IMRN International Mathematics Research Notices
2002, No. 8

Asymptotics of Semiclassical Soliton Ensembles: Rigorous Justification of the WKB Approximation

Peter D. Miller

1 Introduction

Many important problems in the theory of integrable systems and approximation theory can be recast as Riemann-Hilbert problems for a matrix-valued unknown. Via the connection with approximation theory, and specifically the theory of orthogonal polynomials, one can also study problems from the theory of random matrix ensembles and combinatorics. Roughly speaking, solving a Riemann-Hilbert problem amounts to reconstructing a sectionally meromorphic matrix from given homogeneous multiplicative "jump conditions" at the boundary contours of the domains of meromorphy, from "principal part data" given at the prescribed singularities, and from a normalization condition. So, many asymptotic questions in integrable systems (e.g., long time behavior and singular perturbation theory) and approximation theory (e.g., behavior of orthogonal polynomials in the limit of large degree) amount to determining asymptotic properties of the solution matrix of a Riemann-Hilbert problem from given asymptotics of the jump conditions and principal part data.

In recent years a collection of techniques has emerged for studying certain asymptotic problems of this sort. These techniques are analogous to familiar asymptotic methods for expanding oscillatory integrals, and we often refer to them as "steepest-descent" methods. The basic method first appeared in the work of Deift and Zhou [5]. The first applications were to Riemann-Hilbert problems without poles, in which the solution matrix is sectionally holomorphic. Later, some problems were studied in which

Received 7 September 2001.
Communicated by Percy Deift.

there were a number of poles—a number held fixed in the limit of interest—in the solution matrix (see, for example, the paper [2] on the long-time behavior of the Toda lattice with rarefaction initial data). The previous methods were extended to these more complicated problems through the device of making a local change of variable near each pole in some small domain containing the pole. The change of variable is chosen so that it has the effect of removing the pole at the cost of introducing an explicit jump on the boundary of the domain around the pole in which the transformation is made. The result is a Riemann-Hilbert problem for a sectionally holomorphic matrix, which can be solved asymptotically by pre-existing "steepest-descent" methods. Recovery of an approximation for the original sectionally meromorphic matrix unknown involves putting back the poles by reversing the explicit change of variables that was designed to get rid of them to begin with.

Yet another category of Riemann-Hilbert problems consists of those problems where the number of poles is not fixed, but becomes large in the limit of interest, with the poles accumulating on some closed set F in the finite complex plane. A problem of this sort has been addressed [8] by making an explicit transformation of the type described above in a single fixed domain G that contains the locus of accumulation F of all the poles. The transformation is chosen to get rid of all the poles at once. In order to specify it, discrete data related to the residues of the poles must be interpolated at the corresponding poles by a function that is analytic and nonvanishing in all of G. Once the poles have been removed in this way, the Riemann-Hilbert problem becomes one for a sectionally holomorphic matrix, with a jump at the boundary of G given in terms of the explicit change of variables. In this way, the poles are "swept out" from F to the boundary of G resulting in an analytic jump. There is a strong analogy in this procedure with the concept of balayage (meaning "sweeping out") from potential theory.

In establishing asymptotic formulae for such Riemann-Hilbert problems, it is essential that one makes judicious use of the freedom to place the boundary of the domain in which one removes the poles from the problem. Placing this boundary contour in the correct position in the complex plane allows one to convert oscillations into exponential decay in such a way that the errors in the asymptotics can be rigorously controlled. If the poles accumulate with some smooth density on $F \subset G$, the characterization of the correct location of the boundary of G can be determined by first passing to a continuum limit of the pole distribution in the resulting jump matrix on the boundary of G, and then applying analytic techniques or variational methods. The continuum limit is justified as long as the boundary of G remains separated from F.

This idea leads to an interesting question. What happens if the boundary of G, as determined from passing to the continuum limit, turns out to intersect F? Far from being a hypothetical possibility, this situation is known to occur in at least three different problems.

(1) The semiclassical limit of the focusing nonlinear Schrödinger hierarchy with decaying initial data. See [8]. This is an inverse-scattering problem for the nonselfadjoint Zakharov-Shabat operator. On an *ad hoc* basis, one replaces the true spectral data for the given initial condition with a formal WKB approximation. There is no jump in the Riemann-Hilbert problem associated with inverse-scattering for the modified spectral data, but there are poles accumulating asymptotically with the WKB density of states on an interval F of the imaginary axis in the complex plane of the eigenvalue. The methods described above turn out to yield rigorous asymptotics for this modified inverse-scattering problem as long as the independent time variable in the equation is not zero. For $t = 0$, the argument of passing to the continuum limit in the pole density leads one to choose the boundary of G to *coincide* in part with the interval F. Strangely, if one sets $t = 0$ in the problem from the beginning, an alternative method due to Lax and Levermore [10, 11, 12] and extended to the nonselfadjoint Zakharov-Shabat operator with real potentials by Ercolani, Jin, Levermore, and MacEvoy [6] can be used to carry out the asymptotic analysis in this special case; this alternative method is not based on matrix Riemann-Hilbert problems, and therefore when taken together with the methods described in [8] does not result in a uniform treatment of the semiclassical limit for all x and t. At the same time, the Lax-Levermore method that applies when $t = 0$ fails in this problem when $t \neq 0$.

(2) The zero-dispersion limit of the Korteweg-de Vries equation with potential well initial data. As pointed out above, the original treatment of this problem by Lax and Levermore [10, 11, 12] was not based on asymptotic analysis for a matrix-valued Riemann-Hilbert problem. But it is possible to pose the inverse-scattering problem with modified (WKB) spectral data as a matrix-valued Riemann-Hilbert problem and ask whether the “steepest descent” techniques for such problems could be used to reproduce and/or strengthen the original asymptotic results of Lax and Levermore. In particular, we might point out that the Lax-Levermore method only gives weak limits of the conserved densities, and that a modification due to Venakides [13] is required to extract any pointwise asymptotics (i.e., to reconstruct the microstruture of the modulated and rapidly oscillatory wavetrains giving rise to the leading-order weak asymptotics). On the other hand, “steepest descent” techniques for matrix-valued Riemann-Hilbert problems typically give pointwise asymptotics automatically. It would therefore be most useful if these techniques could be applied to provide a new and unified approach to this problem.

If one tries to enclose the locus of accumulation of poles (WKB eigenvalues for the Schrödinger operator with a potential well) with a contour and determine the optimal location of this contour for zero-dispersion asymptotics, it turns out that the contour must contain the support of a certain weighted logarithmic equilibrium measure. It is a well-known consequence of the Lax-Levermore theory that the support of this measure is a subset of the interval of accumulation of WKB eigenvalues. Consequently, the enclosing contour "wants" to lie right on top of the poles in this problem, and the approach fails. In a sense this failure of the "steepest descent" method is more serious than in the analogous problem for the focusing nonlinear Schrödinger equation because the contour is in the wrong place for all values of x and t (the independent variables of the problem), whereas in the focusing nonlinear Schrödinger problem the method fails generically only for $t = 0$.

(3) The large degree limit of certain systems of discrete orthogonal polynomials. Fokas, Its, and Kitaev [7] have shown that the problem of reconstructing the orthogonal polynomials associated with a given continuous weight function can be expressed as a matrix-valued Riemann-Hilbert problem. It is not difficult to modify their construction to the case when the weight function is a sum of Dirac masses. The corresponding matrix-valued Riemann-Hilbert problem has no jump, but has poles at the support nodes of the weight. The solution of this Riemann-Hilbert problem gives in this case the associated family of discrete orthogonal polynomials. If one takes the nodes of support of the discrete weight to be distributed asymptotically in some systematic way, then it is natural to ask whether "steepest descent" methods applied to the corresponding Riemann-Hilbert problem with poles could yield accurate asymptotic formulae for the discrete orthogonal polynomials in the limit of large degree. Indeed, similar asymptotics were obtained in the continuous weight case [3] using precisely these methods.

Unfortunately, when the poles are encircled and the optimal contour is sought, it turns out again to be necessary that the contour contains the support of a certain weighted logarithmic equilibrium measure (see [9] for a description of this measure) which is supported on a subset of the interval of accumulation of the nodes of orthogonalization (i.e., the poles). For this reason, the method based on matrix-valued Riemann-Hilbert problems would appear to fail.

In this paper, we present a new technique in the theory of "steepest descent" asymptotic analysis for matrix Riemann-Hilbert problems that solves all three problems mentioned above in a general framework. We illustrate the method in detail for the first case described above: the inverse-scattering problem for the nonselfadjoint Zakharov-Shabat operator with modified (WKB) spectral data, which amounts to a treatment of the semiclassical limit for the focusing nonlinear Schrödinger equation at the initial instant $t = 0$. This work thus fills in a gap in the arguments in [8] connecting the rigorous asymptotic analysis carried out there with the initial-value problem for the focusing nonlinear Schrödinger equation. Application of the same techniques to the zero dispersion limit of the Korteweg-de Vries equation will be the topic of a future paper, and a study of asymptotics for discrete orthogonal polynomials using these methods is already in preparation [1].

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

$$i\hbar\frac{\partial\psi}{\partial t}+\frac{\hbar^2}{2}\frac{\partial^2\psi}{\partial x^2}+|\psi|^2\psi=0, \tag{1.1}$$

subject to the initial condition $\psi(x,0)=\psi_0(x)$. In [8], this problem is considered for cases when the initial data $\psi_0(x)=A(x)$ where $A(x)$ is some positive real function $\mathbb{R}\to(0,A]$. The function $A(x)$ is taken to decay rapidly at infinity and to be even in x with a single genuine maximum at $x=0$. Thus $A(0)=A$, $A'(0)=0$, and $A''(0)<0$. Also, the function $A(x)$ is taken to be real-analytic. With this given initial data, one has a unique solution of (1.1) for each $\hbar>0$. To study the semiclassical limit then means determining asymptotic properties of the family of solutions $\psi(x,t)$ as $\hbar\downarrow 0$.

This problem is associated with the scattering and inverse-scattering theory for the nonselfadjoint Zakharov-Shabat eigenvalue problem [14]:

$$\hbar\frac{du}{dx}=-i\lambda u+A(x)v,\qquad \hbar\frac{dv}{dx}=-A(x)u+i\lambda v, \tag{1.2}$$

for auxiliary functions $u(x;\lambda)$ and $v(x;\lambda)$. The complex number λ is a spectral parameter. Under the conditions on $A(x)$ described above, it is known only that for each $\hbar>0$ the discrete spectrum of this problem is invariant under complex conjugation and reflection through the origin. However, a formal WKB method applied to (1.2) suggests for small $\hbar$ a distribution of eigenvalues that are confined to the imaginary axis. The same method suggests that the reflection coefficient for scattering states obtained for real λ is small beyond all orders.

It is therefore natural to propose a modification of the problem. Rather than studying the inverse-scattering problem for the true spectral data (which is not known), simply replace the true spectral data by its formal WKB approximation in which the eigenvalues are given by a quantization rule of Bohr-Sommerfeld type, and in which the reflection coefficient is neglected entirely. For each $\hbar>0$, this modified spectral data is the *true* spectral data for some other ($\hbar$-dependent) initial condition $\psi_0^\hbar(x)$. Since there is no reflection coefficient in the modified problem, it turns out that for each $\hbar$ the solution of (1.1) corresponding to the modified initial data $\psi_0^\hbar(x)$ is an exact N-soliton solution, with $N\sim\hbar^{-1}$. We call such a family of N-soliton solutions, all obtained from the same function $A(x)$ by a WKB procedure, a *semiclassical soliton ensemble*, or SSE for short. We will be more precise about this idea in Section 2. In [8], the asymptotic behavior of SSEs was studied for $t \neq 0$. Although the results were rigorous, it was not possible to deduce anything about the *true* initial-value problem for (1.1) with $\psi_0(x) \equiv A(x)$ because the asymptotic method failed for $t = 0$. In this paper, we will explain the following new result.

Theorem 1.1. Let $A(x)$ be real-analytic, even, and decaying with a single genuine maximum at $x = 0$. Let $\psi_0^\hbar(x)$ be for each $\hbar > 0$ the exact initial value of the SSE corresponding to $A(x)$ (see Section 2). Then, there exists a sequence of values of $\hbar$, $\hbar = \hbar_N$ for $N = 1, 2, 3, \ldots$, such that

$$\lim_{N\to\infty} \hbar_N = 0 \tag{1.3}$$

and such that for all $x \neq 0$, there exists a constant $K_x > 0$ such that

$$\left|\psi_0^{\hbar_N}(x) - A(x)\right| \leq K_x \hbar_N^{1/7-\nu}, \quad \text{for } N = 1, 2, 3, \ldots \tag{1.4}$$

for all $\nu > 0$. $\square$

As $\psi_0^\hbar(x)$ is obtained by an inverse-scattering procedure applied to WKB spectral data, this theorem establishes in a sense the validity of the WKB approximation for the Zakharov-Shabat eigenvalue problem (1.2). It says that the true spectral data and the formally approximate spectral data generate, via inverse-scattering, potentials in the Zakharov-Shabat problem that are pointwise close. The omission of $x = 0$ is merely technical; a procedure slightly different from that we will explain in this paper is needed to handle this special case. We will indicate as we proceed the modifications that are necessary to extend the result to the whole real line. The pointwise nature of the asymptotics is important; the variational methods used in [6] suggest convergence only in the L^2 sense. Rigorous statements about the nature of the WKB approximation for the Zakharov-Shabat problem are especially significant because the operator in (1.2) is nonselfadjoint and the spectrum is not confined to any axis; furthermore Sturm-Liouville oscillation theory does not apply.

2 Characterization of SSEs

Each N-soliton solution of the focusing nonlinear Schrödinger equation (1.1) can be found as the solution of a meromorphic Riemann-Hilbert problem with no jumps; that is, a problem whose solution matrix is a rational function of $\lambda \in \mathbb{C}$. The N-soliton solution depends on a set of discrete data. Given N complex numbers $\lambda_0, \ldots, \lambda_{N-1}$ in the upper half-plane (these turn out to be discrete eigenvalues of the spectral problem (1.2)), and N nonzero constants $\gamma_0, \ldots, \gamma_{N-1}$ (which turn out to be related to auxiliary discrete spectrum for (1.2)), and an index $J = \pm 1$, one considers the matrix $\mathbf{m}(\lambda)$ solving the following problem.

Riemann-Hilbert Problem 2.1 (meromorphic problem). Find a matrix $\mathbf{m}(\lambda)$ with the following two properties:

(1) Rationality: $\mathbf{m}(\lambda)$ is a rational function of λ, with simple poles confined to the values $\{\lambda_k\}$ and the complex conjugates. At the singularities

$$\begin{aligned} \operatorname*{Res}_{\lambda=\lambda_k} \mathbf{m}(\lambda) &= \lim_{\lambda\to\lambda_k} \mathbf{m}(\lambda)\sigma_1^{(1-J)/2} \begin{bmatrix} 0 & 0 \\ c_k(x,t) & 0 \end{bmatrix} \sigma_1^{(1-J)/2}, \\ \operatorname*{Res}_{\lambda=\lambda_k^*} \mathbf{m}(\lambda) &= \lim_{\lambda\to\lambda_k^*} \mathbf{m}(\lambda)\sigma_1^{(1-J)/2} \begin{bmatrix} 0 & -c_k(x,t)^* \\ 0 & 0 \end{bmatrix} \sigma_1^{(1-J)/2}, \end{aligned} \tag{2.1}$$

for $k = 0, \ldots, N-1$, with

$$c_k(x,t) := \left(\frac{1}{\gamma_k}\right)^J \frac{\displaystyle\prod_{n=0}^{N-1} (\lambda_k - \lambda_n^*)}{\displaystyle\prod_{\substack{n=0\\ n\neq k}}^{N-1} (\lambda_k - \lambda_n)} \exp\left(\frac{2iJ\left(\lambda_k x + \lambda_k^2 t\right)}{\hbar}\right). \tag{2.2}$$

(2) Normalization:

$$\mathbf{m}(\lambda) \longrightarrow \mathbb{I}, \quad \text{as } \lambda \longrightarrow \infty. \tag{2.3}$$

Here, σ_1 denotes one of the Pauli matrices

$$\sigma_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_2 := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \sigma_3 := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{2.4}$$

The function $\psi(x,t)$ defined from $\mathbf{m}(\lambda)$ by the limit

$$\psi(x,t) = 2i \lim_{\lambda\to\infty} \lambda m_{12}(\lambda) \tag{2.5}$$

is the N-soliton solution of the focusing nonlinear Schrödinger equation (1.1) corresponding to the data $\{\lambda_k\}$ and $\{\gamma_k\}$.

The index J will be present throughout this work, so it is worth explaining its role from the start. It turns out that if $J = +1$, then the solution $\mathbf{m}(\lambda)$ of Riemann-Hilbert Problem 2.1 has the property that for all fixed λ distinct from the poles, $\mathbf{m}(\lambda) \to \mathbb{I}$ as $x \to +\infty$. Likewise, if $J = -1$, then $\mathbf{m}(\lambda) \to \mathbb{I}$ as $x \to -\infty$. So as far as scattering theory is concerned, the index J indicates an arbitrary choice of whether we are performing scattering "from the right" or "from the left." Both versions of scattering theory yield the same function $\psi(x,t)$ via the relation (2.5), and are in this sense equivalent. However, the inverse-scattering problem involves the independent variables x and t for (1.1) as parameters, and it may be the case that for different choices of x and t, different choices of the parameter J may be more convenient for asymptotic analysis of the matrix $\mathbf{m}(\lambda)$ solving Riemann-Hilbert Problem 2.1. That this is indeed the case that was observed and documented in [8]. So we need the freedom to choose the index J, and therefore we need to carry it along in our calculations.

A *semiclassical soliton ensemble* (SSE) is a family of particular N-soliton solutions of (1.1) indexed by $N = 1, 2, 3, 4, \ldots$ that are formally associated with given initial data $\psi_0(x) = A(x)$ via an *ad hoc* WKB approximation of the spectrum of (1.2). Note that the initial data $\psi_0(x) = A(x)$ may not exactly correspond to a pure N-soliton solution of (1.1) for any $\hbar$, and similarly that typically none of the N-soliton solutions making up the SSE associated with $\psi_0(x) = A(x)$ will agree with this given initial data at $t = 0$.

We will now describe the discrete data $\{\lambda_k\}$ and $\{\gamma_k\}$ that generate, via the solution of Riemann-Hilbert Problem 2.1 and the subsequent use of formula (2.5), the SSE associated with a function $\psi_0(x) = A(x)$. We suppose that $A(x)$ is an even function of x that has a single maximum at $x = 0$, and is therefore "bell-shaped." We will need $A(x)$ to be rapidly decreasing for large x, and we will suppose that the maximum $A := A(0)$ is genuine in that $A''(0) < 0$. Most importantly in what follows, we will assume that $A(x)$ is a real-analytic function of x.

The starting point is the definition of the WKB eigenvalue density function $\rho^0(\eta)$

$$\rho^0(\eta) := \frac{\eta}{\pi} \int_{x_-(\eta)}^{x_+(\eta)} \frac{dx}{\sqrt{A(x)^2 + \eta^2}}, \tag{2.6}$$

defined for positive imaginary numbers η in the interval $(0, iA)$, where $x_-(\eta)$ and $x_+(\eta)$ are the (unique by our assumptions) negative and positive values of x for which $iA(x) = \eta$. The WKB eigenvalues asymptotically fill out the interval $(0, iA)$, and $\rho^0(\eta)$ is their asymptotic density. This function inherits analyticity properties in η from those of $A(x)$ via the functions $x_\pm(\eta)$. Our assumption that $A(x)$ is real-analytic makes $\rho^0(\eta)$ an analytic function of η in its imaginary interval of definition. Also, our assumption that $A(x)$ should be rapidly decreasing makes $\rho^0(\eta)$ analytic at $\eta = 0$, and our assumption that $A(x)$

has nonvanishing curvature at its maximum makes $\rho^0(\eta)$ analytic at $\eta = iA$. From this function it is convenient to define a measure of the number of WKB eigenvalues between a point $\lambda \in (0, iA)$ on the imaginary axis and iA:

$$\theta^0(\lambda) := -\pi \int_\lambda^{iA} \rho^0(\eta)\, d\eta. \tag{2.7}$$

Now, each N-soliton solution in the SSE for $A(x)$ will be associated with a particular value $\hbar = \hbar_N$, namely

$$\hbar = \hbar_N := -\frac{1}{N} \int_0^{iA} \rho^0(\eta)\, d\eta = \frac{1}{N\pi} \int_{-\infty}^{\infty} A(x)\, dx, \tag{2.8}$$

where $N \in \mathbb{Z}_+$. In this sense we are taking the values of $\hbar$ themselves to be "quantized." Clearly for any given $A(x)$, $\hbar_N = O(1/N)$ which goes to zero as N becomes large. For each $N \in \mathbb{Z}_+$, we then define the WKB eigenvalues formally associated with $A(x)$ according to the Bohr-Sommerfeld rule

$$\theta^0(\lambda_k) = \pi\hbar_N \left(k + \frac{1}{2}\right), \quad \text{for } k = 0, 1, 2, \ldots, N-1 \tag{2.9}$$

and the auxiliary scattering data by

$$\gamma_k := -i(-1)^K \exp\left(-\frac{i(2K+1)\theta^0(\lambda_k)}{\hbar_N}\right). \tag{2.10}$$

Here, K is an arbitrary integer. Clearly the Bohr-Sommerfeld rule (2.9) implies that choosing different integer values of K in (2.10) will yield the same set of numbers $\{\gamma_k\}$. However, we take the point of view that the right-hand side of (2.10) furnishes an analytic function that interpolates the $\{\gamma_k\}$ at the $\{\lambda_k\}$; for different $K \in \mathbb{Z}$ these are different interpolating functions which is a freedom that we will exploit to our advantage. In fact, we will only need to consider $K = 0$ or $K = -1$.

For $A(x)$ given as above, the SSE is a sequence of exact solutions of (1.1) such that the Nth element $\psi^{\hbar_N}(x,t)$ of the SSE (i) solves (1.1) with $\hbar = \hbar_N$ as given by (2.8) and (ii) is defined as the N-soliton solution corresponding to the eigenvalues $\{\lambda_k\}$ given by (2.9) and the auxiliary spectrum $\{\gamma_k\}$ given by (2.10) via the solution of Riemann-Hilbert Problem 2.1 with $\hbar = \hbar_N$. For each N, we restrict the SSE to $t = 0$ to obtain functions

$$\psi_0^{\hbar_N}(x) := \psi^{\hbar_N}(x, 0). \tag{2.11}$$

It is this sequence of functions that is the subject of Theorem 1.1. In the following sections we will set up a new framework for the asymptotic analysis of SSEs in the limit $N \to \infty$, a problem closely related to the computation of asymptotics of solutions of (1.1) for fixed initial data $\psi_0(x) = A(x)$ in the semiclassical limit.

3 Removal of the poles

The asymptotic method we will now develop for studying Riemann-Hilbert Problem 2.1 for SSEs is especially well adapted to studying the case of $t = 0$, where the method described in detail in [8] fails. To illustrate the new method, we therefore set $t = 0$ in the rest of this paper. Also, we anticipate the utility of tying the value of the parameter $J = \pm 1$ to the remaining independent variable x by setting

$$J := \operatorname{sign}(x). \tag{3.1}$$

In all subsequent formulae in which the index J appears it should be assumed to be assigned a definite value according to (3.1).

We now want to convert Riemann-Hilbert Problem 2.1 into a new Riemann-Hilbert problem for a sectionally holomorphic matrix so that the "steepest-descent" methods can be applied. As mentioned in the introduction, in [8] this transformation can be accomplished by encircling the locus of accumulation of the poles, here the imaginary interval $(0, iA)$, with a loop contour in the upper half-plane and making a specific change of variables based on the interpolation formula (2.10) for some value of $K \in \mathbb{Z}$ in the interior of the region enclosed by the loop and also in the complex-conjugate region. One then tries to choose the position of the loop contour in the complex plane that is best adapted to asymptotic analysis of the resulting holomorphic Riemann-Hilbert problem. The trouble with this approach is that it turns out that for $t = 0$ the "correct" placement of the contour requires that part of it should lie on a subset of the imaginary interval $(0, iA)$, that is, right on top of the accumulating poles! For such a choice of the loop contour, the boundary values taken by the transformed matrix on the outside of the loop would be singular and the "steepest descent" theory would not apply.

So taking the point of view that making any particular choice of $K \in \mathbb{Z}$ in (2.10) leads to problems, we propose to simultaneously make use of *two distinct values of* K in passing to a Riemann-Hilbert problem for a sectionally holomorphic matrix. Consider the contours illustrated in Figure 3.1, arranged such that $\{\lambda_0, \dots, \lambda_{N-1}\} \subset D_L \cup D_R$. For $\lambda \in D_L$, set

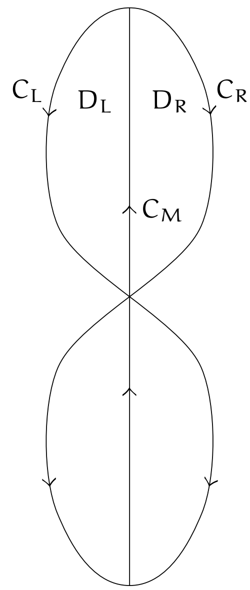

Figure 3.1 The geometry of contours introduced in the complex λ-plane. The uppermost common point of the contours C_L, C_M, and C_R is $\lambda = iA$. The six-fold self-intersection point is the origin $\lambda = 0$.

$$\mathbf{M}(\lambda) := \mathbf{m}(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} 1 & 0 \\ i\left(\prod_{k=0}^{N-1}\frac{\lambda-\lambda_k^*}{\lambda-\lambda_k}\right)\exp\left(\frac{2i\lambda|x| - i\theta^0(\lambda)}{\hbar_N}\right) & 1 \end{bmatrix}\sigma_1^{(1-J)/2}. \quad (3.2)$$

For $\lambda \in D_R$, set

$$\mathbf{M}(\lambda) := \mathbf{m}(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} 1 & 0 \\ -i\left(\prod_{k=0}^{N-1}\frac{\lambda-\lambda_k^*}{\lambda-\lambda_k}\right)\exp\left(\frac{2i\lambda|x| + i\theta^0(\lambda)}{\hbar_N}\right) & 1 \end{bmatrix}\sigma_1^{(1-J)/2}. \quad (3.3)$$

To preserve the conjugation symmetry[1] $\mathbf{m}(\lambda^*) = \sigma_2\mathbf{m}(\lambda)^*\sigma_2$ of the matrix $\mathbf{m}(\lambda)$ that is the unique solution of Riemann-Hilbert Problem 2.1, for $\lambda \in D_L^* \cup D_R^*$ we set $\mathbf{M}(\lambda) := \sigma_2\mathbf{M}(\lambda^*)^*\sigma_2$. Finally, for all other complex λ set $\mathbf{M}(\lambda) = \mathbf{m}(\lambda)$. So rather than enclosing the poles in a loop and making a single change of variables inside, we are splitting the region inside the loop in half, and we are using different interpolants (2.10) of the $\{\gamma_k\}$ at the $\{\lambda_k\}$ in each half of the loop. Some of the properties of the transformed matrix $\mathbf{M}(\lambda)$ are the following.

[1] Note that we are denoting by $\mathbf{A}^*$ the componentwise complex conjugate of the matrix $\mathbf{A}$, and we reserve the notation $\mathbf{A}^\dagger$ for the conjugate-transpose.

Proposition 3.1. The matrix $\mathbf{M}(\lambda)$ is analytic in $\mathbb{C} \setminus \Sigma$ where Σ is the union of the contours C_L, C_R, and C_M, and their complex conjugates. Moreover, $\mathbf{M}(\lambda)$ takes continuous boundary values on Σ. □

Proof. The function $\theta^0(\lambda)$ is analytic in D_L and D_R if C_L and C_R are chosen close enough to the imaginary axis since $\rho^0(\eta)$ is analytic there. By using the residue relation (2.1) and the interpolation formula (2.10) alternatively for $K = 0$ and $K = -1$, one checks directly that the poles of $\mathbf{m}(\lambda)$ are canceled by the explicit Blaschke factors in (3.2) and (3.3). ■

Proposition 3.2. Let $\mathbf{M}_\pm(\lambda)$ denote the boundary values taken on the oriented contour Σ, where the subscript "+" (resp., "−") indicates the boundary value taken from the left (resp., from the right). Then for $\lambda \in C_L$,

$$\mathbf{M}_-(\lambda)^{-1}\mathbf{M}_+(\lambda) = \sigma_1^{(1-J)/2}\begin{bmatrix} 1 & 0 \\ i\left(\prod_{k=0}^{N-1}\frac{\lambda-\lambda_k^*}{\lambda-\lambda_k}\right)\exp\left(\frac{2i\lambda|x| - i\theta^0(\lambda)}{\hbar_N}\right) & 1 \end{bmatrix}\sigma_1^{(1-J)/2}. \tag{3.4}$$

For $\lambda \in C_R$,

$$\mathbf{M}_-(\lambda)^{-1}\mathbf{M}_+(\lambda) = \sigma_1^{(1-J)/2}\begin{bmatrix} 1 & 0 \\ i\left(\prod_{k=0}^{N-1}\frac{\lambda-\lambda_k^*}{\lambda-\lambda_k}\right)\exp\left(\frac{2i\lambda|x| + i\theta^0(\lambda)}{\hbar_N}\right) & 1 \end{bmatrix}\sigma_1^{(1-J)/2}. \tag{3.5}$$

For $\lambda \in C_M$,

$$\mathbf{M}_-(\lambda)^{-1}\mathbf{M}_+(\lambda) = \sigma_1^{(1-J)/2}\begin{bmatrix} 1 & 0 \\ i\left(\prod_{k=0}^{N-1}\frac{\lambda-\lambda_k^*}{\lambda-\lambda_k}\right)\exp\left(\frac{2i\lambda|x|}{\hbar_N}\right)\cdot 2\cos\left(\frac{\theta^0(\lambda)}{\hbar_N}\right) & 1 \end{bmatrix}\sigma_1^{(1-J)/2}. \tag{3.6}$$

On the contours in the lower half-plane the jump relations are determined by the symmetry $\mathbf{M}(\lambda) = \sigma_2\mathbf{M}(\lambda^*)^*\sigma_2$. All jump matrices are analytic functions in the vicinity of their respective contours. □

Proof. This is also a direct consequence of (3.2) and (3.3). The analyticity is clear on C_L and C_R since $\theta^0(\lambda)$ is analytic there, while on C_M one observes that as a consequence of the Bohr-Sommerfeld quantization condition (2.9), the cosine factor precisely cancels the poles on C_M contributed by the product of Blaschke factors. ■

Although we have specified the contour C_M to coincide with a segment of the imaginary axis, the reader will see that the same statements concerning the analyticity of $\mathbf{M}(\lambda)$ and the continuity of the boundary values on Σ also hold when C_M is taken to be absolutely any smooth contour in the upper half-plane connecting $\lambda = 0$ to $\lambda = iA$. Given a choice of C_M, the contours C_L and C_R must be such that the topology of Figure 3.1 is preserved. We also have specified that C_L and C_R should lie sufficiently close to C_M (a distance independent of $\hbar_N$) so that $\theta^0(\lambda)$ is analytic in D_L and D_R. Later we will also exploit the proximity of these two contours to C_M to deduce decay properties of certain analytic functions on these contours from their oscillation properties on C_M by the Cauchy-Riemann equations.

Taken together, Propositions 3.1 and 3.2 indicate that the matrix $\mathbf{M}(\lambda)$ satisfies a Riemann-Hilbert problem without poles, but instead having explicit homogeneous jump relations on Σ given by the matrix functions on the right-hand sides of (3.4), (3.5), and (3.6). The normalization of $\mathbf{M}(\lambda)$ at infinity is the same as that of $\mathbf{m}(\lambda)$ since no transformation has been made outside a compact set, so if $\mathbf{M}(\lambda)$ can be recovered from its jump relations and normalization condition, then the SSE itself can be obtained for $t = 0$ from (2.5) with $\mathbf{m}(\lambda)$ replaced by $\mathbf{M}(\lambda)$.

4 The complex phase function

We now introduce a further change of dependent variable involving a scalar function that is meant to capture the dominant asymptotics for the problem. Let $g(\lambda)$ be a complex-valued function that is independent of $\hbar$, analytic for $\lambda \in \mathbb{C}\setminus(C_M \cup C_M^*)$ taking continuous boundary values, satisfies $g(\lambda) + g(\lambda^*)^* = 0$, and $g(\infty) = 0$. Setting

$$\mathbf{N}(\lambda) := \mathbf{M}(\lambda) \exp\left(-\frac{g(\lambda)\sigma_3}{\hbar}\right), \tag{4.1}$$

we find that for $\lambda \in C_L$,

$$\mathbf{N}_-(\lambda)^{-1}\mathbf{N}_+(\lambda) = \sigma_1^{(1-J)/2} \begin{bmatrix} 1 & 0 \\ a_L(\lambda) & 1 \end{bmatrix} \sigma_1^{(1-J)/2}, \tag{4.2}$$

where

$$a_L(\lambda) := i\left(\prod_{k=0}^{N-1} \frac{\lambda-\lambda_k^*}{\lambda-\lambda_k}\right) \exp\left(\frac{2i\lambda|x| - i\theta^0(\lambda) - 2Jg(\lambda)}{\hbar_N}\right). \tag{4.3}$$

Similarly, for $\lambda \in C_R$, we find

$$\mathbf{N}_-(\lambda)^{-1}\mathbf{N}_+(\lambda) = \sigma_1^{(1-J)/2} \begin{bmatrix} 1 & 0 \\ a_R(\lambda) & 1 \end{bmatrix} \sigma_1^{(1-J)/2}, \tag{4.4}$$

where

$$a_R(\lambda) := i\left(\prod_{k=0}^{N-1} \frac{\lambda - \lambda_k^*}{\lambda - \lambda_k}\right) \exp\left(\frac{2i\lambda|x| + i\theta^0(\lambda) - 2Jg(\lambda)}{\hbar_N}\right). \tag{4.5}$$

Finally, for $\lambda \in C_M$,

$$\mathbf{N}_-(\lambda)^{-1}\mathbf{N}_+(\lambda) = \sigma_1^{(1-J)/2} \begin{bmatrix} \exp\left(\dfrac{i\theta(\lambda)}{\hbar_N}\right) & 0 \\ a_M(\lambda) & \exp\left(-\dfrac{i\theta(\lambda)}{\hbar_N}\right) \end{bmatrix} \sigma_1^{(1-J)/2}, \tag{4.6}$$

where

$$a_M(\lambda) := i\left(\prod_{k=0}^{N-1} \frac{\lambda - \lambda_k^*}{\lambda - \lambda_k}\right) \exp\left(\frac{2i\lambda|x| - Jg_+(\lambda) - Jg_-(\lambda)}{\hbar_N}\right) \cdot 2\cos\left(\frac{\theta^0(\lambda)}{\hbar_N}\right), \tag{4.7}$$

$$\theta(\lambda) := iJ\big(g_+(\lambda) - g_-(\lambda)\big). \tag{4.8}$$

This means that given a function $g(\lambda)$ with the properties described above, one finds that the matrix $\mathbf{N}(\lambda)$ satisfies another holomorphic Riemann-Hilbert problem with jump conditions determined from (4.2), (4.4), and (4.6). Because $g(\infty) = 0$ and $g(\lambda)$ is analytic near infinity, it follows that the correct normalization condition for $\mathbf{N}(\lambda)$ is again that $\mathbf{N}(\lambda) \to \mathbb{I}$ as $\lambda \to \infty$. These same conditions on $g(\lambda)$ show that if $\mathbf{N}(\lambda)$ can be found from its jump conditions and normalization condition, then the SSE can be found via (2.5) with $\mathbf{m}(\lambda)$ replaced by $\mathbf{N}(\lambda)$.

The function $g(\lambda)$ is called a *complex phase function*. The advantage of introducing it into the problem is that by choosing it correctly, the jump matrices (4.2), (4.4), and (4.6) can be cast into a form that is especially convenient for analysis in the semiclassical limit of $\hbar_N \to 0$. The idea of introducing the complex phase function to assist in finding the leading-order asymptotics and controlling the error in this way first appeared in [4] as a modification of the "steepest-descent" method proposed in [5].

5 Pointwise semiclassical asymptotics of the jump matrices

For our purposes, we would like to have each element of the jump matrix for $\mathbf{N}(\lambda)$ of the form $\exp(f(\lambda)/\hbar_N)$ for some appropriate function $f(\lambda)$ that is independent of $\hbar_N$.

While this is not true strictly speaking, it becomes a good approximation in the limit $\hbar_N \to 0$ with λ held fixed (the approximation is not uniform near $\lambda = 0$ or $\lambda = iA$). In this section, we describe the pointwise asymptotics of the jump matrix for $\mathbf{N}(\lambda)$ with the aim of writing all nonzero matrix elements asymptotically in the form $\exp(f(\lambda)/\hbar_N)$ with a small relative error whose magnitude we can estimate.

Roughly speaking, the intuition is that the product over k of Blaschke factors should be replaced with an exponential of a sum over k of logarithms. The latter sum goes over to an integral that scales like $\hbar_N^{-1}$ in the semiclassical limit. On the contour C_M, the cosine that cancels the poles must also be incorporated into the asymptotics.

The branch of the logarithm that is convenient to use here is most conveniently viewed as a function of two complex variables

$$L^0_\eta(\lambda) := \log\big(-i(\lambda-\eta)\big) + \frac{i\pi}{2}. \tag{5.1}$$

As a function of λ for fixed η, it is a logarithm that is cut downwards in the negative imaginary direction from the logarithmic pole at $\lambda = \eta$. Equivalently, $L^0_\eta(\lambda)$ can be viewed as the branch of the multivalued function $\log(\lambda-\eta)$ for which $\arg(\lambda-\eta) \in (-\pi/2, 3\pi/2)$. Suppose $\eta \in C_M$. The boundary value of $L^0_\eta(\lambda)$ taken on C_M as λ approaches from the left (resp., right) side is denoted by $L^0_{\eta+}(\lambda)$ (resp., $L^0_{\eta-}(\lambda)$). The average of these two boundary values is denoted by $\overline{L}^0_\eta(\lambda)$.

All the results we need will come from studying the asymptotic behavior of two quotients:

$$S(\lambda) := \left(\prod_{k=0}^{N-1}\frac{\lambda-\lambda_k^*}{\lambda-\lambda_k}\right)\exp\left(-\frac{1}{\hbar_N}\left(\int_0^{iA} L^0_\eta(\lambda)\rho^0(\eta)\,d\eta + \int_{-iA}^{0} L^0_\eta(\lambda)\rho^0(\eta^*)^*\,d\eta\right)\right), \tag{5.2}$$

$$T(\lambda) := \left(\prod_{k=0}^{N-1}\frac{\lambda-\lambda_k^*}{\lambda-\lambda_k}\right)\exp\left(-\frac{1}{\hbar_N}\left(\int_0^{iA} \overline{L}^0_\eta(\lambda)\rho^0(\eta)\,d\eta + \int_{-iA}^{0} \overline{L}^0_\eta(\lambda)\rho^0(\eta^*)^*\,d\eta\right)\right) \times 2\cos\left(\frac{\theta^0(\lambda)}{\hbar_N}\right). \tag{5.3}$$

The function $S(\lambda)$ is analytic and nonvanishing for $\lambda \in \mathbb{C}_+ \setminus C_M$. We denote by $\Omega \subset \mathbb{C}_+$ the domain of analyticity of $\rho^0(\lambda)$ restricted to the upper half-plane, so that by our assumptions on $A(x)$, $C_M \subset \Omega$. Then, due to the zeros of the cosine on the imaginary axis, which match the poles of the product below $\lambda = iA$ and are not cancelled above $\lambda = iA$, $T(\lambda)$ is analytic and nonvanishing for $\lambda \in \Omega \setminus V$, where V is the vertical ray from $\lambda = iA$ to infinity along the positive imaginary axis. The domain of analyticity for $T(\lambda)$

is a subset of Ω rather than of the whole upper half-plane due to the presence of the averages of the logarithms in the integrand of (5.3). Whereas these are boundary values defined a priori only on C_M, the integrals extend from C_M to analytic functions in the domain $\Omega_+ \setminus V$ via the introduction of the function $\theta^0(\lambda)$ (cf. equation (5.18)).

Lemma 5.1. For all λ in the upper half-plane with $\hbar_N \le |\Re(\lambda)| \le B$, where B is positive and sufficiently small, but fixed as $\hbar_N \to 0$,

$$S(\lambda) = 1 + O\left(\frac{\hbar_N}{|\Re(\lambda)|}\right). \tag{5.4}$$

□

Proof. We define the function $m(\eta)$ by

$$m(\eta) := -\int_0^\eta \rho^0(\xi)\, d\xi. \tag{5.5}$$

This analytic function takes the imaginary interval $[0, iA]$ to the real interval $[0, M]$ where

$$M = m(iA) = \frac{1}{\pi}\int_{-\infty}^{\infty} A(x)\, dx. \tag{5.6}$$

Since $\rho^0(\xi)$ does not vanish on C_M, we have the inverse function $\eta = e(m)$ defined for m near the real interval $[0, M]$. Using these tools, we get the following representation for $S(\lambda)$:

$$\begin{aligned} S(\lambda) &= \exp\left(-\tilde{I}(\lambda)\right), \quad \text{where } \tilde{I}(\lambda) = \sum_{k=0}^{N-1} \tilde{I}_k(\lambda), \\ \tilde{I}_k(\lambda) &:= \frac{1}{\hbar_N}\int_{m_k-\hbar_N/2}^{m_k+\hbar_N/2} \left[L^0_{-e(m)}(\lambda) - L^0_{e(m)}(\lambda)\right] dm \\ &\quad - \left[L^0_{-e(m_k)}(\lambda) - L^0_{e(m_k)}(\lambda)\right], \end{aligned} \tag{5.7}$$

with $m_k := M - \hbar_N(k + 1/2)$. Expanding the logarithms, we find that

$$\tilde{I}_k(\lambda) = \frac{1}{\hbar_N}\int_{m_k-\hbar_N/2}^{m_k+\hbar_N/2} dm \int_{m_k}^{m} d\zeta \int_{m_k}^{\zeta} d\xi \left[\frac{2e''(\xi)\lambda^3 - 2e''(\xi)e(\xi)^2\lambda + 4e'(\xi)^2 e(\xi)\lambda}{(\lambda^2 - e(\xi)^2)^2}\right]. \tag{5.8}$$

This quantity is clearly $O(\hbar_N^2)$ for λ fixed away from C_M. Now, when $|\Re(\lambda)| = o(1)$ as $\hbar_N \downarrow 0$, we can estimate the denominator in the integrand to obtain two different bounds

$$\frac{2e''(\xi)\lambda^3 - 2e''(\xi)e(\xi)^2\lambda + 4e'(\xi)^2 e(\xi)\lambda}{(\lambda^2 - e(\xi)^2)^2} = O\left(\frac{1}{\Re(\lambda)^2}\right), \tag{5.9}$$

$$\frac{2e''(\xi)\lambda^3 - 2e''(\xi)e(\xi)^2\lambda + 4e'(\xi)^2 e(\xi)\lambda}{(\lambda^2 - e(\xi)^2)^2} = O\left(\frac{1}{|i\Im(\lambda) - e(\xi)|^2}\right). \tag{5.10}$$

The idea is to use the estimate (5.9) when $e(m_k)$ is close to $i\Im(\lambda)$ and to use the estimate (5.10) for the remaining terms. Suppose first $\Im(\lambda)$ is bounded between 0 and A, that is, there are small fixed positive numbers δ_1 and δ_2 so that $\delta_1 \le \Im(\lambda) \le A - \delta_2$, and let $\epsilon = \epsilon(\hbar_N)$ be a small positive scale tied to $\hbar$ and satisfying $\hbar_N \ll \epsilon \ll 1$, and let L_1 be chosen from $0, \ldots, N-1$ so that $e(m_{L_1})$ is as close as possible to $i(\Im(\lambda) + \epsilon)$, and likewise let L_2 be chosen from $0, \ldots, N-1$ so that $e(m_{L_2})$ is as close as possible to $i(\Im(\lambda) - \epsilon)$. Using (5.9) we then find that

$$\sum_{k=L_1}^{L_2-1} \tilde{I}_k(\lambda) = O\left(\frac{\hbar_N \epsilon}{\Re(\lambda)^2}\right) \tag{5.11}$$

because the sum contains $O(\epsilon/\hbar_N)$ terms and the volume of the region of integration for each term is $O(\hbar_N^3)$, and we must take into account the overall factor of $1/\hbar_N$. Now in each of the remaining terms $\tilde{I}_k(\lambda)$, we have

$$\frac{1}{|i\Im(\lambda) - e(\xi)|^2} = O\left(\frac{1}{\left(m_k - m(i\Im(\lambda))\right)^2}\right) \tag{5.12}$$

so using (5.10) and summing over k we get both

$$\sum_{k=0}^{L_1-1} \tilde{I}_k(\lambda) = O\left(\frac{\hbar_N}{\epsilon}\right), \qquad \sum_{k=L_2}^{N-1} \tilde{I}_k(\lambda) = O\left(\frac{\hbar_N}{\epsilon}\right). \tag{5.13}$$

The total estimate of $\tilde{I}(\lambda)$ is then optimized by a dominant balance among the three partial sums. This balance requires taking $\epsilon \sim |\Re(\lambda)|$, upon which we deduce that under our assumptions on λ, we indeed have

$$\tilde{I}(\lambda) = O\left(\frac{\hbar_N}{|\Re(\lambda)|}\right) \qquad \text{and consequently} \qquad S(\lambda) - 1 = O\left(\frac{\hbar_N}{|\Re(\lambda)|}\right), \tag{5.14}$$

when $\Im(\lambda)$ is bounded between 0 and A. When $\Im(\lambda) \approx 0$ or $\Im(\lambda) \approx A$, the estimate (5.9) should be used only for those terms that correspond to m near zero or m near M, respectively. In both of these exceptional cases, the same estimate is found. When $\Im(\lambda)$ is bounded below by A, there is no need to use the estimate (5.9) at all, and the relative error is of order $\hbar_N$ uniformly in $\Re(\lambda)$. This completes the proof. ■

We now use this information about $S(\lambda)$ to effectively replace the sums of logarithms by integrals, at least on some portions of the contour Σ.

Proposition 5.2. Suppose that the contour C_L is independent of $\hbar_N$ and that for some sufficiently small positive number B, C_L lies in the strip $-B \le \Re(\lambda) \le 0$ and meets the imaginary axis only at its endpoints and does so transversely. Then

$$a_L(\lambda) = i\exp\left(\frac{1}{\hbar_N}\left(2i\lambda|x| + \int_0^{iA} L_\eta^0(\lambda)\rho^0(\eta)\,d\eta + \int_{-iA}^0 L_\eta^0(\lambda)\rho^0(\eta^*)^*\,d\eta - 2Jg(\lambda)\right)\right)$$
$$\times\exp\left(-\frac{i\theta^0(\lambda)}{\hbar_N}\right)\left(1 + O\left(\frac{\hbar_N}{|\lambda|}\right) + O\left(\frac{\hbar_N}{|\lambda - iA|}\right)\right), \tag{5.15}$$

as $\hbar_N$ goes to zero through positive values, for all $\lambda \in C_L$ with $|\lambda| > \hbar_N$ and $|\lambda - iA| > \hbar_N$. □

Proposition 5.3. Suppose that the contour C_R is independent of $\hbar_N$ and that for some sufficiently small positive number B, C_R lies in the strip $0 \le \Re(\lambda) \le B$ and meets the imaginary axis only at its endpoints and does so transversely. Then

$$a_R(\lambda) = i\exp\left(\frac{1}{\hbar_N}\left(2i\lambda|x| + \int_0^{iA} L_\eta^0(\lambda)\rho^0(\eta)\,d\eta + \int_{-iA}^0 L_\eta^0(\lambda)\rho^0(\eta^*)^*\,d\eta - 2Jg(\lambda)\right)\right)$$
$$\times\exp\left(\frac{i\theta^0(\lambda)}{\hbar_N}\right)\left(1 + O\left(\frac{\hbar_N}{|\lambda|}\right) + O\left(\frac{\hbar_N}{|\lambda - iA|}\right)\right), \tag{5.16}$$

as $\hbar_N$ goes to zero through positive values, for all $\lambda \in C_R$ with $|\lambda| > \hbar_N$ and $|\lambda - iA| > \hbar_N$. □

Proof of Propositions 5.2 and 5.3. These propositions follow directly from Lemma 5.1 upon using the transversality of the intersections with the imaginary axis to replace $O(1/|\Re(\lambda)|)$ by $O(1/|\lambda|) + O(1/|\lambda - iA|)$. ■

We notice that the first factor on the second line in (5.15) and the first factor on the second line in (5.16) are both exponentially small as $\hbar_N$ goes to zero through positive values, as a consequence of the fact that $\rho^0(\eta)d\eta$ is an analytic negative real measure on C_M. This follows from the Cauchy-Riemann equations and the geometry of Figure 3.1. It will be a very useful fact for us shortly.

Now we turn our attention to the function $T(\lambda)$. The result analogous to Lemma 5.1 is the following.

Lemma 5.4. For all λ in the upper half-plane with $\hbar_N \le |\Re(\lambda)| \le B$, where B is positive and sufficiently small, but fixed as $\hbar_N \to 0$,

$$T(\lambda) = 1 + O\left(\frac{\hbar_N}{|\Re(\lambda)|}\right). \tag{5.17}$$

□

Proof. We begin with the jump condition

$$\begin{aligned} &\int_0^{iA} L^0_{\eta+}(\lambda)\rho^0(\eta)\, d\eta + \int_{-iA}^0 L^0_{\eta+}(\lambda)\rho^0(\eta^*)^*\, d\eta \\ &\quad = \int_0^{iA} L^0_{\eta-}(\lambda)\rho^0(\eta)\, d\eta + \int_{-iA}^0 L^0_{\eta-}(\lambda)\rho^0(\eta^*)^*\, d\eta - 2i\theta^0(\lambda) \end{aligned} \tag{5.18}$$

relating the boundary values of the logarithm $L^0_\eta(\lambda)$ on the imaginary axis. Using this jump relation and the definition of $\overline{L}^0_\eta(\lambda)$ as the average of the boundary values of $L^0_{\eta+}(\lambda)$ and $L^0_{\eta-}(\lambda)$, we see that for $\Re(\lambda) < 0$, we have

$$T(\lambda) = S(\lambda)\left(1 + \exp\left(-\frac{2i\theta^0(\lambda)}{\hbar_N}\right)\right), \tag{5.19}$$

while for $\Re(\lambda) > 0$, we have

$$T(\lambda) = S(\lambda)\left(1 + \exp\left(\frac{2i\theta^0(\lambda)}{\hbar_N}\right)\right). \tag{5.20}$$

Now, using the fact that $\rho^0(\eta)$ is an analytic function satisfying $\rho^0(\eta) \in i\mathbb{R}_+$ for $\eta \in C_M$, we see by the Cauchy-Riemann equations that in both cases, the exponential relative error term is of the order $e^{-K|\Re(\lambda)|/\hbar_N}$ for some $K > 0$. Since this is negligible compared with the relative error associated with the asymptotic approximation of $S(\lambda)$ given in Lemma 5.1, the proof is complete. ■

Unfortunately, we need asymptotic information about $T(\lambda)$ right on the imaginary axis, which contains the contour C_M, so we need to improve upon Lemma 5.4. We begin to extract this additional information by noting that under some circumstances, it is easy to show that $T(\lambda)$ remains bounded in the vicinity of the imaginary axis.

Lemma 5.5. If either (i) λ is real or (ii) $|\lambda| = A$ and $\Im(\lambda) > 0$, and if for some $B > 0$ sufficiently small $|\Re(\lambda)| < B$, then $T(\lambda)$ is uniformly bounded as $\hbar_N \to 0$. □

Proof. It suffices to show that $S(\lambda)$ is bounded under the same assumptions, because from (5.19) and (5.20) and the Cauchy-Riemann equations, we see easily that $|T(\lambda)| \leq 2|S(\lambda)|$.

Using the function $m(\cdot)$ and its inverse $e(\cdot)$, we have the following:

$$\hbar_N \log|S(\lambda)| = \sum_{k=0}^{N-1} H(m_k)\hbar_N - \int_0^M H(m)\,dm, \tag{5.21}$$

where

$$H(m) := \log\left|\frac{\lambda + e(m)}{\lambda - e(m)}\right|. \tag{5.22}$$

When $\lambda \in \mathbb{R}$, we see immediately that $H(m) \equiv 0$, and therefore $|S(\lambda)| \equiv 1$ and hence $|T(\lambda)| \leq 2$.

Now consider $\lambda = iAe^{i\theta}$ with θ sufficiently small independent of $\hbar_N$. The idea is that of the terms on the right-hand side of (5.21), the discrete sum is a Riemann sum approximation to the integral. The Riemann sum is constructed using the midpoints of N equal subintervals as sample points. If $H''(m)$ is bounded uniformly, then this sort of Riemann sum provides an approximation to the integral that is of order N^{-2} or equivalently $\hbar_N^2$. In this case, we deduce that $S(\lambda) = 1 + O(\hbar_N)$ and in particular this is bounded as $\hbar_N$ tends to zero. But as λ approaches the imaginary axis, the accuracy of the approximation is lost.

For $\lambda = iAe^{i\theta}$, the function $H(m)$ satisfies $H(0) = H'(M) = 0$ and takes its maximum when $m = M$, with a maximum value

$$H(M) = \log\left|\cot\left(\frac{\theta}{2}\right)\right|. \tag{5.23}$$

Therefore, as θ tends to zero, $H(m)$ becomes unbounded, growing logarithmically in θ. As a consequence of this blowup the approximation of the integral by the Riemann sum based on midpoints for $|\lambda| = A$ fails to be second-order accurate uniformly in θ. However, because the maximum of $H(m)$ always occurs at the right endpoint, it is easy to see that when the error becomes larger than $O(\hbar_N^2)$ in magnitude its sign is such that the Riemann sum is always an *underestimate* of the value of the integral, and consequently the right-hand side of (5.21) is negative. This is concretely illustrated in Figure 5.1 where we have taken the example of the Gaussian function $A(x) = \sqrt{\pi}e^{-x^2}$ in order to supply the function $\rho^0(\eta)$ and therefore the function $e(m)$ needed to build $H(m)$. In this case, $A = \sqrt{\pi}$ and $M = 1$. The error of the Riemann sum is worst when $\theta = 0$. In this case it is easy to see that the discrepancy contributed by only the subinterval adjacent to the

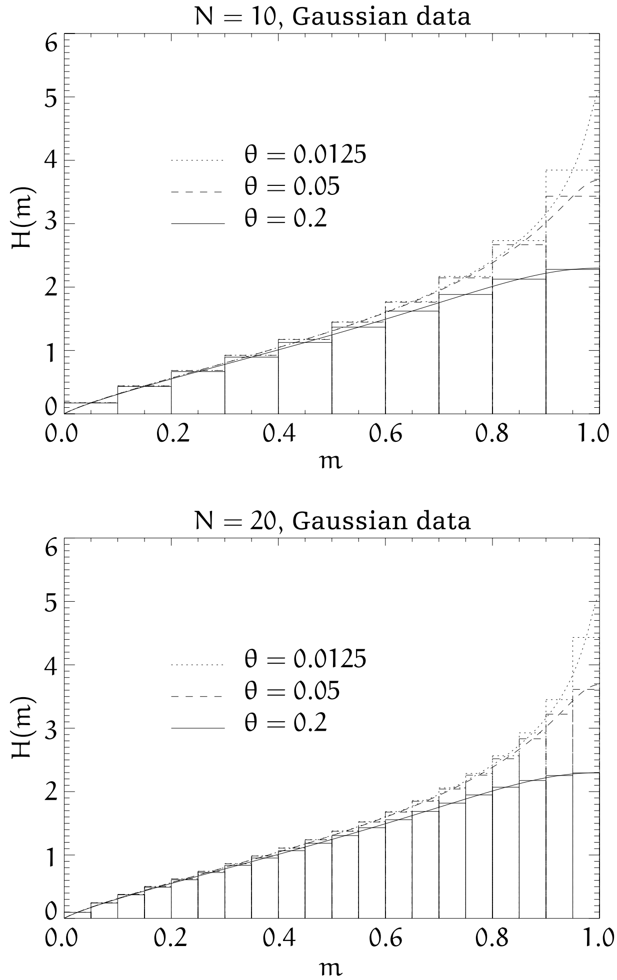

Figure 5.1 The midpoint rule Riemann sums approximating the integral, pictured here for the Gaussian initial data $A(x) = \sqrt{\pi}e^{-x^2}$. When the peak of $H(m)$ becomes underresolved for small θ, the Riemann sums underestimate the value of the integral by an amount that is of the order $\hbar_N$.

logarithmic singularity of $H(m)$ is $(1 - \log 2)\hbar_N + O(\hbar_N^2)$, which clearly dominates the $O(\hbar_N^2)$ error contributed by the majority of the subintervals bounded away from $m = M$. Consequently, for those λ on the circle $|\lambda| = A$ for which $\log|S(\lambda)|$ is not asymptotically small in $\hbar_N$, it is negative, and therefore $S(\lambda)$ is uniformly bounded for $|\lambda| = A$, as is $T(\lambda)$. ∎

Using this information, we can finally extract enough information about $T(\lambda)$ on the imaginary axis to approximate $a_M(\lambda)$ for $\lambda \in C_M$.

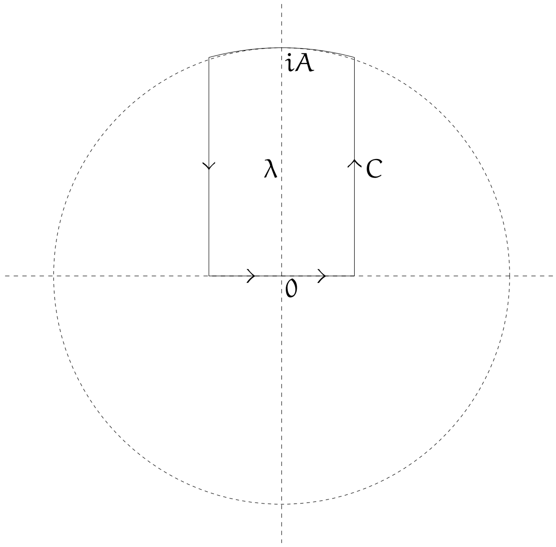

Figure 5.2 The contour C of the Cauchy integral argument.

Proposition 5.6. Let C_M be a fixed contour from $\lambda = 0$ to $\lambda = iA$ lying between C_L and C_R, possibly coinciding with the imaginary axis. Then, for $\mu > 0$ arbitrarily small,

$$a_M(\lambda) = i \exp\left(\frac{1}{\hbar_N}\left(2i\lambda|x| + \int_0^{iA} \overline{L}_\eta^0(\lambda)\rho^0(\eta)\, d\eta + \int_{-iA}^0 \overline{L}_\eta^0(\lambda)\rho^0(\eta^*)^*\, d\eta \right.\right.$$
$$\left.\left. - Jg_+(\lambda) - Jg_-(\lambda)\right)\right)\left(1 + O\left(\frac{\hbar_N^{1-\mu}}{|\lambda|}\right) + O\left(\frac{\hbar_N^{1-\mu}}{|\lambda - iA|}\right)\right), \tag{5.24}$$

as $\hbar_N$ goes to zero through positive values, for all $\lambda \in C_M$ with $|\lambda| > \hbar_N$ and $|\lambda - iA| > \hbar_N$. □

Proof. Let C be the closed contour illustrated in Figure 5.2. This counter-clockwise oriented contour consists of two vertical segments, one horizontal segment that lies on the real axis, and an arc of the circle of radius A centered at the origin. The function $T(\lambda)$ is analytic on the interior of C and is continuous on C itself. In fact it is analytic on most of the boundary, failing to be analytic only at $\lambda = 0$ and $\lambda = iA$. Therefore for any λ in the interior, we may write

$$T(\lambda) = 1 + \frac{1}{2\pi i}\oint_C \frac{T(s) - 1}{s - \lambda}\, ds. \tag{5.25}$$

If we let C_{in} denote the part of C with $|\Re(s)| < \hbar_N$, and let C_{out} denote the remaining portion of C, then we get

$$\left|T(\lambda) - 1\right| \leq \frac{1}{2\pi}\int_{C_{in}} \frac{|T(s)-1|}{|s-\lambda|}|ds| + \frac{1}{2\pi}\int_{C_{out}} \frac{|T(s)-1|}{|s-\lambda|}|ds|. \tag{5.26}$$

Using the estimate guaranteed by Lemma 5.4 in the integral over C_{out}, and the uniform boundedness of $T(s)$ (and therefore of $T(s) - 1$) guaranteed by Lemma 5.5 in the integral over C_{in}, we find

$$\left|T(\lambda) - 1\right| \leq K_{in}\hbar_N \sup_{s\in C_{in}} \frac{1}{|s-\lambda|} - K_{out}\hbar_N \log\hbar_N \sup_{s\in C_{out}} \frac{1}{|s-\lambda|} \tag{5.27}$$

for some positive constants K_{in} and K_{out}. Replacing the logarithm by a slightly cruder estimate of $\hbar_N^{-\mu}$ for arbitrarily small positive μ completes the proof. ■

We have therefore succeeded in showing that, at least away from the self-intersection points of the contour Σ, the jump matrices for $\mathbf{N}(\lambda)$ as defined by (4.2) for $\lambda \in C_L$, (4.4) for $\lambda \in C_R$, and (4.6) for $\lambda \in C_M$ are well approximated in the semiclassical limit $\hbar_N \to 0$ by matrices in which all nonzero matrix elements are of the form $\exp(f(\lambda)/\hbar_N)$ with $f(\lambda)$ being independent of $\hbar_N$. The fact that this approximation is valid even when the “active” contour C_M is taken to be right on top of the poles of the meromorphic Riemann-Hilbert problem for $\mathbf{m}(\lambda)$ is an advantage over the approach taken in [8].

Using these approximations, we can introduce an *ad hoc* approximation of the matrix $\mathbf{N}(\lambda)$. First, define

$$\begin{aligned}\tilde{\phi}(\lambda) := 2i\lambda|x| &+ \int_0^{iA} \overline{L}_\eta^0(\lambda)\rho^0(\eta)\,d\eta \\ &+ \int_{-iA}^0 \overline{L}_\eta^0(\lambda)\rho^0(\eta^*)^*\,d\eta - Jg_+(\lambda) - Jg_-(\lambda), \quad \text{for } \lambda \in C_M,\end{aligned} \tag{5.28}$$

and for $\lambda \in C_L$ or C_R, define

$$\tau(\lambda) := 2i\lambda|x| + \int_0^{iA} L_\eta^0(\lambda)\rho^0(\eta)\,d\eta + \int_{-iA}^0 L_\eta^0(\lambda)\rho^0(\eta^*)^*\,d\eta - 2Jg(\lambda). \tag{5.29}$$

Then we pose the following problem.

Riemann-Hilbert Problem 5.7 (formal continuum limit). Given a complex phase function $g(\lambda)$ find a matrix $\tilde{\mathbf{N}}(\lambda)$ satisfying

(1) Analyticity: $\tilde{\mathbf{N}}(\lambda)$ is analytic for $\lambda \in \mathbb{C} \setminus \Sigma$.

(2) Boundary behavior: $\tilde{\mathbf{N}}(\lambda)$ assumes continuous boundary values on Σ.

(3) Jump conditions: The boundary values taken on Σ satisfy

$$\tilde{\mathbf{N}}_+(\lambda) = \tilde{\mathbf{N}}_-(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} 1 & 0 \\ i\exp\left(\dfrac{\tau(\lambda)-i\theta^0(\lambda)}{\hbar_N}\right) & 1\end{bmatrix}\sigma_1^{(1-J)/2} \tag{5.30}$$

for $\lambda \in C_L$,

$$\tilde{\mathbf{N}}_+(\lambda) = \tilde{\mathbf{N}}_-(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} 1 & 0 \\ i\exp\left(\dfrac{\tau(\lambda)+i\theta^0(\lambda)}{\hbar_N}\right) & 1\end{bmatrix}\sigma_1^{(1-J)/2} \tag{5.31}$$

for $\lambda \in C_R$, and

$$\tilde{\mathbf{N}}_+(\lambda) = \tilde{\mathbf{N}}_-(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} \exp\left(\dfrac{i\theta(\lambda)}{\hbar_N}\right) & 0 \\ i\exp\left(\dfrac{\tilde{\phi}(\lambda)}{\hbar_N}\right) & \exp\left(-\dfrac{i\theta(\lambda)}{\hbar_N}\right)\end{bmatrix}\sigma_1^{(1-J)/2} \tag{5.32}$$

for $\lambda \in C_M$. For all other $\lambda \in \Sigma$ (i.e., in the lower half-plane), the jump is determined by the symmetry $\tilde{\mathbf{N}}(\lambda) = \sigma_2\tilde{\mathbf{N}}(\lambda^*)^*\sigma_2$.

(4) Normalization: $\tilde{\mathbf{N}}(\lambda)$ is normalized at infinity

$$\tilde{\mathbf{N}}(\lambda) \longrightarrow \mathbb{I} \quad \text{as } \lambda \longrightarrow \infty. \tag{5.33}$$

6 Choosing $g(\lambda)$ to arrive at an outer model

Let $R(\lambda)$ be defined by the equation $R(\lambda)^2 = \lambda^2 + A(x)^2$, the fact that $R(\lambda)$ is an analytic function for λ away from the imaginary interval $I := [-iA(x), iA(x)]$, and the normalization that for large λ, $R(\lambda) \sim -\lambda$. For $\eta \in I \cap C_M$, let

$$\rho(\eta) := \rho^0(\eta) + \frac{R_+(\eta)}{\pi i}\int_{-iA}^{-iA(x)} \frac{\rho^0(s^*)^*\,ds}{(\eta-s)R(s)} + \frac{R_+(\eta)}{\pi i}\int_{iA(x)}^{iA} \frac{\rho^0(s)\,ds}{(\eta-s)R(s)}. \tag{6.1}$$

It is easy to check directly that for all $\eta \in I \cap C_M$, we have $\rho(\eta) \in i\mathbb{R}_+$. Also, using the fact that $\rho^0(s)$ is purely imaginary on the imaginary axis, and that $R(s)$ is purely imaginary in the domain of integration, where it satisfies $R(-s) = -R(s)$, we see that

$$\rho(0) = \rho^0(0). \tag{6.2}$$

Furthermore, it follows easily from (6.1) that for all $\eta \in I \cap C_M$, we have

$$0 \le -i\rho(\eta) \le -i\rho^0(\eta), \tag{6.3}$$

with the lower constraint being achieved only at the endpoint[2] of I, $\lambda = iA(x)$, and the upper constraint being achieved only at the origin in accordance with (6.2).

Now, set

$$g(\lambda) := \frac{J}{2}\int_{-iA(x)}^{0} L_\eta^0(\lambda)\rho(\eta^*)^*\, d\eta + \frac{J}{2}\int_0^{iA(x)} L_\eta^0(\lambda)\rho(\eta)\, d\eta. \tag{6.4}$$

This function satisfies all of the basic criteria set out earlier: it is analytic in $\mathbb{C}\setminus(C_M \cup C_M^*)$ and takes continuous boundary values, it satisfies $g(\lambda) + g(\lambda^*)^* = 0$, and it satisfies $g(\infty) = 0$ because

$$\int_{-iA(x)}^{0} \rho(\eta^*)^*\, d\eta + \int_0^{iA(x)} \rho(\eta)\, d\eta = 0. \tag{6.5}$$

Note that $g(\lambda)$ is analytic across C_M for λ above $iA(x)$. Consequently, $\theta(\lambda) = 0$ for all such λ. For $\lambda \in C_M$ below $iA(x)$, $\theta(\lambda)$ becomes (cf. equation (4.8))

$$\theta(\lambda) = -\pi\int_\lambda^{iA(x)} \rho(\eta)\, d\eta. \tag{6.6}$$

We now describe a number of important consequences of our choice of $g(\lambda)$.

Proposition 6.1. For all $\lambda \in I \cap C_M = [0, iA(x)]$, $\tilde{\phi}(\lambda) = 0$. □

To prove the proposition, we first point out that

$$\lim_{\substack{\lambda\to 0\\ \lambda\in C_M}} \tilde{\phi}(\lambda) = 0, \tag{6.7}$$

simply as a consequence of the fact that both $\rho^0(\eta)$ and $\rho(\eta)$ are purely imaginary on C_M. Next we point out that

$$\tilde{\phi}'(\lambda) = 0 \tag{6.8}$$

whenever $\lambda \in [0, iA(x)]$. This follows from a direct calculation in which all integrals are evaluated by residues and the formula (2.6) is used.

[2]It is often convenient to think of the function $\rho(\eta)$ being extended to all of C_M by setting $\rho(\eta) \equiv 0$ for λ above the endpoint $iA(x)$. In this case one views the lower constraint as being active on the whole imaginary interval $[iA(x), iA]$.

Next we consider $\tilde{\phi}(\lambda)$ for $\lambda \in C_M \setminus [0, iA(x)]$, that is, above the endpoint of the support. Clearly, $\tilde{\phi}(\lambda)+i\theta(\lambda)$ is the boundary value on C_M of an analytic function defined near C_M in D_L. Since the boundary value taken below the endpoint is $i\theta(\lambda)$ because $\tilde{\phi}(\lambda) \equiv 0$ there, and the boundary value taken above the endpoint is $\tilde{\phi}(\lambda)$ because $\theta(\lambda) \equiv 0$ there, we obtain the formula

$$\tilde{\phi}(\lambda) = i\theta_+(\lambda) = -i\pi \int_{\lambda}^{iA(x)} \rho_+(\eta)\, d\eta \tag{6.9}$$

valid for $\lambda \in C_M$ above $iA(x)$, where by $\rho_+(\eta)$ for η in the imaginary interval $(iA(x), iA)$ we mean the function $\rho(\eta)$ defined by (6.1) for η in the imaginary interval $(0, iA(x))$, analytically continued from $(0, iA(x))$ in the clockwise direction about the endpoint $\lambda = iA(x)$. In particular, for such λ we have

$$\tilde{\phi}'(\lambda) = i\pi\rho_+(\lambda). \tag{6.10}$$

Carrying out the analytic continuation, we find from (6.1) that for $\eta \in (iA(x), iA)$,

$$\rho_+(\lambda) = \frac{R(\lambda)}{\pi i} \int_{-iA}^{-iA(x)} \frac{\rho^0(s^*)^*\, ds}{(\lambda - s)R(s)} + \frac{R(\lambda)}{\pi i}\, \text{P.V.} \int_{iA(x)}^{iA} \frac{\rho^0(s)\, ds}{(\lambda - s)R(s)}. \tag{6.11}$$

From this formula we see easily that for all λ strictly above the endpoint $iA(x)$, $\rho_+(\lambda)$ is positive real. Consequently, from (6.10) and since $\tilde{\phi}(\lambda) = 0$ for $\lambda = iA(x)$, we get the following result.

Proposition 6.2. The function $\tilde{\phi}(\lambda)$ is negative real and decreasing in the positive imaginary direction for $\lambda \in C_M \setminus [0, iA(x)]$. □

Now we consider the behavior of the function $\tau(\lambda)$ on C_L and C_R. From the definitions of the functions $\tau(\lambda)$ and $\tilde{\phi}(\lambda)$, we see that for $\lambda \in C_L$,

$$\tau(\lambda) = \tilde{\phi}(\lambda) + i\theta(\lambda) - i\theta^0(\lambda), \tag{6.12}$$

and for $\lambda \in C_R$,

$$\tau(\lambda) = \tilde{\phi}(\lambda) - i\theta(\lambda) + i\theta^0(\lambda). \tag{6.13}$$

That is, the analytic function $\tau(\lambda)$ takes boundary values from the left on C_M equal to $\tilde{\phi}(\lambda)+i\theta(\lambda)-i\theta^0(\lambda)$ and from the right on C_M equal to $\tilde{\phi}(\lambda)-i\theta(\lambda)+i\theta^0(\lambda)$. First consider the situation to the left or right of the imaginary interval $[0, iA(x)]$. Since $\tilde{\phi}(\lambda) \equiv 0$ in

$[0, iA(x)]$, the function $\tau(\lambda)$ on C_L will be the analytic continuation of $i\theta(\lambda) - i\theta^0(\lambda)$ from C_M and the function $\tau(\lambda)$ on C_R will be the analytic continuation of $-i\theta(\lambda) + i\theta^0(\lambda)$ from C_M. From (6.3) we see that for $\eta \in [0, iA(x)]$ one has $\rho^0(\eta) - \rho(\eta) \in i\mathbb{R}_+$. Therefore, it follows from the Cauchy-Riemann equations that for λ in portions of C_L and C_R close enough (independently of $\hbar_N$) to the interval $[0, iA(x)]$ one has

$$\Re(\tau(\lambda)) < 0 \tag{6.14}$$

for λ on *both* C_L and C_R. Furthermore, it follows from the fact that $\rho^0(\eta) \in i\mathbb{R}_+$ that $\Re(-i\theta^0(\lambda)) < 0$ for $\lambda \in C_L$ and $\Re(i\theta^0(\lambda)) < 0$ for $\lambda \in C_R$. Therefore,

$$\Re\big(\tau(\lambda) - i\theta^0(\lambda)\big) < 0 \tag{6.15}$$

for $\lambda \in C_L$ near the portion of C_M below $iA(x)$, and

$$\Re\big(\tau(\lambda) + i\theta^0(\lambda)\big) < 0 \tag{6.16}$$

for λ in the analogous portion of C_R. Next consider the situation to the left or right of the portion of C_M lying above the endpoint $\lambda = iA(x)$. Since $\theta(\lambda) \equiv 0$ and $\Re(\tilde{\phi}(\lambda)) < 0$ for $\lambda \in [iA(x), iA]$ we see that for C_L and C_R close enough (again independently of $\hbar_N$) to this part of C_M we again find that we have (6.15) on C_L and (6.16) on C_R. This shows that the jump matrix on both contours C_L and C_R is an exponentially small perturbation of the identity for small positive $\hbar_N$, pointwise in λ bounded away from the origin and iA.

For $\lambda \in [0, iA(x)]$, the jump matrix for $\tilde{\mathbf{N}}(\lambda)$ factors (recall $\tilde{\phi}(\lambda) \equiv 0$ here):

$$\begin{aligned}&\begin{bmatrix} \exp\left(\dfrac{i\theta(\lambda)}{\hbar_N}\right) & 0 \\ i & \exp\left(-\dfrac{i\theta(\lambda)}{\hbar_N}\right)\end{bmatrix} \\ &\quad = \begin{bmatrix} 1 & -i\exp\left(\dfrac{i\theta(\lambda)}{\hbar_N}\right) \\ 0 & 1\end{bmatrix} \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} \begin{bmatrix} 1 & -i\exp\left(-\dfrac{i\theta(\lambda)}{\hbar_N}\right) \\ 0 & 1\end{bmatrix}.\end{aligned} \tag{6.17}$$

Let L_L and L_R be two boundaries of a lens surrounding $[0, iA]$. See Figure 6.1. Using the factorization (6.17), we now define a new matrix function $\mathbf{O}(\lambda)$. In the region between L_L and C_M set

$$\mathbf{O}(\lambda) := \tilde{\mathbf{N}}(\lambda)\sigma_1^{(1-J)/2} \begin{bmatrix} 1 & i\exp\left(-\dfrac{i\theta(\lambda)}{\hbar_N}\right) \\ 0 & 1 \end{bmatrix} \sigma_1^{(1-J)/2}. \tag{6.18}$$

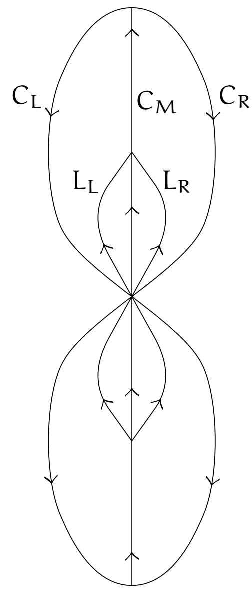

Figure 6.1 Introduction of the lens boundaries L_L and L_R.

In the region between C_M and L_R, set

$$\mathbf{O}(\lambda) := \tilde{\mathbf{N}}(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} 1 & -i\exp\left(\dfrac{i\theta(\lambda)}{\hbar_N}\right) \\ 0 & 1 \end{bmatrix}\sigma_1^{(1-J)/2}. \tag{6.19}$$

Elsewhere in the upper half-plane set $\mathbf{O}(\lambda) := \tilde{\mathbf{N}}(\lambda)$. And in the lower half-plane define $\mathbf{O}(\lambda)$ by symmetry: $\mathbf{O}(\lambda) = \sigma_2\mathbf{O}(\lambda^*)^*\sigma_2$.

These transformations imply jump conditions satisfied by $\mathbf{O}(\lambda)$ on the contours in Figure 6.1 since the jump conditions for $\tilde{\mathbf{N}}(\lambda)$ are given. For $\lambda \in L_L$ we have

$$\mathbf{O}_+(\lambda) = \mathbf{O}_-(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} 1 & -i\exp\left(-\dfrac{i\theta(\lambda)}{\hbar_N}\right) \\ 0 & 1 \end{bmatrix}\sigma_1^{(1-J)/2} \tag{6.20}$$

which is an exponentially small perturbation of the identity except near the endpoints. And for $\lambda \in L_R$ we have

$$\mathbf{O}_+(\lambda) = \mathbf{O}_-(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} 1 & -i\exp\left(\dfrac{i\theta(\lambda)}{\hbar_N}\right) \\ 0 & 1 \end{bmatrix}\sigma_1^{(1-J)/2} \tag{6.21}$$

which is also a jump that is exponentially close to the identity. For $\lambda \in [0, iA(x)]$ we get

$$\mathbf{O}_+(\lambda) = \mathbf{O}_-(\lambda) \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} \tag{6.22}$$

as a consequence of the factorization (6.17). Since $\mathbf{O}(\lambda) := \tilde{\mathbf{N}}(\lambda)$ for all λ in the upper half-plane outside the lens bounded by L_L and L_R, we see that $\mathbf{O}(\lambda)$ satisfies the following jump condition on C_L:

$$\mathbf{O}_+(\lambda) = \mathbf{O}_-(\lambda)\sigma_1^{(1-J)/2} \begin{bmatrix} 1 & 0 \\ i\exp\left(\dfrac{\tau(\lambda) - i\theta^0(\lambda)}{\hbar_N}\right) & 1 \end{bmatrix} \sigma_1^{(1-J)/2}, \tag{6.23}$$

the following jump relation on C_R:

$$\mathbf{O}_+(\lambda) = \mathbf{O}_-(\lambda)\sigma_1^{(1-J)/2} \begin{bmatrix} 1 & 0 \\ i\exp\left(\dfrac{\tau(\lambda) + i\theta^0(\lambda)}{\hbar_N}\right) & 1 \end{bmatrix} \sigma_1^{(1-J)/2}, \tag{6.24}$$

and the following jump relation on the imaginary interval $[iA(x), iA] \subset C_M$:

$$\mathbf{O}_+(\lambda) = \mathbf{O}_-(\lambda)\sigma_1^{(1-J)/2} \begin{bmatrix} 1 & 0 \\ i\exp\left(\dfrac{\tilde{\phi}(\lambda)}{\hbar_N}\right) & 1 \end{bmatrix} \sigma_1^{(1-J)/2}. \tag{6.25}$$

All three of these matrices are exponentially close to the identity matrix pointwise in λ for interior points of their respective contours.

The matrix $\mathbf{O}(\lambda)$ is related to $\tilde{\mathbf{N}}(\lambda)$ by explicit transformations. However, taking the pointwise limit of the jump matrix for $\mathbf{O}(\lambda)$ leads us to the following *ad hoc* Riemann-Hilbert problem.

Riemann-Hilbert Problem 6.3 (outer problem). Find a matrix $\tilde{\mathbf{O}}(\lambda)$ satisfying:

(1) Analyticity: $\tilde{\mathbf{O}}(\lambda)$ is analytic for $\lambda \in \mathbb{C} \setminus I$, where I is the imaginary interval $[-iA(x), iA(x)]$.

(2) Boundary behavior: $\tilde{\mathbf{O}}(\lambda)$ assumes boundary values that are continuous except at $\lambda = \pm iA(x)$, where at worst inverse fourth-root singularities are admitted.

(3) Jump condition: for $\lambda \in I$,

$$\tilde{\mathbf{O}}_+(\lambda) = \tilde{\mathbf{O}}_-(\lambda) \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}. \tag{6.26}$$

(4) Normalization: $\tilde{\mathbf{O}}(\lambda)$ is normalized at infinity:

$$\tilde{\mathbf{O}}(\lambda) \longrightarrow \mathbb{I} \quad \text{as } \lambda \longrightarrow \infty. \tag{6.27}$$

It is not difficult to solve this problem explicitly in terms of algebraic functions.

Proposition 6.4. The unique solution of Riemann-Hilbert Problem 6.3 is

$$\tilde{\mathbf{O}}(\lambda) := \frac{1}{2R(\lambda)\beta(\lambda)}\begin{bmatrix} R(\lambda) - \lambda - iA(x) & R(\lambda) + \lambda + iA(x) \\ R(\lambda) + \lambda + iA(x) & R(\lambda) - \lambda - iA(x) \end{bmatrix}, \tag{6.28}$$

where $R(\lambda)^2 = \lambda^2 + A(x)^2$ and

$$\beta(\lambda)^4 = \frac{\lambda + iA(x)}{\lambda - iA(x)}, \tag{6.29}$$

with both functions $R(\lambda)$ and $\beta(\lambda)$ being analytic in $\mathbb{C} \setminus I$, normalized according to $R(\lambda) \sim -\lambda$ and $\beta(\lambda) \sim 1$ as $\lambda \to \infty$. □

Using the matrix $\tilde{\mathbf{O}}(\lambda)$, we define an "outer" model for the matrix $\mathbf{N}(\lambda)$ as follows. The idea is to recall the relationship between the matrix $\tilde{\mathbf{N}}(\lambda)$ and $\mathbf{O}(\lambda)$, and simply substitute $\tilde{\mathbf{O}}(\lambda)$ for $\mathbf{O}(\lambda)$ in these formulae. For λ in between L_L and C_M, we use (6.18) to set

$$\widehat{\mathbf{N}}_{\text{out}}(\lambda) := \tilde{\mathbf{O}}(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} 1 & -i\exp\left(-\dfrac{i\theta(\lambda)}{\hbar_N}\right) \\ 0 & 1 \end{bmatrix}\sigma_1^{(1-J)/2}. \tag{6.30}$$

For λ in between C_M and L_R, we use (6.19) to set

$$\widehat{\mathbf{N}}_{\text{out}}(\lambda) := \tilde{\mathbf{O}}(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} 1 & i\exp\left(\dfrac{i\theta(\lambda)}{\hbar_N}\right) \\ 0 & 1 \end{bmatrix}\sigma_1^{(1-J)/2}. \tag{6.31}$$

For all other λ in the upper half-plane, set $\widehat{\mathbf{N}}_{\text{out}}(\lambda) := \tilde{\mathbf{O}}(\lambda)$, and in the lower half-plane set $\widehat{\mathbf{N}}_{\text{out}}(\lambda) := \sigma_2\widehat{\mathbf{N}}_{\text{out}}(\lambda^*)^*\sigma_2$. The important properties of this matrix are the following.

Proposition 6.5. The matrix $\widehat{\mathbf{N}}_{\text{out}}(\lambda)$ is analytic for all complex λ except at the contours L_L, L_R, the imaginary interval $[0, iA(x)]$, and their complex-conjugates. It satisfies the following jump conditions:

$$\widehat{\mathbf{N}}_{\mathrm{out},+}(\lambda)=\widehat{\mathbf{N}}_{\mathrm{out},-}(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix}1 & i\exp\left(-\dfrac{i\theta(\lambda)}{\hbar_N}\right)\\ 0 & 1\end{bmatrix}\sigma_1^{(1-J)/2},\quad \text{for } \lambda\in L_L,$$

$$\widehat{\mathbf{N}}_{\mathrm{out},+}(\lambda)=\widehat{\mathbf{N}}_{\mathrm{out},-}(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix}1 & i\exp\left(\dfrac{i\theta(\lambda)}{\hbar_N}\right)\\ 0 & 1\end{bmatrix}\sigma_1^{(1-J)/2},\quad \text{for } \lambda\in L_R,$$

$$\begin{aligned}\widehat{\mathbf{N}}_{\mathrm{out},+}(\lambda)&=\widehat{\mathbf{N}}_{\mathrm{out},-}(\lambda)\sigma_1^{(1-J)/2}\\ &\quad\times\begin{bmatrix}\exp\left(\dfrac{i\theta(\lambda)}{\hbar_N}\right) & 0\\ i & \exp\left(-\dfrac{i\theta(\lambda)}{\hbar_N}\right)\end{bmatrix}\sigma_1^{(1-J)/2},\quad \text{for } \lambda\in\left[0,iA(x)\right],\end{aligned}\tag{6.32}$$

with the jump matrices on the conjugate contours in the lower half-plane being obtained from these by the symmetry $\widehat{\mathbf{N}}_{\mathrm{out}}(\lambda^*)=\sigma_2\widehat{\mathbf{N}}_{\mathrm{out}}(\lambda)^*\sigma_2$. In particular, note that for $\lambda\in[0,iA(x)]$, we have $\widehat{\mathbf{N}}_{\mathrm{out},-}(\lambda)^{-1}\widehat{\mathbf{N}}_{\mathrm{out},+}(\lambda)=\tilde{\mathbf{N}}_-(\lambda)^{-1}\tilde{\mathbf{N}}_+(\lambda)$. Also, if D is any given open set containing the endpoint $\lambda=iA(x)$, then $\widehat{\mathbf{N}}_{\mathrm{out}}(\lambda)$ is uniformly bounded for $\lambda\in\mathbb{C}\setminus(D\cup D^*)$ with a bound that depends only on D and not on $\hbar_N$. □

7 Local analysis

In justifying formally the local model $\widehat{\mathbf{N}}_{\mathrm{out}}(\lambda)$, we ignored the fact that the pointwise asymptotics for the jump matrices for $\mathbf{O}(\lambda)$ that we used to obtain the matrix $\tilde{\mathbf{O}}(\lambda)$ were not uniform near the origin or near the moving endpoint $\lambda=iA(x)$. We also neglected the breakdown of the asymptotics for $a_L(\lambda)$, $a_R(\lambda)$, and $a_M(\lambda)$ near the points $\lambda=0$ and $\lambda=iA$. Consequently, we do not expect the outer model $\widehat{\mathbf{N}}_{\mathrm{out}}(\lambda)$ to be a good approximation to $\mathbf{N}(\lambda)$ near $\lambda=0$, $\lambda=iA(x)$, or $\lambda=iA$. In this section, we examine the neighborhoods of these three points in more detail, and we will obtain accurate local models for $\mathbf{N}(\lambda)$ in the corresponding neighborhoods.

7.1 Local analysis near $\lambda=0$

7.1.1 Local behavior of the matrix elements $a_L(\lambda)$, $a_R(\lambda)$, *and* $a_M(\lambda)$. Let ϵ and δ be small scales tied to $\hbar_N$ such that $\hbar_N\ll\delta\ll\epsilon\ll 1$ as $\hbar_N\downarrow 0$. Let L be defined as the unique integer for which exactly $N-L$ of the numbers $\lambda_0,\dots,\lambda_{N-1}$ lie strictly below $i\epsilon$ on the positive imaginary axis. We want to compute uniform asymptotics for $S(\lambda)$ defined by (5.2) for $\lambda\in C_L\cup C_R$, and for $T(\lambda)$ defined by (5.3) for $\lambda\in C_M$ when $|\lambda|\le\delta$.

Lemma 7.1. When $\Im(\lambda) \geq 0$ and $|\lambda| \leq \delta$ and with L defined as indicated in the preceding paragraph,

$$\exp\left(-\sum_{k=0}^{L-1} \tilde{I}_k(\lambda)\right) = 1 + O\left(\frac{\hbar_N}{\epsilon}\right). \tag{7.1}$$

□

Proof. We recall the integral formula (cf. equation (5.8))

$$\tilde{I}_k(\lambda) = \frac{1}{\hbar_N}\int_{m_k-\hbar_N/2}^{m_k+\hbar_N/2} dm \int_{m_k}^{m} d\zeta \int_{m_k}^{\zeta} d\xi\, g(\lambda,\xi), \tag{7.2}$$

in which we expand the integrand in partial fractions:

$$g(\lambda,\xi) = \frac{e''(\xi)}{\lambda + e(\xi)} + \frac{e''(\xi)}{\lambda - e(\xi)} - \frac{e'(\xi)^2}{(\lambda + e(\xi))^2} + \frac{e'(\xi)^2}{(\lambda - e(\xi))^2}. \tag{7.3}$$

Since $\Im(\lambda) \geq 0$, for $m_k - \hbar_N/2 \leq \xi \leq m_k + \hbar_N/2$ and $k = 0, \ldots, L-1$, we get

$$\begin{aligned} \frac{1}{|\lambda + e(\xi)|} \leq \frac{1}{|\lambda - e(\xi)|} &\leq \frac{1}{|i\delta - e(\xi)|} \\ &\leq \frac{1}{\left|i\delta - e\left(m_k - \frac{\hbar_N}{2}\right)\right|} = O\left(\frac{1}{\left|m(\delta) - m_k + \frac{\hbar_N}{2}\right|}\right). \end{aligned} \tag{7.4}$$

For such ξ we therefore have

$$g(\lambda,\xi) = O\left(\frac{1}{\left|m(\delta) - m_k + \frac{\hbar_N}{2}\right|^2}\right), \tag{7.5}$$

so summing over k gives

$$\begin{aligned} \sum_{k=0}^{L-1} \tilde{I}_k(\lambda) &= O\left(\hbar_N^2 \sum_{k=0}^{L-1} \frac{1}{\left|m(\delta) - m_k + \frac{\hbar_N}{2}\right|^2}\right) \\ &= O\left(\hbar_N \int_{m(\epsilon)}^{M} \frac{dm}{(m - m(\delta))^2}\right) = O\left(\frac{\hbar_N}{\epsilon}\right), \end{aligned} \tag{7.6}$$

because $\delta \ll \epsilon$, which proves the lemma. ■

So only the fraction of terms $\tilde{I}_k(\lambda)$ with $k \geq L$ contribute significantly to the sum for $\tilde{I}(\lambda)$. It is easy to check directly that $\exp(-\tilde{I}_k(\lambda))$ is an analytic function for $|\lambda| \leq \delta$

whenever $0 \le k \le L-1$, so it makes no difference in these terms whether it is $L^0_\eta(\lambda)$ or $\overline{L}^0_\eta(\lambda)$ that appears in the definition of $\tilde{I}_k$. Therefore, the terms in $S(\lambda)$ and $T(\lambda)$ that can be significant for λ near the origin are thus

$$\begin{aligned} S_1^{(0)}(\lambda) &:= \left(\prod_{k=L}^{N-1} \frac{\lambda-\lambda_k^*}{\lambda-\lambda_k}\right) \exp\left(\frac{1}{\hbar_N}\int_0^{m_L+\hbar_N/2} \left(L^0_{e(m)}(\lambda) - L^0_{-e(m)}(\lambda)\right) dm\right), \\ T_1^{(0)}(\lambda) &:= \left(\prod_{k=L}^{N-1} \frac{\lambda-\lambda_k^*}{\lambda-\lambda_k}\right) \exp\left(\frac{1}{\hbar_N}\int_0^{m_L+\hbar_N/2} \left(\overline{L}^0_{e(m)}(\lambda) - \overline{L}^0_{-e(m)}(\lambda)\right) dm\right) \\ &\quad \times 2\cos\left(\frac{\theta^0(\lambda)}{\hbar_N}\right). \end{aligned} \tag{7.7}$$

Here we have written the integrals in the exponent using the change of variables $m = m(\eta)$. So Lemma 7.1 simply says that $S(\lambda) = S_1^{(0)}(\lambda)(1 + O(\hbar_N/\epsilon))$ and $T(\lambda) = T_1^{(0)}(\lambda)(1 + O(\hbar_N/\epsilon))$ uniformly for $|\lambda| < \delta$. When λ is close to the origin along with the points λ_k contributing to $T(\lambda)$, the ladder of discrete nodes appears to become equally spaced. The next lemma shows that this is indeed the case.

Lemma 7.2. Let $\tilde{\lambda}_{N-k}$ for $k = 1, 2, 3, \ldots$ be the sequence of numbers defined by the relation

$$\tilde{\lambda}_{N-k} := -\frac{\hbar_N}{\rho^0(0)}\left(k - \frac{1}{2}\right), \tag{7.8}$$

which results from expanding the Bohr-Sommerfeld relation (2.9) for λ_{N-k} small, and keeping only the dominant terms. Define

$$\begin{aligned} S_2^{(0)}(\lambda) &:= \left(\prod_{k=L}^{N-1} \frac{\lambda-\tilde{\lambda}_k^*}{\lambda-\tilde{\lambda}_k}\right) \exp\left(\frac{1}{\hbar_N}\int_0^{m_L+\hbar_N/2} \left(L^0_{e'(0)m}(\lambda) - L^0_{-e'(0)m}(\lambda)\right) dm\right), \\ T_2^{(0)}(\lambda) &:= \left(\prod_{k=L}^{N-1} \frac{\lambda-\tilde{\lambda}_k^*}{\lambda-\tilde{\lambda}_k}\right) \exp\left(\frac{1}{\hbar_N}\int_0^{m_L+\hbar_N/2} \left(\overline{L}^0_{e'(0)m}(\lambda) - \overline{L}^0_{-e'(0)m}(\lambda)\right) dm\right) \\ &\quad \times 2\cos\left(\frac{\pi\rho^0(0)}{\hbar_N}(iA - \lambda)\right). \end{aligned} \tag{7.9}$$

Then, for $\Im(\lambda) \ge 0$ and $|\lambda| \le \delta$,

$$T_1^{(0)}(\lambda) = T_2^{(0)}(\lambda)\left(1 + O\left(\frac{\epsilon^2}{\hbar_N}\log\left(\frac{\epsilon}{\hbar_N}\right)\right)\right), \tag{7.10}$$

where we suppose that the scale ϵ is further constrained so that the relative error is asymptotically small. If λ is additionally bounded outside of some sector containing the positive imaginary axis, then

$$S_1^{(0)}(\lambda) = S_2^{(0)}(\lambda)\left(1 + O\left(\frac{\epsilon^2}{\hbar_N}\right)\right). \tag{7.11}$$

□

Proof. We begin by observing that for $k = L, \dots, N-1$, the distance between λ_k and $\tilde{\lambda}_k$ is much smaller than the distance between λ_k and λ_{k+1}, as long as $\epsilon \ll \hbar_N^{1/2}$. More precisely, we have

$$\left|\tilde{\lambda}_k - \lambda_k\right| = O\left(\hbar_N^2 (N-k)^2\right). \tag{7.12}$$

Decompose the quotients as follows:

$$\frac{T_1^{(0)}(\lambda)}{T_2^{(0)}(\lambda)} = D(\lambda)C(\lambda)\overline{L}(\lambda), \qquad \frac{S_1^{(0)}(\lambda)}{S_2^{(0)}(\lambda)} = D(\lambda)L(\lambda), \tag{7.13}$$

where

$$\begin{aligned} D(\lambda) &:= \prod_{k=L}^{N-1} \frac{\lambda - \lambda_k^*}{\lambda - \lambda_k} \frac{\lambda - \tilde{\lambda}_k}{\lambda - \tilde{\lambda}_k^*}, \\ C(\lambda) &:= \cos\left(\frac{\pi}{\hbar_N}\int_\lambda^{iA} \rho^0(\eta)\,d\eta\right) \sec\left(-\pi N - \frac{\pi}{\hbar_N}\rho^0(0)\lambda\right), \end{aligned} \tag{7.14}$$

$$\begin{aligned} \overline{L}(\lambda) &:= \exp\left(\frac{1}{\hbar_N}\int_0^{m_L+\hbar_N/2} \left(\left[\overline{L}^0_{e(m)}(\lambda) - \overline{L}^0_{e'(0)m}(\lambda)\right]\right.\right. \\ &\qquad\qquad \left.\left. - \left[\overline{L}^0_{-e(m)}(\lambda) - \overline{L}^0_{-e'(0)m}(\lambda)\right]\right) dm\right), \\ L(\lambda) &:= \exp\left(\frac{1}{\hbar_N}\int_0^{m_L+\hbar_N/2} \left(\left[L^0_{e(m)}(\lambda) - L^0_{e'(0)m}(\lambda)\right]\right.\right. \\ &\qquad\qquad \left.\left. - \left[L^0_{-e(m)}(\lambda) - L^0_{-e'(0)m}(\lambda)\right]\right) dm\right). \end{aligned} \tag{7.15}$$

First we deal with $L(\lambda)$ and $\overline{L}(\lambda)$. Since $e(m)$ is smooth and m is small we have $e(m) - e'(0)m = O(\epsilon^2)$. Also, the interval of integration is $O(\epsilon)$ in length. Although the integrands in (7.15) are not pointwise small, upon integration it follows that

$$L(\lambda) = 1 + O\left(\frac{\epsilon^3}{\hbar_N}\right), \qquad \overline{L}(\lambda) = 1 + O\left(\frac{\epsilon^3}{\hbar_N}\right), \tag{7.16}$$

uniformly for all λ in the upper half-plane satisfying $|\lambda| \leq \delta$. Here we are assuming that $\epsilon \ll \hbar_N^{1/3}$.

For the moment, we drop the conditions $\Im(\lambda) \geq 0$ and $|\lambda| \leq \delta$ and instead consider λ to lie on the sides of the square centered at the origin, one of whose sides is parallel to the real axis and intersects the positive imaginary axis halfway between the points $\lambda = \tilde{\lambda}_L$ and $\lambda = \tilde{\lambda}_{L-1}$. Note that the estimate (7.12) implies that the sides of the square intersect the real and imaginary axes a distance from the origin that is approximately ϵ. Therefore the square asymptotically contains the closed disk $|\lambda| \leq \delta$ because $\delta \ll \epsilon$. We will show that for λ on the four sides of the square, both $D(\lambda)$ and $C(\lambda)$ are very close to one. We write $D(\lambda)$ in the form

$$D(\lambda) = \prod_{k=L}^{N-1} \left(1 + \frac{\tilde{\lambda}_k^* - \lambda_k^*}{\lambda - \tilde{\lambda}_k^*}\right)\left(1 + \frac{\tilde{\lambda}_k - \lambda_k}{\lambda - \tilde{\lambda}_k}\right)^{-1}. \tag{7.17}$$

First consider the top of the square: for $\Im(\lambda) = -i(\tilde{\lambda}_L + \tilde{\lambda}_{L-1})/2$, we easily see that

$$|\lambda - \tilde{\lambda}_k| \geq \frac{i\hbar_N}{\rho^0(0)}\left(k - L + \frac{1}{2}\right), \qquad \frac{1}{|\lambda - \tilde{\lambda}_k^*|} = O\left(\frac{1}{\epsilon}\right), \tag{7.18}$$

for $k = L, \ldots, N-1$. Combining this with (7.12), we get

$$\frac{\tilde{\lambda}_k^* - \lambda_k^*}{\lambda - \tilde{\lambda}_k^*} = O\left(\frac{\hbar_N^2 (N-k)^2}{\epsilon}\right), \qquad \frac{\tilde{\lambda}_k - \lambda_k}{\lambda - \tilde{\lambda}_k} = O\left(\frac{\hbar_N^2 (N-k)^2}{\hbar_N \left(k - L + \frac{1}{2}\right)}\right). \tag{7.19}$$

Summing these estimates over k (it is convenient to approximate sums by integrals in doing so), we find that

$$\begin{aligned} \prod_{k=L}^{N-1} \left(1 + \frac{\tilde{\lambda}_k^* - \lambda_k^*}{\lambda - \tilde{\lambda}_k^*}\right) &= 1 + O\left(\frac{\epsilon^2}{\hbar_N}\right), \\ \prod_{k=L}^{N-1} \left(1 + \frac{\tilde{\lambda}_k - \lambda_k}{\lambda - \tilde{\lambda}_k}\right)^{-1} &= 1 + O\left(\frac{\epsilon^2}{\hbar_N} \log\left(\frac{\epsilon}{\hbar_N}\right)\right). \end{aligned} \tag{7.20}$$

Consequently, for λ on the top of the square,

$$D(\lambda) = 1 + O\left(\frac{\epsilon^2}{\hbar_N} \log\left(\frac{\epsilon}{\hbar_N}\right)\right). \tag{7.21}$$

An estimate of the same form holds when λ is on the bottom of the square, where

$\Im(\lambda) = i(\tilde{\lambda}_L + \tilde{\lambda}_{L-1})/2$. When λ is on the left or right side of the square, so that $|\Re(\lambda)| = -i(\tilde{\lambda}_L + \tilde{\lambda}_{L-1})/2$, both $|\lambda - \tilde{\lambda}_k^*|^{-1}$ and $|\lambda - \tilde{\lambda}_k|^{-1}$ are $O(\epsilon^{-1})$. By the same arguments as above, we then have for such λ that

$$D(\lambda) = 1 + O\left(\frac{\epsilon^2}{\hbar_N}\right). \tag{7.22}$$

Now we look at $C(\lambda)$ on the same square. Generally, for such λ which are of order ϵ in magnitude, we have

$$C(\lambda) = 1 + O\left(\frac{\epsilon^2}{\hbar_N}\right) \sec\left(-\pi N - \frac{\pi}{\hbar_N}\rho^0(0)\lambda\right). \tag{7.23}$$

When λ is on the top or bottom of the square, we have

$$\left|\sec\left(-\pi N - \frac{\pi}{\hbar_N}\rho^0(0)\lambda\right)\right| \leq 1, \tag{7.24}$$

and when λ is on the left or right sides of the square, the same quantity is exponentially small. It follows easily that for λ on any of the sides of the square,

$$C(\lambda) = 1 + O\left(\frac{\epsilon^2}{\hbar_N}\right). \tag{7.25}$$

So uniformly on the four sides of the square, we have

$$D(\lambda)C(\lambda) = 1 + O\left(\frac{\epsilon^2}{\hbar_N}\log\left(\frac{\epsilon}{\hbar_N}\right)\right). \tag{7.26}$$

But the product $D(\lambda)C(\lambda)$ is analytic within the square, so by the maximum principle it follows that the same estimate holds for all λ on the interior of the square, and in particular for all λ in the upper half-plane with $|\lambda| \leq \delta$. This shows that

$$T_1^{(0)}(\lambda) = T_2^{(0)}(\lambda)\left(1 + O\left(\frac{\epsilon^2}{\hbar_N}\log\left(\frac{\epsilon}{\hbar_N}\right)\right)\right) \tag{7.27}$$

holds for all such λ.

Now to control the relationship between $S_1^{(0)}(\lambda)$ and $S_2^{(0)}(\lambda)$ we consider λ to lie outside of some symmetrical sector about the positive imaginary axis, of arbitrarily small nonzero opening angle 2α independent of $\hbar_N$. Since $\Im(\lambda) \geq 0$, we get

$$|\lambda - \tilde{\lambda}_k^*| \geq |\lambda - \tilde{\lambda}_k| \geq \frac{|\tilde{\lambda}_k|}{\sin(\alpha)} = \frac{i\hbar_N\left(N - k - \frac{1}{2}\right)}{\rho^0(0)|\sin(\alpha)|}. \tag{7.28}$$

Combining this result with (7.12), we find

$$\frac{\tilde{\lambda}_k^* - \lambda_k^*}{\lambda - \tilde{\lambda}_k^*} = O(\hbar_N(N-k)), \qquad \frac{\tilde{\lambda}_k - \lambda_k}{\lambda - \tilde{\lambda}_k} = O(\hbar_N(N-k)). \tag{7.29}$$

Summing these estimates over k one finds that

$$D(\lambda) = 1 + O\left(\frac{\epsilon^2}{\hbar_N}\right). \tag{7.30}$$

Combining this with the estimate (7.16) of $L(\lambda) - 1$, we find that

$$S_1^{(0)}(\lambda) = S_2^{(0)}(\lambda)\left(1 + O\left(\frac{\epsilon^2}{\hbar_N}\right)\right), \tag{7.31}$$

for all λ in the upper half-plane with $|\lambda| < \delta$ and bounded outside of the sector of opening angle 2α about the positive imaginary axis. This completes the proof. ∎

Without any approximation, $S_2^{(0)}(\lambda)$ can be rewritten in the form

$$S_2^{(0)}(\lambda) = (-i\zeta)^{-i\zeta}(i\zeta)^{-i\zeta}\frac{\Gamma\left(\frac{1}{2}+i\zeta\right)(\overline{N}+i\zeta)^{\overline{N}+i\zeta}\Gamma\left(\overline{N}+\frac{1}{2}-i\zeta\right)}{\Gamma\left(\frac{1}{2}-i\zeta\right)(\overline{N}-i\zeta)^{\overline{N}-i\zeta}\Gamma\left(\overline{N}+\frac{1}{2}+i\zeta\right)} \tag{7.32}$$

and $T_2^{(0)}(\lambda)$ can be rewritten in the form

$$T_2^{(0)}(\lambda) = \frac{2\pi}{\Gamma\left(\frac{1}{2}-i\zeta\right)^2}(-i\zeta)^{-2i\zeta}\frac{(\overline{N}+i\zeta)^{\overline{N}+i\zeta}\Gamma\left(\overline{N}+\frac{1}{2}-i\zeta\right)}{(\overline{N}-i\zeta)^{\overline{N}-i\zeta}\Gamma\left(\overline{N}+\frac{1}{2}+i\zeta\right)}, \tag{7.33}$$

where $\overline{N} := N - L$ and we are introducing a transformation φ_0 to a local variable ζ given by

$$\zeta = \varphi_0(\lambda) := -\frac{i\rho^0(0)\lambda}{\hbar_N}. \tag{7.34}$$

These formulae come from evaluating the logarithmic integrals exactly, which is possible because $e(m)$ has been replaced by the linear function $e'(0)m$, taking advantage of the equal spacing of the $\tilde{\lambda}_k$ to write the product explicitly in terms of gamma functions, and then using the reflection identity for the gamma function to eliminate the cosine from $T_2^{(0)}(\lambda)$. Now, the integer $\overline{N}$ is large, approximately of size $\epsilon/\hbar_N$. But for $|\lambda| \le \delta$, $\overline{N}$ is asymptotically large compared to ζ because $\delta \ll \epsilon$. These observations allow us to apply Stirling-type asymptotics to $S_2^{(0)}(\lambda)$ and $T_2^{(0)}(\lambda)$.

Lemma 7.3. In addition to all prior hypotheses, suppose that $\delta^2 \ll \epsilon\hbar_N$. Then,

$$S_2^{(0)}(\lambda) = e^{2i\zeta}(-i\zeta)^{-i\zeta}(i\zeta)^{-i\zeta}\frac{\Gamma\left(\frac{1}{2} - i\zeta\right)}{\Gamma\left(\frac{1}{2} + i\zeta\right)}\left(1 + O\left(\frac{\delta^2}{\epsilon\hbar_N}\right)\right),$$
$$T_2^{(0)}(\lambda) = \frac{2\pi e^{2i\zeta}(-i\zeta)^{-2i\zeta}}{\Gamma\left(\frac{1}{2} - i\zeta\right)^2}\left(1 + O\left(\frac{\delta^2}{\epsilon\hbar_N}\right)\right). \tag{7.35}$$

□

Proof. Asymptotically expanding the gamma functions for large $\overline{N}$, we find that

$$S_2^{(0)}(\lambda) = e^{2i\zeta}(-i\zeta)^{-i\zeta}(i\zeta)^{-i\zeta}\frac{\Gamma\left(\frac{1}{2} + i\zeta\right)}{\Gamma\left(\frac{1}{2} - i\zeta\right)} \cdot \Delta(\zeta, \overline{N}) \cdot \left(1 + O\left(\frac{1}{\overline{N}}\right)\right),$$
$$T_2^{(0)}(\lambda) = \frac{2\pi e^{2i\zeta}(-i\zeta)^{-2i\zeta}}{\Gamma\left(\frac{1}{2} - i\zeta\right)^2} \cdot \Delta(\zeta, \overline{N}) \cdot \left(1 + O\left(\frac{1}{\overline{N}}\right)\right), \tag{7.36}$$

where

$$\Delta(\zeta, \overline{N}) := \frac{(\overline{N} + i\zeta)^{\overline{N}+i\zeta}}{\left(\overline{N} + i\zeta + \frac{1}{2}\right)^{\overline{N}+i\zeta}} \frac{\left(\overline{N} - i\zeta + \frac{1}{2}\right)^{\overline{N}-i\zeta}}{(\overline{N} - i\zeta)^{\overline{N}-i\zeta}}. \tag{7.37}$$

Next, expanding $\Delta(\zeta, \overline{N})$, one gets worse error terms

$$\Delta(\zeta, \overline{N}) = 1 + O\left(\left(\frac{\delta}{\hbar_N}\right)^2 \frac{1}{\overline{N}}\right). \tag{7.38}$$

Combining these estimates and noting that $1/\overline{N} = O(\hbar_N/\epsilon)$ completes the proof of the lemma. ■

With these results in hand, we can easily establish the following.

Proposition 7.4. Let λ be in the upper half-plane, with $|\lambda| \leq \hbar_N^\alpha$, where $3/4 < \alpha < 1$, and let λ be bounded outside of some fixed symmetrical sector containing the positive imaginary axis. Then

$$S(\lambda) = e^{2i\zeta}(-i\zeta)^{-i\zeta}(i\zeta)^{-i\zeta}\frac{\Gamma\left(\frac{1}{2}+i\zeta\right)}{\Gamma\left(\frac{1}{2}-i\zeta\right)}\left(1+O\left(\hbar_N^{4\alpha/3-1}\right)\right), \tag{7.39}$$

where $\zeta = \varphi_0(\lambda) := -i\rho^0(0)\lambda/\hbar_N$. □

Proof. According to Lemmas 7.1, 7.2, and 7.3, the total relative error is a sum of three terms

$$O\left(\frac{\hbar_N}{\epsilon}\right), \qquad O\left(\frac{\epsilon^2}{\hbar_N}\right), \qquad O\left(\frac{\delta^2}{\epsilon\hbar_N}\right). \tag{7.40}$$

Note that since $\hbar_N \ll \delta$, the order $\hbar_N/\epsilon$ term is always dominated asymptotically by the order $\delta^2/\epsilon\hbar_N$ term. The error is optimized by picking ϵ so that the two possibly dominant terms are in balance. This forces us to choose $\epsilon \sim \delta^{2/3}$. The proposition follows upon taking $\delta = \hbar_N^\alpha$. ■

Proposition 7.5. Let λ be in the upper half-plane, with $|\lambda| \le \hbar_N^\alpha$, where $3/4 < \alpha < 1$. Then for all $\nu > 0$, however small,

$$T(\lambda) = \frac{2\pi e^{2i\zeta}(-i\zeta)^{-2i\zeta}}{\Gamma\left(\frac{1}{2}-i\zeta\right)^2}\left(1+O\left(\hbar_N^{4\alpha/3-1-\nu}\right)\right), \tag{7.41}$$

where $\zeta = \varphi_0(\lambda) := -i\rho^0(0)\lambda/\hbar_N$. □

Proof. In this case, according to Lemmas 7.1, 7.2, and 7.3, the total relative error is a sum of three different terms

$$O\left(\frac{\hbar_N}{\epsilon}\right), \qquad O\left(\frac{\epsilon^2}{\hbar_N}\log\left(\frac{\epsilon}{\hbar_N}\right)\right), \qquad O\left(\frac{\delta^2}{\epsilon\hbar_N}\right). \tag{7.42}$$

Again, since $\hbar_N \ll \delta$, the order $\hbar_N/\epsilon$ term is always dominated asymptotically by the order $\delta^2/\epsilon\hbar_N$ term. For any $\sigma > 0$, we have

$$\frac{\epsilon^2}{\hbar_N}\log\left(\frac{\epsilon}{\hbar_N}\right) = O\left(\frac{\epsilon^2}{\hbar_N}\left(\frac{\epsilon}{\hbar_N}\right)^\sigma\right). \tag{7.43}$$

So we can eliminate the logarithm at the expense of a slightly larger error. Taking $\delta = \hbar_N^\alpha$ as in the statement of the proposition, and using the cruder estimate (7.43), the nearly optimal value of ϵ to minimize the total relative error is achieved by a dominant balance between the right-hand side of (7.43) and the term of order $\delta^2/\epsilon\hbar_N$. The balance gives $\epsilon = \hbar_N^{\beta}$, with

$$\beta = \frac{2\alpha+\sigma}{3+\sigma}. \tag{7.44}$$

With this choice of ϵ, the total relative error is of the order $\hbar_N^{\gamma}$, with

$$\gamma = 2\alpha - 1 - \beta = \frac{4\alpha + 2(\alpha-1)\sigma - 3}{3+\sigma} < \frac{4}{3}\alpha - 1 \tag{7.45}$$

with the inequality following because $\sigma > 0$ and $\alpha < 1$. The inequality fails in the limit $\sigma \to 0$. Therefore, for each arbitrarily small $\nu > 0$, we can find a $\sigma > 0$ sufficiently small that $\gamma > 4\alpha/3 - 1 - \nu$. This gives us a slightly less optimal estimate of the relative error: simply $O(\hbar_N^{4\alpha/3-1-\nu})$, which completes the proof. ∎

7.1.2 The model Riemann-Hilbert problem. To repair the flaw in our model $\widehat{\mathbf{N}}_{\mathrm{out}}(\lambda)$ for the matrix $\mathbf{N}(\lambda)$ related to the nonuniformity of the approximation of the jump matrices near the origin, we need to provide a different approximation of $\mathbf{N}(\lambda)$ that will be valid when $|\lambda| \le \hbar_N^{\alpha}$ for some $\alpha \in (3/4, 1)$. The local failure of the "outer" approximation is gauged by the deviation of the matrix quotient $\mathbf{N}(\lambda)\widehat{\mathbf{N}}_{\mathrm{out}}(\lambda)^{-1}$ from the identity matrix near the origin. It turns out to be more convenient to study a conjugated form of this matrix (which also deviates from the identity for λ near the origin). Namely, for $|\lambda| \le \hbar_N^{\alpha}$, set

$$\begin{aligned}\mathbf{F}(\lambda) :&= e^{-i\theta(0)\sigma_3/(2\hbar_N)}\,\sigma_1^{(1-J)/2}\,(i\sigma_1)\tilde{\mathbf{O}}(\lambda)^{-1}\mathbf{N}(\lambda)\\ &\quad\times \widehat{\mathbf{N}}_{\mathrm{out}}(\lambda)^{-1}\tilde{\mathbf{O}}(\lambda)(-i\sigma_1)\sigma_1^{(1-J)/2}\,e^{i\theta(0)\sigma_3/(2\hbar_N)}\end{aligned} \tag{7.46}$$

if $\Re(\lambda) < 0$ and

$$\mathbf{F}(\lambda) := e^{-i\theta(0)\sigma_3/(2\hbar_N)}\,\sigma_1^{(1-J)/2}\,\tilde{\mathbf{O}}(\lambda)^{-1}\mathbf{N}(\lambda)\widehat{\mathbf{N}}_{\mathrm{out}}(\lambda)^{-1}\tilde{\mathbf{O}}(\lambda)\sigma_1^{(1-J)/2}\,e^{i\theta(0)\sigma_3/(2\hbar_N)} \tag{7.47}$$

if $\Re(\lambda) > 0$. It is easy to check that as a consequence of the boundary conditions satisfied by the matrix $\tilde{\mathbf{O}}(\lambda)$ on the imaginary axis near the origin, the conjugating factors are analytic throughout the disk $|\lambda| \le \hbar_N^{\alpha}$, and are uniformly bounded there independently of $\hbar_N$.

For later convenience, we assume, without loss of generality, that the auxiliary contours C_L, C_R, L_L, and L_R are straight rays in some $\hbar_N$-independent neighborhood of the origin. It is easy to write down the jump conditions satisfied by $\mathbf{F}(\lambda)$ on these four rays and also on the positive imaginary axis. We find that

$$\begin{aligned}
\mathbf{F}_+(\lambda) &= \mathbf{F}_-(\lambda)\begin{bmatrix} 1 & a_L(\lambda)e^{-i\theta(0)/\hbar_N} \\ 0 & 1 \end{bmatrix}, \quad \text{for } \lambda \in C_L, \\
\mathbf{F}_+(\lambda) &= \mathbf{F}_-(\lambda)\begin{bmatrix} 1 & 0 \\ a_R(\lambda)e^{i\theta(0)/\hbar_N} & 1 \end{bmatrix}, \quad \text{for } \lambda \in C_R, \\
\mathbf{F}_+(\lambda) &= \mathbf{F}_-(\lambda)\begin{bmatrix} 1 & 0 \\ -ie^{-i(\theta(\lambda)-\theta(0))/\hbar_N} & 1 \end{bmatrix}, \quad \text{for } \lambda \in L_L, \\
\mathbf{F}_+(\lambda) &= \mathbf{F}_-(\lambda)\begin{bmatrix} 1 & -ie^{i(\theta(\lambda)-\theta(0))/\hbar_N} \\ 0 & 1 \end{bmatrix}, \quad \text{for } \lambda \in L_R,
\end{aligned} \tag{7.48}$$

and for $\lambda \in C_M$,

$$\begin{aligned}
&\mathbf{F}_+(\lambda) = \mathbf{F}_-(\lambda) \\
&\times \begin{bmatrix} 1 + \left[ia_M(\lambda) + e^{\tilde{\phi}(\lambda)/\hbar_N}\right] & -ie^{i(\theta(\lambda)-\theta(0))/\hbar_N}\left[ia_M(\lambda) + e^{\tilde{\phi}(\lambda)/\hbar_N}\right] \\ -ie^{-i(\theta(\lambda)-\theta(0))/\hbar_N}\left[ia_M(\lambda) + e^{\tilde{\phi}(\lambda)/\hbar_N}\right] & 1 - \left[ia_M(\lambda) + e^{\tilde{\phi}(\lambda)/\hbar_N}\right] \end{bmatrix}.
\end{aligned} \tag{7.49}$$

The jump relations satisfied by $\mathbf{F}(\lambda)$ on the complex conjugate contours in the lower half-plane follow from these by the symmetry $\mathbf{F}(\lambda) = \sigma_2\mathbf{F}(\lambda^*)^*\sigma_2$.

Now, for $\lambda \in C_L$ with $|\lambda| \le \hbar_N^\alpha$,

$$\begin{aligned}
a_L(\lambda)e^{-i\theta(0)/\hbar_N} &= i\exp\left(\frac{\tau(\lambda) - i\theta^0(\lambda) - i\theta(0)}{\hbar_N}\right)\cdot S(\lambda) \\
&= i\exp\left(\frac{\tilde{\phi}(\lambda) + i(\theta(\lambda)-\theta(0)) - 2i\theta^0(\lambda)}{\hbar_N}\right)\cdot S(\lambda) \\
&= i\exp\left(\frac{i(\theta(\lambda)-\theta(0)) - 2i(\theta^0(\lambda)-\theta^0(0))}{\hbar_N}\right)\cdot S(\lambda) \\
&= ie^{\pi\zeta}S(\lambda)\left(1 + O\left(\hbar_N^{2\alpha-1}\right)\right) \\
&= ie^{(2i+\pi)\zeta}(-i\zeta)^{-i\zeta}(i\zeta)^{-i\zeta}\frac{\Gamma\left(\frac{1}{2}+i\zeta\right)}{\Gamma\left(\frac{1}{2}-i\zeta\right)}\left(1 + O\left(\hbar_N^{4\alpha/3-1}\right)\right),
\end{aligned} \tag{7.50}$$

where in the last line $\zeta = \varphi_0(\lambda)$ with the change of coordinate being given by (7.34). In these steps, we used the relation (6.12), the fact that from Proposition 6.1 we get

$\tilde{\phi}(\lambda) \equiv 0$, and, according to (2.7) and the quantization condition (2.8) on $\hbar_N$, $2\theta^0(0)/\hbar_N = 2\pi N \in 2\pi\mathbb{Z}$. We have also used the fact that

$$\frac{1}{\rho^0(0)}\left(2\frac{d\theta^0}{d\lambda}(0) - \frac{d\theta}{d\lambda}(0)\right) = \pi \tag{7.51}$$

which follows directly from the definition (2.7) of $\theta^0(\lambda)$, the definition (6.6) of $\theta(\lambda)$, and the relation (6.2). In a similar way, for $\lambda \in C_R$ with $|\lambda| \le \hbar_N^\alpha$, we get

$$a_R(\lambda)e^{i\theta(0)/\hbar_N} = ie^{(2i-\pi)\zeta}(-i\zeta)^{-i\zeta}(i\zeta)^{-i\zeta}\frac{\Gamma\left(\frac{1}{2}+i\zeta\right)}{\Gamma\left(\frac{1}{2}-i\zeta\right)}\left(1+O\left(\hbar_N^{4\alpha/3-1}\right)\right), \tag{7.52}$$

and for $\lambda \in L_L \cup C_M \cup L_R$ with $|\lambda| \le \hbar_N^\alpha$,

$$e^{\pm i(\theta(\lambda)-\theta(0))/\hbar_N} = e^{\mp\pi\zeta}\left(1+O\left(\hbar_N^{2\alpha-1}\right)\right), \tag{7.53}$$

with $\zeta = \varphi_0(\lambda)$. Finally, when $\lambda \in C_M$ and $|\lambda| \le \hbar_N^\alpha$ we have for arbitrarily small $\nu > 0$,

$$\begin{aligned} ia_M(\lambda) + e^{\tilde{\phi}(\lambda)/\hbar_N} &= e^{\tilde{\phi}(\lambda)/\hbar_N}[1 - T(\lambda)] = 1 - T(\lambda) \\ &= 1 - \frac{2\pi e^{2i\zeta}(-i\zeta)^{-2i\zeta}}{\Gamma\left(\frac{1}{2}-i\zeta\right)^2}\left(1+O\left(\hbar_N^{4\alpha/3-1-\nu}\right)\right) \\ &= 1 - \frac{2\pi e^{2i\zeta}(-i\zeta)^{-2i\zeta}}{\Gamma\left(\frac{1}{2}-i\zeta\right)^2} + O\left(\hbar_N^{4\alpha/3-1-\nu}\right), \end{aligned} \tag{7.54}$$

again with $\zeta = \varphi_0(\lambda)$. The last step follows because $2\pi e^{2i\zeta}(-i\zeta)^{-2i\zeta}/\Gamma(1/2 - i\zeta)^2$ is uniformly bounded on C_M.

Let $\vec{C}_L$, $\vec{C}_R$, $\vec{L}_L$, $\vec{L}_R$, and $\vec{C}_M$ denote the straight rays that agree with the corresponding contours in a fixed neighborhood of the origin in the λ-plane, but lying in the ζ-plane (according to (7.34), ζ is a simple rescaling of λ by a positive number). These rays are oriented contours, with the same orientation as the original contours: $\vec{C}_M$, $\vec{L}_L$, and $\vec{L}_R$ are oriented outwards from the origin toward infinity, and $\vec{C}_L$ and $\vec{C}_R$ are oriented inwards from infinity toward the origin. Let the union of these contours with their complex conjugates be denoted Σ_0. Consider the following Riemann-Hilbert problem.

Riemann-Hilbert Problem 7.6 (local problem near the origin). Find a matrix $\widehat{F}(\zeta)$ with the following properties:

(1) Analyticity: $\widehat{F}(\zeta)$ is analytic for $\zeta \in \mathbb{C} \setminus \Sigma_0$.

(2) Boundary behavior: $\widehat{F}(\zeta)$ assumes continuous boundary values on Σ_0.

(3) Jump conditions: the boundary values taken on Σ_0 satisfy

$$\widehat{F}_+(\zeta) = \widehat{F}_-(\zeta) \begin{bmatrix} 1 & ie^{(2i+\pi)\zeta}(-i\zeta)^{-i\zeta}(i\zeta)^{-i\zeta}\dfrac{\Gamma(1/2+i\zeta)}{\Gamma(1/2-i\zeta)} \\ 0 & 1 \end{bmatrix}, \quad \zeta \in \vec{C}_L,$$

$$\widehat{F}_+(\zeta) = \widehat{F}_-(\zeta) \begin{bmatrix} 1 & 0 \\ ie^{(2i-\pi)\zeta}(-i\zeta)^{-i\zeta}(i\zeta)^{-i\zeta}\dfrac{\Gamma(1/2+i\zeta)}{\Gamma(1/2-i\zeta)} & 1 \end{bmatrix}, \quad \zeta \in \vec{C}_R,$$

$$\widehat{F}_+(\zeta) = \widehat{F}_-(\zeta) \begin{bmatrix} 1 & 0 \\ -ie^{\pi\zeta} & 1 \end{bmatrix}, \quad \zeta \in \vec{L}_L,$$

$$\widehat{F}_+(\zeta) = \widehat{F}_-(\zeta) \begin{bmatrix} 1 & -ie^{-\pi\zeta} \\ 0 & 1 \end{bmatrix}, \quad \zeta \in \vec{L}_R,$$

$$\begin{aligned}\widehat{F}_+(\zeta) &= \widehat{F}_-(\zeta) \\ &\times \begin{bmatrix} 2 - \dfrac{2\pi e^{2i\zeta}(-i\zeta)^{-2i\zeta}}{\Gamma(1/2-i\zeta)^2} & ie^{-\pi\zeta}\left[\dfrac{2\pi e^{2i\zeta}(-i\zeta)^{-2i\zeta}}{\Gamma(1/2-i\zeta)^2} - 1\right] \\ ie^{\pi\zeta}\left[\dfrac{2\pi e^{2i\zeta}(-i\zeta)^{-2i\zeta}}{\Gamma(1/2-i\zeta)^2} - 1\right] & \dfrac{2\pi e^{2i\zeta}(-i\zeta)^{-2i\zeta}}{\Gamma(1/2-i\zeta)^2} \end{bmatrix}, \quad \zeta \in \vec{C}_M.\end{aligned} \tag{7.55}$$

On the contours in the lower half-plane, the jump conditions are implied by the symmetry $\widehat{F}(\zeta^*) = \sigma_2\widehat{F}(\zeta)^*\sigma_2$.

(4) Normalization: $\widehat{F}(\zeta)$ is normalized at infinity

$$\widehat{F}(\zeta) \longrightarrow \mathbb{I} \quad \text{as } \zeta \longrightarrow \infty. \tag{7.56}$$

Unfortunately, we cannot solve Riemann-Hilbert Problem 7.6 explicitly. Luckily, we will not require an explicit solution. However, existence of a solution is an issue that must be resolved, and we need to obtain a decay estimate that quantifies the normalization condition (7.56). These questions are addressed via the abstract theory of Riemann-Hilbert problems.

Proposition 7.7. The local model Riemann-Hilbert Problem 7.6 has a unique solution satisfying $\widehat{F}(\zeta) = \mathbb{I} + O(1/\zeta)$ as $\zeta \to \infty$, uniformly with respect to direction. Also, $\det(\widehat{F}(\zeta)) \equiv 1$. □

Proof. Each Riemann-Hilbert problem is equivalent to an inhomogeneous system of linear singular integral equations. It must be shown that the matrix singular integral operator associated with these equations is of Fredholm type, with index zero. Then it must be shown that there are no homogeneous solutions, at which point one has existence and uniqueness of a solution to the inhomogeneous system. Finally, one maps the solution of the integral equations to the unique solution of the Riemann-Hilbert problem and it remains to verify the rate of decay to the identity matrix as $\zeta \to \infty$.

The theory we will use is the theory of matrix Riemann-Hilbert problems on self-intersecting contours, with boundary values taken in spaces of Hölder continuous functions. This theory is summarized in a self-contained way in the appendix of [8].

The first step is to establish that the operator of the associated system of singular integral equations is Fredholm index zero on an appropriate space of functions. As described in [8], this follows from two facts. First, on each ray of the contour Σ_0 the jump matrix $v_{\widehat{F}}(\zeta) := \widehat{F}_-(\zeta)^{-1}\widehat{F}_+(\zeta)$ is uniformly Lipschitz with respect to ζ, and differs from the identity by a quantity that is $O(1/\zeta)$ for large ζ. Second, the limiting values of the jump matrix, taken as $\zeta \to 0$ along each ray of Σ_0, are consistent with a bounded solution $\widehat{F}(\zeta)$ near $\zeta = 0$. This means the following. Suppose that $\widehat{F}(\zeta)$ has a limiting value, say a matrix $\widehat{F}_0$, as $\zeta \to 0$ inside one of the sectors of $\mathbb{C} \setminus \Sigma_0$. Using the limiting value of the jump matrix $v_{\widehat{F}}(\zeta)$ at the origin along one of the rays of Σ_0 bounding that sector, one can compute the limiting value of $\widehat{F}(\zeta)$ at the origin in the neighboring sector. This procedure can be continued, moving from sector to sector of $\mathbb{C} \setminus \Sigma_0$ in the same direction, until one arrives once again in the original sector, with a matrix $\widehat{F}_1$. The consistency condition is simply that $\widehat{F}_1 = \widehat{F}_0$, which upon elimination of $\widehat{F}_0$ can be viewed as a cyclic relation among the limiting values of the jump matrix $v_{\widehat{F}}(\zeta)$ taken along each ray of Σ_0 as $\zeta \to 0$. It is easily checked that this cyclic relation indeed holds for Riemann-Hilbert Problem 7.6.

The second step is to establish existence and uniqueness of the solution $\widehat{F}(\zeta)$. The fact that the associated singular integral equations are Fredholm index zero means that, in a certain precise sense, the Fredholm alternative applies to our Riemann-Hilbert problem. The inhomogeneity is the normalization to the identity matrix at $\zeta = \infty$. The corresponding homogeneous Riemann-Hilbert problem has exactly the same form except that the normalization condition is replaced by the condition $\widehat{F}(\zeta) \to 0$ as $\zeta \to \infty$. We will have a unique solution of Riemann-Hilbert Problem 7.6 if it can be shown that no such homogeneous solutions exist. For this purpose, it is sufficient that the jump matrix $v_{\widehat{F}}(\zeta)$ should have a certain symmetry with respect to Schwartz reflection through the real axis in the ζ-plane. For the orientation of Σ_0 described above, the required relation is

$$v_{\widehat{F}}(\zeta^*)^{-1} = v_{\widehat{F}}(\zeta)^\dagger \tag{7.57}$$

for $\zeta \in \Sigma_0 \cap \mathbb{C}_+$. It is easily checked, using the symmetry $v_{\widehat{F}}(\zeta^*) = \sigma_2 v_{\widehat{F}}(\zeta)^* \sigma_2$ and structural details of $v_{\widehat{F}}(\zeta)$ in the upper half-plane, that this relation holds, and this means that Riemann-Hilbert Problem 7.6 has a unique solution $\widehat{F}(\zeta)$.

The third step is to establish that the unique solution $\widehat{F}(\zeta)$ decays to the identity for large ζ like $1/\zeta$. The Hölder theory that we have been using generally provides a solution $\widehat{F}(\zeta)$ under these circumstances that takes boundary values on Σ_0 that are Hölder continuous with exponent μ and that differs from the identity matrix by $O(1/\zeta^\mu)$ as $\zeta \to \infty$, for all μ strictly less than 1. This fact can be traced to the compact embedding of each Hölder space into all Hölder spaces with strictly smaller exponents. The compactness is needed to establish the Fredholm property of the Riemann-Hilbert problem. So to obtain the required decay, we need an additional argument. The condition that is required to obtain the $O(1/\zeta)$ decay is that a signed sum of the mean values of $\zeta \cdot (v_{\widehat{F}}(\zeta) - \mathbb{I})$ taken as $\zeta \to \infty$ along each ray of Σ_0 (the signs are related to the orientation of the individual rays) is zero [8]. Now along each ray of Σ_0 except for $\vec{C}_M$ and its conjugate (i.e., the imaginary axis in the ζ-plane), $v_{\widehat{F}}(\zeta)$ decays to the identity exponentially fast as $\zeta \to \infty$. So these rays do not contribute to the sum and it is only necessary to check the imaginary axis. When $\zeta \in \vec{C}_M$,

$$\zeta \cdot \left(v_{\widehat{F}}(\zeta) - \mathbb{I}\right) = \frac{i}{12}\begin{bmatrix} -1 & ie^{-\pi\zeta} \\ ie^{\pi\zeta} & 1 \end{bmatrix} + O\left(\frac{1}{\zeta}\right), \tag{7.58}$$

as $\zeta \to \infty$, and for ζ on the negative imaginary axis oriented upwards,

$$\zeta \cdot \left(v_{\widehat{F}}(\zeta) - \mathbb{I}\right) = \frac{i}{12}\begin{bmatrix} -1 & -ie^{\pi\zeta} \\ -ie^{-\pi\zeta} & 1 \end{bmatrix} + O\left(\frac{1}{\zeta}\right), \tag{7.59}$$

as $\zeta \to \infty$. The limits of these quantities do not exist as $\zeta \to \infty$ due to the oscillations on the off-diagonal. But the mean values exist and are equal, and it turns out that they enter the sum with opposite signs due to the orientation of the contour rays. Thus, the required sum of signed mean values indeed vanishes. This, along with the analyticity of the jump matrix $v_{\widehat{F}}(\zeta)$ along each ray of Σ_0 establishes that $\widehat{F}(\zeta) - \mathbb{I} = O(1/\zeta)$ as $\zeta \to \infty$.

Finally, we check that $\det(\widehat{F}(\zeta)) = 1$. Taking determinants in the jump relations we see that on all rays of the contour, $\det(\widehat{F}_+(\zeta)) = \det(\widehat{F}_-(\zeta))$. Since the boundary values taken by $\widehat{F}(\zeta)$ on Σ_0 are continuous, we discover that $\det(\widehat{F}(\zeta))$ is an entire function. Since this function tends to one at infinity, it follows from Liouville's theorem that $\det(\widehat{F}(\zeta)) \equiv 1$. This completes the proof of the proposition. ■

7.1.3 The local model for $\mathbf{N}(\lambda)$ *near* $\lambda = 0$. From (7.46) and (7.47) we can express $\mathbf{N}(\lambda)$ in terms of $\mathbf{F}(\lambda)$ for $|\lambda| \leq \hbar_N^\alpha$. For $\Re(\lambda) < 0$, we have

$$\begin{aligned} \mathbf{N}(\lambda) = \tilde{\mathbf{O}}(\lambda)(i\sigma_1)\sigma_1^{(1-J)/2} e^{i\theta(0)\sigma_3/(2\hbar_N)} \\ \times \mathbf{F}(\lambda)e^{-i\theta(0)\sigma_3/(2\hbar_N)}\sigma_1^{(1-J)/2}(-i\sigma_1)\tilde{\mathbf{O}}(\lambda)^{-1}\widehat{\mathbf{N}}_{\text{out}}(\lambda) \end{aligned} \tag{7.60}$$

and for $\Re(\lambda) > 0$, we have

$$\mathbf{N}(\lambda) = \tilde{\mathbf{O}}(\lambda)\sigma_1^{(1-J)/2} e^{i\theta(0)\sigma_3/(2\hbar_N)}\mathbf{F}(\lambda)e^{-i\theta(0)\sigma_3/(2\hbar_N)}\sigma_1^{(1-J)/2}\tilde{\mathbf{O}}(\lambda)^{-1}\widehat{\mathbf{N}}_{\text{out}}(\lambda). \tag{7.61}$$

To obtain a local model for $\mathbf{N}(\lambda)$ near the origin, we simply replace $\mathbf{F}(\lambda)$ in these formulae by the approximation $\widehat{\mathbf{F}}(\varphi_0(\lambda))$. With $|\lambda| \leq \hbar_N^\alpha$, we set for $\Re(\lambda) < 0$,

$$\begin{aligned} \widehat{\mathbf{N}}_{\text{origin}}(\lambda) := \tilde{\mathbf{O}}(\lambda)(i\sigma_1)\sigma_1^{(1-J)/2} e^{i\theta(0)\sigma_3/(2\hbar_N)} \\ \times \widehat{\mathbf{F}}(\varphi_0(\lambda))e^{-i\theta(0)\sigma_3/(2\hbar_N)}\sigma_1^{(1-J)/2}(-i\sigma_1)\tilde{\mathbf{O}}(\lambda)^{-1}\widehat{\mathbf{N}}_{\text{out}}(\lambda) \end{aligned} \tag{7.62}$$

and for $\Re(\lambda) > 0$, we set

$$\begin{aligned} \widehat{\mathbf{N}}_{\text{origin}}(\lambda) := \tilde{\mathbf{O}}(\lambda)\sigma_1^{(1-J)/2} e^{i\theta(0)\sigma_3/(2\hbar_N)} \\ \times \widehat{\mathbf{F}}(\varphi_0(\lambda))e^{-i\theta(0)\sigma_3/(2\hbar_N)}\sigma_1^{(1-J)/2}\tilde{\mathbf{O}}(\lambda)^{-1}\widehat{\mathbf{N}}_{\text{out}}(\lambda). \end{aligned} \tag{7.63}$$

The most important properties of this matrix function are easily seen to be the following.

Proposition 7.8. The matrix $\widehat{\mathbf{N}}_{\text{origin}}(\lambda)$ is a piecewise analytic function of λ in the disk $|\lambda| < \hbar_N^\alpha$, with jumps only on the locally straight-line contours C_L, C_R, L_L, L_R, and C_M, and their conjugates in the lower half-disk. The jump relations satisfied by $\widehat{\mathbf{N}}_{\text{origin}}(\lambda)$ on these contours are the following:

$$\begin{aligned}
&\widehat{\mathbf{N}}_{\text{origin},-}(\lambda)^{-1}\widehat{\mathbf{N}}_{\text{origin},+}(\lambda)\\
&\quad=\sigma_1^{(1-J)/2}\begin{bmatrix}1 & 0\\ ie^{i\theta(0)/\hbar_N}\,e^{(2i+\pi)\zeta}(-i\zeta)^{-i\zeta}(i\zeta)^{-i\zeta}\dfrac{\Gamma(1/2+i\zeta)}{\Gamma(1/2-i\zeta)} & 1\end{bmatrix}\sigma_1^{(1-J)/2},\quad \lambda\in C_L,\\
&\widehat{\mathbf{N}}_{\text{origin},-}(\lambda)^{-1}\widehat{\mathbf{N}}_{\text{origin},+}(\lambda)\\
&\quad=\sigma_1^{(1-J)/2}\begin{bmatrix}1 & 0\\ ie^{-i\theta(0)/\hbar_N}\,e^{(2i-\pi)\zeta}(-i\zeta)^{-i\zeta}(i\zeta)^{-i\zeta}\dfrac{\Gamma(1/2+i\zeta)}{\Gamma(1/2-i\zeta)} & 1\end{bmatrix}\sigma_1^{(1-J)/2},\quad \lambda\in C_R,\\
&\widehat{\mathbf{N}}_{\text{origin},-}(\lambda)^{-1}\widehat{\mathbf{N}}_{\text{origin},+}(\lambda)\\
&\quad=\sigma_1^{(1-J)/2}\begin{bmatrix}1 & i(e^{-i\theta(\lambda)/\hbar_N}-e^{\pi\zeta}e^{-i\theta(0)/\hbar_N})\\ 0 & 1\end{bmatrix}\sigma_1^{(1-J)/2},\quad \lambda\in L_L,\\
&\widehat{\mathbf{N}}_{\text{origin},-}(\lambda)^{-1}\widehat{\mathbf{N}}_{\text{origin},+}(\lambda)\\
&\quad=\sigma_1^{(1-J)/2}\begin{bmatrix}1 & i(e^{i\theta(\lambda)/\hbar_N}-e^{-\pi\zeta}e^{i\theta(0)/\hbar_N})\\ 0 & 1\end{bmatrix}\sigma_1^{(1-J)/2},\quad \lambda\in L_R,\\
&\widehat{\mathbf{N}}_{\text{origin},-}(\lambda)^{-1}\widehat{\mathbf{N}}_{\text{origin},+}(\lambda)=\sigma_1^{(1-J)/2}\boldsymbol{v}(\lambda)\sigma_1^{(1-J)/2},\quad \lambda\in C_M,
\end{aligned}\tag{7.64}$$

where

$$\begin{aligned}
v_{11}&:=e^{i\theta(\lambda)/\hbar_N}\left(1+\left(1-e^{-\pi\zeta-i(\theta(\lambda)-\theta(0))/\hbar_N}\right)Z\right),\\
v_{12}&:=iZ\left(e^{\pi\zeta+i(\theta(\lambda)-\theta(0))/\hbar_N}+e^{-\pi\zeta-i(\theta(\lambda)-\theta(0))/\hbar_N}-2\right),\\
v_{21}&:=i+iZ,\\
v_{22}&:=e^{-i\theta(\lambda)/\hbar_N}\left(1+\left(1-e^{\pi\zeta+i(\theta(\lambda)-\theta(0))/\hbar_N}\right)Z\right),
\end{aligned}\tag{7.65}$$

with

$$Z:=\frac{2\pi e^{2i\zeta}(-i\zeta)^{-2i\zeta}}{\Gamma\left(\dfrac{1}{2}-i\zeta\right)^2}-1,\tag{7.66}$$

and where $\zeta=\varphi_0(\lambda)$. The jumps on the corresponding contours in the lower half-plane are obtained from the symmetry $\widehat{\mathbf{N}}_{\text{origin}}(\lambda^*)=\sigma_2\widehat{\mathbf{N}}_{\text{origin}}(\lambda)^*\sigma_2$. The matrix $\widehat{\mathbf{N}}_{\text{origin}}(\lambda)$ is uniformly bounded for $|\lambda|<\hbar_N^\alpha$, with a bound that is independent of $\hbar_N$. Also, $\det(\widehat{\mathbf{N}}_{\text{origin}}(\lambda))\equiv 1$ and when $|\lambda|=\hbar_N^\alpha$,

$$\widehat{\mathbf{N}}_{\text{origin}}(\lambda)\widehat{\mathbf{N}}_{\text{out}}(\lambda)^{-1}=\mathbb{I}+O\left(\hbar_N^{1-\alpha}\right).\tag{7.67}$$

□

7.2 Local analysis near $\lambda = iA$

7.2.1 Local behavior of $a_L(\lambda)$, $a_R(\lambda)$, *and* $a_M(\lambda)$. As before, we suppose that ϵ and δ are small scales satisfying $\hbar_N \ll \delta \ll \epsilon \ll 1$ as $\hbar_N$ tends to zero. We redefine the integer L so that exactly the first L of the numbers $\lambda_0, \ldots, \lambda_{N-1}$ lie on the positive imaginary axis above $i(A-\epsilon)$. We will suppose that $\Im(\lambda) \leq A$, and $|\lambda - iA| \leq \delta$ and we will deduce asymptotic formulae for $T(\lambda)$ given by (5.3) valid for such λ, and for $S(\lambda)$ given by (5.2) when λ is also bounded outside of some downward-opening sector with vertex at iA, in the semiclassical limit $\hbar_N \to 0$. First, we establish a result that is the analogue of Lemma 7.1.

Lemma 7.9. When $\Im(\lambda) \leq A$ and $|\lambda - iA| \leq \delta$ and with L defined as indicated in the preceding paragraph,

$$\exp\left(-\sum_{k=L}^{N-1} \tilde{I}_k(\lambda)\right) = 1 + O\left(\frac{\hbar_N}{\epsilon}\right). \tag{7.68}$$

□

Proof. We again estimate $\tilde{I}_k(\lambda)$ using the integral formula (7.2) with integrand $g(\lambda, \xi)$ given by (7.3). Given our conditions on λ, for $m_k - \hbar_N/2 \leq \xi \leq m_k + \hbar_N/2$ and $k \geq L$ we have

$$\begin{aligned} \frac{1}{|\lambda + e(\xi)|} \leq \frac{1}{|\lambda - e(\xi)|} \leq \frac{1}{|i(A-\delta) - e(\xi)|} &\leq \frac{1}{\left|i(A-\delta) - e\left(m_k + \dfrac{\hbar_N}{2}\right)\right|} \\ &= O\left(\frac{1}{\left|m(iA - i\delta) - m_k - \dfrac{\hbar_N}{2}\right|}\right). \end{aligned} \tag{7.69}$$

For all such ξ we therefore have the estimate

$$g(\lambda, \xi) = O\left(\frac{1}{\left|m(iA - i\delta) - m_k - \dfrac{\hbar_N}{2}\right|}\right). \tag{7.70}$$

Summing over k gives

$$\begin{aligned} \sum_{k=L}^{N-1} \tilde{I}_k(\lambda) &= O\left(\hbar_N^2 \sum_{k=L}^{N-1} \frac{1}{\left|m(iA - i\delta) - m_k - \dfrac{\hbar_N}{2}\right|^2}\right) \\ &= O\left(\hbar_N \int_0^{m(iA-\epsilon)} \frac{dm}{(m(iA - i\delta) - m)^2}\right) \end{aligned} \tag{7.71}$$

which is $O(\hbar_N/\epsilon)$ because $\delta \ll \epsilon$, and the lemma is proved. ■

As was the case when λ was near the origin, only certain terms are important when $\lambda - iA$ is small, as a direct consequence of Lemma 7.9 and the fact that $\exp(-\tilde{I}_k(\lambda))$ is analytic for such λ when $k \geq L$. The important terms when $|\lambda - iA| \leq \delta$ are

$$S_1^{(iA)}(\lambda) := \left(\prod_{k=0}^{L-1} \frac{\lambda - \lambda_k^*}{\lambda - \lambda_k}\right) \exp\left(\frac{1}{\hbar_N}\int_{m_L+\hbar_N/2}^{M} \left(L^0_{e(m)}(\lambda) - L^0_{-e(m)}(\lambda)\right) dm\right),$$
$$T_1^{(iA)}(\lambda) := \left(\prod_{k=0}^{L-1} \frac{\lambda - \lambda_k^*}{\lambda - \lambda_k}\right) \exp\left(\frac{1}{\hbar_N}\int_{m_L+\hbar_N/2}^{M} \left(\overline{L}^0_{e(m)}(\lambda) - \overline{L}^0_{-e(m)}(\lambda)\right) dm\right) \quad (7.72)$$
$$\times 2\cos\left(\frac{\theta^0(\lambda)}{\hbar_N}\right).$$

Lemma 7.9 states that $S(\lambda) = S_1^{(iA)}(\lambda)(1 + O(\hbar_N/\epsilon))$ and $T(\lambda) = T_1^{(iA)}(\lambda)(1 + O(\hbar_N/\epsilon))$ as $\hbar_N \to 0$ for $|\lambda - iA| \leq \delta$.

The analogue of Lemma 7.2 says that for $\lambda - iA$ small the sequence of numbers $\lambda_0, \ldots, \lambda_{L-1}$ contributing to $S(\lambda)$ and $T(\lambda)$ can be replaced essentially by a "straightened-out" sequence with uniform density.

Lemma 7.10. Let $\tilde{\lambda}_k$ for $k = 0, 1, 2, \ldots$ be the sequence of numbers defined by the relation

$$\tilde{\lambda}_k := iA + \frac{\hbar_N}{\rho^0(iA)}\left(k + \frac{1}{2}\right), \quad (7.73)$$

which results from expanding the Bohr-Sommerfeld relation (2.9) for λ_k near iA and keeping only the dominant terms. Define

$$S_2^{(iA)}(\lambda) := \left(\prod_{k=0}^{L-1} \frac{\lambda - \tilde{\lambda}_k^*}{\lambda - \tilde{\lambda}_k}\right)$$
$$\times \exp\left(\frac{1}{\hbar_N}\int_{m_L+\hbar_N/2}^{M} \left(L^0_{iA+e'(M)(m-M)}(\lambda) - L^0_{-iA-e'(M)(m-M)}(\lambda)\right) dm\right),$$
$$T_2^{(iA)}(\lambda) := \left(\prod_{k=0}^{L-1} \frac{\lambda - \tilde{\lambda}_k^*}{\lambda - \tilde{\lambda}_k}\right)$$
$$\times \exp\left(\frac{1}{\hbar_N}\int_{m_L+\hbar_N/2}^{M} \left(\overline{L}^0_{iA+e'(M)(m-M)}(\lambda) - \overline{L}^0_{-iA-e'(M)(m-M)}(\lambda)\right) dm\right)$$
$$\times 2\cos\left(\frac{\pi\rho^0(iA)}{\hbar_N}(iA - \lambda)\right). \quad (7.74)$$

Then, for $\Im(\lambda) \le A$ and $|\lambda - iA| \le \delta$,

$$T_1^{(iA)}(\lambda) = T_2^{(iA)}(\lambda)\left(1 + O\left(\frac{\epsilon^2}{\hbar_N}\log\left(\frac{\epsilon}{\hbar_N}\right)\right)\right), \tag{7.75}$$

where we suppose that the scale ϵ is further constrained so that the relative error is asymptotically small. If λ is additionally bounded outside of some downward opening sector with vertex at iA, then

$$S_1^{(iA)}(\lambda) = S_2^{(iA)}(\lambda)\left(1 + O\left(\frac{\epsilon^2}{\hbar_N}\right)\right). \tag{7.76}$$

□

Proof. The proof of this lemma follows that of Lemma 7.2 almost exactly and will not be repeated here. The only difference is that the square in that proof should be replaced here by the rectangle whose top side is $\Im(\lambda) = A$ and $-\epsilon \le \Re(\lambda) \le \epsilon$ and whose bottom is $\Im(\lambda) = -i(\tilde{\lambda}_{L-1} + \tilde{\lambda}_L)/2$. ■

Without any approximation, $S_2^{(iA)}(\lambda)$ and $T_2^{(iA)}(\lambda)$ can be rewritten in a more transparent form by expressing the products in terms of gamma functions and evaluating the logarithmic integrals exactly. Introduce a local variable ζ in terms of a transformation φ_{iA} given by the relation

$$\zeta = \varphi_{iA}(\lambda) := \rho^0(iA)\frac{\lambda - iA}{i\hbar_N}, \tag{7.77}$$

and let B be the positive constant

$$B := -\frac{2iA\rho^0(iA)}{\hbar_N}. \tag{7.78}$$

In terms of these quantities, one finds that $S_2^{(iA)}(\lambda)$ and $T_2^{(iA)}(\lambda)$ take a simple form

$$\begin{aligned} S_2^{(iA)}(\lambda) &= (-i\zeta)^{i\zeta}\Gamma\left(\frac{1}{2} - i\zeta\right) \cdot V(\zeta, B, L) \cdot W(\zeta, B, L), \\ T_2^{(iA)}(\lambda) &= \frac{2\pi(i\zeta)^{i\zeta}}{\Gamma\left(\dfrac{1}{2} + i\zeta\right)} \cdot V(\zeta, B, L) \cdot W(\zeta, B, L), \end{aligned} \tag{7.79}$$

where

$$
\begin{aligned}
V(\zeta, B, L) &:= \frac{\Gamma\left(B - i\zeta + \frac{1}{2}\right)}{\Gamma\left(B - L - i\zeta + \frac{1}{2}\right)\Gamma\left(L - i\zeta + \frac{1}{2}\right)}, \\
W(\zeta, B, L) &:= \frac{(B - L - i\zeta)^{B-L-i\zeta}(L - i\zeta)^{L-i\zeta}}{(B - i\zeta)^{B-i\zeta}}.
\end{aligned} \tag{7.80}
$$

Now, $B \gg L \gg |\zeta|$ because B is proportional to $\hbar_N^{-1}$ and L is of the order of $\epsilon/\hbar$ while $|\zeta| = O(\delta/\hbar_N)$. So again we can use Stirling's formula to extract the dominant asymptotic contributions to $S_2^{(iA)}(\lambda)$ and $T_2^{(iA)}(\lambda)$ as $\hbar_N$ tends to zero.

Lemma 7.11. As $\hbar_N$ tends to zero through positive values,

$$
\begin{aligned}
S_2^{(iA)}(\lambda) &= \frac{1}{\sqrt{2\pi}} e^{-i\zeta}(-i\zeta)^{-i\zeta}\Gamma\left(\frac{1}{2} - i\zeta\right) \cdot \left(1 + O\left(\frac{\delta^2}{\epsilon\hbar_N}\right)\right), \\
T_2^{(iA)}(\lambda) &= \frac{\sqrt{2\pi} e^{-i\zeta}(i\zeta)^{i\zeta}}{\Gamma\left(\frac{1}{2} + i\zeta\right)} \cdot \left(1 + O\left(\frac{\delta^2}{\epsilon\hbar_N}\right)\right).
\end{aligned} \tag{7.81}
$$

□

Proof. Using Stirling's formula, one expands $V(\zeta, B, L)$ to find

$$
V(\zeta, B, L) = \frac{e^{1/2 - i\zeta}}{\sqrt{2\pi}} \frac{\left(B - i\zeta + \frac{1}{2}\right)^{B-i\zeta}}{\left(B - L - i\zeta + \frac{1}{2}\right)^{B-L-i\zeta}\left(L - i\zeta + \frac{1}{2}\right)^{L-i\zeta}} \left(1 + O\left(\frac{\hbar_N}{\epsilon}\right)\right). \tag{7.82}
$$

The error here is dominated by the fact that $L \sim \epsilon/\hbar_N$ is the smallest large number involved. Now we expand the powers that remain in conjunction with those in $W(\zeta, B, L)$ to find

$$
V(\zeta, B, L)W(\zeta, B, L) = \frac{e^{-i\zeta}}{\sqrt{2\pi}} \left(1 + O\left(\frac{\hbar_N}{\epsilon}\right)\right)\left(1 + O\left(\frac{\delta^2}{\epsilon\hbar_N}\right)\right). \tag{7.83}
$$

Since $\delta \ll \hbar_N$, the relative error is dominated by $O(\delta^2/\epsilon\hbar_N)$. Using this expression for the product VW in (7.79), the lemma is proved. ■

In exactly the same way as in our study of the local behavior near the origin, we may combine Lemmas 7.9, 7.10, and 7.11 and choose the "internal" scale ϵ in terms of δ and $\hbar_N$ to obtain asymptotics for $S(\lambda)$ and $T(\lambda)$ with optimized relative error.

Proposition 7.12. Let $\Im(\lambda) \leq A$, with $|\lambda - iA| \leq \hbar_N^{\alpha}$, where $3/4 < \alpha < 1$, and let λ be bounded outside some fixed symmetrical sector with vertex at iA and opening downward. Then

$$S(\lambda) = \frac{1}{\sqrt{2\pi}} e^{-i\zeta}(-i\zeta)^{-i\zeta}\Gamma\left(\frac{1}{2} - i\zeta\right)\left(1 + O\left(\hbar_N^{4\alpha/3-1}\right)\right), \tag{7.84}$$

where $\zeta = \varphi_{iA}(\lambda)$. □

Proposition 7.13. Let $\Im(\lambda) \leq A$, with $|\lambda - iA| \leq \hbar_N^{\alpha}$, where $3/4 < \alpha < 1$. Then for all $\nu > 0$, however small,

$$T(\lambda) = \frac{\sqrt{2\pi} e^{-i\zeta}(i\zeta)^{i\zeta}}{\Gamma\left(\frac{1}{2} + i\zeta\right)}\left(1 + O\left(\hbar_N^{4\alpha/3-1-\nu}\right)\right), \tag{7.85}$$

where $\zeta = \varphi_{iA}(\lambda)$. □

7.2.2 Why a local model near $\lambda = iA$ is not necessary. In particular, it follows from these considerations that both $S(\lambda)$ and $T(\lambda)$ are uniformly bounded functions on their respective contours in any fixed neighborhood U of $\lambda = iA$. We claim that the quotient of the jump matrices for $\mathbf{N}(\lambda)$ and $\widehat{\mathbf{N}}_{\text{out}}(\lambda)$ is uniformly close to the identity matrix in U as $\hbar_N \to 0$. Since $\widehat{\mathbf{N}}_{\text{out}}(\lambda)$ is, by definition, analytic throughout U, it suffices to show that the jump matrix $\mathbf{v}_N(\lambda) := \mathbf{N}_-(\lambda)^{-1}\mathbf{N}_+(\lambda)$ is uniformly close to the identity in U. This will be the case if $a_L(\lambda)$, $a_R(\lambda)$, and $a_M(\lambda)$ are uniformly small on their respective contours. Now for $\lambda \in C_M$,

$$a_M(\lambda) = i\exp\left(\frac{\tilde{\phi}(\lambda)}{\hbar_N}\right)T(\lambda), \tag{7.86}$$

so, with $\tilde{\phi}(\lambda)$ being real and strictly negative for $\lambda \in C_M \cap U$ according to Proposition 6.2 and $T(\lambda)$ being bounded, we see that $a_M(\lambda)$ is in fact exponentially small as $\hbar_N$ tends to zero through positive values. Similarly, for $\lambda \in C_L$,

$$a_L(\lambda) = i\exp\left(\frac{\tau(\lambda) - i\theta^0(\lambda)}{\hbar_N}\right)S(\lambda) = i\exp\left(\frac{\tilde{\phi}(\lambda) - 2i\theta^0(\lambda)}{\hbar_N}\right)S(\lambda), \tag{7.87}$$

where we have used (6.12) and the fact that $\theta(\lambda) \equiv 0$ on C_M above $\lambda = iA(x)$. Since $\theta^0(iA) = 0$, it is possible to choose the neighborhood U small enough (independent of $\hbar_N$) so that $\Re(\tilde{\phi}(\lambda) - 2i\theta^0(\lambda)) < 0$ throughout U. Since $S(\lambda)$ is bounded, it then follows that for $\lambda \in C_L \cap U$, $a_L(\lambda)$ is exponentially small as $\hbar_N$ tends to zero through positive values. Virtually the same argument using (6.13) in place of (6.12) shows that $a_R(\lambda)$ is also exponentially small for $\lambda \in C_R \cap U$ (it may be necessary to make U slightly smaller to have $\Re(\tilde{\phi}(\lambda) + 2i\theta^0(\lambda)) < 0$ throughout U). It follows that $\mathbf{v}_N(\lambda) - \mathbb{I}$ is exponentially small uniformly for the contours within U.

For this reason we expect that the outer model $\widehat{\mathbf{N}}_{\text{out}}(\lambda)$ will be a good approximation to $\mathbf{N}(\lambda)$ near $\lambda = iA$ even though $S(\lambda) - 1$ and $T(\lambda) - 1$ are not small. We do not need to construct a special-purpose local model for $\mathbf{N}(\lambda)$ in this case.

7.3 Local analysis near $\lambda = iA(x)$

7.3.1 A model Riemann-Hilbert problem and its explicit solution in terms of classical special functions. Let D be a circular disk centered at $\lambda = iA$ of sufficiently small radius (independent of $\hbar_N$) that $0 \notin D$ and $iA \notin D$, and that L_L and L_R each have exactly one intersection with ∂D (of course C_M will have two intersection points with ∂D). This situation is possible as long as $x \neq 0$. The case of $x = 0$ is a degenerate case that we will not treat in detail here.

Since $x \neq 0$ and therefore $A(x) < A$, it follows from the definition (6.1) of $\rho(\eta)$ that as λ tends to $iA(x)$ along I, $\theta(\lambda)$ vanishes like $(\lambda - iA(x))^{3/2}$, and not to higher order. Since $\theta(\lambda)$ may be extended from I to be an analytic function in D except for a branch cut along the part of C_M in D lying above the center, and since $\rho(\lambda)$ extended to this cut domain from I is nonzero, it is easy to see that the function

$$\varphi_{iA(x)}(\lambda) := \left(\frac{\theta(\lambda)}{\hbar_N}\right)^{2/3} \tag{7.88}$$

defines an invertible conformal mapping of all of D to its image. Consider the local variable ζ defined by the relation $\zeta = \varphi_{iA(x)}(\lambda)$. The image of D in the ζ-plane is a neighborhood of $\zeta = 0$ that scales with $\hbar_N$ such that it contains the disk centered at $\zeta = 0$ with radius $C\hbar_N^{-2/3}$ for some constant $C > 0$. The transformation (7.88) maps $I \cap D$ to a ray segment of the positive real ζ-axis, and takes the portion of C_M in D lying above $\lambda = iA(x)$ to a ray segment of the negative real ζ-axis. We suppose that the contours L_L and L_R have been chosen so that $\varphi_{iA(x)}(L_L \cap D)$ and $\varphi_{iA(x)}(L_R \cap D)$ are straight ray segments with angles $-\pi/3$ and $\pi/3$, respectively.

For $\zeta \in \varphi_{iA(x)}(D)$, the matrix $\mathbf{S}(\zeta) := \sigma_1^{(1-J)/2}\mathbf{O}(\varphi_{iA(x)}^{-1}(\zeta))\sigma_1^{(1-J)/2}$ satisfies the following jump relations:

$$\mathbf{S}_+(\zeta) = \mathbf{S}_-(\zeta)\begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} \tag{7.89}$$

for $\zeta \in \mathbb{R}_+ \cap \varphi_{iA(x)}(D)$, oriented from right to left,

$$\mathbf{S}_+(\zeta) = \mathbf{S}_-(\zeta) \begin{bmatrix} 1 & -ie^{-i\zeta^{3/2}} \\ 0 & 1 \end{bmatrix} \tag{7.90}$$

on the part of the ray $\arg(\zeta) = -\pi/3$ in $\varphi_{iA(x)}(D)$, oriented toward the origin,

$$\mathbf{S}_+(\zeta) = \mathbf{S}_-(\zeta) \begin{bmatrix} 1 & -ie^{i\zeta^{3/2}} \\ 0 & 1 \end{bmatrix} \tag{7.91}$$

on the part of the ray $\arg(\zeta) = \pi/3$ in $\varphi_{iA(x)}(D)$, oriented toward the origin, and finally

$$\mathbf{S}_+(\zeta) = \mathbf{S}_-(\zeta) \begin{bmatrix} 1 & 0 \\ ie^{-(-\zeta)^{3/2}} & 1 \end{bmatrix} \tag{7.92}$$

for $\zeta \in \mathbb{R}_- \cap \varphi_{iA(x)}(D)$, oriented from right to left. The jump relation on the negative real ζ-axis follows from the formula (5.32) which applies because $\mathbf{O}(\lambda) = \tilde{\mathbf{N}}(\lambda)$ here. In (5.32) one uses the fact that $\theta(\lambda) \equiv 0$ for $\lambda \in C_M$ above $iA(x)$, and the relation (6.9) giving $\tilde{\phi}(\lambda)$ above $iA(x)$ in terms of the analytic continuation of $\theta(\lambda)$ from I, which one writes in terms of the local coordinate ζ.

As $\zeta \to \infty$ on all of the rays except for $\mathbb{R}_+$, the jump matrix for $\mathbf{S}(\zeta)$ decays exponentially to the identity matrix. These jump conditions were precisely the ones that were neglected in obtaining the outer model. That is, the matrix $\tilde{\mathbf{S}}(\zeta) := \sigma_1^{(1-J)/2}\tilde{\mathbf{O}}(\varphi_{iA(x)}^{-1}(\zeta))\sigma_1^{(1-J)/2}$ defined for $\zeta \in \varphi_{iA(x)}(D)$ is analytic except on the positive real ζ-axis, where it satisfies

$$\tilde{\mathbf{S}}_+(\zeta) = \tilde{\mathbf{S}}_-(\zeta) \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}. \tag{7.93}$$

It follows that the matrix $\tilde{\mathbf{S}}(\zeta)$ can be decomposed into a product of a holomorphic prefactor depending on $\hbar_N$ and a universal (i.e., independent of $\hbar_N$) local factor that takes care of the jump. We therefore may write

$$\tilde{\mathbf{S}}(\zeta) = \tilde{\mathbf{S}}^{\mathrm{hol}}(\zeta)\tilde{\mathbf{S}}^{\mathrm{loc}}(\zeta), \tag{7.94}$$

where $\tilde{\mathbf{S}}^{\mathrm{hol}}(\zeta)$ is holomorphic in $\varphi_{iA(x)}(D)$ and where

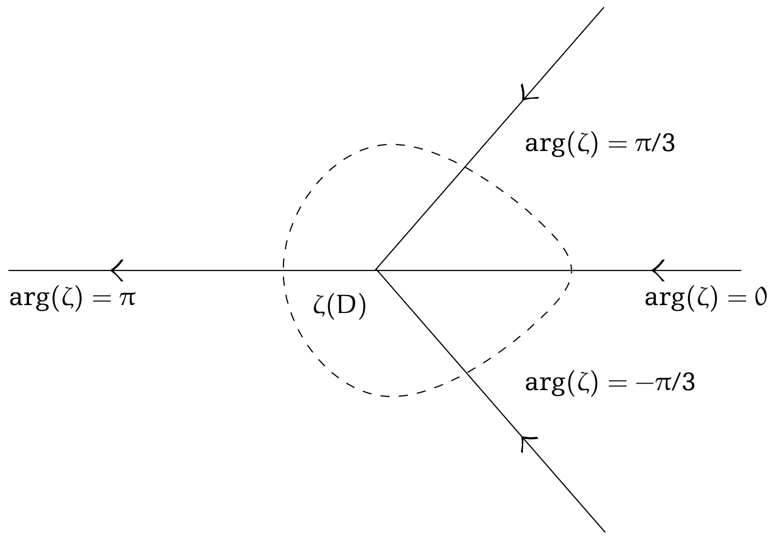

Figure 7.1 The contour $\Sigma_{iA(x)}$ in the ζ-plane for Riemann-Hilbert Problem 7.14. The boundary of the image $\varphi_{iA(x)}(D)$, expanding as $\hbar_N \to 0$, is shown as a dashed curve.

$$\tilde{S}^{\mathrm{loc}}(\zeta) := \frac{1}{\sqrt{2}}(-\zeta)^{\sigma_3/4}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} (-\zeta)^{1/4} & (-\zeta)^{1/4} \\ -(-\zeta)^{-1/4} & (-\zeta)^{-1/4} \end{bmatrix}. \tag{7.95}$$

Note that $\tilde{S}^{\mathrm{hol}}(\zeta)$ has determinant one. Its matrix elements are of size $O(\hbar_N^{-1/6})$ for $\zeta \in \varphi_{iA(x)}(D)$. It is easy to write down an explicit formula for $\tilde{S}^{\mathrm{hol}}(\zeta)$ because both $\tilde{S}(\zeta)$ and $\tilde{S}^{\mathrm{loc}}(\zeta)$ are known.

We will now approximate $S(\zeta)$ by

$$\widehat{S}(\zeta) := \tilde{S}^{\mathrm{hol}}(\zeta)S^{\mathrm{loc}}(\zeta), \tag{7.96}$$

where $S^{\mathrm{loc}}(\zeta)$ is the solution of the following Riemann-Hilbert problem. Let $\Sigma_{iA(x)}$ be the contour shown in Figure 7.1.

Riemann-Hilbert Problem 7.14 (local problem near $\lambda = iA(x)$). Find a matrix $S^{\mathrm{loc}}(\zeta)$ with the following properties:

(1) Analyticity: $S^{\mathrm{loc}}(\zeta)$ is analytic for $\zeta \in \mathbb{C} \setminus \Sigma_{iA(x)}$.

(2) Boundary behavior: $S^{\mathrm{loc}}(\zeta)$ assumes continuous boundary values on $\Sigma_{iA(x)}$.

(3) Jump conditions: the boundary values taken on $\Sigma_{iA(x)}$ satisfy

$$\begin{aligned}
\mathbf{S}^{\mathrm{loc}}_{+}(\zeta) &= \mathbf{S}^{\mathrm{loc}}_{-}(\zeta)\begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \quad \text{for } \arg(\zeta) = 0, \\
\mathbf{S}^{\mathrm{loc}}_{+}(\zeta) &= \mathbf{S}^{\mathrm{loc}}_{-}(\zeta)\begin{bmatrix} 1 & -ie^{-i\zeta^{3/2}} \\ 0 & 1 \end{bmatrix}, \quad \text{for } \arg(\zeta) = -\frac{\pi}{3}, \\
\mathbf{S}^{\mathrm{loc}}_{+}(\zeta) &= \mathbf{S}^{\mathrm{loc}}_{-}(\zeta)\begin{bmatrix} 1 & -ie^{i\zeta^{3/2}} \\ 0 & 1 \end{bmatrix}, \quad \text{for } \arg(\zeta) = \frac{\pi}{3}, \\
\mathbf{S}^{\mathrm{loc}}_{+}(\zeta) &= \mathbf{S}^{\mathrm{loc}}_{-}(\zeta)\begin{bmatrix} 1 & 0 \\ ie^{-(-\zeta)^{3/2}} & 1 \end{bmatrix}, \quad \text{for } \arg(\zeta) = \pi.
\end{aligned} \tag{7.97}$$

(4) Normalization: $\mathbf{S}^{\mathrm{loc}}(\zeta)$ is normalized at infinity so that

$$\mathbf{S}^{\mathrm{loc}}(\zeta)\tilde{\mathbf{S}}^{\mathrm{loc}}(\zeta)^{-1} \longrightarrow \mathbb{I} \quad \text{as } \zeta \longrightarrow \infty. \tag{7.98}$$

We will describe how Riemann-Hilbert Problem 7.14 can be solved explicitly. First, we make an explicit change of variable to a new matrix $\mathbf{T}(\zeta)$ by setting

$$\mathbf{S}^{\mathrm{loc}}(\zeta) = \mathbf{T}(\zeta)\begin{bmatrix} e^{-(-\zeta)^{3/2}/2+i\pi/4} & 0 \\ 0 & e^{(-\zeta)^{3/2}/2-i\pi/4} \end{bmatrix}. \tag{7.99}$$

If $\mathbf{S}^{\mathrm{loc}}(\zeta)$ satisfies Riemann-Hilbert Problem 7.14, then it follows that $\mathbf{T}(\zeta)$ obeys the following jump relations:

$$\begin{aligned}
\mathbf{T}_{+}(\zeta) &= \mathbf{T}_{-}(\zeta)\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \text{for } \arg(\zeta) = 0, \\
\mathbf{T}_{+}(\zeta) &= \mathbf{T}_{-}(\zeta)\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \text{for } \arg(\zeta) = \pm\frac{\pi}{3}, \\
\mathbf{T}_{+}(\zeta) &= \mathbf{T}_{-}(\zeta)\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad \text{for } \arg(\zeta) = \pi.
\end{aligned} \tag{7.100}$$

It is a general fact that $N \times N$ matrix functions that satisfy piecewise constant jump conditions, like the matrix $\mathbf{T}(\zeta)$ does, can be expressed in terms of solutions of Nth order linear differential equations with meromorphic (and often rational, or even polynomial) coefficients. In the 2×2 case, classical special functions therefore play a key role. In this case, we see immediately from the fact that the boundary values taken by $\mathbf{S}^{\mathrm{loc}}(\zeta)$ on

$\Sigma_{iA(x)}$ are continuous that the matrix

$$\begin{aligned}\mathbf{Q}(\zeta) &:= \frac{d\mathbf{T}}{d\zeta}(\zeta)\mathbf{T}(\zeta)^{-1} \\ &= \frac{d\mathbf{S}^{\mathrm{loc}}}{d\zeta}(\zeta)\mathbf{S}^{\mathrm{loc}}(\zeta)^{-1} - \frac{3}{4}(-\zeta)^{1/2}\mathbf{S}^{\mathrm{loc}}(\zeta)\begin{bmatrix}1 & 0 \\ 0 & -1\end{bmatrix}\mathbf{S}^{\mathrm{loc}}(\zeta)^{-1}\end{aligned} \tag{7.101}$$

is analytic for $\zeta \in \mathbb{C}^*$. If we suppose that $\mathbf{T}(\zeta)$ has a bounded derivative near the origin in each sector of $\mathbb{C} \setminus \Sigma_{iA(x)}$—a hypothesis that must be verified later—then we see that in fact $\mathbf{Q}(\zeta)$ is analytic at the origin and is consequently an entire function of ζ.

To work out how $\mathbf{Q}(\zeta)$ behaves for large ζ, we need to use the normalization condition (7.98) for $\mathbf{S}^{\mathrm{loc}}(\zeta)$. We interpret (7.98) to mean both that

$$\mathbf{S}^{\mathrm{loc}}(\zeta) = \left(\mathbb{I} + O\left(\frac{1}{\zeta}\right)\right)\tilde{\mathbf{S}}^{\mathrm{loc}}(\zeta) \tag{7.102}$$

and also that

$$\frac{d\mathbf{S}^{\mathrm{loc}}}{d\zeta}(\zeta) = \left(\mathbb{I} + O\left(\frac{1}{\zeta}\right)\right)\frac{d\tilde{\mathbf{S}}^{\mathrm{loc}}}{d\zeta}(\zeta) + O\left(\frac{1}{\zeta^2}\right)\tilde{\mathbf{S}}^{\mathrm{loc}}(\zeta). \tag{7.103}$$

Both (7.102) and (7.103) are again hypotheses that must be verified once we obtain a solution for $\mathbf{S}^{\mathrm{loc}}(\zeta)$. They are not true a priori by virtue of (7.98) alone; for example the decay rate in (7.98) might not be as fast as $1/\zeta$, and the error term might have rapid oscillations that would make its derivative larger than $1/\zeta^2$ thus violating (7.103). It follows from our hypotheses that $\mathbf{Q}(\zeta)$ must be a polynomial; in fact,

$$\mathbf{Q}(\zeta) = \frac{3}{4}\begin{bmatrix}0 & -\zeta \\ 1 & 0\end{bmatrix}. \tag{7.104}$$

With the matrix $\mathbf{Q}(\zeta)$ explicitly known, we find that the matrix $\mathbf{T}(\zeta)$ solves the linear differential equation

$$\frac{d\mathbf{T}}{d\zeta}(\zeta) = \frac{3}{4}\begin{bmatrix}0 & -\zeta \\ 1 & 0\end{bmatrix}\mathbf{T}(\zeta). \tag{7.105}$$

Upon introducing the new independent variable

$$\xi := -\left(\frac{3}{4}\right)^{2/3}\zeta \tag{7.106}$$

we see that the elements of the second row of $\mathbf{T}$ satisfy Airy's equation

$$\frac{d^2T_{2k}}{d\xi^2} = \xi T_{2k}, \tag{7.107}$$

and that the elements of the first row are given by

$$T_{1k} = -\left(\frac{4}{3}\right)^{1/3}\frac{dT_{2k}}{d\xi}. \tag{7.108}$$

So which solutions of Airy's equation are the appropriate ones for our purposes? The first observation is that we need to specify different solutions of Airy's equation in each simply-connected region of the complex plane where $T_{21}(\xi)$ and $T_{22}(\xi)$ are analytic. The assignment of solutions in these regions must be consistent with the jump conditions and asymptotics for $\mathbf{T}(\zeta)$. From the jump conditions for $\mathbf{T}(\zeta)$, we can see that in fact $T_{21}(\xi)$ is analytic in $\mathbb{C}\backslash\mathbb{R}$, while $T_{22}(\xi)$ is analytic except when $\arg(\xi) = \pm 2\pi/3$ or $\xi \in \mathbb{R}_-$. We now want to use the normalization condition (7.102) to find sectors in which T_{21} and T_{22} are exponentially decaying. Then we will be able to uniquely identify these functions with particular solutions of Airy's equation that also decay. From the presumed asymptotic relation (7.102) we have (in terms of the variable ξ),

$$\begin{aligned} T_{21}(\xi) &= -\left(\frac{3}{32}\right)^{1/6} e^{-i\pi/4}\frac{e^{2\xi^{3/2}/3}}{\xi^{1/4}}\left(1 + O\left(\frac{1}{\xi^{1/2}}\right)\right), \\ T_{22}(\xi) &= \left(\frac{3}{32}\right)^{1/6} e^{i\pi/4}\frac{e^{-2\xi^{3/2}/3}}{\xi^{1/4}}\left(1 + O\left(\frac{1}{\xi^{1/2}}\right)\right), \end{aligned} \tag{7.109}$$

as $\xi \to \infty$ with $-\pi < \arg(\xi) < \pi$. These show that $T_{21}(\xi)$ is exponentially decaying for $-\pi < \arg(\xi) < -\pi/3$ and also for $\pi/3 < \arg(\xi) < \pi$, while $T_{22}(\xi)$ is exponentially decaying for $-\pi/3 < \arg(\xi) < \pi/3$. Significantly, both matrix elements are analytic throughout the sectors where they are exponentially decaying for large ξ. As a basis of linearly independent solutions of Airy's equation we take the functions $\mathrm{Ai}(\xi)$ and $\mathrm{Ai}(\xi e^{2i\pi/3})$. These decay in different sectors, and have the asymptotic expansions

$$\mathrm{Ai}(\xi) = \frac{1}{2\sqrt{\pi}}\frac{e^{-2\xi^{3/2}/3}}{\xi^{1/4}}\left(1 + O\left(\frac{1}{\xi}\right)\right) \tag{7.110}$$

as $\xi \to \infty$ for $-\pi < \arg(\xi) < \pi$, and

$$\mathrm{Ai}\left(\xi e^{2i\pi/3}\right) = \frac{e^{-i\pi/6}}{2\sqrt{\pi}}\frac{e^{2\xi^{3/2}/3}}{\xi^{1/4}}\left(1 + O\left(\frac{1}{\xi}\right)\right) \tag{7.111}$$

as $\xi \to \infty$ for $-\pi < \arg(\xi) < -\pi/3$. Comparing the expansion of $\mathrm{Ai}(\xi e^{2i\pi/3})$ with that of $T_{21}(\xi)$ in the sector $-\pi < \arg(\xi) < -\pi/3$, we find that here

$$T_{21}(\xi) = -e^{-i\pi/12} 6^{1/6} \sqrt{\pi}\, \mathrm{Ai}\left(\xi e^{2i\pi/3}\right). \tag{7.112}$$

Since T_{21} is analytic in the lower half ξ-plane, this relation holds identically for $\Im(\xi) < 0$. Similarly, comparing the expansion of $\mathrm{Ai}(\xi)$ with that of $T_{22}(\xi)$ in the sector $-\pi/3 < \arg(\xi) < \pi/3$, we find that here

$$T_{22}(\xi) = e^{i\pi/4} 6^{1/6} \sqrt{\pi}\, \mathrm{Ai}(\xi). \tag{7.113}$$

Being as T_{22} is analytic for $-2\pi/3 < \arg(\xi) < 2\pi/3$, this identity holds throughout the sector of analyticity.

Restoring the original independent variable ζ, we find that for $\Im(\zeta) > 0$,

$$T_{21}(\zeta) = -e^{-i\pi/12} 6^{1/6} \sqrt{\pi}\, \mathrm{Ai}\left(\left(\frac{3}{4}\right)^{2/3} \zeta e^{-i\pi/3}\right) \tag{7.114}$$

and therefore throughout the same domain,

$$T_{11}(\zeta) = e^{i\pi/12} \left(\frac{32}{3}\right)^{1/6} \sqrt{\pi}\, \mathrm{Ai}'\left(\left(\frac{3}{4}\right)^{2/3} \zeta e^{-i\pi/3}\right). \tag{7.115}$$

For ζ with $\pi/3 < \arg(\zeta) \le \pi$ or $-\pi \le \arg(\zeta) < -\pi/3$,

$$T_{22}(\zeta) = e^{i\pi/4} 6^{1/6} \sqrt{\pi}\, \mathrm{Ai}\left(-\left(\frac{3}{4}\right)^{2/3} \zeta\right), \tag{7.116}$$

and therefore throughout the same domain,

$$T_{12}(\zeta) = e^{3i\pi/4} \left(\frac{32}{3}\right)^{1/6} \sqrt{\pi}\, \mathrm{Ai}'\left(-\left(\frac{3}{4}\right)^{2/3} \zeta\right). \tag{7.117}$$

The sector of $\mathbb{C} \setminus \Sigma_{iA(x)}$ that is contained in both of these domains is $\pi/3 < \arg(\zeta) < \pi$. It is sufficient to have specified the matrix elements of $\mathbf{T}(\zeta)$ in this sector, since it may be consistently obtained in the remaining sectors of $\mathbb{C} \setminus \Sigma_{iA(x)}$ by making use of the jump relations for $\mathbf{T}(\zeta)$. The procedure is consistent because the cyclic product of these jump matrices is the identity

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{-1} = \mathbb{I}. \tag{7.118}$$

Once $\mathbf{T}(\zeta)$ is known for $\zeta \in \mathbb{C} \setminus \Sigma_{iA(x)}$, the original unknown matrix $\mathbf{S}^{\mathrm{loc}}(\zeta)$ solving Riemann-Hilbert Problem 7.14 is obtained directly by the transformation (7.99). It suffices to give a formula that holds for $\pi/3 < \arg(\zeta) < \pi$. We find

$$\begin{aligned} S_{11}^{\mathrm{loc}}(\zeta) &= e^{i\pi/3}\left(\frac{32}{3}\right)^{1/6}\sqrt{\pi}e^{-(-\zeta)^{3/2}/2}\,\mathrm{Ai}'\left(\left(\frac{3}{4}\right)^{2/3}\zeta e^{-i\pi/3}\right),\\ S_{12}^{\mathrm{loc}}(\zeta) &= i\left(\frac{32}{3}\right)^{1/6}\sqrt{\pi}e^{(-\zeta)^{3/2}/2}\,\mathrm{Ai}'\left(-\left(\frac{3}{4}\right)^{2/3}\zeta\right),\\ S_{21}^{\mathrm{loc}}(\zeta) &= e^{-5i\pi/6}6^{1/6}\sqrt{\pi}e^{-(-\zeta)^{3/2}/2}\,\mathrm{Ai}\left(\left(\frac{3}{4}\right)^{2/3}\zeta e^{-i\pi/3}\right),\\ S_{22}^{\mathrm{loc}}(\zeta) &= 6^{1/6}\sqrt{\pi}e^{(-\zeta)^{3/2}/2}\,\mathrm{Ai}\left(-\left(\frac{3}{4}\right)^{2/3}\zeta\right). \end{aligned} \tag{7.119}$$

We would have found a solution to Riemann-Hilbert Problem 7.14 if we can verify the two hypotheses (cf. equations (7.102) and (7.103)) we made regarding the interpretation of the normalization condition (7.98) and the differentiability of $\mathbf{S}^{\mathrm{loc}}(\zeta)$ at $\zeta = 0$. One verifies these directly, using the explicit formulae given here.

7.3.2 The local model for $\mathbf{N}(\lambda)$ *near* $\lambda = iA(x)$. To build a better model for $\mathbf{N}(\lambda)$ in the disk D than $\widehat{\mathbf{N}}_{\mathrm{out}}(\lambda)$, we begin by recalling the exact relationship between the matrix $\mathbf{O}(\lambda)$ and the matrix $\mathbf{S}(\lambda)$:

$$\mathbf{O}(\lambda) = \sigma_1^{(1-J)/2}\,\mathbf{S}(\varphi_{iA(x)}(\lambda))\sigma_1^{(1-J)/2} \tag{7.120}$$

for all $\lambda \in D$. The matrix $\tilde{\mathbf{N}}(\lambda)$ is also explicitly related to $\mathbf{O}(\lambda)$. For $\lambda \in D$ in the part of the lens between the contours L_L and C_M,

$$\tilde{\mathbf{N}}(\lambda) = \mathbf{O}(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} 1 & -i\exp\left(-\dfrac{i\theta(\lambda)}{\hbar_N}\right) \\ 0 & 1 \end{bmatrix}\sigma_1^{(1-J)/2}, \tag{7.121}$$

for $\lambda \in D$ in the part of the lens between the contours C_M and L_R,

$$\tilde{\mathbf{N}}(\lambda) = \mathbf{O}(\lambda)\sigma_1^{(1-J)/2}\begin{bmatrix} 1 & i\exp\left(\dfrac{i\theta(\lambda)}{\hbar_N}\right) \\ 0 & 1 \end{bmatrix}\sigma_1^{(1-J)/2}, \tag{7.122}$$

and for all other $\lambda \in D$, we simply have $\tilde{\mathbf{N}}(\lambda) = \mathbf{O}(\lambda)$. As we do not expect the difference between $\mathbf{N}(\lambda)$ and the formal continuum limit approximation $\tilde{\mathbf{N}}(\lambda)$ to be important in the disk D since it is isolated from the points $\lambda = 0$ and $\lambda = iA$, we can obtain a guess for an approximation for $\mathbf{N}(\lambda)$ that should be valid in D simply by substituting $\widehat{\mathbf{S}}(\varphi_{iA(x)}(\lambda))$ for $\mathbf{S}(\varphi_{iA(x)}(\lambda))$ in these formulae.

Putting these steps together, the model for $\mathbf{N}(\lambda)$ for $\lambda \in D$ that we will use is defined as follows. For $\lambda \in D$ in the lens between L_L and C_M, set

$$\widehat{\mathbf{N}}_{\text{endpoint}}(\lambda) := \sigma_1^{(1-J)/2}\,\widehat{\mathbf{S}}\left(\varphi_{iA(x)}(\lambda)\right)\begin{bmatrix}1 & -i\exp\left(-\dfrac{i\theta(\lambda)}{\hbar_N}\right)\\ 0 & 1\end{bmatrix}\sigma_1^{(1-J)/2}, \tag{7.123}$$

for $\lambda \in D$ in the lens between C_M and L_R, set

$$\widehat{\mathbf{N}}_{\text{endpoint}}(\lambda) := \sigma_1^{(1-J)/2}\,\widehat{\mathbf{S}}\left(\varphi_{iA(x)}(\lambda)\right)\begin{bmatrix}1 & i\exp\left(\dfrac{i\theta(\lambda)}{\hbar_N}\right)\\ 0 & 1\end{bmatrix}\sigma_1^{(1-J)/2}, \tag{7.124}$$

and for all other $\lambda \in D$, set

$$\widehat{\mathbf{N}}_{\text{endpoint}}(\lambda) := \sigma_1^{(1-J)/2}\,\widehat{\mathbf{S}}\left(\varphi_{iA(x)}(\lambda)\right)\sigma_1^{(1-J)/2}, \tag{7.125}$$

where $\widehat{\mathbf{S}}(\zeta)$ is defined by (7.96). The most important properties of the matrix $\widehat{\mathbf{N}}_{\text{endpoint}}(\lambda)$ in the disk D are the following.

Proposition 7.15. The matrix $\widehat{\mathbf{N}}_{\text{endpoint}}(\lambda)$ is piecewise analytic in the left and right half-disks of D. On the imaginary axis (which bisects D) oriented in the positive imaginary direction,

$$\widehat{\mathbf{N}}_{\text{endpoint},-}(\lambda)^{-1}\widehat{\mathbf{N}}_{\text{endpoint},+}(\lambda) = \tilde{\mathbf{N}}_-(\lambda)^{-1}\tilde{\mathbf{N}}_+(\lambda), \tag{7.126}$$

that is, the local model has exactly the same jump as $\tilde{\mathbf{N}}(\lambda)$. For $\lambda \in D$, the matrix function $\widehat{\mathbf{N}}_{\text{endpoint}}(\lambda)$ is bounded by a constant of order $\hbar_N^{-1/3}$. Also, for $\lambda \in \partial D$,

$$\widehat{\mathbf{N}}_{\text{endpoint}}(\lambda)\widehat{\mathbf{N}}_{\text{out}}(\lambda)^{-1} = \mathbb{I} + O\left(\hbar_N^{1/3}\right). \tag{7.127}$$

□

Proof. Computing the jump matrix for $\widehat{\mathbf{N}}_{\text{endpoint}}(\lambda)$ is straightforward. To show the bound on $\widehat{\mathbf{N}}_{\text{endpoint}}(\lambda)$, we recall that $\tilde{\mathbf{S}}^{\text{hol}}(\varphi_{iA(x)}(\lambda))$ is bounded by a quantity of order $\hbar_N^{-1/6}$, and note that the other factor of the same size comes from the factor $\mathbf{S}^{\text{loc}}(\varphi_{iA(x)}(\lambda))$ via the normalization condition on this matrix and the fact that $\varphi_{iA(x)}(\lambda)$ grows like $\hbar_N^{-2/3}$ for $\lambda \in D$. Similar reasoning using (7.102) establishes the error in matching onto $\widehat{\mathbf{N}}_{\text{out}}(\lambda)$ on the boundary of D. ■

8 The parametrix and its error

We are now in a position to put all of our models together to build a guess for a uniformly valid approximation of $\mathbf{N}(\lambda)$. Such a guess is called a parametrix.

8.1 Constructing the parametrix

To build the parametrix $\widehat{\mathbf{N}}(\lambda)$ as a sectionally holomorphic matrix function, we simply combine the outer and local models. For all λ satisfying $|\lambda| \le \hbar_N^{\alpha}$, where the parameter α is to be determined later, set

$$\widehat{\mathbf{N}}(\lambda) := \widehat{\mathbf{N}}_{\mathrm{origin}}(\lambda). \tag{8.1}$$

For $\lambda \in D$, we set

$$\widehat{\mathbf{N}}(\lambda) := \widehat{\mathbf{N}}_{\mathrm{endpoint}}(\lambda), \tag{8.2}$$

and by symmetry for all $\lambda \in D^*$ we set

$$\widehat{\mathbf{N}}(\lambda) := \sigma_2 \widehat{\mathbf{N}}_{\mathrm{endpoint}}(\lambda^*)^* \sigma_2. \tag{8.3}$$

Finally, for all remaining $\lambda \in \mathbb{C}$, set

$$\widehat{\mathbf{N}}(\lambda) := \widehat{\mathbf{N}}_{\mathrm{out}}(\lambda). \tag{8.4}$$

The parametrix $\widehat{\mathbf{N}}(\lambda)$ is holomorphic for $\lambda \in \mathbb{C} \setminus \widehat{\Sigma}$, where $\widehat{\Sigma}$ is the contour illustrated in Figure 8.1.

8.2 Estimating the error

To determine the accuracy of the parametrix, we compare it directly with the original matrix $\mathbf{N}(\lambda)$. That is, we consider the error matrix defined by

$$\mathbf{E}(\lambda) := \mathbf{N}(\lambda)\widehat{\mathbf{N}}(\lambda)^{-1}. \tag{8.5}$$

This matrix is sectionally analytic in the complex λ-plane, with discontinuities on a contour Σ_E that is illustrated in Figure 8.2. Note that as a consequence of the symmetry of $\widehat{\mathbf{N}}(\lambda)$ and $\mathbf{N}(\lambda)$ under complex conjugation, we have $\mathbf{E}(\lambda^*) = \sigma_2\mathbf{E}(\lambda)^*\sigma_2$. If the parametrix

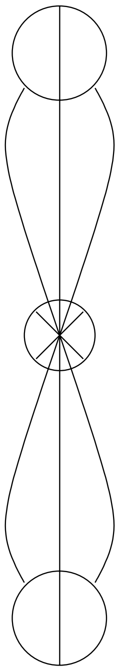

Figure 8.1 The contour $\widehat{\Sigma}$. The circles at the top and bottom of the figure are the boundaries of the disks D and D^*, respectively. The circle at the origin has radius $\hbar_N^\alpha$. The contours to the left and right of the imaginary axis in the upper half-plane are portions of the lens boundaries L_L and L_R, respectively. The remaining small segments present in the upper half-plane for $|\lambda| < \hbar_N^\alpha$ are parts of C_L and C_R.

is indeed a good model for $\mathbf{N}(\lambda)$, then we must be able to show that the matrix $\mathbf{E}(\lambda)$ is uniformly close to the identity matrix in the whole complex plane.

While we do not know $\mathbf{E}(\lambda)$ explicitly like we know $\widehat{\mathbf{N}}(\lambda)$, we know from the normalization condition of both factors that

$$\mathbf{E}(\lambda) \longrightarrow \mathbb{I} \quad \text{as } \lambda \longrightarrow \infty. \tag{8.6}$$

It turns out that we can also calculate explicitly the ratio of boundary values taken by $\mathbf{E}(\lambda)$ from both sides on each arc of Σ_E. That is, we know the jump matrix for $\mathbf{E}(\lambda)$, and can express it explicitly in terms of $\widehat{\mathbf{N}}(\lambda)$ and the jump matrix for $\mathbf{N}(\lambda)$, both of which are known.[3] This means that the matrix $\mathbf{E}(\lambda)$ itself is a solution of a particular Riemann-Hilbert problem for which we know the data. By solving this Riemann-Hilbert problem, we will show that indeed $\mathbf{E}(\lambda)$ is uniformly close to the identity matrix.

[3]Or at least well-understood. We have characterized the parametrix for $|\lambda| \leq \hbar_N^\alpha$ in terms of the matrix function $\widehat{F}(\zeta)$ for which we have an existence proof and a characterization, but not an explicit formula.

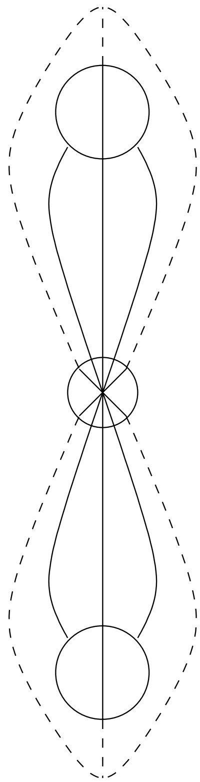

Figure 8.2 The contour Σ_E. We have $\widehat{\Sigma} \subset \Sigma_E$ and the components of $\Sigma_E \setminus \widehat{\Sigma}$ are shown in dashed lines to make a clear comparison with Figure 8.1.

There are two kinds of arcs in the contour Σ_E: "matching" arcs of the circles ∂D, ∂D^*, and $|\lambda| = \hbar_N^\alpha$ where two different components of the parametrix have to match well onto each other, and the remaining arcs within the disks and outside the disks where the jump matrix for $\widehat{\mathbf{N}}(\lambda)$ should be a good approximation to that of $\mathbf{N}(\lambda)$.

Consider one of the arcs of Σ_E oriented in some convenient way, and as usual let the subscript "+" (resp., "−") denote a boundary value taken on the arc from its left (resp., right). We can easily see from the definition (8.5) that for λ on this arc,

$$\mathbf{E}_+(\lambda) = \mathbf{E}_-(\lambda)\boldsymbol{v}_{\mathbf{E}}(\lambda) \quad \text{with} \quad \boldsymbol{v}_{\mathbf{E}}(\lambda) := \widehat{\mathbf{N}}_-(\lambda)\boldsymbol{v}_{\mathbf{N}}(\lambda)\widehat{\boldsymbol{v}}_{\widehat{\mathbf{N}}}(\lambda)^{-1}\widehat{\mathbf{N}}_-(\lambda)^{-1}, \tag{8.7}$$

where $\boldsymbol{v}_{\mathbf{N}}(\lambda)$ and $\boldsymbol{v}_{\widehat{\mathbf{N}}}(\lambda)$ denote the jump matrices on the arc for $\mathbf{N}(\lambda)$ and the parametrix $\widehat{\mathbf{N}}(\lambda)$, respectively. If the arc under consideration is a "matching" arc, then the discontinuity in $\mathbf{E}(\lambda)$ is wholly due to the mismatch of components of the parametrix, and the jump matrix $\boldsymbol{v}_{\mathbf{N}}(\lambda)$ is therefore replaced with the identity matrix in (8.7). It then follows that an equivalent formula for $\mathbf{v}_E(\lambda)$ on a "matching" arc is the following:

$$\mathbf{v}_E(\lambda) = \widehat{\mathbf{N}}_-(\lambda)\widehat{\mathbf{N}}_+(\lambda)^{-1}, \quad \text{for } \lambda \text{ on a "matching" arc of } \Sigma_E. \tag{8.8}$$

In this case, the two boundary values represent different components of the parametrix, for example $\widehat{\mathbf{N}}_{\text{out}}(\lambda)$ would play the role of $\widehat{\mathbf{N}}_+(\lambda)$ and $\widehat{\mathbf{N}}_{\text{endpoint}}(\lambda)$ would play that of $\widehat{\mathbf{N}}_-(\lambda)$ if the "matching" arc under consideration is an arc of ∂D, oriented clockwise.

The key fact that we need now is the following.

Proposition 8.1. The optimal value of the radius parameter α is $\alpha = 6/7$. For this value of α, and for all $\nu > 0$ arbitrarily small,

$$\mathbf{v}_E(\lambda) - \mathbb{I} = O\left(\hbar_N^{1/7-\nu}\right) \tag{8.9}$$

uniformly for all $\lambda \in \Sigma_E$. □

Proof. We begin by considering the "matching" arcs. We take the circle ∂D to be oriented in the clockwise direction. Here we find

$$\mathbf{v}_E(\lambda) = \widehat{\mathbf{N}}_{\text{endpoint}}(\lambda)\widehat{\mathbf{N}}_{\text{out}}(\lambda)^{-1} = \mathbb{I} + O\left(\hbar_N^{1/3}\right), \tag{8.10}$$

with the error estimate coming from Proposition 7.15. An estimate of the same form necessarily holds on the "matching" arcs of ∂D^* according to the conjugation symmetry of $\mathbf{E}(\lambda)$. The remaining "matching" arcs lie on the circle $|\lambda| = \hbar_N^\alpha$, which again we take to be oriented in the clockwise direction. Here we find

$$\mathbf{v}_E(\lambda) = \widehat{\mathbf{N}}_{\text{origin}}(\lambda)\widehat{\mathbf{N}}_{\text{out}}(\lambda)^{-1} = \mathbb{I} + O\left(\hbar_N^{1-\alpha}\right), \tag{8.11}$$

with the error estimate coming from Proposition 7.8.

We continue by considering the arcs of Σ_E with $|\lambda| < \hbar_N^\alpha$. Using the fact recorded in Proposition 7.8 that $\widehat{\mathbf{N}}_{\text{origin}}(\lambda)$ has determinant one and is uniformly bounded, we see from (8.7) that the important quantity to estimate is simply $\mathbf{v}_N(\lambda)\mathbf{v}_{\widehat{N}}(\lambda)^{-1} - \mathbb{I}$, the difference between the jump matrix ratio and the identity. First, consider the portion of the contour C_L with $|\lambda| < 1$. Using Proposition 7.8 and (4.2), we find that here

$$\begin{aligned}
&\mathbf{v}_N(\lambda)\mathbf{v}_{\widehat{N}}(\lambda)^{-1} \\
&\quad = \sigma_1^{(1-J)/2}\begin{bmatrix} 1 & 0 \\ a_L(\lambda) - ie^{i\theta(0)/\hbar_N}e^{(2i+\pi)\zeta}(-i\zeta)^{-i\zeta}(i\zeta)^{-i\zeta}\dfrac{\Gamma(1/2+i\zeta)}{\Gamma(1/2-i\zeta)} & 1\end{bmatrix}\sigma_1^{(1-J)/2},
\end{aligned} \tag{8.12}$$

where $\zeta = \varphi_0(\lambda)$. Using (7.50) and the uniform boundedness of the leading-order term on the right-hand side of (7.50) for $|\lambda| < \hbar_N^\alpha$, we see that the matrix quotient in (8.12) differs from the identity matrix by an order $\hbar_N^{4\alpha/3-1}$ amount. Virtually the same argument using (4.4) and (7.52) in conjunction with Proposition 7.8 establishes that on C_R the matrix quotient $\mathbf{v}_N(\lambda)\mathbf{v}_{\widehat{N}}(\lambda)^{-1}$ differs from the identity by a quantity of order $\hbar_N^{4\alpha/3-1}$. Next, consider the contour L_L for $|\lambda| < \hbar_N^\alpha$. On this contour, there is no jump for $\mathbf{N}(\lambda)$, so $\mathbf{v}_N(\lambda) = \mathbb{I}$ in the formula (8.7) for $\mathbf{v}_E(\lambda)$. But from the formula for $\mathbf{v}_{\widehat{N}}(\lambda)$ for $\lambda \in L_L$ given in Proposition 7.8, we see by ordinary Taylor expansion that for $|\lambda| < \hbar_N^\alpha$, we have $\mathbf{v}_E(\lambda) - \mathbb{I} = O(\hbar_N^{2\alpha-1})$ for $\lambda \in L_L$. Virtually the same argument yields the same estimate for $\mathbf{v}_E(\lambda) - \mathbb{I}$ on L_R with $|\lambda| < \hbar_N^\alpha$. Finally, consider the contour C_M (the positive imaginary axis) with $|\lambda| < \hbar_N^\alpha$. Using (4.6) and the jump matrix $\mathbf{v}_{\widehat{N}}(\lambda)$ for $\lambda \in C_M$ recorded in Proposition 7.8, we find that here

$$\begin{aligned}\mathbf{v}_N(\lambda)\mathbf{v}_{\widehat{N}}(\lambda)^{-1} &= \sigma_1^{(1-J)/2}\begin{bmatrix} 1 + W_+ Z & i e^{i\theta(\lambda)/\hbar_N}\left(W_+ + W_-\right)Z \\ e^{-i\theta(\lambda)/\hbar_N}\left(a_M(\lambda) - i(1+Z)\right) & 1 + i\left(W_+ + W_-\right)a_M(\lambda)Z + W_- Z\end{bmatrix} \\ &\quad \times \sigma_1^{(1-J)/2},\end{aligned} \tag{8.13}$$

where Z is given in terms of $\zeta = \varphi_0(\lambda)$ by (7.66) and where

$$W_\pm := 1 - e^{\pm\pi\zeta + i(\theta(\lambda)-\theta(0))/\hbar_N} . \tag{8.14}$$

Note that W_+ and W_- are both of order $\hbar_N^{2\alpha-1}$ for $|\lambda| < \hbar_N^\alpha$ by Taylor expansion arguments. Also, Z is uniformly bounded in the disk of radius $\hbar_N^\alpha$, and $e^{\pm i\theta(\lambda)/\hbar_N}$ both have modulus one for $\lambda \in C_M$ in this disk. Also, from (7.54) we get $a_M(\lambda) - i(1+Z) = O(\hbar_N^{4\alpha/3-1-\nu})$ for all $\nu > 0$ since $\tilde{\phi}(\lambda) \equiv 0$ on this part of C_M. This error dominates those arising from Taylor approximation, and thus on C_M with $|\lambda| < \hbar_N^\alpha$ we find that $\mathbf{v}_N(\lambda)\mathbf{v}_{\widehat{N}}(\lambda)^{-1} - \mathbb{I} = O(\hbar_N^{4\alpha/3-1-\nu})$. The corresponding estimates hold on the corresponding contours in the lower half-disk, by conjugation symmetry. Putting this information together with the uniform boundedness of $\widehat{\mathbf{N}}_{\text{origin}}(\lambda)$ and its inverse, we find that for all $\lambda \in \Sigma_E$ with $|\lambda| < \hbar_N^\alpha$,

$$\mathbf{v}_E(\lambda) - \mathbb{I} = O\left(\hbar_N^{4\alpha/3-1-\nu}\right), \tag{8.15}$$

for all $\nu > 0$.

Now we proceed to study the jump matrix $\mathbf{v}_E(\lambda)$ inside the disk D centered at the endpoint $\lambda = iA(x)$, assuming $x \neq 0$ is fixed so that D is fixed and bounded away from the origin and from $\lambda = iA$. The only contour we need to consider is the imaginary axis. The jump matrix $\mathbf{v}_\mathbf{E}(\lambda)$ is given by (8.7). This time, the conjugating factors of $\widehat{\mathbf{N}}_{\text{endpoint},-}(\lambda)$ and its inverse are not uniformly bounded in D as $\hbar_N$ tends to zero. According to Proposition 7.15 each conjugating matrix contributes an amplifying factor of $\hbar_N^{-1/3}$. Also according to Proposition 7.15, we have that $\mathbf{v}_{\widehat{\mathbf{N}}}(\lambda)$ is the same as the jump matrix for $\tilde{\mathbf{N}}(\lambda)$. Therefore, using (4.6) and (5.32), we find that

$$\mathbf{v}_\mathbf{N}(\lambda)\mathbf{v}_{\widehat{\mathbf{N}}}(\lambda)^{-1} = \sigma_1^{(1-J)/2}\begin{bmatrix} 1 & 0 \\ e^{-i\theta(\lambda)/\hbar_N}\left(a_M(\lambda) - ie^{i\tilde{\phi}(\lambda)/\hbar_N}\right) & 1 \end{bmatrix}\sigma_1^{(1-J)/2}. \tag{8.16}$$

Using Proposition 5.6 and the fact that $\Re(\tilde{\phi}) \leq 0$ while $e^{-i\theta(\lambda)/\hbar_N}$ has modulus one on C_M within D, we get that $\mathbf{v}_\mathbf{N}(\lambda)\mathbf{v}_{\widehat{\mathbf{N}}}(\lambda)^{-1} = \mathbb{I} + O(\hbar_N^{1-\mu})$ for all $\mu > 0$. Combining this with the bounds on the conjugating factors, we find that within D,

$$\mathbf{v}_\mathbf{E}(\lambda) - \mathbb{I} = O\left(\hbar_N^{1/3-\mu}\right), \tag{8.17}$$

for all $\mu > 0$. By conjugation symmetry, the same estimate holds for the jump matrix $\mathbf{v}_\mathbf{E}(\lambda)$ when $\lambda \in D^*$.

Finally, we consider the parts of Σ_E outside all of the disks, where we have set $\widehat{\mathbf{N}}(\lambda) := \widehat{\mathbf{N}}_{\text{out}}(\lambda)$. On all of these parts of Σ_E, Proposition 6.5 guarantees that the conjugating factors $\widehat{\mathbf{N}}_{\text{out},-}(\lambda)$ and $\widehat{\mathbf{N}}_{\text{out},-}(\lambda)^{-1}$ are uniformly bounded as $\hbar_N$ tends to zero. So it remains to determine the magnitude of the difference between the quotient $\mathbf{v}_\mathbf{N}(\lambda)\mathbf{v}_{\widehat{\mathbf{N}}}(\lambda)^{-1}$ and the identity. First consider the part of C_M between the disk at the origin and the disk D. Using Proposition 6.5 to find $\mathbf{v}_{\widehat{\mathbf{N}}}(\lambda)$ and recalling (4.6), we find

$$\mathbf{v}_\mathbf{N}(\lambda)\mathbf{v}_{\widehat{\mathbf{N}}}(\lambda)^{-1} = \sigma_1^{(1-J)/2}\begin{bmatrix} 1 & 0 \\ e^{-i\theta(\lambda)/\hbar_N}\left(a_M(\lambda) - i\right) & 1 \end{bmatrix}\sigma_1^{(1-J)/2}. \tag{8.18}$$

Using Proposition 5.6 and the fact that $\tilde{\phi}(\lambda) \equiv 0$ while $e^{-i\theta(\lambda)/\hbar_N}$ has modulus one here then allows us to conclude that $\mathbf{v}_\mathbf{N}(\lambda)\mathbf{v}_{\widehat{\mathbf{N}}}(\lambda)^{-1} = \mathbb{I} + O(\hbar_N^{1-\alpha-\mu})$ for all $\mu > 0$. The α appears because we need the estimate down to the outside boundary of the shrinking disk at the origin. Next we look at the contours L_L and L_R. On these contours there is no jump for $\mathbf{N}(\lambda)$, and we see directly from Proposition 6.5 that $\mathbf{v}_\mathbf{N}(\lambda)\mathbf{v}_{\widehat{\mathbf{N}}}(\lambda)^{-1} - \mathbb{I}$ is exponentially small as $\hbar_N$ tends to zero through positive values. The next contour we examine is the portion of C_M lying above the disk D. Here we observe that $\widehat{\mathbf{N}}_{\text{out}}(\lambda)$ has no jump, and because we have on this contour the strict inequality $\Re(\tilde{\phi}(\lambda)) < 0$, the matrix $\mathbf{v}_\mathbf{N}(\lambda)$ is exponentially close to the identity matrix. To see this, note that here

$$\mathbf{v}_{\mathbf{N}}(\lambda)=\sigma_1^{(1-J)/2}\begin{bmatrix}1 & 0\\ a_M(\lambda) & 1\end{bmatrix}\sigma_1^{(1-J)/2} \tag{8.19}$$

because $\theta(\lambda)\equiv 0$. While the relative error in replacing $a_M(\lambda)$ by $ie^{\tilde{\phi}(\lambda)/\hbar_N}$ is not small near $\lambda=iA$, it is bounded. So the exponential decay afforded by the strict inequality on the real part of $\tilde{\phi}(\lambda)$ is maintained. Virtually the same arguments show that on the contours C_L and C_R outside of the disk at the origin, the quotient $\mathbf{v}_{\mathbf{N}}(\lambda)\mathbf{v}_{\widehat{\mathbf{N}}}(\lambda)^{-1}$ is again an exponentially small perturbation of the identity matrix, since on these contours there is again no jump of the parametrix $\widehat{\mathbf{N}}_{\text{out}}(\lambda)$. Therefore, for all $\lambda\in\Sigma_E$ outside all disks, we have

$$\mathbf{v}_{\mathbf{E}}(\lambda)-\mathbb{I}=O\left(\hbar_N^{1-\alpha-\mu}\right), \tag{8.20}$$

for all $\mu>0$.

We come up with an overall estimate for $\mathbf{v}_{\mathbf{E}}(\lambda)-\mathbb{I}$ for $\lambda\in\Sigma_E$ by combining the estimates (8.10), (8.11), (8.15), (8.17), and (8.20), and optimizing the error by choosing the parameter α. The optimal balance among all $\alpha\in(3/4,1)$ comes from taking $\alpha=6/7$, which gives an overall error estimate of

$$\mathbf{v}_{\mathbf{E}}(\lambda)-\mathbb{I}=O\left(\hbar_N^{1/7-\nu}\right) \tag{8.21}$$

uniformly for all $\lambda\in\Sigma_E$, for all $\nu>0$. This proves the proposition. ■

The following is then a consequence of the L^2 theory of Riemann-Hilbert problems (see the analogous discussion in [8]).

Proposition 8.2. For $\hbar_N$ sufficiently small, the Riemann-Hilbert problem for $\mathbf{E}(\lambda)$ has a unique solution. Let $R>0$ be sufficiently large so that Σ_E is contained in the circle of radius R centered at the origin. Then uniformly for all λ outside of this circle, and for all $\mu>0$ however small, the matrix $\mathbf{E}(\lambda)$ satisfies

$$\mathbf{E}(\lambda)-\mathbb{I}=O\left(\hbar_N^{1/7-\nu}\right) \tag{8.22}$$

with the size of the matrix measured in any matrix norm. □

Proof. Only one thing must be verified in order to deduce existence and uniqueness from the general theory: the Cauchy-kernel singular integral operators defined on the contour Σ_E, which depends on $\hbar_N$ because of the shrinking boundary of the circle at the origin, have $L^2(\Sigma_E)$ norms that can be bounded uniformly in N. But this fact follows in this case from the fact that for N sufficiently large, the circle $|\lambda| = \hbar_N^\alpha$ intersects only radial straight-line segments (we chose the contours to all be exactly straight lines in some fixed neighborhood of the origin), so that the portion of Σ_E near the origin simply scales with $\hbar_N^\alpha$. The uniform bound we need can then be established using the fact that the Cauchy operators commute with scaling. A similar result was established under more general conditions in [8]. Once existence and uniqueness have been established, the estimate of $\mathbf{E}(\lambda)$ follows from an integral representation formula for this matrix

$$\mathbf{E}(\lambda) = \mathbb{I} + \frac{1}{2\pi i}\int_{\Sigma_E} (s-\lambda)^{-1}\mathbf{m}(s)\big(\boldsymbol{\nu}_\mathbf{E}(s) - \mathbb{I}\big)\, ds \tag{8.23}$$

in which the matrix function $\mathbf{m}(\lambda)$ for $\lambda \in \Sigma_E$ is an element of $L^2(\Sigma_E)$ with a norm that is bounded independently of $\hbar_N$ (and in fact converges to the identity matrix in $L^2(\Sigma_E)$ as $\hbar_N$ tends to zero). ■

We are now in a position to prove our main result, which we presented as Theorem 1.1 in the introduction.

Proof of Theorem 1.1. We have been setting $t = 0$ all along, so the function $\psi_0^{\hbar_N}(x)$ given by (2.11) can be found from the matrix $\mathbf{N}(\lambda)$ by the relation (cf. equation (2.5))

$$\psi_0^{\hbar_N}(x) = 2i \lim_{\lambda\to\infty} \lambda N_{12}(\lambda). \tag{8.24}$$

Writing $\mathbf{N}(\lambda) = \mathbf{E}(\lambda)\widehat{\mathbf{N}}(\lambda) = \widehat{\mathbf{N}}(\lambda) + (\mathbf{E}(\lambda) - \mathbb{I})\widehat{\mathbf{N}}(\lambda)$, we get

$$\begin{aligned}\psi_0^{\hbar_N}(x) &= 2i \lim_{\lambda\to\infty} \lambda \widehat{N}_{12}(\lambda) + 2i \lim_{\lambda\to\infty} \lambda\big(E_11(\lambda) - 1\big)\widehat{N}_{12}(\lambda) \\ &\quad + 2i \lim_{\lambda\to\infty} \lambda E_{12}(\lambda)\widehat{N}_{22}(\lambda).\end{aligned} \tag{8.25}$$

Now, the first term on the right-hand side of (8.25) can be evaluated explicitly since near $\lambda = \infty$ we have $\widehat{\mathbf{N}}(\lambda) \equiv \widehat{\mathbf{N}}_{\text{out}}(\lambda) \equiv \tilde{\mathbf{O}}(\lambda)$, and we have an explicit formula (cf. equation (6.28)) for $\tilde{\mathbf{O}}(\lambda)$. We find

$$2i \lim_{\lambda\to\infty} \lambda \widehat{N}_{12}(\lambda) = A(x), \tag{8.26}$$

which is the "true" initial data that we started with, before making any modifications based on the WKB approximation of the spectral data. When we consider the second term on the right-hand side of (8.25), we see that as a consequence of the normalization of the matrices $\mathbf{E}(\lambda)$ and $\widehat{\mathbf{N}}(\lambda)$,

$$\mathbf{E}(\lambda) = \mathbb{I} + O\left(\frac{1}{\lambda}\right), \qquad \tilde{\mathbf{N}}(\lambda) = \mathbb{I} + O\left(\frac{1}{\lambda}\right), \tag{8.27}$$

as $\lambda \to \infty$. Therefore the second term on the right-hand side of (8.25) vanishes identically. Finally, for the third term on the right-hand side of (8.25) we can again apply the normalization condition for $\widehat{\mathbf{N}}(\lambda)$ to obtain

$$2i \lim_{\lambda \to \infty} \lambda E_{12}(\lambda) \widehat{N}_{22}(\lambda) = 2i \lim_{\lambda \to \infty} \lambda E_{12}(\lambda). \tag{8.28}$$

Putting these steps together, we have

$$\psi_0^{\hbar_N}(x) - A(x) = 2i \lim_{\lambda \to \infty} \lambda E_{12}(\lambda). \tag{8.29}$$

The proof of the theorem is finished upon using the integral formula (8.23) for $\mathbf{E}(\lambda)$ and Proposition 8.1. ■

9 Discussion

Using the new technique of simultaneous interpolation of residues by two different analytic interpolating functions, combined with "steepest-descent" techniques for matrix-valued Riemann-Hilbert problems, we have established the validity of the formal WKB approximation of the spectrum in the nonselfadjoint Zakharov-Shabat eigenvalue problem (1.2) in the sense of pointwise convergence of the potentials. Strictly speaking, our analysis applies to certain classes of potential functions whose most important property for our purposes is their real analyticity, and then we obtain convergence for all nonzero values of x.

In order to extend the result of Theorem 1.1 to $x = 0$, some different steps are required. Since $A(x) \to A$ as $x \to 0$, the local analysis that we carried out independently for $\lambda \approx iA$ (cf. Section 7.2) and for $\lambda \approx iA(x)$ (cf. Section 7.3) will need to be combined. Consequently, a different local model for $\mathbf{N}(\lambda)$ will need to be constructed near $\lambda = iA = iA(0)$. Due to the presence of the gamma functions in the asymptotics established in Section 7.2 and given in Propositions 7.12 and 7.13, it is likely that the construction of the local model will require knowledge of the solution of a new Riemann-Hilbert problem that, like that for the matrix $\widehat{\mathbf{F}}(\zeta)$ in Section 7.1, cannot be solved explicitly. Nonetheless, one expects that to establish the validity of Theorem 1.1 for $x = 0$ will require only technical modifications of what we have done here.

Understanding the nature of the WKB approximation at the level of the potentials is one step in a larger ongoing program to obtain corresponding information at the level of the (unknown) spectrum itself. Indeed, quantifying the difference between the true spectrum of a given potential $A(x)$ and the WKB approximation of the spectrum, in terms of motion of eigenvalues, will be necessary before it can be proven that the rigorous asymptotic analysis of SSEs is relevant to the problem of semiclassical asymptotics for the initial-value problem for the focusing nonlinear Schrödinger equation (1.1). One imagines that a study of the semiclassical limit for (1.1) should proceed by first generating from the given initial data $\psi(x, 0) = A(x)$ the corresponding well-defined SSE, and then using the fact that by combining the results of [8] with Theorem 1.1 from this paper, one has a complete picture of the limiting behavior of the SSE for an open interval of time t that is independent of $\hbar$ and includes $t = 0$, and moreover that according to Theorem 1.1 the SSE is pointwise close to the given initial data $A(x)$ for $t = 0$. The problem here is that the focusing nonlinear Schrödinger equation is known to have modulational instabilities whose exponential growth rates become arbitrarily large in the semiclassical limit. There is, therefore, the very real possibility that while the SSE is close to the initial data $A(x)$ at $t = 0$, it is not close to the corresponding solution of (1.1) for any positive t. In order to control the difference for positive time, it is necessary to know in advance how much the SSE spectral data differs from the true (unknown) spectral data, as it is the spectral data that is the starting point for analysis (cf. Riemann-Hilbert Problem 2.1). One can imagine obtaining this sort of information from the pointwise estimate given in Theorem 1.1 by Rayleigh-Schrödinger perturbation theory.

Acknowledgments

I would like to thank Jinho Baik, Thomas Kriecherbauer, and Ken McLaughlin for useful feedback. This work was supported in part by the National Science Foundation under award number DMS-0103909.

References

[1] J. Baik, T. Kriecherbauer, K. T.-R. McLaughlin, and P. D. Miller, *Strong asymptotics for discrete orthogonal polynomials*, in preparation.

[2] P. Deift, S. Kamvissis, T. Kriecherbauer, and X. Zhou, *The Toda rarefaction problem*, Comm. Pure Appl. Math. **49** (1996), no. 1, 35–83. MR 97e:34023. Zbl 0857.34025.

[3] P. Deift, T. Kriecherbauer, K. T.-R McLaughlin, S. Venakides, and X. Zhou, *Strong asymptotics of orthogonal polynomials with respect to exponential weights*, Comm. Pure Appl. Math. **52** (1999), no. 12, 1491–1552. MR 2001f:42037.

[4] P. Deift, S. Venakides, and X. Zhou, *The collisionless shock region for the long-time behavior of solutions of the KdV equation*, Comm. Pure Appl. Math. **47** (1994), no. 2, 199–206. MR 95f:35220. Zbl 0797.35143.

[5] P. Deift and X. Zhou, *A steepest descent method for oscillatory Riemann-Hilbert problems. Asymptotics for the MKdV equation*, Ann. of Math. (2) **137** (1993), no. 2, 295–368. MR 94d:35143. Zbl 0771.35042.

[6] N. Ercolani, S. Jin, C. D. Levermore, and W. MacEvoy Jr., *The zero dispersion limit of the NLS/mKdV hierarchy for the nonselfadjoint ZS operator*, preprint, 1993.

[7] A. S. Fokas, A. R. Its, and A. V. Kitaev, *The isomonodromy approach to matrix models in 2D quantum gravity*, Comm. Math. Phys. **147** (1992), no. 2, 395–430. MR 93h:81115. Zbl 0760.35051.

[8] S. Kamvissis, K. T.-R. McLaughlin, and P. D. Miller, *Semiclassical soliton ensembles for the focusing nonlinear Schrödinger equation*, submitted to Ann. Math. Stud., http://arXiv.org/abs/nlin.SI/0012034, 2000.

[9] A. B. J. Kuijlaars and E. A. Rakhmanov, *Zero distributions for discrete orthogonal polynomials*, J. Comput. Appl. Math. **99** (1998), no. 1-2, 255–274, [Corresponding corrigendum: *J. Comput. Appl. Math.* **104** (1999), 213. MR 2000c:42025b]. MR 2000c:42025a. Zbl 0929.33010.

[10] P. D. Lax and C. D. Levermore, *The small dispersion limit of the Korteweg-de Vries equation. I*, Comm. Pure Appl. Math. **36** (1983), no. 3, 253–290. MR 85g:35105a. Zbl 0532.35067.

[11] ———, *The small dispersion limit of the Korteweg-de Vries equation. II*, Comm. Pure Appl. Math. **36** (1983), no. 5, 571–593. MR 85g:35105b. Zbl 0527.35073.

[12] ———, *The small dispersion limit of the Korteweg-de Vries equation. III*, Comm. Pure Appl. Math. **36** (1983), no. 6, 809–829. MR 85g:35105c. Zbl 0527.35074.

[13] S. Venakides, *The Korteweg-de Vries equation with small dispersion: higher order Lax-Levermore theory*, Comm. Pure Appl. Math. **43** (1990), no. 3, 335–361. MR 91k:35236. Zbl 0705.35125.

[14] V. E. Zakharov and A. B. Shabat, *Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media*, Soviet Physics JETP **34** (1972), no. 1, 62–69, translated from Ž. Èksper. Teoret. Fiz. **61** (1971), no. 1, 118–134. MR 53 #9966.

Peter D. Miller: Department of Mathematics, The University of Michigan, East Hall, 525 E. University Avenue, Ann Arbor, MI 48109, USA
E-mail address: millerpd@umich.edu