

REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-09-0139

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for review the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson L. Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE Feb. 27, 2008	3. REPORT TYPE AND DATES COVERED Final Report; 1 June, 2006 - 30 Nov., 2008	
4. TITLE AND SUBTITLE Theoretical Study of the Saturated Stage of a Relativistic magnetron			5. FUNDING NUMBERS FA9550-06-C-0030	
6. AUTHOR(S) D.J. Kaup				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) DJ Kaup Inc. 903 Oak Forest Dr. Winter Springs, FL 32708-4004			8. PERFORMING ORGANIZATION REPORT NUMBER 2009-02	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) AF Office of Scientific Research 875 North Randolph St., Rm 3112 Arlington, VA 22203 DCMA Orlando 3555 Maguire Blvd Orlando, FL 32803-3726			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution unlimited.				12b. DISTRIBUTION CODE Distribution A
13. ABSTRACT (Maximum 200 Word) The saturated stage of a magnetron or a cross-field amplifier is when the device has powered up and is steadily delivering power. Theoretical studies of this stage had suggested that when the dc current would become sufficiently large in comparison with the rf current, then no steady state solution could be expected to exist. Attempts were made to develop a numerical code which would allow one to numerically solve for a stationary solution of the saturated stage, if it existed. However the high order of the ordinary differential equations and their singular nature had prevented any solutions from being found. It has been suggested that a hybrid method involving single particle orbits could be used. Here one would assume that a stationary solution did exist and then one would solve for the particle orbits in that stationary solution. From those orbits, one then could obtain the corresponding fluid quantities by averaging and thereby bypassing any need to solve the high order, singular differential equations. Other work completed under this contract included studies of integrable systems, particularly those integrable optical systems which govern intense laser beams and stability, and also Bose-Einstein condensates.				
14. SUBJECT TERMS				15. NUMBER OF PAGES
				16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT None	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

20090429211

To: technicalreports@afosr.af.mil

Subject: Final Report

Contract/Grant Title: Theoretical Study of the Saturated Stage of a Relativistic Magnetron

Contract/Grant #: FA9550-06-C-0030

Reporting Period: June 1, 2006 to Nov. 30, 2008

Original Objectives:

A) We shall continue our studies of the nonrelativistic, electrostatic, planar magnetron #T266 in the saturation stage. We shall (i) obtain numerical results for the operating characteristics; and (ii) compare these with the known experimental characteristics.

B) We shall extend the above studies to the saturation stage of relativistic, electromagnetic, cylindrical crossed-field devices, particularly the A6 magnetron. We shall (i) obtain results for the operating characteristics; and (ii) seek to detail and understand how each of the relativistic, and electromagnetic, and cylindrical effects affect the resulting density profiles and other operating characteristics in this stage.

C) We shall continue other studies that have potential interest to the AFOSR, such as in the areas of nonlinear optics, instabilities in higher order systems, the development of a quantitative measurement of variational results, and precursor formation in the ionosphere.

Personnel involved:

Dr. Heinz Steudel (Humboldt University, Berlin, Germany; consultant)

Dr. V. Gerjikov (Sofia, Bulgaria; consultant)

Mr. Galen T. Kaup (Research Assistant, mathematician)

Prof. Subash Antani (Edgewood College, Madison, Wisc., consultant, nonlinear interactions in the ionosphere.)

Summary of Accomplishments:

- 1) Saturated stage of magnetrons and CFAs: The main thrust of this contract was to theoretically study the various stages of high power magnetrons (HPM) and cross-field amplifiers (CFA) by means of the cold-fluid plasma equations. The work accomplished on this project has been published in Pub01 and Pub03 listed below, which are a conference proceedings followed by publication in a standard journal. Crossed-field devices have basically two main stages: first there is the initiation stage wherein the power is turned on and the energy output of the device ramps up from zero towards its maximum output. Second is the saturation stage wherein the device has achieved the maximum power output. The growth in the initiation stage is due to the excitation of linear instabilities. The instabilities must vanish when the saturation stage is achieved, which is a stationary state. When the above theoretical results were used in the design of a Fortran code to calculate this saturated stage, it was found that regardless of the approach used (WKB expansions, direct numerical integrations, etc.), unwanted

instabilities invariably occurred. First, a review of the theory in these publications revealed that the theory was defective, in that it attempted to solve the problem of the saturated stage in an unstable manner. The approach that was used was exactly the same approach which had worked successfully for solving the initiation stage: separating the problem into an rf part and a dc part, then numerically iterating back and forth between them until a solution was achieved. But what happens when one reaches the saturation stage is that the interaction between the rf wave and the dc background becomes so strong that one cannot treat these two modes separately. Rather, one must solve the coupled system as a single unit because small changes in one mode can generate larger changes in the other. The difficulty here is that the coupled system then becomes equivalent to a 15th order set of ordinary differential equations (ODEs) which are *singular ODEs*. They are singular because some of the highest derivatives in these ODEs are multiplied by small quantities. Whence any error in the evaluation of these derivatives can become magnified as one integrates forward. This set also has various boundary conditions to be set at the cathode and also at the anode. Whence this problem then becomes a two-point boundary problem. In short, this is the worst of all possible cases to attempt to solve. And all attempts by this PI to numerically integrate this 15th order system have not been successful. Let us note that in the initiation stage, one has only a fourth-order system of ODEs for the rf mode, another second-order system of ODEs for the dc mode and none of these ODEs are singular. This reduction in the order occurs because one has assumed that certain algebraic relations are satisfied between the physical quantities. These initiation ODEs can be successfully numerically integrated mainly due to being only 6th order. However the jump from 6th order to 15th order requires a much more careful formulation of the numerical approach in order to succeed. We note that at this point, we have not been able to analyze whether the instabilities involved are numerical instabilities or actual physical instabilities coming from the cold-fluid equations. If one is not to be able to numerically integrate these ODEs, then it is necessary to search for another approach. There are variational methods that could possibly work. There are methods that use single particle orbits. The same is done in PIC codes, but there one only solves for the particle orbits and averages appropriately to get the macroscopic quantities such as density and average velocities. Our approach has been to solve for that stationary state wherein the device is delivering maximum power, if such a state exists. Near the end of the contract, it was realized that there is another approach which would be somewhere between the sole use of PIC codes and fully solving the appropriate ODEs. This would be a hybrid method, whereby one would calculate the single particle orbits in an assumed stationary dc electric field and an assumed stationary rf electromagnetic field. From a knowledge of the single particle orbits in these fields, one can then obtain the macroscopic density and the average particle velocity, if such a stationary state could exist. If it could exist, then linear stability theory could be used to study its stability features.

- 2) The Camassa-Holm Equation: Certain problems can have instabilities which are related to the presence of eigensolutions of the perturbed equations whose analytical properties are not spatially uniform. This shows up in integrable systems in that the eigenvalue problem of the Lax pair will have this nature. One example of this is the Camassa-Holm equation (CH) when the initial data is sufficiently strong. Other examples are certain degenerate parametric optical

laser problems such as degenerate two-photon propagation. A very good reason to study the CH equation first, instead of the optical problems, is that the CH equation has the advantage it lacks many of the complexities found in these optical problems. Thus its results can therefore be potentially used as a model for understanding how to treat these optical problems. In Pub02, we carry out an analysis of this feature when the initial data for the CH equation is sufficiently strong and is on an infinite spatial interval. We show that one can solve this problem if the infinite interval is broken into semi-infinite and finite regions, inside of which, each region contains its own collection of scattering data. Given the scattering data in these regions, we then demonstrate that one can, in principle, reconstruct the full solution on the infinite interval.

- 3) In Pub04, we have shown how one could solve the problem of the estimation of the accuracy of the solution of a variational problem. Variational methods have often been successful in solving complicated problems. And the method is usually more accurate than one can justify. We note that a variational formalism does exist for the HPM problem [D. J. Kaup and Gary E. Thomas, *J. Plasma Physics* **57**, 765-84 (1997)]. Consequently even if all else would fail, one would still have the option to study the HPM problem variationally. If so, then one would like to have some means for estimating the accuracy of such a solution. The results of this publication show how one can always obtain estimates of the errors in a variational solution and their associated quantities, and without having to know anything about the actual solution.
- 4) Pub05 and Pub07 are the result of joint work between Prof. Gerdjikov and myself on eigenvalue problems. It is cast into the form of studies of integrable equations, but the problem is really one of a study of direct scattering and inverse scattering in one dimension, which are also exactly of the same nature as the problem of solving the perturbed cold-fluid equations in plasma physics. Such studies give the investigators experience in understanding these problems and experience in devising methods of solution for such problems.
- 5) In integrable systems, one has two classes of solutions: solitons (bound states) and radiation (the continuous spectrum). Given the scattering data, in order to reconstruct the potentials, one must solve a set of linear integral equations. When the continuous spectrum is absent, then these linear integral equations can be reduced to a set of only algebraic equations. For certain integrable systems, and on finite or semi-infinite intervals, one can invariably convert any continuous spectrum into bound states, which do generate soliton-like solutions which are called "virtual solitons". Integrable systems of this nature are mainly optical parametric equations which are posed on finite intervals, such as second harmonic generation and related hyperbolic equations. In Pub06, Prof. Steudel and I have demonstrated how to do this and how it is related to earlier work in NMR (nuclear magnetic resonance). What we demonstrate here is that the scattering data for such problems can be reduced to a countable set of data. In the language of integrable equations, we show that for any reasonable initial value problem of such a system, the continuous spectrum can be converted into a countable set of bound states. Thus the inverse scattering problem can then be reduced to an algebraic problem, which is the same problem as constructing an N-soliton solution, where N is countable. In practical terms, it is further demonstrated that typically one can take N to be finite, albeit large.
- 6) Pub08 is joint work with J. Yang and is a solution of the perturbation problem of

the Sasa-Satsuma equation, which is integrable and has a 3×3 eigenvalue problem. This appears to be the first time that anyone has explicitly calculated the squared eigenfunctions for anything beyond the AKNS problem. This result has a breath of complexity which goes well above that of the AKNS and the Schrodinger eigenvalue problems. It also contains certain universal features which one can see will continue into higher order problems. It has just this month appeared in the Journal of Mathematical Physics.

- 7) Pub09 is another publication in collaboration with Prof. Gerdjikov. In this case we have presented a one-dimensional, integrable system which closely models other proposed models for a one-dimensional Bose-Einstein condensate. Its basic structure is a multi-component, nonlinear Schrodinger equation, and has some freedom in assigning values to the coefficients of the nonlinear terms. This is recent work and was first presented in the Summer of 2008.
- 8) Pub10 is work in progress and has been supported by this contract. At the present time, no variational principal has been developed for two-level atoms and the phenomenon of self-induced transparency (SIT). Although two-level SIT is an integrable system, three-level SIT is not integrable except for special level configurations and then not unless the coupling coefficients are certain special values. The advantage of having a variational principle for three-level atoms is obvious. We could execute certain studies of the three-level non-integrable cases.
- 9) Pub11 is also work in progress by an undergraduate student. This work will complement Pub08 and will allow one to study perturbations of the integrable three-wave resonant interactions.
- 10) Pub12 is work in progress and is concerned with ionospheric interactions when the heater wave is tuned to the second harmonic of the electron cyclotron frequency. Experiments at this frequency have recently been conducted at the HAARP transmitter. In our earlier ionospheric work of 1992, this frequency range was inadequately treated since at that time, there was no experimental data for that regime. This work is intended to fill that gap and to compare theory with the experiments for that frequency.

Archival publications (published) during the period of this contract:

Pub01) *Resonances in High Density Nonneutral Plasmas*, D. J. Kaup, "Frontiers of Nonlinear Physics, Proceedings of the Second International Conference", Nizhny Novgorod - St. Petersburg, Russia, 2004, edited by Alexander Litvak (Nizhny Novgorod, Russia, Institute of Applied Physics RAS, 2005), pp. 308-20.

Pub02) *Evolution of the Scattering Coefficients of the Camassa-Holm Equation, for General Initial Data*, D.J. Kaup, *Studies in Appl. Math.* **117**, 149-64 (2006).

Pub03) *Drift Resonance in High Density Non-neutral Plasmas*, D.J. Kaup, *Phys. Plasmas* **13**, 053113 (2006).

Pub04) *Quantitative Measurement of Variational Approximations*, D.J. Kaup and T.K. Vogel, *Phys. Lett. A* **362**, 289-97 (2007).

Pub05) *How many types of soliton solutions do we know?*, Vladimir S. Gerdjikov and D. J. Kaup, "Geometry, Integrability and Quantization", pp. 11-34, Ivailo M. Mladenov and

Manuel De Leon, Editors, (Softex, Sofia, Bulgaria, 2006).

Pub06) *Inverse Scattering for an AKNS Problem with Rational Reflection Coefficients*, H. Steudel & D.J. Kaup, *Inverse Problems* **24**, 025015 (2008).

Pub07) *On classification of soliton solutions of multicomponent nonlinear evolution equations*, V. S. Gerdjikov, D. J. Kaup, N. A. Kostov & T. I. Valchev, *J. Phys. A: Math. Theor.* **41**, 315213 (2008).

Pub08) *Derivation of Squared Eigenfunctions for the Sasa-Satsuma Equation*, Jianke Yang & D. J. Kaup, *J. Math. Phys.* **50**, 023504 (2009).

Archival publications to be completed, whose work was started during the period of this contract:

Pub09) *Bose-Einstein condensates and multi-component NLS models on symmetric spaces of **BD.I**-type. Expansions over squared solutions*, V. S. Gerdjikov, D. J. Kaup, N. A. Kostov and T. I. Valchev, [To appear in the proceedings of the Conference on Nonlinear Science and Complexity, Porto, Portugal, July 28-31 (2008)].

Pub10) *Variational Principle for Self-Induced Transparency*, D.J. Kaup and Galen T. Kaup (work in progress).

Pub11) *Perturbations and Squared Eigenfunctions for the Three-Wave Resonant Interaction*, R. A. Van Gorder and D. J. Kaup (work in progress).

Pub12) *Upper hybrid resonance phenomena at the second cyclotron harmonic*, Subhash Antani & DJ Kaup (work in progress).

Inverse scattering for an AKNS problem with rational reflection coefficients

H Steudel¹ and D J Kaup²

¹ Institut für Physik der Humboldt-Universität Newtonstraße 15, 12489 Berlin, Germany

² Department of Mathematics and Institute for Training and Simulation, University of Central Florida, Orlando, FL 32816-1364, USA

E-mail: steudel@physik.hu-berlin.de and kaup@mail.ucf.edu

Received 29 August 2007, in final form 27 December 2007

Published 21 February 2008

Online at stacks.iop.org/IP/24/025015

Abstract

We study the AKNS (Ablowitz–Kaup–Newell–Segur) inverse scattering problem for rational reflection coefficients on a semi-infinite interval. We demonstrate that the Marchenko integral equations for the AKNS version on such an interval can be solved in a direct and straightforward way by algebraic methods for any set of rational reflection coefficients, vanishing at infinity. The general AKNS scattering problem for this case, as well as the usual symmetry reductions, are discussed. The connection to an alternative procedure by Rourke and Morris (1992 *Phys. Rev. A* **46** 3631) is pointed out. Our procedure is built around a constant matrix, M , which can be constructed from the poles and residues of the rational reflection coefficients. Under certain conditions, which we define as *minimal symmetry* and which then suitably constrains the distribution of eigenvalues, we show that it is always possible to represent the resulting potentials as truncated N -soliton potentials. This procedure is of interest for solving initial-boundary value problems of integrable hyperbolic systems by the inverse scattering transform (IST) applied to a semi-infinite or finite interval.

1. Introduction

Upon the development of the inverse scattering transform for solving *initial value problems* for integrable nonlinear evolution equations, it was natural that efforts would soon be made toward applying this method to *initial-boundary value problems* as well. This was first done by Ablowitz and Segur [1] who developed an inverse scattering transform for a special initial-boundary value problem on the semi-infinite interval for the nonlinear Schrödinger equation. The more general initial-boundary problem for the nonlinear Schrödinger equation was soon realized to be more complex, in that Riemann–Hilbert techniques were found to be necessary for their solution [2–8].

On the other hand, for hyperbolic integrable systems where one may use characteristic coordinates, the above complexities are generally absent. In earlier papers [9–12], an inverse scattering transform for solving hyperbolic initial-boundary problems on a finite interval was developed. The procedure consists of three steps:

- (1) Solve the direct scattering problem for the initial values at $t = 0$ on the interval $0 < x < x_f$.
- (2) Determine the *effective* S -matrix at $t = t_f > 0$ by the use of boundary data at the left end, $x = 0$, only.
- (3) Reconstruct the potentials by solving the inverse scattering problem at $t = t_f$.

In solving these problems [9–11] it was observed that the effective reflection coefficients could always be replaced by an equivalent set of poles and residues. Typically however, there are usually an infinite number of these poles and residues for which there is at least one accumulation point. However by selecting a finite number N of these poles from the infinite set and, with increasing N , one could obtain increasing accurate approximations of the solution. Whence the reconstruction problem for these systems could then be reduced to the reconstruction problem for rational reflection coefficients which then becomes an algebraic problem. In this manner one could bypass the need to solve the Marchenko *integral* equations [13], replacing them instead with a set of equivalent *algebraic* equations. For second harmonic generation and for degenerate two-photon propagation in the low-excitation limit, it has been demonstrated that the procedure developed by Steudel and Kaup (S&K) did indeed work [10, 11].

Related to this but solved by a different approach is the earlier work of Rourke and Morris (R&M) [14] whereby they solved a reconstruction problem which arose in nuclear magnetic resonance. Here they were given the reflection coefficients for a *half-pulse* which vanished for $\tau > 0$ (with τ corresponding to our variable x). From this information they needed to reconstruct the potentials. To solve this problem, they cleverly added another *half-pulse*, but vanishing for $\tau < 0$, which was crafted such that the composite pulse would be reflectionless. This new composite pulse, now a pure multi-soliton solution and extending over the full τ -axis, could then be reconstructed by the algebraic soliton-lattice method due to Calogero and Degasperis [15]. Note that the physical solution in this case was only for $\tau < 0$. Thus by arbitrarily adjusting the shape for $\tau > 0$, in the unphysical regime, they were able to convert the desired solution into the left half of a pure N -soliton pulse and therefore could construct the solution entirely by algebraic means, thereby also bypassing the need to solve the Marchenko integral equations. Thus again, the reflection coefficient on a semi-infinite interval became equivalent to rational reflection coefficients.

The most difficult aspect of the reconstruction problem on the infinite interval is always the presence of a continuous spectrum, in which case one cannot in general reconstruct the potentials without solving the Marchenko *integral* equations [13]. However in the above two cases, we see that when the interval for the potential becomes semi-infinite or finite, interestingly the basic nature of the reconstruction problem changes dramatically, reducing to only an *algebraic* problem instead of the problem of solving a set of *integral* equations.

As we shall see, on a semi-infinite interval or a finite interval, the usual relationship between the structure of the potentials and the scattering data will become different from that which is found on the infinite interval. In the latter case [16], each soliton corresponds to a pair of zeros of a and $\bar{a}$, which for potentials on compact support, also are the poles of the reflection coefficients. On the other hand, the radiation, which is the linear-like part of the solution, has a Fourier transform-like correspondence with the reflection coefficients for real values of the spectral parameter. However on a semi-infinite or a finite interval, these

relations no longer need hold. First, as in the above two examples, the continuous spectra can be replaced by a countable infinity of regular solitons and/or virtual solitons [11] which are represented by poles of the reflection coefficients. Second, the location of these poles in the complex spectral plane is generally *not* located where the soliton's eigenvalue is located in the infinite interval case. These are just two of the unique differences that distinguish the inverse scattering transform (IST) on the finite/semi-infinite interval from the well-known infinite interval case. These differences will be pointed out as we develop this work.

Since we have two distinctly different approaches being used to solve the same basic problem, it then becomes appropriate to compare and discuss the similarities, differences and limitations in each of these two approaches. In the process, we shall also show how one can reformulate this reconstruction problem into a set of ordinary differential equations (ODEs) with constant coefficients, from which all the above features can be obtained.

Section 2 consists of four parts. In subsection 2.1, the method of solution for the general AKNS (Ablowitz–Kaup–Newell–Segur) problem where r is unrelated to q is developed for the case when the reflection coefficients are rational. We show that our previous results [9–12] on the Marchenko equations can be replaced by a system of ordinary differential equations with constant coefficients, (23), whose coefficient matrix M is built from the poles and the residues of the given reflection coefficients. We show that with the definition of two sets of $(N + \bar{N})$ -vectors, the two resulting ODE systems become identical, but differ in their boundary values. With the use of the matrix exponential function an exact solution is obtained for these systems and for the reconstruction problem. A condensed summary of this new method of solution is then given in subsection 2.2. In subsection 2.3 we derive a system of partial differential equations (PDEs) fulfilled by the auxiliary functions introduced in subsection 2.1. In subsection 2.4, we give a reformulation of the characteristic equation of the matrix M . The resulting equation (43) gives us a bridge to connect our procedure to that which was developed by R&M [14]. Section 3 deals with symmetry reductions. In sections 4 and 5, the general AKNS system is restricted to what we call *minimal symmetry*. In section 4 we show how to compute all scattering data from the rational reflection coefficients. In section 5 we show—quite parallel to [14]—how a potential with rational reflection coefficients can be supplemented to become a complete N -soliton potential. (By ‘complete’, we mean that on the interval(s) on which the potential(s) vanish, supplemental potentials have been added such that the total potential on the full infinite interval has become a pure N -soliton solution.) Several examples are given in section ???. In particular we demonstrate there by an explicit example that when r is unrelated to q and where minimal symmetry is not present, the R&M method [14] does not work. Summary and discussions are presented in section 6.

2. The general AKNS system

2.1. The method of solution

From this point on, we shall only consider the AKNS scattering problem [16, 26],

$$\partial_x u = \begin{pmatrix} -i\zeta & q(x) \\ r(x) & i\zeta \end{pmatrix} u, \quad (1)$$

under the conditions that $r(x) = 0 = q(x)$ when $x < 0$. We will solve the reconstruction problem under the assumptions that the reflection coefficients $\sigma(\zeta) \equiv \bar{b}(\zeta)/a(\zeta)$, $\bar{\sigma}(\zeta) \equiv b(\zeta)/\bar{a}(\zeta)$ are rational functions. Here the scattering coefficients, $a, b, \bar{a}, \bar{b}$, are defined according to [16]. (Note that as per the AKNS convention, the overbar does not indicate a complex conjugation, but rather indicates an independent quantity. Complex conjugation will

be indicated instead by an asterisk.) We will write $\mathbb{C}_+(\mathbb{C}_-)$ for the open upper (lower) half ζ -plane. Note that when we take $r(x) = 0 = q(x)$ for $x < 0$, it then follows from appendix 5 of [16] that $b(\zeta)$ becomes analytic in $\mathbb{C}_-$ while $\bar{b}(\zeta)$ becomes analytic in $\mathbb{C}_+$. In general, we shall treat this problem with r and q both independent and unrelated, although we shall also discuss what restrictions and symmetries apply when these potentials are constrained in the usual manners such as $r(x) = \pm q(x)$ and $r(x) = \pm q^*(x)$. The Schrödinger equation variation of this problem has been treated in [17–19].

For solving the inverse scattering problem, the analytical properties of the scattering coefficients $a, b, \bar{a}, \bar{b}$ are decisive. We recall that for any potentials $q(x), r(x)$ vanishing stronger than any power for $|x| \rightarrow \infty$, $a(\zeta)$ is an analytic function in $\mathbb{C}_+$ while $\bar{a}(\zeta)$ is an analytic function in $\mathbb{C}_-$. It has already been pointed out above that under the more stringent condition $q = r = 0$ for $x < 0$, $\bar{b}(\zeta)$ is analytic in $\mathbb{C}_+$ while $b(\zeta)$ is analytic in $\mathbb{C}_-$. Consequently any singularity of $\sigma(\bar{\sigma})$ in $\mathbb{C}_+(\mathbb{C}_-)$ must be a zero of $a(\bar{a})$.

In relation to the above, there are several series of articles dealing with various features of the inverse scattering problem for the Dirac problem (which is another equivalent form of (1)) on the semi-line (see [20–25] and references therein). These papers use the spectral function and the Weyl–Titchmarsh function, where the latter corresponds to the reflection coefficient(s) of (1). This function is more general however because of the weaker conditions used concerning the asymptotics of the potentials. Also $2n$ -dimensional generalizations of the 2×2 scattering problem are treated. However these articles are generally restricted, in AKNS notation, to the $r = \pm q^*$ case.

There are two ways to solve the Marchenko equations: either invert about $x = -\infty$ or invert about $x = +\infty$. On a semi-infinite interval where the potentials vanish for $x < 0$, these two ways are no longer equivalent. One finds that there are both subtle and distinct differences between them. Here we shall treat the case when we invert about $x = -\infty$. At the end of this section we shall make some comments about how inversion about $x = \infty$ would differ from that given here. The Marchenko equations for inversion about $x = +\infty$ were first given in [26]. Inversion about $x = -\infty$ can be equivalently done and was given in [16], equations (4.39a), (4.39b), which is

$$\bar{L}(x, y) + \begin{pmatrix} 1 \\ 0 \end{pmatrix} G(x+y) - \int_{-\infty}^x L(x, s) G(s+y) ds = 0, \quad x > y \quad (2)$$

$$L(x, y) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \bar{G}(x+y) + \int_{-\infty}^x \bar{L}(x, s) \bar{G}(s+y) ds = 0, \quad x > y. \quad (3)$$

where under the conditions considered above,

$$\begin{aligned} G(z) &= \frac{1}{2\pi} \int_{\mathcal{C}} \sigma(\zeta) e^{-i\zeta z} d\zeta \\ \bar{G}(z) &= \frac{1}{2\pi} \int_{\mathcal{C}} \bar{\sigma}(\zeta) e^{i\zeta z} d\zeta. \end{aligned} \quad (4)$$

In (4), the contour $\mathcal{C}$ of integration goes from $\zeta = -\infty$ to $+\infty$ and either on or above the real axis and also above all zeros of $a(\zeta)$ in $\mathbb{C}_+$. Meanwhile $\bar{\mathcal{C}}$ goes from $\zeta = -\infty$ to $+\infty$ and either on or below the real axis and also below all zeros of $\bar{a}(\zeta)$ in $\mathbb{C}_-$. Note that as mentioned before, there will be no poles of $\bar{b}(\zeta)$ in $\mathbb{C}_+$ and no poles of $b(\zeta)$ in $\mathbb{C}_-$. Therefore $\mathcal{C}(\bar{\mathcal{C}})$ as well may be defined as going above (below) all poles of $\sigma(\bar{\sigma})$.

This is a system of linear integral equations for the two-component vector functions $L \equiv (L_1, L_2)^T$, $\bar{L} \equiv (\bar{L}_1, \bar{L}_2)^T$. On an infinite interval, the kernels $G, \bar{G}$ can generally be given as sums over the bound state eigenvalues plus Fourier integrals over the real ζ -axis. On the semi-infinite interval considered here, they may be reduced to a similar form.

From the Marchenko equations, the required potentials can be reconstructed from equations (4.42a), (4.42d) in [16],

$$q(x) = 2\bar{L}_1(x, x), \quad r(x) = -2L_2(x, x). \quad (5)$$

Now we assume that the reflection coefficients $\sigma(\zeta)$ and $\bar{\sigma}(\zeta)$ are rational functions and that they vanish at least as fast as $O(1/\zeta)$ for $|\zeta| \rightarrow \infty$. For $z < 0$, it follows that the contour $\mathcal{C}(\bar{\zeta})$ may be closed in $\mathbb{C}_+(\mathbb{C}_-)$, giving $G(z) = \bar{G}(z) = 0$. For $z > 0$ then the contour $\mathcal{C}(\bar{\zeta})$ may be closed in $\mathbb{C}_-(\mathbb{C}_+)$ so that G and $\bar{G}$ may be written as sums over the poles. Let us assume that there are N distinct poles ζ_j , $j = 1, \dots, N$, of σ and $\bar{N}$ distinct poles $\bar{\zeta}_l$, $l = 1, \dots, \bar{N}$, of $\bar{\sigma}$ with the respective residues σ_j and $\bar{\sigma}_l$. Note that if $\bar{b}(\zeta)$ has a pole or poles in $\mathbb{C}_-$, then some of the ζ_j 's could lie in $\mathbb{C}_-$. Similarly for the $\bar{\zeta}_l$'s. Summarizing then we find

$$G(z) = -i\theta(z) \sum_{j=1}^N \sigma_j e^{-i\zeta_j z} \quad (6)$$

$$\bar{G}(z) = i\theta(z) \sum_{l=1}^{\bar{N}} \bar{\sigma}_l e^{i\bar{\zeta}_l z},$$

where $\theta(z)$ denotes the Heaviside step function.

It is very easy to solve the Marchenko equations for $x < 0$. Since $G(z) = \bar{G}(z) \equiv 0$ for $z < 0$, we see that (2), (3) have the solutions

$$L(x, y) = \bar{L}(x, y) \equiv 0, \quad y \leq x < 0, \quad (7)$$

and, due to (5), it follows that

$$q(x) = r(x) \equiv 0, \quad x < 0. \quad (8)$$

Now we consider the region $x > 0$. Then, since $G(z) = \bar{G}(z) = 0$ for $z < 0$ we may take $-\infty$ instead of $-\infty$ as the lower limits of integration in (2), (3).

A solution of these integral equations may be given by a closed system of ordinary differential equations (ODEs). As a first step to achieve this goal it is useful to replace y by $-y$ in (3), but not so in (2). Since our range of the variable y is $-x < y < +x$, with $x > 0$ understood to be a fixed parameter, this range does not change when y changes sign. Then the substitution of G and $\bar{G}$ from (6) leads to

$$\bar{L}(x, y) + i \sum_{j=1}^N \sigma_j e^{-i\zeta_j y} \left[-\begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta_j x} + \int_{-y}^x L(x, s) e^{-i\zeta_j s} ds \right] = 0 \quad (9)$$

$$L(x, -y) + i \sum_{m=1}^{\bar{N}} \bar{\sigma}_m e^{-i\bar{\zeta}_m y} \left[\begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\bar{\zeta}_m x} + \int_y^x \bar{L}(x, s) e^{i\bar{\zeta}_m s} ds \right] = 0. \quad (10)$$

To solve these Marchenko equations, note that due to the y dependence in the lower limits of the integrals in (9) and (10), one cannot use the usual method of solution for an N -soliton solution, wherein one takes $\bar{L}(x, y)$ and $L(x, y)$ to be sums of separable functions of x and y .

Let us now introduce the two-component column functions

$$l_j(x, y) \equiv e^{-i\zeta_j y} \left[\begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta_j x} - \int_{-y}^x L(x, s) e^{-i\zeta_j s} ds \right], \quad j = 1, \dots, N \quad (11)$$

$$\bar{l}_m(x, y) \equiv -e^{-i\bar{\zeta}_m y} \left[\begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\bar{\zeta}_m x} + \int_y^x \bar{L}(x, s) e^{i\bar{\zeta}_m s} ds \right], \quad m = 1, \dots, \bar{N} \quad (12)$$

fulfilling

$$l_j(x, -x) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \bar{l}_m(x, x) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \quad (13)$$

then (9), (10) may be written as

$$\bar{L}(x, y) = i \sum_{j=1}^N \sigma_j l_j(x, y), \quad L(x, y) = i \sum_{m=1}^{\bar{N}} \bar{\sigma}_m \bar{l}_m(x, -y), \quad (14)$$

and according to (5) the potentials are expressed as

$$q(x) = 2\bar{L}_1(x, x) = 2i \sum_j \sigma_j l_{j1}(x, x) \quad (15)$$

$$r(x) = -2L_2(x, x) = -2i \sum_m \bar{\sigma}_m \bar{l}_{m2}(x, -x). \quad (16)$$

Differentiation of (11), (12) with respect to y then leads to a closed system of constant coefficient ODEs for $l_j, \bar{l}_m$,

$$\begin{aligned} (\partial_y + i\zeta_j)l_j(x, y) &= -L(x, -y) = -i \sum_m \bar{\sigma}_m \bar{l}_m(x, y) \\ (\partial_y + i\bar{\zeta}_m)\bar{l}_m(x, y) &= \bar{L}(x, y) = i \sum_j \sigma_j l_j(x, y) \end{aligned} \quad (17)$$

with x being a parameter. Let us write $l_{jp}, \bar{l}_{mp}$, $p = 1, 2$ for the components of $l_j, \bar{l}_m$ and combine two column vectors

$$\mathbf{l}_p = \begin{pmatrix} l_{jp} \\ \bar{l}_{mp} \end{pmatrix}, \quad p = 1, 2 \quad (18)$$

each of them consisting of $N + \bar{N}$ components. Now we define an $(N + \bar{N}) \times (N + \bar{N})$ -matrix M ,

$$M = \begin{pmatrix} M^{11} & M^{12} \\ M^{21} & M^{22} \end{pmatrix} \quad (19)$$

with the submatrices

$$M^{11} = \text{diag}[\zeta_1 \cdots \zeta_N], \quad M^{22} = \text{diag}[\bar{\zeta}_1 \cdots \bar{\zeta}_{\bar{N}}] \quad (20)$$

$$M^{12} = \begin{pmatrix} \bar{\sigma}_1 & \cdots & \bar{\sigma}_{\bar{N}} \\ \cdots & \cdots & \cdots \\ \bar{\sigma}_1 & \cdots & \bar{\sigma}_{\bar{N}} \end{pmatrix}, \quad N \text{ rows} \quad (21)$$

$$M^{21} = - \begin{pmatrix} \sigma_1 & \cdots & \sigma_N \\ \cdots & \cdots & \cdots \\ \sigma_1 & \cdots & \sigma_N \end{pmatrix}, \quad \bar{N} \text{ rows.} \quad (22)$$

In this way we find that both column vectors $\mathbf{l}_p$ fulfil one and the same ODE system,

$$i\partial_y \mathbf{l}_p(x, y) = M \mathbf{l}_p(x, y), \quad p = 1, 2, \quad -x < y < x \quad (23)$$

which is subject to two-point boundary conditions, cf (13),

$$(a) \quad l_{j1}(x, -x) = 1, \quad \bar{l}_{m1}(x, x) = 0 \quad (24)$$

or

$$(b) \quad l_{j2}(x, -x) = 0, \quad \bar{l}_{m2}(x, x) = -1, \quad (25)$$

respectively.

The general solution of this problem can be given in terms of the matrix exponential function

$$\Gamma(y) \equiv \exp(iMy) \quad (26)$$

so that it holds generally

$$I_p(x, y_2) = \Gamma(y_1 - y_2) I_p(x, y_1). \quad (27)$$

In quite an analogous way to (19) Γ may be divided into four block matrices. Then from (27) together with (24), (25) we find

$$I_{j1}(x, x) = \sum_{j'} ([\Gamma^{11}(2x)]^{-1})_{jj'}, \quad (28)$$

$$\bar{I}_{m2}(x, -x) = - \sum_{m'} ([\Gamma^{22}(-2x)]^{-1})_{mm'}. \quad (29)$$

From (15), (16), we now obtain the potentials $q(x)$ and $r(x)$. Due to (27), (14), we now know L and $\bar{L}$ from which one could also construct the Jost functions ϕ and $\bar{\phi}$ [16].

With the above, the question of the existence of a solution naturally arises. We can briefly address this here by referring to the known solutions of the infinite interval case. The key question is could Γ^{11} and Γ^{22} ever be non-invertible? As discussed in AKNS [16], on the infinite interval and for the case where $r = -q^*$, the solution always exists for any finite number of solitons and for any set of reflection coefficients, vanishing sufficiently rapidly as $\zeta \rightarrow \infty$. Thus in this case, it follows that one could in general expect that the matrices Γ^{11} and Γ^{22} would always exist and be nonsingular. Similarly, when $r = +q^*$, it follows that if the reflection coefficients are always bounded by unity and if there are no bound states, then it was also shown that nonsingular solutions for r and q on the infinite interval would always exist. Again providing that those conditions were satisfied, then one would in general expect the matrices Γ^{11} and Γ^{22} to exist and be nonsingular. However if these conditions are found to *not* exist, then as shown in [16], the presence of even a single bound state in the scattering data when $r = +q^*$ would give a singular solution of the form $1/\sinh \eta x$. Whence the solutions for r and q would be singular at some values of x on the infinite interval. In the more general case, if r and q were generally unrelated, the same would still generally occur (singular solutions). It is found that on a finite or semi-infinite interval, these singularities could well occur only in the *nonphysical* regime, and then leaving nonsingular solutions in physical regime. Thus except in the case where $r = -q^*$, determining when the solution exists in the *physical regime* is an open problem and in need of solution.

Let us now discuss the problem of inversion about $x = +\infty$ for the semi-infinite case. As long as $x > 0$, the reconstruction process would be exactly the same as in the infinite interval with one key exception. First, $q(x) = r(x) = 0$ for $x < 0$ is equivalent to the condition that $\bar{b}(\zeta)$ is analytic in $\mathbb{C}_+$ and $b(\zeta)$ is analytic in $\mathbb{C}_-$. Note that in general there still could be poles of $b(\zeta)$ in $\mathbb{C}_+$ and poles of $\bar{b}(\zeta)$ in $\mathbb{C}_-$. Now according to equations (4.38a), (4.38b) in [16] we would have to use the reflection coefficients $\rho \equiv b/a$, $\bar{\rho} \equiv \bar{b}/\bar{a}$ instead of σ , $\bar{\sigma}$. Then we would have to take a contour $\mathcal{C}$ of integration going above all zeros of a but below all poles of b in $\mathbb{C}_+$ and a contour $\bar{\mathcal{C}}$ going below all zeros of $\bar{a}$ but above all poles of $\bar{b}$ in $\mathbb{C}_-$. (Following the derivation outlined in [16], equations (4.30)–(4.33), it is rather direct to show that only zeros of a ($\bar{a}$) contribute to these contours, and that the poles of b and $\bar{b}$ do not.) Then the reconstruction would proceed as in the infinite interval case.

However, if one was using inversion about $x = +\infty$ and was also interested in verifying that the solutions for r and q were indeed zero when $x < 0$, then that proof would have to

follow along the lines of the earlier parts of this section in order to obtain the solution of the Marchenko equations for the region $x < 0$. The procedure does work, although much more effort would be required just to rediscover that for a simple solution such as a truncated one-soliton potential, one does have indeed r and q vanishing for $x < 0$.

In the case of a finite interval, it would not matter about which direction one would invert since both $\pm\infty$ are excluded from the finite interval and one would therefore need to proceed as in the earlier parts of this section.

2.2. Solution procedure

Here we shall briefly summarize the preceding results into a rather simple procedure. We start with two rational reflection coefficients $\sigma(\zeta)$, $\bar{\sigma}(\zeta)$ vanishing at infinity and with no multiple poles but otherwise arbitrary. Let N , $\bar{N}$ be the respective numbers of poles. Then one proceeds as follows:

- (1) Establish the matrix M from the poles and residues of $\sigma(\zeta)$, $\bar{\sigma}(\zeta)$ according to (19)–(22).
- (2) Take the matrix exponential function $\Gamma(y) \equiv \exp[iMy]$, see (26). There are several methods to get this function. A related discussion can be found, e.g., in [27]. One way is to diagonalize M , compare subsection 2.4. Since the matrix exponential is usually included in any standard mathematical software system such as, e.g., Mathematica, we shall give no further details here.
- (3) Invert the submatrices Γ^{11} , Γ^{22} to get the functions $l_{j1}(x, x)$, $\bar{l}_{m2}(x, -x)$ from (28), (29) so that the potentials $q(x)$, $r(x)$ are finally found according to (15), (16). The pointwise invertibility of these submatrix functions cannot be guaranteed generally except in the case of $r = -q^*$.

2.3. A system of PDEs for the auxiliary functions

The two $2 \times (N + \bar{N})$ functions $l_{jp}(x, y)$, $\bar{l}_{mp}(x, y)$, $j = 1, \dots, N$, $m = 1, \dots, \bar{N}$, $p = 1, 2$ defined by (11), (12) and combined to form two $(N + \bar{N})$ -vectors $\mathbf{l}_1, \mathbf{l}_2$ in (18) do play a key role in the procedure developed above. They are defined in (11), (12) as Fourier-type integrals over the functions $L_p, \bar{L}_p$, $p = 1, 2$, which were introduced in [16]. Vice versa $L_p, \bar{L}_p$ are expressed as sums over $l_{jp}, \bar{l}_{mp}$ with constant coefficients, see (14). Here we will derive a system of PDEs for $\mathbf{l}_1, \mathbf{l}_2$ which resembles such a system for $L_p, \bar{L}_p$ known from [16], and, indeed, the latter can be rediscovered from the former. We believe that it might be of interest to point out this nontrivial structure.

Obviously, together with $\mathbf{l}_p(x, y)$, p fixed, each of the column vectors $\partial_x \mathbf{l}_p, \partial_y \mathbf{l}_p, (\partial_x \pm \partial_y) \mathbf{l}_p$ fulfils the ODE system (23) with the same matrix M for any fixed x . In particular, let us take

$$\mathbf{h}(x, y) \equiv \begin{pmatrix} h_j \\ \bar{h}_m \end{pmatrix} \equiv (\partial_x - \partial_y) \mathbf{l}_1(x, y), \quad j = 1, \dots, N, \quad m = 1, \dots, \bar{N}. \quad (30)$$

Due to (24), (23) and with using (15) we find boundary values

$$\begin{aligned} h_j(x, -x) &= 0 \\ \bar{h}_m(x, x) &= -2\partial_y \bar{l}_{m1}(x, y)|_{y=x} = -2i \sum_j \sigma_j l_{j1}(x, x) = q(x). \end{aligned} \quad (31)$$

From a comparison of (31) with (25) we conclude $\mathbf{h}(x, y) = q(x) \mathbf{l}_2(x, y)$.

Quite similarly we may define

$$\mathbf{f}(x, y) \equiv \begin{pmatrix} f_j \\ \bar{f}_m \end{pmatrix} \equiv (\partial_x + \partial_y) \mathbf{l}_2(x, y) \quad (32)$$

to get

$$\begin{aligned} f_j(x, -x) &= 2\partial_y l_{j1}(x, y) \Big|_{y=-x} = -2i \sum_m \bar{\sigma}_m \bar{l}_{m2}(x, -x) = r(x) \\ \bar{f}_m(x, x) &= 0. \end{aligned} \quad (33)$$

From (33) compared to (25) we get $\mathbf{f}(x, y) = r(x) \mathbf{l}_1(x, y)$. Putting together we may write down the relations

$$\begin{aligned} (\partial_x - \partial_y) \mathbf{l}_1(x, y) - q(x) \mathbf{l}_2(x, y) &= 0, \\ (\partial_x + \partial_y) \mathbf{l}_2(x, y) - r(x) \mathbf{l}_1(x, y) &= 0. \end{aligned} \quad (34)$$

Due to (14) from (34) we get the PDEs

$$\begin{aligned} (\partial_x + \partial_y) L_1(x, y) - q(x) L_2(x, y) &= 0 \\ (\partial_x - \partial_y) L_2(x, y) - r(x) L_1(x, y) &= 0 \\ (\partial_x - \partial_y) \bar{L}_1(x, y) - q(x) \bar{L}_2(x, y) &= 0 \\ (\partial_x + \partial_y) \bar{L}_2(x, y) - r(x) \bar{L}_1(x, y) &= 0 \end{aligned} \quad (35)$$

corresponding to equations (4.46a), (4.46b) in [16], as well as the conditions for consistent boundary values

$$L_1(x, -x) = \bar{L}_2(x, -x) = 0 \quad (36)$$

$$L_2(x, x) = -r(x)/2, \quad \bar{L}_1(x, x) = q(x)/2. \quad (37)$$

As one can see in (35), L_1 and L_2 propagate along the characteristics $x \pm y$. Whence if q and r were suitably bounded on the interval of interest, then, as in [16], it would follow that L_1 and L_2 would exist.

2.4. The characteristic equation

By considering the characteristic equation of the matrix M , we can relate our results to the procedure developed by Rourke and Morris [14]. This equation is

$$P(\zeta) \equiv \det(M - \zeta I) = 0 \quad (38)$$

where I denotes the unit matrix of order $N + \bar{N}$. Let us first give a reformulation of the characteristic polynomial $P(\zeta)$. We divide the j th column of the matrix $(M - \zeta I)$ by $(\zeta - \zeta_j)$ and the $(\bar{N} + k)$ th by $(\zeta - \bar{\zeta}_k)$ so that we may write

$$P(\zeta) = \left[\prod_{j=1}^N (\zeta_j - \zeta) \right] \left[\prod_{k=1}^{\bar{N}} (\bar{\zeta}_k - \zeta) \right] \det W. \quad (39)$$

Here W denotes a matrix consisting of four block matrices,

$$W = \begin{pmatrix} I_N & B \\ C & I_{\bar{N}} \end{pmatrix}, \quad (40)$$

such that I_N and $I_{\bar{N}}$ are unit matrices of orders N and $\bar{N}$, respectively, while B is an $N \times \bar{N}$ -matrix with N identical rows $(b_1, \dots, b_N)$ and C is an $\bar{N} \times N$ -matrix with $\bar{N}$ identical rows $(c_1, \dots, c_{\bar{N}})$ with

$$b_k = \bar{\sigma}_k / (\zeta - \bar{\zeta}_k), \quad c_j = -\sigma_j / (\zeta - \zeta_j). \quad (41)$$

Obviously, according to the presumptions given above, the reflection coefficients $\sigma, \bar{\sigma}$ may be written as the sums

$$\sigma(\zeta) = -\sum_{j=1}^N c_j, \quad \bar{\sigma}(\zeta) = \sum_{k=1}^{\bar{N}} b_k. \quad (42)$$

Due to the lemma given in the appendix together with (41), (42) we find

$$\det W = 1 + \sigma(\zeta)\bar{\sigma}(\zeta). \quad (43)$$

It is obvious that the zeros of (43) are eigenvalues of the matrix M . However, not all eigenvalues of M need to be zeros of (43). It may happen that $\bar{\sigma}(\zeta)$ could just accidentally vanish at one of the poles in σ , say at $\zeta = \zeta_k$. In such a case, in the product $\sigma\bar{\sigma}$, the corresponding pole in σ would be canceled and, due to the factor $(\zeta - \zeta_k)$ extracted in (39), this value of ζ_k would become an eigenvalue of M . More generally: when some of the poles ζ_j in σ and $\bar{\zeta}_j$ in $\bar{\sigma}$ are canceled in the product $\sigma\bar{\sigma}$ then the corresponding subset of $\{\zeta_j, \bar{\zeta}_j\}$ together with the zeros of $1 + \sigma\bar{\sigma}$ is just the eigenvalues of M .

Rourke and Morris [14] determined the eigenvalues of the completed N -soliton states as the zeros of $1 + \sigma\bar{\sigma}$ for the case $r = -q^*$, in which case $\bar{\sigma}(\zeta) = \sigma^*(\zeta^*)$ (see (55), (59)). Thus in their treatment, they found it necessary to exclude such accidental cancelations. These cases of accidental cancelations were later handled by Rourke and Saunders [28]. However one should note that no such difficulty occurs in our procedure as described in subsection 2.1 and continued below.

3. Reductions

In physical applications, the general AKNS system typically is simplified by one or another symmetry. Upon taking (a) $r = \pm q$ or (b) $r = \pm q^*$ we arrive at a Zakharov–Shabat problem. Here of course, we would have $\bar{N} = N$. The respective reductions for these symmetries are already given in [16]. However here we are more interested in the consequences that this has for the procedure developed in the foregoing section.

(a) For $r = \epsilon q, \epsilon = \pm 1$ it holds

$$\bar{\sigma}(\zeta) = -\epsilon\sigma(-\zeta), \quad \bar{G}(z) = \epsilon G(z) \quad (44)$$

and $L_2 = \epsilon \bar{L}_1, \bar{L}_2 = -L_1$. There are $2N$ poles $\zeta_j, \bar{\zeta}_j, j = 1, \dots, N$, and we will assume here that they all are different from one another. (A degenerate set of poles would require a separate treatment.) These poles may be ordered such that

$$\bar{\zeta}_j = -\zeta_j, \quad \bar{\sigma}_j = -\epsilon\sigma_j. \quad (45)$$

Due to (11), (12) we obtain

$$l_{j1}(x, y) = \epsilon \bar{l}_{j2}(x, -y), \quad \bar{l}_{j2}(x, y) = -l_{j1}(x, -y). \quad (46)$$

The $2N \times 2N$ -matrix M has the properties

$$M^{22} = -M^{11}, \quad M^{21} = -\epsilon M^{12}. \quad (47)$$

By the use of the $2N \times 2N$ -matrices

$$E \equiv \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ \epsilon \mathbf{1} & \mathbf{0} \end{pmatrix} \quad (48)$$

with $\mathbf{0}$ and $\mathbf{1}$ denoting the $N \times N$ null matrix and the $N \times N$ unit matrix respectively, the symmetry (47) of M can be characterized by the relation

$$EME^{-1} = -M. \quad (49)$$

For any natural number n , it holds $EM^nE^{-1} = (-M)^n$. Consequently, the matrix exponential function (26) has the symmetry

$$E\Gamma(y)E^{-1} = \Gamma(-y). \quad (50)$$

If λ_1 is an eigenvalue of M , then $-\lambda_1$ is an eigenvalue as well. We shall assume that there are no degenerate eigenvalues.

It then follows that there are $2N$ eigenvalues $\lambda_1, \dots, \lambda_{2N}$ which may be ordered such that $\lambda_{j+N} = -\lambda_j$. Let D be the diagonal matrix

$$D = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_{2N}]. \quad (51)$$

There exists a $2N \times 2N$ -matrix A such that

$$AM = DA. \quad (52)$$

The rows of A are the eigenvectors of M . Multiplying (52) by E from the left and by E^{-1} from the right, using (49) and the obvious relation $EDE^{-1} = -D$, we find that in (52) A may be replaced by EAE^{-1} or, as well, by a linear combination of both matrices. In particular we may replace A by $\tilde{A} \equiv A - EAE^{-1}$ fulfilling the symmetry $E\tilde{A}E^{-1} = -\tilde{A}$. Thus, without loss of generality we may require the diagonalizing matrix A to fulfil the same symmetry as M , cf (49), namely

$$EAE^{-1} = -A. \quad (53)$$

(b) For $r = \epsilon q^*$, $\epsilon = \pm 1$ it holds that

$$\bar{\sigma}(\zeta) = -\epsilon \sigma^*(\zeta^*), \quad \bar{G}(z) = -\epsilon G^*(z), \quad (54)$$

and $L_2 = \epsilon \bar{L}_1^*$, $\bar{L}_2 = -L_1^*$. The poles may be ordered such that

$$\bar{\zeta}_j = \zeta_j^*, \quad \bar{\sigma}_j = -\epsilon \sigma_j^*. \quad (55)$$

$$l_{j1}(x, y) = \epsilon \bar{l}_{j2}^*(x, -y), \quad \bar{l}_{j2}(x, -y) = -l_{j1}^*(x, y). \quad (56)$$

The symmetry of the matrix M is characterized by

$$EME^{-1} = M^* \quad (57)$$

with the same matrix E defined by (48), and in quite an analogous way as for (a) we find the symmetry of Γ ,

$$E\Gamma(y)E^{-1} = \Gamma^*(-y). \quad (58)$$

With the specification (54) the equation (43) reads

$$\det W = 1 - \epsilon \sigma(\zeta) \sigma^*(\zeta^*). \quad (59)$$

For $\epsilon = -1$ the equation $\det W = 0$ corresponds exactly to an equation found in [14], which is located in the text below their equation (24). In comparing these equations we only need to keep in mind (a) that their variable τ corresponds to our variable x and (b) that they are considering a pulse vanishing for $\tau > 0$ while we take the potential to vanish for $x < 0$.

4. Computing all scattering data from the reflection coefficients

It has been demonstrated in appendix 5 of [16] how one can determine all scattering data from the so-called *primordial scattering data*. In particular, for $\sigma, \bar{\sigma}$ being rational functions of ζ and regular at infinity together with the requirement that $\bar{b}(\zeta)(b(\zeta))$ is analytic in $\mathbb{C}_+(\mathbb{C}_-)$, we know the zeros of $a(\bar{a})$ in $\mathbb{C}_+(\mathbb{C}_-)$ and so we are able to determine $a, b, \bar{a}, \bar{b}$ explicitly by the use of this procedure. Particular symmetries such as discussed in section 3 would simplify the procedure. Instead here, more generally, we will require a *minimal symmetry*. It will be seen that this symmetry is a sufficient condition for the R&M method of construction to work.

First we assume $\bar{N} = N$, i.e., that there are as many poles $\zeta_j, j = 1, \dots, N$ of σ as there are poles $\bar{\zeta}_j$ of $\bar{\sigma}(\zeta)$. Moreover we assume that for any fixed j the points $\zeta_j, \bar{\zeta}_j$ are in different half-planes $\mathbb{C}_+, \mathbb{C}_-$. E.g., for the reductions (a) $r = \pm q$ or (b) $r = \pm q^*$ with taking $\bar{\zeta}_j = -\zeta_j$ or $\bar{\zeta}_j = \zeta_j^*$, respectively, these conditions are fulfilled. Note that this minimal symmetry is not a condition on the potentials but on the reflection coefficients $\sigma, \bar{\sigma}$.

We define two functions $f, \bar{f}$,

$$f(\zeta) = a\Pi_1, \quad \bar{f}(\zeta) = \bar{a}/\Pi_1, \quad \Pi_1 \equiv \prod_{\zeta_j \in \mathbb{C}_+} \frac{\zeta - \bar{\zeta}_j}{\zeta - \zeta_j}. \quad (60)$$

Clearly, $\log f$ is analytic in $\mathbb{C}_+$ while $\log \bar{f}$ is analytic in $\mathbb{C}_-$, and both functions vanish at least as fast as $1/\zeta$ as $|\zeta| \rightarrow \infty$. Now we take the integral over $\log[f(\zeta')\bar{f}(\zeta')]/(\zeta' - \zeta)$ along the real ζ' -axis. This is a sum of two integrals. For the term containing $\log f$ ($\log \bar{f}$) the contour of integration may be closed in $\mathbb{C}_+(\mathbb{C}_-)$. Making use of the relation

$$a(\zeta)\bar{a}(\zeta) + b(\zeta)\bar{b}(\zeta) \equiv a(\zeta)\bar{a}(\zeta)[1 + \sigma(\zeta)\bar{\sigma}(\zeta)] = 1 \quad (61)$$

we arrive at

$$\frac{1}{2\pi i} \int_{-\infty}^{\infty} \log[1 + \sigma(\xi')\bar{\sigma}(\xi')] \frac{d\xi'}{\xi' - \zeta} = \begin{cases} -\log f(\zeta), & \zeta \in \mathbb{C}_+ \\ \log \bar{f}(\zeta), & \zeta \in \mathbb{C}_-. \end{cases} \quad (62)$$

The product $\sigma(\zeta)\bar{\sigma}(\zeta)$ has $2N_1 \leq 2N$ poles. This means that σ contains $(N - N_1)$ multiplicative factors of the type $(\zeta - \bar{\zeta}_k)/(\zeta - \zeta_k), k = N_1 + 1, \dots, N$, which are canceled in the product by the reciprocal factors contained in $\bar{\sigma}$. There are $2N_1$ zeros of $1 + \sigma(\zeta)\bar{\sigma}(\zeta)$, say $\lambda_l, \bar{\lambda}_l, l = 1, \dots, N_1$, with an as far arbitrary order. Thus we may write

$$1 + \sigma(\zeta)\bar{\sigma}(\zeta) = \prod_{l=1}^{N_1} \left[\frac{(\zeta - \lambda_l)(\zeta - \bar{\lambda}_l)}{(\zeta - \zeta_l)(\zeta - \bar{\zeta}_l)} \right]. \quad (63)$$

Now we substitute (63) into (62). It is helpful to write down the subsequent formula which is easily verified by partial integration and then closing the contour in $\mathbb{C}_-$. For $\zeta \in \mathbb{C}_+$ it holds that

$$\begin{aligned} \frac{1}{2\pi i} \int_{-\infty}^{+\infty} \log \left[\frac{s - z_1}{s - z_2} \right] \frac{ds}{s - \zeta} \\ = \begin{cases} \log[1 - z_1/\zeta], & z_1 \in \mathbb{C}_-, \quad z_2 \in \mathbb{C}_+ \\ -\log[1 - z_2/\zeta], & z_1 \in \mathbb{C}_+, \quad z_2 \in \mathbb{C}_- \\ \log[(\zeta - z_1)/(\zeta - z_2)], & z_1 \in \mathbb{C}_-, \quad z_2 \in \mathbb{C}_- \\ 0, & z_1 \in \mathbb{C}_+, \quad z_2 \in \mathbb{C}_+. \end{cases} \end{aligned} \quad (64)$$

The denominator in (63) has exactly N_1 zeros in $\mathbb{C}_+$ and, as well, N_1 zeros in $\mathbb{C}_-$. Then the same must hold for the numerator because otherwise—according to (64)—the integration would lead to a singularity at $\zeta = 0$. Without loss of generality, we may take

$\lambda_l \in \mathbb{C}_+$, $\bar{\lambda}_l \in \mathbb{C}_-$, $l = 1, \dots, N_1$. We introduce two discrete functions z_+ , z_- defined on the set $\zeta_j, \bar{\zeta}_j$, $j = 1, \dots, N$, as follows:

$$z_{\pm}(\zeta_j) = z_{\pm}(\bar{\zeta}_j) = \begin{cases} \zeta_j & \text{if } \zeta_j \in \mathbb{C}_{\pm} \\ \bar{\zeta}_j & \text{if } \bar{\zeta}_j \in \mathbb{C}_{\pm}. \end{cases} \quad (65)$$

Choosing $\zeta \in \mathbb{C}_+$ we find

$$f(\zeta) = \prod_{l=1}^{N_1} \frac{\zeta - \bar{\lambda}_l}{\zeta - z_-(\zeta_l)} \quad (66)$$

and combination with (60) leads to

$$a(\zeta) = \prod_{l=1}^{N_1} \frac{\zeta - \zeta_l}{\zeta - \bar{\lambda}_l} \prod_{\{k > N_1, \zeta_k \in \mathbb{C}_+\}} \frac{\zeta - \zeta_k}{\zeta - \bar{\zeta}_k}. \quad (67)$$

Note that this formula does not distinguish which of the ζ_l , $l = 1, \dots, N_1$, is in $\mathbb{C}_+$ and which in $\mathbb{C}_-$. Afterward we may even admit values ζ_l on the real axis. In an analogous way we find

$$\bar{a}(\zeta) = \prod_{l=1}^{N_1} \frac{\zeta - \bar{\zeta}_l}{\zeta - \lambda_l} \prod_{\{k > N_1, \zeta_k \in \mathbb{C}_+\}} \frac{\zeta - \bar{\zeta}_k}{\zeta - \zeta_k}. \quad (68)$$

5. The completed N -soliton potential

In this section again we assume what we have called *minimal symmetry* in the foregoing section. We may combine the potential $q_R(x) \equiv q(x)$ vanishing for $x < 0$ with a potential $q_L(x)$ vanishing for $x > 0$. The respective scattering matrices are defined as

$$S_R(\zeta) \equiv \begin{pmatrix} a_R & -\bar{b}_R \\ b_R & \bar{a}_R \end{pmatrix}, \quad S_L(\zeta) \equiv \begin{pmatrix} a_L & -\bar{b}_L \\ b_L & \bar{a}_L \end{pmatrix}, \quad (69)$$

and the scattering matrix S_C of the combined potential is just the product

$$S_C \equiv \begin{pmatrix} a_C & -\bar{b}_C \\ b_C & \bar{a}_C \end{pmatrix} = S_R S_L = \begin{pmatrix} a_R a_L - \bar{b}_R b_L & a_R \bar{b}_L + \bar{b}_R \bar{a}_L \\ \bar{a}_R b_L + b_R a_L & \bar{a}_R \bar{a}_L - b_R \bar{b}_L \end{pmatrix}. \quad (70)$$

S_R and S_L have to be unimodular matrices, i.e.,

$$(a\bar{a} + b\bar{b})_R = (a\bar{a} + b\bar{b})_L = 1. \quad (71)$$

Note that $a_R, b_R, \dots$ here are the same as $a, b, \dots$ outside this section. Below we use the notation, cf [16],

$$\rho \equiv b/a, \quad \bar{\rho} \equiv \bar{b}/\bar{a}, \quad \sigma \equiv \bar{b}/a, \quad \bar{\sigma} \equiv b/\bar{a} \quad (72)$$

with subscripts R or L to be added. From appendix 5 in [16] we know that $a_R, \bar{b}_R, a_L, b_L$ are analytic functions in $\mathbb{C}_+$ while $\bar{a}_R, b_R, \bar{a}_L, \bar{b}_L$ are analytic functions in $\mathbb{C}_-$.

Now we require that the combined potential should be reflectionless, i.e.,

$$b_C = \bar{b}_C = 0 \text{ or}$$

$$\rho_L = -\bar{\sigma}_R, \quad \bar{\rho}_L = -\sigma_R. \quad (73)$$

From (73) together with (71), (72) we get

$$a_C = a_R a_L (1 + \sigma_R \bar{\sigma}_R) = \frac{a_L}{\bar{a}_R} (a_R \bar{a}_R + b_R \bar{b}_R) = \frac{a_L}{\bar{a}_R} = \frac{a_R}{\bar{a}_L}. \quad (74)$$

From (67), (68) we know a_R , $\bar{a}_R$, and, clearly, from knowledge of $\rho_L(\zeta)$ due to (72) we may determine a_L , $\bar{a}_L$ by a procedure quite analogous to that used in section 4. In particular we get

$$a_L(\zeta) = \prod_{l=1}^{N_1} \frac{\zeta - \bar{\zeta}_l}{\zeta - \bar{\lambda}_l} \prod_{\{k > N_1, \zeta_k \in \mathbb{C}_-\}} \frac{\zeta - \bar{\zeta}_k}{\zeta - \bar{\zeta}_k}. \quad (75)$$

When now we identify

$$\lambda_k = z_+(\zeta_k), \quad \bar{\lambda}_k = z_-(\zeta_k), \quad k = N_1 + 1, \dots, N \quad (76)$$

then, due to (74), we arrive at

$$a_C(\zeta) = a_L/\bar{a}_R = \prod_{j=1}^N \frac{\zeta - \lambda_j}{\zeta - \bar{\lambda}_j}, \quad \bar{a}_C = 1/a_C. \quad (77)$$

This is the typical structure of a , $\bar{a}$ for an N -soliton state.

Conclusion. We see that under the assumption of minimal symmetry, the considered *right* potential (vanishing for $x < 0$) can be combined with a *left* potential such that the combined potential is reflectionless, i.e., it is an N -soliton potential. This statement generalizes the corresponding results in [14, 28].

Combining these results with (38), (43), (63) we see that the zeros λ_j of a_C , $j = 1, \dots, N$, together with the zeros $\bar{\lambda}_j$ of $\bar{a}_C$ are the eigenvalues of our key matrix M , while at the same time, they are also just the eigenvalues of the completed N -soliton potential.

6. Examples

We shall now present a few examples of solutions constructed by the application of the procedure outlined in subsection 2.2. Here N and $\bar{N}$ will be restricted to be equal to 2 at most, in which case the resulting potentials can be given in analytical form.

Example 1. The simplest example is to take—in the case of no symmetry reduction— $N = 1$, $\bar{N} = 0$. The ‘matrix’ M then has one element only,

$$M = M^{11} = \zeta_1. \quad (78)$$

Consequently we have

$$\Gamma(x) = \Gamma^{11}(x) = \exp[i\zeta_1 x], \quad (79)$$

$$l_{11}(x, x) = \exp[-2i\zeta_1 x], \quad (80)$$

$$q(x) = 2i\theta(x)\sigma_1 \exp[-2i\zeta_1 x], \quad r(x) = 0. \quad (81)$$

We have to choose the pole $\zeta_1 \in \mathbb{C}_-$ in order to obtain an exponentially decreasing potential. It is easy to solve the direct scattering problem for the potentials (81). Using scattering data a , b , $\bar{a}$, $\bar{b}$ in AKNS notation we find

$$a = \bar{a} = 1, \quad b = 0, \quad \bar{b} = \frac{\sigma_1}{\zeta - \zeta_1}, \quad (82)$$

and consistently we recover the reflection coefficients

$$\sigma(\zeta) \equiv \frac{\bar{b}}{a} = \frac{\sigma_1}{\zeta - \zeta_1}, \quad \bar{\sigma}(\zeta) \equiv \frac{b}{\bar{a}} = 0. \quad (83)$$

Since the coefficients a and $\bar{a}$ have no zeros, this solution has no solitons. Thus it cannot have a completion. Whence this demonstrates that there are scattering data for which a completion

cannot be done. However, this system also does not have the minimal symmetry introduced above.

Example 2. We choose reduction (a) of section 3 and take $N = \bar{N} = 1$, $\bar{\zeta}_1 = -\zeta_1$, $\bar{\sigma}_1 = -\epsilon\sigma_1$, $\epsilon = \pm 1$ to obtain the matrices

$$M = \begin{pmatrix} \zeta_1 & -\epsilon\sigma_1 \\ -\sigma_1 & -\zeta_1 \end{pmatrix} \quad (84)$$

$$\Gamma(y) \equiv \exp[iMy] = 1 \cosh(\mu_1 y) + \frac{i}{\mu_1} M \sinh(\mu_1 y) \quad (85)$$

with

$$\mu_1 = \sqrt{-\zeta_1^2 - \epsilon\sigma_1^2}. \quad (86)$$

According to (28), (29), (45)–(50) we find

$$l_{11}(x, x) = [\Gamma^{11}(2x)]^{-1}, \quad \bar{l}_{12}(x, -x) = -[\Gamma^{22}(-2x)]^{-1} = -l_{11}(x, x) \quad (87)$$

$$q(x) = \epsilon r(x) = \frac{2i\theta(x)\sigma_1}{\cosh(2\mu_1 x) + (i\zeta_1/\mu_1) \sinh(2\mu_1 x)}. \quad (88)$$

μ_1 must not be purely imaginary in order to get an exponentially decaying potential.

Example 3 (cf [12]). Now we choose reduction (b) of section 5 and take, in particular, $N = \bar{N} = 1$, $\zeta_1 = \xi_1 + i\eta_1$, $\bar{\zeta}_1 = \xi_1 - i\eta_1$, η_1 real, $\bar{\sigma}_1 = -\epsilon\sigma_1^*$, $\epsilon = \pm 1$ to obtain the matrices

$$M = \begin{pmatrix} \xi_1 + i\eta_1 & -\epsilon\sigma_1^* \\ -\sigma_1 & \xi_1 - i\eta_1 \end{pmatrix}, \quad M_0 = \begin{pmatrix} i\eta_1 & -\epsilon\sigma_1^* \\ -\sigma_1 & -i\eta_1 \end{pmatrix} \quad (89)$$

$$\Gamma(y) \equiv \exp[iMy] = e^{i\xi_1 y} \left[1 \cosh(\kappa_1 y) + \frac{i}{\kappa_1} M_0 \sinh(\kappa_1 y) \right] \quad (90)$$

with

$$\kappa_1 = \sqrt{\eta_1^2 - \epsilon|\sigma_1|^2}. \quad (91)$$

Then equations (28), (29) hold as before, and due to (15), (16), we find the potential

$$q(x) = \epsilon r^*(x) = \frac{2i\theta(x)\sigma_1 \exp[-2i\xi_1 x]}{\cosh(2\kappa_1 x) - (\eta_1/\kappa_1) \sinh(2\kappa_1 x)}. \quad (92)$$

Obviously, for $\epsilon = -1$ the potential is nonsingular since we will have $|\eta_1| < |\kappa_1|$. Without loss of generality, we may take the branch in (91) to be positive, i.e., $\kappa_1 > 0$, in which case it is a soliton truncated at $x = 0$. If $\eta_1 > 0$ then there is a maximum of $|q|$ at a point $x = x_0 > 0$. If $\eta_1 < 0$ the untruncated soliton has its maximum value at a point $x = x_0 < 0$. We note that, quite generally, a pole of $\sigma(\zeta)$ in the lower half-plane corresponds to a soliton with its center being in the *unphysical* region. This has been called a *virtual* soliton [9]. It is easy to see that the eigenvalues of the matrix M in (89) are $\xi_1 \pm i\kappa_1$, and that these are just the eigenvalues of the completed soliton.

For $\epsilon = +1$ we must introduce the constraint $|\sigma_1| < |\eta_1|$ in order to have real κ_1 . Then for $\eta_1 < 0$, the potential is nonsingular due to the θ function. However, the untruncated potential will have a singularity in the *unphysical* region, $x < 0$. If $\eta_1 > 0$, then we will have a singular soliton.

We note that in all three examples, cf equations (81), (88), (92), it holds that the potential is discontinuous at the truncation point and the discontinuity is just the residue, σ_1 , multiplied by $2i$ [12], i.e.,

$$q(+0) = 2i\sigma_1. \quad (93)$$

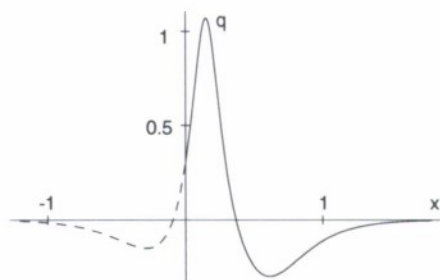


Figure 1. Example 4. Truncated two-soliton potential $q(x)$. The truncated left-hand part is indicated by the dashed line. $C = -1.5i$, $\zeta_1 = 2i$, $\zeta_2 = \lambda_2 = 3i$, $\lambda_1 = 2.5i$.

Example 4. Choosing the reduction $r = -q^*$, $\bar{\sigma}(\zeta) = \sigma^*(\zeta^*)$ we take

$$\sigma(\zeta) = C \frac{\zeta - \zeta_2^*}{(\zeta - \zeta_1)(\zeta - \zeta_2)} \equiv \frac{\sigma_1}{\zeta - \zeta_1} + \frac{\sigma_2}{\zeta - \zeta_2} \quad (94)$$

with

$$\sigma_1 = C \frac{\zeta_2 - \zeta_1^*}{\zeta_1 - \zeta_2}, \quad \sigma_2 = C \frac{\zeta_2 - \zeta_2^*}{\zeta_1 - \zeta_2}. \quad (95)$$

It is the particular feature of this example that σ does contain a factor $(\zeta - \zeta_2^*)/(\zeta - \zeta_2)$ which is canceled in the product $\sigma\bar{\sigma}$ by the reverse factor contained in $\bar{\sigma}$. Now we specify C , ζ_1 , ζ_2 as purely imaginary, $\zeta_1 = i\eta_1$, $\zeta_2 = i\eta_2$. Then the matrix M , cf (19), is

$$M = \begin{pmatrix} i\eta_1 & 0 & \sigma_1^* & \sigma_2^* \\ 0 & i\eta_2 & \sigma_1^* & \sigma_2^* \\ -\sigma_1 & -\sigma_2 & -i\eta_1 & 0 \\ -\sigma_1 & -\sigma_2 & 0 & -i\eta_2 \end{pmatrix}. \quad (96)$$

Further specification $C = -1.5i$, $\eta_1 = 2$, $\eta_2 = 3$ leads to $\sigma_1 = 7.5i$, $\sigma_2 = -9i$. Then the eigenvalues of the matrix M are found to be $\lambda_1 = 2.5i$, $\lambda_2 = 3i$, $\bar{\lambda}_1 = -2.5i$, $\bar{\lambda}_2 = -3i$. Steps (2) and (3) of the procedure are done by a Mathematica program which directly yields the real potential

$$q(x) = -330 \frac{e^{5x}(100 - 108e^x - 12e^{11x} + 9e^{12x})}{900 + 12100e^{10x} - 21600e^{11x} + 9801e^{12x} + 9e^{22x}}, \quad x > 0 \quad (97)$$

as is depicted in figure 1. This is an example where one of the eigenvalues of the truncated potential coincides with an eigenvalue of the completed two-soliton potential, cf, the discussion in the final part of subsection 2.4.

Example 5. Now we assume $r = q^*$ and take $\zeta_1 = -2i$, $\zeta_2 = -4i$, both in $\mathbb{C}_-$, $\sigma_1 = \bar{\sigma}_1 = i$, $\sigma_2 = \bar{\sigma}_2 = -i$. Again the computer program yields a potential in analytic form, as a quotient of sums over exponential functions. The respective formula, however, is not well-suited to be printed. The resulting potential has two singularities in the *unphysical* region $x < 0$ but no

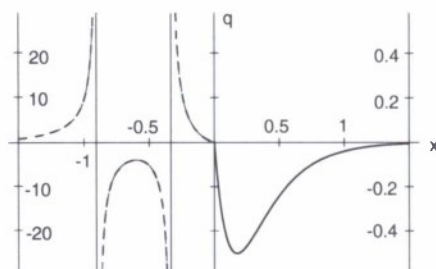


Figure 2. Example 5. Two virtual solitons, $r = q^*$. $\zeta_1 = -2i$, $\zeta_2 = -4i$, $\sigma_1 = \bar{\sigma}_1 = i$, $\sigma_2 = \bar{\sigma}_2 = -i$, $\lambda_1 = 4.040i$, $\lambda_2 = 1.917i$, $\bar{\lambda}_1 = \lambda_1^*$, $\bar{\lambda}_2 = \lambda_2^*$. Note the different scales for the ordinates left and right.

singularities on the right-hand side, see figure 2. This is what one would call a *two-virtual soliton solution*.

7. Summary and discussion

We have developed a method to solve the inverse scattering problem for an AKNS problem with given rational reflection coefficients $\sigma(\zeta) \equiv \bar{b}(\zeta)/a(\zeta)$, $\bar{\sigma}(\zeta) \equiv b(\zeta)/\bar{a}(\zeta)$ on a finite or semi-infinite interval. The requirements for the implementation of this solution are that the matrix exponential function of a constant matrix M , of order $N + \bar{N}$, has to be computed and two matrices, Γ^{11} and Γ^{22} , of orders N and $\bar{N}$ respectively, must be inverted. This must be done for all points at which the solution is desired.

With restriction to so-called minimal symmetry, we have shown that the reflection coefficients are consistent with a *right* potential, i.e., vanishing for $x < 0$, and that a *left* potential can be added so that the combined potential is an N -soliton state. The eigenvalues of the combined potential are just the eigenvalues of the matrix M .

It has been pointed out already in the introduction that the procedure presented here is of importance for solving initial-boundary problems by the use of the inverse scattering method. Indeed such an application already has been done for the second harmonic generation which is mathematically equivalent to degenerating two-photon propagation in the low-excitation limit [10, 11]. There one has to select a number N out of an infinity of poles, and it has been demonstrated in the quoted papers that with growing N the approximation of the solution for an initial-boundary problem becomes better and better. As shown in the above references, one can obtain solutions for N as large as even 60 and depending on the software, one could even go higher. Compared to [10, 11] the solution method here is essentially generalized and presented in a more systematic form so that it is well-suited for further applications.

Acknowledgments

This research has been supported in part by the AFOSR and the NSF grant #DMS0505566. HS would like to thank AS Sakhnovich for fruitful discussions. We are grateful to the referees for valuable comments.

Appendix. The determinant of W

Let us start from some matrix W of a structure as it is given by equation (40) (subsection 2.4). It consists of four block matrices,

$$W = \begin{pmatrix} I_m & B \\ C & I_n \end{pmatrix}, \quad (\text{A.1})$$

such that I_m and I_n are unit matrices of orders m and n , respectively, while B is an $m \times n$ -matrix with m identical rows $(b_1, \dots, b_n)$ and C is an $n \times m$ -matrix with n identical rows $(c_1, \dots, c_m)$. We say that W is of the type (m, n) .

Lemma. *The determinant of this matrix is found as*

$$\det W = 1 - \left(\sum_{j=1}^n b_j \right) \left(\sum_{k=1}^m c_k \right). \quad (\text{A.2})$$

Proof by induction.

- (a) Obviously, the statement is true for $m = n = 1$.
Then we have to prove
- (b) It is true for (m, n) , $n > 1$ provided it holds for $(m, n - 1)$.
- (c) It is true for (m, n) , $m > 1$ provided it holds for $(m - 1, n)$.

In order to prove the claim (b) we take the matrix W and subtract the row immediately above the bottom row from the bottom row, so that the latter then reads $(0, \dots, 0, -1, 1)$. Next we add the right-most column to the one next to it, so that it becomes

$$[(b_{n-1} + b_n), \dots, (b_{n-1} + b_n), 0 \dots 0, 1, 0]^T, \quad (\text{A.3})$$

and the bottom row then is simply $(0, \dots, 0, 1)$. Thus the determinant of W is equal to the determinant of a matrix $\tilde{W}$ of type $(m, n - 1)$ with

$$\tilde{b}_i = b_i, \quad i < n - 1, \quad \tilde{b}_{n-1} = b_{n-1} + b_n, \quad \tilde{c}_k = c_k, \quad (\text{A.4})$$

and point (b) is true simply because $\sum_{j=1}^{n-1} \tilde{b}_j = \sum_{j=1}^n b_j$.

In quite an analogous way the claim (c) can be proven so that the Lemma is true for all natural numbers m, n .

References

- [1] Ablowitz M J and Segur H 1975 *J. Math. Phys.* **16** 1054
- [2] Fokas A S and Ablowitz M J 1989 *Stud. Appl. Math.* **80** 253
- [3] Bikbaev R F and Its A R 1989 *Math. Notes* **45** 349
- [4] Sakhnovich A L 1990 *Ukrainian Math. J.* **42** 316
- [5] Bikbaev R F and Tarasov V O 1991 *J. Phys. A: Math. Gen.* **24** 2507
- [6] Fokas A S and Its A R 1992 *Phys. Rev. Lett.* **68** 3117
- [7] Boutet de Monvel A and Kotlyarov V 2000 *Inverse Problems* **16** 1813
- [8] Fokas A S, Its A R and Sung L Y 2005 *Nonlinearity* **18** 1771
- [9] Steudel H and Kaup D J 1999 *J. Phys. A: Math. Gen.* **32** 6219
- [10] Steudel H and Kaup D J 2000 *J. Phys. A* **33** 1445
- [11] Kaup D J and Steudel H 2001 *Inverse Problems* **17** 959
- [12] Steudel H 2002 *Contemp. Math.* **301** 331–8
- [13] Marchenko V A 1955 *Dokl. Akad. SSSR* **104** 695
- [14] Rourke D E and Morris P G 1992 *Phys. Rev. A* **46** 3631
- [15] Calogero F and Degasperis A 1984 *Physica D* **14** 103

- [16] Ablowitz M J, Kaup D J, Newell A C and Segur H 1974 *Stud. Appl. Math.* **53** 249
- [17] Kay I 1960 *Commun. Pure Appl. Math.* **13** 371
- [18] Kurasov P B 1993 *Inverse Problems in Mathematical Physics* ed L Päiväranta and E Somersalo (Berlin: Springer) pp 126–33
- [19] Aktuson T 2004 *Inverse Problems* **20** 859
- [20] Sakhnovich L A 1994 *St. Petersburg Math. J.* **5** 1
- [21] Alpay D and Gohberg I 1995 *J. Diff. Eqns* **118** 1
- [22] Horvath H 2001 *Transact. Am. Math. Soc.* **353** 4155
- [23] Gohberg I, Kashoek A M and Sakhnovich A L 1998 *J. Diff. Eqns* **146** 375
- [24] Sakhnovich A 2002 *Inverse Problems* **18** 331
- [25] Gohberg I, Kaashoek M A and Sakhnovich A 2002 *Asymptotic Anal.* **29** 1
- [26] Zakharov V E and Shabat A B 1971 *Zh. Eksp. Teor. Fiz.* **61** 118
Zakharov V E and Shabat A B 1972 *Sov. Phys.—JETP* **34** 62
- [27] Leonard I E 1996 *SIAM Rev.* **20** 507
- [28] Rourke D E and Saunders J K 1994 *J. Math. Phys.* **35** 848

Bose-Einstein condensates and multi-component NLS models on symmetric spaces of BD.I-type. Expansions over squared solutions

V. S. Gerdjikov¹, D. J. Kaup², N. A. Kostov¹ and T. I. Valchev¹

¹ *Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences,
72 Tsarigradsko chaussee 1784 Sofia, Bulgaria*

E-mails: gerjikov@inrne.bas.bg, nakostov@inrne.bas.bg, valtchev@inrne.bas.bg,

² *Department of Mathematics, University of Central Florida, Orlando,
Florida 32816-1364, USA*

E-mail: kaup@mail.ucf.edu

A special class of multicomponent NLS equations, generalizing the vector NLS and related to the BD.I-type symmetric are shown to be integrable through the inverse scattering method (ISM). The corresponding fundamental analytic solutions are constructing thus reducing the inverse scattering problem to a Riemann-Hilbert problem. We introduce the minimal sets of scattering data $\mathfrak{T}$ which determines uniquely the scattering matrix and the potential Q of the Lax operator. The elements of $\mathfrak{T}$ can be viewed as the expansion coefficients of Q over the 'squared solutions' that are natural generalizations of the standard exponentials. Thus we demonstrate that the mapping $\mathfrak{T} \rightarrow Q$ is a generalized Fourier transform. Special attention is paid to two special representatives of this MNLS with three-component and five components which describe spinor ($F = 1$ and $F = 2$, respectively) Bose-Einstein condensates.

Keywords: Bose-Einstein condensate, Integrable systems, Recursion operator, squared solutions

1. Introduction

Bose-Einstein condensates (BEC) with the internal degrees of freedom, the so-called spinor BEC have attracted much attention experimentally and theoretically in recent years. Spinor BEC opens up a new paradigm, where the order parameter of condensates is described by a multicomponent vector. This situation can be realized by optically trapping cold atoms where all hyperfine states are liberated, while magnetic trapping freezes its freedom. So far ^{23}Na (the hyperfine state $F = 1$), and ^{87}Rb ($F = 2$) are extensively investigated, see¹⁻¹² and the references therein.

We consider BECs of alkali atoms with hyperfine structure elongated in x direction and confined in the transverse directions y, z by purely optical means. Thus the assembly of atoms in the hyperfine state of spin F can be described by a normalized spinor wave vector with $2F + 1$ components $\Phi(x, t) = (\Phi_F(x, t), \Phi_{F-1}(x, t), \dots, \Phi_{-F}(x, t))^T$ whose components are labeled by the values of $m_F = F, \dots, 1, 0, -1, \dots, -F$.

The main tool for investigating BEC is the Gross-Pitaevski (GPE) equation. In the one-dimensional approximation the GPE system goes into a multicomponent nonlinear Schrödinger (MNLS) equation in 1D x -space:⁶

$$i \frac{\partial \Phi}{\partial t} = \frac{\delta E_{\text{GP}}[\Phi]}{\delta \Phi^*}, \quad (1)$$

where for $F = 1$ the energy functional is given by:

$$E_{\text{GP}} = \int dx \left\{ \frac{\hbar^2}{2m} |\partial_x \Phi|^2 + \bar{c} [|\Phi_1|^4 + |\Phi_{-1}|^4 + 2|\Phi_0|^2(|\Phi_1|^2 + |\Phi_{-1}|^2)] \right. \\ \left. + (\bar{c}_0 - \bar{c}_2)|\Phi_1|^2|\Phi_{-1}|^2 + \frac{\bar{c}_0}{2}|\Phi_0|^4 + \bar{c}_2(\Phi_1^* \Phi_{-1}^* \Phi_0^2 + \Phi_0^{*2} \Phi_1 \Phi_{-1}) \right\} \quad (2)$$

The effective 1D couplings $\bar{c} = (\bar{c}_0 + \bar{c}_2)/2$, $\bar{c}_0 = c_0/2a_\perp^2$, $\bar{c}_2 = c_2/2a_\perp^2$, where $a_\perp$ is the size of the transverse ground state. In this expression, $c_0 = \pi\hbar^2(a_0 + 2a_2)/3m$, $c_2 = \pi\hbar^2(a_2 - a_0)/3m$, where a_f are the s -wave scattering lengths for the channel of total hyperfine spin f and m is the mass of the atom. We consider special (integrable) choice for the coupling constants $\bar{c}_0 = \bar{c}_2 \equiv -c < 0$. This situation corresponds to attractive mean-field interaction and ferromagnetic spin-exchange interaction. We will use dimensionless form: $\Phi \rightarrow \{\Phi_1, \sqrt{2}\Phi_0, \Phi_{-1}\}^T$, where time and length are measured respectively in units of $\bar{t} = \hbar a_\perp / c$ and $\bar{x} = \hbar \sqrt{a_\perp / 2mc}$.

For $F = 2$ the energy functional is defined by^{1-3,5,11}

$$E_{\text{GP}}[\Phi] = \int_{-\infty}^{\infty} dx \left(\frac{\hbar^2}{2m} |\partial_x \Phi|^2 + \frac{\epsilon c_0}{2} n^2 + \frac{c_2}{2} \mathbf{f}^2 + \frac{\epsilon c_4}{2} |\Theta|^2 \right), \quad (3)$$

where $\epsilon = \pm 1$. The number density n and the singlet-pair amplitude Θ are defined by^{3,5,11}

$$n = (\vec{\Phi}, \vec{\Phi}^*) = \sum_{\alpha=-2}^2 \Phi_\alpha \Phi_\alpha^*, \quad \Theta = (\vec{\Phi}, s_0 \vec{\Phi}) = 2\Phi_2 \Phi_{-2} - 2\Phi_1 \Phi_{-1} + \Phi_0^2.$$

The coupling constants c_i are real and can be expressed in terms of a transverse confinement radius and a linear combination of the s -wave scattering lengths of atoms^{6,7,10} and $\mathbf{f}$ describe spin densities.¹¹ Choosing $c_2 = 0$, $c_4 = 1$ and $c_0 = -2$ we obtain an integrable version of the

corresponding GPE equation by means of the inverse scattering transform method.^{6,13}

The aim of present paper is to show that both GPE energy functionals (2), (3) correspond to integrable MNLS models related to symmetric spaces¹⁴ of **BD.I**-type: $\simeq \text{SO}(n+2)/\text{SO}(2) \times \text{SO}(n)$ with $n = 3$ and $n = 5$ respectively. In Section 2 we formulate the Lax representations and the explicit form of MNLS models, generalizing the vector NLS for any n . In Section 3 we construct the fundamental analytic solutions of the corresponding Lax operator L and reduce the inverse scattering problem (ISP) for L to a Riemann-Hilbert problem (RHP). We also introduce the minimal sets of scattering data $\mathfrak{T}_i$ each of which allow one to reconstruct both the scattering matrix $T(\lambda)$ and the corresponding potential $q(x, t)$. In Section 4 we explain that the ISM for this class of Lax operators can also be considered as a generalized Fourier transform. To this end we use the expansions of $q(x)$ and $\text{ad}_J^{-1}\delta q$ over the ‘squared solutions’ of L . In Section 5 we briefly discuss the fundamental properties of these MNLS type equations.

2. MNLS equations for **BD.I** series of symmetric spaces

MNLS equations for the **BD.I** series of symmetric spaces have the Lax representation $[L, M] = 0$ as follows

$$L\psi(x, t, \lambda) \equiv i\partial_x\psi + (q(x, t) - \lambda J)\psi(x, t, \lambda) = 0. \quad (4)$$

$$M\psi(x, t, \lambda) \equiv i\partial_t\psi + (V_0(x, t) + \lambda V_1(x, t) - \lambda^2 J)\psi(x, t, \lambda) = 0, \quad (5)$$

$$V_1(x, t) = q(x, t), \quad V_0(x, t) = i\text{ad}_J^{-1}\frac{dq}{dx} + \frac{1}{2}[\text{ad}_J^{-1}q, q(x, t)]. \quad (6)$$

We consider $n = 2r + 1$ and choose $q(x, t) = \sum_{\alpha \in \Delta_1^+} (q_\alpha E_\alpha + p_\alpha E_{-\alpha})$ where the set of roots $\Delta_1^+ = \{e_1 - e_2, \dots, e_1 - e_r, e_r, e_1 + e_r, \dots, e_1 + e_2\}$. For the typical representation we have the matrix form:

$$q(x, t) = \begin{pmatrix} 0 & \vec{q}^T & 0 \\ \vec{p} & 0 & s_0 \vec{q} \\ 0 & \vec{p}^T s_0 & 0 \end{pmatrix}, \quad J = \text{diag}(1, 0, \dots, 0, -1). \quad (7)$$

The n -component vectors $\vec{q} = (q_2, \dots, q_n)^T$ are formed by the coefficients q_α as follows: $q_k \equiv q_{e_1 - e_k}$, $q_{r+1} \equiv q_{e_1}$ and $q_{n+1-k} \equiv q_{e_1 + e_k}$, $k = 1, \dots, r$; the vector $\vec{p} = (p_2, \dots, p_n)^T$ is formed analogously. The matrix $s_0 = S_0^{(n)}$ enters in the definition of $so(n)$, i.e. $X \in so(n)$, if $X + S_0^{(n)} X^T S_0^{(n)} = 0$, and for $n = 2r + 1$:

$$S_0^{(n)} = \sum_{s=1}^{n+1} (-1)^{s+1} E_{s, n+1-s}^{(n)}, \quad (8)$$

With this definition of orthogonality the Cartan subalgebra generators are represented by diagonal matrices. By $E_{sp}^{(n)}$ above we mean $n \times n$ matrix whose matrix elements are $(E_{sp}^{(n)})_{ij} = \delta_{si}\delta_{pj}$. In terms of these notations the generic MNLS type equations connected to **BD.I** acquire the form

$$\begin{aligned} i\vec{q}_t + \vec{q}_{xx} + 2(\vec{q}, \vec{p})\vec{q} - (\vec{q}, s_0\vec{q})s_0\vec{p} &= 0, \\ i\vec{p}_t - \vec{p}_{xx} - 2(\vec{q}, \vec{p})\vec{p} + (\vec{p}, s_0\vec{p})s_0\vec{q} &= 0, \end{aligned} \quad (9)$$

This equation allows the typical reduction $\vec{p} = \vec{q}^*$. The Hamiltonian for these MNLS equations (9) is given by

$$H_{\text{MNLS}} = \int_{-\infty}^{\infty} dx \left((\partial_x \vec{q}, \partial_x \vec{q}^*) - (\vec{q}, \vec{q}^*)^2 + (\vec{q}, s_0\vec{q})(\vec{q}^*, s_0\vec{q}^*) \right), \quad (10)$$

For $r = 2$ we introduce the variables $\Phi_1 = q_2$, $\Phi_0 = q_3/\sqrt{2}$, $\Phi_{-1} = q_4$; for $r = 3$ we set $\Phi_2 = q_2$, $\Phi_1 = q_3$, $\Phi_0 = q_4$, $\Phi_{-1} = q_5$ and $\Phi_{-2} = q_6$. This reproduces the action functionals E_{GP} for $F = 1$ and $F = 2$.

3. The Inverse scattering problem.

Solving the direct and the inverse scattering problem (ISP) for L uses the Jost solutions which are defined by, see¹⁶ and the references therein

$$\lim_{x \rightarrow -\infty} \phi(x, t, \lambda) e^{i\lambda J x} = \mathbf{1}, \quad \lim_{x \rightarrow \infty} \psi(x, t, \lambda) e^{i\lambda J x} = \mathbf{1} \quad (11)$$

and the scattering matrix $T(\lambda, t) \equiv \psi^{-1}\phi(x, t, \lambda)$. The choice of J and the fact that the Jost solutions and $T(\lambda, t)$ take values in the group $SO(n+2)$ means that we can use the following block-matrix structure of $T(\lambda, t)$

$$T(\lambda, t) = \begin{pmatrix} m_1^+ & -\vec{b}^{-T} & c_1^- \\ \vec{b}^+ & \mathbf{T}_{22} & -s_0\vec{B}^- \\ c_1^+ & \vec{B}^{+T}s_0 & m_1^- \end{pmatrix}, \quad \hat{T}(\lambda, t) = \begin{pmatrix} m_1^- & \vec{b}^{-T} & c_1^- \\ -\vec{B}^+ & \hat{\mathbf{T}}_{22} & s_0\vec{b}^- \\ c_1^+ & -\vec{b}^{+T}s_0 & m_1^+ \end{pmatrix}, \quad (12)$$

where $\vec{b}^\pm(\lambda, t)$ and $\vec{B}^\pm(\lambda, t)$ are n -component vectors, $\mathbf{T}_{22}(\lambda)$ and $\mathbf{m}^\pm(\lambda)$ are $n \times n$ block matrices, and $m_1^\pm(\lambda)$, $c_1^\pm(\lambda)$ are scalars. Such parametrization is compatible with the generalized Gauss decompositions of $T(\lambda)$.

$$T(\lambda, t) = T_J^- D_J^+ \hat{S}_J^+ = T_J^+ D_J^- \hat{S}_J^-, \quad (13)$$

$$T_J^- = \begin{pmatrix} 1 & 0 & 0 \\ \vec{\rho}^+ & \mathbf{1} & 0 \\ c_1^{+,+} & \vec{\rho}^{+,T}s_0 & 1 \end{pmatrix}, \quad S_J^+ = \begin{pmatrix} 1 & \vec{\tau}^{+,T} & c_1^{+,-} \\ 0 & \mathbf{1} & s_0\vec{\tau}^+ \\ 0 & 0 & 1 \end{pmatrix}, \quad (14)$$

$$S_J^-(\lambda) = \left(\hat{S}_J^+(\lambda^*) \right)^\dagger, \quad T_J^+(\lambda) = \left(\hat{T}_J^-(\lambda^*) \right)^\dagger, \quad D_J^-(\lambda) = \left(\hat{D}_J^+(\lambda^*) \right)^\dagger,$$

$$c_1^{\prime\prime,\pm} = \frac{1}{2}(\bar{\rho}^{\pm,T} s_0 \bar{\rho}^{\pm}), \quad c_1^{\prime,\pm} = \frac{1}{2}(\bar{\tau}^{\mp,T} s_0 \bar{\tau}^{\mp}) \quad (15)$$

where the block-diagonal matrix $D_J^+ = \text{b-diag}(m_1^+, m_2^+, 1/m_1^+)$.

$$\bar{\rho}^+ = \frac{\bar{b}^+}{m_1^+}, \quad \bar{\tau}^+ = \frac{\bar{b}^-}{m_1^+}, \quad (16)$$

The matrix elements of $T(\lambda)$ satisfy a number of relations which ensure that both $T(\lambda)$ and its inverse $\hat{T}(\lambda)$ belong to $SO(n+2)$ and that $T(\lambda)\hat{T}(\lambda) = \mathbb{1}$. Some of them take the form:

$$m_1^+ m_1^- + (\bar{b}^-, \bar{B}^+) + c_1^+ c_1^- = 1, \quad \bar{b}^+ \bar{B}^{-T} + \mathbf{T}_{22} s_0 \mathbf{T}_{22}^T s_0 + s_0 \bar{B}^- \bar{b}^{+T} s_0 = \mathbb{1}.$$

Important tools for reducing the ISP to a Riemann-Hilbert problem (RHP) are the fundamental analytic solution (FAS) $\chi^\pm(x, t, \lambda)$. Their construction is based on the generalized Gauss decomposition of $T(\lambda, t)$, see:¹⁷⁻¹⁹

$$\chi^\pm(x, t, \lambda) = \phi(x, t, \lambda) S_J^\pm(t, \lambda) = \psi(x, t, \lambda) T_J^\mp(t, \lambda) D_J^\pm(\lambda). \quad (17)$$

If $q(x, t)$ evolves according to (9) then the scattering matrix and its elements satisfy the following linear evolution equations

$$i \frac{d\bar{\rho}^+}{dt} + \lambda^2 \bar{\rho}^+(t, \lambda) = 0, \quad i \frac{d\bar{\tau}^+}{dt} - \lambda^2 \bar{\tau}^+(t, \lambda) = 0, \quad i \frac{dD^+}{dt} = 0, \quad (18)$$

so the block-diagonal matrices $D^\pm(\lambda)$ can be considered as generating functionals of the integrals of motion. The fact that all $(2r-1)^2$ matrix elements of $m_2^\pm(\lambda)$ for $\lambda \in \mathbb{C}_\pm$ generate integrals of motion reflect the superintegrability of the model and are due to the degeneracy of the dispersion law of (9). We remind that $D_J^\pm(\lambda)$ allow analytic extension for $\lambda \in \mathbb{C}_\pm$ and that their zeroes and poles determine the discrete eigenvalues of L .

The FAS for real λ are linearly related

$$\chi^+(x, t, \lambda) = \chi^-(x, t, \lambda) G_{0,J}(\lambda, t), \quad G_{0,J}(\lambda, t) = \hat{S}_J^-(\lambda, t) S_J^+(\lambda, t). \quad (19)$$

and satisfy the normalization relation $\lim_{\lambda \rightarrow \infty} \chi^\pm(x, t, \lambda) e^{i\lambda Jx} = \mathbb{1}$. Then these FAS satisfy

$$\xi^+(x, t, \lambda) = \xi^-(x, t, \lambda) G_J(x, \lambda, t), \quad G_J(x, \lambda, t) = e^{-i\lambda Jx} G_{0,J}^-(\lambda, t) e^{i\lambda Jx}. \quad (20)$$

Obviously the sewing function $G_J(x, \lambda, t)$ is uniquely determined by the Gauss factors $S_J^\pm(\lambda, t)$.

Lemma 3.1. *Let the potential $q(x, t)$ is such that the Lax operator L has no discrete eigenvalues. Then as minimal set of scattering data which determines uniquely the scattering matrix $T(\lambda, t)$ and the corresponding potential*

$q(x, t)$ one can consider either one of the sets $\mathfrak{T}_i$, $i = 1, 2$

$$\mathfrak{T}_1 \equiv \{\rho_\alpha^+(\lambda, t), \quad \alpha \in \Delta_1^+\}, \quad \mathfrak{T}_2 \equiv \{\tau_\alpha^+(\lambda, t), \quad \alpha \in \Delta_1^+\}, \quad \lambda \in \mathbb{R}.$$

Given the solutions $\chi^\pm(x, t, \lambda)$ one recovers $q(x, t)$ via the formula

$$q(x, t) = \lim_{\lambda \rightarrow \infty} \lambda (J - \chi^\pm J \widehat{\chi}^\pm(x, t, \lambda)). \quad (21)$$

The main goal of the dressing method^{17,20-22} is, starting from a known solutions $\chi_0^\pm(x, t, \lambda)$ of $L_0(\lambda)$ with potential $q_{(0)}(x, t)$ to construct new singular solutions $\chi_1^\pm(x, t, \lambda)$ of L with a potential $q_{(1)}(x, t)$ with two (or more) additional singularities located at prescribed positions $\lambda_1^\pm$; the reduction $\bar{p} = \bar{q}^*$ ensures that $\lambda_1^- = (\lambda_1^+)^*$. It is related to the regular one by a dressing factor $u(x, t, \lambda)$, for details see.¹³

4. The Generalized Fourier Transforms for Non-regular J

It is known that the ‘squared solutions’ $e_\alpha^\pm(x, \lambda) = \pi_{0J}(\chi^\pm E_\alpha \chi^\pm(x, t, \lambda))$, where $\pi_{0J} \equiv \text{ad}_J^{-1} \text{ad}_J$, form complete set of functions in the space of allowed potentials $q(x)$, see.^{16,18,19} Skipping the details we write down the expansions of $q(x)$ and $\text{ad}_J^{-1} \delta q(x)$ assuming L has no discrete spectrum:

$$q(x) = -\frac{i}{\pi} \int_{-\infty}^{\infty} d\lambda \sum_{\alpha \in \Delta_1^+} (\tau_\alpha^+(\lambda) e_\alpha^+(x, \lambda) - \tau_\alpha^-(\lambda) e_\alpha^-(x, \lambda)), \quad (22)$$

$$\text{ad}_J^{-1} \delta q(x) = \frac{i}{\pi} \int_{-\infty}^{\infty} d\lambda \sum_{\alpha \in \Delta_1^+} (\delta \tau_\alpha^+(\lambda) e_\alpha^+(x, \lambda) + \delta \tau_\alpha^-(\lambda) e_\alpha^-(x, \lambda)). \quad (23)$$

These expansions can be viewed as tool to establish the one-to-one correspondence between $q(x)$ (resp. $\text{ad}_J^{-1} \delta q$) and each of the minimal sets of scattering data $\mathcal{T}_i$ (resp. $\delta \mathcal{T}_i$), $i = 1, 2$. To complete the analogy between the standard Fourier transform and the expansions over the ‘squared solutions’ we need the generating operators $\Lambda_\pm$:

$$\Lambda_\pm X(x) \equiv \text{ad}_J^{-1} \left(i \frac{dX}{dx} + i \left[q(x), \int_{\pm\infty}^x dy [q(y), X(y)] \right] \right). \quad (24)$$

for which the ‘squared solutions’ are eigenfunctions:

$$\begin{aligned} (\Lambda_+ - \lambda) e_{-\alpha}^+(x, \lambda) &= 0, & (\Lambda_+ - \lambda) e_{-\alpha}^-(x, \lambda) &= 0, \\ (\Lambda_- - \lambda) e_{\alpha}^+(x, \lambda) &= 0, & (\Lambda_- - \lambda) e_{\alpha}^-(x, \lambda) &= 0. \end{aligned} \quad (25)$$

5. Fundamental properties of the MNLS equations

The expansions (22), (23) and the explicit form of $\Lambda_{\pm}$ and eq. (25) are basic for deriving the fundamental properties of all MNLS type equations related to the Lax operator L . Each of these NLEE is determined by its dispersion law which we choose to be of the form $F(\lambda) = f(\lambda)J$, where $f(\lambda)$ is polynomial in λ . The corresponding NLEE becomes:

$$i \operatorname{ad} J^{-1} q_t + f(\Lambda_{\pm}) q(x, t) = 0. \quad (26)$$

Theorem 5.1. *The NLEE (26) are equivalent to: i) the equations (18) and ii) the following evolution equations for the generalized Gauss factors of $T(\lambda)$:*

$$i \frac{dS_J^+}{dt} + [F(\lambda), S_J^+] = 0, \quad i \frac{dT_J^-}{dt} + [F(\lambda), T_J^-] = 0, \quad \frac{dD_J^+}{dt} = 0. \quad (27)$$

The principal series of integrals is generated by the asymptotic expansion of $\ln m_1^+(\lambda) = \sum_{k=1}^{\infty} I_k \lambda^{-k}$. The first three integrals of motion:

$$I_1 = -\frac{i}{2} \int_{-\infty}^{\infty} dx \langle q(x), q(x) \rangle, \quad I_2 = \frac{1}{2} \int_{-\infty}^{\infty} dx \langle q_x(x), \operatorname{ad} J^{-1} q(x) \rangle, \quad (28)$$

Now iI_1 can be interpreted as the density of the particles, I_2 is the momentum and $I_3 = iH_{\text{MNLS}}$. Indeed, the Hamiltonian equations of motion provided by $H_{(0)} = -iI_3$ with the Poisson brackets

$$\{q_k(y, t), p_j(x, t)\} = i\delta_{kj}\delta(x - y), \quad (29)$$

coincide with the MNLS equations (9). The above Poisson brackets are dual to the canonical symplectic form:

$$\Omega_0 = i \int_{-\infty}^{\infty} dx \operatorname{tr} \left(\delta \bar{p}(x) \wedge \delta \bar{q}(x) \right) = \frac{1}{2i} \left[\left[\operatorname{ad} J^{-1} \delta q(x) \wedge \operatorname{ad} J^{-1} \delta q(x) \right] \right],$$

where $\wedge$ means that taking the scalar or matrix product we exchange the usual product of the matrix elements by wedge-product.

The Hamiltonian formulation of eq. (9) with Ω_0 and H_0 is just one member of the hierarchy of Hamiltonian formulations provided by:

$$\Omega_k = \frac{1}{i} \left[\left[\operatorname{ad} J^{-1} \delta Q \wedge \Lambda^k \operatorname{ad} J^{-1} \delta Q \right] \right], \quad H_k = i^{k+3} I_{k+3}. \quad (30)$$

where $\Lambda = \frac{1}{2}(\Lambda_+ + \Lambda_-)$. We can also calculate Ω_k in terms of the scattering data variations. Imposing the reduction $q(x) = q^\dagger(x)$ we get:

$$\begin{aligned} \Omega_k &= \frac{1}{2\pi i} \int_{-\infty}^{\infty} d\lambda \lambda^k (\Omega_0^+(\lambda) - \Omega_0^-(\lambda)) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\lambda \lambda^k \operatorname{Im} \left(m_1^+(\lambda) \left(\hat{m}_2^+ \delta \bar{\rho}^+(\lambda) \wedge \delta \bar{\tau}^+(\lambda) \right) \right). \end{aligned}$$

This allows one to prove that if we are able to cast Ω_0 in canonical form, then all Ω_k will also be cast in canonical form and will be pair-wise equivalent.

Acknowledgments

One of us (V. S. G.) thanks the organizers of the Conference for their hospitality and for making his participation possible. This research has been supported in part by the National Science Foundation of the USA via grant DMS-0505566 and by the USA Air Force Office of Scientific Research.

References

1. T. Ohmi and K. Machida, *Bose-Einstein condensation with internal degrees of freedom in alkali atom gases* J. Phys. Soc. Jpn. **67** (1998) 1822.
2. T.-L. Ho, *Spinor Bose condensates in optical traps*, Phys. Rev. Lett. **81** (1998) 742.
3. C. V. Ciobanu, S.-K. Yip and T.-L. Ho, *Phase diagrams of $F = 2$ spinor Bose-Einstein condensates* Phys. Rev. A **61** (2000) 033607.
4. Nille N. Klausen, John L. Bohn and Chris H. Greene, *Nature of spinor Bose-Einstein condensates in rubidium*, Phys. Rev. A **64**, 053602 (2001).
5. M. Ueda and M. Koashi, *Theory of spin-2 Bose-Einstein condensates: Spin correlations, magnetic response, and excitation spectra*, Phys. Rev. A **65** (2002) 063602.
6. J. Ieda, T. Miyakawa and M. Wadati, *Exact Analysis of Soliton Dynamics in Spinor Bose-Einstein Condensates*, Phys. Rev Lett. **93**, (2004), 194102.
7. J. Ieda, T. Miyakawa and M. Wadati, *Matter-wave solitons in an $F = 1$ spinor Bose-Einstein condensate*, J. Phys. Soc. Jpn. **73** (2004) 2996.
8. Kuwamoto, T., Araki, K., Eno, T., Hirano, T. *Magnetic field dependence of the dynamics of ^{87}Rb spin-2 Bose-Einstein condensates*. Physical Review A, vol. 69, (2004) 063604
9. L. Li, Z. Li, B. A. Malomed, D. Mihalache and W. M. Liu, *Exact Soliton Solutions and Nonlinear Modulation Instability in Spinor Bose-Einstein Condensates*, Phys. Rev. A **72**, (2005) 033611.
10. M. Uchiyama, J. Ieda and M. Wadati, *Dark solitons in $F = 1$ spinor Bose-Einstein condensate* J. Phys. Soc. Jpn. **75** (2006) 064002.
11. M. Uchiyama, J. Ieda, and M. Wadati, *Multicomponent Bright Solitons in $F = 2$ Spinor Bose-Einstein Condensates*, J. Phys. Soc. Japan, **76**, No. 7, (2007), 74005.
12. Uchino, Shun, Otsuka, Takaharu, Ueda, Masahito. *Dynamical symmetry in spinor Bose-Einstein condensates* **arXiv:0710.5210**.
13. V. S. Gerdjikov, N. A. Kostov, T. I. Valchev. *Solutions of multi-component NLS models and Spinor Bose-Einstein condensates*, Physica D (2008) (In press); **ArXiv:0802.4398 [nlin.SI]**.
14. A. P. Fordy and P. P. Kulish, *Nonlinear Schrodinger Equations and Simple Lie Algebras*, Commun. Math. Phys. **89** (1983) 427-443.

15. S. Helgasson, *Differential Geometry, Lie Groups and Symmetric Spaces*, (Graduate studies in Mathematics, vol.34), AMS, Providence, Rhode Island (2001).
16. V. S. Gerdjikov. *Basic Aspects of Soliton Theory*. In: Eds.: I. M. Mladenov, A. C. Hirshfeld. "Geometry, Integrability and Quantization", pp. 78-125; Softex, Sofia 2005. **nlin.SI/0604004**
17. V. E. Zakharov, S. V. Manakov, S. P. Novikov and L. I. Pitaevskii, *Theory of Solitons. The Inverse Scattering Method*, Plenum Press (Consultant Bureau), N.Y., (1984).
18. Gerdjikov V. S. *Generalized Fourier Transforms for the Soliton Equations. Gauge Covariant Formulation*. Inverse Problems **2**, n. 1 (1986) 51-74.
19. Gerdjikov V. S. *Complete Integrability, Gauge Equivalence and Lax Representations of the Inhomogeneous Nonlinear Evolution Equations*. Theor. Math. Phys. **92** (1992) 374-386.
20. V. S. Gerdjikov, G. G. Grahovski, R. I. Ivanov and N. A. Kostov, *N-wave interactions related to simple Lie algebras. Z_2 - reductions and Soliton Solutions*, Inv. Problems **17** (2001) 999-1015.
21. R. I. Ivanov, *On the dressing method for the generalized Zakharov-Shabat system*, Nucl. Phys. **B 694**, (2004) 509-524.
22. G. G. Grahovski, V. S. Gerdjikov, N. A. Kostov, V. A. Atanasov, *New Integrable Multi-component NLS type Equations on Symmetric Spaces: Z_4 and Z_6 reductions*, In "Geometry, Integrability and Quantization VII", Eds: I. Mladenov and M. De Leon, Softex, Sofia (2006), pp. 154-175.

On classification of soliton solutions of multicomponent nonlinear evolution equations

V S Gerdjikov¹, D J Kaup², N A Kostov¹ and T I Valchev¹

¹ Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences,
72 Tsarigradsko chaussee, 1784 Sofia, Bulgaria

² Department of Mathematics, University of Central Florida, Orlando, FL 32816-1364, USA

E-mail: gerdjikov@inrne.bas.bg, kaup@mail.ucf.edu and valtchev@inrne.bas.bg

Received 21 April 2008, in final form 12 June 2008

Published 4 July 2008

Online at stacks.iop.org/JPhysA/41/315213

Abstract

We consider several ways of how one could classify the various types of soliton solutions related to multicomponent nonlinear evolution equations which are solvable by the inverse scattering method for the generalized Zakharov–Shabat system related to a simple Lie algebra g . In doing so we make use of the fundamental analytic solutions, the Zakharov–Shabat dressing procedure, the reduction technique and other tools characteristic for that method. The multicomponent solitons are characterized by several important factors: the subalgebras of g and the way these subalgebras are embedded in g , the dimension of the corresponding eigensubspaces of the Lax operator L , as well as by additional constraints imposed by reductions.

PACS numbers: 05.45.Yv, 02.30.Zz, 02.20.Sv, 02.30.Ik

1. Introduction

Since the time when the inverse scattering transform was invented the number of integrable nonlinear evolution equations (NLEE) has been increasing significantly [1–3]. There exist different approaches in analyzing these equations: constructing the spectral theory of the so-called recursion operators from Lax scattering operators, studying whether there exist higher order integrals of motion and so on.

Along with the number of integrable equations the diversity of ‘species’ of soliton solutions has bloomed up: breathers, topological solitons, tracons, boomerons, etc. Thus the analogous problem of classifying solutions to nonlinear evolution equations and soliton solutions, in particular, becomes more and more important. It is our opinion that the mentioned problem is still waiting for its solution. Even for some of the best known NLEE such as the N -wave equations, the multicomponent nonlinear Schrödinger equations, multicomponent modified

KdV equations, etc only the simplest types of soliton solutions are known. Our aim is to outline several criteria for the classification of the soliton solutions.

We use the term 'soliton solution' as a special solution to a given NLEE which is solvable by the inverse scattering method [1–3], i.e. it allows a zero curvature representation

$$[L(\lambda), M(\lambda)] = 0, \quad (1)$$

where $L(\lambda)$ and $M(\lambda)$ are two linear differential operators. In what follows we take them to be first-order matrix differential operators,

$$L\psi(x, t, \lambda) \equiv i\partial_x \psi + U(x, t, \lambda)\psi(x, t, \lambda),$$

$$M\psi(x, t, \lambda) \equiv i\partial_t \psi + V(x, t, \lambda)\psi(x, t, \lambda).$$

The soliton solutions are related to a set of several discrete eigenvalues of the Lax operator L . Therefore one first has to describe all configurations of discrete eigenvalues of L , see [4]. The next step in classifying the types of one-soliton solutions is related to the study of their internal degrees of freedom.

In order to make the problem not too difficult we will specify L to be the operator for the generalized Zakharov–Shabat system

$$L(x)\psi(x, \lambda) \equiv i\partial_x \psi + (q(x) - \lambda J)\psi(x, \lambda) = 0,$$

where we take the potential $q(x, t)$ to be an $n \times n$ matrix-valued smooth function of x tending to zero sufficiently rapid as $x \rightarrow \pm\infty$. We also restrict J to be a real constant diagonal matrix with different eigenvalues.

For simplicity we consider the class of Lax operators of Zakharov–Shabat type in most cases with real-valued J . In doing this we will be using the dressing method [5–7], one of the best known methods for constructing reflectionless potentials and soliton solutions. This paper is intended as a natural continuation of the work [8] published several years ago by two of the authors.

In section 2, we outline preliminary facts about the generalized Zakharov–Shabat systems related to the $sl(n)$ algebras. We also review the basic facts for the $sl(2)$ soliton solutions, and especially the topological and the breather solutions.

In section 3, we treat the different types of one-soliton solutions for the $sl(n)$ Zakharov–Shabat systems starting with $n = 3$. Along with the well-known soliton solutions obtained by Zakharov and Manakov [9] and Kaup [10] we show that the 3-wave equations with a $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry allow doublet and quadruplet soliton solutions (analogies to the topological and breather solutions mentioned above). Next we analyze N -wave equations related to $sl(n)$ algebra with $n = 5$; of course nearly all formulae can easily be generalized for any n . Here we use generic projectors of rank $s \geq 1$ and explain why it is enough to consider only $s \leq [n/2]$ where $[n/2]$ is the entire part of $n/2$. We also demonstrate that using special choices for the polarization vectors defining the projector we can get one-soliton solutions $q_{1s}(x, t)$ taking values in a subalgebra of $sl(n)$. The simplest nontrivial type of solitons can be related to the subalgebra $sl(2)$; we call them typical $sl(2)$ solitons. Our next observation is that the $sl(2)$ subalgebra can be embedded in several different inequivalent ways into $sl(5)$. For example we can embed a spin J representation of $sl(2)$, $J \leq 2$ into $sl(5)$; we call them spin J $sl(2)$ solitons. The corresponding construction is done by using the corresponding symmetrized tensor products of the $sl(2)$ dressing factor. The next step is to consider typical $sl(3)$ and $sl(4)$ solitons. Of course, like for the $sl(3)$ algebra, using additional $\mathbb{Z}_2$ -symmetries one can obtain N -wave-type equations allowing doublet and quadruplet solitons.

In section 4, we analyze the structure of the eigenfunctions of $L(\lambda)$ corresponding to the different types of solitons.

In section 5, we discuss the effects of several types of reductions [11] on the different types of one-soliton solutions. The first $\mathbb{Z}_2$ -reduction does not affect the spectral parameter λ and is produced using the exterior automorphism of $sl(5)$. Its effect is that now $q(x, t) - \lambda J$ takes values in the subalgebra $so(5) \subset sl(5)$. Here we briefly recall the soliton solutions of the $so(5)$ 4-wave equations we derived in [12]. Next we apply an additional pair of $\mathbb{Z}_2 \times \mathbb{Z}_2$ reductions we obtain a 4-wave system allowing doublet and quadruplet solutions. At the end of this section we analyze the NLEE related to the symmetric spaces [13–15] of **BD.I** type; the corresponding J in the Lax pair is dual to the vector e_1 in the root space. One easily checks that for such J the N -wave equations become linear. However there are multicomponent mKdV-type equations which may be called vector mKdV equations. Such equations have been shown to possess higher symmetries and have been related to Jordan algebras [16]. The simplest nontrivial example is a three-component mKdV related to the algebra $so(5)$. We apply an additional $\mathbb{Z}_2$ -reduction on it using an automorphism related to a specific Weyl reflection and obtain a two-component vector mKdV which to the best of our knowledge is new. We end the section by deriving for it the doublet and quadruplet solutions.

In section 6 we summarize our results and discuss their possible generalizations. In the appendices, we collect some technical details about the symmetric spaces of **BD.I** type and about the spin J representations of $sl(2)$.

2. Preliminaries

2.1. The generalized Zakharov–Shabat system related to $sl(n)$

In this section, we shall outline some basic features of the mathematical machinery we are about to use for the classification of soliton solutions.

Integrability or more precisely S-integrability of a NLEE means that the NLEE can be presented as a zero curvature condition

$$[L(\lambda), M(\lambda)] = 0 \quad (2)$$

of two first-order linear matrix differential operators $L(\lambda)$ and $M(\lambda)$ of the form

$$L\psi(x, t, \lambda) \equiv (i\partial_x + U(x, t, \lambda))\psi(x, t, \lambda) = 0, \quad (3)$$

$$M\psi(x, t, \lambda) \equiv (i\partial_t + V(x, t, \lambda))\psi(x, t, \lambda) = \psi(x, t, \lambda)C(\lambda). \quad (4)$$

The potentials $U(x, t, \lambda)$ and $V(x, t, \lambda)$ are typically chosen as elements of some semisimple Lie algebra $\mathfrak{g}$ (the fundamental solutions $\psi(x, t, \lambda)$ belong to the corresponding Lie group G). We shall mainly deal with the algebra $sl(n)$.

Remark 1. The compatibility condition (2) means that the Lax operators L and M possess the same eigenfunctions. The matrix $C(\lambda)$ depends on the definition of Jost solutions.

Since the compatibility condition (2) must hold true identically with respect to λ one can verify that

$$i\partial_x V - i\partial_t U + [U(x, t, \lambda), V(x, t, \lambda)] = 0 \quad (5)$$

and it is valid for any choice of $C(\lambda)$. For simplicity we shall restrict our considerations on scattering operators of the Zakharov–Shabat type (GZS)

$$L(\lambda)\psi(x, t, \lambda) \equiv (i\partial_x + q(x, t) - \lambda J)\psi(x, t, \lambda) = 0. \quad (6)$$

The matrix J is a real traceless diagonal matrix, i.e. a real Cartan element of $sl(n)$, while $q(x, t)$ is a matrix with zero diagonal elements. Since J is a real matrix one can introduce an

ordering of its elements $J_1 > J_2 > \dots > J_n$. By carrying out a gauge transformation which commutes with J , we can always take $q(x, t)$ to be of the form $q(x, t) = [J, Q(x, t)]$, i.e. $q_{jj} \equiv 0$. The linear subspace in $sl(n)$ of matrix-valued functions $q(x, t) = [J, Q(x, t)]$ is known in the literature to be the co-adjoint orbit of g passing through J . The co-adjoint orbits can be supplied in a natural way with a non-degenerate symplectic structure which makes them natural choices for the phase spaces $\mathcal{M}_J$ of the corresponding NLEE.

The class of NLEE related to $L(\lambda)$ are systems of equations for the multicomponent functions $Q(x, t)$, which may be written in the compact form [9, 10, 17, 18]

$$i\partial_t Q + 2 \sum_{k=1}^4 \Lambda^k [H_k, Q(x, t)] = 0, \quad (7)$$

where H_k , $\text{tr } H_k = 0$ are constant diagonal matrices and $f(\lambda) = \sum_{k=1}^4 \lambda^k H_k$ is the dispersion law of the NLEE. Here and below we define

$$(\text{ad}_J X)_{ks} \equiv ([J, X])_{ks} = (J_k - J_s)X_{ks}, \quad (\text{ad}_J^{-1} X)_{ks} = \frac{X_{ks}}{J_k - J_s}, \quad (8)$$

for all $X \in \mathcal{M}_J$, i.e. $X_{kk} = 0$. The operator Λ is either one of the recursion operators $\Lambda_{\pm}$, acting on the space $\mathcal{M}_J$ of $n \times n$ off-diagonal matrix-valued functions as follows:

$$\Lambda_{\pm} X \equiv \text{ad}_J^{-1} \left(i\partial_x X + P_0[q, X] + i \sum_{k=1}^5 [Q, E_{k,k}] \int_{\pm\infty}^x dy \text{tr}([Q, X], E_{k,k}) \right). \quad (9)$$

where P_0 is the projector $\text{ad}_J^{-1} \text{ad}_J (E_{k,k})_{jl} = \partial_{k,j} \partial_{j,l}$. Choosing $H_1 = I = \text{diag}(I_1, \dots, I_n) \neq \mathbb{1}$, so that the dispersion law $f(\lambda) = -\lambda I$ is a linear function of λ we get a system, generalizing the well-known N -wave equation

$$i[J, Q_t] - i[I, Q_x] - [[J, Q], [I, Q]] = 0, \quad (10)$$

which contains $N = n(n-1)$ complex-valued functions $Q_{ij}(x, t)$.

In order to describe the soliton solutions we shall use the so-called dressing procedure [7]. For that purpose we need some basic facts on the direct scattering problem for L .

Let $\psi_{\pm}(x, t, \lambda)$ are two fundamental solutions of the GZS system (6). If they satisfy the requirement

$$\lim_{x \rightarrow \pm\infty} \psi_{\pm}(x, t, \lambda) e^{i\lambda J x} = \mathbb{1} \quad (11)$$

they shall be called Jost solutions. The Jost solutions are interrelated via

$$\psi_{-}(x, t, \lambda) = \psi_{+}(x, t, \lambda) T(t, \lambda), \quad (12)$$

where $T(t, \lambda)$ is called a scattering matrix. The scattering matrix is x -independent and its time evolution is driven by the linear equation

$$i\partial_t T + [f(\lambda), T] = 0 \quad \Rightarrow \quad T(t, \lambda) = e^{if(\lambda)t} T(0, \lambda) e^{-if(\lambda)t}. \quad (13)$$

Equation (13) shows how one can recover the time evolution of the scattering data. It is used to solve Cauchy's problem for NLEE as it is displayed below,

$$q(t=0) \rightarrow L(t=0) \rightarrow T(t=0) \rightarrow T(t) \rightarrow L(t) \rightarrow q(t). \quad (14)$$

Due to this we shall skip the t -dependence of all functions (potentials, fundamental solutions, etc) regarding them at a fixed moment $t = t_0$.

The set of matrix elements of $T(\lambda)$ must satisfy a number of relations. Indeed, they are uniquely determined by $Q(x)$, i.e. by $n(n-1)$ complex functions of x , so it seems natural that there should not be more than $n(n-1)$ independent functions among $T_{jk}(\lambda)$ for λ on

the real axis. Of course, $T(\lambda)$ must satisfy the ‘unitarity’ condition $\det T(\lambda) = 1$. The rest of these relations follow from the analyticity properties of certain combinations of the matrix elements of $T(\lambda)$. These analyticity properties must follow naturally from the corresponding fundamental analytic solutions (FAS) $\chi^\pm(x, \lambda)$.

The Jost solutions are well defined only for $\lambda \in \mathbb{R}$, i.e. they do not have necessarily analytic properties beyond the real axis. This can be easily seen if one reformulates the problem (6) in terms of a Volterra-type integral equation

$$\xi_\pm(x, \lambda) = \mathbb{1} + i \int_{\pm\infty}^x dy e^{i\lambda J(y-x)} q(y) \xi_\pm(y, \lambda) e^{i\lambda J(x-y)}, \quad (15)$$

where $\xi_\pm(x, \lambda) = \psi_\pm(x, \lambda) e^{i\lambda Jx}$ represents another set of fundamental solutions but this time to the linear problem

$$i\partial_x \xi(x, \lambda) + q(x) \xi(x, \lambda) - \lambda[J, \xi(x, \lambda)] = 0.$$

It is easy to see that only the first and the last columns of $\psi_+(x, \lambda)$ and $\psi_-(x, \lambda)$ allow analytic extensions in λ off the real axis; generally the other columns do not have analyticity properties. Nevertheless it is possible to introduce FAS [3, 19]. Taking into account the ordering introduced above one is able to construct new fundamental solutions

$$\xi_{kl}^\pm(x, \lambda) = \begin{cases} \delta_{kl} + i \int_{\pm\infty}^x dy e^{i\lambda(J_k - J_l)(y-x)} (q\xi^\pm)_{kl}(y, \lambda), & k \leq l, \\ i \int_{\mp\infty}^x dy e^{i\lambda(J_k - J_l)(y-x)} (q\xi^\pm)_{kl}(y, \lambda), & k > l \end{cases} \quad (16)$$

to possess analytic properties in the half planes $\mathbb{C}_\pm$ of the spectral parameter. This definition can be rewritten using the Gauss factors of the scattering matrix T

$$\chi^\pm(x, \lambda) = \psi_-(x, \lambda) S^\pm(\lambda) = \psi_+(x, \lambda) T^\mp(\lambda), \quad (17)$$

where $T(\lambda) = T^\mp(\lambda)(S^\pm(\lambda))^{-1}$ and $\chi^\pm(x, \lambda) = \xi^\pm(x, \lambda) e^{-i\lambda Jx}$. The matrix elements of $T^\pm(\lambda)$ and $S^\pm(\lambda)$ can be expressed in terms of the minors of $T(\lambda)$. Here we note that their diagonal elements can be given by

$$S_{jj}^+(\lambda) = m_{j-1}^+(\lambda), \quad T_{jj}^-(\lambda) = m_j^+(\lambda), \quad (18)$$

$$T_{jj}^+(\lambda) = m_{n-j}^-(\lambda), \quad S_{jj}^-(\lambda) = m_{n+1-j}^-(\lambda), \quad (19)$$

where $m_0^\pm = m_n^\pm = 1$ and by $m_k^+(\lambda)$ (resp. $m_k^-(\lambda)$) we have denoted the upper (resp. lower) principal minors of $T(\lambda)$ of order k , e.g.,

$$m_k^+(\lambda) = \begin{Bmatrix} 1 & 2 & \dots & k \\ 1 & 2 & \dots & k \end{Bmatrix}, \quad k = 1, \dots, n \quad (20)$$

$$m_k^-(\lambda) = \begin{Bmatrix} n-k+1 & n-k+2 & \dots & n \\ n-k+1 & n-k+2 & \dots & n \end{Bmatrix}_{T(\lambda)}, \quad (21)$$

$$\begin{Bmatrix} i_1 & i_2 & \dots & i_k \\ j_1 & j_2 & \dots & j_k \end{Bmatrix}_{T(\lambda)} \equiv \det \begin{pmatrix} T_{i_1 j_1} & T_{i_1 j_2} & \dots & T_{i_1 j_k} \\ T_{i_2 j_1} & T_{i_2 j_2} & \dots & T_{i_2 j_k} \\ \vdots & \vdots & \ddots & \vdots \\ T_{i_k j_1} & T_{i_k j_2} & \dots & T_{i_k j_k} \end{pmatrix}. \quad (22)$$

As a consequence of the analyticity of the FAS, it follows that the minors $m_k^+(\lambda)$ (resp. $m_k^-(\lambda)$) are analytic functions for $\lambda \in \mathbb{C}_+$ (resp. for $\lambda \in \mathbb{C}_-$).

One can construct the kernel of the resolvent of $L(\lambda)$ in terms of the FAS [17, 18] from which it follows that the resolvent has poles for all $\lambda_k^\pm$ which happen to be zeroes of any of the minors $m_k^\pm(\lambda)$. Therefore what we have now is that each of the minors $m_k^\pm(\lambda)$ may be considered to be an analog of the Evans function, and thus now, there is more than one Evans function, with each Evans function in a one-to-one correspondence with one of the minors, $m_k^\pm(\lambda)$.

There exist different methods to solve a NLEE possessing a Lax representation: Gel'fand–Levitan–Marchenko integral equation, Hirota method, dressing method, etc. We shall use the dressing Zakharov–Shabat method [7]. Let $\psi_0(x, \lambda)$ be a fundamental solution of Zakharov–Shabat's system with a known potential $U_0(x, \lambda) = q_0(x) - \lambda J$. Consider a new function $\psi(x, \lambda) = u(x, \lambda)\psi_0(x, \lambda)$ which is a solution to Zakharov–Shabat's problem with some potential $q(x) - \lambda J$ to be found. This requires that $u(x, \lambda)$ satisfies

$$i\partial_x u + qu - uq_0 - \lambda[J, u] = 0. \quad (23)$$

The dressing procedure transforms the Jost solutions $\psi_{\pm,0}(x, \lambda)$, the scattering matrix $T_0(\lambda)$ and the fundamental solution $\chi_0^\pm(x, \lambda)$ of the generalized Zakharov–Shabat system with a potential $U_0(x, \lambda)$ in the following fashion:

$$\psi_\pm(x, \lambda) = u(x, \lambda)\psi_{\pm,0}(x, \lambda)u_\pm^{-1}(\lambda), \quad (24)$$

$$T(\lambda) = u_+(\lambda)T_0(\lambda)u_-^{-1}(\lambda), \quad (25)$$

$$\chi^\pm(x, \lambda) = u(x, \lambda)\chi_0^\pm(x, \lambda)u_\pm^{-1}(\lambda). \quad (26)$$

The normalizing factors $u_\pm(\lambda) = \lim_{x \rightarrow \pm\infty} u(x, \lambda)$ ensures the proper asymptotics of the dressed solutions $\psi_\pm(x, \lambda)$.

Our aim is by using the structure of the dressing factor to try to classify the soliton solutions of NLEE. A classical ansatz [3] for $u(x, \lambda)$ is the following one:

$$u(x, \lambda) = \mathbb{I} + (c(\lambda) - 1)P(x), \quad c(\lambda) = \frac{\lambda - \lambda^+}{\lambda - \lambda^-}, \quad (27)$$

where P is a projector ($P^2 = P$) which can be expressed via the fundamental analytic solutions (FAS) and λ^+ (resp. λ^-) is an arbitrary complex number in the upper (resp. lower) half plane $\mathbb{C}_+$ (resp. $\mathbb{C}_-$). After applying the dressing procedure $\lambda^\pm$ become a pair of discrete eigenvalues for the dressed Lax operator. The projector $P(x)$ determines the corresponding eigensubspaces—the rank s of P equals the dimension of the eigensubspaces. If we introduce two sets, each containing s vectors

$$|n_1\rangle, \dots, |n_s\rangle; \quad \langle m_1|, \dots, \langle m_s|,$$

we can write P as follows:

$$P = \sum_{a,b=1}^s |n_a\rangle \hat{M}_{ab} \langle m_b|, \quad M_{ab} = \langle m_a | n_b \rangle, \quad \hat{M} \equiv M^{-1}. \quad (28)$$

Obviously the two sets of vectors are the left and right eigenvectors of P , namely

$$P|n_a\rangle = |n_a\rangle, \quad \langle m_b|P = \langle m_b|, \quad (29)$$

for all $a, b = 1, \dots, s$. In the simplest case when $\text{rank } P = 1$ it reads

$$P(x) = \frac{|n(x)\rangle \langle m(x)|}{\langle m(x) | n(x) \rangle}, \quad (30)$$

where

$$|n(x)\rangle = \chi_0^+(x, \lambda^+) |n_0\rangle, \quad \langle m(x)| = \langle m_0 | (\chi_0^-(x, \lambda^-))^{-1}. \quad (31)$$

By taking the limit $\lambda \rightarrow \infty$ in equation (23) we obtain an interrelation between the seed solution q_0 and the new one

$$q = q_0 + (\lambda^- - \lambda^+) [J, P]. \quad (32)$$

Thus starting from a known solution of the NLEE we can find another solution by simply dressing it with some factor $u(x, \lambda)$. An important particular case is when $q_0 = 0$. The dressed solution is called a one-soliton solution. The fundamental analytic solution in the soliton case is given by a plane wave $\chi_0^\pm(x, \lambda) = \exp(-i\lambda Jx)$. Repeating the same procedure one derives step by step the multisoliton solution of the corresponding NLEE, i.e.,

$$0 \rightarrow q^{1s} \rightarrow q^{2s} \rightarrow \dots \rightarrow q^{ns}. \quad (33)$$

Many integrable equations correspond to Lax operators that obey some additional symmetry conditions of algebraic nature. That is why it is worthwhile to outline some aspects of the theory of such Lax operators.

Let an action of a discrete group G_R to be referred to as a reduction group be given on the set of fundamental solutions to the generalized Zakharov–Shabat system (6) as follows:

$$\tilde{\psi}(x, \lambda) = K \psi(x, k(\lambda)) K^{-1}, \quad (34)$$

where $k: \mathbb{C} \rightarrow \mathbb{C}$ is a conformal map. This action yields another action on the potential in the scattering operator L ,

$$KU(x, k(\lambda)) K^{-1} = U(x, \lambda). \quad (35)$$

A common case is when $G_R = \mathbb{Z}_2$. Then the action of $\mathbb{Z}_2$ might involve external automorphisms of $SL(n)$ as well,

$$\tilde{\psi}(x, \lambda) = K(\psi^T(x, k_1(\lambda)))^{-1} K^{-1} \Rightarrow KU^T(x, k_1(\lambda)) K^{-1} = -U(x, \lambda), \quad (36)$$

$$\tilde{\psi}(x, \lambda) = K\psi^*(x, k_2(\lambda)) K^{-1} \Rightarrow KU^*(x, k_2(\lambda)) K^{-1} = -U(x, \lambda). \quad (37)$$

In particular, if $\mathbb{Z}_2$ acts trivially on the complex plane of the spectral parameter, i.e. $k = id$, then the symmetry condition (36) may restricts the potential $U(x, \lambda)$ to a certain subalgebra of $sl(n)$. For example, suppose $K^T = S_0 \widehat{K} S_0$ then $U(x, \lambda)$ belongs to the orthogonal algebra³ $so(n)$. The existence of a $\mathbb{Z}_2$ reduction requires a modification of the dressing factor $u(x, \lambda)$ as follows:

$$u(x, \lambda) = \mathbb{I} + (c(\lambda) - 1)P(x) + \left(\frac{1}{c(\lambda)} - 1\right)\overline{P}(x), \quad (38)$$

where $P(x)$ is a projector of rank 1,

$$P(x) = \frac{|n(x)\rangle\langle m(x)|}{\langle m(x)|n(x)\rangle}, \quad \overline{P}(x) = K S_0 P^T(x) S_0^{-1} K^{-1}. \quad (39)$$

The projector itself can be expressed through the FAS $\chi^\pm(x, \lambda)$,

$$|n(x)\rangle = \chi_0^+(x, \lambda^+) |n_0\rangle, \quad \langle m(x)| = \langle m_0| (\chi_0^-(x, \lambda^-))^{-1}. \quad (40)$$

³ Here we use a slightly modified definition of the orthogonal algebras, see appendix A where the definition and the explicit form of the matrix S_0 are given.

2.2. The Zakharov–Shabat system and $sl(2)$ solitons

The best known examples of NLEE are related to the Zakharov–Shabat system which is associated with the $sl(2)$ algebra as follows:

$$\begin{aligned} L\psi(x, t, \lambda) &\equiv (i\partial_x + q(x, t) - \lambda\sigma_3)\psi(x, t, \lambda), \\ q(x, t) &= q^+\sigma_+ + q^-\sigma_- = \begin{pmatrix} 0 & q^+ \\ q^- & 0 \end{pmatrix}, \end{aligned} \quad (41)$$

where $\sigma_{\pm} = (\sigma_1 \pm i\sigma_2)/2$ and σ_1, σ_2 and σ_3 are the Pauli matrices.

The class of NLEE for the functions $q^{\pm}(x, t)$ related to (41) can be written in the compact form [20–22],

$$i\sigma_3\partial_t q + 2f(\lambda)q(x, t) = 0, \quad (42)$$

where $f(\lambda)$ is the dispersion law of the NLEE and Λ is one of the recursion operators, acting on the space $\mathcal{M}_0$ of off-diagonal matrix-valued functions as follows:

$$\Lambda_{\pm} X \equiv \frac{i}{4} [\sigma_3, \partial_x X] + \frac{i}{2} q(x) \int_{\pm\infty}^x dy \operatorname{tr}(q(y), [\sigma_3, X(y)]). \quad (43)$$

The simplest nontrivial example of NLEE is related to a dispersion law of the type $f(\lambda) = -2\lambda^2$. This is the nonlinear Schrödinger equation

$$\begin{aligned} iq_t^+ + q_{xx}^+ + 2(q^+(x, t))^2 q^-(x, t) &= 0, \\ iq_t^- - q_{xx}^- - 2(q^-(x, t))^2 q^+(x, t) &= 0. \end{aligned} \quad (44)$$

Another well-known example is provided by a cubic dispersion law $f(\lambda) = 4\lambda^3$, one gets the system

$$\begin{aligned} q_t^+ + q_{xxx}^+ + 6q^+(x, t)q^-(x, t)q_x^+ &= 0, \\ q_t^- + q_{xxx}^- + 6q^-(x, t)q^+(x, t)q_x^- &= 0, \end{aligned} \quad (45)$$

directly linking to the Korteweg de Vries equation.

As we discussed in the previous section the scattering theory is based on introducing Jost solutions of $L(\lambda)$, scattering matrix, fundamental solutions, etc. In the $sl(2)$ case the Jost solutions are 2×2 matrix-valued solutions defined by an analog of (11) where the matrix J is simply substituted by σ_3 . Then one introduces the scattering matrix $T(\lambda, t)$ by

$$T(\lambda, t) \equiv (\psi_+(x, t, \lambda))^{-1} \psi_-(x, t, \lambda) = \begin{pmatrix} a^+(\lambda) & -b^-(\lambda, t) \\ b^+(\lambda, t) & a^-(\lambda) \end{pmatrix}, \quad (46)$$

which is x -independent. The t -dependence of the scattering matrix is driven by

$$i\partial_t T + [f(\lambda)\sigma_3, T(\lambda, t)] = 0. \quad (47)$$

Thus, if $q^{\pm}(x, t)$ satisfy the system of equations (42) we get

$$\partial_t a^{\pm}(\lambda) = 0, \quad i\partial_t b^{\pm}(\lambda) \mp 2f(\lambda)b^{\pm}(\lambda) = 0. \quad (48)$$

The matrix elements of $T(\lambda, t)$ are not independent. They satisfy the ‘unitarity’ condition $\det T(\lambda) \equiv a^+a^- + b^+b^- = 1$. Besides the diagonal elements a^+ and a^- allow analytic extension with respect to λ in the upper and lower complex λ -plane respectively. In fact the minimal set of scattering data which uniquely determines both the scattering matrix and the corresponding potential $q(x)$ consists of two types of variables: (i) the reflection coefficients $\rho^{\pm}(\lambda) = b^{\pm}/a^{\pm}$ defined for real $\lambda \in \mathbb{R}$ and (ii) a discrete set of scattering data including the discrete eigenvalues $\lambda_k^{\pm} \in \mathbb{C}_{\pm}$ and the constants $C_k^{\pm}$ which determine the norm of the corresponding Jost solutions [23].

A simple analysis shows that the first column of ψ_+ allows analytic continuation in the lower half plane of the spectral parameter while the last one in the upper half plane (for ψ_- the opposite holds true)

$$\psi_+(x, t, \lambda) = |\psi_+^-, \psi_+^+|, \quad \psi_-(x, t, \lambda) = |\psi_-^+, \psi_-^-|. \quad (49)$$

The superscripts $\pm$ in the columns of the Jost solutions refer to their analyticity properties while the subscripts $\pm$ refer to different Jost solutions (with different limits of x). The fundamental analytic solutions are constructed in the following manner:

$$\chi^+(x, t, \lambda) = |\psi_+^-, \psi_+^+|, \quad \chi^-(x, t, \lambda) = |\psi_-^+, \psi_-^-|. \quad (50)$$

The functions $a^\pm(\lambda) = \det \chi^\pm(x, \lambda)$ are known as the Evans functions [6, 24] of the system $L(\lambda)$. Their importance comes from the fact that they are t -independent (see (48)), and therefore they (or rather $\ln a^\pm$) can be viewed as generating functionals of the (local) integrals of motion. In addition it is known that their zeroes determine the discrete eigenvalues of $L(\lambda)$,

$$a^+(\lambda_k^+) = 0, \quad \lambda_k^+ \in \mathbb{C}_+; \quad a^-(\lambda_k^-) = 0, \quad \lambda_k^- \in \mathbb{C}_-. \quad (51)$$

One can define the soliton solutions of the NLEE as those for which $\rho^\pm(\lambda) = 0$ for all $\lambda \in \mathbb{R}$. Thus the soliton solutions of (42) are parametrized by the discrete eigenvalues and the constants $C_k^\pm$ whose t -dependence is determined from

$$\frac{d\lambda_k^\pm}{dt} = 0, \quad i \frac{dC_k^\pm}{dt} \mp 2f_k^\pm C_k^\pm = 0, \quad f_k^\pm = f(\lambda_k^\pm). \quad (52)$$

In fact we will analyze the various possible types of one-soliton solutions; in our case they are determined by one *pair* of discrete eigenvalues $\lambda^\pm \in \mathbb{C}_\pm$ and one *pair* of norming constants $C^\pm$. Thus for (42) we get just one type of one-soliton solutions. In order to derive its explicit form we shall use the dressing Zakharov–Shabat method [7]. In our case the dressing factor $u(x, t, \lambda)$ is given by a 2×2 matrix of the form (27) where P is a projector of rank 1 (see (30)). Then the following relations hold

$$P|n(x, t)\rangle = |n(x, t)\rangle, \quad |n(x, t)\rangle = \begin{pmatrix} n_1(x, t) \\ n_2(x, t) \end{pmatrix}, \quad (53)$$

$$\langle m(x, t)|P(x, t) = \langle m(x, t)|, \quad \langle m(x, t)| = (m_1(x, t), m_2(x, t)). \quad (54)$$

The transmission coefficients are transformed by the dressing procedure as follows:

$$a^+(\lambda) = c(\lambda)a_0^+(\lambda), \quad a^-(\lambda) = \frac{a_0^-(\lambda)}{c(\lambda)}. \quad (55)$$

The $sl(2)$ analog of (32) reads

$$q(x, t) - q_0(x, t) = -(\lambda^+ - \lambda^-)[\sigma_3, P(x, t)]. \quad (56)$$

By applying the above formulae to properly chosen constant vectors $|n_0\rangle$ and $\langle m_0|$ we can construct the eigenvectors of $P(x, t)$ and as a result, obtain $P(x, t)$ explicitly. It then remains only to insert it into (56) in order to obtain the corresponding potential $q(x, t)$ explicitly. It can be proved that the spectrum of $L(\lambda)$ will differ from the spectrum of $L_0(\lambda)$ only by an additional pair of discrete eigenvalues located at $\lambda^\pm \in \mathbb{C}_\pm$.

A pure soliton solution is obtained by assuming $q_0(x, t) = 0$; as a result we have

$$\begin{aligned} |n(x, t)\rangle &= e^{-i(x\lambda^+ + f^+t)\sigma_3} |n_0\rangle, & \langle m(x, t)| &= \langle m_0| e^{i(x\lambda^- - f^-t)\sigma_3}, \\ P(x, t) &= \frac{1}{2 \cosh \Phi_0(x, t)} \begin{pmatrix} e^{\Phi_0(x, t)} & \kappa_2 e^{-i\Phi(x, t)} \\ \frac{1}{\kappa_2} e^{i\Phi(x, t)} & e^{-\Phi_0(x, t)} \end{pmatrix} \\ \Phi_0(x, t) &= -i(\lambda^+ - \lambda^-)x + i(f^+ - f^-)t - \ln \kappa_1, \\ \Phi(x, t) &= (\lambda^+ + \lambda^-)x - (f^+ + f^-)t, \end{aligned} \quad (57)$$

where $f^\pm$ and the constants κ_1 and κ_2 are given by

$$f^\pm = f(\lambda^\pm), \quad \kappa_1 = \sqrt{\frac{n_{0,1}m_{0,1}}{n_{0,2}m_{0,2}}}, \quad \kappa_2 = \sqrt{\frac{n_{0,1}m_{0,2}}{n_{0,2}m_{0,1}}}. \quad (58)$$

Then the corresponding one-soliton solution takes the form

$$q^+(x, t) = -\frac{\kappa_2(\lambda^+ - \lambda^-) e^{-i\Phi(x, t)}}{\cosh \Phi_0(x, t)}, \quad q^-(x, t) = \frac{\kappa_2(\lambda^+ - \lambda^-) e^{i\Phi(x, t)}}{\kappa_2 \cosh \Phi_0(x, t)}. \quad (59)$$

Remark 2. The eigenvalues $\lambda^\pm$ are two independent complex numbers; therefore, in the denominator in (57) we get cosh of complex argument. This function vanishes whenever its argument becomes equal to $i(\pi/2 + p\pi)$ for some integer p and the generic solitons of (42) may have singularities for finite x and t .

One way to avoid these singularities is to impose on the Zakharov–Shabat system an involution, i.e. if we constrain the potential $q_0(x, t)$ by

$$q(x, t) = q^\dagger(x, t), \quad \text{i.e.} \quad q^+ = (q^-)^* = u(x, t). \quad (60)$$

Such constraint reduces (42) to NLEE for the single function $u(x, t)$; the second equation of the system becomes consequence of the first one. As a result (44) becomes the NLS equation

$$iu_t + u_{xx} + 2|u|^2 u(x, t) = 0, \quad (61)$$

while (45) goes into the MKdV-type equation

$$u_t + u_{xxx} + 6|u(x, t)|^2 u_x = 0. \quad (62)$$

This involution imposes constraints on all the scattering data; in particular we have

$$a^+(\lambda) = (a^-(\lambda^*))^*, \quad b^+(\lambda) = (b^-(\lambda^*))^*. \quad (63)$$

From the first relation above we find that the zeroes of the functions $a^\pm(\lambda)$ which are the eigenvalues of $L_0(\lambda)$ must satisfy

$$\lambda^+ = (\lambda^-)^* = \mu + i\nu, \quad C^+ = (C^-)^*, \quad P(x, t) = P^\dagger(x, t). \quad (64)$$

So now the one-soliton solution corresponds to a pair of eigenvalues which must be mutually conjugated pairs.

As a result we find that the expression for $P(x, t)$ and the one for the one-soliton solution simplifies to

$$P(x, t) = \frac{1}{2 \cosh \Phi_{00}(x, t)} \begin{pmatrix} e^{\Phi_{00}(x, t)} & e^{-i\Phi_{01}(x, t)} \\ e^{i\Phi_{01}(x, t)} & e^{-\Phi_{00}(x, t)} \end{pmatrix} \\ \Phi_{00}(x, t) = 2\nu x - 2ht - \ln \left| \frac{n_0^1}{n_0^2} \right|, \\ \Phi_{01}(x, t) = 2\mu x - 2gt - \arg n_0^1 + \arg n_0^2, \quad (65)$$

where

$$\lambda^\pm = \mu \pm i\nu, \quad f^\pm = g \pm ih. \quad (66)$$

Now both functions $\Phi_{00}(x, t)$ and $\Phi_{01}(x, t)$ become real valued. The denominator now becomes cosh of real argument, so this soliton solution is regular function for all x and t .

One can impose on $q_0(x, t)$ a different involution

$$q(x, t) = -q^\dagger(x, t), \quad \text{i.e.} \quad q^+ = -(q^-)^* = u(x, t). \quad (67)$$

However it is well known that under this involution the Zakharov–Shabat system $L(\lambda)$ becomes equivalent to an eigenvalue problem

$$\mathcal{L}\psi(x, t, \lambda) \equiv i\sigma_3 \partial_x \psi + \sigma_3 q(x, t) \psi(x, t, \lambda) = \lambda \psi(x, t, \lambda), \quad (68)$$

where the operator $\mathcal{L}$ is a self-adjoint one, so its spectrum must be on the real λ -axis. But the continuous spectrum of $\mathcal{L}$ fills up the whole real λ -axis, which leaves no room for solitons.

Finally, the Zakharov–Shabat system can be restricted by a third involution, e.g.,

$$q(x, t) = -q^T(x, t), \quad \text{i.e.} \quad q^+ = -q^- = -i w_x. \quad (69)$$

Such involution is compatible only with those NLEE whose dispersion law is odd function $f(\lambda) = -f(-\lambda)$. Therefore it cannot be applied to the NLS equation; applied to the MKdV equation it gives

$$w_{xt} + w_{xxx} + 6(w_x(x, t))^2 w_x = 0, \quad (70)$$

which can be integrated ones with the result $v = w_x$,

$$v_t + v_{xxx} + 6(v(x, t))^2 v_x = 0, \quad (71)$$

i.e. we get the MKdV equation for the real-valued function $v(x, t)$. It is well known also that the NLEE with dispersion law $f(\lambda) = (2\lambda)^{-1}$ can be explicitly derived under this reduction and comes out to be the famous sine-Gordon equation [25]

$$w_{xt} + \sin(2w(x, t)) = 0. \quad (72)$$

This second involution can be imposed together with the one in (60). The restrictions that it imposes on the scattering data are as follows:

$$a^+(\lambda) = (a^-(\lambda^*))^*, \quad a^+(\lambda) = (a^-(-\lambda)). \quad (73)$$

Now if λ^+ is an eigenvalue of $L(\lambda)$ then $(\lambda^+)^*$, $-\lambda^+$ and $-(\lambda^+)^*$ must also be eigenvalues. This means that we can have two configurations of eigenvalues:

(i) pairs of purely imaginary eigenvalues

$$\lambda^+ = i\nu \equiv -(\lambda^+)^*, \quad \lambda^- = -i\nu \equiv -(\lambda^-)^*; \quad (74)$$

(ii) quadruplets of complex eigenvalues

$$\begin{aligned} \lambda^+ &= \mu + i\nu & -(\lambda^+)^* &= -\mu + i\nu, \\ \lambda^- &= \mu - i\nu & -(\lambda^-)^* &= -\mu - i\nu. \end{aligned} \quad (75)$$

Thus we conclude, that the sine-Gordon and MKdV equations allow *two types* of solitons: type 1 with purely imaginary pairs of eigenvalues and type 2 each corresponding to a quadruplet of eigenvalues. Type 1 solitons are known also as topological solitons, or kinks (for details see [2]). They are parametrized by two real parameters: ν and $|C^+|$ so they have just one degree of freedom corresponding to the uniform motion.

Type 2 solitons are known as the breathers and are parametrized by four real parameters: μ and ν and the real and imaginary parts of C^+ . Therefore they have two degrees of freedom: one corresponds to the uniform motion and the second one describes the internal degree of freedom responsible for the ‘breathing’.

The purpose of presenting the above well-known facts in the above manner was simply to make it clear that the structure, as well as the number of related parameters which determine what different types of solitons can exist, depend strongly on the type of, and the number of, different involutions that can be imposed on the system.

3. The Generalized Zakharov–Shabat system and $sl(n)$ solitons

3.1. N -wave system related to $sl(3)$

In this subsection, we start with the generic N -wave system related to $sl(3)$ and analyze its reductions. The one-soliton solutions for the generic case and for the typical reductions are well known [9, 10]. Next we impose $\mathbb{Z}_2$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ reductions and derive the corresponding one-soliton solutions. Below we will use a notation which exploits the root structure of $sl(3)$, namely $Q_{kn}k$, $n = 1, 2$ stands for the component of Q associated with the root $\alpha = k\alpha_1 + n\alpha_2$ expanded over the simple roots $\alpha_1 = e_1 - e_2$ and $\alpha_2 = e_2 - e_3$. Taking into account that convention the generic N -wave system for $sl(3)$ consists of six equations; the first three of them are given by

$$i(J_1 - J_2)Q_{10,t} - i(I_1 - I_2)Q_{10,x} + 3kQ_{11}Q_{0\bar{1}} = 0, \quad (76)$$

$$i(J_1 - J_3)Q_{11,t} - i(I_1 - I_3)Q_{11,x} - 3kQ_{10}Q_{01} = 0, \quad (77)$$

$$i(J_2 - J_3)Q_{01,t} - i(I_2 - I_3)Q_{01,x} + 3kQ_{1\bar{0}}Q_{11} = 0; \quad (78)$$

the other three equations are obtained from (76) by replacing $J_k \leftrightarrow -J_k$, $I_k \leftrightarrow -I_k$ and $Q_{kn} \leftrightarrow Q_{k\bar{n}}$. We recall that $\sum_{k=1}^3 J_k = 0$, $\sum_{k=1}^3 I_k = 0$ and $k = J_1 I_2 - I_1 J_2$. This system can be solved via a standard dressing procedure [7, 9] with the dressing factor (27). The one-soliton solution obtained that way is given by the following expressions:

$$\begin{aligned} Q_{10}(x, t) &= \frac{\lambda^- - \lambda^+}{\langle m | n \rangle} e^{-i(\lambda^+ z_1 - \lambda^- z_2)} n_{0,1} m_{0,2}, \\ Q_{11}(x, t) &= \frac{\lambda^- - \lambda^+}{\langle m | n \rangle} e^{-i(\lambda^+ z_1 - \lambda^- z_3)} n_{0,1} m_{0,3}, \\ Q_{01}(x, t) &= \frac{\lambda^- - \lambda^+}{\langle m | n \rangle} e^{-i(\lambda^+ z_2 - \lambda^- z_3)} n_{0,2} m_{0,3}, \end{aligned} \quad (79)$$

where $z_\sigma = J_\sigma x + I_\sigma t$ and

$$\langle m | n \rangle = \sum_{j=1}^3 e^{-i(\lambda^+ - \lambda^-)z_j} n_{0,j} m_{0,j}. \quad (80)$$

The expressions for the other three fields can be derived from these by executing the following changes $Q_{kn} \leftrightarrow Q_{k\bar{n}}$, $e^{i\lambda^+ z_j} \leftrightarrow e^{-i\lambda^- z_j}$, $n_{0,j} \leftrightarrow m_{0,j}$.

Next we consider the typical $\mathbb{Z}_2$ reductions of the type

$$K_1 U^\dagger(x, \lambda^*) K_1^{-1} = U(x, \lambda), \quad \Rightarrow K_1 J^* K_1^{-1} = J, \quad K_1 Q^\dagger K_1^{-1} = -Q, \quad (81)$$

where $K_1 = \text{diag}(\epsilon_1, \epsilon_2, \epsilon_3)$, $\epsilon_j^2 = 1$, is an element of the Cartan subgroup $H \subset SL(3)$ which represents an action of $\mathbb{Z}_2$. This results in reducing the number of independent fields since we have

$$Q_{1\bar{0}} = -\epsilon_1 \epsilon_2 Q_{10}^*, \quad Q_{1\bar{1}} = -\epsilon_1 \epsilon_3 Q_{11}^*, \quad Q_{0\bar{1}} = -\epsilon_2 \epsilon_3 Q_{01}^*$$

and therefore the number of equations reduces from 6 to 3 as follows:

$$i(J_1 - J_2)Q_{10,t} - i(I_1 - I_2)Q_{10,x} - 3k\epsilon_2 \epsilon_3 Q_{11}Q_{01}^* = 0, \quad (82)$$

$$i(J_1 - J_3)Q_{11,t} - i(I_1 - I_3)Q_{11,x} - 3kQ_{10}Q_{01} = 0, \quad (83)$$

$$i(J_2 - J_3)Q_{01,t} - i(I_2 - I_3)Q_{01,x} - 3k\epsilon_1 \epsilon_2 Q_{10}^* Q_{11} = 0. \quad (84)$$

The discrete eigenvalues of $\mathbb{Z}_2$ -reduced operator L are complex conjugated, i.e. $\lambda^- = (\lambda^+)^* = \mu - i\nu$ and the polarization vectors are interrelated via $|n\rangle = K_1|m\rangle^*$. The one-soliton solution in this case is

$$Q_{10}(x, t) = -\frac{2i\nu e^{\nu(z_1+z_2)}}{\langle n^*|K_1|n\rangle} e^{-i\mu(z_1-z_2)} n_{0,1} \epsilon_2 n_{0,2}^*, \quad (85)$$

$$Q_{11}(x, t) = -\frac{2i\nu e^{-\nu z_2}}{\langle n^*|K_1|n\rangle} e^{-i\mu(z_1-z_3)} n_{0,1} \epsilon_3 n_{0,3}^*, \quad (86)$$

$$Q_{01}(x, t) = -\frac{2i\nu e^{-\nu z_1}}{\langle n^*|K_1|n\rangle} e^{-i\mu(z_2-z_3)} n_{0,2} \epsilon_3 n_{0,3}^*. \quad (87)$$

Remark 3. In general, the denominator (80) of the expressions for the one-soliton solution can possess zeros for some x and t , i.e. we have singular solitons. The same holds true also for the 3-wave system with the typical involutions for which $K \neq \mathbb{1}$. This is directly related to the well-known 'blow-up' instability [9]. For the typical involution with $K_1 = \mathbb{1}$ the corresponding denominator is a sum real-valued exponentials multiplied by some positive factors which is always positive.

By imposing another $\mathbb{Z}_2$ reduction on the potential $U(x, \lambda)$, namely

$$K_2 U^T(x, -\lambda) K_2^{-1} = -U(x, \lambda), \quad K_2 J^T K_2^{-1} = J, \quad K_2 Q^T K_2^{-1} = Q, \quad (88)$$

where $K_2 \in H$ satisfies $[K_1, K_2] = 0$ we obtain a $\mathbb{Z}_2 \times \mathbb{Z}_2$ -reduced $sl(3)N$ -wave system. As a consequence we have a pair of purely imaginary eigenvalues $\lambda^\pm = \pm i\nu$. Choosing $K_1 = K_2 = \mathbb{1}$ we see that the three independent fields $Q_{10}(x, t)$, $Q_{01}(x, t)$ and $Q_{11}(x, t)$ are purely imaginary while the polarization vector is real, $|n\rangle^* = |n\rangle$. After introducing new variables

$$Q_{10}(x, t) = i q_{10}(x, t), \quad Q_{01}(x, t) = i q_{01}(x, t), \quad Q_{11}(x, t) = i q_{11}(x, t),$$

we derive a real 3-wave system for three real-valued fields

$$\begin{aligned} (J_1 - J_2) q_{10,t} - (I_1 - I_2) q_{10,x} + 3k q_{11} q_{01} &= 0, \\ (J_1 - J_3) q_{11,t} - (I_1 - I_3) q_{11,x} - 3k q_{10} q_{01} &= 0, \\ (J_2 - J_3) q_{01,t} - (I_2 - I_3) q_{01,x} + 3k q_{10} q_{11} &= 0. \end{aligned} \quad (89)$$

Since the dressing factor must satisfy the conditions

$$(u^\dagger(x, \lambda^*))^{-1} = u(x, \lambda), \quad (90)$$

$$(u^T(x, -\lambda))^{-1} = u(x, \lambda), \quad (91)$$

the projector P is real valued. In this case the discrete eigenvalues are purely imaginary, i.e. $\lambda^\pm = \pm i\nu$. The one-soliton solution is

$$q_{kl}^{1s}(x) = -2\nu P_{kl}(x), \quad P = \frac{|n\rangle\langle n|}{\langle n|n\rangle}, \quad k \neq l.$$

Taking into account that $|n\rangle = e^{\nu J_3} |n_0\rangle$ we derive explicitly the following result:

$$\begin{aligned} q_{10}(x, t) &= -\frac{2\nu e^{\nu(z_1+z_2)} n_{0,1} n_{0,2}}{e^{2\nu z_1} n_{0,1}^2 + e^{2\nu z_2} n_{0,2}^2 + e^{-2\nu(z_1+z_2)} n_{0,3}^2}, \\ q_{11}(x, t) &= -\frac{2\nu e^{-\nu z_2} n_{0,1} n_{0,3}}{e^{2\nu z_1} n_{0,1}^2 + e^{2\nu z_2} n_{0,2}^2 + e^{-2\nu(z_1+z_2)} n_{0,3}^2}, \\ q_{01}(x, t) &= -\frac{2\nu e^{-\nu z_1} n_{0,2} n_{0,3}}{e^{2\nu z_1} n_{0,1}^2 + e^{2\nu z_2} n_{0,2}^2 + e^{-2\nu(z_1+z_2)} n_{0,3}^2}. \end{aligned} \quad (92)$$

In the $\mathbb{Z}_2 \times \mathbb{Z}_2$ case there exists another type of soliton solutions—these obtained by using a dressing factor of the form

$$u(x, \lambda) = \mathbb{I} + (c(\lambda) - 1)A(x) + \left(\frac{1}{c(-\lambda)} - 1\right)B(x), \quad c(\lambda) = \frac{\lambda - \lambda^+}{\lambda - \lambda^-}, \quad (93)$$

while its inverse reads

$$u^{-1}(x, \lambda) = \mathbb{I} + \left(\frac{1}{c(\lambda)} - 1\right)C(x) + (c(-\lambda) - 1)D(x). \quad (94)$$

These solutions are associated with four discrete eigenvalues of the scattering operator L : $\pm\lambda^\pm$. In this sense they may be called quadruplet solitons unlike the solutions (92) which being associated with two eigenvalues $\pm i\nu$ represent doublet solitons. The dressing factor (93) is $\mathbb{Z}_2 \times \mathbb{Z}_2$ invariant if the conditions hold true

$$K_1(u^\dagger(x, \lambda^*))^{-1}K_1^{-1} = u(x, \lambda), \quad (95)$$

$$K_2(u^T(x, -\lambda))^{-1}K_2^{-1} = u(x, \lambda). \quad (96)$$

Hence the following relations hold:

$$B = K_1 K_2 A^* K_2^{-1} K_1^{-1}, \quad C = K_1 A^\dagger K_1^{-1}, \quad D = K_2 A^T K_2^{-1}, \quad (97)$$

as well as $\lambda^- = (\lambda^+)^* = \mu - i\nu$.

The matrix-valued function A admits the decomposition

$$A(x) = X(x)F^T(x). \quad (98)$$

By using the equality $uu^{-1} = \mathbb{I}$ one can prove that the factor $X(x)$ can be expressed by $F(x)$ as follows:

$$X = \frac{1}{a^2 - |b|^2} (aK_1 F^* - b^* K_2 F), \quad (99)$$

where

$$a = F^\dagger K_1 F, \quad b = -\frac{i\nu F^T K_2 F}{\mu - i\nu}, \quad F^T(x) = F_0^T(\chi_0^-(x, \lambda^-))^{-1}. \quad (100)$$

To find the one-soliton solution we take the limit $\lambda \rightarrow \infty$ in (23) and put $q_0 \equiv 0$. Thus we obtain the following formula:

$$Q_{jk}^{\text{ls}} = (\lambda^- - \lambda^+)(A + K_1 K_2 A^* K_2 K_1)_{jk}, \quad j \neq k. \quad (101)$$

Let $K_1 = K_2 = \mathbb{I}$. Then $Q^* = -Q$ and using the above notation we have for the one-soliton solution:

$$q_{10} = -\frac{4\nu}{\Delta} f_{12} \sum_{k=1}^3 f_{kk} \left\{ \cos(\phi_1 - \phi_2) - \frac{\nu \cos(2\phi_k - \phi_1 - \phi_2 + \phi_0)}{\sqrt{\mu^2 + \nu^2}} \right\}, \quad (102)$$

$$q_{11} = -\frac{4\nu}{\Delta} f_{13} \sum_{k=1}^3 f_{kk} \left\{ \cos(\phi_1 - \phi_3) - \frac{\nu \cos(2\phi_k - \phi_1 - \phi_3 + \phi_0)}{\sqrt{\mu^2 + \nu^2}} \right\}, \quad (103)$$

$$q_{01} = -\frac{4\nu}{\Delta} f_{23} \sum_{k=1}^3 f_{kk} \left\{ \cos(\phi_2 - \phi_3) - \frac{\nu \cos(2\phi_k - \phi_2 - \phi_3 + \phi_0)}{\sqrt{\mu^2 + \nu^2}} \right\},$$

$$\Delta = a^2 - |b|^2 = \frac{1}{\mu^2 + \nu^2} \left\{ \mu^2 \sum_{k=1}^3 f_{kk}^2 + 2 \sum_{k < p} f_{kp}^2 [\mu^2 + 2\nu^2 \sin^2(\phi_k - \phi_p)] \right\}, \quad (104)$$

$$f_{jk} = |F_{0,j} F_{0,k}| e^{\nu(z_j + z_k)}, \quad \phi_k(x, t) = \mu z_k + \delta_k, \quad \delta_k = \arg F_{0,k},$$

where $\lambda^+ = i\sqrt{\mu^2 + \nu^2} e^{i\phi_0}$.

3.2. *N*-wave systems related to $sl(n)$, $n > 3$

For the sake of simplicity and clarity below, most of our discussions will be restricted to the case $n = 5$ and rank-1 projectors P . They also could easily be reformulated for any other chosen value of n and for higher rank projectors. The corresponding Lax operator $L(\lambda)$ which is a particular case of (3) with

$$L \equiv i\partial_x + U(x, t, \lambda) = i\partial_x + [J, Q(x, t)] - \lambda J,$$

$$J = \text{diag}(J_1, J_2, J_3, J_4, J_5), \quad Q(x, t) = \begin{pmatrix} 0 & Q_{12} & Q_{13} & Q_{14} & Q_{15} \\ Q_{21} & 0 & Q_{23} & Q_{24} & Q_{25} \\ Q_{31} & Q_{32} & 0 & Q_{34} & Q_{35} \\ Q_{41} & Q_{42} & Q_{43} & 0 & Q_{45} \\ Q_{51} & Q_{52} & Q_{53} & Q_{54} & 0 \end{pmatrix}.$$

Furthermore, for definiteness we will assume that

$$\text{tr } J = 0, \quad J_1 > J_2 > J_3 > 0, \quad 0 > J_4 > J_5. \quad (106)$$

The M -operator in the Lax representation for the N -wave equation (10) is given by

$$M\psi(x, t, \lambda) \equiv i\partial_t \psi + ([I, Q(x, t)] - \lambda I)\psi(x, t, \lambda) = -\lambda \psi(x, t, \lambda)I, \quad (107)$$

where $I = \text{diag}(I_1, \dots, I_5)$ is a traceless matrix. The generic one-soliton solution can be derived by using (32),

$$q(x) = \lim_{\lambda \rightarrow \infty} \lambda(J - u(x, \lambda)Ju^{-1}(x, \lambda)) = -(\lambda^+ - \lambda^-)[J, P(x)], \quad (108)$$

with a generic projector P whose rank s can be bigger than 1,

$$P(x, t) = \sum_{a,b=1}^s |n_a(x, t)\rangle M_{ab}^{-1} \langle n_b^\dagger(x, t)|, \quad M_{ab}(x, t) = \langle n_b^\dagger(x, t) | n_a(x, t) \rangle, \quad (109)$$

$$|n_a(x, t)\rangle = \chi_0^+(x, t, \lambda^+) |n_{0,a}\rangle, \quad \langle n_{0,a} | S_0 | n_{0,b} \rangle = 0.$$

Note that the set of s linearly independent polarization vectors $|n_k\rangle$ determine the corresponding eigensubspace of L . Such subspace can be defined either as the image of P or as the kernel of the projector $\tilde{P} = \mathbb{1} - P$ which is defined by a complimentary set of $n - s$ polarization vectors. Therefore studying $sl(n)$ -type Zakharov–Shabat systems it is enough to analyze projectors of rank $s \leq [n/2]$, where $[n/2]$ is the entire part of $n/2$. Thus for $n = 3$ it is enough to study rank-1 projectors, while for $n = 5$ one needs also rank-2 projectors.

For $s = 2$ we have two linearly independent polarization vectors $|n_a\rangle$, $a = 1, 2$ and from (109) we get

$$P(x, t) = \frac{1}{\det M} (|n_1(x, t)\rangle M_{22} \langle n_1^\dagger(x, t)| - |n_2(x, t)\rangle M_{12} \langle n_1^\dagger(x, t)|$$

$$- |n_1(x, t)\rangle M_{21} \langle n_2^\dagger(x, t)| + |n_2(x, t)\rangle M_{11} \langle n_2^\dagger(x, t)|),$$

$$\det M(x, t) = M_{11}M_{22} - M_{12}M_{21}, \quad M_{ab}(x, t) = \langle n_a^\dagger(x, t) | n_b(x, t) \rangle. \quad (110)$$

Let first concentrate on rank-1 projectors, see (93) with

$$|n(x)\rangle = \chi_0^+(x, \lambda^+) |n_0\rangle, \quad \langle m(x)| = \langle m_0 | \hat{\chi}_0^-(x, \lambda^-). \quad (111)$$

The polarization vectors $|n_0\rangle$ and $\langle m_0|$ are constant 5-vectors. Thus these type of one-soliton solutions is parametrized by:

- (i) The discrete eigenvalues $\lambda^\pm = \mu^\pm \pm i\nu^\pm$; $\mu^\pm$ determine the soliton velocity, $\nu^\pm$ determine the amplitude.

(ii) The ‘polarization’ vectors $|n_0\rangle$, $\langle m_0|$ parametrize the internal degrees of freedom of the soliton. Note that $P(x)$ is invariant under the scaling of each of these vectors. Generically each ‘polarization’ has five components, one of which can be fixed, say to 1. So each ‘polarization’ is determined by four independent complex parameters.

We have several options that will lead to different types of solitons:

- (1) generic case when all components of $|n_0\rangle$ are non-vanishing;
- (2) several special subcases when one (or several) of these components vanish. The corresponding solitons will have different structures and properties.

For the generic choice of $|n_0\rangle$ one finds

$$\lim_{x \rightarrow \pm\infty} P(x, t) = P_{\pm}, \quad P_+ = E_{11}, \quad P_- = E_{nn}, \quad (112)$$

where the matrix E_{kj} has only one non-vanishing matrix element equal to 1 at position k, j , i.e. $(E_{kj})_{mp} = \delta_{km}\delta_{jp}$. Therefore both the limiting values $u_{\pm}(\lambda)$ and their inverse $\hat{u}_{\pm}(\lambda)$ are diagonal matrices

$$u_+(\lambda) = \text{diag}(c(\lambda), 1, 1, \dots, 1), \quad u_-(\lambda) = \text{diag}(1, 1, \dots, 1, c(\lambda)). \quad (113)$$

From (25) for $n = 5$ we have

$$T_{1j}(\lambda) = c(\lambda)(T_0)_{1j}(\lambda), \quad j = 1, 2, 3, 4; \quad (114)$$

$$T_{j5}(\lambda) = (T_0)_{j5}(\lambda)/c(\lambda), \quad j = 2, 3, 4, 5;$$

$$T_{ij}(\lambda) = (T_0)_{ij}(\lambda), \quad \text{for all other values of } i, j. \quad (115)$$

This relation allows us to derive the interrelations between the Gauss factors of $T_0(\lambda)$ and $T(\lambda)$. In particular we find for the principal minors of $T(\lambda)$,

$$m_k^+(\lambda) = c(\lambda)m_{0,k}^+(\lambda), \quad m_k^-(\lambda) = m_{0,k}^-(\lambda)/c(\lambda), \quad (116)$$

where $m_k^+(\lambda)$ (resp. $m_k^-(\lambda)$) are the upper (resp. lower) principal minors of $T(\lambda)$. Since $\chi_0^{\pm}(x, t, \lambda)$ are regular solutions of the RHP then $m_{0,k}^{\pm}(\lambda)$ have no zeroes at all, but (116) means all $m_k^{\pm}(\lambda)$ have a simple zero at $\lambda = \lambda^{\pm}$.

The generic one-soliton solution then is obtained by taking $\chi^{\pm}(x, t, \lambda) = \exp(-i\lambda Jx)$. As a result we get

$$(P(x, t))_{ks} = \frac{1}{k(x, t)} n_{0,k} m_{0,s} e^{-i(\lambda^+ z_k - \lambda^- z_s)}, \quad (117)$$

$$k(x, t) = \sum_{p=1}^n n_{0,p} m_{0,p} e^{-i(\lambda^+ - \lambda^-)z_p(x, t)}, \quad (118)$$

$$z_p(x, t) = J_p x + I_p t, \quad q_{ks}^{1s} = -(\lambda^+ - \lambda^-)(P(x, t))_{ks}, \quad (119)$$

i.e. in all channels we have nontrivial waves. The number of internal degrees of freedom is $2(n-1) = 8$. Note that the denominator $k(x, t)$ is a linear combination of exponentials with complex arguments, so it could vanish for certain values of x, t . Thus the generic soliton (117) in this case is a singular solution.

Next we impose on $U(x, t, \lambda)$ the involution

$$K U^{\dagger}(x, t, \lambda^*) K^{-1} = U(x, t, \lambda), \quad K = \text{diag}(\epsilon_1, \dots, \epsilon_n), \quad (120)$$

with $\epsilon_j = \pm 1$. More specifically this means that

$$K q^{\dagger}(x, t) K^{-1} = q(x, t), \quad K u^{\dagger}(x, t, \lambda^*) K^{-1} = u^{-1}(x, t, \lambda), \quad (121)$$

and

$$\lambda^* = (\lambda^-)^* = \mu + i\nu, \quad \langle m_0 | = (K | n_0 \rangle)^\dagger. \quad (122)$$

Thus only $|n_0\rangle$ is independent.

Then the one-soliton solution simplifies to

$$q_{ks}^{1s}(x, t) = -\frac{2iv(J_k - J_s)}{k_{\text{red}}(x, t)} \epsilon_s n_{0,k} n_{0,s}^* e^{v(z_k + z_s)} e^{-i\mu(z_k - z_s)}, \quad (123)$$

$$k_{\text{red}}(x, t) = \sum_{p=1}^n \epsilon_p |n_{0,p}|^2 e^{2vz_p(x,t)}. \quad (124)$$

The number of internal degrees of freedom now is $n - 1 = 4$. If one or more of ϵ_j are different, then this reduced soliton may still have singularities. The singularities are absent only if all ϵ_j are equal.

The analysis of solitons obtained with rank-2 projectors should go along the same lines. It is lengthier than the one above and we may omit it. We just note that even with the canonical reduction with $K = \mathbb{I}$ in (121), one cannot guarantee that $\det M > 0$ for all x and t (see (110)) which means that one may encounter singular solitons.

3.3. Typical $sl(2)$ solitons

The finite-dimensional irreducible representations of the algebra $sl(2)$ are labeled by their spin J and have dimension $2J + 1$; the spin J can take half-integer positive values. The simplest representation with $J = \frac{1}{2}$ is two dimensional and is known also as the typical representation of $sl(2)$. In this subsection, we will address first the simplest possibility when $sl(2)$ is embedded into $sl(5)$ via its typical representation; we will call such solitons typical $sl(2)$ solitons.

From now on we assume that the reduction (120) with $\epsilon_p = 1$ holds. Here $|n_{0,1}\rangle$ has only two non-vanishing components. We consider here three examples with $n = 5$ and three different choices for the polarization vectors

$$(a) \quad |n_0\rangle = \begin{pmatrix} n_{0,1} \\ 0 \\ 0 \\ 0 \\ n_{0,5} \end{pmatrix}; \quad (b) \quad |n_0\rangle = \begin{pmatrix} 0 \\ n_{0,2} \\ 0 \\ n_{0,4} \\ 0 \end{pmatrix}; \quad (c) \quad |n_0\rangle = \begin{pmatrix} n_{0,1} \\ n_{0,2} \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (125)$$

In all these cases the corresponding one-soliton solutions $q(x, t)$ are given by similar analytic expressions, each having only two non-vanishing matrix elements

$$q_{jk}(x, t) = (q_{jk}(x, t))^* = -\frac{iv(J_j - J_k) e^{i(\arg(n_{0,j}) - \arg(n_{0,k}))} e^{-i\mu(J_j - J_k)(x + w_{jk}t)}}{\cosh[v(J_j - J_k)(x + w_{jk}t) + \ln|n_{0,j}| - \ln|n_{0,k}|]}, \quad (126)$$

where we recall that $w_{jk} = (I_j - I_k)/(J_j - J_k)$, $j < k$. For case (a) we have $j = 1, k = 5$; in case (b): $j = 2, k = 4$ and in case (c) $j = 1$ and $k = 2$.

The $sl(2)$ soliton is very much like the NLS soliton (apart from the t -dependence); the NLS soliton has only one internal degree of freedom.

The different choices for the polarization vector result in different asymptotics for the projector $P_1(x)$,

$$\begin{aligned} (a) \quad & \lim_{x \rightarrow \infty} P(x) = E_{11}, & \lim_{x \rightarrow -\infty} P(x) = E_{55}, \\ (b) \quad & \lim_{x \rightarrow \infty} P(x) = E_{22}, & \lim_{x \rightarrow -\infty} P(x) = E_{44}, \\ (c) \quad & \lim_{x \rightarrow \infty} P(x) = E_{11}, & \lim_{x \rightarrow -\infty} P(x) = E_{22}. \end{aligned} \quad (127)$$

In case (a) the results for the limits of $P(x)$ and for $u_{\pm}(\lambda)$ are the same as for the generic case, see (112) and (113). As a consequence, such $sl(2)$ solitons require the vanishing of *all* Evans functions $m_k^{\pm}(\lambda)$ for $\lambda = \lambda^{\pm}$, see (116).

In case (b) from (25) and from the appendices we get that such $sl(2)$ soliton provides for the vanishing of $m_2^{\pm}(\lambda)$ and $m_3^{\pm}(\lambda)$,

$$\begin{aligned} m_2^+(\lambda) &= c(\lambda)m_{0,2}^+(\lambda), & m_3^+(\lambda) &= c(\lambda)m_{0,3}^+(\lambda), \\ m_2^-(\lambda) &= m_{0,2}^-(\lambda)/c(\lambda), & m_3^-(\lambda) &= m_{0,3}^-(\lambda)/c(\lambda), \end{aligned} \quad (128)$$

whereas $m_1^{\pm}(\lambda) = m_{0,1}^{\pm}(\lambda)$ and $m_4^{\pm}(\lambda) = m_{0,4}^{\pm}(\lambda)$ remain regular and do not have zeros at $\lambda = \lambda^{\pm}$.

Likewise in case (c) we get that only $m_1^+(\lambda)$ and $m_4^-(\lambda)$ acquire zeroes,

$$m_1^+(\lambda) = c(\lambda)m_{0,1}^+(\lambda), \quad m_4^-(\lambda) = m_{0,4}^-(\lambda)/c(\lambda), \quad (129)$$

and all the other Evans functions $m_j^+(\lambda)$ with $j = 2, 3, 4$, and $m_p^-(\lambda)$ with $p = 1, 2, 3$ do not have zeroes.

3.4. Higher spin J $sl(2)$ solitons

Here we provide some examples of $sl(2)$ soliton solutions which are embedded into $sl(5)$ via a higher $(2J + 1)$ -dimensional representation of $sl(2)$ which we call spin J $sl(2)$ solitons. Obviously for $g \simeq sl(5)$ the 'spin' of the solitons can take values $J = 1, 3/2$ and 2.

It is well known that spin J representation can be constructed using the completely symmetrized tensor products of the typical 2×2 representation. The details of their derivations are given in appendix B. Here we briefly formulate the results.

Before starting with to calculate symmetrized tensor products of the dressing factor we have to make a slight modification so that it takes values in the group $SL(2)$. Indeed, as it is given by (27) and (65) one easily checks that $\det u(x, \lambda) = c(\lambda)$. Therefore we multiply it by the constant $1/\sqrt{c(\lambda)}$ factor so that its determinant equals 1. Thus we can rewrite the dressing factor for the typical $sl(2)$ soliton in the form

$$u(x, \lambda) = \sqrt{c(\lambda)}P(x) + \frac{1}{\sqrt{c(\lambda)}}\bar{P}(x), \quad (130)$$

where $\bar{P} = \mathbb{1} - P$; in terms of the polarization vectors they take the form

$$P = \frac{1}{n_1m_1 + n_2m_2} \begin{pmatrix} n_1m_1 & n_1m_2 \\ n_2m_1 & n_2m_2 \end{pmatrix}, \quad \bar{P} = \frac{1}{n_1m_1 + n_2m_2} \begin{pmatrix} n_2m_2 & -n_1m_2 \\ -n_2m_1 & n_1m_1 \end{pmatrix}. \quad (131)$$

After some calculations we get the following results for the higher spin dressing factors:

$$\begin{aligned} U^{(3)} &\equiv u \odot u = c(\lambda)\pi_1^{(3)} + \pi_0^{(3)} + \frac{1}{c(\lambda)}\pi_{-1}^{(3)}, \\ U^{(4)} &\equiv u \odot u \odot u = \sum_{l=-3/2}^{3/2} \pi_l^{(4)} c^l(\lambda), \\ U^{(5)} &\equiv u \odot u \odot u \odot u = \sum_{l=-2}^2 \pi_l^{(5)} c^l(\lambda). \end{aligned} \quad (132)$$

By $u \odot u$, $u \odot u \odot u$, etc we have denoted the completely symmetrized part of the corresponding tensor powers of the dressing factor $u(x, \lambda)$, the superscript k of $U^{(k)}$ denotes the dimension $2J + 1$ of the representation. By $\pi_a^{(k)}$ we have denoted mutually orthogonal rank-1 projectors

$$\pi_a^{(k)} = \frac{|N_a^{(k)}\rangle\langle M_a^{(k)}|}{\langle M_a^{(k)}|N_a^{(k)}\rangle}, \quad \pi_a^{(k)}\pi_b^{(k)} = \delta_{ab}\pi_a^{(k)}. \quad (133)$$

The explicit expressions for the vectors $|N_a^{(k)}\rangle$ and $\langle M_a^{(k)}|$ in terms of n_i and m_j are given in appendix A. Here we just mention that

$$\langle M_a^{(k)}| = \langle N_a^{(k)}| S_0^{(k)}, \quad (134)$$

where the matrices $S_0^{(k)}$

$$S_0^{(k)} = \begin{cases} \sum_{s=0}^k (-1)^{k+1} E_{s,k+1-s}^{(k)} & \text{for } k = 2p+1 \\ \sum_{s=0}^{(k-1)/2} (-1)^{k+1} (E_{s,k+1-s}^{(k)} - E_{k+1-s,s}^{(k)}) & \text{for } k = 2p \end{cases} \quad (135)$$

are the ones used to define the orthogonal algebras $so(k)$.

All these properties of the projectors $\pi_a^{(k)}$ allow us to rewrite the dressing factors in the form

$$U^{(k)}(x, \lambda) = \exp \left(\ln c(\lambda) \sum_{s=1}^{(k-1)/2} (\pi_s^{(k)} - \pi_{-s}^{(k)}) + \pi_0^{(k)} \right), \quad (136)$$

for odd values of k and

$$U^{(k)}(x, \lambda) = \exp \left(\ln c(\lambda) \sum_{s=1/2}^{(k-1)/2} (\pi_s^{(k)} - \pi_{-s}^{(k)}) \right) \quad (137)$$

for even values of k . It is important to note that due to (134) the differences $\pi_s^{(k)} - \pi_{-s}^{(k)} \in so(k)$ and also $\pi_0^{(k)} \in so(k)$.

Let us now outline how one can construct spin 3/2 soliton of the N -wave equations related to $sl(5)$. First we have to embed the dressing factor $U^{(4)}$ which is a 4×4 matrix into a 5×5 dressing factor. This can be done in several inequivalent ways, which reflects the fact that the group $SO(4)$ can be embedded into the group $SL(5)$ in different ways. As first of them we choose the following one. First we extend the four-component vectors $|N_a^{(k)}\rangle$ into five-component ones $|\mathbf{N}_a^{(k)}\rangle = (|N_a^{(k)}\rangle, 0)^T$. We will need also the vector $|\mathbf{e}_5\rangle = (0, 0, 0, 0, 1)^T$. Then we can construct the 5×5 dressing factor

$$\mathbf{U}^{(4)} = \pi_{-3/2}^{(4)} c^{-3/2}(\lambda) + \pi_{-1/2}^{(4)} c^{-1/2}(\lambda) + \pi_{1/2}^{(4)} c^{1/2}(\lambda) + \pi_{3/2}^{(4)} c^{3/2}(\lambda), \quad (138)$$

where

$$\pi_a^{(4)} = \frac{|\mathbf{N}_a^{(4)}\rangle \langle \mathbf{M}_a^{(4)}|}{\langle \mathbf{M}_a^{(4)} | \mathbf{N}_a^{(4)} \rangle}, \quad \pi_0^{(4)} = \frac{|\mathbf{N}_0^{(4)}\rangle \langle \mathbf{M}_0^{(4)}|}{\langle \mathbf{M}_0^{(4)} | \mathbf{N}_0^{(4)} \rangle} + |\mathbf{e}_5\rangle \langle \mathbf{e}_5|. \quad (139)$$

In order to calculate the corresponding soliton solution it remains to insert $\mathbf{U}^{(5)}$ as $u(x, \lambda)$ into (108). For $Q_{(0)} = 0$ the result will be

$$Q_{3/2} = \begin{pmatrix} 0 & \sqrt{3}q^+ & 0 & 0 & 0 \\ \sqrt{3}q^- & 0 & 2q^+ & 0 & 0 \\ 0 & 2q^- & 0 & \sqrt{3}q^+ & 0 \\ 0 & 0 & \sqrt{3}q^- & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad (140)$$

where the functions $q^\pm(x, t)$ are given by (59).

It is natural to analyze the structure of the discrete eigenvalues $\lambda^\pm$ corresponding to these types of soliton solutions. The discrete eigenvalues of L are the zeroes of the principal minors

$m_k^\pm(\lambda)$ of the scattering matrix $T(\lambda)$. To this end we have to calculate the diagonal factors $D^\pm(\lambda)$ of the Gauss decomposition of $T(\lambda)$. The result is

$$\begin{aligned} m_1^+(\lambda) &= c^3(\lambda)m_{0,1}^+(\lambda), & m_2^+(\lambda) &= c^4(\lambda)m_{0,2}^+(\lambda), \\ m_3^+(\lambda) &= c^3(\lambda)m_{0,3}^+(\lambda), & m_4^+(\lambda) &= m_{0,4}^+(\lambda). \end{aligned} \quad (141)$$

Comparing with (129) we conclude that going to higher representations of $sl(2)$ leads to multiple zeroes of some of the principle minors of $T(\lambda)$. Therefore the resolvent of L will have poles of higher order at $\lambda^\pm$, though the residues at this points will be expressed by the same polarization vectors as for the $\frac{1}{2}$ -spin solitons. One consequence of this construction is that the higher spin $sl(2)$ solitons have the two degrees of freedom as the standard (spin $\frac{1}{2}$) solitons.

3.5. Typical $sl(3)$ -solitons

Here $|n_0\rangle$ has three non-vanishing components. We consider three examples of such polarization vectors

$$(a) \quad |n_0\rangle = \begin{pmatrix} n_{0,1} \\ 0 \\ n_{0,3} \\ 0 \\ n_{0,5} \end{pmatrix}, \quad (b) \quad |n_0\rangle = \begin{pmatrix} 0 \\ n_{0,2} \\ n_{0,3} \\ n_{0,4} \\ 0 \end{pmatrix}, \quad (c) \quad |n_0\rangle = \begin{pmatrix} n_{0,1} \\ n_{0,2} \\ n_{0,3} \\ 0 \\ 0 \end{pmatrix}. \quad (142)$$

Therefore the $sl(3)$ -solitons have two internal degrees of freedom.

The asymptotics of the projector $P(x)$ read as follows:

$$\begin{aligned} (a) \quad \lim_{x \rightarrow \infty} P(x) &= E_{11}, & \lim_{x \rightarrow -\infty} P(x) &= E_{55}, \\ (b) \quad \lim_{x \rightarrow \infty} P(x) &= E_{22}, & \lim_{x \rightarrow -\infty} P(x) &= E_{44}, \\ (c) \quad \lim_{x \rightarrow \infty} P(x) &= E_{11}, & \lim_{x \rightarrow -\infty} P(x) &= E_{33}. \end{aligned} \quad (143)$$

Note that cases (a) and (b) in (143) coincide with the corresponding cases in (125). Therefore the set of Evans functions that acquire zeroes will be the same as for the corresponding $sl(2)$ solitons. In case (c) of (143) we have

$$\begin{aligned} m_1^+(\lambda) &= c(\lambda)m_{0,1}^+(\lambda), & m_2^+(\lambda) &= c(\lambda)m_{0,2}^+(\lambda), \\ m_4^-(\lambda) &= m_{0,4}^-(\lambda)/c(\lambda), & m_3^-(\lambda) &= m_{0,3}^-(\lambda)/c(\lambda), \end{aligned} \quad (144)$$

whereas the remaining Evans functions $m_j^+(\lambda)$ with $j = 3, 4$, and $m_p^-(\lambda)$ with $p = 1, 2$ remain regular.

In case (a) the corresponding one-soliton solutions acquire the form

$$\begin{aligned} (a) \quad q^{1s} &= \begin{pmatrix} 0 & 0 & q_{13} & 0 & q_{15} \\ 0 & 0 & 0 & 0 & 0 \\ q_{13}^* & 0 & 0 & 0 & q_{35} \\ 0 & 0 & 0 & 0 & 0 \\ q_{15}^* & 0 & q_{35}^* & 0 & 0 \end{pmatrix}, & (b) \quad q^{1s} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{23} & q_{24} & 0 \\ 0 & q_{23}^* & 0 & q_{34} & 0 \\ 0 & q_{24}^* & q_{34}^* & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \\ (c) \quad q^{1s} &= \begin{pmatrix} 0 & q_{12} & q_{23} & 0 & 0 \\ q_{12}^* & 0 & q_{23} & 0 & 0 \\ q_{13}^* & q_{23}^* & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (145)$$

where the matrix elements $q_{ks}(x, t)$ are given by

$$q_{ks}(x, t) = -\frac{i\nu(J_k - J_s)\epsilon_s e^{\nu(\tilde{J}_k + \tilde{J}_s)(x + \tilde{v}_{ks}t)} n_{0,k} n_{0,s}^* e^{-i\mu(J_k - J_s)(x + w_{1s}t)}}{|n_{0,1}|^2 e^{2\nu(\tilde{J}_1 x + \tilde{I}_1 t)} + |n_{0,3}|^2 e^{2\nu(\tilde{J}_3 x + \tilde{I}_3 t)} + |n_{0,5}|^2 e^{2\nu(\tilde{J}_5 x + \tilde{I}_5 t)}}, \quad (146)$$

and

$$\tilde{J}_k = J_k - (J_1 + J_3 + J_5)/3, \quad \tilde{I}_k = I_k - (I_1 + I_3 + I_5)/3, \quad \tilde{v}_{ks} = \frac{\tilde{J}_k + \tilde{J}_s}{\tilde{I}_k + \tilde{I}_s}. \quad (147)$$

This soliton has two internal degrees of freedom and is regular.

Obviously it is by now clear how one can write more complicated solitons like $sl(4)$ which would be characterized by the polarization vectors of the form

$$(a) \quad |n_0\rangle = \begin{pmatrix} n_{0,1} \\ n_{0,2} \\ n_{0,3} \\ n_{0,4} \\ 0 \end{pmatrix}, \quad (b) \quad |n_0\rangle = \begin{pmatrix} n_{0,1} \\ n_{0,2} \\ n_{0,3} \\ 0 \\ n_{0,5} \end{pmatrix}, \quad \dots \quad (148)$$

The $sl(4)$ -solitons will have three internal degrees of freedom.

We note here that due to our choice of J in (106), $sl(4)$ -solitons cannot give rise to generalized eigenfunctions.

Of course one could try to embed $sl(3)$ into g using one of its higher dimensional representations. The algebra $sl(3)$ has a second three-dimensional representation which is obtained related to the typical one used above by an exterior automorphism. The corresponding soliton solutions are obtained from those demonstrated above by a re-parameterization of the polarization vectors. The next irreducible representation of $sl(3)$ which is the adjoint representation, has dimension 8, so it cannot be embedded into $sl(5)$. Of course, considering algebras g of rank high enough it is possible to embed subalgebras g_0 using their non-typical representations.

4. Eigenfunctions and eigensubspaces

The structure of these eigensubspaces and the corresponding solitons becomes more complicated with the growth of n .

In what follows we start with the generic case and split the 'polarization' vector into two parts

$$|n_0\rangle = |p_0\rangle + |d_0\rangle; \quad |p_0\rangle = \begin{pmatrix} n_{0,1} \\ n_{0,2} \\ n_{0,3} \\ 0 \\ 0 \end{pmatrix}, \quad |d_0\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ n_{0,4} \\ n_{0,5} \end{pmatrix}, \quad (149)$$

and therefore

$$|n\rangle = |p\rangle + |d\rangle, \quad |p\rangle = \chi_0^+(x, \lambda^+) |p_0\rangle, \quad |d\rangle = \chi_0^+(x, \lambda^+) |d_0\rangle. \quad (150)$$

This splitting is compatible with (106) and has the advantage: if $\chi_0^+(x, \lambda^+) = \exp(-i\lambda^+ Jx)$ then $|p\rangle$ increases exponentially for $x \rightarrow \infty$ and decreases exponentially for $x \rightarrow -\infty$; $|d\rangle$ decreases exponentially for $x \rightarrow \infty$ and increases exponentially for $x \rightarrow -\infty$, see also lemma 1.

What we will prove below is that one can take a special linear combination of the columns of $\chi_0^+(x, \lambda^+)$ which decreases exponentially for both $x \rightarrow \infty$ and $x \rightarrow -\infty$. Doing this we will use the fact that

$$\chi^+(x, \lambda^+)|n_0\rangle \equiv (\mathbb{1} - P(x))\chi^+(x, \lambda^+)|n_0\rangle = (\mathbb{1} - P(x))|n(x)\rangle = 0. \quad (151)$$

Lemma 1. *The eigenfunctions of L provided by*

$$f^+(x) = \chi^+(x, \lambda^+)|p_0\rangle = -\chi^+(x, \lambda^+)|d_0\rangle, \quad (152)$$

decrease exponentially for both $x \rightarrow \infty$ and $x \rightarrow -\infty$.

Proof. From (151) and (149) there follows that both expressions for $f^+(x, t)$ coincide, so we can use each of them to our advantage, see (152). We will use also the fact that $\mathbb{1} - P(x)$ is a bounded function of x .

We start with

$$\lim_{x \rightarrow \infty} f^+(x) = \lim_{x \rightarrow \infty} \chi'^{+,+}(x, \lambda^+)|d_0\rangle = (\mathbb{1} - P_+) \lim_{x \rightarrow \infty} e^{-i\lambda^+ J x} \mathbb{T}^-(\lambda^+)|d_0\rangle, \quad (153)$$

where $\mathbb{T}^-(\lambda^+)$ is the lower triangular matrix introduced in (17). If the potential is on finite support or is reflectionless then $\mathbb{T}^-(\lambda)$ is rational function well defined for $\lambda = \lambda^+$. If the potential is generic then $\mathbb{T}^-(\lambda)$ does not allow analytic continuation off the real axis. Nevertheless $\mathbb{T}^-(\lambda^+)$ can be understood as lower triangular constant matrix (generalizing the constant C_0^+ of the NLS case). Being lower triangular $\mathbb{T}^-(\lambda^+)$ maps $|d_0\rangle$ onto $|d'_0\rangle = \mathbb{T}_0^-(\lambda^+)|d_0\rangle$ which is again of the form (149), i.e. its first three components vanish. Therefore

$$\lim_{x \rightarrow \infty} e^{va x} f^+(x) = \lim_{x \rightarrow \infty} (\mathbb{1} - P_+) e^{va x} \begin{pmatrix} 0 \\ 0 \\ 0 \\ e^{-i\lambda^+ J_4 x} n'_{0,4} \\ e^{-i\lambda^+ J_5 x} n'_{0,5} \end{pmatrix} = 0, \quad (154)$$

for any constant $a > 0$ such that $a + J_4 < 0$.

Likewise we can calculate the limit for $x \rightarrow -\infty$

$$\lim_{x \rightarrow -\infty} f^+(x) = - \lim_{x \rightarrow -\infty} \chi'^{+,+}(x, \lambda^+)|p_0\rangle = -(\mathbb{1} - P_+) \lim_{x \rightarrow \infty} e^{-i\lambda^+ J x} \mathbb{S}^+(\lambda^+)|p_0\rangle. \quad (155)$$

The upper triangular matrix $\mathbb{S}^+(\lambda^+)$ is treated analogously as $\mathbb{T}^-(\lambda^+)$. In the generic case it is just an upper triangular constant matrix which maps $|p_0\rangle$ onto $|p'_0\rangle = \mathbb{S}^+(\lambda^+)|p_0\rangle$ whose last two components vanish. Therefore

$$\lim_{x \rightarrow -\infty} e^{vb x} f^+(x) = \lim_{x \rightarrow -\infty} e^{vb x} (\mathbb{1} - P_-) \begin{pmatrix} e^{-i\lambda^+ J_1 x} n'_{0,1} \\ e^{-i\lambda^+ J_2 x} n'_{0,2} \\ e^{-i\lambda^+ J_3 x} n'_{0,3} \\ 0 \\ 0 \end{pmatrix} = 0, \quad (156)$$

for any constant $b < 0$ such that $J_3 + b > 0$.

The lemma is proved. $\square$

For the choices (a) and (b) of $|n_0\rangle$ in (125) we define the square integrable discrete eigenfunctions using the splitting (149) and (152).

Remark 4. The choice (c) for $|n_0\rangle$ does not allow for the splitting (149). In this case we can introduce only generalized discrete eigenfunctions, $f_{\text{gen}}(x, t)$, which are not square integrable. But upon multiplying by the exponential factor $e^{-\nu c x}$ with $c = (J_1 + J_2)/2$, we can obtain square integrable functions $f(x) = f_{\text{gen}}(x) e^{-\nu c x}$. See also the discussion in the following subsection.

The generalized eigenfunctions come up in situations when the splitting (149) is not possible, i.e. when either $|p_0\rangle$ or $|d_0\rangle$ vanish. Let us construct the generalized eigenfunction for the polarization vector $|n_0\rangle$ of case (c) in (142). Let $(J_1 + J_2 + J_3)/3 = a'$; then $J'_1 = J_1 - a'$, $J'_2 = J_2 - a'$ and $J'_3 = J_3 - a'$ are such that $J'_1 > J'_2 > J'_3$ and $J'_1 + J'_2 + J'_3 = 0$. Let us assume for definiteness that $J'_1 > J'_2 > 0$ and $0 > J'_3$. Then we can split $|n_0\rangle$ into

$$|n_0\rangle = |p'_0\rangle + |d'_0\rangle, \quad |p'_0\rangle = \begin{pmatrix} n_{0,1} \\ n_{0,2} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |d'_0\rangle = \begin{pmatrix} 0 \\ 0 \\ n_{0,3} \\ 0 \\ 0 \end{pmatrix}. \quad (157)$$

and define

$$f^{+, \prime}(x) = \chi^+(x, \lambda^+) |p'_0\rangle = -\chi^+(x, \lambda^+) |d'_0\rangle. \quad (158)$$

Obviously $f^{+, \prime}(x)$ is an eigenfunction of the dressed operator L corresponding to the eigenvalue λ^+ .

Then we can prove the following lemma.

Lemma 2. *The eigenfunction $f^{+, \prime}(x)$ is such that $e^{\nu_1 a' x} f^{+, \prime}(x)$ decreases exponentially for both $x \rightarrow \pm\infty$.*

Proof. The proof is similar to the one of lemma 1 and we omit it. $\square$

Since the polarization vector $|n_0\rangle$ in case (c) of (142) does not allow the splitting (149) the corresponding discrete eigenfunction will not be square integrable, so it will give rise to a generalized eigenfunction.

5. Effects of reductions on soliton solutions

In this section, we analyze how different kinds of reductions affect the classification of the soliton solutions to a nonlinear equation. This criterion is tightly connected with symmetries imposed on the auxiliary linear problem (the zero curvature condition). We shall consider in the next subsection types of solitons which differ from one another in the number of eigenvalues associated with them: doublet solitons associated with two purely imaginary eigenvalues $\lambda^\pm = \pm i\nu$ and quadruplet solitons associated with four eigenvalues situated symmetrically with respect to the real and the imaginary axis in $\mathbb{C}$. This is the case when a $\mathbb{Z}_2 \times \mathbb{Z}_2$ reduction is in action. Such a type of reduction is compatible with the Lax representation of a NLEE to have a dispersion law obeying $f(-\lambda) = -f(\lambda)$. The N -wave equation is a typical example of NLEE to admit a variety of $\mathbb{Z}_2$ reductions [27] since $f_{N-W}(\lambda) = -\lambda I$.

It is also possible to impose reductions that do not affect the spectral parameter. Thus, imposing such reduction on the generalized ZSS related to the exterior automorphism of $sl(n)$ we can effectively reduce it to the generalized ZSS system for its subalgebras $so(n)$ and to the symmetric space of **BD.I** type.

For further convenience we shall say that the matrix X belongs to the orthogonal algebra $so(n)$ if

$$X + S_0^{(n)} X^T S_0^{(n)} = 0, \quad S_0^{(n)} S_0^{(n)} = \mathbf{1}, \quad (159)$$

where $S_0^{(n)}$ is defined by

$$S_0^{(n)} = \sum_{s=1}^{n+1} (-1)^{s+1} E_{s, n+1-s}^{(n)}, \quad (160)$$

for $n = 2r + 1$ and

$$S_0^{(n)} = \sum_{s=1}^r (-1)^{s+1} (E_{s, n+1-s}^{(n)} + E_{n+1-s, s}^{(n)}) \quad (161)$$

for $n = 2r$. With this definition of orthogonality the Cartan subalgebra generators are represented by diagonal matrices. By $E_{sp}^{(n)}$ above we mean $n \times n$ matrix whose matrix elements are $(E_{sp}^{(n)})_{ij} = \delta_{si} \delta_{pj}$.

In order to get the ZSS related to the symmetric space of **BD.I** type we have to specify $J = \text{diag}(1, 0, \dots, 0, -1)$ and take $q(x, t) = [J, Q(x, t)]$ where $Q(x, t) \in so(n)$. In this case the NLEE with linear dispersion law become trivial. However, one can consider NLEE with quadratic and cubic dispersion laws which are multicomponent generalizations of NLS and mKdV equations respectively.

5.1. N -wave system related to $so(5)$

From now on we shall focus our attention on a N -wave equation related to the $so(5)$ algebra. The corresponding Lax pair is given by (3), (4) where

$$\begin{aligned} U(x, t, \lambda) &= [J, Q(x, t)] - \lambda J, & V(x, t, \lambda) &= [I, Q(x, t)] - \lambda I, \\ J &= \text{diag}(J_1, J_2, 0, -J_2, -J_1), & I &= \text{diag}(I_1, I_2, 0, -I_2, -I_1), \end{aligned} \quad (162)$$

and $Q(x, t)$ is a function taking values in $so(5)$. This algebra has two simple roots $\alpha_1 = e_1 - e_2$, $\alpha_2 = e_2$, and two more positive roots: $\alpha_1 + \alpha_2 = e_1$ and $\alpha_1 + 2\alpha_2 = e_1 + e_2 = \alpha_{\max}$. When they come as indices, e.g. in Q_α we will replace them by sequences of two integers: $\alpha \rightarrow kn$ if $\alpha = k\alpha_1 + n\alpha_2$. Moreover, we are going to use the auxiliary notation $\overline{kn} = -k\alpha_1 - n\alpha_2$. Thus the N -wave system itself consists of eight equations. Half of them read

$$\begin{aligned} i(J_1 - J_2)Q_{10,t}(x, t) - i(I_1 - I_2)Q_{10,x}(x, t) + kQ_{11}(x, t)Q_{\overline{01}}(x, t) &= 0, \\ iJ_1Q_{11,t}(x, t) - iI_1Q_{11,x}(x, t) - k(Q_{10}Q_{01} - Q_{12}Q_{\overline{01}})(x, t) &= 0, \\ i(J_1 + J_2)Q_{12,t}(x, t) - i(I_1 + I_2)Q_{12,x}(x, t) - kQ_{11}(x, t)Q_{01}(x, t) &= 0, \\ iJ_2Q_{01,t}(x, t) - iI_2Q_{01,x}(x, t) + k(Q_{\overline{11}}Q_{12} + Q_{\overline{10}}Q_{11})(x, t) &= 0, \end{aligned} \quad (163)$$

where $k := J_1 I_2 - J_2 I_1$ is a constant describing the wave interaction. The other four equations can be derived from those above by using the formal transformation $Q_{kn} \leftrightarrow Q_{\overline{kn}}$. One is able

to integrate the system by applying the already discussed ideas—dressing method, etc. For that purpose we make use of the dressing factor (38). The one-soliton solution reads

$$\begin{aligned}
 Q_{10}(x, t) &= \frac{\lambda^- - \lambda^+}{\langle m|n \rangle} (e^{-i(\lambda^+ z_1 - \lambda^- z_2)} n_{0,1} m_{0,2} + e^{i(\lambda^+ z_2 - \lambda^- z_1)} n_{0,4} m_{0,5}), \\
 Q_{11}(x, t) &= \frac{\lambda^- - \lambda^+}{\langle m|n \rangle} (e^{-i\lambda^+ z_1} n_{0,1} m_{0,3} - e^{-i\lambda^- z_1} n_{0,3} m_{0,5}), \\
 Q_{12}(x, t) &= \frac{\lambda^- - \lambda^+}{\langle m|n \rangle} (e^{-i(\lambda^+ z_1 + \lambda^- z_2)} n_{0,1} m_{0,4} + e^{-i(\lambda^- z_1 + \lambda^+ z_2)} n_{0,2} m_{0,5}), \\
 Q_{01}(x, t) &= \frac{\lambda^- - \lambda^+}{\langle m|n \rangle} (e^{-i\lambda^+ z_2} n_{0,2} m_{0,3} + e^{-i\lambda^- z_2} n_{0,3} m_{0,4}), \\
 \langle m|n \rangle &= \sum_{k=1}^5 e^{-i(\lambda^+ - \lambda^-) z_k} n_{0,k} m_{0,k}, \quad z_k = J_k x + I_k t, \quad k = 1, 2.
 \end{aligned} \tag{164}$$

The other four field can be formally constructed by doing the following transformation:

$$Q_{kn} \leftrightarrow Q_{\overline{kn}}, \quad e^{-i\lambda^+ z_k} \leftrightarrow e^{i\lambda^- z_k}, \quad n_{0,j} \leftrightarrow m_{0,j}.$$

For the typical $\mathbb{Z}_2$ reduction (120) of course we must choose $K_0 = \text{diag}(\epsilon_1, \epsilon_2, 1, \epsilon_2, \epsilon_1) \in SO(5)$ where $\epsilon_i = \pm 1$. As a result J and I become real valued and $K_0 Q^\dagger K_0^{-1} = -Q$, i.e.,

$$Q_{\overline{10}} = -\epsilon_1 \epsilon_2 Q_{10}^*, \quad Q_{\overline{12}} = -\epsilon_1 \epsilon_2 Q_{12}^*, \quad Q_{\overline{11}} = -\epsilon_1 Q_{11}^*, \quad Q_{01}^* = -\epsilon_2 Q_{01}. \tag{165}$$

The corresponding 4-wave system takes the form

$$\begin{aligned}
 i(J_1 - J_2)Q_{10,t}(x, t) - i(I_1 - I_2)Q_{10,x}(x, t) - k\epsilon_2 Q_{11}(x, t)Q_{01}^*(x, t) &= 0, \\
 iJ_1 Q_{11,t}(x, t) - iI_1 Q_{11,x}(x, t) - k(Q_{10}Q_{01} + \epsilon_2 Q_{12}Q_{01}^*)(x, t) &= 0, \\
 i(J_1 + J_2)Q_{12,t}(x, t) - i(I_1 + I_2)Q_{12,x}(x, t) - kQ_{11}(x, t)Q_{01}(x, t) &= 0, \\
 iJ_2 Q_{01,t}(x, t) - iI_2 Q_{01,x}(x, t) - k\epsilon_1(Q_{11}^* Q_{12} + \epsilon_2 Q_{10}^* Q_{11})(x, t) &= 0.
 \end{aligned} \tag{166}$$

The particular case $\epsilon_1 = \epsilon_2 = 1$ occurs in Raman scattering [28].

The corresponding one-soliton solution is obtained from (164) imposing $\lambda^+ = (\lambda^-)^* = \mu + i\nu$ and $|m_0\rangle = K_0|n_0^*\rangle$. Here we just note that

$$\langle m|n \rangle = \epsilon_1 (e^{-2\nu z_1} |n_{0,1}|^2 + e^{2\nu z_1} |n_{0,5}|^2) + \epsilon_2 (e^{-2\nu z_2} |n_{0,2}|^2 + e^{2\nu z_2} |n_{0,4}|^2) + |n_{0,3}|^2. \tag{167}$$

Taking $\epsilon_1 = \epsilon_2 = 1$ we find that $\langle m|n \rangle$ is positive for all x and t . If $\epsilon_1 \epsilon_2 = -1$ the product $\langle m|n \rangle$ may vanish for finite x and t , i.e. the corresponding soliton is singular.

5.2. $\mathbb{Z}_2$ reduction related to Weyl group elements

Here we consider several $\mathbb{Z}_2$ reductions related to Weyl group elements. The first one is of the type

$$K_1 U^*(\lambda^*) K_1^{-1} = -U(\lambda) \quad \Rightarrow \quad K_1 Q^* K_1^{-1} = Q, \quad K_1 J^* K_1^{-1} = -J, \tag{168}$$

where K_1 corresponds to the Weyl reflection $S_{e_1 - e_2}$,

$$K_1 = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{169}$$

This reduction leads to the requirements $J_2 = -J_1^*$, $I_2 = -I_1^*$ for J and I respectively, and the following ones for Q ,

$$Q_{\overline{10}} = Q_{10}^*, \quad Q_{01} = -Q_{11}^*, \quad Q_{12}^* = -Q_{12}, \quad Q_{\overline{01}} = -Q_{1\overline{1}}^*, \quad Q_{1\overline{2}}^* = -Q_{1\overline{2}}.$$

The $\mathbb{Z}_2$ reduced integrable system consists of the following five equations for three complex Q_{10} , Q_{11} , $Q_{1\overline{1}}$ and two real fields $Q_{12} = i\mathbf{q}_{12}$, $Q_{1\overline{2}} = i\mathbf{q}_{1\overline{2}}$:

$$\begin{aligned} 2J_{10}Q_{10,t}(x, t) - 2I_{1,0}Q_{10,x}(x, t) - k_0Q_{11}(x, t)Q_{1\overline{1}}^*(x, t) &= 0, \\ J_1Q_{11,t}(x, t) - I_1Q_{11,x}(x, t) + k_0(Q_{10}Q_{11}^* - i\mathbf{q}_{12}Q_{1\overline{1}}^*)(x, t) &= 0, \\ 2J_{1,1}\mathbf{q}_{12,t}(x, t) - 2I_{1,1}\mathbf{q}_{12,x}(x, t) - k_0|Q_{11}(x, t)|^2 &= 0, \\ J_1Q_{1\overline{1},t}(x, t) - I_1Q_{1\overline{1},x}(x, t) + k_0(Q_{10}Q_{1\overline{1}}^* - i\mathbf{q}_{1\overline{2}}Q_{11}^*)(x, t) &= 0, \\ 2J_{1,1}Q_{1\overline{2},t}(x, t) - 2I_{1,1}Q_{1\overline{2},x}(x, t) - k_0|Q_{1\overline{1}}(x, t)|^2 &= 0, \end{aligned} \quad (170)$$

where $J_1 = J_{0,1} + iJ_{1,1}$, $I_1 = I_{0,1} + iI_{1,1}$ and the interaction constant $k_0 = -ik$ is real.

Next we consider a $\mathbb{Z}_2$ reduction of the type

$$K_2 U^\dagger(\lambda^*) K_2^{-1} = U(\lambda) \quad \Rightarrow \quad K_2 J^* K_2^{-1} = J, \quad K_2 Q^\dagger K_2^{-1} = -Q,$$

with $K = S_{e_2}$, i.e.,

$$K_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (171)$$

Therefore we have the following relations:

$$\begin{aligned} J_1^* &= J_1, & J_2^* &= -J_2, & Q_{\overline{10}} &= -Q_{12}^*, & Q_{1\overline{2}} &= -Q_{10}^*, \\ Q_{1\overline{1}} &= Q_{11}^*, & Q_{01}^* &= Q_{01}, & Q_{\overline{01}}^* &= Q_{\overline{01}}. \end{aligned}$$

Thus we derive the following 5-wave system:

$$\begin{aligned} (J_1 - J_2)Q_{10,t}(x, t) - (I_1 - I_2)Q_{10,x}(x, t) + k_0Q_{11}(x, t)Q_{\overline{01}}(x, t) &= 0, \\ J_1Q_{11,t}(x, t) - I_1Q_{11,x}(x, t) - k_0(Q_{10}Q_{01} - Q_{12}Q_{\overline{01}})(x, t) &= 0, \\ (J_1 + J_2)Q_{12,t}(x, t) - (I_1 + I_2)Q_{12,x}(x, t) - k_0Q_{11}(x, t)Q_{01}(x, t) &= 0, \\ J_2Q_{01,t}(x, t) - I_2Q_{01,x}(x, t) + k_0(Q_{11}^*Q_{12} - Q_{1\overline{2}}^*Q_{11})(x, t) &= 0, \\ J_2Q_{\overline{01},t}(x, t) - I_2Q_{\overline{01},x}(x, t) - k_0(Q_{11}Q_{10}^* - Q_{10}Q_{11}^*)(x, t) &= 0, \end{aligned} \quad (172)$$

where $k_0 = -ik$ is real.

The last two reductions requested complex valued J and I . As a result the direct and inverse spectral problems for the corresponding Lax operator L become more complicated [26]. In particular, the continuous spectrum of L fills up a bunch of lines in the complex λ -plane intersecting at the origin. The construction of the corresponding fundamental analytic solutions and the dressing factors requires additional care and will be discussed elsewhere.

5.3. Reductions of MNLS-type systems related to $so(5)$

MNLS equations related to $so(n+2)$ algebras have a Lax representation of the following type:

$$L = i\partial_x + q(x, t) - \lambda J, \quad (173)$$

$$M = i\partial_t + V_0(x, t) + V_1(x, t)\lambda - \lambda^2 J, \quad (174)$$

$$V_1(x, t) = q(x, t), \quad V_0(x, t) = \text{iad}_J \partial_x q + \frac{1}{2} [\text{ad}_J q, q]. \quad (175)$$

Here J is the element of the Cartan subalgebra of $so(n)$ dual to e_1 and

$$q := \begin{pmatrix} 0 & \mathbf{q}^T & 0 \\ \mathbf{p} & 0 & s_0 \mathbf{q} \\ 0 & \mathbf{p}^T s_0 & 0 \end{pmatrix}, \quad J := H_1 = \text{diag}(1, 0, \dots, 0, -1), \quad (176)$$

where

$$\vec{q} = (q_2, \dots, q_{n+1})^T, \quad \vec{p} = (p_2, \dots, p_{n+1})^T.$$

Because of the specific choice of J these MNLS equations can be viewed as connected with the **BD.I** symmetric space. In most of this section we will use the simplest case with $n = 3$, though all results can be easily generalized for any rank of the algebra. The MNLS system itself takes the form

$$\text{i}q_{2,t} + q_{2,xx} + 2q_2^2 p_2 + q_3^2 p_4 + 2q_3 q_2 p_3 = 0, \quad (177)$$

$$\text{i}q_{3,t} + q_{3,xx} + 2q_2 q_4 p_3 + 2q_2 p_2 q_3 + 2q_3 q_4 p_4 + q_3^2 p_3 = 0, \quad (178)$$

$$\text{i}q_{4,t} + q_{4,xx} + 2q_4^2 p_4 + q_3^2 p_2 + 2q_3 q_4 p_3 = 0, \quad (179)$$

$$\text{i}p_{2,t} - p_{2,xx} - 2p_2^2 q_2 - p_3^2 q_4 - 2p_3 p_2 q_3 = 0, \quad (180)$$

$$\text{i}p_{3,t} - p_{3,xx} - 2p_2 p_4 q_3 - 2p_2 q_2 p_3 - 2p_4 q_4 p_3 - p_3^2 q_3 = 0, \quad (181)$$

$$\text{i}p_{4,t} - p_{4,xx} - 2p_4^2 q_4 - p_3^2 q_2 - 2p_3 p_4 q_3 = 0. \quad (182)$$

Its one-soliton solution derived via dressing procedure reads

$$q_k = \frac{\lambda^- - \lambda^+}{\Delta} (e^{-\text{i}\lambda^+(x+\lambda^+t)} n_{0,1} m_{0,k} + (-1)^k e^{-\text{i}\lambda^-(x+\lambda^-t)} n_{0,6-k} m_{0,5}), \quad (183)$$

$$p_k = \frac{\lambda^- - \lambda^+}{\Delta} (e^{\text{i}\lambda^-(x+\lambda^-t)} n_{0,k} m_{0,1} + (-1)^k e^{\text{i}\lambda^+(x+\lambda^+t)} n_{0,5} m_{0,6-k}), \quad (184)$$

where

$$\Delta = e^{\text{i}(\lambda^- - \lambda^+)(x+(\lambda^- + \lambda^+)t)} n_{0,1} m_{0,1} + \sum_{k=2}^4 n_{0,k} m_{0,k} + e^{\text{i}(\lambda^+ - \lambda^-)(x+(\lambda^- + \lambda^+)t)} n_{0,5} m_{0,5}.$$

Consider a $\mathbb{Z}_2$ reduction of the form

$$K U^\dagger(x, \lambda^*) K^{-1} = U(x, \lambda), \quad \text{i.e.} \quad K q^\dagger K^{-1} = q, \quad K J^* K^{-1} = J. \quad (185)$$

If J is real there are two inequivalent choices for K satisfying $K J K^{-1} = J$: the first one is the typical reduction via an element of the Cartan subgroup, $K_1 = \text{diag}(\epsilon_1, \epsilon_2, 1, \epsilon_2, \epsilon_1)$, $\epsilon_{1,2}^2 = 1$; the second one is a reduction via a Weyl reflection S_{e_2} , see (171). The reductions obtained with K_1 and K_2 respectively give

$$p_2 = \epsilon_1 \epsilon_2 q_2^*, \quad p_3 = \epsilon_1 q_3^*, \quad p_4 = \epsilon_1 \epsilon_2 q_4^*; \quad (186)$$

$$p_2 = q_4^*, \quad p_3 = -q_3^*, \quad p_4 = q_2^*. \quad (187)$$

The former $\mathbb{Z}_2$ reduction leads to the following three-component system of NLS equation:

$$\text{i}q_{2,t} + q_{2,xx} + 2\epsilon_1(\epsilon_2 |q_2|^2 + |q_3|^2)q_2 + \epsilon_1 \epsilon_2 q_3^2 q_4^* = 0, \quad (188)$$

$$iq_{3,t} + q_{3,xx} + 2\epsilon_1 q_2 q_4 q_3^* + \epsilon_1 (2\epsilon_2 |q_2|^2 + 2\epsilon_2 |q_4|^2 + |q_3|^2) q_3 = 0, \quad (189)$$

$$iq_{4,t} + q_{4,xx} + 2\epsilon_1 (\epsilon_2 |q_4|^2 + |q_3|^2) q_4 + \epsilon_1 \epsilon_2 q_3^2 q_2^* = 0. \quad (190)$$

In order to integrate this system we apply the dressing procedure with the dressing factor (38). Taking into account that in the reduced case we have the relations

$$\lambda^* = (\lambda^-)^* = \mu + i\nu, \quad |m_0\rangle = K_1 |n_0^*\rangle, \quad (191)$$

we find that its soliton solution is given by

$$q_2 = \frac{-2i\nu}{\Delta} e^{-i\mu(x-\nu t)} (\epsilon_2 e^{\nu(x-\nu t)} n_{0,1} n_{0,2}^* + \epsilon_1 e^{-\nu(x-\nu t)} n_{0,4} n_{0,5}^*), \quad (192)$$

$$q_3 = \frac{-2i\nu}{\Delta} e^{-i\mu(x-\nu t)} (e^{\nu(x-\nu t)} n_{0,1} n_{0,3}^* - \epsilon_1 e^{-\nu(x-\nu t)} n_{0,3} n_{0,5}^*), \quad (193)$$

$$q_4 = \frac{-2i\nu}{\Delta} e^{-i\mu(x-\nu t)} (\epsilon_2 e^{\nu(x-\nu t)} n_{0,1} n_{0,4}^* + \epsilon_1 e^{-\nu(x-\nu t)} n_{0,2} n_{0,5}^*), \quad (194)$$

$$\Delta = \epsilon_1 e^{2\nu(x-\nu t)} |n_{0,1}|^2 + \epsilon_2 (|n_{0,2}|^2 + |n_{0,4}|^2) + |n_{0,3}|^2 + \epsilon_1 e^{-2\nu(x-\nu t)} |n_{0,5}|^2, \quad (195)$$

$$\nu = \frac{\nu^2 - \mu^2}{\mu}, \quad u = -2\mu. \quad (196)$$

The latter reduction gives rise to another inequivalent system of three NLS equations

$$iq_{2,t} + q_{2,xx} + 2(q_2 q_4^* - |q_3|^2) q_2 + q_3^2 q_2^* = 0, \quad (197)$$

$$iq_{3,t} + q_{3,xx} - 2q_2 q_4 q_3^* + (2q_2 q_4^* + 2q_4 q_2^* - |q_3|^2) q_3 = 0, \quad (198)$$

$$iq_{4,t} + q_{4,xx} + 2(q_4 q_2^* - |q_3|^2) q_4 + q_3^2 q_4^* = 0. \quad (199)$$

Then we have the following one-soliton solution:

$$q_2 = \frac{-2i\nu}{\Delta} e^{-i\mu(x-\nu t)} (e^{\nu(x-\nu t)} n_{0,1} n_{0,4}^* + e^{-\nu(x-\nu t)} n_{0,4} n_{0,5}^*), \quad (200)$$

$$q_3 = \frac{2i\nu}{\Delta} e^{-i\mu(x-\nu t)} (e^{\nu(x-\nu t)} n_{0,1} n_{0,3}^* + e^{-\nu(x-\nu t)} n_{0,3} n_{0,5}^*), \quad (201)$$

$$q_4 = \frac{-2i\nu}{\Delta} e^{-i\mu(x-\nu t)} (e^{\nu(x-\nu t)} n_{0,1} n_{0,2}^* + e^{-\nu(x-\nu t)} n_{0,2} n_{0,5}^*), \quad (202)$$

$$\Delta = e^{2\nu(x-\nu t)} |n_{0,1}|^2 + (n_{0,2} n_{0,4}^* + n_{0,2}^* n_{0,4}) - |n_{0,3}|^2 + e^{-2\nu(x-\nu t)} |n_{0,5}|^2. \quad (203)$$

As in the previous example the polarization vectors are interrelated via (191) which in this case reads

$$m_{0,k} = \begin{cases} n_{0,k}^*, & k = 1, 5 \\ (-1)^k n_{0,6-k}^*, & k = 2, 3, 4. \end{cases}$$

Next we consider a $\mathbb{Z}_2 \times \mathbb{Z}_2$ reduction, which is a combination of reductions with K_1 and K_2 . This is possible only for $\epsilon_1 = -1$. Then

$$p_{2,4} = -\epsilon_2 q_{2,4}^*, \quad q_2 = -\epsilon_2 q_4, \quad p_3 = -q_3^*, \quad (204)$$

and we obtain the following two-component NLS system:

$$iq_{2,t} + q_{2,xx} - 2(\epsilon_2 |q_2|^2 + |q_3|^2) q_2 + q_3^2 q_2^* = 0, \quad (205)$$

$$iq_{3,t} + q_{3,xx} - (4\epsilon_2 |q_2|^2 + |q_3|^2) q_3 + 2\epsilon_2 (q_2)^2 q_3^* = 0. \quad (206)$$

Its one-soliton solution takes the form

$$q_2 = \frac{2iv}{\Delta} e^{-i\mu(x-vt)} \epsilon_2 (e^{v(x-ut)} n_{0,1} n_{0,2}^* + e^{-v(x-ut)} n_{0,2} n_{0,5}^*), \quad (207)$$

$$q_3 = \frac{2iv}{\Delta} e^{-i\mu(x-vt)} (e^{v(x-ut)} n_{0,1} n_{0,3}^* + e^{-v(x-ut)} n_{0,3} n_{0,5}^*), \quad (208)$$

$$\Delta = e^{2v(x-ut)} |n_{0,1}|^2 - 2\epsilon_2 |n_{0,2}|^2 - |n_{0,3}|^2 + e^{-2v(x-ut)} |n_{0,5}|^2, \quad (209)$$

$$v = \frac{v^2 - \mu^2}{\mu}, \quad u = -2\mu, \quad (210)$$

where we have made use of the following relations:

$$m_{0,k} = \begin{cases} n_{0,k}^*, & k = 1, 5 \\ -\epsilon_2 n_{0,k}^* = n_{0,6-k}^*, & k = 2, 4, \\ -n_{0,3}^*, & k = 3. \end{cases}$$

5.4. MMKdV equations on symmetric spaces of *BD.I* type

Multicomponent MKdV related to $so(n+2)$ algebra possesses a Lax pair with the L operator (173) and M -operator in the form

$$\begin{aligned} M\psi(x, t, \lambda) &\equiv i\partial_t \psi + (V_0(x, t) + \lambda V_1(x, t) + \lambda^2 V_2(x, t) - \lambda^3 J)\psi(x, t, \lambda), \\ V_2(x, t) &= q(x, t), \quad V_1(x, t) = i\text{ad}_J^{-1} \partial_x q + \frac{1}{2} [\text{ad}_J^{-1} q, q(x, t)], \\ V_0(x, t) &= -\partial_{xx}^2 q + \frac{1}{2} [\text{ad}_J^{-1} q, [\text{ad}_J^{-1} q, q(x, t)]] + i[\partial_x q, q]. \end{aligned} \quad (211)$$

The corresponding MMKdV equations can be written down in compact form as

$$\partial_t \bar{q} + \partial_{xxx}^3 \bar{q} + 3(\bar{p}, \bar{q}) \partial_x \bar{q} + 3(\partial_x \bar{q}, \bar{p}) \bar{q} - 3(\partial_x \bar{q} s_0 \bar{q}) s_0 \bar{p} = 0, \quad (212)$$

$$\partial_t \bar{p} + \partial_{xxx}^3 \bar{p} + 3(\bar{p}, \bar{q}) \partial_x \bar{p} + 3(\partial_x \bar{p}, \bar{q}) \bar{p} - 3(\partial_x \bar{p} s_0 \bar{p}) s_0 \bar{q} = 0. \quad (213)$$

Consider a $\mathbb{Z}_2$ reduction of the type

$$U^\dagger(\lambda^*) = U(\lambda), \quad \Rightarrow \quad \bar{p} = \bar{q}^*. \quad (214)$$

Then we obtain the reduced system of equations

$$\partial_t \bar{q} + \partial_{xxx}^3 \bar{q} + 3|\bar{q}|^2 \partial_x \bar{q} + 3(\partial_x \bar{q}, \bar{q}^*) \bar{q} - 3(\partial_x \bar{q} s_0 \bar{q}) s_0 \bar{q}^* = 0. \quad (215)$$

Its one-soliton solution reads

$$\begin{aligned} q_k &= \frac{-iv e^{-i\mu(x-ut-\delta_0)}}{\cosh(2v(x-vt-\xi_0)) + C} (e^{v(x-vt-\xi_0)} c_k^* + (-1)^k e^{-v(x-vt-\xi_0)} c_{n+3-k}), \\ c_k &= \frac{n_{0,k}}{\sqrt{|n_{0,1}| |n_{0,n+2}|}}, \quad k = 2, \dots, n+1 \quad C = \sum_{k=2}^{n+1} \frac{|n_{0,k}|^2}{2|n_{0,1}| |n_{0,n+2}|}, \\ v &= v^2 - 3\mu^2, \quad u = 3v^2 - \mu^2, \quad \delta_0 = \frac{\arg n_{0,1}}{\mu}, \quad \xi_0 = \frac{1}{2v} \ln \frac{|n_{0,n+2}|}{|n_{0,1}|}, \end{aligned} \quad (216)$$

provided we have fixed $\arg n_{0,1} = -\arg n_{0,n+2}$ by using the natural $U(1)$ symmetry of the solution.

As a final example let us consider MKdV related to $so(5)$ with a $\mathbb{Z}_2$ reduction of the type

$$KU^\dagger(-\lambda^*)K^{-1} = U(\lambda) \quad \Rightarrow \quad Kq^\dagger K^{-1} = q, \quad KJK^{-1} = -J. \quad (217)$$

We choose $K = K_0 \circ W_{e_1}$ where

$$W_{e_1} = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \Rightarrow \quad K = \begin{pmatrix} 0 & 0 & 0 & 0 & -\epsilon_1 \\ 0 & \epsilon_2 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & \epsilon_2 & 0 \\ -\epsilon_1 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (218)$$

is the Weyl reflection with respect to the hyperplane orthogonal to e_1 . The following interrelations hold true:

$$q_4 = -\epsilon_1 \epsilon_2 q_2^*, \quad q_3 = -\epsilon_1 q_3^*, \quad p_4 = -\epsilon_1 \epsilon_2 p_2^*, \quad p_3 = -\epsilon_1 p_3^*. \quad (219)$$

As a consequence of the reduction we have

$$\lambda^+ = -(\lambda^-)^*, \quad |m\rangle = K|n\rangle^*, \quad (220)$$

or

$$(\lambda^\pm)^* = -\lambda^\pm, \quad |n\rangle = SK|n\rangle^*, \quad \langle m| = \langle m|^*(SK)^{-1}. \quad (221)$$

Applying another $\mathbb{Z}_2$ reduction of the type

$$U^T(-\lambda) = -U(\lambda), \quad \Rightarrow \quad q^T = -q, \quad (222)$$

we obtain that

$$\lambda^+ = -\lambda^-, \quad |m\rangle = |n\rangle. \quad (223)$$

The corresponding system of MKdV is

$$q_{2,t} + q_{2,xxx} - 3(q_2 q_3)_x q_3 + 3\epsilon_1 \epsilon_2 q_3 q_2^* q_{3,x} - 6q_2^2 q_{2,x} = 0, \quad (224)$$

$$q_{3,t} + q_{3,xxx} + 3\epsilon_1 \epsilon_2 |q_2|^2 q_3 - 3(q_2 q_3)_x q_2 - 3(q_2^* q_3)_x q_2^* - 3q_3^2 q_{3,x} = 0, \quad (225)$$

and is new to the best of our knowledge. Combining both reductions we again may have two types of solitons. The doublet soliton corresponds to purely imaginary $\lambda^\pm = \pm i\nu$ and $|n\rangle = SK|n\rangle^*$ and is given by

$$q_2 = \frac{i\nu e^{i\delta_0}}{\epsilon_1 \cosh 2\nu(x - vt - \xi_0) + C} (e^{\nu(x-vt-\xi_0)} c_2 + e^{-\nu(x-vt-\xi_0)} c_4), \quad (226)$$

$$q_3 = \frac{2i\nu c_3 e^{i\delta_0} \sinh \nu(x - vt - \xi_0)}{\epsilon_1 \cosh 2\nu(x - vt - \xi_0) + C},$$

$$\delta_0 = \arg n_{0,1} = \arg n_{0,5} = \frac{l\pi}{2}, \quad l \in \mathbb{Z}, \quad C = \frac{2\epsilon_2 \operatorname{Re}(n_{0,2} n_{0,4}) + |n_{0,3}|^2}{2|n_{0,1}| |n_{0,5}|}, \quad (227)$$

$$c_1^* = -\epsilon_1 c_1, \quad c_2^* = -\epsilon_2 c_4, \quad c_3^* = -c_3, \quad c_5^* = -\epsilon_1 c_5,$$

where c_k , ν and ξ_0 coincide with those in the previous example when $r = 2$. From (219) it follows that q_3 is either real or purely imaginary valued function.

Yet another possibility to ensure compatibility of both $\mathbb{Z}_2$ reductions is to modify the dressing factor into

$$u(x, t, \lambda) = \mathbb{I} + \frac{A(x, t)}{\lambda - \lambda_0} - \frac{KSA^*(x, t)SK}{\lambda + \lambda_0^*} - \frac{SA(x, t)S}{\lambda + \lambda_0} + \frac{KA^*(x, t)K}{\lambda - \lambda_0^*}, \quad (228)$$

which leads to the quadruplet soliton solution

$$q(x, t) = [J, A - KSA^*SK - SAS + KA^*K](x, t). \quad (229)$$

In order to find it we have to calculate the matrix $A(x, t)$. For that purpose it proves to be convenient to decompose A into $A = XF^T$, where X and F are rectangular matrices of rank $s \leq r$ and $\lambda_0 = \mu + i\nu$. It can be checked that $F(x, t) = e^{i(\lambda_0 x + \lambda_0^2 t)} F_0$, $F_0 = \text{const}$. In the simplest $s = 1$ case for the factor X one can obtain the following:

$$X = \frac{a^* F + bSKF^* - cKF^*}{|a|^2 + b^2 - c^2},$$

where

$$\begin{aligned} a &= \frac{F^T F}{2\lambda_0} = \frac{|F_{0,1} F_{0,5}|}{\lambda_0} (\cosh 2(\phi_R - i\phi_1) + C_a), & C_a &= \frac{F_{0,2}^2 + F_{0,3}^2 + F_{0,4}^2}{2|F_{0,1} F_{0,5}|}, \\ b &= \frac{(F^\dagger SKF)}{2i\nu} = \frac{i|F_{0,1} F_{0,5}|}{\nu} (\cosh 2\phi_R + C_b), & C_b &= \frac{2\text{Re}(F_{0,2}^* F_{0,4}) + |F_{0,3}|^2}{2|F_{0,1} F_{0,5}|}, \\ c &= \frac{(F^\dagger KF)}{2\mu} = \frac{|F_{0,1} F_{0,5}|}{\mu} (\cos 2\phi_1 + C_c), & C_c &= \frac{|F_{0,2}|^2 - |F_{0,3}|^2 + |F_{0,4}|^2}{2|F_{0,1} F_{0,5}|}, \\ \phi_R &= \nu \left(x - \nu t - \frac{1}{2\nu} \ln \frac{|F_{0,1}|}{|F_{0,5}|} \right), & \phi_1 &= \mu \left(x - \nu t - \frac{\arg F_{0,5}}{\mu} \right), \\ & \arg F_{0,1} = -\arg F_{0,5}. \end{aligned}$$

Substituting this result into (229) and choosing $\epsilon_1 = \epsilon_2 = 1$ one derives

$$q_2 = \frac{2\sqrt{|F_{0,1} F_{0,2} F_{0,4} F_{0,5}|}}{|a|^2 + b^2 - c^2} \{ a^* \cosh(\phi_R^- - i\phi_1^-) - b(\cosh(\phi_R^- + i\phi_1^+) + \cosh(\phi_R^+ - i\phi_1^-)) \\ - a \cosh(\phi_R^+ + i\phi_1^+) + c(\cosh(\phi_R^+ + i\phi_1^-) - \cosh(\phi_R^- - i\phi_1^+)) \}, \quad (230)$$

$$q_3 = \frac{2i\sqrt{|F_{0,1} F_{0,5}|}}{|a|^2 + b^2 - c^2} \text{Im}\{(b+c) \sinh(\phi_R + i\phi_1) - a^* \sinh(\phi_R - i\phi_1)\} F_{0,3}, \quad (231)$$

where we have used the following notation:

$$\phi_R^\pm = \phi_R \pm \frac{1}{2} \ln \frac{|F_{0,2}|}{|F_{0,4}|}, \quad \phi_1^\pm = \phi_1 \pm \arg F_{0,4}. \quad (232)$$

6. Discussion and further studies

Here we shall outline some further topics which could be studied and which could lead to a deeper understanding of these soliton properties.

The first obvious remark is that $sl(n)$ contains as subalgebras also $so(p)$ and $sp(p)$ subalgebras. So it will be interesting to specify the conditions under which $L(\lambda)$ has solitons of type $so(p)$ or $sp(p)$.

The second remark of the same nature is that one can start with $L(\lambda)$ related to any simple Lie algebra $\mathfrak{g}$ (e.g., $so(n)$ or $sp(n)$). The analysis of soliton solutions for NLEE related to such systems require deeper knowledge of Lie algebras and their representations.

The explicit form of the corresponding N -wave system related to these algebras has been reported in [4, 5, 13], see also [29, 30]. What could be done is to analyze the structure of its soliton solutions [29, 30] which are more involved, especially in the case when additional reductions are imposed. The construction of the corresponding dressing factors is more complicated, but one can expect that new interesting types of integrable cubic and quartic interactions could be obtained.

Another important question is the study of solitons constructed by projectors of rank 2 and higher. Such solitons have been already used by Wadati *et al* [31] in describing BEC

with hyperfine structure $F = 1$. Now each soliton will be parameterized by two polarization vectors; the corresponding eigensubspace will be two dimensional too. Among the various types of rank-2 one-soliton solutions, there will be various possible configurations for the two polarization vectors. An example of a dressing factor $u(x, t, \lambda)$ constructed by a projector of second rank is the following one:

$$u(x, t, \lambda) = \mathbb{1} + \left(\frac{1}{\sqrt{c(\lambda)}} - 1 \right) P_2 + (\sqrt{c(\lambda)} - 1) \bar{P}_2.$$

Such type of a dressing factor was used in [32] to derive the soliton solutions to a multicomponent Schrödinger equation relate to symplectic algebra $sp(4)$.

It is known in general how the machinery, well understood for the AKNS system such as Wronskian relations, expansions over 'squared solutions', etc can be generalized also for these types of systems. The dressing method, after some modifications, can also be applied, leading to the derivation of their soliton solutions.

An interesting problem is the study of how the different possible reductions (see, e.g., [29]) of these systems will influence the number of one-soliton types.

Soliton interactions for various different types of solitons of these systems also present interesting problems. From the results known for the N -wave systems [9, 10] it is known that new effects in soliton interaction, such as soliton decay and soliton fusion may arise.

Acknowledgments

This research has been supported in part by USA Air Force Office of Scientific Research, the National Science Foundation of the USA via grant DMS-0505566 and by the National Science Foundation of Bulgaria via grant 1410.

Appendix A. The orthogonal algebras and BD.I symmetric spaces

The definition of orthogonality used in equations (159)–(161) has the advantage that the Cartan subalgebras can be represented by diagonal matrices.

It is well known that the two cases of odd and even n are substantially different. The algebras $so(2r+1)$ are known as the $\mathbf{B}_r$ series according to the Cartan classification [33]. Assuming that the reader is familiar with the basics of simple Lie algebras here we just recall that the root systems of these algebras consist of $\Delta^- \cup \Delta^+$ where $\Delta^- = -\Delta^+$ and $\Delta^+ \equiv \{e_i \pm e_j, e_j\}$ with $1 \leq i < j \leq r$. We provide also the Cartan–Weyl basis of these algebras in the typical $(2r+1)$ -dimensional representation

$$\begin{aligned} E_{e_i - e_j} &\equiv E_{ij} - S_0 E_{ji} S_0 = E_{i,j} - (-1)^{i+j} E_{\bar{j},\bar{i}}, \\ E_{e_i + e_j} &\equiv E_{i\bar{j}} - S_0 E_{\bar{j}i} S_0 = E_{i,\bar{j}} - (-1)^{i+j} E_{j,\bar{i}}, \\ E_{e_i} &\equiv E_{i,r+1} - S_0 E_{r+1,i} S_0 = E_{i,r+1} - (-1)^{i+r+1} E_{r+1,\bar{i}}, \\ H_{e_j} &\equiv E_{jj} - S_0 E_{j,j} S_0 = E_{j,j} - E_{\bar{j},\bar{j}}, \quad E_{-\alpha} = E_{\alpha}^T. \end{aligned} \quad (\text{A.1})$$

Each symmetric space is obtained by applying a Cartan involution. This involution splits the group $SO(2r+1)$ into a subgroup $SO(2r-1) \otimes SO(2)$ and a factor space $SO(2r+1)/SO(2r-1) \otimes SO(2)$. Effectively the system of positive roots is split into $\Delta^+ = \Delta_0^+ \cup \Delta_1^+$, where the subsets of roots are defined as follows:

$$\Delta_0^+ = \{e_i \pm e_j, \quad e_i, \quad 2 \leq i < j \leq r\}, \quad \Delta_1^+ = \{e_1 \pm e_j, \quad e_1, \quad 2 \leq j \leq r\}. \quad (\text{A.2})$$

In fact Δ_0^+ contains all positive roots that are orthogonal to e_1 , while Δ_1^+ contains all positive roots that have scalar product equal to 1 with e_1 .

Similarly one can consider the algebras $so(2r)$ which are known as the $\mathbf{D}_r$ series. Their Cartan–Weyl basis in the typical $2r$ -dimensional representation is given by

$$\begin{aligned} E_{e_i - e_j} &\equiv E_{ij} - S_0 E_{ji} S_0 = E_{i,j} - (-1)^{i+j} E_{\bar{j},\bar{i}}, \\ E_{e_i + e_j} &\equiv E_{i\bar{j}} - S_0 E_{\bar{j}i} S_0 = E_{i,\bar{j}} - (-1)^{i+j} E_{j,\bar{i}}, \\ H_{e_j} &\equiv E_{jj} - S_0 E_{j,j} S_0 = E_{j,j} - E_{\bar{j},\bar{j}}, \quad E_{-\alpha} = E_{\alpha}^T. \end{aligned} \quad (\text{A.3})$$

The corresponding symmetric space is $SO(2r+2)/SO(2r) \otimes SO(2)$. The system of positive roots is split into $\Delta^+ = \Delta_0^+ \cup \Delta_1^+$, where

$$\Delta_0^+ = \{e_i \pm e_j, \quad 2 \leq i < j \leq r\}, \quad \Delta_1^+ = \{e_1 \pm e_j, \quad 2 \leq j \leq r\}. \quad (\text{A.4})$$

Again Δ_0^+ contains all positive roots that are orthogonal to e_1 , while Δ_1^+ contains all positive roots that have scalar product equal to 1 with e_1 .

Appendix B. Higher spin representations of $sl(2)$

Here we construct the $sl(2)$ dressing factor (130) for higher spin representation of $sl(2)$. This we do by using completely symmetric tensor powers of u .

Using the way how u acts on basic vectors e_i , $i = 1, 2$ in $\mathbb{C}_2$

$$ue_1 = u_{11}e_1 + u_{21}e_2, \quad ue_2 = u_{12}e_1 + u_{22}e_2. \quad (\text{B.1})$$

The normalized basis in the completely symmetrized tensor product of $\mathbb{C}^2 \odot \mathbb{C}^2$ is given by

$$\epsilon_1 = e_1 \otimes e_1, \quad \epsilon_2 = \frac{1}{\sqrt{2}}(e_2 \otimes e_1 + e_1 \otimes e_2), \quad \epsilon_3 = e_2 \otimes e_2. \quad (\text{B.2})$$

Since $u(x, \lambda)$ belongs to the group $SL(2)$ it must act on the basic elements as follows:

$$U^{(3)}(x, \lambda)(e_i \otimes e_j) = (ue_i) \otimes (ue_j). \quad (\text{B.3})$$

Thus we obtain the following representation for the dressing factor for spin-1 representation

$$U^{(3)} \equiv u \odot u = \begin{pmatrix} u_{11}^2 & \sqrt{2}u_{11}u_{12} & u_{12}^2 \\ \sqrt{2}u_{11}u_{21} & u_{11}u_{22} + u_{12}u_{21} & \sqrt{2}u_{22}u_{12} \\ u_{21}^2 & \sqrt{2}u_{22}u_{21} & u_{22}^2 \end{pmatrix}. \quad (\text{B.4})$$

Now we have to insert the expressions for u_{ij} in terms of c_1, n_k and m_k and thus we derive

$$U^{(3)} \equiv u \odot u = c_1 \pi_1^{(3)} + \pi_0^{(3)} + \frac{1}{c_1} \pi_{-1}^{(3)}, \quad (\text{B.5})$$

where the projectors π_a , $a = -1, 0, 1$ are all rank-1 projectors of the form

$$\pi_1^{(3)} = \frac{|N_1^{(3)}\rangle\langle M_1^{(3)}|}{\langle M_1^{(3)}|N_1^{(3)}\rangle}, \quad \pi_0^{(3)} = \frac{|N_0^{(3)}\rangle\langle M_0^{(3)}|}{\langle M_0^{(3)}|N_0^{(3)}\rangle}, \quad \pi_{-1}^{(3)} = \frac{|N_{-1}^{(3)}\rangle\langle M_{-1}^{(3)}|}{\langle M_{-1}^{(3)}|N_{-1}^{(3)}\rangle}, \quad (\text{B.6})$$

where $\langle M_1^{(3)}|N_1^{(3)}\rangle = \langle M_0^{(3)}|N_0^{(3)}\rangle = \langle M_{-1}^{(3)}|N_{-1}^{(3)}\rangle = (m_1 n_1 + m_2 n_2)^2$

$$\begin{aligned} |N_1^{(3)}\rangle &= \begin{pmatrix} n_1^2 \\ \sqrt{2}n_1 n_2 \\ n_2^2 \end{pmatrix}, \quad |N_{-1}^{(3)}\rangle = \begin{pmatrix} m_2^2 \\ -\sqrt{2}m_1 m_2 \\ m_1^2 \end{pmatrix}, \quad |N_0^{(3)}\rangle = \begin{pmatrix} \sqrt{2}m_2 n_1 \\ n_2 m_2 - n_1 m_1 \\ -\sqrt{2}n_2 m_1 \end{pmatrix}, \\ \langle M_1^{(3)}| &= (m_1^2, \sqrt{2}m_1 m_2, m_2^2), \quad \langle M_{-1}^{(3)}| = (n_2^2, -\sqrt{2}n_1 n_2, n_1^2), \\ \langle M_0^{(3)}| &= (\sqrt{2}n_2 m_1, n_2 m_2 - n_1 m_1, -\sqrt{2}n_1 m_2). \end{aligned} \quad (\text{B.7})$$

Similarly the normalized basis in the completely symmetrized tensor products of $\mathbb{C}^2 \odot \mathbb{C}^2 \odot \mathbb{C}^2$ is given by

$$\begin{aligned}\epsilon'_1 &= e_1 \otimes e_1 \otimes e_1, & \epsilon'_2 &= \frac{1}{\sqrt{3}}(e_1 \otimes e_1 \otimes e_2 + e_1 \otimes e_2 \otimes e_1 + e_2 \otimes e_1 \otimes e_1), \\ \epsilon'_3 &= \frac{1}{\sqrt{3}}(e_1 \otimes e_2 \otimes e_2 + e_2 \otimes e_1 \otimes e_2 + e_2 \otimes e_2 \otimes e_1), & \epsilon'_4 &= e_2 \otimes e_2 \otimes e_2.\end{aligned}\quad (\text{B.8})$$

Therefore the dressing factor obtains the form

$$U^{(4)} = \begin{pmatrix} u_{11}^3 & \sqrt{3}u_{11}^2u_{12} & \sqrt{3}u_{11}u_{12}^2 & u_{12}^3 \\ \sqrt{3}u_{11}^2u_{21} & u_{11}^2u_{22} + 2u_{11}u_{12}u_{21} & u_{12}^2u_{21} + 2u_{11}u_{12}u_{22} & \sqrt{3}u_{12}^2u_{22} \\ \sqrt{3}u_{11}u_{21}^2 & u_{12}u_{21}^2 + 2u_{11}u_{21}u_{22} & u_{11}u_{22}^2 + 2u_{12}u_{21}u_{22} & \sqrt{3}u_{12}u_{22}^2 \\ u_{21}^3 & \sqrt{3}u_{21}^2u_{22} & \sqrt{3}u_{21}u_{22}^2 & u_{22}^3 \end{pmatrix}. \quad (\text{B.9})$$

It can be decomposed into

$$U^{(4)} = \pi_{-3/2}^{(4)}c^{-3/2} + \pi_{-1/2}^{(4)}c^{-1/2} + \pi_{1/2}^{(4)}c^{1/2} + \pi_{3/2}^{(4)}c^{3/2} \quad (\text{B.10})$$

where

$$\pi_a^{(4)} = \frac{|N_a^{(4)}\rangle\langle M_a^{(4)}|}{\langle M_a^{(4)}|N_a^{(4)}\rangle}, \quad a = -3/2, -1/2, 1/2, 3/2. \quad (\text{B.11})$$

The (co) vectors $|N_a^{(4)}\rangle$ ($|M_a^{(4)}\rangle$) are given by

$$\begin{aligned}|N_{-3/2}^{(4)}\rangle &= (m_2^3, -\sqrt{3}m_2^2m_1, \sqrt{3}m_2m_1^2, -m_1^3)^T, \\ \langle M_{-3/2}^{(4)}| &= (n_2^3, -\sqrt{3}n_2^2n_1, \sqrt{3}n_2n_1^2, -n_1^3) \\ |N_{-1/2}^{(4)}\rangle &= (\sqrt{3}n_1m_2^2, m_2(n_2m_2 - 2n_1m_1), m_1(n_1m_1 - 2n_2m_2), \sqrt{3}n_2m_1^2)^T, \\ \langle M_{-1/2}^{(4)}| &= (\sqrt{3}n_2^2m_1, n_2(n_2m_2 - 2n_1m_1), n_1(n_1m_1 - 2n_2m_2), \sqrt{3}n_1^2m_2) \\ |N_{1/2}^{(4)}\rangle &= (\sqrt{3}n_2^2m_2, -n_1(n_1m_1 - 2n_2m_2), n_2(n_2m_2 - 2n_1m_1), -\sqrt{3}n_2^2m_1)^T, \\ \langle M_{1/2}^{(4)}| &= (\sqrt{3}n_2m_1^2, m_1(n_1m_1 - 2n_2m_2), m_2(n_2m_2 - 2n_1m_1), -\sqrt{3}n_1m_2^2) \\ |N_{3/2}^{(4)}\rangle &= (n_1^3, \sqrt{3}n_1^2n_2, \sqrt{3}n_1n_2^2, n_2^3)^T, \\ \langle M_{3/2}^{(4)}| &= (m_1^3, \sqrt{3}m_1^2m_2, \sqrt{3}m_1m_2^2, m_2^3).\end{aligned}$$

References

- [1] Calogero F and Degasperis A 1982 *Spectral Transform and Solitons* vol 1 (Amsterdam: North-Holland)
- [2] Faddeev L D and Takhtadjan L A 1987 *Hamiltonian Methods in the Theory of Solitons* (Berlin: Springer)
- [3] Zakharov V E, Manakov S V, Novikov S P and Pitaevskii L I 1984 *Theory of Solitons: The Inverse Scattering Method* (New York: Plenum)
- [4] Harnad J, Saint-Auben Y and Schneider S 1984 Bäcklund transformations for nonlinear sigma models with values in Riemannian symmetric spaces *Commun. Math. Phys.* **92** 329–87
- [5] Zakharov V E and Mikhailov A V 1978 Relativistically invariant two-dimensional models of field theory which are integrable by means of the inverse scattering problem method *Zh. Eksp. Teor. Fiz.* **74** 1953–73
Zakharov V E and Mikhailov A V 1980 On the integrability of classical spinor models in two-dimensional space-time *Commun. Math. Phys.* **74** 21–40
- [6] Zakharov V E and Shabat A B 1972 Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media *Sov. Phys.—JETP* **34** 62
- [7] Zakharov V E and Shabat A B 1974 A scheme for integrating nonlinear equations of mathematical physics by the method of the inverse scattering transform I *Funct. Anal. Appl.* **8** 43–53
Zakharov V E and Shabat A B 1979 A scheme for integrating nonlinear equations of mathematical physics by the method of the inverse scattering transform II *Funct. Anal. Appl.* **13** 13–23

- [8] Gerdjikov V S and Kaup D J 2006 How many types of soliton solutions do we know? *Proc. 7th Int. Conf. on Geometry, Integrability and Quantization (June 2–10 2005, Varna)* vol 7 ed I Mladenov and M de Leon (Sofia: Softex) pp 11–34
- [9] Zakharov V E and Manakov S V 1975 The theory of resonant interaction of wave packets in nonlinear media *Sov. Phys.—JETP* **69** 1654 (in Russian)
- [10] Kaup D J 1981 The solution of a general initial value problem for the full three dimensional three-wave resonant interaction *Physica D* **3** 374–95
- Kaup D J, Reiman A and Bers A 1979 Space-time evolution of nonlinear three-wave interactions: I. Interactions in an Homogeneous Medium *Rev. Mod. Phys.* **51** 275–310
- [11] Mikhailov A V 1981 The reduction problem and the inverse scattering problem *Physica D* **3D** 73–117
- [12] Gerdjikov V S, Kostov N A and Valchev T I 2007 N -wave equations with orthogonal algebras: $\mathbb{Z}_2$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ reductions and soliton solutions *Sigma* **3** 039
- [13] Fordy A P and Kulish P P 1983 Nonlinear Schrödinger equations and simple Lie algebras *Commun. Math. Phys.* **89** 427–43
- [14] Athorne C and Fordy A 1986 Generalised KdV and MKdV equations associated with symmetric spaces *J. Phys. A: Math. Gen.* **20** 1377–86
- [15] Harnad J, Saint-Aubin Y and Shnider S 1984 Backlund transformations for nonlinear sigma models with values in Riemannian symmetric spaces *Commun. Math. Phys.* **92** 329–67
- Harnad J, Saint-Aubin Y and Shnider S 1984 The soliton correlation matrix and the reduction problem for integrable systems *Commun. Math. Phys.* **93** 33–56
- [16] Sokolov V and Wolf Th 2001 Classification of integrable polynomial vector evolution equations *J. Phys. A: Math. Gen.* **34** 11139–48
- [17] Gerdjikov V S and Kulish P P 1983 Expansions over the “Squared” eigenfunctions of the matrix linear $n \times n$ *J. Sov. Math.* **23** 2400–12
- Gerdjikov V S and Kulish P P 1981 The generating operator for the $n \times n$ linear system *Physica D* **3** 549–64
- [18] Gerdjikov V S 1982 On the spectral theory of the integro-differential operator Λ , generating nonlinear evolution equations *Lett. Math. Phys.* **6** 315–24
- [19] Shabat A B 1975 The inverse scattering problem for a system of differential equations *Funct. Anal. Appl.* **9** 75–8 (in Russian)
- Shabat A B 1979 The inverse scattering problem *Diff. Eqns* **15** 1824–34 (in Russian)
- [20] Ablowitz M J, Kaup D J, Newell A C and Segur H 1974 The inverse scattering transform—Fourier analysis for nonlinear problems *Stud. Appl. Math.* **53** 249–315
- [21] Kaup D J and Newell A C 1979 Soliton equations, singular dispersion relations and moving eigenvalues *Adv. Math.* **31** 67–100
- [22] Gerdjikov V S and Khristov E Kh 1980 On the evolution equations solvable with the inverse scattering problem: I. The spectral theory *Bulg. J. Phys.* **7** 28–41 (in Russian)
- Gerdjikov V S and Khristov E Kh 1980 On the evolution equations solvable with the inverse scattering problem: II. Hamiltonian structures and Bäcklund transformations *Bulg. J. Phys.* **7** 119–33 (in Russian)
- [23] Kaup D J 1976 Closure of the squared Zakharov–Shabat eigenstates *J. Math. Anal. Appl.* **54** 849–64
- Kaup D J 1976 The three wave interaction—a non-dispersive phenomenon *Stud. Appl. Math.* **55** 9–44
- [24] Evans J 1972 Nerve axon equations: I. Linear approximation *Indiana Univ. Math. J.* **21** 877–955
- Evans J 1973 Nerve axon equations: II. Stability at rest *Indiana Univ. Math. J.* **22** 75–90
- Evans J 1975 Nerve axon equations: III. Stability of the nerve impulse *Indiana Univ. Math. J.* **22** 577–94
- Evans J 1975 Nerve axon equations: IV. The stable and unstable impulse *Indiana Univ. Math. J.* **24** 1169–90
- [25] Ablowitz M J, Kaup D J, Newell A C and Segur H 1973 Method for solving the sine-Gordon equation *Phys. Rev. Lett.* **30** 1262–4
- [26] Gerdjikov V S and Yanovski A 1994 Completeness of the eigenfunctions for the Caudrey–Beals–Coifman system *J. Math. Phys.* **35** 3687–725
- [27] Gerdjikov V, Grahovski G and Kostov N 2000 Second order reductions of N -wave interactions related to low-rank simple Lie algebras *Proc. Int. Conf. on Geometry, Integrability and Quantization (Varna)* ed I Mladenov and G Naber (Sofia: Coral Press) pp 55–77
- Gerdjikov V S, Grahovski G G and Kostov N A 2001 Reductions of N -wave interactions related to low-rank simple Lie algebras: I. $\mathbb{Z}_2$ -reductions *J. Phys. A: Math. Gen.* **34** 9425–61
- [28] Gerdjikov V S and Kostov N A 1996 Inverse scattering transform analysis of Stokes–anti Stokes stimulated Raman scattering *Phys. Rev. A* **54** 4339–50 (Preprint pait-sol/9502001)

- [29] Gerdjikov V S, Grahovski G G, Ivanov R I and Kostov N A 2001 N -wave interactions related to simple Lie algebras: $\mathbb{Z}_2$ -reductions and soliton solutions *Inverse Problems* **17** 999–1015
- [30] Ivanov R 2004 On the dressing method for the generalized Zakharov–Shabat system *Nucl. Phys. B* **694** 509–24
- [31] Ieda J, Miyakawa T and Wadati M 2004 Exact analysis of soliton dynamics for Bose–Einstein condensates *Phys. Rev. Lett.* **93** 194102
- [32] Atanasov V A, Gerdjikov V S, Grahovski G G and Kostov N A 2008 Fordy–Kulish Models and spinor Bose–Einstein condensates *Preprint nlin.SI/0802.4405*
- [33] Helgason S 1978 *Differential Geometry, Lie Groups and Symmetric Spaces* (Toronto: Academic)

Squared eigenfunctions for the Sasa–Satsuma equation

Jianke Yang^{1,a)} and D. J. Kaup²

¹*Department of Mathematics and Statistics, University of Vermont, Burlington, Vermont 05401, USA*

²*Department of Mathematics, University of Central Florida, Orlando, Florida 32816, USA*

(Received 5 September 2008; accepted 29 December 2008;

published online 5 February 2009)

Squared eigenfunctions are quadratic combinations of Jost functions and adjoint Jost functions which satisfy the linearized equation of an integrable equation. They are needed for various studies related to integrable equations, such as the development of its soliton perturbation theory. In this article, squared eigenfunctions are derived for the Sasa–Satsuma equation whose spectral operator is a 3×3 system, while its linearized operator is a 2×2 system. It is shown that these squared eigenfunctions are sums of two terms, where each term is a product of a Jost function and an adjoint Jost function. The procedure of this derivation consists of two steps: First is to calculate the variations of the potentials via variations of the scattering data by the Riemann–Hilbert method. The second one is to calculate the variations of the scattering data via the variations of the potentials through elementary calculations. While this procedure has been used before on other integrable equations, it is shown here, for the first time, that for a general integrable equation, the functions appearing in these variation relations are precisely the squared eigenfunctions and adjoint squared eigenfunctions satisfying, respectively, the linearized equation and the adjoint linearized equation of the integrable system. This proof clarifies this procedure and provides a unified explanation for previous results of squared eigenfunctions on individual integrable equations. This procedure uses primarily the spectral operator of the Lax pair. Thus two equations in the same integrable hierarchy will share the same squared eigenfunctions (except for a time-dependent factor). In the Appendix, the squared eigenfunctions are presented for the Manakov equations whose spectral operator is closely related to that of the Sasa–Satsuma equation. © 2009 American Institute of Physics.

[DOI: 10.1063/1.3075567]

I. INTRODUCTION

Squared eigenfunctions are quadratic combinations of Jost functions and adjoint Jost functions which satisfy the linearized equation of an integrable system. This name was derived from the fact that, for many familiar integrable equations such as the Korteweg-de Vries (KdV) and nonlinear Schrödinger (NLS) equations, solutions of the linearized equations are squares of Jost functions of the Lax pairs.^{1–7} Squared eigenfunctions are intimately related to the integrable equation theory. For instance, they are eigenfunctions of the recursion operator of integrable equations.^{8–11} They also appear as self-consistent sources of integrable equations.^{12,13} A more important application of squared eigenfunctions is in the direct soliton perturbation theory, where squared eigenfunctions and their closure relation play a fundamental role.^{5,6,14,15}

Squared eigenfunctions have been derived for a number of integrable equations including the KdV hierarchy, the Ablowitz-Kaup-Newell-Segur (AKNS) hierarchy, the derivative NLS hierarchy, the sine-Gordon equation, the massive Thirring model, the Benjamin-Ono equation, the

^{a)} Author to whom correspondence should be addressed. Electronic mail: jyang@cems.uvm.edu.

matrix NLS equations, and the Kadomtsev–Petviashvili equation.^{4,6,9–11,16–18} Several techniques have been used in those derivations. One is to notice the commutability relation between the linearized operator and the recursion operator of the integrable equation.^{10,11,19} Thus squared eigenfunctions are simply eigenfunctions of the recursion operator. The drawback of this method is that one has to first derive the recursion operator of the integrable equation and determine its eigenfunctions, which can be highly nontrivial for many equations (see Ref. 20, for instance). Another technique is to first calculate variations of the scattering data via the variations of the potentials, then use the Wronskian relations to invert this relation to get the variations of the potentials via variations of the scattering data.⁴ This Wronskian-relation technique involves some ingenious steps and could run into difficulties too if the problem is sufficiently complicated.⁴ The third technique is related to the second one, except that one directly calculates the variations of the potentials via variations of the scattering data by the Riemann–Hilbert method.^{4,6} This third method is general and conceptually simpler. Regarding the second and third methods, however, one question which was never clarified is why the functions appearing in the variation relations are indeed squared eigenfunctions and adjoint squared eigenfunctions which satisfy, respectively, the linearized equation and adjoint linearized equation of an integrable system. In all known examples, this was always found to be true by direct verifications. But whether and why it would remain true for the general case was not known.

Recently, we were interested in the Sasa–Satsuma equation which is relevant for the propagation of ultrashort optical pulses.^{21–23} This equation is integrable.²³ Two interesting features about this equation are that its solitons are embedded inside the continuous spectrum of the equation,^{15,24} and their shapes can be double humped for a wide range of soliton parameters.²³ One wonders how these double-humped solitons evolve when the equation is perturbed—a question which is interesting and significant from both physical and mathematical points of view. This question can be studied by a soliton perturbation theory. Due to the embedded nature of these solitons, external perturbations will generally excite continuous-wave radiation which is in resonance with the soliton. So a key component in the soliton perturbation theory would be to calculate this continuous-wave radiation, for which squared eigenfunctions are needed (a similar situation occurs in the soliton perturbation theory for the Hirota equation¹⁵). A special feature of the Sasa–Satsuma equation is that, while the linearized operator of this equation is 2×2 , its spectral operator is 3×3 . How to build two-component squared eigenfunctions from three-component Jost functions is an interesting and curious question. As we shall see, only certain components of the 3×3 Jost functions and their adjoints are used to construct the squared eigenfunctions and the adjoint squared eigenfunctions of the Sasa–Satsuma equation. To calculate these squared eigenfunctions for the Sasa–Satsuma equation, we have tried to directly use the first method mentioned above. The recursion operator for the Sasa–Satsuma equation has been derived recently.²⁰ However, that operator is quite complicated; thus its eigenfunctions are difficult to obtain. The second method mentioned above is met with difficulties too. Thus we choose the third method for the Sasa–Satsuma equation and demonstrate here certain advantages of this method.

In this paper, we further develop the third method using the Sasa–Satsuma equation as an example. We show that the functions appearing in the expansion of the variations of the potentials are always the squared eigenfunctions which satisfy the linearized equation of an integrable system, and that the functions appearing in the formulas for the variations of scattering data are always the adjoint squared eigenfunctions which satisfy the adjoint linearized equation of an integrable system. In addition, given these two relations between the variations of the potentials and the variations of scattering data, there naturally follows the closure relation for the squared eigenfunctions and their adjoints, as well as all the inner-product relations between the squared eigenfunctions and their adjoints. Thus no longer is it necessary to grind away at calculating these inner products from the asymptotics of the Jost functions. Rather one can just read off the values of the nonzero inner products from these two variation relations. This clarifies the long-standing question regarding squared eigenfunctions in connection with the linearized integrable equation and streamlines the third method as a general and conceptually simple procedure for the derivation of squared eigenfunctions. We apply this method to the Sasa–Satsuma equation and find that it

readily gives the squared eigenfunctions and adjoint squared eigenfunctions. The squared eigenfunctions for the Sasa–Satsuma equation are sums of two terms, with each term being a product of a component of a Jost function and a component of an adjoint Jost function. This two-term structure of squared eigenfunctions is caused by the symmetry properties of scattering data of the Sasa–Satsuma equation. It should be noted that our derivation uses almost exclusively the spectral operator of the Lax pair; thus all integrable equations with the same spectral operator (such as the Sasa–Satsuma hierarchy) will share the same set of squared eigenfunctions as we derived here (except for a time-dependent factor which is equation specific). An additional benefit of the method we used is that for two integrable equations with similar spectral operators, the derivation of their squared eigenfunctions will be essentially the same. Thus one can get squared eigenfunctions for one equation by minor modifications for the other equation. As an example, we demonstrate in the Appendix how squared eigenfunctions for the Manakov equations can be easily obtained by minor modifications of our calculations for the Sasa–Satsuma equation.

II. THE RIEMANN–HILBERT PROBLEM

To start our analysis, we first formulate the Riemann–Hilbert problem for the Sasa–Satsuma equation which will be needed for later calculations. The Sasa–Satsuma equation is

$$u_t + u_{xxx} + 6|u|^2 u_x + 3u(|u|^2)_x = 0. \quad (1)$$

Its spectral (scattering) problem of the Lax pair is²³

$$Y_x = -i\zeta \Lambda Y + QY, \quad (2)$$

where Y is a matrix function, $\Lambda = \text{diag}(1, 1, -1)$,

$$Q = \begin{pmatrix} 0 & 0 & u \\ 0 & 0 & u^* \\ -u^* & -u & 0 \end{pmatrix} \quad (3)$$

is the potential matrix, the “*” represents complex conjugation, and ζ is a spectral parameter. In this paper, we always assume that the potential $u(x)$ decays to zero sufficiently fast as $x \rightarrow \pm\infty$. Notice that this matrix Q has two symmetry properties. One is that it is anti-Hermitian, i.e., $Q^\dagger = -Q$, where the superscript “†” represents the Hermitian of a matrix. The other one is that

$$\sigma Q \sigma = Q^*, \quad \text{where } \sigma = \sigma^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (4)$$

Introducing new variables

$$J = YE^{-1}, \quad E = e^{-i\zeta \Lambda x}, \quad (5)$$

then Eq. (2) becomes

$$J_x = -i\zeta [\Lambda, J] + QJ, \quad (6)$$

where $[\Lambda, J] = \Lambda J - J\Lambda$. The matrix Jost solutions $J_\pm(x, \zeta)$ of Eq. (6) are defined by the asymptotics

$$J_\pm \rightarrow 1 \quad \text{as } x \rightarrow \pm\infty, \quad (7)$$

where “1” is the unit matrix. Here the subscripts in $J_\pm$ refer to which end of the x axis the boundary conditions are set. Since $\text{tr } Q = 0$, using Abel’s formula on Eq. (6), we see that $\det J_\pm = 1$ for all x . In addition, since $J_\pm E$ are both solutions of the linear equations (7), they are not independent and are linearly related by the scattering matrix $S(\zeta) = [s_{ij}(\zeta)]$:

$$J_- = J_+ E S E^{-1}, \quad \det S = 1 \quad (8)$$

for real ζ , i.e., $\zeta \in \mathbb{R}$ (for nonreal ζ , certain elements in S may not be well defined).

Due to the two symmetry properties of the potential matrix Q , the Jost solutions $J_{\pm}$ and the scattering matrix S satisfy the corresponding symmetry properties. One is the involutive property. Since the potential Q is anti-Hermitian, we see that $J_{\pm}^{\dagger}(\zeta^*)$ and $J_{\pm}^{-1}(\zeta)$ both satisfy the adjoint equation of (6). In addition, we see from Eq. (7) that $J_{\pm}^{\dagger}(\zeta^*)$ and $J_{\pm}^{-1}(\zeta)$ have the same large- x asymptotics; thus they are equal to each other:

$$J_{\pm}^{\dagger}(\zeta^*) = J_{\pm}^{-1}(\zeta). \quad (9)$$

Then in view of Eq. (8), we see that

$$S^{\dagger}(\zeta^*) = S^{-1}(\zeta). \quad (10)$$

To derive the other symmetry properties of the Jost solutions and the scattering matrix, we notice that due to the symmetry (4) of the potential Q , it is easy to see that $\sigma J_{\pm}^*(-\zeta^*)\sigma$ also satisfies Eq. (6). Then in view of the asymptotics (7), we find that Jost solutions possess the following additional symmetry:

$$J_{\pm}(\zeta) = \sigma J_{\pm}^*(-\zeta^*)\sigma. \quad (11)$$

From this symmetry and relation (8), we see that the scattering matrix S possesses the additional symmetry

$$S(\zeta) = \sigma S^*(-\zeta^*)\sigma. \quad (12)$$

Analytical properties of the Jost solutions play a fundamental role in the Riemann–Hilbert formulation for the scattering problem (6). For convenience, we express $J_{\pm}$ as

$$J_- = \Phi E^{-1}, \quad \Phi = [\phi_1, \phi_2, \phi_3], \quad (13)$$

$$J_+ = \Psi E^{-1}, \quad \Psi = [\psi_1, \psi_2, \psi_3], \quad (14)$$

where Φ and Ψ are fundamental solutions of the original spectral problem (2) and are related by the scattering matrix as

$$\Phi = \Psi S. \quad (15)$$

Then in view of the boundary conditions (7), the spectral equation (6) for $J_{\pm}$ can be rewritten into the following Volterra-type integral equations:

$$J_-(\zeta; x) = 1 + \int_{-\infty}^x e^{i\zeta\Lambda(y-x)} Q(y) J_-(\zeta; y) e^{i\zeta\Lambda(x-y)} dy, \quad (16)$$

$$J_+(\zeta; x) = 1 - \int_x^{\infty} e^{i\zeta\Lambda(y-x)} Q(y) J_-(\zeta; y) e^{i\zeta\Lambda(x-y)} dy. \quad (17)$$

These integral equations always have solutions when the integrals on their right hand sides converge. Due to the structure (3) of the potential Q , we easily see that Eq. (16) for the first and second columns of J_- contain only the exponential factor $e^{i\zeta(x-y)}$ which decays when ζ is in the upper half-plane $\mathbb{C}_+$, and Eq. (17) for the third column of J_+ contains only the exponential factor $e^{i\zeta(y-x)}$ which also falls off for $\zeta \in \mathbb{C}_+$. Thus these three columns can be analytically extended to $\zeta \in \mathbb{C}_+$. In other words, Jost solutions

$$P_+ = [\phi_1, \phi_2, \phi_3] e^{i\zeta \Lambda x} = J_- H_1 + J_+ H_2 \quad (18)$$

are analytic in $\zeta \in \mathbb{C}_+$, where

$$H_1 = \text{diag}(1, 1, 0), \quad H_2 = \text{diag}(0, 0, 1). \quad (19)$$

Here the subscript in P refers to which half-plane the functions are analytic in. From the Volterra integral equations for P_+ , we see that

$$P_+(x, \zeta) \rightarrow 1 \quad \text{as } \zeta \in \mathbb{C}_+ \rightarrow \infty. \quad (20)$$

Similarly, Jost functions $[\psi_1, \psi_2, \phi_3] e^{i\zeta \Lambda x}$ are analytic in $\zeta \in \mathbb{C}_-$, and their large- ζ asymptotics is

$$[\psi_1, \psi_2, \phi_3] e^{i\zeta \Lambda x} \rightarrow 1 \quad \text{as } \zeta \in \mathbb{C}_- \rightarrow \infty. \quad (21)$$

To obtain the analytic counterpart of P_+ in $\mathbb{C}_-$, we consider the adjoint spectral equation of (6):

$$K_x = -i\zeta[\Lambda, K] - KQ. \quad (22)$$

The inverse matrices $J_{\pm}^{-1}$ satisfy this adjoint equation. Notice that

$$J_-^{-1} = E\Phi^{-1}, \quad J_+^{-1} = E\Psi^{-1}. \quad (23)$$

Let us express Φ^{-1} and Ψ^{-1} as a collection of rows,

$$\Phi^{-1} \equiv \bar{\Phi} = \begin{bmatrix} \bar{\phi}_1 \\ \bar{\phi}_2 \\ \bar{\phi}_3 \end{bmatrix}, \quad \Psi^{-1} \equiv \bar{\Psi} = \begin{bmatrix} \bar{\psi}_1 \\ \bar{\psi}_2 \\ \bar{\psi}_3 \end{bmatrix}, \quad (24)$$

where the overbar refers to the adjoint quantity; then by similar techniques as used above, we can show that the first and second rows of J_-^{-1} and the third row of J_+^{-1} are analytic in $\zeta \in \mathbb{C}_-$, i.e., adjoint Jost solutions

$$P_- = e^{-i\zeta \Lambda x} \begin{bmatrix} \bar{\phi}_1 \\ \bar{\phi}_2 \\ \bar{\psi}_3 \end{bmatrix} = H_1 J_-^{-1} + H_2 J_+^{-1} \quad (25)$$

are analytic in $\zeta \in \mathbb{C}_-$. In addition, their large- ζ asymptotics is

$$P_-(x, \zeta) \rightarrow 1 \quad \text{as } \zeta \in \mathbb{C}_- \rightarrow \infty. \quad (26)$$

Similarly, the first and second rows of J_+^{-1} and the third row of J_-^{-1} , i.e., $e^{-i\zeta x} \bar{\psi}_1$, $e^{-i\zeta x} \bar{\psi}_2$, and $e^{i\zeta x} \bar{\phi}_3$, are analytic in $\zeta \in \mathbb{C}_+$, and their large- ζ asymptotics is

$$e^{-i\zeta \Lambda x} \begin{bmatrix} \bar{\psi}_1 \\ \bar{\psi}_2 \\ \bar{\phi}_3 \end{bmatrix} \rightarrow 1 \quad \text{as } \zeta \in \mathbb{C}_+ \rightarrow \infty. \quad (27)$$

In view of the involution properties (9) of $J_{\pm}$, we see that the analytic solutions $P_{\pm}$ satisfy the involutive property as well:

$$P_+^{\dagger}(\zeta^*) = P_-(\zeta). \quad (28)$$

This property can be taken as a definition of the analytic function P_- from the known analytic function P_+ .

The analytic properties of Jost functions described above have immediate implications on the analytic properties of the scattering matrix S . Let us denote

$$S^{-1}(\zeta) \equiv \bar{S}(\zeta) = [\bar{s}_{ij}(\zeta)]. \quad (29)$$

Then since

$$S = \Psi^{-1}\Phi = \begin{bmatrix} \bar{\psi}_1 \\ \bar{\psi}_2 \\ \bar{\psi}_3 \end{bmatrix} [\phi_1, \phi_2, \phi_3], \quad \bar{S} = \Phi^{-1}\Psi = \begin{bmatrix} \bar{\phi}_1 \\ \bar{\phi}_2 \\ \bar{\phi}_3 \end{bmatrix} [\psi_1, \psi_2, \psi_3], \quad (30)$$

we see immediately that s_{11} , s_{12} , s_{21} , s_{22} , and $\bar{s}_{33}$ can be analytically extended to the upper half-plane $\zeta \in C_+$, while $\bar{s}_{11}$, $\bar{s}_{12}$, $\bar{s}_{21}$, $\bar{s}_{22}$, and s_{33} can be analytically extended to the lower half-plane $\zeta \in C_-$. In addition, their large- ζ asymptotics are

$$\begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} \rightarrow 1, \quad \bar{s}_{33} \rightarrow 1 \quad \text{as } \zeta \in C_+ \rightarrow \infty \quad (31)$$

and

$$\begin{pmatrix} \bar{s}_{11} & \bar{s}_{12} \\ \bar{s}_{21} & \bar{s}_{22} \end{pmatrix} \rightarrow 1, \quad s_{33} \rightarrow 1 \quad \text{as } \zeta \in C_- \rightarrow \infty. \quad (32)$$

Hence we have constructed two matrix functions P_+ and P_- which are analytic in C_+ and C_- , respectively. On the real line, using Eqs. (8), (18), (25), and (29), we easily see that

$$P_-(\zeta)P_+(\zeta) = G(\zeta), \quad \zeta \in R, \quad (33)$$

where

$$G = E(H_1 + H_2S)(H_1 + S^{-1}H_2)E^{-1} = E \begin{pmatrix} 1 & 0 & \bar{s}_{13} \\ 0 & 1 & \bar{s}_{23} \\ s_{31} & s_{32} & 1 \end{pmatrix} E^{-1}. \quad (34)$$

Equation (33) determines a matrix Riemann–Hilbert problem. The normalization condition for this Riemann–Hilbert problem can be seen from (20) and (26) as

$$P_{\pm}(x, \zeta) \rightarrow 1 \quad \text{as } \zeta \in C_{\pm} \rightarrow \infty. \quad (35)$$

If this problem can be solved, then the potential Q can be reconstructed from an asymptotic expansion of its solution for large ζ . Indeed, writing P_+ as

$$P_+(x, \zeta) = 1 + \zeta^{-1}P^{(1)}(x) + \zeta^{-2}P^{(2)}(x) + O(\zeta^{-3}), \quad (36)$$

inserting it into Eq. (6), and comparing terms of the same order in ζ^{-1} , we find at $O(1)$ that

$$Q = i[\Lambda, P^{(1)}]. \quad (37)$$

Thus the potential Q can be reconstructed from $P^{(1)}$. At $O(\zeta^{-1})$, we find that

$$P_x^{(1)} = -i[\Lambda, P^{(2)}] + QP^{(1)}. \quad (38)$$

From the above two equations as well as the large- x asymptotics of $P_+(x, \zeta)$ from Eq. (7), we see that the full matrix $P^{(1)}$ is

$$P^{(1)}(x) = \frac{1}{2i} \begin{bmatrix} \int_{-\infty}^x |u(y)|^2 dy & \int_{-\infty}^x u^2(y) dy & u(x) \\ \int_{-\infty}^x u^{*2}(y) dy & \int_{-\infty}^x |u(y)|^2 dy & u^*(x) \\ u^*(x) & u(x) & 2 \int_x^{\infty} |u(y)|^2 dy \end{bmatrix}. \quad (39)$$

This matrix, together with Eq. (36), gives the leading-order asymptotic expansions for the analytical functions P_+ . Leading-order asymptotic expansions for P_- can be similarly derived.

Elements of the scattering matrix S are not all independent. Indeed, since the potential matrix Q contains only one independent function $u(x)$, matrix S should contain only one independent element as well. To get the dependence of scattering coefficients, we first notice from equation $S\bar{S}=1$ that

$$s_{33}(\zeta)\bar{s}_{33}(\zeta) = 1 - s_{31}(\zeta)\bar{s}_{13}(\zeta) - s_{32}(\zeta)\bar{s}_{23}(\zeta), \quad \zeta \in \mathbb{R}. \quad (40)$$

Since s_{33} and $\bar{s}_{33}$ are analytic in C_- and C_+ , respectively, the above equation defines a (scalar) Riemann–Hilbert problem for $(s_{33}, \bar{s}_{33})$ under the canonical normalization conditions which can be seen from (31) and (32). Thus $(s_{33}, \bar{s}_{33})$ can be uniquely determined by the locations of their zeros in $C_{\pm}$ together with the scattering data $(s_{31}, s_{32}, \bar{s}_{13}, \bar{s}_{23})$ on the real line $\zeta \in \mathbb{R}$ by the Plemelj formula. Similarly, from equation $S\bar{S}=1$ we have

$$\begin{pmatrix} \bar{s}_{11} & \bar{s}_{12} \\ \bar{s}_{21} & \bar{s}_{22} \end{pmatrix} \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} = 1 - \begin{pmatrix} \bar{s}_{13} \\ \bar{s}_{23} \end{pmatrix} (s_{31}, s_{32}), \quad \zeta \in \mathbb{R}, \quad (41)$$

which defines another (matrix) Riemann–Hilbert problem for $(\bar{s}_{11}, \bar{s}_{12}, \bar{s}_{21}, \bar{s}_{22})$ and $(s_{11}, s_{12}, s_{21}, s_{22})$ under the canonical normalization conditions which can be seen from (31) and (32). Thus these analytical scattering elements can be determined by the scattering data $(s_{31}, s_{32}, \bar{s}_{13}, \bar{s}_{23})$ on the real line $\zeta \in \mathbb{R}$ as well. The scattering data $(s_{31}, s_{32}, \bar{s}_{13}, \bar{s}_{23})$ themselves are dependent on each other by the symmetry relations (10) and (12). Specifically, for $\xi \in \mathbb{R}$, $\bar{s}_{13}(\xi) = s_{31}^*(\xi)$, $\bar{s}_{23}(\xi) = s_{32}^*(\xi)$, and $s_{32}(\xi) = s_{31}^*(-\xi)$. Thus the scattering matrix S indeed contains only a single independent element as we would expect.

In the next section, we will derive squared eigenfunctions for the Sasa–Satsuma equation. For the convenience of that derivation, we introduce the following notations:

$$\rho_1 \equiv \frac{s_{31}}{s_{33}}, \quad \rho_2 \equiv \frac{s_{32}}{s_{33}}, \quad \bar{\rho}_1 \equiv \frac{\bar{s}_{13}}{\bar{s}_{33}}, \quad \bar{\rho}_2 \equiv \frac{\bar{s}_{23}}{\bar{s}_{33}}. \quad (42)$$

Due to the symmetry conditions (10) and (12) of the scattering matrix S , we see that ρ_k and $\bar{\rho}_k$ satisfy the symmetry conditions

$$\bar{\rho}_k(\zeta) = \rho_k^*(\zeta), \quad k=1, 2, \quad \bar{\rho}_1(\zeta) = \rho_2(-\zeta), \quad \bar{\rho}_2(\zeta) = \rho_1(-\zeta), \quad \zeta \in \mathbb{R}. \quad (43)$$

In the next section, symmetry properties of Φ and Ψ will also be needed. Due to the symmetry conditions (9) and (11) of the Jost solutions, Φ and Ψ satisfy the symmetry conditions

$$\Phi^+(\zeta^*) = \Phi^{-1}(\zeta), \quad \Psi^+(\zeta^*) = \Psi^{-1}(\zeta) \quad (44)$$

and

$$\Phi_{\pm}(\zeta) = \sigma \Phi_{\pm}^*(-\zeta^*)\sigma, \quad \Psi_{\pm}(\zeta) = \sigma \Psi_{\pm}^*(-\zeta^*)\sigma. \quad (45)$$

Symmetry (44) means that

$$\phi_k^\dagger(\zeta^*) = \bar{\phi}_k(\zeta), \quad \psi_k^\dagger(\zeta^*) = \bar{\psi}_k(\zeta), \quad k = 1, 2, 3, \quad (46)$$

while symmetry (45), together with (46), means that

$$\sigma\phi_1(\zeta) = \bar{\phi}_2^T(-\zeta), \quad \sigma\phi_2(\zeta) = \bar{\phi}_1^T(-\zeta), \quad \sigma\phi_3(\zeta) = \bar{\phi}_3^T(-\zeta). \quad (47)$$

Here the superscript T represents the transpose of a matrix. These symmetry properties will be important for deriving the final expressions of squared eigenfunctions for the Sasa–Satsuma equation in the next section.

III. SQUARED EIGENFUNCTIONS AND THEIR CLOSURE RELATION

In this section, we calculate the variation of the potential via variations of the scattering data, then calculate variations of the scattering data via the variation of the potential. The first step will yield squared eigenfunctions, and it will be done by the Riemann–Hilbert method. The second step will yield adjoint squared eigenfunctions, and it will be done using basic relations of the spectral problem (2). For the ease of presentation, we first assume that s_{33} and $\bar{s}_{33}$ have no zeros in their respective planes of analyticity, i.e., the spectral problem (2) has no discrete eigenvalues. This facilitates the derivation of squared eigenfunctions. The results for the general case of s_{33} and $\bar{s}_{33}$ having zeros will be given at the end of this section.

A. Variation of the potential and squared eigenfunctions

In this subsection, we derive the variation of the potential via variations of the scattering data, which will readily yield the squared eigenfunctions satisfying the linearized Sasa–Satsuma equation. Our derivation will be based on the Riemann–Hilbert method.

To proceed, we define the following matrix functions

$$F_+ = P_+ \operatorname{diag}\left(1, 1, \frac{1}{s_{33}}\right), \quad F_- = P_-^{-1} \operatorname{diag}(1, 1, s_{33}). \quad (48)$$

The reason to introduce diagonal matrices with s_{33} and $\bar{s}_{33}$ in $F_\pm$ is to obtain a new Riemann–Hilbert problem (49) with a connection matrix $\tilde{G}$ which depends on ρ_k and $\bar{\rho}_k$ rather than s_{ij} and $\bar{s}_{ij}$. This way, the variation of the potential will be expressed in terms of variations in ρ_k and $\bar{\rho}_k$. When $\bar{s}_{33}$ and s_{33} have no zeros in C_+ and C_- , respectively, then $F_\pm$ as well as $F_\pm^{-1}$ are analytic in $C_\pm$. On the real line, they are related by

$$F_+(\zeta) = F_-(\zeta)\tilde{G}(\zeta), \quad \zeta \in \mathbb{R}, \quad (49)$$

where

$$\tilde{G} = \operatorname{diag}\left(1, 1, \frac{1}{s_{33}}\right) G \operatorname{diag}\left(1, 1, \frac{1}{s_{33}}\right) = E \begin{pmatrix} 1 & 0 & \bar{\rho}_1 \\ 0 & 1 & \bar{\rho}_2 \\ \rho_1 & \rho_2 & 1 + \rho_1\bar{\rho}_1 + \rho_2\bar{\rho}_2 \end{pmatrix} E^{-1}. \quad (50)$$

Here relation (40) has been used. Equation (49) defines a regular Riemann–Hilbert problem (i.e., without zeros).

Next we take the variation of the Riemann–Hilbert problem (49) and get

$$\delta F_+ = \delta F_- \tilde{G} + F_- \delta \tilde{G}, \quad \zeta \in \mathbb{R}. \quad (51)$$

Utilizing Eq. (49), we can rewrite the above equation as

$$\delta F_+ F_+^{-1} = \delta F_- F_-^{-1} + F_- \delta \tilde{G} F_+^{-1}, \quad \zeta \in \mathbb{R}, \quad (52)$$

which defines yet another regular Riemann–Hilbert problem for $\delta F F^{-1}$. Unlike the previous Riemann–Hilbert problems (33) and (49) which were in matrix product forms, the present Riemann–Hilbert problem (52) can be explicitly solved. Using the Plemelj formula, the general solution of this Riemann–Hilbert problem is

$$\delta F F^{-1}(\zeta; x) = A_0(x) + \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\Pi(\xi; x)}{\xi - \zeta} d\xi, \quad (53)$$

where

$$\Pi(\xi; x) \equiv F_-(\xi; x) \delta \tilde{G}(\xi; x) F_+^{-1}(\xi; x), \quad \xi \in \mathbb{R}. \quad (54)$$

Now we consider the large- ζ asymptotics of this solution. Notice from expansions (36) and (31) and relation (48) that as $\zeta \rightarrow \infty$,

$$F_+(\zeta; x) \rightarrow \begin{bmatrix} 1 + O\left(\frac{1}{\zeta}\right) & O\left(\frac{1}{\zeta}\right) & \frac{u(x)}{2i\zeta} \\ O\left(\frac{1}{\zeta}\right) & 1 + O\left(\frac{1}{\zeta}\right) & \frac{u^*(x)}{2i\zeta} \\ \frac{u^*(x)}{2i\zeta} & \frac{u(x)}{2i\zeta} & 1 + O\left(\frac{1}{\zeta}\right) \end{bmatrix}, \quad \delta F_+(\zeta; x) \rightarrow \begin{bmatrix} O\left(\frac{1}{\zeta}\right) & O\left(\frac{1}{\zeta}\right) & \frac{\delta u(x)}{2i\zeta} \\ O\left(\frac{1}{\zeta}\right) & O\left(\frac{1}{\zeta}\right) & \frac{\delta u^*(x)}{2i\zeta} \\ \frac{\delta u^*(x)}{2i\zeta} & \frac{\delta u(x)}{2i\zeta} & O\left(\frac{1}{\zeta}\right) \end{bmatrix}. \quad (55)$$

In addition, it is easy to see that

$$\int_{-\infty}^{\infty} \frac{\Pi(\xi; x)}{\xi - \zeta} d\xi \rightarrow -\frac{1}{\zeta} \int_{-\infty}^{\infty} \Pi(\xi; x) d\xi, \quad \zeta \rightarrow \infty. \quad (56)$$

When these large- ζ expansions are substituted into Eq. (53), at $O(1)$, we find that $A_0(x) = 0$. At $O(\zeta^{-1})$, we get

$$\delta u(x) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \Pi_{13}(\xi; x) d\xi, \quad \delta u^*(x) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \Pi_{31}(\xi; x) d\xi. \quad (57)$$

Now we calculate the elements Π_{13} and Π_{31} . From Eq. (50), we see that

$$\delta \tilde{G} = E \begin{pmatrix} 0 & 0 & \delta \bar{\rho}_1 \\ 0 & 0 & \delta \bar{\rho}_2 \\ \delta \rho_1 & \delta \rho_2 & \rho_1 \delta \bar{\rho}_1 + \bar{\rho}_1 \delta \rho_1 + \rho_2 \delta \bar{\rho}_2 + \bar{\rho}_2 \delta \rho_2 \end{pmatrix} E^{-1}. \quad (58)$$

Then in view of the $F_{\pm}$ definitions (48) and $P_{\pm}$ expressions (18) and (25), we find from (54) that

$$\Pi(\zeta; x) = J_- E (H_1 + H_2 S)^{-1} \text{diag}(1, 1, s_{33}) E^{-1} \delta \tilde{G} E \text{diag}(1, 1, \bar{s}_{33}) (H_1 + S^{-1} H_2)^{-1} E^{-1} J_-^{-1}, \quad (59)$$

which simplifies to

$$\Pi(\xi; x) = \Phi \begin{pmatrix} 0 & 0 & \delta\bar{\rho}_1 \\ 0 & 0 & \delta\bar{\rho}_2 \\ \delta\rho_1 & \delta\rho_2 & 0 \end{pmatrix} \bar{\Phi}. \quad (60)$$

This relation, together with Eq. (57), shows that the variation of the potential δu can be expanded into quadratic combinations between Jost solutions Φ and adjoint Jost solutions $\bar{\Phi}$. This relation is generic in integrable systems.

Now we calculate explicit expressions for δu . Inserting (60) into (57) and recalling our notations (24), we readily find that

$$\delta u = -\frac{1}{\pi} \int_{-\infty}^{\infty} (\phi_{31}\bar{\phi}_{13}\delta\rho_1 + \phi_{31}\bar{\phi}_{23}\delta\rho_2 + \phi_{11}\bar{\phi}_{33}\delta\bar{\rho}_1 + \phi_{21}\bar{\phi}_{33}\delta\bar{\rho}_2) d\xi. \quad (61)$$

Here the notations are

$$\phi_k = \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \end{bmatrix}, \quad \psi_k = \begin{bmatrix} \psi_{k1} \\ \psi_{k2} \\ \psi_{k3} \end{bmatrix}, \quad \bar{\phi}_k = [\bar{\phi}_{k1}, \bar{\phi}_{k2}, \bar{\phi}_{k3}], \quad \bar{\psi}_k = [\bar{\psi}_{k1}, \bar{\psi}_{k2}, \bar{\psi}_{k3}], \quad k = 1, 2, 3. \quad (62)$$

Utilizing the symmetry relations (43) and (47), the above δu formula reduces to

$$\delta u = -\frac{1}{\pi} \int_{-\infty}^{\infty} [(\phi_{31}\bar{\phi}_{13} + \phi_{33}\bar{\phi}_{12})\delta\rho_1 + (\phi_{11}\bar{\phi}_{33} + \phi_{13}\bar{\phi}_{32})\delta\bar{\rho}_1] d\xi. \quad (63)$$

This is an important step in our derivation, where symmetry conditions play an important role. Regarding δu^* , its formula can be obtained by taking the complex conjugate of the above equation and simplified by using the symmetry relations (46). Defining functions

$$Z_1 = \begin{bmatrix} \phi_{31}\bar{\phi}_{13} + \phi_{33}\bar{\phi}_{12} \\ \phi_{32}\bar{\phi}_{13} + \phi_{33}\bar{\phi}_{11} \end{bmatrix}, \quad Z_2 = \begin{bmatrix} \phi_{11}\bar{\phi}_{33} + \phi_{13}\bar{\phi}_{32} \\ \phi_{12}\bar{\phi}_{33} + \phi_{13}\bar{\phi}_{31} \end{bmatrix}, \quad (64)$$

then the final expression for the variation of the potential $(\delta u, \delta u^*)^T$ is

$$\begin{bmatrix} \delta u(x) \\ \delta u^*(x) \end{bmatrix} = -\frac{1}{\pi} \int_{-\infty}^{\infty} [Z_1(\xi; x)\delta\rho_1(\xi) + Z_2(\xi; x)\delta\bar{\rho}_1(\xi)] d\xi. \quad (65)$$

Notice here that due to the symmetry relations (46), $Z_2(\xi)$ is equal to $Z_1^*(\xi^*)$ with its two components swapped. Also notice that Z_1 and Z_2 are the sum of two terms, where each term is a product of a component of a Jost function and a component of an adjoint Jost function. The two-term feature of these functions is caused by the symmetry properties of the scattering data and Jost functions, which enable us to combine terms together in the general expansion (61). In the massive Thirring model, the counterparts of functions $Z_{1,2}$ are also sums of two terms.⁴ The two-term structure there is not due to symmetry properties but rather due to the asymptotics of its Jost functions in the spectral plane. The feature of each term in Z_1 and Z_2 being a product between a Jost function and an adjoint Jost function, on the other hand, is a generic feature in integrable systems. In previous studies on many integrable equations (such as the KdV, NLS, derivative NLS, and massive Thirring equations), it was found that these functions were often “squares” or products of Jost functions themselves (thus the name “squared eigenfunctions”).^{1,3,4,16} This was so simply because the spectral operators in those systems were 2×2 , for which the adjoint Jost functions (rows of the inverse of the Jost function matrix) are directly proportional to Jost functions themselves. That is not generic, however, and does not hold for the Sasa–Satsuma equation or in general for integrable equations whose spectral operator is 3×3 or higher.

The derivation of (65) for the expansion of $(\delta u, \delta u^*)^T$ is an important result of this subsection. It readily gives the variations in the Sasa–Satsuma fields in terms of variations in the initial data.

To show this, we first restore the time dependence in Eq. (65). An important fact we need to notice here is that the Jost solutions Φ as defined in (13) do not satisfy the time evolution equation of the Lax pair of the Sasa–Satsuma equation. Indeed, this time evolution equation of the Lax pair is²³

$$Y_t = -4i\xi^3 \Lambda Y + V(\xi, u)Y, \quad (66)$$

where the matrix V goes to zero as x approaches infinity. Obviously the large- x asymptotics of Φ [see (7) and (13)] cannot satisfy the above equation as $|x| \rightarrow \infty$ where V vanishes. But this problem can be easily fixed. Defining the “time-dependent” Jost functions

$$\Phi^{(t)} = \Phi e^{-4i\xi^3 \Lambda t}, \quad (67)$$

then these functions satisfy *both* parts of the Lax pair, (2) and (66). The reason they now satisfy the time evolution equation (66) is due to their satisfying the asymptotic time evolution equation (66) as $|x| \rightarrow \infty$ as well as the compatibility relation of the Lax pair. Similarly we define the time-dependent adjoint Jost functions as

$$\bar{\Phi}^{(t)} = e^{4i\xi^3 \Lambda t} \bar{\Phi}, \quad (68)$$

which satisfy both adjoint equations of the Lax pair. Now we use these new Jost functions and adjoint Jost functions to replace those in the definitions (64) and get time-dependent (Z_1, Z_2) functions

$$Z_1^{(t)} = \begin{bmatrix} \phi_{31}^{(t)} \bar{\phi}_{13}^{(t)} + \phi_{33}^{(t)} \bar{\phi}_{12}^{(t)} \\ \phi_{32}^{(t)} \bar{\phi}_{13}^{(t)} + \phi_{33}^{(t)} \bar{\phi}_{11}^{(t)} \end{bmatrix}, \quad Z_2^{(t)} = \begin{bmatrix} \phi_{11}^{(t)} \bar{\phi}_{33}^{(t)} + \phi_{13}^{(t)} \bar{\phi}_{32}^{(t)} \\ \phi_{12}^{(t)} \bar{\phi}_{33}^{(t)} + \phi_{13}^{(t)} \bar{\phi}_{31}^{(t)} \end{bmatrix}. \quad (69)$$

In view of relations (67) and (68), we see that these time-dependent (Z_1, Z_2) functions are related to the original ones as

$$Z_1^{(t)} = Z_1 e^{8i\xi^3 t}, \quad Z_2^{(t)} = Z_2 e^{-8i\xi^3 t}. \quad (70)$$

Another fact we need to notice in conjunction with the time-restored equation of (65) is that for the Sasa–Satsuma equation, the time evolution of the scattering matrix S is given by²³

$$S_t = -4i\xi^3 [\Lambda, S]. \quad (71)$$

Thus

$$s'_{33}(t) = 0, \quad s'_{31}(t) = 8i\xi^3 s_{31}(t), \quad (72)$$

where the prime indicates differentiation with respect to t . Then recalling definition (42) of $(\rho_1, \bar{\rho}_1)$ as well as the symmetry property (10), we obtain the time evolution of the varied scattering data as

$$\delta\rho_1(\xi, t) = \delta\rho_1(\xi, 0) e^{8i\xi^3 t}, \quad \delta\bar{\rho}_1(\xi, t) = \delta\bar{\rho}_1(\xi, 0) e^{-8i\xi^3 t}. \quad (73)$$

Now we insert the above time-dependence relations (70) and (73) into the time-restored expansion (65) and get a new expansion,

$$\begin{bmatrix} \delta u(x, t) \\ \delta u^*(x, t) \end{bmatrix} = -\frac{1}{\pi} \int_{-\infty}^{\infty} [Z_1^{(t)}(\xi; x, t) \delta\rho_1(\xi, 0) + Z_2^{(t)}(\xi; x, t) \delta\bar{\rho}_1(\xi, 0)] d\xi. \quad (74)$$

The advantages of this expansion are that the Jost functions $\Phi^{(t)}$ in $[Z_1^{(t)}, Z_2^{(t)}]$ satisfy both equations of the Lax pair, and the expansion coefficients $[\delta\rho_1, \delta\bar{\rho}_1]$ are time independent.

Since δu and δu^* must satisfy the homogeneous linearized Sasa–Satsuma equation, it follows that the functions $[Z_1^{(t)}, Z_2^{(t)}]$ appearing in expansion (74) must also and precisely satisfy the same linearized Sasa–Satsuma equation. A similar fact for several other integrable equations has been

noted before by direct verification (see Ref. 6, for instance). Here we will give a simple proof of this fact which holds for any integrable equation where an expansion such as (74) exists. Suppose the linearized operator of the Sasa–Satsuma equation for $[u(x, t), u^*(x, t)]^T$ is $\mathcal{L}$. Since the potential $u(x, t)$ satisfies the Sasa–Satsuma equation and the variation of the potential $[\delta u(x, t), \delta u^*(x, t)]^T$ is infinitesimal, it follows that

$$\mathcal{L} \begin{bmatrix} \delta u(x, t) \\ \delta u^*(x, t) \end{bmatrix} = 0. \quad (75)$$

Inserting Eq. (74) into the above equation, we get

$$\int_{-\infty}^{\infty} [\mathcal{L}Z_1^{(t)}(\xi; x, t)\delta\rho_1(\xi, 0) + \mathcal{L}Z_2^{(t)}(\xi; x, t)\delta\bar{\rho}_1(\xi, 0)]d\xi = 0. \quad (76)$$

Since the initial variations of the scattering data $(\delta\rho_1, \delta\bar{\rho}_1)(\xi, 0)$ are arbitrary and linearly independent, it follows that

$$\mathcal{L}Z_1^{(t)}(\xi; x, t) = \mathcal{L}Z_2^{(t)}(\xi; x, t) = 0 \quad (77)$$

for any $\xi \in \mathbb{R}$. Thus $Z_1^{(t)}$ and $Z_2^{(t)}$ satisfy the linearized Sasa–Satsuma equation and are therefore the squared eigenfunctions of this equation.

Now in the derivation of (74), from (52) forward, no constraints have been put on the variations of u and u^* , except for the implied assumption that the scattering data for u and u^* and its variations will exist. Whence we expect that $[\delta u, \delta u^*]^T$ in (74) can be considered to be arbitrary. If so, then due to the equality which we see in (74), we already can expect $\{Z_1^{(t)}(\xi; x, t), Z_2^{(t)}(\xi; x, t), \xi \in \mathbb{R}\}$ to form a complete set in an appropriate functional space (such as L_1).

We remark that the squared eigenfunctions given here are all linearly independent. That can be easily verified upon recognizing that from (2), one may readily construct an integrodifferential eigenvalue problem for which these squared eigenfunctions are the true eigenfunctions. Then in the usual manner it follows that these squared eigenfunctions are linearly independent. Similar for the adjoint squared eigenfunctions. The linear independence of these squared eigenfunctions can also be easily established after we have obtained the dual relations of (74) which give variations of scattering data due to arbitrary variations of the potentials (see next section).

The squared eigenfunctions $[Z_1^{(t)}, Z_2^{(t)}]$ have nice analytic properties. Indeed, recalling the analytic properties of Jost functions discussed in Sec. II, we see that $Z_1^{(t)}(\xi)e^{-2i\xi x}$ is analytic in $\mathbb{C}_-$, while $Z_2^{(t)}(\xi)e^{2i\xi x}$ is analytic in $\mathbb{C}_+$. These analytic properties will be essential when we extend our results to the general case where s_{33} and $\bar{s}_{33}$ have zeros (see end of this section).

It should be noted that in expressions (69) for $[Z_1^{(t)}, Z_2^{(t)}]$, if $\phi_3^{(t)}$ is replaced by other Jost functions (such as $\phi_1^{(t)}$) or if $\bar{\phi}_1^{(t)}$ is replaced by other adjoint Jost functions (such as $\bar{\phi}_2^{(t)}$), the resulting functions would still satisfy the linearized Sasa–Satsuma equation and are thus also squared eigenfunctions of this equation. These facts can be verified directly by inserting such functions into the linearized Sasa–Satsuma equation and noticing that $\Phi^{(t)}$ and $\bar{\Phi}^{(t)}$ satisfy the Lax pair (2) and (66) and the adjoint Lax pair, respectively. Similar results also hold for other integrable equations. However, these other squared eigenfunctions in general do not have nice analytic properties in $\mathbb{C}_\pm$ nor are they linearly independent of set (69).

B. Variations of the scattering data and adjoint squared eigenfunctions

In this subsection, we calculate variations of the scattering data which occur due to arbitrary variations in the potentials. These formulas, together with Eq. (65) or Eq. (74), will give the “adjoint” form of the squared eigenfunctions as well as their closure relation.

We start with the spectral equation (2) or Jost functions Φ or, specifically,

$$\Phi_x = -i\zeta\Lambda\Phi + Q\Phi. \quad (78)$$

Taking the variation to this equation, we get

$$\delta\Phi_x = -i\zeta\Lambda\delta\Phi + Q\delta\Phi + \delta Q\Phi. \quad (79)$$

Recalling that $\Phi \rightarrow e^{-i\zeta\Lambda x}$ as $x \rightarrow -\infty$ [see Eqs. (7) and (13)], thus $\delta\Phi \rightarrow 0$ as $x \rightarrow -\infty$. As a result, the solution of the inhomogeneous equation (79) can be found by the method of variation of parameters as

$$\delta\Phi(\zeta; x) = \Phi(\zeta; x) \int_{-\infty}^x \Phi^{-1}(\zeta; y) \delta Q(y) \Phi(\zeta; y) dy. \quad (80)$$

Here

$$\delta Q = \begin{pmatrix} 0 & 0 & \delta u \\ 0 & 0 & \delta u^* \\ -\delta u^* & -\delta u & 0 \end{pmatrix} \quad (81)$$

is the variation of the potential. Now we take the limit of $x \rightarrow \infty$ in the above equation. Recalling $\Phi = \Psi S$ and the asymptotics of $\Psi \rightarrow E$ as $x \rightarrow \infty$, we see that $\Phi \rightarrow ES$, $\delta\Phi \rightarrow E\delta S$ as $x \rightarrow \infty$. Thus in this limit, the above equation becomes

$$\delta S(\xi) = S(\xi) \int_{-\infty}^{\infty} \Phi^{-1}(\xi; x) \delta Q(x) \Phi(\xi; x) dx, \quad \xi \in \mathbb{R}. \quad (82)$$

Noticing that $S\Phi^{-1} = \Psi^{-1} \equiv \bar{\Psi}$, the above equation can be rewritten as

$$\delta S(\xi) = \int_{-\infty}^{\infty} \bar{\Psi}(\xi; x) \delta Q(x) \Phi(\xi; x) dx, \quad \xi \in \mathbb{R}. \quad (83)$$

This formula gives variations of scattering coefficients δs_{ij} via the variation of the potential δQ . From it, the variation of the scattering data $\delta\rho_1$ can be readily found. This $\delta\rho_1$ formula contains both Jost functions Φ and adjoint Jost functions $\bar{\Psi}$. When we further express Φ in terms of Ψ through relation (15) (so that the boundary conditions of the Jost functions involved are all set at $x = +\infty$), we obtain the expression for $\delta\rho_1$ as

$$\delta\rho_1 = \frac{1}{s_{33}^2} \int_{-\infty}^{\infty} [\delta u(\bar{\psi}_{31}\varphi_3 - \bar{\psi}_{33}\varphi_2) + \delta u^*(\bar{\psi}_{32}\varphi_3 - \bar{\psi}_{33}\varphi_1)] dx, \quad (84)$$

where the column vector φ is defined as

$$\varphi = [\varphi_1, \varphi_2, \varphi_3]^T \equiv \bar{s}_{22}\psi_1 - \bar{s}_{21}\psi_2. \quad (85)$$

The variation $\delta\bar{\rho}_1$ can be derived by taking the complex conjugate of the above equation and utilizing the symmetry relations (43) and (46). Defining functions

$$\bar{\varphi} = [\bar{\varphi}_1, \bar{\varphi}_2, \bar{\varphi}_3] \equiv s_{22}\bar{\psi}_1 - s_{12}\bar{\psi}_2, \quad (86)$$

$$\Omega_1 = \begin{bmatrix} \bar{\psi}_{31}\varphi_3 - \bar{\psi}_{33}\varphi_2 \\ \bar{\psi}_{32}\varphi_3 - \bar{\psi}_{33}\varphi_1 \end{bmatrix}, \quad \Omega_2 = \begin{bmatrix} \psi_{32}\bar{\varphi}_3 - \psi_{33}\bar{\varphi}_1 \\ \psi_{31}\bar{\varphi}_3 - \psi_{33}\bar{\varphi}_2 \end{bmatrix} \quad (87)$$

and inner products

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f^T(x)g(x)dx, \quad (88)$$

then the final expressions for variations of the scattering data $\delta\rho_1$ and $\delta\bar{\rho}_1$ are

$$\delta\rho_1(\xi) = \frac{1}{s_{33}^2(\xi)} \left\langle \Omega_1(\xi; x), \begin{bmatrix} \delta u(x) \\ \delta u^*(x) \end{bmatrix} \right\rangle, \quad (89)$$

$$\delta\bar{\rho}_1(\xi) = \frac{1}{\bar{s}_{33}^2(\xi)} \left\langle \Omega_2(\xi; x), \begin{bmatrix} \delta u(x) \\ \delta u^*(x) \end{bmatrix} \right\rangle, \quad (90)$$

where $\xi \in \mathbb{R}$.

The above two equations are associated with expansion (65) of the previous section, where time is frozen in both cases. A small problem with these equations is that the Jost functions and adjoint Jost functions appearing in definitions (87) of $[\Omega_1, \Omega_2]$ do not satisfy the time evolution equation (66) of the Lax pair (see the previous section). To fix this problem, we repeat the practice of the previous section and introduce time-dependent versions of the functions $[\Omega_1, \Omega_2]$ as

$$\Omega_1^{(t)} = \begin{bmatrix} \bar{\psi}_{31}^{(t)}\varphi_3^{(t)} - \bar{\psi}_{33}^{(t)}\varphi_2^{(t)} \\ \bar{\psi}_{32}^{(t)}\varphi_3^{(t)} - \bar{\psi}_{33}^{(t)}\varphi_1^{(t)} \end{bmatrix}, \quad \Omega_2^{(t)} = \begin{bmatrix} \psi_{32}^{(t)}\bar{\varphi}_3^{(t)} - \psi_{33}^{(t)}\bar{\varphi}_1^{(t)} \\ \psi_{31}^{(t)}\bar{\varphi}_3^{(t)} - \psi_{33}^{(t)}\bar{\varphi}_2^{(t)} \end{bmatrix}, \quad (91)$$

where $\varphi^{(t)}(\xi, x, t)$ and $\bar{\varphi}^{(t)}(\xi, x, t)$ are defined as

$$\varphi^{(t)} \equiv \bar{s}_{22}(\xi)\psi_1^{(t)} - \bar{s}_{21}(\xi)\psi_2^{(t)}, \quad \bar{\varphi}^{(t)} \equiv s_{22}(\xi)\bar{\psi}_1^{(t)} - s_{12}(\xi)\bar{\psi}_2^{(t)}, \quad (92)$$

$\Phi^{(t)}$ and $\bar{\Phi}^{(t)}$ have been defined in (67) and (68), and $\Psi^{(t)}$ and $\bar{\Psi}^{(t)}$ are defined as

$$\Psi^{(t)} = \Psi e^{-4i\xi^3\Lambda t}, \quad \bar{\Psi}^{(t)} = e^{4i\xi^3\Lambda t}\bar{\Psi}. \quad (93)$$

Notice from (71) and the symmetry condition (10) that $(s_{12}, s_{22}, \bar{s}_{21}, \bar{s}_{22})$ are time independent. Similar to what we have done in the previous section, we can show that $\varphi^{(t)}$ and $\Psi^{(t)}$ in definition (91) are Jost functions satisfying both the Lax pair (2) and (66), while functions $\bar{\varphi}^{(t)}$ and $\bar{\Psi}^{(t)}$ satisfy both the adjoint Lax pair. In addition, $[\Omega_1^{(t)}, \Omega_2^{(t)}]$ are related to $[\Omega_1, \Omega_2]$ as

$$\Omega_1^{(t)} = \Omega_1 e^{-8i\xi^3 t}, \quad \Omega_2^{(t)} = \Omega_2 e^{8i\xi^3 t}. \quad (94)$$

Inserting this relation and (73) into the time-restored equations (89) and (90), we get the new relations

$$\delta\rho_1(\xi, 0) = \frac{1}{s_{33}^2(\xi)} \left\langle \Omega_1^{(t)}(\xi; x, t), \begin{bmatrix} \delta u(x, t) \\ \delta u^*(x, t) \end{bmatrix} \right\rangle, \quad (95)$$

$$\delta\bar{\rho}_1(\xi, 0) = \frac{1}{\bar{s}_{33}^2(\xi)} \left\langle \Omega_2^{(t)}(\xi; x, t), \begin{bmatrix} \delta u(x, t) \\ \delta u^*(x, t) \end{bmatrix} \right\rangle. \quad (96)$$

These two new relations are associated with the new expansion (74) in the previous section.

The functions $(\Omega_1^{(t)}, \Omega_2^{(t)})$ appearing in the above variations of scattering data formulas (95) and (96) are precisely adjoint squared eigenfunctions satisfying the adjoint linearized Sasa-Satsuma equation. Analogous facts for several other integrable equations have been noted before by direct verification (see Ref. 6). Below we will present a general proof of this fact which will hold for those integrable equations where we have relations similar to (74), (95), and (96). Along the way, we will also obtain the closure relation, orthogonality relations, and inner products between the squared eigenfunctions and the adjoint squared eigenfunctions.

First, we insert Eqs. (95) and (96) into expansion (74). Exchanging the order of integration, we get

$$\begin{aligned} \begin{bmatrix} \delta u(x, t) \\ \delta u^*(x, t) \end{bmatrix} = & -\frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{1}{s_{33}^2(\xi)} Z_1^{(t)}(\xi; x, t) \Omega_1^{(t)T}(\xi; x', t) + \frac{1}{\bar{s}_{33}^2(\xi)} Z_2^{(t)} \right. \\ & \left. \times (\xi; x, t) \Omega_2^{(t)T}(\xi; x', t) \right] d\xi \begin{bmatrix} \delta u(x', t) \\ \delta u^*(x', t) \end{bmatrix} dx'. \end{aligned} \quad (97)$$

Since we are considering that $(\delta u, \delta u^*)$ are arbitrary localized functions (in an appropriate functional space such as L_1), in order for the above equation to hold, we must have

$$-\frac{1}{\pi} \int_{-\infty}^{\infty} \left[\frac{1}{s_{33}^2(\xi)} Z_1^{(t)}(\xi; x, t) \Omega_1^{(t)T}(\xi; x', t) + \frac{1}{\bar{s}_{33}^2(\xi)} Z_2^{(t)}(\xi; x, t) \Omega_2^{(t)T}(\xi; x', t) \right] d\xi = \delta(x - x') 1, \quad (98)$$

which is the closure relation (discrete-spectrum contributions are absent here due to our assumption of $s_{33}, \bar{s}_{33}$ having no zeros). Here $\delta(x)$ is the Dirac delta function. This closure relation has the usually symmetry in that not only do the squared eigenfunctions $\{Z_1^{(t)}(\xi; x, t), Z_2^{(t)}(\xi; x, t), \xi \in \mathbb{R}\}$ form a complete set but also the adjoint functions $\{\Omega_1^{(t)}(\xi; x, t), \Omega_2^{(t)}(\xi; x, t), \xi \in \mathbb{R}\}$ form a complete set as well.

To obtain the inner products and orthogonality relations between these functions, we reverse the above and insert expansion (74) into Eqs. (95) and (96). Exchanging the order of integration, we get

$$\begin{bmatrix} \delta \rho_1(\xi, 0) \\ \delta \bar{\rho}_1(\xi, 0) \end{bmatrix} = \int_{-\infty}^{\infty} M(\xi, \xi', t) \begin{bmatrix} \delta \rho_1(\xi', 0) \\ \delta \bar{\rho}_1(\xi', 0) \end{bmatrix} d\xi', \quad (99)$$

where

$$M(\xi, \xi', t) = \begin{pmatrix} -\frac{1}{\pi s_{33}^2(\xi)} \langle \Omega_1^{(t)}(\xi; x, t), Z_1^{(t)}(\xi'; x, t) \rangle, & -\frac{1}{\pi \bar{s}_{33}^2(\xi)} \langle \Omega_1^{(t)}(\xi; x, t), Z_2^{(t)}(\xi'; x, t) \rangle \\ -\frac{1}{\pi \bar{s}_{33}^2(\xi)} \langle \Omega_2^{(t)}(\xi; x, t), Z_1^{(t)}(\xi'; x, t) \rangle, & -\frac{1}{\pi s_{33}^2(\xi)} \langle \Omega_2^{(t)}(\xi; x, t), Z_2^{(t)}(\xi'; x, t) \rangle \end{pmatrix}. \quad (100)$$

Since relation (99) is taken to be valid for arbitrary functions of $\delta \rho_1(\xi, 0)$ and $\delta \bar{\rho}_1(\xi, 0)$,

$$M(\xi, \xi', t) = 1 \delta(\xi - \xi'). \quad (101)$$

Consequently we get the inner products and orthogonality relations between squared eigenfunctions $[Z_1^{(t)}, Z_2^{(t)}]$ and functions $[\Omega_1^{(t)}, \Omega_2^{(t)}]$ as

$$\langle \Omega_1^{(t)}(\xi; x, t), Z_1^{(t)}(\xi'; x, t) \rangle = -\pi s_{33}^2(\xi) \delta(\xi - \xi'), \quad (102)$$

$$\langle \Omega_2^{(t)}(\xi; x, t), Z_2^{(t)}(\xi'; x, t) \rangle = -\pi \bar{s}_{33}^2(\xi) \delta(\xi - \xi'), \quad (103)$$

$$\langle \Omega_1^{(t)}(\xi; x, t), Z_2^{(t)}(\xi'; x, t) \rangle = \langle \Omega_2^{(t)}(\xi; x, t), Z_1^{(t)}(\xi'; x, t) \rangle = 0 \quad (104)$$

for any $\xi, \xi' \in \mathbb{R}$.

To show that (Ω_1, Ω_2) are adjoint squared eigenfunctions which satisfy the adjoint linearized equation of the Sasa–Satsuma equation, we first separate the temporal derivatives from the spatial ones in the linearized operator $\mathcal{L}$ as

$$\mathcal{L} = 1\partial_t + L, \quad (105)$$

where L only involves spatial derivatives. The adjoint linearized operator is then

$$\mathcal{L}^A = -1\partial_t + L^A, \quad (106)$$

where L^A is the adjoint operator of L . Now we take the inner product between the equation $\mathcal{L}Z_1^{(t)}(\xi'; x, t) = 0$ and function $\Omega_1^{(t)}(\xi; x, t)$. When Eq. (105) is inserted into it, we find that

$$\langle LZ_1^{(t)}(\xi'), \Omega_1^{(t)}(\xi) \rangle - \langle Z_1^{(t)}(\xi'), \partial_t \Omega_1^{(t)}(\xi) \rangle + \partial_t \langle Z_1^{(t)}(\xi'), \Omega_1^{(t)}(\xi) \rangle = 0. \quad (107)$$

Here and immediately below, the (x, t) dependence of $Z_{1,2}^{(t)}$ and $\Omega_{1,2}^{(t)}$ is suppressed for notational simplicity. Recall that s_{33} is time independent for the Sasa–Satsuma equation; thus from the inner-product equation (102) we see that $\partial_t \langle Z_1^{(t)}(\xi'), \Omega_1^{(t)}(\xi) \rangle = 0$. Using integration by parts, the first term in the above equation can be rewritten as

$$\langle LZ_1^{(t)}(\xi'), \Omega_1^{(t)}(\xi) \rangle = \langle Z_1^{(t)}(\xi'), L^A \Omega_1^{(t)}(\xi) \rangle + W(\xi', \xi)|_{x=-L}^{x=L}, \quad L \rightarrow \infty, \quad (108)$$

where function $W(\xi', \xi; x, t)$ contains terms which are generated during integration by parts. These terms are quadratic combinations of $Z_1^{(t)}(\xi')$, $\Omega_1^{(t)}(\xi)$, and their spatial derivatives. When $x = \pm L \rightarrow \pm \infty$, each term has the form $f(\xi')g(\xi)\exp(\pm i\xi' L \pm i\xi L \pm 4i\xi'^3 t \pm 4i\xi^3 t)$ due to the large- x asymptotics of Jost functions Φ and Ψ . Here $f(\xi')$ and $g(\xi)$ are related to scattering coefficients and are continuous functions. Then in the sense of generalized functions, the last term in Eq. (108) is zero due to the Riemann–Lebesgue lemma. Inserting the above results into Eq. (107), we find that

$$\langle Z_1^{(t)}(\xi'), \mathcal{L}^A \Omega_1^{(t)}(\xi) \rangle = 0, \quad \xi', \xi \in \mathbb{R}, \quad (109)$$

i.e., $\mathcal{L}^A \Omega_1^{(t)}(\xi)$ is orthogonal to all $Z_1^{(t)}(\xi')$. By taking the inner product between the equation $\mathcal{L}Z_2^{(t)}(\xi') = 0$ and function $\Omega_1^{(t)}(\xi)$ and doing similar calculations, we readily find that

$$\langle Z_2^{(t)}(\xi'), \mathcal{L}^A \Omega_1^{(t)}(\xi) \rangle = 0, \quad \xi', \xi \in \mathbb{R}. \quad (110)$$

In other words, $\mathcal{L}^A \Omega_1^{(t)}(\xi)$ is also orthogonal to all $Z_2^{(t)}(\xi')$. Since $\{Z_1^{(t)}(\xi'), Z_2^{(t)}(\xi'), \xi' \in \mathbb{R}\}$ forms a complete set, Eqs. (109) and (110) dictate that $\mathcal{L}^A \Omega_1^{(t)}$ has to be zero for any ξ, x and t , i.e.,

$$\mathcal{L}^A \Omega_1^{(t)}(\xi; x, t) = 0, \quad \xi \in \mathbb{R}. \quad (111)$$

Thus $\Omega_1^{(t)}(\xi)$ is an adjoint squared eigenfunction that satisfies the adjoint linearized Sasa–Satsuma equation for every $\xi \in \mathbb{R}$. Similarly, it can be shown that $\Omega_2^{(t)}(\xi)$ is also an adjoint squared eigenfunction for every $\xi \in \mathbb{R}$. In short, $\{\Omega_1^{(t)}(\xi), \Omega_2^{(t)}(\xi), \xi \in \mathbb{R}\}$ is a complete set of adjoint squared eigenfunctions.

Like squared eigenfunctions $(Z_1^{(t)}, Z_2^{(t)})$, these adjoint squared eigenfunctions $(\Omega_1^{(t)}, \Omega_2^{(t)})$ are also quadratic combinations of Jost functions and adjoint Jost functions. In addition, they have nice analytic properties as well. To see this latter fact, notice from definition (85) that $\varphi e^{i\xi x}$ is analytic in $\mathbb{C}_-$ since $\bar{s}_{21}, \bar{s}_{22}, \psi_1 e^{i\xi x}$ and $\psi_2 e^{i\xi x}$ are analytic in $\mathbb{C}_-$. Then recalling that $\bar{\psi}_3 e^{i\xi x}$ is also analytic in $\mathbb{C}_-$, we see that $\Omega_1^{(t)}(\xi) e^{2i\xi x}$ is analytic in $\mathbb{C}_-$. Similarly, $\Omega_2^{(t)}(\xi) e^{-2i\xi x}$ is analytic in $\mathbb{C}_+$.

It is worthy to point out that in the above adjoint squared eigenfunctions (91), if $\varphi^{(t)}$ is replaced by the Jost function $\psi_1^{(t)}$ (or $\psi_2^{(t)}$) or if $\bar{\varphi}^{(t)}$ is replaced by the adjoint Jost function $\bar{\psi}_1^{(t)}$ (or $\bar{\psi}_2^{(t)}$), the resulting functions would still satisfy the adjoint linearized Sasa–Satsuma equation. Thus one may be tempted to simply take $\varphi^{(t)} = \psi_k^{(t)}$, $\bar{\varphi}^{(t)} = \bar{\psi}_k^{(t)}$ ($k=1$ or 2) rather than (92) as adjoint squared eigenfunctions. The problem with these “simpler” adjoint squared eigenfunctions is that their inner products with squared eigenfunctions (69) are not what they are supposed to be [see (102)–(104)], neither does one get the closure relation (98). Thus the adjoint squared eigenfunctions (91) coming from our systematic procedure of variation calculations are the correct ones to use.

C. Extension to the general case

In the previous subsections, the squared eigenfunctions and their closure relation were established under the assumption that s_{33} and $\bar{s}_{33}$ have no zeros, i.e., the spectral equation (2) has no discrete eigenvalues. In this subsection, we extend those results to the general case where s_{33} and $\bar{s}_{33}$ have zeros in their respective planes of analyticity $\mathbb{C}_{\pm}$. Due to the symmetry properties (10) and (12), we have

$$\bar{s}_{33}(\zeta) = s_{33}^*(\zeta^*), \quad s_{33}(\zeta) = \bar{s}_{33}^*(-\zeta^*). \quad (112)$$

Thus if $\zeta_j \in \mathbb{C}_-$ is a zero of s_{33} , i.e., $s_{33}(\zeta_j) = 0$, then $-\zeta_j^* \in \mathbb{C}_-$ is also a zero of s_{33} , and $\{\zeta_j^*, -\zeta_j\} \in \mathbb{C}_+$ are both zeros of $\bar{s}_{33}$. This means that zeros of s_{33} and $\bar{s}_{33}$ always appear in quadruples. Suppose all the zeros of s_{33} and $\bar{s}_{33}$ are $\zeta_j \in \mathbb{C}_-$ and $\bar{\zeta}_j \in \mathbb{C}_+$, $j = 1, \dots, 2N$, respectively. For simplicity, we also assume that all zeros are simple.

In this general case with zeros, squared eigenfunctions $Z_1^{(t)}, Z_2^{(t)}$ clearly still satisfy the linearized Sasa–Satsuma equation, and adjoint squared eigenfunctions $\Omega_1^{(t)}, \Omega_2^{(t)}$ still satisfy the adjoint linearized Sasa–Satsuma equation—facts which will not change when s_{33} and $\bar{s}_{33}$ have zeros. The main difference from the previous no-zero case is that, the sets of (continuous) squared eigenfunctions $\{Z_1^{(t)}(\xi; x), Z_2^{(t)}(\xi; x), \xi \in \mathbb{R}\}$ and adjoint squared eigenfunctions $\{\Omega_1^{(t)}(\xi; x), \Omega_2^{(t)}(\xi; x), \xi \in \mathbb{R}\}$ are no longer complete, i.e., the closure relation (98) does not hold any more, and contributions from the discrete spectrum must be included now. To account for discrete-spectrum contributions, one could proceed by adding variations in the discrete scattering data in the above derivations. But a much easier way is to simply pick up the pole contributions to the integrals in the closure relation (98), as we will do below. Recall from the previous subsections that $Z_1^{(t)} \times (\zeta; x, t)e^{-2i\zeta x}$ and $\Omega_1^{(t)}(\zeta; x, t)e^{2i\zeta x}$ are analytic in $\mathbb{C}_-$, while $Z_2^{(t)}(\zeta; x, t)e^{2i\zeta x}$ and $\Omega_2^{(t)}(\zeta; x, t)e^{-2i\zeta x}$ are analytic in $\mathbb{C}_+$. In addition, from the large- ζ asymptotics (31), (32), and (35) and symmetry relations (46), we easily find that

$$Z_1^{(t)}(\zeta; x, t) \rightarrow e^{2i\zeta x + 8i\zeta^3 t}(0, 1)^T, \quad Z_2^{(t)}(\zeta; x, t) \rightarrow e^{-2i\zeta x - 8i\zeta^3 t}(1, 0)^T, \quad \zeta \rightarrow \infty, \quad (113)$$

$$\Omega_1^{(t)}(\zeta; x, t) \rightarrow -e^{-2i\zeta x - 8i\zeta^3 t}(0, 1)^T, \quad \Omega_2^{(t)}(\zeta; x, t) \rightarrow -e^{2i\zeta x + 8i\zeta^3 t}(1, 0)^T, \quad \zeta \rightarrow \infty. \quad (114)$$

Thus,

$$\int_{C^-} \frac{1}{s_{33}^2(\zeta)} Z_1^{(t)}(\zeta; x, t) \Omega_1^{(t)T}(\zeta; x', t) d\zeta = - \int_{C^-} e^{2i\zeta(x-x')} d\zeta \text{diag}(0, 1), \quad (115)$$

where the integration path C^- is the lower semicircle of infinite radius in anticlockwise direction. Since the function $e^{2i\zeta(x-x')}$ in the right hand side of the above equation is analytic, its integration path can be brought up to the real axis $\zeta \in \mathbb{R}$. In the sense of generalized functions, that integral is equal to $\pi\delta(x-x')$; thus

$$\int_{C^-} \frac{1}{s_{33}^2(\zeta)} Z_1^{(t)}(\zeta; x, t) \Omega_1^{(t)T}(\zeta; x', t) d\zeta = -\pi\delta(x-x')\text{diag}(0, 1). \quad (116)$$

Doing the same calculation for the integral of $Z_2^{(t)}\Omega_2^{(t)T}/\bar{s}_{33}^2$ along the upper semicircle of infinite radius C^+ in clockwise direction and combining it with the above equation, we get

$$-\frac{1}{\pi} \int_{C^-} \frac{1}{s_{33}^2(\zeta)} Z_1^{(t)}(\zeta; x, t) \Omega_1^{(t)T}(\zeta; x', t) d\zeta - \frac{1}{\pi} \int_{C^+} \frac{1}{\bar{s}_{33}^2(\zeta)} Z_2^{(t)}(\zeta; x, t) \Omega_2^{(t)T}(\zeta; x', t) d\zeta = \delta(x-x')1. \quad (117)$$

Now we evaluate the two integrals in the above equation by bringing down the integration paths to the real axis and picking up pole contributions by the residue theorem. Recalling that zeros of

s_{33} and $\bar{s}_{33}$ are simple, the poles in the above integrals are second order. After simple calculations, we get the closure relation for the general case (with discrete-spectrum contributions) as

$$\begin{aligned}
 & -\frac{1}{\pi} \int_{-\infty}^{\infty} \left[\frac{1}{s_{33}^2(\xi)} Z_1^{(i)}(\xi; x, t) \Omega_1^{(i)T}(\xi; x', t) + \frac{1}{\bar{s}_{33}^2(\xi)} Z_2^{(i)}(\xi; x, t) \Omega_2^{(i)T}(\xi; x', t) \right] d\xi - \sum_{j=1}^{2N} \frac{2i}{s_{33}'^2(\bar{\zeta}_j)} \left[Z_1^{(i)} \right. \\
 & \quad \times (\bar{\zeta}_j; x, t) \Theta_1^{(i)T}(\bar{\zeta}_j; x', t) + \frac{\partial Z_1^{(i)}}{\partial \bar{\zeta}}(\bar{\zeta}_j; x, t) \Omega_1^{(i)T}(\bar{\zeta}_j; x', t) \left. \right] + \sum_{j=1}^{2N} \frac{2i}{\bar{s}_{33}'^2(\zeta_j)} \left[Z_2^{(i)}(\zeta_j; x, t) \Theta_2^{(i)T}(\zeta_j; x', t) \right. \\
 & \quad \left. + \frac{\partial Z_2^{(i)}}{\partial \zeta}(\zeta_j; x, t) \Omega_2^{(i)T}(\zeta_j; x', t) \right] = \delta(x - x') 1, \quad (118)
 \end{aligned}$$

where

$$\Theta_1^{(i)}(\bar{\zeta}_j; x, t) = \frac{\partial \Omega_1^{(i)}}{\partial \bar{\zeta}}(\bar{\zeta}_j; x, t) - \frac{s_{33}''(\bar{\zeta}_j)}{s_{33}'(\bar{\zeta}_j)} \Omega_1^{(i)}(\bar{\zeta}_j; x, t), \quad (119)$$

$$\Theta_2^{(i)}(\zeta_j; x, t) = \frac{\partial \Omega_2^{(i)}}{\partial \zeta}(\zeta_j; x, t) - \frac{\bar{s}_{33}''(\zeta_j)}{\bar{s}_{33}'(\zeta_j)} \Omega_2^{(i)}(\zeta_j; x, t). \quad (120)$$

which have been termed “derivative states”^{1,2} due to the differentiation with respect to ζ . Clearly, $Z_1^{(i)}(\bar{\zeta}_j)$, $\partial Z_1^{(i)} / \partial \bar{\zeta}(\bar{\zeta}_j)$, $Z_2^{(i)}(\zeta_j)$, and $\partial Z_2^{(i)} / \partial \zeta(\zeta_j)$ satisfy the linearized Sasa–Satsuma equation and are thus discrete squared eigenfunctions, while $\Omega_1^{(i)}(\bar{\zeta}_j)$, $\Theta_1^{(i)}(\bar{\zeta}_j)$, $\Omega_2^{(i)}(\zeta_j)$, and $\Theta_2^{(i)}(\zeta_j)$ satisfy the adjoint linearized Sasa–Satsuma equation and are thus discrete adjoint squared eigenfunctions.

In soliton perturbation theories, explicit expressions for squared eigenfunctions under soliton potentials are needed. For such potentials, $s_{31}=s_{32}=\bar{s}_{13}=\bar{s}_{23}=0$; thus $G=1$ in the Riemann–Hilbert problem (33). Solutions of this Riemann–Hilbert problem have been solved completely for any number of zeros and any orders of their algebraic and geometric multiplicities.²⁵ In the simplest case where all zeros of the solitons are simple, solutions $P_{\pm}$ can be found in Eq. 76 of Ref. 25 (setting $r^{(i)}=r^{(j)}=1$). One only needs to keep in mind here that the zeros of s_{33} and $\bar{s}_{33}$ always appear in quadruples due to the symmetries (112). Once solutions $P_{\pm}$ are obtained, the solitons $u(x, t)$ follow from Eq. (37). Taking the large- x asymptotics of $P_{\pm}$ and utilizing the symmetry conditions (10), the full scattering matrix S can be readily obtained. Then together with $P_{\pm}$, explicit expressions for all the Jost functions Φ , Ψ as well as squared eigenfunctions are then derived.

IV. SUMMARY AND DISCUSSION

In this paper, squared eigenfunctions were derived for the Sasa–Satsuma equation. It was shown that these squared eigenfunctions are sums of two terms, where each term is a product of a Jost function and an adjoint Jost function. The procedure of this derivation consists of two steps: One is to calculate the variations of the potentials via variations of the scattering data by the Riemann–Hilbert method. The other one is to calculate variations of the scattering data via variations of the potentials through elementary calculations. It was proved that the functions appearing in these variation relations are precisely the squared eigenfunctions and adjoint squared eigenfunctions satisfying, respectively, the linearized equation and the adjoint linearized equation of the Sasa–Satsuma system. More importantly, our proof is quite general and also holds for other integrable equations. Since the spectral operator of the Sasa–Satsuma equation is 3×3 while its linearized operator is 2×2 , we demonstrated how the two-component squared eigenfunctions are built from only selected components of the three-component Jost functions, and we have seen that symmetry properties of Jost functions and the scattering data play an important role here.

The derivation used in this paper for squared eigenfunctions is a universal technique. Our main contributions to this method are triplefold. First, we showed that for a general integrable

system, if one can obtain the variational relations between the potentials and the scattering data in terms of the components of the Jost functions and their adjoints, then from these variational relations one can immediately identify the squared eigenfunctions and adjoint squared eigenfunctions which satisfy, respectively, the linearized integrable equation and its adjoint equation. Second, we showed that from these variational relations, one can read off the values of nonzero inner products between squared eigenfunctions and adjoint squared eigenfunctions. Thus no longer is it necessary to calculate these inner products from the Wronskian relations and the asymptotics of the Jost functions. Third, we showed that from these variational relations, one can immediately obtain the closure relation of squared eigenfunctions and adjoint squared eigenfunctions. After squared eigenfunctions of an integrable equation are obtained, one then can readily derive the recursion operator for that integrable equation, since the squared eigenfunctions are eigenfunctions of the recursion operator.^{8–11,19}

An important remark about this derivation of squared eigenfunctions is that it uses primarily the spectral operator of the Lax pair. Indeed, the key variation relations (65), (89), and (90) were obtained exclusively from the spectral operator (2), while the time-dependent versions of these relations (74), (95), and (96) were obtained by using the asymptotic form of the time evolution equation in the Lax pair as $|x| \rightarrow \infty$. This means that all integrable equations with the same spectral operator would share the same squared eigenfunctions (except an exponential-in-time factor which is equation dependent). For instance, a whole hierarchy of integrable equations would possess the “same” squared eigenfunctions, since the spectral operators of a hierarchy are the same. This readily reproduces earlier results in Refs. 10, 11, and 19 where such results were obtained by the commutability relations between the linearized operators and the recursion operator of a hierarchy.

Since the derivation in this paper uses primarily the spectral operator of the Lax pair, if two integrable equations have similar spectral operators, derivations of their squared eigenfunctions would be similar. For instance, the spectral operator of the Manakov equations is very similar to that of the Sasa–Satsuma equation [except that the potential term Q in the Sasa–Satsuma spectral operator (20) possesses one more symmetry]. Making very minor modifications to the calculations for the Sasa–Satsuma equation in the text, we can obtain squared eigenfunctions for the Manakov equations. This will be demonstrated in the Appendix.

With the squared eigenfunctions obtained for the Sasa–Satsuma equation, a perturbation theory for Sasa–Satsuma solitons can now be developed. This problem lies outside the scope of the present article and will be left for future studies.

ACKNOWLEDGMENTS

We thank Dr. Taras Lakoba for very helpful discussions. The work of J.Y. was supported in part by the Air Force Office of Scientific Research under Grant No. USAF 9550-05-1-0379. The work of D.J.K. has been supported in part by the U.S. National Science Foundation under Grant No. DMS-0505566.

APPENDIX: SQUARED EIGENFUNCTIONS FOR THE MANAKOV EQUATION

The Manakov equations are²⁶

$$iu_t + u_{xx} + 2(|u|^2 + |v|^2)u = 0, \quad (\text{A1})$$

$$iv_t + v_{xx} + 2(|u|^2 + |v|^2)v = 0. \quad (\text{A2})$$

The spectral operator for the Manakov equations is Eq. (2) with

$$Q = \begin{pmatrix} 0 & 0 & u \\ 0 & 0 & v \\ -u^* & -v^* & 0 \end{pmatrix}. \quad (\text{A3})$$

Below we will use the same notations on Jost functions and the scattering matrix as in the main text. Since Q above is also anti-Hermitian, Jost functions and the scattering matrix satisfy the involution properties (9) and (10) as well. The main difference between the Sasa–Satsuma equation and the Manakov equations is that the potential Q above for the Manakov equations does not possess the additional symmetry (4); thus the Jost functions and the scattering matrix do not possess the symmetries (11) and (12). In addition, the asymptotics (55) also needs minor modification. It is noted that the spectral operator of the Manakov equations is 3×3 , while the linearized operator of these equations is 4×4 . So one needs to build four-component squared eigenfunctions from three-component Jost functions. This contrasts the Sasa–Satsuma equation where one builds two-component squared eigenfunctions from three-component Jost functions.

Repeating the same calculations as in the main text with the above minor modifications kept in mind, we still get Eq. (57) for the variations $(\delta u, \delta u^*)$, while the expressions for variations $(\delta v, \delta v^*)$ are Eq. (57) with Π_{13} and Π_{31} replaced by Π_{23} and Π_{32} . The expression for matrix Π is still given by Eq. (60). Thus variations of the potentials via variations of the scattering data are found to be

$$[\delta u, \delta v, \delta u^*, \delta v^*]^T = -\frac{1}{\pi} \int_{-\infty}^{\infty} [Z_1(\xi; x) \delta \rho_1(\xi) + Z_2(\xi; x) \delta \rho_2(\xi) + Z_3(\xi; x) \delta \bar{\rho}_1(\xi) + Z_4(\xi; x) \delta \bar{\rho}_2(\xi)] d\xi, \quad (\text{A4})$$

where

$$Z_1 = \begin{bmatrix} \phi_{31} \bar{\phi}_{13} \\ \phi_{32} \bar{\phi}_{13} \\ \phi_{33} \bar{\phi}_{11} \\ \phi_{33} \bar{\phi}_{12} \end{bmatrix}, \quad Z_2 = \begin{bmatrix} \phi_{31} \bar{\phi}_{23} \\ \phi_{32} \bar{\phi}_{23} \\ \phi_{33} \bar{\phi}_{21} \\ \phi_{33} \bar{\phi}_{22} \end{bmatrix}, \quad Z_3 = \begin{bmatrix} \phi_{11} \bar{\phi}_{33} \\ \phi_{12} \bar{\phi}_{33} \\ \phi_{13} \bar{\phi}_{31} \\ \phi_{13} \bar{\phi}_{32} \end{bmatrix}, \quad Z_4 = \begin{bmatrix} \phi_{21} \bar{\phi}_{33} \\ \phi_{22} \bar{\phi}_{33} \\ \phi_{23} \bar{\phi}_{31} \\ \phi_{23} \bar{\phi}_{32} \end{bmatrix}. \quad (\text{A5})$$

Variations of the scattering matrix δS are still given by Eq. (83), with δQ modified in view of the form of (A3). Then defining vectors

$$\varphi_1 = (\varphi_{1j}) \equiv \bar{s}_{22} \psi_1 - \bar{s}_{21} \psi_2, \quad \varphi_2 = (\varphi_{2j}) \equiv -\bar{s}_{12} \psi_1 + \bar{s}_{11} \psi_2, \quad (\text{A6})$$

$$\bar{\varphi}_1 = (\bar{\varphi}_{1j}) \equiv s_{22} \bar{\psi}_1 - s_{12} \bar{\psi}_2, \quad \bar{\varphi}_2 = (\bar{\varphi}_{2j}) \equiv -s_{21} \bar{\psi}_1 + s_{11} \bar{\psi}_2, \quad (\text{A7})$$

we find that variations of the scattering data are

$$\delta \rho_1(\xi) = \frac{1}{s_{33}^2(\xi)} \left\langle \Omega_1(\xi; x), \begin{bmatrix} \delta u(x) \\ \delta v(x) \\ \delta u^*(x) \\ \delta v^*(x) \end{bmatrix} \right\rangle, \quad \delta \rho_2(\xi) = \frac{1}{s_{33}^2(\xi)} \left\langle \Omega_2(\xi; x), \begin{bmatrix} \delta u(x) \\ \delta v(x) \\ \delta u^*(x) \\ \delta v^*(x) \end{bmatrix} \right\rangle, \quad (\text{A8})$$

$$\delta \bar{\rho}_1(\xi) = \frac{1}{\bar{s}_{33}^2(\xi)} \left\langle \Omega_3(\xi; x), \begin{bmatrix} \delta u(x) \\ \delta v(x) \\ \delta u^*(x) \\ \delta v^*(x) \end{bmatrix} \right\rangle, \quad \delta \bar{\rho}_2(\xi) = \frac{1}{\bar{s}_{33}^2(\xi)} \left\langle \Omega_4(\xi; x), \begin{bmatrix} \delta u(x) \\ \delta v(x) \\ \delta u^*(x) \\ \delta v^*(x) \end{bmatrix} \right\rangle, \quad (\text{A9})$$

where

$$\Omega_1 = \begin{bmatrix} \bar{\psi}_{31}\varphi_{13} \\ \bar{\psi}_{32}\varphi_{13} \\ -\bar{\psi}_{33}\varphi_{11} \\ -\bar{\psi}_{33}\varphi_{12} \end{bmatrix}, \quad \Omega_2 = \begin{bmatrix} \bar{\psi}_{31}\varphi_{23} \\ \bar{\psi}_{32}\varphi_{23} \\ -\bar{\psi}_{33}\varphi_{21} \\ -\bar{\psi}_{33}\varphi_{22} \end{bmatrix}, \quad \Omega_3 = \begin{bmatrix} -\bar{\varphi}_{11}\psi_{33} \\ -\bar{\varphi}_{12}\psi_{33} \\ \bar{\varphi}_{13}\psi_{31} \\ \bar{\varphi}_{13}\psi_{32} \end{bmatrix}, \quad \Omega_4 = \begin{bmatrix} -\bar{\varphi}_{21}\psi_{33} \\ -\bar{\varphi}_{22}\psi_{33} \\ \bar{\varphi}_{23}\psi_{31} \\ \bar{\varphi}_{23}\psi_{32} \end{bmatrix}. \quad (\text{A10})$$

For Manakov equations, the time evolution operator of the Lax pair is

$$Y_t = -2i\zeta^2 \Lambda Y + V(\zeta, u)Y, \quad (\text{A11})$$

where V goes to zero as x approaches infinity. Introducing time-dependent Jost functions and adjoint Jost functions such as

$$\Phi^{(t)} = \Phi e^{-2i\zeta^2 \Lambda t}, \quad \bar{\Phi}^{(t)} = e^{2i\zeta^2 \Lambda t} \bar{\Phi} \quad (\text{A12})$$

and use them to replace the original Jost functions and adjoint Jost functions in the Z_k and Ω_k definitions (A5) and (A10), the resulting functions $Z_{1,2,3,4}^{(t)}$ and $\Omega_{1,2,3,4}^{(t)}$ would be squared eigenfunctions and adjoint squared eigenfunctions of the Manakov equations. Equivalently, $\{Z_1 e^{-4i\zeta^2 t}, Z_2 e^{-4i\zeta^2 t}, Z_3 e^{4i\zeta^2 t}, Z_4 e^{4i\zeta^2 t}\}$ are squared eigenfunctions and $\{\Omega_1 e^{4i\zeta^2 t}, \Omega_2 e^{4i\zeta^2 t}, \Omega_3 e^{-4i\zeta^2 t}, \Omega_4 e^{-4i\zeta^2 t}\}$ adjoint squared eigenfunctions of the Manakov equations.

It is noted that squared eigenfunctions for the Manakov equations have been given in Ref. 18 under different notations and derivations. Our expressions above are more explicit and easier to use.

¹ D. J. Kaup, *SIAM J. Appl. Math.* **31**, 121 (1976).

² D. J. Kaup, *J. Math. Anal. Appl.* **54**, 849 (1976).

³ A. C. Newell, in *Solitons*, Topics in Current Physics Vol. 17, edited by R. K. Bullough and P. J. Caudrey (Springer-Verlag, New York, 1980), pp. 177–242.

⁴ D. J. Kaup and T. I. Lakoba, *J. Math. Phys.* **37**, 308 (1996).

⁵ X. Chen and J. Yang, *Phys. Rev. E* **65**, 066608 (2002).

⁶ D. J. Kaup, T. I. Lakoba, and Y. Matsuno, *Inverse Probl.* **15**, 215 (1999).

⁷ R. L. Herman, *J. Phys. A* **23**, 2327 (1990).

⁸ M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, *Stud. Appl. Math.* **53**, 249 (1974).

⁹ A. S. Fokas and P. M. Santini, *Stud. Appl. Math.* **75**, 179 (1986).

¹⁰ J. Yang, *J. Math. Phys.* **41**, 6614 (2000).

¹¹ J. Yang, *Stud. Appl. Math.* **108**, 145 (2002).

¹² Y. Zeng, W. Ma, and R. Lin, *J. Math. Phys.* **41**, 5453 (2000).

¹³ H. Wang, X. Hu, and Gegenhasi, *J. Comput. Appl. Math.* **202**, 133 (2007).

¹⁴ D. J. Kaup, *Phys. Rev. A* **42**, 5689 (1990).

¹⁵ J. Yang, *Phys. Rev. Lett.* **91**, 143903 (2003).

¹⁶ J. Yang and X. Chen, in *Trends in Soliton Research*, edited by L. V. Chen (Nova, New York, 2006), pp. 15–27.

¹⁷ D. J. Kaup, *J. Math. Phys.* **25**, 2467 (1984).

¹⁸ V. S. Gerdjikov, in *Geometry, Integrability and Quantization-VI*, edited by I. M. Mladenov and A. C. Hirshfeld (Softex, Sofia, 2005), pp. 78–122.

¹⁹ A. S. Fokas, *J. Math. Phys.* **21**, 1318 (1980).

²⁰ A. Sergyeyev and D. Demskoi, *J. Math. Phys.* **48**, 042702 (2007).

²¹ A. Hasegawa and Y. Kodama, *Solitons in Optical Communications* (Clarendon, Oxford, 1995).

²² G. P. Agrawal, *Nonlinear Fiber Optics* (Academic, San Diego, 1989).

²³ N. Sasa and J. Satsuma, *J. Phys. Soc. Jpn.* **60**, 409 (1991).

²⁴ J. Yang, B. A. Malomed, and D. J. Kaup, *Phys. Rev. Lett.* **83**, 1958 (1999).

²⁵ V. S. Shchesnovich and J. Yang, *J. Math. Phys.* **44**, 4604 (2003).

²⁶ S. V. Manakov, *Zh. Eksp. Teor. Fiz.* **65**, 1392 (1973) [*Sov. Phys. JETP* **38**, 248 (1974)].

**“Frontiers of Nonlinear Physics”
“Proceedings of the Second International Conference”
Nizhny Novgorod – St. Petersburg, Russia, 2004
edited by Alexander Litvak
(Nizhny Novgorod, Russia, Institute of Applied Physics RAS, 2005)**

CONTENTS

<i>Preface</i>	3
Plenary Talks	
<i>L. V. Butov</i> . Condensation and pattern formation in cold exciton gases in coupled quantum wells.....	9
<i>N. J. Fisch</i> . Compression of high power laser pulses in plasma by Raman backscattering.....	22
<i>J. W. Fleischer, O. Cohen, G. Bartal, N. K. Efremidis, H. Buljan, T. Schwartz, T. Carmon, O. Manela, D. N. Christodoulides, M. Segev</i> . Solitons in nonlinear periodic systems.....	34
<i>V. L. Ginzburg</i> . On physics and astrophysics at the beginning of the 21 st century.....	47
<i>E. A. Kuznetsov</i> . Breaking in incompressible fluids and its consequences in turbulence.....	50
<i>A. G. Litvak</i> . Progress in high-power microwaves.....	67
<i>M. Bache, L. A. Lugiato, A. Gatti, E. Brambilla</i> . Classical and entangled ghost imaging schemes.....	80
<i>A. Sergeev, M. Ryabikin, G. Gerber</i> . Engineering of electron wave packets for attosecond pulse production in atoms and molecules.....	93
<i>V. E. Zakharov</i> . Theoretical interpretation of fetch limited wind-driven sea observations.....	107
General Problems of Nonlinear Physics, with Applications to Fluid Dynamics and Environmental Research	
Invited Talks	
<i>I. R. Akylas, A. Tabaei</i> . Resonant self-acceleration and instability of nonlinear internal gravity wavetrains.....	129
<i>S. Goldin</i> . Geodynamics of seismic source.....	136
<i>L. Friedland</i> . Autoresonant excitation and control of nonlinear waves.....	147
<i>L. A. Ostrovsky</i> . Strongly nonlinear internal solitary waves.....	159

<i>E. Pelinovsky, A. Slunyaev, T. Talipova, A. Sergeeva, C. Kharif</i> . Freak waves: Physical mechanisms and experimental data.....	169
<i>V. Shrira, V. Voronovich, I. Sazonov</i> . On quasi-stationary «large amplitude» internal wave solitons.....	179
Contributed Talks	
<i>S. I. Badulin, A. N. Pushkarev, D. Resio, V. E. Zakharov</i> . Self-similar solutions of the Hasselmann equation and experimental scaling of wind-wave spectra.....	191
<i>F. V. Dolzhanskiy, V. M. Ponomarev, I. G. Yakushkin</i> . Quasi-two-dimensional flows and passive scalar transport.....	197
<i>A. I. Dyachenko</i> . On the well-posedness of the free surface hydrodynamics.....	204
<i>O. A. Godin</i> . Wind over fast waves and feasibility of early tsunami detection from space.....	210
<i>V. I. Klyatskin</i> . Coherent phenomena in stochastic dynamical systems.....	216
<i>A. O. Korotkevich, A. I. Dyachenko, V. E. Zakharov</i> . Weak turbulent Kolmogorov spectrum for surface gravity waves: Numerical simulation.....	221
<i>V. Kuvshinov, E. Maslov</i> . On the parametric instability of coherent states of a real nonlinear scalar field.....	226
<i>A. G. Litvak, V. A. Mironov, N. A. Zharova</i> . Self-focusing of the envelope shock waves in resonance media.....	229
<i>W. Pauls, T. Matsumoto</i> . Lagrangian singularities of steady two-dimensional flow.....	236
<i>A. B. Ezersky, V. P. Reutov, G. V. Rybushkina, V. V. Chernov</i> . Simulation of the convective structures in a thin liquid layer in the presence of a wind-driven flow.....	243
<i>N. Romanova, I. Yakushkin</i> . Semi-Lagrangian approach to the investigation of waves in stratified fluid flows.....	248
<i>G. N. Balandina, D. A. Sergeev, Yu. I. Troitskaya</i> . Studyng of the turbulent wake behind bodies of different form in stratified fluid.....	253
<i>A. Shagalov</i> . Emergence and control of vortices in ideal fluids by nonlinear phase synchronization.....	255
<i>S. V. Shagalov</i> . Nonlinear excitation of Rossby waves in a subcritical zonal flow as an analogue of the inversionless amplification in plasma.....	258
<i>A. Celani, S. Musacchio, D. Vincenzi</i> . Single polymer dynamics in random flow.....	264
<i>N. Zubarev</i> . Formation of singularities on the free surface of conducting fluids in an electric field.....	269
<i>O. Zubareva, N. Zubarev</i> . Equilibrium configurations of the charged surface of a conducting liquid in the bounded geometry.....	273

High Power Lasers and Laser-Matter Interactions

Invited Talks

P. Sauer, Y. Dou, B. Torralva, R. E. Allen, SERID: A technique for simulations of complex chemical systems responding to intense fields on short time scales	279
A. Fradkov, M. Ananyevskiy, A. Efimov. Cybernetical physics and control of molecular systems	290
E. S. Fry, X. Z. Zhao, D. Haubrich, J. Musser. Measurement of scattering by micron-size scatterers at angles near zero degrees using two-wave mixing in a barium titanate crystal	299
D. J. Kaup. Resonances in high density nonneutral plasmas	308
K. L. Dvorkin, V. N. Ginzburg, G. I. Freidman, E. V. Katin, E. A. Khazanov, A. V. Kirsanov, A. N. Mal'shakov, V. V. Lozhkarev, G. A. Luchinin, M. A. Martynov, O. V. Palashov, A. K. Poteomkin, A. M. Sergeev, A. A. Shaykin, I. V. Yakovlev. Principles and state of the art in creation of multi-petawatt optical parametric amplifiers	321
J. P. Maranges, S. Gundry, E. Sali, M. P. Ascombe, A. M. Abdulla, S. D. Hogan, P. Kinsler, J. W. G. Tisch, G. H. C. New. Modulation of ultrashort pulses in transiently and adiabatically prepared Raman media	330
C. R. Munterlyn, A. L. Munterlyn, M. E. Arnoldussen, B. A. Logan. Nonlinear ablation in excimer laser vision correction	340
V. Ramanathan, S. Kazamias, F. Weihe, G. Mullot, D. Douillet, F. Augé, O. Albert, J. P. Chambaret, D. Hulin, P. Balcou, D. Reitzé. Enhancement and tuning of high order harmonic generation through adaptive control	351
W. Cheng, Y. Avitzour, Y. Ping, N. J. Fisch, M. S. Hur, J. S. Wurtele, S. Suckewer. Raman amplification of ultrashort laser pulses in the nonlinear regime	360
A. E. Charman, G. Penn, A. Wolski, J. S. Wurtele. Beam conditioning and harmonic generation in free electron lasers	369
Contributed Talks	
O. L. Antipov, O. N. Eremeykin, A. V. Ievlev, A. P. Savikin. Diode-pumped solid-state lasers with reciprocal dynamic holographic cavity	381
A. A. Balakin, G. M. Fraiman, N. J. Fisch. Backward Raman amplification in a partially ionized gas	386
I. V. Izmailov, A. V. Lyachin. Simulation of complex dynamics in the ring resonator with account of multi-photons processes	393
I. Kostyukov, A. Pukhov. Cavity formation in strongly nonlinear regime of laser-plasma interaction	399
V. P. Krainov, V. S. Rastunkov. Stochastic heating of electrons at the relativistic laser-solid interaction	404

V. P. Krainov, A. V. Sofronov. Relativistic electron drift in underdense plasma produced by a super-intense femtosecond laser pulse	409
A. B. Savel'ev, V. M. Gordienko, I. M. Lachko, B. V. Mar'in, D. S. Uryupina, R. V. Volkov. Experimental investigation of ion acceleration from femtosecond laser plasma interaction: Role of surface impurities	414

Gamma-ray Optics and Quantum Nucleonics

Invited Talks

F. J. Agee, A. Y. Matsuura. Review of current challenges and opportunities in quantum nucleonics	421
D. Cline, A. B. Hayes, C. Y. Wu. High-K isomer bands in ^{178}Hf	426
G. R. Hoy. Speculations about a proposed experiment to observe stimulated emission of gamma-radiation	439
S. A. Karamian, J. J. Carroll, J. Adam, N. A. Demekhina. Production of the $^{178\text{m}2}\text{Hf}$ isomer using the 4.5-GeV electron synchrotron and other accelerators	448
L. A. Rivlin. Quantum nucleonics: Outlines and outlook	458
V. I. Vyatskii, A. A. Kornilova, Yu. D. Perfiliev, L. A. Kulikov, A. A. Opatenko. Controlled spontaneous conversion decay of excited nuclei	461
Contributed Talks	
V. M. Cherepanov. Mossbauer spectroscopy of electron-nuclear transitions	471
V. I. Kirischuk, N. V. Strilchuk, V. A. Zheltonozhsky. Accelerated decay of long-lived nuclear isomers and nuclides induced by low-energy photons and electrons	477
E. K. Sadykov, F. G. Vagizov, G. I. Petrov, A. V. Pyataev. The interaction of Mossbauer photons with nuclei showing nonlinear spins dynamics	482
A. B. Savel'ev, O. V. Chutko. Excitation of low-lying nuclear states in two-temperature femtosecond laser produced plasmas	487
R. N. Shakhmuratov, J. Odeurs, P. Mandel. Causality and the color of a single gamma photon	493
E. V. Tkalya. Induced decay of $^{178}\text{Hf}^{\text{m}2}$: Cross section upper bound and comparison with experimental data	499
F. G. Vagizov, R. L. Kolesov, O. A. Kocharovskaya. Laser-induced transformations of ^{57}Fe : Mossbauer spectrum	505
S. V. Adamenko, V. I. Vyatskii. Laboratory collapse of electron-nuclear plasma and full-range nucleosynthesis of stable usual and superheavy nuclei	511
A. A. Zadernovsky. Selection of nuclides for an active medium of gamma-ray laser	517

<i>V. I. Kirischuk, N. V. Strilchuk, V. A. Zheltontozhsky. Isomeric decay through electronic bridges.....</i>	529
---	-----

Nonlinear and Coherent Phenomena in Quantum Systems

Invited Talks

<i>L. Davidovich, R. L. de Matos Filho, F. Toscano. Quantum-classical transition in phase space: Roles of dissipation and diffusion.....</i>	537
<i>L. E. Reichl. Nonlinear dynamics of matter-field systems.....</i>	547

Contributed Talks

<i>R. Akhmedzhanov, L. Gushchin, A. Litvak, Y. Radeonychev, I. Zelensky, R. Kolesov, E. Kuznetsova. Radio-frequency induced transparency in a neon discharge.....</i>	555
<i>R. A. Akhmedzhanov, I. V. Zelensky. Resonance radiation group delay in the rubidium vapor.....</i>	561
<i>V. G. Arkhipkin, I. V. Timofeev. Storage and retrieval of short intense laser pulses.....</i>	567
<i>H. Ishio. Quantum-classical transition for transport through open chaotic cavities.....</i>	573
<i>A. M. Kamchatnov, A. Gammal, R. A. Kraenkel. Formation of dissipationless shocks in Bose-Einstein condensate.....</i>	578
<i>V. Kurin, A. Chiginev. Dynamic phase transitions in lattices of Josephson vortices moving in layered superconductors.....</i>	583
<i>R. Kolesov, E. Kuznetsova, O. Kocharovskaya. Suppression of excited-state absorption in laser crystals.....</i>	588
<i>V. Yu. Argonov, S. V. Prants. Synchronization, riddling, bifurcations, strange attractors, and fractals in nonlinear dynamics of an atom in a standing light wave.....</i>	594
<i>Y. V. Radeonychev, M. Erutimova, O. Kocharovskaya. Lasing without inversion in a doped photonic crystal.....</i>	601

Selected Problems in Astrophysics and Cosmology

Invited Talks

<i>G. S. Bisnovatyi-Kogan. Approximate theory of large-scale structure formation in the Universe.....</i>	609
---	-----

<i>E. Aurell, D. Fanelli, S. N. Gurbatov, A. Yu. Moshkov. The non-linear regime of the gravitational instability. The inner structure of the pancake.....</i>	619
<i>E. V. Derishev, V. V. Kocharovsky, Vl. V. Kocharovsky, F. A. Aharonian, A. A. Belyanin. The highest-energy cosmic rays and the converter acceleration mechanism.....</i>	630

Contributed Talks

<i>E. V. Derishev, V. V. Kocharovsky, Vl. V. Kocharovsky, P. Mészáros. Nonlinear interplay between synchrotron and inverse Compton emission in relativistic astrophysical plasmas.....</i>	641
<i>V. Ju. Martianov, E. V. Derishev, V. V. Kocharovsky, Vl. V. Kocharovsky. Weibel instability in relativistic plasmas and the self-consistent magnetic field in astrophysical jets.....</i>	647
<i>S. N. Gurbatov, A. Yu. Moshkov. Isotropization of multi-dimensional Burgers turbulence.....</i>	654

Selected Problems in Nonlinear Acoustics

Invited Talk

<i>L. A. Crum, M. R. Bailey, R. O. Cleveland, O. A. Sapozhnikov, V. A. Khokhlova. Nonlinear phenomena in modern applications of ultrasound and shock waves in medicine.....</i>	663
---	-----

Contributed Talks

<i>Ph. Blanc-Benon, V. A. Khokhlova, M. V. Averianov, R. O. Cleveland. Propagation of nonlinear acoustic signals through inhomogeneous moving media.....</i>	674
<i>V. A. Khokhlova, M. R. Bailey, J. Reed, P. J. Kaczowski, L. A. Crum. The relative effects of cavitation and nonlinear ultrasound propagation on dynamics of thermal lesion in a tissue phantom.....</i>	679
<i>A. V. Lebedev, L. A. Ostrovsky, A. M. Sutin. Nonlinear acoustic spectroscopy of local defects in solids.....</i>	685
<i>V. Nazarov, A. Radostin. Wave processes in micro-inhomogeneous elastic media with hysteretic nonlinearity and relaxation.....</i>	692

RESONANCES IN HIGH DENSITY NONNEUTRAL PLASMAS

D. J. Kaup

Institute for Simulation and Training & Department of Mathematics,
University of Central Florida, Orlando, FL 32816
Email: kaup@ucf.edu; Tel.: 407-823-2795; Fax: 407-823-6253

Theoretical studies of operation of crossed-field electron vacuum devices, such as magnetrons and crossed-field amplifiers (CFA), have usually centered on their initial growth, taking this as an indication of their operating modes. In such an analysis, one assumes a growth rate, and solves the equations for the density profile, the operating frequency, the growth rate, and other features of these devices. What one obtains here are the conditions for the initial operation of the device. The dominate interaction in this stage is a Rayleigh-like instability which turns into a nonlinear diffusion process, whereby the electron density profile redistributes itself into a profile which will be in an equilibrium with the ponderomotive-like forces produced by the growing RF fields. Eventually the RF fields will saturate and the device will then settle into either a stationary operating regime or some other regime which could be non-operating. A key question is to understand this latter stage of a device's operation, which has been called the «saturation stage». In this saturation stage, wherein one would like the device to simply smoothly deliver RF power, there is a different set of physical interactions occurring. No longer is there a growth rate - rather the amplitudes have saturated, and as a result, the ponderomotive-like forces have also vanished, along with the nonlinear diffusion. Actually there is no further need for a nonlinear diffusion, since the profile has already been reshaped in the initiation stage. To study this saturation stage, it is necessary to return to the classical magnetron equations and re-study the nature of the solutions in a saturation state. What we find is that upon saturation, new RF modes appear. In fact, there are a total of five RF modes in the saturation stage, three of which have very fast oscillations in the vertical direction. One mode corresponds to a plasma oscillation in the vertical direction, which is driven by the vertical current through the device. The other two modes are cyclotron modes. When operating appropriately, the plasma oscillation has a resonance with the potential flow created in the initiation stage, and appears to be the dominant interaction in transferring power from the DC field, to the RF fields. The fast cyclotron modes can interact with the potential flow only under special circumstances, one of which occurs when the CPI T266 operates in an «ultra-low noise» mode. What we will do here is to present the appropriate magnetron equations for the saturation stage, and use the WKB approximation to discuss and describe the nature of these solutions.

1. Introduction

Electronic vacuum devices such as magnetrons and crossed-field amplifiers (CFA) are important generators of high power microwaves. They are also relatively simple in their structure and operation. They consist of a vacuum region, inside of which, a high density electron plasma is created. The theory of their general operating range follows straight-forwardly from electrostatics and magnetostatics [1]. However developing a theory for the RF oscillations has not been so direct. The genesis of the RF theory is the Buneman, Levy, Lindson (BLL) theory of 1966 [2]. Here, by assuming a Brillouin flow, they were able to calculate RF operating characteristics. However, their theory was not successful since it was based on what has been called the magnetron resonances, which occur when the local RF frequency, as seen by the electrons, $\omega_e = \omega - kv_0$, is equal in magnitude to the cyclotron frequency, Ω . (ω is the RF frequency, k is the RF wavevector, and v_0 is the $E \times B$ drift velocity.) The diocotron resonance (the wave-particle resonance at $\omega_e = 0$) did not appear in their theory. The reason for this was found later [3], when it was discovered that the diocotron resonance did occur «inside» the assumed discontinuity at the edge of the Brillouin flow. This became apparent when the assumed discontinuity at the edge of the Brillouin flow was replaced by a steep ramp, inside of which was found the diocotron resonance and its structure [3]. It was found that due to the shear flow, a Raleigh instability would occur inside the steep ramp at the edge of the Brillouin flow, which then excited a nonlinear diffusion process which redistributed the electrons into a new stationary density profile [4], but one which would be in equilibrium with the ponderomotive-like forces of the propagating RF wave.

However, the RF analytical theory has only been developed for the initiation stage, which is where the device is initializing, and the internal RF wave is still growing. In this paper, an extension is presented of the RF theory to the stage that follows the initiation stage, called the «saturation stage». The saturation stage is where the RF amplitudes have grown to their maximum size and have saturated, and this is where the device will steadily deliver RF power. To obtain this extension, we return to the original derivation [5] of the RF theory for the initiation stage. Comparison of the properties of the two stages, it is seen that there are two items that are sufficiently different in these two stages to affect the derivation. The first item is that by the end of the initiation stage, a DC current from the anode to the cathode has been developed, and along with it, an corresponding component to the DC electron $E \times B$ drift velocity in the direction from the cathode to the anode. The second item is that in the saturation stage, since we must have a steady state, there can be no growth rate in the RF

field [6]. In the initiation stage, a nonzero growth rate was 1) necessary for the existence of the nonlinear diffusion, and 2) necessary for the existence of nonsingular solutions of the RF equations. Thus in the saturation stage, the nature of the RF solutions will be significantly different. The purpose of this paper is to present the RF solution in the saturation stage.

In Section 2, we will discuss the geometry used and derive the RF equations for the saturation stage, from the basic cold-fluid, Poisson equations. We will assume planar geometry for simplicity. What we will find here is that due to the nonzero but small value of the component of the electron drift velocity from the cathode to the anode, the saturation RF equations become a fifth-order set of ordinary differential equations (ODEs), whereas in the initiation stage, one only had a second-order set of ODEs. The solution of these ODEs will clearly be nonsingular, in contrast to the initiation RF equations, which would be singular if the growth rate were to vanish. In Section 3 we shall discuss the structure of the solution of the RF saturation equations, and construct approximate solutions. What we find here the well-known situation when a small parameter [7] multiplies the highest derivative(s) in a set of ODEs: one has either a boundary layer or a resonance layer. The solution of this problem then requires a multi-scaling of the boundary layer and/or resonance layer. Here the fifth-order system of ODEs for the RF fields will be studied in the WKB limit. We will see that three of these modes are «fast», and do not occur in the initiation RF equations. The other two modes are «slow» modes, and are the usual two modes that are found in the initiation stage. We will summarize our results in Section 4.

2. Basic equations

The geometry and configuration that we shall be using is shown in Fig. 1, which is a diagram of a «planar magnetron». At the bottom is the cathode, which is an electron emitting material. It is located in the xz -plane and at $y = 0$. We take the ambient magnetic field to be directed along the negative z -axis, so that the electron drift velocity will be mainly in the positive x -direction. (In the saturation stage, this DC drift current will have a small vertical component in the y -direction, as shown.) The electron cyclotron frequency is $\Omega = eB_0/(mc)$ where B_0 is the ambient magnetic field. Next to the cathode is the electron sheath which extends out for some distance from the cathode, as shown. The anode is located at a distance ℓ above the cathode. On the anode is a slow wave structure which is simply a wave guide, along which the RF wave will propagate. This slow wave structure is there to i) slow down the electromagnetic RF wave and ii) to select the effective RF (electrostatic) wave vector, k , that the plasma will see and will respond to.

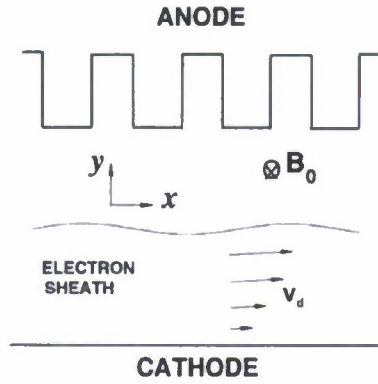


Figure 1. The geometry and configuration of a planar magnetron or CFA. The magnetic field is directed into the paper and the electron drift velocity is to the right. The z -axis is perpendicular to the paper, the x -axis is parallel to the cathode surface and the y -axis is perpendicular to the cathode.

During the initiation stage, the diocotron interaction causes the sheath to readjust its profile, from a pure Brillouin flow to one that is in equilibrium with the growing RF wave. The wavelength of the RF wave will determine the periodicity of the spoke-like structure that evolves from the sheath due to the nonlinear diffusion, which extends from the surface of the sheath out to the anode. A typical plot of this structure is seen in Fig. 2, which is a plot of the electron density vs. x and y . At the back is the Brillouin-like sheath while in the front one can see the spoke of charge which carries the y -component of the DC current to the anode. Note that the structure shown contains only two Fourier components in the x direction: a DC background and a single value of k in e^{ikx} , representing the RF component [8].

The basic cold fluid equations describing this physical system are:

$$\partial_t n + \vec{\nabla} \cdot (\vec{v} n) = 0, \quad (1)$$

$$\partial_t \vec{v} + (\vec{v} \cdot \vec{\nabla}) \vec{v} - \vec{\nabla} \phi + \vec{v} \times \vec{\Omega} = 0, \quad (2)$$

$$\nabla^2 \phi = n. \quad (3)$$

In the above equations, the fluid velocity is in the xy -plane, and all quantities are taken to be independent of the z coordinate. We have also simplified the notation by taking n to be the electron plasma frequency squared ($4\pi e^2 \rho / m$ where $e(m)$ is the electronic charge (mass) and ρ is the electron number density) and ϕ to be (e/m) times the electrostatic potential.

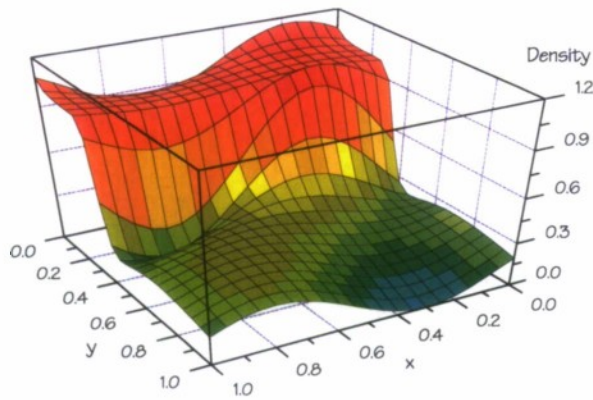


Figure 2. The electron density for one period of a fully developed spoke structure by the initiation stage on the sheath. The anode is in the foreground and the cathode toward the back. Taken from Fig. 5 of Ref [8].

Let us now proceed to expand these equations in the following manner. Let us take

$$n(x, y, t) = n_0 + \left(n_1 e^{i(kx - \omega t)} + c.c. \right), \quad (4)$$

and similar for the other quantities. In the above, the subscript «0» will refer to a DC quantity while the subscript «1» will refer to a RF quantity. Also, k is the wavevector for the RF wave and ω is the RF frequency, both of which are real. Note that we do not include any «second-order DC terms», as has been done for the initiation stage [5]. The reason for this is that by the end on the initiation stage, any such corrections would typically no longer be small, and therefore should be combined with the lower order term as a general DC term. Thus we will only have the DC and the RF Fourier modes. As a consequence, the DC equations will contain second order terms which arise from the averaging of the RF fields, in addition to the usual DC quantities and terms. Also, unlike in the initiation stage, the DC drift velocity will have both an x - and y -component, as:

$$\vec{v} = v_{0x}(y)\hat{x} + v_{0y}(y)\hat{y}, \quad (5)$$

where the y -component will arise from the ponderomotive-like forces in the DC equations.

Using the expansion form in (4), upon expanding (1) - (3), one obtains for the DC components of these equations:

$$\frac{d}{dy}(n_0 v_{0y} + n_1 v_{1y}^* + n_1^* v_{1y}) = 0, \quad (6)$$

$$v_{0y} \left(\frac{dv_{0x}}{dy} - \Omega \right) + \left[v_{1y}^* \frac{dv_{1x}}{dy} + cc \right] = 0, \quad (7)$$

$$v_{0y} \frac{dv_{0y}}{dy} + \Omega v_{0x} - \frac{d\phi_0}{dy} + \left[v_{1y}^* \left(\frac{dv_{1y}}{dy} - ikv_{1x} \right) + cc \right] = 0, \quad (8)$$

$$\frac{d^2 \phi_0}{dy^2} - n_0 = 0. \quad (9)$$

Note the quadratic RF, ponderomotive-like terms present. Meanwhile, from the RF components of the same equation, which are the linear perturbations of (1)-(3), we obtain:

$$\frac{d}{dy}(n_1 v_{0y} + n_0 v_{1y}) + ikn_0 v_{1x} - i\omega_e n_1 = 0, \quad (10)$$

$$v_{1y} \frac{dv_{0x}}{dy} + v_{0y} \frac{dv_{1x}}{dy} - \Omega v_{1y} - i\omega_e v_{1x} - ik\phi_1 = 0, \quad (11)$$

$$\frac{d}{dy}(v_{0y} v_{1y}) - i\omega_e v_{1y} + \Omega v_{1x} - \frac{d\phi_1}{dy} = 0, \quad (12)$$

$$\frac{d^2 \phi_1}{dy^2} - k^2 \phi_1 - n_1 = 0, \quad (13)$$

where

$$\omega_e = \omega - kv_{0x}. \quad (14)$$

Note that when v_{0y} is zero, Eqs. (10) - (13) become the same RF equations as those used for the initiation stage [5]. In this case, the vanishing of v_{0y} causes the derivatives of n_1 , v_{1x} and v_{1y} to vanish from these equations. One thus is left with a second-order set of ODEs and three algebraic relations. However, when v_{0y} is nonzero, one then has all five derivatives of the RF quantities present, and furthermore, Eqs. (10) - (13) can be solved for each derivative of the RF quantities. Doing so, one obtains

$$\begin{aligned} \frac{dn_1}{dy} = \frac{1}{v_{0y}^2} & \left[(n_0 v_{1y} - n_1 v_{0y}) \left(\frac{dv_{0y}}{dy} - i\omega_e \right) - \frac{dn_0}{dy} v_{1y} v_{0y} - \right. \\ & \left. - n_0 \frac{d\phi_1}{dy} + n_0 v_{1x} (\Omega - ikv_{0y}) \right], \end{aligned} \quad (15)$$

$$\frac{dv_{1x}}{dy} = \frac{1}{v_{0y}} \left[v_{1y} \left(\Omega - \frac{dv_{0x}}{dy} \right) + i\omega_e v_{1x} + ik\phi_1 \right], \quad (16)$$

$$\frac{dv_{1y}}{dy} = \frac{1}{v_{0y}} \left[v_{1y} \left(i\omega_e - \frac{dv_{0y}}{dy} \right) - \Omega v_{1x} + \frac{d\phi_1}{dy} \right], \quad (17)$$

$$\frac{d^2\phi_1}{dy^2} = k^2\phi_1 + n_1. \quad (18)$$

Clearly, this is a fifth-order set of nonsingular, linear ODEs for the RF quantities.

These latter equations differ from the RF equations of the initiation stage in two aspects. First, in the initiation stage, $v_{0y} = 0$ which gave us only a second-order set of ODEs. Second, in the initiation stage, we have a nonzero growth rate, in which case, ω_e has a nonzero constant imaginary part. Consequently, one can solve (10) for n_1 , and the combination of these two aspects allows one to obtain nonsingular solutions to the RF equations. In the saturation stage, the growth rate must vanish, and so ω_e must then be real, vanishing identically at the diocotron resonance, which would give a singularity in the RF initiation ODEs. However if we take v_{0y} to be nonzero, this is of no consequence since we then have a fifth-order set of nonsingular ODEs, even at the diocotron resonance. We note that in general, from Eq. (6), which is the same as the conservation of the DC current in the vertical direction, once this current is initiated in the initiation stage and is present and nonzero, provided the fraction of the RF contribution is not too large, we would expect v_{0y} to always be nonzero.

As an example case, we show in Figs. 3 - 5 a typical theoretical solution for the CPI (Communications and Power Industries, Inc., formally Varian Beverly Microwave Division) Tube #T266, which is a crossed-field amplifier (CFA) operating in the GHz range. These figures show the expected initiation solution. One notes in Fig. 1 that the background density consists of a strong Brillouin-like hub with an extended foot, on which the RF density variations oscillate. Fig. 4 shows the parallel drift velocity while Fig. 5 shows the vertical drift velocity. One notes that the vertical drift velocity is rather small, typical about one-thousandth of the parallel drift velocity. (The spike in the vertical drift velocity is due to the cyclotron resonance ($\omega_e = \Omega$ where Ω is the electron cyclotron frequency) just above the cathode. This resonance only occurs when Ω is smaller than the operating frequency, ω .) Except where the cyclotron resonance occurs, this vertical drift current is typically almost constant. Here the important point is that v_{0y} is significantly smaller than the other parameters.

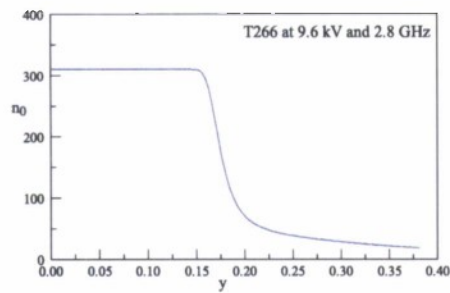


Figure 3. The theoretical background density for the T266 from the initiation equations. The units of n_0 are $(\text{rad.}/\text{ns})^2$ and of y are $\text{m}/100$. The cathode is to the left and the anode is located at $y = 0.381$.

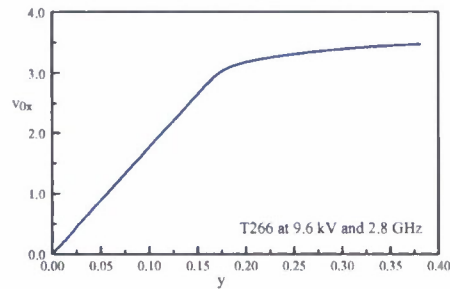


Figure 4. The theoretical parallel drift velocity for the T266 from the initiation equations. The units for v_{0x} are $\text{m}/(100 \text{ ns})$.

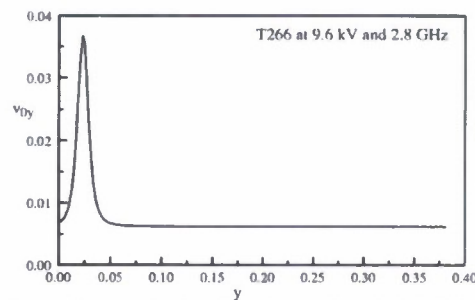


Figure 5. The theoretical vertical drift velocity for the T266 from the initiation equations. The units for v_{0y} are $\text{m}/(100 \text{ ns})$.

Due to the smallness of v_{2y} , one finds it to be essentially impossible to directly numerically integrate the RF equations of the saturation stage. One can see that since three of the derivatives in Eqs. (15) - (18) are inversely proportional to at least v_{0y} , it follows that in general large derivatives will occur, as can be seen in Figs. 6 and 7, which shows a typical example of attempting to integrate these equations directly. As one can see, the density oscillations dominate while the velocity oscillations are an order smaller. (The potential oscillations are another order smaller still.)

Considering that these fast oscillations are present, it therefore becomes appropriate to use WKB methods to separate out the fast variations from the slow ones. In the next section, we shall use a WKB expansion to isolate the three fast modes from the two slow modes and discuss the nature of the solution represented by each mode.

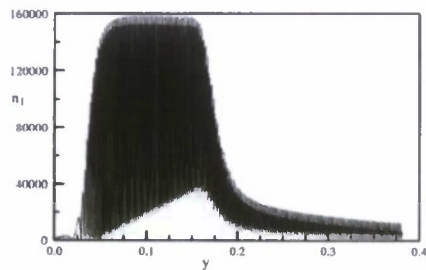


Figure 6. The magnitude of the RF N_1 as a function of y from a direct numerical solution of Eqs. (15)-(18), for the background given in Figs. 3 - 5.

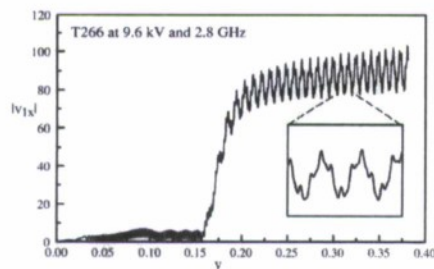


Figure 7. The magnitude of the RF v_{1x} as a function of y from a direct numerical solution of Eqs. (15)-(18), for the background given in Figs. 3 - 5. The insert shows that the solution is actually comprised of more than one fast mode, three in fact.

3. Fast and Slow RF Modes

What we will do here is to start with a given DC background, such as that in Figs. 3 - 5 and proceed to solve the fifth-order set of the RF saturation equations, (10)-(13) by the WKB method, using the smallness of v_{0y} as an expansion parameter. We take v_{0y} to be of order ϵ and assume a fast dependence of the phase of the RF quantities on the coordinate y . Thus we take

$$n_1 = A_n e^{i\theta/\epsilon}, \quad v_{1x} = A_x e^{i\theta/\epsilon}, \quad v_{1y} = A_y e^{i\theta/\epsilon}, \quad \phi = A_\phi e^{i\theta/\epsilon}, \quad (19)$$

where

$$\frac{d\theta}{dy} = \alpha, \quad (20)$$

with α of order unity. Then upon replacing v_{0y} with ϵv_{0y} in Eqs. (10)-(13) and expanding these equations in powers of ϵ , we obtain a set of homogeneous algebraic equations. Requiring the determinant of the coefficients to vanish in leading order of ϵ gives the fifth-order polynomial

$$\alpha^2 (\alpha v_{0y} - \omega_e) [(\alpha v_{0y} - \omega_e)^2 - \lambda^2] = 0, \quad (21)$$

where

$$\lambda = \sqrt{\Omega^2 + n_0 - \Omega \frac{dv_{0x}}{dy}}. \quad (22)$$

The two $\alpha = 0$ modes in (21) are just the two slow modes found in the initiation RF equations. The other three modes are fast modes. Their eigenvalues and unnormalized eigenvectors, $[A_n, A_x, A_y, A_\phi]$, are given by

$$\alpha_1 = (\omega_e + \lambda)/v_{0y};$$

$$[-\alpha_1^2 n_0/k^2, -i\alpha_1(\omega_e - v'_{0x})/k^2, \alpha_1 \lambda/k^2, n_0/k^2], \quad (23)$$

$$\alpha_2 = (\omega_e - \lambda)/v_{0y};$$

$$[-\alpha_2^2 n_0/k^2, -i\alpha_2(\omega_e + v'_{0x})/k^2, -\alpha_2 \lambda/k^2, n_0/k^2], \quad (24)$$

$$\alpha_3 = \omega_e/v_{0y};$$

$$[-\alpha_3^2 \Omega(\Omega - v'_{0x})/k^2, i\alpha_3(\Omega - v'_{0x})/k^2, -i\Omega/k, \Omega(\Omega - v'_{0x})/k^2], \quad (25)$$

where v'_{0x} is the derivative of v_{0x} with respect to y . (In the DC initiation equations, v'_{0x} is equal exactly to n_0/Ω , in which case λ would exactly equal

Ω . However from the DC saturation Eq. (7), we see that in general this need will not be so in the saturation stage.)

The nature of these eigenmodes can be deduced from the eigenvectors and the overall phase. For the first two eigenvectors, since the α 's are inverse proportional to v_{0y} , and as we have already seen in Figs. 6 and 7, the dominate component is the density oscillations, with the velocities being a lower order and consisting of a right- or left-handed elliptical motion. The orientation of the elliptic oscillation depends on the location. In the sheath, since n_0 there is essentially Ω^2 , the motion will mainly be orientated in the y -direction when v'_{0x} is close to n_0/Ω . Outside the sheath, where n_0 is significantly less than Ω^2 , the motion will be approximately circular.

Turning next to the overall phase of these modes,

$$\varphi_j = kx - \omega t + \int^y \alpha_j dy, \quad (26)$$

we see that if we follow the motion of the guiding center of a fluid particle, over a small time interval of time, Δt , the change in the phase is given by

$$\begin{aligned} \Delta\varphi_j &= k\Delta x - \omega\Delta t + \alpha_j\Delta y = \\ &= (kv_{0x} - \omega + \alpha_j v_{0y})\Delta t = \\ &= \pm \lambda\Delta t, \end{aligned} \quad (27)$$

for the first two eigenmodes. Thus following the guiding center, any small motion of the fluid particle about the guiding center is seen to be oscillating at a modified cyclotron frequency, λ . These eigenmodes are therefore the fluid analogy of the well-known «Slater orbits» [1]. On the other hand, if we stay fixed in x , the Eulerian frame, we see rapid oscillations in y , as we have already seen in Figs. 6 and 7. These density oscillations, at a fixed t would be almost parallel to the x -direction, except near a resonance, where α would be zero.

For the third eigenmode, we again have that the dominate component of the eigenvector is the density oscillations, with the motion mainly in the parallel direction, and with the two velocity components in phase. Thus there is no cyclotron-like motion in this mode. Rather, the motion is a back-and-forth motion. If we follow the motion of the guiding center of a fluid particle for this mode, we find that (27) has a zero change along the motion of the guiding center. Thus the fluid particles have a zero oscillation frequency relative to the guiding center. Thus they are drifting. Below the diocotron resonance ($\omega_e > 0$), if the fluid particle moves in the positive x -direction relative to the guiding center, it also moves slightly downward in the y -direction, and the opposite occurs if it is above the diocotron resonance. In the Eulerian frame with x fixed, we again would see the strong

density oscillations in y , and at a fixed t , they would also be almost parallel to the x -direction, except near the diocotron resonance.

When the α 's are well separated, the modes are noninteracting. However when they cross, or are equal, they interact, and can transfer energy from one mode to another. For the two zero modes, they continually interact, and satisfy the well known initiation equations [5], except at the diocotron resonance, where these equations become singular when the growth rate vanishes. However, when they do become singular, at the same time these modes will be interacting with the third mode. It is this interaction which allows a nonsingular solution of the zero modes at the diocotron resonance, with the result of energy being transferred from one mode to another. A plot of these α 's vs. y for the background given in Figs. 3 - 5 is given in Fig 8.

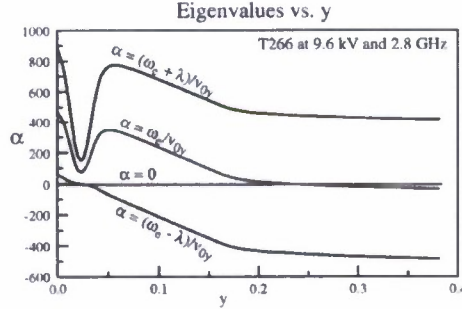


Figure 8. The five modes for the background given in Figs. 3 - 5. Note that the three fast modes never cross and tend to have a fixed separation, except near the resonance between the low cyclotron mode and the slow modes, near the cathode. The main resonance, the diocotron resonance, occurs at the edge of the sheath ($y \approx 0.25$) and corresponds to the crossing of the slow modes ($\alpha = 0$) and the drift mode ($\alpha = \omega_e/v_{0y}$).

Note that the first three modes will never interact, since their α 's could never be equal, unless λ would happen to vanish. This is not expected to ever happen since generally v'_{0x} would be near the value of n_0/Ω . However, any of these three fast modes could cross the zero modes, depending on the parameters, in which case they would interact. For an operational device, the third mode must cross the zero modes; this is the interaction which allows the device to generate RF power. Thus an understanding of the crossing of the third mode with the zero modes is of prime importance. Also of importance is when the second mode crosses the zero modes. This occurs

in the example shown in Figs. 3 - 5, and appears to be responsible for the ultra-low noise operation of this device [9, 10]. No known examples of when the first mode would cross the zero modes are known, nor is it known if this interaction would be of value. However, if it did occur, it would have to occur between the diocotron resonance and the anode.

4. Conclusions

We have shown that the full problem of the theory of the magnetron is basically a fifth-order system, which consists of two slow modes, two fast cyclotron modes (Slater orbits), and one fast drift mode (the diocotron mode). We have also detailed the nature of the WKB solutions of these modes, and have indicated what resonances between these modes would be important. Of course, key questions remain concerning the appropriate boundary conditions for these modes. Future work will be concerned with these, as well as with further details of the nature of these modes, and the calculation of the conversion rates at the resonances.

Acknowledgement

This research was supported in part by AFOSR.

References

1. Slater J.C., *Microwave Electronics*, Van Nostrand, Princeton, NJ, 1959, Chap. 13.
2. Buneman O., Levy R.H., Linson L.M., Stability of crossed-field electron beams, *J. Appl. Phys.*, **37**, 3203-3222, 1966.
3. Kaup D.J., Thomas G.E., Creation of a resonant mode, *Phys. Plasmas*, **3**, 771-780, 1996.
4. Kaup D.J., Thomas G.E., Stationary operating density profiles in a crossed-field amplifier, *J. Plasma Phys.*, **59**, 259-276, 1998.
5. Kaup D.J., Thomas G.E., The third-order singular perturbation expansion of the planar cold-fluid magnetron equation, *Stud. Appl. Math.*, **81**, 57-78, 1989; **83**, 271-272 (E), 1990.
6. Kaup D.J., Initiation and Stationary Operating States in a Crossed-Field Vacuum Electron Device, *Proc. SPIE Aerosense meeting*, **4720**, 67-74, 2002.
7. Bender C.M., Orszag S.A., *Advanced Mathematical Methods for Scientist and Engineers*, McGraw-Hill, N.Y., 1978.
8. Kaup D.J., Theoretical modeling of crossed-field electron vacuum devices, *Phys. Plasmas*, **8**, 2473-2480, 2001.
9. Kaup D.J., Thomas G.E., Chaotic instabilities and density profiles in a crossed-field vacuum device, *Proc. SPIE*, **4031**, 54-64, 2000.
10. Kaup D.J., El-Reedy J.O., Thomas G.E., Parametric Interactions Inside a Magnetron, *J. Plasma Phys.*, **64**, 489-506, 2000.

Evolution of the Scattering Coefficients of the Camassa–Holm Equation, for General Initial Data

By *D. J. Kaup*

We consider the Camassa–Holm equation for general initial data, particularly when the potential in the scattering problem of the Lax pair, $m + \kappa$, becomes negative over a finite region. We show that the direct scattering problem of the eigenvalue problem of the Lax pair for this equation may be solved by dividing the spatial infinite interval into a union of separate intervals. Inside each of these intervals, the initial potential is uniformly either positive or negative. Due to this, one can define Jost functions inside each interval, each of which will have a uniform asymptotic form. We then demonstrate that one can obtain the t -evolution of the scattering coefficients of the scattering matrix of each interval. In the process, we also demonstrate that the evolution of the zeros of $m + \kappa$ can be given entirely in terms of limits of the scattering coefficients at singular points.

1. Introduction

The Camassa–Holm (CH) equation [1]

$$m_t + um_x + 2u_x(m + \kappa) = 0, \quad t > 0, x \in R, \quad (1a)$$

$$m = u - u_{xx}, \quad (1b)$$

Address for correspondence: D. J. Kaup, Department of Mathematics and the Institute for Simulation & Training, University of Central Florida, Orlando, FL 32816-1364, USA; e-mail: kaup@ucf.edu

describes unidirectional propagation of shallow water waves on a flat bottom where $u(x, t)$ is the scaled height of the free surface, $\kappa(>0)$ is the scaled critical shallow water speed, and $m(x, t)$ is the conjugate momentum of u . This equation is integrable, and has the Lax pair [1]

$$v_{xx} - \frac{1}{4}v + \frac{m + \kappa}{2\lambda}v = 0, \quad (2a)$$

$$v_t + uv_x + \lambda v_x + \left(\alpha - \frac{1}{2}u_x\right)v = 0, \quad (2b)$$

where λ is the spectral parameter and α is an arbitrary constant (used to adjust the oscillation frequency of the phase of the Jost functions). The integrability condition for (2) is (1), with λ independent of time.

Among the interesting features of this and related equations is the existence of peaked soliton solutions [1–3] which, for $\kappa = 0$, although the solution is continuous, nevertheless has a discontinuity in the slope at the peak [1]. Due to the interesting nature of these solutions, there have been many studies of the $\kappa = 0$ case [4–8], as well as others, which include applications to geometry [9] and studies of the periodic case [1, 4, 6, 10]. Multisoliton solutions have been given a parametric representation by Matsuno [11]. Interestingly, this equation has also served as a counter-example to the Painlevé test for integrability [12]. More recently, Boyd [13] has studied the transition from a smooth traveling wave to a peaked wave ($\kappa \rightarrow 0$), and A. Parker [14] has shown how to develop a direct method of solution. Water waves are not the only application of this equation, since it has also been found to occur in elasticity [15] and in non-Newtonian fluids [16].

As one can see from (2a), the x -component of this Lax pair has a form where the inverse of the spectral parameter multiplies the potential term, $m + \kappa$, whereas in the Schrödinger equation, the spectral parameter and the potential are additive. This gives rise to a different set of analytical and asymptotic properties for the eigenfunctions of this spectral problem [17]. In particular, whenever the potential has a change in sign, the analytical and asymptotic properties are not uniform in x . The direct and inverse scattering problem for (2a), when the initial data for $m + \kappa$ changes sign, has not been addressed. However, in the case when the initial data for $m + \kappa$ never changes sign, both the direct and inverse scattering problem, on the infinite interval, had been earlier addressed [2], and further described in a series of recent papers [18–20]. In this case, eigenfunctions can always be chosen so that the analytical and asymptotic properties will be uniform in x .

An important feature of the CH equation is that one can transform from an Eulerian coordinate to a Lagrangian coordinate, upon which (2a) transforms into the Schrödinger equation [19]. Then providing $m + \kappa$ remains of one sign, one can proceed as in the solution of the KdV equation [21]. However,

whenever $m + \kappa$ has simple zeros, the direct and inverse scattering problems of (2a) become significantly more complex. In addition to singularities appearing, there is, effectively, a rotation of $\pi/2$ in the analytical and asymptotic properties of the eigenfunction of (2a), with these properties not being uniform in x .

Relevant to the above, it has been pointed out by McKean [22, 23] that whenever the initial data of $m + \kappa$ has any positive region to the left of any negative region, then the solution of (1) will break in a finite time. Thus we can anticipate that when the initial data do contain a negative region, then the solution of the more general initial value problem would contain some new and interesting features. This same class of spectral problems (where the potential multiplies the spectral parameter) has been around for some time, since it also appears in the Lax pair of the Harry–Dym equation [24] and other related equations [2, 25], the oscillating two-stream instability (OTSI) problem [26, 27], as well as in the related degenerate two-photon propagation (DTPP) system [28–30]. The latter two are important physical integrable systems, which with this result, can now be re-attacked.

In this paper, we define a set of scattering coefficients, and also present the time evolution of those scattering coefficients, for the spectral problem (2a), when the potential of that spectral problem changes sign. This therefore covers the general class of initial value problems for the CH equation, as well as the same for the OTSI and the DTPP systems. The general formulation of an Inverse Scattering Transform (IST) for the CH equation, valid for general initial data on the infinite interval, provided only that the initial data for m vanish sufficiently rapid at infinity (the usual Faddeev conditions), will not be addressed in this publication. Rather, here we shall only address the direct scattering problem and the evolution of the scattering coefficients.

Our approach will be as follows. Given the initial data for m , we break the infinite interval into intervals of positive and negative values of the initial data for $m + \kappa$. This is necessary if one is to have uniform analytical and asymptotical properties for the Jost functions, inside any interval. In the absence of such uniform analytical properties, the usual method of solution for the inverse scattering problem appears intractable. Inside each of these intervals, we can solve the direct and inverse scattering problems, independent of the other intervals, and define appropriate scattering data for the initial data inside that interval [29]. The real problem now becomes how to evolve these independent sets of scattering data in time. This problem can indeed be solved. By the use of the second component of the Lax pair, (2b), one can determine the evolution in time of the scattering coefficients of each interval. (The scattering coefficients are the coefficients, a , $\bar{a}$, b , and $\bar{b}$ to be introduced below. These coefficients are not the “scattering data.” But the scattering data are a reduced set which can be obtained from the scattering coefficients.) How to correctly evolve and couple the t -evolutions of the various scattering coefficients will be shown below. Following upon this procedure, one could

then formulate and solve the inverse scattering problem inside each interval at any later time, and each could be solved independent of the other intervals, with an independent reconstruction of m inside each interval. The union of the solutions in each interval, over all intervals, would give the total solution for m on the infinite interval.

Of course, given this result, one could construct the complete scattering matrix for the entire infinite interval, as a product of the individual scattering matrices from each interval. However how useful this could be in constructing an inverse scattering solution for the entire interval remains to be seen. Our approach is to use known techniques to solve the full problem. Then in hindsight, perhaps a general solution for this inverse scattering problem on the full infinite interval could be formulated.

In Section 2, we shall describe a key important feature, which can be gleaned directly from the CH equation. This is the fact that the zeros of $m + \kappa$ will move with the flow. In Section 3, we will show that one can obtain the evolution of the flow, u , at a zero of $m + \kappa$, in terms of two functions of t . This will be important for determining the time evolution of the individual scattering coefficients. In Section 4, we will describe the Jost functions and the solution of the direct scattering problem for all intervals. In Section 5, we will obtain the evolution of the scattering coefficients in each of these intervals. In Section 6, we will obtain the evolution of the zeros of $m + \kappa$, from simply a knowledge of the scattering coefficients, and will also related this evolution directly to m itself. In Section 7, we make some concluding remarks and outline future work.

2. Background

Let us next expand on some key points essential to the development of a general IST for the CH equation. First, for the CH equation, zeros of $m + \kappa$ will move with the flow [19]. This follows from (1a), which can be rewritten as

$$(\partial_t + u\partial_x)(m + \kappa) + 2(m + \kappa)\partial_x u = 0. \quad (3)$$

Note that whenever $m + \kappa = 0$, that point moves with the flow. As a consequence of this, Lagrangian coordinates, (χ, τ) , can be defined [19] where,

$$\partial_x \chi = \sqrt{|m + \kappa|}, \quad \tau = t. \quad (4)$$

These coordinates have the property that the χ value of a zero of $m + \kappa$ remains invariant in τ . Thus in the Lagrangian frame, assuming only simple zeros of $m + \kappa$, one may label the zeros of the initial data of $m + \kappa$ with $\chi_a < \chi_b < \dots < \chi_z$ providing that all zeros are simple (i.e., $\partial_\chi m|_{\chi_j} \neq 0$). (Note that since we take m to vanish at $x = \pm\infty$, then we must have an even number of simple zeros.) Furthermore, the χ -separation between these zeros is maintained, since $\int_{x_a}^{x_b} \sqrt{|m + \kappa|} dx$ is a constant of the motion, where x_a and x_b

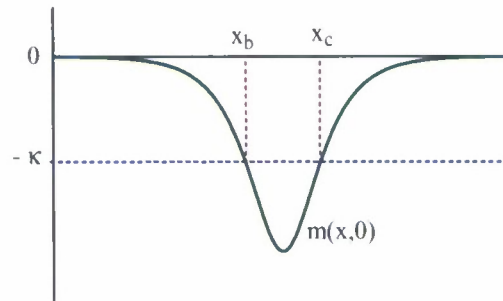


Figure 1. An example of an initial profile for $m(x, 0)$ wherein $m + \kappa$ has two zeros at x_a and x_b .

are the Eulerian values of any two successive zeros. Given the Lagrangian values of the zeros of the initial data for $m + \kappa$, it follows that one may transform the location of these zeros back to the Eulerian frame [19]. Thus the infinite interval in either frame can be broken up into a set of successive intervals, each of a fixed χ -length, and consisting of alternating intervals of positive and negative $m + \kappa$. Inside each of these intervals, we can solve the direct scattering problem for the scattering coefficients for that interval.

We note that since the Lagrangian variable, χ , is simply a coordinate transformation on x , one can actually treat this problem with either coordinate, depending on one's preference. If one uses Lagrangian coordinate, χ , then the zeros will be stationary. However, the signs of the coefficients in the transformed Lax pair, (2), will vary from interval to interval, due to the magnitudes in (4). If one uses the Eulerian coordinate, x , then the zeros will move, but they will never cross. However, the structure of, and the signs in, the Lax pair, (2), will be uniform throughout all intervals. We shall choose to use the Eulerian form. In this case, the zeros of $m + \kappa$ will be designated by $x_a(t) < x_b(t) < \dots < x_z(t)$. Their values at $t = 0$ can be obtained from zeros of the initial data of $m + \kappa$.

To illustrate the above, consider a simple nontrivial example where $m + \kappa$ has only two simple zeros, such as shown in Figure 1, which is a plot of $m(x, 0)$ vs. x . Note that we shall always consider the physical case where, as $x \rightarrow \pm\infty$, $m(x, t)$ vanishes. Here we see that we have three intervals; $-\infty < x \leq x_a$ where $m + \kappa \geq 0$, $x_a \leq x \leq x_b$ where $m + \kappa \leq 0$, and $x_b \leq x < \infty$ where again $m + \kappa \geq 0$.

3. Evolution of the flow at a zero of $m + \kappa$

The evolution of u at a zero of $m + \kappa$ follows from the CH equation. We assume that $u(x, t)$ to be regular at a zero of $m + \kappa$, and therefore expand u in a Taylor series as

$$u(x, t) = \eta_0 + (x - x_0)\eta_1 + \frac{1}{2!}(x - x_0)^2\eta_2 + \dots \quad (5)$$

where $x_0(t)$ is the location of the zero concerned, as a function of t , and the η 's are functions of t also. Taking this, and then by (1b), we have

$$m(x, t) = \eta_{0,2} + \eta_{1,3}(x - x_0) + \frac{1}{2!}\eta_{2,4}(x - x_0)^2 + \dots \quad (6)$$

where we have defined

$$\eta_{\ell, \ell+2} = \eta_{\ell} - \eta_{\ell+2}. \quad (7)$$

Now we determine the evolution of these coefficients. First, since $m(x_0(t), t) + \kappa = 0$ in order to be a zero, it follows from (6) that $\eta_{0,2} + \kappa = 0$, or

$$\eta_2 = \eta_0 + \kappa. \quad (8)$$

The time evolution of the remaining coefficients follow from (1b). Inserting (5) and (6) into (1b) then gives

$$\partial_t x_0 = \eta_0 \quad (9)$$

$$\partial_t \eta_{1,3} + 3\eta_1 \eta_{1,3} = 0 \quad (10a)$$

$$\partial_t \eta_{2,4} + 4\eta_1 \eta_{2,4} + 5\eta_2 \eta_{1,3} = 0 \quad (10b)$$

etc. Clearly, all the coefficients in (5) can be determined in terms of two functions, η_0 and η_1 , and the initial data for u .

4. Direct scattering problem

In this section, we shall set up the direct scattering problem and define the scattering coefficients. There are several manners in which this could be done. On the finite intervals, one could set up the Jost functions to be defined as in a periodic problem. Or one could set them up with independent boundary conditions like $(1, 0)$ and $(0, 1)$. However, what we have found to be best is to set them up exactly as plane waves, like one would do for the infinite interval. To do this, we take the boundary conditions on each Jost function to be exactly those of a plane wave with some wave vector k , when $m = 0$. We will need to relate this wave-vector to the spectra parameter λ , and to do this, we simply consider the limit of $x \rightarrow \pm\infty$ where m vanishes. Then inserting such a plane wave into (2a), upon setting $m = 0$, we have that k and λ will be related by

$$\lambda = \frac{2\kappa}{4k^2 + 1}. \quad (11)$$

Here, we shall treat the example case, such as shown in Figure 1, where $m + \kappa$ has only two zeros at $x = x_b$ and $x = x_c$. Extension of the method to more general cases is obvious. Let us first consider the left interval where $x < x_b$. At the left end, we define the Jost functions ϕ_1 and $\bar{\phi}_1$ by

$$\phi_1(x \rightarrow -\infty) \rightarrow e^{-ikx}, \quad (12)$$

$$\bar{\phi}_1(x \rightarrow -\infty) \rightarrow e^{ikx}, \quad (13)$$

where the subscripts on the Jost functions, and on the scattering coefficients, will refer to those quantities in this interval ($-\infty < x \leq x_b$), which we shall refer to as "Interval #1".

At the right boundary of this interval, at $x = x_b$, we define two other Jost functions, ψ_1 and $\bar{\psi}_1$, such that they appear, at the boundary, to be plane waves. Thus we take

$$\psi_1(x_b) = e^{ikx_b}, \quad \partial_x \psi_1(x_b) = ike^{ikx_b}, \quad (14)$$

$$\bar{\psi}_1(x_b) = e^{-ikx_b}, \quad \partial_x \bar{\psi}_1(x_b) = -ike^{-ikx_b}. \quad (15)$$

With these definitions, using the Green's function for (2a), standard techniques [18–20, 31] and the Faddeev conditions, one can show that inside Interval #1, ($-\infty < x < x_b$)

1. $\phi_1 e^{ikx}$ and $\psi_1 e^{-ikx}$ are analytic in the upper half complex k -plane,
2. $\bar{\phi}_1 e^{-ikx}$ and $\bar{\psi}_1 e^{ikx}$ are analytic in the lower half complex k -plane.

Furthermore, each set of Jost functions is a linearly independent set on the real k axis (except for the usual singularity at $k = 0$). Thus we may relate the two sets as follows, which defines the "scattering coefficients" on this interval.

$$\phi_1 = a_1(k, t)\bar{\psi}_1 + b_1(k, t)\psi_1, \quad (16)$$

$$\bar{\phi}_1 = \bar{a}_1(k, t)\psi_1 + \bar{b}_1(k, t)\bar{\psi}_1. \quad (17)$$

The inverse of this is

$$\psi_1 = a_1(k, t)\bar{\phi}_1 - \bar{b}_1(k, t)\phi_1, \quad (18)$$

$$\bar{\psi}_1 = \bar{a}_1(k, t)\phi_1 - b_1(k, t)\bar{\phi}_1. \quad (19)$$

From the Wronskian of (2a), we have that, for real k ,

$$a_1\bar{a}_1 - b_1\bar{b}_1 = 1. \quad (20)$$

From the above relations, and the fact that the interval is a left semi-infinite interval, it follows that [31]

1. Both a_1 and b_1 are analytic in the upper-half complex k -plane,
2. Both $\bar{a}_1$ and $\bar{b}_1$ are analytic in the lower-half complex k -plane.

One should note that on the infinite interval, b_1 and $\bar{b}_1$ are not analytic and usually only exist on the real k axis (and with the usual $1/k$ singularity at $k = 0$). However since we have only a semi-infinite interval where a_1 and b_1 can be defined at a finite value of x , and since $\phi_1 e^{ikx}$ and $(\partial_x \phi_1) e^{ikx}$ are both analytic

in the upper-half complex k -plane, then so are these scattering coefficients. Similarly for $\bar{a}_1$ and $\bar{b}_1$ in the lower half plane.

In a similar manner, we can define the scattering coefficients for the other intervals. First, we define our intervals.

$$\begin{aligned}\text{Interval \#1: } & -\infty < x \leq x_b(t) \\ \text{Interval \#2: } & x_b(t) \leq x \leq x_c(t) \\ \text{Interval \#3: } & x_c(t) \leq x < \infty.\end{aligned}\tag{21}$$

Let us now define the remainder of our Jost functions. In general, we will take $\phi(\psi)$ to be the Jost function with boundary conditions on the left (right). We shall continue to use subscripts as needed to label the Jost functions and scattering coefficients according to the above intervals. Two of these intervals are semi-infinite. We have already defined the Jost functions for Interval #1. For Interval #3, the ψ Jost function will be defined by

$$\begin{aligned}\psi_3(x \rightarrow +\infty) & \rightarrow e^{ikx}, \\ \bar{\psi}_3(x \rightarrow +\infty) & \rightarrow e^{-ikx}.\end{aligned}\tag{22}$$

All other boundary conditions will be at a zero of $m + \kappa$. At these boundaries, we take

$$\begin{aligned}\phi_2(x_b) &= e^{-ikx_b}, & \partial_x \phi_2(x_b) &= -ike^{-ikx_b}, \\ \bar{\phi}_2(x_b) &= e^{ikx_b}, & \partial_x \bar{\phi}_2(x_b) &= ike^{ikx_b},\end{aligned}\tag{23}$$

$$\begin{aligned}\psi_2(x_c) &= e^{ikx_c}, & \partial_x \psi_2(x_c) &= ike^{ikx_c}, \\ \bar{\psi}_2(x_c) &= e^{-ikx_c}, & \partial_x \bar{\psi}_2(x_c) &= -ike^{-ikx_c},\end{aligned}\tag{24}$$

$$\begin{aligned}\phi_3(x_c) &= e^{-ikx_c}, & \partial_x \phi_3(x_c) &= -ike^{-ikx_c}, \\ \bar{\phi}_3(x_c) &= e^{ikx_c}, & \partial_x \bar{\phi}_3(x_c) &= ike^{ikx_c}.\end{aligned}\tag{25}$$

Similar to the above statements for Interval #1, we have the following properties for the Jost functions inside Interval #2, ($x_b < x < x_c$)

1. $\phi_2 e^{ikx}$, $\bar{\phi}_2 e^{-ikx}$, $\psi_2 e^{-ikx}$, $\bar{\psi}_2 e^{ikx}$ are all analytic in the entire complex k -plane,

For the Jost functions inside Interval #3, ($x_c < x < +\infty$), we have

1. $\phi_3 e^{ikx}$ and $\psi_3 e^{-ikx}$ are analytic in the upper half complex k -plane,
2. $\bar{\phi}_3 e^{-ikx}$ and $\bar{\psi}_3 e^{ikx}$ are analytic in the lower half complex k -plane.

The scattering coefficients are again defined as in Interval #1. For the middle interval, Interval #2, we take

$$\begin{aligned}\phi_2 &= a_2 \bar{\psi}_2 + b_2 \psi_2 \\ \bar{\phi}_2 &= \bar{a}_2 \psi_2 + \bar{b}_2 \bar{\psi}_2,\end{aligned}\tag{26}$$

where

- $a_2, \bar{a}_2, b_2, \bar{b}_2$ are entire functions of k .

From the Wronskian,

$$\bar{a}_2 a_2 - \bar{b}_2 b_2 = 1, \quad (27)$$

we obtain the inverse relations

$$\begin{aligned} \psi_2 &= a_2 \bar{\phi}_2 - \bar{b}_2 \phi_2, \\ \bar{\psi}_2 &= \bar{a}_2 \phi_2 - b_2 \bar{\phi}_2. \end{aligned} \quad (28)$$

On Interval #3, we take

$$\begin{aligned} \psi_3 &= a_3 \bar{\phi}_3 - \bar{b}_3 \phi_3, \\ \bar{\psi}_3 &= \bar{a}_3 \phi_3 - b_3 \bar{\phi}_3. \end{aligned} \quad (29)$$

where

1. a_3 and $\bar{b}_3$ are analytic in the upper-half complex k -plane,
2. $\bar{a}_3$ and b_3 are analytic in the lower-half complex k -plane.

From the Wronskian,

$$\bar{a}_3 a_3 - \bar{b}_3 b_3 = 1, \quad (30)$$

which gives the inverse of (29) to be

$$\begin{aligned} \phi_3 &= a_3 \bar{\psi}_3 + b_3 \psi_3, \\ \bar{\phi}_3 &= \bar{a}_3 \psi_3 + \bar{b}_3 \bar{\psi}_3. \end{aligned} \quad (31)$$

The above have defined the scattering coefficients inside each interval. From these scattering coefficients, one could define the scattering data [31], and then inside any interval, proceed to solve the appropriate Marchenko equations for either the semi-infinite interval or the finite interval [29]. Thus the potentials could be reconstructed from only the scattering data in that interval, and this reconstruction could be done at any instant of time. However, it is also certainly clear that these scattering coefficients must be coupled in some way, at least in their time evolution. Otherwise the separate intervals would evolve independent of each other, in violation of the known nature of characteristics of PDEs. Thus it is essential to address the time evolution of these scattering coefficients, which we do in the next section.

5. The time evolution of the scattering coefficients

By the use of (2b), one can obtain the time evolution of the scattering coefficients. Starting with Interval #1, as $x \rightarrow -\infty$, in order to satisfy (12) and (13), we must take

$$\alpha_\phi = i\lambda\kappa, \quad \alpha_{\bar{\phi}} = -i\lambda\kappa \quad (32)$$

where α_ϕ is the value of α for ϕ_1 and $\alpha_{\bar{\phi}}$ is the same for $\bar{\phi}_1$.

At the right end, we evaluate (2b) at $x = x_b(t)$, which is where $m + \kappa = 0$, by means of the expansion (5), the values in (14) and (15), and the definitions of the scattering coefficients in (15), to express ϕ in terms of ψ , $\bar{\psi}$ and the scattering coefficients. One obtains the following evolutions for the scattering coefficients a_1 and b_1 .

$$\partial_t a_1 = i\eta_{0b} \frac{\kappa}{2k\lambda} a_1 + \frac{1}{2} \left(\eta_{1b} + \frac{i}{2k} \eta_{0b} \right) b_1 e^{2ikx_b}, \quad (33)$$

$$\partial_t b_1 = \frac{1}{2} \left(\eta_{1b} - \frac{i}{2k} \eta_{0b} \right) a_1 e^{-2ikx_b} + \frac{i}{2k} \left[\lambda - 2\kappa - \frac{\eta_{0b}\kappa}{\lambda} \right] b_1, \quad (34)$$

where η_{0b} and η_{1b} are the values of η_0 and η_1 at the left zero, x_b .

The evolutions for $\bar{a}_1$ and $\bar{b}_1$ are exactly the same as for (33) and (34), except that one replace a_1 and b_1 by $\bar{a}_1$ and $\bar{b}_1$, and i by $-i$. Thus

$$\partial_t \bar{a}_1 = -i\eta_{0b} \frac{\kappa}{2k\lambda} \bar{a}_1 + \frac{1}{2} \left(\eta_{1b} - \frac{i}{2k} \eta_{0b} \right) \bar{b}_1 e^{-2ikx_b}, \quad (35)$$

$$\partial_t \bar{b}_1 = \frac{1}{2} \left(\eta_{1b} + \frac{i}{2k} \eta_{0b} \right) \bar{a}_1 e^{2ikx_b} - \frac{i}{2k} \left[\lambda - 2\kappa - \frac{\eta_{0b}\kappa}{\lambda} \right] \bar{b}_1. \quad (36)$$

In exactly the same manner, one can find the evolutions of the other scattering coefficients. Letting η_{0c} and η_{1c} be the values of η_0 and η_1 at the right zero, x_c , we then find that

$$\partial_t a_2 = i(\eta_{0c} - \eta_{0b}) \frac{\kappa}{2k\lambda} a_2 + \frac{2\eta_{1c}k + i\eta_{0c}}{4k} e^{2ikx_c} b_2 - \frac{2\eta_{1b}k - i\eta_{0b}}{4k} e^{-2ikx_b} \bar{b}_2, \quad (37)$$

$$\begin{aligned} \partial_t b_2 = & \frac{1}{2} \left(\eta_{1c} - \frac{i}{2k} \eta_{0c} \right) a_2 e^{-2ikx_c} - \frac{1}{2} \left(\eta_{1b} - \frac{i}{2k} \eta_{0b} \right) \bar{a}_2 e^{-2ikx_b} \\ & + \frac{i}{2k} \left[\lambda - 2\kappa - \kappa \frac{\eta_{0b} + \eta_{0c}}{\lambda} \right] b_2, \end{aligned} \quad (38)$$

$$\partial_t \bar{a}_2 = -i(\eta_{0c} - \eta_{0b}) \frac{\kappa}{2k\lambda} \bar{a}_2 - \frac{2\eta_{1b}k + i\eta_{0b}}{4k} e^{2ikx_b} b_2 + \frac{2\eta_{1c}k - i\eta_{0c}}{4k} e^{-2ikx_c} \bar{b}_2, \quad (39)$$

$$\begin{aligned} \partial_t \bar{b}_2 = & -\frac{1}{2} \left(\eta_{1b} + \frac{i}{2k} \eta_{0b} \right) a_2 e^{2ikx_b} + \frac{1}{2} \left(\eta_{1c} + \frac{i}{2k} \eta_{0c} \right) \bar{a}_2 e^{2ikx_c} \\ & - \frac{i}{2k} \left[\lambda - 2\kappa - \kappa \frac{\eta_{0b} + \eta_{0c}}{\lambda} \right] \bar{b}_2, \end{aligned} \quad (40)$$

and

$$\partial_t a_3 = -i\eta_{0c} \frac{\kappa}{2k\lambda} a_3 - \frac{1}{2} \left(\eta_{1c} - \frac{i}{2k} \eta_{0c} \right) \bar{b}_3 e^{-2ikx_c}, \quad (41)$$

$$\partial_t b_3 = -\frac{1}{2} \left(\eta_{1c} - \frac{i}{2k} \eta_{0c} \right) \bar{a}_3 e^{-2ikx_c} + \frac{i}{2k} \left[\lambda - 2\kappa - \frac{\eta_{0c}\kappa}{\lambda} \right] b_3, \quad (42)$$

$$\partial_t \bar{a}_3 = i\eta_{0c} \frac{\kappa}{2k\lambda} \bar{a}_3 - \frac{1}{2} \left(\eta_{1c} + \frac{i}{2k} \eta_{0c} \right) b_3 e^{2ikx_c}, \quad (43)$$

$$\partial_t \bar{b}_3 = -\frac{1}{2} \left(\eta_{1c} + \frac{i}{2k} \eta_{0c} \right) a_3 e^{2ikx_c} - \frac{i}{2k} \left[\lambda - 2\kappa - \frac{\eta_{0c}\kappa}{\lambda} \right] \bar{b}_3. \quad (44)$$

In the above, the evolution of the four η 's are still unstated, and how this needs to be stated will be described in the next section.

6. The time evolutions of the η 's

At this point, the only thing remaining before one can determine the evolution of the scattering coefficients are the four functions η_{0b} , η_{1b} , η_{0c} , and η_{1c} . These quantities are nothing more than the values of u and u_x at the zeros of $m + \kappa$. Normally, one could not determine these until after the solution for $u(x, t)$ itself was determined. However, we shall now find that these functions can also be determined by the evolutions of the scattering coefficients as given by (33)–(44), combined with the analytical properties of the scattering coefficients, as detailed earlier. To see this, let us first obtain the structure of the scattering coefficients about the singular points.

Consider (11) when $k = \pm i/2$. At these points in the complex plane, $1/\lambda = 0$. Then according to (2a), the potential term vanishes, and the solution for all Jost functions becomes trivial. One then can ascertain that in the upper-half complex k -plane, at $k = \pm i/2$, we have $a = 1 + \mathcal{O}(\lambda^{-1})$ and $b = \mathcal{O}(\lambda^{-1})$ in all three intervals, provided that the appropriate a and b are known to exist in that half-plane. In particular, let us designate these limits as follows. First for the limit of $k \rightarrow +i/2$, let us take

$$a_1 \rightarrow 1 - \frac{1}{2\lambda} I_{ab}^{(0)} + \mathcal{O}(\lambda^{-2}), \quad b_1 \rightarrow \frac{1}{2\lambda} I_{ab}^{(+1)} + \mathcal{O}(\lambda^{-2}), \quad (45)$$

$$a_2 \rightarrow 1 - \frac{1}{2\lambda} I_{bc}^{(0)} + \mathcal{O}(\lambda^{-2}), \quad b_2 \rightarrow \frac{1}{2\lambda} I_{bc}^{(+1)} + \mathcal{O}(\lambda^{-2}), \quad (46)$$

$$\bar{a}_2 \rightarrow 1 + \frac{1}{2\lambda} I_{bc}^{(0)} + \mathcal{O}(\lambda^{-2}), \quad \bar{b}_2 \rightarrow -\frac{1}{2\lambda} I_{bc}^{(-1)} + \mathcal{O}(\lambda^{-2}), \quad (47)$$

$$a_3 \rightarrow 1 - \frac{1}{2\lambda} I_{cd}^{(0)} + \mathcal{O}(\lambda^{-2}), \quad \bar{b}_3 \rightarrow -\frac{1}{2\lambda} I_{cd}^{(-1)} + \mathcal{O}(\lambda^{-2}), \quad (48)$$

while in the lower half-plane, in the limit of $k \rightarrow -i/2$, we will take

$$\bar{a}_1 \rightarrow 1 - \frac{1}{2\lambda} I_{ab}^{(0)} + \mathcal{O}(\lambda^{-2}), \quad \bar{b}_1 \rightarrow \frac{1}{2\lambda} I_{ab}^{(+1)} + \mathcal{O}(\lambda^{-2}), \quad (49)$$

$$a_2 \rightarrow 1 + \frac{1}{2\lambda} I_{bc}^{(0)} + \mathcal{O}(\lambda^{-2}), \quad b_2 \rightarrow -\frac{1}{2\lambda} I_{bc}^{(-1)} + \mathcal{O}(\lambda^{-2}), \quad (50)$$

$$\bar{a}_2 \rightarrow 1 - \frac{1}{2\lambda} I_{bc}^{(0)} + \mathcal{O}(\lambda^{-2}), \quad \bar{b}_2 \rightarrow \frac{1}{2\lambda} I_{bc}^{(+1)} + \mathcal{O}(\lambda^{-2}), \quad (51)$$

$$\bar{a}_3 \rightarrow 1 - \frac{1}{2\lambda} I_{cd}^{(0)} + \mathcal{O}(\lambda^{-2}), \quad b_3 \rightarrow -\frac{1}{2\lambda} I_{cd}^{(-1)} + \mathcal{O}(\lambda^{-2}), \quad (52)$$

where all the I 's are defined by these limits. Thus these I 's are seen to be obtainable from the scattering data.

On the other hand, we may return to (2a), construct the appropriate Green's function, and obtain a solution for the Jost functions, and thereby the scattering coefficients, about the singular points $k = \pm i/2$, in an asymptotic series of $1/\lambda$. Doing so, one finds that these I 's can also be related to the dynamical variable m by

$$I_{ij}^{(\gamma)} = \int_{x_i}^{x_j} m(x, t) e^{\gamma x} dx, \quad (53)$$

where $\gamma = \pm 1$ or 0 , $x_a = -\infty$, $x_d = +\infty$, and as before, x_b and x_c are the two zeros of $m + \kappa$. So, the I 's may be interpreted either as scattering data, or integrals of m , as in (53).

Now, let us put our attention on Interval #1. There we have that both a_1 and b_1 are analytic in the upper-half complex k -plane. Therefore they must possess a Taylor series expansion about any point in this half plane. (Note that we are assuming that the Faddeev conditions will continue to be satisfied, at least for some finite time during the evolution, and that the solution will not break at $t = 0$.) Inserting such an expansion into (33) and (34), one finds that the evolution of b_1 will in general develop a nonzero value at the singular point $k = i/2$. This is in violation of the above expansion for b_1 and will immediately lead to the instantaneous creation of an essential singularity at the singular point, in violation of the stated analytical properties of b_1 . This can be avoided only if, in the limit of $1/\lambda \rightarrow 0$, the right-hand side of (34) vanishes. Using the above expansions, one can show that this coefficient must be

$$e^{x_b}(\eta_{1b} - \eta_{0b}) + I_{ab}^{(+1)} = 0. \quad (54)$$

If one would carry out the same for the singularity in the lower-half plane, one would find that the exact same condition also arises.

The same can also be done in the other intervals, noticing that in the second interval, all coefficients must be entire, and thus b_2 must not be allowed to

develop essential singularities are either of these points, $k = i/2$ and $k = -i/2$. This then gives the conditions

$$(\eta_{1c} - \eta_{0c})e^{x_c} - (\eta_{1b} - \eta_{0b})e^{x_b} + I_{bc}^{(+1)} = 0, \quad (55)$$

$$(\eta_{1c} + \eta_{0c})e^{-x_c} - (\eta_{1b} + \eta_{0b})e^{-x_b} + I_{bc}^{(-1)} = 0, \quad (56)$$

while in the third interval, we find

$$e^{x_c}(\eta_{1c} + \eta_{0c}) - I_{cd}^{(-1)} = 0. \quad (57)$$

Solving these four conditions for the η 's, we obtain

$$\eta_{0b} = \frac{1}{2}e^{-x_b}I_{ab}^{(+1)} + \frac{1}{2}e^{x_b}I_{bd}^{(-1)}, \quad (58)$$

$$\eta_{1b} = -\frac{1}{2}e^{-x_b}I_{ab}^{(+1)} + \frac{1}{2}e^{x_b}I_{bd}^{(-1)}, \quad (59)$$

$$\eta_{0c} = \frac{1}{2}e^{-x_c}I_{ac}^{(+1)} + \frac{1}{2}e^{x_c}I_{cd}^{(-1)}, \quad (60)$$

$$\eta_{1c} = -\frac{1}{2}e^{-x_c}I_{ac}^{(+1)} + \frac{1}{2}e^{x_c}I_{cd}^{(-1)}. \quad (61)$$

Thus by obtaining the I 's from the limits of the scattering coefficients about the singular points, we have the η 's. Once we have the η 's, we then have the evolution of the scattering coefficients, entirely in terms of the same scattering coefficients and their initial values. Note that by (9), the evolution of the x_j 's are given by (58) and (60), which therefore can also be obtained from the scattering coefficients.

To make this point further, consider solving an arbitrary initial value problem by means of this composite IST. Taking the initial data, we solve (2a) for the initial scattering coefficients. Then we integrate the scattering coefficients forward in time, using (33)–(44), taking care that the η 's in the coefficients of these equations always satisfy (58)–(61), where the I 's are obtained from the limits in (45)–(48) or (49)–(52). Then clearly, one can obtain the scattering coefficients at any later time, unless the solution for u happens to break. We then obtain the evolution of the scattering coefficients.

We note in passing that due to (53), which is obtained from (2a), the I 's, for $\gamma = \pm 1$, are exactly integrable, because $me^{\pm x} = \partial_x[e^{\pm x}(\pm u - u_x)]$, due to (1b). Using this relation and (5), one may readily verify that (54)–(57) are indeed correct, which were obtained instead by requiring the scattering coefficients to retain their appropriate analytical properties as they evolved. This is not surprising, since by requiring the scattering coefficients to retain their analytical properties during their evolution, we are also requiring the potential in (1b) to continue to remain as it should be.

7. Conclusions

Let us now review and summarize the situation. As demonstrated above, inside each interval, one can obtain a set of first-order differential equations for the evolution of the scattering coefficients, as in (33) and (34). These equations will be dependent on the position of the zeros, the x_n 's, and the first two η 's at each zero. One then finds that on each semi-infinite interval, there will be exactly one relation, as in (54) and (57). On every internal finite interval, one will obtain two other such independent relations, as in (33) and (34). A simple counting then shows that one has sufficient data to determine all functions, as well as the evolution of the scattering coefficients.

Another feature of these evolution equations for the scattering coefficients that should be noted is that they are linear in the scattering coefficients, provided one is not in a neighborhood of either of the singular points $k = \pm i/2$. However in a neighborhood of these singular points, the equations actually become nonlinear, since the values of the η 's in the coefficients of these equations are to be given in terms of the limits of the b 's, as given in (45)–(48) or (49)–(52).

Although we do now have the equations for these evolutions, how practical they will be as a means for solving the time evolution of the scattering data, as well the same of the CH equation, remains to be seen. However, the above does demonstrate that in principle, it is workable. The least that one can say, is that when this method of solution is finally and fully detailed, we will obtain a new understanding of the CH equation, as well as a new and deeper understanding of the method of the IST. Perhaps we will also be able to better understand why the CH equation satisfies only the weak Painlevé property [12].

In concluding this section, we should make some comments on some possible complications that will appear in deriving the inverse scattering equations for this general initial value problem of the CH equation. For example, in any neighborhood of the zeros of $m + \kappa$, the asymptotics of the Jost functions are not as clean and simple as for the regular Schrödinger equation, where one has the simple result that $\phi e^{ikx} \rightarrow 1$ uniformly in x as $|k| \rightarrow \infty$ in the upper half k -plane. Rather, considering (2a) when one is in a neighborhood of a zero of $m + \kappa$, for $|k|$ large, one sees that any Jost function will have a boundary layer of the structure of an Airy function. At any x away from such a zero, one does have ϕe^{ikx} approaching a constant limit [17] as $|k| \rightarrow \infty$ for fixed x . However, if one is exactly at such a zero, then one is always inside the boundary layer. Thus the limit is not uniform in x for the closed interval. We note that this may be an important consideration in deriving the Marchenko equations for this system, as well as what has been called the "linear dispersion relations" [31], which is the same as the solution of the associated Riemann–Hilbert problem. The key point here is that the scattering coefficients are defined by the values of the Jost functions exactly at these zeros, which is always inside these boundary layers.

Acknowledgments

The author thanks Daryl Holm for his encouragement to investigate this problem. This research has been supported in part by the National Science Foundation, through Grant No. 0129714, and the Air Force Office of Scientific Research.

References

1. R. CAMASSA and D. HOLM, An integrable shallow water equation with peaked solitons, *Phys. Rev. Letters* 71:1661–1664 (1993).
2. M. WADATI, Y. H. ICHIKAWA, and T. SHIMUIZU, Cusp soliton of a new integrable nonlinear evolution equation, *Prog. Theor. Phys.* 64:1959–1967 (1980).
3. R. S. JOHNSON, On solutions of the Camassa-Holm equation, *Proc. R. Soc. Lond. A* 459:1687–1708 (2003).
4. A. CONSTANTIN and W. A. STRAUSS, Stability of the Camassa-Holm solitons, *J. Nonlinear Sci.* 12:415–422 (2002).
5. H. P. MCKEAN, Breakdown of a shallow water equation, *Asian J. Math.* 2:867–879 (1998).
6. A. CONSTANTIN and J. ESCHER, Global existence and blow-up for a shallow water equation, *Annali Sc. Norm. Sup. Pisa* 26:303–328 (1998).
7. Y. LIU and W. WANG, Global existence of solution to the Camassa-Holm equation, *Nonl. Analysis* 60:945–953 (2005).
8. A. CONSTANTIN, Finite propagation speed for the Camassa-Holm equation, *J. Math. Phys.* 46:023506 (2005).
9. G. MISIOLEK, A shallow water equation as a geodesic flow on the Bott-Virasoro group, *J. Geom. Phys.* 24:203–208 (1998).
10. A. CONSTANTIN and H. P. MCKEAN, A shallow water equation on the circle, *Commun. Pure Appl. Math.* 52:949–982 (1999).
11. Y. MATSUNO, Parametric representation for the multisoliton solution of the Camassm-Holm equation. [To appear in *J. Phys. Soc. Jpn.* 74:(July, 2005).]
12. C. GILSON and A. PICKERING, Factorization and Painlevé analysis of a class of nonlinear third-order partial differential equations, *J. Phys. A* 28:2871–2888 (1995).
13. J. P. BOYD, Near-corner waves of the Cammassa-Holm equation, *Phys. Lett. A* 336:342–348 (2005).
14. A. PARKER, On the Camassa-Holm equation and a direct method of solution I. Bilinear form and solitary waves, *Proc. R. Soc. Lond. A* 460:2929–2957 (2004).
15. H.-H. DAI, Model equations for nonlinear dispersive waves in a compressible Mooney-Rivlin rod, *Acta Mechanica* 127:293–308 (1998).
16. V. BUSUIOC, On second grade fluids with vanishing viscosity, *C. R. Acad. Sci. Paris* 328:1241–1246 (1999).
17. T. SHIMIZU and M. WADATI, A new integrable nonlinear evolution equation, *Prog. Theor. Phys.* 63:808–820 (1980).
18. A. CONSTANTIN, On the scattering problem for the Camassa-Holm equation, *Proc. R. Soc. Lond. A* 457:953–970 (2001).
19. J. LENELLS, The scattering approach for the Camassa-Holm equation, *J. Nonlinear Math. Phys.* 9:389–393 (2002).

20. A. CONSTANTIN and J. LENELL, On the inverse scattering approach to the Camassa-Holm equation, *J. Nonlinear Math. Phys.* 10:252–255 (2003).
21. C. S. GARDNER, J. M. GREENE, M. D. KRUSKAL, and R. M. MIURA, Method for solving the Korteweg-deVries equation, *Phys. Rev. Lett.* 19:1095–1097 (1967).
22. H. P. MCKEAN, Correction to breakdown of a shallow water equation, *Asian J. Math.* 3:3 (1999).
23. H. P. MCKEAN, Breakdown of the Camassa-Holm equation, *Commun. Pure Appl. Math.* LVII:0416–0418 (2004).
24. M. D. KRUSKAL, Nonlinear wave equations, *Lect. Notes Phys.* 38:310–354 (1975), [see pg. 340].
25. M. WADATI, K. KONNO, and Y. H. ICHIKAWA, New integrable nonlinear evolution equations, *J. Phys. Soc. Japan* 47:1698–1700 (1979).
26. D. J. KAUP, The Estabrook-Wahlquist method with examples of applications, *Physica* 1D:391–411 (1980).
27. D. J. KAUP, An exact solution for the oscillating two-stream instability, *Phys. Fluids* 23:1383–1387 (1980).
28. H. STEUDEL and D. J. KAUP, Degenerate two-photon propagation and the oscillating two-stream instability: The general solution for amplitude-modulated pulses, *J. Modern Optics* 43:1851–1866 (1996).
29. H. STEUDEL and D. J. KAUP, The inverse scattering transform on a finite interval, *J. Phys. A: Math. Gen.* 32:6219–6231 (1999).
30. H. STEUDEL and D. J. KAUP, Inverse scattering method applied to degenerate two-photon propagation in the low-excitation limit, *J. Phys. A: Math. Gen.* 33:1445–1457 (2000).
31. M. J. ABLOWITZ, D. J. KAUP, A. C. NEWELL, and H. SEGUR, The inverse scattering transform: Fourier analysis for nonlinear problems, *Stud. Appl. Math.* 53:249 (1974).

UNIVERSITY OF CENTRAL FLORIDA

(Received July 15, 2005)

Drift resonance in high density non-neutral plasmas

D. J. Kaup^{a)}

*Institute for Simulation and Training and Department of Mathematics, University of Central Florida,
Orlando, Florida 32816*

(Received 9 July 2005; accepted 31 March 2006; published online 24 May 2006)

Theoretical studies of the operation of crossed-field electron vacuum devices such as magnetrons and crossed-field amplifiers (CFA) have usually centered on their initial growth, taking this as an indication of their operating modes. In such an analysis one solves the equations for the density profile, the operating frequency, the growth rate, and other features of these devices. What one really obtains then are only the conditions for the device to turn on. The dominant interaction in this stage is a Rayleigh-type instability which initiates a quasilinear diffusion process whereby the electron density profile redistributes itself into a profile which will be in equilibrium with the ponderomotive-like forces produced by the growing rf fields. Eventually the rf fields will saturate and an operating device will settle into a stationary operating regime. This stage of a device's operation is called the "saturation stage." This latter stage involves a different set of physical interactions from the initiation stage. No longer is there a growth rate; rather the rf amplitudes have saturated and as a result, the ponderomotive-like forces have also vanished along with the quasilinear diffusion. In this saturation stage, we find that new rf modes appear. In fact, there are a total of five rf modes, two of which are the usual slow modes of the initiation stage, and three of which have fast oscillations in the vertical direction. One fast mode corresponds to a drift plasma oscillation while the other two fast modes are drift cyclotron modes. In this paper, we will describe how the drift plasma oscillation interacts and couples with the slow rf modes at the diocotron resonance. © 2006 American Institute of Physics. [DOI: 10.1063/1.2199227]

I. INTRODUCTION

Crossed-field devices such as magnetrons and crossed-field amplifiers (CFA) are important generators of high power microwaves. They are also relatively simple in their structure and operation. They consist of a vacuum region inside of which a high density electron plasma is created. The theory of their general operating range follows straightforwardly from electrostatics and magnetostatics.¹ However developing a theory for the rf oscillations has not been so direct. The genesis of the rf theory is the Buneman, Levy, Linson theory of 1966.² This theory is based on a linear instability which at best, for the operating mode, lies deep inside the electron sheath and does not correspond to the parameter regime used by the design engineers. The predicted instability is indeed there,^{3,4} however it seems to have nothing to do with the initiation of power generation in a CFA or magnetron. The dominant effect of the instability appears to be to simply diffuse the edge of the sheath with no growth in the generation of power. We now know that this instability is related to the "magnetron resonances," which occur when the local rf frequency, as seen by the electrons, $\omega_r (= \omega - kv_0)$, is equal in magnitude to the cyclotron frequency, Ω . (Here ω is the rf frequency, k is the rf wave vector, and v_0 is the $E \times B$ drift velocity.)

It is also now well known that a linear theory is insufficient to describe magnetrons and CFAs, and it is necessary to include ponderomotive effects in order to understand an op-

erating device. Davidson was the first to point out the importance of the diocotron resonance, which occurs whenever $\omega_r = 0$, in the operation of this device. He showed that this resonance leads to the reshaping of any positive density gradients into negative density gradients.⁵ Also he had early on recognized that the quasilinear theory would be an important component for any understanding of magnetron operation. Following this, other works have also pursued the study of nonlinear interactions in magnetron operation, with the use of multiple scales.⁶⁻⁹

Let us note that typically such devices are initiated by first applying the ambient magnetic field (and for a CFA, also applying the rf input signal), and then secondly ramping up the dc voltage. Under these circumstances one finds that within a couple of cyclotron oscillations a near classical Brillouin sheath will form next to the cathode.^{4,10-13} Meanwhile, as the dc voltage rises, this near-Brillouin sheath slowly expands its thickness, typically taking hundreds of cyclotron oscillations for the dc voltage to ramp up to its full value. As the sheath expands, so does the value of the $E \times B$ drift velocity, v_0 , at the edge of the sheath. If the operating parameters are taken to be inside the device's operating range then at some time shortly before the voltage reaches its full value, the electron drift velocity will equal the phase velocity of some unstable rf mode in the slow-wave structure. At this point the electrons at the edge of the sheath will be strongly accelerated by the unstable rf mode, initiating it as a linear instability on a shear flow¹⁴ described by a Rayleigh-type equation for a shear flow.¹⁵ Once this linear instability initializes, the electrons at the edge of the sheath will undergo

^{a)}Telephone: 407-823-2795. Fax: 407-823-6253. Electronic mail: kaup@ucf.edu

the Davidson "quasilinear" diffusion.^{5,6} The growth rate of this nonlinear instability is found to be directly proportional to the density gradient at the edge of the sheath,^{5,14} which for the near-Brillouin flow is very large. The second-order theory then predicts that the dc electron density profile will change and evolve. As it evolves, the density profile relaxes to a new density profile, one which is in equilibrium with the propagating and growing rf wave in the slow-wave structure and one which also extends all the way up to the anode.¹⁴ This instability of a near-Brillouin sheath is not surprising since earlier kinetic studies in magnetically insulated diodes^{16,17} had revealed that stochastic electron motion occurs as the Brillouin limit is approached. Thus electrons in a Brillouin flow can be easily perturbed and kicked out of the flow and into new states.

Knowing the ordering in the physical processes involved and due to the time scales involved, one can use an adiabatic approximation to construct an analytical qualitative description of the final reshaped density profiles in the presence of a slow-wave structure by means of the WKB approximation.¹⁸ To go beyond the WKB approximation one has to turn to numerical solutions^{5,6} which can be done in a self-consistent manner, producing reshaped density profiles which were in equilibrium with a rf wave propagating in the slow-wave structure.^{19,20} These solutions were also shown to agree quite well with the operating range of the device #T266, a CFA manufactured by CPI (Communications and Power Industries, Inc., formerly Varian Beverly Microwave Division) operating in the GHz range. A review and summary of this entire theory has been given.⁹

The above analysis was for the nonrelativistic cold-fluid equations in the electrostatic approximation. The key approximation in it was to assume only three modes; the background dc fields, only one frequency for the oscillating rf fields, and second-order dc ponderomotive terms generated by the oscillating rf fields. Many of the results of this analysis have since been corroborated by, or have corroborated, results from other approaches such as particle-in-cell (PIC) code results or results from a guiding center (GC) approach. We note that in comparing these other approaches with the cold-fluid approach, one must keep in mind that the resolution afforded by PIC codes is limited due to the relatively large grid size required as well as the relative low number of particles that can be handled. The approximations used in the guiding center approach are most closely satisfied for low density electron plasmas where the plasma frequency is much smaller than the electron cyclotron frequency, Ω .²¹ With regard to our cold-fluid approach, it needs to be noted that these solutions are obtained in the adiabatic approximation, are therefore equilibrium solutions and are valid in the limit of slowly growing rf fields. Also we note that Lau and Chernin²² has pointed out that eigenfunction expansions such as those which have been used in our cold-fluid approach have definite advantages over other methods.

Summarizing some of the results of other approaches, the second-order vertical drift^{6,9} has also been found in the GC approximation by Riyopoulos.²³ This is a very important feature about which we will have more to say later. This drift has also been observed in PIC code calculations⁴ as well.

Riyopoulos has also used the GC approximation to study the initiation of the instability of a classical Brillouin sheath^{7,24} which was then confirmed also in the cold-fluid approach.¹⁴ He has also showed that in order to fit experimental data, it is necessary for the electron density at the diocotron resonance to be taken on the order of one-half of the maximum Brillouin value^{8,25} with the density in the spokes being on the same order of magnitude. This is indeed what happens in the self-consistent cold-fluid theory¹⁸⁻²⁰ as one can also see from figures given below. Along with this, it is to be noted that many PIC code results^{4,26} now show hub and spoke features similar to that obtained by our self-consistent cold-fluid treatment.^{9,18-20} We also point out that the ultralow noise feature of the CPI #T266 (Ref. 20) has also been found in other crossed-field devices.²⁷ We note that all these mechanisms used to create ultralow noise devices seem to have the consequence that the normally (assumed) laminar electron flow just above the cathode has been disturbed somehow into a more turbulent flow. Lastly, we note that except for the experimental knowledge of the operating regimes of magnetrons and CFAs, there has been no other corroborating evidence concerning an explanation as to why the operating voltage regime of these devices is experimentally limited to less than 20% or so above the Hartree value²⁸ and certainly never ever really approaching the higher Hull cutoff voltage.

However, the above theory applies only to what we shall call the "initiation stage" which is where the device is initializing and where the internal rf wave is still growing. The "saturation stage" follows the initiation stage and is the stage where the rf wave will have reached its maximum amplitude, have saturated and the device will be steadily delivering rf power. These two stages differ significantly in two distinctly different features. First, by the end of the initiation stage a second-order dc flow of electrons from the cathode to the anode will have come into existence. The second feature is that when one arrives at the saturation stage, the growth rate of the rf fields will have vanished.²⁹ [Note that consistent with this, in the initiation stage one does see the growth rate smoothly decreasing toward zero as the rf amplitude rises, as shown in Figs. 4(b) and 6(b) of Ref. 20.] These two features have significant physical and mathematical consequences. Physically, the quasilinear diffusion ceases in the saturation stage and the additional dc current between the cathode and the anode can no longer be neglected. Mathematically, one finds three new rf modes appearing and three narrow resonance layers located at the positions of the magnetron and diocotron resonances.³⁰ The latter are sufficiently narrow that a new scaling must be introduced in order to obtain solutions for the rf fields. Furthermore they are so narrow that one would not expect to see them in PIC code results. This is not to say that their effects are not contained in PIC code results, but only that the statistics of PIC codes do not have the resolution to see such narrow resonances.

In this paper, we will study the structure of the rf solutions about the diocotron resonance in the saturation stage which contains the major mechanism for the transfer of dc power into rf power. In an operating device, optimally this resonance will occur somewhere along the edge of the electron sheath where the density gradient is large and negative.

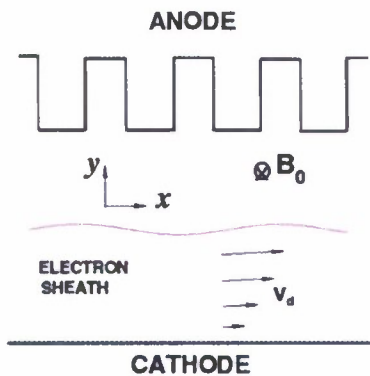


FIG. 1. The geometry and configuration of a planar magnetron or CFA. The magnetic field is directed into the paper and the electron drift velocity is to the right. The z axis is perpendicular to the paper, the x axis is parallel to the cathode surface, and the y axis is perpendicular to the cathode.

Using two multiple scales, we will obtain the inner and the outer solutions around this resonance to first order in an expansion parameter (taken to be the square root of the vertical dc velocity). From these solutions we will obtain the coefficients that asymptotically match the inner and outer solutions, thereby obtaining the mode conversion coefficients.

In Sec. II we will discuss the geometry used and give the nonrelativistic rf equations for the saturation stage in planar geometry and the electrostatic approximation.²⁹ In Sec. III we will quickly treat and dispense with the cyclotron modes since they will be noninteracting at the diocotron resonance. In Sec. IV we will obtain the “outer” solutions which are valid outside of, but near the diocotron resonance. In Sec. V we shall proceed to use a multiple scale expansion to solve for the “inner” solutions inside the diocotron resonance. An expansion will be found for the inner solutions and the solutions will be given in terms of Fresnel-type integrals. In Sec. VI we shall find the asymptotic form of the connection coefficients that connect the inner and outer solutions. We shall then use these coefficients to obtain the conversion coefficients which will give how any one mode, as it passes through the diocotron resonance, converts into the other modes in passing to the other side. In Sec. VII we will make some concluding remarks.

II. BASIC EQUATIONS AND WKB SOLUTIONS

The geometry and configuration that we shall be using is shown in Fig. 1 which is a diagram of a “planar magnetron.” At the bottom is the cathode which is an electron emitting material. It is located in the xz plane and at $y=0$. We take the ambient magnetic field to be directed along the negative z axis so that the electron drift velocity will be mainly in the positive x direction. (In the saturation stage this dc drift current will have a small vertical component in the y direction, as shown.) The electron cyclotron frequency is $\Omega = eB_0/(mc)$, where B_0 is the ambient magnetic field. Next to the cathode is the electron sheath which extends out for some distance from the cathode as shown. The anode is located at a distance ℓ above the cathode. On the anode is a slow-wave structure which is simply a wave guide along which the rf wave will propagate. This slow-wave structure is there to (i)

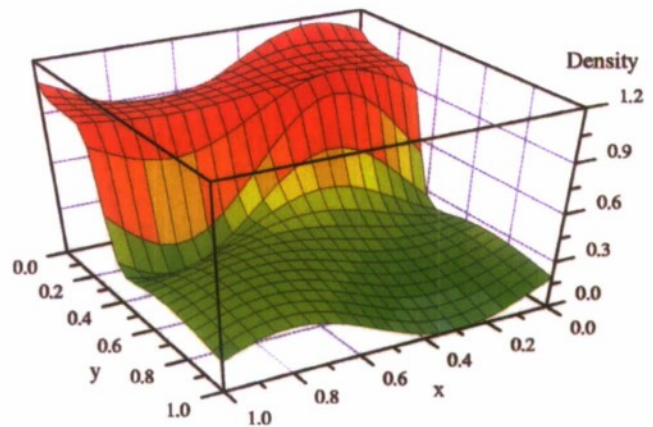


FIG. 2. (Color) The electron density for one period of a fully developed spoke structure in the initiation stage. The anode is in the foreground and the cathode toward the back. Taken from Fig. 5 of Ref. 9.

slow down the electromagnetic rf wave and (ii) to select the effective rf (electrostatic) wave vector, $\mathbf{k}$, the plasma will see and respond to.

During the initiation stage the diocotron interaction causes the sheath to readjust its profile from a pure Brillouin flow to one in equilibrium with the growing rf wave. The wavelength of the rf wave will determine the periodicity of the spoke-like structure which evolves from the sheath due to the nonlinear diffusion and will extend from the surface of the sheath out to the anode. A typical plot of this structure is seen in Fig. 2 which is a plot of the electron density vs x and y . At the back is the Brillouin-type sheath while in the front one can see the spoke of charge which carries the y component of the dc current to the anode. Note that the structure shown contains only two Fourier components in the x direction: a dc background and a single value of k in e^{ikx} representing the rf component.⁹

The basic cold-fluid equations coupled with Poisson's equation that we will use are:

$$\partial_t n + \nabla \cdot (n\mathbf{v}) = 0, \quad (1)$$

$$\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v} - \nabla \phi + \mathbf{v} \times \boldsymbol{\Omega} = 0, \quad (2)$$

$$\nabla^2 \phi = n. \quad (3)$$

In the above equations the fluid velocity is in the xy plane and all quantities are taken to be independent of the z coordinate. We have also simplified the notation by taking n to be the electron plasma frequency squared [$4\pi e^2 \rho/m$ where $e(m)$ is the electronic charge (mass) and ρ is the electron number density] and ϕ to be (e/m) times the electrostatic potential.

For the saturation stage one expands these equations significantly different from that for the initiation stage. Here we take

$$n(x, y, t) = n_0 + (n_1 e^{i(kx - \omega t)} + \text{c.c.}), \quad (4)$$

and similarly for the other quantities. In the above the subscript “0” will refer to a dc quantity while the subscript “1” will refer to a rf quantity. Also, k is the wave vector for the rf wave and ω is the rf frequency, both of which are real. Note

that we do not include any “second-order dc terms,” by which this expansion differs from that used in the initiation stage.⁶ The reason for this change in the form of the expansion for the saturation stage is to incorporate the dc vertical velocity into lowest order. By the end of the initiation stage, second-order corrections would typically no longer be small, and therefore in the saturation stage should be directly combined with the lower order dc terms into a general dc term. As a consequence the dc equations of the saturation stage will contain quadratic rf terms in addition to the usual dc terms of the initiation stage. Thus unlike the initiation stage, the dc drift velocity in the saturation stage will have both an x and y component, as:

$$\mathbf{v}_0 = v_{0x}(y)\hat{x} + v_{0y}(y)\hat{y}. \quad (5)$$

To obtain the equations for the saturation stage we insert expansions for the density, velocity and potential, as in (4), into (1)–(3) and expand. Note that in the saturations stage we will not use a small amplitude expansion; instead we only separate on the basis of Fourier components. First we have the dc components, which are:

$$\frac{d}{dy}(n_0 v_{0y} + n_1 v_{1y}^* + n_1^* v_{1y}) = 0, \quad (6)$$

$$v_{0y} \left(\frac{dv_{0x}}{dy} - \Omega \right) + \left[v_{1y}^* \frac{dv_{1x}}{dy} + \text{c.c.} \right] = 0, \quad (7)$$

$$v_{0y} \frac{dv_{0y}}{dy} + \Omega v_{0x} - \frac{d\phi_0}{dy} + \left[v_{1y}^* \left(\frac{dv_{1y}}{dy} - ikv_{1x} \right) + \text{c.c.} \right] = 0, \quad (8)$$

$$\frac{d^2 \phi_0}{dy^2} - n_0 = 0. \quad (9)$$

Note the quadratic ponderomotive-like rf terms present. Lastly, from the components of the rf fundamental, $e^{i(kx - \omega t)}$, of the same equations [which are also the linear perturbations of (1)–(3)] we obtain:

$$\frac{d}{dy}(n_1 v_{0y} + n_0 v_{1y}) + ikn_0 v_{1x} - i\omega_e n_1 = 0, \quad (10)$$

$$v_{1y} \frac{dv_{0x}}{dy} + v_{0y} \frac{dv_{1x}}{dy} - \Omega v_{1y} - i\omega_e v_{1x} - ik\phi_1 = 0, \quad (11)$$

$$\frac{d}{dy}(v_{0y} v_{1y}) - i\omega_e v_{1y} + \Omega v_{1x} - \frac{d\phi_1}{dy} = 0, \quad (12)$$

$$\frac{d^2 \phi_1}{dy^2} - k^2 \phi_1 - n_1 = 0, \quad (13)$$

where

$$\omega_e = \omega - kv_{0x}. \quad (14)$$

When v_{0y} is zero, Eqs. (10)–(13) become the same rf equations as those used for the initiation stage^{2,6} upon taking ω to be complex. In the initiation stage the vanishing of v_{0y} causes the derivatives of n_1 , v_{1x} , and v_{1y} to vanish from these

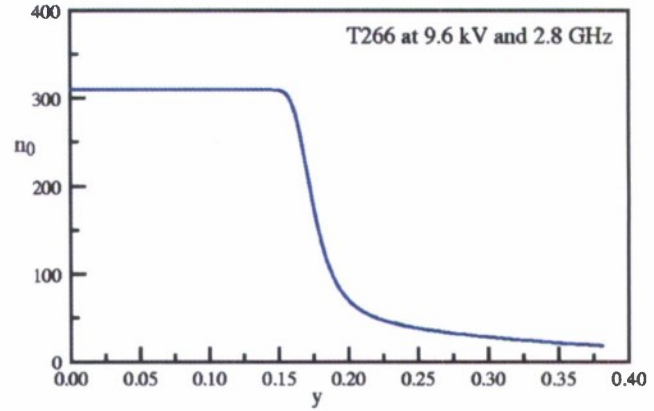


FIG. 3. The theoretical background density for the T266 from the initiation equations. The units of n_0 are $(\text{rad/ns})^2$ and of y are $m/100$. The cathode is to the left and the anode is located at $y=0.381$.

equations. Thus the above set of five ODEs of the saturation stage can be reduced to the second-order set of ODEs and three algebraic relations of the initiation stage.

When v_{0y} is nonzero as in the saturation stage, one then has all five derivatives of the rf quantities present and furthermore, Eqs. (10)–(13) can be solved for each derivative of the rf quantities. Doing so, one obtains

$$\frac{dn_1}{dy} = \frac{1}{v_{0y}^2} \left[(n_0 v_{1y} - n_1 v_{0y}) \left(\frac{dv_{0y}}{dy} - i\omega_e \right) - \frac{dn_0}{dy} v_{1y} v_{0y} - n_0 \frac{d\phi_1}{dy} + n_0 v_{1x} (\Omega - ikv_{0y}) \right], \quad (15)$$

$$\frac{dv_{1x}}{dy} = \frac{1}{v_{0y}} \left[v_{1y} \left(\Omega - \frac{dv_{0x}}{dy} \right) + i\omega_e v_{1x} + ik\phi_1 \right], \quad (16)$$

$$\frac{dv_{1y}}{dy} = \frac{1}{v_{0y}} \left[v_{1y} \left(i\omega_e - \frac{dv_{0y}}{dy} \right) - \Omega v_{1x} + \frac{d\phi_1}{dy} \right], \quad (17)$$

$$\frac{d^2 \phi_1}{dy^2} = k^2 \phi_1 + n_1. \quad (18)$$

This is a fifth-order set of nonsingular, linear ODEs for the rf quantities when v_{0y} is nonzero and all the dc quantities exist. By nonsingular we mean that the derivatives of the rf quantities always exist and are finite. (The equations are linear in the rf variables and in the saturation stage there are no denominator terms which could vanish and give rise to singularities.)

The above rf equations of the saturation stage differ from the rf equations of the initiation stage in two aspects. First, v_{0y} is nonzero in the saturation stage and is zero in the initiation stage and second, the growth-rate is zero (which means that ω is real) in the saturation stage and is nonzero (which means that ω is complex) in the initiation stage.

As an example of the solution in the *initiation* stage, is shown in Figs. 3–5 a typical numerical solution for the CPI device #T266. One notes in Fig. 3 that the background density consists of a strong Brillouin-type hub with an extended foot on which the rf density variations would oscillate. Fig-

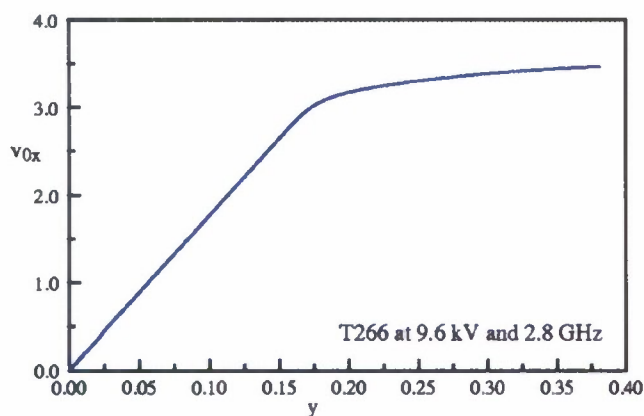


FIG. 4. The theoretical parallel drift velocity for the T266 from the initiation equations. The units for v_{0x} are $m/(100 \text{ ns})$.

ure 4 shows the parallel dc drift velocity while Fig. 5 shows the vertical dc velocity. One notes that the vertical dc velocity is rather small, typical about one-thousandth of the parallel velocity. (The spike in the vertical dc velocity is due to the drift-cyclotron resonance at $\omega_e = \Omega$ located just above the cathode.) Except where the drift-cyclotron resonance occurs, this vertical dc current is essentially constant. Here the important point is that v_{0y} is significantly smaller than the other dc variables. [The reader should note that at the present time there exist no self-consistent solutions of the saturation stage (6)–(13). That will be the object of future work. The above plots are for the initiation stage. However one would expect those of the saturation stage to be similar to these.]

Due to the smallness of v_{0y} , it becomes difficult and time-consuming to numerically integrate the rf equations of the saturation stage. One notes that three of the derivatives in Eqs. (15)–(18) are inversely proportional to v_{0y} or v_{0y}^2 . Thus it follows that in general large values of the derivatives can occur and do occur, as can be seen in Figs. 6 and 7 which shows a typical solution that occurs when one integrates (15)–(18) using the *initiation* stage dc values with sample initial data. As one can see, the density oscillations dominate while the velocity oscillations are an order of magnitude

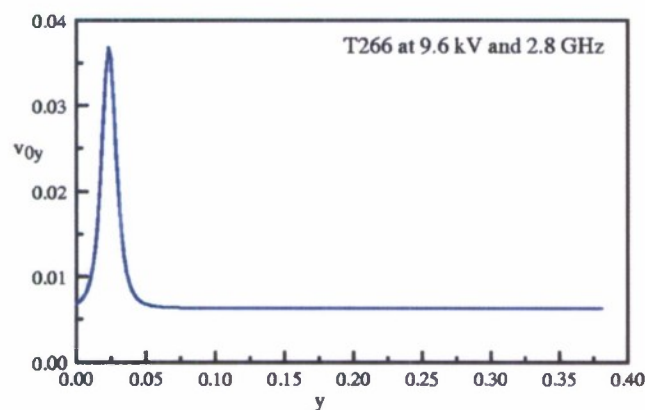


FIG. 5. The theoretical vertical drift velocity for the T266 from the initiation equations. The units for v_{0y} are $m/(100 \text{ ns})$.

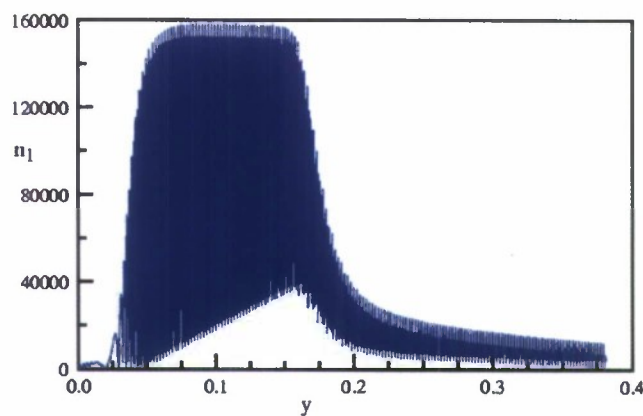


FIG. 6. The magnitude of the rf N_1 as a function of y from a direct numerical solution of Eqs. (15)–(18) for the background given in Figs. 3–5.

smaller. (The oscillations in the potential are even another order smaller.)

In Ref. 30 the WKB solutions of the above equations, (15)–(18), were presented. Assuming fast phase variations of the form

$$n_1 \approx A_n e^{i\int \alpha dy}, \text{ etc.}, \quad (19)$$

where A_n is a slowly varying amplitude and α is a fast WKB wave vector, it was found that α could take on five possible values of which two were zero and corresponded to the usual (slow) modes of the initiation stage. The other three modes were fast modes and their eigenvalues, α , and unnormalized eigenvectors, $[n_1, v_{1x}, v_{1y}, \phi_1]$, in lowest order were found to be

$$\alpha_1 = (\omega_e + \lambda)/v_{0y}; [-\alpha_1^2 n_0/k^2, -i\alpha_1(\Omega - v'_{0x})/k^2, \alpha_1 \lambda/k^2, n_0/k^2], \quad (20)$$

$$\alpha_2 = (\omega_e - \lambda)/v_{0y}; [-\alpha_2^2 n_0/k^2, -i\alpha_2(\Omega - v'_{0x})/k^2, -\alpha_2 \lambda/k^2, n_0/k^2], \quad (21)$$

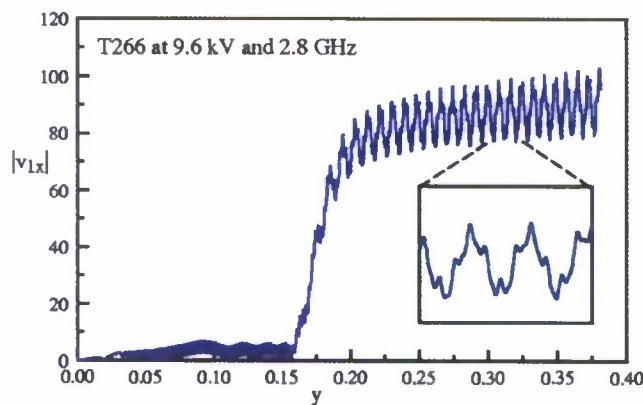


FIG. 7. The magnitude of the rf v_{1x} as a function of y from a direct numerical solution of Eqs. (15)–(18) for the background given in Figs. 3–5. The insert shows that the solution is actually comprised of more than one fast mode, three in fact.

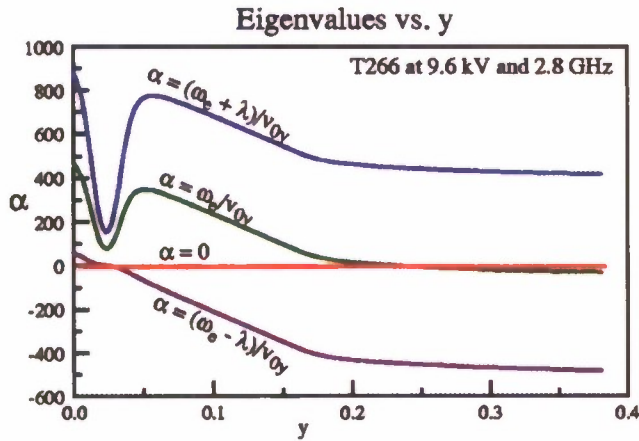


FIG. 8. (Color) The five rf modes for the background given in Figs. 3–5. Note that the three fast modes never cross and tend to have a fixed separation except near the resonance between the lower cyclotron mode and the slow modes which is near the cathode. The main resonance which is the diocotron resonance, occurs at the edge of the sheath ($y \approx 0.25$) and corresponds to the crossing of the slow modes ($\alpha=0$) and the drift mode ($\alpha = \omega_e/v_{0y}$).

$$\alpha_3 = \omega_e/v_{0y}; [-\alpha_3^2 \Omega(\Omega - v'_{0x})/k^2, i\alpha_3(\Omega - v'_{0x})/k^2, -i\Omega/k, \Omega(\Omega - v'_{0x})/k^2], \quad (22)$$

where

$$\lambda = \sqrt{\Omega^2 + n_0 - \Omega v'_{0x}}, \quad (23)$$

and v'_{0x} is the derivative of v_{0x} with respect to y . [In the dc initiation equations v'_{0x} is equal exactly to n_0/Ω in which case λ would be exactly equal to Ω . However from dc saturation Eq. (7), we see that in general this need not necessarily be true in the saturation stage.] A plot of these α s vs y for the initiation stage background is given in Figs. 3–5 is given in Fig. 8. Note that the typical values for α in a fast mode are in the hundreds, which is large, whereas the typical value for the dc quantities is unity. Thus as seen from the eigenvectors in (20)–(22) the density oscillations dominate in the fast modes and the potential oscillations are the weakest.

From the above eigenvalues, eigenvectors and the overall phase, which is

$$\varphi_j = kx - \omega t + \int^y \alpha_j dy, \quad (24)$$

one can determine the nature of these three fast modes. Following a fluid particle one finds that in the first two modes the fluid particle executes circular oscillations at the modified cyclotron frequency, λ , about the guiding center. These eigenmodes are thus the fluid analogy of the well-known “Slater orbits”¹ which are the single-particle cycloidal orbits for electrons in a Brillouin sheath. In the Eulerian frame one sees rapid oscillations in y , as already shown in Figs. 6 and 7. We shall refer to these modes as the “drift-cyclotron modes.”

For the third eigenmode we again have that the dominant component of the eigenvector is the density oscillation with the motion of the fluid particle being mainly in the parallel

direction and with the two velocity components in phase. The motion is thus a drifting back-and-forth motion and will therefore be referred to as the “drift mode.”

For any two modes when the α s are well separated, the modes are noninteracting. However when they cross they interact and energy can transfer from one mode to another. For the two slow modes, they are of course continually interacting and satisfy the well-known initiation equations⁶ except near the cyclotron resonances and the diocotron resonance. At the diocotron resonance, the slow modes cross the drift mode, interact and exchange energy with it.

For an operational device the diocotron mode must cross the slow modes near the Brillouin edge. This is the interaction which allows the device to generate rf power. Thus an understanding of the interactions at the crossing of the drift mode and the slow modes is of prime importance. Also of importance is when the second cyclotron mode crosses the zero modes. This occurs in the example shown in Figs. 3–5 and appears to be responsible for the ultralow noise operation of this device.²⁰ No known examples of when the first cyclotron mode crosses the slow modes are known nor is it known if this interaction would be of interest.

To treat this problem it is necessary to use multiscale methods to separate the fast variations from the slow ones. In the next few sections we shall use a multiscale expansion to obtain the rf solutions for all five modes in the region of the diocotron resonance, which is near $y \approx 0.25$ in Fig. 8.

III. THE DRIFT-CYCLOTRON MODES NEAR THE DIOCOTRON RESONANCE

Let us first handle the two drift-cyclotron modes near the diocotron resonance. We introduce the following notation for our rf eigenvector:

$$V = [n_1, v_{1x}, v_{1y}, \phi'_1, \phi_1]. \quad (25)$$

This eigenvector is to satisfy the fifth-order set of ODEs given by (15)–(18). From the WKB analysis given above we have that the solution for the drift-cyclotron modes near the diocotron resonance will be of the form given in (20) and (21). This is:

$$V_1 = \mathcal{N} \left[n_0 \frac{\omega_e + \lambda}{\lambda}, i(\Omega - v'_{0x}) \frac{v_{0y}}{\lambda}, -v_{0y}, -i \frac{n_0 v_{0y}}{\lambda}, -\frac{n_0 v_{0y}^2}{\lambda(\omega_e + \lambda)} \right] e^{i \int^y \alpha_1 dy}, \quad (26)$$

where the subscript 1 indicates that this is the eigenvector for the first drift-cyclotron mode. This expression also contains an overall amplitude factor which is found from the first-order correction to the WKB approximation. This amplitude factor, suitably normalized to be unitless, for this mode is

$$\mathcal{N} = \sqrt{\frac{\Omega^4}{k^3 v_{0y}^3 |\lambda|}}. \quad (27)$$

The eigenvector for the second drift-cyclotron mode will be the same as for the first drift-cyclotron mode except that λ is to be replaced by $-\lambda$ and α_1 replaced by α_2 .

As is obvious from the above, for $\omega_e \rightarrow 0$, no components are singular and α_1 and α_2 are well separated from the other three modes. Thus the WKB solution for the drift-cyclotron modes will simply continue across the diocotron resonance without any significant mixing or conversion. In the next section we treat the outer solutions for the remaining three modes and following that we will obtain the inner solutions.

IV. THE OUTER SOLUTIONS NEAR THE DIOCOTRON RESONANCE

The outer solutions are the solutions for the eigenvectors when one is approaching the diocotron resonance from outside. From the WKB solution for the drift mode, (22), we have the scalings for the various components in the eigenvector. Thus this mode will be of the form

$$V_{3o} = \left[p_1 \Omega^2, p_2 v_{0y}, p_3 \frac{kv_{0y}^2}{\Omega}, p_4 v_{0y} \Omega, p_5 v_{0y}^2 \right] e^{i \int y \alpha_3 dy}, \quad (28)$$

where the p_j s are unitless functions, of order unity, to be determined. Also the subscript 3o on the eigenvector simply denotes that this is the outer solution for the third eigenvector. Inserting (28) into (15)–(18) and taking the lowest order in v_{0y} for each equation, one readily obtains the solution

$$[p_1, p_2, p_3, p_4, p_5] = \frac{\Omega}{kv_{0y}} \left[1, -i \frac{\Omega}{\omega_e}, i \frac{\Omega^2}{\omega_e^2}, -i \frac{\Omega}{\omega_e}, -\frac{\Omega^2}{\omega_e^2} \right] + O(1), \quad (29)$$

where the prefactor $\frac{\Omega}{kv_{0y}}$ is the amplitude factor that comes from the WKB first-order correction for this mode. Then from (28) and (29) we construct the outer solution, which in lowest order is

$$V_{3o} = \mathcal{A}_{3o} \frac{\Omega}{kv_{0y}} \left[\Omega^2, -i \frac{\Omega v_{0y}}{\omega_e}, i \frac{k \Omega v_{0y}^2}{\omega_e^2}, -i \frac{\Omega^2 v_{0y}}{\omega_e}, -\frac{\Omega^2 v_{0y}^2}{\omega_e^2} \right] e^{i \int y \alpha_3 dy}, \quad (30)$$

where $\mathcal{A}_{3o}$ is a constant which we will specify later.

Next we address the outer solutions for the slow modes. In this case we can simply set $v_{0y}=0$ in (10)–(13) and proceed to solve. The resulting equations are the same as the rf equations for the initiation stage except that (i) there is no growth rate and (ii) we can no longer equate $\partial_y v_{0x}$ to n_0/Ω [see (8)]. Taking this into account we then have:

$$\partial_y v_{1x} = i v_{1y} \left\{ -k - \frac{n'_0}{\Omega \omega_e} + \frac{2k}{A} \left[\frac{(\lambda^2 - n_0)^2}{\Omega^2} + n_0 - \omega_e^2 \right] + \frac{2\lambda' \lambda}{\Omega \omega_e A} \right\}, \quad (31)$$

$$\partial_y v_{1y} = -\frac{A'}{A} v_{1y} - i k v_{1x}, \quad (32)$$

where A is

$$A = \lambda^2 - \omega_e^2. \quad (33)$$

Once solutions are obtained for v_{1x} and v_{1y} , then the other components can be obtained by:

$$n_1 = i v_{1y} \left(\frac{2k n_0}{A} - \frac{n'_0}{\omega_e} + \frac{2\lambda' \lambda n_0}{\omega_e A} \right), \quad (34)$$

$$\phi'_1 = v_{1x} \Omega - i v_{1y} \omega_e, \quad (35)$$

$$\phi_1 = -v_{1x} \frac{\omega_e}{k} + i v_{1y} \frac{v'_{0x} - \Omega}{k}. \quad (36)$$

Since the quantities v_{1x} and v_{1y} continue to interact on the slow scale, it is not possible to separate these two slow modes into separate eigenvectors. Rather we should simply represent them as a single eigenvector with two degrees of freedom. Thus with v_{1x} and v_{1y} satisfying (31) and (32), we take this eigenvector to be

$$[n_1, v_{1x}, v_{1y}, \phi'_1, \phi_1]_{45o} = v_{1x} \left[0, 1, 0, \Omega, -\frac{\omega_e}{k} \right] + v_{1y} \left[i \left(\frac{2k n_0}{A} - \frac{n'_0}{\omega_e} + \frac{2\lambda' \lambda n_0}{\omega_e A} \right), 0, 1, -i \omega_e, i \frac{v'_{0x} - \Omega}{k} \right], \quad (37)$$

where the subscript 45o indicates that this is the eigenvector for the 4,5 slow modes.

The above solutions are accurate to the next order in v_{0y} and are valid solutions providing one is not too near the diocotron resonance. Within a distance of $\mathcal{O}(\sqrt{\frac{v_{0y}}{k\Omega}})$ around the diocotron resonance, these two slow modes will interact with the drift mode. To match these two outer solutions to the inner solutions, we shall require the limit of these solutions as one approaches the diocotron resonance.

To obtain this limit we will scale the variables as follows. We take our scale parameter to be the small unitless quantity, ϵ , defined by

$$\epsilon = \sqrt{v_{0y} k / \Omega}, \quad (38)$$

where we will take the value of v_{0y} (see Fig. 5) to be that at the diocotron resonance. The scaled coordinate, z , will be defined from ω_e by

$$z = -(\omega_e / \Omega) / \epsilon, \quad (39)$$

so that $z=0$ at the resonance. Let ν be the unitless derivative of the parallel dc drift velocity, defined by

$$\nu = v'_{0x} / \Omega, \quad (40)$$

then we have

$$\partial_y = \frac{k \nu}{\epsilon} \partial_z. \quad (41)$$

Now if we take

$$\mu = \lambda / \Omega, \quad (42)$$

then it follows that

$$n_0 = \Omega^2(\mu^2 + \nu - 1), \quad (43)$$

$$n'_0 = k\Omega^2(2\mu_\chi\mu + \nu_\chi), \quad (44)$$

$$\lambda' = k\Omega\mu_\chi, \quad (45)$$

$$v''_{0x} = k\Omega\nu_\chi, \quad (46)$$

and so on. The subscript χ in the above indicates the derivative with respect to ky . This is so that we do not have to convert these dc quantities into functions of z since they are supposedly of order unity functions of ky . Whence we have introduced the variable $\chi = ky$ in order to have a unitless derivative of order unity for these quantities.

Let us comment that ν is essentially the scaled density profile at the diocotron resonance. Since this resonance usually occurs at or near the Brillouin edge, we can expect the derivative of ν to become large but not so large as to become of order $1/\epsilon$. Similarly since μ is the scaled value of λ , which is the "effective" cyclotron frequency inside the resonance, we can expect its derivative to be generally small.

First we scale the two slow modes and the equations that they satisfy, (31) and (32), so that we can obtain approximate solutions for v_{1x} and v_{1y} near the resonance. We do this so that we will have functional forms for the solution, required in the matching process. From (38)–(46), for this scaling, the ODE's (31) and (32) become

$$\partial_z p_3 + \epsilon \left(i \frac{p_2}{\nu} + 2 \frac{p_3 \mu_\chi}{\mu \nu} \right) + O(\epsilon^2) = 0, \quad (47)$$

$$\partial_z p_2 + i p_3 \frac{\beta_0}{z} - i \epsilon p_3 \left(\frac{1}{\nu} - \frac{1-\nu}{\mu^2} \right) + O(\epsilon^2) = 0, \quad (48)$$

where $p_2 = \frac{kv_{1x}}{\Omega}$ and $p_3 = \frac{kv_{1y}}{\Omega}$ are unitless functions of z . Also β_0 is a constant coefficient. It and two other coefficients (to appear later) are defined as

$$\beta_0 = 2 \frac{\nu-1}{\nu} \frac{\mu_\chi}{\mu} - \frac{\nu_\chi}{\nu}, \quad (49)$$

$$\beta_1 = \frac{2\nu-1}{\nu} \frac{\mu_\chi}{\mu} - \frac{\nu_\chi}{\nu}, \quad (50)$$

$$\beta_2 = \frac{2\nu-1}{\nu} \frac{\mu_\chi}{\mu}. \quad (51)$$

The general solution of (47) and (48), through order ϵ , may be verified to be

$$p_2 = C_4 - \frac{i}{2} \beta_0 C_5 \ln z^2 + O(\epsilon), \quad (52)$$

$$p_3 = C_5 - i \epsilon C_4 \frac{z}{\nu} - 2 \epsilon C_5 \frac{z \mu_\chi}{\mu \nu} - \epsilon C_5 \frac{\beta_0 z}{2 \nu} (\ln z^2 - 2) + O(\epsilon^2), \quad (53)$$

where C_4 and C_5 are the constants of integration. From these solutions one may now construct the eigenvectors for these slow modes. From the above and (34)–(46), taking the eigen-

vectors to be the appropriate coefficients of C_4 and C_5 , we obtain

$$V_{4o} = \left[-\beta_0 \Omega^2, \frac{\Omega}{k}, -i \epsilon \frac{z \Omega}{k \nu}, \frac{\Omega^2}{k}, \epsilon \frac{\Omega^2 z}{k^2 \nu} \right], \quad (54)$$

$$V_{5o} = \left[-i \beta_0 \frac{\nu}{z} \Omega^2, -i \epsilon \beta_0 \frac{\Omega}{2k} \ln z^2, \epsilon \frac{\Omega}{k} - \epsilon^2 \beta_0 \frac{\Omega z}{2k \nu} \right. \\ \times (\ln z^2 - 2) - \epsilon^2 \frac{2 \Omega z \mu_\chi}{k \mu \nu}, \\ \left. -i \epsilon \beta_0 \frac{\Omega^2}{2k} \ln z^2, i \epsilon \frac{\Omega^2}{k^2 \nu} (1-\nu) - i \epsilon^2 \beta_0 \right. \\ \left. \times \frac{\Omega^2 z}{2k^2 \nu} (\ln z^2 - 2 + 2\nu) - 2i \epsilon^2 \frac{\Omega^2 z}{k^2 \mu \nu} (1-\nu) \mu_\chi \right], \quad (55)$$

where the error in each component is the next order in ϵ . Eigenvector 4 corresponds to the component of the slow mode where the potential vanishes at the resonance while eigenvector 5 corresponds to the component where, in lowest order, the potential is nonzero at the resonance. Thus we will refer to these two modes respectively as the "potential gradient mode" and the "potential mode" due to their structure at this resonance.

Upon taking the same limit for the third outer eigenvector, one obtains

$$V_{3o} = \left[\Omega^2, -i \frac{\epsilon \Omega}{z k}, i \frac{\epsilon^2 \Omega}{z^2 k}, -i \frac{\epsilon \Omega^2}{z k}, -\frac{\epsilon^2 \Omega^2}{z^2 k^2} \right] e^{i \int^y \alpha_3 dy}, \quad (56)$$

where we have taken $A_{3o} = \epsilon^2$ so that the largest component of this eigenvector would be of order unity.

V. THE INNER SOLUTIONS AT THE DIOCOTRON RESONANCE

To construct inner solutions we start with the scalings for the quantities $[n_1, v_{1x}, v_{1y}, \phi_1]$ as given by (20) and (21) and by (38)–(46). Thus we take

$$V = [\Omega^2 \psi_1, \epsilon \Omega \psi_2/k, \epsilon^2 \Omega \psi_3/k, \epsilon \Omega^2 \psi_4/k, \epsilon^2 \Omega^2 \psi_5/k^2], \quad (57)$$

where the ψ s are other unitless functions of z to be determined. Substituting the above into (15)–(18) and using (41) to convert y derivatives to z derivatives as needed, one obtains, through first order in ϵ , the five scaled equations

$$\nu \partial_z \psi_1 = i(\psi_2 - z \psi_3)(1 - \mu^2 - \nu) - \epsilon \psi_3(2\mu_\chi\mu + \nu_\chi) \\ + (\psi_4 - \psi_2)(1 - \mu^2 - \nu)/\epsilon^2 + O(\epsilon^2), \quad (58)$$

$$\nu \partial_z \psi_2 = -iz \psi_2 + \psi_3(1 - \nu) + i \psi_5 + O(\epsilon^2), \quad (59)$$

$$\nu \partial_z \psi_3 = -iz \psi_3 + (\psi_4 - \psi_2)/\epsilon^2 + O(\epsilon^2), \quad (60)$$

$$\nu \partial_z \psi_4 = \psi_1 + O(\epsilon^2), \quad (61)$$

$$\nu \partial_z \psi_5 = \psi_4. \quad (62)$$

The last equation, being the definition of ψ_4 , is exact to all orders.

These equations reduce to a third-order system. Starting with (58) and (60), one deduces

$$\psi_4 = \psi_2 + \epsilon^2 p_4 + O(\epsilon^4), \quad (63)$$

where p_4 is another function of z to be determined. From (59) and using (63) to eliminate ψ_4 from (61), one obtains

$$\psi_1 = -iz\psi_2 + \psi_3(1-\nu) + i\psi_5 + O(\epsilon^2). \quad (64)$$

Using this to eliminate ψ_1 from (58), we have

$$p_4 = i(z\psi_3 - \psi_2) - 2\epsilon\psi_3\mu_\chi/\mu + O(\epsilon^2). \quad (65)$$

Then upon eliminating p_4 from (60), it follows from (62) that

$$\psi_5 = i\psi_3 + C_5 + \epsilon p_5 + O(\epsilon^2), \quad (66)$$

where C_5 is a constant and p_5 is another unknown function of z . From (60) we obtain ψ_2

$$\psi_2 = i\nu\partial_z\psi_3 + 2i\epsilon\psi_3\mu_\chi/\mu + O(\epsilon^2). \quad (67)$$

Lastly we use the above to eliminate all variables except ψ_3 and p_5 from (59) and (62), to obtain the third-order system

$$\nu^2\partial_z^2\psi_3 + [iz + \epsilon(2\mu_\chi/\mu + \nu_\chi/\nu)]\nu\partial_z\psi_3 + 2i\epsilon\psi_3\mu_\chi/\mu - \epsilon p_5 = C_5 + O(\epsilon^2), \quad (68)$$

$$\nu\partial_z p_5 = 2i\psi_3\mu_\chi/\mu + O(\epsilon). \quad (69)$$

The solution of (68) and (69) for ψ_3 and p_5 may be verified to be

$$p_5 = C_4 \frac{iz^2}{\mu\nu} \mu_\chi + C_3 [ize^{i\theta} + (i - z^2/\nu)F_1] \frac{\mu_\chi}{\mu\nu} - C_5 \frac{2z\mu_\chi}{\mu\nu^2} + O(\epsilon), \quad (70)$$

$$\begin{aligned} \psi_3 = & C_4 \left[z + \epsilon \left(i\beta_0 - \beta_2 \frac{z^2}{\nu} \right) \right] + C_3 \left\{ e^{i\theta} + \frac{iz}{\nu} F_1 + \epsilon \left[-\frac{\beta_1}{\nu} \left(i \frac{z^2}{\nu} \right. \right. \right. \\ & \left. \left. \left. + 1 \right) F_1 + ze^{i\theta} \right] \right\} + C_5 \frac{i}{\nu} \left\{ 1 + \epsilon \left[\frac{iz}{\nu^2} (\nu_\chi - \mu_\chi/\mu) \right. \right. \\ & \left. \left. + \frac{\beta_0}{\nu} \left(\frac{z}{\nu} F_2 - iF_1^* e^{i\theta} \right) - i \frac{\beta_1 z}{\nu} \right] \right\} + O(\epsilon^2), \end{aligned} \quad (71)$$

where C_3 and C_4 are additional constants, F_1^* is the complex conjugate of F_1 and

$$\theta = \int_0^z \frac{z}{\nu} dz, \quad (72)$$

$$F_1 = \int_0^z e^{i\theta} dz, \quad (73)$$

$$F_2 = \int_0^z e^{i\theta} F_1^* dz. \quad (74)$$

In lowest order F_1 is a Fresnel integral. The other ψ s may be obtained from the above and (63)–(67).

With this we have three linearly independent inner solutions which correspond to the drift mode and the two slow modes. Letting the independent solutions be the coefficients of C_3 , C_4 , and C_5 , we have

$$\begin{aligned} V_{3i} = & \left\{ \left[-\nu e^{i\theta} + \epsilon\beta_0 F_1 + \epsilon z e^{i\theta} \frac{\nu_\chi}{\nu} \right] \Omega^2, \right. \\ & \times \left[-\epsilon F_1 + \epsilon^2 \beta_0 \frac{z}{\nu} F_1 - 2i\epsilon^2 \left(\beta_1 - \frac{\mu_\chi}{\mu} \right) \right] \frac{\Omega}{k}, \\ & \left[\epsilon^2 \left(e^{i\theta} + iF_1 \frac{z}{\nu} \right) - \epsilon^3 \beta_1 \left(\frac{z}{\nu} + F_1 \left(i \frac{z^2}{\nu^2} + \frac{1}{\nu} \right) \right) \right] \frac{\Omega}{k}, \left[-\epsilon F_1 \right. \\ & \left. + \epsilon^2 \beta_0 \frac{z}{\nu} F_1 - 2i\epsilon^2 \left(\beta_1 - \frac{\mu_\chi}{\mu} \right) \right] \frac{\Omega^2}{k}, \left[\epsilon^2 \left(i e^{i\theta} - F_1 \frac{z}{\nu} \right) \right. \\ & \left. + \epsilon^3 \left(\beta_1 - \frac{\mu_\chi}{\mu} \right) + \epsilon^3 \left(F_1 \frac{z^2 - i\nu}{\nu^2} - i \frac{z}{\nu} e^{i\theta} \right) \right] \frac{\Omega^2}{k^2} \right\}, \end{aligned} \quad (75)$$

$$\begin{aligned} V_{4i} = & \left\{ -i\epsilon\beta_0 \nu \Omega^2, \left[i\epsilon\nu + 2i\epsilon^2 \frac{z\mu_\chi}{\mu\nu} (1-\nu) \right] \frac{\Omega}{k}, \right. \\ & \left[\epsilon^2 z - \epsilon^3 \left(\beta_2 \frac{z^2}{\nu} - \beta_0 \right) \right] \frac{\Omega}{k}, \left[i\epsilon\nu + 2i\epsilon^2 \frac{z\mu_\chi}{\mu\nu} (1-\nu) \right] \frac{\Omega^2}{k}, \\ & \left[i\epsilon^2 z + i\epsilon^3 \left(\frac{z^2\mu_\chi}{\mu\nu^2} (1-\nu) - \beta_0 \right) \right] \frac{\Omega^2}{k^2} \right\}, \end{aligned} \quad (76)$$

$$\begin{aligned} V_{5i} = & \left\{ \frac{i}{\nu} \beta_0 F_1^* \Omega^2, \epsilon \left[\frac{2}{\nu} \left(\beta_1 - \frac{\mu_\chi}{\mu} \right) + i \frac{\beta_0}{\nu^2} F_2 \right] \frac{\Omega}{k}, \right. \\ & \times \left[i \frac{\epsilon}{\nu} + \frac{\epsilon^2}{\nu^2} \beta_0 \left(\frac{z}{\nu} F_2 - iF_1^* e^{i\theta} \right) - 2i\epsilon^2 \frac{z}{\nu^2} \left(\frac{\mu_\chi}{\mu} \right. \right. \\ & \left. \left. - \frac{\nu_\chi}{\nu} \right) \right] \frac{\Omega}{k}, \epsilon \left[\frac{2}{\nu} \left(\beta_1 - \frac{\mu_\chi}{\mu} \right) + i \frac{\beta_0}{\nu^2} F_2 \right] \frac{\Omega^2}{k}, \left[-\frac{\epsilon}{\nu} (1-\nu) \right. \\ & \left. - 2\epsilon^2 \frac{z\nu_\chi}{\nu^3} + \frac{\epsilon^2}{\nu^2} \beta_0 \left(\frac{iz}{\nu} F_2 + F_1^* e^{i\theta} \right) \right] \frac{\Omega^2}{k^2} \right\}, \end{aligned} \quad (77)$$

where the subscript i refers to the inner solution. In the next section we will match the inner and outer solutions and determine the connection coefficients.

VI. THE CONNECTION COEFFICIENTS

In order to match the inner solutions to the outer solutions, we need the asymptotics of the inner solutions in the limit of z large. For this we need the asymptotics of θ , F_1 and F_2 . One may easily verify from (38)–(40) that the asymptotic of θ for large z , up to an additive constant, is just the phase of the drift mode, $\int^y \alpha_3 dy$. The asymptotics of F_1 will be required to first order in ϵ . Thus it will be necessary to include the χ dependence of ν in the evaluation of F_1 . However the asymptotics of F_2 will only be required to zeroth order in ϵ , in which case we may ignore the variation of ν with respect to χ .

As shown in the Appendix, in the limit of $|z|$ large, the asymptotics of these integrals are:

$$\begin{aligned} F_1 \rightarrow & \sqrt{\frac{\pi\nu_0}{2}} \frac{z}{|z|} e^{i\frac{\pi}{4}} - \frac{2i\epsilon}{3\nu_0} \nu_\chi + \left(\frac{i\nu}{z} + \frac{\epsilon}{z^2} \nu_\chi \right) e^{i\theta} + O(\epsilon^2) \\ & + O(1/z^3), \end{aligned} \quad (78)$$

$$F_2 \rightarrow \frac{-i}{2} \nu \left[i \frac{\pi}{2} + \gamma + \ln \left(\frac{4z^2}{\nu} \right) \right] + O \left(\frac{1}{z} \right), \quad (79)$$

where ν_0 is the value of ν at the resonance ($z=0$) and $\gamma=0.57721\dots$ is Euler's constant.

Let us evaluate the connection coefficients in the following manner. Start with a linear superposition of the three outer solutions below the resonance ($z < 0$) of the form

$$V = C_3^{(-)} V_{3o} + C_4^{(-)} V_{4o} + C_5^{(-)} V_{5o}, \quad (80)$$

where $C_n^{(-)}$ is the amplitude of the n th eigenvector and the superscript $(-)$ indicates that this is the amplitude below the resonance. We take the form of the interior solution to be

$$V = D_3 V_{3i} + D_4 V_{4i} + D_5 V_{5i}, \quad (81)$$

where D_n is the amplitude of the n th interior solution. Then upon inserting (78) and (79) into (75)–(77), for z large and negative, and upon comparing the result with (54)–(56), one finds that in lowest order of ϵ ,

$$D_3 = -\frac{C_3^{(-)}}{\nu_0} - \beta_0 C_5^{(-)} \sqrt{\frac{\pi}{2\nu_0}} e^{i\pi/4}, \quad (82)$$

$$D_4 = -\frac{C_3^{(-)}}{\nu_0} \sqrt{\frac{\pi}{2\nu_0}} e^{i\pi/4} - i \frac{C_4^{(-)}}{\nu_0} + \frac{C_5^{(-)}}{2\nu_0} \mathcal{E}, \quad (83)$$

$$D_5 = -i\nu_0 C_5^{(-)}, \quad (84)$$

where

$$\mathcal{E} = \beta_0 \left[\gamma + 1 + \ln \left(\frac{4}{\nu_0} \right) - i \frac{\pi}{2} \right] - 2 \frac{\nu_\chi}{\nu_0}. \quad (85)$$

Repeating the same for z large and positive, we take the outer solution to be the linear superposition

$$V = C_3^{(+)} V_{3o} + C_4^{(+)} V_{4o} + C_5^{(+)} V_{5o}, \quad (86)$$

where the superscript $(+)$ indicates that this coefficient is for $z > 0$ and large. Then we obtain

$$D_3 = -\frac{C_3^{(+)}}{\nu_0} + \beta_0 C_5^{(+)} \sqrt{\frac{\pi}{2\nu_0}} e^{i\pi/4}, \quad (87)$$

$$D_4 = \frac{C_3^{(+)}}{\nu_0} \sqrt{\frac{\pi}{2\nu_0}} e^{i\pi/4} - i \frac{C_4^{(+)}}{\nu_0} + \frac{C_5^{(+)}}{2\nu_0} \mathcal{E}, \quad (88)$$

$$D_5 = -i\nu_0 C_5^{(+)}. \quad (89)$$

Combining (82)–(84) with (87)–(89), upon eliminating the D s we then have

$$C_3^{(+)} = C_3^{(-)} + \beta_0 C_5^{(-)} \sqrt{2\pi\nu_0} e^{i\pi/4} + O(\epsilon), \quad (90)$$

$$C_4^{(+)} = C_4^{(-)} + C_3^{(-)} \sqrt{\frac{2\pi}{\nu_0}} e^{-i\pi/4} + \pi\beta_0 C_5^{(-)} + O(\epsilon), \quad (91)$$

$$C_5^{(+)} = C_5^{(-)} + O(\epsilon). \quad (92)$$

Some physical understanding to these results can be obtained by noting the following points. Recall that β_0 , (49), is essentially the negative of the logarithmic derivative of the

density with respect to χ , and as such, it is only significantly nonzero at the edge of the sheath. Away from the edge we may take β_0 to be essentially zero. Thus when $\beta_0 \approx 0$, we see that the amplitudes of the drift mode and potential mode would be transmitted across the resonance without change. Meanwhile the potential gradient mode would always be modified, picking up an additional part proportional to the amplitude of the drift mode. On the other hand when the resonance is near the edge such that β_0 is significantly nonzero, then the potential mode always modifies the other two modes, although itself is transmitted across the resonance unchanged.

Let us review the nature of these three modes in the outer region ($|z| \gg 1$). From (56) we see that the drift mode basically only has fast density oscillations as long as z is in the outer region. The potential gradient mode, (54), has a rf density “bulge” at the resonance, on the slow scale, proportional to β_0 . It also has a slow rf potential gradient and essentially a vanishing value for the potential at the resonance. On the other hand the potential mode, (55), is more complex, but basically is the only mode with a nonzero value for the rf potential at the resonance.

Now consider the following. Potentials are more difficult to shift than potential gradients, which could then be taken as an explanation for why the potential mode tends to be so resilient and transmitted across the resonance unchanged. (To justify this explanation, note that without work being done, one can shift the charge distribution inside the plasma such that one shifts the potential gradient but not the overall potential.) Consider when we have an absence of a density gradient at the resonance and thus $\beta_0 \approx 0$. The drift mode, containing essentially only fast density oscillations, would tend to simply transmit unchanged across the resonance. However as this drift mode crosses the resonance, since the WKB wave vector α_3 will cross zero here, its fast oscillations will become slower and slower and eventually reverse. Thus at the resonance, this mode would leave behind a nonzero average for the rf potential gradient, thereby contributing to the potential gradient mode. On the other hand since its density average is zero it would not significantly contribute to the potential mode. Now let the resonance be located near the edge of the sheath so that $\beta_0 \neq 0$. Here we have an additional term coming into play which is the coupling provided by the dc density gradient at the edge of the sheath as in (52) and (53). This dc density gradient, combining with the potential mode, then generates contributions to both the drift mode and the potential gradient mode.

VII. CONCLUSIONS

We have shown that in the saturation stage the rf component of the non-neutral plasma inside a magnetron becomes a fifth-order system. This fifth-order system consists of two slow modes, two fast drift-cyclotron modes (Slater orbits) and one fast drift mode. The three fast modes are not found in the initiation stage since their existence requires the dc drift velocity to have a component directed from the cathode to the anode. Although the current generated by this vertical dc velocity is only a small fraction of the total cur-

rent circulating in the device, it is exactly this smallness which creates a dramatic change in this structure of the rf solutions. In particular this vertical dc current causes resonance layers to occur at the position of any resonance in the plasma, inside of which mode conversions between these five modes can occur. These resonances were already in the initiation stage: (1) the diocotron resonance at $\omega_e=0$, and (2) the two cyclotron resonances at $\omega_e=\pm\Omega$. The structure of these resonances occurs due to the small parameter (vertical velocity) multiplying three of the derivatives of the fifth-order set of ODEs.³¹

As we have already seen in Fig. 6 these fast modes prevent one from obtaining rapid numerical solutions in this stage. Now that the structure and location of these resonances are known, and the structure and widths of the resonance layers, it will be possible to develop numerical code to rapidly solve this fifth-order system. For example if there was no magnetron resonance inside the plasma then the drift-cyclotron modes never cross the other resonances. In this case the fifth-order system can be reduced to a third-order system and two first-order systems resulting in a very significant decrease in the required computational time for the drift-cyclotron modes. In effect what one can do is to use the WKB approximation (followed by a perturbation expansion) to separate the fifth-order system into three first-order ODE's and one second-order ODE, each of which could be rapidly computed since the fast variations could be analytically separated out. The only problem would be in the region of any resonances where one would have to appropriately couple the ODEs and carefully numerically integrate through the resonance. However such can be done: one can separately propagate each fast mode from the cathode until it either reaches the anode or crosses the two slow modes. When the latter occurs one simply does a separate computation around the resonance.

Then there is the question of boundary conditions on the rf fields. In the initiation stage one only requires two boundary conditions: (i) the vanishing of the rf potential at the cathode and (ii) specifying the magnitude of the rf density oscillations at the anode. The former follows whenever the cathode is a smooth conductor and the latter is generally determined by the dc density profile. (Note that since the electron density can never be negative, the amplitude of the rf density oscillations must always be less than the dc electron density.²⁰ Taking these to be equal has the effect of coupling the rf amplitude at the anode to the dc amplitude at the anode.) For the saturation stage three additional boundary conditions will be required. It seems reasonable that one should take these three conditions to be the vanishing of all fast modes at the cathode. This would follow from recent results^{10-13,32} on the stability of various non-neutral flows. However it should also be mentioned that various recent methods for breaking up the laminar flow at the cathode, as in Ref. 27, could be found to be a method for creating non-zero sources of fast modes at the cathode.

In the usual situation of a smooth cathode with no magnetron resonance present, the rf structure in the saturation stage becomes simpler since only the diocotron resonance exists. Taking the fast modes to be zero at the cathode, it

follows that both drift-cyclotron modes will be absent. Also the drift mode will be zero between the cathode and the location of the diocotron resonance. Now we apply (90)–(92) and find that above the diocotron resonance the drift mode will exist with $C_3^{(+)} \propto \beta_0 C_5^{(-)}$. Thus the total rf field at the anode would then consist of a drift mode and two slow modes and the ponderomotive terms in (6)–(9) would consist of two parts: the usual contribution from the slow modes and an additional contribution from the drift-cyclotron mode created at the diocotron resonance from the slow modes.

How important are these three fast modes to the generation of the rf power of a magnetron? At the moment we cannot assess this quantitatively since we still are to obtain self-consistent saturation stage solutions for dc and the rf fields [i.e. satisfying (6)–(13)]. We do know that the growth rate vanishes when the device is in the saturation stage. We do know that if we start with the dc field of the initiation stage that we obtain the rf solutions discussed herein. We would expect the dc solution of the saturation stage, satisfying (6)–(9), to be qualitatively similar to that of the initiation stage as given in Figs. 3–5. We do know that these devices generally operate only when the diocotron resonance occurs at or near the middle of the density profile and where the density gradient is large. We do know that if the parameters are outside this range then the device will not operate. In the latter case $\beta_0 \ll 1$ and per (90)–(92), essentially no drift mode would be produced. What we do not know is the range of values for β_0 in (90). One can argue from (30) and (37) that for $\beta_0 \gg 1$, the amplitude of the rf electric field of the drift mode at the anode could even exceed that of the slow modes. It all depends on the sharpness of the edge of the electron sheath in the saturation stage. All these points are also consistent with the statement: a magnetron will only operate when a significant drift mode has been generated at the diocotron resonance from the slow modes. Thus it certainly appears that the drift mode is a key component in the operation of a magnetron. We will understand more about this after self-consistent solutions for the saturation stage are obtained following along the lines of the self-consistent calculations done for the initiation stage.⁹

In this paper we have detailed the five independent solutions in the region of the diocotron resonance. We have determined the conversion coefficients for the three modes concerned and have given a general explanation for their structure. In particular, the amplitude of the potential mode remains unaffected in crossing the resonance while the drift mode and the potential gradient mode pick up additional contributions to their amplitudes from the potential mode with the coupling being proportional to the dc density gradient. The drift mode will always add a contribution to the potential gradient mode due to the break in the potential gradient of the drift mode at the resonance.

In summary, we have studied the nature of the rf solutions of the steady-state magnetron when it is steadily delivering rf power, particularly the rf modes in the non-neutral plasma in the neighborhood of the diocotron resonance. We have found that in this neighborhood the ordinary differential equations for the rf fields are nonsingular and that nonsingular solutions do exist for the rf quantities. This resonance

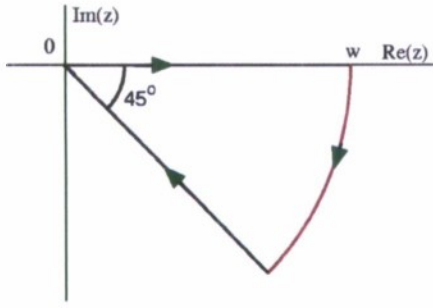


FIG. 9. The contour in the complex plane for evaluating $F_1(z)$ in the limit of large z .

layer is relatively thin and in general the rf fields will rapidly vary as they pass through the resonance in a manner similar to what occurs at boundary layers. During the passage of the slow modes and/or the drift mode through this resonance, conversion from one mode to another will occur with the conversion coefficients being given by (90)–(92).

ACKNOWLEDGMENT

This research was supported in part by the AFOSR.

APPENDIX: ASYMPTOTIC LIMITS OF INTEGRALS

In this Appendix we will derive the asymptotic limits of F_1 and F_2 for $|z| \rightarrow \infty$.

Starting with F_1 , note that we will need this result to order ϵ . Thus we cannot just simply replace F_1 by a Fresnel integral, but we must consider the first-order variations of ν with respect to χ . As long as ν does not significantly change over the region where $|z| < \approx 1/\sqrt{\epsilon}$, we may consider ν to be an analytic function of z and use a contour integral to evaluate the asymptotic value of F_1 . For clarity we introduce dummy variables w and u and rewrite (72) and (73) as

$$\theta(z) = \int_0^z \frac{udu}{\nu(u)}, \quad F_1(w) = \int_0^w e^{i\theta(z)} dz. \quad (\text{A1})$$

Consider the integral of $\int e^{i\theta(z)} dz$ around the contour shown in Fig. 9 for w large and positive. It can be seen that $F_1(w)$ is the integral along the real-axis segment of the contour. By assumption θ will be analytic inside the contour. Thus there will be no residue and the value of the contour integral will be zero. Thus the value of the real-axis segment will be the negative of the contributions along the other two segments.

Along the arc, the major contribution to the integral will come from the region of the arc next to the real axis. To evaluate the integral along this arc we expand $\theta(z)$ in a Taylor series, expanded about $z=w$. Then evaluation of this integral gives

$$\left[-\frac{i}{w} - \frac{\epsilon\nu_\chi}{w^2} + O\left(\frac{1}{w^3}\right) \right] e^{i\theta(w)} + O(\epsilon^2). \quad (\text{A2})$$

Along the ray the major contribution to that segment of the contour integral will come from the region next to the origin. Here we expand $\theta(z)$ in a Taylor series, expanded about z

$=0$, and the value of the integral is then found to be

$$-\sqrt{\frac{\pi\nu_0}{2}} e^{-i\pi/4} + \frac{2i\epsilon\nu_\chi}{3\nu_0} + O(\epsilon^2), \quad (\text{A3})$$

where ν_0 is the value of ν at the center of the resonance. Since F_1 will be the negative of the sum of these two integrals, we have that in the limit where z is large and positive,

$$F_1(z) \rightarrow \sqrt{\frac{\pi\nu_0}{2}} e^{-i\pi/4} - \frac{2i\epsilon\nu_\chi}{3\nu_0} + \left[\frac{i}{z} + \frac{\epsilon\nu_\chi}{z^2} + O\left(\frac{1}{z^3}\right) \right] e^{i\theta(z)} + O(\epsilon^2). \quad (\text{A4})$$

To evaluate the limit of $F_2(z)$ for large $|z|$, we only need the limit to order unity and thus we take ν to be a constant. In this case, from (74) and (A1), $F_2(z)$ may be given by

$$F_2(z) = \int_0^z du \int_0^u e^{-i\left(\frac{u^2-v^2}{2\nu}\right)} dv. \quad (\text{A5})$$

This double integral may be evaluated by transforming to a hyperbolic coordinate system. Let us take $u=w \cosh \varphi$ and $v=w \sinh \varphi$. Then the above integral can be transformed into

$$F_2(z) = I_1(z) + I_2(z), \quad (\text{A6})$$

where

$$I_1(z) = \int_0^z w e^{\frac{-iw^2}{2\nu}} \ln\left(\frac{z}{w}\right) dw, \quad (\text{A7})$$

$$I_2(z) = \int_0^z w e^{\frac{-iw^2}{2\nu}} \ln\left(1 + \sqrt{1 - \frac{w^2}{z^2}}\right) dw. \quad (\text{A8})$$

Integrating by parts on the first integral, to eliminate the $\ln(z/w)$ term gives

$$I_1(z) = \frac{i}{2}\nu \left(-i\frac{\pi}{2} - \int_0^{\frac{z^2}{2\nu}} \frac{1-e^{-t}}{t} dt \right), \quad (\text{A9})$$

where the last integral is listed in Ref. 33. In the limit of large $|z|$ this gives

$$I_1(z) = \frac{i}{2}\nu \left[-i\frac{\pi}{2} - \gamma - \ln\left(\frac{z^2}{\nu}\right) \right], \quad (\text{A10})$$

up to exponentially small terms. In the above, $\gamma = 0.57721\dots$ is Euler's constant.

For I_2 , we first transform variables from w to p where $w^2 = z^2(1-p^2)$ and then we perform an integration by parts to eliminate the $\ln$ term. This gives

$$I_2(z) = -i\nu \ln 2 + i\nu e^{-\frac{iz^2}{2\nu}} \int_0^1 \frac{e^{i\frac{z^2 p^2}{2\nu}}}{1+p} dp. \quad (\text{A11})$$

The major contribution to the remaining integral comes from the region near $p=0$, so we replace $\frac{1}{1+p}$ by $1 - \frac{p}{1+p}$ under the integral. When that is done, the first term becomes a Fresnel integral and can be evaluated as with $F_1(z)$ above. The second term, $\frac{p}{1+p}$, under the integral can be shown to be bounded by an order of $1/z^2$ for large $|z|$. Thus we have

$$I_2(z) = -i\nu \left[\ln 2 - \sqrt{\frac{\pi\nu}{2z^2}} e^{-\frac{iz^2}{2\nu}} e^{i\frac{\pi}{4}} + O\left(\frac{1}{z^2}\right) \right]. \quad (\text{A12})$$

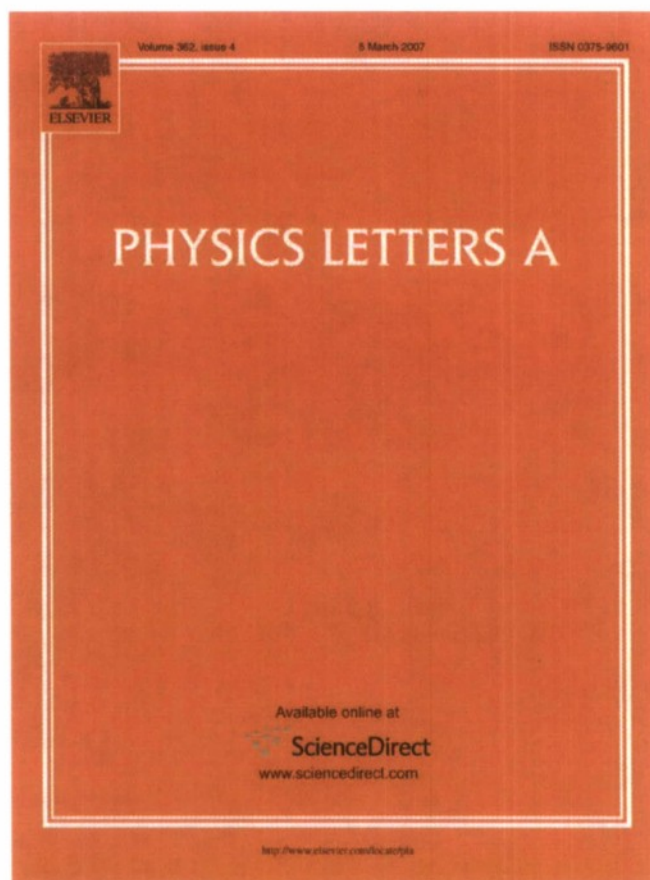
Combining the above results for (A10) and (A12) into (A6) gives

$$F_2(z) \rightarrow \frac{-i}{2} \nu \left[i\frac{\pi}{2} + \gamma + \ln\left(\frac{4z^2}{\nu}\right) - e^{-\frac{iz^2}{2\nu}} e^{i\frac{\pi}{4}} \sqrt{\frac{2\pi\nu}{z^2}} \right] + O\left(\frac{1}{z^2}\right) + O(\epsilon), \quad (\text{A13})$$

for $|z|$ large.

- ¹J. C. Slater, *Microwave Electronics* (Van Nostrand, Princeton, 1959), Chap. 13.
- ²O. Buneman, R. H. Levy, and L. M. Linson, *J. Appl. Phys.* **37**, 3203 (1966).
- ³Mark K. Jablonski, "A comparison of linear perturbation analysis with computer simulation results for a planar magnetron," M. Sc. thesis, Massachusetts Institute of Technology (1992).
- ⁴K. L. Cartwright, P. J. Christenson, J. P. Verboncoeur, and C. K. Birdsall, *Phys. Plasmas* **7**, 1740 (2000).
- ⁵Ronald C. Davidson, *Phys. Fluids* **28**, 1937 (1985).
- ⁶D. J. Kaup and G. E. Thomas, *Stud. Appl. Math.* **81**, 57 (1989); D. J. Kaup and G. E. Thomas, *Stud. Appl. Math.* **83**, 271 (1990).
- ⁷Spilios Riyopoulos, *Phys. Plasmas* **2**, 935 (1995).
- ⁸Spilios Riyopoulos, *Phys. Plasmas* **3**, 1137 (1996).
- ⁹D. J. Kaup, *Phys. Plasmas* **8**, 2473 (2001).
- ¹⁰P. J. Christenson and Y. Y. Lau, *Phys. Plasmas* **1**, 3725 (1994).
- ¹¹P. J. Christenson and Y. Y. Lau, *Phys. Rev. Lett.* **76**, 3324 (1996).
- ¹²P. J. Christenson, D. P. Chernin, A. L. Gardner, and Y. Y. Lau, *Phys. Plasmas* **3**, 4455 (1996).

- ¹³A. L. Gardner, Y. Y. Lau, and D. P. Chernin, *Phys. Plasmas* **5**, 2447 (1998).
- ¹⁴D. J. Kaup and Gary E. Thomas, *Phys. Plasmas* **3**, 771 (1996).
- ¹⁵P. G. Drazin and W. H. Reid, *Hydrodynamic Stability* (Cambridge University Press, New York, 1981), Chap. 4.
- ¹⁶M. P. Desjarlais and R. N. Sudan, *Phys. Fluids* **29**, 1245 (1986).
- ¹⁷M. P. Desjarlais and R. N. Sudan, *Phys. Fluids* **30**, 1536 (1987).
- ¹⁸D. J. Kaup and Gary E. Thomas, *J. Plasma Phys.* **58**, 145 (1997).
- ¹⁹D. J. Kaup and Gary E. Thomas, *J. Plasma Phys.* **59**, 259 (1998).
- ²⁰D. J. Kaup and Gary E. Thomas, *Proc. SPIE* **4031**, 54 (2000).
- ²¹R. C. Davidson, *Phys. Fluids* **27**, 1804 (1984).
- ²²Y. Y. Lau and D. Chernin, *Phys. Fluids B* **4**, 3473 (1992).
- ²³Spilios A. Riyopoulos, *IEEE Trans. Plasma Sci.* **22**, 626 (1994).
- ²⁴Spilios Riyopoulos, *Phys. Plasmas* **6**, 323 (1999).
- ²⁵Spilios Riyopoulos, *IEEE Trans. Plasma Sci.* **26**, 755 (1998).
- ²⁶Hunter L. McDowell, *IEEE Trans. Plasma Sci.* **30**, 980 (2002).
- ²⁷J. W. Luginsland, Y. Y. Lau, V. B. Neculaes, R. M. Gilgenbach, M. C. Jones, M. H. Frese, and J. J. Watrous, *Appl. Phys. Lett.* **84**, 5425 (2004).
- ²⁸D. J. Kaup, J. O. El-Reedy, and Gary E. Thomas, *Proc. SPIE* **3702**, 57 (1999).
- ²⁹D. J. Kaup, *Proc. SPIE* **4720**, 67 (2002).
- ³⁰D. J. Kaup, "Resonances in High Density Nonneutral Plasmas," *Frontiers of Nonlinear Physics*, in Proceedings of the Second International Conference, Nizhny Novgorod, St. Petersburg, Russia, 2004, edited by Alexander Litvak (Nizhny Novgorod, Russia, Institute of Applied Physics RAS, 2005), pp. 308–320.
- ³¹Carl M. Bender and Steven A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978), Chap. 9.
- ³²Naoufel Ben Abdallah, Pierre Degond, and Florian Mèhats, *Phys. Plasmas* **5**, 1522 (1998).
- ³³*Handbook of Mathematical Functions*, edited by Abramowitz and Stegun (Dover, New York, 1972).



This article was originally published in a journal published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues that you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>



Quantitative measurement of variational approximations[☆]

D.J. Kaup^{a,b,*}, T.K. Vogel^a

^a Department of Mathematics, University of Central Florida, Orlando, FL 32816-1364, USA

^b Institute of Simulation and Training (IST), University of Central Florida, Orlando, FL 32816-0544, USA

Received 25 April 2006; accepted 8 December 2006

Available online 22 December 2006

Communicated by A.P. Fordy

Abstract

Variational problems have long been used to mathematically model physical systems. Their advantage has been the simplicity of the model as well as the ability to deduce information concerning the functional dependence of the system on various parameters embedded in the variational trial functions. However, the only method in use for estimating the error in a variational approximation has been to compare the variational result to the exact solution. In this Letter, we demonstrate that one can computationally obtain estimates of the errors in a one-dimensional variational approximation, without any a priori knowledge of the exact solution. Additionally, this analysis can be done by using only linear techniques. The extension of this method to multidimensional problems is clearly possible, although one could expect that additional difficulties could arise.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

In the theory of modern dynamics, there exist two paradigmatic views on formulating the mathematical model for a physical system. These paradigms arrive at a completely equivalent result. The first perspective, commonly viewed as the Newtonian approach, argues that the evolution of the state of a process can be completely described in terms of the *forces* involved with the process. This approach is a manifestation of the Aristotelian “cause and effect” view of the universe. The mathematical framework for the second perspective is based on two publications [1,2] in the 1830s by William Hamilton. Within these papers, Hamilton proposes a theory of dynamics which describes all of classical (Newtonian) mechanics *without the use of forces*. Instead, the physical process under consideration will evolve in such a way as to extremize the integral of the difference between the kinetic and potential energies. Philosophically, this is in stark contrast to the Newtonian perspective. Instead of viewing the evolution of a process as a result of ex-

ternal influences, the Hamiltonian approach suggests there is an *internal intent* on the part of the process to evolve in a particular manner.

Variational problems have existed in mathematics throughout history. The earliest that we have located was in the *Aeneid* [3], where there is a legend of a woman who fled to Africa after her husband was murdered. She arrived somewhere along the North African coast where she pleaded with the local ruler, King Jambus, for land. The King granted the woman as much land as could be enclosed with the hide of a bull. Legend says that the clever lady cut the hide into very thin and long strips and laid it out to enclose the maximum possible area (a circle). The land later became the city of Carthage (c. 814 BC) and the woman its ruler, Queen Dido. While legend and fact may not necessarily agree, this illustrates that the idea of extremum principles has existed for millennia. Eight centuries later, Hero of Alexandria (c. 10–70) proved the first recorded scientific minimum principle. He was able to show that the trajectory of a reflected light ray is a minimum if the angles of incidence and reflection are equal [4]. This idea was later formulated as a least time principle by Fermat in the early part of the seventeenth century [4].

The mathematical machinery necessary to investigate such problems was found to require more than the elementary

[☆] Research supported in part by the National Science Foundation Grant #DMS-0505566 and the Air Force Office of Scientific Research.

* Corresponding author.

E-mail address: kaup@mail.ucf.edu (D.J. Kaup).

calculus developed by Newton [5] and Leibniz in the mid-seventeenth century. In 1696, Johann Bernoulli proposed "... *brachistochrone problem which can be summarized as 'If two points are connected by a wire whose shape is given by an unknown function $y(x)$ in a vertical plane, what shape function minimizes the time of descent of a bead sliding without friction from the higher to the lower point.'*" [6]. This problem was addressed by several of Bernoulli's contemporaries, and a new type of calculus arose from those investigations. This calculus is known today as the *calculus of variations* and was developed into a full mathematical theory by Euler around 1744 [7]. The mathematics developed by Euler was extended by Joseph-Louis Lagrange (1736–1813). Lagrange discovered that Euler's equation (later to be named the Euler–Lagrange equation) for minimizing a functional integral could be expressed in a compact way by simply using integration by parts. It was Lagrange who introduced the integrand of the functional appropriate to mechanics, i.e., the difference between kinetic and potential energies. Euler had essentially only considered the kinetic energy [8]. This approach amounted to requiring additional conditions to get a complete formulation of classical mechanics.

Since the solution(s) of a dynamical system is one which extremizes this quantity called the action, it therefore follows from this principle that one can also devise methods for finding approximate solutions. Thus, one can reduce the number of degrees of freedom in which the function can be varied and Hamilton's Principle will lead one to a solution which, in some sense, will be as close as possible to the correct solution. Due to the reduction in the number of degrees of freedom, one cannot generally expect the function to achieve the actual extremum of the full problem. "... *these methods aim to "project" the full dynamics onto a finite-mode space, or, in other words, truncate a system with infinitely many degrees of freedom to a finite-dimensional one*" [9]. However, one can still expect to find an extremum which should be "near" the exact solution. This method has long been used in various applications ranging from geometry [10] to quantum mechanics [11,12]. In particular, it was in this period that we have the "Raleigh–Ritz method" being developed independently by Raleigh and Ritz,¹ for finding eigenfunctions and eigenvalues of linear differential equations. This method is known in modern mathematical literature as the Raleigh–Ritz method [13,14]. In fact, this usage of variational methods continued into the first half of the 20th century, and even up into the late 1980's. During this time, quantum mechanics and all of modern physics were born and with these, there came a critical need to be able to obtain numerical values to compare with experiments. Up until the 1950s there were no electronic computers; even into the latter half of the 20th century, such electronic devices were generally only available at large government labs or universities where their usage was generally quite limited. Otherwise, one only had slide-rules or

crude, electric-operated, mechanical calculators for use. Thus, there was a definite need for the development of methods to obtain some type of approximate solutions consistent with the constraints imposed by these computational devices. Consequently the power of analytical methods of mathematics was put into use with these devices. As an answer to this need, there were major efforts to utilize variational methods for developing approximate solutions, as well as perturbation expansions of such, for various physical systems [15]. We note that in this time period nonlinear problems were generally not studied as such, except to the degree that one could expand about some solvable linear problem, or some known analytical solution. Thus, it is not surprising that all the work in this period mainly concentrated on linear eigenvalue problems and their perturbations, with variational methods playing leading roles.

The need for this development of variational methods began to decline after the 1970s; the availability of calculators and computers and particularly the advent of the personal computer and the accompanying increase in computational power that came with these devices and their larger brothers. This marked the end of an era and the beginning of a new one. Within a decade, it would no longer be important to carefully and analytically expand solutions of equations of motion, in order to obtain a three-to-four place accuracy in their numerical values, when with the touch of a few keys, the value desired would appear almost instantly and with an 8-to-14 place accuracy.

Nevertheless, even though general computations no longer needed to utilize variational methods, there have always been other aspects of variational methods which current computational power cannot hope to match. In particular, variational methods have always been adept at providing functional dependences and trends, and thereby a single, simple variational result could contain more information than dozens of pages of numerical computations. For example, when a nonlinear problem contains three parameters and is also a differential system of twelfth-order, one would like to be able to localize (as much as possible) the parameter regime wherein one would need to search for soliton solutions. This can be done with variational methods [16]. Furthermore, variational techniques developed in numerical methods (such as the Galerkin method) have also been applied to develop variational methods for dissipative systems [17] and also to study the parameter space of electron devices [18]. Indeed, variational methods have been used to find approximate solutions in various cases, particularly in cases where closed-form solutions would not generally exist [16,17,19–23]. The most recent comprehensive review of current applications and successes of the variational method has been given by Malomed [9] and is particularly recommended as a reference on variational methods. However, in spite of the successes that have been achieved, variational methods can also lead one to uncertain or wrong results [18, 22,24]² and thus, there is also a need for an even deeper and

¹ The authors were not able to identify historical records of the work done by Raleigh and Ritz, only that the name of the method seemed to have been adopted sometime during the late 19th Century, which was the time period in which Raleigh and Ritz performed their work.

² We note that Ref. [24] had used the variational approximation to make predictions on the stability of gap solitons, and had found that all models, including the integrable Massive Thirring model, contained parameter regimes in which the solitary wave solutions were unstable. Integrable systems are not normally

more through understanding of these methods. In particular, there is a need to be able to discern when a variational method is not “sufficiently close” to serve as an adequate approximation.

At this point, let us define some nomenclature. By “Hamilton’s Principle”, we mean the general principle which states that the equations of interest can be obtained by the extremization of some functional quantity, called the “action”. The equations obtained are called the Euler–Lagrange equations. The action is obtained by integrating (summing) over an appropriate Lagrangian. By “variational method” we mean a method that makes use of Hamilton’s Principle in some manner. When one uses Hamilton’s Principle to find approximate solutions, we shall refer to the solution found as the “variational approximation” (VA). We note that the Raleigh–Ritz method also uses Hamilton’s Principle to obtain a VA. The term VA is intended to be an inclusive term and to include applications to nonlinear systems as well. More general variational approaches, such as those used in Ref. [17], should perhaps be distinguished and called “Galerkin variational methods” due to the approach used to construct the Lagrangian. We shall not otherwise discuss those methods here. However, the reader should be aware that what we shall present here can easily be extended, at least in principle, to all variational methods.

The major criticism of any use of the VA has generally concentrated on the fact that there was no general method for estimating the quantitative validity of the VA, except by comparison with the exact solution. With the VA, one has obtained an approximate solution. But how could one determine whether or not this approximate solution was of sufficient accuracy to be of value? In the absence of having the exact solution, one had no way to answer this. This fact has always caused one to hesitate, even if only momentarily, before using or quoting variational results, despite its many frequent successes. This hesitation has even been addressed in literature pertaining to variational approximations. “... As for the necessity to verify the validity of the results against direct simulation, this does not devalue the VA, since it is frequently sufficient to perform the comparison at a few different values of the problem’s control parameters. If the comparison at several benchmark points corroborates the applicability of VA, then its (semi-)analytical predictions are reliable enough to describe solitons in broad parametric regions” [9]. What is lacking is a general technique for the quantitative evaluation of any VA.

As mentioned above, in the VA one expects to find a solution “close” to the exact solution. But how close is “close”? The

key to determining this is to return to the early 20th century when variational methods were being used with great intensity. This was so much the case that even perturbation methods for VAs were being developed [15]. Certainly it would have been generally known that the size of a term at any one given order would have given an estimate of the accuracy at that order. And certainly this would have been used to determine to how many orders one needed to expand. At the same time, this clearly tells us how to quantitatively estimate any VA. What has to be done is to formulate the VA as the leading term in a perturbation expansion. Then provided one is expanding about a function which is sufficiently close to the exact solution, any term in the perturbation expansion will indicate the resulting error if the series is terminated at that term.

Note that we are discussing functional accuracy, not numerical accuracy. This is a step higher than what was used in the Raleigh–Ritz method, wherein the criteria was only the accuracy of the eigenvalue. At the same time, there is no one single number that can represent “functional accuracy”. The purpose of this Letter is to present a method for *quantitatively* estimating the functional error in any VA by using only linear techniques and without having the exact solution at hand. By utilizing the functional accuracy, one will be able to determine the numerical accuracy in any quantity of interest.

The method of the VA is quite straightforward: establish the Lagrangian for the governing system, choose a reasonable trial function (ansatz) containing certain arbitrary parameters, calculate the action for that ansatz and then determine the optimum parameters of the ansatz by solving the Euler–Lagrange equations [25]. Given an optimized ansatz, one can then quantitatively estimate the error in the VA without needing the exact solution. To do this, all that is required is for one to be able to solve a linear problem, to be defined below.

To illustrate this method, two physical systems will be analyzed. These example systems have been chosen to be simple so that the method can be easily illustrated. These examples are also only one-dimensional for the same reason. However, it is clear that an extension of this method to multidimensional problems can be developed, although it could possibly be fraught with additional difficulties. But here, the first system that we shall study is the fundamental mode of a vibrating string. Using a sawtooth ansatz and a parabolic ansatz, we will demonstrate how to quantitatively determine how “close” the VA is to the correct solution. The second model will be the well-known soliton solution of the Korteweg–de Vries (KdV) equation. Here we will use a Gaussian ansatz, and demonstrate how to quantitatively estimate two types of errors which this VA could introduce into another calculation. In one case, the Gaussian ansatz is an acceptable approximation while in the other case, it is unacceptable.

2. Theory

Let us first outline the theory behind this method of quantitatively estimating the error in a VA. Suppose the Lagrangian, L , corresponding to the physical system of interest is known

unstable, whence their result on the Massive Thirring model was quite surprising. Afterwards, Ref. [22] pointed out how a variational approximation could indeed produce an unstable result for a stable system. Thus the blame for the mistake seemed to be due to the variational method. However later Ref. [27] verified that the instability results of Ref. [24] were indeed true, up to the one exception: that being the integrable Massive Thirring model, whose solitons were all found to be actually stable. Thus Ref. [22] was indeed correct in that the VA of Ref. [24] did mispredict the stability of the integrable Massive Thirring model and demonstrated how such could occur. On the other hand, and not in conflict with this, Ref. [24] did indeed correctly predict the general stability properties of those gap solitons belonging to the non-integrable models.

and that this Lagrangian is a functional of the variable $u(t)$. The variable $u(t)$ may be a scalar, vector or tensor quantity. For now we shall only consider a one-dimensional scalar case, where the integration variable is t .

The action, $S[u]$, is defined by

$$S = \int_{\mathcal{D}} L[u] dt, \quad (1)$$

where $\mathcal{D}$ is the domain of support of the function $u(t)$. Hamilton's Principle states that the governing equations of the variable $u(t)$ will be such that

$$\lim_{\epsilon \rightarrow 0} \frac{S[u(t) + \epsilon \delta u(t)] - S[u(t)]}{\epsilon} = \int_{\mathcal{D}} \frac{\delta L}{\delta u} [u(t)] \delta u(t) dt = 0, \quad (2)$$

for any bounded variation $\delta u(t)$, provided that this variation vanishes at all end points of the domain $\mathcal{D}$. Note that this also defines the quantity $\frac{\delta L}{\delta u}(t)$, which is called the (first) variational derivative of L .

Now, let us define the (nonlinear) operator Φ to be the first variational derivative of L :

$$\Phi[\bullet] = \frac{\delta L[\bullet]}{\delta u}. \quad (3)$$

This gives us the Euler–Lagrange equation,

$$\Phi[u_e] = 0, \quad (4)$$

where u_e will be the exact solution(s). (There could be more than one solution, depending on the boundary conditions imposed on u_e .) Thus, the problem is to find u_e . Let us now use Hamilton's Principle to obtain a variational approximation to $u_e(t)$. The problem is then to find the “best” solution for whatever ansatz is chosen (the VA) and then to obtain a quantitative estimate of its resulting error(s).

As mentioned earlier, we want to expand u_e in a perturbation expansion about the ansatz plus some error. Let $u_0(t; q)$ be the ansatz, where q is some finite collection of parameters, on which u_0 is dependent. Then we should expand u_e as

$$u_e = u_0(t; q) + \epsilon u_1(t) + \epsilon^2 u_2(t) + \dots, \quad (5)$$

where ϵ is an expansion parameter whose size will be defined shortly. The collection of parameters, q , on which the trial function is dependent, is assumed to be finite and could include an amplitude, width, position, velocity, chirp, phase, etc., depending on the problem. The variation of u_0 will then consist only of the changes that occur in u_0 when these parameters are varied. In varying these parameters, from (2), we will obtain a finite number of independent variations of S .

Let us observe what happens when we use this expansion. Substituting the expansion given by (5) into (1), and then expanding about $u = u_0$, yields

$$\begin{aligned} S &= \int_{\mathcal{D}} L[u_e] dt = \int_{\mathcal{D}} L[u_0 + \epsilon u_1 + \epsilon^2 u_2] dt + O(\epsilon^3) \\ &= \int_{\mathcal{D}} L[u_0] dt + \epsilon \int_{\mathcal{D}} \left(\frac{\delta L}{\delta u} \right)_0 (u_1 + \epsilon u_2) dt \\ &\quad + \epsilon^2 \frac{1}{2} \int_{\mathcal{D}} \int_{\mathcal{D}} \left(\frac{\delta^2 L}{\delta u(t) \delta u(t')} \right)_0 u_1(t) u_1(t') dt dt' + O(\epsilon^3), \end{aligned} \quad (6)$$

where the subscript “0” indicates evaluation at $u = u_0$. Now consider the expression above. By Hamilton's Principle, we are to find a u_e such that for small variations about u_e , we will have $\delta S = 0$. In making this variation, we can only vary the parameters q and the functions u_1, u_2 , etc.

Consider the leading order term in Eq. (6). Since the functional form of u_0 has been fixed, we can only vary the parameters q . In lowest order ($\epsilon = 0$), variations of the q 's will give the set of equations

$$\frac{\partial S}{\partial q} = \frac{\partial}{\partial q} \int_{\mathcal{D}} L[u_0] dt = 0, \quad (7)$$

which determine the lowest order solution (and structure) of the q 's. Once this is done, we have the q 's determined, which is the result of applying the VA.

Let us consider the next order, the ϵ term. By varying the function u_1 we can impose additional linearly independent variations, beyond what was imposed in zeroth order. Varying u_1 in the ϵ term will give us:

$$\left(\frac{\delta L}{\delta u} \right)_0 = 0, \quad (8)$$

which is nothing more than the equation(s) of motion evaluated at the ansatz ($u = u_0$). However, unless u_0 happens to be an exact solution, it is impossible for this condition to be satisfied. But it can be almost satisfied if u_0 is close to an exact solution, in which case the left-hand side of (8) will actually be “small” but nonzero. And since it is nonzero, it must be balanced by something. The only possible terms that could balance it are the next higher order terms, which are the ϵ^2 terms in (6). This we can naturally do if we take the size of ϵ to be determined by the order of magnitude of $(\frac{\delta L}{\delta u})_0$. Since (8) cannot generally be satisfied, we then must take

$$\left(\frac{\delta L}{\delta u(t)} \right)_0 = \epsilon R[q](t), \quad (9)$$

which defines the functional $R[q]$ and at the same time also defines ϵ . Inserting (9) into (6) then gives us an expansion in which no ϵ terms are present. Rather, the next lowest order terms are the ϵ^2 terms and upon varying u_1 in these ϵ^2 terms, one obtains

$$\int_{\mathcal{D}} \left(\frac{\delta^2 L}{\delta u(t) \delta u(t')} \right)_0 u_1(t') dt' = -R[q](t). \quad (10)$$

Note that this equation is linear in u_1 , and thus to determine u_1 , one only has to solve a linear nonhomogeneous problem.

In hindsight, this result is not surprising. Although we have used Hamilton's Principle to derive this result, one just as easily could have derived it from the Euler–Lagrange equations. From (3)–(5), upon using (9), one can easily obtain (10). One can also interpret what is occurring here, and give a simple statement of its procedure. One starts with an ansatz, u_0 , and uses the VA to calculate the best values for its parameters. Then we calculate the first-order correction to the ansatz, u_1 , based on the linearly perturbed Euler–Lagrange equations, (10). The smallness of various features of this correction will be a basis for a quantitative measure of the validity of the VA.

3. The vibrating string

The vibrating string of a finite length ℓ , fixed at both ends, has a simplicity which allows one to readily understand the approach used. If we decompose its evolution into normal modes, then we may use a VA to determine the modal shapes and the modal eigenfrequencies. The modal problem is a linear eigenvalue problem, with a well known solution. We take the Euler–Lagrange equation to be

$$\Phi[u(x)] = 0, \quad \Phi[\bullet] = \frac{d^2}{dx^2}[\bullet] + k^2[\bullet], \quad (11)$$

where $\Phi[\bullet]$ is the (linear) operator, k^2 is the eigenvalue, the interval is $[0 \leq x \leq \ell]$ and the integration variable is now changed to x . Our boundary conditions are $u(0) = 0 = u(\ell)$.

Since this problem is linear, the amplitude of the solution is arbitrary. To eliminate this irrelevant degree of freedom, we shall require the $\mathcal{L}^{(2)}$ norm of the eigenmode, $u(x)$, to be unity.

For the system described in (11), we may take the Lagrangian to be

$$L(u, u_x) = \frac{1}{2} \left(\frac{du}{dx} \right)^2 - \frac{1}{2} \nu u^2, \quad (12)$$

where $\nu = k^2$ is the eigenvalue. Note that due to translational invariance, we only need to treat the case of the fundamental mode, since the n th-order mode for a string of length ℓ , can be obtained from the fundamental mode of a string of length ℓ/n .

3.1. The variational approximation

We shall consider two different approximations for the fundamental eigenfunction of (11): a sawtooth approximation and an upside-down parabola, each of which is given by

$$u_0^{(\text{sawtooth})}(x; A, \ell) = \begin{cases} \frac{2A}{\ell}x, & 0 \leq x < \frac{\ell}{2}, \\ \frac{2A}{\ell}(\ell - x), & \frac{\ell}{2} \leq x < \ell, \end{cases} \quad (13)$$

$$u_0^{(\text{parabolic})}(x; A, \ell) = A \left[1 - \left(\frac{2x}{\ell} - 1 \right)^2 \right], \quad (14)$$

where A is the amplitude of the eigenfunction. Requiring the norm of each ansatz to be unity, gives

$$A^{(\text{sawtooth})} = \sqrt{\frac{3}{\ell}}, \quad A^{(\text{parabolic})} = \sqrt{\frac{15}{8\ell}}. \quad (15)$$

Diagrams of both ansätze are shown in Fig. 1.

As outlined in Section 2, once we have the ansatz, we then evaluate the action for that ansatz. However for eigenvalue problems, it is standard to set up the variational problem with the “action” being the eigenvalue [14,15]. Taking this approach,

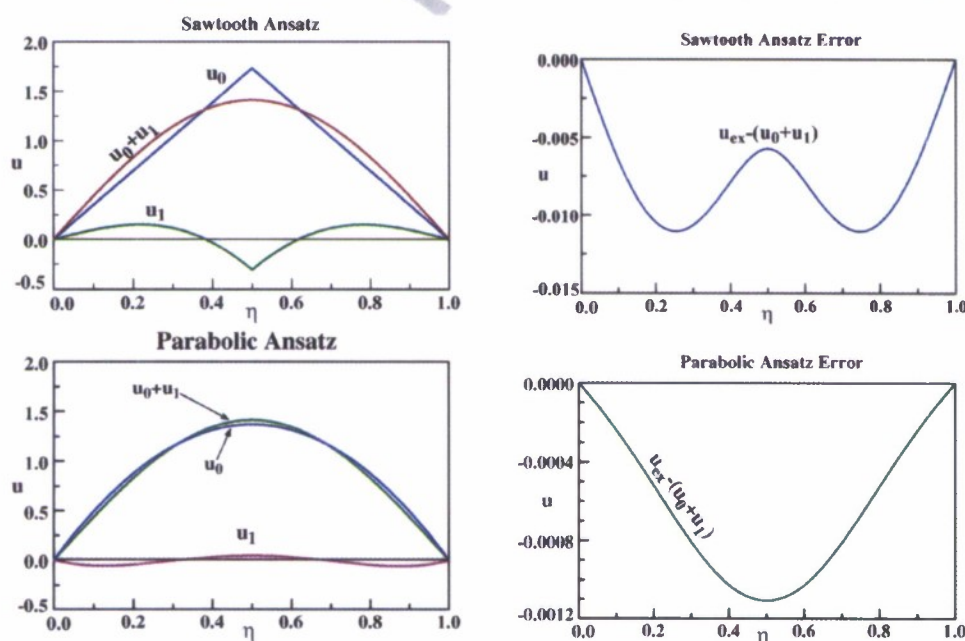


Fig. 1. The above figures depict two variational approximations for the vibrating string vs. $\eta = x/\ell$. On the left is the variational ansatz, its first-order correction, and the sum. On the right we have plotted the difference between the exact solution and the sum of variational ansatz and the first order correction.

we then vary the eigenvalue, v , where

$$v = \frac{\int_0^\ell \left(\frac{du}{dx}\right)^2 dx}{\int_0^\ell u^2 dx}. \quad (16)$$

It is trivial to verify that varying u and requiring the action to be an extremum then gives us (11). Inserting (13) and (14) into (16) respectively, we obtain

$$v_0^{(\text{sawtooth})} = \frac{12}{\ell^2} = 1.216 \dots \pi^2 / \ell^2, \quad (17)$$

$$v_0^{(\text{parabolic})} = \frac{10}{\ell^2} = 1.013 \dots \pi^2 / \ell^2. \quad (18)$$

In this case, the only parameter was the amplitude, which was set by the normalization condition. Furthermore, since (11) is linear, the amplitude drops out anyway and the action has only a fixed value and no further variation can be done. Thus, the VA for the eigenvalue for each case takes on the value shown above. As for accuracy, we see that the sawtooth ansatz is less accurate with the eigenvalue within 22% of the exact value. On the other hand, the parabolic approximation is much better giving an eigenvalue that is within 1.3% of the correct value. However, we know this only because we are comparing with the exact solution. Let us now see how we can quantify the level of accuracy without knowing the correct answer.

3.2. The correction

In solving for the first-order correction to an eigenvalue problem, there will also be an eigenvalue shift as well as a shift in the function. Rather than rederiving our result for eigenvalue problems, we may simply recognize that this is the case and that a shift is needed. Then directly using (9) and (10), we obtain:

$$\int_D \left(\frac{\delta^2 L}{\delta u(x) \delta u(x')} \right) u_1(x') dx' + R(x) + \delta v u_0(x) = 0, \quad (19)$$

where we have taken δv to be the shift in the eigenvalue and defined

$$R(x) = \frac{d^2 u_0}{dx^2} + v_0 u_0. \quad (20)$$

Since $v_0 = k^2$, then in general

$$v = k^2 + \delta v, \quad (21)$$

in which case, from (12), (19) becomes

$$\frac{d^2 u_1}{dx^2} + k^2 u_1 = U(x), \quad (22)$$

where

$$U(x) = -R(x) - \delta v u_0(x). \quad (23)$$

Eq. (19) is a linear problem, even when the analogy of (11) is nonlinear. Thus, we can always solve it with linear techniques, and in this case, the solution of (22) is

$$u_1 = -i C_+ e^{ikx} - i C_- e^{-ikx} + \int_0^x U(x') \frac{\sin(k(x-x'))}{k} dx', \quad (24)$$

where $C_\pm$ are the two constants of integration. As one may verify by integrating by parts,

$$\begin{aligned} & \int_0^x U(x') \frac{\sin(k(x-x'))}{k} dx' \\ &= \frac{du_0}{dx}(0) \frac{\sin(kx)}{k} - u_0(x) \\ & \quad - \frac{\delta v}{k} \int_0^x u_0(x') \sin(k(x-x')) dx'. \end{aligned} \quad (25)$$

Now we impose our boundary conditions and the normalization condition. At the left end, $u(0) = 0$ gives $C_+ = -C_-$. At the right end, $u(\ell) = 0$ gives

$$\begin{aligned} C_+ &= -C_- \\ &= -\frac{1}{2k} \frac{du_0}{dx}(0) \\ & \quad + \frac{\delta v}{2k \sin(k\ell)} \int_0^\ell u_0(x') \sin(k(\ell-x')) dx'. \end{aligned} \quad (26)$$

Imposing the condition that the norm be unity to order ϵ will give

$$\int_0^\ell u_0(x') u_1(x') dx' = 0, \quad (27)$$

which along with (26) will give us the value for δv .

Let us now consider our two ansätze. With the above, it is straightforward to evaluate the integrals analytically as well as numerically. We will not give the analytical expressions due to their complexity. The numerical values are

$$\begin{aligned} C_+^{(\text{sawtooth})} &= 0.0873 \dots A, \\ \delta v^{(\text{sawtooth})} &= -0.219 \dots \pi^2 / \ell^2, \end{aligned} \quad (28)$$

$$\begin{aligned} C_+^{(\text{parabolic})} &= -0.119 \dots A, \\ \delta v^{(\text{parabolic})} &= -0.013 \dots \pi^2 / \ell^2 \end{aligned} \quad (29)$$

whereby, we note that the corrections to the eigenvalues are indeed essentially what we had estimated from our knowledge of the exact solution. Thus δv does indeed give a quantitative estimate of the accuracy of these two VAs. As we had pointed out earlier, the parabolic ansatz is more accurate than the sawtooth and the above numbers give a quantitative justification of that statement.

Naturally, the corrected eigenvalues will be more accurate, and these values are

$$\begin{aligned} (v_0 + \delta v)^{(\text{sawtooth})} &= 0.99675 \dots \pi^2 / \ell^2, \\ (v_0 + \delta v)^{(\text{parabolic})} &= 0.99998 \dots \pi^2 / \ell^2, \end{aligned} \quad (30)$$

which gives an almost five 9s agreement for the corrected parabolic approximation.

The correction to an eigenvalue is just one possible quantitative estimate of VA accuracy. Another quantitative estimate is the RMS error of the ansatz. Taking the above expressions for (24) and defining the relative RMS error, E_{rms} , by

$$E_{\text{rms}}^2 = \frac{\int_0^\ell (u_1)^2 dx}{\int_0^\ell (u_0)^2 dx}. \quad (31)$$

we find that for each of our approximations,

$$E_{\text{rms}}^{(\text{sawtooth})} = 0.124\dots, \quad E_{\text{rms}}^{(\text{parabolic})} = 0.038\dots \quad (32)$$

Obviously, again the parabolic ansatz is the most accurate.

4. The Korteweg–de Vries equation

This equation is a well-known nonlinear equation [26] and we will use it to demonstrate how one can obtain estimates of the error of a VA in a nonlinear system. We shall apply this method to finding soliton (localized traveling wave) solutions of the Korteweg–de Vries (KdV) equation, which is

$$u_t + 6uu_x + u_{xxx} = 0. \quad (33)$$

Soliton solutions of the form $u(x, t) = U(\xi)$ where $\xi = x - ct$, exist where c is the wavespeed. The amplitude of such a wave, U , then satisfies the ODE

$$-cU_\xi + 6U_\xi U + U_{\xi\xi\xi} = 0. \quad (34)$$

Since these soliton solutions vanish as $\xi \rightarrow \pm\infty$, we may integrate this equation once to obtain the equation

$$-cU + 3U^2 + U_{\xi\xi} = 0. \quad (35)$$

From (35) one may construct the Lagrangian

$$L = -\frac{1}{2}cU^2 + U^3 - \frac{1}{2}(U_\xi)^2. \quad (36)$$

Let us take a Gaussian ansatz of the form

$$U_0(\xi) = A \exp(-\xi^2/\rho^2), \quad (37)$$

where U_0 has two parameters: an amplitude A and a width ρ . Given this ansatz, we may construct the action, S :

$$\begin{aligned} S &= \int_{-\infty}^{+\infty} L[u_0(\xi)] d\xi \\ &= -\frac{\sqrt{2\pi}}{4} \left(cA^2\rho - \frac{4}{\sqrt{6}}A^3\rho + \frac{A^2}{\rho} \right). \end{aligned} \quad (38)$$

Note that in this case, c is a control parameter and is not to be varied. Varying the action with respect to the trial function parameters will result in two Euler–Lagrange equations whose solution(s) will give “the closest fit” to the exact solution. These equations reduce to

$$\frac{\delta S}{\delta A} = 0 = c\rho^2 - \sqrt{6}A\rho^2 + 1, \quad (39)$$

$$\frac{\delta S}{\delta \rho} = 0 = c\rho^2 - \frac{4}{\sqrt{6}}A\rho^2 - 1, \quad (40)$$

whose solutions are

$$A = \frac{\sqrt{6}}{5}c, \quad \rho^2 = \frac{5}{c}. \quad (41)$$

4.1. First-order correction

We shall now calculate the first-order correction to the variational ansatz. As before, we take an expansion as in (5),

$$U_e = U_0(\xi; q) + \epsilon U_1(\xi) + \dots, \quad (42)$$

and insert it into (35). In the above, ϵ is a small ordering parameter, $U_0(\xi)$ is the variational ansatz and $U_1(\xi)$ is the first-order correction. Again we note that if (37) together with (41) is a reasonably accurate approximation, then we would expect the quantity

$$\epsilon R(\xi) = -cU_0 + 3U_0^2 + U_{0,\xi\xi\xi}, \quad (43)$$

to be quite small. This then should generate a reasonably small U_1 , which would be determined by the linear problem

$$-cU_1 + 6U_0U_1 + U_{1,\xi\xi\xi} = -R(\xi). \quad (44)$$

The problem of solving (44) can be simplified if we scale our coordinates and amplitudes to be unitless. Let us define a new coordinate, y , and new amplitudes, H_0 and H_1 , where

$$\xi = \rho y, \quad H_0(y) = cU_0(\xi), \quad H_1(y) = cU_1(\xi), \quad (45)$$

in which case (43) and (44) become

$$\frac{d^2 H_1}{dy^2} + (6\sqrt{6}e^{-y^2} - 5)H_1 = -R'(y), \quad (46)$$

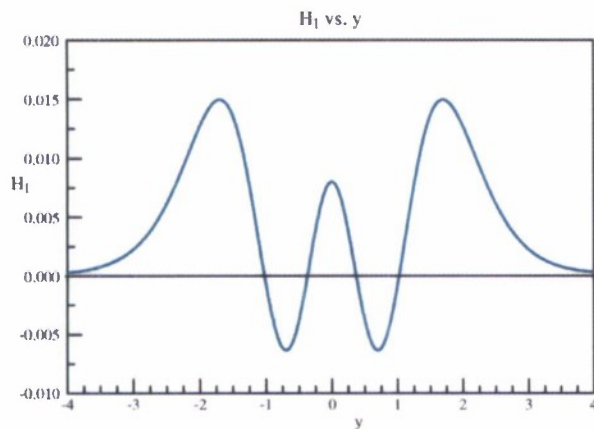
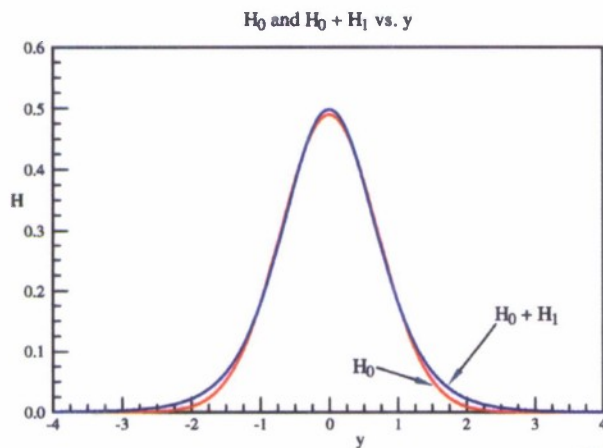
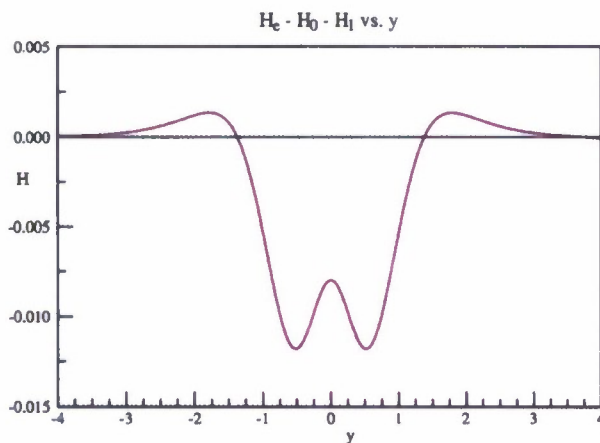
where

$$R' = \frac{\sqrt{6}}{5}e^{-y^2}[(4y^2 - 7) + 3\sqrt{6}e^{-y^2}]. \quad (47)$$

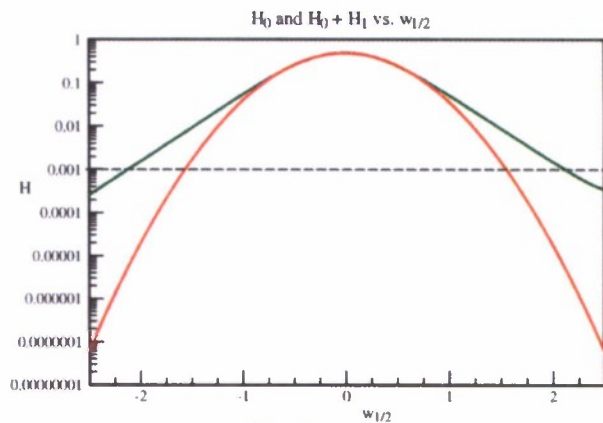
The boundary conditions on H_1 are such that $H_1 \rightarrow 0$ as $y \rightarrow \pm\infty$. There will be two homogeneous solutions of (47) and one particular solution. The two homogeneous solutions approach a linear combination of $\exp(\pm\sqrt{5}y)$ as $y \rightarrow \pm\infty$. The boundary conditions will give a unique solution for H_1 , provided that neither homogeneous solution is a bound state (which is the case).

Since this ODE does not seem to have an analytical solution, it has been solved numerically. Using the shooting method, we integrate (46) from a large negative value. We have to integrate two homogeneous solutions and one particular solution and find the appropriate linear combination at the right end such that the total solution vanishes as one approaches positive infinity. The result is the unique, bounded solution shown in Fig. 2.

In Fig. 3, we show the ansatz H_0 compared to the sum $H_0 + H_1$. As one can see, the first-order correction mainly corrects for the more rapid decay of the Gaussian at large y . In Fig. 4, we show the difference between the exact solution, H_e , and $H_0 + H_1$, which will indicate what the second-order correction would be. Here we note that the largest remaining correction is in the central region of the pulse. [Since the KdV

Fig. 2. The first-order correction, $H_1(y)$, to the Gaussian ansatz.Fig. 3. $H_0(y)$ and $H_0(y) + H_1(y)$.Fig. 4. The exact solution, $H_e(y)$, minus the sum of the ansatz and the first-order correction, $H_0(y) + H_1(y)$.

equation is nonlinear, after the first-order correction we would have a nonzero second-order correction, which would therefore be largest wherever the amplitude was largest. Here we see that this is indeed the case.]

Fig. 5. A $\log(H)$ plot showing the amplitudes of the tails vs. the width at half-height, with H_0 being the lower curve and with $H_0(y) + H_1(y)$ being the upper curve. The dashed line is an example—see the text.

As already mentioned, there are several ways in which one could quantitatively measure the accuracy of an ansatz, depending on the quantity in which one was interested. We shall consider two possibilities. First, the standard rms error (31) can again be calculated. This would be appropriate as a measure of the global error in the pulse shape. Here we find that $E_{\text{rms}} = 0.0380 \dots$, which indicates that this is quite good for a global approximation. However, if one were interested in calculating soliton–soliton interactions wherein the overlap of the tails were important, then the situation would be different. So the quantitative measure of a VA will depend on the application to which it will be applied.

Let us now be definite and calculate a quantitative measure for this case. For the Gaussian approximation, we have from Fig. 3 that the width at half-height is $w_{1/2} = 1.594 \dots$. If one were to plot the tail amplitude versus this width, as in Fig. 5, one would readily see that the Gaussian ansatz is an inferior approximation for the tails. For example, if one chose to have the solitons separated such that the tail amplitude in the region of the next soliton would be no more than 0.001, the dashed line in Fig. 5 shows that a soliton–soliton separation of only 1.56 widths would be sufficient. However if we add the first-order correction to the ansatz, $H_0 + H_1$, which is also shown in Fig. 5, then we would come to a different conclusion. The curve $H_0 + H_1$ shows that the ansatz plus correction, at a width of 1.56, would have an amplitude of ≈ 0.008 , which is eight times too large. Looking at where the separation should actually be, we see that the tails of $H_0 + H_1$ do not decay to 0.001 until at least 2.1 widths. Thus since the ratio of these widths is $2.1/1.59 = 1.32$, a measure of the VA accuracy in this case would be $\pm 32\%$, considerably more than the 3.8% indicated by the rms error.

5. Conclusion

In this Letter we have established a method for quantifying the accuracy of variational approximations which does not require an exact solution of the governing equation(s) of mo-

tion. The method possesses an inherent simplicity in that the measure of accuracy is obtained via solving an inhomogeneous linear equation.

We have also pointed out that the quantitative measure of any variational approximation must be based on the intended use of the solution. Depending on the use, the VA could either be quite acceptable or could even be unacceptable. The smallness of the correction, or the smallness of its deviation from some important property of the exact solution, should be the quantitative measure of the validity of any VA.

This work opens up additional approaches to the utilization of variational approximations as a modeling technique for physical and mathematical systems. There is no longer the necessity for having exact (or close to exact) solutions a priori in order to determine the validity of a variational approximation. The advantage of this for nonlinear problems is that to obtain an estimate of the correction, one only needs to solve a linear problem, not the fully nonlinear problem. The steps in this method for quantifying the quality of a VA consists of choosing an analytical model for the ansatz which will be close to that of the exact solution, obtain a VA from that ansatz and use the technique outlined here to determine the accuracy of those results. This latter is accomplished by estimating the first-order correction to the VA. This first-order correction may then be used to also improve upon the predictions of the VA. Naturally, one could develop an iterative process from this for improving the accuracy of a variational technique to higher orders.

Note added in proof

It needs to be mentioned that an earlier method to estimate the validity of a variational approximation had been given in Ref. [28], where Dexais, Anderson and Lasik investigated how well a variational solution would preserve the next higher order invariant beyond the Hamiltonian.

Acknowledgements

This research was supported in part by the National Science Foundation Grant #DMS-0505566 and the Air Force Office of Scientific Research.

References

- [1] W.R. Hamilton, On a General Method in Dynamics, *Philosophical Transactions of the Royal Society of London*, part II, 1834, pp. 247–308.
- [2] W.R. Hamilton, Second Essay on a General Method in Dynamics, *Philosophical Transactions of the Royal Society of London*, part I, 1835, pp. 95–144.
- [3] Homer, *Aeneid*, R. Fitzgerald (translator), reissue ed., Vintage, New York, 1996.
- [4] D.S. Lemons, *Perfect Form*, Princeton Univ. Press, Princeton, 1997.
- [5] I. Newton, *Principia*, A. Mott (translator), Prometheus Books, UK, 1995.
- [6] R.K. Nesbet, *Variational Principles and Methods in Theoretical Physics and Chemistry*, IBM Almaden Research Center, Cambridge Univ. Press, Cambridge, 2003.
- [7] W.W. Rouse Ball, *A Short Account of the History of Mathematics*, Martino, 2004.
- [8] <http://www-history.mcs.st-andrews.ac.uk/Biographies/Lagrange.html>.
- [9] B.A. Malomed, *Prog. Opt.* 43 (2002) 71.
- [10] J. Steiner, *Gesammelte Werke*, Prussian Academy of Sciences, 1881.
- [11] C. Lubich, *Math. Comput.* 74 (2005) 765.
- [12] G. Rosen, *Formulations of Classical and Quantum Dynamical Theory*, Academic Press, New York, 1969.
- [13] I. Christov, *Wavelet-Galerkin Methods for Partial Differential Equations*, Massachusetts Institute of Technology Matrix Analysis and Wavelets REU, TAMU, 2003.
- [14] G.B. Arfken, H.J. Weber, *Mathematical Methods for Physicists*, fourth ed., Academic Press, New York, 1995.
- [15] P.M. Morse, H. Feshbach, *Methods of Theoretical Physics*, part II, McGraw-Hill, New York, 1953, Chapter 9.
- [16] M. Chen, D.J. Kaup, B. Malomed, *Phys. Rev. E* 69 (2004) 056605.
- [17] D.J. Kaup, B.A. Malomed, *Physica D* 87 (1995) 155.
- [18] D.J. Kaup, G.E. Thomas, *J. Plasma Phys.* 57 (1997) 765.
- [19] D.J. Kaup, B.A. Malomed, *Physica D* 84 (1995) 319.
- [20] D.J. Kaup, *Math. Comput. Simul.* 69 (2005) 322.
- [21] B.A. Malomed, M. Weinstein, *Phys. Lett. A* 220 (1996) 91.
- [22] D.J. Kaup, T.I. Lakoba, *J. Math. Phys.* 37 (1996) 3442.
- [23] J.A.D. Wattis, *Nonlinearity* 9 (1996) 1583.
- [24] B.A. Malomed, R.S. Tasgal, *Phys. Rev. E* 49 (1994) 5787.
- [25] A. Bondeson, M. Lisak, D. Anderson, *Phys. Scr.* 20 (1979) 479.
- [26] C.S. Gardner, J.M. Greene, M.D. Kruskal, R.M. Miura, *Phys. Rev. Lett.* 19 (1967) 1095.
- [27] I.V. Barashenkov, D.E. Pelinovsky, E.V. Zemlyanaya, *Phys. Rev. Lett.* 80 (1998) 5117.
- [28] M. Dexais, D. Anderson, M. Lasik, *Phys. Rev. A* 40 (1989) 2441.

HOW MANY TYPES OF SOLITON SOLUTIONS DO WE KNOW?

VLADIMIR S. GERDJKOV¹ AND DAVID KAUP²

¹*Institute for Nuclear Research and Nuclear Energy,
Bulgarian Academy of Sciences,*

72 Tsarigradsko chaussee, 1784 Sofia, BULGARIA

²*Department of Mathematics, University of Central Florida
Orlando, Florida 32816-1364, USA*

Abstract. We discuss several ways of how one could classify the various types of soliton solutions related to NLEE solvable with the generalized $n \times n$ Zakharov-Shabat system. In doing that we make use of the fundamental analytic solutions, the dressing procedure and other tools characteristic for the inverse scattering method. We propose to relate to each subalgebra $sl(p)$, $2 \leq p \leq n$ of $sl(n)$, a type of one-soliton solutions which have $p-1$ internal degrees of freedom.

1. Introduction

It is our impression that the question in the title has not been answered satisfactorily even for some of the best known type of soliton equations such as the N -wave equations, the multicomponent NLS equations and others.

We are using the term 'soliton solution' as a special solution to a given nonlinear evolution equation (NLEE) which is solvable by the inverse scattering method [20, 4]. That means that the NLEE allows Lax representation:

$$[L(\lambda), M(\lambda)] = 0, \quad (1)$$

where $L(\lambda)$ and $M(\lambda)$ are two linear operators. In what follows we take them to be first order matrix differential operators:

$$L\psi(x, t, \lambda) \equiv i \frac{d\psi}{dx} + U(x, t, \lambda)\psi(x, t, \lambda) = 0, \quad (2)$$

$$M\psi(x, t, \lambda) \equiv i \frac{d\psi}{dt} + V(x, t, \lambda)\psi(x, t, \lambda) = \psi(x, t, \lambda)C(\lambda). \quad (3)$$

The compatibility condition (1) which must hold true identically with respect to λ takes the form:

$$i \frac{\partial V}{\partial x} - i \frac{\partial U}{\partial t} + [U(x, t, \lambda), V(x, t, \lambda)] = 0 \quad (4)$$

and is valid for any choice of $C(\lambda)$.

The one-soliton solutions are related to one or a set of several discrete eigenvalues of the Lax operator L . Therefore one first has to study the different configurations of discrete eigenvalues of L , see [12]. The next step in classifying the types of one-soliton solutions is related to the study of their internal degrees of freedom.

In order to make the problem reasonably difficult we will specify L to be the generalized Zakharov-Shabat system:

$$L(\lambda)\psi(x, \lambda) \equiv i \frac{d\psi}{dx} + (q(x) - \lambda J)\psi(x, \lambda) = 0. \quad (5)$$

where we consider the potential $q(x, t)$ to be $n \times n$ matrix-valued which is smooth function of x tending to zero fast enough for $x \rightarrow \pm\infty$. Another important step is fixing up J to be real constant diagonal matrix with different eigenvalues. Without restrictions we can assume that:

$$J = \text{diag}(J_1, \dots, J_n), \quad J_1 > J_2 > \dots > J_n, \quad \text{tr } J = 0. \quad (6)$$

Using a proper gauge transformation commuting with J we can always take $q(x)$ to be of the form $q(x) = [J, q'(x)]$, i.e. $q_{ij} \equiv 0$. The linear subspace in $sl(n)$ of matrix-valued functions $q(x) = [J, q'(x)]$ are known in the literature to be the co-adjoint orbit of $\mathfrak{g}$ passing through J . The co-adjoint orbits can be supplied in a natural way with non-degenerate symplectic structure which makes them natural choices for the phase spaces $\mathcal{M}_J$ and Hamiltonian structures of the corresponding NLEE.

We will try to answer the question in the title first for the simplest class of Lax operators of the type (5) with real-valued J . In doing this we will be using the dressing method, one of the best known methods for constructing reflection-less potentials and soliton solutions. The choice of $\mathcal{M}_J$ determines the number of independent matrix elements in $q(x)$ that will satisfy the NLEE.

In Section 2 below we first outline the well known facts about the soliton types of NLEE solvable by the $sl(2)$ Zakharov-Shabat system. In the next Section 3 we treat the different one-soliton solutions for the $sl(n)$ Zakharov-Shabat systems related to the subalgebras $sl(p)$. Most of our results are illustrated for the $sl(5)$ system, but it is not difficult to extend them to any $sl(n)$ system. The structure of the eigenfunctions of $L(\lambda)$ corresponding to the different types of solitons is outlined in Section 4. In the last Section we discuss possible generalizations to other Zakharov-Shabat systems having additional symmetry properties.

2. Zakharov-Shabat system and $sl(2)$ solitons

To each choice of the Lax operator $L(\lambda)$, i.e. for each specific $U(x, t, \lambda)$ one can relate a class of NLEE. The best known examples of such NLEE are related to the Zakharov-Shabat system $L_0(\lambda)$ with

$$U_0(x, t, \lambda) = q_0(x, t) - \lambda \sigma_3, \quad q_0(x, t) = \begin{pmatrix} 0 & q_0^+ \\ q_0^- & 0 \end{pmatrix}. \quad (7)$$

The class of NLEE related to L_0 are systems of equations for the functions $q_0^\pm(x, t)$ are written in the compact form [2, 17, 10]:

$$i\sigma_3 \frac{\partial q_0}{\partial t} + 2f(\Lambda_0)q_0(x, t) = 0, \quad (8)$$

where $f(\lambda)$ is the dispersion law of the NLEE and Λ_0 is (anyone of) the recursion operators, acting on the space $\mathcal{M}_0$ of off-diagonal matrix-valued functions as follows:

$$\Lambda_{0,\pm} X \equiv \frac{i}{4} \left[\sigma_3, \frac{dX}{dx} \right] + \frac{i}{2} q(x) \int_{\pm\infty}^x dy r(q(y), [\sigma_3, X(y)]). \quad (9)$$

Choosing $f(\lambda)$ to be linear function we get simple linear system of equations for $q_0^\pm(x, t)$. For $f(\lambda) = -2\lambda^2$ the NLEE (8) reduces to the system:

$$\begin{aligned} iq_{0,t}^+ + q_{0,xx}^+ + 2(q_0^+(x, t))^2 q_0^-(x, t) &= 0, \\ iq_{0,t}^- - q_{0,xx}^- - 2(q_0^-(x, t))^2 q_0^+(x, t) &= 0; \end{aligned} \quad (10)$$

for $f(\lambda) = 4\lambda^3$, one gets the system:

$$\begin{aligned} q_{0,t}^+ + q_{0,xxx}^+ + 6q_0^+(x, t)q_0^-(x, t)q_{0,x}^+ &= 0, \\ q_{0,t}^- + q_{0,xxx}^- + 6q_0^-(x, t)q_0^+(x, t)q_{0,x}^- &= 0. \end{aligned} \quad (11)$$

The main idea of the ISM is based on the possibility to analyze and solve the direct and the inverse scattering problems for $L_0(\lambda)$. More precisely we introduce the Jost solutions of $L_0(\lambda)$ as 2×2 matrix-valued solutions:

$$L_0(\lambda) \psi_0(x, t, \lambda) = 0, \quad L_0(\lambda) \phi_0(x, t, \lambda) = 0, \quad (12)$$

defined by their asymptotic behavior for $x \rightarrow \pm\infty$ respectively:

$$\begin{aligned} \psi_0(x, t, \lambda) &= |\psi_0^+, \psi_0^-|, & \lim_{x \rightarrow \infty} \psi_0(x, t, \lambda) e^{i\lambda\sigma_3 x} &= \mathbb{1}, \\ \phi_0(x, t, \lambda) &= |\phi_0^+, \phi_0^-|, & \lim_{x \rightarrow -\infty} \phi_0(x, t, \lambda) e^{i\lambda\sigma_3 x} &= \mathbb{1}, \end{aligned} \quad (13)$$

The superscripts $\pm$ in the columns of the Jost solutions refer, as we shall see below, to their analyticity properties.

The definitions (13) hold true for all t if we fix up appropriately $C_0(\lambda)$ as follows:

$$C_0(\lambda) \equiv \lim_{x \rightarrow \pm\infty} V(x, t, \lambda) = f_0(\lambda) \sigma_3. \quad (14)$$

Then one introduces the scattering matrix $T_0(\lambda, t)$ by:

$$T_0(\lambda, t) \equiv (\psi(x, t, \lambda))^{-1} \phi(x, t, \lambda) = \begin{pmatrix} a_0^+(\lambda) & -b_0^-(\lambda, t) \\ b_0^+(\lambda, t) & a_0^-(\lambda) \end{pmatrix}, \quad (15)$$

which is x -independent. Its t -dependence can be derived from the Lax representation to be:

$$i \frac{dT_0}{dt} + [f_0(\lambda) \sigma_3, T(\lambda, t)] = 0, \quad (16)$$

Thus, if $q_0^\pm(x, t)$ satisfy the system of equations (8) we get:

$$\frac{da_0^\pm(\lambda)}{dt} = 0, \quad i \frac{db_0^\pm(\lambda)}{dt} \mp 2f_0(\lambda) b_0^\pm(\lambda) = 0, \quad (17)$$

The matrix elements of $T_0(\lambda, t)$ are not independent. They satisfy the 'unitarity' condition $\det T_0(\lambda) \equiv a_0^+ a_0^- + b_0^+ b_0^- = 1$. Besides the diagonal elements a_0^+ and a_0^- allow analytic extension with respect to λ in the upper and lower complex λ -plane respectively. In fact the minimal set of scattering data which uniquely determines both the scattering matrix and the corresponding potential $q_0(x)$ consists of two types of variables: i) the reflection coefficients $\rho_0^\pm(\lambda) = b_0^\pm/a_0^\pm$ defined for real $\lambda \in \mathbb{R}$ and ii) a discrete set of scattering data including the discrete eigenvalues $\lambda_k^\pm \in \mathbb{C}_\pm$ and the constants $C_k^\pm$ which determine the norm of the corresponding Jost solutions [14].

Along with the Jost solutions we can introduce the fundamental analytic solutions (FAS) as follows:

$$\chi_0^+(x, t, \lambda) = |\phi_0^+, \psi_0^+|, \quad \chi_0^-(x, t, \lambda) = |\psi_0^-, \phi_0^-|. \quad (18)$$

Indeed, one can prove [22, 2] that $\chi_0^\pm(x, \lambda) e^{i\lambda\sigma_3 x}$ are analytic functions of λ for $\lambda \in \mathbb{C}_\pm$.

The functions $a_0^\pm(\lambda) = \det \chi^\pm(x, \lambda)$ are known as the Evans functions [22, 3] of the system $L_0(\lambda)$. Their importance comes from the fact that they are t -independent (see eq. (17)), and therefore they (or rather $\ln a_0^\pm$) can be viewed as generating functionals of the (local) integrals of motion. In addition it is known that their zeroes determine the discrete eigenvalues of $L_0(\lambda)$:

$$a_0^+(\lambda_{0,k}^+) = 0, \quad \lambda_{0,k}^+ \in \mathbb{C}_+; \quad a_0^-(\lambda_{0,k}^-) = 0, \quad \lambda_{0,k}^- \in \mathbb{C}_-. \quad (19)$$

ERROR: typecheck
OFFENDING COMMAND: restore

STACK:

0
1
110
63
-63
69
209
69
278
-69
-savelevel-