Contemporary Mathematics

Riemann-Hilbert Problems with Lots of Discrete Spectrum

Peter D. Miller

This paper is dedicated to Percy Deift, from whom we have all learned so much, in honor of his 60th birthday.

Abstract. We review some recent work on steepest-descent asymptotics for Riemann-Hilbert problems involving a large number of isolated singularities that accumulate in a reasonable fashion in the limit of interest.

1. Introduction

No doubt one of the great successes in the application of integrable systems theory to wave propagation problems was the analysis of Lax and Levermore [**10**] of the zero-dispersion limit of the initial-value problem for the Korteweg-de Vries equation on the line:

$$u_t + uu_x + \epsilon^2 u_{xxx} = 0\,, \qquad x \in \mathbb{R}, \quad t > 0\,, \tag{1.1}$$

subject to ϵ-independent initial data $u(x,0;\epsilon) = u_0(x)$. Here the phrase "zero-dispersion limit" refers to the asymptotic analysis of the solution $u(x,t;\epsilon)$ of this well-posed Cauchy problem in the limit $\epsilon \to 0$.

To be more precise, Lax and Levermore considered positive initial data $u_0(x) > 0$ decaying rapidly to zero as $x \to \pm\infty$ with only one critical point (a local maximum). The first step in the solution of the Cauchy problem by inverse-scattering is then to analyze the stationary Schrödinger equation

$$-6\epsilon^2\psi_{xx} + V(x)\psi = E\psi \tag{1.2}$$

where E is the spectral (eigenvalue) parameter, and $V(x) := -u_0(x)$ is a potential well. The (real) spectrum consists of a continuous part for $E > 0$ and a bounded discrete part for $E < 0$. As ϵ is considered to be small, this is a semiclassical spectral problem, and many facts (originating from various calculations based on the WKB approximation) about the spectral asymptotics were available to Lax and Levermore in their analysis. In particular:

1991 *Mathematics Subject Classification.* Primary 30E25, 37K15, 37K40; Secondary 35Q53, 35Q55, 42C05.

Key words and phrases. Riemann-Hilbert problems, semiclassical asymptotics.

The author gratefully acknowledges the support of the National Science Foundation under a Focused Research Group grant (DMS-0354373) in the preparation of this paper.

- The reflection coefficient for $E > 0$ fixed in the continuous spectrum is "as small in ϵ as V is smooth". In particular, if V is infinitely differentiable, then the reflection coefficient is small beyond all orders: $O(\epsilon^p)$ for all $p > 0$.
- The number $N(\epsilon)$ of discrete eigenvalues (all simple) is large, proportional to ϵ^{-1}. The eigenvalues are approximately located according to the Bohr-Sommerfeld quantization rule: $E_n = E_n^0 + O(\epsilon^2)$ where

$$\Phi(E_n^0) = \pi\epsilon\left(n + \frac{1}{2}\right), \qquad n = 0, 1, 2, \ldots, N(\epsilon) - 1, \tag{1.3}$$

and the phase integral is

$$\Phi(E) := \frac{1}{\sqrt{6}} \int_{x_-(E)}^{x_+(E)} (E - V(s))^{1/2}\, ds\,. \tag{1.4}$$

Here $x_-(E) < x_+(E)$ are the turning points (branches of the inverse function V^{-1}: $V(x_\pm(E)) = E$). The asymptotic number of eigenvalues is

$$N(\epsilon) = \left\lfloor \frac{1}{2} + \frac{1}{\pi\epsilon\sqrt{6}} \int_{-\infty}^{+\infty} \sqrt{-V(x)}\, dx \right\rfloor. \tag{1.5}$$

Lax and Levermore therefore neglected the reflection coefficient, assuming it to be zero. The zero-dispersion limit solution of the Korteweg-de Vries equation in these circumstances is well-approximated as a reflectionless potential, a pure ensemble of (a large number of) solitons.

A multisoliton solution of the Korteweg-de Vries equation (or, for that matter, of practically any integrable equation) is specified from the point of view of the inverse-scattering transform by a collection of discrete eigenvalues and for each, a corresponding norming constant giving information about the eigenfunction. Whether it is formulated via Gel'fand-Levitan-Marcenko equations or via a Riemann-Hilbert problem, the inverse-scattering algorithm for reflectionless (multisoliton) potentials reduces to a problem of $N(\epsilon)$-dimensional linear algebra. Clever symmetrization of the solution (by Cramer's rule) of such a problem leads to the Kay-Moses determinantal formula

$$u(x,t;\epsilon) = 12\epsilon^2 \frac{\partial^2}{\partial x^2} \log(\tau), \quad \text{where} \quad \tau := \det\left(\delta_{mn} + \frac{F_m F_n}{\kappa_m + \kappa_n}\right). \tag{1.6}$$

Here $\kappa_n := \sqrt{-E_n}$ and $F_n := e^{(\kappa_n x - 4\kappa_n^3 t + \beta_n)/\epsilon}$, and the real numbers $\{\beta_n\}$ amount to the norming constants for the problem. A natural approach to the analysis of $u(x,t;\epsilon)$ is then to consider the expansion of τ in principal minors:

$$\tau = 1 + \sum_{\text{subsets } S \text{ of } \{0,\ldots,N-1\}} \det\left(\left.\frac{F_\alpha F_\beta}{\kappa_\alpha + \kappa_\beta}\right|_{\alpha,\beta\in S}\right). \tag{1.7}$$

Lax and Levermore made the crucial observation that each of the terms in this sum is positive and hence argued that the sum would be asymptotically dominated by its largest term. This argument leads to a discrete variational problem that is further approximated by a variational problem for absolutely continuous measures when ϵ is small. By carefully establishing the validity of these approximations and analyzing the variational problem that results, Lax and Levermore showed that $u(x,t;\epsilon)$ has a weak limit $\overline{u}(x,t)$. By extending this analysis along similar lines but keeping track of the details to higher order in ϵ (in particular by keeping some

discreteness in the measures considered in the variational problem), Venakides [**15**] was able to extract strong asymptotic information about $u(x,t;\epsilon)$, including the formation of fully nonlinear oscillations modeled by algebro-geometric multiphase wave local solutions of the Korteweg-de Vries equation.

A prominent component of the work of Percy Deift celebrated in this conference has been the systematic development, together with Xin Zhou and others, of a "steepest-descent" asymptotic technique applicable to matrix Riemann-Hilbert problems involving a small or large parameter. Part of what makes this method attractive is that it provides in a single step the level of accuracy analogous to that which Venakides obtained in the zero-dispersion Korteweg-de Vries problem by continuing the detailed analysis of Lax and Levermore to higher order. In the simplest formulation, a Riemann-Hilbert problem consists of seeking a matrix-valued function $\mathbf{M}$ of a complex variable z that is supposed to (i) be analytic in z away from a system Σ of oriented contour arcs, (ii) satisfy a "jump condition": $\mathbf{M}_+(z) = \mathbf{M}_-(z)\mathbf{V}(z)$ for $z \in \Sigma$ where $\mathbf{M}_+(z)$ and $\mathbf{M}_-(z)$ refer respectively to the boundary values taken on Σ from its left and right and where $\mathbf{V}(z)$ is a "jump matrix" prescribed on Σ, and (iii) have a prescribed value $\mathbf{M}_0$ at a prescribed point z_0 (usually $\mathbf{M}_0 = \mathbb{I}$ and $z_0 = \infty$). The steepest-descent method then involves a sequence of changes of dependent variable whose composition is a linear substitution of the form $\mathbf{M}(z) = \mathbf{E}(z)\hat{\mathbf{M}}(z)$ where $\hat{\mathbf{M}}(z)$ is an explicit global "parametrix" and $\mathbf{E}(z)$ is a new unknown satisfying its own Riemann-Hilbert problem deduced from the original one through the explicit substitution. The steepest-descent method amounts to a systematic construction of the parametrix $\hat{\mathbf{M}}(z)$ and the aim of the construction is to ensure that the error $\mathbf{E}(z)$ satisfies a "small norm" Riemann-Hilbert problem (*i.e.*, one with a near-identity jump matrix and normalization) that can be proved to have a unique solution asymptotically approaching the identity matrix $\mathbb{I}$ in a suitably uniform (with respect to z) sense.

In Riemann-Hilbert problems arising in inverse-scattering applications, the contour Σ is associated with the continuous spectrum of the associated scattering problem, and the jump matrix $\mathbf{V}(z)$ differs from the identity matrix $\mathbb{I}$ only at points $z \in \Sigma$ where the reflection coefficient is nonzero. If there is any discrete spectrum, the simple formulation of the Riemann-Hilbert problem must be augmented; the matrix unknown $\mathbf{M}(z)$ is now permitted to have poles at the points z of the discrete spectrum, and homogeneous conditions involving the norming constants are imposed to relate the principal part of the Laurent expansion of $\mathbf{M}(z)$ at each pole to the regular part. For a typical example of such a condition, consider the 2×2 case and suppose that $\mathbf{M}(z)$ is allowed a simple pole at z_p such that for some constant $c \neq 0$:

$$\operatorname*{Res}_{z=z_p} \mathbf{M}(z) = \lim_{z\to z_p} \mathbf{M}(z) \begin{bmatrix} 0 & c \\ 0 & 0 \end{bmatrix}. \tag{1.8}$$

This imposes the condition that the pole of $\mathbf{M}(z)$ is in the second column only, and the residue of the second column is c times the value at z_p of the (regular) first column. If the singularity at z_p remains separated from the rest of the spectrum in the asymptotic limit of interest, the pole may be removed at the expense of

augmenting the jump contour Σ by making the substitution:

$$\mathbf{N}(z) := \begin{cases} \mathbf{M}(z) \begin{bmatrix} 1 & -c(z-z_p)^{-1} \\ 0 & 1 \end{bmatrix}, & |z-z_p| < \delta \\ \mathbf{M}(z), & |z-z_p| > \delta. \end{cases} \tag{1.9}$$

It is easy to check that $\mathbf{N}(z)$ is then analytic for $z \in \mathbb{C} \setminus \Sigma \cup \Sigma_p$ where Σ_p is a small, counterclockwise-oriented circle about $z = z_p$ of radius δ. Clearly the jump condition for $\mathbf{N}(z)$ then becomes $\mathbf{N}_+(z) = \mathbf{N}_-(z)\mathbf{V}(z)$ for $z \in \Sigma$ and

$$\mathbf{N}_+(z) = \mathbf{N}_-(z) \begin{bmatrix} 1 & -c(z-z_p)^{-1} \\ 0 & 1 \end{bmatrix}, \quad z \in \Sigma_p. \tag{1.10}$$

(In this latter formula, + refers to the boundary value from the inside of the circle, and − to the value from the outside.)

This procedure for removing poles becomes impractical in a spectral situation like that arising in the Lax-Levermore theory due to the asymptotic accumulation of discrete spectrum. Nonetheless, it is of some interest to be able to handle such problems within the framework of the steepest-descent method. In part this is because the steepest-descent method encodes in a unified framework both the first-order (weak) and second-order (strong) asymptotics. This was shown by Deift, Venakides, and Zhou [**5**] by analyzing a zero-dispersion limit problem for the Korteweg-de Vries equation in which assumptions on initial data $u_0(x)$ different from those imposed by Lax and Levermore yield a problem with no discrete spectrum but for which the reflection coefficient is easily calculated in the semiclassical limit. However, it is also of some interest to be able to analyze Riemann-Hilbert problems with accumulating discrete spectrum because there are situations in which such problems arise but where the Lax-Levermore method does not apply. This paper is a review of some recent extensions of the Deift-Zhou steepest-descent method to handle fully discrete Riemann-Hilbert problems like these having no reflection coefficient but (asymptotically) lots of discrete spectrum.

2. Klaus-Shaw Initial Data for the Semiclassical Focusing Nonlinear Schrödinger Equation

A problem that has been the subject of much research in recent years is the semiclassical limit for the focusing nonlinear Schrödinger equation. One is interested in the asymptotic behavior of $\psi = \psi(x,t;\epsilon)$ in the limit $\epsilon \to 0$, where for each $\epsilon > 0$, ψ is the solution of the equation

$$i\epsilon\psi_t + \frac{\epsilon^2}{2}\psi_{xx} + |\psi|^2\psi = 0, \qquad x \in \mathbb{R}, \quad t > 0, \tag{2.1}$$

subject to initial data $\psi(x,0;\epsilon) = A(x)e^{iS(x)/\epsilon}$ where $A(\cdot)$ and $S(\cdot)$ are ϵ-independent real-valued functions. One reason that this problem is interesting is that the formal limit $\epsilon \to 0$ yields a model problem that is ill-posed. Indeed, if one introduces "fluid-dynamical" variables

$$\rho := |\psi|^2, \qquad u := \epsilon\left[\Im\{\log(\psi)\}\right]_x, \tag{2.2}$$

then one may, in the spirit of Madelung's approach to the Schrödinger equation of quantum mechanics as a quantum-corrected fluid motion, write the initial-value

problem without approximation as a coupled system for ρ (fluid density) and u (fluid velocity):

$$\begin{aligned} \rho_t + u\rho_x + \rho u_x &= 0 \\ u_t - \rho_x + uu_x &= \epsilon^2 F[\rho]\,. \end{aligned} \tag{2.3}$$

Here $F[\rho]$ is a certain expression in ρ and its x-derivatives that only depends on ϵ through ρ. This system is to be solved subject to ϵ-independent initial data $\rho(x,0;\epsilon) = A(x)^2$ and $u(x,0;\epsilon) = S'(x)$. It would appear that the limiting dynamics would be governed by the approximate system obtained by simply neglecting the "quantum correction" term $\epsilon^2 F[\rho]$, subject to the same initial data. However, the classical fluid problem that results is an Euler system of gas dynamics with a pressure that is a decreasing function of the density ρ. Physically, this leads one to expect phenomena like spontaneous condensation of the gas. Mathematically, the quasilinear system with $\epsilon^2 F[\rho]$ neglected is of elliptic type, and the Cauchy problem is ill-posed[1]; it can only be solved at all if the initial data functions $A(\cdot)$ and $S(\cdot)$ are analytic at each point $x \in \mathbb{R}$. Such a requirement of analyticity must be regarded as overly restrictive for any physically significant theory. The main question (which remains almost completely unanswered despite much work on the problem) is: if $A(\cdot)$ and $S(\cdot)$ are only required to have sufficient smoothness to guarantee a unique global solution to the initial-value problem for all $\epsilon > 0$ (it is enough for $A(\cdot)e^{iS(\cdot)/\epsilon}$ to lie in a weighted Sobolev space), what can be said about the asymptotic behavior of the solution $\psi(x,t;\epsilon)$ as $\epsilon \to 0$?

For each $\epsilon > 0$, the initial-value problem can be solved by inverse-scattering. The associated spectral problem is the eigenvalue problem for the nonselfadjoint Zakharov-Shabat (or Dirac) operator. The continuous spectrum consists of the real axis in the complex λ-plane, and the discrete spectrum consists of eigenvalues that come in complex-conjugate pairs but generally have no further symmetries. It is easy to see that if $S(\cdot) \equiv 0$, the discrete spectrum is further invariant under inversion through the origin $\lambda = 0$, which makes it symmetric with respect to reflection through the imaginary axis. For many years the literature was filled with faulty statements that certain additional requirements on $A(\cdot)$ made the discrete spectrum purely imaginary, and the question was finally settled by Klaus and Shaw [**9**], who proved that if $A(\cdot)$ is a positive, rapidly decreasing function of x that has a single critical point (a local maximum) then the discrete spectrum lies exactly on the imaginary axis. We refer to this type of initial data ($A(\cdot)$ positive and "single-hump," and $S(\cdot) \equiv 0$) for the focusing nonlinear Schrödinger equation as being of *Klaus-Shaw type*.

Formal WKB calculations can be carried out for the nonselfadjoint Zakharov-Shabat eigenvalue problem with Klaus-Shaw potential, with results quite similar to those used by Lax and Levermore in their analysis of the semiclassical Schrödinger operator. The reflection coefficient is as small in ϵ as $A(\cdot)$ is smooth, uniformly for $\lambda \in \mathbb{R}$ bounded away from $\lambda = 0$. The discrete spectrum consists of $N(\epsilon) = O(\epsilon^{-1})$ eigenvalues accumulating on the positive imaginary axis in the interval $[0, i \max A(x)]$ according to a Bohr-Sommerfeld rule giving approximate eigenvalues

[1]Here one should compare with the corresponding calculation in the theory of the zero-dispersion limit of the Korteweg-de Vries equation. The formal $\epsilon \to 0$ limit of (1.1) is simply the inviscid Burgers equation $u_t + uu_x = 0$ which is hyperbolic, and the Cauchy problem with smooth initial data $u(x,0) = u_0(x)$ is locally (*i.e.*, for t sufficiently small) well-posed.

$\lambda_n \approx \lambda_n^0$ satisfying

$$\Psi(\lambda_n^0) = \pi\epsilon\left(n+\frac{1}{2}\right), \quad n = 0, 1, 2, \ldots, N(\epsilon) - 1, \tag{2.4}$$

where the phase integral is

$$\Psi(\lambda) := \int_{x_-(\lambda)}^{x_+(\lambda)} \left(A(s) + \lambda^2\right)^{1/2}\, ds \tag{2.5}$$

and the radicand is positive exactly in the interval between the two turning points $x_-(\lambda) < x_+(\lambda)$. (Of course there are corresponding approximations for the conjugate eigenvalues on the negative imaginary axis.) The asymptotic number of eigenvalues is

$$N(\epsilon) = \left\lfloor \frac{1}{2} + \frac{1}{\pi\epsilon}\int_{-\infty}^{+\infty} A(x)\, dx \right\rfloor. \tag{2.6}$$

These results suggest an approach completely parallel to that applied to the Korteweg-de Vries equation by Lax and Levermore. Namely, one could neglect the reflection coefficient and consider the asymptotic behavior of the "semiclassical soliton ensemble" given by the reflectionless potential associated with the WKB approximations to the Zakharov-Shabat discrete spectrum. One can derive from inverse scattering a formula analogous to the Kay-Moses formula (see [**16**]):

$$|\psi(x,t;\epsilon)|^2 = \epsilon^2 \frac{\partial^2}{\partial x^2}\log(\tau), \quad \text{where} \quad \tau := \det(\mathbb{I} + \mathbf{B}^*\mathbf{B}), \tag{2.7}$$

where the asterisk denotes componentwise complex conjugation and

$$B_{mn} := \frac{E_m E_n^*}{i(\lambda_m - \lambda_n)}. \tag{2.8}$$

Here $\{\lambda_n\}$ are the eigenvalues in the upper half-plane (on the positive imaginary axis in the Klaus-Shaw case), and $E_n := e^{i(\lambda_n x + \lambda_n^2 t + \beta_n)/\epsilon}$ where $\{\beta_n\}$ are analogues of the norming constants. Despite this strong analogy at the formal level with the zero-dispersion Korteweg-de Vries problem, the Lax-Levermore method grinds to a halt at this point because the principal-minors expansion of τ consists of both positive and negative terms. The discrete Laplace-type argument used by Lax and Levermore simply does not apply to such a sum; indeed the limiting behavior is apparently achieved by a subtle kind of cancellation[2], rather than domination by a single term.

The asymptotics of semiclassical soliton ensembles of the type described above were considered by Kamvissis, McLaughlin, and Miller [**8**] from the point of view of the formulation of inverse-scattering as a Riemann-Hilbert problem. More precisely, in [**8**], Klaus-Shaw type initial data is considered with three further assumptions:

- The parameter ϵ is restricted to a discrete sequence of values tending to zero:

$$\epsilon = \epsilon_N := \frac{1}{\pi N}\int_{-\infty}^{+\infty} A(x)\, dx, \quad N = 1, 2, 3, \ldots. \tag{2.9}$$

[2]In fact, it is an open problem to interpret what is now understood (by means of methods other than analyzing τ) about the asymptotic behavior of this initial-value problem at the level of the expansion of τ.

This has the effect of making the reflection coefficient small uniformly for all $\lambda \in \mathbb{R}$ (one does not have to delete a neighborhood of the origin). Obviously, for such $\epsilon = \epsilon_N$ we have $N(\epsilon) = N$.

- The potential $A(\cdot)$ is assumed to be an even function of x. This allows one to deduce from formal WKB arguments a simple formula for the proportionality constants (related to norming constants): $\gamma_n^0 := (-1)^{n+1}$, $n = 0, \ldots, N-1$.
- The potential $A(\cdot)$ is assumed to be an analytic function of x. In such a situation one can find a solution (by the Cauchy-Kovaleskaya method) of the limiting elliptic model system. This fact is of no particular importance in the method of analysis used in [**8**] however; of crucial importance is rather the fact that the phase integral $\Psi(\lambda)$ is an analytic function of λ in the interior of the imaginary interval of existence of the eigenvalues.

Neglecting the reflection coefficient and taking the WKB eigenvalues $\{\lambda_n := \lambda_n^0\}$ and proportionality constants $\{\gamma_n := \gamma_n^0\}$ as exact spectral data, the Riemann-Hilbert problem of inverse scattering takes the following form. Let

$$c_n(x,t) := \frac{1}{\gamma_n} \operatorname*{Res}_{\lambda=\lambda_n} W(\lambda)\,, \qquad W(\lambda) := e^{2i(\lambda x + \lambda^2 t)/\epsilon} \prod_{n=0}^{N-1} \frac{\lambda - \lambda_n^*}{\lambda - \lambda_n}\,, \tag{2.10}$$

and seek a 2×2 matrix $\mathbf{m}(\lambda)$, $\lambda \in \mathbb{C}$, with the following properties:

Rationality: $\mathbf{m}(\lambda)$ is a rational function of λ with simple poles confined to $\{\lambda_n, \lambda_n^*\}_{n=0}^{N-1}$ such that for $n = 0, \ldots, N-1$:

$$\operatorname*{Res}_{\lambda=\lambda_n} \mathbf{m}(\lambda) = \lim_{\lambda\to\lambda_n} \mathbf{m}(\lambda) \begin{bmatrix} 0 & 0 \\ c_n(x,t) & 0 \end{bmatrix}, \tag{2.11}$$

$$\operatorname*{Res}_{\lambda=\lambda_n^*} \mathbf{m}(\lambda) = \lim_{\lambda\to\lambda_n^*} \mathbf{m}(\lambda) \begin{bmatrix} 0 & -c_n(x,t)^* \\ 0 & 0 \end{bmatrix}. \tag{2.12}$$

Normalization: The matrix $\mathbf{m}(\lambda)$ is normalized at infinity in the sense that

$$\lim_{\lambda\to\infty} \mathbf{m}(\lambda) = \mathbb{I}\,. \tag{2.13}$$

From the solution of this Riemann-Hilbert problem the semiclassical soliton ensemble supposed to model the solution of the initial-value problem is given by

$$\psi(x,t;\epsilon) = 2i \lim_{\lambda\to\infty} \lambda m_{12}(\lambda)\,. \tag{2.14}$$

This is not a Riemann-Hilbert problem in the traditional sense: there is no jump discontinuity of the matrix unknown across any contour. Indeed, it seems at first to be quite simple. One could solve for $\mathbf{m}(\lambda)$ by using a partial-fractions ansatz:

$$\mathbf{m}(\lambda) = \mathbb{I} + \sum_{n=0}^{N-1} \frac{\mathbf{a}_n}{\lambda - \lambda_n} + \sum_{n=0}^{N-1} \frac{\mathbf{b}_n}{\lambda - \lambda_n^*}\,, \tag{2.15}$$

where $\mathbf{a}_n$ and $\mathbf{b}_n$ are 2×2 matrix-valued coefficients to be determined. This ansatz clearly satisfies the required normalization condition (2.13), and it is a rational function of λ with simple-pole singularities at all allowed points. Substitution of the ansatz into the residue conditions (2.11) and (2.12) that $\mathbf{m}(\lambda)$ is required to satisfy leads directly to a linear algebra problem governing the matrix elements of the coefficients $\mathbf{a}_n$ and $\mathbf{b}_n$. It is, however, not advisable to proceed in this manner,

since the determinant of this linear algebra problem is the function τ encountered earlier, and a direct asymptotic evaluation of this determinant (at least via the Lax-Levermore method) leads to a dead end.

On the other hand, the system of linear equations that results from the partial-fractions approach can be solved numerically for fixed N by independent calculations carried out on a grid of (x,t)-values. While the condition numbers of the matrices involved become large very rapidly with N, calculations like this have been carried out (see [**13**] and [**11**]) for N sufficiently large to discern marvelous patterns in the space-time structure of the solution $\psi(x,t;\epsilon)$, and these experiments have proven to be useful both in inspiring curiosity and suggesting new phenomena to be verified by subsequent rigorous asymptotic analysis. An image from such an experiment is shown in Figure 1. From these numerical calculations, one learns to expect

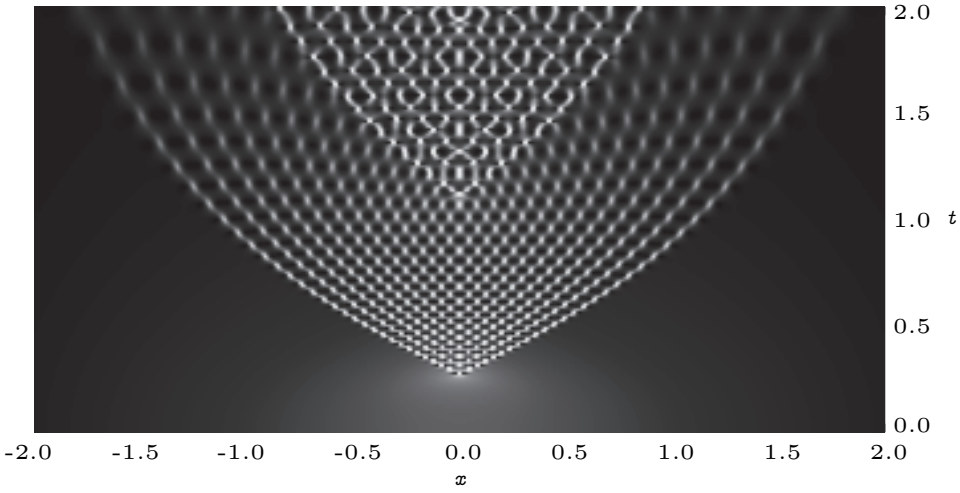

FIGURE 1. A density plot of the semiclassical soliton ensemble corresponding to $A(x) = 2\,\mathrm{sech}(x)$ and $\epsilon = 2/N$ with $N = 40$. Brighter areas correspond to larger values of $|\psi|^2$ (black corresponds to $|\psi|^2 = 0$). The apparent phase transition that begins near $t = 0.25$ is called the primary caustic, and that beginning near $t = 1.15$ is called the secondary caustic.

asymptotically sharp *caustic curves* to separate regions of space-time containing qualitatively different modulated oscillations having wavelengths and frequencies proportional to ϵ. A successful analysis of the semiclassical limit must correctly predict such salient features of the dynamics.

To begin such an analysis, we avoid dealing with the determinant τ by working instead with the conditions of the Riemann-Hilbert problem for $\mathbf{m}(\lambda)$. The key idea we need to begin to apply the Deift-Zhou steepest-descent method to such a completely discrete problem is to remove the poles from the problem at the expense of jump discontinuities across contours. As mentioned in the introduction it is not feasible to remove the poles "one at a time." Instead we remove them all at once by finding an analytic function that interpolates the proportionality constants $\{\gamma_n\}$ at the corresponding eigenvalues $\{\lambda_n\}$. It is easy to see that such a function can be

built from the WKB phase integral; indeed we may observe that as a consequence of the eigenvalue condition (2.4),

$$\gamma_n = -ie^{-i\Psi(\lambda_n)/\epsilon}\,, \qquad n = 0, 1, 2, \ldots, N-1\,. \tag{2.16}$$

Therefore, the poles may be removed from the problem by choosing a fixed Jordan contour Σ lying in the upper half-plane that contains all of the positive-imaginary eigenvalues in its interior (such a contour necessarily meets the origin because the eigenvalues are accumulating there from above) and defining a new matrix unknown $\mathbf{M}(\lambda)$ relative to Σ by setting

$$\mathbf{M}(\lambda) := \mathbf{m}(\lambda) \begin{bmatrix} 1 & 0 \\ iW(\lambda)e^{i\Psi(\lambda)/\epsilon} & 1 \end{bmatrix}\,, \quad \text{inside } \Sigma\,, \tag{2.17}$$

and outside of Σ with $\Im(\lambda) > 0$, $\mathbf{M}(\lambda) := \mathbf{m}(\lambda)$. To define $\mathbf{M}(\lambda)$ for $\Im(\lambda) < 0$, set

$$\mathbf{M}(\lambda) := \sigma_2 \mathbf{M}(\lambda^*)^* \sigma_2\,, \quad \text{where} \quad \sigma_2 := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}\,. \tag{2.18}$$

This transformation preserves a natural complex-conjugation symmetry enjoyed by the original matrix $\mathbf{m}(\lambda)$.

It is easy to check that the matrix $\mathbf{M}(\lambda)$ is analytic at each of the poles $\{\lambda_n\}$ and their complex conjugates. The only points of nonanalyticity of $\mathbf{M}(\lambda)$ are along the contour Σ and its complex conjugate where $\mathbf{M}(\lambda)$ experiences jump discontinuities. It follows that the matrix $\mathbf{M}(\lambda)$ satisfies a Riemann-Hilbert problem of the more traditional type, and from here on one expects some version of the steepest-descent method of Deift and Zhou to apply.

The jump condition satisfied by $\mathbf{M}(\lambda)$ across the contour Σ involves the function $W(\lambda)$ evaluated on this contour. Noting that $W(\lambda)$ may be expressed as an exponential

$$W(\lambda) = \exp\left(\frac{1}{\epsilon}\left[2i\lambda x + 2i\lambda^2 t + \sum_{n=0}^{N-1} \epsilon \log(\lambda - \lambda_n^*) - \sum_{n=0}^{N-1} \epsilon \log(\lambda - \lambda_n)\right]\right)\,, \tag{2.19}$$

we may consider a continuum limit in which the (Riemann) sums are approximated by integrals to write $W(\lambda)$ in the form $W(\lambda) = S(\lambda)e^{F(\lambda)/\epsilon}$ where $F(\lambda)$ is a function independent of ϵ whose analytic continuation from Σ to its interior has a logarithmic branch cut in the imaginary interval where there once were accumulating poles, and where $S(\lambda) = 1 + O(\epsilon)$ for fixed nonzero $\lambda \in \Sigma$. (The size of the error is partially a consequence of the choice of a discrete sequence of $\epsilon = \epsilon_N$, which makes the Riemann sums "midpoint rule" approximants of the corresponding integrals.)

We should view the freedom of choice of the contour Σ as being analogous to the choice of contour in evaluating exponential integrals with analytic integrands by the (classical) method of steepest descent. We want to choose Σ to make the jump matrix for the problem have convenient properties for analysis. We must keep in mind, however, that whatever mechanism we work out to prescribe the "correct" location for the contour Σ must be compatible with our original assumption: namely that Σ should encircle all of the eigenvalues of the problem.

One continues with the Deift-Zhou method by attempting to stabilize the Riemann-Hilbert problem for $\mathbf{M}(\lambda)$ with the help of a "g-function", a complex potential first introduced by Deift, Venakides, and Zhou [**5**] to encorporate the measure solving the Lax-Levermore variational problem into the framework of the

Riemann-Hilbert problem of inverse scattering for the Korteweg-de Vries equation. To apply the method here, one assumes that $g(\lambda)$ is a function analytic in $\mathbb{C}\setminus\Sigma\cup\Sigma^*$ with $g(\lambda)\to 0$ as $\lambda\to\infty$ and satisfying the symmetry condition $g(\lambda)+g(\lambda^*)^*=0$. With the help of any such function $g(\lambda)$, one may define a new matrix unknown $\mathbf{N}(\lambda)$ by setting

$$\mathbf{N}(\lambda) := \mathbf{M}(\lambda)\begin{bmatrix} e^{-g(\lambda)/\epsilon} & 0 \\ 0 & e^{g(\lambda)/\epsilon}\end{bmatrix}. \tag{2.20}$$

The jump condition to be satisfied by $\mathbf{N}(\lambda)$ on Σ is easily seen to be

$$\mathbf{N}_+(\lambda) = \mathbf{N}_-(\lambda)\begin{bmatrix} e^{i\theta(\lambda)/\epsilon} & 0 \\ iS(\lambda)e^{\phi(\lambda)/\epsilon} & e^{-i\theta(\lambda)/\epsilon}\end{bmatrix}, \qquad \lambda\in\Sigma, \tag{2.21}$$

where $\theta(\lambda) := i(g_+(\lambda)-g_-(\lambda))$ and $\phi(\lambda) := F(\lambda)-g_+(\lambda)-g_-(\lambda)$, and where the subscript $+$ (respectively $-$) refers to a boundary value taken from inside (respectively outside) the contour Σ.

The jump condition for $\mathbf{N}(\lambda)$ involves the undetermined function $g(\lambda)$ in an explicit way. It also involves the geometry of the contour Σ in a more subtle way, through the small ϵ behavior (rapidly oscillatory, exponentially large, or exponentially small) of the analytic jump matrix elements when restricted to the contour. The fundamental principle is to try to choose the scalar function $g(\lambda)$ and the contour Σ to make the jump matrix as simple as possible in the limit $\epsilon\to 0$.

One might expect that the identity matrix would be the best possible target, but this turns out to be unachievable. The best that one can settle for is a piecewise constant asymptotic behavior for the jump matrix, which turns out to be just good enough. Therefore, we try to choose $g(\lambda)$ and the contour Σ so that Σ splits into two kinds of intervals:

Bands: in which $\phi(\lambda)$ is an imaginary constant, and $\theta(\lambda)$ is real decreasing in the positive (counterclockwise) direction[3].

Gaps: in which $\theta(\lambda)$ is a real constant, and $\Re\{\phi(\lambda)\}<0$.

Then (skipping many steps) $\mathbf{N}(\lambda)$ can be approximately built (for small ϵ) from Riemann Θ functions of the hyperelliptic surface over the complex λ-plane with branch points at the band endpoints. There are two important implications:

- The semiclassical asymptotics of $\psi(x,t;\epsilon)$ in between the caustic curves are precisely described by modulated multiphase waves written in terms of Θ, where the number of phases is related to the genus of the surface.
- Caustic curves in the (x,t)-plane are genus transitions.

Analysis of caustics therefore boils down to finding $g(\lambda)$ and the contour Σ, and determining the way the number of bands and gaps varies with x and t. In some problems (*e.g.* zero-dispersion Korteweg-de Vries, orthogonal polynomials) the conditions that determine $g(\lambda)$ are encoded in a convex variational principle. Such a problem corresponds physically to determining the distribution of a unit charge in electrostatic equilibrium in the presence of an external field. The genus corresponds to the number of components of support of the equilibrium charge distribution. The support of the equilibrium distribution is exactly the union of the bands defined above. For the focusing nonlinear Schrödinger problem, there is

[3]The jump matrix for $\mathbf{N}$ in a band is rapidly oscillatory on the diagonal, and an additional factorization and deformration in a lens-shaped region about the band is required to obtain piecewise-constant asymptotics of the jump matrix.

a variational formulation for $g(\lambda)$, but the problem is nonconvex, so for detailed calculations it is more lucrative to construct $g(\lambda)$ by ansatz. The ansatz-based method for finding $g(\lambda)$ consists of the following steps. First, one makes a guess of an even (due to conjugation symmetry in the complex λ-plane) genus G. Then it turns out that for each configuration of $2G+2$ complex band/gap endpoints (in conjugate pairs), $g'(\lambda)$ is necessarily given by a well-defined Cauchy-type integral formula. One then enforces on this formula the band/gap conditions on θ and ϕ. This has two effects. Firstly, the endpoints are implicitly determined as functions of x and t through a system of nonlinear algebraic equations. Secondly, certain inequalities become evident that must be satisfied if the correct genus (for some fixed x and t) is indeed G.

The problem of finding $g(\lambda)$ may therefore be approached computationally. By contrast with direct numerical simulation of the focusing nonlinear Schrödinger equation for small ϵ (a notoriously stiff and unstable problem), or use of numerical linear algebra to carry out inverse-scattering for large N (a problem hindered by rapidly growing condition numbers), the conditions determining g are independent of any large or small parameter. This makes the problem of finding $g(\lambda)$ especially attractive for numerical computation. Indeed, finding the endpoints from the nonlinear equations that implicitly define them as functions of x and t amounts to root-finding, and once the endpoints are known, the band intervals of Σ are determined by solving simple ordinary differential equations to determine paths in the complex plane connecting pairs of endpoints. The (x,t)-plane can then be explored to search for genus transitions (caustics); the task is to seek topological changes in the region of the complex λ-plane where the inequality $\Re\{\phi\} < 0$ allows the gaps to reside as x and t are varied. (Unlike the bands, whose locations are precisely specified by the procedure, the gaps are constrained only by the inequality $\Re\{\phi\} < 0$ and the condition that with the bands they complete a loop contour Σ encircling the eigenvalues in the upper half-plane.)

With this kind of analysis, it is shown in [**8**] that the $G = 0$ ansatz captures the semiclassical dynamics as long as $t \neq 0$ and $|t|$ is sufficiently small but independent of ϵ. The mechanism for the formation of the primary caustic curve turns out to be the "pinching off" of the region where $\Re\{\phi\} < 0$, and it is shown in [**8**] that by passing from genus $G = 0$ to $G = 2$ with the insertion of a new small band near the pinch-off point, the genus $G = 2$ ansatz succeeds in describing the semiclassical dynamics just beyond the primary caustic curve.

The analysis exactly at $t = 0$ requires a modification of the techniques briefly described above because the ansatz-based theory of the g-function predicts that the single band present for $G = 0$ should coincide with an interval of the imaginary axis; in other words, the contour Σ no longer surrounds the eigenvalues. A technique for solving the problem exactly at $t = 0$ was proposed in [**12**]. This technique is based on generalizing the interpolation formula (2.16) and in fact making use of *two simultaneous interpolants* of the proportionality constants $\{\gamma_n\}$ at the eigenvalues $\{\lambda_n\}$. Perhaps such a complicated construction seems at this point to be a technical inconvenience, but it turns out to be a fundamentally important idea in several other problems, including the prediction of the secondary caustic curve (as is evidently present in the dynamics shown in Figure 1). But before discussing that application, we digress briefly to consider another problem that is simpler in that the location

of the contour Σ is never in question, but for which the dual interpolant idea is necessary from the start.

3. Discrete Orthogonal Polynomials

Probably the most far-reaching application of the Deift-Zhou steepest-descent method to date is the analysis by Deift, Kriecherbauer, McLaughlin, Venakides, and Zhou [**3**] of the asymptotic behavior of orthogonal polynomials with exponentially varying weights. Here one considers the large-N asymptotics of the polynomial of degree N in the system of orthonormal polynomials on the real line with respect to a weight function (also involving N, hence a "varying weight") of the form $w(x) = e^{-NV(x)}$ for some analytic function V increasing rapidly enough for large $|x|$. This problem is of great relevance in both classical approximation theory and the mathematical physics of random matrix theory, but it remained unsolved until the publication of [**3**], at least in terms of the degree of generality of the potential $V(x)$ and of the remarkable precision of the asymptotics obtained. What made this work possible was a combination of two things: the observation by Fokas, Its, and Kitaev [**7**] that the orthogonal polynomial of degree N on $\mathbb{R}$ can be obtained via the solution of a simple-looking Riemann-Hilbert problem with the real line as the contour of discontinuity, and the availability of a recently developed tool (the steepest-descent method) for the asymptotic analysis of such problems.

In some related applications, including the continuum limit of the Toda lattice and the statistical combinatorics of certain planar tilings, one is interested also in the large degree asymptotics of systems of orthogonal polynomials where the measure of orthogonality consists of a finite but large number N of point masses of different weights. There are exactly N independent *discrete orthogonal polynomials*, and in the applications mentioned above one is interested in the asymptotic behavior as $N \to \infty$ of the polynomial of degree $k = cN$, where c is a fixed number in $(0, 1)$. This problem was solved in [**1**] by relating the discrete orthogonal polynomials to a fully discrete Riemann-Hilbert problem that is remarkably similar to that occurring in the theory of the semiclassical focusing nonlinear Schrödinger equation.

To be more precise, in [**1**] one considers polynomials orthogonal with respect to the discrete measure

$$\mu(x) = \sum_{n=0}^{N-1} w_{N,n}\delta(x - x_{N,n}) \tag{3.1}$$

where the nodes of orthogonality $\{x_{N,n}\}$ are prescribed for each N by giving a fixed real analytic function $\rho^0(x) > 0$ on an interval $[a, b]$ with

$$\int_a^b \rho^0(x)\,dx = 1 \tag{3.2}$$

and then imposing the relation

$$\int_a^{x_{N,n}} \rho^0(x)\,dx = \frac{2n+1}{2N}, \qquad n = 0, 1, 2, \ldots, N-1 \tag{3.3}$$

which resembles a Bohr-Sommerfeld quantization rule. The positive weights $\{w_{N,n}\}$ are assumed to be of the form

$$w_{N,n} = e^{-NV(x_{N,n})} \prod_{\substack{m=0 \\ m\neq n}}^{N-1} |x_{N,n} - x_{N,m}|^{-1} \tag{3.4}$$

for some real analytic function $V(x)$ that is allowed to vary with N only very little if at all. Let $\mathbb{Z}_N := \{0, 1, 2, \ldots, N-1\}$ and choose a subset $\Delta \subset \mathbb{Z}_N$ with cardinality $0 \leq \#\Delta \leq N$. Let ∇ denote the complementary subset in $\mathbb{Z}_N$: $\nabla = \mathbb{Z}_N \setminus \Delta$. The role of the subset Δ is related to the properties of the appropriate equilibrium measure corresponding to $\rho^0(\cdot)$, $V(\cdot)$, and the ratio $c = k/N$, and goes beyond the scope of this brief discussion. Finally, set

$$W(z) := e^{-NV(z)} \frac{\displaystyle\prod_{n\in\Delta} (z - x_{N,n})}{\displaystyle\prod_{n\in\nabla} (z - x_{N,n})} \,. \tag{3.5}$$

The discrete orthogonal polynomial of degree k may be obtained via the solution of the following (fully discrete) Riemann-Hilbert problem. Seek a 2×2 matrix $\mathbf{Q}(z)$, $z \in \mathbb{C}$, with the following properties:

Rationality: $\mathbf{Q}(z)$ is a rational function of z with simple poles confined to the nodes $\{x_{N,n}\}_{n=0}^{N-1}$ such that

$$\operatorname*{Res}_{z=x_{N,n}} \mathbf{Q}(z) = \lim_{z\to x_{N,n}} \mathbf{Q}(z) \begin{bmatrix} 0 & (-1)^{N-1-n} \operatorname*{Res}_{\zeta=x_{N,n}} W(\zeta) \\ 0 & 0 \end{bmatrix}, \quad n \in \nabla \,, \tag{3.6}$$

and

$$\operatorname*{Res}_{z=x_{N,n}} \mathbf{Q}(z) = \lim_{z\to x_{N,n}} \mathbf{Q}(z) \begin{bmatrix} 0 & 0 \\ (-1)^{N-1-n} \operatorname*{Res}_{\zeta=x_{N,n}} W(\zeta)^{-1} & 0 \end{bmatrix}, \quad n \in \Delta \,. \tag{3.7}$$

Normalization: The matrix $\mathbf{Q}(z)$ is normalized at infinity in the sense that

$$\lim_{z\to\infty} \mathbf{Q}(z) \begin{bmatrix} z^{\#\Delta - k} & 0 \\ 0 & z^{k-\#\Delta} \end{bmatrix} = \mathbb{I} \,. \tag{3.8}$$

From the solution of this Riemann-Hilbert problem, the monic discrete orthogonal polynomial of degree $k < N$ is given by

$$\pi_{N,k}(z) = z^k + \cdots = Q_{11}(z) \prod_{n\in\Delta} (z - x_{N,n}) \,. \tag{3.9}$$

The first step in the analysis of such a fully discrete Riemann-Hilbert problem is to remove the poles, and it should be clear based on the discussion of the inverse-scattering problem for the focusing nonlinear Schrödinger equation in Section 2 that the key to the problem is to find an analytic function that interpolates the signs $\{(-1)^{N-1-n}\}$ at the nodes $\{x_{N,n}\}$. Setting

$$\theta^0(z) := 2\pi \int_z^b \rho^0(s)\, ds \,, \tag{3.10}$$

(this is an analogue of the phase integral of WKB theory) it is not difficult to see from the definition of the nodes $\{x_{N,n}\}$ that the identities

$$ie^{-iN\theta^0(x_{N,n})/2} = (-1)^{N-1-n} \quad \text{and} \quad -ie^{iN\theta^0(x_{N,n})/2} = (-1)^{N-1-n} \tag{3.11}$$

both hold for $N = 1, 2, 3, \ldots$ and $n \in \mathbb{Z}_N$. Therefore, there are clearly (at least) two analytic interpolants, and the question that arises in connection with the approach in Section 2 is: which one should be used to remove the poles?

The answer is that we need both of the interpolants. The discrete orthogonal polynomial problem has an associated convex variational problem with a unique equilibrium measure, and the support of this measure is a subset of the interval $[a, b]$ where the nodes of orthogonality are accumulating as $N \to \infty$. So, if one were to choose a single interpolant and remove the poles from within a single closed contour as in the approach of Section 2, one should expect to be led to allow the contour to coincide, at least in part, with some subsets of the interval $[a, b]$. Such a location for the contour would not be consistent with the role of the contour in containing the poles of the discrete Riemann-Hilbert problem. The discrete orthogonal polynomial problem is therefore quite like the special case of the semiclassical focusing nonlinear Schrödinger problem exactly at $t = 0$, and to solve that problem one may use two distinct analytic interpolants [**12**].

Let us describe how this dual-interpolant procedure works. In practice, the sets Δ and ∇ are both taken to consist of unions of contiguous sequences of indices of lengths proportional to N, and more precisely the nodes $x_{N,n}$ with $n \in \Delta$ lie in a finite union Σ_0^Δ of N-independent subintervals of $[a, b]$ (and the nodes $x_{N,n}$ with $n \in \nabla$ lie in the complementary system Σ_0^∇ of subintervals). Let $\Omega_\pm^{\Delta,\nabla}$ be unions of rectangular domains of the z-plane defined by the inequalities $\Re(z) \in \Sigma_0^{\Delta,\nabla}$, $|\Im(z)| < h$ for some $h > 0$ independent of N, and $\pm\Im(z) > 0$. Consider using both analytic interpolants of $(-1)^{N-1-n}$ given in (3.11) to define a new matrix unknown $\mathbf{R}(z)$ as follows:

$$\mathbf{R}(z) := \mathbf{Q}(z) \begin{bmatrix} 1 & \mp i e^{\mp iN\theta^0(z)/2} W(z) \\ 0 & 1 \end{bmatrix}, \quad \text{for } z \in \Omega_\pm^\nabla, \tag{3.12}$$

$$\mathbf{R}(z) := \mathbf{Q}(z) \begin{bmatrix} 1 & 0 \\ \mp i e^{\mp iN\theta^0(z)/2} W(z)^{-1} & 1 \end{bmatrix}, \quad \text{for } z \in \Omega_\pm^\Delta, \tag{3.13}$$

and for all other $z \in \mathbb{C}$ set $\mathbf{R}(z) := \mathbf{Q}(z)$. The main observations to make at this point are the following:

- The matrix $\mathbf{R}(z)$ takes continuous boundary values on the contour Σ illustrated in Figure 2 from both sides. In particular, there are no singularities of the boundary values taken by $\mathbf{R}(z)$ on the interval $[a, b]$ despite the fact that $\mathbf{Q}(z)$ has (many) poles there.
- Let $\mathbf{V}(z)$ denote the jump matrix relating the boundary values of $\mathbf{R}(z)$ from the upper and lower half-planes via $\mathbf{R}_+(z) = \mathbf{R}_-(z)\mathbf{V}(z)$, $z \in (a, b)$. Then $\mathbf{V}(z)$ is analytic in the interior of Σ_0^Δ and of Σ_0^∇.

The point is that the matrix $\mathbf{R}(z)$ satisfies a Riemann-Hilbert problem of the more traditional type, with a piecewise analytic unknown matrix satisfying jump conditions across a given system Σ of contours. Moreover, part of the contour Σ is the interval $[a, b]$ itself, and this is exactly where the support of the associated equilibrium measure lies.

In [**1**], the equilibrium measure is used to concoct an appropriate g-function for the Riemann-Hilbert problem for $\mathbf{R}(z)$, after which the steepest-descent method of Deift and Zhou is used to determine very precise asymptotic information about the discrete orthogonal polynomials and associated quantities like three-term recurrence

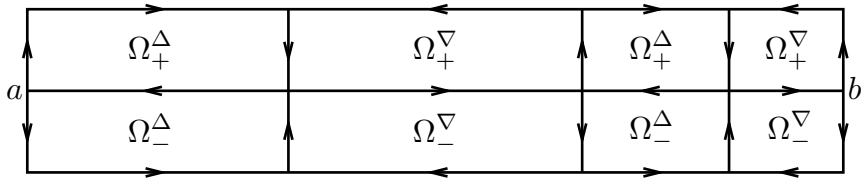

FIGURE 2. The oriented contour Σ for removing the poles from the discrete orthogonal polynomial Riemann-Hilbert problem.

coefficients. The results of this analysis have been applied to a number of problems including the refinement (from weak to strong asymptotics) of the analysis of Deift and McLaughlin [**4**] on the continuum limit of the Toda lattice as well as the proof of the Tracy-Widom law for fluctuations of the boundary of the "frozen" region in random rhombus tilings of large hexagons. While the Lax-Levermore method can also be applied to this fully discrete problem, the dual interpolant approach and the steepest-descent method have proven to be indispensible for establishing all of these results in a unified fashion.

4. Multiple Interpolants and Caustics

Now let us return to the consideration of the semiclassical limit for the focusing nonlinear Schrödinger equation with initial data of Klaus-Shaw type. In Section 2 it was noted that after getting beyond the initial instant of $t = 0$ and at least until a time beyond the formation of the primary caustic at $t = T_1(x)$ (a transition from genus zero to genus two), a single analytic interpolant of the proportionality constants $\{\gamma_n\}$ at the eigenvalues $\{\lambda_n\}$ suffices to convert the fully discrete Riemann-Hilbert problem of inverse-scattering into one of the more traditional type that, crucially, can be analyzed by the steepest-descent method. In other words, whenever $0 < t < T_1(x) + \delta$, removal of the poles yields a problem for which g-function can be found relative to a contour Σ that indeed encircles all of the eigenvalues in the upper half-plane. (These statements actually hide some complicated details. For example, while one interpolant suffices, different interpolants may be required for different x and t. Also, to have a problem well-conditioned for semiclassical asymptotics, for some x and t it is additionally necessary to formulate a different but equivalent fully discrete Riemann-Hilbert problem of inverse-scattering by steps analogous to making the choice of $\Delta = \mathbb{Z}_N$ rather than $\Delta = \emptyset$ in the Riemann-Hilbert problem for the discrete orthogonal polynomials considered in Section 3.)

In the recent paper [**11**], it is shown that as t increases further toward the apparent secondary caustic $t = T_2(x) > T_1(x)$ clearly visible in the plot in Figure 1, it becomes necessary once again to use multiple interpolants to remove the poles from the problem. More importantly, it is shown that the mechanism generating the secondary caustic is the failure of a new variational inequality generalizing the condition $\Re\{\phi\} < 0$ and made necessary exactly by the presence in the analysis of an additional interpolant.

Figure 3 shows the results of using the ansatz-based procedure described in Section 2 with $G = 2$ at two different times with x fixed. The left-hand frame corresponds to a time t just beyond the primary caustic at $t = T_1(x)$. The wavy line is the locus of accumulation of the eigenvalues, which becomes a logarithmic

branch cut as far as the conditions to determine $g(\lambda)$ are concerned. The solid arcs are the two bands of the contour Σ, which when completed with two gap arcs lying in the shaded regions where $\Re\{\phi\} < 0$ completely encircles the eigenvalues in the upper half λ-plane as desired. The right-hand frame corresponds to a larger value of t, and it is clear that something is about to go terribly wrong with the approach because the conditions determining $g(\lambda)$ have driven the band labeled I_1 to collide with the branch cut.

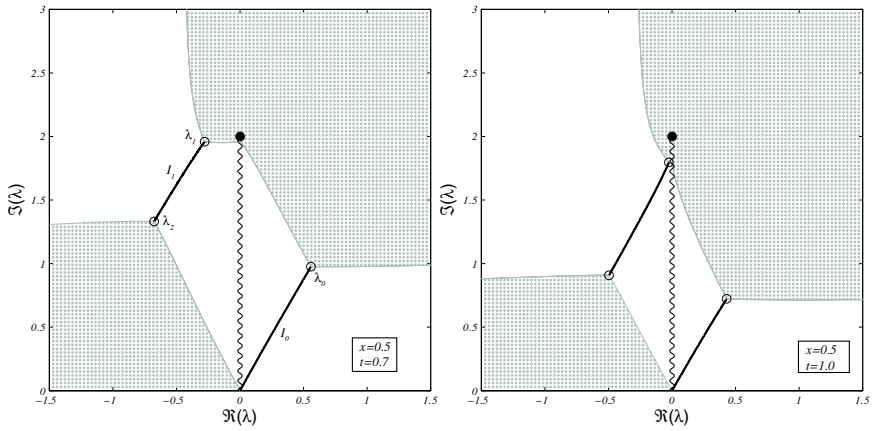

FIGURE 3. Computed bands and the inequality $\Re\{\phi\} < 0$ for the genus two ansatz as described in Section 2. The band endpoints labeled λ_0, λ_1 and λ_2 should not be confused with the discrete eigenvalues, all of which lie on the imaginary axis.

From one point of view, the branch cut is an artificial obstruction to analytic continuation, and it could perhaps easily be deformed out of the way of the moving band I_1. The contour Σ created from the bands I_0 and I_1 and the intervening gaps would, however, no longer encircle all of the eigenvalues $\{\lambda_n\}$ and so some poles of the original discrete Riemann-Hilbert problem of inverse-scattering would not have been removed at the start!

The interpolation formula (2.16) generalizes to a family of interpolants indexed by an integer j in the following way:

$$\gamma_n = -i(-1)^j e^{-i(2j+1)\Psi(\lambda_n)/\epsilon}, \qquad n = 0, 1, 2, \ldots, N-1\,. \tag{4.1}$$

(The interpolant considered in Section 2 corresponds to $j = 0$.) To repair the bad situation about to occur on the right-hand frame of Figure 3, consider the sequence of steps illustrated in Figure 4. It turns out that the introduction of a new contour lobe and the use of a new interpolant does not significantly change the conditions satisfied by g on the common boundary of the $j = 0$ and $j = 1$ regions. However, the new contour lobe, which contains no bands, must lie within a region in which an inequality different from $\Re\{\phi\} < 0$ holds. The inequality to be satisfied on the new contour lobe corresponds to adding to ϕ a certain multiple of the WKB phase integral $\Psi(\lambda)$.

In Figure 5, the numerically computed evolution of the g-function is continued for larger t, using the additional interpolant formulation described briefly above.

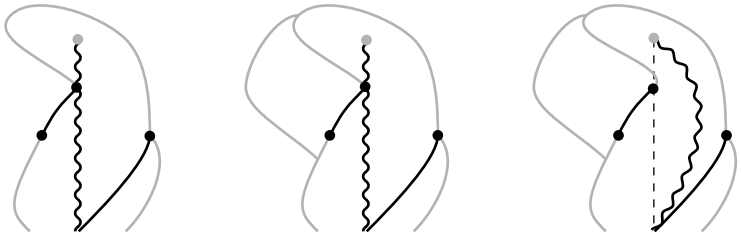

FIGURE 4. Left: the loop contour Σ in the upper half-plane joining the origin to the negative real axis and (barely) encircling the eigenvalues that are condensed (as far as the conditions on g are concerned) into a branch cut illustrated with a wavy line. Within Σ the interpolant corresponding to $j = 0$ is used, and (for technical reasons not relevant to the current discussion, see [**11**]) in the smaller region in the right half-plane the interpolant corresponding to $j = -1$ is used. Center: a new lobe is added to the contour in the left half-plane, and within this new region the interpolant corresponding to $j = 1$ is used. Right: with this change, the common boundary between the $j = 0$ and $j = 1$ regions may be allowed to pass through the eigenvalue locus, which may be genuinely treated as a movable branch cut.

Two regions are shaded with different intensities: that corresponding to the "original" inequality $\Re\{\phi\} < 0$ (lighter shading) and that corresponding to the "modified" inequality (darker shading). It is evident that something goes wrong with this

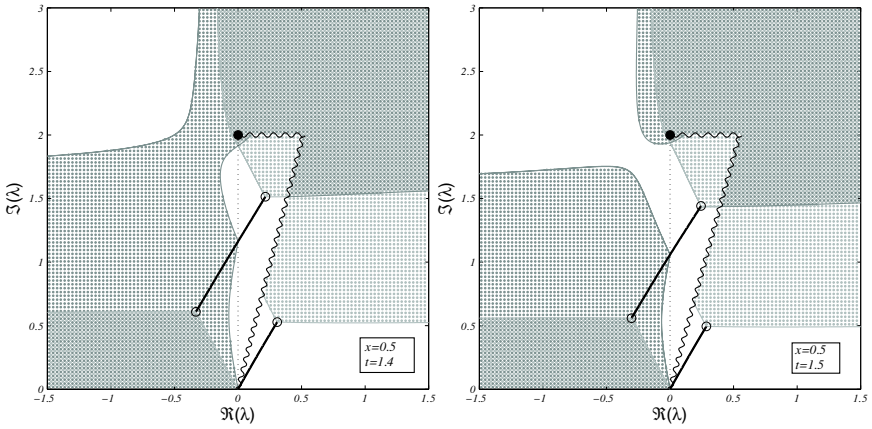

FIGURE 5. The genus two ansatz modified with an additional interpolant.

newer formulation at a time just before that corresponding to the second frame. The region corresponding to the modified inequality has "pinched off," which should be viewed as a topological condition that makes it impossible to locate the new contour lobe so as to satisfy all of the band/gap conditions for genus two. It should be said that this phenomenon occurs precisely at the time when one observes the secondary caustic shown in Figure 1.

In [**11**] it is shown that the situation in the right-hand frame of Figure 5 is remedied without the introduction of further interpolants (at least at this time) by the introduction of a small band on the new contour lobe. The genus just beyond the secondary caustic turns out to be $G = 4$. It is a matter for speculation as to whether interpolants corresponding to higher values of j will be required to capture the dynamics for yet larger values of t.

5. Other Problems

To briefly touch on some ongoing work, there are (at least) two other applications of asymptotic analysis for Riemann-Hilbert problems with lots of discrete spectrum (and no contribution from continuous spectrum):

- Strong asymptotic analysis for the continuum limit of the Ablowitz-Ladik equations. Here one is interested in the coupled system of ordinary differential equations

$$i\frac{dq_n}{dt} + (1 - |q_n|^2)(q_{n+1} + q_{n-1}) = 0\,, \quad 1 \le n \le N-1\,, \quad t > 0\,, \tag{5.1}$$

  subject to initial data of the form $q_n(0) = A(x_n)e^{iNS(x_n)}$, where $x_n := n/N$, $A(x)$ is a smooth N-independent real function with $A(0) = A(1) = 1$ and $A(x) < 1$ for $|x| < 1$, and $S(x)$ is a real N-independent function with $S(0) = 0$. The boundary conditions are $q_0(t) = 1$ and $q_N(t) = e^{i\theta}$ for some angle $\theta = NS(1)$. Weak asymptotics for this problem were obtained by Shipman [**14**] by the Lax-Levermore method. Computation of strong asymptotics by means of a fully discrete Riemann-Hilbert problem related to (discrete) orthogonal polynomials on the unit circle is the subject of current work [**6**].
- Semiclassical asymptotics for the sine-Gordon equation in laboratory coordinates. Here one is interested in the equation

$$\epsilon^2 u_{tt} - \epsilon^2 u_{xx} + \sin(u) = 0\,, \quad x \in \mathbb{R}\,, \quad t > 0\,, \tag{5.2}$$

  subject to initial data of the form $u(x,0) = f(x)$, $\epsilon u_t(x,0) = g(x)$ where f and g are given ϵ-independent real-valued functions, and the boundary conditions $u \to 2\pi n_\pm$ as $x \to \pm\infty$ with $n_\pm \in \mathbb{Z}$ are imposed. Recent calculations [**2**] show that if the initial data are given in the form

$$\sin\left(\frac{1}{2}f(x)\right) = \operatorname{sech}(x)\,, \quad \cos\left(\frac{1}{2}f(x)\right) = \tanh(x)\,, \quad g(x) = -2\mu\operatorname{sech}(x)\,, \tag{5.3}$$

  where μ is an arbitrary real parameter, then the corresponding spectral problem may be solved exactly for all ϵ and for all values of the spectral parameter. From this one finds that as long as ϵ lies in the discrete sequence

$$\epsilon = \epsilon_N := \frac{\sqrt{1+\mu^2}}{2N+1} \tag{5.4}$$

  the reflection coefficient vanishes identically, so the corresponding solution of (5.2) is a pure semiclassical soliton ensemble. An interesting effect here is that the ensemble consists of definite μ-dependent fractions of breathers and kink/antikinks. With the exact fully discrete spectral data in hand, one can apply the kind of methods described in this article to extract

semiclassical asymptotics for the sine-Gordon equation (5.2) from a fully discrete Riemann-Hilbert problem.

In both of these problems, as in the problems discussed previously, there are natural interpolating functions arising from WKB phase calculations. It should also be mentioned that the motivating example given in the introduction for asymptotic analysis of fully discrete Riemann-Hilbert problems, namely the zero-dispersion limit of the Korteweg-de Vries equation with positive initial data, has not yet been revisited from this more recent perspective. Perhaps there is yet something to be discovered in the context of this prototypical example in the asymptotic theory of integrable nonlinear waves.

References

[1] J. Baik, T. Kriecherbauer, K. T.-R. McLaughlin, and P. D. Miller, *Discrete orthogonal polynomials: asymptotics and applications*, Volume 164, Annals of Mathematics Studies Series, Princeton University Press, Princeton, NJ, 2007.

[2] R. Buckingham and P. D. Miller, *Exact solutions of semiclassical non-characteristic Cauchy problems for the sine-Gordon equation*, `arXiv:0705.3159`, 2007.

[3] P. A. Deift, T. Kriecherbauer, K. T.-R. McLaughlin, S. Venakides, and X. Zhou, *Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory*, Comm. Pure Appl. Math., **52** (1999), 1335–1425.

[4] P. A. Deift and K. T.-R. McLaughlin, *A continuum limit of the Toda lattice*, Memoirs Amer. Math. Soc., **624** (1998).

[5] P. A. Deift, S. Venakides, and X. Zhou, *New results in small dispersion KdV by an extension of the steepest-descent method for Riemann-Hilbert problems*, Int. Math. Res. Not., **6** (1997), 286–299.

[6] J. DiFranco, P. D. Miller, and G. Teran-Lozano, in preparation.

[7] A. S. Fokas, A. R. Its, and A. V. Kitaev, *Discrete Painlevé equations and their appearance in quantum gravity*, Comm. Math. Phys., **142** (1991), 313–344.

[8] S. Kamvissis, K. T.-R. McLaughlin, and P. D. Miller, *Semiclassical soliton ensembles for the focusing nonlinear Schrödinger equation*, Volume 154, Annals of Mathematics Studies Series, Princeton University Press, Princeton, NJ, 2003.

[9] M. Klaus and J. K. Shaw, *Purely imaginary eigenvalues of Zakharov-Shabat systems*, Phys. Rev. E, **65** (2002), 36607–36611.

[10] P. D. Lax and C. D. Levermore, *The small dispersion limit of the Korteweg-de Vries equation, I, II, III*, Comm. Pure Appl. Math., **36** (1983), 253–290, 571–593, 809–830.

[11] G. D. Lyng and P. D. Miller, *The N-soliton of the focusing nonlinear Schrödinger equation for N large*, Comm. Pure Appl. Math., **60** (2007), 951–1026.

[12] P. D. Miller, *Asymptotics of semiclassical soliton ensembles: rigorous justification of the WKB approximation*, Int. Math. Res. Not. (2002), 383–454.

[13] P. D. Miller and S. Kamvissis, *On the semiclassical limit of the focusing nonlinear Schrödinger equation*, Phys. Lett. A, **247** (1998), 75–86.

[14] S. P. Shipman, *The spectral transform in the semiclassical limit of a finite discrete NLS chain*, Physica D, **162** (2002), 95–129.

[15] S. Venakides, *The Korteweg-de Vries equation with small dispersion: higher order Lax-Levermore theory*, Comm. Pure Appl. Math., **43** (1990), 335–361.

[16] V. E. Zakharov and A. B. Shabat, *Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media*, Sov. Phys. JETP, **34** (1972), 62–69.

Department of Mathematics, University of Michigan, 530 Church St., Ann Arbor, Michigan 48109

E-mail address: `millerpd@umich.edu`