

CONTEMPORARY MATHEMATICS

301

The Legacy of the Inverse Scattering Transform in Applied Mathematics

Proceedings of an AMS-IMS-SIAM
Joint Summer Research Conference
on the Legacy of Inverse Scattering Transform
in Nonlinear Wave Propagation
June 17–21, 2001
Mount Holyoke College, South Hadley, MA

Jerry Bona
Roy Choudhury
David Kaup
Editors



American Mathematical Society

Licensed to Univ of Arizona. Prepared on Wed Jul 30 10:49:31 EDT 2014 for download from IP 128.196.226.62.

License or copyright restrictions may apply to redistribution; see <http://www.ams.org/publications/ebooks/terms>

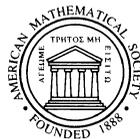
CONTEMPORARY MATHEMATICS

301

The Legacy of the Inverse Scattering Transform in Applied Mathematics

Proceedings of an AMS-IMS-SIAM
Joint Summer Research Conference
on the Legacy of Inverse Scattering Transform
in Nonlinear Wave Propagation
June 17–21, 2001
Mount Holyoke College, South Hadley, MA

Jerry Bona
Roy Choudhury
David Kaup
Editors



American Mathematical Society
Providence, Rhode Island

Editorial Board

Dennis DeTurck, managing editor

Andreas Blass Andy R. Magid Michael Vogelius

This volume contains the proceedings of an AMS-IMS-SIAM Joint Summer Research Conference on The Legacy of Inverse Scattering Transform in Nonlinear Wave Propagation, held at Mount Holyoke College, South Hadley, MA on June 17–21, 2001, with support from the National Science Foundation, grant DMS-9973450.

2000 *Mathematics Subject Classification*. Primary 35Q51, 35Q53, 35Q55, 35Q58, 35A20, 35C05.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Library of Congress Cataloging-in-Publication Data

The legacy of the inverse scattering transform in applied mathematics : proceedings of an AMS-IMS-SIAM joint summer research conference on the legacy of inverse scattering transform in nonlinear wave propagation, Mount Holyoke College, South Hadley, MA, June 17–21, 2001 / Jerry Bona, J. L. Roy Choudhury, David Kaup, editors.

p. cm. — (Contemporary mathematics, ISSN 0271-4132 ; 301)

Includes bibliographical references.

ISBN 0-8218-3161-5 (alk. paper)

1. Inverse scattering transform—Congresses. 2. Nonlinear wave equations—Congresses. I. Bona, J. L. II. Choudhury, Roy, 1956– III. Kaup, David, 1939– IV. Contemporary mathematics (American Mathematical Society) ; v. 301.

QA927 .L395 2002
531'.1133/01515355—dc21

2002027974

Copying and reprinting. Material in this book may be reproduced by any means for educational and scientific purposes without fee or permission with the exception of reproduction by services that collect fees for delivery of documents and provided that the customary acknowledgment of the source is given. This consent does not extend to other kinds of copying for general distribution, for advertising or promotional purposes, or for resale. Requests for permission for commercial use of material should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294, USA. Requests can also be made by e-mail to reprint-permission@ams.org.

Excluded from these provisions is material in articles for which the author holds copyright. In such cases, requests for permission to use or reprint should be addressed directly to the author(s). (Copyright ownership is indicated in the notice in the lower right-hand corner of the first page of each article.)

© 2002 by the American Mathematical Society. All rights reserved.

The American Mathematical Society retains all rights
except those granted to the United States Government.
Printed in the United States of America.

⊗ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.

Visit the AMS home page at <http://www.ams.org/>

10 9 8 7 6 5 4 3 2 1 07 06 05 04 03 02

Contents

Reviews

The legacy of the IST D. J. KAUP	1
Application of inverse scattering method to problems of differential geometry V. E. ZAKHAROV	15
Algebraic and analytic aspects of soliton type equations V. S. GERDJIKOV	35
Differential forms, spectral theory, and boundary value problems A. S. FOKAS	69
Chaos in partial differential equations Y. LI	93
Multi-soliton complexes N. N. AKHMEDIEV, A. A. SUKHORUKOV, AND A. ANKIEWICZ	117
A unified approach to integrable systems via Painlevé analysis R. CHOUDHURY	139

Articles

Asymptotic stability of solitary waves for nonlinear Schrödinger equations V. S. BUSLAEV AND C. SULEM	163
Finite-time blowup in the additive supercritical stochastic nonlinear Schrödinger equations A. DE BOUARD AND A. DEBUSSCHE	183
Method of symmetry transforms for ideal magnetohydrodynamics equilibrium equations O. I. BOGOYAVLENSKIJ	195
The p -system I: The Riemann problem R. YOUNG	219
Statistical analysis of collision-induced timing shifts in a wavelength-division-multiplexed optical soliton-transmission system G. J. MORROW AND S. CHAKRAVARTY	235

Cuspons and peakons vis-a-vis regular solitons and collapse in a 3-wave system R. GRIMSHAW, G. A. GOTTWALD, AND B. A. MALOMED	249
First integrals and gradient flow for a generalized Darboux-Halphen system S. CHAKRAVARTY AND R. G. HALBURD	273
Blow-ups of the Toda lattices and their intersections with the Bruhat cells L. CASIAN AND Y. KODAMA	283
Superposition principle for oscillatory solutions of integrable systems M. KOVALYOV	311
Scattering at truncated solitons and inverse scattering on the semiline H. STEUDEL	331

Preface

This volume contains the proceedings of the conference on the Legacy of the Inverse Scattering Transform which was held at Mount Holyoke College in Massachusetts from June 17–21, 2001.

Current progress in the area of Solitons and the Inverse Scattering Transform continues to be rapid and new applications are also multiplying, with current nonlinear optical technology moving so rapidly, larger and larger intensities becoming more available, pulse widths becoming smaller and smaller, and relaxation times and damping rates becoming less and less significant. As this limit is approached, the exactly integrable soliton equations, such as 3-wave resonant interactions and second harmonic generation become more and more relevant to experimental applications. Experimental techniques are currently being developed to use these interactions to frequency convert high intensity sources into frequency regimes where there are no lasers. Other experiments are using these interactions to develop intense variable frequency sources, opening up even more new experimental possibilities.

However, in regard to the mathematics of this area, the ‘easy’ problems have been solved long since, and the field has attained a kind of intellectual adolescence. As such, it was felt that it was a good time for taking stock of the current situation, and seeing where the area might go next. We believe that such a reconsideration of the numerous strands of activity which are the legacy of the Inverse Scattering Transform can reveal much about where the field can go in the future, and can even re-energize the field significantly. It will also bring lines of research which currently are somewhat independent of each other closer together, and possibly even open up new avenues of enquiry.

Thus, the conference provided a forum for the more general exposition and assessment of recent developments in Nonlinear Waves and related areas and of their potential applicability in various fields, and this is clearly reflected in the articles in these proceedings. The present volume is thus expected to be of strong interest to experienced and beginning researchers in the Mathematics, Physics, and Engineering communities.

We express our sincere thanks to the American Mathematical Society for their support of the Joint Summer Research Conference Series and the publication of this volume, and to the U.S. Army Research Office for a generous supplemental travel grant that enabled us to cover the travel expenses of all participants. Our greatest single debt of gratitude is undoubtedly to Wayne Drady for his thoroughness and sang-froid at all stages of the conference organization. We knew that we could leave all the practical details in his capable hands and concentrate on enjoying the

meeting. Our grateful thanks also to Chris Thivierge and Gil Poulin for all their help and support on the many steps in the process of putting this volume together.

The Legacy of the IST

David J. Kaup

ABSTRACT. We provide a brief review of some of the major research results arising from the method of the Inverse Scattering Transform.

1. Introduction

I will give a brief review of several items in the Legacy of the Inverse Scattering Transform. In no way is this to be a complete review, since the Legacy has become so vast. However, I will treat those items with which I am most familiar, and try to detail their significance and importance.

There is no doubt that the most important contribution was the famous classical Gardner, Greene, Kruskal and Miura (GGKM) work [1] of 1967 on the KdV equation. This was the starting point. They had found a very strange and new method for solving the initial value problem of a nonlinear evolution equation, the KdV. At that time, and even for several years later, this strange new method was considered to be only a novelty, since it would only work for that one equation, the KdV. Shortly thereafter, as a prelude to what was to follow, Peter Lax [2] showed that if given an appropriate linear operator, L , dependent on a potential, $u(x)$, then one could always construct an infinite sequence of evolution operators, B , each of which would satisfy

$$(1.1) \quad BL - LB = \partial_t L.$$

This sequence of evolution operators could be generated by simply increasing the order of the spatial differentials contained in B . Then from (1.1) one would obtain additional nonlinear evolution equations, each of the form

$$(1.2) \quad \partial_t u = K(u)$$

where K was some (nonlinear) operator. All these additional higher order evolution equations would be solvable by this same technique. This collection is now known as the KdV hierarchy.

1991 *Mathematics Subject Classification*. Primary 01A65; Secondary 35Q51.

Key words and phrases. Solitons, Inverse Scattering Transform.

The author was supported in part by NSF Grant #0129714. The author thanks an anonymous referee for his comments, and also H. Steudel for his comments.

If we consider the eigenvalue problem for L ,

$$(1.3a) \quad L\psi = -\lambda\psi,$$

where λ is the eigenvalue, and append to it the condition

$$(1.3b) \quad \partial_t\psi = B\psi,$$

then it is easy to see that (1.1) is simply the integrability condition for (1.3). Furthermore, as Lax pointed out for the KdV case, (1.1)-(1.3) also implies that the eigenvalues, λ , in (1.3a) would be stationary,

$$(1.4) \quad \partial_t\lambda = 0,$$

a relation that would occur time-and-time again as the study of integrable equations would expand in the decades to follow.

It was also about this time that the term “radiation” was introduced. We haven’t said anything yet about solitons or solitary waves, but more will be said later. For now, let us note that a remarkable feature of the GGKM method of solution was the appearance of fully nonlinear solitary wave solutions, called solitons. The other part of the solution has been called “radiation”, and is essentially linear-like in its behavior. The asymptotics (long-time behavior) of the total solution are generally that the radiation does disperse away, leaving the solitons traveling in a sea of decaying radiation.

As to nomenclature, we shall refer to $(L + \lambda)\psi$ as the eigenvalue problem, ψ as the eigenfunctions, and B as the evolution operation. The pair $[L + \lambda, B]$ is known as the “Lax pair”. For the KdV equation, the Lax pair is

$$(1.5a) \quad (L + \lambda)\psi = (\partial_x^2 + u + \lambda)\psi = 0$$

$$(1.5b) \quad \partial_t\psi = b\psi = (\alpha - 4\partial_x^3 - 6u\partial_x - 3u_x)\psi$$

where α is an arbitrary constant and the integrability condition is the KdV equation:

$$(1.5c) \quad \partial_t u + \partial_x^3 u + 6u\partial_x u = 0.$$

We note here, given L and B , it follows that one can then obtain $K(u)$. However, an important problem is given $K(u)$, construct L and B . The solution of this inverse problem is still an area of active research. One method that sometimes works for this is called “Painlevé Analysis”. For a description of this aspect of the Legacy, the reader is referred to Choudhury’s article in this same issue.

It was not until 1971, that the next physically significant integrable system was uncovered by V. Zakharov and A.B. Shabat (ZS) [3], which was the focusing Nonlinear Schrödinger Equation (NLS)

$$(1.6) \quad iq_t = -q_{xx} - 2q^*q^2.$$

This equation required a different eigenvalue problem,

$$(1.7a) \quad v_{1x} + i\zeta v_1 = qv_2$$

$$(1.7b) \quad v_{2x} - i\zeta v_2 = rv_1$$

where (v_1, v_2) is the eigenvector, ζ is the eigenfunction, and q and r are the “potentials”. For the focusing NLS case, one has $r = -q^*$ and $r = +q^*$ for the defocusing

case. The time evolution operator, B , is given by [3]:

$$(1.8a) \quad i\partial_t v_1 = -i(4\zeta^2 + 2qr)v_1 + (4\zeta q + 2iq_x)v_2$$

$$(1.8b) \quad i\partial_t v_2 = (4\zeta r - 2ir_x)v_1 + i(4\zeta^2 + 2qr)v_2.$$

Exactly as was shown by Lax [2] for the KdV, one also has a hierarchy here, which can be obtained by generalizing (1.8) to higher orders in ζ . In 1972, Wadati [4] found the next member of this hierarchy, the “modified KdV” (mKdV)

$$(1.9) \quad q_t + q_{xxx} + 6q^2 q_x = 0$$

which was also integrable. Its eigenvalue problem was again the ZS eigenvalue problem, (1.7), but where now $r = -q$, and q real. Also, (1.8) had to be generalized to be cubic in ζ .

By this time, it was becoming apparent to many researchers, that this strange method found by GGKM was not simply a novelty. Rather, there was something very significant underlying all of this. This became even more obvious when Ablowitz, Kaup, Newell and Segur (AKNS) presented a method of solution of both the Goursat and Cauchy initial value problems of the sine-Gordon equation [5]. This was also based on the ZS eigenvalue problem, but with a very different form for the B in (1.8): it was now inversely proportional to the spectral parameter, ζ . The sine-Gordon equation was well known at that time. It had a long history, first occurring in 1853 in differential geometry, and was the first equation for which Bäcklund transformations and N-solitons solutions were found. It was known in solid state physics in the 1930’s, and in 1965 had found applications in optics.

The IST solution of the sine-Gordon equation was shortly followed by another letter [6] pointing out how one could generate a large number of integrable equations, each of which were physically significant and important. With one general approach, AKNS were able to reproduce all the Lax pairs found up to that time, and were able to connect the form of the dependence of B on ζ to the linear dispersion relation, $\omega(k)$. (The linear dispersion relation relates how the frequency, ω , depends on the wave vector, k , in the linear limit, where plane waves, $e^{i(kx-\omega t)}$, are the natural solutions.) In 1974, they published their classic AKNS paper [7], wherein they described in detail this new method of solution, calling it the method of the Inverse Scattering Transform (IST). One of the major points of this classic was that the IST could be viewed as a nonlinear extension of the method of the Fourier Transform.

This was also the start of the explosion in research on solitons and integrable systems, because unbeknownst to most westerners, Faddeev, Zakharov and their students were all very busy in the same direction. In the next few years, many important papers were to be published on the IST and related issues.

2. The Legacy

Beginning in 1974, it becomes difficult to try to detail all the results. Nevertheless, we will now discuss in general terms, the legacy which followed from this. In the following, we will list the general areas of the legacy, and briefly describe the importance and the major contributions made to each one.

2.1. Method of Solution — the IST. Above all, the IST is a method of solution for integrable nonlinear equations. It was the pioneering work of GGKM [1], ZS [3] and AKNS [7] which made the most significant impact and set the tone which

followed. Consider what the IST allows one to do. One may take any reasonable initial data, and by means of the direct scattering transform (the eigenvalue problem), transform the initial data into “scattering data”. For the KdV, the scattering data consists of a reflection coefficient $[\rho(k); -\infty < k < \infty]$, bound state eigenvalues $[\kappa_j; j = 1, 2, \dots, N]$ and bound state normalization coefficients $[C_j; j = 1, 2, \dots, N]$ where N is the total number of bound states, usually finite. Now, Lax [2] showed that the eigenvalues, κ_j , would be independent of time, due to (1.1). GGKM showed that if $u(x, t)$ evolved according to the KdV equation, the reflection coefficient, ρ , and the normalization coefficients, C_j , would evolve according to

$$(2.1a) \quad \partial_t \rho(k; t) = \rho(k; 0) e^{8ik^3 t},$$

$$(2.1b) \quad \partial_t C_j(t) = C_j(0) e^{-8\kappa_j^3 t}.$$

Thus it becomes a very simple matter to determine the scattering data at any later time.

Next, one used the solution of the inverse scattering problem, which is the core of this method of solution, to reconstruct the potential(s). One transforms (with the IST) from the scattering data, at time t , back to the potential(s), at time t . For the KdV, the necessary steps are to first construct

$$(2.2) \quad F(z; t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \rho(k; t) e^{ikz} dk + \sum_{j=1}^N C_j(t) e^{-\kappa_j z},$$

then one solves the linear integral equation

$$(2.3) \quad K(x, y) + F(x + y) + \int_x^{\infty} K(x, s) F(s + y) ds = 0,$$

for $K(x, y; t)$. Lastly $u(x, t)$ is constructed from

$$(2.4) \quad u(x, t) = -2 \frac{dK(x, x; t)}{dx}.$$

All integrable systems are solved by an IST of the above format, although there can be a wide variation in the form of the formats. Some eigenvalue problems are higher order and/or even multidimensional. But there is always some scattering problem which maps the potential(s) into a set of scattering data (the Direct Scattering Transform). There is always some evolution of the scattering data, as in (2.1). There is always an Inverse Scattering Transform that allows one to map from the scattering data back to the potential(s), as in (2.2)-(2.4).

One may describe this method of solution as a method for solving the initial value problem of a nonlinear equation by using only linear techniques. Furthermore, one may also say that one solves these nonlinear problems by not solving the nonlinear problem. Instead one solves two related linear problems.

To understand these comments, consider

$$(2.5) \quad L(u, \zeta) \cdot V = 0,$$

where L is the eigenvalue problem in the Lax pair, as in (1.1)-(1.3) and ζ is a spectral parameter. Let the second component of the Lax pair be $B(u, u_x, \dots, \zeta, \partial_x, \dots)$, where

$$(2.6) \quad B \cdot V = \partial_t V.$$

Now note that (2.5) determines the x -dependence of $V(x, t; \zeta)$, whereas (2.6) determines the t -dependence. Thus one function, $V(x, t; \zeta)$, is being determined by two equations. In general, such would overdetermine V , and in order for a mutual solution to exist, certain consistency conditions must be satisfied. This condition is the single integrability condition, (1.2). When (1.2) is satisfied, then a mutual solution exists for (2.3)-(2.6).

But if this integrability condition is nothing more than the nonlinear evolution equation, it therefore follows that if, by some means, we can construct a single-valued solution for $V(x, t; \zeta)$, for some $u(x, t)$, which satisfies each component of the Lax pair, then it follows that $u(x, t)$ must satisfy the nonlinear evolution equation. So, one could say that the entirety of the method of the IST is based on not solving integrable nonlinear equations (at least, not directly). Instead, we solve them indirectly, exactly by the same format used in any transform method. This is as follows. We satisfy (2.5) by mapping $u(x, t)$ into scattering data and constructing the eigenfunctions. We can do this for certain classes of potentials (i.e. $L_1 \cap L_2$). Then we always will have reflection coefficient(s), as a function of the spectral parameter, ζ , and certain bound state data (eigenvalues and normalization coefficients). This is a linear problem. We satisfy (2.6) by requiring the scattering data to evolve appropriately (as in (2.1) for the KdV). This is usually easy to do, since scattering data is typically defined for $x \rightarrow \pm\infty$, where potentials normally approach specified values (usually zero) in that limit. So one really only needs $B(x \rightarrow \pm\infty)$. This is also a linear problem. Therefore, by fixing the scattering data to evolve appropriately, we have effectively forced $u(x, t)$ to evolve by the nonlinear evolution equation. To solve for $u(x, t)$, we need to solve the inverse scattering problem, which is also a linear problem. The construction of the kernel(s), $F(z)$, as in (2.2), is another linear problem. Equation (2.3) is a linear integral equation for $K(x, y)$. Then u is reconstructed as in (2.4) by a linear operation. Thus with only linear techniques, we are able to solve these nonlinear evolution equations.

2.2. Soliton Solutions. One of the unique features of the IST is that it allows one to construct an infinity of exact nonlinear solutions, called the N -soliton solutions. These are also called reflectionless potentials because the scattering data consists of only bound state scattering data, with all reflection coefficients set equal to zero. In the case of the KdV equation, when the reflection coefficients vanish, the function $F(x + y; t)$ in Eq. (2.3), then separates into a finite sum of products of known functions of x and y , allowing one to obtain $K(x, y)$, also as a finite sum of known functions. The same is true for all other cases integrable by the IST.

The value of these solutions is tremendous, since they allow one to study and obtain exact results for these systems. The most important solution in this class is always the one-soliton solution, since it is the basic building block of these solutions, and also of any interactions between solitons and between solitons and radiation. The 1-soliton solution of the KdV hierarchy is

$$(2.7) \quad u = \frac{2\eta^2}{\cosh^2\{\eta[x - x_0(t)]\}}$$

and that of the NLS hierarchy is

$$(2.8) \quad q = \frac{2\eta e^{-2i\xi[x - x_0(t)]} e^{-i\phi_0(t)}}{\cosh\{2\eta[x - x_0(t)]\}}$$

where the form of $x_0(t)$ and $\phi_0(t)$ depends on the member of the hierarchy. The 2- and 3-solitons solutions demonstrate soliton interactions and collisions. Typically what happens in any soliton collision is that after the collision, the solitons separate out according to their individual velocities, completely unscathed, except for a possible phase shift in their positions and/or phase. There are two well known exceptions to this. In the soliton decay case of the 3WRI [8], an initial soliton in the high frequency wave can decay into its two daughter waves, transferring its identity to them. In the vector NLS [9], similar inelastic type collisions occur whereby a soliton in one mode (polarization), can flip to another combination of modes. Bound states can occur in the sine-Gordon field, where one can have stable 2-soliton bound states, called “breathers” [5], which are localized oscillations of the sine-Gordon field.

2.3. Hierarchies. Given any eigenvalue problem, Lax [2] had noted that one could always take the eigenvalue problem (the first component of the Lax Pair), and by simply extending the order of the evolution operator, B , one could generate another nonlinear integrable equation. All these nonlinear integrable equations which have a common eigenvalue problem in the Lax Pair, but different evolution operators, B , is called a “hierarchy”. Thus, there is a hierarchy for every one of these eigenvalue problems. For the Schrödinger equation, the most important members are the KdV equation, a 5th order KdV equation [10] and one-dimensional “caviton” equation [11] (the analogy of the sine-Gordon equation for the Schrödinger eigenvalue problem). In the ZS problem, we have a “workhorse” as far as physical equations are concerned. If we allow r in (1.7) to be in general independent from q , then in addition to the NLS [3], the hierarchy contains the modified KdV [4], the sine-Gordon [5], the sinh-Gordon equation [5], coherent pulse propagation and self-induced transparency (SIT) [12], stimulated Raman scattering (SRS) [13], and the defocusing NLS [14]. The hierarchy containing the three-wave resonant interaction (3WRI) includes all three forms of this interaction (explosive, soliton decay, and stimulated backscatter (SBS)) [8, 15, 16], as well as the Manakov vector NLS [9].

2.4. Inverse Scattering. Another aspect of the legacy is the wide variety of inverse scattering problems solved. In 1967, the IST of the Schrödinger equation had just recently been obtained [17, 18], and it was only in 1972 that the IST of the ZS eigenvalue problem had been solved [3]. Since that time, there has been a multitude of other and even more complex scattering problems solved.

The next one was the solution of the third-order eigenvalue problem for the three wave resonant interaction (3WRI) [15, 16], which interaction we shall return to later. This one differed from the ZS significantly only in the additional order of the problem. An important subcase of this was the inverse scattering solution for the eigenvalue problem for the vector nonlinear Schrödinger (VNLS) equation, solved by Manakov [9]. This VNLS equation is a very important and key equation for several studies of nonlinear optical pulses propagating in optical fibers [19, 20]. The generalization of the order 3 problem to order n has been done by Gerdjikov and Kulish [21].

More complex forms have also been done. The first one of these was the eigenvalue problem for the sine-Gordon equation in laboratory coordinates [22, 23]. Here one has the spectral parameter distributed among various potential terms,

and is

$$(2.9a) \quad v_{1x} + \left(\frac{i}{2}\zeta - \frac{i}{8\zeta} \cos u \right) v_1 = \left[\frac{i}{8\zeta} \sin u - \frac{1}{4}(u_x + u_t) \right] v_2$$

$$(2.9b) \quad v_{2x} - \left(\frac{i}{2}\zeta - \frac{i}{8\zeta} \cos u \right) v_1 = \left[\frac{i}{8\zeta} \sin u - \frac{1}{4}(u_x + u_t) \right] v_1.$$

One notes that due to this form, one cannot easily express this equation in the standard form as $L \cdot V = \lambda V$ where L is a nondegenerate differential operator, V is the eigenvector and λ is the eigenvalue. However, in spite of this, one still can solve the direct and inverse scattering problems for such systems. Further examples of such eigenvalue problems are the eigenvalue problems for the massive Thirring model [24], and the derivative NLSL [25]. A more standard form is the cubic generalization of the Schrödinger equation [10, 26],

$$(2.10) \quad \psi_{xxx} + 6Q\psi_x + 6R\psi = \lambda\psi$$

which appeared as the eigenvalue problem for the Boussinesq, Sawada-Kotera equation and the Kaup-Kuperschmidt equation. There is also the inverse scattering for multidimensional problems, such as the 3D form of the 3WRI [27], the KPI and II equations [28], as well as the DSI and II equations [29]. Several aspects of these are still of current research interest.

2.5. Perturbations and Closure. Once one has an exact method for the solution of a system, it then becomes possible to develop perturbation methods, to study nearby systems. This work was first done in 1976 for the ZS eigenvalue problem for the one-soliton solution [30], with a general summary of the perturbation method being given in 1978 [31].

Key to this, is the concept called “closure” or “completeness”, which itself arises from the one-to-one nature of the direct scattering transform and the IST [32]. What this simply means is that given any potential in the appropriate class, there exist a unique set of scattering data that can be associated with it by the eigenvalue problem of the Lax pair, and vice versa for the IST. Thus for any potential whose evolution is slightly perturbed away from its integrable value, by the direct scattering transform, it will be mapped into some other scattering data near the initial integrable scattering data. The perturbation problem is then to determine how this scattering data in the perturbed case evolves in time. Once its evolution is known, then by the IST, one may map back to the potentials and then obtain their evolution.

One solves this by relating variations in the potentials to variations in the scattering data, with the transformation from the former to the latter being accomplished by the so-called “squared eigenfunctions”. Then with these squared eigenfunctions, one may obtain the evolution of the scattering data under a given perturbation.

Under perturbations, one no longer has the simple evolution of the scattering data as in (2.1). Rather, one has a slow mixing of the various elements of the scattering data: solitons will decay and/or be pumped, transforming some of their energy into radiation and/or absorbing energy from the perturbations. In addition, radiation modes will similarly grow and/or decay. Many aspects of this have been covered in the review of soliton perturbations done by Kivshar and Malomed [33] as well as in Ref. [31].

However, whenever the eigenvalue problem has a singular structure, one has difficulties. The classical example is the perturbation theory of the KdV, where the Schrödinger eigenvalue problem is singular at $k = 0$. What happens is that for a pure one-soliton solution, $\rho(k) = 0$ at $k = 0$. But if there is just even the smallest amount of radiation present, we have $\rho(k = 0) = -1$. Now, the width of this region in k -space can be very small (proportional to the radiation density), so this region could be vanishingly small and one might have to look very close to $k = 0$ in order to see it. However, due to this singular behavior, any perturbations of the KdV soliton which will create any radiation, will always have secular terms [34]. Physically what is occurring is that a shelf or a depression is forming due to the perturbation [31], and this action does generate a finite shift in the scattering data as the perturbation vanishes.

2.6. General Integrable Evolution Equations. Given a linear dispersion relation, what are the possible integrable evolution equations for that system? Well, for any integrable system solvable by the ZS IST or the Schrödinger IST, that answer has been given by Newell and Kaup [35]. They showed that the most general system is any of the AKNS polynomial forms [7], coupled with a generalized SIT system. Using the properties of the ZS squared eigenstates and their closure, they were able to construct the most general evolution equation, given the linear dispersion relation. In other words, the dispersion relation of the linear theory determines the nonlinear terms, given the eigenvalue problem. The same could be done for any other hierarchy.

Now, a burning question has always been “Is it ever possible for an exactly integrable system to have evolving eigenvalues (i.e. for soliton amplitudes to evolve)?” In general, the answer to this is “No”. However, Kaup and Newell did find exceptions. Such equations can indeed be constructed, but potential applications for them seem to be remote.

2.7. Optical Systems. One question which arose naturally after a few years is “Why are so many of the physical integrable systems related to nonlinear optical systems?” As one goes down the list, one has the sine-Gordon equation (which applies to two-level atoms and is the sharp line limit of SIT), nonlinear Schrödinger (focusing and defocusing — the NLS is almost always the weakly nonlinear limit of any almost monochromatic envelope [36]), self-induced transparency (SIT — also more generally referred to as two-level coherent propagation), three-level coherent propagation [37], three-wave resonant interactions (3WRI) [15, 16, 8], second harmonic generation (SHG) [38], the three-dimensional form of the 3WRI [39, 40, 41], stimulated Raman scattering (SRS) [13, 42, 43], two photon propagation (TPP) [44, 45, 46], and degenerate TPP (DTPP) [47]. This is not meant to be a comprehensive list, but it does include the major integrable nonlinear optical systems.

Of these, the 3WRI was the first system to demonstrate a major departure from the accepted and expected soliton behavior. The first deviation was that the radiation would never asymptotically vanish. Also after any and all collisions, nor would the radiation separate out from the solitons. The reason for this is that the 3WRI has no dispersion, and therefore solitons will never separate from the radiation present. They each have the same velocity. One could now ask what is the significance of the 3WRI solitons, when they differ so much from KdV or NLS

solitons? To answer that the best we can say is that in the 3WRI, any solitons seem to simply represent a “packet” of something. It is a unit which cannot be broken up. Although there are certain exchange rules for the exchanging of solitons between the three envelopes [16, 8], nevertheless solitons in the 3WRI do seem to be some robust and coherent part of the envelope. On the other hand, the radiation component of any envelope has no such finite “packet” size, but rather can always be subdivided and redistributed among the three envelopes, subject only to the Conservation of Action laws [8].

It was the SRS system that first brought to the forefront certain ambiguities with the IST on finite intervals. The first solution of the finite interval case, by using an infinite interval IST, was given in [45], where the general IST for SRS was developed, and features of the solution were discussed. Numerics of SRS have been studied by Hilfer and Menyuk [48], the asymptotical form of the solution was described by Kaup [49], and was later solved as a Riemann-Hilbert problem by Fokas and Menyuk [43]. In the meantime, by analytically solving a model initial value problem, Menyuk and Kaup essentially found that for all these integrable nonlinear optical problems (except NLS and 3WRI), one could just as easily describe the solution as “an infinity of solitons with no radiation”, as well as by the usual description of “a finite number of solitons in a sea of radiation”. Briefly, why this could happen is basically the same reason as why a Fourier transform on a finite interval has a variety of forms. There one could use either a cosine series, or a sine series, or the exponential series to represent the function. For a given function, the coefficients in each of these series is quite different. Another way to look at this, is that on a finite interval, there is an infinite number of ways to take and combine plane waves to reconstruct some function inside the finite interval. Similarly for the IST, on a finite interval, there is no unique form for the scattering data.

Pursuing this further, consider the case where solitons are forbidden, as in the ZS $r = +q^*$ case. Now, what is going to happen in this case where no solitons are allowed on the infinite interval, if we try to represent the solution on the finite interval with no reflection coefficients, and only with solitons? What happens to this system on a finite interval? Well, we again find something surprising. Taking SHG as an example case of these defocusing systems, we find that the solution of a simple initial value problem can also be given in terms of no radiation, but now (since regular solitons are forbidden) an infinity of what is called “virtual solitons” [50]. These are indeed interesting objects.

The possibility of their existence was noted way back in 1974 [7]. However, no known application was then known for them. In the ZS IST, the typical soliton for the focusing case ($r = -q^*$) is of the form

$$(2.11) \quad q = \frac{2\eta e^{-2i\xi(x-x_0)} e^{-i\phi_0}}{\cosh[2\eta(x-x_0)]}.$$

Now, if one simply assumes that the defocusing case ($r = +q^*$) has one bound state in the scattering data, one obtains the “virtual soliton” solution

$$(2.12) \quad q = \frac{2\eta}{\sinh[2\eta(x-x_0)]},$$

which clearly will always be singular on any infinite interval. However, on a finite or a semi-infinite interval, as long as x_0 lies outside the physical region, the solution is perfectly valid. Thus a virtual soliton is the singular, ZS, $r = +q^*$, soliton.

(For certain technical reasons, they also were first found in the lower half complex plane [50], which is another reason for the prefix “virtual”. For more details on this, see Steudel’s contribution in this same series.)

As to my first question as to why so many nonlinear optical systems are integrable, perhaps the principle reason for this is the vast orders of magnitude difference between the speed of light and other velocities in these physical systems (which are typically acoustic velocities, and/or electronic drift velocities). Due to this disparity in velocities, one can then expect a multiple-scale expansion to give quite good results. Also one would then expect to see the higher order terms scaling like some power of this velocity ratio, and therefore rapidly vanishing.

2.8. Benjamin-Ono Equation. Although we could perhaps have included the Benjamin-Ono (BO) equation in the section on eigenvalue problems, it is unique enough to justify some additional comments. First it is a one dimension eigenvalue problem but it is also a nonlocal eigenvalue problem [51]. Actually it can be formulated as an electrostatics problem (Poisson’s equation) in two dimensions, since it can be stated in terms of Hilbert transforms and their analytical properties. This is probably the simplest explanation as to why Fokas and Ablowitz [51] could term the BO equation as a “pivotal equation for multidimensional problems”. It indeed does contain this multidimensional flavor. It also should be noted that there are now two versions of its IST [51, 52]. This is not entirely surprising, since the multidimensional 3D-3WRI [41] also contains a multitude of different forms of the IST. The multidimensional flavor of the BO equation stands out even more when one considers its closure property, its perturbation theory, and its Hamiltonian structure [53].

2.9. Reduction Problems. It seems that almost any integrable system can be found in the Yang-Mills field, if one knows how to find the right reduction [54]. However, there is another quite useful direction that one can take in reducing integrable systems. Let us first note that if one simplifies, or reduces the number of degrees of freedom of an integrable system by some set of constraints, consistent with the integrable flows, then the reduced system will also be integrable. As one example, consider the propagation of N solitons in an optical fiber. It is very important to maintain the spacings of these N solitons over long distances. So a key question is the stability of such an arrangement, which is simply an N -soliton solution. However, to try to analyze a 100-soliton state, where every soliton has approximately the same amplitude, is not an easy thing to analyze, since the transmission coefficient, a , has a zero of order 100 at the pole. Another approach for N larger than 2 or 3 is clearly needed. One way to do this is to approximate the system as a lattice, since in general, one expects the solitons to have an almost equally spacing (or being absent if representing a zero) and of equal amplitude. Then it turns out that when the N -soliton system is reduced to a lattice system, it reduces to a Complex Toda Lattice of N points, which is integrable. One now can study the stability problem of the reduced system, the Complex Toda Lattice, and transfer the results to the N -soliton optical pulse [55]. From this, one obtains general criteria about how the phase of each successive soliton should be adjusted, to maximize the stability of the pulse train, and etc.

2.10. The Fokas Method. Another new aspect of the IST has been recently developed by Prof. Fokas [56]. There are complex practical problems requiring

the solution of 2-dimensional, linear, constant coefficient, PDEs but with complex boundaries, such as wedges and polygons. Before his work, there was no general analytical technique for constructing solutions with such complicated boundaries. However, it is a simple matter to create a Lax pair for almost any linear PDE with constant coefficients [57]. Once this is done, one then can approach the solution of these linear two-dimensional, boundary value problems from the IST point of view, whereby one solves (or satisfies) both components of the Lax pair, thereby also satisfying the integrability condition, which is just the PDE to be solved. The method can also be extended to integrable nonlinear PDEs and evolution equations, however the solution then frequently requires the solution of a Riemann-Hilbert problem for the reflection coefficients.

2.11. Unsolved Problems. Now we will discuss what the hard problems are. They are hard because they haven't been solved. This is not a complete list, but is the start of a list of problems in need of solutions.

Although we have a solution for the IST of the cubic eigenvalue problem, (2.20), there still is no equivalent of the GLM equations for this system. What we do have is the solution of the Riemann-Hilbert problem for the eigenfunctions. What is next needed is a representation of the eigenfunctions in terms of some transformation kernels (like the $K(x, y; t)$ for the KdV). Once these are known, then the equivalent GLM equations will follow. Some progress in this direction has recently been made by A. Parker [58].

There are still aspects of KPI and II that are of interest. Numerical simulations [59] provide valuable insight into the evolution of these equations. I would also say that there are probably still some unanswered questions about DSI and II. However, the unfortunate thing about these two equations is that potential applications seem to be almost lacking, due to the scales involved. (Typical parameters for DS solitons in water require meter-like distances horizontally, but only centimeter-like water depths, and even smaller wave amplitudes.)

Questions about perturbation theory and closure relations seem to have become quite well understood, based on the works of Gerdjikov, Ivanov and Kulish [60], Beals and Coifman [61], and J. Yang [62]. They are also quite well understood for the 3D3WRI and DSI, since they both use the same eigenvalue problem. However, I know of no published work in this area.

Of the optical problems there is still interest in SHG and DTPP, and the latter has even the direct scattering problem to be detailed, as well as the inverse scattering problem.

I also strongly suspect that there are other integrable systems still to be found. Here one would have to apply Painlevé analysis and see what will result.

References

- [1] C.S. Gardner, J.M. Greene, M.D. Kruskal and R.M. Miura, Method for solving the Korteweg-deVries equation, *Phys. Rev. Lett.* **19** (1967), 1095-1097.
- [2] P.D. Lax, Integrals of nonlinear equations of evolution and solitary waves, *Comm. Pure Appl. Math.* **21** (1968), 467-490.
- [3] V.E. Zakharov and A.B. Shabat, Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media, *Zh. Eksp. Teor. Fiz* **61**, (1971), 118 [*Sov. Phys. JETP* **34**, (1972), 62].
- [4] M. Wadati, The exact solution of the modified Korteweg-de Vries equation, *J. Phys. Soc. Jap.* **32** (1972), 1681.

- [5] M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Method for solving the sine-Gordon equation, *Phys. Rev. Letters*, **30**, 25 (1973), 1262-1264.
- [6] M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Nonlinear-evolution equations of physical significance, *Phys. Rev. Letters* **31**, 2 (1973), 125-127.
- [7] M.J. Ablowitz, D.J. Kaup, A.C. Newell and H. Segur, The inverse scattering transform-Fourier analysis for nonlinear problems, *Studies in Appl. Math.* **LIII**, 4 (1974), 249-315.
- [8] D.J. Kaup, A. Reiman, and A. Bers, Space-time evolution of nonlinear three-wave interactions. I. Interaction in a homogeneous medium, *Rev. Mod. Phys.* **51** (1979), 275-310.
- [9] S.V. Manakov, On the theory of two-dimensional stationary self-focusing of electromagnetic waves, *Zh. Eksp. Teor. Fiz* **65** (1973), 505-516 [*Sov. Phys.-JETP* **38**, 2 (1974), 248-253].
- [10] D.J. Kaup, On the inverse scattering problem for cubic eigenvalue problems of the class $\psi_{xxx} + 6Q\psi_x + 6R\psi = \lambda\psi$, *Studies in Appl. Math.* **62** (1980), 189-216.
- [11] D.J. Kaup, Cavitons are solitons: An integrable ponderomotive system, *Phys. Fluids* **31**, 6 (1988), 1465-1470.
- [12] M.J. Ablowitz, D.J. Kaup, and A.C. Newell, Coherent pulse propagation, a dispersive, irreversible phenomenon, *J. Math. Phys.* **15**, 11 (1974), 1852-1858.
- [13] F.Y.F. Chu and A.C. Scott, Inverse scattering transform for wave-wave scattering, *Phys. Rev. A* **12** (1975), 2060.
- [14] V.E. Zakharov and A.B. Shabat, Interaction between solitons in a stable medium, *Zh. Eksp. Teor. Fiz* **64** (1973), 1627-1639.
- [15] V.E. Zakharov and S.V. Manakov, Resonant interaction of wave packets in nonlinear media, *Zh. Eksp. Teor. Fiz. Pis'ma Red.* **18**, (1973), 413 [*Sov. Phys.-JETP Lett.* **18** (1973), 243].
- [16] D.J. Kaup, The three-wave interaction — A nondispersive phenomenon, *Studies in Appl. Math.* **55** (1976), 9-44.
- [17] I.M. Gel'fand and B.M. Levitan, On the determination of a different equation from its spectral function, *Izvest. Akad. Nauk S.S.S.R. ser. matem* **15** (1951), 309-360 [*Amer. Math. Soc. Transl.*, Ser. 2, 1 (1955), 259-309].
- [18] L.D. Faddeyev and B. Seckler, The inverse problem in quantum theory of scattering, *J. Math. Phys.* **4** 1 (1963), 72-104.
- [19] T.I. Lakoba and D.J. Kaup, Perturbation theory for the Manakov soliton and its applications to pulse propagation in randomly birefringent fibers, *Phys. Rev. E* **56**, 5 (1997), 6147-6165.
- [20] H.A. Haus, W.S. Wong, and F.I. Khatri, Continuum generation by perturbation of soliton, *J. Opt. Soc. Am. B* **14**, 2 (1997), 304-313.
- [21] V.S. Gerdjikov and P.P. Kulish, The generating operator for the $n \times n$ linear system, *Physica* **3D** (1981), 549-564.
- [22] L.D. Faddeev and L.A. Takhtajan, Essentially nonlinear one-dimensional model of the classical field theory, *Teor. Mat. Fiz.* **21** (1974), 160-174.
- [23] D.J. Kaup, Method for solving the sine-Gordon equation in laboratory coordinates, *Studies in Appl. Math.* **54**, 2 (1975), 165-179.
- [24] E.A. Kuzhetsov and A.V. Mikhailov, On the complete integrability of the two-dimensional classical Thirring model, *Teor. Mat. Fiz.* **30**, 3 (1977), 303-314.
- [25] D.J. Kaup and A.C. Newell, An exact solution for a derivative nonlinear Schrödinger equation, *J. Math. Phys.* **19** (1978), 798-801.
- [26] V. Zakharov, On the stochastization of one-dimensional chains of nonlinear oscillators, *Sov. Phys. JETP* **38** (1974), 108-110.
- [27] A. Reiman and D.J. Kaup, Multi-shock solutions of random phase three-wave interactions, *Physica* **3D** (1981), 389-394.
- [28] A.S. Fokas and M.J. Ablowitz, On the inverse scattering of the time-dependent Schrödinger equation and the associated Kadomtsev-Petviashvili (I) equation, *Stud. Appl. Math* **69** (1983), 211-228.
- [29] A.S. Fokas and P.M. Santini, Dromions and a boundary value problem for the Davey-Stewartson 1 equation, *Physica D* **44** (1990), 99-130.
- [30] D.J. Kaup, A perturbation expansion for the Zakharov-Shabat inverse scattering transform, *SIAM J. Appl. Math.* **31**, 1 (1976), 121-133.
- [31] D.J. Kaup and A.C. Newell, Solitons as particles, oscillators, and in slowly changing media: A singular perturbation theory, *Proc. R. Soc. Lond. A* **361** (1978), 413-446.
- [32] D.J. Kaup, Closure of the squared Zakharov-Shabat eigenstates, *J. Math. Anal. Appl.* **54**, 3 (1976), 849-864.

- [33] Y.S. Kivshar and B.A. Malomed, Dynamics of solitons in nearly integrable systems, *Rev. Mod. Phys.* **61**, 4 (1989), 763-915.
- [34] V.I. Karpman and E.M. Maslov, Perturbation theory for solitons, *Zh. Eksp. Teor. Fiz.* **73** (1977), 537-559.
- [35] D.J. Kaup and A.C. Newell, Evolution equations, singular dispersion relations, and moving eigenvalues, *Adv. Math.* **31**, 1 (1979), 67-100.
- [36] A.C. Newell, The inverse scattering transform, in *Solitons*, Ed. R.K. Bullough and P.J. Caudrey, pp. 176-242 [Springer-Verlag, New York (1980)].
- [37] A.I. Maimistov, Rigorous theory of self-induced transparency in the case of a double resonance in a three-level medium, *Sov. J. Quantum Electron.* **14**, 3 (1984), 385-389.
- [38] D.J. Kaup, Simple harmonic generation: An exact method of solution, *Studies in Appl. Math.* **59** (1978), 25-35.
- [39] V.E. Zakharov and S.V. Manakov, The theory of resonance interaction of wave packets in nonlinear media, *Zh. Eksp. Teor. Fiz.* **69** (1975), 1654 [*Sov. Phys.-JETP* **42** (1976), 842].
- [40] H. Cornille, Solutions of the nonlinear 3-wave equations in three spatial dimensions, *J. Math. Phys.* **20** (1979), 1653.
- [41] D.J. Kaup, The solution of the general initial value problem for full three dimensional three-wave resonant interaction, *Physica D* **3** (1981), 374-395.
- [42] D.J. Kaup, The method of solution for stimulated Raman scattering and two-photon propagation, *Physica D* **6** (1983), 143-154.
- [43] A.S. Fokas and C.R. Menyuk, Integrability and self-similarity in transient stimulated Raman scattering, *J. Nonlinear Sci.* **9** (1999), 1-31.
- [44] H. Steudel, Stimulated Raman scattering with ultrashort pulses, *Exp. Tech. Phys.* **20** 5 (1972), 409-415
- [45] H. Steudel, Solitons in stimulated Raman scattering and resonant two-photon propagation, *Physica* **6 D** (1983), 155-178.
- [46] H. Steudel, N -soliton solutions to degenerate self-induced transparency, *J. Mod. Opt.* **35**, 4 (1988), 693-702.
- [47] H. Steudel and D.J. Kaup, Degenerate two-photon propagation and the oscillating two-stream instability: The general solution for amplitude-modulated pulses, *J. Mod. Opt.* **43**, 9 (1996), 1851-1866.
- [48] G. Hilfer and C.R. Menyuk, Stimulated Raman scattering in the transient limit, *J. Opt. Soc. Am. B* **7** (1990), 739.
- [49] D.J. Kaup, The asymptotic solution of the stimulated Raman-scattering equation, in *Painlevé Transcendents*, Ed. D. Levi and P. Winternitz, Plenum Press, New York, 1992, 345-351.
- [50] D.J. Kaup and H. Steudel, Virtual solitons and the asymptotics of second harmonic generation, *Inv. Prob.* **17** (2001), 959-970.
- [51] A.S. Fokas and M.J. Ablowitz, The inverse scattering transform for the Benjamin-Ono equation — A pivot to multidimensional problems, *Stud. Appl. Math* **68** (1983), 1-10.
- [52] D.J. Kaup and Y. Matsuno, The inverse scattering transform for the Benjamin-Ono equation, *Stud. Appl. Math.* **101** (1998), 73-98.
- [53] D.J. Kaup, T.I. Lakoba and Y. Matsuno, Complete integrability of the Benjamin-Ono equation by means of action-angle variables, *Phys. Letters A* **238** (1998), 123-133.
- [54] M.J. Ablowitz, S. Chakravarty and R. Halburd, The generalized Chazy equation from the self-duality equations, Preprint (2002).
- [55] V.S. Gerdjikov, E.G. Evstatiev, D.J. Kaup, G.L. Diankov and I.M. Uzunov, Stability and quasi-equidistant propagation of NLS soliton trains, *Phys. Letters A* **241** (1998), 323-328.
- [56] A.S. Fokas, Two-dimensional linear PDEs in a convex polygon, *Proc. R. Soc. Lond. A* (2001).
- [57] D. ben-Avraham and A.S. Fokas, The solution of the modified Helmholtz equation in a triangular domain and an application to diffusion-limited coalescence (Preprint, 2001)
- [58] A. Parker, A reformulation of the “dressing method” for the Sawada-Kotera equation, *Inverse Problems*, submitted.
- [59] E. Infeld, A. Senatorski and A.A. Skorupski, Numerical simulations of Kadomtsev-Petviashvili soliton interactions, *Phys. Rev. E* **51** (1995), 3183-3191.
- [60] V.S. Gerdzhikov, M.I. Ivanov and P.P. Kulish, Quadratic bundle and nonlinear equations, *Teor. Mat. Fiz.* **44**, 3 (1980), 342-357.
- [61] R. Beals and R.R. Coifman, Scattering and inverse scattering for first order systems, *Comm. Pure Appl. Math.* **37** (1984), 38-90.

[62] J. Yang, Private communication (2002).

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CENTRAL FLORIDA, ORLANDO, FL 32816-1364

E-mail address: `kaup@ucf.edu`

Application of Inverse Scattering Method to Problems of Differential Geometry

V. Zakharov

ABSTRACT. This paper summarizes the results on application of the Inverse Scattering Method (ISM) to classical problems of Differential Geometry. Some results are new. The most important one is the following - each space of diagonal curvature can be approximated by integrable spaces of flat connection.

1. Introduction

This article presents, in a brief form, the recent results on application of the Inverse scattering method to some problems of Differential geometry. A connection between the theory of Integrable systems and Differential geometry is not a new concept. The Sine–Gordon equation was introduced in the theory of surfaces of constant negative curvature around 1860. Actually, it should be called the “Bonnet equation”. The Backlund transformations appeared in Differential geometry also in nineteenth century.

At present, it is established that some important integrable equations, found in the last three decades in the theory of solitons (Bullough–Dodd equation, Dawey–Stewartson stationary equation, *etc.*), have a geometrical interpretation. All of the equations mentioned above are in $1 + 1$ dimensions.

In this article we explore some geometrical applications of integrable systems in $2 + 1$ dimensions. The closest relative of the famous “three–wave equation”,

$$(1.1) \quad \begin{aligned} \frac{\partial \Psi_1}{\partial x_1} &= i \Psi_2 \Psi_3, \\ \frac{\partial \Psi_2}{\partial x_2} &= i \Psi_1 \Psi_3^*, \\ \frac{\partial \Psi_3}{\partial x_3} &= i \Psi_1 \Psi_2^*, \end{aligned}$$

plays the central role in our study. Equation (1.1) describes a resonant interaction of three quasimonochromatic wave trains in a nonlinear media. This interaction is an induced Raman scattering of the wave 1 to waves 2 and 3, and the inverse process. There is a special reason to present this article to this particular collection

1991 *Mathematics Subject Classification.* Primary 53A07, 53A35.

Key words and phrases. Differential geometry, spaces of diagonal curvature.

of papers. One of the pioneers in the study of system (1.1) was David Kaup [1]. (See also [2], [3].)

In this article we study the n -dimensional overdetermined system,

$$(1.2) \quad \frac{\partial Q_{ij}}{\partial u^j} = Q_{ij} Q_{jk},$$

$$i, j, k = 1, \dots, n, \quad i \neq j \neq k.$$

This system is nontrivial if $n \geq 3$. For $n = 3$, equation (1.1) is a special case of (1.2), if Q_{ij} is a complex-valued matrix. We will study mostly the case of real-valued Q_{ij} .

Equation (1.2) have appeared in Differential geometry in the middle of last century, long before the concept of Raman scattering was formulated. It's origin was connected with the problem of classification of n -orthogonal curvilinear coordinate systems in R^n . This problem, which was formulated almost two hundred years ago, was considered for a long time as one of the central in Differential geometry. In 1910, Gaston Darboux devoted to this problem (to the case $n = 3$ only), the monograph of 546 pages [4]. If an expert on the Inverse scattering technique looks into this book, he will be surprised: how many formulae are familiar to him!

After the First World War, the problem of n -orthogonal coordinate systems was almost forgotten. An interest to this problem was revived ten years ago, when Dubrovin and Novikov [5, 6], then Tzarev [7], have published their papers on integrable systems of hydrodynamic type. The complete solution of this problem was found only in 1998 [8].

In the theory of n -orthogonal coordinate systems, equation (1.2) is considered together with the additional constrain:

$$(1.3) \quad E_{ij} = \frac{\partial Q_{ij}}{\partial u^j} + \frac{\partial Q_{ji}}{\partial u^i} + \sum_{i \neq j \neq k} Q_{ik} Q_{jk} = 0.$$

This constrain is a "reduction" imposed on (1.2). However, equation (1.2) itself has a nice geometrical interpretation. It describes n -dimensional Riemann spaces of diagonal curvature [9, 10].

The spaces of diagonal curvature include very interesting class of Riemann spaces: 2 - D surfaces, spaces of constant curvature, and spaces of flat connection. They all are defined by the reduction more complicated than (1.3). In this article we will show how the Inverse scattering method could be implemented for description of the spaces of flat connection. We will show also that the closure of this class of spaces coincides with the whole class of the diagonal curvature. This question was formulated to the author by E. Ferapontov [11]. In this article we will use the Inverse scattering method in the form of "dressing method" [2, 9, 12, 13].

We will discuss also one important problem. Some classes of the Einstein spaces (solutions of Einstein equations of general relativity) belong to the class of spaces of diagonal curvature. The Schwartzschild's metric around a black hole is in this list. Can one find the class of reduction, which separates the Einstein spaces from the whole class of spaces of diagonal curvature? This extremely interesting question is still unanswered.

2. N -orthogonal coordinate systems

The problem of classification of n -orthogonal curvilinear coordinate systems was formulated by Dupin and Binet in 1810. The problem is the following: Find

in R^n all coordinate systems,

$$(2.1) \quad u^i = u^i(x^1, \dots, x^n), \quad i = 1, \dots, n$$

$$(2.2) \quad \det \left\| \frac{\partial u^i}{\partial x^j} \right\| \neq 0,$$

satisfying the condition of orthogonality,

$$(2.3) \quad \sum_{k=1}^n \frac{\partial u^i}{\partial x^k} \frac{\partial u^j}{\partial x^k} = 0, \quad i \neq j.$$

In virtue of (2.2) one can resolve equations (2.1),

$$(2.4) \quad x^i = x^i(u^1, \dots, u^n),$$

and introduce Lamé coefficients,

$$(2.5) \quad H_i^2 = \sum_k \left(\frac{\partial x^i}{\partial u^k} \right)^2.$$

In the coordinate system u^i the metric tensor in R^n is diagonal,

$$(2.6) \quad ds^2 = \sum_{i=1}^n H_i^2 (du^i)^2,$$

and Christoffel's coefficients for the Levi-Civita connection are the following:

$$(2.7) \quad \begin{aligned} \Gamma_{lm}^i &= 0 \quad (i \neq l \neq m), \\ \Gamma_{il}^i &= \frac{1}{H_i} \frac{\partial H_i}{\partial u^l}, \end{aligned}$$

$$(2.8) \quad \Gamma_{ll}^i = -\frac{H_l}{H_i^2} \frac{\partial H_l}{\partial u^i}.$$

One can calculate the elements of Riemann curvature tensor R_{ijkl} . They are:

$$(2.9) \quad \begin{aligned} R_{ij,kl} &= 0, \quad i \neq j \neq k \neq l, \\ R_{ik,jk} &= -H_i H_j \left(\frac{\partial Q_{ij}}{\partial u^k} - Q_{ik} Q_{kj} \right), \quad i \neq j \neq k, \end{aligned}$$

$$(2.10) \quad R_{ij,ij} = -H_i H_j E_{ij}, \quad i \neq j,$$

where rotation coefficients Q_{ij} are

$$(2.11) \quad Q_{ij} = \frac{1}{H_j} \frac{\partial H_i}{\partial u^j}, \quad i \neq j,$$

and

$$(2.12) \quad E_{ij} = \frac{\partial Q_{ij}}{\partial u^j} + \frac{\partial Q_{ji}}{\partial u^i} + \sum_{k \neq i,j} Q_{ik} Q_{jk}, \quad i \neq j.$$

As far as R^n is flat, the Riemann curvature tensor is identically equal zero. Hence Q_{ij} satisfy the following two systems of equations:

$$(2.13) \quad \frac{\partial Q_{ij}}{\partial u^k} = Q_{ik} Q_{kj},$$

$$(2.14) \quad E_{ij} = \frac{\partial Q_{ij}}{\partial u^j} + \frac{\partial Q_{ji}}{\partial u^i} + \sum_{k \neq i,j} Q_{ik} Q_{jk} = 0.$$

If Q_{ij} are known, the Lamé coefficients H_i can be found as solutions of the linear system

$$(2.15) \quad \frac{\partial H_i}{\partial u^j} = Q_{ij} H_j, \quad i \neq j.$$

Two different solutions of system (2.15), H_i and \tilde{H}_i , are *Combescure equivalent*, $H_i \leftrightarrow \tilde{H}_i$, at the same Q_{ij} . The Lamé coefficients satisfy to Gauss-Lamé equations,

$$(2.16) \quad \frac{\partial^2 H_i}{\partial u^l \partial u^m} = \frac{1}{H_l} \frac{\partial H_l}{\partial u^m} \frac{\partial H_i}{\partial u^l} + \frac{1}{H_m} \frac{\partial H_m}{\partial u^l} \frac{\partial H_i}{\partial u^m},$$

and to the additional system,

$$(2.17) \quad \frac{\partial}{\partial u^l} \frac{\partial H_i}{H_l \partial u^l} + \frac{\partial}{\partial u^i} \frac{1}{H_i} \frac{\partial H_l}{\partial u^l} + \sum \frac{1}{(H_m)^2} \frac{\partial H_i}{\partial u^m} \frac{\partial H_l}{\partial u^m} = 0, \quad (i \neq l).$$

One can construct the adjoint linear system,

$$(2.18) \quad \frac{\partial \Psi_i}{\partial u^k} = Q_{ik} \Psi_k, \quad i \neq k,$$

where Ψ_i are adjoint Lamé coefficients. They satisfy to equation (2.16) and to the additional system,

$$(2.19) \quad \frac{\partial}{\partial u^l} \frac{\partial \Psi_i}{\Psi_l \partial u^l} + \frac{\partial}{\partial u^i} \frac{1}{\Psi_i} \frac{\partial \Psi_l}{\partial u^l} + \frac{1}{\Psi_i \Psi_l} \sum_{n \neq i, j} \frac{\partial \Psi_m}{\partial u^i} \frac{\partial \Psi_m}{\partial u^j} = 0$$

Both systems (2.15), (2.18) are overdetermined and are compatible in virtue of (2.13).

Note that

$$\frac{\partial}{\partial u^i} H_k \Psi_k = \frac{\partial}{\partial u^k} H_i \Psi_i,$$

hence

$$(2.20) \quad H_k \Psi_k = \frac{\partial h}{\partial u^k}.$$

One can check that the potential h satisfies the system of Laplace equations,

$$(2.21) \quad \frac{\partial^2 h}{\partial u^k \partial u^l} = \frac{1}{H_k} \frac{\partial H_k}{\partial u^l} \frac{\partial h}{\partial u^k} + \frac{1}{H_l} \frac{\partial H_l}{\partial u^k} \frac{\partial h}{\partial u^l},$$

$$(2.22) \quad \frac{\partial^2 h}{\partial u^k \partial u^l} = \frac{1}{\Psi_k} \frac{\partial \Psi_k}{\partial u^l} \frac{\partial h}{\partial u^k} + \frac{1}{\Psi_l} \frac{\partial \Psi_l}{\partial u^k} \frac{\partial h}{\partial u^l}.$$

Systems (2.21), (2.22) are overdetermined linear equations imposed on the potential h . They can be treated as an analog of the "Lax pair" for Gauss-Lamé equations (2.16).

Equation (2.21) can be rewritten as follow:

$$(2.23) \quad \frac{\partial^2 h}{\partial u^k \partial u^l} = \Gamma_{kl}^k \frac{\partial h}{\partial u^k} + \Gamma_{lk}^l \frac{\partial h}{\partial u^l},$$

and one can prove that functions $x^i = x^i(u^1, \dots, u^n)$ are solutions of (2.23):

$$(2.24) \quad \frac{\partial^2 x^i}{\partial u^k \partial u^l} = \Gamma_{kl}^k \frac{\partial x^i}{\partial u^k} + \Gamma_{lk}^l \frac{\partial x^i}{\partial u^l}.$$

Moreover, they satisfy the equations

$$(2.25) \quad \frac{\partial^2 x^i}{(\partial u^l)^2} = \sum_k \Gamma_{ll}^k \frac{\partial x^l}{\partial u^k}.$$

Equations (2.24), (2.25) stem from the fact that a straight line defined by conditions $x^i = c^i = \text{const}$, $i \neq l$, is the geodesic line in R^n .

We see now that the problem of description of n -orthogonal coordinate systems can be solved in the following several steps.

- (1) Solve system (2.13).
- (2) Find the solutions for linear system (2.15); on this stage one finds all Combescure equivalent metrics connected with the given rotation coefficients Q_{ij} satisfying equation (2.14).
- (3) Solve the adjoint system (2.18).
- (4) Integrate the relation (2.20); on this stage one finds the general solution of Laplace equations (2.21), (2.22).
- (5) Find the solutions of (2.13), satisfying additional conditions (2.11).
- (6) Find $x^i(u^1, \dots, u^n)$, the array of solutions of (2.21) and (2.24) satisfying additional conditions (2.25). In virtue of the Bonnet theorem, x^i are defined uniquely up to motions in R^n . Note, that systems (2.24), (2.25) are compatible only if equation (2.11) is satisfied.

The first five points of this program pertain to the intrinsic geometry of R^n in a new curvilinear coordinate system. On this stage R^n appears as a flat Riemann space with diagonal metric (2.6). The last point realizes embedding of this space to Cartesian coordinate system $x^i(u^1, \dots, u^n)$.

3. Spaces of diagonal curvature

In this chapter we introduce a new geometrical object - the space of diagonal curvature. By definition, it is a Riemann space of n dimensions, G^n , satisfying the following two conditions:

Condition 1. One can introduce in G^n (in some simple-connected domain) a diagonal coordinate system, such that the metric tensor g_{ik} is diagonal,

$$(3.1) \quad g_{ik} = H_i^2 \delta_{ik}.$$

Condition 2. In this coordinate system the non-diagonal elements of Riemann's tensor are zero,

$$(3.2) \quad R_{ik,jk} = 0, \quad i \neq j \neq k.$$

Note, that the diagonal coordinate system (3.1) could be introduced by many different ways. G^n is the space of diagonal curvature if at least in one diagonal coordinate system condition (3.2) is satisfied.

The most trivial example of the space of diagonal curvature is a flat space R^n or torus R^n/Z^n . In this case, description of diagonal metric tensors is exactly the problem of classification of n -orthogonal systems in R^n . One can display other examples of the spaces of diagonal curvature.

1. Adjoint Lamé metrics.

Let H_i be Lamé coefficients for an n -orthogonal curvilinear in R^n and Ψ_i - adjoint Lamé coefficients. Let us consider a Riemann space with the following metric:

$$(3.3) \quad ds^2 = \sum \Psi_i^2 (du^i)^2.$$

It is a space of diagonal curvature. However, it is not flat in a general case, because the additional constrain (2.29) imposed on Ψ_i is different from condition (2.17)

imposed on H_i . Only in a very special case of "Egorov's metric", when Q_{ij} is a symmetric matrix, $Q_{ik} = Q_{ki}$, we will have

$$H_i = \Psi_i, \quad H_i^2 = \frac{\partial h}{\partial u^i},$$

and conditions (2.16), (2.178) will coincide.

2. Spaces of constant curvature.

In the spaces of constant curvature, the Riemann's tensor is:

$$(3.4) \quad R_{ij,kl} = \epsilon(g_{ik}g_{jl} - g_{il}g_{jk}).$$

Here ϵ is the curvature of space.

All such spaces admit diagonal metrics, $g_{ik} = H_i^2 \delta_{ik}$. In this case,

$$(3.5) \quad E_{ij} = \epsilon H_i H_j.$$

By a trivial rescaling one can get $\epsilon = \pm 1$.

3. Spaces of flat connection.

Spaces of flat connection, $G^{n,N}$, are generalizations of spaces of constant curvature. Let $H_i^{(k)}$, $k = 1, \dots, N$ be a set of Combescure equivalent metrics, corresponding to the given rotation coefficients Q_{ij} . In the space of flat connection $R_{ij,kj} = 0$, and

$$(3.6) \quad E_{ij} = \sum_{k=1}^N \epsilon_k H_i^{(k)} H_j^{(k)}, \quad \epsilon_k = \pm 1.$$

Spaces of flat connection appear in the following geometrical problem. Let us consider a special class of n -orthogonal curvilinear coordinates in $n+N$ dimensional Euclidean space R^{n+N} . Suppose, that coordinates in this space can be separated in two classes: u^1, \dots, u^n and y^1, \dots, y^N ; and the metric is

$$(3.7) \quad ds^2 = \sum_{i=1}^n H_i^2 (du^i)^2 + \sum_{i=1}^N (dy^i)^2.$$

In this case H_i are linear functions on y ,

$$(3.8) \quad H_i = P_i + \sum_{a=1}^N H_i^a y^a,$$

where P_i, H_i^a are functions on coordinates u^i only.

One can introduce rotation coefficients,

$$(3.9) \quad Q_{ij}(u^1, \dots, u^n) = \frac{1}{P_j} \frac{\partial P_i}{\partial u^j},$$

which satisfy the system of equations

$$(3.10) \quad \frac{\partial Q_{ij}}{\partial u^k} = Q_{ik} Q_{kj}.$$

Hence the n -dimensional Riemann space $G^{n,N}$ with the metric

$$(3.11) \quad ds^2 = \sum_{i=1}^n H_i^2 (du^i)^2$$

is the space of diagonal curvature. Moreover, H_i^a satisfies the equations

$$(3.12) \quad \frac{\partial H_i^a}{\partial u^k} = Q_{ik} H_k^a.$$

Thus all arrays H_i^a, H_i^b are Combescure equivalent to each other and to P_i . One can easily check that

$$(3.13) \quad E_{ij} = \frac{\partial Q_{ij}}{\partial u^j} + \frac{\partial Q_{ji}}{\partial u^i} + \sum_{k \neq i, j} Q_{ik} Q_{jk} = - \sum_a H_i^a H_j^a.$$

It means that $G^{n, N}$ is a space of flat connection.

All two-dimensional Riemannian spaces (surfaces) are spaces of flat connection. Let Γ be a surface in R^3 . One can introduce coordinates x_1, x_2 on Γ such that both the first and the second quadratic forms of the surface are diagonal,

$$(3.14) \quad \begin{aligned} \omega_1 &= p^2 dx_1^2 + q^2 dx_2^2, \\ \omega_2 &= pA dx_1^2 + qB dx_2^2. \end{aligned}$$

Coordinates x_1, x_2 are defined up to the trivial transformation $x_1 = x_1(u_1), x_2 = x_2(u_2)$. The coefficients of these two quadratic forms ω_1, ω_2 cannot be chosen independently. They are connected by three nonlinear PDE's known as Gauss-Codazzi equations (GCE). These equations can be written in a nice and compact form after introducing new functions α, β :

$$(3.15) \quad \alpha = \frac{1}{q} \frac{\partial p}{\partial x_2}, \quad \beta = \frac{1}{p} \frac{\partial q}{\partial x_1}.$$

Thus

$$(3.16) \quad \frac{\partial p}{\partial x_2} = \alpha q, \quad \frac{\partial q}{\partial x_1} = \beta p,$$

then

$$(3.17) \quad \frac{\partial A}{\partial x_2} = \alpha B, \quad \frac{\partial B}{\partial x_1} = \beta A,$$

and

$$(3.18) \quad \frac{\partial \alpha}{\partial x_2} + \frac{\partial \beta}{\partial x_1} + \alpha \beta = 0.$$

Let us embed the surface Γ in R^3 . One can do this by constructing in vicinity of Γ a special three-orthogonal coordinate system, such that

$$(3.19) \quad ds^2 = H_1^2 dx_1^2 + H_2^2 dx_2^2 + H_3^2 dx_3^2,$$

where

$$(3.20) \quad H_1 = p + Ax_3, \quad H_2 = q + Bx_3, \quad H_3 = 1.$$

Apparently x_3 is directed along the normal vectors to Γ .

One can check that rotation coefficients Q_{ij} are:

$$(3.21) \quad \begin{aligned} Q_{31} &= Q_{32} = 0, \\ Q_{13} &= A, \quad Q_{23} = B, \quad Q_{12} = \alpha, \quad Q_{21} = \beta. \end{aligned}$$

The rotation coefficients Q_{ij} satisfy the system (3.10), which is reduced now to system (3.17). Thus, a two-dimensional surface is the space of diagonal curvature. Moreover, comparing (3.16) and (3.17) one can see that A^2, B^2 can be treated as elements of an orthogonal metric that is Combescure equivalent to the elements of

first quadratic form p^2, q^2 . Hence equation (3.18) can be interpreted as a special case of equation (3.13), and $2-D$ surface is the space of flat connection. In a general case, equations (3.10), (3.12) and (3.13) can be called Gauss-Codazzi equations as well.

4. Einstein spaces of diagonal curvature

So far we discussed Riemann spaces of signature $(n, 0)$ only, but the whole theory can be extended to pseudo-Riemann spaces of arbitrary signature (p, q) . One has just to assume that some Lamé coefficients H_i are pure imaginary. In this chapter we will discuss spaces of diagonal curvature $G^{(4)}$ satisfying Einstein equations of general relativity.

These equations read:

$$(4.1) \quad \Lambda_k^i = \frac{8\pi k}{c^4} T_k^i,$$

where k is the gravity constant, c is the light velocity, and

$$(4.2) \quad \Lambda_k^i = R_k^i - \frac{1}{2} R \delta_k^i.$$

Here R_k^i is the Ricci tensor, R is a scalar curvature, and T_k^i is an energy-momentum tensor. Thereafter we will use a system of units such that $k/c^4 = 1$. Thus,

$$\Lambda_k^i = 8\pi T_k^i.$$

We will also use the covariant version of Einstein equations,

$$(4.3) \quad R_{ik} - \frac{1}{2} R g_{ik} = 8\pi T_{ik}.$$

Due to the Bianchi identity, the Einstein's tensor Λ_k^i satisfies the condition

$$(4.4) \quad \Lambda_{k,i}^i = 0.$$

In virtue of (4.1), the same condition is imposed on the energy-momentum tensor,

$$(4.5) \quad T_{k,i}^i = 0.$$

In general, (4.5) is the condition imposed on the energy-momentum spectrum. If one interprets Einstein equations in that broad sense, any Riemann space of signature $(1, 3)$ can be treated as a solution of Einstein equations (Einstein space) for a proper energy-momentum tensor. Condition (4.4) is satisfied automatically.

Suppose that g_{ik} is diagonal. In this case $g^{ik} = 1/H_i^2 \delta_{ik}$, and

$$(4.6) \quad R_{kl} = \frac{1}{H_i^2} R_{ki,li}.$$

In the space of diagonal curvature the Ricci tensor is diagonal too:

$$(4.7) \quad R_{kl} = \tilde{R}_k \delta_{kl},$$

$$(4.8) \quad \tilde{R}_k = -H_k \sum_{i \neq k} \frac{E_{ik}}{H_i}.$$

In the same way:

$$(4.9) \quad R_l^k = R_k \delta_l^k,$$

$$(4.10) \quad R_k = - \sum_{i \neq k} \alpha_{ik}, \quad \alpha_{ik} = \frac{E_{ik}}{H_i H_k},$$

and

$$(4.11) \quad \Lambda_k^i = \Lambda_k \delta_k^i, \quad \Lambda = \sum \alpha_{lm}, \quad (l \neq i, m \neq i, l < m).$$

The energy–momentum tensor T_k^i is diagonal as well:

$$(4.12) \quad T_k^i = T_k \delta_k^i,$$

and Einstein equations can be written in the form:

$$(4.13) \quad \Lambda_k = 8\pi T_k.$$

Here T_k are arbitrary solutions of the following equation:

$$(4.14) \quad \frac{\partial T_k}{\partial u^k} + \sum_{i \neq k} T_i \frac{\partial \ln H_k}{\partial u^i} = 0.$$

If Λ_k are components of the Einstein tensor for some space of diagonal curvature, equations (4.14) are satisfied automatically.

Suppose that $T_k \equiv 0$, and Einstein equations describe metrics in vacuum. In this case:

$$(4.15) \quad \alpha_{12} = \alpha_{34} = \alpha, \quad \alpha_{13} = \alpha_{24} = \beta, \quad \alpha_{14} = \alpha_{23} = \gamma,$$

and $\alpha + \beta + \gamma = 0$.

At the moment, one can enlist more than a dozen exact solutions of Einstein equations with a diagonal metric tensor. It is interesting that most of them describe spaces with a diagonal curvature. We present here a list of such spaces, that is far from being complete.

1. "One–dimensional" metrics

Let

$$(4.16) \quad ds^2 = -H_0^2 dx_0^2 + H_1^2 dx_1^2 + H_2^2 dx_2^2 + H_3^2 dx_3^2,$$

where H_i are functions on one variable only, suppose on x_0 . In this case the only nontrivial rotational coefficients are

$$(4.17) \quad Q_{i0} = \frac{1}{H_0} \frac{\partial H_i}{\partial x_0} = Q_{i0}(x_0),$$

and apparently, all equations (3.10) are satisfied. In a general case, metric (4.16) presumes existence of matter. In vacuum, it turns to the Kasner metric:

$$(4.18) \quad \begin{aligned} H_1 &= x_0^{2p_1}, \quad x_2 = x_0^{2p_2}, \quad x_3 = x_0^{2p_3}, \\ p_1 + p_2 + p_3 &= 1, \quad p_1^2 + p_2^2 + p_3^2 = 1. \end{aligned}$$

2. Spherically–symmetric Einstein spaces

In the theory of general relativity, this is a very important class of spaces, which includes the Schwartzschild's space outside of a black hole, the basic cosmological models, and Tolman's space describing the collapse of dust matter. In all these cases the metric is

$$(4.19) \quad ds^2 = -H_0^2 dx_0^2 + H_1^2 dx_1^2 + x_1^2(dx_2^2 + \sin^2 x_2 dx_3^2),$$

where H_0, H_1 are functions on x_0, x_1 only.

In this case, the nonzero rotation coefficients are:

$$(4.20) \quad \begin{aligned} Q_{01} &= \frac{1}{H_1} \frac{\partial H_0}{\partial x_1}, \quad Q_{10} = \frac{1}{H_0} \frac{\partial H_1}{\partial x_0}, \quad Q_{32} = \cos x_2, \\ Q_{21} &= \frac{1}{x_1}, \quad Q_{31} = \frac{1}{x_3} \sin x_2. \end{aligned}$$

Among equations (3.10), the only one is non-trivial,

$$(4.21) \quad \frac{\partial Q_{31}}{\partial x_2} = Q_{32} Q_{21},$$

and is satisfied in virtue (4.20).

Hence all the spherically symmetric Einstein spaces, both in vacuum and in the presence of matter, are spaces of diagonal curvature.

3. Bianchi III model

In this case,

$$(4.22) \quad -ds^2 = -dx_0^2 + a^2 e^{-2mx_3} dx_e^2 + b^2 dx_2^2 + c^2 dx_3^2,$$

where a, b, c are functions on x_0 only.

One can check that metric (4.22) describes the space of diagonal curvature if $c(x_0) = \lambda a(x_0)$, (λ is constant).

4. Bianchi V model

In this model the metric has the form

$$(4.23) \quad -ds^2 = -dx_0^2 + a^2 e^{2x_3} dx_1^2 + b^2 e^{2x_3} dx_2^2 + c^2 dz^2,$$

where a, b, c are functions on x_0 only.

This model describes the space of diagonal curvature if $a = \lambda_1 c$, $b = \lambda_2 c$ (λ_1, λ_2 are constants).

5. Bianchi VI model

Now

$$(4.24) \quad -ds^2 = dx_0^2 + a^2 e^{-2mx_3} dx_1^2 + b^2 e^{2x_3} dx_2^2 + c^3 dx_3^2 \quad (m \neq 0, -1),$$

where again a, b, c are functions on x_0 only. Like in the previous case, this model belongs to the spaces of diagonal curvature if $a = \lambda_1 c$, $b = \lambda_2 c$.

These examples show that some important known solutions of Einstein equations are the spaces of diagonal curvature. Note, that it is not clear so far how unique is diagonal coordinatization of Einstein spaces corresponding to the enlisted Bianchi models. It might happen that in some other diagonal coordinate system equations (3.10) are satisfied for a more broad class of spaces.

5. Dressing method in application to spaces of diagonal curvature

In this chapter we describe a procedure of integration of equations (3.10), a dressing method described in 1974 [2]. We present here some important developments of this method.

Let $F(s, s', u)$ be a matrix $n \times n$ -valued kernel of an integral operator \hat{F} acting on vector-functions on the real axis $-\infty < s < \infty$, and $u = u^1, \dots, u^n$ be

n -dimensional parameter. One can study two simultaneous factorizations of the operator \hat{F} :

$$(5.1) \quad 1 + \hat{F} = (1 + \hat{K}^+)^{-1}(1 + \hat{K}^-),$$

$$(5.2) \quad 1 + \hat{F} = (1 + \hat{M}^+)(1 + \hat{M}^-)^{-1}$$

Here \hat{K}^\pm, \hat{M}^\pm are triangle operators. Their kernels, $K^\pm(s, s', u)$ and $M^\pm(s, s', u)$, satisfy the conditions:

$$(5.3) \quad \begin{aligned} K^+(s, s') &= 0, & s' < s, \\ M^+(s, s') &= 0, & s' < s, \\ K^-(s, s') &= 0, & s' > s, \\ M^-(s, s') &= 0, & s' > s. \end{aligned}$$

Apparently,

$$1 + M^+ = (1 + K^+)^{-1},$$

and K^+, M^+ are connected by the relation:

$$(5.4) \quad K^+(s, s') + M^+(s, s') + \int_s^{s'} K^+(s, s'') M^+(s'', s') ds'' = 0.$$

In a half plane $s > s'$, following integral equations hold:

$$(5.5) \quad K^+(s, s') + F(s, s') + \int_s^\infty K^+(s, s'') F(s'', s') ds'' = 0,$$

$$(5.6) \quad M^-(s, s') + F(s, s') + \int_{s'}^\infty F(s, s'') M^-(s'', s') ds'' = 0.$$

Let I_i be a set of projective operators acting in n -dimensional linear space and satisfying the conditions

$$(5.7) \quad I_i I_j = I_i \delta_{ij}.$$

This set can be interpreted as a matrix,

$$I_i = \text{diag}(0, \dots, \underbrace{1}_i, \dots, 0).$$

One can introduce a set of differential operators,

$$(5.8) \quad D_i F = \frac{\partial}{\partial u^i} + I_i \frac{\partial F}{\partial s} + \frac{\partial F}{\partial s'} I_i,$$

and require

$$(5.9) \quad D_i F = 0.$$

We assume also that $1 + F$ is an invertible operator. In other words, the relation

$$(5.10) \quad (1 + F)\Omega = 0$$

implies $\Omega = 0$. If conditions (5.9), (5.10) are fulfilled, the kernels K^+ and M^- satisfy equations

$$(5.11) \quad \begin{aligned} \tilde{D}_i K^+ &= D_i K^+ + [I_i, Q] K^+ = 0, \\ \hat{D}_i M^- &= D_i M^- + M^- [I_i, Q] = 0, \end{aligned}$$

or

$$(5.12) \quad \begin{aligned} \frac{\partial K^+}{\partial u^i} + I_i \frac{\partial K^+}{\partial s} + \frac{\partial K^+}{\partial s'} I_i + [I_i, Q(s)] K^+(s, s') &= 0, \\ \frac{\partial M^-}{\partial u^i} + I_i \frac{\partial M^-}{\partial s} + \frac{\partial M^-}{\partial s'} I_i - M^-(s, s') [I_i, Q(s')] &= 0. \end{aligned}$$

Here $Q(s) = K^+(s, s) = M^-(s, s)$. Moreover, matrix elements $Q_{ij}(s, u)$, ($i \neq j$) satisfy equation (3.10),

$$(5.13) \quad \frac{\partial Q_{ij}}{\partial u^k} = Q_{ik} Q_{kj}.$$

We will not prove this fact, which is almost trivial. Another two points are important:

1. Let $\phi_i = \phi_i(s - u_i)$ be a set of n arbitrary functions of one variable. Then a set of functions,

$$(5.14) \quad H_i(s, u) = \phi_i + \int_s^\infty K_i^+(s, s', u_i) \phi_k(s' - u_k),$$

presents the set of Lamé coefficients H_i at any s . In other words, the metric

$$ds^2 = \sum_{i=1}^n H_i^2 (du^i)^2,$$

is a metric of certain space of diagonal curvature. A different choice of ϕ_i gives different Combescure equivalent metrics.

2. Let the following set of functions,

$$(5.15) \quad \Psi_i(s', u) = \psi_i + \int_s^\infty \phi_k(s' - u_k) M_{k_1}^-(s', s) ds',$$

be a set of adjoint Lamé coefficients satisfying equation (2.18). Then the metric

$$ds^2 = \sum_{i=1}^N \Psi_i^2 (du^i)^2$$

presents the adjoint space of diagonal curvature. In Egorov's case, $F(s, s')$ satisfy the additional constrain,

$$(5.16) \quad F_{ik}(s, s') = F_{ki}(s', s).$$

In this case $H_i = \Psi_i$.

6. Differential reduction

Following the approach of the dressing method, we will call $F(s, s', u)$, satisfying equation (5.8), a "dressing function". Relation (5.16) is an example of additional constrain, which could be imposed on F without violation of basic equation (3.10). One can easily check that relation (5.16) implies the relation

$$(6.1) \quad M^-(s, s') = [K^+(s', s)]^{tr},$$

or

$$(6.2) \quad M_{ij}^-(s, s') = K_{ji}^+(s', s).$$

In this case,

$$(6.3) \quad Q_{ij}(s, u) = Q_{ji}(s, u).$$

The additional relations imposed on F can be called reductions. The key role in the theory of n -orthogonal coordinate systems plays the differential reduction,

$$(6.4) \quad \frac{\partial F(s, s')}{\partial s'} + \frac{\partial F^{tr}(s', s)}{\partial s} = 0,$$

which was introduced in [8]. One can note that this reduction leads to an algebraic relation, which connects the kernels K^+ and M^- . Also, it proves that differential constraint (6.4) stipulates the differential relation between K^+ and M^- . Omitting a very simple proof, we will just formulate this relation:

$$(6.5) \quad \left(\frac{\partial K^+(s', s)}{\partial s} \right)^{tr} + \frac{\partial M^-(s, s')}{\partial s'} + M^-(s, s') [Q(s') - Q^{tr}(s')] = 0.$$

Using this relation one can prove that on the diagonal $s = s'$,

$$(6.6) \quad \left. \frac{\partial K^+(s, s')}{\partial s'} + \frac{\partial K^{+tr}(s', s)}{\partial s} \right|_{s'=s} = -Q(s) Q^{tr}(s'),$$

and finally, that Q_{ij} satisfy additional equations,

$$(6.7) \quad \frac{\partial Q_{ij}}{\partial u^j} + \frac{\partial Q_{ji}}{\partial u^i} + \sum_{k \neq i, j} Q_{ik} Q_{jk} = 0.$$

So, relation (6.4) is the reduction, which gives a solution for the problem of construction of n -orthogonal coordinate systems.

7. One-soliton solution; a general case.

The most striking point of the dressing method is an opportunity to construct exact solutions of integrable systems in a close form. Let us consider a general case of spaces of diagonal curvature and suppose that the "dressing" function F is a product of two matrixes:

$$(7.1) \quad F(s, s', u) = A(s - u) B(s' - u).$$

We will call the corresponding solution of equation (3.19) as "one-soliton solution". In fact, this is a very complicated solution, which includes many interesting Riemann spaces. To satisfy the basic condition (5.8) one should put

$$(7.2) \quad \begin{aligned} A_{ij} &= A_{ij}(s - u^i), \\ B_{ij} &= B_{ij}(s' - u^j). \end{aligned}$$

Integral equations (5.5) and (5.6) can be solved immediately:

$$(7.3) \quad \begin{aligned} K^+(s, s') &= K^+(s) B(s') = \\ M^-(s, s') &= A(s) M^-(s'). \end{aligned}$$

Here

$$(7.4) \quad \begin{aligned} K^+(s) &= -A(s) (1 + \Delta(s))^{-1}, \\ M^-(s') &= -(1 + \Delta(s'))^{-1} B(s'), \end{aligned}$$

$$(7.5) \quad \Delta_{ik} = \sum_p \int_{-\infty}^{u^p - s} B_{ip}(\xi) A_{pk}(\xi) d\xi.$$

The rotation coefficients Q_{ij} take the form

$$(7.6) \quad \begin{aligned} Q_{ij} &= Q_{ij}(u^1 - s, \dots, u^n - s) = \\ &= -A_{ip} [(1 + \Delta)^{-1}]_{pq} B_{qj} \Big|_{s=s'} \end{aligned}$$

In this equation one can put $s = 0$. In a general case, Q_{ij} is parametrized by $2n^2$ functions of one variable.

Let $f_i = f_i(u^i - s)$ be a set of arbitrary functions of one variable. All Combescure equivalent metrics of spaces of diagonal curvature corresponding to the rotation coefficients (7.6) are given as

$$(7.7) \quad H_i = f_i - A_{ip} [(1 + \Delta)^{-1}]_{pq} R_q,$$

where

$$(7.8) \quad R_q = \sum_k \int_{-\infty}^{u^k - s} B_{qk}(\xi) f_k(\xi) d\xi.$$

For the given Q_{ij} , they are parametrized by functions f_n .

The adjoint Lamé coefficients Ψ_i are parametrized by another set of functions on one variable, $g_k(u^k - s)$. Now

$$(7.9) \quad \Psi_i = g_i - T_p [(1 + \Delta)^{-1}]_{pq} B_{qi},$$

$$(7.10) \quad T_p = \sum_k \int_{-\infty}^{u^p - s} g_k(\xi) A_{kp}(\xi) d\xi.$$

A general solution of the Laplace equations (2.21), (2.22) is given in the form

$$(7.11) \quad h = h_0 - T_p [(1 + \Delta)^{-1}] R_q,$$

where

$$(7.12) \quad h_0 = \sum_k \int_{-\infty}^{u^k - s} f_k(\xi) g_k(\xi) d\xi.$$

Diagonal elements of the curvature tensor can be presented in the form

$$(7.13) \quad \begin{aligned} E_{ij} &= \frac{\partial Q_{ij}}{\partial u^j} + \frac{\partial Q_{ji}}{\partial u^i} + \sum_{k \neq i, j} Q_{ik} Q_{jk} = \\ &= -A_{ip} [(1 + \Delta)^{-1}]_{pq} B'_{qj} - A_{jp} [(1 + \Delta)^{-1}]_{jp} B'_{qi} + \\ &+ A_{ip} A_{jk} [(1 + \Delta)^{-1}]_{pl} [(1 + \Delta)^{-1}]_{qm} B_{lk} B_{mk}. \end{aligned}$$

8. One-soliton solutions; special cases.

One-soliton solutions exist for all special classes of spaces of diagonal curvature. They are separated from general one-soliton solutions by imposing of some constrain on matrixes $A(s), B(s)$.

The most interesting case is a flat Euclidean space R^n . In this case A and B are connected:

$$(8.1) \quad A = B^{tr'} \Lambda \text{ or } A_{ij} = B'_{ki} \Lambda_{kj}.$$

Here $\Lambda_{kj} = -\Lambda_{jk}$ are constant antisymmetric matrixes. One can check directly (while it is a hard procedure) that, if conditions (8.1) are satisfied, $E_{ij} = 0$ and the metric is flat.

The problem of embedding of the n -orthogonal metric to R^n can be solved efficiently in the soliton case. Suppose, that the potential h satisfying the Laplace equation (2.23), satisfy also the equation

$$(8.2) \quad \frac{\partial^2 h}{(\partial u^l)^2} = \sum_k \Gamma_{ll}^k \frac{\partial h}{\partial u^k}.$$

In virtue of (2.20), equation (8.2) is equivalent to the relation

$$(8.3) \quad \frac{\partial \Psi_l}{\partial x_l} = - \sum_{k \neq l} Q_{lk} \Psi_k,$$

which is compatible with the definition of Ψ_l

$$(8.4) \quad \frac{\partial \Psi_l}{\partial u^k} = Q_{lk} \Psi_k,$$

if and only if the metric is flat and $E_{ik} \equiv 0$.

Apparently, equation (8.4) is satisfied only for a very special choice of function $g_i(\xi)$. The following theorem holds:

THEOREM 8.1. *Equations (8.3), (8.4) are satisfied if and only if $g_i(\xi)$ are constants.*

The proof of this theorem is straightforward but cumbersome. It will be published in another article. It should be noted that solutions h in (2.23), (8.3) are defined up to an arbitrary constant. If one takes $n - 1$ arbitrary common solution of this system, equations

$$(8.5) \quad h^i(u^1, \dots, u^n) = c^i, \quad i = 1, \dots, n - 1,$$

define all possible geodesics (straight lines) in R^n . By a proper orthogonalization we could define $x^i = x^i(u^1, \dots, u^n)$ and accomplish introducing of n -orthogonal coordinates for the soliton case. In the same way, one can find the constrains connecting A and B for the spaces of constant curvature and for the spaces of flat connection.

It is a much more difficult problem to determine reductions, which separate solutions of Einstein equations for the given equation of matter state. We discuss here the vacuum case only. In this case one should satisfy equations (4.10), (4.15).

Let us study an infinitesimal dressing $F \rightarrow 0$. In this limit

$$(8.6) \quad Q(u^1 - s, \dots, u^n - s) = K(s, s) = -F(s, s),$$

and

$$(8.7) \quad -E_{ij} = \frac{\partial F_{ij}(u^i - s, u^j - s')}{\partial s'} + \frac{\partial F_{ji}(u^j - s', u^i - s)}{\partial s}.$$

At $s = 0$, $E_{ij} = E_{ij}(u^i, u^j)$. According to (4.10) and (4.15), $E_{ij} = \alpha_{ij} H_i H_j$, and

$$\alpha_{12} = \alpha_{34} = \alpha, \quad \alpha_{13} = \alpha_{23} = \beta, \quad \alpha_{14} = \alpha_{23} = \gamma.$$

As far as we have $H_i = H_i(u^i)$, $H_j = H_j(u^j)$ in the linear approximation, α_{ij} must be constants. Finally, the dressing matrix F should satisfy the condition

$$(8.8) \quad \frac{\partial F_{ij}(u^i - s, u^j - s')}{\partial s'} + \frac{\partial F_{ji}(u^j - s', u^i - s)}{\partial s} = \alpha_{ij} H_i(u^i - s) H_j(u^j - s'),$$

where $\alpha_{ij} = \alpha_{ji}$ is a symmetric constant matrix satisfying conditions (4.15).

One must remember that conditions (8.8) are necessary. The sufficient conditions, which must be imposed on F to find the exact solution of Einstein equations in vacuum, are unknown so far.

9. Dressing via $\bar{\partial}$ -problem

The dressing procedure described above will lead to the same results, if one performs the following replacement in (5.5), (5.6):

$$(9.1) \quad \int_s^\infty u \, ds \rightarrow \frac{1}{2} \left(\int_s^\infty u \, ds - \int_{-\infty}^s u \, ds \right).$$

Let us introduce:

$$(9.2) \quad F(s, s') = \int F(\lambda, \mu) e^{-i(\lambda s - \mu s')} \, d\lambda \, d\mu,$$

$$(9.3) \quad K^+(s, s') = \int K(s, \mu) e^{i\mu s'} \, d\mu,$$

$$(9.4) \quad M^+(s, s') = \int M(\lambda, s') e^{-i\lambda s} \, d\lambda.$$

One can check that equations (5.5), (5.6) are equivalent to integral equations:

$$(9.5) \quad \phi(s, \lambda) + \int_{-\infty}^\infty f(\eta, \lambda) \, d\eta - i \int_{-\infty}^\infty \frac{\phi(s, \xi)}{\eta - \xi} f(\eta, \lambda) \, d\xi \, d\eta = 0,$$

$$(9.6) \quad \psi(\lambda, s') + \int_{-\infty}^\infty f(\lambda, \eta) \, d\eta + i \int_{-\infty}^\infty \frac{f(\lambda, \xi)}{\xi - \eta} \psi(\eta, s') \, d\xi \, d\eta = 0,$$

where

$$(9.7) \quad \begin{aligned} \phi(s, \lambda) &= K(s, \lambda) e^{i\lambda s}, \\ \psi(\lambda, s') &= M(\lambda, s') e^{-i\lambda s'}, \\ f(\lambda, \eta) &= F(\lambda, \eta) e^{i(\lambda - \eta)s}. \end{aligned}$$

Later on we will omit sometimes the notations s, s' . In equations (9.5), (9.6)

$$(9.8) \quad \frac{1}{\eta - \xi} = \lim_{\epsilon \rightarrow 0} \frac{\bar{\eta} - \bar{\xi}}{|\eta - \xi|^2 + \epsilon^2}.$$

Taking into account (9.8), one can understand the integration performed in (9.5), (9.6) in a more broad sense. So far, all functions are defined in the real axis $-\infty < \lambda < \infty$, $-\infty < \eta < \infty$. The dressing procedure does not change if one assumes that ϕ, λ are quasianalytic functions defined on the whole complex plane C^1 . In this case, $F(\eta, \lambda)$ is defined on C^2 . One can define:

$$(9.9) \quad \begin{aligned} \chi(\lambda, \bar{\lambda}) &= 1 - i \int \frac{\phi(\xi, \bar{\xi})}{\lambda - \xi} \, d\xi \, d\bar{\xi}, \\ \bar{\xi}(\lambda, \bar{\lambda}) &= 1 + i \int \frac{\psi(\eta, \bar{\eta})}{\lambda - \eta} \, d\eta \, d\bar{\eta}. \end{aligned}$$

Using the standard formula,

$$(9.10) \quad \frac{\partial}{\partial \bar{\lambda}} \frac{1}{\lambda} = \pi \delta(\lambda) \delta(\bar{\lambda}),$$

one obtains

$$(9.11) \quad \phi(\lambda, \lambda) = -\frac{1}{\pi i} \frac{\partial \chi}{\partial \bar{\lambda}}, \quad \psi(\lambda, \bar{\lambda}) = \frac{1}{\pi i} \frac{\partial \bar{\chi}}{\partial \lambda}.$$

Quasianalytic functions $\chi, \bar{\chi}$ satisfy the equations:

$$(9.12) \quad \frac{\partial \chi}{\partial \bar{\lambda}} = -\pi i \int \chi(\xi, \bar{\xi}) f(\xi, \bar{\xi}, \lambda \bar{\lambda}) d\xi d\bar{\xi},$$

$$(9.13) \quad \frac{\partial \bar{\chi}}{\partial \lambda} = \pi i \int f(\lambda, \bar{\lambda}, \xi, \bar{\xi}) \bar{\chi}(\xi, \bar{\xi}) d\xi d\bar{\xi},$$

which define the dual "non-local $\bar{\partial}$ -problem", and are accomplished by the following normalization:

$$\chi \rightarrow 1, \quad \bar{\chi} \rightarrow 1 \quad \text{at} \quad \lambda \rightarrow \infty.$$

In the limit $\lambda \rightarrow \infty$ one has:

$$(9.14) \quad \chi = 1 + \frac{Q}{i\lambda} + \frac{P}{(i\lambda)^2} + \frac{R}{(i\lambda)^3} + \dots,$$

$$(9.15) \quad Q = K^+(s, s) = \int K(s, \mu) e^{i\mu s} d\mu = \int \phi(\xi, \bar{\xi}) d\xi d\bar{\xi},$$

$$(9.16) \quad P = \left. \frac{\partial K}{\partial s} \right|_{s=s'} = i \int \xi \phi(\xi, \bar{\xi}) d\xi d\bar{\xi}.$$

In the same way,

$$(9.17) \quad \bar{\chi} = 1 - \frac{Q}{i\lambda} + \frac{\bar{P}}{(i\lambda)^2} + \dots$$

Equation (5.8) imposes on function $f(\lambda, \eta)$ the following condition:

$$(9.18) \quad f(\lambda, \eta) = e^{i\lambda\Phi} F(\lambda, \eta) e^{-i\eta\Phi},$$

$$(9.19) \quad \Phi = s + \sum_{i=1}^n I_i u^i.$$

In virtue of (9.12) and (9.13), functions $\chi, \bar{\chi}$ satisfy the linear systems:

$$(9.20) \quad I_k \left(\frac{\partial \chi}{\partial u^i} + i\lambda \chi I_i - Q I_i \chi \right) = 0, \quad i \neq k,$$

$$(9.21) \quad \left(\frac{\partial \bar{\chi}}{\partial u^i} - i\lambda I_i \bar{\chi} - \bar{\chi} I_i Q \right) I_k = 0, \quad i \neq k.$$

Let us expand equation (9.20) in powers of $1/i\lambda$. The first term of the expansion reads:

$$(9.22) \quad I_k \frac{\partial Q}{\partial u^i} - I_k Q I_i Q + I_k P = 0.$$

Multiplication by I_j ($j \neq i, k$) from the right side gives

$$(9.23) \quad I_k \frac{\partial Q}{\partial u^i} I_j = I_k Q I_i Q I_j.$$

This is just another notation for equation (5.13). Multiplication of (9.22) by I_i leads to definition of P ,

$$(9.24) \quad I_k P I_i = I_k Q I_i Q I_i - I_k \frac{\partial Q}{\partial u^i} I_i.$$

The second term of the expansion leads to the equation:

$$(9.25) \quad I_k \frac{\partial P}{\partial u^i} I_i = I_k Q I_i P I_j.$$

Let us introduce

$$\chi = X e^{-i\lambda\Phi}, \quad \tilde{\chi} = e^{i\lambda\Phi} Y.$$

Apparently, X and Y satisfy equations

$$(9.26) \quad I_k \frac{\partial X}{\partial u^i} = I_k Q I_i X,$$

$$(9.27) \quad \frac{\partial Y}{\partial u^k} I_i = Y I_i Q I_k.$$

Suppose, that $\phi_i = \phi_i(\lambda, \bar{\lambda})$, $\psi_i = \psi_i(\lambda, \bar{\lambda})$ are two arbitrary sets of functions of some complex variable, not necessary analytical one. One can see that the sets

$$(9.28) \quad H_i(u^1, \dots, u^n) = \sum_k \int X_{ik}(u^1, \dots, u^n, \xi, \bar{\xi}) \phi_k(\xi, \bar{\xi}) d\xi d\bar{\xi},$$

$$(9.29) \quad \Psi_i(u^1, \dots, u^n) = \sum_k \int Y(u^{1*}, \dots, u^n, \xi, \bar{\xi}) \psi_i(\xi, \bar{\xi}) d\xi d\bar{\xi},$$

give the arrays of Lamé coefficients and adjoint Lamé coefficients.

Now the reductions are some restrictions imposed on $f(\lambda, \eta)$. The fundamental reduction (6.4) reads:

$$(9.30) \quad \eta F(\lambda, \eta) = \lambda F^{tr}(-\eta, -\lambda).$$

Let us impose a more general reduction,

$$(9.31) \quad \eta F(\lambda, \eta) - \lambda F^{tr}(-\eta, -\lambda) = e^{i\lambda\Phi} R(\lambda, \eta) e^{-i\lambda\Phi},$$

where

$$(9.32) \quad R(\lambda, \eta) = R^{tr}(-\eta, -\lambda)$$

is some matrix function on C^2 and does not depend on u^i . From (9.12) one gets:

$$(9.33) \quad \frac{\partial}{\partial \lambda} \chi^{tr}(-\lambda, -\bar{\lambda}) = \pi i \int f^{tr}(-\xi, -\bar{\xi}, -\lambda, -\bar{\lambda}) \chi^{tr}(-\xi, -\bar{\xi}) d\xi d\bar{\xi}.$$

Using (9.12), (9.32) and (9.30) one can obtain the following bilinear identity:

$$(9.34) \quad \int \frac{\partial}{\partial \lambda} \lambda \chi(\lambda, \bar{\lambda}) \chi^{tr}(-\lambda, -\bar{\lambda}) d\lambda d\bar{\lambda} = -\pi i \int \chi(\xi, \bar{\xi}) e^{i\xi\Phi} R(\xi, \eta) e^{-i\eta\Phi} \chi^{tr}(-\eta, -\bar{\eta}) d\xi d\bar{\xi} d\eta d\bar{\eta}.$$

The integral in the left part of (9.34) is proportional to the residue of $\lambda \chi \chi^{tr}$ at infinity. Using the asymptotic expansion (9.14) one gets:

$$(9.35) \quad \int \frac{\partial}{\partial \lambda} \lambda \chi(\lambda, \bar{\lambda}) \chi^{tr}(-\lambda, -\bar{\lambda}) d\lambda d\bar{\lambda} = 2\pi i (P - P^{tr} - QQ^{tr}).$$

Suppose, that $R(\lambda, \eta)$ is presented in the form:

$$(9.36) \quad R_{ij}(\lambda, \eta) = -2 \sum_{k=1}^N \phi_i^{(k)}(\lambda, \bar{\lambda}) \phi_j^{(k)}(-\eta, -\bar{\eta}) (-1)^{\alpha(k)}.$$

Here $\phi_i^{(k)}(\lambda, \bar{\lambda})$, $i = 1, \dots, N$ are some functions on C^1 , and $\alpha(k) = 0, 1$. Then formula (9.34) reads:

$$(9.37) \quad P - P^{tr} - QQ^{tr} = \sum_{k \pm 1}^N H_i^{(k)} H_j^{(k)} (-1)^{\alpha(k)}.$$

Here

$$(9.38) \quad H_i^k = \sum_q \chi_{iq}(\xi, \bar{\xi}) e^{i\xi u^k} \Psi_q^k(\xi, \bar{\xi}) d\xi d\bar{\xi}.$$

H_i^k are the arrays of Combescure equivalent metrics at different k . Combining (9.36) with (9.35), one easily obtains:

$$(9.39) \quad E_{ij} = \sum_k H_i^{(k)} H_j^{(k)} (-1)^{\alpha(k)}.$$

In other words, reduction (9.31) describes a space of flat connection. If $N \rightarrow \infty$, one can present any matrix function $R_{ij}(\lambda, \eta)$ in the form (9.36). That means that the spaces of flat connection are dense in the set of all spaces of diagonal curvature.

There are some other, more sophisticated methods that allow to separate the spaces of flat connection from the total pool of diagonal curvature spaces. One of them is describes in [14].

Most probably, the Einstein spaces of diagonal curvature are among the spaces of flat connection.

References.

- (1) D. Kaup, The three-wave interaction — a non-dispersive phenomenon. *Studies Appl. Math.*, **55** (1976), pp. 9–44.
- (2) V. Zakharov and A. Shabat, An integration scheme for the nonlinear equations of mathematical physics by the method of the Inverse scattering problem. *Funk. Anal. Appl.*, **8** (1974), pp. 43–54, 226–235.
- (3) V. Zakharov and S. Manakov, The theory of resonant interaction of wave packets in nonlinear media. *Zh. Eksper. Teor. Fiz.*, **69** (1975), pp. 1654–1673; *Soviet Phys. JETP*, **42** (1976), pp. 842–850.
- (4) G. Darboux, *Lecons sur le Systemes Orthogonaux et les Coordonnees Curvilignes*, Gauthier–Villars, Paris, 1910.
- (5) B. Dubrovin, Integrable systems in topological field theory. *Nuclear Phys.*, **B 379** (1992), pp. 627–689; On the differential geometry of strongly integrable systems of hydrodynamic type, *Funk. Anal. Appl.* **24** (1990), pp. 280–285.
- (6) B. Dubrovin and S. Novikov, Hamiltonian formalism of one-dimensional systems of hydrodynamic type and the Bogolyubov–Whitham averaging method. *Dokl. Acad. Nauk SSSR* **270** (1983), pp. 781–785.
- (7) S. Tsarev, Poisson’s brackets and one-dimensional Hamiltonian systems of hydrodynamic type, *Dokl. Akad. Nauk SSSR*, **282** (1985), pp. 534–537.
- (8) V. Zakharov, Description of n -orthogonal curvilinear coordinate systems and Hamiltonian integrable systems of hydrodynamic type. *Duke Mathematical J.*, **94:1** (1998), pp. 103–139.
- (9) V. Zakharov and S. Manakov, Reductions in systems integrable by the method of inverse scattering problem. *Doclady Math.* **57:3** (1998), pp. 471–474.

- (10) V. Zakharov, How classical physics helps mathematics. GAFA Geometrical and Functional Analysis, Special Volume GAFA 2000, pp. 859–879.
- (11) E. Ferapontov, Private communication.
- (12) V. Zakharov, On the dressing method, in "Inverse methods in action", Springer–Verlag, Berlin, 1990, pp. 602–623.
- (13) V. Zakharov and S. Manakov, Construction of multidimensional nonlinear integrable systems and their solitons. Funk. Anal. Appl. **19** (1985), pp. 11–25.
- (14) V. Zakharov, Integrating of the Gauss–Codazzi equations. Theor. Math. Phys. **128** (2001), pp. 946–956.

LANDAU INSTITUTE FOR THEORETICAL PHYSICS, MOSCOW, 117334, RUSSIA;
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ARIZONA, TUCSON, AZ 85721, USA
E-mail address: `zakharov@itp.ac.ru`, `zakharov@math.arizona.edu`

Algebraic and Analytic Aspects of Soliton Type Equations

Vladimir S. Gerdjikov

ABSTRACT. This is a review of two of the fundamental tools for analysis of soliton equations: i) the algebraic ones based on Kac-Moody algebras, their central extensions and their dual algebras which underlie the Hamiltonian structures of the NLEE; ii) the construction of the fundamental analytic solutions (FAS) of the Lax operator and the Riemann-Hilbert problem (RHP) which they satisfy. The fact that the inverse scattering problem for the Lax operator can be viewed as a RHP gave rise to the dressing Zakharov-Shabat, one of the most effective ones for constructing soliton solutions. These two methods when combined may allow one to prove rigorously the results obtained by the abstract algebraic methods. They also allow to derive spectral decompositions for non-self-adjoint Lax operators.

1. INTRODUCTION

We start with three examples of integrable nonlinear evolution equations (NLEE). The first one is the well known N -wave equation [49, 48, 35]:

$$(1.1) \quad i[I, Q_t] - i[J, Q_x] + [[I, Q], [J, Q(x, t)]] = 0, \quad \lim_{x \rightarrow \pm\infty} Q(x, t) = 0,$$

where $Q(x, t)$ is a smooth $n \times n$ matrix-valued function, $Q(x, t) = -BQ^\dagger B$ and I and J are constant diagonal matrices; $B_{ij} = \delta_{ij}\epsilon_j$, $\epsilon_j = \pm 1$.

The second example is the 2-dimensional affine Toda chain [41]:

$$(1.2) \quad \frac{\partial^2 Q_k}{\partial x \partial t} = e^{Q_{k+1} - Q_k} - e^{Q_k - Q_{k-1}}, \quad k = 1, \dots, n,$$

where we assume that $e^{Q_{n+1}} \equiv e^{Q_1}$.

The third example belongs to the same family as (1.2) and is of the form:

$$(1.3) \quad i \frac{\partial \psi_k}{\partial t} + \gamma \coth \frac{\pi k}{n} \frac{\partial^2 \psi_k}{\partial x^2} + i\gamma \sum_{p=1}^{n-1} \frac{d}{dx} (\psi_p \psi_{k-p}) = 0, \quad k = 1, \dots, n,$$

and $k - p$ is understood modulo n and $\psi_0 = \psi_n = 0$.

2000 *Mathematics Subject Classification.* Primary 37K15, 17B70; Secondary 37K10, 17BB80.
Key words and phrases. graded algebras, soliton equations.

Financial support in part from Gruppo collegato di INFN at Salerno, Italy and project PRIN 2000 (contract 323/2002) is acknowledged.

© 2002 American Mathematical Society

The integrability of these equations is based on their Lax representations. This means that each of the NLEE (1.1)–(1.3) can be represented as the compatibility condition

$$(1.4) \quad [L(\lambda), M(\lambda)] = 0,$$

of two linear matrix differential operators depending on the spectral parameter λ . Below we will use as Lax operator $L(\lambda)$

$$(1.5) \quad L(\lambda)\psi(x, t, \lambda) = \left(i \frac{d}{dx} + q(x, t) - \lambda J \right) \psi(x, t, \lambda) = 0;$$

as examples of $M(\lambda)$ -operators we use:

$$(1.6a) \quad M(\lambda)\psi = \left(i \frac{d}{dt} + V_0(x, t) + \lambda I \right) \psi(x, t, \lambda) = \lambda \psi(x, t, \lambda) I;$$

$$(1.6b) \quad M_1(\lambda)\psi = \left(i \frac{d}{dt} + V_0(x, t) + \lambda V_1(x, t) + \lambda^2 V_2 \right) \psi(x, t, \lambda) = \lambda^2 \psi(x, t, \lambda) V_2^{\text{as}};$$

$$(1.6c) \quad M_2(\lambda)\psi = \left(i \frac{d}{dt} + V_0(x, t) + \frac{1}{\lambda} V_{-1}(x, t) \right) \psi(x, t, \lambda) = \frac{1}{\lambda} \psi(x, t, \lambda) V_{-1}^{\text{as}};$$

where $V_2^{\text{as}} = \lim_{x \rightarrow \pm\infty} V_2(x, t)$ and $V_{-1}^{\text{as}} = \lim_{x \rightarrow \pm\infty} V_{-1}(x, t)$.

Choosing the form of $L(\lambda)$ in (1.5) we fixed up the gauge by assuming that J is constant diagonal matrix and $q(x, t) = [J, Q(x, t)]$, i.e. $q_{jj} = 0$.

The system (1.5) with $q(x, t)$ and J 2×2 -matrices (i.e., $\mathfrak{g} \simeq sl(2)$) is known as the Zakharov-Shabat (ZSs) system; the same system with $n \times n$ matrices will be referred to below as the generalized Zakharov-Shabat system (GZSs).

The Lax representation of the N -wave equation is provided by $L(\lambda)$ (1.5) and $M(\lambda)$ (1.6a). If the potentials in these operators are $n \times n$ -matrix ones we may assume that the Lie algebra underlying the Lax representation is $\mathfrak{g} \simeq sl(n)$. The set of independent fields $Q_{ij}(x, t)$ equals $n(n - 1)$ and may be restricted by the involution [49, 52, 48]:

$$(1.7) \quad q(x, t) = Bq^\dagger(x, t)B^{-1}, \quad J = J^\dagger,$$

Often by N -wave equations in the literature people mean eq. (1.1) with the involution (1.7). Such systems with $n = 3$ and $n = 4$ find applications in describing wave-wave interactions, see [48, 49, 35].

The Lax representation of the \mathbb{Z}_n -NLS eq. (1.3) is provided by (1.5) and (1.6b) but with rather specific restrictions imposed on $q(x, t)$ and J :

$$(1.8) \quad \tilde{q}(x, t) = i \sum_{k=1}^n \psi_k(x, t) K_0^k, \quad \tilde{J} = c_0 \sum_{k=1}^n \omega^{k-1/2} E_{kk}, \quad K_0 = \sum_{k=1}^n E_{k,k+1},$$

Here and below we will denote by E_{jk} the $n \times n$ -matrices equal to $(E_{jk})_{mn} = \delta_{jm} \delta_{kn}$; the indices should be taken modulo n , i.e. $E_{n,n+1} \equiv E_{n,1}$ and the constant c_0 will be defined below.

The affine Toda chain (1.2) has several equivalent Lax representations. We mention here two of them. Their Lax operators are:

$$(1.9) \quad \tilde{L}_{\text{Toda}} = i \frac{d}{dx} - \frac{i}{2} \sum_{k=1}^n \frac{dQ_k}{dx} E_{kk} + i\lambda \sum_{k=1}^n e^{(Q_{k+1} - Q_k)/2} E_{k,k+1},$$

and its gauge equivalent:

$$(1.10) \quad \tilde{L}_{\text{ Toda}} = i \frac{d}{dx} - i \sum_{k=1}^n \frac{dQ_k}{dx} E_{kk} + i\lambda K_0.$$

The corresponding M -operators are of the form (1.6c). Both choices (1.9) and (1.10) are not of the form (1.5), but are adjusted to the grading of the Lie algebra $sl(n, \mathbb{C})$ we introduce in the next subsection, see formulae (1.20)–(1.25) below.

The operator $\tilde{L}_{\text{ Toda}}$ (1.10) after a similarity transformation with the constant matrix u_0 , such that $u_0^{-1} K_0 u_0 = \sum_{k=1}^n \omega^k E_{kk}$ can be cast into the form of (1.5) in which both $q(x, t)$ and J have a special form:

$$(1.11) \quad \tilde{q}(x, t) = -i \sum_{k=1}^n \frac{dQ_k}{dx} \omega^{kp} K_0^p, \quad \tilde{J} = c_0 \sum_{j=1}^n \omega^{k-1/2} E_{kk}.$$

where $\omega = \exp(2\pi i/n)$. The special form of $q(x, t)$ and J in both (1.8) and (1.11) shows that both models have only $n - 1$ independent fields. This special form can also be made compatible with the structure of the graded and Kac-Moody algebras [11, 33, 31] and is best understood with the method of the reduction group proposed by Mikhailov [41].

The idea of the ISM is based on the possibility to linearize the NLEE [53, 2, 9, 13, 49, 48, 35]. To this end we consider the solution to the NLEE $q(x, t)$ as a potential in $L(\lambda)$ (1.5). In order to solve the direct scattering problem for $L(\lambda)$ we introduce the Jost solutions $\psi_{\pm}(x, t, \lambda)$ and the scattering matrix $T(\lambda, t)$ as follows:

$$(1.12) \quad i \frac{d\psi_{\pm}}{dx} + (q(x, t) - \lambda J)\psi_{\pm}(x, t, \lambda) = 0,$$

$$(1.13) \quad \lim_{x \rightarrow \pm\infty} \psi_{\pm}(x, t, \lambda) e^{i\lambda Jx} = \mathbf{1},$$

$$(1.14) \quad T(\lambda, t) = \psi_+^{-1} \psi_-(x, t, \lambda).$$

The Jost solutions of $L(\lambda)$ are also eigenfunctions of the operator $M(\lambda)$. We can use this fact to determine the t -dependence of the scattering matrix:

$$(1.15) \quad i \frac{dT}{dt} + [f(\lambda), T(\lambda, t)] = 0,$$

which can be easily solved as follows:

$$(1.16) \quad T(\lambda, t) = e^{if(\lambda)t} T(\lambda, 0) e^{-if(\lambda)t}.$$

By $f(\lambda) \in \mathfrak{h}$ above we mean the dispersion law of the NLEE; for the N -wave system we have $f_{N-w}(\lambda) = \lambda I$.

Thus the solution of the NLEE for a given initial condition $q(x, t)|_{t=0} = q_0(x)$ can be performed in three steps, see [48, 9, 13]:

- (1) insert $q(x, 0)$ as a potential in $L(\lambda)$ and determine the corresponding scattering matrix $T(\lambda, 0)$;
- (2) Given $T(\lambda, 0)$ and the dispersion law $f(\lambda)$ find $T(\lambda, t)$ from eq. (1.16);
- (3) Given $T(\lambda, t)$ reconstruct the corresponding potential $q(x, t)$ of $L(\lambda)$ which will be also the solution of the NLEE.

Step 2 is trivial. Steps 1 and 3 involve solving the direct and inverse scattering problem for (1.5) which can be reduced to linear integral equations. The most difficult step 3 provided for the name of the method. The most effective method to solve it for operators like $L(\lambda)$ is based on the equivalence to a RHP [45].

Along with solving the inverse scattering problem in step 3) we will construct also the minimal set of scattering data \mathfrak{T} . Indeed the scattering matrix $T(\lambda, t)$ has n^2 matrix elements with only one obvious constraint $\det T(\lambda, t) = 1$ while the potential $q(x, t)$ has only $n(n-1)$ matrix elements. Therefore there must be additional interrelations between the matrix elements of $T(\lambda, t)$.

The analysis of the mapping between $q(x, t)$ and \mathfrak{T} allows one to interpret it as a generalized Fourier transform [2, 36, 25, 21, 27, 28, 29, 30]. The proof of all these facts and the effective solving of the ISP for the GZSs (1.5) is based on the possibility to construct fundamental solutions of (1.5) which are section-analytic functions of the spectral parameter λ .

Algebraic structures: Kac-Moody and graded Lie algebras

Let us now briefly outline the first basic tool inherent in the Lax representation – its algebraic structure. Indeed, $L(\lambda)$ and $M(\lambda)$ above are polynomial in λ and/or $1/\lambda$ whose coefficients take values in some simple Lie algebra \mathfrak{g} .

Let us take generic Lax operators in the form:

$$(1.17) \quad L(\lambda)\psi \equiv \left(i \frac{d}{dx} + U(x, t, \lambda) \right) \psi(x, t, \lambda) = 0,$$

$$(1.18) \quad M(\lambda)\psi \equiv \left(i \frac{d}{dt} + V(x, t, \lambda) \right) \psi(x, t, \lambda) = \psi(x, t, \lambda) V^{\text{as}}(\lambda),$$

$$(1.19) \quad U(x, t, \lambda) = \sum_k U_k(x, t) \lambda^k, \quad V(x, t, \lambda) = \sum_k V_k(x, t) \lambda^k,$$

where the potentials $U(x, t, \lambda)$ and $V(x, t, \lambda)$ are polynomials in λ and/or $1/\lambda$. Such potentials can be viewed as elements of a Kac-Moody algebra $\widehat{\mathfrak{g}}_C$. Roughly speaking the construction of $\widehat{\mathfrak{g}}_C$ involves a simple Lie algebra \mathfrak{g} and an automorphism C of finite order, i.e. there exist such an integer h that $C^h = \mathbf{1}$. Then we can split \mathfrak{g} into a direct sum of linear subspaces

$$(1.20) \quad \mathfrak{g} = \bigoplus_{k=0}^{h-1} \mathfrak{g}^{(k)},$$

which are eigensubspaces of C , i.e. if

$$(1.21) \quad X^{(k)} \in \mathfrak{g}^{(k)} \quad \Leftrightarrow \quad C(X^{(k)}) = \omega^{-k} X^{(k)},$$

where $\omega = \exp(2\pi i/h)$. The decomposition (1.20) satisfies the grading condition:

$$(1.22) \quad [X^{(k)}, X^{(m)}] = X^{(k+m)} \in \mathfrak{g}^{(k+m)},$$

where the superscript $k+m$ in $\mathfrak{g}^{(k+m)}$ is understood modulo h . Then the elements of the corresponding Kac-Moody algebra $\widehat{\mathfrak{g}}_C$ have the form:

$$(1.23) \quad X(\lambda) = \sum_{k \leq N_1} \lambda^k X^{(k)}, \quad X^{(k)} \in \mathfrak{g}^{(k)},$$

Obviously due to (1.22) the commutator of any two elements $X(\lambda), Y(\lambda)$ of the form (1.23) will also have the form (1.23).

The classification and the theory of the Kac-Moody algebras can be found in [33, 31]. Their simplest realization can be obtained from a pair (\mathfrak{g}, C) with a few special choices of the automorphism of finite order C , namely:

a) $C = \mathbb{1}$; then each of the subspaces $\mathfrak{g}^{(k)} \simeq \mathfrak{g}$. This leads to a generic GZS system if J is real and to a generic CBC system if J is complex.

b) $C^h = 1$ where C is the Coxeter automorphism of \mathfrak{g} and h is the Coxeter number. This leads to a CBC system with \mathbb{Z}_h -reduction and will be used in analyzing the NLEE (1.2) and (1.3).

c) CV where V is a nontrivial external automorphism of \mathfrak{g} . Such gradings also lead to interesting NLEE but will not be used in this paper.

The Kac-Moody algebras are obtained from the constructions a)-c) with additional central extensions; they are split into three classes: of height 1 (cases a) and b)) and of height 2 and 3 depending on the order of V .

The potential $U(x, t, \lambda)$ for the N -wave equations equals $[J, Q(x, t)] - \lambda J$ belongs to a Kac-Moody algebra with $\mathfrak{g} \simeq sl(n)$ and $C = \mathbb{1}$. The potential $\tilde{U}(x, t, \lambda) = \tilde{q}(x, t) - \lambda \tilde{J}$ of the form (1.8) and (1.11) gives rise to the NLEE (1.2) and (1.3) is related to Kac-Moody algebra of the class b) with $\mathfrak{g} \simeq sl(n)$. The Coxeter number then is $h = n$; the Coxeter automorphism can be realized as inner automorphism of the form:

$$(1.24) \quad C(X) = C_0 X C_0^{-1}, \quad C_0 = \sum_{k=1}^n \omega^k E_{kk}, \quad \omega = e^{2\pi i/n},$$

where C obviously satisfies $C^n = \mathbb{1}$. With this choice of C we can easily check that the linear subspaces $\mathfrak{g}^{(k)}$ are spanned by

$$(1.25) \quad \mathfrak{g}^{(k)} \equiv \text{l.c.} \{E_{j,j+k}, \quad j, k = 1, \dots, n\},$$

and $j + k$ is considered modulo n . Comparing (1.8), (1.10) with (1.25) below we find that $\tilde{q}(x, t) \in \mathfrak{g}^{(0)}$ and $\tilde{J} \in \mathfrak{g}^{(1)}$. Note that now the condition $X^{(k)} \in \mathfrak{g}^{(k)}$ imposes a set of nontrivial constraints on $X^{(k)}$.

The idea to use finite order automorphisms for the reductions of the NLEE was proposed first by Mikhailov [41] who introduced also the notion of the reduction group. The \mathbb{Z}_n -reduction condition according to [41] is introduced by:

$$(1.26) \quad C(\tilde{U}(x, t, \lambda\omega)) = \tilde{U}(x, t, \lambda), \quad C(\tilde{V}(x, t, \lambda\omega)) = \tilde{V}(x, t, \lambda),$$

where we have chosen the simplest possible realization of the \mathbb{Z}_n group on the complex λ -plane: $\lambda \rightarrow \lambda\omega$ with $\omega = \exp(2\pi i/n)$.

The Kac-Moody algebras, like the semi-simple Lie algebras have an important property which ensures the solvability of the inverse scattering problem for $L(\lambda)$ and the non-degeneracy of the Hamiltonian structures of the NLEE. While the semi-simple Lie algebras possess just one invariant bilinear form $\langle X, Y \rangle \equiv \text{tr}(\text{ad}_X, \text{ad}_Y)$ the Kac-Moody algebras possess a family of invariant bilinear forms:

$$(1.27) \quad \langle\langle X(\lambda), Y(\lambda) \rangle\rangle^{(p)} = \text{Res}_{\lambda=0} \lambda^{-p-1} \langle X(\lambda), Y(\lambda) \rangle,$$

for all integer values of p .

We will need also a central extension of the Kac-Moody algebras $\tilde{\mathfrak{g}} = \hat{\mathfrak{g}} \oplus c$ where the central element is generated by the 2-cocycle:

$$(1.28) \quad \omega_p(X(x, \lambda), Y(x, \lambda)) = \int_{-\infty}^{\infty} dx \langle\langle X(x, \lambda), \frac{dY(x, \lambda)}{dx} \rangle\rangle^{(p)}.$$

This means that each element of $\tilde{\mathfrak{g}}$ is a pair $(X(x, \lambda), c_X)$ where c_X is a constant. The commutation relation in $\tilde{\mathfrak{g}}$ is defined by:

$$(1.29) \quad [(X(x, \lambda), c_X), (Y(x, \lambda), c_Y)] = ([X(x, \lambda), Y(x, \lambda)], \omega_p(X(x, \lambda), Y(x, \lambda))).$$

Important role for the Hamiltonian formulation of the NLEE is played by the dual algebras $\hat{\mathfrak{g}}^*$, $\tilde{\mathfrak{g}}^* = \hat{\mathfrak{g}}^* \oplus c$ and their splittings into direct sums of Borel-like subalgebras. These splittings for $\hat{\mathfrak{g}} = \hat{\mathfrak{g}}_+ \oplus \hat{\mathfrak{g}}_-$ look like:

$$(1.30) \quad \hat{\mathfrak{g}}_+ \equiv \left\{ \sum_{k=0}^{N_1} u_k(x) \lambda^k \right\}, \quad \hat{\mathfrak{g}}_- \equiv \left\{ \sum_{k=-\infty}^{-1} u_k(x) \lambda^k \right\},$$

and for the dual $\hat{\mathfrak{g}}^* = \hat{\mathfrak{g}}_+^* \oplus \hat{\mathfrak{g}}_-^*$:

$$(1.31) \quad \hat{\mathfrak{g}}_+^* \equiv \left\{ \sum_{k=-N_1}^p u_k(x) \lambda^k \right\}, \quad \hat{\mathfrak{g}}_-^* \equiv \left\{ \sum_{k=p+1}^{\infty} u_k(x) \lambda^k \right\}.$$

The co-adjoint orbits of $\tilde{\mathfrak{g}}$ on $\tilde{\mathfrak{g}}^*$ in fact are isomorphic to the space of coefficients for which the NLEE is written. Thus they are natural candidate for the phase space of these equations. The freedom provided by the parameter p is directly related to the existence of hierarchy of Hamiltonian structures for the NLEE.

Fundamental analytic solutions

The second important tool in this scheme is the fundamental analytic solution (FAS) of $L(\lambda)$. We will see that using the FAS one is able to:

- reduce the solving of the ISP for $L(\lambda)$ to an equivalent Riemann-Hilbert problem (RHP) for the FAS [45, 48, 52];
- construct the kernel of the resolvent for $L(\lambda)$ and derive the spectral decomposition for $L(\lambda)$ [26, 18, 30];
- construct the ‘squared’ solutions of $L(\lambda)$ which allow the interpretation of the ISM as a generalized Fourier transform (GFT) [2, 34, 36, 25, 28, 29, 30];
- construct the Green function for the recursion operators Λ_{\pm} and prove the completeness relation for the ‘squared’ solutions. This property ensures the uniqueness of the solution of the ISM [25, 27, 19, 30].

The existence of FAS is ensured by the analytic dependence of both $U(x, t, \lambda)$ and $V(x, t, \lambda)$ on λ . The properties of FAS depend crucially on the boundary conditions imposed on the potential $q(x, t)$. For simplicity here we consider the class of potentials $q(x, t)$ that are sufficiently smooth functions of x and tend to zero fast enough for $x \rightarrow \pm\infty$ for any fixed value of t .

The FAS for the Zakharov-Shabat system (i.e. $\mathfrak{g} \simeq sl(2)$) can easily be constructed due to the fact that each of the columns of the Jost solutions

$$(1.32) \quad L(\lambda)\psi_{\pm}(x, t, \lambda) = 0, \quad \lim_{x \rightarrow \pm\infty} e^{iJ\lambda x} \psi_{\pm}(x, t, \lambda) = \mathbf{1},$$

allow analytic extension either for $\lambda \in \mathbb{C}_+$ or for $\lambda \in \mathbb{C}_-$, see [2].

If we analyze the analyticity properties of the Jost solutions $\psi_{\pm}(x, t, \lambda)$ related to algebras of higher rank one finds that only the first and the last columns of $\psi_{\pm}(x, t, \lambda)$ allow analytic extensions off the real λ -axis. An important result of Zakharov and Manakov [49, 48] consisted in showing that a FAS for the GZS with $\mathfrak{g} \simeq sl(n)$ and real-valued J can be constructed by taking proper linear combinations of the columns of $\psi_{\pm}(x, t, \lambda)$.

The construction is more complicated for the Caudrey-Beals-Coifman (CBC) systems when the eigenvalues of J are complex [5, 6, 8]. The generalization of this construction for CBC systems related to any simple Lie algebra \mathfrak{g} was done in [30].

We make attempt to outline the construction and the properties of each of these tools. Then we show how the FAS can be used to construct the kernel of the resolvent of $L(\lambda)$ and to exhibit its spectral properties and the structure of its discrete spectrum. Finally we illustrate how these tools can be used in the analysis of the NLEE and their fundamental properties and finish with some conclusions.

Both these aspects are rather broad; they have been widely discussed in hundreds of papers. Therefore inevitably the list of references consists mainly of reviews and monographs and bears an illustrative character. The thorough reader is advised to consult also the papers referred to in these references.

2. CONSTRUCTION OF THE FAS

Preliminaries: Jost solutions and scattering matrix

The direct and the inverse scattering problem for the Lax operator (1.5) will be done for fixed t and in most of the corresponding formulae t will be omitted.

The crucial fact that determines the spectral properties of the operator L is the choice of the class of functions where from we shall choose the potential $q(x)$. Below for simplicity we assume that the potential $q(x)$ is defined on the whole axis and satisfies the following conditions:

C.1: By $q(x) \in \mathfrak{M}_S$ we mean that $q(x)$ possesses smooth derivatives of all orders and falls off to zero for $|x| \rightarrow \infty$ faster than any power of x :

$$\lim_{x \rightarrow \pm\infty} |x|^k q(x) = 0, \quad \forall k = 0, 1, 2, \dots$$

C.2: $q(x)$ is such that the corresponding operator L has only a finite number of simple discrete eigenvalues.

Below we will use along with $L\psi(x, \lambda) = 0$ also the following equivalent formulations of the system (1.5):

$$(2.1) \quad i \frac{d\xi}{dx} + q(x, t)\xi(x, \lambda) - \lambda[J, \xi(x, \lambda)] = 0, \quad \xi(x, \lambda) = \psi(x, \lambda)e^{i\lambda Jx},$$

$$(2.2) \quad i \frac{d\hat{\psi}}{dx} - \hat{\psi}(x, \lambda)q(x, t) + \lambda\hat{\psi}(x, \lambda)J = 0, \quad \hat{\psi}(x, \lambda) = (\psi(x, \lambda))^{-1},$$

$$(2.3) \quad i \frac{d\hat{\xi}}{dx} - \hat{\xi}(x, \lambda)q(x, t) + \lambda[\hat{\xi}(x, \lambda), J] = 0, \quad \hat{\xi}(x, \lambda) = e^{-i\lambda Jx}\hat{\psi}(x, \lambda).$$

where by ‘hat’ we denote the inverse matrix, $\hat{X} \equiv X^{-1}$. The Jost solutions $\xi_{\pm}(x, \lambda)$, $\hat{\chi}_{\pm}(x, \lambda)$ and $\hat{\xi}_{\pm}(x, \lambda)$ for the systems (2.1)–(2.3) can be introduced by:

$$\lim_{x \rightarrow \pm\infty} \xi_{\pm}(x, \lambda) = \mathbf{1}, \quad \lim_{x \rightarrow \pm\infty} \psi_{\pm}(x, \lambda)e^{-i\lambda Jx} = \mathbf{1}, \quad \lim_{x \rightarrow \pm\infty} \hat{\xi}_{\pm}(x, \lambda) = \mathbf{1},$$

in analogy to (1.13); then their scattering matrices are:

$$T_2(\lambda) \equiv e^{i\lambda Jx}(\xi_{-}(x, \lambda))^{-1}\xi_{+}(x, \lambda)e^{-i\lambda Jx} = T(\lambda),$$

$$T_3(\lambda) \equiv \hat{\psi}_{+}(x, \lambda)(\hat{\psi}_{-}(x, \lambda))^{-1} = T^{-1}(\lambda),$$

$$T_4(\lambda) \equiv e^{i\lambda Jx}\hat{\xi}_{+}(x, \lambda)(\hat{\xi}_{-}(x, \lambda))^{-1}e^{-i\lambda Jx} = T^{-1}(\lambda),$$

Below we will consider two specific reductions of the Lax operator: the GZSs with \mathbb{Z}_2 -reduction:

$$(2.4) \quad B(U^\dagger(x, t, \epsilon\lambda^*))B^{-1} = U(x, t, \lambda), \quad B^2 = \mathbf{1}, \quad \epsilon = \pm 1.$$

The first possible choice for $B = \text{diag}(\epsilon_1, \dots, \epsilon_n)$, $\epsilon_j = \pm 1$ with $\epsilon = 1$ leads to the classical N -wave equations [49, 48] with

$$(2.5) \quad q_{kj}^*(x, t) = \epsilon_k \epsilon_j q_{jk}(x, t), \quad J = \text{diag}(a_1, \dots, a_n), \quad a_k = \epsilon a_k^*.$$

Since all eigenvalues of J are real ($\epsilon = 1$), or purely imaginary ($\epsilon = -1$), the Lax operator becomes a GZSs. The second choice for B :

$$(2.6) \quad B = \sum_{k=1}^n E_{k\bar{k}}, \quad \bar{k} = n + 1 - k, \quad \epsilon = -1,$$

will be used in combination with the \mathbb{Z}_n -reduction:

$$(2.7) \quad \tilde{C}_0(U(x, t, \omega\lambda))\tilde{C}_0^{-1} = U(x, t, \lambda), \quad \tilde{C}_0^n = \mathbf{1}.$$

which leads to the CBC system. For the sake of convenience in doing the spectral problem of CBCs we choose $C_0 = \sum_{k=1}^n E_{k,k+1}$; then $L(\lambda)$ has the form (1.5) with diagonal complex-valued J given by (1.8) or (1.11) where $c_0 = 1$ (resp. $c_0 = i$) if $\epsilon = 1$ (resp. $\epsilon = -1$). Both Lax operators will have similar spectral properties.

In solving the NLEE (1.2) and (1.3) we will need to apply both reductions (2.4) and (2.7) simultaneously. An attempt for classification of the \mathbb{Z}_2 -reductions is made in [23].

The FAS of the GZSs with \mathbb{Z}_2 -reduction.

Let us outline without proofs the construction of the FAS for the GZSs with real J , see [49, 48, 5, 8, 30]. For definiteness we assume that the real eigenvalues of J are pair-wise different and ordered as follows:

$$(2.8) \quad J = \text{diag}(a_1, \dots, a_n), \quad a_1 > a_2 > \dots > a_n.$$

PROPOSITION 2.1. *Let the potential of (1.5) $q(x) \in \mathfrak{M}_S$ satisfies conditions (C.1), (C.2) and (2.5). Then:*

a) *the Jost solutions $\xi_\pm(x, \lambda)$ and $\hat{\xi}_\pm(x, \lambda)$ of (2.1), (2.2) exist and are well defined functions for $\lambda \in \mathbb{R}$;*

b) *the matrix elements of the scattering matrix $T(\lambda)$ and its inverse $\hat{T}(\lambda)$ are Schwartz-type functions of λ for $\lambda \in \mathbb{R}$.*

REMARK 2.2. The proposition 2.1 concerns the Jost solutions as fundamental solutions. One can prove that the first and the last columns $\xi_\pm^{[1]}(x, \lambda)$ and $\xi_\pm^{[n]}(x, \lambda)$ of the Jost solutions allow analytic extension with respect to λ as follows:

$$\begin{array}{l} \text{Column} \quad \xi_+^{[1]}(x, \lambda) \quad \xi_+^{[n]}(x, \lambda) \quad \xi_-^{[1]}(x, \lambda) \quad \xi_-^{[n]}(x, \lambda) \\ \text{Analytic for} \quad \lambda \in \mathbb{C}_- \quad \lambda \in \mathbb{C}_+ \quad \lambda \in \mathbb{C}_+ \quad \lambda \in \mathbb{C}_- \end{array},$$

Analogously the first and the last rows of the Jost solutions $\hat{\xi}_\pm^{[1]}(x, \lambda)$ and $\hat{\xi}_\pm^{[n]}(x, \lambda)$ allow analytic extension with respect to λ as follows:

$$\begin{array}{l} \text{Row} \quad \hat{\xi}_+^{[1]}(x, \lambda) \quad \hat{\xi}_+^{[n]}(x, \lambda) \quad \hat{\xi}_-^{[1]}(x, \lambda) \quad \hat{\xi}_-^{[n]}(x, \lambda) \\ \text{Analytic for} \quad \lambda \in \mathbb{C}_+ \quad \lambda \in \mathbb{C}_- \quad \lambda \in \mathbb{C}_- \quad \lambda \in \mathbb{C}_+ \end{array},$$

All the other columns of $\xi_\pm(x, \lambda)$ and rows of $\hat{\xi}_\pm(x, \lambda)$ are defined only for $\lambda \in \mathbb{R}$ and do not allow analytic extensions off the real axis.

We start by explaining the construction of the FAS $\chi^\pm(x, \lambda)$ or rather of the solutions

$$(2.9) \quad \xi^\pm(x, \lambda) = \chi^\pm(x, \lambda)e^{i\lambda Jx}.$$

to equation (2.1) which allow analytic extensions for $\lambda \in \mathbb{C}_\pm$. Skipping the details (see [45, 48, 49]) we formulate the answer and determine $\xi^+(x, \lambda)$ as the solution of the following set of integral equations:

$$(2.10a) \quad \xi_{ij}^+(x, \lambda) = \delta_{ij} + i \int_{-\infty}^x dy e^{-i\lambda(a_i - a_j)(x-y)} \sum_{p=1}^h q_{ip}(y) \xi_{pj}^+(y, \lambda), \quad i \geq j;$$

$$(2.10b) \quad \xi_{ij}^+(x, \lambda) = i \int_{\infty}^x dy e^{-i\lambda(a_i - a_j)(x-y)} \sum_{p=1}^h q_{ip}(y) \xi_{pj}^+(y, \lambda), \quad i < j;$$

Analogously we define $\xi^-(x, \lambda)$ as the solution of the set of integral equations:

$$(2.11a) \quad \xi_{ij}^-(x, \lambda) = i \int_{\infty}^x dy e^{-i\lambda(a_i - a_j)(x-y)} \sum_{p=1}^h q_{ip}(y) \xi_{pj}^-(y, \lambda), \quad i > j;$$

$$(2.11b) \quad \xi_{ij}^-(x, \lambda) = \delta_{ij} + i \int_{-\infty}^x dy e^{-i\lambda(a_i - a_j)(x-y)} \sum_{p=1}^h q_{ip}(y) \xi_{pj}^-(y, \lambda), \quad i \leq j;$$

THEOREM 2.3. *Let $q(x) \in \mathfrak{M}_S$ satisfies conditions (C.1), (C.2) and let J satisfy (2.8). Then the solution $\xi^+(x, \lambda)$ of the eqs. (2.10) (resp. $\xi^-(x, \lambda)$ of the eqs. (2.11)) exists and allows analytic extension for $\lambda \in \mathbb{C}_+$ (resp. for $\lambda \in \mathbb{C}_-$).*

REMARK 2.4. Due to the fact that in eq. (2.10) we have both ∞ and $-\infty$ as lower limits the equations are rather of Fredholm than of Volterra type. Therefore we have to consider also the Fredholm alternative, i.e. there may exist finite number of values of $\lambda = \lambda_k^\pm \in \mathbb{C}_\pm$ for which the solutions $\xi^\pm(x, \lambda)$ have zeroes and pole singularities in λ . The points λ_k^\pm in fact are the discrete eigenvalues of $L(\lambda)$ in \mathbb{C}_\pm .

The reduction condition (2.4) with $\epsilon = 1$ means that the FAS and the scattering matrix $T(\lambda)$ satisfy:

$$(2.12) \quad B(\chi^+(x, \lambda^*))^\dagger B^{-1} = (\chi^-(x, \lambda))^{-1}, \quad B(T(\lambda^*))^\dagger B^{-1} = (T(\lambda))^{-1}.$$

Each fundamental solution of the Lax operator is uniquely determined by its asymptotic for $x \rightarrow \infty$ or $x \rightarrow -\infty$. Therefore in order to determine the linear relations between the FAS and the Jost solutions for $\lambda \in \mathbb{R}$ we need to calculate the asymptotics of FAS for $x \rightarrow \pm\infty$. Taking the limits in the right hand sides of the integral equations (2.10) and (2.11) we get:

$$(2.13a) \quad \lim_{x \rightarrow -\infty} \xi_{ij}^+(x, \lambda) = \delta_{ij}, \quad \text{for } i \geq j; \quad \lim_{x \rightarrow -\infty} \xi_{ij}^+(x, \lambda) = 0, \quad \text{for } i < j;$$

$$(2.13b) \quad \lim_{x \rightarrow -\infty} \xi_{ij}^-(x, \lambda) = \delta_{ij}, \quad \text{for } i \leq j; \quad \lim_{x \rightarrow -\infty} \xi_{ij}^-(x, \lambda) = 0, \quad \text{for } i > j;$$

This can be written in compact form using (2.9):

$$(2.14) \quad \chi^\pm(x, \lambda) = \psi_-(x, \lambda)S^\pm(\lambda) = \psi_+(x, \lambda)T^\mp(\lambda)D^\pm(\lambda),$$

where the matrices $S^\pm(\lambda)$, $D^\pm(\lambda)$ and $T^\pm(\lambda)$ are of the form:

$$(2.15a) \quad S^+(\lambda) = \begin{pmatrix} 1 & S_{12}^+ & \dots & S_{1n}^+ \\ 0 & 1 & \dots & S_{2n}^+ \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \quad T^+(\lambda) = \begin{pmatrix} 1 & T_{12}^+ & \dots & T_{1n}^+ \\ 0 & 1 & \dots & T_{2n}^+ \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

$$(2.15b) \quad D^+(\lambda) = \text{diag}(D_1^+, D_2^+, \dots, D_n^+), \quad D^-(\lambda) = \text{diag}(D_1^-, D_2^-, \dots, D_n^-),$$

$$(2.15c) \quad S^-(\lambda) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ S_{21}^- & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1}^- & S_{n2}^- & \dots & 1 \end{pmatrix}, \quad T^-(\lambda) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ T_{21}^- & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ T_{n1}^- & T_{n2}^- & \dots & 1 \end{pmatrix},$$

Let us now relate the factors $T^\pm(\lambda)$, $S^\pm(\lambda)$ and $D^\pm(\lambda)$ to the scattering matrix $T(\lambda)$. Comparing (2.14) with (1.14) we find

$$(2.16) \quad T(\lambda) = T^-(\lambda)D^+(\lambda)\hat{S}^+(\lambda) = T^+(\lambda)D^-(\lambda)\hat{S}^-(\lambda),$$

i.e. $T^\pm(\lambda)$, $S^\pm(\lambda)$ and $D^\pm(\lambda)$ are the factors in the Gauss decomposition of $T(\lambda)$.

It is well known how given $T(\lambda)$ one can construct explicitly its Gauss decomposition, see the Appendix A. Here we need only the expressions for $D^\pm(\lambda)$:

$$(2.17) \quad D_j^+(\lambda) = \frac{m_j^+(\lambda)}{m_{j-1}^+(\lambda)}, \quad D_j^-(\lambda) = \frac{m_{n-j+1}^-(\lambda)}{m_{n-j}^-(\lambda)},$$

where m_j^\pm are the principal upper and lower minors of $T(\lambda)$ of order j .

COROLLARY 2.5. *The upper (resp. lower) principal minors $m_j^\pm(\lambda)$ (resp. $m_j^-(\lambda)$) of $T(\lambda)$ are analytic functions of λ for $\lambda \in \mathbb{C}_+$ (resp. for $\lambda \in \mathbb{C}_-$).*

PROOF. Follows directly from theorem 2.3, from the limits:

$$(2.18) \quad \lim_{x \rightarrow \infty} \xi_{jj}^+(x, \lambda) = D_j^+(\lambda), \quad \lim_{x \rightarrow \infty} \xi_{jj}^-(x, \lambda) = D_j^-(\lambda),$$

and from (2.15b) and (2.17). \square

COROLLARY 2.6. *The following relations hold:*

$$a) \lim_{\lambda \rightarrow \infty} \xi^\pm(x, \lambda) = \mathbf{1}; \quad b) \lim_{\lambda \rightarrow \infty} m_j^\pm(\lambda) = 1.$$

PROOF. a) follows from the integral equations (2.10), (2.11) taking into account that the integrands in their right hand sides vanish for $\lambda \rightarrow \infty$. b) follows from a), (2.18) and (2.15b). \square

In what follows we will also assume that the set of minors $m_k^\pm(\lambda)$ have only finite number of simple zeroes located at the points

$$(2.19) \quad \mathfrak{Z} \equiv \{\lambda_j^\pm \in \mathbb{C}_\pm, \quad j = 1, \dots, N.\}$$

Generically each of the λ_j^\pm can be a zero of a string of minors, e.g.:

$$(2.20) \quad m_k^+(\lambda) = (\lambda - \lambda_j^+) \dot{m}_{k,j}^+ + \mathfrak{D}((\lambda - \lambda_j^+)^2),$$

for $1 \leq I_j < F_j \leq n$. Let us introduce the quantities b_{jk} as follows:

$$(2.21) \quad b_{jk} = \begin{cases} 1 & \text{if } \lambda_j^+ \text{ is a zero of } m_k^+(\lambda); \\ 0 & \text{if } \lambda_j^+ \text{ is not a zero of } m_k^+(\lambda). \end{cases}$$

and note that the reduction (2.4) means that the Gauss factors of $T(\lambda)$ satisfy ($\epsilon = 1$):

$$(2.22a) \quad B \left(\hat{S}^+(\lambda^*) \right)^\dagger B^{-1} = S^-(\lambda), \quad B \left(\hat{T}^+(\lambda^*) \right)^\dagger B^{-1} = T^-(\lambda),$$

$$(2.22b) \quad \left(\hat{D}^+(\lambda^*) \right)^\dagger = D^-(\lambda).$$

The relations (2.22a) are strictly valid only for $\lambda \in \mathbb{R}$ while (2.22b) together with (2.15b) and (2.17) leads to the following constraints on the minors $m_k^\pm(\lambda)$:

$$(2.23) \quad \left(m_k^+(\lambda^*) \right)^* = m_{n-k}^-(\lambda).$$

Thus if λ_k^+ is a zero of $m_k^+(\lambda)$ then $\lambda_k^- = (\lambda_k^+)^*$ is a zero of $m_{n-k}^-(\lambda)$.

The FAS of the CBCs with \mathbb{Z}_n -reduction.

The crucial difference with the \mathbb{Z}_2 -case treated above consists in the fact that now J is given by (1.8) or (1.11) and has complex eigenvalues. Skipping the details (see [5, 6, 8, 30]) we just outline the procedure of constructing the FAS.

First we have to determine the regions of analyticity. For potentials $q(x)$ satisfying the conditions (C.1) and (C.2) and subject to the \mathbb{Z}_n -reduction (2.7) these regions are the $2n$ sectors Ω_ν separated by the rays l_ν on which $\text{Im } \lambda(a_j - a_k) = 0$. We remind that if we assume also the \mathbb{Z}_2 -reduction (2.6) with $c_0^* = \epsilon c_0$ then $a_k = c_0 \omega^{k-1/2}$. Then the rays l_ν are given by:

$$(2.24) \quad l_\nu: \arg(\lambda) = \phi_0 + \frac{\pi(\nu - 1)}{n}, \quad \nu = 1, \dots, 2n,$$

where $\phi_0 = \pi/(2n)$ only if $\epsilon = 1$ and n is odd; in all other cases $\phi_0 = 0$. Thus the neighboring rays l_ν and $l_{\nu+1}$ close angles equal to π/n .

The next step is to construct the set of integral equations analogous to (2.10) whose solution will be analytic in Ω_ν . To this end we associate with each sector Ω_ν the relations (orderings) $\underset{\nu}{>}$ and $\underset{\nu}{<}$ by:

$$(2.25) \quad \begin{matrix} i \underset{\nu}{>} j \\ i \underset{\nu}{<} j \end{matrix} \quad \text{if} \quad \begin{matrix} \text{Im } \lambda(a_i - a_j) < 0 & \text{for } \lambda \in \Omega_\nu, \\ \text{Im } \lambda(a_i - a_j) > 0 & \text{for } \lambda \in \Omega_\nu. \end{matrix}$$

Then the solution of the system (2.10)

$$(2.26a) \quad \xi_{ij}^\nu(x, \lambda) = \delta_{ij} + i \int_{-\infty}^x dy e^{-i\lambda(a_i - a_j)(x-y)} \sum_{p=1}^h q_{ip}(y) \xi_{pj}^\nu(y, \lambda), \quad i \underset{\nu}{\geq} j;$$

$$(2.26b) \quad \xi_{ij}^\nu(x, \lambda) = i \int_{\infty}^x dy e^{-i\lambda(a_i - a_j)(x-y)} \sum_{p=1}^h q_{ip}(y) \xi_{pj}^\nu(y, \lambda), \quad i \underset{\nu}{\leq} j;$$

will be the FAS of the CBCs in the sector Ω_ν . The asymptotics of $\xi^\nu(x, \lambda)$ and $\xi^{\nu-1}(x, \lambda)$ along the ray l_ν can be written in the form:

$$(2.27a) \quad \lim_{x \rightarrow -\infty} e^{i\lambda Jx} \xi^\nu(x, \lambda e^{i0}) e^{-i\lambda Jx} = S_\nu^+(\lambda), \quad \lambda \in l_\nu,$$

$$(2.27b) \quad \lim_{x \rightarrow -\infty} e^{i\lambda Jx} \xi^{\nu-1}(x, \lambda e^{-i0}) e^{-i\lambda Jx} = S_\nu^-(\lambda), \quad \lambda \in l_\nu,$$

$$(2.27c) \quad \lim_{x \rightarrow \infty} e^{i\lambda Jx} \xi^\nu(x, \lambda e^{i0}) e^{-i\lambda Jx} = T_\nu^- D_\nu^+(\lambda), \quad \lambda \in l_\nu,$$

$$(2.27d) \quad \lim_{x \rightarrow \infty} e^{i\lambda Jx} \xi^{\nu-1}(x, \lambda e^{-i0}) e^{-i\lambda Jx} = T_\nu^+ D_\nu^-(\lambda), \quad \lambda \in l_\nu,$$

where the matrices S_ν^+ , T_ν^+ (resp. S_ν^- , T_ν^-) are upper-triangular (resp. lower-triangular) with respect to the ν -ordering. As in the previous case they provide the Gauss decomposition of the scattering matrix with respect to the ν -ordering, i.e.:

$$(2.28) \quad T_\nu(\lambda) = T_\nu^-(\lambda) D_\nu^+(\lambda) \hat{S}_\nu^+(\lambda) = T_\nu^+(\lambda) D_\nu^-(\lambda) \hat{S}_\nu^-(\lambda), \quad \lambda \in l_\nu.$$

More careful analysis shows [30] that in fact $T_\nu(\lambda)$ belongs to a subgroup \mathfrak{G}_ν of $SL(n, \mathbb{C})$. Indeed, with each ray l_ν one can relate a subalgebra $\mathfrak{g}_\nu \subset sl(n, \mathbb{C})$.

If \mathbb{Z}_n -symmetry is present each of these subalgebras \mathfrak{g}_ν is a direct sum of $sl(2)$ -subalgebras. Each such $sl(2)$ -subalgebra can be specified by a pair of indices (k, s) and is generated by:

$$(2.29) \quad h^{(k,s)} = E_{kk} - E_{ss}, \quad e^{(k,s)} = E_{ks}, \quad f^{(k,s)} = E_{sk}, \quad k < \underset{\nu}{s}.$$

Then the scattering matrix $T_\nu(\lambda)$ will be a product of mutually commuting matrices $T_\nu^{(k,s)}$ of the form:

$$(2.30) \quad T_\nu^{(k,s)} = \mathbb{1} + (a_{\nu;ks}^+(\lambda) - 1)E_{kk} + (a_{\nu;ks}^-(\lambda) - 1)E_{ss} - b_{\nu;ks}^-(\lambda)E_{ks} + b_{\nu;ks}^+(\lambda)E_{sk},$$

where $k < \underset{\nu}{s}$, with only 4 non-trivial matrix elements, just like the ZS (or AKNS) system.

The \mathbb{Z}_n -symmetry imposes the following constraints on the FAS and on the scattering matrix and its factors:

$$(2.31a) \quad C_0 \xi^\nu(x, \lambda \omega) C_0^{-1} = \xi^{\nu-2}(x, \lambda), \quad C_0 T_\nu(\lambda \omega) C_0^{-1} = T_{\nu-2}(\lambda),$$

$$(2.31b) \quad C_0 S_\nu^\pm(\lambda \omega) C_0^{-1} = S_{\nu-2}^\pm(\lambda), \quad C_0 D_\nu^\pm(\lambda \omega) C_0^{-1} = D_{\nu-2}^\pm(\lambda),$$

where the index $\nu - 2$ should be taken modulo $2n$. Consequently we can view as independent only the data on two of the rays, e.g. on l_1 and $l_{2n} \equiv l_0$; all the rest will be recovered from (2.31).

If in addition we impose the \mathbb{Z}_2 -symmetry (2.4), (2.6) with $\epsilon = -1$ then we will have also $a_k = i\omega^{k-1/2}$ and:

$$(2.32) \quad B(\xi^\nu(x, -\lambda^*))^\dagger B^{-1} = (\xi^{\nu+1-\nu}(x, \lambda))^{-1}, \quad B(S_\nu^\pm(\lambda^*)) B^{-1} = (S_{n+1-\nu}^\mp(\lambda))^{-1},$$

and analogous relations for $T_\nu^\pm(\lambda)$ and $D_\nu^\pm(\lambda)$. Another interesting subcase takes place for even values of n and \mathbb{Z}_2 -reduction (2.4), (2.6) with $\epsilon = 1$; then $a_k = \omega^{k-1/2}$ and the FAS satisfy:

$$(2.33) \quad B(\xi^\nu(x, \lambda^*))^\dagger B^{-1} = (\xi^{2n+1-\nu}(x, \lambda))^{-1}, \quad B(S_\nu^\pm(\lambda^*)) B^{-1} = (S_{2n+1-\nu}^\mp(\lambda))^{-1},$$

In both cases the rays l_ν are defined by (2.24) with $\phi_0 = 0$. The pairs of indices $\{k_\nu, m_\nu\}$ specifying the imbeddings of the $sl(2)$ -subalgebras related to the ray l_ν are defined as follows:

$$(2.34) \quad \begin{aligned} \text{a) for } \epsilon = 1 & \quad k_\nu + m_\nu = \left[\frac{n}{2} \right] + 2 - \nu \pmod{n}, \\ \text{b) for } \epsilon = -1 & \quad k_\nu + m_\nu = 2 - \nu \pmod{n}, \end{aligned}$$

One can prove also that $D_\nu^+(\lambda)$ (resp. $D_\nu^-(\lambda)$) allows analytic extension for $\lambda \in \Omega_\nu$ (resp. for $\lambda \in \Omega_{\nu-1}$, compare with corollary 2.5. Another important fact is [30] that $D_\nu^+(\lambda) = D_{\nu+1}^-(\lambda)$ for all $\lambda \in \Omega_\nu$.

The inverse scattering problem and the Riemann-Hilbert problem.

The next important step is the possibility to reduce the solution of the ISP for the GZSS to a (local) RHP. Indeed the relation (2.14) can be rewritten as:

$$(2.35a) \quad \xi^+(x, t, \lambda) = \xi^-(x, t, \lambda)G(x, t, \lambda), \quad \lambda \in \mathbb{R},$$

$$(2.35b) \quad G(x, t, \lambda) = e^{-i(\lambda Jx - f(\lambda)t)}G_0(\lambda)e^{i(\lambda Jx - f(\lambda)t)},$$

$$(2.35c) \quad G_0(\lambda) = \hat{S}^-(\lambda)S^+(\lambda)\Big|_{t=0};$$

in other words the sewing function $G(x, t, \lambda)$ satisfies the equations:

$$(2.36) \quad i\frac{dG}{dx} - \lambda[J, G(x, t, \lambda)] = 0, \quad i\frac{dG}{dt} + [f(\lambda), G(x, t, \lambda)] = 0,$$

Here $f(\lambda) \in \mathfrak{h}$ determines the dispersion law of the NLEE. Together with

$$(2.37) \quad \lim_{\lambda \rightarrow \infty} \xi^\pm(x, \lambda) = \mathbb{1},$$

eq. (2.35) is known as the RHP with canonical normalization.

THEOREM 2.7 ([45]). *Let $\xi^+(x, t, \lambda)$ and $\xi^-(x, t, \lambda)$ be solutions to the RHP (2.35), (2.37) allowing analytic extension in λ for $\lambda \in \mathbb{C}_\pm$ respectively. Then $\chi^\pm(x, t, \lambda) = \xi^\pm(x, t, \lambda)e^{i\lambda Jx}$ are fundamental analytic solutions of both operators L and M , i.e. satisfy eqs. (1.5), (1.6) with*

$$(2.38) \quad q(x, t) = \lim_{\lambda \rightarrow \infty} \lambda \left(J - \xi^\pm(x, t, \lambda)J\hat{\xi}^\pm(x, t, \lambda) \right).$$

PROOF. Let us assume that $\xi^\pm(x, t, \lambda)$ are regular solutions to the RHP and let us introduce the function:

$$(2.39) \quad g^\pm(x, t, \lambda) = i\frac{d\xi^\pm}{dx}\hat{\xi}^\pm(x, t, \lambda) + \lambda\xi^\pm(x, t, \lambda)J\hat{\xi}^\pm(x, t, \lambda).$$

If $\xi^\pm(x, t, \lambda)$ are regular then neither $\xi^\pm(x, t, \lambda)$ nor their inverse $\hat{\xi}^\pm(x, t, \lambda)$ have singularities in their regions of analyticity $\lambda \in \mathbb{C}_\pm$. Then the functions $g^\pm(x, t, \lambda)$ also will be regular for all $\lambda \in \mathbb{C}_\pm$. Besides:

$$(2.40) \quad \lim_{\lambda \rightarrow \infty} g^+(x, t, \lambda) = \lim_{\lambda \rightarrow \infty} g^-(x, t, \lambda) = \lambda J.$$

The crucial step in the proof of [52] is based on the chain of relations:

$$\begin{aligned} g^+(x, t, \lambda) &\stackrel{(2.35)}{=} i\frac{d(\xi^-G)}{dx}\hat{G}\hat{\xi}^-(x, t, \lambda) + \lambda\xi^-GJG\hat{\xi}^-(x, t, \lambda) \\ &= i\frac{d\xi^-}{dx}\hat{\xi}^-(x, t, \lambda) + \xi^- \left(i\frac{dG}{dx}\hat{G} + \lambda GJG\hat{G}(x, t, \lambda) \right) \hat{\xi}^-(x, t, \lambda) \\ &\stackrel{(2.36)}{=} i\frac{d\xi^-}{dx}\hat{\xi}^-(x, t, \lambda) + \xi^- \left(\lambda[J, G]\hat{G} + \lambda GJG\hat{G}(x, t, \lambda) \right) \hat{\xi}^-(x, t, \lambda) \\ &= i\frac{d\xi^-}{dx}\hat{\xi}^-(x, t, \lambda) + \lambda\xi^-J\hat{\xi}^-(x, t, \lambda) \\ (2.41) \quad &\equiv g^-(x, t, \lambda), \quad \lambda \in \mathbb{R}. \end{aligned}$$

Thus we conclude that $g^+(x, t, \lambda) = g^-(x, t, \lambda)$ is a function analytic in the whole complex λ -plane except in the vicinity of $\lambda \rightarrow \infty$ where $g^+(x, t, \lambda)$ tends to λJ , (2.40). Next from Liouville theorem we conclude that the difference $g^+(x, t, \lambda) - \lambda J$ is a constant with respect to λ ; if we denote this ‘constant’ by $-q(x, t)$ we get:

$$(2.42) \quad g^+(x, t, \lambda) - \lambda J = -q(x, t).$$

It remains to remember the definition of $g^+(x, t, \lambda)$ (2.39) to find that $\xi^\pm(x, t, \lambda)$ satisfy (2.1), i.e. that $\chi^\pm(x, t, \lambda)$ is a fundamental solution to L . The relation between $q(x, t)$ and $\xi^\pm(x, t, \lambda)$ (2.38) can be obtained by taking the limit of the left-hand sides of (2.42) for $\lambda \rightarrow \infty$.

Arguments along the same line applied to the functions $h^\pm(x, t, \lambda)$

$$(2.43) \quad h^\pm(x, t, \lambda) = i \frac{d\xi^\pm}{dt} \hat{\xi}^\pm(x, t, \lambda) - \xi^\pm(x, t, \lambda) f(\lambda) \hat{\xi}^\pm(x, t, \lambda),$$

can be used to prove that $\chi^\pm(x, t, \lambda)$ are fundamental solutions also of the operator M ; equivalently it satisfies ($V'(x, t, \lambda) = V(x, t, \lambda) - f(\lambda)$):

$$(2.44) \quad i \frac{d\xi^\pm}{dt} + V'(x, t, \lambda) \xi^\pm(x, t, \lambda) + [f(\lambda), \xi^\pm(x, t, \lambda)] = 0,$$

and one finds that $h^+(x, t, \lambda) = h^-(x, t, \lambda)$ is a function analytic everywhere in \mathbb{C} except at $\lambda \rightarrow \infty$ where it has a polynomial behavior of order $N - 1$. Denoting the polynomial as $V(x, t, \lambda)$ we derive (2.43).

To conclude the proof of the theorem we have to account for possible zeroes and pole singularities of $\xi^\pm(x, t, \lambda)$ at the points \mathfrak{Z} (2.19). Below we derive the structure of these singularities which is such that they do not influence the functions $g^\pm(x, t, \lambda)$ and $h^\pm(x, t, \lambda)$. The theorem is proved. \square

The analyticity properties of $m_k^\pm(\lambda)$ allow one to reconstruct them from the sewing function $G(\lambda)$ (2.35c) and from the locations of their zeroes through (see Appendix B):

$$(2.45) \quad \mathfrak{D}_k(\lambda) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{d\mu}{\mu - \lambda} \ln \left\{ \begin{matrix} 1, 2, \dots, k \\ 1, 2, \dots, k \end{matrix} \right\}_{G(\mu)} + \sum_{j=1}^N b_{jk} \ln \frac{\lambda - \lambda_j^+}{\lambda - \lambda_j^-},$$

where

$$(2.46) \quad \mathfrak{D}_k(\lambda) = \begin{cases} \ln m_k^+(\lambda), & \lambda \in \mathbb{C}_+ \\ -\ln m_{n-k}^-(\lambda), & \lambda \in \mathbb{C}_-. \end{cases}$$

One can view $\mathfrak{D}_k(\lambda)$ as generating functionals of the conserved quantities for the related N -wave-type equations; the relevant expressions for them in terms of the scattering data can be obtained from the right hand sides of (2.45).

Quite analogously we can treat also the CBCs with \mathbb{Z}_n -symmetry. More precisely, we have:

$$(2.47) \quad \begin{aligned} \xi^\nu(x, t, \lambda) &= \xi^{\nu-1}(x, t, \lambda) G_\nu(x, t, \lambda), & \lambda \in l_\nu, \\ G_\nu(x, t, \lambda) &= e^{-i\lambda Jx + if(\lambda)t} G_{0,\nu}(\lambda) e^{i\lambda Jx - if(\lambda)t}, & G_{0,\nu}(\lambda) = \hat{S}_\nu^-(\lambda) S_\nu^+(\lambda) \Big|_{t=0} \end{aligned}$$

The collection of all relations (2.47) for $\nu = 1, 2, \dots, 2n$ together with

$$(2.48) \quad \lim_{\lambda \rightarrow \infty} \xi^\nu(x, t, \lambda) = \mathbf{1},$$

can be viewed as a local RHP posed on the collection of rays $\Sigma \equiv \{l_\nu\}_{\nu=1}^{2n}$ with canonical normalization. Rather straightforwardly we can reformulate the results for the GZSs for the CBCs and prove that if $\xi^\nu(x, \lambda)$ is a solution of the RHP (2.47), (2.48) then $\chi^\nu(x, \lambda) = \xi^\nu(x, \lambda) e^{i\lambda Jx}$ satisfy the CBC with potential

$$(2.49) \quad q(x, t) = \lim_{\lambda \rightarrow \infty} \left(J - \xi^\nu(x, t, \lambda) J \hat{\xi}^\nu(x, t, \lambda) \right).$$

We finish this subsection by formulating the dispersion relations for the functions $\ln m_{\nu,k}^+(\lambda)$ which allows us to reconstruct them from their analyticity properties: (2.50)

$$\ln m_{\nu,k}^+(\lambda) = \sum_{\eta=1}^{2n} \frac{(-1)^\eta}{2\pi i} \int_{l_\nu} \frac{d\mu}{\mu - \lambda} \ln \left\{ \begin{matrix} 1 & \dots & k \\ 1 & \dots & k \end{matrix} \right\}_{G_\eta(\mu)}^\eta + \sum_{\eta=1}^n \sum_{j=1}^N b_{kj}^\eta \ln \frac{\lambda - \lambda_{j,k}^+ \omega^\eta}{\lambda - \lambda_{j,k}^- \omega^\eta},$$

where $\lambda \in \Omega_\nu$ and the superscript η in the integrand shows that we should use the ordering characteristic for the sector Ω_η ; b_{kj}^η are the analogs for b_{kj} (2.21) in Ω_η .

Both dispersion relations (2.45) and (2.50) can be used to derive the so-called trace identities (see [48, 13]) for the GZSs and CBCs respectively. Indeed, $\mathfrak{D}_k(\lambda)$ and $\ln m_{\nu,k}^+(\lambda)$ allow asymptotic expansions

$$(2.51) \quad \mathfrak{D}_k(\lambda) = \sum_{s=1}^{\infty} \mathfrak{D}_k^{(s)} \lambda^{-s}, \quad \ln m_{\nu,k}^+(\lambda) = \sum_{s=1}^{\infty} M_{\nu,k}^{(s)} \lambda^{-s}.$$

The expansion coefficients $\mathfrak{D}_k^{(s)}$ and $M_{\nu,k}^{(s)}$ are local integrals of motion, i.e. their densities depend only on $q(x, t)$ and its derivatives with respect to x . Their explicit calculation is done via recurrent procedure. We illustrate it by the two first integrals of motion of the \mathbb{Z}_n -NLS equation (1.3):

$$(2.52) \quad M_{1,1}^{(1)} = \frac{1}{2\omega} \int_{-\infty}^{\infty} dx \sum_{p=1}^n \psi_p \psi_{n-p}(x, t),$$

$$(2.53) \quad M_{1,1}^{(2)} = \frac{1}{2\omega^2} \int_{-\infty}^{\infty} dx \left\{ \sum_{p=1}^n i \cotan \left(\frac{\pi p}{n} \right) \left(\frac{d\psi_p}{dx} \psi_{n-p} - \psi_p \frac{d\psi_{n-p}}{dx} \right) - \frac{2}{3} \sum_{p+k+l=n} \psi_p \psi_k \psi_l(x, t) \right\},$$

One can also expand the right hand sides of the dispersion relations (2.45) and (2.50) over the inverse powers of λ which allows to express $\mathfrak{D}_k^{(s)}$ and $M_{\nu,k}^{(s)}$ also in terms of the scattering data of GZSs and CBCs.

The dressing Zakharov-Shabat method

One of the most fruitful ideas for the explicit constructing of the NLEE's solutions is based on the possibility starting from a given regular solutions $\xi_0^\pm(x, t, \lambda)$ to the RHP to construct new singular solutions $\xi^\pm(x, t, \lambda)$ having zeroes and pole singularities at the prescribed points $\lambda_j^\pm \in \mathbb{C}_\pm$. The structure of these singularities are determined by the dressing factor $u_j(x, t, \lambda)$:

$$(2.54) \quad \xi^\pm(x, t, \lambda) = u_j(x, t, \lambda) \xi_0^\pm(x, t, \lambda) u_{j,-}^{-1}(\lambda),$$

which for the $SL(n)$ -group has a simple fraction-linear dependence on λ :

$$(2.55) \quad u_j(x, t, \lambda) = \mathbf{1} + (c_j(\lambda) - 1) P_j(x, t), \quad c_j(\lambda) = \frac{\lambda - \lambda_j^+}{\lambda - \lambda_j^-},$$

$$(2.56) \quad u_{j,-}^{-1} = \lim_{x \rightarrow -\infty} u_j(x, t, \lambda).$$

Here $P_j(x, t)$ is a projector $P_j^2 = P_j$ which for simplicity is chosen to be of rank 1; then it can be written down as:

$$(2.57) \quad P_j(x) = \frac{|n_j\rangle\langle m_j|}{\langle m_j|n_j\rangle},$$

where the bra- and ket- eigenvectors $\langle m_j|$ and $|n_j\rangle$ are the ‘left’ and ‘right’ eigenvectors of the projector.

From (2.54) there follows that the dressing factor $u(x, t, \lambda)$ satisfies the equation:

$$(2.58) \quad i \frac{du}{dx} + q(x, t)u(x, t, \lambda) - u(x, t, \lambda)q_0(x, t) - \lambda[J, u(x, t, \lambda)] = 0.$$

The main advantage of the dressing method is in the fact that one can determine the x and t -dependence of $\langle m_j|$ and $|n_j\rangle$ through the regular solution $\chi_0^\pm(x, t, \lambda)$ as follows:

$$(2.59) \quad |n_j\rangle = \chi_{0j}^+(x, t)|n_j^0\rangle, \quad \langle m_j| = \langle m_j^0|\hat{\chi}_{0j}^-(x, t), \quad \chi_{0j}^\pm(x, t) = \chi_0^\pm(x, t, \lambda_j^\pm)$$

or equivalently these vectors are solutions to the equations:

$$(2.60) \quad i \frac{d|n_j\rangle}{dx} + U^{(0)}(x, t, \lambda_j^+)|n_j\rangle = 0, \quad i \frac{d|n_j\rangle}{dt} + V^{(0)}(x, t, \lambda_j^+)|n_j\rangle = 0,$$

$$(2.61) \quad i \frac{d\langle m_j|}{dx} - \langle m_j|U^{(0)}(x, t, \lambda_j^-) = 0, \quad i \frac{d\langle m_j|}{dt} - \langle m_j|V^{(0)}(x, t, \lambda_j^-) = 0,$$

$$(2.62) \quad U^{(0)}(x, t, \lambda) = q_0(x, t) - \lambda J, \quad V^{(0)}(x, t, \lambda) = V(x, t, \lambda)|_{q=q_0}.$$

Here $q_0(x, t)$ is the potential corresponding to the regular solutions $\chi_0^\pm(x, t, \lambda)$ to the RHP and $V^{(0)}(x, t, \lambda)$ is obtained from $V(x, t, \lambda)$ (see (3.35), (3.36)) replacing $q(x, t)$ by $q_0(x, t)$. This construction is well defined also in the case when $\chi_0^\pm(x, \lambda)$ are singular solutions to the RHP, provided they are regular for $\lambda = \lambda_j^\pm$.

If $q(x, t)$ is the potential corresponding to the singular solution $\chi^\pm(x, t, \lambda)$ then:

$$(2.63) \quad \begin{aligned} q(x, t) &= q_0(x, t) + \lim_{\lambda \rightarrow \infty} \lambda(J - u_j(x, t, \lambda)J\hat{u}_j(x, t, \lambda)) \\ &= q_0(x, t) - (\lambda_j^+ - \lambda_j^-)[J, P_j(x, t)]. \end{aligned}$$

Thus starting from a given regular solution of the RHP (and related solution $q_0(x, t)$ to the NLEE) we can construct a singular solution to the RHP and a new solution $q(x, t)$ of the NLEE depending on the λ_j^\pm and on the eigenvectors of $P_j(x)$. If we start from the trivial solution $q_0(x, t) = 0$ of the NLEE then we will get the one-soliton solution of the NLEE. Repeating the procedure N times we can get the N -soliton solution of the NLEE.

With the explicit formulae for $P_j(x)$ and using (2.54) we can establish the relationship between the scattering data of the regular RHP and the corresponding singular one. The dressing factor $u_j(x, \lambda)$ is determined by the constant vectors $\langle m_j^0|$ and $|n_j^0\rangle$ can not be quite arbitrary. The condition that $q(x)$ vanishes for $x \rightarrow \pm\infty$ requires that if $(n_j^0)_s = 0$ for all $1 \leq s < I_j$ and $F_j < s \leq n$ then also $(m_j^0)_s = 0$ for all $1 \leq s < I_1$ and $F_1 < s \leq n$. Thus we derive that:

$$(2.64) \quad \lim_{x \rightarrow \infty} P_j(x) = E_{I_j I_j}, \quad \lim_{x \rightarrow -\infty} P_j(x) = E_{F_j F_j},$$

and therefore

$$(2.65) \quad u_{j,+}(\lambda) = \mathbf{1} + (c_j(\lambda) - 1)E_{I_j I_j}, \quad u_{j,-}(\lambda) = \mathbf{1} + (c_j(\lambda) - 1)E_{F_j F_j}.$$

The interrelation between the Gauss factors of the corresponding scattering matrices are:

$$(2.66) \quad S^\pm(\lambda) = u_{j,-}(\lambda)S_0^\pm(\lambda)u_{j,-}^{-1}(\lambda), \quad T^\pm(\lambda) = u_{j,+}(\lambda)T_0^\pm(\lambda)u_{j,+}^{-1}(\lambda),$$

and

$$(2.67) \quad D^\pm(\lambda) = u_{j,+}(\lambda)D_0^\pm(\lambda)u_{j,-}^{-1}(\lambda).$$

Comparing these last relations with (2.17) we find for the principal minors of $T(\lambda)$ and $T_0(\lambda)$:

$$(2.68a) \quad m_s^+(\lambda) = \frac{\lambda - \lambda_j^+}{\lambda - \lambda_j^-} m_{0,s}^+(\lambda), \quad \text{for } I_j \leq s < F_j, \quad \lambda \in \mathbb{C}_+ \cup \mathbb{R},$$

$$(2.68b) \quad m_s^-(\lambda) = \frac{\lambda - \lambda_j^-}{\lambda - \lambda_j^+} m_{0,s}^+(\lambda), \quad \text{for } n - F_j < s \leq n - I_j, \quad \lambda \in \mathbb{C}_- \cup \mathbb{R},$$

and $m_s^\pm(\lambda) = m_{0,s}^\pm(\lambda)$ for the other values of s . Thus the result of the dressing is that the string of upper principle minors $m_s^+(\lambda)$, $I_j \leq s < F_j$ acquire simple zero at $\lambda = \lambda_j^+$ while the string of lower principle minors $m_s^-(\lambda)$, $n - F_j < s \leq n - I_j$ acquire simple zero at $\lambda = \lambda_j^-$.

Obviously if we impose on $L(\lambda)$ the \mathbb{Z}_2 -reduction then we should restrict also the dressing factor by:

$$(2.69) \quad B(u(x, t, \epsilon\lambda^*))^\dagger B^{-1} = u(x, t, \lambda).$$

The ansatz (2.55) satisfies (2.69) if $\lambda_j^- = \epsilon(\lambda_j^+)^*$ and the vectors $|n_{0j}\rangle, \langle m_{0j}|$ are related by:

$$(2.70) \quad \langle m_{0j}| = B|n_{0j}^\dagger\rangle.$$

If we impose the \mathbb{Z}_n -reduction (2.7) then $u(x, t, \lambda)$ must satisfy:

$$(2.71) \quad C_0 u(x, t, \omega\lambda) C_0^{-1} = u(x, t, \lambda).$$

Such conditions require generalizations of the ansatz (2.55) [41]:

$$(2.72) \quad u_j(x, t, \lambda) = \mathbb{1} + \sum_{s=0}^{n-1} (c_j(\omega^s\lambda) - 1) C_0^s P_j(x) C_0^{-s}.$$

A slightly different approach treating also multi-soliton solutions of the \mathbb{Z}_n -symmetric NLEE is given in [5].

Up to now we dealt with the algebra $\mathfrak{g} \simeq sl(n, \mathbb{C})$. Treating the other simple Lie algebras (orthogonal or symplectic) needs additional care especially in constructing the dressing factors [51, 23].

In fact $u_j(x, \lambda)$ (2.55) must be an element of the corresponding group. From the ansatz (2.55) it follows that $u_j(x, \lambda)$ belongs to $GL(n, \mathbb{C})$, but one can always multiply $u(x, \lambda)$ by an appropriate x - and t -independent scalar and to adjust its determinant to 1. Such a multiplication goes through the whole scheme outlined above but is adequate only for the $sl(n, \mathbb{C})$ case. However the ansatz (2.55) can not be used, e.g. for the case $so(n, \mathbb{C})$. The adequate ansatz is formulated below [23].

THEOREM 2.8. *Let $\mathfrak{g} \sim \mathbf{B}_r$ or \mathbf{D}_r and let the dressing factor $u(x, \lambda)$ be of the form:*

$$(2.73) \quad u_j(x, \lambda) = \mathbb{1} + (c_j(\lambda) - 1)P_j(x) + (c_j^{-1}(\lambda) - 1)P_{-j}(x), \quad P_{-j} = S_0 P_j^T S_0^{-1},$$

where S_0 is introduced in (A.11) and $P_j(x)$ is a rank 1 projector (2.57). Let the constant vectors $|n_0\rangle$ and $\langle m_0|$ satisfy the condition

$$(2.74) \quad \langle m_0|S|m_0\rangle = \langle n_0|S|n_0\rangle = 0.$$

Then $u_j(x, \lambda)$ (2.73) satisfies the equation (2.58) with a potential

$$(2.75) \quad q(x) = q_0(x) - (\lambda_j^+ - \lambda_j^-)[J, p_j(x)], \quad p_j(x) = P_j(x) - P_{-j}(x).$$

PROOF. Due to the fact that $\chi_0^\pm(x, \lambda)$ take values in the corresponding orthogonal group we find that from (2.74) it follows $\langle m|S|m\rangle = 0$, $\langle m|JS|m\rangle = 0$ and analogous relations for the vector $|n\rangle$. As a result we get that

$$(2.76) \quad P_j(x)P_{-j}(x) = P_{-j}(x)P_j(x) = 0, \quad P_j(x)JP_{-j}(x) = P_{-j}(x)JP_j(x) = 0.$$

Let us now insert (2.73) into (2.58) and take the limit of the r.h.side of (2.58) for $\lambda \rightarrow \infty$. This immediately gives eq. (2.75). In order that Eq. (2.58) be satisfied identically with respect to λ we have to put to 0 also the residues of its r.h.side at $\lambda \rightarrow \lambda_j^+$ and $\lambda \rightarrow \lambda_j^-$. This gives us the following system of equation for the projectors $P_j(x)$ and $P_{-j}(x)$:

$$(2.77) \quad i \frac{dP_j}{dx} + q(x)P_j(x) - P_j(x)q_0(x) - \lambda_j^- [J, P_j(x)] = 0,$$

$$(2.78) \quad i \frac{dP_{-j}}{dx} + q(x)P_{-j}(x) - P_{-j}(x)q_0(x) - \lambda_j^+ [J, P_{-j}(x)] = 0,$$

where we have to keep in mind that q is given by (2.75). Taking into account (2.76) and the relation between $P_j(x)$ and $P_{-j}(x)$ eq. (2.77) reduces to:

$$(2.79) \quad i \frac{dP_j}{dx} + [q_0(x), P_j(x)] + \lambda_j^- P_j(x)J - \lambda_j^+ JP_j(x) - (\lambda_j^- - \lambda_j^+) P_j(x)JP_j(x) = 0.$$

One can check by a direct calculation that (2.57) satisfies identically (2.79). The theorem is proved. □

3. THE RESOLVENT AND SPECTRAL PROPERTIES OF GZSS AND CBCS

The FAS $\chi^\pm(x, \lambda)$ of $L(\lambda)$ allows one to construct the resolvent of the operator L and then to investigate its spectral properties. By resolvent of $L(\lambda)$ we understand the integral operator $R(\lambda)$ with kernel $R(x, y, \lambda)$ which satisfies

$$(3.1) \quad L(\lambda)(R(\lambda)f)(x) = f(x),$$

where $f(x)$ is an n -component vector function in \mathbb{C}^n with bounded norm, i.e. $\int_{-\infty}^{\infty} dy(f^T(y)f(y)) < \infty$.

From the general theory of linear operators [4, 12, 46] we know that the point λ in the complex λ -plane is a regular point if $R(\lambda)$ is a bounded integral operator. In each connected subset of regular points $R(\lambda)$ is analytic in λ .

The points λ which are not regular constitute the spectrum of $L(\lambda)$. Roughly speaking the spectrum of $L(\lambda)$ consist of two types of points:

- i) the continuous spectrum of $L(\lambda)$ consists of all points λ for which $R(\lambda)$ is an unbounded integral operator;
- ii) the discrete spectrum of $L(\lambda)$ consists of all points λ for which $R(\lambda)$ develops pole singularities.

Let us now show how the resolvent $R(\lambda)$ can be expressed through the FAS of $L(\lambda)$. Indeed, if we write down $R(\lambda)$ in the form:

$$(3.2) \quad R(\lambda)f(x) = \int_{-\infty}^{\infty} R(x, y, \lambda)f(y),$$

the kernel $R(x, y, \lambda)$ of the resolvent is given by:

$$(3.3) \quad R(x, y, \lambda) = \begin{cases} R^+(x, y, \lambda) & \text{for } \lambda \in \mathbb{C}^+, \\ R^-(x, y, \lambda) & \text{for } \lambda \in \mathbb{C}^-, \end{cases}$$

where

$$(3.4) \quad R^\pm(x, y, \lambda) = \pm i\chi^\pm(x, \lambda)\Theta^\pm(x - y)\hat{\chi}^\pm(y, \lambda),$$

$$\Theta^\pm(z) = \theta(\mp z)\Pi_0 - \theta(\pm z)(\mathbb{1} - \Pi_0), \quad \Pi_0 = \sum_{s=1}^{k_0} E_{ss},$$

where k_0 is the number of positive eigenvalues of J ; namely:

$$(3.5) \quad a_1 > a_2 > \dots > a_{k_0} > 0 > a_{k_0+1} > \dots > a_n.$$

Due to the condition $\text{tr } J = \sum_{s=1}^n a_s = 0$, k_0 is fixed up uniquely.

The next theorem establishes that $R(x, y, \lambda)$ is indeed the kernel of the resolvent of $L(\lambda)$.

THEOREM 3.1. *Let $q(x)$ satisfy conditions (C.1) and (C.2) and let λ_j^\pm be the simple zeroes of the minors $m_k^\pm(\lambda)$. Then*

- (1) $R^\pm(x, y, \lambda)$ is an analytic function of λ for $\lambda \in \mathbb{C}_\pm$ having pole singularities at $\lambda_j^\pm \in \mathbb{C}_\pm$;
- (2) $R^\pm(x, y, \lambda)$ is a kernel of a bounded integral operator for $\text{Im } \lambda \neq 0$;
- (3) $R(x, y, \lambda)$ is uniformly bounded function for $\lambda \in \mathbb{R}$ and provides a kernel of an unbounded integral operator;
- (4) $R^\pm(x, y, \lambda)$ satisfy the equation:

$$(3.6) \quad L(\lambda)R^\pm(x, y, \lambda) = \mathbb{1}\delta(x - y).$$

IDEA OF THE PROOF. (1) is obvious from the fact that $\chi^\pm(x, \lambda)$ are the FAS of $L(\lambda)$;

- (2) Assume that $\text{Im } \lambda > 0$ and consider the asymptotic behavior of $R^+(x, y, \lambda)$ for $x, y \rightarrow \infty$. From equations (2.9) we find that

$$(3.7) \quad R_{ij}^+(x, y, \lambda) = \sum_{p=1}^n \xi_{ip}^+(x, \lambda)e^{-i\lambda a_p(x-y)}\Theta_{pp}^+(x - y)\hat{\xi}_{pj}^+(y, \lambda)$$

Due to the fact that $\chi^+(x, \lambda)$ has triangular asymptotics for $x \rightarrow \infty$ and $\lambda \in \mathbb{C}_+$ and for the correct choice of $\Theta^+(x - y)$ (3.4) we check that the right hand side of (3.7) falls off exponentially for $x \rightarrow \infty$ and arbitrary choice of y . All other possibilities are treated analogously.

- (3) For $\lambda \in \mathbb{R}$ the arguments of 2) can not be applied because the exponentials in the right hand side of (3.7) $\text{Im } \lambda = 0$ only oscillate. Thus we conclude that $R^\pm(x, y, \lambda)$ for $\lambda \in \mathbb{R}$ is only a bounded function and thus the corresponding operator $R(\lambda)$ is an unbounded integral operator.

(4) The proof of eq. (3.6) follows from the fact that $L(\lambda)\chi^+(x, \lambda) = 0$ and

$$(3.8) \quad \frac{d\Theta^\pm(x-y)}{dx} = \mp \mathbf{1}\delta(x-y).$$

□

PROPOSITION 3.2. Let $q(x)$ satisfy the conditions (C.1) and (C.2), let $\mathfrak{Z} \equiv \{\lambda_j^\pm, j = 1, \dots, N\}$ be the set of simple zeroes (2.20) of the minors $m_s^\pm(\lambda)$ and let $I_j \leq k_0 < F_j$ for all $j = 1, \dots, N$. Then the kernel of the resolvent $R^+(x, y, \lambda)$ (resp. $R^-(x, y, \lambda)$) has simple poles for $\lambda = \lambda_j^+$ (resp. for $\lambda = \lambda_j^-$) with residues given by:

$$(3.9a) \quad \text{Res}_{\lambda=\lambda_j^\pm} R^\pm(x, y, \lambda) = \mp 2i\nu_j |\mathbf{n}_j^\pm(x)\rangle \langle \mathbf{m}_j^\pm(y)|,$$

$$(3.9b) \quad |\mathbf{n}_j^+(x)\rangle = (\mathbf{1} - P_j(x))\chi_{0,j}^+(x)\Pi_0|n_{0,j}\rangle, \quad \langle \mathbf{m}_j^+(y)| = \frac{\langle m_j(y)|}{\langle m_j(y)|n_j(y)\rangle},$$

$$(3.9c) \quad |\mathbf{n}_j^-(x)\rangle = \frac{|n_j(x)\rangle}{\langle m_j(x)|n_j(x)\rangle}, \quad \langle \mathbf{m}_j^-(y)| = \langle m_{0,j}|\Pi_0\hat{\chi}_{0,j}^-(y)(\mathbf{1} - P_1(y)),$$

where $\lambda_j^\pm = \mu_j \pm i\nu_j$ and $\chi_{0,j}^\pm(x) = \chi_0^\pm(x, \lambda_j^\pm)$ are the FAS corresponding to the potential q_0 satisfying (C.1) and (C.2) and whose set of simple zeroes is $\mathfrak{Z}_0 \equiv \mathfrak{Z} \setminus \{\lambda_j^+, \lambda_j^-\}$.

PROOF. Let $\chi_0^\pm(x, \lambda)$ be the FAS of $L_0(\lambda)$ with potential $q_0(x)$; then $\chi_0^\pm(x, \lambda)$ are regular for $\lambda = \lambda_j^\pm$. Now we apply the dressing method choosing λ_j^\pm as the locations of the singularities and construct the projector $P_j(x)$ using the constant vectors $|n_{0,j}\rangle$ and $\langle m_{0,j}|$. The normalizing factor $u_{j,-}^{-1}(\lambda)$ in the right hand side of (2.54) is a diagonal matrix that commutes with Π_0 . Then we insert $\chi^\pm(x, \lambda) = u_j(x, \lambda)\chi_0^\pm(x, \lambda)$ in (3.4) and note that the pole singularity of $R^+(x, y, \lambda)$ at $\lambda = \lambda_j^+$ (resp. $R^-(x, y, \lambda)$ at $\lambda = \lambda_j^-$) can come up only from the factor $u_{j,-}^{-1}(y, \lambda)$ (resp. $u(x, \lambda)$). To derive the expressions in (3.9) one needs the explicit form of the projectors $P_j(x)$ and $P_j(y)$ (2.57) and (2.59).

The right hand sides of (3.9) do not vanish if the following conditions

$$(3.10) \quad \begin{aligned} &\Pi_0|n_{0,j}\rangle \neq |n_{0,j}\rangle, \quad \text{or} \quad \Pi_0|n_{0,j}\rangle \neq 0, \\ &\langle m_{0,j}|\Pi_0 \neq \langle m_{0,j}|, \quad \text{or} \quad \langle m_{0,j}|\Pi_0 \neq 0. \end{aligned}$$

hold. In other words if (3.10) hold then the residues (3.9) do not vanish, $R^\pm(x, y, \lambda)$ have simple poles at $\lambda = \lambda_j^\pm$ and by definition λ_j^\pm are discrete eigenvalues of $L(\lambda)$. Eq. (3.10) is equivalent to the condition $I_j \leq k_0 < F_j$. Indeed violating this condition we get either $(\mathbf{1} - \Pi_0)|n_{0,j}\rangle = 0$ or $\Pi_0|n_{0,j}\rangle = 0$ and as a result – vanishing right hand sides in (3.9).

To finish the proof one must check that from the definitions (3.9b) the relations (2.68) follow. Besides $|\mathbf{n}_j^\pm\rangle$ and $\langle \mathbf{m}_j^\pm|$ satisfy:

$$(3.11) \quad i \frac{d|\mathbf{n}_j^\pm\rangle}{dx} + (q(x) - \lambda_j^\pm J)|\mathbf{n}_j^\pm\rangle = 0, \quad i \frac{d\langle \mathbf{m}_j^\pm|}{dx} - \langle \mathbf{m}_j^\pm|(q(x) - \lambda_j^\pm) = 0,$$

where $q(x)$ is given by (2.63). □

COROLLARY 3.3. The discrete spectrum of the Lax operator (1.5) consists of the zeroes of the principal minors $m_j^+(\lambda)$ for $\lambda \in \mathbb{C}_+$ and $m_j^-(\lambda)$ for $\lambda \in \mathbb{C}_-$ provided the conditions (3.10) are satisfied.

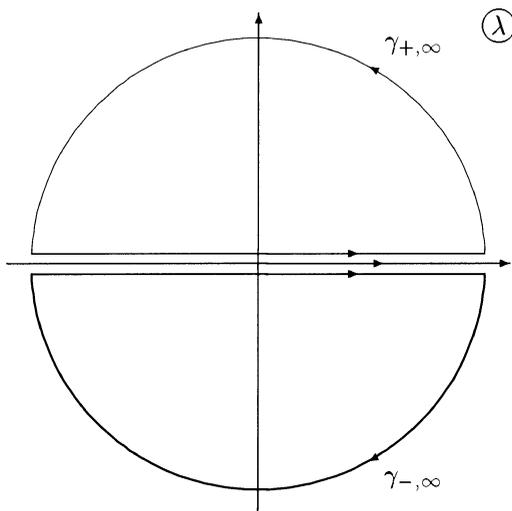


FIGURE 1. The contours $\gamma_{\pm} = \mathbb{R} \cup \gamma_{\pm\infty}$.

Now we can derive the completeness relation for the eigenfunctions of the Lax operator (1.5) by applying the contour integration method (see e.g. [26, 27, 2]) to the integral:

$$(3.12) \quad \mathfrak{J}(x, y) = \frac{1}{2\pi i} \oint_{\gamma_+} d\lambda R^+(x, y, \lambda) - \frac{1}{2\pi i} \oint_{\gamma_-} d\lambda R^-(x, y, \lambda),$$

where the contours γ_{\pm} are shown on the Figure 1. Skipping the details we get:

$$(3.13) \quad \begin{aligned} \delta(x - y) \sum_{s=1}^n \frac{1}{a_s} E_{ss} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\lambda \left\{ \sum_{s=1}^{k_0} |\chi^{[s]+}(x, \lambda)\rangle \langle \hat{\chi}^{[s]+}(y, \lambda)| \right. \\ &\quad \left. - \sum_{s=k_0+1}^n |\chi^{[s]-}(x, \lambda)\rangle \langle \hat{\chi}^{[s]-}(y, \lambda)| \right\} \\ &\quad + 2i \sum_{j=1}^N \nu_j \{ |\mathbf{n}_j^+(x)\rangle \langle \mathbf{m}_j^+(y)| - |\mathbf{n}_j^-(x)\rangle \langle \mathbf{m}_j^-(y)| \}. \end{aligned}$$

This relation (3.13) allows one to expand any vector-function $|z(x)\rangle \in \mathbb{C}^n$ over the eigenfunctions of the system (1.5). Indeed, let us multiply (3.13) on the right by $J |z(y)\rangle$ and integrate over y . This gives:

$$(3.14) \quad \begin{aligned} |z(x)\rangle &= \frac{i}{2\pi} \int_{-\infty}^{\infty} d\lambda \left\{ \sum_{s=1}^{k_0} |\chi^{[s]+}(x, \lambda)\rangle \cdot \zeta_s^+(\lambda) - \sum_{s=k_0+1}^n |\chi^{[s]-}(x, \lambda)\rangle \cdot \zeta_s^-(\lambda) \right\} \\ &\quad + \sum_{j=1}^N \nu_j (|\mathbf{n}_j^+(x)\rangle \zeta_j^+ - |\mathbf{n}_j^-(x)\rangle \zeta_j^-), \end{aligned}$$

where the expansion coefficients are of the form:

$$(3.15) \quad \zeta_s^{\pm}(\lambda) = -i \int_{-\infty}^{\infty} dx \langle \hat{\chi}^{[s]\pm}(x, \lambda) | J | z(x) \rangle, \quad \zeta_j^{\pm}(\lambda) = -i \int_{-\infty}^{\infty} dx \langle \mathbf{m}_j^{\pm} | J | z(x) \rangle.$$

REMARK 3.4. If $q(x) \simeq 0$ then $\chi^+(x, \lambda) \simeq \chi^-(x, \lambda) \simeq \exp(-i\lambda Jx)$ the set \mathfrak{Z} is empty and (3.14) goes into the usual Fourier transform for the space \mathbb{C}^n .

REMARK 3.5. Here we used also the fact that all eigenvalues of J are non-vanishing. In the case when one (or several) of them vanishes we can prove completeness of the eigenfunctions only in a certain subspace of \mathbb{C}^n .

The resolvent for the CBCs is defined quite analogously:

$$(3.16) \quad \begin{aligned} R(x, y, \lambda) &= R_\nu(x, y, \lambda), \quad \lambda \in \Omega_\nu, \\ R_\nu(x, y, \lambda) &= i\chi_\nu(x, \lambda)\Theta^\nu(x - y)\check{\chi}_\nu(x, \lambda), \\ \Theta^\nu(z) &= \theta(-z)\Pi_0^\nu - \theta(z)(\mathbb{1} - \Pi_0^\nu), \quad \Pi_0^\nu = \sum_{s \leq k_{0,\nu}} E_{ss}, \end{aligned}$$

where $\chi_\nu(x, \lambda) = \xi_\nu(x, \lambda)e^{i\lambda Jx}$ and $k_{0,\nu}$ is the number of positive eigenvalues of $\text{Im}(\lambda J)$ in the ν -th ordering.

The following theorem is a specific case of one in [30].

THEOREM 3.6. *Let $q(x)$ satisfy the conditions (C.1) and (C.2) and let $\mathfrak{Z} = \cup_{p=1}^n (\mathfrak{Z}_{2p-1} \cup \mathfrak{Z}_{2p})$ where*

$$(3.17) \quad \begin{aligned} \mathfrak{Z}_{2p-1} &\equiv \{ \lambda_j^+ \omega^{p-1} \in \Omega_{2p-1}, \quad j = 1, \dots, N \}, \\ \mathfrak{Z}_{2p} &\equiv \{ \lambda_j^- \omega^p \in \Omega_{2p}, \quad j = 1, \dots, N \}, \end{aligned}$$

are the sets of zeroes and poles of the minors $m_{\nu,k}(\lambda)$ in the sectors Ω_ν . Then

- (1) $R_\nu(x, y, \lambda)$ is an analytic function of λ for $\lambda \in \Omega_\nu$ having pole singularities at \mathfrak{Z}_ν ;
- (2) $R_\nu(x, y, \lambda)$ is a kernel of a bounded integral operator for $\lambda \in \Omega_\nu$;
- (3) For $\lambda \in l_\nu \cup l_{\nu+1}$ $R_\nu(x, y, \lambda)$ is an uniformly bounded function which is a kernel of an unbounded integral operator;
- (4) $R_\nu(x, y, \lambda)$ satisfies the equation:

$$(3.18) \quad L(\lambda)R_\nu(x, y, \lambda) = \mathbb{1}\delta(x - y).$$

The next natural step is to establish the structure of the singularities of $R_\nu(x, y, \lambda)$ at the points of \mathfrak{Z} . This is done quite analogously by using the dressing factor (2.72). Note that in these matters the symmetry complicates the calculations.

One of the effects of the \mathbb{Z}_n -symmetry is that the sets \mathfrak{Z}_ν are determined uniquely by \mathfrak{Z}_1 and \mathfrak{Z}_0 :

$$(3.19) \quad \mathfrak{Z}_1 = \{ \lambda_j^+ \in \Omega_1, \quad j = 1, \dots, N \}, \quad \mathfrak{Z}_0 = \{ \lambda_j^- \in \Omega_{2n}, \quad j = 1, \dots, N \}.$$

The residue of $R_\nu(x, y, \lambda)$ at the point $\lambda = \lambda_j^\nu$ can be cast into the form:

$$(3.20) \quad \begin{aligned} \text{Res}_{\lambda=\lambda_j^+} R_1(x, y, \lambda) &= -2i\text{Im} \lambda_j^+ |\mathbf{n}_j^+(x)\rangle \langle \mathbf{m}_j^+(x)|, \\ \text{Res}_{\lambda=\lambda_j^-} R_{2n}(x, y, \lambda) &= 2i\text{Im} \lambda_j^- |\mathbf{n}_j^-(x)\rangle \langle \mathbf{m}_j^-(x)|, \end{aligned}$$

where $|\mathbf{n}_j^\pm(x)\rangle$ and $\langle \mathbf{m}_j^\pm(x)|$ are properly normalized eigenvectors of the Lax operator corresponding to the eigenvalues $\lambda_j^\pm \in \Omega_{\pm 1}$. The residues in the other sectors Ω_ν with $\nu \neq 0, 1 \pmod{2n}$ are evaluated from (3.20) by employing eq. (2.31). Here we also have the analog of the condition (3.10).

The derivation of the completeness relation of the eigenfunctions for CBCs with \mathbb{Z}_n -reduction follows the same lines but needs some modifications. Instead of $\tilde{\mathfrak{J}}(x, y)$ (3.12) we should consider

$$(3.21) \quad \tilde{\mathfrak{J}}(x, y) = \sum_{\nu=1}^{2n} \frac{(-1)^{\nu-1}}{2\pi i} \oint_{\gamma_\nu} d\lambda R_\nu(x, y, \lambda),$$

where the contours γ_ν are defined by:

$$(3.22) \quad \gamma_{2\nu-1} = l_{2\nu-1} \cup \gamma_{2\nu-1}^\infty \cup \bar{l}_{2\nu}, \quad \gamma_{2\nu} = \bar{l}_{2\nu} \cup l_{2\nu+1}^\infty \cup \tilde{\gamma}_{2\nu}.$$

Here l_ν are the rays (2.24) oriented from 0 to ∞ ; γ_ν^∞ is the ‘infinite’ arc $R_0 e^{i\varphi_0}$ with $R_0 \gg 1$ and $\pi(\nu-1)/n \leq \varphi_0 \leq \pi\nu/n$; by overbar we denote the same contour with opposite orientation. Thus all the contours $\gamma_{2\nu-1}$ (resp. $\gamma_{2\nu}$) are positively (resp. negatively) oriented.

Now we apply again the contour integration method and get two answers for $\tilde{\mathfrak{J}}(x, y)$. The first, according to Cauchy residue theorem is:

$$(3.23) \quad \tilde{\mathfrak{J}}(x, y) = \sum_{p=1}^{2n} \sum_{j=1}^N \left(\operatorname{Res}_{\lambda=\lambda_j^+ \omega^p} R_{2p+1}(x, y, \lambda) + \operatorname{Res}_{\lambda=\lambda_j^- \omega^p} R_{2p}(x, y, \lambda) \right).$$

Integration along the contours taking into account that $\lim_{\lambda \rightarrow \infty} \chi^\nu(x, \lambda) = \mathbb{1}$ gives:

$$(3.24) \quad \tilde{\mathfrak{J}}(x, y) = \sum_{\nu=1}^{2n} \frac{(-1)^{\nu-1}}{2\pi i} \int_{l_\nu} dx (R_\nu(x, y, \lambda) - R_{\nu-1}(x, y, \lambda)) + J^{-1} \delta(x - y).$$

The completeness relation follows after equating both expressions and taking into account that (compare with (3.20) and (2.31)):

$$(3.25) \quad \begin{aligned} \operatorname{Res}_{\lambda=\lambda_j^+ \omega^p} R_{2p+1}(x, y, \lambda) &= -2i \operatorname{Im} \lambda_j^+ |\mathbf{n}_j^{(2p+1)}(x)\rangle \langle \mathbf{m}_j^{(2p+1)}(x)|, \\ \operatorname{Res}_{\lambda=\lambda_j^- \omega^p} R_{2p}(x, y, \lambda) &= 2i \operatorname{Im} \lambda_j^+ |\mathbf{n}_j^{(2p)}(x)\rangle \langle \mathbf{m}_j^{(2p)}(x)|, \end{aligned}$$

where $|\mathbf{n}_j^{(2p)}(x)\rangle$, $|\mathbf{n}_j^{(2p+1)}(x)\rangle$ (resp. $\langle \mathbf{m}_j^{(2p)}(x)|$, $\langle \mathbf{m}_j^{(2p+1)}(x)|$) are properly normalized discrete eigenfunctions of the CBCs (1.5) (resp. of the adjoint CBCs (2.2)) corresponding to the discrete eigenvalues $\lambda_j^- \omega^{2p}$ and $\lambda_j^+ \omega^{2p}$. For the lack of space we can not provide all the details of the calculations. The final result is similar to the one for GZSs. Namely, any vector-function $|z(x)\rangle \in \mathbb{C}^n$ can be expanded over the eigenfunctions of the CBCs as follows:

$$(3.26) \quad |z(x)\rangle = \sum_{\nu=1}^{2n} \frac{(-1)^{\nu-1}}{2\pi} \int_{l_\nu} d\lambda \left\{ \sum_{s <_\nu k_{0,\nu}} \zeta_{\nu,s}^+(\lambda) |\chi^{\nu,[s]}(x, \lambda)\rangle - \sum_{s >_\nu k_{0,\nu}} \zeta_{\nu,s}^-(\lambda) |\chi^{\nu-1,[s]}(x, \lambda)\rangle \right\} + \sum_{j=1}^N \sum_{\nu=1}^{2n} \operatorname{Im} \lambda_j^+ [\zeta_{\nu,j}^+ |\mathbf{n}_j(x)^{\nu,+}\rangle - \zeta_{\nu,j}^- |\mathbf{n}_j(x)^{\nu,-}\rangle],$$

where the expansion coefficients are given by:

$$\zeta_{\nu,s}^+(\lambda) = -i \int_{-\infty}^{\infty} dx \langle \chi^{\nu,[s]}(x, \lambda) | J | z(x) \rangle,$$

$$(3.27) \quad \zeta_{\nu,s}^-(\lambda) = -i \int_{-\infty}^{\infty} dx \langle \hat{\chi}^{\nu-1,[s]}(x, \lambda) | J | z(x) \rangle,$$

$$\zeta_{\nu,j}^+ = -i \int_{-\infty}^{\infty} dx \langle \mathbf{m}_j^{\nu}(x) | J | z(x) \rangle, \quad \zeta_{\nu,j}^- = -i \int_{-\infty}^{\infty} dx \langle \mathbf{m}_j^{\nu-1}(x) | J | z(x) \rangle.$$

The completeness relations derived for GZSs and CBCs above can be viewed as the spectral decompositions for the generically non-self-adjoint operators $L(\lambda)$.

REMARK 3.7. The special case of a CBCs with \mathbb{Z}_n -symmetry is equivalent to n -th order scalar differential operator [11]. Indeed, one can easily check that the system L (1.5), (1.8) can be written down as:

$$(3.28) \quad L\chi \equiv i \left[\frac{d}{dx} + \sum_{k=1}^n \psi_k(x) K_0^k + i\lambda c_0 \omega^{-1/2} \sum_{k=1}^n \omega^k E_{kk} \right] \chi(x, \lambda) = 0.$$

After similarity transformation with $u_0 = \sum_{s,j=1}^n \omega^{sj} E_{sj}$ goes into:

$$(3.29) \quad \tilde{L}\tilde{\chi} \equiv \frac{1}{i} u_0^{-1} L u_0 \tilde{\chi} \equiv \left[\frac{d}{dx} + \sum_{s=1}^n \phi_s(x) E_{ss} - \tilde{\lambda} \sum_{s=1}^n E_{s,s+1} \right] \tilde{\chi}(x, \lambda) = 0,$$

$$\phi_s(x) = \sum_{k=1}^n \psi_k(x) \omega^{ks}, \quad \tilde{\lambda} = i\lambda c_0 \omega^{-1/2},$$

and can be rewritten as the scalar operator

$$(3.30) \quad L^{(n)}\chi_1 \equiv d_n d_{n-1} \cdots d_2 d_1 \chi_1(x, \lambda) = \tilde{\lambda}^n \chi_1(x, \lambda),$$

where $d_k X(x, \lambda) = dX/dx + \phi_k(x)X(x, \lambda)$. If $\phi_k(x)$ are real functions (additional \mathbb{Z}_2 -reduction of the type (2.6) ensures this) then $L^{(n)}$ is a self-adjoint operator.

REMARK 3.8. The author is aware that these type of derivations need additional arguments to be made rigorous. One of the real difficulties is to find explicit conditions on the potential $q(x)$ that are equivalent to the condition (C.2) or equivalently, to the conditions that $m_k^{\pm}(\lambda)$ have only finite number of simple zeroes. Nevertheless there are situations (e.g., the reflectionless potentials) when all these conditions are fulfilled and all eigenfunctions of $L(\lambda)$ can be explicitly calculated. Another advantage of this approach is the possibility to apply it to Lax operators with more general dependence on λ , e.g., quadratic or polynomial in λ .

The ‘diagonal’ of the resolvent

By the diagonal of the resolvent one usually means $R(x, y, \lambda)$ evaluated at $y = x$. However the definition (3.3) is not continuous for $y = x$ and needs regularization. The simplest possibility is to consider as the diagonal of the resolvent:

$$R(x, \lambda) = \frac{1}{2} (R(x + 0, x, \lambda) + R(x, x + 0, \lambda)).$$

In fact we will consider as the a somewhat more general expression:

$$(3.31) \quad R_P(x, \lambda) = i\chi^{\pm}(x, \lambda) P \hat{\chi}^{\pm}(x, \lambda),$$

where P is a constant diagonal matrix. Obviously $R_P(x, \lambda)$ satisfies

$$(3.32) \quad i \frac{dR_P(x, \lambda)}{dx} + [q(x) - \lambda J, R_P(x, \lambda)] \equiv [L(\lambda), R_P(x, \lambda)] = 0.$$

Thus $R_P(x, \lambda)$ belongs to the kernel of the operator $[L(\lambda), \cdot]$. Due to the fact that $\chi^\pm(x, \lambda)$ is the FAS and satisfies a RHP with canonical normalization we find:

$$(3.33) \quad R_P(x, \lambda) = iP + \sum_{k=1}^{\infty} R_P^{(k)}(x)\lambda^{-k}.$$

The coefficients $R_P^{(k)}(x)$ can be expressed through $q(x)$ using the recursion relations generalizing the ones of AKNS [2, 18, 37, 27]. These relations are solved by the recursion operators Λ_\pm which have the form:

$$(3.34) \quad \Lambda_\pm X = \text{ad}_J^{-1} \left(i \frac{dX}{dx} + P_0([q_x], X(x)) + i \left[q(x), \int_{\pm\infty}^x dy [q(y), X(y)] \right] \right),$$

where P_0 is the projector onto the off-diagonal part of the matrix $P_0 X = X_0^f$, the matrix $X(x)$ in (3.34) satisfies $X \equiv P_0 X$ and

$$(\text{ad}_J^{-1} X_0^f)_{ij} = \frac{(X_0^f)_{ij}}{a_i - a_j}.$$

The coefficients $R_P^{(k)}(x)$ can be expressed in compact form through Λ_\pm as follows:

$$(3.35) \quad R_P^{k+1f} = \Lambda_\pm R_P^{kf} = -\Lambda_\pm^k \text{ad}_J^{-1} [iP, q(x, t)],$$

$$(3.36) \quad R_P^{(k)d} = i \int_{\pm\infty}^x dy (\mathbb{1} - P_0) \left([q(y, t), R_P^{(k)f}] \right) + \lim_{x \rightarrow \pm\infty} R_P^{(k)d}(x, t).$$

Quite naturally these coefficients, or rather the diagonal of the resolvent generates [17, 10, 18]:

- the class of NLEE. Given the dispersion law, e.g., $f(\lambda) = \lambda^N P$ of the NLEE we can write down the equation itself by:

$$(3.37) \quad -i \frac{dq}{dt} + i \frac{(dR_P^{(N)})^f}{dx} + P_0([q(x, t), R_P^{(N)}(x, t)]) = 0.$$

- the corresponding Lax representations, or in other words, the M -operators for each of these NLEE as follows:

$$(3.38) \quad V_P^{(N)}(x, \lambda) = \sum_{k=0}^N R_P^{(k)}(x)\lambda^{N-k}.$$

- the integrals of motion of the corresponding NLEE. This follows from

THEOREM 3.9 ([18]). *The quantities*

$$(3.39) \quad R_{\Pi^{(k)}}^\pm(x, \lambda) = i\chi^\pm(x, \lambda)\Pi^{(k)}\hat{\chi}^\pm(x, \lambda), \quad \Pi^{(k)} = \sum_{s=1}^k E_{ss} - \frac{k}{n} \mathbb{1},$$

satisfy the relations

$$(3.40) \quad \int_{-\infty}^{\infty} dx \text{tr} \left(R_{\Pi^{(k)}}^\pm(x, \lambda) J - i\Pi^{(k)} J \right) = -\frac{d}{d\lambda} \mathfrak{D}_k(\lambda),$$

where $\mathfrak{D}_k^\pm(\lambda)$ is defined by (2.45).

Combined with the (3.34) we can deduce that the diagonal of the resolvent and the recursion operator

$$(3.41) \quad (\Lambda_\pm - \lambda)R_P^\pm(x, \lambda) = i[P, \text{ad}_J^{-1}q(x)],$$

directly reproduce the generating functionals of the conserved quantities.

The termin ‘squared’ solutions and recursion operator do not reflect properly the algebraic properties of these objects. The recursion operators Λ_{\pm} can be understood as the Lax operator $L(\lambda)$ in the adjoint representation. One of the definitions of the adjoint representation means that we should replace each element $U(x, \lambda) \in \mathfrak{g}$ by $\text{ad}_{U(x, \lambda)} \cdot = [U(x, \lambda), \cdot]$. Therefore due to (3.32) we can view the diagonal of the resolvent $R_{\mathcal{P}}^{\pm}(x, \lambda)$ as the eigenfunction of $L(\lambda)$ in the adjoint representation. It remains to project out the kernel of ad_J in order to derive Λ_{\pm} from $L(\lambda)$.

The ‘squared’ solutions are eigenfunctions of Λ_{\pm} and belong to a linear space, which is the co-adjoint orbit of $\hat{\mathfrak{g}}_+^*$ determined by J . The gauge covariant way to introduce them involves the FAS of $L(\lambda)$ and is:

$$(3.42) \quad e_{ij}^{\pm}(x, \lambda) = P_0(\chi^{\pm}(x, \lambda) E_{ij} \hat{\chi}^{\pm}(x, \lambda)), \quad h_j^{\pm}(x, \lambda) = P_0(\chi^{\pm}(x, \lambda) H_j \hat{\chi}^{\pm}(x, \lambda)),$$

where $\chi^{\pm}(x, \lambda)$ are the FAS of $L(\lambda)$ GZSs. The similarity transformation by $\chi^{\pm}(x, \lambda)$ is the adjoint action of the group \mathfrak{G} on the algebra \mathfrak{g} ; therefore $e_{\alpha}^{\pm}(x, \lambda)$ and $h_j^{\pm}(x, \lambda)$ are elements again of \mathfrak{g} . The projection $\Pi_0 = \text{ad}_J^{-1} \text{ad}_J$ is a natural linear operator on \mathfrak{g} . Besides the ‘squared’ solutions are analytic functions of λ having both poles and zeroes at λ_j^{\pm} .

More detailed analysis based on the Wronskian relations reveals several other important aspects [36, 19, 30] of the ‘squared’ solutions of GZSs. First, the sets

$$\{e_{ij}^+(x, \lambda), e_{ji}^-(x, \lambda)\}, e_{ij;k}^+(x), e_{ji;k}^-(x), \dot{e}_{ij;k}^+(x), \dot{e}_{ji;k}^-(x), \quad i < j, k = 1, \dots, N\}$$

and

$$\{e_{ji}^+(x, \lambda), e_{ij}^-(x, \lambda)\}, e_{ji;k}^+(x), e_{ij;k}^-(x), \dot{e}_{ji;k}^+(x), \dot{e}_{ij;k}^-(x), \quad i < j, k = 1, \dots, N\}$$

form complete sets of functions on \mathfrak{M} that realize the mapping $\mathfrak{M} \leftrightarrow \mathfrak{T}$. Here by $e_{ji;k}^{\pm}(x)$ and $\dot{e}_{ji;k}^{\pm}(x)$ we have denoted:

$$e_{ij;k}^{\pm}(x) = e_{ij}^{\pm}(x, \lambda_k^{\pm}), \quad \dot{e}_{ij;k}^{\pm}(x) = \left. \frac{de_{ij}^{\pm}(x, \lambda)}{d\lambda} \right|_{\lambda=\lambda_k^{\pm}}.$$

Second, it is possible to expand the potential $[P, \text{ad}_J^{-1} q(x, t)]$ and its variation $\text{ad}_J^{-1} \delta q(x)$ over each of the complete sets shown above. The corresponding expansion coefficients are expressed through \mathfrak{T} and their variations. These facts constitute the grounds on which one can show that the ISM can be understood as a generalized Fourier transform. The important difference as compare to the standard Fourier transform is in the fact that the operator L (as well as the operators Λ_{\pm}) allows for discrete eigenvalues. Therefore the completeness relations involve both integrals along the continuous spectrum and sum over the discrete eigenvalues. In the usual Fourier transform the discrete eigenvalues are absent.

Hamiltonian properties of the NLEE

Here we briefly formulate the Hamiltonian properties of the NLEE paying more attention to its algebraic structure. This has been widely studied problem, see [3, 11, 39, 14, 17, 10, 13, 48, 18, 19] and the numerous references therein.

In doing so we follow mainly the ideas of [39] with a natural generalization from $sl(2)$ to $sl(n)$ -algebras. The main idea in these papers is the possibility to write down the Lax equation (1.4) in explicitly Hamiltonian form as the co-adjoint

action of $\tilde{\mathfrak{g}}$ on its dual $\tilde{\mathfrak{g}}^*$. Obviously any non-trivial grading in \mathfrak{g} (resp. $\tilde{\mathfrak{g}}, \hat{\mathfrak{g}}$) will reflect into a corresponding grading of the dual algebra \mathfrak{g}^* (resp. $\tilde{\mathfrak{g}}^*, \hat{\mathfrak{g}}^*$).

Below we will need also the Cartan-Weyl basis of $sl(n)$. Choosing for definiteness the typical $n \times n$ representation we fix it up by:

$$(3.43) \quad \mathfrak{h} \equiv \text{l.c.} \{H_i = E_{ii} - E_{i+1,i+1}, \quad i = 1, \dots, n-1\}, \quad \{E_{ij}, \quad i \neq j\}.$$

As invariant bilinear form we can use $\langle X, Y \rangle = \text{tr}(XY)$. Then the commutation relations can be written in the form:

$$(3.44) \quad \begin{aligned} [H_i, E_{jk}] &= (e_i - e_{i+1}, e_j - e_k)E_{jk}, & j \neq k, \\ [E_{jk}, E_{kl}] &= E_{jl}, & [E_{jk}, E_{lj}] = -E_{lk}, & l \neq j, \\ [E_{jk}, E_{kj}] &= \sum_{s=j}^{k-1} H_s, & j < k. \end{aligned}$$

By e_k above we mean an orthonormal basis in the n -dimensional Euclidean space with a standard scalar product: $(e_j, e_k) = \delta_{jk}$. Those, who are familiar with Lie algebras will recognize $e_i - e_{i+1}$ as the simple roots of $sl(n)$ and the set of $e_j - e_k, j \neq k$ as the root system of $sl(n)$.

If $C = \mathbf{1}$ (i.e. with the trivial grading) each of the matrices $U_k(x)$ in (1.19) is of generic form:

$$(3.45) \quad U_k(x) = \sum_{j=1}^{n-1} u_j^{(k)} H_j + \sum_{j \neq p} u_{jp}^{(k)} E_{jp}.$$

The coefficients $u_j^{(k)}(x), u_{jp}^{(k)}(x)$ can be viewed as linear functionals on $u_k(x)$ and thus they belong to \mathfrak{g}^* . Using the bilinear form (1.27) they can be interpreted as linear functionals on $\hat{\mathfrak{g}}$ and thus as elements also of $\hat{\mathfrak{g}}^*$. The algebraic structure on $\tilde{\mathfrak{g}}^*$ can be introduced in analogy with the commutation relations (3.44), namely:

$$(3.46) \quad \begin{aligned} \left\{ u_i^{(s)}(x), u_{j,j+k}^{(m)}(y) \right\}_p &= (e_i - e_{i+1}, e_j - e_k) u_{j,j+k}^{(s+m-p)}(x) \delta(x-y), \\ \left\{ u_{i,i+k}^{(s)}(x), u_{i+k,j}^{(m)}(y) \right\}_p &= u_{i,j}^{(s+m-p)}(x) \delta(x-y), \\ \left\{ u_{i,i+k}^{(s)}(x), u_{j+k,i}^{(m)}(y) \right\}_p &= -u_{j+k,i+k}^{(s+m-p)}(x) \delta(x-y), \\ \left\{ u_{i,i+k}^{(s)}(x), u_{i+k,i}^{(m)}(y) \right\}_p &= \sum_{l=i}^{i+k-1} u_l^{(s+m-p)}(x) \delta(x-y) + i \delta_{s+m,p} \delta'(x-y). \end{aligned}$$

The derivation of these relations follows [39] in a rather straightforward manner; though a bit tedious, it can be generalized also to any simple Lie algebra.

Note that if $p = -1$ then the term with $\delta'(x-y)$ disappears and the Poisson brackets (3.46) become ultralocal. Then we can rewrite them in a compact form using the classical r -matrix [13]:

$$(3.47) \quad \left\{ U(x, \lambda) \otimes U(y, \mu) \right\}_{-1} = [r(\lambda - \mu), U(x, \lambda) \otimes \mathbf{1} + \mathbf{1} \otimes U(x, \mu)] \delta(x-y),$$

$$(3.48) \quad r(\lambda - \mu) = \frac{\Pi_0}{\lambda - \mu}, \quad \Pi_0 = \sum_{i,j=1}^n E_{ij} \otimes E_{ji}.$$

The left hand side of (3.47) has the structure of the usual tensor product of $n \times n$ matrices, but instead of taking the product one should rather take the Poisson bracket between the corresponding matrix elements of $U(x, \lambda)$ and $U(y, \mu)$.

The relations (3.47) are local in the sense that for the evaluation of the left hand side of (3.47) we need to use only the Poisson brackets between the matrix elements of $U(x, \lambda)$ and do not need the boundary conditions on the potentials. The effectiveness of the r -matrix, when it exists, is in the possibility to evaluate the Poisson brackets between the matrix elements of the scattering matrix $T(\lambda)$. To do this we need to ‘integrate’ (3.47) which needs to take into account also the boundary conditions. For periodic boundary conditions on $q(x)$ this gives:

$$(3.49) \quad \left\{ T(\lambda) \otimes T(\mu) \right\}_{-1} = [r(\lambda - \mu), T(\lambda) \otimes T(\mu)].$$

For vanishing boundary conditions on $q(x)$ and $J = J^*$ the calculations need some additional considerations with the result (see [13]):

$$(3.50) \quad \left\{ T(\lambda) \otimes T(\mu) \right\}_{-1} = r_+(\lambda - \mu)T(\lambda) \otimes T(\mu) - T(\lambda) \otimes T(\mu)r_-(\lambda - \mu),$$

$$r_{\pm}(\lambda - \mu) = \frac{1}{\lambda - \mu} \sum_{j=1}^n E_{jj} \otimes E_{jj} \mp i\pi\delta(\lambda - \mu) \sum_{i \neq j=1}^n E_{ij} \otimes E_{ji}.$$

From both relations (3.49) and (3.50) there follows that the principal minors $m_k^{\pm}(\lambda)$ commute with respect to the Poisson brackets (3.46) [19], i.e.:

$$(3.51) \quad \{ \mathfrak{D}_k(\lambda), \mathfrak{D}_j(\mu) \}_{-1} = 0.$$

Since $\mathfrak{D}_k(\lambda)$ are the generating functionals of integrals of motion $\mathfrak{D}_k^{(s)}$ (see eq. (2.51)), then eq. (3.51) means that all these integrals are in involution with respect to these Poisson brackets.

The \mathbb{Z}_n -symmetry may modify substantially some of the above results. Indeed, it can be viewed as a set of constraints on the phase space \mathfrak{M} and on the generic Poisson brackets (3.46). Then one should evaluate the corresponding Dirac brackets on the reduced phase space. However in the case of the \mathbb{Z}_n -NLS equation (1.3) with Lax operator L given by (1.5), (1.8) somewhat surprisingly the approach of [39] gives us directly the correct answer. If we define $\psi_j(x, t)$ as linear functionals of $U(x, t, \lambda) = q(x, t) - \lambda J$ by:

$$(3.52) \quad \psi_j(x, t) = \frac{1}{n} \text{tr} (U(x, t, \lambda) K^{n-j}),$$

and make use of (1.8) then the set of Poisson brackets in (3.46) simplify to

$$(3.53) \quad \{ \psi_j(x, t), \psi_k(y, t) \} = \delta_{k+j-n} \delta'(x - y).$$

Together with the Hamiltonian $H = \omega^2 M_{1,1}^{(2)}$ (2.53) they provide the Hamiltonian formulation of (1.3). Unfortunately this Poisson brackets are not ultra-local and the corresponding Lax operator does not allow classical r -matrix of the form (3.47).

For the affine Toda chain (1.2) the simplest Poisson brackets are provided by:

$$(3.54) \quad \left\{ \frac{dQ_j}{dx}, Q_k(y, t) \right\} = \delta_{kj} \delta(x - y).$$

The corresponding Lax operator (1.9) unlike the previous case allows classical r -matrix satisfying (3.47) which however has more complicated dependence on $\lambda - \mu$; it is known as the trigonometric r -matrix [38].

Another special property of the \mathbb{Z}_n -symmetric CBCs concerns the existence of the so-called symplectic basis [25]. The elements of these bases are special linear combinations of the ‘squared solutions’ (3.42) which are also complete in \mathfrak{M} and which are such that the expansion coefficients of $\delta q(x, t)$ over it produce the variations of the action-angle variables of the corresponding set of NLEE. In [25] this basis was worked out for the Zakharov-Shabat system related to the $sl(2)$ algebra. For GZSs related to algebras of higher rank such basis is yet unknown although it must exist since the action-angle variables for them are known [40, 7].

For the \mathbb{Z}_n -symmetric CBCs the construction of the symplectic basis is very much like the one in [25] due to the fact that the subalgebras \mathfrak{g}_ν related to each of the rays l_ν are direct sums of $sl(2)$ subalgebras. It is a complete set of functions on the phase space of the corresponding \mathbb{Z}_n -symmetric NLEE (1.1) and (1.2). Skipping the details we just give the explicit expressions for the set \mathfrak{A} of action-angle variables of the \mathbb{Z}_n -NLS equation in terms of the scattering data of its Lax operator. Obviously \mathfrak{A} will consist of two sets of functions $\mathfrak{A} = \mathfrak{A}_0 \cup \mathfrak{A}_1$ each set defined on the ray l_0 and l_1 respectively:

$$\begin{aligned}\mathfrak{A}_0 &\equiv \{\pi_{ij}(\lambda), \kappa_{ij}(\lambda), \quad \lambda \in l_0, \quad i + j = 2 \pmod{n}\}, \\ \mathfrak{A}_1 &\equiv \{\pi_{ij}(\lambda), \kappa_{ij}(\lambda), \quad \lambda \in l_1, \quad i + j = 1 \pmod{n}\},\end{aligned}$$

where

$$(3.55) \quad \pi_{ij}(\lambda) = -\frac{1}{\pi} \ln(1 + \rho_{ij}^+ \rho_{ij}^-), \quad \kappa_{ij}(\lambda) = -\frac{i}{2} \frac{b_{ij}^+(\lambda)}{b_{ij}^-(\lambda)}, \quad \rho_{ij}^+(\lambda) = \frac{b_{ij}^+(\lambda)}{a_{ij}^+(\lambda)},$$

and the coefficients $a_{ij}^+(\lambda)$, $b_{ij}^+(\lambda)$ were introduced in (2.30).

Quite analogous are the expressions for the action-angle variables for the two-dimensional affine Toda chain provided we use the scattering data of the Lax operator (1.10).

4. CONCLUSION

The restricted space did not allow us to give more details or explanations on these and related problems. We only mention some of them below.

One such important to our mind result is the interpretation of the ISM as a generalized Fourier transform. In its derivation for the GZSs and CBCs [27, 19, 30] both algebraic methods and analytic ones were used. As a result the pair-wise equivalence of the symplectic structures in the hierarchy becomes obvious.

The approach based on the Kac-Moody algebras is a natural basis for the Hamiltonian hierarchies. If one can derive a bi-Hamiltonian formulation of a given NLEE then there is a whole hierarchy of them related by a recursion operator Λ . Here we mention the paper [15] where the operator Λ was derived as the ‘ratio’ of two such Hamiltonian structures for the N -wave equations. The result, of course coincides with the natural expression for Λ obtained with the AKNS recursion method and whose spectral theory was constructed by other means in [27, 18].

The method based on the diagonal of the resolvent of the Lax operator started by Gel’fand and Dickey [17, 10] can be viewed also as a formal algebraic one.

The authors studied by algebraic means the ring of operators, commuting with L . They expressed most of the quantities, including the diagonal of the resolvent of L , as series over fractional powers of L and did not investigate the existence and convergence of these series. Once identified with the expression (3.31) in terms of the FAS these problems find their natural and positive solution.

Besides the classical r -matrix corresponding to the ultralocal Poisson brackets there exist also dynamical r -matrices depending on the fields $q_{ij}(x)$ in the NLEE. One of the problems, that is still not solved is to find the interrelation between the dynamical r -matrices, r and the recursion operator Λ .

Finally, we should mention that both approaches have been further generalized. For example, the analytic approach was generalized from a local RHP to a nonlocal RHP and to ∂ -bar problem (also local and nonlocal), see [1, 50, 37]. This allowed to treat NLEE of soliton type in $2 + 1$ dimensions.

Another direction is to study Lax operators with more general λ -dependence such as polynomial, or rational [51].

Obviously all results concerning spectral decompositions can be formulated in a gauge covariant way thus allowing to treat also gauge equivalent NLEE [28, 29, 19].

The algebraic approach was also generalized to use as a basis infinite dimensional algebras such as Virasoro algebra, $W_{1+\infty}$ etc. which lead to the important construction of the Japanese τ -function and its relation to the soliton theory, see [32, 14].

Thus we just outlined the beginning of all this and so it is time to stop.

Acknowledgements

The author wishes to thank Professor D. J. Kaup and Professor R. Choudhury for financial support making his participation in the conference possible as well as for useful discussions. I am grateful to Professors G. Vilasi, S. de Filippo and M. Salerno for useful discussions and warm hospitality at Salerno University where this paper was finished.

Appendix A. Gauss decompositions

The Gauss decompositions mentioned above have natural group-theoretical interpretation and can be generalized to any semi-simple Lie algebra. It is well known that if given group element allows Gauss decompositions then its factors are uniquely determined. Below we write down the explicit expressions for the matrix elements of $T^\pm(\lambda)$, $S^\pm(\lambda)$, $D^\pm(\lambda)$ through the matrix elements of $T(\lambda)$:

$$(A.1) \quad T_{pj}^-(\lambda) = \frac{1}{m_j^+(\lambda)} \left\{ \begin{matrix} 1, 2, \dots, j-1, p \\ 1, 2, \dots, j-1, j \end{matrix} \right\}_{T(\lambda)}^{(j)},$$

$$(A.2) \quad \hat{T}_{jp}^-(\lambda) = \frac{(-1)^{j+p}}{m_{j-1}^+(\lambda)} \left\{ \begin{matrix} 1, 2, \dots, \check{p}, \dots, j \\ 1, 2, \dots, p, \dots, j-1 \end{matrix} \right\}_{T(\lambda)}^{(j-1)},$$

$$(A.3) \quad S_{pj}^+(\lambda) = \frac{(-1)^{p+j}}{m_{j-1}^+(\lambda)} \left\{ \begin{matrix} 1, 2, \dots, p, \dots, j-1 \\ 1, 2, \dots, \check{p}, \dots, j \end{matrix} \right\}_{T(\lambda)}^{(j)},$$

$$(A.4) \quad \hat{S}_{jp}^+(\lambda) = \frac{1}{m_j^+(\lambda)} \left\{ \begin{matrix} 1, 2, \dots, j-1, j \\ 1, 2, \dots, j-1, p \end{matrix} \right\}_{T(\lambda)}^{(j-1)},$$

$$(A.5) \quad T_{pj}^+(\lambda) = \frac{1}{m_{n-j+1}^-(\lambda)} \left\{ \begin{matrix} p, j+1, \dots, n-1, n \\ j, j+1, \dots, n-1, n \end{matrix} \right\}_{T(\lambda)}^{(n-j+1)},$$

$$(A.6) \quad \hat{T}_{jp}^+(\lambda) = \frac{(-1)^{p+j}}{m_{n-j}^-(\lambda)} \left\{ \begin{matrix} j, j+1, \dots, \check{p}, \dots, n \\ j+1, j+2, \dots, p, \dots, n \end{matrix} \right\}_{T(\lambda)}^{(n-j)},$$

$$(A.7) \quad S_{pj}^-(\lambda) = \frac{(-1)^{p+j}}{m_{n-j}^-(\lambda)} \left\{ \begin{matrix} j+1, j+2, \dots, p, \dots, n \\ j, j+1, \dots, \check{p}, \dots, n \end{matrix} \right\}_{T(\lambda)}^{(n-j)},$$

$$(A.8) \quad \hat{S}_{jp}^-(\lambda) = \frac{1}{m_{n-j+1}^-(\lambda)} \left\{ \begin{matrix} j, j+1, \dots, n-1, n \\ p, j+1, \dots, n-1, n \end{matrix} \right\}_{T(\lambda)}^{(n-j+1)},$$

where

$$(A.9) \quad \left\{ \begin{matrix} i_1, i_2, \dots, i_k \\ j_1, j_2, \dots, j_k \end{matrix} \right\}_{T(\lambda)}^{(k)} = \det \begin{vmatrix} T_{i_1 j_1} & T_{i_1 j_2} & \dots & T_{i_1 j_k} \\ T_{i_2 j_1} & T_{i_2 j_2} & \dots & T_{i_2 j_k} \\ \vdots & \vdots & \ddots & \vdots \\ T_{i_k j_1} & T_{i_k j_2} & \dots & T_{i_k j_k} \end{vmatrix}$$

is the minor of order k of $T(\lambda)$ formed by the rows i_1, i_2, \dots, i_k and the columns j_1, j_2, \dots, j_k ; by \check{p} we mean that p is missing.

From the formulae above we arrive to the following

COROLLARY A.1. *In order that the group element $T(\lambda) \in SL(n, \mathbb{C})$ allows the first (resp. the second) Gauss decomposition (2.16) is necessary and sufficient that all upper- (resp. lower-) principle minors $m_k^+(\lambda)$ (resp. $m_k^-(\lambda)$) are not vanishing.*

These formulae hold true also if we need to construct the Gauss decomposition of an element of the orthogonal $SO(n)$ group. Here we just note that if $T(\lambda) \in SO(n)$ then

$$(A.10) \quad S_0(T(\lambda))^T S_0^{-1} = T^{-1}(\lambda),$$

where

$$(A.11) \quad S_0 = \sum_{k=1}^{n_0} (-1)^{k+1} (E_{k, n+1-k} + E_{n+1-k, k}), \quad \text{if } n = 2n_0,$$

$$S_0 = \sum_{k=1}^{n_0} (-1)^{k+1} (E_{k, n+1-k} + E_{n+1-k, k}) + (-1)^{n_0} E_{n_0+1, n_0+1}, \quad \text{if } n = 2n_0 + 1.$$

One can check that if $T(\lambda)$ satisfies (A.10) then each of the factors $T^\pm(\lambda)$, $S^\pm(\lambda)$ and $D^\pm(\lambda)$ also satisfy (A.10) and thus belong to the same group \mathfrak{G} . In addition we have the following interrelations between the principal minors of $T(\lambda)$:

$$(A.12) \quad \begin{aligned} m_j^\pm(\lambda) &= m_{n-j}^\pm(\lambda), & \text{for } SO(n), \\ m_j^\pm(\lambda) &= m_{n-j}^\pm(\lambda), & \text{for } SP(n), \end{aligned}$$

Appendix B. Dispersion relations for $\mathfrak{D}_k(\lambda)$ and $\ln m_{\nu, k}^+(\lambda)$

Let us introduce the functions $f_k^\pm(\lambda)$:

$$f_k^+(\lambda) = \frac{m_k^+(\lambda)}{R_k(\lambda)}, \quad f_{n-k}^-(\lambda) = R_k(\lambda) m_{n-k}^-(\lambda), \quad R_k(\lambda) = \prod_{j=1}^N \left(\frac{\lambda - \lambda_j^+}{\lambda - \lambda_j^-} \right)^{b_{jk}},$$

which like $m_k^\pm(\lambda)$ are: i) analytic for $\lambda \in \mathbb{C}_\pm$; ii) satisfy $\lim_{\lambda \rightarrow \infty} f_k^\pm(\lambda) = 1$. Besides, $f_k^\pm(\lambda)$ have no zeroes in their regions of analyticity and therefore the functions $\ln f_k(\lambda)$ are analytic for $\lambda \in \mathbb{C}_\pm$ and tend to 0 for $\lambda \rightarrow \infty$. This allows one to apply the Plemelj-Sokhotzky formula with the result:

$$(B.1) \quad \tilde{\mathfrak{D}}_k(\lambda) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{d\mu}{\mu - \lambda} \ln (f_k^+(\mu) f_{n-k}^-(\mu)),$$

where

$$(B.2) \quad \tilde{\mathfrak{D}}_k(\lambda) = \begin{cases} \ln f_k^+(\lambda), & \lambda \in \mathbb{C}_+ \\ -\ln f_{n-k}^-(\lambda), & \lambda \in \mathbb{C}_-. \end{cases}$$

It remains to insert the above definitions of $f_k^\pm(\lambda)$ into (B.1) and (B.2) to derive Eqs. (2.45), (2.46).

The dispersion relation (2.50) is derived analogously treating the integral

$$(B.3) \quad \tilde{\mathfrak{J}}(\lambda) = \sum_{\nu=1}^{2n} \frac{(-1)^{\nu-1}}{2\pi i} \oint_{\gamma_i} \frac{d\mu}{\mu - \lambda} \ln \left\{ m_{\nu,k}^+(\mu) \prod_{\eta=1}^n \prod_{j=1}^N \left(\frac{\mu - \lambda_{j,k}^+}{\mu - \lambda_{j,k}^-} \right)^{b_{k,j}^\eta} \right\},$$

with $\lambda \in \Omega_\nu$ and the contours γ_i as in (3.22).

References

- [1] M. J. Ablowitz, A. S. Fokas, R. Anderson. *The direct linearising transform and the Benjamin-Ono equation*. Phys. Lett. **93A**, n.8, 375–378, 1983.
- [2] M. J. Ablowitz, D. J. Kaup, A. C. Newell, H. Segur. *The inverse scattering transform – Fourier analysis for nonlinear problems*. Studies in Appl. Math. **53**, n. 4, 249–315, 1974.
- [3] M. Adler. *On a trace functional for formal pseudo-differential operators and the symplectic structure of the Korteweg-de Vries equations*. Inv. Math. **50**, 219–248 (1979).
- [4] N. I. Akhiezer, I. M. Glazman. *Theory of Linear operators in Hilbert space*. Translated from Russian, New York, F. Ungar (1961–1963).
- [5] R. Beals, R. R. Coifman. *Scattering and inverse scattering for first order systems*. Commun. Pure & Appl. Math. **37**, 39 (1984).
- [6] R. Beals, R. R. Coifman. *Inverse scattering and evolution equations*. Commun. Pure & Appl. Math. **38**, 29 (1985).
- [7] R. Beals, D. H. Sattinger. *On the complete integrability of completely integrable systems*. Commun. Math. Phys. **138**, 409 (1991).
- [8] P. J. Caudrey. *The inverse problem for the third order equation $u_{xxx} + q(x)u_x + r(x)u = -i\zeta^3 u$* . Phys. Lett. A **79A**, 264 (1980);
— *The inverse problem for a general $n \times n$ spectral equation*. Physica D **D6**, 56 (1982).
- [9] F. Calogero, A. Degasperis. *Spectral transform and solitons. Vol. I*. North Holland, Amsterdam, 1982.
- [10] L. A. Dickey. *Soliton equations and Hamiltonian systems*. Advanced series in Math. Phys., **12**, World Scientific, (1991).
- [11] V. Drinfel'd, V. V. Sokolov. *Lie Algebras and equations of Korteweg - de Vries type*. Sov. J. Math. **30**, 1975–2036 (1985).
- [12] N. Dunford, J. T. Schwartz. *Linear operators. vol. 2, Spectral theory. Self-adjoint operators in Hilbert space*. (1963), Interscience Publishers, Inc., NY.
- [13] L. D. Faddeev, L. A. Takhtadjan. *Hamiltonian methods in the theory of solitons*. (Springer Verlag, Berlin, 1987).
- [14] H. Flaschka, A. C. Newell, T. Ratiu. *Kac-Moody Lie algebras and soliton equations. II. Lax equations associated with $A_1^{(1)}$* . Physica D **9D**, 300–323 (1983).
- [15] A. Fokas, R. L. Andersson. *On the use of isospectral eigenvalue problems for obtaining hereditary symmetries for Hamiltonian systems*. J. Math. Phys. **23**, No 6, 1066–1073 (1982).
- [16] F. D. Gakhov. *Boundary value problems*. Translated from Russian ed. I. N. Sneddon, (Oxford, Pergamon Press, 1966).

- [17] I. M. Gel'fand, L. A. Dickey. *Funct. Anal. Appl.* **11**(2), 11 (1977) (In Russian).
- [18] V. S. Gerdjikov. *On the spectral theory of the integro-differential operator Λ , generating nonlinear evolution equations.* *Lett. Math. Phys.* **6**, n. 6, 315–324, (1982).
- [19] V. S. Gerdjikov. *Generalized Fourier transforms for the soliton equations. Gauge covariant formulation.* *Inverse Problems* **2**, n. 1, 51–74, 1986.
— *Generating operators for the nonlinear evolution equations of soliton type related to the semisimple Lie algebras.* Doctor of Sciences Thesis, 1987, JINR, Dubna, USSR, (In Russian).
- [20] V. S. Gerdjikov. *Z_N -reductions and new integrable versions of derivative nonlinear Schrödinger equations.* In *Nonlinear evolution equations: integrability and spectral methods*, Ed. A. P. Fordy, A. Degasperis, M. Lakshmanan, Manchester University Press, (1981), p. 367–379.
- [21] V. S. Gerdjikov. *Generalised Fourier transforms for the soliton equations. Gauge covariant formulation.* *Inverse Problems* **2**, n. 1, 51–74, (1986).
- [22] V. S. Gerdjikov. *Complete integrability, gauge equivalence and Lax representations of the inhomogeneous nonlinear evolution equations.* *Theor. Math. Phys.* **92**, 374–386 (1992).
- [23] V. S. Gerdjikov, G. G. Grahovski, R. I. Ivanov and N. A. Kostov. *N -wave interactions related to simple Lie algebras.*
— *\mathbb{Z}_2 -reductions and soliton solutions.* *Inverse Problems* **17**, 999–1015 (2001).
- [24] V. S. Gerdjikov, M. I. Ivanov. *Expansions over the “squared” solutions and the inhomogeneous nonlinear Schrödinger equation.* *Inverse Problems* **8**, 831–847 (1992).
- [25] V. S. Gerdjikov, E. Kh. Khristov. *On the evolution equations solvable with the inverse scattering problem. I. The spectral theory.* *Bulgarian J. Phys.* **7**, No.1, 28–41, (1980). (In Russian);
— *On the evolution equations solvable with the inverse scattering problem. II. Hamiltonian structures and Backlund transformations.* *Bulgarian J. Phys.* **7**, No.2, 119–133, (1980) (In Russian).
- [26] V. S. Gerdjikov, P. P. Kulish. *Complete integrable Hamiltonian systems related to the non-self-adjoint Dirac operator.* *Bulgarian J. Phys.* **5**, No.4, 337–349, (1978), (In Russian).
- [27] V. S. Gerdjikov, P. P. Kulish. *The generating operator for the $n \times n$ linear system.* *Physica D*, **3D**, n. 3, 549–564, 1981.
- [28] V. S. Gerdjikov, A. B. Yanovsky. *Gauge covariant formulation of the generating operator. II. Systems on homogeneous spaces.* *Phys. Lett. A*, **110A**, n. 1, 53–58, 1985.
- [29] V. S. Gerdjikov, A. B. Yanovsky. *Gauge covariant formulation of the generating operator. I.* *Commun. Math. Phys.* **103A**, n. 4, 549–568, 1986.
- [30] V. S. Gerdjikov, A. B. Yanovski. *Completeness of the eigenfunctions for the Caudrey–Beals–Coifman system.* *J. Math. Phys.* **35**, no. 7, 3687–3725 (1994).
- [31] S. Helgasson. *Differential geometry, Lie groups and symmetric spaces.* Academic Press, 1978.
- [32] M. Jimbo, T. Miwa. *Solitons and infinite dimensional algebras.* *Publications RIMS* **19**, 943–1000, 1983.
- [33] V. G. Kac. *Infinite dimensional Lie algebras.* *Progress in Mathematics*, vol. **44**, Boston, Birkhauser, 1983.
V.G.Kac, A. K. Raina. *Bombay lectures on highest weight representations of infinite dimensional Lie algebras.* *Advanced series in Math. Phys.* vol. **2**, (1987).
- [34] D. J. Kaup. *Closure of the squared Zakharov–Shabat eigenstates.* *J. Math. Anal. Appl.* **54**, n. 3, 849–864, 1976.
- [35] D. J. Kaup. *The three-wave interaction – a non-dispersive phenomenon.* *Studies in Appl. Math.* **55**, 9–44 (1976);
D. J. Kaup, A. Reiman, A. Bers. *Rev. Mod. Phys. Space-time evolution of nonlinear three-wave interactions. I. Interaction in a homogeneous medium.* **51**, 275–310 (1979).
- [36] D. J. Kaup, A. C. Newell. *Soliton equations, singular dispersion relations and moving eigenvalues.* *Adv. Math.* **31**, 67–100, 1979.
- [37] B. G. Konopelchenko *Solitons in Multidimensions. Inverse Spectral Transform Method.* World Scientific, Singapore, 1993.
- [38] P. P. Kulish. *Quantum difference nonlinear Schrodinger equation.* *Lett. Math. Phys.* **5**, 191–197, 1981.
- [39] P. P. Kulish, A. G. Reiman *Hamiltonian structure of polynomial bundles.* *Sci. Notes. LOMI seminars* **123** 67 – 76, (1983) (In Russian); Translated in *J. Sov. Math.* **28**, 505–513 (1985);
M. A. Semenov-Tyan-Shanskii. *Classical r -matrices and the method of orbits.* *Sci. Notes.*

- LOMI seminars **123** 77 - 91, (1983) (In Russian); Translated in J. Sov. Math. **28**, 513-523 (1985).
- [40] S. V. Manakov. *An example of a completely integrable nonlinear wave field with non-trivial dynamics (Lee model)*. Teor. Mat. Phys. **28**, 172-179 (1976).
- [41] A. V. Mikhailov *The reduction problem and the inverse scattering problem*. Physica D, **3D**, n. 1/2, 73-117, 1981.
- [42] R. Miura, (editor). *Bäcklund transformations*. Lecture Notes in Math., vol. **515**, Berlin, Springer (1979).
- [43] A. G. Reyman, M. A. Semenov-Tian Shanski. *The jets algebra and nonlinear partial differential equations*. DAN USSR (Reports of the USSR Academy), **251**, No 6, p.1310-1314, (1980) (In Russian).
- [44] G. Segal, G. Wuilson. *Loop groups and equations of KdV type*. Publ. IHES, vol. **61**, 5-65 (1985).
- [45] A. B. Shabat. *The inverse scattering problem for a system of differential equations*. Functional Annal. & Appl. **9**, n.3, 75 (1975) (In Russian);
— *The inverse scattering problem*. Diff. Equations **15**, 1824 (1979) (In Russian).
- [46] E. C. Titchmarsh. *Eigenfunctions expansions associated with second order differential equations. Part I. Expansions over the eigenfunctions of $(d/dx)^2 + \lambda - q(x)$* . (Oxford, Clarendon Press, 1958).
- [47] N. P. Vekua. *Systems of singular integral equations*. Translated from Russian by A. G. Gibs and G. M. Simmons, (Gröningen, P. Noordhoff Ltd., The Netherlands, 1967).
- [48] V. E. Zakharov, S. V. Manakov, S. P. Novikov, L. I. Pitaevskii. *Theory of solitons: the inverse scattering method*. (Plenum, N.Y.: Consultants Bureau, 1984).
- [49] V. E. Zakharov, S. V. Manakov. *The theory of resonant interaction of wave packets in nonlinear media*. Sov. Phys. JETP **69**, 1654 (1975) (In Russian).
- [50] V. E. Zakharov, S. V. Manakov. *Multidimensional nonlinear integrable systems and methods for constructing their solutions*. Sci. Notes. LOMI seminars **133** 77 - 91, (1984) (In Russian); Translated in J. Sov. Math. **31**, 3307-3316 (1985).
- [51] V. E. Zakharov, A. V. Mikhailov. *On the integrability of classical spinor models in two-dimensional space-time*. Commun. Math. Phys. **74**, n. 1, 21-40, 1980;
— *Relativistically invariant two-dimensional models of field theory which are integrable by means of the inverse scattering problem method*. Zh. Eksp. Teor. Fiz. **74** 1953, (1978).
- [52] V. E. Zakharov, A. B. Shabat. *A scheme for integrating nonlinear equations of mathematical physics by the method of the inverse scattering transform. I*. Funct. Annal. and Appl. **8**, no. 3, 43-53 (1974);
— *A scheme for integrating nonlinear equations of mathematical physics by the method of the inverse scattering transform. II*. Funct. Anal. Appl. **13**(3) 13-23, (1979).
- [53] V. E. Zakharov, A. B. Shabat. *Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media*. Sov. Phys. JETP **34**, 62 (1972).
- [54] S. Zhang. *Classical Yang-Baxter equation and low-dimensional triangular Lie bialgebras*. Phys. Lett. A **246** 71-81 (1998).

INSTITUTE FOR NUCLEAR RESEARCH AND NUCLEAR ENERGY, TZARIGRADSKO CHAUSSEE 72,
1784 SOFIA, BULGARIA

Current address: Institute for Nuclear Research and Nuclear Energy, Tzarigradsko chaussee
72, 1784 Sofia, Bulgaria

E-mail address: gerdjikov@inrne.bas.bg

Differential Forms, Spectral Theory, and Boundary Value Problems

A.S. Fokas

ABSTRACT. We review a new method for studying boundary value problems for integrable PDEs in two dimensions. Examples of integrable PDEs are linear PDEs with constant coefficients and the usual nonlinear integrable PDEs such as the Korteweg-deVries equation. The starting point of the method is formulating the given PDE as the condition that an appropriate differential 1-form $W(x_1, x_2, k)$, $k \in \mathbb{C}$, is exact. The fundamental properties of an exact form W are the existence of a 0-form, and the vanishing of the integral of W around a closed contour. The spectral analysis of the associated 0-form gives rise to a Riemann-Hilbert (RH) problem with explicit exponential (x_1, x_2) dependence, while the vanishing of the integral of W around the boundary of the domain gives rise to a global relation. The RH problem and the global relation form the basis of this method. As illustrative examples, we discuss boundary value problems for: (a) an evolution equation with third order spatial derivative on the half-line; (b) the modified Helmholtz equation on a convex polygon; (c) the defocusing nonlinear Schrödinger equation on the half-line.

1. Introduction

A general approach to solving boundary value problems for *two-dimensional integrable PDEs* was announced in [1] and developed in several publications, see the review [2]. Examples of integrable PDEs are linear PDEs with constant coefficients and the usual nonlinear integrable PDEs such as the nonlinear Schrödinger equation.

This method provides a unification as well as a significant extension of the following topics: (a) The classical integral transform and Green's function methods for solving linear PDEs and several of their variants such as the Wiener-Hopf technique; (b) the integral representation of the solution of linear PDEs in terms of the Ehrenpreis fundamental principle; (c) the inverse scattering method for solving initial value problems for nonlinear integrable evolution equations. In addition it has interesting implications for: (A) The numerical solution of linear elliptic PDEs; (B) the spectral theory of linear differential operator; (C) the investigation of nonlinear

2000 *Mathematics Subject Classification.* 35G15, 35G30, 35Q15, 35Q55, 37K15.

This work was partially supported by the EPSRC.

I am deeply grateful to several of my colleagues who have contributed substantially to this new method, in particular, A.R. Its, B. Pelloni, and L.Y. Sung.

non integrable PDEs. It has already been used by more than forty researchers for the analysis of boundary value problems for linear evolution equations with spatial derivatives of arbitrary order, for linear elliptic PDEs including the Laplace, the bi-harmonic and the modified Helmholtz equations, and for several nonlinear integrable PDEs. These boundary value problems are formulated either in a polygonal domain or in a time dependent domain. Applications include fluid mechanics, acoustics, elasticity, pattern formation and statistical mechanics.

We first present a brief review of the topics (a)-(c) mentioned above, then we introduce the new method, and then discuss its relation with (a)-(c) and with (A)-(C).

A. Transform methods for linear PDEs

The proper transform for a given boundary value problem is specified by the PDE, by the domain, and by the boundary conditions. For some simple boundary value problems, there exists an algorithmic procedure for deriving the associated transform, see for example [3], [4]. This procedure is based on the analysis of either of the two ODE's obtained by separation of variables and it involves constructing the Green's function of a single eigenvalue equation and of integrating this Green's function in the complex k -plane, where k denotes the eigenvalue. An alternative procedure, based on a Riemann-Hilbert (RH) or a d-bar problem, has been recently introduced in [5].

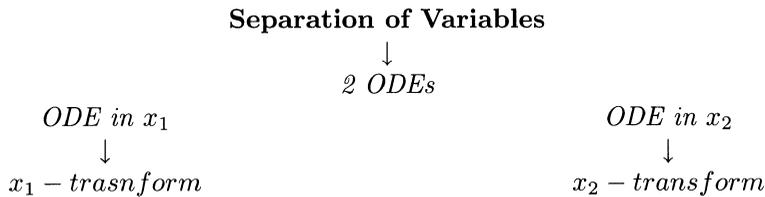


FIGURE 1.1

For an evolution equation, the proper transform for the initial value problem on the line is the Fourier transform. The proper transform for the Dirichlet problem on the half-line for an evolution equation with second order spatial derivative, is the sine transform:

EXAMPLE 1.1. Let $w(k)$ be a polynomial of order n with real coefficients.

The solution of the initial value problem

$$(1.1) \quad \begin{aligned} (\partial_t + iw(-i\partial_x))q(x, t) &= 0, \quad -\infty < x < \infty, \quad t > 0, \\ q(x, 0) &= q_0(x) \in H^{\tilde{n}}(\mathbb{R}), \end{aligned}$$

where $\tilde{n} = n/2$ for n even, $\tilde{n} = (n + 1)/2$ for n odd, is given by

$$(1.2a) \quad q(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx - iw(k)t} \hat{q}_0(k) dk,$$

$$(1.2b) \quad \hat{q}_0(k) = \int_{-\infty}^{\infty} e^{-ikx} q_0(x) dx.$$

EXAMPLE 1.2. The solution of the initial value problem

$$(1.3) \quad \begin{aligned} iqt + q_{xx} &= 0, \quad 0 < x < \infty, \quad t > 0 \\ q(x, 0) &= q_0(x) \in S(\mathbb{R}^+), \quad q(0, t) = g_0(t) \in C^1 \end{aligned}$$

is given by

$$(1.4a) \quad q(x, t) = \frac{2}{\pi} \int_0^\infty \sin(kx) e^{-ik^2 t} \left(\hat{q}_0(k) + ik \int_0^t e^{ik^2 \tau} g_0(\tau) d\tau \right) dk,$$

$$(1.4b) \quad \hat{q}_0(k) = \int_0^\infty \sin(kx) q_0(x) dx.$$

REMARK 1.3.

- Equation (1.2a) provides the spectral decomposition of $q(x, t)$. Indeed, this equation involves (x, t) in an explicit exponential form, and it also involves the spectral function $\hat{q}_0(k)$. This is to be contrasted with equation (1.4a) where the initial condition $q_0(x)$ gives rise to the spectral function $\hat{q}_0(k)$, but the boundary condition $g_0(t)$ gives rise to $\int_0^t \exp(ik^2 \tau) g_0(\tau) d\tau$ which is a function of both k and t .
- The explicit exponential dependence on (x, t) of equation (1.2a) implies that it is straightforward to compute the large t behavior of equation (1.2a). The analogous computation for equation (1.4a) is less straightforward.
- For an applied mathematician, the derivation of equations (1.4), constitutes the solution of the problem, but for an analyst it is just the first step. Indeed, the rigorous investigation of the above IBV problem involves the following: Given $q_0(x)$ define $\hat{q}_0(k)$ by equation (1.4b); given $\hat{q}_0(k)$ and $g_0(t)$, define $q(x, t)$ by equation (1.4a). Then prove that $q(x, t)$ solves equation (1.3) and that $q(x, 0) = q_0(x)$, $q(0, t) = g_0(t)$. The proof of the former equality is a direct consequence of the sine-transform but the proof of the latter equality is less straightforward since the relevant integral is *not* uniformly convergent at $x = 0$.
- The transform method has been enormously successful for solving a great variety of initial and boundary value problems. However, for sufficiently complicated problems the classical transforms fail. For example, there does *not* exist an x -transform for solving evolution equations with a third order derivative on the half line, such as

$$(1.5) \quad q_t + q_x + q_{xxx} = 0, \quad 0 < x < \infty, \quad t > 0;$$

in this case there exists a t -transform (the Laplace transform) but it involves solving a cubic algebraic equation, and also assumes boundary conditions which decay as $t \rightarrow \infty$ (otherwise one has to use certain causality arguments).

Similarly there do not exist proper transforms for solving BVP's for elliptic equations even of second order and in simple domains. The failure of transforms led to the development of several ingenious but ad-hoc techniques, which include conformal mappings for the Laplace and the bi-harmonic equations, as well as the formulation of the Wiener-Hopf factorization problem [6].

B. Euler-Ehrenpreis-Palamodov representations

There exist appropriate generalizations of the Fourier transform which are capable of capturing the general solution of linear PDEs in a smooth convex domain [7], [8]. An elementary implication of this general result is that there exists a measure $d\rho(k)$ such that the solution of any IBV problem of equation (1.1) on the half

line is given by

$$(1.6) \quad q(x, t) = \int_L e^{ikx - iw(k)t} d\rho(k),$$

where L is an appropriate curve in the complex k -plane. However, in general it is not clear how to compute $d\rho(k)$, although recently some progress has been made for some particular types of domains [9], [10].

As a verification of the above result we note that it is possible to rewrite equation (1.4a) in the form (1.6) where $d\rho$ is supported on the real axis and on the positive imaginary axis.

C. The Inverse Scattering Transform

There exist nonlinear evolution equations, whose initial value problem on the line can be solved by a certain nonlinear Fourier transform called the inverse scattering transform [11], [12].

EXAMPLE 1.4. (The defocusing NLS)

$$(1.7) \quad \begin{aligned} iq_t + q_{xx} - 2|q|^2q &= 0, \quad -\infty < x < \infty, \quad t > 0 \\ q(x, 0) &= q_0(x) \in S(\mathbb{R}). \end{aligned}$$

This initial value problem can be solved by the inverse scattering transform pair. However, this pair, in contrast to the Fourier transform pair, cannot be written in terms of explicit integral representations. Instead, the map

$$q_0(x) \rightarrow \hat{q}_0(k)$$

is defined via the solution of a *linear Volterra integral equation*, and the inverse map,

$$\hat{q}_0(k) \rightarrow q_0(x)$$

is defined via the solution of a *matrix RH problem* [13].

Recall that the separation of variables of a linear evolution equation gives rise to two ODE's. The spectral analysis of the x -ODE gives rise to the Fourier transform. The distinctive property of an integrable nonlinear PDE is that it can also be associated with two ODE's (called the Lax pair [14]); the spectral analysis of the x -ODE (i.e. of the x -part of the Lax pair) gives rise to the inverse scattering transform pair.

The evolution of $\hat{q}(k, t)$ is determined by the nonlinear PDE itself, or equivalently by the t -part of the Lax pair.

REMARK 1.5.

Consider the linearized version of equation (1.7), i.e. equation (1.3). The two ODEs obtained by separation of variables of this equation are

$$(1.8) \quad \frac{\partial^2 X(x, \lambda)}{\partial x^2} - \lambda X(x, \lambda) = 0, \quad \frac{\partial T(t, \lambda)}{\partial t} - i\lambda T(t, \lambda) = 0.$$

It turns out that equation (1.3) can also be associated with two other ODEs which, in analogy with the nonlinear PDE (1.7), we call a Lax pair [15]

$$(1.9) \quad \frac{\partial \mu}{\partial x} - ik\mu = q, \quad \frac{\partial \mu}{\partial t} + ik^2\mu = iq_x - kq.$$

There exist two different ways of obtaining the Fourier transform. The classical one involves analyzing equation (1.8a), while the alternative one involves analyzing equation (1.9a) [15]. We emphasize that it is only the latter analysis that can be generalized to the nonlinear equation (1.7). This suggests that perhaps equations

(1.9) are more fundamental than equations (1.8). It will be shown in this paper that this is indeed the case. In this relation we note that, just like the Fourier transform, the sine transform can also be obtained by analyzing either of equations (1.8a) or (1.9a). However, if one wants to obtain the Ehrenpreis form of the solution of the IBV of Example 1.2, then one must analyze *both equations (1.9) simultaneously*. This suggests another reason why equations (1.9) are more fundamental than equations (1.8): They provide the appropriate framework for performing a novel type of analysis, namely the simultaneous spectral analysis of compatible linear equations. It turns out that differential forms provide a most convenient formalism for this purpose.

D. A New Method

DEFINITION 1.6. An equation in two dimensions (x_1, x_2) is called integrable if and only if it is equivalent to the condition that an appropriate differential 1-form $W(x_1, x_2, k)$ is closed, where $k \in \mathbb{C}$.

For linear PDEs with constant coefficients, W can be found algorithmically, see Appendix A. For nonlinear integrable PDEs the existence of W is a direct consequence of the Lax pair.

EXAMPLE 1.7. Equation (1.5) is associated with the closed form W ,

$$(1.10) \quad W = e^{-ikx+iw(k)t} \{qdx - [q_{xx} + ikq_x + (1 - k^2)q]dt\}, \quad w(k) = k - k^3.$$

Indeed, if $e = \exp[-ikx + iw(k)t]$ then

$$\begin{aligned} dW &= (eq)_t dt \wedge dx - \{e[q_{xx} + ikq_x + (1 - k^2)q]\}_x dx \wedge dt \\ &= ((eq)_t + \{e[q_{xx} + ikq_x + (1 - k^2)q]\}_x) dt \wedge dx \\ &= e[q_t + q_{xxx} + q_x]dt \wedge dx. \end{aligned}$$

Thus $dW = 0$ iff q satisfies equation (1.5).

The fundamental properties of an exact differential 1-form W are the existence of a 0-form $\mu^{(0)}(x_1, x_2, k)$, and the vanishing of the integral of W around a closed contour. These two properties form the basis of the new method. Indeed:

(1). The spectral analysis of the associated 0-form yields the solution $q(x_1, x_2)$ in terms of the solution of either a *Riemann-Hilbert problem* or a *d-bar problem*. These problems are formulated in the complex k -plane and are determined in terms of a certain function of k called the *spectral function* and denoted by $\hat{q}(k)$. This function in turn is defined in terms of the boundary values of $q(x_1, x_2)$ and of its derivatives. Since for a well posed boundary value problem only some of the boundary values are prescribed as boundary conditions, part of $\hat{q}(k)$ is unknown.

(2). The vanishing of the integral of W around the boundary of the given domain gives rise to a simple *global algebraic relation* satisfied by the spectral function. The analysis of this relation determines the unknown part of the spectral function in terms of the given boundary conditions. For linear PDEs, the relevant Riemann-Hilbert and d-bar problems can be solved in closed form, thus step (1) yields an *explicit integral representation* of $q(x_1, x_2)$ in terms of the spectral function. For nonlinear integrable PDEs the investigation of the solvability of these problems must be carried out for each equation separately.

The constructions (1) and (2) are summarized in Figure 1.2.

The most difficult step of the methodology outlined above is the analysis of the global relation. It turns out that for linear evolution equations in $\{0 < x < \infty, t >$

$0\}$ [16] or in $\{0 < x < 1, t > 0\}$ [17], this step involves only algebraic manipulations; for linear evolution equations in the time-dependent domain $\{l(t) < x < \infty, t > 0\}$ [18], it involves solving a system of linear Volterra integral equations; for linear elliptic equations in a polygonal domain, it involves either algebraic manipulations, or solving an auxiliary matrix Riemann-Hilbert problem [19]–[23]; for nonlinear integrable evolution equations in $\{0 < x < \infty, t > 0\}$ [24]–[25] or $\{0 < x < 1, t > 0\}$ [26], it involves solving a system of nonlinear Volterra integral equations.

We now discuss the relation of the new method with the three topics (a)–(c) reviewed earlier.

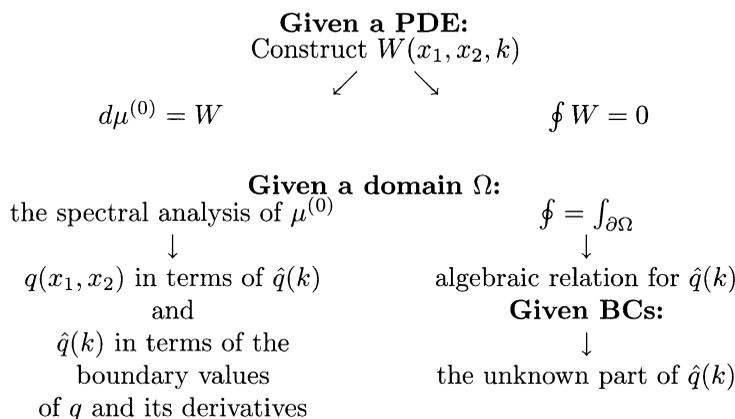


FIGURE 1.2

(a). Suppose that $q(x_1, x_2)$ satisfies a linear PDE. The existence of the differential 1-form $W(x_1, x_2, k)$ is equivalent to the existence of a Lax pair. Performing the spectral analysis of the x_1 -part of the Lax pair corresponds to constructing an x_1 -transform, similarly performing the spectral analysis of the x_2 -part corresponds to constructing a x_2 -transform. The advantage of the 1-form W is that it provides the tool for performing the *simultaneous* spectral analysis. This gives rise to a new transform, which in contrast to both the x_1 and x_2 -transforms is “custom made” for the given PDE and the given domain. In this sense the new method provides the synthesis of separation of variables.

Suppose that $q(x_1, x_2)$ satisfies a linear PDE in a convex polygon. In this case, step (1) yields for $q(x_1, x_2)$ an integral representation in the complex k plane, which has an explicit x_1 and x_2 dependence in the form of an exponential and which involves the spectral function $\hat{q}(k)$, $k \in \mathbb{C}$. This function can be computed by analyzing the global equation. For evolution equations and for elliptic equations with simple boundary conditions, this involves the solution of a system of algebraic equations, while for elliptic equations with arbitrary boundary conditions, it involves the solution of an *auxiliary Riemann-Hilbert problem*. For simple polygons, this Riemann-Hilbert problem is formulated on the infinite line, thus it is equivalent to a Wiener-Hopf problem. This explains the central role played by the Wiener-Hopf technique in many earlier works.

(b). For linear equations in a convex domain, the explicit x_1, x_2 dependence of $q(x_1, x_2)$ is consistent with the Ehrenpreis formulation of the solution. Thus

this method provides the concrete implementation as well as the generalization to concave domains of this fundamental principle.

(c) An important advantage of the new method is that it can be nonlinearized. Indeed, the results for linear PDEs obtained by this method can be generalized to integrable nonlinear PDEs. In the nonlinear case, the relevant RH and d-bar problems *cannot* be solved in closed form. However, their (x_1, x_2) dependence is of the same exponential form as the one occurring in the associated linear equations. In this sense, the new method provides the nonlinearization of the Ehrenpreis principle. For the Cauchy problem, the solution representation obtained by this method coincides with the one obtained by the inverse scattering transform.

Regarding (A)-(C) we note:

(A) *A numerical method*

The new method has led to the formulation of a new numerical scheme for solving elliptic PDEs. This is based on the numerical solution of the global relation [27]–[28].

(B) *Spectral Theory*

The new method has motivated the study of certain classes of linear differential operators which in general are non-self-adjoint. In these studies the RH and the d-bar problems play a crucial role [5].

(C) *Nonlinear non-integrable PDEs*

It should be emphasized that although this method is directly applicable only to integrable PDEs, it nevertheless has important implications for non integrable PDEs. Indeed, by formulating such equations as “forced” linear PDEs and by combining the new method with standard PDE techniques, it is possible to prove the well posedness of boundary value problems for a large class of nonlinear PDEs [29].

This paper is organized as follows: In §2 we solve an initial-boundary value problem for equation (1.5). In §3 we first present the formulae for $q(x_1, x_2)$ and for $\hat{q}(k)$ associated with the modified Helmholtz equation in an arbitrary convex polygon, and then discuss boundary value problems for the semi-strip. In §4 we discuss the defocusing nonlinear Schrödinger equation on the half-line.

2. Linear Evolution PDEs on the Half-Line

We will illustrate the three steps needed for the rigorous implementation of the method by using equation (1.5). This equation is a particular case of equation (1.1), where

$$(2.1) \quad w(k) = k - k^3. \text{label2.1}$$

Equation (1.1) is analyzed in [16].

THEOREM 2.1. *Let $q(x, t)$ satisfy*

$$q_t + q_x + q_{xxx} = 0 \quad 0 < x < \infty, \quad 0 < t < T$$

$$q(x, 0) = q_0(x) \in H^2(\mathbb{R}^+), \quad q(0, t) = g_0(t) \in H^1([0, T]), \quad q_0(0) = g_0(0)$$

where T is a positive constant. The unique solution of this IBV problem is given by

$$(2.2) \quad q(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx - iw(k)t} \hat{q}_0(k) dk + \frac{1}{2\pi} \int_{\partial D_+} e^{ikx - iw(k)t} \hat{g}(k) dk,$$

where the curve ∂D_+ and the spectral function $\hat{q}(k) = \{\hat{q}_0(k), \hat{g}(k)\}$, are defined as follows:

$$(2.3) \quad \partial D_+ : \operatorname{Im} w(k) = 0, \quad \operatorname{Im} k > 0.$$

$$(2.4) \quad \hat{q}_0(k) = \int_0^\infty e^{-ikx} q_0(x) dx, \quad \operatorname{Im} k \geq 0$$

$$(2.5) \quad \hat{g}(k) = (1 - 3k^2)\hat{g}_0(k) + \frac{\nu_1 - k}{\nu_2 - \nu_1}\hat{q}_0(\nu_2) + \frac{k - \nu_2}{\nu_2 - \nu_1}\hat{q}_0(\nu_1),$$

$$(2.6) \quad \hat{g}_0(k) = \int_0^T e^{iw(k)t} g_0(t) dt$$

$\nu_1(k), \nu_2(k)$ are the two nontrivial roots of $w(k) = w(\nu(k))$.

The rigorous investigation of the above IBV problem involves the following steps, see [16] for details.

Step 1 Assuming existence: (a) construct the integral representations for $q(x, t)$ and for $\hat{q}(k)$; (b) find the global relation.

(a) For linear equations there exist several ways of obtaining the relevant representations. The simplest one is to use the Fourier transform and contour deformation, see [30] and Appendix B. Among these different approaches, the only one which nonlinearizes is the one based on the spectral analysis of the equation

$$(2.7) \quad d \left(e^{-ikx + iw(k)t} \mu(x, t, k) \right) = W(x, t, k).$$

This approach will be illustrated in §4 as an introduction to the analysis of the NLS. All of these approaches imply that $q(x, t)$ is given by equation (2.2) where $\hat{q}_0(k)$ is defined by equation (2.4), while $\hat{g}(k)$ is defined by

$$(2.8) \quad \hat{g}(k) = (1 - k^2)\hat{g}_0(k) + ik\hat{g}_1(k) + \hat{g}_2(k),$$

$$(2.9) \quad \hat{g}_j(k) = \int_0^T e^{iw(k)t} g_j(t) dt, \quad j = 0, 1, 2, \quad k \in \mathbb{C},$$

and

$$(2.10) \quad g_j(t) = \partial_x^j q(0, t), \quad j = 0, 1, 2.$$

(b) The equation

$$\oint_{\partial\Omega} W(x, t, k) = 0,$$

where $\partial\Omega$ is the boundary of the domain $\{0 < x < \infty, 0 < t < T\}$ yields, see Figure 2.1,

$$(2.11) \quad \hat{q}_0(k) + \hat{g}(k) = e^{iw(k)T} c(k),$$

where

$$c(k) = \int_0^\infty e^{-ikx} q(x, T) dx, \quad \operatorname{Im} k \leq 0.$$

Step 2. Assuming the validity of the global relation, prove existence: Namely assume that there exist functions $q_0(x), \{g_j(t)\}_0^2$, such that the functions $\hat{q}(k)$ and $\hat{g}(k)$ defined by equations (2.4), (2.8), (2.9), satisfy equation (2.11), where $c(k)$ is some function holomorphic for $\operatorname{Im} k < 0$ and of $O(1/k)$ as $k \rightarrow \infty$. Then prove

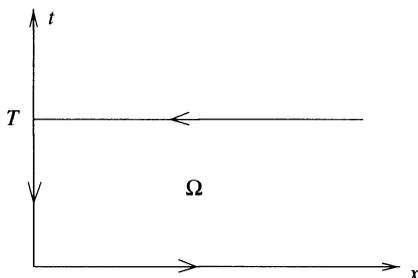


FIGURE 2.1

that if $q(x, t)$ is defined by equation (2.2), (a) $q(x, t)$ solves equation (1.5); (b) $q(x, 0) = q_0(x)$; $\partial_x^j q(0, t) = g_j(t)$, $j = 0, 1, 2$.

The proof of (a) is a direct consequence of the exponential dependence of (x, t) . The proof of (b) follows from the fact that $\exp(-iw(k)t)\hat{g}(k)$ is analytic and bounded in D_+ ,

$$D_+ = \{k \in \mathbb{C}, \quad \text{Im } w(k) > 0, \quad \text{Im } k > 0\}.$$

The proof of (c) is based on the global relation and on appropriate contour deformations.

Step 3. *Given boundary conditions, analyze the global relation*

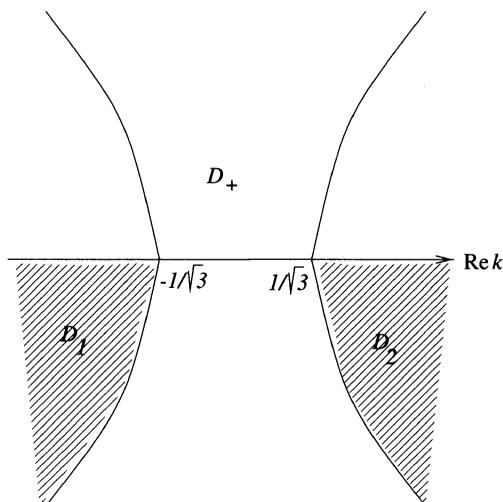


FIGURE 2.2

Using the definition of $\hat{g}(k)$, the global relation (2.11) becomes

$$(2.12) \quad \hat{q}_0(k) + (1 - k^2)\hat{g}_0(k) + ik\hat{g}_1(k) + \hat{g}_2(k) = e^{iw(k)T}c(k), \text{Im } k \leq 0.$$

The crucial observation is that $\hat{g}_j(k)$, $j = 0, 1, 2$, depend on k only through $w(k)$. Thus these functions are invariant if $k \rightarrow \nu(k)$, where $\nu(k)$ is defined by

$$w(k) = w(\nu(k)).$$

This equation has two nontrivial roots: If $\nu_1(k) \in D_1$ then $k \in D_+$, and if $\nu_2(k) \in D_2$, then $k \in D_+$. Thus evaluating equation (2.12) at $\nu_1(k)$ and $\nu_2(k)$ we find

$$\begin{aligned} \hat{q}_0(\nu_j(k)) + (1 - \nu_j^2(k))\hat{g}_0(k) + i\nu_j(k)\hat{g}_1(k) + \hat{g}_2(k) &= \\ &= e^{iw(k)T}c(\nu_j(k)), \quad j = 1, 2, \quad k \in D_+. \end{aligned}$$

Solving these two equations for $\hat{g}_1(k)$, $\hat{g}_2(k)$ and substituting the resulting expressions in equation (2.8) we find that $\hat{g}(k)$ is given by equation (2.5) plus an additional term involving $e^{iw(k)T}$ multiplied by a certain combination of $c(\nu_1(k))$ and $c(\nu_2(k))$. However, this additional term does *not* contribute to $q(x, t)$. Indeed, $\exp[ikx + iw(k)(T - t)]$, as well as $c(\nu_j(k))$, are bounded and analytic for $k \in D_+$, thus Cauchy’s theorem implies that this additional term vanishes.

REMARK 2.2. Let $\hat{g}(k, t)$ be defined by equation (2.5) where $\hat{g}_0(k)$ is replaced by

$$\hat{g}_0(k, t) = \int_0^t e^{iw(k)\tau}g_0(\tau)d\tau.$$

It is straightforward to show that $q(x, t)$ is also given by an expression similar to the rhs of equation (2.2) where $\hat{g}(k)$ is replaced by $\hat{g}(k, t)$. This is consistent with causality.

REMARK 2.3. The representation (2.2) is very convenient for computing the asymptotic properties of $q(x, t)$. These include the long time asymptotics [31] as well as the small dispersion limit [32].

3. Linear Elliptic PDEs

It is shown in appendix A that the equation

$$(3.1) \quad q_{z\bar{z}} + \alpha q = 0, \quad z = x + iy, \quad \alpha \text{ constant}$$

is associated with the closed differential 1-form

$$(3.2) \quad W(z, \bar{z}, k) = e^{-ikz - \frac{i\alpha}{k}\bar{z}} \left(q_z dz - \frac{i\alpha}{k} q d\bar{z} \right).$$

If equation (3.1) is valid in a convex polygon Ω , it is straightforward using equation (3.2) to construct the integral representations for $q(z, \bar{z})$ and for $\hat{q}(k)$ [33]. If α is positive, the relevant contour in the complex k -plane consists of a union of rays and of circular arches; if α is negative the contour involves only rays. For simplicity we consider the latter case.

PROPOSITION 3.1. *Let $q(z, \bar{z})$ satisfy*

$$(3.3) \quad q_{z\bar{z}} - \alpha^2 q = 0, \quad z \in \Omega$$

Ω is a bounded convex polygon specified by z_1, z_2, \dots, z_n : on side (j):

$$(3.4) \quad \cos \beta_j q_s + \sin \beta_j q_N + \gamma_j q = g_j,$$

where $\{\beta_j, \gamma_j\}_1^n$ are constants, $\{g_j\}_1^n$ are smooth functions, and q_s, q_N denote the tangential, normal derivatives. Assume that there exists a smooth global solution. Then $q(z, \bar{z})$ is given by

$$(3.5) \quad q(z, \bar{z}) = \frac{1}{2\pi i} \sum_{j=1}^n \int_{l_j} e^{ikz - \frac{i\alpha^2}{k}\bar{z}} \hat{g}_j(k) \frac{dk}{k}, \quad z \in \Omega$$

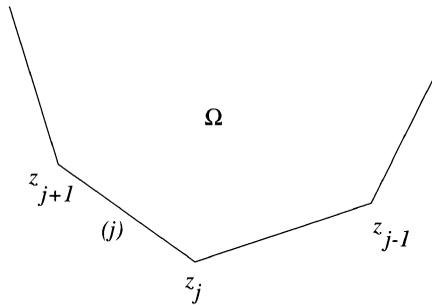


FIGURE 3.1

where the contours $\{l_j\}_1^n$ and the spectral function $\hat{g}(k) = \{\hat{g}_j\}_1^n$, are defined as follows: l_j is the ray from 0 to ∞ making an angle $-\arg(z_j - z_{j+1})$ with the real k -axis;

$$(3.6) \quad \hat{g}_j(k) = \int_{z_{j+1}}^{z_j} e^{-ikz + \frac{i\alpha^2}{k} \bar{z}} \left(q_z dz + \frac{i\alpha^2}{k} q d\bar{z} \right), \quad k \in \mathbb{C}.$$

Furthermore, the spectral function satisfies the global relation

$$(3.7) \quad \sum_{j=1}^n \hat{g}_j(k) = 0, \quad k \in \mathbb{C}.$$

The derivation of equations (3.5), (3.6), using the spectral analysis of the equation

$$(3.8a) \quad d \left(e^{-ikz + \frac{i\alpha^2}{k} \bar{z}} \mu \right) = W$$

is given in [33]. Equation (3.7) is a direct consequence of

$$(3.8b) \quad \oint_{\partial\Omega} W = 0;$$

an alternative derivation of equations (3.5), (3.6) is given in [34] using the fundamental differential form, which is a slight generalization of W .

REMARK 3.2. If Ω is an unbounded polygon with $z_1 = z_n = \infty$, then the summation in equations (3.5), (3.7) are only up to $n - 1$; also equation (3.7) is not valid for all k but only in a certain domain of the complex k -plane [33].

REMARK 3.3. The above proposition is step 1 of the new method. Step 2 is also valid, namely it is possible to show that the global relation (3.7) is not only a necessary but also a sufficient condition for existence [27]. However, step 3, namely the analysis of the global relation, is now more complicated: In general it involves the formulation of an auxiliary matrix RH problem; for some simple polygons and for simple boundary conditions this RH problem can either be solved in closed form or can be bypassed and $\hat{g}(k)$ can be computed via the algebraic manipulation of the global relation.

EXAMPLE 3.4.

Let $q(z, \bar{z})$ satisfy equation (3.3) in the semistrip $\{0 < x < \infty, 0 < y < l\}$ depicted in Figure 3.2 with the boundary conditions (3.4) on each side. It is shown

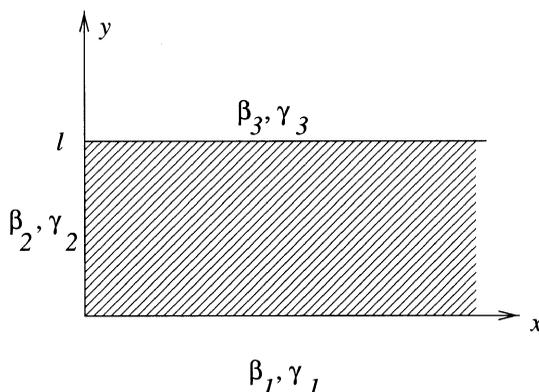


FIGURE 3.2

in [22] that for arbitrary values of β_j, γ_j , the spectral function can be computed via a 2×2 matrix RH problem. However, if either of the conditions

$$(3.9a) \quad (2\alpha^2 - \gamma_2^2) \sin 2\beta_1 = (\gamma_1^2 - 2\alpha^2) \sin 2\beta_2,$$

or

$$(3.9b) \quad (2\alpha^2 - \gamma_3^2) \sin 2\beta_2 = (\gamma_2^2 - 2\alpha^2) \sin 2\beta_3$$

are valid, then the RH problem becomes triangular and can be solved in closed form. If both equations (3.9) are valid then the part of $\hat{g}(k)$ contributing to $q(z, \bar{z})$ can be found via the algebraic manipulation of the global relation.

The derivation of the above results can be found in [22]. Here we only note that just as in the case of linear evolution equations, the *invariant properties* of the global relation play a crucial role: Using the definition of \hat{g}_j (equation (3.6)), the boundary conditions (equation (3.4)), and integration by parts, it follows that each \hat{g}_j involves an unknown function. Using that on the sides (1), (2), (3), z is given by $x, iy, x + il$, it follows that these unknown functions are given by

$$\begin{aligned} \Psi_1(-ik) &= \int_0^\infty e^{-ikx + \frac{i\alpha^2}{k}x} q_1(x) dx, & \Psi_2(k) &= \int_0^l e^{ky + \frac{\alpha^2}{k}y} q_2(y) dy, \\ \Psi_3(-ik) &= \int_0^\infty e^{-ikx + \frac{i\alpha^2}{k}x} q_3(x) dx, \end{aligned}$$

where $q_1(x) = q(x, 0)$, $q_2(y) = q(0, y)$, $q_3(x) = q(x, l)$. Thus the global relation (3.7) becomes a relation with known coefficients among

$$(3.10a) \quad \Psi_1(-ik), \quad \Psi_2(k), \quad \Psi_3(-ik), \quad \text{Im}k \leq 0.$$

The complex conjugate and the transformation $k \rightarrow \bar{k}$ of this relation yields a relation among

$$(3.10b) \quad \Psi_1(ik), \quad \Psi_2(k), \quad \Psi_3(ik), \quad \text{Im}k \geq 0.$$

Equations (3.10) involve the unknown vector functions

$$\{\Psi_1(-ik), \Psi_1(ik)\}, \quad \{\Psi_2(k), \Psi_2(k)\}, \quad \{\Psi_3(-ik), \Psi_3(ik)\}.$$

The first and the third unknown vectors are invariant under the transformation $k \rightarrow -k$. Thus we supplement equations (3.10), with the equations obtained from

(3.10) by the substitution $k \rightarrow -k$. We denote these equations by (3.10)'. The 4 equations (3.10) and (3.10)' are the basic equations needed for the determination of the unknown functions Ψ_j .

Both equations (3.10) are valid for $k \in \mathbb{R}$; eliminating $\Psi_2(k)$ from these two equations we obtain a relationship between $\Psi_1(\pm ik)$ and $\Psi_3(\pm ik)$. Using the substitution $k \rightarrow -k$ in this equation (or equivalently eliminating $\Psi_2(-k)$ from equations (3.10)') we obtain a second relation between $\Psi_1(\pm ik)$ and $\Psi_3(\pm ik)$. These two relations together with the fact that $\{\Psi_j(ik), \Psi_j(-ik)\}$, $j = 1, 3$ are sectionally holomorphic functions with a cut along the real axis, and of $O(1/k)$ as $k \rightarrow \infty$, define a 2×2 matrix RH problem.

4. Integrable Nonlinear Evolution Equations on the Half-Line

The rigorous implementation of the new method to the nonlinear Schrödinger (NLS) equation on the half line is presented in [24]. The Korteweg-deVries (KdV) equation with dominant surface tension, and the sine-Gordon (sG) equation in laboratory coordinates can be treated similarly [25]. In what follows we discuss the three steps (analogues to the three steps presented in §2) needed for the analysis of the defocusing NLS equation on the half line:

$$(4.1) \quad iqt + q_{xx} - 2|q|^2q = 0, \quad 0 < x < \infty, \quad 0 < t < T$$

$$(4.2) \quad q(x, 0) = q_0(x) \in S(\mathbb{R}^+), \quad q(0, t) = g_0(t) \in C^1(0, T), \quad q_0(0) = g_0(0),$$

where T is a given positive constant.

Step 1. *Assuming existence: (a) Construct the integral representations of $q(x, t)$ and of the spectral function $\hat{q}(k)$; the former involves the formulation of a RH problem and the latter involves the solution of certain linear Volterra integral equations. (b) Derive the global relation satisfied by $\hat{q}(k)$.*

If A is a 2×2 matrix, define $\hat{\sigma}_3 A$ by $[\sigma_3, A]$, $\sigma_3 = \text{diag}(1, -1)$; then it follows that

$$e^{\hat{\sigma}_3 A} = e^{\sigma_3 A} e^{-\sigma_3}.$$

Step 1 is based on the fact that the defocusing NLS equation (4.1) is equivalent to

$$(4.3) \quad d \left[e^{(ikx + 2ik^2t)\hat{\sigma}_3} \mu(x, t, k) \right] = W(x, t, k), \quad k \in \mathbb{C},$$

where μ is a 2×2 matrix, and the differential 1form W is defined by

$$(4.4) \quad W = e^{(ikx + 2ik^2t)\hat{\sigma}_3} \left(Q(x, t)\mu(x, t, k)dx + \tilde{Q}(x, t, k)\mu(x, t, k)dt \right),$$

$$(4.5a) \quad Q(x, t) = \begin{pmatrix} 0 & q(x, t) \\ \bar{q}(x, t) & 0 \end{pmatrix},$$

$$(4.5b) \quad \tilde{Q}(x, t, k) = 2kQ - iQ_x\sigma_3 - i|q|^2\sigma_3.$$

The derivation of (a) involves the spectral analysis of equation (4.3). For pedagogical reasons we first consider the spectral analysis of the equation

$$(4.6a) \quad d \left(e^{ikx + ik^2t} \mu(x, t, k) \right) = W(x, t, k), \quad k \in \mathbb{C},$$

$$(4.6b) \quad W(x, t, k) = e^{ikx + ik^2t} (qdx + (iq_x + kq)dt),$$

which corresponds to the linearized version of equation (4.1), i.e. to equation

$$iq_t + q_{xx} = 0.$$

Performing the spectral analysis of equation (4.6a) means: *Construct a function $\mu(x, t, k)$ which for $(x, t) \in \Omega = \{0 < x < \infty, 0 < t < T\}$, is bounded in k for all $k \in \mathbb{C}$.*

We claim that such a μ is given by

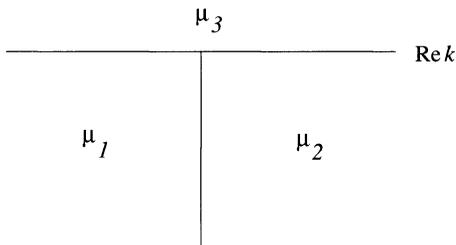


FIGURE 4.1

$$(4.7) \quad \mu = \begin{cases} \mu_1 & \pi \leq \arg k \leq \frac{3\pi}{2} \\ \mu_2 & \frac{3\pi}{2} \leq \arg k \leq 2\pi \\ \mu_3, & \text{Im } k \geq 0 \end{cases}$$

Indeed, equation (4.6a) implies

$$(4.8) \quad \mu_*(x, t, k) = e^{-ikx - ik^2t} \int_{(x_*, t_*)}^{(x, t)} W(\xi, \tau, k),$$

where (x_*, t_*) is any point in the domain Ω .

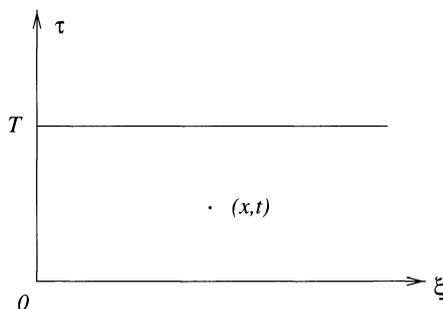


FIGURE 4.2

The properties of μ with respect to k depend on the particular choice of (x_*, t_*) . It was shown in [33] that if Ω is a polygon there exists a canonical way of choosing (x_*, t_*) , namely the corners of the polygon Ω which we denote by (x_j, t_j) . The collection of the corresponding functions μ_j define a function $\mu = \{\mu_j\}_1^n$ which is

sectionally holomorphic in the complex k -plane. In this particular example, there exist three corners,

$$(x_1, t_1) = (0, T), \quad (x_2, t_2) = (0, 0), \quad (x_3, t_3) = (\infty, t).$$

Thus we define μ_1, μ_2, μ_3 , and we also choose the contours shown in Figure 4.3

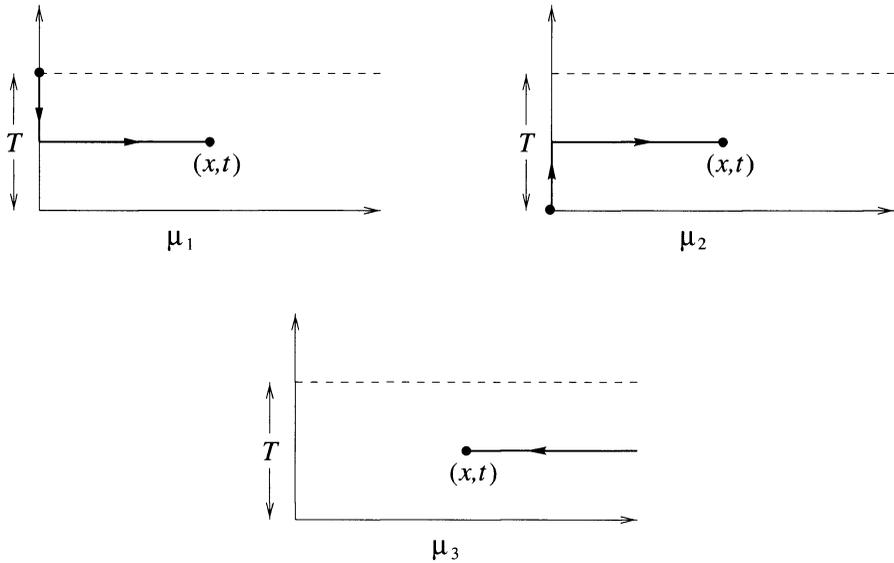


FIGURE 4.3

Equation (4.8) involves the exponential $\exp[ik(\xi - x) + ik^2(\tau - t)]$. In the case of μ_1 ,

$$\xi - x \leq 0 \quad \text{and} \quad \tau - t \geq 0;$$

thus the exponential is bounded in the intersection of $\arg k \leq 0$ and of $\arg k \in [0, \pi/2] \cup [\pi, 3\pi/2]$, i.e. in $\pi \leq \arg k \leq 3\pi/2$. Thus μ_1 is bounded and analytic for $\pi < \arg k < 3\pi/2$. Similarly for μ_2, μ_3 .

These considerations together with the estimate $\mu = O(1/k), k \rightarrow \infty$ (which follows from equation (4.8)), imply that μ is a sectionally holomorphic function in $\mathbb{C} \setminus \{\mathbb{R} \cup i\mathbb{R}^-\}$. Equation (4.8) implies that the “jumps” of μ are of the “Ehrenpreis form” $\exp[-ikx - ik^2t]\rho(k)$. For example

$$\begin{aligned} \mu_2(x, t, k) - \mu_1(x, t, k) &= e^{-ikx - ik^2t} \hat{g}(k), \\ \hat{g}(k) &= \int_0^T e^{ik^2\tau} (iq_x(0, \tau) + kq(0, \tau)) d\tau. \end{aligned}$$

Similarly

$$\begin{aligned} \mu_2(x, t, k) - \mu_3(x, t, k) &= e^{-ikx - ik^2t} \hat{q}_0(k), \\ \hat{q}_0(k) &= \int_0^\infty e^{ik\xi} q(\xi, 0) d\xi. \end{aligned}$$

Thus μ can be expressed in terms of $\{\hat{q}_0(k), \hat{g}(k)\}$:

$$(4.9) \quad \mu = -\frac{1}{2i\pi} \int_{-\infty}^\infty e^{-ilx - il^2t} \frac{\hat{q}_0(l)}{l - k} dl - \frac{1}{2i\pi} \int_L e^{-ilx - il^2t} \frac{\hat{g}(l)}{l - k} dl,$$

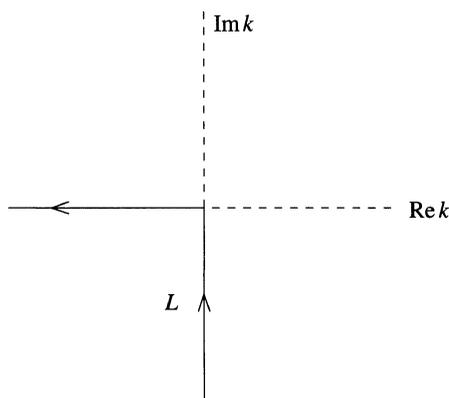


FIGURE 4.4

where L is depicted in Figure 4.4. Equation (4.9) together with

$$q = \mu_x + ik\mu,$$

yield

$$(4.10) \quad q = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikx - ik^2t} \hat{q}_0(k) dk + \frac{1}{2\pi} \int_L e^{-ikx - ik^2t} \hat{g}(k) dk.$$

We note that

$$(4.11) \quad \hat{q}_0(k) = -\mu_3(0, 0, k), \quad \hat{g}(k) = e^{ik^2T} \mu_2(0, T, k).$$

We now return to the NLS equation. In analogy with equation (4.8) we now have

$$(4.12) \quad \mu_*(x, t, k) = I + e^{-(ikx + 2ik^2t)\sigma_3} \int_{(x_*, t_*)}^{(x, t)} W(\xi, \tau, k),$$

where I is the 2×2 identity matrix. Again associated with the three corners we define μ_1, μ_2, μ_3 . These matrices are simply related by the matrix analogues of $\hat{q}_0(k)$ and of $\hat{g}(k)$. Due to certain symmetries these matrices have the form

$$(4.13) \quad \hat{q}_0(k) = \begin{pmatrix} \overline{a(\bar{k})} & b(k) \\ \overline{b(\bar{k})} & a(k) \end{pmatrix}, \quad \hat{g}(k) = \begin{pmatrix} \overline{A(\bar{k})} & B(k) \\ \overline{B(\bar{k})} & A(k) \end{pmatrix}.$$

The matrices $\mu_3(x, 0, k)$ and $\mu_2(0, t, k)$ satisfy linear integral equations, thus the functions $\{a(k), b(k), A(k), B(k)\}$ cannot be written in closed form. Similarly, since μ is a 2×2 matrix, the associated RH problem is not a scalar RH problem, thus it cannot be solved in closed form.

Using $\int_{\partial\Omega} W = 0$, with $\mu = \mu_3$ in the definition of W , it is straightforward to derive the global relation satisfied by the spectral function.

Step 2. *Existence under the assumption that the spectral functions satisfy the global relation.*

Given $q_0(x) \in \mathcal{S}(\mathbb{R}^+)$, the space of Schwartz functions on the positive real axis, define $\{a(k), b(k)\}$. Assume that there exist smooth functions $g_0(t)$ and $g_1(t)$ such that if $\{A(k), B(k)\}$ are defined in terms of them, then $\{a(k), b(k), A(k), B(k)\}$ satisfy the global relation. Define $q(x, t)$ through the solution of the RH problem

formulated in Step 1. Then prove that: (a) $q(x, t)$ is defined for all $0 < x < \infty$, $t > 0$; (b) $q(x, t)$ solves the NLS; (c) $q(x, 0) = q_0(x)$, $0 < x < \infty$ and $q(0, t) = g_0(t)$, $q_x(0, t) = g_1(t)$, $t > 0$.

We give the definitions of $\{a(k), b(k), A(k), B(k)\}$ and the main theorem.

Definition of $a(k), b(k)$. Let $q_0(x) \in \mathcal{S}(\mathbb{R}^+)$. The map

$$(4.14) \quad \mathbf{S} : \{q_0(k)\} \rightarrow \{a(k), b(k)\}$$

is defined as follows:

$$(4.15) \quad \begin{pmatrix} b(k) \\ a(k) \end{pmatrix} = \varphi(0, k),$$

where the vector-valued function $\varphi(x, k)$ is defined in terms of $q_0(x)$ by

$$(4.16) \quad \begin{aligned} \partial_x \varphi(x, k) + 2ik \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \varphi(x, k) &= \begin{pmatrix} 0 & q_0(x) \\ \bar{q}_0(x) & 0 \end{pmatrix} \varphi(x, k), \\ 0 < x < \infty, \quad \text{Im} k \geq 0, \quad \lim_{x \rightarrow \infty} \varphi(x, k) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned}$$

Definition of $A(k), B(k)$. Let $\{g_0(t), g_1(t)\}$ be smooth functions for $0 < t < T$. The map

$$(4.17) \quad \tilde{\mathbf{S}} : \{g_0(t), g_1(t)\} \rightarrow \{A(k), B(k)\}$$

is defined as follows

$$(4.18) \quad \begin{pmatrix} -e^{-4ik^2 T} B(k) \\ A(\bar{k}) \end{pmatrix} = \Phi(T, k),$$

where the vector-valued function $\Phi(t, k)$ is defined by

$$(4.19) \quad \begin{aligned} \partial_t \Phi(t, k) + 4ik^2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \Phi(t, k) &= \tilde{Q}(t, k) \Phi(t, k), \quad 0 < t < T, \quad k \in \mathbb{C}, \\ \Phi(0, k) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \end{aligned}$$

and $\tilde{Q}(t, k)$ is given by:

$$\tilde{Q}(t, k) = 2k \begin{pmatrix} 0 & g_0(t) \\ \bar{g}_0(t) & 0 \end{pmatrix} - i \begin{pmatrix} 0 & g_1(t) \\ \bar{g}_1(t) & 0 \end{pmatrix} \sigma_3 - i|g_0(t)|^2 \sigma_3.$$

THEOREM 4.1. *Given $q_0(x) \in \mathcal{S}(\mathbb{R}^+)$ define $\{a(k), b(k)\}$ according to the definition (4.15). Suppose that there exist smooth functions $\{g_0(t), g_1(t)\}$ satisfying $g_0(0) = q_0(0)$, $g_1(0) = \partial_x q(0)$, such that the functions $\{A(k), B(k)\}$ which are defined from $\{g_l(t)\}_0^1$ according to definition (4.18) satisfy the global relation*

$$(4.20) \quad a(k)B(k) - b(k)A(k) = e^{4ik^2 T} c(k), \quad \text{Im} k \geq 0,$$

where $c(k)$ is analytic and bounded for $\text{Im} k > 0$ and is of $O(1/k)$, $k \rightarrow \infty$.

Define $M(x, t, k)$ as the solution of the following 2×2 matrix RH problem:

M is holomorphic for k in $\mathbb{C} \setminus \mathcal{L}$, where \mathcal{L} is the union of the real and of the imaginary axes of the complex k -plane.

$$M_-(x, t, k) = M_+(x, t, k)J(x, t, k), \quad k \in \mathcal{L},$$

where J is defined in terms of a, b, A, B by (see Figure 4.5)

$$J_1 = \begin{pmatrix} 1 & 0 \\ \Gamma(k)e^{2i\theta} & 1 \end{pmatrix}, \quad J_3 = \begin{pmatrix} 1 & -\overline{\Gamma(\bar{k})}e^{-2i\theta} \\ 0 & 1 \end{pmatrix},$$

$$J_4 = \begin{pmatrix} 1 & -\gamma(k)e^{-2i\theta} \\ \bar{\gamma}(k)e^{2i\theta} & 1 - |\gamma(k)|^2 \end{pmatrix},$$

where

$$\gamma(k) = \frac{b(k)}{a(k)}, k \in \mathbb{R}; \Gamma(k) = \frac{\overline{B(\bar{k})}}{a(k)d(k)}, \frac{\pi}{2} \leq \arg k \leq \pi; \theta(x, t, k) = kx + 2k^2t,$$

$$d(k) = a(k)\overline{A(\bar{k})} - b(k)\overline{B(\bar{k})}$$

$$M(x, t, k) = I + O(1/k), \quad k \rightarrow \infty.$$

Then $M(x, t, k)$ exists and is unique.

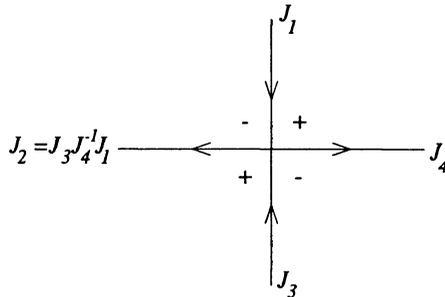


FIGURE 4.5

Define $q(x, t)$ by

$$q(x, t) = 2i \lim_{k \rightarrow \infty} (kM(x, t, k))_{12}.$$

Then $q(x, t)$ solves the NLS equation with

$$q(x, 0) = q_0(x), \quad q(0, t) = g_0(t), \quad q_x(0, t) = g_1(t).$$

The global relation plays a crucial role in the proof of this theorem. Indeed, $q(0, t)$ and $q_x(0, t)$ are defined through $M(0, t, k)$ whose jump matrix involves $\exp[4ik^2t]$, $\{a(k), b(k), A(k), B(k)\}$. On the other hand $g_0(t)$ and $g_1(t)$ are defined through the inverse of the map $\tilde{\mathbf{S}}$, (4.17). It can be shown that this inverse map can be expressed in terms of a RH problem for a 2×2 matrix $M^{(t)}(t, k)$ whose jump matrix involves $\exp[4ik^2t]$ and $\{A(k), B(k)\}$. It can be shown that these two RH problems are equivalent iff the global relation is valid.

Step 3. Analyze the Global Relation

The global relation together with the definition of $\{A(k), B(k)\}$ yield a *nonlinear Volterra integral equation* for $g_1(t)$ in terms of $g_0(t)$ and $q_0(t)$. It is shown in [24] that this nonlinear equation has a global solution.

We recall that the analogous step for linear evolution equations was solved by algebraic manipulations. This was based on the invariance of the global relation under $k \rightarrow \nu(k)$. Unfortunately, the global relation now involves $\Phi(t, k)$ which in

general breaks this invariance. However, for a *particular* class of boundary conditions this invariance survives. This is precisely the class of “linearizable problems”, namely a class of problems for which $\{A(k), B(k)\}$ can be explicitly written in terms of $\{a(k), b(k)\}$.

Some linearizable cases are given below: The basic RH problem has a jump matrix which is uniquely defined in terms of the scalar functions $a(k)$, $b(k)$, and $\Gamma(k)$, where $\Gamma(k)$ involves $a(k)$, $b(k)$, and $B(k)/A(k)$,

$$\overline{\Gamma(\bar{k})} = \frac{\rho \frac{B(k)}{A(k)}}{a(\bar{k}) \left[a(\bar{k}) - \rho \overline{b(\bar{k})} \frac{B(k)}{A(k)} \right]}.$$

The basic RH problems for the KdV with dominant surface tension and for the sine Gordon have a similar form [25], where $\rho = \pm 1$ for the NLS, $\rho = 1$ for the KdV, $\rho = -1$ for the sine Gordon (sG).

In [25] the following concrete linearizable cases are solved.

NLS:

$$q_x(0, t) - \chi q(0, t) = 0, \quad \chi \text{ constant}, \quad \chi \geq 0.$$

sG:

$$q(0, t) = \chi, \quad \chi \text{ constant.}$$

KdV:

$$q(0, t) = \chi, \quad q_{xx}(0, t) = \chi + 3\chi^2, \quad \chi \text{ constant.}$$

For each of these cases, B/A , and hence $\Gamma(k)$, can be explicitly given in terms of $a(k), b(k)$:

$$\text{NLS : } \frac{B(k)}{A(k)} = -\frac{2k + i\chi b(-k)}{2k - i\chi a(-k)},$$

$$\text{KdV, sG : } \frac{B(k)}{A(k)} = \frac{f(k)b(\nu(k)) - a(\nu(k))}{f(k)a(\nu(k)) - b(\nu(k))},$$

where for the sG,

$$\nu(k) = \frac{1}{k}, \quad f(k) = i \frac{k^2 + 1}{k^2 - 1} \frac{\sin \chi}{\cos \chi - 1},$$

while for the KdV,

$$\nu^2 + k\nu + k^2 + \frac{1}{4} = 0, \quad f(k) = \frac{\nu + k}{\nu - k} \left(1 - \frac{4\nu k}{\chi} \right).$$

We emphasize that since $\{a(k), b(k)\}$ are determined in terms of the initial conditions and since $B(k)/A(k)$ and therefore $\Gamma(k)$ is explicitly written in terms of $\{a(k), b(k)\}$, it follows that linearizable initial boundary value problems on the half line are solved as effectively as initial value problems on the line.

REMARK 4.2. We discuss the general features of this method for an integrable evolution equation with spatial derivatives of order n .

- (1) The “jump matrices” of the relevant RH problem have *explicit* x, t dependence of the form $\exp[if_1(k)x + if_2(k)t]$, and they depend on the calar functions $\{a(k), b(k), A(k), B(k)\}$ (compare with Theorem4.1). This means that the associated expression for $q(x, t)$ provides the proper non-linearization of the Euler-Ehrenpreis-Palamodov representation as well as the proper *spectral representation* of the solution. This representation involves the direct and the inverse map between the values of $q(x, t)$ on the boundary, i.e. $\{q(x, 0), \{\partial_x^l q(0, t)\}_0^{n-1}\}$, and the spectral functions

$\{a(k), b(k), A(k), B(k)\}$. We emphasize that for a proper spectral decomposition (since the value of $q(x, t)$ on the boundary are functions of one variable only) the spectral functions must be functions of only one variable.

- (2) Precisely because the solution is given in the above spectral representation form, it is possible to study effectively the asymptotic properties of the solution, such as its long t behavior. For the NLS, sG and KdV equations on the half line this has been done in [35], [36], [37] respectively. The relevant analysis is based on the basic RH problem and on the Deift-Zhou method [38]. The latter method is an elegant nonlinearization of the steepest descent method and it yields rigorous asymptotic results for RH problems with exponential x, t dependence. In our opinion this result is one of the most important developments in the theory of integrable systems in particular and in the theory of RH problems in general, thus it is quite satisfying that the new method gives rise to RH problems precisely of the type that can be analyzed by the Deift-Zhou method. We also note that recently a highly nontrivial generalization of the Deift-Zhou method has been developed which is able to analyze the zero-dispersion limit of the Cauchy problem on the line [39]. Since this method is also based on the analysis of a RH problem with exponential x, t dependence, we expect that the method of [39] applied to our RH problem will yield an effective description of the zero dispersion limit of initial-boundary value problems on the half-line [32].
- (3) It is the authors opinion that the most remarkable fact about boundary value problems for integrable nonlinear PDEs is the *simplicity of the global relation*. Indeed, although the relation between the initial and the boundary values of q is very complicated, this relation takes a simple algebraic form in the k -space, see equation (4.20). The simplicity of the global relation has two important consequences: (a) Under the assumption that there exist spectral functions satisfying this relation, it is possible to prove that the associated $q(x, t)$ exists, satisfies the given nonlinear PDE, and $q(x, 0) = q_0(x)$, $\{\partial_x^l q(0, t) = g_l(t)\}_0^{n-1}$. (b) Given initial conditions and a subset of $\{g_l(t)\}_0^{n-1}$ it is possible to prove the *global* existence of the remaining part of this set. We emphasize that the global relation is a simple algebraic relation between the two components of an eigensolution of the t -part of the Lax pair evaluated at $x = 0$. Thus since the components satisfy a *linear* eigenvalue equation, the derivation of appropriate estimates for their large k behavior is based on the analysis of a linear problem. Thus, although the global relation is a nonlinear equation its rigorous investigation involves mostly the analysis of a linear equation.
- (4) In recent years there have been important developments in the analysis of boundary value problems of nonlinear PDEs using PDE techniques [40], [41]. It is remarkable that some of these techniques yield *global* results. It is satisfying that there exist now a rigorous theory using the integrability machinery, so that it is possible to make comparisons between these different approaches. Although at the moment the PDE results are proven in less restrictive functional spaces, the advantage of our method

is that it yields rigorous asymptotic results. We reiterate that this is a consequence of the Deift-Zhou theory and of our simple RH problem.

Appendix A

An explicit formula for $W(x_1, x_2, k)$ associated with an arbitrary two dimensional linear operator with constant coefficients is given in [34]. Here we present two examples.

EXAMPLE A.1. A closed differential 1-form associated with equation (1.1) is

$$(A.21) \quad W(x, t, k) = e^{-ikx+iw(k)t} \left[qdx - \frac{w(k) - w(-i\partial_x)}{k + i\partial_x} qdt \right].$$

Indeed, let

$$W = e(qdx - Adt), \quad e = \exp[-ikx + iw(k)t].$$

Then

$$dW = [(eq)_t + (eA)_x]dt \wedge dx = -ie[(w(k) - w(-i\partial_x))q + (k + i\partial_x)A]dt \wedge dx.$$

Hence $dW = 0$ provided that

$$A = \frac{w(k) - w(-i\partial_x)}{k + i\partial_x} q.$$

We note that $(k + i\partial_x)$ is a factor of $w(k) - w(-i\partial_x)$, thus A involves q and its spatial derivatives. For example, if $w(k) = k - k^3$, then

$$\frac{w(k) - w(l)}{k - l} = 1 - (k^2 + l^2 + kl),$$

thus

$$A = (1 - k^2)q + ikq_x + q_{xx}.$$

EXAMPLE A.2. A closed differential 1-form associated with equation (3.1) is given by (3.2).

Indeed, let

$$W = e(Adz - Bd\bar{z}), \quad e = \exp[-ikz - \frac{i\alpha}{k}\bar{z}].$$

Then

$$(A.22) \quad dW = [(eA)_{\bar{z}} + (eB)_z] d\bar{z} \wedge dz = e \left[\left(\partial_{\bar{z}} - \frac{i\alpha}{k} \right) A + (\partial_z - ik)B \right] d\bar{z} \wedge dz.$$

Writing equation (3.1) in the form

$$\left(\partial_{\bar{z}} - \frac{i\alpha}{k} \right) q_z + (\partial_z - ik) \frac{i\alpha}{k} q = 0,$$

it follows that equation (A.22) can be rewritten as

$$dW = e \left\{ \left(\partial_{\bar{z}} - \frac{i\alpha}{k} \right) (A - q_z) + (\partial_z - ik) \left(B - \frac{i\alpha}{k} q \right) \right\} d\bar{z} \wedge dz,$$

which implies $A = q_z$ and $B = i\alpha q/k$.

Appendix B

Solving equation (1.5) with a Fourier transform we find

$$(B.23) \quad q(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx - iw(k)t} \hat{q}_0(k) dk + \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx} \left(\int_0^t e^{iw(k)(\tau-t)} \tilde{q}(0, \tau, k) d\tau \right) dk,$$

where $\hat{q}_0(k)$ is defined by equation (2.4) and

$$(B.24) \quad w = k - k^3, \quad \tilde{q}(x, t, k) = q_{xx} + ikq_x + (1 - k^2)q.$$

The simplest way to derive equation (B.23) is to note that equation (1.5) is equivalent to (compare with (1.10))

$$(B.25) \quad \left(e^{-ikx + iw(k)t} q \right)_t = - \left(e^{-ikx + iw(k)t} \tilde{q} \right)_x.$$

Let

$$(B.26) \quad \hat{q}(k, t) = \int_0^{\infty} e^{-ikx} q(x, t) dx.$$

Then

$$\left(e^{iw(k)t} \hat{q}(k, t) \right)_t = \int_0^{\infty} \left(e^{-ikx + iw(k)t} q(x, t) \right)_t dx = e^{iw(k)t} \tilde{q}(0, t, k),$$

where we have used (B.25) to compute the above integral. Hence

$$\hat{q}(k, t) = e^{-iw(k)t} \hat{q}_0(k) + \int_0^t e^{-iw(k)(\tau-t)} \tilde{q}(0, \tau, k) d\tau,$$

and the inverse Fourier transform of equation (B.26) implies (B.23).

Equation (B.23) can be rewritten in the form

$$(B.27) \quad q(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx - iw(k)t} \hat{q}_0(k) dk + \frac{1}{2\pi} \int_{\partial D_+} e^{ikx - iw(k)t} \hat{g}(k, t) dk,$$

where ∂D_+ is defined in equation (2.3) and

$$(B.28) \quad \hat{q}(k, t) = (1 - k^2) \hat{g}_0(k, t) + ik \hat{g}_1(k, t) + \hat{g}_2(k, t),$$

$$\hat{g}_j(k, t) = \int_0^t e^{iw(k)\tau} g_j(\tau) d\tau, \quad g_j(t) = \partial_x^j q(0, t), \quad j = 0, 1, 2.$$

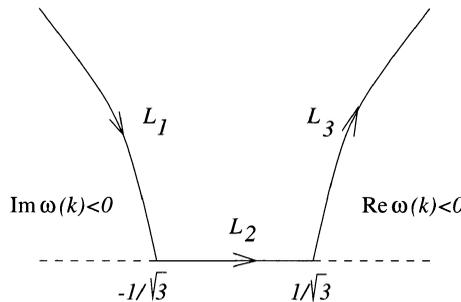


FIGURE B.1

Indeed, we write ∂D_+ as the union of the three contours L_j , $j = 1, 2, 3$, depicted in Figure B.1. The function $\exp[ikx - iw(k)t]\hat{g}(k, t)$ is analytic and bounded in the region bounded by L_1 and by the ray $\text{Re } k \in [-\infty, -1/\sqrt{3}]$. Thus the integral along this ray can be deformed along L_1 . Similar considerations apply to the ray $\text{Re } k \in [1/\sqrt{3}, \infty)$.

We note that if $\hat{g}(k)$ is defined by equation (2.8) then $\hat{g}(k)$ equals $\hat{g}(k, t)$ plus a term involving an integral from t to T . This term is analytic and bounded in the domain D_+ thus, by Cauchy's theorem this term does not contribute to equation (B.27). Hence $q(x, t)$ is given by either equation (B.27) or equation (2.2).

References

- [1] A.S. Fokas, Proc. R. Soc. London, Ser. A **53**, 1411 (1997).
- [2] A.S. Fokas, On the integrability of linear and nonlinear partial differential equations, J. Math. Phys., **41**, 4188-4237 (2000).
- [3] I. Stakgold, *Green's Functions and Boundary Value Problems*, Wiley-Interscience, New York, 1979.
- [4] B. Friedman, *Principles and Techniques of Applied Mathematics*, Wiley, New York, 1956.
- [5] A.S. Fokas and B. Pelloni, Proc. Roy. Soc. Lon. A **456**, 805-833 (2000).
- [6] B. Noble, *Methods Based on the Wiener-Hopf Technique*, Pergamon, New York, 1958.
- [7] L. Ehrenpreis, *Fourier analysis in several complex variables*, Wiley-Interscience Publishers, New York-London-Sydney, 1970.
- [8] V.P. Palamodov, *Linear differential operators with constant coefficients*, Springer Verlag, 1970.
- [9] G. Henkin, Method of integral representations in complex analysis, *Encyclopedia of Mathematical Sciences*, volume 7: several complex variables, Springer-Verlag, 1990.
- [10] Berndtsson and M. Passare, J. Funct. Anal. **84**, 358 (1989); A. Yger, *Formules de Division at Prolongement Meromorphe*, Lecture Notes in Math. **1295**, Springer-Verlag, Berlin, 1987; S. Rigat, J. Math. Pures Appl. **76**, 777 (1997).
- [11] G.S. Gardner, J.M. Greene, M.D. Kruskal, R.M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
- [12] A.S. Fokas and V.E. Zakharov, Eds, *Important Developments in Soliton Theory*, Springer Verlag (1994).
- [13] P. Deift, *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach*, Courant Institute Lecture Notes, 1999.
- [14] P.D. Lax, Integrals of nonlinear equations of evolution and solitary waves, Commun. Pure Appl. Math. **21**, 467-490 (1968).
- [15] A.S. Fokas and I.M. Gelfand, Lett. Math. Phys. **32**, 189 (1994).
- [16] A.S. Fokas, L.Y. Sung, Initial boundary value problems for linear evolution equations on the half line, preprint <http://www.ma.ic.ac.uk/asf/>.
- [17] A.S. Fokas, B. Pelloni, Two-point boundary value problems for linear evolution equations, Math. Proc. Camb. Phil. Soc., 1-23 (2001).
- [18] A.S. Fokas, B. Pelloni, Method for solving moving boundary value problems for linear evolution equations, Phys. Rev. Lett. **84**, 4785-4789 (2000); Integrable evolution equations in time dependent domains, Inv. Prob. **17**, 919-935 (2001).
- [19] D. ben-Avraham, A.S. Fokas, The solution of the modified Helmholtz equation in a wedge and an application to diffusion-limited coalescence, Phys. Lett. A. **263**, 355 (1999).
- [20] D. ben-Avraham, A.S. Fokas, The Modified Helmholtz Equation in a Triangular Domain and an Application to Diffusion-Limited Coalescence, Phys. Rev. E **36**, 1614 (2001).
- [21] A.S. Fokas, A.A. Kapaev, A., Riemann-Hilbert approach to the Laplace equation, J. Math. Anal. Appl. **251**, 770-804 (2000).
- [22] Y. Antipov and A.S. Fokas, The Modified Helmholtz Equation on a Semi-Strip (preprint) 2002.
- [23] D. Crowdy, A.S. Fokas, Explicit integral solutions for the plane elastostatic semi-strip (preprint) 2002.
- [24] A.S. Fokas, A.R. Its, L.Y. Sung, The nonlinear Schrödinger equation on the half line (preprint) 2001.

- [25] A.S. Fokas, Nonlinear Integrable Evolution Equations on the Half-Line, *Comm. Math. Phys.* (to appear).
- [26] A.S. Fokas and A.R. Its, The nonlinear Schrödinger on the interval (preprint) 2002.
- [27] S. Fulton, A.S. Fokas, C. Xenophontos, An analytical method for linear elliptic PDEs and its numerical implementation (preprint) 2001.
- [28] S. Brenner, A.S. Fokas, L.Y. Sung (in preparation).
- [29] J. Bona, L.Y. Sung (private communication).
- [30] A.S. Fokas, A New Transform Method for Evolution Equations on the Half Line, *IMA J. Appl. Math.* (to appear).
- [31] A.S. Fokas and P. Schultz, The long Time Asymptotics of Moving Boundary Problems using an Ehrenpreis-Type Representation and its Riemann-Hilbert Nonlinearization (preprint).
- [32] A.S. Fokas and S. Kamvisis (in preparation).
- [33] A.S. Fokas, Two dimensional linear PDE in a convex polygon, *Proc. R. Soc. Lond. A* **457**, 371-393 (2001).
- [34] A.S. Fokas and M. Zyskin, The Fundamental Differential Form and Boundary Value Problems, *Q. J. Mech. Appl. Math.*, 1-23 (2002)
- [35] A.S. Fokas and A.R. Its, *SIAM J. Math. Anal.*, **27**, 738 (1996).
- [36] A.S. Fokas, A.R. Its, *Theoretical and Mathematical Physics*, **92**, 387-403 (1992).
- [37] A.S. Fokas, A.R. Its, *Mathematics and Computers in Simulation*, **37**, 293-321 (1994).
- [38] P.A. Deift, X. Zhou, A steepest descent method for oscillatory Riemann-Hilbert problems, *Ann. Math.*, **137**, 245-338 (1993); *Bull. Am. Math. Soc.*, **20**, 119 (1992).
- [39] P.A. Deift, S. Venakides and X. Zhou, New results in the small dispersion KdV by an extension of the method of steepest descent for Riemann-Hilbert problems, *IMRN* **6**, 285-299 (1997).
- [40] J.E. Colliander and C.E. Kenig, The Generalized KdV Equation on the Half Line (preprint) 2000.
- [41] J. Bona, S. Sun and B.Y. Zhang, A Non-Homogeneous Boundary Value Problem for the KdV Equation on a Quarter-Plane (preprint) 2000.

DEPARTMENT OF APPLIED MATHEMATICS AND THEORETICAL PHYSICS UNIVERSITY OF CAMBRIDGE, CAMBRIDGE, CB3 9EW, UNITED KINGDOM

E-mail address: T.Fokas@damtp.cam.ac.uk

URL: <http://www.damtp.cam.ac.uk/user/tf227/>

Chaos in Partial Differential Equations

Yanguang (Charles) Li

ABSTRACT. This is a survey on Chaos in Partial Differential Equations. First we classify soliton equations into three categories: 1. (1+1)-dimensional soliton equations, 2. soliton lattices, 3. (1+n)-dimensional soliton equations ($n \geq 2$). A systematic program has been established by the author and collaborators, for proving the existence of chaos in soliton equations under perturbations. For each category, we pick a representative to present the results. Then we review some initial results on 2D Euler equation.

1. Introduction

It is well-known that the theory of chaos in finite-dimensional dynamical systems has been well-developed. That includes both discrete maps and systems of ordinary differential equations. Such theory has produced important mathematical theorems and led to important applications in physics, chemistry, biology, and engineering etc. [8] [26]. On the contrary, the theory of chaos in partial differential equations has not been well-developed. On the other hand, the demand for such a theory is much more stronger than for finite-dimensional systems. Mathematically, studies on infinite-dimensional systems pose much more challenging problems. For example, as phase spaces, Banach spaces possess much more structures than Euclidean spaces. In terms of applications, most of important natural phenomena are described by partial differential equations, nonlinear wave equations, Yang-Mills equations, and Navier-Stokes equations, to name a few.

Nonlinear wave equations are the most important class of equations in natural sciences. They describe a wide spectrum of phenomena; motion of plasma, nonlinear optics (laser), water waves, vortex motion, to name a few. Among these nonlinear wave equations, there is a class of equations called soliton equations. This class of equations describes a variety of phenomena. In particular, the same soliton equation describes several different phenomena. For references, see for example [3] [1]. Mathematical theories on soliton equations have been well developed. Their

1991 *Mathematics Subject Classification.* Primary 35Q55, 35Q30; Secondary 37L10, 37L50, 35Q99.

Key words and phrases. Homoclinic orbits, chaos, Lax pairs, Darboux transformations, invariant manifolds.

This work was partially supported by a Guggenheim Fellowship.

Cauchy problems are completely solved through inverse scattering transforms. Soliton equations are integrable Hamiltonian partial differential equations which are the natural counterparts of finite-dimensional integrable Hamiltonian systems.

To set up a systematic study on chaos in partial differential equations, we started with the perturbed soliton equations. We classify the perturbed soliton equations into three categories:

- (1) Perturbed (1+1)-Dimensional Soliton Equations,
- (2) Perturbed Soliton Lattices,
- (3) Perturbed (1 + n)-Dimensional Soliton Equations ($n \geq 2$).

For each category, we chose a candidate to study. The integrable theories for every members in the same category are parallel, and for members in different categories are substantially different. The theorem on the existence of chaos for each candidate can be parallelly generalized to the rest members of the same category.

- The candidate for Category 1 is the perturbed cubic focusing nonlinear Schrödinger equation [22] [21] [14],

$$i\partial_t q = \partial_x^2 q + 2[|q|^2 - \omega^2]q + \text{Perturbations},$$

under periodic and even boundary conditions $q(x+1) = q(x)$ and $q(-x) = q(x)$, ω is a real constant.

- The candidate for Category 2 is the perturbed discrete cubic focusing nonlinear Schrödinger equation [11] [23] [24],

$$iq'_n = \frac{1}{h^2}[q_{n+1} - 2q_n + q_{n-1}] \\ + |q_n|^2(q_{n+1} + q_{n-1}) - 2\omega^2 q_n + \text{Perturbations},$$

under periodic and even boundary conditions $q_{n+N} = q_n$ and $q_{-n} = q_n$.

- The candidate for Category 3 is the perturbed Davey-Stewartson II equations [15],

$$\begin{cases} i\partial_t q = [\partial_x^2 - \partial_y^2]q + [2(|q|^2 - \omega^2) + u_y]q + \text{Perturbations}, \\ [\partial_x^2 + \partial_y^2]u = -4\partial_y |q|^2, \end{cases}$$

under periodic and even boundary conditions

$$q(t, x + l_x, y) = q(t, x, y) = q(t, x, y + l_y), \\ u(t, x + l_x, y) = u(t, x, y) = u(t, x, y + l_y),$$

and

$$q(t, -x, y) = q(t, x, y) = q(t, x, -y), \\ u(t, -x, y) = u(t, x, y) = u(t, x, -y).$$

We have established a standard program for proving the existence of chaos in perturbed soliton equations, with the machineries:

- (1) Darboux Transformations for Soliton Equations.
- (2) Isospectral Theory for Soliton Equations Under Periodic Boundary Condition.
- (3) Persistence of Invariant Manifolds and Fenichel Fibers.
- (4) Melnikov Analysis.
- (5) Smale Horseshoes and Symbolic Dynamics Construction of Conley-Moser Type.

The most important implication of the theory on chaos in partial differential equations in theoretical physics will be on the study of turbulence. For that goal, we chose the 2D Navier-Stokes equations under periodic boundary conditions to begin a dynamical system study on 2D turbulence. Since they possess a Lax pair [17] and Darboux transformation [25], the 2D Euler equations are the starting point for an analytical study. The high Reynolds number 2D Navier-Stokes equations are viewed as a singular perturbation of the 2D Euler equations through the perturbation parameter $\varepsilon = 1/Re$ which is the inverse of the Reynolds number. Corresponding singular perturbations of nonlinear Schrödinger equation have been studied in [31] [30] [19] [20]. We have studied the linearized 2D Euler equations and obtained a complete spectra theorem [16]. In particular, we have identified unstable eigenvalues. Then we found the approximate representations of the hyperbolic structures associated with the unstable eigenvalues through Galerkin truncations [18].

2. Existence of Chaos in Perturbed Soliton Equations

By existence of chaos, we mean that there exist a Smale horseshoe and the Bernoulli shift dynamics for certain Poincaré map. For lower dimensional systems, there have been a lot of theorems on the existence of chaos [8] [26]. For perturbed soliton equations under dissipative perturbations, we first establish the existence of a Silnikov homoclinic orbit. And then we define a Poincaré section which is transversal to the Silnikov homoclinic orbit, and the Poincaré map on the Poincaré section induced by the flow. Finally we construct the Smale horseshoe for the Poincaré map. In establishing the existence of the Silnikov homoclinic orbit, we need to build a Melnikov analysis through Darboux transformations to generate the explicit representation for the unperturbed heteroclinic orbit, the isospectral theory for soliton equations to generate the Melnikov vectors, and the persistence of invariant manifolds and Fenichel fibers. We also need to utilize the properties of the Fenichel fibers to build a second measurement inside a slow manifold, together with normal form techniques. The Melnikov measurement and the second measurement together lead to the existence of the Silnikov homoclinic orbit through implicit function arguments. In establishing the existence of Smale horseshoes for the Poincaré map, we first need to establish a smooth linearization in the neighborhood of the saddle point (i.e. the asymptotic point of the Silnikov homoclinic orbit). Then the dynamics in the neighborhood of the saddle point is governed by linear partial differential equations which are explicitly solvable. The global dynamics in the tubular neighborhood of the Silnikov homoclinic orbit away from the above neighborhood of the saddle point, can be approximated by linearized flow along the Silnikov homoclinic orbit due to finiteness of the passing time. Finally we can obtain a semi-explicit representation for the Poincaré map. Then we establish the existence of fixed points of the Poincaré map under certain except-one-point conditions. And we study the action of the Poincaré map in the neighborhood of these fixed points, and verify the Conley-Moser criteria to establish the existence of Smale horseshoes and Bernoulli shift dynamics.

2.1. Existence of Chaos in Perturbed (1+1)-Dimensional Soliton Equations. For this category of the perturbed soliton equations, we chose the candidate to be the perturbed cubic nonlinear Schrödinger equation. The cubic nonlinear Schrödinger equation describes self-focusing phenomena in nonlinear optics, deep water surface wave, vortex filament motion etc.. Recently, more and more interests

are on perturbed nonlinear Schrödinger equations describing new nonlinear optical effects, for example, the works of the Laser Center at Oklahoma State University.

2.1.1. *Dissipative Perturbations.* In a series of three papers [22] [21] [14], we proved the existence of chaos in the cubic nonlinear Schrödinger equation under dissipative perturbations. We study the following perturbed nonlinear Schrödinger equation:

$$(2.1) \quad iq_t = q_{xx} + 2\left[|q|^2 - \omega^2\right]q + i\varepsilon\left[-\alpha q + \hat{D}^2q + \Gamma\right],$$

under even periodic boundary conditions

$$q(-x) = q(x), \quad q(x + 1) = q(x);$$

where $i = \sqrt{-1}$, q is a complex-valued function of two variables (x, t) , (ω, α, Γ) are positive constants, ε is the positive perturbation parameter, \hat{D}^2 is a “regularized” Laplacian specifically defined by

$$\hat{D}^2q \equiv -\sum_{j=1}^{\infty} \beta_j k_j^2 \hat{q}_j \cos k_j x,$$

in which $k_j = 2\pi j$, \hat{q}_j is the Fourier transform of q , $\beta_j = \beta$ for $j \leq N$, $\beta_j = \alpha_* k_j^{-2}$ for $j > N$, β and α_* are positive constants, and N is a large fixed positive integer.

THEOREM 2.1 (Homoclinic Orbit Theorem). *There exists a positive number ε_0 such that for any $\varepsilon \in (0, \varepsilon_0)$, there exists a codimension 1 hypersurface E_ε in the external parameter space $\{\omega, \alpha, \Gamma, \beta, \alpha_*\}$. For any external parameters $(\omega, \alpha, \Gamma, \beta, \alpha_*) \in E_\varepsilon$, there exists a symmetric pair of homoclinic orbits $h_k = h_k(t, x)$ ($k = 1, 2$) in $H_{\varepsilon, p}^1$ (the Sobolev space H^1 of even and periodic functions) for the PDE (2.1), which are asymptotic to a fixed point q_ε . The symmetry between h_1 and h_2 is reflected by the relation that h_2 is a half-period translate of h_1 , i.e. $h_2(t, x) = h_1(t, x + 1/2)$. The hypersurface E_ε is a perturbation of a known surface $\beta = \kappa(\omega)\alpha$, where $\kappa(\omega)$ is shown in Figure 1.*

For the complete proof of the theorem, see [22] and [21]. The main argument is a combination of a Melnikov analysis and a geometric singular perturbation theory for partial differential equations. The Melnikov function is evaluated on a homoclinic orbit of the nonlinear Schrödinger equation, generated through Darboux transformations. For more details on this, see the later section on the Darboux transformations for the discrete nonlinear Schrödinger equation.

THEOREM 2.2 (Horseshoe Theorem). *Under certain generic assumptions for the perturbed nonlinear Schrödinger system (2.1), there exists a compact Cantor subset Λ of Σ (a Poincaré section transversal to the homoclinic orbit), Λ consists of points, and is invariant under P (the Poincaré map induced by the flow on Σ). P restricted to Λ , is topologically conjugate to the shift automorphism χ on four symbols $1, 2, -1, -2$. That is, there exists a homeomorphism*

$$\phi : \mathcal{W} \mapsto \Lambda,$$

where \mathcal{W} is the topological space of the four symbols, such that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{W} & \xrightarrow{\phi} & \Lambda \\ \chi \downarrow & & \downarrow P \\ \mathcal{W} & \xrightarrow{\phi} & \Lambda \end{array}$$

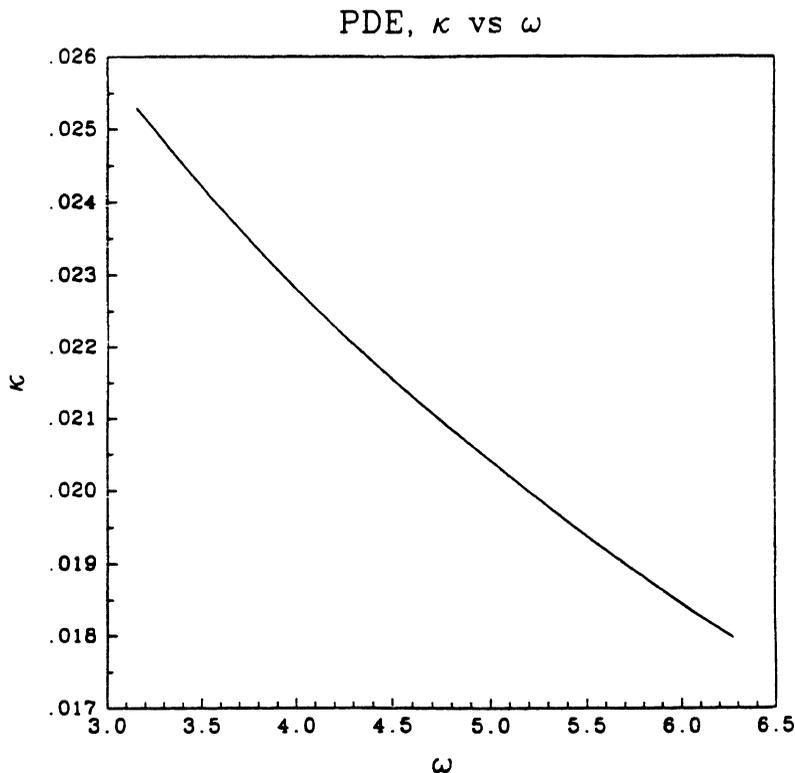


FIGURE 1. The curve of $\kappa = \kappa(\omega)$.

For the complete proof of the theorem, see [14]. The construction of horseshoes is of Conley-Moser type for partial differential equations.

2.1.2. *Singular Perturbations.* Recently, singular perturbation, i.e. replacing $\hat{D}^2 q$ by $\partial_x^2 q$, has been studied [31] [30] [19] [20]. Consider the singularly perturbed nonlinear Schrödinger equation,

$$(2.3) \quad iq_t = q_{xx} + 2[|q|^2 - \omega^2]q + i\varepsilon[-\alpha q + \beta q_{xx} + \Gamma],$$

where $q = q(t, x)$ is a complex-valued function of the two real variables t and x , t represents time, and x represents space. $q(t, x)$ is subject to periodic boundary condition of period 1, and even constraint, i.e.

$$q(t, x + 1) = q(t, x), \quad q(t, -x) = q(t, x).$$

ω is a positive constant, $\alpha > 0$, $\beta > 0$, and Γ are constants, and $\varepsilon > 0$ is the perturbation parameter. The main difficulty introduced by the singular perturbation $\varepsilon \partial_x^2$ is that it breaks the spectral gap condition of the unperturbed system. Therefore, standard invariant manifold results will not apply. Nevertheless, it turns out that certain invariant manifold results do hold. The regularity of such invariant manifolds at $\varepsilon = 0$ is controlled by the regularity of $e^{\varepsilon \partial_x^2}$ at $\varepsilon = 0$.

THEOREM 2.3 (Li, [19]). *There exists a $\varepsilon_0 > 0$, such that for any $\varepsilon \in (0, \varepsilon_0)$, there exists a codimension 1 surface E_ε in the space of $(\omega, \alpha, \beta, \Gamma) \in \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+$, where $\omega \in (\pi, 2\pi)/S$, S is a finite subset. For any external parameters*

on the codimension-one surface, the perturbed nonlinear Schrödinger equation (2.3) possesses a symmetric pair of homoclinic orbits $h_k = h_k(t, x)$ ($k = 1, 2$) in $C_{\varepsilon, p}^\infty[0, 1]$ (the space of C^∞ even and periodic functions on the interval $[0, 1]$), which is asymptotic to a saddle fixed point q_ε . The symmetry between h_1 and h_2 is reflected by the relation that h_2 is a half-period translate of h_1 , i.e. $h_2(t, x) = h_1(t, x + 1/2)$. The hypersurface E_ε is a perturbation of a known surface $\beta = \kappa(\omega)\alpha$, where $\kappa(\omega)$ is shown in Figure 1.

2.1.3. Hamiltonian Perturbations. The problem on the existence of chaos in the cubic nonlinear Schrödinger equations under Hamiltonian perturbations is largely open. The right objects to investigate should be “homoclinic tubes” rather than “homoclinic orbits” due to the non-dissipative nature and infinite-dimensionality of the perturbed system. Transversal homoclinic tubes are objects of large dimensional generalization of transversal homoclinic orbits. As Smale’s theorem indicates, structures in the neighborhood of a transversal homoclinic orbit are rich, structures in the neighborhood of a transversal homoclinic tube are even richer. Especially in high dimensions, dynamics inside each invariant tubes in the neighborhoods of homoclinic tubes are often chaotic too. We call such chaotic dynamics “*chaos in the small*”, and the symbolic dynamics of the invariant tubes “*chaos in the large*”. Such cascade structures are more important than the structures in a neighborhood of a homoclinic orbit, when high or infinite dimensional dynamical systems are studied. Symbolic dynamics structures in the neighborhoods of homoclinic tubes are more observable than in the neighborhoods of homoclinic orbits in numerical and physical experiments. When studying high or infinite dimensional Hamiltonian system (for example, the cubic nonlinear Schrödinger equation under Hamiltonian perturbations), each invariant tube contains both KAM tori and stochastic layers (chaos in the small). Thus, not only dynamics inside each stochastic layer is chaotic, all these stochastic layers also move chaotically under Poincaré maps.

There have been a lot of works on the KAM theory of soliton equations under Hamiltonian perturbations [29] [5] [9] [4] [28]. For perturbed nonlinear Schrödinger equations, we are interested in the region of the phase space where there exist hyperbolic structures. Thus, the relevant KAM tori are hyperbolic. In finite dimensions, the relevant work on such tori is that of Graff [7]. In infinite dimensions, the author is not aware of such work yet.

In the paper [13], the author studied the cubic nonlinear Schrödinger equation under Hamiltonian perturbations:

$$(2.4) \quad iq_t = q_{xx} + 2[|q|^2 - \omega^2]q + \varepsilon[\alpha_1 + 2\alpha_2\bar{q}],$$

under even periodic boundary conditions $q(-x) = q(x)$ and $q(x+1) = q(x)$; where $i = \sqrt{-1}$, q is a complex-valued function of two variables (t, x) , $(\omega, \alpha_1, \alpha_2)$ are real constants, ε is the perturbation parameter. The system (2.4) can be written in the Hamiltonian form:

$$iq_t = \delta H / \delta \bar{q},$$

where $H = H_0 + \varepsilon H_1$,

$$H_0 = \int_0^1 [|q|^4 - 2\omega^2|q|^2 - |q_x|^2] dx,$$

$$H_1 = \int_0^1 [\alpha_1(q + \bar{q}) + \alpha_2(q^2 + \bar{q}^2)] dx.$$

DEFINITION 2.4. Denote by $W^{(c)}$ a normally hyperbolic center manifold, by $W^{(cu)}$ and $W^{(cs)}$ the center-unstable and center-stable manifolds such that $W^{(c)} = W^{(cu)} \cap W^{(cs)}$, and by F^t the evolution operator of the partial differential equation. Let \mathcal{H} be a submanifold in the intersection between the center-unstable and center-stable manifolds $W^{(cu)}$ and $W^{(cs)}$, such that for any point $q \in \mathcal{H}$, $\text{distance}\{F^t(q), W^{(c)}\} \rightarrow 0$, as $|t| \rightarrow \infty$. We call \mathcal{H} a transversal homoclinic tube asymptotic to $W^{(c)}$ under the flow F^t if the intersection between $W^{(cu)}$ and $W^{(cs)}$ is transversal at \mathcal{H} . Let Σ be an Poincaré section which intersects \mathcal{H} transversally, and P is the Poincaré map induced by the flow F^t ; then $\mathcal{H} \cap \Sigma$ is called a transversal homoclinic tube under the Poincaré map P .

THEOREM 2.5 (Homoclinic Tube Theorem). *There exist a positive constant $\varepsilon_0 > 0$ and a region \mathcal{E} for $(\alpha_1, \alpha_2, \omega)$, such that for any $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$ and any $(\alpha_1, \alpha_2, \omega) \in \mathcal{E}$, there exists a codimension 2 transversal homoclinic tube asymptotic to a codimension 2 center manifold $W^{(c)}$.*

For a complete proof of this theorem, see [13].

2.2. Chaos in Perturbed Soliton Lattices. For this category, we chose the candidate to be the perturbed cubic nonlinear Schrödinger lattice.

2.2.1. *Dissipative Perturbations.* In a series of three papers [11] [23] [24], we proved the existence of chaos in the discrete cubic nonlinear Schrödinger equation under a concrete dissipative perturbation.

We study the perturbed discrete cubic nonlinear Schrödinger equation

$$(2.5) \quad \begin{aligned} i\dot{q}_n &= \frac{1}{h^2} \left[q_{n+1} - 2q_n + q_{n-1} \right] + |q_n|^2 (q_{n+1} + q_{n-1}) - 2\omega^2 q_n \\ &+ i\varepsilon \left[-\alpha q_n + \frac{\beta}{h^2} (q_{n+1} - 2q_n + q_{n-1}) + \Gamma \right], \end{aligned}$$

under even periodic boundary conditions ($q_{N-n} = q_n$) and ($q_{n+N} = q_n$) for arbitrary N ; where $i = \sqrt{-1}$, q_n 's are complex variables, $h = 1/N$, $(\omega, \alpha, \beta, \Gamma)$ are positive constants, ε is the positive perturbation parameter.

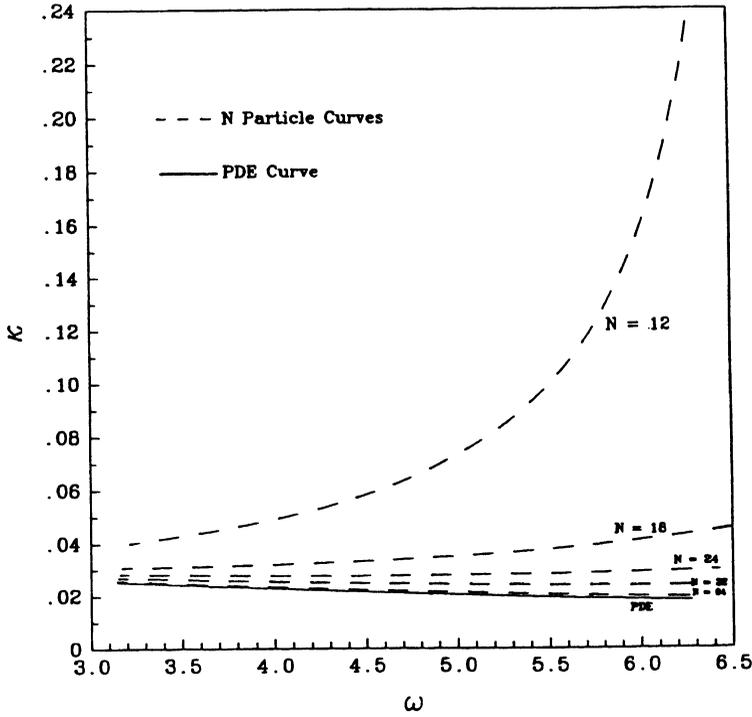
Denote by Σ_N ($N \geq 7$) the external parameter space,

$$\begin{aligned} \Sigma_N &= \left\{ (\omega, \alpha, \beta, \Gamma) \mid \omega \in (N \tan \frac{\pi}{N}, N \tan \frac{2\pi}{N}), \right. \\ &\quad \Gamma \in (0, 1), \alpha \in (0, \alpha_0), \beta \in (0, \beta_0); \\ &\quad \left. \text{where } \alpha_0 \text{ and } \beta_0 \text{ are any fixed positive numbers.} \right\} \end{aligned}$$

THEOREM 2.6. *For any N ($7 \leq N < \infty$), there exists a positive number ε_0 , such that for any $\varepsilon \in (0, \varepsilon_0)$, there exists a codimension 1 submanifold E_ε in Σ_N ; for any external parameters $(\omega, \alpha, \beta, \Gamma)$ on E_ε , there exists a homoclinic orbit asymptotic to a fixed point q_ε . The submanifold E_ε is in an $O(\varepsilon^\nu)$ neighborhood of the hyperplane $\beta = \kappa \alpha$, where $\kappa = \kappa(\omega; N)$ is shown in Figures 2 and 3, $\nu = 1/2 - \delta_0$, $0 < \delta_0 \ll 1/2$.*

REMARK 2.7. In the cases ($3 \leq N \leq 6$), κ is always negative as shown in Figure 3. Since we require both dissipation parameters α and β to be positive, the relation $\beta = \kappa \alpha$ shows that the existence of homoclinic orbits violates this positivity. For $N \geq 7$, κ can be positive as shown in Figure 2. When N is even and ≥ 7 , there is in fact a pair of homoclinic orbits asymptotic to a fixed point q_ε at the same values

Comparison of N Particle Curves
to PDE Curve, κ vs ω



N Particle Curve, κ vs ω
($N > N_c$)

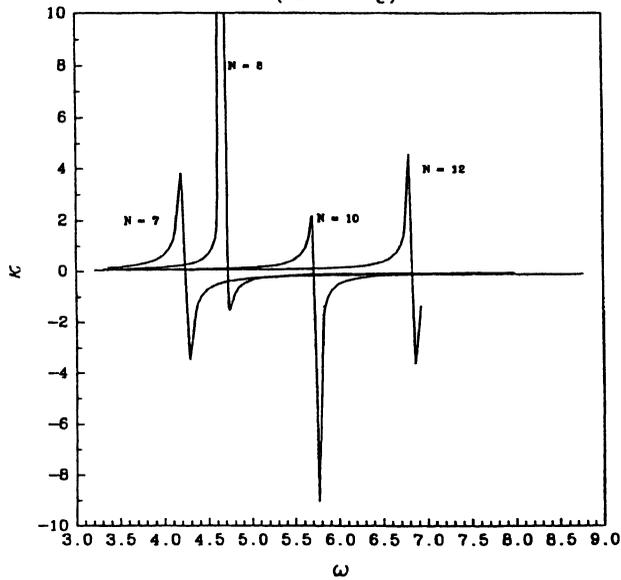


FIGURE 2. The curve of $\kappa = \kappa(\omega; N)$.

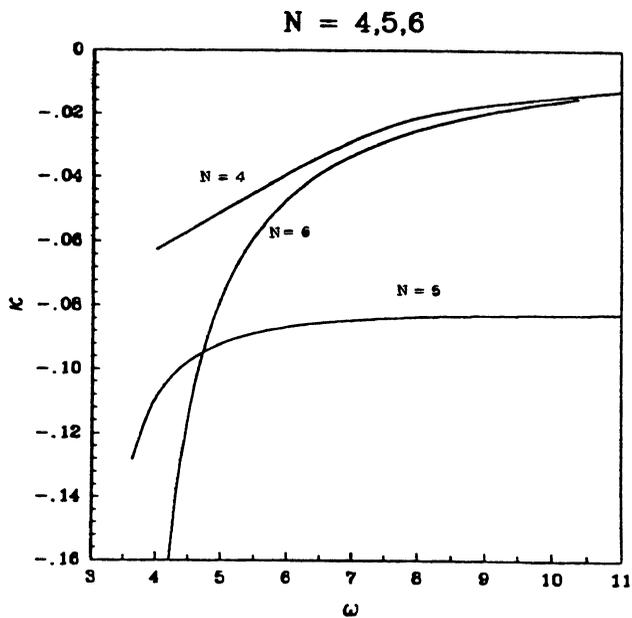
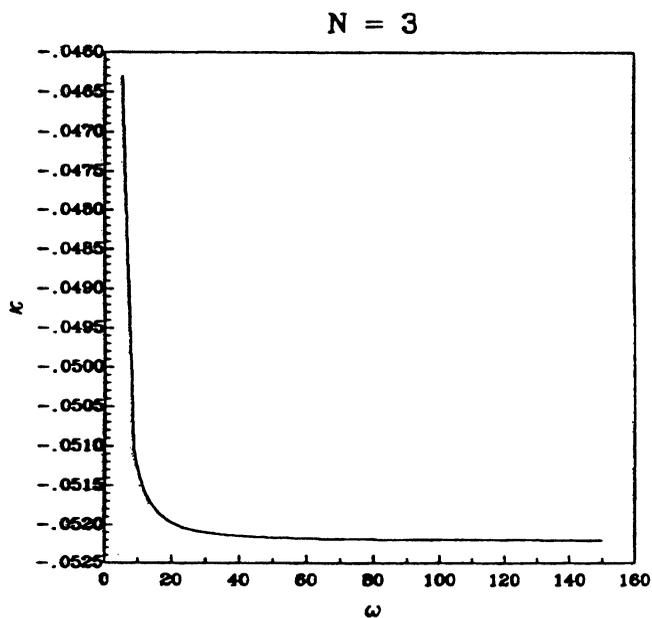


FIGURE 3. The curve of $\kappa = \kappa(\omega; N)$.

of the external parameters; since for even N , we have the symmetry: If $q_n = f(n, t)$ solves (2.5), then $q_n = f(n + N/2, t)$ also solves (2.5). When N is odd and ≥ 7 , the study can not guarantee that two homoclinic orbits exist at the same value of the external parameters.

For the complete proof of this theorem, see [23].

THEOREM 2.8 (Horseshoe Theorem). *Under certain generic assumptions for the perturbed discrete nonlinear Schrödinger system (2.5), there exists a compact Cantor subset Λ of Σ (a Poincaré section transversal to the homoclinic orbit), Λ consists of points, and is invariant under P (the Poincaré map induced by the flow on Σ). P restricted to Λ , is topologically conjugate to the shift automorphism χ on four symbols $1, 2, -1, -2$. That is, there exists a homeomorphism*

$$\phi : \mathcal{W} \mapsto \Lambda,$$

where \mathcal{W} is the topological space of the four symbols, such that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{W} & \xrightarrow{\phi} & \Lambda \\ \chi \downarrow & & \downarrow P \\ \mathcal{W} & \xrightarrow{\phi} & \Lambda \end{array}$$

For the complete proof of the theorem, see [24].

The unperturbed homoclinic orbits for the discrete nonlinear Schrödinger equation

$$(2.7) \quad i\dot{q}_n = \frac{1}{h^2} \left[q_{n+1} - 2q_n + q_{n-1} \right] + |q_n|^2(q_{n+1} + q_{n-1}) - 2\omega^2 q_n,$$

was constructed through the Darboux transformations which will be presented below in details. The discrete nonlinear Schrödinger equation is associated with the following discrete Zakharov-Shabat system [2]:

$$(2.8) \quad \varphi_{n+1} = L_n^{(z)} \varphi_n,$$

$$(2.9) \quad \dot{\varphi}_n = B_n^{(z)} \varphi_n,$$

where

$$L_n^{(z)} \equiv \begin{pmatrix} z & ihq_n \\ ih\bar{q}_n & 1/z \end{pmatrix},$$

$$B_n^{(z)} \equiv \frac{i}{h^2} \begin{pmatrix} 1 - z^2 + 2i\lambda h - h^2 q_n \bar{q}_{n-1} + \omega^2 h^2 & -zihq_n + (1/z)ihq_{n-1} \\ -izh\bar{q}_{n-1} + (1/z)ih\bar{q}_n & 1/z^2 - 1 + 2i\lambda h + h^2 \bar{q}_n q_{n-1} - \omega^2 h^2 \end{pmatrix},$$

and where $z \equiv \exp(i\lambda h)$.

Fix a solution $q_n(t)$ of the system (2.7), for which the linear operator L_n has a double point z^d of geometric multiplicity 2, which is not on the unit circle. We denote two linearly independent solutions (Bloch functions) of the discrete Lax pair (2.8;2.9) at $z = z^d$ by (ϕ_n^+, ϕ_n^-) . Thus, a general solution of the discrete Lax pair (2.8;2.9) at $(q_n(t), z^d)$ is given by

$$\phi_n(t; z^d, c) = \phi_n^+ + c\phi_n^-,$$

where c is a complex parameter called Bäcklund parameter. We use ϕ_n to define a transformation matrix Γ_n by

$$\Gamma_n = \begin{pmatrix} z + (1/z)a_n & b_n \\ c_n & -1/z + zd_n \end{pmatrix},$$

where,

$$\begin{aligned} a_n &= \frac{z^d}{(\bar{z}^d)^2 \Delta_n} \left[|\phi_{n2}|^2 + |z^d|^2 |\phi_{n1}|^2 \right], \\ d_n &= -\frac{1}{z^d \Delta_n} \left[|\phi_{n2}|^2 + |z^d|^2 |\phi_{n1}|^2 \right], \\ b_n &= \frac{|z^d|^4 - 1}{(\bar{z}^d)^2 \Delta_n} \phi_{n1} \bar{\phi}_{n2}, \\ c_n &= \frac{|z^d|^4 - 1}{z^d \bar{z}^d \Delta_n} \bar{\phi}_{n1} \phi_{n2}, \\ \Delta_n &= -\frac{1}{\bar{z}^d} \left[|\phi_{n1}|^2 + |z^d|^2 |\phi_{n2}|^2 \right]. \end{aligned}$$

Then we define Q_n and Ψ_n by

$$(2.10) \quad Q_n \equiv \frac{i}{h} b_{n+1} - a_{n+1} q_n$$

and

$$(2.11) \quad \Psi_n(t; z) \equiv \Gamma_n(z; z^d; \phi_n) \psi_n(t; z)$$

where ψ_n solves the discrete Lax pair (2.8;2.9) at $(q_n(t), z)$. Formulas (2.10) and (2.11) are the Bäcklund-Darboux transformations for the potential and eigenfunctions, respectively. We have the following theorem [11].

THEOREM 2.9 (Bäcklund-Darboux Transformations). *Let $q_n(t)$ denote a solution of the system (2.7), for which the linear operator L_n has a double point z^d of geometric multiplicity 2, which is not on the unit circle and which is associated with an instability. We denote two linearly independent solutions (Bloch functions) of the discrete Lax pair (2.8;2.9) at (q_n, z^d) by (ϕ_n^+, ϕ_n^-) . We define $Q_n(t)$ and $\Psi_n(t; z)$ by (2.10) and (2.11). Then*

- (1) $Q_n(t)$ is also a solution of the system (2.7). (The evenness of Q_n can be guaranteed by choosing the complex Bäcklund parameter c to lie on an certain curve.)
- (2) $\Psi_n(t; z)$ solves the discrete Lax pair (2.8;2.9) at $(Q_n(t), z)$.
- (3) $\Delta(z; Q_n) = \Delta(z; q_n)$, for all $z \in C$, where Δ is the Floquet discriminant.
- (4) $Q_n(t)$ is homoclinic to $q_n(t)$ in the sense that $Q_n(t) \rightarrow e^{i\theta_{\pm}} q_n(t)$, exponentially as $\exp(-\sigma|t|)$ as $t \rightarrow \pm\infty$. Here θ_{\pm} are the phase shifts, σ is a nonvanishing growth rate associated to the double point z^d , and explicit formulas can be developed for this growth rate and for the phase shifts θ_{\pm} .

Next we consider a concrete example. Let

$$(2.12) \quad q_n = q, \quad \forall n; \quad q = a \exp\{-2i[(a^2 - \omega^2)t + i\gamma]\},$$

where $N \tan \frac{\pi}{N} < a < N \tan \frac{2\pi}{N}$ for $N > 3$, $3 \tan \frac{\pi}{3} < a < \infty$ for $N = 3$. Then Q_n defined in (2.10) has the explicit representation:

$$(2.13) \quad Q_n \equiv Q_n(t; N, \omega, \gamma, r, \pm) = q \left[\frac{G}{H_n} - 1 \right],$$

where,

$$\begin{aligned} G &= 1 + \cos 2P - i \sin 2P \tanh \tau, \\ H_n &= 1 \pm \frac{1}{\cos \vartheta} \sin P \operatorname{sech} \tau \cos 2n\vartheta, \end{aligned}$$

$$\tau = 4N^2 \sqrt{\rho} \sin \vartheta \sqrt{\rho \cos^2 \vartheta - 1} t + r,$$

where r is a real parameter. Furthermore,

$$P = \arctan \frac{\sqrt{\rho \cos^2 \vartheta - 1}}{\sqrt{\rho} \sin \vartheta},$$

$$\vartheta = \frac{\pi}{N}, \quad \rho = 1 + \frac{|q|^2}{N^2}.$$

As $\tau \rightarrow \pm\infty$, $Q_n \rightarrow qe^{\mp i2P}$. Therefore, Q_n is homoclinic to the circle $|q_n| = a$, and heteroclinic to points on the circle which are separated in phase of $-4P$.

2.2.2. *Hamiltonian Perturbations.* In the paper [12], the author studied the discrete nonlinear Schrödinger equation under Hamiltonian perturbations:

$$(2.14) \quad i\dot{q}_n = \frac{1}{h^2} [q_{n+1} - 2q_n + q_{n-1}] + |q_n|^2 (q_{n+1} + q_{n-1}) - 2\omega^2 q_n$$

$$+ \varepsilon \left\{ [\alpha_1 (q_n + \bar{q}_n) + \alpha_2 (q_n^2 + \bar{q}_n^2)] q_n + [\alpha_1 + 2\alpha_2 \bar{q}_n] \frac{\rho_n}{h^2} \ln \rho_n \right\},$$

where $i = \sqrt{-1}$, q_n 's are complex variables, $n \in Z$, $(\omega, \alpha_1, \alpha_2)$ are real constants, ε is the perturbation parameter, h is the step size, $h = 1/N$, $N \geq 3$ is an integer, $\rho_n = 1 + h^2 |q_n|^2$, and $q_{n+N} = q_n$, $q_{-n} = q_n$. The system (2.14) can be written in the Hamiltonian form:

$$i\dot{q}_n = \rho_n \frac{\partial H}{\partial \bar{q}_n},$$

where $H = H_0 + \varepsilon H_1$,

$$H_0 = \frac{1}{h^2} \sum_{n=0}^{N-1} [\bar{q}_n (q_{n+1} + q_{n-1}) - \frac{2}{h^2} (1 + \omega^2 h^2) \ln \rho_n],$$

$$H_1 = \frac{1}{h^2} \sum_{n=0}^{N-1} [\alpha_1 (q_n + \bar{q}_n) + \alpha_2 (q_n^2 + \bar{q}_n^2)] \ln \rho_n.$$

THEOREM 2.10 (Homoclinic Tube Theorem). *There exist a positive constant $\varepsilon_0 > 0$ and a region \mathcal{E} for $(\alpha_1, \alpha_2, \omega)$, such that for any $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$ and any $(\alpha_1, \alpha_2, \omega) \in \mathcal{E}$, there exists a codimension 2 transversal homoclinic tube asymptotic to a codimension 2 center manifold $W^{(c)}$.*

For a complete proof of this theorem, see [12].

2.3. Chaos in Perturbed $(1 + n)$ -Dimensional Soliton Equations ($n \geq 2$). For this category of the perturbed soliton equations, we chose the candidate to be the perturbed Davey-Stewartson II equations. The Davey-Stewartson II equations describe nearly one-dimensional water surface wave train [6]. There have been a lot of studies on the inverse scattering transforms for this set of equations [1] [3]. The inverse scattering transforms for $(1+n)$ -dimensional soliton equations ($n \geq 2$) are substantially different from those for $(1+1)$ -dimensional soliton equations and soliton lattices. In fact, the Davey-Stewartson II equations possess finite-time singularities [27]. For the perturbed Davey-Stewartson II equations, the theory on chaos is largely unfinished. So far, its Melnikov theory has been successfully built.

Although the inverse spectral theory for the DSII equations is very different from those for $(1+1)$ -dimensional soliton equations and there is no Floquet spectral theory, its Bäcklund-Darboux transformation is as simple as those for $(1+1)$ -dimensional soliton equations, e.g. the cubic nonlinear Schrödinger equation. These

Bäcklund-Darboux transformations are successfully utilized to construct heteroclinic orbits of Davey-Stewartson II equations through an elegant iteration of the transformations. In [22], we successfully built Melnikov vectors for the focusing cubic nonlinear Schrödinger equation with the gradients of the invariants F_j defined through the Floquet discriminants evaluated at critical spectral points. The invariants F_j 's Poisson commute with the Hamiltonian, and their gradients decay exponentially as time approaches positive and negative infinities – these two properties are crucial in deriving and evaluating Melnikov functions. Since there is no Floquet discriminant for Davey-Stewartson equations (in contrast to nonlinear Schrödinger equations [22]), the Melnikov vectors here are built with the novel idea of replacing the gradients of Floquet discriminants by quadratic products of Bloch functions. Such Melnikov vectors still maintain the properties of Poisson commuting with the gradient of the Hamiltonian and exponential decay as time approaches positive and negative infinities. This solves the problem of building Melnikov vectors for Davey-Stewartson equations without using the gradients of Floquet discriminant. Melnikov functions for perturbed Davey-Stewartson II equations evaluated on the above heteroclinic orbits are built.

2.3.1. *Darboux Transformations.* First we study the Darboux transformations for the Davey-Stewartson II (DSII) equations:

$$(2.15) \quad \begin{aligned} i\partial_t q &= [\partial_x^2 - \partial_y^2]q + [2(|q|^2 - \omega^2) + u_y]q, \\ [\partial_x^2 + \partial_y^2]u &= -4\partial_y|q|^2; \end{aligned}$$

under periodic boundary conditions $q(t, x + l_x, y) = q(t, x, y + l_y) = q(t, x, y)$, where q and u are a complex-valued and a real-valued functions of three variables (t, x, y) . To simplify the study, we may also pose even conditions in both x and y . The DSII equations are associated with a Lax pair and a congruent Lax pair. The Lax pair is:

$$(2.16) \quad L\psi = \lambda\psi,$$

$$(2.17) \quad \partial_t\psi = A\psi,$$

where $\psi = (\psi_1, \psi_2)^T$, and

$$L = \begin{pmatrix} D^- & q \\ r & D^+ \end{pmatrix},$$

$$A = i \left[2 \begin{pmatrix} -\partial_x^2 & q\partial_x \\ r\partial_x & \partial_x^2 \end{pmatrix} + \begin{pmatrix} r_1 & (D^+q) \\ -(D^-r) & r_2 \end{pmatrix} \right].$$

Here we denote by

$$(2.18) \quad D^+ = \alpha\partial_y + \partial_x, \quad D^- = \alpha\partial_y - \partial_x.$$

where $r = \bar{q}$, $\alpha^2 = -1$,

$$r_1 = \frac{1}{2}[-U + iV], \quad r_2 = \frac{1}{2}[U + iV], \quad U = 2(|q|^2 - \omega^2) + u_y.$$

The congruent Lax pair is:

$$(2.19) \quad \hat{L}\hat{\psi} = \lambda\hat{\psi},$$

$$(2.20) \quad \partial_t\hat{\psi} = \hat{A}\hat{\psi},$$

where $\hat{\psi} = (\hat{\psi}_1, \hat{\psi}_2)^T$, and

$$\hat{L} = \begin{pmatrix} -D^+ & q \\ r & -D^- \end{pmatrix},$$

$$\hat{A} = i \left[2 \begin{pmatrix} -\partial_x^2 & q\partial_x \\ r\partial_x & \partial_x^2 \end{pmatrix} + \begin{pmatrix} -r_2 & -(D^-q) \\ (D^+r) & -r_1 \end{pmatrix} \right].$$

Let $(q, r = \bar{q}, r_1, r_2)$ be a solution to the DSII equation, and let λ_0 be any value of λ . Denote by $\psi = (\psi_1, \psi_2)^T$ the eigenfunction solving the Lax pair (2.16, 2.17) at $(q, r = \bar{q}, r_1, r_2; \lambda_0)$. Define the matrix operator:

$$\Gamma = \begin{bmatrix} \wedge + a & b \\ c & \wedge + d \end{bmatrix},$$

where $\wedge = \alpha\partial_y - \lambda$, and a, b, c, d are functions defined as:

$$a = \frac{1}{\Delta} [\psi_2 \wedge_2 \bar{\psi}_2 + \beta \bar{\psi}_1 \wedge_1 \psi_1],$$

$$b = \frac{1}{\Delta} [\bar{\psi}_2 \wedge_1 \psi_1 - \psi_1 \wedge_2 \bar{\psi}_2],$$

$$c = \frac{\beta}{\Delta} [\bar{\psi}_1 \wedge_1 \psi_2 - \psi_2 \wedge_2 \bar{\psi}_1],$$

$$d = \frac{1}{\Delta} [\bar{\psi}_2 \wedge_1 \psi_2 + \beta \psi_1 \wedge_2 \bar{\psi}_1],$$

in which $\wedge_1 = \alpha\partial_y - \lambda_0$, $\wedge_2 = \alpha\partial_y + \bar{\lambda}_0$, and

$$\Delta = -[\beta|\psi_1|^2 + |\psi_2|^2].$$

Define a transformation as follows:

$$\begin{cases} (q, r = \beta\bar{q}, r_1, r_2) & \rightarrow (Q, R, R_1, R_2), \\ \phi & \rightarrow \Phi; \end{cases}$$

$$Q = q - 2b,$$

$$R = \beta\bar{q} - 2c,$$

$$(2.21) \quad R_1 = r_1 + 2(D^+a),$$

$$R_2 = r_2 - 2(D^-d),$$

$$\Phi = \Gamma\phi;$$

where ϕ is an eigenfunction solving the Lax pair (2.16, 2.17) at $(q, r = \bar{q}, r_1, r_2; \lambda)$, D^+ and D^- are defined in (2.18),

THEOREM 2.11 ([15]). *The transformation (2.21) is a Bäcklund-Darboux transformation. That is, the functions $(Q, R = \bar{Q}, R_1, R_2)$ defined through the transformation (2.21) are also a solution to the Davey-Stewartson II equations. The function Φ defined through the transformation (2.21) solves the Lax pair (2.16, 2.17) at $(Q, R = \bar{Q}, R_1, R_2; \lambda)$.*

A concrete example with two iterations of the Darboux transformations has been worked out in [15].

2.3.2. *Melnikov Vectors.* The DSII equations can be put into the Hamiltonian form,

$$(2.22) \quad \begin{cases} iq_t &= \delta H / \delta \bar{q} , \\ i\bar{q}_t &= -\delta H / \delta q , \end{cases}$$

where

$$H = \int_0^{l_y} \int_0^{l_x} [|q_y|^2 - |q_x|^2 + \frac{1}{2}(r_2 - r_1) |q|^2] dx dy .$$

Let $\psi = (\psi_1, \psi_2)^T$ be an eigenfunction solving the Lax pair (2.16, 2.17), and $\hat{\psi} = (\hat{\psi}_1, \hat{\psi}_2)^T$ be an eigenfunction solving the corresponding congruent Lax pair (2.19, 2.20); then

LEMMA 2.12. *The inner product of the vector*

$$\mathcal{U} = \begin{pmatrix} \psi_2 \hat{\psi}_2 \\ \psi_1 \hat{\psi}_1 \end{pmatrix}^- + S \begin{pmatrix} \psi_2 \hat{\psi}_2 \\ \psi_1 \hat{\psi}_1 \end{pmatrix} ,$$

where $S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, with the vector field $J\nabla H$ given by the right hand side of (2.22) vanishes,

$$\langle \mathcal{U}, J\nabla H \rangle = 0 .$$

where

$$\langle f, g \rangle = \int_0^{l_y} \int_0^{l_x} \{ \bar{f}_1 g_1 + \bar{f}_2 g_2 \} dx dy .$$

and

$$J = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

Consider the perturbed DSII equations

$$(2.24) \quad \begin{cases} i\partial_t q = [\partial_x^2 - \partial_y^2]q + [2(|q|^2 - \omega^2) + u_y]q + \varepsilon i f , \\ [\partial_x^2 + \partial_y^2]u = -4\partial_y |q|^2 , \end{cases}$$

where f is the perturbation which can depend on q and \bar{q} and their derivatives and t, x and y . Let $\vec{G} = (f, \bar{f})^T$. Then the Melnikov function has the expression,

$$(2.25) \quad \begin{aligned} M &= \int_{-\infty}^{\infty} \langle \mathcal{U}, \vec{G} \rangle dt \\ &= 2 \int_{-\infty}^{\infty} \int_0^{l_y} \int_0^{l_x} Re \left\{ (\psi_2 \hat{\psi}_2) f + (\psi_1 \hat{\psi}_1) \bar{f} \right\} dx dy dt , \end{aligned}$$

where the integrand is evaluated on an unperturbed heteroclinic orbit obtained through the Bäcklund-Darboux transformations given in Theorem 2.11. A concrete example has been worked out in [15].

3. Two-Dimensional Euler Equations

One of the most important implications of chaos theory of partial differential equations in theoretical physics will be on the study of turbulence. For that goal, the author choose the 2D Navier-Stokes equations under periodic boundary conditions to begin a dynamical system study.

$$(3.1) \quad \begin{aligned} \frac{\partial \Omega}{\partial t} &= -u \frac{\partial \Omega}{\partial x} - v \frac{\partial \Omega}{\partial y} + \varepsilon \left[\Delta \Omega + f \right], \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0; \end{aligned}$$

under periodic boundary conditions in both x and y directions with period 2π , where Ω is vorticity, u and v are respectively velocity components along x and y directions, $\varepsilon = 1/Re$, and f is the body force. When $\varepsilon = 0$, we have the 2D Euler equations,

$$(3.2) \quad \begin{aligned} \frac{\partial \Omega}{\partial t} &= -u \frac{\partial \Omega}{\partial x} - v \frac{\partial \Omega}{\partial y}, \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0. \end{aligned}$$

The relation between vorticity Ω and stream function Ψ is,

$$\Omega = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \Delta \Psi,$$

where the stream function Ψ is defined by,

$$u = -\frac{\partial \Psi}{\partial y}, \quad v = \frac{\partial \Psi}{\partial x}.$$

3.1. Lax Pair and Darboux Transformation. The main breakthrough in this project is the discovery of the Lax pair for 2D Euler equation [17]. The philosophical significance of the existence of a Lax pair for 2D Euler equation is beyond the particular project undertaken here. If one defines integrability of an equation by the existence of a Lax pair, then 2D Euler equation is integrable. More importantly, 2D Navier-Stokes equation at high Reynolds numbers is a near integrable system. Such a point of view changes our old ideology on Euler and Navier-Stokes equations.

Starting from Lax pairs, homoclinic structures can be constructed through Darboux transformations [15]. Indeed, in [25], the Darboux transformation for the Lax pair of 2D Euler equation has been found. Our general program is to first identify the figure eight structures of 2D Euler equation, and then study their consequence in 2D Navier-Stokes equation. The high Reynolds number 2D Navier-Stokes equation is viewed as a singular perturbation of the 2D Euler equation through the perturbation $\varepsilon \Delta$, where $\varepsilon = 1/Re$ is the inverse of the Reynolds number. As mentioned above, singular perturbations have been investigated for nonlinear Schrödinger equations.

We consider the 2D Euler equation,

$$(3.3) \quad \frac{\partial \Omega}{\partial t} + \{\Psi, \Omega\} = 0,$$

where the bracket $\{ , \}$ is defined as

$$\{f, g\} = (\partial_x f)(\partial_y g) - (\partial_y f)(\partial_x g) , \text{ and } \Omega = \Delta \Psi .$$

THEOREM 3.1 ([17]). *The Lax pair of the 2D Euler equation (3.3) is given as*

$$(3.4) \quad \begin{cases} L\varphi = \lambda\varphi , \\ \partial_t \varphi + A\varphi = 0 , \end{cases}$$

where

$$L\varphi = \{\Omega, \varphi\} , \quad A\varphi = \{\Psi, \varphi\} ,$$

and λ is a complex constant, and φ is a complex-valued function.

In [25], A Bäcklund-Darboux transformation is found for the above Lax pair. Consider the Lax pair (3.4) at $\lambda = 0$, i.e.

$$(3.5) \quad \{\Omega, p\} = 0 ,$$

$$(3.6) \quad \partial_t p + \{\Psi, p\} = 0 ,$$

where we replaced the notation φ by p .

THEOREM 3.2. *Let $f = f(t, x, y)$ be any fixed solution to the system (3.5, 3.6), we define the Gauge transform G_f :*

$$(3.7) \quad \tilde{p} = G_f p = \frac{1}{\Omega_x} [p_x - (\partial_x \ln f)p] ,$$

and the transforms of the potentials Ω and Ψ :

$$(3.8) \quad \tilde{\Psi} = \Psi + F , \quad \tilde{\Omega} = \Omega + \Delta F ,$$

where F is subject to the constraints

$$(3.9) \quad \{\Omega, \Delta F\} = 0 , \quad \{\Omega, F\} = 0 .$$

Then \tilde{p} solves the system (3.5, 3.6) at $(\tilde{\Omega}, \tilde{\Psi})$. Thus (3.7) and (3.8) form the Darboux transformation for the 2D Euler equation (3.3) and its Lax pair (3.5, 3.6).

3.2. Linearized 2D Euler Equations. Under the periodic boundary condition and requiring that both u and v have means zero,

$$\int_0^{2\pi} \int_0^{2\pi} u \, dx dy = \int_0^{2\pi} \int_0^{2\pi} v \, dx dy = 0 ,$$

expanding Ω into Fourier series, $\Omega = \sum_{k \in Z^2 \setminus \{0\}} \omega_k e^{ik \cdot X}$, where $\omega_{-k} = \overline{\omega_k}$, $k = (k_1, k_2)$, $X = (x, y)$, the system (3.2) can be rewritten as the following kinetic system,

$$(3.10) \quad \dot{\omega}_k = \sum_{k=p+q} A(p, q) \omega_p \omega_q ,$$

where $A(p, q)$ is given by,

$$(3.11) \quad \begin{aligned} A(p, q) &= \frac{1}{2} [|q|^{-2} - |p|^{-2}] (p_1 q_2 - p_2 q_1) \\ &= \frac{1}{2} [|q|^{-2} - |p|^{-2}] \begin{vmatrix} p_1 & q_1 \\ p_2 & q_2 \end{vmatrix} , \end{aligned}$$

where $|q|^2 = q_1^2 + q_2^2$ for $q = (q_1, q_2)$, similarly for p . To understand the hyperbolic structures of the 2D Euler equations, we first investigate the linearized 2D Euler

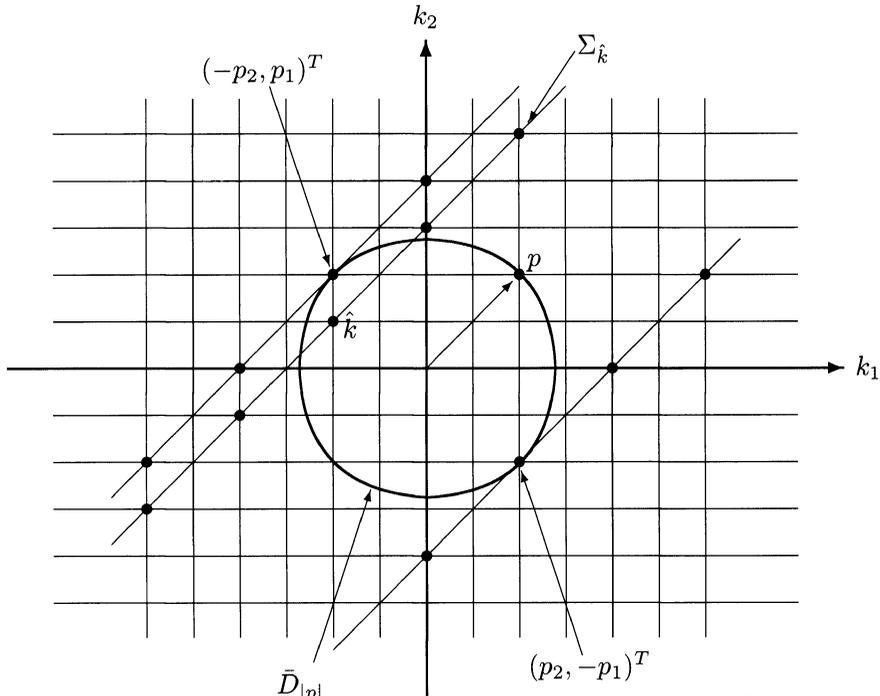


FIGURE 4. An illustration of the classes $\Sigma_{\hat{k}}$ and the disk $\bar{D}_{|p|}$.

equations at a stationary solution. Denote $\{\omega_k\}_{k \in \mathbb{Z}^2 / \{0\}}$ by ω . Consider the simple fixed point ω^* :

$$(3.12) \quad \omega_p^* = \Gamma, \quad \omega_k^* = 0, \text{ if } k \neq p \text{ or } -p,$$

of the 2D Euler equation (3.10), where Γ is an arbitrary complex constant. The linearized two-dimensional Euler equation at ω^* is given by,

$$(3.13) \quad \dot{\omega}_k = A(p, k - p) \Gamma \omega_{k-p} + A(-p, k + p) \bar{\Gamma} \omega_{k+p}.$$

DEFINITION 3.3 (Classes). For any $\hat{k} \in \mathbb{Z}^2 / \{0\}$, we define the class $\Sigma_{\hat{k}}$ to be the subset of $\mathbb{Z}^2 / \{0\}$:

$$\Sigma_{\hat{k}} = \left\{ \hat{k} + np \in \mathbb{Z}^2 / \{0\} \mid n \in \mathbb{Z}, p \text{ is specified in (3.12)} \right\}.$$

See Figure 4 for an illustration of the classes. According to the classification defined in Definition 3.3, the linearized two-dimensional Euler equation (3.13) decouples into infinite many invariant subsystems:

$$(3.14) \quad \begin{aligned} \dot{\omega}_{\hat{k}+np} &= A(p, \hat{k} + (n - 1)p) \Gamma \omega_{\hat{k}+(n-1)p} \\ &+ A(-p, \hat{k} + (n + 1)p) \bar{\Gamma} \omega_{\hat{k}+(n+1)p}. \end{aligned}$$

DEFINITION 3.4 (The Disk). The disk of radius $|p|$ in $\mathbb{Z}^2 / \{0\}$, denoted by $D_{|p|}$, is defined as

$$D_{|p|} = \left\{ k \in \mathbb{Z}^2 / \{0\} \mid |k| < |p| \right\}.$$

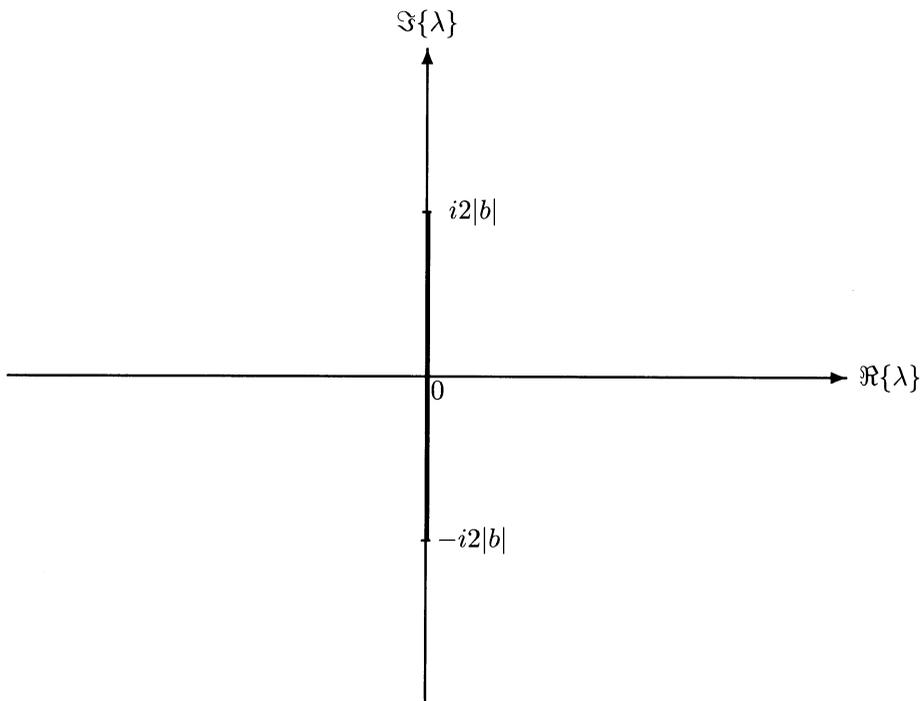


FIGURE 5. The spectrum of \mathcal{L}_A in case 1.

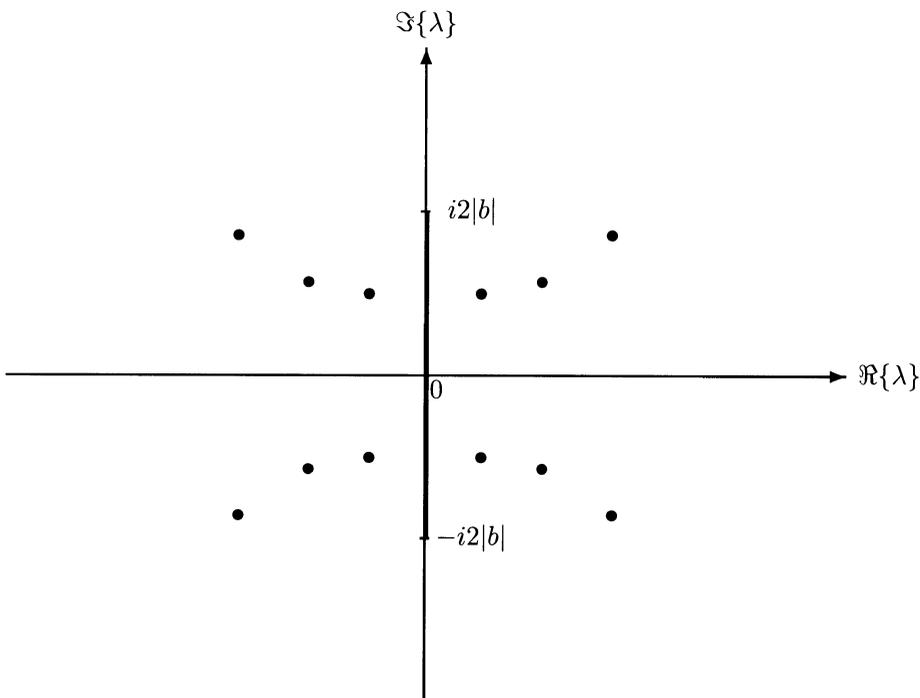


FIGURE 6. The spectrum of \mathcal{L}_A in case 2.

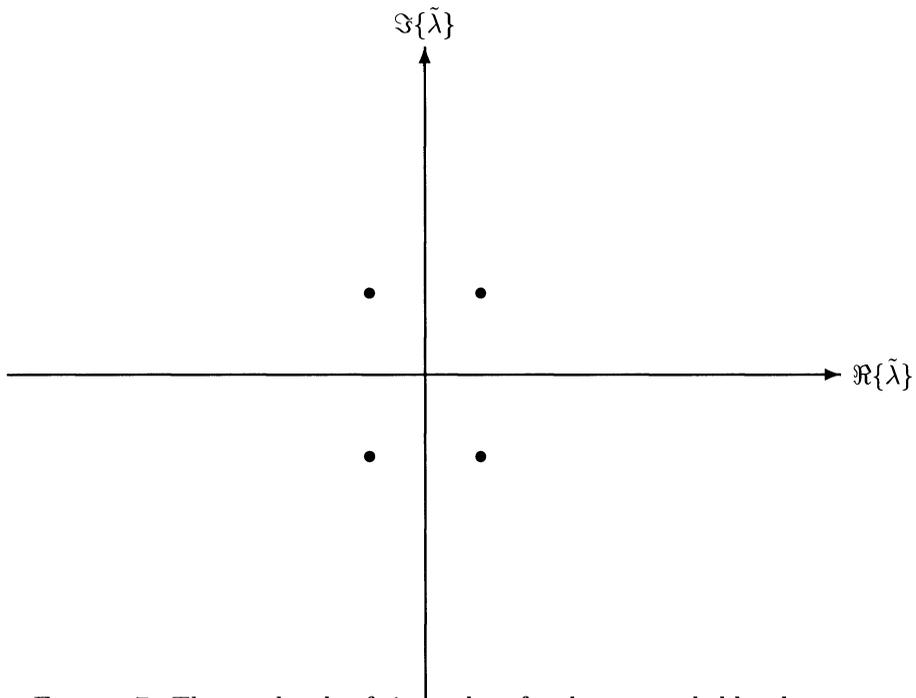


FIGURE 7. The quadruple of eigenvalues for the system led by the class $\Sigma_{\hat{k}}$ labeled by $\hat{k} = (1, 0)^T$, when $p = (1, 1)^T$.

The closure of $D_{|p|}$, denoted by $\bar{D}_{|p|}$, is defined as

$$\bar{D}_{|p|} = \left\{ k \in \mathbb{Z}^2 / \{0\} \mid |k| \leq |p| \right\}.$$

THEOREM 3.5 (Unstable Disk Theorem). *If $\Sigma_{\hat{k}} \cap \bar{D}_{|p|} = \emptyset$, then the invariant subsystem (3.14) is Liapunov stable for all $t \in \mathbb{R}$, in fact,*

$$\sum_{n \in \mathbb{Z}} \left| \omega_{\hat{k}+np}(t) \right|^2 \leq \sigma \sum_{n \in \mathbb{Z}} \left| \omega_{\hat{k}+np}(0) \right|^2, \quad \forall t \in \mathbb{R},$$

where

$$\sigma = \left[\max_{n \in \mathbb{Z}} \{-\rho_n\} \right] \left[\min_{n \in \mathbb{Z}} \{-\rho_n\} \right]^{-1}, \quad 0 < \sigma < \infty.$$

THEOREM 3.6. *The eigenvalues of the linear system (3.14) are of four types: real pairs $(c, -c)$, purely imaginary pairs $(id, -id)$, quadruples $(\pm c \pm id)$, and zero eigenvalues.*

THEOREM 3.7 (The Spectral Theorem). (1) *If $\Sigma_{\hat{k}} \cap \bar{D}_{|p|} = \emptyset$, then the entire ℓ_2 spectrum of the linear operator \mathcal{L}_A (defined by the right-hand side of the invariant subsystem) is its continuous spectrum. See Figure 5.*
 (2) *If $\Sigma_{\hat{k}} \cap \bar{D}_{|p|} \neq \emptyset$, then the entire essential ℓ_2 spectrum of the linear operator \mathcal{L}_A is its continuous spectrum. That is, the residual spectrum of \mathcal{L}_A is empty, $\sigma_r(\mathcal{L}_A) = \emptyset$. The point spectrum of \mathcal{L}_A is symmetric with respect to both real and imaginary axes. See Figure 6.*

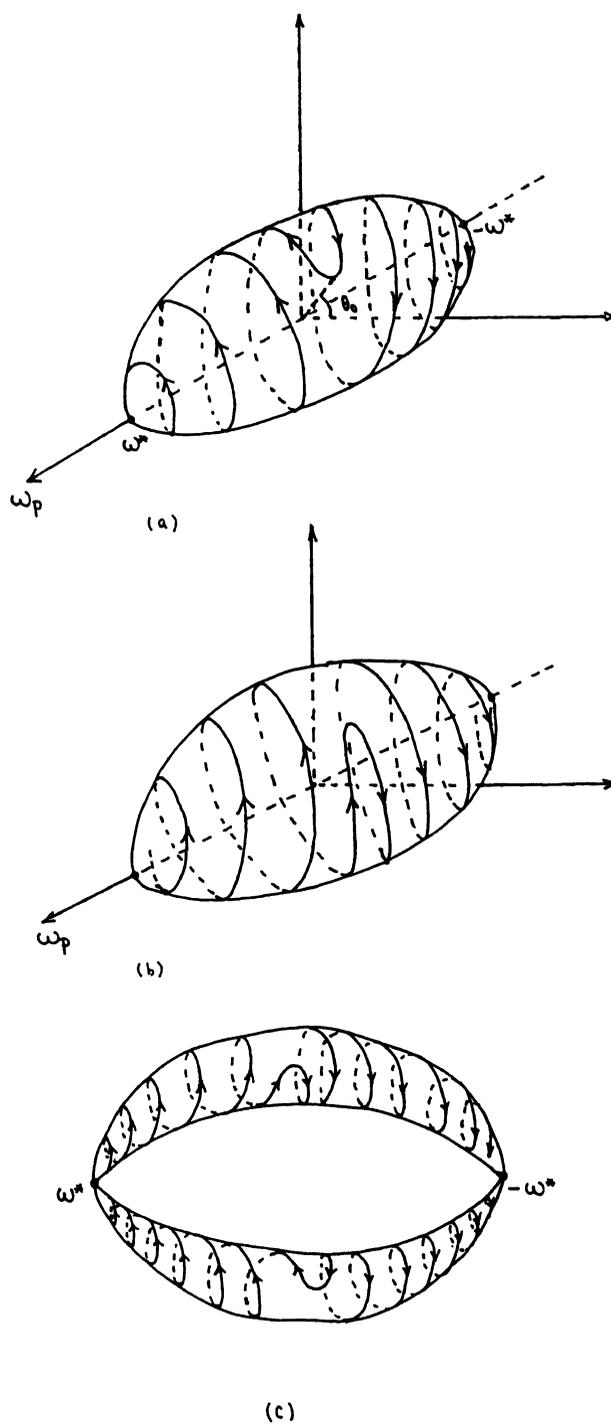


FIGURE 8. The heteroclinic orbits and unstable manifolds of the Galerkin truncation.

We can calculate the eigenvalues through continued fractions. Let $p = (1, 1)^T$, in this case, only one class $\Sigma_{\hat{k}}$ labeled by $\hat{k} = (1, 0)^T$ has no empty intersection with $\bar{D}_{|p|}$ (the other class labeled by $\hat{k} = (0, 1)^T$ gives the complex conjugate of the system led by the class labeled by $\hat{k} = (1, 0)^T$). For this class, there is no real eigenvalue. Numerical calculation through continued fractions gives the eigenvalue:

$$\tilde{\lambda} = 0.24822302478255 + i 0.35172076526520 .$$

Thus we have a quadruple of eigenvalues, see Figure 7 for an illustration. Denote by L the right hand side of (3.13), the spectral mapping theorem holds.

THEOREM 3.8 ([10]).

$$\sigma(e^{tL}) = e^{t\sigma(L)}, t \neq 0.$$

Moreover, the number of eigenvalues has a sharp upper bound. Let ζ denote the number of points $q \in Z^2/\{0\}$ that belong to the open disk of radius $|p|$, and such that q is not parallel to p .

THEOREM 3.9 ([10]). *The number of nonimaginary eigenvalues of L (counting the multiplicities) does not exceed 2ζ .*

3.3. Approximate Explicit Representations of the Hyperbolic Structures of 2D Euler Equations. From Figure 7, we see that the simple fixed point given by $p = (1, 1)$, has unstable eigenvalues. Our interest is to obtain representations of the corresponding hyperbolic structures for 2D Euler equations. In [18], through Galerkin truncation, we obtained the approximate explicit representation. Figure 8 shows the heteroclinic orbits and unstable manifolds of the Galerkin truncation.

4. Conclusion and Discussion

We have reported the status of chaos in nonlinear wave equations and of study on 2D Euler equations. In particular, we have summarized the most recent results on Lax pair and Darboux transformations for 2D Euler equations.

References

- [1] M.J. Ablowitz and P. A. Clarkson, *Solitons, Nonlinear Evolution Equations and Inverse Scattering*, London Mathematical Society Lecture Note Series 149, 1991.
- [2] M.J. Ablowitz and J. F. Ladik, A nonlinear difference scheme and inverse scattering, *Stud. Appl. Math.* **55** (1976), 213.
- [3] M.J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform*, SIAM, Philadelphia, 1981.
- [4] J. Bourgain, Construction of quasi-periodic solutions for Hamiltonian perturbations of linear equations and applications to nonlinear PDE, *International Mathematics Research Notes* **11** (1994), 1.
- [5] W. Craig and C. E. Wayne, Newton's method and periodic solutions of nonlinear wave equation, *Comm. Pure Appl. Math.* **46** (1993), 1409.
- [6] A. Davey and K. Stewartson, On three-dimensional packets of surface waves, *Proc. R. Soc. Lond.* **A338** (1974), 101.
- [7] S. M. Graff, On the construction of hyperbolic invariant tori for Hamiltonian systems, *J. Diff. Eqs.* **15** (1974), 1.
- [8] J. Guckenheimer and P. J. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, Applied Mathematical Sciences, vol.42, 1983.
- [9] S. B. Kuksin, *Nearly Integrable Infinite-Dimensional Hamiltonian Systems*, Springer-Verlag, Lecture Notes Math. 1556, 1993.

- [10] Y. Latushkin and Y. Li and M. Stanislavova, The spectrum of a linearized 2D Euler operator, *Submitted* (2001).
- [11] Y. Li, Backlund transformations and homoclinic structures for the NLS equation, *Phys. Letters A* **163** (1992), 181.
- [12] Y. Li, Homoclinic tubes and symbolic dynamics discrete nonlinear Schrödinger equation under Hamiltonian perturbations, *Submitted* (1998).
- [13] Y. Li, Homoclinic tubes in nonlinear Schrödinger equation under Hamiltonian perturbations, *Progress of Theoretical Physics* **101**, No. **3(4)** (1999), 559.
- [14] Y. Li, Smale horseshoes and symbolic dynamics in perturbed nonlinear Schrödinger equations, *Journal of Nonlinear Sciences* **9**, no.4 (1999), 363.
- [15] Y. Li, Bäcklund-Darboux transformations and Melnikov analysis for Davey-Stewartson II equations, *Journal of Nonlinear Sciences* **10**, no.1 (2000), 103.
- [16] Y. Li, On 2D Euler equations. I. On the energy-Casimir stabilities and the spectra for a linearized two dimensional Euler equation, *Journal of Mathematical Physics* **41**, no.2 (2000), 728.
- [17] Y. Li, A Lax pair for the 2D Euler equation, *Journal of Mathematical Physics* **42**, no.8 (2001), 3552.
- [18] Y. Li, On 2D Euler equations: part II. Lax pairs and homoclinic structures, *Submitted* (2001).
- [19] Y. Li, Persistent homoclinic orbits for nonlinear Schrödinger equation under singular perturbation, *Submitted* (2001).
- [20] Y. Li, Singularly perturbed vector and scalar nonlinear Schrödinger equation with persistent homoclinic orbits, *Studies in Applied Mathematics, to appear* (2001).
- [21] Y. Li et al., Persistent homoclinic orbits for a perturbed nonlinear Schrödinger equation, *Comm. Pure Appl. Math.* **XLIX** (1996), 1175.
- [22] Y. Li and D. W. McLaughlin, Morse and Melnikov functions for NLS pdes, *Comm. Math. Phys.* **162** (1994), 175.
- [23] Y. Li and D. W. McLaughlin, Homoclinic orbits and chaos in discretized perturbed NLS system, part I. homoclinic orbits, *Journal of Nonlinear Sciences* **7** (1997), 211.
- [24] Y. Li and S. Wiggins, Homoclinic orbits and chaos in discretized perturbed NLS system, part II. symbolic dynamics, *Journal of Nonlinear Sciences* **7** (1997), 315.
- [25] Y. Li and A. Yurov, Lax pairs and Darboux transformations for Euler equations, **Preprint, available at: <http://xxx.lanl.gov/abs/math.AP/0101214>, or <http://www.math.missouri.edu/~cli>** (2000).
- [26] A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion*, Springer-Verlag, Applied Mathematical Sciences, vol.38, 1983.
- [27] T. Ozawa, Exact blow-up solutions to the Cauchy problem for the Davey-Stewartson systems, *Proc. R. Soc. Lond.* **436** (1992), 345.
- [28] J. Poschel, A KAM-theorem for some nonlinear PDEs, *Ann. Scuola Norm. Sup. Pisa, Cl. Sci., IV Ser.* **15** **23** (1996), 119.
- [29] C. E. Wayne, Periodic and quasi-periodic solutions of nonlinear wave equations via KAM theory, *Commun. Math. Phys.* **127** (1990), 479.
- [30] C. Zeng, Erratum: Homoclinic orbits for a perturbed nonlinear Schrödinger equation, *Comm. Pure Appl. Math.* **53**, no.12 (2000), 1603.
- [31] C. Zeng, Homoclinic orbits for a perturbed nonlinear Schrödinger equation, *Comm. Pure Appl. Math.* **53**, no.10 (2000), 1222.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MISSOURI, COLUMBIA, MO 65211
 E-mail address: cli@math.missouri.edu

This page intentionally left blank

Multi-Soliton Complexes

Nail N. Akhmediev, Andrey A. Sukhorukov, and Adrian Ankiewicz

ABSTRACT. This paper reviews the latest advances in the area of multi-soliton complexes (MSCs). We present exact analytical solutions of coupled nonlinear Schrödinger equations, which describe multi-soliton complexes and their interactions on top of a background in media with self-focusing or self-defocusing Kerr-like nonlinearities. We present numerical examples illustrating the remarkable properties of MSCs, such as their reshaping after collisions. This occurs because the fundamental solitons composing an MSC can acquire different lateral shifts. We also obtain an accurate estimate for the peak intensities of stationary and interacting MSCs, by establishing a rigorous relationship between the eigenvalues of incoherently-coupled fundamental solitons and the range of admissible intensities.

1. Introduction

Dynamic nonlinear systems have properties which were initially surprising to scientists [FPU55, Fer65, Akh01]. The concept of ‘solitons’, first introduced in [ZK65], helped to demystify at least some of these surprises. The inverse scattering technique, developed later in a number of works [GGKM67, ZS71, AKNS74], gave scientists a powerful tool for understanding and for investigating the properties of nonlinear systems with an infinite number of degrees of freedom. The theory has far-reaching consequences which allow us not only to solve specific problems, but also to understand the situation qualitatively. For example, this was the case with the so called multi-soliton complexes (MSCs).

A multi-soliton complex is a self-localized state which is a nonlinear superposition of several fundamental solitons [AA00]. In optics, it can be a single beam or pulse created by a nonlinear superposition of fundamental solitons, where each has the same velocity. The nonlinear superposition can be either coherent or incoherent, in the sense that the phases of the separate solitons in the collection can be either related or independent. The solitons in the group may be bound together, or at least may stay close to each other, simply because they have the same initial speed. A large class of problems involve MSCs. They arise from diverse applications in optics. The first review on this subject [AA00] appeared more than a year ago. The

1991 *Mathematics Subject Classification.* Primary: 35Q51; Secondary 35Q55, 37K15.

Key words and phrases. Inverse Scattering, Solitons, Multi-Soliton Complexes.

The first author was supported in part by AROFE Grant # N62649-01-1-0002.

present article can be considered as a continuation of [AA00] which summarizes new material published since that time.

MSCs have common properties which are of interest in broad areas of physics. These include systems of fermions in one space dimension in the Hartree - Fock approximation, envelope solitons of random phase waves [Has75, Has77], multicomponent Bose-Einstein condensates at zero temperature [BV97], self-confinement of optical pulses in multimode glass fibers [Has80] and short pulses in multi-core optical fibers [BA95]. Gap solitons [ds94], Manakov solitons [Man73] and vector solitons [CJ88, TS88] are particular examples of MSCs. A similar case is a soliton and its 'shadow' [Men87, Men88]. A parametric interaction between two waves at different frequencies can result in their coupling and the formation of a parametric soliton [KS74, KS75], which is another example of an MSC. One more widely - explored model of nonlinear superposition of high frequency and low frequency vibrations is that of Davydov solitons in solid state physics, molecular physics and biology [Dav91, Sco92]. Recently, it has been shown that spatial incoherent solitons [MCSS96, SS98, CCJ97, CCMS97] can propagate in photorefractive materials [SSV⁺95, DSS⁺93, ICMASM⁺94, ZAMS96, MSAZ96]. In many cases, multisoliton complexes appear in conservative systems which may be Hamiltonian. However, generalization to nonconservative systems is also possible [ASCC⁺98].

The number of components, M , in a complex can be arbitrary, and can go up to infinity for incoherent solitons [SM98, SA98]. We assume, throughout the rest of this paper, that the *components have independent phases*. The difference between the two types of phase relationship means that, when the interactions between the components are phase-dependent, stationary self-trapping can occur only if the relative phases of all the components are fixed, so that the soliton solutions which form are one-parameter families and they can be represented on plots of Hamiltonian-versus-energy [AA97]. On the other hand, when the phases are independent, the MSC is a multi-parameter family. In the latter case, coherent four wave mixing (FWM) terms (which are also known as energy exchange terms) average out and effectively disappear [KSAA96, Has80]. This is the case for spatial incoherent solitons which can be excited in photorefractive materials [SC01].

In general, MSCs can be described by a set of M coupled nonlinear Schrödinger (NLS) equations. For example, evolution of spatial solitons along the propagation direction can be modeled, in the parabolic approximation, by a system of NLS equations for the set of modes, where the equations are coupled through the change of refractive index. In the case of (1+1)-D spatial geometry, the normalized equations describing the propagation of M self-trapped, mutually-incoherent wave packets in a medium with a Kerr-like nonlinearity are [SC01]

$$(1.1) \quad i \frac{\partial u_m}{\partial z} + \frac{1}{2} \frac{\partial^2 u_m}{\partial x^2} + \mathcal{F}(I) u_m = 0,$$

where $m = 1, 2, \dots, M$, while M is the number of modes (or components), u_m is the complex amplitude of the m -th mode, x is the transverse coordinate, z is the coordinate along the direction of propagation, $I = \sum_{j=1}^M |u_j|^2$ is the total intensity, and $\mathcal{F}(I)$ is the normalized change of refractive index profile created by all the incoherent components of the light beam. The response time of the nonlinearity is assumed to be long compared with temporal variations of the mutual phases of all the components, so the medium responds to the average light intensity, and this is just a simple sum of modal intensities I .

Ordinary solitons are known to behave like single particles. Thus, the difference between a single ordinary soliton and an incoherent soliton could be compared to that between an elementary particle and a complicated structure, such as an atom. Indeed, detailed analysis has shown that MSCs are multi-parameter families of solutions [AKS98], as distinct from single-parameter families, such as NLSE-solitons [AA97]. Moreover, MSCs behave like multi-particle objects in collisions [AKS98]. A simple example of a one-parameter family of “optical mesons” has been considered in [DH97].

Solitons belonging to different components can couple together through the cross-phase modulation (XPM) effect, since the presence of one component results in a modification of the effective refractive index for the other components according to Eq. (1.2). The possibilities include various combinations of bright and dark solitons [Man73, Ino76, TS88, TWWS88, ADPS88, Chr88]. For example, a coupled dark-bright soliton pair has been observed in a self-defocusing nonlinear medium [SB92].

Since a multi-soliton complex is, by definition, a composite structure, it can behave in a more complicated way than a conventional one-component soliton. For example, the MSC shape is not fixed, so it can change after collisions with other solitons [AKS98]. In the following, we study the features of bright and dark MSCs. These can be linked to experimental observations in electrically biased photorefractive crystals.

Let us briefly outline some general properties of Eqs. (1.1) and (1.2). The set of equations (1.1) has M quantities

$$Q_i = \int_{-\infty}^{\infty} |u_i|^2 dx,$$

which are conserved separately from the conservation of the total energy

$$Q = \sum_{i=1}^M Q_i.$$

This occurs because there is no energy transfer mechanism between the components. In fact, this is the main difference from the phase-dependent components case, where only the total energy is conserved.

We now consider the *low saturation* case, where the photorefractive medium response approximately follows the Kerr-law dependence [DCD01], and in normalized units we have

$$(1.2) \quad \mathcal{F}(I) = sI,$$

where $s = +1$ in the self-focusing case and $s = -1$ in the defocusing case.

In some special cases, the coupled NLS equations are found to be *integrable* [AC91, MP82, GI82, NPSSM98, KL01]. In particular, Eqs. (1.1), with the nonlinear response function defined in Eq. (1.2), are, in fact, a generalized Manakov set. This set of coupled equations is *completely integrable* by means of the inverse scattering technique (IST). This technique was first developed for a one-component ($M = 1$) NLS equation [ZS71], then extended to the case of two ($M = 2$) coupled equations [Man73], and it was later demonstrated that the equations are integrable for arbitrary M [GK83, MP82]. As a consequence, any solution of Eq. (1.1) and (1.2) can be represented as a nonlinear superposition of a finite number of *solitary waves* and *radiation modes* which correspond to the discrete and

continuous parts of the linear (L, A) operators, respectively [Man73, ZS71]. The soliton part of the solution accounts for wave localization, while radiation waves appear if the background is present. The property of integrability allows a simple qualitative approach to the problem, and we can also find exact solutions in explicit analytical forms.

Every *fundamental soliton* (labeled j) is characterized by (i) a complex wavenumber $k_j = r_j + i\mu_j$, (ii) a shift in the coordinate plane (x_j, z_j) , and (iii) a polarization vector $\mathbf{p}^{(j)}$ in the function space, normalized to unity as $\sum_{m=1}^M |p_m^{(j)}|^2 = 1$. The simplest bright single-soliton solution in a self-focusing medium ($s = +1$) can be written as:

$$(1.3) \quad u_m(x, z) = p_m^{(j)} r_j \operatorname{sech}(\beta_j) e^{i\gamma_j},$$

where $\beta_j = r_j(\bar{x}_j - \mu_j \bar{z}_j)$, $\gamma_j = \mu_j \bar{x}_j + (r_j^2 - \mu_j^2) \bar{z}_j/2$, and $(\bar{x}_j, \bar{z}_j) = (x - x_j, z - z_j)$ are the shifted coordinates. The peak soliton intensity and its inverse width are determined by the real part of the wavenumber, r_j , while the imaginary part, μ_j , characterizes the tangent of the inclination angle of the soliton (or the velocity in the transverse direction). Moreover, each fundamental soliton can be “spread out” into several incoherent components, as defined by the polarization vector. We note that the term “polarization” is used because Eqs. (1.1), in the case $M = 2$, can describe coupling of two components with orthogonal polarizations of the electric field [Man73]. However, in our case, the polarization parameters $\mathbf{p}^{(j)}$ are not related to the orientation of the electric field.

The solution for a single radiation mode, in the form of a plane wave, can be characterized by a similar set of parameters,

$$(1.4) \quad u_m(x, z) = p_m^{(j)} r_j e^{i\sigma_j},$$

where $\sigma_j = \mu \bar{x}_j + (2sr_j^2 - \mu_j^2) \bar{z}_j/2$. Such a plane wave exists for either sign of nonlinearity, $s = \pm 1$, and it is stable in a self-defocusing medium. Moreover, an incoherent superposition of a large number of plane waves can be stable, even in a self-focusing medium, as was shown in Ref. [SSC⁺00]. In the presence of solitons, the plane waves are distorted, but due to the integrability of the original equations, the corresponding solutions can be obtained in an explicit form, as we demonstrate in Sec. 3 below.

2. Bright multi-soliton complexes

2.1. General solution. A *stationary MSC* can only be formed by *incoherently coupled* fundamental solitons with identical angle tangents (velocities) μ_m , and radiation waves. In the framework of the integrable model given by Eqs. (1.1) and (1.2), such a structure is asymptotically unstable, since a small perturbation can result in a change of the fundamental soliton angles. Nevertheless, the break-up of multi-soliton complexes into individual fundamental solitons can be neglected if we consider their propagation over finite distances, since the instability mode only grows linearly with distance.

For multi-soliton complexes existing in photorefractive crystals, all the solitons and radiation modes should have orthogonal polarization vectors, i.e. $\sum_{m=1}^M p_m^{(j)} p_m^{(n)} = \delta_{jn}$. Indeed, if this condition is not satisfied, an intensity profile will experience periodic beating due to the difference in the phase velocities of the fundamental solitons and radiation waves (see, e.g., Refs. [AA97, PS99, SA99]),

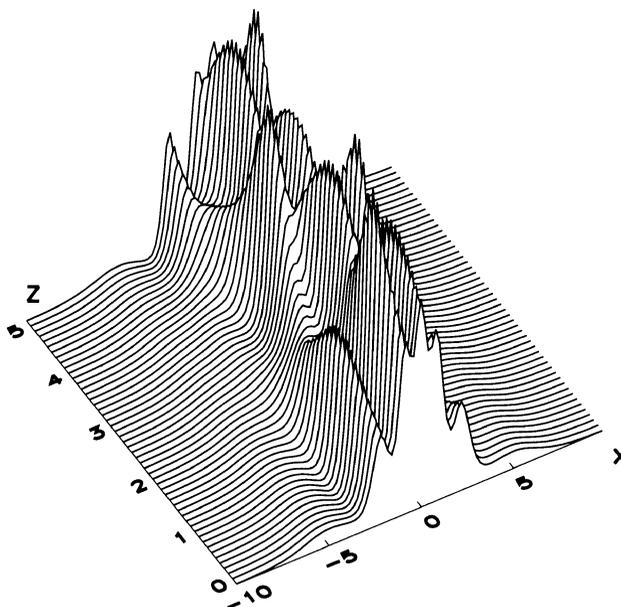


FIGURE 1. Evolution of an MSC intensity profile with multi-scale periodic "beating" due to internal coherent interactions.

as illustrated in Fig. 1. However, such behavior is not consistent with the fact that the model Eqs. (1.1) was introduced for time-averaged fields.

The mathematical description can be simplified in the case of orthogonal polarizations if we use the *rotational symmetry in the functional space* of the original Eqs. (1.1). Indeed, it is sufficient to find solutions u_j where each fundamental non-linear eigenmode belongs to a different component, $p_m^{(j)} = \delta_{mj}$, and then the full family of solutions can be determined using the following transformation:

$$(2.1) \quad \bar{u}_m = \sum_{j=1}^M \mathbf{R}_{mj} u_j,$$

where the matrix \mathbf{R}_{mj} defines a rotation in the M -dimensional space (characterized by $M - 1$ angles), which preserves the MSC intensity profile $\sum_m |\bar{u}_m|^2 \equiv \sum_m |u_m|^2$.

Bright MSC solutions of Eqs. (1.1) and (1.2), existing in the self-focusing case ($s = +1$), can be found from the set of linear equations [NW76]:

$$(2.2) \quad \sum_{m=1}^M \frac{e_j e_m^* u_m}{k_j + k_m^*} + \frac{1}{2r_j} u_j = -e_j,$$

where $e_j = \chi_j \exp(\beta_j + i\gamma_j)$. The coefficients χ_j can be arbitrary, but, in order to simplify the analysis, it is convenient to choose them in a special way [SA99],

$$(2.3) \quad \chi_j = \prod_{m \neq j} \sqrt{b_{jm}},$$

where $b_{jm} = (k_j + k_m^*) / (k_j - k_m)$, and the square root value is taken on the branch with argument in the range $(-\pi/2, \pi/2]$.

For the particular choice of coefficients given by (2.3), it is possible to derive an explicit analytical solution of Eqs. (2.2) for the mode amplitudes in multi-soliton complexes. Using the mathematical induction approach, we obtain the following result [SA99]:

$$(2.4) \quad \begin{aligned} u_j &= \frac{e^{i\gamma_j}}{U} \sum_{\{1, \dots, j-1, j+1, \dots, M\} \rightarrow L} C_L^j F_L^j(x, z), \\ U &= \sum_{\{1, \dots, M\} \rightarrow L} C_L F_L(x, z), \end{aligned}$$

where

$$(2.5) \quad \begin{aligned} C_L &= T_{mb}, \quad C_L^j = 2r_j \chi_j T_{mb}, \\ F_L &= \cosh(S_b), \quad F_L^j = \cosh(S_b^j). \end{aligned}$$

Here L denotes sets of indices (L_1, L_2) , and the summation goes over all possible permutations of soliton numbers between the two sets. Then, the variables for each realization of L are found to be:

$$(2.6) \quad \begin{aligned} T_{mb} &= \prod_{j \in L_1; m \in L_2} c_{jm}, \quad S_b = \sum_{m \in L_1} \beta_m - \sum_{m \in L_2} \beta_m, \\ S_b^j &= S_b + i \sum_{m \in L_1} \varphi_{jm} - i \sum_{m \in L_2} \varphi_{jm}, \end{aligned}$$

where $c_{jm} = |b_{jm}|$, $\varphi_{jm} = \arg(1/b_{jm})/2$, with the function \arg providing an argument value in the interval $[0, 2\pi)$.

We note that only β_j and γ_j depend on the coordinates (x, z) . All the other coefficients are expressed in terms of the complex wave numbers k_j and constant shifts in positions (x_j, z_j) of M fundamental solitons. Since Eqs. (1.1) possess a translational symmetry along the x and z axes, the soliton solution can be shifted as a whole. Therefore, the number of independent parameters controlling the multi-soliton complex is $2M - 1$.

2.2. Soliton interactions. Suppose we have an initial field distribution consisting of fundamental solitons whose positions along the x axis are d_j . We assume that each pair of solitons (numbered j and m) are either (i) well separated, i.e. $|d_j - d_m| \gg (r_j^{-1} + r_m^{-1})$, or (ii) compose a symmetric MSC, i.e. $d_j \equiv d_m$ and $\mu_j = \mu_m$. Then, the intensity profiles of all the solitons are symmetric, and the shifts x_j in the general solution, are:

$$(2.7) \quad x_j = d_j + \sum_{m \neq j} \delta x_{jm} / 2 + z_j \mu_j,$$

where z_j is an arbitrary parameter which determines the soliton phase, and

$$(2.8) \quad \begin{aligned} \delta x_{jm} &= \text{sign}(d_m - d_j) \ln(c_{jm}) / r_j, \\ \text{sign}(\xi) &= \begin{cases} 1, & \xi > 0, \\ 0, & \xi = 0, \\ -1, & \xi < 0. \end{cases} \end{aligned}$$

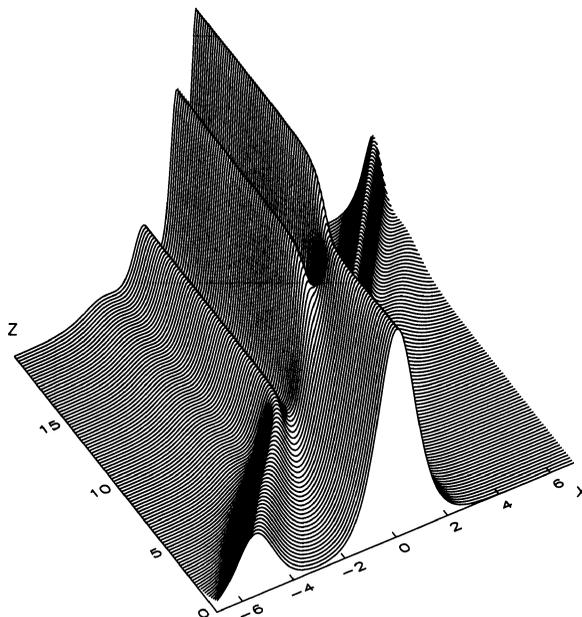


FIGURE 2. Interaction of two multisoliton complexes, resulting in subsequent reshaping. The stationary (zero-velocity) MSC is composed of four fundamental solitons with amplitudes $r_m = 1, 2, 3, 4$ ($\mu_m = 0$), while the MSC at an angle (moving MSC) consists of two solitons with $r_m = 1, 2$ and angle tangent (velocity) $\mu_m = 0.5$.

In general, the fundamental soliton profiles are asymmetric, and simple analytical expressions for the shift parameters cannot be obtained.

It follows from Eq. (2.7) that, after the collisions, the fundamental solitons acquire lateral shifts (i.e. shifts along the x axis). For a single soliton, the translational shift is found to be

$$(2.9) \quad \delta x_j = \sum_m \delta x_{jm}.$$

Here the summation involves the fundamental solitons which participate in the collisions with the soliton number j . This result agrees with the expression found in the $M = 2$ case [Man73]. According to Eq. (2.9), the shift is different for each soliton in an MSC. As a result, the intensity profile of an MSC changes after a collision [AKS98] (see an example in Fig. 2).

3. Multi-soliton complexes on a background

3.1. General solution. We now study the properties of MSCs existing on top of a background which is composed of radiation modes (see Fig. 3). A simple case with one component in the radiation field was studied in [AA99]. In general, there can be an arbitrary number (M_r) of radiation modes, and, to be specific, we

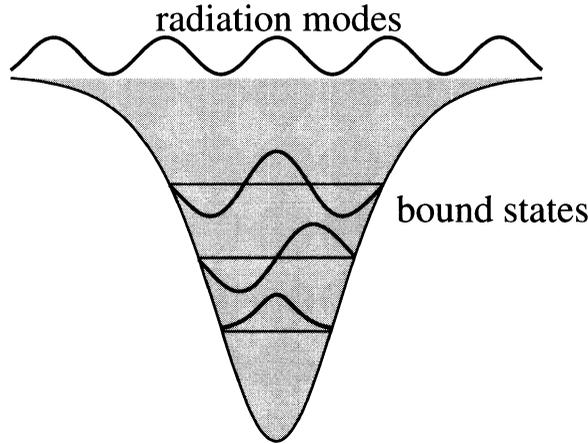


FIGURE 3. Modal structure of an MSC on a background: (i) discrete levels — *fundamental solitons* and (ii) continuum spectrum — *radiation waves*.

assume that they belong to the components with $M_s + 1 \leq m \leq M_s + M_r$, while the fundamental solitons are numbered so that $1 \leq m \leq M_s$.

In order to reveal the basic properties of radiation modes in the presence of an MSC, we first perform a *linear analysis*, assuming that the radiation wave amplitudes are vanishingly small, and that they do not contribute to the intensity profile. We note that, in the limit $r_j \rightarrow 0$, Eqs. (1.3) and (1.4) coincide. Therefore, the low-amplitude radiation mode profile can be found by taking Eqs. (2.4), which define soliton profiles in bright MSCs, and considering the limit

$$(3.1) \quad r_m \rightarrow +0.$$

Then, the profile in component number m will approach that of a dark mode, provided that the limiting transformation is done properly. It is now convenient to return to the system of linear equations (2.2), and after applying the limit (3.1) we have (up to a constant phase which can be neglected) [SA00, SAA01]:

$$(3.2) \quad u_m = r_m(1 + J_m)e^{i\mu_m(\bar{x}_m - \mu_m\bar{z}_m/2)},$$

where

$$(3.3) \quad J_m = \sum_{j=1}^{M_s} \frac{u_j e_j^*}{k_j^* + i\mu_m}.$$

This sum depends only on the amplitudes of the bright components, which in turn are found from an independent system of M_s linear equations (2.2). Note that solution (3.2), which is valid in the limit $r_m \rightarrow 0$, reduces to a simple plane-wave profile given by Eq. (1.4) in the absence of bright components.

When the radiation wave amplitudes r_m are not small, both the radiation modes and the bright soliton profiles defined by Eq. (2.2) are distorted according to the nonlinear superposition principle. In the following, we develop a special technique for constructing solutions for MSCs with a non-zero background. We recall that the self-induced waveguide depends only on the mode intensities. Thus, important information can be obtained by analyzing the normalized intensity profile of the low-amplitude dark mode defined by Eq. (3.2), $|u_m/r_m|^2 = 1 + J_m + J_m^* + |J_m|^2$.

In order to calculate this value, we multiply Eqs. (2.2) by $u_j^*/(k_j - i\mu)$, add the complex conjugate and sum over the fundamental soliton numbers $1 \leq j \leq M_s$. On comparing the resulting expression with Eqs. (3.2) and (3.3), we obtain the following relation:

$$(3.4) \quad \sum_{j=1}^{M_s} \left| \frac{u_j}{k_j^* + i\mu_m} \right|^2 + \left| \frac{u_m}{r_m} \right|^2 = 1,$$

where the subscript m indicates a radiation mode. This remarkable result demonstrates an intrinsic relation between the intensities of bright and dark solitons. Moreover, relation (3.4) opens up an opportunity to introduce a *scaling transformation* and construct solutions for MSCs on a background, with the dark components having non-zero amplitudes [SA00, SAA01]. Indeed, let us scale the bright soliton intensities with the following coefficients,

$$(3.5) \quad |U_m|^2 = s + \sum_{j=M_s+1}^{M_s+M_r} \frac{r_j^2}{|k_m^* + i\mu_j|^2},$$

while for the radiation modes we put $U_m = 1$. Then, the full intensity of the re-scaled solution, \tilde{u}_m , including the contribution of the finite-amplitude radiation waves, is found to be

$$(3.6) \quad \tilde{I} = I_b + sI.$$

Here I is the intensity profile of the bright MSC on a zero background, and $I_b = \sum_{j=M_s+1}^{M_s+M_r} |r_j|^2$ is the background intensity. Quite remarkably, this procedure can be used for both signs of nonlinearity ($s = \pm 1$). We note that the nonlinearly-induced waveguide profiles defined by Eq. (1.2) coincide for the original [$\mathcal{F}(I) = I(x, z)$] and re-scaled [$\mathcal{F}(\tilde{I}) = sI_b + I(x, z)$] solutions, up to a constant background. Therefore, the *self-consistency condition is preserved*, and this is the principal feature of the introduced transformation. The presence of the background can be taken into account by modifying the propagation constants, so that the resulting functions satisfy the original Eqs. (1.1) and (1.2). We finally obtain:

$$(3.7) \quad \tilde{u}_m(x, z) = U_m e^{isI_b z} u_m(x, z).$$

At this point, the derivation of the analytical solutions for MSCs existing on top of several radiation modes is complete, and the component profiles are defined by Eqs. (3.5) and (3.7), together with Eqs. (2.2) and (3.2). Each solution of this type corresponds to a multi-parameter family which can be generated with the help of the rotation transformation (2.1).

Let us now extend the analytical results to the case where the background is composed of a continuum set of radiation modes, i.e. $M_r \rightarrow +\infty$. This situation corresponds to spatial optical solitons excited by an incoherent light source [SC01]. Here the plane waves in the background can be characterized by an angular distribution function, $R(\mu) (\geq 0)$, so that $R(\mu)d\mu$ is the wave intensity corresponding to the tangents of inclination angles in the interval $(\mu, \mu + d\mu)$. Therefore, the full background intensity is $I_b = \int_{-\infty}^{+\infty} R(\mu)d\mu$, and the scaling coefficients for bright components are

$$(3.8) \quad |U_m|^2 = s + \int_{-\infty}^{+\infty} \frac{R(\mu)}{|k_m^* + i\mu|^2} d\mu.$$

We note that, for a finite number of radiation modes, the distribution function can be written as $R(\mu) = \sum_{j=M_s+1}^{M_s+M_r} r_j^2 \delta(\mu - \mu_j)$, and then expression (3.8) reduces to Eq. (3.5).

We stress that the above results are valid for both self-focusing ($s = +1$) and self-defocusing ($s = -1$) media. As follows from Eq. (3.6), the qualitative difference is that, in the former case we have bright complexes on a constant background while in the latter case dark dips are formed.

3.1.1. *Modulation of background components.* According to the general relation (3.6), the intensity profile is uniquely determined by the eigenvalues of the bright fundamental solitons and the background intensity I_b , and does not depend on the the angular distribution of radiation waves. However, the total intensity of the soliton components,

$$I_s = \sum_{m=1}^{M_s} |U_m|^2 |u_m|^2,$$

and the intensity of the radiation modes,

$$(3.9) \quad I_r = I_b - \sum_{m=1}^{M_s} (|U_m|^2 - s) |u_m|^2,$$

both depend on the scaling coefficients U_m , defined in Eq. (3.8). As follows from Eq. (3.9), each fundamental soliton creates a dark hole in the background, and the corresponding modulation depth is proportional to the *bright-dark coupling coefficient*, given by the value $(|U_m|^2 - s)$. Interestingly enough, the radiation mode profiles are the same in self-focusing and self-defocusing media, provided the distribution function and soliton eigenvalues remain unchanged.

However, there are some key differences between solitons in self-focusing and self-defocusing media. In the former case, any modulation of the background is compensated by the bright components having larger amplitudes (since $|U_m|^2 > s$, and $s = +1$). On the other hand, in a self-defocusing medium, a dark soliton creates an effective waveguide, which in turn can trap bright components. Such a self-trapping mechanism results in the restriction that there is a *minimum for the dark soliton width*. This happens because the maximum intensity contrast is limited by the value of the background intensity. As a matter of fact, the limitation can be even stricter, since the maximum modulation depth, $\mathcal{M} = \max_x (I_b - I) / I_b \leq 1$, cannot always reach the value of 1. Then, according to Eq. (1.3), the characteristic width corresponding to one fundamental soliton cannot exceed the value $(\mathcal{M} I_b)^{-1/2}$. The actual limit can be determined by solving the existence conditions, which follow from the requirement that the right-hand-side of Eq. (3.8) be non-negative, since, by definition, $|U_m|^2 \geq 0$. It is interesting to note that these conditions involve only the individual wavenumbers of fundamental solitons, and that they are automatically satisfied for interacting solitons forming MSCs.

The radiation modes are characterized by a non-trivial phase modulation. For practical applications, it is especially important to know the *phase jump*, or the additional phase shift which appears due to the presence of bright fundamental solitons. Using Eqs. (2.2) and (3.2), we find the following relation,

$$e^{i\phi(\mu)} = \prod_{m=1}^{M_s} \frac{i\mu - k_m}{i\mu + k_m^*},$$

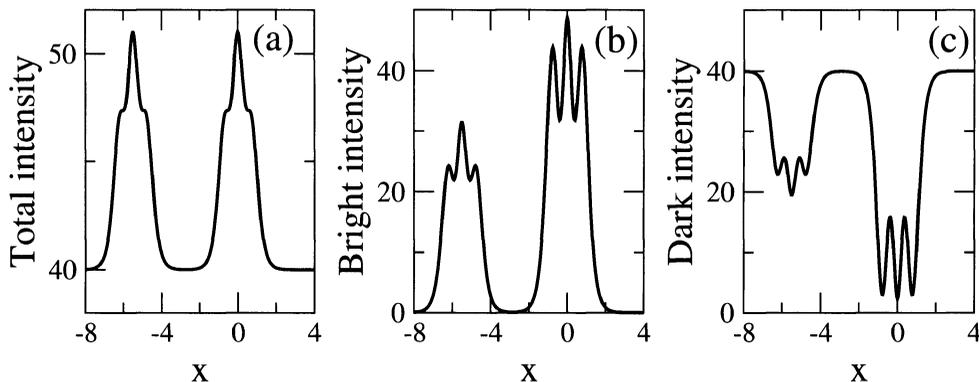


FIGURE 4. (a) Total intensity profiles for multisoliton complexes, and the corresponding decomposition between (b) bright and (c) dark components in a self-focusing medium. Each MSC consists of three fundamental solitons with $r_m = 2, 3, 4$; the MSC on the right is stationary ($\mu_m = 0$), while the MSC on the left has a non-zero (positive) incidence angle ($\mu_m = 3$). Angular width of background distribution is $\rho = 1$.

where ϕ is the phase jump, and μ defines the inclination angle of the radiation mode. Then, the phase jump can be found as a sum over the phase shifts associated with individual fundamental solitons,

$$(3.10) \quad \phi(\mu) = \sum_{m=1}^{M_s} \phi_m(\mu) = \sum_{m=1}^{M_s} 2 \arctan \left(\frac{r_m}{\mu - \mu_m} \right).$$

We see that the absolute values of the individual phase shifts are limited to π . However, the total phase jump can become larger than π if $M > 1$, i.e. if the MSC is composed of several fundamental solitons.

3.1.2. *Special case of dark-only solitons.* Pure dark solitons, supported by a defocusing nonlinearity ($s = -1$), were extensively investigated earlier in one [AA97] and two-component [SK97] cases. Our solution (3.7) can be reduced to describe such cases, if we choose the soliton wave numbers in such a way that $|U_m| = 0$ in Eq. (3.5). Then, the amplitudes of all the bright solitons reduce to zero, and the resulting expression gives a multi-dark soliton solution. In the case of a single dark component, the condition reduces to a simple relation [ZS71] $I_b = |k_m|^2$, where $1 \leq m \leq M_s$.

3.2. General results for a Gaussian angular distribution. Let us now analyse the features of the bright-dark decomposition for a Gaussian-type angular distribution of the radiation waves which compose the background,

$$(3.11) \quad R(\mu) = I_b \frac{1}{\rho\sqrt{\pi}} e^{-\mu^2/\rho^2},$$

where $\rho(> 0)$ is the characteristic angular width. Since the integral in Eq. (3.8) cannot be expressed in elementary functions for arbitrary ρ , we first consider the limiting cases. Specifically, for a narrow angular distribution, i.e. $\rho \ll 1$, we have $(|U_m|^2 - s) \simeq I_b/(r_m^2 + \mu_m^2)$, while for $\rho \gg 1$ (and $\rho \gg \mu_m$) we obtain $(|U_m|^2 - s) \simeq I_b\sqrt{\pi}/(\rho r_m) \rightarrow 0$. Therefore, we expect that, for a fixed background intensity

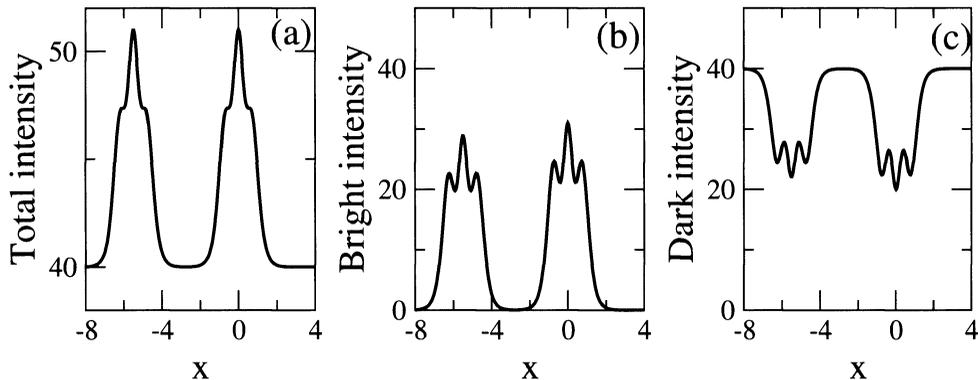


FIGURE 5. Intensity distributions for $\rho = 7$ in a self-focusing medium. Parameters and notation are the same as in Fig. 4.

I_b , the modulation of the radiation waves should be reduced (i) for wider angular distributions, i.e. larger ρ , and (ii) for MSCs having higher velocities $|\mu_m|$.

On the other hand, since the phase jump depends on the radiation mode wave number μ , the excitation of solitons can be more difficult in cases of wider angular spectra of radiation modes, i.e. larger ρ . Additionally, for an MSC at an angle, i.e. when $\mu_m = \text{const.} \neq 0$, the dependence $\phi(\mu)$ becomes asymmetric, unless $R(\mu - \mu_m) = R(\mu_m - \mu)$.

3.3. Bright solitons in a self-focusing medium. For a self-focusing non-linearity ($s = +1$), MSCs exist in the form of bright localized waves having higher intensity than the background, as follows from Eq. (3.6). Examples of the total intensity profiles and the bright-dark mode decomposition are shown in Figs. 4 and 5 for different μ values of ρ . The MSC on the right has zero angle (with the corresponding $\mu_m = 0$), while the other MSC (on the left) has a positive angle; thus in the latter case, the corresponding background modulation is smaller, as predicted in Sec. 3.2.

A collision between two MSCs is illustrated in Fig. 6. This example corresponds to the initial conditions shown in Fig. 4. A remarkable fact is that the total intensity profile does not depend on the value of ρ , provided that I_b is preserved. The intensity profile for the collision will be the same for other values of ρ or for other distribution functions. In these examples, the MSC actually has an intensity which is relatively small compared with the background level.

Note that the shape of each MSC changes after the collision, for the reasons discussed in Sec. 2.2. In particular, a symmetric MSC becomes asymmetric after a collision. The presence of radiation does not influence this process. Another feature of a collision is that the lateral shift of the MSCs is relatively large. For example, it can easily be seen on the scale of Fig. 6. In contrast to single solitons, MSCs experience larger shifts during collisions, due to the multiple contributions from all the constituent fundamental solitons.

3.4. Dark solitons in self-defocusing medium. To describe MSCs on a background in a self-defocusing medium ($s = -1$), we first have to determine the existence conditions, as outlined in Sec. 3.1.1. By considering the case of the Gaussian distribution given by Eq. (3.11), we find that, for a narrow angular spectrum, in the lowest-order approximation, the existence condition is $|k_m|^2 = r_m^2 + \mu_m^2 \leq I_b$.

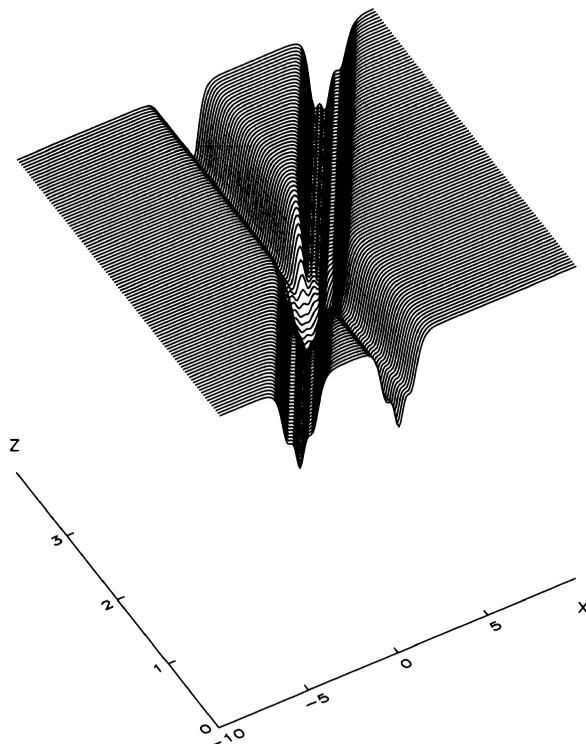


FIGURE 6. Interaction of two multisoliton complexes existing on a multi-component background in a self-focusing medium. Input profile corresponds to Fig. 4.

Therefore, the minimum soliton width, which is of order r_m^{-1} , can be achieved if the soliton angle is zero. In the other limit where $\rho \gg 1$ and $\rho \gg \mu_m$, we have $r_m \leq I_b \sqrt{\pi}/\rho$, i.e. the minimum width increases linearly with an increase in ρ . Numerically-calculated existence regions are shown in Fig. 7 for two values of ρ in Eq. (3.11). Figure 7(a) clearly shows that the existence region is very similar to that in the case of a single component radiation field when ρ is relatively small. However, the existence regions become visibly different when ρ is large, as seen in Fig. 7(b).

The range of possible soliton widths for various ρ is given by the shaded region in Fig. 8. This result shows that the distribution function for the radiation field influences the properties of an MSC, in that it changes the limiting parameters for the existence of the MSC, although the intensity profile of the MSC is not directly influenced by the properties of the radiation field.

Figure 9 shows an example of the intensity distribution for a self-defocusing medium (i.e. dark MSCs on a background). We have chosen the soliton eigenvalues to be the same as those in Figs. 4 and 6. According to our general expression (3.6), the total intensity profiles in self-defocusing and self-focusing media are “mirror-images” relative to the level of the background. Even the radiation mode intensities coincide in these two cases — cf. Figs. 4(c) and 9(c). However, the bright component intensities are different, as is clearly seen in Fig. 4(b) and 9(b). This is a manifestation of the nontrivial nature of the nonlinear superposition of the solitons and the background components.

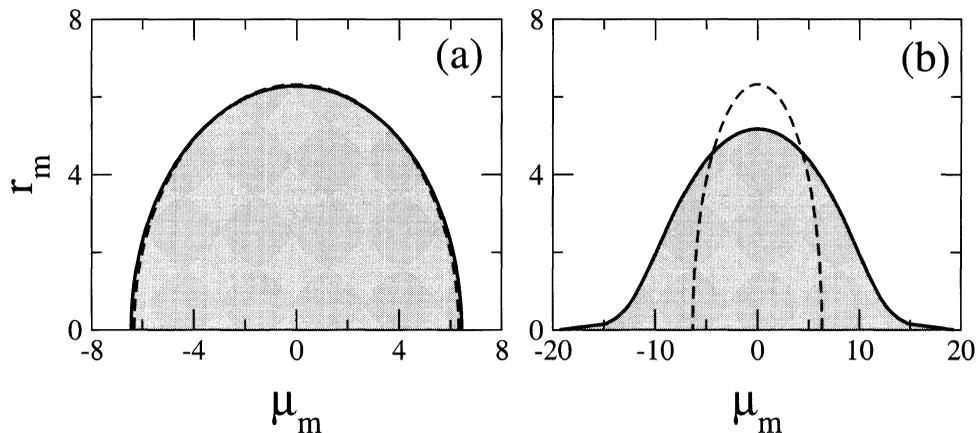


FIGURE 7. Grey shading marks existence regions in the parameter space of fundamental soliton eigenvalues (r_m, μ_m) . The angular distribution function is given by Eq. (3.11) with $I_b = 40$ and (a) $\rho = 1$ or (b) $\rho = 7$. Dashed lines correspond to the case of a single component background, when $\rho \rightarrow 0$.

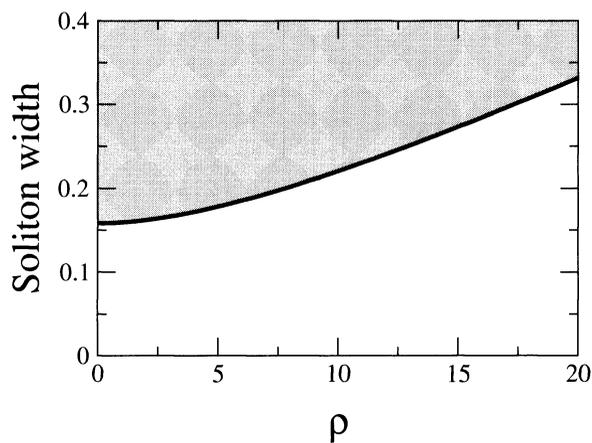


FIGURE 8. Dependence of the range of possible soliton widths on the parameter ρ for the angular distribution function of radiation waves given by Eq. (3.11) with $I_b = 40$.

Figure 10 shows a collision of two MSCs on a background. Again, we can see that the nonstationary intensity profile created by the soliton interaction during collision is the "mirror image" of that for bright MSCs in a self-focusing medium, as shown in Fig. 6. The symmetry relation is mathematically exact. Correspondingly, the lateral shift is also governed by the same rules as those for a bright MSC.

An important consequence is that the change of refractive index induced by incoherent MSCs has exactly the same pattern in cases of self-focusing and self-defocusing media with Kerr-type nonlinearity.

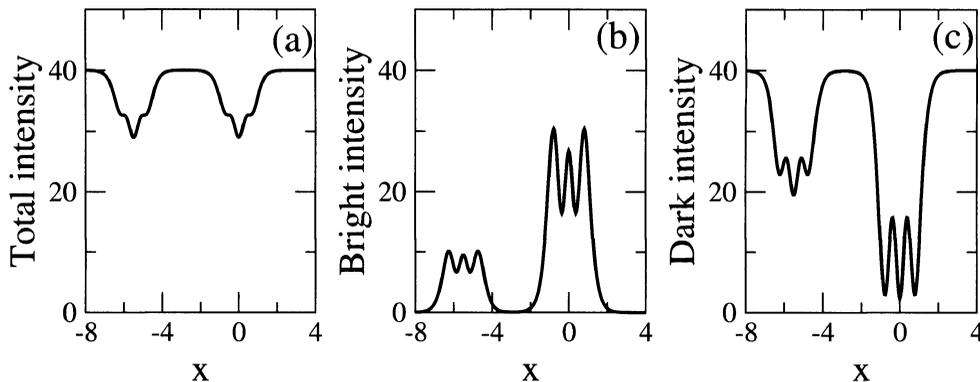


FIGURE 9. Intensity distributions in a self-defocusing medium. Parameters and notation are the same as in Fig. 4.

4. Intensity limits in multi-soliton complexes

4.1. Estimate of the peak intensity. For a single NLS describing the evolution of a coherent field, soliton interactions are phase-sensitive, so that constructive or destructive nonlinear interference can be observed. Since the profiles of interacting solitons are distorted according to the nonlinear superposition principle, large variations of the peak intensities can occur. Indeed, it has been demonstrated that the peak intensity can vary by a factor of N^2 for N interacting solitons [AM91]. In some sense, phase sensitivity can be “amplified” due to nonlinearity, so that it is greater than that occurring for linear interference between mutually coherent sources. Knowledge of how the maximal beam intensity of stationary and interacting MSCs changes during and after collisions can be important for the development of switching devices based on incoherent solitons. In what follows, we derive a rigorous relationship between the parameters of incoherently-coupled solitons and the range of admissible intensities for MSCs.

We perform the analysis for bright MSCs in a self-focusing medium, when the background is absent. However, with the use of Eq. (3.6), the results can be readily applied to bright solitons in a self-focusing medium existing on top of a background, and also to dark solitons which can exist in media with a self-defocusing Kerr-type nonlinearity.

Although the general solution for the MSC profile can be obtained in an explicit form (2.4), it is not possible to find an explicit analytical expression for the maximum intensity, and it can only be determined by numerically solving transcendental equations. Therefore, a different approach is needed to make an analytical estimate for the peak intensity levels. In order to do this, we turn Eq. (3.4) into an inequality,

$$(4.1) \quad \sum_{j=1}^M \left| \frac{u_j}{k_j^* + i\mu} \right|^2 \leq 1,$$

which is valid for any (real) μ . This key result makes it possible to estimate the limitations on the total intensity. Indeed, it follows that the upper boundary is

$$(4.2) \quad I_{max}(z) = \max_x I(x, z) \leq I_{inf} = \max_j [r_j^2 + (\mu_j - \mu)^2],$$

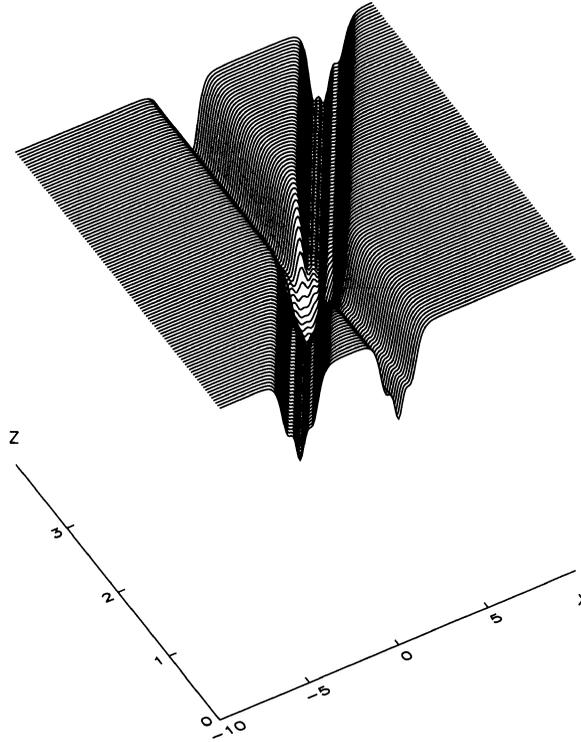


FIGURE 10. Interaction of two multisoliton complexes existing on a multi-component background in a self-defocusing medium. Input profile corresponds to Fig. 9.

and μ is chosen to minimize the value of I_{inf} . The above estimate is invariant relative to a change of all the fundamental soliton angle tangents (velocities) by a constant ($\mu_j \rightarrow \mu_j + \Delta\mu$). This property is due to the Galilean symmetry of original equations (1.1) [AA97]. We shall now consider several examples illustrating the physical meaning of Eqs. (4.1) and (4.2).

4.2. Stationary solitons. Let us first analyse the properties of a single MSC, composed of several solitons with identical angle tangents, μ_m . In such a case, the optimal choice for the free parameter in Eq. (4.2) is $\mu = \mu_m$, and we have $I_{\text{inf}} = \max_j(r_j^2)$. Note that this value can be interpreted as the minimal squared radius of a circle which has its center at the point $(0, \mu_m)$ in the parameter space (r, μ) and which contains all the soliton eigenvalues within it [see Fig. 11(a)]. Thus I_{inf} is proportional to the area (πr_j^2) of this circle. On the other hand, we note that, for a stationary MSC, the amplitude profiles satisfy a self-consistent eigenvalue problem, viz. $d^2U_m/dx^2 - r_m^2U_m + 2IU_m = 0$, where the functions $U_m = u_m \exp(-i\gamma_m)$ are real. Then, since solutions should be localized in the transverse direction (x), we conclude that

$$(4.3) \quad I_{\text{max}}(z) > I_{\text{sup}} = \max_j(r_j^2/2).$$

This means that the variations of the peak intensity are strictly limited by the largest eigenvalue in an MSC, as illustrated in Fig. 11. Note that $I_{\text{max}} = I_{\text{inf}}$ if there is only one fundamental soliton (Fig. 11, left). The peak intensity decreases if several

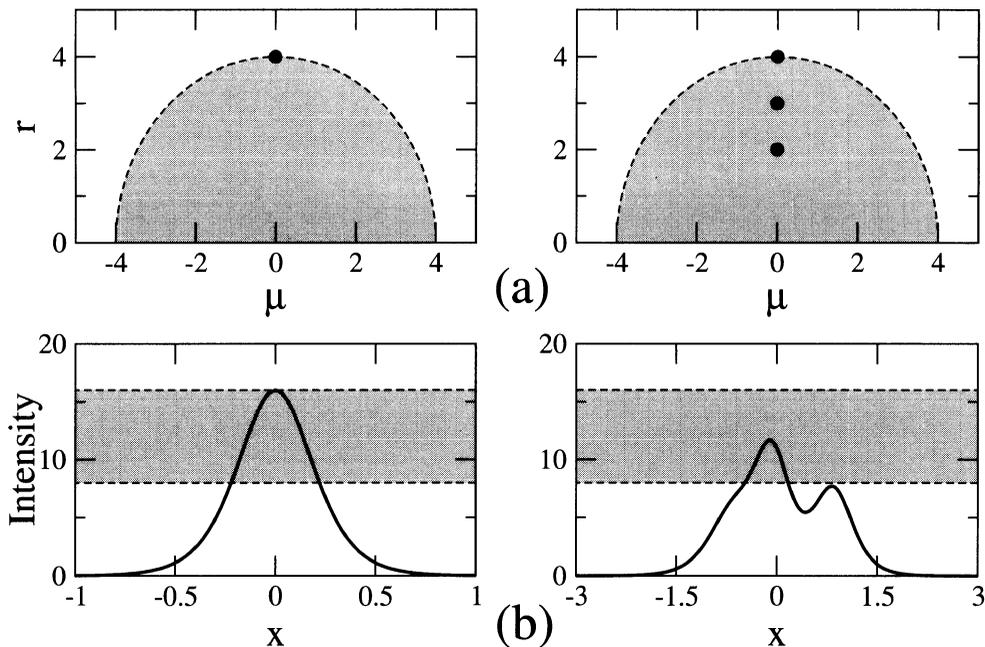


FIGURE 11. (a) Geometrical illustration of the maximum intensity criterion for a stationary MSC: black dots mark the eigenvalues of fundamental solitons in the (r, μ) parameter space, and the radius of the shaded semicircle determines the maximum amplitude. (b) Intensity profiles corresponding to upper plots; the shaded bands show the allowed ranges for the peak intensities.

solitons compose an MSC, but always remains above the lower limit (Fig. 11, right). This occurs despite the fact that the individual fundamental soliton intensities are always superimposed, i.e. destructive interference is not possible, in contrast to the case of coherent interactions. The observed *decrease of total intensity* underlines the complicated nature of the *nonlinear superposition* phenomenon, and occurs because the profiles of individual solitons are strongly distorted due to the nonlinear self-action effect.

4.3. Interacting solitons. Our general results can also be applied to the case when solitons have different angle tangents, μ_n . In other words, we can estimate the peak intensity changes during the collision of several MSCs. As follows from the form of Eq. (4.2), the limiting value I_{inf} depends only on the *maximum eigenvalue in each of the colliding MSCs*. Again, Eq. (4.2) has a clear geometrical interpretation: the optimal value of μ in Eq. (4.2) must be chosen to minimize the area of a semicircle which has its center at the point $(0, \mu)$, and which contains all the soliton eigenvalues. Two examples, corresponding to a collision between a single fundamental soliton and an MSC with different angle tangents (μ_n), are shown in Fig. 12. When the relative angle is small (Fig. 12, left), the intensity at the impact area of the collision decreases. We have already observed this effect for a stationary MSC when the relative angle is zero. However, for larger relative angles, the peak intensity increases (Fig. 12, right).

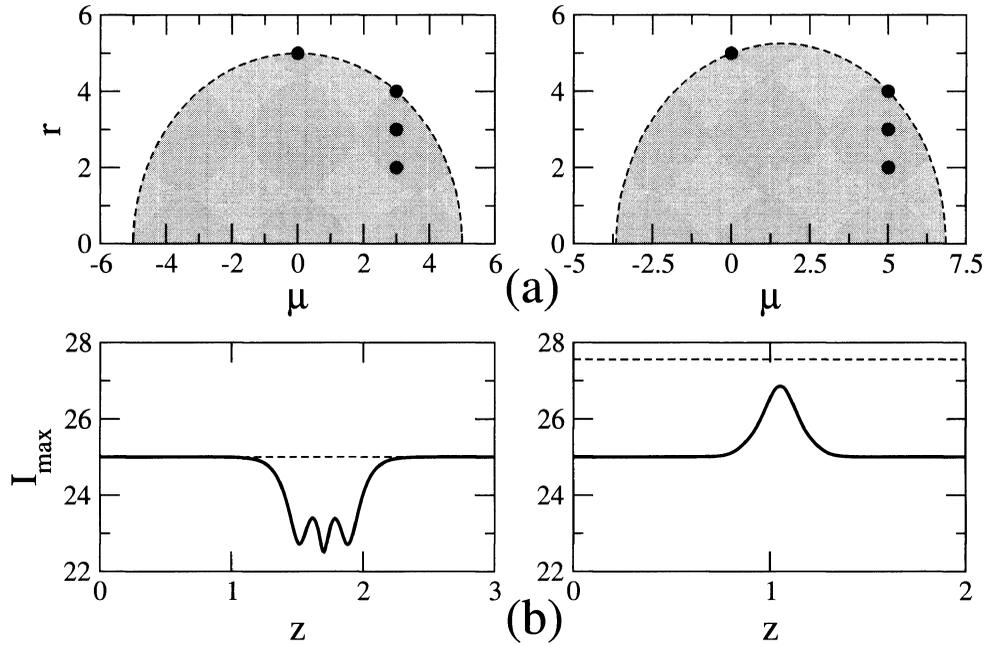


FIGURE 12. (a) Geometrical illustration of the maximum intensity criterion for interacting MSCs: black dots mark the eigenvalues of fundamental solitons in the (r, μ) parameter space, and radius of the shaded semicircle determines the maximum amplitude. (b) Dependencies of the peak intensities, corresponding to the upper plots, on the propagation distance; dashed line shows the maximum possible value.

In order to understand the differences in the interaction pattern, we study the dependence of the peak intensities on the relative angle tangent of the colliding MSCs. To illustrate the key features, we consider interactions of two identical MSCs with a relative angle tangent $2\mu_1$ [see Fig. 13(a)]. Although Eq. (4.2) can be used to obtain an estimate for the peak intensities, as in the previous examples, we find that the results are not optimal for large μ_1 . To obtain a more accurate estimate, we recall that Eq. (4.1) is satisfied for all (real) μ simultaneously. We now choose $\mu = \pm\mu_1$, add the corresponding inequalities together, and obtain the following upper limit,

$$(4.4) \quad I_{\text{inf}} = \max_j \frac{r_j^2(r_j^2 + 4\mu_1^2)}{r_j^2 + 2\mu_1^2}.$$

We thus see that $I_{\text{inf}}(\mu_1 \rightarrow +\infty) \rightarrow 2 \max_j(r_j^2)$, which is a simple sum of the upper bounds for individual MSCs. This result means that the interaction of solitons with large relative angles is weak, and the MSC intensities are added together similarly to the linear case. As a matter of fact, this is a general property of optical solitons [SS99]. Results of numerical simulations for soliton collisions are presented in Fig. 13(b). We also illustrate the evolution of the intensity profiles at small and large relative angles in Figs. 13(c) and (d), respectively.

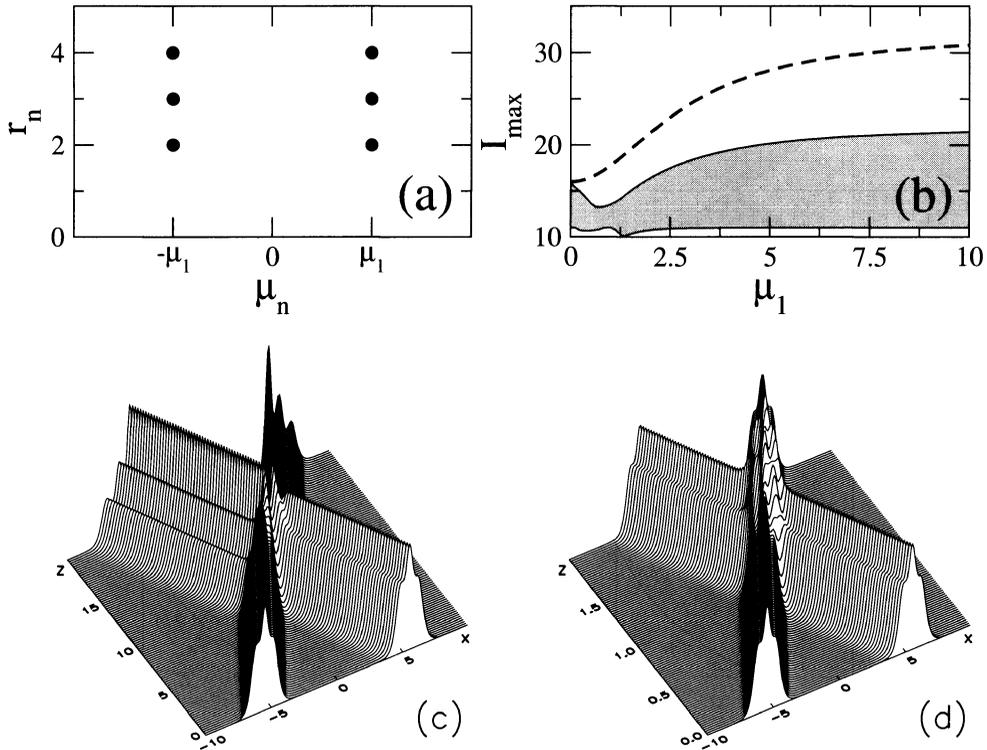


FIGURE 13. (a) Location of the eigenvalues of fundamental solitons in the (r, μ) parameter space (black dots). (b) Ranges of the peak intensity variations during the soliton collision vs. relative angles shown with shading; dashed line gives the maximum possible value. (c,d) Intensity profiles illustrating soliton collisions at small ($\mu_1 = 0.5$) and large ($\mu_1 = 5$) relative angles.

5. Conclusions

In conclusion, we have obtained an exact solution for multisoliton complexes on top of a multi-component background composed of radiation waves in Kerr-type nonlinear media. We have identified similarities and differences between bright and dark MSCs which exist in self-focusing and self-defocusing media, respectively. In particular, we have found that the intensity profiles in these two cases are “mirror-images” relative to the level of the background, and that they depend only on the eigenvalues of the fundamental solitons. For example, the reshaping of MSCs after collisions is determined by the lateral shifts of the fundamental solitons and is not affected by the background components. On the other hand, the width of dark solitons has a minimum, and this value depends strongly on the angular distribution of the radiation waves. We have performed a detailed analysis of the key soliton characteristics for the case of a Gaussian angular distribution of radiation waves, and presented numerical examples illustrating the principal features of bright and dark MSCs.

We have also obtained an accurate estimate for the peak intensities of multi-soliton complexes. We have demonstrated that incoherent coupling can result in a *decrease of the peak intensity* when the relative soliton angles are small (or zero)

and nonlinear interaction is strong. On the other hand, solitons with large relative angles can roughly be superimposed as linear waves, since the interaction is weak.

References

- [AA97] N. N. Akhmediev and A. Ankiewicz, *Solitons: Nonlinear pulses and beams*, Chapman and Hall, London, 1997.
- [AA99] N. Akhmediev and A. Ankiewicz, *Partially coherent solitons on a finite background*, Phys. Rev. Lett. **82** (1999), no. 13, 2661–2664.
- [AA00] ———, *Multi-soliton complexes*, Chaos **10** (2000), no. 3, 600–612.
- [AC91] M. J. Ablowitz and P. A. Clarkson, *Solitons, Nonlinear Evolution Equations and Inverse Scattering*, London Mathematical Society Lecture Notes Series, vol. 149, Cambridge University Press, Cambridge, 1991, p. 53.
- [ADPS88] V. V. Afanas'ev, E. M. Dianov, A. M. Prokhorov, and V. N. Serkin, *Nonlinear pairing of light and dark optical solitons*, Pis'ma Zh. Éksp. Teor. Fiz. **48** (1988), no. 11, 588–592 [English Translation: JETP Lett. **48** (1988), no. 11, 638–642].
- [Akh01] N. N. Akhmediev, *Nonlinear physics - Deja vu in optics*, Nature **413** (2001), no. 6853, 267–268.
- [AKNS74] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, *Inverse scattering transform-Fourier analysis for nonlinear problems*, Stud. Appl. Math. **53** (1974), no. 4, 249–315.
- [AKS98] N. Akhmediev, W. Krolikowski, and A. W. Snyder, *Partially coherent solitons of variable shape*, Phys. Rev. Lett. **81** (1998), no. 21, 4632–4635.
- [AM91] N. N. Akhmediev and N. V. Mitzkevich, *Extremely high degree of n -soliton pulse-compression in an optical fiber*, IEEE J. Quantum Electron. **27** (1991), no. 3, 849–857.
- [ASCC+98] N. N. Akhmediev, J. M. Soto-Crespo, S. T. Cundiff, B. C. Collings, and W. H. Knox, *Phase locking and periodic evolution of solitons in passively mode-locked fiber lasers with a semiconductor saturable absorber*, Opt. Lett. **23** (1998), no. 11, 852–854.
- [BA95] A. V. Buryak and N. N. Akhmediev, *Stationary pulse-propagation in n -core nonlinear fiber arrays*, IEEE J. Quantum Electron. **31** (1995), no. 4, 682–688.
- [BV97] E. P. Bashkin and A. V. Vagov, *Instability and stratification of a two-component Bose-Einstein condensate in a trapped ultracold gas*, Phys. Rev. B **56** (1997), no. 10, 6207–6212.
- [CCJ97] D. N. Christodoulides, T. H. Coskun, and R. I. Joseph, *Incoherent spatial solitons in saturable nonlinear media*, Opt. Lett. **22** (1997), no. 14, 1080–1082.
- [CCMS97] D. N. Christodoulides, T. H. Coskun, M. Mitchell, and M. Segev, *Theory of incoherent self-focusing in biased photorefractive media*, Phys. Rev. Lett. **78** (1997), no. 4, 646–649.
- [Chr88] D. N. Christodoulides, *Black and white vector solitons in weakly birefringent optical fibers*, Phys. Lett. A **132** (1988), no. 8-9, 451–452.
- [CJ88] D. N. Christodoulides and R. I. Joseph, *Vector solitons in birefringent nonlinear dispersive media*, Opt. Lett. **13** (1988), no. 1, 53–55.
- [Dav91] A. S. Davydov, *Solitons in molecular systems*, 2nd ed., Reidel, Dordrecht, 1991.
- [DCD01] Eugenio DelRe, Bruno Crosignani, and Paolo Di Porto, *Photorefractive spatial solitons*, Spatial Optical Solitons (S. Trillo and W. E. Torruellas, eds.), Springer series in optical sciences, vol. 82, Springer-Verlag, New York, 2001, pp. 61–85.
- [DH97] P. D. Drummond and H. He, *Optical mesons*, Phys. Rev. A **56** (1997), no. 2, R1107–R1109.
- [dS94] C. M. de Sterke and J. E. Sipe, *Gap solitons*, Progress in Optics (E. Wolf, ed.), vol. XXXIII, North-Holland, Amsterdam, 1994, pp. 203–260.
- [DSS+93] G. C. Duree, J. L. Shultz, G. J. Salamo, M. Segev, A. Yariv, B. Crosignani, P. Diporto, E. J. Sharp, and R. R. Neurgaonkar, *Observation of self-trapping of an optical beam due to the photorefractive effect*, Phys. Rev. Lett. **71** (1993), no. 4, 533–536.
- [Fer65] E. Fermi, *Collected Papers*, vol. II, p. 978, Univ. of Chicago Press, Chicago, 1965.
- [FPU55] E. Fermi, J. Pasta, and S. Ulam, *Studies of Nonlinear Problems. I*, Tech. Report LA-1940, Los Alamos National Laboratory, 1955.

- [GGKM67] Clifford S. Gardner, John M. Greene, Martin D. Kruskal, and Robert M. Miura, *Method for solving the Korteweg-de Vries equation*, Phys. Rev. Lett. **19** (1967), 1095–1097.
- [GI82] V. S. Gerdzhikov and M. I. Ivanov, *Hamiltonian structure of multicomponent nonlinear Schrödinger equations in difference form*, Teor. Matem. Fiz. **52** (1982), no. 1, 89–104 [English Translation: Theor. Math. Phys. **52** (1982), no. 1, 676–685].
- [GK83] V. S. Gerdjikov and P. P. Kulish, *Multicomponent nonlinear Schrödinger equation in case of nonzero boundary conditions*, Zapiski LOMI **131** (1983), 34–46 (in Russian).
- [Has75] A. Hasegawa, *Dynamics of an ensemble of plane waves in nonlinear dispersive media*, Phys. Fluids **18** (1975), no. 1, 77–79.
- [Has77] ———, *Envelope soliton of random phase waves*, Phys. Fluids **20** (1977), no. 12, 2155–2156.
- [Has80] ———, *Self-confinement of multimode optical pulse in a glass fiber*, Opt. Lett. **5** (1980), no. 10, 416–417.
- [ICMASM⁺94] M. D. Iturbe Castillo, P. A. Marquez Aguilar, J. J. Sanchez-Mondragon, S. Stepanov, and V. Vysloukh, *Spatial solitons in photorefractive Bi₁₂TiO₂₀ with drift mechanism of nonlinearity*, Appl. Phys. Lett. **64** (1994), no. 4, 408–410.
- [Ino76] Y. Inoue, *Nonlinear coupling of polarized plasma-waves*, J. Plasma Phys. **16** (1976), no. DEC, 439–459.
- [KL01] T. Kanna and M. Lakshmanan, *Exact soliton solutions, shape changing collisions, and partially coherent solitons in coupled nonlinear Schrödinger equations*, Phys. Rev. Lett. **86** (2001), no. 22, 5043–5046.
- [KS74] Yu. N. Karamzin and A. P. Sukhorukov, *Nonlinear interaction of diffracting beams of light in a medium of quadratic nonlinearity, mutual focusing of beams and restriction of optical frequency convertor efficiency*, Pis'ma Zh. Éksp. Teor. Fiz. **20** (1974), no. 11, 734–739 (in Russian) [English Translation: JETP Lett. **20** (1974), no. 11, 339–342].
- [KS75] ———, *Mutual focusing of intense light-beams in media with quadratic nonlinearity*, Zh. Éksp. Teor. Fiz. **68** (1975), no. 3, 834–847 (in Russian) [English Translation: JETP **41** (1975), 414–420].
- [KSAA96] J. U. Kang, G. I. Stegeman, J. S. Aitchison, and N. Akhmediev, *Observation of Manakov spatial solitons in AlGaAs planar waveguides*, Phys. Rev. Lett. **76** (1996), no. 20, 3699–3702.
- [Man73] S. V. Manakov, *On the theory of two-dimensional stationary self focussing of electromagnetic waves*, Zh. Éksp. Teor. Fiz. **65** (1973), no. 2, 505–516 (in Russian) [English Translation: JETP **38** (1974), 248–253].
- [MCSS96] M. Mitchell, Z. G. Chen, M. F. Shih, and M. Segev, *Self-trapping of partially spatially incoherent light*, Phys. Rev. Lett. **77** (1996), no. 3, 490–493.
- [Men87] C. R. Menyuk, *Stability of solitons in birefringent optical fibers .1. Equal propagation amplitudes*, Opt. Lett. **12** (1987), no. 8, 614–616.
- [Men88] ———, *Stability of solitons in birefringent optical fibers .2. Arbitrary amplitudes*, J. Opt. Soc. Am. B **5** (1988), no. 2, 392–402.
- [MP82] V. G. Makhankov and O. K. Pashaev, *Nonlinear Schrödinger equation with non-compact isogroup*, Teor. Matem. Fiz. **53** (1982), no. 1, 55–67 [English Translation: Theor. Math. Phys. **53** (1982), no. 1, 979–987].
- [MSAZ96] A. V. Mamaev, M. Saffman, D. Z. Anderson, and A. A. Zozulya, *Propagation of light beams in anisotropic nonlinear media: From symmetry breaking to spatial turbulence*, Phys. Rev. A **54** (1996), no. 1, 870–879.
- [NPSSM98] K. Nakkeeran, K. Porsezian, P. Shanmugha Sundaram, and A. Mahalingam, *Optical solitons in N-coupled higher order nonlinear Schrödinger equations*, Phys. Rev. Lett. **80** (1998), no. 7, 1425–1428.
- [NW76] Y. Nogami and C. S. Warke, *Soliton solutions of multicomponent nonlinear Schrödinger equation*, Phys. Lett. A **59** (1976), no. 4, 251–253.
- [PS99] Q. H. Park and H. J. Shin, *Parametric control of soliton light traffic by cw traffic light*, Phys. Rev. Lett. **82** (1999), no. 22, 4432–4435.
- [SA98] V. V. Shkunov and D. Z. Anderson, *Radiation transfer model of self-trapping spatially incoherent radiation by nonlinear media*, Phys. Rev. Lett. **81** (1998), no. 13, 2683–2686.

- [SA99] A. A. Sukhorukov and N. N. Akhmediev, *Coherent and incoherent contributions to multisoliton complexes*, Phys. Rev. Lett. **83** (1999), no. 23, 4736–4739.
- [SA00] ———, *Multisoliton complexes on a background*, Phys. Rev. E **61** (2000), no. 5, 5893–5899.
- [SAA01] A. A. Sukhorukov, A. Ankiewicz, and N. N. Akhmediev, *Multi-soliton complexes in a sea of radiation modes*, Opt. Commun. **195** (2001), no. 1-4, 293–302.
- [SB92] M. Shalaby and A. J. Barthelemy, *Observation of the self-guided propagation of a dark and bright spatial soliton pair in a focusing nonlinear medium*, IEEE J. Quantum Electron. **28** (1992), no. 12, 2736–2741.
- [SC01] M. Segev and D. N. Christodoulides, *Incoherent solitons: self-trapping of weakly-correlated wave-packets*, Spatial Optical Solitons (S. Trillo and W. E. Torruellas, eds.), Springer series in optical sciences, vol. 82, Springer-Verlag, New York, 2001, pp. 87–125.
- [Sco92] A. Scott, *Davydovs soliton*, Phys. Rep. **217** (1992), no. 1, 1–67.
- [SK97] A. P. Sheppard and Yu. S. Kivshar, *Polarized dark solitons in isotropic Kerr media*, Phys. Rev. E **55** (1997), no. 4, 4773–4782.
- [SM98] A. W. Snyder and D. J. Mitchell, *Big incoherent solitons*, Phys. Rev. Lett. **80** (1998), no. 7, 1422–1424.
- [SS98] M. Segev and G. Stegeman, *Self-trapping of optical beams: Spatial solitons*, Phys. Today **51** (1998), no. 8, 42–48.
- [SS99] G. I. Stegeman and M. Segev, *Optical spatial solitons and their interactions: Universality and diversity*, Science **286** (1999), no. 5444, 1518–1523.
- [SSC+00] M. Soljagic, M. Segev, T. Coskun, D. N. Chlistodoulides, and A. Vishwanath, *Modulation instability of incoherent beams in noninstantaneous nonlinear media*, Phys. Rev. Lett. **84** (2000), no. 3, 467–470.
- [SSV+95] M. F. Shih, M. Segev, G. C. Valley, G. Salamo, B. Crosignani, and P. Diporto, *Observation of 2-dimensional steady-state photorefractive screening solitons*, Electron. Lett. **31** (1995), no. 10, 826–827.
- [TS88] M. V. Tratnik and J. E. Sipe, *Bound solitary waves in a birefringent optical fiber*, Phys. Rev. A **38** (1988), no. 4, 2011–2017.
- [TWWS88] S. Trillo, S. Wabnitz, E. M. Wright, and G. I. Stegeman, *Optical solitary waves induced by cross-phase modulation*, Opt. Lett. **13** (1988), no. 10, 871–873.
- [ZAMS96] A. A. Zozulya, D. Z. Anderson, A. V. Mamaev, and M. Saffman, *Self-focusing and soliton formation in media with anisotropic nonlocal material response*, Europhys. Lett. **36** (1996), no. 6, 419–424.
- [ZK65] N. J. Zabusky and M. D. Kruskal, *Interaction of “solitons” in a collisionless plasma and the recurrence of initial states*, Phys. Rev. Lett. **15** (1965), 240–243.
- [ZS71] V. E. Zakharov and A. B. Shabat, *Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media*, Zh. Éksp. Teor. Fiz. **61** (1971), no. 1, 118–134 (in Russian) [English Translation: JETP **34** (1972), no. 1, 62–69].

AUSTRALIAN PHOTONICS CRC, OPTICAL SCIENCES CENTRE, RESEARCH SCHOOL OF PHYSICAL SCIENCES AND ENGINEERING, AUSTRALIAN NATIONAL UNIVERSITY, CANBERRA, ACT 0200, AUSTRALIA

E-mail address: nna124@rsphysse.anu.edu.au

URL: <http://rsphysse.anu.edu.au/~nna124>

NONLINEAR PHYSICS GROUP, RESEARCH SCHOOL OF PHYSICAL SCIENCES AND ENGINEERING, AUSTRALIAN NATIONAL UNIVERSITY, CANBERRA, ACT 0200, AUSTRALIA

E-mail address: ans124@rsphysse.anu.edu.au

URL: <http://rsphysse.anu.edu.au/nonlinear/ans>

APPLIED PHOTONICS GROUP, RESEARCH SCHOOL OF PHYSICAL SCIENCES AND ENGINEERING, AUSTRALIAN NATIONAL UNIVERSITY, CANBERRA, ACT 0200, AUSTRALIA

E-mail address: ana124@rsphysse.anu.edu.au

A Unified Approach to Integrable Systems via Painlevé Analysis

S. Roy Choudhury

ABSTRACT. We review an algorithmic procedure based on truncated Painlevé expansions to derive various features of integrable equations, including Lax Pairs, Hirota's Tau Function, Darboux Transformations and multisoliton solutions, and Miura Transformations to related equations.

1. Introduction

The techniques of Painlevé analysis [1] are by now well-known in the area of testing nonlinear systems for integrability. The purpose of the present review is to focus on one of the developments which have taken place in the field over the past decade or so. These developments have been in several directions. Since the objective here is to be more discursive than in regular research articles, we shall primarily be concerned with one of these, i.e. the formulation of a technique for algorithmically deriving from the Painlevé analysis all properties relevant to integrable systems including auto-Bäcklund Transformations, Lax Pairs, Miura Transformations to related systems (if any), Darboux Transformations, Hirota's Tau function, and multisoliton solutions. In fact, the method has also been applied to the derivation of similarity reductions of nonlinear PDEs (NLPDEs), although we shall not consider this here. It will thus be seen that the techniques of Painlevé analysis are by now much broader than as mere tests for the integrability of a system. In fact, the area has developed to a point where it may be considered as one which yields a major unifying perspective on integrable systems, as well as one which complements the perspectives afforded by other approaches.

In order to keep the treatment to a manageable length while still being reasonably complete, we shall concentrate on the sub-area mentioned above. However, for the sake of readability as well as for the purpose of introducing some relevant concepts and terminology, we shall briefly summarize some earlier work in the area in Section 2. This will hopefully orient the reader better as to where the work detailed in this paper fits within the overall field of Painlevé analysis. Some key references

1991 *Mathematics Subject Classification*. Primary 35Q51, 35Q55; Secondary 35A10, 35A20, 35C05.

Key words and phrases. Painlevé analysis, integrable systems, algorithms for Lax pairs, Darboux transformations, tau functions.

are mentioned in Section 2, although the list is not meant to be exhaustive. Following this, we shall develop the main theme of the paper in Section 3, using integrable (2+1) generalizations of NLS-type systems (and of the Kaup equation in particular) as examples. Section 4 will develop the theme further by considering the additional example of the well-known AKNS system(s). Finally, Section 5 summarizes the main results of the paper. It also contains some comments on the current status and future prospects of this particular sub-area of Painlevé analysis.

2. Background and Basic Concepts

Although not yet fully proven, the Painlevé tests [1] seem to provide extremely useful necessary conditions for identifying the completely integrable cases of a wide variety of families of nonlinear ordinary and partial differential equations, as well as integrodifferential equations. Originally, Ablowitz et al. [2] conjectured that a nonlinear partial differential equation is integrable if *all* its exact reductions to ordinary differential equations have the Painlevé property. This approach poses the obvious operational difficulty of finding *all* exact reductions. This difficulty was circumvented by Weiss et al. [3] by postulating that a partial differential equation has the Painlevé property if its solutions are single-valued about a movable singular manifold. In this paper, we follow this latter approach to perform the Painlevé analysis of several nonlinear evolution equations.

There is now a compelling body of evidence that if an equation possesses the Painlevé property it is likely to be integrable, i.e., the Painlevé test is a necessary test for integrability. In the cases where the criteria for the Painlevé test are met, the analysis may have failed to detect an essential singularity and further analysis would be needed to rigorously prove integrability by:

- (a) constructing the full set of integrals of the motion [5], or
- (b) linearizing the equations, e.g., by the inverse scattering transform [6], or
- (c) reducing them to one of Painlevé's transcendental equations [1, 4, 7].

The usefulness of the Painlevé approach is not limited to integrability prediction, and use of the generalized Weiss algorithm [4, 8] yields auto-Bäcklund transformations and Lax pairs for the integrable cases. Painlevé analysis also yields a systematic procedure for obtaining special solutions when the equation possesses only the conditional Painlevé property [9]-[14], when the compatibility conditions of the Painlevé analysis result in constraint equations for the movable singular manifold which is no longer completely arbitrary.

Weiss' original technique [3, 8] was extensively developed by others (see [15, 16] for instance). This approach, which will be briefly reviewed in this section, involves the Weiss strategy of truncating the Painlevé singularity expansion for the solution of the system of NLPDEs at the constant term, thereby imposing a specific choice of singular manifold function which has come to be called 'the singular manifold'. This singular manifold function and the truncated (singular part) of the Painlevé expansion are then used to semi-algorithmically derive an auto-Bäcklund transformation between two different solutions of the NLPDE(s), and also to derive the associated linear scattering problem or Lax Pair. The latter step is not completely algorithmic since it involves linearizing the overdetermined system of PDEs connecting various derivatives of the singularity manifold by employing a 'Weiss substitution' which may often involve prior, extraneous knowledge about the NLPDE(s) under consideration. References 15 and 16 also discuss the connections

between Painlevé analysis and other properties of, and approaches to, integrable systems such as Lie symmetries and Hirota's method. However, the original semi-algorithmic character of the Weiss SMM persists.

A second recent approach, which has opened up a whole new sub-field, involves making the entire process of singularity analysis invariant under the homographic or Möbius transformation [17, 18]. This significantly simplifies the testing for integrability [18], the derivation of Lax Pairs [19, ?], as well as the derivation of special families of analytic solutions (see [21]-[24] for instance). Some of these special families of analytic solutions have also been employed in tandem with Melnikov theory to analytically investigate the breakdown of coherent structure solutions and the onset of chaos in NLPDEs under forcing [25, 26]. Note that the invariant analysis yields a fully algorithmic procedure for finding Lax pairs, but none for auto-BTs, tau functions, and multisoliton solutions. We shall not consider this approach at all in this review.

A third approach [27, 28] involves significant extensions of the original Weiss procedure to derive the 'Weiss substitution' and the Lax Pair completely algorithmically. In addition, this technique algorithmically derives many other important features of integrable systems such as Miura Transformations, Darboux Transformations, multisoliton solutions, and Hirota's tau function. Much of this work is motivated by the connections sought to be made between the various properties of integrable systems in [15] and [16]. Earlier work along these lines includes [29]. We shall develop this approach systematically in the next two sections.

There has also been other activity in the area in recent years, including investigations of why the Painlevé test works, and on higher-order truncations and so on. We do not refer to these at all here since they do not directly impact the topic of this article.

In the remainder of this section, we briefly review Weiss' original approach [8] in the context of the simplest of the many integrable systems which he considered, viz. the KdV equation

$$(2.1) \quad u'_t + 6u'u'_x + u'_{xxx} = 0$$

We use primed variables here for reasons which will become clear subsequently. The behavior of the solutions of (2.1) around a movable singular manifold [1, 3, 4] $\phi(x, t) = 0$ is determined by a 'leading-order analysis' whereby one makes the ansatz

$$(2.2) \quad u'(x, t) = u_0(x, t)[\phi(x, t)]^{-\alpha}$$

and balances the most singular or dominant terms (the ones with the most negative powers of ϕ ; usually the highest derivatives and some or all of the most nonlinear terms). For (2.1), balancing the most singular parts of the nonlinear and dispersive terms yields [1, 3, 4]

$$(2.3) \quad \alpha = 2, \quad u_0 = -2\phi_x^2$$

At this point, if testing for integrability [1, 3, 4], one determines the so-called resonances or Kowaleskaya exponents or indices, and then attempts to construct a Laurent expansion solution of the NLPDE with the full complement of arbitrary functions. We shall instead follow what has come to be called 'the Singular Manifold Method (SMM)' by following Weiss' [8] original prescription of constructing an expansion in inverse powers of ϕ and truncating it at the constant term. It should be noted that, once this is done and in contrast to the situation while testing for

integrability, the singular manifold is no longer an arbitrary function. Instead, it is a well-defined function which has come to be called ‘the singular manifold’. Weiss’ original idea was that information about the integrable system was encoded in this singular manifold. In particular, he was able to derive auto-Bäcklund Transformations (auto-BTs) between two distinct solutions of the NLPDE(s), as well as the associated Lax Pair, in a semi-algorithmic fashion from the equations satisfied by ϕ and its derivatives. Let us briefly review the steps associated with this next.

Inserting the truncated expansion (truncated at the constant term) in terms of powers of ‘the singular manifold function’

$$(2.4) \quad u' = \frac{-2\phi_x^2}{\phi^2} + \frac{u_1(x, t)}{\phi} + u(x, t)$$

the $O(\phi^{-4})$ terms yield

$$(2.5) \quad u_1 = 2\phi_{xx}$$

Note that the coefficients (u_0 and u_1) of the singular part of the truncated expansion (2.4) are thus fully determined and it is necessary to ensure this before proceeding to the next step. The reason for using the primed variables in the original equation (2.1) should also now be apparent from inspecting (2.4).

Once the coefficients in the singular part of the truncated expansion are determined, we then re-insert the truncated expansion (2.4), together with these explicit singular coefficients (2.5), into (2.1) to obtain the so-called Painlevé-Bäcklund equations

$$(2.6a) \quad u'(x, t) = 2\frac{\partial^2}{\partial x^2} \ln \phi + u$$

$$(2.6b) \quad \phi_x \phi_t + 4\phi_x \phi_{xxx} - 3\phi_{xx}^2 + 6u\phi_x^2 = 0$$

$$(2.6c) \quad \phi_{xt} + \phi_{xxxx} + 6u\phi_{xx} = 0$$

$$(2.6d) \quad u_t + 6uu_x + u_{xxx} = 0$$

The first thing we note from (2.6d), and Weiss proved that this is a generic feature, is that the final coefficient (that for the constant term) in (2.4) satisfies the original KdV equation (2.1). Thus, (2.6a) provides an auto-BT between two solutions u' and u of the KdV equation. This can be used as an explicit auto-BT if one can determine ‘the singularity manifold function’ ϕ . There are several strategies for doing this, and we shall consider one of them in the next section. Thus, the Weiss singularity manifold method (SMM) provides an algorithmic method for finding auto-BT’s.

Next, one might be inclined to think that the remaining Painlevé-Bäcklund equations (2.6 b-d) constitute an overdetermined system of three equations for the two variables $u(x, t)$ and $\phi(x, t)$. They are, however, self-consistent, with the last one being the solvability condition for the other two. Also, they may always be linearized to yield the associated linear scattering problem or Lax Pair for the system. For (2.1), Weiss and his co-workers used two different strategies [3, 8], and other strategies have included those in [19] and [?]. While each of the Weiss strategies has semi-algorithmic features, neither is fully algorithmic and self-consistent in the sense that each utilizes some a priori information about (2.1). We shall use the simpler Weiss strategy here, and develop a more systematic and algorithmic version of the other one in the next two sections. Utilizing the well-known AKNS-type

substitution (Weiss et al [3] used v instead of ψ in their original work)

$$(2.7) \quad \phi_x = \psi^2(x, t)$$

transforms (2.6 b,c) into the well-known Lax Pair

$$(2.8a) \quad \psi_{xx} + (u + \lambda)\psi = 0$$

$$(2.8b) \quad \psi_t + 4\psi_{xxx} + 6u\psi_x + 3u_x\psi = 0$$

for the KdV equation.

One might sum up the results above by saying that this original Weiss SMM thus provided an algorithmic method for obtaining an auto-BT and semi-algorithmic methods for obtaining the Lax Pair for integrable NLPDEs via the use of truncated Painlevé expansions.

It should also be noted that various other aspects of the Weiss SMM have been investigated (see [8], [19], and [20] for instance). In particular, Conte and Musette have used the singular part of the truncated expansion in a so-called ‘singular part transformation’ to algorithmically derive Darboux Transformations for numerous integrable systems.

In the next two sections, we shall follow the third approach mentioned at the beginning of this section and develop the Weiss SMM further into an algorithmic method for deriving various properties of integrable systems. We shall illustrate the requisite techniques using two integrable (2+1) versions of the Kaup equation in the next section and the well-known AKNS system(s) in Section 4. We shall notice some interesting contrasts between what the SMM yields for the systems in Section 3 and that in Section 4.

3. Unified Treatment of Integrable (2+1) Generalizations of the Kaup Equation

In this section, we first develop the analysis using two members of the integrable (2+1) NLS Type systems considered by Mikhailov and his co-workers [30]-[32] as typical examples. We choose systems in (2+1) dimensions intentionally so as to demonstrate both the algorithmic nature of the analysis as well its direct applicability to systems in more than one spatial dimension. As we shall see, the analysis in (2+1) is, as one might expect, somewhat more involved than for (1+1) systems, but in a fashion which may be developed algorithmically and systematically. We shall mention appropriate references as we proceed, but two background papers of general relevance are those by Estevez and her co-workers [27, 28].

In particular, we shall consider the following two integrable generalizations of the Kaup equation [33]:

$$(3.1a) \quad u'_t = u'_{xx} + 2p'u'_x$$

$$(3.1b) \quad -v'_t = v'_{xx} - 2p'v'_x$$

$$(3.1c) \quad p'_y = (u + v)_x$$

and

$$(3.2a) \quad u'_t = u'_{yy} + (u'^2 + u'v')_y + q'$$

$$(3.2b) \quad -v'_t = v'_{yy} - (v'^2 + u'v')_y + q'$$

$$(3.2c) \quad q'_x = (v'u'_x - u'v'_x)_y.$$

3.1. Analysis of (3.1). We shall detail the calculations and the procedure for (3.1) first, and subsequently summarize similar computations for (3.2).

As usual, we first perform the leading-order analysis as in Section 2 and assume

$$(3.3) \quad u' \sim u_0 \phi^{-\alpha}, \quad v' \sim v_0 \phi^{-\beta}, \quad p' \sim p_0 \phi^{-\gamma}.$$

Balancing the most singular second derivative and nonlinear terms in the first two equations yields:

$$(3.4a) \quad \alpha + \beta = -2$$

$$(3.4b) \quad \gamma = 1$$

$$(3.4c) \quad p_0 = (\alpha + 1)\phi_x/2.$$

At this point, it is tempting to look at the apparently symmetric way in which the variables u and v occur in (3.1) and thus assume that $\alpha = \beta = -1$. However, it is straightforward to check that this choice leads to a contradiction. One may obtain consistent choices by a. balancing the left hand side of (3.1c) with the first term on the right, with the other term being less singular, or by b. balancing the left side of (3.1c) with the second term on the right. These correspond respectively to:

$$(3.5) \quad \alpha = 1, \quad \beta = -3$$

or

$$(3.6) \quad \alpha = -3, \quad \beta = 1.$$

We shall detail the case corresponding to (3.4)/(3.5) and summarize the analogous results for (3.4)/(3.6) subsequently. As discussed in Section 2, we shall next invoke the Weiss SMM by substituting expansions for our variables truncated at the constant term (and with coefficients of all singular terms explicitly expressed in terms of derivatives of the singular manifold function), and use the resulting expansions to develop an algorithmic method for deriving various properties of the integrable system (3.1). For (3.4)/(3.5), the leading order $O(\phi^{-3})$ terms in (3.1a,b) yield the coefficients of the singular terms in u and p explicitly as

$$(3.7) \quad u_0 = \phi_y, \quad p_0 = \phi_x.$$

Using these, we substitute the truncated expansion

$$(3.8a) \quad u' = \frac{\phi_y}{\phi} + u$$

$$(3.8b) \quad v' = v_0 \phi^3 + v_1 \phi^4 + \dots$$

$$(3.8c) \quad p' = \frac{\phi_x}{\phi} + p.$$

Substituting these in (3.1) yields equations at various orders in ϕ , the Painlevé-Bäcklund equations, which are contained in Appendix A. Notice that (A9) and (A10) show that

$$(3.9) \quad v_0 = v_1 = 0$$

Thus, (A4) to (A6), and (A8) to (A11), are identically satisfied. The only non-trivial equations surviving are (A1) to (A3) and (A7) which are given below for

ease of comprehension in the following calculations:

$$\begin{aligned}
 (3.10a) \quad & \phi_t - 2p\phi_x - \phi_{xx} = 0 \\
 (3.10b) \quad & -\phi_{yt} + 2\phi_x u_x + 2p\phi_{xy} + \phi_{xxy} = 0 \\
 (3.10c) \quad & p_y = u_x \\
 (3.10d) \quad & -u_t + 2pu_x + u_{xx} = 0
 \end{aligned}$$

Substituting (3.10c) in (3.10b) and integrating with respect to y yields

$$(3.11) \quad -\phi_t + 2p\phi_x + \phi_{xx} = \lambda_*(x, t)$$

which is the same as (3.10a) if the ‘constant’ of integration on the right is taken to be zero. Thus, we may essentially just ignore (3.10b) since it is really the y derivative of (3.10a).

We now work with the remaining Painlevé-Bäcklund equations in (3.10) to derive the so-called singular manifold equation (SME). The essential idea in deriving the SME is to express all physical or field variables (or potentials in the language of scattering) in terms of functions of the singularity manifold and, using these, to derive a consistency condition on this singularity manifold which is the SME. The motivation for this is that analysis of the SME yields an algorithmic method for deriving the Weiss substitution and thus linearizing the Painlevé-Bäcklund equations to obtain the Lax Pair. The details vary from case to case, but the essential ideas in deriving the SME and analyzing it are common to all examples. For this purpose, we also define the quantities [17]-[19]

$$\begin{aligned}
 (3.12a) \quad & V \equiv \phi_{xx}/\phi_x \\
 (3.12b) \quad & C_1 \equiv \phi_t/\phi_x \\
 (3.12c) \quad & C_3 \equiv \phi_y/\phi_x
 \end{aligned}$$

which satisfy the compatibility conditions

$$\begin{aligned}
 (3.13a) \quad & V_t = (C_{1x} + C_1V)_x \quad (\text{from } \phi_{xxt} = \phi_{txx}) \\
 (3.13b) \quad & V_y = (C_{3x} + VC_3)_x \quad (\text{from } \phi_{xxy} = \phi_{yxx}) \\
 (3.13c) \quad & C_{3t} = C_{1y} + C_1C_{3x} - C_3C_{1x} \quad (\text{from } \phi_{yt} = \phi_{ty})
 \end{aligned}$$

Using (3.10c) in (3.10d) yields

$$\begin{aligned}
 (3.14) \quad & u_t = \frac{\partial}{\partial y} [p^2 + p_x] \\
 & u_x = p_y
 \end{aligned}$$

Integrating the consistency condition $u_{xt} = u_{tx}$ with respect to y yields

$$(3.15) \quad p_t = \partial_x [p^2 + p_x] + \lambda(x, t)$$

From (3.10a) and (3.12), we have

$$(3.16) \quad p = \frac{1}{2}(C_1 - V)$$

Using this in (3.15) yields the SME

$$(3.17) \quad \frac{1}{2}(C_1 - V)_t = \partial_x \left[\frac{1}{4}(C_1^2 - 2C_1V + V^2) + \frac{1}{2}(C_{1x} - V_x) \right] + \lambda(x, t)$$

The key to linearizing the Painlevé-Bäcklund equations in algorithmic fashion is to perform a leading-order singularity analysis of the SME and the consistency

conditions (3.13), treated as an NLPDE in C_1 and V in a manner analogous to Section 2. In other words, we apply the first part of the SMM to the SME. Assuming

$$(3.18) \quad C_1 \sim c_0 \chi^a, \quad V \sim v_0 \chi^b$$

and balancing the most singular terms (those within the square bracket) in (3.17) yields

$$(3.19) \quad a = b = -1$$

and

$$(3.20) \quad c_0 = v_0 \quad \text{or} \quad c_0 - v_0 = 2\chi_x$$

Next, using (3.18)/(3.19) and balancing the most singular terms in (3.13a) yields

$$(3.21) \quad v_0 = \chi_x$$

which, with (3.20), implies

$$(3.22) \quad c_0 = 3\chi_x \quad \text{or} \quad c_0 = \chi_x.$$

Once this leading-order analysis of the SME is complete, we follow an approach due to Musette and Conte [20, 34, 35] and assign a separate singularity manifold, i.e., two distinct χ 's, to each of the two branches for C_1 and V in (3.18) to (3.22). Denoting these as ψ^+ and ψ^- (the connection of these to the original singularity manifold ϕ will become apparent in the following step), (3.18) through (3.22) yield the following leading behaviors:

$$(3.23) \quad V \equiv \frac{\phi_{xx}}{\phi_x} = \frac{\psi_x^+}{\psi^+} + \frac{\psi_x^-}{\psi^-}$$

and

$$(3.24) \quad C_1 \equiv \frac{\phi_t}{\phi_x} = \frac{3\psi_x^+}{\psi^+} + \frac{\psi_x^-}{\psi^-}.$$

Integrating (3.23) with respect to x and using the result in (3.24) yields the connection of the original singularity manifold variable ϕ to the ψ s, i.e.

$$(3.25) \quad \phi_x = \psi^+ \psi^-$$

and

$$(3.26) \quad \phi_t = 3\psi_x^+ \psi^- + \psi_x^- \psi^+.$$

These last two equations are in fact the analogues of the Weiss substitutions. Note that, unlike Weiss' original procedure, they have been derived here completely algorithmically and self-consistently from the singularity analysis. More specifically, in Weiss' original procedure [3, 8], such substitutions were based on either guesswork or information regarding the order of the underlying linear scattering problem, both of which were based on extraneous knowledge about the system. As in Section 2, these substitutions will be key to linearizing the Painlevé-Bäcklund equations (3.10a,c,d) to yield the Lax Pair for the system, and we proceed next to this step.

Using (3.23)/(3.24) in (3.16) yields

$$(3.27) \quad \psi_x^+ = p\psi^+.$$

Using (3.27) for $p(x, t)$ in (3.10c) and interchanging the order of the derivatives, the resulting equation may be integrated with respect to x to yield

$$(3.28) \quad \psi_y^+ = [u - \lambda_1(y, t)]\psi^+.$$

The last two equations comprise the spatial part of the Lax Pair for (3.1) (with unprimed variables instead of primed ones). It is straightforward to check that the compatibility condition for (3.27)/(3.28) yields the governing equation (3.1c) for the system (remember that $v = 0$ for this branch of the analysis, as is apparent from (3.8)/(3.9)). Next, solving for u from (3.28) and using the result in the first term in (3.10d) yields

$$(3.29) \quad \lambda_{1t} + \frac{d}{dt} \left[\frac{\psi_y^+}{\psi^+} \right] = 2pu_x + u_{xx}.$$

This constitutes the temporal part of the Lax Pair and it is straightforward to verify that the compatibility condition for (3.28) and (3.29) yields the first governing equation (3.1a) for the system (with unprimed variables), while the compatibility of (3.27) and (3.29) simply yields the x derivative of (3.1a).

Notice that, since (3.10b) is redundant and (3.10a) was used to obtain (3.16) and hence (3.27), we have linearized all the Painlevé-Bäcklund equations (3.10) to obtain the Lax Pair for (3.1) (with unprimed variables). Notice too that (3.1b) is trivially satisfied for this branch of the analysis since $v = 0$ by (3.8b) and (3.9).

At this point, we remind ourselves that the above branch of the singularity analysis of (3.1) corresponds to (3.4)/(3.5). Performing an exactly analogous analysis for the other branch corresponding to (3.4)/(3.6) results in (the ψ^\pm functions in (3.30) to (3.37) are different from those in (3.23) to (3.29)):

a. (3.8)/(3.9) are replaced by

$$(3.30) \quad \begin{aligned} u' &= 0 \\ v' &= -\frac{\phi_y}{\phi} + v \\ p' &= -\frac{\phi_x}{\phi} + p \end{aligned}$$

b. (3.23) (3.26) are replaced by

$$(3.31) \quad V = \frac{\psi_x^+}{\psi^+} + \frac{\psi_x^-}{\psi^-}$$

$$(3.32) \quad C_1 = -\frac{\psi_x^+}{\psi^+} - \frac{3\psi_x^-}{\psi^-}$$

$$(3.33) \quad \phi_x = \psi^+\psi^-$$

$$(3.34) \quad \phi_t = -\psi_x^+\psi^- - 3\psi_x^-\psi^+$$

c. the resulting Lax Pair is

$$(3.35) \quad \psi_x^- = -p\psi^-$$

$$(3.36) \quad \psi_y^- + [v - \lambda_2(y, t)]\psi^- = 0$$

$$(3.37) \quad \lambda_{2t} - \frac{d}{dt} \left[\frac{\psi_y^-}{\psi^-} \right] = 2pv_x - v_{xx}.$$

Note that the consistency of the last three equations recovers the second and third governing equations (3.1b) and (3.1c), while $u = 0$ for this branch of the analysis and so (3.1a) is trivially satisfied.

Thus, the Lax system (3.27)-(3.29) recover the governing equations (3.1a,c), while (3.35)-(3.37) recover (3.1b,c).

Once we have the Lax Pair, the next step in the analysis is to proceed to derive Darboux Transformations [36], i.e. transformations of the potentials u, v , and p and the eigenfunctions ψ which leave the Lax Pair(s) invariant. Once again, a systematic procedure may be formulated from the Weiss SMM. If non-trivial Darboux Transformations (DTs) result, they may then be iterated [36] in the usual manner starting from relatively simple seminal solutions of the governing PDEs following the Crum procedure to generate more complex families of multisoliton solutions. It is worth commenting here that, for many systems, the iteration of DTs appears to work better than the iteration of auto-BTs where one often remains confined to the same family of solutions after a single iteration. In addition, the procedure for deriving DTs may be iterated to generate Hirota's tau function. We shall lay out the basic ideas for the derivation of DTs next.

The key idea in deriving DTs is due to Konopelchenko and Stramp [37] and involves treating the Lax Pair itself as a system of NLPDEs in the field variables (potentials) u, p and the ψ s. We shall primarily follow [27, 28] here. Assuming a singular manifold ϕ_1 , spectral parameter λ_1 and +/- Lax Pair eigenfunctions ψ_1 associated to starting (or seminal) solutions u, v , and p of (3.1) yields the equations:

$$(3.38) \quad \phi_{1x} = \psi_1^+ \psi_1^-$$

$$(3.39) \quad \phi_{1t} = 3\psi_{1x}^+ \psi_1^- + \psi_{1x}^- \psi_1^+$$

$$(3.40) \quad \psi_{1x}^+ = p\psi_1^+$$

$$(3.41) \quad \psi_{1y}^+ = [u - \lambda_1(y, t)]\psi_1^+$$

$$(3.42) \quad \lambda_{1t} + \frac{d}{dt} \left[\frac{\psi_{1y}^+}{\psi_1^+} \right] = 2p u_x + u_{xx}.$$

Here, we have used (3.25) to (3.29). New solutions u' and p' may be constructed using the auto-BTs (3.8a,c) (with ϕ replaced by ϕ_1 corresponding to the seminal solutions), and associating a singular manifold ϕ'_2 , spectral parameter λ_2 and +/- Lax Pair eigenfunctions ψ'_2 to these yields the analogous equations:

$$(3.43) \quad \phi'_{2x} = \psi'^+_2 \psi'^-_2$$

$$(3.44) \quad \phi'_{2t} = 3\psi'^+_{2x} \psi'^-_2 + \psi'^-_{2x} \psi'^+_2$$

$$(3.45) \quad \psi'^+_{2x} = p'\psi'^+_2$$

$$(3.46) \quad \psi'^+_{2y} = [u' - \lambda_2(y, t)]\psi'^+_2$$

$$(3.47) \quad \lambda_{2t} + \frac{d}{dt} \left[\frac{\psi'^+_{2y}}{\psi'^+_2} \right] = 2p' u'_x + u'_{xx}.$$

Next, following [37] and treating the Lax Pair (3.45) to (3.47) as a coupled system of NLPDEs in u', p' and the ψ'_2 s, we may apply the SMM to this system of NLPDEs and thus add the following truncated expansion for the ψ'^+_2 to those in (3.8a,c)

(with ϕ replaced by ϕ_1 for the seminal solutions) to obtain:

$$(3.48) \quad \psi_2'^+ = \psi_2^+ - \frac{\psi_1^+ \theta^+}{\phi_1}$$

$$(3.49) \quad u' = \frac{\phi_{1y}}{\phi_1} + u$$

$$(3.50) \quad p' = \frac{\phi_{1x}}{\phi_1} + p.$$

Now, for a DT, the transformation of potentials and eigenfunctions given by (3.48) to (3.50) must preserve the Lax Pair. In other words, the original starting solutions corresponding to u, p , and ψ_2^+ must satisfy the same Lax Pair equations (3.45) to (3.47) for the same eigenvalue λ_2 , i.e.,

$$(3.51) \quad \psi_{2x}^+ = p\psi_2^+$$

$$(3.52) \quad \psi_{2y}^+ = [u - \lambda_2(y, t)]\psi_2^+$$

$$(3.53) \quad \lambda_{2t} + \frac{d}{dt} \left[\frac{\psi_{2y}^+}{\psi_2^+} \right] = 2pu_x + u_{xx}.$$

Substituting the truncated expansions (3.48) to (3.50) in (3.45) to (3.47) and using (3.38)/(3.39) and (3.51) to (3.53) yields, after some computer algebra with MATHEMATICA, the trivial result:

$$(3.54) \quad \theta^+ = 0.$$

Also, a leading-order singularity analysis of (3.13c), in a manner similar to that performed on (3.13a) while analyzing the SME (3.17) to derive (3.23)/(3.24), shows that

$$(3.55) \quad \phi_y = k(3\psi_x^+ \psi^- + \psi_x^- \psi^+)$$

for some arbitrary k . Using (3.38), (3.54), and (3.55) (with $\phi = \phi_1$) in (3.38) to (3.50) yields the following DT under which the Lax Pair(s) are invariant (and corresponding to $v = 0$ as discussed earlier)

$$(3.56) \quad u' = \frac{k(3\psi_{1x}^+ \psi_1^- + \psi_{1x}^- \psi_1^+)}{\int \psi_1^+ \psi_1^- dx} + u$$

$$(3.57) \quad p' = \frac{\psi_1^+ \psi_1^-}{\int \psi_1^+ \psi_1^- dx} + p$$

$$(3.58) \quad \psi_2'^+ = \psi_2^+.$$

Note that this DT may be iterated starting from simple seminal solutions of (3.1) (with $v = 0$) and using the Crum procedure [36]. In order to do this, one would substitute the simple seminal solutions for u and p in (3.40) to (3.42) to obtain the first iterate for ψ_1^+ . This may then be substituted in (3.56)/(3.57) to yield a second iterate for the potentials u and p , and the process may then be iterated as long as closed-form solutions may still be readily obtained. Before attempting this, we make one other comment. It is possible to iterate the singular manifold function itself to obtain Hirota's tau function. However, (3.58) makes it apparent that, for the present example (3.1), only trivial or identity iterates result for the ψ functions, and hence for the ϕ s (see (3.25)). We therefore postpone the discussion of Hirota's method to the next section, where the situation will turn out to be different from

the one we just considered. For the same reason, we shall postpone consideration of the iteration of DTs to Section 4.

In order to complete the treatment of (3.1), we finally turn to a discussion of the iteration of the auto-BT (3.8) for this equation in order to derive analytic solutions of (3.1). The relevant equations here will be (3.1), (3.8) (with $v_0 = v_1 = 0$), and (3.10a,c). Starting from the simplest vacuum solutions $u = p = 0$ ($v = 0$ anyway for the branch corresponding to (3.10)) as seminal solutions, (3.10a) yields the heat equation (in t and x) for the first iterate of ϕ . Thus,

$$(3.59) \quad \phi(x, y, t) = \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t} c_1(y) + c_2(y).$$

Using this and the seminal solutions in (3.8) yields the next iterate for the solutions, i.e.

$$(3.60) \quad u' = \frac{c_1'(y) + 2\sqrt{\pi t} c_2'(y) e^{x^2/4t}}{c_1(y) + 2\sqrt{\pi t} c_2(y) e^{x^2/4t}}$$

$$(3.61) \quad p' = \frac{xc_1(y)}{-2tc_1(y) - 4\sqrt{\pi t^3} c_2(y) e^{x^2/4t}}.$$

It is straightforward to check that these are indeed solutions of (3.1) for $v = 0$ and arbitrary $c_1(y)$ and $c_2(y)$. One may try and iterate the process by using the last two equations in (3.10a) to obtain a second iterate for ϕ , but the solution becomes complicated and so we shall stop at this point. Figures 1 to 3 show plots of the solutions (3.60) and (3.61) for:

- a. plot of p for $c_1(y) = \pi, c_2(y) = y^5$ at $t = 1$;
- b. plot of u for $c_1(y) = \exp(-y^2/4t), c_2(y) = y^5$ at $t = 1$;
and
- c. plot of p for $c_1(y) = c_2(y) = \exp(-y^2/4t)$ at $t = 1$;

These solutions will be extensively discussed elsewhere, but note in particular the strong x and y modulations in Figure 1, as well as the y/x modulations respectively in Figures 2 and 3.

This concludes our treatment of (3.1), and we turn next to a relatively brief treatment of (3.2). In order to illustrate other features of the SMM method under consideration, we shall refer to features of (3.2) which are analogous to those seen above for (3.1) only briefly. Our main concentration will be on features dissimilar to those discussed for (3.1)

3.2. Brief analysis of (3.2). Attempting a leading-order analysis of (3.2) by substituting

$$(3.62) \quad u' \sim u_0 \phi^{-\alpha}, \quad v' \sim v_0 \phi^{-\beta}, \quad q' \sim q_0 \phi^{-\gamma}$$

it is straightforward to check the possible consistent dominant balances and conclude the following:

- a. as for (3.1) (see (3.5)/(3.6)), consistent dominant balances exist with α and β having unequal values. We do not consider these cases or branches of the singularity analysis further as they are similar to the treatment in Section 3.1 and will be detailed elsewhere.

and

b. unlike (3.1), (3.2) admits a consistent dominant balance with

$$(3.63) \quad \begin{aligned} \alpha &= \beta = 1 \\ \gamma &= 3. \end{aligned}$$

We shall concentrate on this branch as it illustrates somewhat different features of the analysis from those discussed in Section 3.1.

The leading-order analysis for the branch discussed in b above yields

$$(3.64) \quad \{u_0, v_0, 0\} = \begin{cases} (0, 0, 0) \\ \text{or} \\ (0, -\phi_y, 0) \\ \text{or} \\ (\phi_y, 0, 0) \end{cases} .$$

Using the last of these together with (3.62)/(3.62), and substituting the resulting truncated expansions

$$(3.65a) \quad u' = \frac{\phi_y}{\phi} + u$$

$$(3.65b) \quad v' = v$$

$$(3.65c) \quad q' = \frac{q_1}{\phi^2} + \frac{q_2}{\phi} + q$$

into (3.2) results in equations at different orders in powers of ϕ (analogous to those in Appendix A for (3.1)). Solving these as in Section (3.1) yields

$$(3.66) \quad v = v(y, t)$$

$$(3.67) \quad q_1 = -v\phi_y^2$$

$$(3.68) \quad q_2 = \phi_y v_y + v\phi_{yy}$$

together with the conditions

$$(3.69) \quad -2v\phi_y^2 + \phi_t\phi_y - 2u\phi_y^2 - \phi_y\phi_{yy} = 0$$

$$(3.70) \quad 2\phi_y v_y + 2v\phi_{yy} + 2\phi_y u_y - \phi_{yt} + 2u\phi_{yy} + \phi_{yyy} = 0.$$

It is straightforward to check that (3.70) is the y partial of (3.69) and this ‘apparent overdeterminedness’ might seem reminiscent of that observed in (2.6b-d) for the KdV equation. There is however an important difference from that case. Careful inspection of (3.66) to (3.69) (and (3.65c)) reveals an insufficient number of equations to eliminate all field variables (or potentials) u, v , and q and derive an SME. In fact, this is characteristic of a singular branch of the Painlevé analysis. Such a branch may not be used to algorithmically derive the various properties of the integrable system (3.2) as was done using a general or regular branch of (3.1) in the previous subsection. However, it may still be used to derive special analytic solutions (these are usually referred to as ‘singular’ solutions, but in the sense of solutions not contained in the general solution and not necessarily in the sense of possessing singularities). We shall use the governing equations above for the chosen singular branch of the Painlevé analysis to derive special analytic solutions of (3.2). The procedure used will be iteration of the auto-BT (3.65), as was done for (3.1) at the end of Section 3.1, and we proceed to this next.

The relevant equations are (3.65) to (3.69). Starting with vacuum solutions $u = v = q = 0$ of (3.2), (3.69) yields the heat equation (in t and y) for ϕ . Solving this yields

$$(3.71) \quad \phi(x, y, t) = \frac{e^{-y^2/4t}d_1(x)}{\sqrt{4\pi t}} + d_2(x).$$

Using this and (3.66) to (3.68) in (3.65) yields the next iterate

$$(3.1) \quad u' = \frac{y d_1(x)}{-2t d_1(x) - 4\sqrt{\pi t^3} d_2(x) e^{y^2/4t}}$$

$$(3.72) \quad q' = 0$$

for solutions of (3.2). It is straightforward to check that these satisfy (3.2). A typical plot of the solution in (3.72) is shown in Figure 4 for:

$$\text{plot of } u \text{ for } d_1(x) = \exp(-x^2/4t), \quad c_2(x) = x^5 \text{ at } t = 10.$$

This concludes our discussion of the $(2 + 1)$ dimensional generalizations (3.1) and (3.2) of the Kaup equation. We proceed next to a consideration of the well-known AKNS equation(s) in the following section.

4. The AKNS Equation in $(1 + 1)$ -dimensions

We shall use the AKNS system [38] in $1 + 1$

$$(4.1) \quad M_{yxxx} + 4M_y M_{xx} + 8M_x M_{xy} = 0,$$

to illustrate further features of the method. This section of the review follows [27]. Some of the steps which are similar to Section 4 are omitted.

The leading-order analysis and truncation at the constant level yields

$$(4.2) \quad M' = M + \frac{\phi_x}{\phi}.$$

The substitution of the truncated expansion (4.2) into equation (4.1) provides the following results (details are omitted):

- M as well as M' should be solutions of (4.1). This means that (4.2) could be considered as an auto-Bäcklund transformation between two solutions M' and M of the same equation.

- The solution M can be written in terms of the singular manifold in the following way:

$$(4.3) \quad M_x = - \left(\frac{1}{4} \right) \left(V_x + \frac{V^2}{2} + 2\lambda \right),$$

$$(4.4) \quad M_y = \frac{1}{2} (-V_y + 2\lambda C_3).$$

These are the Painlevé-Bäcklund equations and λ is an arbitrary constant of integration that, as in Section 3, plays the role of the spectral parameter.

- The singular manifold equations. The equations that the truncation procedure implies for ϕ are:

$$(4.5) \quad S_y = 4\lambda C_{3x},$$

where S is the schwartzian derivative defined as [17]:

$$(4.6) \quad S = V_x - \frac{V^2}{2}.$$

Also, the compatibility condition $\phi_{xxt} = \phi_{txx}$ between definitions (3.12a) and (3.12c) requires:

$$(4.7) \quad S_y = C_{3xxx} + 2SC_{3x} + C_3S_x.$$

It is straightforward to show that the singular manifold equations are just the AKNS system once again. In fact with the change of variables

$$(4.8) \quad S = 4p_x + 2\lambda,$$

$$(4.9) \quad C_3 = \frac{p_y}{\lambda},$$

(4.5) is trivially satisfied and (4.7) yields

$$0 = p_{yxxx} + 4p_y p_{xx} + 8p_x p_{xy}$$

which is the AKNS system once again.

As we have seen above, the singular manifold equations, written in terms of V and C_3 are (see (3.13b)):

$$(4.10) \quad V_{xy} - VV_y = 4\lambda C_{3x},$$

$$(4.11) \quad V_y = (C_{3x} + C_3V)_x$$

which can be considered as a new system of nonlinear equations as in Section 4. If we apply the Painlevé analysis to this system to derive the Weiss substitution, the leading terms are (using ψ for the singularity manifold)

$$V \sim V_0\psi^a, \quad C_3 \sim C_{30}\psi^b.$$

Using these in (4.10-4.11) yields

$$a = -1, \quad b = -2, \quad V_0 = 2\psi_x, \quad C_{30} = -\frac{1}{\lambda}\psi_x\psi_y.$$

As in Section 4, these leading terms provide the key for the linearization of the truncated solutions (4.3-4.4). If we replace V by its dominant term

$$(4.12) \quad \frac{\phi_{xx}}{\phi_x} = V_0\psi^a = 2\frac{\psi_x}{\psi} \Rightarrow \phi_x = \psi^2.$$

Thus, (4.3) becomes

$$(4.13) \quad 0 = \psi_{xx} + (2M_x + \lambda)\psi$$

and from (4.12), (4.4) and (4.7) we obtain

$$2\frac{\psi_y}{\psi} = C_{3x} + C_3V = \frac{1}{2\lambda}(2M_{xy} + V_{xy} + 2VM_y + VV_y)$$

or (using $V = V_0\psi^a = 2\psi_x/\psi$)

$$(4.14) \quad 0 = 2\lambda\psi_y + M_{xy}\psi - 2M_y\psi_x.$$

(4.13) and (4.14) are precisely the Lax pair for the AKNS system.

As in Section 4, we can consider the Lax pair itself as a pair of coupled nonlinear equations between M and ψ . Let us now explain how to proceed to find a Darboux Transformation and Hirota's tau function using this.

Since M' is also a solution of (4.1), an associated singular manifold ϕ'_2 linked to a spectral parameter λ_2 can be defined just by defining (from (4.12))

$$(4.15) \quad \phi'_{2x} = \psi'^2_2,$$

and a Lax pair for M' can be written as (from (4.13) and (4.14))

$$(4.16) \quad 0 = \psi'_{2xx} + (2M'_x + \lambda_2)\psi'_2,$$

$$(4.17) \quad 0 = 2\lambda_2\psi'_{2y} + M'_{xy}\psi'_2 - 2M'_y\psi'_{2x},$$

where the notation means that ψ'_2 is an eigenfunction corresponding to M' with eigenvalue λ_2 . If we call ϕ_1 and ϕ_2 two singular manifolds for M attached to spectral parameters λ_1 and λ_2 respectively the corresponding eigenfunctions are defined from (4.12) as:

$$(4.18) \quad \phi_{1x} = \psi_1^2,$$

$$(4.19) \quad \phi_{2x} = \psi_2^2.$$

Using (4.13)/(4.14), the Lax pairs take the form

$$(4.20) \quad 0 = \psi_{1xx} + (2M_x + \lambda_1)\psi_1,$$

$$(4.21) \quad 0 = 2\lambda_1\psi_{1y} + M_{xy}\psi_1 - 2M_y\psi_{1x},$$

$$(4.22) \quad 0 = \psi_{2xx} + (2M_x + \lambda_2)\psi_2,$$

$$(4.23) \quad 0 = 2\lambda_2\psi_{2y} + M_{xy}\psi_2 - 2M_y\psi_{2x}.$$

If we use the singular manifold ϕ_1 to construct the truncated Painlevé expansion (ϕ_2 may be used with $\lambda = \lambda_2$ instead)

$$(4.24) \quad M' = M + \frac{\phi_{1x}}{\phi_1}$$

and we then treat (4.16-4.17) as a system of nonlinear coupled equations, a similar expansion should be performed for ψ'_2 . In other words,

$$(4.25) \quad \psi'_2 = \psi_2 + \frac{\Theta}{\psi_1}.$$

The substitution of the truncated expansions (4.24-4.25) in (4.16)/(4.17) provides the functional form for Θ . The result is

$$(4.26) \quad \Theta = -\psi_1\Omega(\psi_1, \psi_2),$$

where

$$(4.27) \quad \Omega(\psi_1, \psi_2) = \left(\frac{1}{\lambda_1 - \lambda_2} \right) (\psi_1\psi_{2x} - \psi_2\psi_{1x}).$$

The expansions (4.24)/(4.25) leave invariant the Lax pair (4.13-4.14), or are Darboux transformations. The eigenfunctions and singular manifolds are trivially related through (4.18) and we therefore have the following: with two eigenfunctions ψ_1 and ψ_2 for M , we can construct an eigenfunction ψ'_2 for the iterated solution M' using (4.25)-(4.27). Hence they provide a standard Darboux transformation.

Furthermore, (4.15) is a nonlinear equation that relates ϕ'_2 and ψ'_2 . This means that the singular manifold ϕ'_2 itself could also be expanded in terms of ϕ_1

$$(4.28) \quad \phi'_2 = \phi_2 + \frac{\Delta}{\phi_1}$$

and by substituting this expansion in (4.15) and using (4.25)-(4.27), we obtain

$$(4.29) \quad \Delta = -[\Omega(\psi_1, \psi_2)]^2.$$

The procedure described above may be easily iterated. The singular manifold ϕ'_2 for M' can be used to construct a new solution

$$(4.30) \quad M'' = M' + \frac{\phi'_{2x}}{\phi'_2}$$

that combined with (4.24) can be written as

$$(4.31) \quad M'' = M + \frac{\tau_{12x}}{\tau_{12}},$$

where

$$(4.32) \quad \tau_{12} = \phi'_2 \phi_1$$

and by using (4.28) and (4.29)

$$(4.33) \quad \tau_{12} = \phi_2 \phi_1 - [\Omega(\psi_1, \psi_2)]^2.$$

Note that the function τ_{12} for the second iteration is not a singular manifold but it can be constructed from two singular manifolds of the first iteration. Thus, the SMM algorithmically yields Hirota's bilinear method [39]. It also provides the algorithm to construct solutions for the τ -function, as we will see below.

The easiest nontrivial solutions can be obtained from the seminal solution

$$(4.34) \quad M = a_0 y$$

(so that M_x and M_{xy} are zero in (4.20-4.23). For this solution, exponential solutions of (4.20-4.23) are

$$(4.35) \quad \psi_i = \exp\left(k_i x - \frac{a_0}{k_i} y\right),$$

where

$$(4.36) \quad \lambda_i = -k_i^2$$

and, integrating (4.18)/(4.29) after using (4.35), the corresponding manifolds are

$$(4.37) \quad \phi_i = \frac{1}{2k_i}(\alpha_i + \psi_i^2),$$

where α_i are arbitrary constants. Now, (4.27) implies

$$(4.38) \quad \Omega(\psi_1, \psi_2) = \frac{1}{k_1 + k_2} \psi_1 \psi_2$$

and (4.33) also yields

$$(4.39) \quad \tau_{12} = \frac{1}{4k_1 k_2} (\alpha_1 + \psi_1^2)(\alpha_2 + \psi_2^2) - \frac{\psi_1^2 \psi_2^2}{(k_1 + k_2)^2}.$$

Using (4.24) and (4.31), we can write the first and second iterates as:

$$(4.40) \quad M' = a_0 y + \frac{\phi_{1x}}{\phi_1},$$

$$(4.41) \quad M'' = a_0 y + \frac{\tau_{12x}}{\tau_{12}},$$

where

$$(4.42) \quad \phi_1 = \frac{\alpha_1}{2k_1}(1 + F_1),$$

$$(4.43) \quad \tau_{12} = \frac{\alpha_1\alpha_2}{4k_1k_2}\{1 + F_1 + F_2 + A_{12}F_1F_2\},$$

$$(4.44) \quad \alpha_i = \exp(2k_ix_{0i}),$$

$$F_i = \exp\left(2k_i\left(x - \frac{a_0}{k_i^2}y - x_{0i}\right)\right),$$

$$(4.45) \quad A_{12} = \left(\frac{k_1 - k_2}{k_1 + k_2}\right)^2.$$

(4.40) corresponds to the one-soliton solution and (4.41) to the interaction of two solitons for the AKNS system in $1 + 1$.

5. Conclusions and Prospects

In this review, we have considered a technique which has evolved over the last decade or so and which provides a method of algorithmically deriving various properties of integrable systems from truncated Painleve expansions. As should be apparent from the examples we have considered, the technique has by now evolved to a point where it affords one form of unifying perspective on integrable systems, and also provides an algorithmic method for investigating new integrable systems such as new integrable hierarchies of equations.

Future work will probably seek to develop and refine the method further. In addition, it will probably continue to be used to investigate new integrable equations or hierarchies.

References

- [1] A. Ramani, B. Grammaticos and T. Bountis, The Painlevé property and singularity analysis of integrable and non-integrable systems, *Phys. Rep.* **180** (1989), 160.
- [2] M.J. Ablowitz, A. Ramani and H. Segur, A connection between nonlinear evolution equations and ODEs of P-type: I and II, *J. Math. Phys.* **21** (1980), 715, 1006.
- [3] J. Weiss, M. Tabor and G. Carnevale, The Painlevé property for partial differential equations, *J. Math. Phys.* **24** (1983), 522.
- [4] M. Tabor, *Chaos and Integrability in Nonlinear Dynamics*, Wiley, New York, 1989; and in *Soliton Theory: A Survey of Results*, A.P. Fordy (Ed.), Manchester U. Press, Manchester, 1990.
- [5] J. Hietarinta, Direct methods for the search of the second invariant, *Phys. Rep.* **147** (1987), 87.
- [6] M.J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform*, SIAM, Philadelphia, 1981; R.K. Dodd, J.C. Eilbec, J.D. Gibbon and H.C. Morris, *Solitons and Nonlinear Wave Equations*, Academic Press, New York, 1982.
- [7] E.L. Ince, *Ordinary Differential Equations*, Dover, New York, 1956.
- [8] J. Weiss, The Painlevé property for PDEs II: BTs, Lax pairs, and the Schwarzian derivative; On classes of integrable systems and the Painlevé property, *J. Math. Phys.* **24** (1983), 1405; **25** (1984), 13, 2226; **26** (1985), 258, 2174; **27** (1986), 1293, 2647; **28** (1987), 2025.
- [9] N.A. Kudryashov, Exact solutions of the generalized KS equation, *Phys. Lett. A* **147** (1990), 287.
- [10] R. Conte and M. Musette, Painlevé analysis and BT in the Kuramoto-Sivashinsky equation, *J. Phys. A* **22** (1989), 169.
- [11] F. Cariello and M. Tabor, Painlevé expansions for nonintegrable evolution equations, *Physica D* **39** (1989), 77.

- [12] S. Roy Choudhury, Painlevé analysis and special solutions of two families of reaction-diffusion equations, *Phys. Lett. A* **159** (1991), 311.
- [13] B. Yu Guo and Z. Xiong, Analytic solutions of the Fisher equation, *J. Phys. A* **24** (1991), 645.
- [14] S. Roy Choudhury, BTs, truncated Painlevé expansions and special solutions of nonintegrable long-wave evolution equations, *Can. J. Phys.* **70** (1992), 595; Painlevé analysis and partial integrability of a class of reaction-diffusion equations, *Nonlin. Anal. Theory, Meth. & Appl.* **18** (1992), 445.
- [15] A.C. Newell, M. Tabor and Y.B. Zeng, A unified approach to Painlevé expansions, *Physica D* **29** (1987), 1.
- [16] H. Flaschka, A.C. Newell and M. Tabor, Monodromy- and spectrum-preserving deformations, in *What is Integrability*, V.E. Zakharov (Ed.), Springer, Berlin, 1991.
- [17] E. Hille, *ODEs in the Complex Domain*, Wiley, New York, 1976.
- [18] R. Conte, Invariant Painlevé analysis of PDEs, *Phys. Lett. A* **140** (1989), 383.
- [19] M. Musette and R. Conte, Algorithmic method for deriving Lax pairs from the invariant Painlevé analysis of NLPDEs, *J. Math. Phys.* **32** (1991), 1450.
- [20] M. Musette and R. Conte, The two-singular-manifold method: I. MKdV and sine-Gordon equations, *J. Phys. A: Math. Gen.* **27** (1994), 3895.
- [21] S. Roy Choudhury, Invariant Painlevé analysis and coherent structures of two families of reaction-diffusion equations, *J. Math. Phys.* **40** (1999), 3643.
- [22] S. Roy Choudhury, One and 2D coherent structures of the Zakharov-Kuznetsov equations, *Problems of Nonlin. Anal.* **6** (2000), 1.
- [23] N. Isldore and W. Malfliet, New special solutions of the ‘Brusselator’ reaction model, *J. Phys. A: Math. Gen.* **30** (1997), 5151.
- [24] R. Conte and M. Musette, Linearity inside nonlinearity: Exact solutions of the 1D Quintic CGL equation, *Physica D* **69** (1993), 1.
- [25] S. Roy Choudhury, A novel analytic technique for the breakdown of soliton solutions of NLPDEs and integrodifferential equations under forcing, *J. Math. Phys.*, submitted.
- [26] S. Roy Choudhury, A novel analytic technique for the disintegration of coherent structures of the ZK equation under forcing, *J. Phys. A: Math. Gen.*, submitted.
- [27] P.G. Estevez, E. Conde and P.R. Gordoa, Unified approach to Miura, Bäcklund and Darboux transformations for NLPDEs, *J. Math. Phys.* **5** (1998), 82.
- [28] P.G. Estevez and P.R. Gordoa, Darboux transformations via Painlevé analysis, *Inverse Problems* **13** (1997), 939.
- [29] M. Wadati, H. Sanuki, and K. Konno, Relationships among the inverse method, BTs, and an infinite number of conservation laws, *Prog. Theor. Phys.* **53** (1975), 419; A.C. Newell, *Solitons in Math. and Physics*, SIAM, Philadelphia, 1985; A.C. Newell, T. Ratiu, M. Tabor and Z. Yunbo, *Seminaire de Mathematiques Superieures*, Les Presses de L’Université de Montreal, Montreal, 1986.
- [30] A.V. Mikhailov and R.I. Yamilov, On integrable 2-D generalizations of NLS type equations, *Phys. Lett. A* **230** (1997), 295.
- [31] A.V. Mikhailov, A.B. Shabat and R.I. Yamilov, Extension of the module of invertible transformations. Classification of integrable systems, *Comm. Math. Phys.* **115** (1988), 1.
- [32] A.B. Shabat and R.I. Yamilov, To a transformation theory of two-dimensional integrable systems, *Phys. Lett. A* **227** (1997), 15.
- [33] D.J. Kaup, Finding eigenvalue problems for solving nonlinear evolution equations, *Progr. Theor. Phys.* **54** (1975), 72.
- [34] M. Musette and R. Conte, Factorization of the ‘classical Boussinesq’ system, *J. Phys. A: Math. Gen.* **27** (1994), 3895.
- [35] R. Conte, M. Musette and A. Pickering, The two-singular manifold method: II classical Boussinesq system, *J. Phys. A: Math. Gen.* **28** (1995), 179.
- [36] V.B. Matveev and M.A. LaSalle, *Darboux Transformations and Solitons*, Springer, Berlin, 1991.
- [37] B.G. Konopelchenko and W. Stramp, On the structure and properties of the singular manifold equations of the KP hierarchy, *J. Math. Phys.* **32** (1991), 40.
- [38] M.J. Ablowitz, D.J. Kaup, A.C. Newell and H. Segur, The inverse scattering transform—Fourier analysis for nonlinear problems, *Stud. Appl. Math.* **53** (1974), 249.

- [39] R. Hirota, Direct methods in soliton theory, in *Topics in Current Physics* **17** (R. Bullough and P. Caudrey, Eds.) (1980), 157.

Appendix A

Substituting (3.8) in (3.1) yields:

From (3.1a):

$$(A.1) \quad O\left(\frac{1}{\phi^2}\right) : \phi_y[\phi_t - 2p_1\phi_x - \phi_{xx}] = 0$$

$$(A.2) \quad O\left(\frac{1}{\phi}\right) : -\phi_{yt} + 2\phi_x u_{1x} + 2p_1\phi_{xy} + \phi_{xxy} = 0$$

$$(A.3) \quad O(1) : -u_{1t} + 2p_1u_{1x} + u_{1xx} = 0$$

From (3.1b):

$$(A.4) \quad O(1) : v_{1t} - 2p_1v_{1x} + v_{1xx} = 0$$

$$(A.5) \quad O(\phi) : 4v_1\phi_t + v_{0t} - 8p_1v_1\phi_x - 2p_1v_{0x} + 6\phi_x v_{1x} + 4v_1\phi_{xx} + v_{0xx} = 0$$

$$(A.6) \quad O(\phi^2) : 4\phi_x(v_1\phi_x + v_{0x}) + 3v_0\{\phi_t - 2p_1\phi_x + \phi_{xx}\} = 0$$

From (3.1c):

$$(A.7) \quad O(1) : -p_{1y} + u_{1x} = 0$$

$$(A.8) \quad O(\phi) : 0 = 0$$

$$(A.9) \quad O(\phi^2) : 3v_0\phi_x = 0$$

$$(A.10) \quad O(\phi^3) : 4v_1\phi_x + v_{0x} = 0$$

$$(A.11) \quad O(\phi^4) : v_{1x} = 0$$

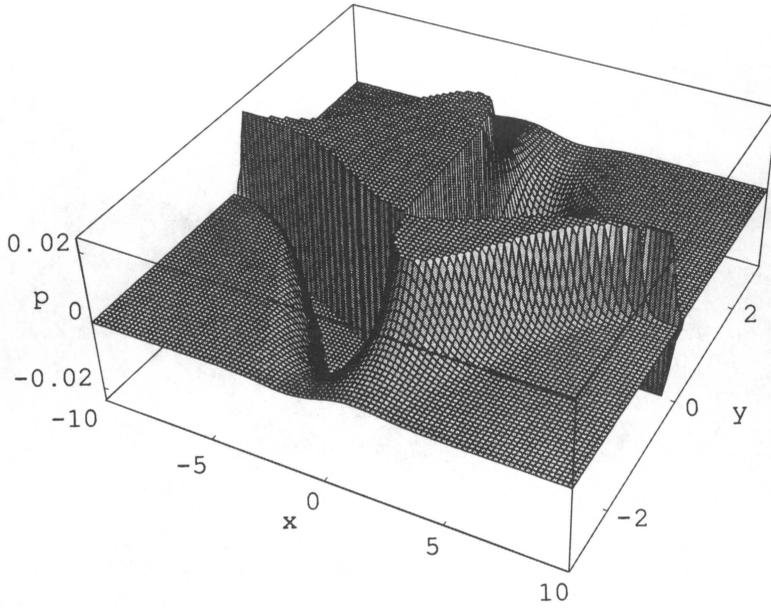


Fig. 1 Plot of p for $c_1(y) = \pi, c_2(y) = y^5$ at $t = 1$ for Eq. (3.1)

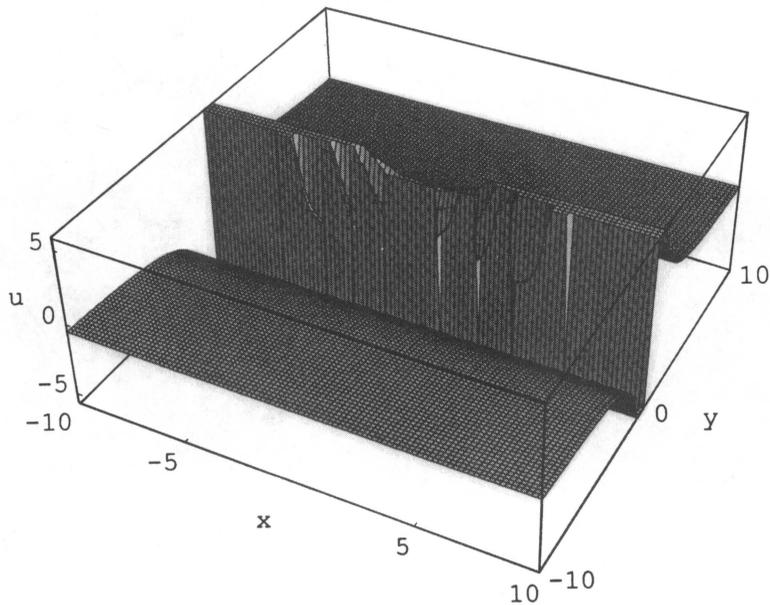


Fig. 2 Plot of u for $c_1(y) = \exp(-y^2/4t), c_2(y) = y^5$ at $t = 1$ for Eq. (3.1)

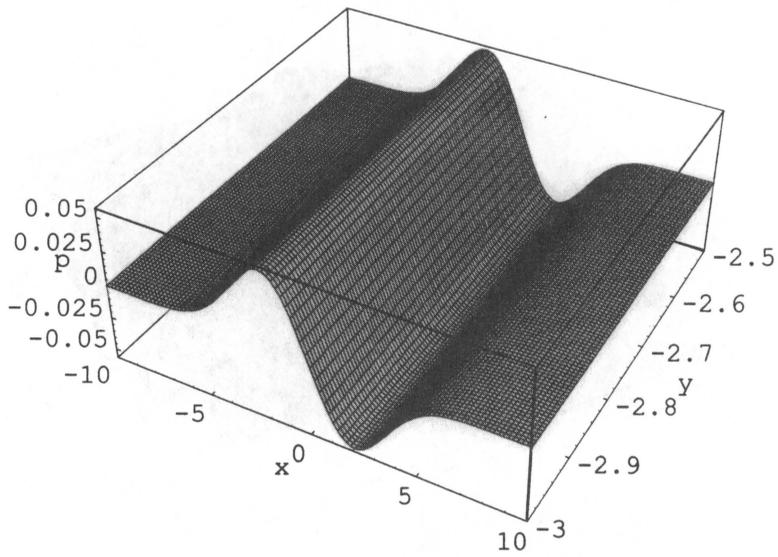


Fig. 3 Plot of p for $c_1(y) = c_2(y) = \exp(-y^2/4t)$ at $t = 1$ for Eq. (3.1)

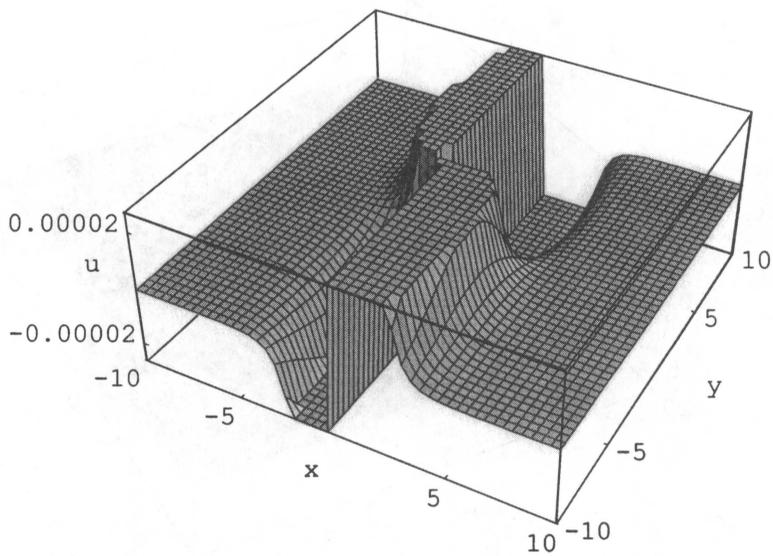


Fig. 4 Plot of u for $d_1(x) = \exp(-x^2/4t)$, $c_2(x) = x^5$ at $t = 10$ for Eq. (3.2)

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CENTRAL FLORIDA, ORLANDO, FL 32816-1364

E-mail address: `choudhur@longwood.cs.ucf.edu`

This page intentionally left blank

Asymptotic stability of solitary waves for Nonlinear Schrödinger equations

Vladimir S. Buslaev and Catherine Sulem

ABSTRACT. This review deals with long-time behavior of solutions of nonlinear Schrödinger equations for initial conditions in a small neighborhood of a stable solitary wave. Under some hypothesis on the structure of the spectrum of the linearized operator near the soliton, the solution decomposes, asymptotically in time, into a solitary wave with slightly modified parameters and a dispersive part described by the free Schrödinger equation. Time behavior of the correction is explicitly calculated.

1. Introduction

This survey deals with the scattering theory of the Nonlinear Schrödinger (NLS) equation in one space dimension

$$(1.1) \quad i\psi_t = -\psi_{xx} + F(|\psi|^2)\psi, \quad x \in \mathbb{R}$$

$$(1.2) \quad \psi(x, 0) = \psi_0(x)$$

where $\psi(x, t)$ is complex-valued function. We suppose that it possesses solitary wave solutions of the form

$$(1.3) \quad \psi(x, t) = e^{i\omega t}\varphi(x, \omega)$$

where φ is the positive solution of the equation

$$(1.4) \quad \varphi'' - \omega\varphi - F(\varphi^2)\varphi = 0$$

vanishing exponentially at infinity. The problem of (orbital) stability of solitary waves for nonlinear dispersive equations has been the object of numerous works [22] [4] [11] [19], [5].

The question we address here is their asymptotic stability that is the long-time behavior of solutions whose initial conditions are close to a stable solitary wave. In the case of integrable nonlinear equations (such as Korteweg-de Vries equation, cubic Schrödinger equation, Benjamin-Ono equation...), the inverse scattering method, under certain conditions, decouples the localized part and the dispersive part and provides an asymptotic decomposition of the solution into a sum of solitary

1991 *Mathematics Subject Classification.* Primary 35Q55.

Key words and phrases. Nonlinear Schrödinger Equations, Solitary Waves, Asymptotic Stability.

The first author acknowledges support from INTAS grant 97-10812.

The second author acknowledges support from NSERC operating grant OGP0046179.

waves and a dispersive component. Here, we deal with non-integrable equations and the approach is completely different and local.

The scattering theory of the NLS equation in \mathbb{R}^n

$$(1.5) \quad i\psi_t + \Delta\psi - \epsilon|\psi|^l\psi = 0, \quad \epsilon > 0$$

$l > 1 + 2/n$ with respect to the free Schrödinger equation

$$(1.6) \quad i\psi_t + \Delta\psi = 0,$$

has been extensively studied by various authors [9], [17], [12]. This case corresponds to the absence of bound states. In dimension $n = 1$, the case $l = 2$ is somehow critical, with the potential term $|\psi|^2 \sim 1/t$ being long range, leading to a $\log t$ phase shift for the asymptotic behavior of the solution [23], [7]. Deift and Zhou [8] considered a perturbation

$$(1.7) \quad i\psi_t + \psi_{xx} - 2|\psi|^2\psi - \epsilon|\psi|^l\psi = 0, \quad l > 2, \quad \epsilon > 0$$

of the defocusing one-dimensional cubic Schrödinger equation

$$(1.8) \quad i\psi_t + \psi_{xx} - 2|\psi|^2\psi = 0.$$

Since (1.8) is completely integrable, they viewed (1.7) as a perturbation of an infinite dimensional integrable system on the line. They proved that as $t \rightarrow \infty$, solutions of (1.7) behave like solutions of (1.8) and that the long-time behavior is universal for a large class of initial data.

In [15], Soffer and Weinstein considered the NLS equation with a potential term

$$(1.9) \quad i\psi_t + \Delta\psi = (V(x) + \lambda|\psi|^{m-1})\psi,$$

for $x \in \mathbb{R}^n$, and $1 < m < (n+2)/(n-2)$. Under the assumptions that $V(x)$ decays fast enough at infinity and that the operator $-\Delta + V$ has exactly one bound state (isolated eigenvalue) in $L^2(\mathbb{R}^n)$, with strictly negative eigenvalue E_* , they proved that for a class of initial conditions, the solution of (1.9) is given by $\psi = e^{-i\Theta(t)}\varphi_{E(t)} + f(t)$, $\Theta = \int_0^t E(s)ds - \gamma(t)$, where φ_E is a spatially localized solitary wave and f a purely dispersive wave. As $t \rightarrow \pm\infty$, $E(t) \rightarrow E_\pm$ and $\gamma(t) \rightarrow \gamma_\pm$. The case where the operator $-\Delta + V$ has 2 bound states was investigated recently by Tsai and Yau, using ideas of [16] developed in the context of resonance solutions of the nonlinear Klein-Gordon equation. In [20], they proved that, in the three-dimensional case, if the initial condition is sufficiently small and near a nonlinear ground state, then the solution approaches a certain nonlinear ground state as time goes to infinity. In subsequent papers [21], they consider the case of initial conditions near an excited state and give a sufficient condition on the initial data so that the solution approaches a nonlinear ground state as time goes to infinity. For certain finite codimension subset in the space of initial data, they also construct solutions converging to the excited states.

The analysis presented here, was initiated in [1] and is based on the spectral decomposition of the solution on the eigenspaces associated to the discrete and continuous spectrum of the linearized operator near the solitary wave. It is convenient to rewrite the NLS equation in the vectorial form

$$(1.10) \quad j\psi_t = -\psi_{xx} + F(|\psi|^2)\psi,$$

$$(1.11) \quad \psi(x, 0) = \psi_0(x)$$

where $j = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$, $\psi_1 = \operatorname{Re} \psi$, $\psi_2 = \operatorname{Im} \psi$. For simplicity, we restrict ourselves to *even solutions*.

Assumption (NL): We suppose that the nonlinearity $F(s)$ is a C^r -function of $s \geq 0$, such that $s = 0$ is a root of multiplicity r with $r \geq 4$, and that for $s > 1$, it satisfies the lower estimate

$$(1.12) \quad F(s) \geq -F_1 s^q, \quad \text{with } F_1 > 0, q < 2.$$

Assumption (NL) ensures that for an initial condition ψ_0 in the Sobolev space $H^1(\mathbb{R})$, the solution $\psi(x, t)$ exists for all time as a continuous function of t with value in $H^1(\mathbb{R})$. In addition, if initially $x\psi_0 \in L^2(\mathbb{R})$, then $x\psi(x, t)$ remains in $L^2(\mathbb{R})$ for all time [9].

Assumption (SL): : Further assumption is made in terms of

$$(1.13) \quad U(\varphi) = -\frac{\omega}{2}\varphi^2 - \frac{1}{2} \int_0^{\varphi^2} F(s) ds.$$

We assume that, for all ω in an interval centered at some ω_0 , the mapping $\varphi \rightarrow U(\varphi)$ has a positive root and the smallest positive root φ_0 is simple, with $U'(\varphi_0) \neq 0$.

Under this assumption there exists a unique, even solution $\varphi(x, \omega)$ of $\varphi_{xx} = -U_\varphi$, decreasing like $A(\omega)e^{-\sqrt{\omega}x}$ as $x \rightarrow +\infty$. Equation (1.10) has thus solutions in the form of solitary waves $e^{j\omega t}\phi$, $\phi = \begin{pmatrix} \varphi(x, \omega) \\ 0 \end{pmatrix}$.

The linearized operator near the solitary wave $e^{j\omega t}\phi$ is

$$(1.14) \quad Bu = -\partial_{xx}u + \omega u + F(|\phi|^2)u + 2F'(|\phi|^2)(u, \phi)\phi,$$

where (\cdot, \cdot) denotes the usual scalar product in \mathbb{C}^2 defined by $(u, v) = u_1\bar{v}_1 + u_2\bar{v}_2$.

Let $C = j^{-1}B = j^{-1}(-\partial_{xx} + \omega) + V$. In general, the spectrum of C is located on the real and imaginary axis. It is composed of the continuous spectrum located on the two half axis $(-i\infty, -i\omega] \cup [i\omega, i\infty)$, and a finite number of discrete eigenvalues. The corresponding invariant spaces are of finite dimension. The point 0 belongs to the discrete spectrum and the dimension of its invariant subspace is at least 2. Recall that we restrict the operator C to even solutions. We now assume more specific conditions:

Assumption (SP): : There is no real eigenvalue except $\lambda = 0$, and the invariant subspace associated to the eigenvalue $\lambda = 0$ is of dimension exactly 2. In addition, there are 2 simple eigenvalues $\pm i\mu$, which satisfy the property $2\mu > \omega$. Their corresponding eigenspaces are of dimension 1. We assume the generic condition that the edges of the continuous spectrum $\pm i\omega$ are not resonances, or equivalently, that there are no solutions, bounded at infinity (virtual levels), nor bound states of $Cu = \pm i\omega u$. We also assume that there are no embedded eigenvalues in the continuous spectrum.

If assumptions (SL) and (SP) are true for a fixed value ω_0 , they are also true for values of ω in a small interval centered at ω_0 . A detailed analysis of the spectral theory of the operator C was developed in [1]. A key point in the analysis of the non self-adjoint operator C is that the coefficients of the matrix-potential V

decrease exponentially fast at infinity. The hypothesis on the spectrum of $C(\omega)$ ensures orbital stability of solitary waves.

We consider initial conditions ψ_0 in the form:

$$(1.15) \quad \psi_0(x) = \phi(x, \omega_0) + (z_0 u(x, \omega_0) + \bar{z}_0 u^*(x, \omega_0)) + f_0(x)$$

where $u(x, \omega_0)$ and $u^*(x, \omega_0)$ are the eigenvectors of $C(\omega_0)$ associated to the eigenvalues $\pm i\mu(\omega_0)$, and f_0 belongs to the eigenspace associated to the continuous spectrum of $C(\omega_0)$. We also assume a non-degeneracy condition. Let $\langle \cdot, \cdot \rangle$ denotes the scalar product in L^2 of \mathbb{C}^2 -valued functions: $\langle u, v \rangle = \int_{\mathbb{R}} (u, v) dx$, and $E_2[f, f]$ be the quadratic terms coming from the Taylor expansion of the nonlinearity:

$$(1.16) \quad E_2[f, f] = F'(|\phi|^2)|f|^2\phi + 2F''(|\phi|^2)|\phi\rangle(\phi, f)^2\phi + 2F'(|\phi|^2)(\phi, f)f.$$

The condition has the form

$$(1.17) \quad \langle E_2[u, u], u(2i\mu_0) \rangle \neq 0,$$

where $u(2i\mu_0)$ is the eigenfunction associated to $\lambda = 2i\mu_0 = 2i\mu(\omega_0)$ of the continuous spectrum. This condition expresses that the interaction of the term of double frequency $2\mu_0$ generated by the nonlinearity with the continuous spectrum is non trivial. It is sometimes referred to as a nonlinear version of the Fermi Golden rule.

Let $|z_0| = \epsilon^{1/2}$ and $N \equiv \|f_0\|_{H^1} + \|(1+x^2)f_0\|_2 \leq c\epsilon^{3/2}$ where c is a constant. For ϵ sufficiently small, we construct a solution in the form

$$(1.18) \quad \psi(x, t) = e^{j(\int_0^t \omega(s) ds + \gamma(t))} \left(\phi(x, \omega(t)) + w(x, t) + f(x, t) \right),$$

where $w(x, t) = z(t)u(x, \omega(t)) + \bar{z}(t)u^*(x, \omega(t))$, and $f(x, t)$ belongs to the subspace associated to the continuous spectrum of $C(\omega(t))$. The dependency on t of ω and γ is defined by the structure of the solution. We show that, as $t \rightarrow +\infty$, $\omega(t) \rightarrow \omega_+$, and

$$(1.19) \quad \psi(x, t) = e^{j\Phi_+(t)} [\phi(x, \omega_+) + z_+(t)u(x, \omega_+) + \bar{z}_+(t)u^*(x, \omega_+)] + e^{j^{-1}Lt}h_+ + o(1)$$

in L^2 , where $o(1)$ is taken with respect to the variable t ,

$$(1.20) \quad \Phi_+(t) = \omega_+t + c_+ \log(1 + k_+\epsilon t) + \gamma_+,$$

ω_+, c_+, k_+ and γ_+ are constants, $k_+ > 0$, $L = -\frac{\partial^2}{\partial x^2}$,

$$(1.21) \quad z_+(t) = \epsilon^{1/2} \frac{\zeta_+ e^{i\mu_+t}}{(1 + k_+\epsilon t)^{1/2 - i\delta}},$$

$\mu_+ = \mu(\omega_+)$, δ a real constant, $\zeta_+ = O(1)$ as $\epsilon \rightarrow 0$, and $h_+ \in L^2$ is independent of t . Asymptotically $f(x, t)$ reduces to a purely dispersive wave. We have, as $t \rightarrow \infty$,

$$(1.22) \quad \|f\|_2 = \|e^{j^{-1}Lt}h_+\|_2 + o(1).$$

If x is bounded (i.e when f is estimated in a norm with a decreasing weight) then $f = O\left(\left(\frac{\epsilon}{1+\epsilon t}\right)^{3/2}\right)$, while for large x (f is then estimated in L^∞ -norm), $f = O\left(\left(\frac{\epsilon}{1+\epsilon t}\right)^{1/2}\right)$. Furthermore, we have, for the conservation of mass

$$(1.23) \quad \|\psi(\cdot, t)\|_2^2 = \|\varphi_{\omega_+}\|_2^2 + \|h_+\|_2^2,$$

and an analogous formula for the conservation of the energy. The result shows that in the neighborhood of a stable soliton state, the system is equivalent to the

direct sum of two systems: the first one is a Hamiltonian system with one degree of freedom and the second is the free Schrödinger equation.

One of the difficulties is that the operator $C(\omega)$ depends on ω and thus is a operator depending slowly on time. To overcome this problem, we consider the problem in a finite interval of time $[0, T]$, and replace the operator $C(\omega(t))$ by $C(\omega(T))$, including the correction in the remainder. We then establish estimates independent of T , which allows us to consider limiting values as $t \rightarrow \infty$. Another difficulty is the fact that the projection P_c on the continuous spectrum does not commute with j . The object of Proposition 2.2 is to extract the leading terms that commute and estimate the correction.

The case where the discrete spectrum is reduced to $\lambda = 0$ was studied in [1]. In a subsequent paper [2], the authors proved the asymptotic stability of solitary waves when the operator $C(\omega_0)$ satisfies the spectral properties (SP). Following their ideas, we developed in [3] a more transparent calculation of the splitting of motions, and give a detailed description and explicit formulas for the correction terms. In particular, we calculated the period of oscillations of the phase as well as of the amplitude of the solution. An heuristic analysis of the phenomenon of amplitude oscillations was developed in [14].

Cuccagna [6] extended the analysis of [1] to the case of spatial dimension larger or equal to 3. The method of decomposition of motion has been also used to investigate the blow-up properties of the NLS equation with critical power nonlinearity in one space dimension [13] (see [18] for a review of the properties of blowing-up solutions).

In this paper, we present the main ideas of the analysis, putting emphasis on the explicit calculation of the leading terms and their correction, and leaving aside the precise estimates of the remainders that can be found in [3].

In Section 2, we recall basic facts about the decomposition of motions and the linearized operator, and we derive a system of equations for the various components of the solution in the form

$$(1.24) \quad \dot{\omega} = \bar{\Omega}(\omega, z, f), \quad \dot{\gamma} = \bar{\Gamma}(\omega, z, f),$$

$$(1.25) \quad \dot{z} = i\mu z + \bar{Z}(\omega, z, f), \quad \dot{f} = C(\omega)f + \bar{F}(\omega, z, f).$$

It consists in 3 scalar ordinary differential equations for $\omega(t)$, $\gamma(t)$ and $z(t)$, and a vector partial differential equation for $f = (f_1(x, t), f_2(x, t))$. We then separate the leading terms and the remainders in these equations. In Section 3, we transform the evolution equations to a simpler, canonical form using ideas of normal coordinates, with the purpose of keeping unchanged the estimates for the remainders. In Section 4, we introduce the notion of majorants defined in terms of norms of $\omega(t)$, $\gamma(t)$, $z(t)$, and $f(x, t)$, with appropriate time dependent weights in a fixed interval of time $[0, T]$ and state (without proof) uniform bounds independent of T , for initial conditions sufficiently close to a solitary wave. In Section 5, we write the precise long time behavior of the various components of the solution.

We conclude this introduction by discussing some natural extensions to this work. One could drop the restriction to even solutions and consider general solutions. This would lead to 2 additional equations to (1.24)-(1.25) for the center and the velocity of the solitary wave. One could replace the hypothesis $2\mu > \omega$ of Assumption (SP) by $n\mu > \omega$. This would imply that the resonance occurs at higher order terms and thus modify the rate of decay. One can also allow more

than one pair of eigenvalues $\pm i\mu$. A more difficult problem is to allow the presence of resonances at the edge of the continuous spectrum.

Finally, let us notice that in general, it is not easy in practice to check the spectral properties of a given operator. A numerical investigation of this problem was done by Grikurov [10] who studied the spectral properties of the linearized operator near a solitary wave associated to the NLS equation with a nonlinear damping term:

$$(1.26) \quad iu_t + u_{xx} + |u|^{2p}u - \alpha|u|^{2q}u = 0,$$

with $p = 3$, $q = 6$, for various small values of the coefficient α . He observed that the point spectrum is composed, in addition to the eigenvalue $\lambda = 0$, of two opposite real eigenvalues when α is smaller than a specific value α_* , while for $\alpha > \alpha_*$, it is composed of two complex conjugate imaginary eigenvalues.

Notations All integrals are taken over \mathbb{R} unless indicated otherwise. Norms in $L^p(\mathbb{R})$ spaces are denoted $\|\cdot\|_p$ and $\|f\|_\rho = \|\rho f\|_2$ denotes the weighted norm in $L^2(\rho)$ with the decreasing weight $\rho(x) = (1 + x^2)^{-\alpha}$, where $\alpha > 0$ will be fixed later.

2. Decomposition of motion

2.1. Linearization near the soliton. The linearized operator near the solitary wave $e^{j\omega t}\phi$, $\phi = \begin{pmatrix} \varphi \\ 0 \end{pmatrix}$, where φ is the positive solution, decreasing like $A(\omega)e^{-\sqrt{\omega}|x|}$ at infinity, of

$$(2.1) \quad \left(-\frac{d^2}{dx^2} + \omega\right)\varphi + F(\varphi^2)\varphi = 0,$$

is

$$(2.2) \quad Bu = \left(-\frac{\partial^2}{\partial x^2} + \omega + F(|\phi|^2)\right)u + 2F'(|\phi|^2)(u, \phi)\phi.$$

Equivalently,

$$Bu = \begin{pmatrix} \mathcal{D}_1 u_1 \\ \mathcal{D}_2 u_2 \end{pmatrix},$$

where $\mathcal{D}_1 = -\frac{\partial^2}{\partial x^2} + \omega + F(|\phi|^2) + 2F'(|\phi|^2)|\phi|^2$, and $\mathcal{D}_2 = -\frac{\partial^2}{\partial x^2} + \omega + F(|\phi|^2)$. It is useful to define the operator

$$(2.3) \quad C = j^{-1}B = j^{-1}\left(-\frac{\partial^2}{\partial x^2} + \omega\right) + V = \begin{pmatrix} \mathcal{D}_2 u_2 \\ -\mathcal{D}_1 u_1 \end{pmatrix}.$$

The spectrum of C has the structure described in Assumption (SP). We denote by X_0 the invariant space associated to $\lambda = 0$, and X_1 and X_c , the eigenspaces associated to $\lambda = \pm i\mu$, and the continuous spectrum respectively. Let $X_d = X_0 + X_1$. Note that, if C^* is the adjoint operator of C , we have $C^*j = -jC$, while $-C\sigma_3 = \sigma_3C$, where $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Define $\chi_0 = j\phi$. We have $C\chi_0 = 0$. In addition, $\chi_1 = \frac{\partial\phi}{\partial\omega} = \phi_\omega$ satisfies $C\chi_1 = \chi_0$. The invariant space X_0 associated to $\lambda = 0$ is spanned by χ_0 and χ_1 . The spectral projection P_0 of a vector valued function f on X_0 is defined by

$$(2.4) \quad P_0 f = \frac{1}{\langle\phi, \phi_\omega\rangle} \left(\langle f, j\phi_\omega\rangle j\phi + \langle f, \phi\rangle \phi\right).$$

Let $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ be the eigenvector of C associated to $i\mu$. We have

$$(2.5) \quad i\mu u_1 = \mathcal{D}_2 u_2, \quad \text{and} \quad i\mu u_2 = -\mathcal{D}_1 u_1.$$

This implies that $\mathcal{D}_2 \mathcal{D}_1 u_1 = \mu^2 u_1$. Since $\mathcal{D}_2 \mathcal{D}_1$ is a real operator, it is possible to choose the function $u_1(x)$ real. From (2.5), we see that u_2 is then purely imaginary. This will be our choice throughout this paper. We denote by $u^* = \begin{pmatrix} u_1 \\ -u_2 \end{pmatrix}$, the eigenvector associated to $-i\mu$. The spectral projection P_1 of an arbitrary vector-valued function f on X_1 is

$$(2.6) \quad P_1 f = \frac{\langle f, ju \rangle}{\langle u, ju \rangle} u + \frac{\langle f, ju^* \rangle}{\langle u^*, ju^* \rangle} u^*.$$

Finally, the spectral projection P_c on X_c is $P_c = I - P_d = I - P_0 - P_1$. It is easy to see that the projection operators satisfy the property

$$(2.7) \quad jP_0 = P_0^* j, \quad jP_1 = P_1^* j, \quad jP_c = P_c^* j.$$

Denote by $u(i\lambda)$ and $u^*(i\lambda)$ the solutions of

$$(2.8) \quad Cu = i\lambda u, \quad Cu^* = -i\lambda u^*,$$

where $\lambda > \omega$. For $f \in X_c$, we have the spectral representation

$$(2.9) \quad f = \int_{\omega}^{\infty} d\lambda \left(\theta_+(\lambda) \langle f, ju(i\lambda) \rangle u(i\lambda) + \theta_-(\lambda) \langle f, ju^*(i\lambda) \rangle u^*(i\lambda) \right).$$

The measures $\theta_{\pm}(\lambda)$ are calculated as follows. By orthogonality

$$(2.10) \quad \begin{aligned} \langle u(i\lambda), ju(i\lambda') \rangle &= \delta(\lambda - \lambda'), \\ \langle u^*(i\lambda), ju^*(i\lambda') \rangle &= \delta(\lambda - \lambda'). \end{aligned}$$

In the limit $|x| \rightarrow \infty$, up to terms exponentially decreasing at infinity, we have, for $\lambda > \omega$,

$$(2.11) \quad u_1(i\lambda) \sim N(\lambda) \cos(\sqrt{\lambda - \omega}|x| - \vartheta(\lambda))$$

where $N(\lambda)$ is a real normalization constant, and $\vartheta(\lambda)$ a phase factor. For large x , $u_2(i\lambda) \sim iu_1(i\lambda)$. The smallest exponential rate $|\beta|$ of the decaying remainders $\exp(-|\beta||x|)$ is equal to $(2\omega)^{1/2}$. The normalization factors in (2.9) can be calculated if we take as f the functions $u(i\lambda)$ and $u^*(i\lambda)$ and compute the singularities (in the sense of distributions) of the corresponding diverging integrals, with the help of the asymptotic formula (2.11). As a result, we get

$$(2.12) \quad \theta_-(\lambda) = \theta_+(\lambda) = \frac{1}{i} \theta(\lambda), \quad \text{with} \quad \theta(\lambda) = \frac{1}{2\pi N^2(\lambda) \sqrt{\lambda - \omega}}.$$

The spectral decomposition (2.9) takes the form

$$(2.13) \quad P_c f = \frac{1}{i} \int_{\omega}^{\infty} \theta(\lambda) d\lambda \left(\langle f, ju(i\lambda) \rangle u(i\lambda) + \langle f, ju^*(i\lambda) \rangle u^*(i\lambda) \right).$$

2.2. The dynamical equations. We look for a solution in the form

$$(2.14) \quad \psi(x, t) = e^{j(\int_0^t \omega(s) ds + \gamma(t))} \Psi(x, t),$$

with

$$(2.15) \quad \Psi = \phi + \chi.$$

In (2.15), $\phi = (\varphi, 0)$, φ solution of (2.1) and $\chi = w(x, t) + f(x, t)$ where $w = z(t)u + \bar{z}(t)u^* \in X_1$ and $f \in X_c$. Notice that u and u^* depend on ω and thus on t . Substituting (2.15) into (1.1), one gets

$$(2.16) \quad -\dot{\gamma}\Psi + j\dot{\Psi} = B\chi + E[\chi],$$

where B is the linearized operator defined in (2.2) and $E[\chi]$ contains all the remainder terms which are at least quadratic in χ as $\chi \rightarrow 0$. Defining $Q[\chi] = j^{-1}E[\chi]$, (2.16) is rewritten

$$(2.17) \quad \dot{\gamma}j\Psi + \dot{\Psi} = C\chi + Q[\chi].$$

Applying successively the spectral projections P_0, P_1 and P_c to (2.17) and using orthogonality relations, we get a system of coupled equations for $\omega(t), \gamma(t), z(t)$ and $f(x, t)$ in the form

PROPOSITION 2.1. *The functions $\omega(t), \gamma(t), z(t)$ and $f(x, t)$ satisfy the system*

$$(2.18) \quad \dot{\omega} = \frac{\langle P_0 Q, \Psi \rangle}{\langle (\phi_\omega - P_{0\omega} \chi), \Psi \rangle}$$

$$(2.19) \quad \dot{\gamma} = \frac{\langle jP_0(\phi_\omega - P_{0\omega} \chi), P_0 Q \rangle}{\langle (\phi_\omega - P_{0\omega} \chi), \Psi \rangle}$$

$$(2.20) \quad \langle u, ju \rangle (\dot{z} - i\mu z) = \langle Q, ju \rangle - \langle w_\omega - P_{1\omega} f, ju \rangle \dot{\omega} - \langle \chi, u \rangle \dot{\gamma}$$

$$(2.21) \quad \dot{f} = Cf + P_c Q[\chi] + \dot{\omega} P_{c\omega} \chi - \dot{\gamma} P_c(j\chi).$$

2.3. Effective equations. We separate in (2.18)–(2.20) the leading terms and the remainders. The nonlinear term $F(|\psi|^2)\psi$ near the solitary wave ϕ is expanded in the form

$$(2.22) \quad F(|\psi|^2)\psi = F(|\phi|^2)\phi + F(|\phi|^2)\chi + 2F'(|\phi|^2)(\chi, \phi)\phi + E[\chi],$$

Recalling that r is the order of the zero of $F(s)$ at $s = 0$, we expand $E[\chi]$ as

$$(2.23) \quad E[\chi] = E_2 + \dots + E_{2r} + E_R$$

where E_j is of order j in χ . The quadratic terms have the form (1.16). It is also useful to define $E_2[\chi_1, \chi_2]$ as a symmetric quadratic form

$$(2.24) \quad \begin{aligned} E_2[\chi_1, \chi_2] = & \frac{1}{2} F'(|\phi|^2) \left((\chi_1, \chi_2) + (\chi_2, \chi_1) \right) \phi + 2F''(|\phi|^2) (\phi, \chi_1) (\phi, \chi_2) \phi \\ & + F'(|\phi|^2) \left((\phi, \chi_2) \chi_1 + (\phi, \chi_1) \chi_2 \right), \end{aligned}$$

and to notice that it satisfies

$$(2.25) \quad \langle E_2[X, Y], Z \rangle = \langle X, E_2[Y^*, Z] \rangle,$$

where X, Y, Z are complex valued vector functions and $X^* = (\bar{X}_1, \bar{X}_2)$.

After expansion of numerators and denominators in (2.18) and (2.19), one has

$$(2.26) \quad \begin{aligned} \dot{\gamma} = & -\frac{\langle E_2[w, w] + 2E_2[w, f], \phi_\omega \rangle}{\langle \phi, \phi_\omega \rangle} \\ & - \langle \phi, \phi_\omega \rangle^{-2} \left[(\langle E_3[w, w, w], \phi_\omega \rangle - \langle E_2[w, w], P_{0\omega} w \rangle) \langle \phi, \phi_\omega \rangle \right. \\ & \left. + \langle E_2[w, w], \phi_\omega \rangle (\langle P_{0\omega} w, \phi \rangle - \langle w, \phi_\omega \rangle) \right] + \Gamma_R, \end{aligned}$$

and

$$(2.27) \quad \begin{aligned} \dot{\omega} = & \frac{\langle E_2[w, w] + 2E_2[w, f], j\phi \rangle}{\langle \phi, \phi_\omega \rangle} \\ & + \langle \phi, \phi_\omega \rangle^{-2} \left[(\langle E_3[w, w, w], j\phi \rangle + \langle E_2[w, w], P_0 j\chi \rangle) \langle \phi, \phi_\omega \rangle \right. \\ & \left. + \langle E_2[w, w], j\phi \rangle (\langle P_{0\omega} w, \phi \rangle - \langle \phi_\omega, w \rangle) \right] + \Omega_R. \end{aligned}$$

Notice that

$$(2.28) \quad \langle E_2[w, w], j\phi \rangle = z^2 \langle E_2[u, u], j\phi \rangle + \bar{z}^2 \langle E_2[u^*, u^*], j\phi \rangle + 2z\bar{z} \langle E_2[u, u^*], j\phi \rangle.$$

Using the definition of E_2 , and that $u = (u_1, u_2)$ with u_1 real and u_2 pure imaginary, we have that $\langle E_2[u, u^*], j\phi \rangle = 0$, and

$$(2.29) \quad \begin{aligned} \langle E_2[w, w], j\phi \rangle &= (z^2 - \bar{z}^2) \langle E_2[u, u], j\phi \rangle \\ &= 2(z^2 - \bar{z}^2) \frac{1}{\langle \phi, \phi_\omega \rangle} \int F'(|\phi|^2)(u, \phi)(u, j\phi) dx \end{aligned}$$

is purely imaginary.

Finally, we rewrite (2.20) in the form:

$$(2.30) \quad \begin{aligned} \dot{z} - i\mu z = & -\frac{\langle E_2[w, w] + 2E_2[w, f] + E_3[w, w, w], u \rangle}{\langle u, ju \rangle} \\ & - \frac{\langle w_\omega, ju \rangle \langle E_2[w, w], j\phi \rangle - \langle w, u \rangle \langle E_2[w, w], \phi_\omega \rangle}{\langle \phi, \phi_\omega \rangle \langle u, ju \rangle} + Z_R. \end{aligned}$$

It is important to notice that

$$(2.31) \quad \langle u, ju \rangle = i\delta, \text{ with } \delta > 0.$$

Indeed,

$$(2.32) \quad \begin{aligned} \langle u, ju \rangle &= \int_{\mathbb{R}} u_1(u_2 - u_2^*) dx = 2 \int_{\mathbb{R}} u_1 u_2 dx \\ &= \frac{2}{i\mu} \int u_2 \mathcal{D}_2 u_2 dx = \frac{2}{i\mu} \int u_2 \left(-\frac{d^2}{dx^2} + \omega + F(|\phi|^2) \right) u_2 dx. \end{aligned}$$

Since u_2 is purely imaginary and ϕ is the only eigenfunction of the operator $-\frac{d^2}{dx^2} + \omega + F(|\phi|^2)$ corresponding to the minimal spectral point 0, the integral is strictly negative. We know also, from the stability condition of the solitary wave that, $\frac{d}{d\omega} \|\phi\|_2^2 = 2\langle \phi, \phi_\omega \rangle > 0$ for $\omega = \omega_0$ and consequently for ω close to ω_0 .

It can be proved that the remainders Γ_R , Ω_R , and Z_R are of the form

$$(2.33) \quad \mathcal{R}(\omega, |z| + \|f\|_\infty) (|z|^2 + \|f\|_\rho)^2,$$

where $\mathcal{R}(\omega, |z| + \|f\|_\infty)$ is a quantity that remains bounded as long as ω is in the vicinity of ω_0 and $|z| + \|f\|_\infty$ is bounded. Estimates of the same type, with

sometimes different norms involved, can be obtained for all terms that we treat as remainders. We do not provide details here and refer to [3] for the detailed calculations.

We now turn to equation (2.21) for f . One difficulty is that the projection P_c does not commute with j . We rewrite it in the form

$$(2.34) \quad \dot{f} = Cf - P_c j E_2[w, w] + \dot{\gamma} i(P_+ - P_-)f + F_R$$

where P_{\pm} are respectively the projection on the positive and negative parts of the continuous spectrum, and F_R is the remainder

$$(2.35) \quad F_R = -P_c j (E[\chi] - E_2[w, w]) - \dot{\omega} P_{d\omega} \chi + \dot{\gamma} P_c j^{-1} w + \dot{\gamma} (P_c j^{-1} - i(P_+ - P_-)) f.$$

The last term in the expression for F_R is estimated in terms of weighted norms:

PROPOSITION 2.2. [2] For $f \in X_c$,

$$(2.36) \quad \|(1 + x^2) (P_c j^{-1} f - i(P_+ - P_-) f)\|_2 \leq K(\omega) \|f\|_{\rho},$$

where $\rho(x) = (1 + x^2)^{-\alpha}$, $\alpha > 0$ arbitrary, and K is a constant depending on ω .

3. Transformations of the equations

Our goal is to transform the evolution equations for γ , ω , z and f to a more simple, canonical form. It is based on the idea of normal coordinates where one extracts from the solution terms that can be explicitly calculated, trying at the same time, to keep unchanged the estimates for the remainders.

3.1. Equation for ω . Equation (2.27) for ω has the form

$$(3.1) \quad \begin{aligned} \dot{\omega} = & \Omega_{20}(\omega)z^2 + \Omega_{11}(\omega)z\bar{z} + \Omega_{02}(\omega)\bar{z}^2 + \Omega_{30}(\omega)z^3 + \Omega_{21}(\omega)z^2\bar{z} + \Omega_{12}(\omega)z\bar{z}^2 \\ & + \Omega_{03}\bar{z}^3 + z\langle f, \Omega'_{10} \rangle + \bar{z}\langle f, \Omega'_{01} \rangle + \Omega_R. \end{aligned}$$

Notice that $\Omega_{ij} = \bar{\Omega}_{ji}$. Also,

$$(3.2) \quad \Omega_{20} = \bar{\Omega}_{02} = \frac{\langle E_2[u, u], j\phi \rangle}{\langle \phi, \phi_\omega \rangle} = \frac{2}{\langle \phi, \phi_\omega \rangle} \int dx F'(|\phi|^2)(\phi, u)(u, j\phi)$$

is purely imaginary and

$$(3.3) \quad \Omega_{11} = -\frac{\langle E_2[u, u^*], j\phi \rangle}{\langle \phi, \phi_\omega \rangle} = 0.$$

Using property (2.25), we find that the coefficients Ω'_{10} and Ω'_{01} are given by

$$(3.4) \quad \Omega'_{10} = \bar{\Omega}'_{01} = 2 \frac{E_2[u^*, j\phi]}{\langle \phi, \phi_\omega \rangle}.$$

Following the classical method of normal coordinates, one proves that

PROPOSITION 3.1. *There exist coefficients $b_{ij}(\omega)$, $0 \leq i, j \leq 3$, and vector-functions $b'_{ij}(x, \omega)$, such that the new function ω_1 defined as*

$$(3.5) \quad \omega_1 = \omega + b_{20}(\omega)z^2 + b_{02}(\omega)\bar{z}^2 + b_{30}(\omega)z^3 + b_{21}(\omega)z^2\bar{z} + b_{12}z\bar{z}^2 + b_{03}\bar{z}^3 + z\langle f, b'_{10} \rangle + \bar{z}\langle f, b'_{01} \rangle$$

obeys the differential equation

$$(3.6) \quad \dot{\omega}_1 = \widehat{\Omega}_R$$

and $\widehat{\Omega}_R$ will satisfy the same estimate (2.33) as Ω_R .

If we apply the same calculation to equation (2.30) for z , a change of variables similar to (3.5) leads to a system of equations for coefficients c_{ij} associated to z . The equation for c'_{01} has the form

$$(3.7) \quad (C^* + 2i\mu)c'_{01} = -iZ'_{10}.$$

Note that $-2i\mu$ is a point in the continuous spectrum of C^* , thus the function c'_{01} does not vanish at infinity. This implies that in the expression for the new variable z_1 , the term $\bar{z}\langle f, c'_{01} \rangle$ can have a complicated structure. The function f itself does not decrease well enough at infinity. To proceed with the equation for z , we have to analyze carefully the behavior of f .

3.2. Transformation of the equation for f . A technical difficulty in the treatment of the evolution equation for f is that the operator $C = C(\omega(t))$ depends on time. This dependency is, naturally, quite slow and we can overcome this problem by fixing an interval of time $[0, T]$, and approximate $C(\omega)$ by its value at time $t = T$. This will allow us to estimate the value of the important quantity $\omega(T)$, and as a result, to get a final closed system of the estimated quantities on the whole time axis for all the components of the solution.

We decompose f into its projection on the discrete and continuous spectrum at time T as

$$(3.8) \quad f = g + h \quad g \in X_T^d, \quad h \in X_T^c,$$

where $X_T^d = P_T^d X$ and $X_T^c = P_T X$ are the spectral spaces associated to the discrete and continuous spectrum respectively, at time T , and we have denoted the corresponding projections $P_T = P_c(\omega(T))$ and $P_T^d = I - P_T$. The operator $C(\omega)$ in the equation for f will now be replaced by $C_T = C(\omega_T)$ with $\omega_T = \omega(T)$, leading to additional terms in the remainders. Applying P_T to (2.34), we get

$$(3.9) \quad \dot{h} = C_T h + \sigma(t)P_T j^{-1}h - P_T j E_2[w, w] + H'_R$$

with $\sigma(t) = \omega - \omega_T + \dot{\gamma}$. In the light of Proposition 2.2, we rewrite it as

$$(3.10) \quad \dot{h} = C_M h - P_T j E_2[w, w] + H_R$$

where

$$(3.11) \quad C_M = C_M(t) = C_T + i\sigma(t)(P_T^+ - P_T^-),$$

or more explicitly, as

$$(3.12) \quad \dot{h} = C_M h + H_{20} z^2 + H_{11} z \bar{z} + H_{02} \bar{z}^2 + H_R.$$

Here, the coefficients H_{ij} are defined by

$$(3.13) \quad H_{20} = -P_T j E_2[u, u], \quad H_{11} = -2P_T j E_2[u, u^*], \quad H_{02} = -P_T j E_2[u^*, u^*].$$

We now introduce a new function h_1 defined by

$$(3.14) \quad h = h_1 + k + k_1$$

where

$$(3.15) \quad k = a_{20} z^2 + a_{11} z \bar{z} + a_{02} \bar{z}^2, \quad k_0 = k|_{t=0}, \quad \text{and} \quad k_1 = -e^{(\int_0^t C_M(\tau) d\tau)} k_0,$$

with some $a_{ij} \equiv a_{ij}(\omega, x)$ satisfying $a_{ij} = \bar{a}_{ji}$. The purpose here is to extract from h the contribution which is quadratic in z . Note that $h_{10} = h_1(t=0) = h_0$.

We look for coefficients a_{ij} such that the resulting equation for h_1 has the form

$$(3.16) \quad \dot{h}_1 = C_M(t)h_1 + \widehat{H}_R.$$

It is also convenient to replace in the equations for a_{ij} , the coefficient $\mu(\omega)$ by $\mu_T = \mu(\omega_T)$ and include the correction in the remainder. This will avoid later the differentiation of $(C_T - 2i\mu)^{-1}$ with respect to t . After substitution of (3.15) into (3.12), we get

$$\begin{aligned}
 H_{20} - 2i\mu_T a_{20} &= -C_T a_{20} \\
 H_{11} &= -C_T a_{11} \\
 H_{02} + 2i\mu_T a_{02} &= -C_T a_{02}.
 \end{aligned}
 \tag{3.17}$$

The dependency in x appears here through the coefficients $a_{ij} = a_{ij}(\omega, x)$. Notice that the coefficients $H_{ij} \in X_T^c$ are smooth, exponentially vanishing functions at infinity. The coefficient a_{11} is obtained as

$$a_{11} = -C_T^{-1} H_{11}.
 \tag{3.18}$$

As a function of x , it is smooth and exponentially decreasing at infinity. Furthermore, we have to invert $C_T \pm 2i\mu_T$, with both points $\pm 2i\mu_T$ in the continuous spectrum of C_T . In general, the functions

$$a_{20} = -(C_T - 2i\mu_T)^{-1} H_{20}, \quad \text{and} \quad a_{02} = \bar{a}_{20} = -(C_T + 2i\mu_T)^{-1} H_{02}
 \tag{3.19}$$

do not decrease at infinity. They behave like solutions of the homogeneous equation $(C_T - 2i\mu_T)a = 0$, and thus oscillate at infinity. Nevertheless, there exists a special choice for these inverse operators that leads to preferable properties of h_1 . This choice is

$$a_{20} = -(C_T - 2i\mu_T - 0)^{-1} H_{20}, \quad a_{02} = \bar{a}_{20} = -(C_T + 2i\mu_T - 0)^{-1} H_{02}.
 \tag{3.20}$$

Such property is reflected in Lemma 4.2 that claims that for $t \geq 0$,

$$\|e^{C_T t} (C_T \pm 2i\mu - 0)^{-1} P_T^\pm h\|_\rho \leq c(1+t)^{-3/2} (\|h\|_2 + \|(1+x^2)^{3/2} h\|_1).
 \tag{3.21}$$

It corresponds to the classical fact that $(i\lambda - 0)^{-1} e^{i\lambda t} \rightarrow 0$ as $t \rightarrow +\infty$, in the sense of distributions. The weight $\rho(x) = (1+x^2)^{-\alpha}$ here must satisfy the condition $\alpha \geq 2$.

3.3. Transformation of the equation for z . When substituting in (2.30) the contribution of f in terms of z , one gets

$$\begin{aligned}
 \dot{z} &= i \left(\mu z + Z_{20} z^2 + Z_{11} z \bar{z} + Z_{02} \bar{z}^2 + Z_{30} z^3 + Z_{21} z^2 \bar{z} + Z_{12} z \bar{z}^2 + Z_{03} \bar{z}^3 \right) \\
 &+ i \left(Z'_{30} z^3 + Z'_{21} z^2 \bar{z} + Z'_{12} z \bar{z}^2 + Z'_{03} \bar{z}^3 \right) + \widetilde{Z}_R
 \end{aligned}
 \tag{3.22}$$

where the coefficients Z_{ij} can be calculated explicitly and are real. We are especially interested in the coefficient Z'_{21} , which is given by the formula

$$iZ'_{21} = -\langle C_T^{-1} 2P_T j E_2[u, u^*], 2 \frac{E_2[u^*, u]}{\langle u, ju \rangle} \rangle - \langle (C_T - 2i\mu_T - 0)^{-1} P_T j E_2[u, u], 2 \frac{E_2[u, u]}{\langle u, ju \rangle} \rangle.
 \tag{3.23}$$

Applying the method of normal coordinates like in Section 3.1, we have

PROPOSITION 3.2. *There exist coefficients c_{ij} such that the new function z_1 defined by*

$$z_1 = z + c_{20} z^2 + c_{11} z \bar{z} + c_{02} \bar{z}^2 + c_{30} z^3 + c_{12} z \bar{z}^2 + c_{03} \bar{z}^3,
 \tag{3.24}$$

satisfies an equation of the form

$$\dot{z}_1 = i\mu(\omega) z_1 + iK(\omega) |z_1|^2 z_1 + \widehat{Z}_R
 \tag{3.25}$$

where \widehat{Z}_R satisfies estimates of the same type as \widetilde{Z}_R , and

$$(3.26) \quad \operatorname{Re} iK = \operatorname{Re} iZ'_{21}.$$

Notice that, by this change of variables, one can eliminate all quadratic and cubic terms in z except the term in the form of $|z|^2 z$.

PROPOSITION 3.3. *Suppose that the non-degeneracy condition $|\langle E_2[u, u], u(2i\mu) \rangle|^2 \neq 0$ is satisfied, then*

$$(3.27) \quad \operatorname{Re} iZ'_{21} < 0.$$

This proposition is a key point in the analysis. It ensures that, when integrating the evolution equation for z_1 , the function z_1 and thus z , remains bounded. We present below the main elements of the proof.

PROOF. We first notice that the coefficient $\langle C_T^{-1} P_T j E_2[u, u^*], E_2[u, u^*] \rangle$ that appears in the expression (3.23) for iZ'_{21} is real. Indeed, $E_2[u, u^*]$ is real, and all the operators in the above scalar product are also real. We know from (2.31) that $\langle u, ju \rangle = i\delta$, with $\delta > 0$. It follows that

$$(3.28) \quad \operatorname{Re} iZ'_{21} = -\frac{2}{\delta} \operatorname{Im} \langle (C_T - 2i\mu_T - 0)^{-1} P_T j E_2[u, u], E_2[u, u] \rangle.$$

Using that P_T commutes with C_T^{-1} , we have $C_T^{-1} P_T = P_T C_T^{-1} P_T$. We have also that $P_T^* = j^{-1} P_T j$. Denoting $\alpha = P_T j E_2[u, u]$, we thus have

$$(3.29) \quad \operatorname{Re} iZ'_{21} = -\frac{2}{\delta} \operatorname{Im} \langle (C_T - 2i\mu_T - 0)^{-1} \alpha, j\alpha \rangle.$$

From the spectral representation (2.9), we have, since $\alpha \in X_T$,

$$(3.30) \quad \begin{aligned} & \langle (C_T - 2i\mu - 0)^{-1} \alpha, j\alpha \rangle \\ &= \frac{1}{i} \int_{\omega}^{\infty} \theta(\lambda) d\lambda \left(\frac{\langle u(i\lambda), j\alpha \rangle \overline{\langle u(i\lambda), j\alpha \rangle}}{i\lambda - 2i\mu - 0} + \frac{\langle u^*(i\lambda), j\alpha \rangle \overline{\langle u^*(i\lambda), j\alpha \rangle}}{-i\lambda - 2i\mu - 0} \right) \\ &= - \int_{\omega}^{\infty} \theta(\lambda) d\lambda \left(\frac{\langle u(i\lambda), j\alpha \rangle \overline{\langle u(i\lambda), j\alpha \rangle}}{\lambda - 2\mu + i0} - \frac{\langle u^*(i\lambda), j\alpha \rangle \overline{\langle u^*(i\lambda), j\alpha \rangle}}{\lambda + 2\mu - i0} \right) \end{aligned}$$

Using that $\frac{1}{\lambda + i0} = \text{p.v.} \frac{1}{\lambda} - i\pi\delta(\lambda)$, with p.v. the Cauchy principal value and δ the Dirac distribution, we have

$$(3.31) \quad \begin{aligned} & \langle (C_T - 2i\mu - 0)^{-1} \alpha, j\alpha \rangle \\ &= - \int_{\omega}^{\infty} \theta(\lambda) d\lambda \left(\frac{\langle u(i\lambda), j\alpha \rangle \overline{\langle u(i\lambda), j\alpha \rangle}}{\lambda - 2\mu} - \frac{\langle u^*(i\lambda), j\alpha \rangle \overline{\langle u^*(i\lambda), j\alpha \rangle}}{\lambda + 2\mu} \right) \\ & \quad + i\pi\theta(2\mu) \langle u(i2\mu), j\alpha \rangle \overline{\langle u(i2\mu), j\alpha \rangle}. \end{aligned}$$

The integral term in (3.31) is real. Thus,

$$(3.32) \quad \operatorname{Im} \langle (C_T - 2i\mu - 0)^{-1} \alpha, j\alpha \rangle = \operatorname{Im}(i\pi\theta(2\mu)) |\langle u(2i\mu), j\alpha \rangle|^2,$$

with

$$(3.33) \quad \theta(2\mu) = \frac{1}{2\pi\sqrt{2\mu - \omega} N^2(2\mu)}.$$

Assuming the non-degeneracy condition $|\langle E_2[u, u], u(2i\mu) \rangle|^2 \neq 0$, we get

$$(3.34) \quad \operatorname{Re} iZ'_{21} = -\frac{2}{\delta} \pi\theta(2\mu) |\langle E_2[u, u], u(2i\mu) \rangle|^2 < 0.$$

□

We now replace the coefficient $K(\omega)$ in (3.25) (which depends on t) by $K_T = K(\omega_T)$, and put the resulting additional term in the remainder. Equation (3.25) becomes

$$(3.35) \quad \dot{z}_1 = i\mu z_1 + iK_T |z_1|^2 z_1 + \widehat{Z}_R.$$

It is easier to deal with $y = |z_1|^2$ rather than z_1 . Indeed, while both functions decrease at infinity, the function z_1 oscillates at infinity. The equation satisfied by y is simply obtained by multiplying (3.25) by \bar{z}_1 and taking the real part:

$$(3.36) \quad \dot{y} = \operatorname{Re}(iK_T)y^2 + Y_R.$$

3.4. Summary of transformed equations and canonical system. We summarize the main formulas of Sections 3.1-3.3. We have obtained a canonical system in the form

$$(3.37) \quad \dot{\omega}_1 = \widehat{\Omega}_R,$$

$$(3.38) \quad \dot{y} = \operatorname{Re}(iK_T)y^2 + Y_R,$$

$$(3.39) \quad \dot{h} = C_M(t)h - P_T j E_2[w, w] + H_R,$$

$$(3.40) \quad \dot{h}_1 = C_M(t)h_1 + \widehat{H}_R.$$

where the operator $C_M = C_T + i\sigma(t)(P_T^+ - P_T^-)$. The operator $C_T = C(\omega_T)$ does not depend on t and the structure of its spectrum is known. The function σ is a smooth, real-valued function of t and $\operatorname{Re}(iK_T) < 0$.

An important part of the analysis that we have omitted in this presentation is the precise estimates of the remainders. Detailed calculations are given in [3].

3.5. Initial conditions. We suppose that at $t = 0$,

$$(3.41) \quad z(0) = z_0, \quad |z_0| = \epsilon^{1/2}$$

$$(3.42) \quad f(x, 0) = f_0(x), \quad N \equiv \|f_0\|_{H^1} + \|(1 + x^2)f_0\|_2 \leq c\epsilon^{3/2}.$$

Since $|z_1|^2 \leq |z|^2 + \mathcal{R}(\omega, z)|z|^3$, we have also, denoting $z_{10} = z_1(0)$,

$$(3.43) \quad y_0 = |z_{10}|^2 \leq \epsilon + \mathcal{R}(\omega, |z_0|)\epsilon^{3/2}.$$

From the formula $h = P_T f = f + (P_d - P_T^d)f$, we have that

$$(3.44) \quad \|h_0\|_{H^1} + \|(1 + x^2)h_0\|_2 \leq c\epsilon^{3/2} + \mathcal{R}(\omega)|\omega_T - \omega| \|f_0\|_\rho.$$

4. Estimates of majorants

4.1. Linear evolution. Defining the weighted norms

$$(4.1) \quad \|h\|_{W_1} = \|(1 + x^2)h\|_1, \quad \text{and} \quad \|h\|_{W_2} = \|(1 + x^2)h\|_2,$$

we have two important lemmas on the evolution operator $e^{C_T t}$ [1] that are used to estimate the functions h and h_1 solutions of (3.39) and (3.40).

LEMMA 4.1.

$$(4.2) \quad \|e^{C_T t} P_T^\pm \alpha\|_2 \leq c(\omega_T) \|\alpha\|_2,$$

$$(4.3) \quad \|e^{C_T t} P_T^\pm \alpha\|_\infty \leq c(\omega_T) \begin{cases} t^{-1/2} (\|\alpha\|_2 + \|\alpha\|_W) \\ (1+t)^{-1/2} (\|\alpha\|_{H^1} + \|\alpha\|_W) \end{cases}$$

where $\|h\|_W$ stands for either $\|h\|_{W_1}$ or $\|h\|_{W_2}$, and the constant $c(\omega_T)$ depends on C_T and thus on ω_T .

We have also

$$(4.4) \quad \|e^{C_T t} P_T^\pm h\|_\rho \leq c(\omega_T) (1+t)^{-3/2} [\|h\|_2 + \|h\|_W],$$

where $\rho(x) = (1+x^2)^{-q}$, $q \geq 2$.

LEMMA 4.2.

$$(4.5) \quad \|e^{C_T t} C_T^{-1} P_T^\pm h\|_\rho \leq c(1+t)^{-3/2} \|(1+x^2)h\|_2,$$

$$(4.6) \quad \|e^{C_T t} (C_T \pm 2i\mu - 0)^{-1} P_T^\pm h\|_\rho \leq c(1+t)^{-3/2} [\|h\|_2 + \|(1+x^2)^{3/2} h\|_1],$$

where $\rho(x) = (1+x^2)^{-q}$, $q \geq 2$.

4.2. Uniform estimates. We define the quantities

$$(4.7) \quad \mathbb{M}_0(T) = \sup_{0 \leq t \leq T} |\omega_T - \omega(t)| \left(\frac{\epsilon}{1+\epsilon t} \right)^{-1},$$

$$(4.8) \quad \mathbb{M}_1(T) = \sup_{0 \leq t \leq T} |z(t)| \left(\frac{\epsilon}{1+\epsilon t} \right)^{-1/2},$$

$$(4.9) \quad \mathbb{M}_2(T) = \sup_{0 \leq t \leq T} \|f(t)\|_\infty \left(\frac{\epsilon}{1+\epsilon t} \right)^{-1/2} \log^{-1}(2+\epsilon t),$$

$$(4.10) \quad \mathbb{M}_3(T) = \sup_{0 \leq t \leq T} \|h_1(t)\|_\rho \left(\frac{\epsilon}{1+\epsilon t} \right)^{-3/2} \log^{-1}(2+\epsilon t),$$

referred to as ‘majorants’, and denote \mathbb{M} the 4-dimensional vector $(\mathbb{M}_0, \dots, \mathbb{M}_3)$. The goal is to prove that if ϵ is sufficiently small, all these quantities are bounded uniformly in T . All the tools are in place now and the analysis proceeds in several steps that we state below.

First, one has to write the remainders in (3.37)-(3.40) in terms of the \mathbb{M}_i . Then, one has to study the canonical system (3.37)-(3.40) with assumptions on the r.h.s. in the form of the remainder estimates. These estimates are a little technical and we refer to Sections 5.1 and 5.2 of [3] for the details. Notice that eq.(3.38) is a Riccati equation. The Duhamel principle combined with Lemmas 4.1 and 4.2 provides estimates on h and h_1 . This leads to upper bounds for the weighted quantities \mathbb{M}_i , that can be combined in the form:

$$(4.11) \quad \mathbb{M}^2 \leq \mathcal{R}(\epsilon^{1/2} \mathbb{M})(1 + \epsilon^{1/2} F(\mathbb{M}))$$

where $F(r)$ is a function with finite power growth, and $\mathcal{R}(\epsilon^{1/2} \mathbb{M})$ is a quantity that remains bounded as long as its argument is bounded. From this inequality, it follows, that, \mathbb{M} is either bounded independently of ϵ or \mathbb{M} belongs to a set separated from the origin $|\mathbb{M}| > k(\epsilon)$, with $k(\epsilon) \rightarrow \infty$ as $\epsilon \rightarrow 0$. However, since at $t = 0$, $\mathbb{M}(0)$ is bounded independently of ϵ , the second possibility cannot happen,

because it would lead to a discontinuity of \mathbb{M} as a function of time. We conclude that, for ϵ sufficiently small,

$$(4.12) \quad \mathbb{M}(T) \leq M$$

with a constant M independent of T .

PROPOSITION 4.3. *The function $\omega(t)$ has a limit ω_+ as $t \rightarrow \infty$. Furthermore, we have the estimates for all $t > 0$,*

$$(4.13) \quad |\omega_+ - \omega(t)| \leq \mathcal{M}_0 \frac{\epsilon}{1 + \epsilon t},$$

$$(4.14) \quad |z(t)| \leq \mathcal{M}_1 \left(\frac{\epsilon}{1 + \epsilon t} \right)^{1/2},$$

$$(4.15) \quad \|f(t)\|_\infty \leq \mathcal{M}_2 \left(\frac{\epsilon}{1 + \epsilon t} \right)^{1/2} \log(2 + \epsilon t),$$

$$(4.16) \quad \|h_1(t)\|_\rho \leq \mathcal{M}_3 \left(\frac{\epsilon}{1 + \epsilon t} \right)^{3/2} \log(2 + \epsilon t).$$

In particular, $|\omega_T - \omega(t)| \leq \frac{\epsilon M}{1 + \epsilon t}$ and thus $|\omega_T - \omega(t)|$ is a decreasing function of t . Applying this result to $|\omega(t_1) - \omega(t_2)|$, we see that $\omega(t)$ is a Cauchy sequence. It thus has a limit, denoted ω_+ . Let \mathcal{M}_i the limiting value of $\mathbb{M}_i(T)$ as $T \rightarrow \infty$.

Notice that in the decomposition (3.8), a fixed time T has been chosen, and all the components depend on $\omega(T)$. From the above proposition, $\omega(t)$ has a limit ω_+ as $t \rightarrow \infty$. We can thus reformulate the decomposition by choosing $T = \infty$ and have the dependency of the various components of f on ω_+ . Let us denote $P_\infty = P_c(\omega_+)$, $P_\infty^d = I - P_\infty$. All the estimates previously obtained under the hypothesis that T is a fixed finite time can be carried out without modification with $T = \infty$ and $\omega_T = \omega_+$.

THEOREM 4.4. *Consider the NLS equation (1.1).*

- (i) *Assume that the nonlinearity satisfies assumptions (NL) and (SL) and that there exist solutions in the form of solitary waves $e^{j\omega_0 t} \phi(\omega_0)$.*
- (ii) *Denoting B the linearized operator near the solitary wave, assume that $C = j^{-1}B$ satisfies the condition (SP) describing the structure of its spectrum.*
- (iii) *Assume the non-degeneracy condition $\langle E_2[u, u], u(2i\mu_0) \rangle \neq 0$.*
- (iv) *Assume an initial condition ψ_0 in the form of a perturbation of the solitary wave $\psi_0 = \phi(x, \omega_0) + (z_0 u(x, \omega_0) + \bar{z}_0 u^*(x, \omega_0)) + f_0(x)$, satisfying $|z_0| = \epsilon^{1/2}$ and $N \equiv \|f_0\|_{H^1} + \|(1 + x^2)f_0\|_2 \leq \epsilon^3$.*

For ϵ small enough, one can write the solution in the form

$$(4.17) \quad \psi(x, t) = e^{j(\int^t \omega(s) ds + \gamma(t))} \left(\phi(x, \omega) + z(t)u(x, \omega) + \bar{z}(t)u^*(x, \omega) + f(x, t) \right)$$

with the following properties. There exists a constant ω_+ such that $\omega_+ = \lim_{t \rightarrow \infty} \omega(t)$. In addition, for all $t > 0$, there exist constants $\mathcal{M}_0, \dots, \mathcal{M}_3$ such that

$$(4.18) \quad |\omega_+ - \omega(t)| \leq \mathcal{M}_0 \frac{\epsilon}{1 + \epsilon t},$$

$$(4.19) \quad |z(t)| \leq \mathcal{M}_1 \left(\frac{\epsilon}{1 + \epsilon t} \right)^{1/2},$$

$$(4.20) \quad \|f(t)\|_\infty \leq \mathcal{M}_2 \left(\frac{\epsilon}{1 + \epsilon t} \right)^{1/2} \log(2 + \epsilon t),$$

Furthermore, we write $f = g + h$, where $g = P_\infty^d f$, $h = P_\infty f = k + k_1 + h_1$, $k = a_{20}z^2 + a_{11}z\bar{z} + a_{02}\bar{z}^2$, $k_0 = k|_{t=0}$, and $k_1 = -\exp\left(\int_0^t C_+(\tau) d\tau\right) k_0$. In the above equations, $a_{ij} = a_{ij}(\omega_+, x)$ are defined as in (3.18) and (3.20), and

$$(4.21) \quad C_+ = C(\omega_+) + i(\omega(t) - \omega_+ + \dot{\gamma})(P_\infty^+ - P_\infty^-).$$

The function h_1 satisfies the estimate

$$(4.22) \quad \|h_1(t)\|_\rho \leq \mathcal{M}_3 \left(\frac{\epsilon}{1 + \epsilon t} \right)^{3/2} \log(2 + \epsilon t).$$

5. Asymptotic behavior for large time

Until now, the function γ did not play an essential role in our computations. However, we are interested in its long-time behavior. We can repeat the calculation performed for ω in Section 3.1. Equation (2.26) is rewritten

$$(5.1) \quad \begin{aligned} \dot{\gamma} = & \Gamma_{20}(\omega)z^2 + \Gamma_{11}(\omega)z\bar{z} + \Gamma_{02}(\omega)\bar{z}^2 + \Gamma_{30}(\omega)z^3 + \Gamma_{21}(\omega)z^2\bar{z} + \Gamma_{12}(\omega)z\bar{z}^2 \\ & + \Gamma_{03}\bar{z}^3 + z\langle f, \Gamma'_{10} \rangle + \bar{z}\langle f, \Gamma'_{01} \rangle + \Gamma_R \end{aligned}$$

and Γ_R satisfies the same estimate as Ω_R . The only difference between the equations for ω and γ is that, in general the coefficient $\Gamma_{11}(\omega) \neq 0$. There exist coefficients $d_{ij}(\omega)$, $0 \leq i, j \leq 3$, and vector functions $d'_{ij}(x, \omega)$ such that the function γ_1 defined as

$$(5.2) \quad \gamma_1 = \gamma + d_{20}z^2 + d_{02}\bar{z}^2 + d_{30}z^3 + d_{21}z^2\bar{z} + d_{12}z\bar{z}^2 + d_{03}\bar{z}^3 + z\langle f, d'_{10} \rangle + \bar{z}\langle f, d'_{01} \rangle$$

with $d_{ij} = \bar{d}_{ji}$, is a solution of the differential equation

$$(5.3) \quad \dot{\gamma}_1 = \Gamma_{11}(\omega)z\bar{z} + \widehat{\Gamma}_R$$

and $\widehat{\Gamma}_R$ satisfies the same estimate (2.33) as Γ_R . Notice that $\Gamma_{11}(\omega) = -\frac{2E_2[u, u^*]}{\langle \phi, \phi_\omega \rangle}$ is real.

PROPOSITION 5.1. *Under the hypotheses of Theorem 4.4, the functions $\omega(t)$, $z(t)$ and $\gamma(t)$ have the following asymptotic behavior as $t \rightarrow \infty$*

$$(5.4) \quad \omega \sim \omega_+ + \frac{b_+ \epsilon}{1 + k_+ \epsilon t} \cos(2\mu_+ t + b_1 \log(1 + k_+ \epsilon t) + b_2),$$

$$(5.5) \quad z(t) \sim \zeta_+ \epsilon^{1/2} \frac{e^{i\mu_+ t}}{(1 + k_+ \epsilon t)^{\frac{1}{2} - i\delta}}, \quad |\zeta_+| = 1 + O(\epsilon^{1/2}),$$

$$(5.6) \quad \gamma(t) = \gamma_\infty + c_+ \log(1 + k_+ \epsilon t) + O\left(\frac{\epsilon}{1 + \epsilon t}\right).$$

where $b_+, b_1, b_2, \gamma_+, c_+, k_+, \gamma_\infty$ are some constants, $k_+ > 0$, and $\mu_+ = \mu(\omega_+)$.

We finally turn to the function f . We know from (4.20) and (4.22) that in norms $\|\cdot\|_\rho$ and $\|\cdot\|_\infty$, f tends to 0 as $t \rightarrow \infty$. However, this radiative part carries non-zero energy and other integrals of motion, and from this point of view it is not negligible. To control the contribution of f to the integrals of motion, we have to study its asymptotic behavior in the usual L^2 norm. We recall the representation

$$(5.7) \quad f = g + h, \quad g = P_d(\omega_+)f, \quad h = P_c(\omega_+)f.$$

and

$$(5.8) \quad \|f\|_2 = \|h\|_2 + O\left(\left(\frac{\epsilon}{1+\epsilon t}\right)^2\right).$$

The function h is solution of

$$(5.9) \quad \dot{h} = C_+h - P_c(\omega_+)jE_2[w, w] + H_R,$$

where the operator C_+ is defined in (4.21).

PROPOSITION 5.2. *We have the asymptotic formula*

$$(5.10) \quad e^{C(\omega_+)t}\tilde{h} = e^{C_0t}h_+ + o(1)$$

as $t \rightarrow \infty$, where $h_+ = W\tilde{h}$ and W is a bounded operator in $L^2(\mathbb{R})$, that can be seen as a wave operator. It is the strong limit in L^2 of $e^{-C_0t}e^{C(\omega_+)t}$, with $C_0 = j^{-1}(-\partial_{xx} + \omega_+)$.

Combining the asymptotic formulas for ω, γ, z , and f , we get

THEOREM 5.3. *Suppose that the hypothesis of Theorem 4.4 are satisfied. Then for ϵ small enough the solution of the NLS equation (1.10) have the following behavior as $t \rightarrow \infty$*

$$(5.11) \quad \psi(x, t) = e^{j(\omega_+t + \gamma_+(t) + \kappa)} [\phi(x, \omega_+) + z_+(t)u(x, \omega_+) + \bar{z}_+(t)u^*(x, \omega_+)] + e^{j^{-1}Lt}h_+ + o(1)$$

in L^2 , where $L = -\partial_{xx}$ and

$$(5.12) \quad \kappa = \int_0^\infty (\omega(t_1) - \omega_+)dt_1 + \gamma_\infty, \quad \gamma_+(t) = c_+ \log(1 + k_+\epsilon t).$$

This result can be rewritten in terms of the original complex notation. As $t \rightarrow \infty$, the solution of the NLS equation (1.1) behaves as follows

$$(5.13) \quad \psi = e^{i(\omega_+t + \gamma_+(t) + \kappa)} [\varphi(x, \omega_+) + z_+(t)v_+(x, \omega_+) + \bar{z}_+(t)v_-(x, \omega_+)] + e^{-iLt}h_+ + o(1)$$

with $v_\pm(x, \omega) = u_1(x, \omega) \pm iu_2(x, \omega)$ and $h_+ = (h_+)_1 + i(h_+)_2$.

References

- [1] Buslaev, V.S and Perelman, G., *Scattering for the nonlinear Schrödinger equation : states close to a soliton*, St. Petersburg Math. J. **4**(1993), 1111–1142.
- [2] Buslaev, V.S and Perelman, G., *On the stability of solitary waves for nonlinear Schrödinger equations*, Amer. Math. Soc. Transl. **164**(1995), 75–98.
- [3] Buslaev, V.S and Sulem, C., *On the asymptotics stability of solitary waves of nonlinear Schrödinger equations*, to appear in Annales Institut H. Poincaré, Analyse nonlinéaire, 2002.
- [4] Cazenave, T. and Lions, P.L., *Orbital stability of standing waves for some nonlinear Schrödinger equations*, Comm. Math. Phys. **85**(1982), 549–561.

- [5] Comech, A., *On orbital stability of quasistationary solitary waves of minimal energy*, preprint.
- [6] Cuccagna, S., *Stabilization of solutions to nonlinear Schrödinger equations*, Comm. Pure Appl. Math. **LIV**(2001) 1110-1145.
- [7] Deift, P. and Zhou, X., *Long-time asymptotics for integrable Nonlinear wave equations. Important developments in Soliton theory 1980-1990*, Springer-Verlag, 1993 181–204.
- [8] Deift, P. and Zhou, X., *Perturbation theory for infinite dimensional integrable systems on the line*, Preprint.
- [9] Ginibre, J. and Velo, G., *On a class of Schrödinger equations. I: The Cauchy problem, General case; II: Scattering theory, General case*, J. Funct. Anal. **32**(1979), 1–32, 33–71.
- [10] Grikurov, V.E., *Perturbation of unstable solitons for generalized NLS with saturating nonlinearity*, Intern. Seminar ‘Day on Diffraction-97’, 1997, 170-179.
- [11] Grillakis, M., Shatah, J., and Strauss, W., *Stability theory of solitary waves in the presence of symmetry*, Part I, J. Funct. Anal. **74**(1987), 160–197; Part II, J. Funct. Anal. **94**(1990), 308–348.
- [12] McKean, H.P. and Shatah, J., *The nonlinear Schrödinger equation and the nonlinear heat equation reduction to linear form*, Comm. Pure Appl. Math. **44**(1991), 1067–1083.
- [13] Perelman, G., *On the formation of singularities in solutions of the critical nonlinear Schrödinger equation*, Ann. Henri Poincaré, **2**(2001), 605-673.
- [14] Pelinovsky, D., Kivshar, Yu., and Afanasjev, V.V., *Internal modes of envelope solitons* Physica **D 116**, 121–142.
- [15] Soffer, A. and Weinstein, M., *Multichannel nonlinear scattering for non-integrable equations*, Comm. Math. Phys. **133**(1990), 119–146. *II. The case of anisotropic potentials and data*, J. Diff. Eq., **98**(1992), 376–390.
- [16] Soffer, A. and Weinstein, M., *Resonances, radiation damping and instability in Hamiltonian nonlinear wave equations*, Invent. Math, **136**, 1999, 9–74.
- [17] Strauss, W., *Nonlinear scattering theory at low energy*, J. Funct. Anal. **41**(1981), 110–133; J. Funct. Anal. **43**(1981), 281–293.
- [18] Sulem, C. and Sulem, P.-L., *The Nonlinear Schrödinger Equation: Self-focusing and Wave Collapse*, Applied Mathematical Sciences, **139**, 1999, Springer.
- [19] Weinstein, M., *Lyapunov stability of ground states of nonlinear dispersive evolution equations*, Comm. Pure Appl. Math. **39**(1986), 51–68.
- [20] Tsai, T.-P. and Yau H.-T, *Asymptotic dynamics of nonlinear Schrödinger equations: resonance dominated and radiation dominated solutions*, Comm. Pure Appl. Math. **LV**(2002) 1-64.
- [21] Tsai, T.-P. and Yau H.-T, *Relaxation of excited states in nonlinear Schrödinger equations*, preprint; *Stable directions for excited states of nonlinear Schrödinger equations*, preprint.
- [22] Vakhitov, N.G. and Kolokolov, A.A., *Stationary solutions of the wave equation in a medium with nonlinearity saturation*, Radiophys. Quantum Electronics, **16**(1973), 783–789.
- [23] Zakharov, V.E and Manakov, S.V, *Asymptotic behavior of nonlinear wave systems integrated by the inverse method*, Sov. Phys. JETP, **44**(1976), 106-112.

INSTITUTE FOR PHYSICS, ST. PETERSBURG STATE UNIVERSITY, ST. PETERSBURG-PETRODVORETS,
198904, RUSSIA

E-mail address: buslaev@mph.phys.spbu.ru

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TORONTO, TORONTO, M5S 3G3 CANADA

E-mail address: sulem@math.toronto.edu

This page intentionally left blank

Finite-time Blow-up in the Additive Supercritical Stochastic Nonlinear Schrödinger Equation : the Real Noise Case

A. de Bouard and A. Debussche

ABSTRACT. We review some results concerning the apparition of finite time singularities in nonlinear Schrödinger equations with a Gaussian additive noise which is white in time and correlated in space. We then extend the results to the case where the noise is real valued, which is the case in some physical situations.

1. Introduction

The nonlinear Schrödinger (NLS) equation is a generic equation describing the propagation of weakly nonlinear waves in strongly dispersive media. It is found in diverse fields of physics, such as hydrodynamics, plasma physics, nonlinear optics, or molecular biology, where it appears to be the continuum limit of certain discrete systems (see [2] and the references therein).

Recently, interest has grown up in the influence of Gaussian white noise on the dynamical behaviour of solutions of this equation; especially, in the focusing case, propagation of soliton solutions in the presence of noise has been the subject of several investigations.

A one dimensional NLS equation with additive Gaussian space-time white noise is e.g. considered in [7], with the aim of computing error probability in signal transmissions.

Another example of NLS equation with noise is given in [1] and [2], where it describes energy transfer in monolayer molecular aggregates, and where the noise stands for thermal fluctuations. As explained in [2], this noise may be multiplicative if it describes process where excitation is not being created or destroyed and in this case the noise appears in the equation as a linear potential. It may also be additive in the case of an exciton that creates or absorbs a photon. In both cases, the noise is real valued and depends on space and time variables.

Here, we consider the stochastic nonlinear Schrödinger equation

$$(1.1) \quad i\partial_t\psi - (\Delta\psi + |\psi|^{2\sigma}\psi) = \xi$$

1991 *Mathematics Subject Classification.* 35Q55, 60H15, 76B35.

Key words and phrases. Nonlinear Schrödinger Equations, Stochastic Partial Differential Equations, White Noise, Blow up, Variance Identity.

in general dimension n . The noise ξ is an additive real valued Gaussian noise, which is white in time and correlated in space. The nonlinear term $|\psi|^{2\sigma}\psi$ is a supercritical power of the solution, and our aim is to investigate the possible blow-up of solutions.

It is well known indeed that when there is no noise, i.e. $\xi = 0$ in equation (1.1), and when σ satisfies $\frac{2}{n} \leq \sigma < \frac{2}{n-2}$ ($\sigma \geq \frac{2}{n}$ in dimension $n = 1$ or 2), a solution of (1.1) (with $\xi = 0$) starting from ψ_0 with a finite negative energy, that is with

$$H(\psi_0) = \frac{1}{2} \int |\nabla\psi_0(x)|^2 dx - \frac{1}{2\sigma+2} \int |\psi_0(x)|^{2\sigma+2} dx < 0$$

cannot be globally well defined. More precisely, there is a positive t^* such that

$$(1.2) \quad \lim_{t \nearrow t^*} \int_{\mathbb{R}^n} |\nabla\psi(t, x)|^2 dx = +\infty.$$

The ingredient of the proof of such a fact uses what is sometimes called the “variance identity” (see [8], [11], [12]) which consists in computing the second order time derivative of the quantity

$$V(\psi(t)) = \int_{\mathbb{R}^n} |x|^2 |\psi(t, x)|^2 dx.$$

Using the equation satisfied by ψ and the fact that the energy H is a conserved quantity for the deterministic equation, it is indeed possible to show that under the preceding conditions on σ ,

$$(1.3) \quad V(\psi(t)) \leq V(\psi_0) + t \frac{d}{dt} V(\psi(t))|_{t=t_0} + 8H(\psi_0)t^2.$$

$V(\psi)$ being a nonnegative quantity, this inequality cannot remain true for all time if e.g. the energy $H(\psi_0)$ is negative, and it leads to (1.2). The condition $H(\psi_0) < 0$ is of course far from necessary in order that the solution blows up, and some much more precise criteria may be exhibited (see [10]).

We have generalized in [5] this identity to the stochastic equation (1.1), where ξ is a complex valued noise which is correlated in space and white in time. In this case, the solution is a random process, which is defined on a random time interval $[0, \tau^*(\psi_0))$, provided that ψ is sufficiently correlated in space, as was proved in [4]. Assuming that for some deterministic $t > 0$, one has $t < \tau^*(\psi_0)$ almost surely and that

$$\mathbb{E} \int_0^t \left(\int_{\mathbb{R}^n} |\nabla\psi(s, x)|^2 dx + \left(\int_{\mathbb{R}^n} |\psi(s, x)|^{2\sigma+2} dx \right)^{\frac{2\sigma+1}{\sigma+1}} ds \right) < +\infty,$$

we have proved that $\mathbb{E}(V(\psi(t)))$ satisfies an inequality of the form (1.3) where the right hand side is replaced by a third order polynomial in time; here, the expression $\mathbb{E}(v)$ stands for the mathematical expectation, or mean value, of the random variable v . In this third order polynomial, the coefficient of t^3 depends only on the covariance operator of the noise, while the coefficient of t^2 is $8\mathbb{E}(H(\psi_0))$. Hence, by choosing ψ_0 such that $\mathbb{E}(H(\psi_0))$ is sufficiently negative, again the inequality cannot remain true for all positive time because the right hand side takes negative values, and the solution necessarily blows up (see Proposition 2.3 for a precise definition of blow-up).

This result does not make use of the fact that the noise is complex valued, and it is true with exactly the same proof in the present case of a real valued noise.

We then made use in [5] of a control argument to show that if the noise is nondegenerate, the stochastic equation is irreducible in the sense that for any time

$T > 0$, initial data ψ_0 and final data ψ_T , the solution of (1.1) with $\psi(0) = \psi_0$ is close at time T to ψ_T with a positive probability. Choosing then ψ_T with sufficiently negative energy allowed us to apply the stochastic variance identity, to the solution of (1.1) starting from ψ_T at time T , so that this solution finally blows up. Hence, blow-up occurs for any initial data.

Contrary to the variance identity, the control argument leading to the irreducibility of the equation strongly uses the fact that the noise is complex valued, since in this case it is sufficient to control the equation with a complex valued deterministic forcing term. The aim of the present note is first to review the existence and blow-up results that were previously obtained and then to show that the control argument is still valid in the real valued case, implying the same result as in the complex valued case, that is any solution of (1.1) blows up in finite time if the noise is sufficiently correlated.

Note that on the opposite case of a completely uncorrelated noise in space – that is a space-time white noise – even though we are not able to prove any theoretical result, some numerical computations have been performed in [6], which seem to indicate that if the noise is multiplicative and arises as a Stratonovitch potential then it will tend to prevent the blow-up phenomenon.

We now describe more precisely the noise that we consider. We introduce a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, endowed with a filtration $(\mathcal{F}_t)_{t \geq 0}$, and a sequence $(\beta_k)_{k \in \mathbb{N}}$ of independent real valued Brownian motions on \mathbb{R}^+ associated to the filtration $(\mathcal{F}_t)_{t \geq 0}$. We then consider a complete orthonormal system $(e_k)_{k \in \mathbb{N}}$ in the space of real valued square integrable functions on \mathbb{R}^n , and a bounded linear operator Φ on this space. The process

$$W(t, x, \omega) = \sum_{k=0}^{\infty} \beta_k(t, \omega) \Phi e_k(x), \quad t \geq 0, \quad x \in \mathbb{R}^n, \quad \omega \in \Omega,$$

is then a Wiener process on the space of real valued square integrable functions on \mathbb{R}^n , with covariance operator $t\Phi\Phi^*$. We then set $\xi = \frac{\partial W}{\partial t}$. Note that if Φ is defined through a real valued kernel \mathcal{K} , which means that for any real valued square integrable function u ,

$$\Phi u(x) = \int_{\mathbb{R}^n} \mathcal{K}(x, y) u(y) dy,$$

then the correlation function of the noise is given by

$$\mathbb{E} \left(\frac{\partial W}{\partial t}(t, x) \frac{\partial W}{\partial t}(s, y) \right) = c(x, y) \delta_{t-s}$$

with

$$c(x, y) = \int_{\mathbb{R}^n} \mathcal{K}(x, z) \mathcal{K}(y, z) dz.$$

We then write equation (1.1) as

$$(1.4) \quad i d\psi - (\Delta\psi + |\psi|^{2\sigma}\psi) dt = dW.$$

Note that in the physical situations described at the beginning, the correlation function $c(x, y)$ is a Dirac delta function, corresponding to space-time white noise (in this case, Φ is the identical operator). We are not able to treat that case for

two reasons. To understand them, one should consider the linear equation

$$\begin{cases} idz - \Delta z dt = dW \\ z(0) = 0 \end{cases}$$

whose solution is given by the stochastic integral

$$(1.5) \quad z(t) = \int_0^t S(t-s) dW(s)$$

where $S(t) = e^{-it\Delta}$ is the group associated with the linear Schrödinger equation. Since $S(t)$ is a unitary group in any Sobolev space $H^s(\mathbb{R}^n)$, it is easy to see that $z(t)$ lies in $H^s(\mathbb{R}^n)$ almost surely if and only if Φ is a Hilbert-Schmidt operator from $L^2(\mathbb{R}^n)$ into $H^s(\mathbb{R}^n)$. Note indeed the identity

$$|z(t)|_{H^s(\mathbb{R}^n)}^2 = \left| \int_0^t S(t-s) dW(s) \right|_{H^s(\mathbb{R}^n)}^2 = t \|\Phi\|_{HS(L^2, H^s)}^2$$

where $\|\Phi\|_{HS(L^2, H^s)}^2 \equiv \sum_k |\Phi e_k|_{H^s(\mathbb{R}^n)}^2$ is the Hilbert-Schmidt norm of Φ as an operator from $L^2(\mathbb{R}^n)$ into $H^s(\mathbb{R}^n)$.

However, it is easy to see that a convolution operator – i.e. an operator defined through a kernel $\mathcal{K}(x, y) = k(x - y)$ – will never be Hilbert-Schmidt from $L^2(\mathbb{R}^n)$ into $H^s(\mathbb{R}^n)$, even if s is largely negative. This proves that the integral $z(t)$ cannot live in $H^s(\mathbb{R}^n)$ if the noise is homogeneous. This is the first reason : homogeneity of the noise.

The second reason is the irregularity of the correlations : even if one adds some localization in the correlations of the noise – e.g. if Φ is given by a kernel $\mathcal{K}(x, y) = k(x)\delta_{x-y}$ – there is no hope that $z(t)$ lies in a more regular space than $H^{-n/2}(\mathbb{R}^n)$. However, it has been proved (see [9]) that the deterministic conservative NLS equation is ill posed in any $H^s(\mathbb{R}^n)$ with negative s .

This implies in particular that treating the stochastic term as a perturbation by using the integral $z(t)$ will never lead to the existence of a strong solution of the stochastic equation with a space-time white noise, as long as we deal with H^s Sobolev spaces ; note that the H^s spaces have revealed to be very natural spaces to handle the deterministic NLS equation.

Anyway, we only consider correlated noise in this note, which means that we will require from Φ sufficient regularization properties, and the above mentioned problem will not appear here.

The note is organized as follows : in Section 2, we recall the results proved in [4] and [5] concerning the existence of solutions and blow-up for some initial data. Those results were proved in the context of a complex valued noise, but they hold with exactly the same proof for equation (1.4) with a real valued noise, so that we do not recall the proofs. we will give for each particular result the minimal assumptions required on Φ , and on the initial data. In Section 3, we prove that the controlability problem allowing to deduce the irreducibility of equation (1.4) has a solution in the real valued case – and here the proof is different and more complicated than in the complex valued case. We then deduce the irreducibility as in [5] from this result and from the continuity with respect to the forcing term in the equation. Finally, irreducibility together with the blow-up result of Section 2 implies as in [5] the blow-up for any initial data (see Theorem 3.1).

2. Review of existence and blow-up for a restricted class of initial data

We start with some local and global existence results. All these results are proved in [4].

2.1. Local and global existence results.

THEOREM 2.1. *Assume that $0 \leq \sigma < 2/(n - 2)$ if $n \geq 3$ or $0 \leq \sigma$ for $n = 1, 2$, that Φ is Hilbert-Schmidt from $L^2(\mathbb{R}^n)$ into $H^1(\mathbb{R}^n)$ and that the initial data ψ_0 is a \mathcal{F}_0 measurable random variable with values in $H^1(\mathbb{R}^n)$; then there exists a unique solution ψ to (1.4) with continuous H^1 valued paths, such that $\psi(0) = \psi_0$. This solution is defined on a random interval $[0, \tau^*(\psi_0))$, where $\tau^*(\psi_0)$ is a stopping time such that*

$$\tau^*(\psi_0) = +\infty \text{ or } \lim_{t \nearrow \tau^*(\psi_0)} \|\psi(t)\|_{H^1(\mathbb{R}^n)} = +\infty.$$

Furthermore, τ^* is almost surely lower semicontinuous with respect to ψ_0 .

In order to prove the global existence result in the subcritical case $\sigma < 2/n$, the following invariant quantities of the deterministic NLS equation have been used in [4] : the momentum

$$M(\psi) = \int_{\mathbb{R}^n} |\psi(x)|^2 dx$$

and the Hamiltonian

$$H(\psi) = \frac{1}{2} \int_{\mathbb{R}^n} |\nabla \psi(x)|^2 dx - \frac{1}{2\sigma + 2} \int_{\mathbb{R}^n} |\psi(x)|^{2\sigma+2} dx.$$

The evolution of these quantities along the solutions of the stochastic equation (1.4) is described in the next proposition.

PROPOSITION 2.1. *Let ψ_0, σ and Φ be as in Theorem 2.1. For any stopping time τ such that $\tau < \tau^*(\psi_0)$ a.s., we have*

$$(2.1) \quad M(\psi(\tau)) = M(\psi_0) - 2Im \sum_{\ell \in \mathbb{N}} \int_0^\tau \int_{\mathbb{R}^n} \psi(x) \Phi_{e_\ell}(x) dx d\beta_\ell(s) + \tau \sum_{\ell} |\Phi_{e_\ell}|_{L^2}^2$$

where ψ is the solution of (1.4) given by Theorem 2.1 with $\psi(0) = \psi_0$.

Moreover, for any $k \in \mathbb{N}$,

$$(2.2) \quad \mathbb{E} \left[\sup_{t \in [0, \tau]} M^k(\psi(t)) \right] \leq C_k \mathbb{E} [M^k(\psi_0)]$$

for a constant $C_k \geq 0$.

In the same way, for any τ such that $\tau < \tau^*(\psi_0)$ a.s. we have

$$(2.3) \quad \begin{aligned} H(\psi(\tau)) &= H(\psi_0) - Im \int_{\mathbb{R}^n} \int_0^\tau (\Delta \bar{\psi} + |\psi|^{2\sigma} \bar{\psi}) dW dx \\ &+ \frac{1}{2} \sum_{\ell \in \mathbb{N}} \int_0^\tau \int_{\mathbb{R}^n} |\nabla \Phi_{e_\ell}|^2 dx ds \\ &- \frac{1}{2} \sum_{\ell \in \mathbb{N}} \int_0^\tau \int_{\mathbb{R}^n} \left[|\psi|^{2\sigma} |\Phi_{e_\ell}|^2 + 2\sigma |\psi|^{2\sigma-2} (Re(\bar{\psi} \Phi_{e_\ell}))^2 \right] dx ds \end{aligned}$$

where $\psi(\cdot)$ is the solution of (1.4) given by Theorem 2.1 with $\psi(0) = \psi_0$.

Using the preceding proposition, the following global existence result was proved in [4] in the subcritical case.

THEOREM 2.2. *If in addition to the assumptions of Theorem 2.1, $\sigma < 2/n$, then for any \mathcal{F}_0 -measurable ψ_0 , the solution of (1.4) with $\psi(0) = \psi_0$ given by Theorem 2.1 is global, i.e. $\tau^*(\psi_0) = +\infty$ a.s.*

Note that the result of Theorem 2.1 is still true with L^2 solutions instead of H^1 solutions, if $0 \leq \sigma \leq 2/n$ and if Φ is only Hilbert-Schmidt in $L^2(\mathbb{R}^n)$. In this case, the solutions are global, due to the estimate (2.1) on the L^2 norm.

2.2. Blow-up for some initial data. The blow-up result for a restricted class of initial data, which is proved exactly as in [5], is based on Proposition 2.1, together with another identity, which we call the “stochastic variance identity”. This identity is proved in [5] in the case of a complex valued noise – the real noise case is proved exactly in the same way – and requires slightly more regularity on Φ and ψ_0 . In order to state precisely the assumptions we need, we introduce the space

$$\Sigma = \left\{ v \in H^1(\mathbb{R}^n), \int_{\mathbb{R}^n} |x|^2 |v(x)|^2 dx < \infty \right\}.$$

PROPOSITION 2.2. *Let ψ_0, σ and Φ be as in Theorem 2.1, and assume furthermore that Φ is Hilbert-Schmidt form $L^2(\mathbb{R}^n)$ into Σ and that ψ_0 lies almost surely in Σ . Then for any stopping time τ such that $\tau < \tau^*(\psi_0)$ a.s. the solution ψ of (1.4) with $\psi(0) = \psi_0$ belongs to $L^\infty(0, \tau; \Sigma)$ a.s. and satisfies*

$$\begin{aligned} V(\psi(\tau)) = & V(\psi_0) + 4G(\psi_0)\tau + 8H(\psi_0)\tau^2 + 4\frac{2 - \sigma n}{\sigma + 1} \int_0^\tau (\tau - s) |\psi|_{L^{2\sigma+2}}^{2\sigma+2} ds \\ & + c_\Phi^\Sigma \tau + \frac{4}{3} c_\Phi^1 \tau^3 \\ & - 4 \sum_{\ell \in \mathbb{N}} \int_0^\tau (\tau - s)^2 \int_{\mathbb{R}^n} |\psi|^{2\sigma} |\Phi e_\ell|^2 + 2\sigma |\psi|^{2\sigma-2} (\text{Re}(\bar{\psi} \Phi e_\ell))^2 dx ds \\ & + 2 \text{Im} \int_{\mathbb{R}^n} \int_0^\tau |x|^2 \bar{\psi} dW dx \\ & - 16 \text{Im} \int_{\mathbb{R}^n} \int_0^\tau \int_{s_0}^{s_2} \int_0^{s_1} (\Delta \bar{\psi} + |\psi|^{2\sigma} \bar{\psi}) dW(r) ds_1 ds_2 dx \\ & + 4 \text{Re} \sum_{\ell \in \mathbb{N}} \int_0^\tau \int_0^\tau \int_{\mathbb{R}^n} \bar{\psi} (2x \cdot \nabla \Phi e_\ell + n \Phi e_\ell) dx d\beta_\ell ds \end{aligned}$$

with

$$G(v) = \text{Im} \int_{\mathbb{R}^n} v(x) x \cdot \nabla \bar{v}(x) dx$$

for $v \in \Sigma$, and with

$$c_\Phi^\Sigma = \sum_{\ell \in \mathbb{N}} \int_{\mathbb{R}^n} |x|^2 |\Phi e_\ell|^2 dx \text{ and } c_\Phi^1 = \sum_{\ell \in \mathbb{N}} |\nabla \phi e_\ell|_{L^2}^2.$$

Note that the first four terms in this identity already occur in the deterministic identity, and that the other terms vanish in the absence of noise. The last three terms are stochastic integrals and are responsible for technical difficulties. In the particular case where $\tau = T$ is a deterministic time, the mean value of these last three terms vanishes. Let us denote $\mathbb{E}_{\Omega_0}(f) = \mathbb{E}(f \mathbf{1}_{\Omega_0}) / \mathbb{P}(\Omega_0)$ for $f \in L^1(\Omega_0)$ or f measurable and nonnegative on Ω_0 , and Ω_0 a \mathcal{F}_0 measurable set.

In the preceding case of a deterministic $\tau = T$, and if we assume moreover that

$$\mathbb{E}_{\Omega_0} \left(\int_0^T |\nabla\psi(s)|_{L^2}^2 + |\psi(s)|_{L^{2\sigma+2}}^{2\sigma+2} ds \right)$$

is finite, then Proposition 2.1 implies that

$$\mathbb{E}_{\Omega_0}(V(\psi(t))) \leq \mathbb{E}_{\Omega_0}(V(\psi_0)) + (4\mathbb{E}_{\Omega_0}(|G(\psi_0)|) + c_{\Phi}^{\Sigma})t + 8\mathbb{E}_{\Omega_0}(H(\psi_0))t^2 + \frac{4}{3}c_{\Phi}^1 t^3$$

for any $t \in [0, T]$. Now, it is possible to choose ψ_0 in such a way that the right hand side polynomial takes negative values for some $t \in [0, T]$, leading to a contradiction since the left hand side is nonnegative. More precisely, we have

PROPOSITION 2.3. *Assume that $2/n \leq \sigma < 2/(n - 2)$, and that Φ is Hilbert-Schmidt from $L^2(\mathbb{R}^n)$ into Σ . For each $\bar{T} > 0$, $\bar{V} > 0$, $\bar{G} > 0$, $\bar{H}_1 > 0$, let \bar{H}_2 be such that*

$$(2.4) \quad \bar{V} + (4\bar{G} + c_{\Phi}^{\Sigma})\bar{T} + 4 \left(\bar{H}_1 - \frac{1}{\sigma + 1} \bar{H}_2 \right) \bar{T}^2 + \frac{4}{3}c_{\Phi}^1 \bar{T}^3 < 0;$$

then for each \mathcal{F}_0 measurable ψ_0 with values in Σ and for any $\Omega_0 \in \mathcal{F}_0$ with $\mathbb{P}(\Omega_0) > 0$ such that

$$(2.5) \quad \begin{aligned} \mathbb{E}_{\Omega_0} (|\nabla\psi_0|_{L^2}^2) &\leq \bar{H}_1, & \mathbb{E}_{\Omega_0} (|G(\psi_0)|) &\leq \bar{G} \\ \mathbb{E}_{\Omega_0} (V(\psi_0)) &\leq \bar{V} & \text{and } \mathbb{E}_{\Omega_0} (|\psi_0|_{L^{2\sigma+2}}^{2\sigma+2}) &\geq \bar{H}_2, \end{aligned}$$

then either

$$(2.6) \quad \mathbb{P}(\tau^*(\psi_0) < \bar{T}) \text{ is positive}$$

or

$$(2.7) \quad \mathbb{E}_{\Omega_0} \left(\int_0^{\bar{T}} (|\nabla\psi(s)|_{L^2}^2 + |\psi(s)|_{L^{2\sigma+2}}^{4\sigma+2}) ds \right) = +\infty,$$

where ψ is the solution of (1.4) with $\psi(0) = \psi_0$ given by Theorem 2.1.

The possibility that (2.7) occurs instead of (2.6) is due to the fact that we had to choose a deterministic time T in the argument above in order to cancel the mean value of the stochastic terms in the stochastic variance identity. If instead we use a stopping time, then the expectation of those terms do not vanish. Under more restrictive assumptions on σ and Φ , however, they may be handled and lead to the following result.

PROPOSITION 2.4. *Assume that $2/n < \sigma < \min(\frac{2}{3}, \frac{2}{n-2})$, that Φ is Hilbert-Schmidt from $L^2(\mathbb{R}^n)$ into Σ , with moreover $\sum_{\ell \in \mathbb{N}} |\Phi e_{\ell}|_{L^{4\sigma+2}}^2 < \infty$, and bounded from $L^2(\mathbb{R}^n)$ into $H^2(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$. Let $\bar{T}, \bar{V}, \bar{G}, \bar{H}_1, \bar{H}_2, \Omega_0$ and ψ_0 be as in Proposition 2.2, with moreover, $\mathbb{E}_{\Omega_0}(M(\psi_0)^{1/(1-\sigma)}) < \infty$; then*

$$\mathbb{P}(\tau^*(\psi_0) < \bar{T}) > 0.$$

Note that the assumptions on the power σ in Proposition 2.4 are compatible only when $n \geq 4$.

3. The controlability problem and blow-up for any initial data

In this section, we prove the irreducibility of equation (1.4), or equivalently of the following equation :

$$(3.1) \quad v(t) = S(t)\psi_0 - i \int_0^t S(t-s) (|v+z|^{2\sigma}(v+z)) ds$$

where z is given by (1.5), in the case where the noise is nondegenerate. (Note that if z is given by (1.5) and v satisfies (3.1), then $\psi = v + z$ is a solution of (1.4) with $\psi(0) = \psi_0$.) This is done by using the following controlability result, which was already proved in [5] in the case of a complex valued noise, together with some continuity property . The controlability result is more difficult to prove in the real valued case, because we have to control the NLS equation by using a real valued forcing term instead of a complex valued one, which means that we can act only on one of the two components of the solution. The result is also weaker in the sense that it requires more smoothness on the forcing term and on ψ_0 . Of course, this will affect the main result in consequence (see Theorem 3.1).

We define some other spaces which are necessary to state the assumptions : for $k, \ell \in \mathbb{N}$ we define

$$\Sigma^{k,\ell} = \{v \in H^{k+\ell}(\mathbb{R}^n), |x|^\ell v \in H^k(\mathbb{R}^n)\}$$

endowed with the natural norm

$$|v|_{\Sigma^{k,\ell}}^2 = |v|_{H^{k+\ell}}^2 + ||x|^\ell v|_{H^k}^2$$

and let S_n be the following space (recall that n is here the dimension of the space variable) :

$$S_1 = \Sigma, \quad S_2 = \cup_{\alpha>0} \Sigma^{\alpha,1}, \quad S_n = \Sigma^{1,2} \text{ if } n \geq 3.$$

In order to understand how Proposition 3.1 below is related to the controlability of the NLS equation by a real valued forcing, one only needs to note that \tilde{u} is a solution of

$$(3.2) \quad \begin{cases} i \frac{d\tilde{u}}{dt} - (\Delta\tilde{u} + |\tilde{u}|^{2\sigma}\tilde{u}) = \frac{df}{dt} \\ \tilde{u}(0) = \psi_0 \end{cases}$$

if and only if $\tilde{u} = v + \tilde{z}$ with \tilde{z} satisfying

$$(3.3) \quad \begin{cases} i \frac{d\tilde{z}}{dt} - \Delta\tilde{z} = \frac{df}{dt}, \\ \tilde{z}(0) = 0, \end{cases}$$

and v satisfies equation (3.1) with z replaced by \tilde{z} . In addition, $f(t)$ is real valued if and only if $i\tilde{z}(t) - \int_0^t \Delta\tilde{z}(s)ds$ is real valued.

PROPOSITION 3.1. *For any $T_1 > 0$, $\psi_0 \in S_n \cap H^s$, $s > n/2$, for any real valued function b_1 in S_n , there exists a \tilde{z} in $C([0, T_1]; S_n)$ such that $\tilde{z}(0) = 0$, $i\tilde{z}(t) - \int_0^t \Delta\tilde{z}(s)ds$ is real valued for any $t \in [0, T_1]$ and such that the solution $v(\tilde{z}, \psi_0, \cdot)$ of (3.1) exists on $[0, T_1]$, with*

$$Im(\tilde{z}(T_1) + v(\tilde{z}, \psi_0, T_1)) = b_1.$$

Moreover, for any $\delta > 0$, \tilde{z} may be chosen so that

$$|Re(\tilde{z}(T_1) + v(\tilde{z}, \psi_0, T_1)) - Re(\psi_0)|_{S_n} \leq \delta.$$

REMARK 3.1. *This result is weaker than what we proved in [5] for a complex valued noise in two ways. First, we are not able to choose $\tilde{z}(\cdot)$ such that $\tilde{z}(T_1) + v(\tilde{z}, \psi_0, T_1) = u_1$ for some fixed u_1 ; only the imaginary part can be controled exactly. However this is not a problem for our purpose as will be made clear later. Also, here we have to assume that ψ_0, b_1 have extra smoothness assumptions if $n \geq 2$: they are assumed to be in $S_n \cap H^s(\mathbb{R}^n)$. This is the reason why Theorem 3.1 below is restricted to initial data in $S_n \cap H^s(\mathbb{R}^n)$.*

Proof of Proposition 3.1. It is convenient in this proof to decompose the solution of the nonlinear Schrödinger equation into its real and imaginary parts. We first construct the forcing f in (3.2) and then deduce \tilde{z} solution of (3.3) by the formula

$$\tilde{z}(t) = -if(t) - \int_0^t S(t-s)(i\Delta)f(s)ds.$$

We note that \tilde{u} solves

$$i \frac{d\tilde{u}}{dt} - (\Delta\tilde{u} + |\tilde{u}|^{2\sigma}\tilde{u}) = \frac{df}{dt}$$

for a real valued f if its real and imaginary parts a and b solve

$$(3.4) \quad \frac{da}{dt} - \Delta b - (a^2 + b^2)^\sigma b = 0,$$

and

$$(3.5) \quad -\frac{db}{dt} - \Delta a - (a^2 + b^2)^\sigma a = \frac{df}{dt}.$$

The idea is to construct b explicitly such that $b(0) = \text{Im}(\psi_0)$, $b(T_1) = b_1$ and $b = 0$ on a large interval in $(0, T_1)$. In that way $\frac{da}{dt} = 0$ in that interval and $a(T_1)$ is close to $a(0)$.

Take $k_1 \in \mathbb{N}$ and denote by $U(t)$ the semigroup on Σ associated to the linear equation

$$\begin{cases} \frac{dw}{dt} + (-\Delta)^{k_1}w + |x|^{2k_1}w = 0, & x \in \mathbb{R}^n, \\ w(0) = \psi_0. \end{cases}$$

For $\varepsilon > 0$ to be chosen, we set

$$b(t) = \begin{cases} \frac{\varepsilon-t}{\varepsilon}U(t) \text{Im}(\psi_0), & t \in [0, \varepsilon], \\ 0, & t \in [\varepsilon, T_1 - \varepsilon] \\ \frac{t-T_1+\varepsilon}{\varepsilon}U(T_1-t)b_1, & t \in [T_1 - \varepsilon, T_1]. \end{cases}$$

Clearly, for k_1 large enough, b is in $C([0, T_1]; S_n) \cap L^1(0, T_1; H^s(\mathbb{R}^n))$ and Δb is in $L^1(0, T_1; S_n \cap H^s(\mathbb{R}^n))$. Then, for any $\delta > 0$, there exists $\varepsilon > 0$ depending on ψ_0 and b_1 such that (3.4) has a solution a in $C([0, \varepsilon]; S_n \cap H^s(\mathbb{R}^n))$ such that

$$a(0) = \text{Re}(\psi_0)$$

and

$$|a - a(0)|_{C([0, \varepsilon]; S_n \cap H^s)} \leq \frac{\delta}{2}.$$

This can be proved by a fixed point argument.

Similarly, there exists a solution a of (3.4) on $[T_1 - \varepsilon, T_1]$ such that $a(T_1 - \varepsilon) = a(\varepsilon)$ and

$$|a - a(\varepsilon)|_{C([T_1 - \varepsilon, T_1]; S_n \cap H^s)} \leq \frac{\delta}{2}.$$

Then, since $b = 0$ on $[\varepsilon, T_1 - \varepsilon]$, setting $a(t) = a(\varepsilon)$ on $[\varepsilon, T_1 - \varepsilon]$ we obtain a solution of (3.4) on $[0, T_1]$ such that

$$|a - \operatorname{Re}(\psi_0)|_{C([0, T_1]; S_n \cap H^s)} \leq \delta.$$

We now define

$$f(t) = -b(t) + b(0) - \int_0^t \Delta a + (a^2 + b^2)^\sigma \, ads$$

and

$$\tilde{z}(t) = -f(t) - i \int_0^t S(t-s) \Delta f(s) \, ds$$

so that $v = a + ib - \tilde{z}$ solves

$$i \frac{dv}{dt} - \Delta v = |\tilde{u}|^{2\sigma} \tilde{u},$$

with $\tilde{u} = a + ib$, and for any $t \in [0, T_1]$, $i\tilde{z}(t) - \int_0^t \Delta \tilde{z}(s) \, ds = f(t)$ is real valued.

We can choose k_1 in the definition of $U(\cdot)$ sufficiently large to ensure that $|b|^{2\sigma+1} \in L^1(0, T_1; S_n)$. Moreover, since $S_n \cap H^s$ is an algebra, we also have $|a|^{2\sigma+1} \in L^1(0, T_1; S_n)$. It follows that $|\tilde{u}|^{2\sigma} \tilde{u} \in L^1(0, T_1; S_n)$ and since $(S(t))_{t \in \mathbb{R}}$ is strongly continuous in S_n and $v(0) \in S_n$, $v \in C([0, T_1]; S_n)$. Since we know that \tilde{u} also belongs to this space, we have proved :

$$\tilde{z} \in C([0, T_1]; S_n),$$

and this ends the proof of Proposition 3.1.

The following corollary of Proposition 3.1 shows that it is possible to reach a state which leads to blow-up by controlling only the imaginary part of the solution.

COROLLARY 3.1. *For any $\bar{T}_1, \bar{T} > 0$, $\psi_0 \in S_n \cap H^s(\mathbb{R}^n)$ for some $s > n/2$, there exist $\bar{V} > 0$, $\bar{G} > 0$, $\bar{H}_1 > 0$, $\bar{H}_2 > 0$ satisfying (2.4) and there exists $\tilde{z} \in C([0, T_1]; S_n)$ such that $\tilde{z}(0) = 0$, $i\tilde{z}(t) - \int_0^t \Delta \tilde{z}(s) \, ds$ is real valued for any $t \in [0, T_1]$, $v(\tilde{z}, \psi_0, \cdot)$ exists on $[0, T_1]$ and $\tilde{u}_1 = \tilde{z}(T_1) + v(\tilde{z}, \psi_0, T_1)$ verifies*

$$|\nabla \tilde{u}_1|_{L^2}^2 \leq \frac{1}{2} \bar{H}_1, \quad |G(\tilde{u}_1)| \leq \frac{1}{2} \bar{G}, \quad V(\tilde{u}_1) \leq \frac{1}{2} \bar{V}, \quad \text{and} \quad |\tilde{u}_1|_{L^{2\sigma+2}}^{2\sigma+2} \geq 2\bar{H}_2.$$

Proof. Assume first that $\operatorname{Im}(\psi_0) \neq 0$; set $u_\lambda = \operatorname{Re}(\psi_0) + i\lambda \operatorname{Im}(\psi_0)$. Then the expression

$$4V(u_\lambda) + (16G(u_\lambda) + c_\Phi^\Sigma) \bar{T} + 4 \left(4|\nabla u_\lambda|_{L^2}^2 - \frac{1}{2(2\sigma+2)} |u_\lambda|_{L^{2\sigma+2}}^{2\sigma+2} \right) \bar{T}^2 + \frac{4}{3} c_\Phi^1 \bar{T}^3$$

is negative for λ large enough, say larger than λ_0 . Then, by Proposition 3.1 with $b_1 = \lambda_0 \operatorname{Im}(\psi_0)$, given $\delta > 0$, we can find \tilde{z} with $\tilde{z}(0) = 0$, $i\tilde{z}(t) - \int_0^t \Delta \tilde{z}(s) \, ds$ is real valued for any $t \in [0, T_1]$, $v(\tilde{z}, \psi_0, \cdot)$ exists on $[0, T_1]$ and with moreover

$$\operatorname{Im}(\tilde{z}(T_1) + v(\tilde{z}, \psi_0, T_1)) = \lambda_0 \operatorname{Im}(\psi_0),$$

and

$$|\tilde{z}(T_1) + v(\tilde{z}, \psi_0, T_1) - u_{\lambda_0}|_{S_n} \leq \delta.$$

Now, the quantities $|\nabla u|_{L^2}^2$, $|G(u)|$, $V(u)$, $|u|_{L^{2\sigma+2}}^{2\sigma+2}$ depend continuously on $u \in S_n$, so that choosing δ small enough, and setting $\tilde{u}(t) = \tilde{z}(t) + v(\tilde{z}, \psi_0, t)$, we have

$$|\nabla \tilde{u}(T_1)|_{L^2}^2 \leq 2|\nabla u_{\lambda_0}|_{L^2}^2, \quad |G(\tilde{u}(T_1))| \leq 2|G(u_{\lambda_0})|, \quad V(\tilde{u}(T_1)) \leq 2V(u_{\lambda_0})$$

and

$$|u(T_1)|_{L^{2\sigma+2}}^{2\sigma+2} \geq \frac{1}{2}|u_{\lambda_0}|_{L^{2\sigma+2}}^{2\sigma+2}.$$

Hence we obtain the result with $\bar{H}_1 = 4|\nabla u_{\lambda_0}|_{L^2}^2$, $\bar{G} = 4|G(u_{\lambda_0})|$, $\bar{V} = 4V(u_{\lambda_0})$ and $\bar{H}_2 = \frac{1}{4}|u_{\lambda_0}|_{L^{2\sigma+2}}^{2\sigma+2}$.

Assume now that $\text{Im}(\psi_0) = 0$, and let us consider the solution \tilde{u} of (3.2) with $f = 0$; it is not difficult to see that the imaginary part of \tilde{u} cannot be identically zero on a whole time interval $[0, t_0]$ with $t_0 > 0$; indeed, writting \tilde{u} as $a + ib$ with a and b satisfying respectively (3.4) and (3.5), this would imply that a is a stationary solution of the deterministic NLS equation on $[0, t_0]$, but it is well known that there is no such solution in $H^1(\mathbb{R}^n)$. Hence, for some positive t_0 , one has $\text{Im}(\tilde{u}(t_0)) \neq 0$, and we can reproduce the preceding argument by starting at time t_0 from $\tilde{u}(t_0)$ (which corresponds to taking $\tilde{z} \equiv 0$ on $[0, t_0]$).

We have now all the tools in hand to state and prove the main result.

THEOREM 3.1. *Assume that $2/n \leq \sigma < \frac{2}{n-2}$ if $n \geq 3$, or $2/n \leq \sigma$ if $n = 1$ or 2 , that Φ is Hilbert-Schmidt from $L^2(\mathbb{R}^n)$ into S_n and that the null space of Φ^* is reduced to $\{0\}$. Then for any $\psi_0 \in S_n \cap H^s$ for some $s > n/2$, and for any $t > 0$ the solution $\psi(t)$ starting from ψ_0 and given by Theorem 2.1 satisfies either*

$$\mathbb{P}(\tau^*(\psi_0) < \bar{T}) > 0$$

or

$$\mathbb{E} \left(\int_0^{\bar{T}} \left(|\nabla \psi(s)|_{L^2}^2 + |\psi(s)|_{L^{2\sigma+2}}^{4\sigma+2} \right) ds \right) = +\infty.$$

If furthermore Φ satisfies the assumptions of Proposition 2.4, then ψ blows up with a positive probability.

Proof of Theorem 3.1. The proof follows exactly the same lines as the proof of Theorem 2.1 in [5], once we have Corollary 3.1 in hand. We repeat shortly the arguments for the sake of completeness. Let $\bar{T}_1, \bar{T} > 0$, and $\psi_0 \in S_n \cap H^s(\mathbb{R}^n)$ with $s > n/2$. Applying Corollary 3.1, we get $\bar{V}, \bar{G}, \bar{H}_1$, and \bar{H}_2 satisfying (2.4), and $\tilde{z} \in C([0, T_1]; S_n)$ such that if we set $\tilde{u}(t) = \tilde{z}(t) + v(\tilde{z}, \psi_0, t)$ then

$$|\nabla \tilde{u}(T_1)|_{L^2}^2 \leq \frac{1}{2}\bar{H}_1, \quad |G(\tilde{u}(T_1))| \leq \frac{1}{2}\bar{G}, \quad V(\tilde{u}(T_1)) \leq \frac{1}{2}\bar{V}, \quad \text{and} \quad |\tilde{u}(T_1)|_{L^{2\sigma+2}}^{2\sigma+2} \geq 2\bar{H}_2.$$

Now, for $t \in [0, T_1]$, the mapping $z \mapsto v(\psi_0, z, t)$ is continuous on a neighbourhood of \tilde{z} in $C([0, T_1], \Sigma) \cap L^r(0, T_1; W^{1,2\sigma+2}) \cap L^1(0, T_1; S_n)$ with values in $H^1(\mathbb{R}^n)$, and lower semi-continuous with values in Σ (see Proposition 3.4 and 3.5 in [5]). Hence, there is a ball B centered at \tilde{z} in the preceding space, such that for any $z \in B$, $u = z + v(z, \psi_0, \cdot)$ exists on $[0, T_1]$ and satisfies

(3.6)

$$|\nabla u(T_1)|_{L^2}^2 \leq \bar{H}_1, \quad |G(u(T_1))| \leq \bar{G}, \quad |V(u(T_1))| \leq \bar{V}, \quad \text{and} \quad |u(T_1)|_{L^{2\sigma+2}}^{2\sigma+2} \geq \bar{H}_2.$$

The solution of (1.4) with $\psi(0) = \psi_0$ is given by $\psi(t) = z(t) + v(z, \psi_0, t)$ with

$$z(t) = \int_0^t S(t-s)dW(s)$$

almost surely on $[0, \tau^*(\psi_0))$. Since Φ is Hilbert-Schmidt from $L^2(\mathbb{R}^n)$ into S_n , z is almost surely in $C([0, T_1]; S_n)$ (see [3], Theorem 6.10). Moreover, it is shown in [4] that z is almost surely in $L^r(0, T; W^{1, 2\sigma+2})$. Since the null space of Φ^* is equal to $\{0\}$, Φ has dense range in S^n and in $W^{1, 2\sigma+2}$, we deduce that z is non-degenerate (note that z is a Gaussian process with values in S_n) and $\mathbb{P}(z \in B) > 0$; therefore the probability that $\tau^*(\psi_0) \geq T_1$ and $\psi(T_1)$ satisfies (3.6) is positive. We now set

$$\Omega_1 = \{\omega \in \Omega, \tau^*(u_0) \geq T_1 \text{ and } \psi(T_1) \text{ satisfies (3.6)}\}$$

and note that $\psi(T_1), \Omega_1, \bar{T}, \bar{V}, \bar{G}, \bar{H}_1, \bar{H}_2$ satisfy the condition of Proposition 3.1, or Proposition 3.2. The result follows.

References

- [1] O. Bang, P.L. Christiansen, F. If, K.O. Rasmussen, Y.B. Gaididei, *Temperature effects in a nonlinear model of monolayer Scheibe aggregates*, Phys. Rev. E **49**, 4627–4636 (1994).
- [2] O. Bang, P.L. Christiansen, F. If, K.O. Rasmussen, Y.B. Gaididei, *White Noise in the Two-dimensional Nonlinear Schrödinger Equation*, Appl. Anal. **57**, 3–15 (1995).
- [3] G. Da Prato, J. Zabczyk, *Stochastic equations in infinite dimensions*, Encyclopedia of Mathematics and its Applications, Cambridge University Press (1992).
- [4] A. de Bouard, A. Debussche, *The stochastic nonlinear Schrödinger equation in H^1* , to appear in Stoch. Anal. and Appl.
- [5] A. de Bouard, A. Debussche, *On the effect of a noise on the solutions of supercritical nonlinear Schrödinger equation*, to appear in Probab. Theory Relat. Fields.
- [6] A. Debussche, L. Di Menza *Numerical simulation of focusing stochastic nonlinear Schrödinger equations*. Preprint.
- [7] G.E. Falkovich, I. Kolokolov, V. Lebedev, S.K. Turitsyn, *Statistics of soliton-bearing systems with additive noise*, Phys. Rev. E **63** (2001).
- [8] R.T. Glassey, *On the blowing-up of solutions to the Cauchy problem for the nonlinear Schrödinger equation*, J. Math. Phys. **18** 1794–1797 (1977).
- [9] C.E. Kenig, G. Ponce, L. Vega, *On the ill-posedness of some canonical dispersive equations*, Duke Math. J. **106**, 617–633 (2001).
- [10] C. Sulem, P.L. Sulem, *The Nonlinear Schrödinger Equation, Self-Focusing and Wave Collapse*, Appl. Math. Sciences, Springer Verlag: New York, 1999.
- [11] S.N. Vlasov, V.A. Petrishchev, V.I. Talanov, *Average description of wave beams in linear and nonlinear media (The method of moments)*, Radiophys. Quantum Electron. **14** 1062–1070 (1974).
- [12] V.E. Zakharov, *Collapse of Langmuir waves*, Sov. Phys. JETP **35**, 908–914 (1972).

CNRS ET UNIVERSITÉ PARIS-SUD., UMR 8628, BÂT. 425, UNIVERSITÉ DE PARIS-SUD., 91405 ORSAY CEDEX, FRANCE

E-mail address: Anne.deBouard@math.u-psud.fr

ENS DE CACHAN, ANTENNE DE BRETAGNE,, CAMPUS DE KER LANN, AV. R. SCHUMAN,, 35170 BRUZ, FRANCE

E-mail address: Arnaud.Debussche@bretagne.ens-cachan.fr

Method of symmetry transforms for ideal MHD equilibrium equations

Oleg I. Bogoyavlenskij

ABSTRACT. Method of symmetry transforms is developed for constructing ideal magnetohydrodynamics equilibria. The method is applicable to the subsonic divergence-free plasma flows and to the ideal gas plasma. The symmetry transforms are presented in explicit algebraic form; they depend on all three spatial variables x, y, z and form infinite-dimensional abelian groups G_m . Applying the symmetry transforms gives both the exact MHD equilibria with non-collinear vector fields \mathbf{B} and \mathbf{V} and the global non-symmetric MHD equilibria that model astrophysical jets.

1. Introduction

As known, the inverse scattering transform method [1] was discovered as a result of the studies of the plasma physics problems connected with the controlled thermonuclear fusion, see the papers [2 - 4]. In this paper, we study the main equations that are used to model controlled thermonuclear fusion and numerous astrophysical phenomena - the nonlinear system of magnetohydrodynamics equilibrium equations. We present a new method of symmetry transforms for constructing ideal MHD equilibria. The method uses the recently discovered [5] continuous symmetries for these equations. By applying the symmetries to any known equilibrium solution we obtain infinite families of new MHD equilibria.

The method of symmetry transforms is different from the method of Backlund transforms for the soliton equations such as the Korteweg - de Vries equation [1], the Kadomtsev - Petviashvili equation [6], the Sine - Gordon equation [7], etc. The method of Backlund transforms is based on resolution of certain auxiliary differential equations which usually can not be solved explicitly. All soliton equations depend only on a part of spatial variables (one for KdV and SG and two for KP).

The method of symmetry transforms has the following features that distinguish it from that of Backlund transforms:

- i)* The method of symmetry transforms produces new solutions in explicit algebraic form,
- ii)* The symmetry transforms depend on all three spatial variables $\mathbf{x} = x, y, z$,
- iii)* The generic symmetry transforms break the geometrical symmetries of the field-aligned MHD equilibria,

- iv)* The symmetries form infinite-dimensional abelian groups G_m that depend on the topology of the given MHD equilibria,
v) The groups of symmetries G_m have an additional algebraic structure of modules over the associative algebras of functions.

The existence of some Backlund transform usually implies integrability of the corresponding soliton equation. The existence of the symmetry transforms allows us to reduce the order of the ideal MHD equilibrium equations. However, the resulting equations are not integrable because they contain as special cases the non-integrable Grad-Shafranov equation [8, 9] and the non-integrable JFKO equation [10].

The symmetry transforms break the axial symmetry of the magnetic analog of Hill's spherical vortex [11] and produce a continuous family of non-symmetric MHD equilibria with closed magnetic field lines. These exact solutions depend on two arbitrary functions and have no geometrical symmetries. This result shows that Grad's 1985 conjecture [12], p.35, that states: "The proper formulation of the nonexistence statement is that, other than the stated symmetric exceptions, there are no *families* of solutions depending smoothly on a parameter" is not applicable to the MHD equilibria. By applying the method to the exact solutions of the papers [13, 14] we get the MHD equilibria that model ball lightning with dynamics of plasma inside the fireball.

Applying the symmetry transforms to the exact axially and helically symmetric plasma equilibria constructed in [5, 15, 16] we obtain infinite families of physically significant MHD equilibria. The constructed exact non-symmetric solutions satisfy the necessary physical conditions and model astrophysical jets outside of their accretion disks, for example the jet in the elliptical galaxy Messier 87 [17 - 19]. In the papers [20, 21], the non-symmetric plasma equilibria were presented that blow up as $x^2 + y^2 \rightarrow \infty$. In Section 11, we derive non-symmetric and bounded exact solutions that describe the non-steady dynamics of viscous plasma.

2. The symmetry transforms

The system of ideal magnetohydrodynamics equilibrium equations has the form

$$(2.1) \quad \rho(\mathbf{V} \cdot \text{grad})\mathbf{V} + \frac{1}{\mu}\mathbf{B} \times \text{curl} \mathbf{B} = -\text{grad} P,$$

$$(2.2) \quad \text{div}(\rho\mathbf{V}) = 0, \quad \text{div} \mathbf{B} = 0, \quad \text{curl}(\mathbf{V} \times \mathbf{B}) = 0,$$

where \mathbf{B} is the magnetic vector field, μ is the constant magnetic permeability, \mathbf{V} is the plasma velocity, $\rho = \rho(\mathbf{x})$ is its density, $\mathbf{x} = (x, y, z)$, and P is the pressure.

We consider both the incompressible and compressible plasma flows. The condition of incompressibility $\text{div} \mathbf{V} = 0$ is widely used in the MHD literature [22 - 26]. For example, it is applicable with a high accuracy to the subsonic plasma flows with Mach number $M \ll 1$, $M^2 = V^2/(\gamma P/\rho)$. Another application of the condition $\text{div} \mathbf{V} = 0$ is to the dynamics of a perfectly conducting incompressible fluid with a variable mass density $\rho(\mathbf{x})$. Then the continuity equation $\text{div}(\rho\mathbf{V}) = 0$ implies $(\mathbf{V} \cdot \text{grad} \rho(\mathbf{x})) = 0$. Hence the plasma density $\rho(\mathbf{x})$ is constant on the plasma streamlines.

For the compressible plasma flows, we suppose that plasma satisfies the ideal gas equation of state

$$(2.3) \quad P = \rho^\gamma \exp(S/c_v), \quad \mathbf{V} \cdot \text{grad} S = 0,$$

where $S(\mathbf{x})$ is the density of entropy and $\gamma > 1$ is the adiabatic exponent. We consider also the more general equations of state $P = \rho^\gamma f(S)$.

The third equation (2.2) implies the existence of either magnetic surfaces or a magnetic foliation. Indeed, equation $\text{curl}(\mathbf{V} \times \mathbf{B}) = 0$ yields

$$(2.4) \quad \mathbf{V} \times \mathbf{B} = \text{grad } \psi(\mathbf{x}), \quad \psi(\mathbf{x}) = \int_{x_0}^x (\mathbf{V} \times \mathbf{B}) \cdot d\mathbf{s}.$$

Hence we obtain that in any simply connected domain $E \subset \mathbb{R}^3$ the surfaces $\psi(\mathbf{x}) = \text{const}$ are magnetic surfaces because

$$(2.5) \quad \mathbf{B} \cdot \text{grad } \psi = 0, \quad \mathbf{V} \cdot \text{grad } \psi = 0.$$

In a non-simply connected domain D , the function $\psi(\mathbf{x})$ (2.4) is multivalued in general. However, the differential form $d\psi$ is uniquely defined. Indeed, we have

$$(2.6) \quad d\psi(\mathbf{Y}(\mathbf{x})) = (\mathbf{V}(\mathbf{x}) \times \mathbf{B}(\mathbf{x}) \cdot \mathbf{Y}(\mathbf{x}))$$

for any tangent vector $\mathbf{Y}(\mathbf{x}) \in \mathbb{R}^3$. In view of the equation $\text{curl}(\mathbf{V} \times \mathbf{B}) = 0$, the differential form $d\psi$ (2.6) is closed. Hence the equation $d\psi(\mathbf{Y}(\mathbf{x})) = 0$ defines an integrable foliation in \mathbb{R}^3 that is generated by the vector fields \mathbf{B} and \mathbf{V} . For any solution to equations (2.1) - (2.2), the commutator of the vector fields \mathbf{B} and \mathbf{V} is proportional to \mathbf{B} :

$$(2.7) \quad [\mathbf{B}, \mathbf{V}] = -\rho^{-1}(\mathbf{V} \cdot \text{grad } \rho)\mathbf{B}.$$

Formula (2.7) is proven in Section 10.

For the case of toroidal domains D (which are the most important for the tokamak applications) the function $\psi(\mathbf{x})$ is defined up to a constant nI , where n is an integer and I is the integral (2.4) over the shortest non-contractable loop of the torus.

Any given smooth non-field-aligned MHD equilibrium $\mathbf{B}, \mathbf{V}, \rho, P$ in \mathbb{R}^3 (\mathbf{B} and \mathbf{V} are non-collinear) defines a distribution of magnetic surfaces $\psi(\mathbf{x}) = \text{const}$ in \mathbb{R}^3 [24]. Let E_m be the set of all divergence-free equilibria that have the same magnetic surfaces as the given one. We introduce the transforms $S: E_m \rightarrow E_m$ that depend on the arbitrary functions $a(\mathbf{x}) \neq 0, b(\mathbf{x}), c(\mathbf{x})$ that are constant on the magnetic surfaces $\psi(\mathbf{x}) = \text{const}$. The $a(\mathbf{x}), b(\mathbf{x}), c(\mathbf{x})$ and $\rho(\mathbf{x})$ are either arbitrary functions of $\psi(\mathbf{x})$ or, more generally, satisfy arbitrary equations $F(a(\mathbf{x}), \psi(\mathbf{x})) = 0, G(b(\mathbf{x}), \psi(\mathbf{x})) = 0, H(c(\mathbf{x}), \psi(\mathbf{x})) = 0, R(\rho(\mathbf{x}), \psi(\mathbf{x})) = 0$. For the MHD equilibria in a toroidal domain D (a tokamak), functions $a = a(\psi), b = b(\psi), c = c(\psi)$ and $\rho(\psi)$ are periodic with period I defined above.

THEOREM 1. *The magnetohydrodynamics equilibrium equations (2.1) - (2.2) for the divergence-free flows are invariant under the following transforms:*

$$(2.8) \quad \mathbf{B}_1 = b(\mathbf{x})\mathbf{B} + c(\mathbf{x})\sqrt{\mu\rho(\mathbf{x})}\mathbf{V}, \quad \mathbf{V}_1 = \frac{c(\mathbf{x})}{a(\mathbf{x})\sqrt{\mu\rho(\mathbf{x})}}\mathbf{B} + \frac{b(\mathbf{x})}{a(\mathbf{x})}\mathbf{V},$$

$$\rho_1(\mathbf{x}) = a^2(\mathbf{x})\rho(\mathbf{x}), \quad P_1 = CP + (C\mathbf{B}^2 - \mathbf{B}_1^2)/(2\mu), \quad b^2(\mathbf{x}) - c^2(\mathbf{x}) = C,$$

where $C \neq 0$ is an arbitrary constant.

Proof. We prove that transforms (2.8) define new solutions to the equilibrium equations (2.1) - (2.2). Let $h(\mathbf{x})$ be any function that is constant on the magnetic surfaces, for example $a(\mathbf{x}), b(\mathbf{x}), c(\mathbf{x})$, or $\rho(\mathbf{x})$. Hence we have

$$(2.9) \quad \mathbf{V} \cdot \text{grad } h(\mathbf{x}) = 0, \quad \mathbf{B} \cdot \text{grad } h(\mathbf{x}) = 0.$$

Applying the classical identity $\mathbf{B} \times \text{curl } \mathbf{B} = -(\mathbf{B} \cdot \text{grad})\mathbf{B} + \text{grad}(\mathbf{B}^2/2)$, we present the equation (2.1) in the form

$$(2.10) \quad \rho(\mathbf{V} \cdot \text{grad})\mathbf{V} - (\mathbf{B} \cdot \text{grad})\mathbf{B}/\mu = -\text{grad}(P + \mathbf{B}^2/(2\mu)).$$

Substituting formulae (2.8) and (2.9), we derive

$$\begin{aligned} &\rho_1(\mathbf{V}_1 \cdot \text{grad})\mathbf{V}_1 - (\mathbf{B}_1 \cdot \text{grad})\mathbf{B}_1/\mu + \text{grad}(P_1 + \mathbf{B}_1^2/(2\mu)) = \\ &(b^2 - c^2) (\rho(\mathbf{V} \cdot \text{grad})\mathbf{V} - (\mathbf{B} \cdot \text{grad})\mathbf{B}/\mu + \text{grad}(P + \mathbf{B}^2/(2\mu))) = 0. \end{aligned}$$

Thus the functions $\rho_1, \mathbf{B}_1, \mathbf{V}_1, P_1$ satisfy equation (2.10) and therefore equation (2.1). \mathbf{x}

Equations $\text{div}(\rho_1\mathbf{V}_1) = 0$ and $\text{div } \mathbf{B}_1 = 0$ easily follow from equations (2.9) and $\text{div}(\rho\mathbf{V}) = 0, \text{div } \mathbf{B} = 0$. Substituting (2.8), we obtain

$$\text{curl}(\mathbf{V}_1 \times \mathbf{B}_1) = \text{curl} \left(\frac{C}{a(\mathbf{x})} \mathbf{V} \times \mathbf{B} \right) = \text{grad} \frac{C}{a(\mathbf{x})} \times (\mathbf{V} \times \mathbf{B}) + \frac{C}{a(\mathbf{x})} \text{curl}(\mathbf{V} \times \mathbf{B}).$$

Applying equations (2.9), we find that $\text{grad}(C/a(\mathbf{x}))$ is collinear with the vector field $\mathbf{V} \times \mathbf{B}$. Hence $\text{grad}(C/a(\mathbf{x})) \times (\mathbf{V} \times \mathbf{B}) = 0$. This fact and equation $\text{curl}(\mathbf{V} \times \mathbf{B}) = 0$ proves that $\text{curl}(\mathbf{V}_1 \times \mathbf{B}_1) = 0$. Hence formulae (2.8) define a new solution to equations (2.1) - (2.2). \square

Remark 1. The new vector fields \mathbf{B}_1 and \mathbf{V}_1 (2.8) are linearly dependent on the original \mathbf{B} and \mathbf{V} . Hence the new MHD equilibrium $\mathbf{B}_1, \mathbf{V}_1, \rho_1, P_1$ has the same magnetic surfaces as the original one $\mathbf{B}, \mathbf{V}, \rho, P$. This property shows that the transforms (2.8) map the set E_m into itself. For $C \neq 0$, the transforms (2.8) are invertible. We shall refer to the transforms S (2.8) as the symmetries of the divergence-free MHD equilibrium equations (2.1) - (2.2). In Section 4, we apply the symmetry transforms (2.8) to Grad’s “transverse” flows [27] and obtain the more complex non-field-aligned MHD equilibria.

The symmetries (2.8) have the following physical meaning. The difference between the plasma kinetic and magnetic energies is changed by a scalar multiplication. Indeed, the transform (2.8) implies

$$(2.11) \quad \frac{1}{2}\rho_1\mathbf{V}_1^2 - \frac{1}{2\mu}\mathbf{B}_1^2 = C \left(\frac{1}{2}\rho\mathbf{V}^2 - \frac{1}{2\mu}\mathbf{B}^2 \right),$$

where $C = b^2(\mathbf{x}) - c^2(\mathbf{x}) = \text{const}$. Another consequence of transform (2.8) is the equation

$$\mathbf{V}_1 \times \mathbf{B}_1 = \frac{C}{a(\mathbf{x})} \mathbf{V} \times \mathbf{B}.$$

Hence we obtain $\sqrt{\rho_1}\mathbf{V}_1 \times \mathbf{B}_1 = C\sqrt{\rho}\mathbf{V} \times \mathbf{B}$, or $\sqrt{\rho_1}\mathbf{E}_1 = C\sqrt{\rho}\mathbf{E}$. Here $\mathbf{E} = -\mathbf{V} \times \mathbf{B}/c_0$ is the electric field for the plasma with a perfect electric conductivity; c_0 is the speed of light. Hence the symmetries (2.8) have the vector field invariant

$$(2.12) \quad \frac{\sqrt{\rho_1}\mathbf{V}_1 \times \mathbf{B}_1}{\rho_1\mathbf{V}_1^2 - \mathbf{B}_1^2/\mu} = \frac{\sqrt{\rho}\mathbf{V} \times \mathbf{B}}{\rho\mathbf{V}^2 - \mathbf{B}^2/\mu}.$$

Remark 2. In his 1962 paper [28], Newcomb proved that the (time-dependent) MHD equations (2.1) for $\text{div } \mathbf{V} = 0$ follow from the variational principle

$$(2.13) \quad \delta \int_{t_1}^{t_2} dt \int L(\mathbf{B}, \mathbf{V}, \rho) d^3x = 0, \quad L(\mathbf{B}, \mathbf{V}, \rho) = \frac{1}{2}\rho\mathbf{V}^2 - \frac{1}{2\mu}\mathbf{B}^2,$$

provided that the (time-dependent) equations (2.2) are satisfied. The symmetry transformations (2.8) preserve the equations (2.2) because the functions $a(\mathbf{x})$, $b(\mathbf{x})$ and $c(\mathbf{x})$ are constant on magnetic surfaces. Equation (2.11) implies the following relation between the Lagrangians

$$(2.14) \quad L(\mathbf{B}_1, \mathbf{V}_1, \rho_1) = CL(\mathbf{B}, \mathbf{V}, \rho).$$

Equation (2.14) means that the symmetry transforms (2.8) for $\text{div } \mathbf{V} = 0$ preserve the Lagrangian of the Newcomb variational principle (2.13) up to a constant factor. Hence any extremum of the principle (2.13) is transformed into a new extremum. Thus we obtain the second proof of Theorem 1. The first proof above is direct and independent of Newcomb’s 1962 variational principle [28].

3. The infinite-dimensional abelian groups of symmetries

For $C \neq 0$, the transform (2.8) has the following inverse transform:

$$(3.1) \quad C\mathbf{B} = b(\mathbf{x})\mathbf{B}_1 - c(\mathbf{x})\sqrt{\mu\rho_1(\mathbf{x})}\mathbf{V}_1, \quad C\mathbf{V} = \frac{-c(\mathbf{x})}{a_1(\mathbf{x})\sqrt{\mu\rho_1(\mathbf{x})}}\mathbf{B}_1 + \frac{b(\mathbf{x})}{a_1(\mathbf{x})}\mathbf{V}_1,$$

where $a_1(\mathbf{x}) = 1/a(\mathbf{x})$. For $C = 0$, or $b(\mathbf{x}) = \pm c(\mathbf{x})$, the transform (2.8) is not invertible and its range consists of the field-aligned solutions

$$(3.2) \quad \mathbf{B}_1 = \pm\sqrt{\mu\rho_1(\mathbf{x})}\mathbf{V}_1, \quad P_1 + \frac{1}{2\mu}\mathbf{B}_1^2 = C_0 = \text{const},$$

that are the known Chandrasekhar equipartition equilibria [22].

We consider the set G_m of all transforms (2.8) with $C \neq 0$ for which the smooth functions $a(\mathbf{x}) \neq 0$, $b(\mathbf{x})$ and $c(\mathbf{x})$ are constant on the magnetic surfaces for a given MHD equilibrium. It is evident that the set G_m is infinite-dimensional. These transforms (2.8) are defined on the domain E_m that consists of all divergence-free MHD equilibria that have the same magnetic surfaces as the given one. The invertibility of the transforms (2.8) for $C \neq 0$ and Remark 1 above prove that the range of these transforms is the same as their domain, E_m . Hence the composition of the transforms is well defined. Let us show that the composition assigns on the set G_m the structure of an abelian group. For the transforms (2.8), we write

$$(3.3) \quad a(\mathbf{x}) = \tau \exp(\alpha(\mathbf{x})), \quad b^2(\mathbf{x}) - c^2(\mathbf{x}) = C = \sigma k^2,$$

where $\alpha(\mathbf{x})$ is an arbitrary smooth function that is constant on the magnetic surfaces and $\tau = \pm 1$, $\sigma = \pm 1$, $k > 0$. The second equation (3.3) is resolved in the form: $\sigma = 1$: $b(\mathbf{x}) = \eta k \text{ch } \beta(\mathbf{x})$, $c(\mathbf{x}) = \eta k \text{sh } \beta(\mathbf{x})$; $\sigma = -1$: $b(\mathbf{x}) = \eta k \text{sh } \beta(\mathbf{x})$, $c(\mathbf{x}) = \eta k \text{ch } \beta(\mathbf{x})$, where $\eta = \pm 1$ and $\beta(\mathbf{x})$ is an arbitrary smooth function that is constant on the magnetic surfaces. Hence all transforms (2.8) are represented by the sextuples $(\alpha(\mathbf{x}), \beta(\mathbf{x}), k, \tau, \sigma, \eta)$.

The composition of the transforms (2.8) is equivalent to the 3×3 matrix multiplication

$$\begin{pmatrix} a_2 & 0 & 0 \\ 0 & b_2 & c_2\sqrt{\mu\rho_1} \\ 0 & \frac{c_2}{\sqrt{\mu\rho_2}} & b_2\sqrt{\frac{\rho_1}{\rho_2}} \end{pmatrix} \times \begin{pmatrix} a_1 & 0 & 0 \\ 0 & b_1 & c_1\sqrt{\mu\rho} \\ 0 & \frac{c_1}{\sqrt{\mu\rho_1}} & b_1\sqrt{\frac{\rho}{\rho_1}} \end{pmatrix} = \begin{pmatrix} a_1a_2 & 0 & 0 \\ 0 & f & g\sqrt{\mu\rho} \\ 0 & \frac{g}{\sqrt{\mu\rho_2}} & f\sqrt{\frac{\rho}{\rho_2}} \end{pmatrix}.$$

Here $f = b_1 b_2 + c_1 c_2$, $g = c_1 b_2 + b_1 c_2$ and $f^2 - g^2 = (b_1^2 - c_1^2)(b_2^2 - c_2^2)$. In view of the known identities for the hyperbolic functions $\operatorname{ch} t$ and $\operatorname{sh} t$, these formulae are equivalent to the following multiplication of the sextuples

$$(3.4) \quad (\alpha_1(\mathbf{x}), \beta_1(\mathbf{x}), k_1, \tau_1, \sigma_1, \eta_1) \cdot (\alpha_2(\mathbf{x}), \beta_2(\mathbf{x}), k_2, \tau_2, \sigma_2, \eta_2) = \\ (\alpha_1(\mathbf{x}) + \alpha_2(\mathbf{x}), \beta_1(\mathbf{x}) + \beta_2(\mathbf{x}), k_1 k_2, \tau_1 \tau_2, \sigma_1 \sigma_2, \eta_1 \eta_2).$$

The sextuple $(0, 0, 1, 1, 1, 1)$ is the unit for the multiplication (3.4). The inverse transform (3.1) corresponds to the sextuple

$$(3.5) \quad (\alpha(\mathbf{x}), \beta(\mathbf{x}), k, \tau, \sigma, \eta)^{-1} = (-\alpha(\mathbf{x}), -\beta(\mathbf{x}), k^{-1}, \tau, \sigma, \eta).$$

Formulae (3.4) - (3.5) mean that the composition of transforms (2.8) defines on the set G_m the structure of an abelian group. Formula (3.4) proves that the group G_m is the direct sum

$$(3.6) \quad G_m = A_m \oplus A_m \oplus R^+ \oplus Z_2 \oplus Z_2 \oplus Z_2.$$

Here R^+ is the multiplicative group of positive numbers $k > 0$. The A_m is the additive abelian group of smooth functions on \mathbb{R}^3 that are constant on the magnetic surfaces for the given MHD equilibrium. The group A_m is also a linear space and an associative algebra with respect to the multiplication of functions.

The group G_m (3.6) depends on the topology of the distribution of magnetic surfaces because the functions $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ (3.4) are constant on them. Hence there are infinitely many different infinite-dimensional groups G_m of symmetries of the MHD equilibrium equations.

Remark 3. An additional algebraic structure. The group G_m has an additional structure that does not exist for the groups of symmetries of the soliton equations [1, 6, 7]. This group is a module over the associative algebra $A_m \oplus A_m$ with the multiplication defined by the multiplication of functions in A_m :

$$(3.7) \quad (\gamma(\mathbf{x}), \zeta(\mathbf{x})) \cdot (\alpha(\mathbf{x}), \beta(\mathbf{x}), k, \tau, \sigma, \eta) = (\gamma(\mathbf{x})\alpha(\mathbf{x}), \zeta(\mathbf{x})\beta(\mathbf{x}), k, \tau, \sigma, \eta).$$

This operation results in the transform: $a(\mathbf{x}) \longrightarrow \tau |a(\mathbf{x})|^{\gamma(\mathbf{x})}$, $b(\mathbf{x}) = \eta k \operatorname{ch} \beta(\mathbf{x}) \longrightarrow \eta k \operatorname{ch}(\zeta(\mathbf{x})\beta(\mathbf{x}))$, $c(\mathbf{x}) = \eta k \operatorname{sh} \beta(\mathbf{x}) \longrightarrow \eta k \operatorname{sh}(\zeta(\mathbf{x})\beta(\mathbf{x}))$.

Remark 4. Subgroups of the group G_m . The group G_m (3.6) has eight periodic elements which form the subgroup $\Gamma_p = Z_2 \oplus Z_2 \oplus Z_2 \subset G_m$ and are defined by the sextuples $(0, 0, 1, \pm 1, \pm 1, \pm 1)$. All other elements are non-periodic. The subgroup Γ_p elements represent the eight components $G_{m,j}$, $j = 1, \dots, 8$, of the group G_m . The scale and reflection symmetries form the subgroup $\Gamma_{sr} = R \oplus R^+ \oplus Z_2 \oplus Z_2 \oplus Z_2 \subset G_m$ that consists of the sextuples $(r, 0, k, \tau, \sigma, \eta)$, $k > 0$. The subgroup Γ_{sr} does not depend on the considered MHD equilibria.

Thus we have shown that the divergence-free MHD equilibrium equations (2.1) - (2.2) have the infinite-dimensional abelian groups of symmetries G_m (3.6) and that these groups have the additional structure (3.7) of modules over the associative algebras of functions.

4. Applications of the symmetry transforms

I. In this Section, we apply the symmetry transforms (2.8) to obtain the new MHD equilibria with non-collinear vector fields \mathbf{B} and \mathbf{V} . We consider in the cylindrical coordinates (r, z, ϕ) the differential rotation of a perfectly conducting ideal gas plasma around the axis z . Let $\omega(r)$ be its angular velocity and $H(r)$ be the vertical

magnetic field. Equations (2.1) - (2.2) have the exact solutions [27] that depend on the three arbitrary functions $\omega(r), H(r), \rho(r) \geq 0$:

$$(4.1) \quad \mathbf{B} = H(r)\hat{e}_z, \quad \mathbf{V} = \omega(r)(-y\hat{e}_x + x\hat{e}_y), \quad P(r) = F(r) - H^2(r)/(2\mu),$$

Here $F(r) = \int_0^r t\rho(t)\omega^2(t)dt$. The vector fields \mathbf{B} and \mathbf{V} (4.1) are orthogonal. Solutions (4.1) satisfy the ideal gas equation of state (2.3) where density of entropy $S(r)$ is defined from equations (2.3) and (4.1) for the arbitrary gas density $\rho(r)$.

Applying symmetry transform (2.8) to solutions (4.1), we obtain the new exact MHD equilibria

$$(4.2) \quad \rho_1(r) = a^2(r)\rho(r), \quad \mathbf{B}_1 = c\sqrt{\mu\rho}\mathbf{V} + bH\hat{e}_z, \quad \mathbf{V}_1 = \frac{b}{a}\mathbf{V} + \frac{c}{a\sqrt{\mu\rho}}H\hat{e}_z,$$

where $b^2(r) - c^2(r) = C = \text{const}$ and $P_1 = CF - b^2H^2/(2\mu) - \rho(cr\omega)^2/2$. Exact solutions (4.2) also satisfy the ideal gas equation of state of the form (2.3) and depend on four arbitrary functions $a(r), b(r), \omega(r), H(r)$. Their magnetic field lines and plasma streamlines are helices which rotate on the cylindrical magnetic surfaces $x^2 + y^2 = r^2 = \text{const}$.

Solutions (4.2) describe, for example, the helical dynamics of plasma inside a cylinder $0 \leq r \leq R$, provided that $c(R) = 0, \rho(R) = 0, \omega(R) = 0, H^2(R) = 2\mu F(R), b^2H^2(r) \leq 2\mu CF(r) - \mu\rho(cr\omega)^2, c(r)/\rho(r) < \text{const}$. Then at the wall $r = R$, the plasma pressure $P_1(R) = 0$, the density $\rho_1(R) = 0$ and velocity $\mathbf{V}_1(R) = 0$. The exact MHD equilibria (4.2) belong to the generic class of solutions with non-collinear vector fields \mathbf{B} and \mathbf{V} .

II. Let us consider the 2-dimensional steady ideal fluid dynamics

$$(4.3) \quad (\mathbf{V} \cdot \text{grad})\mathbf{V} = -\text{grad } P_0, \quad \text{div } \mathbf{V} = 0$$

with constant density $\rho = 1$. As is known, any solution to equations (4.3) has the form

$$\mathbf{V} = -\psi_y\hat{e}_x + \psi_x\hat{e}_y, \quad \psi_{xx} + \psi_{yy} = \frac{dF(\psi)}{d\psi}, \quad P_0 = F(\psi) - \frac{1}{2}(\psi_x^2 + \psi_y^2)$$

with an arbitrary smooth function $F(\psi)$. Let $f(x, y)$ be any first integral of the flow: $(\mathbf{V} \cdot \text{grad } f) = f_y\psi_x - f_x\psi_y = 0$. Then the orthogonal vector fields $\mathbf{B} = f\hat{e}_z$ and \mathbf{V} satisfy the equilibrium equations (2.1) - (2.2) with $P = P_0 - f^2/(2\mu)$ and $\rho = 1$ [27]. Applying the intrinsic symmetries (2.8) to these solutions, we derive the new MHD equilibria

$$(4.4) \quad \rho_1 = a^2, \quad \mathbf{B}_1 = (-c\sqrt{\mu}\psi_y, c\sqrt{\mu}\psi_x, bf), \quad \mathbf{V}_1 = \left(-\frac{b}{a}\psi_y, \frac{b}{a}\psi_x, \frac{c}{a\sqrt{\mu}}f\right),$$

where $b^2 - c^2 = C = \text{const}, P_1 = CP_0 - \mathbf{B}_1^2/(2\mu)$. Here $a(x, y), b(x, y)$ and $f(x, y)$ are any first integrals of the flow $\mathbf{V}: a_y\psi_x - a_x\psi_y = 0$, etc. Thus the symmetry transforms (2.8) produce from any 2-dimensional steady flow (4.3) with $\rho = 1$ the 3-dimensional non-field-aligned MHD equilibria (4.4) which depend on three functions a, b, f and have a variable plasma density $\rho_1 = a^2(x, y)$.

5. The symmetry transforms for the field-aligned solutions

For the field-aligned equilibria, the vector fields \mathbf{B} and \mathbf{V} are collinear and satisfy the equation

$$(5.1) \quad \mathbf{V} = \frac{\lambda(\mathbf{x})}{\sqrt{\mu\rho(\mathbf{x})}}\mathbf{B},$$

where $\lambda(\mathbf{x})$ is some smooth function on \mathbb{R}^3 . The first two equations (2.2) and equation $\operatorname{div} \mathbf{V} = 0$ imply $\mathbf{B} \cdot \operatorname{grad} \rho = 0$ and $\mathbf{B} \cdot \operatorname{grad} \lambda = 0$. The flows (5.1) are sub-Alfvénic for $|\lambda(\mathbf{x})| < 1$ and super-Alfvénic for $|\lambda(\mathbf{x})| > 1$. The Chandrasekhar equipartition solutions (3.2) [22] correspond to $|\lambda(\mathbf{x})| \equiv 1$, $\rho(\mathbf{x}) \equiv \text{const}$.

THEOREM 2. *For the field-aligned solutions (5.1), the magnetohydrodynamics equilibrium equations (2.1) - (2.2) for $\operatorname{div} \mathbf{V} = 0$ possess the following intrinsic symmetries:*

$$(5.2) \quad \rho_1(\mathbf{x}) = a_1^2(\mathbf{x})\rho(\mathbf{x}), \quad \mathbf{B}_1 = b_1(\mathbf{x})\mathbf{B}, \quad \mathbf{V}_1 = \frac{c_1(\mathbf{x})}{a_1(\mathbf{x})\sqrt{\mu\rho(\mathbf{x})}}\mathbf{B},$$

$$P_1 + \mathbf{B}_1^2/(2\mu) = C_1(P + \mathbf{B}^2/(2\mu)), \quad \frac{b_1^2(\mathbf{x}) - c_1^2(\mathbf{x})}{1 - \lambda^2(\mathbf{x})} = C_1 = \text{const},$$

where smooth functions $a_1(\mathbf{x}) \neq 0$, $b_1(\mathbf{x})$, $c_1(\mathbf{x})$, $\lambda(\mathbf{x})$, $\rho(\mathbf{x})$ are constant on the magnetic field lines.

Proof. It is evident that symmetries (5.1) preserve equations (2.2). For the field-aligned solutions (5.1), equation (2.1), (2.10) takes the form

$$(5.3) \quad \mu^{-1}(\lambda^2(\mathbf{x}) - 1)(\mathbf{B} \cdot \operatorname{grad})\mathbf{B} + \operatorname{grad}(P + \mathbf{B}^2/(2\mu)) = 0.$$

Applying transformation (5.2) and using formula (5.3), we obtain

$$(5.4) \quad \rho_1(\mathbf{x})(\mathbf{V}_1 \cdot \operatorname{grad})\mathbf{V}_1 - (\mathbf{B}_1 \cdot \operatorname{grad})\mathbf{B}_1/\mu + \operatorname{grad}(P_1 + \mathbf{B}_1^2/(2\mu)) =$$

$$\mu^{-1}(c_1^2(\mathbf{x}) - b_1^2(\mathbf{x}))(\mathbf{B} \cdot \operatorname{grad})\mathbf{B} + C_1 \operatorname{grad}(P + \mathbf{B}^2/(2\mu)) = 0.$$

Hence functions \mathbf{B}_1 , \mathbf{V}_1 , ρ_1 and P_1 (5.2) satisfy equation (2.1) and therefore define the new exact MHD equilibria. \square

Remark 5. The symmetries (5.2) for $C_1 \neq 0$ form the infinite-dimensional groups G_m that have the same structure as (3.6) where A_m is the algebra of smooth functions on \mathbb{R}^3 that are constant on the magnetic field lines (5.1).

The functions $a_1(\mathbf{x})$, $b_1(\mathbf{x})$, $c_1(\mathbf{x})$, $\lambda(\mathbf{x})$, $\rho(\mathbf{x})$ that are constant on the magnetic field lines are first integrals of the magnetic dynamical system

$$(5.5) \quad \dot{x} = B_x(x, y, z), \quad \dot{y} = B_y(x, y, z), \quad \dot{z} = B_z(x, y, z),$$

where B_x , B_y , B_z are components of the magnetic field \mathbf{B} . The following three possibilities can be realized in different domains $E \subset \mathbb{R}^3$:

1) The magnetic field lines (5.5) are dense on some closed magnetic surfaces - then these surfaces topologically are tori \mathbb{T}^2 [2] and first integrals $a_1(\mathbf{x})$, $b_1(\mathbf{x})$, $c_1(\mathbf{x})$, $\lambda(\mathbf{x})$, $\rho_1(\mathbf{x})$ are constant on \mathbb{T}^2 . Hence they are actually functions of one transversal variable.

2) All magnetic field lines go to infinity or are closed curves - then the first integrals $a_1(\mathbf{x})$, $b_1(\mathbf{x})$, $c_1(\mathbf{x})$, $\lambda(\mathbf{x})$, $\rho_1(\mathbf{x})$ are arbitrary smooth functions of the two transversal variables and magnetic surfaces are not uniquely defined.

3) The magnetic field lines are dense in some 3-dimensional domains $E \subset \mathbb{R}^3$ - then in the domains E there are no magnetic surfaces and no first integrals of

system (5.5). In the paper [24], Moffatt proved that this can occur only for the “force-free” equilibria

$$(5.6) \quad \mathbf{V} = c_1 \mathbf{B}, \quad \rho = c_2, \quad \text{curl } \mathbf{B} = c_3 \mathbf{B}, \quad P + \rho \mathbf{V}^2/2 = c_4,$$

where c_j are constants. For these solutions, the intrinsic symmetries (5.2) degenerate into the constant scaling transformations.

Remark 6. In Section 10 below, we present a general construction of the force-free magnetic fields that satisfy the equations $\text{curl } \mathbf{B} = \alpha \mathbf{B}$, $\text{div } \mathbf{B} = 0$, where $\alpha = \frac{1}{r}$ const.

Formulae (5.2) imply $\mathbf{V}_1 = \lambda_1(\mathbf{x}) \mathbf{B}_1 / \sqrt{\mu \rho_1(\mathbf{x})}$ where $\lambda_1(\mathbf{x}) = c_1(\mathbf{x})/b_1(\mathbf{x})$. The last formula (5.2) yields $\lambda_1^2(\mathbf{x}) = 1 - C_1(1 - \lambda^2(\mathbf{x}))/b_1^2(\mathbf{x})$. Hence for $C_1 \neq 0$ we have $|\lambda_1(\mathbf{x})| = 1$ if and only if $|\lambda(\mathbf{x})| = 1$. Therefore the Chandrasekhar solutions (3.2), $|\lambda(\mathbf{x})| = 1$, are invariant under transforms (5.2).

Example 1. For any solution to the plasma equilibrium equations

$$(5.7) \quad \text{curl } \mathbf{B} \times \mathbf{B} = \mu \text{grad } P, \quad \text{div } \mathbf{B} = 0, \quad \mathbf{V} = 0,$$

the surfaces of constant pressure $P(\mathbf{x}) = \text{const}$ are magnetic surfaces [2]. Applying transformations (5.2) to any plasma equilibrium (5.7) ($\lambda(\mathbf{x}) \equiv 0$), we obtain new field-aligned solutions

$$(5.8) \quad \mathbf{B}_1 = b_1(\mathbf{x}) \mathbf{B}, \quad \mathbf{V}_1 = \frac{c_1(\mathbf{x})}{a_1(\mathbf{x})\sqrt{\mu}} \mathbf{B}, \quad P_1 = C_1 P - \frac{c_1^2(\mathbf{x})}{2\mu} \mathbf{B}^2,$$

where $b_1^2(\mathbf{x}) - c_1^2(\mathbf{x}) = C_1 = \text{const}$. Here $\rho_1(\mathbf{x}) = a_1^2(\mathbf{x})$ and $b_1(\mathbf{x})$ are arbitrary smooth functions that are constant on the magnetic field lines.

Solutions (5.8) have the following physical property: the ratio of the plasma magnetic and kinetic energy

$$(5.9) \quad \frac{\mu \rho_1 \mathbf{V}_1^2}{\mathbf{B}_1^2} = \frac{b_1^2(\mathbf{x}) - C_1}{b_1^2(\mathbf{x})}$$

is constant on the magnetic surfaces $P(\mathbf{x}) = \text{const}$; but it is variable in the space \mathbb{R}^3 . Note that neither magnetic nor kinetic energy separately are constant on the magnetic surfaces. If $C_1 > 0$ then the kinetic energy is everywhere smaller than the magnetic energy (sub-Alfvénic flows); if $C_1 < 0$, the converse is true (super-Alfvénic flows). The case $C_1 = 0$ and $\rho_1(\mathbf{x}) = \text{const}$ corresponds to the Chandrasekhar equipartition solutions [22], see (3.2).

6. Ball lightning model with dynamics of plasma

In the papers [13, 14], a model of ball lightning is developed where a steady plasma with velocity $\mathbf{V} = 0$ fills a spherical ball and the magnetic field \mathbf{B} is axially symmetric inside the ball and vanishes outside of it. The model is based on an exact solution of the paper [13] which is given in terms of the spherical Bessel functions and the Legendre functions. In what follows, we generalize this model for the non-zero plasma velocity $\mathbf{V} \neq 0$.

In the cylindrical coordinates r, z, ϕ , the axially symmetric magnetic field \mathbf{B} has the form [2]

$$(6.1) \quad \mathbf{B} = \frac{\psi_z}{r} \hat{\mathbf{e}}_r - \frac{\psi_r}{r} \hat{\mathbf{e}}_z + \frac{I}{r} \hat{\mathbf{e}}_\phi,$$

where $\psi(r, z)$ is a flux function, $\psi_z = \partial\psi/\partial z$, $\psi_r = \partial\psi/\partial r$, $I = I(r, z)$ describes the electric current density and $\hat{\mathbf{e}}_r, \hat{\mathbf{e}}_z, \hat{\mathbf{e}}_\phi$ are the coordinate unit ords. For the axially

symmetric solutions (6.1), the plasma equilibrium equations (5.7) are equivalent to the Grad - Shafranov equation [8, 9]

$$\psi_{rr} - \psi_r/r + \psi_{zz} + I(\psi)I'(\psi) + \mu r^2 p'(\psi) = 0,$$

where $I = I(\psi)$ and $P = p(\psi)$ are arbitrary functions of ψ and prime means the derivative with respect to ψ . The authors of [13, 14] assume $I(\psi) = \alpha\psi$, $p(\psi) = p_1 - c\psi$ and construct a solution $\psi(r, z)$ that satisfies the overdetermined boundary conditions $\psi|_{\partial V} = 0$, $\text{grad } \psi|_{\partial V} = 0$ on a spherical boundary ∂V : $R = a$, $R = \sqrt{r^2 + z^2}$. Hence $\mathbf{B}|_{\partial V} = 0$, $P|_{\partial V} = p_1$ and the solution is continued in the outer space $R > a$ by the trivial solution $\mathbf{B}(\mathbf{x}) = 0$, $P(\mathbf{x}) = p_1$ [13, 14]. The exact solution of [13] can be presented also in the form

$$(6.2) \quad \psi(r, z) = mr^2 \left(1 + \frac{3 - x_1^2}{\cos x_1} \frac{1}{\alpha^2 R^2} \left(\cos(\alpha R) - \frac{\sin(\alpha R)}{\alpha R} \right) \right),$$

where $m = c\mu/\alpha^2$, $(3 - x_1^2)/\cos x_1 \approx -34.8145$ and $0 \leq R \leq a = x_1/\alpha$. Here $x_1 \approx 5.763459$ is the smallest positive root of the equation $\tan x = 3x/(3 - x^2)$ where $x = \alpha R$. Inside the ball $R \leq a$, the generic magnetic surfaces $\psi(r, z) = \text{const}$ are toroidal. The singular magnetic surfaces are the segment $r = 0$, $-a \leq z \leq a$ and the magnetic axis $r = r_1$, $z = 0$, $0 \leq \phi \leq 2\pi$, that is defined by the conditions $\psi_r = 0$, $\psi_z = 0$.

Let $\beta(\psi)$ be an arbitrary smooth function of ψ , $b_1 = \text{ch } \beta(\psi)$, $c_1 = \text{sh } \beta(\psi)$, and $a_1(\psi) \neq 0$ be another smooth function. By applying the symmetry transforms (5.2) to the plasma equilibrium (6.1), (6.2) we obtain new field-aligned sub-Alfvénic MHD equilibria

$$(6.3) \quad \mathbf{B}_1 = \text{ch } \beta(\psi) \mathbf{B}, \quad \mathbf{V}_1 = \frac{\text{sh } \beta(\psi)}{a_1(\psi) \sqrt{\mu}} \mathbf{B}, \quad P_1 = p_1 - c\psi - \frac{1}{2\mu} \text{sh}^2 \beta(\psi) \mathbf{B}^2,$$

with the plasma density $\rho_1(\mathbf{x}) = a_1^2(\psi(r, z))$. The MHD equilibria (6.3) are defined inside the ball $R \leq a$ and satisfy the boundary conditions $\mathbf{B}_1|_{\partial V} = 0$, $\mathbf{V}_1|_{\partial V} = 0$, $P_1|_{\partial V} = p_1$. Hence the equilibria are continued in the outer space $R > a$ by the trivial solution $\mathbf{B}(\mathbf{x}) = 0$, $\mathbf{V}(\mathbf{x}) = 0$, $P(\mathbf{x}) = p_1 = \text{const}$, $\rho_1(\mathbf{x}) = a_1^2(0) = \text{const}$. The field-aligned equilibria (6.3) model ball lightning with the variable plasma density and sub-Alfvénic dynamics of plasma inside the fireball. The MHD equilibria (6.3) manifest also neutral modes for the stability analysis [14] of the ball lightning model [13, 14].

7. Breaking of the geometrical symmetries

The symmetry transforms (5.2) have important applications connected with the breaking of the geometrical symmetries of the field-aligned MHD equilibria. Suppose that a field-aligned MHD equilibrium (5.1) possesses some geometrical symmetry - translational, axial or helical and that all magnetic field lines in a domain D either are closed curves or go to infinity. Then the symmetries (5.2) are defined by the first integrals $a_1(\mathbf{x})$, $b_1(\mathbf{x})$, $c_1(\mathbf{x})$ of the system (5.5), which are functions of the two transversal variables. These functions generically are not invariant with respect to the above geometrical symmetries. Hence the new equilibrium ρ_1 , \mathbf{B}_1 , \mathbf{V}_1 , P_1 (5.2) is non-symmetric. This means that the symmetry transforms

(2.8), (5.2) break the geometrical symmetry of the original field-aligned equilibrium $\rho, \mathbf{B}, \mathbf{V}, P$ (5.1). In Section 8, we present the global non-symmetric MHD equilibria obtained by breaking of the helical symmetry.

In this Section, we give an example of the axial symmetry breaking for a plasma equilibrium with toroidal magnetic surfaces and closed magnetic field lines. We consider the magnetic analog of Hill’s spherical vortex [11] for which the magnetic field \mathbf{B} is axially symmetric and pure poloidal:

$$(7.1) \quad \mathbf{B} = \frac{\psi_z}{r} \hat{\mathbf{e}}_r - \frac{\psi_r}{r} \hat{\mathbf{e}}_z.$$

The notations are the same as in (6.1), $I \equiv 0$. Inside the ball $R \leq a$, $R = \sqrt{r^2 + z^2}$, Hill’s solution has the flux function

$$(7.2) \quad \psi(r, z) = c\mu r^2(R^2 - a^2)/10, \quad p(\psi) = p_1 - c\psi.$$

Outside of the ball, $R > a$, the flux function is

$$(7.3) \quad \hat{\psi}(r, z) = Ar^2(R^{-3} - a^{-3}), \quad A = -a^5 c\mu/15, \quad p(\psi) = p_1.$$

The formulae (7.1) - (7.3) imply $\mathbf{B}(\mathbf{x}) = \hat{\mathbf{B}}(\mathbf{x})$ for $|\mathbf{x}| = a$. In the outer space $R > a$, the magnetic field $\hat{\mathbf{B}}$ (7.1), (7.3) is potential: $\hat{\mathbf{B}} = \text{grad}(Az(R^{-3} + 2a^{-3}))$, and has a constant asymptotics $\hat{\mathbf{B}} \rightarrow 2a^{-3}A\hat{\mathbf{e}}_z$ as $R \rightarrow \infty$.

The magnetic field lines (7.1) have two first integrals: the $\psi(r, z)$ and the angle ϕ . Hence any smooth function $f(\psi(r, z), \phi)$ also is their first integral. Inside the ball $R \leq a$, the magnetic field lines are either closed curves $\psi(r, z) = C_1, \phi = C_2$ or the separatrix $r = 0, -a < z < a$ or the rest points: $r = a/\sqrt{2}, z = 0, 0 \leq \phi \leq 2\pi$, where $\psi = -\ell, \ell = a^4 c\mu/40$.

Let $a(\psi, \phi) > 0$ and $\beta(\psi, \phi)$ be any smooth functions on the annulus $-\ell \leq \psi \leq 0, 0 \leq \phi \leq 2\pi$ such that $a(0, \phi) = 1$ and $\beta(0, \phi) = 0$. Applying the symmetry transform (5.2) with functions $a_1(\mathbf{x}) = a(\psi, \phi), b_1(\mathbf{x}) = \text{ch } \beta(\psi, \phi), c_1(\mathbf{x}) = \text{sh } \beta(\psi, \phi)$ where $\psi = \psi(r, z)$ (7.2) to Hill’s solution (7.1), we obtain new sub-Alfvénic MHD equilibria

$$(7.4) \quad \mathbf{B}_1 = \text{ch } \beta(\psi, \phi)\mathbf{B}, \quad \mathbf{V}_1 = \frac{\text{sh } \beta(\psi, \phi)}{a(\psi, \phi)\sqrt{\mu}}\mathbf{B},$$

$$\rho_1(\mathbf{x}) = a^2(\psi, \phi)\rho_0, \quad P_1(\mathbf{x}) = p_1 - c\psi - \text{sh}^2 \beta(\psi, \phi)\mathbf{B}^2/(2\mu)$$

inside the ball $R \leq a$. On the sphere $R = a$, the equilibria (7.4) coincide with Hill’s solution (7.1), (7.2) and hence are continued in the outer space $R > a$ by the same Hill’s plasma equilibrium (7.1), (7.3). It is evident that the equilibria (7.4) have toroidal magnetic surfaces $\psi(r, z) = \text{const}$ and closed magnetic field lines and are non-symmetric. Hence the symmetry transform (5.2) breaks the axial symmetry of Hill’s solution.

8. Graphs Γ and cellular complexes C^2 defined by the MHD equilibria

I. Any smooth MHD equilibrium in \mathbb{R}^3 with non-collinear vector fields \mathbf{B} and \mathbf{V} uniquely defines a family of magnetic surfaces $\psi(\mathbf{x}) = \text{const}$ (2.4) which fibre the Euclidean space \mathbb{R}^3 . For a single-valued flux function $\psi(\mathbf{x})$ (2.4), the regular magnetic surfaces are either smooth tori or cylinders or planes. The singular magnetic surfaces are points of bifurcations of the family of magnetic surfaces. A singular magnetic surface is either a 1-dimensional magnetic axis or a 2-dimensional surface with singularities or an asymptotic magnetic surface at infinity. Any complete



FIGURE 1. Graph Γ for the non-collinear MHD equilibria (4.1) and (4.2).

family of magnetic surfaces is equivalent to a graph Γ . Indeed, let us denote each singular magnetic surface as a vertex of a graph Γ . Any 1-parametric family of regular magnetic surfaces connects certain two singular magnetic surfaces. Such 1-parametric families form the edges of the graph Γ . For the symmetry transforms (2.8), the functions $a(\mathbf{x})$, $b(\mathbf{x})$, $c(\mathbf{x})$ and $\rho(\mathbf{x})$ are constant on the magnetic surfaces. Hence they are actually continuous functions on the graph Γ that are smooth outside of the vertices.

The cylindrically symmetric exact solutions (4.1) and (4.2) define the simplest possible graph Γ , see Figure 1. Here two vertices correspond to the singular magnetic surfaces that are the 1-dimensional magnetic axis $r = 0$, $-\infty < z < +\infty$, and the asymptotic magnetic surface at $r \rightarrow \infty$. The edge corresponds to the 1-parametric family of regular magnetic surfaces - cylinders $r = r_0$ where $0 < r_0 < \infty$.

For the exact field-aligned MHD equilibria (5.1), the magnetic surfaces are not uniquely defined in the domains E_j where all magnetic field lines go to infinity or are closed curves. Indeed, any smooth first integral $f(r, z, \phi)$ of the magnetic dynamical system (5.5) in the domain E_j defines invariant magnetic surfaces $f(r, z, \phi) = \text{const}$. The set of magnetic field lines in such a domain E_j is a 2-dimensional cell C_j whose boundary corresponds either to a singular magnetic surface S_k or to a 1-parametric family of the magnetic field lines at infinity. In the domains T_m where the generic magnetic field lines are bounded and are not closed curves, the magnetic surfaces are uniquely defined as the closures of the magnetic field lines. In the domains T_m , the magnetic surfaces form 1-parametric families I_m of tori \mathbb{T}^2 . Therefore the field-aligned MHD equilibria define a 2-dimensional cellular complex C^2 that consists of the vertices, the edges and the 2-dimensional cells. The vertices correspond to the singular magnetic surfaces S_k . The edges correspond to the 1-parametric families I_m of the toroidal magnetic surfaces and the 1-parametric families of the magnetic field lines at infinity. The 2-dimensional cells C_j correspond to the 2-parametric families of magnetic field lines in the domains E_j where all lines either go to infinity or are closed curves.

For the symmetry transforms (5.2) for the field-aligned MHD equilibria, the smooth functions $a_1(\mathbf{x})$, $b_1(\mathbf{x})$, $c_1(\mathbf{x})$ and $\rho(\mathbf{x})$ are constant on the magnetic field lines. Hence they are actually continuous functions on the corresponding 2-dimensional cellular complex C^2 that are smooth outside of the vertices. In what follows, we present the non-isomorphic 2-dimensional cellular complexes C^2 that correspond to the exact field-aligned MHD equilibria derived in [5, 15].

II. In [15], we derived the following exact axially symmetric solutions to the plasma equilibrium equations (5.7):

$$(8.1) \quad \mathbf{B} = (\psi_z \hat{\mathbf{e}}_r - \psi_r \hat{\mathbf{e}}_z + \sqrt{8\beta N} \psi \hat{\mathbf{e}}_\phi) / r, \quad P = p_1 - 2\beta^2 \psi^2(r, z) / \mu,$$

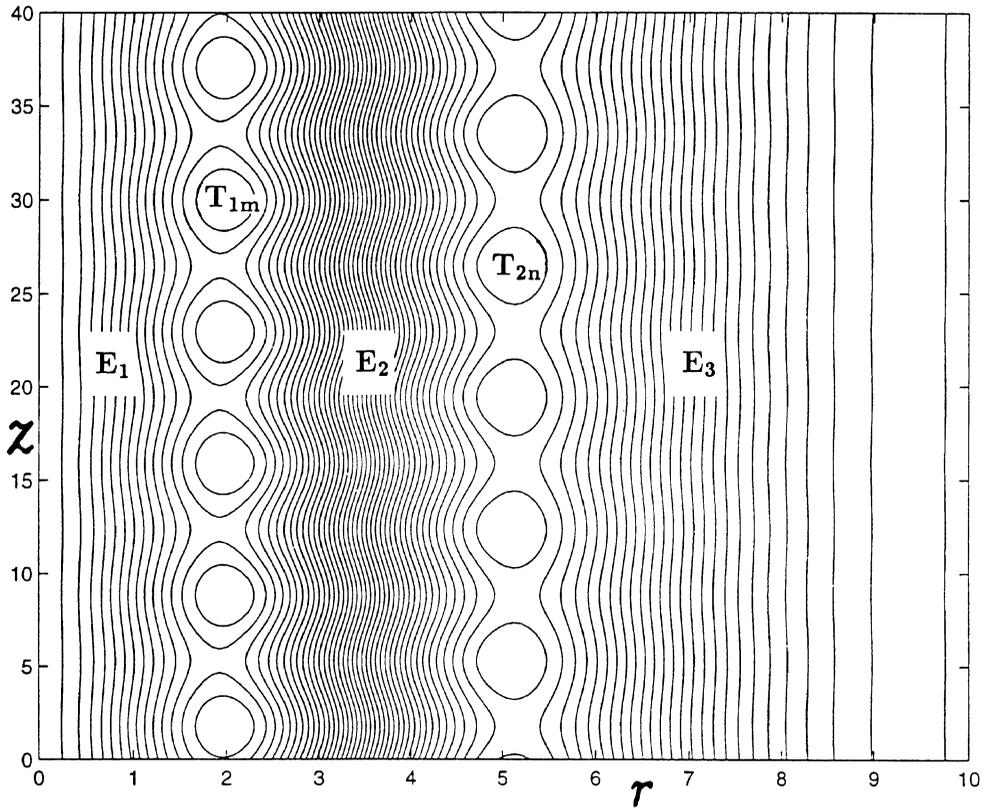


FIGURE 2. Periodic in the variable z level curves $\psi_2(r, z) = \text{const}$ (8.4).

where $\hat{e}_r, \hat{e}_z, \hat{e}_\phi$ are unit orfts in the cylindrical system of coordinates r, z, ϕ and flux functions $\psi(r, z)$ are parametrized by the integers $N \geq 0$ and have the form

$$(8.2) \quad \psi_N(r, z) = e^{-\beta r^2} (a_N L_N^*(y) + \sum_{n=1}^{N-1} L_n^*(y) (a_n \cos(\omega_n z) + b_n \sin(\omega_n z))).$$

Here $\omega_n = \sqrt{8\beta(N-n)}$, $y = 2\beta r^2$, $\beta > 0$ and $L_n^*(y)$ are the primitive functions of the Laguerre polynomials [30], a_n and b_n are arbitrary constants. Applying the symmetry transform (2.8) or (5.2) to these solutions, we obtain an infinite family of new exact MHD equilibria for $C = k^2$:

$$(8.3) \quad \mathbf{B}_1 = k \text{ch } \beta(\mathbf{x}) \mathbf{B}, \quad \mathbf{V}_1 = \frac{k \text{sh } \beta(\mathbf{x})}{\sqrt{\mu \rho_1(\mathbf{x})}} \mathbf{B}, \quad P_1 = k^2 P - \frac{k^2}{2\mu} \text{sh}^2 \beta(\mathbf{x}) \mathbf{B}^2.$$

Here $\beta(\mathbf{x})$ and plasma density $\rho_1(\mathbf{x}) = a^2(\mathbf{x}) > 0$ are arbitrary smooth functions that are constant on the magnetic field lines (8.1).

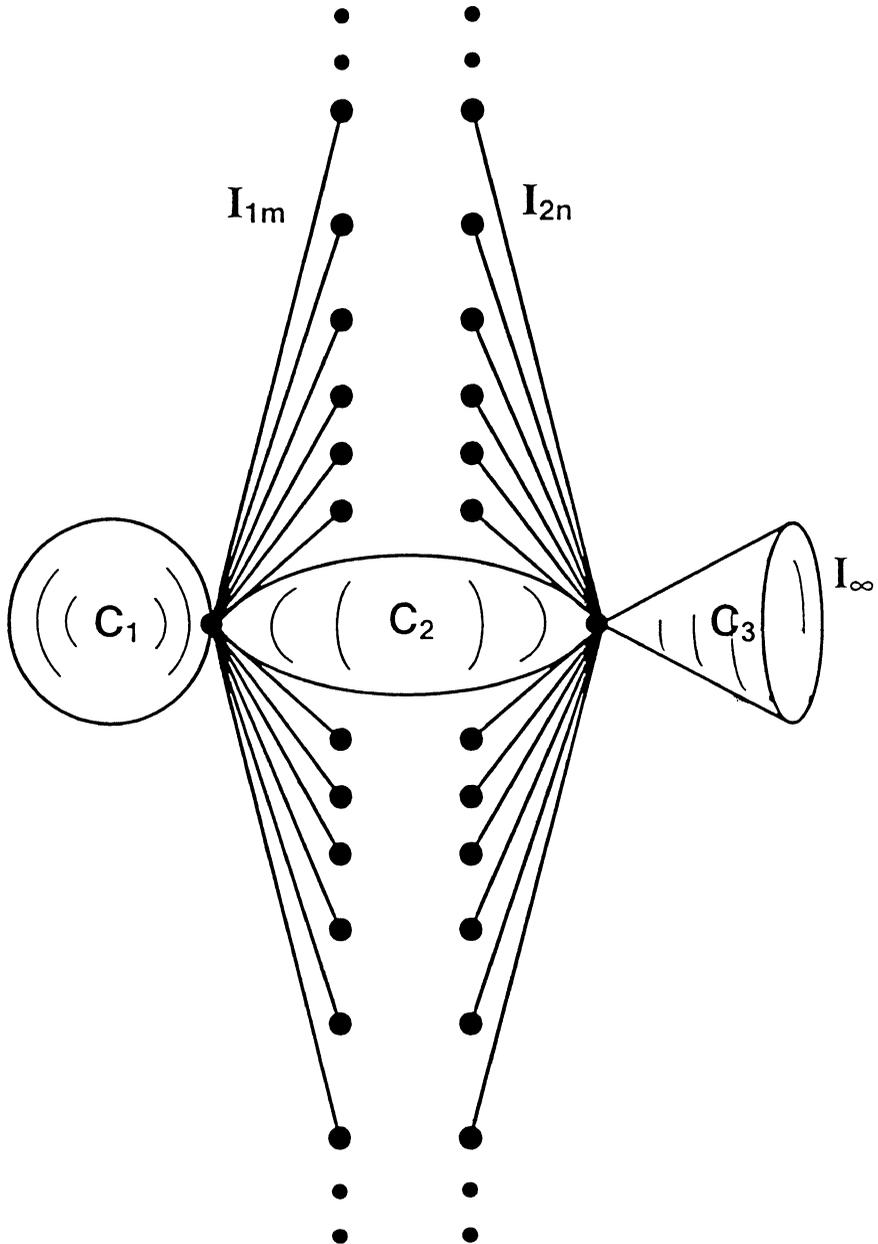


FIGURE 3. Cellular complex C^2 for the z -periodic MHD equilibrium (8.3), (8.4).

Example 3. The simplest z -periodic plasma equilibrium (8.1) is defined for $N = 2$ and has the flux function (8.2):

$$(8.4) \quad \psi_2(r, z) = e^{-\beta r^2} (-\beta r^4 + r^2 + a_1 \sin(\sqrt{8\beta z} r^2)).$$

The corresponding level curves $\psi_2(r, z) = \text{const}$ are shown in Figure 2 above for $\beta = 0.1, a_1 = 0.1$. There are three infinite in z domains E_1, E_2, E_3 , which are

filled with the unbounded magnetic field lines that go to the infinities $z \rightarrow \pm\infty$ and two infinite families of bounded domains T_{1m} and T_{2n} which are filled with the magnetic field lines that lie on the toroidal magnetic surfaces \mathbb{T}^2 . These tori \mathbb{T}^2 are axially symmetric and are obtained from the closed curves in Figure 2 by rotating around the vertical axis z . The two families of domains T_{1m} and T_{2n} are bounded by the two singular surfaces S_1 and S_2 . Each of the surfaces S_1 and S_2 contains an infinite set of saddle points of function $\psi_2(r, z)$ (8.4). The exterior domain E_3 is bounded by the 1-parametric family of the magnetic field lines at $r \rightarrow \infty$.

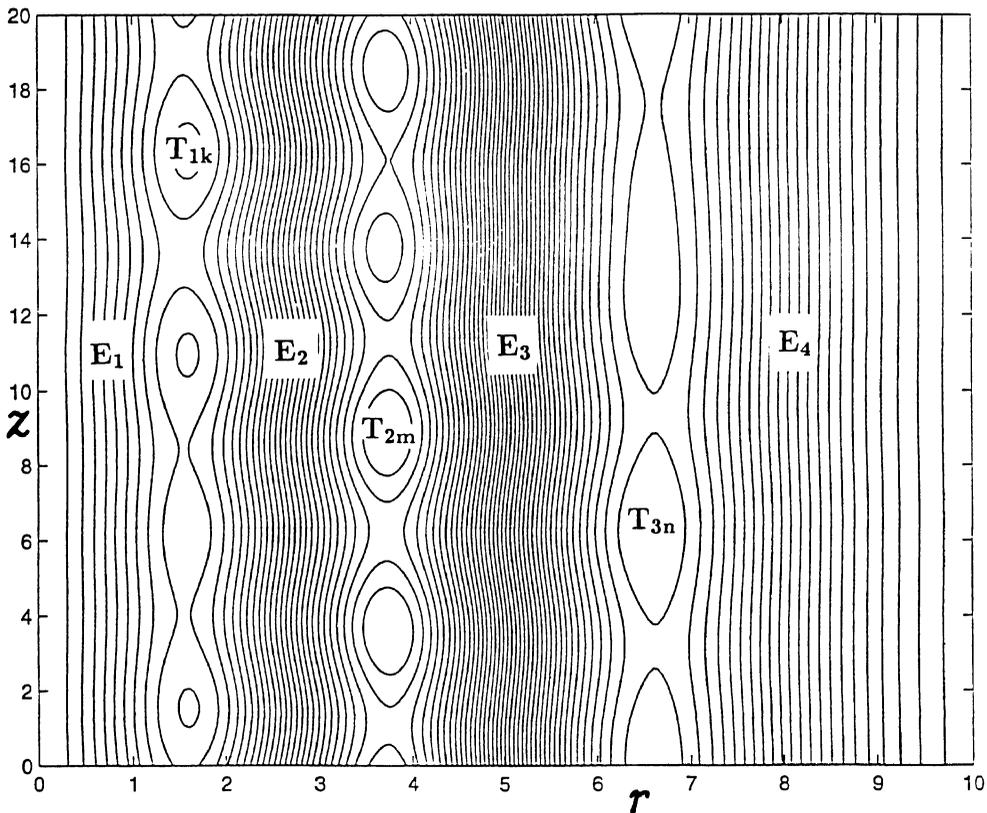


FIGURE 4. Quasi-periodic in the variable z level curves $\psi_3(r, z) = \text{const}$ (8.5).

In the domains T_{1m} and T_{2n} , the generic toroidal magnetic surfaces are uniquely defined as the closures of the magnetic field lines (8.1). The cellular complex C^2 describing the magnetic surfaces and the unbounded magnetic field lines for the exact solution (8.3) is shown in Figure 3. Here the 2-dimensional cells C_1, C_2, C_3 correspond to the 2-parametric families of the unbounded magnetic field lines in the domains E_1, E_2, E_3 . The boundaries of the cells C_1 and C_2 are the vertices that correspond to the two singular magnetic surfaces S_1 and S_2 . The boundary of

the cell C_3 is the vertex S_2 and the edge I_∞ that corresponds to the 1-parametric family of the magnetic field lines at $r \rightarrow \infty$.

Each edge I_{1m} and I_{2n} in Figure 3 corresponds to a 1-parametric family of toroidal magnetic surfaces that connect one of the singular surfaces S_1 or S_2 with a singular "surface" S_{1m} or S_{2n} that is the 1-dimensional inner circular magnetic axis. The magnetic axes S_{1m} and S_{2n} correspond to the maxima and minima of the function $\psi_2(r, z)$ (8.4).

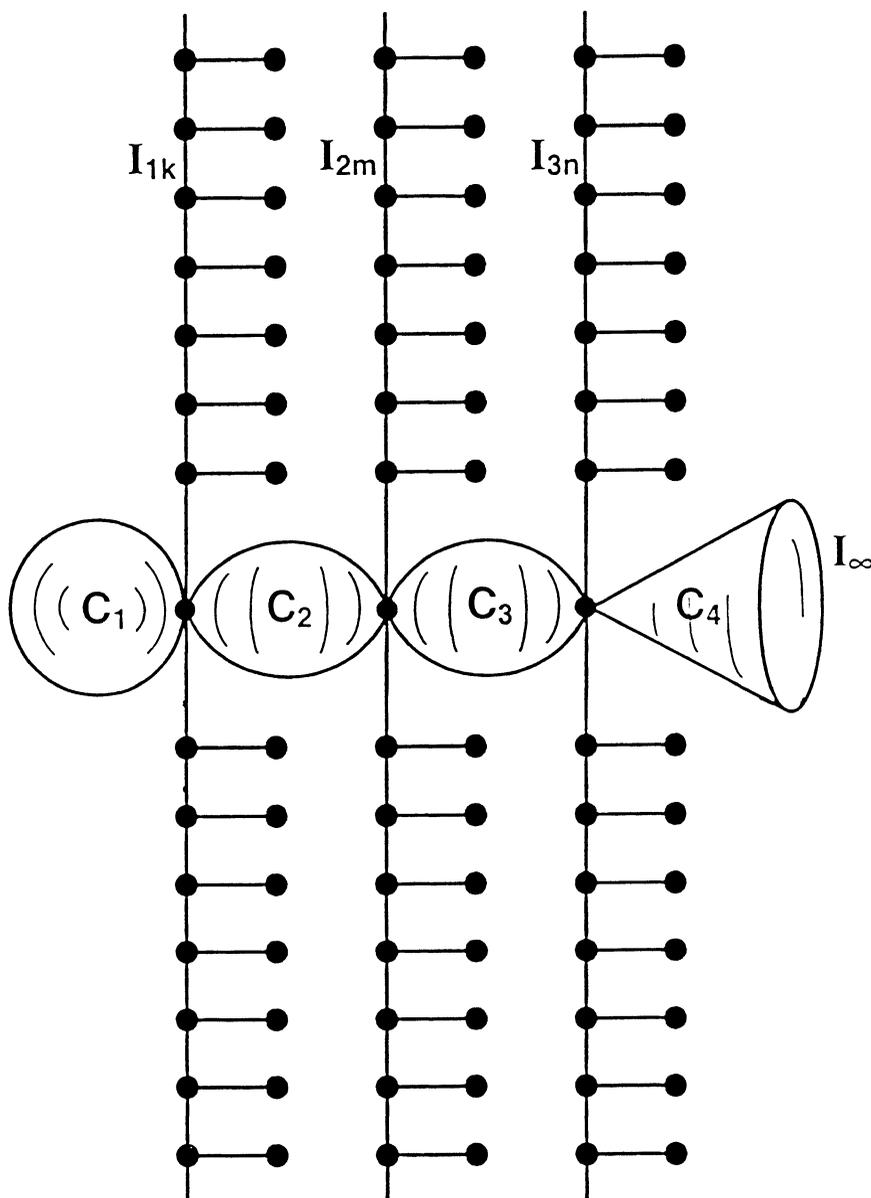


FIGURE 5. Cellular complex C^2 for the MHD equilibrium (8.3), (8.5).

Example 4. The generic flux functions (8.2) are quasi-periodic in variable z because the frequencies $\omega_n = \sqrt{8\beta(N-n)}$ are rationally independent. The simplest z -quasi-periodic solution (8.2) is defined for $N = 3$ and has the form

$$(8.5) \quad \psi_3(r, z) = e^{-\beta r^2} (L_3^*(y) + \frac{1}{20} \sin(2\sqrt{2\beta}z)L_2^*(y) + \frac{1}{20} \sin(4\sqrt{\beta}(z-1))L_1^*(y)).$$

There are four infinite domains E_1, E_2, E_3, E_4 in Figure 4 that are filled with the unbounded magnetic field lines going to the infinity in the variable z . The magnetic field lines in these domains corresponds to the 2-dimensional cells C_1, C_2, C_3, C_4 of the cellular complex C^2 in Figure 5.

The edges I_{1k}, I_{2m}, I_{3n} of the cellular complex C^2 correspond to the 1-parametric families of the toroidal magnetic surfaces in the domains T_{1k}, T_{2m}, T_{3n} and the edge I_∞ corresponds to the magnetic field lines at infinity $r \rightarrow \infty$. The toroidal magnetic surfaces are obtained by rotating the closed curves in Figure 4 around the axis z . The non-periodicity in z of the exact solution (8.5) causes the appearance of infinitely many new singular 2-dimensional magnetic surfaces $S_{1\alpha}, S_{2\beta}, S_{3\gamma}$, instead of the two singular surfaces S_1 and S_2 for the z -periodic solution (8.4). The corresponding vertices lie on the three vertical lines in Figure 5.

The smooth functions $a_1(\mathbf{x}), b_1(\mathbf{x}), c_1(\mathbf{x})$ and $\rho(\mathbf{x})$ that define the symmetry transforms (5.2) for the field-aligned MHD equilibria (8.3), are constant on the magnetic field lines (8.1). Hence they are actually functions on the cellular complexes C^2 shown in Figure 3 and Figure 5.

The groups of symmetries G_m (3.6) depend on the topology of the graph Γ or the 2-dimensional complex C^2 defined by the given MHD equilibrium because the functions $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ (3.4) are continuous functions on the graph Γ or on the cellular complex C^2 that are smooth outside of the vertices. Hence we have infinitely many different infinite-dimensional abelian groups G_m of symmetries of the MHD equilibrium equations which depend on the topological properties of the given equilibrium solutions, for example the exact solutions (8.1) - (8.2).

9. Astrophysical jets as the global non-symmetric MHD equilibria

I. We consider the following helically symmetric [10, 29] magnetic fields:

$$(9.1) \quad \mathbf{B}_h = \frac{\psi_u}{r} \hat{\mathbf{e}}_r + B_1 \hat{\mathbf{e}}_z + B_2 \hat{\mathbf{e}}_\phi, \quad B_1 = \frac{\alpha\gamma\psi - r\psi_r}{r^2 + \gamma^2}, \quad B_2 = \frac{\alpha r\psi + \gamma\psi_r}{r^2 + \gamma^2},$$

where $\hat{\mathbf{e}}_r, \hat{\mathbf{e}}_z, \hat{\mathbf{e}}_\phi$ are the unit ords in the cylindrical coordinates r, z, ϕ and $\psi = \psi(r, u)$ is the flux function, $u = z - \gamma\phi, \alpha = \text{const}, \gamma = \text{const}$. In [5], we derived the exact helically symmetric plasma equilibria (9.1) that correspond to the flux functions

$$(9.2) \quad \psi_{Nmn} = e^{-\beta r^2} (a_N B_{0N}(y) + r^m B_{mn}(y)(a_{mn} \cos(mu/\gamma) + b_{mn} \sin(mu/\gamma))),$$

where N, m, n are arbitrary integers ≥ 0 satisfying the inequality $2N > 2n + m$, and $y = 2\beta r^2, \beta > 0$. The pressure P in the plasma equilibrium equations (5.7) is $P = p_0 - 2\beta^2 \psi^2 / \mu$. The polynomials $B_{mn}(y)$ have the form

$$B_{mn}(y) = \frac{d^m}{dy^m} L_{m+n}(y) - k_{mn} y \frac{d^{m+1}}{dy^{m+1}} L_{m+n}(y),$$

where $L_p(y)$ are the Laguerre polynomials [30]

$$L_p(y) = \frac{1}{p!} e^y \frac{d^p}{dy^p} (e^{-y} y^p) = \sum_{k=0}^p \frac{(-1)^k p!}{(k!)^2 (p-k)!} y^k.$$

For the exact solutions (9.1) - (9.2), the helically symmetric magnetic surfaces are defined by the equations $\psi_{Nmn}(r, u) = \text{const}$, where $u = z - \gamma\phi$.

The simplest exact solution (9.2) is defined for $N = 1, m = 1, n = 0$. The corresponding flux function (9.2) has the form

$$(9.3) \quad \psi_{110}(r, z, \phi) = e^{-\beta r^2} (1 - 4\beta r^2 + a_1 r \cos(z/\gamma - \phi)).$$

II. Applying the symmetry transforms (5.2) to the exact solutions (9.1) - (9.2), we obtain an infinite family of new field-aligned MHD equilibria for $C = k^2$:

$$(9.4) \quad \mathbf{B}_1 = k \operatorname{ch} f(\mathbf{x}) \mathbf{B}_h, \quad \mathbf{V}_1 = \frac{k \operatorname{sh} f(\mathbf{x})}{\sqrt{\mu \rho_1(\mathbf{x})}} \mathbf{B}_h, \quad P_1 = k^2 P - \frac{k^2}{2\mu} \operatorname{sh}^2 f(\mathbf{x}) \mathbf{B}_h^2.$$

Here $f(\mathbf{x})$ and the plasma density $\rho_1(\mathbf{x}) = a_1^2(\mathbf{x})$ are arbitrary smooth functions that are constant on the magnetic field lines (9.1). These lines go to infinity in the variable z . Hence functions $f(\mathbf{x})$ and $a_1(\mathbf{x})$ are arbitrary functions of the two transversal variables and are not helically symmetric in general. Therefore the obtained exact solutions (9.4) are non-symmetric.

For the exact solutions (9.1) that satisfy condition $B_1 > 0$ for $z = 0$ and arbitrary r and ϕ , all magnetic field lines are unbounded in variable z and are uniquely defined by their intersections with the plane $z = 0$. Hence the set of all distinct magnetic field lines is the plane \mathbb{R}^2 . The cellular complex C^2 for these helically symmetric solutions is the 2-dimensional disk shown in Figure 6 where the boundary circle corresponds to the magnetic field lines at infinity $r \rightarrow \infty$.

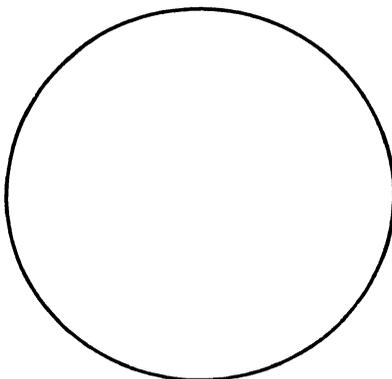


FIGURE 6. Cellular complex C^2 for the helically symmetric field-aligned MHD equilibria.

For the exact solutions (9.4), the magnetic surfaces are not uniquely defined. However there are special helically symmetric magnetic surfaces defined by the

constant levels of the flux function $\psi_{Nmn}(r, z, \phi) = \text{const}$. Figure 7 shows the level curves $\psi_{110}(r, z, \phi) = \text{const}$ at $z = 0$ for $a_1 = -1$, $\beta = 0.1$, $\gamma = \sqrt{5/2}$. The function $\psi_{110}(r, 0, \phi)$ (9.3) achieves its maximum at $r = 0.8968$, $\phi = \pi$ and minimum at $r = 3.0168$, $\phi = 0$. The 2-dimensional helically symmetric magnetic surfaces are obtained from the curves in Figure 7 by simultaneous rotation in ϕ with angular speed 1 and translation in z with speed γ .

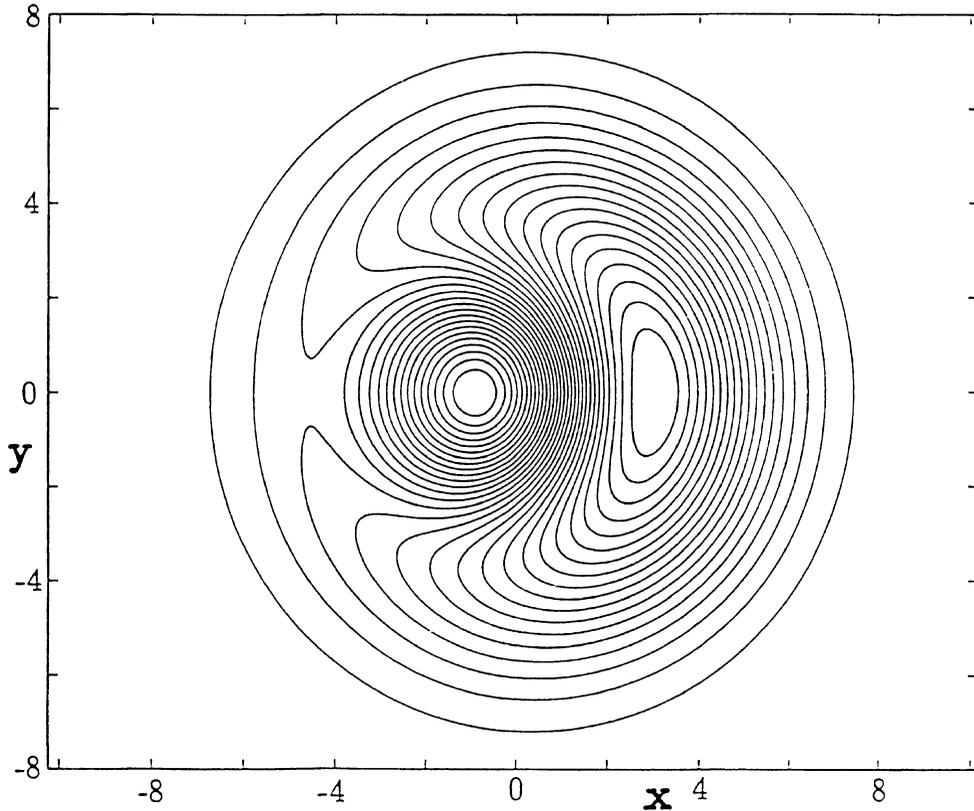


FIGURE 7. Section $z = 0$ of the helically symmetric magnetic surfaces (9.3).

III. Solutions (9.4) are sub-Alfvénic because the ratio of the plasma kinetic and magnetic energies is

$$(9.5) \quad \mu\rho_1 \mathbf{V}_1^2 / \mathbf{B}_1^2 = \text{th}^2 f(\mathbf{x}) < 1.$$

This ratio is variable in the space \mathbb{R}^3 but is constant on the helically symmetric magnetic surfaces $\psi_{Nmn}(r, z) = \text{const}$. For $f(\mathbf{x}) = c_\ell \psi_{Nmn}^{2\ell+m}(\mathbf{x})$, $\rho_1(\mathbf{x}) = b_\ell \psi_{Nmn}^{2\ell}(\mathbf{x})$ (9.4), $\ell > 0$, $m \geq 0$, the plasma magnetic and kinetic energies and its mass are finite in any layer $c_1 \leq z \leq c_2$ because $\psi_{Nmn}(r, u) \approx c_N r^{2N} \exp(-\beta r^2)$ at

$r \rightarrow \infty$. All magnetic field lines and plasma streamlines are bounded in the radial variable r because the leading term of the flux function $\psi_{Nmn}(r, u)$ (9.2) is $b_N(-2\beta)^N \exp(-\beta r^2)r^{2N}/N!$ at $r \gg 1$, $b_N = a_N(1 + k_{0N})$. In view of the rapid decreasing of the magnetic field $\mathbf{B}_h(r, u)$ at $r \rightarrow \infty$, the plasma magnetic and kinetic energies and its mass are concentrated near the z -axis $r = 0$. Therefore the obtained exact MHD equilibria (9.4) model the astrophysical jets outside of their accretion disks; for example the jet in the elliptic galaxy Messier 87 [17]. These exact solutions satisfy the necessary physical conditions at $r \rightarrow \infty$. The total plasma kinetic and magnetic energy and its mass are finite in any layer $c_1 < z < c_2$.

10. Symmetry transforms for the compressible gas plasma

THEOREM 3. 1) Equations (2.1) - (2.3) possess the following intrinsic symmetries

$$(10.1) \quad \rho_1(\mathbf{x}) = a^2(\mathbf{x})\rho(\mathbf{x}), \quad \mathbf{B}_1 = b\mathbf{B}, \quad \mathbf{V}_1 = \frac{b}{a(\mathbf{x})}\mathbf{V},$$

$$P_1 = b^2P, \quad S_1 = S + 2c_v(\ln |b| - \gamma \ln |a(\mathbf{x})|),$$

where $a(\mathbf{x}) \neq 0$ is an arbitrary smooth function that is constant on the magnetic surfaces for a given MHD equilibrium $\mathbf{B}, \mathbf{V}, \rho, P$ and $b = \text{const} \neq 0$. Transforms (10.1) preserve the equation of state (2.3).

2) Suppose that for some equation of state $P = \rho^\gamma f(S)$ magnetic surfaces are closed in some domain E . Then the entropy density S is constant on the magnetic surfaces and there exist a symmetry (10.1) that transforms the MHD equilibrium in the domain E into an isentropic equilibrium satisfying equation $P_1 = b^2\rho_1^\gamma$.

Proof. 1) Equation $(\mathbf{V} \cdot \text{grad } a(\mathbf{x})) = 0$ implies $\rho_1(\mathbf{V}_1 \cdot \text{grad})\mathbf{V}_1 = b^2\rho(\mathbf{V} \cdot \text{grad})\mathbf{V}$ and $\text{div}(\rho_1\mathbf{V}_1) = a(\mathbf{x})b \text{div}(\rho\mathbf{V}) = 0$. Hence equation (2.1) and the first equation (2.2) are satisfied. Equation $\text{curl}(\mathbf{V}_1 \times \mathbf{B}_1) = 0$ follows from the identity $\text{curl } f\mathbf{U} = \text{grad } f \times \mathbf{U} + f \text{curl } \mathbf{U}$ and the equation $\text{grad}(a(\mathbf{x})) \times (\mathbf{V} \times \mathbf{B}) = 0$ that is true for $a(\mathbf{x})$ is constant on the magnetic surfaces.

2) Let us prove that for any variable density $\rho(\mathbf{x})$ vector fields $\rho^{-1}\mathbf{B}$ and \mathbf{V} commute, $[\rho^{-1}\mathbf{B}, \mathbf{V}] = 0$. Indeed, the known identity

$$\text{curl}(\mathbf{X} \times \mathbf{Y}) = (\text{div } \mathbf{Y})\mathbf{X} - (\text{div } \mathbf{X})\mathbf{Y} + [\mathbf{Y}, \mathbf{X}]$$

and the equation $\text{curl}(\mathbf{V} \times \mathbf{B}) = 0$ imply

$$(10.2) \quad \text{curl}(\rho\mathbf{V} \times \rho^{-1}\mathbf{B}) = \text{div}(\rho^{-1}\mathbf{B})\rho\mathbf{V} - \text{div}(\rho\mathbf{V})\rho^{-1}\mathbf{B} + [\rho^{-1}\mathbf{B}, \rho\mathbf{V}] = 0.$$

Substituting equation $\text{div}(\rho\mathbf{V}) = 0$ into (10.2), we obtain

$$(10.3) \quad [\rho^{-1}\mathbf{B}, \rho\mathbf{V}] + \rho \text{div}(\rho^{-1}\mathbf{B})\mathbf{V} = 0.$$

Equation $\text{div } \mathbf{B} = 0$ implies

$$(10.4) \quad \text{div}(\rho\rho^{-1}\mathbf{B}) = \rho \text{div}(\rho^{-1}\mathbf{B}) + \rho^{-1}(\mathbf{B} \cdot \text{grad } \rho) = 0.$$

Equation (10.4) yields

$$\rho[\rho^{-1}\mathbf{B}, \mathbf{V}] = [\rho^{-1}\mathbf{B}, \rho\mathbf{V}] - \rho^{-1}(\mathbf{B} \cdot \text{grad } \rho)\mathbf{V} = [\rho^{-1}\mathbf{B}, \rho\mathbf{V}] + \rho \text{div}(\rho^{-1}\mathbf{B})\mathbf{V}.$$

Hence using equation (10.3) we obtain $[\rho^{-1}\mathbf{B}, \mathbf{V}] = 0$. If for the given MHD equilibrium $\mathbf{B}, \mathbf{V}, \rho, P$ the magnetic surfaces are closed in some domain E than they are tori \mathbb{T}^2 [2] and the commuting vector fields $\rho^{-1}\mathbf{B}$ and \mathbf{V} define the dense quasi-periodic trajectories on almost all tori \mathbb{T}^2 . Hence any function that is constant along

the plasma streamlines is constant on the tori \mathbb{T}^2 . Therefore equation $(\mathbf{V} \cdot \text{grad } S) = 0$ (2.3) implies that entropy S is constant on the magnetic surfaces.

Suppose that plasma satisfies an equation $P = \rho^\gamma f(S)$. The function $a^{2\gamma}(\mathbf{x}) = f(S(\mathbf{x}))$ is constant on the magnetic surfaces \mathbb{T}^2 . The corresponding intrinsic symmetry (10.1) transforms the equation of state $P = \rho^\gamma f(S)$ into the isoentropic equation $P_1 = b^2 \rho_1^\gamma$. □

Remark 7. The commutativity equation $[\rho^{-1}\mathbf{B}, \mathbf{V}] = 0$ implies $[\mathbf{B}, \mathbf{V}] = -\rho^{-1}(\mathbf{V} \cdot \text{grad } \rho)\mathbf{B}$. This proves the formula (2.7).

The symmetry transforms (10.1) form the subgroup $G_{0m} = A_m \oplus R^+ \oplus Z_2 \oplus Z_2 \subset G_m$, see (3.6). Elements of G_{0m} are the sextuples $(\alpha(\mathbf{x}), 0, k, \tau, 1, \eta)$. The subgroup G_{0m} has an additional structure of a module over the associative algebra A_m with the multiplication induced by (3.7).

11. Exact solutions to viscous MHD and Navier-Stokes equations

The previous results are connected with the steady dynamics of an ideal plasma. In this Section, we derive exact solutions for the non-steady dynamics of a non-ideal viscous plasma. The system of viscous magnetohydrodynamics equations has the form

$$(11.1) \quad \rho \frac{\partial \mathbf{V}}{\partial t} + \rho(\mathbf{V} \cdot \text{grad})\mathbf{V} = \mathbf{f} - \text{grad } P + \nu_k \nabla^2 \mathbf{V} + \mathbf{J} \times \mathbf{B},$$

$$(11.2) \quad \frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{V}) = 0,$$

$$(11.3) \quad \frac{\partial \mathbf{B}}{\partial t} = \text{curl}(\mathbf{V} \times \mathbf{B}) + \frac{1}{\mu\sigma} \nabla^2 \mathbf{B}, \quad \text{div } \mathbf{B} = 0,$$

where \mathbf{V} is the vector field of plasma velocity, \mathbf{B} is the magnetic vector field, \mathbf{J} is the electric current density, \mathbf{f} is the vector of external forces, P is the pressure, ρ is the plasma density and ν_k is the kinematic viscosity. The first equation of (11.3) is a consequence (for a constant product $\mu\sigma$) of the Faraday's, Ampère's and Ohm's laws:

$$(11.4) \quad \frac{\partial \mathbf{B}}{\partial t} + \text{curl } \mathbf{E} = 0, \quad \mathbf{J} = \frac{1}{\mu} \text{curl } \mathbf{B}, \quad \mathbf{J} = \sigma(\mathbf{E} + \mathbf{V} \times \mathbf{B}).$$

Here \mathbf{E} is the electric field, μ is the magnetic permeability and σ is the electric conductivity.

We assume that the plasma density ρ , the viscosity ν_k and the product $\mu\sigma$ are constant and $\mathbf{f} = -\rho \text{grad } \Phi$ where $\Phi(x, y, z)$ is an external gravitational potential. For $\mathbf{B} = \mathbf{E} = \mathbf{J} = 0$, equations (11.1) - (11.3) are reduced to the Navier-Stokes (NS) equations that describe the dynamics of viscous fluid. The viscous MHD equations (11.1) - (11.3) and the NS equations were intensively studied, see [31 - 35]. However only highly symmetric exact solutions have been found. We derive new exact solutions to these equations that depend on all four variables t, x, y, z and are non-symmetric in general.

THEOREM 4. *The viscous MHD equations with arbitrary constant parameters ρ , ν_k , $\mu\sigma$ have infinite-dimensional linear spaces S_α of exact solutions*

$$(11.5) \quad \mathbf{B}(t, \mathbf{x}) = C_1 \exp\left(-\frac{\alpha^2}{\mu\sigma}t\right) (\operatorname{curl} \mathbf{B}_1 + \alpha^{-1} \operatorname{curl} \operatorname{curl} \mathbf{B}_1),$$

$$\mathbf{J}(t, \mathbf{x}) = \frac{\alpha}{\mu} \mathbf{B}(t, \mathbf{x}), \quad \mathbf{E}(t, \mathbf{x}) = \frac{\alpha}{\mu\sigma} \mathbf{B}(t, \mathbf{x}),$$

$$\mathbf{V}(t, \mathbf{x}) = C_2 \exp\left(-\frac{\alpha^2 \nu_k}{\rho}t\right) (\operatorname{curl} \mathbf{B}_1 + \alpha^{-1} \operatorname{curl} \operatorname{curl} \mathbf{B}_1),$$

$$(11.6) \quad \mathbf{B}_1(\mathbf{x}) = \int \int_{S^2} \cos(\alpha \mathbf{k} \cdot \mathbf{x}) \mathbf{A}(\mathbf{k}) dS, \quad \rho\Phi + P + \rho \mathbf{V}^2/2 = \text{const},$$

where $\alpha \neq 0$ is an arbitrary parameter, $\mathbf{x} = (x, y, z)$, $\mathbf{k} = (k_1, k_2, k_3)$ and $\mathbf{A}(\mathbf{k})$ is a smooth vector field and the integral in (11.6) is taken with respect to an arbitrary measure dS on the 2-dimensional unit sphere S^2 : $\mathbf{k} \cdot \mathbf{k} = 1$. For $C_1 = 0$, the formulae (11.5) - (11.6) define exact solutions to the NS equations.

Proof. For the vector fields $\mathbf{B}_1(x, y, z)$ (11.6), we have

$$(11.7) \quad \nabla^2 \mathbf{B}_1(\mathbf{x}) = -\alpha^2 \int \int_{S^2} \cos(\alpha \mathbf{k} \cdot \mathbf{x}) (\mathbf{k} \cdot \mathbf{k}) \mathbf{A}(\mathbf{k}) dS = -\alpha^2 \mathbf{B}_1(\mathbf{x}),$$

because $\mathbf{k} \cdot \mathbf{k} = 1$. Hence we get

$$(11.8) \quad \nabla^2 \operatorname{curl} \mathbf{B}_1(\mathbf{x}) = -\alpha^2 \operatorname{curl} \mathbf{B}_1(\mathbf{x}).$$

We consider divergence-free vector fields

$$(11.9) \quad \mathbf{B}_2(\mathbf{x}) = \operatorname{curl} \mathbf{B}_1 + \alpha^{-1} \operatorname{curl} \operatorname{curl} \mathbf{B}_1.$$

Applying the identity $\operatorname{curl}(\operatorname{curl} \mathbf{U}) = \operatorname{grad}(\operatorname{div} \mathbf{U}) - \nabla^2 \mathbf{U}$ to the vector field $\mathbf{U} = \operatorname{curl} \mathbf{B}_1$ and using equation (11.8), we derive $\operatorname{curl} \operatorname{curl} \operatorname{curl} \mathbf{B}_1 = \alpha^2 \operatorname{curl} \mathbf{B}_1$. Hence vector fields (11.9) satisfy the Beltrami equation

$$(11.10) \quad \operatorname{curl} \mathbf{B}_2 = \alpha \mathbf{B}_2.$$

Equations (11.7) and (11.10) imply for the vector fields $\mathbf{B}(t, \mathbf{x})$ and $\mathbf{V}(t, \mathbf{x})$:

$$(11.11) \quad \frac{\partial \mathbf{B}}{\partial t} = -\frac{\alpha^2}{\mu\sigma} \mathbf{B}, \quad \nabla^2 \mathbf{B} = -\alpha^2 \mathbf{B}, \quad \operatorname{curl} \mathbf{B} = \alpha \mathbf{B}, \quad \operatorname{div} \mathbf{B} = 0,$$

$$(11.12) \quad \frac{\partial \mathbf{V}}{\partial t} = -\frac{\alpha^2 \nu_k}{\rho} \mathbf{V}, \quad \nabla^2 \mathbf{V} = -\alpha^2 \mathbf{V}, \quad \operatorname{curl} \mathbf{V} = \alpha \mathbf{V}, \quad \operatorname{div} \mathbf{V} = 0.$$

As is known, the viscous MHD equations (11.1) have also the form

$$(11.13) \quad \rho \frac{\partial \mathbf{V}}{\partial t} = \rho \mathbf{V} \times \operatorname{curl} \mathbf{V} - \operatorname{grad}(\rho\Phi + P + \rho \mathbf{V}^2/2) + \nu_k \nabla^2 \mathbf{V} + \frac{1}{\mu} \operatorname{curl} \mathbf{B} \times \mathbf{B}.$$

Substituting the formulae (11.11) and (11.12), we see that the vector fields $\mathbf{B}(t, \mathbf{x})$ and $\mathbf{V}(t, \mathbf{x})$ satisfy equations (11.1) - (11.3). For each $\alpha \neq 0$, the vector fields

$\mathbf{B}(t, \mathbf{x}), \mathbf{V}(t, \mathbf{x})$ (11.5) - (11.6) form an infinite-dimensional linear space S_α and $S_\alpha \cap S_\beta = 0$ for $\alpha \neq \beta$. □

Remark 8. The vector fields $\mathbf{B}_2(\mathbf{x})$ (11.9) have the form

$$(11.14) \quad \mathbf{B}_2(\mathbf{x}) = \int \int_{S^2} (\sin(\alpha \mathbf{k} \cdot \mathbf{x}) \mathbf{T}(\mathbf{k}) + \cos(\alpha \mathbf{k} \cdot \mathbf{x}) \mathbf{k} \times \mathbf{T}(\mathbf{k})) dS,$$

where $\mathbf{T}(\mathbf{k}) = -\alpha \mathbf{k} \times \mathbf{A}(\mathbf{k})$ are arbitrary smooth vector fields tangent to the unit sphere S^2 . Using the Fourier method, we proved that any smooth solution to the Beltrami equation (11.10) has form (11.14). Formula (11.14) gives a more explicit form for the exact solutions $\mathbf{B}(t, \mathbf{x}) = \exp(-\alpha^2 t / \mu \sigma) \mathbf{B}_2(\mathbf{x})$. For any smooth vector field $\mathbf{A}(\mathbf{k})$ and any smooth measure dS , the vector field $\mathbf{B}_2(\mathbf{x})$ (11.14) and the corresponding exact solutions (11.5) - (11.6) are smooth and bounded and decrease as $C/|\mathbf{x}|$ when $|\mathbf{x}| \rightarrow \infty$.

Remark 9. For $t = 0$, the formulae (11.5), (11.6) represent the arbitrary aligned Beltrami fields $C_{1,2} \mathbf{B}_2(\mathbf{x})$ (11.10). Hence we have derived the exact solutions to the Cauchy problem for the viscous MHD equations provided that the Cauchy initial data for $t = 0$ are aligned Beltrami fields and $\rho \Phi + P + \rho \mathbf{V}^2 / 2 = \text{const}$.

Remark 10. The derived exact solutions imply that any force-free plasma equilibrium is realized as a relaxation limit of a viscous MHD flow. Indeed, any force-free magnetic field \mathbf{B} , $\text{curl } \mathbf{B} = \alpha \mathbf{B}$, has form (11.14). The corresponding exact solution (11.5) - (11.6) for viscous plasma $\nu_k \neq 0$ with perfect electric conductivity $\sigma = \infty$ describes this relaxation. This result confirms the Taylor theory of plasma relaxation [34, 35]. For $C_2 = 0$ and $0 < \sigma < \infty$, the solutions (11.5) - (11.6) describe the force-free magnetic relaxation.

Acknowledgements. The work was supported by a Killam Research Fellowship of Canada and an Alexander von Humboldt Research Award of Germany.

References

- [1] C. S. Gardner, J. M. Greene, M. D. Kruskal and R. M. Miura, Method for solving the Korteweg-de Vries equation, *Phys. Rev. Lett.* **19** (1967), 1095-1097.
- [2] M. D. Kruskal and R. M. Kulsrud, Equilibrium of a magnetically confined plasma in a toroid, *Phys. Fluids*, **1** (1958), 265-274.
- [3] M. D. Kruskal and C. R. Oberman, On the stability of plasma in static equilibrium, *Phys. Fluids*, **1** (1958), 275-280.
- [4] I. B. Bernstein, E. A. Frieman, M. D. Kruskal and R. M. Kulsrud, An energy principle for hydromagnetic stability problems, *Proceedings of the Royal Society of London*, **244** (1958), 17-40.
- [5] O. I. Bogoyavlenskij, Helically symmetric astrophysical jets, *Physical Review E*, **62** (6-B) (2000), 8616-8627.
- [6] B. B. Kadomtsev and V. P. Petviashvili, On the stability of solitary waves in weakly dispersing media, *Sov. Phys. Dokl.* **15** (1970), 539-541.
- [7] M. J. Ablowitz, D. J. Kaup, A. C. Newell and H. Segur, Nonlinear-evolution equations of physical significance, *Phys. Rev. Lett.* **31** (1973), 125-127.
- [8] H. Grad and H. Rubin, Hydromagnetic equilibria and force-free fields, in: *Proceedings of the second United Nations International Conference on the peaceful uses of atomic energy*, United Nations, Geneva, 1958, **31**, pp. 190-197.
- [9] V. D. Shafranov, On magnetohydrodynamical equilibrium configurations, *Soviet Physics JETP* **6** (1958), 545-554.
- [10] J. L. Johnson, C. R. Oberman, R. M. Kulsrud and E. A. Frieman, Some stable hydromagnetic equilibria, *Phys. Fluids* **1** (1958), 281-296.

- [11] J. Hill, On a spherical vortex, *Philos. Trans. R. Soc. London*, Ser. A **185** (1894), 213-245.
- [12] H. Grad, Theory and applications of the nonexistence of simple toroidal plasma equilibrium, *Intern. J. of Fusion Energy* **3** (2) (1985), 33-46.
- [13] H. Wu and Y. Chen, Magneto-hydrodynamics equilibrium of plasma ball lightning, *Phys. Fluids B* **1** (1989), 1753-1755.
- [14] R. Kaiser and D. Lortz, Ball lightning as an example of a magneto-hydrodynamics equilibrium, *Physical Review E* **52** (1995), 3034-3044.
- [15] O. I. Bogoyavlenskij, Astrophysical jets as exact plasma equilibria, *Physical Review Letters*, **84** (2000), 1914-1917.
- [16] O. I. Bogoyavlenskij, Bogoyavlenskij replies, *Physical Review Letters*, v.85 (20) (2000), 4406.
- [17] D. C. Hines, F. N. Owen and J. A. Eilek, Filaments in the radio lobes of M87, *Astroph. J.*, **347** (1989), 713-726.
- [18] A. Ferrari, Modeling extragalactic jets, *Annu. Rev. Astron. Astroph.* **36** (1998), 5390598.
- [19] T. P. Ray, Jets: A star formation perspective, in: *Astrophysical Jets. Open Problems*, Gordon and Breach, Amsterdam, 1998, pp. 173-189.
- [20] A. Salat and R. Kaiser, Three-dimensional closed field line magneto-hydrodynamic equilibria without symmetries, *Phys. Plasmas* **2** (1995), 3777-3781.
- [21] R. Kaiser and A. Salat, Analytic three-dimensional solutions of the magneto-hydrodynamic equations with twisted field lines, *Phys. Rev. Letters* **77** (1996), 3133-3136.
- [22] S. Chandrasekhar, On the stability of the simplest solution of the equations of hydromagnetics, *Proc. Natl. Acad. Sci. U.S.A.* **42** (1956), 273-276.
- [23] H. K. Moffatt, Magnetostatic equilibria and analogous Euler flows of arbitrary complex topology. Part 1. Fundamentals, *J. Fluid Mech.* **159** (1985), 359-378.
- [24] H. K. Moffatt, On the existence, structure and stability of MHD equilibrium states, in: *Turbulence and nonlinear dynamics in MHD flows*, North Holland Publ. C., Amsterdam, 1989, pp. 185-195.
- [25] U. Gebhardt and M. Kiessling, The structure of ideal magneto-hydrodynamics with incompressible steady flow, *Phys. Fluids B* **4** (1992), 1689-1701.
- [26] P. Spiess and K. Elsasser, Two variational principles for incompressible plasmas, *Phys. Plasmas* **6** (1999), 4208-4214.
- [27] H. Grad, Reducible problems in magneto-fluid dynamic steady flows, *Rev. Mod. Phys.* **32** (1960), 830-847.
- [28] W. A. Newcomb, Lagrangian and Hamiltonian methods in magneto-hydrodynamics, *Nuclear Fusion Suppl.* **2** (1962), 451-463.
- [29] B. B. Kadomtsev, Equilibrium of a plasma with helical symmetry, *Soviet Physics JETP* **10** (1960), 962-963.
- [30] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D. C., 1964.
- [31] A. N. Kolmogorov, The local structure of turbulence in incompressible viscous fluid for very large Reynolds' numbers, *C. R. Acad. Sci. URSS* **30** (1941), 301-305.
- [32] G. K. Batchelor, Kolmogoroff's theory of locally isotropic turbulence, *Proc. Camb. Phil. Soc.* **43** (1947), 533-559.
- [33] G. K. Batchelor, *The theory of homogeneous turbulence*, Cambridge University Press, New York, 1959.
- [34] J. B. Taylor, Relaxation of toroidal plasma and generation of reverse magnetic fields, *Phys. Rev. Letters* **33** (1974), 1139-1141.
- [35] J. B. Taylor, Relaxation and magnetic reconnection in plasmas, *Reviews of Modern Physics* **58** (1986), 741-763.

DEPARTMENT OF MATHEMATICS, QUEEN'S UNIVERSITY, KINGSTON, K7L 3N6 CANADA

The p -system I: The Riemann problem

Robin Young

ABSTRACT. The p -system is the prototypical system of nonlinear hyperbolic conservation laws. It is the simplest nontrivial system, and appears as a subsystem of nearly all larger systems of physical importance. Thus a good understanding of the p -system is critical to understanding most interesting systems. The Riemann problem is the building block of general solutions, and many features of solutions are apparent in the simplified context of Riemann solutions. In this paper we solve the Riemann problem for the p -system with a minimum of constitutive assumptions; in particular we do not assume convexity. We deliberately avoid the use of Riemann invariants in the construction, in the expectation of extending our methods to larger systems.

1. Introduction

One of the major open questions in hyperbolic conservation laws is the global existence of solutions having large data. This question is highly dependent on the number of equations: when there are three or more equations, solutions generally exist only for finite times. For 2×2 systems, the celebrated Glimm-Lax theory states that solutions decay, and provides strong global existence results. Our ultimate goal is to extend this theory to larger physical systems of conservation laws. Almost all of these physical systems contain the p -system as a subsystem.

The Glimm-Lax theory [6] relies heavily on the use of Riemann invariant coordinates, which are not available for general systems. Our intention is to rewrite the theory in a manner that can be extended to larger systems. As such, we avoid the use of Riemann invariants, instead preferring to work in physical variables. Moreover, we wish to deal with strong waves, which are ruled out by the requirement that the supnorm of the data be small.

All constructive methods for finding solutions of systems essentially depend on Glimm-type estimates for the total variation [5]. Apart from certain special classes of systems, these in turn rely on a potential which captures

1991 *Mathematics Subject Classification.* 35L65, 35Q35.

This work was partly supported by the NSF under grant number DMS-0104485.

future interaction effects. The central estimate in Glimm's method is an asymptotic estimate of the effects of wave interactions. This is the essential reason that the supnorm of the Cauchy data must be small in Glimm's method.

Here we propose a new point of view in studying wave interactions, namely to treat interactions *exactly* rather than asymptotically. In particular, this will allow us to treat strong waves. As a first step in this program, in this paper we solve the Riemann problem for the p -system with arbitrarily large initial jump.

Apart from the structural simplicity implied by the p -system, we work as generally as possible. The system consists of the two equations

$$v_t - u_x = 0, \quad u_t + p_x = 0,$$

where v is the specific volume of the fluid, u the velocity, and p the pressure. Here we are in the Lagrangian frame of reference, so x is a material coordinate. The system is closed by prescribing a constitutive relation $p = p(v)$. We make a minimum of assumptions on p : namely, we assume p is a C^1 function which decreases as a function of v . Specifically, we assume only

$$p'(v) < 0, \quad \lim_{v \rightarrow 0} p(v) = \infty, \quad \text{and} \quad \lim_{v \rightarrow \infty} p(v) = 0.$$

These assumptions describe hyperbolicity, no infinite compression, and a vacuum state, respectively. The Riemann problem is the Cauchy problem with piecewise constant initial data,

$$U(x, 0) = \begin{cases} U_l, & \text{if } x < 0, \\ U_r, & \text{if } x > 0, \end{cases}$$

where $U = (v \ u)^t$.

All physical systems are endowed with a convex entropy function, which makes the system symmetric hyperbolic [4], and appears to be enough to rule out catastrophic blowup in amplitude [16]. In our case, the structure that helps us is somewhat stronger, and can be loosely described as giving a 'Hamiltonian-like' structure to the system. In fact, the entropy resembles a Hamiltonian energy,

$$H(u, v) = \frac{1}{2} u^2 + V(v),$$

and leads to a system which, although fully nonlinear, is linear in one of the two variables.

Although we do not assume convexity of the pressure, the role of convexity becomes clear in our analysis. Indeed, our construction resembles Oleinik's construction for a scalar nonconvex conservation law [10]. Given volumes v_1 and v_2 to be separated by a wave, we consider the upper or lower convex envelope of the function p , as appropriate. The slope of this envelope, which is monotone, is then the square of the local wave speed, and

the velocity change across the wave is the integrated wavespeed. All waves are concisely described by the relation

$$(1.1) \quad u_r - u_l = g(v_a, v_b),$$

where the subscripts refer to the left, right, ahead and behind states, respectively, and g is a given nonlinear function. By combining forward and backward waves, we get our first existence theorem.

THEOREM 1. *Given constant left and right states $(v_l, u_l)^t$ and $(v_r, u_r)^t$, respectively, there is a unique solution to the corresponding Riemann problem, provided that the condition*

$$u_r - u_l < g(v_l, \infty) + g(v_r, \infty)$$

holds. Moreover, the intermediate state $(v_, u_*)^t$ (and hence wave strength) is a C^1 function of the data.*

The failure of the one-sided condition above heralds the appearance of the vacuum: physically the particles are moving apart so rapidly that a vacuum is formed. Thus the density ρ vanishes, and $v = 1/\rho$ becomes infinite. In fact, to make sense of the vacuum consistent with the Eulerian picture [12, 14], we allow the specific volume v to be a Radon measure, whose singular part is supported on the vacuum region. Since the sound speed vanishes at vacuum, this is stationary, and the singular part takes the form

$$\nu_S = w(t) \delta(x),$$

where the weight $w(t)$ represents the width of the vacuum in physical space. By extending our class of solutions so that the specific volume is a Radon measure, while all other functions remain bounded, we get existence of solutions to the Riemann problem for any choice of initial states.

THEOREM 2. *Given arbitrary states $(v_l, u_l)^t$ and $(v_r, u_r)^t$, there is a unique self-similar solution $(v(x, t), u(x, t))^t$ to the Riemann problem, where the velocity $u(x, t)$ is bounded, while $v(x, t)$ is a Radon measure. Moreover this solution is Lipschitz continuous in time as a distribution in L^1_{loc} .*

The advantage of obtaining an elementary description of waves and the solution of the Riemann problem is that we can use it to describe the results of wave interactions. Thus the states across a pair of waves which are about to interact are related by (1.1), and the interaction problem is the resolution of the Riemann problem connecting the extreme states. The interaction of a forward wave with a backward wave, say, thus is determined by the single nonlinear equation

$$g(v_W, v_N) + g(v_E, v_N) = g(v_S, v_W) + g(v_S, v_E),$$

in which v_S, v_W and v_E are known, and we are solving for v_N . Detailed studies of the interaction problem for large interacting waves are carried out in [18, 15]. It is expected that by our methods of exactly describing large waves, we will be able to reproduce the Glimm-Lax theory for systems with

large data. Also, We believe our approach is suitable for larger classes of systems, a study of which has begun in [20, 17].

2. The system

We are interested in the Cauchy problem for the fully nonlinear wave equation

$$(2.1) \quad w_{tt} = (c^2(w_x))_x, \quad \text{with } w(x, 0) = w_0(x)$$

for $w : \mathbf{R} \times \mathbf{R}_+ \rightarrow \mathbf{R}$, and the data w_0 is large in some appropriate space.

It is convenient to write our wave equation in the equivalent first order form

$$(2.2) \quad \begin{pmatrix} v \\ u \end{pmatrix}_t + \begin{pmatrix} -u \\ p(v) \end{pmatrix}_x = 0,$$

where p is a C^1 function satisfying $-p'(v) = c^2(v)$. We shall refer to (2.2), which are the equations for isentropic gas dynamics in Lagrangian coordinates, as the p -system. Here u is the fluid velocity, $v = 1/\rho$ is specific volume, p is the pressure, and c is the sound speed. The detailed dynamics and effects of nonlinearity are determined by the pressure law $p = p(v)$ (or $p = \tilde{p}(\rho)$ where $\rho = 1/v$ is density).

In this paper, we make only the following general assumptions, which are justified on physical grounds. First, the pressure increases with increasing density, or decreases as a function of specific volume,

$$p'(v) < 0, \quad \text{so that } c(v) = \sqrt{-p'(v)}$$

is defined and positive for all v . Second, as the specific volume vanishes, the pressure becomes infinite,

$$(2.3) \quad p(v) \rightarrow \infty \quad \text{as } v \rightarrow 0,$$

so that infinite compression can only be obtained by applying infinite force. Our final assumption is that the pressure vanishes at vacuum, when the specific volume becomes infinite (or density vanishes), as does the sound speed,

$$\lim_{v \rightarrow \infty} p(v) = 0 \quad \text{so also } \lim_{v \rightarrow \infty} c(v) = 0.$$

It is occasionally convenient to assume that the pressure is convex, $p''(v) > 0$, but once the role of convexity has been established the condition is dropped with no penalty.

Our assumption $p'(v) < 0$ ensures that the system is everywhere hyperbolic, and the solution can thus be decomposed into superpositions of nonlinear waves. There are some examples in which the pressure may be non-monotone, as in a van der Waals gas or some models of nonlinear elasticity. These equations of state have been used to model phase transitions, although their physical basis is not yet completely understood. The equation is elliptic in regions in which the pressure is increasing, and states within

this region are unstable. Moreover, the Riemann problem may fail to have a unique solution near the elliptic region [11, 1].

2.1. Further examples. The p -system is the basic prototype for systems, analogous to Burgers' equation for scalar conservation laws. It is the simplest nonlinear system, and possesses features common to the most important physical examples. Indeed, it can be regarded as a subsystem of many larger systems, and as such is fundamental to their study.

The most obvious system of which it is a reduction is that of Lagrangian gas dynamics,

$$(2.4) \quad \begin{pmatrix} v \\ u \\ \frac{1}{2}u^2 + e \end{pmatrix}_t + \begin{pmatrix} -u \\ p \\ up \end{pmatrix}_x = 0,$$

where e is the internal energy, and thermodynamic principles yield relations

$$e = e(v, S), \quad p = p(v, S) \quad \text{and} \quad c^2 = c^2(v, S).$$

Here S is the entropy, which for smooth solutions satisfies the extra conservation law $S_t = 0$. The p -system is obtained by assuming the flow is isentropic (S constant) and dropping the last equation.

The p -system is equivalent to the Euler equations of isentropic gas dynamics [14], which is in turn a subsystem of the full 3×3 Euler equations. Apart from the obvious addition of an extra family of waves, the nonlinearity here appears as a function of two thermodynamic variables, which further complicates the analysis. Indeed, some structure needs to be assumed to obtain a unique solution to the Riemann problem [11]. On the other hand, it is well known that the third family is linearly degenerate and that it weakly decouples (through the equation for entropy), so it is reasonable to expect that solutions to this system will generally be closely related to that of the p -system. In particular, we expect solutions to 3×3 gas dynamics to be stable [19].

Another system which reduces to the p -system is that of nonlinear elasticity. By considering plane deformations in 3-D elasticity, one obtains the 6×6 system

$$(2.5) \quad \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}_t - \begin{pmatrix} \mathbf{v} \\ \mathbf{T}(\mathbf{u}) \end{pmatrix}_x = 0,$$

where \mathbf{u} and \mathbf{v} are vectors, and $\mathbf{T}(\mathbf{u})$ is the reduced stress-strain relation [7, 3]. A nonlinear string is modeled by the particular relation

$$\mathbf{T}(\mathbf{u}) = T(u) \hat{\mathbf{u}}, \quad \text{where} \quad \mathbf{u} = u \hat{\mathbf{u}}$$

is polar decomposition, and restricting to longitudinal motions of the string ($\hat{\mathbf{u}}$ constant) again yields the p -system [20, 17].

In the string system there are four wave families (two of which are degenerate), but the nonlinearity again appears only as a scalar function of a single variable, $T = T(u)$. Indeed, one can regard the evolution of the

'stretch' u in the full string system as satisfying a p -system with a collection of point sources on the right hand side [20].

We can similarly obtain (2.2) as a subsystem of many other physical systems, for example Maxwell's equations in a nonlinear medium, nozzle flow, and combinations of these, like MHD, etc. The point of view we adopt here is that we wish to extend our methods to these larger systems, so we develop methods that we expect will extend in a natural way. In particular, we avoid the use of Riemann invariants, and appeal to only one convex entropy as necessary. Previous studies of (2.2) rely either on Riemann coordinates or on infinitely many entropies [6, 2], which are unavailable for larger systems. Indeed, the use of physical variables leads to a better intuition for nonlinear wave phenomena.

3. The nonlinear waves

When written in quasilinear form, our system (2.2) becomes

$$(3.1) \quad \begin{pmatrix} v \\ u \end{pmatrix}_t + A(v) \begin{pmatrix} v \\ u \end{pmatrix}_x = 0,$$

with flux matrix

$$(3.2) \quad A(v) = \begin{pmatrix} 0 & -1 \\ p'(v) & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -c^2(v) & 0 \end{pmatrix}.$$

The eigenvalues of A are $\pm c$, with corresponding right eigenvectors $(1 \mp c)^t$. As such, the system is strictly hyperbolic, but degenerates at the vacuum. We will refer to those waves with positive speed $c > 0$ as forward waves, and those with negative speed as backward waves. The left eigenvectors are also easy to calculate: they are the gradients of the Riemann invariants

$$h(v) \pm u, \quad \text{where} \quad h(v) = \int c(v) dv.$$

Although we do not use the Riemann invariants, the integrated sound speed $h(v)$ will play a major part in our analysis.

3.1. Rarefactions. The rarefaction curves consist of those states that can be connected to a given fixed state by a centered rarefaction wave, and are the integrals of the eigenvectors [8, 13]. The state is constant along characteristics, which are the curves

$$\frac{dx}{dt} = \pm c(v(x, t)),$$

while across the characteristics we have

$$(3.3) \quad \frac{d}{d\epsilon} \begin{pmatrix} v \\ u \end{pmatrix} = \begin{pmatrix} 1 \\ \mp c(v) \end{pmatrix},$$

with $c(v(\epsilon)) = \pm x/t$. We use v as the parameter, and we get

$$(3.4) \quad u - u_0 = \mp \int_{v_0}^v c(v) dv.$$

The wavespeed must increase from left to right across the wave, so that the backward wave with left state $(v_0, u_0)^t$ must satisfy

$$(3.5) \quad -c(v_0) \leq -c(v) \leq -c(v),$$

for v between v_0 and v . In case $p(v)$ is convex, so $c'(v) < 0$, this yields $v_0 \leq v$. Similarly, the forward rarefactions (with p convex) are given by (3.4) with $v_0 \geq v$. In either case, the (absolute) sound speed c decreases as we traverse the wave from ahead state to behind state.

We note that this explicit solution of the rarefaction curves does not depend on the size of the wave $v - v_0$, and because (3.3) is autonomous, we can exactly piece together small rarefactions to get large waves.

3.2. Shocks. In general the wavespeed cannot always increase from ahead states to behind, and we expect shocks to form. Jump discontinuities satisfy the Rankine-Hugoniot conditions [8, 13], which in this case are

$$(3.6) \quad \begin{aligned} \sigma [v] &= -[u] \quad \text{and} \\ \sigma [u] &= [p(v)] \end{aligned}$$

where as usual, $[\cdot]$ denotes the jump in a quantity and σ is the shock speed, so a centered jump is located on the ray $x/t = \sigma$. We easily solve to get

$$(3.7) \quad \begin{aligned} u - u_0 &= -\sigma(v_0, v) (v - v_0), \quad \text{where} \\ \sigma(v_0, v) &= \pm \sqrt{\frac{p(v_0) - p(v)}{v - v_0}} \end{aligned}$$

which we again regard as being parameterized by v . In the case of convex pressure, we use the Lax entropy condition, which states that the characteristic speed be faster on the left,

$$-c(v_0) > -\sqrt{\frac{p(v_0) - p(v)}{v - v_0}} > -c(v),$$

which yields $v_0 > v$ for backward waves. Similarly, the forward shock curve is the branch of (3.7) with $v_0 < v$. In both cases, the sound speed c is larger behind the shock.

3.3. The convex case. We now combine the shock and rarefaction curves under the assumption that p is convex. It is convenient to define the function $g : \mathbf{R}^2 \rightarrow \mathbf{R}$ by

$$(3.8) \quad g(v_1, v_2) = \begin{cases} \int_{v_1}^{v_2} c(v) \, dv, & \text{for } v_1 \leq v_2, \\ -\sqrt{(p(v_2) - p(v_1))(v_1 - v_2)}, & \text{for } v_1 \geq v_2. \end{cases}$$

The forward and backward wave curves can then be described by

$$(3.9) \quad u_r - u_l = g(v_a, v_b),$$

where the subscripts l and r refer to the left and right states, while a and b refer to the states ahead of and behind the wave, respectively (so $a = l$

for a backward wave). As expected, the wave is a shock if the sound speed behind the wave is faster,

$$c(v_a) < c(v_b), \quad \text{that is } v_a \geq v_b,$$

while the opposite inequalities hold for rarefactions. Again we remark that this description holds exactly for waves having arbitrary strength.

3.4. Composite waves. For general pressure which is not convex, the simple combination of shocks and rarefactions is not enough to fully describe the Riemann problem. In this case we refer to Liu's entropy criterion [9], and to Oleinik's construction for scalar equations [10, 2]. Analyzing the change of state across the integral curve (3.4), we see that (3.5) is not satisfied in general. This necessitates introduction of a 'characteristic shock', which propagates with speed σ which is equal to the sound speed on one or both sides of the jump [9]. For backward jumps, Liu's entropy condition is

$$\sigma(v_0, \nu) \geq \sigma(v_0, v) \geq \sigma(\nu, v),$$

for ν between v_0 and v , where σ is given in (3.7).

The composite waves consist of rarefaction waves adjacent to characteristic shocks, pieced together in such a way that the wave speed increases continuously from left to right across the interior of the wave. If the wave has finite width, the extreme edges are one-sided shocks or characteristics.

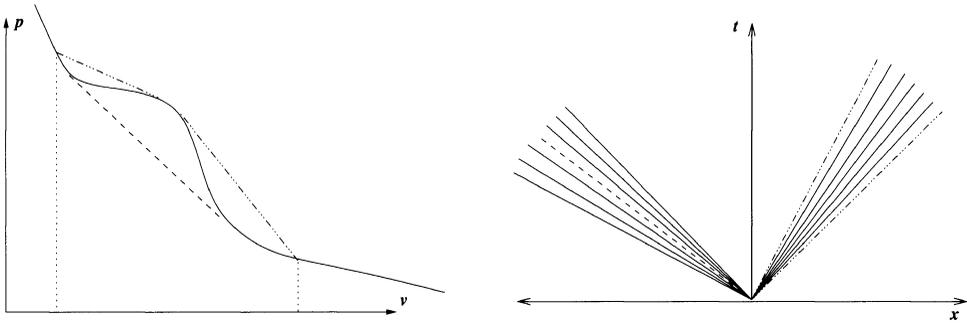


FIGURE 1. Composite wave

We now give a more geometric description of the composite wave, and relate it to Oleinik's construction of solutions for scalar nonconvex equations. Dafermos describes a similar construction in his book [2]. To do this, we recall the definition of the wave speeds,

$$(3.10) \quad -c^2(v) = p'(v) \quad \text{and} \quad -\sigma^2(v_0, v) = \frac{p(v_0) - p(v)}{v - v_0},$$

and observe that these are the slopes of the tangent and secant to the graph of $p(v)$, respectively, as in Fig. 1. Now the requirement that the absolute wavespeed increases across the composite wave from the behind state to the ahead state becomes monotonicity of the slope as the wave is traversed. As

the state varies across the wave, these slopes (which are squared wavespeeds) trace out the upper or lower convex envelope of the graph of p .

Moreover, referring to (3.8), for $v_1 \geq v_2$ we rewrite

$$g(v_1, v_2) = |\sigma(v_1, v_2)| (v_2 - v_1) = \int_{v_1}^{v_2} |\sigma(v_1, v_2)| d\nu,$$

where σ is given in (3.7). Here we again interpret $-\sigma^2$ as the slope of the upper convex envelope of p between v_1 and v_2 .

For general pressure law, it is now clear how we should define the function g : namely identify the correct convex envelope, determine the sound speed, and integrate. Thus given v_1 and v_2 , denote by p^\cap and p_\cup the upper and lower convex envelopes of p between v_1 and v_2 , respectively, and set

$$(3.11) \quad g(v_1, v_2) = \int_{v_1}^{v_2} \sqrt{-\hat{p}'(\nu)} d\nu = \int_{v_1}^{v_2} \hat{c}(\nu) d\nu,$$

where $\hat{p} = p^\cap$ (resp. $\hat{p} = p_\cup$) if $v_1 \geq v_2$ (resp. $v_1 \leq v_2$). Here the local sound speed $\hat{c}(v)$ is defined by

$$\hat{c}^2(v) = -\hat{p}'(v),$$

and our modified characteristics are given by

$$\frac{dx}{dt} = \pm \hat{c}(v).$$

Note that on the intervals that p agrees with \hat{p} , the composite wave will be a rarefaction, while if p and \hat{p} differ on an interval, then \hat{c} is the shock speed σ , which is constant on that interval. According to our construction, \hat{c} (and so also v) is monotone, and the composite wave fills a fan centered at the origin.

Similarly, across the composite wave, the change in state $(v \ u)^t$ satisfies

$$\frac{d}{d\epsilon} \begin{pmatrix} v \\ u \end{pmatrix} = \begin{pmatrix} 1 \\ \mp \hat{c}(v) \end{pmatrix},$$

which agrees with (3.4) and (3.7) across rarefactions and jumps, respectively. As before, we solve this equation to get

$$(3.12) \quad u - u_0 = \mp \int_{v_0}^v \hat{c}(\nu) d\nu,$$

where \hat{p} is either p^\cap or p_\cup , as appropriate. Once \hat{p} has been fixed, (3.12) is used to find the states inside the composite wave.

Again, we concisely describe both forward and backward waves by the relation

$$(3.13) \quad u_r - u_l = g(v_a, v_b),$$

where g is given by (3.11) and the subscripts refer to right, left, ahead and behind states, respectively. Recall that g depends on the choice of p^\cap or p_\cup , which in turn depends on whether $v_a \leq v_b$ or $v_a \geq v_b$, respectively.

We remark that on intervals that \hat{p} is different from p , the state $(v \ u)^t$ may change continuously, but this is not seen in the solution because the

modified characteristics coincide, all having speed $\pm\hat{c}$, which is constant because \hat{p} is affine on the interval. We thus see that this description of composite waves is a direct generalization of our earlier description (3.9), (3.4), (3.7). Again note that this construction is global in that it works for waves having arbitrary strength.

Our description of the forward and backward composite waves is a natural extension of Oleinik’s construction for nonconvex scalar conservation laws, and it clarifies the role that convexity plays in describing nonlinear waves. Indeed, when viewed from this point of view, we need not distinguish between shocks and rarefactions. In a future paper, we show that the global effect of nonlinearity can be quantified by the difference between the upper and lower convex envelopes of the pressure [15].

4. The Riemann problem

We now combine the above descriptions of forward and backward waves to solve the Riemann problem with arbitrary left and right states. Thus we are given two states $(v_l \ u_l)^t$ and $(v_r \ u_r)^t$, and we must identify a middle state $(v_* \ u_*)^t$, so that a backward wave joins $(v_l \ u_l)^t$ to $(v_* \ u_*)^t$, and a forward wave joins $(v_* \ u_*)^t$ to $(v_r \ u_r)^t$. To carry this out, we use the description of composite waves developed in the previous section.

According to (3.13), if $(v_l \ u_l)^t$ is joined to $(v_* \ u_*)^t$ by a backward wave, then we have

$$u_* - u_l = g(v_l, v_*),$$

while if $(v_* \ u_*)^t$ is joined to $(v_r \ u_r)^t$ by a forward wave, then

$$u_r - u_* = g(v_r, v_*).$$

Eliminating u_* , we get the equation

$$(4.1) \quad u_r - u_l = g(v_l, v_*) + g(v_r, v_*) \equiv \phi(v_*),$$

and we wish to solve for v_* . It is clear from (3.11) that, for fixed v_l and v_r , the function ϕ defined in (4.1) is differentiable and that

$$\phi'(v_*) = \sqrt{-\hat{p}'_l(v_*)} + \sqrt{-\hat{p}'_r(v_*)} > 0,$$

where \hat{p}_l and \hat{p}_r are appropriately chosen. Thus (4.1) can be uniquely solved provided $u_r - u_l$ is in the range of ϕ , which is clearly determined by that of g .

Fix v_0 , and first consider $v < v_0$. In this case $\hat{p} = p^\cap$, the upper convex envelope, and so \hat{p}' is decreasing on the interval $[v, v_0]$. Thus for $\nu \in [v, v_0]$, we have

$$\sqrt{-\hat{p}'(\nu)} \leq \sqrt{-\hat{p}'(v_0)},$$

and

$$\begin{aligned} \int_v^{v_0} \sqrt{-\hat{p}'(\nu)} \, d\nu &\geq \int_v^{v_0} \frac{-\hat{p}'(\nu)}{\sqrt{-\hat{p}'(v_0)}} \, d\nu \\ &= \frac{\hat{p}(v) - \hat{p}(v_0)}{\sqrt{-\hat{p}'(v_0)}}, \end{aligned}$$

and therefore

$$g(v_0, v) \leq \frac{p(v_0) - p(v)}{\sqrt{-\hat{p}'(v_0)}} \rightarrow -\infty \quad \text{as } v \rightarrow 0.$$

On the other hand, for $v > v_0$, we have $\hat{p} = p_{\cup}$, and

$$g(v_0, v) = \int_{v_0}^v \sqrt{-p_{\cup}'(\nu)} \, d\nu.$$

As $v \rightarrow \infty$, this integral may converge or diverge. If it diverges, then the range of the function ϕ is the entire real line, and (4.1) can always be solved. If the integral converges, which is more interesting physically, then the range of ϕ is bounded from above,

$$\phi(v_*) < g(v_l, \infty) + g(v_r, \infty).$$

We have thus proved the following theorem.

THEOREM 1. *Given constant left and right states $(v_l \ u_l)^t$ and $(v_r \ u_r)^t$, respectively, there is a unique solution to the corresponding Riemann problem, provided that the condition*

$$(4.2) \quad u_r - u_l < g(v_l, \infty) + g(v_r, \infty)$$

holds. Moreover, the intermediate state $(v_ \ u_*)^t$ (and hence wave strength) is a C^1 function of the data.*

The fact that the intermediate states are C^1 functions of the data follows in the standard way by the implicit function theorem. It is clear that g (and so also ϕ) is C^1 , and indeed g will be C^2 (but not C^3) if p is, as is well known. Note that even if the pressure p is not assumed to be differentiable, its convex envelope is absolutely continuous, and so we would gain some regularity in g .

4.1. The vacuum. Physically, if the velocity difference $u_r - u_l$ is large, particles move away from each other and the gas rarefies in between the two outgoing waves. However, if this difference is so large that the one-sided condition (4.2) is violated, then there is not enough gas ‘between’ the left and right states, and a vacuum develops. Due to scale invariance, this vacuum develops immediately and is present for all positive times. We extend our solution to include the vacuum, as follows.

The vacuum corresponds to infinite specific volume, or equivalently vanishing density, pressure or sound speed. In our self-similar solution for the

Riemann problem, the vacuum will lie on the positive t -axis, *i.e.* along $x = 0$. For $x \neq 0$, and positive times, the specific volume should be given by

$$\pm \hat{c}(v(x, t)) = x/t,$$

which is our equation for characteristics, and the velocity $u(x, t)$ should be given by integrating the wavespeed

$$\hat{c}(v) = \sqrt{-p_{\mathcal{U}}'(v)},$$

given by the lower convex envelope.

We thus suppose that (4.2) is violated, and construct a solution containing the vacuum as follows. For $x < 0$, the gas rarefies along a (composite) backward wave, the volume and velocity is given in the wedge

$$-\hat{c}(v_l) t \leq x < 0$$

by

$$(4.3) \quad \begin{aligned} \hat{c}(v(x, t)) &= -x/t \quad \text{and} \\ u(x, t) - u_l &= g(v_l, v(x, t)) = \int_{v_l}^{v(x, t)} \hat{c}(v) \, dv. \end{aligned}$$

For fixed $t > 0$, we therefore have

$$v_x = \frac{-1}{t \hat{c}'} > 0 \quad \text{and} \quad u_x = \hat{c} v_x > 0,$$

so both specific volume and velocity increase as functions of $x < 0$.

Similarly, the forward wave is given in the wedge $0 < x \leq \hat{c}(v_r) t$ by

$$(4.4) \quad \begin{aligned} \hat{c}(v(x, t)) &= x/t, \quad \text{and} \\ u_r - u(x, t) &= g(v_r, v(x, t)) = \int_{v_r}^{v(x, t)} \hat{c}(v) \, dv, \end{aligned}$$

so that

$$v_x = \frac{1}{t \hat{c}'} < 0 \quad \text{and} \quad u_x = -\hat{c} v_x > 0,$$

and the specific volume decreases while the velocity increases for $x > 0$.

On the t -axis $x = 0$, the velocity has a jump with left and right limits given by

$$(4.5) \quad \begin{aligned} u(0-, t) &= u_- = u_l + g(v_l, \infty) \quad \text{and} \\ u(0+, t) &= u_+ = u_r - g(v_r, \infty) \end{aligned}$$

respectively. Note that because (4.2) fails, we always have

$$u_- = u(0-, t) \leq u(0+, t) = u_+,$$

and so u is monotone increasing and bounded as a function of x .

However, the specific volume is infinite, and to take into account the jump in u , we take v to be a Radon measure, while the other variables remain bounded. This measure is singular only at the vacuum, so its singular part

is supported on the t -axis $x = 0$. Since v must be locally integrable, this singular part must be of the form

$$(4.6) \quad \nu_S = w(t) \delta(x).$$

Since the velocity u is defined (a.e.) above, we can find the weight $w(t)$ by solving the equation

$$v_t - u_x = 0$$

in the sense of distributions. On the t -axis, this reduces to

$$\frac{dw}{dt} = u_+ - u_-,$$

so that the singular part of $v(x, t)$ is the measure

$$(4.7) \quad \nu_S = (u_+ - u_-) t \delta(x).$$

Note that ν_S is self-similar since $\delta(\alpha x) = \delta(x)/\alpha$.

Although the continuous part of v is unbounded, it is locally integrable. Indeed, for t fixed,

$$(4.8) \quad \begin{aligned} \int_{-\hat{c}(v_l)t}^0 v(x, t) dx &= x v(x, t) \Big|_{-\hat{c}(v_l)t}^0 - \int_{v_l}^{\infty} x dv \\ &= -t \hat{c} v \Big|_{v_l}^{\infty} + t \int_{v_l}^{\infty} \hat{c} dv, \end{aligned}$$

where we have integrated by parts and used (4.3). This last integral is bounded, and since \hat{c} is decreasing as a function of $v > v_0$, for large $V > v_0$ we have

$$\begin{aligned} \hat{c}(V) V &\leq \hat{c}(v_l) v_l + \hat{c}(V) (V - v_l) \\ &\leq \hat{c}(v_l) v_l + \int_{v_l}^V \hat{c} dv \\ &\leq \hat{c}(v_l) v_l + g(v_l, \infty), \end{aligned}$$

so the right hand side of (4.8) is bounded. The integral across the forward wave is similarly bounded, and since ν_S has mass $(u_+ - u_-) t$, we have proved the following theorem.

THEOREM 2. *Given states $(v_l, u_l)^t$ and $(v_r, u_r)^t$ such that (4.2) fails, there is a unique self-similar solution $(v(x, t), u(x, t))^t$ to the Riemann problem, where the velocity*

$$u(x, t) \in L^\infty \cap BV \cap L^1_{loc}$$

is a bounded monotone increasing function, while $v(x, t)$ is a Radon measure whose singular part is the Dirac measure (4.7) and whose absolutely continuous part is locally integrable. Moreover this solution is Lipschitz continuous in time as a distribution in L^1_{loc} .

We compare our solution to that of the Euler equations of isentropic gas dynamics,

$$\begin{pmatrix} \rho \\ \rho u \end{pmatrix}_t + \begin{pmatrix} \rho u \\ \rho u^2 + u p(\rho) \end{pmatrix}_X = 0,$$

where X is the spatial coordinate. Equivalence of this system to the p -system is given by the map

$$x = \int^X \rho(\chi, t) d\chi = \int^X \frac{d\chi}{v(\chi, t)},$$

so that x is correctly interpreted as a material coordinate. Weak solutions of the two systems are shown to be equivalent in [14], provided care is taken at the vacuum.

We compare our solution to those constructed in [12] which also contain a vacuum. At the vacuum, the map $X \rightarrow x$ from Eulerian to Lagrangian coordinates is no longer 1-1, and so the inverse map $x \rightarrow X$ from Lagrangian back to Eulerian coordinates is not well defined. Since the density ρ vanishes, the solution of the Euler equations remains in L^∞ . The weight $w(t)$ on the delta-function in Lagrangian coordinates corresponds to the width of the vacuum in Eulerian coordinates. According to our solution, the vacuum thus has width $(u_+ - u_-)t$ in Eulerian coordinates. This agrees with the solution obtained directly (for convex p) in [12, 11].

5. Concluding remarks

Although our pressure law is not convex, our results are anticipated by the known solutions of the Riemann problem for Eulerian gas dynamics, together with the equivalence of Eulerian and Lagrangian formalisms established in [14]. Our interest lies not in the actual solution of the Riemann problem, as much as in the elementary description thereof in terms of the scalar function $g(v_1, v_2)$. In a forthcoming paper, we will use this function to exactly analyze wave interactions and obtain global bounds for solutions to the Cauchy problem [15].

Our solution of the Riemann problem for arbitrary data is a step towards the well-posedness of the Cauchy problem for the p -system with arbitrary data. In order to get results with large data, it is necessary to understand the Riemann problem and interactions of finite waves exactly, rather than asymptotically, as has been done in the past. This is intractable for general systems, and it is essential to use the structure of the equations as given by the physics. Indeed, many pathological phenomena have been found for Riemann problems with large data, even for physical systems. This is generally because the Hugoniot locus may have complicated global geometry [11].

In the case of the p -system, the Hugoniot curves are well-behaved and allow us to express all waves in the simple functional form (3.13). Indeed, we can describe a general interaction of two waves of arbitrary strength: by resolving the left and right states in terms of both the incoming and outgoing waves, and eliminating the velocities, we get a single equation for

the interaction. Thus, the interaction of a forward and backward wave is described by

$$(5.1) \quad g(v_s, v_w) + g(v_s, v_e) = g(v_w, v_n) + g(v_e, v_n),$$

which is a single equation for the single variable v_n . This equation is exact for waves of arbitrary strength. Using such expressions, we can analyze many interactions and get bounds for general solutions which will allow us to obtain convergence and existence of solutions [15].

Equation (5.1) is not always solvable near the vacuum, in which case v_n would be infinite and some δ -function generated. However, it turns out that the vacuum does not form in finite time, so this difficulty can be resolved. Indeed, the vacuum can be characterized by vanishing sound speed \hat{c} , and as this sound speed gets smaller, so the time during which the interaction takes place becomes larger, and in the limit becomes infinite [18].

References

- [1] Fumioki Asakura, *Stability of Maxwell states in thermo-elasticity*, preprint, 2000.
- [2] Constantine M. Dafermos, *Hyperbolic conservation laws in continuum physics*, Springer, 2000.
- [3] W. Domański and R. Young, *Interactions of plane waves in nonlinear elasticity*, In preparation.
- [4] K.O. Friedrichs and P.D. Lax, *Systems of conservation laws with a convex extension*, Proc. Nat. Acad. Sci. **68** (1971), 1686–1688.
- [5] J. Glimm, *Solutions in the large for nonlinear hyperbolic systems of equations*, Comm. Pure Appl. Math. **18** (1965), 697–715.
- [6] J. Glimm and P.D. Lax, *Decay of solutions of systems of nonlinear hyperbolic conservation laws*, Memoirs Amer. Math. Soc. **101** (1970).
- [7] Fritz John, *Formation of singularities in one-dimensional nonlinear wave propagation*, Comm. Pure Appl. Math. **27** (1974), 377–405.
- [8] P.D. Lax, *Hyperbolic systems of conservation laws, II*, Comm. Pure Appl. Math. **10** (1957), 537–566.
- [9] Tai-Ping Liu, *Admissible solutions of hyperbolic conservation laws*, Memoirs Amer. Math. Soc. **240** (1981).
- [10] O. Oleinik, *On the uniqueness of the generalized solution of a Cauchy problem for a nonlinear system of equations occurring in mechanics*, Usp. Mat. Nauk. **73** (1957), 169–176, In Russian.
- [11] R.G. Smith, *The Riemann problem in gas dynamics*, Trans. Amer. Math. Soc. **249** (1979), 1–50.
- [12] J. Smoller, *On the solution of the Riemann problem with general step data for an extended class of hyperbolic systems*, Michigan Math. J. **16** (1969), 201–210.
- [13] ———, *Shock waves and reaction-diffusion equations*, Springer-Verlag, New York, 1982.
- [14] D. Wagner, *Equivalence of the Euler and Lagrangian equations of gas dynamics for weak solutions*, Jour. Diff. Equations **68** (1987), 118–136.
- [15] Robin Young, *The p -system with large data*, In preparation.
- [16] ———, *On elementary interactions for hyperbolic conservation laws*, Preprint, 1993.
- [17] ———, *Nonstrictly hyperbolic waves in elasticity*, J. Diff. Eq. (2001), submitted.
- [18] ———, *The p -system II: The vacuum*, Nonlinear Evolution Equations, Warsaw, 2001, submitted.
- [19] ———, *Sustained solutions for conservation laws*, Comm. PDE **26** (2001), 1–32.

- [20] ———, *Wave interactions in nonlinear elastic strings*, Arch. Rat. Mech. Anal. (2001), in press.

DEPARTMENT OF MATHEMATICS & STATISTICS, UNIVERSITY OF MASSACHUSETTS
E-mail address: `young@math.umass.edu`

Statistical analysis of collision-induced timing shifts in a wavelength-division-multiplexed optical soliton-transmission system

G.J. Morrow and S. Chakravarty

ABSTRACT. A statistical analysis of optical soliton collisions in a two-channel wavelength-division multiplexed transmission system is performed. It is shown that the variance of the collision-induced timing shifts due to interactions between pulses in the adjacent channel can grow significantly with the length of the transmission line due to a certain resonance effect. In addition, a statistical model is developed which takes into account the effect on the main channel timing shifts arising from the interference due to four-wave mixing components generated during a two-channel collision.

1. Introduction

Considerable progress has been made in recent years in the field of fiber-optic long-distance soliton data transmission with wavelength-division multiplexing (WDM). When compared to conventional single-channel systems, WDM offers the potential for a large increase in the total capacity of data transmission. Indeed in recent experiments trans-oceanic data transmission of over 1 Tbit/s on each fiber has been realized [7, 6]. However, the use of WDM raises a number of important issues. For example, due to the periodic distribution of amplifiers, a resonant instability created by the four-wave mixing (FWM) interactions can seriously degrade the signal. However, proper use of dispersion management can alleviate the negative effects of FWM to a certain extent. This issue was studied in Refs. [9, 1] for the classical solitons and more recently in Refs. [8, 4] for the dispersion-managed solitons. Another serious problem that arises in WDM soliton systems is caused by the permanent frequency shifts and the associated displacements in pulse arrival times created by the nonuniform soliton collisions in the presence of periodic loss and amplification [11]. For classical solitons a statistical theory was developed in Refs. [10, 3] to account for the cumulative effects of the timing shifts due to the large number of collisions taking place in a long-distance transmission link, and explicit expressions for the RMS timing jitter and bit error rate of a two-channel

1991 *Mathematics Subject Classification.* Primary 35Q55, 35Q60; Secondary 78A60.

Key words and phrases. Optical solitons, Timing jitter, Four-wave mixing.

The second author was partially supported by a Grigsby-Sensenbaugh Foundation grant from the University of Colorado.

WDM system were calculated for both with and without weak dispersion management. Recently, using a similar approach a statistical analysis of collision-induced timing jitter in a two-channel WDM system with dispersion management was reported in Ref. [12]. However, in addition to the impairment caused by the timing jitter between the main channels in a practical WDM system, it is also necessary to consider the effect of the FWM components colliding with the main channels. Even though an individual collision event may impart a very small temporal displacement to the main channel pulses, it is possible that the accumulated effect of a significantly large number of such collisions on a main channel pulse may not be negligible in a high-capacity, multi-channel WDM system. To our knowledge, no comprehensive statistical analysis has been undertaken which takes into account the large number of collisions between the FWM components and the main channel pulses that occur in the fiber. Furthermore, no analytical expressions are known for the collision-induced timing jitter due to such interactions.

In this paper we study the statistical properties of the timing jitter induced by the main channel collisions as well as the collisions between a main channel and the FWM pulses in a two-channel WDM system. The purpose for revisiting the statistical theory of timing jitter due to adjacent channel collisions is that we find a correction term in the root mean square timing jitter that was neglected in previous studies. The correction term is especially dominant if the average distance between successive collisions is an integer multiple (or very close to it) of the amplifier spacing. In order to simplify the discussion and to focus primarily on the statistical aspects of timing jitter we consider the case of classical soliton pulses in the presence of loss and amplification only. However the statistical framework developed here is equally applicable to any WDM system which includes the effects of dispersion management, frequency filtering or any other in-line control mechanisms.

The paper is planned as follows. Section 2 consists of background material based on previously known results. Here we briefly discuss the set up for a two-channel WDM system in the presence of loss and periodic amplification. Then we derive the frequency and timing shift formulas for a two-soliton collision with large channel spacing using standard soliton perturbation theory. In section 3, we perform a statistical analysis of the cumulative effect of the large number of collisions between the main channel pulses in a two-channel transmission line. We give explicit formulas for the mean and the root mean square timing jitter which exhibits a previously unnoticed resonance phenomenon. Finally in section 4, we set up a statistical model to estimate the net timing jitter of a main pulse due to its interaction with the large number of FWM components produced in a two-channel high bit-rate transmission link.

2. Timing shifts for collisions between main channels

In this section we recall the analytical results for the timing shift resulting from a single collision between two solitons in adjacent channels in a fiber with loss and amplification. The dynamics of the dimensionless field amplitude q is governed by the perturbed nonlinear Schrödinger equation (NLS) with damping and amplification:

$$(2.1) \quad iq_z + (1/2)D(z)q_{tt} + |q|^2q = iP(z)q,$$

where $P(z)$ is the periodic damping/amplification term to be described below, and the function $D(z)$ describes the particular choice of dispersion. The variables z and t are the usual non-dimensional space and retarded time, normalized to the dispersion length z_* and the characteristic time t_* : $z_* = 2\pi ct_*^2/(\lambda^2 \bar{D})$ and $t_* = \tau/1.763$, respectively, where $\lambda = 1.550 \mu\text{m}$ is the central wavelength, \bar{D} is the average dispersion parameter, τ is the full width at half maximum of the pulse intensity and c is the speed of light in vacuum. The typical values of various physical quantities are $\tau = 20 \text{ ps}$, $\bar{D} = 0.5 \text{ ps}/(\text{nm} \cdot \text{km})$ and $z_* = 201.94 \text{ km}$.

The effects of damping and amplification are described by taking $P(z)$ to be

$$(2.2) \quad P(z) = -\Gamma + (e^{\Gamma z_a} - 1) \sum_{n=-\infty}^{\infty} \delta(z - nz_a),$$

where $\Gamma = \gamma z_*$ is the dimensionless damping coefficient and $z_a = l_a/z_*$ is the dimensionless amplifier spacing, while $\delta(z)$ is the Dirac delta function. Typical experimental values for γ and l_a are $2\gamma = 0.20 \text{ dB/km} = 0.046 \text{ km}^{-1}$ and $l_a = 25 \text{ km}$. For $\tau = 20 \text{ ps}$ these values yield $\Gamma = 4.62$ and $z_a = 0.12$. As usual we rescale the field amplitude as $q(z, t) = [g(z)]^{1/2}u(z, t)$, where the function $g(z)$ denotes the periodic energy gain/loss cycle and varies on the length scale of the amplifier distance—which is small compared to the dispersion distance, since $z_a \ll 1$. That is, $g(z)$ is the periodic function

$$(2.3) \quad g(z) = a_0^2 \exp[-2\Gamma(z - nz_a)], \quad nz_a \leq z < (n + 1)z_a.$$

The quantity a_0^2 is chosen so that the average of $g(z)$ is unity over an amplification cycle; i.e.,

$$a_0^2 = \frac{2\Gamma z_a}{1 - \exp(-2\Gamma z_a)}.$$

With the substitution $q(z, t) = [g(z)]^{1/2}u(z, t)$ eq. (2.1) becomes a perturbed NLS equation

$$(2.4) \quad iu_z + \frac{D(z)}{2}u_{tt} + g(z)|u|^2u = 0.$$

The unperturbed NLS equation corresponds to the ideal case (lossless, constant dispersion): $g(z) = D(z) = 1$. In this article, we will only consider the case of constant dispersion, and set $D(z) \equiv 1$ in in eq. (2.4). We note that a weakly dispersion managed system can be handled in a similar fashion by introducing the transformed coordinate $\zeta(z) = \int_0^z D(z) dz$ so that eq. (2.4) becomes $iu_\zeta + (1/2)u_{t\zeta\zeta} + G(\zeta(z))|u|^2u = 0$, where $G(\zeta(z)) = g(z)/D(z)$ and $D(z)$ is normalized so that its average $\langle D(z) \rangle = 1$ over one amplification period. However, we will not pursue dispersion-management in this article. Instead we simply restrict our discussion to the constant dispersion $D(z) \equiv 1$ case.

When the solitons are widely separated in frequency space, to leading order we decompose u as $u(z, t) \sim u_1(z, t) + u_2(z, t) + O(\epsilon)$ (cf. Refs. [5, 2]), where

$$(2.5) \quad u_j = A_j \operatorname{sech} S_j \exp(i\chi_j), \quad S_j = A_j[t - \Omega_j(z - z_o)], \quad \chi_j = \Omega_j t - (\Omega_j^2 - A_j^2)z/2,$$

with S_j chosen so that two solitons collide at $z = z_o$. We assume A_j and Ω_j to be slowly varying functions of z with respect to the characteristic amplification period z_a . We also set initially $\Omega_2 = \Omega$ and $\Omega_1 = 0$, so that the dimensionless frequency separation is Ω . We also assume $\epsilon = 1/\Omega \ll 1$.

Next, by using soliton perturbation theory we obtain the variation (slow with respect to the amplifier period) of the quantities A_j and Ω_j . Then, using the explicit expressions for $u_{1,2}$ from eq. 2.5 and requiring that to leading order the A_j be constant (normalized to 1) yields the following equation for Ω

$$(2.6) \quad \frac{d\Omega}{dz} = g(z) \int_{-\infty}^{\infty} |u_1|^2 \frac{\partial |u_2|^2}{\partial t} dt = g(z) \frac{d}{dz} f(z - z_0),$$

where $f(z) = [\Omega z \cosh(\Omega z) - \sinh(\Omega z)] / \Omega \sinh^3(\Omega z)$ expresses the frequency shift in an ideal collision. Here we compute the frequency shift for the slower soliton u_1 ; the frequency shift for the soliton u_2 is the negative of that of u_1 and can be obtained by interchanging u_1 and u_2 in the integral appearing in eq. (2.6). Equation (2.6) can be integrated exactly to obtain the collision-induced frequency shift:

$$(2.7) \quad \Delta\Omega(z) = \int_{-\infty}^z g(z') \frac{d}{dz'} f(z' - z_0) dz'.$$

Note that the residual frequency shift defined as $\Delta\Omega(\infty)$ is given by $\Delta\Omega(\infty) = \int g(z) f'(z - z_0) dz$. The frequency shift $\Delta\Omega(z)$ due to collision at $z = z_0$ induces a net shift in the pulse arrival time at a given position L in the fiber given by

$$(2.8) \quad \delta t(z_0) = \int_{-\infty}^L \Delta\Omega(z) dz = \int_{-\infty}^L dz \int_{-\infty}^z dz' g(z') \frac{d}{dz'} f(z' - z_0).$$

The relevant integrals in eq. (2.8) can be readily calculated. To a good approximation we can take Ω to be constant in the integrals. Since $g(z)$ is periodic, all integrals involving $g(z)$ can be written in terms of Fourier series. The Fourier coefficients are

$$(2.9) \quad g_m = \frac{a_0^2}{z_a} \int_0^{z_a} e^{-2\Gamma z + 2m\pi iz/z_a} dz = \frac{\Gamma z_a}{\Gamma z_a - m\pi i}, \quad m = 0, \pm 1, \pm 2, \dots$$

Next we substitute eq. (2.3) into eq. (2.8) and use eq. (2.9), together with the integral

$$\int e^{2n\pi iz/z_a} f(z) dz = (n^2 \pi^4 / 4\Omega^4 z_a^2) \operatorname{csch}^2(n\pi^2 / 2\Omega z_a).$$

In this way, after rearranging the terms in the summation, we can express the dominant contribution to $\delta t(z_0)$ for $L \gg 1$ as a sine Fourier series:

$$(2.10) \quad \delta t(z_0) = (L - z_0) \sum_{m=1}^{\infty} a_m \sin[2m\pi z_0/z_a - \phi_m],$$

where

$$a_m = c_m |g_m|, \quad \phi_m = \arg[g_m], \quad c_m = (\pi^5 / \Omega^4 z_a^3) m^3 \operatorname{csch}^2(m\pi^2 / 2\Omega z_a),$$

and where

$$|g_m| = \Gamma z_a / \sqrt{(\Gamma z_a)^2 + (m\pi)^2}, \quad \arg[g_m] = \arctan[m\pi / (\Gamma z_a)].$$

Note that the timing shift depends on the position of the collision point z_0 , a fact that is important when computing the statistical averages. In the next section we discuss a statistical theory for computing the mean and RMS timing jitter in a WDM soliton system.

3. Statistical analysis of timing shifts

The transmission of binary data is accomplished by sending trains of soliton pulses in each channel of the 2-channel WDM system considered here. We assume independent, random binary transmission sequences in each of channels 1 and 2, where the presence of a pulse is denoted by a 1 and the absence of a pulse by a 0. We denote the sequence of pulses in channel 1 by $\{\alpha_j\}_{j=1}^\infty$ and in channel 2 by $\{\beta_j\}_{j=1}^\infty$. Thus we have two independent sequences of independent Bernoulli (0 and 1 valued) random variables that have a common fair-coin-tossing distribution for all j : $P(\alpha_j = 1) = P(\beta_j = 1) = 1/2$. It is assumed that the bit period T is the same for the data stream in each channel. Hence the binary data streams in channels 1 and 2 can be represented respectively as

$$(3.1) \quad \sum_{k=1}^\infty \alpha_k u_1(t + kT, z) \quad \text{and} \quad \sum_{k=1}^\infty \beta_k u_2(t + kT + t_0, z)$$

where $u_1(t, z)$ and $u_2(t, z)$ are given by eq. (2.5). We assume also that the initial time delay t_0 between the two channels is a random variable which is uniformly distributed on $[0, T]$.

For the sake of concreteness, let us consider the first non-zero pulse in the slower channel 1 to be a reference pulse. By re-indexing if necessary we can always label it as α_1 . (Note that here and in the following we refer without ambiguity to the various pulses by their designations as binary data). The information bit β_1 in the faster channel 2 will interact with the reference pulse α_1 at a location z_1 given in terms of the initial time delay as $z_1 = t_0/\Omega$. Note that since t_0 is random (within 1 bit period), the collision site z_1 is also a random variable. Once the location z_1 is given, any two subsequent adjacent bits β_k and β_{k+1} will interact with the reference pulse α_1 at equally spaced locations z_k and z_{k+1} respectively along the fiber. The distance between the successive interactions is given by $z_{k+1} - z_k := z_s = T/\Omega$ where T is the bit period and Ω is the channel spacing. The total number of collisions in the fiber can be estimated as $N = L/z_s$, where L is the non-dimensional system length (in physical units $L \sim 10^4\text{km}$). Thus the site z_k where the interaction between β_k and α_1 takes place can be expressed as

$$(3.2) \quad z_k = [U + (k - 1)]z_s, k = 1 \dots N,$$

where U is a uniformly distributed random variable on $[0, 1]$. Then the total timing shift imparted to the reference pulse α_1 by the β_k in channel 2 is the sum of all the individual shifts due to collisions occurring over the total length L of the system:

$$(3.3) \quad \Delta t = \sum_{k=1}^N \beta_k \delta t(z_k).$$

Sometimes the expression for Δt is called the absolute timing jitter in contrast to the relative timing jitter studied for main channel collisions in Refs. [3, 12]. In eq. (3.3) we treat β_k and $\delta t(z_k)$ as independent random variables. For β_k the randomness stems from the arbitrary encoding of data in the channels as mentioned earlier; for $\delta t(z_k)$ it comes from the arbitrariness of the location of the first collision point z_1 depending on the initial time delay between the bit streams in the two channels. That is, each collision is characterized by a random value of z_k as given in eq. (3.2). As a result, the values of $\delta t(z_k)$ will also be random.

First we calculate the mean of the total timing jitter Δt . By the independence of the random variables β_k and z_k and the fact that the mean $\langle \beta_k \rangle = 1/2$, we have

$\langle \beta_k \delta t(z_k) \rangle = \langle \beta_k \rangle \langle \delta t(z_k) \rangle = \langle \delta t(z_k) \rangle / 2$. Hence the mean timing shift is given by

$$\langle \Delta t \rangle = \frac{1}{2} \sum_{k=1}^N \langle \delta t(z_k) \rangle = \frac{1}{2} \sum_{k=1}^N \int_0^1 \delta t[(U + k - 1)z_s] dU$$

which is to be calculated by using the expression of $\delta t(z_k)$ from eq. (2.10) (with z_o replaced by z_k). To illustrate the statistical computations involved we choose only one Fourier component from the infinite sum in eq. (2.10) in our subsequent analyses. That is, we simply take

$$\delta t(z_k) = (L - z_k) a_m \sin[2m\pi z_k / z_a - \phi_m]$$

corresponding to the m^{th} Fourier mode. We emphasize that our choice is motivated by simplicity since the method developed here can be easily extended to the full Fourier series in eq. (2.10).

Substituting the (simplified) formula for $\delta t(z_k)$ into the equation for $\langle \Delta t \rangle$ above, changing the integration variable inside the sum and using the relation $L = Nz_s$, we obtain

$$\langle \Delta t \rangle = \frac{z_s a_m}{2} \sum_{k=1}^N \int_{k-1}^k (N - x) \sin(2\pi \rho x - \phi_m) dx = \frac{z_s a_m}{2} \int_0^N (N - x) \sin(2\pi \rho x - \phi_m) dx,$$

where $\rho = mz_s / z_a$. The last integral can be explicitly evaluated to yield the following expression for the mean timing jitter:

$$(3.4) \quad \langle \Delta t \rangle = \frac{z_s a_m}{4\pi^2 \rho^2} [2\pi N \rho \cos \phi_m - \sin(2\pi N \rho - \phi_m) + \sin \phi_m].$$

Note that the mean value of the timing jitter is nonzero; in fact we find (from the first term in eq. (3.4)) that $\langle \Delta t \rangle \sim O(N)$ so that the mean timing jitter grows linearly with system length L .

We next calculate the mean square of the total timing jitter $\langle (\Delta t)^2 \rangle$ and then the variance $var(\Delta t) = \langle (\Delta t)^2 \rangle - \langle \Delta t \rangle^2$. From eq. (3.3), we note that Δt is a sum of products of pairwise independent random variables. So $\langle (\Delta t)^2 \rangle$ can be expanded as a double sum of the expectation values of quadratic terms as follows:

$$\begin{aligned} \langle (\Delta t)^2 \rangle &= \sum_{k=1}^N \sum_{l=1}^N \langle \beta_k \beta_l \delta t(z_k) \delta t(z_l) \rangle = \sum_{k=1}^N \sum_{l=1}^N \langle \beta_k \beta_l \rangle \langle \delta t(z_k) \delta t(z_l) \rangle \\ &= \frac{1}{2} \sum_{k=1}^N \langle \delta t(z_k)^2 \rangle + \frac{1}{2} \sum_{\substack{k=1, \\ k < l}}^N \sum_{l=1}^N \langle \delta t(z_k) \delta t(z_l) \rangle \end{aligned}$$

where we have used the standard facts $\langle \beta_k \beta_l \rangle = \langle \beta_l \beta_k \rangle = 1/4, k \neq l$ and $\langle \beta_k^2 \rangle = 1/2$ for a Bernoulli random variable. In order to compute the remaining expectations involving the uniformly distributed random variable z_k we use eq. (3.2) in above to compute the sum over k . After making appropriate change of variables, we obtain

$$\langle (\Delta t)^2 \rangle = \frac{1}{2} \int_0^N \delta t(x)^2 dx + \frac{1}{2} \sum_{j=1}^{N-1} \int_0^N \delta t(x) \delta t(x + j) dx$$

where $\delta t(x) = z_s a_m (N - x) \sin(2\pi \rho x - \phi_m)$ and $\rho = mz_s / z_a$ as defined earlier. After evaluating the above integrals and using eq. (3.4) for $\langle \Delta t \rangle$ we finally obtain the variance of timing jitter

$$(3.5) \quad var(\Delta t) = \frac{z_s^2 a_m^2}{24} [2N^3 + \sum_{j=1}^{N-1} (N - j)^2 (2N + j) \cos(2\pi \rho j) + O(N^2)]$$

The first two terms in eq. (3.5) are the dominant contributions to the variance for large system length (equivalently for large number of collisions, $N \gg 1$). There are two cases to be distinguished here.

Non-resonant case: When ρ is sufficiently away from an integer value, the sum of the cosines in the second term of eq. (3.5) is small ($O(1)$) compared to N . The variance in this case turns out to be

$$(3.6) \quad \text{var}(\Delta t)_{nr} = \frac{z_s^2 a_m^2}{24} [N^3 + O(N^2)]$$

and the root mean square timing jitter $(\Delta t)_{rms} \sim O(L^{3/2})$ which is consistent with earlier results.

Resonant case: If ρ is equal or close to an integral value, $\cos(2\pi\rho j) \approx 1$ and the sum of the cosines becomes $O(N)$. Specifically, when ρ is exactly an integer,

$$(3.7) \quad \text{var}(\Delta t)_r = \frac{z_s^2 a_m^2}{32} [N^4 + O(N^3)].$$

The root mean square timing jitter $(\Delta t)_{rms} \sim O(L^2)$, so it grows considerably faster than the non-resonant case for large system length. For a high data rate transmission system ($T \ll 1$) with large to moderate channel spacing Ω , it is possible that $\rho = mT/\Omega z_a \approx 1$ is close to the resonance condition even for the first Fourier mode ($m = 1$) (cf. eq. (2.10)). In such a situation, the variance may be unusually large.

4. Statistical model for FWM-main channel collisions

In addition to causing permanent frequency shifts, collisions between the soliton pulses in adjacent channels give rise to spurious frequency components known as four-wave mixing (FWM) sidebands. In the lossless case, the four-wave contributions grow temporarily from a vanishing background during soliton interactions, and then decay back to zero when the solitons re-emerge from the collision. However, when loss and amplification are included in the theory, the amplitude of the FWM terms grows to become an order of magnitude larger than in the lossless case, and it saturates at its maximum value after the collision is completed. In Ref. [1] the growth and saturation of four-wave mixing components in the presence of periodic damping and amplification was studied analytically. It was shown that the final amplitude and the frequency location of the four-wave mixing terms depend on the underlying periodicity of the amplification process and on the frequency separation between channels. In particular, for certain combinations of amplifier spacing and channel spacing, a resonance appears.

In this section we develop a statistical model for the timing shifts of the main channel pulses due to their interactions with the four-wave-mixing (FWM) components. The collision between a pulse from channel 1 with frequency $\Omega_1 = 0$ with a pulse from channel 2 with frequency $\Omega_2 = \Omega$ generates two FWM components with center frequencies $\Omega_{112} = 2\Omega_1 - \Omega_2 = -\Omega$ and $\Omega_{221} = 2\Omega_2 - \Omega_1 = 2\Omega$. Recall from section 3 that binary data is encoded in the two channels according to eq. (3.1). For simplicity, we will only consider the timing jitter effect due to FWM collisions with the α -pulses in channel 1. As before, we fix α_1 at the leading edge of the data stream to be the reference pulse and note that it will only be affected by the faster moving (frequency = 2Ω) FWM components. In order to estimate the total number of FWM pulses hitting the reference pulse α_1 we first need to carefully examine the

launching points and the subsequent trajectories of FWM pulses generated due to the interaction of the α and β pulses in the fiber. It is convenient to introduce the following schematic diagram of the collisions indicating the various trajectories of the main channel and FWM pulses.

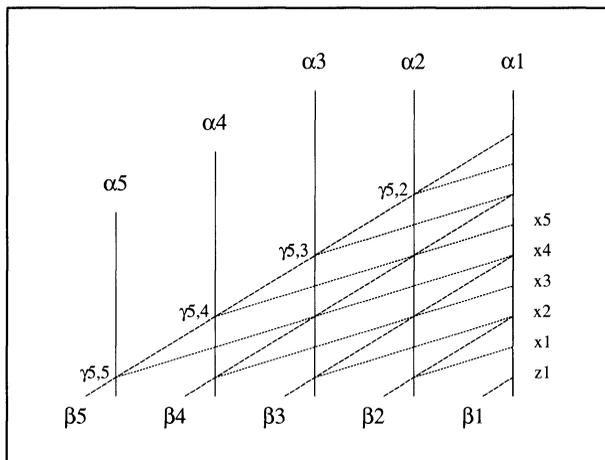


FIGURE 1. Pulse trajectories in a two-channel collision. Solid lines: channel 1 pulses; dashed lines: channel 2 pulses; dotted lines: FWM pulses. z_1 is the collision point of β_1 with the reference pulse α_1 ; x_1, x_2, \dots are the collision sites of the FWM pulses with α_1 .

A FWM pulse is generated at the collision point $z = z_{pq}$ of the pulses α_p and β_q where their phases S_j (cf. eqs. (2.5,3.1)) coincide. This amounts to

$$(4.1) \quad t + pT = t + qT + t_0 - \Omega z \Rightarrow z = z_{pq} = z_1 + (q - p)z_s,$$

where $z_1 = t_0/\Omega$ and $z_s = T/\Omega$ were defined in section 3. We denote the corresponding FWM pulse by its bit value $\gamma_{pq} = \alpha_p\beta_q$ and its phase is given by

$$(4.2) \quad S_{fwm} = t + pT - 2\Omega(z - z_{pq}).$$

where z_{pq} is given by eq. (4.1). Note that $p \leq q$ since the trailing β -pulses which are to the left of α_p (see figure above) can only collide with α_p . The FWM pulse γ_{pq} with frequency 2Ω travels from its origin at z_{pq} along the fiber and collides with the reference pulse α_1 . These collision sites are determined by equating the phase S_{fwm} of γ_{pq} from eq. (4.2) with the phase $t + T$ of α_1 , and are given by

$$(4.3) \quad x(p, q) = z_1 + (2q - p - 1)z_f, p \leq q,$$

where $z_f = T/2\Omega$. The collision sites are equally spaced starting from $x(1, 1) = z_1$ and the spacing between the successive sites is z_f . The total number of FWM collision sites for a system length L is then estimated to be $N_4 = L/z_f$ which is *twice* the number of main channel collisions $N = L/z_s$. However, the actual number of FWM collisions with the reference pulse α_1 is in fact much more than N_4 because multiple FWM collisions occur at a given site. Here we do not take the first site $x(1, 1)$ into account because in the present analysis we neglect any “self-interaction” contribution of the FWM pulse γ_{11} generated at that site to the timing shift of the reference pulse α_1 . Thus if we relabel the collision sites simply

as $x_k = z_1 + kz_f$, $k = 1, 2, \dots, N_4$ then the total number of FWM pulses γ_{pq} colliding with the reference pulse α_1 at the site x_k is determined by the condition $x(p, q) = x_k$. Using eq. (4.3) we see that the possible values of p, q for the FWM pulse γ_{pq} at x_k satisfy the constraint

$$(4.4) \quad \begin{aligned} 2q - p &= k + 1 \quad \text{or,} \\ q &= k + 1 - j, \quad p = k + 1 - 2j, \quad j = 0, 1, \dots, m - 1, \quad m = \lfloor \frac{k + 1}{2} \rfloor \end{aligned}$$

Thus there are m FWM collisions at x_k , $k = 1, 2, \dots, N_4$.

The permanent residual frequency shift and timing jitter of the reference pulse α_1 due to its collision with a single FWM component can be calculated using soliton perturbation theory. In the presence of loss and periodic amplification, the timing shift is periodic with period z_a , which is represented by a Fourier series as in the case of main channel collisions. In this paper, we omit the details of perturbation analysis and derivation of the timing shift formula for the main channel-FWM interaction; they will be reported elsewhere. Moreover, as in the main channel case we proceed to obtain the qualitative statistics of timing jitter due to FWM interference by assuming the simplified case of just one Fourier mode. That is, once again we take

$$(4.5) \quad \delta t(x_k) = (L - x_k)b_n \sin[2n\pi x_k/z_a - \theta_n]$$

corresponding to the n^{th} Fourier mode. Unlike the inter-channel collisions, the FWM-main channel interaction is dispersive. Therefore the FWM pulse will lose a fraction of its energy during each collision with a main channel pulse. If the FWM pulse collides with α_1 at position x after having made a total of μ (non-zero) collisions with main pulses from either channel 1 or channel 2, then we assume the timing shift to α_1 will be $\varepsilon^\mu \delta t(x)$, for some damping parameter $0 < \varepsilon \leq 1$. The parameter ε depends on the net energy loss of the FWM pulse during a collision and is assumed to be constant for all collisions. The details of the energy transfer mechanism are not necessary in the present context of statistical analysis, it will be discussed elsewhere. To take into account the damping effect we need to calculate the number of ‘‘pre-collisions’’ μ experienced by a FWM pulse γ_{pq} before it reaches the reference pulse α_1 . From the phase diagram it can be easily seen that before it arrives at the collision site x_k , the FWM pulse γ_{pq} collides with all channel 1 pulses: α_i , $i = 2 \dots p - 1$ which are to its right. But it does not collide with *all* the channel 2 pulses to the right of β_p . Using eqs. (4.2) and (4.4), we find that the collision between such a γ_{pq} (with phase S_{fwm}) and a channel 2 pulse β_i with phase $S_2 = t + iT - \Omega(z - z_1)$ occurs at the locations

$$z = z_1 + (k - i)z_s \quad \text{where} \quad z_{pq} < z < x_k,$$

along the fiber. The inequality above is due to the fact that the γ_{pq} can only collide after it is generated at z_{pq} and before it arrives at x_k . From this condition, it follows that $(k/2) + 1 < i < q$ for those β_i that are colliding with γ_{pq} before it reaches x_k . When k is even, $k = 2m$, β_i , $i = m + 2, m + 3, \dots, q - 1$ while in the odd case, $k = 2m - 1$, β_i , $i = m + 1, m + 2, \dots, q - 1$. Thus the total number of pre-collisions for the FWM pulse γ_{pq} arriving at the collision site x_k is given by

$$(4.6) \quad \mu(p, q) = \sum_{i=2}^{p-1} \alpha_i + \sum_{i=k-m+2}^{q-1} \beta_i$$

where $m = \lfloor (k + 1)/2 \rfloor$. Finally, the total timing jitter Δt_4 experienced by the reference pulse α_1 due to all the FWM pulses:

$$\begin{aligned}
 \Delta t_4 &= \sum_{k=1}^N \delta t(x_k) \sum_{p,q} \varepsilon^{\mu(p,q)} \alpha_p \beta_q \\
 &= \sum_{k=1}^N \delta t(x_k) \sum_{j=0}^{m-1} \varepsilon^{\mu(j,k)} \alpha_{k+1-2j} \beta_{k+1-j}
 \end{aligned}
 \tag{4.7}$$

where the second equality and the expression for the quantity $\mu(j, k)$ is obtained from eq. (4.6) with p, q parametrized according to eq. (4.4). Using the explicit form of $\mu(j, k)$ in eq. (4.7), we can express the mean timing jitter as

$$\langle \Delta t_4 \rangle = \sum_{k=1}^{N_4} \langle \delta t(x_k) \rangle \sum_{j=0}^{m-1} \langle X(j, k) \rangle \langle Y(j, k) \rangle,$$

where $X(j, k)$ and $Y(j, k)$ are independent random variables defined as

$$X(j, k) = \alpha_{k+1-2j} \varepsilon^{\mu_\alpha}, \quad \mu_\alpha = \sum_{i=2}^{k-2j} \alpha_i \quad \text{and} \quad Y(j, k) = \beta_{k+1-j} \varepsilon^{\mu_\beta}, \quad \mu_\beta = \sum_{i=m-2}^j \beta_{k-i}.$$

We first compute the inner sum in the above expression for $\langle \Delta t_4 \rangle$. This is the expected number of FWM pulses arriving at the collision site x_k . Denoting this average by ν_k and defining $\xi := (1 + \varepsilon)/2$, we get

$$\nu_k := \sum_{j=0}^{m-1} \langle X(j, k) \rangle \langle Y(j, k) \rangle = \frac{\xi^{k+m-2}}{4} \sum_{j=0}^{m-1} \xi^{-3j},$$

so that,

$$\begin{aligned}
 \nu_k &= \frac{\xi^{k-2m+1}(1 - \xi^{3m})}{4(1 - \xi^3)}, \quad \text{when } \xi < 1 \\
 &= \frac{m}{4} = \frac{\lfloor (k + 1)/2 \rfloor}{4}, \quad \text{when } \xi = 1.
 \end{aligned}
 \tag{4.8}$$

In order to calculate $\langle \delta t(x_k) \rangle$ next, we recall that $x_k = z_1 + kz_f = (k + 2U)z_f$ where $U = t_0/T$ is the uniformly distributed random variable on $[0, 1]$. Thus substituting eq. (4.8) and eq. (4.5) into eq. (4.7) yields,

$$\begin{aligned}
 \langle \Delta t_4 \rangle &= \sum_{k=1}^{N_4} \nu_k \int_0^1 \delta t[(k + 2U)z_f] dU = \sum_{k=1}^{N_4} \nu_k (u_{k+2} - u_k) \quad \text{where,} \\
 u_k &= \frac{z_f b_n}{2} \int_0^k (N_4 - x) \sin(2\pi\eta x - \theta_n), \quad \eta = \frac{nz_f}{z_a} \quad \text{and} \quad L = N_4 z_f.
 \end{aligned}
 \tag{4.9}$$

The sum over k in eq. (4.9) for $\langle \Delta t_4 \rangle$ can be rearranged by summation by parts as follows

$$\langle \Delta t_4 \rangle = \sum_{k=1}^{N_4} (\nu_{k-2} - \nu_k) u_k + (\nu_{N_4} u_{N_4+2} + \nu_{N_4-1} u_{N_4+1} - \nu_0 u_2 - \nu_{-1} u_1).$$

We note from eq. (4.8) that $\nu_0 = \nu_{-1} = 0$ for all values of ξ . Consequently the last two terms vanish identically in the above expression for $\langle \Delta t_4 \rangle$. We split the remaining calculations into two cases. For convenience we also take $\theta_n = 0$ ($\theta_n \neq 0$ case can be handled in a similar way). We consider first the case $\xi = \varepsilon = 1$. From eq. (4.8) we obtain $\nu_0 = \nu_{-1} = 0$, $\nu_{N_4} = \lfloor (N_4 + 1)/2 \rfloor / 4$ and that $\nu_{k-2} - \nu_k = -1/4$ for all k when $\xi = 1$. Substituting these into the expression for $\langle \Delta t_4 \rangle$ below eq. (4.9)

and calculating the integral for u_k explicitly, we obtain the mean timing shift for $\xi = 1$:

$$(4.10) \quad \langle \Delta t_4 \rangle = \frac{z_f b_n}{16\pi\eta} \left[\sum_{k=1}^{N_4} (N_4 - k) \cos 2\pi\eta k + S(N_4) \right],$$

where

$$S(N_4) = \lfloor (N_4 + 1)/2 \rfloor \left(2 \cos 2\pi\eta(N_4 + 2) - \frac{\sin 2\pi\eta(N_4 + 2)}{2\pi\eta} \right) + \lfloor (N_4/2) \rfloor \left(\cos 2\pi\eta(N_4 + 1) - \frac{\sin 2\pi\eta(N_4 + 1)}{2\pi\eta} \right) + \frac{1}{2\pi\eta} \sum_{k=1}^{N_4} \sin 2\pi\eta k$$

Note that $S(N_4)$ is an $O(N_4)$ term for all values of η . Moreover the trigonometric sum in eq. (4.10) is precisely

$$\sum_{k=1}^{N_4} (N_4 - k) \cos(2\pi\eta k) = -\frac{N_4}{2} + \frac{1}{2} \left(\frac{\sin \pi\eta(N_4)}{\sin \pi\eta} \right)^2$$

Thus when η is not an integer value (away from resonance), the above sum is $O(N_4)$ and we have $\langle \Delta t_4 \rangle = O(N_4)$. Therefore, the mean timing jitter in this case increases linearly with system length L . But for integer values of η , the quantity $\sin \pi\eta(N_4)/\sin \pi\eta \sim O(N_4)$. So at resonance the mean timing jitter turns out to be

$$\langle \Delta t_4 \rangle = \frac{N_4^2 b_n z_f}{32\pi\eta} + O(N_4).$$

Next we consider the case when $\xi < 1$. In this case eq. (4.8) gives $\nu_{k-2} - \nu_k = -\xi^{3m-2}/4$ when $k = 2m$ and $\nu_{k-2} - \nu_k = -\xi^{3m-3}/4$ when $k = 2m - 1$. Hence it is convenient to decompose the sum over k in the expression for $\langle \Delta t_4 \rangle$ into even and odd parts. Furthermore, when $N_4 \gg 1$, ν_{N_4} and ν_{N_4-1} become $O(1)$ terms after we neglect the exponentially small contribution due to $\xi^{3\lfloor (N_4+1)/2 \rfloor}$. Thus we have

$$\langle \Delta t_4 \rangle = \frac{z_f b_n}{16\pi\eta\xi^3} \left[\xi \sum_{m=1}^{\lfloor N_4/2 \rfloor} (N_4 - 2m) A_m + \sum_{m=1}^{\lfloor (N_4+1)/2 \rfloor} (N_4 - 2m + 1) B_m + O(1) \right]$$

where $A_m = \xi^{3m} \cos 4\pi\eta m$ and $B_m = \xi^{3m} \cos 2\pi\eta(2m - 1)$. The Dirichlet type sums involving A_m, B_m and mA_m, mB_m can be computed in a straight-forward way. After dropping all the exponentially small terms involving $\xi^{3\lfloor (N_4+1)/2 \rfloor}$ and $\xi^{3\lfloor N_4/2 \rfloor}$ from these sums, they become $O(1)$ quantities for all values of $\eta \neq 0$ (integer or non-integer). Therefore the mean timing jitter in the $\xi < 1$ case is given by

$$(4.11) \quad \langle \Delta t_4 \rangle = \frac{z_f b_n N_4}{16\pi\eta} \left(\frac{(1 - \xi^3) \cos 2\pi\eta + \xi \cos 4\pi\eta - \xi^4}{1 - 2\xi^3 \cos 4\pi\eta + \xi^6} \right) + O(1).$$

Thus we find that the mean timing jitter does not exhibit any resonance and grows linearly with system length L if the damping effect due to energy loss of the FWM pulses is taken into consideration.

Next we need to calculate the variance of the total timing shift Δt_4 . Proceeding in a similar way as for the main channel collisions (cf. section 3), we first compute

the mean square timing jitter using eq. (4.7). This is given by the formula

$$\langle(\Delta t_4)^2\rangle = \sum_{k,l=1}^{N_4} \langle\delta t(x_k)\delta t(x_l)\rangle \sum_{j_1=0}^{m_k-1} \sum_{j_2=0}^{m_l-1} \langle X(j_1,k)X(j_2,l)\rangle \langle Y(j_1,k)Y(j_2,l)\rangle,$$

where $m_k = \lfloor(k+1)/2\rfloor$, $m_l = \lfloor(l+1)/2\rfloor$ and X, Y are the independent variables introduced earlier. However, since the calculations involved are extremely lengthy we omit all the details and summarize only the main results:

We find that in the absence of damping ($\xi = \varepsilon = 1$), the variance $\text{var}(\Delta t_4) \sim O(N_4^4)$ in the non-resonant case, that is, when η is not an integer. However, $\text{var}(\Delta t_4) \sim O(N_4^6)$ when η is an integer. Consequently, there is a sharp increase in the variance in the resonant case. Finally, when damping is taken into account (that is, when $\xi < 1$), $\text{var}(\Delta t_4) \sim O(N_4^3)$ in the non-resonant case, whereas $\text{var}(\Delta t_4) \sim O(N_4^4)$ at resonance.

5. Conclusion.

We have studied a statistical model of collision-induced timing shifts generated by both main channel and four wave mixing in a two channel WDM optical communication system with damping and amplification. In this model we have assumed that the collisions between the pulses have a random component which is inherited from the randomness of the encoded data and the initial time delay between the bit streams in the adjacent channels. In addition to the main channel interactions, we have also considered the interference between a main channel pulse with the FWM noise. Such events are estimated to be twice as frequent as the main channel collisions. We have calculated the mean and the variance of the timing jitter experienced by a pulse propagating through the FWM noise field. The mean and variance have a power law dependence with system length L in both cases of inter-channel and FWM-main channel interactions that are considered here. However, we have also found a resonance phenomenon that takes place when the distance between successive collisions (z_s for main channels and z_f for FWM) with a given pulse is an integer multiple of the amplifier spacing. At resonance, the mean and variance grow faster (larger exponent in the power law) with system length than in the non-resonant case. In this article we have concentrated mainly on the statistical computations involved in a two-channel collision process. We have simplified our calculations by replacing the actual Fourier sum expressions for the timing shifts by only one Fourier mode although it does not affect our statistical analyses in any essential way. We expect to report the details omitted here in a future publication.

References

1. M. J. Ablowitz, G. Biondini, S. Chakravarty, R. B. Jenkins and J. R. Sauer, *Four-wave mixing in wavelength-division multiplexed soliton systems: Damping and amplification*, Opt. Lett. **21** (1996), 1646–1648.
2. M. J. Ablowitz, G. Biondini, S. Chakravarty, R. B. Jenkins and J. R. Sauer, *Four-wave mixing in wavelength-division-multiplexed soliton systems: ideal fibers*, J. Opt. Soc. Am. B **14** (1997)
3. M. J. Ablowitz, G. Biondini, S. Chakravarty, R. Horne, *On timing jitter in wavelength-division multiplexed soliton systems*, Opt. Commun. **150** (1998), 305–318.
4. M. J. Ablowitz, G. Biondini, S. Chakravarty, R. L. Horne and E. Spiller, *Four-wave mixing in strong dispersion-managed WDM soliton systems*, in OSA/IEEE/LEOS Conf. in Nonlinear Optics: Materials, Fundamentals and Applications 2000, OSA Technical Digest Series (2000), 338–340.

5. S. Chakravarty, M. J. Ablowitz, J. R. Sauer and R. B. Jenkins, *Multi-soliton interactions and wavelength-division multiplexing*, Opt. Lett. **20** (1995), 136–138.
6. K. Fukuchi, M. Kakui, A. Sasaki, T. Ito, Y. Inada, T. Tsuzaki, T. Shitomi, K. Fujii, S. Shikii, H. Sugahara, and A. Hasegawa, *1.1 Tb/s (55×20 Gb/s) dense WDM soliton transmission over 3,020 Km widely dispersion-managed transmission line employing 1.55-1.58 μ m hybrid repeaters*, in 25th European Conference on Optical Communication, Nice, France, paper PD2-10, (1999).
7. D. L. Guen, S. D. Burgo, M. L. Moulinard, D. Grot, M. Henry, F. Favre, and T. Georges, *Narrow band 1.02 Tbit/s (51×20 Gbit/s) soliton DWDM transmission over 1,000 Km of standard fiber with 100 Km amplifier span*, in Optical Fiber Communications Conference, 1999 OSA Technical Digest Series, (Optical Society of America, Washington, D.C.) (1999), paper PD4.
8. A. H. Liang, H. Toda and A. Hasegawa, *Dense periodic fibers with ultra-low four-wave mixing over a broad wavelength range*, Opt. Lett. **24** (1999), 1094–1096.
9. P. V. Mamyshev and L. F. Mollenauer, *Pseudo-phase-matched four-wave mixing in soliton wavelength-division multiplexed transmission*, Opt. Lett. **21** (1996), 396–398.
10. A. Mecozzi, *Timing jitter in wavelength-division multiplexed filtered soliton transmission*, J. Opt. Soc. Am. B **13** (1998), 152–.
11. L. F. Mollenauer, S. G. Evangelides and J. P. Gordon, *Wavelength Division Multiplexing with Solitons in Ultra-Long Distance Transmission Using Lumped Amplifiers*, J. Lightwave Technol. **9** (1991), 362–367.
12. H. Sugahara and A. Maruta, *Collision-induced timing-jitter analysis in a wavelength-division-multiplexed optical soliton-transmission system with dispersion management*, J. Opt. Soc. Am. B **18** (2001), 419–431.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF COLORADO, COLORADO SPRINGS, CO 80933-7150, USA

E-mail address: gjmorrow@math.uccs.edu, chuck@math.uccs.edu

This page intentionally left blank

Cuspons and peakons vis-a-vis regular solitons and collapse in a three-wave system

Roger Grimshaw, Georg A. Gottwald, and Boris A. Malomed

ABSTRACT. We introduce a general model of a one-dimensional three-component wave system with cubic nonlinearity. Linear couplings between the components prevent intersections between the corresponding dispersion curves, which opens two gaps in the system's linear spectrum. Detailed analysis is performed for zero-velocity solitons, in the reference frame in which the group velocity of one wave is zero. Disregarding the self-phase-modulation (SPM) term in the equation for that wave, we find an analytical solution which shows that there simultaneously exist two different families of generic solitons: regular ones, which may be regarded as a smooth deformation of the usual *gap solitons* in the two-wave system, and *cuspons* with a singularity in the first derivative at the center, while their energy is finite. Even in the limit when the linear coupling of the zero-group-velocity wave to the other two components is vanishing, the soliton family remains drastically different from that in the linearly uncoupled system: in this limit, regular solitons whose amplitude exceeds a certain critical value are replaced by *peakons*. While the regular solitons, cuspons, and peakons are found in an exact analytical form, their stability is tested numerically, showing that they all may be stable. In the case when the cuspons are unstable, the instability may trigger onset of *spatio-temporal collapse* in the system. If the SPM terms are retained, we find that there again simultaneously exist two different families of generic stable soliton solutions, which are regular ones and peakons. The existence of the peakons depends, in this case, on the sign of certain parameters of the system. Direct simulations show that both types of the solitons may be stable in this most general case too.

1. Introduction

1.1. Gap-soliton models. Gap solitons (GS) is a common name for solitary waves in nonlinear systems which feature one or more gaps in their linear spectrum [1]. A soliton may exist if its frequency belongs to the gap, as then it does not decay into linear waves.

Gaps in the linear spectrum are a generic phenomenon in two- or multicomponent systems, as intersection of dispersion curves belonging to different components is, generically, prevented by a linear coupling between the components. Excluding cases when the zero solution in the system is unstable [2], the intersection avoidance

2000 *Mathematics Subject Classification.* Primary 35Q51; Secondary 35Q55.

Key words and phrases. Bragg grating, gap soliton, massive Thirring model.

alters the spectrum so that a gap opens in place of the intersection. Approximating the two dispersion curves, that would intersect in the absence of the linear coupling, by straight lines, and assuming a generic cubic nonlinearity, one arrives at a *generalized massive Thirring model* (GMTM) for two wave fields $u_{1,2}(x, t)$:

$$(1.1) \quad i\left(\frac{\partial u_1}{\partial t} - \frac{\partial u_1}{\partial x}\right) + u_2 + (\sigma|u_1|^2 + |u_2|^2) u_1 = 0,$$

$$(1.2) \quad i\left(\frac{\partial u_2}{\partial t} + \frac{\partial u_2}{\partial x}\right) + u_1 + (\sigma|u_2|^2 + |u_1|^2) u_2 = 0,$$

where the group velocities of the two waves are normalized to be ± 1 , the linear-coupling constant and the coefficient of the nonlinear *cross-phase-modulation* (XPM) coupling may also be normalized to be 1, and $\sigma \geq 0$ is the *self-phase-modulation* (SPM) coefficient.

The model based on Eqs. (1.1) and (1.2) with $\sigma = 1/2$ has a direct, and very important, application to nonlinear optics, describing co-propagation of left- and right-traveling electromagnetic waves in a fiber with a resonant Bragg grating (BG) written on it [3, 4, 1]. The version of the model corresponding to $\sigma \rightarrow \infty$, i.e., with the SPM nonlinearity only,

$$(1.3) \quad i\left(\frac{\partial u_1}{\partial z} - \frac{\partial u_1}{\partial \tau}\right) + u_2 + |u_1|^2 u_1 = 0,$$

$$(1.4) \quad i\left(\frac{\partial u_2}{\partial z} + \frac{\partial u_2}{\partial \tau}\right) + u_1 + |u_2|^2 u_2 = 0,$$

may also be realized in terms of nonlinear fiber optics, describing propagation of light in a dual-core fiber with a group-velocity mismatch between the cores (which is normalized to be 1), while the intrinsic dispersion of the cores is neglected [5]. In Eqs. (1.3) and (1.4), the evolutionary variable is not time, but rather the propagation distance z , while the role of x is played by the so-called reduced time, $\tau \equiv t - z/V_0$, where V_0 is the mean group velocity of the carrier wave.

It had been demonstrated more than twenty years ago that the massive Thirring model proper, which corresponds to Eqs. (1.1) and (1.2) with $\sigma = 0$, is *exactly integrable* by means of the inverse scattering transform, and, moreover, it can be explicitly transformed into the sine-Gordon equation [6]. On the other hand, it was also demonstrated that GMTM with any $\sigma \neq 0$ is *not* integrable (this conclusion follows, for instance, from an early observation that collisions between solitons are inelastic if $\sigma \neq 0$ [4]). Nevertheless, the general model (1.1), (1.2) with an arbitrary value of σ has a family of exact GS solutions that completely fill the gap in its spectrum. Gap solitons, first predicted theoretically [3, 4], were observed in experiments with light pulses launched into a short piece of the BG-equipped fiber [7] (in fact, optical solitons that were first observed in the BG fiber [8] were, strictly speaking, not of the GS type, but more general ones, whose central frequency did not belong to the fiber's bandgap).

Models giving rise to GSs are known not only in optics but also in other areas, for instance, in hydrodynamics of density-stratified fluids, where dispersion curves pertaining to different internal-wave modes can readily intersect. Taking into regard the nonlinearity, one can easily predict the occurrence of GS in density-stratified fluids [10].

1.2. Introducing a three-wave model. In this work, we aim to study GSs in a system of *three* coupled waves, assuming that the corresponding three dispersion curves are close to intersection at a single point, unless linear couplings are taken into regard. Of course, the situation with three curves passing through a single point is degenerate. Our objective is to investigate GS not for this special case, but in its vicinity in the parameter space. We will demonstrate that families of GS solutions in the three-wave systems is drastically different from that in the two-wave GMTM. In particular, generic solutions will include not only regular solitons, similar to those known in GMTM, but also *cuspons* and *peakons*, i.e., solitons with a divergence or jump of the first derivative, but, nevertheless, with finite amplitude and energy. Moreover, we will demonstrate that a part of the cuspon and peakon solutions are completely stable ones. Another principal difference of the three-wave system from its two-wave counterpart is that the former one may give rise to *spatio-temporal collapse*, i.e., formation of a singularity of the wave fields in finite time. We will demonstrate that, in the cases when cuspons or peakons are unstable, their instability may easily provoke the onset of the collapse [9].

Three-wave systems of this type can readily occur in the above-mentioned density-stratified flows [11], and are also possible in optics. For instance, this case takes place in a *resonantly absorbing* BG, which are arranged as a system of thin (~ 100 nm) parallel layers of two-level atoms, with the spacing between them equal to half the wavelength of light. This system combines the resonant Bragg reflection and self-induced transparency (SIT), see Ref. [12] and references therein. A model describing the BG-SIT system includes equations for three essential fields, viz., local amplitudes of right- and left-traveling electromagnetic waves, and the inversion rate of the two-level atoms (which, obviously, has zero group velocity in the laboratory reference frame). This model indeed produces a linear spectrum with three dispersion curves close to intersecting at one point, so that two gaps open in the system's spectrum.

Another realization of gaps between three dispersion curves is possible in terms of stationary optical fields in a planar nonlinear waveguide equipped with BG in the form of parallel scores [13]. In this case, the resonant Bragg reflection linearly couples waves propagating in two different directions. To induce linear couplings between all the three waves in the system, it is necessary to have a planar waveguide with *two* different BG systems of parallel scores, oriented in different directions. Postponing a consideration of this rather complicated model to another work, we here give a simple example for a case when the single BG is aligned along the axis x , perpendicular to the propagation direction z . Two waves $u_{1,2}$ have opposite incidence angles with respect to the BG, while the third wave u_3 has its wave vector parallel to x , see Fig. 1 in Ref. [13]. Then, assuming that the size of the sample is much smaller than the diffraction length of a broad spatial beam, but is larger than a characteristic length induced by strong artificial diffraction induced by BG, normalized equations governing the spatial evolution of the fields in the planar waveguide with the usual Kerr nonlinearity are

$$(1.5) \quad i\left(\frac{\partial u_1}{\partial z} - \frac{\partial u_1}{\partial x}\right) + u_2 + \left(\frac{1}{2}|u_1|^2 + |u_2|^2 + |u_3|^2\right)u_1 = 0,$$

$$(1.6) \quad i\left(\frac{\partial u_2}{\partial z} + \frac{\partial u_2}{\partial x}\right) + u_1 + \left(\frac{1}{2}|u_2|^2 + |u_1|^2 + |u_3|^2\right)u_2 = 0,$$

$$(1.7) \quad i \frac{\partial u_3}{\partial z} + \left(\frac{1}{2} |u_3|^2 + |u_1|^2 + |u_2|^2 \right) u_3 = k_0 u_3,$$

where k_0 is a wavenumber mismatch between the third and first two waves.

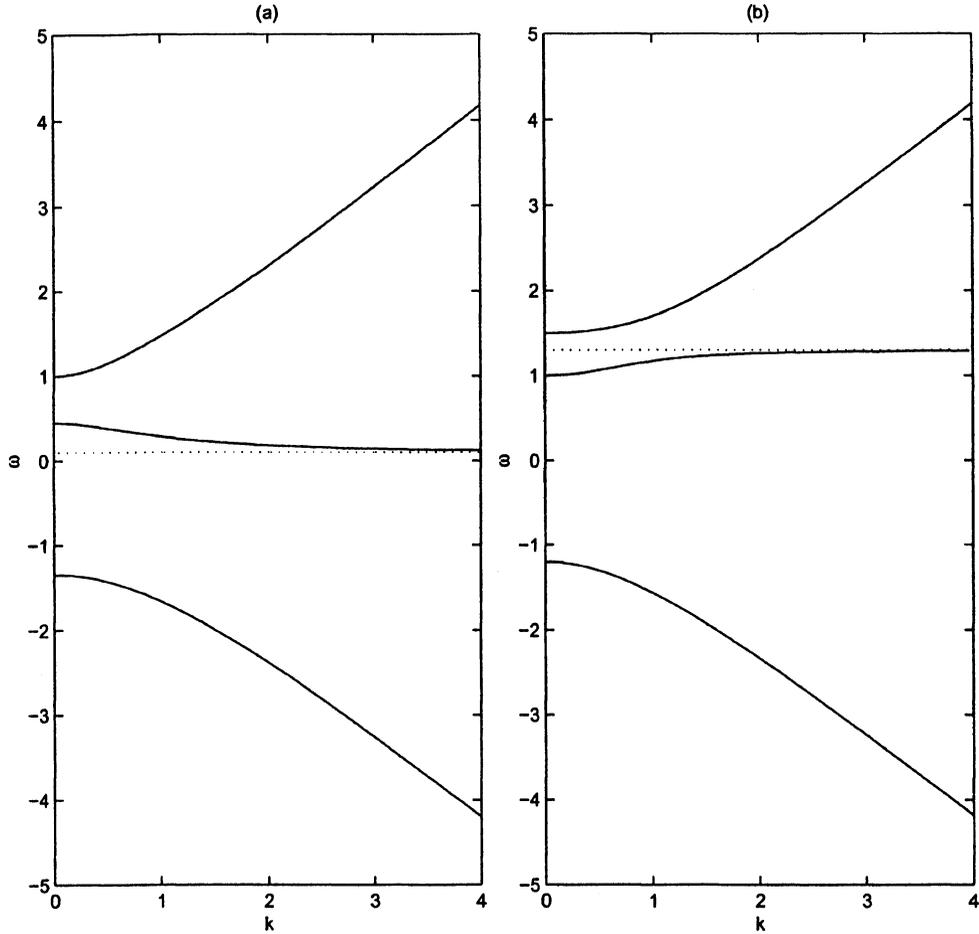


FIGURE 1. Dispersion curves produced by Eq. (2.1) in the case $\kappa = 0.5$: (a) $\omega_0 < 1 - \kappa^2$; (b) $\omega_0 > 1$. The dashed line in each panel is $\omega = \omega_0$. The case with $1 - \kappa^2 < \omega_0 < 1$ is similar to the case (a) but with the points ω_+ and 1 at $k = 0$ interchanged.

The model based on Eqs. (1.5) - (1.7) represents a particular case only, as it does not include linear couplings between the waves $u_{1,2}$ and u_3 . We aim to introduce a generic model describing a nonlinear system of three waves with linear couplings between all of them. We assume that the system can be derived from a Hamiltonian, and confine attention to the case of cubic nonlinearities. Taking into regard these restrictions, and making use of scaling invariances to diminish the number of free parameters, we arrive at a system

$$(1.8) \quad i \left(\frac{\partial u_1}{\partial t} - \frac{\partial u_1}{\partial x} \right) + u_2 + \kappa u_3 + \alpha \left(\alpha \sigma_1 |u_1|^2 + \alpha |u_2|^2 + |u_3|^2 \right) u_1 = 0,$$

$$(1.9) \quad i\left(\frac{\partial u_2}{\partial t} + \frac{\partial u_2}{\partial x}\right) + u_1 + \kappa u_3 + \alpha(\alpha\sigma_1|u_2|^2 + \alpha|u_1|^2 + |u_3|^2)u_2 = 0,$$

$$(1.10) \quad i\frac{\partial u_3}{\partial t} + \kappa(u_1 + u_2) + (\sigma_3|u_3|^2 + \alpha|u_1|^2 + \alpha|u_2|^2)u_3 = \omega_0 u_3.$$

Here, we consider the evolution in the temporal domain, unlike the spatial-domain evolution in Eqs. (1.5) - (1.7), and without loss of generality, we use a reference frame in which the third wave u_3 has zero group velocity. Note that the coefficient of the linear coupling between the first two waves is normalized to be 1, while κ accounts for their linear coupling to the third wave, and it may always be defined to be positive.

We assume full symmetry between the two waves $u_{1,2}$, following the pattern of the GMT model; in particular, the group velocities of these waves are normalized to be ∓ 1 . However, we note that this assumption is not essential, and we shall comment later on the case when the group-velocity terms in Eqs. (1.8) and (1.9) are generalized as follows:

$$(1.11) \quad -\frac{\partial u_1}{\partial x} \rightarrow -c_1 \frac{\partial u_1}{\partial x}, \quad +\frac{\partial u_2}{\partial x} \rightarrow +c_2 \frac{\partial u_1}{\partial x},$$

where c_1 and c_2 are different, but have the same sign. Note that the symmetry of the system's dispersion law $\omega = \omega(k)$ is assumed with respect to the sign of k , but not of ω . To this end, the parameter ω_0 was added to Eq. (3). This parameter breaks the " ω -symmetry", that, unlike the " k -symmetry", does not have any natural cause to exist.

The coefficients $\sigma_{1,3}$ and α in Eqs. (1.8) - (1.10) account for the nonlinear SPM and XPM nonlinearities, respectively. In particular, α is defined as a relative XPM coefficient between the first two and the third waves, hence it is an irreducible parameter. As for the SPM coefficients, both σ_1 and σ_3 may be normalized to be ± 1 , unless they are equal to zero; however, it will be convenient to keep them as free parameters, see below (note that the SPM coefficients are always positive in the optical models, but in those describing stratified fluids they may have either sign).

Equations (1.8) - (1.10) conserve the norm, which is frequently called energy in optics,

$$(1.12) \quad N \equiv \sum_{n=1,2,3} \int_{-\infty}^{+\infty} |u_n(x)|^2 dx,$$

the Hamiltonian,

$$(1.13) \quad H \equiv H_{\text{grad}} + H_{\text{coupl}} + H_{\text{focus}},$$

$$(1.14) \quad H_{\text{grad}} \equiv \frac{i}{2} \int_{-\infty}^{+\infty} \left(u_1^* \frac{\partial u_1}{\partial x} - u_2^* \frac{\partial u_1}{\partial x} \right) dx + \text{c.c.},$$

$$(1.15) \quad H_{\text{coupl}} \equiv - \int_{-\infty}^{+\infty} [u_1^* u_2 + \kappa u_3^* (u_1 + u_2)] dx + \text{c.c.},$$

$$(1.16) \quad H_{\text{focus}} \equiv - \int_{-\infty}^{+\infty} \left[\frac{1}{2} \alpha^2 \sigma_1 (|u_1|^4 + |u_2|^4) + \frac{1}{2} \sigma_3 |u_3|^4 \right.$$

$$(1.17) \quad \left. + \alpha^2 |u_1|^2 |u_2|^2 + \alpha |u_3|^2 (|u_1|^2 + |u_2|^2) \right] dx,$$

and the momentum, which will not be used here. In these expressions, the asterisk and c.c. both stand for complex conjugation, H_{grad} , H_{coupl} and H_{focus} being the gradient, linear-coupling, and self-focusing parts of the Hamiltonian.

1.3. Solitons in the three-wave models. Our objective is to find various types of solitons existing in the generic three-wave system (1.8) - (1.10) and investigate their stability. The existence of various types of the solitons is considered in section 3. Focusing first on the case (suggested by the analogy with GMTM) when the SPM term in Eq. (1.10) may be neglected (i.e., $\sigma_3 = 0$), we will find a general family of zero-velocity solitons in an exact analytical form. We will demonstrate that the family contains solutions of two drastically different types: regular GSs, and *cuspons*, i.e., solitons with a cusp singularity at the center, while their energy is finite (this singularity assumes that the function remains finite at the cusp point, while its first derivative diverges). Cuspons are known to exist in degenerate models without linear terms (except for the evolution term such as $\partial u/\partial t$), a well-known example being the Camassa-Holm (CH) equation [17, 18]. As well as the massive Thirring model (1.1), (1.2) with $\sigma = 0$, the CH equation is *exactly integrable* by means of the inverse scattering transform. Our nonintegrable model, as well as the CH one, gives rise to coexisting solutions in the form of regular solitons and cuspons. However, the cause for the existence of cuspons in our model is very different: looking for a zero-velocity soliton solution, one may eliminate the field u_3 by means of an algebraic relation following, in this case, from Eq. (1.10). The subsequent substitution of this result into the first two equations (1.8) and (1.9) produces a *non-polynomial* (in fact, rational) nonlinearity in them. The corresponding rational functions feature a singularity at some (critical) value of the soliton's amplitude. If the amplitude of a formal regular-soliton solution exceeds the critical value, it actually cannot exist, and, in the case when $\sigma_3 = 0$, it is replaced by a cuspon, whose amplitude is exactly equal to the critical value.

In the limit $\kappa \rightarrow 0$, which corresponds to the vanishing linear coupling between the first two and third waves, the cuspon resembles a *peakon*, which is a finite-amplitude solitary wave with a jump of its first derivative at the center. Note that peakon solutions, coexisting with regular solitons (they also coexist in our model), are known in a slightly different version of the CH equation (which is also integrable by means of the inverse scattering transform), see, e.g., Ref. [17, 19, 20].

Then, we show that, when the SPM term in Eq. (1.10) is restored in Eq. (1.10) (i.e., $\sigma_3 \neq 0$; the presence or absence of the SPM term $\propto \sigma_1$ in Eqs. (1.8) and (1.9) is not crucially important), the system supports a different set of soliton families. These are regular GSs and, depending on the sign of certain parameters, a family of peakons, which, this time, appear as generic solutions, unlike the case $\sigma_3 = 0$, when they only exist as a limiting form of the solutions corresponding to $\kappa \rightarrow 0$. As far as we know, the model formulated in the present work is the first non-degenerate one (i.e., a model with a nonvanishing linear part) which yields both cuspons and peakons.

1.4. Stability of the solitons and spatiotemporal collapse. As concerns the dynamical stability of the various solitons in the model (1.8) -(1.10), in this work we limit ourselves to direct simulations, as a more rigorous approach, based on numerical analysis of the corresponding linear stability-eigenvalue problem [21],

is technically difficult in the case of cuspons and peakons. In fact, direct simulations of perturbed cuspons and peakons is a hard problem too, but we have checked that identical results concerning the stability are produced (see section 3 below) by high-accuracy finite-difference and pseudo-spectral methods (each being implemented in more than one particular form), which lends the results credibility. A general conclusion is that the regular solitons are always stable. As for the cuspons and peakons, they may be both stable or unstable. If the cusp is strong enough, instability of the cuspon initiates formation of a genuine singularity, i.e., onset of a *spatiotemporal collapse* [9] in the present one-dimensional model.

Note that a simple virial-type estimate for the possibility of collapse can be made, assuming that the field focuses itself in a narrow spot with a size $L(t)$, amplitude $\aleph(t)$, and a characteristic value $K(t)$ of the field's wavenumber [9]. The conservation of the norm (1.12) imposes a restriction $\aleph^2 L \sim N$, i.e., $L \sim N/\aleph^2$. Next, the self-focusing part (1.13) of the Hamiltonian (1.13), which drives the collapse, can be estimated as

$$(1.18) \quad H_{\text{focus}} \sim -\aleph^4 L \sim -N\aleph^2.$$

On the other hand, the collapse can be checked by the gradient term (1.14) in the full Hamiltonian, that, in the same approximation, can be estimated as $H_{\text{grad}} \sim \aleph^2 KL \sim NK$. Further, Eqs. (1.8) - (1.10) suggest an estimate $K \sim \aleph^2$ for a characteristic wavenumber of the wave field (the same estimate for K follows from an expression (2.7) for the exact stationary-soliton solution given below), thus we have $H_{\text{grad}} \sim N\aleph^2$. Comparing this with the expression (1.18), one concludes that the parts of the Hamiltonian promoting and inhibiting the collapse scale the same way as $\aleph \rightarrow \infty$ (or $L \rightarrow 0$), hence a *weak collapse* [9] may be possible (but not necessarily) in systems of the present type. In the models of GSs studied thus far and based on GMTM, collapse has never been reported. The *real existence* of the collapse in the present one-dimensional three-wave GS model is therefore a novel dynamical feature, and it seems quite natural that cuspons and peakons, in the case when they are unstable, play the role of catalysts stimulating the onset of the collapse.

2. Analytical solutions for solitons

2.1. The dispersion relation. The first step in the investigation of the system is to understand its linear spectrum. Substituting $u_{1,2,3} \sim \exp(ikx - i\omega t)$ into Eqs. (1.8 -1.10), and omitting nonlinear terms, we arrive at a dispersion equation,

$$(2.1) \quad (\omega^2 - k^2 - 1)(\omega - \omega_0) = 2\kappa^2(\omega - 1).$$

If $\kappa = 0$, the third wave decouples, and the coupling between the first two waves produces a commonly known gap, so that the solutions to Eq. (2.1) are $\omega_{1,2} = \pm\sqrt{1+k^2}$ and $\omega_3 = \omega_0$. If $\kappa \neq 0$, the spectrum can be easily understood by treating κ as a small parameter. However, the following analysis is valid for all values of κ in the range $0 < \kappa^2 < 1$.

First, consider the situation when $k = 0$. Then, three solutions of Eq. (2.1) are

$$(2.2) \quad \omega = 1, \omega = \omega_{\pm} \equiv (\omega_0 - 1)/2 \pm \sqrt{(\omega_0 + 1)^2/4 + 2\kappa^2}.$$

It can be easily shown that $\omega_- < \min\{\omega_0, -1\} \leq \max\{\omega_0, -1\} < \omega_+$, so that one always has $\omega_- < -1$, while $\omega_+ > 1$ if $1 - \omega_0 > \kappa^2$, and vice versa. Next, it is readily seen that, as $k^2 \rightarrow \infty$, either $\omega^2 \approx k^2$, or $\omega \approx \omega_0$. It can also be shown that each

branch of the dispersion relation generated by Eq. (2.1) is a monotonic function of k^2 . Generic examples of the spectrum are shown in Fig. 1, where the panels (a) and (b) pertain, respectively, to the cases $\omega_0 < 1 - \kappa^2$ with $\omega_+ < 1$, and $\omega_0 > 1$ with $\omega_+ > 1$. The intermediate case, $1 - \kappa^2 < \omega_0 < 1$, is similar to that shown in panel (a), but with the points ω_+ and 1 at $k = 0$ interchanged. When $\omega_0 < 1$, the upper gap in the spectrum is $\min\{\omega_+, 1\} < \omega < \max\{\omega_+, 1\}$, while the lower gap is $\omega_- < \omega < \omega_0$. When $\omega_0 > 1$, the upper gap is $\omega_0 < \omega < \omega_+$, and the lower one is $\omega_- < \omega < 1$.

2.2. A generic family of gap solitons. The next step is to search for GS solutions to the full nonlinear system. In this work, we confine ourselves to the case of zero-velocity GS, substituting into Eqs. (1.8) - (1.10)

$$(2.3) \quad u_n(x, t) = U_n(x) \exp(-i\omega t), \quad n = 1, 2, 3,$$

where it is assumed that the soliton's frequency ω belongs to one of the gaps. In fact, even the description of zero-velocity solitons is quite complicated. Note, however, that if one sets $\kappa = 0$ in Eqs. (1.8) - (1.10), keeping nonlinear XPM couplings between the first two and third waves, the gap which exists in the two-wave GMT model remains unchanged, and the corresponding family of GS solutions does not essentially alter, in accord with the principle that nonlinear couplings cannot alter gaps or open a new one if the linear coupling is absent [14]; nevertheless, the situation is essentially different if κ is vanishingly small, but not exactly equal to zero, see below.

First, the substitution of (2.3) into Eqs. (1.8) and (1.9) leads to a system of two ordinary differential equations for $U_1(x)$ and $U_2(x)$,

$$(2.4) \quad iU_1' = \omega U_1 + U_2 + \kappa U_3 + \alpha (\alpha \sigma_1 |U_1|^2 + \alpha |U_2|^2 + |U_3|^2) U_1,$$

$$(2.5) \quad -iU_2' = \omega U_2 + U_1 + \kappa U_3 + \alpha (\alpha \sigma_1 |U_2|^2 + \alpha |U_1|^2 + |U_3|^2) U_2,$$

where the prime represents d/dx . To solve these equations, we substitute $U_{1,2} = A_{1,2}(x) \exp(i\phi_{1,2}(x))$ with real A_n and ϕ_n . After simple manipulations, it can be found that $(A_1^2 - A_2^2)' = 0$ and $(\phi_1 + \phi_2)' = 0$. With regard to the condition that the soliton fields vanish at infinity, we immediately conclude that

$$(2.6) \quad A_1^2(x) = A_2^2(x) \equiv S(x);$$

as for the constant value of $\phi_1 + \phi_2$, it may be set equal to zero without restriction of the generality, so that $\phi_1(x) = -\phi_2(x) \equiv \phi(x)/2$, where $\phi(x)$ is the relative phase of the two fields. After this, we obtain two equations for $S(x)$ and $\phi(x)$ from Eqs. (2.4) and (2.5),

$$(2.7) \quad \phi' = -2\omega - 2 \cos \phi - 2\alpha^2 (1 + \sigma_1) S - S^{-1} U_3^2 (\omega_0 - \omega - \sigma_3 U_3^2),$$

$$(2.8) \quad S' = -2S \sin \phi - 2\kappa \sqrt{S} U_3 \sin(\phi/2),$$

and Eq. (1.10) for the third wave U_3 takes the form of a cubic algebraic equation

$$(2.9) \quad U_3 (\omega_0 - \omega - 2\alpha S - \sigma_3 |U_3|^2) = 2\kappa \sqrt{S} \cos(\phi/2),$$

from which it follows that U_3 is a real-valued function.

This analytical consideration can be readily extended for more general equations (1.8) and (1.9) that do not assume the symmetry between the waves u_1 and u_2 , i.e., with the group-velocity terms in the equations altered as per Eq. (1.11).

In particular, the relation (2.6) is then replaced by $c_1 A_1^2(x) = c_2 A_2^2(x) \equiv S(x)$. It can be checked that results for the asymmetric model are not qualitatively different from those presented below for the symmetric one.

Equations (2.7) and (2.8) have a Hamiltonian structure, as they can be represented in the form

$$(2.10) \quad \frac{dS}{dx} = \frac{\partial H}{\partial \phi}, \quad \frac{d\phi}{dx} = -\frac{\partial H}{\partial S},$$

with the Hamiltonian

$$(2.11) \quad H = 2S \cos \phi + \alpha^2 (1 + \sigma_1) S^2 + 2\omega S + U_3^2 (\omega_0 - \omega - 2\alpha S) - \frac{3}{2} \sigma_3 U_3^4,$$

which is precisely a reduction of the Hamiltonian (1.13) of the original system (1.8) - (1.10) for the solutions of the present type. Note that H is here regarded as a function of S and ϕ , and the relation (2.9) is regarded as determining U_3 in terms of S and ϕ . For soliton solutions, the boundary conditions at $x = \pm\infty$ yield $H = 0$, so that the solutions can be obtained in an implicit form,

$$(2.12) \quad 2S \cos \phi + \alpha^2 (1 + \sigma_1) S^2 + 2\omega S + U_3^2 (\omega_0 - \omega - 2\alpha S) - (3/2) \sigma_3 U_3^4 = 0.$$

In principle, one can use the relations (2.9) and (2.12) to eliminate U_3 and ϕ and so obtain a single equation for S . However, this is not easily done unless $\sigma_3 = 0$ [no SPM term in Eq. (1.10)], therefore we proceed to examine this special, but important, case first. Recall that the zero-SPM case also plays an important role in the case of the two-wave GMTM based on Eqs. (1.1) and (1.2), as precisely in this case (which corresponds to the massive Thirring model proper) the equations are exactly integrable by means of the inverse scattering transform [6].

2.3. Cuspions in the zero-self-phase-modulation case ($\sigma_3 = 0$). Setting $\sigma_3 = 0$ makes it possible to solve Eq. (2.9) for U_3 explicitly in terms of S and ϕ ,

$$(2.13) \quad U_3 = \frac{2\kappa\sqrt{S} \cos(\phi/2)}{\omega_0 - \omega - 2\alpha S}.$$

For simplicity, we also set $\sigma_1 = 0$ in Eqs. (1.8) and (1.9) and subsequent equations, although the latter assumption is not crucially important for the analysis developed below. If $\sigma_1 \neq 0$ is restored, the conclusions of this subsection will not be substantially altered.

As the next step, one can also eliminate ϕ , using Eqs. (2.12) and (2.13), to derive a single equation for S ,

$$(2.14) \quad (dS/dx)^2 = 4S^2 F(S),$$

$$(2.15) \quad F(S) \equiv (1 - \omega - \frac{1}{2}\alpha^2 S) \left[2 \left(1 + \frac{\kappa^2}{\omega_0 - \omega - 2\alpha S} \right) - (1 - \omega - \frac{1}{2}\alpha^2 S) \right].$$

The function $F(S)$ has either one or three real zeros S_0 . One is

$$(2.16) \quad S_{01} = 2(1 - \omega)/\alpha^2,$$

and the remaining two, if they exist, are real roots of the quadratic equation,

$$(2.17) \quad (2 + 2\omega + \alpha^2 S_0)(\omega_0 - \omega - 2\alpha S_0) + 4\kappa^2 = 0.$$

Only the smallest positive real root of Eq. (2.17), to be denoted S_{02} (if such exists), will be relevant below. Note, incidentally, that $F(S)$ cannot have double roots. A consequence of this fact is that Eq. (2.14) cannot generate kink solutions, which

have different limits as $x \rightarrow \pm\infty$. Indeed, if $S(x) \rightarrow \text{const} \equiv \bar{S}$ as $x \rightarrow \pm\infty$, then one needs to have $dS/dx \sim (S - \bar{S})$ in the same limit, which implies that the function $F(s)$ in Eq. (2.14) must have a double zero at $S = \bar{S}$.

For a soliton solution of (2.14), we need first that $F(0) > 0$, which can be shown to be exactly equivalent to requiring that ω belongs to either the upper or the lower gap of the linear spectrum. We note that the coupling to the third wave gives rise to the rational nonlinearity in the expression (2.15), despite the fact that the underlying system (1.8) - (1.10) contains only linear and cubic terms. Even if the coupling constant κ is small, it is clear that the rational nonlinearity may produce a strong effect in a vicinity of a *critical value* of the squared amplitude at which the denominator in the expression (2.15) vanishes,

$$(2.18) \quad S_{\text{cr}} = (\omega_0 - \omega) / 2\alpha.$$

As it follows from this expression, one must have $\alpha(\omega_0 - \omega) > 0$ for the existence of the critical value.

If S_{cr} exists, the structure of the soliton crucially depends on whether, with an increase of S , the function $F(S)$ defined by Eq. (2.15) first reaches zero at $S = S_0$, or, instead, it first reaches the singularity at $S = S_{\text{cr}}$, i.e., whether $0 < S_0 < S_{\text{cr}}$, or $0 < S_{\text{cr}} < S_0$. In the former case, the existence of S_{cr} plays no role, and the soliton is a regular one, having the amplitude $\sqrt{S_0}$. This *regular soliton* may be regarded as obtained by a smooth deformation from the usual GS known in GMTM at $\kappa = 0$.

As the soliton cannot have an amplitude larger than $\sqrt{S_{\text{cr}}}$, in the case $0 < S_{\text{cr}} < S_0$ the squared amplitude takes the value S_{cr} , rather than S_0 . The soliton is singular in this case, being a *cuspon* [see Eqs. (2.24) and (2.25) below], but, nevertheless, it is an absolutely relevant solution. If $S_{\text{cr}} < 0$ and $S_0 > 0$ or vice versa, then the soliton may only be, respectively, regular or singular, and no soliton exists if both S_0 and S_{cr} are negative. Further, it is readily shown that for all these soliton solutions, $S(x)$ is symmetric about its center, which may be set at $x = 0$, that is, $S(x)$ is an even function of x . For the cuspon solutions, and for those regular solutions whose squared amplitude is S_{01} , it can also be shown that the phase variable $\psi(x) = \phi(x) - \pi$ and $U_3(x)$ are odd functions of x , while for those regular solutions whose squared amplitude is S_{02} the phase variable $\phi(x)$ and $U_3(x)$ are, respectively, odd and even functions of x .

It is now necessary to determine which parameter combinations in the set $(\omega, \omega_0, \alpha)$ permit the options described above. The most interesting case occurs when $\omega_0 > \omega$ (so that ω belongs to the lower gap, see Fig. 1) and $\alpha > 0$ (the latter condition always holds in the applications to nonlinear optics). In this case, it can be shown that the root S_{02} of Eq. (2.17) is not relevant, and the options are determined by the competition between S_{01} and S_{cr} . The soliton is a cuspon ($0 < S_{\text{cr}} < S_{01}$) if

$$(2.19) \quad \alpha(\omega_0 - \omega) < 4(1 - \omega).$$

In effect, the condition (2.19) sets an upper bound on α for given ω_0 and ω . In particular, the condition is always satisfied if $0 < \alpha < 4$.

If, on the other hand, the condition (2.19) is not satisfied (i.e., $0 < S_{01} < S_{\text{cr}}$), we obtain a regular soliton. In a less physically relevant case, when again $\omega_0 > \omega$ but $\alpha < 0$, cuspons do not occur [as this time $S_{\text{cr}} < 0$, see Eq. (2.18)], and only regular solitons may exist.

Next we proceed to the case $\omega_0 < \omega$, so that ω is located in the upper gap of the linear spectrum. For $\alpha > 0$, we have $S_{\text{cr}} < 0$, hence only regular solitons may occur, and indeed it can be shown that there is always at least one positive root S_0 , so a regular soliton exists indeed. If $\alpha < 0$, then we have $S_{\text{cr}} > 0$, but it can be shown that, if $\omega_0 < 1 - \kappa^2$ (when also $\omega < 1$), there is at least one positive root $S_0 < S_{\text{cr}}$; thus, only a regular soliton can exist in this case too. On the other hand, if $\alpha < 0$ and $\omega_0 > 1 - \kappa^2$ (and then $\omega > 1$), there are no positive roots S_0 , and so only cuspons occur.

Let us now turn to a detailed description of the cuspon's local structure near its center, when S is close to S_{cr} . From the above analysis, one sees that cuspons occur whenever ω lies in the lower gap, with $\omega_0 > \omega$ and $\alpha > 0$, so that the criterion (2.19) is satisfied, or when ω lies in the upper gap with $1 - \kappa^2 < \omega_0 < \omega$ and $\alpha < 0$. To analyze the structure of the cuspon, we first note that, as it follows from Eq. (2.12), one has $\cos \phi = -1$ (i.e., $\phi = \pi$) when $S = S_{\text{cr}}$, which suggest to set

$$(2.20) \quad S_{\text{cr}} - S \equiv \delta \cdot \kappa^2 R, \quad 1 + \cos \phi \equiv \delta \cdot \rho,$$

where δ is a small positive parameter, and the stretched variables R and ρ are positive. At the leading order in δ , it then follows from Eq. (2.12) that $\rho = \rho_0 R$, where

$$(2.21) \quad \rho_0 \equiv \alpha^3 (S_{01} - S_{\text{cr}}).$$

As it follows from the above analysis, ρ_0 is always positive for a cuspon. We also stretch the spatial coordinate, defining $x \equiv \delta^{3/2} \kappa^2 y$, the soliton center being at $x = 0$. Since $S(x)$ is an even function of x , it is sufficient to set $x > 0$ in this analysis. Then, on substituting the first relation from Eq. (2.20) into Eq. (2.14), we get, to the leading order in δ , an equation

$$(2.22) \quad R(dR/dy)^2 = \rho_0 S_{\text{cr}}^2 / \alpha^2 \equiv K^2,$$

so that

$$(2.23) \quad R = (3Ky/2)^{2/3}.$$

In the original unstretched variables, the relation (2.23) shows that, near the cusp,

$$(2.24) \quad S_{\text{cr}} - S(x) \approx (3K\kappa x/2)^{2/3},$$

$$(2.25) \quad dS/dx \approx (2/3)^{1/3} (K\kappa)^{2/3} \cdot x^{-1/3},$$

and it follows from Eq. (2.13) that U_3 is unbounded near the cusp,

$$(2.26) \quad U_3 \approx (S_{\text{cr}}/\alpha)(2\alpha\rho_0 K^2/3\kappa x)^{1/3}.$$

In particular, Eq. (2.25) implies that, as $K\kappa$ decreases, the cusp gets localized in a narrow region where $|x| \lesssim K^2 \kappa^2$ (outside this region, $|dS/dx|$ is bounded and shows no cusp). Note that this limit can be obtained either as $\kappa^2 \rightarrow 0$, or as $\rho_0 \rightarrow 0$ [recall ρ_0 is defined in Eq. (2.21)].

It is relevant to mention that, very close to the cusp, the underlying physical model, which is based on the paraxial approximation in the application to optical systems, or on a long-wave expansion in the case of internal waves in stratified fluids, may become irrelevant. However, this circumstance will lead to a modification of the structure of the physical fields inside the cuspon only in a very small vicinity of the singular point (for instance, on a scale of the order of the light wavelength

in optical systems, or the layer's depth in the fluids). Thus, the cuspon solutions are quite relevant to applications, provided that they are stable.

An example of the cuspon is shown in Fig. 2. Although the first derivative in the cuspon is singular at its center, as follows from Eq. (2.25) [see also Fig. 2(a)], it is easily verified that the Hamiltonian (1.13) (and, obviously, the norm (1.12) too) are finite for the cuspon solution. These solitons are similar to cuspons found as exact solutions to the Camassa-Holm (CH) equation [17, 18], which have a singularity of the type $|x|^{1/3}$ or $|x|^{2/3}$ as $|x| \rightarrow 0$, cf. Eqs. (2.24) and (2.25). The CH equation is integrable, and it is degenerate in the sense that it has no linear terms except for $\partial u/\partial t$ (which makes the existence of the solution with a cusp singularity possible). Our three-wave system (1.8) - (1.10) is not degenerate in that sense; nevertheless, the cuspon solitons are possible in it because of the model's multicomponent structure: the elimination of the third component generates the non-polynomial nonlinearity in Eqs. (2.4), (2.5), and, finally, in Eqs. (2.8) and (2.14), which gives rise to the cusp. It is noteworthy that, as well as the CH model, ours gives rise to two different *coexisting* families of solitons, viz., regular ones and cuspons. It will be shown below that the solitons of both types may be stable.

In the special case $\kappa \ll 1$, when the third component is weakly coupled to the first two ones in the linear approximation, a straightforward perturbation analysis shows that the cuspons look like *peakons*; that is, except for the above-mentioned narrow region of the width $|x| \sim \kappa^2$, where the cusp is located, they have the shape of a soliton with a discontinuity in the first derivative of $S(x)$ and a jump in the phase $\phi(x)$, which are the defining features of peakons ([17, 19]). An important result of our analysis is that the family of solitons obtained in the limit $\kappa \rightarrow 0$ is drastically different from that in the model where one sets $\kappa = 0$ from the very beginning. In particular, in the most relevant case, with $\omega_0 > \omega$ and $\alpha > 0$, the family corresponding to $\kappa \rightarrow 0$ contains regular solitons whose amplitude is smaller than $\sqrt{S_{\text{cr}}}$; however, the solitons whose amplitude at $\kappa = 0$ is larger than $\sqrt{S_{\text{cr}}}$, i.e., the ones whose frequencies belong to the region (2.19) [note that the definition of S_{cr} does not depend on κ at all, see Eq. (2.18)], are replaced by the peakons which are constructed in a very simple way: drop the part of the usual soliton above the critical level $S = S_{\text{cr}}$, and bring together the two symmetric parts which remain below the critical level, see Fig. 2(b). It is interesting that peakons are known as exact solutions to a version of the integrable CH equation slightly different from that which gives rise to the cuspons. As well as in the present system, in that equation the peakons coexist with regular solitons [19]. In the next subsection, we demonstrate that the peakons, which are found only as limit-form solutions in the zero-SPM case $\sigma_3 = 0$, become generic solutions in the case $\sigma_3 \neq 0$.

2.4. Peakons, the case $\sigma_3 \neq 0$. Before proceeding to the consideration of dynamical stability of various soliton solutions found above, it is relevant to address another issue, viz., *structural stability* of the cuspon solutions. To this end, we restore the SPM term in Eq. (1.10), that is, we now set $\sigma_3 \neq 0$, but assume that it is a small parameter. Note that, in the application to nonlinear optics, one should expect that $\sigma_3 > 0$, but there is no such a restriction on the sign of σ_3 in the application to the flow of a density-stratified fluid. We still keep $\sigma_1 = 0$, as the inclusion of the corresponding SPM terms in Eqs. (1.8) and (1.9) amounts to straightforward changes in details of both the above analysis, and that presented below. On the other hand, we show below that the inclusion of the SPM term in

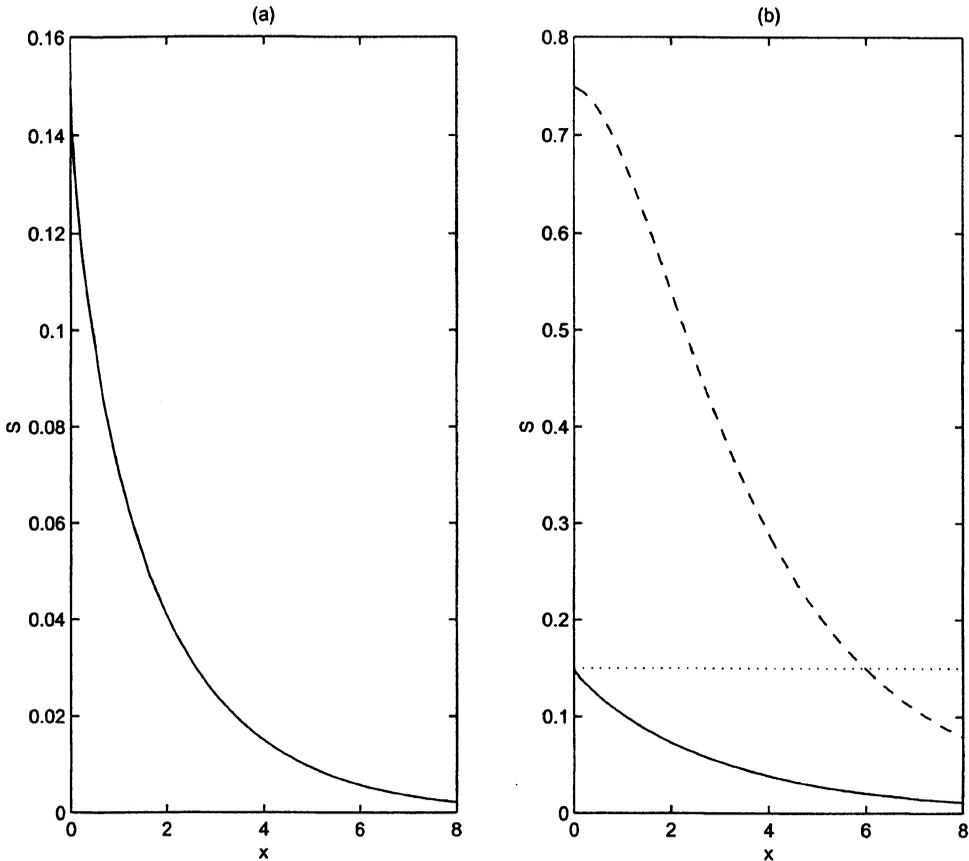


FIGURE 2. The shape of the cuspon for $\alpha = 2.0$, $\omega_0 = 0.1$, $\omega = -0.5$, and (a) $\kappa = 0.5$, i.e., in the general case, and (b) $\kappa = 0.1$, i.e., for small κ . In the case (b) we also show the usual gap soliton (by the dashed line), the part of which above the critical value $S = S_{\text{cr}}$ (shown by the dotted line) should be removed and the remaining parts brought together to form the peakon corresponding to $\rho_0 \kappa^2 \rightarrow 0$.

Eq. (1.10) is a structural perturbation which drastically changes the character of the soliton solutions.

In view of the above results concerning the cuspons, we restrict our discussion here to the most interesting case when $S(x)$ is an even function of x , while $\psi(x) = \phi(x) - \pi$ and $U_3(x)$ are odd functions. In principle, one can use the relations (2.9) and (2.12) to eliminate ϕ and U_3 and so obtain a single equation for S (a counterpart to Eq. (2.14)), as it was done above when $\sigma_3 = 0$. However, when $\sigma_3 \neq 0$, this cannot be done explicitly. Instead, we shall develop an asymptotic analysis valid for $x \rightarrow 0$, which will be combined with results obtained by direct numerical integration of Eqs. (2.7) and (2.8), subject, of course, to the constraints (2.9) and (2.12). Since singularities only arise at the center of the soliton (i.e., at $x = 0$) when $\sigma_3 = 0$, it is clear that the introduction of a small $\sigma_3 \neq 0$ will produce

only a small deformation of the soliton solution in the region where x is bounded away from zero.

First, we consider regular solitons. Because the left-hand side of Eq. (2.9) is not singular at any x , including the point $x = 0$ when $\sigma_3 = 0$, we expect that regular solitons survive a perturbation induced by $\sigma_3 \neq 0$. Indeed, if there exists a regular soliton, with $S_0 \equiv S(x = 0)$, and $\phi(x = 0) = \pi$ and $U_3(x = 0) = 0$, it follows from Eq. (2.12) that the soliton's amplitude remains exactly the same as it was for $\sigma_3 = 0$, due to the fact that the regular soliton has $U_3(x = 0) = 0$.

Next, we turn to the possibility of singular solutions, that is, cuspons or peakons. Since we are assuming that $S_0 = S(x = 0)$ is finite, and that $\phi(x = 0) = \pi$, it immediately follows from Eq. (2.9) that when $\sigma_3 \neq 0$, U_3 must remain finite for all x , taking some value $U_0 \neq 0$, say, as $x \rightarrow +0$. As it has been established above that U_3 is an odd function of x , and $U_3(x = +0) \equiv U_0 \neq 0$, there must be a discontinuity in U_3 at $x = 0$, i.e., a jump from U_0 to $-U_0$. This feature is in marked contrast to the cuspons for which U_3 is infinite at the center, see Eq. (2.26). Further, it then follows from Eq. (2.8) that, as $x \rightarrow 0$, there is also a discontinuity in dS/dx , with a jump from $2\kappa U_0 \sqrt{S_0}$ to $-2\kappa U_0 \sqrt{S_0}$. Hence, if we can find soliton solutions of this type, with $U_0 \neq 0$, they are necessarily *peakons*, and we infer that cuspons do *not* survive the structural perturbation induced by $\sigma_3 \neq 0$.

Further, if we assume that $U_0 \neq 0$, then Eq. (2.9), taken in the limit $x \rightarrow 0$, immediately shows that

$$(2.27) \quad 2\alpha(S_{\text{cr}} - S_0) = \sigma_3 U_0^2$$

(recall that S_{cr} is defined by Eq. (2.18)). Next, the Hamiltonian relation (2.12), also taken in the limit $x \rightarrow 0$, shows that

$$(2.28) \quad -\frac{\rho_0}{\alpha} S_0 - \alpha^2 S_0 (S_{\text{cr}} - S_0) = \frac{1}{2} \sigma_3 U_0^4,$$

where we have used Eq. (2.27) (recall that ρ_0 is defined by Eq. (2.21)). Elimination of U_0 from (2.27, 2.28) yields a quadratic equation for S_0 , whose positive roots represent the possible values of the peakon's amplitude.

We recall that for a cuspon which exists at $\sigma_3 = 0$ one has $\rho_0 > 0$, i.e., the amplitude of the corresponding formal regular soliton exceeds the critical value of the amplitude, see Eq. (2.21). Then, if we retain the condition $\rho_0 > 0$, it immediately follows from Eqs. (2.27) and (2.28) that no peakons may exist if the SPM coefficient in Eq. (1.10) is positive, $\sigma_3 > 0$. Indeed, Eq. (2.27) shows that $S_{\text{cr}} - S_0 > 0$ if $\sigma_3 > 0$, which, along with $\rho_0 > 0$, leads to a contradiction in the relation (2.28).

Further, it is easy to see that a general condition for the existence of peakons following from Eqs. (2.27) and (2.28) is

$$(2.29) \quad \sigma_3 \rho_0 < 0,$$

hence peakons are possible if $\sigma_3 < 0$, or if we keep $\sigma_3 > 0$ but allow $\rho_0 < 0$. In the remainder of this subsection, we will show that peakons may exist only if $\rho_0 > 0$. Hence, it follows from the necessary condition (2.29) that peakons may indeed be possible solely in the case $\sigma_3 < 0$. On the other hand, regular solitons do exist in the case $\sigma_3 > 0$ (i.e., in particular, in nonlinear-optics models), as they have $U_0 = 0$, hence neither Eq. (2.27) nor its consequence in the form of the inequality (2.29) apply to regular solitons. The existence of (stable) peakons for $\sigma_3 < 0$, and

of (also stable) regular solitons for $\sigma_3 > 0$ will be confirmed by direct numerical results presented in the next section.

To obtain a necessary condition (which will take the form of $\rho_0 > 0$) for the existence of the peakons, we notice that existence of any solitary wave implies the presence of closed dynamical trajectories in the phase plane of the corresponding dynamical system, which is here based on the ordinary differential equations (2.7) and (2.8), supplemented by the constraint (2.9). Further, at least one stable fixed point (FP) must exist inside such closed trajectories, therefore the existence of such a stable FP is, finally, a necessary condition for the existence of any solitary wave.

The FP is found by equating to zero the right-hand sides of Eq. (2.7) and (2.8), which together with Eq. (2.9) give three equations for the three coordinates ϕ , S and U_3 of the FP. First of all, one can find a trivial unstable FP of the dynamical system,

$$\cos \phi = -\frac{\omega + \kappa^2/(\omega_0 - \omega)}{1 + \kappa^2/(\omega_0 - \omega)}, \quad S = 0,$$

which does not depend on σ_3 . Then, three nontrivial FPs can be found, with their coordinates ϕ_* , S_* and U_{3*} given by the following expressions:

$$(2.30) \quad \phi_*^{(1)} = \pi, \quad S_*^{(1)} = \frac{1 - \omega}{\alpha^2} = \frac{1}{2}S_{01}, \quad U_{3*}^{(1)} = 0,$$

$$(2.31) \quad \phi_*^{(2)} = \pi, \quad (2 - \sigma_3)S_*^{(2)} = 2S_{\text{cr}} - \frac{\sigma_3}{2}S_{01}, \quad (2 - \sigma_3) \left[\alpha U_{3*}^{(2)} \right]^2 = \rho_0 - \alpha^3 S_{\text{cr}},$$

$$(2 - \sigma_3)S_*^{(3)} = 2S_{\text{cr}} - \frac{1}{2}\sigma_3 S_{01} + \frac{\kappa^2}{\alpha}, \quad (2 - \sigma_3) \left[\alpha U_{3*}^{(3)} \right]^2 = \rho_0 - \alpha^3 S_{\text{cr}} - \alpha^2 \kappa^2,$$

$$(2.32) \quad \cos \left(\phi_*^{(3)} / 2 \right) = -\frac{1}{2} \kappa U_{3*}^{(3)} / \sqrt{S_*^{(3)}}.$$

To be specific, we now consider the case of most interest, when both $S_{01} > 0$ and $S_{\text{cr}} > 0$. In this case, the FP given by Eqs. (2.30) exists for all σ_3 and all ρ_0 . However, for small σ_3 (in fact $\sigma_3 < 2$ is enough) and small κ , the FPs given by Eqs. (2.31) and (2.32) exist only when $\rho_0 > 0$. Indeed, they exist only for $\rho_0 > \alpha^3 S_{01}$ and $\rho_0 > \alpha^3 S_{01} + \kappa^2$, respectively, or, on using the definition (2.21) of ρ_0 , when $S_{01} > 2S_{\text{cr}}$ and $S_{01} > 2S_{\text{cr}} + \kappa^2/\alpha$, respectively.

Let us first suppose that $\rho_0 < 0$. Then there is only the single non-trivial FP, namely the one given by Eqs. (2.30). This FP is clearly associated with the regular solitons, whose squared amplitude is S_{01} . Hence, we infer that for $\rho_0 < 0$ there are no other solitary-wave solutions, and in particular, no peakons (and no cuspons either when $\sigma_3 = 0$, in accordance with what we have already found in subsection 2.3 above). Combining this with the necessary condition (2.29) for the existence of peakons, we infer that there are no peakons when $\sigma_3 > 0$, thus excluding peakons from applications to the nonlinear-optics models, where this SPM coefficient is positive. However, peakons may occur in density-stratified fluid flows, where there is no inherent restriction on the sign of σ_3 . This case is considered below, but first we note that in the case $\rho_0 < 0$ and $\sigma_3 > 0$ (which includes the applications to nonlinear optics), the same arguments suggest that there may be *periodic* solutions with peakon-type discontinuities; indeed, our numerical solutions of the system (2.7,2.8) (not displayed in this paper) show that this is the case.

Next, we suppose that $\rho_0 > 0$. First, if $S_{01} < 2S_{\text{cr}}$, then there again exists the single non-trivial FP given by (2.30). But now, by analogy with the existence of

cuspons when $\rho_0 > 0$ and $\sigma_3 = 0$, we infer that the solitary-wave solution which is associated with this fixed point is a *peakon*, whose squared amplitude S_0 for small σ_3 is close to S_{cr} , while the FP has $S_*^{(1)} = S_{01}/2 < S_{\text{cr}}$.

If, on the other hand, $S_{01} > 2S_{\text{cr}}$, the FPs given by Eqs.(2.31) and (2.32) become available as well. We now infer that the peakon solitary-wave solution continues to exist, and for sufficiently small σ_3 and κ it is associated with the FP given by Eq. (2.31). Although Eq. (2.31) implies that $S_*^{(2)} \approx S_{\text{cr}}$, and the peakon's squared amplitude S_0 , determined by Eqs. (2.27) and (2.28), is also approximately equal to S_{cr} , we nevertheless have $S_0 > S_*^{(2)}$ as required. Note that, in the present case, the FPs given by Eqs. (2.30) and (2.32) lie outside the peakon's homoclinic orbit. In Fig. 3, we show a plot of a typical peakon obtained, in this case, by numerical solution of Eqs. (2.7) and (2.8).

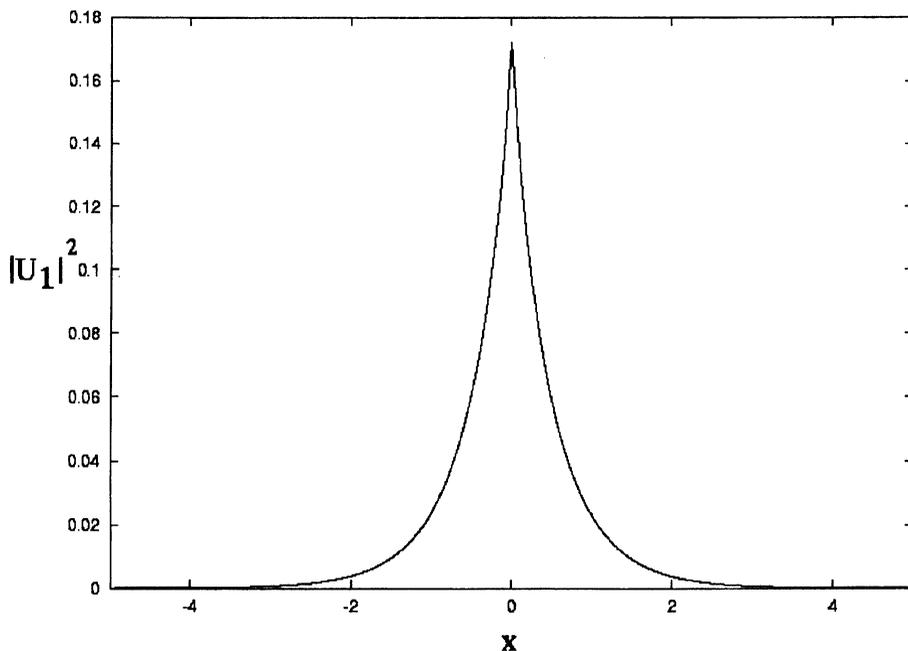


FIGURE 3. The shape of the peakon in for the case when $\sigma_3 < 0$. The parameters are $\sigma_3 = -0.01$, $\kappa = 0.1$, $\alpha = 2.0$, $\omega_0 = 0.1$, and $\omega = -0.5$. In this case, $\rho_0 = 4.8$.

3. Numerical results

3.1. Simulation techniques. The objectives of direct numerical simulations of the underlying equations (1.8) - (1.10) were to check the dynamical stability of regular solitons, cuspons, and peakons in the case $\sigma_3 = 0$, and the existence and stability of peakons in the more general case, $\sigma_3 \neq 0$. Both finite-difference and pseudo-spectral numerical methods have been used, in order to check that identical results are obtained by methods of both types. We used semi-implicit Crank-Nicholson schemes, in which the nonlinear terms were treated by means of the Adams-Bashforth algorithm.

The presence of singularities required a careful treatment of cusp and peakon solutions. To avoid numerical instabilities due to discontinuities, we sometimes introduced a weak artificial high-wavenumber viscosity into the pseudospectral code. We have found that viscosities $\sim 10^{-5}$ were sufficient to avoid the Gibbs' phenomenon in long-time simulations. When instabilities occur at a singular point (cusp or peak), it is hard to determine whether the instability is a real one or a numerical artifact. Therefore, we checked the results by means of a finite-difference code which used an adaptive staggered grid; motivated by the analysis of the vicinity of the point $x = 0$ reported above, we introduced the variable $\xi \equiv x^{2/3}$ to define an adaptive grid, and also redefined $U_3 \equiv \sqrt{\xi}\tilde{U}_3$. In these variables, the cusp seems like a regular point. We stress that this approach was solely used to check the possible occurrence of numerical instabilities.

In the following subsections we present typical examples of the numerical results for both cases considered above, viz., $\sigma_3 = 0$ and $\sigma_3 < 0$, when, respectively, the cuspons and peakons are expected.

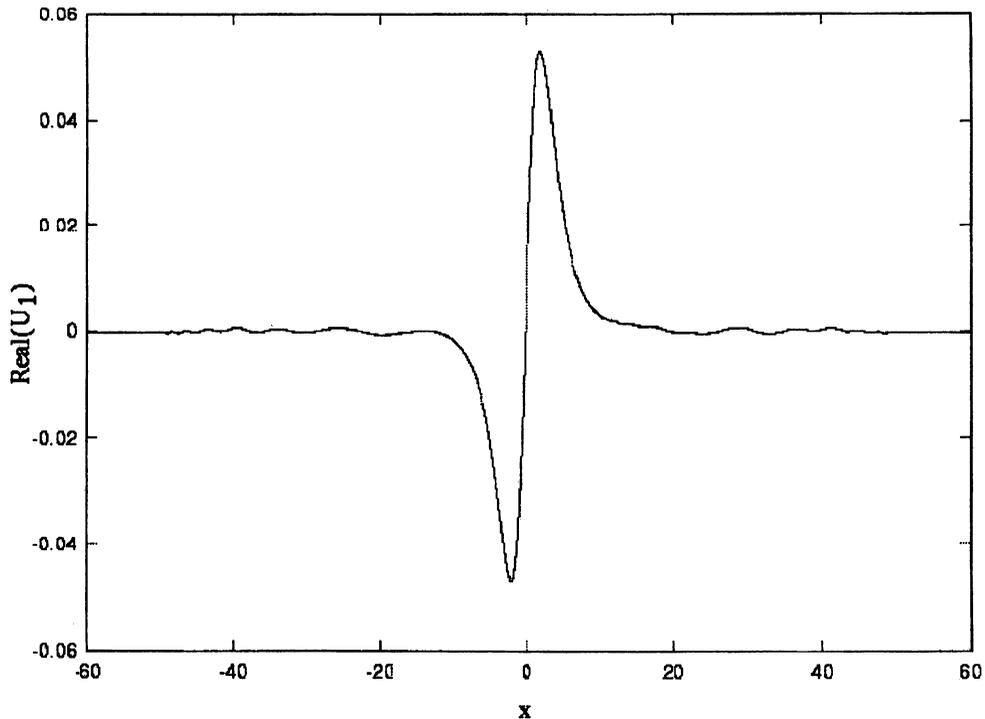


FIGURE 4. The shape of an initially perturbed regular soliton in the case $\sigma_3 = 0$ at $t = 5$, which illustrates the stabilization of the soliton via the shedding of small-amplitude radiation waves. The plot displays the field $\text{Re} U_1(x)$. The parameters are $\kappa = 0.01$, $\alpha = 1.0$, $\omega_0 = 0.2$, and $\omega = 0.9$.

3.2. The case $\sigma_3 = 0$. First, we report results obtained for the stability of regular solitary waves in the case $\sigma_3 = 0$. As initial configurations, we used the corresponding stationary solutions to Eqs. (2.7) and (2.8). To test the stability

of the regular solitary waves, we added small perturbations to them. As could be anticipated, the regular solitary wave sheds off a small-amplitude dispersive wave (radiation) and relaxes to a stationary soliton, see Fig. 4. If, however, regular solitons are taken at parameter values close to the border of the cuspon region, an initial perturbation does not make the soliton unstable, but it excites persistent internal vibrations in the soliton, see an example in Fig. 5. These and many other simulations clearly show that the regular soliton is *always stable*, and, close to the parameter border with cuspons, it has a persistent *internal mode*.

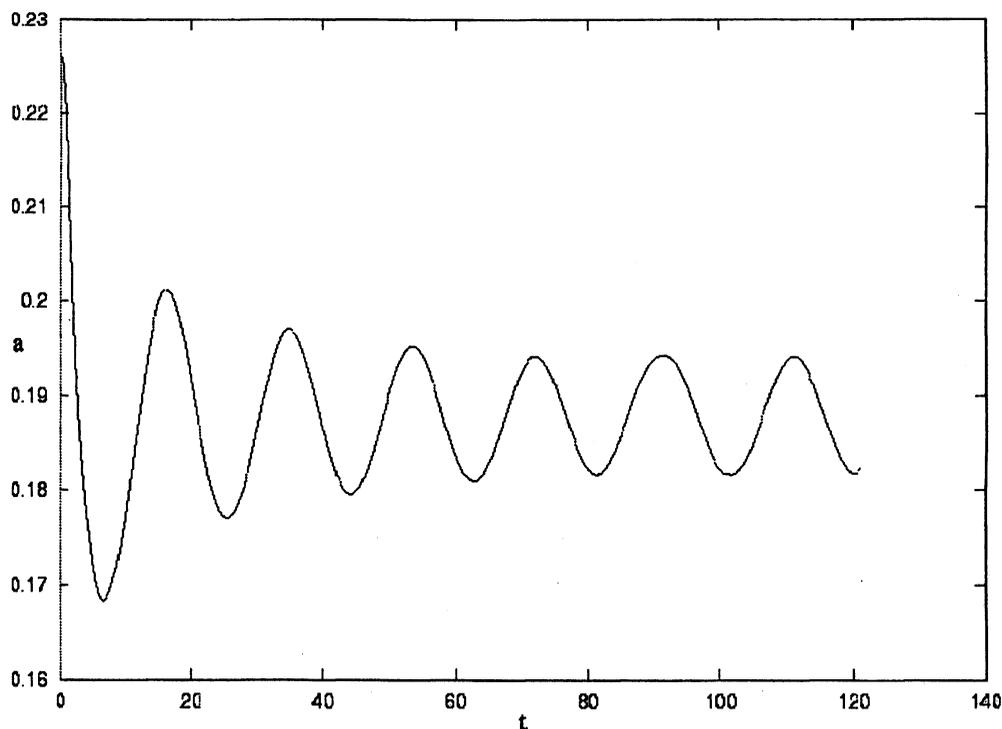


FIGURE 5. Internal vibrations of an initially-perturbed regular soliton, which was taken close to the border of the cuspon region. The plot shows the squared amplitude $a \equiv |U_1(x=0)|^2$ of the $U_1(x)$ field vs. time. The parameters are $\kappa = 0.01$, $\alpha = 1.9$, $\omega_0 = 1.5$, and $\omega = 0.5$, with $\rho_0 = 0.095$ [see Eq. (2.21)].

It was shown analytically above that Eqs. (2.4) and (2.5) (with $\sigma_3 = 0$) support peakons when $\rho_0 > 0$ and $\rho_0\kappa^2$ is very small. Direct simulations show that the peakons do exist in this case, and are *stable*. In Fig. 6, we display the time evolution of a typical stable peakon.

An essential result revealed by the simulations is that cuspons may also be *stable*, a typical example being displayed in Fig. 7. A moving weak singularity seen in this figure is, actually, a small shock wave which is initially generated at the cuspon's crest. It seems plausible that this shock wave is generated by some initial perturbation which could be a result of the finite mesh size in the finite-difference numerical scheme employed for the simulations. We have observed that

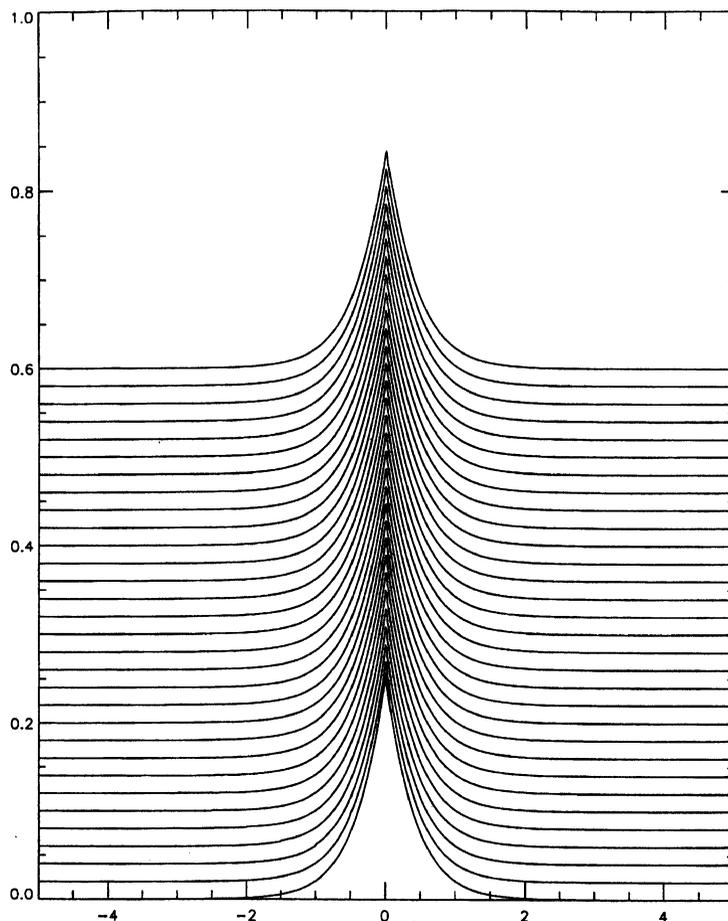


FIGURE 6. An example of a stable peakon. The plot shows the field $\text{Im } U_1$ vs. x and t . The parameters are $\kappa = 1.0$, $\alpha = 1.95$, $\omega_0 = 1.5$, and $\omega = 0.5$, with $\rho_0 = 0.04875$.

the emission of a small-amplitude shock wave is quite a typical way of the relaxation of both cuspons and peakons to a final stable state.

However, unlike the regular solitons, which were found to be always stable, the cuspons are sometimes unstable. Typically, their instability triggers onset of the spatiotemporal collapse, i.e., formation of a singularity in a finite time (see a discussion of the feasible collapse in systems of the present type, given in the Introduction). Simulations of the collapse were possible with the use of an adaptive grid. A typical example of the collapse is shown in Fig. 8, the inset illustrating the fact that the amplitude of the solution indeed diverges in a finite time. In some other cases, which are not displayed here, the instability of peakons could be quite weak, giving rise to their rearrangement into regular solitons by shedding small amounts of radiation.

3.3. The case $\sigma_3 \neq 0$. The predictions of the analysis developed above for the most general case, when the SPM terms are present in the model ($\sigma_3 \neq 0$),

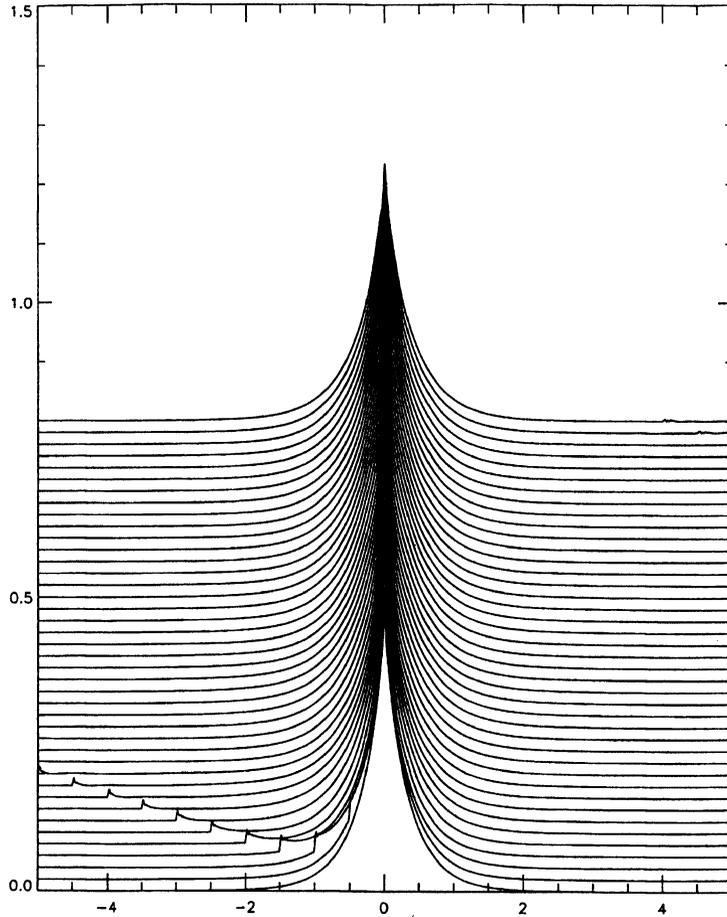


FIGURE 7. An example of a stable cuspon. The plot shows the field $\text{Im } U_1$ vs. x and t . The parameters are $\kappa = 1.0$, $\alpha = 1.0$, $\omega_0 = 1.5$, and $\omega = 0.5$, with $\rho_0 = 0.5$. The moving “defect” is a small-amplitude shock wave.

were also checked against direct simulations. As a result, we have found, in accord with the predictions, that only regular solitons exist in the case $\sigma_3 > 0$, while in the case $\sigma_3 < 0$ both regular solitons and peakons have been found as generic solutions. Further simulations, details of which are not shown here, demonstrate that both regular solitons *and* peakons are stable in this case.

4. Conclusion

In this work, we have introduced a generic model of three waves coupled by linear and nonlinear terms, which describes a situation when three dispersion curves are close to intersection at one point. The model was cast into the form of a system of two waves with opposite group velocities that, by itself, gives rise to the usual gap solitons, which is further coupled to a third wave with the zero group velocity (in the laboratory reference frame). Situations of this type are possible in various models of nonlinear optics and density-stratified fluid flows. The consideration was focussed

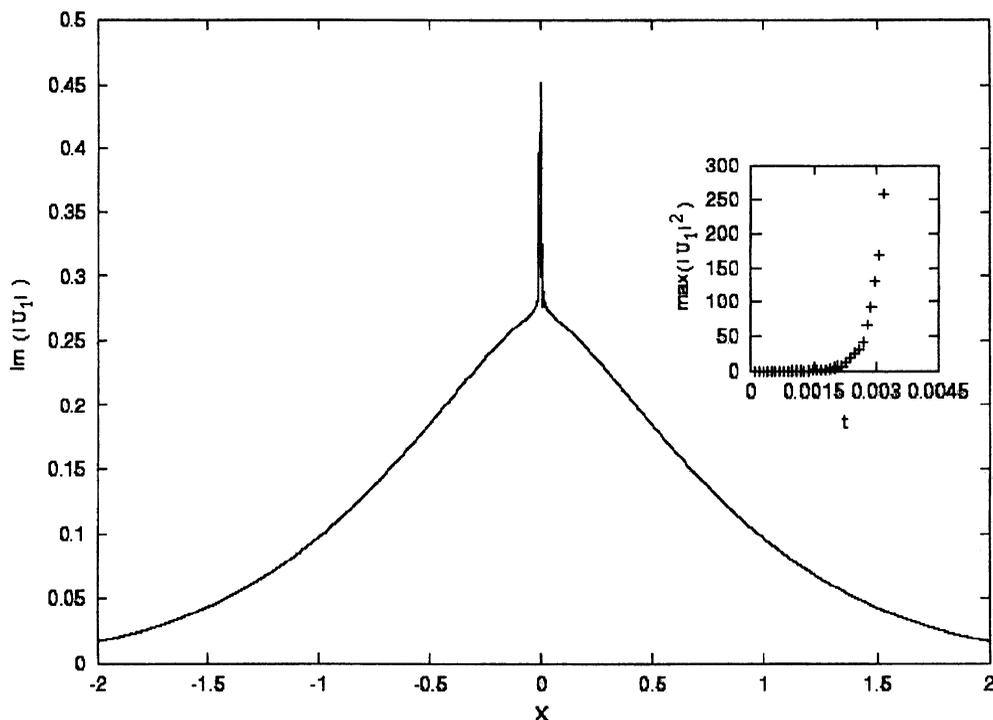


FIGURE 8. The spatial profile is shown for an unstable (collapsing) cuspon in terms of $\text{Im } U_1$ at $t = 10^{-3}$. The transition to collapse is additionally illustrated by the inset which shows the growth of the amplitude of the field $|U_1|^2$ with time. The parameters are $\kappa = 0.01$, $\alpha = 1.1$, $\omega_0 = 0.1$, and $\omega = -0.3$, with $\rho_0 = 2.618$.

on zero-velocity solitons. In a special case when the self-phase modulation (SPM) is absent in the equation for the third wave, soliton solutions were found in an exact form. It was shown that there are two coexisting generic families of solitons: regular solitons and cuspons. In the special case when the coefficient of the linear coupling between the first two waves and the third one vanishes, cuspons are replaced by peakons. Direct simulations have demonstrated that the regular solitons are stable (in the case when the regular soliton is close to the border of the cuspon region, it has a persistent internal mode). The cuspons and peakons may be both stable and unstable. If they are unstable, they either shed off some radiation and rearrange themselves into regular solitons, or, in most typical cases, the development of the cuspon's instability initiates onset of spatiotemporal collapse. To the best of our knowledge, the present system gives the first explicit example of the collapse in one-dimensional gap-soliton models.

The most general version of the model, which includes the self-phase modulation term in the equation for the third wave, has also been considered. Analysis shows that cuspons cannot exist in this case, i.e., cuspons, although being dynamically stable, are structurally unstable. However, depending on the signs of the SPM coefficient and some combination of the system's parameters, it was shown that a generic family of peakon solutions may exist instead. In accord with this prediction,

the peakons have been found in direct simulations. The peakons, as well as the regular solitons, are stable in the system including the SPM term.

The next step in the study of this system should be consideration of moving solitons, which is suggested by the well-known fact that the usual two-wave model gives rise to moving gap solitons too [1]. However, in contrast to the two-wave system, one may expect a drastic difference between the zero-velocity and moving solitons in the present three-wave model. This is due to the reappearance of a derivative term in Eq. (1.10), when it is written for a moving soliton, hence solitons which assume a singularity or jump in the U_3 component, i.e., both cuspons and peakons, cannot exist if the velocity is different from zero. Nevertheless, one may expect that slowly moving solitons will have approximately the same form as the cuspons and peakons, with the singularity at the central point replaced by a narrow transient layer with a large gradient of the U_3 field. However, detailed analysis of the moving solitons is beyond the scope of this work.

Acknowledgements

We would like to thank T. Bridges, G. Derks and S. Reich for valuable discussions. B.A.M. and G.A.G. appreciate hospitality of the University of Loughborough (UK). B.A.M. acknowledges financial support from the American Mathematical Society. The work of G.A.G. is supported by a European Commission Grant, contract number HPRN-CT-2000-00113, for the Research Training Network *Mechanics and Symmetry in Europe* (MASIE).

References

- [1] C.M. de Sterke and J.E. Sipe, in *Progress in Optics*, edited by E. Wolf (Elsevier, North-Holland, 1994), Vol. XXXIII, Chap. 3, pp. 205.
- [2] R. Grimshaw, J. He, and B.A. Malomed, *Physica D* **113**, 26 (1998).
- [3] D.N. Christodoulides and R.I. Joseph, *Phys. Rev. Lett.* **2**, 1746 (1989).
- [4] A. Aceves and S. Wabnitz, *Phys. Lett. A* **141**, 37 (1989).
- [5] B.A. Malomed, *Phys. Rev. E* **51**, R864 (1995).
- [6] E.A. Kuznetsov and A.V. Mikhailov, *Teor. Mat. Fiz.* **30**, 193 (1977); D.J. Kaup and A.C. Newell, *Lett. Nuovo Cimento* **20**, 325 (1977).
- [7] N.G.R. Broderick, D.J. Richardson, R.I. Laming, and M. Ibsen, *Opt. Lett.* **23**, 328 (1998).
- [8] B.J. Eggleton *et al.*, *Phys. Rev. Lett.* **76**, 1627 (1996); C.M. de Sterke, B.J. Eggleton, and P.A. Krug, *J. Lightwave Technol.* **15**, 1494 (1997).
- [9] L. Bergé, *Phys. Rep.* **303**, 259 (1998).
- [10] R. Grimshaw and B.A. Malomed, *Phys. Rev. Lett.* **72**, 949 (1994).
- [11] J. Gear and R. Grimshaw, *R. Stud. Appl. Math.*, **70**, 235 (1984).
- [12] T. Opatrný, B.A. Malomed, and G. Kurizki, *Phys. Rev. E* **60**, 6137 (1999).
- [13] W.C.K. Mak, B.A. Malomed, and P.L. Chu, *Phys. Rev. E* **58**, 6708 (1998).
- [14] R. Grimshaw and B.A. Malomed, *Phys. Lett. A* **198**, 205 (1995).
- [15] H. Aratyn, *J. Phys. A* **14**, 1313 (1981).
- [16] T. Peschel *et al.*, *Phys. Rev. E* **55**, 4730 (1997); J. Yang, B.A. Malomed, and D.J. Kaup, *Phys. Rev. Lett.* **83**, 1958 (1999).
- [17] R. Camassa and D.D. Holm, *Phys. Rev. Lett.* **71**, 1661 (1993).
- [18] R.A. Kraenkel and A. Zenchuk, *J. Phys. A* **32**, 4733 (1999); M.C. Ferreira, R.A. Kraenkel, and A.I. Zenchuk, *J. Phys. A* **32**, 8665 (1999).
- [19] R. Beals and D.H. Sattinger, *Inverse Problems* **15**, L1 (1999).
- [20] T. Qian and M. Tang, *Chaos, Solitons and Fractals* **12**, 1347 (2001).
- [21] I.V. Barashenkov, D.E. Pelinovsky, and E.V. Zemlyanaya, *Phys. Rev. Lett.* **80**, 5117 (1998); A. De Rossi, C. Conti, and S. Trillo, *ibid.* **81**, 85 (1998).

DEPARTMENT OF MATHEMATICAL SCIENCES, LOUGHBOROUGH UNIVERSITY, LOUGHBOROUGH,
LE11 3TU, UK

E-mail address: e-mail R.H.J.Grimshaw@lboro.ac.uk

DEPARTMENT OF MATHEMATICS AND STATISTICS, SURREY UNIVERSITY, GUILDFORD, GU2
7XH, UK

E-mail address: e-mail g.gottwald@surrey.ac.uk

DEPARTMENT OF INTERDISCIPLINARY STUDIES, FACULTY OF ENGINEERING, TEL AVIV UNIVER-
SITY, TEL AVIV 69978, ISRAEL

E-mail address: malomed@eng.tau.ac.il

This page intentionally left blank

First integrals and gradient flow for a generalized Darboux-Halphen system

S. Chakravarty and R.G. Halburd

ABSTRACT. First integrals are explicitly constructed for a third-order system of ODEs that arises as a reduction of the self-dual Yang-Mills equations and in the theory of hypercomplex manifolds. These first integrals are branched functions of the phase space variables, even in cases for which the general solution is single-valued. This branching is characterized in terms of the monodromy of the hypergeometric equations. The first integrals are then used to formulate a Nambu-Poisson structure of the system. A representation of the generalized Darboux-Halphen system as a gradient flow is also given.

1. Introduction

The system

$$(1.1) \quad \begin{aligned} \dot{\omega}_1 &= \omega_2\omega_3 - \omega_1(\omega_2 + \omega_3) + \tau^2, \\ \dot{\omega}_2 &= \omega_3\omega_1 - \omega_2(\omega_3 + \omega_1) + \tau^2, \\ \dot{\omega}_3 &= \omega_1\omega_2 - \omega_3(\omega_1 + \omega_2) + \tau^2, \end{aligned}$$

where

$$\tau^2 = \alpha^2(\omega_1 - \omega_2)(\omega_3 - \omega_1) + \beta^2(\omega_2 - \omega_3)(\omega_1 - \omega_2) + \gamma^2(\omega_3 - \omega_1)(\omega_2 - \omega_3),$$

was first studied by Halphen [14] as a natural generalization of the classical Darboux-Halphen (DH) system, which corresponds to setting $\tau = 0$ in equation (1.1). The classical Darboux-Halphen system first arose in Darboux's study of triply orthogonal surfaces [11] and was later solved by Halphen [15]. The classical DH system has also appeared in studies of self-dual Bianchi-IX metrics with Euclidean signature [4, 13] and in reductions of the associativity equations on a three-dimensional Frobenius manifold [12]. Furthermore, if $(\omega_1, \omega_2, \omega_3)$ is a solution of the classical Darboux-Halphen system, then

$$(1.2) \quad y := -2(\omega_1 + \omega_2 + \omega_3)$$

1991 *Mathematics Subject Classification.* Primary 34A34, 34A05; Secondary 37K10.

Key words and phrases. Darboux-Halphen, Schwarzian.

The first author was partially supported by a Grigsby-Sensenbaugh Foundation grant from the University of Colorado.

The second author was partially supported by grant number NAL/00344/G from the Nuffield Foundation.

satisfies the Chazy equation,

$$(1.3) \quad \frac{d^3y}{dt^3} = 2y \frac{d^2y}{dt^2} - \left(\frac{dy}{dt}\right)^2.$$

In [3] it was shown that y defined by equation (1.2) solves the generalized Chazy equation,

$$(1.4) \quad \frac{d^3y}{dt^3} - 2y \frac{d^2y}{dt^2} + \left(\frac{dy}{dt}\right)^2 = \frac{4}{36 - n^2} \left(6 \frac{dy}{dt} - y^2\right)^2,$$

where the ω_i 's solve the generalized Darboux-Halphen system for the special choices of parameters (α, β, γ) given by $(2/n, 2/n, 2/n)$ and $(1/3, 1/3, 2/n)$, et cyc. Note that equation (1.3) corresponds to the limit $n \rightarrow \infty$ in equation (1.4). Equations (1.3) and (1.4) were first studied by Chazy in [8, 9, 10].

The system (1.1) arises in the study of the equation

$$(1.5) \quad \dot{M} = (\text{adj } M)^T + M^T M - (\text{Tr } M)M,$$

for a 3×3 matrix valued function $M(t)$ where $\text{adj } M$ is the adjoint of M satisfying $(\text{adj } M)M = (\det M)I$, M^T is the transpose of M and the dot denotes differentiation with respect to t . Equation (1.5) was obtained in [6] as a reduction of the self-dual Yang-Mills equations with an infinite dimensional gauge group of diffeomorphisms $\text{Diff}(S^3)$ of the three-sphere. Equation (1.5) also describes a class of self-dual Weyl Bianchi IX space-times with Euclidean signature [5]. More recently, equation (1.5) was used to describe $\text{SU}(2)$ invariant hypercomplex 4-manifolds [16].

In section 2 we will review the reduction of equation (1.5) to the generalized DH system (1.1). The general solution will be constructed and a special case will be studied. In section 3, first integrals and ‘‘action-angle’’ variables are given for equation (1.1). The first integrals involve hypergeometric functions and are non-meromorphic, even in cases where the general solution is single-valued.

2. The solution of the generalized Darboux-Halphen system

In this section the solution of equation (1.5) is given by a factorization method which first appeared in [1]. The solution can also be obtained via associated linear problems. In [2], the solution was obtained via an evolving monodromy problem that arises as a reduction of the isospectral problem for the self-dual Yang-Mills equations. In [16], the solution was obtained via an isomonodromy problem which describes the Riccati solutions of the sixth Painlevé equation. Degenerate cases were discussed in [3].

We begin by decomposing the matrix M into its symmetric (M_s) and antisymmetric (M_a) parts. Furthermore, the eigenvalues of M_s are assumed to be distinct, so that it can be diagonalized using a complex orthogonal matrix. Thus we have

$$(2.1) \quad M = M_s + M_a = P(d + a)P^{-1},$$

where $P \in \text{SO}(3, \mathbf{C})$, $d = \text{diag}(\omega_1, \omega_2, \omega_3)$ with $\omega_i \neq \omega_j$, $i \neq j$, and the elements of the skew-symmetric matrix a are given by $a_{ij} = \sum_{k=1}^3 \varepsilon_{ijk} \tau_k$, where ε_{ijk} is totally skew-symmetric in its indices and $\varepsilon_{123} = 1$. Using the transformation (2.1), the

diagonal part of equation (1.5) yields the system (1.1), where $\tau^2 = \tau_1^2 + \tau_2^2 + \tau_3^2$. The skew-symmetric part gives

$$(2.2) \quad \dot{\tau}_1 = -\tau_1(\omega_2 + \omega_3), \quad \dot{\tau}_2 = -\tau_2(\omega_3 + \omega_1), \quad \dot{\tau}_3 = -\tau_3(\omega_1 + \omega_2),$$

and the off-diagonal symmetric part gives

$$(2.3) \quad \dot{P} = -Pa,$$

which is a linear equation for P .

Taking differences of equations in system (1.1) gives

$$(2.4) \quad \frac{d}{dt}(\omega_1 - \omega_2) = -2\omega_3(\omega_1 - \omega_2), \quad \text{et cyc.}$$

Using equations (2.2) and (2.4), we can solve the τ_i 's in terms of the ω_i 's as

$$(2.5) \quad \begin{aligned} \tau_1^2 &= \alpha^2(\omega_1 - \omega_2)(\omega_3 - \omega_1), \\ \tau_2^2 &= \beta^2(\omega_2 - \omega_3)(\omega_1 - \omega_2), \quad \tau_3^2 = \gamma^2(\omega_3 - \omega_1)(\omega_2 - \omega_3), \end{aligned}$$

where $\alpha, \beta,$ and γ are integration constants.

In terms of the cross ratio

$$(2.6) \quad s := \frac{\omega_1 - \omega_3}{\omega_2 - \omega_3},$$

it follows from equation (1.1) that the ω_i 's can be parameterized as

$$(2.7) \quad \omega_1 = -\frac{1}{2} \frac{d}{dt} \ln \frac{\dot{s}}{s(s-1)}, \quad \omega_2 = -\frac{1}{2} \frac{d}{dt} \ln \frac{\dot{s}}{s-1}, \quad \omega_3 = -\frac{1}{2} \frac{d}{dt} \ln \frac{\dot{s}}{s}.$$

where s satisfies

$$(2.8) \quad \frac{d}{dt} \left(\frac{\ddot{s}}{\dot{s}} \right) - \frac{1}{2} \left(\frac{\ddot{s}}{\dot{s}} \right)^2 + \frac{\dot{s}^2}{2} V(s) = 0,$$

with

$$\{s, t\} := \frac{d}{dt} \left(\frac{\ddot{s}}{\dot{s}} \right) - \frac{1}{2} \left(\frac{\ddot{s}}{\dot{s}} \right)^2$$

being the Schwarzian derivative and V is given by

$$(2.9) \quad V(s) = \frac{1 - \beta^2}{s^2} + \frac{1 - \gamma^2}{(s - 1)^2} + \frac{\beta^2 + \gamma^2 - \alpha^2 - 1}{s(s - 1)}.$$

Equation (2.8) is the Schwarzian equation, which describes the conformal mappings of the upper-half s -plane to the interior of a region of the complex sphere bounded by three regular circular arcs. If $\alpha, \beta,$ and γ are non-negative real numbers such that $\alpha + \beta + \gamma < 1$, then the angles subtended at the vertices $s = 0, s = 1,$ and $s = \infty$ of this triangle are $\alpha\pi, \beta\pi,$ and $\gamma\pi$. Furthermore, if $\alpha, \beta,$ and γ are chosen to be either reciprocals of integers or zero, then s is analytic on the interior of a circle on the complex sphere and cannot be analytically extended across this circle, which is a natural barrier [19].

The general solution of equation (2.8) is given implicitly by

$$(2.10) \quad t(s) = \frac{u_1(s)}{u_2(s)},$$

where $u_1(s)$ and $u_2(s)$ are independent solutions to the Fuchsian equation

$$(2.11) \quad \frac{d^2 u}{ds^2} + \frac{1}{4} V(s) u = 0.$$

Equation (2.11) is equivalent to the hypergeometric equation,

$$(2.12) \quad s(1-s)\frac{d^2\chi}{ds^2} + [c - (a+b+1)s]\frac{d\chi}{ds} - ab\chi = 0,$$

where $a = (1 + \alpha - \beta - \gamma)/2$, $b = (1 - \alpha - \beta - \gamma)/2$, $c = 1 - \beta$, and

$$(2.13) \quad u(s) = s^{c/2}(1-s)^{(a+b-c+1)/2}\chi(s).$$

A fifth order reduction

We will now consider the case in which M has the special form

$$(2.14) \quad M = \begin{pmatrix} M_{11} & M_{12} & 0 \\ M_{21} & M_{22} & 0 \\ 0 & 0 & M_{33} \end{pmatrix}.$$

This special form of M was considered in [5, 2], where equation (1.5) was analyzed using an associated evolving monodromy problem. Here we will show that quantities that arise naturally from this monodromy analysis can be obtained in a straightforward manner from the factorization method described above.

Consider the factorization of M given by equation (2.1) where M is given by equation (2.14). Due to special block structure of M , its symmetric part can be diagonalized by an orthogonal matrix of the form

$$(2.15) \quad P = \begin{pmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where ψ is a (generally complex) function of t to be determined. That is, $M_s = PdP^{-1}$, where $d = \text{diag}(\omega_1, \omega_2, \omega_3)$. Furthermore, the skew-symmetric part of M is unchanged by the adjoint action of P . So

$$(2.16) \quad a = P^{-1}M_aP = M_a = \frac{1}{2}(M - M^T) = \begin{pmatrix} 0 & \tau_3(t) & 0 \\ -\tau_3(t) & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $\tau_3(t) = \frac{1}{2}(M_{12} - M_{21})$. Since $\tau_1 = \tau_2 = 0$ in this case, the ω_i 's are given by equation (2.7) where s solves equation (2.8) with $\alpha = \beta = 0$. From equation (2.5) and equation (2.7), we have

$$\tau_3(t) = \frac{i\gamma}{2} \frac{\dot{s}}{\sqrt{s(s-1)}},$$

where γ is a constant. With the $\tau_3(t)$ given above, equation (2.3) can be readily integrated to give

$$\psi = \frac{i\gamma}{2} \log \left(\frac{\sqrt{s} - 1}{\sqrt{s} + 1} \right) + \psi_0,$$

where ψ_0 is a constant. Finally, the matrix M in (2.14) is reconstructed from the various components P , d and a according to equation (2.1). Note that in order to obtain any solution of equation (1.5) where M is given by (2.14) we must fix the two constants γ and ψ_0 and choose a solution to equation (2.8) with $\alpha = \beta = 0$ and the fixed value of γ .

In [5] the general solution of equation (1.5) with M of the form (2.14) was found via a different method which involved the analysis of a certain evolving monodromy problem. This led to the following combination of the matrix elements of M

$$\alpha_{\pm} = (M_{11} - M_{22}) \mp i(M_{12} + M_{21}), \quad R^2 = \alpha_+\alpha_-,$$

$$\beta_{\pm} = \omega \pm i(M_{21} - M_{12}), \quad \omega = M_{11} + M_{22} - 2M_{33},$$

together with the conserved quantity

$$a^2 = \frac{(M_{12} - M_{21})^2}{R^2 - \omega^2}.$$

In the factorization method outlined above, these variables arise naturally from the component matrices P , d and a as follows

$$R = \omega_1 - \omega_2, \quad \alpha_{\pm} = R e^{\mp 2i\psi}, \quad \omega = (\omega_1 - \omega_3) + (\omega_2 - \omega_3), \quad M_{12} - M_{21} = 2\tau_3,$$

and $a = \gamma$.

3. First Integrals

In this section, following [7], we will use the method of solution given in section 2 to construct first integrals for equation (1.1). We begin by constructing explicit first integrals for the Schwarzian equation (2.8), as the formulas are much simpler in this case. Let u_1 and u_2 be two linearly independent solutions of equation (2.11) satisfying the Wronskian condition $W(u_1, u_2) = u_1 u_2' - u_2 u_1' = 1$. Then any solution to equation (2.8) is given implicitly by

$$(3.1) \quad t(s) = \frac{J_2 u_1(s) - J_1 u_2(s)}{I_2 u_1(s) - I_1 u_2(s)},$$

where I_k and J_k , $k = 1, 2$, are constants satisfying

$$(3.2) \quad I_1 J_2 - I_2 J_1 = 1.$$

Hence any three of the constants I_1, I_2, J_1, J_2 can be taken as independent first integrals.

Differentiating equation (3.1) with respect to t gives

$$(3.3) \quad I_2 u_1 - I_1 u_2 = s^{1/2}.$$

Differentiation of equation (3.3) gives

$$(3.4) \quad I_2 u_1' - I_1 u_2' = \frac{1}{2} s^{-3/2} \ddot{s}.$$

Solving the system (3.3–3.4) for I_1 and I_2 gives

$$(3.5) \quad I_k = \frac{d\phi_k}{dt}, \quad \phi_k = s^{-1/2} u_k(s), \quad k = 1, 2.$$

The constants J_1 and J_2 are given by the solution of equations (3.1), (3.5) and the normalization condition (3.2). This yields $J_k = t I_k - \phi_k$, $k = 1, 2$.

In terms of the gDH variables, we have

$$(3.6) \quad \begin{aligned} \phi_k &= \sqrt{2} \left(\frac{(\omega_2 - \omega_3)}{(\omega_1 - \omega_2)(\omega_1 - \omega_3)} \right)^{1/2} u_k \left(\frac{\omega_1 - \omega_3}{\omega_2 - \omega_3} \right), \\ I_k &= \sqrt{2} \left(\frac{(\omega_1 - \omega_3)(\omega_1 - \omega_2)}{(\omega_2 - \omega_3)} \right)^{1/2} u_k' \left(\frac{\omega_1 - \omega_3}{\omega_2 - \omega_3} \right) \\ &\quad - (\omega_1 - \omega_2 - \omega_3) \left(\frac{(\omega_2 - \omega_3)}{2(\omega_1 - \omega_3)(\omega_1 - \omega_2)} \right)^{1/2} u_k \left(\frac{\omega_1 - \omega_3}{\omega_2 - \omega_3} \right). \end{aligned}$$

Using the new variables ϕ_k and I_k instead of the gDH variables ω_i 's, equation (1.1) can be formulated as a Hamiltonian system

$$(3.7) \quad \dot{\phi}_k = \frac{\partial H}{\partial I_k} = I_k, \quad \dot{I}_k = -\frac{\partial H}{\partial \phi_k} = 0, \quad H = \frac{I_1^2 + I_2^2}{2}, \quad k = 1, 2,$$

together with the constraint

$$(3.8) \quad \phi_1 I_2 - \phi_2 I_1 = W(u_1, u_2) = 1.$$

Although I_1 and I_2 are constant functions of t , they are multivalued functions of $\{\omega_1, \omega_2, \omega_3\}$ and of the Schwarzian variables $\{s, \dot{s}, \ddot{s}\}$. In terms of the solutions χ_1, χ_2 of the hypergeometric equation (2.12), the first integrals are given by

$$(3.9) \quad [I_1 \ I_2] = \sigma [\lambda \ 1] \begin{bmatrix} \chi_1(s) & \chi_2(s) \\ \chi_1'(s) & \chi_2'(s) \end{bmatrix},$$

where

$$\sigma(s, \dot{s}) = s^{c/2}(1-s)^{(a+b-c+1)/2}\dot{s}^{1/2}, \quad \text{and} \quad \lambda(s, \dot{s}, \ddot{s}) = \frac{a+b+1-cs}{2s(1-s)} - \frac{\ddot{s}}{2\dot{s}^2}.$$

Next we will discuss the dependence of I_1 and I_2 on s, \dot{s} , and \ddot{s} . Clearly $I_k, k = 1, 2$, is single-valued as a function of \dot{s} and has square-root branch points as a function of \dot{s} about $\dot{s} = 0$ and $\dot{s} = \infty$. In fact, the conserved quantities I_1^2 and I_2^2 are single-valued as functions of \dot{s} . Holding \dot{s} and \ddot{s} fixed, I_μ can only admit branch points at $s = 0, s = 1$, and $s = \infty$. Let γ_0 and γ_1 be two closed curves with a common base point in the finite complex s -plane enclosing the points $s = 0$ and $s = 1$ respectively, and traversed once in the positive direction. Analytic continuation of σ along γ_0 and γ_1 gives

$$\gamma_0 : \sigma \mapsto e^{i\pi c}\sigma, \quad \gamma_1 : \sigma \mapsto e^{i\pi(a+b-c)}\sigma.$$

Analytic continuation along γ_0 and γ_1 transforms the fundamental matrix of solutions of equation (2.12) according to

$$\gamma_\mu : \begin{pmatrix} \chi_1(s) & \chi_2(s) \\ \chi_1'(s) & \chi_2'(s) \end{pmatrix} \mapsto \begin{pmatrix} \chi_1(s) & \chi_2(s) \\ \chi_1'(s) & \chi_2'(s) \end{pmatrix} M_\mu, \quad \mu = 0, 1.$$

For generic values of a, b, c , and for the choice of basis solutions: $\chi_1 = F(a, b, c; s)$, $\chi_2 = F(a, b, a + b - c + 1; 1 - s)$ of the hypergeometric equation, the monodromy matrices M_μ are given by [20]

$$M_0 = \begin{pmatrix} 1 & e^{-2\pi ib} - e^{-2\pi ic} \\ 0 & e^{-2\pi ic} \end{pmatrix} \quad \text{and} \quad M_1 = \begin{pmatrix} e^{-2\pi i(a+b-c)} & 0 \\ 1 - e^{-2\pi i(a-c)} & 1 \end{pmatrix}.$$

So under analytic continuation, the first integrals I_1, I_2 transform as

$$\gamma_0 : [I_1 \ I_2] \mapsto [I_1 \ I_2] M_0 e^{i\pi c}, \quad \gamma_1 : [I_1 \ I_2] \mapsto [I_1 \ I_2] M_1 e^{i\pi(a+b-c)}.$$

Analytic continuation around $s = \infty$ is equivalent to a loop around $s = 0$ and $s = 1$. Hence the branching of the first integrals I_1 and I_2 is characterized in terms of the monodromy group for the hypergeometric equation.

The first integrals in equation (3.9) for the classical DH system ($\alpha = \beta = \gamma = 0$) are expressed in terms of the special hypergeometric equation (2.12) with $a = b = 1/2, c = 1$. In this case, the monodromy matrices are given by

$$M_0 = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad M_1 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

relative to the choice of basis $\chi_1 = F(1/2, 1/2, 1; s)$ and $\chi_2 = iF(1/2, 1/2, 1; 1 - s)$. Note that in this case, I_1 and I_2 are still branched, even though the solution itself is single-valued. The non-existence of meromorphic first integrals for the classical Darboux-Halphen system was proved in [17]. We show by explicit construction that the first integrals do indeed exist but they are non-algebraic and multi-valued.

4. Nambu-Poisson Structure and Gradient flow

The gDH system (1.1) can also be viewed as a complex dynamical system given by $(\dot{\omega}_1, \dot{\omega}_2, \dot{\omega}_3) = \mathbf{X}$ where the vector field

$$\mathbf{X} = (\omega_2\omega_3 - \omega_1\omega_2 - \omega_3\omega_1 + \tau^2, \omega_3\omega_1 - \omega_2\omega_3 - \omega_1\omega_2 + \tau^2, \omega_1\omega_2 - \omega_3\omega_1 - \omega_2\omega_3 + \tau^2).$$

The generalized DH flow given by the integral curves of \mathbf{X} lie on the intersection of the level sets of the first integrals ($I_1 = \text{constant}$ and $I_2 = \text{constant}$) in a three-dimensional complex manifold $M^3 := \mathbf{C} \setminus \{\omega_i \neq \omega_j, i \neq j\}$. Since the I_k 's are conserved under the gDH flow,

$$\frac{dI_k}{dt} = \mathbf{X} \cdot \nabla I_k = 0, \quad k = 1, 2,$$

it follows that the vector field \mathbf{X} is proportional to $\nabla I_1 \times \nabla I_2$. Explicit calculation shows that,

$$\mathbf{X} = \frac{1}{4\Delta} \nabla I_1 \times \nabla I_2$$

where

$$\Delta = (\omega_2 - \omega_3)(\omega_3 - \omega_1)(\omega_1 - \omega_2).$$

The gDH equations can be expressed as

$$\dot{\omega}_j = \mathbf{X} \cdot \nabla \omega_j = \begin{cases} (4\Delta)^{-1} \nabla I_2 \cdot (\nabla \omega_j \times \nabla I_1) =: \{\omega_j, I_1\}_1, \\ -(4\Delta)^{-1} \nabla I_1 \cdot (\nabla \omega_j \times \nabla I_2) =: \{\omega_j, I_2\}_2. \end{cases}$$

It can be verified that the brackets

$$\begin{aligned} B_1(g, h) = \{g, h\}_1 &= (4\Delta)^{-1} \nabla I_2 \cdot (\nabla g \times \nabla h), \\ B_2(g, h) = \{g, h\}_2 &= -(4\Delta)^{-1} \nabla I_1 \cdot (\nabla g \times \nabla h), \end{aligned}$$

are Poisson (i.e., they are bi-linear, anti-symmetric and satisfy the Jacobi identity). So \mathbf{X} is a Hamiltonian vector field with respect to the two Poisson structures B_1 and B_2 with Hamiltonians I_1 and I_2 respectively.

The Poisson structures B_1 and B_2 are degenerate (rank 2) and admit Casimir functions I_2 and I_1 respectively. This is easily verified by using the vector triple product identity

$$\begin{aligned} \{I_2, g\}_1 &= (4\Delta)^{-1} \nabla I_2 \cdot (\nabla I_2 \times \nabla g) = (4\Delta)^{-1} \nabla g \cdot (\nabla I_2 \times \nabla I_2) = 0, \\ \{I_1, g\}_2 &= (4\Delta)^{-1} \nabla I_1 \cdot (\nabla I_1 \times \nabla g) = (4\Delta)^{-1} \nabla g \cdot (\nabla I_1 \times \nabla I_1) = 0, \end{aligned}$$

for any smooth function g on M . Furthermore, since $\{I_1, I_2\}_1 = \{I_1, I_2\}_2 = 0$, the integrals of the gDH system are in involution. The Poisson structures B_1 and B_2 are compatible in the sense that

$$B := \lambda_1 B_1 + \lambda_2 B_2, \quad \lambda_i = \lambda_i(I_1, I_2)$$

is also a Poisson structure for gDH with a Hamiltonian $H(I_1, I_2)$ satisfying

$$\lambda_1 \frac{\partial H}{\partial I_1} + \lambda_2 \frac{\partial H}{\partial I_2} = 0.$$

Thus the gDH system is *bi-Hamiltonian*.

The symmetric representation of the gDH system using *both* Hamiltonians I_1, I_2 is

$$\dot{\omega}_j = (4\Delta)^{-1} \nabla \omega_j \cdot (\nabla I_1 \times \nabla I_2) = (4\Delta)^{-1} \frac{\partial(\omega_i, I_1, I_2)}{\partial(\omega_1, \omega_2, \omega_3)} =: \{\omega_j, I_1, I_2\}.$$

This is an example of a Nambu-Poisson bracket similar to rigid body dynamics in three dimensions [18, 21].

The Darboux-Halphen system (1.1) is also a gradient flow. In terms of local coordinates ω_i , it can be written as

$$\dot{\omega}_i = \sum_{j=1}^3 g^{ij} \frac{\partial \Phi}{\partial \omega_j},$$

where g^{ij} is a constant contravariant metric and Φ is a potential function. The metric is given by

$$(g^{ij}) = \begin{pmatrix} m(\alpha, \beta, \gamma) & \kappa + 4\gamma^2 & \kappa + 4\beta^2 \\ \kappa + 4\gamma^2 & m(\beta, \gamma, \alpha) & \kappa + 4\alpha^2 \\ \kappa + 4\beta^2 & \kappa + 4\alpha^2 & m(\gamma, \alpha, \beta) \end{pmatrix},$$

where

$$\kappa = (\alpha + \beta + \gamma)(\alpha + \beta - \gamma)(\alpha - \beta + \gamma)(\alpha - \beta - \gamma) - 1,$$

and

$$m(\alpha, \beta, \gamma) = (1 - \alpha^2 + (\beta + \gamma)^2)(1 - \alpha^2 + (\beta - \gamma)^2).$$

The potential Φ is a homogeneous polynomial of degree 3 in the ω_i 's and is invariant under the *simultaneous* cyclic permutation of $\{\omega_1, \omega_2, \omega_3\}$ and $\{\alpha, \beta, \gamma\}$. In terms of the function

$$F(\omega_1, \omega_2, \omega_3; \alpha, \beta, \gamma) = [(1 - \alpha^2)(3\alpha^2 - 2\beta^2 - 2\gamma^2 + 1) + (\beta^2 - \gamma^2)^2] \times \omega_1 [\alpha^2 \omega_1^2 + 3\beta^2 \omega_2^2 + 3\gamma^2 \omega_3^2 + (1 - \alpha^2 - 3\beta^2 - 3\gamma^2)\omega_2 \omega_3],$$

the potential is expressed as

$$\Phi = -\frac{4}{3 \det(g^{ij})} [F(\omega_1, \omega_2, \omega_3; \alpha, \beta, \gamma) + F(\omega_2, \omega_3, \omega_1; \beta, \gamma, \alpha) + F(\omega_3, \omega_1, \omega_2; \gamma, \alpha, \beta)].$$

With respect to the metric g^{ij} , the constant Φ surfaces are orthogonal to the curves obtained by the intersection of the constant I_1 and I_2 surfaces.

In the classical Darboux-Halphen case ($\alpha = \beta = \gamma = 0$), we have

$$g^{ij} = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix} \quad \text{and} \quad \Phi = \omega_1 \omega_2 \omega_3.$$

For the fifth order reduction ($\alpha = \beta = 0$) discussed in section 2,

$$g^{ij} = \begin{pmatrix} (1 + \gamma^2)^2 & \gamma^4 + 4\gamma^2 - 1 & \gamma^4 - 1 \\ \gamma^4 + 4\gamma^2 - 1 & (1 + \gamma^2)^2 & \gamma^4 - 1 \\ \gamma^4 - 1 & \gamma^4 - 1 & (1 - \gamma^2)^2 \end{pmatrix}$$

and the corresponding potential function is given by

$$\Phi = \frac{\gamma^2 \omega_3^2 [3(1 - \gamma^2)(\omega_1 + \omega_2) + (1 + 3\gamma^2)\omega_3] + 3(1 - \gamma^2)^2 \omega_1 \omega_2 \omega_3}{3(1 - \gamma^2)^3}.$$

References

1. M.J. Ablowitz, S. Chakravarty, and R. Halburd, *Darboux-Halphen systems and the singularity structure of its solutions*, Proc. Fourth Int. Conf. on Mathematical and Numerical Aspects of Wave Propagation (Philadelphia) (J.A. DeSanto, ed.), SIAM, 1998, pp. 408–412.
2. ———, *On Painlevé and Darboux-Halphen type equations*, The Painlevé Property, One Century Later (Berlin) (R. Conte, ed.), CRM Series in Mathematical Physics, Springer, 1998.
3. ———, *The generalized Chazy equation from the self-duality equations*, Stud. Appl. Math. **103** (1999), 75–88.
4. M.F. Atiyah and N.J. Hitchin, *The geometry and dynamics of magnetic monopoles*, Princeton University Press, Princeton, New Jersey, 1988.
5. S. Chakravarty and M.J. Ablowitz, *Integrability, monodromy, evolving deformations, and self-dual Bianchi IX systems*, Phys. Rev. Lett. **76** (1996), 857–860.
6. S. Chakravarty, M.J. Ablowitz, and L.A. Takhtajan, *Self-dual Yang-Mills equation and new special functions in integrable systems*, Nonlinear Evolution Equations and Dynamical Systems (Singapore) (M. Boiti, L. Martina, and F. Pempinelli, eds.), World Scientific, 1992.
7. S. Chakravarty and R. Halburd, *First integrals of a generalized Darboux-Halphen system*, Loughborough University Preprint (2002).
8. J. Chazy, *Sur les équations différentielles dont l'intégrale générale est uniforme et admet des singularités essentielles mobiles*, C.R. Acad. Sc. Paris **149** (1909), 563–565.
9. ———, *Sur les équations différentielles dont l'intégrale générale possède une coupure essentielle mobile*, C.R. Acad. Sc. Paris **150** (1910), 456–458.
10. ———, *Sur les équations différentielles du troisième et d'ordre supérieur dont l'intégrale générale à ses points critiques fixés*, Acta Math. **34** (1911), 317–385.
11. G. Darboux, *Sur la théorie des coordonnées curvilignes et les systèmes orthogonaux*, Ann. Ec. Normale Supér. **7** (1878), 101–150.
12. B. Dubrovin, *Geometry of 2D topological field theories*, Integrable systems and quantum groups (Montecatini Terme, 1993), Springer, Berlin, 1996, pp. 120–348.
13. G.W. Gibbons and C.N. Pope, *The positive action conjecture and asymptotically Euclidean metrics in quantum gravity*, Commun. Math. Phys. **66** (1979), 267–290.
14. G. Halphen, *Sur certains systèmes d'équations différentielles*, C. R. Acad. Sci. Paris **92** (1881), 1404–1406.
15. ———, *Sur un système d'équations différentielles*, C. R. Acad. Sci. Paris **92** (1881), 1101–1103.
16. N. Hitchin, *Hypercomplex manifolds and the space of framings*, The geometric universe (Oxford, 1996), Oxford Univ. Press, Oxford, 1998, pp. 9–30.
17. A.J. Maciejewski and J. Strelcyn, *On the algebraic non-integrability of the Halphen system*, Phys. Lett. A **201** (1995), 161–166.
18. Y. Nambu, *Generalized Hamiltonian dynamics*, Phys. Rev. D (3) **7** (1973), 2405–2412.
19. Z. Nehari, *Conformal mapping*, McGraw-Hill, New York (1952).
20. J. Plemelj, *Problems in the sense of Riemann and Klein*, Interscience Publishers John Wiley & Sons Inc. New York-London-Sydney, 1964.
21. L. Takhtajan, *On foundation of the generalized Nambu mechanics*, Comm. Math. Phys. **160** (1994), 295–315.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF COLORADO, COLORADO SPRINGS, CO 80933-7150, USA

E-mail address: chuck@math.uccs.edu

DEPARTMENT OF MATHEMATICAL SCIENCES, LOUGHBOROUGH UNIVERSITY, LOUGHBOROUGH, LEICESTERSHIRE, LE11 3TU, UK

E-mail address: R.G.Halburd@lboro.ac.uk

This page intentionally left blank

Blow-ups of the Toda lattices and their Intersections with the Bruhat Cells

Luis Casian and Yuji Kodama

ABSTRACT. We study the topology of the set of singular points (blow-ups) in the solution of the nonperiodic Toda lattice defined on real split semisimple Lie algebra \mathfrak{g} . The set of blow-ups is called the Painlevé divisor. The isospectral manifold of the Toda lattice is compactified through the companion embedding which maps the manifold to the flag manifold associated with the underlying Lie algebra \mathfrak{g} . The Painlevé divisor is then given by the intersections of the compactified manifold with the Bruhat cells in the flag manifold. In this paper, we give explicit description of the topology of the Painlevé divisor for the cases of all the rank two Lie algebra, A_2, B_2, C_2, G_2 , and A_3 type. The results are obtained by using the Mumford system and the limit matrices introduced originally for the periodic Toda lattice. We also give a Lie theoretic description of the Painlevé divisor of codimension one case, and propose several conjectures for the general case.

1. Introduction

It is well-known that the generalized (nonperiodic) Toda lattices associated with semisimple Lie algebra \mathfrak{g} of rank l possess l polynomial invariants, *the Chevalley invariants*, which provide their integrability [4, 12]. The isospectral manifold determined by those polynomials defines a l -dimensional affine variety, and it can be compactified by adding the points associated with the *blow-ups* in the solution of the Toda lattice. Those points are defined as the zeros of τ -functions giving an explicit solution of the Toda lattice [13], and the set of zeros is sometimes called the *Painlevé divisor*. The number of τ -functions is given by the rank of the algebra, and each τ -function can be labeled by a dot in the corresponding Dynkin diagram. Then the Painlevé divisor consists of l components $\{\Theta_{\{k\}} : k = 1, \dots, l\}$, and each $\Theta_{\{k\}}$ is associated with a root α_k in the Dynkin diagram. As in the case of periodic Toda [1], the singularities of the divisor are canonically associated with the Dynkin diagrams, i.e. $\Theta_J = \bigcap_{k \in J} \Theta_{\{k\}}$ for a subdiagram $J \subset \{1, \dots, l\}$.

In [8], Flaschka and Haine considered a *companion* embedding map of the isospectral manifold into a flag manifold, and identified the Painlevé divisor as the

2000 *Mathematics Subject Classification.* 37K10 ; Secondary 22E46.

Both authors are supported in part by NSF Grant #DMS0071523.

intersection with certain Bruhat cells in the Bruhat decomposition,

$$G/B^+ = \bigcup_{w \in W} N^- w B^+ / B^+,$$

where G is the Lie group with $\mathfrak{g} = \text{Lie}(G)$, B^+ the Borel subgroup, N^- the unipotent subgroup and W the Weyl group of G . Then the compactification of the isospectral manifold can be obtained by gluing the Painlevé divisor in the flag manifold.

On the other hand, the real part of the compactified isospectral manifold was studied in [5], where the manifold was constructed by extending the work of Kostant in [12]. Theorem 3.2 in [12] describes part of the isospectral manifold of the Toda lattice in terms of one connected component of a split Cartan subgroup of G . There is a total of 2^l connected components which are labeled by a set of signs $\epsilon = (\epsilon_1, \dots, \epsilon_l)$. In [5] instead all the connected components of a Cartan subgroup are involved. The upshot is that now a split Cartan subgroup $H_{\mathbb{R}}$, with all its connected components, becomes an open dense subset in the compactified isospectral manifold. This manifold is then described as a union of convex polytopes Γ_{ϵ} glued as in [6], and each connected component with the sign ϵ of the Cartan subgroup is the interior of the corresponding polytope Γ_{ϵ} . The convexity of the polytope Γ_{ϵ} can be shown by Atiyah's convexity theorem [3] with the torus embedding (conjugate to the companion embedding) in the flag manifold.

In this paper, we study the topological structure of the Painlevé divisor as the blow-ups of the Toda lattice on the polytopes. In Section 2, we provide a background information necessary for the present study which includes the isospectral manifold, the companion embedding to the flag manifold, the τ -functions and the Painlevé divisor.

In Section 3, we define the *limit matrices* to parametrize the Painlevé divisor. The limit matrix was first introduced in [2] for the periodic Toda lattice for a parametrization of the Birkhoff strata of the hyperelliptic Jacobi variety, and the existence of the limit matrix was shown based on Sato's theory of universal Grassmannians. We here give a direct proof of the existence of the limit matrix by using a factorization of the unipotent subgroup N^- (Proposition 3.2), and show that the companion embedding maps the limit matrix to the corresponding Bruhat cell.

In Section 4, we define the Mumford system for the A_l Toda lattice, which may be considered as an extension of the system used to parametrize the moduli space associated with the hyperelliptic Riemann surface and its Jacobian. The Mumford system gives an explicit coordinate for the Painlevé divisor through the limit matrix. Then we prove a topological equivalence between the top cell of A_k and certain Painlevé divisor of A_j with $j > k$ (Proposition 4.2).

In Section 5, we provide several explicit results for the Toda lattices on the Lie algebra \mathfrak{g} of all rank 2 cases, A_2, B_2, C_2, G_2 , and of type A_3 .

Then in Section 6, we give a Lie theoretic description of the Painlevé divisor based on the results in [5, 6]. We first review the details of the construction of the compactified manifold by gluing the polytopes Γ_{ϵ} of the Cartan subgroup $H_{\mathbb{R}}$, and it is worth keeping in mind that $H_{\mathbb{R}}$ is not necessarily a Cartan subgroup in G but rather in another Lie group \tilde{G} defined in Notation 6.2. We then define an "algebraic" version of the Painlevé divisor, denoted by $\tilde{\Theta}_{\{i\}}^2$, in terms of the simple root character χ_{α_i} defined on the Cartan subgroup. The characters χ_{α_k} can be expressed in terms of the characters χ_{ω_i} associated to the fundamental weights

ω_i which have similar properties to the τ -functions. Then we give Conjectures 6.5 and 6.18 that $\Theta_{\{i\}}$ and $\tilde{\Theta}_{\{i\}}^a$ become homeomorphic if small modifications on $\tilde{\Theta}_{\{i\}}^a$ are introduced. These conjectures about the structure of the Painlevé divisors are verified in all the rank 2 cases as well as in A_3 discussed in Section 5. The homology of the spaces constructed in terms of the root characters is computable with the same methods used in [5]. Conjectures 6.5 and 6.18 then would allow the computation of the homology of the Painlevé divisors.

2. Toda lattices and Painlevé divisor

The generalized (nonperiodic) Toda lattice equation related to real split semisimple Lie algebra \mathfrak{g} of rank l is defined by the Lax equation, [4, 12],

$$(2.1) \quad \frac{dL}{dt} = [A, L]$$

where the Lax pair (L, A) are given by

$$(2.2) \quad \begin{cases} L(t) = \sum_{i=1}^l b_i(t)h_{\alpha_i} + \sum_{i=1}^l (a_i(t)e_{-\alpha_i} + e_{\alpha_i}) \\ A(t) = -\sum_{i=1}^l a_i(t)e_{-\alpha_i} \end{cases}$$

Here $\{h_{\alpha_i}, e_{\pm\alpha_i}\}$ is the Cartan-Chevalley basis of the algebra \mathfrak{g} with the positive simple roots $\Pi = \{\alpha_1, \dots, \alpha_l\}$ which satisfy the relations,

$$[h_{\alpha_i}, h_{\alpha_j}] = 0, \quad [h_{\alpha_i}, e_{\pm\alpha_j}] = \pm C_{j,i} e_{\pm\alpha_j}, \quad [e_{\alpha_i}, e_{-\alpha_j}] = \delta_{i,j} h_{\alpha_j},$$

where $(C_{i,j})$ is the $l \times l$ Cartan matrix of the Lie algebra \mathfrak{g} . The Lax equation (2.1) then gives

$$(2.3) \quad \begin{cases} \frac{db_i}{dt} = a_i \\ \frac{da_i}{dt} = -\left(\sum_{j=1}^l C_{i,j} b_j\right) a_i \end{cases}$$

The integrability of the system can be shown by the existence of the Chevalley invariants, $\{I_k(L) : k = 1, \dots, l\}$, which are given by the homogeneous polynomial of $\{(a_i, b_i) : i = 1, \dots, l\}$. Then in this paper we are concerned with the topology of the *real* isospectral manifold defined by

$$Z(\gamma)_{\mathbb{R}} = \{(a_1, \dots, a_l, b_1, \dots, b_l) \in \mathbb{R}^{2l} : I_k(L) = \gamma_k \in \mathbb{R}, k = 1, \dots, l\}.$$

The manifold $Z(\gamma)_{\mathbb{R}}$ can be compactified by adding the set of points corresponding to the *blow-ups* of the solution. The set of blow-ups has been shown to be characterized by the intersections with the Bruhat cells of the flag manifold G/B^+ , which are referred to as the *Painlevé divisors*, and the compactification is described in the flag manifold. In order to explain some details of this fact, we first define the set \mathcal{F}_{γ} ,

$$\mathcal{F}_{\gamma} := \{L \in e_+ + \mathcal{B}^- : I_k(L) = \gamma_k, k = 1, \dots, l\},$$

where $e_+ = \sum_{i=1}^l e_{\alpha_i}$, and \mathcal{B}^- is the Lie algebra of B^- . Then there exists a unique element $n_0 \in N^-$ such that $L \in \mathcal{F}_{\gamma}$ can be conjugated to the normal form C_{γ} ,

$L = n_0 C_\gamma n_0^{-1}$ [11]. In the case of $\mathfrak{g} = \mathfrak{sl}(l + 1, \mathbb{R})$, C_γ is the companion matrix given by

$$C_\gamma = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \ddots & 0 & 1 \\ (-1)^l \gamma_l & \cdots & \cdots & -\gamma_1 & 0 \end{pmatrix},$$

where the Chevalley invariants are given by the elementary symmetric polynomials of the eigenvalues of L . Then we define:

DEFINITION 2.1. [8]: The companion embedding of \mathcal{F}_γ is defined as the map,

$$\begin{aligned} c_\gamma : \mathcal{F}_\gamma &\longrightarrow G/B^+ \\ L &\longmapsto n_0^{-1} \bmod B^+ \end{aligned}$$

where $L = n_0 C_\gamma n_0^{-1}$ with $n_0 \in N^-$.

The isospectral manifold $Z(\gamma)_\mathbb{R}$ can be considered as a subset of \mathcal{F}_γ with the element L in the form of (2.2). The Toda lattice (2.1) then defines a flow on \mathcal{F}_γ which is embedded as follows:

PROPOSITION 2.1. [8] *The Toda flow maps to the flag manifold as*

$$\begin{aligned} c_\gamma : L(t) &\longmapsto n_0^{-1} n(t) \bmod B^+ \\ &= n_0^{-1} e^{tL^0} \bmod B^+ \end{aligned}$$

where $L^0 = n_0 C_\gamma n_0^{-1}$, and $n(t) \in N^-, b(t) \in B^+$ are given by the factorization of $e^{tL^0} = n(t)b(t)$.

This Proposition is based on the solution formula using the factorization, i.e.

$$(2.4) \quad L(t) = n(t)^{-1} L^0 n(t) = b(t) L^0 b(t)^{-1}.$$

However one should note that the factorization is not always possible, and the general form is given by the Bruhat decomposition,

$$G = \bigcup_{w \in W} N^- w B^+.$$

It has been also shown in [8, 1] that for a subset J of $\{1, \dots, l\}$ the blow-up of the solution $L(t)$ at $t = t_J$ corresponds to the case

$$e^{t_J L^0} \in N^- w_J B^+, \quad \text{where } w_J \neq id,$$

where w_J is the longest element of the Weyl subgroup W_J associated with the Dynkin diagram labeled by J . Thus the Toda flow meets only those Bruhat cells, and we see that the Painlevé divisor, denoted by \mathcal{D}_J , characterizes the intersection of the Bruhat cell corresponding to the longest element $w_J \in W$ with the compactified isospectral manifold $\tilde{Z}(\gamma)_\mathbb{R}$, i.e.

$$(2.5) \quad \mathcal{D}_J = \tilde{Z}(\gamma)_\mathbb{R} \cap N^- w_J B^+ / B^+, \quad \text{with } w_\emptyset = id.$$

Here $\tilde{Z}(\gamma)_{\mathbb{R}}$ is the closure of the image of the isospectral manifold under the companion embedding c_γ in (2.1), and it has a decomposition (intersection with the Bruhat decomposition),

$$\tilde{Z}(\gamma)_{\mathbb{R}} = \overline{c_\gamma(Z(\gamma)_{\mathbb{R}})} = \bigsqcup_{J \subset \{1, \dots, l\}} \mathcal{D}_J.$$

The analytical structure of the blow-ups can be obtained by the τ -functions, which are defined by

$$(2.6) \quad b_k = \frac{d}{dt} \ln \tau_k, \quad a_k = a_k^0 \prod_{j=1}^l (\tau_j)^{-C_{k,j}},$$

From (2.3), the tau-functions then satisfy the bilinear equations,

$$(2.7) \quad \tau_k \tau_k'' - (\tau_k')^2 = \prod_{j \neq k} (\tau_j)^{-C_{k,j}},$$

where $\tau_k'' = d^2 \tau_k / dt^2$ and $\tau_k' = d \tau_k / dt$, and $\tau_0 = 1, \tau_{l+2} = 0$. Then the Painlevé divisor \mathcal{D}_J can be defined as

$$(2.8) \quad c_\gamma(L(t)) \in \mathcal{D}_J \stackrel{\text{def}}{\iff} \tau_k(t) = 0, \text{ iff } k \in J.$$

We also define the set Θ_J as a disjoint union of $\mathcal{D}_{J'}$,

$$\Theta_J := \bigsqcup_{J' \supseteq J} \mathcal{D}_{J'}.$$

Then we have a stratification of $\tilde{Z}(\gamma)_{\mathbb{R}}$,

$$\tilde{Z}(\gamma)_{\mathbb{R}} = \Theta^{(0)} \supset \Theta^{(1)} \supset \dots \supset \Theta^{(l)} = c_\gamma(C_\gamma), \quad \text{with } \Theta^{(k)} = \bigcup_{|J|=k} \Theta_J.$$

The irreducibility of the Painlevé divisors $\Theta_{\{k\}}$ was shown in [7], where the analog of Riemann’s singularity theorem for the compactified complex manifold $\tilde{Z}(\gamma)_{\mathbb{C}}$ was also discussed.

In the case of a given matrix (adjoint) representation, one can construct an explicit solution for $\{a_j(t)\}$ in the matrix $L(t)$. First we have the following Lemma:

LEMMA 2.1. *The diagonal element $b_{j,j}$ of the upper triangular matrix $b \in B^+$ in the factorization (2.1) is expressed by*

$$b_{j,j}(t) = \frac{D_j[\exp(tL^0)]}{D_{j-1}[\exp(tL^0)]}$$

where $D_j[\exp(tL^0)]$ is the determinant of the j -th principal minor of $\exp(tL^0)$, i.e.

$$D_j[\exp(tL^0)] = \left(e^{tL^0} v_1 \wedge \dots \wedge v_j, v_1 \wedge \dots \wedge v_j \right).$$

with the standard basis $\{v_i\}_{i=1}^l$ of \mathbb{R}^n with some n .

Then using the formula in (2.4), we can obtain the solution $a_j(t)$ and the explicit representation of the τ -functions in terms of the determinants $D_j[\exp(tL^0)]$. Thus the τ -functions are the entire functions of t given by polynomials of exponential

functions $\exp(\lambda_k t)$ with the eigenvalues λ_k of L^0 . In fact, one can show that the $D_j[\exp(tL^0)]$ can be expressed as the Hankel determinant,

$$D_j[\exp(tL^0)] = \begin{vmatrix} D_1 & D'_1 & \cdots & D_1^{(j-1)} \\ D'_1 & D''_1 & \cdots & D_1^{(j)} \\ \vdots & \ddots & \ddots & \vdots \\ D_1^{(j-1)} & \cdots & \cdots & D_1^{(2j-2)} \end{vmatrix}, \quad j = 1, 2, \dots, n,$$

where $D_1 = D_1[\exp(tL^0)] = \sum_{i=1}^n \rho_i \exp(\lambda_i t)$ for some $\rho_i \in \mathbb{R} \setminus \{0\}$. With this formula, one can study a detailed behavior of the τ functions [9].

REMARK 2.2. On any Cartan subgroup of G there is another set of functions having similar properties to the τ functions. These are the root characters χ_{ω_i} associated to fundamental weights ω_i . For example, the simple root characters $\chi_i := \chi_{\alpha_i}$ can be expressed in terms of the χ_{ω_i} with the inverse of the Cartan matrix of the Lie algebra \mathfrak{g} . This is the same relation that exists between the a_i in (2.2) and the τ functions. The signs of the characters χ_i change when chamber walls $\chi_i = -1$ are crossed in a Cartan subgroup in analogy to what happens to the signs of the a_i when a Painlevé divisor is crossed. If χ_i^* denotes the root character of the simple root α_i corresponding to each separate chamber in the Cartan subgroup, then χ_i^* is continuous through α_i walls and through some α_j walls. The points on a Cartan subgroup where $\chi_i^* + 1 = 0$ are called the α_i -negative wall [5], which defines an “algebraic” version of the Painlevé divisors $\Theta_{\{i\}}$ in terms of the functions χ_i . This set is compactified and gives rise to a topological space $\tilde{\Theta}_{\{i\}}^a$ (see Section 6).

3. Limit matrices, Painlevé divisors and the companion embedding

Here we show that Painlevé divisors can be parametrized using limit matrices. These were first introduced in [2] for the case of the periodic Toda lattice. The main result in [2] is to show the existence of the limit matrix, say L_J , which is constructed by conjugating the Lax matrix $L(t)$ with a matrix in N^- and taking the limit $t \rightarrow t_J$ corresponding to the factorization $e^{tJL^0} = \hat{n}(t_J)w_J\hat{b}(t_J)$ for $\hat{n} \in N^-$ and $\hat{b} \in B^+$. In our case of the nonperiodic Toda lattice limit matrices arise as a consequence of Theorem 3.3 of [8].

DEFINITION 3.1. For fixed $J \subset \{1, \dots, l\}$ we let P_J denote the parabolic subgroup of G containing B^+ and associated to J . One can define a projection

$$\pi_J : G/B^+ \rightarrow G/P_J.$$

The group N^- factors as $N^- = N_J^- N_J^+$ with $N_J^\pm := N^- \cap w_J N^\pm w_J^{-1}$. Hence any $n \in N^-$ can be written as $n = uy$ with $u \in N_J^-$ and $y \in N_J^+$ unique elements. We thus obtain factorizations (notation of Proposition 2.1): $n_0^{-1}n(t) = u(t)y(t)$, and $\pi_J(u(t)y(t)B^+) = u(t)P_J$.

Since the limit $n_0^{-1}n(t)B^+$ as $t \rightarrow t_J$ exists (see Proposition 2.1), it is of the form $\hat{u}(t_J)w_J B^+$ for some $\hat{u}(t_J) \in N_J^-$. Then we have

PROPOSITION 3.1. *With notation as in Definition 3.1, the limit of $u(t)$ as $t \rightarrow t_J$ exists,*

$$\lim_{t \rightarrow t_J} u(t) = \hat{u}(t_J) \in N_J^-.$$

Proof. Since $\lim_{t \rightarrow t_J} n_0^{-1}n(t)B^+ = \hat{u}(t_J)w_JB^+$ with $\hat{u}(t_J) \in N_J^-$, by applying π_J we obtain

$$\lim_{t \rightarrow t_J} n_0^{-1}n(t)P_J = \hat{u}(t_J)P_J.$$

On the other hand $\pi_J(n_0^{-1}n(t)B^+) = \pi_J(u(t)y(t)B^+) = u(t)P_J$. Therefore, since the top N^- orbit in G/P_J can be identified with N_J^- we then obtain a limit inside this group: $\lim_{t \rightarrow t_J} u(t) = \hat{u}(t_J)$. \square

DEFINITION 3.2. A limit matrix of L is an element L_J in the set \mathcal{F}_γ of the form,

$$L_J = Ad(\hat{u}^{-1}(t_J))C_\gamma, \quad \text{with } \hat{u}(t_J) \in N_J^-.$$

Let $u(t) = \hat{u}(t_J)\bar{u}(t)$. Then $\lim_{t \rightarrow t_J} \bar{u}(t) = e$, with e the identity, and we have

PROPOSITION 3.2. *The limit matrix is also expressed as*

$$\lim_{t \rightarrow t_J} Ad(y(t))L(t) = L_J(t_J).$$

Proof. We have

$$L(t) = Ad(n^{-1}(t)n_0)C_\gamma = Ad(y^{-1}(t)u^{-1}(t))C_\gamma = Ad(y^{-1}(t)\bar{u}^{-1}(t)\hat{u}^{-1}(t_J))C_\gamma$$

Hence $Ad(y(t))L(t) = Ad(\bar{u}(t))L_J$. We now take limit and use that $\bar{u}(t) \rightarrow e$ to conclude. \square

The result can be summarized in the diagram,

$$\begin{array}{ccc} L(t) & \xrightarrow{c_\gamma} & n_0^{-1}n(t) \bmod B^+ \\ \downarrow & & \downarrow \\ Ad(y(t))L(t) & \longrightarrow & u(t)y(t) \bmod B^+ \\ \downarrow_{t \rightarrow t_J} & & \downarrow_{t \rightarrow t_J} \\ L_J(t_J) & \xrightarrow{c_\gamma} & \hat{u}(t_J)w_J \bmod B^+ \end{array}$$

REMARK 3.3. For each set J we can define a function $\phi_J : Z(\gamma)_\mathbb{R} \rightarrow Ad(N_J^-)C_\gamma$ given by $\phi_J(L) = Ad(y)L$. A limit matrix L_J is then an element in the boundary $\overline{\phi_J(Z(\gamma)_\mathbb{R})} \setminus \phi_J(Z(\gamma)_\mathbb{R})$. The closure takes place inside $Ad(N_J^-)C_\gamma$. This gives another description of \mathcal{D}_J which allows one to define the compactification of the isospectral manifold $Z(\gamma)_\mathbb{R}$ using only the limit matrices. The companion embedding then takes a simple form. First note that any limit matrix L_J is contained in the N_J^- orbit of C_γ . Hence $L_J = Ad(\hat{u}^{-1}(t_J))C_\gamma$ where $\hat{u}(t_J) \in N_J^-$ is unique. For $J = \emptyset$, we just set $\hat{u}(t) = n_0^{-1}n(t)$. The companion embedding then maps L_J to $\hat{u}(t_J)w_JB^+$.

REMARK 3.4. In all our examples $y(t) = y_J(t)$ can be replaced with $x_J^{-1}(t)$ an element in N_J^- defined below in terms of a companion matrix associated to a Levi factor.

In the following, we determine the limit matrices for the case of $\mathfrak{g} = \mathfrak{sl}(l+1, \mathbb{R})$ (the general case will be discussed elsewhere). Let consider the set J be given by s consecutive numbers, say $\{i+1, \dots, i+s\}$, ($i+s \leq l$). Then from (2.8) this implies that the divisor \mathcal{D}_J consists of the points corresponding to the zeros of τ -functions, $\tau_k = 0$ for all $k \in J$. On the other hand, from (2.7), we can show

LEMMA 3.1. *For each $j \in J = \{i+1, \dots, i+s\}$, $\tau_j(t)$ has the following form near its zero $t = t_J$,*

$$(3.1) \quad \tau_{i+k}(t) \simeq (t - t_J)^{m_k} + \dots, \quad \text{with } m_k = k(s+1-k), \quad 1 \leq k \leq s.$$

Proof. Substituting (3.1) into (2.7), and using $\tau_i(t_J) \neq 0$, we have $m_k = k(m_1 + 1 - k)$. Then $\tau_{i+s+1}(t_J) \neq 0$ implies $m_1 = s$. \square

Then using (2.6) one can find the blow-up structure of the functions (a_j, b_j) . We note here that this structure is the same as the case of the smaller system $\mathfrak{sl}(s+1, \mathbb{R})$ with the total blow-up. The Lax matrix of this smaller system is just the submatrix (here the b -variables are modified from the original form in (2.2), e.g. $b_k - b_{k-1} \rightarrow b_k$),

$$L' = \begin{pmatrix} b_{i+1} & 1 & 0 & \dots & 0 \\ a_{i+1} & b_{i+2} & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \ddots & b_{i+s} & 1 \\ 0 & \dots & \dots & a_{i+s} & b_{i+s+1} \end{pmatrix}$$

Then one can put this matrix into a companion matrix by a unique element $x'_J \in N^-$, the set of $(s+1) \times (s+1)$ lower triangular matrices with 1's on the diagonals. The companion matrix $C'_J = x'^{-1}_J L' x'_J$ and x'_J are given by

$$C'_J = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \ddots & 0 & 1 \\ (-1)^s \xi_{s+1} & \dots & \dots & -\xi_2 & \xi_1 \end{pmatrix}, \quad x'_J = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ * & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & 0 \\ * & \dots & \dots & * & 1 \end{pmatrix},$$

where ξ_k 's are the polynomials of (a_j, b_j) in the Lax matrix. Since the Toda lattice is isospectral, those polynomials stays constants even when all of the elements (a_j, b_j) blows up. Then the limit matrix L_J is obtained by the limit of the conjugation of L with $x_J \in N^+_J$,

$$L_J = \lim_{t \rightarrow t_J} Ad(x_J^{-1}(t))L(t), \quad \text{with } x_J = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \dots & \ddots & \vdots \\ \vdots & \vdots & \vdots & x'_J & \vdots & \vdots \\ \vdots & \ddots & \ddots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & 1 \end{pmatrix}.$$

Let us now give an example to illustrate the construction:

EXAMPLE 3.5. The A_3 Toda lattices: The Lax matrix is given by

$$L = \begin{pmatrix} b_1 & 1 & 0 & 0 \\ a_1 & b_2 & 1 & 0 \\ 0 & a_2 & b_3 & 1 \\ 0 & 0 & a_3 & b_4 \end{pmatrix}, \quad \sum_{k=1}^4 b_k = 0.$$

The limit matrices L_J are determined as follows:

- a) $J = \{1\}$: Then $\tau_1(t) \sim t_* = (t - t_{\{1\}})$ implies that $a_1 \sim t_*^{-2}, a_2 \sim t_*, b_1 \sim t_*^{-1}, b_2 \sim t_*^{-1}$ and others are regular. The limit matrix is then obtained by the limit $x_{\{1\}}^{-1} L x_{\{1\}} \rightarrow L_{\{1\}}$ as $t_* \rightarrow 0$,

$$L_{\{1\}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\xi_2 & \xi_1 & 1 & 0 \\ \eta_1 & 0 & b_3 & 1 \\ 0 & 0 & a_3 & b_4 \end{pmatrix}, \quad \text{with } x_{\{1\}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -b_1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where $\xi_1 = b_1 + b_2, \xi_2 = b_1 b_2 - a_1, \eta_1 = -a_2 b_1$ are the parameters for the divisor $\mathcal{D}_{\{1\}}$.

- b) $J = \{2\}$: With $\tau_2 \sim t_* = (t - t_{\{2\}})$, we have $a_1 \sim t_*, a_2 \sim t_*^{-2}, a_3 \sim t_*, b_2 \sim t_*^{-1}, b_3 \sim t_*^{-1}$, and the limit matrix is given by

$$L_{\{2\}} = \begin{pmatrix} b_1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \eta_1 & -\xi_2 & \xi_1 & 1 \\ 0 & \eta_2 & 0 & b_4 \end{pmatrix}, \quad \text{with } x_{\{2\}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -b_2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where $\xi_1 = b_2 + b_3, \xi_2 = b_2 b_3 - a_2, \eta_1 = -a_1 b_3, \eta_2 = -a_3 b_2$ are the parameters for the divisor $\mathcal{D}_{\{2\}}$.

- c) $J = \{3\}$: This case is similar to the case $J = \{1\}$, and we have

$$L_{\{3\}} = \begin{pmatrix} b_1 & 1 & 0 & 0 \\ a_1 & b_2 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \eta_1 & -\xi_2 & \xi_1 \end{pmatrix}, \quad \text{with } x_{\{3\}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -b_3 & 1 \end{pmatrix}$$

where $\xi_1 = b_3 + b_4, \xi_2 = b_3 b_4 - a_3, \eta_1 = -a_2 b_4$ are the parameters for the divisor $\mathcal{D}_{\{3\}}$.

- d) $J = \{1, 2\}$: We construct $L_{\{1,2\}}$ from $L_{\{1\}}$ with $x_{\{1,2\}}$,

$$L_{\{1,2\}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \xi'_3 & -\xi'_2 & \xi'_1 & 1 \\ \eta'_1 & 0 & 0 & b_4 \end{pmatrix}, \quad \text{with } x_{\{1,2\}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \xi_2 & -\xi_1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where $\xi'_1 = \xi_1 + b_3, \xi'_2 = \xi_2 + \xi_1 b_3, \eta'_1 = \eta_1 + \xi_2 b_3$ with ξ_1, ξ_2, η_1 in $L_{\{1\}}$ are then the parameters for the divisor $\mathcal{D}_{\{1,2\}}$. This can be of course done with a matrix $x_{\{2,1\}}$ from $L_{\{2\}}$.

- e) $J = \{2, 3\}$: This is similar to the previous case d), and we have

$$L_{\{2,3\}} = \begin{pmatrix} b_1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \eta'_1 & \xi'_3 & -\xi'_2 & \xi'_1 \end{pmatrix}, \quad \text{with } x_{\{2,3\}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & \xi_2 & -\xi_1 & 1 \end{pmatrix}$$

where $\xi'_1 = \xi_1 + b_2$, $\xi'_2 = \xi_2 + \xi_1 b_2$, $\eta'_1 = \eta_1 + \xi_2 b_2$ with ξ_1, ξ_2, η_1 in $L_{\{2\}}$ are then the parameters for the divisor $\mathcal{D}_{\{2,3\}}$.

f) $J = \{1, 3\}$: We construct the limit matrix $L_{\{1,3\}}$ from $L_{\{1\}}$ by using $x_{\{1,3\}} = x_{\{3\}}$.

$$L_{\{1,3\}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\xi_2 & \xi_1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \eta'_1 & 0 & -\xi'_2 & \xi'_1 \end{pmatrix},$$

where $\xi'_1 = b_3 + b_4$, $\xi'_2 = b_3 b_4 - a_3$, $\eta'_1 = -\eta_1 b_3$.

4. The A_l Toda lattice and the Mumford system

In [14], Mumford gave a parametrization of the theta divisor for a hyperelliptic Jacobian with triples of polynomials determined by the factorization of the corresponding hyperelliptic curve. This is related to the periodic Toda lattice, but the idea can be also applied to the present case of nonperiodic Toda lattice on $\mathfrak{g} = \mathfrak{sl}(l + 1, \mathbb{R})$:

DEFINITION 4.1. The Mumford system for the spectral curve $F_l(\lambda) = \det(\lambda I - L)$ of degree $l + 1$ is the triples of polynomials $(u_d(\lambda), v_d(\lambda), w_d(\lambda))$ determined by

$$F_l(\lambda) = u_d(\lambda)w_d(\lambda) + v_d(\lambda),$$

where u_d is a monic polynomial of degree d , v_d is a polynomial of degree less than d with the condition $v_d(\mu_k) = F_l(\mu_k)$ for the roots of $u_d(\lambda) = 0$, and w_d is a monic polynomial of degree $l + 1 - d$.

One can write u_d and v_d in the form,

$$\begin{cases} u_d(\lambda) &= \prod_{k=1}^d (\lambda - \mu_k), \\ v_d(\lambda) &= \sum_{k=1}^d F_l(\mu_k) \prod_{j \neq k} \frac{\lambda - \mu_j}{\mu_k - \mu_j}. \end{cases}$$

When $d = l$ (the rank of the matrix), the μ -variables can globally parametrize the isospectral manifold $Z(\gamma)_{\mathbb{R}}$ by taking an explicit relation with the original variables (a_k, b_k) in L , for example, choose the l -th principal minor of L to be $u_l(\lambda)$. One can also define an integrable system for the Mumford system as

$$(4.1) \quad \begin{cases} \frac{du}{dt} &= v, \\ \frac{dv}{dt} &= u \left[\frac{vw}{u} \right]_+ - vw, \\ \frac{dw}{dt} &= - \left[\frac{vw}{u} \right]_+, \end{cases}$$

where $[f(\lambda)]_+$ indicates the polynomial part of $f(\lambda)$ (see [14, 15] for the periodic case). The integrability is a direct consequence of the isospectrality, i.e. fixing the curve $F_l(\lambda) = uw + v$. It is also interesting to note that the system has a Lax form,

$$\frac{dM}{dt} = [M, B], \quad \text{with} \quad M = \begin{pmatrix} h & u \\ w & -h \end{pmatrix}, \quad B = \frac{1}{2h} \begin{pmatrix} 0 & v \\ b & 0 \end{pmatrix},$$

where $h^2 = v$ and $b = [vw/u]_+$. Then the first equation in (4.1) gives the system,

$$\frac{d\mu_k}{dt} = -\frac{F_l(\mu_k)}{\prod_{j \neq k} (\mu_k - \mu_j)}, \quad k = 1, \dots, d.$$

Using the Lagrange interpolation formula,

$$\sum_{k=1}^d \frac{\mu_k^n}{\prod_{j \neq k} (\mu_k - \mu_j)} = \begin{cases} 0 & \text{if } n < d - 1, \\ 1 & \text{if } n = d - 1. \end{cases}$$

we obtain, after integration,

$$\sum_{k=1}^d \int_{\mu_0}^{\mu_k} \frac{\lambda^n d\lambda}{F_l(\lambda)} = \begin{cases} c_n & n < d - 1, \\ -t + c_{d-1} & n = d - 1. \end{cases}$$

with some constants μ_0 and $c_k, k = 1, \dots, d - 1$. In particular, the system with $d = 1$ gives

$$\frac{d\mu_1}{dt} = -F_l(\mu_1),$$

whose solution has $l + 1$ fixed points at $\mu_1 = \lambda_k$ for $k = 1, \dots, l + 1$, and blows up when $\mu_1 > \max_k(\lambda_k)$ or $\mu_1 < \min_k(\lambda_k)$. One can also show the following Proposition on the topology of certain 1-dimensional Painlevé divisors $\Theta_J(A_k)$ of the A_k Toda lattice:

PROPOSITION 4.1. *Let $J_{k-1} \subset \{1, \dots, k\}$ be either $\{1, \dots, k-1\}$ or $\{2, \dots, k\}$. Then the Painlevé divisors $\Theta_{J_{k-1}}(A_k)$ are all homeomorphic to circle, i.e.*

$$\Theta_{J_{k-1}}(A_k) \cong S^1, \quad \text{for } k = 1, 2, \dots,$$

where $J_0 = \emptyset$.

Proof. Since the homeomorphism between the divisors with $J = \{1, \dots, k - 1\}$ and $J = \{2, \dots, k\}$ is obvious, we consider the case with $J = \{1, \dots, k - 1\}$. In this case, the limit matrix has the form,

$$L_J = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 1 & 0 \\ (-1)^{k-1} \xi_k & \dots & \dots & \dots & \xi_1 & 1 \\ \eta & 0 & \dots & \dots & 0 & b_{k+1} \end{pmatrix},$$

where ξ_i are the coefficients of the polynomial $|\lambda I - L'| = \lambda^k + \sum_{i=1}^k (-1)^i \xi_i \lambda^{k-i}$ with L' given by the first $k \times k$ part of the Lax matrix L , and $\eta = -a_k b_1 \cdots b_{k-1}$ (in the limit $t \rightarrow t_J$). Then from the Mumford system $F_k(\lambda) := |\lambda I - L_J| = u_1 w_1 + v_1$, we have

$$\eta = -F_k(\mu_1), \quad \text{with } \mu_1 = b_{k+1},$$

where $v_1 = -\eta$. This indicates that the Painlevé divisor $\mathcal{D}_J(A_k)$ has just one connected component of \mathbb{R} , and adding the highest divisor $\Theta_{\{1, \dots, k\}}(A_k)$ we see that the closure $\Theta_J(A_k)$ is homeomorphic to S^1 . This completes the proof. \square

We can also show the following on higher dimensional divisors,

PROPOSITION 4.2. *Let $J_n^k \subset \{1, \dots, k+n\}$ be either $\{1, \dots, n\}$ or $\{k+1, \dots, k+n\}$. The Painlevé divisors $\mathcal{D}_{J_n^k}(A_{k+n})$ are all homeomorphic to the top cell of the A_k Toda lattice, i.e.*

$$\mathcal{D}_\emptyset(A_k) \cong \mathcal{D}_{J_n^k}(A_{k+n}), \quad \text{for } n \geq 1.$$

Proof. Let J be $\{1, \dots, n\}$. Then the limit matrix L_J is given by

$$L_J = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix},$$

where A_1 is the $(n + 1) \times (n + 1)$ companion matrix of the corresponding block in the matrix L , A_2 is the $(n + 1) \times k$ matrix having zero entries except 1 at the bottom left corner (i.e. $(k, 1)$ -entry), A_3 is the $k \times (n + 1)$ matrix having zero entries except η at the top left corner ($(1, 1)$ -entry), and A_4 is the $k \times k$ submatrix of Lax matrix,

$$A_4 = \begin{pmatrix} b_{n+2} & 1 & \cdots & \cdots & 0 \\ a_{n+2} & b_{n+3} & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & b_{k+n} & 1 \\ 0 & \cdots & \cdots & a_{k+n} & b_{k+n+1} \end{pmatrix}.$$

Then from the factorization of $F_{k+n}(\lambda) = |\lambda I - L_J| = u_k(\lambda)w_k(\lambda) + v_k(\lambda)$, we have the Mumford system,

$$u_k = |\lambda I - A_4|, \quad v_k = -\eta|\lambda I - B_4|, \quad w_k = |\lambda I - A_1|,$$

where B_4 is the $(k - 1) \times (k - 1)$ submatrix of A_4 by deleting the first row and column vectors. Thus we have

$$\eta|\mu_j I - B_4| = -F_{k+n}(\mu_j), \quad \text{for } j = 1, \dots, k,$$

The left-hand side of this equation has the same form for all the cases with fixed k , and the right-hand side gives a real one-dimensional affine curve for each $\mu_j \in \mathbb{R}$ of degree $k + n$. This implies that all the divisors $\mathcal{D}_{J_n^k}(A_{k+n})$ have the same parametrization, so that they are all homeomorphic. \square

Since the boundaries of each $\mathcal{D}_{J_n^k}(A_{k+n})$ seems to have the same structure for $n \geq 1$, we expect

CONJECTURE 4.2. *The Painlevé divisors $\Theta_{J_n^k}(A_{k+n})$ for $n \geq 1$ are all homeomorphic, i.e.*

$$\Theta_{J_1^k}(A_{k+1}) \cong \cdots \cong \Theta_{J_n^k}(A_{k+n}).$$

5. Examples for rank 2 and 3

5.1. **The A_2 -Toda lattice.** The Lax matrix is a 3×3 matrix given by

$$L = \begin{pmatrix} b_1 & 1 & 0 \\ a_1 & b_2 & 1 \\ 0 & a_2 & b_3 \end{pmatrix}, \quad \text{with } \sum_{k=1}^3 b_k = 0,$$

and the spectral curve $F_2(\lambda) = \det(\lambda I - L)$ is

$$F_2(\lambda) = \lambda^3 + I_1\lambda - I_2,$$

where the Chevalley invariants $I_k(L)$ are given by

$$I_1(L) = b_1b_2 + b_2b_3 + b_1b_3 - a_1 - a_2, \quad I_2(L) = b_1b_2b_3 - a_1b_3 - a_2b_1.$$

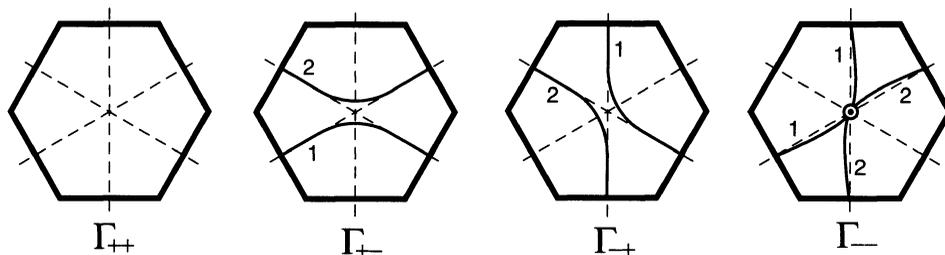


FIGURE 1. The A_2 hexagons $\Gamma_{\epsilon_1\epsilon_2}$. The numbers 1 and 2 mark the Painlevé divisors $\Theta_{\{1\}}$ and $\Theta_{\{2\}}$. The center point (double circle) in Γ_{--} indicates the divisor $\Theta_{\{1,2\}}$.

To parametrize the isospectral manifold $Z(\gamma)_{\mathbb{R}}$, we consider for example the following Mumford system with the choice of the triples,

$$\begin{cases} u_2(\lambda) &= \prod_{k=1}^2(\lambda - \mu_k) = \begin{vmatrix} \lambda - b_2 & -1 \\ -a_2 & \lambda - b_3 \end{vmatrix}, \\ v_2(\lambda) &= F_2(\mu_1)\frac{\lambda - \mu_2}{\mu_1 - \mu_2} + F_2(\mu_2)\frac{\lambda - \mu_1}{\mu_2 - \mu_1}, \\ w_2(\lambda) &= \frac{1}{u_2(\lambda)}(F_2(\lambda) - v_2(\lambda)) = \lambda + w_0. \end{cases}$$

Then in terms of (μ_1, μ_2) the Chevalley invariants are given by

$$I_1 = -(\mu_1 + \mu_2)^2 + \mu_1\mu_2 - a_1, \quad I_2 = -\mu_1\mu_2(\mu_1 + \mu_2) - a_1b_3,$$

which leads to

$$a_1(\mu_k - b_3) = -F_2(\mu_k), \quad k = 1, 2.$$

Also from (4.1), we have the Toda flow in the variable (μ_1, μ_2) ,

$$\frac{d\mu_k}{dt} = (-1)^k \frac{F_2(\mu_k)}{\mu_1 - \mu_2}, \quad k = 1, 2,$$

which is also obtained by setting $a_1(\mu_k - b_3) = (-1)^{k+1}(\mu_1 - \mu_2)d\mu_k/dt$. The system has 6 fixed points with $(\mu_1, \mu_2) = (\lambda_i, \lambda_j)$, $1 \leq i \neq j \leq 3$, and for each set of the signs (ϵ_1, ϵ_2) with $\epsilon_i = \text{sign}(a_i(0))$ the integral manifold gives a hexagon, denoted by $\Gamma_{\epsilon_1\epsilon_2}$ as in Fig.1. In particular, one can easily see that there is no blow-up in Γ_{++} (note that $I_1(L) = \gamma_1$ makes all the variables be bounded, if both a_1 and a_2 are positive). Those four hexagons are glued together along with their boundaries according to the standard action of the Weyl group S_3 on the signs (ϵ_1, ϵ_2) , and the compactified manifold is topologically equivalent to a connected sum of two Klein bottles \mathbb{K} [9]. This can be seen by counting the Euler characteristic, $6(\text{vertices}) - 12(\text{edges}) + 4(\text{hexagons}) = -2$ and the nonorientability (see [5] for the general argument on the compactification based on the Weyl group action).

The Painlevé divisor $\mathcal{D}_{\{1\}}$ corresponding to $\tau_1 = 0$ can be parametrized by the limit matrix,

$$(5.1) \quad L_{\{1\}} = \begin{pmatrix} 0 & 1 & 0 \\ -\xi_2 & \xi_1 & 1 \\ \eta_1 & 0 & b_3 \end{pmatrix},$$

where $\xi_1 = b_1 + b_2, \xi_2 = b_1 b_2 - a_1$ and $\eta_1 = -a_2 b_1$. The matrix $L_{\{1\}}$ is obtained by the limit,

$$L_{\{1\}} = \lim_{t \rightarrow t_1} x_{\{1\}}^{-1}(t)L(t)x_{\{1\}}(t), \quad \text{with } x_{\{1\}}(t) = \begin{pmatrix} 1 & 0 & 0 \\ -b & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then the spectral curve $F_2(\lambda)$ gives the algebraic relations (the Chevalley invariants),

$$\xi_1 + b_3 = 0, \quad I_1(L_{\{1\}}) = \xi_2 + \xi_1 b_3, \quad I_2(L_{\{1\}}) = \eta_1 + \xi_2 b_3,$$

which leads to

$$\mathcal{D}_{\{1\}} = \left\{ \begin{aligned} & \{(\xi_1, \xi_2, \eta_1, b_3) \in \mathbb{R}^4 : \xi_1 = -b_3, I_1 = \gamma_1, I_2 = \gamma_2\} \\ & \cup \{(\eta_1, b_3) \in \mathbb{R}^2 : \eta_1 = -F_2(b_3) = -b_3^3 - \gamma_1 b_3 + \gamma_2\} \end{aligned} \right\}.$$

We thus show that the closure of $\mathcal{D}_{\{1\}}$ is homeomorphic to a circle S^1 , and it intersects with three subsystems corresponding to $(a_2 = 0, b_3 = \lambda_k)$ for $k = 1, 2, 3$. The Mumford equation (4.1) can be used to provide a dynamics on $\mathcal{D}_{\{1\}}$ with $\mu_1 = b_3$ and $\eta_1 = d\mu_1/dt$,

$$\frac{d\mu_1}{dt} = -F_2(\mu_1).$$

In Figure 1, $\Theta_{\{1\}}$ is shown as a curve with the label “1”. The $\Theta_{\{2\}}$ has the similar structure. Thus we obtain:

PROPOSITION 5.1. *The compactified manifold $\tilde{Z}(\gamma)_{\mathbb{R}}$ and the Painlevé divisor have the following topology,*

$$\tilde{Z}(\gamma)_{\mathbb{R}} = \Theta_{\emptyset} \cong \mathbb{K} \# \mathbb{K}, \quad \Theta_{\{1\}} \cong \Theta_{\{2\}} \cong S^1.$$

We also note by taking out the divisors $\Theta_{\{1\}}$ and $\Theta_{\{2\}}$ from $\tilde{Z}(\gamma)_{\mathbb{R}}$ that the top cell $\mathcal{D}_{\emptyset} = \tilde{Z}(\gamma)_{\mathbb{R}} \cap N^- B^+ / B^+$ is diffeomorphic to a torus \mathbb{T} with a hole of a disk \mathbb{D} , i.e.

$$\mathcal{D}_{\emptyset} \cong \mathbb{T} \setminus \mathbb{D}.$$

5.2. The C_2 Toda lattice. Since the B_2 Toda lattice has the same structure as the C_2 case, we discuss only the latter one. The Lax matrix for C_2 Toda lattice is given by a 4×4 matrix,

$$L = \begin{pmatrix} b_1 & 1 & 0 & 0 \\ a_1 & b_2 & 1 & 0 \\ 0 & 2a_2 & -b_2 & 1 \\ 0 & 0 & a_1 & -b_1 \end{pmatrix}$$

whose spectral curve $F_2(\lambda) = \det(\lambda I - L)$ is

$$F_2(\lambda) = \lambda^4 - I_1 \lambda^2 + I_2$$

with the Chevalley invariants $I_k(L)$,

$$I_1 = b_1^2 + b_2^2 + 2a_1 + 2a_2, \quad I_2 = (b_1 b_2 - a_1)^2 + 2b_1^2 a_2.$$

The corresponding polytope $\Gamma_{\epsilon_1 \epsilon_2}$ with the signs $\epsilon_k = \text{sign}(a_k)$ is given by a octagon with eight vertices associated with the fixed point of the system, $a_1 = a_2 = 0$. Those vertices are expressed as $(b_1, b_2) = (\sigma_1 \lambda_i, \sigma_2 \lambda_j)$ for $\sigma_k \in \{\pm\}, i \neq j \in \{1, 2\}$. Gluing those octagons along their boundaries, we find that the compactified manifold $\tilde{Z}(\gamma)_{\mathbb{R}}$ is topologically equivalent to a connected sum of three Klein bottles \mathbb{K} . Again just

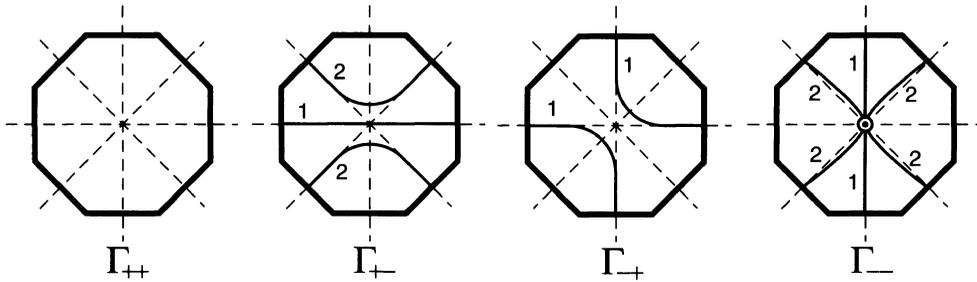


FIGURE 2. The C_2 octagons $\Gamma_{\epsilon_1\epsilon_2}$ and the Painlevé divisors $\Theta_{\{1\}}, \Theta_{\{2\}}$ and $\Theta_{\{1,2\}}$.

count the Euler characteristic, $8(\text{vertices}) - 16(\text{edges}) + 4(\text{octagons}) = -4$, and the nonorientability leads to the result.

The Painlevé divisor $\Theta_{\{1\}}$ is now parametrized by the limit matrix

$$L_{\{1\}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\xi_2 & \xi_1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \eta_1 & 0 & -\xi_2 & -\xi_1 \end{pmatrix}$$

where $\xi_1 = b_1 + b_2, \xi_2 = b_1b_2 - a_1$ and $\eta_1 = -2a_2b_1^2$. Then the Chevalley invariants $I_k(L)$ are expressed by

$$I_1 = \xi_1^2 - 2\xi_2, \quad I_2 = \xi_2^2 - \eta_1$$

from which we obtain

$$\eta_1 = \frac{1}{4}((\xi_1^2 - I_1)^2 - 4I_2).$$

This implies that the $\Theta_{\{1\}}$ is homeomorphic to S^1 and intersects with four subsystems corresponding to $\xi_1 = \sigma(\lambda_1 \pm \lambda_2)$ with $\sigma \in \{\pm 1\}$ and with the divisor $\Theta_{\{2\}}$ in Γ_{--} (see Figure 2),

Unlike the case of A_2 Toda lattice, the divisor $\Theta_{\{2\}}$ has a different structure. The corresponding limit matrix $L_{\{2\}}$ is given by

$$L_{\{2\}} = \begin{pmatrix} b_1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \eta_2 & \xi_3 & 0 & 1 \\ 0 & -\eta_2 & 0 & -b_1 \end{pmatrix}$$

where $\xi_3 = b_2^2 + 2a_2, \eta_2 = a_1b_2$. The invariants I_k are then given by

$$I_1 = b_1^2 + \xi_3, \quad I_2 = \xi_3b_1^2 - 2\eta_2b_1,$$

and we obtain

$$\eta_2 = -\frac{1}{b_1}F_2(b_1).$$

Because of the singularity in this equation at $b_1 = 0$, the $\Theta_{\{2\}}$ is shown to be homeomorphic to a figure eight, where each circle intersects two subsystems corresponding to either $b_1 = |\lambda_k|$ or $b_1 = -|\lambda_k|$ with $k = 1, 2$. The node of the figure eight corresponds to the divisor $\Theta_{\{1,2\}}$ (see Figure 2). We thus obtain,

PROPOSITION 5.2. *The topology of the isospectral manifold of C_2 and the Painlevé divisor is given by*

$$\tilde{Z}(\gamma)_{\mathbb{R}} \cong \mathbb{K} \# \mathbb{K} \# \mathbb{K}, \quad \Theta_{\{1\}} \cong S^1, \quad \Theta_{\{2\}} \cong S^1 \vee S^1.$$

The B_2 Toda lattice has the same structure, but $\Theta_{\{1\}}$ and $\Theta_{\{2\}}$ have the opposite structure.

5.3. The G_2 Toda lattice. We use the following one for the Lax matrix,

$$L = \begin{pmatrix} b_1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ a_1 & b_2 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & a_2 & b_1 - b_2 & 1 & 0 & \cdot & 0 \\ 0 & 0 & 2a_1 & 0 & 1 & 0 & 0 \\ 0 & \cdot & 0 & 2a_1 & -b_1 + b_2 & 1 & 0 \\ 0 & \cdot & \cdot & 0 & a_2 & -b_2 & 1 \\ 0 & \cdot & \cdot & \cdot & 0 & a_1 & -b_1 \end{pmatrix}.$$

The spectral curve is then given by

$$F_2(\lambda) = \lambda(\lambda^2(\lambda^2 + I_1)^2 + I_2),$$

where I_1 and I_2 are the Chevalley invariants given by homogeneous polynomials of (a_1, \dots, b_2) . Each polygon $\Gamma_{\epsilon_1, \epsilon_2}$ in the isospectral manifold has 12 vertices corresponding to $a_1 = a_2 = 0$ which is also the order of the Weyl group. Those polygons are glued to obtain the compactified manifold which is topologically equivalent to a sum of five Klein bottles. The Euler characteristic is $12(\text{vertices}) - 24(\text{edges}) + 4(\text{polygons}) = -8$.

The Painlevé divisor $\mathcal{D}_{\{1\}}$ is parametrized by the limit matrix $L_{\{1\}}$,

$$L_{\{1\}} = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ -\xi_2 & \xi_1 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdot & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \eta & 0 & 0 & \xi_1^2 - 4\xi_2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -\eta & 0 & 0 & -\xi_2 & -\xi_1 \end{pmatrix},$$

where $\xi_1 = b_1 + b_2$, $\xi_2 = b_1 b_2 - a_1$ and $\eta = 2b_1 a_1 a_2$ in the limit $t \rightarrow t_{\{1\}}$ with $\tau_1(t) \sim t - t_{\{1\}}$. Here we have used the conjugating matrix $x_{\{1\}}$ as,

$$x_{\{1\}} = \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ -b_1 & 1 & 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdot & 0 \\ 0 & 0 & b_2 - b_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -2a_1 & b_2 - b_1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -b_1 & 1 \end{pmatrix},$$

which can be obtained from the structure of τ -functions in (2.6). Then the invariants I_1, I_2 are given by

$$I_1 = 3\xi_2 - \xi_1^2, \quad I_2 = (4\xi_2 - \xi_1^2)\xi_2^2 + 2\eta\xi_1.$$

Eliminating ξ_2 , we obtain

$$\eta = \frac{1}{2\xi_1} \left(-\frac{1}{27}(\xi_1^2 + I_1)^2(\xi_1^2 + 4I_1) + I_2 \right).$$

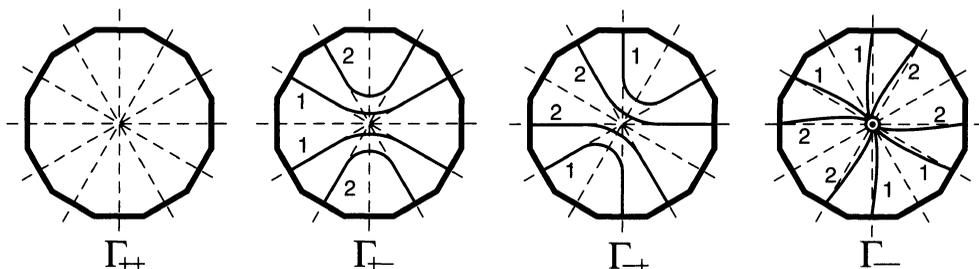


FIGURE 3. The G_2 polygons $\Gamma_{\epsilon_1\epsilon_2}$ with the Painlevé divisors $\Theta_{\{1\}}$, $\Theta_{\{2\}}$ and $\Theta_{\{1,2\}}$.

which has two connected components, and each component intersects three times with the boundaries of the polytopes Γ_{+-} , Γ_{-+} and Γ_{--} (see Figure 3).

The limit matrix corresponding to the Painlevé divisor $\mathcal{D}_{\{2\}}$ is given by

$$L_{\{2\}} = \begin{pmatrix} \xi_1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ \eta & -\xi_2 & \xi_1 & 1 & 0 & \cdot & 0 \\ 0 & -2\eta & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2\eta & -\xi_2 & -\xi_1 & 1 \\ 0 & 0 & 0 & 0 & -\eta & 0 & -\xi_1 \end{pmatrix}$$

with the conjugating matrix $x_{\{2\}}$,

$$x_{\{2\}} = \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & 0 & \cdot & \cdot & 0 \\ 0 & -b_2 & 1 & 0 & 0 & \cdot & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & b_1 - b_2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Here the new variables are $\xi_1 = b_1$, $\xi_2 = b_2(b_1 - b_2) - a_2$ and $\eta = a_1b_2$ in the corresponding limit with $\tau_2(t) \rightarrow 0$. With those variables, the invariants are

$$I_1 = \xi_2 - \xi_1^2, \quad I_2 = 3\eta^2 - 2\xi_1\eta(\xi_2 + 2\xi_1^2) - \xi_1^2\xi_2^2,$$

from which we have two curves $\eta = \eta_{\pm}(\xi_1)$,

$$\eta_{\pm} = \frac{1}{3} \left(\xi_1(3\xi_1^2 + I_1) \pm \sqrt{2\xi_1^2(5\xi_1^4 + 4I_1\xi_1^2 + I_1^2) + I_2} \right).$$

Those curves indicate that there are two connected components of the divisor $\mathcal{D}_{\{2\}}$ and each component has three intersections with the subsystems. Topologically then the divisors $\mathcal{D}_{\{1\}}$ and $\mathcal{D}_{\{2\}}$ are the same, and adding the divisor $\mathcal{D}_{\{1,2\}}$ one can conclude that the closure of both divisors are topologically equivalent to a figure eight. Thus we have

PROPOSITION 5.3. *The topology of the G_2 Toda isospectral manifold and the divisor is given by*

$$\tilde{Z}(\gamma)_{\mathbb{R}} \cong \overbrace{\mathbb{K} \# \cdots \# \mathbb{K}}^5, \quad \Theta_{\{1\}} \cong \Theta_{\{2\}} \cong S^1 \vee S^1.$$

5.4. The A_3 Toda lattice. In the example 3.5, we gave the limit matrices for the Painlevé divisors. Here we discuss the topology of the divisors by computing explicitly the isospectral sets of those matrices, i.e.

$$F_3(\lambda) = \lambda^4 + I_1\lambda^2 - I_2\lambda + I_3,$$

where the Chevalley invariants $I_k(L_J), k = 1, 2, 3$ are now expressed in terms of the parameters in the limit matrices. Here we use the same parametrizations in Example 3.5:

- a) $J = \{1\}$: We take the polynomial $u_2(\lambda)$ in the Mumford system as $u_2(\lambda) = \begin{vmatrix} \lambda - b_3 & -1 \\ -a_3 & \lambda - b_4 \end{vmatrix}$, i.e. $\mu_1 + \mu_2 = b_3 + b_4 = -\xi_1, \mu_1\mu_2 = b_3b_4 - a_3$. Then the Chevalley invariants are given by

$$\begin{cases} I_1 &= \xi_2 - (\mu_1 + \mu_2)^2 + \mu_1\mu_2, \\ I_2 &= \eta_1 - \mu_1\mu_2(\mu_1 + \mu_2) + \xi_2(\mu_1 + \mu_2), \\ I_3 &= \xi_2\mu_1\mu_2 + \eta_1b_4. \end{cases}$$

Eliminating ξ_2 , we find

$$\eta_1(\mu_k - b_4) = -F_3(\mu_k), \quad k = 1, 2.$$

As was shown in Proposition 4.2, comparing this with the top cell of the A_2 Toda lattice in Subsection 5.1, one can see

$$\mathcal{D}_{\{1\}} \cong \mathbb{T} \setminus \mathbb{D}.$$

- b) $J = \{2\}$: We take $u_2(\lambda) = (\lambda - b_1)(\lambda - b_4)$, i.e. $\mu_1 = b_1, \mu_2 = b_4$. Then we have, using $\xi_1 = -(\mu_1 + \mu_2)$,

$$\begin{cases} I_1 &= \mu_1\mu_2 - (\mu_1 + \mu_2)^2 + \xi_2, \\ I_2 &= \xi_2(\mu_1 + \mu_2) - \mu_1\mu_2(\mu_1 + \mu_2) + \eta_1 + \eta_2, \\ I_3 &= \mu_1\mu_2\xi_2 + \mu_1\eta_2 + \mu_2\eta_1, \end{cases}$$

which lead to

$$\eta_k = (-1)^k \frac{F_3(\mu_k)}{\mu_1 - \mu_2}.$$

- c) $J = \{3\}$: This case is similar to the one with $J = \{1\}$, and we have the same formulae of the Chevalley invariants in the variables μ_1, μ_2 which are defined as $u_2(\lambda) = \begin{vmatrix} \lambda - b_1 & -1 \\ -a_3 & \lambda - b_2 \end{vmatrix}$, i.e. $\mu_1 + \mu_2 = b_1 + b_2 = -\xi_1, \mu_1\mu_2 = b_1b_2 - a_1$.

- d) $J = \{1, 2\}$: Here we take $u_1(\lambda) = \lambda - b_4$ for the Mumford system, i.e. $\mu_1 = b_4$ and $v_1 = -\eta'_1$. Then the Chevalley invariants are

$$I_1 = \xi'_2 - \mu_1^2, \quad I_2 = \xi'_3 + \xi'_2\mu_1, \quad I_3 = \xi'_3\mu_1 - \eta'_1,$$

and from $v_1 = -\eta'_1$, we obtain

$$\eta'_1 = -F_3(\mu_1),$$

which implies that $\mathcal{D}_{\{1,2\}}$ intersects with four boundaries of the polytopes, and the closure, $\Theta_{\{1,2\}}$ is homeomorphic to a circle.

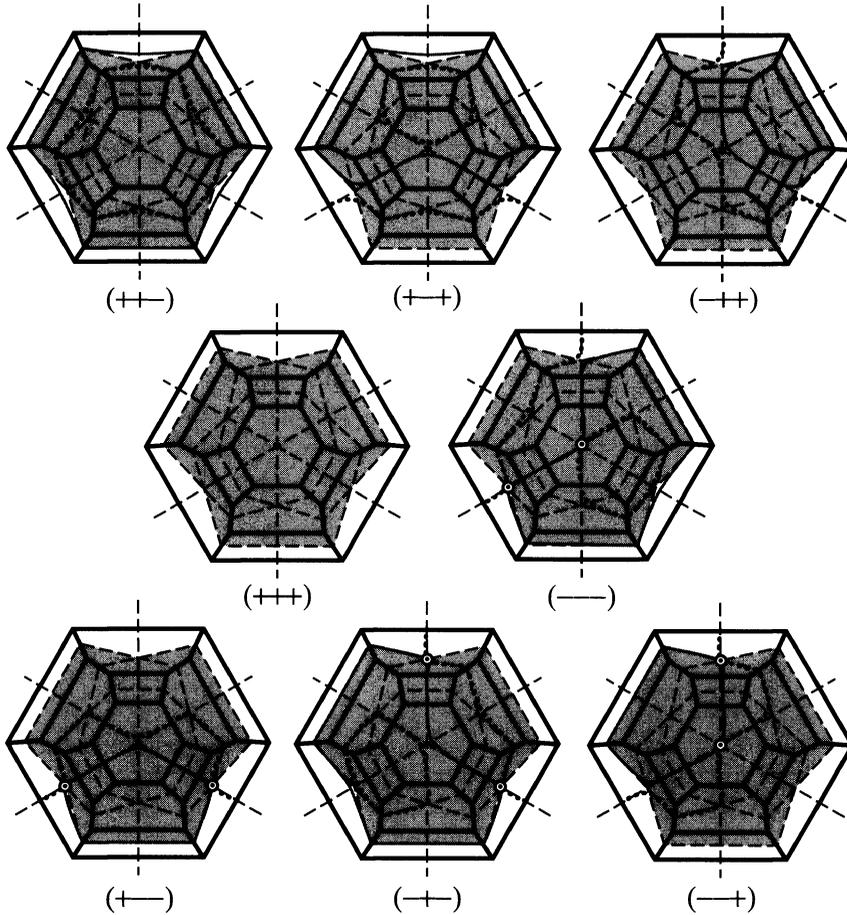


FIGURE 4. The A_3 polytopes Γ_ϵ marked by $\epsilon = (\epsilon_1\epsilon_2\epsilon_3)$ and the Painlevé divisors $\Theta_{\{1\}}$ (the solid grey curves), $\Theta_{\{2\}}$ (the dotted curves) and $\Theta_{\{1,2\}}$ (the double circles).

- e) $J = \{2, 3\}$: We get exactly the same result as the previous case with $\mu_1 = b_1$.
- f) $J = \{1, 3\}$: The Chevalley invariants are given by

$$I_1 = \xi_1\xi'_1 + \xi_2 + \xi'_2, \quad I_2 = \xi_1\xi'_2 + \xi_2\xi'_1, \quad I_3 = \xi_2\xi'_2 - \eta'_1.$$

Using $\xi_1 + \xi'_1 = 0$ and eliminating ξ'_1, ξ'_2 , we obtain

$$\eta'_1 = \frac{1}{4\xi_1^2} (\xi_1^2(\xi_1^2 + I_1)^2 - I_2^2) - I_3.$$

This equation indicates that $\mathcal{D}_{\{1,3\}}$ has two connected components, each of which intersects with three boundaries of the polytopes. Each boundary corresponds to a point $(\eta'_1 = 0, \lambda_i + \lambda_j)$ for $i \neq j$. Then we can see

$$\Theta_{\{1,3\}} \cong S^1 \vee S^1.$$

The results are summarized in Figure 4 where the Painlevé divisors $\Theta_{\{1\}}, \Theta_{\{2\}}$ and $\Theta_{\{1,2\}}$ are shown as the solid grey curves, the dotted curves and the double circles. The $\Theta_{\{3\}}$ has a similar structure to the $\Theta_{\{1\}}$. One can see from this Figure that

each portion of the $\Theta_{\{1\}}$ on a Γ_ϵ is homeomorphic to either hexagon or octagon, and we have 4 hexagons in Γ_ϵ with $\epsilon = (+ + -), (- + +), (+ - -), (- - +)$, and 3 octagons with $\epsilon = (+ - +), (- + -), (- - -)$. Then the Euler characteristic can be computed as follows: The total number of vertices are given by $12 = (4 \times 6 + 3 \times 8)/4$ by identifying 4 vertices of the polygons, the edges are $24 = (4 \times 6 + 3 \times 8)/2$ in total, and we have 7 faces, i.e. the Euler characteristic is $12 - 24 + 7 = -5$. One can also see the non-orientability of the divisor, so that the $\Theta_{\{1\}}$ is topologically equivalent to a connected sum of 7 real projective planes \mathbb{P} (or 3 Klein bottles plus a projective plane). For $\Theta_{\{2\}}$, we have 4 squares and 4 hexagons. However two squares in Γ_{---} are attached at a point of the divisor $\Theta_{\{1,2,3\}}$, and thus the $\Theta_{\{2\}}$ gives a singular variety. By detaching those two squares, one can compute the Euler characteristic in the same way as above, and we obtain $12 - 24 - 10 = -2$. This shows that the desingularized variety of $\Theta_{\{2\}}$ is homeomorphic to the compactified manifold $\tilde{Z}(\gamma)_{\mathbb{R}}$ for the A_2 Toda lattice (in the next section we give a further discussion on the desingularization in Lie theoretic point of view). Thus we have

PROPOSITION 5.4. *The Painlevé divisors for the A_3 Toda lattice have the following topology,*

$$\begin{aligned} \Theta_{\{1\}} \cong \Theta_{\{3\}} \cong \mathbb{K} \# \mathbb{K} \# \mathbb{K} \# \mathbb{P}, & \quad \Theta_{\{2\}}^\vee \cong \mathbb{K} \# \mathbb{K}, \\ \Theta_{\{1,2\}} \cong \Theta_{\{2,3\}} \cong S^1, & \quad \Theta_{\{1,3\}} \cong S^1 \vee S^1, \end{aligned}$$

where $\Theta_{\{2\}}^\vee$ is the desingularization of $\Theta_{\{2\}}$ by a resolution at the divisor $\Theta_{\{1,2,3\}}$.

The singular structure on the divisor $\Theta_{\{1,3\}}$ has been also found in the case of periodic Toda lattice [10].

6. An algebraic version of the Painlevé divisor

Here we discuss the Painlevé divisor in the framework of the Lie theory. We first review and summarize some Lie theoretic notation.

6.1. Notations and Definitions.

NOTATION 6.1. Lie algebras: Recall that \mathfrak{g} denote a real split semisimple Lie algebra of rank l and we are fixing a split Cartan subalgebra \mathfrak{h} with root system Δ , a positive system Δ_+ determining the Borel subgroup B^+ of G . The corresponding set of simple roots is $\Pi := \{\alpha_i : i = 1, \dots, l\}$ as in Section 2 where we just denoted $\Pi = \{k = 1, \dots, l\}$.

The Weyl group W is thus generated by the simple reflections $s_{\alpha_i}, i = 1, \dots, l$. For any $S \subset \Pi$, we define the subgroup generated by S ,

$$W_S = \langle s_{\alpha_i} : \alpha_i \in S \rangle$$

This is the Weyl group of a parabolic Lie subgroup and it is standard to refer to W_S as a *parabolic subgroup* of W .

NOTATION 6.2. Lie groups: We let $G_{\mathbb{C}}$ denote the connected adjoint Lie group with Lie algebra $\mathfrak{g}_{\mathbb{C}}$ and G the connected Lie subgroup corresponding to \mathfrak{g} . Denote by \tilde{G} the Lie group $\{g \in G_{\mathbb{C}} : Ad(g)\mathfrak{g} \subset \mathfrak{g}\}$. A split Cartan of \tilde{G} with Lie algebra \mathfrak{h} will be denoted by $H_{\mathbb{R}}$; this Cartan subgroup has exactly 2^l connected components and the component of the identity is denoted by $H = \exp(\mathfrak{h})$. We let $\chi_i := \chi_{\alpha_i}$ denote the roots characters defined on $H_{\mathbb{R}}$.

EXAMPLE 6.3. If $G = Ad(SL(n, \mathbb{R}))$, then \tilde{G} is isomorphic to $SL(n, \mathbb{R})$ for n odd and to $Ad(SL(n, \mathbb{R})^\pm)$ for n even. This example is the underlying Lie group for the Toda lattices as shown in [5].

DEFINITION 6.4. The negative walls: Recall that the compactified isospectral manifold $\tilde{Z}(\gamma)_\mathbb{R}$ of the Toda lattice is described in [5] as a closure in G/B^+ of a generic $H_\mathbb{R}$ orbit. Hence there is an embedding $f : H_\mathbb{R} \rightarrow \tilde{Z}(\gamma)_\mathbb{R} \subset G/B^+$.

The exponential map $\exp : \mathfrak{h} \rightarrow H$ separates H , and consequently every connected component of $H_\mathbb{R}$, into chambers. If χ_i is a simple root characters relative to a fixed dominant chamber then χ_{α_i} can be extends to an adjacent chamber by $s_{\alpha_j} \chi_{\alpha_i} = \chi_{\alpha_i} \chi_{\alpha_j}^{-C_{i,j}}$. This defines a single function χ_i^* on an open dense subset of $H_\mathbb{R}$ which equals $\chi_{w(\alpha_i)}$ on each w -chamber for $w \in W$ (denoted by $\phi_{w,i}$ in Definition 5.4 of [5]). The functions $|\chi_i^*|$ are well defined and continuous throughout $H_\mathbb{R}$ and the χ_i^* are well defined and continuous at all the α_i walls and some of the α_j walls. For example, if σ is a permutation, then the corresponding chamber in $SL(3, \mathbb{R})$ looks like $\{(r_{\sigma(1)}, r_{\sigma(2)}, r_{\sigma(3)}) : |r_1| > |r_2| > |r_3|\}$ and $\chi_1^* = r_{\sigma(1)} r_{\sigma(2)}^{-1}$, $\chi_2^* = r_{\sigma(2)} r_{\sigma(3)}^{-1}$.

The functions $\chi_i^* + 1 = 0$ on $H_\mathbb{R}$ then determine a topological subspace of $\tilde{Z}(\gamma)_\mathbb{R}$ whose closure we denote $\tilde{\Theta}_{\{i\}}$. Similarly a subset $J \subset \Pi$ determines a topological space $\tilde{\Theta}_J$ by equations $\chi_i^* + 1 = 0$ for $\alpha_i \in J$. We call $\tilde{\Theta}_J$ the *negative wall* associated with the set J (see Subsection 6.3 for another definition in the language of [6] which does not explicitly involve the Cartan subgroup).

CONJECTURE 6.5. *There is a surjective continuous map $f : \Theta_J \rightarrow \tilde{\Theta}_J^a$. This map is a homeomorphism in an open dense subset of Θ_J . Whenever $\tilde{\Theta}_J^a$ happens to be homeomorphic to a non-singular manifold then f is a homeomorphism.*

EXAMPLE 6.6. In the case of $\mathfrak{sl}(3)$ all $\Theta_{\{i\}}$ and $\tilde{\Theta}_{\{i\}}^a$ are homeomorphic. They are both homeomorphic to a circle (see Example 6.15 below). For $\mathfrak{sl}(4)$ again $\Theta_{\{i\}}$ is homeomorphic to $\tilde{\Theta}_{\{i\}}^a$ for $i = 1, 3$ (details in Example 6.12 and Proposition 5.4). However $\Theta_{\{2\}}$ and $\tilde{\Theta}_{\{2\}}^a$ are not homeomorphic. The situation is described in Example 6.19 together with Proposition 5.4 and is as follows. It is possible to desingularize $\Theta_{\{2\}}$ so that the compact connected surface $\hat{\Theta}_{\{2\}}$ which is obtained is non-orientable with Euler characteristic -2 . Then there are maps $\hat{\Theta}_{\{2\}} \rightarrow \Theta_{\{2\}} \rightarrow \tilde{\Theta}_{\{2\}}$ and $\hat{\Theta}_{\{2\}}$ now resolves the singularities of both $\Theta_{\{2\}}$ and $\tilde{\Theta}_{\{2\}}$. Conjecture 6.5 needs to be sharpened by modifying $\tilde{\Theta}_J^a$ slightly so that one always has homeomorphisms. Below we propose such a modification for the case when J consists of one simple root.

It is now easy to see that $\Theta_{\{1\}}$ and $\tilde{\Theta}_{\{1\}}^a$ agree in the case of A_3 . Figure 4 shows the eight polytopes Γ_ϵ corresponding to 2^l connected components of $H_\mathbb{R}$. In fact what is shown is the boundary of each polytope and the intersection of $\Theta_{\{i\}}$ for $i = 1, 2$. However, the negative walls are also depicted by the same picture. The only modification consists in drawing the dotted lines or the solid grey lines through the center of the hexagons. The actual negative walls are obtained by joining the dotted line or solid grey line to the center of the polytope through straight lines generating cones. Hence $\tilde{\Theta}_{\{1\}}^a$ intersected with each polytope consists of a disk in the form of a cone joining the center of the polytope with the path described on

the boundary of the polytope by the solid grey line. Gluings are described in detail in Definition 6.9. What results is a smooth compact surface.

In order to introduce modifications to $\tilde{\Theta}_J^q$ we need to describe its structure in more detail. We do this by using the description of a manifold M given in [6] which is homeomorphic to $\tilde{Z}(\gamma)_{\mathbb{R}}$. We review the construction of M and then define new topological spaces Θ_J^q in the case when J consists of one simple root.

DEFINITION 6.7. Let \mathcal{E} be the set of signs $\mathcal{E} = \{(\epsilon_1, \dots, \epsilon_l) : \epsilon_k \in \{\pm\}\}$. Then we define an action of W on \mathcal{E} by setting $s_i \epsilon = \epsilon'$ where

$$\epsilon'_j = \epsilon_j \epsilon_i^{C_j, i},$$

which can be deduced from the W -action on the root character χ_i with $\epsilon_i = \text{sign}(\chi_i)$. The fact that this defines an action which corresponds to the action of W on the set of connected components of a split Cartan subgroup of the real semisimple Lie group \tilde{G} can be found in [5].

For any $S \subset \Pi$ we let $\mathbb{D}(S)$ denote the set of all Dynkin diagrams that have the simple roots in S marked by $+$ or $-$. We also define an action of the group W_S on this set by making $w \in W_S$ act on the signs associated to the simple roots in S as prescribed above. For example $\circ_- - \circ_+ - \circ \in \mathbb{D}(S)$ with $S = \{\alpha_1, \alpha_2\}$ and $s_1(\circ_- - \circ_- - \circ) = \circ_- - \circ_+ - \circ$.

We now obtain actions of W_S on $\mathcal{E} \times W$ and on $\mathbb{D}(S) \times W$ given by $\sigma(\epsilon, w) = (\sigma\epsilon, w\sigma^{-1})$ and $\sigma(\delta, w) = (\sigma\delta, w\sigma^{-1})$. The orbits of the W_S action on $(\epsilon, w) \in \mathcal{E} \times W$ and $(\delta, w) \in \mathbb{D}(S) \times W$ are denoted by $[\epsilon, w]_S$ and $[\delta, w]_S$ respectively with the sub-index S dropped when the set S is clear from the context. These W_S orbits in the case of $\mathbb{D}(S) \times W$ are the *full set of colored Dynkin diagrams* introduced in section 4 of [5]. The orbits of W_S on $\mathcal{E} \times W$ with $S = \{\alpha_i\}$ are used below to parametrize the walls $\chi_i^* \pm 1 = 0$ intersected with a fixed polytope. The walls $\chi_i^* \pm 1 = 0$ in $\tilde{Z}(\gamma)_{\mathbb{R}}$ can be parametrized by $\mathbb{D}(S) \times W$ with $S = \{\alpha_i\}$.

6.2. Review of the description of $\tilde{Z}(\gamma)_{\mathbb{R}}$ in terms of the polytopes Γ_ϵ . Here we discuss the detailed description of negative walls in the connection to the Painlevé divisors Θ_J . Let us first summarize the construction of the isospectral manifold of the Toda lattice given in [6]. Starting with a polytope Γ , other polytopes Γ_ϵ are constructed where $\epsilon \in \mathcal{E}$. These polytopes then form a compact smooth manifold when they are glued together through their boundaries. We now review the details.

In terms of the description given in [5], each Γ_ϵ has interior that can be made to correspond to a connected component of a split Cartan subgroup of the real semisimple split Lie group \tilde{G} . *Chambers* and *walls* then refer to the action of W on a Cartan subgroup, and the internal chamber walls in the polytopes Γ_ϵ correspond to walls of this action ($\chi_i^* \pm 1 = 0$). When $\chi_i^* = -1$ then the chamber at the other side of the wall need not be the one obtained by application of $s_{\alpha_i} = s_i$.

DEFINITION 6.8. Consider Γ a convex polytope consisting of the convex hull of a W orbit of a regular element x_o in \mathfrak{h} . We first denote C'_e the dominant chamber in \mathfrak{h} intersected with Γ and \overline{C}'_e the corresponding closure, and also denote $C'_w = w(C'_e)$. We define $C_w = \{w\} \times C'_w$ and its closure $\overline{C}_w = \{w\} \times \overline{C}'_w$. The \dots will refer to subsets of Γ , and we have the convention:

$$\{\dots\}^{\dots} = \{w\} \times \{\dots\}^{\dots}$$

in all our notation concerning walls. For each simple root α_i we may consider the corresponding α_i (internal) chamber wall intersected with \overline{C}'_w . Denote this set by $[w]'^{\alpha_i, IN}$. Each external wall of the convex hull of Wx_o is parametrized by a simple roots α_i . We denote an *external* wall of Γ by $[w]'^{\alpha_i, OUT}$ if it intersects all the *internal* chamber walls except for $[w]'^{\alpha_i, IN}$.

For any $J \subset \Pi$ we define the subsets of \overline{C}'_w of dimension $|\Pi \setminus J|$,

$$\begin{cases} [w]'^{J, \Theta} = \bigcap_{\alpha_i \in J} [w]'^{\alpha_i, \Theta}, & \text{if } J \neq \emptyset, \\ [w]'^{J, \Theta} = C'_w, & \text{if } J = \emptyset, \end{cases}$$

where Θ is either *OUT* or *IN*. Thus we have the decomposition,

$$\overline{C}'_w = \bigcup_{\substack{J \subset \Pi \\ \Theta \in \{OUT, IN\}}} [w]'^{J, \Theta}.$$

DEFINITION 6.9. We will need to use the action of W on the set of signs \mathcal{E} of Definition 6.7. We now define gluing maps between the chamber walls denoted by $\{\epsilon\} \times [w]'^{\dots} = \{\epsilon\} \times \{w\} \times [w]'^{\dots}$ as follows: For the internal walls, we define

$$g_{w,i,IN} : \begin{array}{ccc} \{\epsilon\} \times [w]^{\alpha_i, IN} & \longrightarrow & \{s_{\alpha_i} \epsilon\} \times [ws_{\alpha_i}]^{\alpha_i, IN} \\ (\epsilon, w, x) & \longmapsto & (s_{\alpha_i} \epsilon, ws_{\alpha_i}, x) \end{array}$$

where note $ws_{\alpha_i}w^{-1}x = x$. For the external walls, we define

$$g_{w,i,OUT} : \begin{array}{ccc} \{\epsilon\} \times [w]^{\alpha_i, OUT} & \longrightarrow & \{\epsilon^{(i)}\} \times [w]^{\alpha_i, OUT} \\ (\epsilon, w, x) & \longmapsto & (\epsilon^{(i)}, w, x) \end{array}$$

where $\epsilon^{(i)} = (\epsilon_1, \dots, -\epsilon_i, \dots, \epsilon_l)$.

We denote \tilde{M} the disjoint union of all the chambers endowed with different signs,

$$\tilde{M} = \bigcup_{\substack{w \in W \\ \epsilon \in \mathcal{E}}} \{w\epsilon\} \times \overline{C}_{w^{-1}}.$$

We also denote M the topological space obtained from the disjoint union in \tilde{M} by gluing along the internal and external walls using the maps $g_{w,i,IN}$ and $g_{w,i,OUT}$. There is then a map

$$z : \tilde{M} \rightarrow M.$$

6.3. The negative walls. We now give a precise definition of the negative wall. Let us first define:

DEFINITION 6.10. First denote

$$\tilde{\Gamma}_\epsilon = \bigcup_{w \in W} \{w\epsilon\} \times \overline{C}_{w^{-1}}.$$

We now let Γ_ϵ denote the image of $\tilde{\Gamma}_\epsilon$ in M . Then after the identifications in M , this space becomes a copy of Γ .

NOTATION 6.11. Set $\epsilon' = w^{-1}\epsilon$ and recall the action of W and its subgroups on pairs (ϵ', w) where $\epsilon \in \mathcal{E}$ and $w \in W$ (Definition 6.7). Note that an α_i wall $[w]^{\alpha_i, IN}$ which is the intersection of two closed chambers $\{\epsilon'\} \times \overline{C}_w$ and $\{s_i \epsilon'\} \times \overline{C}_{ws_i}$ can be simply parametrized by the coset of w in $[w] \in W / \langle s_i \rangle$. To keep track of signs

we need to consider the two pairs (ϵ', w) and $(s_i \epsilon', ws_i)$. This constitutes the orbit of (ϵ', w) under the action of $W_{\{s_i\}}$. We have already denoted this orbit by $[\epsilon', w]$ in Definition 6.7 and now this orbit $[w^{-1}\epsilon, w]$ will also be used as a parameter to denote the corresponding internal wall in Γ_ϵ . This wall is called *negative* for α_i if $\epsilon' = (\epsilon'_1, \dots, \epsilon'_l)$ has $\epsilon'_i = -1$.

For a set $J \subset \Pi$ one can also consider the orbit of W_J denoted $[\epsilon', w]_J$ which will now denote the intersection:

$$[\epsilon', w]_J = z \left(\bigcap_{\sigma \in W_J} \{\sigma \epsilon'\} \times \overline{C}_{w\sigma^{-1}} \right)$$

This parametrizes an intersection of several walls. We call this J multi-wall intersection J -negative or just negative if it is such that for all $\alpha_i \in J$ there is $\sigma \in W_J$ such that $(\sigma \epsilon')_i = -1$ where (ϵ', w) is a representative where w has minimal length in its coset in W/W_J .

For any $\alpha_i \in \Pi$ we consider the set $R_{\epsilon, J}$ given by

$$R_{\epsilon, J} = \{[\epsilon', w] : [\epsilon', w] \text{ is negative} \}$$

When $J = \{\alpha_i\}$ we will just write $R_{\epsilon, i}$

Consider the subspace of M given by

$$\tilde{\Theta}_J^a = z \left(\bigcup_{\epsilon \in \mathcal{E}, w \in R_{\epsilon, J}} [w^{-1}\epsilon, w]_J \right).$$

EXAMPLE 6.12. We can now describe the topology of $\tilde{\Theta}_{\{1\}}^a$. Since $\tilde{\Theta}_{\{1\}}^a$ is smooth it will suffice to compute its Euler characteristic. That $\tilde{\Theta}_{\{1\}}^a$ is not orientable will follow.

Walls in a fixed Γ_ϵ are parametrized as in Notation 6.11. If we want to parametrize walls independently of each separate polytope, we consider colored Dynkin diagrams as in [5]. Intersections of walls are obtained by coloring more simple roots with $-s$ or $+s$. Thus we consider walls as parametrized by the full set of colored Dynkin diagrams (Definition 6.7).

The negative walls in $\Theta_{\{1\}}$ can then be listed: $[o_- - o - o, e]$, $[o_- - o - o, 2]$, $[o_- - o - o, 3]$, $[o_- - o - o, 23]$, $[o_- - o - o, 12]$, $[o_- - o - o, 32]$, $[o_- - o - o, 312]$, $[o_- - o - o, 123]$, $[o_- - o - o, 232]$, $[o_- - o - o, 1232]$, $[o_- - o - o, 2312]$, $[o_- - o - o, 12132]$.

Now boundaries must be considered. For example the boundaries of $[o_- - o - o, e]$ are $[o_- - o_- - o, e]$, $[o_- - o - o_-, e]$, $[o_- - o_+ - o, e]$, $[o_- - o - o_+, e]$. Therefore all these walls are part of $\Theta_{\{1\}}$. However $[o_- - o - o, 2]$ produces $[o_- - o_- - o, 2]$. Since the Weyl group W_S now includes s_2 then we can write this wall as $[o_+ - o_- - o, e]$ because $s_2(o_- - o_- - o) = o_+ - o_- - o$. Therefore the wall $[o_+ - o_- - o, e]$ must also be included in $\tilde{\Theta}_{\{1\}}^a$. Taking this into account we can easily count all the cells of $\tilde{\Theta}_{\{1\}}^a$. All the 1-cells of the form $[o_{\epsilon_1} - o_{\epsilon_2} - o, w]$ except $\epsilon_1 = \epsilon_2 = +$ appear. This gives $3 \times |W/W_{\{s_1, s_2\}}| = 3 \times 4$ such cells. We also get all the cells of the form $[o_- - o - o_\pm, w]$, a total of $2 \times |W/W_{\{s_1, s_3\}}| = 2 \times 6$. Hence a total of 24 cells of dimension one. Finally there are 7 cells of dimension 0. The Euler characteristic obtained is $12 - 24 + 7 = -5$. Since $\tilde{\Theta}_{\{1\}}^a$ is homeomorphic to a smooth compact surface, this completely describes its topology.

Note that the actual boundary maps involved in a homology computation are as described in section 4 of [5].

6.4. A graph associated to the negative walls in Γ_ϵ . Let us define a graph to describe the negative walls for a fixed $\alpha_i \in \Pi$.

DEFINITION 6.13. The graph $\mathbb{G}(\epsilon)$:

We consider a graph $\mathbb{G}(\epsilon)$ having vertices (ϵ', w) with $\epsilon' = w^{-1}\epsilon$.

(a) If all $\epsilon_j = -$ then all the pairs $(\epsilon', w), (s_j\epsilon', ws_j)$ are edges.

We now describe the edges when not all ϵ_j are negative.

First for all semisimple Lie algebras of rank $l \leq 3$:

$(\epsilon', w), (s_j\epsilon', ws_j)$ is an edge if and only if one of the following is satisfied:

(b) $i \neq j, C_{j,i} \neq -2, \epsilon'_j = +$

(c) $i \neq j, C_{j,i} = -2, \epsilon'_j = -$

(d) $i \neq j, C_{i,j} = 0, \epsilon'_i = -$

Note that if we fix i and j then the corresponding subdiagram of the Dynkin diagram has rank two. We will show below that these conditions lead to the correct description of the divisors Θ_i in the rank two cases. The condition d) of Definition 6.13 will correspond to the case of $A_1 \times A_1$. If we consider only Lie algebra of types A, D, E and G_2 , the conditions simplify to:

(a') $\alpha_j \in \Pi(\epsilon)$

(b') $i \neq j, \epsilon'_j = +$

(c') $i \neq j, s_j$ commutes with s_i and $\epsilon'_i = -$

The case $\alpha_i = \alpha_2$ in B_2 is some kind of exception which requires a separate rule given in condition c).

In general, if the rank is n given a subset $S \subset \Pi$ and $\epsilon \in \mathcal{E}$ we denote ϵ_S the restriction formed by the ordered $|S|$ -tuple consisting only of the ϵ_k with $\alpha_k \in S$.

We assume that all the edges of the graph have been defined for rank $|S| < n$. The pair $(\epsilon', w), (s_j\epsilon', ws_j)$ is an edge if there is $S \subset \Pi$ with $|S| < n, s_i, s_j \in S$ and there is $\sigma \in W_S$ with $w = w_1\sigma, \ell(w_1) + \ell(\sigma) = \ell(w)$ and $(\epsilon'_S, \sigma), (s_j\epsilon'_S, \sigma s_j)$ form an edge in the case of the split Lie subalgebra determined by S .

We now break up $R_{\epsilon,i}$ as a disjoint union of subsets consisting of negative walls belonging to the same connected component of the graph $\mathbb{G}(\epsilon)$. We thus obtain a set $I(\epsilon)$ consisting of subsets of $R_{\epsilon,i}$. The disjoint union $\bigcup_{\alpha \in I(\epsilon)} \alpha$ equals $R_{\epsilon,i}$.

DEFINITION 6.14. The graph \mathbb{G} : We now define a graph whose vertices are the elements $\alpha \in I(\epsilon)$ for $\epsilon \in \mathcal{E}$. If $\alpha \in I(\epsilon_1)$ and $\beta \in I(\epsilon_2)$, then there is an edge joining α to β if and only if there is w such that

(i) $[w^{-1}\epsilon_1, w] \in \alpha, [w^{-1}\epsilon_2, w] \in \beta$

(ii) Denoting $w^{-1}\epsilon_1 = \epsilon'$ then we have: $w^{-1}\epsilon_2 = (\epsilon')^{(i)}$

EXAMPLE 6.15. Consider the case of A_2 and $J = \{\alpha_1\}$. If $\epsilon = (--)$ condition a) in Definition 6.13 applies; however, as it turns out, condition b) alone will suffice in this case. We have the following edges indicated by \rightarrow connecting the only two negative walls. $((--), e) \rightarrow ((-+), s_1) \rightarrow ((-+), s_1s_2)$. We have the following set of negative walls $R_{(--),1} = \{[(--), e], [(-+), s_1s_2]\}$. We obtain that $I(--)$ consists of one single element $\alpha^{-} = \{[(--), e], [(-+), s_1s_2]\}$.

For $\epsilon = (-+)$ we obtain $R_{(-+),1} = \{[(-+), e], [(-+), s_2]\}$ and $I(-+)$ consists of one single element $\alpha^{-+} = \{[(-+), e], [(-+), s_2]\}$

For $\epsilon = (+-)$ we have $R_{(+-,)1} = \{[(-), s_2] \rightarrow [(-), s_1 s_2]\}$ and again there is one single element α^{+-} .

The graph \mathbb{G} consists of a “cycle” $\alpha^{--} \rightarrow \alpha^{-+} \rightarrow \alpha^{+-} \rightarrow \alpha^{--}$. For example, there is an edge $\alpha^{--} \rightarrow \alpha^{-+}$ because $[(-), e] \in \alpha^{--} \cap \alpha^{-+}$. If one consider the topological space consisting of the corresponding walls then what results is a circle in agreement with what was found in Proposition 5.1. This corresponds to Figure 1 where the divisor indicated by the number 1 is replaced with two walls -straight lines- joining at the center of the hexagons. The edges of \mathbb{G} correspond to intersections with the boundaries of the hexagons Γ_ϵ , that is with “subsystems”.

EXAMPLE 6.16. Consider the case of G_2 and $J = \{s_1\}$ with $\epsilon = (+-)$. The negative walls are parametrized by

$$[(-), s_2], [(-+), s_2 s_1 s_2], [(-), s_1, s_2], [(-+), s_1 s_2 s_1 s_2].$$

Note that $[(-), s_2], [(-+), s_2 s_1 s_2]$ are in the same connected component of $\mathbb{G}(\epsilon)$ since

$$(-, s_2) \rightarrow (-+, s_2 s_1) \rightarrow (-+, s_2 s_1 s_2);$$

where \rightarrow indicates an edge. However this process reaches a dead-end when we apply s_1 once more since one obtains $(-, s_2 s_1 s_2 s_1)$ but s_2 cannot be applied at this point because $\epsilon_2 = -1$. The connected component of the graph $\mathbb{G}(+-)$ which contains $[(-), s_2]$ then consists of

$$(-, s_2) \rightarrow (-+, s_2 s_1) \rightarrow (-+, s_2 s_1 s_2) \rightarrow (-, s_2 s_1 s_2 s_1).$$

From here

$$\alpha^{+-} = \{[(-), s_2], [(-+), s_2 s_1 s_2]\}$$

and similarly there is another set of negative walls

$$\beta^{+-} = \{[(-), s_1 s_2], [(-+), s_1 s_2 s_1 s_2]\}.$$

We have $I(+-) = \{\alpha^{+-}, \beta^{+-}\}$.

For $\epsilon = (-+)$, $\Pi(\epsilon) = \{s_1\}$ one obtains $\alpha^{-+} = \{[(-+), e], [(-+), s_2]\}$, $\beta^{+-} = \{[(-), s_1 s_2 s_1 s_2], [(-), s_2 s_1 s_2 s_1 s_2]\}$. For $\epsilon = (-)$ all the negative walls are in a single connected component. However, here, unlike what happens in the A_2 example one requires condition a) of Definition 6.13 with $\Pi(-) = \{s_1, s_2\}$. This allows the application of s_2 independently of the sign ϵ'_2 . We have $\alpha^{--} = \{[(-), e], [(-+), s_1, s_2], [(-), s_2 s_1 s_2], [(-+), s_2 s_1 s_2 s_1 s_2]\}$ and $I(-) = \{\alpha^{--}\}$.

The graph \mathbb{G} has nodes given by $\{\alpha^{+-}, \beta^{+-}, \alpha^{-+}, \beta^{+-}, \alpha^{--}\}$. The edges are $\alpha^{+-} \rightarrow \alpha^{-+}$, $\beta^{+-} \rightarrow \beta^{+-}$ and $\alpha^{--} \rightarrow x$ for $x = \alpha^{+-}, \beta^{+-}, \alpha^{-+}, \beta^{+-}$. This gives a total of six edges.

When one considers the topology of the sets of walls involved and the edges are regarded as the only gluings: α^{--} consists of two intersecting line segments and all the others consist of segments. What then results is a figure 8. The 6 edges are the intersection of this figure 8 with the boundaries of the polytopes Γ_ϵ (subsystems). Note that segments associated to α^{+-} and β^{+-} are being regarded as *disjoint*. However the two segments forming α^{--} are not disjoint because they form part of one single connected component of $\mathbb{G}(-)$. The topological space associated to these graphs and the negative walls will be made precise below for a general semisimple Lie algebra.

EXAMPLE 6.17. We now consider the case of B_2 , $J = \{s_2\}$ and $\epsilon = (+, -)$. We have edges $(+, -) \rightarrow (+, -, s_1) \rightarrow (+, -, s_1 s_2)$ but $(+, -, s_1 s_2)$ is a dead-end because s_1 cannot be applied since $\epsilon_1 = +$ and $C_{1,2} = -2$. We also have an edge $(+, -) \rightarrow (+, -, s_2)$ which leads to a dead-end for the same reason. This gives a set $\alpha^{+-} = \{[(+, -), e], [(+, -), s_1]\}$. Another connected component of the graph produces $\beta^{+-} = \{[(+, -), s_2 s_1], [(+, -), s_1 s_2 s_1]\}$.

For $\epsilon = (--)$ one obtains $\alpha^{--} = \{[(--), e], [(--), s_1 s_2 s_1]\}$ and for $\epsilon = (-+)$ $\alpha^{-+} = \{[(-+), s_1], [(-+), s_2 s_1]\}$. Hence a graph results with edges $\alpha^{--} \rightarrow x$ and $\alpha^{-+} \rightarrow x$ where $x = \alpha^{+-}, \beta^{+-}$ giving a total of four edges in \mathbb{G} . Again we consider the topology of the sets of walls involved and, as in the previous examples, the edges in \mathbb{G} are regarded as the only gluings between these segments. We obtain four segments corresponding to the elements in $I(\epsilon)$ giving rise to a circle that intersects the boundaries of the Γ_ϵ at four points (the edges of the graph \mathbb{G}). This corresponds to $\Theta_{\{1\}}$ in Proposition 5.2.

6.5. The spaces of negative walls Θ_i^a . Fix an element $\alpha \in I(\epsilon)$. We consider the disjoint union

$$\bigcup_{\alpha \in I(\epsilon), [\epsilon', w] \in \alpha, \epsilon \in \mathcal{E}} \{\alpha\} \times [w^{-1}\epsilon, w].$$

We define gluings for any pair α, β which are joined by an edge of the graph \mathbb{G} .

$$\begin{aligned} g : (\alpha, [\epsilon', w]) &\longrightarrow [(\beta, \epsilon'^{(i)} w \sigma^{-1})] \\ (\alpha, \epsilon', w, x) &\longmapsto (\beta, \epsilon'^{(i)}, w, x) \end{aligned}$$

CONJECTURE 6.18. *There is a homeomorphism $g : \Theta_i \rightarrow \Theta_i^a$.*

EXAMPLE 6.19. The topology of $\Theta_{\{2\}}^a$ in the case of A_3 can be computed explicitly and shown to correspond to $\Theta_{\{2\}}$. We first compute the Euler characteristic of $\tilde{\Theta}_{\{2\}}^a$ using the method in Example 6.12. One obtains twelve 2-cells, twenty four 2 cells and seven 1-cells giving the same Euler characteristic as in the case of $\tilde{\Theta}_{\{1\}}^a$. However the sets $I(\epsilon)$ for $\epsilon = (+ - +)$ and $\epsilon = (- + -)$ contain two elements. This can be seen in Figure 4 where the corresponding paths of dotted lines are disconnected. The recipe for the construction of $\Theta_{\{2\}}^a$ corresponds to separating the two cones obtained by joining these paths to the center of each of these polytopes. This introduces two additional points! Hence the Euler characteristic for $\Theta_{\{2\}}^a$ becomes -3.

One now notes that $\Theta_{\{2\}}^a$ remains singular as can be seen in Figure 4 where in the boundary of the polytope Γ_{---} there are two disconnected paths of dotted lines. It is possible to resolve this singularity by simply separating the center of this polytope into two separate points. This gives rise to a compact surface of Euler characteristic -2 since one additional point is added. The resulting surface can be seen to be non-orientable. We thus obtain that $\Theta_{\{2\}}^a$ is homeomorphic to $\Theta_{\{2\}}$. The compactification of the isospectral manifold of A_2 reappears but only as a resolution of singularities of the Painlevé divisor.

References

- [1] M. Adler and P. van Moerbeke, The Toda lattice, Dynkin diagrams, singularities and Abelian varieties, *Invent. math.* **103** (1991) 223-278.
- [2] M. Adler, L. Haine and P. van Moerbeke, Limit matrices for the Toda flow and periodic flags for loop groups, *Math. Ann.* **296** (1993) 1-33.

- [3] M. F. Atiyah, Convexity and commuting hamiltonians, *Bull. London Math. Soc.* **14** (1982) 1-15.
- [4] O. I. Bogoyavlensky, On perturbations of the periodic Toda lattice, *Comm. math. Phys.*, **51** (1976) 201-209.
- [5] L. Casian and Y. Kodama, Toda lattice and toric varieties for real split semisimple Lie algebras. Preprint OSU-MRI 99-17 (math-SG/9912021), *Pacific J. of Math.* (2001) in press (math.SG/9912021).
- [6] L. Casian and Y. Kodama, Twisted Tomei manifolds and Toda lattice to appear in *Contemporary Mathematics* (math.GT/0104054).
- [7] N. M. Ercolani, H. Flaschka and L. Haine, Painlevé balances and dressing transformations. in "Painlevé transcendents, their asymptotics and physical applications", Eds. P. Levi and P. Winternitz, NATO ASI Series, Series B: Physics Vol. 278 (Plenum Press, 1992) 249-260.
- [8] H. Flaschka and L. Haine. Varietes de drapeaux et reseaux de Toda. *Mathematische Zeitschrift*, **208** (1991), 545-556.
- [9] Y. Kodama and J. Ye, Toda Lattices with indefinite metric II: Topology of the iso-spectral manifolds. *Physica D*, **121** (1998), 89-108.
- [10] Y. Kodama, Topology of the real part of hyperelliptic Jacobian associated with the periodic Toda lattice, (nlin.SI/0201009).
- [11] B. M. Kostant, On Whittaker Vectors and Representation Theory, *Invent. Math.*, **48** (1978), 101-184.
- [12] B. M. Kostant, The solution to a generalized Toda lattice and representation theory, *Adv. Math.* **34** (1979) 195-338.
- [13] M. Jimbo and T. Miwa, Soliton and infinite dimensional Lie algebras, *Publ. RIMS, Kyoto Univ.*, **19** (1983) 943-1001.
- [14] D. Mumford, *Tata Lectures on Theta II*, Progress in Mathematics **43** (Birkhäuser, Boston 1984).
- [15] P. Vanhaecke, Integrable systems and symmetric products of curves. *Math. Z.*, **227** (1998) 93-127.

DEPARTMENT OF MATHEMATICS, OHIO STATE UNIVERSITY, COLUMBUS, OH 43210
E-mail address: `casian@math.ohio-state.edu`

DEPARTMENT OF MATHEMATICS, OHIO STATE UNIVERSITY, COLUMBUS, OH 43210
E-mail address: `kodama@math.ohio-state.edu`

Superposition Principle for Oscillatory Solutions of Integrable Systems.

Mikhail Kovalyov

ABSTRACT. Linear waves are often represented as linear superposition of harmonics $e^{i(kx-\omega t)}$. In this paper nonlinear analogues of both the functions $e^{i(kx-\omega t)}$ and the concept of superposition are considered. The parameters introduced provide a generalization of the concept of scattering data to solutions of integrable systems for which classical scattering data do not make sense.

Mathematics subject classification index: 37K15, 35Q53, 42A99

Key words: KdV, KP, integrable systems, solitons, harmonic solitons, breathers, harmonic breathers, positons.

0. Introduction: linear waves and their properties

The most important aspect in the theory of linear differential equations is that of *linear superposition*. Its formulation is quite simple: if u_1 and u_2 are solutions of a linear partial differential equation then so are their linear combinations. That naturally leads to the problem of finding a minimal set of solutions of a partial differential equation whose linear superpositions would produce all or, at least, a sufficiently large class of solutions of that equation. Such a minimal set is often referred to as *basis*. For a sufficiently large class of equations used to describe miscellaneous *linear waves*, the basis is often comprised of oscillatory solutions of the type

$$\cos(\beta(\boldsymbol{\lambda}) + \boldsymbol{\lambda} \cdot \boldsymbol{x} + \omega t) \tag{0.1a}$$

where $\beta(\boldsymbol{\lambda}) \in \mathbb{R}$ is the phase, $\boldsymbol{x} \in \mathbb{R}^n$ is the space coordinate, $\boldsymbol{\lambda} \in \mathbb{R}^n$ is the momentum, t is time and ω is the energy which itself is often a function of $\boldsymbol{\lambda}$ whose exact form

$$\omega = \omega(\boldsymbol{\lambda}) \tag{0.1b}$$

is known as the *dispersion relationship*.

By taking finite and infinite linear combinations of (0.1)

$$\sum_{\mathbf{k} \in \mathcal{N}} \cos(\beta(\boldsymbol{\lambda}_{\mathbf{k}}) + \boldsymbol{\lambda}_{\mathbf{k}} \cdot \boldsymbol{x} + \omega(\boldsymbol{\lambda}_{\mathbf{k}})t) u_{\mathbf{k}}, \quad \mathcal{N} \subset \mathbb{N}^n, \boldsymbol{\lambda}_{\mathbf{k}} \in \mathbb{R}^n, u_{\mathbf{k}} \in \mathbb{R} \tag{0.2a}$$

© 2002 American Mathematical Society

one may construct a fairly large class of real functions in the form of trigonometric sums/series (0.2a). Taking

$$\mathcal{N} = \mathbb{N}^n, \quad \lambda_{\mathbf{k}} = \mathbf{k}\Delta\lambda, \quad u_{\mathbf{k}} = \hat{u}(\lambda_{\mathbf{k}})\Delta\lambda^n$$

with a small parameter $\Delta\lambda$ and some, possibly generalized, function $\hat{u}(\lambda)$ and passing to the limit as $\Delta\lambda \rightarrow 0$, one obtains Fourier integrals

$$\int_{\mathbb{R}^n} \cos(\beta(\lambda) + \lambda \cdot \mathbf{x} + \omega(\lambda)t) \hat{u}(\lambda) d\lambda, \quad d\lambda = d\lambda_1 \dots d\lambda_n. \quad (0.2b)$$

Linear superposition and formulas (0.2a,c) are mathematical facets of what in physics is known as *interference* and/or *diffraction* of linear waves, [Fey 1], which in this paper will be referred to as *linear interference* and *linear diffraction*.

Although nowadays linear interference and diffraction are often taken for granted without much forethought, their role in physics and natural sciences cannot be underestimated. At the dawn of quantum mechanics, when physicists had to come to grips with the concept of wave-particles, the phenomena of linear interference and diffraction essentially served to define a wave as a motion exhibiting linear interference and diffraction. Wave-particles then were defined as localized packets of waves or *wave-packets*.

Linear diffraction was first observed for light by the primeval man in the form of a rainbow, then for X-rays by Max Von Laue and later for electrons by Clinton J. Davisson and George P. Thomson. It occurs when a wave interacts with an object producing waves of the type $\rho e^{i(\lambda \cdot x + \omega t + \varphi)}$, $\rho > 0$. Two such waves

$$\tilde{\rho} e^{i(\tilde{\lambda} \cdot x + \tilde{\omega} t + \tilde{\varphi})} \quad \text{and} \quad \tilde{\tilde{\rho}} e^{i(\tilde{\tilde{\lambda}} \cdot x + \tilde{\tilde{\omega}} t + \tilde{\tilde{\varphi}})} \quad (0.3a)$$

in turn interact with each other, provided

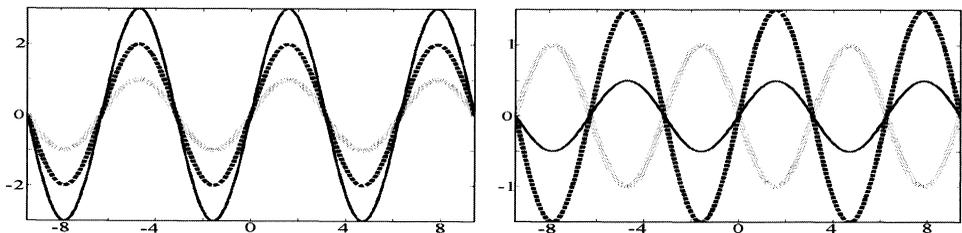
$$\tilde{\tilde{\lambda}} = \tilde{\lambda}, \quad \tilde{\tilde{\omega}} = \tilde{\omega}, \quad \text{and} \quad \tilde{\tilde{\varphi}} - \tilde{\varphi} = \pi n \quad (0.3b)$$

to produce a wave

$$[\tilde{\tilde{\rho}} + (-1)^n \tilde{\rho}] e^{i(\tilde{\lambda} \cdot x + \tilde{\omega} t + \tilde{\varphi})}. \quad (0.3c)$$

The latter interaction of the two waves, which is exactly what linear interference is, can result in amplification or annihilation of the wave motion in certain regions of space providing numerous diffraction patterns known in physics [Fey 1].

Fig. 2



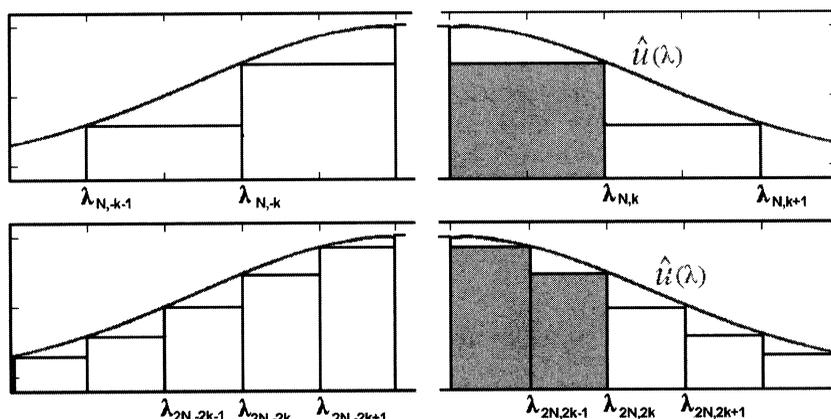
That the resulting wave (0.3c) is an exact sum of the interacting waves (0.3a) subject to conditions (0.3b) is far from obvious and there is no apriori reason why it should be so. It has been observed experimentally for some waves and serves as

a physical proof that the equations describing such waves must be linear. Fig. 2 illustrates linear interference with waves (0.3a) shown in thicker broken lines and the resulting wave (0.3c) shown in thinner solid line.

Linear interference, as given by formulas (0.3), underlies the very existence of the Fourier integrals (0.2b). Indeed, for a sufficiently smooth $\hat{u}(\lambda)$ the integral (0.2b) is defined to be the limit of Fourier sums, which in one-dimensional case may be taken in the form

$$S_N = \sum_{k=1}^{2^N N} \cos(\beta(\lambda_{N,k}) + \lambda_{N,k}x + \omega t) \hat{u}(\lambda_{N,k}) \Delta\lambda, \quad \lambda_{N,k} = 2^{-N}k, \quad \Delta\lambda = 2^{-N}. \quad (0.4a)$$

Fig. 3



For the sequence S_N of Fourier sums to converge as $N \rightarrow \infty$, one must have

$$\lim_{N \rightarrow \infty} \left[\cos(\beta(\lambda_{N,k}) + \lambda_{N,k}x + \omega t) \hat{u}(\lambda_{N,k}) - \frac{1}{2} \sum_{q=0}^1 \cos(\beta(\lambda_{2N,2k-q}) + \lambda_{2N,2k-q}x + \omega t) \hat{u}(\lambda_{2N,2k-q}) \right] = 0 \quad (0.4b)$$

as shown in Fig. 3, i. e. for large N the term $\cos(\beta(\lambda_{N,k}) + \lambda_{N,k}x + \omega t) \hat{u}(\lambda_{N,k})$ must be linear superposition of $\frac{1}{2} \cos(\beta(\lambda_{2N,2k}) + \lambda_{2N,2k}x + \omega t) \hat{u}(\lambda_{2N,2k})$ and $\frac{1}{2} \cos(\beta(\lambda_{2N,2k-1}) + \lambda_{2N,2k-1}x + \omega t) \hat{u}(\lambda_{2N,2k-1})$, in accordance with (0.3).

About the most popular of all linear equations describing a variety of waves arising in mechanics of fluids and solids, electro-magnetic theory, theory of relativity, etc. is the classical wave equation

$$\frac{\partial^2 u}{\partial t^2} - \sum_{j=1}^n \frac{\partial^2 u}{\partial x_j^2} = 0. \quad (0.5a)$$

Functions (0.1) subject to an implicit dispersion relation

$$\omega^2 = |\boldsymbol{\lambda}|^2 \quad (0.5b)$$

satisfy (0.5a). A broader class of solutions of (0.5a) is furnished by (0.2a,c) subject to (0.5b).

Note that for the wave-equation, as for most other equations describing linear waves, multiplication of the basic solution (0.1a) by energy ω (momentum λ_j , resp.) and application of the operator $\frac{1}{i} \frac{\partial}{\partial t}$ ($\frac{1}{i} \frac{\partial}{\partial x_j}$, resp.) yield the same result, i.e.

$$\omega = \frac{1}{i} \frac{\partial}{\partial t} \quad \text{and} \quad \lambda_j = \frac{1}{i} \frac{\partial}{\partial x_j}, \quad \text{when applied to} \quad e^{i(\boldsymbol{\lambda} \cdot \mathbf{x} + \omega t)} \quad (0.6)$$

Unfortunately, once one moves from linear to nonlinear phenomena the governing equations become nonlinear and the linear superposition principle as well as the concepts of linear interference, linear diffraction and even that of a basic solution lose their meanings. Remarkably, a class of nonlinear partial differential equations known as integrable systems, albeit nonlinear, exhibit behavior that in many aspects is identical to that of linear systems. For, at least, some of them nonlinear analogues of basic solutions (0.1a), linear interference described by (0.3) and the superposition formulas (0.2a,c) exist.

1. Derivation of the KP and KdV equations

Most integrable equations/systems appear in the same manner, that will be illustrated here on the KP/KdV equations. To derive them one may start with a two-dimensional long wave of the form

$$e^{i(\omega t + \lambda_1 x + \lambda_2 y)}, \quad \sqrt{\lambda_1^2 + \lambda_2^2} \ll 1 \quad (1.1a)$$

that is almost one-dimensional, that is

$$\lambda_2 \ll \lambda_1. \quad (1.1b)$$

The dispersion relationship is assumed to be of the form $\omega^2 = \Omega(|\boldsymbol{\lambda}|^2)$ which generalizes the dispersion relationship (0.5b) for the classical wave-equation. Requirement that the function Ω be sufficiently smooth to be well-approximated for small $\boldsymbol{\lambda}$ by its McLaurin series $\sum_{n=0}^{\infty} \Omega_n |\boldsymbol{\lambda}|^{2n}$ and $\Omega_0 = \Omega(0) = 0$ allow one to simplify the dispersion relationship to the form

$$\omega^2 \approx \Omega_1 (\lambda_1^2 + \lambda_2^2) + \Omega_2 (\lambda_1^2 + \lambda_2^2)^2, \quad \Omega_1 > 0.$$

Since ω is a two-valued function of $|\boldsymbol{\lambda}|$, for a fixed value of $\boldsymbol{\lambda}$ the dispersion relationship incorporates two waves of the form (1.1a) moving with speed $|\boldsymbol{\lambda}| \omega^{-1}$ in the directions of $\pm \frac{1}{|\boldsymbol{\lambda}|} \boldsymbol{\lambda}$. After appropriate re-scaling

$$\omega^2 \rightarrow \frac{2\Omega_1^2}{3|\Omega_2|} \omega^2, \quad \lambda_j^2 \rightarrow \frac{2\Omega_1}{3|\Omega_2|} \lambda_j^2, \quad \alpha^2 = \frac{\Omega_2}{|\Omega_2|},$$

if necessary, and restriction to the wave moving towards $-\infty$ along the x - axis, the dispersion relationship becomes

$$\lambda_1 \omega \approx \lambda_1 \sqrt{\lambda_1^2 + \lambda_2^2 + \frac{2}{3\alpha^2}(\lambda_1^2 + \lambda_2^2)^2} \approx \lambda_1 \left(\lambda_1 + \frac{1}{3\alpha^2} \lambda_1^3 \right) + \frac{\lambda_2^2}{2} \tag{1.1c}$$

Due to (0.6), a wave (0.1) satisfies the dispersion relationship (1.1c) if and only if it satisfies the equation

$$(u_t - u_x - \frac{1}{3\alpha^2} u_{xxx})_x = \frac{1}{2} u_{yy}. \tag{1.1d}$$

To accommodate certain conservation laws usually arising in applications a non-linear term $(6uu_x)_x$ is added to the left hand side yielding

$$(u_t - u_x + 6uu_x - \frac{1}{3\alpha^2} u_{xxx})_x = \frac{1}{2} u_{yy}. \tag{1.1e}$$

The change of variables

$$u_{old} = \frac{1}{3\alpha^2} u_{new}, \quad t_{old} = -3\alpha^2 t_{new}, \quad x_{old} = x_{new} + 3\alpha^2 t_{new}, \quad y_{old} = \frac{1}{\sqrt{2}} y_{new} \tag{1.1f}$$

turns equations (1.1d,e) into

$$(u_t + u_{xxx})_x = -3\alpha^2 u_{yy}, \quad \alpha^2 = \pm 1. \tag{IKP}$$

$$(u_t + 6uu_x + u_{xxx})_x = -3\alpha^2 u_{yy}, \quad \alpha^2 = \pm 1. \tag{KP}$$

known respectively as the *linear Kadomtsev-Petviashvili* equation and the *Kadomtsev-Petviashvili* equation. Properties of some solutions of KP are determined by the sign of α^2 so to distinguish between the two cases, the KP (*IKP*, resp.) equation is often called KPI (*IKPI*, resp.) if $\alpha^2 = -1$, i.e. $\alpha = i$ and KP II (*IKPII*, resp.) if $\alpha^2 = +1$, i.e. $\alpha = 1$. As far as the solutions considered in this paper are concerned, the difference between KPI and KP II is minimal and the insertion of α into formulas will take care of it. So no distinction will be made between KPI and KP II.

Solutions of KP and *IKP*, that are independent of y and decay to 0 as $x \rightarrow \infty$ satisfy correspondingly

$$u_t + u_{xxx} = 0, \tag{IKdV}$$

$$u_t + 6uu_x + u_{xxx} = 0, \tag{KdV}$$

known respectively as the *linear Korteweg-de Vries* equation and the *Korteweg-de Vries* equation. Equations *IKdV* and *KdV* may be viewed as specific cases of *IKP* and *KP* equations corresponding to exactly one-dimensional waves, i.e. when (1.1b) is replaced with $\lambda_2 = 0$.

The basic solutions (0.1a) of *IKP* take form

$$\cos 2\lambda \left[\beta_\lambda + x + 4(\lambda^2 - 3\alpha^2 \mu^2)t \right] \cos 4\lambda \mu \left[\beta_\mu + y \right], \quad \text{phases } \beta_\lambda, \beta_\mu \in \mathbb{R} \tag{1.2a}$$

with simplified notations

$$\mathbb{R}^2 \ni \lambda = \{ \lambda_1, \lambda_2 \} = \{ \lambda, \mu \}, \text{ i.e. } \lambda_1, \lambda_2 \text{ are written as simply } \lambda, \mu.$$

Basic solutions (0.1a) of *IKdV*

$$\cos 2\lambda \left[\beta_\lambda + x + 4\lambda^2 t \right], \quad \text{phase } \beta_\lambda \in \mathbb{R} \tag{1.2b}$$

are obtained by taking

$$\mathbb{R}^1 \ni \lambda = \{\lambda_1\} = \{\lambda\}$$

or by setting $\mu = 0$ in (1.2a).

Since, in applications at least, u is small, the term $6uu_x$ is much smaller than the rest of the terms in KP (KdV, resp.) and thus KP (KdV, resp.) itself may be viewed as a nonlinear perturbation of l KP (l KdV, resp.) by the term $6uu_x$. Since the very derivation of KP (KdV, resp.) starts with functions (1.1a), which are, of course, equivalent to (1.2a) ((1.2b), resp.), for a physical model described by KP (KdV, resp.) to be valid it must admit solutions whose local behavior, at least, is very close to that of the functions (1.2a) ((1.2b), resp.). Such solutions may be viewed as nonlinear perturbations of (1.2a) ((1.2b), resp.) by $6uu_x$. Note also that linear combinations of (1.2a) ((1.2b), resp.) are solutions of l KP (l KdV, resp.) and, since the very derivation of KP (KdV, resp.) is based on l KP (l KdV, resp.), for a physical model described by KP (KdV, resp.) to be valid, it must admit solutions whose local behavior, at least, is very close to that of linear combinations of (1.2a) ((1.2b), resp.). Such solutions may be viewed as nonlinear perturbations of solutions of l KP (l KdV, resp.) representable as linear combinations of (1.2a) ((1.2b), resp.).

2. Basic solutions of KdV

Since (1.2b) play the role of basic solutions of l KdV so should their nonlinear analogues for KdV and thus the latter should be sought in the form of the simplest components comprising more general solutions of KdV.

Decaying at infinity solutions of KdV are usually constructed by first setting a function, [Kov 1],

$$F(x, t) = \sum_{j=1}^N 2\mu_j e^{\mu_j \phi_j + 8\mu_j^3 t - \mu_j x} + \frac{1}{2\pi} \int_{-\infty}^{+\infty} r(k) e^{i(8k^3 t + kx)} dk, \tag{2.1a}$$

subject to

$$\mu_j > 0, \quad \mu_j \phi_j \text{ are either real or real } \pm \frac{\pi i}{2}, \quad |r(k)| < 1, \quad r(-k) = \overline{r(k)}, \tag{2.1b}$$

and then solving the integral equation

$$K(x, y, t) + F(x + y, t) + \int_x^{+\infty} K(x, z, t) F(y + z, t) dz = 0. \tag{2.1c}$$

The corresponding solution of KdV then is obtained via

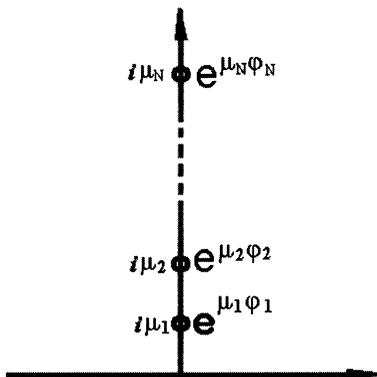
$$u(t, x) = 2 \frac{d}{dx} K(x, x, t). \tag{2.1d}$$

If $r(k) \equiv 0$ the procedure yields the well-known class of multi-soliton solutions

$$u^{(N\text{-sol})} = 2 \frac{d^2}{dx^2} \ln \det A^{(N\text{-sol})}, \tag{2.2a}$$

$$A_{mn}^{(N\text{-sol})} = \delta_{mn} + \frac{2\sqrt{\mu_n \mu_m}}{\mu_n + \mu_m} e^{\mu_n \phi_n + \mu_m \phi_m + 4(\mu_n^3 + \mu_m^3)t - (\mu_n + \mu_m)x}, \quad n, m = 1, \dots, N \tag{2.2b}$$

Fig. 4



with each term of the sum

$$F(x, t) = \sum_{j=1}^N 2\mu_j e^{8\mu_j^3 t - \mu_j x} e^{\mu_j \phi_j} \quad (2.2c)$$

generating a soliton. The multi-soliton solution (2.2a,b) itself may be interpreted as a nonlinear superposition of N solitons. It is convenient to picture each soliton as generated by a charge $e^{\mu_j \phi_j}$ placed at $i\mu_j$ as shown in Fig. 4. Each soliton corresponds to a term in the sum in the right hand side of (2.2c) and thus nonlinear superposition of solitons is equivalent to addition of corresponding terms

in (2.2c). The solution (2.2a,b) is obtained by substituting (2.2c) into the integral equation (2.1c) and using ansatz $K(x, y) = \sum_{n=1}^N K_n(x) e^{-\mu_n y}$ to solve (2.1c).

If all $\phi_j = -\infty$, the sum in (2.1a) drops out and only the integral term is left:

$$F(x, t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} r(k) e^{i(8k^3 t + kx)} dk, \quad |r(k)| < 1, \quad r(-k) = \overline{r(k)}. \quad (2.3)$$

By analogy with multi-soliton case, one may view the corresponding solution of KdV as a nonlinear superposition of some basic solutions generated by functions

$$F_\varepsilon(x, t) = \varepsilon \left[e^{p\varepsilon + 2i\lambda\gamma + 8(i\lambda + \varepsilon)^3 t} + e^{p\varepsilon - 2i\lambda\gamma + 8(-i\lambda + \varepsilon)^3 t} \right], \quad (2.4a)$$

$$e^{p\varepsilon} = |r(\lambda)|, \quad 2\lambda\gamma = -arg(r(\lambda)). \quad (2.4b)$$

as $\varepsilon \rightarrow 0$. Substituting (2.4) into (2.1c), solving the obtained integral equation just the same way it is done for the multi-soliton case and letting $\varepsilon \rightarrow 0$, one obtains, [Kov 1]:

$$u(t, x) = 8\lambda \left[\frac{\lambda \sin 2\lambda(\gamma - 4\lambda^2 t - x)}{2\lambda(p - 12\lambda^2 t - x) - \sin 2\lambda(\gamma - 4\lambda^2 t - x)} \right] - 8\lambda^2 \left[\frac{\cos 2\lambda(\gamma - 4\lambda^2 t - x) - 1}{2\lambda(p - 12\lambda^2 t - x) - \sin 2\lambda(\gamma - 4\lambda^2 t - x)} \right]^2. \quad (2.5)$$

Formula (2.5) gives the required nonlinear analogues of basic solutions (1.2b) of $IKdV$, the functions (2.5) will be referred to as *harmonic breathers*. Evolution of a typical harmonic breather is shown in Fig. 5. To construct nonlinear superposition of basic solutions (2.5) one should note that such superposition should be consistent with that for multi-soliton solutions. So it is only reasonable to define nonlinear superposition of the solutions of KdV generated by

$$F_\varepsilon(k) = \sum_{j=1}^N \varepsilon \left[e^{p_j \varepsilon + 2i\lambda_j \gamma_j + 8(i\lambda_j + \varepsilon)^3 t} + e^{p_j \varepsilon - 2i\lambda_j \gamma_j + 8(-i\lambda_j + \varepsilon)^3 t} \right] \quad (2.6a)$$

$$e^{p_j \varepsilon} = |r(\lambda_j)|, \quad 2\lambda_j \gamma_j = -arg(r(\lambda_j)) \quad (2.6b)$$

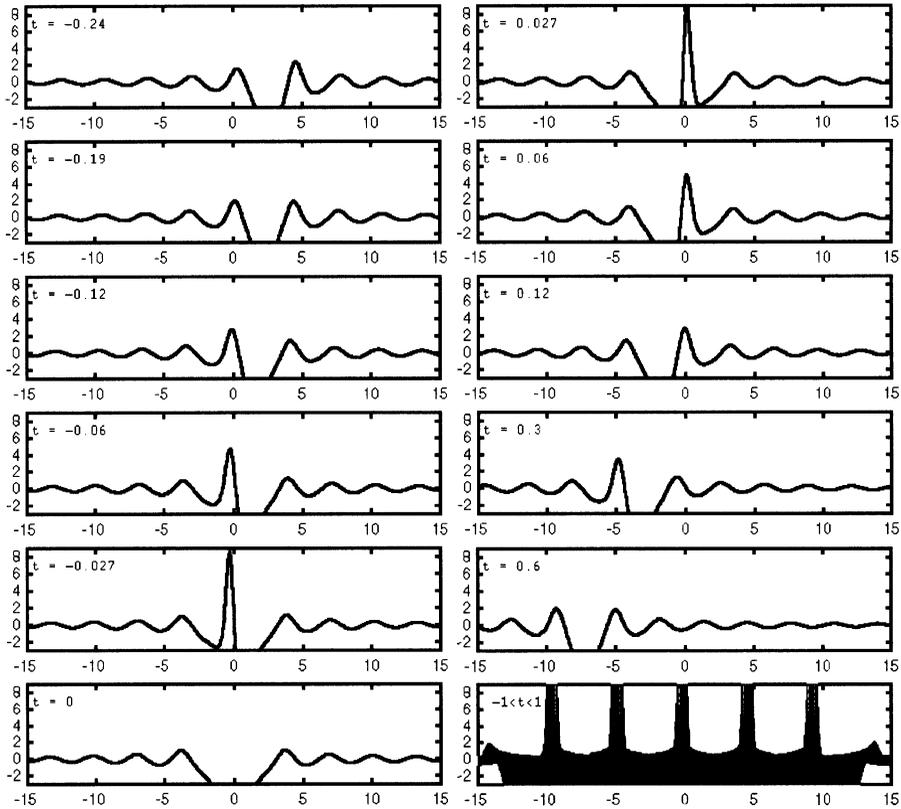
as $\varepsilon \rightarrow 0$. Substituting (2.6) into (2.1c), solving the obtained integral equation just the same way it is done for the multi-soliton case and letting $\varepsilon \rightarrow 0$ one obtains

$$u(t, x) = 2 \frac{d^2}{dx^2} \ln \det \mathcal{D}, \quad \mathcal{D} \text{ is an } N \times N \text{ matrix,} \tag{2.7a}$$

$$\mathcal{D}_{nm} = \begin{cases} \frac{\sin(\Gamma_n - \Gamma_m)}{\lambda_n - \lambda_m} - \frac{\sin(\Gamma_n + \Gamma_m)}{\lambda_n + \lambda_m}, & n \neq m, \\ \tau_n = p_n - 12\lambda_n^2 t - x - \frac{\sin 2\Gamma_n}{2\lambda_n}, & n = m, \end{cases} \tag{2.7b}$$

$$\Gamma_n = \lambda_n(\gamma_n - 4\lambda_n^2 t - x). \tag{2.7c}$$

Fig. 5



Time evolution of a single harmonic breather with $\lambda = 1, \gamma = 0, p = 0$. The last frame contains graphs of the solution for all values of time from the interval $-1 < t < 1$ superimposed on the same frame.

Formulas (2.7) describe superposition of N harmonic breathers. It is convenient to picture each harmonic breather in (2.7) as generated by a pair of charges $e^{p_j \varepsilon + 2\lambda_j \gamma_j i}$, $e^{p_j \varepsilon - 2\lambda_j \gamma_j i}$ placed correspondingly at $i\varepsilon_j - \lambda_j$, $i\varepsilon_j + \lambda_j$ as $\varepsilon \rightarrow 0$, as shown in Fig. 6. For the case of arbitrary $r(k)$ in (2.3), the sums (2.6) may be viewed as approximations of the integral in (2.3), as shown in Fig. 7. One would expect the solutions generated by (2.6) to approximate, in a certain sense, solutions

generated by arbitrary $r(k)$, subject, of course, to the standard restrictions usually imposed on this function [Dei 2], such as conditions of smoothness, $r(0) = -1$, etc. For illustrative purposes the reflection coefficient is shown as a real function, while it is in fact complex.

Fig. 6

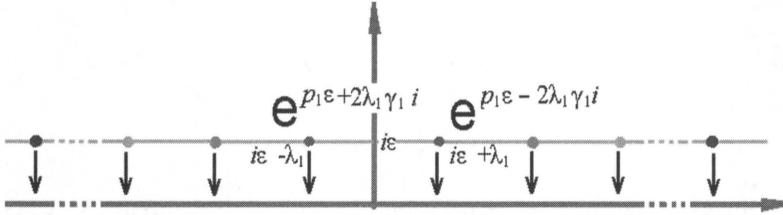
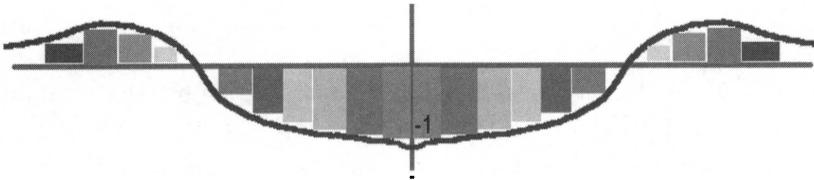


Fig. 7



For $N \geq 1$ and large $p_j - 12\lambda_j^2 t - x$'s, (2.7) immediately implies, [Kov 1]:

$$u = 4 \sum_{j=1}^N \frac{\lambda_j}{p_j - 12\lambda_j^2 t - x} \sin 2\lambda_j(\gamma_j - 4\lambda_j^2 t - x) + o\left(\frac{1}{p_j - 12\lambda_j^2 t - x}\right) \quad (2.8a)$$

If in addition t, x are restricted to some bounded space-time domain whose size is much smaller than the absolute values of all p_j 's, the previous estimate can be further simplified to

$$u = 4 \sum_{j=1}^N \frac{\lambda_j}{p_j} \sin 2\lambda_j(\gamma_j - 4\lambda_j^2 t - x) + o\left(\frac{1}{p_j}\right). \quad (2.8b)$$

The formula shows that if all $|p_j|$'s are large, in the space-time domain whose dimensions are much smaller than all $|p_j|$'s, the functions given by the superposition formula (2.7) essentially become Fourier sums. A sequence of such functions, may, under appropriate conditions, converge to a nonlinear analogue of the Fourier integral, similarly to how finite Fourier sums (0.2a) converge to Fourier integrals (0.2b). Recall that as shown in (0.4) the physical phenomenon behind such convergence of Fourier sums to Fourier integrals is linear interference described by (0.3). Remarkably, a similar phenomenon exists for (2.7). To see it consider (2.7) with $N = 2$:

$$u = 2 \frac{d^2}{dx^2} \ln \left\{ \left[p_1 - 12\lambda_1^2 t - x - \frac{\sin 2\lambda_1(\gamma_1 - 4\lambda_1^2 t - x)}{2\lambda_1} \right] \left[p_2 - 12\lambda_2^2 t - x - \frac{\sin 2\lambda_2(\gamma_2 - 4\lambda_2^2 t - x)}{2\lambda_2} \right] - \left[\frac{\sin(\lambda_1(\gamma_1 - 4\lambda_1^2 t - x) - \lambda_2(\gamma_2 - 4\lambda_2^2 t - x))}{\lambda_1 - \lambda_2} - \frac{\sin(\lambda_1(\gamma_1 - 4\lambda_1^2 t - x) + \lambda_2(\gamma_2 - 4\lambda_2^2 t - x))}{\lambda_1 + \lambda_2} \right]^2 \right\}. \quad (2.9a)$$

As $\lambda_2 \rightarrow \lambda_1$, $\lambda_2\gamma_2 \rightarrow \lambda_1\gamma_1$, solution (2.9a) degenerates into a single harmonic breather

$$u = 2 \frac{d^2}{dx^2} \ln \left[p_{12} - 12\lambda_1^2 t - x - \frac{\sin 2\lambda_1(\gamma_1 - 4\lambda_1^2 t - x)}{2\lambda_1} \right] \tag{2.9b}$$

where p_{12} satisfies

$$\frac{1}{p_{12} - \lim_{\lambda_2 \rightarrow \lambda_1} \frac{\lambda_2\gamma_2 - \lambda_1\gamma_1}{\lambda_2 - \lambda_1}} = \frac{1}{p_2 - \lim_{\lambda_2 \rightarrow \lambda_1} \frac{\lambda_2\gamma_2 - \lambda_1\gamma_1}{\lambda_2 - \lambda_1}} + \frac{1}{p_1 - \lim_{\lambda_2 \rightarrow \lambda_1} \frac{\lambda_2\gamma_2 - \lambda_1\gamma_1}{\lambda_2 - \lambda_1}}. \tag{2.9c}$$

If $|p_1|$, $|p_2|$, $|p_{12}|$ are large and the term $\lim_{\lambda_2 \rightarrow \lambda_1} \frac{\lambda_2\gamma_2 - \lambda_1\gamma_1}{\lambda_2 - \lambda_1}$ can be neglected, the formula simplifies to

$$\frac{1}{p_{12}} \approx \frac{1}{p_1} + \frac{1}{p_2}. \tag{2.9d}$$

Formula (2.9a) is a nonlinear superposition of two harmonic breathers whose asymptotic behavior for large $|p_j - 12\lambda_j^2 t - x|$ (large $|p_j|$ and small x, t , resp.) is given by (2.8):

$$u_j \approx \frac{4\lambda_j \sin 2\lambda_j(\gamma_j - 4\lambda_j^2 t - x)}{p_j - 12\lambda_j^2 t - x} \approx \frac{4\lambda_j \sin 2\lambda_j(\gamma_j - 4\lambda_j^2 t - x)}{p_j}, \quad j = 1, 2. \tag{2.10a}$$

Due to (2.9b) as $\lambda_2 \rightarrow \lambda_1$, $\lambda_2\gamma_2 \rightarrow \lambda_1\gamma_1$ their superposition has the asymptotic behavior for large $|p_{12} - 12\lambda_1^2 t - x|$ (large $|p_{12}|$ and small x, t , resp.)

$$u_{12} \approx \frac{4\lambda_1 \sin 2\lambda_1(\gamma_1 - 4\lambda_1^2 t - x)}{p_{12} - 12\lambda_1^2 t - x} \approx \frac{4\lambda_1 \sin 2\lambda_1(\gamma_1 - 4\lambda_1^2 t - x)}{p_{12}} \tag{2.10b}$$

Formula (2.8b) suggests that $-\frac{4\lambda_j}{p_j}$ play the role of nonlinear Fourier coefficients, in which case (2.9a) becomes the nonlinear analogue of (0.3c), that is, formulas (2.9) represent nonlinear analogue of linear interference described in §0, which will be referred to as *nonlinear interference*.

Tails of harmonic breathers seem to possess all the properties required for modulation, that is creation of more general functions by means of nonlinear superposition. Some examples of that were constructed in [Kov 2]. Yet the use of the tails of harmonic breathers for such constructions is not very satisfactory due to the following reasons:

- 1) the amplitude of harmonic breathers does not stay constant but decays in x as $x \rightarrow \pm\infty$;
- 2) the standard Riemann-Hilbert problem approach to finding solutions of KdV does not relate to the method of harmonic breathers;
- 3) there is an ambiguity in the formula $u = 4 \sum_{n=1}^N \frac{\lambda_n}{p_n} \sin 2\lambda_n(\gamma_n - 4\lambda_n^2 t - x) + o\left(\frac{1}{p_n}\right)$, as changing $p_k \rightarrow -p_k$, $\lambda_k\gamma_k \rightarrow \lambda_k\gamma_k + \frac{\pi}{2}$ leaves the right hand-side invariant, leading to non-uniqueness of the parameters $p_j, \lambda_j\gamma_j$.

3. Double layers of harmonic breathers

To remedy the problems mentioned at the end of the previous section, define a *double layer of harmonic breathers* to be superposition of harmonic breathers obtained from (2.7) by

replacing N with $2M$, (3.1a)

choosing $\lambda_{2n}\gamma_{2n} = \lambda_{2n-1}\gamma_{2n-1} + \frac{\pi}{2}$, (3.1b)

choosing $p_{2n} = -p_{2n-1}$. (3.1c)

In addition assume that $\lambda_{2n} - \lambda_{2n-1}$ are sufficiently small to allow

free interchange of λ_{2n-1} and λ_{2n} , as well as $\Gamma_{2n-1} + \frac{\pi}{2}$ and Γ_{2n} . (3.1d)

If the values of all $|p_j|$'s are sufficiently large and $t \ll \min_j |p_j|$, then the real axis can be broken up into part $\mathbb{R} = T_l \cup S_l \cup K \cup S_r \cup T_r$ as shown in Fig. 8, where S_l, S_r are correspondingly the left and right *singular sets* containing all singularities of a double layer given by (2.7, 3.1), T_l, T_r support the left and right *tails* that satisfy asymptotics (2.8a), the set K will be referred to as *core* and G is the buffer zone between the singular sets S_l, S_r and the core.

Fig. 8



Within the core, double layers are smooth, do not decay in x and behave very much like functions (1.2b). Consider for simplicity's sake the simplest double layer of harmonic breathers that consists of just two components and may be called a *harmonic couple*:

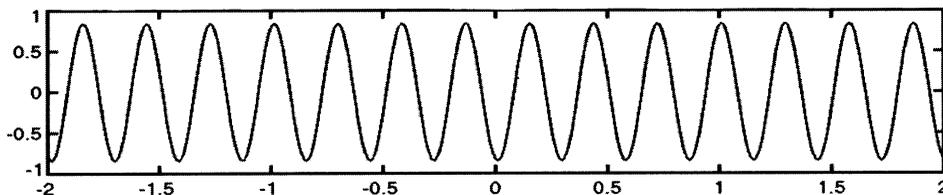
$$u = 2 \frac{d^2}{dx^2} \ln \left\{ \left[p_1 - 12\lambda_1^2 t - x - \frac{\sin 2\lambda_1(\gamma_1 - 4\lambda_1^2 t - x)}{2\lambda_1} \right] \left[p_1 + 12\lambda_2^2 t + x + \frac{\sin 2\lambda_2(\gamma_2 - 4\lambda_2^2 t - x)}{2\lambda_2} \right] + \left[\frac{\sin(\lambda_1(\gamma_1 - 4\lambda_1^2 t - x) - \lambda_2(\gamma_2 - 4\lambda_2^2 t - x))}{\lambda_1 - \lambda_2} - \frac{\sin(\lambda_1(\gamma_1 - 4\lambda_1^2 t - x) + \lambda_2(\gamma_2 - 4\lambda_2^2 t - x))}{\lambda_1 + \lambda_2} \right]^2 \right\}. \quad (3.2a)$$

Due to (3.1) asymptotic, (2.8a) in K takes form:

$$u = \frac{8\lambda_1}{p_1} \sin 2\lambda_1(\gamma_1 - 4\lambda_1^2 t - x) + o\left(\frac{1}{p_1}\right) + o(|\lambda_2 - \lambda_1|) \quad (3.2b)$$

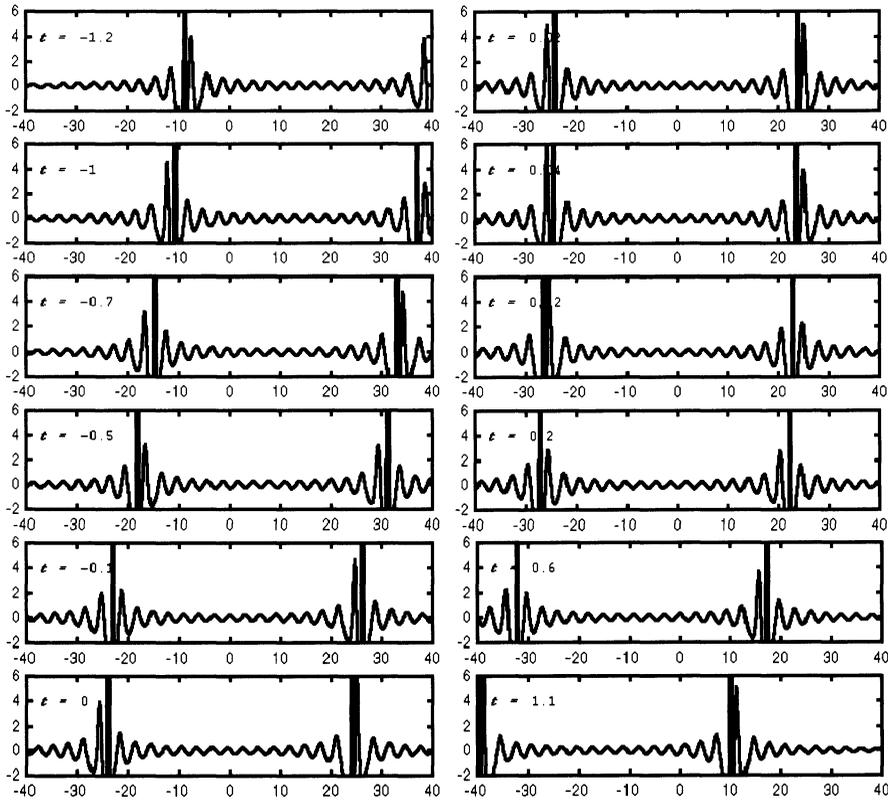
The graph of a typical harmonic couple inside core is shown in Fig. 9 and time evolution of the core of a typical harmonic breather is shown in Fig. 10. A double layer in general may be viewed as nonlinear superposition of harmonic couples. Since a single harmonic couple is just a nonlinear superposition of two harmonic

Fig. 9



A harmonic couple with $\lambda_1 = 11, \lambda_2 = 11.01, \gamma_1 = 0, \gamma_2 = \frac{\pi}{2}, p_1 = 33, p_2 = -33$.

Fig. 10



Time evolution of a single harmonic breather with $\lambda = 1, \gamma = 0, p = 0$.

breathers, harmonic couples also exhibit nonlinear interference and satisfy asymptotic expansion (2.8a), which, due to (3.1) takes form

$$u = 8 \sum_{j=1}^N \frac{\lambda_{2j-1}}{p_{2j-1}} \sin 2\lambda_{2j-1}(\gamma_{2j-1} - 4\lambda_{2j-1}^2 t - x) + \text{lower order terms} \quad (3.3)$$

that is somewhat similar to (2.8b) but is valid on larger space-time domains.

As shown in [Kov 1] the operator

$$\frac{\partial^2}{\partial x^2} \psi(k, t, x) + u(t, x)\psi(k, t, x) = -k^2\psi(k, t, x), \quad \lim_{x \rightarrow \infty} \psi(k, t, x) = 0 \quad (3.4a)$$

with u given by (2.7),(3.1) has $2M$ eigenfunctions $\psi_n(t, x)$ that satisfy the equation

$$\sum_{m=1}^{2M} \mathcal{D}_{mn} \psi_m = \sin \Gamma_n, \quad n = 1, \dots, 2M \quad (3.4b)$$

where \mathcal{D}_{mn} is given by (2.7b) subject to conditions (3.1).

The potential u itself is expressible in terms of the eigenfunctions as

$$\begin{aligned}
 u &= 2 \frac{d}{dx} \ln \det \mathcal{D} = 2 \sum_{n,m=1}^{2M} (\mathcal{D}^{-1})_{mn} \frac{d}{dx} \mathcal{D}_{mn} = -4 \sum_{n,m=1}^{2M} (\mathcal{D}^{-1})_{mn} \sin \Gamma_n \sin \Gamma_m = \\
 &-4 \sum_{m=1}^{2M} \psi_m \sin \Gamma_m = -4 \frac{d}{dx} \sum_{m=1}^M \left[\sin \Gamma_{2m-1} \psi_{2m-1} + \sin \Gamma_{2m} \psi_{2m} \right] \tag{3.4c}
 \end{aligned}$$

As shown in [Kov 1], formulas (3.4) may be viewed as discretization of the operator

$$\frac{\partial^2}{\partial x^2} \phi + v(t, x) \phi = -k^2 \phi \tag{3.5a}$$

with continuous spectrum, the Riemann-Hilbert problem

$$\phi(k, t, x) = e^{ikx} + \frac{1}{2\pi i} \int_{-\infty}^{+\infty} \frac{r(\zeta) \phi(\zeta) e^{ik^3 t + i(k+\zeta)x}}{k + \zeta + i0} d\zeta, \quad \phi(k, t, x) \sim e^{ikx} \text{ as } x \rightarrow +\infty. \tag{3.5b}$$

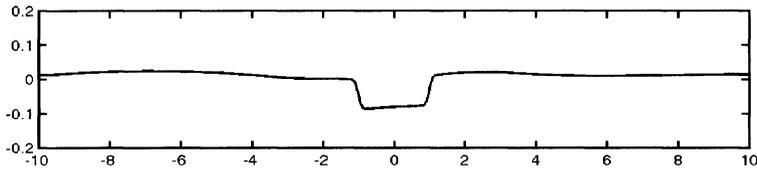
corresponding to (3.5a) and the formula expressing the potential v in terms of the solutions ϕ 's of the Riemann-Hilbert problem

$$v(t, x) = \frac{\partial}{\partial x} \left\{ \frac{1}{\pi} \int_{-\infty}^{+\infty} r(\zeta) \phi(x, \zeta) e^{ik^3 t + i\zeta x} d\zeta \right\}. \tag{3.5c}$$

Thus inside the core K , double layers of harmonic breathers do not have the drawbacks mentioned at the end of the previous section. For a given physical space-time domain of finite but otherwise arbitrary dimensions, one can construct double layers whose cores contain the physical domain. Within the physical domain the double layers can modulate more complicated functions just like basic functions (1.2b) do so by means of the Fourier sums/integrals (0.2a,c). Figs. 11-14 show examples of such nonlinear modulation. Nonlinear modulation by harmonic breathers/ harmonic couples is remarkably similar to linear modulation by functions (1.2b), exhibiting even nonlinear analogue of the Gibbs phenomenon, [Kov 1]. Note that the modulation is valid only inside bounded physical domains, that, though, is not a drawback for in applications all physical domains are bounded.

For an arbitrary function u satisfying the standard conditions of the Inverse Scattering Theory, [Dei 2], one may construct a sequence u_ε of double layers generated by (2.6) subject to (3.1) with $\varepsilon = \frac{1}{\log_2 2N}$ and all other parameters determined by (2.6) and (3.1). The sequence u_ε should, in a certain sense, converge to u . The nature of such convergence is far from clear and, at the moment, there is no theoretical justification behind it. Yet numerical experiments suggest that for, at least some, u the approximating sequence does converge to u and it does so somewhat similarly to how the sequence of trigonometric sums (0.4a) converges to the corresponding trigonometric integral. As $\varepsilon \rightarrow 0$, the singular sets S_r, S_l move to $\pm\infty$, the core of u_ε gets larger and larger approaching \mathbb{R} and the space-time domain where u_ε well approximates u gets larger and larger approaching $\mathbb{R}^2 \ni \{x, t\}$. Since the elements u_ε of the approximating sequence are given by (2.7) subject to (3.1), u itself may be viewed as some sort of nonlinear integral whose role is similar to that of (0.2b) in linear theory.

Fig. 11



A nonlinear analogue of an approximation to $-0.1\chi(x)$ where $\chi(x)$ is the characteristic function of $[-1,1]$. Obtained by taking $N = 56$, $\lambda_n = 0.25n$, $\lambda_{2n-1}\gamma_{2n-1} = \lambda_{2n}\gamma_{2n} - \frac{\pi}{2} = \frac{\pi}{4}$, $p_n = (-1)^{n+1} \times \frac{10\pi n^2}{\sin 0.5n}$.

Fig. 12

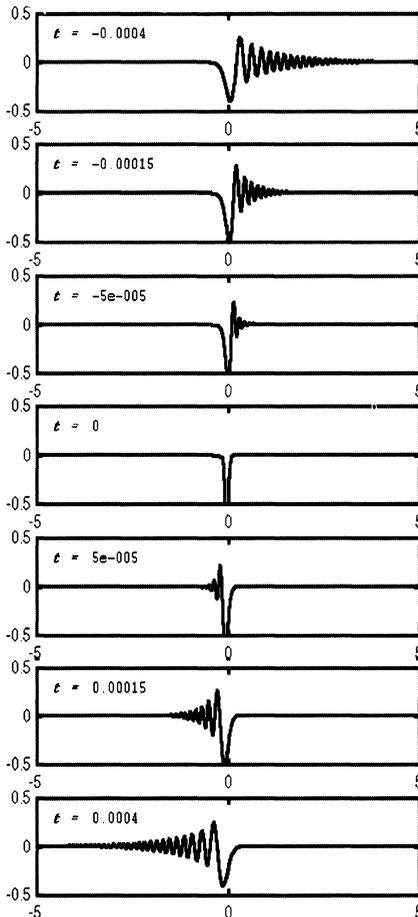
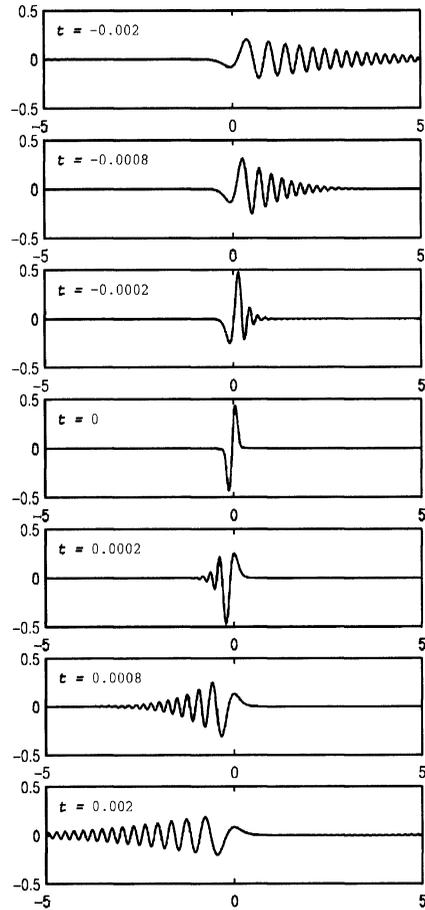
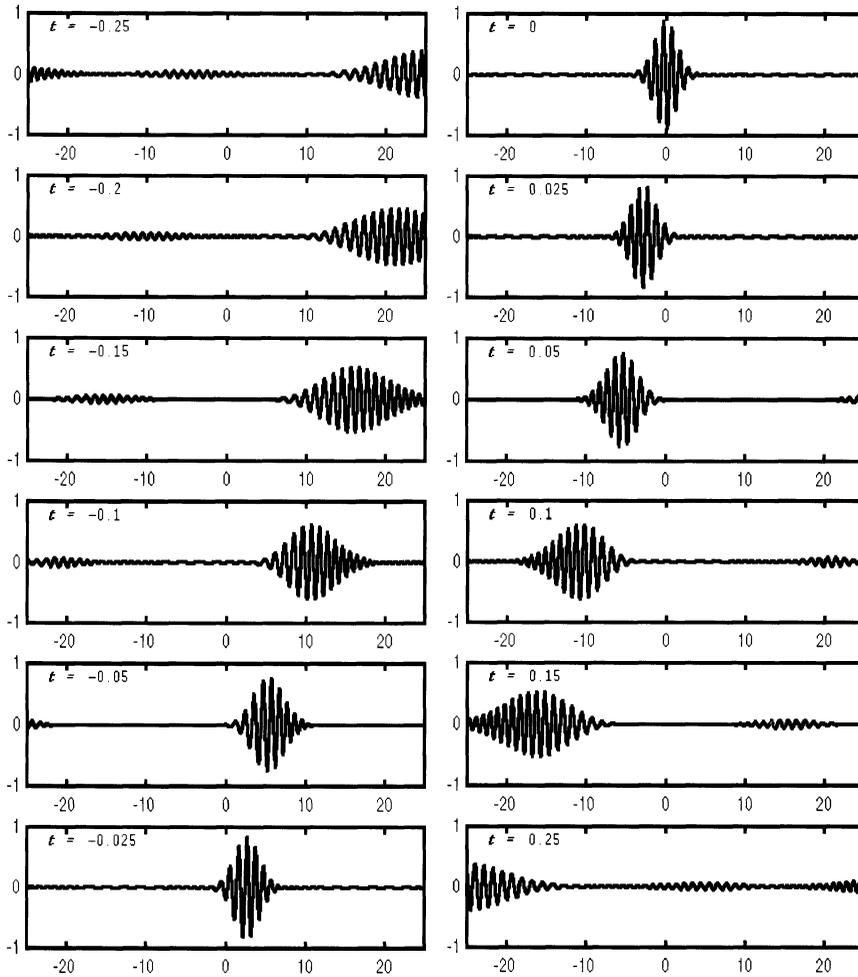


Fig. 13



Time evolution of resonances with short life-span. The left column is obtained by taking $N = 120$, $\lambda_n = 0.25n$, $\lambda_n\gamma_n = \frac{\pi}{4}$, $p_n = \frac{2n^2}{\sin 0.05\lambda_n} \left(1 - \frac{n}{N + 1/3}\right)^{-1}$. The right column is obtained by taking $N = 90$, $\lambda_n = 0.3n$, $\lambda_n\gamma_n = 0$, $p_n = \frac{2.7n^2}{1 - \cos 0.15\lambda_n} \left(1 - \frac{n}{N + 1/3}\right)^{-1}$.

Fig. 14



Time evolution of a wave-packet with $N = 64$, $\lambda_n = 3 + 0.05(n - 31.5)$, $\lambda_{2n-1}\gamma_{2n-1} = 0$, $\lambda_{2n}\gamma_{2n} = \frac{\pi}{2}$, $p_n = (-1)^n \times 80\lambda_n e^{4(\lambda_n - 3)^2}$.

4. KP equation.

Following the ideas of the previous section one can compute two-dimensional version of (2.7), [Kov 3]:

$$u(x, y, t) = 2 \frac{\partial^2}{\partial x^2} \ln \det \mathcal{K}, \quad \mathcal{K} = \begin{pmatrix} K_{11} & K_{12} & \dots & K_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ K_{N1} & K_{N2} & \dots & K_{NN} \end{pmatrix}, \quad (4.1a)$$

$$K_{nn} = -(\varrho_n + x - 2\mu_n y + 12[\lambda_n^2 - \alpha^2 \mu_n^2]t) + \frac{\cos 2\Gamma_n}{2\lambda_n} \quad (4.1b)$$

$$K_{nk} = \left[\frac{(\lambda_n + \lambda_k) \cos(\Gamma_n + \Gamma_k)}{\alpha^2(\mu_n - \mu_k)^2 + (\lambda_n + \lambda_k)^2} - \frac{(\lambda_n - \lambda_k) \sin(\Gamma_n - \Gamma_k)}{\alpha^2(\mu_n - \mu_k)^2 + (\lambda_n - \lambda_k)^2} \right] + \alpha \left[\frac{(\mu_n - \mu_k) \sin(\Gamma_n + \Gamma_k)}{\alpha^2(\mu_n - \mu_k)^2 + (\lambda_n + \lambda_k)^2} + \frac{(\mu_n - \mu_k) \cos(\Gamma_n - \Gamma_k)}{\alpha^2(\mu_n - \mu_k)^2 + (\lambda_n - \lambda_k)^2} \right], \quad (4.1c)$$

$$\Gamma_n = \gamma_n + \lambda_n x - 2\lambda_n \mu_n y + 4\lambda_n (\lambda_n^2 - 3\alpha^2 \mu_n^2) t. \quad (4.1d)$$

subject to the physical conditions (1.1a,b) used in the derivation of KP in the form

$$\max_j \mu_j \ll \min_j \lambda_j \ll 1. \quad (4.1e)$$

Functions given by formulas (4.1a-d) satisfy KP for arbitrary values of the constants λ_n 's, μ_n 's, γ_n 's, ϱ_n 's. However, in order for (4.1a-c) to be nonlinear analogues of the Fourier sums of basic solutions (1.2a) of lKP

$$u = \sum_{j,k} u_{j,k} \cos 2\lambda_j \left[\beta_{\lambda,j} + x + 4(\lambda_j^2 - 3\alpha^2 \mu_k^2) t \right] \cos 4\lambda_j \mu_k \left[\beta_{\mu,k} + y \right], \quad (4.2)$$

their parameters must satisfy the additional assumptions that

$$N = 2LM \quad (4.3a)$$

and for all $l = 1, \dots, L$, $m = 1, \dots, M$, [Kov 3],

$$\lambda_{(2m-2)L+l} = \lambda_{(2m-1)L+l} = \lambda_l, \quad \varrho_{(2m-2)L+l} = \varrho_{(2m-1)L+l}, \quad (4.3b)$$

$$\mu_{(2m-2)L+l} = -\mu_{(2m-1)L+l} = \mu_{(2m-2)L+1} \text{ which, for simplicity's sake will be denoted hereafter as simply } \mu_m, \quad (4.3c)$$

$$\gamma_{(2m-2)L+l} = \lambda_l \beta_{\lambda_l} + 2\lambda_l \mu_m \beta_{\mu_m}, \quad \gamma_{(2m-1)L+l} = \lambda_l \beta_{\lambda_l} - 2\lambda_l \mu_m \beta_{\mu_m}. \quad (4.3d)$$

Also define

$$\Lambda_{lm} = \lambda_l \beta_{\lambda_l} + \lambda_l x + 4\lambda_l (\lambda_l^2 - 3\alpha^2 \mu_m^2) t, \quad M_m = \mu_m (\beta_{\mu_m} - y). \quad (4.3e)$$

A nonlinear analogue of the two-dimensional basic motion (1.2a), that will be referred to as a *two-dimensional harmonic breather*, is obtained by taking $L = M = 1$ in (4.2). Fig. 16 shows a picture of a typical two-dimensional harmonic breather. For $L, M > 1$ formula (4.2), subject to (4.3), may be viewed as nonlinear superposition of LM two-dimensional harmonic breathers.

If x, y, t are restricted to a finite region

$$\Omega_{X,Y,T} = \left\{ (x, y, t) \mid |x| \leq X, |y| \leq Y, |t| \leq T \right\} \quad (4.4a)$$

and

$$|\varrho_n| \gg X, Y, T, \frac{1}{2\lambda_n}, \mathcal{K}_{nk}, \text{ for all } n, k \quad (4.4a)$$

then the functions (4.1) subject to (4.3) satisfy

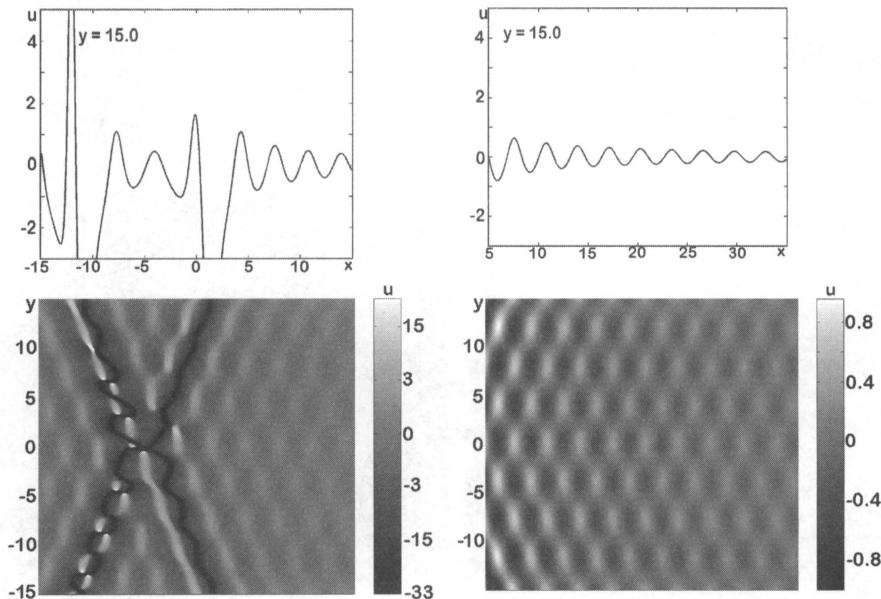
$$u(x, y, t) = \sum_{l,m=1}^{L,M} \frac{8\lambda_l}{\varrho_{2(m-1)L+l}} \cos 2\Lambda_{lm} \cos 4\lambda_l M_{lm} + O\left(\frac{1}{\varrho_{2(m-1)L+l}^2}\right) \quad (4.4c)$$

which is the two-dimensional analogue of (2.8b). Just like corresponding solutions of KdV, basic motions of KP exhibit nonlinear interference as shown in [Kov 3] and thus can be used to generate more complicated solutions of KP, two examples are shown in Figs. 17, 18.

Just like for KdV, due to the presence of singularities, formula (4.1) cannot be used to construct localized non-singular solutions of KP in all of $\Omega_{\infty,\infty,\infty} =$

$\{(x, y, t) \mid |x| \leq +\infty, |y| \leq +\infty, |t| \leq +\infty\}$. Instead, for an apriory given domain $\Omega_{X,Y,T}$ of finite but otherwise arbitrary size, one may select parameters in

Fig. 15



A two-dimensional harmonic breather of KPII equation with $\lambda_1 = \lambda_2 = 1.0$, $\mu_1 = -\mu_2 = 0.2$, $\gamma_1 = \gamma_2 = 0.0$, $\varrho_1 = \varrho_2 = 5.0$ at time $t = 0$.

(4.1), (4.3) to keep the singularities outside the domain while producing a solution localized within $\Omega_{X,Y,T}$. Formula (4.1) subject to (4.3) thus provides a method of constructing solutions of KP localized within $\Omega_{X,Y,T}$. Since, except for a few trivial ones, there are no solutions of KP localized in $\Omega_{\infty,\infty,\infty}$, localization within $\Omega_{X,Y,T}$ is second best, for in applications all physical space-time domains are bounded.

Generally, the solutions generated by (4.1) subject only to conditions (4.3) will have drawbacks similar to those of the solutions of KdV described at the end of §2 so additional conditions similar to (3.1) need to be imposed to remedy these drawbacks. Such additional conditions would lead to two-dimensional analogues of double layers (2.7),(3.1). The construction of two-dimensional analogues of double layers though has not been carried out.

Fig. 16

$\alpha = 1, L = M = 7, \text{ all } \beta_{\lambda l} = \beta_{\mu m} = 0,$
 $\text{all } \chi_n = 0, \mu_{(2m-1)L+l} = -\mu_{(2m-2)L+l}$
 $= 0.4 \frac{m}{l}, \lambda_{(2m-1)L+l} = \lambda_{(2m-2)L+l}$
 $= \frac{l}{4}, \varrho_{(2m-1)L+l} = \varrho_{(2m-2)L+l} =$
 $50l \frac{L+1}{L+1-l} \frac{M+1}{M+1-m}.$

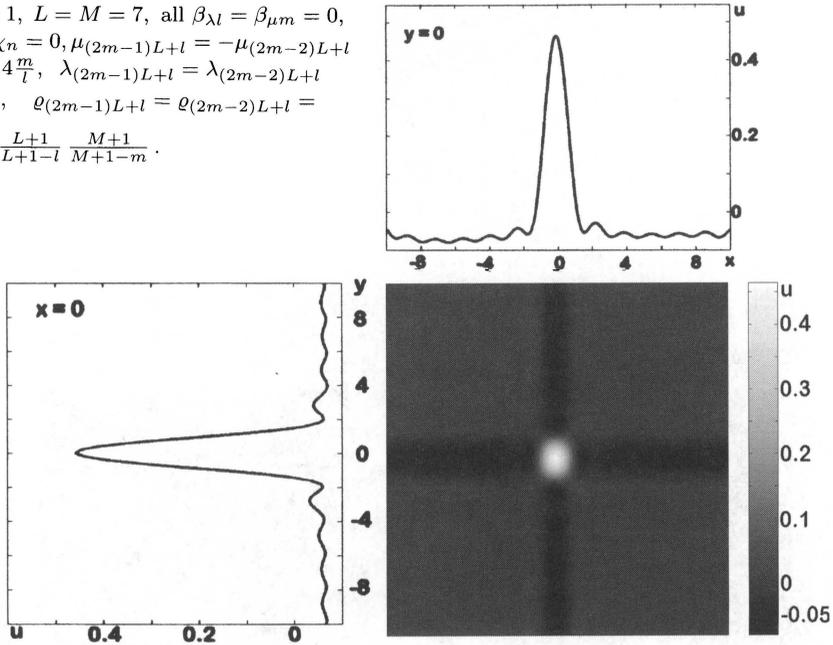
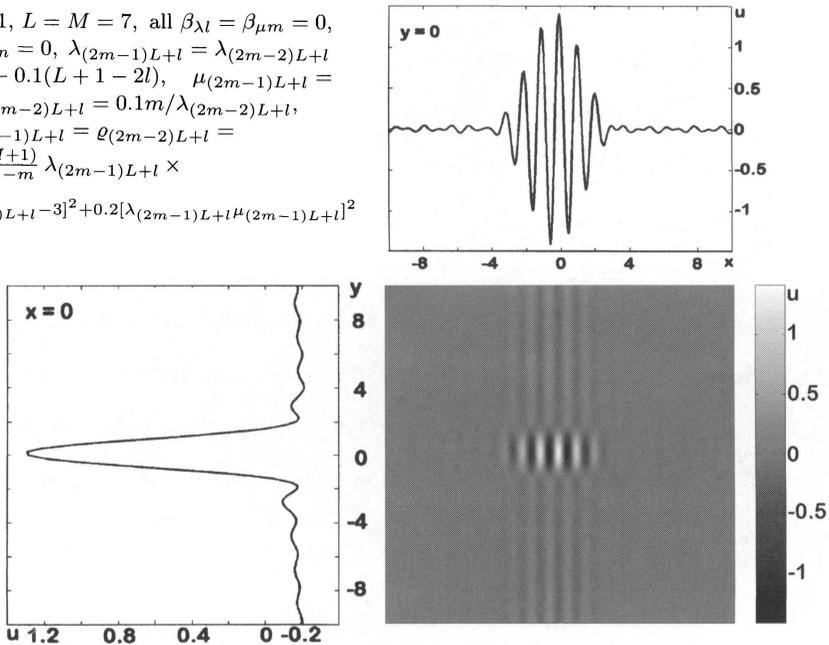


Fig. 17

$\alpha = 1, L = M = 7, \text{ all } \beta_{\lambda l} = \beta_{\mu m} = 0,$
 $\text{all } \chi_n = 0, \lambda_{(2m-1)L+l} = \lambda_{(2m-2)L+l}$
 $= 3 - 0.1(L + 1 - 2l), \mu_{(2m-1)L+l} =$
 $-\mu_{(2m-2)L+l} = 0.1m/\lambda_{(2m-2)L+l},$
 $\varrho_{(2m-1)L+l} = \varrho_{(2m-2)L+l} =$
 $\frac{80(M+1)}{M+1-m} \lambda_{(2m-1)L+l} \times$
 $e^{4[\lambda_{(2m-1)L+l}-3]^2 + 0.2[\lambda_{(2m-1)L+l}\mu_{(2m-1)L+l}]^2}$



Acknowledgements

The author would like to thank the organizers of the Conference for the invitation and a very nice and productive conference. The current paper is essentially based on [Kov 1,3], Figs. 16-18 are taken from [Kov 3]. Part of the work on KP was done together with I. Bica and became a part of his thesis. Some aspects of harmonic breathers can be found in [Mat 1], where they are derived by means of the Darboux transformation, and in [Novi 1]. Most references in this work are made to the author's papers [Kov 1,2,3] simply because the necessary details are provided there in the appropriate form. References to the original sources are generally omitted in this paper but may be found in [Kov 1,2,3].

References

- [Dei1] P. Deift, S. Venakides, and X. Zhou, *An extension of the steepest descent method for Riemann-Hilbert problem: the small dispersion limit of the Korteweg-de Vries (KdV) equation*, Proc. Natl. Acad. Sci. USA, **95**(2), (1998), 450-454.
- [Dei2] P. Deift, and E. Trubowitz, *Communication on Pure and Applied Mathematics*, Vol. XXXII, 1979, pp. 121-251.
- [Fey1] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, V. I, Addison-Wesley Publishing Company, 1977.
- [Kor1] D.J. Korteweg, and G. de Vries, *On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves*, Phil. Mag., **39** (5), (1895), 422-43.
- [Kov1] M. Kovalyov, *On a class of solutions of KdV*, (Submitted for publication, 2001).
- [Kov2] M. Kovalyov, *Modulating properties of harmonic breather solutions of KdV equation*, Journal of Physics A: Math. Gen., **31** (1998), 5117-5128.
- [Kov3] M. Kovalyov, and I. Bica, *Some properties of slowly decaying oscillatory solutions of KP*, (Submitted for publication, 2001).
- [Mat1] V.B. Matveev, *Asymptotics of the multipositon-soliton τ -function of the Korteweg-de Vries equation and the supertransparency*, Journal of Math. Phys., **35** (1994), 2955.
- [Nov1] S. Novikov, S.V. Manakov, L.P. Pitevskii, and V.E. Zakharov, *Theory of Solitons, The Inverse Scattering Method*, Consultants Bureau, Plenum, New York (in Russian), November 1984.
- [Nov1] R. Novikov, and G. Khenkin, *Oscillating weakly localized solutions of the Korteweg-de Vries equation*, *Teoreticheskaya i Matematicheskaya Fizika*, **61** (2), (November 1984), pp.199-213, (in Russian).
- [Ven1] S. Venakides, *The zero dispersion limit of the KdV equation with nontrivial reflection coefficient*, Communications on Pure and Applied Mathematics, **38** (1985), 125-155.
- [Ven2] S. Venakides, *The generation of modulated wavetrains in the solution of the KdV equation*, Communications on Pure and Applied Mathematics, **38** (1985), 883-909.
- [Zak1] V.E. Zakharov, and A.B. Shabat, *A scheme for integrating the nonlinear equations of Mathematical Physics by the method of the inverse scattering problem, I*, Funct. Anal. Appl., **8** (1974), 226-235.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ALBERTA, EDMONTON, CANADA T6G 2G1
 E-mail address: mkovalyo@gpu.srv.ualberta.ca

This page intentionally left blank

Scattering at truncated solitons and inverse scattering on the semiline

H. Steudel

ABSTRACT. Consider a Zakharov–Shabat potential $q(x)$ with one complex eigenvalue in the upper half–plane of the spectral parameter, and with a zero reflection coefficient. We then truncated the potential at $x = 0$, and set the potential equal to zero for $x < 0$. The potential truncated in this way gives rise to a rational reflection coefficient with exactly one pole. However, the location of this pole depends on the location of the soliton center, and may even move into the lower half plane. It is explicitly shown how such a truncated potential can be recovered by solving the Gelfand–Levitan–Marchenko equations. More generally, we show that a truncated N –soliton potential leads to a rational reflection coefficient with exactly N poles. This explains how a recently developed method for solving initial–boundary problems for hyperbolic integrable evolution equations can be understood as an approximation in terms of truncated N –soliton solutions.

1. Introduction

In its classical form – see, e.g., [1, 2] – the inverse scattering method solves the initial value problem on the full x –axis for some evolution equations. However for many physical systems, one does not require the entire interval. Instead one has to solve the initial–boundary value problems on a finite x –interval or on the semiline, and there has been considerable mathematical effort in this direction, see [3, 4] and references therein. However, when restricted to a certain subclass of integrable evolution equations the problem becomes much easier, and a constructive solution method was established in a series of recent papers [5, 6, 7, 8] which are based on some former suggestion by Kaup [9]. In order to characterize this subclass we recall the pioneering paper of Ablowitz et al. [2] which starts from a simultaneous system of differential equations $\phi_x = U(q, r, \zeta)\phi$, $\phi_t = V(q, r, \zeta)\phi$ where U is a matrix function of the typical AKNS–form, see eq.(1) below. The matrix function V may be of a much more general form. With the restricting assumption that it is either a) polynomial in the spectral parameter ζ or b) polynomial in $1/\zeta$ the compatibility of those linear equations leads to the simplest but, nevertheless, most

1991 *Mathematics Subject Classification*. Primary 37K15.

It is a pleasure to thank David J. Kaup for stimulating discussions.

famous integrable evolution equations. The Korteweg–de Vries and the nonlinear Schrödinger equations belong to the subclass a) while the sin–Gordon equation and the equations of second harmonic generation [10] belong to class b). The method of solving initial–boundary value problems that we have in mind here applies to the subclass b) and cannot be applied to the subclass a) of the AKNS class of integrable evolution equations. Here we will not give a review of this method but only stress some new features. By this method [5, 6], the solution in a finite space–time region typically is given in terms of an N –pole solution approximating the exact solution for $N \rightarrow \infty$. It is the motivation of the present paper to achieve a better understanding of these N –pole approximations. We have the result that they are just N –soliton solutions in the usual sense but truncated at $x = 0$ and are physical for $x > 0$ only.

The scattering at a truncated one–soliton potential is considered in detail in section 2. We will see how the motion of the pole depends on the location of the soliton center. A related discussion of scattering at a truncated soliton was given by Lamb [11] but otherwise, apparently, the direct and inverse scattering for truncated potentials has not been discussed elsewhere in the literature. Quite independent of section 2 in section 3 we solve the inverse problem, i.e., we construct a potential whose reflection coefficient $c(\zeta)$ is rational in the spectral parameter ζ with exactly one pole ζ_1 and a prescribed residuum c_1 . In section 4 it is shown that a truncated N –soliton solution gives rise to a rational reflection coefficient. Some conclusion with respect to N –pole approximations for solving initial–boundary problems is drawn in the final section 5.

2. The direct scattering problem

First, we will establish the scattering matrix for a truncated soliton. We do this by taking the solution on the infinite interval and then obtaining the scattering data for the truncated potential. Let us write the AKNS scattering problem in the form

$$(1) \quad \phi_x = (J\zeta + Q(x))\phi \equiv U\phi$$

with

$$(2) \quad J = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}, \quad Q(x) = \begin{pmatrix} 0 & q(x) \\ r(x) & 0 \end{pmatrix}$$

and, more particularly, specify to the Zakharov–Shabat problem $r = \epsilon q^*$, $\epsilon = \pm 1$ where the star denotes the complex conjugate value. In order to treat both types characterized by the sign ϵ on an equal footing it is of advantage to introduce two functions,

$$(3) \quad \begin{aligned} A_\epsilon(y) &= \frac{e^y + \epsilon e^{-y}}{2} = \begin{cases} \cosh y & , \quad \epsilon = +1 \\ \sinh y & , \quad \epsilon = -1 \end{cases} \\ B_\epsilon(y) &= A_\epsilon(y)/A_{-\epsilon}(y) = \begin{cases} \coth y & , \quad \epsilon = +1 \\ \tanh y & , \quad \epsilon = -1 \end{cases} . \end{aligned}$$

Then the well known potential with one complex eigenvalue $\zeta = \zeta_0 \equiv \xi_0 + i\eta_0$, $\eta_0 > 0$ is given as

$$(4) \quad q^{[1]}(x) = 2\eta_0 e^{i(\alpha - 2\xi_0 x)} / A_{-\epsilon}(2\eta_0(x - x_0))$$

Clearly, on the infinite interval the potential is singular for $\epsilon = +1$. Next, the wave function corresponding to (4) may be derived from the Darboux transformation

(see [12, 13]) and can be written explicitly. It is convenient to introduce the matrix function

$$(5) \quad W(\zeta, x) = (\zeta - \xi_0)J - \eta_0 B_\epsilon(2\eta_0(x - x_0))\mathbf{1} + Q^{[1]}(x)/2$$

where $\mathbf{1}$ is the unit matrix. Below we also use the abbreviations

$$(6) \quad E(y) \equiv \begin{pmatrix} \exp(y) & 0 \\ 0 & \exp(-y) \end{pmatrix},$$

$$(7) \quad Z \equiv (\zeta - \xi_0)J - \eta_0 \mathbf{1}.$$

As it can be confirmed by direct verification, the matrix-valued function

$$(8) \quad \Phi^{[1]}(\zeta, x) = W(\zeta, x)E(-i\zeta x)$$

then solves (1) with the potential $Q^{[1]}$ and fulfils

$$(9) \quad \Phi^{[1]} \rightarrow Z(\zeta)E(-i\zeta x) \text{ for } x \rightarrow +\infty.$$

Now, let us consider how to obtain the scattering data for the truncated potential. Generally, for any potential $q(x)$ that vanishes outside of some interval $x_i < x < x_f$ and for $\Phi(x)$ being any non-singular matrix-valued solution the scattering matrix may be defined as

$$(10) \quad S(\zeta) = E(i\zeta x_f)\Phi(\zeta, x_f)\Phi^{-1}(\zeta, x_i)E(-i\zeta x_i).$$

It is independent of the particular solution $\Phi(\zeta, x)$. If we write

$$(11) \quad S(\zeta) \equiv \begin{pmatrix} a & -\bar{b} \\ b & \bar{a} \end{pmatrix}.$$

our notation is in agreement with that of [2] though the S -matrix has not been used explicitly there.

Now we introduce the *truncated soliton* as

$$(12) \quad q_{tru}^{[1]}(x) = \theta(x)q^{[1]}(x)$$

where $\theta(x)$ denotes the Heaviside step function. Clearly $q_{tru}^{[1]}$ is nonsingular for $\epsilon = -1$ and also nonsingular for $\epsilon = 1, x_0 < 0$. For this potential, from (8,9,10) we obtain the scattering matrix for the truncated potential. It is

$$(13) \quad S_{tru}^{[1]}(\zeta) = \left[E(i\zeta x)\Phi^{[1]}(\zeta, x) \right]_{x \rightarrow \infty} \left[\Phi^{[1]}(\zeta, 0) \right]^{-1} = ZW^{-1}(\zeta, 0).$$

For definiteness we take the limit for real ζ and afterwards continue $S_{tru}^{[1]}$ to the complex plane.

The reflection coefficient for this truncated potential is found to be

$$(14) \quad c_{tru}^{[1]}(\zeta) \equiv \left(\frac{\bar{b}(\zeta)}{a(\zeta)} \right)_{tru}^{[1]} = \frac{i\epsilon\eta_0 e^{i\alpha}}{(\zeta - \xi_0)A_{-\epsilon}(2\eta_0 x_0) - i\eta_0 A_\epsilon(2\eta_0 x_0)}.$$

Obviously there is a pole at

$$(15) \quad \zeta_1 \equiv \xi_1 + i\eta_1$$

$$(16) \quad \xi_1 = \xi_0, \eta_1 = \eta_0 B_\epsilon(2\eta_0 x_0)$$

with its residuum being

$$(17) \quad c_1 = \text{Res}(\zeta_1) = \frac{i\epsilon\eta_0 e^{i\alpha}}{A_{-\epsilon}(2\eta_0 x_0)} = -\frac{i}{2}q^{[1]}(0).$$

Note that – up to a constant factor – the residuum is just the value of the potential at the truncation point. From (16) and (17) with the definitions (3) one obtains the connection

$$(18) \quad \eta_1^2 - \eta_0^2 = \epsilon |c_1|^2 .$$

The scattering matrix (13) now can be written explicitly in the form

$$(19) \quad S_{tru}^{[1]}(\zeta) = \begin{pmatrix} \frac{\zeta - \zeta_1}{\zeta - \zeta_0^*} & \frac{-c_1}{\zeta - \zeta_0^*} \\ \frac{-\epsilon c_1^*}{\zeta - \zeta_0} & \frac{\zeta - \zeta_1}{\zeta - \zeta_0} \end{pmatrix} .$$

A complete soliton is characterized by the phase α , the frequency ξ_0 , the location x_0 of its center and the inverse width η_0 . The truncated soliton – truncated at $x = 0$ and taken for $x > 0$ – is characterized by the same parameters. From (16,17) we get a biunique mapping between the soliton parameters and the scattering data for the truncated soliton,

$$(20) \quad (\zeta_0 = \xi_0 + i\eta_0, \eta_0 > 0, x_0) \longleftrightarrow (c_1, \zeta_1 = \xi_0 + i\eta_1, \eta_1^2 - \epsilon |c_1|^2 > 0) .$$

Thus the parameters of the truncated soliton are also characterized just as well by (ζ_1, c_1, α) .

Summarizing we state:

For $x_0 \rightarrow +\infty$, $q(x)$ approaches an untruncated soliton while c_1 goes to zero. When x_0 goes from $+\infty$ to $-\infty$, the eigenvalue $\zeta_1 = \xi_1 + i\eta_1$ of the truncated soliton is moving from the pole ζ_0 of the untruncated soliton to the complex conjugated value ζ_0^* . For $\epsilon = -1$ this motion is downward, crossing the real axis for $x_0 = 0$. On the other hand, for $\epsilon = +1$ the motion goes upward to $\xi_0 + i\infty$ and then from $\xi_0 - i\infty$ to ζ_0^* . Earlier, see [7], solutions corresponding to poles in the lower ζ -plane were called *virtual solitons*. From the present considerations we see that for both cases of $\epsilon = \pm 1$ a virtual soliton corresponds to a soliton with its center being outside of the *physical domain* $x > 0$. For $\epsilon = +1$ such a conclusion is contained already in [7].

3. The inverse problem

In the above, we constructed the scattering data for a truncated 1-soliton potential. Now, we will address the inverse problem. Given the scattering data for the above truncated 1-soliton potential, can we reconstruct the potential? Specifically we will now consider the problem: Can we construct a potential $Q(x)$ in the scattering problem (1), such that the reflection coefficient, $c(\zeta)$, is analytic in the entire ζ -plane, except for one pole at ζ_1 , and with its residue being c_1 ? Let us try to take ζ_1 and c_1 as arbitrary complex numbers. Thus we write

$$(21) \quad c(\zeta) = \frac{c_1}{\zeta - \zeta_1} ,$$

and we define the functions

$$(22) \quad G(z) = \frac{1}{2\pi} \int_c c(\zeta) e^{-i\zeta z} d\zeta = -i\theta(z) c_1 e^{-i\zeta_1 z} ,$$

$$(23) \quad \bar{G}(z) = -\epsilon G^*(z)$$

which appear as kernels in the Gelfand–Levitan–Marchenko equations (4.39a,b) from [2]. For inversion about $x = -\infty$, these are

$$\begin{aligned}
 \bar{L}_1(x, y) + G(x + y) - \int_{-\infty}^x L_1(x, s)G(s + y)ds &= 0 \\
 (24) \quad L_1(x, y) + \int_{-\infty}^x \bar{L}_1(x, s)\bar{G}(s + y)ds &= 0.
 \end{aligned}$$

Restricting the above to our present example at first we see that for $x < 0, y < -x$ both solution functions $L(x, y), \bar{L}(x, y)$ vanish identically so that we get the potential

$$(25) \quad q_1(x) = 2\bar{L}(x, x) \equiv 0, \quad x < 0.$$

Now we assume $x > 0, y > -x$ and find that then the system (24) may be rewritten as

$$\begin{aligned}
 e^{i\zeta_1 y} \bar{L}_1(x, y) + ic_1 \left[-e^{-i\zeta_1 x} + \int_{-y}^x L_1(x, s)e^{-i\zeta_1 s} ds \right] &= 0 \\
 (26) \quad e^{-i\zeta_1^* y} L_1(x, y) + i\epsilon c_1^* \int_{-y}^x \bar{L}_1(x, s)e^{i\zeta_1^* s} ds &= 0.
 \end{aligned}$$

It has to be solved with the initial conditions

$$\begin{aligned}
 \bar{L}_1(x, -x) &= ic_1 \\
 (27) \quad L_1(x, -x) &= 0.
 \end{aligned}$$

We introduce the notation

$$(28) \quad \mathbf{L}(x, y) = (\bar{L}_1(x, y), L_1(x, -y))^T,$$

and take $\zeta_1 = \xi_1 + i\eta_1$. Then by differentiation of (26) with respect to y we arrive at a homogeneous system of ordinary differential equations,

$$(29) \quad (i\partial_y - \xi_1)\mathbf{L}_y = M\mathbf{L},$$

with M denoting the constant matrix

$$(30) \quad M = \begin{pmatrix} i\eta_1 & c_1 \\ \epsilon c_1^* & -i\eta_1 \end{pmatrix}.$$

Its general solution is easily found as

$$(31) \quad \mathbf{L}(x, y) = e^{-i\xi_1 y} \left(\mathbf{1} \cosh \eta_0 y - \frac{i}{\eta_0} M \sinh \eta_0 y \right) \mathbf{L}_0(x).$$

with

$$(32) \quad \eta_0 = \sqrt{\eta_1^2 - \epsilon|c_1|^2}$$

\mathbf{L}_0 is determined by the initial conditions (27). For $\epsilon = +1$ now we introduce the constraint $|c_1|^2 < \eta_1^2$ in order to get real η_0 . Finally we find the potential to be

$$\begin{aligned}
 (33) \quad q_1(x) &= 2\bar{L}(x, x) \\
 &= \frac{-2ic_1 \exp(-2i\xi_1 x)}{\cosh(2\eta_0 x) - (\eta_1/\eta_0) \sinh(2\eta_0 x)}, \quad x > 0.
 \end{aligned}$$

One may readily verify that the potentials (33) and (4) coincide for $x > 0$, upon using (16,17) and the *addition theorem*

$$(34) \quad A_\epsilon(x - y) = A_\epsilon(x) \cosh y - A_{-\epsilon}(x) \sinh y .$$

Thus we have recovered exactly the truncated solitons of section 1. Excluding singular potentials from our present discussion we state that nonsingular potentials are found

- i) for $\epsilon = -1$, $c_1 \neq 0$, η_1 arbitrary and
- ii) for $\epsilon = +1$, $c_1 \neq 0$, $\eta_1 < -|c_1|$.

Alternatively, the inversion can also be done just as well and in a straightforward and explicit manner by inverting about $x = +\infty$, upon using those GLM equations given in [2]. We will not give the details here but only point out the essential qualitative features. What now is required as the *reflection coefficient* is b/a , i.e., the quotient of the (12)– and the (11)–elements of the S –matrix (19). This gives

$$(35) \quad \frac{b(\zeta)}{a(\zeta)} = \frac{-\epsilon c_1^* \zeta - \zeta_0^*}{\zeta - \zeta_1 \zeta - \zeta_0} .$$

In stark contrast to \bar{b}/a , given in (14), the singularities of b and of a do not cancel. $\zeta = \zeta_0$ is a singularity of b/a , but it is not a zero of $a(\zeta)$ and, therefore, is not an eigenvalue of the potential. Thus it and its residue must not be used in constructing the kernel F . It is not difficult to rediscover $q_{tru}^{[1]}(x)$ for $x > 0$. A bit more effort is required to prove that the potential vanishes for $x < 0$.

Note that the manner of construction of a truncated soliton from its scattering data differs essentially from the construction of a genuine (untruncated) soliton, where the reflection coefficient is zero everywhere.

4. Scattering at a truncated N –soliton state

Now we will draw some conclusions for N –soliton states as they may be generated from the vacuum solution by successive application of Darboux transformations. According to the polynomial method of Neugebauer and Meinel [12] the resulting wave function $\Phi^{[N]}$ may be written in the form

$$(36) \quad \Phi^{[N]}(\zeta, x) = P_N(\zeta, x)\Phi(\zeta, x)$$

where P_N is a polynomial of N –th order in ζ . By Darboux transformation the vacuum potential $q \equiv 0$ is mapped to a potential which asymptotically for $x \rightarrow \infty$ approaches vacuum. The Darboux transformation (36) works for any vacuum solution Φ with the same P_N . In particular we may choose for Φ the diagonal matrix $E(-i\zeta x)$ with E defined by (6). The general matrix wave function fulfilling (1) with vacuum potential then is $E(-i\zeta x)$ multiplied from the right by a constant matrix C . Thus we get a relation

$$(37) \quad [P_N]_{x \rightarrow \infty} = E(-i\zeta x)CE(i\zeta x) .$$

The off–diagonal elements of the matrix at the righthand side of this equation are $C_{12} \exp(2i\zeta x)$ and $C_{21} \exp(-2i\zeta x)$. Since P_N is assumed to be a polynomial in ζ , it follows $C_{12} = C_{21} = 0$ or, in other words, P_N asymptotically becomes diagonal. In order to get the S –Matrix for an N –soliton potential which is truncated both at $x = x_i$ and at $x = x_f > x_i$ we have to substitute (36) into (10). Then we

specify $x_i = 0$, take the limit $x_f \rightarrow \infty$ and take into account that the S -matrix for the seed solution is the unity matrix. In this way we arrive at

$$(38) \quad S^{[N]}(\zeta) = [P_N(\zeta, x)]_{x \rightarrow +\infty} P_N^{-1}(\zeta, 0).$$

Because the first factor at the right-hand side is diagonal it does not affect the reflection coefficient and we get

$$(39) \quad c_{tru}^{[N]}(\zeta) = [P_N(\zeta, 0)]_{12} / [P_N(\zeta, 0)]_{22}.$$

Thus the reflection coefficient $c_{tru}^{[N]}$ is rational in ζ . More precisely, due to [12], we may state that it is a polynomial of order $N - 1$ over an N -th order polynomial so that it is vanishing of order $1/\zeta$ at infinity.

5. Conclusion – Connection to inverse scattering on a semiline

Treating a soliton truncated from the left at $x = 0$, we have found the rational reflection coefficient $c(\zeta) = c_1/(\zeta - \zeta_1)$ and have shown how the potential may be recovered by inversion about $x = -\infty$.

For N -soliton states truncated at $x = 0$ and taken to be nonzero only at the semiline $0 < x < \infty$ the S -matrix elements all become rational functions of ζ , and the reflection coefficient $c(\zeta)$ is regular at infinity.

In the papers [5, 6, 7, 8] quoted in the introduction a method for solving initial-boundary problems in a finite or semi-infinite interval has been developed. In short, the temporal dependence of an *effective* scattering matrix is determined by taking into account only the known boundary values at the left-hand end of the considered interval, say at $x = 0$. The corresponding *effective* reflection coefficient then typically has an infinity of moving poles in the complex ζ -plane. For a finite number N of such poles together with their respective residues then the Gelfand–Levitan–Marchenko equations can be solved explicitly. In general, there is no guarantee that this procedure converges with $N \rightarrow \infty$. However, for some problems with given model initial data, we were able to show that it works. In particular we investigated second harmonic generation or – with another physical interpretation – resonant two-photon propagation in the low-excitation limit [6, 7], and there we were able to demonstrate that with increasing N indeed the numerical solution is approached.

In the light of our present consideration we get a very natural interpretation of those N -pole approximation. It means that at any fixed time, the solution of the considered initial-boundary problem is approximated by states of N truncated solitons with increasing N .

References

- [1] S. P. Novikov, S. V. Manakov, L. P. Pitaevski and V. E. Zakharov, *Theory of solitons*, Consultants Bureau, New York 1984.
- [2] M. J. Ablowitz, D. J. Kaup, A. C. Newell and H. Segur, *Stud. Appl. Math.* **53** (1974) 249.
- [3] A. S. Fokas, *J. Math. Phys.* **41** (2000) 4188.
- [4] A. Boutet de Monvel and V. Kotlyarov, *Inverse Problems* **16** (2000) 1813.
- [5] H. Steudel and D. J. Kaup, *J. Phys. A: Math. Gen.* **32** (1999) 6219.
- [6] H. Steudel and D. J. Kaup, *J. Phys. A: Math. Gen.* **33** (2000) 1445.
- [7] D. J. Kaup and H. Steudel, *Inverse Problems* **17** (2001) 959.

- [8] D. J. Kaup and H. Steudel, *Recent Results on Second Harmonic Generation*, (Submitted to the Proceedings of the Nov. 2000 AMS sectional meeting, Birmingham, Alabama. To appear as an AMS Contemporary Mathematics Series.)
- [9] D. J. Kaup, *Physica* **6D**(1983) 143 .
- [10] D. J. Kaup, *Stud. Appl. Math.* **59** (1978) 25.
- [11] G. L. Lamb Jr., *Elements of Soliton Theory*, section 2.11, J. Wiley & Sons, New York 1980.
- [12] G. Neugebauer and R. Meinel, *Phys. Lett.* **100A** (1984) 467.
- [13] V. B. Matveev and M. A. Salle, *Darboux Transformations and Solitons*, Springer-Verlag, Berlin 1991.

INSTITUT FÜR PHYSIK DER HUMBOLDT-UNIVERSITÄT, INVALIDENSTRASSE 110, 10115 BERLIN,
GERMANY

E-mail address: `steudel@physik.hu-berlin.de`

This volume contains new developments and state-of-the-art research arising from the conference on the “Legacy of the Inverse Scattering Transform” held at Mount Holyoke College (South Hadley, MA). Unique to this volume is the opening section, “Reviews”. This part of the book provides reviews of major research results in the inverse scattering transform (IST), on the application of IST to classical problems in differential geometry, on algebraic and analytic aspects of soliton-type equations, on a new method for studying boundary value problems for integrable partial differential equations (PDEs) in two dimensions, on chaos in PDEs, on advances in multi-soliton complexes, and on a unified approach to integrable systems via Painlevé analysis.

This conference provided a forum for general exposition and discussion of recent developments in nonlinear waves and related areas with potential applications to other fields. The book will be of interest to graduate students and researchers interested in mathematics, physics, and engineering.

ISBN 0-8218-3161-5



9 780821 831618

CONM/301

AMS on the Web
www.ams.org