a fundamental solution matrix of (4.1) with asymptotic behavior $\tilde{\Phi}$ in S_1. The solution $\bar{\psi}\sim\binom{0}{1}e$ is recessive (asymptotically decaying whereas $\psi\sim\binom{1}{0}e^{-\theta}$ is dominant (it is unique as it is defined on the initial ray $\zeta=0$ where θ is imaginary). Continuing to sector S_2, the recessive solution $\bar{\psi}$ becomes the dominant solution $\bar{\psi}_2$ in that sector; however, in order that the dominant solution ψ in S_1 become the recessive solution ψ_2 in S_2, we must add a constant factor (the stokes multiplier) times the recessive solution $\bar{\psi}$. In general, the fundamental solution matrix Φ_{j+1} which has asymptotic behavior $\tilde{\Phi}$ in S_{j+1} is related to its preceding neighbor by

$$\Phi_{j+1} = \Phi_j M_j, \tag{4.5}$$

where $M_j = \begin{pmatrix} 1 & s_j' \\ s_j & 1 \end{pmatrix}$. If ψ_j is dominant, then $s_j'=0$, $s_j\neq 0$. If ψ_j is recessive and $\bar{\psi}_j$ dominant, then the nonzero off-diagonal element of M_j occupies the (1,2) position.

We call the set of Stokes multipliers $\{s_j\}_{j=1}^{2n}$ the monodromy data M for (4.1). Since the total monodromy around $\zeta=\infty$ is the identity (as it equals that about $\zeta=0$, an ordinary point), only (2n-2) of the set M are independent. For reasons of symmetry we will choose to work with the set $\{s_k\}$ for $k\epsilon Z=\{1,2,---n-1,n+1,--2n-1\}$.

One of the central results of our previous papers is: given Φ_j, j=1,--2n with asymptotic behaviors $\tilde{\Phi}$ which satisfy (3.1a), (3.1b), (3.1c), and $M_j(s_j)$ defined by (4.5), then the stokes multipliers are constants, independent of $t_0,x,---t_{n-2}$. This result is again proved in Section 5.

We therefore have a map from old variables

$$q(\bar{b}_2,---\bar{b}_n,x,t_2,---t_{n-2}),\ p(\bar{c}_n,--\bar{c}_2,X,T_2,---T_{n-2})$$

to new variables

$$Q(f_1(s_k),---f_{n-1}(s_k),x_1t\ ,--t_{n-2}),\ P(g_1(s_k),---g_{n-1}(s_k),H_1,---H_{n-2})$$

where the functions f_j and g_j are functions of the Stokes multipliers s_k,$k\epsilon Z$. In the new variables the equations are

$$\begin{aligned} &s_j,_{t_k} = 0,\quad j\epsilon Z,\quad k = 0,1,--n-2,t_1 = x.\\ &H_j,_{t_k} = 0,\quad j = 0,--n-2,\ k = 0,1,--n-2.\\ &t_j,_{t_k} = \delta_{jk}\quad j,k = 0,---n-2. \end{aligned} \tag{4.6}$$

Our next goal is to show that this map is canonical. In order to do this, it is necessary to express the infinitesimal variations δQ, δP in terms of δq, δp.

5. INFINITESIMAL VARIATIONS AND THE t_j DEPENDENCE of s_k.

5a. Infinitesimal variations: Take the infinitesimal variation of

$$\zeta\Phi = \begin{pmatrix} -a & b \\ c & a \end{pmatrix}\Phi,\ \Phi = \begin{pmatrix} \psi_1 & \bar{\psi}_1 \\ \psi_2 & \bar{\psi}_2 \end{pmatrix} \tag{5.1}$$

and solve by variation of parameters. Integrate the result between ζ_1, which lies near $\zeta=\infty$ on R_1, and ζ_2, defined similarly. Since Φ is entire, any path will do. We find

$$\Phi^{-1}\delta\Phi|_{\zeta_2} - \Phi^{-1}\delta\Phi|_{\zeta_1} = \int_{\zeta_1}^{\zeta_2} \frac{1}{\zeta}\Phi^{-1}\begin{pmatrix} -\delta a & \delta b \\ \delta c & \delta a \end{pmatrix}\Phi d\zeta. \tag{5.2}$$

Now use (4.3) and the facts that on R_1, $\Phi\sim\tilde{\Phi}$, on R_2 $\Phi\sim\tilde{\Phi}M_1^{-1}$, and compare the (2,1) elements of (5.2) to obtain

$$-\delta s_1 - 2i s_1\Big(\sum_{j=2}^{n-2}\delta t_j\zeta_2^j + \delta x\zeta_2 + \delta t_0\Big) - 2s_1\delta\alpha_{n+1}\ell n\zeta_2 = \int_{\infty_1}^{\zeta_2}(2\delta a\psi_1\psi_2 - \delta b\psi_2^2 + \delta c\psi_1^2)\frac{d\zeta}{\zeta}, \tag{5.3}$$

where we have let $\zeta_1\to\infty_1$, the point at ∞ on R_1. We cannot set $\zeta_2=\infty_2$, for neither side

of (5.3) converges. However, we make use of the following identities derived from (3.1):

$$\frac{1}{2}(2a_x\psi_1\psi_2-b_x\psi_2^2+c_x\psi_1^2) = \frac{d}{d\zeta}(2i\zeta\psi_1\psi_2-q\psi_2^2+r\psi_1^2), \tag{5.4a}$$

$$\frac{1}{2}(2a_{t_j}\psi_1\psi_2-b_{t_j}\psi_2^2+c_{t_j}\psi_1^2) = \frac{d}{d\zeta}(2a^{(j)}\psi_1\psi_2-b^{(j)}\psi_2^2+c^{(j)}\psi_1^2).$$

Also from (A.2), we may show that the arguments of the right hand sides of (5.6a) and (5.6b) have asymptotic behaviors

$$-2i\zeta_2 s_1+o(1) \text{ and } -2i\zeta_2^j s_1+o(1) \text{ as } \zeta_2\to\infty \text{ on } R_2$$

respectively. Therefore we may write

$$-\delta s_1 = \int_{\infty_1}^{\zeta_2}(2da\psi_1\psi_2-db\psi_2^2+dc\psi_1^2)\frac{d\zeta}{\zeta}+2s_1\delta\alpha_{n+1}\ell n\zeta_2 \tag{5.5}$$

where $da = \delta a-\sum_{j=o}^{n-2} a_{t_j}\delta t_j, (t_1=x)$ and db, dc are defined similarly.

The integral on the right hand side of (5.5) has an asymptotic expansion which consists of (a) terms like $e^{\pm 2\theta\zeta^p}$ when $p<n-1$, which are integrable along the rays R_1 and R_2 and (b) a term proportional to $1/\zeta$ on R_2. When integrated, this term is exactly cancelled by $2s_1\ \delta\alpha_{n+1}\ \ell n\ 2$. Thus the limit $\zeta_2\to\infty$ along R_2 may be taken and $-\delta s_1$ is well defined.

The calculation may be repeated for any sector. In the odd numbered sectors, one compares the (2,1) elements of (5.2); for the even onescompare the (1,2) elements. This results in a change of sign. If $\zeta_R(\infty_R)$ is a point which tends to $\zeta=\infty$ along R_k, then,

$$(-1)^k\delta s_k = \int_{\infty_k}^{\zeta_{k+1}}(2da\psi_1\psi_2-db\psi_2^2+dc\psi_1^2)\frac{d\zeta}{\zeta}+2s_k\delta\alpha_{n+1}\ell n\zeta_{k+1}, \tag{5.6}$$

where $\psi = \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}$ is always the dominant solution in S_k.

Our next task is to rewrite (5.6) taking account of the dependence of a_j on the sequence of b's and c's. Let us begin with the case n=3.

$$(-1)^k\delta s_k = dc_2\left(\int_{\infty_k}^{\zeta_{n+1}}(\zeta\psi_1^2+ib_2\psi_1\psi_2)d\zeta-is_kb_3\ell n\zeta_{k+1}\right)$$

$$+ db_2\left(-\int_{\infty_k}^{\zeta_{k+1}}(\zeta\psi_2^2-ic_2\psi_1\psi_2)d -is_kc_3\ell n\zeta_{k+1}\right)$$

$$+dc_3\left(\int_{\infty_k}^{\zeta_{k+1}}\psi_1^2d\zeta-is_kb_2\ell n\zeta_{k+1}\right)$$

$$+db_3\left(-\int_{\infty_k}^{\zeta_{k+1}}\psi_2^2 d\zeta - is_k c_2 \ell n\zeta_{k+1}\right) \tag{5.7}$$

Define

$$V_k = \begin{pmatrix} \int_{\infty_k}^{\zeta_{k+1}} \psi_2^2 d\zeta + i\ {}_k c_2 \ell n\zeta_{k+1} \\ \int_{\infty_k}^{\zeta_{k+1}} \psi_1^2 d\zeta - is_k b_2 \ell n\zeta_{k+1} \\ \int_{\infty_k}^{\zeta_{n+1}} (\zeta\psi_2^2 - ic_2\psi_1\psi_2) d\zeta + is_k c_2 \ell n\zeta_{k+1} \\ \int_{\infty_k}^{\zeta_{k+1}} (\zeta\psi_2^2 + ib_2\psi_1\psi_2) d\zeta - is_k b_3 \ell n\zeta_{k+1} \end{pmatrix} \tag{5.8}$$

Defining J as in (3.31) and $dR = (d\bar{c}_2, d\bar{b}_2, \cdots d\bar{c}_n, d\bar{b}_n)^T$, where $(\bar{c}_j, \bar{b}_j)$ are the conjugate variables of Section 3, we have

$$(-1)^k \delta s_k = \langle dR, JV_k \rangle,\ k\epsilon Z, \tag{5.9}$$

where $\langle u,v\rangle = \sum_1^n u_i v_i$. By rewriting the identity

$$\int_{\zeta_1}^{\zeta_2} (b\psi_2^2 + c\psi_1^2)\frac{d\zeta}{\zeta} = (\psi_1\psi_2)_{\zeta_1}^{\zeta_2} = -s_1 \tag{5.10}$$

in terms of $(\bar{b}_k, \bar{c}_k)$, we find

$$-s_k = \langle FR, JV_k' \rangle,\ k\epsilon Z \tag{5.11}$$

where F is the matrix $\{(-1)^{i+1}\delta_{ij}\}$. In Appendix 2, we find the orthogonality relations for the sequence of vectors $V = \{V_k\}$, $k\epsilon Z$,

$$M_k = \langle JV_k, V_\ell \rangle = \pi(\delta_{k,\ell-1} + s_k s_\ell),\ k<\ell, \tag{5.12}$$

The set V forms a basis in R_{2n-2} and the relations (5.9), (5.11) and (5.12) allow us to expand dR and FR in the basis V. Let $N=M^{-1}$; then

$$dR = \sum_{j,\ell} (-1)^j N_{\ell j}\ \delta s_j\ V_\ell \tag{5.13}$$

and

$$FR = \sum_{j,\ell} N_{\ell j}\ s_j\ V_\ell \tag{5.14}$$

where in both cases, the summation is over $j,\ell\in Z$.

Several remarks are now in order.

1. We can, by the Gram-Schmidt process, choose a new set of vectors $U=U_k$, $k\in Z$, such that $\langle JU_k, U_\ell\rangle = \delta_{k,\ell-1}$, $k<\ell$. The transformation between U and V will depend on the Stokes multipliers. We will not do this here but will indicate, in the results which follow, the effects of such a transformation.

2. Equation (5.9) shows us that if R changes in such a way such that $dR=0$ then $\delta s_k=0$. But $dR=0$ implies

$$\delta R = -J\nabla H_o \delta t_o - J\nabla H_1 \delta x - \sum_{k=2}^{n-2} J\nabla H_k \delta t_k$$

that is, R changes as a linear combination of the flows generated by Ω. Conversely if $\delta s_k=0$, then $dR=0$.

3. Notice that if

$$dR = -\alpha FR$$

then $\delta s_k = (-1)^k \alpha s_k$. However, the right hand side still is jsut the scaling flow and is proportional to $J\nabla H_o$. The time t dependence ($\delta = d/dt$ say) of s_k ($=s_k(t=0)e^{(-1)^k\alpha t}$.

can always be removed by correctly normalizing the asymptotic expansion (4.4).

4. Equation (5.9) is the s[illegible]ting point for a perturbation theory. Using it we may calculate how a perturbation

$$dR = \varepsilon F(R),\ o<\varepsilon<<1, \qquad (5.15)$$

affects the Stokes multipliers. In particular, (5.15) may no longer have the special property that as functions of the phases, the only moving singularities of q,r are poles. How is this reflected in the change of s_j? More generally, it is very natural to ask if an analogue to the KAM (Kolmogoroff, Arnold, Moser) theorem obtains. The finite dimensional solution manifold for these flows is not necessarily compact, is not a torus and so the KAM theorem does not directly apply. The potential connection between a possible preservation of the solution manifold and the preservation of the Painlevé property is an intriguing one. Also, what would be the analogue of the tori which are not preserved? Are there resonances? The only frequencies which appear in the unperturbed problem are constants, either zero or one.

We now return briefly to some further details in the derivation of (5.9). Why, in general, does the expression

$$E = 1/\zeta\ (2da\psi_1\psi_2 - db\psi_2^2 + dc\psi_1^2)$$

naturally involve the conjugate variables $d\bar{c}_j$, $d\bar{b}_j$? The answer is that the basis vectors V_k must be carefully constructed in order that their inner products can be calculated. For example, when n=4, ignoring the terms arising from δt_j,

$$E = dc_2(\zeta^2\psi_2^2 + ib_2\zeta\psi_1\psi_2 + ib_3\psi_1\psi_2). - db_2(\zeta^2\psi_1^2 - ic_2\zeta\psi_1\psi_2 - ic_3\psi_1\psi_2)$$

$$+ dc_3(\zeta\psi_1^2 + ib_2\psi_1\psi_2) - db_3(\zeta\psi_2^2 - ic_2\psi_1\psi_2) + dc_4\psi_1^2 - db_4\psi_2^2.$$ If one were to define V_k by

writing E as an inner product between $(dc_2,db_2,dc_3,db_3,dc_4,db_4)^T$ and JV_k, the formula (A.14 of Appendix 2 would not apply. It turns out we must add $-\frac{1}{2}a_3\psi_1^2 - \frac{1}{8}b_2^2\psi_2^2$ and $-\frac{1}{2}a_3\psi_2^2 - \frac{1}{8}c_2^2\psi_1^2$ to the coefficients of dc_2 and $-db_2$. Subtracting these terms from the rest of the expression E leads to a redefinition of b_4 and c_4 to $\bar{b}_4$ and $\bar{c}_4$. Also there are terms proportional to δt_j left over which combine in just the correct way to make (5.9) hold. It now also becomes clear why α_{n+1} must factor into a sum of products of the conjugate variables. These terms cancel the potential logarithmic divergences. For n=4, the basis vectors are

$$V_k = \begin{matrix} \int_{\infty_k}^{\zeta_{k+1}} d\zeta\{(\zeta^2 - ia_{3/2})\psi_2^2 - ic_2\psi_1\psi_2 - ic_3\psi_1\psi_2 - \frac{1}{8}c_2^2\psi_1^2\} - is_k\bar{c}_4\ell n\zeta_{k+1} \\ \int_{\infty_k}^{\zeta_{k+1}} d\zeta\{(\zeta^2 - \frac{ia_3}{2})\psi_1^2 + ib_2\zeta\psi_1\psi_2 + ib_3\psi_1\psi_2 - \frac{1}{8}b_2^2\psi_2^2\} - is_k\bar{b}_4\ell n\zeta_{k+1} \\ \int_{\infty_k}^{\zeta_{k+1}} d\zeta\{(\psi_2^2\theta - ic_2\psi_1\psi_2 - is_k\bar{c}_3\ell n\zeta_{k+1} \\ \int_{\infty_k}^{\zeta_{k+1}} d\zeta\{\zeta\psi_1^2 + ib_2\psi_1\psi_2\} - is_k\bar{b}_3\ell n\zeta_{k+1} \\ \int_{\infty_k}^{\zeta_{k+1}} d\zeta\psi_2^2 - is_k\bar{c}_2\ell n\zeta_{k+1} \\ \int_{\infty_k}^{\zeta_{k+1}} d\zeta\psi_1^2 - is_k\bar{b}_2\ell n\zeta_{k+1}. \end{matrix} \qquad (5.16)$$

Recall that the vector $\psi = (\psi_1,\psi_2)^T$ is the dominant solution in sector S_k.

5b. Poisson brackets of the Stokes multipliers.

We have already defined the Poisson bracket in (3.29). Since δs_k does not depend on δX, δT_2,---δT_{n-2}, we have

$$\begin{aligned} [s_k,s_\ell] &= \sum_{2}^{n} \frac{\partial s_k}{\partial b_j}\frac{\partial s_\ell}{\partial \bar{c}_{n+2-j}} - \frac{\partial s_k}{\partial \bar{c}_{n+2-j}}\frac{\partial s_\ell}{\partial b_j} \\ &= -\langle \nabla s_k, J\nabla s_\ell\rangle \\ &= -\langle (-1)^k JV_k, (-1)^\ell J^2 V_\ell\rangle \text{ from (5.9)}, \\ &= (-1)^{k+\ell} M_{k\ell}. \end{aligned} \qquad (5.17)$$

AD-A119 843 CLARKSON COLL OF TECHNOLOGY POTSDAM NY DEPT OF MATHE--ETC F/G 12/1
NONLINEAR WAVES.(U)
SEP 82 M J ABLOWITZ, D J KAUP, A C NEWELL N00014-76-C-0867
UNCLASSIFIED NL

6 of 7

Then the (k,ℓ) element of $P = \{[s_k,s_\ell]\}$, the matrix of Poisson brackets, is $(-1)^{k+\ell}M_{k\ell}$. Define $N=M^{-1}$; then the (k,ℓ) element of $(P^{-1})^T$ is $(-1)^{k+\ell}N_{\ell k}$

5c. Preservation of the symplectic form.

We now calculate the symplectic form

$$\omega = \sum_2^n \delta\bar{b}_j\wedge\delta\bar{c}_{n+2-j} + \delta x\wedge\delta X + \sum_2^{n-2}\delta t_j\wedge\delta T_j$$

$$= -\langle\delta_1 R, J\delta_2 R\rangle + \delta x\wedge\delta X + \sum_2^{n-2}\delta t_j\wedge\delta T_j \tag{5.18}$$

in terms of the infinitesimal variations of the new coordinates $s_k,\delta x,\delta t_j,\delta H_1,\delta H_j$. First, recall that

$$\delta R = dR - J\nabla H_o\delta t_o - J\nabla H_1\delta x - \sum_2^{n-2} J\nabla H_k\delta t_k. \tag{5.19}$$

Second, observe that since

$X = H_1-\hat{H}_1(\bar{b}_k,\bar{c}_k,x,t_k)$, $T_j = H_j-\hat{H}_j(\bar{b}_k,\bar{c}_k,x,t_k)$, we have

$$\delta X = \delta H_1 - \langle\nabla H_1,\delta R\rangle - H_{1x}\delta x - \sum_2^{n-2} H_{1t_k}\delta t_k \tag{5.20}$$

$$= \delta H_1 - \langle\nabla H_1, dR\rangle + (\langle\nabla H_1, J\nabla H_1\rangle - H_{1x})\delta x$$

$$+ \sum_2^{n-2}(\langle\nabla H_1, J\nabla H_k\rangle - H_{1t_k})\delta t_k,$$

$$\delta T_j = \delta H_j - \langle\nabla H_j, dR\rangle + (\langle\nabla H_j, J\nabla H_1 - H_{jx}(\delta x$$

$$+ \sum_2^{n-2}(\langle\nabla H_j, J\nabla H_k - H_{jt_k})\delta t_k. \tag{5.21}$$

In (5.20), (5.21) H_{1x} means $\frac{\partial H_1}{\partial x}$, the partial derivative of H_1 with respect to explicit x. Also in both $\delta X,\delta T_j$, the terms proportional to δt_o are of the form $\langle J\nabla H_o,\nabla H_k\rangle = [H_k,H_o] = 0$ (since H_o is independent of t_k and t_o never explicitly appears anywhere). In these equations we have also replaced $\nabla\hat{H}_k$, $\hat{H}_{kx}$ by ∇H_k and H_{kx} since these are respectively equal. Also we recall that since $J^T=-J$, $\langle Ju,v\rangle = \langle v,Ju\rangle = -\langle u,Jv\rangle$. After a little calculation we now find,

$$\omega = \langle -d_1R, Jd_2R\rangle + \delta x\wedge\delta H_1 + \sum_2^{n-2}\delta t_j\wedge\delta H_j. \tag{5.22}$$

The terms proportional to $\delta x\wedge\delta t_j$ and $\delta t_j\wedge\delta t_k$ have as coefficients $[H_j,H_1]$ and $[H_k,H_j]$ respectively, which are zero.

But from (5.13),

$$-\langle d_1R,\, Jd_2R\rangle = -\langle d_1R,\, {}_{j,\ell}\Sigma(-1)^j N_{\ell j}\delta_2 s_j V_\ell\rangle$$

$$= {}_{j,\ell\in Z}\Sigma(-1)^{j+\ell} N_{j\ell}\delta_1 s_\ell \delta_2 s_j \text{ from (5.9)},$$

$$= (\delta_1 s)^T (P^{-1})^T \delta_2 s$$

$$= \langle P^{-1}\delta_1 s, \delta_2 s\rangle$$

where P is the matrix of Poisson brackets. By a judicious choice of basis U such that $\langle JU_k, U_\ell\rangle = \delta_{k,\ell-1}$, $k<\ell$, we can arrange combinations $f_j(s_k)$, $g_j(s_k)$ $j\in Z$ such that

$$\omega = {}_{j=1}^{n-1}\Sigma \delta f_j(s_k)\wedge\delta g_j(s_k) + \delta x\wedge\delta H_1 + {}_{2}^{n-2}\Sigma\, \delta t_j\wedge\delta H_j. \tag{5.23}$$

In (5.23), the only variables which depend on x,t_j are the variables x,t_j $j=2,\text{--}n-2$ themselves. For the t_k flow ($t_{1=x}$) $f_j, g_j, H_1, H_k, k=2,\text{--}n-2$ are constant and

$$t_k = \frac{\partial H_k}{\partial H_k} = 1.$$

5d. *Fourier expansions and contour integral representations*.

We focus on the case n=3 and take $t_o=0$. The equations (3.13) are

$$q_{xx}-2q^2r = xq, \quad r_{xx}-2qr^2 = rxr. \tag{5.24}$$

From (5.14),

$$FR = -\Sigma_{j,\ell} N_{\ell j} s_j V_\ell \tag{5.25}$$

with V given by (5.8). In this case, there are six sectors at $\zeta=\infty$. We will choose V as being formed from sectors 1,2, 4 and 5 and the relevant Stokes multipliers are s_1, s_2, s_4, s_5. In this case

$$N = \frac{1}{\pi(1+s_1s_2)(1+s_4s_5)} \begin{pmatrix} 0 & -1-s_4s_5 & s_2s_5 & -s_2s_4 \\ 1+s_4s_5 & 0 & -s_1s_5 & s_1s_4 \\ -s_2s_5 & s_1s_5 & 0 & -1-s_1s_2 \\ s_2s_4 & -s_1s_4 & 1+s_1s_2 & 0 \end{pmatrix}$$

and ($\alpha_1=i$)

$$\pi\begin{Bmatrix} c_2 \\ -b_2 \\ c_3 \\ -b_3 \end{Bmatrix} = \pi\begin{Bmatrix} r \\ -q \\ \frac{-ir_x}{2} \\ \frac{-iq_x}{2} \end{Bmatrix} = \frac{s_2}{1+s_1s_2}V_1 - \frac{s_1}{1+s_1s_2}V_2 + \frac{s_5}{1+s_4s_5}V_4 - \frac{s_4}{1+s_4s_5}V_5. \tag{5.26}$$

Case 1. The connection with the squared eigenfunction expansions of inverse scattering.

Originally we were looking for a four parameter (complex) family of solutions for (5.24). Let us look at the two parameter family formed by setting $s_2=s_5=0$. Then, from (5.26)

$$\begin{pmatrix} q \\ -r \end{pmatrix} = \frac{s_1}{\pi}\int_{\infty_2}^{\infty_3} \begin{pmatrix} \psi_1^2 \\ \psi_2^2 \end{pmatrix} d\zeta + \frac{s_4}{\pi}\int_{\infty_4}^{\infty_5} \begin{pmatrix} \psi_1^2 \\ \psi_2^2 \end{pmatrix} d\zeta \tag{5.27}$$

where ψ refers to the uniquely defined dominant solution in each of the Sectors 2 and 5. But the ψ in S_2 is simply the recessive solution in S_1 and, since $s_2=0$, also the recessive solution in S_3. Hence $\psi e^{-i\zeta^3/3}$ is a solution of both (3.1a) and (3.1c) which is analytic for $Im\zeta>0$ and asymptotes to $\begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta x}$ as $\zeta\to\infty$ in $Im\zeta>0$. This is precisely the solution ψ we defined in papers [9,10] in connection with the scattering problem (3.1a). Similarly the solution $\psi e^{i\zeta^3/3}$ of S_5 is what we called $\bar{\psi}$ in [9] and [10]. Now since the ψ of $S_2(S_5)$ is recessive in S_1,S_3 (S_4,S_6), we may extend the end points of the integration paths in (5.32) to $-\infty$ and ∞ on the real ζ axis. Thus (5.27) becomes

$$\begin{pmatrix} q \\ -r \end{pmatrix} = \frac{-1}{\pi}\int_{-\infty}^{\infty} s_1 e^{\frac{2}{3}\zeta^3} \begin{pmatrix} \psi_1^2 \\ \psi_2^2 \end{pmatrix} d\zeta + \frac{1}{\pi}\int_{-\infty}^{\infty} s_4 e^{\frac{-2}{3}\zeta^3} \begin{pmatrix} \bar{\psi}_1^2 \\ \bar{\psi}_2^2 \end{pmatrix} d\zeta \tag{5.28}$$

which is precisely equation 6.55 of reference [10] with

$$\frac{b}{a}(\zeta) = +s_1 e^{2\zeta\frac{3}{3}} \qquad \frac{\bar{b}}{\bar{a}} = s_4 e^{\frac{-2}{3}\zeta^3} \tag{5.29}$$

But if we do define $b/_a$, $\bar{b}/_{\bar{a}}$ in the usual way for (3.1a), then (3.1c) shows that

$$\frac{b}{a}(\zeta) = \frac{b}{a}(0)e^{2\zeta^3/3}, \quad \frac{\bar{b}}{\bar{a}}(\zeta) = \frac{\bar{b}}{\bar{a}}(0)e^{-2\zeta^3/3} \tag{5.30}$$

and so the Stokes multipliers s_1 and s_4 are simply $b/_a(\zeta=0)$ and $-\bar{b}/_{\bar{a}}(\zeta=0)$ respectively. We showed in [1], that if $r=q$, $s_4=s_1$. We also know in this case that $\bar{b}(\zeta)=-b(-\zeta)$ and $\bar{a}(\zeta)=a(-\zeta)$ which is consistent.

Case 2. Contour integral representations of Painlevé functions.

Again we let r=q. From previous work [1] we know $s_4=s_1$, $s_5=s_2$ and the ψ of $S_4(S_5)$ is $\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\bar{\psi}(-\zeta)$ where the latter refers to the recessive solution of sector 1 (2). Changing the integration variable in V_4 and V_5 of (5.26), we obtain

$$q = \frac{s_2}{1+s_1s_2}\int_{\infty_1}^{\zeta_2}(\psi_1^2-\psi_2^2)d\zeta - 2is_1q\,\ell n\zeta_2$$

$$= \frac{s_1}{1+s_1s_2}\int_{\infty_2}^{\zeta_3}(\bar{\psi}_1^2-\bar{\psi}_2^2)d\zeta - 2is_2q\,\ell n\zeta_3 \qquad (5.31)$$

where $\psi,(\bar{\psi})$ are the dominant (recessive) solutions of S_1. Now the contour of the first integral can be extended back to ∞_6 as ψ is recessive in S_6. Similarly, we can set the ∞_2 of the second integral to ∞_1. Now the contours are the same as those used in the integral definitions of Airy functions. In particular, if $s_2=0$ which we have shown is equivalent to picking that class of solutions of (5.24) which decay at $x=\pm\infty$ and admit construction by inverse scattering,

$$q = +s_1\int_{-\infty}^{\infty}(\bar{\psi}_1^2-\bar{\psi}_2^2)d\zeta \qquad (5.32)$$

In the limit of small amplitudes,

$$\bar{\psi} \sim \begin{pmatrix}0\\ 1\end{pmatrix} e^{i\zeta^3/3+i\zeta x}$$

and (5.) is

$$q = -s_1\int_{-\infty}^{\infty} e^{2i\zeta x+2i\zeta^3/3}\,d\zeta$$

which is $-s_1\pi\,A_i(x/4)$.

These representations will be very important when we treat the equivalence between Painlevé equations and particle systems which parallels the analogy between the finite gap solutions and particle systems [5].

Appendix I: Asymptotic Expansions.

We write equation (3.1c) in the form

$$V_\zeta = \frac{1}{\zeta}(-aY_1+bY_2+cY_3)V \qquad (A.1)$$

in the notation (2.3). Let

$$V \sim (I + \sum_1^\infty \frac{C_r}{\zeta^r})e^{-\theta Y_1}\psi \qquad (A.2)$$

with

$$\theta_\zeta = \alpha_1\zeta^{n-1}+\hat{\alpha}_3\zeta^{n-3}+\ldots+(\hat{\alpha}_n+ix) + \frac{1}{\zeta}\hat{\alpha}_{n+1}+\cdots \qquad (A.3)$$

$$\psi_3 = \frac{\delta_{n+3}}{\zeta^3} + \text{----}. \tag{A.4}$$

In (A.3), the integration constant is it_o, in (A.4) it is zero. By induction, we show that for $k \leq n$

$$C_k = -\frac{1}{2}(\hat{b}_{k+1}\gamma_2 - \hat{c}_{k+1}\gamma_3) \tag{A.5}$$

where $\hat{b}_{k+1}$, $\hat{c}_{k+1}$ are defined recursively as

$$\hat{b}_{k+1} = b_{k+1} + \frac{1}{2}\sum_{\ell=3}^{k}(a_\ell + \alpha_\ell)\hat{b}_{k-\ell+2}, \hat{c}_{k+1} = c_{k+1} + \frac{1}{2}\sum_{3}^{k}(a_\ell + \alpha_\ell)\hat{c}_{k-\ell+2} \tag{A.6}$$

$$\hat{b}_2 = b_2, \hat{b}_3 = b_3, c_2 = \hat{c}_2, \hat{c}_3 = c_3.$$

Also

$$\hat{\alpha}_{n+1-j} = \alpha_{n+1-j} \quad j=2,\text{---}n. \tag{A.7}$$

After this stage, the terms from differentiating the bracket term on the right hand side of (A.2) enters and disrupts the pattern.

$$C_{n+1} = -\frac{1}{2}(\hat{b}_{n+2} - \frac{1}{2}\hat{b}_2)\gamma_2 + \frac{1}{2}(\hat{c}_{n+2} + \frac{1}{2}\hat{c}_2)\gamma_3,$$

$$C_{n+2} = -\frac{1}{2}(\hat{b}_{n+3} - i\hat{b}_3)\gamma_2 + \frac{1}{2}(\hat{c}_{n+3} + i\hat{c}_3)\gamma_3,$$

$$C_{n+3} = -\frac{1}{2}(\hat{b}_{n+4} - \frac{3i}{2}\hat{b}_4 - \frac{1}{2}\delta_{n+3}\hat{b}_2)\gamma_2$$

$$+ \frac{1}{2}(\hat{c}_{n+4} + \frac{3i\hat{}}{2}c_4 + \frac{1}{2}\delta_{n+3}\hat{c}_2)\gamma_3,$$

with $\delta_{n+3} = \frac{b_2c_2}{4}$, $\delta_{n+4} = 3/8(b_2c_3 + b_3c_2)$,

$$\hat{\alpha}_{n+4} = \alpha_{n+4} + 1/8(b_2c_3 - b_3c_2).$$

In particular we often have occasion to need to calculate $\psi_1\psi_2$ on R_1 and R_2 as $\zeta \to \infty$. On R_1, $\psi \sim \binom{1}{0}e$, and thus $\psi_1\psi_2 \sim 0$. On R_2, however, $\psi \sim \binom{1}{0}e^{\theta} - s_1\binom{0}{1}e^{-\theta}$ and thus $\psi_1\psi_2 \sim -s_1$. In order to calculate the behavior of the components of V_k, we need to keep more terms in the expansions.

Appendix II: The inner product (5.12).

We will illustrate the means of proof in the case n=3. Let us examine

$$\langle JV_1, V_2\rangle = \left\{ \begin{array}{l} \int_{\zeta_1}^{\zeta_2}(\zeta\psi_1^2+ib_2\psi_1\psi_2)d\zeta - is_1b_3\ell n\zeta_2 \\ -\int_{\zeta_1}^{\zeta_2}(\zeta\psi_2^2-ic_2\psi_1\psi_2)d\zeta - is_1c_3\ell n\zeta_2 \\ \int_{\zeta_1}^{\zeta_2}\psi_1^2 d\zeta - is_1b_2\ell n\zeta_2 \\ -\int_{\zeta_1}^{\zeta_2}\psi_2^2 d\zeta - is_1b_2\, n\ 2 \end{array} \right\} \cdot \left\{ \begin{array}{l} \int_{\zeta_2^+}^{\zeta_3}\bar{\psi}_2^2 d\zeta + is_2c_2\ell n\zeta_3^+ \\ \int_{\zeta_3^+}^{\zeta_3^+}\bar{\psi}_1^2 d\zeta - is_2b_2\ell n\zeta_3^+ \\ \int_{\zeta_2^+}^{\zeta_3^+}(\zeta\bar{\psi}_2^2-ic_2\bar{\psi}_1\bar{\psi}_2)d\zeta + is_2c_3\ell n_3 \\ \int_{\zeta_2^+}^{\zeta_3^+}(\zeta\bar{\psi}_1^2+ib_2\bar{\psi}_1\bar{\psi}_2)d\zeta - is_2b_3\ell n\zeta_3^+ \end{array} \right\} \qquad (A.9)$$

ψ is the dominant solution of S_1 and $\bar{\psi}$, the recessive solution of S_1, is the dominant solution of S_2. We take ζ_2^+ to be closer to $\zeta=\infty$ than ζ_2 on R_2. The product in in (A.9) has three sets of terms. First, the terms in products of logorithms cancel. Next, the terms proportional to $+is_2\ell og\ \zeta_3^+$ are

$$\int_{\zeta_1}^{\zeta_2}(c_2\zeta\cdot c_3)\psi_1^2 + (b_2\zeta+b_3)\psi_2^2)d\zeta = (\psi_1\psi_2)_{\zeta_1}^{\zeta_2}$$

which in the limit of ζ_1,ζ_2 tending to $\zeta=\infty$ on R_1,R_2 respectively is $-s_1$ (use Appendix I). Similarly the terms proportional to $-is_1\ell n\zeta_2$ are $-s_2$. Hence these two terms together contribute $-i\,s_1\,s_2\ell n^{\zeta_3^+}/_{\zeta_2}$. Next we examine the integral terms which may be written

$$\int_{\zeta_2^+}^{\zeta_3^+}\int_{\zeta_1}^{\zeta_2} d\zeta d\zeta'\{(\zeta+\zeta')(\psi_1^2\bar{\psi}_2^2-\bar{\psi}_1^2\psi_2^2)-ib_2(\psi_2^2\bar{\psi}_1\bar{\psi}_2-\bar{\psi}_2^2\psi_1\psi_2) \qquad (A.11)$$

$$-ic_2(\psi_1^2\bar{\psi}_1\bar{\psi}_2-\bar{\psi}_1^2\psi_1\psi_2)\}$$

where ζ lies in the path between ζ_1,ζ_2 and ζ' on the path between ζ_2^+,ζ_3^+. Now the integral is

$$-i\ \frac{(a/\zeta - a'/\zeta')}{\zeta-\zeta'}(\psi_1^2\bar{\psi}_2^2 - \bar{\psi}_1^2\psi_2^2)\left|\frac{b}{\zeta} - \frac{b'}{\zeta'}\right|_{+\ \zeta-\zeta'}(\psi_2^2\bar{\psi}_1\bar{\psi}_2 - \bar{\psi}_2^2\psi_1\psi_2)$$

$$+\left(\frac{\frac{c}{\zeta}-\frac{c'}{\zeta'}}{\zeta-\zeta'}\right)(\psi_1^2\bar{\psi}_1\bar{\psi}_2 - \bar{\psi}_1^2\psi_1\psi_2) \qquad (A.12)$$

We have an identity: If A,B,C satisfy

$$A_\zeta = \beta C+\gamma B,\ B_\zeta+2\alpha B = 2\beta A,\ C_\zeta-2\alpha C = 2\gamma A \tag{A.13}$$

and A',B',C' satisfy the same equations with ζ replaced by ζ', then

$$(\alpha-\alpha')(BC'-B'C) + (\beta-\beta')(A'C-AC') + (\gamma-\gamma')(A'B-AB')$$

$$= \frac{1}{2}\left(\frac{d}{d\zeta} + \frac{d}{d\zeta'}\right)(2AA'-B'C-BC'). \tag{A.14}$$

Now, in our case $A=\psi_1\psi_2$, $B=\psi_1^2$, $C=\psi_2^2$, $\alpha=\frac{a}{\zeta}$, $\beta=b/\zeta$ $\gamma=c/\zeta$ and thus (A.11) is

$$-\frac{i}{2}\int_{\zeta_3^+}^{\zeta_3^+}\int_{\zeta_1}^{\zeta_2} d\zeta d\zeta' \left(\frac{d}{d\zeta} + \frac{d}{d\zeta'}\right)\frac{w(\zeta,\zeta')}{\zeta-\zeta'} \tag{A.15}$$

where

$$W(\zeta,\zeta') = -(\psi_1(\zeta)\bar{\psi}_2(\zeta') - \psi_2(\zeta)\bar{\psi}_1(\zeta'))^2.$$

Integrating we obtain (A.11) is

$$-\frac{1}{2}\int_{\zeta_1}^{\zeta_2}\left(\frac{W(\zeta,\zeta_3^+}{\zeta-\zeta_3^+} - \frac{W(\zeta,\zeta_2^+)}{\zeta-\zeta_2^+}\right)d\zeta - \frac{1}{2}\int_{\zeta_2^+}^{\zeta_3^+}\left(\frac{W(\zeta_2,\zeta')}{\zeta_2-\zeta'} - \frac{W(\zeta_1,\zeta')}{\zeta_1-\zeta'}\right)d\zeta'. \tag{A.16}$$

Now, we take the path between ζ_1 and ζ_2 to be in along R_1 to the $\zeta=0$ and out along R_2. To integrate the third integral, we recognize that $\bar{\psi}$ is recessive in S_1 and bring the contour from ζ_2^+ from Arg $\zeta = \frac{\pi}{3}$ to Arg $\zeta=0$, then in along R_1 and out R_3 to ζ^+. This means we must take account of the pole at $\zeta'=S_2$ and since

Res. $\frac{W(\zeta_2,\zeta')}{\zeta'=\zeta_2\ \zeta_2-\zeta'} = -W(\zeta_2,\zeta_2) = -1$, since $\psi_1\bar{\psi}_2-\bar{\psi}_1\psi_2$ is the Wronskian. Hence from the third integral we obtain the extra term $-1/2(-2\pi i)(-1) = \pi$. Now on the neutral rays, the only terms which will contribute as $\zeta_1,\zeta_2,\zeta_2^+,\zeta_3^+$ tend to infinity are those along rays where $W(\zeta,\zeta')$ tends to a constant. This only happens when ζ lies on R_2, ζ' on R_3 whence $W(\zeta,\zeta\nearrow+\infty) \to +2s_1s_2$, and thus the remaining terms in (A.16) tend asymptotically to

$$\pi - is_1s_2\int_0^{\zeta_2}\frac{d\zeta}{\zeta-\zeta_3^+} - is_1s_2\int_0^{\zeta_3^+}\frac{d\zeta'}{\zeta_2-\zeta'}.$$

Hence

$$\langle JV_1,V_2\rangle = \pi - is_1s_2\ln\frac{\zeta_2-\zeta_3^+}{-\zeta_3^+} + is_1s_2\ln\frac{\zeta_3^+-\zeta_2}{-\zeta_2} - is_1s_2\ln\zeta_3^2/\zeta_2 = \pi(1+s_1s_2).$$

It is easy to see that if S_k and S_ℓ $(k<\ell)$ are not contiguous sectors, the contribution from the pole disappears. Hence we have

$$\langle JV_k, V_\ell \rangle = (\pi\delta_{k,\ell-1} + s_k s_\ell) \qquad (A.17)$$

This formula also holds for all n>3.

Acknowledgements:

The authors gratefully acknowledge support for this work Grant #DAAG29-81-K-0025, N00014-76-C-0867 and MCS-7903498-01.

References

[1] Flaschka, H. and Newell, A.C., Monodromy and Spectrum Preserving Deformation I, Comm. Math. Phys. 76 (1980) 65-116.

[2] Flaschka, H. and Newell, A.C., Multiphase Similarity Solutions of Integrable Evolution Equations, Lectures in Pure and Applied Mathematics 54 (1980) 373-395.
Also, Proc. Kiev Conference on Solitons, September 1979, to appear Physica D July, 1981.

[3] Lax, P.D., Periodic Solutions of the Korteweg deVries equation, Lectures in Applied Math. 15 (1974) 85-96.

[4] Novikov, S.P., The Periodic Problem for the Korteweg deVries Equation, Funkts. Anal. Prilozen 8 (1974) 54-66.

[5] Flaschka, H. and Newell, A.C., Monodromy and Spectrum Preserving Deformations II, III. In preparation.

[6] Newell, A.C., The General Structure of Integrable Evolution Equations, Proc. Roy. Sco. A 365, (1979) 283-311.

[7] Jimbo, M., Miwa, T., Mori, Y. and Sato, J., Density Matrix of an Inpenetrable Bose Gas and the Fifth Painlevé Transcendent, Physica D 1, (1980) 80-158.

[8] Gel'fand, I.M. and Dikii, L.A., Integrable Nonlinear Equations and the Liouville Theorem, Funkt. Analiz Priloz. 1, (1979) 8-20.

[9] Ablowitz, M.J., Kaup, D.J., Newell, A.C. and Segur, H., The Inverse Scattering Transform, Stud. Appl. Math. 53, (1974) 249.

[10] Newell, A.C., The Inverse Scattering Transform, Topics in Current Physics 17, (1980) 177-242. Solitons, Eds. R. Bullough and P. Caudrey, Springer-Verlag.

1.3

Physica 3D (1981) 1 & 2, 203-221
© North-Holland Publishing Company

MULTIPHASE SIMILARITY SOLUTIONS OF INTEGRABLE EVOLUTION EQUATIONS

H. Flaschka
Department of Mathematics
University of Arizona
Tucson, AS 87521

A.C. Newell
Department of Mathematics
and Computer Science
Clarkson College
Potsdam, NY 13676

We introduce a new class of solutions to integrable nonlinear evolution equations which are thought to have a deep connection with the n-point correlation functions of exactly solvable models in statistical mechanics.

1. INTRODUCTION

In our earlier paper [1], we used deformation theory to study the Painlevé equations which govern the self-similar solutions of the modified KdV (MKdV) and sinh-Gordon equations. Here we introduce a wider class of solutions of these and related equations: the multiphase similarity solutions. These are supposed to be to the standard (one-phase) similarity solutions what multi-soliton or finite-gap periodic solutions are to solitons and cnoidal waves.

Our study of multiphase similarity solutions originated in attempts to understand the work of Sato, Miwa and Jimbo [2] on the scaling limit of the n-point correlation functions of the Ising model. The 2-point function can be expressed in terms of the 3rd Painlevé transcendent [3]; the same Painlevé function also governs the selfsimilar solution of the sinh-Gordon equation. It was shown in [2] that the Ising n-point function is expressible by what appears to be a multivariable generalization of Painlevé transcendents. We conjecture that these many phase Painlevé functions are somehow analogous to the θ-functions which afford formulas for finite-gap solutions of integrable systems, and that correspondingly the concept of similarity should have a natural generalization to many phases. The present note reports some preliminary work in this direction.

Our earlier approach [1] to Painlevé equations leads to a very natural definition of multiphase similarity solutions. We explain the definition, and the philosophy behind it, for the MKdV family; it will be clear how to adapt it to other systems. The Ising n-point functions fit into the same framework, as we describe below. The analytical properties of multiphase similarity solutions, and the geometric theory underlying their definition, are not yet understood. We limit ourselves to a description of the basic facts and of some very suggestive analogies with the Riemann-surface theory of finite-gap solutions. In particular, we will write down an equation which describes the shape of the multiphase similarity solution as function of x; we will also give the equation which describes the interaction of a similarity solution with a soliton.

We believe that a thorough study of similarity solutions of integrable systems will require the development of interesting new mathematical ideas. A hint of the results to be discovered may be found in some early papers of Garnier [4,5]. These contain the beginning of a theory of partial differential equations with fixed critical points; the solutions of such equations appear to be generalizations of abelian functions (if one adopts the point of view from which Painlevé

transcendents are seen as generalization of elliptic functions). As far as we know, Garnier's ideas have not been developed during the past 60 years; some reinterpretation of his work in the context of soliton theory will be given in [6] and in a paper (in preparation) which will elaborate on the present note.

2. REVIEW OF FINITE-GAP SOLUTIONS.

We consider nonlinear evolution equations for the elements of an N x N matrix P, arising as integrability conditions for the linear systems [7]

$$V_x(x,\bar{t}) = (\zeta R_o + P(x,\bar{t}))V(x,\bar{t}) \tag{2.1}$$

$$V_{t_j}(x,\bar{t}) = Q^{(j)}(x,\bar{t},\zeta)V(x,\bar{t}) \tag{2.2}$$

Here $\bar{t}=(t_1,\ldots,t_n)$, V is an N-vector, R_o is a constant diagonal matrix $(i\beta_j\delta_{jk})$ and the $Q^{(j)}$ are matrices which depend polynomially on the parameter ζ and on $P,P_x,\ldots$. For simplicity, we do not at this time allow dependence of $Q^{(j)}$ on ζ^{-1}; see Section 5 for a more general situation. For a definition of $Q^{(j)}$, see the appendix.

The conditions $V_{xt_j} = V_{t_jx}$ etc. imply

$$P_{t_j} - Q^{(j)}{}_x + [P,Q^{(j)}] = 0 \tag{2.3}$$

$$Q^{(i)}{}_{,t_j} - Q^{(j)}{}_{,t_i} + [Q^{(i)},Q^{(j)}] = 0 \tag{2.4}$$

Example: $R_o = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}$, $P = \begin{bmatrix} 0 & q \\ q & 0 \end{bmatrix}$, $Q^{(1)} = P$,

$$Q^{(3)} = \begin{bmatrix} -4i\zeta^3 - 2iq^2\zeta & 4\zeta^2 q + 2iq_x\zeta - q_{xx} + 2q^3 \\ 4\zeta^2 q - 2i\zeta q_x - q_{xx} + 2q^3 & 4i\zeta^3 + 2iq^2\zeta \end{bmatrix}$$

Then $(2.3)_1$ and $(2.3)_3$ are the first two equations in the MKdV hierarchy,

$$q_{t_1} = q_x, \quad q_{t_3} = 6q^2q_x - q_{xxx}.$$

A somewhat more complicated choice for $Q^{(5)}$ leads to the next MkdV equation,

$$q_{t_5} = (q_{xxxx} - 10q^2q_{xx} - 10\,qq_x^2 + 6q^5)_x.$$

Conditions (2.4) hold automatically, because the MKdV flows commute [8]. Thus we may and do think of q as function of x and $t_1,\ldots,t_n$ simultaneously, satisfying the j^{th} equation of the MKdV hierarchy in the variable t_j.

We are interested in special solutions of (2.3), solutions which in some sense have finitely many degrees of freedom. For MKdV, for example, one has rational

[9] or multisoliton solutions [10], both of which are special cases of the finite-gap solutions. We summarize the facts needed for our later discussion.

Finite-gap solutions are obtained by subjecting the vector V in (2.1), (2.2) to a further (algebraic) constraint:

$$\sum_1^n u_j V_{t_j} = \lambda V \tag{2.5}$$

or equivalently

$$\left(\sum_1^n u_j Q^{(j)} - \lambda \right) V = 0. \tag{2.6}$$

(2.6) has a nontrivial solution only if the determinant of $\Sigma u_j Q^{(j)} - \lambda I$ vanishes. Cross-differentiation of (2.5) and (2.1) leads to

$$\sum_1^n u_j P_{t_j} = 0. \tag{2.7}$$

This is a further restriction on the matrix P, supplementing $(2.3)_j$ and consistent with it. Specifically, from $(2.3)_j$ one sees that P_{t_j} is a polynomial in $P, P_x, \ldots$, so that (2.7) is an ordinary differential equation in x for the matrix P. The finite-dimensional solution manifold of this equation is left invariant by the flows (2.3). (2.7) is often called a Lax-Novikov equation [11,12].

Solutions of the equation (2.7) (as well as their time evolution under $(2.3)_j$ can be constructed explicitly; they are abelian functions. The connection with Riemann surfaces comes about as follows. The vanishing of the determinant in (2.6) imposes a polynomial relation $\Gamma(\zeta,\lambda) = 0$ between ζ and λ. The points of the Riemann surface

$$R: \quad \Gamma(\zeta,\lambda) = 0 \tag{2.8}$$

parametrize $V(x,t)$ in (2.1), (2.3), (2.6). By proper normalization of V as eigenvector of (2.7), one can arrange that at the point(s) at infinity of (2.8),

$$V \sim \left[1 + O\left(\frac{1}{k}\right)\right] \exp\left(kx + q_1(k)t_1 + \ldots + q_n(k)t_n\right) \tag{2.9}$$

k being a local parameter. The $q_j(k)$ are polynomials, usually the dispersion relation of the linearized equation $(2.3)_j$. In addition, V as function of (ζ,λ) on R has poles μ_k which are independent of x and t_j. Specification of the surface R, the poles μ_k, the asymptotics (2.9), and of a certain normalization condition determines V uniquely. Knowing V one can, by differentiation with respect to x, recover $P(x,t)$ from (2.1).

For details of this theory, we refer to [13]. In the next section, we develop a difinition of multiphase similarity solution along the same lines.

3. MULTIPHASE SIMILARITY SOLUTIONS

We again impose a supplementary constraint on the solution of the linear system (2.1), (2.2). The constraint is now not algebraic, but is given by a differential equation in ζ:

$$\zeta V_\zeta = \sum_1^n j t_j V_{t_j} + x V_x \tag{3.1}$$

or

$$\zeta V_\zeta = \left(\sum_1^n j t_j Q^{(j)} + x(\zeta R_o + P) \right) V. \tag{3.2}$$

Equation (3.2) has coefficients which are rational functions of ζ; all dependence on x and the t_j is only parametric as far as ζ is concerned.

The integrability conditions $(2.3)_j$, (2.4) are now supplemented by ($V_{\zeta x} = V_{x\zeta}$ etc.)

$$0 = (xP)_x + \Sigma j t_j P_{t_j} \tag{3.3}$$

Just as in Section 2, we are specifying a particular class of solutions by constraints on the ζ-dependence of $V(x,t,\zeta)$. In Section 2, V was seen to be meromorphic on a Riemann surface $\Gamma(\zeta,\lambda)=0$. Equation (3.2), as we review below (see [1] for more discussion) also imposes certain global behavior on the function $\zeta \to V(\zeta)$. *This behavior will be seen to characterize V.*

(3.3) is the analog of the Lax-Novikov equation (2.7). For fixed $\bar{t}$, it is a non-autonomous o.d.e. in x. A solution of this o.d.e., after evolving under the flows $(2.3)_j$, will at any later t_j again satisy (3.3) (with new coefficients t_j).

To see in what sense (3.3) exhibits similarity behavior, consider the MKdV example, and the simplest case of (3.3),

$$(xP)_x + 3t_3 P_{t_3} = 0. \tag{3.4}$$

Written out, this is:

$$(xq)_x + 3t_3(6q^2 q_x - q_{xxx}) = 0. \tag{3.5}$$

The characteristics of (3.4) are given by

$$\frac{dx}{x} = \frac{dt_3}{3t_3}$$

and $q(x,t_3)$ has the form $(3t_3)^{-1/3}\bar{q}(x(3t_3)^{-1/3})$. So $q(x,t_3)$ is self-similar, and by (3.5) it satisfies the MKdV equation, $q_{t_3} = 6q^2 q_x - q_{xxx}$. $\bar{q}$ as function of the similarity variable $x(3t_3)^{-1/3} = \xi$ satisfies the second Painlevé equation,

$$\bar{q}_{\xi\xi} = 2\bar{q}^3 + \xi\bar{q} - \nu$$

which we studied in [1].

The general equation (3.3) has characteristics

$$\frac{dx}{x} = \frac{dt_1}{t_1} = \ldots = \frac{dt}{nt_n}$$

and its solution will be of the form

$$P = x^{-1}\,\bar{P}\,(xt_1^{-1},\ldots,\, x(t_n)^{-1/n}).$$

We now explain in what sense (3.3) is a natural generalization of a standard self similarity condition such as (3.5). The difficulty is that there is not, in the literature, a well-developed concept of multiphase similarity. The notion of multiply periodic (quasiperiodic or almost periodic) wave *trains* is, on the other hand, well understood. Our definition is based on certain properties of the solutions of (3.2), which are analogous to properties of $V(x,t,(\zeta,\lambda))$ described in Section 2 in connection with the theory of finite-gap potentials.

(3.2) is a system with an irregular singular point at $\zeta = \infty$ and a regular singular point at $\zeta = 0$. The solution $V(\zeta)$ of such a system is completely characterized by the so-called monodromy data [1]. For clarity, we describe these data for the special example

$$\zeta V_\zeta = (t_3 Q^{(3)} + t_5 Q^{(5)} + xP)\,V \qquad (3.6)$$

in the MKdV case.

1) Near $\zeta = \infty$, (3.6) has a formal fundamental solution

$$\tilde{\Psi}(x,t,\zeta) = \hat{\Psi}(x,t,\zeta) \begin{bmatrix} e^{-\theta} & 0 \\ 0 & e^{+\theta} \end{bmatrix},$$

$\theta = i(5t_5\zeta^5 + 3t_3\zeta^3 + x\zeta)$, where $\hat{\Psi} = \sum_0^\infty C_j\zeta^{-j}$ is a formal Laurent series. In each sector S_j, $\frac{\pi}{5}(j-1) \leq \arg\zeta \ll \frac{\pi}{5}j$, there exists a true solution Ψ_j for which $\tilde{\Psi}$ is an asymptotic expansion (in S_j). Then $\Psi_{j+1} = \Psi_j A_j$; the A_j are called <u>Stokes matrices</u>.

2) Near $\zeta = 0$, (3.6) has a solution of the form

$$\Phi(x,t,\zeta) = \hat{\Phi}(x,t,\zeta) \begin{bmatrix} \zeta^{-\nu} & 0 \\ 0 & \zeta^{+\nu} \end{bmatrix}.$$

$\nu \notin \mathbb{Z}$ is a constant of integration of (3.3) [1]; if ν = half-integer, there may be logarithms in the series as well. The monodromy matrix at $\zeta = 0$ is J,

$$\Phi(\zeta e^{2\pi i}) = \Phi(\zeta)\,J.$$

3) A connection matrix is needed to specify the relation between Φ (normalized at $\zeta = 0$) and Ψ_1 (normalized at $\zeta = \infty$ in sector S_1):

$$\Phi = \Psi_1 A.$$

One then has the following result (just as in [1]):

All matrices $A_1,\ldots,A_{10}$, J, A are independent of x, t_3, t_5, iff $V_x = (\zeta R_0 + P)V$,

$V_{t_3} = Q^{(3)}V$, and $V_{t_5} = Q^{(5)}V$. Equivalently, the self-similar equation

$$xq + 3t_3(2q^3 - q_{xx}) + 5t_5(q_{xxxx} - 10q^2q_{xx} - 10qq_x^2 + 6q^5) = \nu \qquad (3.7)$$

is the condition that a deformation of the system (3.6) leave the monodromy of its solutions invariant.

Following [1], it is possible to reconstruct Ψ_1, given $A_1,\ldots,A_{10}$, J, A and $\theta(x,t_3,t_5)$, *through the solution of a system of* <u>*linear*</u> singular integral equation. Knowing Ψ_1, one can reconstruct q from the equation $\Psi_{1x} = (\zeta R_0 + P)\Psi_1$. In principle, the solution of (3.7) has thereby been reduced to a linear problem.

It will be clear that analogous considerations apply to the general equation (3.3). The number of sectors (and hence of Stokes matrices) will increase, and explicit computations will become quite unmanageable. In principle, however, the fundamental solution of (3.1) (and hence the potential P) can always be reconstructed from knowledge of the monodromy data and of the exponent θ *in the* asymptotic expansion.

It is of particular interest in this connection that the only occurrence of x,t_j among the data characterizing the fundamental solution of (3.1) is in the exponent θ. In the general MKdV case, for example,

$$x \qquad \tilde{\Psi}(x,t,\zeta) = \hat{\Psi}(x,t,\zeta) \begin{bmatrix} e^{-\theta} & 0 \\ 0 & e^{+\theta} \end{bmatrix},$$

where $\theta = \Sigma\, jt_j\zeta^j$. This, it will be remembered from Section 2, is exactly the situation encountered in the finite-gap theory (see equation (2.9)). Indeed, the relationship between the analyticity properties of $V(\zeta)$ discussed in Sections 2,3 goes quite far. In Section 2, V was meromorphic on a Riemann surface, except at ∞ where it was seen to have an essential singularity, with behavior e^{θ} times a convergent power series (i.e., Stokes matrices are all the identity). In Section 3, V is meromorphic on a possibly infinite-sheeted covering of the sphere (determined by the nature of the regular singular point at $\zeta = 0$), except at ∞ where it has an essential singularity, with a behavior of e^{θ} times an asymptotic power series (Stokes matrices are not the identity).

This point of view, on the one hand, places our definition of "multiphase *similarity*" *in contact with the well-developed quasiperiodic theory*, and makes it more motivated than it might have seemed at first glance. On the other hand, this perspective suggests further generalizations of the solutions presented here. Namely, one may consider systems of differential equations on compact Riemann surfaces, with various regular and irregular singular points. The deformations preserving monodromy will lead to nonlinear equations including, but extending, the equations described above. One such equation, based on work of K. Ueno [14], is given in Section 6.

4. THE PERIODIC LIMIT

There is a remarkable principle, so far not really explained but supported by all examples known to us. The nonlinear equations arising from mondromy-preserving deformations are not autonomous (the independent variables, such as x,t_j above, enter explicitly). Replace the explicit independent variables by constants. <u>The resulting autonomous equations are solvable in terms of abelian functions</u> (and in

particular by Krichever's method [13]). One very interesting example, due to Garnier [5], is studied in [6]. Here we describe, in a formal way, the situation for equation (3.3).

In equations (3.2), (3.3)

$$\zeta V_\zeta = (\Sigma\, jt_j Q^{(j)} + x(\zeta R_o + P))V,$$

$$(xP)_x + \Sigma j t_j P_{t_j} = 0,$$

take $t_j = j^{-1} u_j t$, and let $t \to \infty$. Then (3.3) becomes the Lax-Novikov equation (2.7). In (3.2), one now has the small parameter t^{-1} in front of the derivatives, so a WKBJ approximation

$$V = V_o \exp S,\ S = t \int^{\zeta} \lambda(\zeta')\, d\zeta'$$

is natural. The eikonal equation is

$$(Q - \lambda I)V_o = 0, \tag{4.1}$$

where $Q = \Sigma u_j Q^j$. This is precisely (2.7): the vanishing of the determinant in (4.1) defines the Riemann surface of the finite-gap theory.

The large t behavior of the solutions of (3.3) is therefore formally given by abelian functions; this is related to the familiar fact (cf. [15]) that Painlevé transcendents are asymptotically described by modulated elliptic functions. Such approximations are known not to be uniformly valid even in the single-phase case, and precise results will require much more study.

5. FIRST-ORDER MATRIX SYSTEMS

Here we consider briefly the similarity solutions of the systems arising as integrability condition of

$$V_{x_j} = (\zeta R_j + P_j)V \tag{5.1}$$

$$V_{t_j} = \zeta^{-1} Q_j V, \tag{5.2}$$

where R_j = diag $(0,\ldots,1,\ldots 0)$ (1 in the j-th place), $P_j = [R_j, P]$ for some fixed n x n matrix P, and Q_j is independent of ζ.

Cross-differentiation results in the following set of equations.

$$\begin{aligned}
&\text{a)}\ \ P_{j,x_k} - P_{k,x_j} + [P_j, P_k] = 0,\\
&\text{b)}\ \ Q_{k,t_j} = Q_{j,t_k},\\
&\text{c)}\ \ [Q_j, Q_k] = 0,\\
&\text{d)}\ \ P_{j,t_k} = -[R_j, Q_k],\\
&\text{e)}\ \ Q_{k,t_j} = [P_j, Q_k].
\end{aligned} \tag{5.3}$$

For the case n=2, with $x=x_1-x_2$, $t=t_1-t_2$, (5.3) includes the sine- or sinh-Gordon equation. General results about evolution equations of the form (5.3) have been obtained by Newell [7] amd Dubrovin [16]. Dubrovin studies finite-gap potentials for the system (5.1). In his theory, there is no distinguished "eigenvalue problem" in x. Rather, the n variables x_j are on equal footing; if one insists on taking a linear combination $\Sigma a_j V_{x_j}$ of (5.1), with $x=\Sigma a_j x_j$, as a basic scattering problem, the other x_j play the role of time parameters for commuting flows which define a manifold of finite-gap potentials.

The algebraic constraint (cf. Section 2) leading to finite-gap potentials is, in this example,

$$\Sigma u_j V_{x_j} - \Sigma u_j' V_{t_j} = \lambda V.$$

The non-algebraic constraint (cf. Section 3) leading to similarity solutions is

$$\zeta V_\zeta = \Sigma x_j V_{x_j} - t_j V_{t_j} = \Sigma(x_j(\zeta R_j + P_j) - \zeta^{-1} t_j Q_j)V. \tag{5.4}$$

Compatibility of (5.4) and (5.1), (5.2) requires

$$P_k + \Sigma(x_j P_{k,x_j} - t_j P_{k,t_j}) = 0, \tag{5.5}$$

$$\Sigma\,(x_j Q_{k,x_j} - t_j Q_{k,t_j}) = 0. \tag{5.6}$$

The monodromy structure of (5.4) is now due to two irregular singular points, at $\zeta=0$ and $\zeta=\infty$. The behavior of V at these points will be

$$\exp(\zeta x_j R_j)V_\infty \quad \text{at } \infty\,,$$

$$\exp(\zeta^{-1} t_j R_j)V_0 \quad \text{at } 0.$$

Stokes multiplier matrices will be defined across the anti-Stokes lines $\mathrm{Im}(x_j-x_k)\zeta=0$, and again equations (5.1), (5.2) and (5.5), (5.6) are a consequence of the requirement that the Stokes matrices, as well as the $\zeta=0$ to $\zeta=\infty$ connection matrix, be independent of x_j, t_j.

This example is particularly interesting because the deformation equations (5.3), (5.5), (5.6) are just the equations whose solutions were shown by Sato, Miwa and Jimbo to govern the scaling limit of the Ising n-point funtions $\tau_n((x_1,t_1),\ldots,(x_n,t_n))$. It appears from the above sketch that the theory of these functions is closely related to conventional soliton mathematics. Various aspects of this connection are now under investigation.

6. ADDITIONAL EXAMPLES

In [1], we derived formulae for the rational solutions of Painlevé II and of the other equations in the Painlevé II hierarchy (i.e., the self-similar solutions of the higher MKdV equations). These solutions are characterized by the following monodromy data: all Stokes matrices are the identity, the exponent ν at $\zeta=0$ is an integer, and the polynomial θ in the asymptotic expansion is

$\theta = \sum_0^n (2j+1)t_{2j+1}\zeta^{2j+1}$. The resulting solutions are rational functions of $t_1, t_3, \ldots$. For N=2, for example, with the notation $t_1=x$, $t_3=t$, $t_5=y$, we have

$$q(x,t,y) = \frac{1}{x} - \frac{3x^2}{x^3+3t} + \frac{y+tx^2+2/15\,x^5}{xy-t^2+tx^3/3+x^6/15}$$

Such rational solutions are intermediate between finite-gap potentials and multiphase-selfsimilar solutions, being a degenerate case of both types. A rather more interesting example is afforded by results of Ueno [14]. We rephrase Ueno's construction along the lines developed by Manin [17]. When q(x) is a one-soliton potential for MKdV, the eigenfunction $V(x,\zeta)$ of the scattering problem (2.1) is meromorphic on a rational curve with three double points (Ueno uses a characterization of Date according to which the fundamental matrix of (2.1) assumes a Jordan form at those three points; the points are $\pm$ the soliton eigenvalue). At $\zeta=\infty$, V has the asymptotic behavior ζ expi$(\zeta x+\zeta^3 t)$ familiar from scattering theory. It follows that $\zeta \to V(\zeta)$ satisfies on the rational curve a system of differential equations with an irregular singular point at $\zeta=\infty$, the Stokes matrices being the identity, or equivalently, on the Riemann sphere, a system with apparent (removable; see appendix) singularities at the double points, and again an irregular singular point at ∞.

One can modify Ueno's linear system for V to produce a system, still with coefficients rational in ζ, whose solutions may have nonzero Stokes multipliers at ∞. The condition that the mondromy of V be preserved under deformation is again the MKdV scattering equation (2.1). Referring to the general system introduced in the appendix, we take

$$F = D_R^3 + \eta^2 D_R, \quad \Omega = D_R^5 + \eta^2 D_R^3 + \beta D_R$$

whence the equation is

$$(xq)_{xx} + 4Aq - 4\eta^2 xq = \tfrac{1}{4}(q_{xx}-2q^3)_{xx} - q(qq_{xx} - \tfrac{1}{2}q_x^2 - \tfrac{3}{2}q^4)$$

$$- \eta^2(q_{xx}-2q^3) + 4\beta q - 4K, \ K \text{ a constant} \qquad (6.1)$$

$$A_x = -(xq)_x q, \ A(\infty) = 0.$$

The corresponding equation for V as function of ζ has the property that if (a) the Stokes multipliers at ∞ are zero, (b) $\zeta=0$ is an ordinary point (i.e. K=0), and (c) the indicial roots at the apparent singularities $\zeta = \pm i\eta$ are -1,0, then the solution of (6.1) is precisely the soliton $q = 2i\eta$ Sech $2\eta(x-x_0)$ of the MKdV family. If the Stokes multipliers are not zero, we conjecture, and support this conjecture below with some corroborative evidence, that the shape q(x) describes the interaction of a soliton with a one-phase similarity solution of MKdV.

First, as a heuristic guide, let us consider the case K=0 and imagine that q(x) decays to zero at $\pm\infty$ so that the scattering data connected with the eigenvalue problem (2.1) are defined. Using (A.5) we find

$$a_\zeta = \frac{-iA(-\infty)}{\zeta^2+\eta^2}\,a, \quad b_\zeta = 2i\zeta^2 b + i\,\frac{2\beta-A(-\infty)}{\zeta^2+\eta^2}\,b. \qquad (6.2)$$

The choices $A(-\infty) = -2\eta = 2B$ gives us

$$a = a\frac{\zeta - i\eta}{0\zeta + i\eta}, \quad b = b_0 e^{2i\zeta^3/3} \tag{6.3}$$

which is the scattering data for the superposition of a soliton of amplitude $2i\eta$ and a single phase similarity solution. Further if we take the limit $\eta = 0$ and again assume $q \to 0$ as $x \to +\infty$ we find

$$xq = 1/4(q_{xx} - 2q^3); \tag{6.4}$$

that is, q satisfies the one phase selfsimilar equation. Note that if (6.4) holds,

$$4Aq = q(qq_{xx} - 1/2q_x^2 - 3/2q^4).$$

We also observe that as $\eta \to \infty$, (6.4) is again satisfied. We may interpret this as follows. Consider the shape $q(x)$ to consist of a well separated soliton $q = 2i\eta$ Sech $2\eta(x-x_0)$ and a solution of (6.4). If η is large we may neglect all the terms not multiplied by η^2 when $x)x_0=0(1)$. Thus, away from the soliton $q_s=2i\eta$Sech $2\eta(x-x_0)$ (which is an exact solution of (6.1) with K=0), the reduced equation (6.4) describes the shape of $q(x)$. Consider $q(x)=q_s(x)+\hat{q}(x)$. Then $A(x)=A_s(x)+\hat{A}(x) = -xq_s^2/2 + \eta(\tanh 2\eta(x-x_0)-1) - 1/4(\hat{q}\hat{q}_{xx} - 1/2\hat{q}_x^2 - 3/2\hat{q}^4)$ and $\hat{q}$ satisfies (6.4).

Furthermore, if we take $K=\eta^2$, we may interpret one solution of (6.1) as describing the shape of a soliton $q_s=2i\eta$Sech $2\eta(x-x_0)$ and a rational solution $\hat{q}(x)=1/x-c$. The corresponding $A(x)= -xq_s^2/2 + \eta(\tanh 2\eta(x-x)-1) - \frac{c}{2(x-c)^2}$. Now, if $x_0 \gg c$, then in the neighborhood of $x=c$, $A(x) \approx -2\eta - \frac{c}{2(x-c)^2}$ and substitution of this and $\hat{q}(x) = \frac{1}{x-c}$ into (6.1) gives $c = -1/\eta$. On the other hand, if $x_0 \ll c$, then in the neighborhood of $x=c$, $A(x) \approx -\frac{c}{2(x-c)^2}$ and $c = 1/\eta$. Thus the phase shift experienced by the rational solution as the soliton goes from $x_0 = -\infty$ to $x_0 = +\infty$ is $2/\eta$.

This analysis suggests that the structures which are described by the ordinary differential equations (3.3), (3.5), (3.7) are intimately connected to the nature of the singular points of the constraining V_ζ equation (3.1). Associated with the regular singular point is a phase $\theta = \Sigma(2j+1)t_{2j+1}\zeta^{2j+1} + \zeta x$ which, through a rescaling of ζ, is a function of $x(t_r)^{-1/r}, r=1,\ldots n$, and a set of Stokes multipliers which prescribe the amplitudes of the multiphase similarity solutions. Associated with the apparent singular points $\zeta = \pm i\eta_k$ and $\zeta = 0$ are soliton structures and rational solutions (which are a limiting case of solitons) respectively. A more complicated choice of $F(\zeta)$ (see appendix; in (3.2), $F(\zeta) = \zeta$) in the constraining equation (3.2) has the effect of adding more singular points. The interpretation of the corresponding solution $q(x)$ in these cases is still an open question.

7. A BÄCKLUND TRANSFORMATION

We now consider Ueno's monodromy problem [14], and our generalization, from the point of view of Bäcklund transformations. (BT) It is easy to see that in Manin's language [17], the addition of a soliton by a BT amounts to the identification of an additional pair of points on the Riemann sphere. That is, in the pure-soliton case, a BT creates an apparent singularity in the ζ-behavior of the Jost functions

of scattering theory. We will see that the same situation obtains in the Painlevé setting.

Consider the usual scattering problem for MKdV,

$$\begin{aligned} V_{1_x} + i\zeta V_1 &= qV_2 \\ V_{2_x} - i\zeta V_2 &= qV_1. \end{aligned} \tag{7.1}$$

Our considerations are local in x, and boundary conditions play no role. Fix a particular $\zeta=i\eta$, and a solution $V(x,i\eta)$ of (7.1). Define

$$\Gamma = \frac{V_1(x,i\eta)}{V_2(x,i\eta)}.$$

The x-part of the BT transforms q(x) into the new potential Q,

$$Q = q - 2\frac{d}{dx}\tanh\Gamma = q - \frac{2\Gamma x}{1-\Gamma^2} \tag{7.2}$$

The eigenfunctions $V(x,\zeta)$ of (1) yield eigenfunctions $W(x,\zeta)$ of

$$\begin{aligned} W_{1_x} + i\zeta W_1 &= QW_2 \\ W_{2_x} - i\zeta W_2 &= QW_1 \end{aligned} \tag{7.3}$$

by the formula

$$\begin{aligned} W_1 &= \left(i\zeta + \eta + \frac{2\eta}{\Gamma^2-1}\right) V_1 + \frac{2\eta\Gamma}{1-\Gamma^2} V_2 \\ W_2 &= \frac{2\eta\Gamma}{1-\Gamma^2} V_1 + \left(-i\zeta + \eta + \frac{2\eta}{\Gamma^2-1}\right) V_2. \end{aligned} \tag{7.4}$$

The verification is straightforward.

If we know the ζ-behavior of solutions of (7.1), we can read off the ζ-behavior of solutions of (7.3) from formulas (7.4). The effect of the transformation (7.4) is easy to ascertain. No poles or branch points are added, and the Stokes multipliers of V remain unchanged. The formal monodromy at $\zeta=\infty$ is increased by 1. The only other change is the addition of apparent singular points at $\zeta=\pm i\eta$. This is clear because the determinant of the transformation (7.4) vanishes at $\zeta=\pm i\eta$; two solutions of (7.1) which are independent at all ζ will be transformed into two solutions of (7.3) which are independent at all ζ except $\pm i\eta$. This is the symptom of an apparent singular point.

For example, start with q=0, and let $V = \begin{pmatrix} c_1 e^{\eta x} \\ c_2 e^{-\eta x} \end{pmatrix}$. Then

$$\Gamma = \frac{c_1}{c_2} e^{2\eta x} = e^{2\eta(x-\bar{x})} = e^{2\eta\tilde{x}},$$

and

$$Q = \eta \operatorname{cosech} 2\eta\bar{x}$$

is the (singular) one-soliton potential.

The eigenfunctions of (7.3), as computed from the independent solutions $e^{-i\zeta x}\begin{pmatrix}-i\\0\end{pmatrix}$, $e^{i\zeta x}\begin{pmatrix}0\\i\end{pmatrix}$ of (7.1) have:

(i) zero Stokes multipliers;
(ii) formal monodromy at $\infty, \ell=1$;
(iii) apparent singular points (with exponents 0,1) at $\zeta=\pm i\eta$.

One can check this on the explicit formulas:

$$(1) = \begin{bmatrix} \zeta - i\eta + \dfrac{2i\eta}{1-e^{4\eta\bar{x}}} \\ -\dfrac{2i\eta e^{2\eta\bar{x}}}{1-e^{4\eta\bar{x}}} \end{bmatrix} e^{-i\zeta x},$$

$$(2) = \begin{bmatrix} \dfrac{2i\eta e^{2\eta\bar{x}}}{1-e^{4\eta\bar{x}}} \\ \zeta + i\eta - \dfrac{2i\eta}{1-e^{4\eta\bar{x}}} \end{bmatrix} e^{i\zeta x}.$$

This is exactly the kind of normalized solution used by Ueno [14].

To complete the monodromy information, one must set up the solutions normalized at $\zeta=\pm i\eta$, and compute the connection matrices from $\pm i\eta$ to ∞. Following [14], we require the solutions near $\zeta=\pm i\eta$ to look like

$$\psi^{(\pm)}(x,\zeta) = T^{(\pm)}(x)\begin{pmatrix}\zeta\mp i\eta & 0\\ 0 & 1\end{pmatrix} \Phi^{(\pm)}(x,\zeta) = T^{(\pm)}(x)Y^{(\pm)}(x,\zeta) \qquad (7.5)$$

where

$$\Phi^{(\pm)}(x,\pm i\eta) = I$$

and $Y^{(\pm)}$ is the solution of a diagonalized ζ-equation,

$$Y^{(\pm)}_{\zeta} = \left\{ \frac{\begin{pmatrix}1&0\\0&0\end{pmatrix}}{\zeta\mp i\eta} + \text{holomorphic} \right\} Y^{(\pm)}.$$

Near $\zeta=i\eta$, for example, one finds $\Psi^{(+)} = \Psi A^{(+)}$ ($\Psi = (\psi^{(1)},\psi^{(2)})$), where the connection matrix $A^{(+)}$ has the form

$$A^{(+)} = \begin{pmatrix} C_1 & a \\ -C_2 & b \end{pmatrix}.$$

C_1 C_2 are as in the definition of Γ, $C_1/C_2 = e^{-2\eta x_0}$. a,b are functions of x, in the ratio $a/b = -ie^{-4\eta x}e^{2\eta x_0}$. It is important to note that the whole connection matrix $A^{(+)}$ is <u>not</u> required to be independent of x; the correct condition [14] is

$$JA_x^{(+)} A^{(+)^{-1}} J = 0, \tag{7.6}$$

with $J = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. The deformation condition (7.6) says that $\left(\frac{C_1}{C_2}\right)_x = 0$. Note that the ratio C_1/C_2 determines the single free parameter in the one-soliton potential. It is easy to see that it is precisely the ratio C_1/C_2 which is fixed by the normalization (7.5); there is some freedom in the determination of T, in that each column may be multiplied by a constant without affecting the other conditions.

It is now evident that a similar procedure will add apparent singular points at $\pm i\eta$, with $\pm i\eta$-to-∞ connection matrices

$$\begin{pmatrix} C_1 & a_{\pm} \\ -C_2 & b_{\pm} \end{pmatrix}$$

to the eigenfunctions of any potential q. In particular, we may take for q a solution of Painlevé II,

$$q_{xx} = 2q^3 + xq - \nu.$$

All of the monodromy data listed in [1] are unchanged after (4), with the exception of the monodromy at ∞ (which was zero, and is now 1), and of course the $\pm i\eta$ data are added.

The condition that the monodromy be independent of x then leads to the nonlinear system (6.1).

One computable example is afforded by the rational Painlevé solution $q(x) = \frac{1}{x}$. With

$$V(x,\eta) = a \begin{bmatrix} -\frac{1}{2\eta x} \\ 1 + \frac{1}{2\eta x} \end{bmatrix} e^{-\eta x} + b \begin{bmatrix} 1 - \frac{1}{2\eta x} \\ \frac{1}{2\eta x} \end{bmatrix} e^{\eta x},$$

we find

$$Q = \eta\left[\frac{a^2 - b^2e^{4\eta x} - 4ab\eta^2 xe^{2\eta x}}{(a+be^{2\eta x})^2 + \eta x(a^2-b^2e^{4\eta x})}\right]$$

Note that Q has the limiting forms

$$Q = \frac{1}{x - \frac{1}{\eta}}, \quad a = 0$$

$$\frac{1}{x + \frac{1}{\eta}}, \quad b = 0$$

which reconfirms our earlier computation.

This sort of explicit result is reassuring, but the main points of interest are more theoretical.

(1) (6.1) is the 5^{th} -order equation. Our example covers only a 3-parameter family of solutions. The place of the other solutions in the isospectral framework is still unknown.
(2) The equations (6.1) must be degenerate forms of deformation equations of a monodromy problem set on an elliptic (or hyperelliptic; the MKdV traveling wave is associated with a genus 2 Riemann surface). Riemann surface, with irregular singularity at the point at infinity. Those equations have yet to be analyzed.

8. CONCLUSION

The main purpose of this note is to draw attention to the mathematical problems connected with the notion of "multiphase similarity solution," and to the interesting physical applications which are still to be worked out. The theory is in an embryonic stage; indeed, not even the correct general definition of this class of solutions has been found. This is evident from the lack of information about deformation equations for systems with coefficients rational on a Riemann surface of genus greater than 1. Another point of interest is the connection of our work with that of Fokas [18] who obtains one class of multiphase similarity solutions contained in our formalism by considering Lie-Backlund symmetries.

A paper in preparation will address some of these questions, primarily for the MKdV family and for the Ising system of Section 5, but the general theory remains open.

APPENDIX:

Here we quote previous results [7] which are used throughout this paper. The most general equation with x- dependent coefficients which can be solved using the N^{th} order system

$$V_x = (\zeta R_o + P(x,t))V, \tag{A-1}$$

$R_o = i(\beta_j\delta_{jk}), P = (P_{jk}), P_{jj} = 0$, V an N-vector, is

$$G(D_R,t)P_t + F(D_R,t) \times [R_o,P] = \Omega(D_R,t)[C,P]. \qquad (A\text{-}2)$$

In (A-2), G, Ω and F are entire functions of an integro- differential operator D_R, C is a diagonal matrix and the bracket denotes the commutator. The operator D_R, which is the linear limit is $\partial/\partial x$ and which operates on off-diagonal matrices $H_F(h_{\ell m})$, $h_{\ell\ell} = 0$, is defined as follows. First, introduce the matrix $(H_F)_R = \left(\frac{h_{\ell m}}{i(\beta_\ell - \beta_m)}\right)$, the solution of $[R_o,(H_F)_R] = H_F$. The diagonal counterpart of H_F is defined to be $H_D = \int_x^\infty [H_F,P]_D dy$. Second, define the operator D acting on H_F to be

$$DH_F = \frac{\partial}{\partial x} H_F + [H_F,P]_F + \left[\int_x^\infty [H_F,P]_D dy, P\right]_F$$

$$= \frac{\partial}{\partial x} H + [H,P], \quad H = H_F + H_D. \qquad (A\text{-}3)$$

The subscripts F and D denote the off-diagonal and diagonal parts of the designated matrices respectively. Finally, $D_R H_F = D(H_F)_R$. The corresponding x and t dependence of $V(x,t,\zeta)$ is given by

$$G(\zeta,t)V_t + F(\zeta,t)V = (xR_o F(\zeta,t) + Q + S - T)V. \qquad (A\text{-}4)$$

If the potential $P(x)$ decays at $x = \pm\infty$, we can define fundamental matrix solutions $\Phi(x,t,\zeta)$, $\Psi(x,t,\zeta)$ with the asymptotic behavior $(e^{i\beta_j \zeta x}\delta_{jk})$ as $x \to -\infty$ respectively. Then, the scattering matrix, $A(\zeta,t) = \Phi^{-1}\Psi$ satisfies the equation

$$G(\zeta,t)A_t + F(\zeta,t)A_\zeta = W_+A - AW_- \qquad (A\text{-}5)$$

where $W_\pm = \underset{x\to\pm\infty}{Lt}\ (Q+S-T)$. The quantities Q,S and T are defined as follows. Let $\Omega(D_R,t) = \sum_0 \omega_m(t) D_R^m$, then $Q = \sum_0 \omega_m Q^{(m)}$ where $Q^{(m)} = \sum_0^m Q_r \zeta^{m-r}$, $Q_o = C$ and

$$Q_{SF} = (D_R^{S-1}[C,P])_R, \quad S = 1,\ldots m$$

$$Q_{SD} = \int_x^\infty [Q_{SF},P]_D dy, \quad Q_S = Q_{SF} + Q_{SD}. \qquad (A\text{-}6)$$

Define the sequences $\{T_k\}_{k=1}$, $\{S_\ell\}_{\ell=1}$ in a similar way:

$$T_{kF} = (D_R^{k-1}P_t)_R, T_{kD} = \int_x^\infty [T_{kF},P]_D dy, \quad T_k = T_{kF} + T_{kD}, \quad k \geq 1$$

$$S_{\ell F} = (D_R^{\ell-1}x[R_o,P])_R, \quad S_{\ell D} = \int_x^\infty [S_{\ell F},P]_D dy, \quad S_\ell = S_{\ell F} + S_{\ell D}, \quad \ell \geq 1. \qquad (A\text{-}7)$$

Then, if $G(D_R,t) = \sum_0 q_k(t) D_R^k$, $F(D_R,t) = \sum_0 f_\ell(t) D_R^\ell$,

$$T = \sum_1 g_k(t) T^{(k)},\quad S = \sum_1 f_\ell(t)\, S^{(\ell)},$$

$$T^{(k)} = \sum_1^k T_r \zeta^{k-r},\quad S^{(\ell)} = \sum_1^\ell S_r \zeta^{\ell-r}. \tag{A-8}$$

The flows (2.3) can be written in the form $P_{t_j} = D_R^j[C,P]$. Clearly $P_{t_j t_i} = P_{t_i t_j}$ and using this in (2.3) we find $\frac{\partial Q}{\partial x} = [P,Q]$, $Q = Q^{(i)}_{t_j} - Q^{(j)}_{t_i} + [Q^{(i)},Q^{(j)}]$. Therefore $Q = GQ_oG^{-1}$ where $G_x = PG$, i.e. G is the fundamental matrix solution to (2.1) at $\zeta=0$, Thus if Q is zero for any x, as it is if $q \to o$ at either ∞, then Q is zero for all x. However the result holds more generally. Q is a function of ζ and the equation $\frac{\partial Q}{\partial x} = [P,Q]$ holds for each ζ. But this is a system of homogeneous linear o.d.e.'s of order N^2 and thus has only N^2 degrees of freedom. Thus Q is identically zero and the flows commute.

The example used in Section 2,3,4, and 6 are derived from (A-2) as follows. Define $Y_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $Y_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $Y_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, and take $R_o = C = -iY_1$, $P = q$ $P = q(Y_2 + Y_3)$. Then $Q_1 = q(Y_2 + Y_3)$, $Q_2 = \frac{iq_x}{2}(Y_2 - Y_3)\ \frac{iq^2}{2} Y_1$,

$Q_3 = -1/4(q_{xx} - 2q^3)(Y_2 + Y_3)$, $Q_4 = -i/8(q_{xx} - 2q^3)_x(Y_2 - Y_3)$

$-1/4(qq_{xx} - 1/2q_x^2 - 3/2q^4)Y_1$, $Q_5 = [1/16(q_{xx} - 2q^3)_{xxx}$

$-1/4(q^2q_{xx} - 1/2qq_x^2 - 3/2q^5)_x](Y_2 + Y_3)$. Recall $Q^{(n)} =$

$\sum_0^n Q_r \zeta^{n-r}$, $Q_0 = -iY_1$. $S_1 = xq(Y_2 + Y_3)$, $S_2 = i/2(xq)_x(Y_2 - Y_3)$

$+ iA(x)Y_1$, $A_x = -(xq)_xq$, $A(\infty) = 0$, $S_3 = (-1/4(xq)_{xx} - Aq)\ (Y_2 + Y_3)$

and $S^{(n)} = \sum_1^n S_r \zeta^{n-r}$. Equation (2.3), is found by setting $\Omega = D_R$; $P_{t_1} = DQ_1 = Q_{1x}$ implies $q_{t_1} = q_x$: equation $(2.3)_3$ is obtained by taking $\Omega = 4D_R^3$; $P_{t_3} = 4DQ_3 = 4Q_{3x}$ implies that $q_{t_3} = -q_{xxx} + 6q^2q_x$: equation $(2.3)_5$ is obtained by taking $\Omega = 16D_R^5$; $P_{t_5} = 16DQ_5 = 16Q_{5x}$ implies that $q_{t_5} = (q_{xxxx} - 10q^2q_{xx} - 10qq_x^2 + 6q^5)_x$: equation $(2.3)_{-1}$ is obtained by taking $\Omega = D_R^{-1}$. One can show $D_R^{-1}[C,P] = 2\sinh u\ (Y_2 + Y_3)$

where $q = -u_{x/2}$ and $u(\infty) = 0$. Hence $q_{t_{-1}} = 2 \sinh u$ or $u_{xt_{-1}} = -4 \sinh u$. Also the constraint (3.2) is found by taking P and V independent of t (but not of $t_1 \cdots t_n$) and $F = D_R$, $\Omega = \Sigma\, j t_j D_R^j$ for odd j.

The equation (6.1) which describes the interaction of a soliton and a similarity solution comes about by choosing $F = D_R^3 + \eta^2 D_R$, $\Omega = D_R^5 + \eta^2 D_R^3 + \beta D_R$. The corresponding equation for V_ζ is

$$\zeta(\zeta^2+\eta^2)V_\zeta = (-ix\zeta(\zeta^2+\eta^2)Y_1 + Q + S - \lambda)V \tag{A-9}$$

where $Q = Q^{(5)} + \eta^2 Q^{(3)} + \beta Q^{(1)}$ and $S = S^{(3)} + \eta^2 S^{(1)}$. The equation (6.1) which is the compatibility condition of (A-9) and (A-1) is, when once integrated,

$$Q_5 + \eta^2 Q_3 + \beta Q_1 + S_3 + \eta^2 S_1 = K. \tag{A-10}$$

The scalar λ is chosen to normalize the solution $V(\zeta,x)$. If we were to choose $\lambda=0$, then, if $q = 2i\eta \operatorname{sech} 2\eta(x-x_0)$, the two linearly independent solutions of (A-9) near $\zeta = \pm i\eta$ have the form $(\zeta \mp i\eta)^{-1/2} F_1(\zeta)$ and $(\zeta \mp i\eta)^{1/2} F_2(\zeta)$ where $F_1(\zeta)$ and $F_2(\zeta)$ are analytic at $\zeta = \pm i\eta$. Even though the indicial roots differ by an integer, these are no logarithm terms. It is convenient to make $\lambda = \frac{1}{2}\left(\frac{1}{\zeta - i\eta} + \frac{1}{\zeta + i\eta}\right)\zeta(\zeta^2+\eta^2)$ as this makes the indicial roots -1 and 0 which means that $V(\zeta)$ has only a pole singularity at $\zeta = \pm i\eta$. Choosing the opposite sign for λ would give the indicial roots 0,1. A singularity which can be removed by normalizing the solution $V(\zeta)$ is called apparent.

ACKNOWLEDGMENTS

The authors are grateful for support by NSF grants MCS75-07548 A01, MPS75-07530, MCS79-03533, ONR grant N00014-76-C-0867, and U.S. Army Contract DAAG29-78-G-0059.

H. Flaschka thanks Prof. M. Satō, Drs. T. Miwa and M. Jimbo, and Mr. K Ueno for many stimulating discussions; these were made possible by a U.S.-Japan Cooperative Science Program grant, and by the kind hospitality of the Kyoto University RIMS in January 1978 and April 1979.

REFERENCES:

[1]. Flaschka, H. and Newell, A.C., Monodromy and Spectrum Preserving Deformations, I. Preprint. To appear Comm. Math. Phys. (1980).

[2]. Sato, M., Miwa, T., Jimbo, M., Publ. RIMS, Kyoto Univ., Holonomic Quantum Fields I, Vol. 14, pp. 223-267 (1978), Holonomic Quantum Fields II, Vol. 15, pp. 201-278 (1978), Holonomic Quantum Fields III, RIMS Preprint 260, (1978), Holonomic Quantum Fields IV, RIMS Preprint 263, (1978), Holonomic Quantum Fields V, RIMS Preprint 267, (1978).

[3]. WU, T.T., McCoy, B.M., Tracy, C.A., Barouch, E., Phys. Rev. B 13 Spin-Spin Correlation Functions for the Two-Dimensional Ising Model: Exact Theory in the Scaling Region, pp. 316-371, (1976).

[4]. Garnier, R., Ann. Ec. Norm. Sup. 29, pp. 1-126, Differential Equations of Third Order.

[5]. Garnier, R., Rend. Circ. Mat. Palermo 43 pp. 155-191 (1919), Sur une classe de Systemes Differentiels Abeliens deduits de la Theorie des Equations Lineaires.

[6]. Flaschka, H. and Newell, A.C., Mondromy and Spectrum Preserving Deformations, II, in preparation.

[7]. Newell, A.C., Proc. Roy. Soc. London A 365 pp. 283-311, The general structure of integrable evolution equations, (1979).

[8]. Flaschka, H. and Newell, A.C., Springer Lecture Notes in Physics 38 pp. 355-440 (1975).

[9]. Ablowitz, M.J. and Satsuma, J., J. Math. Phys. 19, (10) October (1978) p. 2180, Solitons and Rational Solutions of Nonlinear Evolution Equations.

[10]. Wadati, M., J. Phys. Soc. Japan 34 pp. 1289-1296 (1973) The modified Korteweg-deVries equation.

[11]. Lax, P.D., Lectures in Applied Math. 15, pp. 85-96 (1974), Periodic Solution of the Korteweg-deVries equations.

[12]. Novikov, S.P., Funkts. Anal. Prilozen. 8 pp. 54-66 (1974), The periodic problem for the Korteweg-deVries equation.

[13]. Krichever, I.M., Funkts. Anal. Prilozen., 11 pp. 15-31 (1977), Integration of nonlinear equations by methods of algebraic geometry.

[14]. Ueno, K., Kyoto University RIMS Master's Thesis, December, 1978, (written in Japanese), RIMS preprints (1979) 301-Monodromy Preserving Deformations of Linear Differential Equations with Irregular Singular Points, 302-Monodromy Preserving Deformations and its Application to Soliton Theory.

[15]. Hille, E., Ordinary Differential Equations in the Complex Domain, Pure & Applied Math Series, (1976) Wiley-Interscience NY.

[16]. Dubrovin, B.A., Funkts. Anal. Prilozen. 11 pp. 28-41 (1977) Completely Integrable Hamiltonian Systems Connected with Matrix Operators and Abelian Varieties.

[17]. Manin, Yu I., J. Soviet Math. Vol. 11, No1 (1979), Algebraic Aspects of Nonlinear Differential Equations.

[18]. Fokas, T., Ph.D. Thesis, Cal Tech, (1979), Invariants, Lie-Bäcklund Operators and Bäcklund Transformations.

2·1

J. Fluid Mech. (1980), *vol.* 98, *part* 4, *pp.* 803–818
Printed in Great Britain

Shelves and the Korteweg–de Vries equation

By C. J. KNICKERBOCKER AND ALAN C. NEWELL

Department of Mathematics and Computer Science, Clarkson College,
Potsdam, NY 13676

(Received 14 November 1978 and in revised form 1 October 1979)

An extension of the analytical results of Kaup & Newell (1978) concerning the effect of a perturbation on a solitary wave of the Korteweg–de Vries equation is given and numerical studies are conducted to verify the conclusions. In all cases, the numerical results agree with the results predicted by the theory. The most striking feature of the perturbed flow is the presence of a shelf in the lee of the solitary wave whose role is to absorb (provide) the extra mass which is created (depleted) by the perturbation.

1. Introduction and discussion

The problem of the propagation of a shallow-water solitary wave in a canal of slowly varying depth has been the subject of several papers in the recent literature. Although the changing depth causes reflexions (Peregrine 1967; Miles 1979); to a good approximation the unidirectional propagation is well described by the perturbed Korteweg–de Vries equation (PKdV) (Johnson 1973*a*; Kakutani 1971),

$$q_t + 6qq_x + q_{xxx} = -\Gamma(t)\,q, \quad 0 < \Gamma \ll 1. \tag{1.1}$$

In the context of water waves, the local depth is $hD(\epsilon^{\frac{1}{2}}X/h) + \epsilon hN(\epsilon^{\frac{1}{2}}X/h,\ \epsilon^{\frac{1}{2}}g^{\frac{1}{2}}h^{-\frac{1}{2}}T)$ with X and T the dimensional space and time co-ordinates respectively; the co-ordinate x is the local retarded time

$$x = \frac{1}{\epsilon}\int_0^{\epsilon^{\frac{1}{2}}X/h} \frac{ds}{D^4(s)} - \frac{(gh\epsilon)^{\frac{1}{2}}}{h}T; \quad t = \frac{1}{6}\int_0^{\epsilon^{\frac{1}{2}}X/h} D^4(s)\,ds$$

is a measure of the distance along the channel from the point where the depth begins to change. The right-going component of the disturbed elevation N is given by $\frac{1}{2}D^2q(x,t)$ and $\Gamma(t)$ is $9D_t/4D$ which is assumed small; namely D changes slowly with respect to the length scale of the disturbance.

It is natural to exploit the smallness of $\Gamma(t)$ and to write as a first approximation the solution to (1.1) in terms of the solution of the unperturbed problem allowing those quantities which are constants of the latter to vary slowly in time. Ott & Sudan (1970), assuming that the basic solution has the form of a soliton

$$q(x,t) = 2\eta^2 \operatorname{sech}^2 \eta(x - \bar{x}), \quad \bar{x}_t = 4\eta^2 \tag{1.2}$$

and using the conservation law (energy)

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} q^2\,dx = -2\Gamma(t)\int_{-\infty}^{\infty} q^2\,dx, \tag{1.3}$$

0022-1120/80/4463-1790 $02.00 © 1980 Cambridge University Press

found that $\eta(t)$ satisfies

$$\eta_t = -\tfrac{1}{8}\Gamma\eta, \tag{1.4}$$

which we shall see is a correct result. They did not explain, however, the fact that neither of the expressions for conservation of mass (here we speak of conservation of mass in the sense of the KdV equation; in the water wave context some of the actual mass is reflected),

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} q\,dx = -\Gamma(t)\int_{-\infty}^{\infty} q\,dx, \tag{1.5}$$

nor of first moment

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} xq\,dx = 3\int_{-\infty}^{\infty} q^2\,dx - \Gamma(t)\int_{-\infty}^{\infty} xq\,dx, \tag{1.6}$$

are satisfied to leading order. Each of the relations (1.3), (1.5), (1.6) may be integrated exactly (Leibovich & Randall 1971),

$$M(t) = \int_{-\infty}^{\infty} q\,dx = M(t_0)\exp\left(-\int_{t_0}^{t}\Gamma(s)\,ds\right), \tag{1.7a}$$

$$E(t) = \int_{-\infty}^{\infty} q^2\,dx = E(t_0)\exp\left(-2\int_{t_0}^{t}\Gamma(s)\,ds\right) \tag{1.7b}$$

and

$$G(t) = \int_{-\infty}^{\infty} xq\,dx = \left[G(t_0) + 3E(t_0)\int_{t_0}^{t} ds\exp\left(-\int_{t_0}^{s}\Gamma(r)\,dr\right)\right]\exp\left(-\int_{t_0}^{t}\Gamma(s)\,ds\right). \tag{1.7c}$$

Several authors (Grimshaw 1970, 1971; Johnson 1973*b*; Leibovich & Randall 1973) have examined the propagation of a solitary wave over an uneven bottom topography in some detail. Johnson (1973*b*) and Leibovich & Randall (1973) work with (1.1) directly and attempt to find an asymptotic representation of the solution in the form

$$q(x,t) = q_0(x,t) + \sigma q_1(x,t) + \ldots, \tag{1.8}$$

where σ, $0 < \epsilon \ll \sigma \ll 1$, is a measure of the amplitude of $\Gamma(t)$ and $q_0(x,t)$, the leading approximation, is given by (1.2). By demanding that the asymptotic series (1.8) remains a uniformly valid description of the solution $q(x,t)$ over long time (σ^{-1}), they found that $\eta(t)$ obeys (1.4). However, they were unable to find a solution $q_1(x,t)$ which tends to zero both as $x\to\infty$ and $x\to-\infty$. In fact, they found that, as $x\to-\infty$, $q_1\to\Gamma/3\eta\sigma$, which renders (1.8) non-uniform and seems to indicate that the mass

$$\int_{-\infty}^{\infty} q\,dx$$

is infinite. Clearly there is something non-uniform about the expansion (1.8), a point to which we will return. Nevertheless, as we shall see, the results are almost correct and indeed a shelf does form behind the solitary wave. However, it has a finite range, a fact observed numerically by Leibovich & Randall (1973) but not explained theoretically.

The dilemma of infinite mass and the role of the shelf was first explained by Kaup & Newell (1978) (hereinafter referred to as KN) who used a totally different method (also used by Karpman & Maslov (1977); we should also mention that some of these results were obtained by Ko & Kuehl (1978) by a direct method after they were familiar with the results of Kaup & Newell). They exploited the fact that the unper-

turbed equation (1.1) is exactly integrable. By exactly integrable, we mean the following. The KdV equation is an infinite-dimensional Hamiltonian system (Gardner 1971; Zakharov & Faddeev 1971; Flaschka & Newell 1975) which may be written

$$q_t = \frac{\partial}{\partial x}\frac{\delta H}{\delta q}, \quad H = \frac{1}{2}\int_{-\infty}^{\infty}(q_x^2 - 2q^3)\,dx \tag{1.9}$$

and $\delta/\delta q$ is the variational derivative. The inverse scattering transform (Gardner *et al.* 1967, 1974), is a canonical transformation (preserves the form of Hamilton's equations) which carries the old co-ordinates $q(x,t)$, $-\infty < x < \infty$, to new ones which are defined by the scattering data S,

$$S = \{(\gamma_k, \zeta_k = i\eta_k)|_{k=1}^{N}, R(\zeta), \zeta \text{ real}\} \tag{1.10a}$$

of the eigenvalue problem

$$\psi_{xx} + (\zeta^2 + q(x,t))\,\psi = 0, \quad -\infty < x < \infty. \tag{1.10b}$$

In S, the quantities η_k are the bound-state eigenvalues, the γ_k are the normalization constants for the corresponding eigenfunctions and $R(\zeta)$ is the reflexion coefficient. If $q(x,t)$ evolves according to the KdV equation (or any member of the KdV family), then the functions $-2\eta_k^2$ and $(-2\zeta/\pi)\ln(1-|R|^2)$ are action variables and therefore constants of the motion; their angle counterparts, which change linearly in time, are proportional to $\ln\gamma_k$ and $\arg R(\zeta)$ respectively, and the Hamiltonian H in the new co-ordinates is an additive function of the action variables. Each η_k gives rise to a soliton which when physically separated from the other solution components has the form $2\eta_k^2\,\mathrm{sech}^2\,\eta_k(x-\bar{x}_k)$ with $\bar{x}_{kt} = 4\eta_k^2$, where $\gamma_k = 2i\eta_k\exp[2\eta_k\bar{x}_k]$. Thus the action variable η prescribes the constant amplitude, shape and speed of the soliton; the angle variable γ or $\bar{x}$ defines its position. The function $R(\zeta)$, the reflexion coefficient, measures the degree to which the continuous spectrum is excited. For a pure soliton state or reflexionless potential $q(x,t)$, $R(\zeta) \equiv 0$. In general, however, $q(x,t)$ is expressed as a series in terms of the squared eigenfunctions of (1.10*b*), wherein contributions from both the continuous and discrete spectra are included,

$$q(x,t) = \frac{2}{i\pi}\int_{-\infty}^{\infty}\zeta R(\zeta)\,\psi^2(x,t,\zeta)\,d\zeta - 4\sum_{1}^{N}\gamma_k\zeta_k\psi_k^2(x,t,\zeta_k). \tag{1.11}$$

The solution component corresponding to the continuous spectrum gives rise to that portion of the solution which is oscillatory and dispersive in nature. We can therefore think of the KdV equation as being separated into its various normal modes (ψ^2, ψ_k^2) by the inverse scattering transform.

The effect of a perturbation is to render (1.1) no longer exactly separable. Instead the normal modes can become mixed; an initial state consisting only of solitons can stimulate radiation and in certain cases vice versa. If the perturbation term is small, then it is natural to treat the system by writing down the equations for the rates of change of the action variables and allowing the leading-order approximations of the latter to vary slowly so as to suppress any non-uniformities appearing in the perturbation expansions for these quantities. This is the method used by KN to analyse the effects of various typical perturbations on the canonical equations of inverse scattering theory and is a natural generalization of classical perturbation methods for finite-dimensional Hamiltonian systems.

The principal new result of the KN approach is the resolution of the infinite mass dilemma. What happens is that the continuous spectrum is excited by a resonance due to the interaction between the soliton and the perturbation. The quantity $\zeta R(\zeta)$ develops a Dirac delta-function behaviour at $\zeta = 0$. The corresponding structure in physical space is a shelf of almost constant height which stretches between the position ($x = 0$) of the soliton when the perturbation was just switched on and the solitary wave's present position. In the original co-ordinate frame, the shelf stretches from the soliton to the position to which an infinitesimal disturbance would have travelled from the point where the topography first changed. As the solitary wave moves, new shelf (if $\Gamma > 0$, it is a shelf of depression, if $\Gamma < 0$ of elevation) is continuously formed and the extra mass depleted or created is exactly the amount needed to satisfy (1.7*a*) and (1.7*c*). Although it was not originally noted by KN, the amplitude of the shelf continues to evolve after its initial formation due to the influence of the perturbation, a point observed by the present authors (Newell 1978) and independently by Miles (1979). The transition between the shelf and the $q = 0$ state at $x = 0$ is achieved through a series of decaying oscillations (the integral of an Airy function).

Before we give the analytical and numerical results in the next two sections, we make the following remarks. The first concerns the connexion between the KN approach and the straightforward method of perturbing (1.1) directly to obtain

$$q_{0t} + 6q_0 q_{0x} + q_{0xxx} = 0, \tag{1.12a}$$

$$q_{1t} + 6q_0 q_{1x} + 6q_1 q_{0x} + q_{1xxx} = -(1/\sigma)\,\Gamma(t)\,q_0, \tag{1.12b}$$

and so on. The question now arises as to how to solve (1.12*b*). If $q_0(x,t)$ is a solitary wave with phase $\theta = x - \int 4\eta^2\,dt$ and if one asks, as Johnson and Leibovich & Randall did, for solutions $q_1(x,t)$ which depend on the fast scale x and t only through the combination θ and a slow scale σx or σt, then (1.12*b*) is an ordinary differential equation whose solution has the property that $q_1 \to \Gamma/3\eta\sigma$ as x and $\theta \to -\infty$. In other words, if one simplifies (1.12*b*) by making the ansatz $q = q(\theta, \sigma x \text{ or } \sigma t)$, one looks at the problem from the frame of reference of the solitary wave, from which vantage point the shelf looks infinite, and one therefore loses information about the initial onset of the perturbation. Thus the ansatz $q = q(\theta, \sigma x \text{ or } \sigma t)$ fails because, although the shelf amplitude is slowly varying, its range is not. It also fails at the point where q_1 must make the reverse transition from $-\Gamma/3\eta\sigma$ to zero. Near that point, and indeed away from the solitary wave, q_0 is asymptotically zero and q_1 satisfies $q_{1t} + q_{1xxx} = 0$, which together with the local boundary conditions $q_1(x \to -\infty, t) = 0$, $q_1(x \to +\infty, t) = -\Gamma/3\eta\sigma$ is satisfied by the integral of an Airy function, in fact, exactly (2.5). However as we shall see in §§ 2 and 3, for long times, account must be taken of the further evolution of the shelf after its initial creation; this will mean in fact that the second boundary condition on $q_1(x,t)$ will be $q_1(x \to \infty, t) = -(\Gamma/3\eta(0)\,\sigma)\exp\left(-\int_0^t \Gamma(s)\,ds\right)$. One can calculate the position at which the reverse transition must occur by using the exact mass-balance equation (1.7*a*).

There is no doubt, then, that a direct perturbation can be successful, provided one understands *a priori* the nature of the solution. The important thing to check is whether the slow change of the soliton parameter (or parameters) can simultaneously satisfy all the conservation relations; in general, it will not. When it does not, the

solution will no longer be adiabatic; i.e. a slowly varying solitary wave. Nevertheless, as we have seen, using a judicious combination of the perturbation equations and the conservation laws one can obtain the solution by direct methods. Indeed this is the only approach available if the problem is not exactly integrable in the first approximation. However, when it is, we emphasize that all these results appear quite naturally when one takes advantage of the exact integrability of the leading-order equation and inverse scattering theory.

Furthermore, if $q_0(x,t)$ is not simply a soliton but a more complicated solution of (1.12*a*), then it is extremely difficult to find, by direct means, the appropriate basis for which the left-hand side of (1.12*b*), which is then a partial differential equation, separates. The answer is provided by inverse-scattering theory (Newell 1980), which tells us that the correct basis for expanding $q_1(x,t)$ is

$$E = \left\{\frac{\partial\phi^2}{\partial x}(x,\zeta), \zeta \text{ real}; \left(\frac{\partial\phi_k^2}{\partial x}, \frac{\partial^2\phi_k^2}{\partial x\,\partial\zeta}\right)_1^N\right\},$$

which is adjoint to the set $F = \{\psi^2(x,\zeta), \zeta \text{ real}; (\psi_k^2, \partial\psi_k^2/\partial\zeta)_1^N\}$, where $\phi(x,\zeta)$ ($\psi(x,\zeta)$) is the solution of (1.10*b*) which behaves as $e^{-i\zeta x}$ ($e^{i\zeta x}$) as $x \to -\infty$ ($+\infty$). Indeed multiplying (1.12*b*) by $\psi^2(x,\zeta,t)$, integrating over $(-\infty,\infty)$ in x, using the expression

$$\psi_t = (q_{0x} - 4i\zeta^2)\,\psi + (4\zeta^2 - 2q_0)\,\psi_x,$$

for the time dependence of ψ, gives us exactly the expressions we would have obtained by the KN approach. Thus in order to separate the perturbed system, we are led back to the same expressions for the perturbed action-angle variables as we would have obtained using the KN method.

Our second remark emphasizes the point that in the KN method no *a priori* ansatz is made concerning the solution structure. All we do is give the initial values of the action and angle variables (e.g. the amplitude and position of the soliton before the perturbation is switched on). Then, no matter how $q(x,t)$ evolves, the scattering data S, see (1.10*a*), is always uniquely defined, the eigenfunctions $\psi(x,t,\zeta)$ always computable in principle and $q(x,t)$ may be written down through (1.11). We stress that the structure of the kth soliton is given by $q = -4\gamma_k\zeta_k\psi_k^2$. The initial shape is

$$q = 2\eta_k^2 \operatorname{sech}^2\eta_k(x - \bar{x}_k)$$

but one is not guaranteed that the soliton shape is always given by the hyperbolic secant. In the present case, the long-time behaviour of the solitary wave structure can be written as $2\eta_k^2\operatorname{sech}^2\eta_k(x-\tilde{x}_k)$ except that now $\tilde{x}_k$ is a modification of $\bar{x}_k$, the original angle variable.

Finally we mention that the height of the reflected shelf has been calculated by Miles (1979).

2. Analytical results

For $t < 0$, a soliton (1.2) with $\eta = \eta_0$ travels unperturbed. At $t = 0$, the soliton arrives at $x = 0$ and the perturbation is switched on. Our goal is to monitor the subsequent evolution of $q(x,t)$ to leading order. It is stressed that even though the shelf has amplitude of order Γ, over long time it makes an order one contribution to both the

mass and first moment balance. By examining the change in the scattering data, KN found to leading order that

$$\eta_t = -\tfrac{2}{3}\Gamma\eta, \tag{2.1a}$$

$$\tilde{x}_t = 4\eta^2 + O(\Gamma), \tag{2.1b}$$

where $\tilde{x}$ is a measure of the position of the solitary wave (e.g. the position of its maximum).

As already mentioned, the interaction of the soliton with the perturbation give rise to a resonance which leads to a non-decaying component of the solution connected with the continuous spectrum. Indeed for order one times, KN found that this contribution (calculated by using (1.11)) is given by

$$q_c(x,t) = \frac{\Gamma}{6\pi\eta}\tanh^2\eta(x-\tilde{x})\int_{-\infty}^{\infty}\frac{\sin 2\zeta(x-\tilde{x}) - \sin 2\zeta(x+4\zeta^2 t)}{2\zeta}d(2\zeta) \tag{2.2}$$

$$= \frac{\Gamma}{6\pi\eta}\tanh^2\eta(x-\tilde{x})\left[\pi\,\mathrm{sgn}\,(x-\tilde{x}) - \frac{\pi}{3} - 2\pi\int_0^{x/(3t)^{\frac{1}{3}}}\mathrm{Ai}(s)\,ds\right], \tag{2.3}$$

where $\mathrm{sgn}\,(x) = 1$ for $x \geqslant 0$, $\mathrm{sgn}\,(x) = -1$ for $x < 0$. For $x > \tilde{x} > 0$, the term in the brackets is zero. For $0 < x < \tilde{x}$ (and for those x where $\tanh^2\eta(x-\tilde{x}) \simeq 1$),

$$q_c(x,t) \sim -\frac{\Gamma}{3\eta} + \frac{\Gamma}{6\pi^{\frac{1}{2}}\eta}\left(\frac{x}{(3t)^{\frac{1}{3}}}\right)^{-\frac{1}{4}}\exp\left(-\frac{2}{3}\left(\frac{x}{(3t)^{\frac{1}{3}}}\right)^{\frac{3}{2}}\right); \tag{2.4}$$

for $x < 0$,

$$q_c(x,t) \sim -\frac{\Gamma}{3\pi^{\frac{1}{2}}\eta}\left(\frac{-x}{(3t)^{\frac{1}{3}}}\right)^{-\frac{1}{4}}\cos\left(\frac{2}{3}\left(\frac{-x}{(3t)^{\frac{1}{3}}}\right)^{\frac{3}{2}} + \frac{\pi}{4}\right). \tag{2.5}$$

Thus between $x = 0$ and $x = \tilde{x}$, the position of the solitary wave, a shelf of height $-\Gamma/3\eta$ is created. At both $x = 0$ and $x = \tilde{x}$ the transitions to the respective solitary wave and zero states are smooth. KN pointed out how this shelf accounts for the *rate* of the extra mass created ($\Gamma < 0$; depth-decreasing case) per unit distance, mass which is not absorbed by the amplifying soliton:

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} q\,dx = \frac{\partial}{\partial t}\int_{-\infty}^{\infty} q_s\,dx + \frac{\partial}{\partial t}\int_0^{\tilde{x}} q_c\,dx = 4\eta_t + q_c(\tilde{x})\,\tilde{x}_t$$

$$= -\tfrac{2}{3}\Gamma(4\eta) - \tfrac{1}{3}\Gamma(4\eta) = -\Gamma(4\eta),$$

the exact result. However as pointed out by the authors (Newell 1978; this point has also been noted by Miles 1979) the calculation of shelf height for times $1/\Gamma$ is only valid immediately behind the solitary wave. The subsequent evolution of the shelf is most easily calculated from (1.1) directly; the nonlinear and dispersion terms are negligible. Then at $\bar{t}$ when the solitary wave is at $\tilde{x}(\bar{t}) = \int_0^{\bar{t}} 4\eta^2\,dt$, the height of the shelf $q_c(x,\bar{t})$ at the point x is

$$q(x,\bar{t}) = \frac{-\Gamma}{3\eta(\bar{t})}\exp\left(\int_{\bar{t}}^{t(x)}\Gamma(s)\,ds\right), \quad 0 < x < \tilde{x}, \tag{2.6}$$

$$= 0 \quad \text{otherwise},$$

where $t = t(x)$ through the integration of $x_t = 4\eta^2$. Our numerical results presented in §3 agree almost precisely with this formula.

Let us now see how these approximations allow us to balance the exact relations (1.7*a*) and (1.7*c*) almost precisely. To leading order, the mass in the shelf is given by

$$M_c(\tilde{t}) = \int_0^{\tilde{x}} q_c(x, \tilde{t})\, dx. \tag{2.7a}$$

By using (2.6) and the transformation $x = \int_0^t 4\eta^2\, dt$ on (2.7), we find

$$\begin{aligned} M_c(\tilde{t}) &= -4\eta_0 \exp\left(-\int_0^{\tilde{t}} \Gamma(s)\, ds\right)\left\{\int_0^{\tilde{t}} dt\, \frac{\Gamma(t)}{3} \exp\frac{1}{3}\int_0^t \Gamma(s)\, ds\right\} \\ &= 4\eta_0 \exp\left(-\int_0^{\tilde{t}} \Gamma(s)\, ds\right) - 4\eta_0 \exp\left(-\int_0^{\tilde{t}} \tfrac{2}{3}\Gamma(s)\, ds\right). \end{aligned} \tag{2.7b}$$

But the last term is simply $M_s(\tilde{t}) = \int_{-\infty}^{\infty} q_s(x, \tilde{t})\, dx$, the mass in the solitary wave, and thus to leading order

$$M(\tilde{t}) = M_c(\tilde{t}) + M_s(\tilde{t}) = 4\eta_0 \exp\left(-\int_0^{\tilde{t}} \Gamma(s)\, ds\right), \tag{2.8}$$

the exact result. A similar calculation can be carried out for $G(\tilde{t})$. When Γ is constant it can be done explicitly. We find, again to leading order, that

$$G(\tilde{t}) = (-16\eta_0^3/\Gamma) \exp(-\Gamma\tilde{t})\,[\exp(-\Gamma\tilde{t}) - 1], \tag{2.9}$$

which is the exact result (1.7*c*).

Finally we mention the consequences of these results in the context of shallow water waves. Integration of (2.1) gives

$$\eta/\eta_0 = (D_0/D)^{\frac{1}{2}}, \tag{2.10}$$

whence the soliton amplitude is

$$\tfrac{1}{2}\epsilon h \eta_0^2 D_0^2/D. \tag{2.11}$$

Thus the solitary wave amplitude is inversely proportional to the depth. On the other hand the shelf height (here it is convenient to express the height as a function of t, which in the water wave case measures position, and $\tilde{t}$ the present position of the solitary wave) is $\frac{1}{2}\epsilon h D^2(\tilde{t})\, q_c(x(t), \tilde{t})$, which, using (2.6) with $\Gamma = +9D_t/(4D)$, is

$$-3\epsilon h D' \frac{D^{\frac{1}{2}}(t)}{D_0^{\frac{1}{2}}} \frac{1}{D^{\frac{1}{2}}(\tilde{t})}, \tag{2.12}$$

which is inversely proportional to the fourth root of the depth estimated at the present solitary wave position. D' refers to the derivative of the depth with respect to the argument $\epsilon^{\frac{1}{2}}X/h$.

3. Numerical results

In order to verify the theoretical predictions, we numerically simulated the differential equation (1.1). For a number of reasons we chose an explicit finite-difference scheme suggested by Vliegenthart (1971). This scheme discretizes the components of (1.1) by

$$\left.\begin{aligned} q_t &= \frac{q(j,n+1)-q(j,n-1)}{2\Delta t}+O((\Delta t)^2),\\ q &= \frac{q(j+1,n)+q(j,n)+q(j-1,n)}{3}+O((\Delta x)^2),\\ q_x &= \frac{q(j+1,n)-q(j-1,n)}{2\Delta x}+O((\Delta x)^2),\\ q_{xxx} &= \frac{q(j+2,n)-2q(j+1,n)+2q(j-1,n)-q(j-2,n)}{2(\Delta x)^3}+O((\Delta x)^2), \end{aligned}\right\} \tag{3.1}$$

where $t = n\Delta t$ and $x = j\Delta x$.

We define $\Gamma(t) = -f_t(t)/f(t)$ and we will present the results of two numerical cases. Case (i) consists of defining $f(t) = e^{\sigma t}$ so that $\Gamma(t) = -\sigma =$ constant. For this case the computations used $\sigma = -\frac{1}{40}$ and considered times $0 < t < 100$. For case (ii) we took $f(t) = \sigma t - 1$, yielding $\Gamma(t) = -\sigma/(\sigma t - 1)$. In these computations we chose $\sigma = \frac{1}{40}$ and looked at times $0 < t < 40$. The choice of case (ii) is of interest because of its potential applicability as a model to the problem of an internal solitary wave travelling on a thermocline in the neighbourhood of the point where the coefficient of the nonlinear terms vanishes (approximately where the depths of the upper and lower layers are equal).

The time at which the perturbation procedure breaks down is when $\Gamma/3\eta$, the shelf height, is of the same order as the amplitude $2\eta^2$. For case (i) this occurs when t is $O(\ln|\sigma|/\sigma)$, which for $\sigma = -\frac{1}{40}$ is approximately 100 time units. We monitored times up to 100 time units for case (i) and did not observe any divergence between the numerical and perturbation results. For case (ii), the breakdown occurs when time is $\sigma^{-1} - O(\sigma^{-2})$, which for $\sigma = \frac{1}{40}$ is approximately 33 time units. After this time we noticed that the perturbation solution began to diverge from the numerical solution.

Figure 1 gives an overall picture of the total motion of the system. The major features of the solution are: (*a*) a slightly distorted solitary wave, (*b*) the formation of the shelf, its finite range and its subsequent evolution and (*c*) the decaying oscillatory tail. Figure 1 shows the numerical solution for all x at five different times t.

In order to check on the accuracy of the numerical results, we continuously monitored the values of the total mass, energy and the centre of gravity and compared the numerical results with the exact relations (1.7). First, with $\Gamma(t) = -f_t(t)/f(t)$ the mass is given by the relation

$$\int q\,dx = 4\eta_0|f|. \tag{3.2}$$

Second, the energy is given by

$$\int q^2\,dx = \tfrac{16}{3}\eta_0^3 f^2, \tag{3.3}$$

and, third, the centre of gravity is defined by

$$\bar{y} = \frac{\int xq\,dx}{\int q\,dx} = \frac{16\eta_0^3\{\int|f|\,dt\}}{4\eta_0|f|}. \tag{3.4}$$

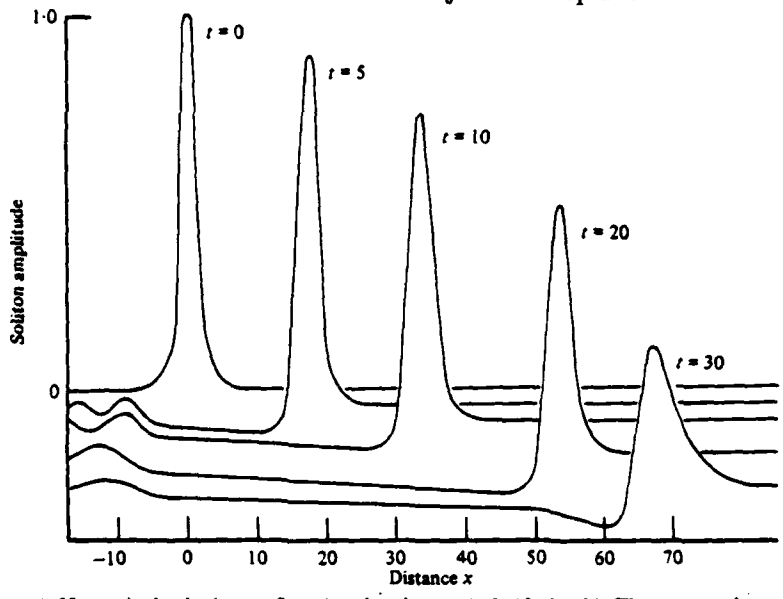


FIGURE 1. Numerical solution at five time levels $t = 0, 5, 10, 20, 30$. The curves shown graph the distance $x\ (-10 < x < 70)$ versus the scaled amplitude $(q(x, t)/2\eta_0^2)^{\frac{1}{2}}$.

	Case (i)	Case (ii)
Mass	0·0003 %	0·48 %
Energy	1·9 %	2·28 %
Centre of gravity	0·7 %	1·07 %

TABLE 1. The maximum error obtained in comparing the numerical results to the exact results, over the intervals (i) $0 < t < 100$ and (ii) $0 < t < 32$.

As the table 1 indicates, we obtained very close agreement with the exact results, which is to be anticipated since the Vliegenthart scheme is designed to conserve both mass and energy in the unperturbed equation.

Having established the accuracy of the numerical results, we now compare them with the results of the perturbation theory.

In analysing the solitary wave portion of the perturbation solution, two features were checked numerically, the amplitude and the position. The amplitude evolution as derived from the perturbation theory is given by

$$2\eta^2 = 2\eta_0^2\,|f|^{\frac{1}{2}}. \tag{3.5}$$

Figure 2 gives a graphical representation of the comparison of the numerical results and (3.5) for case (i). The comparison shown in figure 2 yields a maximum error of 3 %.

The second feature of the solitary wave examined is its position, which, to leading order, is given by

$$\begin{aligned} x(\tilde{t}) &= \int_0^{\tilde{t}} 4\eta^2\,dt + O(\Gamma) \\ &= 4\eta_0^2 \int_0^{\tilde{t}} |f|^{\frac{1}{2}}\,dt + O(\Gamma). \end{aligned} \tag{3.6}$$

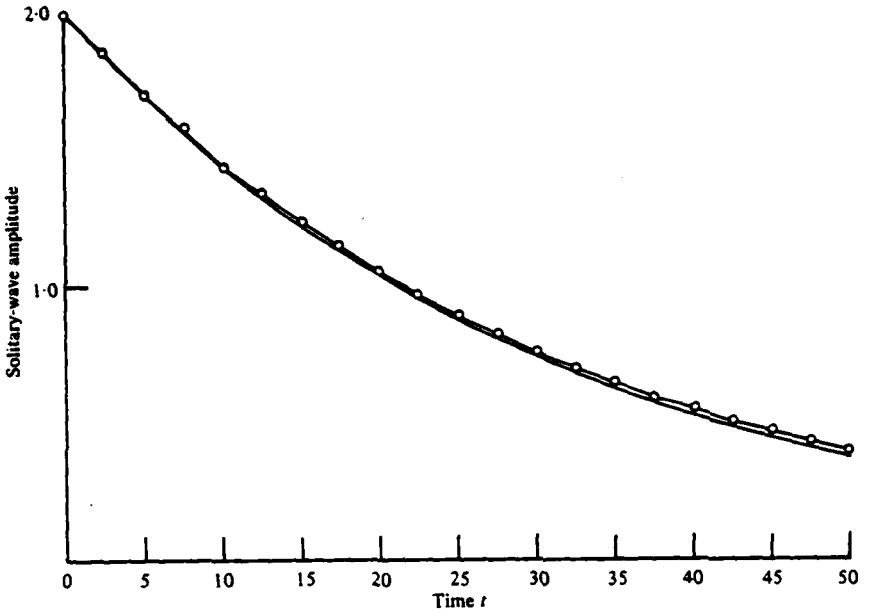


FIGURE 2. Numerical solution (—o—) versus perturbation theory (——) for the amplitude variation of the solitary wave. The results graph the amplitude ($0 < q(x, t) < 2$) versus time ($0 < t < 50$) for case (i).

The comparison between the integration of (3.6) and the numerical results for case (i) is given in figure 3.

Before we proceed there are two points to be made concerning the results shown in figure 3. First, from the numerical experiment, one observes that the solitary wave slows down and stops. Remarkably the integration of (3.6),

$$x(\bar{t}) = (+3\eta_0^2/\Gamma)\,(1-\exp(-\tfrac{4}{3}\Gamma\bar{t})),$$

follows the entire experimental trajectory of the solitary wave maximum to within 2 %, which is of order Γ. We say remarkably, for one might expect the approximate solution to be valid only for times when $t \ll \Gamma^{-1}\ln\Gamma^{-1}$ when $-\Gamma/3\eta$, the shelf height, is much less than $2\eta^2$, the soliton amplitude. Second, we note in figure 3 that the difference between theory and numerical experiment for the trajectory of the centre of gravity $\bar{y} = (4\eta_0^2/\Gamma)\,(1-\exp(-\Gamma\bar{t}))$ is less than 0·7 %. In view of the fact that the approximate theory follows both $\int_{-\infty}^{\infty} q\,dx$ and $\int_{-\infty}^{\infty} xq\,dx$ so closely (cf. (2.8) and (2.9)), it is not surprising that $\bar{y}(t)$ can be followed so closely.

The comparison between the integration of (3.6) and the numerical results for case (ii) is given in figure 4. The maximum error for case (ii) was 2·3 %.

The second major feature of the general solution which we checked was the shelf and again the numerical and perturbation results were extremely close.

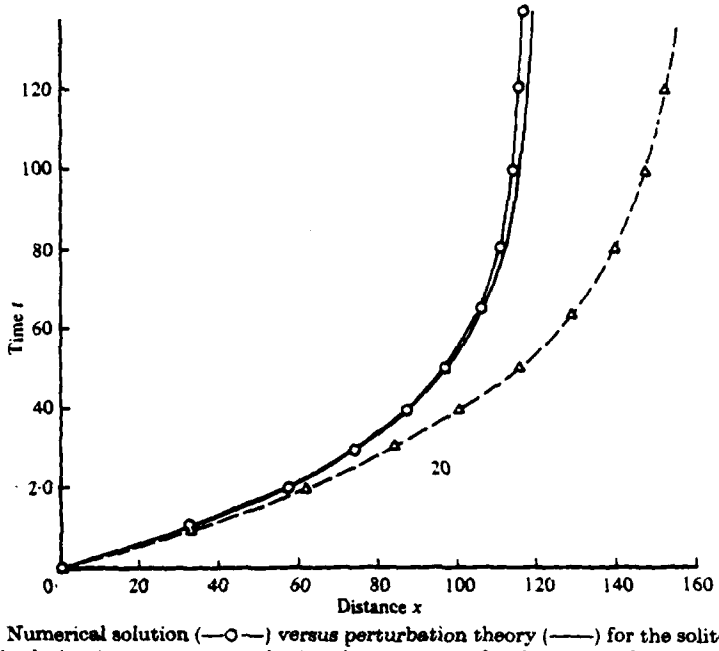


FIGURE 3. Numerical solution (—○—) versus perturbation theory (——) for the soliton position. Numerical solution (△) versus perturbation theory (– – – –) for the centre of gravity. The results graph distance ($0 < x < 160$) versus time ($0 < t < 140$) for case (i).

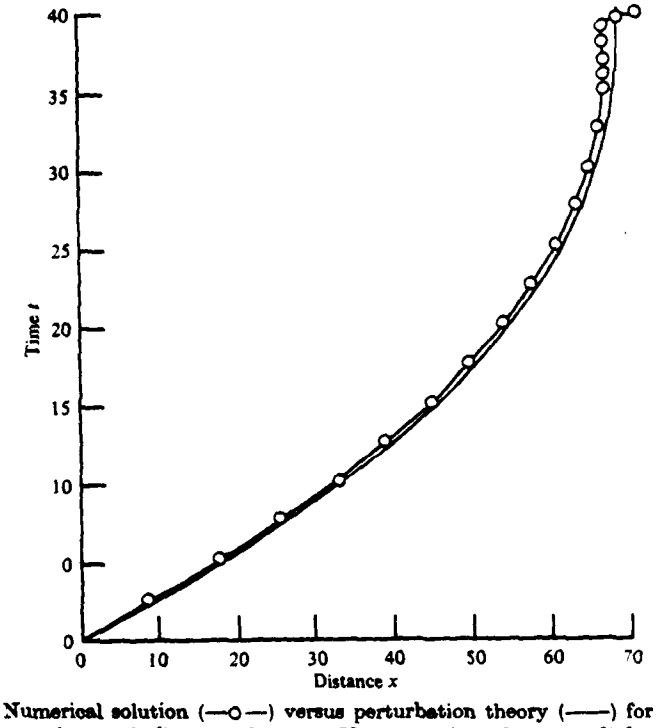


FIGURE 4. Numerical solution (—○—) versus perturbation theory (——) for soliton position. The results graph distance ($0 < x < 70$) versus time ($0 < t < 40$) for case (ii).

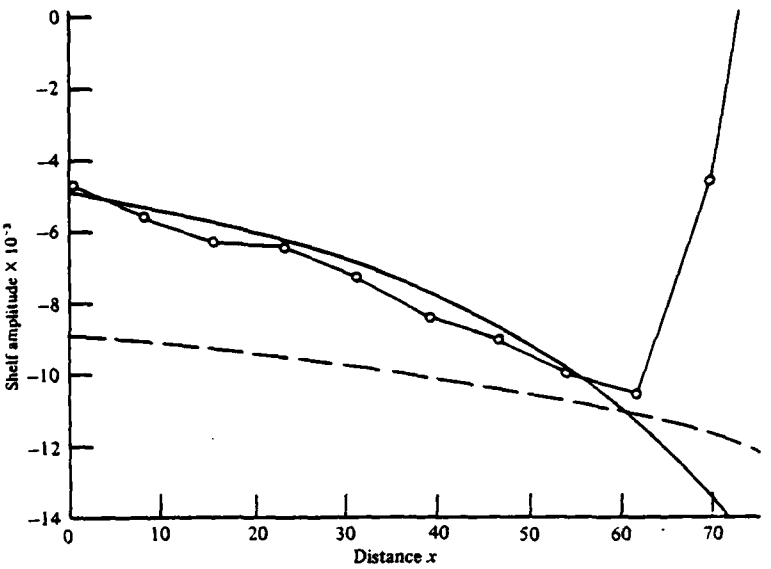


FIGURE 5. Numerical solution (—○—) versus perturbation theory (——) for shelf amplitude. The curves relate distance ($0 < x < 70$) and amplitude ($-14 \times 10^{-3} < q < 0$) for case (i) at $t = 27{\cdot}75$ time units. The dashed line (– – –) represents the maximum amplitude of the shelf for each point x. This maximum occurs at $t = t_c(x)$ the creation time of the shelf at the point x. The rear portion of the solitary wave may be seen at the right of the figure.

From (2.6) with $\Gamma(t) = -f_t(t)/f(t)$, we find

$$q_c(x, \bar{t}) = \frac{q_c(x, t_c(x))}{f(t_c(x))} f(\bar{t}), \tag{3.7}$$

where $\bar{t}$ measures the current time and $t_c(x)$ represents the time at which the shelf was created at each point x. By writing the creation time as a function of x through the integration of (3.6) we arrived at the following results. For case (i), (3.7) yields

$$q_c(x, \bar{t}) = \frac{\sigma e^{\sigma \bar{t}}}{3\eta_0(1 + \sigma x/3\eta_0^2)^{\frac{1}{2}}}, \tag{3.8}$$

where $\eta_0 = \eta(t = 0)$.

Similarly for case (ii), it is easy to verify that

$$q_c(x, \bar{t}) = \frac{\sigma(\sigma \bar{t} - 1)}{3\eta_0[1 - 7\sigma x/12\eta_0^2]^{\frac{1}{7}}}. \tag{3.9}$$

In figure 5 we look at an enlarged version of the solution for case (i) at $\bar{t} = 27$ time units, and we focus our attention on the shelf portion of that solution. The dashed line just below the shelf represents the maximum amplitude of the shelf at each point x. This maximum, which is $-\Gamma(t_c)/3\eta(t_c)$, occurs at the time of its creation. The numerical results agree almost precisely with the perturbation results given by (3.8) and (3.9), and the comparison is equally good at all other times.

In order to study further the formation and evolution of the shelf, we focused our attention on several positions and monitored the evolution of the shelf at those points.

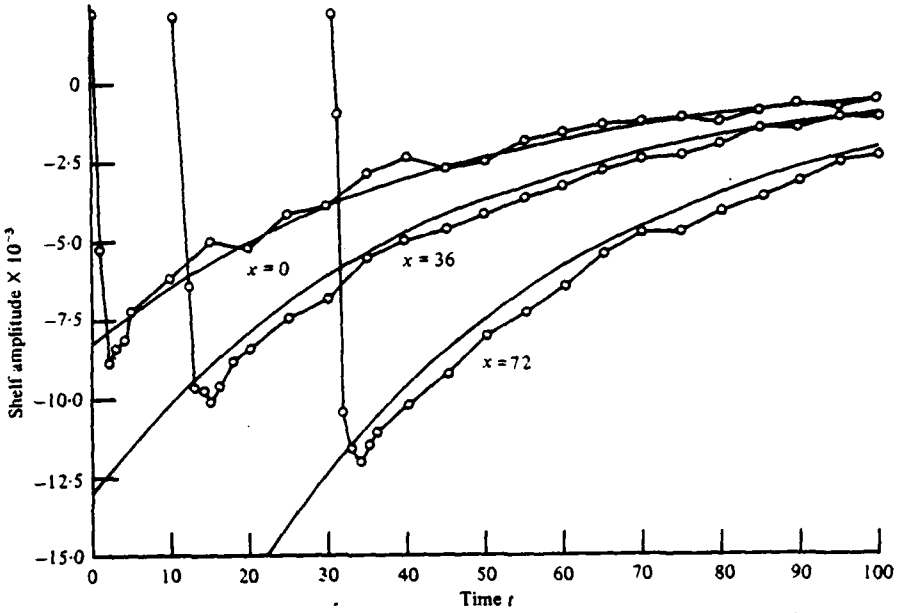


FIGURE 6. Numerical solution (—o—) versus perturbation theory (——) for shelf amplitude. The results which are quoted at $x = 0$ (initial soliton position), $x = 36$, and $x = 72$, graph shelf amplitude ($-15{\cdot}0 \times 10^{-3} < q(x, t) \leqslant 0$) versus time ($0 < t < 100$) for case (i). The rear portion of the solitary wave appears at the left of the figure.

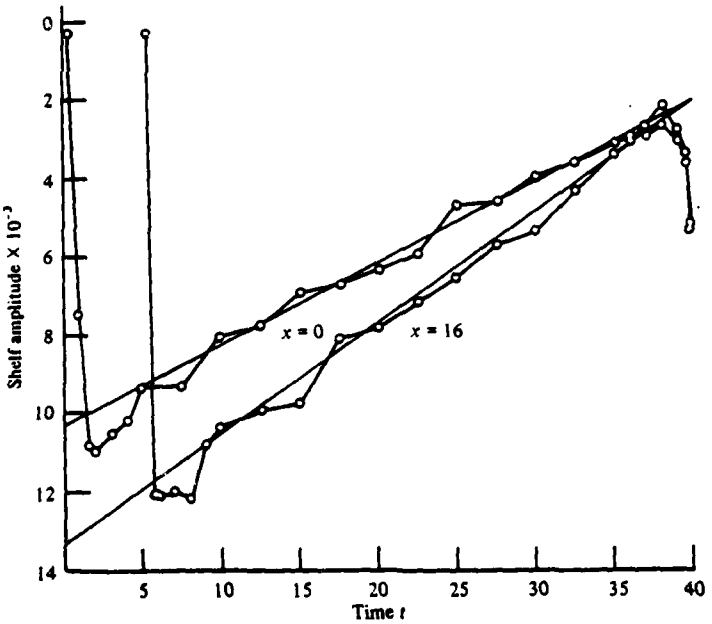


FIGURE 7. Numerical solution (—o—) versus perturbation theory (——) for shelf amplitude. The results which are quoted at $x = 0$ (initial soliton position) and $x = 16$ graph shelf amplitude ($-12{\cdot}0 \times 10^{-3} < q(x, t) < 2 \times 10^{-3}$) versus time ($0 < t < 40$) for case (ii). The rear portion of the solitary wave appears at the left of the figure.

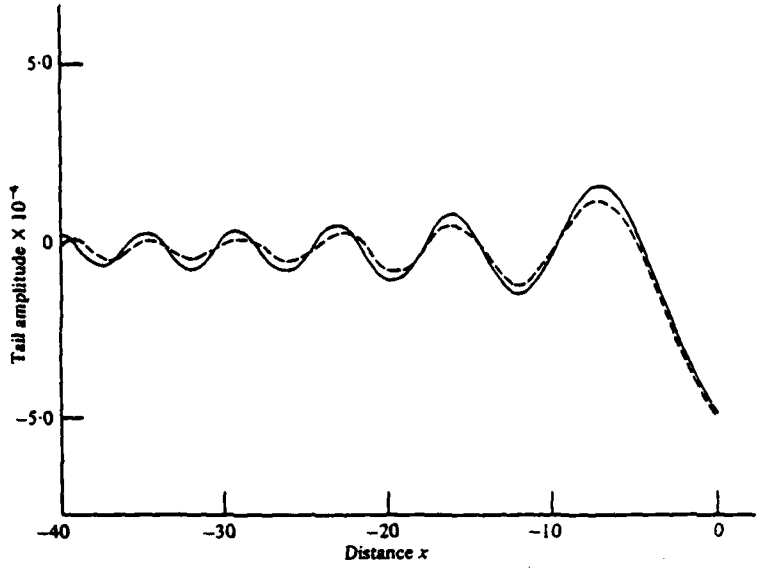


FIGURE 8. Numerical solution (—o—) versus perturbation theory (——) for oscillatory tail. The results graph tail amplitude ($-5{\cdot}0\times10^{-4} < q(x,t) < 5{\cdot}0\times10^{-4}$) *versus* distance ($-40 < x < 0$) for case (i) at $t = 7$ time units.

Figures 6 and 7 show in detail the evolution of the shelf after its initial formation and the perturbation and numerical results agree almost precisely. Although the slight differences between theory and experiment fall within the range of the numerical error, there is some suggestion of long waves propagating along the shelf from the solitary wave.

The results shown in figure 7 which are for case (ii) are qualitatively the same as case (i).

We also checked case (i) with $\sigma = +\frac{1}{40}$, the amplified case corresponding to an upward-facing shelf. After checking (3.8) against the numerical results we again found close agreement. Note that while the shelf amplitude decreases with x ($-\Gamma/3\eta\,(t_c)$ decreases with x), nevertheless at each point the shelf eventually grows with time (see (2.6), (3.8), with $\sigma = +\frac{1}{40}$). We remind the reader that the result (3.8) for the shelf is obtained by the balance $q_t = \sigma q$. Both qq_x and q_{xxx} are of smaller order. However, as the shelf grows, the nonlinear term again becomes important and indeed the shelf will begin to break and form order-one spatial derivatives on the time scale $\sigma^{-1}\ln\sigma^{-1}$. At this stage the dispersion becomes important; the shelf breaks into a solitary wave train, each pulse of which is weakly amplified by the σq term.

The last major portion of the solution considered was the oscillatory tail. This tail can be viewed in figure 1 for various time levels. In figure 8 we show a more detailed comparison of the perturbation results (2.3) and the numerical results for case (i) at $t = 7$ time units. The amplitude $\Gamma/3\eta$ in (2.3) has been adj usted to the value

$$(\Gamma(t_0)/3\eta(t_0))\exp\left(\int_{\tilde{t}}^{t_0}\Gamma(s)\,ds\right)$$

in line with (2.6) and (3.7); that is

$$q_{\text{tail}}(x,\tilde{t}) = \frac{\sigma e^{\sigma t}}{3\eta_0}\left[-\tfrac{2}{3} + \int_0^{x/(3\tilde{t})^{\frac{1}{3}}} \mathrm{Ai}\,(s)\,ds\right]. \tag{3.10}$$

The phase of the theoretical solution is not known exactly (the discrepancy is of the order of the width of the solitary wave) due to the lack of precise information as to where the shelf is formed with respect to the solitary wave. Therefore within this latitude we have chosen the phase of the theoretical solution (2.3) so that it agrees with the numerical solution at the point where the oscillatory tail attaches to the shelf. We believe that the slight remaining discrepancy between the theory and numerical experiment is due to the presence of very low-amplitude long waves which are continuously created at the solitary wave. As already noted some evidence for these waves is seen in figures 5–7. Indeed, a very careful examination of the time dependence in the oscillatory tail seems to indicate the presence of these long waves. However our present numerical scheme is not sufficiently accurate to study this very small effect in further detail.

The authors are grateful to NSF (MCS75-07548 A01) and ONR (N00014-76-C-0867) for support.

REFERENCES

Djordjevic, V. D. & Redekopp, L. G. 1978 The fission and disintegration of internal solitary waves moving over two-dimensional topography. *J. Phys. Oceanog.* **8**, 1016–1024.

Flaschka, H. & Newell, A. C. 1975 *Dynamical Systems, Theory and Applications* (ed. J. Moser), Lecture Notes in Physics, vol. 38, p. 335. Springer.

Gardner, C. S. 1971 The Korteweg–de Vries equation as a Hamiltonian system. *J. Math. Phys.* **12**, 1548–1551.

Gardner, C. S., Greene, J. M., Kruskal, M. D. & Miura, R. M. 1967 Method for solving the Korteweg deVries equation. *Phys. Rev. Lett.* **19**, 1095–1097.

Gardner, C. S., Greene, J. M., Kruskal, M. D. & Miura, R. M. 1974 Korteweg–de Vries equation and generalizations. VI. Methods for exact solution. *Comm. Pure Appl. Math.* **27**, 97–133.

Grimshaw, R. 1970 The solitary wave in water of variable depth. *J. Fluid Mech.* **42**, 639–656.

Grimshaw, R. 1971 The solitary wave in water of variable depth. Part 2. *J. Fluid Mech.* **46**, 611–622.

Johnson, R. S. 1973*a* Asymptotic solution of the Korteweg–de Vries with slowly varying coefficients. *J. Fluid Mech.* **60**, 813–824.

Johnson, R. S. 1973*b* On the development of a solitary wave moving over an uneven bottom. *Proc. Camb. Phil. Soc.* **73**, 183–203.

Kakutani, T. 1971 Effects of an uneven bottom on gravity waves. *J. Phys. Soc. Japan* **30**, 272–276.

Karpman, V. I. & Maslov, E. M. 1977 A perturbation theory for the Korteweg–de Vries equation. *Phys. Lett.* A **60**, 307–308.

Kaup, D. J. & Newell, A. C. 1978 Solitons as particles and oscillators and in slowly varying media: a singular perturbation theory. *Proc. Roy. Soc.* A **361**, 413–446.

Ko, K. & Kuehl, H. H. 1978 Korteweg–de Vries soliton in a slowly varying medium. *Phys. Rev. Lett.* **4**, 233–236.

Leibovich, S. & Randall, J. D. 1971 Dissipative effects on nonlinear waves in rotating fluids. *Phys. Fluids* **14**, 2559–2561.

Leibovich, S. & Randall, J. D. 1973 Amplification and decay of long nonlinear waves. *J. Fluid Mech.* **58**, 481–493.

Miles, J. W. 1979 On the Korteweg–de Vries equation for a gradually varying channel. *J. Fluid Mech.* **91**, 181–190.

NEWELL, A. C. 1978 Soliton perturbation and nonlinear focussing, symposium on nonlinear structure and dynamics in condensed matter. In *Solid State Physics*, vol. 8, pp. 52–68. Oxford University Press.

NEWELL, A. C. 1980 The inverse scattering transform. *Solitons* (ed. R. K. Bullough) *Topics in Current Physics*, vol. 17, Springer.

OTT, E. & SUDAN, R. N. 1970 Damping of solitary waves. *Phys. Fluids* **13**, 1432–1434.

PEREGRINE, D. H. 1967 Long waves on a beach. *J. Fluid Mech.* **27**, 815–827.

TAPPERT, F. & ZABUSKY, N. J. 1971 Gradient-induced fission of solitons. *Phys. Rev. Lett.* **27**, 1774–1776.

VLIEGENTHART, A. C. 1971 On finite difference methods for the Korteweg–de Vries equation. *J. Engng Math.* **5**, 137–155.

ZABUSKY, N. J. & KRUSKAL, M. D. 1965 Interactions of solitons in a collisionless plasma and the recurrence of initial states. *Phys. Rev. Lett.* **15**, 240–243.

ZAKHAROV, V. E. & FADDEEV, L. D. 1971 The Korteweg–de Vries equation. A completely integrable Hamiltonian system. *Anal. Priloz.* **5**, 18.

2.2.

INTERNAL SOLITARY WAVES NEAR A TURNING POINT

C.J. KNICKERBOCKER and Alan C. NEWELL
Department of Mathematics, Clarkson College of Technology, Potsdam, NY 13676, USA

Received 28 November 1979

We report an unexpected result and illustrate the mechanism by which a solitary wave propagating on a thermocline reverses its polarity as it passes through a turning point near the shoreline.

1. Introduction. Long solitary waves in lakes and estuaries, propagating on the thermocline separating two shallow layers of fluid of almost equal densities, are approximately described by the Korteweg–de Vries equation. The effect of the change in depth of the bottom layer, which the wave feels as it approaches the shore, results in the coefficient of the nonlinear term being a slowly varying function of position which has a zero at the point (the turning point) where the depths of the top and bottom layers are approximately equal. Since a solitary wave represents a balance between quadratic nonlinearity and linear dispersion, in the present context the sign of the nonlinear term determines that the wave must always face into the deeper layer. It has therefore been argued [1] that the solitary wave ceases to exist at the turning point and disintegrates into a train of dispersive nonlinear waves there. Although this does indeed occur, it is far from the whole story. What happens in addition is that the solitary wave, which initially faces into the bottom layer, develops in its wake a long shelf of the opposite polarity as it approaches the turning point. After the turning point, the original downward facing solitary wave disintegrates, but the upward facing shelf forms a new upward facing solitary wave which propagates through to the shoreline.

2. The model. Consider the situation as shown in fig. 1. Following the ideas of refs. [1,2], we find the equation which describes the nondimensional elevation,

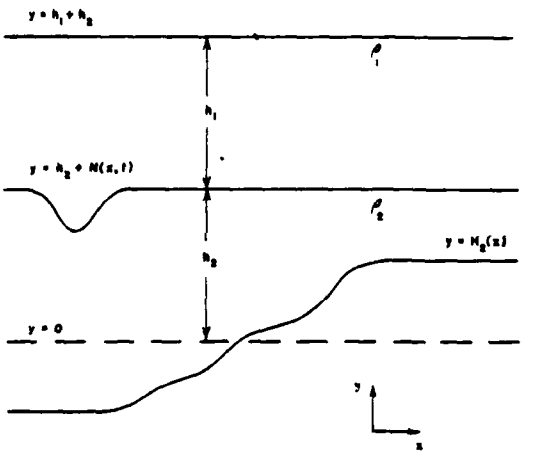


Fig. 1. Physical system being considered, where $N(x, t)$ is the dimensional elevation of the thermocline and the densities of the upper and lower fluids are ρ_1 and ρ_2, respectively.

$$V(X, \theta) = D(X)N(x, t)/\{[H_2(X)/h_2 - 1](\epsilon h_2)\},$$

of the thermocline separating two shallow layers of fluid (the effects of surface waves are neglected) of slightly different densities ρ_1, ρ_2 $(\rho_1 < \rho_2)$ is

$$V_X + A(X)VV_\theta + B(X)V_{\theta\theta\theta} + C(X)V = 0, \qquad (2.1)$$

where $X = \epsilon^{3/2}x/h_2$ measures position,

$$\theta = \epsilon^{-1}\int^X D^{-1/2}(r)\,dr - (g\epsilon/h_2)^{1/2}t$$

is the negative of the retarded time,

$$D(X) = \frac{(1-\rho_1/\rho_2)[1-H_2(X)/h_2]}{1+\rho_1[h_2-H_2(X)]/\rho_2 h_1},$$

x and t are the dimensional space and time coordinates and $0 < \epsilon \ll 1$ is the ratio of wave amplitude to depth. The coefficients $A(X)$, $B(X)$ and $C(X)$ are defined by

$$A(X) = 3(1-\rho_1 l_2^2/\rho_2 l_1^2)/2D^{1/2}(1-\rho_1/\rho_2)l_2,$$

$$B(X) = l_1 l_2(l_2+\rho_1 l_1/\rho_2)/6D^{3/2}(l_1+\rho_1 l_2/\rho_2),$$

$$C(X) = -D_X/4D + l_{2X}/2l_2,$$

where $l_1 = h_1/h_2$ and $l_2 = 1 - H_2(X)/h_2$ are considered to be slowly varying functions of X (i.e. $H_2^{-1}H_{2X} = O(\sigma)$), so that the change in the bottom slope occurs over many solitary wave widths; however, the terms reflecting changes in A, B and C are more important than the next terms (higher order nonlinearity and dispersion) which would appear in eq. (2.1) (i.e. $\epsilon^2 \ll \epsilon\sigma \ll \epsilon$). Note that the coefficient $A(X)$ of the nonlinear term changes sign where $\rho_2 l_1^2 = \rho_1 l_2^2$, which when $\rho_2 - \rho_1 = 0.06\rho_1$ occurs when l_1 and l_2 are approximately equal.

Using the transformation $W(\tau,\theta) = A(X)V(X,\theta)/B(X)$, where $\tau = \int^X B(r)\,dr$ we obtain

$$W_\tau + WW_\theta + W_{\theta\theta\theta} + (C + B_\tau/B - A_\tau/A)W = 0. \quad (2.2)$$

Once $H_2(X)$ begins to change, the terms $A^{-1}A_\tau$, $B^{-1}B_\tau$ and C are all important; however, as the turning point ($\tau = 1/\sigma$) is approached, the term $A^{-1}A_\tau$ is dominant. This allows us to simplify our model further by neglecting $C(\tau)$, changes in $B(\tau)$ and writing $A(\tau) = \sigma K(\tau - 1/\sigma)$, where

$$\sigma K = -3\rho_1 l_{2\tau}(\tau)/l_1^2(\rho_2-\rho_1)D^{1/2}(\tau)B(\tau),$$

evaluated at $\tau = 1/\sigma$. Setting $u(\tau,\theta) = KW(\tau,\theta)/A(\tau)$, we obtain what we shall call the TKdV (transitional Korteweg–de Vries) equation,

$$u_\tau + f(\tau)uu_\theta + u_{\theta\theta\theta} = 0, \quad (2.3)$$

where $f(\tau) = -1 + \sigma\tau$. We exploit the fact that $f(\tau)$ changes slowly with τ by setting $u(\tau,\theta) = 6q(\tau,\theta)/f(\tau)$, whence eq. (2.3) becomes

$$q_\tau + 6qq_\theta + q_{\theta\theta\theta} = (f_\tau/f)q. \quad (2.4)$$

3. Perturbation theory. There have been many analyses [2,3,6,8] which discuss a perturbation theory for eq. (2.4). However, it was only recently [4,5] that the role of the shelf in creating or destroying mass was appreciated. The global structure of the shelf, which is needed to show that $\int_{-\infty}^{\infty} u\,d\theta$ is indeed a constant (in τ) of the motion, was first given in ref. [7]. Here we follow the prescription given in refs. [6,7] and show how, by a judicious use of the conservation laws of eq. (2.4), we can obtain all the leading order behavior for distances τ and retarded times θ of order $1/\sigma$.

We begin at the point $\tau = 0$ when the perturbation is switched on with the pure soliton state,

$$q_s(\tau,\theta) = 2\eta^2\,\mathrm{sech}^2\eta(\theta-\bar\theta), \quad \bar\theta_\tau = 4\eta^2, \quad \eta = \eta_0.$$

First we calculate the slow change in the parameter $\eta(\tau)$ induced by the perturbation. By using the conservation of energy

$$\frac{\partial}{\partial\tau}\int_{-\infty}^{\infty} q_s^2\,d\theta = \frac{2f_\tau}{f}\int_{-\infty}^{\infty} q_s^2\,d\theta,$$

we obtain $\eta(\tau) = \eta_0 f^{2/3}(\tau)$. Second, from the conservation of center of gravity

$$\frac{\partial}{\partial\tau}\int_{-\infty}^{\infty}\theta q\,d\theta = 3\int_{-\infty}^{\infty} q^2\,d\theta + \frac{f_\tau}{f}\int_{-\infty}^{\infty}\theta q\,d\theta,$$

we find that to leading order, the solitary wave velocity $\bar\theta_\tau$ is $4\eta^2(\tau)$ [the true velocity, in the physical coordinates x, t is

$$x_t = -\theta_t/(\theta_x - \bar\theta_x)$$

$$\approx (gDh_2)^{1/2} + \frac{2(gDh_2)^{1/2}B^2(X)KD^{3/2}(X)N(x,t)}{A(X)[H_2(X)-h_2]}\,]$$

which can be integrated to give $\bar\theta$ as a function of τ:

$$\bar\theta(\tau) = 12\eta_0^2\{1 - [-f(\tau)]^{7/3}\}/7\sigma.$$

Next, we calculate the shape of the secondary structure. Because the solitary wave parameter has been already modulated to satisfy the conservation of energy requirement (the only consistent choice), the local mass conservation relation cannot be satisfied. Only two-thirds of the extra mass created (depleted) per unit distance by the perturbation can be absorbed (lost) by the solitary wave. Therefore a shelf $q_c(\tau,\theta)$ is created between $\theta = 0$ and $\theta = \bar\theta$, the present posi-

tion of the solitary wave. Its height upon creation can be found by balancing the expression for the local conservation of mass,

$$\frac{\partial}{\partial\tau}\int_{-\infty}^{\infty} q_s\,d\theta + \frac{\partial}{\partial\tau}\int_{0}^{\bar\theta} q_c\,d\theta \approx \frac{f_\tau}{f}\int_{-\infty}^{\infty} q_s\,d\theta.$$

Using $\bar\theta_\tau = 4\eta^2(\tau)$ yields

$$q_c(\bar\tau,\bar\theta) = [f_\tau/3\eta(\tau)f(\tau)]|_{\tau=\bar\tau},$$

which may be written as a function of $\bar\theta$ using the relation $f(\bar\tau) = \sigma\bar\tau - 1 = -(1 - 7\sigma\bar\theta/12\eta_0^2)^{3/7}$, the last expression obtained by integrating $\bar\theta_\tau = 4\eta^2$. Since the shelf amplitude is small and since it varies slowly with respect to θ, its subsequent evolution is given by balancing q_τ and $f_\tau q/f$. Thus, when the solitary wave is at $\hat\tau$, at retarded time $\hat\theta$, the height of the shelf at $\bar\theta$ is

$$q_c(\bar\theta,\hat\tau) = q_c(\bar\theta,\bar\tau)f(\hat\tau)/f(\bar\tau)$$

$$= \sigma(\sigma\hat\tau - 1)/3\eta_0(1 - 7\sigma\bar\theta/12\eta_0^2)^{8/7}, \quad 0 < \bar\theta < \hat\theta.$$

It is important to stress that while the shelf amplitude is order σ, its mass content is order one. Therefore the shelf plays a crucial role in the leading order description of the system. For example, from our results we note that solitary wave for TKdV is given by

$$u_s(\theta,\tau) = 12\eta_0^2[-f(\tau)]^{-1/3}$$

$$\times \operatorname{sech}^2\{\eta_0[-f(\tau)]^{2/3}(\theta - \bar\theta)\},$$

its mass content is $-24\eta_0(-f)^{-1/3}$; the shelf $u_c(\theta,\tau)$ is $2\sigma/\eta_0(1 - \sigma\theta/12\eta_0^2)^{8/7}$, $0 < \theta < \bar\theta$, and zero elsewhere and its mass content is $24\eta_0(-f)^{-1/3} - 24\eta_0$. Note the conservation of total mass $\int_{-\infty}^{\infty} u\,d\theta$ to leading order depends crucially on the existence of the shelf. We also observe that while the solitary wave decreases in amplitude, its mass content and consequently that of the shelf increases without bound as the turning point is approached. Indeed, the perturbation theory breaks down at a point where the amplitudes of the solitary wave and the shelf are of the same order, when $\tau = (1/\sigma)[1 - O(\sigma^{1/3})]$. Nevertheless, as we see from the numerical results which follow, the perturbation theory gives us qualitatively accurate results beyond the breakdown point; the solitary wave (and consequently the shelf) continues to gain negative (positive) mass. We remark that the perturbation procedure shown here can be used quite universally and results from other important examples are listed in the appendix.

4. Numerical results. Here we focus on fig. 2 which gives the results of a numerical integration of eq. (2.3) from $\tau = 0$, with $\sigma = 1/20$ and initial conditions

$$u_s(\theta,0) = -12\eta_0^2\operatorname{sech}^2\eta_0\theta.$$

We used a modification of the Vliegenthart scheme [9] and checked accuracy by continuously monitoring conservation of mass ($\int u\,d\theta = -24\eta_0$) and energy ($\int u^2 d\theta = 192\eta_0^3$). In fig. 2, we drew, at each fixed position τ, the negative (and retarded; the phase depends on location) time history θ of the pulse. For locations between $\tau = 0$ and 15 ($\tau = 16$ is the approximate breakdown position), the results of perturbation theory and numerical experiment were compared and are in close agreement. After the breakdown of the perturbation theory and before the turning point $15 < \tau < 20$, we observe that the solitary wave continues to gain (negative) mass. However it has also noticeably slowed down and, just before $\tau = 20$, has begun to travel backwards which, in physical coordinates, means that it is now traveling with a velocity less than that of the longest linear dispersive wave. The result of the slowing down is that the compensating gain of (positive) mass of the shelf takes place over smaller intervals and a peak begins to appear at the front of the shelf. At the turning point $\tau = 20$, the peak is quite pronounced. As the turning point is passed, the peak and the remaining shelf breaks up into a train of pulses. These pulses, if they are to be solitary waves, must still pass through the valley created by the remains of the original solitary wave. The criterion (determined numerically; one can also make a plausible analytical argument using inverse scattering theory) that a solitary wave forms is that for some τ, the amplitude of the leading pulse is at least twice that of the valley. At $\tau = 30$, we can see that, in this case, the leading pulse satisfies this criterion and is beginning to separate from the others which form part of a dispersive wave train. At $\tau = 40$, at which point we take $f(\tau) = 1$, the new upward facing solitary wave is about to emerge from the effects of the disintegrating original solitary wave. We continued the numerical calculation till $\tau = 200$ in order to determine that the emerging solitary wave has indeed the character of a Korteweg–de Vries soliton, $12k^2 \times \operatorname{sech}^2 k(\theta - \theta_0 - 4k^2\tau)$.

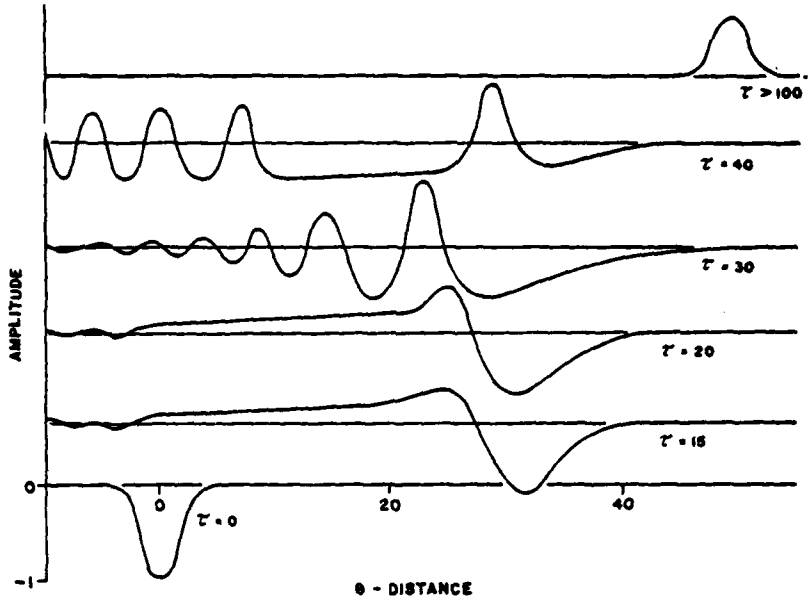


Fig. 2. Numerical integration of eq. (2.3) with $\eta_0 = 1$ and $\sigma = 1/20$. The turning point occurs at $\tau = 20$.

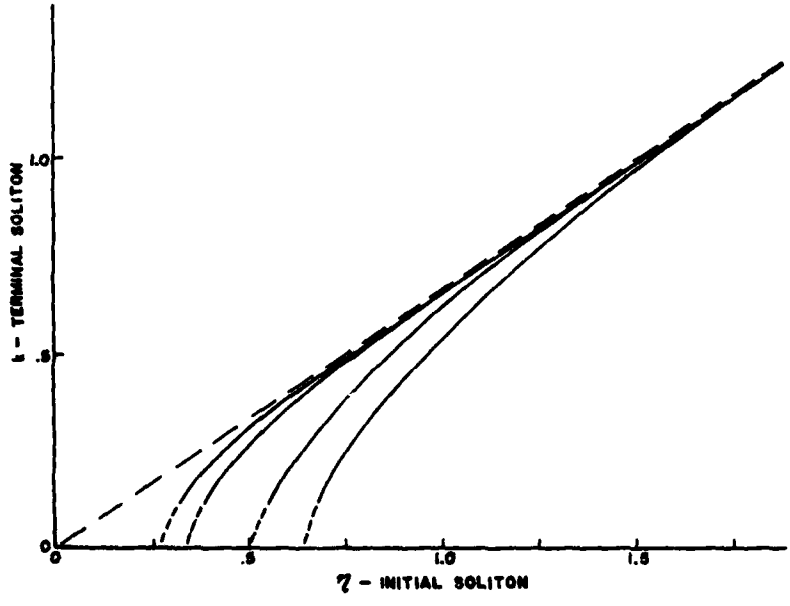


Fig. 3. This graph shows (——) the relationship between η ($= \eta_0$) of the initial soliton and k of the terminal soliton for various values of σ ($\sigma = 1/40, 1/20, 1/5, 1/2$ reading from left to right). The line (– – –) represents the relation $k = 2\eta_0/3$. The lines (- - -) appearing at the lower portion of the graph are due to the inability of the numerical scheme to give precise information for small k.

Fig. 3 is a graph of k versus η_0 for various σ. Note that as long as η_0 is large enough, the amplitude of the first upward facing pulse will be sufficient to create a solitary wave. For larger values of η_0, we note the remarkable result, for which we have no analytical explanation at the present time, that $k = 2\eta_0/3$, independent of σ.

Appendix. Perturbation results for various equations.

$$q_t + 6q^r q_X + q_{XXX} = F,$$

$$q_s = \alpha \operatorname{sech}^{2/r}\eta(X - \bar{X}),$$

$$\alpha^r = (r+1)(r+2)\eta^2/3r^2.$$

Example 1: $r = 1, F = \sigma q; \eta = \eta_0 \exp(2\sigma t/3)$, $\bar{X}_t \approx 4\eta^2$ and

$$q_c = \sigma \exp(\sigma t)/3\eta_0(1 + \sigma X/3\eta_0^2)^{5/4}, \quad \text{for } 0 < X < \bar{X}.$$

Example 2: $r = 1, F = \sigma q_{XX}; \eta = \eta_0(1 + 16\eta_0^2\sigma t/15)^{-1/2}, \bar{X}_t \approx 4\eta^2$ and

$$q_c = (8\eta_0\sigma/15)\exp(-2\sigma X/15), \quad \text{for } 0 < X < \bar{X}.$$

Example 3: $r = 2, F = \sigma q; \eta = \eta_0 \exp(2\sigma t), \bar{X}_t \approx \eta^2$ and

$$q_c = \pi\sigma \exp(2\sigma t)/\eta_0^2(1 + 4\sigma X/\eta_0^2), \quad \text{for } 0 < X < \bar{X}.$$

References

[1] V.D. Djordjevic and L.G. Redekopp, J. Phys. Ocean. 8 (1978) 1016.
[2] R.S. Johnson, Proc. Camb. Philos. Soc. 73 (1973) 183.
[3] S. Leibovich and J.D. Randall, J. Fluid Mech. 53 (1973) 481.
[4] V.I. Karpman and E.M. Maslov, Phys. Lett. 60A (1977) 307.
[5] D.J. Kaup and A.C. Newell, Proc. Roy. Soc. London A361 (1978) 413.
[6] C.J. Knickerbocker and A.C. Newell, Shelves and the Korteweg–de Vries equation, J. Fluid Mech. (1979), to be published.
[7] A.C. Newell, Soliton perturbation and nonlinear focussing, Symp. on Nonlinear structure and dynamics in condensed matter (Oxford Univ., Solid State Physics), (Springer, 1978) Vol. 8, pp. 52–68.
[8] E. Ott and R.N. Sudan, Phys. Fluids 13 (1970) 1432.
[9] A.C. Vliegenthart, J. Eng. Math. 5 (1971) 137.

3.1

From Pattern Formation & Recognition
Ed. H. Haken Springer 1979.

Bifurcation and Nonlinear Focusing

A.C. Newell

Clarkson College of Technology
Potsdam, NY 13676, USA

Abstract:

We suggest that the phenomenon of nonlinear focusing can play an important role in the transition of a system from one wavelike state to another.

1. Introduction

Nature provides an abundance of physical situations which are not simply conservative and, in addition to the balance of forces which give rise to wave motions, involve nonconservative effects such as diffusion and external influences which provide a reservoir of potential energy. The balance between the external forces and dissipation or restoring forces is usually characterized in the form of a parameter (or parameters), such as the Reynolds, Rayleigh or Taylor numbers in fluid mechanics, the north-south temperature gradient in meteorology, the loading parameter in elastic shell theory, the inversion number in lasers, the temperature in superconductors, the time step in a finite difference algorithm used to solve a partial differential equation. At certain critical values of these parameters, a fundamental and nonanalytic change in the nature of the solution occurs (it will be necessary to qualify this statement in what follows); for example, the change can be from a stationary state to a steady or wavelike, regular or irregular motion. When the resulting motion is ordered, like in a fluid heated from below (one-dimensional convection cells), or in a laser

---

This work is supported by NSF grant #MCS75-07548 A01 and ONR grant #N00014-76-C-0867. The lecture was given during the International Symposium on Synergetics, Schloss Elmau, April 30 - May 5, 1979.

(synchronized light emission) or a superconductor (boson like behavior of superconducting electrons - the Cooper pairs), then we call the new state a cooperative phenomenon [1]. On the other hand, it is often the case (e.g. instability of the Blasius flow, certain plasma instabilities) that while at the onset of instability the flow appears regular, it quickly degenerates and exhibits local turbulent bursts. It is the purpose of this article to suggest a possible explanation for the concentrated patches of irregular behavior.

Before we do this, it will be useful to review some background material on transition. Our first goal will be to identify those circumstances under which a system in which many normal modes are potentially present exhibits a behavior in which one particular configuration or pattern dominates all others. This happens in open systems, far from equilibrium, because one configuration can draw on the reservoir of potential energy more efficiently than others and, having grown to a finite amplitude state, can then inhibit the others from further growth. Present evidence seems to indicate that systems with degeneracy (or symmetry) have trouble deciding which of the many configurations, each of which can draw on the source of potential energy at exactly the same rate, are dominant. Therefore, we expect that in situations such as gravitational convection between two horizontal planes in which rolls of a fixed wavelength but arbitrary alignment can grow initially at the same rate, the dominant configuration, if at all realized, must arise because of the effects of sidewalls or some external forcing which gives preference to some particular roll alignment.

Conversely, one expects that in systems in which one mode is slightly preferred over the others, then this pattern will eventually dominate the flow and an ordered state will result. In many cases, this is indeed what happens. However, as we point out in sections 4 and 5, there is another mechanism which can work against the realization of the ordered state. This mechanism is dynamic in character and is a property of the wavelike behavior of the excited state. Briefly stated, in certain circumstances, nonlinear dispersive wavetrains do not wish to remain monochromatic with one characteristic wavevector $\vec{k}_c$. Instead, they often break up and focus into pulses. Sometimes the

process stops after the pulses achieve a certain amplitude. In other cases, the pulses continue to focus until the local finite amplitude effects cause the system to become widely irregular in local patches.

2. Nondegenerate Transition

Consider a mechanical system in which the so-called critical parameter (henceforth called the Reynolds number) R is near its lowest critical value R_c. At this value, linear stability analysis suggests that one of the normal modes of the system is about to make a transition from a damped (and perhaps oscillatory) or purely oscillating state to one which grows exponentially. Provided that the system is nondegenerate, that is, only one mode is about to destabilize, we can describe the behavior of the system near $R = R_c$ by an equation for the amplitude A of the mode in transition,

$$\delta A = I + \chi A + \alpha A^2 - \beta A^3 = F(\chi, A) \ . \tag{2.1}$$

In (2.1) δ stands for d/dt if the transition is from a damped to excited state, and $\frac{d^2}{dt^2}$ if the transition is from a neutral (oscillatory) to excited state. The parameter χ measures $R-R_c$ and for our discussion we will assume it to be real; α and β, both taken positive, measure the nonlinear reaction of the system. In the context of elastic shells, the quadratic term results from the influence of a nonlinear elastic foundation; the cubic term usually arises from a self-modal interaction. The parameter I we call the geometric imperfection after Koiter [2], who introduced the term to account for the imperfections which may be present in the shell before loading. As we point out later in our narrative, this constant term can result from many factors. It plays two very important roles.

Consider Figures 1 and 2. Figure 1 is the curve $F(\chi=R-R_c, A) = 0$ with $\alpha = 0$ and represents what is called a supercritical bifurcation. If $I = 0$, there is only one root $A = 0$ for $\chi = R-R_c < 0$ and three $A = 0, A = \pm \sqrt{\chi/\beta}$ for $\chi > 0$. The curves CD and CE represent stable solutions of (2.1); CF is

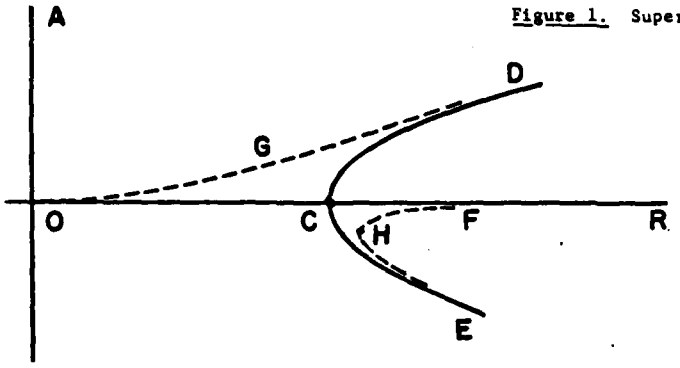


Figure 1. Supercritical Bifurcation

unstable. The transition at the bifurcation point is non-analytic; $\chi < 0$, $A = 0$; $\chi > 0$, $A = \pm \sqrt{\chi/\beta}$. If $I > 0$, then the curve $F(\chi,A)$ is the dotted curve OGD (stable) and EHF (EH stable, HF unstable). Note that now the transition is smooth and analytic. This is the first important role of the geometric imperfection. In [3], Benjamin describes how its presence (due to the effect of cylinders of finite length) affects the onset of Taylor vortices in the supercritical bifurcation of a flow between rotating cylinders. The Taylor vortices can be seen as ghostly apparitions at subcritical values of R, the Taylor number; however, they are amplified rather rapidly, albeit smoothly, when R is close to R_c.

Figure 2 shows the curve of $\alpha \neq 0$. This kind of bifurcation is called transcritical. Again, if $I = 0$, the "parabola" ECD in Figure 1 is simply displaced so that its vertex C is at $\chi = -\alpha^2/4\beta$, $A = \alpha/2\beta$. The portion of the curve R_cC is unstable; CD and R_cE are stable; R_cF is unstable. The stability properties of the various branches can be simply understood. Let A_o be an equilibrium solution of (2.1). Then if $A = A_o + \rho$, $d\rho/dt = (\partial F/\partial A)_o \rho$ to first order. But, $\frac{\partial F}{\partial \chi} + \frac{\partial F}{\partial A} \cdot \frac{dA}{d\chi} = 0$ and since $\frac{\partial F}{\partial \chi} = A$ we have, $\frac{d\rho}{dt} = -A_o \left(\frac{dA}{d\chi}\right)^{-1}_o \rho$. Hence, for $A_o > 0$ (< 0), the branch is unstable (stable) if A decreases with χ, stable (unstable) if A increases with χ. In the situation depicted in Figure 2, one can have a subcritical bifurcation; that is, for values of $\chi < 0$, there is a possibility that if perturbations are large enough,

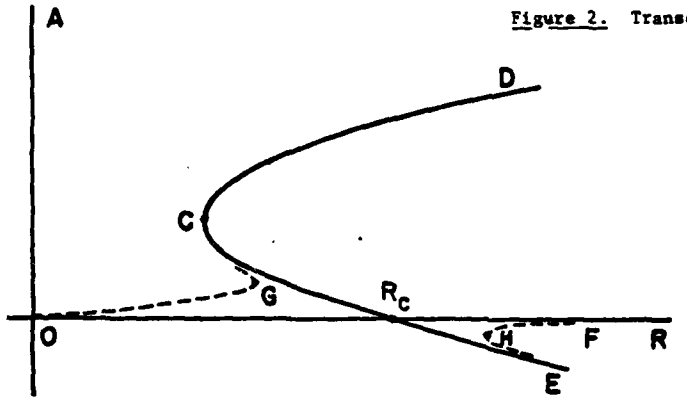


Figure 2. Transcritical Bifurcation

the system can transfer from the stable solution OR_c to the stable solution CD. The phenomenon of hysteresis is also present. In the case of transcritical bifurcation, the geometric imperfection I plays a second important role. When $I > 0$, the curve $F(\chi,A) = 0$ changes as shown by the dotted line in Figure 2. The curve has two separate branches OGCD (OG stable, GC unstable, CD stable) and EHF (EH stable, HF unstable). Moreover, we also note that if I is sufficiently large, the curve OGCD describes a single valued relation between the amplitude A and the Reynolds number R. However, the important point is that the imperfection provides a means for the system to reach the neighborhood of the unstable saddle points GC without the benefit of large disturbances. If χ or R is increased beyond the value corresponding to G, the system will be attracted to the only possible stable configuration on CD. It should be noted, however, that in the shell buckling problem there is no branch CD and therefore once G is reached, the shell begins a total collapse which is only halted when the shell assumes a completely new configuration (large buckles with sharp corners to absorb the energy). The role of the geometric imperfection in lowering the critical buckling load from R_c to G was one of the many fundamental contributions of Koiter [2] to our understanding of bifurcation phenomena.

In this lecture, I will suggest another means by which a system can realize locally the unstable branch GC without benefit of large perturbations. The mechanism is dynamic and relies on the fact that the transition is almost

degenerate in the sense that whereas one normal mode has the fastest growth, there is a continuum of others which destablize at only slightly higher values of R_c.

3. Degenerate Transitions

Degenerate transitions most often occur in infinite dimensional systems with symmetry. The problem of gravitational convection in a fluid contained between two horizontal planes in an adverse temperature gradient is a typical example. In this situation, a linear stability analysis of the purely conductive solution fixes the wavelength $\lambda_c = 2\pi/|\vec{k}_c|$ but not the direction $\frac{\vec{k}_c}{|\vec{k}_c|}$ of the normal modes in which convective motion can first take place. For the problem of an axially loaded cylindrical shell, the symmetry allows any mode with a wavevector lying on the locus $k_x^2 + k_y^2 = \pm k_x$ ($\vec{k} = (k_x,k_y)$, x is the axial, the circumferential coordinate) to appear. Moreover, the linear analysis permits any linear combination of these normal modes to occur; we call this the planform. For example, if the motion has only one $\vec{k}$ vector, then the motion is a roll whose axis lies in the direction perpendicular to $\vec{k}$. If the motion has three wavevectors $\vec{k}_1$, $\vec{k}_2$, $\vec{k}_3$, each of which satisfies $2\pi/|\vec{k}| = \lambda_c$ and each separated by 120°, then the planform is hexagonal. In the buckling of a cylindrical shell, the planform $\vec{k}_1 = \frac{2\pi}{\lambda_c}(-1,0)$, $\frac{2\pi}{\lambda_c}(k_x,k_y)$, $\frac{2\pi}{\lambda_c}(1-k_x,-k_y)$ has a diamond shape.

The question naturally arises: which of the various configurations are realized in a real experiment? In the neighborhood of $R = R_c$ each normal mode, if left on its own, would grow to a stage until the finite amplitude (nonlinear) effects sufficiently modify the excess adverse temperature gradient so as the motion settles down to a steady (or in other examples a limit cycle) state. This situation is essentially described by (2.1) with $I = \alpha = 0$. However, the modes, once they have reached a finite amplitude state, do not behave independently; rather they compete and out of this competition, a dominant planform may (or indeed may not) emerge. One way to gain some insight into

what might be the preferred mode is to consider the stability of the various steady solutions (each realized without interference from the others) when perturbed by all other allowable configurations. This approach was pioneered by Schlüter, Lortz and Busse [4] and extended by Busse (the Busse "balloon"; see [5]) and much of the work is either summarized or referenced in a recent review by Busse [6]. For example, using this approach, one can infer that if the Oberbeck-Boussinesq equations are modified slightly so as to include the viscosity dependence on temperature or quadratic effects in the density temperature relation, the hexagonal planform is preferred. The reason for this is worth elaborating (which we do below) because the resulting planform is due to the presence of quadratic terms ($\alpha \neq 0$ in (2.1)) in the amplitude relations and it is useful to understand how such terms naturally arise. In one sense they are atypical although catastrophy theory experts like to state that the transcritical bifurcation is generic (that is, more typical than the supercritical one). I will leave this semantic point to the reader to decide.

Let us now review the reasons for the existence of various terms in the amplitude equation (2.1) and discuss in what sense the quadratic terms are typical. (2.1) is derived by a perturbation analysis of the underlying equations written schematically as

$$L(\frac{\partial}{\partial t}\ ,\ \frac{\partial}{\partial x_j}\ ,\ R)u = N(\frac{\partial}{\partial t}\ ,\ \frac{\partial}{\partial x_j}\ ,\ R)[u^2,u^3,\ldots] \quad . \tag{3.1}$$

In (3.1), L is a linear operator acting on the scalar field $u(x_j,t)$ which describes the difference between the actual state of the system $V(x_j,t)$ and some known background state $V_s(x_j,t)$ whose stability is about to be lost as the Reynolds number is increased. The RHS represents all the nonlinear terms which become important when the solution $u(x_j,t)$ reaches finite amplitude.

One begins the analysis by investigating the linear stability problem. Simply ignore the RHS of (3.1) and seek solutions of

$$L(\frac{\partial}{\partial t}\ ,\ \frac{\partial}{\partial x_j}\ ,\ R)u = 0 \tag{3.2}$$

in the form

$$u(x_j,t) \propto e^{i(\tilde{k}_j\cdot\tilde{x}_j-\sigma t)}\phi(k'_j,x'_j) , \qquad \sigma = \omega + i\nu \quad . \tag{3.3}$$

In (3.3) $\tilde{x}_j$, j = 1,...M, corresponds to those spatial coordinates whose domain is infinite (or at least large compared with λ_c) and which are associated with a continuous spectrum k_j, and x'_j, j = M+1,...N corresponds to those coordinates which are finite in extent and lead to a quantized and discrete spectrum. In the case of gravitational convection in a fluid layer between two horizontal plates, N = 3, M = 2, $\tilde{x}_1$, $\tilde{x}_2$ correspond to x, y the horizontal coordinates $-\infty < x,y < \infty$ and z to the vertical coordinate $0 < z < 1$. Substitution of (3.3) into (3.2) with the appropriate choice of eigenfunctions $\phi(k'_j,x'_j)$ leads to the equation

$$L(-i\omega+\nu,\ ik_j,\ R) = 0 \tag{3.4}$$

(recall some of the k_j are discrete) which is going to be central in our future discussion. In general, (3.4) is a complex relation and allows us to solve for the <u>dispersion</u> ω and <u>growth rate</u> ν as functions of k_j and R.

$$\omega = \omega(k_j,R) \quad , \tag{3.5}$$

$$\nu = \nu(k_j,R) \quad . \tag{3.6}$$

Now the solution (3.3) is linearly unstable in the region of the (k_j,R) plane where $\nu(k_j,R) > 0$ and stable when $\nu(k_j,R) < 0$. This defines a <u>neutral surface</u>

$$\nu(k_j,R) = 0 \tag{3.7}$$

or solving for R assuming $\partial\nu/\partial R \neq 0$,

$$R = R(k_j) \quad . \tag{3.8}$$

Next, let us choose k_j such that R is minimal. One usually can choose among the discrete modes to achieve this end by inspection. At this point we fix

the values of k_j' corresponding to discrete modes at the values which lead to minimum R and call the corresponding lowest eigenfunction ϕ_o. Henceforth we omit the tilde on the continuous spectrum. From (3.7),

$$\frac{\partial \nu}{\partial k_j} + \frac{\partial \nu}{\partial R}\frac{\partial R}{\partial k_j} = 0 \tag{3.9}$$

and we choose k_{jc} (the $\tilde{k}_{jc}$), such that

$$\frac{\partial \nu}{\partial k_j} = \frac{\partial R}{\partial k_j} = 0 \quad . \tag{3.10}$$

As we have already pointed out, in those situations with built-in symmetries, these equations may lead to a surface

$$K(k_1,\ldots k_M) = 0 \quad , \tag{3.11}$$

rather than a unique wavevector $\vec{k}_{jc}$.

One now proceeds to look at the nonlinear problem in the neighborhood of R_c: that is, $R = R_c(1+\varepsilon^2\chi)$, by seeking a solution for (3.1) in the form

$$u(x_j,t) = \varepsilon(u_o + \varepsilon u_1 + \varepsilon^2 u_2 + \ldots) \tag{3.12}$$

where

$$u_o = \sum_r A_r(t) e^{i(\vec{k}_r\cdot\vec{x}-\omega_r t)} \phi_o(\vec{x}') + (*) \tag{3.13}$$

In (3.13), the subscript r refers to a particular wavevector $\vec{k}_r = (k_{1r}, k_{2r},\ldots k_{Mr})$ and not to the component. The coordinate $\vec{x} = (x_1,\ldots x_M)$, $\vec{x}' = (x_{M+1},\ldots x_N)$. Each $\vec{k}_r$ lies on the surface (3.11) and $\omega_r = \omega(\vec{k}_r, R_c)$ from (3.5). We know $L(-i\omega_c, i\vec{k}_c, R_c) = 0$. To find u_1, we will solve a linear inhomogeneous equation; the inhomogeneity will arise from the terms Nu_o^2 on the RHS of (3.1). Each of these terms will have the form

$$e^{i(\vec{k}_1+\vec{k}_2)\cdot\vec{x} - i(\omega_1+\omega_2)t}\, N\cdot\phi_o^2 \tag{3.14}$$

where N is some operator in the x_i' (e.g. $\phi_o \dfrac{\partial^2\phi_o}{\partial z^2}$).

In order for u_1 to have a solution, certain solvability conditions (the Fredholm alternative theorem) must be satisfied. Crudely speaking, in the case of self-adjont operators L and self-adjont boundary conditions, the term (3.14) is secular if it is an eigenfunction for the operator $L(\frac{\partial}{\partial t}, \frac{\partial}{\partial x_j}, R_c)$ corresponding to the eigenvalue R_c. In order for this to happen, we must have that (a) the term $N\phi_o^2$, when projected into the basis of which ϕ_o is the first member, must have a non-zero component of ϕ_o ; (b). $\vec{k}_1 + \vec{k}_2$ must belong to the critical surface (3.11); $\omega_1 + \omega_2 = \omega(\vec{k}_1) + \omega(\vec{k}_2)$ must be the natural frequency for the wavevector $\vec{k}_1+\vec{k}_2$ that is $\pm\,\omega(\vec{k}_1+\vec{k}_2)$. Only when these conditions are satisfied is there a quadratic term in the amplitude equation.

In order to gain some feeling for when they might be satisfied, consider the simplest† problem of convection. Here, the onset of instability is steady and $\omega(\vec{k}) = 0$. Hence, (c) is automatically satisfied. Since for this case the surface K is a circle $|\vec{k}|$ = constant, (b) is satisfied for three vectors $\vec{k}_1$, $\vec{k}_2$, $\vec{k}_3$ at angle 120°. If there is no vertical asymmetry in the problem, the first eigenfunction ϕ_o will be $\mathrm{Sin}\pi z$ which is symmetric about z = 1/2 and its square will have a zero projection into itself ($\int_o^1 \mathrm{Sin}^2\pi z \mathrm{Sin}\pi z = 0$). However, any slight vertical asymmetry will lead to an eigenfunction ϕ_o which is not symmetric about z = 1/2 and this will result in a term like $A_2^*A_3^*$ in the time evolution of A_1 ($\vec{k}_1+\vec{k}_2+\vec{k}_3 = 0$). Because of phase differences, such terms almost always (for a counterexample, see [7]) lead to the possibility of the onset of convection at subcritical values of R; the relevant bifurcation in these cases is transcritical.

The quadratic terms are also responsible for the existence of a subcritical branch in the (A,R) curves (Figure 2) in the case of shell buckling (see, for example, [8,9]).

Whereas quadratic terms can sometimes appear in (2.1), cubic terms always do. They will result from terms on the RHS of (3.1) of the form Nu_ou_1 which are typically of the form

† Free boundaries at z = 0,1.

$$e^{\pm i\theta_j \pm i\theta_\ell \pm i\theta_m} N\phi_o^3 \quad , \qquad \theta_j = \vec{k}_j \vec{x} - \omega_j t \quad , \tag{3.15}$$

over all (ℓ,m). But if $\theta_m = \mp \theta_\ell$ and $N\phi_o^3$ has a non-zero projection into ϕ_o (cf. $\int_o^1 \sin^3 \pi z \mathrm{Sin}\pi z dz \neq 0$), then the term $\sum_\ell A_1 A_\ell A_\ell^*$, which is called a modal interaction, will appear in the equation for the growth of A_1.

The equation for the growth of the amplitude of a typical wavevector $\vec{k}_1$ can now be written down:

$$C_1 \frac{dA_1}{dt} = (R-R_c)A_1 - \alpha_{123} A_2^* A_3^* - A_1 \sum_\ell \beta_{1\ell} A_\ell A_\ell^* \tag{3.16}$$

where C_1, α_{123} and $\beta_{1\ell}$ are constants. The solution of (3.16) and the companion equations for the other A_j do not have the property that a unique configuration emerges asymptotically in time which is independent of the initial conditions. There is an extremum principle due to Busse [10] but the principle describes a local property and does not exclude the possibility that both hexagons and rolls are stable solutions. Indeed, for certain ranges of $R-R_c$, that is the case.

It turns out, then, that the final state depends not only on the relative stability properties of the various configurations, but is also sensitive to initial conditions and most important to sidewall effects. Indeed, if the aspect ratio is less than 30, sidewall effects seem to dominate in that they help choose the particular roll alignments. For example, in the experiments of Koschmieder [11] done in a cylindrical dish, the rolls are circular; in Krishnamurti's [12] experiments, the rolls align themselves with the sidewalls of the rectangular layer.

For large aspect ratios, what happens is that the various configurations form locally in a fairly random fashion, according to whatever particular roll alignment was favored at the time at a particular location. Then configurations diffuse outwards (and some useful description of this process may be possible by an equation [4.3] derived below) until they interact with disturbances for other configurations. Since there is no real way for the various configurations

to decide which is dominant, the resulting flow will tend to exhibit irregular behavior. This fact is dramatically emphasized in the experiments of Ahlers and Behringer [13] who find that if the aspect ratio is 57, the fluid flow is turbulent in the sense that it has a nonperiodic time dependence for all $R > R_c$. It must be emphasized that their results are not a contradiction to the experimental results of Busse and Whitehead [5] nor to the theoretical results of Busse [14]. These authors do not allow spontaneous onset of any rolls but rather by the use of carefully controlled experiments (the controls in their experiments can be modelled by a geometric imperfection term in (3.16)), cause certain roll alignments to be favored over others. Then because they are locally stable, they remain dominant, at least until disturbances from distant boundaries can affect them.

To summarize then: the existence of too many allowable states causes the system to appear disordered. There is order in the sense that the typical wavelength of the flow is approximately λ_c but the roll orientation is distributed over many directions. Therefore, one might expect that if a system is to realize a natural ordered state, it must be almost one dimensional (that is, dominated by a single wavevector $\vec{k}_c$). However, as we shall see in the sections to come, there are other mechanisms lurking in the wings which can frustrate such a realization.

Before we discuss these, let us emphasize that disorder or turbulence can appear in systems of small dimension (specifically three) and does not require an infinity of competing modes [15,16,17,18]. For example, in those situations where the first transition is to an ordered state (cf. in circular Couette flow), typically what happens is as follows. As the Reynolds number (Taylor number in the circular Couette flow) is raised, one sees another bifurcation to periodic (in time) behavior (a 1 torus). When the Reynolds number is increased further, a second frequency appears (flow on a 2-torus). At a value of the Reynolds number at which these two frequencies fall into a rational relationship, there is some evidence of aperiodic flow, the spectrum broadens and the flow appears chaotic. Thus, a complete disorder is possible

even though only three modes of the system are present. However, all this occurs at values of the Reynolds number considerably beyond the value at which the first bifurcation takes place (in the circular Couette problem, the Taylor number is approximately 20 times the linear critical value) which is the parameter range we are investigating in this paper.

4. The Generalized Newell-Whitehead Equation and Focusing

If the system is truly one-dimensional and nondegenerate in the sense that the motion can be characterized by a single wavevector $\vec{k}_c$, then the system will evolve according to an equation like (2.1). If in addition $\alpha = 0$, then for supercritical Reynolds numbers, the system saturates ($A^2 = \frac{R-R_c}{\beta}$). However, if the geometry admits a continuum of wavevectors in the neighborhood of $\vec{k}_c$ (such is the case if the aspect ratio in the convection problem is infinite or if the cylinders in the circular Couette problem are infinitely long), then for a given $R > R_c$, a finite bandwidth of order $\sqrt{R-R_c}$ of wavenumbers can be excited. Even if the spectrum is discrete (due to the effects of finite but distant boundaries), it is necessary to incorporate in the description of the flow those modes whose wavevectors lie in a $O((R-R_c)^{1/2})$ neighborhood of $\vec{k}_c$. One does this by allowing the amplitude A to be a slowly varying function of both position and time. The envelope equation we will derive was first developed by Newell and Whitehead [19] and Segel [20] for the case of transition from a zero to a steady state. If one allows only dependence on the direction parallel to $\vec{k}_c$, the appropriate bandwidth is $(R-R_c)^{1/2}$ and the equation has Ginzburg-Landau form. If wavevectors in the direction perpendicular to $\vec{k}_c$ are included, the relevant bandwidth is $O(R-R_c)^{1/4}$. For more details, see the above references or the book by Joseph [21]. When the transition is to a state of growing oscillations (the "overstable" case), then the envelope equation includes the effects of group velocity and dispersion and was first given in [22].

We seek solutions of the form (3.12) to the equation (3.1) with

$$u_o(\vec{x},\vec{x}',t) = A(\vec{X},T_1,T_2,..)e^{i(\vec{k}_c\cdot\vec{x}-\omega(\vec{k}_c,R_c)t)}\phi_o(\vec{x}') + (*) \qquad (4.1)$$

where $\vec{k}_c$ is the most critical wavevector with corresponding frequency $\omega_c = \omega(\vec{k}_c,R_c)$, $\vec{x}'$ are those coordinates of finite extent, $\vec{X} = \varepsilon\vec{x}$, $T_1 = \varepsilon t$, $T_2 = \varepsilon^2 t$ and $R = R_c(1 + \varepsilon^2\chi)$. The effect of differentiating the product of a slowly varying envelope A and a fast varying exponential phase by the independent variables can be modelled by treating $\vec{x}$, t, $\vec{X}$, T_1 and T_2 as independent variables and making the following transformations:

$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} + \varepsilon\frac{\partial}{\partial T_1} + \varepsilon^2\frac{\partial}{\partial T_2}$, $\frac{\partial}{\partial x_j} \to \frac{\partial}{\partial x_j} + \varepsilon\frac{\partial}{\partial X_j}$. With this formal structure, equation (3.1) is now written:

$$\{L(\frac{\partial}{\partial t}, \frac{\partial}{\partial x_j}, R_c) + \varepsilon(L_o\frac{\partial}{\partial T_1} + L_j\frac{\partial}{\partial X_j}) + \varepsilon^2(L_o\frac{\partial}{\partial T_2} + \frac{1}{2}L_{oo}\frac{\partial^2}{\partial T_1^2}$$

$$+ L_{oj}\frac{\partial^2}{\partial X_j\partial T_1} + L_{jl}\frac{\partial^2}{\partial X_j\partial X_l} + L_R R_c\chi)\}(u_o + \varepsilon u_1 + \varepsilon^2 u_2)$$

$$= \varepsilon N u_o^2 + \varepsilon^2 N(u_o u_1, u_o^3) \quad . \qquad (4.2)$$

where we have expanded $L(\frac{\partial}{\partial t} + \varepsilon\frac{\partial}{\partial T_1} + \varepsilon^2\frac{\partial}{\partial T_2}, \frac{\partial}{\partial x_j} + \varepsilon\frac{\partial}{\partial X_j}, R_c(1+\varepsilon^2\chi))$ in a Taylor expansion. In (4.2), the summation convention is implied. To order one, (4.2) is exactly satisfied by the choice (4.1). At order ε, the only secular terms are $(L_o\frac{\partial}{\partial T_1} + L_j\frac{\partial}{\partial X_j})u_o$ since unless $\omega(2\vec{k}_c) = 2\omega(\vec{k}_c)$, a very unlikely occurrence, Nu_o^2 only contributes second harmonic terms to u_1. Removing the secular terms and using (A.8) we find

$$\frac{\partial A}{\partial T_1} + (\frac{\partial\omega}{\partial k_j})_c\frac{\partial A}{\partial X_j} = 0 \qquad (4.3)$$

which shows that the envelope $A(\vec{X}, T_1, T_2)$ depends on $\vec{x}$ and t through the

combinations $\hat{X} = \varepsilon(\vec{x} - \nabla\omega t)$ and $T = \varepsilon^2 t$. Henceforth, when we write $\vec{X}$, we will mean $\varepsilon(\vec{x} - \nabla\omega t)$. $\nabla\omega$ is the group velocity of the most unstable wave.

At order ε^2, there will be secular terms produced by the RHS of (4.2) in the form of a self modal interaction with coefficient $-\beta L_o A^2 A^*$, $\beta = \beta_r + i\beta_i$. All the other secular terms come from the ε^2 terms in the expansion of L. Using (4.3) and the relations (A.9) and (A.10) of Appendix 1, we obtain that the solvability condition for u_2 is the envelope equation

$$\frac{\partial A}{\partial T} - \frac{i}{2}\sum_{j,\ell}\frac{\partial^2\omega}{\partial k_j \partial k_\ell}\frac{\partial^2 A}{\partial X_j \partial X_\ell} - \frac{1}{2}\frac{\partial \nu}{\partial R}\sum_{j,\ell}\frac{\partial^2 R}{\partial k_j \partial k_\ell}\frac{\partial^2 A}{\partial X_j \partial X_\ell}$$

$$= (\frac{\partial \nu}{\partial R} - i\frac{\partial \omega}{\partial R})R_c \chi A - (\beta_r + i\beta_i)A^2 A^* \quad . \tag{4.4}$$

The term $-i\frac{\partial \omega}{\partial R} R_c \chi A$ may be removed by taking the ω_c in (3.1) to be $\omega(\vec{k}_c, R)$ rather than $\omega(\vec{k}_c, R_c)$. All the coefficients in (4.4) are estimated at critical. Also, the second partial derivatives of ω keep $R = R_c$ fixed in (3.5): The main difference between (4.4) and (2.1) with $\alpha = I = 0$ is the presence of the dispersion term $-\frac{1}{2}\sum \frac{\partial^2\omega}{\partial k_j \partial k_\ell}\frac{\partial^2 A}{\partial X_j \partial X_\ell}$ and diffusion term $-\frac{\partial \nu}{\partial R}\sum \frac{\partial^2 R}{\partial k_j \partial k_\ell}\frac{\partial^2 A}{\partial X_j \partial X_\ell}$. The latter acts as a diffusion term; the matrix $\frac{\partial^2 R}{\partial k_j \partial k_\ell}$ is positive definite by definition of $\vec{k}_c$ and helps the system to approach an ordered state in which A is independent of $\vec{X}$. Notice that a sideband solution $A = e^{i\vec{K}\cdot\vec{X}} B$ (in which case the effective wavevector $\vec{k} = \vec{k}_c + \varepsilon\vec{K}$) will grow at the rate $\frac{\partial \nu}{\partial R}(R_c\chi - \frac{\partial^2 R}{\partial k_j \partial k_\ell}\cdot \vec{K}_j \vec{K}_\ell)$ which is less than the growth rate of $\vec{k}_c$ itself.

If χ, $\beta_r > 0$ one might expect that A will tend asymptotically in time to the ordered state

$$A(\vec{X},T) = \sqrt{\frac{\chi'}{\beta_r}}\exp\left(-i\left(\frac{\beta_i \chi'}{\beta_r}\right)T\right) \quad , \quad \chi' = \frac{\partial \nu}{\partial R} R_c \chi \, , \tag{4.5}$$

whence $u_o(\vec{x},\vec{x}',t)$ consists of a monochromatic wavetrain with wavenumber $\vec{k}_c$ and the nonlinearly adjusted frequency $\omega = \omega(\vec{k}_c, R) + \varepsilon^2 \beta_i \chi'/\beta_r$. However, there is a destabilizing mechanism, first discovered by Benjamin and Feir [23] in

connection with water waves, which involves only the dispersive terms of the equation and the imaginary part of the coefficient β. In fact, it can readily be shown that the solution (4.5) is only stable when the matrix

$$M = (m_{j\ell}) = \beta_i \frac{\partial^2 \omega}{\partial k_j \partial k_\ell} + \beta_r \frac{\partial \nu}{\partial R} \frac{\partial^2 R}{\partial k_j \partial k_\ell} \tag{4.6}$$

is positive definite. This criterion reflects the battle between the "cooperative" tendency of the system (namely, an initial power spectrum of the spatial correlation function $\langle u(\vec{x})u(\vec{x}+\vec{r})\rangle - \langle u\rangle^2$ narrows around $\vec{k} = \vec{k}_c$ and the system develops an ordered structure) and the "dispersive" tendency in which nonlinear dispersive wavetrains prefer not to remain monochromatic but rather to focus into pulses.

If the matrix M is not positive definite, then certain sidebands of $\vec{k}_c$ will grow at its expense. The resulting broadening of the power spectrum around $\vec{k}_c$ is associated with the formation of pulses in $\vec{X}$ space.

In one dimension and in the absence of the real terms in (4.3), i.e. $\beta_r = \frac{\partial \nu}{\partial R} = 0$, (4.4) is the nonlinear Schrodinger equation and the pulses form solitons. Also in one dimension, Lange and Newell [24] investigated the statistical initial value problem for (4.3) and showed that if $\beta_i\omega'' + \beta_r \frac{\partial \nu}{\partial R} R'' > 0$, the system does indeed realize the ordered state in which

$$\langle A \rangle = A = \sqrt{\chi/\beta_r}\, \exp -i(\beta_i/\beta_r)\chi^{\frac{1}{2}}T$$

and all the higher cumulants tend to zero. If $\beta_i\omega'' + \beta_r \frac{\partial \nu}{\partial R} R'' < 0$, the higher cumulants diverge with time.

However, if the dimension of the system is greater than one, then the work of Zakharov and Synakh [25] suggests that a much more dramatic phenomenon can occur. In order to gain some understanding of what can happen, we again neglect the terms with real coefficients in (4.4) and obtain the higher dimensional nonlinear Schrodinger equation [26] of which the canonical form in two dimensions is $\phi_t - i(a\phi_{xx} + b\phi_{yy}) - 2ic\phi^2\phi^* = 0$. If $a = b = c = 1$, the solution $\phi(x,y,t)$ collapses in a finite time in a self-similar manner for a sufficiently large value of the motion constant $\int_0^\infty r|\phi|^2 dr$, $r = \sqrt{x^2+y^2}$. If $a = -b = c = 1$,

the case of deep water gravity waves, then y independent solutions are unstable essentially as a result of the weakly nonlinear quartet resonance mechanism of the underlying carrier wave. In fact, by setting $x = r\cosh\theta$, $y = r\sinh\theta$, $x^2 > y^2$, we see that the solution will collapse onto the resonance curves $x^2 = y^2$ (the portion of the Phillips [27] figure of eight curve near the vertex). If $a = b = -c = 1$, then the system disperses and the long time behavior is given by the two dimensional version of the Benney-Newell [26] similarity solution $\phi(x,y,t) \sim \frac{1}{t} B(r/t)\exp(ir^2/4t + 2i/t\, B^2(\frac{r}{t}) + \ldots)$.

How are these results likely to apply to (4.3)? The first point we make is that if M is non-positive, the system will not be cooperative in the usual sense. The second and more important point is that if M is negative definite, then the solution begins to collapse. If $\beta_r > 0$ then the collapse is eventually stopped but the system begins to oscillate and can create local collapsing points elsewhere and the process is repeated. Furthermore, if β_r is very small (as it is in the case of Blasius flow), then the large local amplitudes can give rise to secondary (e.g. inflexionai point) instabilities. If $\beta_r < 0$, then even though $\chi < 0$, the subcritical case, the collapse can overcome the initial damping of the system locally and reach amplitudes (corresponding to the branch GC in Fig. 2)† at which the nonlinear instability occurs. Focusing then provides a mechanism whereby the amplitude of linearly stable but nonlinearly unstable waves can reach the critical amplitude without the benefit of imperfections, end effects or large initial perturbations. One of the key factors in the focusing mechanism is the strong, nonlinear interaction between neighboring wavenumbers.

It is, of course, necessary that the sideband modes are not damped out before the strong focusing can take place. One might conjecture, then, that the focusing mechanism is a very important feature in systems where the transition is from a neutral state to an excited one for then at subcritical values of the critical parameter, all the sideband modes compete on an equal footing. We discuss the focusing phenomenon in this context in the following section.

†In this case due to cubic terms.

5. Transition from Neutral States and Focusing

Although much of what we have to say in this section carries over to the buckling of elastic shells, the model we choose is an explicit momentum conserving finite difference algorithm for solving the modified Korteweg de Vries equation, $u_t + u_{xxx} + 6u^2u_x = 0$. Specifically, if m and n are the time and space steps respectively, the algorithm is

$$u(m+1,n) - u(m-1,n) - \alpha(u(m,n+2) - 2u(m,n+1) + 2u(m,n-1) - u(m,n-2))$$
$$= \gamma(u^3(m,n+1) - u^3(m,n-1)) \quad , \tag{5.1}$$

where $\alpha = \Delta t/(\Delta x)^3$ and $\gamma = -2\Delta t/\Delta x$. A linear stability analysis of the $u = 0$ solution of the form

$$u \propto \exp(i\mu nh - i\theta mk) \quad , \quad \mu \text{ real} \quad , \quad \theta = \theta_r + i\theta_i \tag{5.2}$$

shows that instability will set in at

$$\alpha = \frac{1}{4\sin\mu h \sin^2\frac{\mu h}{2}} \tag{5.3}$$

with $\theta_r k = \pi/2\ \mathrm{Sgn}\mu$. The minimum α is realized when $\mu = \mu_c$, $\mu_c h = 2\pi/3$ and $\alpha_c = .385$. We stress that for $\alpha < \alpha_c$, the modes (5.2) simply oscillate.

A nonlinear analysis in which we take

$$\alpha = \alpha_c(1 + \varepsilon^2\chi) \tag{5.4}$$

$$u = \varepsilon(u_o + \varepsilon u_1 + \dots) \tag{5.5}$$

where

$$u_o = A(m)\exp(i\mu_c nh - im\pi/2) + (*) \tag{5.6}$$

yields an amplitude equation

$$\Delta_T^2 A = 2\chi A + \delta A^2 A^* \tag{5.7}$$

where $\Delta_T^2 A = A(m+1) - 2A(m) + A(m-1)$ and $\delta = -6\gamma\sin\mu_c h > 0$. Note that from (5.7) we can infer that finite amplitude subcritical instabilities are possible.

Namely, if $\chi < 0$ the solution $AA^* = -2\chi/\delta$ is unstable for the initial amplitude of A (not $\sqrt{2|\chi|/\delta}$, because A is complex, but a little below it) which we call $\bar{A}$. If we perturb the system (5.1) with a mode with the critical structure ($u(o,n) = 2B\cos\mu_c nh$) then from (5.7) we would expect the envelope A of (5.7) to oscillate if $B < \bar{A}$ and grow rapidly if $B > \bar{A}$.

When we tested the results suggested by (5.7) in a numerical experiment, we found to our surprise that the nonlinear instability could be triggered for values of B well below the critical threshold. What happens is that the solution of (5.7) coupled with the underlying "carrier wave" structure $\exp(i\mu_c nh - im\pi/2)$ is unstable to perturbations with different wavenumbers μ. This instability is analogous to the Benjamin-Feir instability (which was the mechanism for the onset of focusing in the nonlinear Schrodinger equation discussed in the previous section). In fact, in order to account for the finite bandwidth, one may assume $A = A(m,n)$ is a slowly varying function of time (m) and space (n). This adds the term $-2\Delta_x^2 A = -2(A(m,n+1) - 2A(m,n) + A(m,n-1))$ to the left-hand side of (5.7) and one can show that the periodic in time solutions of (5.7) are unstable to X (or n) dependent disturbances. The spread of energy in wavenumber space is manifested as a focusing of the envelope A in X (or n) space. Thus, whereas the initial A may be below the critical threshold at every point in space, its subsequent evolution is such that it decreases in most areas and focuses at a few points. It can continue to focus (if the initial amplitude is much smaller than threshold, this can take a very long time! See [28]) until locally the critical threshold for nonlinear destabilization is attained. At this point, the calculation explodes rapidly.

The details of this work are reported in the literature in reference [28].

6. Conclusions

We suggest that dynamic focusing may be an important factor in bifurcation phenomena when the transition is from a wave-like (neutral or weakly damped) state to a state of growing waves. In particular, it may cause local patches of wildly irregular behavior in otherwise regular patterns. It provides a

mechanism by which the unstable subcritical branch can be reached dynamically without the aid of large initial disturbances or material imperfections. It may be extremely relevant in elastic shell buckling. It may also be responsible for a phenomenon familiar to some numerical analysts, namely, the sudden appearance of local breakdown of a partial difference equation, which initially was well within stability margins, after a very long time.

APPENDIX I

The Dispersion Relation

We begin with (3.4)

$$L(-i\omega(k_j,R) + \nu(k_j,R),\ ik_j,\ R) = 0 \tag{A.1}$$

which holds for all k_j and R. Differentiate with respect to k_j, the result with respect to k_ℓ and R to obtain:

$$L_o(-i\frac{\partial\omega}{\partial k_j} + \frac{\partial\nu}{\partial k_j}) + iL_j = 0 \quad , \tag{A.2}$$

$$L_o(-i\frac{\partial^2\omega}{\partial k_j\partial k_\ell} + \frac{\partial^2\nu}{\partial k_j\partial k_\ell}) + L_{oo}(-i\frac{\partial\omega}{\partial k_j} + \frac{\partial\nu}{\partial k_j})(-i\frac{\partial\omega}{\partial k_\ell} + \frac{\partial\nu}{\partial k_\ell})$$
$$+ L_{o\ell}(\frac{\partial\omega}{\partial k_j} + i\frac{\partial\nu}{\partial k_j}) + L_{jo}(\frac{\partial\omega}{\partial k_\ell} + i\frac{\partial\nu}{\partial k_\ell}) - L_{j\ell} = 0 \quad , \tag{A.3}$$

$$L_o(-i\frac{\partial\omega}{\partial R} + \frac{\partial\nu}{\partial R}) + L_R = 0 \quad , \tag{A.4}$$

where L_o is the derivative of L with respect to its first argument, L_j to its j^{th} (that is with respect to $\partial/\partial x_j$), L_R with respect to R. L_{oo}, L_{oj} and $L_{j\ell}$ are the second partial derivatives which we assume continuous in the neighborhood of $\vec{k}_c$ and R_c. Now, from (3.6) we have

$$\frac{\partial\nu}{\partial k_j} + \frac{\partial\nu}{\partial R}\frac{\partial R}{\partial k_j} = 0 \tag{A.5}$$

and the critical wavevector $\vec{k}_c$ is chosen so that

$$\frac{\partial \nu}{\partial k_j} = \frac{\partial R}{\partial k_j} = 0 \quad , \quad j = 1,\ldots M \quad . \tag{A.6}$$

Take the derivative of (A.5) with respect to k_ℓ and evaluate at $\vec{k}_c$ to obtain

$$\left.\frac{\partial^2 \nu}{\partial k_j \partial k_\ell}\right|_{\vec{k}_c} = - \frac{\partial \nu}{\partial R} \left.\frac{\partial^2 R}{\partial k_j \partial k_\ell}\right|_{\vec{k}_c} \quad . \tag{A.7}$$

From (A.2), (A.4), we have that at critical $\vec{k}_c$

$$L_j = \frac{\partial \omega}{\partial k_j} L_o \quad , \tag{A.8}$$

$$L_R = (i \frac{\partial \omega}{\partial R} - \frac{\partial \nu}{\partial R}) L_o \quad , \tag{A.9}$$

and (summation convention implied) from (A.3) and (A.7)

$$- L_{oo} \frac{\partial \omega}{\partial k_j} \frac{\partial \omega}{\partial k_\ell} + \frac{\partial \omega}{\partial k_j} L_{o\ell} + \frac{\partial \omega}{\partial k_\ell} L_{oj} - L_{j\ell}$$

$$= L_o (i \frac{\partial^2 \omega}{\partial k_j \partial k_\ell} + \frac{\partial \nu}{\partial R} \frac{\partial^2 R}{\partial k_j \partial k_\ell}) \quad . \tag{A.10}$$

REFERENCES

1. H. Haken. Synergetics. Springer-Verlag (1978).
2. W.T. Koiter. Elastic stability and post-buckling behavior. Proc. Sym. on Nonlinear Problems. R.E. Langer, ed., Univ. of Wisconsin Press, Madison (1963).
3. T.B. Benjamin. Univ. of Essex. Fluid Mech. Report 76 (1976).
4. A. Schlüter, D. Lortz and F.H. Busse. J. Fluid Mech. 23, 129 (1965).
5. F.H. Busse and J.A. Whitehead. J. Fluid Mech. 47, 308 (1971).
6. F.H. Busse. Reports on Progress in Physics 41, 1929 (1978).
7. A.C. Newell. J. Atmos. Sci. 29, 64 (1972).
8. C.G. Lange and A.C. Newell. SIAM Journal of Appl. Math. 21, 608 (1971).
9. C.G. Lange and A.C. Newell. J. Appl. Mech., 575, June (1973).
10. F.H. Busse. J. Fluid Mech. 30, 625 (1967).
11. E.L. Koschmieder. Beitr. Phys. Atmos. 39, 1 (1966).

3.2

From Chaos and Order in Nature
Ed H. Haken Springer 1987

The Mechanism by Which Many Partial Difference Equations Destabilize

W. Briggs, A.C. Newell, and T. Sarie

Department of Mathematics and Computer Science, Clarkson College of Technology
Potsdam, NY 13676, USA

The motivation for this work was to attempt to understand the reason that certain classes of explicit numerical algorithms for computing solutions to partial differential equations develop intense local patches of instability after long times, even when all the linear stability criteria are satisfied. Now, numerical analysts are the academic world's greatest plumbers, and so practitioners of the art of numerical computation have invented many ingenious schemes to circumvent these problems (see, for example [1]). However, beyond declaring that the instabilities are nonlinear in character and beyond a few careful analyses of the problem (for example [2], [3]), the numerical analyst has not investigated in any great detail the nature of the breakdown.

Our thesis is that the phenomenon has a universal character and has features in common with the instability of monochromatic surface gravity waves on the sea, the development of Langmuir turbulence in plasmas and the intense spots which occur when a laser pulse passes through a nonlinear dielectric [4]. We suggest it will occur in those situations in which:

(i) The algorithm has a linear instability at some value $(\Delta t)_c$ of the time step Δt which has the character of a transition from a neutral (not damped) state to an exponentially growing one.

(ii) The algorithm has a finite amplitude subcritical nonlinear instability threshold which may be represented by a graph of "initial amplitude" A versus Δt which begins at $A=0$, $\Delta t=(\Delta t)_c$ and rises (in a more or less parabolic fashion) as Δt decreases. Here, A is the amplitude of a potentially unstable mode. For values of A and Δt below the critical curve, the algorithm should be neutrally stable. Above the critical curve the nonlinear instability rapidly (within 10 time steps) sets in. The existence of the threshold is due to nonlinear terms which represent either a cubic self-interaction of the mode with itself or a quadratic term which represents the interaction of the mode with its subharmonic.

Whereas it would appear that the ultimate reason for the breakdown of the algorithm is that the system has reached (at least locally) the critical threshold, one must explain how it can do this since the major source of perturbations is round-off error and these can be extremely well controlled. Therefore, large perturbations simply do not occur spontaneously. The answer to this difficulty and indeed the answer to the questions (a) why does it take so long? (b) why is it local? is that the system possesses

(iii) the focusing property.

We will explain this property with the aid of a simple example. Let us imagine that we are attempting to use a leap-frog scheme in order to compute the constant solution U of

$$u_t + uu_x = 0, \tag{1}$$

or the u = 0 solution for

$$u_t + (U+u)u_x = 0. \tag{2}$$

Nothing could be simpler! The algorithm is

$$u(m+1,n) - u(m-1,n) + \frac{\theta\gamma}{2}\left(u^2(m,n+1) - u^2(m,n-1)\right) \tag{3}$$

$$+((1-\theta)\gamma u(m,n) + \alpha)(u(m,n+1) - u(m,n-1)) = 0$$

where $\alpha = \frac{U\Delta t}{\Delta x}$, $\gamma = \frac{\Delta t}{\Delta x}$ and θ is chosen to have the value 2/3 in order to conserve the discrete conservation property

$$\sum_{n=1}^{N} u(m,n)u(m+1,n) \text{ is independent of } m\ , \tag{4}$$

a property necessary to suppress a fast-acting instability [3].

A straightforward linear stability analysis quickly reveals that the mode

$$u(m,n) = ae^{i(n-m)\pi/2} + (*),$$

(*) is complex conjugate, is the most unstable mode. Indeed if we set

$$u(m,n) = a(m)e^{i(n-m)\pi/2} + (*) + b(m)e^{i(n-m)\pi/2} \tag{5}$$

in (3), we obtain exactly

$$a(m+1) - 2a(m) + a(m-1) = 2(\alpha-1)a(m) + 2/3\gamma a^*(m)b(m), \tag{6}$$

$$b(m+1) - b(m-1) = +\frac{2}{3}\gamma i(a^2(m) - a^{*2}(m)). \tag{7}$$

Note that the linear stability result is contained in (6). For $\alpha<1$, the $\pi/2$ mode is neutrally stable; for $\alpha>1$, it grows exponentially. The subcritical nonlinear instability is due to the excitation of its second harmonic b(m). If b(m) is initially small, then for $\alpha<1$ a(m) is sinusoidal, and a^2-a^{*2} contains a constant term in addition to the second harmonic. Thus b(m) initially grows linearly and if the initial a(0), a(1) are sufficiently large (as measured by some parameter A), then the nonlinear term a^*b in (6) overcomes the linear term.

Likewise, we may obtain an exact result if we take three modes,

$$u(m,n) = A(m)e^{i\pi/3 n} + B(m)e^{i\frac{2\pi}{3}n} + C(m)e^{i\pi n} + (*), \tag{8}$$

whence A,B,C satisfy

$$A(m+1) - A(m-1) + \frac{i\gamma}{\sqrt{3}}(2A^*B+B(C+C^*))(m) + i\alpha\sqrt{3}A(m) = 0, \qquad (9)$$

$$B(m+1) - B(m-1) + \frac{i\gamma}{\sqrt{3}}(2A^2+A^*(C+C^*))(m) + i\alpha\sqrt{3}B(m) = 0, \qquad (10)$$

$$C(m+1) - C(m-1) + \frac{2i\gamma}{\sqrt{3}}A(m)B(m) = 0. \qquad (11)$$

Again, using (9)-(11), we can obtain a critical stability curve A (representing the initial conditions A(0), A(1), B(0), B(1), C(0), C(1)) vs. α.

If we take values A, below the critical stability curve, the solutions A,B,C of (9),(10),(11) simply oscillate. Since (9)-(11) are exact, we should expect to get the same result if we insert these initial values into (8) to obtain u(0,n), u(1,n) and then integrate via (3). The experiment which is shown below in Figs. 1-5 shows what happened when we took $\alpha = .9$, $N = 60$, and $A = .06$. The dashed line in Figs. 1,3,5 is the critical threshold. At $m = 200$, Fig. 1, the solution still retains its period 6 structure, its amplitude well below critical everywhere. Figure 2 shows the spectral revolution. The first noticeable changes occur at about $m = 1600$, Fig. 3, when it becomes clear that the envelope of the solution is not constant but has begun to develop a structure of its own. The spectral resolution, Fig. 4, shows that energy has leaked to the sidebands of $\pi/3$ and π. The envelope continues to focus and at $m = 8004$, Fig. 5, one value of u(m,n) is about to cross the critical threshold. Very soon thereafter, the solution exhibits rapid growth and the spectrum becomes that of a Drac delta function. During all this time, the area $\sum_0^{N-1} u(m,n)$ and the functional $\sum_0^{N-1} u(m,n)u(m+1,n)$ are conserved exactly.

So what has happened? It would appear that amplitude alone is not the criterion for stability. If the system has the property that the monochromatic wavetrain (the pure $(\frac{\pi}{2},\pi)$ or $(\frac{\pi}{3},\frac{2\pi}{3},\pi)$ mode solutions) is unstable to sideband disturbances, and if this envelope instability continues to grow, then the critical threshold can be reached locally without the benefit of large initial perturbations. In our experiment, the envelope instability was induced purely by round-off error. The initial conditions were the pure $(\frac{\pi}{3},\frac{2\pi}{3},\pi)$ mode to 8 decimal places; double precision simply delayed but did not stop the focusing. In short, the solutions of (9)-(11) are exact but unstable solutions of the full equation (3).

In summary, then, we have found that if the numerical algorithm has the properties (i) its potentially unstable modes are neutrally stable according to linear theory (ii) a subcritical nonlinear stability threshold and (iii) the focusing property, then the algorithm will ultimately exhibit unstable behavior no matter what the time step is. Furthermore, we conjecture that the time $t = m\Delta t$ on which the instability occurs is relatively independent of Δt. We do not yet know what constitutes a sufficient condition that the focusing is sufficiently intense for u(m,n) to cross threshold for some m,n but we suspect it is an "initial area" criterion. Finally, we emphasize that this mechanism has a universal nature and appears to occur in all these algorithms which satisfy properties (i),(ii),(iii). Some of these ideas were already discussed in [5] and their relevance to onset of turbulence in other continuum mechanical contexts is suggested in [6]

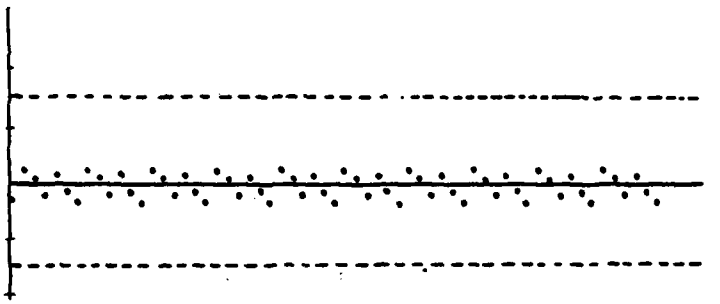


Fig. 1 u(200,n)



Fig. 2 The spectral resolution of u(200,n)

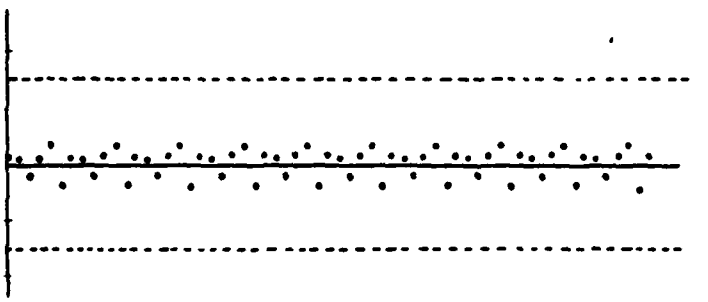


Fig. 3 u(1600,n)

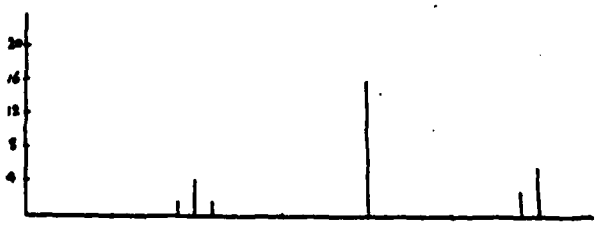


Fig. 4 The spectral resolution of u(1600,n)

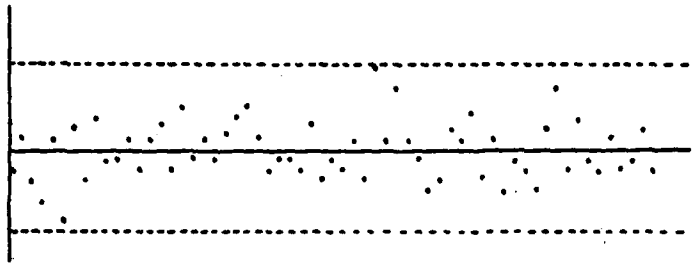


Fig. 5 u(8004,n).

[1] J.M. Hyman, Advances in Computer Methods for Partial Differential Equations III. IMACS, 313-321 (1979).

[2] H.O. Kriss and J. Oliger, Tellus, 24, 199-215 (1972), GARP Publ. No. 10, (1973).

[3] B. Fornberg, Math. of Comp., 27, 45-57 (1973).

[4] Y.S. Sigov and V.E. Zakharov, Journal de Physique, 40, C7-63 (1979).

[5] A.C. Newell, SIAM J. Appl. Math. A, 33, 133-160 (1977).

[6] A.C. Newell: "Bifurcation and Nonlinear Focusing", in *Pattern Formation by Dynamic Systems and Pattern Recognition*, ed. by H. Haken, Springer Series in Synergetics, Vol. 5 (Springer Berlin, Heidelberg, New York 1979) pp. 244-265

3.2

FOCUSING : A MECHANISM FOR INSTABILITY ON NONLINEAR FINITE DIFFERENCE EQUATIONS.

William L. Briggs[+]

Alan C. Newell*

Talib Sarie[+]

[+] Department of Mathematics and Computer Science
Clarkson College of Technology
Potsdam, New York 13676

* Department of Mathematics
University of Arizona
Tucson, Arizona 85721

INTRODUCTION AND GENERAL DISCUSSION

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

$$u_t + uu_x = 0.$$

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

to suppress instability and sustain computations over long times. In particular, we will discuss, in the context of the example used in this paper, ways in which one might judiciously choose the frequency at which one must apply these remedies.

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

$$\exp(i2\pi\ell_1 + \frac{n}{N})\exp(i2\pi\ell_2 + \frac{n}{N}) \rightarrow \exp(i2\pi(\ell_1 + \ell_2)\frac{n}{N})$$

$$\text{where } \ell_1 + \ell_2 < \frac{N}{2}$$

and to "indirect" interactions which involve aliasing error[6,9] or (what crystal physicists would call) Umklapp processes in which a wavenumber $\ell = \ell_1 + \ell_2 > \frac{N}{2}$ is misrepresented by the wavenumber $\ell = N - \ell_1 - \ell_2$ due to the inability of the grid to resolve wavelengths smaller than $2\Delta x$. It is evident that the wavenumber sets $\{\pm \frac{N}{4}, \pm \frac{N}{2}\}$ and $\{\pm \frac{N}{6}, \pm \frac{N}{3}, \pm \frac{N}{2}\}$ are closed under quadratic interactions (e.g. $\frac{N}{4} + \frac{N}{2} = \frac{3N}{4} = N - \frac{N}{4}$).

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

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

perturbations.

This instability mechanism evolves as follows. The solutions of the ordinary difference equations which correspond to values of (α,E) in the neutrally stable region of Figure 1a and which are exact solutions of the original partial difference equations, are unstable. They are unstable to modes which are their immediate neighbors in wavenumber space. The instability, which results from the nonlinear interaction between the original modes and their sidebands, manifests itself as a distortion of the envelope of the noise. The exact solution will have a spatial period of the order of the grid length ($4\Delta x$ - $6\Delta x$ in our examples) depending on which set of N_1 modes is used. The envelope of the exact solution is constant in space and oscillates in time. In our experiment, the initial noise in the sideband modes triggering the instability is due to roundoff error. In real calculations, there would generally be some energy already in these modes. The instability mechanism itself is a noise amplifier. Its character (initial growth rate and wavelength) is independent of the size of the grid and the degree of precision used in the calculations. Its wavelength is chosen dynamically as being the one of optimal growth. We understand the initial stages of this process. For the later stages, we have developed an envelope equation which appears to describe the subsequent growth reasonably well. The envelope begins to distort and slowly develops sharp peaks (focuses) at isolated points along the grid. When the local amplitude reaches the critical threshold given by Figure 1a, the noise level accelerates drammatically and becomes unbounded within relatively few time steps.

Thus the process by which the partial difference equation destabilizes is a two-fold one. At first the noise level in the potentially unstable modes (introduced in real computations by nonlinear cascade) is not large. The noise begins to focus and, if the spatial grid is large enough, can eventually reach

the critical threshold locally. Depending on the initial noise amplitude, this focusing process can take a long time (often on the order of 10^3-10^4 time steps) to develop. At this point, the conditions for nonlinear finite amplitude instability are satisfied and the noise grows without bound. The critical parameter in determining whether a partial difference equation is unstable is a combination of both noise level and grid size. In this sense, for large enough grids, the Leapfrog method is always unstable!

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

The contents of the paper are as follows. In Section 2, we present the stability diagrams on which the discussion of nonlinear instability is usually based. Along the way, we see, in the context of our example, the nonlinear instabilities described by Phillips and Kreiss and Oliger. In Section 3, we illustrate the instability of the solutions used in Section 2 and display through numerical experiments, the focusing property. We include careful experiments showing that this behavior is not simply due to lack of computational

precision, but rather is a genuine instability whose initial growth rate is independent of the precision and the size of the grid. In Section 4, we introduce an envelope equation as an attempt to give a universal equation which would describe the focusing process for a larger class of schemes. In the last section, we discuss some ideas about how to apply various remedies to inhibit instability and we advance some conjectures concerning the parameters on which the focusing property might depend.

II. DIFFERENCE SCHEME AND AMPLITUDE EQUATIONS

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

$$u_t + uu_x = 0$$

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

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

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

$$u(m+1,n) - u(m-1,n) + \frac{\theta\gamma}{2}\left[u^2(m,n+1) - u^2(m,n-1)\right] + \left[(1-\theta)\gamma u(m,n) + \alpha\right]\left[u(m,n+1) - u(m,n-1)\right] = 0 \qquad (2.1)$$

$$\text{for } 0 \le n \le N-1, \quad m \ge 1$$

$$u(m,0) = u(m,N).$$

where $\theta \in \mathbb{R}$, $\gamma = \frac{k}{h}$, $\alpha = \frac{kU}{h}$.

The nonlinear term has been discretized in two different ways. It is not difficult to show that with $\theta = 2/3$ the scheme satifies the conservation properties that

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

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

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

$$u(m+1,n) - u(m-1,n) + \alpha[u(m,n+1) - u(m,n-1)] = 0$$

$$0 \leq n \leq N-1$$

have normal mode solutions of the form

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

The frequencies ϕ_1 and ϕ_2 are real and given by

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

provided that α is less than the critical value

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

The frequency ϕ_1 is associated with the physical mode and converges to the exact solution, while ϕ_2 belongs to a spurious or computational mode[6]. Note that $\alpha_p^* \geq 1$ and that if $\alpha \leq \alpha_p^*$, then mode p is neutral. However, if $\alpha > \alpha_p^*$, then mode p grows exponentially in m (or time). The smallest critical value of α occurs for $p = \frac{N}{4}$ when $\alpha_p^* = 1$. This corresponds to the spatial mode

$e^{in\pi/2}$ which has a wavelength of 4h. This is identified as the most unstable mode. With p = 0, the corresponding spatial mode $e^{in\pi}$ has wavelength 2h and is neutrally stable for all α. Finally with $\alpha = 1$, $\phi_{1,2} = \frac{2p\pi}{N}$, $\pi - \frac{2p\pi}{N}$ one choice of which gives the dispersion relation of the continuous problem.

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

A. ONE MODE SOLUTION

A solution of the form

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

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

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

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

gives $u(m,0) = u(m,3) = 0$, $u(m,1) = -u(m,2) = -\sqrt{3}\, a(m)$ where $a(m+1) = a(m-1) - a^2(m)$. This pattern leads to unbounded growth and the choice $\theta = 2/3$ is again advisable in order to suppress this fast acting instability.

B. TWO MODE SOLUTION

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

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

and obtain the exact amplitude equations

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

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

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

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

whence (2.5) becomes

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

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

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

The amplitude equations (2.5a,b) are very revealing and deserve careful analysis. First notice that the A*B term which appears in the $\pi/2$ mode

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

Equations (2.5a,b) also show clearing the role which the computational mode plays in the development of finite amplitude instability. Assume that initially the amplitudes A(m) and B(m) are small, in which case the linear portions of (2.5a,b) will determine their growth. We then have

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

$$B(m) = B_p(-1)^m + B_c$$

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

$$B(m+1) - B(m-1) = -\frac{2}{3} i\gamma(A_c^2 - A_p^{*2})e^{2im\phi} - \frac{2}{3} i\gamma(A_p^2 - A_c^{*2})e^{-2im\phi}$$

$$+ \frac{8}{3}\gamma(-1)^m \mathrm{Im}(A_p A_c).$$

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

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

where a, b are constant independent of m.

The nonlinear term of the right hand side of the A equation (2.5a) will now reflect this growth in B(m):

$$A^*(m)B(m) = e^{-im\phi}\{-\frac{2}{3}\gamma \mathrm{Im}(A_p A_c)A_c^* m + bA_p^* + A_c^* B_p + A_c^* B_c(-1)^m\}$$

$$+ e^{im\phi}\{-\frac{2}{3}\gamma \mathrm{Im}(A_p A_c)A_p^* m + aA_c^* + A_p^* B_c + A_p^* B_p(-1)^m\}$$

$$+ \text{higher harmonics.}$$

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

We compute the nonlinear stability threshold as follows. Let

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

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

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

In Figure 1, we show the regions of the (α,E)

plane which correspond to bounded (for 2×10^4 time steps, the solution oscillates) and unbounded (usually overflow occurs in less than 10^2 time steps) solutions. The transition in the (α,E) plane from bounded to unbounded solutions is not smooth. Given recent experience with mappings, it is not surprising that the boundary is irregular and that the domains of attraction of the bounded and unbounded solutions are interspersed. We stress, however, that when we examine the stability of solutions in Section 3, we begin with initial conditions which belong to the stable region of the amplitude equations.

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

Part of this behavior is due to the fact that initial phase (which we have chosen to be fixed) is also important in determining the final disposition of the solution. Our choice of weighting the three modes $e^{\pm i\pi n/2}$, $e^{i\pi n}$ equally does not significantly affect the average position of the stability boundary nor the qualitative features of Figure 1b. It does change quantitatively the complicated patterns seen near the boundary and it does change the actual number of steps needed to reach instability. We confirmed this by choosing different weightings while keeping the "energy"

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

fixed.

The curves, in Figure 1, can be considered to be representative. It is interesting to note that the stability boundary reaches a maximum at approximately $\alpha = .5$ and then returns to $E = 0$ at $\alpha = 0$. Recall that since we keep γ fixed

(equal to one) in these experiments, α getting smaller means that the size of the solution U about which we perturb is getting smaller. One might argue from (2.6a) that the smaller α is, the larger is the linear restoring force which the nonlinearity must overcome. However this thinking is really only of value when α is close to one and we can take the continuous time limit of (2.6). It is better to consider Eqs. (2.5). If we write $A(m) = x_m + iy_m$, $B(m) = b_m$, then (2.5a) reads for $\alpha = 0$, $\gamma = 1$,

$$x_{m+1} = x_{m-1} - \frac{2}{3} b_m y_m$$

$$y_{m+1} = y_{m-1} - \frac{2}{3} b_m x_m$$

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

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

and therefore always increases. This result is not significantly affected by a change of initial conditions. For example, if the energy is redistributed in a different manner among $x_0, x_1, y_0, y_1, b_0, b_1$ the stability threshold at $\alpha = 0$ can increase to as much as + .05. Thus the role of α, for α small, is to dephase x_m and y_m which inhibits the monotonic growth of b_m.

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

C. THREE MODE SOLUTION

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

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

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

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

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

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

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

D. FOUR MODE SOLUTION

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

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

The amplitudes must satisfy the ordinary difference equations

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

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

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

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

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

III. FOCUSING IN THE PARTIAL DIFFERENCE EQUATIONS

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

Consider the specific case of the exact three mode solution

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

According to the stability curve of Figure 1, a value of $\alpha = .9$ and an initial amplitude of $E = .1$ should produce a stable solution of the amplitude equations. Yet when the partial difference equations are solved with $\alpha = .9$ and $E = .1$, something unexpected happens. The results of this calculation are shown in Figure 2. With N=300 grid points on the interval $0 \le x \le 1$, the solution is plotted at time steps m = 400, 1000, 2000, 2200, 2400, 2680. The dashed

lines indicate the critical amplitude at which finite amplitude instability sets in according to the stability curve of Figure 1. (In this case the critical amplitude is about .2). Clearly, the initial amplitude in this case is subcritical. At m = 400 and m = 1000 (Figure 2a,b), the solution still retains the periodic structure of the initial conditions; its amplitude is well below critical everywhere. By m = 2000 (Figure 2c), the constant envelope of the initial profile begins to vary slowly in x. The solution remains well-contained through m = 2400 (Figure 2e) although local amplitudes have exceeded their initial value. At m = 2680 (Figure 2f) the solution exceeds the threshold value at a single grid point. This completes the first stage of the development of the instability. It is characterized by the slow gathering or focusing of the solution locally. Once the solution reaches the critical threshold at even a single grid point, the second stage of the development takes place swiftly. By m = 2700, finite amplitude instability, as predicted by the amplitude equations, has taken over and the solution grows without bound. During the integration, the two quantities $\sum_n u(m,n)$ and $\sum_n u(m+1,n)u(m,n)$ are conserved exactly.

Some understanding of this process may be gained by looking at the Fourier spectrum of the solution at the same time steps shown in Figure 2. On a grid of N = 300 points there are 150 distinct modes with mode j of the form $e^{i\frac{\pi j}{150}}$ having wavelength of $(\frac{300}{j})h$. After m = 1000 time steps (Figure 3a) the energy is still in the three modes of the initial conditions. By m = 2000 time steps (Figure 3b) the energy has spread to the sidebands with wave numbers $\mu = 45,55$, $\mu = 95,105$ and $\mu = 145$ ($\mu = 50$ is the period 6 or $\frac{\pi}{3}$ mode). This corresponds to an envelope modulation of wavelength $\frac{300}{5} = 60$ and the constant envelope of the initial profile begins to vary slowly in x. <u>In short, the exact solutions of the amplitude equations are unstable solutions of the full</u>

partial difference equations. This is the beginning of the focusing process. The slow modulation of the envelope is triggered only by the presence of errors ther in the initial conditions or in computation. Figure 3c at m = 2200 time steps shows a further spreading of the energy in wave number space corresponding to a continued enhancement of the modulation in the envelope. During time steps $m \geq 2400$ (Figure 3d) the energy is distributed through all wave numbers approaching a uniform distribution. In physical space this corresponds to the envelope of the solution having focused into a local peak with subcritical amplitude. Once focusing elevates the maximum amplitude above the threshold, finite amplitude instability sets in, leading to a rapid deterioration of the solution.

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

Figures 4 show another sequence of experimental results. We choose initial conditions consisting of the $\pi/2$ and π modes and N, the number of grid points, is 300. With $\alpha = .9$, the critical amplitude (by Figure 1) is $E_c = 0.36$ and is marked by a dashed line in the figures. For these calculations, an initial amplitude of E = .15 was chosen. For early times the envelope oscillates in a manner almost independent of x and in precise agreement with the motion predicted by Eqs. (2.5a,b). One can think of the system as consisting of a chain of coupled oscillators in a nonlinear potential. For early times, their orbits are almost synchronized. However, a careful analysis of the spectrum reveals that the envelope has begun to deform and one can already see

the long wave modulation at times m = 100, 400 (Figures 4a,b). The spectral decompositions of u(100,n) (Figure 5b) shows that the sidebands k = 72, 78 (k = 75 is the period 4 or $\pi/2$ mode) are excited. This corresponds to an envelope modulation of wavelength 300/3 = 100. This wavelength is chosen dynamically, is a function of the initial amplitude but is not a function of N, the number of grid points. This fact was verified by taking values of N ranging from 60 to 600. The fact that there is a most unstable sideband and that the wavelength of the most unstable mode is inversely proportional to the initial amplitude is consistent with parallel theories of modulation instabilities.[8,11] Returning to the experiment shown in Figures 4, we note that by m = 400, the deformation of the envelope into a wave of wavelength 100 is clear to the eye although by this step some energy has also been transferred to the sidebands k = 71 and 79 (Figure 5c). By m = 800 (Figure 4e), the envelope has deformed so that in several locations, it is about to exceed the critical threshold. Within fifty more time steps (Figure 4f), the solution becomes rapidly unbounded. Note that the maximum negative peak (Figures 4e, 4f) travels with a speed of almost one consistent with the envelope description discussed in the next section.

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

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

where u is the solution of the full partial difference equation and $\hat{u}$ is the solution of the partial difference equation as reconstructed from the solution of the amplitude equations. The quantity e_1 measures the deviation of the

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

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

measures the rate at which the amplitude of the envelope modulation grows. When this quantity is monitored, a growth rate is obtained which not only agrees in single and double precision, but also agrees with the growth rate obtained from e_1. It seems reasonable to conclude that the mechanism which is responsible for focusing resides in the difference equations and is not an artifact of finite precision arithmetic.

IV. ENVELOPE DESCRIPTION

Since the instability which leads to focusing involves wave numbers in the immediate neighborhood of the primary modes, it is natural to seek an envelope description of the process. We carry out this analysis for the situation in which the energy is initially in the $\pi/2$ and π modes. In (2.6) let the amplitudes a(m,n), b(m,n) be slowly varying functions of both the time and space variable. When substituted into the full partial difference equations (2.1) the following envelope equations for the amplitudes are obtained.

$$[a(m+1,n) - 2a(m,n) + a(m-1,n)] - [a(m,n+1) - 2a(m,n) + a(m,n-1)]$$

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

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

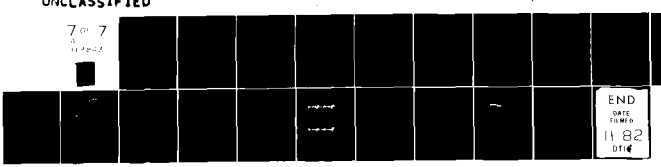
The approximation used in obtaining these equations are valid provided that the spatial gradient $a(m,n+1) - a(m,n-1)$ of $a(m,n)$ is small with respect to the amplitude $a(m,n)$ itself. The advantage of Eqs. (4.1) is that they are universal and will apply to a broad class of partial difference equations.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

In Figure 7, we show the result of comparing $u(m,n)$ as calculated from (2.1) and as constructed from a solution of the envelope Eqs. (4.1). The initial conditions consist of the $\pi/2$ and π modes only with an amplitude $E = .15$ modulated by a long wave perturbation with an amplitude of $.05E$. The parameter values $\alpha = .9$ and $N = 60$ grid points were used. For $m < 200$ (Figures 7a,b) the two computations produce identical results. When $m > 300$ (Figure 7c) the approximations used to derive the envelope equations cease to be valid. For example, a universal term, such as $2a^*(m,n)b(m,n)$, no longer represents $a^*(m,n+1)b(m,n+1) + a^*(m,n-1)b(m,n-1)$, a term which is peculiar to the particular partial difference equation under study. Nevertheless the envelope equations do exhibit the focusing property and display qualitatively similar behavior to the full difference equations even though, in this computation, the critical threshold is reached much sooner (at $m = 650$) by the envelope equations. The full difference equations exhibit focusing behavior which reaches the critical threshold at about $m = 1800$ time steps.

AD-A119 843 CLARKSON COLL OF TECHNOLOGY POTSDAM NY DEPT OF MATHE--ETC F/G 12/1
NONLINEAR WAVES.(U)
SEP 82 M J ABLOWITZ, D J KAUP, A C NEWELL N00014-76-C-0867

UNCLASSIFIED NL

7 of 7



END
DATE FILMED
11 82
DTIC

V. CONCLUSIONS AND CONJECTURES

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

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

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

(iii) the focusing property.

We believe that any difference scheme possessing these properties will be susceptible to instability through this mechanism. Of these three properties, the focusing mechanism is the most difficult to predict. One necessary criterion for focusing is that the envelope equations (4.1) possess n-independent solutions which are always unstable. Although certainly accessible, an analytic result to this effect has not been proved, but, to date, experimental evidence strongly suggests that this instability is always present. A second, more difficult, question is whether the focusing envelope always attains the critical threshold. In order to gain some insight into this question, we plot, in Figures 8 and 9, the number of time steps needed for the critical threshold to be reached (M) as a function of the number of spatial grid points (N). The different curves are parameterized by E, the amplitude of the solution from which the envelope starts to deform. Figure 8 refers to the case of two mode $(\pi/2,\pi)$ initial conditions in which, with $\alpha = .9$, the critical amplitude is $E_c = .36$. Figure 9 refers to the case of three mode $(\pi/3, 2\pi/3, \pi)$ initial conditions in which, with $\alpha = .9$, the critical amplitude is $E_c = .15$. These results also exhibit some interesting features.

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

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

This raises the question of how instability due to focusing can be avoided. In Section 2 an argument was given to show the role of the computational mode in the onset of finite amplitude instability. It appears that the presence of spurious, neutral modes also contributes to the focusing mechanism. A number of strategies have been developed to eliminate the computational mode from calculations that use non-dissipative schemes. Among these strategies we tested the following and reached some conclusions.

(i) Averaging the solution on two consecutive time levels at regular

intervals effectively eliminates the development of envelope instability and keeps the solution intact for any number of time steps. However, averaging amounts to a non-physical "time" step and, not surprisingly, the conservation of quantities such as M and E is badly violated.

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

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

critical threshold has been reduced to about 75% of its value.

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

It would be useful to derive a rough prescription for the frequency with which forward steps should be inserted. There are some assumptions in such a calculation which may mitigate its usefulness as a general result, but it does show that damping and envelope growth can be made to compensate each other in an effective way. The growth rate of the focusing instability can be estimated either from Figures 6 or from the flat portions (N > 200) of the curves in Figure 8. At the same time a linear analysis gives the amount of damping associated with one Matsuno step. For example, the most highly damped mode ($\pi/2$ mode) is damped by a factor of .92 when $\alpha = .9$. Assuming a constant exponential growth of the envelope, it is possible to determine how often a forward step should be inserted to exactly cancel the growth of the envelope. With N = 300, such a calculation yields a frequency of 400 time steps which agrees well with empirically determined strategies. This calculation is certainly oversimplified. The growth rate is not uniform and is somewhat amplitude dependent. Furthermore, the linear estimate of the damping factor is not exactly correct, especially in the later stages of the computation. Nevertheless, the

the argument does explain qualitatively the success of a forward step in inhibiting the instability.

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

The qualitative similarity between all of the features reported above and the properties of focusing envelopes of partial differential equations leads us to conjecture that the strength of the focusing mechanism increases with the dimension of the problem. Indeed, preliminary calculations with analogous two-dimensional equations have borne out this conclusion. We expect that this two step instability process will be potentially present in all large scale computations. Our goal, in this and future work, is to understand the nature of the breakdown of numerical algorithms to the point that we can (a) appreciate why certain ad hoc instability inhibitors (such as filtering and the introduction of artificial viscosity) work and (b) to devise new and more enlightened ways to control instabilities without sacrificing accuracy.

ACKNOWLEDGEMENT

This research was supported in part by the National Science Foundation, Grants PHY77-27084 and MCS75-07548 A01. One of us (ACN) is very appreciative of the hospitality and scientific atmosphere he enjoyed while visiting the Institute for Theoretical Physics at the University of California, Santa Barbara. The work was also supported in part by U.S. Army Contract DAAG29-78-C-0059 and ONR Contract N00014-76-C-0867. The authors are also particularly grateful to David Rand whose healthy scepticism and insightful comments encouraged us to examine Figure 1 in greater detail.

REFERENCES

[1] A. Arakawa and V.R. Lamb, "Computational design of the basic dynamical process of the UCLA general circulation model", Methods in Computational Physics 173-265, Academic Press (1977).

[2] B. Fornberg, "On the instability of leapfrog and Crank-Nicloson approximations of a nonlinear partial differential equation, Math. of Comp., 27, 45 (1973).

[3] R.H.G. Helleman, "Self-generated chaotic behavior in nonlinear mechanics", Fundamental Problems in Statistical Mechanics, Vol. 5, ed. E.G.D. Cohen, North-Holland, 165 (1980).

[4] H.O. Kreiss and J. Oliger, "Comparison of accurate methods for the integration of hyperbolic equations", Tellus, 24, 199 (1972).

[5] H.O. Kriss and J. Oliger, "Methods for the approximate solution of time dependent problems", GARP Publication No. 10 (1973).

[6] F. Mesinger and A. Arakawa, "Numerical methods used in atmospheric models", GARP Publication Series 17, Volume 1.

[7] A.C. Newell, "Bifurcation and nonlinear focusing, pattern formation and pattern recognition", ed. H. Haken, Springer-Verlag, 244 (1979).

[8] A.C. Newell, "Finite amplitude instabilities of partial difference equations", SIAM App. Math., 33, 133 (1977).

[9] N.A. Phillips, "An example of nonlinear computational instability", The Atmosphere and the Sea in Motion, ed. B. Bolin, Rockefeller Institute, New York, 501, (1959).

[10] Y.S. Sigov and V.E. Zakharov, "Strong turbulence and its computer simulation", Journal de Physique, 40 (1979).

[11] G.B. Whitham, "Linear and nonlinear waves", Wiley Interscience (1974).

FIGURE CAPTIONS

1a. Stability curves in (α,E) plane as determined from amplitude equations: 2. Two mode solution $(\pi/2, \pi)$. 3. Three mode solution $(\pi/3, 2\pi/3, \pi)$. 4. Four mode solution $(\pi/4, \pi/2, 3\pi/4, \pi)$.

1b. Enlargement of (α,E) parameter plane for two mode solution. · • Stable solution to 2×10^4 time steps. X Unstable solution.

2a-f. Solution to the partial difference equations: 3 modes, N=300, α=.9, E=.1, m=400, 1000, 2000, 2200, 2400, 2680.

3a-d. Spectral resolution of u(1000,n), u(2000,n), u(2200,n), u(2400,n), of Figure 2.

4a-f. Solution to partial difference equations: 2 modes, N=300, α=.9, E=.15, m=100, m=400, m=450, m=500, m=800, m=850.

5a-f. Spectral resolution of solutions of Figure 4 at m=0, 100, 400, 450, 500, 550.

6a,b. Growth rate curves. Single and double precision, N=240.

7a-c. Solution of partial difference equations (·) and envelope equations (*). 2 modes, N=60, α=.9, E=.15, m=100, 200, 300.

8. Number of grid points (N) vs. number of time steps to critical threshold (M): 2 modes, α=.9, various E=(1).3, (2).27, (3).21, (4).18, (5).15.

9. Number of grid points (N) vs. number of time steps to critical threshold (M): 3 modes, α=.9, various E=(1).14, (2).12, (3).1, (4).09, (5).106, (6).05.

10. $\ell n(\ell nN)$ vs. $\ell n(E)$, α=.9, 3 modes for M = 2000, 3000, 6000.

11a,b. The effect of one Matsuno step. N=300, α=.9, m=853, 855.

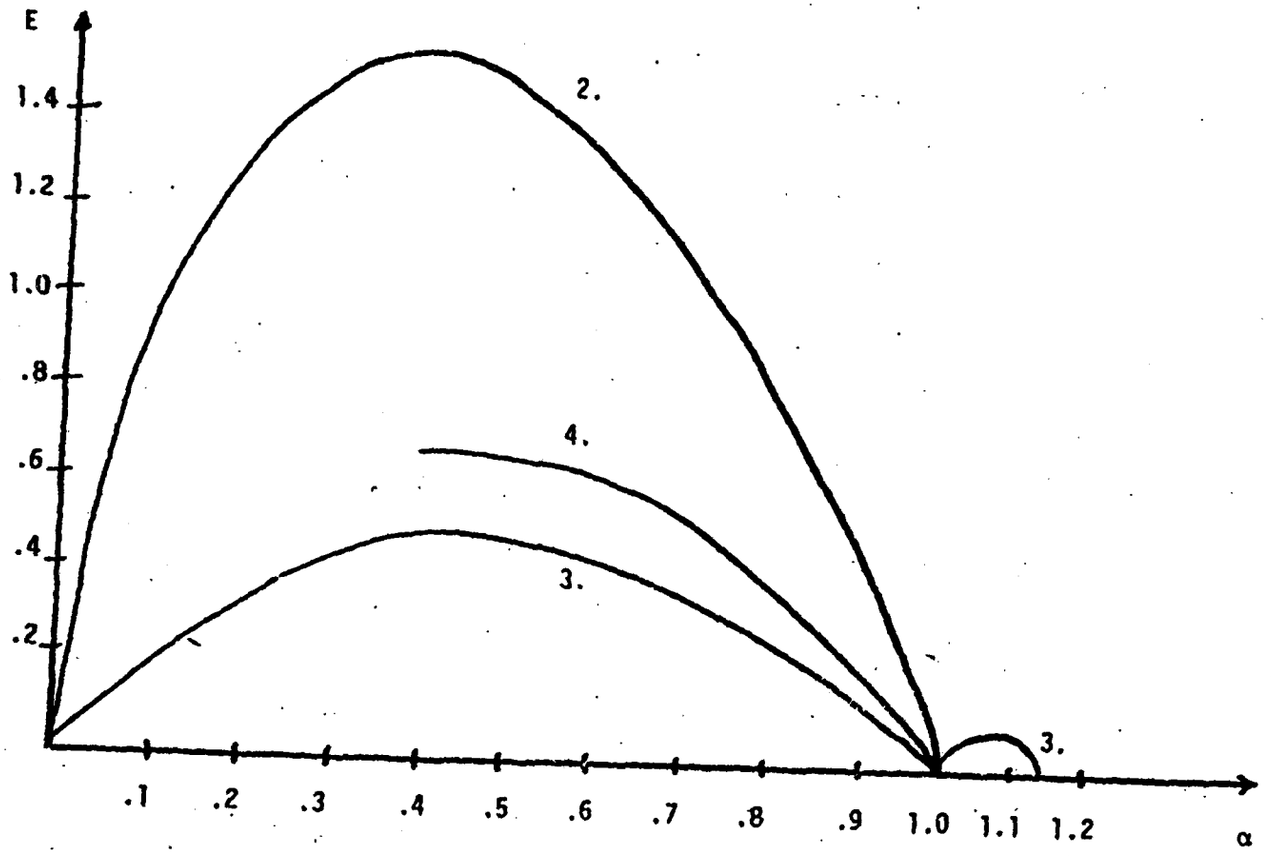


Figure 1a. Stability curves in (α, E) plane as determined from amplitude equations:

2. Two mode solution ($\pi/2$, π).
3. Three mode solution ($\pi/3$, $2\pi/3$, π).
4. Four mode solution ($\pi/4$, $\pi/2$, $3\pi/4$, π).

All modes with equal initial weight.

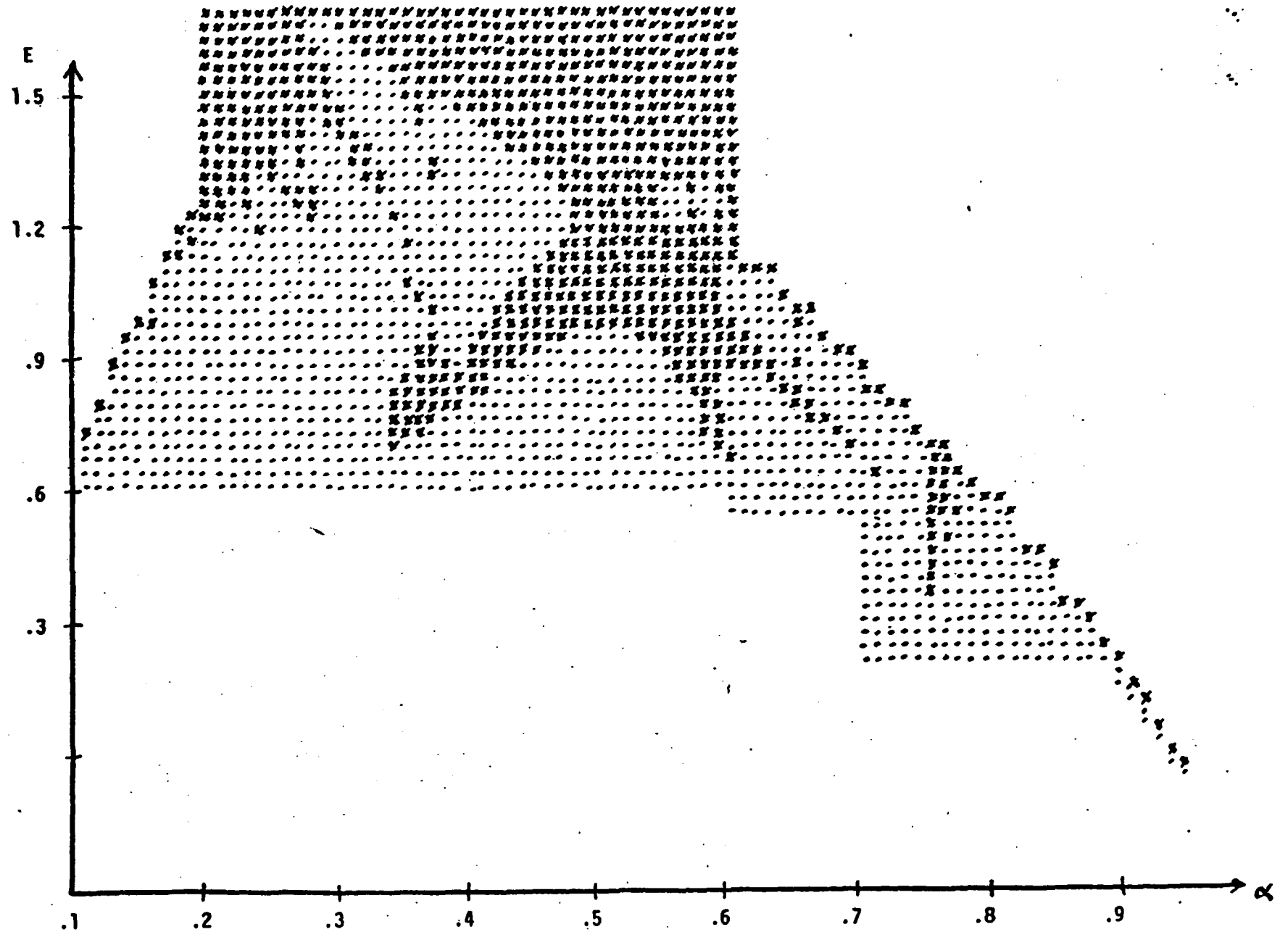


Figure 1b. Enlargement of (α, E) parameter plane for two mode solution.
• Stable solution to 2 x 10^4 time steps. X Unstable solution.

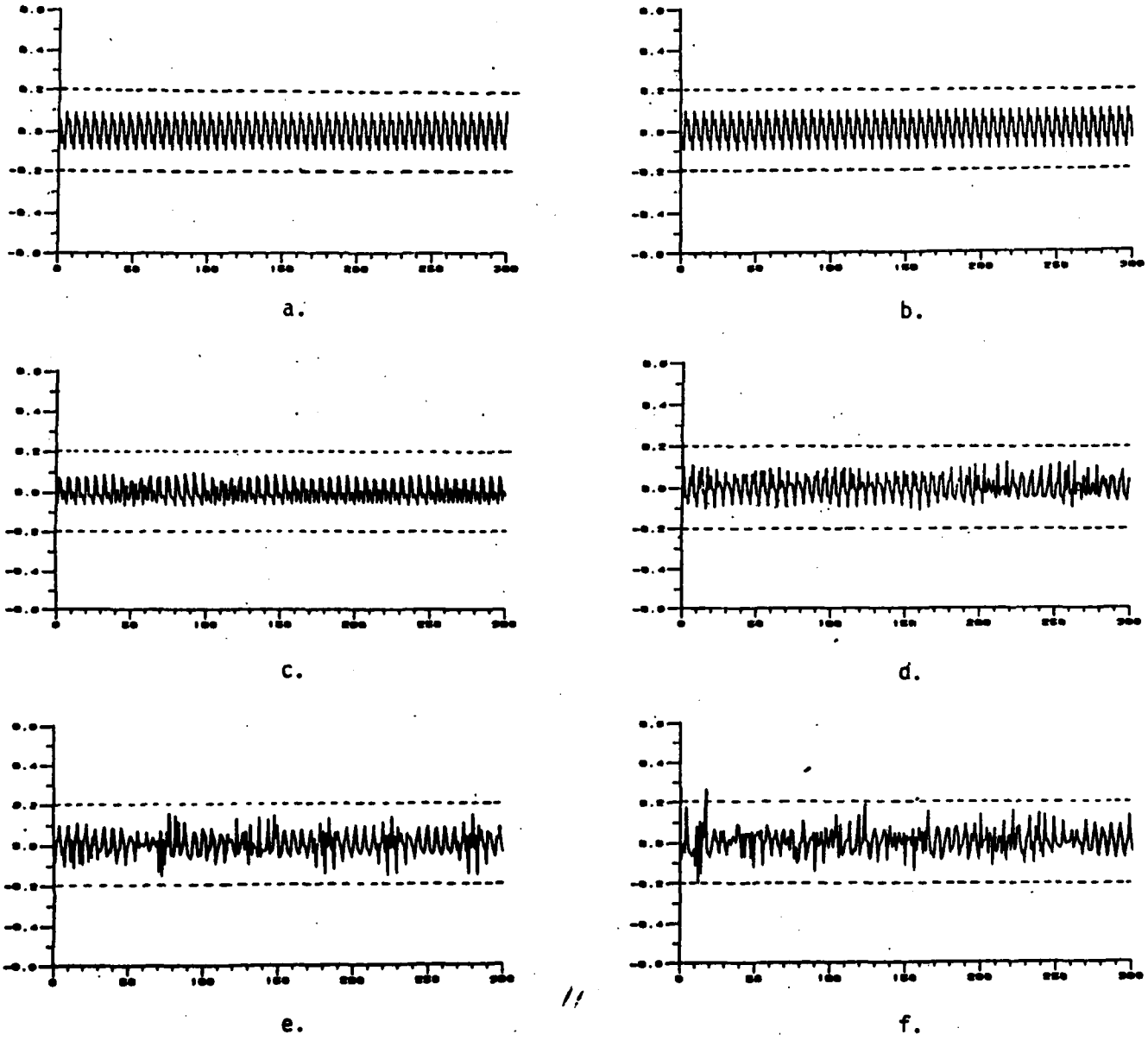


Figure 2. Solution to partial difference equations
3 Modes, N=300, α=.9, E=.1 .
(a) m=400 (b) m=1000 (c) m=2000
(d) m=2200 (e) m=2400 (f) m=2680.

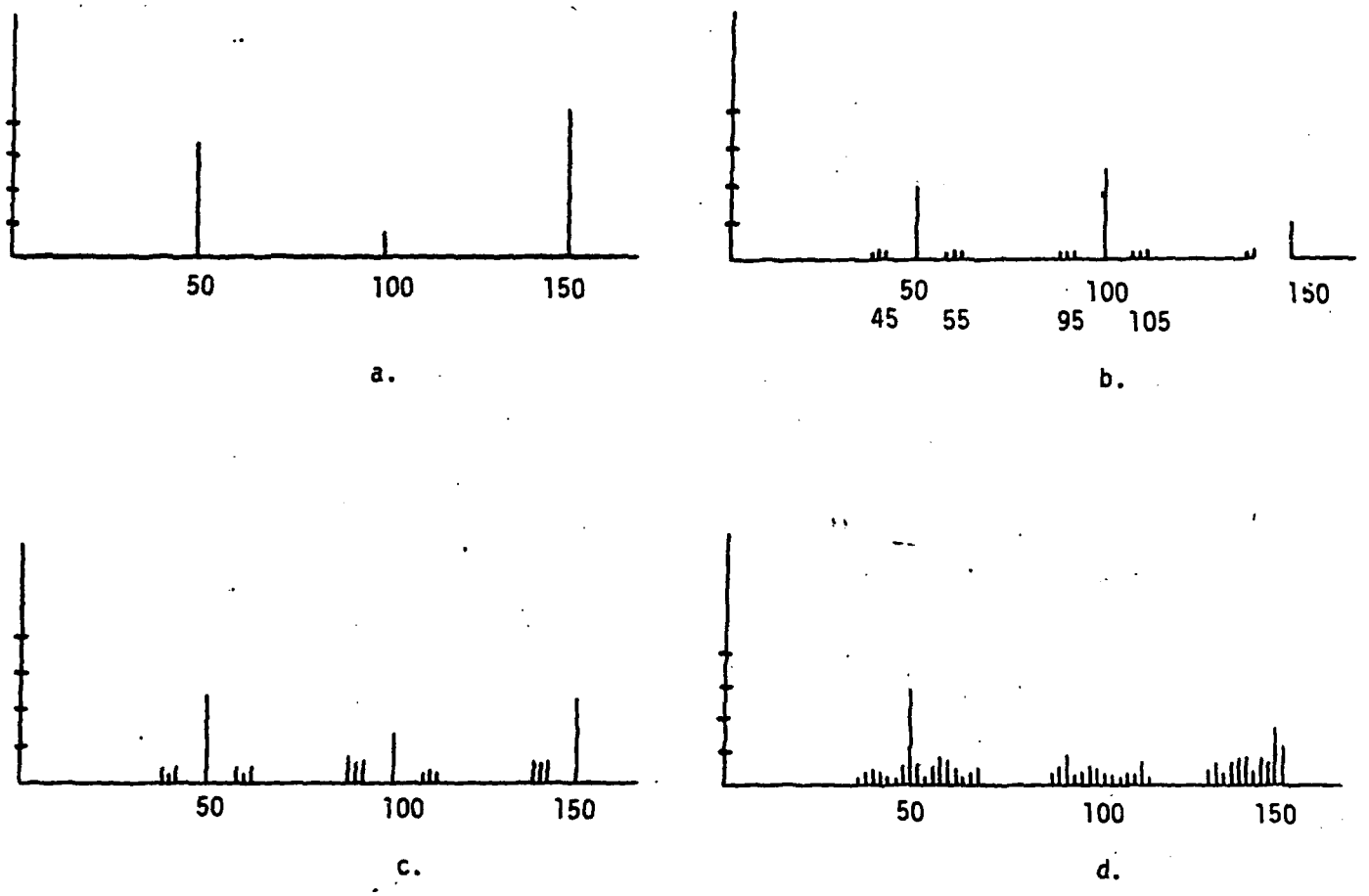


Figure 3. Spectral resolution of solution of Figure 2.

(a) m=1000 (b) m=2000 (c) m=2200 (d) m=2400

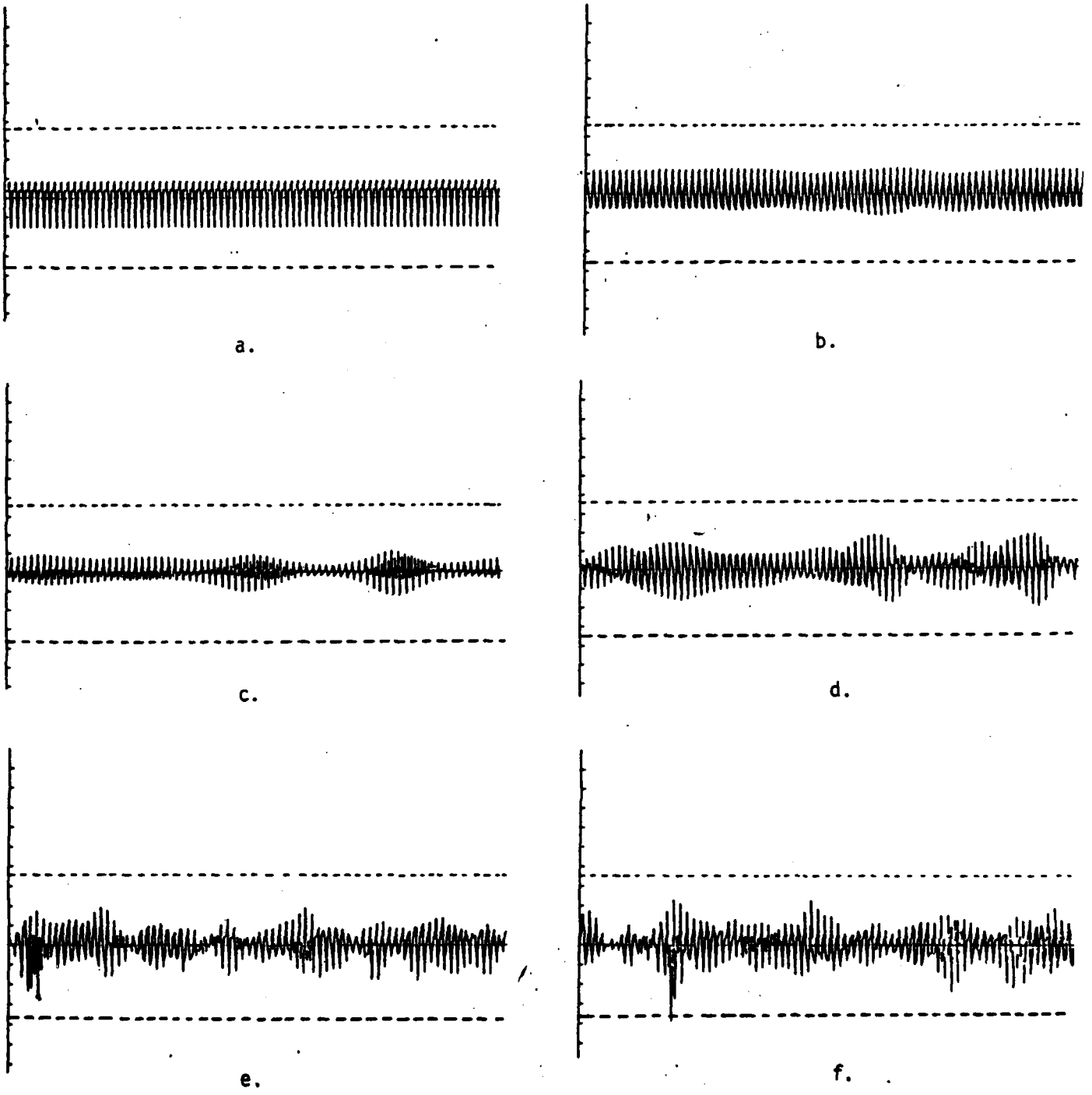


Figure 4. Solution of partial difference equations
2 Modes, N=300, α=.9, E=.15
(a) m=100 (b) m=400 (c) m=450 (d) m=500 (e) m=800 (f) m=850

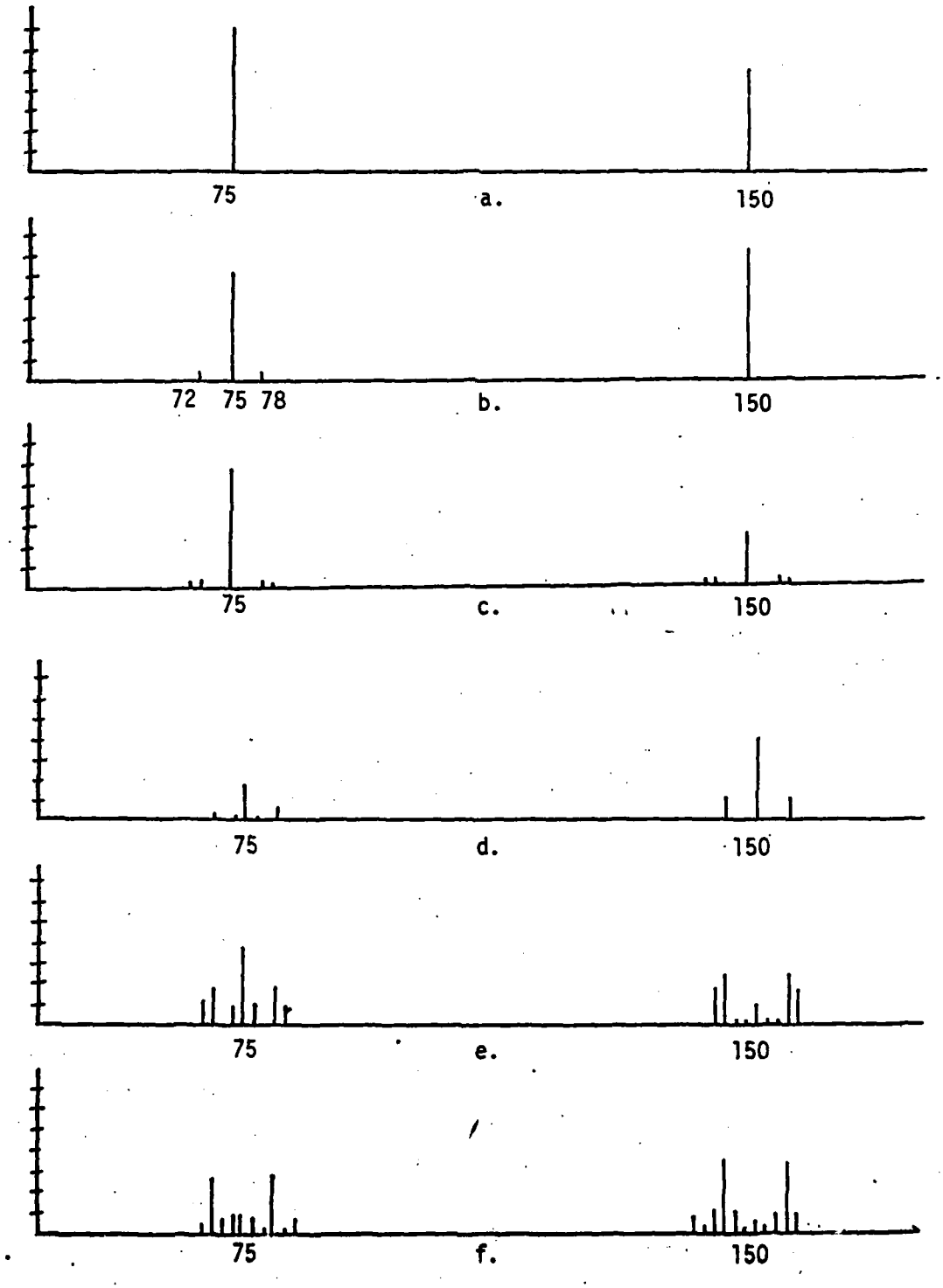


Figure 5. Spectral resolution of solution of Figure 4.
(a) m=0 (b) m=100 (c) m=400 (d) m=450 (e) m=500 (f) m=550

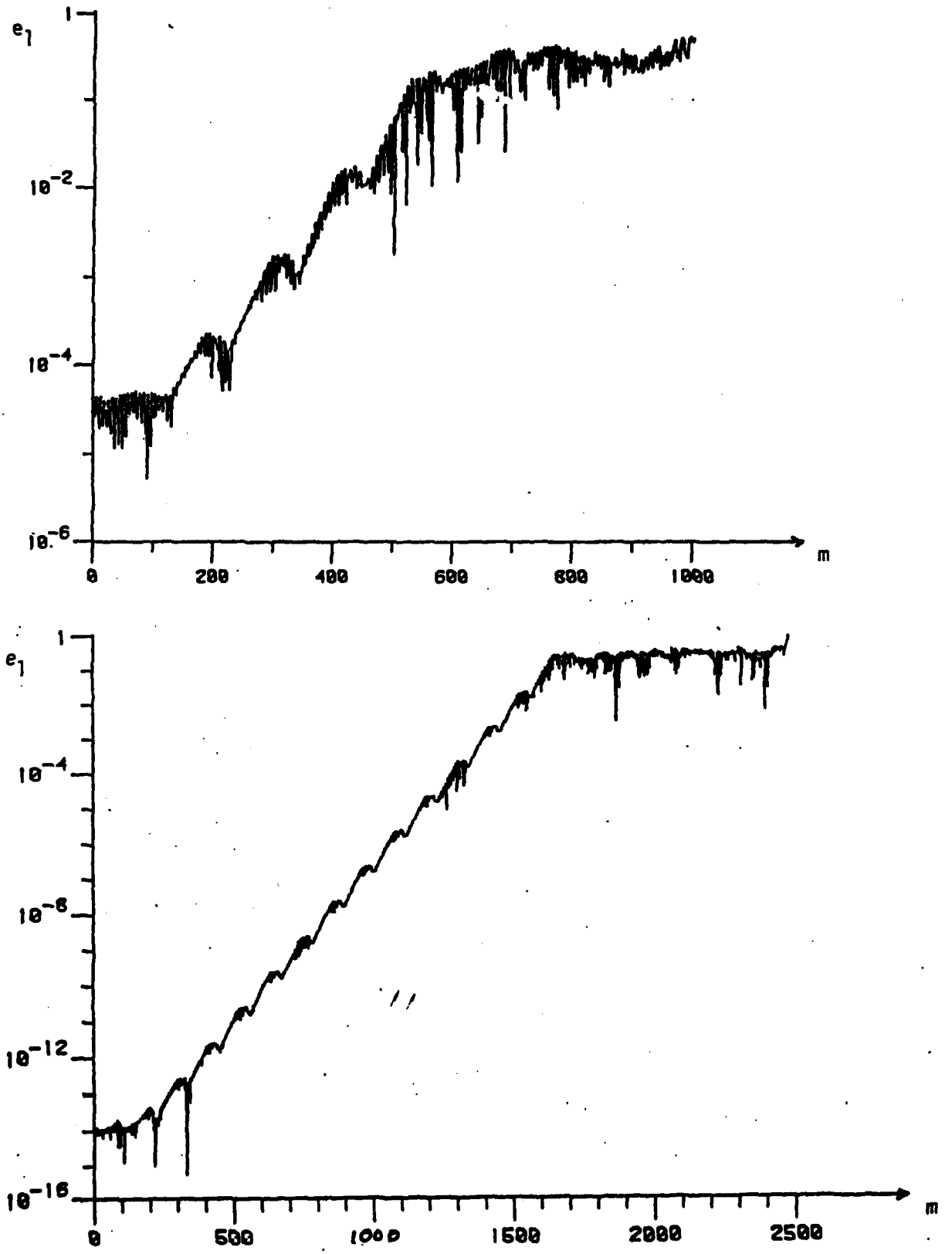


Figure 6. Growth rate curves, e_1 vs. m (as defined in text). N=240.
Upper graph-single precision. Lower graph-double precision.

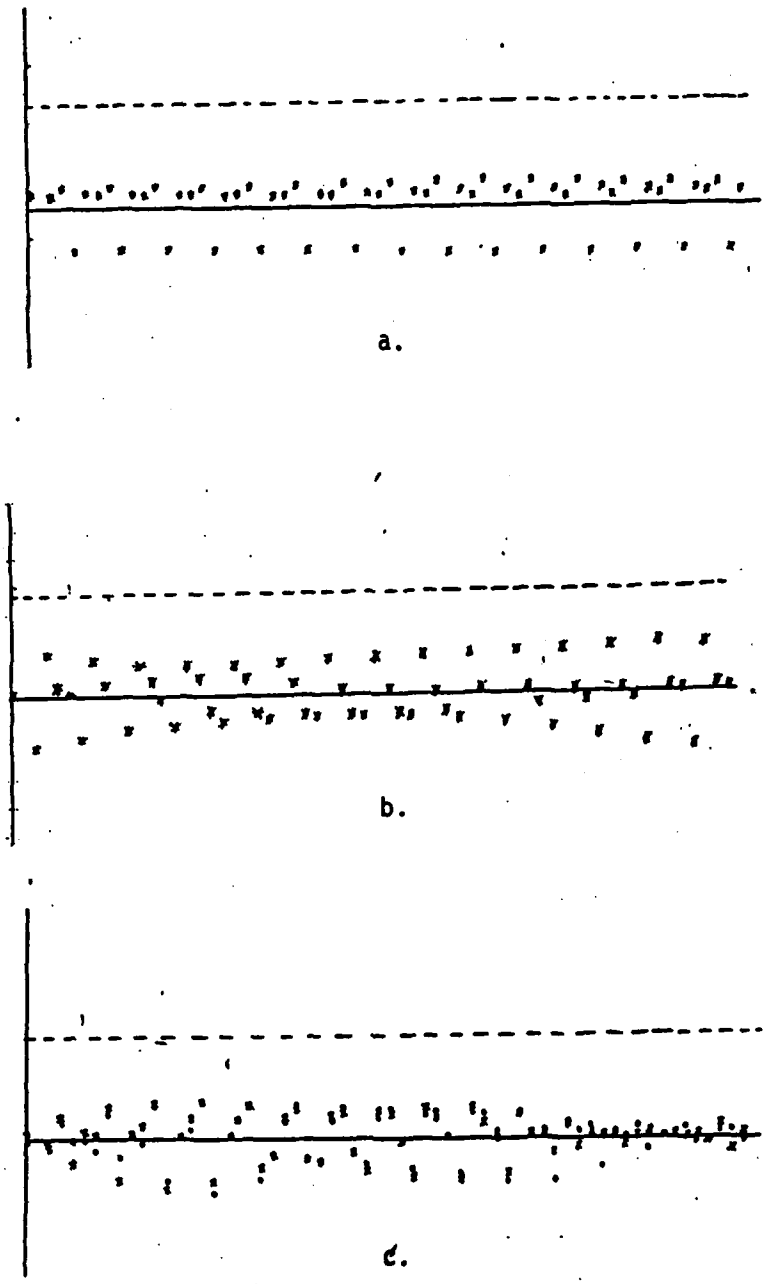


Figure 7. Solution of partial difference equations (·) and envelope equations (*). 2 Modes, N = 60, α=.9, E=.15 (a) m=100 (b) m=200 (c) m=300.

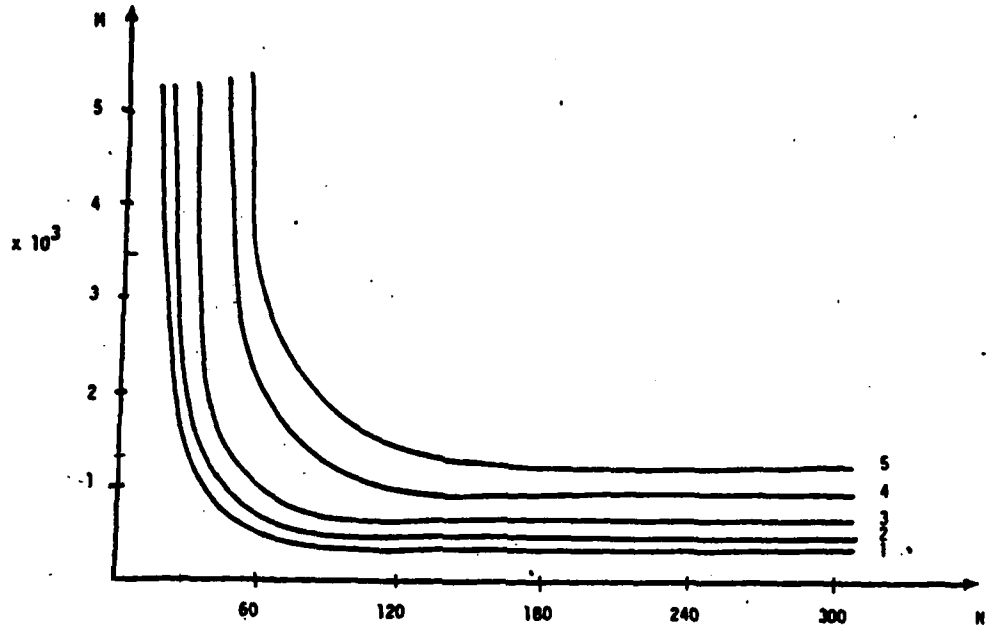


Figure 8. Number of grid points (N) vs. number of time steps to critical threshold (M). 2 Modes, α=.9 (1) E=.3 (2) E=.27 (3) E=.21 (4) E=.18 (5) E=.15.

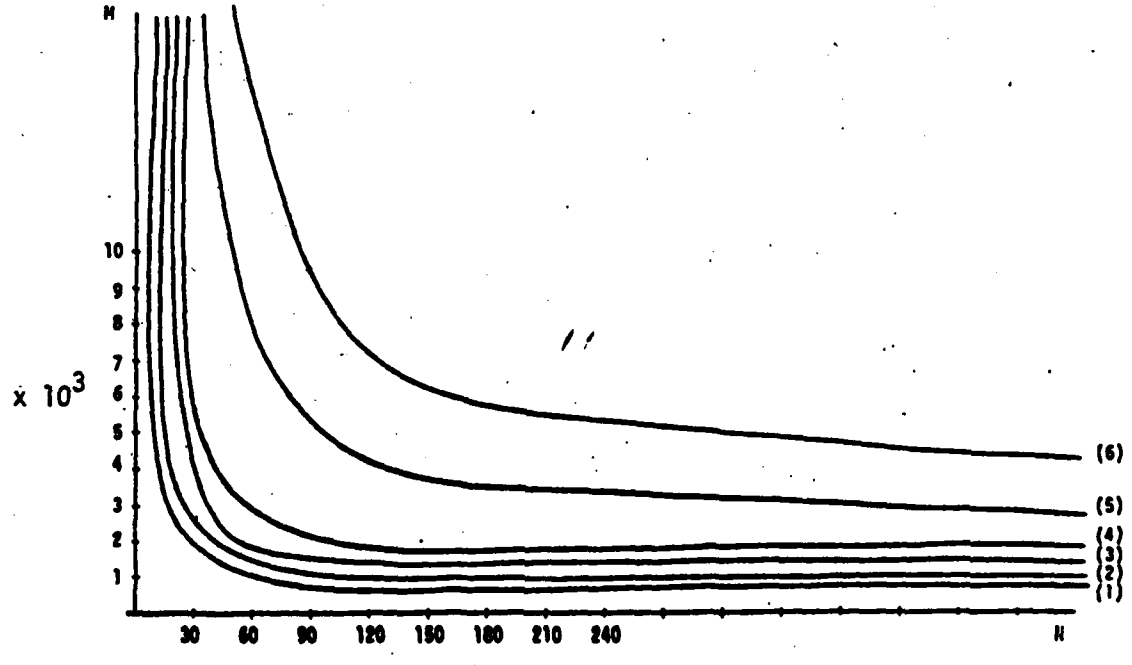


Figure 9. Number of grid points (N) vs. number of time steps to critical threshold (M). 3 Modes, α=.9 (1) E=.14 (2) E=.12 (3) E=.1 (4) E=.09 (5) E=.06 (6) E=.05.

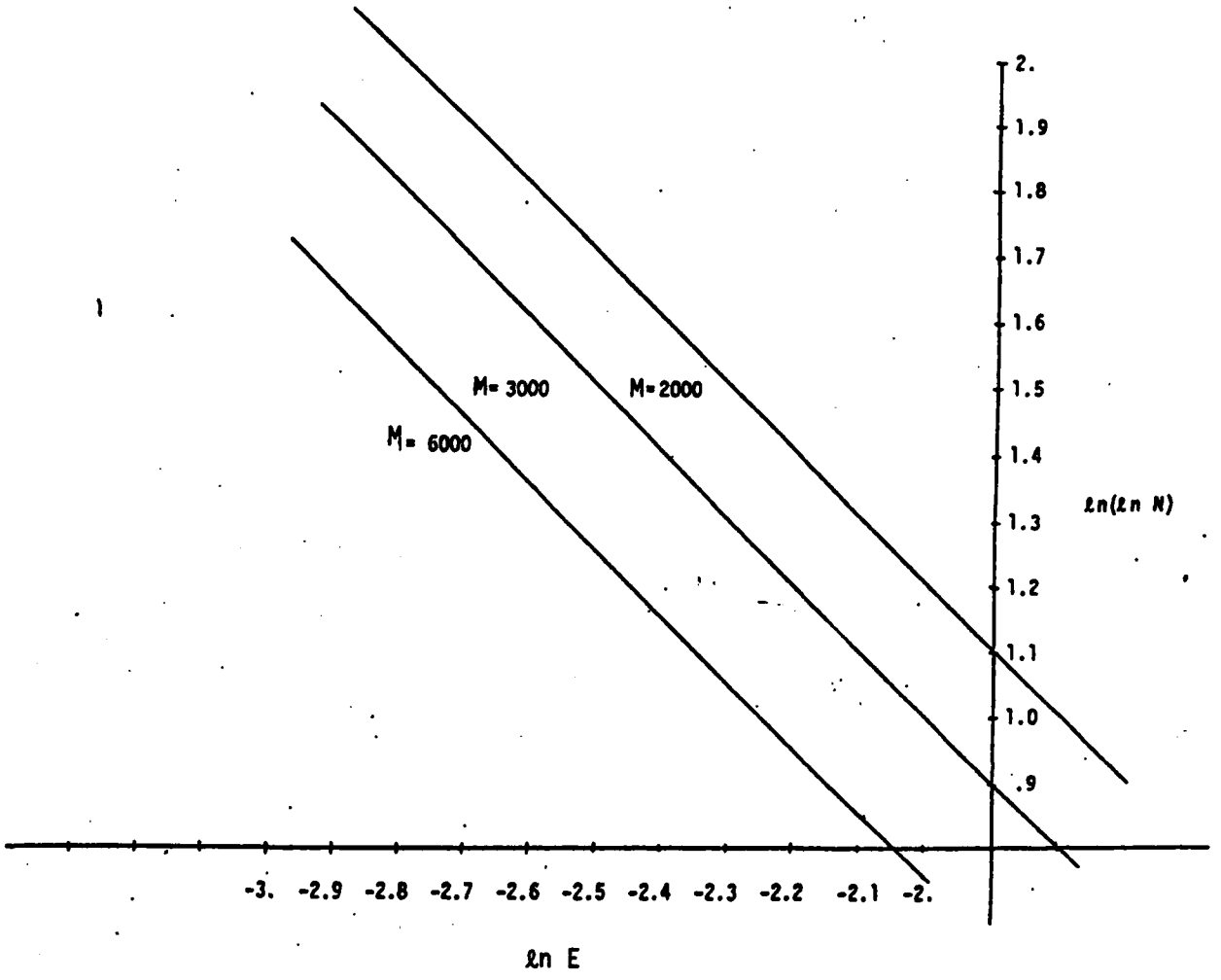


Figure 10. ℓn(ℓnN) vs. ℓnE for 3 Modes, α=.9 with M = 2000, 3000, 6000.

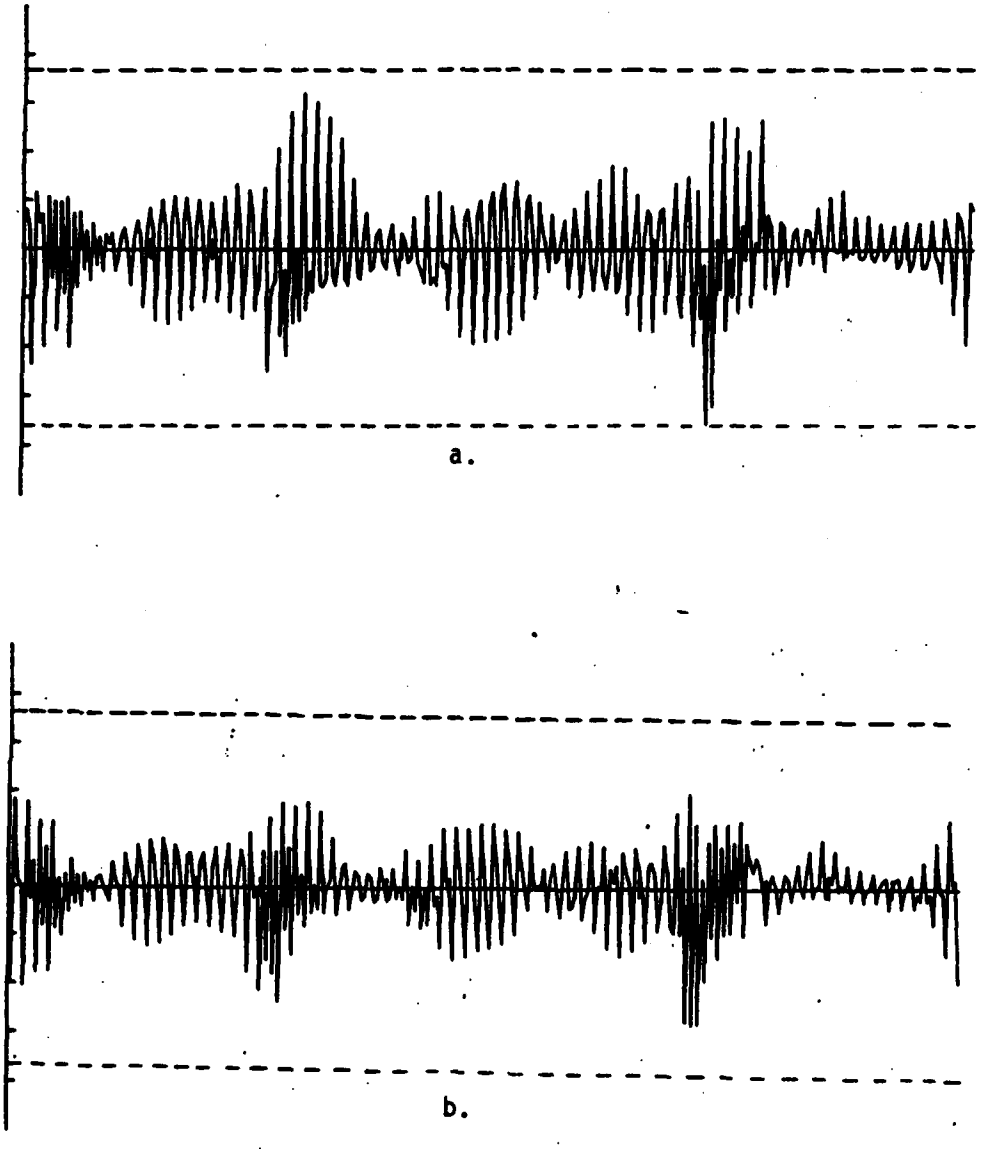


Figure 11. The effect of one Matsuno step. N=300, α=.9

(a) m=853 (b) m=855

4.1

Volume 87A, number 1,2 PHYSICS LETTERS 21 December 1981

CHAOS IN THE INHOMOGENEOUSLY DRIVEN SINE-GORDON EQUATION

J.C. EILBECK
Department of Mathematics, Heriot-Watt University, Riccarton, Currie, Edinburgh EH14 4AS, UK

P.S. LOMDAHL
Laboratory of Applied Mathematical Physics, The Technical University of Denmark, DK-2800 Lyngby, Denmark

and

A.C. NEWELL
Program in Applied Mathematics, University of Arizona, Tuscon, AZ 85721, USA

Received 22 October 1981

Some numerical studies are reported of solutions of the sine-Gordon equation with damping and a mixture of two types of driving force: a constant (dc) force and a spatially inhomogeneous force varying harmonically (ac) in time. For some ranges of parameters the excitation of kink–antikink pairs is chaotic.

Recently there has been a considerable interest in the occurrence of chaotic states in deterministic systems (c.f. ref. [1]). In particular, the appearance of unusually high noise-temperatures in Josephson parametric amplifiers has been ascribed to transitions from coherent to chaotic states of the underlying dynamical system

$$\phi'' + \alpha\phi' + \sin\phi = \eta_{ac}\cos(\omega_{ac}t). \tag{1}$$

This is also the equation of motion for a damped pendulum with an applied torque varying harmonically in time. Analog computer simulations by Huberman et al. [2], and digital computer calculations and electronic experiments by Pedersen and Davidson [3], have shown that eq. (1) exhibits chaotic behaviour for certain ranges of parameter values.

In this letter we report a preliminary investigation of the effect of adding a spatial dependence of ϕ to eq. (1), i.e. instead of a Josephson point-diode or a single pendulum, we consider a long Josephson junction or an array of coupled pendulums. The equation of motion in suitably normalized units is

$$\phi_{tt} - \phi_{xx} + \alpha\phi_t + \sin\phi = F(x,t), \tag{2}$$

where $F(x,t)$ represents a general driving term.

Without damping ($\alpha = 0$) and driving terms, eq. (2) is the well known sine-Gordon (SG) equation (c.f. ref. [4]). The fundamental nonlinear normal modes of the SG equation are the kink/antikink ($\phi_\pm$) and breather or bion (ϕ_b) solutions

$$\phi_\pm = 4\tan^{-1}\{\exp[\pm\gamma(x - x_0 - ut)]\} \tag{3}$$

and

$$\phi_b = 4\tan^{-1}[(1/\omega_b^2 - 1)^{1/2}\sin\theta_I \operatorname{sech}\theta_R], \tag{4}$$

where $\theta_I = \gamma\omega_b[t - u(x - x_0)]$, $\theta_R = \gamma(1-\omega_b)^{1/2} \times (x - x_0 - ut)$, and $\gamma = (1-u^2)^{-1/2}$, u being the velocity. The parameter ω_b determines the internal oscillation frequency of the breather. The energy carried by the kink/antikink and the breather are

$$E_\pm = 8\gamma, \quad E_b = 2E_\pm(1-\omega_b^2)^{1/2}, \tag{5}$$

respectively. These solutions contrast with the small-amplitude harmonic (linear-phonon or plasmon) solutions $\phi(x,t) \propto \mathrm{Re}\{\exp[i(kx - \omega_k t)]\}$ with the dispersion relation $\omega_k^2 = 1 + k^2$. It is well known that the breather can be considered as a bound state kink–antikink pair, where the energy goes to zero as $\omega_b \to 1$ and

0 031-9163/81/0000–0000/$ 02.75 © 1981 North-Holland

approaches $2E_{\pm}$ as $\omega_b \to 0$. For $\omega_b \to 1$ (small amplitude) the breather can be viewed as a coherent spatially confined excitation of linear plasmons. Although there is no clear distinction between a linear localized plasmon and a nonlinear breather, we shall below use the term plasmon in a loose sense to mean a low energy breather.

It is important to note that there is no threshold energy required for breather creation, in contrast to the creation of a kink or antikink. This suggests that a relatively small spatially inhomogeneous driving force tuned to unit frequency will be successful in exciting a low energy breather mode. An obvious conjecture is that such a driving force will continue to feed energy into the breather until it eventually breaks up into a kink–antikink pair. Numerical experiments support this hypothesis, and also suggest that the time evolution of this scenario is chaotic in nature.

We used the finite difference scheme of Ablowitz et al. [5] to integrate eq. (2), with step size $\Delta x = \Delta t = 0.05$. As initial conditions we took $\phi(x, 0) = 0$ and $\phi_t(x, 0) = 0$ throughout. At $x = 0, 40$ we used linear outflow boundary conditions, $\phi_x \pm \phi_t = 0$. In order to investigate the chaotic nature (or otherwise) of the results, we calculated a power spectrum obtained as the autocorrelation function of $\phi_t(20, t)$ computed by the fast fourier transform using 2^{14} values of $\phi_t(20, t)$. The time series $\phi_t(20, t)$ was shaped with a cosine bell window in order to reduce frequency components introduced by truncation.

Initially we used the following driving force

$$\begin{aligned} F(x, t) &= 0, && 0 \leqslant x \leqslant 15, \\ &= \eta_{ac} \cos t, && 15 < x < 25, \\ &= 0, && 25 \leqslant x \leqslant 40. \end{aligned} \tag{6}$$

In fig. 1 we show $\phi(x, t)$ for the case $\eta_{ac} = 0.5$, and $\alpha = 0.01$ for $0 < t < 130$. Initially plasmons are created, but at $t \approx 100$ a breather is formed which eventually aquires enough energy to break up into a kink–antikink pair. Further calculations, not shown in the diagram, reveal that this process continues in an aperiodic manner. In fig. 2 we show the power spectrum $S(\omega)$ obtained for the time interval $50 < t < 5000$. A broad band noise level is seen, characterizing the cha-

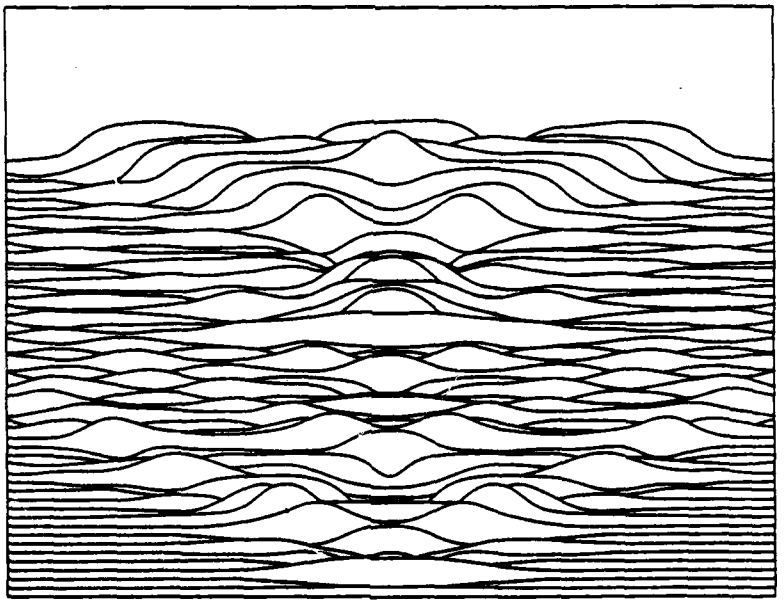


Fig. 1. Numerical solution of eq. (2) and eq. (6) with $\eta_{ac} = 0.5$, $\alpha = 0.01$, shown at discrete times, $0 < t < 130$, time running upwards.

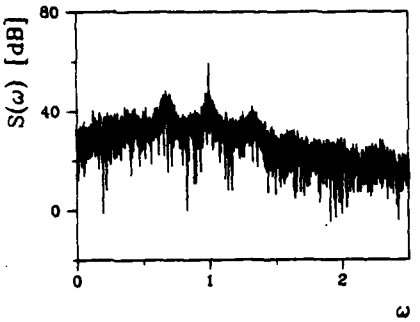


Fig. 2. Power spectrum $S(\omega)$, for the case $\eta_{ac} = 0.5$, $\alpha = 0.01$, showing the 1956 lowest frequencies.

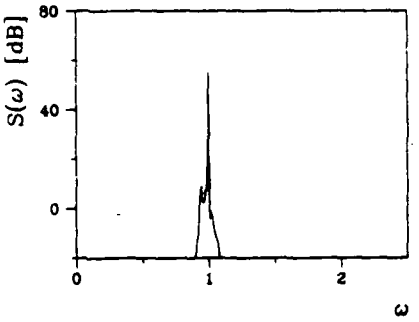


Fig. 3. Power spectrum $S(\omega)$, for the case $\eta_{ac} = 0.01$, $\alpha = 0.01$, showing the 1956 lowest frequencies.

otic nature of the process. For comparison, the power *spectrum of a nonchaotic* case obtained for $\eta_{ac} = 0.01$, and $\alpha = 0.01$ is shown in fig. 3. We find no creation of high energy breathers in this case, and only a slight broadening (60 dB below) around the driving frequency is observed. For $\eta_{ac} \approx 0.1$ the background noise begins to rise.

A constant (in space and time) applied torque on the pendulum chain will force a breather to break up into a kink–antikink pair. The final velocity of the kink and antikink is determined as a balance between the loss and the driving force. Thus we expect the addition of a constant force in eq. (6) to enhance the breakup of breathers and to accelerate the movement of the kink–antikink pairs from the central driven region. We carried out calculations with the following form

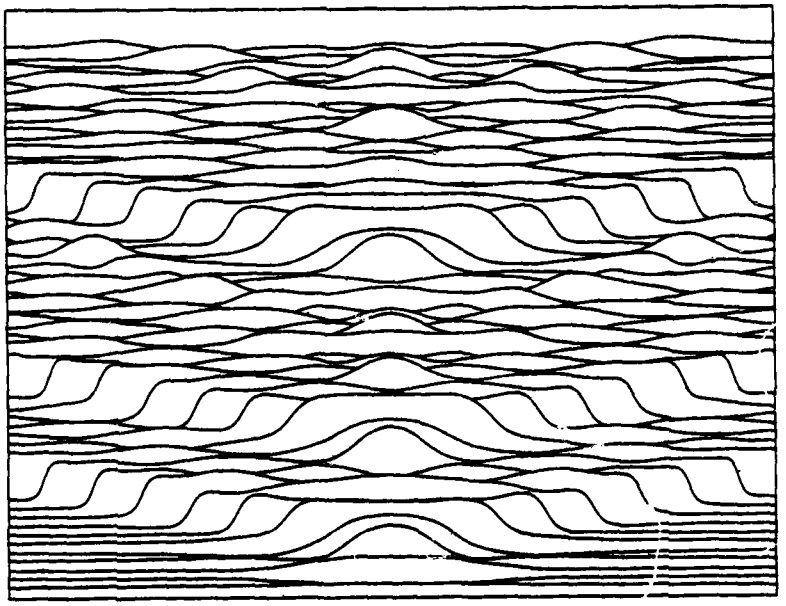


Fig. 4. Numerical solution of eq. (2) and eq. (7) with $\eta_{dc} = 0.3$, $\eta_{ac} = 0.5$, $\alpha = 0.01$, shown at discrete times, $0 < t < 130$, time running upwards.

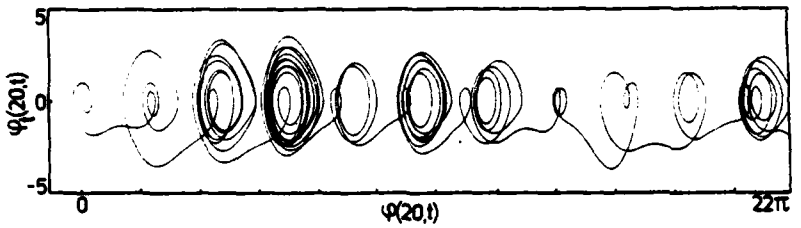


Fig. 5. Plot of $\phi_t(20, t)$ versus $\phi(20, t)$ for the case $\eta_{dc} = 0.3$, $\eta_{ac} = 0.5$, and $\alpha = 0.01$ for $0 \leq t \leq 310$.

of $F(x, t)$

$$F(x,t) = \eta_{dc} \qquad 0 \leqslant x \leqslant 15,$$
$$= \eta_{dc} + \eta_{ac} \cos t, \quad 15 < x < 25, \qquad (7)$$
$$= \eta_{dc}, \qquad 25 \leqslant x \leqslant 40.$$

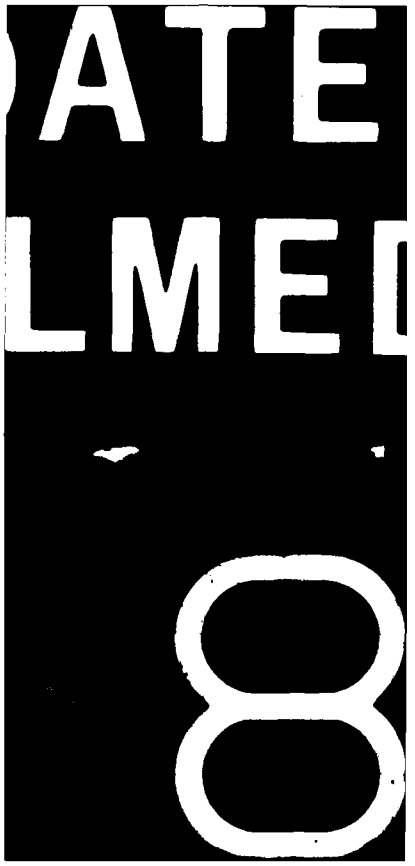
The result for $\eta_{dc} = 0.3$, $\eta_{ac} = 0.5$, and $\alpha = 0.01$ is shown in fig. 4 for $0 < t < 130$. As before, we find plasmons being created, but now three breather and kink–antikink pairs are found, supporting the enhancement conjecture above. The power spectrum is qualitatively similar to that of fig. 2. In fig. 5 we show $\phi_t(20, t)$ versus $\phi(20, t)$ for $0 < t < 310$. The chaotic nature of the process is clearly seen – the creation of plasmons is accompanied by successive 2π jumps in a random way.

In conclusion, we have shown that kink–antikink pairs of the damped SG equation are created in a chaotic way by the introduction of a spatially inhomogeneous time-harmonic driving term. Including further a constant driving term tends to accelerate the process. Because the excitation energy of breathers is near zero, we find the transition to chaos for a lower magnitude of driving force than in the case of a spatially independent model such as eq. (1). However we have not yet attempted a detailed study of the parameter space such as that carried out in [4,5].

This work was started at the Schloss Elmau meeting in 1981, and two of us (JCE and ACN) are grateful to Professor Dr. H. Haken for his hospitality, and to Professor B.A. Huberman for interesting discussions. Further work was done while PSL was at Heriot-Watt University during a visit supported by the Danish Council for Scientific and Technical Research. One of us (JCE) is grateful for computational support from the SRC Interactive Computing Facility, and another (ACN) was supported by US Army Contract DAAG29-78-G-0059, ONR Contract N00014-76-C-0867 and NSF Grant MCS75-07548 A01.

References

[1] H. Haken, ed. Chaos and order in nature (Springer, Berlin, 1981).
[2] B.A. Huberman, J.P. Crutchfield and N.H. Packard, Appl. Phys. Lett. 37 (1980) 750.
[3] N.F. Pedersen and A. Davidson, Chaos and noise rise in Josephson junctions, IBM T.J. Watson Research Center Report (7/8/81) submitted for publication.
[4] P.J. Caudrey, J.C. Eilbeck and J.D. Gibbon, Nuovo Cimento 25B (1975) 497.
[5] M.J. Ablowitz, M.D. Kruskal and J.F. Ladik, SIAM J. Appl. Math. 36 (1979) 428.


ATE
LME
8