Chapter One

Introduction and Overview

1.1 BACKGROUND

The initial-value problem for the focusing nonlinear Schrödinger equation is

$$i\hbar\partial_t\psi + \frac{\hbar^2}{2}\partial_x^2\psi + |\psi|^2\psi = 0, \qquad \psi(x,0) = \psi_0(x). \tag{1.1}$$

We are interested in studying the behavior of solutions of this initial-value problem in the so-called semiclassical limit. To make this precise, the initial data is given in the form

$$\psi_0(x) = A(x)\exp(iS(x)/\hbar), \tag{1.2}$$

where $A(x)$ is a positive real amplitude function that is rapidly decreasing for large $|x|$ and where $S(x)$ is a real phase function that decays rapidly to constant values for large $|x|$. Studying the semiclassical limit means: fix once and for all the functions $A(x)$ and $S(x)$, and then for each sufficiently small value of $\hbar > 0$, solve the initial-value problem (1.1) subject to the initial data (1.2), obtaining the solution $\psi(x, t; \hbar)$. Describe the collection of solutions $\psi(x, t; \hbar)$ in the limit of $\hbar \downarrow 0$.

The initial-value problem (1.1) is a key model in modern nonlinear optical physics and its increasingly important applications in the telecommunications industry. On the one hand, it describes the stationary profiles of high-intensity paraxial beams propagating in materials with a nonlinear response, the so-called Kerr effect. This is the realm of *spatial solitons*, which are envisioned as self-guided beams that can form the fundamental components of an all-optical switching system. In this context, the semiclassical scaling $\hbar \ll 1$ of (1.1) corresponds to the joint limit of paraxial rays and geometrical optics in the presence of nonlinear effects. On the other hand, (1.1) also describes the propagation of time-dependent envelope pulses in optical fibers operating at carrier wavelengths in the anomalous dispersion regime (usually infrared wavelengths near 1550 nm). These envelope pulses are known as *temporal solitons* and are envisioned as robust bits in a digital signal traveling through the fiber. In these fiber-optic applications, the semiclassical scaling $\hbar \ll 1$ is particularly appropriate for modeling propagation in certain dispersion-shifted fibers that are increasingly common. See [FM98] for a careful discussion of this point leading to a similarly scaled *defocusing* equation when the fiber parameters are indicative of weak normal dispersion; similar arguments with slightly adjusted parameters can lead to the focusing problem (1.1) just as easily. Of course in neither of these optical applications is the small parameter actually Planck's constant, but we write it as $\hbar$ in formal analogy with the quantum-mechanical interpretation of the linear terms in (1.1), which also gives rise to the description of the limit of interest as "semiclassical."

The independent variables x and t parametrize the semiclassical limit, and one certainly does not expect a pointwise asymptotic description of the solution to be uniform with respect to these parameters. The statement of the problem becomes more precise when one further constrains these parameters. For example, one might set $x = X/\hbar$ and $t = T/\hbar$ for X and T fixed as $\hbar \downarrow 0$. In this limit, several studies [B96, BK99] have suggested that for initial data with $|S(x)|$ sufficiently large, the field consists of trains of separated solitons, with the remarkable property that there is a well-defined relationship between the soliton amplitude and velocity (nonlinear dispersion relation) that is determined from the initial functions $A(x)$ and $S(x)$ via the asymptotic distribution of eigenvalues of the Zakharov-Shabat scattering problem. In general, solitons can have arbitrary amplitudes and velocities, so the observed correlation is a direct consequence of the semiclassical limit.

Here, we are concerned with a different asymptotic parametrization. Namely, we consider the sequence of functions $\psi(x, t; \hbar)$ in a fixed but arbitrary compact set of the (x, t)-plane in the limit $\hbar \downarrow 0$. In this scaling, the large number of individual solitons present in the initial data are strongly nonlinearly superposed, and interesting spatiotemporal patterns have been observed [MK98, BK99].

This choice of scaling has several features in common with similar limits studied in other integrable systems, for example, the zero-dispersion limit of the Korteweg–de Vries equation analyzed by Lax and Levermore [LL83], the continuum limit of the Toda lattice studied by Deift and K. T.-R. McLaughlin [DM98], and the semiclassical limit of the *defocusing* nonlinear Schrödinger equation studied by Jin, Levermore, and D. W. McLaughlin [JLM99]. In all of these cases, the challenge is to use the machinery of the inverse-scattering transform to prove convergence in some sense to a complicated asymptotic description that necessarily consists of two disparate space and time scales. One scale (the *macrostructure*) is encoded in the initial data, and the other scale (the *microstructure*) is introduced by the small parameter (the dispersion parameter in the Korteweg–de Vries equation, the lattice spacing in the Toda lattice, and Planck's constant, $\hbar$, in the nonlinear Schrödinger equation).

In Lax and Levermore's analysis of the zero-dispersion limit for the Korteweg–de Vries equation [LL83], a fundamental role was played by an explicit, albeit complicated, formula for the exact solution of the initial-value problem for initial data that has been modified in an asymptotically negligible sense. This formula directly represents the solution $u(x, t)$ of the problem in terms of the second logarithmic derivative of a determinant. When the determinant is expanded as a sum of principal minors, the minors are all positive, and the sum is shown to be asymptotically dominated by its largest term. This leads directly to a discrete maximization problem (the problem is discrete because the number of minors is finite but large when the dispersion parameter is small) in which the independent variables x and t appear as parameters. This maximization problem characterizes the determinant up to a controllable error. Leading-order asymptotics are obtained by letting the dispersion parameter go to zero and observing that the discrete maximization problem goes over into a variational problem in a space of admissible functions. It turns out that the weak limit of each member of the whole hierarchy of conserved local densities for the Korteweg–de Vries equation can be directly expressed in terms of the solution of the variational problem and its derivatives.

In all of the problems where the method of Lax and Levermore has been successful, the macrostructure parameters (or equivalently weak limits of various conserved local densities) have been shown to evolve locally in space and time as solutions of a hyperbolic system known as the *Whitham equations* or the *modulation equations*. The global picture consists of several regions of the (x, t)-plane in each of which the microstructure is qualitatively uniform (periodic or quasiperiodic) and the macrostructure obeys a system of modulation equations. The size of this system of equations (number of unknowns) is related to the complexity of the microstructure. The variational method of Lax and Levermore amounts to the global analysis showing how the solutions of the modulation equations are patched together at the boundaries of these various regions. By hyperbolicity and the corresponding local well-posedness of the modulation equations, it follows that, for example, the small-time behavior (sufficiently small, but independent of the size of the limit parameter) of the limit is connected with prescribed initial data in a stable fashion.

The modulation equations may be derived formally, without reference to initial data. For the focusing nonlinear Schrödinger equation, these quasilinear equations are *elliptic* [FL86], which makes the Cauchy initial-value problem for them ill-posed in common spaces. To illustrate this ill-posedness for the Whitham equations in their simplest version (genus zero), one might make the assumption that the microstructure in the solution of (1.1) resembles the modulated rapid oscillations present in the initial data. That is, one could suppose that for some order-one time the solution can be represented in the form

$$\psi(x, t) = A(x, t) \exp(iS(x, t)/\hbar), \tag{1.3}$$

where $A(x, 0) = A(x)$ and $S(x, 0) = S(x)$. Setting $\rho(x, t) = A(x, t)^2$ and $\mu(x, t) = A(x, t)^2 \partial_x S(x, t)$, one finds that the initial-value problem (1.1) implies

$$\partial_t \rho + \partial_x \mu = 0, \qquad \partial_t \mu + \partial_x \left(\frac{\mu^2}{\rho} - \frac{\rho^2}{2} \right) = \frac{\hbar^2}{4} \partial_x \big(\rho \partial_x^2 \log(\rho) \big), \tag{1.4}$$

with initial data $\rho(x, 0) = A(x)^2$ and $\mu(x, 0) = A(x)^2 S'(x)$. The modulation equations corresponding to our assumption about the microstructure are obtained by simply neglecting the terms that are formally of order $\hbar^2$ in these equations. That is, one supposes that for some finite time $\rho(x, t)$ and $\mu(x, t)$ are uniformly close, respectively, to functions $\rho_\circ(x, t)$ and $\mu_\circ(x, t)$ as $\hbar \downarrow 0$, where these latter two functions solve the system

$$\partial_t \rho_\circ + \partial_x \mu_\circ = 0, \qquad \partial_t \mu_\circ + \partial_x \left(\frac{\mu_\circ^2}{\rho_\circ} - \frac{\rho_\circ^2}{2} \right) = 0, \tag{1.5}$$

with initial data $\rho_\circ(x, 0) = A(x)^2$ and $\mu_\circ(x, 0) = A(x)^2 S'(x)$. This is a quasilinear nonlinear system, and it is easy to check that it is of elliptic type; that is, the characteristic velocities $\mu_\circ/\rho_\circ \pm i\sqrt{\rho_\circ}$ are complex at every point where $\rho_\circ$ is nonzero. This implies that the Cauchy problem posed here for the modulation equations is ill-posed.

This fact immediately makes the interpretation of the semiclassical limit of the initial-value problem (1.1) complicated; even if it turns out that one can prove convergence to the solutions of the modulation equations for some initial data, it is not clear that one can deduce anything at all about the asymptotics for "nearby" initial data. In this sense, the formal semiclassical limit of (1.1) is very unstable.

One feature that both the hyperbolic and elliptic modulation equations have in common is the possibility of singularities that develop in finite time from smooth initial data. This singularity formation seems physically correct in the context of spatial optical solitons, where the Kerr effect has been known for some time to lead to self-focusing of light beams and, in two transverse dimensions (the independent variable x), to the total collapse of the beam in finite propagation distance (the independent variable t). As long ago as 1966, this led Akhmanov, Sukhorukov, and Khokhlov [ASK66] to propose a certain exact solution of the modulation equations (1.5) as a model for the self-focusing phenomenon in one transverse dimension. They did not try to solve any initial-value problem for these equations; indeed they were clearly aware of the ellipticity of the system (1.5) and the coincident ill-posedness of its Cauchy problem. Rather, they introduced a clever change of variables (some insight into their possible reasoning was proposed by Whitham [W74]) and obtained a set of two real equations implicitly defining two real unknowns as functions of x and t. After the fact, they noted that their solution matched onto the initial data $A(x) = A\,\mathrm{sech}(x)$ and $S(x) \equiv 0$. The original paper of Akhmanov, Sukhorukov, and Khokhlov contains drawings of the solution at various times up to the formation of a finite-amplitude singularity (i.e., the singularity forms in the derivatives) at the time $t = t_{\mathrm{crit}} = 1/(2A)$. The authors even plotted their solution beyond the singularity, showing the onset of multivaluedness. They understood that the model solution cannot possibly be valid beyond the singularity and, in the physical context of interest in their study, ascribed this as much to the breakdown of the paraxial approximation leading to the nonlinear Schrödinger equation (1.1) as a beam propagation model in the first place as to the failure of the formal geometrical optics (semiclassical) limit for (1.1).

As is the case in all of the problems for which the method of Lax and Levermore has been successful, careful analysis of the semiclassical limit $\hbar \downarrow 0$ for (1.1) is possible in principle because the problem can be solved for each $\hbar$ by the inverse-scattering transform, as was first shown by Zakharov and Shabat [ZS72]. The small parameter necessarily enters the problem both in the forward-scattering step and in the inverse-scattering step. It is significant that the analysis of the semiclassical limit for (1.1) is frustrated in both steps. In the forward-scattering step, the difficulties are related to the nonselfadjointness of the scattering problem associated with (1.1). By contrast, in each of the cases mentioned previously, where calculations of this type were successfully carried out, the associated scattering problem is selfadjoint. In the inverse-scattering step, the difficulties are related to the limit being attained by a kind of furious cancellation in which no single term in the expansion of the solution is apparently dominant. In fact, in Zakharov and Shabat's paper [ZS72], there appears an explicit formula for the function $\rho(x, t)$ solving (1.4) that is qualitatively very similar to that solving the Korteweg–de Vries equation and taken as the starting point in Lax and Levermore's analysis. When $t = 0$, this formula has all of the properties required by the Lax-Levermore theory. Namely, the determinant can be expanded as a sum of positive terms, which is controlled by its largest term as $\hbar \downarrow 0$. This calculation is carried out in the paper of Ercolani, Jin, Levermore, and MacEvoy [EJLM93]. But when t is fixed at any nonzero value, the principal minors lose their positive definiteness, and it can no longer be proved that the sum is dominated by its largest term. If the weak limit exists, then all that can be said from this approach is that it arises out of subtle cancellation. In particular, from this point of view it appears that there is no obvious variational principle characterizing the limit.

1.2 APPROACH AND SUMMARY OF RESULTS

This book is primarily concerned with the semiclassical analysis of the inverse-scattering step. For simplicity, we restrict attention from the start to the case of initial data that satisfy $S(x) \equiv 0$. In this case it was observed already in Zakharov and Shabat's paper [ZS72] that while not strictly selfadjoint for any $\hbar > 0$ the scattering problem formally goes over into a semiclassically scaled selfadjoint linear Schrödinger operator in the limit $\hbar \downarrow 0$. In [EJLM93], this observation was exploited to propose WKB formulae that were subsequently used to study the zero-dispersion limit of the modified Korteweg–de Vries equation, an equation associated with the same scattering problem as (1.1), but whose inverse-scattering step is more straightforward because there is no cancellation of the type mentioned previously. (As already mentioned, this cancellation is also absent for the focusing nonlinear Schrödinger problem when $t = 0$, and the calculations in [EJLM93] hold in this case as well.) The WKB approximation amounts to the neglect of the reflection coefficient and the replacement of the true eigenvalues with a sequence of purely imaginary numbers that are obtained from an explicit Bohr-Sommerfeld-type quantization rule. These WKB formulae have not been rigorously established to date; their justification in [EJLM93] rests upon the fact that they reproduce the exact initial data when t is set to zero in the inverse-scattering step. There is, however, one function $A(x)$ for which all of the exact scattering data is known (assuming $S(x) \equiv 0$) exactly: $A(x) = A\,\mathrm{sech}(x)$. The spectrum corresponding to this potential in the nonselfadjoint Zakharov-Shabat scattering problem was computed exactly for all $\hbar$ by Satsuma and Yajima [SY74] and published in 1974. At face value this is a remarkable coincidence: the same initial data for which Akhmanov, Sukhorukov, and Khokhlov found (after the fact!) that they had an exact solution of the modulation equations turns out to be data for which the forward-scattering problem was later shown to be exactly solvable for all $\hbar$. Some additional special cases of potentials where the spectrum can be obtained exactly for all $\hbar$, including some cases with $S(x) \not\equiv 0$, have been found recently by Tovbis and Venakides [TV00].

It turns out that the exact scattering data for the special initial condition $\psi_0(x) = A\,\mathrm{sech}(x)$ coincides with the formal WKB approximation to the scattering data, as long as one restricts attention to a particular sequence of positive values of $\hbar \in \{\hbar_N\}$ converging to zero. For these special values of $\hbar$, the initial data is exactly reflectionless, there are exactly N eigenvalues all purely imaginary, and also the distance between the most excited state (the eigenvalue with the smallest magnitude) and the continuous spectrum is exactly half of the distance between each adjacent pair of eigenvalues. In particular, for $\hbar = \hbar_N$, there is no error incurred in reconstructing the corresponding solution of (1.1) using inverse-scattering theory *without reflection coefficient*; the true solution for these values of $\hbar$ is a pure ensemble of N solitons.

In this book, we develop a method that yields detailed strong asymptotics for the inverse-scattering problem corresponding to the scattering data just briefly described. Since this scattering data is the true scattering data corresponding to the Satsuma-Yajima potential, our results imply rigorous asymptotics for the corresponding initial-value problem (1.1). But since the scattering data for this case agrees with its WKB approximation, we prefer to approach the problem from the more general perspective of computing rigorous asymptotics for the inverse problem corresponding to a general family of WKB scattering data. Thus, our approach to the semiclassical limit for the initial-value problem (1.1) for quite general data satisfying $S(x) \equiv 0$ is essentially the familiar step of introducing modified reflectionless initial data whose scattering data is that predicted by the formal WKB approximation. This sort of modification was the first step in the pioneering work of Lax and Levermore [LL83]. Of course, for the Satsuma-Yajima initial data, no modification is necessary as long as $\hbar \in \{\hbar_N\}$.

The main idea that allows our analysis of the inverse-scattering problem to proceed for $t \neq 0$ where the Lax-Levermore method fails is to avoid the direct connection of the discrete scattering data with the solution of the problem via an explicit determinant formula and instead to introduce an intermediate object, namely, an appropriately normalized eigenfunction of the Zakharov-Shabat scattering problem. In general, this eigenfunction satisfies a certain matrix Riemann-Hilbert problem with poles encoding the discrete spectrum and a jump on the real axis of the eigenvalue corresponding to the reflection coefficient on the continuous spectrum. The solution of the nonlinear Schrödinger equation is obtained in turn from the solution of this Riemann-Hilbert problem. This is the essential content of inverse-scattering theory [FT87]. While it of course turns out that in the reflectionless case the Riemann-Hilbert problem may be explicitly solved in terms of meromorphic functions and ratios of determinants, leading to the formula that is the starting point for Lax-Levermore-type analysis, there is some advantage to ignoring this explicit solution and instead trying to obtain uniform asymptotics for the eigenfunction that is the solution of the Riemann-Hilbert problem. Only in studying this intermediate problem do we recover a variational principle that is a generalization of the one from Lax and Levermore's method.

The method we develop in this book to study the asymptotic behavior of the eigenfunction generalizes the steepest-descent method for matrix Riemann-Hilbert problems first proposed by Deift and Zhou in [DZ93] and subsequently developed and further applied in several papers [DVZ94, DZ95]. The generalization of the steepest-descent method that we present here has its basic features in common with the recent application of the method to the Korteweg–de Vries equation in [DVZ97], with recent applications in the theory of orthogonal polynomials and random matrices [DKMVZ97, DKMVZ99A, DKMVZ99B], and also with some applications to long-time asymptotics for soliton-free initial data in the focusing nonlinear Schrödinger equation [K95, K96]. These latter papers make use of an idea that was first introduced in [DVZ94]—using the special choice of a complex phase function to enable the asymptotic reduction of the Riemann-Hilbert problem to a simple form. Our work generalizes this approach because it turns out that an appropriate complex phase function typically does not exist at all relative to a given contour in the complex plane, unless this contour satisfies some additional conditions. In fact, we show that the existence of an appropriate complex phase function *selects portions of the contour on which the Riemann-Hilbert problem should be posed to begin with*. In this sense, the generalization of the method proposed in [DVZ94] that we present here further develops the analogy with the classical asymptotic method of steepest descent; the problem must be solved on a particular contour in the complex plane. In problems previously treated by the steepest-descent method of Deift, Zhou, and others, the problem of finding this special contour has simply not arisen because there is an obvious contour, often implied by the selfadjointness of a related scattering problem, for which the additional conditions that select the contour are *automatically* satisfied. The specification of this special contour can be given a variational interpretation that is the correct generalization of the Lax-Levermore variational principle.

Among our primary results are the following.

1. Strong leading-order semiclassical asymptotics for solutions of the focusing nonlinear Schrödinger equation corresponding to sequences of initial data whose spectral data are reflectionless and have a discrete spectrum obtained from a Bohr-Sommerfeld quantization rule. These asymptotics are valid even after wave breaking and come with a rigorous error bound. The explicit model we obtain—to which the semiclassical solutions are asymptotically close pointwise in x and t—displays qualitatively different behavior before and after wave breaking and in particular exhibits violent oscillations after breaking, confirming phenomena that have been observed in numerical experiments.
2. Formulae explicitly involving the initial data that solve the elliptic Whitham modulation equations. These formulae consequently provide the complete solution to the initial-value problem for the Whitham equations in the category of analytic initial data.
3. The characterization of the caustic curves in the (x, t)-plane where the nature of the microstructure changes suddenly. We also provide what amount to "connection formulae" describing the phase transition that occurs at the caustic. In particular our analysis shows that at first wave breaking there is a spontaneous transition from fields with smooth amplitude (genus zero) to oscillatory fields with intermittent concentrations in amplitude (genus two).
4. A significant extension of the steepest-descent method for asymptotic analysis of Riemann-Hilbert problems introduced by Deift and Zhou. For problems with analytic jump matrices, we show how the freedom of placement of the jump contour in the complex plane can be systematically exploited to asymptotically reduce the norms of the singular integral operators involved in the solution of the Riemann-Hilbert problem. Ultimately this expresses the solution as an explicit contribution modified by a Neumann series involving small bounded operators.

5. A new generalization of Riemann-Hilbert methods allowing the analysis of inverse-scattering problems in which there is an asymptotic accumulation of an unbounded number of solitons.
6. An interpretation of our asymptotic solution of the Riemann-Hilbert problem in terms of a new variational principle that generalizes the quadratic programming problem of Lax and Levermore and explicitly encodes the contour-selection mechanism. This interpretation also makes a strong connection with approximation theory, where variational problems of the same type appear when one tries to find sets of minimal weighted Green's capacity in the plane.
7. A proof that the systematic selection of an appropriate contour is guaranteed to succeed under certain generic conditions. Finding the correct contour amounts to solving a problem of geometric function theory, namely, the construction of "trajectories of quadratic differentials." We show that the existence of such trajectories is an open condition with respect to the independent variables x and t.

1.3 OUTLINE AND METHOD

We begin in chapter 2 by expressing the function $\psi(x, t; \hbar_N)$ in terms of the solution of a holomorphic matrix Riemann-Hilbert problem posed relative to a contour that surrounds the locus of accumulation of eigenvalues but is otherwise arbitrary a priori. The scattering data is introduced in chapter 3, where we present the formal WKB formulae for initial data satisfying $S(x) \equiv 0$ and appropriate functions $A(x)$. We carry out some detailed asymptotic calculations starting from the WKB approximations to the discrete eigenvalues that we require later in the book, and we compare these general calculations with the specific exact formulas of Satsuma and Yajima. With this WKB data in hand, we proceed in chapter 4 to study the asymptotics of the inverse-scattering problem for this (generally) approximate data. We introduce in §4.1 a certain complex scalar phase function, and in §4.2 we show how to choose it to capture the essentially wild asymptotic behavior of the solution of the Riemann-Hilbert problem. Factoring off a proper choice of the complex phase leads to a simpler Riemann-Hilbert problem whose leading-order asymptotics can be described explicitly. In §4.3 we solve this leading-order Riemann-Hilbert problem, the *outer model problem*, in terms of Riemann theta functions (and in fact for small time in terms of exponentials). Subject to proving the validity of this asymptotic reduction, the solution $\psi(x, t; \hbar_N)$ is then also given at leading order in terms of theta functions and exponentials.

Assuming the existence of the complex phase function on an appropriate contour, we continue with some detailed local analysis in §4.4, building local approximations near certain exceptional points in the complex plane. Patching these local approximations together with the outer approximation yields a uniform approximation of the solution of the Riemann-Hilbert problem that we prove is valid in §4.5.

This detailed error analysis is completely vacuous *unless* we can establish the existence of the complex phase function and its support contour. We carry out this construction in chapter 5 using a modification of the finite-gap ansatz familiar from the Lax-Levermore method. Temporarily tossing out the inequalities that the phase function must ultimately satisfy, we show how to write down equations for the endpoints of the bands and gaps along the contour and how the bands of the contour can be viewed as heteroclinic orbits of a particular explicit differential equation for contours in the complex plane (or trajectories of a quadratic differential). Some of the conditions we impose on the endpoints of the bands and gaps are precisely those that are necessary for the existence of the correct number of heteroclinic orbits. There is a finite-gap ansatz corresponding to any number of bands and gaps, and the idea is to choose this number so that the phase function satisfies certain inequalities as well. This choice then determines the local complexity (genus of the Riemann theta function) of the approximate solution of the initial-value problem (1.1). In §5.3 we show that in fixed neighborhoods of fixed x and t, the macrostructure parameters of the solutions (moduli of an associated hyperelliptic Riemann surface) satisfy a quasilinear system of partial differential equations that we believe to be the elliptic modulation (Whitham) equations for multiphase wavetrains [FL86].

In chapter 6, we investigate the simplest possible ansatz (i.e., genus zero), showing that for small time independent of $\hbar$ it does indeed satisfy all necessary inequalities. For the Satsuma-Yajima initial data, this completes the proof of convergence to the semiclassical limit for small time, ultimately justifying the geometrical optics approximation made by Akhmanov, Sukhorukov, and Khokhlov in 1966. For semiclassical soliton ensembles corresponding to more general data, we still obtain rigorous strong asymptotics, but the connection to initial data is more tenuous. The asymptotics formally recover the initial data and the successful ansatz persists for small time, but our scheme of essentially uniformly approximating the eigenfunction in the complex plane of the eigenvalue breaks down near $t = 0$, when the regions of the complex plane where the description of the eigenfunction requires detailed local analysis come into contact with the locus of accumulation of poles.[1] On the other hand, we know that asymptotics for $t = o(1)$ can be obtained by bypassing the Riemann-Hilbert problem and applying the Lax-Levermore method to the determinant solution formula [EJLM93]. Of course, even if the error is controlled uniformly near $t = 0$, the error present at $t = 0$ in cases where the WKB approximation is not exact can in principle be amplified by this unstable problem in ways that are not possible in "selfadjoint" integrable problems where the semiclassical limit is "hyperbolic." Using a computer program to construct the ansatz for finite times (as opposed to a perturbative calculation based at $t = 0$), we verify the ansatz in the special case of the Satsuma-Yajima data right up to the phase transition to more complicated local behavior termed the "primary caustic" in [MK98]. These computer simulations clearly demonstrate both the selection of the special contour and the breakdown of the ansatz when inequalities fail and/or integral curves of the differential equation determining the contour bands become disconnected. We use perturbation theory in chapter 7 to show that when the genus-zero ansatz fails at the primary caustic, the genus-two ansatz takes over. At such a transition, the smooth wave field "breaks" and gives way to a hexagonal spatiotemporal lattice of maxima.

[1] Note added: This difficulty for t near zero has been surmounted recently using a modification of the methods presented in this book. See [M02] for details.

The conditions that we use to specify the complex phase function are naturally obtained in chapter 8 as the Euler-Lagrange variational conditions describing a particular type of critical point for a certain functional related to potential theory in the upper half-plane. This makes the problem of computing the semiclassical limit equivalent to solving a certain problem of extreme Green's capacity and establishing regularity properties of the solution. Solving the variational problem can be given the physical interpretation of finding unstable electrostatic equilibria of a certain system of electric charges under the influence of an externally applied field which has an attractive component that is exactly the potential of the WKB eigenvalue distribution. The significance of variational problems in the characterization of singular limits of solutions of completely integrable partial differential equations was first established by Lax and Levermore [LL83] for the Korteweg–de Vries equation, and the method was subsequently extended to the Toda lattice [DM98] and the entire hierarchy of the *defocusing* nonlinear Schrödinger equation [JLM99].

The calculations presented in §4.4 and §4.5 rely on certain technical details of the Fredholm theory of Riemann-Hilbert problems posed in Hölder spaces and small-norm theory for Riemann-Hilbert problems in L^2 that we present in the appendices. In particular, the Hölder theory that we summarize unites some very classical results of the Georgian school of Muskhelishvili and others with the treatment of matrix Riemann-Hilbert problems posed on self-intersecting contours given by Zhou [Z89]. The Hölder theory appears to have fallen by the wayside in inverse-scattering applications, possibly because these problems are often posed from the start in L^p or Sobolev spaces. However, in local analysis one is always dealing with explicit piecewise-analytic jump relations on piecewise-smooth contours, and at the same time one requires uniform control on the solutions right up to the contour. In such cases, the compactness required for Fredholm theory comes almost for free (and significantly in a contour-independent way) in Hölder spaces at the cost of an arbitrarily small loss of smoothness. At the same time, once existence is established in a Hölder space, the required control up to the contour is built-in as a property of the solution. On the other hand, in the bigger L^p or Sobolev spaces, compactness depends on a rational approximation argument that can be a lot of work to establish (and in particular it seems that the argument must be tailored for each particular contour configuration). And then having established existence in these spaces, one must put in extra effort to obtain the required control up to the contour, with special care needing to be taken near self-intersection points.

In summary, our primary mathematical techniques include the following.

1. Techniques for the asymptotic analysis of matrix Riemann-Hilbert problems, including the steepest-descent method of Deift and Zhou.
2. The Fredholm theory of Riemann-Hilbert problems in the class of functions with Hölder-continuous boundary values on self-intersecting contours.
3. The use of Cauchy integrals (or Hilbert transforms) to solve certain scalar boundary value problems for sectionally analytic functions in the plane.
4. Careful perturbation theory to establish the semiclassical limit for small times and then to study the phase transition that occurs at a caustic curve in the (x, t)-plane.
5. Some theory of logarithmic potentials with external fields.

1.4 SPECIAL NOTATION

We use several different branches of the logarithm, distinguished one from another by notation. We only use the lowercase $\log(z)$ to refer to a generic branch (cut anywhere) when it makes no difference in an expression, that is, when it appears in an exponent or when its real part is considered. The uppercase $\mathrm{Log}(z)$ always refers to the standard cut of the principal branch, defined for $z \in \mathbb{C}\backslash\mathbb{R}_-$ by the integral

$$\mathrm{Log}(z) := \int_1^z \frac{dw}{w}\,. \tag{1.6}$$

All other branches of the function $\log(\lambda - \eta)$ considered as a function of λ for η fixed are written with notation like $L_\eta^s(\lambda)$. Each of these is also defined for $z = \lambda - \eta$ by (1.6), but with a particular well-defined branch cut in the λ-plane that is associated with the logarithmic pole η and the superscript s. Each of these branches is clearly defined when it first appears in the text. Exponential functions will *always* refer to the principal branch, $a^u = e^{u\,\mathrm{Log}(a)}$.

We use the Pauli matrices throughout the book. They are defined as

$$\sigma_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_2 := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \sigma_3 := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{1.7}$$