

# Initial-Boundary Value Problems for the Defocusing Nonlinear Schrödinger Equation in the Semiclassical Limit

By Peter D. Miller and Zhenyun Qin

---

Initial-boundary value problems for integrable nonlinear partial differential equations have become tractable in recent years due to the development of so-called unified transform techniques. The main obstruction to applying these methods in practice is that calculation of the spectral transforms of the initial and boundary data requires knowledge of too many boundary conditions, more than are required to make the problem well-posed. The elimination of the unknown boundary values is frequently addressed in the spectral domain via the so-called global relation, and types of boundary conditions for which the global relation can be solved are called *linearizable*. For the defocusing nonlinear Schrödinger equation, the global relation is only known to be explicitly solvable in rather restrictive situations, namely homogeneous boundary conditions of Dirichlet, Neumann, and Robin (mixed) type. General nonhomogeneous boundary conditions are not known to be linearizable. In this paper, we propose an explicit approximation for the nonlinear Dirichlet-to-Neumann map supplied by the defocusing nonlinear Schrödinger equation and use it to provide approximate solutions of general nonhomogeneous boundary value problems for this equation posed as an initial-boundary value problem on the half-line. Our method sidesteps entirely the solution of the global relation. The accuracy of our method is proven in the semiclassical limit, and we provide explicit asymptotics for the solution in the interior of the quarter-plane space-time domain.

---

Address for correspondence: Peter D. Miller, Department of Mathematics, University of Michigan, East Hall, 530 Church St., Ann Arbor, MI 48109, USA; e-mail: millerpd@umich.edu

[Correction added on 2 March 2015, after first online publication: 17 February 2015: Several formatting errors have been corrected.]

DOI: 10.1111/sapm.12075

*STUDIES IN APPLIED MATHEMATICS* 134:276–362

© 2015 Wiley Periodicals, Inc., A Wiley Company

## 1. Introduction

Consider the following initial-boundary value problem for the defocusing nonlinear Schrödinger equation on the positive half-line

$$i\epsilon \frac{\partial q}{\partial t} + \epsilon^2 \frac{\partial^2 q}{\partial x^2} - 2|q|^2 q = 0, \quad q = q(x, t), \quad x > 0, \quad t > 0, \quad (1)$$

with given initial data:

$$q(x, 0) = q_0(x), \quad x > 0, \quad (2)$$

and with a given (generally nonhomogeneous) Dirichlet boundary condition at  $x = 0$ :

$$q(0, t) = Q^D(t), \quad t > 0. \quad (3)$$

Here,  $\epsilon > 0$  is an arbitrary parameter. Assuming  $q_0 \in H^2(\mathbb{R}_+)$ ,  $Q^D \in C^2(\mathbb{R}_+)$ , and that the compatibility condition  $q_0(0) = Q^D(0)$  holds, Carroll and Bu [1] have established the existence of a unique classical global solution of this problem that is a continuously differentiable map from  $t \in \mathbb{R}_+$  to  $q \in L^2(\mathbb{R}_+)$  and that is a continuous map from  $t \in \mathbb{R}_+$  to  $q \in H^2(\mathbb{R}_+)$ .

The defocusing nonlinear Schrödinger equation (1) is an integrable equation, being the compatibility condition for the existence of a simultaneous general solution  $\psi$  of the equation

$$\epsilon \frac{\partial \psi}{\partial x} = \mathbf{U}\psi, \quad \mathbf{U} := \begin{bmatrix} -ik & q \\ q^* & ik \end{bmatrix} \quad (4)$$

and also of the equation

$$\epsilon \frac{\partial \psi}{\partial t} = \mathbf{V}\psi, \quad \mathbf{V} := \begin{bmatrix} -2ik^2 - i|q|^2 & 2kq + i\epsilon q_x \\ 2kq^* - i\epsilon q_x^* & 2ik^2 + i|q|^2 \end{bmatrix}, \quad q_x := \frac{\partial q}{\partial x}. \quad (5)$$

Here,  $k$  is a complex spectral parameter, and the compatibility condition is independent of  $k$ . These two linear equations for  $\psi$  comprise the *Lax pair* for (1). One of the earliest applications of the Lax pair representation of integrable equations was the development of a transform technique based on the spectral theory of the spatial equation (4) of the Lax pair, the *inverse-scattering transform*, for solving initial-value problems posed for  $x \in \mathbb{R}$  with initial data given at  $t = 0$ ; see [2] for a pedagogical description. More recently, a *unified transform method* has been developed involving the simultaneous use of both equations of the Lax pair to study mixed initial-boundary value problems of various types. As a general reference for these methods that includes the specific details we will need in this paper, we refer to [3]; there is also a website [4] that summarizes the salient features of the technique and has links to many original references.

For the defocusing nonlinear Schrödinger equation (1) on the half-line  $x > 0$ , the unified transform method first advanced in [5] and also described in [3] amounts to the following algorithm. Recall the Pauli spin matrices

$$\sigma_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \text{and} \quad \sigma_3 := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad (6)$$

and let  $Q^N(t) := \epsilon q_x(0, t)$ . First, define the following special solutions of the Lax pair:

$$\epsilon \frac{d\mathbf{X}}{dx}(x; k) = \begin{bmatrix} -ik & q_0(x) \\ q_0(x)^* & ik \end{bmatrix} \mathbf{X}(x; k), \quad \lim_{x \rightarrow +\infty} \mathbf{X}(x; k) e^{ikx\sigma_3/\epsilon} = \mathbb{I}, \quad (7)$$

and

$$\epsilon \frac{d\mathbf{T}}{dt}(t; k) = \begin{bmatrix} -2ik^2 - i|Q^D(t)|^2 & 2kQ^D(t) + iQ^N(t) \\ 2kQ^D(t)^* - iQ^N(t)^* & 2ik^2 + i|Q^D(t)|^2 \end{bmatrix} \mathbf{T}(t; k),$$

$$\lim_{t \rightarrow +\infty} \mathbf{T}(t; k) e^{2ik^2t\sigma_3/\epsilon} = \mathbb{I}. \quad (8)$$

The spectral transforms of  $q_0$ ,  $Q^D$ , and  $Q^N$  are then given by

$$a(k) := X_{22}(0; k), \quad b(k) := X_{12}(0; k), \quad A(k) := T_{22}(0; k), \quad B(k) := T_{12}(0; k), \quad (9)$$

and by elementary symmetries one also has that

$$a(k^*)^* = X_{11}(0; k), \quad b(k^*)^* = X_{21}(0; k),$$

$$A(k^*)^* = T_{11}(0; k) \quad B(k^*)^* = T_{21}(0; k). \quad (10)$$

The second column of  $\mathbf{X}(x; k)$  is analytic and bounded in  $k$  for  $\Im\{k\} > 0$  whenever  $x \geq 0$  (and hence the same is true of  $a(k)$  and  $b(k)$ ). The second column of  $\mathbf{T}(t; k)$  is analytic and bounded in  $k$  for  $\Im\{k^2\} > 0$  whenever  $t \geq 0$  (and hence the same is true of  $A(k)$  and  $B(k)$ ).

Next, given these functions of  $k$ , one formulates a Riemann–Hilbert problem. Let  $\Sigma$  denote the contour  $\Im\{k^2\} = 0$  with each of the four half-line arcs of  $\Sigma \setminus \{0\}$  assigned an orientation such that the domain  $\Im\{k^2\} > 0$  lies on the left. On each of the four arcs we define a *jump matrix* as follows:

$$\mathbf{J}(k) := \begin{bmatrix} 1 - |\gamma(k)|^2 & \gamma(k)e^{-2i\theta(k;x,t)/\epsilon} \\ -\gamma(k)^*e^{2i\theta(k;x,t)/\epsilon} & 1 \end{bmatrix}, \quad \arg(k) = 0, \quad (11)$$

$$\mathbf{J}(k) := \begin{bmatrix} 1 & 0 \\ -\Gamma(k)e^{2i\theta(k;x,t)/\epsilon} & 1 \end{bmatrix}, \quad \arg(k) = \pi/2, \quad (12)$$

$$\begin{aligned}
 \mathbf{J}(k) &:= \begin{bmatrix} 1 & \Gamma(k^*)^* e^{-2i\theta(k;x,t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad \arg(k) = -\pi/2, \quad \text{and} \quad (13) \\
 \mathbf{J}(k) &:= \begin{bmatrix} 1 & (\Gamma(k)^* - \gamma(k))e^{-2i\theta(k;x,t)/\epsilon} \\ (\gamma(k)^* - \Gamma(k))e^{2i\theta(k;x,t)/\epsilon} & 1 - |\gamma(k)|^2 - |\Gamma(k)|^2 + \Gamma(k)\gamma(k) + \Gamma(k)^*\gamma(k)^* \end{bmatrix}, \\
 &\quad \arg(-k) = 0. \quad (14)
 \end{aligned}$$

Here, the spectral coefficients in the jump matrix are

$$\begin{aligned}
 \gamma(k) &:= \frac{b(k)}{a(k)^*} \quad \text{and} \quad \Gamma(k) := \frac{B(k^*)^*}{a(k)d(k)}, \\
 \text{where} \quad d(k) &:= a(k)A(k^*)^* - b(k)B(k^*)^*, \quad (15)
 \end{aligned}$$

and all of the dependence on  $x$  and  $t$  appears explicitly through the function

$$\theta(k; x, t) := kx + 2k^2t. \quad (16)$$

The Riemann–Hilbert problem is then the following.

**Riemann–Hilbert problem 1.** Find a  $2 \times 2$  matrix  $\mathbf{M}(k)$  with the following properties:

**Analyticity:**  $\mathbf{M}$  is analytic and uniformly bounded for  $k \in \mathbb{C} \setminus \Sigma$ , taking boundary values  $\mathbf{M}_\pm(k)$  on each of the four rays of  $\Sigma$  from the domain where  $\pm\Im\{k^2\} > 0$ .

**Jump Condition:** The boundary values are related on each ray of  $\Sigma$  by the jump condition

$$\mathbf{M}_+(k) = \mathbf{M}_-(k)\mathbf{J}(k), \quad k \in \Sigma \setminus \{0\}, \quad (17)$$

where  $\mathbf{J} : \Sigma \setminus \{0\} \rightarrow \text{SL}(2, \mathbb{C})$  is defined by (11)–(14).

**Normalization:**  $\mathbf{M}(k) \rightarrow \mathbb{I}$  as  $k \rightarrow \infty$ .

From the solution of Riemann–Hilbert problem 1, which depends parametrically on  $x, t$ , and  $\epsilon$ , one obtains a solution of the defocusing nonlinear Schrödinger equation by taking the limit

$$q(x, t) = 2i \lim_{k \rightarrow \infty} kM_{12}(k). \quad (18)$$

This procedure is derived assuming the existence of a solution  $q(x, t)$  satisfying the initial and boundary conditions in addition to some other technical assumptions. It produces the solution to the initial-boundary value problem under two conditions:

- The given boundary data  $(Q^D, Q^N)$  used to compute the spectral transforms from (7)–(10) are consistent. That is,  $Q^N(t)$  must agree with  $(\epsilon$  times) the

Neumann boundary value of the solution of the Dirichlet problem whose well-posedness was established by Carroll and Bu [1].

- The function  $d(k)$  must have no zeros in the closed second quadrant of the complex  $k$ -plane. This is a technical condition as otherwise Riemann–Hilbert Problem 1 must be formulated differently to allow  $\mathbf{M}(k)$  to have poles at these points and their complex conjugates, with prescribed residue relations. It is conjectured that in fact  $d(k)$  is nonvanishing for consistent boundary data, but to our knowledge there is no proof<sup>1</sup> of this in the literature.

Of course the problem is that if the boundary data functions  $Q^D$  and  $Q^N$  are both independently specified as is required to calculate the spectral transforms and hence the jump matrices, then the initial-boundary value problem is overdetermined and the solution of the equation produced by the method cannot generally satisfy the initial and boundary conditions (although it will solve the differential equation in the interior of the domain). On the other hand, the procedure is sadly incomplete if only the Dirichlet data (the function  $Q^D$ ) is specified in which case the jump matrices (11)–(14) are indeterminate as  $\Gamma(k)$  cannot be calculated at all from (8)–(10) and (15).

A central role in the unified transform theory is therefore played by the *global relation*, an identity satisfied by the spectral transforms of consistent boundary data ( $Q^D$ ,  $Q^N$ ) that encodes in the transform domain the Dirichlet-to-Neumann map giving  $\epsilon_{q_x}(0, t)$  in terms of  $q(0, t)$  and  $q(x, 0)$ . Under certain conditions on the Dirichlet data, the global relation can be effectively solved, and hence the unknown Neumann data is eliminated. The class of boundary conditions for which the global relation can be solved by symmetries in the complex  $k$ -plane is called the class of *linearizable* boundary conditions. Unfortunately, the only type of Dirichlet boundary condition known to be linearizable is the homogeneous boundary condition  $Q^D(t) \equiv 0$ . Of course, this special case can also be handled via the usual inverse scattering transform on the whole line  $x \in \mathbb{R}$  simply by extending the initial data  $q_0(x)$  to  $x < 0$  as an odd function.

Another approach to general nonhomogeneous Dirichlet boundary conditions that avoids the global relation entirely may be based on the observation that under mild conditions, given the spectral transforms  $\{a(k), b(k), A(k), B(k)\}$ , Riemann–Hilbert problem 1 has a unique solution for almost all  $(x, t) \in \mathbb{R}^2$  by analytic Fredholm theory (see proposition 4.3 of [7]) combined with steepest descent asymptotics for large  $x$  and  $t$ . The exceptional set is the zero locus of an entire scalar function of  $(x, t) \in \mathbb{C}^2$  that does not vanish identically, i.e., a complex curve in  $\mathbb{C}^2$  that may or may not have real points but that would be at worst a closed and nowhere-dense union of analytic arcs in the real  $(x, t)$ -plane. This in turn implies (by the standard arguments of the *dressing*

---

<sup>1</sup>After this paper was accepted for publication, a preprint [6] was made public that evidently contains a proof of this conjecture.

*method*, see also the proof of Proposition 3 below) that for those  $(x, t)$  for which a solution exists the function  $q(x, t)$  produced by taking the limit (18) is necessarily *some* solution of the defocusing nonlinear Schrödinger equation (1). The question is whether this solution satisfies also the (three in total) initial and boundary conditions that were used to generate the spectral transforms  $\{a(k), b(k), A(k), B(k)\}$  in the first place. This line of reasoning suggests an iteration procedure for solving the Dirichlet initial-boundary value problem for the defocusing nonlinear Schrödinger equation on the half-line with general nonhomogeneous data: begin by making an initial guess for the (unknown) Neumann boundary data, say  $Q_0^N(t)$  for  $t > 0$ . Set  $n = 0$  and then:

1. Take  $Q^N(t) = Q_n^N(t)$ , and with the given Dirichlet data  $Q^D(t)$  and  $q_0(x)$  calculate the spectral transforms  $\{a(k), b(k), A(k), B(k)\} = \{a(k), b(k), A_n(k), B_n(k)\}$  from (7)–(10).
2. Formulate Riemann–Hilbert problem 1 with these spectral transforms and solve it. Denote the function obtained from (18) as  $q_n(x, t)$ . It is necessarily a solution of the defocusing nonlinear Schrödinger equation (1) on the quarter plane  $x > 0$  and  $t > 0$ .
3. Define  $Q_{n+1}^N(t) := \epsilon \partial_x q_n(0, t)$  for  $t > 0$ .
4. Set  $n := n + 1$  and go to step 1.

No doubt the reader can imagine various other iterative approaches like this one. It is not the purpose of this paper to study the convergence of this algorithm, but in the spirit of the principle that finite truncations of a convergent iteration (or infinite series) can often provide accuracy in various asymptotic limits, we wish to explore the possibility of using *just one iteration* of the algorithm (actually a slightly modified version of the first iteration, see Section 3 for details) to provide an asymptotic approximation of the solution of the Dirichlet initial-boundary value problem in the *semiclassical limit*  $\epsilon \downarrow 0$ . The key to the success of this procedure is to make a very good initial guess for the unknown Neumann data, one that is asymptotically accurate in the semiclassical limit (as we will rigorously prove after the fact, see Theorem 2). That is, what we need is an explicit approximation of the Dirichlet-to-Neumann map for (1).

### 1.1. The semiclassical Dirichlet-to-Neumann map

Let us now explain the approximation of the Dirichlet-to-Neumann map for the defocusing nonlinear Schrödinger equation (1) that we plan to study in this article. Without loss of generality, we represent the complex field  $q(x, t)$  in real phase-amplitude form:

$$q(x, t) = \eta(x, t)e^{i\sigma(x, t)/\epsilon}, \quad \eta(x, t) := |q(x, t)|. \quad (19)$$

Substituting into (1), dividing by the common factor  $e^{i\sigma(x,t)/\epsilon}$ , and separating real and imaginary parts yields the following system of equations:

$$\begin{aligned} \frac{\partial \eta}{\partial t} + 2 \frac{\partial \sigma}{\partial x} \frac{\partial \eta}{\partial x} + \eta \frac{\partial^2 \sigma}{\partial x^2} &= 0 \\ \frac{\partial \sigma}{\partial t} + \left( \frac{\partial \sigma}{\partial x} \right)^2 + 2\eta^2 &= \frac{\epsilon^2}{\eta} \frac{\partial^2 \eta}{\partial x^2}. \end{aligned} \quad (20)$$

This coupled system is equivalent to (1). It is useful to introduce notation for the phase gradient:

$$u(x, t) := \frac{\partial \sigma}{\partial x}(x, t). \quad (21)$$

Now, in terms of  $\eta$  and  $u$ , the exact ratio between the unknown Neumann data and the given Dirichlet data at  $x = 0$  takes the form

$$\begin{aligned} -i \frac{Q^N(t)}{Q^D(t)} &= \frac{-i\epsilon}{q(0, t)} \frac{\partial q}{\partial x}(0, t) \\ &= u(0, t) - \frac{i\epsilon}{\eta(0, t)} \frac{\partial \eta}{\partial x}(0, t). \end{aligned} \quad (22)$$

Consider the second equation of the system (20) at  $x = 0$  along with (22) in the formal semiclassical limit  $\epsilon \rightarrow 0$ , assuming that  $\eta(0, t) \neq 0$ . This means that we simply neglect the terms explicitly proportional to  $\epsilon$  or  $\epsilon^2$  in each case, yielding the formal approximations:

$$\frac{\partial \sigma}{\partial t}(0, t) + u(0, t)^2 + 2\eta(0, t)^2 \approx 0 \quad \text{and} \quad -i \frac{Q^N(t)}{Q^D(t)} \approx u(0, t). \quad (23)$$

Our approach is to assume that the known Dirichlet boundary data is specified in the form

$$Q^D(t) := H(t)e^{iS(t)/\epsilon} \quad (24)$$

where  $H(\cdot) > 0$  and  $S(\cdot)$  are given real-valued functions independent of  $\epsilon$ . Obviously we then have  $\eta(0, t) = H(t)$  and  $\sigma_t(0, t) = S'(t)$ , so we may rewrite the approximate relations (23) as

$$S'(t) + u(0, t)^2 + 2H(t)^2 \approx 0 \quad \text{and} \quad -i \frac{Q^N(t)}{Q^D(t)} \approx u(0, t). \quad (25)$$

Assuming further that

$$S'(t) < -2H(t)^2, \quad t > 0, \quad (26)$$

we solve the first of these relations for the unknown phase derivative  $u(0, t)$  at the boundary:

$$u(0, t) \approx U(t) := \sqrt{-S'(t) - 2H(t)^2} > 0, \quad (27)$$

that is,  $U(t)$  is the (real valued) formal semiclassical approximation of the exact phase derivative  $u(0, t)$ . Finally, for Dirichlet boundary data (24) satisfying the condition (26) we use the second equation of (25) to approximate the Dirichlet-to-Neumann map as follows.

DEFINITION 1. *Suppose that the Dirichlet boundary data  $Q^D(t)$  of the form (24) satisfies  $S'(t) + 2H(t)^2 < 0$ . The semiclassical approximation of the Dirichlet-to-Neumann map is defined by*

$$Q_0^N(t) := iU(t)Q^D(t), \quad t > 0, \quad (28)$$

where  $U(t)$  is given in terms of the phase and amplitude of the known Dirichlet data by (27).

The key point of our approach is that by neglecting the formally small dispersive terms in (20) we obtain a system that is first-order in  $x$  and hence allows the unknown Neumann data to be *explicitly eliminated* in favor of  $t$ -derivatives that may be computed along the boundary from the given Dirichlet data. This approximation is a purely local relation between the two boundary values, and in particular the approximate Dirichlet-to-Neumann map is independent of initial data  $q_0$ .

We have selected the positive square root in (27) for a specific reason, which we now explain. Differentiation of the second equation of (20) with respect to  $x$  produces the equivalent system

$$\frac{\partial}{\partial t} \begin{bmatrix} \eta \\ u \end{bmatrix} + \mathbf{C}(\eta, u) \frac{\partial}{\partial x} \begin{bmatrix} \eta \\ u \end{bmatrix} = \epsilon^2 \frac{\partial}{\partial x} \begin{bmatrix} 0 \\ \eta^{-1} \eta_{xx} \end{bmatrix}, \quad \mathbf{C}(\eta, u) := \begin{bmatrix} 2u & \eta \\ 4\eta & 2u \end{bmatrix}. \quad (29)$$

Obviously, (29) is a formally small perturbation of a quasilinear system obtained by simply setting  $\epsilon$  to zero. The *characteristic velocities* of the limiting system are the eigenvalues  $c(\eta, u)$  of the coefficient matrix  $\mathbf{C}(\eta, u)$ :

$$c(\eta, u) := 2u \pm 2\eta. \quad (30)$$

As the characteristic velocities are real and distinct (for  $\eta \neq 0$ ), the limiting quasilinear system is of hyperbolic type. Causality and local well-posedness for the Cauchy problem of the hyperbolic approximating system in the quarter plane  $x > 0$  and  $t > 0$  requires that the boundary  $x = 0$  be a space-like curve. In other words, we require *both* characteristic velocities to be strictly positive at the boundary. This means that we will require that  $U(t) > H(t)$  for all  $t > 0$ . Since  $H(t) \geq 0$  for  $t > 0$  it is clear that well-posedness of the limiting hyperbolic boundary-value problem requires in particular  $U(t) > 0$ . In fact, we will ensure the condition  $U(t) > H(t)$  by imposing the stronger condition  $U(t) > 2H(t)$ ; the latter condition appears to be necessary to recover the Dirichlet boundary data at  $x = 0$  for all  $t > 0$  (see Remark 2 below).



### 1.2. Outline of the paper and description of main results

For convenience we restrict our attention to the already nontrivial and physically interesting case of zero initial data:  $q_0(x) = 0$  for all  $x > 0$ . However, we fix rather general nonhomogeneous Dirichlet boundary data of the form (24) for  $t > 0$  (and satisfying several additional conditions allowing our procedure to succeed, see Assumption 1 below), and attempt to solve the corresponding initial-boundary value problem. The first step is the calculation of the spectral transforms  $A_0(k)$  and  $B_0(k)$  corresponding to the Dirichlet data  $Q^D(t)$  given by (24) and the formally approximate Neumann data  $Q_0^N(t)$  given by Definition 1. The direct spectral analysis is made possible in practice because the parameter  $\epsilon > 0$  is presumed small, so the equations of the Lax pair become singularly perturbed differential equations that may be studied by classical methods. The results of this analysis are summarized in Section 2.2, with the corresponding proofs appearing in two appendices. Despite the rigor of these results, there are certain difficulties that remain with directly formulating the inverse problem for the exact scattering data, so rather than calculate the solution of the defocusing nonlinear Schrödinger equation corresponding to the exact spectral transforms of the (generally incompatible) Dirichlet-Neumann pair  $(Q^D, Q_0^N)$  we modify the spectral functions in an ad-hoc fashion, but one inspired by the rigorous direct spectral analysis of the temporal problem of the Lax pair. This allows us to formulate a simpler and completely explicit version of Riemann–Hilbert problem 1 for a matrix  $\tilde{\mathbf{M}}(k)$ ; see Section 3.

The simpler Riemann–Hilbert problem explicitly encodes the given Dirichlet boundary data (24) through two integral transforms denoted  $\tau$  and  $\Phi$  (these are really the semiclassical analogues of the amplitude and phase of the spectral function  $\Gamma$ ; see (56)–(57)), and its solution produces, for each  $\epsilon > 0$ , a solution  $q = \tilde{q}^\epsilon(x, t)$  of the defocusing nonlinear Schrödinger equation (1). The rest of the paper is concerned with analyzing this solution, paying particular attention to the semiclassical asymptotic behavior of  $\tilde{q}^\epsilon(x, t)$  at the initial time  $t = 0$  for  $x > 0$  and at the boundary  $x = 0$  for  $t > 0$ . Our first result is the following.

**THEOREM 1** (Approximation of the initial condition). *The solution  $q = \tilde{q}^\epsilon(x, t)$  of the defocusing nonlinear Schrödinger equation (1) obtained from Riemann–Hilbert problem 2 satisfies*

$$\tilde{q}^\epsilon(x, 0) = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2}) \quad x > 0, \quad (31)$$

where the error term is uniform on  $x \geq x_0$  for each  $x_0 > 0$ .

Thus, the function  $\tilde{q}^\epsilon(x, t)$  nearly satisfies the given homogeneous initial condition  $q_0(x) = 0$  for  $x > 0$ . Our rigorous proof of this result is based on the steepest descent method for Riemann–Hilbert problems, combined with a natural generalization of the method involving  $\bar{\partial}$ -problems [8]. After

establishing some preliminary results in Section 4.1.1, we give the proof of Theorem 1 in Section 4.1.2.

Our next main result is the following. The points  $t_a$  and  $t_b$  are defined as part of Assumption 1 below.

**THEOREM 2** (Approximation of boundary conditions). *Suppose that  $t > 0$  and  $t \neq t_a$ ,  $t \neq t_b$ . The solution  $q = \tilde{q}^\epsilon(x, t)$  of the defocusing nonlinear Schrödinger equation (1) obtained from Riemann–Hilbert problem 2 satisfies*

$$\tilde{q}^\epsilon(0, t) = H(t)e^{iS(t)/\epsilon} + \mathcal{O}((\log(\epsilon^{-1}))^{-1/2}), \tag{32}$$

and

$$\epsilon \tilde{q}_x^\epsilon(0, t) = iU(t)H(t)e^{iS(t)/\epsilon} + \mathcal{O}((\log(\epsilon^{-1}))^{-1/2}), \tag{33}$$

where the error terms are uniform for  $t$  in compact subintervals of  $(0, +\infty) \setminus \{t_a, t_b\}$ .

Equation (32) shows that the same solution  $\tilde{q}^\epsilon(x, t)$  very nearly satisfies the given nonhomogeneous Dirichlet boundary condition (24) at  $x = 0$  for  $t > 0$ . Moreover, from (33) we see directly that the true Neumann data at the boundary is indeed asymptotically consistent with the formal approximation given by Definition 1. The proof of this result is again based on the steepest descent method, this time augmented with the use of a complex phase function  $g$ . After describing the general methodology and constructing the function  $g$  in Section 4.2.1, we present the proof of Theorem 2 in Section 4.2.2.

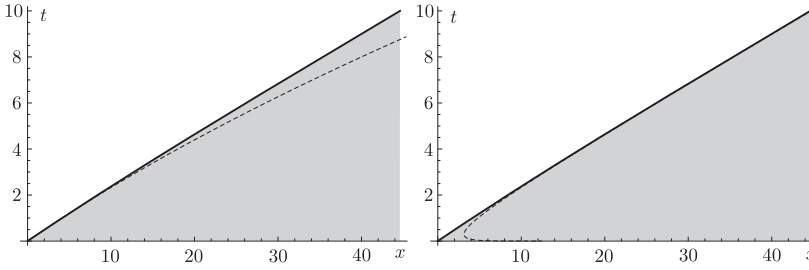
The main point, however, is that the solution  $\tilde{q}^\epsilon(x, t)$  is represented also for  $(x, t)$  not on the boundary of the quarter plane  $x > 0, t > 0$  via exactly the same Riemann–Hilbert problem (see Riemann–Hilbert problem 2). This means that one may use steepest descent methods to calculate  $\tilde{q}^\epsilon(x, t)$  for small  $\epsilon > 0$  for positive  $t$  and away from the boundary. For  $(x, t)$  close to the boundary of the quarter plane the analysis is virtually the same as it is exactly on the boundary, with similar results. For example, a corollary of the proof of Theorem 1 is the following. The points  $k_a$  and  $k_b$  are specified in terms of the functions characterized by Assumption 1 below, and  $\Phi$  is explicitly given by (57).

**COROLLARY 1** (Existence of a vacuum domain). *Let  $t \geq 0$ , and let  $X(t)$  be the smallest nonnegative value of  $x_0$  for which the inequality  $x + 4tk - \Phi'(k) \geq 0$  holds for all  $k \in (k_a, k_b)$  whenever  $x \geq x_0$ . Then the solution  $q = \tilde{q}^\epsilon(x, t)$  of the defocusing nonlinear Schrödinger equation (1) obtained from Riemann–Hilbert Problem 2 satisfies  $\tilde{q}^\epsilon(x, t) = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  as  $\epsilon \downarrow 0$  whenever  $x > X(t)$ .*

In the case that  $f(\cdot) := -\Phi'(\cdot)$  is convex, we may characterize  $X(t)$  explicitly in terms of the Legendre dual  $f^*$  as follows:

$$X(t) := f^*(-4t) = [-\Phi']^*(-4t), \quad t > 0, \tag{34}$$

where  $f^*(p) := \sup_{k_a < k < k_b} (pk - f(k)).$



**Figure 1.** The curve  $x = X(t)$  (solid black curve) and the  $\epsilon$ -independent vacuum domain (shaded) calculated from  $\Phi$  corresponding to the explicit boundary data illustrated in Figure 2 below. Also shown (dashed curves) are explicit asymptotes to  $x = X(t)$  for small  $t$  (left panel, the asymptote  $x = X_0(t)$ ) and large  $t$  (right panel, the asymptote  $x = X_\infty(t)$ ). These asymptotes are described in Section 4.1.3 (see (126) and (130)).

The proof of Corollary 1 is given in Section 4.1.3. We call the domain  $x > X(t)$  the *vacuum domain* corresponding to the Dirichlet boundary data  $H(t)e^{iS(t)/\epsilon}$ . In the vacuum domain the solution is influenced predominantly by the homogeneous initial data rather than the nonhomogeneous boundary data in the semiclassical limit. A concrete calculation of the vacuum domain for a particular choice of Dirichlet boundary data is shown in Figure 1.

Another result is the following, which is essentially a corollary of the proof of Theorem 2. Here,  $\mathfrak{a}(\cdot)$  and  $\mathfrak{b}(\cdot)$  are defined in terms of the Dirichlet data by (40).

**COROLLARY 2** (Existence of a plane-wave domain). *Each point  $(0, t_0)$  with  $t_0 \in (0, +\infty) \setminus \{t_a, t_b\}$  has a neighborhood  $D_{t_0}$  in the  $(x, t)$ -plane in which there exist unique differentiable functions  $\alpha = \alpha(x, t)$  and  $\beta = \beta(x, t)$  satisfying  $\alpha(0, t) = \mathfrak{a}(t)$  and  $\beta(0, t) = \mathfrak{b}(t)$  and the partial differential equations*

$$\begin{aligned} \frac{\partial \alpha}{\partial t} - (3\alpha + \beta) \frac{\partial \alpha}{\partial x} &= 0 \\ \frac{\partial \beta}{\partial t} - (\alpha + 3\beta) \frac{\partial \beta}{\partial x} &= 0, \end{aligned} \tag{35}$$

and such that the solution  $q = \tilde{q}^\epsilon(x, t)$  of the defocusing nonlinear Schrödinger equation (1) obtained from Riemann–Hilbert problem 2 satisfies

$$\tilde{q}^\epsilon(x, t) = \eta(x, t)e^{i\sigma(x, t)/\epsilon} + \mathcal{O}((\log(\epsilon^{-1}))^{-1/2}) \tag{36}$$

uniformly for  $(x, t) \in D_{t_0}$  as  $\epsilon \downarrow 0$ , where

$$\begin{aligned} \eta(x, t) &:= \frac{1}{2}(\beta(x, t) - \alpha(x, t)) \quad \text{and} \\ u(x, t) &:= \frac{\partial \sigma}{\partial x}(x, t) = -(\alpha(x, t) + \beta(x, t)) \end{aligned} \tag{37}$$

and  $\sigma(0, t) = S(t)$ .

The proof of Corollary 2 is given in Section 4.2.3. Note that eliminating  $\alpha$  and  $\beta$  from (35) in favor of  $\eta$  and  $u$  using (37) yields

$$\frac{\partial}{\partial t} \begin{bmatrix} \eta \\ u \end{bmatrix} + \mathbf{C}(\eta, u) \frac{\partial}{\partial x} \begin{bmatrix} \eta \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{C}(\eta, u) := \begin{bmatrix} 2u & \eta \\ 4\eta & 2u \end{bmatrix}, \quad (38)$$

which should be compared with (29), the defocusing nonlinear Schrödinger equation written without approximation in terms of amplitude  $\eta$  and phase derivative  $u$ . Therefore, we observe that for small positive  $x$ ,  $\tilde{q}^\epsilon(x, t)$  resembles a modulated plane wave of the form (19) for amplitude  $\eta$  and phase  $\sigma$  independent of  $\epsilon$ , and the modulation is described by the dispersionless nonlinear Schrödinger system (38), or equivalently the Whitham (Riemann–invariant form) system (35). This shows consistency with, and adds yet more weight to, our approximate formula for the Dirichlet-to-Neumann map given in Definition 1. Indeed, the latter was formally derived under the *initially unjustified assumption* that the solution resembles a modulated plane wave near the boundary  $x = 0$ .

We call the union of the neighborhoods of the  $(x, t)$  plane for  $x > 0$  in which  $\tilde{q}^\epsilon(x, t)$  is described by Corollary 2 the *plane-wave domain* for  $\tilde{q}^\epsilon(x, t)$ . We therefore see that the quarter-plane  $x > 0$  and  $t > 0$  is split up into several regions in which the approximate solution  $\tilde{q}^\epsilon(x, t)$  of the Dirichlet boundary-value problem<sup>2</sup> behaves quite differently. So far we have observed the vacuum domain, in which  $\tilde{q}^\epsilon(x, t)$  simply decays to zero with  $\epsilon > 0$ , and the plane-wave domain, in which  $\tilde{q}^\epsilon(x, t)$  resembles a modulated plane wave with nonzero amplitude. It is to be expected that these two domains do not exhaust the quarter plane. While we do not pursue the topic further in this paper, the methodology presented in Section 4.2.1 below also allows one to calculate the semiclassical behavior of  $\tilde{q}^\epsilon(x, t)$  for  $(x, t)$  in domains not contiguous to the boundary of the quarter plane, in which (in principle) more complicated local behavior of  $\tilde{q}^\epsilon(x, t)$  can occur, with microstructure modeled by higher transcendental functions (e.g., dispersive shock waves described by modulated elliptic functions). See Remark 5 for more information.

With the proofs of our results complete, we conclude the body of our paper with some further remarks, some indicating directions for future work, in Section 5.

### 1.3. Related work

Our paper represents a further contribution to the literature on the use of the unified transform to study nonlinearizable boundary value problems in various asymptotic limits. A key observation that was made fairly early in the development of the theory was that regardless of whether the spectral functions  $A(k)$  and  $B(k)$  actually correspond to a compatible Dirichlet-Neumann pair  $(Q^D, Q^N)$ , Riemann–Hilbert problem 1 yields to asymptotic analysis in the

<sup>2</sup>We wish to stress that while  $q = \tilde{q}^\epsilon(x, t)$  only approximately satisfies the given initial and boundary conditions, it is an exact solution of the defocusing nonlinear Schrödinger equation (1) for every  $\epsilon > 0$ .

limit of large  $t$  (with  $x = vt$  for some nonnegative velocity  $v$ ) by the steepest descent method. General asymptotic properties of the solution that can be observed by such analysis therefore necessarily also describe the physical solutions of the Dirichlet problem that simply correspond to the special case that the spectral functions satisfy the global relation. As a representative of this type of analysis (for the focusing case of the nonlinear Schrödinger equation), we cite a paper of Boutet de Monvel, Its, and Kotlyarov [9], where time-periodic boundary conditions are analyzed. Such analysis does not require any preliminary asymptotic analysis of the spectral functions, as they are independent of the asymptotic parameter  $t$ . Another approach to the asymptotic solution of nonlinearizable boundary value problems is to consider the situation in which the initial and boundary data are small, in which case a perturbation scheme based on the amplitude as a small parameter can be developed in detail, and significantly this allows the global relation to be solved order-by-order. This means that the asymptotic results obtained are guaranteed to correspond to a compatible Dirichlet-Neumann pair  $(Q^D, Q^N)$  even though only  $Q^D$  is given. The recent papers of Fokas and Lenells [10], [11] pursue this approach and obtain new convergence results showing that for small simple harmonic Dirichlet boundary data, the solution is eventually periodic with the same period, at least to third order in the small amplitude.

The semiclassical limit is, in a sense, the exact opposite to the weakly-nonlinear small-amplitude limit. Indeed, the formal semiclassical limit is given by the strongly nonlinear hyperbolic system (38). There is at least one other paper in the literature on the subject of semiclassical analysis of the Dirichlet initial-boundary-value problem on the half-line for the defocusing nonlinear Schrödinger equation, namely a paper of Kamvissis [12], which directly stimulated our interest in this problem. Like we do, Kamvissis considers general nonhomogeneous Dirichlet boundary data together with homogeneous initial conditions, and he applies the steepest descent methodology for Riemann–Hilbert problems to deduce general properties of the solution in the semiclassical limit. Our Corollary 2 is consistent with theorem 5 of [12] (the main result of that paper) albeit in the simplest case of genus  $N = 0$ . On the other hand, it is less clear whether the vacuum domain  $x > X(t)$  described by our Corollary 1 is a special case of Kamvissis’ theorem 5.

While we study the same problem, and apply similar methods, the approach in [12] is fundamentally different from ours, being based solely on the abstract existence result for the unknown Neumann data  $Q^N(t)$  corresponding to the given Dirichlet data  $Q^D(t)$ . While Kamvissis’ assumption that  $Q^D(t)$  is independent of  $\epsilon$  is quite reasonable and physically interesting,<sup>3</sup> his subsequent analysis of the direct spectral problem for the  $t$ -part of the Lax pair (theorems 2

---

<sup>3</sup>In the setting of (24), Dirichlet boundary data that is independent of  $\epsilon$  corresponds to taking  $S(t) \equiv 0$ . Therefore, in a sense our results cannot be compared well with those of [12], because we require  $S'(t)$  to be strictly negative (see (26)).

and 3 of [12], of which our Propositions 1 and 2 are analogues) apparently rests upon the additional hidden assumption that the implicitly defined function  $q_x(0, t)$  is *also* independent of  $\epsilon$ ; otherwise the WKB methodology cited in Section III of [12] does not apply. Since the defocusing nonlinear Schrödinger equation involves the parameter  $\epsilon$  in a singular way, whether this assumption is justified is certainly not obvious. Indeed one might worry that a slowly varying Dirichlet boundary condition might give rise to a Neumann boundary value with rapid variations in amplitude or phase of period proportional to  $\epsilon$ . For example, our Theorem 2 shows that some bounded Dirichlet data  $q(0, t)$  can lead to rapidly oscillatory Neumann data  $q_x(0, t)$  that moreover is large of size  $\epsilon^{-1}$ .

Our approach is to avoid abstract assumptions, and instead make a very explicit assumption, based on the modulated plane-wave ansatz, concerning the unknown Neumann data as described in Definition 1. This allows us to justify our steepest descent analysis by ultimately tying the solution generated back to the hypothesized initial and boundary data (Theorems 1 and 2) in an explicit fashion. This same approach leads to a very concrete description of the semiclassical dynamics of the solution in the full domain  $x > 0$  and  $t > 0$ , as in the characterization of the vacuum domain presented in Corollary 1.

Another paper that we wish to mention is work of Degasperis, Manakov, and Santini [13] that presents an alternate approach to the general initial-boundary value problem for the defocusing nonlinear Schrödinger equation. The method described in [13] avoids using the  $t$ -part of the Lax pair to formulate the inverse problem and instead uses the inverse theory of the spatial part of the Lax pair only, at the cost of a more implicit nonlinear description of the time evolution of the jump matrices on the real line. The fact that the inverse problem is ultimately formulated as a Riemann–Hilbert problem relative to the real axis may be a crucial benefit in our view (see Remark 6). In the future, we plan to explore the possibility of using semiclassical asymptotic techniques to analyze this alternate method of studying initial-boundary value problems.

## 2. A class of Dirichlet boundary-value problems

### 2.1. Characterization of the boundary data

For simplicity,<sup>4</sup> we consider the case of vanishing initial data:

$$q_0(x) = 0, \quad x > 0. \quad (39)$$

This immediately implies that the spectral transforms defined from the differential equation (7) satisfy  $a(k) \equiv 1$  and  $b(k) \equiv 0$  for all  $k \in \mathbb{C}$ . We take

---

<sup>4</sup>See Remark 9.

the Dirichlet boundary data in the form (24), and for convenience we impose several conditions on the functions  $H(t)$  and  $S(t)$  for  $t \geq 0$ . These are specified in terms of an auxiliary function  $U$  as follows:

ASSUMPTION 1. *The functions  $H : \mathbb{R}_+ \rightarrow \mathbb{R}$  and  $U : \mathbb{R}_+ \rightarrow \mathbb{R}$  satisfy the following conditions:*

- $H : \mathbb{R}_+ \rightarrow \mathbb{R}$  is real analytic for  $t > 0$ , strictly positive for all  $t > 0$ , and  $t^p H^{(q)}(t) \rightarrow 0$  as  $t \rightarrow +\infty$  for all  $p \geq 0$  and  $q = 0, 1, 2, \dots$ . Also, there is a positive number  $h_0$  such that  $H(t) = h_0 t^{1/2}(1 + o(1))$  and  $H'(t) = \frac{1}{2} h_0 t^{-1/2}(1 + o(1))$  hold as  $t \rightarrow 0$  with  $\Re\{t\} \geq 0$ .
- $U : \mathbb{R}_+ \rightarrow \mathbb{R}$  is real analytic for  $t > 0$ , satisfying  $U(t) \geq 2H(t) + \delta$  for some  $\delta > 0$ , and  $t^p U^{(q)}(t) \rightarrow 0$  as  $t \rightarrow +\infty$  for all  $p \geq 0$  and  $q = 1, 2, 3, \dots$ . Also, there is a positive number  $U_0$  such that  $U(t) = U_0 + o(t^{1/2})$  and  $U'(t) = \mathcal{O}(t^{-1/2})$  hold as  $t \rightarrow 0$  with  $\Re\{t\} \geq 0$ .
- *The functions*

$$a(t) := -\frac{1}{2}U(t) - H(t) \quad \text{and} \quad b(t) := -\frac{1}{2}U(t) + H(t) \quad (40)$$

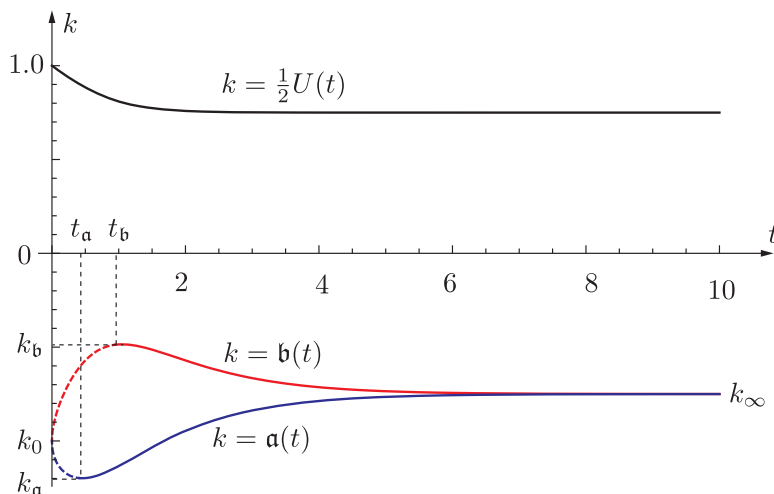
each have precisely one critical point in  $(0, \infty)$ , corresponding to a nondegenerate maximum for  $b$  at a point  $t = t_b$  and a nondegenerate minimum for  $a$  at a point  $t = t_a$ . Nondegeneracy means that  $a''(t_a) > 0$  and  $b''(t_b) < 0$ .

*Remark 1.* The square-root behavior of the amplitude  $H(t)$  that is specified in Assumption 1 evidently violates the conditions for the proof of Carroll and Bu [1] to guarantee the existence of a solution of the initial-boundary value problem. Nonetheless this behavior leads to additional smoothness of the integral transform  $\Phi$  defined in (57) below that is useful in the proof of Theorem 1. See Remark 4.

We may avoid this difficulty as follows. Let  $\mathcal{B}(t)$  be a  $C^\infty(\mathbb{R}_+)$  “bump function” satisfying  $\mathcal{B}(t) = 0$  for  $0 < t < 1$  and  $\mathcal{B}(t) = 1$  for  $t > 2$ . Replacing  $H(t)$  by  $H^\epsilon(t) := \mathcal{B}(\epsilon^{-1}t)H(t)$ , by [1] there is a unique solution  $q = q^\epsilon(x, t)$  of (1) for each  $\epsilon > 0$  satisfying  $q^\epsilon(x, 0) = 0$  for  $x > 0$  and  $q^\epsilon(0, t) = H^\epsilon(t)e^{iS(t)/\epsilon}$  for  $t > 0$ . We may view our results as a comparison between  $\tilde{q}^\epsilon(x, t)$  and the function  $q^\epsilon(x, t)$ , the latter of which exactly satisfies the given boundary condition (24) for every  $t > 0$  as long as  $\epsilon > 0$  is sufficiently small (given  $t$ ).

We now use (27) to define  $S(t)$  in terms of functions  $H$  and  $U$  satisfying the conditions of Assumption 1 as

$$S(t) := S(0) - \int_0^t [U(s)^2 + 2H(s)^2] ds. \quad (41)$$



**Figure 2.** The turning point curve consists of three branches, curves along which  $\lambda^2 = 0$  (see (50)). Here, the turning point curve is shown for boundary data  $H(t) := \frac{1}{2}t^{1/2} \operatorname{sech}(t)$  and  $U(t) := 2 - \frac{1}{2} \tanh(t)$  consistent with Assumption 1. The functions  $\mathfrak{a}(\cdot)$  and  $\mathfrak{b}(\cdot)$  are given in terms of  $H(\cdot)$  and  $U(\cdot)$  by (40). The dashed (solid) curves in the interval  $k_a < k < k_b$  correspond to the function  $t = t_-(k)$  (the function  $t = t_+(k)$ ).

Note that as  $U$  is real, the inequality (26) is automatically satisfied. We introduce the following notation:

$$k_0 := -\frac{1}{2}U_0, \quad U_\infty := \lim_{t \rightarrow +\infty} U(t), \quad \text{and} \quad k_\infty := -\frac{1}{2}U_\infty \quad (42)$$

( $U_\infty$  is well-defined as  $U'(\cdot) \in L^1(\mathbb{R}_+)$ ), and we set  $k_a := \mathfrak{a}(t_a)$  and  $k_b := \mathfrak{b}(t_b)$ . Note that the assumption  $H(t) > 0$  guarantees that  $\mathfrak{a}(t) < \mathfrak{b}(t)$ , and the assumption that  $U(t) \geq 2H(t) + \delta$  guarantees that  $k_a < k_b < 0$ . The points  $k_0$  and  $k_\infty$  lie in the interval  $(k_a, k_b)$ , and we see that  $\mathfrak{a}(0) = \mathfrak{b}(0) = k_0$  while  $\lim_{t \rightarrow +\infty} \mathfrak{a}(t) = \lim_{t \rightarrow +\infty} \mathfrak{b}(t) = k_\infty$ . These definitions are illustrated for boundary data satisfying Assumption 1 in Figure 2.

With the semiclassical approximation of the Dirichlet-to-Neumann map given in Definition 1, the direct scattering problem encoding the boundary data is

$$\begin{aligned} \epsilon \frac{d\mathbf{T}_0}{dt}(t; k) &= \begin{bmatrix} -2ik^2 - i|Q^D(t)|^2 & 2kQ^D(t) + iQ_0^N(t) \\ 2kQ^D(t)^* - iQ_0^N(t)^* & 2ik^2 + i|Q^D(t)|^2 \end{bmatrix} \mathbf{T}_0(t; k) \\ &= \begin{bmatrix} -2ik^2 - iH(t)^2 & (2k - U(t))H(t)e^{iS(t)/\epsilon} \\ (2k - U(t))H(t)e^{-iS(t)/\epsilon} & 2ik^2 + iH(t)^2 \end{bmatrix} \mathbf{T}_0(t; k), \\ \lim_{t \rightarrow +\infty} \mathbf{T}_0(t; k)e^{2ik^2 t \sigma_3 / \epsilon} &= \mathbb{I}. \end{aligned} \quad (43)$$



The oscillatory factors  $e^{\pm iS(t)/\epsilon}$  can be removed from the coefficient matrix by means of a simple substitution:

$$\mathbf{T}_0(t; k) = e^{iS(t)\sigma_3/(2\epsilon)} \mathbf{F}(t; k). \tag{44}$$

Indeed, making use of (41), this substitution leads to the equivalent system of equations

$$\epsilon \frac{d\mathbf{F}}{dt}(t; k) = \mathbf{B}(t; k)\mathbf{F}(t; k), \tag{45}$$

with  $\epsilon$ -independent coefficient matrix given by

$$\begin{aligned} \mathbf{B}(t; k) &:= \frac{1}{2} \begin{bmatrix} -4ik^2 + iU(t)^2 & 2H(t)(2k - U(t)) \\ 2H(t)(2k - U(t)) & 4ik^2 - iU(t)^2 \end{bmatrix} \\ &= \frac{1}{2}(2k - U(t)) \begin{bmatrix} -i(2k + U(t)) & 2H(t) \\ 2H(t) & i(2k + U(t)) \end{bmatrix}, \end{aligned} \tag{46}$$

that we need to solve subject to the boundary condition

$$\lim_{t \rightarrow +\infty} \mathbf{F}(t; k) e^{i(4k^2 t + S(t))\sigma_3/(2\epsilon)} = \mathbb{I}. \tag{47}$$

The corresponding spectral transforms are given for  $\Im\{k^2\} \leq 0$  by

$$A_0(k^*)^* := T_{0,11}(0; k) = e^{iS(0)/(2\epsilon)} f_1(0; k) \tag{48}$$

and

$$B_0(k^*)^* := T_{0,21}(0; k) = e^{-iS(0)/(2\epsilon)} f_2(0; k), \tag{49}$$

where  $\mathbf{f}(t; k) = (f_1(t; k), f_2(t; k))^T := (F_{11}(t; k), F_{21}(t; k))^T$  denotes the first column of  $\mathbf{F}(t; k)$ .

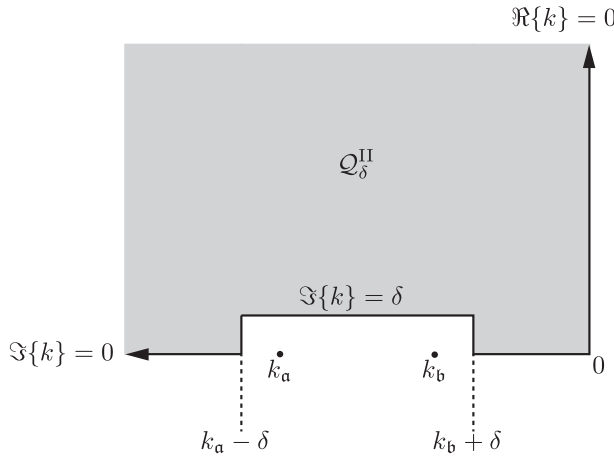
### 2.2. Semiclassical behavior of the spectral functions $A_0(k)$ and $B_0(k)$

Since  $\epsilon > 0$  appears both in the data  $(Q^D, Q_0^N)$  and in the differential equation (43), the spectral functions  $A_0(\cdot)$  and  $B_0(\cdot)$  will also depend on this small parameter. We now study this dependence rigorously in the limit  $\epsilon \downarrow 0$ .

Given any sufficiently small number  $\delta > 0$  (not necessarily related to the constant in Assumption 1) we define  $Q_\delta^{\text{II}}$  to be the closed unbounded subset of the  $k$ -plane characterized by the inequalities  $\pi/2 \leq \arg(k) \leq \pi$  and one of the three inequalities:  $\Re\{k\} \leq k_a - \delta$  or  $\Re\{k\} \geq k_b + \delta$  or  $\Im\{k\} \geq \delta$ . See Figure 3.

The eigenvalues  $\lambda$  of  $\mathbf{B}(t; k)$  satisfy

$$\begin{aligned} \lambda^2 &= \frac{1}{4}(2k - U(t))^2 [4H(t)^2 - (2k + U(t))^2] \\ &= (2k - U(t))^2 (k - \mathfrak{a}(t))(\mathfrak{b}(t) - k). \end{aligned} \tag{50}$$



**Figure 3.** The closed unbounded subset  $\mathcal{Q}_\delta^{\text{II}}$  of the second quadrant of the complex  $k$ -plane.

Given  $k$  with  $\Im\{k^2\} \leq 0$ , a positive real number  $t > 0$  is called a *turning point* for (45) if the two eigenvalues of  $\mathbf{B}(t; k)$  degenerate (at  $\lambda = 0$ ). We have the following basic fact.

**LEMMA 1 (Existence of turning points).** *Suppose that Assumption 1 holds and that  $\pi/2 \leq \arg(k) \leq \pi$ . Then there exist turning points  $t > 0$  precisely when  $k$  lies in the negative real interval  $k_a \leq k \leq k_b < 0$ . Moreover, for each  $k \in (k_a, k_b)$  there exist precisely two turning points  $t_-(k) < t_+(k)$ . The two turning points coalesce as  $k \downarrow k_a$  and as  $k \uparrow k_b$ :  $t_-(k_a) = t_+(k_a) = t_a$  and  $t_-(k_b) = t_+(k_b) = t_b$ . Also,  $t_-(k) \rightarrow 0$  as  $k \rightarrow k_0$  while  $t_+(k) \rightarrow +\infty$  as  $k \rightarrow k_\infty$ . Finally, given any  $\delta > 0$ , the condition that  $k \in \mathcal{Q}_\delta^{\text{II}}$  bounds  $\lambda$  away from zero uniformly for  $t > 0$ .*

We omit the proof as it is a direct and easy consequence of the conditions on the functions  $U(t)$  and  $H(t)$  and formula (50). Given a value of  $k$ , the presence or absence of turning points determines the nature of the spectral functions in the semiclassical limit.

**2.2.1. Analysis in the absence of turning points.** According to Lemma 1, there are no turning points if  $k \in \mathcal{Q}_\delta^{\text{II}}$ . This implies a certain triviality of the spectral functions in this region of the  $k$ -plane. We have the following result.

**PROPOSITION 1.** *Assume that  $q_0(x) = 0$ . Let a number  $\delta > 0$  be given, and suppose that the functions  $H : \mathbb{R}_+ \rightarrow \mathbb{R}$  and  $U : \mathbb{R}_+ \rightarrow \mathbb{R}$  satisfy Assumption 1. Then:*

- *For bounded  $\epsilon > 0$ , any zeros in  $\mathcal{Q}_\delta^{\text{II}}$  of the analytic function  $d_0(k) := A_0(k^*)^*$  lie in an  $\epsilon$ -independent bounded subset.*

- The analytic function  $d_0(k)$  has no zeros in  $\mathcal{Q}_\delta^{\text{II}}$  if  $\epsilon > 0$  is sufficiently small.
- The function  $\Gamma_0(k) := B_0(k^*)^*/A_0(k^*)^*$  satisfies a bound of the form

$$\Gamma_0(k) = \mathcal{O}\left(\frac{\epsilon^{1/2}}{1 + |k|^3}\right), \quad k \in \mathcal{Q}_\delta^{\text{II}}, \quad \epsilon > 0, \tag{51}$$

where the constant implicit in the estimate depends only on the functions  $H$  and  $U$ .

In other words, when  $\epsilon > 0$  is small, then for  $k \in \mathcal{Q}_\delta^{\text{II}}$ ,  $\Gamma_0(k)$  has no poles and is uniformly small. The proof of this proposition is given in Appendix A.

2.2.2. *Analysis in the presence of turning points.* We now study the asymptotic behavior of the function  $\Gamma_0(k) := B_0(k^*)^*/A_0(k^*)^*$  for real  $k$  in the interval  $k_a < k < k_b$ . For each such  $k$ , as can be seen in Figure 2, the eigenvalues  $\pm\lambda$  of the coefficient matrix  $\mathbf{B}(t; k)$  satisfy  $\lambda^2 > 0$  for  $t_-(k) < t < t_+(k)$ , while  $\lambda^2 < 0$  for  $0 < t < t_-(k)$  and for  $t > t_+(k)$ . Considering  $\epsilon > 0$  as being very small, one is reminded of the language of the WKB method, in which the interval  $(t_-(k), t_+(k))$  is analogous to a “classically forbidden region” separating two “classically allowed regions.” Thus, we have an analogue of a quantum tunneling problem. Rather than use the WKB method, which is well-known to fail near the turning points  $t_\pm(k)$ , in the proof of the following results we use the method of Langer transformations to uniformly handle the neighborhoods of the two turning points while simultaneously maintaining full accuracy when  $t$  is not close to either turning point. The presence of turning points leads to nontrivial behavior of the spectral functions in the limit  $\epsilon \downarrow 0$ , as the following result shows.

PROPOSITION 2. *Let  $k \in (k_a, k_b)$  with  $k \neq k_0$  and  $k \neq k_\infty$ , and suppose that the functions  $H : \mathbb{R}_+ \rightarrow \mathbb{R}$  and  $U : \mathbb{R}_+ \rightarrow \mathbb{R}$  satisfy Assumption 1. Then in the limit  $\epsilon \downarrow 0$ ,*

$$A_0(k) = -e^{\tau(k)/\epsilon} e^{-i(\ell(k)\text{sgn}(k^2 - k_\infty^2) - \frac{1}{2}S_\infty)/\epsilon} \left[ e^{-i\Phi(k)/\epsilon} + \mathcal{O}(\epsilon) \right] \tag{52}$$

and

$$B_0(k) = -e^{\tau(k)/\epsilon} e^{-i(\ell(k)\text{sgn}(k^2 - k_\infty^2) - \frac{1}{2}S_\infty)/\epsilon} \left[ e^{i\Phi(k)/\epsilon} + \mathcal{O}(\epsilon) \right] \tag{53}$$

where

$$\begin{aligned} \ell(k) := & \int_{t_+(k)}^{+\infty} \left[ (U(t) - 2k)\sqrt{(k - \mathbf{a}(t))(k - \mathbf{b}(t))} - 2|k^2 - k_\infty^2| \right] dt \\ & - 2|k^2 - k_\infty^2|_{t_+(k)}, \end{aligned} \tag{54}$$

$$S_\infty := \lim_{t \rightarrow +\infty} (S(t) + U_\infty^2 t) = \lim_{t \rightarrow +\infty} (S(t) + 4k_\infty^2 t), \tag{55}$$

$$\tau(k) := \int_{t_-(k)}^{t_+(k)} (U(t) - 2k)\sqrt{(k - \mathbf{a}(t))(\mathbf{b}(t) - k)} dt, \tag{56}$$

and

$$\Phi(k) := \frac{1}{2}S(0) + \operatorname{sgn}(k^2 - k_0^2) \int_0^{t_-(k)} (U(t) - 2k)\sqrt{(k - \mathfrak{a}(t))(k - \mathfrak{b}(t))} dt. \quad (57)$$

The error terms are uniform in  $k$  in compact subintervals of  $(k_a, k_b) \setminus \{k_0, k_\infty\}$ .

The proof of Proposition 2 is given in Appendix B. An immediate corollary is the following.

**COROLLARY 3.** *Suppose that  $q_0(x) = 0$ . Under the same conditions and with the same characterization of the error terms as in Proposition 2, we have*

$$\Gamma_0(k) = \frac{B_0(k)^*}{A_0(k)^*} = \sqrt{1 - e^{-2\tau(k)/\epsilon}} e^{-2i\Phi(k)/\epsilon} + \mathcal{O}(\epsilon) \quad \text{and} \\ 1 - |\Gamma_0(k)|^2 = e^{-2\tau(k)/\epsilon} (1 + \mathcal{O}(\epsilon)). \quad (58)$$

*Proof:* Since  $q_0(x) = 0$  we have  $\gamma(k) = 0$  and hence  $\Gamma_0(k) = B_0(k)^*/A_0(k)^*$  for all real  $k < 0$ . The formula for  $1 - |\Gamma_0(k)|^2$  follows from the identity  $1 - |\Gamma_0(k)|^2 = 1/|A_0(k)|^2$  equivalent to the condition that  $\det(\mathbf{T}_0(0; k)) = 1$ . The factor  $\sqrt{1 - e^{-2\tau(k)/\epsilon}}$  is exponentially close to 1 (except near  $k_a$  and  $k_b$ , points excluded from consideration) and is included in the formula for  $\Gamma_0$  to ensure that the jump matrix we shall construct from this approximation has unit determinant. ■

### 3. Formulation of the inverse problem

Propositions 1 and 2 and Corollary 3 give a rigorous characterization of the spectral functions associated with vanishing initial data  $q_0(x)$  and with a class of boundary data  $(Q^D, Q_0^N)$  given (in terms of the functions  $H(\cdot)$  and  $U(\cdot)$ ) described by Assumption 1) by (24) and (28) subject to the equation (41) giving  $S(t)$  in terms of  $H(\cdot)$  and  $U(\cdot)$ . To summarize:

- From  $q_0 = 0$  we have  $\gamma(k) = 0$  for all  $k \in \mathbb{R}$ . Therefore, Riemann–Hilbert problem 1 has no jump on the positive real axis, and the remaining jump matrices only involve  $\Gamma_0(k)$ , which is simply a ratio of the spectral functions  $A_0(k^*)^*$  and  $B_0(k^*)^*$  arising from the approximate boundary data.
- If the function  $\Gamma_0(k)$  has any poles in the second quadrant of the complex plane, they must lie very close (in the limit  $\epsilon \downarrow 0$ ) to the negative real interval  $[k_a, k_b]$ .
- On the imaginary  $k$ -axis, as well as on the negative real  $k$ -axis away from the interval  $[k_a, k_b]$ ,  $\Gamma_0(k)$  is small in the limit  $\epsilon \downarrow 0$ .
- In the interior of the negative real interval  $(k_a, k_b)$  and away from the special points  $k_0$  and  $k_\infty$ ,  $\Gamma_0(k)$  has an accurate explicit approximation given by Corollary 3.

However, this information alone is insufficient to properly formulate and analyze Riemann–Hilbert problem 1 associated with the exact spectral transforms  $A_0(k)$  and  $B_0(k)$  corresponding to the approximate Neumann boundary data  $Q_0^N(t)$ . Indeed, to formulate the Riemann–Hilbert problem without poles one would need to know a priori that there cannot be any poles of  $\Gamma_0(k)$  whatsoever in the second quadrant, and it is not enough to know that any poles have to move toward  $[k_a, k_b]$  as  $\epsilon \downarrow 0$ . Another issue is that our results do not provide approximations for  $\Gamma_0(k)$  near the real points  $k_a, k_0, k_\infty$ , or  $k_b$ . In fact, the analytical methodology based on Langer transformations used in the proof of Proposition 2 either requires substantial modification or breaks down entirely in neighborhoods of these points.

These arguments<sup>5</sup> suggest making a further modification of the first step of the proposed iteration algorithm: we will reformulate the inverse problem by:

- Assuming that the Riemann–Hilbert problem can be formulated without poles,
- Neglecting  $\Gamma_0(k)$  and  $\Gamma_0(k)^*$  completely on the imaginary axis,
- Neglecting  $\Gamma_0(k)$  on the real axis for  $k < k_a$  and  $k > k_b$ , and
- Replacing  $\Gamma_0(k)$  in the whole interval  $(k_a, k_b)$  by the formulae recorded in Corollary 3 with the  $\mathcal{O}(\epsilon)$  error terms set to zero.

The resulting Riemann–Hilbert problem has the negative real interval  $[k_a, k_b]$  as its only jump contour. For convenience we will re-orient this contour from left to right, which requires the inversion of the jump matrix written in (14).

We therefore formulate the following Riemann–Hilbert problem. Let  $\tilde{\Gamma}(k)$  be defined by

$$\tilde{\Gamma}(k) := \chi_{(k_a, k_b)}(k) Y^\epsilon(k) e^{-2i\Phi(k)/\epsilon}, \quad Y^\epsilon(k) := \sqrt{1 - e^{-2\tau(k)/\epsilon}}, \quad k \in \mathbb{R}, \quad (59)$$

where  $\chi_{(k_a, k_b)}$  denotes the characteristic function of the interval  $(k_a, k_b)$ , and where  $\tau : (k_a, k_b) \rightarrow \mathbb{R}_+$  is defined by (56) while  $\Phi : (k_a, k_b) \rightarrow \mathbb{R}$  is defined by (57). It can be shown that  $\tilde{\Gamma} : \mathbb{R} \rightarrow \mathbb{C}$  is Hölder continuous with every exponent  $0 < \alpha \leq 1$ .

**Riemann–Hilbert problem 2.** Seek a  $2 \times 2$  matrix function  $\tilde{\mathbf{M}} : \mathbb{C} \setminus \mathbb{R} \rightarrow \mathbb{C}^{2 \times 2}$  with the following properties:

**Analyticity:**  $\tilde{\mathbf{M}}$  is analytic in  $\mathbb{C}_+$  and  $\mathbb{C}_-$  and Hölder continuous for some exponent  $0 < \alpha < 1$  in  $\overline{\mathbb{C}}_+$  and  $\overline{\mathbb{C}}_-$ , taking boundary values  $\tilde{\mathbf{M}}_\pm : \mathbb{R} \rightarrow \mathbb{C}^{2 \times 2}$  on  $\mathbb{R}$  from  $\mathbb{C}_\pm$ .

---

<sup>5</sup>A more serious reason for making this modification, especially the step of setting the jump matrix on the imaginary axis to the identity, is discussed in detail in Remark 6.

**Jump Condition:** The boundary values are related by

$$\tilde{\mathbf{M}}_+(k) = \tilde{\mathbf{M}}_-(k) \begin{bmatrix} 1 - |\tilde{\Gamma}(k)|^2 & -\tilde{\Gamma}(k)^* e^{-2i\theta(k;x,t)/\epsilon} \\ \tilde{\Gamma}(k) e^{2i\theta(k;x,t)/\epsilon} & 1 \end{bmatrix},$$

$$k \in \mathbb{R}. \tag{60}$$

**Normalization:** The matrix function  $\tilde{\mathbf{M}}$  satisfies

$$\lim_{k \rightarrow \infty} \tilde{\mathbf{M}}(k) = \mathbb{I}, \tag{61}$$

where the limit is uniform with respect to direction in the complex plane.

The following is a standard result.

**PROPOSITION 3.** *Riemann–Hilbert problem 2 has a unique solution for every  $(x, t) \in \mathbb{R}^2$  and for every  $\epsilon > 0$ . The function*

$$q = \tilde{q}^\epsilon(x, t) := 2i \lim_{k \rightarrow \infty} k \tilde{M}_{12}(k) \tag{62}$$

*is infinitely differentiable with respect to  $(x, t)$ , and for every  $\epsilon > 0$ ,  $q = \tilde{q}^\epsilon(x, t)$  satisfies the defocusing nonlinear Schrödinger equation (1).*

*Proof:* To see the uniqueness of the solution of Riemann–Hilbert problem 2 (assuming existence), one first notes that  $\det(\tilde{\mathbf{M}}(k)) = 1$  necessarily holds as an identity for any solution, and therefore  $\tilde{\mathbf{M}}(k)^{-1}$  is also analytic for  $k \in \mathbb{C} \setminus \mathbb{R}$ . Given two solutions, say  $\tilde{\mathbf{M}}$  and  $\tilde{\mathbf{N}}$ , one considers the matrix ratio  $\mathbf{R}(k) := \tilde{\mathbf{M}}(k)\tilde{\mathbf{N}}(k)^{-1}$ , which is analytic for  $k \in \mathbb{C} \setminus \mathbb{R}$  and satisfies  $\mathbf{R}(k) \rightarrow \mathbb{I}$  as  $k \rightarrow \infty$ . A simple calculation using the jump condition (60) satisfied by both  $\tilde{\mathbf{M}}$  and  $\tilde{\mathbf{N}}$  shows that the continuous boundary values taken on  $\mathbb{R}$  agree:  $\mathbf{R}_+(k) = \mathbf{R}_-(k)$  for all  $k \in \mathbb{R}$ . It follows that  $\mathbf{R}(k)$  is an entire analytic (matrix-valued) function of  $k$  that tends to  $\mathbb{I}$  as  $k \rightarrow \infty$ , so by Liouville’s theorem  $\mathbf{R}(k) = \mathbb{I}$  for all  $k$ , i.e.,  $\tilde{\mathbf{N}}(k) = \tilde{\mathbf{M}}(k)$  holds for all  $k \in \mathbb{C} \setminus \mathbb{R}$ .

To establish existence of a solution, one observes that Riemann–Hilbert problem 2 is equivalent to a system of linear singular integral equations for which the relevant operator is Fredholm with zero index on an appropriate space of Hölder-continuous functions (see [14] and [7]). It therefore suffices to show that the kernel of this Fredholm operator is trivial. Triviality of the kernel is equivalent to the assertion that the only solution of the homogeneous form of Riemann–Hilbert problem 2, in which the normalization condition (61) is replaced with a limit of  $\mathbf{0}$  as  $k \rightarrow \infty$ , is the zero matrix. Zhou’s *vanishing lemma* (theorem 9.3 of [7]) shows that this latter assertion holds true in the present case because the jump contour is the real axis and the jump matrix has a positive semidefinite real part as a consequence of the inequality  $|\tilde{\Gamma}(k)|^2 \leq 1$  holding for  $k \in \mathbb{R}$ .

The infinite differentiability of the matrix  $\tilde{\mathbf{M}}$  with respect to  $(x, t)$ , and hence that of  $\tilde{q}^\epsilon$ , follows from the compact support  $[k_a, k_b]$  of  $\tilde{\Gamma}$ . Finally, let us show that  $q = \tilde{q}^\epsilon$  satisfies (1). We begin by defining a matrix  $\mathbf{L}(k)$  from the solution  $\tilde{\mathbf{M}}(k)$  by setting

$$\mathbf{L}(k) := \tilde{\mathbf{M}}(k)e^{-i\theta(k;x,t)\sigma_3/\epsilon}. \tag{63}$$

One verifies that  $\mathbf{L}(k)$  is analytic for  $k \in \mathbb{C} \setminus \mathbb{R}$ , and that

$$\mathbf{L}_+(k) = \mathbf{L}_-(k) \begin{bmatrix} 1 - |\tilde{\Gamma}(k)|^2 & -\tilde{\Gamma}(k)^* \\ \tilde{\Gamma}(k) & 1 \end{bmatrix}, \quad k \in \mathbb{R}. \tag{64}$$

It follows from the fact that this jump condition is independent of  $x$  and  $t$ , that the matrices  $\mathbf{U}(k) := \epsilon \mathbf{L}_x(k)\mathbf{L}(k)^{-1}$  and  $\mathbf{V}(k) := \epsilon \mathbf{L}_t(k)\mathbf{L}(k)^{-1}$  have no jump across the real axis and since  $\det(\mathbf{L}(k)) = 1$ ,  $\mathbf{U}$  and  $\mathbf{V}$  are entire functions of  $k$ . Moreover, from the asymptotic behavior of  $\mathbf{L}(k)$  near  $k = \infty$  one can check that  $\mathbf{U}$  is a linear function of  $k$  while  $\mathbf{V}$  is a quadratic polynomial in  $k$ . In fact, using (62) one sees that  $\mathbf{U}$  is given by (4) with  $q = \tilde{q}^\epsilon$ . Moreover, using the fact that  $\mathbf{L}(k)$  satisfies the differential equation  $\epsilon \mathbf{L}_x = \mathbf{U}\mathbf{L}$  (by definition of  $\mathbf{U}$ ), one sees that  $\mathbf{V}$  is given by (5) with  $q = \tilde{q}^\epsilon$ . The fact that  $\mathbf{L}$  is a simultaneous fundamental solution matrix of the Lax pair equations  $\epsilon \mathbf{L}_x = \mathbf{U}\mathbf{L}$  and  $\epsilon \mathbf{L}_t = \mathbf{V}\mathbf{L}$  means that these equations are consistent, that is, the zero-curvature condition

$$\epsilon \mathbf{U}_t - \epsilon \mathbf{V}_x + [\mathbf{U}, \mathbf{V}] = \mathbf{0} \tag{65}$$

holds, and substitution from (4) and (5) yields the equation (1) for  $q = \tilde{q}^\epsilon$  (and the complex conjugate of that equation). ■

We note that this proof implies that  $\tilde{\mathbf{M}}(k)$  has a convergent Laurent series in descending powers of  $k$  for  $|k|$  sufficiently large:

$$\tilde{\mathbf{M}}(k) = \mathbb{I} + k^{-1}\tilde{\mathbf{M}}_1 + k^{-2}\tilde{\mathbf{M}}_2 + \mathcal{O}(k^{-3}), \quad k \rightarrow \infty, \tag{66}$$

and that  $\tilde{q}^\epsilon(x, t)$  and  $\epsilon \tilde{q}_x^\epsilon(x, t)$  can be expressed in terms of the coefficients  $\tilde{\mathbf{M}}_1$  and  $\tilde{\mathbf{M}}_2$  as follows:

$$\begin{aligned} \tilde{q}^\epsilon(x, t) &= 2i\tilde{M}_{1,12} \quad \text{and} \quad \epsilon \tilde{q}_x^\epsilon(x, t) = 4\tilde{M}_{2,12} + 2i\tilde{q}^\epsilon(x, t)\tilde{M}_{1,22} \\ &= 4\tilde{M}_{2,12} - 4\tilde{M}_{1,12}\tilde{M}_{1,22}. \end{aligned} \tag{67}$$

The question that remains is what, if anything, does the family of functions  $\tilde{q}^\epsilon$  have to do with the exact solution  $q^\epsilon$  of the Dirichlet initial-boundary value problem with  $q^\epsilon(x, 0) = 0$  for  $x > 0$  and  $q^\epsilon(0, t) = H^\epsilon(t)e^{iS(t)/\epsilon}$  for  $t > 0$  (recall the modified amplitude function  $H^\epsilon(t)$  defined in Remark 1). This is the topic we take up next.

*Remark 2.* The values of  $\Gamma(k)$  for  $k$  real and positive are irrelevant to the inverse problem, as the jump matrix for  $k > 0$  generally only involves the function  $\gamma(k)$  obtained from the initial data for  $x > 0$  (see (11)), and in

the present case that  $q_0(x) = 0$ ,  $\gamma(k) = 0$ . The condition  $U(t) > 2H(t)$  implied by Assumption 1 ensures that  $k_b < 0$ , and hence that the full asymptotic support of  $\Gamma$  on  $\mathbb{R}$  (and hence by definition the exact support of  $\tilde{\Gamma}$  on  $\mathbb{R}$ ) contributes to the jump matrix for the inverse problem. If on the contrary we had  $k_b > 0$ , then some information about the boundary data would be lost from the inverse problem in the semiclassical limit. It therefore seems that it is possible to reconstruct the boundary data only if  $U(t) > 2H(t)$ . As pointed out earlier, this is a stronger condition than the necessary condition  $U(t) > H(t)$  for the boundary to be a spacelike curve for the hyperbolic dispersionless system (38).

*Remark 3.* One may observe that Riemann–Hilbert problem 2 is of exactly the same form as that which occurs in the treatment of the initial-value problem for the defocusing nonlinear Schrödinger equation formulated on an appropriate space of decaying functions of  $x \in \mathbb{R}$  instead of the half-line. The function  $\tilde{\Gamma}(k)$ , here obtained from Dirichlet boundary data via the temporal part of the Lax pair, plays the role usually played by the reflection coefficient calculated from initial data via the spatial part of the Lax pair. This means that the “reflection coefficient”  $\tilde{\Gamma}(k)$  corresponds to some initial data given on the whole line  $x \in \mathbb{R}$ , a fact that has been made quite rigorous in [15]. In this case, according to Theorem 1, the initial data is very small for  $x > 0$ ; however to reproduce the nontrivial boundary data described by Theorem 2 the initial data must not be small for  $x < 0$ . Thus, the formula (59) for  $\tilde{\Gamma}(k)$  makes a connection in the transform domain between (i) a problem on the half-line with zero initial data and nontrivial boundary data and (ii) a problem on the whole line with initial data supported on the negative half-line. The latter initial data is then defined implicitly in terms of the boundary data for the former problem via the solution of Riemann–Hilbert problem 2.

#### 4. Semiclassical analysis of $\tilde{q}^\epsilon(x, t)$

##### 4.1. Asymptotic behavior of $\tilde{q}^\epsilon(x, 0)$ for $x > 0$ and related analysis

4.1.1. *Implications of Assumption 1 for the functions  $\Phi$  and  $\tau$ .* In light of Assumption 1, the definition (56) of  $\tau$  easily implies the following.

LEMMA 2. *Under Assumption 1, the function  $\tau : (k_a, k_b) \setminus \{k_0, k_\infty\} \rightarrow \mathbb{R}$  is real analytic and it extends by continuity to a function of class  $C^0(k_a, k_b)$  satisfying  $\tau(k) > 0$ .*

We will also require detailed information about the behavior of the function  $\Phi$  near the point  $k = k_0$ . In this direction we have the following.

LEMMA 3. *If the functions  $U$  and  $H$  satisfy Assumption 1, then  $\Phi : (k_a, k_b) \setminus \{k_0, k_\infty\} \rightarrow \mathbb{R}$  defined by (57) extends by continuity to a function*



analytic in  $(k_a, k_b) \setminus \{k_0\}$ , and of class  $C^3$  in a neighborhood of  $k = k_0$ . Also,  $\Phi'(k) \leq 0$  for  $k_a < k < k_b$  with equality only for  $k = k_0$ .

*Proof:* Using analyticity of  $U$  and  $H$ , which implies that of  $\mathfrak{a}$  and  $\mathfrak{b}$ , we may write  $\Phi(k)$  in terms of a contour integral. Indeed,

$$\Phi(k) = \frac{1}{2}S(0) + \frac{1}{2}i \operatorname{sgn}(k^2 - k_0^2) \oint_C (U(t) - 2k)(k - \mathfrak{a}(t))^{1/2}(\mathfrak{b}(t) - k)^{1/2} dt, \quad (68)$$

where the fractional powers denote the principal branches. The integrand has a branch cut connecting  $t = 0$  with  $t = t_-(k)$  (due to the factor  $(k - \mathfrak{a}(t))^{1/2}$  when  $k_a < k < k_0$  and due to the factor  $(\mathfrak{b}(t) - k)^{1/2}$  when  $k_0 < k < k_b$ ). The contour  $C$  is a positively oriented loop; it begins at  $t = 0$  on the lower edge of the branch cut, encloses the cut once passing through the real axis at a point  $t = t_-(k) + \delta < t_+(k)$ , and terminates at  $t = 0$  on the upper edge of the cut. In the neighborhood of a fixed value of  $k$  the contour  $C$  may be taken to be independent of  $k$ , and it follows easily that  $\Phi(k)$  is analytic in  $k$  separately in the intervals  $(k_a, k_0)$  and  $(k_0, k_b)$ .

For  $k \neq k_0$ , all derivatives of  $\Phi$  may be calculated by differentiation under the integral sign. Thus,

$$\Phi'(k) = i \operatorname{sgn}(k^2 - k_0^2) \oint_C \frac{2k^2 + U(t)k - H(t)^2}{(k - \mathfrak{a}(t))^{1/2}(\mathfrak{b}(t) - k)^{1/2}} dt, \quad k \neq k_0, \quad (69)$$

$$\Phi''(k) = \frac{1}{4}i \operatorname{sgn}(k^2 - k_0^2) \oint_C \frac{2H(t)^2(6k + U(t)) - (2k + U(t))^3}{(k - \mathfrak{a}(t))^{3/2}(\mathfrak{b}(t) - k)^{3/2}} dt, \quad k \neq k_0, \quad (70)$$

$$\Phi'''(k) = \frac{3}{2}i \operatorname{sgn}(k^2 - k_0^2) \oint_C \frac{H(t)^2(2H(t)^2 - U(t)(2k + U(t)))}{(k - \mathfrak{a}(t))^{5/2}(\mathfrak{b}(t) - k)^{5/2}} dt, \quad k \neq k_0, \quad (71)$$

and

$$\begin{aligned} &\Phi^{(4)}(k) \\ &= \frac{3}{2}i \operatorname{sgn}(k^2 - k_0^2) \oint_C \frac{H(t)^2(H(t)^2(10k + 3U(t)) - 2U(t)(2k + U(t))^2)}{(k - \mathfrak{a}(t))^{7/2}(\mathfrak{b}(t) - k)^{7/2}} dt, \\ &k \neq k_0. \end{aligned} \quad (72)$$

We consider  $k$  to be a real number close to, but not equal to,  $k_0$ . Assumption 1 ensures that  $\mathfrak{a}(t) = k_0 - h_0 t^{1/2} + o(t^{1/2})$  and  $\mathfrak{b}(t) = k_0 + h_0 t^{1/2} + o(t^{1/2})$  for small  $t$ . This implies that when  $k - k_0$  is small,  $t_-(k)$  is proportional to  $(k - k_0)^2$ . Based on this observation, we scale the contour  $C$  as  $C = h_0^{-2}(k - k_0)^2 D$ , where  $D$  is a suitable contour that we will hold fixed as  $k \rightarrow k_0$ , and we make the substitution  $t = h_0^{-2}(k - k_0)^2 s$  in the above integrals. In each case, the

integrand considered as a function of  $s$  has uniform asymptotic behavior on the contour  $D$  in the limit  $k \rightarrow k_0$ , being determined from the local behavior of the functions  $U$  and  $H$  near  $t = 0$  as specified in Assumption 1. Thus, uniformly for  $s \in D$  one has  $U(t) = U(h_0^{-2}(k - k_0)^2 s) = -2k_0 + o(k - k_0)$  and  $H(t) = H(h_0^{-2}(k - k_0)^2 s) = |k - k_0|s^{1/2}(1 + o(1))$  in the limit  $k \rightarrow k_0$ , and it follows that

$$\Phi(k) = \frac{1}{2}S(0) + 2ik_0h_0^{-2}(k - k_0)^3 \left[ \oint_D (s - 1)^{1/2} ds + o(1) \right], \quad k \rightarrow k_0, \quad (73)$$

$$\Phi'(k) = -2ik_0h_0^{-2}(k - k_0)^2 \left[ \oint_D \frac{ds}{(s - 1)^{1/2}} + o(1) \right], \quad k \rightarrow k_0, \quad (74)$$

$$\Phi''(k) = -2ik_0h_0^{-2}(k - k_0) \left[ \oint_D \frac{s ds}{(s - 1)^{3/2}} + o(1) \right], \quad k \rightarrow k_0, \quad (75)$$

$$\Phi'''(k) = -6ik_0h_0^{-2} \left[ \oint_D \frac{s ds}{(s - 1)^{5/3}} + o(1) \right], \quad k \rightarrow k_0, \quad (76)$$

and

$$\Phi^{(4)}(k) = -6ik_0h_0^{-2}(k - k_0)^{-1} \left[ \oint_D \frac{s^2 + 4s}{(s - 1)^{7/2}} ds + o(1) \right], \quad k \rightarrow k_0. \quad (77)$$

The contour  $D$  begins and ends at  $s = 0$  on opposite sides of the branch cut (all fractional powers of  $s - 1$  are understood as principal branches) and encircles the branch point  $s = 1$  once in the counterclockwise sense. It is now obvious that  $\Phi(k)$  tends to  $\frac{1}{2}S(0)$  while  $\Phi'(k)$  and  $\Phi''(k)$  both vanish as  $k \rightarrow k_0$  and hence all three extend by continuity to  $k = k_0$ . It is also obvious that  $\Phi'''(k)$  has a finite limit as  $k \rightarrow k_0$ ; by computing an integral we find the limiting value

$$\Phi'''(k_0) = 16k_0h_0^{-2}. \quad (78)$$

This completes the proof that  $\Phi(k)$  is of class  $C^3$  near  $k = k_0$ . We note that the fourth derivative  $\Phi^{(4)}(k)$  appears to be singular at  $k = k_0$ , but in reality the issue is subtle because the explicit leading term in the square brackets in (77) is an integral that vanishes identically, and therefore the asymptotic behavior of  $\Phi^{(4)}(k)$  in the limit  $k \rightarrow k_0$  cannot be determined without making further hypotheses on  $U$  and  $H$  sufficient to provide leading-order asymptotics for the  $o(1)$  error term in (77).

It remains to determine the sign of  $\Phi'(k)$  for  $k \neq k_0$ . We go back to the real integral representation (57) for  $\Phi(k)$ , which admits differentiation by Leibniz' rule because either  $k - \mathfrak{a}(t_-(k)) = 0$  or  $k - \mathfrak{b}(t_-(k)) = 0$  with the result:

$$\Phi'(k) = -\operatorname{sgn}(k^2 - k_0^2) \int_0^{t_-(k)} \frac{4k^2 + 2U(t)k - 2H(t)^2}{\sqrt{(k - \mathfrak{a}(t))(k - \mathfrak{b}(t))}} dt, \quad (79)$$

which can also be written in the form

$$\Phi'(k) = -\operatorname{sgn}(k^2 - k_0^2) \int_0^{t_-(k)} \frac{2(\alpha(t) - k)(\mathfrak{b}(t) - k) + \frac{1}{2}(U(t) - 2k)((\alpha(t) - k) + (\mathfrak{b}(t) - k))}{\sqrt{(\alpha(t) - k)(\mathfrak{b}(t) - k)}} dt \tag{80}$$

If  $k_0 < k < k_b$ , then we use the form (79) and factor the quadratic in the numerator as  $4(k - k_+(t))(k - k_-(t))$  with

$$k_{\pm}(t) := \frac{1}{4} \left( \alpha(t) + \mathfrak{b}(t) \pm \sqrt{(\alpha(t) + \mathfrak{b}(t))^2 + (\mathfrak{b}(t) - \alpha(t))^2} \right). \tag{81}$$

Since  $k_+(t) > 0$  and  $k < k_b < 0$ , obviously  $k - k_+(t) < 0$ . Also, since  $k_-(t) < \frac{1}{2}(\alpha(t) + \mathfrak{b}(t)) \leq \mathfrak{b}(t)$  we have  $k - k_-(t) > k - \mathfrak{b}(t)$ , a lower bound which is nonnegative for  $0 < t < t_-(k)$  because  $k > k_0$ . Hence,  $\Phi'(k) < 0$  for  $k_0 < k < k_b$ . On the other hand, for  $k_a < k < k_0$  we instead use the form (80), because in this range of  $k$  we have  $\alpha(t) - k > 0$  and  $\mathfrak{b}(t) - k > 0$  for  $0 < t < t_-(k)$  so combining this with the inequality  $U(t) - 2k > 0$  shows that also for  $k_a < k < k_0$  we have  $\Phi'(k) < 0$ .

*Remark 4.* Although it may seem counterintuitive, assuming that  $H$  is smoother at  $t = 0$ , say vanishing linearly rather than like  $t^{1/2}$  as  $t \downarrow 0$ , leads to less smoothness of  $\Phi(k)$  at  $k = k_0$ . Linear vanishing of  $H$  implies continuity of  $\Phi$  and  $\Phi'$  at  $k_0$ , but  $\Phi''$  will have a jump discontinuity. The point is that it should be the inverse function  $t_-(k)$  that is smooth at  $k = k_0$ , not the functions  $\alpha(t)$  and  $\mathfrak{b}(t)$  at  $t = 0$ .

The part of Assumption 1 concerning the nondegeneracy of the extrema of  $\mathfrak{a}$  and  $\mathfrak{b}$  allows us to obtain the following result.

LEMMA 4. *If the functions  $U$  and  $H$  satisfy Assumption 1, then  $\tau : (k_a, k_b) \rightarrow \mathbb{R}_+$  is analytic at  $k = k_a$  and  $k = k_b$ , with  $\tau(k_a) = 0$  and  $\tau'(k_a) > 0$  while  $\tau(k_b) = 0$  and  $\tau'(k_b) < 0$ .*

*Proof:* Using analyticity of  $U$  and  $H$  for  $t$  near  $t_a$  and  $t_b$ , we can express  $\tau(k)$  as a contour integral:

$$\tau(k) = \frac{1}{2} \oint_C (U(t) - 2k)R(t; k) dt, \tag{82}$$

where  $C$  is a closed contour enclosing the interval  $[t_-(k), t_+(k)]$  once in the positive sense and where  $R(t; k)$  is the function analytic in  $D \setminus [t_-(k), t_+(k)]$ , where  $D$  is a domain containing  $C$ , that satisfies  $R(t; k)^2 = (k - \alpha(t))(\mathfrak{b}(t) - k)$  and that the boundary value  $R_+(t; k)$  taken on the upper edge of the branch cut  $[t_-(k), t_+(k)]$  is negative.

To analyze  $\tau$  for  $k$  near  $k_a$  and  $k_b$ , we may in each case choose the contour  $C$  to be fixed, and then since the integrand is analytic in  $k$  at each point  $t \in C$  it follows that  $\tau$  extends from a function defined for real  $k$  in a right (left) neighborhood of  $k = k_a$  ( $k = k_b$ ) to an analytic function of  $k$  at  $k_a$  ( $k_b$ ). Since  $C$  is fixed and since  $R_k(t; k) = (a(t) + b(t) - 2k)/(2R(t; k)) = -(U(t) + 2k)/(2R(t; k))$  by differentiation of  $R(t; k)^2$ ,

$$\tau'(k) = - \oint_C R(t; k) dt - \frac{1}{4} \oint_C \frac{(U(t) - 2k)(U(t) + 2k) dt}{R(t; k)}. \tag{83}$$

Now near  $t_a$ , the function  $R(t; k_a)^2$  has the Taylor expansion  $R(t; k_a)^2 = -\frac{1}{2}a''(t_a)(b(t_a) - k_a)(t - t_a)^2 + \mathcal{O}((t - t_a)^3)$ , and since  $R(t; k)$  is positive imaginary to the right of  $t_+(k)$  and negative imaginary to the left of  $t_-(k)$ , it follows that when  $k = k_a$  so that the branch cut collapses to a point  $t = t_a$ , we have

$$R(t; k_a) = i\sqrt{\frac{1}{2}a''(t_a)(b(t_a) - k_a)(t - t_a)} + \mathcal{O}((t - t_a)^2), \quad t \rightarrow t_a. \tag{84}$$

Similar arguments show that

$$R(t; k_b) = i\sqrt{-\frac{1}{2}b''(t_b)(k_b - a(t_b))(t - t_b)} + \mathcal{O}((t - t_b)^2), \quad t \rightarrow t_b. \tag{85}$$

In both cases the indicated square roots are positive numbers. In particular, since  $R(t; k_a)$  and  $R(t; k_b)$  are analytic functions of  $t$  within  $C$  it follows from (82) that  $\tau(k_a) = \tau(k_b) = 0$ . We may now use (84)–(85) in (83) to compute  $\tau'(k_a)$  and  $\tau'(k_b)$  by residues:

$$\tau'(k_a) = -\frac{\pi}{2} \frac{(U(t_a) - 2k_a)(U(t_a) + 2k_a)}{\sqrt{\frac{1}{2}a''(t_a)(b(t_a) - k_a)}} = \frac{\pi}{2} \frac{(U(t_a) - 2k_a)\sqrt{b(t_a) - k_a}}{\sqrt{\frac{1}{2}a''(t_a)}} > 0 \tag{86}$$

and

$$\tau'(k_b) = -\frac{\pi}{2} \frac{(U(t_b) - 2k_b)(U(t_b) + 2k_b)}{\sqrt{-\frac{1}{2}b''(t_b)(k_b - a(t_b))}} = -\frac{\pi}{2} \frac{(U(t_b) - 2k_b)\sqrt{k_b - a(t_b)}}{\sqrt{-\frac{1}{2}b''(t_b)}} < 0. \tag{87}$$

This completes the proof. ■

The nondegeneracy of the extrema of  $a$  and  $b$  also leads to the following result.

LEMMA 5. *If the functions  $U$  and  $H$  satisfy Assumption 1, then  $\Phi : (k_a, k_b) \rightarrow \mathbb{R}$  has an analytic continuation into the complex plane from a right neighborhood of  $k_a$  and from a left neighborhood of  $k_b$  and*

$$\begin{aligned} \Phi(k) &= \Phi(k_\alpha) + C_\alpha(k - k_\alpha) \log(k - k_\alpha) + \mathcal{O}(k - k_\alpha), \quad k \rightarrow k_\alpha, \\ C_\alpha &:= \frac{1}{4} \frac{(U(t_\alpha) - 2k_\alpha)\sqrt{b(t_\alpha) - k_\alpha}}{\sqrt{\frac{1}{2}a''(t_\alpha)}}, \end{aligned} \tag{88}$$

while

$$\begin{aligned} \Phi(k) &= \Phi(k_b) + C_b(k - k_b) \log(k_b - k) + \mathcal{O}(k - k_b), \quad k \rightarrow k_b, \\ C_b &:= \frac{1}{4} \frac{(U(t_b) - 2k_b)\sqrt{k_b - a(t_b)}}{\sqrt{-\frac{1}{2}b''(t_b)}}. \end{aligned} \tag{89}$$

*Proof:* This can be shown with the help of the contour integral formula (68). In particular, note that  $\Phi$  is continuous in the limits  $k \downarrow k_\alpha$  and  $k \uparrow k_b$ , but formulae (88) and (89) hold in full neighborhoods of the indicated limit point with only local branch cuts of the logarithms omitted. Note that  $2\pi C_\alpha = \tau'(k_\alpha)$  and  $2\pi C_b = -\tau'(k_b)$  so both  $C_\alpha$  and  $C_b$  are positive. ■

Combining Lemma 5 with Lemma 4 yields the following result.

LEMMA 6. *The function  $\tilde{\Gamma}(k) = Y^\epsilon(k)e^{-2i\Phi(k)/\epsilon}$  has an analytic continuation into the complex plane from right and left neighborhoods of  $k_\alpha$  and  $k_b$ , respectively, and for each  $\theta \in (0, \pi/2)$  and each  $\delta > 0$  sufficiently small,  $\tilde{\Gamma}(k_\alpha + re^{i\theta}) = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  and  $\tilde{\Gamma}(k_b - re^{-i\theta}) = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  both hold in the limit  $\epsilon \rightarrow 0$  with  $\epsilon > 0$ , uniformly for  $0 \leq r \leq \delta$ . (Of course by Schwarz reflection  $\tilde{\Gamma}(k^*)^*$  satisfies similar estimates along segments in the lower half-plane with endpoints  $k_\alpha$  and  $k_b$ .)*

*Proof:* This follows from the fact that  $Y^\epsilon(k)$  has an analytic continuation satisfying  $Y^\epsilon(k) = \mathcal{O}(\epsilon^{-1/2}\tau(k)^{1/2})$  near  $k_\alpha$  and  $k_b$  as long as  $\Re\{\tau(k)\} > 0$ . By Lemma 4 this estimate holds locally near  $k_\alpha$  or  $k_b$  as long as  $\Re\{k - k_\alpha\} > 0$  or  $\Re\{k_b - k\} > 0$  holds, respectively, and in each case we may replace  $\tau(k)$  in the estimate by either  $|k - k_\alpha|$  or  $|k - k_b|$ .

Using Lemma 5 shows that the problem boils down to estimating functions of a real variable,  $x \in (0, 1)$ , having the form

$$f^\epsilon(x) := \frac{x^{1/2}}{\epsilon^{1/2}} e^{x \log(x)/\epsilon}, \quad x \in (0, 1), \quad \epsilon > 0. \tag{90}$$

This function vanishes as  $x \downarrow 0$  for each  $\epsilon > 0$ , and it has two critical points for  $x > 0$ , only the smaller of the two being relevant for bounded  $x$ . This critical point is the global maximizer on  $(0, 1)$  and it satisfies  $x = x_c(\epsilon) \sim \epsilon/(2 \log(\epsilon^{-1}))$  as  $\epsilon \downarrow 0$ . It then follows that  $f^\epsilon(x_c(\epsilon)) = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  as  $\epsilon \downarrow 0$  by direct calculation. ■

The intuition behind this result is that while the exponential decay of  $e^{-2i\Phi(k)/\epsilon}$  in the upper half-plane is not uniform near  $k_\alpha$  or  $k_b$  (and in fact there

is no decay at all exactly at these two points), the factor  $Y^\epsilon(k)$  vanishes at these points, with the result being that the product is locally uniformly small with  $\epsilon > 0$ , albeit exhibiting a very slow rate of decay to zero.

4.1.2 *Proof of Theorem 1.* We now give the proof of Theorem 1. The strategy is to open a single lens about the entire interval  $(k_a, k_b)$  based upon the natural upper-lower factorization of the jump matrix:

$$\begin{aligned} & \begin{bmatrix} 1 - |\tilde{\Gamma}(k)|^2 & -\tilde{\Gamma}(k)^* e^{-2i\theta(k;x,t)/\epsilon} \\ \tilde{\Gamma}(k) e^{2i\theta(k;x,t)/\epsilon} & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -\tilde{\Gamma}(k)^* e^{-2i\theta(k;x,t)/\epsilon} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \tilde{\Gamma}(k) e^{2i\theta(k;x,t)/\epsilon} & 1 \end{bmatrix}, \quad k_a < k < k_b. \end{aligned} \quad (91)$$

However, technical modifications of the steepest descent method will be required because  $\tau$  has no analytic continuation from the real axis near the points  $k_0$  and  $k_\infty$ , and  $\Phi$  fails to be analytic at  $k = k_0$ .

We will in particular need a way to extend the three-times differentiable but nonanalytic function  $\Phi(k)$  into the complex plane from a real neighborhood of  $k = k_0$ . Let  $\Re\{k\} = k_r$  and  $\Im\{k\} = k_i$  denote the real and imaginary parts of the complex variable  $k$ . We follow the approach of [8] and first define a nonanalytic extension of  $\Phi(k_r)$  by the formula

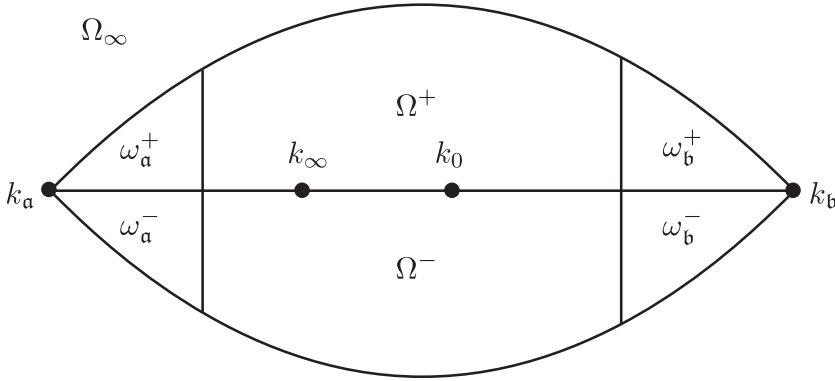
$$\hat{\Phi}_0(k_r, k_i) := \Phi(k_r) + ik_i \Phi'(k_r) + \frac{1}{2}(ik_i)^2 \Phi''(k_r). \quad (92)$$

Note that  $\hat{\Phi}_0(k_r, k_i)$  is nearly analytic close to the real axis  $k_i = 0$  in the sense that

$$\bar{\partial} \hat{\Phi}_0(k_r, k_i) := \frac{1}{2} \left( \frac{\partial}{\partial k_r} + i \frac{\partial}{\partial k_i} \right) \hat{\Phi}_0(k_r, k_i) = \frac{1}{4}(ik_i)^2 \Phi'''(k_r) = \mathcal{O}(k_i^2) \quad (93)$$

according to Lemma 3. Also according to Lemma 3, we may identify  $\Phi(k)$  with two distinct analytic functions,  $\Phi_a(k)$  denoting the analytic continuation of  $\Phi(k)$  from  $(k_a, k_0)$  and  $\Phi_b(k)$  denoting the analytic continuation of  $\Phi(k)$  from  $(k_0, k_b)$ . Note that for each fixed  $k_r \in (k_a, k_0)$  we have  $\hat{\Phi}_0(k_r, k_i) - \Phi_a(k) = \mathcal{O}(k_i^3)$  and for each fixed  $k_r \in (k_0, k_b)$  we have  $\hat{\Phi}_0(k_r, k_i) - \Phi_b(k) = \mathcal{O}(k_i^3)$ , with the error terms being uniform by Taylor's theorem for  $k_r$  in compact subsets of  $(k_a, k_b)$  bounded away from  $k_0$ . Now let  $\delta > 0$  be so small that  $[k_0 - 2\delta, k_0 + 2\delta] \subset (k_a, k_b)$ , and define the smooth bump function  $\mathcal{B} : \mathbb{R} \rightarrow [0, 1]$  such that  $\mathcal{B}$  is of class  $C^\infty$  and

$$\mathcal{B}(u) = \begin{cases} 1, & |u - k_0| < \delta \\ 0, & |u - k_0| > 2\delta. \end{cases} \quad (94)$$



**Figure 4.** The lens domains about the interval  $[k_a, k_b]$ .

Then we define an extension of  $\Phi$  into the upper half-plane near  $(k_a, k_b)$  as follows:

$$\hat{\Phi}(k_r, k_i) := \begin{cases} \mathcal{B}(k_r)\hat{\Phi}_0(k_r, k_i) + (1 - \mathcal{B}(k_r))\Phi_a(k), & k_a < k_r \leq k_0 \\ \mathcal{B}(k_r)\hat{\Phi}_0(k_r, k_i) + (1 - \mathcal{B}(k_r))\Phi_b(k), & k_0 \leq k_r < k_b. \end{cases} \quad (95)$$

By direct calculation,

$$\bar{\partial}\hat{\Phi}(k_r, k_i) = \begin{cases} \mathcal{B}(k_r)\bar{\partial}\hat{\Phi}_0(k_r, k_i) + \bar{\partial}\mathcal{B}(k_r) \cdot (\hat{\Phi}_0(k_r, k_i) - \Phi_a(k)), & k_a < k_r \leq k_0 \\ \mathcal{B}(k_r)\bar{\partial}\hat{\Phi}_0(k_r, k_i) + \bar{\partial}\mathcal{B}(k_r) \cdot (\hat{\Phi}_0(k_r, k_i) - \Phi_b(k)), & k_0 \leq k_r < k_b. \end{cases} \quad (96)$$

It follows that  $\bar{\partial}\hat{\Phi}(k_r, k_i)$  is  $\mathcal{O}(k_i^2)$  uniformly for  $k_r$  in compact subsets of  $(k_a, k_b)$ . Note that  $\hat{\Phi}(k_r, -k_i) = \hat{\Phi}(k_r, k_i)^*$  holds, as the nonanalytic analogue of the Schwarz reflection symmetry  $\Phi_{a,b}(k^*) = \Phi_{a,b}(k)^*$  of the real analytic functions  $\Phi_{a,b}$ .

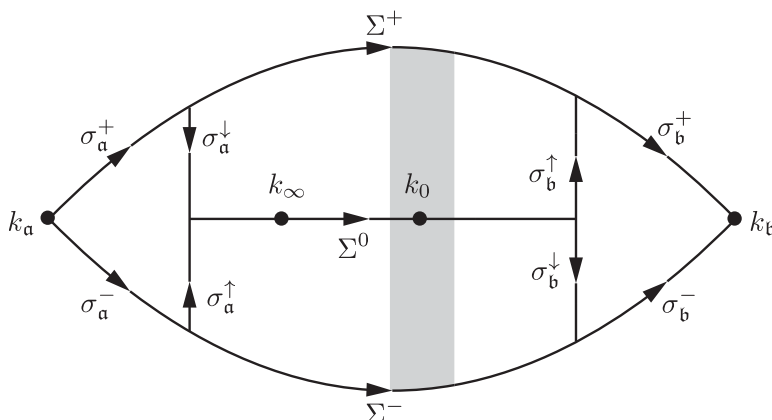
Based on the extension  $\hat{\Phi}$  we now make an explicit transformation of  $\tilde{\mathbf{M}}(k)$  to open lenses about  $(k_a, k_b)$ . Consider the domains illustrated in Figure 4. We define a new matrix unknown  $\mathbf{O}(k_r, k_i)$  as follows:

$$\mathbf{O}(k_r, k_i) := \tilde{\mathbf{M}}(k) \begin{bmatrix} 1 & 0 \\ -e^{2i(\theta(k;x,t) - \hat{\Phi}(k_r, k_i))/\epsilon} & 1 \end{bmatrix}, \quad k \in \Omega^+, \quad (97)$$

$$\mathbf{O}(k_r, k_i) := \tilde{\mathbf{M}}(k) \begin{bmatrix} 1 & -e^{2i(\hat{\Phi}(k_r, k_i) - \theta(k;x,t))/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \Omega^-, \quad (98)$$

$$\mathbf{O}(k_r, k_i) := \tilde{\mathbf{M}}(k) \begin{bmatrix} 1 & 0 \\ -Y_a^\epsilon(k)e^{2i(\theta(k;x,t) - \Phi_a(k))/\epsilon} & 1 \end{bmatrix}, \quad k \in \omega_a^+, \quad (99)$$

$$\mathbf{O}(k_r, k_i) := \tilde{\mathbf{M}}(k) \begin{bmatrix} 1 & 0 \\ -Y_b^\epsilon(k)e^{2i(\theta(k;x,t) - \Phi_b(k))/\epsilon} & 1 \end{bmatrix}, \quad k \in \omega_b^+, \quad (100)$$



**Figure 5.** The jump contour for the matrix function  $\mathbf{O}(k_r, k_i)$  is the union of eleven oriented arcs, labeled as shown. The domain in which  $\mathbf{O}(k_r, k_i)$  fails to be analytic is shaded.

$$\mathbf{O}(k_r, k_i) := \tilde{\mathbf{M}}(k) \begin{bmatrix} 1 & -Y_a^\epsilon(k)e^{2i(\Phi_a(k)-\theta(k;x,t))/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \omega_a^-, \quad (101)$$

$$\mathbf{O}(k_r, k_i) := \tilde{\mathbf{M}}(k) \begin{bmatrix} 1 & -Y_b^\epsilon(k)e^{2i(\Phi_b(k)-\theta(k;x,t))/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \omega_b^-, \quad (102)$$

and in the unbounded domain  $\Omega_\infty$  we set  $\mathbf{O}(k_r, k_i) := \tilde{\mathbf{M}}(k)$ . Here, the notation  $Y_{a,b}^\epsilon(k)$  refers to the two distinct analytic functions  $\tau(k) = \tau_a(k)$  defined near  $k = k_a$  and  $\tau(k) = \tau_b(k)$  defined near  $k = k_b$  (see Lemma 2 and Lemma 4). Indeed, the domains  $\omega_a^\pm$  and  $\omega_b^\pm$  are chosen small enough to exclude both  $k_0$  and  $k_\infty$ , the two points of nonanalyticity of  $\tau : (k_a, k_b) \rightarrow \mathbb{R}_+$ . By making  $\delta > 0$  smaller if necessary, we also ensure that these domains have no intersection with the vertical strip  $|k_r - k_0| \leq 2\delta$ , in which  $\mathbf{O}(k_r, k_i)$  fails to be analytic. The matrix  $\mathbf{O}(k_r, k_i)$  has jump discontinuities across a contour  $\Sigma$  illustrated in Figure 5. Note that the real segment common to the boundary of the domains  $\omega_a^\pm$  and the real segment common to the boundary of the domains  $\omega_b^\pm$  are not part of the jump contour  $\Sigma$  as it is easy to check that  $\mathbf{O}(k_r, k_i)$  is continuous across these segments.

We claim that the matrix  $\mathbf{O}(k_r, k_i)$  satisfies the conditions of a hybrid Riemann–Hilbert- $\bar{\partial}$  problem of small-norm type. This problem is the following.

**Riemann–Hilbert- $\bar{\partial}$  problem 3.** Find a  $2 \times 2$  matrix  $\mathbf{O}(k_r, k_i)$  with the following properties:

- Continuity:**  $\mathbf{O}$  is continuous in each connected component of  $\mathbb{R}^2 \setminus \Sigma$  and takes continuous boundary values  $\mathbf{O}_+$  and  $\mathbf{O}_-$  on each oriented arc of  $\Sigma$  from the left and right, respectively.



**Jump Condition:** On each oriented arc of  $\Sigma$  the boundary values are related by  $\mathbf{O}_+(k_r, k_i) = \mathbf{O}_-(k_r, k_i)\mathbf{J}_0(k_r, k_i)$  (see below for the explicit definition of  $\mathbf{J}_0(k_r, k_i)$ ).

**Nonanalyticity:** In each connected component of  $\mathbb{R}^2 \setminus \Sigma$  the matrix  $\mathbf{O}$  satisfies  $\bar{\partial}\mathbf{O}(k_r, k_i) = \mathbf{O}(k_r, k_i)\mathbf{W}(k_r, k_i)$  (see below for the explicit definition of  $\mathbf{W}(k_r, k_i)$ ).

**Normalization:**  $\mathbf{O}(k_r, k_i) \rightarrow \mathbb{I}$  as  $(k_r, k_i) \rightarrow \infty$  in  $\mathbb{R}^2$ .

The jump matrix  $\mathbf{J}_0(k_r, k_i)$  is defined explicitly on each arc of  $\Sigma$  simply by using the definitions (97)–(102) and the jump condition satisfied by  $\tilde{\mathbf{M}}$  across the segment  $(k_a, k_b)$  according to Riemann–Hilbert problem 2. The result is the following:

$$\mathbf{J}_0(k_r, k_i) := \begin{bmatrix} 1 & 0 \\ Y_{a,b}^\epsilon(k) e^{2i(\theta(k;x,t) - \Phi_{a,b}(k))/\epsilon} & 1 \end{bmatrix}, \quad k \in \sigma_{a,b}^+, \quad (103)$$

$$\mathbf{J}_0(k_r, k_i) := \begin{bmatrix} 1 & -Y_{a,b}^\epsilon(k) e^{2i(\Phi_{a,b}(k) - \theta(k;x,t))/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \sigma_{a,b}^-, \quad (104)$$

$$\mathbf{J}_0(k_r, k_i) := \begin{bmatrix} 1 & 0 \\ (Y_{a,b}^\epsilon(k) - 1) e^{2i(\theta(k;x,t) - \Phi_{a,b}(k))/\epsilon} & 1 \end{bmatrix}, \quad k \in \sigma_a^\downarrow, \sigma_b^\uparrow, \quad (105)$$

$$\mathbf{J}_0(k_r, k_i) := \begin{bmatrix} 1 & (1 - Y_{a,b}^\epsilon(k)) e^{2i(\Phi_{a,b}(k) - \theta(k;x,t))/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \sigma_a^\uparrow, \sigma_b^\downarrow, \quad (106)$$

$$\mathbf{J}_0(k_r, k_i) := \begin{bmatrix} 1 & 0 \\ e^{2i(\theta(k;x,t) - \hat{\Phi}(k_r, k_i))/\epsilon} & 1 \end{bmatrix}, \quad k \in \Sigma^+, \quad (107)$$

$$\mathbf{J}_0(k_r, k_i) := \begin{bmatrix} 1 & -e^{2i(\hat{\Phi}(k_r, k_i) - \theta(k;x,t))/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \Sigma^-, \quad (108)$$

and, finally, using the fact that  $\hat{\Phi}(k_r, 0) = \Phi(k_r)$ ,

$$\mathbf{J}_0(k_r, k_i) := \begin{bmatrix} 1 & (1 - Y^\epsilon(k_r)) e^{2i(\Phi(k_r) - \theta(k_r;x,t))/\epsilon} \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ (Y^\epsilon(k_r) - 1) e^{2i(\theta(k_r;x,t) - \Phi(k_r))/\epsilon} & 1 \end{bmatrix}, \quad k \in \Sigma^0 \quad (109)$$

Consider the jump matrix defined by (103)–(104). Note that since  $\Im\{\theta(k;x,t)\} = o(\Im\{\Phi_{a,b}\})$  near  $k = k_{a,b}$  according to Lemma 5, it follows from Lemma 6 that for all  $(x,t) \in \mathbb{R}^2$ ,  $\mathbf{J}_0 - \mathbb{I} = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  holds uniformly on the four contour arcs  $\sigma_{a,b}^{+,-}$ , provided the lens opens with an acute nonzero

angle and the vertical contours  $\sigma_{a,b}^{\uparrow,\downarrow}$  are placed close enough to the respective endpoint  $k_{a,b}$ . Similarly, since  $\tau(k) > 0$  is bounded away from zero while  $\Phi(k)$  and  $\theta(k; x, t)$  are real for  $k \in \Sigma^0$ , it is easy to see from (109) that  $\mathbf{J}_0 - \mathbb{I}$  is uniformly exponentially small on  $\Sigma^0$  in the limit  $\epsilon \downarrow 0$ , again independently of  $(x, t) \in \mathbb{R}^2$ . Controlling the jump matrix  $\mathbf{J}_0$  on the remaining arcs of  $\Sigma$  requires conditions on  $(x, t) \in \mathbb{R}^2$  as we will see below.

The matrix  $\mathbf{W}(k_r, k_i)$  is defined explicitly by applying the operator  $\bar{\partial}$  to the formulae (97)–(98). The result is

$$\mathbf{W}(k_r, k_i) := \begin{bmatrix} 0 & 0 \\ 2i\epsilon^{-1}\bar{\partial}\hat{\Phi}(k_r, k_i) \cdot e^{2i(\theta(k;x,t)-\hat{\Phi}(k_r,k_i))/\epsilon} & 0 \end{bmatrix}, \quad k \in \Omega^+, \quad (110)$$

and

$$\mathbf{W}(k_r, k_i) := \begin{bmatrix} 0 & -2i\epsilon^{-1}\bar{\partial}\hat{\Phi}(k_r, k_i) \cdot e^{2i(\hat{\Phi}(k_r,k_i)-\theta(k;x,t))/\epsilon} \\ 0 & 0 \end{bmatrix}, \quad k \in \Omega^-, \quad (111)$$

and in all other connected components of  $\mathbb{R}^2 \setminus \Sigma$ ,  $\mathbf{W}(k_r, k_i) := \mathbf{0}$ . In particular,  $\mathbf{W}$  has compact support.

Riemann–Hilbert- $\bar{\partial}$  problem 3 is of small-norm type in the sense that, as a consequence of the conditions  $t = 0$  and  $x > 0$ , the jump matrix  $\mathbf{J}_0$  defined on the compact contour  $\Sigma$  satisfies  $\|\mathbf{J}_0 - \mathbb{I}\|_{L^\infty(\Sigma)} = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  as  $\epsilon \rightarrow 0$ , and at the same time the matrix  $\mathbf{W}$  defined on  $\mathbb{R}^2 \setminus \Sigma$  satisfies  $\|\mathbf{W}\|_{L^\infty(\mathbb{R}^2 \setminus \Sigma)} = \mathcal{O}(\epsilon)$  as  $\epsilon \rightarrow 0$ . Indeed,

$$\theta'(k; x, 0) - \Phi'(k) = x - \Phi'(k) \geq x > 0, \quad k_a < k < k_b, \quad (112)$$

according to Lemma 3. This immediately implies, by the Cauchy–Riemann equations applied to the real analytic functions  $\theta(k; x, t) - \Phi_{a,b}(k)$  near the real  $k$ -axis, that the exponential factors appearing in the formulae (105)–(106) are bounded in modulus by 1 provided the lens is sufficiently thin (independent of  $\epsilon$ ). Since the factors  $Y_{a,b}^\epsilon(k) - 1$  are exponentially small as they are on the real axis if the lens is thin enough, we conclude that for  $x > 0$  and  $t = 0$ ,  $\mathbf{J}_0 - \mathbb{I}$  is uniformly exponentially small for  $k \in \sigma_a^\downarrow \cup \sigma_b^\uparrow \cup \sigma_a^\uparrow \cup \sigma_b^\downarrow$ . Finally, from (95) and the Cauchy–Riemann equations, we see that

$$\Im\{\hat{\Phi}(k_r, k_i)\} = \Phi'(k_r)k_i + \mathcal{O}(k_i^2), \quad k_i \rightarrow 0, \quad (113)$$

a fact which, taken together with (107)–(108) for  $x > 0$  and  $t = 0$  shows that  $\mathbf{J}_0 - \mathbb{I}$  is uniformly exponentially small in the limit  $\epsilon \downarrow 0$  for  $k \in \Sigma^\pm$ . Combining these estimates yields the claimed  $L^\infty(\Sigma)$  bound for  $\mathbf{J}_0 - \mathbb{I}$ . Similarly, for  $t = 0$  and for each  $x > 0$  we obtain exponential decay of the exponential factors in (110)–(111), so combining this fact with the fact that  $\bar{\partial}\hat{\Phi}(k_r, k_i) = \mathcal{O}(k_i^2)$  holds for  $k \in \Omega^+ \cup \Omega^-$ , we obtain ( $\|\cdot\|$  denotes any matrix norm)

$$\|\mathbf{W}(k_r, k_i)\| \leq K \frac{k_i^2}{\epsilon} e^{-C|k_i|/\epsilon}, \quad k \in \Omega^+ \cup \Omega^-, \quad (114)$$

where  $C = C(x_0) > 0$  for  $x > x_0$  and  $K > 0$  is independent of  $x$  and  $\epsilon$ . The claimed  $L^\infty(\mathbb{R}^2 \setminus \Sigma)$  estimate of  $\mathbf{W}$  follows immediately because  $r^2 e^{-Cr}$  is uniformly bounded for all  $r > 0$  (recall  $\mathbf{W}(k_r, k_i) = 0$  for  $k \notin \Omega^+ \cup \Omega^-$ ).

One makes use of the estimates  $\|\mathbf{J}_0 - \mathbb{I}\|_{L^\infty(\Sigma)} = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  and  $\|\mathbf{W}\|_{L^\infty(\mathbb{R}^2 \setminus \Sigma)} = \mathcal{O}(\epsilon)$  as follows. The strategy is to solve the hybrid Riemann–Hilbert– $\bar{\partial}$  problem by first solving the “ $\bar{\partial}$  part” and then using the result to obtain a standard Riemann–Hilbert problem of  $L^2$  small-norm type. We therefore consider the following auxiliary  $\bar{\partial}$  problem:

$\bar{\partial}$  **Problem 4.** Find a  $2 \times 2$  matrix  $\dot{\mathbf{O}}(k_r, k_i)$  with the following properties:

**Continuity:**  $\dot{\mathbf{O}} : \mathbb{R}^2 \rightarrow \mathbb{C}^{2 \times 2}$  is continuous.

**Nonanalyticity:**  $\bar{\partial} \dot{\mathbf{O}}(k_r, k_i) = \dot{\mathbf{O}}(k_r, k_i) \mathbf{W}(k_r, k_i)$  holds in the distributional sense.

**Normalization:**  $\dot{\mathbf{O}}(k_r, k_i) \rightarrow \mathbb{I}$  as  $(k_r, k_i) \rightarrow \infty$  in  $\mathbb{R}^2$ .

We solve for  $\dot{\mathbf{O}}$  by setting up an integral equation involving the solid Cauchy transform:

$$\begin{aligned} \dot{\mathbf{O}}(k_r, k_i) &= \mathbb{I} + \mathcal{K} \dot{\mathbf{O}}(k_r, k_i), \\ \mathcal{K} \mathbf{F}(k_r, k_i) &:= -\frac{1}{\pi} \iint_{\Omega^+ \cup \Omega^-} \frac{\mathbf{F}(k'_r, k'_i) \mathbf{W}(k'_r, k'_i) dA(k'_r, k'_i)}{k' - k}, \end{aligned} \quad (115)$$

where  $dA(k_r, k_i) = dk_r dk_i$  denotes the area element. It can be shown that the integral equation (115) is in fact equivalent to the formulation of  $\bar{\partial}$  Problem 4. The operator norm of  $\mathcal{K}$  acting on  $L^\infty(\mathbb{R}^2)$  is easily estimated based on the fact that Cauchy kernel is locally integrable in two dimensions. Thus,

$$\|\mathcal{K}\|_{L^\infty(\mathbb{R}^2) \circlearrowleft} \leq \frac{1}{\pi} \|\mathbf{W}\|_{L^\infty(\mathbb{R}^2)} \sup_{(k_r, k_i) \in \mathbb{R}^2} \iint_{\Omega^+ \cup \Omega^-} \frac{dA(k'_r, k'_i)}{|k' - k|}. \quad (116)$$

The latter supremum is finite and depends only on the bounded domain  $\Omega^+ \cup \Omega^-$  as the double integral is continuous and decays as  $|k|^{-1}$  as  $(k_r, k_i) \rightarrow \infty$  in  $\mathbb{R}^2$ . From the bound  $\|\mathbf{W}\|_{L^\infty(\mathbb{R}^2)} = \mathcal{O}(\epsilon)$  it follows also that  $\|\mathcal{K}\|_{L^\infty(\mathbb{R}^2) \circlearrowleft} = \mathcal{O}(\epsilon)$  as  $\epsilon \downarrow 0$ . It follows that for  $\epsilon > 0$  sufficiently small, the operator  $1 - \mathcal{K}$  is invertible by Neumann series convergent in  $L^\infty(\mathbb{R}^2)$ . Since each term of the series is continuous, so is the sum  $\dot{\mathbf{O}}(k_r, k_i)$  of the series for  $(1 - \mathcal{K})^{-1}$  applied to  $\mathbb{I}$ . Furthermore, we obtain the important estimate  $\|\dot{\mathbf{O}} - \mathbb{I}\|_{L^\infty(\mathbb{R}^2)} = \mathcal{O}(\epsilon)$ , which in particular implies that  $\dot{\mathbf{O}}(k_r, k_i)^{-1}$  exists for sufficiently small  $\epsilon$  as a continuous function on  $\mathbb{R}^2$  that satisfies  $\|\dot{\mathbf{O}}^{-1} - \mathbb{I}\|_{L^\infty(\mathbb{R}^2)} = \mathcal{O}(\epsilon)$ . Finally, compact support of  $\mathbf{W}$  ensures that  $\dot{\mathbf{O}}(k_r, k_i)$  has a convergent Laurent series in descending powers of  $k$  for  $|k|$  sufficiently large, and in particular we obtain  $\dot{\mathbf{O}}(k_r, k_i) = \mathbb{I} + k^{-1} \dot{\mathbf{O}}_1 + \mathcal{O}(k^{-2})$  as  $k \rightarrow \infty$ , where  $\|\dot{\mathbf{O}}_1\| = \mathcal{O}(\epsilon)$ .

We now use the unique solution  $\dot{\mathbf{O}}(k_r, k_i)$  of  $\bar{\partial}$  Problem 4 as obtained above to convert the hybrid Riemann–Hilbert– $\bar{\partial}$  problem 3 into a standard

Riemann–Hilbert problem that we can show is of  $L^2$  small norm type. Indeed, consider the matrix function  $\mathbf{E}(k_r, k_i)$  defined in terms of  $\mathbf{O}(k_r, k_i)$  solving Riemann–Hilbert- $\bar{\partial}$  problem 3 and  $\dot{\mathbf{O}}(k_r, k_i)$  solving  $\bar{\partial}$  problem 4 by

$$\mathbf{E}(k_r, k_i) := \mathbf{O}(k_r, k_i)\dot{\mathbf{O}}(k_r, k_i)^{-1}, \quad (k_r, k_i) \in \mathbb{R}^2 \setminus \Sigma. \quad (117)$$

By direct calculation,

$$\begin{aligned} \bar{\partial}\mathbf{E}(k_r, k_i) &= [\bar{\partial}\mathbf{O}(k_r, k_i)]\dot{\mathbf{O}}(k_r, k_i)^{-1} \\ &\quad - \mathbf{O}(k_r, k_i)\dot{\mathbf{O}}(k_r, k_i)^{-1} [\bar{\partial}\dot{\mathbf{O}}(k_r, k_i)]\dot{\mathbf{O}}(k_r, k_i)^{-1} \\ &= \mathbf{O}(k_r, k_i)\mathbf{W}(k_r, k_i)\dot{\mathbf{O}}(k_r, k_i)^{-1} - \mathbf{O}(k_r, k_i)\mathbf{W}(k_r, k_i)\dot{\mathbf{O}}(k_r, k_i)^{-1} \\ &= \mathbf{0}, \quad (k_r, k_i) \in \mathbb{R}^2 \setminus \Sigma, \end{aligned} \quad (118)$$

and hence  $\mathbf{E}$  is analytic in each connected component of  $\mathbb{R}^2 \setminus \Sigma$ . In light of this result, we will henceforth use the notation  $\mathbf{E} = \mathbf{E}(k)$  with  $k = k_r + ik_i$ . It is a direct matter to calculate the jump conditions satisfied by  $\mathbf{E}$  across the arcs of the contour  $\Sigma$  in terms of the jump matrix  $\mathbf{J}_0$  for  $\mathbf{O}$  and the function  $\dot{\mathbf{O}}$  restricted to  $\Sigma$ , and to calculate the asymptotic behavior of  $\mathbf{E}$  as  $k \rightarrow \infty$ . We deduce that  $\mathbf{E}$  satisfies the following (pure) Riemann–Hilbert problem:

**Riemann–Hilbert problem 5.** Find a  $2 \times 2$  matrix  $\mathbf{E}(k)$  with the following properties:

**Analyticity:**  $\mathbf{E}$  is analytic in each connected component of  $\mathbb{C} \setminus \Sigma$ , and takes continuous boundary values  $\mathbf{E}_+(k)$  ( $\mathbf{E}_-(k)$ ) from the left (right) at each non-self-intersection point  $k$  of  $\Sigma$ .

**Jump condition:** On each oriented arc of  $\Sigma$  the boundary values are related by the jump condition  $\mathbf{E}_+(k) = \mathbf{E}_-(k)\mathbf{J}(k_r, k_i)$ , where

$$\mathbf{J}(k_r, k_i) := \dot{\mathbf{O}}(k_r, k_i)\mathbf{J}_0(k_r, k_i)\dot{\mathbf{O}}(k_r, k_i)^{-1}. \quad (119)$$

**Normalization:**  $\mathbf{E}(k) \rightarrow \mathbb{I}$  as  $k \rightarrow \infty$  in  $\mathbb{C}$ .

Since  $\dot{\mathbf{O}}(k_r, k_i)$  and  $\dot{\mathbf{O}}(k_r, k_i)^{-1}$  are uniformly bounded independent of  $\epsilon$  for  $\epsilon > 0$  sufficiently small, it follows immediately from the estimate  $\|\mathbf{J}_0 - \mathbb{I}\|_{L^\infty(\Sigma)} = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  that also  $\|\mathbf{J} - \mathbb{I}\|_{L^\infty(\Sigma)} = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  as  $\epsilon \downarrow 0$ . Since  $\Sigma$  is compact, this condition implies the unique solvability of Riemann–Hilbert problem 5 as a small-norm problem in the  $L^2$  sense. See [16] or Appendix B of [17] for details. In particular,  $\mathbf{E}(k) = \mathbb{I} + k^{-1}\mathbf{E}_1 + \mathcal{O}(k^{-2})$  as  $k \rightarrow \infty$  with  $\|\mathbf{E}_1\| = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$ .

Finally, we consider the matrix  $\tilde{\mathbf{M}}(k)$  solving Riemann–Hilbert problem 2 for large  $|k|$ . We obtain the exact formula

$$\tilde{\mathbf{M}}(k) = \mathbf{O}(k_r, k_i) = \mathbf{E}(k)\dot{\mathbf{O}}(k_r, k_i), \quad \text{for } |k| \text{ sufficiently large,} \quad (120)$$

from which we compute

$$\begin{aligned} \tilde{q}^\epsilon(x, 0) &= 2i \lim_{k \rightarrow \infty} k[\mathbf{E}(k)\dot{\mathbf{O}}(k_r, k_i)]_{12} = 2i[\mathbf{E}_1 + \dot{\mathbf{O}}_1]_{12} \\ &= \mathcal{O}((\log(\epsilon^{-1}))^{-1/2}), \quad \epsilon \downarrow 0. \end{aligned} \tag{121}$$

As the error term is uniform for  $x \geq x_0 > 0$ , the proof is complete.

*4.1.3 Proof of Corollary 1.* To prove Corollary 1, we simply observe that the only dependence on  $(x, t) \in \mathbb{R}^2$  in the proof of Theorem 1 given in Section 4.1.2 involved the inequality (112), and it is not hard to see that this inequality holds also for some nonzero  $t$ . More generally,

$$\theta'(k; x, t) - \Phi'(k) = x + 4tk - \Phi'(k), \tag{122}$$

and as  $\Phi'(\cdot)$  is a function with maximum value zero at  $k = k_0$  only and tending to  $-\infty$  as  $k \downarrow k_a$  and  $k \uparrow k_b$ , given  $t > 0$  there will be some finite  $X(t) > 0$  such that  $\theta'(k; x, t) - \Phi'(k) > 0$  holds strictly on  $k_a < k < k_b$  for  $x > X(t)$  but fails for some  $k \in (k_a, k_b)$  if  $x \leq X(t)$ . The rest of the proof of Theorem 1 then goes through unchanged, with the same result, and the proof of Corollary 1 is complete.

We conclude this short section by obtaining explicit and simple asymptotic formulae for the boundary curve  $x = X(t)$  valid for small and large  $t > 0$ . We note firstly that  $-\Phi'(k)$  is certainly locally convex (i) for  $k - k_0$  small, because  $\Phi'''(k_0) = 16k_0h_0^{-2} < 0$  (see (78)) and according to Lemma 3,  $\Phi'''(\cdot)$  is continuous on  $(k_a, k_b)$ , and also (ii) for  $k - k_a$  small, where according to Lemma 5 we have  $\Phi'(k) \sim C_a \log(k - k_a)$  where  $C_a > 0$  is given in (88). In the former case, the slope of the tangent line of  $-\Phi'(k)$  is small, while in the latter case the slope is large. Therefore, for small or large positive  $t$  we can apply the following steps to obtain  $X(t)$ : first solve the equation  $\Phi''(k) = 4t$  for  $k = k(t)$ , and then obtain  $X(t) = \Phi'(k(t)) - 4tk(t)$ .

When  $t > 0$  is small, we expect  $k - k_0$  to be small, and since

$$\Phi''(k) = \Phi'''(k_0)(k - k_0) + o(k - k_0) = 16k_0h_0^{-2}(k - k_0) + o(k - k_0), \quad k \rightarrow k_0, \tag{123}$$

from  $\Phi''(k) = 4t$  we obtain

$$k(t) = k_0 + \frac{4}{\Phi'''(k_0)}t + o(t) = k_0 + \frac{h_0^2}{4k_0}t + o(t), \quad t \downarrow 0. \tag{124}$$

Since  $\Phi'(k_0) = 0$ , integrating (123) and substituting from (124) yields

$$X(t) = \Phi'(k(t)) - 4tk(t) = X_0(t) + o(t^2), \quad t \downarrow 0, \tag{125}$$

where the asymptote to  $X(t)$  for small  $t$  is defined by

$$X_0(t) := -4k_0t - \frac{h_0^2}{2k_0}t^2. \tag{126}$$

On the other hand, if  $t > 0$  is large, then we expect  $k - k_a$  will be small. The equation to be solved for  $k = k(t)$  in this case is then  $\Phi''(k) = 4t$ , where now

$$\Phi''(k) = \frac{C_a}{k - k_a} + \mathcal{O}(1), \quad k \downarrow k_a. \tag{127}$$

Therefore,

$$k(t) = k_a + \frac{1}{4}C_a t^{-1} + \mathcal{O}(t^{-2}), \quad t \rightarrow +\infty. \tag{128}$$

Using  $\Phi'(k) = C_a \log(k - k_a) + \mathcal{O}(1)$  as  $k \downarrow k_a$ , we then have

$$X(t) = \Phi'(k(t)) - 4tk(t) = X_\infty(t) + \mathcal{O}(1), \quad t \rightarrow +\infty, \tag{129}$$

where the asymptote to  $X(t)$  for large  $t$  is given by

$$X_\infty(t) := -4k_a t - C_a \log(t). \tag{130}$$

#### 4.2. Asymptotic behavior of $\tilde{q}^\epsilon(0, t)$ for $t > 0$ and related analysis

4.2.1. *General methodology: the complex phase function  $g$ .* A general strategy to the analysis of the solution of Riemann–Hilbert problem 2 in the semiclassical limit  $\epsilon \downarrow 0$  follows the basic approach outlined in [18], which is based on the introduction of a scalar *complex phase function*  $g = g(k)$  having the following basic properties:

- $g$  is analytic for  $k \in \mathbb{C} \setminus [k_a, k_b]$  and takes continuous boundary values  $g_\pm$  on  $(k_a, k_b)$  from  $\mathbb{C}_\pm$ ,
- $g(k) \rightarrow 0$  as  $k \rightarrow \infty$ ,
- $g(k^*) = g(k)^*$  (Schwarz symmetry), and very importantly,
- $g$  is independent of  $\epsilon$  (although it will generally depend on  $x$  and  $t$ ).

One introduces such a function  $g$  into Riemann–Hilbert problem 2 by making the substitution

$$\tilde{\mathbf{M}}(k) := \mathbf{N}(k)e^{ig(k)\sigma_3/\epsilon}. \tag{131}$$

The basic properties of  $g$  listed above are by no means sufficient to determine  $g$  (this is why we do not formulate them as a proper Riemann–Hilbert problem), and the point is that one should use the freedom of choice of  $g$  to try to bring  $\mathbf{N}(k)$  into a form amenable for asymptotic analysis in the limit  $\epsilon \downarrow 0$ .

The transformation (131) implies that  $\mathbf{N}$  is analytic where  $\tilde{\mathbf{M}}$  is, takes boundary values in the same way, satisfies exactly the same normalization condition as  $k \rightarrow \infty$  as does  $\tilde{\mathbf{M}}$ , and satisfies a modified jump condition:

$$\mathbf{N}_+(k) = \mathbf{N}_-(k) \begin{bmatrix} e^{2(\Delta(k)-\tau(k))/\epsilon} & -Y^\epsilon(k)e^{-2i\phi(k)/\epsilon} \\ Y^\epsilon(k)e^{2i\phi(k)/\epsilon} & e^{-2\Delta(k)/\epsilon} \end{bmatrix}, \quad k_a < k < k_b, \tag{132}$$

where

$$\begin{aligned}\phi(k) &:= \theta(k; x, t) - \Phi(k) - \frac{1}{2}(g_+(k) + g_-(k)) \in \mathbb{R} \quad \text{and} \\ \Delta(k) &:= -i\frac{1}{2}(g_+(k) - g_-(k)) \in \mathbb{R}.\end{aligned}\tag{133}$$

The strategy of [18] is to try to choose  $g$  so that the interval  $(k_a, k_b)$  splits into a finite number of subintervals of three distinct types:

**Voids:** intervals in which  $\Delta(k) \equiv 0$  and  $\phi'(k) > 0$ .

**Bands:** intervals in which  $0 < \Delta(k) < \tau(k)$  and  $\phi'(k) \equiv 0$ .

**Saturated regions:** intervals in which  $\Delta(k) \equiv \tau(k)$  and  $\phi'(k) < 0$ .

In each void interval  $V$ , the modified jump matrix admits an “upper-lower” factorization because  $\Delta(k) \equiv 0$ :

$$\begin{aligned}& \begin{bmatrix} e^{2(\Delta(k)-\tau(k))/\epsilon} & -Y^\epsilon(k)e^{-2i\phi(k)/\epsilon} \\ Y^\epsilon(k)e^{2i\phi(k)/\epsilon} & e^{-2\Delta(k)/\epsilon} \end{bmatrix} \\ &= \begin{bmatrix} 1 & -Y^\epsilon(k)e^{-2i\phi(k)/\epsilon} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ Y^\epsilon(k)e^{2i\phi(k)/\epsilon} & 1 \end{bmatrix}, \quad k \in V,\end{aligned}\tag{134}$$

and the monotonicity condition  $\phi'(k) > 0$  suggests that the first (second) factor has a continuation into the lower (upper) half-plane that is exponentially close to the identity matrix. In each band interval  $B$ , the modified jump matrix is obviously exponentially close to a constant off-diagonal matrix due to the inequalities  $0 < \Delta(k) < \tau(k)$ :

$$\begin{aligned}& \begin{bmatrix} e^{2(\Delta(k)-\tau(k))/\epsilon} & -Y^\epsilon(k)e^{-2i\phi(k)/\epsilon} \\ Y^\epsilon(k)e^{2i\phi(k)/\epsilon} & e^{-2\Delta(k)/\epsilon} \end{bmatrix} = \begin{bmatrix} 0 & -e^{-2i\phi_B/\epsilon} \\ e^{2i\phi_B/\epsilon} & 0 \end{bmatrix} \\ & \quad + \text{exponentially small terms}, \quad k \in B,\end{aligned}\tag{135}$$

where  $\phi_B$  denotes the constant value of  $\phi(k)$  in the band  $B$ . Finally, in each saturated interval  $S$ , the modified jump matrix admits a “lower-upper” factorization because  $\Delta(k) \equiv \tau(k)$ :

$$\begin{aligned} & \begin{bmatrix} e^{2(\Delta(k)-\tau(k))/\epsilon} & -Y^\epsilon(k)e^{-2i\phi(k)/\epsilon} \\ Y^\epsilon(k)e^{2i\phi(k)/\epsilon} & e^{-2\Delta(k)/\epsilon} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ Y^\epsilon(k)e^{2i\phi(k)/\epsilon} & 1 \end{bmatrix} \begin{bmatrix} 1 & -Y^\epsilon(k)e^{-2i\phi(k)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in S. \end{aligned} \quad (136)$$

The inequality  $\phi'(k) < 0$  then suggests that the first (second) factor can be continued into the lower (upper) half-plane, becoming an exponentially small perturbation of the identity matrix  $\mathbb{I}$ .

The function  $g$  may be constructed by temporarily setting aside the inequalities involved with the voids, bands, and saturated regions. We suppose that there are  $N + 1$  bands in  $(k_a, k_b)$  that we denote by  $B_j := [\alpha_j, \beta_j]$  with  $k_a < \alpha_0 < \beta_0 < \alpha_1 < \beta_1 < \dots < \alpha_N < \beta_N < k_b$ . The complementary intervals are either voids or saturated regions. The boundary values of the function  $g(k)$  then satisfy

- $g_+(k) - g_-(k) = 0$  which implies  $g'_+(k) - g'_-(k) = 0$  for  $k$  in voids.
- $g'_+(k) + g'_-(k) = 2\theta'(k; x, t) - 2\Phi'(k)$  for  $k$  in bands.
- $g_+(k) - g_-(k) = 2i\tau(k)$  which implies  $g'_+(k) - g'_-(k) = 2i\tau'(k)$  for  $k$  in saturated regions.

Therefore, we know the value of  $g'_+(k) - g'_-(k)$  everywhere in the interval  $(k_a, k_b)$  with the exception of the band intervals, where instead we know  $g'_+(k) + g'_-(k)$ . Denoting by  $r(k)$  the function analytic in the domain  $\mathbb{C} \setminus (B_0 \cup B_1 \cup \dots \cup B_N)$  that satisfies  $r(k)^2 = (k - \alpha_0)(k - \beta_0) \dots (k - \alpha_N)(k - \beta_N)$  and  $r(k) = k^{N+1} + \mathcal{O}(k^N)$  as  $k \rightarrow \infty$ , we may consider instead of  $g'(k)$  the related function  $m(k) := g'(k)/r(k)$ . This function is analytic for  $k \in \mathbb{C} \setminus [k_a, k_b]$  and since  $r$  changes sign across the band intervals and is otherwise analytic,  $m$  satisfies

$$\begin{aligned} & m_+(k) - m_-(k) \\ &= \begin{cases} 0, & k \text{ in voids (and outside of } [k_a, k_b]) \\ \frac{2\theta'(k; x, t) - 2\Phi'(k)}{r_+(k)}, & k \text{ in bands} \\ \frac{2i\tau'(k)}{r(k)}, & k \text{ in saturated regions.} \end{cases} \end{aligned} \quad (137)$$

Note that  $g'(k)$  must decay as  $\mathcal{O}(k^{-2})$  as  $k \rightarrow \infty$  because  $g(k) \rightarrow 0$  in this limit; this implies that  $m(k) = \mathcal{O}(k^{-(N+3)})$  for large  $k$ , and in particular  $m(k) = o(1)$ . Therefore,  $m$  is necessarily given in terms of the difference of its boundary values explicitly written in (137) by the Plemelj formula, which implies that

$$g'(k) = \frac{r(k)}{\pi i} \int_B \frac{(\theta'(l; x, t) - \Phi'(l)) dl}{r_+(l)(l - k)} + \frac{r(k)}{\pi} \int_S \frac{\tau'(l) dl}{r(l)(l - k)}, \quad (138)$$



where here  $B$  denotes the union of all bands and  $S$  denotes the union of all saturated regions. Expanding the Cauchy kernel  $(l - k)^{-1}$  in geometric series for large  $k$ , we see that the condition  $g'(k) = \mathcal{O}(k^{-2})$  as  $k \rightarrow \infty$  is equivalent to the following *moment conditions*:

$$m_n := \int_B \frac{(\theta'(k; x, t) - \Phi'(k))k^{n-1} dk}{ir_+(k)} + \int_S \frac{\tau'(k)k^{n-1} dk}{r(k)} = 0, \quad n = 1, 2, \dots, N + 2. \tag{139}$$

Subject to these conditions,  $g'(k)$  is integrable at infinity, and  $g(k)$  may be expressed as a contour integral:

$$g(k) = \int_{\infty}^k g'(l) dl, \tag{140}$$

where  $g'(k)$  is explicitly given by (138). Equations (139) are  $N + 2$  conditions on the  $2N + 2$  unknown endpoints of the bands  $B_0, \dots, B_N$ . In general, additional conditions arise in order to get the integration constants right so that instead of just  $g'_+(k) - g'_-(k) = 0$  in voids we actually have  $g_+(k) - g_-(k) = 0$ , and so that instead of just  $g'_+(k) - g'_-(k) = 2i\tau'(k)$  in saturated regions we actually have  $g_+(k) - g_-(k) = 2i\tau(k)$ . Since  $\tau(k_{\alpha, \beta}) = 0$ , no additional conditions are required if there is only one band, i.e.,  $N = 0$ , in which case the moment conditions (139) may determine the band endpoints  $\alpha = \alpha_0$  and  $\beta = \beta_0$ . For the purposes of the proofs of Theorem 2 and Corollary 2, we only consider this case.

Because  $\theta'(k; x, t)$  is entire in  $k$ , the corresponding integral in the expression (139) for  $m_n$  can always be expressed in closed form by a residue calculation at  $k = \infty$ :

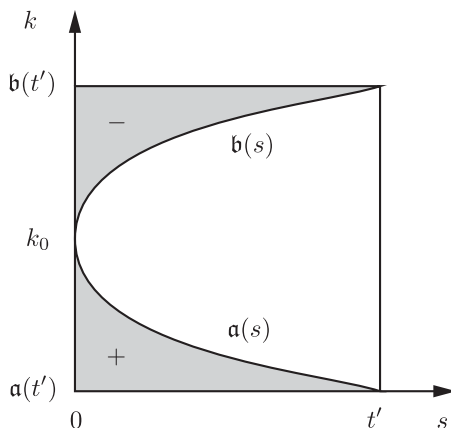
$$\int_B \frac{\theta'(k; x, t)k^{n-1} dk}{ir_+(k)} = -\frac{1}{2i} \oint_L \frac{(x + 4kt)k^{n-1} dk}{r(k)}, \tag{141}$$

where  $L$  is a large, positively oriented circular contour that encloses all of the bands. With the help of the expansion

$$\frac{1}{r(k)} = \frac{1}{k} + \frac{\alpha + \beta}{2k^2} + \frac{3\alpha^2 + 2\alpha\beta + 3\beta^2}{8k^3} + \mathcal{O}(k^{-4}), \quad k \rightarrow \infty, \quad N = 0, \tag{142}$$

we therefore obtain in the case  $N = 0$  that the moment conditions (139) take the form

$$\begin{aligned} m_1(\alpha, \beta) &= I_1(\alpha, \beta) - \pi(x + 2(\alpha + \beta)t) = 0 \\ m_2(\alpha, \beta) &= I_2(\alpha, \beta) - \frac{\pi}{2}((\alpha + \beta)x + (3\alpha^2 + 2\alpha\beta + 3\beta^2)t) = 0, \end{aligned} \tag{143}$$



**Figure 6.** The region of integration in the  $(s, k)$ -plane in the VBV case.

where

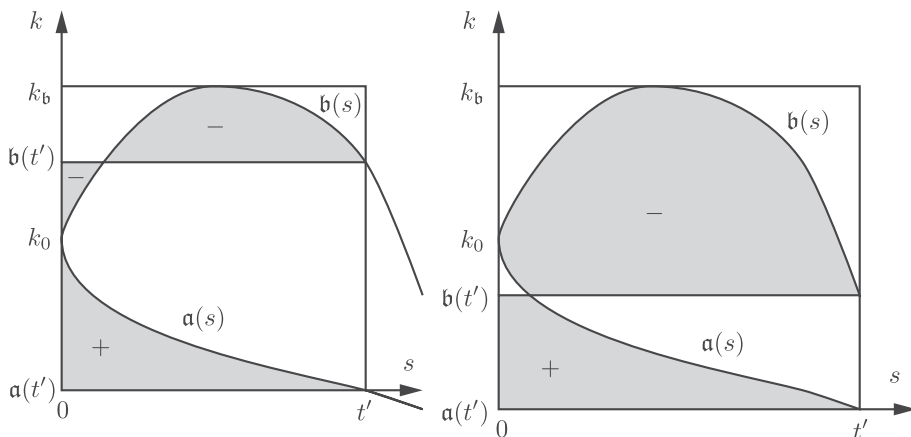
$$I_p(\alpha, \beta) := \int_S \frac{\tau'(k)k^{p-1} dk}{r(k)} - \int_\alpha^\beta \frac{\Phi'(k)k^{p-1} dk}{ir_+(k)}, \quad p = 1, 2. \quad (144)$$

**LEMMA 7.** *Suppose that  $x = 0$ . The equations (143) are satisfied for  $t > 0$  by  $\alpha = \alpha(t)$  and  $\beta = \beta(t)$  as long as  $(k_\alpha, \alpha(t))$  is a void (saturated region) if  $\alpha'(t) < 0$  ( $\alpha'(t) > 0$ ) and  $(\beta(t), k_\beta)$  is a void (saturated region) if  $\beta'(t) > 0$  ( $\beta'(t) < 0$ ).*

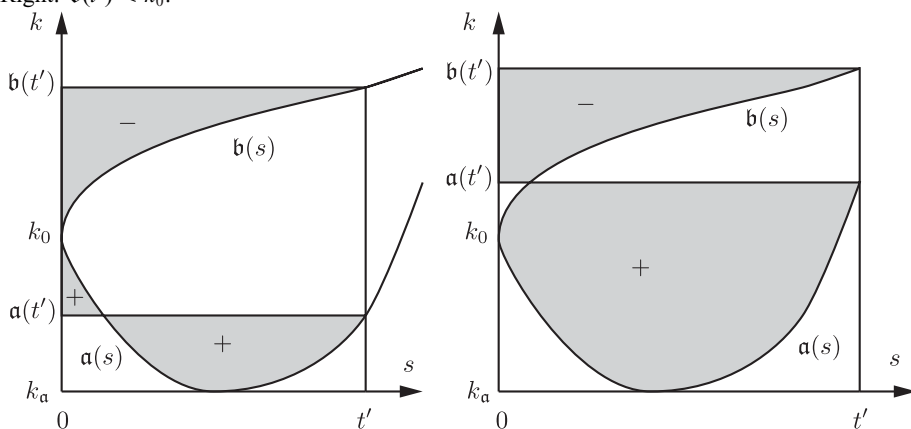
*Proof.* Suppose that for some  $t' > 0$  we have  $\alpha = \alpha(t')$  and  $\beta = \beta(t')$ . Let us evaluate  $I_p(\alpha(t'), \beta(t'))$  in four cases:

- VBV:** in this case we assume that  $\alpha'(t') < 0$ ,  $\beta'(t') > 0$ , and both intervals  $(k_\alpha, \alpha(t'))$  and  $(\beta(t'), k_\beta)$  are voids.
- VBS:** in this case we assume that  $\alpha'(t') < 0$  and  $\beta'(t') < 0$ , and that  $(k_\alpha, \alpha(t'))$  is a void but  $(\beta(t'), k_\beta)$  is a saturated region.
- SBV:** in this case we assume that  $\alpha'(t') > 0$  and  $\beta'(t') > 0$ , and that  $(\beta(t'), k_\beta)$  is a void but  $(k_\alpha, \alpha(t'))$  is a saturated region.
- SBS:** in this case we assume that  $\alpha'(t') > 0$  and  $\beta'(t') < 0$ , and that both intervals  $(k_\alpha, \alpha(t'))$  and  $(\beta(t'), k_\beta)$  are saturated regions.

We may substitute into (144) from (56) (after differentiating  $\tau(k)$  by Leibniz' rule taking into account that the integrand vanishes at both endpoints) and from (79). We may also simplify  $ir_+(k)$  for  $\alpha < k < \beta$  as  $-\sqrt{(k - \alpha)(\beta - k)} < 0$ , while in the integral over S we have either  $r(k) = \sqrt{(k - \alpha)(k - \beta)} > 0$  if  $S = (\beta, k_\beta)$  or  $r(k) = -\sqrt{(\alpha - k)(\beta - k)} < 0$  if  $S = (k_\alpha, \alpha)$ . The first remarkable fact is that the result can be written in a uniform way in all four cases (including all sub-cases related to where  $k_0$  falls with respect to  $\alpha = \alpha(t')$  and  $\beta = \beta(t')$  as illustrated in Figures 6–9).



**Figure 7.** The region of integration in the  $(s, k)$ -plane in the VBS case. Left:  $b(t') > k_0$ . Right:  $b(t') < k_0$ .

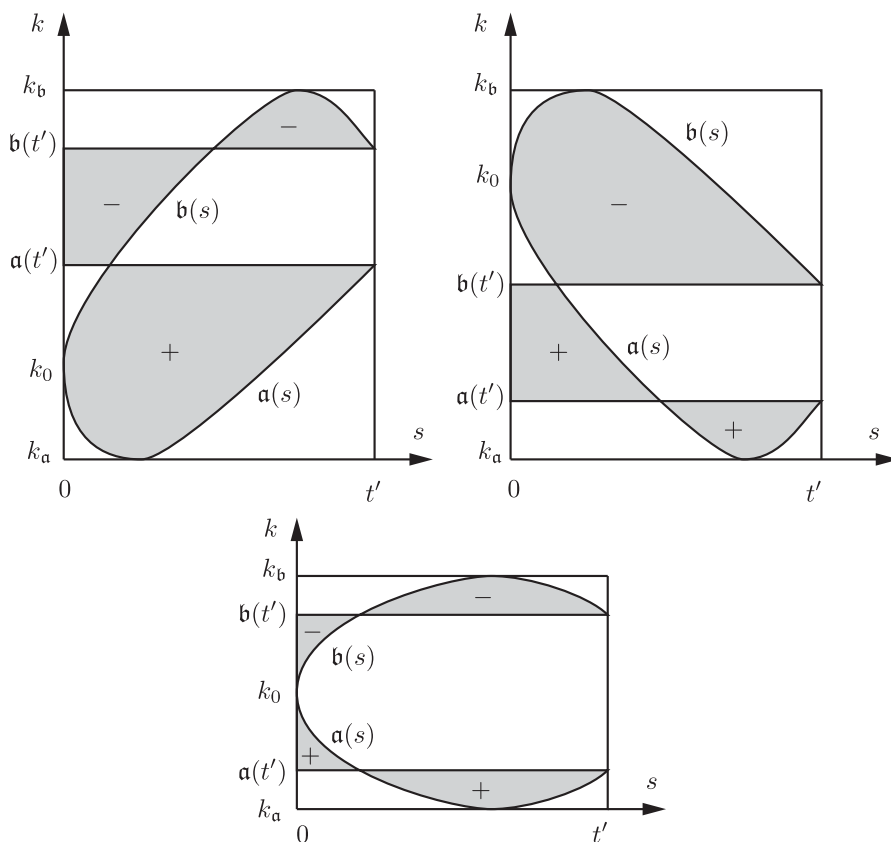


**Figure 8.** The region of integration in the  $(s, k)$ -plane in the SBV case. Left:  $a(t') < k_0$ . Right:  $a(t') > k_0$ .

Namely, we have

$$\begin{aligned}
 I_p(a(t'), b(t')) = & \iint_{D_+} \frac{4k^{p+1} + 2U(s)k^p - 2H(s)^2k^{p-1}}{\sqrt{-(k - a(t'))(k - a(s))(k - b(t'))(k - b(s))}} dA(s, k) \\
 & - \iint_{D_-} \frac{4k^{p+1} + 2U(s)k^p - 2H(s)^2k^{p-1}}{\sqrt{-(k - a(t'))(k - a(s))(k - b(t'))(k - b(s))}} dA(s, k),
 \end{aligned}
 \tag{145}$$

where  $dA(s, k)$  denotes the positive area element and where the domains  $D_+$  and  $D_-$  are the indicated shaded regions in Figures 6–9. It is now obvious that the original order of integration is easily reversed in all cases, with the outer  $s$  integral over the interval  $0 < s < t'$  and the inner integrals over

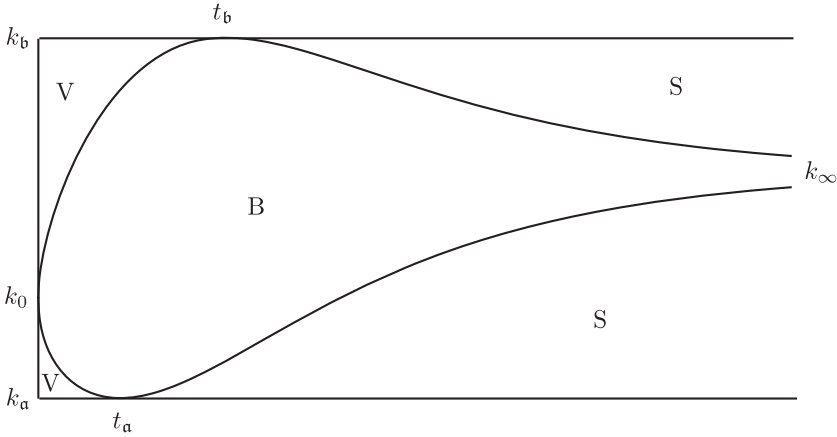


**Figure 9.** The region of integration in the  $(s, k)$ -plane in the SBS case. Top left:  $a(t') > k_0$ . Top right:  $b(t') < k_0$ . Bottom:  $a(t') < k_0 < b(t')$ .

the intervals with endpoints being the two most negative (for the domain  $D_+$ ) and the two least negative (for the domain  $D_-$ ) among the four values  $a(t')$ ,  $a(s)$ ,  $b(t')$ , and  $b(s)$ . Carrying out this reinterpretation of the formula (145), we may further combine the inner  $k$  integrals with the introduction of the function  $S(k; s, t')$  that is analytic for  $k$  in the complex plane with branch cuts lying in the two intervals of integration omitted, whose square is  $S(k; s, t')^2 = -(k - a(t'))(k - a(s))(k - b(t'))(k - b(s))$ , and that satisfies  $S(k; s, t') = -ik^2 + \mathcal{O}(k)$  as  $k \rightarrow \infty$ . The result is

$$I_p(a(t'), b(t')) = \int_0^{t'} \oint_L \frac{2k^{p+1} + U(s)k^p - H(s)^2 k^{p-1}}{S(k; s, t')} dk ds, \quad (146)$$

where  $L$  is a positively oriented loop that encloses both branch cuts of  $S$ . The inner  $k$  integral may now be computed by residues for each  $s \in (0, t')$ . The second remarkable fact is that for  $p = 1, 2$  the inner  $k$ -integral is independent



**Figure 10.** The curves  $k = a(t)$  and  $k = b(t)$  from Figure 2 and the corresponding voids (V), bands (B), and saturated regions (S).

of  $s$ :

$$I_0(a(t'), b(t')) = \int_0^{t'} 2\pi(a(t') + b(t')) ds = 2\pi(a(t') + b(t'))t', \quad (147)$$

and

$$\begin{aligned} I_1(a(t'), b(t')) &= \int_0^{t'} \frac{\pi}{2}(3a(t')^2 + 2a(t')b(t') + 3b(t')^2) dt' \\ &= \frac{\pi}{2}(3a(t')^2 + 2a(t')b(t') + 3b(t')^2)t'. \end{aligned} \quad (148)$$

It therefore follows by inspection that in all four cases, the equations  $m_1(\alpha, \beta) = 0$  and  $m_2(\alpha, \beta) = 0$  written in the form (143) are satisfied for  $x = 0$  and  $t > 0$  by taking  $\alpha = a(t)$  and  $\beta = b(t)$ . ■

The locations of the voids, bands, and saturated regions are indicated for the boundary data from Figure 2 in Figure 10. We therefore find for each  $t > 0$  a well-defined candidate for  $g(k)$  that we will denote by  $g(k; t)$ , with corresponding functions  $\phi(k; t)$  and  $\Delta(k; t)$  defined by (133), and it remains only to confirm the inequalities that were dropped earlier. Indeed, we will now prove the following.

**LEMMA 8.** *Let  $x = 0$  and  $t > 0$ , and let the function  $g(k; t)$  be determined from the values  $\alpha = a(t)$  and  $\beta = b(t)$  and the configuration of voids and saturated regions as described in Lemma 7. The functions  $\phi(k; t)$  and  $\Delta(k; t)$  given by (133) in terms of  $g$  satisfy the following inequalities:*

$$0 < \Delta(k; t) < \tau(k), \quad a(t) < k < b(t), \quad (149)$$

$$\phi'(k; t) > 0, \quad k \in V, \quad (150)$$

and

$$\phi'(k; t) < 0, \quad k \in S, \tag{151}$$

where  $V$  denotes the union of the voids,  $S$  denotes the union of the saturated regions, and the prime denotes differentiation in  $k$  for fixed  $t$ .

*Proof.* The proof of these three statements involves the same object, namely the partial derivative of  $g$  in  $t$  for fixed  $k$ , denoted  $g_t(k; t)$ . We may construct  $g_t(k; t)$  explicitly as follows. Consider differentiation with respect to  $t$  of the three equations  $g_+(k; t) - g_-(k; t) = 0$  for  $k \in V$ ,  $g_+(k; t) - g_-(k; t) = 2i\tau(k)$  for  $k \in S$ , and  $g_+(k; t) + g_-(k; t) = 2\theta(k; 0, t) - 2\Phi(k) - 2\phi_B(t)$  for  $k \in B = (\alpha(t), \beta(t))$ . Since neither  $\tau$  nor  $\Phi$  depend on  $t$ , we find that  $g_t(k; t)$  is analytic for  $k \in \mathbb{C} \setminus [\alpha(t), \beta(t)]$ , and on the cut the equation

$$g_{t+}(k; t) + g_{t-}(k; t) = 4k^2 - 2\phi'_B(t), \quad \alpha(t) < k < \beta(t) \tag{152}$$

holds. Keeping in mind the condition  $g_t(k; t) = \mathcal{O}(k^{-1})$  as  $k \rightarrow \infty$ , it therefore follows by similar arguments as led to (138) that

$$g_t(k; t) = \frac{r(k; t)}{2\pi i} \int_{\alpha(t)}^{\beta(t)} \frac{4l^2 - 2\phi'_B(t)}{r_+(l; t)(l - k)} dl, \tag{153}$$

where the notation  $r(k; t)$  reminds us that the branch points are  $\alpha = \alpha(t)$  and  $\beta = \beta(t)$ , and the integral can be evaluated in closed form by residues at  $l = k$  and  $l = \infty$ :

$$g_t(k; t) = 2k^2 - \phi'_B(t) + (U(t) - 2k)r(k; t), \quad k \in \mathbb{C} \setminus [\alpha(t), \beta(t)], \tag{154}$$

where we used the identity  $\alpha(t) + \beta(t) = -U(t)$ . Now we prove (149). From (133) and (154) it follows that

$$\begin{aligned} \Delta_t(k; t) &= -\frac{1}{2}i(g_{t+}(k; t) - g_{t-}(k; t)) \\ &= -i(U(t) - 2k)r_+(k; t) \\ &= (U(t) - 2k)\sqrt{(k - \alpha(t))(\beta(t) - k)}, \quad \alpha(t) < k < \beta(t). \end{aligned} \tag{155}$$

For a given  $k \in (\alpha(t), \beta(t))$ , we use the fundamental theorem of calculus to write

$$\Delta(k; t) = \Delta(k; t_-(k)) + \int_{t_-(k)}^t (U(s) - 2k)\sqrt{(k - \alpha(s))(\beta(s) - k)} ds, \tag{156}$$

a formula that makes use of the fact that  $\alpha(s) < k < \beta(s)$  holds for  $t_-(k) < s < t$ . But since  $t = t_-(k)$  always corresponds to the boundary between a band and a void, we have  $\Delta(k; t_-(k)) = 0$ . Taking this into account and comparing (156)

with (56) completes the proof of (149), since for  $a(t) < k < b(t)$  we have

$$\begin{aligned} \Delta(k; t) &= \int_{t_-(k)}^t (U(s) - 2k)\sqrt{(k - a(s))(b(s) - k)} ds > 0 \\ \tau(k) - \Delta(k; t) &= \int_t^{t_+(k)} (U(s) - 2k)\sqrt{(k - a(s))(b(s) - k)} ds > 0. \end{aligned} \tag{157}$$

Next, we consider together (150) and (151). From (133) and (154) we have

$$\begin{aligned} \phi_t(k; t) &= \theta_t(k; 0, t) - \frac{1}{2}(g_{t+}(k; t) + g_{t-}(k; t)) \\ &= 2k^2 - g_t(k; t), \quad k \in (k_a, a(t)) \cup (b(t), k_b), \\ &= \phi'_B(t) - (U(t) - 2k)r(k; t). \end{aligned} \tag{158}$$

Differentiation with respect to  $k$  using  $r(k; t)^2 = (k - a(t))(k - b(t))$  then gives

$$\phi'_t(k; t) = \frac{4k^2 + 2U(t)k - 2H(t)^2}{r(k; t)}, \quad k \in (k_a, a(t)) \cup (b(t), k_b). \tag{159}$$

Now we apply the fundamental theorem of calculus to obtain

$$\phi'(k; t) = \phi'(k; t_{\pm}(k)) + \int_{t_{\pm}(k)}^t \frac{4k^2 + 2U(s)k - 2H(s)^2}{r(k; s)} ds. \tag{160}$$

Here, the turning point  $t_{\pm}$  is selected so that either  $k < a(s)$  or  $k > b(s)$  holds for all  $s$  in the interval of integration. Since  $k$  lies on the boundary of the band  $(a(t_{\pm}(k)), b(t_{\pm}(k)))$ , we have  $\phi'(k; t_{\pm}(k)) = 0$ , so it remains to determine the sign of the integral. First observe that  $4k^2 + 2U(s)k - 2H(s)^2$  and  $r(k; s)$  always have opposite signs, regardless of whether  $k < a(s)$  or  $k > b(s)$ . Indeed, if  $k < a(s)$ , then  $r(k; s) = -\sqrt{(a(s) - k)(b(s) - k)} < 0$ , and by the identity

$$\begin{aligned} 4k^2 + 2U(s)k - 2H(s)^2 &= 2(a(s) - k)(b(s) - k) \\ &+ \frac{1}{2}(U(s) - 2k)((a(s) - k) + (b(s) - k)) \end{aligned} \tag{161}$$

one has  $4k^2 + 2U(s)k - 2H(s)^2 > 0$ . On the other hand, if  $k > b(s)$ , then  $r(k; s) = \sqrt{(k - a(s))(k - b(s))} > 0$  and by the identity  $4k^2 + 2U(s)k - 2H(s)^2 = 4(k - k_+(s))(k - k_-(s))$  where  $k_{\pm}(\cdot)$  are defined by (81) one sees that  $4k^2 + 2U(s)k - 2H(s)^2 < 0$ , because  $k_+(s) > 0$  which implies  $k - k_+(s) < 0$ , and also  $k_-(s) < b(s)$  which implies  $k - k_-(s) > k - b(s) > 0$ . Therefore, we conclude that  $\phi'(k; t) > 0$  if  $t < t_-(k)$ , which characterizes  $k$  lying in a void, and  $\phi'(k; t) < 0$  if  $t > t_+(k)$ , which characterizes  $k$  lying in a saturated region. This completes the proof of the inequalities (150) and (151). ■

The expression (154) for  $g_t(k; t)$  from this proof actually leads to a complete characterization of the constant  $\phi_B(t)$  as the following result shows.

LEMMA 9. *Let  $x = 0$  and  $t > 0$ . Then  $\phi_B(t) = -\frac{1}{2}S(t)$ .*

*Proof:* Since  $g(k; t) = \mathcal{O}(k^{-1})$  as  $k \rightarrow \infty$  for all  $t > 0$  (by (140)), it follows that also  $g_t(k; t) = \mathcal{O}(k^{-1})$  as  $k \rightarrow \infty$ . But if we examine the explicit expression for  $g_t(k; t)$  given by (154), we observe that there is a constant leading term in the Laurent series of  $g_t(k; t)$  for large  $|k|$ . This constant term therefore must vanish, and this gives rise to an identity expressing  $\phi'_B(t)$  explicitly in terms of  $\alpha(t)$  and  $\mathfrak{b}(t)$  (which can be simplified further with the help of (40) and (41)):

$$\phi'_B(t) = \frac{1}{4} (3\alpha(t)^2 + 2\alpha(t)\mathfrak{b}(t) + 3\mathfrak{b}(t)^2) = \frac{1}{2}U(t)^2 + H(t)^2 = -\frac{1}{2}S'(t). \tag{162}$$

Therefore, it remains to determine an integration constant. Suppose that  $t > 0$  is sufficiently small that both  $\alpha'(t) < 0$  and  $\mathfrak{b}'(t) > 0$ , i.e., we have a VB configuration for  $g(k; t)$ . Then  $g(k) = g(k; t)$  is analytic for  $k \in \mathbb{C} \setminus [\alpha, \beta]$ , and for  $\alpha < k < \beta$  we have  $\phi(k) = \phi_B$ , so using (133) we get

$$g_+(k) + g_-(k) = 2\theta(k; x, t) - 2\Phi(k) - 2\phi_B, \quad \alpha < k < \beta. \tag{163}$$

It follows from these considerations that  $g(k)$  must be given by the formula

$$g(k) = \frac{1}{r(k)} \left[ g_1 + \frac{1}{\pi i} \int_{\alpha}^{\beta} \frac{(\theta(l; x, t) - \Phi(l) - \phi_B)r_+(l)}{l - k} dl \right], \tag{164}$$

where  $g_1$  is an additional constant ( $g(k) = g_1k^{-1} + \mathcal{O}(k^{-2})$  as  $k \rightarrow \infty$ ). But  $g(k)$  is known to be bounded at the band endpoints  $k = \alpha$  and  $k = \beta$ , so the expression in square brackets must be made to vanish for these values of  $k$ , resulting in a system of linear equations for  $g_1$  and  $\phi_B$ :

$$\begin{bmatrix} 1 & -\frac{1}{\pi i} \int_{\alpha}^{\beta} \frac{r_+(l)dl}{l-\alpha} \\ 1 & -\frac{1}{\pi i} \int_{\alpha}^{\beta} \frac{r_+(l)dl}{l-\beta} \end{bmatrix} \begin{bmatrix} g_1 \\ \phi_B \end{bmatrix} = \begin{bmatrix} \frac{1}{\pi i} \int_{\alpha}^{\beta} \frac{(\Phi(l)-\theta(l;x,t))r_+(l)dl}{l-\alpha} \\ \frac{1}{\pi i} \int_{\alpha}^{\beta} \frac{(\Phi(l)-\theta(l;x,t))r_+(l)dl}{l-\beta} \end{bmatrix}. \tag{165}$$

The integrals in the coefficient matrix can be calculated by residues at  $l = \infty$  and the system solved for  $\phi_B$ :

$$\phi_B = \frac{1}{\pi i} \int_{\alpha}^{\beta} \frac{\Phi(l) - \theta(l; x, t)}{r_+(l)} dl. \tag{166}$$

Likewise, the integral involving  $\theta(l; x, t)$  can be evaluated by a residue at  $l = \infty$ , yielding

$$\phi_B = -\frac{1}{2}tS'(t) - \frac{1}{2}xU(t) + \frac{1}{\pi i} \int_{\alpha}^{\beta} \frac{\Phi(l) dl}{r_+(l)}, \tag{167}$$

where we have also used (40) and (41). Finally, we set  $x = 0$  and consider the limit  $t \downarrow 0$ , in which  $\alpha = \alpha(t)$  and  $\beta = \mathfrak{b}(t)$  converge to  $k_0$ . Using the fact, as shown in the proof of Lemma 3, that  $\Phi(k) = \frac{1}{2}S(0) + \mathcal{O}((k - k_0)^3)$  as  $k \rightarrow k_0$ , we replace  $\Phi(l)$  in the integral by its limiting value and calculate the resulting



integral by a residue at  $l = \infty$ . Therefore,

$$\lim_{t \downarrow 0} \phi_B(t) = -\frac{1}{2}S(0) \tag{168}$$

and the proof is complete. ■

LEMMA 10. *Let  $x = 0$  and  $t > 0$ . The function  $\phi(k; t)$  is analytic for  $k_a < k < a(t)$  and for  $b(t) < k < k_b$ .*

*Proof:* Suppose first that the interval containing  $k$  is a void. Then  $g_+(k; t) = g_-(k; t)$ , so  $g(k; t)$  is analytic at  $k$ , and we may write  $\phi(k; t) = \theta(k; 0, t) - \Phi(k) - g(k; t)$ , which is clearly analytic except at  $k = k_0$ , according to Lemma 3. Since  $a(0) = b(0) = k_0$ , and since  $a(t)$  is decreasing if  $(k_a, a(t))$  is a void while  $b(t)$  is increasing if  $(b(t), k_b)$  is a void, it follows that the point  $k_0$  cannot lie in a void for any  $t > 0$ .

Next suppose that the interval containing  $k$  is a saturated region. Then  $g_+(k; t) - g_-(k; t) = 2i\tau(k)$ , so we may write  $\phi(k; t)$  in two alternate forms:

$$\phi(k; t) = \theta(k; 0, t) - g_{\pm}(k; t) - (\Phi(k) \mp i\tau(k)). \tag{169}$$

Therefore,  $\phi(k; t)$  will be the boundary value of a function analytic in  $\mathbb{C}_{\pm}$  if this is true of the function  $\Phi(k) \mp i\tau(k)$ . From (56) and (57), we have

$$\begin{aligned} \Phi(k) \mp i\tau(k) &= \frac{1}{2}S(0) + \operatorname{sgn}(k^2 - k_0^2) \int_0^{t_-(k)} (U(t) - 2k)\sqrt{(k - a(t))(k - b(t))} dt \\ &\quad \mp i \int_{t_-(k)}^{t_+(k)} (U(t) - 2k)\sqrt{(k - a(t))(b(t) - k)} dt \\ &= \frac{1}{2}S(0) - \int_0^{t_-(k)} (U(t) - 2k)r(k; t) dt - \int_{t_-(k)}^{t_+(k)} (U(t) - 2k)r_{\pm}(k; t) dt \\ &= \frac{1}{2}S(0) - \int_0^{t_+(k)} (U(t) - 2k)r_{\pm}(k; t) dt. \end{aligned} \tag{170}$$

But the only point of nonanalyticity of  $t_+(k)$  in  $(k_a, k_b)$  is  $k = k_{\infty}$ , so if  $k \neq k_{\infty}$ , then  $\Phi(k) \mp i\tau(k)$  is the boundary value of a function analytic for  $k$  in the half-plane  $\mathbb{C}_{\pm}$  near  $k$ . It follows that if  $k$  is in a saturated region and  $k \neq k_{\infty}$  then  $\phi(k; t)$  is analytic at  $k$ , i.e., it can be continued into both half-planes. But since  $a(t)$  and  $b(t)$  both tend to  $k_{\infty}$  as  $t \rightarrow \infty$  and  $a(t)$  is increasing if  $(k_a, a(t))$  is a saturated region while  $b(t)$  is decreasing if  $(b(t), k_b)$  is a saturated region, it is impossible for  $k_{\infty}$  to lie in a saturated region for any  $t > 0$ . ■

We will refer to the analytic function  $\phi(k; t)$  defined in the interval  $(k_a, a(t))$  (respectively, in the interval  $(b(t), k_b)$ ) as  $\phi_a(k; t)$  (respectively,  $\phi_b(k; t)$ ). Finally, we require an analogue of Lemma 6.

LEMMA 11. *The functions  $Y^\epsilon(k)e^{\pm 2i\phi(k;t)/\epsilon}$  have analytic continuations into the complex plane from right and left neighborhoods of  $k_\alpha$  and  $k_\beta$ , respectively, and along small segments with one endpoint  $k_\alpha$  or  $k_\beta$  and the other endpoint having real part in the interior of  $(k_\alpha, k_\beta)$  and nonzero imaginary part of the appropriate sign so that  $|e^{\pm 2i\phi(k;t)/\epsilon}| \leq 1$  along the segment, the uniform estimate  $Y^\epsilon(k)e^{\pm 2i\phi(k;t)/\epsilon} = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  holds.*

*Proof:* Applying the nondegeneracy of the extrema of  $\mathfrak{a}$  and  $\mathfrak{b}$  guaranteed by Assumption 1 to the formula (160), it is easy to see that  $\phi'(k; t)$  always diverges logarithmically as  $k \downarrow k_\alpha$  and as  $k \uparrow k_\beta$ . Hence, by integration in  $k$  one sees that up to a nonzero constant factor plus an integration constant, the leading-order behavior of  $\phi(k; t)$  is the same as that of  $\Phi(k)$  as established in Lemma 5. The rest of the proof is then exactly the same as that of Lemma 6, playing off the exponential decay of  $e^{\pm 2i\phi(k;t)/\epsilon}$  into the appropriate half-plane away from  $k_\alpha$  or  $k_\beta$  against the linear vanishing of  $\tau_\alpha$  or  $\tau_\beta$  to establish the claimed uniform estimate. ■

4.2.2 *Proof of Theorem 2.* To prove Theorem 2, we apply the steepest descent method to Riemann–Hilbert problem 2 with the help of the complex phase function  $g = g(k; t)$  introduced in Section 4.2.1, that is, we exploit the transformation (131) from  $\tilde{\mathbf{M}}(k)$  to  $\mathbf{N}(k)$  and use (134)–(136) to handle the jump condition (132) by opening lenses about the voids and saturated regions. Some minor modifications are required because  $Y^\epsilon(k)$  is not analytic at  $k = k_0$  (which may lie in a saturated region but not a void) or  $k = k_\infty$  (which may lie in a void but not a saturated region), however it will not be necessary to introduce any nonanalyticity or deal with  $\bar{\partial}$  problems as in the proof of Theorem 1.

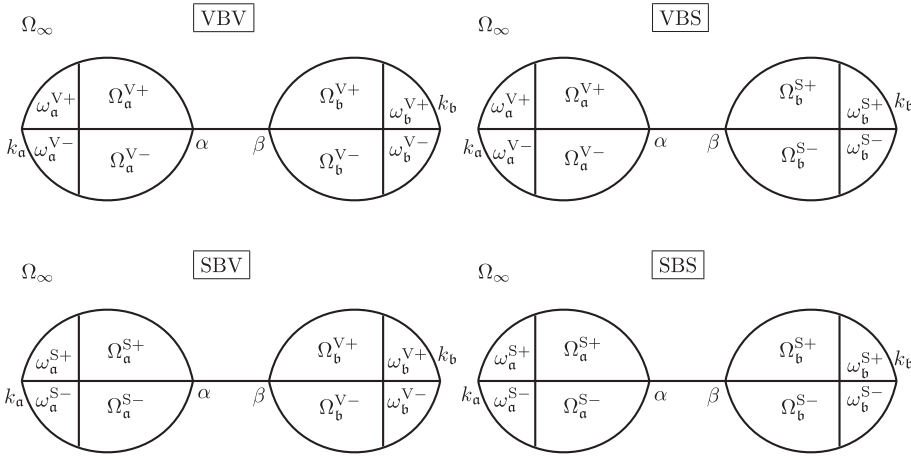
To open the lenses, we define domains of the complex plane as illustrated in Figure 11 and make the following explicit substitution:

$$\mathbf{O}(k) := \mathbf{N}(k) \begin{bmatrix} 1 & 0 \\ -e^{2i\phi_{\mathfrak{a}, \mathfrak{b}}(k;t)/\epsilon} & 1 \end{bmatrix}, \quad k \in \Omega_{\mathfrak{a}, \mathfrak{b}}^{V+}, \tag{171}$$

$$\mathbf{O}(k) := \mathbf{N}(k) \begin{bmatrix} 1 & -e^{-2i\phi_{\mathfrak{a}, \mathfrak{b}}(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \Omega_{\mathfrak{a}, \mathfrak{b}}^{V-}, \tag{172}$$

$$\mathbf{O}(k) := \mathbf{N}(k) \begin{bmatrix} 1 & e^{-2i\phi_{\mathfrak{a}, \mathfrak{b}}(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \Omega_{\mathfrak{a}, \mathfrak{b}}^{S+}, \tag{173}$$

$$\mathbf{O}(k) := \mathbf{N}(k) \begin{bmatrix} 1 & 0 \\ e^{2i\phi_{\mathfrak{a}, \mathfrak{b}}(k;t)/\epsilon} & 1 \end{bmatrix}, \quad k \in \Omega_{\mathfrak{a}, \mathfrak{b}}^{S-}, \tag{174}$$



**Figure 11.** The lens domains about the interval  $[k_a, k_b]$  in the four configurations of the complex phase function  $g$ . Void intervals cannot contain the point  $k_0$ , and saturated regions cannot contain the point  $k_\infty$ . The various domains labeled with the letter  $\omega$  are so small as to exclude both of these points, because only near the endpoints of  $[k_a, k_b]$  do we need to exploit analyticity of  $\tau$  in order to obtain decay along the lens boundaries without installing unusual parametrices but rather by using Lemma 11.

$$\mathbf{O}(k) := \mathbf{N}(k) \begin{bmatrix} 1 & 0 \\ -Y_{a,b}^\epsilon(k)e^{2i\phi_{a,b}(k;t)/\epsilon} & 1 \end{bmatrix}, \quad k \in \omega_{a,b}^{V+}, \quad (175)$$

$$\mathbf{O}(k) := \mathbf{N}(k) \begin{bmatrix} 1 & -Y_{a,b}^\epsilon(k)e^{-2i\phi_{a,b}(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \omega_{a,b}^{V-}, \quad (176)$$

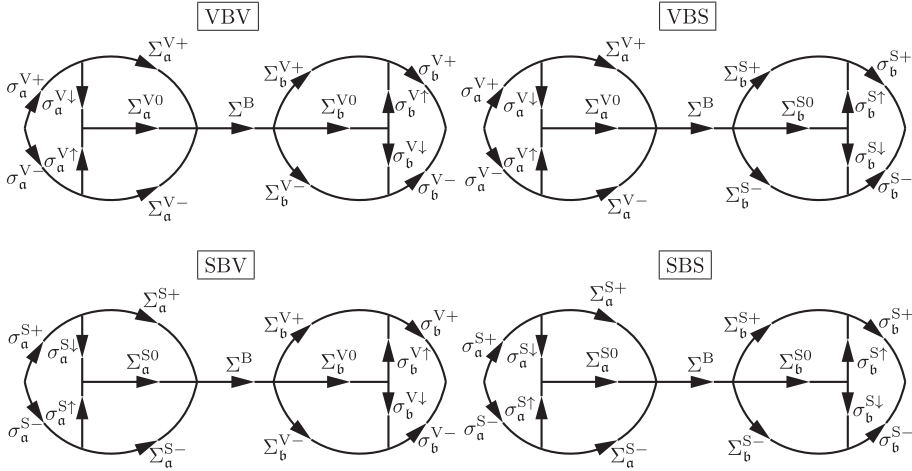
$$\mathbf{O}(k) := \mathbf{N}(k) \begin{bmatrix} 1 & Y_{a,b}^\epsilon(k)e^{-2i\phi_{a,b}(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \omega_{a,b}^{S+}, \quad (177)$$

$$\mathbf{O}(k) := \mathbf{N}(k) \begin{bmatrix} 1 & 0 \\ Y_{a,b}^\epsilon(k)e^{2i\phi_{a,b}(k;t)/\epsilon} & 1 \end{bmatrix}, \quad k \in \omega_{a,b}^{S-}, \quad (178)$$

and in the unbounded domain  $\Omega_\infty$  we set  $\mathbf{O}(k) := \mathbf{N}(k)$ . The matrix  $\mathbf{O}(k)$  satisfies the conditions of the following Riemann–Hilbert problem.

**Riemann–Hilbert problem 6.** Find a  $2 \times 2$  matrix  $\mathbf{O}(k)$  with the following properties:

**Analyticity:**  $\mathbf{O}(k)$  is analytic for  $k \in \mathbb{C} \setminus \Sigma$ , where  $\Sigma$  is the contour illustrated in Figure 12 and  $\mathbf{O}$  takes continuous boundary values on each oriented arc of  $\Sigma$ ,  $\mathbf{O}_+(k)$  from the left and  $\mathbf{O}_-(k)$  from the right.



**Figure 12.** The oriented arcs of the contour  $\Sigma$  for Riemann–Hilbert problem 6 in the four cases for the complex phase function  $g$ .

**Jump Condition:** The boundary values on each oriented arc of  $\Sigma$  are related by  $\mathbf{O}_+(k) = \mathbf{O}_-(k)\mathbf{J}(k)$  (see below for the definition of  $\mathbf{J}$ ).

**Normalization:**  $\mathbf{O}(k) \rightarrow \mathbb{I}$  as  $k \rightarrow \infty$ .

The jump matrix  $\mathbf{J}$  is defined on  $\Sigma$  as follows:

$$\mathbf{J}(k) := \begin{bmatrix} 1 & 0 \\ Y_{a,b}^\epsilon(k)e^{2i\phi_{a,b}(k;t)/\epsilon} & 1 \end{bmatrix}, \quad k \in \sigma_{a,b}^{V+} \cup \sigma_{a,b}^{S-}, \quad (179)$$

$$\mathbf{J}(k) := \begin{bmatrix} 1 & -Y_{a,b}^\epsilon(k)e^{-2i\phi_{a,b}(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \sigma_{a,b}^{V-} \cup \sigma_{a,b}^{S+}, \quad (180)$$

$$\mathbf{J}(k) := \begin{bmatrix} 1 & 0 \\ e^{2i\phi_{a,b}(k;t)/\epsilon} & 1 \end{bmatrix}, \quad k \in \Sigma_{a,b}^{V+} \cup \Sigma_{a,b}^{S-}, \quad (181)$$

$$\mathbf{J}(k) := \begin{bmatrix} 1 & -e^{-2i\phi_{a,b}(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \Sigma_{a,b}^{V-} \cup \Sigma_{a,b}^{S+}, \quad (182)$$

$$\mathbf{J}(k) := \begin{bmatrix} 1 & 0 \\ (Y_{a,b}^\epsilon(k) - 1)e^{2i\phi_{a,b}(k;t)/\epsilon} & 1 \end{bmatrix}, \quad k \in \sigma_{a,b}^{V\downarrow} \cup \sigma_{a,b}^{S\downarrow}, \quad (183)$$

$$\mathbf{J}(k) := \begin{bmatrix} 1 & (1 - Y_{a,b}^\epsilon(k))e^{-2i\phi_{a,b}(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \sigma_{a,b}^{V\uparrow} \cup \sigma_{a,b}^{S\uparrow}, \quad (184)$$

$$\mathbf{J}(k) := \begin{bmatrix} 1 & (1 - Y^\epsilon(k))e^{-2i\phi(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ (Y^\epsilon(k) - 1)e^{2i\phi(k;t)/\epsilon} & 1 \end{bmatrix},$$

$$k \in \Sigma_a^{V0} \cup \Sigma_b^{V0}, \tag{185}$$

$$\mathbf{J}(k) := \begin{bmatrix} 1 & 0 \\ (Y^\epsilon(k) - 1)e^{2i\phi(k;t)/\epsilon} & 1 \end{bmatrix} \begin{bmatrix} 1 & (1 - Y^\epsilon(k))e^{-2i\phi(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix},$$

$$k \in \Sigma_a^{S0} \cup \Sigma_b^{S0}, \tag{186}$$

and

$$\mathbf{J}(k) := \begin{bmatrix} e^{2(\Delta(k;t) - \tau(k))/\epsilon} & -Y^\epsilon(k)e^{-2i\phi_B(t)/\epsilon} \\ Y^\epsilon(k)e^{2i\phi_B(t)/\epsilon} & e^{-2\Delta(k;t)/\epsilon} \end{bmatrix}, \quad k \in \Sigma^B. \tag{187}$$

It follows from (150) and (151) in Lemma 8, from Lemma 10, and from Lemma 11 that  $\mathbf{J} - \mathbb{I}$  is uniformly small on  $\sigma_{a,b}^{V\pm} \cup \sigma_{a,b}^{S\pm} \cup \Sigma_{a,b}^{V\pm} \cup \Sigma_{a,b}^{S\pm}$  (that is, all nonvertical and nonhorizontal arcs of  $\Sigma$ ) omitting only neighborhoods of the band endpoints  $\alpha = a(t)$  and  $\beta = b(t)$ . The rate of decay is determined from neighborhoods of  $k_a$  and  $k_b$  according to Lemma 11, namely  $\mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$ , but away from these points one has exponential decay. Likewise, from the fact that  $\tau(k) > 0$  is uniformly bounded away from zero on compact subsets of  $(k_a, k_b)$  means that  $\mathbf{J} - \mathbb{I}$  is also uniformly exponentially small on  $\Sigma_{a,b}^{V0} \cup \Sigma_{a,b}^{S0} \cup \sigma_{a,b}^{V\downarrow} \cup \sigma_{a,b}^{V\uparrow} \cup \sigma_{a,b}^{S\downarrow} \cup \sigma_{a,b}^{S\uparrow}$ , that is, on all vertical arcs of  $\Sigma$  and on all horizontal arcs except the band B.

Riemann–Hilbert problem 6 is *not*, however, a small-norm problem in the semiclassical limit  $\epsilon \downarrow 0$ , because  $\mathbf{J} - \mathbb{I}$  is not decaying with  $\epsilon$  on the band B, nor is the decay on the nonreal arcs of  $\Sigma$  that meet at  $k = \alpha$  and  $k = \beta$  uniform near these band endpoints. We will now remedy this situation by constructing an explicit parametrix for  $\mathbf{O}(k)$  in a standard fashion. First, we exhibit a matrix solving the limiting form of the jump condition on the band B:

$$\dot{\mathbf{O}}^{(\text{out})}(k) := e^{-i\phi_B(t)\sigma_3/\epsilon} \mathbf{S} \left( \frac{(k - \beta)^{1/4}}{(k - \alpha)^{1/4}} \right)^{\sigma_3} \mathbf{S}^{-1} e^{i\phi_B(t)\sigma_3/\epsilon}, \quad \mathbf{S} := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -i \\ -i & 1 \end{bmatrix}, \tag{188}$$

where  $(k - \beta)^{1/4}$  and  $(k - \alpha)^{1/4}$  denote the principal branches (and hence the ratio may be considered to be well-defined on the interval  $k < \alpha$  common to both branch cuts). It is easy to confirm that this *outer parametrix* has the following properties:

- $\dot{\mathbf{O}}^{(\text{out})}(k)$  is analytic for  $k \in \mathbb{C} \setminus [\alpha, \beta]$ ,
- $\dot{\mathbf{O}}^{(\text{out})}(k)$  satisfies the jump condition

$$\dot{\mathbf{O}}_+^{(\text{out})}(k) = \dot{\mathbf{O}}_-^{(\text{out})}(k) \begin{bmatrix} 0 & -e^{-2i\phi_B(t)/\epsilon} \\ e^{2i\phi_B(t)/\epsilon} & 0 \end{bmatrix}, \quad \alpha < k < \beta, \quad (189)$$

- $\det(\dot{\mathbf{O}}^{(\text{out})}(k)) = 1$ ,
- $\dot{\mathbf{O}}^{(\text{out})}(k) \rightarrow \mathbb{I}$  as  $k \rightarrow \infty$ , and
- For  $k$  bounded away from  $\{\alpha, \beta\}$ ,  $\dot{\mathbf{O}}^{(\text{out})}(k)$  is uniformly bounded independent of  $\epsilon$ .

The outer parametrix will turn out to be a good approximation of  $\mathbf{O}(k)$  away from the points  $k = \alpha, \beta$ , but it blows up at these two points and fails to even approximately satisfy the nonnegligible jump conditions on the complex contours nearby. For now, we record what will be a useful formula later on:

$$\dot{\mathbf{O}}^{(\text{out})}(k) = \mathbb{I} + k^{-1}\dot{\mathbf{O}}_1 + k^{-2}\dot{\mathbf{O}}_2 + \mathcal{O}(k^{-3}), \quad k \rightarrow \infty, \quad (190)$$

where

$$\begin{aligned} \dot{\mathbf{O}}_1 &:= \frac{\beta - \alpha}{4} \begin{bmatrix} 0 & -ie^{-2i\phi_B(t)/\epsilon} \\ ie^{2i\phi_B(t)/\epsilon} & 0 \end{bmatrix} \\ \dot{\mathbf{O}}_2 &:= \frac{\beta - \alpha}{32} \begin{bmatrix} \beta - \alpha & -4i(\beta + \alpha)e^{-2i\phi_B(t)/\epsilon} \\ 4i(\beta + \alpha)e^{2i\phi_B(t)/\epsilon} & \beta - \alpha \end{bmatrix}. \end{aligned} \quad (191)$$

Let  $D_\alpha$  and  $D_\beta$  be open disks centered at  $k = \alpha$  and  $k = \beta$ , respectively, with radius  $\delta > 0$  sufficiently small, but independent of  $\epsilon$ . We shall construct an *inner parametrix* in each of these disks, an approximation that will locally be far superior to the outer parametrix.

First consider the disk  $D_\alpha$ . If the interval  $(k_\alpha, \alpha)$ ,  $\alpha = a(t)$ , is a void  $V$ , then we claim that the function  $w_\alpha^V(k; t)$  defined for  $k \in B$  near  $\alpha$  by  $w_\alpha^V(k; t) := (2\Delta(k; t))^{2/3}$  (positive  $2/3$  power) can be analytically continued to a full complex neighborhood of  $k = \alpha$  as a function that satisfies  $w_\alpha^V(\alpha; t) > 0$ . This is a simple consequence of the fact that  $\Delta(a(t); t) = 0$  and that  $\Delta'(k; t)$  vanishes like a square root and no higher power at  $k = a(t)$ . It follows that  $w_\alpha^V(\cdot; t)$  defines a conformal mapping from  $D_\alpha$  onto a neighborhood of the origin. The outer parametrix may be represented locally near  $k = \alpha$  in terms of  $w_\alpha^V(k; t)$  as follows:

$$\dot{\mathbf{O}}^{(\text{out})}(k) = \mathbf{H}_\alpha^V(k)(-w_\alpha^V(k; t))^{-\sigma_3/4} \mathbf{S}^{-1} e^{i\phi_B(t)\sigma_3/\epsilon}, \quad (192)$$

where  $\mathbf{H}_\alpha^V(k)$  is a well-defined unimodular matrix function holomorphic near  $k = \alpha$  that is obviously uniformly bounded in  $D_\alpha$  independent of  $\epsilon$ . It will be

useful later to write this in the equivalent form

$$\dot{\mathbf{O}}^{(\text{out})}(k) = \mathbf{H}_a^V(k) \epsilon^{-\sigma_3/6} (-\zeta)^{-\sigma_3/4} \mathbf{S}^{-1} e^{i\phi_B(t)\sigma_3/\epsilon}, \quad k \in D_\alpha, \quad \zeta := \epsilon^{-2/3} w_a^V(k; t). \quad (193)$$

Let an auxiliary matrix function  $\mathbf{P}(\zeta)$  be defined as follows ( $\xi := (\frac{3}{4})^{2/3}\zeta$ ):

$$\begin{aligned} \mathbf{P}(\zeta) = & \sqrt{2\pi} \left(\frac{3}{4}\right)^{\frac{1}{6}\sigma_3} e^{-\frac{1}{4}\pi i\sigma_3} \begin{bmatrix} -ie^{\frac{2}{3}\pi i} \text{Ai}(\xi e^{\frac{2}{3}\pi i}) & -ie^{-\frac{2}{3}\pi i} \text{Ai}(\xi e^{-\frac{2}{3}\pi i}) \\ e^{-\frac{2}{3}\pi i} \text{Ai}'(\xi e^{\frac{2}{3}\pi i}) & e^{\frac{2}{3}\pi i} \text{Ai}'(\xi e^{-\frac{2}{3}\pi i}) \end{bmatrix} \\ & \cdot e^{-\frac{2}{3}i(-\xi)^{3/2}\sigma_3}, \quad |\arg(-\zeta)| < \frac{\pi}{3}, \end{aligned} \quad (194)$$

$$\begin{aligned} \mathbf{P}(\zeta) := & \sqrt{2\pi} \left(\frac{3}{4}\right)^{\frac{1}{6}\sigma_3} e^{-\frac{1}{4}\pi i\sigma_3} \begin{bmatrix} i\text{Ai}(\xi) & -ie^{-\frac{2}{3}\pi i} \text{Ai}(\xi e^{-\frac{2}{3}\pi i}) \\ -\text{Ai}'(\xi) & e^{\frac{2}{3}\pi i} \text{Ai}'(\xi e^{-\frac{2}{3}\pi i}) \end{bmatrix} \\ & \cdot e^{\frac{2}{3}\xi^{3/2}\sigma_3}, \quad 0 < \arg(\zeta) < \frac{2\pi}{3}, \end{aligned} \quad (195)$$

and

$$\begin{aligned} \mathbf{P}(\zeta) := & \sqrt{2\pi} \left(\frac{3}{4}\right)^{\frac{1}{6}\sigma_3} e^{-\frac{1}{4}\pi i\sigma_3} \begin{bmatrix} -ie^{\frac{2}{3}\pi i} \text{Ai}(\xi e^{\frac{2}{3}\pi i}) & i\text{Ai}(\xi) \\ e^{-\frac{2}{3}\pi i} \text{Ai}'(\xi e^{\frac{2}{3}\pi i}) & -\text{Ai}'(\xi) \end{bmatrix} \\ & \cdot e^{-\frac{2}{3}\xi^{3/2}\sigma_3}, \quad -\frac{2\pi}{3} < \arg(\zeta) < 0. \end{aligned} \quad (196)$$

Using well-documented formulae involving the Airy function  $\text{Ai}$  and its derivative [19], it follows that  $\mathbf{P}(\zeta)$  is analytic in the three sectors of its definition, that across the rays bounding the sectors one has

$$\lim_{\mu \downarrow 0} \mathbf{P}(\zeta e^{i\mu}) = \lim_{\mu \downarrow 0} \mathbf{P}(\zeta e^{-i\mu}) \begin{bmatrix} 0 & -1 \\ 1 & e^{-\zeta^{3/2}} \end{bmatrix}, \quad \arg(\zeta) = 0, \quad (197)$$

$$\lim_{\mu \downarrow 0} \mathbf{P}(\zeta e^{i\mu}) = \lim_{\mu \downarrow 0} \mathbf{P}(\zeta e^{-i\mu}) \begin{bmatrix} 1 & 0 \\ -e^{\zeta^{3/2}} & 1 \end{bmatrix}, \quad \arg(\zeta) = \frac{2\pi}{3}, \quad (198)$$

and

$$\lim_{\mu \downarrow 0} \mathbf{P}(\zeta e^{i\mu}) = \lim_{\mu \downarrow 0} \mathbf{P}(\zeta e^{-i\mu}) \begin{bmatrix} 1 & e^{\zeta^{3/2}} \\ 0 & 1 \end{bmatrix}, \quad \arg(\zeta) = -\frac{2\pi}{3}, \quad (199)$$

and that

$$\mathbf{P}(\zeta) \mathbf{U}(-\zeta)^{\sigma_3/4} = \mathbb{I} + \begin{bmatrix} \mathcal{O}(\zeta^{-3}) & \mathcal{O}(\zeta^{-2}) \\ \mathcal{O}(\zeta^{-1}) & \mathcal{O}(\zeta^{-3}) \end{bmatrix}, \quad \zeta \rightarrow \infty, \quad (200)$$

with the asymptotics being uniform with respect to direction in the complex plane, including along the sector boundary rays. Then, in terms of  $\mathbf{P}(\zeta)$  and  $\mathbf{H}_\alpha^V(k)$  we define an inner parametrix near  $k = \alpha$  by setting

$$\dot{\mathbf{O}}_\alpha^V(k) := \mathbf{H}_\alpha^V(k)\epsilon^{-\sigma_3/6}\mathbf{P}(\epsilon^{-2/3}w_\alpha^V(k;t))e^{i\phi_B(t)\sigma_3/\epsilon}, \quad k \in D_\alpha. \quad (201)$$

Assuming without loss of generality that for  $k \in D_\alpha$  the contours  $\Sigma_\alpha^{V\pm}$  coincide with radial segments in the  $w_\alpha^V$ -plane with angles  $\pm 2\pi/3$ , one can check that the following facts hold true:

- $\dot{\mathbf{O}}_\alpha^V(k)$  is analytic for  $k \in D_\alpha \setminus (\Sigma_\alpha^{V+} \cup \Sigma_\alpha^{V-} \cup \Sigma^B)$  (there is no jump across  $\Sigma_\alpha^{V0}$ ).
- $\dot{\mathbf{O}}_\alpha^V(k)$  satisfies the jump conditions

$$\dot{\mathbf{O}}_{\alpha+}^V(k) = \dot{\mathbf{O}}_{\alpha-}^V(k) \begin{bmatrix} 1 & 0 \\ e^{2i\phi_\alpha(k;t)/\epsilon} & 1 \end{bmatrix}, \quad k \in \Sigma_\alpha^{V+} \cap D_\alpha, \quad (202)$$

$$\dot{\mathbf{O}}_{\alpha+}^V(k) = \dot{\mathbf{O}}_{\alpha-}^V(k) \begin{bmatrix} 1 & -e^{-2i\phi_\alpha(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \Sigma_\alpha^{V-} \cap D_\alpha, \quad (203)$$

and

$$\dot{\mathbf{O}}_{\alpha+}^V(k) = \dot{\mathbf{O}}_{\alpha-}^V(k) \begin{bmatrix} 0 & -e^{-2i\phi_B(t)/\epsilon} \\ e^{2i\phi_B(t)/\epsilon} & e^{-2\Delta(k;t)/\epsilon} \end{bmatrix}, \quad k \in \Sigma^B \cap D_\alpha. \quad (204)$$

- Inner and outer parametrices match well on the disk boundary:

$$\dot{\mathbf{O}}_\alpha^V(k)\dot{\mathbf{O}}^{(\text{out})}(k)^{-1} = \mathbb{I} + \mathcal{O}(\epsilon), \quad k \in \partial D_\alpha. \quad (205)$$

- $\det(\dot{\mathbf{O}}_\alpha^V(k)) = 1$  and  $\dot{\mathbf{O}}_\alpha^V(k) = \mathcal{O}(\epsilon^{-1/6})$  holds uniformly for  $k \in D_\alpha$ .

To check the jump conditions one should first express the analytic function  $\phi_\alpha(k;t)$  on the arcs  $\Sigma_\alpha^{V\pm}$  in terms of the conformal coordinate  $w_\alpha^V(k;t)$  (this has already been done for  $\Delta(k;t)$ , really by definition). This is accomplished by noting that

$$\begin{aligned} g_\pm(k;t) + \Phi(k) - \theta(k;x,t) &= \frac{1}{2}(g_+(k;t) + g_-(k;t)) \pm \frac{1}{2}(g_+(k;t) - g_-(k;t)) \\ &\quad + \Phi(k) - \theta(k;x,t) \\ &= -\phi(k;t) \pm i\Delta(k;t) \end{aligned} \quad (206)$$

holds at every point of  $(k_\alpha, k_b)$ . Consequently, since  $(k_\alpha, \mathfrak{a}(t))$  is a void  $V$ , then for  $t > 0$ ,  $\Phi$  is analytic at  $k = \mathfrak{a}(t)$  and it follows that the function  $-\phi_\alpha(k;t)$  for  $k_\alpha < k < \mathfrak{a}(t)$  is the analytic continuation through  $\mathbb{C}_\pm$  of the function  $-\phi_B(t) \pm i\Delta(k;t)$  for  $\mathfrak{a}(t) < k < \mathfrak{b}(t)$ . Thus, for  $k \in \Sigma^B \cap D_\alpha$  we have  $2\Delta(k;t) = w_\alpha^V(k;t)^{3/2}$ , and for  $k \in \Sigma_\alpha^{V\pm} \cap D_\alpha$  we have  $\pm 2i(\phi_\alpha(k;t) - \phi_B(t)) = w_\alpha^V(k;t)^{3/2}$ . To confirm the matching between the inner



and outer parametrices, one uses the asymptotic condition (200) and the fact that  $k \in \partial D_\alpha$  means  $|\zeta| \sim \epsilon^{-2/3}$ . The  $\mathcal{O}(\epsilon^{-1/6})$  bound within the disk can be proved similarly.

If instead  $(k_\alpha, \alpha)$  is a saturated region S, then one defines a conformal mapping  $w_\alpha^S(k; t)$  taking  $D_\alpha$  to a neighborhood of the origin by the analytic continuation from  $k \in \Sigma^B$  of  $w_\alpha^S(k; t) := (2\tau(k) - 2\Delta(k; t))^{2/3} > 0$ . One next defines a uniformly bounded unimodular holomorphic matrix function  $\mathbf{H}_\alpha^S(k)$  near  $k = \alpha$  by writing the outer parametrix in the form

$$\mathbf{O}^{(\text{out})}(k) = \mathbf{H}_\alpha^S(k)(-w_\alpha^S(k; t))^{-\sigma_3/4} \mathbf{S}^{-1} e^{i\phi_B \sigma_3/\epsilon}, \quad k \in D_\alpha. \tag{207}$$

Finally, one defines

$$\dot{\mathbf{O}}_\alpha^S(k) := \mathbf{H}_\alpha^S(k) \epsilon^{-\sigma_3/6} (i\sigma_3) \mathbf{P}(\epsilon^{-2/3} w_\alpha^S(k; t)) (i\sigma_2) e^{i\phi_B(t)\sigma_3/\epsilon}, \quad k \in D_\alpha. \tag{208}$$

Under the assumption that for  $k \in D_\alpha$  the contours  $\Sigma_\alpha^{\pm}$  coincide with  $\arg(w_\alpha^S(k; t)) = \pm 2\pi/3$ , one has the following.

- $\dot{\mathbf{O}}_\alpha^S(k)$  is analytic for  $k \in D_\alpha \setminus (\Sigma_\alpha^{S+} \cup \Sigma_\alpha^{S-} \cup \Sigma^B)$  (there is no jump across  $\Sigma_\alpha^{S0}$ ).
- $\dot{\mathbf{O}}_\alpha^S(k)$  satisfies the jump conditions

$$\dot{\mathbf{O}}_{\alpha+}^S(k) = \dot{\mathbf{O}}_{\alpha-}^S(k) \begin{bmatrix} 1 & -e^{-2i\phi_\alpha(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \Sigma_\alpha^{S+} \cap D_\alpha, \tag{209}$$

$$\dot{\mathbf{O}}_{\alpha+}^S(k) = \dot{\mathbf{O}}_{\alpha-}^S(k) \begin{bmatrix} 1 & 0 \\ e^{2i\phi_\alpha(k;t)/\epsilon} & 1 \end{bmatrix}, \quad k \in \Sigma_\alpha^{S-} \cap D_\alpha, \tag{210}$$

and

$$\dot{\mathbf{O}}_{\alpha+}^S(k) = \dot{\mathbf{O}}_{\alpha-}^S(k) \begin{bmatrix} e^{2(\Delta(k;t) - \tau(k))/\epsilon} & -e^{2i\phi_B(t)/\epsilon} \\ e^{2i\phi_B(t)/\epsilon} & 0 \end{bmatrix}, \quad k \in \Sigma^B \cap D_\alpha. \tag{211}$$

- Inner and outer parametrices match well on the disk boundary:

$$\dot{\mathbf{O}}_\alpha^S(k) \mathbf{O}^{(\text{out})}(k)^{-1} = \mathbb{I} + \mathcal{O}(\epsilon), \quad k \in \partial D_\alpha. \tag{212}$$

- $\det(\dot{\mathbf{O}}_\alpha^S(k)) = 1$  and  $\dot{\mathbf{O}}_\alpha^S(k) = \mathcal{O}(\epsilon^{-1/6})$  holds uniformly for  $k \in D_\alpha$ .

The proof is virtually the same as before, with the main difference being that the identity (206) implies that if  $(k_\alpha, \mathbf{a}(t))$  is a saturated region S, then the function  $-\phi_\alpha(k; t) \pm i\tau(k)$  for  $k \in S$  is the analytic continuation through  $\mathbb{C}_\pm$  of the function  $-\phi_B(t) \pm i\Delta(k; t)$ .

Now consider the disk  $D_\beta$ . If  $(\beta, k_\beta)$  is a void V, we define a conformal map  $w_\beta^V : D_\beta \rightarrow \mathbb{C}$  by continuation from B of the formula  $w_\beta^V(k; t) := (2\Delta(k; t))^{2/3} > 0$ . Introduce a uniformly bounded, unimodular, and holomorphic matrix  $\mathbf{H}_\beta^V(k)$  for  $k \in D_\beta$  by writing

$$\mathbf{O}^{(\text{out})}(k) = \mathbf{H}_\beta^V(k)(-w_\beta^V(k; t))^{\sigma_3/4} \mathbf{S}^{-1} e^{i\phi_B(t)\sigma_3/\epsilon}, \quad k \in D_\beta. \tag{213}$$

Then define an inner parametrix in  $D_\beta$  by the formula

$$\dot{\mathbf{O}}_b^V(k) := \mathbf{H}_b^V(k)\epsilon^{\sigma_3/6}(-i\sigma_1)\mathbf{P}(\epsilon^{-2/3}w_b^V(k))(i\sigma_1)e^{i\phi_B(t)\sigma_3/\epsilon}, \quad k \in D_\beta. \quad (214)$$

Assuming that for  $k \in D_\beta$  the contours  $\Sigma_b^{V\pm}$  coincide with segments with angles  $\arg(w_b^V) = \mp 2\pi/3$ , this parametrix satisfies the following conditions.

- $\dot{\mathbf{O}}_b^V(k)$  is analytic for  $k \in D_\beta \setminus (\Sigma_b^{V+} \cup \Sigma_b^{V-} \cup \Sigma^B)$  (there is no jump across  $\Sigma_b^{V0}$ ).
- $\dot{\mathbf{O}}_b^V(k)$  satisfies the jump conditions

$$\dot{\mathbf{O}}_{b+}^V(k) = \dot{\mathbf{O}}_{b-}^V(k) \begin{bmatrix} 0 & -e^{-2i\phi_B(t)/\epsilon} \\ e^{2i\phi_B(t)/\epsilon} & e^{-2\Delta(k;t)/\epsilon} \end{bmatrix}, \quad k \in \Sigma^B \cap D_\beta, \quad (215)$$

$$\dot{\mathbf{O}}_{b+}^V(k) = \dot{\mathbf{O}}_{b-}^V(k) \begin{bmatrix} 1 & 0 \\ e^{2i\phi_b(k;t)/\epsilon} & 1 \end{bmatrix}, \quad k \in \Sigma_b^{V+} \cap D_\beta, \quad (216)$$

and

$$\dot{\mathbf{O}}_{b+}^V(k) = \dot{\mathbf{O}}_{b-}^V(k) \begin{bmatrix} 1 & -e^{-2i\phi_b(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \Sigma_b^{V-} \cap D_\beta. \quad (217)$$

- Inner and outer parametrices match well on the disk boundary:

$$\dot{\mathbf{O}}_b^V(k)\dot{\mathbf{O}}^{(\text{out})}(k)^{-1} = \mathbb{I} + \mathcal{O}(\epsilon), \quad k \in \partial D_\beta. \quad (218)$$

- $\det(\dot{\mathbf{O}}_b^V(k)) = 1$  and  $\dot{\mathbf{O}}_b^V(k) = \mathcal{O}(\epsilon^{-1/6})$  holds uniformly for  $k \in D_\beta$ .

If instead  $(\beta, k_b)$  is a saturated region S, then start with the conformal map defined in  $D_\beta$  by the continuation from B of the formula  $w_b^S(k; t) := (2\tau(k) - 2\Delta(k; t))^{2/3} > 0$ . Introduce the matrix  $\mathbf{H}_b^S(k)$  bounded, holomorphic, and unimodular in  $D_\beta$ , by

$$\dot{\mathbf{O}}^{(\text{out})}(k) = \mathbf{H}_b^S(k)(-w_b^S(k; t))^{\sigma_3/4}\mathbf{S}^{-1}e^{i\phi_B\sigma_3/\epsilon}, \quad k \in D_\beta. \quad (219)$$

Then set

$$\dot{\mathbf{O}}_b^S(k) := \mathbf{H}_b^S(k)\epsilon^{\sigma_3/6}(-i\sigma_2)\mathbf{P}(\epsilon^{-2/3}w_b^S(k; t))(i\sigma_3)e^{2i\phi_B(t)\sigma_3/\epsilon}, \quad k \in D_\beta. \quad (220)$$

Assuming that for  $k \in D_\beta$  the contours  $\Sigma_b^{S\pm}$  coincide with segments with angles  $\arg(w_b^S) = \mp 2\pi/3$ , this parametrix satisfies these conditions.

- $\dot{\mathbf{O}}_b^S(k)$  is analytic for  $k \in D_\beta \setminus (\Sigma_b^{S+} \cup \Sigma_b^{S-} \cup \Sigma^B)$  (there is no jump across  $\Sigma_b^{S0}$ ).
- $\dot{\mathbf{O}}_b^S(k)$  satisfies the jump conditions

$$\dot{\mathbf{O}}_{b+}^S(k) = \dot{\mathbf{O}}_{b-}^S(k) \begin{bmatrix} e^{2(\Delta(k;t)-\tau(k))/\epsilon} & -e^{-2i\phi_B(t)/\epsilon} \\ e^{2i\phi_B(t)/\epsilon} & 0 \end{bmatrix}, \quad k \in \Sigma^B \cap D_\beta, \quad (221)$$

$$\dot{\mathbf{O}}_{b^+}^S(k) = \dot{\mathbf{O}}_{b^-}^S(k) \begin{bmatrix} 1 & -e^{-2i\phi_b(k;t)/\epsilon} \\ 0 & 1 \end{bmatrix}, \quad k \in \Sigma_b^{S^+} \cap D_\beta, \quad (222)$$

and

$$\dot{\mathbf{O}}_{b^+}^S(k) = \dot{\mathbf{O}}_{b^-}^S(k) \begin{bmatrix} 1 & 0 \\ e^{2i\phi_b(k;t)/\epsilon} & 1 \end{bmatrix}, \quad k \in \Sigma_b^{S^-} \cap D_\beta. \quad (223)$$

- Inner and outer parametrices match well on the disk boundary:

$$\dot{\mathbf{O}}_b^S(k) \dot{\mathbf{O}}^{(\text{out})}(k)^{-1} = \mathbb{I} + \mathcal{O}(\epsilon), \quad k \in \partial D_\beta. \quad (224)$$

- $\det(\dot{\mathbf{O}}_b^S(k)) = 1$  and  $\dot{\mathbf{O}}_b^S(k) = \mathcal{O}(\epsilon^{-1/6})$  holds uniformly for  $k \in D_\beta$ .

Now we combine the outer and inner parametrices into a *global parametrix*  $\dot{\mathbf{O}}(k)$  defined as follows. If  $\alpha'(t) < 0$  and  $b'(t) > 0$  (case VBV), then

$$\dot{\mathbf{O}}(k) := \begin{cases} \dot{\mathbf{O}}_\alpha^V(k), & k \in D_\alpha \\ \dot{\mathbf{O}}_b^V(k), & k \in D_\beta \\ \dot{\mathbf{O}}^{(\text{out})}(k), & k \in \mathbb{C} \setminus (\overline{D}_\alpha \cup \overline{D}_\beta). \end{cases} \quad (225)$$

If  $\alpha'(t) < 0$  and  $b'(t) < 0$  (case VBS), then

$$\dot{\mathbf{O}}(k) := \begin{cases} \dot{\mathbf{O}}_\alpha^V(k), & k \in D_\alpha \\ \dot{\mathbf{O}}_b^S(k), & k \in D_\beta \\ \dot{\mathbf{O}}^{(\text{out})}(k), & k \in \mathbb{C} \setminus (\overline{D}_\alpha \cup \overline{D}_\beta). \end{cases} \quad (226)$$

If  $\alpha'(t) > 0$  and  $b'(t) > 0$  (case SBV), then

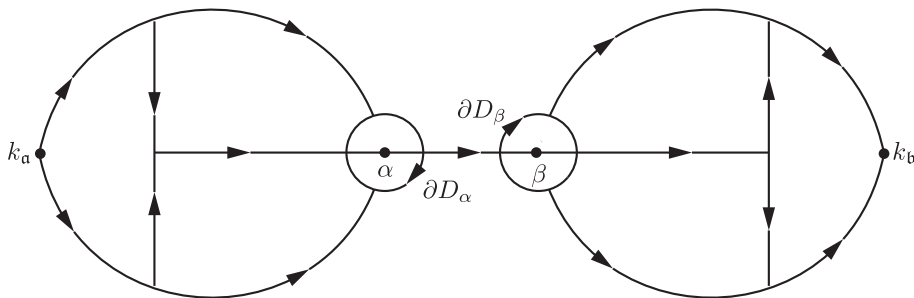
$$\dot{\mathbf{O}}(k) := \begin{cases} \dot{\mathbf{O}}_\alpha^S(k), & k \in D_\alpha \\ \dot{\mathbf{O}}_b^V(k), & k \in D_\beta \\ \dot{\mathbf{O}}^{(\text{out})}(k), & k \in \mathbb{C} \setminus (\overline{D}_\alpha \cup \overline{D}_\beta). \end{cases} \quad (227)$$

Finally, if  $\alpha'(t) > 0$  and  $b'(t) < 0$  (case SBS), then

$$\dot{\mathbf{O}}(k) := \begin{cases} \dot{\mathbf{O}}_\alpha^S(k), & k \in D_\alpha \\ \dot{\mathbf{O}}_b^S(k), & k \in D_\beta \\ \dot{\mathbf{O}}^{(\text{out})}(k), & k \in \mathbb{C} \setminus (\overline{D}_\alpha \cup \overline{D}_\beta). \end{cases} \quad (228)$$

The global parametrix is intended to be a good model for  $\mathbf{O}(k)$  in the whole complex plane. To evaluate this claim, consider the error  $\mathbf{E}(k)$  defined as

$$\mathbf{E}(k) := \mathbf{O}(k) \dot{\mathbf{O}}(k)^{-1}. \quad (229)$$



**Figure 13.** The jump contour  $\Sigma^E$  for the error matrix  $\mathbf{E}(k)$ . All contour arcs are labeled as in Figure 12 (different labels in the four cases) with the exception of the disk boundaries  $\partial D_\alpha$  and  $\partial D_\beta$  (both oriented clockwise). Note that the subarcs of  $\Sigma_a^{V\pm}$  or  $\Sigma_a^{S\pm}$  within  $D_\alpha$  and the subarcs of  $\Sigma_b^{V\pm}$  or  $\Sigma_b^{S\pm}$  within  $D_\beta$  are absent from  $\Sigma^E$ , because the inner parametrices satisfy exactly the jump conditions of  $\mathbf{O}(k)$  on these subarcs.

Because  $\mathbf{O}(k)$  satisfies the conditions of Riemann–Hilbert problem 6 while  $\dot{\mathbf{O}}(k)$  is known explicitly,  $\mathbf{E}(k)$  solves a Riemann–Hilbert problem equivalent to that for  $\mathbf{O}(k)$ . This problem is the following.

**Riemann–Hilbert problem 7.** Find a  $2 \times 2$  matrix  $\mathbf{E}(k)$  with the following properties:

**Analyticity:**  $\mathbf{E}(k)$  is analytic for  $k \in \mathbb{C} \setminus \Sigma^E$ , where  $\Sigma^E$  is the contour illustrated in Figure 13, and  $\mathbf{E}(k)$  takes continuous boundary values  $\mathbf{E}_+(k)$  and  $\mathbf{E}_-(k)$  on each oriented arc of  $\Sigma^E$  from the left and right, respectively.

**Jump Condition:** The boundary values on each oriented arc of  $\Sigma^E$  are related by  $\mathbf{E}_+(k) = \mathbf{E}_-(k)\mathbf{J}^E(k)$  (see below for a complete explicit characterization of the jump matrix  $\mathbf{J}^E(k)$ ).

**Normalization:**  $\mathbf{E}(k) \rightarrow \mathbb{I}$  as  $k \rightarrow \infty$ .

On all arcs of  $\Sigma^E$  with the exception of (i) the disk boundaries  $\partial D_\alpha$  and  $\partial D_\beta$ , (ii) the real arcs of  $\Sigma_a^{V0}$  or  $\Sigma_a^{S0}$  within  $D_\alpha$ , (iii) the real arcs of  $\Sigma_b^{V0}$  or  $\Sigma_b^{S0}$  within  $D_\beta$ , and (iv) the real arc  $\Sigma^B$  (including parts inside and outside the disks), the jump matrix  $\mathbf{J}^E(k)$  is given by

$$\mathbf{J}^E(k) := \dot{\mathbf{O}}^{(\text{out})}(k)\mathbf{J}(k)\dot{\mathbf{O}}^{(\text{out})}(k)^{-1}. \tag{230}$$

Here,  $\mathbf{J}(k)$  is the jump matrix for Riemann–Hilbert problem 6 characterizing  $\mathbf{O}(k)$ , and it has already been shown that  $\mathbf{J}(k) - \mathbb{I}$  is uniformly  $\mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  on these arcs of  $\Sigma^E$ . Since  $\dot{\mathbf{O}}^{(\text{out})}(k)$  is unimodular and bounded independently of  $\epsilon$  away from  $k = \alpha$  and  $k = \beta$ , it follows that also  $\mathbf{J}^E(k) - \mathbb{I} = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  holds uniformly on these arcs.

The discontinuity of  $\mathbf{E}(k)$  across the boundary of the two disks is caused by the mismatch of the inner and outer parametrices, since  $\mathbf{O}(k)$  has no jump.

Letting  $\dot{\mathbf{O}}_\alpha(k)$  denote the inner parametrix installed in  $D_\alpha$  and  $\dot{\mathbf{O}}_\beta(k)$  denote the inner parametrix installed in  $D_\beta$ , a calculation shows that

$$\mathbf{J}^E(k) = \dot{\mathbf{O}}_{\alpha, \beta}(k) \dot{\mathbf{O}}^{(\text{out})}(k)^{-1}, \quad k \in \partial D_{\alpha, \beta}. \tag{231}$$

But using (205) or (213) for  $k \in \partial D_\alpha$ , and using (219) or (225) for  $k \in \partial D_\beta$ , we conclude that  $\mathbf{J}^E(k) - \mathbb{I} = \mathcal{O}(\epsilon)$  holds uniformly on both  $\partial D_\alpha$  and  $\partial D_\beta$ .

If  $k$  lies on the real segment of  $\Sigma_\alpha^{V0}$  or  $\Sigma_\alpha^{S0}$  within  $D_\alpha$ , then

$$\mathbf{J}^E(k) = \dot{\mathbf{O}}_\alpha(k) \mathbf{J}(k) \dot{\mathbf{O}}_\alpha(k)^{-1}, \quad k \in (\Sigma_\alpha^{V0} \cup \Sigma_\alpha^{S0}) \cap D_\alpha, \tag{232}$$

and similarly

$$\mathbf{J}^E(k) = \dot{\mathbf{O}}_\beta(k) \mathbf{J}(k) \dot{\mathbf{O}}_\beta(k)^{-1}, \quad k \in (\Sigma_\beta^{V0} \cup \Sigma_\beta^{S0}) \cap D_\beta, \tag{233}$$

since the inner parametrix has no jump. In all four cases we have shown that  $\mathbf{J}(k) - \mathbb{I}$  is exponentially small in the limit  $\epsilon \downarrow 0$ , so even though each conjugating factor amplifies this by  $\epsilon^{-1/6}$ , we still have decay beyond all orders in  $\epsilon$ .

If  $k$  lies on the band  $\Sigma^B$  within either disk, then both  $\mathbf{O}(k)$  and the (inner) parametrix  $\dot{\mathbf{O}}(k)$  are discontinuous. If the band edge abuts a void, a calculation shows that

$$\begin{aligned} \mathbf{J}^E(k) &= \dot{\mathbf{O}}_{\alpha, \beta-}^V(k) \begin{bmatrix} Y^\epsilon(k) + e^{-2\tau(k)/\epsilon} & e^{-2i\phi_B(t)/\epsilon} e^{2(\Delta(k;t) - \tau(k))/\epsilon} \\ (Y^\epsilon(k) - 1)e^{2i\phi_B(t)/\epsilon} e^{-2\Delta(t;k)/\epsilon} & Y^\epsilon(k) \end{bmatrix} \\ &\cdot \dot{\mathbf{O}}_{\alpha, \beta-}^V(k)^{-1}. \end{aligned} \tag{234}$$

The central factor is an exponentially small perturbation of  $\mathbb{I}$  because  $Y^\epsilon(k) - 1$  is exponentially small and both  $\tau(k)$  and  $\tau(k) - \Delta(k;t)$  are strictly positive. The conjugating factors amplify this by  $\epsilon^{-1/3}$ , but this remains beyond all orders small. Similarly, if the band edge abuts a saturated region, then

$$\begin{aligned} \mathbf{J}^E(k) &= \dot{\mathbf{O}}_{\alpha, \beta-}^S(k) \begin{bmatrix} Y^\epsilon(k) & (1 - Y^\epsilon(k))e^{-2i\phi_B(t)/\epsilon} e^{2(\Delta(k;t) - \tau(k))/\epsilon} \\ -e^{2i\phi_B(t)/\epsilon} e^{-2\Delta(k;t)/\epsilon} & Y^\epsilon(k) + e^{-2\tau(k)/\epsilon} \end{bmatrix} \\ &\cdot \dot{\mathbf{O}}_{\alpha, \beta-}^S(k)^{-1}, \end{aligned} \tag{235}$$

and again  $\mathbf{J}^E - \mathbb{I}$  is small beyond all orders despite the conjugating factors algebraically large size.

Finally, consider  $k \in \Sigma^B$  outside both disks. Again both  $\mathbf{O}(k)$  and the (outer) parametrix  $\dot{\mathbf{O}}(k)$  have jump discontinuities across this segment, and a calculation shows that

$$\mathbf{J}^{\mathbb{E}}(k) = \dot{\mathbf{O}}_-^{(\text{out})}(k) \begin{bmatrix} Y^\epsilon(k) & -e^{-2i\phi_B(t)/\epsilon} e^{-2\Delta(k;t)/\epsilon} \\ e^{2i\phi_B(t)/\epsilon} e^{2(\Delta(k;t)-\tau(k))/\epsilon} & Y^\epsilon(k) \end{bmatrix} \dot{\mathbf{O}}_-^{(\text{out})}(k)^{-1}, \tag{236}$$

which is exponentially close to  $\mathbb{I}$  because the outer parametrix and its inverse are uniformly bounded (away from  $k = \alpha$  and  $k = \beta$ ), while  $Y^\epsilon(k) - 1$  is exponentially small and  $\Delta(k; t)$  and  $\tau(k) - \Delta(k; t)$  are both strictly positive as  $k$  lies on a compact subset of the interior of the band  $B$ .

These considerations prove that  $\mathbf{J}^{\mathbb{E}}(k) - \mathbb{I}$  is uniformly  $\mathcal{O}((\log(\epsilon^{-1}))^{-1/2})$  on the  $\epsilon$ -independent contour  $\Sigma^{\mathbb{E}}$ . Therefore, in the  $\epsilon \downarrow 0$  limit, Riemann–Hilbert problem 7 is a small-norm Riemann–Hilbert problem in the  $L^2(\Sigma^{\mathbb{E}})$  sense. This implies that  $\mathbf{E}(k)$  exists for  $\epsilon > 0$  sufficiently small and has a convergent (because  $\Sigma^{\mathbb{E}}$  is bounded) Laurent expansion for sufficiently large  $|k|$  of the form

$$\mathbf{E}(k) = \mathbb{I} + k^{-1}\mathbf{E}_1 + k^{-2}\mathbf{E}_2 + \mathcal{O}(k^{-3}), \quad k \rightarrow \infty, \tag{237}$$

with the first two moments satisfying

$$\mathbf{E}_1 = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2}) \quad \text{and} \quad \mathbf{E}_2 = \mathcal{O}((\log(\epsilon^{-1}))^{-1/2}), \quad \epsilon \downarrow 0. \tag{238}$$

At last, we obtain the exact formula for  $\tilde{\mathbf{M}}(k)$  valid for sufficiently large  $|k|$ :

$$\tilde{\mathbf{M}}(k) = \mathbf{N}(k)e^{ig(k)\sigma_3/\epsilon} = \mathbf{O}(k)e^{ig(k)\sigma_3/\epsilon} = \mathbf{E}(k)\dot{\mathbf{O}}_-^{(\text{out})}(k)e^{ig(k)\sigma_3/\epsilon}. \tag{239}$$

Taking into account the Laurent expansion of  $g$  given by  $g(k; t) = k^{-1}g_1(t) + k^{-2}g_2(t) + \mathcal{O}(k^{-3})$  as  $k \rightarrow \infty$ , using (190) with (191), and using (237) with (238), it therefore follows from (67) that in the limit  $\epsilon \downarrow 0$ ,

$$\begin{aligned} \tilde{q}^\epsilon(0, t) &= \frac{\beta - \alpha}{2} e^{-2i\phi_B(t)/\epsilon} + \mathcal{O}((\log(\epsilon^{-1}))^{-1/2}) \\ \epsilon \tilde{q}_x^\epsilon(0, t) &= -\frac{1}{2}i(\beta^2 - \alpha^2)e^{-2i\phi_B(t)/\epsilon} + \mathcal{O}((\log(\epsilon^{-1}))^{-1/2}), \end{aligned} \tag{240}$$

where we have used (190) and (238). Since  $\alpha = a(t)$  and  $\beta = b(t)$ , using (40) and Lemma 9 we obtain

$$\begin{aligned} \tilde{q}^\epsilon(0, t) &= H(t)e^{iS(t)/\epsilon} + \mathcal{O}((\log(\epsilon^{-1}))^{-1/2}) \\ \epsilon \tilde{q}_x^\epsilon(0, t) &= iU(t)H(t)e^{iS(t)/\epsilon} + \mathcal{O}((\log(\epsilon^{-1}))^{-1/2}). \end{aligned} \tag{241}$$

The error terms are uniform on compact subintervals of  $t \in (0, +\infty) \setminus \{t_a, t_b\}$ . Indeed, the  $L^\infty(\Sigma^{\mathbb{E}})$  estimate on  $\mathbf{J}^{\mathbb{E}}(k) - \mathbb{I}$  fails only as  $t \rightarrow 0$  or  $t \rightarrow \infty$  (in which case the band  $B$  shrinks to a point), as  $t \rightarrow t_a$  (in which case the void or saturated region on the left of the band shrinks to a point), and as  $t \rightarrow t_b$  (in which case the void or saturated region on the right of the band shrinks to a point). Furthermore, by a minor modification of the preceding arguments in which the disks  $D_\alpha$  and  $D_\beta$  are allowed to be slightly off-center from the points  $\alpha$  and  $\beta$ , it can be shown that for  $t$  in a given compact subinterval of  $(0, +\infty) \setminus \{t_a, t_b\}$ , Riemann–Hilbert problem 7 may be formulated with a finite

number of different contours  $\Sigma^E$ . Therefore, the norm of the Cauchy projection operator  $C_{-}^{\Sigma^E} : L^2(\Sigma^E) \rightarrow L^2(\Sigma^E)$  is uniformly bounded (independent of  $\epsilon$ , of course) as a maximum over a finite number of values (see proposition 3 in Section 3.6 of [17] and Section 4.6 of [17] for further information and more details about this procedure). This completes the proof of Theorem 2.

*4.2.3 Proof of Corollary 2.* The main idea of the proof is to observe that the construction of  $g(k; t)$  for  $x = 0$  and  $t > 0$  presented in Section 4.2.1 may be subjected to continuation for small  $x > 0$ . Consider the prospect of continuation of the solution  $(\alpha, \beta) = (\alpha(0, t), \beta(0, t)) = (\mathfrak{a}(t), \mathfrak{b}(t))$  of the equations (143) for  $x = 0$  to nonzero  $x$  by means of the implicit function theorem. This requires calculating the Jacobian of the system (143). We first establish the following.

LEMMA 12. *The functions  $m_j(\alpha, \beta; x, t)$  for  $j = 1, 2$  defined by (143) satisfy the identities*

$$\frac{\partial m_2}{\partial \alpha} = \frac{1}{2}m_1 + \alpha \frac{\partial m_1}{\partial \alpha} \quad \text{and} \quad \frac{\partial m_2}{\partial \beta} = \frac{1}{2}m_1 + \beta \frac{\partial m_1}{\partial \beta}. \tag{242}$$

*Proof:* First, one establishes the related identities

$$\frac{\partial I_2}{\partial \alpha} = \frac{1}{2}I_1 + \alpha \frac{\partial I_1}{\partial \alpha} \quad \text{and} \quad \frac{\partial I_2}{\partial \beta} = \frac{1}{2}I_1 + \beta \frac{\partial I_1}{\partial \beta} \tag{243}$$

by replacing the integrals in (144) in a standard way by contour integrals over contours locally independent of  $\alpha$  and  $\beta$  that are bounded away from these points, and then differentiating under the integral sign. Then one eliminates  $I_1(\alpha, \beta)$  and  $I_2(\alpha, \beta)$  in favor of  $m_1(\alpha, \beta; x, t)$  and  $m_2(\alpha, \beta; x, t)$  using (143) to finish the proof. ■

Whenever  $(\alpha, \beta)$  satisfy the equations (143), as is the case for  $x = 0$  and  $t > 0$  with  $(\alpha, \beta) = (\mathfrak{a}(t), \mathfrak{b}(t))$  by Lemma 7, the Jacobian is easily calculated with the help of Lemma 12:

$$\frac{\partial m_1}{\partial \alpha} \frac{\partial m_2}{\partial \beta} - \frac{\partial m_1}{\partial \beta} \frac{\partial m_2}{\partial \alpha} = (\beta - \alpha) \frac{\partial m_1}{\partial \alpha} \frac{\partial m_1}{\partial \beta}. \tag{244}$$

When  $x = 0$  and  $t > 0$ , we have  $\beta - \alpha = 2H(t) > 0$ . It can be shown that in this situation also

$$\frac{\partial m_1}{\partial \alpha}(\mathfrak{a}(t), \mathfrak{b}(t); 0, t) \neq 0 \quad \text{and} \quad \frac{\partial m_1}{\partial \beta}(\mathfrak{a}(t), \mathfrak{b}(t); 0, t) \neq 0, \quad t \in (0, +\infty) \setminus \{t_a, t_b\}. \tag{245}$$

Indeed, it turns out that  $m_{1\alpha} = 0$  if and only if the function  $\Delta'(k)$  or  $\tau'(k) - \Delta'(k)$  vanishes as  $k \downarrow \alpha$  to higher order than  $(k - \alpha)^{1/2}$  depending on whether  $(k_\alpha, \alpha)$  is a void  $V$  or a saturated region  $S$ . Similarly,  $m_{1\beta} = 0$  detects higher-order vanishing of either  $\Delta'(k)$  or  $\tau'(k) - \Delta'(k)$  at  $k = \beta$ . It is easy to see from the explicit construction of  $g$  given in Section 4.2.1 that as long as

$t \neq t_a$  or  $t_b$ , then  $m_{1\alpha} \neq 0$  and  $m_{1\beta} \neq 0$ . Therefore, the Jacobian is nonzero, and so the following is true.

LEMMA 13. *Suppose  $t_0 > 0$  and  $t \neq t_a$ ,  $t \neq t_b$ , so that the complex phase function  $g(k; t_0)$  is well-defined in a particular configuration,  $VBV$ ,  $VBS$ ,  $SBV$ , or  $SBS$ . Then there is a neighborhood of the point  $(0, t_0)$  in the  $(x, t)$ -plane in which there is a unique solution  $(\alpha(x, t), \beta(x, t))$  of the equations (143) formulated for the same configuration that satisfies the boundary condition  $(\alpha(0, t), \beta(0, t)) = (a(t), b(t))$ . The solution is differentiable and the partial derivatives satisfy the identities (35).*

*Proof.* It only remains to establish the identities (35). However, this can be accomplished simply by implicit differentiation of the equations  $m_1(\alpha(x, t), \beta(x, t); x, t) = m_2(\alpha(x, t), \beta(x, t); x, t) = 0$  with respect to  $x$  and  $t$ , using Lemma 12 with  $m_1 = m_2 = 0$ . ■

This result allows us to define a candidate complex phase function  $g = g(k; x, t)$  given  $\alpha$  and  $\beta$  as described in Section 4.2.1. Moreover, it is easy to see that the inequalities, guaranteed for  $x = 0$  in the particular configuration of voids and saturated regions valid for  $t = t_0$  according to Lemma 8, persist for  $g(k; x, t)$  in the same configuration for small  $x \neq 0$ . The existence of an appropriate complex phase function  $g$  was the only essential ingredient in the proof of Theorem 2. Therefore, the same result holds, with  $\alpha(x, t)$  and  $\beta(x, t)$  taking the place of  $a(t)$  and  $b(t)$ , respectively, and the proof of Corollary 2 is complete.

## 5. Concluding remarks

Here, we collect together a few additional remarks, some indicating natural generalizations of our results, and some pointing toward directions for future research.

*Remark 5.* The methodology of the complex phase function as described in Section 4.2.1 is capable of detecting and describing asymptotic behavior of  $\tilde{q}^\epsilon(x, t)$  that neither corresponds to the vacuum domain nor the plane-wave domain. A typical scenario would be that as one tunes  $(x, t)$  through the plane-wave domain by continuing the complex phase function  $g$  along some path, at some point  $(x_0, t_0)$  one of the inequalities associated with the complex phase function  $g = g(k; x, t)$  fails for some  $k \in (k_a, k_b)$  (a less generic scenario involves the Jacobian of the system (143) vanishing). Further continuation becomes impossible without resorting to the device of increasing  $N$ , where we recall that  $N + 1$  is the number of bands in  $(k_a, k_b)$ . Thus, one is witnessing the birth of a new band, void, or saturated region from the point  $k$  as one continues beyond  $(x_0, t_0)$ . Once  $N > 0$ , instead of an asymptotic formula like (36), one arrives at an asymptotic formula written in terms of



hyperelliptic functions of genus  $N$ , a formula that exhibits rapid variations of the amplitude on space and time scales proportional to  $\epsilon$ . Thus, if  $N > 0$ ,  $\tilde{q}^\epsilon(x, t)$  no longer resembles a slowly modulated plane wave, but rather has a more complicated microstructure. The point  $(x_0, t_0)$  therefore lies along the common boundary between domains corresponding to  $N = 0$  and (in the next simplest case)  $N = 1$ . In general, one expects the quarter plane to be tiled with  $\epsilon$ -independent domains corresponding to various values of  $N \geq 0$ , with the only general statement being that the positive  $t$ -axis abuts the (plane-wave) domain with  $N = 0$  and the positive  $x$ -axis abuts the vacuum domain (which is not associated with any value of  $N$ ). These phase transitions may in principal be computed, with the details depending on the particular boundary data given at  $x = 0$  and  $t > 0$ . We leave such considerations for future studies. Note, however, that under the assumptions in force in this paper, it is impossible for a dispersive shock wave region with  $N > 0$  that is generated from a gradient catastrophe in the solution of the dispersionless system (38) at some time  $(x, t) = (x_c, t_c)$  to reach the boundary  $x = 0$  in finite time, since such behavior would be inconsistent with Theorem 2.

*Remark 6.* We wish to briefly explain the key obstruction in our opinion that arises in the semiclassical analysis of Riemann–Hilbert problem 1 posed in Section 1 relative to a cross contour  $\Sigma = \mathbb{R} \cup i\mathbb{R}$ , and hence our actual motivation in introducing the modified approach described in Section 3.

A calculation starting from (138) in the VB configuration shows that  $g'(k; t) \rightarrow 0$  as  $t \downarrow 0$  (for  $x = 0$ ), and hence from (140) we obtain  $g(k; 0) = 0$  (this is consistent with our approach for  $x > 0$  and  $t = 0$ , which basically assumed  $g(k) = 0$ ). Therefore, for  $x = 0$  and  $t > 0$  we have

$$g(t; k) = \int_0^t g_t(k; s) ds. \tag{246}$$

Using (154) and assuming that  $k$  is purely imaginary gives

$$\Im\{g(t; k)\} = \int_0^t \Im\{g_t(k; s)\} ds = \int_0^t \Im\{(U(s) - 2k)r(k; s)\} ds, \quad k \in i\mathbb{R}. \tag{247}$$

Note that

$$\begin{aligned} (U(t) - 2k)r(k; t) &= -2k^2 + \frac{1}{2}U(t)^2 + H(t)^2 \\ &\quad - H(t)^2U(t)k^{-1} + \mathcal{O}(k^{-2}), \quad k \rightarrow \infty, \end{aligned} \tag{248}$$

which together with (247) implies that  $\Im\{g(t; k)\} > 0$  holds whenever  $k$  is positive imaginary and of sufficiently large magnitude, because  $H(t)^2U(t) > 0$  for  $t > 0$  by Assumption 1.

Consider now the effect of introducing  $g$  not into Riemann–Hilbert problem 2 for  $\tilde{\mathbf{M}}(k)$  having (a subset of)  $\mathbb{R}$  as the jump contour, but rather into Riemann–Hilbert problem 1 for  $\mathbf{M}(k)$  having  $\Sigma = \mathbb{R} \cup i\mathbb{R}$  as the jump contour. In the latter problem one has a lower-triangular jump matrix on the positive imaginary axis given by (12). Thus, if one introduces the complex phase function  $g(k; t)$  by means of a formula analogous to (131), the resulting jump condition for  $\mathbf{N}(k)$  on the positive imaginary axis (oriented toward the origin) would read

$$\mathbf{N}_+(k) = \mathbf{N}_-(k) \begin{bmatrix} 1 & 0 \\ -\Gamma_0(k)e^{2i\theta(k;0,t)/\epsilon}e^{-2ig(k;t)/\epsilon} & 1 \end{bmatrix}, \quad \arg(k) = \frac{\pi}{2}. \quad (249)$$

The factor  $e^{2i\theta(k;0,t)/\epsilon}$  is purely oscillatory for imaginary  $k$ , and by Proposition 1 we have an upper bound for  $|\Gamma_0(k)|$  that decays algebraically to zero with  $\epsilon$ . However, we now see that the factor  $e^{-2ig(k;t)/\epsilon}$  is exponentially large in the limit  $\epsilon \downarrow 0$  for sufficiently large  $|k|$  with  $\arg(k) = \pi/2$ . Therefore, the same complex phase function  $g$  that works so well to control the modified problem for  $\tilde{\mathbf{M}}(k)$  produces uncontrollable errors if used to study  $\mathbf{M}(k)$  itself. This difficulty originates in the jump discontinuities along the imaginary axis, which appear to be negligible before  $g$  is introduced but that seem impossible to either neglect afterwards or include in a parametrix without explicit beyond-all-orders information about  $\Gamma_0(k)$ .

*Remark 7.* As Corollary 2 is formulated, the points  $t = t_a$  and  $t = t_b$  appear to present an obstruction to continuation of the plane-wave approximation to positive  $x$  from the boundary. While the details are not easy to explain, the fact is that the complex phase function  $g$  can indeed be continued away from the boundary near such points, which become curves  $t = t_a(x)$  and  $t = t_b(x)$  for  $x > 0$  along which one of the band endpoints is fixed to an endpoint of the interval  $[k_a, k_b]$  (see, for example, Section 4.3.2 of [20]). There still remains, however, a technical issue in that it appears that a new type of inner parametrix is required near the corresponding point in the  $k$ -plane, and to our knowledge this has not been worked out. Although we are therefore prevented from proving convergence to a plane wave for  $(x, t)$  near the curves  $t = t_a(x)$  and  $t = t_b(x)$ , one can check that the plane wave formulae that are indeed valid as given in Corollary 2 on either side of these curves actually match on the curves themselves. Thus, one does not expect any new leading-order asymptotic behavior for  $\tilde{q}^\epsilon(x, t)$  along these curves, which nonetheless present an obstruction to rigorous analysis.

*Remark 8.* The reader will observe that the very slow rate of decay of the error terms in our results (proportional to  $(\log(\epsilon^{-1}))^{-1/2}$ ) is in all cases due to the truncation introduced in Section 3 of the jump matrix on  $\mathbb{R}$  to one whose difference from  $\mathbb{I}$  is compactly supported in the interval  $[k_a, k_b]$ . Indeed this truncation leads to dominant contributions to the error generated from small

neighborhoods of the endpoints of this support interval. This suggests that it may be possible to further modify the formula for  $\tilde{\Gamma}(k)$  near these two points with the aim of reducing the magnitude of these local contributions to the error without having any other significant effect.

*Remark 9.* In our view, the incorporation of nonzero initial conditions together with a nonlinearizable boundary condition presents substantial new difficulties. Physically, one expects a nontrivial interaction between the initial and boundary data, and from the “hyperbolic” point of view of the dispersionless nonlinear Schrödinger system (38) one may expect competing influences at a given point  $(x, t)$  from boundary and initial data propagating along characteristics. Mathematically, the initial and boundary conditions get mixed up in the complicated jump matrix on the negative real axis as indicated in (14). The fact that at points  $k < 0$  where both  $\gamma(k)$  and  $\Gamma(k)$  are nonzero the elements of the jump matrix are differences of oscillatory exponentials in the semiclassical approximation of  $\gamma$  and  $\Gamma$  makes analysis by steepest descent methods quite challenging, and we hope to consider such issues in the future.

*Remark 10.* We may consider several ways to generalize Assumption 1 while still maintaining the form (24) of the Dirichlet boundary data.

- One could drop the condition  $U(t) > 2H(t)$  in favor of the weaker condition  $U(t) > H(t)$ . In this situation the boundary  $x = 0$  is still a spacelike curve for the hyperbolic dispersionless defocusing nonlinear Schrödinger system (38), so one has a local solution of this approximate system near the boundary for all  $t > 0$ . On the other hand, it seems that it is not possible to recover the boundary data from the Riemann–Hilbert problem in this case as pointed out in Remark 2. Is this a merely technical issue, or does this imply that a dispersive shock wave generated for positive  $x$  reaches the boundary at some later time  $t$ , ruining the local plane-wave microstructure?
- One could go further and drop the condition  $U(t) > H(t)$ . If some bicharacteristics of the hyperbolic system (38) point outside of the domain at the boundary  $x = 0$ , is it still possible for the approximate Dirichlet-to-Neumann map given in Definition 1 to be valid? Should one perhaps expect a kind of boundary layer to form near  $x = 0$ ?
- Finally, one could even drop the inequality (26). In this case the phase gradient  $u$  is apparently imaginary at the boundary. What does this mean?

It would also be very interesting to consider the nonlinear Schrödinger equation in the focusing case, in which case the analogue of the dispersionless system (38) is a quasilinear system of elliptic type, so all intuition involving propagation along characteristics is lost.

### Acknowledgments

P. D. Miller was partially supported by the National Science Foundation under grant number DMS-1206131 and by a fellowship from the Simons Foundation, grant number 267106. Part of this work was done during two visits Miller made to Fudan University in Shanghai, China, and both authors are grateful for the support Fudan University offered. The authors' collaboration began while Z. Qin visited the University of Michigan for an extended period under the support of the China Scholarship Council. Z. Qin was also supported by a grant from the Key Laboratory of Mathematics for Nonlinear Science, Fudan University.

### Appendix A: Proof of Proposition 1

We begin with a key lemma.

**LEMMA 14** (Bound on  $\Re\{\lambda\}$ ). *Suppose that Assumption 1 holds and that either  $\Im\{k\} > 0$  and  $\Re\{k\} \leq 0$ , or that  $k < k_a$ , or that  $k_b < k \leq 0$ . Then there is an eigenvalue  $\lambda$  of the matrix  $\mathbf{B}(t; k)$  defined by (46) that depends smoothly on  $t > 0$  and satisfies  $\Re\{\lambda\} \leq 0$ ,  $\forall t > 0$ .*

*Proof.* The fact that  $\lambda$  can be chosen to be a smooth function of  $t$  follows from Lemma 1 since there are no turning points and  $\lambda^2$  is a smooth function of  $t > 0$ . If  $k$  is real and  $k < k_a$  or  $k > k_b$ , then  $\lambda^2$  is real and strictly negative for all  $t > 0$ , so both eigenvalue functions  $\lambda = \lambda(t)$  will be purely imaginary for  $t > 0$ . On the other hand, if  $\Im\{k\} > 0$  but  $\Re\{k\} \leq 0$ , then the lemma can be proven by showing that  $\lambda^2$  is *not* negative real for any  $t > 0$ . Let  $k_r := \Re\{k\}$  and  $k_i := \Im\{k\}$ , and suppose that  $\lambda^2$  is real for some  $k_r \leq 0$ , some  $k_i > 0$ , and some  $t > 0$ . Using (50), we therefore have

$$\Im\{\lambda^2\} = 4k_i [4k_r k_i^2 + [2k_r - U(t)](H(t)^2 - k_r[2k_r + U(t)])] = 0. \quad (\text{A.1})$$

If  $k_r = 0$ , then as also  $k_i > 0$ , (A.1) implies that  $U(t)H(t)^2 = 0$ , which is a contradiction with  $t > 0$  by Assumption 1. On the other hand, if  $k_r < 0$ , then we may solve (A.1) for  $k_i^2$  and hence from (50) we obtain

$$\begin{aligned} \Re\{\lambda^2\} &= \left(\frac{1}{2}[2k_r - U(t)]^2 - 2k_i^2\right) \left(2H(t)^2 - \frac{1}{2}[2k_r + U(t)]^2 + 2k_i^2\right) \\ &\quad + 4k_i^2[2k_r - U(t)][2k_r + U(t)] \\ &= \frac{1}{4k_r^2} [4k_r^2 - H(t)^2]^2 [4k_r^2 - U(t)^2]. \end{aligned} \quad (\text{A.2})$$

But, at the same time (A.1) implies the inequality

$$[2k_r - U(t)](H(t)^2 - k_r[2k_r + U(t)]) > 0, \quad (\text{A.3})$$

which can be written in the equivalent form (because  $k_r < 0$  and  $U(t) > 0$  by Assumption 1)

$$U(t)^2 - 4k_r^2 < \frac{U(t) - 2k_r}{k_r} H(t)^2 < 0. \tag{A.4}$$

Obviously, (A.4) implies that  $\Re\{\lambda^2\} > 0$ . Hence,  $\lambda^2 < 0$  is not possible for any  $t > 0$ . ■

We denote by  $\lambda = \Lambda(t; k)$  the eigenvalue branch characterized by Lemma 14, so that  $\Re\{\Lambda(t; k)\} \leq 0$  for all  $t > 0$ . If  $k \in \mathcal{Q}_\delta^{\text{II}} \cap \mathbb{R}$ , then  $\Lambda(t; k) \in i\mathbb{R}$  is ambiguous up to a sign and we choose the sign so that  $\Lambda : \mathbb{R}_+ \times \mathcal{Q}_\delta^{\text{II}} \rightarrow \mathbb{C}$  is continuous. The matrix  $\mathbf{R}_0(t; k)$  given by

$$\mathbf{R}_0(t; k) := \left( \frac{1}{2}(2k - U(t))(2k + U(t)) + i\Lambda(t; k) \right) i\sigma_3 - (2k - U(t))H(t)\sigma_1 \tag{A.5}$$

is an eigenvector matrix for the coefficient matrix  $\mathbf{B}(t; k)$  defined by (46), i.e., the identity  $\mathbf{B}(t; k)\mathbf{R}_0(t; k) = \Lambda(t; k)\mathbf{R}_0(t; k)\sigma_3$  holds. Since  $\Lambda(t; k) \neq 0$  for  $k \in \mathcal{Q}_\delta^{\text{II}}$  and  $t > 0$  by Lemma 1, the eigenvalues are distinct and therefore  $\mathbf{R}_0(t; k)$  is invertible as long as its columns are both nonzero. But  $\Re\{-(2k - U(t))H(t)\} = -(2\Re\{k\} - U(t))H(t) \geq 2U(t)H(t)$  as  $\Re\{k\} \leq 0$ . It follows from Assumption 1 that the off-diagonal entries of  $\mathbf{R}_0(t; k)$  are nonzero for all  $t > 0$  and therefore  $\mathbf{R}_0(t; k)$  has nonzero columns for all  $k \in \mathcal{Q}_\delta^{\text{II}}$  and for all  $t > 0$ , so for such  $k$  and  $t$  we have  $\det(\mathbf{R}_0(t; k)) \neq 0$ .

Moreover,  $\det(\mathbf{R}_0(t; k))$  is bounded away from zero on  $\mathbb{R}_+ \times \mathcal{Q}_\delta^{\text{II}}$ . Indeed, one can show that  $\Lambda(t; k) = -2ik^2 + \mathcal{O}(k)$  as  $k \rightarrow \infty$  uniformly for  $\pi/2 \leq \arg(k) \leq \pi$  and  $t > 0$ , and it follows that  $\det(\mathbf{R}_0(t; k)) = 16k^4 + \mathcal{O}(k^3)$  also holds as  $k \rightarrow \infty$  with the same uniform nature of the error term. On the other hand if  $|k| \leq L$  for some (large)  $L > 0$ , then  $k$  lies in a compact subset of  $\mathcal{Q}_\delta^{\text{II}}$ , so since  $\det(\mathbf{R}_0(t; k))$  is a continuous nonzero function of its arguments it follows that it is uniformly bounded away from zero for all such  $k$  and  $t$  lying in any compact subset of  $(0, \infty)$ . It remains to analyze  $\det(\mathbf{R}_0(t; k))$  in the limits  $t \downarrow 0$  and  $t \uparrow +\infty$ , in which  $H(t) \rightarrow 0$  by Assumption 1. Since  $\Lambda(t; k) \rightarrow -i\frac{1}{2}(4k^2 - U_0^2)$  as  $t \downarrow 0$  and  $\Lambda(t; k) \rightarrow -i\frac{1}{2}(4k^2 - U_\infty^2)$  as  $t \uparrow +\infty$ , we have  $\mathbf{R}_0(t; k) \rightarrow i(4k^2 - U_0^2)\sigma_3$  as  $t \downarrow 0$  and  $\mathbf{R}_0(t; k) \rightarrow i(4k^2 - U_\infty^2)\sigma_3$  as  $t \uparrow +\infty$ . Both limiting values are nonzero for  $k \in \mathcal{Q}_\delta^{\text{II}}$ , and the convergence is uniform for  $|k| \leq L$ , so the argument is complete.

Let  $N(t; k)$  be the continuous function defined by  $N(t; k)^2 = \det(\mathbf{R}_0(t; k))$  and the asymptotic condition (which selects an unambiguous branch of the square root) that  $N(t; k) \rightarrow 4k^2 - U_\infty^2$  as  $t \rightarrow +\infty$ . By a homotopy argument taking the function  $H(t)$  to zero by scaling, it is easy to see that  $N(0; k) = 4k^2 - U_0^2$ . Then  $N(t; k)^{-1}$  is a uniformly bounded nonvanishing function for  $t > 0$  and  $k \in \mathcal{Q}_\delta^{\text{II}}$  that satisfies  $N(t; k)^{-1} = \mathcal{O}(k^{-2})$  as  $k \rightarrow \infty$  in  $\mathcal{Q}_\delta^{\text{II}}$  uniformly for

$t > 0$ . Let an eigenvector matrix  $\mathbf{R}(t; k)$  be defined for  $t > 0$  and  $k \in \mathcal{Q}_\delta^\Pi$  by

$$\mathbf{R}(t; k) := \frac{1}{N(t; k)} \mathbf{R}_0(t; k). \tag{A.6}$$

Of course  $\det(\mathbf{R}(t; k)) \equiv 1$ . Also, it is easy to see that

$$\lim_{t \downarrow 0} \mathbf{R}(t; k) = \lim_{t \uparrow +\infty} \mathbf{R}(t; k) = i\sigma_3. \tag{A.7}$$

Let the first column  $\mathbf{f}(t; k)$  of the matrix  $\mathbf{F}(t; k)$  satisfying (45) and (47) be written in the form

$$\mathbf{f}(t; k) = \mathbf{R}(t; k)\mathbf{g}(t; k) \tag{A.8}$$

for some new unknown vector function  $\mathbf{g}(t; k)$ . It is clear from (47) and (A.7) that

$$\lim_{t \rightarrow +\infty} \mathbf{g}(t; k) e^{i(4k^2 t + S(t))/(2\epsilon)} = -i \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{A.9}$$

and from (48)–(49) and (A.7) that  $A_0(k^*)^*$  and  $B_0(k^*)^*$  can be expressed in terms of  $\mathbf{g}(t; k)$  by

$$A_0(k^*)^* = ie^{iS(0)/(2\epsilon)} g_1(0; k) \quad \text{and} \quad B_0(k^*)^* = -ie^{-iS(0)/(2\epsilon)} g_2(0; k). \tag{A.10}$$

Moreover, a direct calculation using (45) shows that  $\mathbf{g}(t; k)$  satisfies the differential equation

$$\begin{aligned} \epsilon \frac{d\mathbf{g}}{dt}(t; k) &= \Lambda(t; k)\sigma_3 \mathbf{g}(t; k) - \epsilon \mathbf{R}(t; k)^{-1} \frac{d\mathbf{R}}{dt}(t; k) \mathbf{g}(t; k) \\ &= \Lambda(t; k)\sigma_3 \mathbf{g}(t; k) - \epsilon \rho(t; k)\sigma_2 \mathbf{g}(t; k), \end{aligned} \tag{A.11}$$

where

$$\rho(t; k) := \frac{1}{iN(t; k)^2} \left[ R_{0,11}(t; k) \frac{dR_{0,21}}{dt}(t; k) - R_{0,21}(t; k) \frac{dR_{0,11}}{dt}(t; k) \right]. \tag{A.12}$$

Assumption 1 implies that  $H'$  and  $U'$  are absolutely integrable on  $\mathbb{R}_+$ , and this in turn implies that  $\rho(\cdot; k) \in L^1(\mathbb{R}_+)$  for each  $k \in \mathcal{Q}_\delta^\Pi$ . Moreover there is a constant  $C_\rho$  depending only on the functions  $H$  and  $U$  satisfying Assumption 1 such that

$$\|\rho(\cdot; k)\|_1 := \int_0^{+\infty} |\rho(t; k)| dt \leq \frac{C_\rho}{1 + |k|}, \quad k \in \mathcal{Q}_\delta^\Pi. \tag{A.13}$$

Making the rescaling

$$\mathbf{g}(t; k) = -ie^{-iS(0)/(2\epsilon)} e^{iM(k)/(2\epsilon)} \mathbf{h}(t; k) \exp\left(\frac{1}{\epsilon} \int_0^t \Lambda(s; k) ds\right), \tag{A.14}$$

where

$$M(k) := \int_0^\infty [2i\Lambda(s; k) - 4k^2 + U(s)^2 + 2H(s)^2] ds \tag{A.15}$$

results in the system

$$\epsilon \frac{d\mathbf{h}}{dt}(t; k) = \begin{bmatrix} 0 & 0 \\ 0 & -2\Lambda(t; k) \end{bmatrix} \mathbf{h}(t; k) - \epsilon \rho(t; k) \sigma_2 \mathbf{h}(t; k), \tag{A.16}$$

and the boundary condition

$$\lim_{t \rightarrow +\infty} \mathbf{h}(t; k) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{A.17}$$

Note that  $M(k)$  is well defined due to the asymptotic behavior of  $\Lambda(t; k)$  for large positive  $t$  and that the condition (A.17) is a consequence of (41). Also, from (A.10) we have that

$$A_0(k^*)^* = e^{iM(k)/(2\epsilon)} h_1(0; k) \quad \text{and} \quad B_0(k^*)^* = -e^{-iS(0)/\epsilon} e^{iM(k)/(2\epsilon)} h_2(0; k). \tag{A.18}$$

Now we solve (A.16) and (A.17). Writing  $h_1(t; k) = 1 + y(t; k)$  and  $h_2(t; k) = z(t; k)$  and introducing an appropriate integrating factor, (A.16) takes the form

$$\frac{dy}{dt}(t; k) = i\rho(t; k)z(t; k)$$

$$\frac{d}{dt} [e^{2\varphi(0,t;k)/\epsilon} z(t; k)] = -ie^{2\varphi(0,t;k)/\epsilon} \rho(t; k)y(t; k) - ie^{2\varphi(0,t;k)/\epsilon} \rho(t; k), \tag{A.19}$$

where  $\varphi$  is defined as

$$\varphi(t_0, t_1; k) := \int_{t_0}^{t_1} \Lambda(s; k) ds. \tag{A.20}$$

Note that since  $\Re\{\Lambda(s; k)\} \leq 0$  by Lemma 14, the factor  $e^{2\varphi(0,t;k)/\epsilon}$  remains bounded as  $t \rightarrow +\infty$ , so we may build in the boundary conditions (A.17) on  $\mathbf{h}(t; k)$  by integrating equations (A.19) from  $t$  to  $+\infty$ :

$$y(t; k) = -i \int_t^{+\infty} \rho(t_1; k)z(t_1; k) dt_1$$

$$z(t; k) = i \int_t^{+\infty} e^{2\varphi(t,t_1;k)/\epsilon} \rho(t_1; k)y(t_1; k) dt_1 + F^\epsilon(t; k), \quad \text{where}$$

$$F^\epsilon(t; k) := i \int_t^{+\infty} e^{2\varphi(t,t_1;k)/\epsilon} \rho(t_1; k) dt_1. \tag{A.21}$$

This system of integral equations is equivalent to the differential equations (A.16) and boundary conditions (A.17).

To solve (A.21), substitute the first equation into the second and exchange the order of integration to get a closed equation for  $z(t; k)$ :

$$z(t; k) = \int_t^{+\infty} K^\epsilon(t, t_1; k)z(t_1; k) dt_1 + F^\epsilon(t; k), \tag{A.22}$$

where the kernel is given by

$$K^\epsilon(t, t_1; k) := \rho(t_1; k) \int_t^{t_1} e^{2\varphi(t, t_2; k)/\epsilon} \rho(t_2; k) dt_2. \tag{A.23}$$

Again, since  $\Re\{\Lambda(s; k)\} \leq 0$ , we have the  $\epsilon$ -independent estimate

$$|K^\epsilon(t, t_1; k)| \leq |\rho(t_1; k)| \int_t^{t_1} |\rho(t_2; k)| dt_2 \leq \|\rho(\cdot; k)\|_1 |\rho(t_1; k)|, \quad t \leq t_1. \tag{A.24}$$

We first seek  $z(\cdot; k)$  in the Banach space  $X$  of continuous functions for  $t \geq 0$  decaying to zero as  $t \rightarrow +\infty$ , equipped with the supremum norm. Note that  $F^\epsilon(\cdot; k) \in X$  because  $\rho(\cdot; k) \in L^1(\mathbb{R}_+)$ . Defining a sequence of iterates by setting  $z_0(t; k) \equiv 0$  and then

$$z_{n+1}(t; k) := \int_t^{+\infty} K^\epsilon(t, t_1; k)z_n(t_1; k) dt_1 + F^\epsilon(t; k), \quad n \geq 0, \tag{A.25}$$

one easily sees that  $z_n(t; k)$  takes the form of a sum of multiple integrals:

$$z_n(t; k) = \sum_{m=0}^n S_m(t; k), \tag{A.26}$$

where  $S_0(t; k) := F^\epsilon(t; k)$  and for  $m \geq 1$ ,

$$S_m(t; k) := \int_t^{+\infty} \int_{t_1}^{+\infty} \cdots \int_{t_{m-1}}^{+\infty} K^\epsilon(t, t_1; k)K^\epsilon(t_1, t_2; k) \cdots K^\epsilon(t_{m-1}, t_m; k)F^\epsilon(t_m; k) dt_m dt_{m-1} \cdots dt_1. \tag{A.27}$$

Obviously, for all  $m \geq 0$  we have

$$\begin{aligned} |S_m(t; k)| &\leq \|\rho(\cdot; k)\|_1^m \|F^\epsilon\|_\infty \int_t^{+\infty} \int_{t_1}^{+\infty} \cdots \int_{t_{m-1}}^{+\infty} |\rho(t_1; k)||\rho(t_2; k)| \cdots |\rho(t_m; k)| dt_m dt_{m-1} \cdots dt_1 \\ &= \frac{\|\rho(\cdot; k)\|_1^m \|F^\epsilon\|_\infty}{m!} \left[ \int_t^{+\infty} |\rho(t_1; k)| dt_1 \right]^m \\ &\leq \frac{\|\rho(\cdot; k)\|_1^{2m}}{m!} \|F^\epsilon\|_\infty, \end{aligned} \tag{A.28}$$

where  $\|F\|_\infty$  denotes the norm in  $X$ :

$$\|F\|_\infty := \sup_{t \geq 0} |F(t)|. \tag{A.29}$$



It follows by comparison with the exponential series that the infinite series

$$z(t; k) = \lim_{n \rightarrow \infty} z_n(t; k) = \sum_{m=0}^{\infty} S_m(t; k) \tag{A.30}$$

converges uniformly on  $\mathbb{R}_+$  and hence as each partial sum vanishes as  $t \rightarrow +\infty$  it converges in  $X$ . Moreover,

$$\|z(\cdot; k)\|_{\infty} \leq \sum_{m=0}^{\infty} \frac{\|\rho(\cdot; k)\|_1^{2m}}{m!} \|F^\epsilon\|_{\infty} = e^{\|\rho(\cdot; k)\|_1^2} \|F^\epsilon\|_{\infty}. \tag{A.31}$$

It then follows from the first equation of (A.21) that  $y(\cdot; k)$  also lies in  $X$ , and that

$$\|y(\cdot; k)\|_{\infty} \leq \|\rho(\cdot; k)\|_1 \|z(\cdot; k)\|_{\infty} \leq \|\rho(\cdot; k)\|_1 e^{\|\rho(\cdot; k)\|_1^2} \|F^\epsilon\|_{\infty}. \tag{A.32}$$

Once it is known that both  $y$  and  $z$  lie in  $X$ , it follows directly from the integral equations (A.21) that both functions are continuously differentiable, so  $h_1(t; k) := 1 + y(t; k)$  and  $h_2(t; k) := z(t; k)$  constitute the unique classical solution of the differential equations (A.16) subject to the boundary conditions (A.17).

It remains to estimate  $\|F^\epsilon\|_{\infty}$ . The part of Assumption 1 that describes the local behavior of  $U$  and  $H$  near  $t = 0$  implies that  $\rho(t; k)$  is bounded by  $t^{-1/2}$  near  $t = 0$ . This integrable singularity determines the rate of decay of  $F^\epsilon(t; k)$  for small  $t$ . While integration by parts shows that  $F^\epsilon(t; k) = \mathcal{O}(\epsilon)$  for  $t$  bounded away from zero, letting  $t \downarrow 0$  results in a slower uniform rate of decay, namely  $\mathcal{O}(\epsilon^{1/2})$ . Therefore,  $\|F^\epsilon\|_{\infty} = \mathcal{O}(\epsilon^{1/2})$  under the conditions on  $U$  and  $H$  in Assumption 1. Taking into account that  $\Lambda$  scales as  $|k|^2$  for large  $k$ , while  $\rho$  scales as  $|k|^{-1}$  for large  $k$ , a more precise statement is that for some constant  $C_F > 0$ , the inequality

$$\|F^\epsilon\|_{\infty} \leq \frac{C_F \epsilon^{1/2}}{1 + |k|^3} \tag{A.33}$$

holds for  $k \in \mathcal{Q}_\delta^{\text{II}}$  and all  $\epsilon > 0$  sufficiently small.

Combining (A.13), (A.31), (A.32), and (A.33), we see that (since the exponential factors are bounded),

$$|y(0; k)| = \mathcal{O}\left(\frac{\epsilon^{1/2}}{1 + |k|^4}\right) \quad \text{and} \quad |z(0; k)| = \mathcal{O}\left(\frac{\epsilon^{1/2}}{1 + |k|^3}\right) \tag{A.34}$$

both hold for  $k \in \mathcal{Q}_\delta^{\text{II}}$  and  $\epsilon > 0$ . Then, from  $h_1(0; k) = 1 + y(0; k)$  and  $h_2(0; k) = z(0; k)$  and (A.18) we have

$$\begin{aligned}
 A_0(k^*)^* &= e^{iM(k)/(2\epsilon)} \left[ 1 + \mathcal{O} \left( \frac{\epsilon^{1/2}}{1 + |k|^4} \right) \right] \\
 B_0(k^*)^* &= -e^{-iS(0)/\epsilon} e^{iM(k)/(2\epsilon)} \mathcal{O} \left( \frac{\epsilon^{1/2}}{1 + |k|^3} \right).
 \end{aligned}
 \tag{A.35}$$

In particular, it follows that  $d_0(k) := A_0(k^*)^*$  has no zeros for  $k \in \mathcal{Q}_\delta^{\text{II}}$  if  $\epsilon > 0$  is sufficiently small, and that for  $\epsilon > 0$  fixed any zeros of  $d_0(k)$  in  $\mathcal{Q}_\delta^{\text{II}}$  must lie in a bounded subset (depending only on the functions  $H$  and  $U$  satisfying Assumption 1). Also, it is clear that (51) holds true. This completes the proof of Proposition 1.

### Appendix B: Proof of Proposition 2

#### B.1. Langer transformation to a perturbed Airy equation

Returning to (45) subject to the boundary condition (47), we first consider making a gauge transformation; that is we consider a given invertible matrix  $\mathbf{G}(t; k)$  and transform (45) to a new unknown matrix  $\mathbf{Y}(t; k)$  by the substitution

$$\mathbf{F}(t; k) = \mathbf{G}(t; k)\mathbf{Y}(t; k),
 \tag{B.1}$$

which transforms (45) into the form

$$\epsilon \frac{d\mathbf{Y}}{dt}(t; k) = \left[ \mathbf{G}(t; k)^{-1}\mathbf{B}(t; k)\mathbf{G}(t; k) - \epsilon \mathbf{G}(t; k)^{-1} \frac{d\mathbf{G}}{dt}(t; k) \right] \mathbf{Y}(t; k).
 \tag{B.2}$$

The main idea of the method is to try to choose  $\mathbf{G}(t; k)$  so that the leading term on the right-hand side takes a particularly simple form that can be the basis for a perturbation expansion. Unfortunately, conjugation by  $\mathbf{G}$  does not allow the determinant to be changed; however an additional scalar factor can be introduced by making a change of the independent variable. So let  $z = z(t)$  be a smooth strictly monotone transformation of the independent variable (hence invertible with smooth inverse map  $t = t(z)$ ), and let  $\bar{\mathbf{Y}}(z; k) = \mathbf{Y}(t(z); k)$ . The desired scalar factor then comes from the chain rule:

$$\epsilon \frac{d\bar{\mathbf{Y}}}{dz}(z; k) = \left( \frac{dt}{dz} \mathbf{G}(t; k)^{-1}\mathbf{B}(t; k)\mathbf{G}(t; k) - \epsilon \frac{dt}{dz} \mathbf{G}(t; k)^{-1} \frac{d\mathbf{G}}{dt}(t; k) \right) \bar{\mathbf{Y}}(z; k).
 \tag{B.3}$$

The target form for the leading matrix coefficient on the right-hand side is (for a problem with a single turning point in the  $t$ -interval of interest) the coefficient matrix of the first-order form of the Airy equation; therefore one tries to set

$$\frac{dt}{dz} \mathbf{G}(t; k)^{-1}\mathbf{B}(t; k)\mathbf{G}(t; k) = \mathbf{A}(z) := \begin{bmatrix} 0 & 1 \\ z & 0 \end{bmatrix}.
 \tag{B.4}$$

Taking determinants of both sides of (B.4) leads to a differential equation for  $z = z(t)$ :

$$z \left( \frac{dz}{dt} \right)^2 = -\det(\mathbf{B}(t; k)) = \lambda^2 = (2k - U(t))^2(k - a(t))(b(t) - k). \quad (\text{B.5})$$

For  $k_a < k < k_b$  and assuming that  $k \neq k_0$  and  $k \neq k_\infty$ , the right-hand side has simple roots at the turning points  $t = t_\pm(k)$ . We consider building two different transformations, denoted  $z = z(t) = z_\pm(t; k)$ , mapping neighborhoods of the turning points  $t = t_\pm(k)$  to corresponding intervals of  $z$ . If  $z = z_\pm(t; k)$  is to be a smooth invertible transformation, it is necessary that  $z_\pm(t_\pm(k); k) = 0$ , which fixes the integration constant in each case. The solution to (B.5) is obtained (by separating the variables) as follows:

$$z_\pm(t; k) = \mp \text{sgn}(t - t_\pm(k)) \left| \frac{3}{2} \int_{t_\pm(k)}^t (U(s) - 2k) \sqrt{|(k - a(s))(b(s) - k)|} ds \right|^{2/3}. \quad (\text{B.6})$$

This formula<sup>6</sup> defines  $z_+(t; k)$  in the interval  $t_-(k) < t < +\infty$ , and it defines  $z_-(t; k)$  in the interval  $0 < t < t_+(k)$ . Moreover,  $z_-(t; k)$  is monotone increasing, while  $z_+(t; k)$  is monotone decreasing on their respective domains. It is obvious that within their intervals of definition,  $z_\pm(t; k)$  have exactly one more continuous derivative than do  $H(\cdot)$  and  $U(\cdot)$ , as long as  $t$  is bounded away from the corresponding turning point  $t_\pm(k)$ . However, a local analysis of  $z_\pm(t; k)$  for  $t$  near  $t = t_\pm(k)$  shows that  $z_\pm(t; k)$  has only the same number of continuous derivatives as do  $H(\cdot)$  and  $U(\cdot)$  at the turning point. Given that  $H$  and  $U$  are analytic according to Assumption 1, we conclude that  $z_\pm(t; k)$  are real analytic monotone functions in their respective domains of definition. Note that even though (B.6) appears to allow  $z_+(t; k)$  to be defined also for  $t \leq t_-(k)$  and for  $z_-(t; k)$  to be defined also for  $t \geq t_+(k)$ , there will be an essential loss of smoothness of  $z_\pm(t; k)$  at the ‘‘other’’ turning point  $t_\mp(k)$ , so in fact our approach will be to cover the positive  $t$ -axis  $t > 0$  with the two overlapping intervals  $(0, t_+(k))$  and  $(t_-(k), +\infty)$  and hence use two different changes of independent variable to obtain the desired smoothness. (Moreover, it is easy to check that  $z_-(t; k)$  as defined by (B.6) fails to satisfy (B.5) for  $t > t_+(k)$  and that  $z_+(t; k)$  as defined by (B.6) fails to satisfy (B.5) for  $t < t_-(k)$ .) Finally, we record here that the image of  $(t_-(k), +\infty)$  under  $z_+(\cdot; k)$  is the interval  $(-\infty, z_+(t_-(k); k))$  where  $z_+(t_-(k); k) > 0$  while the image of  $(0, t_+(k))$  under  $z_-(\cdot; k)$  is the interval  $(z_-(0; k), z_-(t_+(k); k))$  where  $z_-(0; k) < 0 < z_-(t_+(k); k)$ . Also,  $z_+(t; k) \sim -t^{2/3}$  as  $t \rightarrow +\infty$ .

---

<sup>6</sup>The nonlinear mapping  $t \mapsto z_\pm(t; k)$  of the independent variable, along with the linear gauge transformation (B.1) and its higher-order correction (B.10), constitute the Langer transformation of the system (45) in a neighborhood of the turning point  $t = t_\pm(k)$ .

With  $z_{\pm}(t; k)$  determined by (B.6), we can solve (B.4) for the gauge transformation matrix  $\mathbf{G}(t; k)$ . Since  $\mathbf{B}(t; k)^2 = \lambda^2 \mathbb{I} = -\det(\mathbf{B}(t; k))\mathbb{I}$ , we obtain the general solution in the form

$$\mathbf{G}(t; k) = \begin{bmatrix} B_{11}(t; k)p_1(t; k) + B_{12}(t; k)p_2(t; k) & z'(t; k)p_1(t; k) \\ B_{21}(t; k)p_1(t; k) + B_{22}(t; k)p_2(t; k) & z'(t; k)p_2(t; k) \end{bmatrix}, \quad (\text{B.7})$$

where  $p_1(t; k)$  and  $p_2(t; k)$  are (at this moment) arbitrary functions of  $t$ .

Whether  $t \in (0, t_+(k))$  and  $z = z_-(t; k)$  or whether  $t \in (t_-(k), +\infty)$  and  $z = z_+(t; k)$ , due to (B.6) and (B.7), Equation (B.3) takes the form

$$\epsilon \frac{d\bar{\mathbf{Y}}}{dz}(z; k) = (\mathbf{A}(z) + \epsilon \mathbf{P}(t; k)) \bar{\mathbf{Y}}(z; k), \quad (\text{B.8})$$

where

$$\mathbf{P}(t; k) := -\frac{1}{z'(t; k)} \mathbf{G}(t; k)^{-1} \frac{d\mathbf{G}}{dt}(t; k). \quad (\text{B.9})$$

Of course we think of  $\mathbf{P}(t; k)$  as a function of  $z$  by means of the invertible transformation defined by (B.6).

It turns out that the error term  $\epsilon \mathbf{P}(t; k)$  in the coefficient matrix is too large to be controllable directly when  $\epsilon \ll 1$  due to the corresponding factor of  $\epsilon$  on the left-hand side of (B.8). However, the terms proportional to  $\epsilon$  in the coefficient matrix can be removed by an explicit near-identity transformation. For some  $\epsilon$ -independent matrix  $\mathbf{H} = \mathbf{H}(t; k)$  to be determined, consider the effect of making the substitution

$$\bar{\mathbf{Y}}(z; k) = (\mathbb{I} + \epsilon \mathbf{H}(t; k)) \mathbf{W}(z; k) \quad (\text{B.10})$$

in Equation (B.8). In order to obtain an equivalent differential equation for  $\mathbf{W}(z; k)$ , it will be necessary to invert  $\mathbb{I} + \epsilon \mathbf{H}(t; k)$ , and while this can always be accomplished pointwise for sufficiently small  $\epsilon$  by Neumann series, it is especially convenient if the inversion can be carried out explicitly. Hence, we assume at this point that  $\mathbf{H}(t; k)$  is a nilpotent matrix of the general form

$$\mathbf{H}(t; k) = i\sigma_2 \mathbf{h}(t; k) \mathbf{h}(t; k)^\top = \begin{bmatrix} h_1(t; k)h_2(t; k) & h_2(t; k)^2 \\ -h_1(t; k)^2 & -h_1(t; k)h_2(t; k) \end{bmatrix}, \quad (\text{B.11})$$

where  $\mathbf{h}(t; k)$  is a general vector function of  $t$  and  $k$ . Thus  $\mathbf{H}(t; k)^2 = \mathbf{0}$ , and it follows that the Neumann series for the inverse truncates:  $(\mathbb{I} + \epsilon \mathbf{H}(t; k))^{-1} = \mathbb{I} - \epsilon \mathbf{H}(t; k)$ . Under (B.10) and (B.11), Equation (B.8) becomes

$$\epsilon \frac{d\mathbf{W}}{dz}(z; k) = (\mathbf{A}(z) + \epsilon \mathbf{C}_1(t; k) + \epsilon^2 \mathbf{C}_2(t; k) + \epsilon^3 \mathbf{C}_3(t; k)) \mathbf{W}(z; k), \quad (\text{B.12})$$

where

$$\begin{aligned}
 \mathbf{C}_1(t; k) &:= [\mathbf{A}(z(t; k)), \mathbf{H}(t, k)] + \mathbf{P}(t; k), \\
 \mathbf{C}_2(t; k) &:= [\mathbf{P}(t, k), \mathbf{H}(t; k)] - \mathbf{H}(t; k)\mathbf{A}(z(t; k))\mathbf{H}(t; k) - \frac{1}{z'(t; k)} \frac{d\mathbf{H}}{dt}(t; k), \\
 \mathbf{C}_3(t; k) &:= \frac{1}{z'(t; k)} \mathbf{H}(t; k) \frac{d\mathbf{H}}{dt}(t; k) - \mathbf{H}(t; k)\mathbf{P}(t; k)\mathbf{H}(t; k).
 \end{aligned}
 \tag{B.13}$$

Here,  $[\mathbf{A}, \mathbf{B}] := \mathbf{AB} - \mathbf{BA}$  denotes the matrix commutator.

The question now arises as to how the arbitrary functions  $p_j(t; k)$  and  $h_j(t; k)$  for  $j = 1, 2$  can be chosen to ensure that  $\mathbf{C}_1(t; k) = \mathbf{0}$ . In fact, the condition  $\mathbf{C}_1(t; k) = \mathbf{0}$  is equivalent to the two conditions on the matrix  $\mathbf{P}(t; k)$  (really, on the two functions  $p_j(t; k)$  for  $j = 1, 2$ ):

$$P_{21}(t; k) + z(t; k)P_{12}(t; k) = 0 \quad \text{and} \quad P_{11}(t; k) + P_{22}(t; k) = 0 \tag{B.14}$$

and the two equations relating  $h_j(t; k)$ ,  $j = 1, 2$  to the functions  $p_j(t; k)$ ,  $j = 1, 2$  in the matrix  $\mathbf{P}(t; k)$ :

$$h_1(t; k)h_2(t; k) = \frac{1}{2}P_{12}(t; k) \quad \text{and} \quad h_1(t; k)^2 + z(t; k)h_2(t; k)^2 = -P_{22}(t; k). \tag{B.15}$$

It turns out to be consistent to assume, in addition to (B.14) that

$$P_{12}(t; k) = 0, \tag{B.16}$$

in which case (B.14) implies that also

$$P_{21}(t; k) = 0. \tag{B.17}$$

Then, conditions (B.15) require that either  $h_1(t; k) = 0$  or  $h_2(t; k) = 0$ . Taking

$$h_2(t; k) = 0, \tag{B.18}$$

conditions (B.15) reduce to

$$h_1(t; k)^2 = -P_{22}(t; k) = P_{11}(t; k). \tag{B.19}$$

It follows that the matrix  $\mathbf{H}(t; k)$  can be expressed in terms of  $\mathbf{P}(t; k)$  as

$$\mathbf{H}(t; k) = \begin{bmatrix} 0 & 0 \\ P_{22}(t; k) & 0 \end{bmatrix}, \tag{B.20}$$

while  $\mathbf{P}(t; k)$  has the form

$$\mathbf{P}(t; k) = -P_{22}(t; k)\sigma_3. \tag{B.21}$$

Assuming that  $P_{22}(t; k)$  is differentiable with respect to  $t$ , it follows from (B.20) and (B.21) that

$$\mathbf{C}_2(t; k) = \begin{bmatrix} 0 & 0 \\ Q(z(t; k); k) & 0 \end{bmatrix}, \tag{B.22}$$

where

$$Q(z(t; k); k) := P_{22}(t; k)^2 - \frac{1}{z'(t; k)} \frac{dP_{22}}{dt}(t; k), \tag{B.23}$$

and that

$$\mathbf{C}_3(t; k) = \mathbf{0}. \tag{B.24}$$

Therefore, under the gauge transformation (B.10), the differential equation (B.8) becomes a perturbed Airy equation:

$$\epsilon \frac{d\mathbf{W}}{dz}(z; k) - \mathbf{A}(z)\mathbf{W}(z; k) = \begin{bmatrix} 0 & 0 \\ \epsilon^2 Q(z; k) & 0 \end{bmatrix} \mathbf{W}(z; k). \tag{B.25}$$

Let us consider finding  $p_j(t; k)$ ,  $j = 1, 2$ , so that  $P_{12}(t; k) = 0$ ,  $P_{21}(t; k) = 0$ , and also  $P_{11}(t; k) + P_{22}(t; k) = 0$ . These calculations are based on the formula (B.9) and the representation (B.7). Assuming for the moment that  $\det(\mathbf{G}(t; k)) \neq 0$ , it is easy to see that the condition  $P_{11}(t; k) + P_{22}(t; k) = 0$  implies that in fact  $\det(\mathbf{G}(t; k))$  must be independent of  $t$ ; as the conditions on the elements of  $\mathbf{P}(t; k)$  are all linear in  $p_j(t; k)$ ,  $j = 1, 2$ , we will assume that they are chosen so that  $\det(\mathbf{G}(t; k)) = 1$ . It is then easy to see from (B.9) and (B.7) that the condition  $P_{12}(t; k) = 0$  implies that the ratio  $p_2(t; k)/p_1(t; k)$  is independent of  $t$ ; we therefore write

$$p_2(t; k) = cp_1(t; k), \tag{B.26}$$

where  $c$  is independent of  $t$ . With the further assumption that  $p_1(t; k) \neq 0$ , one checks that because  $U(t) - 2k > 0$  for all  $t > 0$  when  $k \in (k_a, k_b)$ , the condition  $P_{21}(t; k) = 0$  implies that

$$\frac{d}{dt} \left( \frac{2H(t) + ic(2k + U(t))}{-i(2k + U(t)) + 2cH(t)} \right) = 0. \tag{B.27}$$

This equation is obviously solved by choosing either  $c = i$  or  $c = -i$ . Finally, we return to the condition  $\det(\mathbf{G}(t; k)) = 1$ . This condition yields the identity

$$p_1(t; k)^2 = [2z'(t; k)(U(t) - 2k)(H(t) \mp (k + \frac{1}{2}U(t)))]^{-1}, \quad c = \pm i. \tag{B.28}$$

Recalling the definitions of  $a(t)$  and  $b(t)$ , this can be written in the form

$$p_1(t; k)^2 = \begin{cases} [2z'(t; k)(U(t) - 2k)(b(t) - k)]^{-1}, & c = i, \\ [2z'(t; k)(U(t) - 2k)(k - a(t))]^{-1}, & c = -i. \end{cases} \tag{B.29}$$

In each case,  $0 < t < t_+(k)$  and  $t_-(k) < t < +\infty$ , we choose the value of  $c = \pm i$  so that the turning point contained in the corresponding open interval is *not* a root of the factor  $(b(t) - k)$  or  $(k - a(t))$  in the denominator of  $p_1(t; k)^2$ . For example, if  $t_-(k)$  satisfies  $b(t_-(k)) = k$ , then  $(k - a(t))$  will be nonzero for  $t \in (0, t_+(k))$ , so we choose  $c = -i$ . A calculation using the fact

that  $\det(\mathbf{G}(t; k)) = 1$  then shows that

$$P_{22}(t; k) = \begin{cases} -\frac{z''(t; k)}{2z'(t; k)^2} + \frac{1}{2z'(t; k)} \frac{d}{dt} \log((U(t) - 2k)(b(t) - k)), & c = i, \\ -\frac{z''(t; k)}{2z'(t; k)^2} + \frac{1}{2z'(t; k)} \frac{d}{dt} \log((U(t) - 2k)(k - a(t))), & c = -i. \end{cases} \tag{B.30}$$

(It is not necessary to resolve the sign of  $p_1(t; k)$  to determine  $P_{22}(t; k)$  uniquely.) The corresponding function  $Q(z; k)$  is then determined from (B.23) and (B.6).

### B.2. Solution of the perturbed Airy equation

LEMMA 15. *Let  $k_a < k < k_b$  with  $k \neq k_0$  and  $k \neq k_\infty$ . Then there exists a constant  $K = K(k)$  such that for  $z = z_-(t; k) \in (z_-(0; k), z_-(t_+(k); k))$ ,  $|Q(z; k)| \leq K$ , while for  $z = z_+(t; k) \in (-\infty, z_+(t_-(k); k))$ ,  $Q(z; k)$  satisfies an estimate of the form*

$$|Q(z; k)| \leq \frac{K}{1 + z^2}. \tag{B.31}$$

*Proof:* Since  $z'_-(\cdot; k)$  and  $z'_+(\cdot; k)$  are nonzero on  $(0, t_+(k))$  and  $(t_-(k), +\infty)$ , respectively, it follows that  $P_{22}(t; k)$  is in each case an analytic function on the corresponding interval. It then follows from (B.23) that the same is true of  $Q(z(t); k)$ . Since  $z = z_\pm(t; k)$  is in each case a real analytic bijection,  $Q(z; k)$  is smooth on either  $(z_-(0; k), z_-(t_+(k); k))$  or  $(-\infty, z_+(t_-(k); k))$ . Therefore, it only remains to prove that  $Q(z; k) = \mathcal{O}(z^{-2})$  as  $z \rightarrow -\infty$  in the case when  $z = z_+(t; k)$ . But from (B.30) and Assumption 1 it follows that  $|P_{22}(t; k)| \sim t^{-2/3}$  and  $|P'_{22}(t; k)| \sim t^{-5/3}$  as  $t \rightarrow +\infty$ , so the desired decay follows from (B.23) by composition with  $z_+(t; k) \sim -t^{2/3}$  as  $t \rightarrow +\infty$ . ■

We now use the estimate of  $Q$  recorded in Lemma 15 to solve the perturbed Airy equation (B.25) separately in the intervals  $z = z_-(t; k) \in (z_-(0; k), z_-(t_+(k); k))$  and  $z = z_+(t; k) \in (-\infty, z_+(t_-(k); k))$ . The idea in both cases is the same: consider the perturbation proportional to  $\epsilon^2$  on the right-hand side as a forcing term, and convert the differential equation into an integral equation with the help of the fundamental matrix

$$\mathbf{W}_0(z) := \begin{bmatrix} \text{Ai}(\epsilon^{-2/3}z) & \text{Bi}(\epsilon^{-2/3}z) \\ \epsilon^{1/3} \text{Ai}'(\epsilon^{-2/3}z) & \epsilon^{1/3} \text{Bi}'(\epsilon^{-2/3}z) \end{bmatrix} \tag{B.32}$$

of the unforced problem (we refer to [19] for the definitions and properties of the Airy functions  $\text{Ai}(\cdot)$  and  $\text{Bi}(\cdot)$ ). Indeed, setting

$$\mathbf{W}(z; k) = \mathbf{W}_0(z)\mathbf{U}(z; k) \tag{B.33}$$

in (B.25) where  $\mathbf{U}(z; k)$  denotes a new matrix unknown, we obtain the equivalent equation (after canceling a factor of  $\epsilon$ )

$$\frac{d\mathbf{U}}{dz}(z; k) = \epsilon Q(z; k) \mathbf{W}_0(z)^{-1} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \mathbf{W}_0(z) \mathbf{U}(z; k). \tag{B.34}$$

Using the Wronskian identity  $\text{Ai}(\cdot)\text{Bi}'(\cdot) - \text{Bi}(\cdot)\text{Ai}'(\cdot) = 1/\pi$  to invert  $\mathbf{W}_0(z)$ , we may write this system in the form

$$\frac{d\mathbf{U}}{dz}(z; k) = \epsilon^{2/3} \pi Q(z; k) \begin{bmatrix} -\text{Ai}(\epsilon^{-2/3}z)\text{Bi}(\epsilon^{-2/3}z) & -\text{Bi}(\epsilon^{-2/3}z)^2 \\ \text{Ai}(\epsilon^{-2/3}z)^2 & \text{Ai}(\epsilon^{-2/3}z)\text{Bi}(\epsilon^{-2/3}z) \end{bmatrix} \mathbf{U}(z; k). \tag{B.35}$$

Assuming  $\mathbf{U}(z; k)$  is known at some value  $z_0$ , we integrate to obtain the Volterra equation

$$\begin{aligned} \mathbf{U}(z; k) &= \mathbf{U}(z_0; k) + \epsilon^{2/3} \pi \\ &\cdot \int_{z_0}^z Q(\zeta; k) \begin{bmatrix} -\text{Ai}(\epsilon^{-2/3}\zeta)\text{Bi}(\epsilon^{-2/3}\zeta) & -\text{Bi}(\epsilon^{-2/3}\zeta)^2 \\ \text{Ai}(\epsilon^{-2/3}\zeta)^2 & \text{Ai}(\epsilon^{-2/3}\zeta)\text{Bi}(\epsilon^{-2/3}\zeta) \end{bmatrix} \mathbf{U}(\zeta; k) d\zeta. \end{aligned} \tag{B.36}$$

This equation may be analyzed with the help of a weight function  $\omega$  defined by

$$\omega(s) := \begin{cases} 1, & s \leq 0 \\ e^{-4s^{3/2}/3}, & s > 0. \end{cases} \tag{B.37}$$

The weight  $\omega$  is strictly positive, continuous, and nonincreasing on  $\mathbb{R}$ . Also, there is a constant  $C > 0$  such that the three inequalities

$$\begin{aligned} |\text{Ai}(s)\text{Bi}(s)| &\leq \frac{C}{|s|^{1/2}}, \quad |\text{Ai}(s)|^2 \omega(s)^{-1} \leq \frac{C}{|s|^{1/2}}, \quad \text{and} \\ |\text{Bi}(s)|^2 \omega(s) &\leq \frac{C}{|s|^{1/2}} \quad \text{hold for all } s \in \mathbb{R}. \end{aligned} \tag{B.38}$$

We now introduce the weight  $\omega$  into (B.36) in two different ways. First, let

$$\mathbf{U}(z; k) = \begin{bmatrix} \omega(\epsilon^{-2/3}z)^{-1} & 0 \\ 0 & 1 \end{bmatrix} \mathbf{V}_>(z; k) \tag{B.39}$$

define a new unknown  $\mathbf{V}_>(z; k)$ . In terms of  $\mathbf{V}_>(z; k)$ , the Volterra equation (B.36) becomes the equivalent equation

$$(1 - \mathcal{K}_>) \mathbf{V}_>(z; k) = \mathbf{R}_>(z; k), \quad \mathcal{K}_> \mathbf{V}(z) := \int_{z_0}^z \mathbf{K}_>(z, \zeta; k) \mathbf{V}(\zeta) d\zeta \tag{B.40}$$



where the matrix kernel  $\mathbf{K}_>(z, \zeta; k)$  is given by

$$\mathbf{K}_>(z, \zeta; k) := \epsilon^{2/3} \pi Q(\zeta; k) \begin{bmatrix} -\omega(\epsilon^{-2/3}z)\omega(\epsilon^{-2/3}\zeta)^{-1}\text{Ai}(\epsilon^{-2/3}\zeta)\text{Bi}(\epsilon^{-2/3}\zeta) & -\text{Bi}(\epsilon^{-2/3}\zeta)^2\omega(\epsilon^{-2/3}z) \\ \text{Ai}(\epsilon^{-2/3}\zeta)^2\omega(\epsilon^{-2/3}\zeta)^{-1} & \text{Ai}(\epsilon^{-2/3}\zeta)\text{Bi}(\epsilon^{-2/3}\zeta) \end{bmatrix}, \tag{B.41}$$

and where the forcing term is the matrix function

$$\mathbf{R}_>(z; k) := \begin{bmatrix} \omega(\epsilon^{-2/3}z)\omega(\epsilon^{-2/3}z_0)^{-1} & 0 \\ 0 & 1 \end{bmatrix} \mathbf{V}_>(z_0; k). \tag{B.42}$$

We will use this form when  $z > z_0$  in which case  $\omega(\epsilon^{-2/3}z) \leq \omega(\epsilon^{-2/3}\zeta)$  holds for the first row of the kernel. Alternately, let

$$\mathbf{U}(z; k) = \begin{bmatrix} 1 & 0 \\ 0 & \omega(\epsilon^{-2/3}z) \end{bmatrix} \mathbf{V}_<(z; k) \tag{B.43}$$

define a new unknown  $\mathbf{V}_<(z; k)$ . In terms of  $\mathbf{V}_<(z; k)$ , the Volterra equation (B.36) becomes

$$(1 - \mathcal{K}_<)\mathbf{V}(z; k) = \mathbf{R}_<(z; k), \quad \mathcal{K}_<\mathbf{V}(z) := \int_z^{z_0} \mathbf{K}_<(z, \zeta; k)\mathbf{V}(\zeta) d\zeta \tag{B.44}$$

where the matrix kernel  $\mathbf{K}_<(z, \zeta; k)$  is given by

$$\mathbf{K}_<(z, \zeta; k) := -\epsilon^{2/3} \pi Q(\zeta; k) \begin{bmatrix} -\text{Ai}(\epsilon^{-2/3}\zeta)\text{Bi}(\epsilon^{-2/3}\zeta) & -\text{Bi}(\epsilon^{-2/3}\zeta)^2\omega(\epsilon^{-2/3}\zeta) \\ \text{Ai}(\epsilon^{-2/3}\zeta)^2\omega(\epsilon^{-2/3}z)^{-1} & \omega(\epsilon^{-2/3}z)^{-1}\omega(\epsilon^{-2/3}\zeta)\text{Ai}(\epsilon^{-2/3}\zeta)\text{Bi}(\epsilon^{-2/3}\zeta) \end{bmatrix}, \tag{B.45}$$

and the matrix-valued forcing term is

$$\mathbf{R}_<(z; k) := \begin{bmatrix} 1 & 0 \\ 0 & \omega(\epsilon^{-2/3}z_0)\omega(\epsilon^{-2/3}z)^{-1} \end{bmatrix} \mathbf{V}_<(z_0; k). \tag{B.46}$$

We will use this form when  $z < z_0$  in which case  $\omega(\epsilon^{-2/3}z)^{-1} \leq \omega(\epsilon^{-2/3}\zeta)^{-1}$  holds for the second row of the kernel. Let  $t_0$  be a fixed number in the open interval  $(t_-(k), t_+(k))$ . We first use (B.39)–(B.41) along with the mapping  $t \mapsto z = z_+(t; k)$  to analyze  $\mathbf{F}(t; k)$  for  $t \in [t_0, +\infty)$ . Then, we use (B.43)–(B.45) along with the mapping  $t \mapsto z = z_-(t; k)$  to analyze  $\mathbf{F}(t; k)$  for  $t \in (0, t_0]$ .

**Analysis of  $\mathbf{F}(t; k)$  for  $t_0 \leq t < +\infty$ .**

Under the real-analytic bijection  $t \mapsto z = z_+(t; k)$  given by (B.6), the interval  $t_0 \leq t < +\infty$  is identified with the interval  $-\infty < z \leq z_+(t_0; k)$  where  $z_+(t_0; k) > 0$ , and the mapping reverses orientation, i.e.,  $z'_+(t; k) < 0$ . We begin

by resolving the only remaining indeterminacy in the Langer transformation by first noting that in the formula (B.29) we must take  $c = \text{isgn}(k^2 - k_\infty^2)$  (because  $\mathfrak{a}(t_+(k)) = k$  for  $k_\alpha < k < k_\infty < 0$  while  $\mathfrak{b}(t_+(k)) = k$  for  $k_\infty < k < k_\flat < 0$ ). Taking into account that  $z'_+(t; k) < 0$ , we then choose the sign of the square root to obtain  $p_1(t; k)$  as

$$p_1(t; k)^{-1} = \begin{cases} i\sqrt{-2z'_+(t; k)(U(t) - 2k)(\mathfrak{b}(t) - k)}, & k_\alpha < k < k_\infty \\ i\sqrt{-2z'_+(t; k)(U(t) - 2k)(k - \mathfrak{a}(t))}, & k_\infty < k < k_\flat. \end{cases} \tag{B.47}$$

In both cases, the positive square root is meant, and  $p_1(t; k)$  is a purely imaginary nonvanishing analytic function of  $t$  in the interval  $t_0 \leq t < +\infty$ . The definition (B.47) unambiguously determines the matrices  $\mathbf{G}(t; k) = \mathbf{G}_>(t; k)$  and  $\mathbf{H}(t; k) = \mathbf{H}_>(t; k)$ . Then, combining the transformations (B.1), (B.10), and (B.33) with (B.39), we obtain the exact relation linking  $\mathbf{F}(t; k)$  and  $\mathbf{V}_>(z; k)$  for  $z = z_+(t; k)$ :

$$\mathbf{V}_>(z; k) = \begin{bmatrix} \omega(\epsilon^{-2/3}z) & 0 \\ 0 & 1 \end{bmatrix} \mathbf{W}_0(z)^{-1} (\mathbb{I} + \epsilon \mathbf{H}_>(t; k))^{-1} \mathbf{G}_>(t; k)^{-1} \mathbf{F}(t; k), \tag{B.48}$$

$t \geq t_0.$

Since  $\omega(\epsilon^{-2/3}z_+(t; k)) = 1$  for  $t > t_+(k)$  (which implies  $z_+(t; k) < 0$ ), from the asymptotic normalization condition (47) on  $\mathbf{F}(t; k)$ , we expect the following limit to exist for each  $\epsilon > 0$ :

$$\mathbf{V}_>(-\infty; k) := \lim_{t \rightarrow +\infty} \mathbf{W}_0(z_+(t; k))^{-1} (\mathbb{I} + \epsilon \mathbf{H}_>(t; k))^{-1} \mathbf{G}_>(t; k)^{-1} e^{-i(4k^2t + S(t))\sigma_3/(2\epsilon)}. \tag{B.49}$$

This limit does indeed exist, and it may be computed with the help of (i) the estimate  $\mathbf{H}_>(t; k) = \mathcal{O}(t^{-2/3})$ , (ii) the asymptotic relations

$$z_+(t; k) = -3^{2/3}|k^2 - k_\infty^2|^{2/3}t^{2/3}(1 + \mathcal{O}(t^{-1})), \quad t \rightarrow +\infty, \tag{B.50}$$

$$\frac{2}{3}(-z_+(t; k))^{3/2} = 2|k^2 - k_\infty^2|t + \ell(k) + o(1), \quad t \rightarrow +\infty, \tag{B.51}$$

where  $\ell(k)$  is defined by (54) and

$$z'_+(t; k) = -\frac{2}{3^{1/3}}|k^2 - k_\infty^2|^{2/3}t^{-1/3}(1 + \mathcal{O}(t^{-1})), \quad t \rightarrow +\infty, \tag{B.52}$$

and (iii) known asymptotic formulae for Airy functions and their derivatives for large negative  $z$ . The result is:

$$\mathbf{V}_>(-\infty; k) = i\epsilon^{-1/6} \sqrt{\frac{\pi}{2}} e^{i\pi \operatorname{sgn}(k^2 - k_\infty^2)\sigma_3/4} \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} e^{i(\ell(k)\operatorname{sgn}(k^2 - k_\infty^2) - \frac{1}{2}S_\infty)\sigma_3/\epsilon}, \tag{B.53}$$

where  $S_\infty$  is defined by (55).

We now consider the Volterra integral Equation (B.40) with  $z_0 = -\infty$ , on the space of matrix-valued functions defined on  $-\infty < z < z_+(t_0; k)$  equipped with the supremum norm (and based on some matrix norm). It is a simple consequence of Lemma 15, the inequality  $\omega(\epsilon^{-2/3}z) \leq \omega(\epsilon^{-2/3}\zeta)$  holding for  $z \geq \zeta$  and the estimates (B.38) that the Volterra integral operator  $\mathcal{K}_>$  with matrix kernel  $\mathbf{K}_>(z, \zeta; k)$  is bounded on this space with operator norm  $\mathcal{O}(\epsilon)$ , because

$$\int_{-\infty}^{z_+(t_0; k)} \frac{d\zeta}{|\zeta|^{1/2}(1 + \zeta^2)} < \infty. \tag{B.54}$$

It follows easily that  $1 - \mathcal{K}_>$  is invertible for sufficiently small  $\epsilon$ , and that the operator norm of  $(1 - \mathcal{K}_>)^{-1} - 1$  is  $\mathcal{O}(\epsilon)$ . Since  $\omega(\epsilon^{-2/3}z_0) = 1$  for  $z_0 = -\infty$ , it is easy to see that  $\mathbf{R}_>(z; k)$  defined by (B.42) is a bounded continuous function on  $(-\infty, z_+(t_0; k))$ . Therefore we learn that  $(\|\cdot\|)$  is a matrix norm

$$\sup_{-\infty < z < z_+(t_0; k)} \|\epsilon^{1/6}(\mathbf{V}_>(z; k) - \mathbf{R}_>(z; k))\| = \mathcal{O}(\epsilon), \quad \epsilon \rightarrow 0, \tag{B.55}$$

because  $\epsilon^{1/6}\mathbf{R}_>(-\infty; k)$  is bounded independently of  $\epsilon$  according to (B.53) and the fact that  $0 < \omega(\epsilon^{-2/3}z) \leq 1$ . It follows that we may solve (B.48) for  $\mathbf{F}(t; k)$ , set  $t = t_0$  and take into account that  $z_+(t_0; k) > 0$  to obtain an asymptotic formula for  $\mathbf{F}(t_0; k)$ ; here we need to use the fact that  $\omega(\epsilon^{-2/3}z_+(t_0; k)) = e^{-4z_+(t_0; k)^{3/2}/(3\epsilon)}$  along with asymptotic expansions of Airy functions for large positive arguments. The key point is that

$$\mathbf{W}_0(z_+(t_0; k)) \begin{bmatrix} \omega(\epsilon^{-2/3}z_+(t_0; k))^{-1} & 0 \\ 0 & 1 \end{bmatrix} = \frac{\epsilon^{1/6}}{\sqrt{\pi}} e^{2z_+(t_0; k)^{3/2}/(3\epsilon)} z_+(t_0; k)^{-\sigma_3/4} \cdot \left( \begin{bmatrix} \frac{1}{2} & 1 \\ -\frac{1}{2} & 1 \end{bmatrix} + \mathcal{O}(\epsilon) \right), \quad \epsilon \rightarrow 0. \tag{B.56}$$

Combining this with (B.48), (B.53), and (B.55) gives

$$\mathbf{F}(t_0; k) = e^{2z_+(t_0; k)^{3/2}/(3\epsilon)} (\mathbf{Z}_>(k) + \mathcal{O}(\epsilon)) e^{i(\ell(k)\operatorname{sgn}(k^2 - k_\infty^2) - \frac{1}{2}S_\infty)\sigma_3/\epsilon}, \quad \epsilon \rightarrow 0, \tag{B.57}$$

where

$$\mathbf{Z}_>(k) := -\frac{i}{\sqrt{2}} e^{-i\pi \operatorname{sgn}(k^2 - k_\infty^2)/4} \mathbf{G}_>(t_0; k) z_+(t_0; k)^{-\sigma_3/4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{B.58}$$

**Analysis of  $\mathbf{F}(t; k)$  for  $0 < t \leq t_0$ .**

Now, we consider the interval  $0 < t \leq t_0$ , which is mapped by the real-analytic bijection  $t \mapsto z_-(t; k)$  to the interval  $z_-(0; k) < z \leq z_-(t_0; k)$  with orientation preserved, i.e.,  $z'_-(t; k) > 0$ . To fully determine the Langer transformation in this interval, we need to take  $c = \text{isgn}(k^2 - k_0^2)$  (because  $\mathfrak{a}(t_-(k)) = k$  for  $k_a < k < k_0$  while  $\mathfrak{b}(t_-(k)) = k$  for  $k_0 < k < k_b$ ). Choosing a sign for the square root of  $p_1(t; k)^2$  we then obtain

$$p_1(t; k)^{-1} = \begin{cases} \sqrt{2z'_-(t; k)(U(t) - 2k)(\mathfrak{b}(t) - k)}, & k_a < k < k_0 \\ \sqrt{2z'_-(t; k)(U(t) - 2k)(k - \mathfrak{a}(t))}, & k_0 < k < k_b. \end{cases} \tag{B.59}$$

Therefore,  $p_1(t; k)$  is a strictly positive nonvanishing analytic function of  $t$  for  $0 < t \leq t_0$ . This choice determines the matrices  $\mathbf{G}(t; k)$  and  $\mathbf{H}(t; k)$ , here denoted  $\mathbf{G}_<(t; k)$  and  $\mathbf{H}_<(t; k)$ . Then, according to (B.1), (B.10), (B.33), and (B.43), the exact relation between  $\mathbf{F}(t; k)$  and  $\mathbf{V}_<(z; k)$  for  $z = z_-(t; k)$  is

$$\mathbf{V}_<(z; k) = \begin{bmatrix} 1 & 0 \\ 0 & \omega(\epsilon^{-2/3}z)^{-1} \end{bmatrix} \mathbf{W}_0(z)^{-1} (\mathbb{I} + \epsilon \mathbf{H}_<(t; k))^{-1} \mathbf{G}_<(t; k)^{-1} \mathbf{F}(t; k), \tag{B.60}$$

$0 < t \leq t_0$ .

If we consider in particular  $t = t_0$ , then  $\mathbf{F}(t_0; k)$  is given by (B.57), so using the fact that  $z_-(t_0; k) > 0$  to simplify the weight  $\omega$  and using asymptotic formula for Airy functions of large positive arguments, we obtain the following:

$$\mathbf{V}_<(z_-(t_0; k); k) = \epsilon^{-1/6} e^{\frac{2}{3}(z_-(t_0; k)^{3/2} + z_+(t_0; k)^{3/2})/\epsilon} \cdot (\mathbf{Z}_<(t_0; k) + \mathcal{O}(\epsilon)) e^{i(\ell(k) \text{sgn}(k^2 - k_\infty^2) - \frac{1}{2} S_\infty) \sigma_3 / \epsilon}, \tag{B.61}$$

where

$$\mathbf{Z}_<(t_0; k) := -i \sqrt{\frac{\pi}{2}} e^{-i\pi \text{sgn}(k^2 - k_\infty^2)/4} \begin{bmatrix} 1 & -1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} z_-(t_0; k)^{\sigma_3/4} \cdot \mathbf{G}_<(t_0; k)^{-1} \mathbf{G}_>(t_0; k) z_+(t_0; k)^{-\sigma_3/4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{B.62}$$

We claim that the explicit terms in the formula (B.61) are independent of  $t_0 \in (t_-(k), t_+(k))$ . Indeed, we have the following results.

LEMMA 16. *Suppose that  $k \in (k_a, k_b)$  with  $k \neq k_0$  and  $k \neq k_\infty$ , and that  $t_-(k) < t_0 < t_+(k)$ . Then:*

$$\frac{2}{3} z_-(t_0; k)^{3/2} + \frac{2}{3} z_+(t_0; k)^{3/2} = \tau(k), \tag{B.63}$$

where  $\tau : (k_a, k_b) \rightarrow \mathbb{R}_+$  is defined by (56), and

$$\mathbf{Z}_<(t_0; k) = -\sqrt{2\pi} e^{-i\pi \text{sgn}(k^2 - k_0^2)/4} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}. \tag{B.64}$$

In particular,  $\mathbf{Z}_<$  is independent of  $t_0$  and piecewise constant in  $k \in (k_a, k_b)$ .

*Proof:* The proof of (B.63) follows directly from (B.6). To prove (B.64) one first computes the product  $\mathbf{G}_>(t_0; k)z_+(t_0; k)^{-\sigma_3/4}$  from (B.7) using  $z(t; k) = z_+(t; k)$  and (B.47) along with  $p_2 = cp_1$  with  $c = \text{isgn}(k^2 - k_\infty^2)$  and then one computes  $\mathbf{G}_<(t_0; k)z_-(t_0; k)^{-\sigma_3/4}$  from (B.7) using  $z(t; k) = z_-(t; k)$  and (B.59) along with  $p_2 = cp_1$  with  $c = \text{isgn}(k^2 - k_0^2)$ . The resulting formulae may be simplified using the positive one-fourth root of (B.5). The result then follows by a direct calculation.  $\blacksquare$

Henceforth, we will write  $\mathbf{Z}_<(k)$  for  $\mathbf{Z}_<(t_0; k)$  in light of this calculation. We now may consider the Volterra integral equation (B.44) with  $z_0 = z_-(t_0; k)$  and  $z < z_0$ . Fixing a matrix norm  $\|\cdot\|$  and defining the corresponding supremum norm over the interval  $z_-(0; k) < z < z_-(t_0; k)$ , it follows from Lemma 15, the inequality  $\omega(\epsilon^{-2/3}z)^{-1} \leq \omega(\epsilon^{-2/3}\zeta)^{-1}$  holding for  $z \leq \zeta$  and the estimates (B.38) that the Volterra integral operator  $\mathcal{K}_<$  with matrix kernel  $\mathbf{K}_<(z, \zeta; k)$  is bounded with operator norm  $\mathcal{O}(\epsilon)$  because

$$\int_{z_-(0; k)}^{z_-(t_0; k)} \frac{d\zeta}{|\zeta|^{1/2}} < \infty. \quad (\text{B.65})$$

Hence  $1 - \mathcal{K}_<$  is invertible for sufficiently small  $\epsilon$ , and the norm of  $(1 - \mathcal{K}_<)^{-1} - 1$  is  $\mathcal{O}(\epsilon)$ . Therefore,

$$\sup_{z_-(0; k) < z < z_-(t_0; k)} \|\epsilon^{1/6} e^{-\tau(k)/\epsilon} (\mathbf{V}_<(z; k) - \mathbf{R}_<(z; k))\| = \mathcal{O}(\epsilon), \quad \epsilon \rightarrow 0, \quad (\text{B.66})$$

because  $\epsilon^{1/6} e^{-\tau(k)/\epsilon} \mathbf{R}_<(z; k)$  is bounded independently of  $\epsilon$  according to (B.61) and taking into account the fact that  $\omega(\epsilon^{-2/3}z_-(t_0; k))\omega(\epsilon^{-2/3}z)^{-1} \leq 1$ .

Finally, we set  $t = 0$  and solve (B.60) for  $\mathbf{F}(0; k)$  taking into account that  $\omega(\epsilon^{-2/3}z_-(0; k)) = 1$  because  $z_-(0; k) < 0$ . Using asymptotic expansions of Airy functions for large negative arguments, and using the fact that the second row of  $\mathbf{R}_<(z_-(0; k); k)$  is exponentially small compared with the first row due to the factor  $\omega(\epsilon^{-2/3}z_-(t_0; k))$  as  $z_-(t_0; k) > 0$ , we obtain

$$\mathbf{F}(0; k) = -e^{\tau(k)/\epsilon} (\mathbf{X}(k; \epsilon) + \mathcal{O}(\epsilon)), \quad \epsilon \rightarrow 0, \quad (\text{B.67})$$

where

$$\mathbf{X}(k; \epsilon) := e^{\frac{2}{3}\text{isgn}(k^2 - k_0^2)(-z_-(0; k))^{3/2}\sigma_3/\epsilon} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} e^{i(\ell(k)\text{sgn}(k^2 - k_\infty^2) - \frac{1}{2}S_\infty)\sigma_3/\epsilon}. \quad (\text{B.68})$$

This calculation also uses the identity

$$\mathbf{G}_<(0; k)(-z_-(0; k))^{-\sigma_3/4} = \frac{1}{\sqrt{2}} \begin{bmatrix} \text{isgn}(k^2 - k_0^2) & -1 \\ 1 & -\text{isgn}(k^2 - k_0^2) \end{bmatrix}. \quad (\text{B.69})$$

This completes the rigorous asymptotic calculation of  $F(0; k)$  by the method of Langer transformations. From (44) we then obtain  $T_0(0; k)$  for small  $\epsilon$ , and then from (9) and (10) we complete the proof of Proposition 2.

## References

1. R. CARROLL and C. BU, Solution of the forced NLS equation using PDE techniques, *Appl. Anal.* 41:33–51 (1991).
2. L. FADDEEV and L. A. TAKHTAJAN, *Hamiltonian Methods in the Theory of Solitons*, Classics in Mathematics. Springer, Berlin, 2007.
3. A. S. FOKAS, *A Unified Approach to Boundary Value Problems*, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, 2008.
4. D. A. SMITH, “Unified Transform Method for Boundary Value Problems” <http://unifiedmethod.azurewebsites.net/>, 2014.
5. A. S. FOKAS, A. R. ITS, and L. Y. SUNG, The nonlinear Schrödinger equation on the half-line, *Nonlinearity* 18:1771–1822 (2005).
6. J. LENELLS, Absence of solitons for the defocusing NLS equation on the half-line, arXiv:1412.3495, 2014.
7. X. ZHOU, The Riemann–Hilbert problem and inverse scattering, *SIAM J. Math. Anal.* 20:966–986 (1989).
8. K. T.-R. McLAUGHLIN and P. D. MILLER, The  $\bar{\partial}$  steepest descent method for orthogonal polynomials on the real line with varying weights, *Intern. Math. Res. Notices* 2008:1–66 (2008).
9. A. BOUTET DE MONVEL, A. R. ITS, and V. KOTLYAROV, Long-time asymptotics for the focusing NLS equation with time-periodic boundary condition, *C. R. Math. Acad. Sci. Paris* 345:615–620 (2007).
10. A. S. FOKAS and J. LENELLS, The unified method: I. Nonlinearizable problems on the half-line, *J. Phys. A* 45:195201 (2012).
11. A. S. FOKAS and J. LENELLS, The unified method: II. NLS on the half-line with  $t$ -periodic boundary conditions, *J. Phys. A* 45:195202 (2012).
12. S. KAMVISSIS, Semiclassical nonlinear Schrödinger on the half line, *J. Math. Phys.* 44:5849–5868 (2003).
13. A. DEGASPERIS, S. V. MANAKOV, and P. M. SANTINI, Initial-boundary value problems for linear and soliton PDEs, *Theor. Math. Phys.* 133:1475–1489 (2002).
14. N. MUSKHELISHVILI, *Singular Integral Equations (2nd ed.)*, Dover Publications, Mineola, NY 1992.
15. X. ZHOU,  $L^2$ -Sobolev space bijectivity of the scattering and inverse scattering transforms, *Comm. Pure Appl. Math.* 51:697–731 (1998).
16. P. A. DEIFT, *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach*, Courant Lecture Notes 3, American Mathematical Society, Providence, RI, 2002.
17. R. J. BUCKINGHAM and P. D. MILLER, Large-degree asymptotics of rational Painlevé-II functions. Noncritical behavior, *Nonlinearity* 27:2489–2577 (2014).
18. P. DEIFT, S. VENAKIDES, and X. ZHOU, New results in small dispersion KdV by an extension of the steepest descent method for Riemann-Hilbert problems, *Internat. Math. Res. Notices* 1997:285–299 (1997).
19. NIST Digital Library of Mathematical Functions. <http://dlmf.nist.gov/>, Release 1.0.8 of 2014-04-25. Online companion to [21].

20. R. J. BUCKINGHAM and P. D. MILLER, The sine-Gordon equation in the semiclassical limit: Dynamics of fluxon condensates, *Memoirs Am. Math. Soc.* 225:1–136 (2013).
21. F. W. J. OLVER, D. W. LOZIER, R. F. BOISVERT, and C. W. CLARK, editors. *NIST Handbook of Mathematical Functions*, Cambridge University Press, New York, NY, 2010. Print companion to [19].

UNIVERSITY OF MICHIGAN  
FUDAN UNIVERSITY

(Received July 8, 2014)