On the Zero-Dispersion Limit of the Benjamin-Ono Cauchy Problem for Positive Initial Data

PETER D. MILLER
University of Michigan

ZHENGJIE XU
University of Michigan

To Dave Levermore on the occasion of his sixtieth birthday.

Abstract

We study the Cauchy initial-value problem for the Benjamin-Ono equation in the zero-dispersion limit, and we establish the existence of this limit in a certain weak sense by developing an appropriate analogue of the method invented by Lax and Levermore to analyze the corresponding limit for the Korteweg–de Vries equation. © 2010 Wiley Periodicals, Inc.

Contents

1 Introduction

The Benjamin-Ono (BO) equation

$$(1.1) \qquad \frac{\partial u_\varepsilon}{\partial t} + 2u_\varepsilon \frac{\partial u_\varepsilon}{\partial x} + \varepsilon \mathcal{H}\left[\frac{\partial^2 u_\varepsilon}{\partial x^2}\right] = 0, \quad x \in \mathbb{R}, \ t > 0,$$
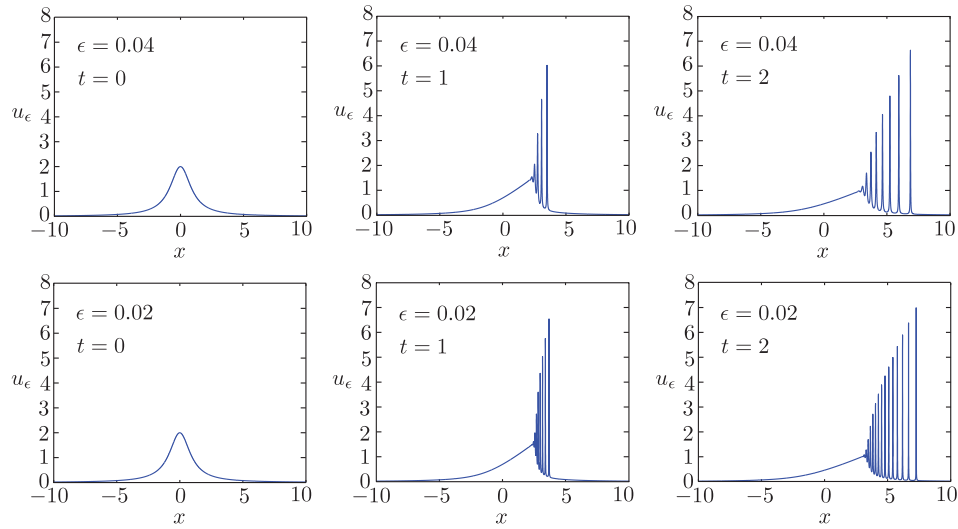
FIGURE 1.1. The evolution of a pulse under the BO equation. Top row: $\varepsilon = 0.04$. Bottom row: $\varepsilon = 0.02$. In both cases the initial condition is the same: $u_0(x) = 2(1 + x^2)^{-1}$.

where $\varepsilon > 0$ is a constant and $\mathcal{H}$ is the Hilbert transform operator defined by the Cauchy principal value integral

(1.2)
$$\mathcal{H}[f](x) := \frac{1}{\pi} \fint_{\mathbb{R}} \frac{f(y)}{y - x} \, dy$$

is a model for weakly nonlinear dispersive waves on the interface between two ideal immiscible fluids, one of which may be considered to be infinitely deep. Applications include the modeling of internal waves in deep water [1, 4, 7, 29], and also the modeling of atmospheric waves like the dramatic "morning glory" phenomenon of northeastern Australia [30]. The relevant Cauchy problem is to determine the solution $u_\varepsilon(x, t)$ of (1.1) subject to a suitable initial condition $u_\varepsilon(x, 0) = u_0(x)$ given for all $x \in \mathbb{R}$.

The parameter $\varepsilon > 0$ is a measure of the relative strength of the dispersive and nonlinear effects in the system. In many applications one thinks of ε as a small parameter in part because numerical experiments show that in this situation the finite-time formation of a shock wave (gradient catastrophe) in the formal limiting equation (obtained simply by setting $\varepsilon = 0$ in (1.1)) is dispersively regularized by the generation of a smoothly modulated train of approximately periodic traveling waves, which correspond to so-called undular bores, frequently observed in the evolution of physical internal waves. Snapshots from the solution of a Cauchy problem for (1.1) illustrating the averted shock and onset of an undular bore are shown in Figure 1.1. These figures clearly show that the mathematical description

of the undular bore consists of waves of amplitude independent of ε and wavelength approximately proportional to ε. We refer to the asymptotic analysis of the solution of the Cauchy problem with ε-independent initial data $u_0(x)$ in the limit $\varepsilon \downarrow 0$ as the *zero-dispersion limit*.

1.1 A Related Problem and Its History

A more famous nonlinear dispersive wave equation is the Korteweg–de Vries (KdV) equation

$$(1.3) \qquad \frac{\partial v_\varepsilon}{\partial t} + 2v_\varepsilon \frac{\partial v_\varepsilon}{\partial x} + \frac{1}{3} \varepsilon^2 \frac{\partial^3 v_\varepsilon}{\partial x^3} = 0, \quad x \in \mathbb{R}, \ t > 0,$$

a model for long surface waves on shallow water among a wide variety of other physical phenomena. When $\varepsilon > 0$ is small, this equation displays qualitatively similar behavior to that just illustrated for the BO equation: the dispersive term arrests the shock in the $\varepsilon = 0$ equation with the formation of a train of waves of amplitude approximately independent of ε and wavelength proportional to ε.

The modeling of the zero-dispersion limit for the KdV equation has a long history going back to the work of Whitham [32], who used the method of averaging to propose a nonlinear hyperbolic system of three partial differential equations to describe the modulational variables (e.g., slowly varying amplitude, mean, and wavelength of the wavetrain). Whitham noted that the system of modulation equations he obtained had the nongeneric property that by choosing special dependent variables v^1, v^2, and v^3, it could be written in so-called Riemann invariant form, in which the three equations are only coupled through the characteristic velocities:

$$(1.4) \qquad \frac{\partial v^i}{\partial t} + c^i(v^1, v^2, v^3) \frac{\partial v^i}{\partial x} = 0, \quad i = 1, 2, 3.$$

Later, Gurevich and Pitaevskii [14] considered the problem of patching together solutions of Whitham's modulational system with solutions of the formal limiting equation (obtained by setting $\varepsilon = 0$ in (1.3)) at two moving boundary points that delineate the oscillation zone; their goal was to provide a reasonable global approximation scheme for the solution of the initial value problem for the KdV equation (1.3) in the zero-dispersion limit subject to given initial data $v_\varepsilon(x, 0) = v_0(x)$ independent of ε.

In the meantime, the KdV equation was discovered to be a completely integrable system possessing a compatible structure now called a *Lax pair* and a coincident solution procedure for addressing the Cauchy (initial value) problem: the *inverse scattering transform*. This development suggested that the methodology invented by Whitham could perhaps be placed on a completely rigorous mathematical footing. After the exact periodic (and quasi-periodic) solutions of the KdV equation (1.3) were given a spectral interpretation by Its and Matveev [15] and Dubrovin, Matveev, and Novikov [11], the Whitham modulation equations themselves were reinterpreted within the framework of integrability by Flaschka, Forest, and McLaughlin [12]. (In particular, this work made clear why Whitham's

equations could be placed in Riemann invariant form; it is a consequence of integrability.)

The task that remained in the use of integrable machinery to study the zero-dispersion limit of the KdV equation was to rigorously analyze the Cauchy problem using the inverse scattering transform. The first step in this program was taken by Lax and Levermore [20], who considered positive initial data $v_0(x)$ rapidly decaying to 0 for large $|x|$. They used Wentzel-Kramer-Brillouin (WKB) methods to argue that the Schrödinger operator with potential $-v_0(x)$ that arises in the scattering theory is approximately reflectionless in the limit $\varepsilon \downarrow 0$. On an ad hoc basis they replaced the true scattering data by its WKB analogue, retaining only contributions from a set of $N(\varepsilon) \sim \varepsilon^{-1}$ discrete eigenvalues. These eigenvalues are approximated by a Bohr-Sommerfeld quantization rule, which amounts to replacing the solution $v_\varepsilon(x, t)$ of the Cauchy problem with another solution $\tilde{v}_\varepsilon(x, t)$ of (1.3) having ε-dependent initial data close to v_0. In this situation, the inverse scattering procedure reduces to finite-dimensional (of dimension $N(\varepsilon)$) linear algebra, and in fact the solution obtained by Cramer's rule can be reduced to the determinantal formula

$$(1.5) \qquad \tilde{v}_\varepsilon(x, t) = 2\varepsilon^2 \, \frac{\partial^2}{\partial x^2} \log(\tau(x, t)), \quad \tau(x, t) = \det(\mathbb{I} + \mathbf{G}(x, t)),$$

where $\mathbf{G}(x, t)$ is a positive-definite real symmetric matrix of dimension $N(\varepsilon) \times N(\varepsilon)$. Lax and Levermore then established the existence of the limit, uniform on compact subsets of the (x, t)-plane, of $2\varepsilon^2 \log(\tau(x, t))$ as $\varepsilon \downarrow 0$. This yields weak convergence of $\tilde{v}_\varepsilon(x, t)$ by differentiation of the limit function with respect to x. The Lax-Levermore method is to expand the determinant $\tau(x, t)$ in principal minors indexed by subsets of the set of eigenvalues; noting that each term is positive, they showed that the sum of terms is asymptotically dominated by its largest term, and they further approximated this discrete optimization problem with an ε-independent (limiting) convex variational problem, explicitly parametrized by x and t, for measures. The weak zero-dispersion limit of the Cauchy problem for the KdV equation is therefore encoded implicitly in the solution of this variational problem. The Lax-Levermore method reproduces the specified initial data $v_0(x)$ at $t = 0$ as $\varepsilon \downarrow 0$, which establishes validity, in a certain sense, of the WKB-based spectral approximation procedure in the first step.

Later, Venakides [31] was able to extend the method of Lax and Levermore to higher order, capturing the form of the oscillations that are averaged out in the weak limit. This work at last made clear that the solution of the Cauchy problem for the KdV equation with smooth, ε-independent initial data $v_0(x)$ really does generate after some fixed breaking time a train of high-frequency waves of exactly the kind originally considered without complete justification by Whitham. More recently the powerful steepest-descent method for matrix Riemann-Hilbert problems developed by Deift and Zhou was used to analyze the zero-dispersion limit for the KdV equation [8]. This technique is best viewed as a tool for converting weak

asymptotics (the solution of the Lax-Levermore variational problem) into strong asymptotics (an improvement of the Venakides asymptotics in which the phase of the waveform is accurate to very high order).

1.2 Zero-Dispersion Limit of the Benjamin-Ono Equation

It turns out that the BO equation (1.1) is also an integrable equation in the sense that it has a representation as the compatibility condition of an overdetermined Lax pair of linear problems [2]. In fact, both BO and KdV equations may be viewed as limiting cases (as depth of a fluid layer tends to infinity and zero, respectively) of the so-called intermediate long-wave equation [18], itself an integrable system for arbitrary layer depth. However, the integrable structure of the BO equation is markedly different from that of the KdV equation. In particular, the nonlocality in the equation due to the presence of the Hilbert transform is mirrored in a certain nonlocality of the scattering and inverse scattering problems. In place of the spectral theory of the self-adjoint Schrödinger (Sturm-Liouville) differential operator, one has to work with the spectral theory of the nonlocal operator

$$(1.6) \quad \mathcal{L} := -i\varepsilon \frac{\partial}{\partial x} - \mathcal{C}_+ \circ u_\varepsilon \circ \mathcal{C}_+, \quad \mathcal{C}_+[f](x) := \lim_{\delta\downarrow 0} \frac{1}{2\pi i} \int_{\mathbb{R}} \frac{f(y)}{y - x - i\delta}\, dy.$$

Here the operator $\mathcal{C}_+$ is the self-adjoint orthogonal projection from $L^2(\mathbb{R})$ onto the Hardy space of the upper half-plane, the Hilbert space on which $\mathcal{L}$ is self-adjoint, and u_ε denotes the operator of multiplication by $u_\varepsilon(\cdot, t)$.

Certainly a key step forward in the theory of the zero-dispersion limit was taken by Dobrokhotov and Krichever [9], who noted that the second (time evolution) equation in the Lax pair for the BO equation (see (2.2) below) is simply a time-dependent Schrödinger equation whose potential is a function with an analytic continuation from the real x-axis into the upper half-plane; they were able to adapt a pre-existing construction of "integrable" potentials for this equation to the appropriate Hardy-space setting. This allowed them to construct, from the Lax pair, a large family of periodic traveling wave solutions of the BO equation (1.1), along with quasi-periodic generalizations. Remarkably, unlike the corresponding exact solutions of the KdV equation (1.3), which are highly transcendental objects constructed from Riemann theta functions of hyperelliptic curves of arbitrary genus, the periodic and quasi-periodic solutions of the BO equation turn out to be simple rational functions of P exponential phases $e^{i(k_j x - \omega_j t)/\varepsilon}$.

In the same paper, Dobrokhotov and Krichever also carried out for the BO equation the analogue of the calculation of Flaschka, Forest, and McLaughlin [12], deriving by multiphase averaging a system of equations governing the modulational variables for a slowly varying train of P-phase waves. Here we arrive at a second remarkable fact: not only can the modulation equations be written in Riemann

invariant form, they are completely diagonal:

$$(1.7) \qquad \frac{\partial u^i}{\partial t} + 2u^i \frac{\partial u^i}{\partial x} = 0, \quad i = 1, 2, \ldots, 2P + 1$$

(the case of $P = 1$ corresponds to simple traveling waves). This again should be contrasted with the situation for the KdV equation in which the characteristic velocities not only provide coupling among the fields but also are transcendental functions of the fields written in terms of ratios of complete hyperelliptic integrals.

The analogue for the BO equation of the matching procedure developed by Gurevich and Pitaevskii [14] to describe the evolution of a dispersive shock was independently described by Matsuno [25, 26] and by Jorge, Minzoni, and Smyth [16]. This matching procedure provides a reasonable approach to the Cauchy problem for the Benjamin-Ono equation (1.1) with fixed initial data $u_\varepsilon(x, 0) = u_0(x)$ when $\varepsilon \ll 1$, but it is based on formal asymptotics. In [25, p. 7939], Matsuno writes:

> From a rigorously mathematical point of view, however, the various results presented in this paper should be justified on the basis of an exact method of solution such as [the inverse scattering transform], or an analog of the Lax-Levermore theory for the KdV equation.

It is our intention in this paper to provide exactly such a justification, by developing a new method that does for the BO Cauchy problem exactly what the Lax-Levermore method does for the KdV Cauchy problem.

The main result of our analysis is remarkably easy to state, but first we need to recall some basic facts concerning the equation obtained from (1.1) simply by setting $\varepsilon = 0$. Recall that while for general sufficiently smooth initial data $u^B(x, 0) = u_0(x)$ the inviscid Burgers equation

$$(1.8) \qquad \frac{\partial u^B}{\partial t} + 2u^B \frac{\partial u^B}{\partial x} = 0$$

does not have a global solution as a function due to gradient catastrophe (shock formation) in finite time, it does have a global solution as a real multisheeted surface over the (x, t)-plane; indeed, this is the construction of the method of characteristics. The sheets of this surface are obtained as the real solutions of the implicit equation

$$(1.9) \qquad u^B = u_0(x - 2u^B t),$$

and by implicit differentiation it is easy to verify that away from singularities each sheet of the surface is a function $u^B = u^B(x, t)$ that satisfies (1.8). A simple consequence of the implicit function theorem is that for sufficiently small $|t|$ there is a unique solution of (1.9) for all $x \in \mathbb{R}$. New sheets of the multivalued solution are born from *breaking points* in the (x, t)-plane that are in one-to-one correspondence with generic inflection points ξ of u_0 for which $u_0'(\xi) \neq 0$ but $u_0''(\xi) = 0$. If $\xi \in \mathbb{R}$
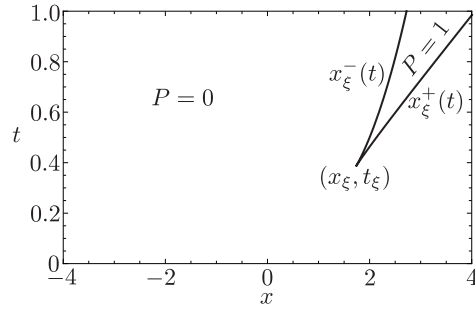
FIGURE 1.2. Except along the caustic curves $x = x_\xi^-(t)$ and $x = x_\xi^+(t)$, the number of solutions of (1.9) is of the form $2P + 1$, and these solutions are simple roots. For this figure, $u_0(x) := 2(1 + x^2)^{-1}$.

is such a point, then the corresponding breaking point is given by

$$(1.10) \qquad (x_\xi, t_\xi) := \left(\xi - \frac{u_0(\xi)}{u_0'(\xi)}, -\frac{1}{2u_0'(\xi)} \right).$$

Each such breaking point is the location of a pitchfork bifurcation for u^B with respect to t holding $x - 2u_0(\xi)t = \xi$ fixed, with two new branches emerging as $|t|$ increases. Thus, assuming that u_0' is a bounded function of total integral zero, the solution of the Cauchy problem for (1.8) is classical for

$$(1.11) \qquad T_- := -\frac{1}{2 \max_{x \in \mathbb{R}} u_0'(x)} < t < -\frac{1}{2 \min_{x \in \mathbb{R}} u_0'(x)} =: T_+.$$

Note that under our assumptions on u_0' we have $T_- < 0 < T_+$. Also, T_- is the supremum of all $t_\xi < 0$ while T_+ is the infimum of all $t_\xi > 0$. When we consider the Cauchy problem for $t > 0$, we will refer to $T := T_+$ as the *breaking time*.

For $t/t_\xi > 1$ there are *caustic curves* $x_\xi^-(t) < x_\xi^+(t)$ with limiting values as $t \to t_\xi$ given by $x_\xi^-(t_\xi) = x_\xi^+(t_\xi) = x_\xi$ that bound the triple-folded region emerging from (x_ξ, t_ξ). The caustic curves correspond to double roots of (1.9), and crossing one of them at a generic point results in a change in the number of sheets by exactly 2. Except along the union of the caustic curves and the breaking points from which they emerge, the number of solutions of (1.9) is always odd, and all are simple roots. See Figure 1.2.

For the initial data $u_0(x) = 2(1 + x^2)^{-1}$ used in Figure 1.1, the breaking time before which there is a unique solution for all $x \in \mathbb{R}$ and after which there is an expanding interval in which there are three solutions, is exactly $T = 2\sqrt{3}/9 \approx 0.3849$. Snapshots of the evolution of the multivalued solution of (1.8) for this initial data are shown in Figure 1.3. Our result is then the following:
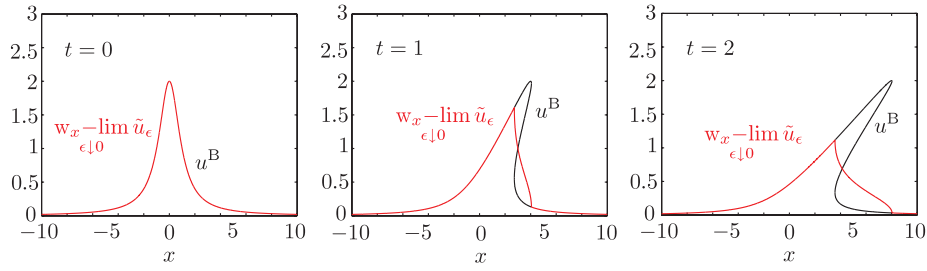
FIGURE 1.3. The multivalued solution (black) of (1.8) and the signed sum of branches (red) corresponding to $u_0(x) = 2(1 + x^2)^{-1}$. Left: $t = 0$. Middle: $t = 1$. Right: $t = 2$. Before the breaking time as well as afterwards but outside the oscillation interval there is only one solution branch and hence no difference between the red and black curves.

THEOREM 1.1 *Let $u_0^B(x,t) < u_1^B(x,t) < \cdots < u_{2P(x,t)}^B(x,t)$ be the branches of the multivalued* (*method of characteristics*) *solution of the inviscid Burgers equation* (1.8) *subject to an admissible initial condition $u^B(x,0) = u_0(x)$. Then the weak $L^2(\mathbb{R})$* (*in x*) *limit of $\tilde{u}_\varepsilon(x,t)$ is given by*

$$(1.12) \qquad \mathrm{w}_x\text{--}\lim_{\varepsilon\downarrow 0} \tilde{u}_\varepsilon(x,t) = \sum_{n=0}^{2P(x,t)} (-1)^n u_n^B(x,t),$$

uniformly for t in arbitrary bounded intervals. Note that the right-hand side extends by continuity to the caustic curves.

The signed sum of branches that is the weak limit is illustrated with red curves in Figure 1.3 for the same initial data as in Figure 1.1. Of course, convergence in the weak $L^2(\mathbb{R})$ (in x) topology means that for every $v \in L^2(\mathbb{R})$, we have

$$(1.13) \qquad \lim_{\varepsilon\downarrow 0} \int_{\mathbb{R}} \tilde{u}_\varepsilon(x,t)v(x)dx = \int_{\mathbb{R}} \left[\sum_{n=0}^{2P(x,t)} (-1)^n u_n^B(x,t) \right] v(x)dx$$

with the limit being uniform with respect to t in arbitrary bounded intervals. Thus the weak limit essentially smooths out the rapid oscillations seen in Figure 1.1 and (if we think of v as the indicator function of a mesoscale interval) represents a kind of local average in x. What it means for an initial condition to be admissible will be explained later (see Definition 3.1). Here $\tilde{u}_\varepsilon(x,t)$ is not exactly the solution of the Cauchy problem for the BO equation (1.1) with fixed initial data $u_0(x)$, but it is for every $\varepsilon > 0$ an exact solution of (1.1) that satisfies an ε-dependent initial condition that converges (in the strong L^2 sense; see Corollary 1.2 below) to $u_0(\cdot)$ as $\varepsilon \downarrow 0$. See Definition 3.3 for more details. This modification of the initial data is an analogue of the replacement of the true scattering data by its reflectionless WKB approximation in the Lax-Levermore theory.

For t before the breaking time T for Burgers' equation, the weak limit guaranteed by Theorem 1.1 may be strengthened as follows.

COROLLARY 1.2 *Suppose that $0 \leq t < T$ so that $P(x,t) = 0$ for all $x \in \mathbb{R}$ (that is, the solution $u^{\text{B}} = u_0^{\text{B}}(x,t)$ of Burgers' equation with initial data $u_0(x)$ is classical). Then*

$$(1.14) \qquad \lim_{\varepsilon \downarrow 0} \tilde{u}_\varepsilon(x,t) = u_0^{\text{B}}(x,t)$$

with the limit being in the (strong) $L^2(\mathbb{R}_x)$ topology.

It should be pointed out that the weak limit formula (1.12) is much more explicit than the corresponding formula found by Lax and Levermore [20] for the weak zero-dispersion limit of the Cauchy problem for the KdV equation. Indeed, the latter requires the solution, for each x and t, of a constrained functional variational problem, which can be solved in closed form only for the simplest initial data. As we will now see, there are several classical wave propagation problems whose asymptotic behavior can be reduced to the multivalued solution of Burgers' equation; however, even these simple problems involve more complicated schemes for combining the solution branches than that exhibited in the simple formula (1.12).

After we introduce the necessary framework for our study (the inverse scattering transform for the BO equation) in Section 2, we will analyze the direct scattering map in the zero-dispersion limit in Section 3. Then we will prove Theorem 1.1 and Corollary 1.2 in Section 4 by carrying out a detailed analysis of the inverse scattering map applied to the asymptotic formulae for scattering data obtained in Section 3. In Section 5 we will illustrate our results with numerical calculations and address the relation between $\tilde{u}_\varepsilon(x,t)$ and the true solution $u_\varepsilon(x,t)$ of the Cauchy problem for the BO equation with initial data u_0. We compare the zero-dispersion asymptotics of the BO equation with some elementary examples from linear and nonlinear wave theory in Section 6, and some comments about our continuing work can be found in the conclusion, Section 7.

2 Relevant Aspects of the Inverse Scattering Transform for the BO Cauchy Problem

2.1 The Lax Pair for the BO Equation and Its Basic Properties

The Lax pair [2], whose compatibility condition is the BO equation (1.1), consists of the two equations

$$(2.1) \qquad i\varepsilon \frac{\partial w^+}{\partial x} + \lambda(w^+ - w^-) = -u_\varepsilon w^+,$$

$$(2.2) \qquad i\varepsilon \frac{\partial w^\pm}{\partial t} - 2i\lambda\varepsilon \frac{\partial w^\pm}{\partial x} + \varepsilon^2 \frac{\partial^2 w^\pm}{\partial x^2} - 2i\mathcal{C}_\pm \left[\varepsilon \frac{\partial u_\varepsilon}{\partial x}\right] w^\pm = 0,$$

where $\lambda \in \mathbb{C}$ is a spectral parameter, $u_\varepsilon = u_\varepsilon(x,t)$ is a solution of (1.1), and $w^\pm = w^\pm(x,t;\lambda)$ are functions that are required to be, for each fixed t and λ,

the boundary values on the real x-axis of functions analytic in the upper $(+)$ and lower $(-)$ half complex x-plane. Also, $\pm\mathcal{C}_\pm = \frac{1}{2}\mathbb{I} \mp \frac{1}{2}i\mathcal{H}$ are the orthogonal and complementary $(\mathcal{C}_+ - \mathcal{C}_- = \mathbb{I}$, the Plemelj formula) projections from $L^2(\mathbb{R})$ onto its upper and lower Hardy subspaces $\mathbb{H}^\pm(\mathbb{R})$. From the point of view of the inverse scattering transform, i.e., using the Lax pair as a tool to solve the Cauchy problem, equation (2.1) may be considered for fixed time t and defines the scattering data associated with $u_\varepsilon(x,t)$ at time t. The function w^- may be viewed as a kind of Lagrange multiplier present to satisfy the constraint that w^+ be an "upper" function. In fact, if $w^\pm \in \mathbb{H}^\pm(\mathbb{R})$, then by applying $\mathcal{C}_+$ to (2.1) and using the projective identities $\mathcal{C}_+[w^+] = w^+$ and $\mathcal{C}_+[w^-] = 0$, (2.1) can be written in the form of an eigenvalue problem

$$(2.3) \qquad \mathcal{L}w^+ = \lambda w^+, \quad w^+ \in \mathbb{H}^+(\mathbb{R}),$$

where $\mathcal{L}$ is the nonlocal self-adjoint operator (1.6). Equation (2.2) determines the (trivial, as we will recall) time dependence of the scattering data.

2.2 Scattering Data

The theory of the inverse scattering transform solution of the Cauchy problem for the BO equation was first developed by Fokas and Ablowitz [13]. Certain analytical details of the theory were clarified by Coifman and Wickerhauser [6], and more recently Kaup and Matsuno [17] found conditions on the scattering data consistent with real-valued solutions of (1.1). As an operator on $\mathbb{H}^+(\mathbb{R})$, the essential spectrum of $\mathcal{L}$ is the positive real λ-axis (for suitable u_ε, $\mathcal{L}$ is a relatively compact perturbation of the "free" operator corresponding to $u_\varepsilon \equiv 0$). For each fixed t and each real $\lambda > 0$, there exists a unique solution $w^+ = M$ of (2.1) with the property that (remarkably, despite the nonlocal nature of the problem) it is determined by its asymptotic behavior as $x \to -\infty$ on the real line: $M(x,t;\lambda) = 1 + o(1)$ as $x \to -\infty$. As the problem is nonlocal, M cannot be characterized by a Volterra-type integral equation, but Fokas and Ablowitz [13] gave a Fredholm-type equation whose unique solution is M. The *reflection coefficient* for the problem is then defined for positive real λ by the formula [13, 17]

$$(2.4) \qquad \beta(\lambda,t) := \frac{i}{\varepsilon} \int_{\mathbb{R}} u_\varepsilon(x,t)M(x,t;\lambda)e^{-i\lambda x/\varepsilon}\, dx, \quad \lambda > 0.$$

For each fixed x and t the function $M(x,t;\lambda)$ can be shown to be the boundary value taken on the positive half-line $\lambda \in \mathbb{R}_+$ from the upper half λ-plane of a function $W(x,t;\lambda)$ that is meromorphic in the complex λ-plane with $\mathbb{R}_+$ (a branch cut) deleted. Fokas and Ablowitz refer to the boundary value taken by $W(x,t;\lambda)$ on the positive half-line from the *lower* half-plane as $\overline{N}(x,t;\lambda)$. The poles of $W(x,t;\lambda)$ are all on the negative real λ-axis (by self-adjointness of $\mathcal{L}$) and correspond to the point spectrum of $\mathcal{L}$. It turns out that one of the consequences of the Lax pair equation (2.2) is that the point spectrum is independent of time t. In [13] it is shown that in the generic case when $\lambda_n < 0$ is a simple pole of W, the first two terms in

the Laurent expansion of W at $\lambda = \lambda_n$ are both proportional to the same function $\Phi_n(\cdot, t) \in \mathbb{H}^+(\mathbb{R})$, which is an eigenfunction of $\mathcal{L}$ with eigenvalue $\lambda = \lambda_n$. The ratio of these two terms is in fact linear in x:

$$(2.5) \quad W(x, t; \lambda) = -i\varepsilon \frac{\Phi_n(x, t)}{\lambda - \lambda_n} + (x + \alpha_n(t))\Phi_n(x, t) + O(\lambda - \lambda_n), \quad \lambda \to \lambda_n.$$

Kaup and Matsuno [17] showed that for real u_ε, the complex-valued phase shift $\alpha_n(t)$ may be written in the form

$$(2.6) \qquad\qquad \alpha_n(t) = \gamma_n(t) - \frac{i}{2\lambda_n}, \quad \gamma_n(t) \in \mathbb{R}.$$

The set of *scattering data* corresponding to the potential $u_\varepsilon(\cdot, t)$ then consists of

- the reflection coefficient $\beta(\lambda, t)$ for $\lambda > 0$ (we write $\beta(\lambda)$ for $\beta(\lambda, 0)$),
- the negative discrete eigenvalues $\{\lambda_n\}_{n=1}^N$, $\lambda_1 < \lambda_2 < \cdots < \lambda_N < 0$, and
- the real *phase constants* $\{\gamma_n(t)\}_{n=1}^N$ (we write γ_n for $\gamma_n(0)$).

2.3 Time Dependence of Scattering Data and the Inverse Scattering Transform

As time varies, one may expect the scattering data to vary, but the time dependence as implied by (2.2) turns out to be very simple. As pointed out above, the discrete eigenvalues $\{\lambda_n\}_{n=1}^N$ are constants of the motion, and Fokas and Ablowitz [13] showed that

$$(2.7) \qquad\qquad \beta(\lambda, t) = \beta(\lambda)e^{i\lambda^2 t/\varepsilon}, \quad \lambda > 0,$$

and

$$(2.8) \qquad\qquad \gamma_n(t) = \gamma_n + 2\lambda_n t, \quad n = 1, 2, \ldots, N.$$

The inverse scattering procedure for solving the Cauchy problem for the BO equation with suitable real initial data $u_0(x)$ is then to calculate the scattering data at time $t = 0$ from u_0, evolve the scattering data forward in time t by the explicit formulae (2.7) and (2.8), and then solve the inverse problem of constructing $u_\varepsilon(\cdot, t)$ from the scattering data at time t. Generally, this requires solving a scalar Riemann-Hilbert problem for $W(x, t; \lambda)$ in the complex λ-plane. This Riemann-Hilbert problem is quite interesting as it involves a jump condition across the continuous spectrum $\lambda > 0$ in which the boundary value from above, $W = M(x, t; \lambda)$, is proportional to an integral from $\lambda' = 0$ to $\lambda' = \lambda$ of the boundary value from below, $W = \overline{N}(x, t; \lambda')$. Thus the jump condition is nonlocal, a fact that makes the inverse problem almost completely analogous to the direct problem (2.1), which after integration becomes a nonlocal Riemann-Hilbert problem of exactly the same type in the complex x-plane. This fact should perhaps be contrasted with the situation for the KdV equation, where the direct and inverse problems are of quite different natures. This remarkable symmetry between the forward and inverse problems for the BO equation is a theme that will be touched upon again in this paper in some detail.

2.4 Reflectionless Inverse Scattering Transform

If $\beta(\lambda) \equiv 0$ (i.e., the problem is *reflectionless*), then the boundary values taken by $W(x,t;\lambda)$ on the positive half-line agree, so $W(x,t;\lambda)$ is a meromorphic function on the whole complex λ-plane with simple poles at the negative real eigenvalues. Condition (2.5) and the normalization condition that $W(x,t;\lambda) \to 1$ as $\lambda \to \infty$ then provides sufficient information to reconstruct $W(x,t;\lambda)$ from the discrete data $\{\lambda_n\}_{n=1}^N$ and $\{\gamma_n\}_{n=1}^N$. Via a partial fractions ansatz for $W(x,t;\lambda)$, this amounts to a solving a linear algebra problem in dimension N. Once W is determined in this way, one obtains $\mathcal{C}_+[u_\varepsilon(\cdot,t)]$ by the formula

$$(2.9) \qquad \mathcal{C}_+[u_\varepsilon(\cdot,t)](x) = \lim_{\lambda \to \infty} \lambda(1 - W(x,t;\lambda)).$$

Since u_ε is real, one then has

$$(2.10) \qquad u_\varepsilon(x,t) = 2\operatorname{Re}\{\mathcal{C}_+[u_\varepsilon(\cdot,t)](x)\}.$$

This procedure clearly leads to a determinantal formula for $u_\varepsilon(x,t)$ in the reflectionless case. It turns out to be the same (multisoliton) formula that Matsuno [21] had obtained, before the relevant inverse scattering transform was discovered, by applying Hirota's bilinear method to the BO equation:

$$(2.11) \qquad u_\varepsilon(x,t) = 2\varepsilon \frac{\partial}{\partial x}\operatorname{Im}\{\log(\tau_\varepsilon(x,t))\},$$

with the "tau function"

$$(2.12) \qquad \tau_\varepsilon(x,t) := \det(\mathbb{I} + i\varepsilon^{-1}\mathbf{A}_\varepsilon),$$

where $\mathbf{A}_\varepsilon = \mathbf{A}_\varepsilon(x,t)$ is an $N \times N$ Hermitian matrix with constant off-diagonal elements

$$(2.13) \qquad A_{nm} := \frac{2i\varepsilon\sqrt{\lambda_n\lambda_m}}{\lambda_n - \lambda_m}, \quad n \neq m,$$

and diagonal elements depending explicitly on x and t,

$$(2.14) \qquad A_{nn} := -2\lambda_n(x + 2\lambda_n t + \gamma_n).$$

In (2.13) we mean the positive square root of the positive product $\lambda_n\lambda_m$. For the purposes of this paper, we will only require this reflectionless version of the inverse scattering transform.

In his paper [21], Matsuno noted that regardless of the value of N, the complex determinant $\tau_\varepsilon(x,t)$ satisfies the real equation (Hirota bilinear form of the BO equation)

$$(2.15) \quad \left(i\varepsilon\frac{\partial\tau_\varepsilon}{\partial t} + 2\varepsilon^2\frac{\partial^2\tau_\varepsilon}{\partial x^2}\right)\tau_\varepsilon^* + \left(-i\varepsilon\frac{\partial\tau_\varepsilon^*}{\partial t} + 2\varepsilon^2\frac{\partial^2\tau_\varepsilon^*}{\partial x^2}\right)\tau_\varepsilon =$$

$$\varepsilon\frac{\partial}{\partial x}\left(\varepsilon\frac{\partial\tau_\varepsilon}{\partial x}\tau_\varepsilon^* + \varepsilon\frac{\partial\tau_\varepsilon^*}{\partial x}\tau_\varepsilon\right).$$

The terms on the left-hand side should be compared with the linear Schrödinger equation (this connection is explored further in Section 6.2). If one makes a formal WKB ansatz of the form $\tau_\varepsilon(x,t) = A(x,t)e^{iS(x,t)/\varepsilon}$, then the terms on the right-hand side of (2.15) are formally small compared with those on the left-hand side, and to leading order in ε (2.15) simply reduces to the inviscid Burgers equation (1.8) with $u^B = 2\partial S/\partial x$ (as is consistent with (2.11)).

2.5 Conservation Laws and Trace Formulae

As with the KdV equation, the time evolution of the BO equation preserves an infinite number of functionals of u_ε. These were first found by Nakamura [28]. The equivalent representation of these functionals in terms of the time-independent portion of the scattering data, i.e., the eigenvalues $\{\lambda_n\}_{n=1}^N$ and the modulus of the reflection coefficient $|\beta(\lambda)|^2$ for $\lambda > 0$, was obtained by Kaup and Matsuno [17]. These identities amount to a hierarchy of *trace formulae* for the operator $\mathcal{L}$.

The conservation laws take the form $dI_k/dt = 0$, $k = 1, 2, 3, \ldots$. The integrals I_k may be generated by the following recursive procedure: first set $\rho_1 := 1$ and then define

$$(2.16) \quad \rho_{k+1}(x,t) := \mathcal{C}_+[u_\varepsilon(\cdot,t)\rho_k(\cdot,t)](x) + i\varepsilon\frac{\partial\rho_k}{\partial x}(x,t), \quad k = 1, 2, 3, \ldots.$$

Then, the integrals of motion are

$$(2.17) \quad I_k(t) := \int_{\mathbb{R}} u_\varepsilon(x,t)\rho_k(x,t)dx, \quad k = 1, 2, 3, \ldots.$$

The equivalent spectral representation given in [17] is

$$(2.18) \quad I_k(t) = 2\pi\varepsilon\sum_{n=1}^N(-\lambda_n)^{k-1}$$
$$+ \frac{(-1)^k\varepsilon}{2\pi}\int_0^{+\infty}|\beta(\lambda)|^2\lambda^{k-2}\,d\lambda, \quad k = 1, 2, 3, \ldots.$$

In view of the results presented in Section 2.3, the latter representation makes clear the fact that $dI_k/dt = 0$.

The first two conserved quantities are quite simple, and in fact they are the only ones in the hierarchy having local densities:

$$(2.19) \quad I_1 := \int_{\mathbb{R}} u_\varepsilon(x,t)dx \quad \text{and} \quad I_2 := \frac{1}{2}\int_{\mathbb{R}} u_\varepsilon(x,t)^2\,dx.$$

3 Scattering Data in the Zero-Dispersion Limit

In this section we consider the following problem. Given a suitable function $u_0(x)$ representing the initial condition for the BO equation, we wish to determine an asymptotic approximation, valid when $\varepsilon > 0$ is small, to the scattering data $\{\beta(\lambda), \{\lambda_n\}_{n=1}^N, \{\gamma_n\}_{n=1}^N\}$ corresponding to u_0. Even though u_0 is held fixed as

ε tends to zero, the scattering data will depend on ε as this parameter appears in equation (2.1). As the operator $\mathcal{L}$ is nonlocal, we cannot rely on the WKB method as is so useful for analysis of differential operators (for example, the analysis of Lax and Levermore [20] was based on the WKB analysis of the Schrödinger operator that arises in the scattering theory for the KdV equation).

3.1 Admissible Initial Conditions

The type of initial data for the BO equation (1.1) that we will consider for the rest of this paper is the following. Many of these conditions are imposed for our convenience; we make no claim that they are necessary.

DEFINITION 3.1 A function $u_0 : \mathbb{R} \to \mathbb{R}$ will be called an *admissible initial condition* if it has the following properties:

- *Smoothness*: $u_0 \in C^3(\mathbb{R})$.
- *Positivity*: $u_0(x) > 0$ for all $x \in \mathbb{R}$.
- *Existence of a unique critical point*: There is a unique point $x_0 \in \mathbb{R}$ for which $u_0'(x_0) = 0$. Moreover,

$$u_0''(x_0) < 0, \tag{3.1}$$

  making x_0 the global, nondegenerate maximizer of u_0.
- *Tail behavior*: $\lim_{x \to \pm\infty} u_0(x) = 0$ and

$$\lim_{x \to \pm\infty} |x|^{q+1} u_0'(x) = C_\pm \quad \text{for some } q > 1, \tag{3.2}$$

  where $C_+ < 0$ and $C_- > 0$ are constants. These two conditions together imply that an admissible initial condition u_0 also satisfies

$$\lim_{x \to \pm\infty} |x|^q u_0(x) = \mp \frac{C_\pm}{q}. \tag{3.3}$$

- *Inflection points*: In each bounded interval there exist at most finitely many points $x = \xi$ at which $u_0''(\xi) = 0$, and each is a simple inflection point: $u_0'''(\xi) \neq 0$.

Corresponding to an admissible initial condition u_0 we define a positive constant L by

$$L := \max_{x \in \mathbb{R}} u_0(x), \tag{3.4}$$

and we let the *mass M* be defined by

$$M := \frac{1}{2\pi} \int_{\mathbb{R}} u_0(x) dx. \tag{3.5}$$

Note that the mass is guaranteed to be finite according to (3.3) since u_0 is bounded. Also, if u_0 is an admissible initial condition, we can define *turning points* $x_\pm :$ $[-L, 0) \to \mathbb{R}$ that are two monotone branches of the inverse function of u_0: $u_0(x_\pm(\lambda)) = -\lambda$ and $x_-(\lambda) \leq x_0 \leq x_+(\lambda)$ for $-L \leq \lambda < 0$. See Figure 3.1.
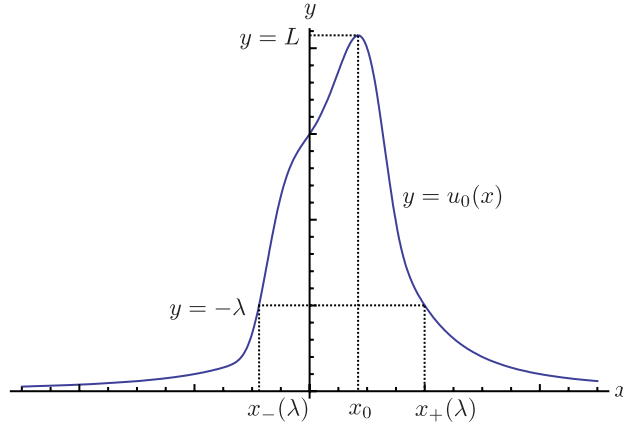
FIGURE 3.1. The graph of an admissible initial condition and the turning points $x_{\pm}(\lambda)$.

3.2 Matsuno's Method

In two papers [22, 23], Matsuno proposed a remarkable method to approximate, in the limit $\varepsilon \downarrow 0$, the time-independent components of the scattering data for suitable u_0. His method was based on the conservation laws for the quantities (2.17). With the use of the more recently obtained trace formulae equating I_k as given by (2.17) with the equivalent formulae (2.18) [17], several heuristic aspects of the original method given in [22, 23] can be placed on more rigorous footing.

The first key observation made in [22, 23] is that if u_0 is a smooth function independent of ε, then by evaluating the integrals $I_k(t)$ at time $t = 0$, one sees that they have limiting values as $\varepsilon \downarrow 0$. These limits may be obtained simply by solving the recurrence relation (2.16) with $\varepsilon = 0$:

(3.6)
$$\lim_{\varepsilon \downarrow 0} I_k = \int_{\mathbb{R}} u_0(x)\mathcal{C}_+[u_0\mathcal{C}_+[u_0\mathcal{C}_+[\cdots u_0\mathcal{C}_+[u_0]\cdots]]](x)dx,$$
$$k = 1, 2, 3, \ldots,$$

where the Cauchy projector $\mathcal{C}_+$ occurs $k - 1$ times in the integrand. With the use of an identity valid for reasonable complex-valued functions $u_0(\cdot)$ and suggested by comparing the conserved quantities generated from the Kaup-Matsuno iteration scheme (2.16) with those generated via the older scheme of Nakamura [28], one sees that the right-hand side of (3.6) can be equivalently written in the simple form

(3.7)
$$\lim_{\varepsilon \downarrow 0} I_k = \frac{1}{k} \int_{\mathbb{R}} u_0(x)^k \, dx, \quad k = 1, 2, 3, \ldots.$$

A direct proof of the equivalence of (3.6) and (3.7) is given in the Appendix. On the basis of heuristic physical arguments, in [22, 23] Matsuno supposed that for

smooth *positive* initial data u_0, all moments of the reflection coefficient remain bounded as $\varepsilon \downarrow 0$. Adopting this hypothesis, a comparison of (3.7) with (2.18) then shows that

$$(3.8) \qquad \lim_{\varepsilon \downarrow 0} \varepsilon \sum_{n=1}^{N} (-\lambda_n)^{k-1} = \frac{1}{2\pi k} \int_{\mathbb{R}} u_0(x)^k \, dx, \quad k = 1, 2, 3, \ldots.$$

In particular, taking $k = 1$ one obtains

$$(3.9) \qquad \lim_{\varepsilon \downarrow 0} \varepsilon N = M,$$

where the mass M is defined by (3.5), so the number of eigenvalues is asymptotically proportional to $1/\varepsilon$.

These calculations suggest that the normalized counting measure of eigenvalues may have a limit in a certain sense as $\varepsilon \downarrow 0$, perhaps as an absolutely continuous measure with density $F(\lambda)$. Matsuno calculated this density by replacing the left-hand side of (3.8) with an integral against the unknown density $F(\lambda)$:

$$(3.10) \qquad \int_{-\infty}^{0} (-\lambda)^{k-1} F(\lambda) d\lambda = \frac{1}{2\pi k} \int_{\mathbb{R}} u_0(x)^k \, dx, \quad k = 1, 2, 3, \ldots.$$

The problem that remains is then the classical one of constructing the density $F(\lambda)$ from its moments, which are known if the initial condition u_0 is given.

Matsuno showed that, remarkably, this moment problem can be solved explicitly. He introduced the characteristic function (Fourier transform) of F,

$$(3.11) \qquad \widehat{F}(\xi) := \int_{-\infty}^{0} F(\lambda) e^{-i\xi\lambda} \, d\lambda,$$

in terms of which the moment relations (3.10) become

$$(3.12) \qquad \frac{d^{k-1}\widehat{F}}{d\xi^{k-1}}(0) = \frac{i^{k-1}}{2\pi k} \int_{\mathbb{R}} u_0(x)^k \, dx, \quad k = 1, 2, 3, \ldots.$$

Recalling the constants L and M defined by (3.4) and (3.5) respectively, it is easy to obtain the estimate

$$(3.13) \qquad \left| \frac{1}{(k-1)!} \frac{d^{k-1}\widehat{F}}{d\xi^{k-1}}(0) \right| \leq \frac{ML^{k-1}}{k!},$$

from which it follows that $\widehat{F}(\xi)$ is an entire function and hence is equal to its Taylor series about $\xi = 0$:

$$\widehat{F}(\xi) = \sum_{k=1}^{\infty} \frac{1}{(k-1)!} \frac{d^{k-1}\widehat{F}}{d\xi^{k-1}}(0)\xi^{k-1} = \sum_{k=1}^{\infty} \frac{(i\xi)^{k-1}}{2\pi k!} \int_{\mathbb{R}} u_0(x)^k \, dx$$

(3.14)

$$= \frac{1}{2\pi i \xi} \sum_{k=1}^{\infty} \int_{\mathbb{R}} \frac{[i\xi u_0(x)]^k}{k!} \, dx.$$

The combined sum and integral is absolutely convergent for all $\xi \in \mathbb{C}$, so the order of operations may be reversed:

$$\widehat{F}(\xi) = \frac{1}{2\pi i \xi} \int_{\mathbb{R}} \left(\sum_{k=1}^{\infty} \frac{[i\xi u_0(x)]^k}{k!} \right) dx$$

(3.15)

$$= \frac{1}{\pi \xi} \int_{\mathbb{R}} e^{i\xi u_0(x)/2} \sin\left(\frac{1}{2} \xi u_0(x) \right) dx.$$

By Fourier inversion,

$$F(\lambda) = \frac{1}{2\pi} \lim_{R \uparrow \infty} \int_{-R}^{+R} \widehat{F}(\xi) e^{i\xi\lambda} \, d\xi$$

(3.16)

$$= \lim_{R \uparrow \infty} \int_{-R}^{+R} \int_{\mathbb{R}} \frac{e^{i\xi(\lambda + u_0(x)/2)}}{2\pi^2 \xi} \sin\left(\frac{1}{2} \xi u_0(x) \right) dx \, d\xi.$$

Applying Fubini's theorem to reverse the order of integration and then passing to the limit $R \uparrow \infty$, we can evaluate the integral over ξ as the indicator function of an interval:

(3.17)
$$F(\lambda) = \frac{1}{2\pi} \int_{\mathbb{R}} \chi_{[-u_0(x),0]}(\lambda) dx.$$

This formula shows that $F(\lambda) \equiv 0$ for $\lambda > 0$ or $\lambda < -L$. By a "layer cake" argument we may simplify this formula for $\lambda \in (-L, 0)$ as

(3.18)
$$F(\lambda) = \frac{1}{2\pi} \int_{\{x \in \mathbb{R}, \, u_0(x) > -\lambda\}} dx, \quad -L < \lambda < 0.$$

This is Matsuno's remarkable result. We have presented Matsuno's method in some detail because it turns out that a key calculation in our analysis of the *inverse* problem in the zero-dispersion limit reduces to almost the same steps, as we will see shortly. This is worth emphasizing because it provides further evidence that for the BO equation, scattering and inverse scattering are mathematically very similar operations.

Matsuno's formula (3.18) could perhaps be compared with the Weyl formula that gives the density of eigenvalues of the self-adjoint Schrödinger operator in the zero-dispersion theory of the KdV equation [20]; aside from a constant factor the Weyl formula replaces the unit integrand in (3.18) with the positive square root $\sqrt{u_0(x) + \lambda}$. So, if u_0 is an admissible initial condition with tail decay parameter $q > 2$, then the Weyl density is finite at $\lambda = -L$ and $\lambda = 0$ while the Matsuno density vanishes at $\lambda = -L$ and blows up as $\lambda \uparrow 0$. This suggests that a typical initial condition will generate a much larger number of broad, low-amplitude, and slowly moving solitons under the BO equation than under the KdV equation when the dispersion is small.

While quite severe hypotheses on u_0 are required for all of the arguments to go through, formula (3.18) makes sense under much weaker conditions. In particular, we may interpret (3.18) for an admissible initial condition, in which case we may express $F(\lambda)$ directly in terms of the turning points $x_\pm(\lambda)$:

$$(3.19) \qquad F(\lambda) := \frac{1}{2\pi}(x_+(\lambda) - x_-(\lambda)), \quad -L \leq \lambda < 0.$$

We take (3.19) as a definition valid for admissible initial conditions u_0. Note that

$$(3.20) \qquad \int_{-L}^{0} F(\lambda)d\lambda = M,$$

where the mass M is defined by (3.5).

3.3 Formula for Phase Constants

The WKB methods applied by Lax and Levermore [20] to the analysis of the Schrödinger equation in the forward problem for the zero-dispersion limit of the KdV equation were sufficiently powerful to provide asymptotic formulae for both the discrete spectrum (via Bohr-Sommerfeld quantization of the Weyl formula that is the analogue in the KdV theory of the function $F(\lambda)$ obtained by Matsuno) and also for the "norming constants" that in the KdV theory are the analogues of the phase constants $\{\gamma_n\}_{n=1}^{N}$ in the BO theory. However, we have not found a way to apply these methods to the nonlocal operator $\mathcal{L}$, and unfortunately Matsuno's method does not provide approximations of the phase constants $\{\gamma_n\}_{n=1}^{N}$ since they do not enter into the trace formulae.

Our contribution to the theory of the spectral analysis of the nonlocal operator $\mathcal{L}$ in the zero-dispersion limit is to provide a new asymptotic formula for the phase constants. It is difficult to motivate the formula as it arises from the analysis of the inverse problem that we will describe in the next section, but it is nonetheless quite easy to present. If $\lambda < 0$ is an eigenvalue of $\mathcal{L}$ with potential u given by an admissible initial condition u_0, then our approximation to the corresponding phase constant is given in terms of the turning points $x_\pm(\lambda)$ as follows:

$$(3.21) \qquad \gamma \approx \gamma(\lambda) := -\frac{1}{2}(x_+(\lambda) + x_-(\lambda)), \quad -L \leq \lambda < 0.$$

Remark 3.2. Our choice of $\gamma(\lambda)$ in terms of u_0 is specifically designed to ensure the convergence of $\tilde{u}_\varepsilon(x, t)$ (to be defined precisely in Definition 3.3 below) at $t = 0$ to the given ε-independent initial condition u_0.

3.4 Modification of the Cauchy Data

Based on the above considerations, we may now make very precise definitions of formal (not rigorously justified) approximations of the scattering data corresponding to an admissible condition u_0. The first approximation is to neglect the reflection coefficient by setting

$$\tilde{\beta}(\lambda) := 0, \quad \lambda > 0. \tag{3.22}$$

Next we define the exact number of approximate eigenvalues (hopefully also the approximate number of exact eigenvalues) by setting

$$N(\varepsilon) := \left\lfloor \frac{M}{\varepsilon} \right\rfloor, \tag{3.23}$$

which in particular implies that

$$\lim_{\varepsilon \downarrow 0} \varepsilon N(\varepsilon) = M. \tag{3.24}$$

Then we define approximations to the eigenvalues themselves as an ordered set of numbers $\{\tilde{\lambda}_n\}_{n=1}^{N(\varepsilon)} \subset (-L, 0)$ obtained by quantizing the Matsuno eigenvalue density given by (3.19):

$$\int_{-L}^{\tilde{\lambda}_n} F(\lambda) d\lambda = \varepsilon \left(n - \frac{1}{2} \right), \quad n = 1, 2, \ldots, N(\varepsilon). \tag{3.25}$$

Finally, we define approximations to the corresponding phase constants as numbers $\{\tilde{\gamma}_n\}_{n=1}^{N(\varepsilon)}$ given precisely by

$$\tilde{\gamma}_n := \gamma(\tilde{\lambda}_n), \quad n = 1, 2, \ldots, N(\varepsilon). \tag{3.26}$$

where $\gamma(\cdot)$ is defined by (3.21).

Now in our analysis of the Cauchy problem for the BO equation with admissible initial data u_0 we take a sideways step that is not a priori justified: we simply replace the true solution $u_\varepsilon(x, t)$ of the Cauchy problem with a family $\tilde{u}_\varepsilon(x, t)$ of exact solutions of the BO equation (1.1) with the property that for each $\varepsilon > 0$ the scattering data for $\tilde{u}_\varepsilon(x, t)$ at time $t = 0$ is *exactly* the approximate scattering data just defined. This step was also an important part of the method of Lax and Levermore [20]. We formalize this modification of the initial data in the following definition.

DEFINITION 3.3 Let u_0 be an admissible initial condition. Then, by $\tilde{u}_\varepsilon(x, t)$ we mean the exact solution of the BO equation (1.1) given for each $\varepsilon > 0$ by the reflectionless inverse scattering formula

$$\tilde{u}_\varepsilon(x, t) := 2\varepsilon \frac{\partial}{\partial x} \text{Im}\{\log(\tilde{\tau}_\varepsilon(x, t))\}, \tag{3.27}$$

where

$$\tilde{\tau}_\varepsilon(x,t) := \det(\mathbb{I} + i\varepsilon^{-1}\tilde{\mathbf{A}}_\varepsilon) \tag{3.28}$$

and where $\tilde{\mathbf{A}}_\varepsilon = \tilde{\mathbf{A}}_\varepsilon(x,t)$ is an $N(\varepsilon) \times N(\varepsilon)$ Hermitian matrix with elements

$$\tilde{A}_{nm} := \frac{2i\varepsilon\sqrt{\tilde{\lambda}_n\tilde{\lambda}_m}}{\tilde{\lambda}_n - \tilde{\lambda}_m}, \quad n \neq m, \tag{3.29}$$

and

$$\tilde{A}_{nn} := -2\tilde{\lambda}_n(x + 2\tilde{\lambda}_n t + \tilde{\gamma}_n) = -2\tilde{\lambda}_n(x + 2\tilde{\lambda}_n + \gamma(\tilde{\lambda}_n)). \tag{3.30}$$

Here the number $N(\varepsilon)$ is defined by (3.23), and the components of the scattering data $\{\tilde{\lambda}_n\}_{n=1}^{N(\varepsilon)}$ and $\{\tilde{\gamma}_n\}_{n=1}^{N(\varepsilon)}$ are given explicitly by (3.25) and (3.26), respectively.

While it is not the case that $\tilde{u}_\varepsilon(x,0) = u_0(x)$ in general, the relevance of this definition in connection with the Cauchy problem with initial condition u_0 is a consequence of Corollary 1.2, which guarantees convergence in the mean square sense of $\tilde{u}_\varepsilon(\cdot,0)$ to $u_0(\cdot)$ as $\varepsilon \downarrow 0$.

The proof of Theorem 1.1 will be given below in Section 4. Before embarking on that we note that Definition 3.1 implies a number of properties of the functions F and γ that will be useful later, so we take the opportunity to record these here. Note that F and γ will frequently occur in the context of the following functions:

$$D(\lambda; x,t) := -2\lambda(x + 2\lambda t + \gamma(\lambda)), \quad -L < \lambda < 0, \tag{3.31}$$

and

$$\varphi(\lambda) := \sqrt{-\lambda F(\lambda)}, \quad -L < \lambda < 0. \tag{3.32}$$

LEMMA 3.4 *Let u_0 be an admissible initial condition with decay exponent $q > 1$, and let $F : [-L,0) \to \mathbb{R}$ be defined by (3.19) and $\gamma : [-L,0) \to \mathbb{R}$ be defined by (3.21). Then F and γ both belong to $C^1(-L,0)$, and F and F' are strictly positive on this open interval. Also, there exists a sufficiently small constant $\delta > 0$ and positive constants C_{-L} and C_0 such that*

$$\frac{1}{2}C_{-L}\sqrt{L+\lambda} < F(\lambda) < C_{-L}\sqrt{L+\lambda} \tag{3.33}$$

and

$$\frac{1}{4}\frac{C_{-L}}{\sqrt{L+\lambda}} < F'(\lambda) < \frac{1}{2}\frac{C_{-L}}{\sqrt{L+\lambda}} \tag{3.34}$$

both hold for $-L < \lambda < -L + \delta$, while

$$\frac{1}{2}C_0(-\lambda)^{-1/q} < F(\lambda) < C_0(-\lambda)^{-1/q} \tag{3.35}$$

and

$$\frac{1}{2}\frac{C_0}{q}(-\lambda)^{-1/q-1} < F'(\lambda) < \frac{C_0}{q}(-\lambda)^{-1/q-1} \tag{3.36}$$

both hold for $-\delta < \lambda < 0$. Also,

$$(3.37) \qquad |\gamma(\lambda) + x_0| \leq \pi F(\lambda) \quad and \quad |\gamma'(\lambda)| \leq \pi F'(\lambda), \quad -L \leq \lambda < 0,$$

inequalities that when combined with (3.33)–(3.36) imply obvious upper bounds for $|\gamma(\lambda) + x_0|$ and $|\gamma'(\lambda)|$.

In particular, these estimates show that $F(\lambda)$ is integrable, $\varphi(\lambda)$ and $D(\lambda; x, t)$ (and hence also $\lambda x_\pm(\lambda)$) are bounded, and $\varphi(\cdot)$ is Hölder-continuous with exponent $\frac{\sigma}{2}$ with $\sigma = \min(\frac{1}{2}, 1 - \frac{1}{q}) \in (0, 1)$, while $D(\cdot; x, t)$ is Hölder-continuous with exponent σ uniformly for (x, t) in compact sets on $(-L, 0)$.

PROOF: The turning points $x_\pm(\lambda)$ are clearly of class $C^1(-L, 0)$; by definition we have $x_+(\lambda) > x_0 > x_-(\lambda)$ on this open interval, and moreover $x_+(\lambda)$ is strictly increasing while $x_-(\lambda)$ is strictly decreasing on $(-L, 0)$. These facts immediately imply the desired basic smoothness properties of F and γ, and the positivity and monotonicity of F, as well as the inequalities (3.37).

Since $u_0(x_0) = L$ and $u_0'(x_0) = 0$, the $C^2(\mathbb{R})$ function u_0 satisfies

$$(3.38) \qquad \lim_{x \to x_0} \frac{u_0(x) - L}{(x - x_0)^2} = \frac{u_0''(x_0)}{2} \quad and \quad \lim_{x \to x_0} \frac{u_0'(x)}{x - x_0} = u_0''(x_0).$$

Using (3.38) together with the inequality $u_0''(x_0) < 0$, we see from the definition of $x_\pm(\lambda)$ as branches of the inverse function of u_0 that

$$(3.39) \qquad \lim_{\lambda \downarrow -L} \frac{\pm(x_\pm(\lambda) - x_0)}{\sqrt{L + \lambda}} = \sqrt{\frac{2}{-u_0''(x_0)}},$$

$$\lim_{\lambda \downarrow -L} \pm x_\pm'(\lambda)\sqrt{L + \lambda} = \sqrt{\frac{1}{-2u_0''(x_0)}}.$$

Using these relations in (3.19) and (3.21) establishes the existence of the limits

$$(3.40) \qquad \lim_{\lambda \downarrow -L} \frac{F(\lambda)}{\sqrt{L + \lambda}} = \frac{1}{\pi}\sqrt{\frac{2}{-u_0''(x_0)}},$$

$$\lim_{\lambda \downarrow -L} F'(\lambda)\sqrt{L + \lambda} = \frac{1}{2\pi}\sqrt{\frac{2}{-u_0''(x_0)}},$$

which prove the two-sided estimates (3.33) and (3.34).

Next, note that the decay conditions (3.2) and (3.3) for u_0 and its derivative together imply that

$$(3.41) \qquad \lim_{\lambda \uparrow 0} x_\pm(\lambda)(-\lambda)^{1/q} = \pm\left(\mp\frac{C_\pm}{q}\right)^{1/q},$$

$$\lim_{\lambda \uparrow 0} x_\pm'(-\lambda)^{1/q+1} = \pm\frac{1}{q}\left(\mp\frac{C_\pm}{q}\right)^{1/q},$$

where $\mp C_\pm$ are the positive constants in (3.2) and (3.3). It follows from (3.19) that

$$(3.42) \qquad \lim_{\lambda \uparrow 0} F(\lambda)(-\lambda)^{1/q} = \frac{1}{2\pi}\left[\left(-\frac{C_+}{q}\right)^{1/q} + \left(\frac{C_-}{q}\right)^{1/q}\right],$$

which proves (3.35) and

$$(3.43) \qquad \lim_{\lambda \uparrow 0} F'(\lambda)(-\lambda)^{\frac{1}{q}+1} = \frac{1}{2\pi q}\left[\left(-\frac{C_+}{q}\right)^{1/q} + \left(\frac{C_-}{q}\right)^{1/q}\right],$$

which proves (3.36). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

4 Inverse Scattering Problem in the Zero-Dispersion Limit

In this section, we provide the proofs of Theorem 1.1 and of Corollary 1.2.

4.1 Basic Strategy: Outline of Proof of Theorem 1.1

According to Definition 3.3, $\tilde{u}_\varepsilon(x,t)$ is expressed in terms of the determinant $\tilde{\tau}_\varepsilon$ as follows:

$$(4.1) \qquad \tilde{u}_\varepsilon(x,t) = \frac{\partial \tilde{U}_\varepsilon}{\partial x}(x,t), \quad \tilde{U}_\varepsilon(x,t) = 2\varepsilon \operatorname{Im}\{\log(\tilde{\tau}_\varepsilon(x,t))\}.$$

As the logarithm of a complex-valued quantity is involved, $\tilde{U}_\varepsilon(x,t)$ is only defined modulo $4\pi\varepsilon$ for each (x,t), and naturally one should choose the appropriate branch for each (x,t) to achieve continuity. We do this concretely in equation (4.3) below.

At this very early point our analysis must take a very different path than that followed by Lax and Levermore [20] in their study of the zero-dispersion limit for the KdV equation. Indeed, the expansion of $\tilde{\tau}_\varepsilon$ in principal minors that is at the heart of the Lax-Levermore method would be a poor choice in this situation. One reason for this is simply that the principal-minors expansion of $\tilde{\tau}_\varepsilon(x,t)$ consists of complex-valued terms of indefinite phase, so the sum cannot be easily estimated by its largest term. But a more important reason is that formula (4.1) for $\tilde{U}_\varepsilon(x,t)$ involves not $\log(\tilde{\tau}_\varepsilon)$ but rather $\operatorname{Im}\{\log(\tilde{\tau}_\varepsilon)\}$; that is, we require an estimate of the *phase* of the determinant and are not interested in its magnitude.

So instead of expanding the determinant as a sum, we write it as a product. Let $\{\alpha_n\}_{n=1}^{N(\varepsilon)}$ be the real eigenvalues of $\tilde{\mathbf{A}}_\varepsilon(x,t)$. Then the corresponding eigenvalues of $\mathbb{I} + i\varepsilon^{-1}\tilde{\mathbf{A}}_\varepsilon(x,t)$ are of course $\{1 + i\varepsilon^{-1}\alpha_n\}_{n=1}^{N(\varepsilon)}$, so we may expand $\tilde{\tau}_\varepsilon$ as a product over eigenvalues in the form

$$(4.2) \qquad \tilde{\tau}_\varepsilon(x,t) = \prod_{n=1}^{N(\varepsilon)}(1 + i\varepsilon^{-1}\alpha_n).$$

This yields a suggestive formula for $\tilde{U}_\varepsilon(x,t)$ in terms of the eigenvalues of $\tilde{\mathbf{A}}_\varepsilon$:

$$(4.3) \qquad \tilde{U}_\varepsilon(x,t) := \varepsilon \sum_{n=1}^{N(\varepsilon)} 2\arctan(\varepsilon^{-1}\alpha_n).$$
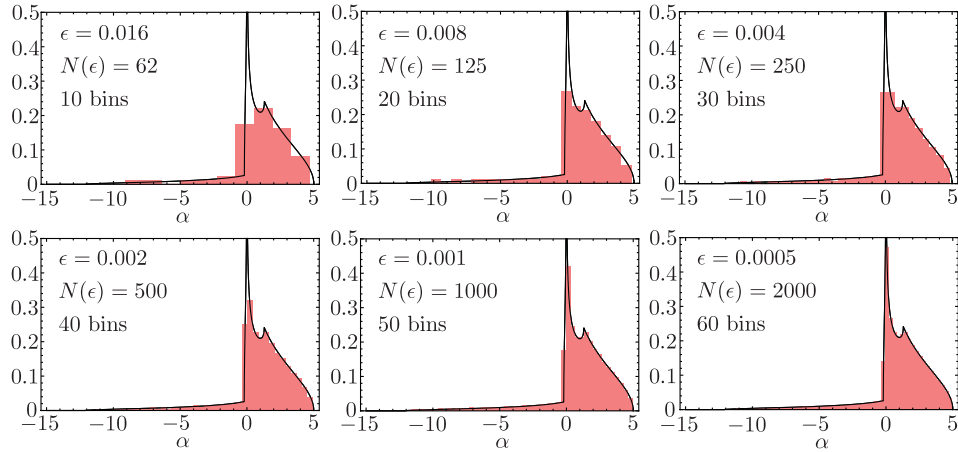
FIGURE 4.1. Histograms of eigenvalues of $\tilde{\mathbf{A}}_\varepsilon$ corresponding to the initial condition $u_0(x) := 2(1 + x^2)^{-1}$, $x = 5$, and $t = 2$, normalized to have total area $M = 1$, compared with the density $G(\alpha; x, t)$ of the limiting absolutely continuous measure μ.

Here $-\pi/2 < \arctan(\cdot) < \pi/2$, so in particular by this definition we have made an unambiguous choice of the branch of the logarithm. This formula seems at first not to be of much use because, unlike the principal minor determinants in the Lax-Levermore method, which can be written explicitly in terms of the matrix elements, the eigenvalues of $\tilde{\mathbf{A}}_\varepsilon$ are only implicitly known. However, numerical experiments suggest that some structure emerges in the limit $\varepsilon \downarrow 0$. Indeed, the plots shown in Figure 4.1 provide good evidence that the normalized (to mass M) counting measures μ_ε given for $\varepsilon > 0$ by

$$(4.4) \qquad \mu_\varepsilon := \frac{M}{N(\varepsilon)} \sum_{n=1}^{N(\varepsilon)} \delta_{\alpha_n}, \quad \{\alpha_n\}_{n=1}^{N(\varepsilon)} \text{ eigenvalues of } \tilde{\mathbf{A}}_\varepsilon,$$

might converge in some sense to a measure μ having a density $G(\alpha; x, t)$. This convergence suggests further that formula (4.3) could be interpreted as a Riemann sum for the integral of $\pi \operatorname{sgn}(\alpha)$ (the pointwise limit as $\varepsilon \downarrow 0$ of the summand) against the limiting measure μ. We will prove that indeed $\tilde{U}_\varepsilon(x, t)$ converges, uniformly with respect to x and t in compact sets, to a limit function $U(x, t)$ given by such an integral in the limit $\varepsilon \downarrow 0$.

To obtain an effective formula for $U(x, t)$, we need to analyze the asymptotic behavior of the measures μ_ε. This part of our analysis is modeled after the work of Wigner [33, 34] on the statistical distribution of eigenvalues of random Hermitian matrices with independent and identically distributed matrix elements. Like Wigner, we use the method of moments because while the measures themselves

are not easy to express in terms of the matrix elements, their moments are. Indeed, the moments are expressed in terms of traces of powers of $\tilde{\mathbf{A}}_\varepsilon$ in the following way:

$$(4.5) \qquad \int_{\mathbb{R}} \alpha^p \, d\mu_\varepsilon(\alpha) = \frac{M}{N(\varepsilon)} \sum_{n=1}^{N(\varepsilon)} \alpha_n^p = \frac{M}{N(\varepsilon)} \text{tr}(\tilde{\mathbf{A}}_\varepsilon^p), \quad p = 0, 1, 2, \dots.$$

We prove the existence of the limit of the right-hand side in equation (4.5) as $\varepsilon \downarrow 0$ for every p using the fact that for small ε the matrix $\tilde{\mathbf{A}}_\varepsilon$ concentrates near the diagonal, where it can be approximated by the product of a diagonal matrix and the Toeplitz matrix corresponding to the symbol $f(\theta) := i(\pi - \theta), 0 < \theta < 2\pi$ (of singular Fisher-Hartwig type due to jump discontinuities). The result of this asymptotic analysis of moments is the following proposition, the proof of which will be given in Section 4.2.

PROPOSITION 4.1 *For each nonnegative integer p,*

$$(4.6) \qquad \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} \alpha^p \, d\mu_\varepsilon(\alpha) = Q_p,$$

with the limit being uniform with respect to (x, t) in any compact set where

$$(4.7) \qquad Q_p := \frac{1}{2\pi(p+1)} \int_{-L}^0 \big[(x + 2\lambda t - x_-(\lambda))^{p+1} \\ - (x + 2\lambda t - x_+(\lambda))^{p+1}\big](-2\lambda)^p \, d\lambda.$$

Given these limiting moments, the next task is to establish the existence of a corresponding limiting measure μ with these moments and to prove the existence of the limit $\tilde{U}_\varepsilon(x, t) \to U(x, t)$. A remarkable feature of this analysis is that the solution of the moment problem for μ is carried out by virtually the same procedure as Matsuno used to obtain the function $F(\lambda)$ from u_0 (see Section 3.2). Our result is the following proposition, which will be proved in all details in Section 4.3.

PROPOSITION 4.2 *Uniformly for (x, t) in compact sets,*

$$(4.8) \qquad \lim_{\varepsilon \downarrow 0} \tilde{U}_\varepsilon(x, t) = U(x, t),$$

where

$$(4.9) \qquad U(x, t) := \int_{\mathbb{R}} \pi \, \text{sgn}(\alpha) d\mu(\alpha)$$

and μ is an absolutely continuous measure of mass M with density $G(\alpha; x, t)$ given by

$$(4.10) \quad G(\alpha; x, t) := -\frac{1}{4\pi} \int_{-L}^0 \chi_{[-2\lambda(x+2\lambda t - x_+(\lambda)), -2\lambda(x+2\lambda t - x_-(\lambda))]}(\alpha) \frac{d\lambda}{\lambda}.$$

Here $\chi_{[a,b]}(z)$ denotes the indicator function of the interval $[a, b]$.
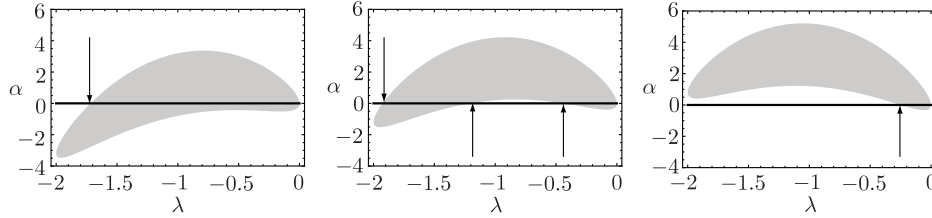
FIGURE 4.2. The region of integration $-2\lambda(x + 2\lambda t - x_+(\lambda)) < \alpha < -2\lambda(x + 2\lambda t - x_-(\lambda))$ for $u_0(x) = 2(1 + x^2)^{-1}$ with $t = 0.7$. Left: $x = 2$ (to the left of the oscillatory region for $u_\varepsilon(x, t)$). Center: $x = 2.5$ (within the oscillatory region for $u_\varepsilon(x, t)$). Right: $x = 3$ (to the right of the oscillatory region for $u_\varepsilon(x, t)$). The line $\alpha = 0$ of discontinuity of the integrand is superimposed, and the intersections of the boundary with this line are indicated with arrows.

The limiting measure μ is the closest analogue in the zero-dispersion theory of the BO equation of the equilibrium (or extremal) measure arising in the Lax-Levermore theory of the KdV equation. But a significant difference is that in this case the measure μ is specified *explicitly* rather than implicitly as the solution of a variational problem.

The region of integration in the double integral obtained by combining (4.10) with (4.9) is illustrated for three different values of (x, t) in Figure 4.2. The points where the boundary curves of this region intersect the line $\alpha = 0$ (where the integrand is discontinuous) obviously will play an important role in the differentiation of $U(x, t)$ with respect to x. Moreover, these intersection points correspond (simply by changing the sign) to the branches of the multivalued solution of Burgers' equation with initial data u_0. This explains their appearance in the formula for the weak limit of $u_\varepsilon(x, t)$. All details of this calculation will be given in Section 4.4, which will complete the proof of Theorem 1.1. Ingredients from this calculation will also set the stage for the proof of Corollary 1.2, which will be given in Section 4.5.

4.2 Asymptotics of Traces of Powers of $\tilde{A}_\varepsilon$: Proof of Proposition 4.1

Definition (3.25) implies that where $F(\lambda)$ is bounded and bounded away from zero, the numbers $\{\tilde{\lambda}_n\}_{n=1}^{N(\varepsilon)}$ are locally nearly equally spaced, but they are more dilute near the "soft edge" of the spectrum $\lambda = -L$ and denser near the "hard edge" of the spectrum $\lambda = 0$. Taking into account the soft edge behavior, we may obtain a uniform estimate:

LEMMA 4.3 *There is a constant $C_\lambda > 0$ independent of ε such that*

$$(4.11) \qquad |\tilde{\lambda}_n - \tilde{\lambda}_m| \leq C_\lambda \varepsilon^{2/3} |n - m|^{2/3}$$

holds for all n and m between 1 and $N(\varepsilon)$.

PROOF: Since F is a monotone increasing function with $F(-L) = 0$, it is bounded away from zero except in a right neighborhood of $\lambda = -L$. Using the

lower bound given in (3.33) from Lemma 3.4, we obtain a lower bound $F(\lambda) \geq C\sqrt{L+\lambda}$ valid uniformly for $-L < \lambda < 0$ with $0 < C \leq C_{-L}/2$. Then, using definition (3.25), we have (assuming $n \geq m$ without loss of generality)

$$
\varepsilon|n - m| = \int_{\tilde{\lambda}_m}^{\tilde{\lambda}_n} F(\lambda) d\lambda \geq C \int_{\tilde{\lambda}_m}^{\tilde{\lambda}_n} \sqrt{L+\lambda}\, d\lambda
$$

(4.12)

$$
\geq C \int_0^{|\tilde{\lambda}_n - \tilde{\lambda}_m|} \sqrt{\xi}\, d\xi = \frac{2C}{3}|\tilde{\lambda}_n - \tilde{\lambda}_m|^{3/2},
$$

so the desired inequality follows with $C_\lambda := (2C/3)^{-2/3}$. $\qquad\square$

We decompose the matrix $\tilde{\mathbf{A}}_\varepsilon$ into a sum $\tilde{\mathbf{A}}_\varepsilon = \mathbf{D} + \mathbf{H}$ of its diagonal part

$$
(4.13) \qquad \mathbf{D} := \mathrm{diag}(D_1, D_2, \ldots, D_{N(\varepsilon)}), \quad D_k := D(\tilde{\lambda}_k; x, t),
$$

where $D(\lambda; x, t)$ is defined by (3.31), and its off-diagonal part $\mathbf{H}$ whose matrix elements are given by

$$
(4.14) \qquad H_{nm} = \frac{2i\varepsilon\sqrt{\tilde{\lambda}_n \tilde{\lambda}_m}}{\tilde{\lambda}_n - \tilde{\lambda}_m} \quad \text{for } n \neq m \text{ and } H_{nn} = 0.
$$

We also will soon need the quantities $\{\varphi_n\}_{n=1}^{N(\varepsilon)}$ defined by

$$
(4.15) \qquad \varphi_n := \varphi(\tilde{\lambda}_n), \quad n = 1, 2, \ldots, N(\varepsilon),
$$

where $\varphi(\lambda)$ is given by (3.32).

LEMMA 4.4 *There is a constant $C_\varphi > 0$ and for each $R > 0$ there is a constant $C_{D,R} > 0$ such that*

$$
(4.16) \qquad |\varphi_n| \leq C_\varphi
$$

and

$$
(4.17) \qquad \sup_{x^2+t^2 \leq R^2} |D_n| \leq C_{D,R};
$$

both hold for all $\varepsilon > 0$ and all n between 1 and $N(\varepsilon)$. Also,

$$
(4.18) \qquad |\varphi_n - \varphi_m| \leq C_\varphi \varepsilon^{\sigma/3}|n - m|^{\sigma/3}
$$

and

$$
(4.19) \qquad \sup_{x^2+t^2 \leq R^2} |D_n - D_m| \leq C_{D,R}\varepsilon^{\sigma/3}|n - m|^{\sigma/3};
$$

both hold for all $\varepsilon > 0$ and for all n and m between 1 and $N(\varepsilon)$. Here σ is the positive Hölder exponent of Lemma 3.4.

PROOF: This is an easy consequence of the Hölder continuity of $\varphi(\cdot)$ and $D(\cdot\,; x, t)$ guaranteed by Lemma 3.4, and of the spacing estimate for the approximate eigenvalues $\{\tilde{\lambda}_k\}_{k=1}^{N(\varepsilon)}$ given in Lemma 4.3. In fact, since D is Hölder-continuous with exponent σ while φ has exponent $\sigma/2$, the most natural bound for $|D_n - D_m|$ is proportional to $\varepsilon^{2\sigma/3}|n - m|^{2\sigma/3}$, and to obtain (4.19) we use the fact that $\varepsilon|n - m| \leq 2\varepsilon N(\varepsilon)$ is uniformly bounded to reduce the exponent to $\sigma/3$. $\qquad\square$

LEMMA 4.5 *There is a constant $C_H > 0$ such that*

$$(4.20) \qquad\qquad |(n - m)H_{nm}| \leq C_H$$

and

$$(4.21) \qquad |(n - m)H_{nm} - 2i\varphi_n\varphi_m| \leq C_H \varepsilon^{\sigma/3}|n - m|^{\sigma/3}$$

both hold for all $\varepsilon > 0$ and all $n \neq m$ between 1 and $N(\varepsilon)$. Again, $\sigma > 0$ is the Hölder exponent of Lemma 3.4.

PROOF: Suppose without loss of generality that $n > m$, implying that $\tilde{\lambda}_m < \tilde{\lambda}_n < 0$. Then

$$
\begin{aligned}
-i(n - m)H_{nm} &= 2\sqrt{-\tilde{\lambda}_n}\sqrt{-\tilde{\lambda}_m}\,\frac{\varepsilon(n - m)}{\tilde{\lambda}_n - \tilde{\lambda}_m} \\
(4.22) \qquad &\leq ([-\tilde{\lambda}_n] + [-\tilde{\lambda}_m])\frac{\varepsilon(n - m)}{\tilde{\lambda}_n - \tilde{\lambda}_m} \\
&= \varepsilon(n - m) - 2\tilde{\lambda}_n\frac{\varepsilon(n - m)}{\tilde{\lambda}_n - \tilde{\lambda}_m}.
\end{aligned}
$$

Now, recalling definition (3.25) of the numbers $\{\tilde{\lambda}_k\}_{k=1}^{N(\varepsilon)}$ and applying the mean value theorem, we may write the latter difference quotient as $F(\xi)$ for some ξ with $\tilde{\lambda}_m \leq \xi \leq \tilde{\lambda}_n$, and since F is increasing we have $F(\xi) \leq F(\tilde{\lambda}_n)$, so

$$(4.23) \qquad -i(n - m)H_{nm} \leq 2\varepsilon(n - m) - 2\tilde{\lambda}_n F(\tilde{\lambda}_n) = 2\varepsilon(n - m) + 2\varphi_n^2,$$

where we have also replaced $\varepsilon(n - m)$ with $2\varepsilon(n - m)$.

On the other hand, we may write

$$(4.24) \quad -i(n - m)H_{nm} = 2\varepsilon(n - m)\frac{\sqrt{-\tilde{\lambda}_n}\sqrt{-\tilde{\lambda}_m} + \tilde{\lambda}_m}{\tilde{\lambda}_n - \tilde{\lambda}_m} - 2\tilde{\lambda}_m\frac{\varepsilon(n - m)}{\tilde{\lambda}_n - \tilde{\lambda}_m}.$$

Again the difference quotient may be replaced by $F(\xi) \geq F(\tilde{\lambda}_m)$, and since

$$(4.25) \qquad \frac{\sqrt{-\tilde{\lambda}_n}\sqrt{-\tilde{\lambda}_m}}{\tilde{\lambda}_n - \tilde{\lambda}_m} = -\frac{\sqrt{-\tilde{\lambda}_m}}{\sqrt{\tilde{\lambda}_n} + \sqrt{\tilde{\lambda}_m}} \geq -1,$$

we obtain

$$(4.26) \quad -i(n - m)H_{nm} \geq -2\varepsilon(n - m) - 2\tilde{\lambda}_m F(\tilde{\lambda}_m) = -2\varepsilon(n - m) + 2\varphi_m^2.$$

Combining (4.23) and (4.26) gives

$$(4.27) \qquad |(n-m)H_{nm} - 2i\varphi_n\varphi_m| \le 2\varepsilon|n-m| + 2\max\{\varphi_n, \varphi_m\}|\varphi_n - \varphi_m|,$$

and then applying Lemma 4.4 we obtain

$$(4.28) \qquad |(n-m)H_{nm} - 2i\varphi_n\varphi_m| \le 2\varepsilon|n-m| + 2C_\varphi^2\varepsilon^{\sigma/3}|n-m|^{\sigma/3}.$$

Now, $0 \le \varepsilon|n-m| \le 2\varepsilon N(\varepsilon)$, and this upper bound has a limit as $\varepsilon \downarrow 0$, so $\varepsilon|n-m|$ is nonnegative and bounded. Since $\sigma \le 3$ we have therefore proved (4.21). Since $\varphi_n\varphi_m$ and $\varepsilon|n-m|$ are bounded, (4.20) then follows from (4.21). □

For any nonnegative integer power p, the pth moment of the measure μ_ε can be written in terms of $\mathbf{D}$ and $\mathbf{H}$ with the use of (4.5):

$$(4.29) \qquad \int_{\mathbb{R}} \alpha^p \, d\mu_\varepsilon(\alpha) = \sum_{j=0}^{p} Z_{pj},$$

where Z_{pj} contains the contribution to the trace coming from products of matrices involving exactly j factors of $\mathbf{H}$:

$$(4.30) \qquad Z_{pj} := \frac{M}{N(\varepsilon)} \sum_{\substack{d_1+d_2+\cdots+d_s=p-j \\ h_1+h_2+\cdots+h_s=j}} \operatorname{tr}(\mathbf{D}^{d_1}\mathbf{H}^{h_1}\mathbf{D}^{d_2}\mathbf{H}^{h_2}\cdots\mathbf{D}^{d_s}\mathbf{H}^{h_s}),$$

and where $d_1 \ge 0$ and $h_s \ge 0$, while $d_k > 0$ for $2 \le k \le s$ and $h_k > 0$ for $1 \le k \le s-1$. Since p is a fixed number, it will suffice to compute the limit of Z_{pj} as $\varepsilon \downarrow 0$ for $j = 0, 1, \ldots, p$. Actually, it will be enough to consider even values of j as the following result shows.

LEMMA 4.6 *If j is an odd number, then $Z_{pj} = 0$.*

PROOF: Since $\operatorname{tr}(\mathbf{M}) = \operatorname{tr}(\mathbf{M}^\mathsf{T})$ for all square matrices $\mathbf{A}$,

$$(4.31) \qquad \begin{aligned} \frac{N(\varepsilon)}{M}Z_{pj} &= \sum_{\substack{d_1+d_2+\cdots+d_s=p-j \\ h_1+h_2+\cdots+h_s=j}} \operatorname{tr}((\mathbf{D}^{d_1}\mathbf{H}^{h_1}\cdots\mathbf{D}^{d_s}\mathbf{H}^{h_s})^\mathsf{T}) \\ &= (-1)^j \sum_{\substack{d_1+d_2+\cdots+d_s=p-j \\ h_1+h_2+\cdots+h_s=j}} \operatorname{tr}(\mathbf{H}^{h_s}\mathbf{D}^{d_s}\cdots\mathbf{H}^{h_1}\mathbf{D}^{d_1}) \end{aligned}$$

where in the second line we have used the facts that $\mathbf{D}^\mathsf{T} = \mathbf{D}$ and $\mathbf{H}^\mathsf{T} = -\mathbf{H}$. By relabeling the terms in the sum we therefore obtain

$$(4.32) \qquad \frac{N(\varepsilon)}{M}Z_{pj} = (-1)^j \frac{N(\varepsilon)}{M}Z_{pj}.$$

Since $N(\varepsilon) > 0$ and $M < \infty$, the desired result follows. □

An important role will be played below by the Toeplitz (discrete convolution) operator $\mathcal{T}_f : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ defined by

$$(4.33) \qquad (\mathcal{T}_f c)_n := \sum_{m \in \mathbb{Z}} f_{n-m} c_m, \quad \{c_m\}_{m \in \mathbb{Z}} \in \ell^2(\mathbb{Z}),$$

where $\{f_n\}_{n \in \mathbb{Z}} \in \ell^2(\mathbb{Z})$ is the sequence

$$(4.34) \qquad f_n := \begin{cases} n^{-1}, & n \neq 0, \\ 0, & n = 0. \end{cases}$$

LEMMA 4.7 *For any even positive integer j, we have*

$$(4.35) \qquad \sum_{n_2,\ldots,n_j \in \mathbb{Z}} f_{-n_2} \Big[\prod_{\ell=2}^{j-1} f_{n_\ell - n_{\ell+1}} \Big] f_{n_j} = \frac{(i\pi)^j}{j+1},$$

where the $(j-1)$-fold infinite sum converges absolutely.

PROOF: Note that since $\{f_n\}_{n \in \mathbb{Z}} \in \ell^2(\mathbb{Z})$, $\{g_n\}_{n \in \mathbb{Z}} \in \ell^2(\mathbb{Z})$ as well, where $g_n := |f_n|$ for all $n \in \mathbb{Z}$. The corresponding Fourier series converge in the mean square sense to functions $f(\cdot)$ and $g(\cdot)$ in $L^2(0, 2\pi)$:

$$(4.36) \qquad f(\theta) := \sum_{n \in \mathbb{Z}} f_n e^{in\theta} = i(\pi - \theta), \quad 0 < \theta < 2\pi,$$

and

$$(4.37) \qquad g(\theta) := \sum_{n \in \mathbb{Z}} g_n e^{in\theta} = -\log\big(2(1 - \cos(\theta))\big), \quad 0 < \theta < 2\pi.$$

First we establish the absolute convergence of the series on the left-hand side of (4.35). Using (4.33), observe that

$$(4.38) \qquad \sum_{n_2,n_3,\ldots,n_j \in \mathbb{Z}} |f_{-n_2}| \Big[\prod_{\ell=2}^{j-1} |f_{n_\ell - n_{\ell+1}}| \Big] |f_{n_j}| = (\mathcal{T}_g^{j-1} g)_0$$

where $\mathcal{T}_g$ is the Toeplitz operator associated with the sequence $\{g_n\}_{n \in \mathbb{Z}}$. Now, $g(\cdot)$ has a logarithmic singularity at $\theta = 0 \pmod{2\pi}$, but this is sufficiently mild that $g(\cdot)^m \in L^2(0, 2\pi) \subset L^1(0, 2\pi)$ for any positive integer power m. Now for any function $k(\cdot) \in L^2(0, 2\pi)$, the corresponding Fourier coefficients are

$$(4.39) \qquad k_n := \frac{1}{2\pi} \int_0^{2\pi} k(\theta) e^{-in\theta} \, d\theta,$$

so in particular we see that $(\mathcal{T}_g^{j-1} g)_0$ is the average value of the function whose Fourier coefficients are $\{(\mathcal{T}_g^{j-1} g)_n\}_{n \in \mathbb{Z}}$. But by the convolution theorem

$$(4.40) \qquad w_n := \sum_{m \in \mathbb{Z}} u_{n-m} v_m \quad \Longleftrightarrow \quad w(\theta) = u(\theta) v(\theta),$$

so it follows that

$$(4.41) \qquad (\mathcal{T}_g^{j-1} g)_0 = \frac{1}{2\pi} \int_0^{2\pi} g(\theta)^j \, d\theta,$$

which is finite because $g(\cdot)^j \in L^1[0, 2\pi]$.

Now we find the exact value of the $(j-1)$-fold infinite sum by the same reasoning:

$$(4.42) \qquad \sum_{n_2, \ldots, n_j \in \mathbb{Z}} f_{-n_2} \left[\prod_{\ell=2}^{j-1} f_{n_\ell - n_{\ell+1}} \right] f_{n_j} = (\mathcal{T}_f^{j-1} f)_0 = \frac{1}{2\pi} \int_0^{2\pi} f(\theta)^j \, d\theta,$$

and by direct calculation using (4.36),

$$(4.43) \qquad \frac{1}{2\pi} \int_0^{2\pi} f(\theta)^j \, d\theta = \frac{1}{2\pi} \int_0^{2\pi} [i(\pi - \theta)]^j \, d\theta = \frac{(i\pi)^j}{j+1}$$

for j even (the integral vanishes by symmetry for j odd). $\qquad \square$

Now we consider separately each of the terms in Z_{pj} for j even.

LEMMA 4.8 *If j is an even number and $h_1 + h_2 + \cdots + h_s = j$ while $d_1 + d_2 + \cdots + d_s = p - j$, then*

$$(4.44) \qquad \lim_{\varepsilon \downarrow 0} \frac{M}{N(\varepsilon)} \operatorname{tr}(\mathbf{D}^{d_1} \mathbf{H}^{h_1} \cdots \mathbf{D}^{d_s} \mathbf{H}^{h_s}) =$$

$$\frac{(2\pi)^j}{j+1} \int_{-L}^0 D(\lambda; x, t)^{p-j} \varphi(\lambda)^{2j} F(\lambda) \, d\lambda,$$

with the limit being uniform with respect to (x, t) in any compact set.

PROOF: Recalling the matrix elements D_n and H_{nm} of $\mathbf{D}$ and $\mathbf{H}$, respectively, we have

$$(4.45) \quad \operatorname{tr}(\mathbf{D}^{d_1} \mathbf{H}^{h_1} \cdots \mathbf{D}^{d_s} \mathbf{H}^{h_s}) = \sum_{a_1, a_2, \ldots, a_j = 1}^{N(\varepsilon)} \left[\prod_{i=1}^j D_{a_i}^{m_i} \right] \left[\prod_{\ell=1}^{j-1} H_{a_\ell a_{\ell+1}} \right] H_{a_j a_1},$$

where the exponents $m_1, m_2, \ldots, m_j$ are given by

$$(4.46) \qquad m_i := \begin{cases} d_1, & i = 1, \\ d_{b+1}, & i = 1 + h_1 + h_2 + \cdots + h_b \text{ for some } 0 < b < s, \\ 0, & \text{otherwise.} \end{cases}$$

Note that $m_1 + m_2 + \cdots + m_j = d_1 + d_2 + \cdots + d_s = p - j$.

Now, the matrix element H_{nm} is relatively small unless $n \approx m$, and this suggests that the j-fold sum in (4.45) should concentrate near the diagonal, where $a_k = a_1$ for all k. Making this precise, given any $r > 0$ we will first show that

$$(4.47) \qquad \lim_{\varepsilon \downarrow 0} Z_{\mathrm{OD}}(\varepsilon) = 0,$$

where

$$(4.48) \qquad Z_{\text{OD}}(\varepsilon) := \frac{M}{N(\varepsilon)} \sum_{\substack{a_1,a_2,\ldots,a_j=1 \\ \exists k:|a_k-a_1|>\varepsilon^{-r}}}^{N(\varepsilon)} \Big[\prod_{i=1}^{j} D_{a_i}^{m_i}\Big]\Big[\prod_{\ell=1}^{j-1} H_{a_\ell a_{\ell+1}}\Big] H_{a_j a_1},$$

with the limit being uniform for (x,t) in compact sets. Indeed, if $x^2 + t^2 \le R^2$, then using (4.17) from Lemma 4.4 and (4.20) from Lemma 4.5, we obtain

$$(4.49) \quad \begin{aligned} |Z_{\text{OD}}(\varepsilon)| &\le \frac{MC_{D,R}^{p-j}C_H^j}{N(\varepsilon)} \sum_{\substack{a_1,a_2,\ldots,a_j=1 \\ \exists k:|a_k-a_1|>\varepsilon^{-r}}}^{N(\varepsilon)} \Big[\prod_{\ell=1}^{j-1} |f_{a_\ell-a_{\ell+1}}|\Big]|f_{a_j-a_1}| \\ &= \frac{MC_{D,R}^{p-j}C_H^j}{N(\varepsilon)} \sum_{a_1=1}^{N(\varepsilon)} \sum_{\substack{a_2,a_3,\ldots,a_j=1 \\ \exists k:|a_k-a_1|>\varepsilon^{-r}}}^{N(\varepsilon)} \Big[\prod_{\ell=1}^{j-1} |f_{a_\ell-a_{\ell+1}}|\Big]|f_{a_j-a_1}| \\ &\le \frac{MC_{D,R}^{p-j}C_H^j}{N(\varepsilon)} \sum_{a_1=1}^{N(\varepsilon)} \sum_{\substack{a_2,a_3,\ldots,a_j\in\mathbb{Z} \\ \exists k:|a_k-a_1|>\varepsilon^{-r}}} \Big[\prod_{\ell=1}^{j-1} |f_{a_\ell-a_{\ell+1}}|\Big]|f_{a_j-a_1}|. \end{aligned}$$

With the inner sum extended over $\mathbb{Z}^{j-1}$ in this way, it becomes independent of the outer sum index a_1 as can be seen by the substitution $n_k = a_k - a_1$ for $k = 2, 3, \ldots, j$. Thus

$$(4.50) \quad |Z_{\text{OD}}(\varepsilon)| \le MC_{D,R}^{p-j}C_H^j \sum_{\substack{n_2,n_3,\ldots,n_j\in\mathbb{Z} \\ \exists k:|n_k|>\varepsilon^{-r}}} |f_{-n_2}|\Big[\prod_{\ell=2}^{j-1} |f_{n_\ell-n_{\ell+1}}|\Big]|f_{n_j}|,$$

and the latter upper bound is of course independent of (x,t) with $x^2 + t^2 \le R^2$ and tends to zero for $r > 0$ by Lemma 4.7.

It follows from (4.47) that

$$(4.51) \qquad \lim_{\varepsilon\downarrow 0} \frac{M}{N(\varepsilon)} \operatorname{tr}(\mathbf{D}^{d_1}\mathbf{H}^{h_1}\cdots\mathbf{D}^{d_s}\mathbf{H}^{h_s}) = \lim_{\varepsilon\downarrow 0} Z_{\text{D}}(\varepsilon)$$

where the diagonally concentrated terms are

$$(4.52) \qquad Z_{\text{D}}(\varepsilon) := \frac{M}{N(\varepsilon)} \sum_{\substack{a_1,a_2,\ldots,a_j=1 \\ \forall k:|a_k-a_1|\le\varepsilon^{-r}}}^{N(\varepsilon)} \Big[\prod_{i=1}^{j} D_{a_i}^{m_i}\Big]\Big[\prod_{\ell=1}^{j-1} H_{a_\ell a_{\ell+1}}\Big] H_{a_j a_1}.$$

We will analyze $Z_{\text{D}}(\varepsilon)$ under the additional assumption that $r < 1$.

The first step is show that if $r < 1$ each occurrence of H_{nm} in (4.52) may be replaced by $2i\varphi_n\varphi_m f_{n-m}$ without affecting the limiting value of $Z_{\text{D}}(\varepsilon)$ as $\varepsilon \downarrow 0$. Indeed, suppose we make this substitution j times in succession, each time keeping

track of the error using Lemma 4.5 along with the estimates (4.16) and (4.17) from Lemma 4.4. Then defining $K_R > 0$ by

$$(4.53) \qquad K_R := C_{D,R}^{p-j} \sum_{k=1}^{j} (2C_\varphi^2)^{k-1} C_H^{j-k+1},$$

we see that, for all j-tuples of integers $a_1, a_2, \ldots, a_j$ between 1 and $N(\varepsilon)$ satisfying $|a_k - a_1| \le \varepsilon^{-r}$ for all k, the following inequality holds true:

$$(4.54) \qquad \begin{aligned} &\left| \left[\prod_{i=1}^{j} D_{a_i}^{m_i} \right] \left[\prod_{\ell=1}^{j-1} H_{a_\ell a_{\ell+1}} \right] H_{a_j a_1} \right. \\ &\left. - (2i)^j \left[\prod_{i=1}^{j} D_{a_i}^{m_i} \varphi_{a_i}^2 \right] \left[\prod_{\ell=1}^{j-1} f_{a_\ell - a_{\ell+1}} \right] f_{a_j - a_1} \right| \\ &\le K_R \varepsilon^{(1-r)\sigma/3} \left[\prod_{\ell=1}^{j-1} |f_{a_\ell - a_{\ell+1}}| \right] |f_{a_j - a_1}|. \end{aligned}$$

Therefore, if we define a modification of $Z_D(\varepsilon)$ by

$$(4.55) \quad Z_D^I(\varepsilon) := \frac{(2i)^j M}{N(\varepsilon)} \sum_{\substack{a_1, a_2, \ldots, a_j = 1 \\ \forall k: |a_k - a_1| \le \varepsilon^{-r}}}^{N(\varepsilon)} \left[\prod_{i=1}^{j} D_{a_i}^{m_i} \varphi_{a_i}^2 \right] \left[\prod_{\ell=1}^{j-1} f_{a_\ell - a_{\ell+1}} \right] f_{a_j - a_1},$$

we have

$$(4.56) \qquad \begin{aligned} &|Z_D(\varepsilon) - Z_D^I(\varepsilon)| \\ &\le \frac{M K_R \varepsilon^{(1-r)\sigma/3}}{N(\varepsilon)} \sum_{\substack{a_1, a_2, \ldots, a_j = 1 \\ \forall k: |a_k - a_1| \le \varepsilon^{-r}}}^{N(\varepsilon)} \left[\prod_{\ell=1}^{j-1} |f_{a_\ell - a_{\ell+1}}| \right] |f_{a_j - a_1}| \\ &\le \frac{M K_R \varepsilon^{(1-r)\sigma/3}}{N(\varepsilon)} \sum_{a_1=1}^{N(\varepsilon)} \left(\sum_{a_2, a_3, \ldots, a_j \in \mathbb{Z}} \left[\prod_{\ell=1}^{j-1} |f_{a_\ell - a_{\ell+1}}| \right] |f_{a_j - a_1}| \right). \end{aligned}$$

By the substitution $n_\ell = a_\ell - a_1$ one sees that the inner sum is independent of a_1, and it is finite by Lemma 4.7. Since $\sigma > 0$ and $r < 1$, we therefore have

$$(4.57) \qquad \lim_{\varepsilon \downarrow 0} Z_D(\varepsilon) = \lim_{\varepsilon \downarrow 0} Z_D^I(\varepsilon)$$

uniformly for $x^2 + t^2 \le R^2$.

The second step is to show that if $r < 1$ we may replace $D_{a_i}^{m_i} \varphi_{a_i}^2$ with $D_{a_1}^{m_i} \varphi_{a_1}^2$ for each i in (4.55) without changing the limiting value of $Z_D^I(\varepsilon)$. Indeed, applying Lemma 4.4 and defining a constant $K_R^I > 0$ by

$$(4.58) \qquad K_R^I := (p + j) C_{D,R}^{p-j} C_\varphi^{2j},$$

we see that for all j-tuples of integers $a_1, a_2, \ldots, a_j$ between 1 and $N(\varepsilon)$ satisfying $|a_k - a_1| \leq \varepsilon^{-r}$ for all k,

$$(4.59) \qquad \left| \prod_{i=1}^{j} D_{a_i}^{m_i} \varphi_{a_i}^2 - D_{a_1}^{p-j} \varphi_{a_1}^{2j} \right| \leq K_R^{\mathrm{I}} \varepsilon^{(1-r)\sigma/3}.$$

Hence, defining a subsequent modification of $Z_{\mathrm{D}}^{\mathrm{I}}(\varepsilon)$ by

$$(4.60) \quad Z_{\mathrm{D}}^{\mathrm{II}}(\varepsilon) :=$$

$$\frac{(2i)^j M}{N(\varepsilon)} \sum_{a_1=1}^{N(\varepsilon)} D_{a_1}^{p-j} \varphi_{a_1}^{2j} \sum_{\substack{a_2,a_3,\ldots,a_j=1 \\ \forall k: |a_k-a_1| \leq \varepsilon^{-r}}}^{N(\varepsilon)} \left[\prod_{\ell=1}^{j-1} f_{a_\ell - a_{\ell+1}} \right] f_{a_j - a_1},$$

we see that

$$|Z_{\mathrm{D}}^{\mathrm{I}}(\varepsilon) - Z_{\mathrm{D}}^{\mathrm{II}}(\varepsilon)|$$

$$(4.61) \qquad \leq \frac{2^j M K_R^{\mathrm{I}} \varepsilon^{(1-r)\sigma/3}}{N(\varepsilon)} \sum_{a_1=1}^{N(\varepsilon)} \sum_{\substack{a_2,a_3,\ldots,a_j=1 \\ \forall k |a_k-a_1| \leq \varepsilon^{-r}}}^{N(\varepsilon)} \left[\prod_{\ell=1}^{j-1} |f_{a_\ell - a_{\ell+1}}| \right] |f_{a_j - a_1}|$$

$$\leq \frac{2^j M K_R^{\mathrm{I}} \varepsilon^{(1-r)\sigma/3}}{N(\varepsilon)} \sum_{a_1=1}^{N(\varepsilon)} \left(\sum_{a_2,a_3,\ldots,a_j \in \mathbb{Z}} \left[\prod_{\ell=1}^{j-1} |f_{a_\ell - a_{\ell+1}}| \right] |f_{a_j - a_1}| \right),$$

and so exactly as before

$$(4.62) \qquad \lim_{\varepsilon \downarrow 0} Z_{\mathrm{D}}^{\mathrm{I}}(\varepsilon) = \lim_{\varepsilon \downarrow 0} Z_{\mathrm{D}}^{\mathrm{II}}(\varepsilon)$$

uniformly for $x^2 + t^2 \leq R^2$.

The third step is to show that if $r < 1$ one may neglect a small fraction of the terms in the outer sum corresponding to $a_1 \leq 1 + \varepsilon^{-r}$ and $a_1 \geq N(\varepsilon) - \varepsilon^{-r}$ without changing the limiting value of $Z_{\mathrm{D}}^{\mathrm{II}}(\varepsilon)$. Indeed, defining the index set

$$(4.63) \qquad S_\varepsilon := \{ n \in \mathbb{Z} : 1 + \varepsilon^{-r} < n < N(\varepsilon) - \varepsilon^{-r} \},$$

and then setting

$$(4.64) \quad Z_{\mathrm{D}}^{\mathrm{III}}(\varepsilon) :=$$

$$\frac{(2i)^j M}{N(\varepsilon)} \sum_{a_1 \in S_\varepsilon} D_{a_1}^{p-j} \varphi_{a_1}^{2j} \sum_{\substack{a_2,a_3,\ldots a_j=1 \\ \forall k: |a_k-a_1| \leq \varepsilon^{-r}}}^{N(\varepsilon)} \left[\prod_{\ell=1}^{j-1} f_{a_\ell - a_{\ell+1}} \right] f_{a_j - a_1},$$

we easily obtain from (4.16) and (4.17) in Lemma 4.4 that

$$|Z_D^{II}(\varepsilon) - Z_D^{III}(\varepsilon)|$$

(4.65)

$$\leq \frac{2^j M C_{D,R}^{p-j} C_\varphi^{2j}}{N(\varepsilon)} \sum_{\substack{a_1=1 \\ a_1 \notin S_\varepsilon}}^{N(\varepsilon)} \sum_{\substack{a_2,a_3,\ldots,a_j=1 \\ \forall k:|a_k-a_1|\leq\varepsilon^{-r}}}^{N(\varepsilon)} \left[\prod_{\ell=1}^{j-1} |f_{a_\ell-a_{\ell+1}}|\right]|f_{a_j-a_1}|$$

$$\leq \frac{2^j M C_{D,R}^{p-j} C_\varphi^{2j}}{N(\varepsilon)} \sum_{\substack{a_1=1 \\ a_1 \notin S_\varepsilon}}^{N(\varepsilon)} \sum_{a_2,a_3,\ldots,a_j\in\mathbb{Z}} \left[\prod_{\ell=1}^{j-1} |f_{a_\ell-a_{\ell+1}}|\right]|f_{a_j-a_1}|.$$

But the inner sum is independent of a_1 and is convergent by Lemma 4.7, and the outer sum has $O(\varepsilon^{-r})$ terms while $N(\varepsilon)$ is proportional to ε^{-1}, so with $r < 1$ we have

(4.66)
$$\lim_{\varepsilon\downarrow 0} Z_D^{II}(\varepsilon) = \lim_{\varepsilon\downarrow 0} Z_D^{III}(\varepsilon)$$

uniformly for $x^2 + t^2 \leq R^2$.

The next step in analyzing $Z_D(\varepsilon)$ is to deal with the inner sum in definition (4.64) of $Z_D^{III}(\varepsilon)$. Taking into account the conditions on a_1 in the outer sum, it is obvious that the conditions $1 \leq a_k \leq N(\varepsilon)$ are superfluous in the inner sum:

(4.67) $Z_D^{III}(\varepsilon) =$

$$\frac{(2i)^j M}{N(\varepsilon)} \sum_{a_1\in S_\varepsilon} D_{a_1}^{p-j} \varphi_{a_1}^{2j} \sum_{\substack{a_2,a_3,\ldots,a_j\in\mathbb{Z} \\ \forall k:|a_k-a_1|\leq\varepsilon^{-r}}} \left[\prod_{\ell=1}^{j-1} f_{a_\ell-a_{\ell+1}}\right] f_{a_j-a_1}.$$

By introducing the differences $n_k = a_k - a_1$ it now becomes clear that the inner sum is independent of a_1:

(4.68) $Z_D^{III}(\varepsilon) =$

$$\frac{(2i)^j M}{N(\varepsilon)} \left(\sum_{a_1\in S_\varepsilon} D_{a_1}^{p-j} \varphi_{a_1}^{2j}\right) \left(\sum_{\substack{n_2,n_3,\ldots,n_j\in\mathbb{Z} \\ \forall k:|n_k|\leq\varepsilon^{-r}}} f_{-n_2}\left[\prod_{\ell=2}^{j-1} f_{a_\ell-a_{\ell+1}}\right] f_{n_j}\right).$$

Now, according to Lemma 4.7, the latter sum has the limit $(i\pi)^j/(j+1)$ as $\varepsilon \downarrow 0$ with $r > 0$, so

(4.69)
$$\lim_{\varepsilon\downarrow 0} Z_D^{III}(\varepsilon) = \lim_{\varepsilon\downarrow 0} Z_D^{IV}(\varepsilon),$$

uniformly for $x^2 + t^2 \leq R^2$, where

(4.70)
$$Z_D^{IV}(\varepsilon) := \frac{(2\pi)^j}{j+1} \cdot \frac{M}{N(\varepsilon)} \sum_{a_1\in S_\varepsilon} D_{a_1}^{p-j} \varphi_{a_1}^{2j}.$$

The final step in the analysis of $Z_D(\varepsilon)$ is simply to evaluate the limit on the right-hand side of (4.69) by recognizing the sum as a Riemann sum for an integral:

$$(4.71) \quad \lim_{\varepsilon \downarrow 0} Z_D(\varepsilon) = \lim_{\varepsilon \downarrow 0} Z_D^{IV}(\varepsilon) = \frac{(2\pi)^j}{j+1} \int_{-L}^{0} D(\lambda; x, t)^{p-j} \varphi(\lambda)^{2j} F(\lambda) d\lambda.$$

Note that since the summand $D_{a_1}^{p-j} \varphi_{a_1}^{2j}$ is polynomial in x and t, the convergence of the Riemann sum is uniform for (x, t) in compact sets. Comparing with (4.51) we see that the proof is complete. $\qquad \square$

Now we may complete the proof of Proposition 4.1. Lemma 4.8 shows that each of the terms in formula (4.30) for Z_{pj} has the same limit as $\varepsilon \downarrow 0$. Therefore, for all even j,

$$(4.72) \quad \begin{aligned} \lim_{\varepsilon \downarrow 0} Z_{pj} &= \sum_{\substack{d_1+d_2+\cdots+d_s=p-j \\ h_1+h_2+\cdots+h_s=j}} \frac{(2\pi)^j}{j+1} \int_{-L}^{0} D(\lambda; x, t)^{p-j} \varphi(\lambda)^{2j} F(\lambda) d\lambda \\ &= \binom{p}{j} \frac{(2\pi)^j}{j+1} \int_{-L}^{0} D(\lambda; x, t)^{p-j} \varphi(\lambda)^{2j} F(\lambda) d\lambda. \end{aligned}$$

Combining this result with Lemma 4.6 and formula (4.29) for the pth moment, we obtain

$$(4.73) \quad \begin{aligned} Q_p &= \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} \alpha^p d\mu_\varepsilon(\alpha) \\ &= \sum_{k=0}^{\lfloor p/2 \rfloor} \binom{p}{2k} \frac{(2\pi)^{2k}}{2k+1} \int_{-L}^{0} D(\lambda; x, t)^{p-2k} \varphi(\lambda)^{4k} F(\lambda) d\lambda, \end{aligned}$$

uniformly for (x, t) in compact sets. Now we apply the identity

$$(4.74) \quad \sum_{k=0}^{\lfloor p/2 \rfloor} \frac{1}{2k+1} \binom{p}{2k} a^{2k} b^{p-2k} = \frac{(b+a)^{p+1} - (b-a)^{p+1}}{2a(1+p)},$$

which holds for any integer $p \geq 0$ and real numbers a and b. (This identity can be most easily obtained by expanding the binomials on the right-hand side.) Recalling definitions (3.31) and (3.32) of $D(\lambda; x, t)$ and $\varphi(\lambda)$ and using the fact that $x_\pm(\lambda) = \pm \pi F(\lambda) - \gamma(\lambda)$ then completes the proof of Proposition 4.1.

4.3 Convergence of Measures and Locally Uniform Convergence of $\tilde{U}_\varepsilon$: Proof of Proposition 4.2

Recall the measures μ_ε defined by (4.4).

LEMMA 4.9 *For each nonnegative integer p,*

$$(4.75) \quad \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} \alpha^p d\mu_\varepsilon(\alpha) = \int_{\mathbb{R}} \alpha^p d\mu(\alpha)$$

where μ is the absolutely continuous (with respect to Lebesgue measure on $\mathbb{R}$) measure defined by $d\mu(\alpha) = G(\alpha; x, t)d\alpha$, and the compactly supported integrable density function $G(\alpha; x, t)$ is given by (4.10). The limit is uniform with respect to (x, t) in compact sets. Also, like each μ_ε, μ is a measure with mass M.

PROOF: Recalling Proposition 4.1, we first show that the given measure μ satisfies

$$(4.76) \qquad \int_{\mathbb{R}} \alpha^p \, d\mu(\alpha) = Q_p,$$

where Q_p is given by (4.7) for all nonnegative $p \in \mathbb{Z}$. Equivalently, we may construct a measure with the desired moments as follows: the characteristic function of the measure μ is the Fourier transform

$$(4.77) \qquad \widehat{G}(\xi; x, t) := \int_{\mathbb{R}} G(\alpha; x, t)e^{-i\alpha\xi} \, d\alpha,$$

and this function necessarily has the desired moments $\{Q_p\}_{p=0}^\infty$ as its derivatives at $\xi = 0$:

$$(4.78) \qquad \frac{d^p \widehat{G}}{d\xi^p}(0; x, t) = (-i)^p Q_p.$$

So $\widehat{G}(\xi; x, t)$ has the Taylor series

$$(4.79) \qquad \widehat{G}(\xi; x, t) = \sum_{p=0}^\infty \frac{(-i\xi)^p}{p!} Q_p.$$

Now from the obvious inequality $|x + 2\lambda t - x_\pm(\lambda)| \leq |x - x_0| + 2L|t| + 2\pi F(\lambda)$, we obtain

$$
\begin{aligned}
(4.80) \qquad |Q_p| &\leq \frac{1}{\pi(p+1)} \int_{-L}^0 (|x - x_0| + 2L|t| + 2\pi F(\lambda))^{p+1}(-2\lambda)^p \, d\lambda \\
&\leq \frac{1}{\pi(p+1)} \int_{-L}^0 (2L|x - x_0| + 4L^2|t| - 4\pi\lambda F(\lambda))^p \\
&\qquad \cdot (|x - x_0| + 2L|t| + 2\pi F(\lambda))d\lambda.
\end{aligned}
$$

Also, from Lemma 3.4, there is a constant $K > 0$ such that $0 \leq -\lambda F(\lambda) \leq K$, so for $(x - x_0)^2 + t^2 \leq R^2$,

$$
\begin{aligned}
(4.81) \qquad |Q_p| &\leq \frac{(2LR + 4L^2R + 4\pi K)^p}{\pi(p+1)} \int_{-L}^0 (|x - x_0| + 2L|t| + 2\pi F(\lambda))d\lambda \\
&\leq \frac{(LR + 2L^2R + 2\pi M)}{\pi(p+1)}(2LR + 4L^2R + 4\pi K)^p \\
&\leq \frac{1}{\pi}(LR + 2L^2R + 2\pi M)(2LR + 4L^2R + 4\pi K)^p,
\end{aligned}
$$

where in the last step we used (3.20). This inequality implies that the Taylor series (4.79) converges for all $\xi \in \mathbb{C}$ to an entire function of exponential type.

Now we will sum the Taylor series (4.79) in closed form by substituting from formula (4.7) and exchanging the order of summation and integration. Indeed, since

$$(4.82) \quad \sum_{p=0}^{\infty} \frac{(-i\xi)^p}{p!} \cdot \frac{(-2\lambda)^p(x + 2\lambda t - x_{\pm}(\lambda))^{p+1}}{p+1} =$$

$$\frac{e^{2i\xi\lambda[x+2\lambda t - x_{\pm}(\lambda)]} - 1}{2i\xi\lambda},$$

we obtain the formula

$$(4.83) \quad \widehat{G}(\xi; x, t) = \int_{-L}^{0} \frac{e^{2i\xi\lambda[x+2\lambda t - x_{-}(\lambda)]} - e^{2i\xi\lambda[x+2\lambda t - x_{+}(\lambda)]}}{4\pi i \xi \lambda} \, d\lambda.$$

Computing the inverse Fourier transform

$$(4.84) \quad G(\alpha; x, t) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{G}(\xi; x, t) e^{i\alpha\xi} \, d\xi$$

by exchanging the order of integration leads directly to the claimed formula (4.10).

It is obvious that $G(\alpha; x, t)$ is a nonnegative function, and since by Lemma 3.4

$$(4.85) \quad \begin{aligned} \inf_{-L<\lambda<0} -2\lambda(x + 2\lambda t - x_{+}(\lambda)) &> -\infty, \\ \sup_{-L<\lambda<0} -2\lambda(x + 2\lambda t - x_{-}(\lambda)) &< +\infty, \end{aligned}$$

for every (x, t), it is clear that $G(\alpha; x, t)$ has compact support. It is also straightforward to verify that μ has mass M:

$$(4.86) \quad \begin{aligned} \int_{\mathbb{R}} d\mu(\alpha) &= \int_{\mathbb{R}} G(\alpha; x, t) d\alpha \\ &= -\frac{1}{4\pi} \int_{\mathbb{R}} \int_{-L}^{0} \chi_{[-2\lambda(x+2\lambda t - x_{+}(\lambda)), -2\lambda(x+2\lambda t - x_{-}(\lambda))]}(\alpha) \frac{d\lambda}{\lambda} \, d\alpha \\ &= -\frac{1}{4\pi} \int_{-L}^{0} \frac{1}{\lambda} \int_{\mathbb{R}} \chi_{[-2\lambda(x+2\lambda t - x_{+}(\lambda)), -2\lambda(x+2\lambda t - x_{-}(\lambda))]}(\alpha) d\alpha \, d\lambda \\ &= -\frac{1}{4\pi} \int_{-L}^{0} \frac{1}{\lambda} \int_{-2\lambda(x+2\lambda t - x_{+}(\lambda))}^{-2\lambda(x+2\lambda t - x_{-}(\lambda))} d\alpha \, d\lambda \\ &= \int_{-L}^{0} F(\lambda) d\lambda \\ &= M, \end{aligned}$$

according to (3.20). Therefore μ is indeed an absolutely continuous, compactly supported (nonnegative) measure of mass M. $\qquad\square$

Note that the reconstruction of the measure μ from its moments is virtually the same calculation as took place on the direct scattering side in our discussion of Matsuno's method in Section 3.2.

LEMMA 4.10 *There is a compact interval $\Omega \subset \mathbb{R}$ containing the support of all of the measures $\{\mu_\varepsilon\}_{\varepsilon>0}$ as well as that of the measure μ, and Ω may be chosen independent of (x, t) in any given compact set.*

PROOF: Since μ has compact support certainly contained within the interval

$$(4.87) \quad \inf_{-L<\lambda<0}[2\lambda x_+(\lambda)] - 2L|x| - 4L^2|t| \leq \alpha \leq$$

$$\sup_{-L<\lambda<0}[2\lambda x_-(\lambda)] + 2L|x| + 4L^2|t|$$

that is clearly bounded uniformly for (x, t) in any compact set, it is enough to show that the support of μ_ε is uniformly bounded as $\varepsilon \downarrow 0$. But by definition of μ_ε this is equivalent to showing that the eigenvalue of $\mathbf{A}_\varepsilon$ with the largest magnitude remains uniformly bounded as $\varepsilon \downarrow 0$.

Since the matrix $\tilde{\mathbf{A}}_\varepsilon$ is Hermitian, we have

$$(4.88) \quad \|\tilde{\mathbf{A}}_\varepsilon\|_2 = \max_{1 \leq j \leq N(\varepsilon)} |\alpha_j|,$$

so to prove that the eigenvalue of $\tilde{\mathbf{A}}_\varepsilon$ with the largest magnitude remains uniformly bounded, it is completely equivalent to prove that the ℓ^2 (induced) matrix norm of $\tilde{\mathbf{A}}_\varepsilon$ is uniformly bounded as $\varepsilon \downarrow 0$ independent of (x, t) in any given compact set.

Recalling the decomposition $\tilde{\mathbf{A}}_\varepsilon = \mathbf{D} + \mathbf{H}$ from the proof of Proposition 4.1 given in Section 4.2, the triangle inequality gives $\|\tilde{\mathbf{A}}_\varepsilon\|_2 \leq \|\mathbf{D}\|_2 + \|\mathbf{H}\|_2$, and since $\mathbf{D}$ is diagonal,

$$(4.89) \quad \begin{aligned} \|\mathbf{D}\|_2 &= \max_{1 \leq n \leq N(\varepsilon)} |2\tilde{\lambda}_n(x + 2\tilde{\lambda}_n t + \gamma(\tilde{\lambda}_n))| \\ &\leq \sup_{-L<\lambda<0} |2\lambda(x + 2\lambda t + \gamma(\lambda))| \\ &\leq \sup_{-L<\lambda<0} |2\lambda\gamma(\lambda)| + 2L|x| + 4L^2|t|, \end{aligned}$$

so since $\lambda\gamma(\lambda)$ is bounded according to Lemma 3.4, and $\mathbf{H}$ is independent of x and t, it is sufficient to show that $\|\mathbf{H}\|_2$ remains bounded as $\varepsilon \downarrow 0$.

To estimate $\|\mathbf{H}\|_2$, we write $\mathbf{H}$ in the following form: $\mathbf{H} = \mathbf{BTB} + \mathbf{E}$ where

$$(4.90) \quad \mathbf{B} = \text{diag}\left(e^{i\pi/4}\sqrt{-2\tilde{\lambda}_1 F(\tilde{\lambda}_1)}, \ldots, e^{i\pi/4}\sqrt{-2\tilde{\lambda}_{N(\varepsilon)} F(\tilde{\lambda}_{N(\varepsilon)})}\right),$$

and $\mathbf{T}$ is the $N(\varepsilon) \times N(\varepsilon)$ Toeplitz matrix with elements $T_{nm} = f_{n-m}$, where the sequence $\{f_n\}_{n\in\mathbb{Z}}$ is defined by (4.34). Of course $\mathbf{E} := \mathbf{H} - \mathbf{BTB}$. Therefore $\|\mathbf{H}\|_2 \leq \|\mathbf{B}\|_2^2 \|\mathbf{T}\|_2 + \|\mathbf{E}\|_2$. Because $\mathbf{B}$ is diagonal,

$$(4.91) \quad \|\mathbf{B}\|_2^2 \leq \max_{1 \leq n \leq N(\varepsilon)} [-2\tilde{\lambda}_n F(\tilde{\lambda}_n)] \leq \sup_{-L<\lambda<0} [-2\lambda F(\lambda)],$$

which is finite by Lemma 3.4. The Toeplitz matrix $\mathbf{T}$ can be written as $\mathbf{T} = \mathcal{P}\mathcal{T}_f\mathcal{P}$, where $\mathcal{P}$ is the orthogonal projection from $\ell^2(\mathbb{Z})$ onto $\mathbb{C}^N$ viewed as a subset of $\ell^2(\mathbb{Z})$ associated with components having indices $\{1, 2, \ldots, N(\varepsilon)\} \subset \mathbb{Z}$, and where $\mathcal{T}_f : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ is the Toeplitz operator defined by (4.33) from Section 4.2. The $\ell^2(\mathbb{Z})$ operator norm of $\mathcal{P}$ is clearly equal to one, and since

$$(4.92) \qquad \sum_{l \in \mathbb{Z}} f_l e^{il\theta} = i(\pi - \theta), \quad 0 < \theta < 2\pi,$$

the Pythagorean theorem in $L^2(0, 2\pi)$ gives

$$
\begin{aligned}
(4.93) \qquad \sum_{n \in \mathbb{Z}} |(\mathcal{T}c)_n|^2 &= \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{n \in \mathbb{Z}} (\mathcal{T}c)_n e^{in\theta} \right|^2 d\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} f_{n-m} c_m e^{in\theta} \right|^2 d\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{m \in \mathbb{Z}} c_m e^{im\theta} \sum_{n \in \mathbb{Z}} f_{n-m} e^{i[n-m]\theta} \right|^2 d\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi} (\pi - \theta)^2 \left| \sum_{m \in \mathbb{Z}} c_m e^{im\theta} \right|^2 d\theta \\
&\leq \pi^2 \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{m \in \mathbb{Z}} c_m e^{im\theta} \right|^2 d\theta \\
&= \pi^2 \sum_{m \in \mathbb{Z}} |c_m|^2,
\end{aligned}
$$

so the $\ell^2(\mathbb{Z})$ operator norm of $\mathcal{T}_f$ is bounded by π. It follows that $\|\mathbf{H}\|_2 \leq \pi + \|\mathbf{E}\|_2$, so it suffices to show that $\|\mathbf{E}\|_2$ remains bounded as $\varepsilon \downarrow 0$.

So far, we have exploited the special structure of the dominant parts of the matrix $\tilde{\mathbf{A}}_\varepsilon$ and applied correspondingly specialized norm estimates to these terms. The error term $\mathbf{E}$ has less structure, but it is smaller; to estimate its norm it will be sufficient to use the rather crude inequality $\|\mathbf{E}\|_2 \leq \|\mathbf{E}\|_{\mathrm{HS}}$ and work with the Hilbert-Schmidt norm

$$(4.94) \qquad \|\mathbf{E}\|_{\mathrm{HS}}^2 := \sum_{n=1}^{N(\varepsilon)} \sum_{m=1}^{N(\varepsilon)} |E_{nm}|^2,$$

where the elements of $\mathbf{E}$ are explicitly given by

$$(4.95) \qquad E_{nm} := 2i \left[\frac{\varepsilon \sqrt{\tilde{\lambda}_n \tilde{\lambda}_m}}{\tilde{\lambda}_n - \tilde{\lambda}_m} - \frac{\sqrt{\tilde{\lambda}_n F(\tilde{\lambda}_n) \tilde{\lambda}_m F(\tilde{\lambda}_m)}}{n - m} \right]$$
$$\text{for } n \neq m \text{ and } E_{nn} = 0.$$

If we introduce continuous variables $a := (n - \frac{1}{2})\varepsilon$ and $b := (m - \frac{1}{2})\varepsilon$, then it is easy to see that the square of the Hilbert-Schmidt norm of $\mathbf{E}$ is a Riemann sum approximation of a certain double integral:

$$(4.96) \qquad \lim_{\varepsilon \downarrow 0} \|\mathbf{E}\|_{\mathrm{HS}}^2 = \iint\limits_{[0,M]^2} e_0(a,b)da\, db,$$

provided the double integral exists, where

$$(4.97) \qquad e_0(a,b) := 4\left[\frac{\sqrt{m^{-1}(a)m^{-1}(b)}}{m^{-1}(a) - m^{-1}(b)} - \frac{\sqrt{m^{-1}(a)F(m^{-1}(a))m^{-1}(b)F(m^{-1}(b))}}{a - b} \right]^2,$$

and where $m^{-1}(\cdot)$ denotes the inverse function to the monotone function $m(\cdot)$ given by

$$(4.98) \qquad m(\lambda) := \int_{-L}^{\lambda} F(\lambda')d\lambda'.$$

By changing variables to $\kappa = m^{-1}(a)$ and $\lambda = m^{-1}(b)$,

$$(4.99) \qquad \iint\limits_{[0,M]^2} e_0(a,b)da\, db = \iint\limits_{[-L,0]^2} e(\kappa,\lambda)d\kappa\, d\lambda,$$

where

$$(4.100) \qquad e(\kappa,\lambda) := 4\left[\frac{\sqrt{\kappa\lambda}}{\kappa - \lambda} - \frac{\sqrt{\kappa F(\kappa)\lambda F(\lambda)}}{m(\kappa) - m(\lambda)} \right]^2 F(\kappa)F(\lambda).$$

Note that since $F \geq 0$ by Lemma 3.4, $e(\kappa,\lambda) \geq 0$ for $(\kappa,\lambda) \in [-L,0]^2$. To complete the proof of the lemma, it is enough to show that the double integral on the right-hand side of (4.99) is finite.

In order to estimate the double integral, we divide the square $[-L,0]^2$ into polygonal regions as follows (see Figure 4.3):

- The square $[-L, -L + \delta]^2$ contains those ordered pairs (κ, λ) for which both κ and λ are near the "soft edge" of the eigenvalue spectrum at $-L$. We divide this square into diagonal and off-diagonal parts according to whether $(\kappa + L)/2 \leq \lambda + L \leq 2(\kappa + L)$ (the diagonal part, S_{D}) or not (the off-diagonal parts, S_{OD}).
- The square $[-\delta, 0]^2$ contains those ordered pairs (κ, λ) for which both κ and λ are near the "hard edge" of the eigenvalue spectrum at zero. We divide this square into diagonal and off-diagonal parts according to whether $2\kappa < \lambda < \kappa/2$ (the diagonal part, H_{D}) or not (the off-diagonal parts H_{OD}).
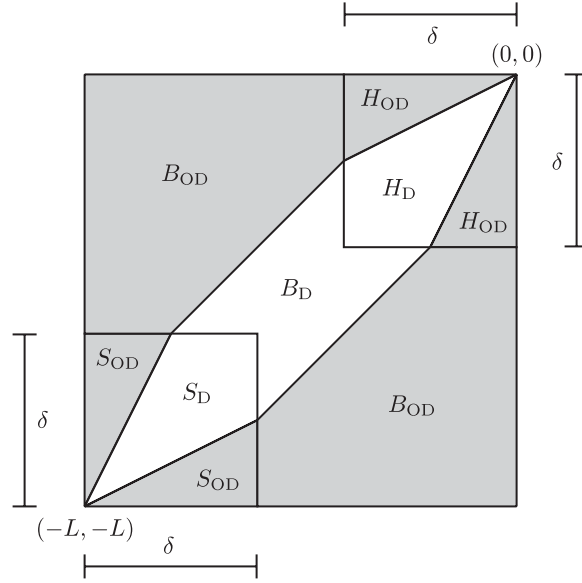
FIGURE 4.3. The square $[-L, 0]^2$ in the (κ, λ)-plane is covered by the six regions S_D, S_{OD}, B_D, B_{OD}, H_D, and H_{OD}.

- The remaining part of $[-L, 0]^2$ contains those ordered pairs (κ, λ) for which at least one of the coordinates lies in the "bulk" of the eigenvalue spectrum, bounded away from both edges. This is divided into a diagonal part B_D and two off-diagonal parts B_{OD} along two straight line segments parallel to the diagonal as indicated in Figure 4.3.

Here the constant $\delta > 0$ is as specified in Lemma 3.4. As $e(\kappa, \lambda) = e(\lambda, \kappa)$ it will be enough to show integrability of e over the part of $[-L, 0]^2$ with $\kappa < \lambda$, an inequality that we will assume tacitly below.

First we consider integrating $e(\kappa, \lambda)$ over the "off-diagonal" shaded regions S_{OD}, B_{OD}, and H_{OD} shown in Figure 4.3. An upper bound for $e(\kappa, \lambda)$ useful in these regions is easily obtained from the inequality $(a - b)^2 \leq 2a^2 + 2b^2$:

(4.101)
$$e(\kappa, \lambda) \leq 8\kappa F(\kappa)\lambda F(\lambda)\left[\frac{1}{(\kappa - \lambda)^2} + \frac{F(\kappa)F(\lambda)}{(m(\kappa) - m(\lambda))^2}\right],$$
$$(\kappa, \lambda) \in (-L, 0)^2.$$

Applying the mean value theorem to this estimate yields

(4.102)
$$e(\kappa, \lambda) \leq \frac{8\kappa F(\kappa)\lambda F(\lambda)}{(\kappa - \lambda)^2}\left[1 + \frac{F(\kappa)F(\lambda)}{F(\xi)^2}\right],$$

where $\kappa \leq \xi \leq \lambda$. Finally, we obtain

(4.103)
$$\begin{aligned}
e(\kappa, \lambda) &\leq \frac{8\kappa F(\kappa)\lambda F(\lambda)}{(\kappa - \lambda)^2}\left[1 + \frac{F(\lambda)}{F(\kappa)}\right] \\
&= \frac{8\kappa F(\kappa)\lambda F(\lambda)}{(\kappa - \lambda)^2} + \frac{8\kappa\lambda F(\lambda)}{(\kappa - \lambda)^2}F(\lambda)
\end{aligned}$$

since F is monotone increasing according to Lemma 3.4.

Now, for $(\kappa, \lambda) \in B_{\mathrm{OD}}$, we have that $\kappa - \lambda$ is bounded away from zero while by Lemma 3.4 $\kappa F(\kappa)$ and $\lambda F(\lambda)$ are bounded (and of course $|\kappa| < L$) while $F(\lambda)$ is integrable. Hence we easily conclude that $e(\kappa, \lambda)$ is integrable on B_{OD}.

If $(\kappa, \lambda) \in H_{\mathrm{OD}}$ with $\kappa < \lambda$, then we have the inequality

(4.104)
$$(\kappa - \lambda)^2 = \left(\left[\lambda - \frac{\kappa}{2}\right] + \left[-\frac{\kappa}{2}\right]\right)^2 \geq \frac{\kappa^2}{4},$$

and also since both $-\delta < \kappa < 0$ and $-\delta < \lambda < 0$, we may use the upper bound for F given in (3.35) from Lemma 3.4 to replace (4.103) with

(4.105) $e(\kappa, \lambda) \leq 32C_0^2(-\kappa)^{-1-1/q}(-\lambda)^{1-1/q} + 32C_0^2(-\kappa)^{-1}(-\lambda)^{1-2/q}$,

where $C_0 > 0$ and $q > 1$ are the constants in (3.35). This estimate is easily seen to be integrable on the component of H_{OD} with $\kappa < \lambda$ by direct calculation of the iterated integrals.

If $(\kappa, \lambda) \in S_{\mathrm{OD}}$ with $\kappa < \lambda$, then we have the inequality

(4.106) $$(\kappa - \lambda)^2 = \left(\left[\frac{\lambda + L}{2}\right] + \left[\frac{\lambda + L}{2} - (\kappa + L)\right]\right)^2 \geq \frac{(\lambda + L)^2}{4},$$

and also since both $-L < \kappa < -L + \delta$ and $-L < \lambda < -L + \delta$ we may use the upper bound for F given in (3.33) from Lemma 3.4 along with the inequalities $|\kappa| < L$ and $|\lambda| < L$ to replace (4.103) with

(4.107) $e(\kappa, \lambda) \leq 32L^2 C_{-L}^2(\kappa + L)^{1/2}(\lambda + L)^{-3/2} + 32L^2 C_{-L}^2(\lambda + L)^{-1}$.

This upper bound is obviously integrable on the component of S_{OD} with $\kappa < \lambda$.

Now we consider integrating $e(\kappa, \lambda)$ over the "diagonal" unshaded regions S_{D}, B_{D}, and H_{D} shown in Figure 4.3. By the mean value theorem and the monotonicity of F guaranteed by Lemma 3.4, we obtain an upper bound more useful when $\kappa \approx \lambda$:

(4.108) $$e(\kappa, \lambda) \leq 4\kappa F(\kappa)\lambda F(\lambda)\left[\frac{F(\kappa) - F(\lambda)}{m(\kappa) - m(\lambda)}\right]^2, \quad (\kappa, \lambda) \in (-L, 0)^2.$$

Again using the mean value theorem and monotonicity of F we may make the upper bound larger for $\kappa < \lambda$:

(4.109) $$e(\kappa, \lambda) \leq \frac{4\kappa\lambda F(\lambda)F'(\xi)^2}{F(\kappa)},$$

where $\kappa \leq \xi \leq \lambda$.

For $(\kappa, \lambda) \in B_D$ with $\kappa < \lambda$, both κ and λ are bounded away from the soft and hard edges of the eigenvalue spectrum, so Lemma 3.4 guarantees that F and F' are bounded, and F is also bounded away from zero by strict monotonicity and the boundary condition $F(-L) = 0$. It follows from (4.109) that $e(\kappa, \lambda)$ is bounded and hence integrable on B_D.

If $(\kappa, \lambda) \in H_D$ then we may use estimates (3.35) and (3.36) from Lemma 3.4 to replace (4.109) with

$$
\begin{aligned}
e(\kappa, \lambda) &\leq \frac{8C_0^2}{q^2}(-\kappa)^{1+1/q}(-\lambda)^{1-1/q}(-\xi)^{-2/q-2} \\
&\leq \frac{8C_0^2}{q^2}(-\kappa)^{1+1/q}(-\lambda)^{-1-3/q}.
\end{aligned}
$$

(4.110)

The double integral of this upper bound over the region H_D with $\kappa < \lambda$ is easily computed by iterated integration and is clearly finite as a consequence of the fact that $q > 1$.

Finally, if $(\kappa, \lambda) \in S_D$ with $\kappa < \lambda$, then we may use estimates (3.33) and (3.34) from Lemma 3.4 together with the inequalities $|\kappa| < L$ and $|\lambda| < L$ to replace (4.109) with

$$
\begin{aligned}
e(\kappa, \lambda) &\leq 2L^2 C_{-L}^2 (\kappa + L)^{-1/2}(\lambda + L)^{1/2}(\xi + L)^{-1} \\
&\leq 2L^2 C_{-L}^2 (\kappa + L)^{-3/2}(\lambda + L)^{1/2},
\end{aligned}
$$

(4.111)

an upper bound that is clearly integrable over the part of S_D with $\kappa < \lambda$. $\square$

LEMMA 4.11 *The measure μ_ε converges in the weak-$*$ sense to μ, uniformly for (x, t) in compact sets. That is, for each continuous function $f : \mathbb{R} \to \mathbb{C}$,*

$$
(4.112) \qquad \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} f(\alpha)d\mu_\varepsilon(\alpha) = \int_{\mathbb{R}} f(\alpha)d\mu(\alpha),
$$

with the limit being uniform with respect to (x, t) in compact sets.

PROOF: According to Lemma 4.9, for each polynomial $p(\alpha)$ we have the following limit, uniform for (x, t) in compact sets:

$$
(4.113) \qquad \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} p(\alpha)d\mu_\varepsilon(\alpha) = \int_{\mathbb{R}} p(\alpha)d\mu(\alpha).
$$

But by Lemma 4.10 we can equivalently integrate over the compact interval Ω (independent of (x, t) in any given compact set) with the same result. Now by the Weierstrass approximation theorem, given any continuous function $f : \mathbb{R} \to \mathbb{C}$ and any $\rho > 0$ there is a polynomial $p_\rho^f(\alpha)$ for which

$$
(4.114) \qquad \sup_{\lambda \in \Omega} |f(\alpha) - p_\rho^f(\alpha)| < \frac{\rho}{M},
$$

so for any measure ν of mass M with support in Ω (like μ_ε and μ),

$$(4.115) \qquad \left| \int_\mathbb{R} [f(\alpha) - p_\rho^f(\alpha)] d\nu(\alpha) \right| \leq \int_\Omega |f(\alpha) - p_\rho^f(\alpha)| d\nu(\alpha) < \rho.$$

Let $\omega > 0$ be an arbitrarily small positive number. Then if we write

$$(4.116) \qquad \nu[g] := \int_\Omega g(\alpha) d\nu(\alpha),$$

we have

$$\left| \int_\mathbb{R} f(\alpha) d\mu_\varepsilon(\alpha) - \int_\mathbb{R} f(\alpha) d\mu(\alpha) \right|$$

$$= \left| \mu_\varepsilon[f] - \mu[f] \right|$$

$$(4.117) \qquad = \left| [\mu_\varepsilon[p_{\omega/3}^f] - \mu[p_{\omega/3}^f]] + \mu_\varepsilon[f - p_{\omega/3}^f] - \mu[f - p_{\omega/3}^f] \right|$$

$$\leq \left| \mu_\varepsilon[p_{\omega/3}^f] - \mu[p_{\omega/3}^f] \right| + \left| \mu_\varepsilon[f - p_{\omega/3}^f] \right| + \left| \mu[f - p_{\omega/3}^f] \right|$$

$$< \left| \mu_\varepsilon[p_{\omega/3}^f] - \mu[p_{\omega/3}^f] \right| + \frac{2}{3}\omega,$$

with the last inequality following from (4.115). But with $\omega > 0$ fixed, (4.113) implies that $\varepsilon > 0$ may be chosen sufficiently small, independently of (x, t) in any given compact set, that

$$(4.118) \qquad \left| \mu_\varepsilon[p_{\omega/3}^f] - \mu[p_{\omega/3}^f] \right| < \frac{1}{3}\omega,$$

which implies

$$(4.119) \qquad \left| \int_\mathbb{R} f(\alpha) d\mu_\varepsilon(\alpha) - \int_\mathbb{R} f(\alpha) d\mu(\alpha) \right| < \omega$$

thereby completing the proof.   □

Now we are in a position to complete the proof of Proposition 4.2. We begin by writing $\tilde{U}_\varepsilon(x, t)$ as defined by (4.3) in terms of the normalized (to mass M) counting measure μ_ε:

$$(4.120) \qquad \tilde{U}_\varepsilon(x, t) = \left[\frac{\varepsilon N(\varepsilon)}{M} \right] \int_\mathbb{R} 2 \arctan(\varepsilon^{-1}\alpha) d\mu_\varepsilon(\alpha).$$

Define the continuous functions

$$(4.121) \quad a_+(\alpha) := \pi + 4H(-\alpha) \arctan(\alpha), \quad a_-(\alpha) := -a_+(-\alpha), \quad \alpha \in \mathbb{R},$$
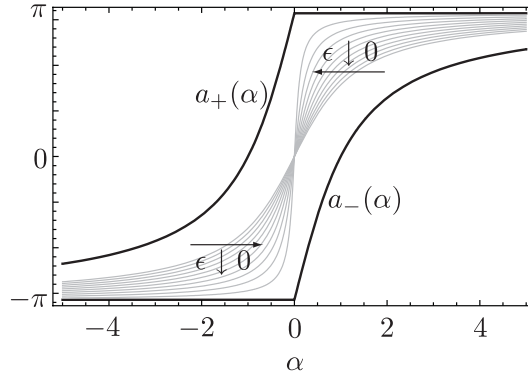
FIGURE 4.4. The graphs of $a_-(\alpha) < a_+(\alpha)$ (black) and several graphs of $2\arctan(\varepsilon^{-1}\alpha)$ for $\varepsilon \leq 1$ (gray).

where $H(\cdot)$ denotes the Heaviside step function. It is then easy to check (see Figure 4.4) that for any $E > 0$,

$$(4.122) \quad 0 < \varepsilon \leq E \quad \Longrightarrow \quad a_-(E^{-1}\alpha) \leq 2\arctan(\varepsilon^{-1}\alpha) \leq a_+(E^{-1}\alpha), \quad \alpha \in \mathbb{R}.$$

Therefore, for any $E > 0$ and all $0 < \varepsilon < E$,

$$(4.123) \quad \begin{aligned} \int_{\mathbb{R}} a_-(E^{-1}\alpha)d\mu_\varepsilon(\alpha) &\leq \int_{\mathbb{R}} 2\arctan(\varepsilon^{-1}\alpha)d\mu_\varepsilon(\alpha) \\ &\leq \int_{\mathbb{R}} a_+(E^{-1}\alpha)d\mu_\varepsilon(\alpha). \end{aligned}$$

Using Lemma 4.11 we may pass to the limit $\varepsilon \downarrow 0$ in the lower and upper bounds to obtain

$$(4.124) \qquad \liminf_{\varepsilon \downarrow 0} \int_{\mathbb{R}} 2\arctan(\varepsilon^{-1}\alpha)d\mu_\varepsilon(\alpha) \geq \int_{\mathbb{R}} a_-(E^{-1}\alpha)d\mu(\alpha)$$

and also

$$(4.125) \qquad \limsup_{\varepsilon \downarrow 0} \int_{\mathbb{R}} 2\arctan(\varepsilon^{-1}\alpha)d\mu_\varepsilon(\alpha) \leq \int_{\mathbb{R}} a_+(E^{-1}\alpha)d\mu(\alpha).$$

In these statements, $E > 0$ is an arbitrary parameter, and the limits are uniform for (x, t) in compact sets. But $a_\pm(E^{-1}\alpha)$ are uniformly bounded functions that both tend pointwise for $\alpha \neq 0$ to the same limit function $\pi\,\mathrm{sgn}(\lambda)$ as $E \downarrow 0$, while μ is a fixed measure that is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, so by the Lebesgue dominated convergence theorem,

$$(4.126) \qquad \lim_{E \downarrow 0} \int_{\mathbb{R}} a_\pm(E^{-1}\alpha)d\mu(\alpha) = \int_{\mathbb{R}} \pi\,\mathrm{sgn}(\alpha)d\mu(\alpha).$$

By letting $E \downarrow 0$, it then follows from (4.124) and (4.125) that

$$(4.127) \qquad \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} 2 \arctan(\varepsilon^{-1}\alpha) d\mu_\varepsilon(\alpha) = \int_{\mathbb{R}} \pi \operatorname{sgn}(\alpha) d\mu(\alpha)$$

with the limit being uniform for (x, t) in any given compact set. Finally, according to (3.24), we have (independent of x and t)

$$(4.128) \qquad \lim_{\varepsilon \downarrow 0} \frac{\varepsilon N(\varepsilon)}{M} = 1,$$

so combining this result with (4.127) and noting that $d\mu(\alpha) = G(\alpha; x, t) d\alpha$ completes the proof of Proposition 4.2.

4.4 Differentiation of $\tilde{U}_\varepsilon$: Burgers' Equation, Weak Convergence of $\tilde{u}_\varepsilon$, and Completion of Proof of Theorem 1.1

Let $\phi \in \mathscr{D}(\mathbb{R})$ be a test function. Then by integration by parts and the uniform convergence of $\tilde{U}_\varepsilon(x, t)$ to $U(x, t)$ on compact sets in the (x, t)-plane guaranteed by Proposition 4.2,

$$\lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} \tilde{u}_\varepsilon(x, t)\phi(x) dx = \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} \frac{\partial \tilde{U}_\varepsilon}{\partial x}(x, t)\phi(x) dx$$

$$(4.129) \qquad\qquad = -\lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} \tilde{U}_\varepsilon(x, t)\phi'(x) dx$$

$$= -\int_{\mathbb{R}} U(x, t)\phi'(x) dx.$$

LEMMA 4.12 *The limit function $U(x, t)$ is continuously differentiable with respect to x, and if (x, t) is a point for which there are $2P(x, t) + 1$ solutions $u_0^B(x, t) < u_1^B(x, t) < \cdots < u_{2P(x,t)}^B(x, t)$ of the implicit equation (4.134),*

$$(4.130) \qquad \frac{\partial U}{\partial x}(x, t) = \sum_{n=0}^{2P(x,t)} (-1)^n u_n^B(x, t),$$

and the above formula is extended to nongeneric (x, t) by continuity.

PROOF: Exchanging the order of integration in the double-integral formula for $U(x, t)$ obtained by substituting $d\mu(\alpha) = G(\alpha; x, t) d\alpha$ with G given by (4.10) into (4.9), we obtain

$$(4.131) \qquad U(x, t) = \int_{-L}^{0} J(\lambda; x, t) d\lambda,$$

where

$$(4.132) \qquad J(\lambda; x, t) := -\frac{1}{4\lambda} \int_{-2\lambda(x+2\lambda t - x_+(\lambda))}^{-2\lambda(x+2\lambda t - x_-(\lambda))} \operatorname{sgn}(\alpha) d\alpha.$$

Note that for $\lambda \in [-L, 0]$ the upper limit of integration is greater than or equal to the lower limit. Moreover, the integral in $J(\lambda; x, t)$ is easily evaluated; for $-L < \lambda < 0$,

$$(4.133) \quad J(\lambda; x, t) = \begin{cases} -\pi F(\lambda), & x + 2\lambda t - x_-(\lambda) < 0, \\ x + 2\lambda t + \gamma(\lambda), & x + 2\lambda t - x_+(\lambda) \leq 0 \leq x + 2\lambda t - x_-(\lambda), \\ \pi F(\lambda), & x + 2\lambda t - x_+(\lambda) > 0. \end{cases}$$

It follows from the relations $x_\pm(\lambda) = \pm\pi F(\lambda) - \gamma(\lambda)$ that for an admissible initial condition u_0, J is a continuous function of x for each fixed t, uniformly with respect to $\lambda \in [-L, 0]$, and hence also from (4.131) that $U(\cdot, t)$ is continuous on $\mathbb{R}$ for each t. To prove that $U(\cdot, t)$ is continuously differentiable, it will therefore suffice to establish continuous differentiability on the complement of a finite set of points and that the resulting piecewise formula for $\partial U/\partial x$ extends continuously to the whole real line.

To use formula (4.133) in the representation (4.131), we therefore need to know those points $\lambda \in (-L, 0)$ at which one of the two quantities $x + 2\lambda t - x_+(\lambda) < x + 2\lambda t - x_-(\lambda)$ changes sign. Under the variable substitution $\lambda = -u^B$, the definition of the turning points $x_\pm(\lambda)$ as branches of the inverse function of u_0 implies that the union of solutions of the two equations $x + 2\lambda t - x_\pm(\lambda) = 0$ is exactly the totality of solutions of the implicit equation

$$(4.134) \qquad\qquad u^B = u_0(x - 2u^B t).$$

In other words, the transitional points λ for formula (4.133) correspond under the sign change $u^B = -\lambda$ to the branches of the multivalued solution of Burgers' equation

$$(4.135) \qquad\qquad \frac{\partial u^B}{\partial t} + 2u^B \frac{\partial u^B}{\partial x} = 0$$

subject to the admissible initial condition $u^B(x, 0) = u_0(x)$.

Note that admissibility of u_0 implies (see Definition 3.1) that given any $t \in \mathbb{R}$ there exist only a finite number of breaking points (x_ξ, t_ξ) with t_ξ in the closed interval between 0 and t. Indeed, the breaking points correspond to values of $\xi \in \mathbb{R}$ for which $u_0''(\xi) = 0$ but $u_0'''(\xi) \neq 0$, and the breaking times are $t_\xi = (-2u_0'(\xi))^{-1}$; since $u_0'(\xi)$ decays to zero for large ξ, bounded breaking times t_ξ correspond to bounded ξ, and there are only finitely many of these by hypothesis. Moreover, each breaking point (x_ξ, t_ξ) generates a new fold in the solution surface lying between two caustic curves emerging in the direction of increasing $|t|$ from (x_ξ, t_ξ), and because $u'''(\xi) \neq 0$ there are exactly two more sheets of the multivalued solution of Burgers' equation born within the fold as a result of a simple pitchfork bifurcation. Therefore, the union of caustic curves and breaking points meets any

line of constant t in the (x,t)-plane in a finite set of points $\{x_j^{\mathrm{crit}}(t)\}$, and on every connected component of the set $S_t := \{(x,t) : x \in \mathbb{R} \setminus \{x_j^{\mathrm{crit}}(t)\}\}$, there is a finite, odd, and constant (with respect to x) number $2P(x,t) + 1$ of roots of equation (4.134), and all roots are simple (and hence differentiable with respect to x).

If $t \geq 0$, then by admissibility of u_0 the quantity $b_-(\lambda;x,t) := x + 2\lambda t - x_-(\lambda)$ is strictly increasing as a function of λ on the interval $(-L, 0)$, and therefore in this interval there can exist at most one root of $b_-(\lambda;x,t)$, regardless of the value of $x \in \mathbb{R}$. Moreover, $b_-(\lambda;x,t) \to +\infty$ as $\lambda \uparrow 0$, so there will be exactly one root in $(-L, 0)$ if $b_-(-L;x,t) = x - x_0 - 2Lt < 0$ and no root in $(-L, 0)$ if $x - x_0 - 2Lt > 0$. Since $b_+(\lambda;x,t) := x + 2\lambda t - x_+(\lambda) < b_-(\lambda;x,t)$ for $-L < \lambda < 0$, if $x - x_0 - 2Lt < 0$, all roots of $b_+(\lambda;x,t)$ in $(-L, 0)$ must lie to the right of the root of $b_-(\lambda;x,t)$. Thus, for $x \in S_t \setminus \{x_0 + 2Lt\}$, we either have

$$
\begin{aligned}
U(x,t) = &\int_{-u_0^{\mathrm{B}}}^{0} (x + 2\lambda t + \gamma(\lambda))d\lambda \\
&+ \sum_{p=1}^{P(x,t)} \left[\pi \int_{-u_{2p-1}^{\mathrm{B}}}^{-u_{2p-2}^{\mathrm{B}}} F(\lambda)d\lambda + \int_{-u_{2p}^{\mathrm{B}}}^{-u_{2p-1}^{\mathrm{B}}} (x + 2\lambda t + \gamma(\lambda))d\lambda \right] \\
&+ \pi \int_{-L}^{-u_{2P(x,t)}^{\mathrm{B}}} F(\lambda)d\lambda, \quad x \in S_t, \quad x > x_0 + 2Lt,
\end{aligned}
\tag{4.136}
$$

in which case

$$
u_0^{\mathrm{B}}(x,t) < u_1^{\mathrm{B}}(x,t) < \cdots < u_{2P(x,t)}^{\mathrm{B}}(x,t)
$$

are all roots of $b_+(-u^{\mathrm{B}};x,t)$, or

$$
\begin{aligned}
U(x,t) = &\int_{-u_0^{\mathrm{B}}}^{0} (x + 2\lambda t + \gamma(\lambda))d\lambda \\
&+ \sum_{p=1}^{P(x,t)} \left[\pi \int_{-u_{2p-1}^{\mathrm{B}}}^{-u_{2p-2}^{\mathrm{B}}} F(\lambda)d\lambda + \int_{-u_{2p}^{\mathrm{B}}}^{-u_{2p-1}^{\mathrm{B}}} (x + 2\lambda t + \gamma(\lambda))d\lambda \right] \\
&- \pi \int_{-L}^{-u_{2P(x,t)}^{\mathrm{B}}} F(\lambda)d\lambda, \quad x \in S_t, \quad x < x_0 + 2Lt,
\end{aligned}
\tag{4.137}
$$

in which case

$$
u_0^{\mathrm{B}}(x,t) < u_1^{\mathrm{B}}(x,t) < \cdots < u_{2P(x,t)-1}^{\mathrm{B}}(x,t)
$$

are roots of $b_+(-u^{\mathrm{B}};x,t)$ while the top branch $u_{2P(x,t)}^{\mathrm{B}}(x,t)$ with $u_{2P(x,t)}^{\mathrm{B}}(x,t) > u_{2P(x,t)-1}^{\mathrm{B}}(x,t)$ is a root of $b_-(-u^{\mathrm{B}};x,t)$. In both cases, the condition $x \in S_t$

guarantees that all roots are differentiable with respect to x, so we may calculate $\partial U/\partial x$ by Leibniz' rule:

(4.138)
$$
\begin{aligned}
\frac{\partial U}{\partial x}(x,t) = {} & b_+(-u^B_{2P}(x,t); x,t)\frac{\partial u^B_{2P}}{\partial x}(x,t) \\
& + \sum_{n=0}^{2P-1}(-1)^n b_+(-u^B_n(x,t); x,t)\frac{\partial u^B_n}{\partial x}(x,t) \\
& + \sum_{n=0}^{2P}(-1)^n u^B_n(x,t), \quad x \in S_t, \ x > x_0 + 2Lt,
\end{aligned}
$$

or

(4.139)
$$
\begin{aligned}
\frac{\partial U}{\partial x}(x,t) = {} & b_-(-u^B_{2P}(x,t); x,t)\frac{\partial u^B_{2P}}{\partial x}(x,t) \\
& + \sum_{n=0}^{2P-1}(-1)^n b_+(-u^B_n(x,t); x,t)\frac{\partial u^B_n}{\partial x}(x,t) \\
& + \sum_{n=0}^{2P}(-1)^n u^B_n(x,t), \quad x \in S_t, \ x < x_0 + 2Lt,
\end{aligned}
$$

where in both cases $P = P(x,t)$ is a constant nonnegative integer on each connected component of S_t. The terms on the first line in each of these formulae arise from differentiating the limits of integration and using $x_\pm(\lambda) = \pm\pi F(\lambda) - \gamma(\lambda)$, while the terms on the second line arise from the explicit partial differentiation of the integrand $x + 2\lambda t + \gamma(\lambda)$ with respect to x. It follows from our division of the solutions of (4.134) among the roots of b_+ and b_- that in both cases the terms on the first line vanish identically, with the result that

(4.140)
$$
\frac{\partial U}{\partial x}(x,t) = \sum_{n=0}^{2P(x,t)}(-1)^n u^B_n(x,t), \quad x \in S_t \setminus \{x_0 + 2Lt\}.
$$

This expression is clearly continuous in x on each connected component of $S_t \setminus \{x_0 + 2Lt\}$. Moreover, it extends continuously to the finite complement in $\mathbb{R}_x$ (at fixed $t \geq 0$) because at caustic curves pairs of solution branches entering into (4.140) with opposite signs simply coalesce. Therefore $U(\cdot, t)$ is indeed continuously differentiable for $t \geq 0$ and its derivative is given by the desired simple formula (4.130). Virtually the same argument applies to $t \leq 0$ with the roles of $b_\pm(\lambda; x,t)$ reversed, and the resulting formula for $\partial U/\partial x$ is the same. $\qquad \square$

It follows from this result that we may integrate by parts in (4.129) and obtain

(4.141)
$$
\lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}}\tilde{u}_\varepsilon(x,t)\phi(x)dx = \int_{\mathbb{R}}\frac{\partial U}{\partial x}(x,t)\phi(x)dx
$$

for every test function $\phi \in \mathscr{D}(\mathbb{R})$. Now let $v \in L^2(\mathbb{R})$. Since $\mathscr{D}(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, for each $\sigma > 0$ there exists a test function $\phi_\sigma \in \mathscr{D}(\mathbb{R})$ such that

$$(4.142) \qquad \|\phi_\sigma - v\|_2^2 := \int_{\mathbb{R}} |\phi_\sigma(x) - v(x)|^2 \, dx < \sigma^2.$$

Then,

$$
\begin{aligned}
\int_{\mathbb{R}} &\left[\tilde{u}_\varepsilon(x,t) - \frac{\partial U}{\partial x}(x,t) \right] v(x) dx \\
&= \int_{\mathbb{R}} \left[\tilde{u}_\varepsilon(x,t) - \frac{\partial U}{\partial x}(x,t) \right] \phi_\sigma(x) dx \\
&\quad + \int_{\mathbb{R}} \frac{\partial U}{\partial x}(x,t)[\phi_\sigma(x) - v(x)] dx \\
&\quad - \int_{\mathbb{R}} \tilde{u}_\varepsilon(x,t)[\phi_\sigma(x) - v(x)] dx.
\end{aligned}
$$

(4.143)

Observe that, according to the definition (see Definition 3.3) of $\tilde{u}_\varepsilon(x,t)$ in terms of the modified scattering data, it follows from (2.19) and (2.18) that

$$(4.144) \qquad \int_{\mathbb{R}} \tilde{u}_\varepsilon(x,t)^2 \, dx = -4\pi\varepsilon \sum_{n=1}^{N(\varepsilon)} \tilde{\lambda}_n.$$

This Riemann sum converges as $\varepsilon \downarrow 0$:

$$(4.145) \qquad \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} \tilde{u}_\varepsilon(x,t)^2 \, dx = -4\pi \int_{-L}^0 \lambda F(\lambda) d\lambda = \int_{\mathbb{R}} u_0(x)^2 \, dx,$$

where the second equality follows from the identities (3.10), which essentially define $F(\lambda)$ in terms of the admissible initial condition u_0. Therefore, $\|\tilde{u}_\varepsilon(\cdot,t)\|_2$ is bounded for sufficiently small ε, independently of t.

Also, $\partial U/\partial x$ is independent of ε and from formula (4.130) it is easy to check that it is positive and bounded above by the constant L for all (x,t). Therefore

$$(4.146) \qquad \left\| \frac{\partial U}{\partial x}(\cdot,t) \right\|_2^2 \leq L \int_{\mathbb{R}} \frac{\partial U}{\partial x}(x,t) dx.$$

By formula (4.130), the latter integral is equal to the area between the graph of the multivalued solution curve for Burgers' equation and the x-axis. Since points on the graph at the same height move with the same speed, this area is independent of time t, and hence we have

$$(4.147) \qquad \left\| \frac{\partial U}{\partial x}(\cdot,t) \right\|_2^2 \leq 2\pi LM,$$

where the mass M is defined in terms of the initial condition u_0 by (3.5). In fact, for $0 \leq t < T$, where T is the breaking time, it follows from the fact that $\partial U/\partial x$ as given by (4.130) reduces to the classical solution $u_0^B(x,t)$ of Burgers' equation with initial data u_0, which conserves exactly the $L^2(\mathbb{R}_x)$ norm, that

$$(4.148) \qquad \left\| \frac{\partial U}{\partial x}(\cdot, t) \right\|_2^2 = \|u_0^B(\cdot, t)\|_2^2 = \int_{\mathbb{R}} u_0(x)^2 \, dx, \quad 0 \leq t < T.$$

We will use this fact below in Section 4.5 when we prove Corollary 1.2. In any case, these considerations show that for all $\varepsilon > 0$ sufficiently small there exists a constant $K > 0$ independent of t such that

$$(4.149) \qquad \left\| \frac{\partial U}{\partial x}(\cdot, t) \right\|_2 + \|\tilde{u}_\varepsilon(\cdot, t)\|_2 \leq K$$

holds for all $t \geq 0$.

Now, by Cauchy-Schwarz it follows that

$$(4.150) \qquad \left| \int_{\mathbb{R}} \frac{\partial U}{\partial x}(x, t)[\phi_\sigma(x) - v(x)]dx \right.$$
$$\left. - \int_{\mathbb{R}} \tilde{u}_\varepsilon(x, t)[\phi_\sigma(x) - v(x)]dx \right| \leq K\|\phi_\sigma - v\|_2.$$

Given $\omega > 0$ arbitrarily small, we then choose $\sigma = \omega/(2M)$ and then (4.143) implies that

$$(4.151) \quad \left| \int_{\mathbb{R}} \left[\tilde{u}_\varepsilon(x, t) - \frac{\partial U}{\partial x}(x, t) \right] v(x)dx \right| \leq$$
$$\left| \int_{\mathbb{R}} \left[\tilde{u}_\varepsilon(x, t) - \frac{\partial U}{\partial x}(x, t) \right] \phi_{\omega/(2M)}(x)dx \right| + \frac{\omega}{2}.$$

Finally, since $\phi_{\omega/(2M)}$ is a test function independent of ε, we may use (4.141) to choose $\varepsilon > 0$ so small that the first term on the right-hand side is less than $\omega/2$.

This proves that

$$(4.152) \qquad \mathrm{w}_x\text{--}\lim_{\varepsilon \downarrow 0} \tilde{u}_\varepsilon(x, t) = \frac{\partial U}{\partial x}(x, t)$$

(weak L^2-convergence) uniformly for t in bounded intervals. Combining (4.130) with (4.152) completes the proof of Theorem 1.1.

4.5 Strong Convergence before Breaking: Proof of Corollary 1.2

To prove Corollary 1.2 we follow closely Lax and Levermore (see theorem 4.5 in part II of [20]). Starting from the identity

$$\|\tilde{u}_\varepsilon(\cdot,t) - u_0^B(\cdot,t)\|_2^2 = \int_{\mathbb{R}} \tilde{u}_\varepsilon(x,t)^2 \, dx + \int_{\mathbb{R}} u_0^B(x,t)^2 \, dx$$

(4.153)

$$- 2 \int_{\mathbb{R}} \tilde{u}_\varepsilon(x,t) u_0^B(x,t) dx,$$

we note that for $0 \le t < T$, where T is the breaking time, (4.145) and (4.148) imply that

(4.154) $$\lim_{\varepsilon\downarrow 0} \|\tilde{u}_\varepsilon(\cdot,t) - u_0^B(\cdot,t)\|_2^2 =$$

$$2 \int_{\mathbb{R}} u_0(x)^2 \, dx - 2 \lim_{\varepsilon\downarrow 0} \int_{\mathbb{R}} \tilde{u}_\varepsilon(x,t) u_0^B(x,t) dx.$$

But $u_0^B(\cdot,t) \in L^2(\mathbb{R})$ is independent of ε, so by Theorem 1.1,

(4.155) $$\lim_{\varepsilon\downarrow 0} \int_{\mathbb{R}} \tilde{u}_\varepsilon(x,t) u_0^B(x,t) dx = \int_{\mathbb{R}} u_0^B(x,t)^2 \, dx = \int_{\mathbb{R}} u_0(x)^2 \, dx,$$

with the second equality following from (4.148) for $0 \le t < T$. Therefore

(4.156) $$\lim_{\varepsilon\downarrow 0} \|\tilde{u}_\varepsilon(\cdot,t) - u_0^B(\cdot,t)\|_2 = 0$$

as desired, and the proof is complete.

5 Numerical Verification

To illustrate the weak convergence of $\tilde{u}_\varepsilon(x,t)$ as guaranteed by Theorem 1.1 and to attempt to empirically quantify the rate of convergence, we have directly used the exact formula (4.3) for $\tilde{U}_\varepsilon(x,t)$, having first chosen the modified scattering data corresponding to the admissible initial condition $u_0(x) = 2(1 + x^2)^{-1}$ as specified in Definition 3.3, and compared the result for several different values of ε with the limiting formula (4.9) for $U(x,t)$.

Our results are shown in Figure 5.1. These plots clearly display the locally uniform convergence specified in Proposition 4.2. An interesting feature is the apparent regular "staircase" form of the graph of $\tilde{U}_\varepsilon(x,t)$ as a function of x; that the steps have nearly equal height is a consequence of the fact that near the leading edge of the oscillation zone for u_ε (which lies approximately in the range $4 < x < 16$ in these plots) the undular bore wavetrain that is generated from the smooth
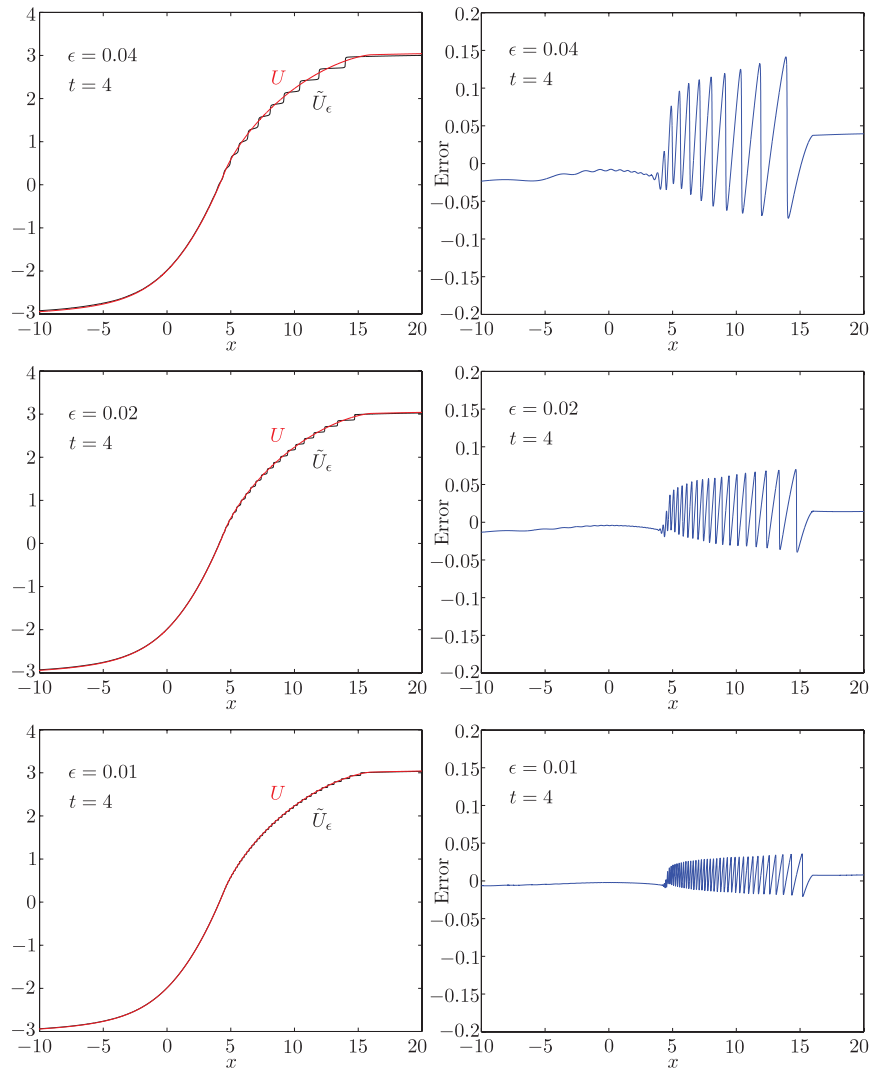
FIGURE 5.1. Left: plots of $\tilde{U}_\varepsilon(x, t)$ (black) and its locally uniform limit $U(x, t)$ (red) at $t = 4$ for various values of ε. For these plots, $u_0(x) := 2(1 + x^2)^{-1}$. Right: corresponding plots of the error $U(x, t) - \tilde{U}_\varepsilon(x, t)$.

initial data resolves into a train of solitons of the BO equation, each of which has a fixed mass proportional to ε (independent of amplitude and velocity).

To the eye, the size of the error between $\tilde{U}_\varepsilon(x, t)$ and $U(x, t)$ appears to scale with ε. To confirm this more quantitatively, we collected numerical data from several experiments, each performed with a different value of ε at the fixed time $t = 4$. The supremum norm, calculated over the interval $-10 < x < 20$, of the
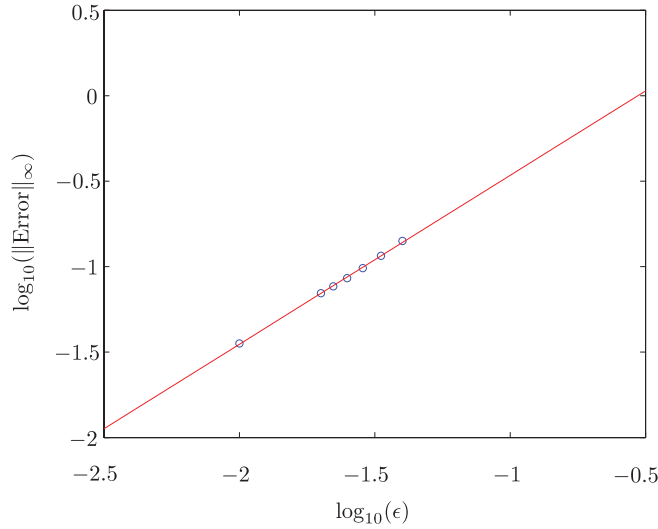
FIGURE 5.2. Circles: $\log_{10}(\|\tilde{U}_\varepsilon(\cdot, 4) - U(\cdot, 4)\|_\infty)$ for $\varepsilon = 1/25$, $1/30$, $1/35$, $1/40$, $1/45$, $1/50$, and $1/100$, as a function of $\log_{10}(\varepsilon)$. In red: The least squares linear fit.

error resulting from each of these experiments is plotted in Figure 5.2. On this plot with logarithmic axes, the data points appear to lie along a straight line, and we calculated the least squares linear fit to the data to be given by

$$(5.1) \qquad \log_{10}(\|\tilde{U}_\varepsilon(\cdot, 4) - U(\cdot, 4)\|_\infty) = 0.988 \log_{10}(\varepsilon) + 0.523$$

where the slope and intercept are given to three significant digits. This strongly suggests a linear rate of convergence in which the error is asymptotically proportional to ε as $\varepsilon \downarrow 0$.

The initial data $u_0(x) = 2(1 + x^2)^{-1}$ was chosen for these experiments because it is the only initial condition (up to a constant multiple) for which the *exact* scattering data is known for a sequence of values of ε tending to zero. This is the result of a calculation of Kodama, Ablowitz, and Satsuma [18], who showed that if $u_0(x) = 2(1 + x^2)^{-1}$, then the reflection coefficient $\beta(\lambda)$ vanishes identically if $\varepsilon = 1/N$ for any positive integer N. Moreover, there are in this case exactly N eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_N$ of the operator $\mathcal{L}$ defined by (1.6), and they are given implicitly by the equation

$$(5.2) \qquad L_N\left(-\frac{2\lambda_n}{\varepsilon}\right) = L_N(-2N\lambda_n) = 0,$$

where L_N is the Laguerre polynomial of degree N.[1] The corresponding phase constants γ_n all vanish exactly. The approximate eigenvalues determined from the initial condition u_0 via formula (3.25) do not agree exactly with the scaled roots of the Laguerre polynomial of degree N (although the approximate phase constants agree exactly with the true phase constants), so it is a worthwhile exercise to compare the function $\tilde{u}_\varepsilon(x, t)$ as specified by Definition 3.3 with the true solution $u_\varepsilon(x, t)$ of the Cauchy problem for the BO equation with initial data $u_0(x) = 2(1 + x^2)^{-1}$. Of course, Corollary 1.2 guarantees strong convergence in L^2 at $t = 0$ (that is, $\tilde{u}_\varepsilon(\cdot, 0)$ is L^2-close to $u_0(\cdot)$), but this alone does not guarantee that $\tilde{u}_\varepsilon(x, t)$ approximates $u_\varepsilon(x, t)$ in any sense for $t > 0$. We made the comparison for several values of $\varepsilon > 0$ corresponding to a reflectionless exact solution of the Cauchy problem constructed[2] from the determinantal formula (2.11) at the time $t = 4$, which is well beyond the breaking time.

Our results are shown in Figure 5.3. These plots show that the modification of the scattering data used to construct $\tilde{u}_\varepsilon(x, t)$ results in a phase shift relative to $u_\varepsilon(x, t)$ that is proportional to ε, the approximate wavelength of the oscillations in the undular bore structure. In particular, $\tilde{u}_\varepsilon(x, t)$ does not remain close to $u_\varepsilon(x, t)$ after the breaking time in any strong sense, although it appears highly likely that convergence is restored in the weak topology.

6 Comparison with Elementary Examples

The key role played in the zero-dispersion limit of the BO Cauchy problem by the multivalued solution of equation (1.8) with the same initial data is reminiscent of two basic example problems from the theory of linear and nonlinear waves.

6.1 Zero-Viscosity Limit of the Viscous Burgers Equation

The Burgers equation with viscosity $\varepsilon > 0$ is the nonlinear wave equation

$$(6.1) \qquad \frac{\partial w_\varepsilon}{\partial t} + 2w_\varepsilon \frac{\partial w_\varepsilon}{\partial x} - \varepsilon \frac{\partial^2 w_\varepsilon}{\partial x^2} = 0, \quad x \in \mathbb{R}, \ t > 0,$$

---

[1] The asymptotic ($N \to \infty$) density of zeros of the scaled Laguerre polynomial $L_N(-2N\lambda)$ is well-known:

$$F(\lambda) = \frac{1}{\pi} \sqrt{\frac{2 + \lambda}{-\lambda}}, \quad -2 < \lambda < 0,$$

a distribution also known in random matrix theory as the Marchenko-Pastur law. This asymptotic formula agrees exactly with Matsuno's formula for $F(\lambda)$ in the case when $u_0(x) = 2(1 + x^2)^{-1}$, which gives some independent justification for its validity.

[2] In fact, this is the numerical method we used to create the plots in Figure 1.1. This has a tremendous advantage over taking a more traditional numerical approach to the Cauchy problem for the BO equation (that is, one involving time stepping) since the calculations necessary to find the solution for any two given values of t are completely independent, so errors do not propagate (and to find the solution for any given time t it is not necessary to perform any calculations at all for intervening times from the initial instant). The only source of error in the use of the determinantal formula (2.11), at least if the differentiation is carried out explicitly resulting in a sum of N determinants, is due to roundoff.
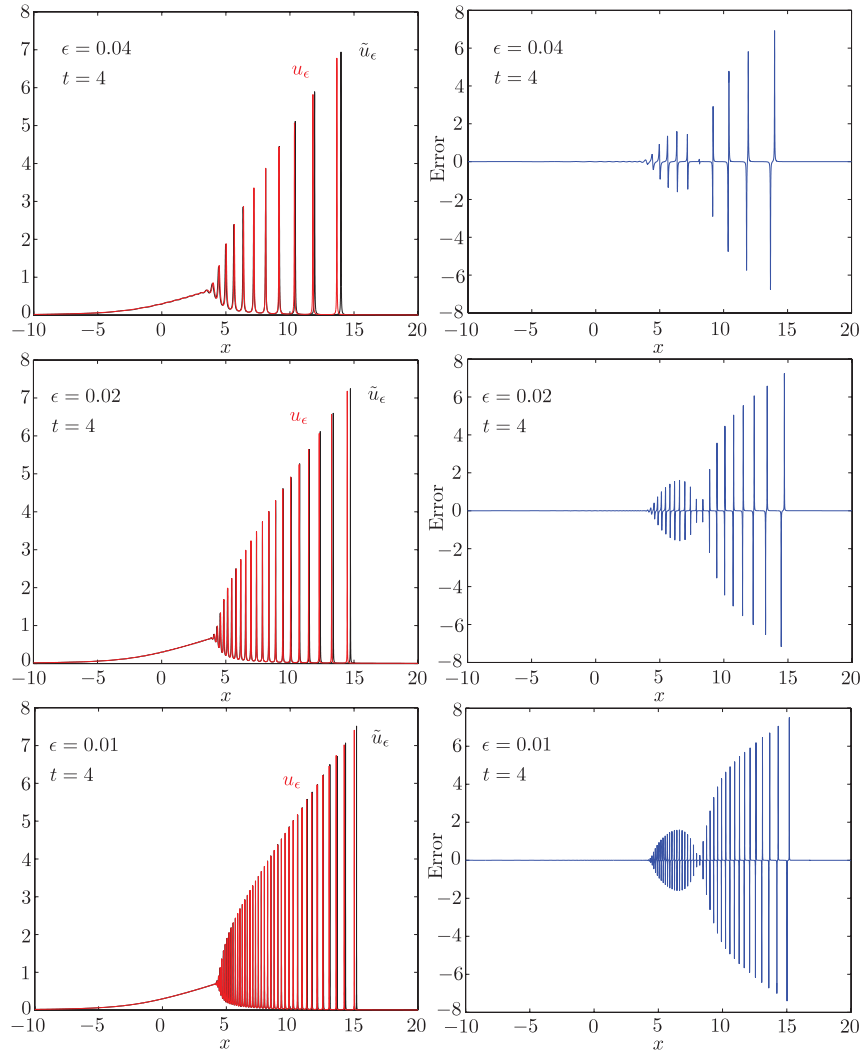
FIGURE 5.3. Left: plots of $\tilde{u}_\varepsilon(x,t)$ (black) shown together with $u_\varepsilon(x,t)$ (red) for the initial data $u_0(x) = 2(1+x^2)^{-1}$ shown for several values of ε at $t = 4$. Right: the error $u_\varepsilon(x,t) - \tilde{u}_\varepsilon(x,t)$.

and we take fixed initial data $w_\varepsilon(x,0) = u_0(x)$. As is well-known, this Cauchy problem is solved by the Cole-Hopf transformation, leading to the exact solution formula

$$(6.2) \qquad w_\varepsilon(x,t) = \frac{1}{2t} \frac{\int_{\mathbb{R}} e^{R(\xi;x,t)/\varepsilon}(x-\xi)d\xi}{\int_{\mathbb{R}} e^{R(\xi;x,t)/\varepsilon}\,d\xi}, \quad t > 0,$$

where the exponent function is defined as

$$(6.3) \qquad R(\xi; x, t) := - \int_0^\xi u_0(\eta)d\eta - \frac{(x - \xi)^2}{4t}.$$

One examines the asymptotic behavior in the limit $\varepsilon \downarrow 0$ by using Laplace's method to analyze the integrals (see [27, sec. 3.6]). The dominant contributions to the integrals come from neighborhoods of points $\xi = \xi(x, t) \in \mathbb{R}$ at which $R(\xi; x, t)$ achieves its maximum value. The critical points of R satisfy $\xi = x - 2u_0(\xi)t$. Writing $u^B = u_0(\xi)$ and applying u_0 to both sides gives equation (1.9), so the critical points correspond to the sheets of the multivalued solution of the (inviscid) Burgers equation (1.8) with initial data u_0. It is easy to check that if x and t are such that there is just one sheet, then the unique critical point is the global maximizer of R, and Laplace's method gives the result that $w_\varepsilon(x, t)$ converges (strongly, pointwise in x and t) to $u^B(x, t)$.

On the other hand, if there are $2P + 1 > 1$ sheets, then for generic (x, t) exactly one of them corresponds to the global maximum of R, and Laplace's method predicts that $w_\varepsilon(x, t)$ will converge to the maximizing sheet. Shocks appear in the small viscosity limit as curves in the (x, t)-plane along which there are jump discontinuities of the pointwise limit corresponding to sudden changes in the choice of sheet that maximizes the exponent R.

To summarize, we have the formula

$$\lim_{\varepsilon \downarrow 0} w_\varepsilon(x, t) = u_n^B(x, t),$$

$$(6.4)$$

$$n = \operatorname*{argmax}_{m = 0, 1, \ldots, 2P(x,t)+1} R(x - 2u_m^B(x, t)t; x, t),$$

for (x, t) not on a shock.

Thus, one sees that for the zero-viscosity limit of the viscous Burgers equation, different sheets of the multivalued solution of the formal limiting Cauchy problem (set $\varepsilon = 0$) provide the strong limit of $w_\varepsilon(x, t)$ for different x and t. However, the choice of sheet requires the solution of a discrete maximization problem parametrized by x and t, making the limiting behavior harder to calculate than the weak zero-dispersion limit of the BO equation.

6.2 Semiclassical Limit of the Free Linear Schrödinger Equation

In this problem, one considers the equation

$$(6.5) \qquad i\varepsilon \frac{\partial \psi_\varepsilon}{\partial t} + 2\varepsilon^2 \frac{\partial^2 \psi_\varepsilon}{\partial x^2} = 0, \quad x \in \mathbb{R}, \ t > 0,$$

for small $\varepsilon > 0$, subject to initial data of WKB form

$$(6.6) \qquad \psi_\varepsilon(x, 0) = A(x)e^{iS(x)/\varepsilon}$$

with A and S real-valued and independent of ε. For suitable A and S, the solution to this problem can be written as an integral

$$\psi_\varepsilon(x,t) = \frac{e^{-i\pi/4}}{\sqrt{8\pi\varepsilon t}} \int_{\mathbb{R}} e^{iI(\xi;x,t)/\varepsilon} A(\xi)\,d\xi,$$

(6.7)

$$t > 0, \quad I(\xi;x,t) := S(\xi) + \frac{(x-\xi)^2}{8t}.$$

The dominant contributions to the solution are calculated via the method of stationary phase (see [27, sec. 5.6]), and these come from small neighborhoods of points ξ satisfying $I'(\xi;x,t) = 0$, that is, solutions ξ of the implicit equation $\xi = x - 4S'(\xi)t$. Evaluating the function $2S'(\cdot)$ on both sides of this equation and making the substitution $u^{\mathrm{B}} = 2S'(\xi)$, one arrives at the equivalent form (1.9) where $u_0(x) := 2S'(x)$. Thus, the branches of the multivalued solution of Burgers' equation (1.8) with initial condition u_0 correspond to stationary phase points ξ that yield the leading term of the solution $\psi_\varepsilon(x,t)$ in the semiclassical limit $\varepsilon \downarrow 0$. Unlike in the analysis of Laplace-type integrals, where only the critical points corresponding to maxima matter in the limit, for oscillatory integrals all stationary phase points contribute to the leading-order behavior, and therefore we have an asymptotic representation of $\psi_\varepsilon(x,t)$ as a sum over branches $u_n^{\mathrm{B}}(x,t)$ of the multivalued solution of Burgers' equation with initial data u_0:

$$(6.8) \qquad \psi_\varepsilon(x,t) = \sum_{n=0}^{2P(x,t)} M_n(x,t) e^{i\theta_n(x,t;\varepsilon)} + O(\varepsilon), \quad t > 0,$$

where $M_n(x,t)$ are slowly varying positive amplitudes given by

$$(6.9) \qquad M_n(x,t) := A(x - 2u_n^{\mathrm{B}}(x,t)t)\sqrt{\left|1 - 2t\frac{\partial u_n^{\mathrm{B}}}{\partial x}(x,t)\right|},$$

and $\theta_n(x,t;\varepsilon)$ are rapidly varying real phases given by

$$(6.10)\ \ \theta_n(x,t;\varepsilon) := \frac{1}{\varepsilon} I(x - 2u_n^{\mathrm{B}}(x,t)t; x,t) + \frac{\pi}{4}\left(\mathrm{sgn}\left(1 - 2t\frac{\partial u_n^{\mathrm{B}}}{\partial x}(x,t)\right) - 1\right)$$

for $n = 0, 1, \ldots, 2P(x,t)$.

A more explicit connection with the multivalued solution of Burgers' equation may be obtained by introducing the quantity

$$(6.11) \qquad w_\varepsilon(x,t) := 2\varepsilon\frac{\partial}{\partial x}\mathrm{Im}\{\log(\psi_\varepsilon(x,t))\},$$

which is the fluid velocity in Madelung's interpretation of the wave function ψ_ε as describing a quantum-corrected fluid motion. Under the condition that the error term in (6.8) becomes $O(1)$ after differentiation with respect to x, some easy

calculations show that (6.8) implies

$$(6.12) \qquad w_\varepsilon(x,t) = \mathrm{Re}\left\{ \frac{\sum_{n=0}^{2P(x,t)} u_n^B(x,t) M_n(x,t) e^{i\theta_n(x,t;\varepsilon)}}{\sum_{n=0}^{2P(x,t)} M_n(x,t) e^{i\theta_n(x,t;\varepsilon)}} \right\} + O(\varepsilon).$$

It is then easy to see that if $P(x,t) = 0$, $w_\varepsilon(x,t)$ converges strongly pointwise to $u_0^B(x,t)$, the unique solution (for this x and t, anyway) of Burgers' equation. On the other hand, if $P(x,t) > 0$, then there are interference effects among the terms in the sums, and these lead to rapid oscillations with the effect that $w_\varepsilon(x,t)$ no longer converges in the pointwise sense as $\varepsilon \downarrow 0$.

However, it does converge in the weak topology. The weak limit may be computed by multiphase averaging, which we illustrate in the case $P(x,t) = 1$. The procedure is to average the leading term in $w_\varepsilon(x,t)$ over an interval in x centered at the point of interest of radius, say, ε^p for some $p \in (0,1)$, and then pass to the limit $\varepsilon \downarrow 0$. This produces the desired local average over rapid oscillations of wavelength or period proportional to ε. Under an ergodic hypothesis that is valid on a set of full measure in the (x,t)-plane, this procedure is equivalent to holding u_n^B and M_n fixed and averaging (with uniform measure) over the torus of relative angles $\phi_1 := \theta_1 - \theta_0$ and $\phi_2 := \theta_2 - \theta_0$. The double integrals can be evaluated explicitly, with the result that

$$(6.13) \qquad \mathrm{w}_x{-}\lim_{\varepsilon \downarrow 0} w_\varepsilon(x,t) = \sum_{n=0}^{2} c_n(x,t) u_n^B(x,t),$$

where $c_n(x,t)$, $n = 0,1,2$, are nonnegative weights having the property that $c_0(x,t) + c_1(x,t) + c_2(x,t) = 1$.

Specifically, the coefficients only depend on x and t through $M_0(x,t)$, $M_1(x,t)$, and $M_2(x,t)$. If any of these, say M_n, exceeds the sum of the other two, then $c_n = 1$ and the two other coefficients vanish. Thus the weak limit produces in this case exactly the branch $u_n^B(x,t)$ through multiphase averaging.[3] On the other hand, if none of the M_n exceeds the sum of the other two, then M_0, M_1, and M_2

---

[3] The strict inequality $M_0(x,t) > M_1(x,t) + M_2(x,t)$ or a permutation thereof, defining this situation, is an open condition on $(x,t) \in \mathbb{R}^2$, and therefore (depending on initial conditions) there can exist open domains in the (x,t)-plane on which the weak limit of $w_\varepsilon(x,t)$ is given by a single branch of the solution of the inviscid Burgers equation while $w_\varepsilon(x,t)$ itself exhibits wild oscillations.

Interestingly, this is precisely the conjecture made by von Neumann regarding grid-scale oscillations observed in the numerical solution of Burgers' equation via a finite-difference scheme (which may be viewed as a dispersive regularization of the equation). While many model equations for finite-difference schemes (the KdV equation is one example) do not yield such a simple interpretation of the weak limit [19], it would seem that von Neumann's conjecture can hold true if the Schrödinger equation is viewed as a dispersive correction to Burgers' equation.

are the side lengths of a triangle, and the weak limit is a genuine weighted average of the three branches, with weights proportional to the opposite angles:

(6.14)
$$c_0 = \frac{1}{\pi} \arccos\left(\frac{M_1^2 + M_2^2 - M_0^2}{2M_1 M_2}\right),$$
$$c_1 = \frac{1}{\pi} \arccos\left(\frac{M_0^2 + M_2^2 - M_1^2}{2M_0 M_2}\right),$$
$$c_2 = \frac{1}{\pi} \arccos\left(\frac{M_0^2 + M_1^2 - M_2^2}{2M_0 M_1}\right).$$

The most significant aspect of this analysis is that the weak limit depends on information other than just the initial condition u_0 for Burgers' equation since the functions $M_n(x,t)$ also involve the initial wave function amplitude A. This makes the evaluation of the weak limit a more complicated procedure than in the case of the BO equation.

A further connection between the BO equation (1.1) and the linear Schrödinger equation (6.5) in the zero-dispersion limit has already been pointed out in Section 2.4.

7 Conclusion

In this paper, we have obtained the first rigorous results regarding the zero-dispersion limit of the Cauchy problem for the BO equation. As suggested by the formal multiphase averaging of modulated N-phase wavetrain solutions carried out by Dobrokhotov and Krichever [9], the scalar inviscid Burgers' equation *and its multivalued solution after wave breaking* characterize the limit.

To analyze the BO Cauchy problem, we used a remarkable formula of Matsuno [22, 23] for the density of eigenvalues of the nonlocal operator $\mathcal{L}$ appearing in the scattering theory, and we have proposed a new asymptotic formula (3.21) for the corresponding phase constants necessary to set up the inverse scattering problem. We then developed an analogue of the Lax-Levermore method [20] to study the inverse scattering problem, and we obtained an explicit formula (4.10) for a measure μ with density $G(\alpha; x, t)$ that is the BO equivalent of the extremal measure in Lax-Levermore theory. By contrast with the KdV case, the formula we obtain for the weak limit from this measure is remarkably simple and explicit, writing the weak limit as a signed sum of branches of the multivalued solution of Burgers' equation.

A useful generalization of the weak limit given in this paper is to consider the weak limits of all of the variational derivatives

(7.1)
$$K_n := \frac{\delta I_n}{\delta u}, \quad n = 3, 4, 5, \ldots,$$

where the I_n is the nth conserved quantity given by (2.17). These functional gradients generate the Benjamin-Ono hierarchy as Hamiltonian equations of the form

$$(7.2) \qquad \frac{\partial u}{\partial t_n} + \frac{\partial K_{n+2}}{\partial x} = 0, \quad n = 1, 2, 3, \ldots .$$

See [3] for more details.

It is a consequence of some concrete calculations that can be found in [24] that at least for the first several values of n, the quantities K_n corresponding to the solution $\tilde{u}_\varepsilon(x, t)$ can be written in the determinantal form

$$(7.3) \qquad K_{n+2} = \frac{\partial}{\partial t_n} \tilde{U}_\varepsilon(x, t_1, t_2, \ldots, t_n)$$

where after the differentiation t_1 is set equal to t and all t_k for $k > 1$ are set to zero (to consider the evolution of K_{n+2} as $\tilde{u}_\varepsilon$ varies according to the first BO equation, (7.2) with $n = 1$). Here $\tilde{U}_\varepsilon(x, t_1, t_2, \ldots, t_n)$ is given by (4.3) with each occurrence of $2t$ in the expression $-2\tilde{\lambda}_k(x + 2\tilde{\lambda}_k t)$ in the diagonal matrix elements of $\tilde{\mathbf{A}}_\varepsilon$ replaced by $2t_1 - 3\tilde{\lambda}_k t_2 + 4\tilde{\lambda}_k^2 t_3 - 5\tilde{\lambda}_k^3 t_4 + \cdots + (-1)^{n+1}(n + 1)\tilde{\lambda}_k^{n-1} t_n$. It is a reasonable conjecture that (7.3) in fact holds for all $n \geq 1$. Formulae for these higher weak limits can also be obtained within the framework of our method and will be published in a subsequent paper.

We are also currently investigating prospects for strengthening the limit after wave breaking occurs. The goal here is to rigorously establish an asymptotic formula for $\tilde{u}_\varepsilon(x, t)$ that is valid *pointwise* for (x, t) in the oscillation zone. Such a formula should accurately resolve the microscopic (wavelength proportional to ε) oscillations, including finding the phase up to error terms that are bounded by a vanishingly small fraction of the wavelength. One expects the asymptotic form of the wavetrain to be given by the rational-exponential formulae found by Dobrokhotov and Krichever [9]. For the KdV equation such pointwise asymptotics have been obtained [8] using the Deift-Zhou steepest-descent technique for matrix-valued Riemann-Hilbert problems. We are working to extend this kind of methodology to the context of scalar Riemann-Hilbert problems with *nonlocal* jump conditions, as occurs in the inverse scattering transform (generally with nonvanishing reflection coefficient) for the BO equation.

It has been recently conjectured by Dubrovin [10] that near the earliest breaking point (x_ξ, t_ξ) the solution of the Cauchy problem for quite general weakly dispersive Hamiltonian perturbations of Burgers' equation should exhibit a universal form expressed in terms of Painlevé transcendents. This conjecture has been confirmed for general initial data for the KdV equation (as a particular case of a perturbation considered by Dubrovin) by Claeys and Grava [5]. It would be interesting to determine by direct calculation of the solution near the breaking point (x_ξ, t_ξ) whether the BO equation should be considered to fall within the universality class of equations conjectured by Dubrovin or whether the BO equation represents a new universality class.

Appendix: Proof of a Cauchy Integral Identity

Here we prove that whenever $f \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$, the identity

$$(A.1) \qquad \int_{\mathbb{R}} f(x)\mathcal{C}_+\big[f\mathcal{C}_+[f\mathcal{C}_+[\cdots f\mathcal{C}_+[f]\cdots]]\big](x)dx = \frac{1}{k}\int_{\mathbb{R}} f(x)^k\,dx$$

holds for all $k = 1, 2, 3, \ldots$, where the Cauchy projector $\mathcal{C}_+$ defined in (1.6) occurs $k - 1$ times in the integrand on the left-hand side.

Let H_k denote the left-hand side of (A.1). We write H_k in the form

$$(A.2) \qquad H_k = \int_{\mathbb{R}} f(x)J_k^+(x)dx, \quad k = 1, 2, 3, \ldots,$$

where the function $J_k^+(x)$ is given by the following recurrence relation:

$$(A.3) \qquad J_1^+(x) := 1; \quad J_{k+1}^+(x) := \mathcal{C}_+[fJ_k^+](x), \quad k = 1, 2, 3, \ldots.$$

Clearly, $J_k^+(x)$ is an element of the Hardy space $\mathbb{H}^+(\mathbb{R})$ of the upper half-plane for $k > 1$. A corresponding sequence of functions belonging to the Hardy space $\mathbb{H}^-(\mathbb{R})$ of the lower half-plane for $k > 1$ is generated by the following recurrence:

$$(A.4) \qquad J_1^-(x) := 1, \quad J_{k+1}^-(x) := \mathcal{C}_-[fJ_k^-](x), \quad k = 1, 2, 3, \ldots,$$

where the operator $\mathcal{C}_-$ is defined by

$$(A.5) \qquad \mathcal{C}_-[f](x) := \lim_{\delta \downarrow 0} \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{f(y)}{y - x + i\delta}\,dy.$$

Therefore, we may equivalently express H_k in the form

$$(A.6) \qquad H_k = \int_{\mathbb{R}} J_1^-(x)f(x)J_k^+(x)dx, \quad k = 1, 2, 3, \ldots.$$

The Plemelj formula $\mathcal{C}_+ - \mathcal{C}_- = \mathbb{I}$ yields that for any $u, v \in L^2(\mathbb{R})$,

$$(A.7) \qquad \int_{\mathbb{R}} u(x)\mathcal{C}_+[v](x)dx + \int_{\mathbb{R}} \mathcal{C}_-[u](x)v(x)dx =$$

$$\int_{\mathbb{R}} \mathcal{C}_+[u](x)\mathcal{C}_+[v](x)dx - \int_{\mathbb{R}} \mathcal{C}_-[u](x)\mathcal{C}_-[v](x)dx,$$

and the latter two integrals vanish by deformation of the contours to infinity in the respective half-planes of analyticity where the integrands are $O(1/x^2)$. Therefore, for $1 \le s \le k - 1$,

$$(A.8) \qquad \int_{\mathbb{R}} J_s^-(x)f(x)J_{k+1-s}^+(x)dx = -\int_{\mathbb{R}} J_{s+1}^-(x)f(x)J_{k-s}^+(x)dx.$$

Comparing this equation with (A.6), we infer that for any s in the range $1 \leq s \leq k$, H_k may also be written in the form

$$(A.9) \qquad H_k = (-1)^{s-1} \int_{\mathbb{R}} J_s^-(x) f(x) J_{k+1-s}^+(x) dx.$$

Averaging over s then gives

$$(A.10) \qquad H_k = \frac{1}{k} \int_{\mathbb{R}} f(x) S_k(x) dx$$

where S_k denotes the sum

$$(A.11) \qquad S_k(x) := \sum_{j=1}^{k} (-1)^{k-j} J_j^+(x) J_{k+1-j}^-(x).$$

We will now give an inductive proof that $S_k = f^{k-1}$ for $k = 1, 2, 3, \ldots$ (from now on since we are no longer dealing with integrals, we suppress all x dependence). It is obvious that $S_1 = 1$, and it follows directly from the Plemelj formula that $S_2 = f$. We let $k \geq 2$ and invoke the inductive hypothesis that $S_m = f^{m-1}$ for $m = 1, 2, \ldots, k$. Extracting the last term in the sum and using the recursion relation (A.4), we write S_{k+1} in the form

$$(A.12) \qquad \begin{aligned} S_{k+1} &= J_{k+1}^+ + \sum_{j=1}^{k} (-1)^{k+1-j} J_j^+ J_{k+2-j}^- \\ &= J_{k+1}^+ + \sum_{j=1}^{k} (-1)^{k+1-j} \mathcal{C}_-[f J_{k+1-j}^-] J_j^+. \end{aligned}$$

We will now write the term J_{k+1}^+ as a certain sum. Let $1 \leq s \leq k - 1$. Then by the Plemelj formula applied to $f J_s^-$,

$$(A.13) \qquad \begin{aligned} \mathcal{C}_+[f J_s^- J_{k+1-s}^+] &= \mathcal{C}_+[\mathcal{C}_+[f J_s^-] J_{k+1-s}^+] - \mathcal{C}_+[\mathcal{C}_-[f J_s^-] J_{k+1-s}^+] \\ &= \mathcal{C}_+[\mathcal{C}_+[f J_s^-] J_{k+1-s}^+] - \mathcal{C}_+[J_{s+1}^- J_{k+1-s}^+], \end{aligned}$$

where in the second line we have recalled the recursion relation (A.4). Now, $\mathcal{C}_+[f J_s^-] J_{k+1-s}^+ \in \mathbb{H}^+(\mathbb{R})$, on which the projection $\mathcal{C}_+$ acts as the identity, so (A.13) becomes

$$(A.14) \qquad \begin{aligned} \mathcal{C}_+[f J_s^- J_{k+1-s}^+] &= \mathcal{C}_+[f J_s^-] J_{k+1-s}^+ - \mathcal{C}_+[J_{s+1}^- J_{k+1-s}^+] \\ &= \mathcal{C}_+[f J_s^-] J_{k+1-s}^+ - \mathcal{C}_+[J_{s+1}^- \mathcal{C}_+[f J_{k-s}^+]] \\ &= -\mathcal{C}_+[f J_{s+1}^- J_{k-s}^+] + \mathcal{C}_+[f J_s^-] J_{k+1-s}^+ \\ &\quad - \mathcal{C}_+[J_{s+1}^- \mathcal{C}_-[f J_{k-s}^+]], \end{aligned}$$

where in the second line we have replaced J^+_{k+1-s} in the second term according to the recursion relation (A.3), and in the third line we have applied the Plemelj formula to fJ^+_{k-s}. Now $J^-_{s+1}\mathcal{C}_-[fJ^+_{k-s}] \in \mathbb{H}^-(\mathbb{R})$, the kernel of the projection $-\mathcal{C}_+$, so (A.14) yields the recursion identity

$$(A.15) \qquad \mathcal{C}_+[fJ^-_s J^+_{k+1-s}] = -\mathcal{C}_+[fJ^-_{s+1}J^+_{k-s}] + \mathcal{C}_+[fJ^-_s]J^+_{k+1-s}.$$

Recalling (A.3) and (A.4),

$$(A.16) \qquad\qquad \mathcal{C}_+[fJ^-_s J^+_{k+1-s}] = J^+_{k+1} \quad \text{if } s = 1$$

and

$$(A.17) \qquad\qquad \mathcal{C}_+[fJ^-_{s+1}J^+_{k-s}] = \mathcal{C}_+[fJ^-_k]J^+_1 \quad \text{if } s = k - 1.$$

From the recursion (A.15) and the boundary conditions (A.16)–(A.17), we obtain the identity

$$(A.18) \qquad\qquad J^+_{k+1} = \sum_{s=1}^{k}(-1)^{s-1}\mathcal{C}_+[fJ^-_s]J^+_{k+1-s}.$$

Changing the index of summation by $j = k + 1 - s$ and substituting into (A.12) yields

$$
\begin{aligned}
S_{k+1} &= \sum_{j=1}^{k}(-1)^{k-j}(\mathcal{C}_+[fJ^-_{k+1-j}] - \mathcal{C}_-[fJ^-_{k+1-j}])J^+_j \\
&= f\sum_{j=1}^{k}(-1)^{k-j}J^+_j J^-_{k+1-j} \\
&= fS_k
\end{aligned}
$$
(A.19)

where in the second and third lines, respectively, we have used the Plemelj formula and recalled definition (A.11) of S_k. Since by the inductive hypothesis $S_m = f^{m-1}$ for $1 \le m \le k$, we have proved that $S_k = f^{k-1}$ for all $k \ge 1$.

Substitution of $S_k(x) = f(x)^{k-1}$ into (A.10) then completes the proof of the identity (A.1).

Bibliography

[1] Benjamin, T. B. Internal waves of permanent form in fluids of great depth. *J. Fluid Mech.* **29** (1967), part 3, 559–592.

[2] Bock, T. L.; Kruskal, M. D. A two-parameter Miura transformation of the Benjamin-Ono equation. *Phys. Lett.* **74** (1979), no. 3-4, 173–176.

[3] Case, K. M. Benjamin-Ono-related equations and their solutions. *Proc. Nat. Acad. Sci. U.S.A.* **76** (1979), no. 1, 1–3.

[4] Choi, W.; Camassa, R. Weakly nonlinear internal waves in a two-fluid system. *J. Fluid. Mech.* **313** (1996), 83–103.

[5] Claeys, T.; Grava, T. Universality of the break-up profile for the KdV equation in the small dispersion limit using the Riemann-Hilbert approach. *Comm. Math. Phys.* **286** (2009), no. 3, 979–1009.

[6] Coifman, R. R.; Wickerhauser, M. V. The scattering transform for the Benjamin-Ono equation. *Inverse Problems* **6** (1990), no. 5, 825–861.

[7] Davis, R. E.; Acrivos, A. Solitary internal waves in deep water. *J. Fluid Mech.* **29** (1967), part 3, 593–607.

[8] Deift, P.; Venakides, S.; Zhou, X. New results in small dispersion KdV by an extension of the steepest descent method for Riemann-Hilbert problems. *Internat. Math. Res. Notices* **1997** (1997), no. 6, 286–299.

[9] Dobrokhotov, S. Yu.; Krichever, I. M. Multi-phase solutions of the Benjamin-Ono equation and their averaging. *Mat. Zametki* **49** (1991), no. 6, 42–58; English translation in *Math. Notes* **49** (1991), no. 6, 583–594.

[10] Dubrovin, B. A. On universality of critical behaviour in Hamiltonian PDEs. *Geometry, topology, and mathematical physics*, 59–109. American Mathematical Society Translations, Series 2, 224. American Mathematical Society, Providence, R.I., 2008,

[11] Dubrovin, B. A.; Matveev, V. B.; Novikov, S. P. Non-linear equations of Korteweg–de Vries type, finite-zone linear operators, and Abelian varieties. *Russ. Math. Surv*. **31** (1976), no. 1, 59–146.

[12] Flaschka, H.; Forest, M. G.; McLaughlin, D. W. Multiphase averaging and the inverse spectral solution of the Korteweg–de Vries equation. *Comm. Pure Appl. Math.* **33** (1980), no. 6, 739–784.

[13] Fokas, A. S.; Ablowitz, M. J. The inverse scattering transform for the Benjamin-Ono equation—a pivot to multidimensional problems. *Stud. Appl. Math.* **68** (1983), no. 1, 1–10.

[14] Gurevich, A. V.; Pitaevskii, L. P. Nonstationary structure of a collisionless shock wave. *Zh. Èksper. Teoret. Fiz.* **65** (1973), 590–604; English translation in *Sov. Phys. JETP* **38** (1974), 291–297.

[15] Its, A. R.; Matveev, V. B. Hill operators with a finite number of lacunae. *Funkcional. Anal. i Priložen*. **9** (1975), no. 1, 69–70; English translation: *Functional Anal. Appl.* **9** (1975), no. 1, 65–66.

[16] Jorge, M. C.; Minzoni, A. A.; Smyth, N. F. Modulation solutions for the Benjamin-Ono equation. *Phys. D* **132** (1999), no. 1-2, 1–18.

[17] Kaup, D. J.; Matsuno, Y. The inverse scattering transform for the Benjamin-Ono equation. *Stud. Appl. Math.* **101** (1998), no. 1, 73–98.

[18] Kodama, Y.; Ablowitz, M. J.; Satsuma, J. Direct and inverse scattering problems of the nonlinear intermediate long wave equation. *J. Math. Phys.* **23** (1982), no. 4, 564–576.

[19] Lax, P. D. On dispersive difference schemes. Solitons and coherent structures (Santa Barbara, Calif., 1985). *Phys. D* **18** (1986), no. 1-3, 250–254.

[20] Lax, P. D.; Levermore, C. D. The small dispersion limit of the Korteweg–de Vries equation. I. II. III. *Comm. Pure Appl. Math.* **36** (1983), no. 3, 253–290; no. 5, 571–593; no. 6, 809–929.

[21] Matsuno, Y. Exact multi-soliton solution of the Benjamin-Ono equation. *J. Phys. A: Math. Gen.* **12** (1979), 619–621.

[22] Matsuno, Y. Number density function of Benjamin-Ono solitons. *Phys. Lett. A* **87** (1981/82), no. 1-2, 15–17.

[23] Matsuno, Y. Asymptotic properties of the Benjamin-Ono equation. *J. Phys. Soc. Japan* **51** (1982), 667–674.

[24] Matsuno, Y. *Bilinear transformation method*. Mathematics in Science and Engineering, 174. Academic, Orlando, Fla., 1984.

[25] Matsuno, Y. Nonlinear modulation of periodic waves in the small dispersion limit of the Benjamin-Ono equation. *Phys. Rev. E (3)* **58** (1998), no. 6, part B, 7934–7940.

[26] Matsuno, Y. The small dispersion limit of the Benjamin-Ono equation and the evolution of a step initial condition. *J. Phys. Soc. Japan* **67** (1998), no. 6, 1814–1817.

[27] Miller, P. D. *Applied asymptotic analysis.* Graduate Studies in Mathematics, 75. American Mathematical Society, Providence, R.I., 2006.

[28] Nakamura, A. Bäcklund transform and conservation laws of the Benjamin-Ono equation. *J. Phys. Soc. Japan* **47** (1979), no. 4, 1335–1340.

[29] Ono, H. Algebraic solitary waves in stratified fluids. *J. Phys. Soc. Japan* **39** (1975), no. 4, 1082–1091.

[30] Porter, A.; Smyth, N. F. Modelling the morning glory of the Gulf of Carpentaria. *J. Fluid Mech.* **454** (2002), 1–20.

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

[32] Whitham, G. B. Non-linear dispersive waves. *Proc. Roy. Soc. Ser. A* **283** (1965), 238–261.

[33] Wigner, E. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)* **62** (1955), 548–564.

[34] Wigner, E. P. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)* **67** (1958), 325–327.

PETER D. MILLER
University of Michigan
Department of Mathematics
East Hall
530 Church Street
Ann Arbor, MI 48109
E-mail: millerpd@umich.edu

ZHENGJIE XU
University of Michigan
Department of Mathematics
East Hall
530 Church Street
Ann Arbor, MI 48109
E-mail: zhengjxu@umich.edu