ZERO DISPERSION LIMITS OF INTEGRABLE EQUATIONS

Spyridon Kamvissis

The present is an extended version of a Seminar given at the Max-Planck Institute on February 28, 2002.

I. THE PROBLEM

We are interested in limits of integrable equations as the "dispersion" term goes to zero. More precisely, we will be concerned with the solutions of the following model problems for small values of the parameter ϵ.

1. The small dispersion KdV problem

$$u_t - 6uu_x + \epsilon^2 u_{xxx} = 0, \tag{1}$$

and

2. The "semiclassical" limit of the Nonlinear Schrödinger Equation (focusing (+) and defocusing (-) cases)

$$i\epsilon u_t + \frac{\epsilon^2}{2} u_{xx} \pm |u|^2 u = 0. \tag{2}$$

For convenience we will consider initial value problems. For KdV, we take initial data

$$u(x,0) = u_0(x) \tag{1a}$$

that belong, say, to the Schwartz class. For the NLS equations, we take

$$u_0(x) = A(x) exp(\frac{iS(x)}{\epsilon}), \tag{2a}$$

where $A(x)$ may again taken to be a Schwartz function, while $S(x)$ is a smooth function converging rapidly to limits $S_{\pm}$, as $x \to \pm\infty$, respectively.

II. HISTORY OF THE PROBLEM

The study of zero dispersion limits, as above, was initiated by Lax and Levermore around 1980. Lax was motivated by a problem posed by John von Neumann in the 1950s. Von Neumann wanted to use the first powerful computers to calculate solutions of the hyperbolic system for gas dynamics. For this, he used a numerical scheme, discrete in time and space. The numerical experiments showed the following:

(i) For small times, the numerical solution seemed to converge, as the space-grid length and the time-step tended to zero.

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX

(ii) For larger times, fast oscillations appeared. The frequency of the oscillations was of order equal to the inverse of the space-grid length used, while the amplitude of the oscillations was of order 1. Thus, the scheme did not converge, at least in the strong sense.

Von Neumann conjectured that (i) the scheme converged in a weak sense (i.e. after averaging the oscillations) and that (ii) it converged to the solution of the PDE system.

It turns out that, at least for special cases, i.e. special laws governing the relation of pressure and volume, the numerical scheme is completely integrable. The complete integrability of the scheme can be used to study its solution as the grid length goes to zero. In such integrable cases, one can indeed prove that von Neumann's first conjecture is correct, but his second conjecture is wrong.

Ths study of such integrable numerical schemes is analogous to the study of integrable equations in the dispersive limit, such as (1) and (2), to which we now turn our attention.

Problem (1) was considered in a series of three papers [LL] in 1983, under the assumption that the data (1a) are negative, single-minimum, rapidly decaying at infinity. Venakides eventually dealt with the positive data case, and also with the periodic case. A review article surveying these results, but also the related problems for discrete integrable chains (and hence numerical schemes) can be found in [LLV].

In all cases studied it was found that as long as the solution of the formal limiting equation

$$\begin{aligned} u_t + u u_x &= 0, \\ u(x,0) &= u_0(x) \end{aligned} \tag{3}$$

exists and is smooth, then the limit of the solution of (1) exists (in the strong sense) and indeed solves equation (3). After the time where the solution of (3) suffers a shock, the limit of (1) is only guaranteed to exist in a weak sense, and it does not (in general) satisfy eqution (3). The solution of (1) develops high frequency oscillations in a sub-area of the (x,t)-plane; its weak limit can be described in terms of the solution of a hyperbolic system of three equations (known as the Whitham system). This system can also suffer a shock, in which case the weak limit of (1) is now described in terms of the solution of a new five equation hyperbolic Whitham system, and so on.

III. NUMERICAL RESULTS FOR NLS

It has long been known that semiclassical Schrödinger equations can be viewed as singular perturbations of compressible Euler-type fluid dynamical systems in the density and momentum variables. For example, setting $\rho = |u|^2$ and $\mu = Im(\epsilon u^* u_x)$, the focusing unperturbed problem (2) can be written as

$$\begin{aligned} \rho_t + \mu_x &= 0, \\ \mu_t + (\frac{\mu^2}{\rho} \mp \frac{\rho^2}{2})_x &= \frac{\epsilon^2}{4}(\rho_x^2 \log\rho)_{xx}, \end{aligned} \tag{4}$$

with initial data $\rho(x,0) = (A(x))^2$ and $\mu(x,0) = 0$. The formal limit of (2) when ϵ goes to 0 is an Euler system of compressible fluid dynamics, which in this particular case is elliptic.

$$\begin{aligned} \rho_t + \mu_x &= 0, \\ \mu_t + (\frac{\mu^2}{\rho} \mp \frac{\rho^2}{2})_x &= 0. \end{aligned} \tag{4a}$$

If the initial data is analytic, the Cauchy-Kovalevskaya theorem guarantees the existence of a unique solution for small times. In general, this is not true.

As in the case of KdV, numerical results show that the formal limit (4a) does govern the strong limit of (4), as long as (4a) has a smooth solution.

Now, we have to distinguish between the focusing and defocusing cases, however. In the defocusing case (sign "-" in (2), "+" in (4)), the formal limit (4a) is hyperbolic. Such a case is very similar to (1). After the shock time, the weak limits of ρ, μ exist and can be expressed in terms of the solution of a hyperbolic Whitham system. The analysis of the semiclassical defocusing NLS problem was done in the 90s by Jin, Levermore and D.McLaughlin (see [JLM]). It is in fact very similar to the analysis of the KdV problem.

In the focusing case (sign "+" in (2), "-" in (4)) the limiting Euler system is elliptic. It too suffers a singularity at some time. After "break" there is an elliptic Whitham system which describes the semiclassical limit. This problem was analyzed in [KMM]. It was shown that in the case of analytic data, the Cauchy-Kovalevskaya solution of system (4a) gives the strong limit of the semiclassical problem, for small times. For larger times, one observes violent oscillations, similar to the ones observed in the defocusing case, but harder to analyze rigorously. The rigorous theory was provided and a connection was made with an unstable variational problem for electrostatic Green potentials over contours in the full complex plane. Even though the formal limit (4a) does not yield the semiclassical limit of (2) for all time, there is a whole hierarchy of Whitham equations, which governs the weak semiclassical limit, and which is also elliptic. It is vital to emphasize that the most important effect of the ellipticity of the formal limit of the Euler-Whitham system is instability. Even though for fixed ϵ, equation (2) is well-posed, the formal limit is an ill-posed unstable initial value problem for an elliptic system (or more precisely, a hierarchy of systems).

We will return to the focusing/defocusing dichotomy in a later section. At this point we just note that [KMM] is in a sense the culmination of Lax's program on the study of zero dispersion completely integrable equations, because it extends the existing theory to the case of non-self-adjoint Lax operators.

IV. INTEGRABILITY

Semiclassical or zero dispersion problems can be posed for non-integrable equations as well. For example, we can change the nonlinear term in (2), or we can consider a higher-dimensional NLS problem. Numerically, one observes similar behavior. Before the "shock" general WKB-type methods have been used by Gérard and Grenier to show that (at least in some cases) the formal limit of NLS (or more precisely the equivalent Euler system) governs the semiclassical (strong) limit of the solution to (2) (or its higher dimension version). After the "shock" no rigorous result is known.

Integrability machinery is based on the scattering and inverse scattering theory for equations like (1) and (2). The Lax operator associated to (1) is the Schrödinger operator

$$-\epsilon^2 \frac{d^2}{dx^2} + u(x) \tag{5}$$

while the NLS equations are associated to the Dirac operators

$$\begin{pmatrix} i\epsilon\partial_x & \mp iu(x) \\ -iu^*(x) & -i\epsilon\partial_x \end{pmatrix}, \tag{6}$$

where sign "+" corresponds to the defocusing case, while sign "-" corresponds to the focusing case. * denotes complex conjugation. It is important to note that, although (5) gives a self-adjoint operator (here u is always real), (6) is self-adjoint in the defocusing case and non-self-adjoint in the focusing case.

V. THE DIRECT SCATTERING PROBLEM

For the Schrödinger operator (5): assume for simplicity that the initial data $u_0(x)$ is a strictly negative function with a single minimum, say at $x = 0$. Assume without loss of generality, $u(0) = -1$. Then,

(i) the continuous spectrum consists of $[0, \infty)$

(ii) the discrete spectrum consists of a finite number of negative eigenvalues; their number is of order $1/\epsilon$ as $\epsilon \to 0$ and they accumulate in the interval $[-1, 0)$. The density of eigenvalues is given explicitly in terms of $u_0(x)$, via an integral formula (see [LL]).

The reflection coefficient is exponentially small, away from $k = 0$.

If the initial data is positive, then no eigenvalues exist. On the other hand the reflection coefficient is no more trivial. It is given by a simple integral formula in terms of the initial data (as $\epsilon \to 0$).

The above results are asymptotic, and can be derived by WKB-techniques.

For the self-adjoint Dirac operator the analysis is similar. No eigenvalues exist and the reflection coefficient is oscillatory; it is given by an asymptotic formula involving the initial data.

In the focusing case, results only exist in the case where $S = 0$, i.e. the initial data is real, independent of ϵ. Then WKB theory works and one has:

(i) The continuous spectrum consists of the real line. The reflection coefficient is exponentially small (at least away from $k = 0$, k being the spectral variable).

(ii) The eigenvalues accumulate near a subinterval of the imaginary axis. If, for simplicity, $A(x)$ is a bell-shaped function with a single maximum at $x = 0$, and say $A(0) = A$, then the asymptotic density of eigenvalues is

$$\begin{aligned} d\mu^{WKB}(\eta) = \rho^0(\eta)\chi_{[0,iA]}d\eta + (\rho^0(\eta))^*\chi_{[-iA,0]}d\eta, \\ for\ \eta \in [-iA, iA], \end{aligned}$$

where χ_I is the characteristic function of the complex interval I, $d\eta$ is the Lebesgue measure and

$$\rho^0(\eta) = \frac{\eta}{\pi}\int_{x_-(\eta)}^{x_+(\eta)} \frac{dx}{(A(x)^2 + \eta^2)^{1/2}},$$

and $x_\pm(\eta)$ are the positive and negative solutions of $A(x)^2 + \eta^2 = 0$ respectively.

VI. RIGOROUS ANALYSIS OF THE SELF-ADJOINT PROBLEM

We consider the case where the potential is strictly negative. In such a case the reflection coefficient is negligible, so we can take it to be zero.[1] For any ϵ then, the solution of (1) is a pure soliton solution and can be easily written in terms of the following determinant formula.

$$\begin{aligned} u(x, t; \epsilon) &= -2\epsilon^2\partial_{xx} log\tau(x, t; \epsilon), \\ \tau(x, t; \epsilon) &= det(I + G), \end{aligned} \tag{7}$$

where G is a matrix of dimension $O(\epsilon^{-1})$. It turns out that τ can be expanded as a finite series in powers of ϵ where each coefficient is a real exponential. It thus follows

[1] At least away from $k = 0$. The rigorous justification can be found in [LL].

easily that the series can be approximated by a dominant term. Furthermore, this dominat term can be conveniently expressed in terms of a solution of a variational problem for a logarithmic functional on the real line.

One has

$$\tau(x,t;\epsilon) \sim \frac{\epsilon^{some\ power}}{2} exp(\frac{-Q}{2\epsilon^2}),$$

where the actual exponent of ϵ is immaterial in view of (7) and Q minimizes the functional

$$Q_\psi = \frac{4}{\pi}\int_0^1 a(\eta,x,t)\psi(\eta)d\eta - \frac{1}{\pi^2}\int_0^1\int_0^1 log(\frac{\eta-\mu}{\eta+\mu})^2\psi(\mu)\psi(\eta)d\mu d\eta \tag{8}$$

over $\psi \in L^1[0,1]$ such that $0 \le \psi(\eta) \le \phi(\eta)$, where here a and ϕ depend on the initial data via simple formulae.

The hard analysis of the work of Lax and Levermore is concerned with the solution of the above variational problem.

VII. RIEMANN-HILBERT PROBLEMS

The method of Lax and Levermore will not work for non-self-adjoint problems. Consider for example, the semiclassical focusing NLS equation (2). One can still write the solution in terms of a formula like (7), for any given ϵ. However, the expansion of determinant τ involves imaginary exponentials. As a result, it is not approximated by a dominant term. One needs completely different techniques.

The most powerful asymptotic technique for integrable systems relies on the use of Riemann-Hilbert problems. The solution of an integrable system via the inverse scattering transform can always be reduced to the solution of a Riemann-Hilbert factorization problem. The asymptotic analysis of the given equation is thus reduced to the asymptotic analysis of a Riemann-Hilbert problem.

Consider, for example, the nonlinear Schrödinger equation (defocusing case), that is equation (2) with the "-" sign. Assume $\epsilon = 1$ for the moment.

The associated linear system is

$$\psi_x = \begin{pmatrix} i\lambda & -iu(x) \\ iu^*(x) & -i\lambda \end{pmatrix} \psi, \tag{9}$$

where * denotes conjugation. λ is the spectral variable. Jost functions $\psi_1, \psi_2, \phi_1, \phi_2$ are defined on the real line as (column vector) solutions of (9) satisfying the asymptotic conditions

$$\begin{aligned} \psi_1(x,\lambda) &= \begin{pmatrix} e^{i\lambda x} \\ 0 \end{pmatrix} + o(1),\ as\ x \to +\infty, \\ \psi_2(x,\lambda) &= \begin{pmatrix} 0 \\ e^{-i\lambda x} \end{pmatrix} + o(1),\ as\ x \to +\infty, \\ \phi_1(x,\lambda) &= \begin{pmatrix} e^{i\lambda x} \\ 0 \end{pmatrix} + o(1),\ as\ x \to -\infty, \\ \phi_2(x,\lambda) &= \begin{pmatrix} 0 \\ e^{-i\lambda x} \end{pmatrix} + o(1),\ as\ x \to -\infty. \end{aligned} \tag{10}$$

Furthermore, ψ_1 and ϕ_2 can be meromorphically extended to the upper half-plane, while ψ_2 and ϕ_1 can be meromorphically extended to the lower half-plane. Indeed, one has

$$\begin{aligned} (\ \psi_1(x,\lambda)e^{-i\lambda x} \quad \phi_2(x,\lambda)e^{i\lambda x}\ ) &\to I,\ as\ \lambda \to \infty,\ Im\lambda \ge 0, \\ (\ \phi_1(x,\lambda)e^{-i\lambda x} \quad \psi_2(x,\lambda)e^{i\lambda x}\ ) &\to I,\ as\ \lambda \to \infty,\ Im\lambda < 0. \end{aligned} \tag{10}$$

I is here the identity matrix.

We point out the symmetry

$$\psi^*(x,\lambda^*) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \psi(x,\lambda), \tag{11}$$

where $\psi = (\ \psi_1 \ \ \psi_2\ )$; the same symmetry is satisfied by $\phi = (\ \phi_1 \ \ \phi_2\ )$.

On the real line, ϕ and ψ are solutions of system (9). Hence there exist 'scattering coefficients' $a(\lambda), b(\lambda)$ such that

$$\begin{aligned} \phi_2(x,\lambda) &= a(\lambda)\psi_2(x,\lambda) + b(\lambda)\psi_1(x,\lambda) \\ \phi_1(x,\lambda) &= \bar{a}(\lambda)\psi_1(x,\lambda) - \bar{b}(\lambda)\psi_2(x,\lambda). \end{aligned} \tag{12}$$

Although $\bar{a}$ and $\bar{b}$ are a priori independent of a and b, one can see from the symmetries above that they are after all their conjugates. Furthermore one can show

$$|a(\lambda)|^2 - |b(\lambda)|^2 = 1. \tag{13}$$

The time evolution of the scattering coefficients is given by

$$\begin{aligned} a(\lambda,t) &= a(\lambda,0), \\ b(\lambda,t) &= b(\lambda,0)e^{4i\lambda^2 t}. \end{aligned} \tag{14}$$

It turns out that $a(\lambda)$ can be analytically extended to the upper half-plane, while $\bar{a}(\lambda)$ can be analytically extended to a function $a^*(\lambda) = \bar{a}(\lambda^*)$ in the lower half-plane. In general, this is not also true for b. It follows from (13) that a has no zeros.

We define the 2×2-matrix-valued function Ψ as follows. Let

$$\begin{aligned} \Psi &= \begin{pmatrix} \psi_1 e^{-i\lambda x} & \frac{\phi_2 e^{i\lambda x}}{a(\lambda)} \end{pmatrix}, \ for\ Im\lambda > 0, \\ &= \begin{pmatrix} \frac{\phi_1 e^{-i\lambda x}}{a^*(\lambda)} & \psi_2 e^{i\lambda x} \end{pmatrix}, \ for\ Im\lambda < 0. \end{aligned} \tag{15}$$

Letting Ψ_+ and Ψ_- denote the limits of Ψ on the real line from above and below respectively, we have (after a few calculations)

$$\Psi_+(\lambda) = \Psi_-(\lambda) \begin{pmatrix} 1 & r(\lambda)e^{2i\lambda x + 4i\lambda^2 t} \\ -r^*(\lambda)e^{-2i\lambda x - 4i\lambda^2 t} & 1 - |r(\lambda)|^2 \end{pmatrix}, \ Im\lambda = 0, \tag{16}$$

where $r(\lambda) = \frac{b(\lambda,t=0)}{a(\lambda,t=0)}$. Note that the jump matrix has determinant 1.

Finally, one can prove that $a(\lambda) = 1 + O(1/\lambda)$ as $\lambda \to \infty, Im\lambda \geq 0$. Hence

$$\Psi(\lambda = \infty) = I. \tag{17}$$

We thus end up with a Riemann-Hilbert factorization problem: Ψ is a matrix function, analytic in the complement of the real line, satisfying the jump condition (16) and the asymptotic condition (17).

Conversely, the solution of the Riemann-Hilbert problem enables us to recover $q(x,t)$. Indeed

$$q(x,t) = -2lim_{\lambda\to\infty}\lambda\Psi_{12}. \tag{18}$$

Thus, the initial value problem for (2) is reduced to the above Riemann-Hilbert. For given $\epsilon \neq 1$ we now have

$$\Psi_+(\lambda) = \Psi_-(\lambda) \begin{pmatrix} 1 & r(\lambda)e^{\frac{2i\lambda x+4i\lambda^2 t}{\epsilon}} \\ -r^*(\lambda)e^{\frac{-2i\lambda x-4i\lambda^2 t}{\epsilon}} & 1-|r(\lambda)|^2 \end{pmatrix}, \; Im\lambda = 0, \tag{19}$$
$$\Psi(\lambda=\infty) = I.$$

As $\epsilon \to 0$, the asymptotic behavior of r can be found through WKB techniques. Suppose now the initial data is positive and bell-shaped. Suppose also that their maximum value is 1. We have

$$r(\lambda;\epsilon) \sim -iexp(\frac{-2iR(\lambda)}{\epsilon})\chi_{[0,1]}, \tag{20}$$

where

$$R(\lambda) = \lambda x_+(\lambda) + \int_{x_+}^{\infty} [\lambda - (\lambda^2 - A(x)^2)^{1/2}]dx,$$

and $x_\pm$ are the positive and negative solutions of $A(x)^2 - \lambda^2 = 0$ respectively.

We also have

$$1 - |r|^2 \sim e^{\frac{-2T(\lambda)}{\epsilon}},$$
$$where\ T(\lambda) = \int_{x_-}^{x_+} (A(x)^2 - \lambda^2)^{1/2}dx.$$

VIII. RIEMANN-HILBERT DEFORMATION

We next indicate how the Riemann-Hilbert problem (19) can be "deformed" to a problem that is explicitly solvable. We are essentially repeating [DVZ].

The first idea involves the so-called "g-function". We introduce a scalar function $g(\lambda)$ which is to be analytic in $\mathbb{C} \setminus \mathbb{R}$ and decay like $O(\frac{1}{\lambda})$ at infinity. This function will beuniquely specified eventually.

Let $\Psi_1 = \Psi exp(\frac{-ig\sigma_3}{\epsilon})$. If Ψ satisfies (19), then Ψ_1 solves a Riemann-Hilbert problem with jump matrix

$$v_1(\lambda) = \begin{pmatrix} e^{\frac{-i(g_+ + g_-)}{\epsilon}} & -ie^{\frac{iH}{\epsilon}} \\ -ie^{\frac{-iH}{\epsilon}} & e^{\frac{(ig_+ - ig_- 2T)}{\epsilon}} \end{pmatrix}, \; Im\lambda = 0, \tag{21}$$

where $H = g_+ + g_- - 2R + 2\lambda x + \lambda^2 t$.

We seek to reduce the jump (21) to a jump of one of the three following types:

$$\begin{pmatrix} 0 & -ie^{\frac{iH}{\epsilon}} \\ -ie^{\frac{-iH}{\epsilon}} & 1 \end{pmatrix}, \begin{pmatrix} 0 & -ie^{\frac{iH}{\epsilon}} \\ -ie^{\frac{-iH}{\epsilon}} & 0 \end{pmatrix}, \begin{pmatrix} 1 & -ie^{\frac{iH}{\epsilon}} \\ -ie^{\frac{-iH}{\epsilon}} & 0 \end{pmatrix}. \tag{22}$$

The motivation is the following. We expect that (19) will be deformable to a RH problem that can be explicitly solvable in terms of finite genus theta functions. Such a problem will have a "finite gap" structure. This means that the real line will be divided into a finite number of subintervals. In some of them the jump matrix has to look like

$$\begin{pmatrix} 0 & -ie^{\frac{iH}{\epsilon}} \\ -ie^{\frac{-iH}{\epsilon}} & 0 \end{pmatrix}$$

and in others it has to be the identity. We know however (through lens-type arguments, see [DVZ]) that matrices of the first or third form in (22) can be reduced to the identity. Hence the ansatz (22).

To arrive from (21) to (22) we impose the seemingly artificial conditions on the functions g, H.

$$
\begin{aligned}
-2iT &= g_+ - g_- \;\; \textit{and} \;\; H' < 0, \;\; \textit{or} \\
-2iT < g_+ - g_- &< 0, \;\; \textit{and} \;\; H' = 0, \;\; \textit{or} \\
g_+ - g_- &= 0, \;\; \textit{and} \;\; H' > 0.
\end{aligned}
\tag{23}
$$

H' denotes the derivative of H with respect to λ. It is easy to see that conditions (23) amount to a scalar Riemann-Hilbert problem that can be solved explicitly, plus a set of algebraic conditions on the endpoints λ_j defining the gap structure, plus a set of inequalities which essentially pick up the appropriate number of λ_j's.

Conditions (23) reduce to the Euler-Lagrange conditions of a variational problem. This is the same variational problrm introduced by Lax and Levermore.

IX. SEMICLASSICAL FOCUSING NLS [KMM]

The problem described in section VIII above is still a self-adjoint problem and can actually be tackled via techniques that go back to Lax-Levermore, by using some ingenious tricks of Venakides.

The focusing NLS problem requires a more sophisticated technique, however. Riemann-Hilbert methods are essential. In [KMM] the corrresponding Riemann-Hilbert problem is reduced to a variational problem, albeit of a more complicated nature than the ones described so far. Instead of looking for a g-function (or a measure) defined on the real line, we are looking for two things: a contour C in the complex plane and a measure μ defined on C. The problem is a mini-max (unstable) problem rather than a simple maximum problem on a given contour. For each contour we look for the equilibrium measure that minimizes the electrostatic energy (under the presence of an external field). But we also have to maximize over all contours to find THE appropriate contour on which our model Riemann-Hilbert problem will be defined.

The derivation and discussion of the minimax variational problem can be found in Chapter 8 of [KMM].

It is fair to say that [KMM] also presents the culmination of the Deift-Its-Zhou program on Riemann-Hilbert deformation techniques. The determination of the optimal contour on which the model Riamnn-Hilbert problem lies is neither trivial (as in the first works on this method) nor depends on the solution of an algebraic system (as in [DVZ]). It relies instead on the solution of a system of ODEs and inequalities. This system is no more the Euler-Lagrange system of a maximum problem for a logarithmic potential on a given contour (as it is in [DVZ] say) but instead it comes from a mini-max variational problem set on the whole complex plane.

The variational problem is described in the last section of these notes. We sketch a proof of existence of a solution to this variational problem. Standard potential theory techniques can then be used to show the regularity of the solution. The support of the maximizing measure can be proved to consist of a finite union of analytic arcs. This justifies the finite-gap ansatz.

X. THE VARIATIONAL PROBLEM [KMMR]

1. THE SETTING

The space we have to work with is the upper half-plane without the spike: $\mathbb{H} = \{z : Imz > 0\} \setminus \{z : Rez = 0, 0 < Imz \leq A\}$. In the closure of this space, $\bar{\mathbb{H}}$, we consider the points ix_+ and ix_-, where $0 \leq x < A$ as different. In other words, we cut a slit in the upper half-plane along the segment $(0, iA)$ and distinguish between the two sides of the slit. The point infinity belongs to $\bar{\mathbb{H}}$, but not $\mathbb{H}$.

The set of distinguished points $\mathbb{A}$ contains the two points 0_+ and 0_-. We define $\mathbb{F}$ to be the set of all continua F in $\bar{\mathbb{H}}$ (i.e. connected compact sets containing more than one point) without interior points, containing the distinguished points $0_+, 0_-$.

We need to introduce a topology on $\mathbb{F}$. We define the Hausdorff metric as follows. Let $F^\delta = \{z \in \mathbb{C} : \rho(z, F) < \delta\}$, where ρ is the usual distance between compact sets $\rho(E, F) = min_{z \in E, \zeta \in F} |z - \zeta|$, and $|z - \zeta|$ is the chordal distance since we want $\bar{\mathbb{H}}$ to be compact. Then, define the Hausdorff distance

$$\rho_{\mathbb{H}}(E, F) = inf\{\delta : E \subset F^\delta, F \subset E^\delta\}. \tag{1}$$

It is easily seen (cf. [D], p.61, with trivial modifications) that, since $\bar{\mathbb{H}}$ is compact and complete, so is $\mathbb{F}$.

Our external field ϕ is the function defined as follows. Let $G(\lambda; \eta)$ be the Green's function for the upper half-plane

$$G(\lambda; \eta) = log \frac{|\lambda - \eta^*|}{|\lambda - \eta|}$$

and $d\mu^0(\eta)$ be the nonnegative measure $-\rho^0(\eta) d\eta$ on the segment $[0, iA]$ oriented from 0 to iA. Then

$$\phi(\lambda) = -\int G(\lambda; \eta) d\mu^0(\eta) - Re(i\pi J \int_\lambda^{iA} \rho^0(\eta) d\eta + 2iJ(\lambda x + \lambda^2 t)). \tag{3}$$

Here we assume $t \geq 0$, so that $\sigma = J$ in formula (8.3) of [KMM]. J is a priori 1 or -1, but eventually we will choose $J = 1$.

ϕ is a real analytic function in $\mathbb{H}$, and continuous in $\bar{\mathbb{H}}$, except at infinity.

Next, let $\mathbb{M}$ be the set of all positive mesaures on $\mathbb{H}$. Define the (free) energy

$$E(\mu) = \int \int G(x, y) d\mu(x) d\mu(y), \ \mu \in \mathbb{M}. \tag{4}$$

Also, let

$$V^\mu(z) = \int G(z, x) d\mu(x), \ \mu \in \mathbb{M}. \tag{5}$$

The weighted energy of the field ϕ is

$$E_\phi(\mu) = E(\mu) + 2 \int \phi d\mu, \tag{6}$$

for any $\mu \in \mathbb{M}$.

Now, given any $F \in \mathbb{F}$, the equilibrium measure λ^F supported in F is defined by

$$E_\phi(\lambda^F) = min_{\mu \in M(F)} E_\phi(\mu), \tag{7}$$

where $M(F)$ is the set of positive measures supported in F. It is not always true that such an equilibrium measure exists. The Gauss-Frostman theorem ([ST], p.135) guarantees the existence of the equilibrium measure when F does not touch the boundary of the domain $\mathbb{H}$. This is not the case here. Still, in the particular case of our special external field we can see that, for any given x, t and for any continuum F, the weighted energy is bounded below and (see Theorem 2 in section 3) λ^F exists. [So, in particular, we do know that the supremum of the equilibrium weighted energies over all continua is finite.]

2. CONTINUITY OF ENERGY

We consider the functional that takes a given continuum F to the equilibrium energy on this continuum:

$$\mathbb{E} : F \to E_\phi[F] = E_\phi(\lambda^F) = \inf_{\mu \in M(F)} (E(\mu) + 2 \int \phi d\mu) \tag{8}$$

and we want to show that it is continuous, if ϕ is continuous in $\bar{\mathbb{H}}$. Note that this is not the case for the field given by (3), since it has a singularity at ∞. We will see how to circumvent this difficulty later. For the present section, ϕ is just assumed to be a continuous function in $\bar{\mathbb{H}}$.

So, suppose $G \in \mathbb{F}$, with $\rho_{\mathbb{H}}(F, G) < \rho$, a small positive constant. Let $\lambda = \lambda_\phi^F$ be the equilibrium measure on F and $\mu = \lambda_\phi^G$ be the equilibrium measure on G. (It exists because of the remark at the end of section 1.)

We consider the Green's balayage of μ on F, say $\hat{\mu}$. Then *supp* $\hat{\mu} \in F$ and

$$\begin{aligned} V_{\mathbb{H}}^{\hat{\mu}} &= V_{\mathbb{H}}^{\mu} \ on \ F, \\ \int u d\hat{\mu} &= \int u d\mu, \end{aligned} \tag{9}$$

for any function u that is harmonic in $\mathbb{H} \setminus F$ and continuous in $\bar{\mathbb{H}}$.

Similarly consider the balayage of λ to G. We trivially have

$$\begin{aligned} E_\phi[G] &\leq E_\phi(\hat{\lambda}), \\ E_\phi[F] &\leq E_\phi(\hat{\mu}). \end{aligned}$$

LEMMA 1. Suppose K is a compact set [with positive capacity], μ some positive measure in the plane and $\hat{\mu}$ is the Green's balayage to K. Then

$$\begin{aligned} V_{\mathbb{H}}^{\hat{\mu}} &= V_{\mathbb{H}}^{\mu} - V_{\mathbb{H} \setminus K}^{\mu}, \\ E(\hat{\mu}) &= E(\mu) - E_{\mathbb{H} \setminus K}(\mu), \end{aligned} \tag{10}$$

where $E_{\mathbb{H} \setminus K}(\mu)$ is the unweighted Green energy with respect to $\mathbb{H} \setminus K$.

PROOF OF LEMMA 1: The first identity follows from the fact that $V_{\mathbb{H}}^{\hat{\mu}} - V_{\mathbb{H}}^{\mu}$ vanishes on K and $\partial \mathbb{H}$, and is harmonic in $\mathbb{H} \setminus K$ and superharmonic in $\mathbb{H}$.

Integrating $E(\hat\mu) = \int V_{\mathbb{H}}^{\hat\mu} d\hat\mu = \int V_{\mathbb{H}}^{\mu} d\hat\mu - \int V_{\mathbb{H}\setminus K}^{\mu} d\hat\mu = \int (V_{\mathbb{H}}^{\mu} - V_{\mathbb{H}\setminus K}^{\mu}) d\mu = E(\mu) - E_{\mathbb{H}\setminus K}(\mu)$. The proof of the Lemma follows.

In particular, since unweighted energies are nonnegative,

$$E(\hat\mu) \le E(\mu). \tag{11}$$

PROOF OF CONTINUITY: Let u_ϕ be a function harmonic in $\mathbb{H} \setminus F$ such that $u_\phi = \phi$ on F and $u_\phi = 0$ on $\partial\mathbb{H} \setminus F$. By the definition of balayage one has $\int \phi d\hat\mu = \int u_\phi d\mu$.

We have $E_\phi[F] \le E_\phi(\hat\mu) = E(\hat\mu) + 2\int \phi d\hat\mu \le E(\mu) + 2\int \phi d\mu + 2\int (u_\phi - \phi) d\mu = E_\phi(\mu) + 2\int (u_\phi - \phi) d\mu = E_\phi[G] + 2\int (u_\phi - \phi) d\mu$.

In a small neighbourhood of F, $\bar F_\rho = \{z : \rho(z, F) \le \rho\}$, we have

$$|2\int (u_\phi - \phi)(x) d\mu(x)| \le C max_{x\in \bar F_\rho} |u_\phi(x) - \phi(x)|. \tag{12}$$

ASSUMPTION: We assumed here that the equilibrium measures on continua near F are bounded above. This assumption can be justified a posteriori.

Now given $x \in \bar F_\rho$, choose $z \in F$ such that $|z - x| = \rho$. The above expression is

$$\le C max_{\bar F_\rho} |u_\phi(x) - \phi(x) - u_\phi(z) + \phi(z)| \le o(1) + C max_{\bar F_\rho} |u_\phi(x) - u_\phi(z)|. \tag{13}$$

It remains to bound $|u_\phi(x) - u_\phi(z)|$ by an $o(1)$ quantity.

The next Theorem is due to Milloux and can be found in [G].

THEOREM 1. Suppose D is an open disc of radius R, with center z_0; let x be a point in D, F a continuum in $\mathbb{C}$, containing z_0, and Ω be the connected component of $D \setminus F$ containing x. Let $w(z)$ be a function harmonic in Ω such that

$$\begin{aligned} w(z) &= 0,\ z \in F \cap \partial\Omega, \\ w(z) &= 1,\ z \in \partial\Omega \setminus F. \end{aligned} \tag{14}$$

Then $w(x) \le C(\frac{|x - z_0|}{R})^{1/2}$.

PROOF: Omitted (see [G], p.347).

Now, select a disc of radius $\rho^{1/2}$, centered on z. We have $|u_\phi(x) - u_\phi(z)| = o(1)$ on the part of F lying in the disc, while $|u_\phi(x) - u_\phi(z)|$ is bounded by some positive constant M on the disc boundary.

LEMMA 2. Let Ω be a domain, $\partial\Omega = F_1 \cup F_2$ and

$$\begin{aligned} w_1 &= 0,\ z \in F_1, \\ &= 1,\ z \in F_2; \\ w_2 &= 1,\ z \in F_1, \\ &= 0,\ z \in F_2. \end{aligned}$$

Suppose u is harmonic in Ω and

$$\begin{aligned} u(z) &\le \epsilon,\ z \in F_1, \\ u(z) &\le M,\ z \in F_2. \end{aligned}$$

Then $u(z) \leq \epsilon w_2(z) + M w_1(z)$.
PROOF OF LEMMA 2: Maximum principle.

Now, using Milloux's Theorem, we get $|u_\phi(x) - u_\phi(z)| \leq o(1) w_2(z) + M w_1(z) \leq o(1) + MC\frac{|x-z|}{\rho}^{1/2} \leq o(1) + MC\rho^{1/2}$. Hence result.

3. PROOF OF EXISTENCE OF A MAXIMIZING CONTINUUM

THEOREM 2. Let F be a continuum in $\bar{\mathbb{H}}$. Suppose

$$inf_{\mu \in M(F)} E_\phi(\mu) > -\infty. \tag{15}$$

Then the equilibrium measure on F exists, that is there is a non-negative measure λ^F with support contained in F such that $E_\phi[F] = E_\phi(\lambda^F) = inf_{\mu \in M(F)} E_\phi(\mu)$.
PROOF: Omitted. Apparently this is a well known theorem, at least for logarithmic rather than Green's energy and for unit measures. But the proof goes through anyway. The fact that we have Green's energy actually makes it a bit easier, since the Green's function is nonnegative evrywhere.

We now recall that the energy continuity proof was based on the continuity of ϕ. In our case, ϕ is discontinuous at ∞. Still we can prove that the energy is upper semicontinuous and that will be enough.
THEOREM 3. For the field given by (3), the energy functional defined in (8) is upper semicontinuous.
PROOF: Let F be a continuum. If ∞ is not in F, then we're done. If $\infty \in F$, let $\lambda = \lambda^F$ be the equilibrium measure. We can assume that on the equilibrium measure ϕ is bounded by some positive constant M, say. Indeed, on the support of the equilibrium measure λ, we have

$$V_{\mathbb{H}}^\lambda + \phi = M,$$

where M is some constant. But $V_{\mathbb{H}}^\lambda \geq 0$, so $\phi \leq M$.
This means that we can change ϕ to $\phi' = min(\phi, M)$, which is an upper semicontinuous function.
Let ϕ_n be a sequence of continuous functions such that $\phi_n \downarrow \phi$. Each functional $E_{\phi_n}[F]$ is continuous, and $E_{\phi_n}[F] \downarrow E_\phi[F]$. So, $E_\phi[F]$ is upper semicontinuous.

THEOREM 4. There exists a contour F such that the equilibrium measure λ^F exists and $E_\phi[F] = E_\phi(\lambda^F) = max_{F \in \mathbb{F}} min_{\mu \in M(F)} E_\phi(\mu)$.
PROOF: We know that there is at least one contour (namely the one wrapped around the spike $(0, iA)$) for which the equilibrium measure exists and $E_\phi(\lambda^F) > -\infty$. Hence the supremum over continua in $\mathbb{F}$ is finite (and trivially nonpositive), since $\mathbb{F}$ is compact. Call it L.
We can now take a sequence F_n such that $E_\phi[F_n] \to L$. Choose a convergent subsequence of continua $F_n \to F$. By upper semicontinuity of the weighted energy functional,

$$\bar{lim} E_\phi[F_n] \leq E_\phi[F] \leq L = lim E_\phi[F_n].$$

So $L = E_\phi[F]$. The theorem is proved.

4. ACCEPTABILITY OF THE CONTOUR

We have thus shown that a solution of the maximum-minimum problem exists. We do not know yet that it is an analytic contour. All we know it is a "continuum".

The fact that it is indeed an analytic contour and that the support of the measure is a finite union of analytic arcs will follow from the analyticity properties of the external field.

It is important however, that the maxi-minimizing continuum does not touch the boundary of the space $\mathbb{H}$ except of course at the points $0^-, 0^+$. This is to guarantee that variations with respect to the maximizing contour can be properly taken.

The proof of the acceptability of the continuum requires three things.

(i) The continuum does not touch the real negative axis.

(ii) The continuum does not touch the real positive axis.

(iii) The continuum does not touch the spike $\{z : Rez = 0, 0 < Imz \leq A\}$, except of course at time $t = 0$.

(iv) It also remains to consider the point at infinity. We cannot prove that the continuum does not hit this point. In fact, our numerics ([KMM], Chapter 6) show that, depending on the sign of J it may well do so. However, we will eventually show that even if we allow the continuum to pass through infinity, a modification of the argument in [KMM] will work.

PROOFS: (i) If $\lambda < 0$, then $\phi(\lambda) > 0$. But we can always delete the strictly positive measure lying in a region where the field is positive and make the energy smaller. So even the solution of the "inner" minimizinng problem has to lie away from the real negative axis.

(ii) If $\lambda > 0$ and $J > 0$, then $\frac{d\phi}{dIm\lambda} > 0$ (since J =sgn(x), see [KMM] (6.12)). So, suppose the maximizing continuum touches the axis. We can always push the measures up away from the real axis and end up with a continuum that has greater minimal energy, thus arriving at a contradiction.

(iii) To show that the maximizing continuum stays away from the linear segment that connects 0 and iA (the 'spike'): we can see that this is indeed the case by comparing to what we know is true near time zero. Begin by observing that for t small and the appropriate choice of J, the contour (which at time zero coincides with the piece of the boundary of $\mathbb{H}$ that joins points 0_- and 0_+) is repelled by the spike.

The key observation is that ON THE SPIKE the field is independent of time. So, if at time zero+ the field repells the max-min contour, it has to do the same at all time.

(iv) At first, it might seem that the maximizing continuum (which is eventually seen to be a contour) should not pass through infinity. Indeed, the transformations (2.17) and (4.1) of [KMM] implicitly assume that the contour C lies in $\mathbb{C}$. Otherwise, one would lose the appropriate normalization for M at infinity. However, one has to simply notice that infinity is just an arbitrary choice of normalizing point, once we view our Riemann-Hilbert problems in the compact Riemann Sphere. The important observation is that the composition of transformations (2.17) and (4.1) [which are purely formal, i.e. no estimates are required and no approximation is needed] does not introduce any bad (essential) singularities. In the end, the asymptotic behavior of $\tilde{N}^\sigma$ is still the identity at ∞_- and non-singular at ∞_+. So, in the end it is acceptable for a contour to go through the point infinity.

XI. REFERENCES

[D] J.Dieudonné, Foundations of Modern Analysis, Academic Press, 1969.

[DVZ] P.Deift, S.Venakides, X.Zhou, New Results in Small Dispersion KdV by an Extension of the Steepest Descent Method for Riemann-Hilbert Problems, IMRN, n.6, 1997, pp.286-299.

[G] G.M. Goluzin, Geometric Theory of Functions of a Complex Variable, Translations of Mathematical Monographs, v.26, AMS 1969.

[KMM] S.Kamvissis, K.McLaughlin, P.Miller, Semiclassical Soliton Ensembles for the Focusing Nonlinear Schrödinger Equation, Annals of Mathematics Studies, v.154, 2003; also nlin.SI/0012034.

[KMMR] S.Kamvissis, K.McLaughlin, P.Miller, E.Rakhmanov, A Variational Problem Related to Semiclassical NLS, in preparation.

[L] N.S.Landkof, Foundations of Modern Potential Theory, Springer 1972.

[LL] P.D.Lax, C.D.Levermore, The Small Dispersion Limit of the Korteweg-de Vries Equation, I, II, and III, Comm Pure Appl.Math, v.36, 1983.

[LLV] P.D.Lax, C.D.Levermore, S.Venakides, The Generation and Propagation of Oscillations in Dispersive Initial Value Problems and their Limiting Behavior, Important Developments in Soliton Theory 1980-1990, ed. by A.Fokas, V.Zakharov, Springer, Berlin, 1993, pp.205-241.

[ST] E.B.Saff, V.Totik, Logarithmic Potentials with External Fields, Springer Verlag, 1997.