HIGHER ORDER MODULATION EQUATIONS FOR A BOUSSINESQ EQUATION

C. EUGENE WAYNE AND J. DOUGLAS WRIGHT

ABSTRACT. In order to investigate corrections to the common KdV approximation to long waves, we derive modulation equations for the evolution of long wavelength initial data for a Boussinesq equation. The equations governing the corrections to the KdV approximation are explicitly solvable and we prove estimates showing that they do indeed give a significantly better approximation than the KdV equation alone. We also present the results of numerical experiments which show that the error estimates we derive are essentially optimal.

1. INTRODUCTION

Modulation, or amplitude, equations are approximate, often explicitly solvable, model equations derived—usually through asymptotic analysis and the method of multiple time scales—to model more complicated physical situations. Although these equations have been been used for over a century, only lately has there been an attempt to rigorously relate solutions of the modulation equations to the original physical problem. In particular, through the work of Craig [7], Kano and Nishida [13], Kalyakin [12], Schneider [21], Ben Youssef and Colin [1] and Schneider and Wayne [22], [24], the validity of Korteweg-de Vries (KdV) equations as a leading order approximation to the evolution of long wavelength water waves and to a number of other dispersive partial differential equations has been established.

While the KdV approximation is extremely useful due to its simplicity and the fact that the KdV equation can be explicitly solved by the inverse scattering transform, both experimentally and numerically one observes departures from the predictions of the KdV equation. Our goal in this paper is to derive modulation equations which govern corrections to the KdV model. In the present paper we will not work with the full water wave problem but rather study modulation equations for long wavelength solutions of the Boussinesq equation:

$$\theta_{tt} - \theta_{xx} = (\theta^2)_{xx} + \theta_{ttxx},$$
(1)
$$x \in \mathbb{R},\ t \geq 0,\ \theta(x,t) \in \mathbb{R}.$$

Our motivation for studying this equation is twofold. First, the Boussinesq equation was originally derived as a model equation for water waves, and as our ultimate goal is to derive corrections to the KdV approximation to water waves, we regard the study of (1) as a useful first step in understanding the much more complicated water wave situation. We note that Schneider's analysis of the KdV approximation for (1) in [21] served as a template for the analysis of the water wave problem in [22].

Our second justification for deriving second order modulation equations for (1) is that these modulation equations serve as a sort of normal form for more complicated PDEs, and as such we expect that the modulation equations which describe corrections to the KdV approximation to (1) will also govern corrections to the KdV approximation in more complicated situations as well. Thus the results on existence, uniqueness and other properties of the modulation equations we derive in this paper should also be of use in more complicated situations that we plan to treat in the future.

We now describe in more detail our results. It is convenient to rewrite (1) as a system of two first order equations. As in [21] we introduce new variables,

$$
\begin{aligned}
u(x,t) &= \frac{1}{2}(\theta(x,t) - \lambda^{-1}\theta_t(x,t)) \\
v(x,t) &= \frac{1}{2}(\theta(x,t) + \lambda^{-1}\theta_t(x,t))
\end{aligned}
\tag{2}
$$

where λ is a skew-symmetric multiplication operator in Fourier space defined by $\widehat{\lambda u} = (ik/\sqrt{1+k^2})\hat{u}$. Note that for $\lambda^{-1}\theta_t$ to be well-defined, we must have $\hat{\theta}_t(0,t) = 0$. That is, the average value of θ_t should be zero. We note that for $\theta(x,t)$ a solution of (1), we have that the average value of $\theta_t(x,t)$ is a constant of the motion. Thus, if the initial condition for θ_t has zero average, $\hat{\theta}_t(0,t) = 0$ will remain zero for all time. Furthermore, as discussed in [21], assuming that the initial condition $\theta_t(x,0)$ has zero average is not unnatural considering the origin of (1). Thus, we will make that assumption so that the change of variables (2) is well-defined.

Taking time derivatives of u and v we find:

$$
\partial_t \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -\lambda & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \frac{1}{2}\begin{pmatrix} -\lambda(u+v)^2 \\ \lambda(u+v)^2 \end{pmatrix}
\tag{3}
$$

Not only is (3) convenient from a mathematical point of view but, as we shall see, u and v have the physical interpretation of being the left and right moving parts of the solution.

We turn now to the assumptions on the initial conditions of (3). The KdV equation is an approximation of small amplitude and long wavelength motions, and thus we will assume that the initial conditions of (1) are of this form. More precisely, fix a constant $C_I > 0$ and assume:

Hypothesis 1. *There exist U_0, V_0 with*

$$
\max\left\{\|U_0\|_{H^s(4)\cap H^{s+9}}, \|V_0\|_{H^s(4)\cap H^{s+9}}\right\} < C_I
$$

such that the initial conditions of (3) are of the form:

$$
u(x,0) = \epsilon^2 U_0(\epsilon x), \quad v(x,0) = \epsilon^2 V_0(\epsilon x),
$$

where ϵ is small.

Here, $H^s(m) = \{f|(1+x^2)^{m/2}f \in H^s\}$. The norm on this weighted Sobolev space is given by $\|f\|_{H^s(m)} = \|(1+x^2)^{m/2}f\|_{H^s}$. We use these spaces because we are interested in solutions which are in some (weak) sense "localized". In particular, any small perturbation of the known soliton solutions to the KdV equation will satisfy this localization property.

Remark 1. *We can, of course, recover the initial conditions for (1) from Hypothesis 1 via (2), and we see that the initial conditions expressed in the θ variables are also of small amplitude, long-wavelength form.*

According to the KdV approximation results of [21], long wavelength solutions of (1) split up into two pieces, one a right moving wave train and one a left moving wave train. Each of these wave trains evolves according to a KdV equation and there is no interaction between the left and right moving pieces. One might expect two types of corrections to such an approximation:

- corrections due the fact that the left and right moving wave trains will interact at higher order.
- corrections due to the fact that even in the case of a purely right (or left) moving wave train, solutions to the Boussinesq equation are not exactly described by solutions to the KdV equation.

Both of these types of corrections are apparent in our results and in fact the corrections to the KdV approximation are a sum of solutions of two types of modulation equations; an inhomogeneous transport equation and a linearized KdV equation which can be seen (roughly speaking) as reflecting these two sources of corrections.

To incorporate these two types of corrections, we add to the KdV wavetrains, which we denote by U and V (since they represent the leading terms in u and v respectively), additional functions A and B and F and G. These functions then satsisfy the modulation equations:

$$(4) \qquad \begin{aligned} \partial_T U &= -\frac{1}{2}\partial_{X_-}^3 U - \frac{1}{2}\partial_{X_-} U^2 \\ \partial_T V &= \frac{1}{2}\partial_{X_+}^3 V + \frac{1}{2}\partial_{X_+} V^2 \end{aligned}$$

$$(5) \qquad \begin{aligned} \partial_\tau A + \partial_X A &= -\frac{1}{2}\partial_{X_+} V^2(X+\tau,\epsilon^2\tau) - \partial_X U(X-\tau,\epsilon^2\tau)V(X+\tau,\epsilon^2\tau) \\ \partial_\tau B - \partial_X B &= \frac{1}{2}\partial_{X_-} U^2(X-\tau,\epsilon^2\tau) + \partial_X U(X-\tau,\epsilon^2\tau)V(X+\tau,\epsilon^2\tau) \end{aligned}$$

and

$$(6) \qquad \begin{aligned} \partial_T F &= -\partial_{X_-}(UF) - \frac{1}{2}\partial_{X_-}^3 F + J^1 \\ \partial_T G &= \partial_{X_+}(VG) + \frac{1}{2}\partial_{X_+}^3 G + J^2 \end{aligned}$$

where $T = \epsilon^3 t$, $\tau = \epsilon\tau$, $X = \epsilon x$ and $X_\pm = X \pm \tau$.

The first of these pairs of equations is simply the KdV approximation. The second and third pairs give rise to the corrections to the KdV approximation. We note that the terms J^1 and J^2 which appear in (6) are inhomogeneous terms which are made up of a combination of sums and products of the solutions to (4), (5) and their derivatives (see equations (15)-(16) below).

There is some freedom in how we choose the initial data for the modulation equations. For simplicity we assume that $U(X,0) = U_0(X)$, $V(X,0) = V_0(X)$ and choose zero initial data for (5) and (6), i.e. $A(X,0) = B(X,0) = F(X,0) = G(X,0) = 0$.

That the KdV equation has solutions for all times with initial data of the type described is well known. In particular one has (see [22]):

Theorem 1. *Let $s \geq 4$. Then for all $C_0, T_0 > 0$ there exists $C_1 > 0$ such that if U, V satisfy (4) with initial conditions U_0, V_0 and*

$$\max\{\|U_0\|_{H^s(4)\cap H^{s+9}}, \|V_0\|_{H^s(4)\cap H^{s+9}}\} < C_0$$

then

$$(7) \qquad \sup_{T \in [0, T_0]} \left\{ \|U(\cdot, T)\|_{H^s(4) \cap H^{s+8}}, \|V(\cdot, T)\|_{H^s(4) \cap H^{s+8}} \right\} < C_1$$

On the other hand it is less clear that solutions of (5) and (6) will remain bounded over the very long time scales necessary for the KdV approximation. Thus, the first significant technical result of this paper is:

Proposition 1. *Fix $T_0 > 0$. Suppose, $U_0, V_0 \in H^\sigma(4)$ and U, V, A, B, F and G satisfy (4)-(6), then there exists a constant C_2 such that the solutions of (5) and (6) satisfy the following estimates:*

$$\sup_{\tau \in [0, T_0 \epsilon^{-2}]} \left\{ \|A(\cdot, \tau)\|_{H^{\sigma-3}}, \|B(\cdot, \tau)\|_{H^{\sigma-3}} \right\} \leq C_2$$

$$\sup_{T \in [0, T_0]} \left\{ \|F(\cdot, T)\|_{\tilde{H}}, \|G(\cdot, T)\|_{\tilde{H}} \right\} \leq C_2$$

where $\tilde{H} = H^{\sigma-5} \cap H^{\sigma-9}(2)$.

With this preliminary result in hand, we can now state our principal result.

Theorem 2. *Fix T_0, $C_I > 0$, $\sigma \geq 13$. Suppose U, V, A, B, F and G satisfy equations (4)-(6). Then there exists $\epsilon_0 > 0$ and $C_F > 0$ such that if the initial conditions for (3) satisfy Hypothesis 1 then for $\epsilon \in (0, \epsilon_0)$, we have that the unique solution to (3) satisfies*

$$\|\bar{u}(\cdot, t) - \bar{w}(\cdot, t)\|_{H^{\sigma-13} \times H^{\sigma-13}} \leq C_F \epsilon^{11/2}$$

for $t \in [0, T_0 \epsilon^{-3}]$, where,

$$\bar{w}(x, t) = \epsilon^2 \begin{pmatrix} U(X_-, T) \\ V(X_+, T) \end{pmatrix} + \epsilon^4 \begin{pmatrix} A(X, \tau) + F(X_-, T) \\ B(X, \tau) + G(X_+, T) \end{pmatrix}$$

Given this result, and the change of variables (2), we can immediately rewrite this approximation theorem in terms of the original variables. Define

$$\begin{aligned} \theta_{app}(x, t) = \ & \epsilon^2 \left(U(\epsilon(x-t), \epsilon^3 t) + V(\epsilon(x+t), \epsilon^3 t) \right) \\ & + \epsilon^4 \left(A(\epsilon x, \epsilon t) + B(\epsilon x, \epsilon t) \right) \\ & + \epsilon^4 \left(F(\epsilon(x-t), \epsilon^3 t) + G(\epsilon(x+t), \epsilon^3 t) \right). \end{aligned}$$

Corollary 1. *Fix T_0, $C_I > 0$, $\sigma \geq 13$. Suppose U, V, A, B, F and G satisfy equations (4)-(6). Then there exists $\epsilon_0 > 0$ and $C_F > 0$ such that if the initial conditions for (3) satisfy Hypothesis 1 then for $\epsilon \in (0, \epsilon_0)$, the unique solution $\theta(x, t)$ to (1) satisfies*

$$\|\theta(x, t) - \theta_{app}(x, t)\|_{H^{\sigma-13}} \leq C_F \epsilon^{11/2}$$

for $t \in [0, T_0 \epsilon^{-3}]$.

Remark 2. *The initial conditions for equation (1) are obtained from those of (3) simply by inverting the transformation (2).*

The remainder of the paper is devoted to the proof of Theorem 2 and Proposition 1. In the next section we give a formal derivation of equations (4)-(6). In Section 3 we study the existence of solutions to equations (5) and (6) and prove Proposition 1. Section 4 is the technical heart of the paper and contains the proof of Theorem 2. The proof follows the general approach for justifying modulation equations laid out in [14], but controlling the higher order approximation requires fairly extensive technical modifications. In Section 5 we present the results of a variety of numerical

computations related to Corollary 1. These computations give insight into several aspects of the second order approximation. First of all, it allows us to estimate how large the values of ϵ_0 and C_F in Corollary 1 are. They also show that the order of ϵ in the error estimates (i.e. 11/2) is apparently optimal. Finally, in the concluding section we discuss other work on second order corrections to the KdV approximation, both rigorous and non-rigorous, and how it relates to our own results.

2. Formal Derivation of the Modulation Equations

One can derive from (3) a system of KdV equations via the method of multiple time scales – this was done in [21], for example. We extend that calculation in this section to include the approximating equations for the next order correction.

To derive the modulation equations we first make the *Ansatz*

$$(8) \qquad \begin{pmatrix} u(x,t) \\ v(x,t) \end{pmatrix} = \epsilon^2 \begin{pmatrix} U(X_-,T) \\ V(X_+,T) \end{pmatrix} + \epsilon^4 \begin{pmatrix} A(X,\tau) + F(X_-,T) \\ B(X,\tau) + G(X_+,T) \end{pmatrix} + O(\epsilon^6)$$

where $\tau = \epsilon t$, $T = \epsilon^3 t$, $X = \epsilon x$, $X_- = X - \tau$, and $X_+ = X + \tau$. The two new time variables are the "multiple time scales" spoken of earlier. For convenience we will also denote $\bar{u} = (u,v)^t$, $\bar{U} = (U,V)^t$, $\bar{A} = (A,B)^t$ and $\bar{F} = (F,G)^t$.

It may seem somewhat odd that the $O(\epsilon^4)$ correction consists of a sum of functions, as opposed to a single function. The reason for this is that for our first order approximation terms, U and V, we are assuming that u and v exhibit only uni-directional motion (right and left, respectively). A and B, loosely, correct for the effect of the interaction of right and left moving waves, and they evolve on the fast time scale, τ. There are also unidirectional second order effects, which we represent with F and G. Their functional form is the same as that of the first order terms

In a moment, we will insert (8) into (3), but first we compute the effect of the operator λ on long wavelength data. Define a function W by $w(x) = W(X)$. We wish to compute $\lambda w(x)$, and see how that relates to $W(X)$. The function $W(X)$ represents solutions of long wavelength and hence in the Fourier domain we expect the frequency content of these waves to be concentrated near zero. Thus, we will (formally) approximate the effect of $\hat{\lambda}(k)$ by the first few terms of its Maclaurin series.

$$\begin{aligned} \lambda w(x) &= \mathfrak{F}^{-1}\left\{ \hat{\lambda}(k)\hat{w}(k) \right\}(x) \\ &= \int e^{ikx} \hat{\lambda}(k) \epsilon^{-1} \hat{W}(k/\epsilon) dk, \quad K = k/\epsilon \\ &= \int e^{iKX} \hat{\lambda}(\epsilon K) \hat{W}(K) dK \\ &= \mathfrak{F}^{-1}\left\{ \hat{\lambda}(\epsilon K) \hat{W}(K) \right\}(X) \\ &= \mathfrak{F}^{-1}\left\{ \left(\epsilon(iK) + \frac{1}{2}\epsilon^3(iK)^3 + \frac{3}{8}\epsilon^5(iK)^5 + O(\epsilon^7) \right) \hat{W}(K) \right\}(X) \\ &= (\epsilon\partial_X + \frac{1}{2}\epsilon^3\partial_X^3 + \frac{3}{8}\epsilon^5\partial_X^5 + O(\epsilon^7))W(X) \end{aligned}$$

It is important to note that this approximation is only formally good to $O(\epsilon^7)$.

Now we insert this approximation for λ and the *Ansatz* into (3). This is a necessarily messy procedure. To reduce the notation, anything formally $O(\epsilon^9)$ or

higher is (more or less) disregarded. Also an additional term is added to the *Ansatz* of the form:

$$(9) \qquad \epsilon^6 \bar{S} = \epsilon^6 \begin{pmatrix} S^1(X, \tau) \\ S^2(X, \tau) \end{pmatrix}$$

While this term will be treated in much the same way as the other terms in the *Ansatz*, it should be noted that this is not truly part of the next order correction. It will, however, be quite useful when we prove the approximation is a good one.

We must re-express the partial derivatives in (3) in terms of the new coordinates. By the chain rule, we have

$$\partial_t = -\epsilon \partial_{X_-} + \epsilon \partial_{X_+} + \epsilon \partial_\tau + \epsilon^3 \partial_T.$$

Spatial derivatives of terms of the form $f(X_-)g(X_+)$ or $f(X)g(X_\pm)$ are denoted by ∂_X, though all other spatial derivatives are denoted with respect to the appropriate coordinate.

So we get on the left hand side of (3)

$$\begin{aligned}
(10) \qquad \partial_t \begin{pmatrix} u(x,t) \\ v(x,t) \end{pmatrix} &= \epsilon^3 \begin{pmatrix} -\partial_{X_-} U(X_-, T) \\ \partial_{X_+} V(X_+, T) \end{pmatrix} \\
&+ \epsilon^5 \begin{pmatrix} \partial_T U(X_-, T) + \partial_\tau A(X, \tau) - \partial_{X_-} F(X_-, T) \\ \partial_T V(X_+, T) + \partial_\tau B(X, \tau) + \partial_{X_+} G(X_+, T) \end{pmatrix} \\
&+ \epsilon^7 \begin{pmatrix} \partial_T F(X_-, T) + \partial_\tau S^1(X, \tau) \\ \partial_T G(X_+, T) + \partial_\tau S^2(X, \tau) \end{pmatrix}
\end{aligned}$$

Now we must compute the right hand side of (3). A routine calculations yields:

$$\begin{aligned}
(11) \qquad RHS &= \epsilon^3 \begin{pmatrix} -\partial_{X_-} U \\ \partial_{X_+} V \end{pmatrix} \\
&+ \epsilon^5 \begin{pmatrix} -\frac{1}{2}\partial_{X_-}^3 U - \partial_X A - \partial_{X_-} F \\ \frac{1}{2}\partial_{X_+}^3 V + \partial_X B + \partial_{X_+} G \end{pmatrix} \\
&+ \epsilon^5 \begin{pmatrix} -\frac{1}{2}\partial_{X_-} U^2 - \frac{1}{2}\partial_{X_+} V^2 - \partial_X(UV) \\ \frac{1}{2}\partial_{X_-} U^2 + \frac{1}{2}\partial_{X_+} V^2 + \partial_X(UV) \end{pmatrix} \\
&+ \epsilon^7 \begin{pmatrix} -\frac{1}{2}\partial_X^3 A - \frac{1}{2}\partial_X^3 F - \frac{3}{8}\partial_{X_-}^5 U \\ \frac{1}{2}\partial_X^3 B + \frac{1}{2}\partial_{X_+}^3 G + \frac{3}{8}\partial_{X_+}^5 V \end{pmatrix} \\
&+ \epsilon^7 \begin{pmatrix} -\partial_X(UA) - \partial_{X_-}(UF) - \partial_X(UB) - \partial_X(UG) \\ +\partial_X(UA) + \partial_{X_-}(UF) + \partial_X(UB) + \partial_X(UG) \end{pmatrix} \\
&+ \epsilon^7 \begin{pmatrix} -\partial_X(VA) - \partial_X(VF) - \partial_X(VB) - \partial_{X_+}(VG) \\ +\partial_X(VA) + \partial_X(VF) + \partial_X(VB) + \partial_{X_+}(VG) \end{pmatrix} \\
&+ \epsilon^7 \begin{pmatrix} -\frac{1}{4}\partial_{X_-}^3 U^2 - \frac{1}{4}\partial_{X_+}^3 V^2 - \frac{1}{2}\partial_X^3(UV) - \partial_X S^1 \\ +\frac{1}{4}\partial_{X_-}^3 U^2 + \frac{1}{4}\partial_{X_+}^3 V^2 + \frac{1}{2}\partial_X^3(UV) + \partial_X S^2 \end{pmatrix} + O(\epsilon^9)
\end{aligned}$$

So we see that we can satisfy (3) formally to $O(\epsilon^5)$ by taking

$$\begin{aligned}
(4) \qquad \partial_T U &= -\frac{1}{2}\partial_{X_-}^3 U - \frac{1}{2}\partial_{X_-} U^2 \\
\partial_T V &= \frac{1}{2}\partial_{X_+}^3 V + \frac{1}{2}\partial_{X_+} V^2
\end{aligned}$$

and

$$(5) \quad \begin{aligned} \partial_\tau A + \partial_X A &= -\frac{1}{2}\partial_{X_+}V^2(X+\tau,\epsilon^2\tau) - \partial_X U(X-\tau,\epsilon^2\tau)V(X+\tau,\epsilon^2\tau) \\ \partial_\tau B - \partial_X B &= \frac{1}{2}\partial_{X_-}U^2(X-\tau,\epsilon^2\tau) + \partial_X U(X-\tau,\epsilon^2\tau)V(X+\tau,\epsilon^2\tau) \end{aligned}$$

Equations (4) are a pair of uncoupled Korteweg-De Vries Equations. That their solutions provide the first-order approximation to long wavelength solutions of (1) was proven in [21]. Solutions to the KdV equations are known to exist and be bounded over a long time scale (see Theorem 1 above).

Equations (5) are a set of inhomogeneous transport equations driven by the solutions to the KdV equations, for which we can write down an explicit formula for the solutions to these equations.

Corollary 2. *The solutions to equations (5) are given by*

$$(12) \quad \begin{aligned} A(X,\tau) &= \frac{1}{4}V^2(X-\tau,0) - \frac{1}{4}V^2(X+\tau,\epsilon^2\tau) \\ &\quad + \alpha(X-\tau,\epsilon^2\tau) + A_1(X,\tau) \\ B(X,\tau) &= \frac{1}{4}U^2(X+\tau,0) - \frac{1}{4}U^2(X-\tau,\epsilon^2\tau) \\ &\quad + \beta(X+\tau,\epsilon^2\tau) + B_1(X,\tau) \end{aligned}$$

where

$$(13) \quad \begin{aligned} A_1(X,\tau) &= \frac{\epsilon^2}{4}\int_0^\tau \partial_T V^2(X-\tau+2s,\epsilon^2 s)ds \\ B_1(X,\tau) &= \frac{\epsilon^2}{4}\int_0^\tau \partial_T U^2(X+\tau-2s,\epsilon^2 s)ds \end{aligned}$$

$$(14) \quad \begin{aligned} \alpha(X_-,T) &= -\epsilon^{-2}\int_0^T \partial_{X_-}\left(U(X_-,s)V(X_-+2\epsilon^{-2}s,s)\right)ds \\ \beta(X_+,T) &= \epsilon^{-2}\int_0^T \partial_{X_+}\left(U(X_+-2\epsilon^{-2}s,s)V(X_+,s)\right)ds \end{aligned}$$

Proof. The proof follows directly from Lemmas 2 and 3, which appear in the next section. They can also be verified by inserting the expressions for A and B back into (5). Furthermore, as we prove below, in spite of the prefactor of ϵ^{-2}, α and β remain $O(1)$ for all $0 \le T \le T_0$, for any T_0. $\quad\square$

The terms of $O(\epsilon^7)$ in (11) give rise to two sets of linear evolution equations, one for F and G and one for S^1 and S^2. Both are inhomogeneous systems of equations due to the presence of terms involving U, V, A, B and their derivatives. We have some freedom in the way we split up the inhomogeneous terms between these equations and we attempt to group them in such a way that it is easy to estimate the resulting solutions over the long time scales relevant for the approximation problem. In particular, we will break A and B up as in the explicit solutions above. We have:

$$\text{(6)} \qquad \begin{aligned} \partial_T F &= -\partial_{X_-}(UF) - \frac{1}{2}\partial_{X_-}^3 F + J^1 \\ \partial_T G &= \partial_{X_+}(VG) + \frac{1}{2}\partial_{X_+}^3 G + J^2 \end{aligned}$$

where the inhomogeneous terms J^1 and J^2 are given by:

$$\text{(15)} \qquad \begin{aligned} J^1(X_-,T) =& -\frac{3}{8}\partial_{X_-}^5 U(X_-,T) - \frac{1}{4}\partial_{X_-}^3 U^2(X_-,T) \\ &+ \frac{1}{4}\partial_{X_-}U^3(X_-,T) - \frac{1}{8}\partial_{X_-}^3 V^2(X_-,0) \\ &- \partial_{X_-}\left(U(X_-,T)(\frac{1}{4}V^2(X_-,0) + \alpha(X_-,T))\right) \\ &- \frac{1}{2}\partial_{X_-}^3 \alpha(X_-,T) \end{aligned}$$

$$\text{(16)} \qquad \begin{aligned} J^2(X_+,T) =& \frac{3}{8}\partial_{X_+}^5 V(X_+,T) + \frac{1}{4}\partial_{X_+}^3 V^2(X_+,T) \\ &- \frac{1}{4}\partial_{X_+}V^3(X_+,T) + \frac{1}{8}\partial_{X_+}^3 U^2(X_+,0) \\ &+ \partial_{X_+}\left(V(X_+,T)(\frac{1}{4}U^2(X_+,0) + \beta(X_+,T))\right) \\ &+ \frac{1}{2}\partial_{X_+}^3 \beta(X_+,T) \end{aligned}$$

The additional terms S^1 and S^2 should satisfy

$$\text{(17)} \qquad \begin{aligned} \partial_\tau S^1 + \partial_X S^1 &= J_{ct}^1 + J_d^1 + J_{sp}^1 \\ \partial_\tau S^2 - \partial_X S^2 &= J_{ct}^2 + J_d^2 + J_{sp}^2 \end{aligned}$$

where

$$\begin{aligned} J_{ct}^1 =& -\partial_X U(X_-,\epsilon^2\tau)\left(G(X_+,\epsilon^2\tau) + \beta(X_+,\epsilon^2\tau)\right) \\ &- \partial_X U(X_-,\epsilon^2\tau)\left(\frac{1}{4}U^2(X_+,0) - \frac{1}{4}V^2(X_+,\epsilon^2\tau)\right) \\ &- \partial_X\left(V(X_+,\epsilon^2\tau)F(X_-,\epsilon^2\tau)\right) - \frac{1}{2}\partial_X^3\left(U(X_-,\epsilon^2\tau)V(X_+,\epsilon^2\tau)\right) \\ &- \partial_X\left(V(X_+,\epsilon^2\tau)(\frac{1}{4}V^2(X_-,0) + \alpha(X_-,\epsilon^2\tau) - \frac{1}{4}U^2(X_-,\epsilon^2\tau))\right) \end{aligned}$$

$$J_d^1 = -\partial_X\left(V(X_+,\epsilon^2\tau)G(X_+,\epsilon^2\tau)\right) - \frac{1}{8}\partial_X^3 V^2(X_+,\epsilon^2\tau) + \frac{1}{4}\partial_X V^3(X_+,\epsilon^2\tau)$$

$$\begin{aligned} J_{sp}^1 =& -\frac{1}{2}\partial_X^3 A_1(X,\tau) - \partial_X\left((U(X_-,\epsilon^2\tau) + V(X_+,\epsilon^2\tau)(A_1(X,\tau) + B_1(X,\tau)))\right) \\ &- \partial_X\left(V(X_+,\epsilon^2\tau)(\frac{1}{4}U^2(X_+,0) + \beta(X_+,\epsilon^2\tau))\right) \end{aligned}$$

$$\begin{aligned}
J_{ct}^2 &= \partial_X V(X_+, \epsilon^2 \tau)\left(F(X_-, \epsilon^2 \tau) + \alpha(X_-, \epsilon^2 \tau)\right) \\
&\quad \partial_X(X_+, \epsilon^2 \tau)\left(\frac{1}{4}V^2(X_-, 0) - \frac{1}{4}U^2(X_-, \epsilon^2 \tau)\right) \\
&\quad + \partial_X\left(U(X_-, \epsilon^2 \tau)G(X_+, \epsilon^2 \tau)\right) + \frac{1}{2}\partial_X^3\left(U(X_-, \epsilon^2 \tau)V(X_+, \epsilon^2 \tau)\right) \\
&\quad + \partial_X\left(U(X_-, \epsilon^2 \tau)(\frac{1}{4}U^2(X_+, 0) + \beta(X_+, \epsilon^2 \tau) - \frac{1}{4}V^2(X_+, \epsilon^2 \tau))\right)
\end{aligned}$$

$$J_d^2 = \partial_X\left(U(X_-, \epsilon^2 \tau)F(X_-, \epsilon^2 \tau)\right) + \frac{1}{8}\partial_X^3 U^2(X_-, \epsilon^2 \tau) - \frac{1}{4}\partial_X U^3(X_-, \epsilon^2 \tau)$$

$$\begin{aligned}
J_{sp}^2 &= \frac{1}{2}\partial_X^3 B_1(X, \tau) + \partial_X\left(U(X_-, \epsilon^2 \tau) + V(X_+, \epsilon^2 \tau)(A_1(X, \tau) + B_1(X, \tau))\right) \\
&\quad + \partial_X\left(U(X_-, \epsilon^2 \tau)(\frac{1}{4}V^2(X_-, 0) + \alpha(X_-, \epsilon^2 \tau))\right)
\end{aligned}$$

The equations (6) are our second set of modulation equations for the terms of $O(\epsilon^4)$ in our long wavelength approximation. Since they are linearized, inhomogeneous KdV equations, linearized about a KdV solution, they are in principle explicitly solvable [18]. However, the form of the solution that results is quite complicated (see [19] and [11]) and thus it requires some effort to show that these solutions remain uniformly bounded in the norms which we use to bound the errors. As we noted above, the functions S^1 and S^2 do not actually form a part of the approximation at $O(\epsilon^4)$; however, we will show that they remain bounded over the time scales of interest as a part of controlling the error in our approximation.

3. Estimates on the Solutions to the Modulation Equations

Before showing that the approximation is a good one, we must first show that the solutions to the modulation equations are tractable in the their own right. Keeping in mind that our goal is show the approximation to (3) is good for a long time, we need to show that solutions to the modulation equations are bounded on the appropriate time scale, that is, for $t \sim O(\epsilon^{-3})$. First we remark on Theorem 1, above.

Notice that since $T = \epsilon^3 t$, this theorem states that we have bounded solutions of (4) for $t \in [0, T_0/\epsilon^3]$, as we had hoped. Moreover, since the solutions to (4) appear in the other modulation equations (often as inhomogeneities), that they are reasonably smooth and of rapid decay is crucial to showing that the other modulation equations are solvable over a long time, and of appropriate size. In particular, we will henceforth take U_0, $V_0 \in H^\sigma(4)$, where σ will be suitably large. We now state and prove a number of lemmas.

The first set of lemmas concerns the solutions to inhomogeneous transport equations with zero initial conditions. From the method of charateristics, we have explicit formulas for solutions.

Lemma 1. *Suppose*

$$\partial_\tau u \pm \partial_X u = f(X, \tau), \quad u(X, 0) = 0,$$

with $\|f(X, \tau)\|_{H^s} \leq C$ for $\tau \in [0, T_0 \epsilon^{-2}]$. Then $\|u(\cdot, \tau)\|_{H^s} \leq C\epsilon^{-2}$ for $\tau \in [0, T_0 \epsilon^{-2}]$.

Proof. We have
$$u(X, \tau) = \int_0^\tau f(X \mp \tau \pm s, s) \ ds.$$
The integrand is bounded by C by the Sobolev embedding theorem. A naive estimate on the integral proves the result. $\square$

Lemma 2. *Suppose*
$$\partial_\tau u \pm \partial_X u = \partial_X f(X \pm \tau, \epsilon^2 \tau), \quad u(X, 0) = 0.$$
Then
(18)
$$u(X, \tau) = \pm \frac{1}{2} \left(f(X \pm \tau, \epsilon^2 \tau) - f(X \mp \tau, 0) \right) \mp \frac{\epsilon^2}{2} \int_0^\tau \partial_T f(X \mp \tau \pm 2s, \epsilon^2 s) \ ds$$
Also, if $\|f(\cdot, T)\|_{H^s} \leq C$ *and* $\|\partial_T f(\cdot, T)\|_{H^s} \leq C$ *for* $T \in [0, T_0]$, *then* $\|u(\cdot, \tau)\|_{H^s} \leq C$ *for* $\tau \in [0, T_0 \epsilon^{-2}]$.

Proof. One can check this result explicitly. The estimate on the norm follows as in Lemma 1. $\square$

Lemma 3. *Suppose*
$$\partial_\tau u \pm \partial_X u = l(X + \tau, \epsilon^2 t) r(X - \tau, \epsilon^2 \tau), \quad u(X, 0) = 0.$$
with $\|l(\cdot, T)\|_{H^s(4)} \leq C$ *and* $\|r(\cdot, T)\|_{H^s(4)} \leq C$ *for* $T \in [0, T_0]$, *then*
$$u(X, \tau) = \upsilon(X \mp \tau, \epsilon^2 \tau)$$
with $\|\upsilon(\cdot, T)\|_{H^s(2)} \leq C$ *for* $T \in [0, T_0]$ *(that is for* $\tau \in [0, T_0 \epsilon^{-2}]$*).*

Proof. See appendix. $\square$

Remark 3. *If* l *and* r *are taken to be in* $H^s(2)$, *a similar proof shows that* υ *is in* H^s *over the long time scale.*

Remark 4. *Since the proof of the Lemma does not make explicit use of the slow time scale dependence of the inhomogeneous factors* l *and* r, *the proof is still valid if the right hand is of the form* $l(X + \tau) r(X - \tau, \epsilon^2 \tau)$, $l(X + \tau, \epsilon^2 \tau) r(X - \tau)$ *or* $l(X + \tau) r(X - \tau)$.

Remark 5. *A general study of the growth of solutions of the transport equation and related linear equations that arise in the justification of modulation equations was recently completely by D. Lannes* [15].

With these results, we may now prove the estimate for A and B in Proposition 1. That is, we have the following:

Corollary 3. *If* $U_0, V_0 \in H^\sigma(4)$, *and* U *and* V *satisfy (4), then*
$$\sup_{\tau \in [0, T_0 \epsilon^{-2}]} \{ \|A(\cdot, \tau)\|_{H^{\sigma - 3}}, \|B(\cdot, \tau)\|_{H^{\sigma - 3}} \} \leq C.$$

Proof. From Corollary 2 we know the form of A and B. By Lemma 3, we have α and β uniformly bounded in $H^{\sigma - 1}(2)$ over the long time scale. Also, by Lemma 2 we see that A_1 and B_1 are in the same space as $\partial_T U^2$ and $\partial_T V^2$. U and V satisfy the KdV equations (4), so we lose three space derivatives for the one time derivative here. That is, A_1 and B_1 are uniformly bounded in $H^{\sigma - 3}$ for the long time scale. $\square$

We will occasionally be using an alternate, but equivalent norm, on $H^s(2)$. It is:

$$|f|_{H^s(2)} = \sum_{j=0}^{s} \|(1 + x^2)\partial_x^j f(x)\|_{L^2}$$

The associated inner product is denoted by $\langle \cdot, \cdot \rangle_{H^s(2)}$

Lemma 4. *For $s > 3/2$, if $u \in H^s$ then*

$$\langle u\partial_x f, f \rangle_{H^s} \leq C|u|_{H^s}|f|_{H^s}^2.$$

Proof. The proof is similar to and simpler than that of Lemma 5, which follows. □

Lemma 5. *For $s > 3/2$, if $u \in H^s(2)$ then*

$$\langle u\partial_x f, f \rangle_{H^s(2)} \leq C|u|_{H^s(2)}|f|_{H^s(2)}^2.$$

Proof. See appendix.   □

Lemma 6. *For $f \in H^s(2) \cap H^{s+4}$,*

$$(f, \partial_x^3 f)_{H^s(2)} \leq C(\|f\|_{H^s(2)}^2 + \|f\|_{H^{s+4}}^2)$$

Proof. See appendix.   □

We may now prove the estimates on F and G in Proposition 1. That is, we have the following Lemma:

Lemma 7. *If $U_0, V_0 \in H^\sigma(4)$ and U, V, A, B, F and G satisfy equations (4)-(6), then F and G satisfy the estimates:*

$$\sup_{T \in [0, T_0]} \{\|F(\cdot, T)\|_{\tilde{H}}, \|G(\cdot, T)\|_{\tilde{H}}\} \leq C$$

where $\tilde{H} = H^{\sigma-5} \cap H^{\sigma-9}(2)$.

Proof. The proof follows from Lemmas 4 - 6 and Gronwall's inequality. We show the details for F. The case for G is entirely analogous. We take the definition of the inner product on $\tilde{H}$ to be $(\cdot, \cdot)_{\tilde{H}} = (\cdot, \cdot)_{H^{\sigma-9}(2)} + (\cdot, \cdot)_{H^{\sigma-5}}$.

The inhomogeneity J^1 is in $H^{\sigma-5}(2)$ (the term $\partial_{X_-}^5 U$ causes the loss of derivatives). So we take the inner product of (6) with F and apply Lemmas 4 - 6 and arrive at:

$$\partial_T \|F\|_{\tilde{H}}^2 \leq C(\|F\|_{\tilde{H}} + \|F\|_{\tilde{H}}^2) \leq C(1 + \|F\|_{\tilde{H}}^2)$$

An application of Gronwall's inequality yields, for $T \in [0, T_0]$:

$$\|F\|_{\tilde{H}}^2(T) \leq CTe^{CT}$$

which concludes the proof of the Lemma and also of Proposition 1.   □

We now turn our eyes to the set of equations (17). As there are many terms driving these equations, many different techniques are used to show that the equations do not blow up over the long time scale. We are aided in this task by the above lemmas, though certain terms will need special consideration.

Lemma 8. *Suppose U, V, A, B, F, G, S^1 and S^2 satisfy equations (4)-(6) and (17), then S^1 and S^1 satisfy the estimates:*

$$\sup_{\tau \in [0, T_0 \epsilon^{-2}]} \{\|S^1(\cdot, \tau)\|_{H^{\sigma-10}}, \|S^2(\cdot, \tau)\|_{H^{\sigma-10}}\} \leq C$$

Proof. We shall treat the equation for S^1 here. The situation for S^2 is completely analogous. Since equations (17) are linear, we can consider the inhomogeneity term by term. First we notice that we can apply Lemma 3 to bound the growth coming from all terms in J_{ct}^1, while Lemma 2 suffices to control all terms coming from J_d^1. Thus these terms cause no growth over the long time scale. We now take a moment to discuss the smoothness of these terms. The least smooth term in J_d^1 is $\partial_X(V(X+\tau,\epsilon^2\tau)G(X+\tau,\epsilon^2\tau))$. When we apply Lemma 2 we need to examine the smoothness of $\partial_T(V(X+\tau,\epsilon^2\tau)G(X+\tau,\epsilon^2\tau))$. Now, G is uniformly bounded in $H^{\sigma-5}$, (from Lemma 7), and since G satisfies a linearized KdV equation, we have $\partial_T(V(X+\tau,\epsilon^2\tau)G(X+\tau,\epsilon^2\tau))$ uniformly bounded in $H^{\sigma-8}$. However, the least smooth term in J_{ct}^1 is the term $\partial_X(U(X-\tau,\epsilon^2\tau)G(X+\tau,\epsilon^2\tau))$, which is uniformly bounded in $H^{\sigma-10}(2)$. Thus at best S^1 is uniformly bounded in $H^{\sigma-10}$.

Each term in J_{sp}^1 will require some special consideration. These terms are:

(19) $\qquad -\quad \partial_X(V(X+\tau,\epsilon^2\tau)\beta(X+\tau,\epsilon^2\tau))$

(20) $\qquad -\quad \partial_X(V(X+\tau,\epsilon^2\tau)U^2(X+\tau,0))$

(21) $\qquad C\quad \partial_X^3 A_1(X,\tau)$

(22) $\qquad C\quad \partial_X\left((U(X-\tau,\epsilon^2\tau)+V(X+\tau,\epsilon^2\tau))(A_1(X,\tau)+B_1(X,\tau))\right)$

Terms (19) and (20) are treated with by slight variations on Lemmas 2 and 3. The technique by which (21) and (22) are dealt with relies primarily on the prefactor of ϵ^2 which appears in the definition of the functions A_1 and B_1. Unfortunately, each computation is rather messy.

In the case of the first of these, we apply Lemma 2 and get

$$S(X,\tau) = -\frac{1}{2}(V(X+\tau,\epsilon^2\tau)\beta(X+\tau,\epsilon^2\tau) - V(X-\tau,0)\beta(X-\tau,0))$$
$$+ \frac{\epsilon^2}{2}\int_0^\tau \partial_T V(X-\tau+2s,\epsilon^2 s)\beta(X-\tau+2s,\epsilon^2 s)ds$$
$$+ \frac{\epsilon^2}{2}\int_0^\tau V(X-\tau+2s,\epsilon^2 s)\partial_T\beta(X-\tau+2s,\epsilon^2 s)ds$$

The first three terms are easily bounded by the techniques discussed previously (namely we replace $\partial_T V$ with the right hand side of the KdV equation and use naive bounds). However, when we replace $\partial_T\beta$, we lose the prefactor of ϵ^2. That is, from equation (32) in the proof of Lemma 3, we have

$$\partial_T\beta(X_+,T) = \epsilon^{-2}\partial_X(U(X_+ - 2T\epsilon^{-2},T)V(X_+,T)).$$

We make this substitution into the last term of (3) to get:

$$\frac{1}{2}\int_0^\tau V(X-\tau+2s,\epsilon^2 s)\partial_X(U(X-\tau,\epsilon^2 s)V(X-\tau+2s,\epsilon^2 s))ds$$

Notice that in this integral we have only terms that lie in the weighted Sobolev spaces, and we can use the same techniques used in the proof of Lemma 3 to control this term.

The term (21) is very nearly of the form needed to apply Lemma 2. The only difference is that there is no dependence on $\epsilon^2\tau$ in one of the terms. The ideas are essentially the same here as in the proof of Lemma 2. Consider,

$$\partial_\tau S + \partial_X S = -\partial_X(V(X+\tau,\epsilon^2\tau)U^2(X+\tau,0)).$$

The solution to this equation is given by:

$$S(X, \tau) = -\frac{1}{2} \left\{ U^2(X + \tau, 0)V(X + \tau, \epsilon^2 \tau) - U^2(X - \tau, 0)V(X - \tau, \epsilon^2 \tau) \right\}$$
$$+ \frac{\epsilon^2}{2} \int_0^\tau U^2(X - \tau + 2s, 0)\partial_T V(X - \tau + 2s, \epsilon^2 s) ds$$

If one replaces $\partial_T V(X - \tau + 2s, \epsilon^2 \tau)$ in the integral by the right hand side of the KdV equation, and then takes naive norms, we find that this term is also controllable.

We now turn our attention to the final two terms which involve the functions A_1 and B_1. The calculations here are quite messy, though the ideas are straightforward. We replace $\partial_T V$ with the right hand side of the KdV equation and then apply a number of the same techniques used in proving Lemmas 2 and 3. The factor of ϵ^2 present in the definitions of A_1 and B_1 is crucial. Consider

$$\partial_\tau S + \partial_X S$$
$$= C \partial_X^3 A_1(X, \tau)$$
$$= C\epsilon^2 \partial_X^3 \int_0^\tau \partial_T V^2(X - \tau + 2s, \epsilon^2 s) ds$$
$$= C\epsilon^2 \partial_X^2 \int_0^\tau \left(\partial_s (\partial_T V^2(X - \tau + 2s, \epsilon^2 s)) - \epsilon^2 \partial_T^2 V(X - \tau + 2s, \epsilon^2 s) \right) ds$$
$$= C\epsilon^2 \partial_X^2 \left\{ \partial_T V^2(X + \tau, \epsilon^2 \tau) - \partial_T V^2(X - \tau, 0) \right\}$$
$$+ C\epsilon^4 \partial_X^2 \int_0^\tau \partial_T^2 V^2(X - \tau + 2s, \epsilon^2 s) ds$$

For ease of notation, we will let $P(X, \tau) = C\epsilon^2 \partial_X^2 \int_0^\tau \partial_T^2 V^2(X - \tau + 2s, \epsilon^2 s) ds$. Notice that by taking naive estimates on this function, we have that $\|P\|_{H^s} \leq C$ for $\tau \in [0, T_0 \epsilon^{-2}]$. Thus we apply Lemma 1 to this equation to find that S is bounded on the long time interval.

In order to deal with (22), we will rewrite $\partial_T V^2$ and $\partial_T U^2$. That is,

$$\partial_T V^2 = 2V \partial_T V$$
$$= V(\partial_X^3 V + \partial_X V^2)$$
$$= \partial_X \left(V \partial_X^2 V - \frac{1}{2}(\partial_X V)^2 + \frac{2}{3}V^3 \right)$$
$$= \partial_X \tilde{V}.$$

Where $\tilde{V} = V \partial_X^2 V - 1/2(\partial_X V)^2 + 2/3V^3$. A similar calculation yields $\partial_T U^2 = \partial_X \tilde{U}$, where $\tilde{U} = -U \partial_X^2 U + 1/2(\partial_X U)^2 - 2/3U^3$. Notice that $\tilde{U}, \tilde{V} \in H^{\sigma - 2}$ for $T \in [0, T_0]$, since they lose at most two derivatives in comparison with U and V. Similarly, we

have $\partial_T \tilde{U} \in H^{\sigma-5}$. So consider the equation,

$$\partial_\tau S + \partial_X S$$
$$= C\partial_X \left(U(X - \tau, \epsilon^2\tau) + V(X + \tau, \epsilon^2\tau) \right) (A_1(X, \tau) + B_1(X, \tau))$$
$$= C\epsilon^2 \partial_X \Big[\left(U(X - \tau, \epsilon^2\tau) + V(X + \tau, \epsilon^2\tau) \right)$$
$$\times \int_0^\tau \partial_T \left(V^2(X - \tau + 2s, \epsilon^2 s) + U^2(X + \tau - 2s, \epsilon^2 s) \right) ds \Big]$$
$$= C\epsilon^2 \Big[\partial_X \left(U(X - \tau, \epsilon^2\tau) + V(X + \tau, \epsilon^2\tau) \right)$$
$$\times \int_0^\tau \partial_X \left(\tilde{V}(X - \tau + 2s, \epsilon^2 s) + \tilde{U}(X + \tau - 2s, \epsilon^2 s) \right) ds \Big]$$
$$= C\epsilon^2 \partial_X \Big[\left(U(X - \tau, \epsilon^2\tau) + V(X + \tau, \epsilon^2\tau) \right)$$
$$\times \Big(\tilde{V}(X + \tau, \epsilon^2\tau) - \tilde{V}(X - \tau, 0) - \epsilon^2 \int_0^\tau \partial_T \tilde{V}(X - \tau + 2s, \epsilon^2 s) ds$$
$$+ \tilde{U}(X - \tau, \epsilon^2\tau) - \tilde{U}(X + \tau, 0) - \epsilon^2 \int_0^\tau \partial_T \tilde{U}(X + \tau - 2s, \epsilon^2 s) ds \Big) \Big]$$

Notice that by taking naive estimates, the terms $Q^1(X, \tau) = \epsilon^2 \int_0^\tau \partial_T \tilde{V}(X - \tau + 2s, \epsilon^2 s) \, ds$ and $Q^2(X, \tau) = \epsilon^2 \int_0^\tau \partial_T \tilde{U}(X + \tau - 2s, \epsilon^2 s) \, ds$ are uniformly bounded in $H^{\sigma-5}$ over the long time scale. Thus we apply Lemma 1 to the above equation and find that this term is well-behaved over the long time scale.   □

4. The Validity of the Approximation

In this section we set prove that the approximation to a true solution of (3) made by the *Ansatz* is in fact a good one by completing the proof of Theorem 2.

Proof. (of Theorem 2.)

To prove this theorem we shall need a number of lemmas.

Lemma 9. *If $\Phi \in H^{s+1}$, then for $\epsilon < 1$,*

$$\|\lambda\Phi(\epsilon\cdot)\|_{H^s} \leq C\epsilon^{1/2}\|\Phi\|_{H^{s+1}}.$$

Proof. The proof here is analogous to the proof of the following lemma.   □

Lemma 10. *Let $T_1(y) = y$, $T_3(y) = y + 1/2y^3$, and $T_5(y) = y + 1/2y^3 + 3/8y^5$. Then for $j = 1, 3, 5$ if $\Phi(X) \in H^{s+j+2}$ we have, for $\epsilon < 1$,*

$$\|\lambda\Phi(\epsilon\cdot) - T_j(\epsilon\partial_X)\Phi(\epsilon\cdot)\|_{H^s} \leq C\epsilon^{j+3/2}\|\Phi\|_{H^{s+j+2}}$$

Proof. See appendix.   □

Now suppose that there is a solution to (3) of the form,

$$(23) \qquad \bar{u}(x, t) = \epsilon^2 \bar{\Psi}(x, t) + \epsilon^{11/2} \bar{R}(x, t)$$

where

$$(24) \qquad \epsilon^2 \bar{\Psi}(x, t) = \epsilon^2 \bar{U} + \epsilon^4 (\bar{A} + \bar{F}) + \epsilon^6 \bar{S},$$

and $\bar{R} = (R^1(x,t), R^2(x,t))^t$. We consider the term $\bar{R}$ to be the error in our approximation. Substituting (23) into (3), we find that $\bar{R}$ must satisfy the equation,

(25)
$$\partial_t \begin{pmatrix} R^1 \\ R^2 \end{pmatrix} = \begin{pmatrix} -\lambda & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} R^1 \\ R^2 \end{pmatrix} + \epsilon^2 \begin{pmatrix} -\lambda(\Psi^1 + \Psi^2)(R^1 + R^2) \\ \lambda(\Psi^1 + \Psi^2)(R^1 + R^2) \end{pmatrix}$$
$$+ \frac{\epsilon^{11/2}}{2} \begin{pmatrix} -\lambda(R^1 + R^2)^2 \\ \lambda(R^1 + R^2)^2 \end{pmatrix} + \epsilon^{-11/2} \mathrm{Res}[\epsilon^2 \bar{\Psi}]$$

where
(26)
$$\mathrm{Res}[\epsilon^2 \bar{\Psi}] = -\partial_t \begin{pmatrix} \epsilon^2 \Psi^1 \\ \epsilon^2 \Psi^2 \end{pmatrix} + \begin{pmatrix} -\lambda & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} \epsilon^2 \Psi^1 \\ \epsilon^2 \Psi^2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} -\lambda(\epsilon^2 \Psi^1 + \epsilon^2 \Psi^2)^2 \\ \lambda(\epsilon^2 \Psi^1 + \epsilon^2 \Psi^2)^2 \end{pmatrix}$$

We have selected our modulation equations precisely so that this term is small. By taking the time derivative of $\bar{\Psi}$ and making then making substitutions from the modulation equations, we find that

$$\mathrm{Res}[\epsilon^2 \bar{\Psi}]$$

(27)
$$= \epsilon^2 \begin{pmatrix} (T_5(\epsilon \partial_X) - \lambda)U \\ -(T_5(\epsilon \partial_X) - \lambda)V \end{pmatrix}$$
$$+ \epsilon^4 \begin{pmatrix} (T_3(\epsilon \partial_X) - \lambda)\left(A + F + \frac{1}{2}(U+V)^2\right) \\ -(T_3(\epsilon \partial_X) - \lambda)\left(B + G + \frac{1}{2}(U+V)^2\right) \end{pmatrix}$$
$$+ \epsilon^6 \begin{pmatrix} (T_1(\epsilon \partial_X) - \lambda)\left((U+V)(A+F+B+G) + S^1\right) \\ -(T_1(\epsilon \partial_X) - \lambda)\left((U+V)(A+F+B+G) + S^1\right) \end{pmatrix}$$
$$+ \epsilon^8 \begin{pmatrix} -\lambda\left(2(U+V)(S^1+S^2) + (A+F+B+G)^2\right) \\ \lambda\left(2(U+V)(S^1+S^2) + (A+F+B+G)^2\right) \end{pmatrix}$$
$$+ 2\epsilon^{10} \begin{pmatrix} -\lambda\left((A+F+B+G)(S^1+S^2)\right) \\ \lambda\left((A+F+B+G)(S^1+S^2)\right) \end{pmatrix}$$
$$+ \epsilon^{12} \begin{pmatrix} -\lambda\left((S^1+S^2)^2\right) \\ \lambda\left((S^1+S^2)^2\right) \end{pmatrix}$$

While the algebra that goes into showing this is lengthy, it should be noted that this step is accomplished by undoing to algebra that goes into deriving the modulation equations formally.

Notice that in the above expression, all functions are of long wavelength form. Thus we can apply Lemmas 9 and 10 to prove the following result.

Lemma 11. *Under the hypotheses of Theorem 2, the residual satisfies the estimate:*

$$\sup_{t \in [0, T_0 \epsilon^{-3}]} \|Res[\epsilon^2 \bar{\Psi}]\|_{H^{\sigma - 13} \times H^{\sigma - 13}} \le C \epsilon^{17/2}$$

Notice that the loss of three more derivatives is caused by the application of Lemma 10 to the term in the fourth line of equation (27), since S^j, $j = 1, 2$ are uniformly bounded in $H^{\sigma - 10}$.

We also need the following fact,

Lemma 12.
$$\left(R^2 - R^1, \lambda[(\Psi^1 + \Psi^2)(R^1 + R^2)]\right)_{H^s}$$
$$\le -\left(\partial_t(R^1 + R^2), (\Psi^1 + \Psi^2)(R^1 + R^2)\right)_{H^s} + C\epsilon^3 \|\bar{R}\|_{H^s \times H^s}$$

Proof. See appendix.   □

We wish to keep the norm of $\bar{R}$ from growing too much over the long time scale. That is, if we can show that $\|\bar{R}\|$ is $O(1)$ for $t \in [0, T_0 \epsilon^{-3}]$, we will have shown that our approximation is good.

The first term on the right hand side of equation (25) will not cause any growth in the norm, since $(f, \lambda f)_{H^s} = 0$. The third term has the prefactor of $\epsilon^{11/2}$, which will assist in controlling it, and we know from the above Lemma 11 that the residual is small.

If we tried to control solutions of (25) by applying a Gronwall type estimate to the time derivative of $(f, f)_{H^s}$, the second term would result in growth of the norm which would destroy our estimate over the time scale of interest. To avoid this problem we introduce a new energy functional which yields a norm equivalent to the $H^s \times H^s$ norm, but which does not suffer from this sort of uncontrolled growth.

Thus we define

$$(28) \qquad E_s^2(\bar{R}) = \frac{1}{2} \left(\|\bar{R}\|_{H^s \times H^s}^2 + \epsilon^2 (R^1 + R^2, (\Psi^1 + \Psi^2)(R^1 + R^2))_{H^s} \right)$$

That this norm is equivalent to the standard norm on $H^s \times H^s$ can be seen by applying the Cauchy-Schwarz inequality to the inner product, provided that we have $\epsilon^2 \|\Psi^1 + \Psi^2\|_{H^s} < 1$. Thus we use without further comment

$$\frac{1}{C} \|\bar{R}\|_{H^s \times H^s} \leq E_s(\bar{R}) \leq C \|\bar{R}\|_{H^s \times H^s}.$$

We now state and prove a useful lemma.

Lemma 13. *Set $s > 0$. Suppose $f(x), g(x) \in H^s$ and $\gamma(X) \in H^{s+1}$ where $X = \epsilon x$. Then*

$$|(f(x), \gamma(\epsilon x) g(x))_{H^s} - (g(x), \gamma(\epsilon x) f(x))_{H^s}| \leq C \epsilon \|f\|_{H^s} \|g\|_{H^s} \|\gamma\|_{W^{s,\infty}}.$$

Proof. See appendix.   □

We now have all the tools needed to finish the proof of the theorem.

$$\partial_t E^2_{\sigma-13}(\bar{R})$$

$$= \frac{1}{2}\partial_t \|\bar{R}\|^2_{H^{\sigma-13}\times H^{\sigma-13}} + \frac{\epsilon^2}{2}\partial_t \left(R^1 + R^2, (\Psi^1 + \Psi^2)(R^1 + R^2)\right)_{H^{\sigma-13}}$$

$$= (R^1, \partial_t R^1)_{H^{\sigma-13}} + (R^2, \partial_t R^2)_{H^{\sigma-13}}$$

$$+\frac{\epsilon^2}{2}\left(\partial_t(R^1 + R^2), (\Psi^1 + \Psi^2)(R^1 + R^2)\right)_{H^{\sigma-13}}$$

$$+\frac{\epsilon^2}{2}\left((R^1 + R^2), (\Psi^1 + \Psi^2)\partial_t(R^1 + R^2)\right)_{H^{\sigma-13}}$$

$$+\frac{\epsilon^2}{2}\left((R^1 + R^2), \partial_t(\Psi^1 + \Psi^2)(R^1 + R^2)\right)_{H^{\sigma-13}}$$

$$\leq (R^1, \partial_t R^1)_{H^{\sigma-13}} + (R^2, \partial_t R^2)_{H^{\sigma-13}}$$

$$+\epsilon^2\left(\partial_t(R^1 + R^2), (\Psi^1 + \Psi^2)(R^1 + R^2)\right)_{H^{\sigma-13}}$$

$$+C\epsilon^3 \|\bar{R}\|^2_{H^{\sigma-13}\times H^{\sigma-13}}$$

$$+C\epsilon^3 \|\bar{R}\|_{H^{\sigma-13}\times H^{\sigma-13}} \|\partial_t\bar{R}\|_{H^{\sigma-13}\times H^{\sigma-13}}$$

$$= \epsilon^2\left(R^2 - R^1, \lambda[(\Psi^1 + \Psi^2)(R^1 + R^2)]\right)_{H^{\sigma-13}}$$

$$+\epsilon^2\left(\partial_t(R^1 + R^2), (\Psi^1 + \Psi^2)(R^1 + R^2)\right)_{H^{\sigma-13}}$$

$$+\epsilon^{11/2}\left(R^2 - R^1, \lambda[(R^1 + R^2)^2]\right)_{H^{\sigma-13}}$$

$$+\epsilon^{-11/2}(\bar{R}, \mathrm{Res}[\bar{\Psi}])_{H^{\sigma-13}\times H^{\sigma-13}}$$

$$+C\epsilon^3 \|\bar{R}\|^2_{H^{\sigma-13}\times H^{\sigma-13}}$$

$$+C\epsilon^3 \|\bar{R}\|_{H^{\sigma-13}\times H^{\sigma-13}} \|\partial_t\bar{R}\|_{H^{\sigma-13}\times H^{\sigma-13}}$$

$$\leq +C\epsilon^{11/2} \|\bar{R}\|^3_{H^{\sigma-13}\times H^{\sigma-13}}$$

$$+C\epsilon^3 \|\bar{R}\|_{H^{\sigma-13}\times H^{\sigma-13}}$$

$$+C\epsilon^3 \|\bar{R}\|^2_{H^{\sigma-13}\times H^{\sigma-13}}$$

$$\leq C\epsilon^3 E^2_{\sigma-13}(\bar{R}) + C\epsilon^3 E_{\sigma-13}(\bar{R}) + C\epsilon^{11/2}E^3_{\sigma-13}(\bar{R})$$

We now state another lemma, which proves that the approximation is good over the long time interval.

Lemma 14. *Given $C > 0$, $T_0 > 0$, there exists $\epsilon_0 > 0$ such that if $\epsilon \in (0, \epsilon_0)$ and*

$$(29) \qquad \dot{\eta}(T) \leq C(1 + \eta(T) + \epsilon^{5/2}\eta^{3/2}(T)), \quad \eta(0) = 0,$$

for $T \in [0, T_0]$ then $\eta(T) \leq 2CT_0 e^{2CT_0}$ for $T \in [0, T_0]$.

Proof. See appendix.   □

We apply Lemma 14 to the equation for the energy of the remainder and find that

$$(30) \qquad\qquad\qquad E_{\sigma-13}(\bar{R}(\cdot, t)) \leq C$$

for $t \in [0, T_0\epsilon^{-3}]$. Note that given this *a priori* estimate proving the existence and uniqueness of solutions of (25) is a standard exercise.

We now use the equivalence of $E_{\sigma-13}$ to the typical norm on $H^{\sigma-13}$, equation (30), and Lemma 9 to find, for $t \in [0, T_0 \epsilon^{-3}]$:

$$
\begin{aligned}
\|\bar{u}(\cdot, t) - \bar{w}(\cdot, t)\|_{H^{\sigma-13} \times H^{\sigma-13}} &= \|\bar{u}(\cdot, t) - \epsilon^2 \bar{\Psi}(\cdot, t) + \epsilon^6 \bar{S}(\epsilon\cdot, \epsilon t)\|_{H^{\sigma-13} \times H^{\sigma-13}} \\
&= \|\epsilon^{11/2} \bar{R}(\cdot, t) + \epsilon^6 \bar{S}(\epsilon\cdot, \epsilon t)\|_{H^{\sigma-13} \times H^{\sigma-13}} \\
&\leq C \epsilon^{11/2}
\end{aligned}
$$
(31)

This completes the proof. $\square$

5. Some Numerics

In this section we show the results of some numerical simulations. We performed these numerics to gain insight into the qualitative nature of the higher order approximation, to estimate the values of the constants C_F and ϵ_0 that appear in Theorem 2, and to validate the results of said theorem.

We will choose the initial conditions of the system so that we can use the known solitary wave solutions to the KdV equation. We shall solve the Boussinesq equation (3) numerically. Though techniques are known for finding explicit solutions to the linearized KdV equation (see [11] and [18]), the resulting expressions are quite complicated, and so we also solve (6) numerically.

One may wonder why we should even bother computing higher order modulation equations, if we have to solve them numerically. In our situation, numerically computing solutions to the Boussinesq equation is not particularly more complicated or time intensive than finding solutions to the linearized KdV equations. However, our goal is to apply these same ideas to derive corrections to the KdV approximation for the water wave problem, whose numerical solution is a much more difficult task. We expect the same modulation equations to hold in these more general and complicated systems. Thus for the water wave problem, numerically solving the modulation equations should result in a great reduction in the complexity of the numerics.

The solutions of (1) and (6) are numerically computed using methods which are largely based around the pseudo-spectral techniques for Matlab used in [20]. Since our equations are relatively simple, Matlab, though slower than other languages (C or Fortran, for example), performs adequately rapidly. The techniques used are largely built around the use of the fast Fourier transform (FFT) to compute the various operators and derivatives, and an iterative technique to compute the nonlinear terms in (1) and the term $\partial_X(UF)$ in (6). It is implicit in the time step.

As noted previously, we use the known explicit solutions to (4). Where possible, we find explicit solutions for the various terms of A and B. The notable exception to this is in the computation of α and β, which we compute via routine trapezoidal rule techniques.

We first consider the head-on collision of two solitary waves. Note that the head-on collision will take place in the initial variable θ, not in either the u or v variables. This is because we have (formally) decomposed the system into left and right moving waves when we rewrite the system as (3). We therefore take initial conditions such that U and V will evolve as the well-known sech-squared solitary
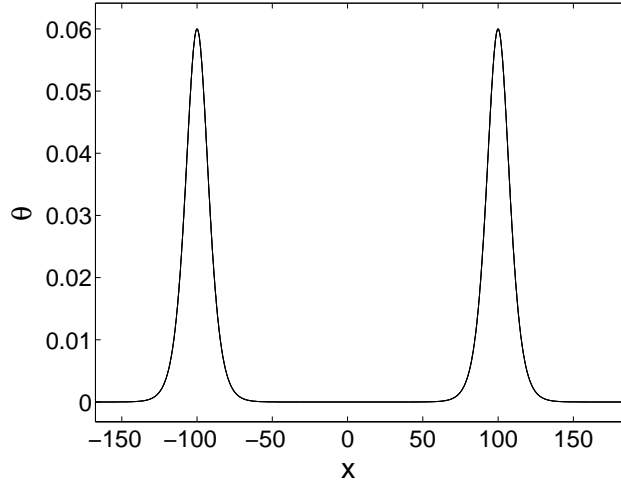
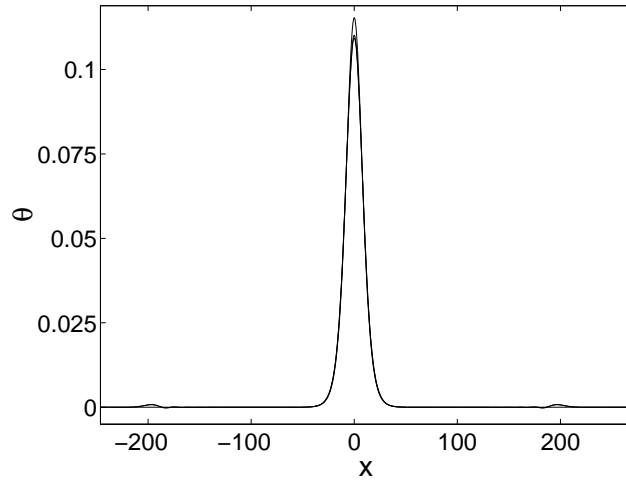FIGURE 1. Initial profile for head-on collision. $\epsilon = 0.1$

FIGURE 2. Head-on collision. $\epsilon = 0.1$

wave solutions to (4). That is we take:

$$u(x, 0) = 6\epsilon^2 \text{sech}^2(\epsilon x - 10)$$
$$v(x, 0) = 6\epsilon^2 \text{sech}^2(\epsilon x + 10)$$

as initial conditions.

Figures 1 and 2 show the solution to the Boussinesq equation, as well as the KdV approximation and the second order correction on the same plot, at the start and at the collision. Here $\epsilon = 0.1$. We remark on several features of the Boussinesq equation that are not reflected in the KdV approximation, but are present in the second order correction.
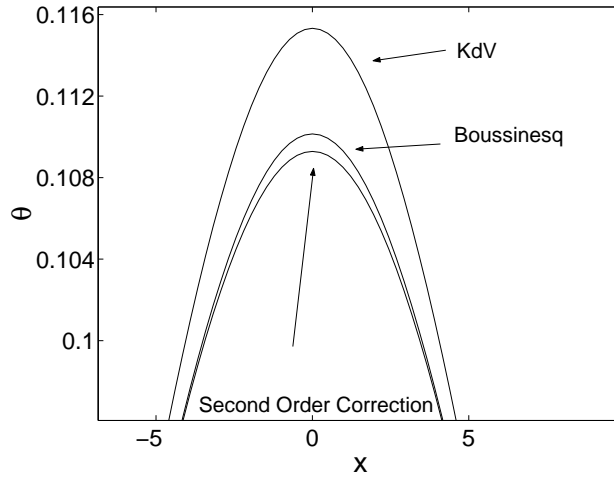
FIGURE 3. Close-up of peak of waves, during the head-on collision. $\epsilon = 0.1$
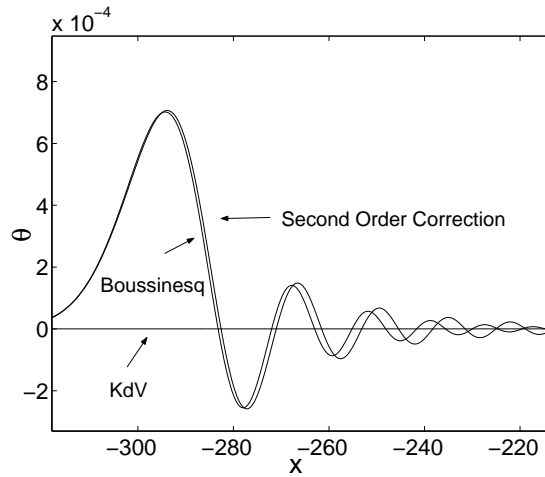
FIGURE 4. A "shadow wave" and dispersive wave train in the head-on collision. $\epsilon = 0.1$

First, in the KdV approximation, during the collision, the two waves add in linear superposition (this can be seen as the the KdV equations evolve independently). However, the solutions to the Boussinesq equation do not display this simple linear property during the collision; the total height of the wave is slightly less than the sum of the two heights of the two waves independently. The second order correction does a notably better job at displaying this feature (see Figure 3). The second feature we notice is the presence of "shadow waves" with dispersive wave trains (see Figure 4) in the solution to (1). These are not present in the KdV approximation, but are seen in the second order correction.
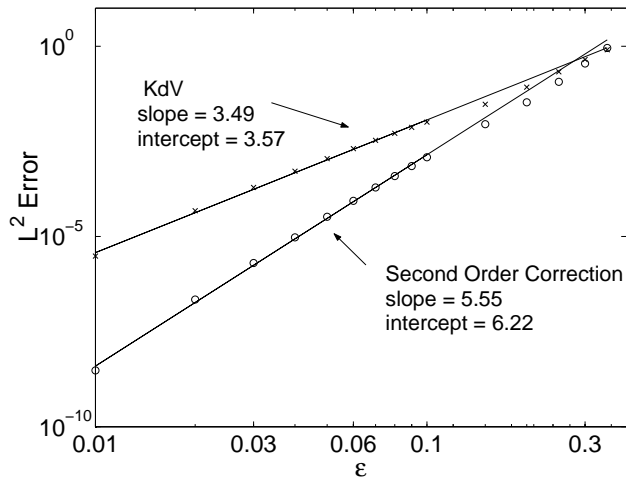
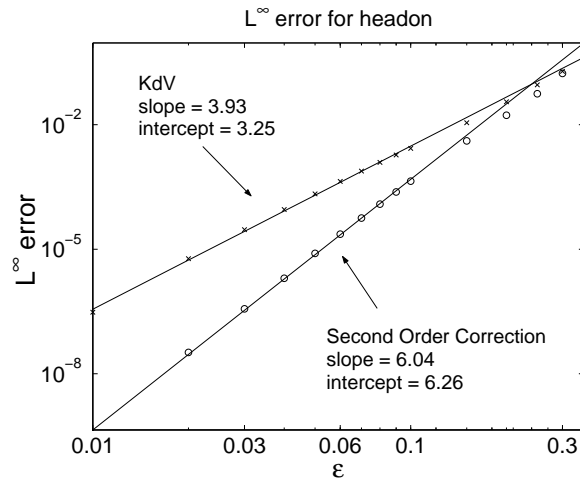FIGURE 5. $\sup \|u - w\|_{L^2}$ vs. ϵ for head-on collision.

FIGURE 6. $\sup \|u - w\|_{L^\infty}$ vs. ϵ for head-on collision.

From these pictures, we see that the second order correction is in fact doing a better job than simply the KdV approximation alone. In order to quantify this, we computed the solution for variety of values of ϵ, and computed the value of the L^2 and L^∞ error of the KdV and second order approximations. The time to collision is of $O(\epsilon^{-1})$, and on this time scale and slightly beyond, the maximum error occurs during the collision.

Figures 5 and 6 display log-log plots of the L^2 and L^∞ error versus ϵ respectively. The slopes of these lines are the order of the correction. We note that we have used only those values of $\epsilon \leq 0.1$ in computing these slopes, as we expect the error estimates to hold if ϵ is sufficiently small. Table 1 summarizes the results.

	L^2	L^∞
KdV	3.49	3.93
KdV + second order correction	5.55	6.04

TABLE 1. Order of the approximation, numerically computed, for the head-on collision

	L^2	L^∞
KdV	35.5	25.8
KdV + second order correction	503	523

TABLE 2. Value of C_F, numerically computed, for the head-on collision

From this we see that our estimate of the error made in approximating the true solution by the second order approximation is optimal, in terms of powers of ϵ. By taking note of the y-intercept of these lines, we can get an estimate on the value of the constant C_F in each case, see Table (2). Unfortunately, these values of the constant are quite large for the second order correction. We also note that it is not so much the actual value of the leading coefficient that matters, as is the location (in ϵ) at which the second order correction and the KdV correction return the same error. That is, graphically, where the lines in figures (5) and (6) cross.

The next simulation was that of right moving overtaking waves. We take initial data such that U will evolve as the famous two soliton solution to (4). Since we are not interested in left moving waves, we take initial data for v to be zero. Note that v does not remain zero, however, due to the coupling.

Unlike the previous situation, the time scale of the overtaking wave collision is $O(\epsilon^{-3})$. To observe the entirety of the collision, we take $T_0 = 8$. We also observe that the error in the approximation is largest at the end of the interval $[0, T_0\epsilon^{-3}]$. From the proof of Lemma 14, one can see that as T_0 increases, ϵ_0 decreases. This requires smaller values of ϵ, which in turn necessitates running the simulation for a longer period of time.

Figures 7, 8 and 9 display the values of u and the approximations at various times during the collision. As in the case of the head-on collision, the second order correction picks up the presence of a dispersive wave, which is not seen in the KdV approximation (see Figure 10)

It is well known that in the two soliton interaction, the waves are phase-shifted after the collision (that is, the faster wave is further ahead after the collision than it would have been had no interaction taken place, and the slower wave falls behind in a similar fashion). Overtaking waves in the Boussinesq equation share this feature, though with a different phase shift. This can be seen in Figure 9, where the KdV approximation is leading the Boussinesq solution. The second order correction noticeably "fixes" this problem. In Figures 11 we plot the locations of the peaks. Note that these figures reflect the fact that the numerics are computed in a moving reference frame (moving to the right with unit velocity).

In Figure 12 we plot the error in the phase shifts for the two approximations versus ϵ. Notice that the slope for the second order correction is steeper than that of the KdV approximation
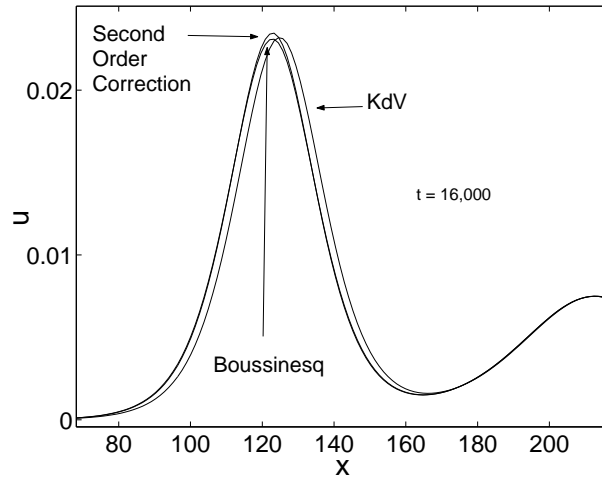
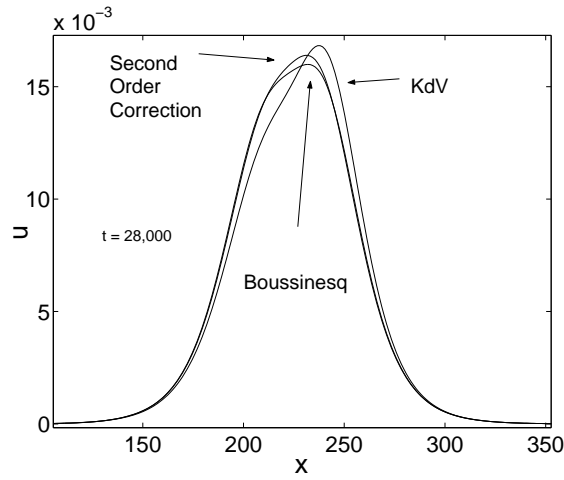FIGURE 7. Overtaking Wave before the Collision, $\epsilon = 0.05$

FIGURE 8. Overtaking Wave at the Collision, $\epsilon = 0.05$

	L^2	L^∞
KdV	3.43	3.88
KdV + second order correction	5.42	5.86

TABLE 3. Order of the approximation, numerically computed, for the overtaking collision

In Figures 13 and 14 we plot the maximum of the L^2 and L^∞ error for the two approximations versus ϵ on a log-log plot (as we did for the head-on interaction earlier). We summarize the results in Tables (3) and (4).
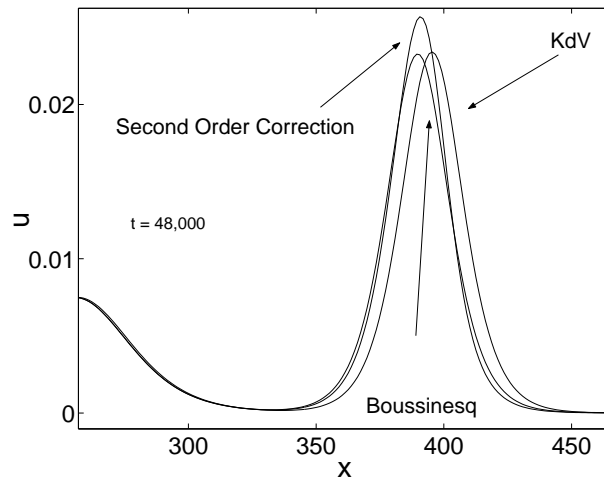
FIGURE 9. Overtaking Wave after the Collision, $\epsilon = 0.05$
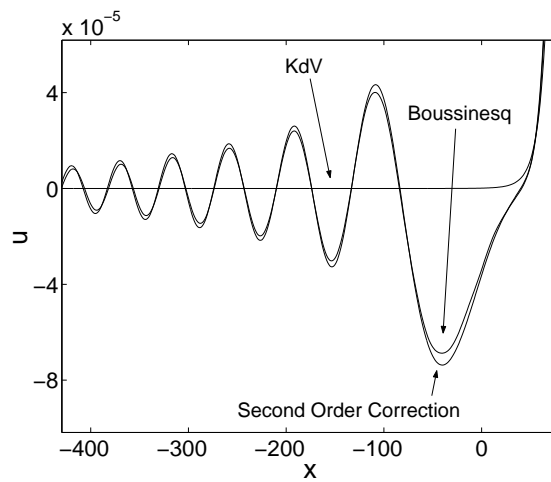
FIGURE 10. Dispersive Wave for the Overtaking Wave

	L^2	L^∞
KdV	1300	860
KdV + second order correction	19,500	19,900

TABLE 4. Value of C_F, numerically computed, for the overtaking collision

6. CONCLUSIONS

We conclude by briefly surveying other work on the derivation of higher order modulation equations for water waves and related systems.
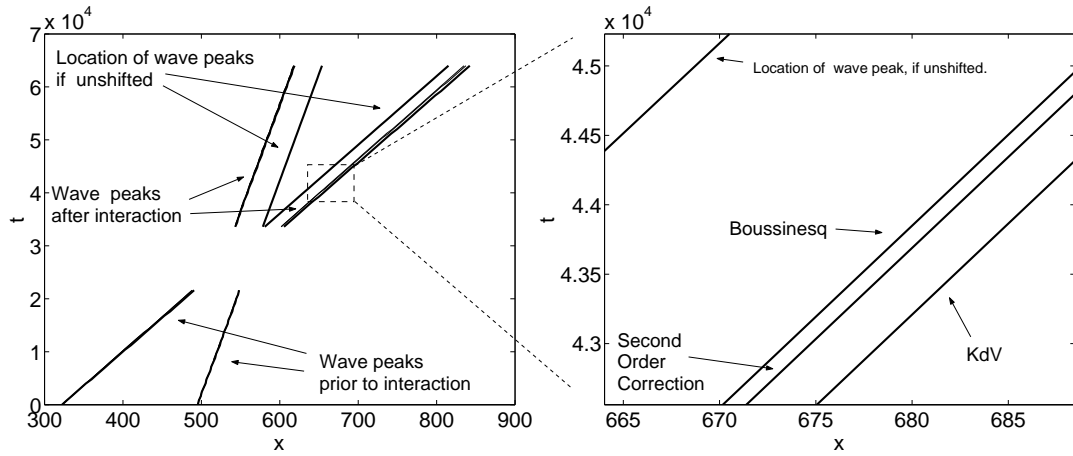
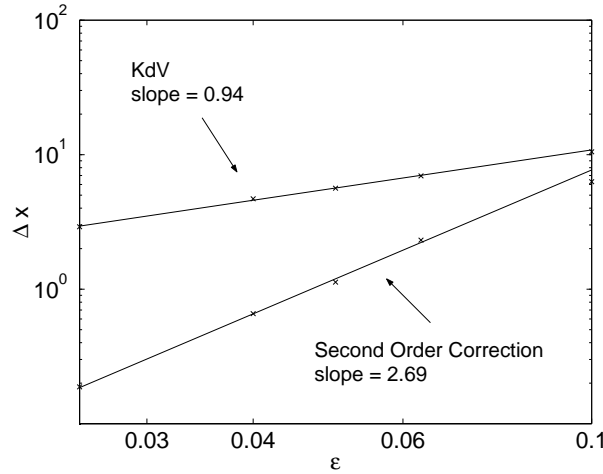FIGURE 11. Wave Peak Locations, $\epsilon = 0.05$

FIGURE 12. Error in Phase Shift vs. ϵ

For the actual water wave equations there have been a number of studies of corrections to the KdV approximation to water waves spanning the spectrum from non-rigorous asymptotic expansions [3], [4], [25] to numerical solutions of the equations of motion and comparison with the KdV predictions [2],[26], [9] to experimental investigations [16], [6]. We concentrate here on the theoretical studies since they have the closest connection to our work. In the investigations of Byatt-Smith [3], [4] and Su & Marie [25] the focus is on the head-on collision of solitary waves. This has several consequences. First of all the authors assume that the initial conditions are of a special form, namely a pair of counter-propagating solitary waves. The higher-order corrections to the solution then exploit this special form by including not only a correction to the amplitude of the solution, but a phase shift for each
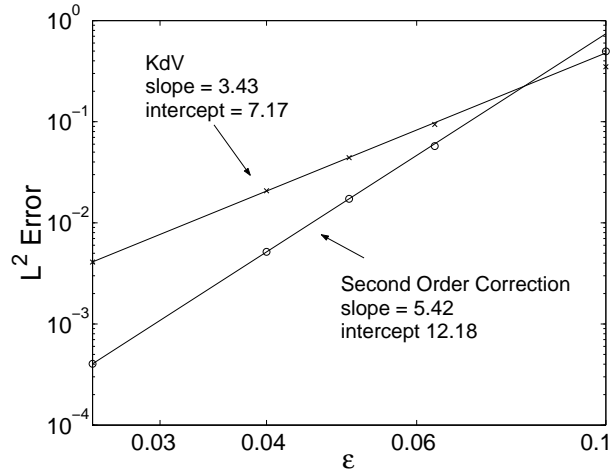
FIGURE 13. $\sup \|u - w\|_{L^2}$ vs. ϵ for Overtaking Wave Collision
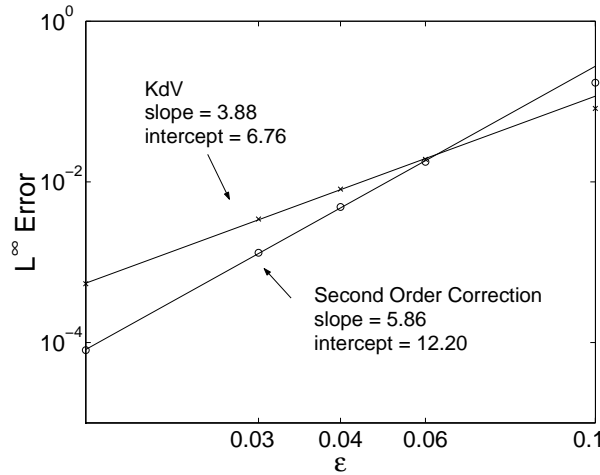
FIGURE 14. $\max \|u - w\|_{L^\infty}$ vs. ϵ for Overtaking Wave Collision

wave as it undergoes the collision. This is a very reasonable hypothesis in these physical circumstances, but one which can't easily be adapted to the more general type of initial conditions considered in our work. Furthermore, since these papers consider specifically the head-on collision of solitary waves they are concerned with events which occur on relatively short time scales (i.e. time scales of $O(\frac{1}{\epsilon})$ in our scaling.) As noted in ([4], p. 503) these expansions are not uniformly valid in time and it is not clear whether or not their solutions could be controlled over time scales of $O(\frac{1}{\epsilon^3})$. It is worth noting that in spite of the differences between our approach and those discussed here, Byatt-Smith [4] also finds that corrections to the amplitude of the solitary wave evolve according to the linearized KdV equation.

Another set of papers by Sachs, [19], Zho and Su [27], and Hărăguş-Courcelle, Sattinger, and Nicholls [10] considers corrections to the KdV approximation for unidirectional motion. The first two of these papers study this question in the context of water waves, while [10] studies the KdV approximation to solutions of the Euler-Poisson equations. The focus of these papers (particularly [19] and [10]) is rather different than ours, however. Both derive an inhomogeneous linearized KdV equation for the correction to the KdV approximation. However, rather than deriving rigorous estimates of the difference between the approximate solutions provided by the model equations and the true solutions they focus on the nature of the solutions of the linearized, inhomogeneous KdV equation. In particular, Sachs [19] shows that if one linearizes about the N-soliton solution of the KdV equation the resulting inhomogeneous equation has solutions which have no secular growth. In [10] the authors obtain explicit solutions of the linearized KdV equation, particularly for the case in which one linearizes about the two-soliton solution of the KdV equation. Hărăguş-Courcelle, et al. then compare the approximation they obtain to numerically computed solutions of the Euler-Poisson equation and they find that the addition of the solution of the linearized KdV equation to the approximation given by the two-soliton solution of the KdV equation does a significantly better job of approximating the solution of the Euler-Poisson equation. In particular, they note that the prediction of the phase shift that occurs when a "fast" traveling wave overtakes a slower one is significantly better when the second order correction is included. This effect is also present in our approximation – see Figure (12).

7. Appendix

Proof. **For Lemma 3**: We consider the case with the "minus" sign on the left hand sign for simplicity. The other case is analogous. First we change variables to $X_+ = X + \tau$, $T = \epsilon^2 \tau$, and $v(X_+, T) = u(X, \tau)$. Under this change we get the equation:

$$(32) \qquad \partial_T v(X_+, T) = \epsilon^{-2} l(X_+, T) r(X_+ - 2T\epsilon^{-2}, T)$$

This can solved by integrating with respect to the variable T. We get,

$$v(X_+, T) = \epsilon^{-2} \int_0^T l(X_+, s) r(X_+ - 2s\epsilon^{-2} 2, s) ds$$

Now we multiply by the appropriate weight and take norms:

$$(1 + X_+^2)|v(X_+, T)|$$

$$\leq \epsilon^{-2} \int_0^T (1 + X_+^2)|l(X_+, s)||r(X_+ - 2s\epsilon^{-2}, s)| \, ds$$

$$\leq \epsilon^{-2} \int_0^T (1 + X_+^2)|l(X_+, s)|(1 + (X_+ - 2s\epsilon^{-2})^2)|r(X_+ - 2s\epsilon^{-2}, s)| \, ds$$

$$= \epsilon^{-2} \int_0^T \frac{(1 + X_+^2)^2|l(X_+, s)|(1 + (X_+ - 2s\epsilon^{-2})^2)^2|r(X_+ - 2s\epsilon^{-2}, s)|}{(1 + X_+^2)(1 + (X_+ - 2s\epsilon^{-2})^2)} \, ds$$

$$\leq \epsilon^{-2} \int_0^T \frac{(1 + X_+^2)^2|l(X_+, s)|(1 + (X_+ - 2s\epsilon^{-2})^2)|r(X_+ - 2s\epsilon^{-2}, s)|}{(1 + (2s\epsilon^{-2})^2)} \, ds$$

Now take the H^s norm of each side of this equation and find that:

$$\|v(\cdot, T)\|_{H^s(2)} \leq \|l\|_{H^s(4)}\|r\|_{H^s(4)}\epsilon^{-2} \int_0^T \frac{1}{(1 + (2s\epsilon^{-2})^2)} ds$$

$$\leq C\|l\|_{H^s(4)}\|r\|_{H^s(4)} \arctan(2T\epsilon^{-2})$$

$$\leq C\|l\|_{H^s(4)}\|r\|_{H^s(4)}$$

$\square$

Proof. **For Lemma 5**:

$$\langle u\partial_x f, \rangle_{H^s(2)} = \langle u\partial_x f, f\rangle_{H^{s-1}(2)} + ((1 + x^2)\partial_x^s(u\partial_x f), (1 + x^2)\partial_x^s f)_{L^2}$$

$$\leq |u\partial_x f|_{H^{s-1}(2)}|f|_{H^{s-1}(2)} + \sum_{j=0}^{s-1} c_{sj}((1 + x^2)^2\partial_x^{s-j}u\partial_x^{j+1}f, \partial_x^s f)_{L^2}$$

$$+ \int (1 + x^2)^2 u\partial_x^{s+1}f \partial_x^s f \, dx$$

$$\leq C|u|_{H^s(2)}|f|_{H^s(2)}^2 + \frac{1}{2}\int (1 + x^2)^2 u\partial_x(\partial_x^s f)^2 \, dx$$

$$\leq C|u|_{H^s(2)}|f|_{H^s(2)}^2 - \frac{1}{2}\int \partial_x((1 + x^2)^2 u)(\partial_x^s f)^2 \, dx$$

$$\leq C|u|_{H^s(2)}|f|_{H^s(2)}^2 - \frac{1}{2}\int \partial_x u((1 + x^2)\partial_x^s f)^2 \, dx$$

$$- 2\int xu(1 + x^2)(\partial_x^s f)^2 \, dx$$

$$\leq C|u|_{H^s(2)}|f|_{H^s(2)}^2$$

$\square$

Proof. **For Lemma 6**: In this proof we use the standard norm on $H^s(2)$.

$$(f, \partial_x^3 f)_{H^s(2)} = -6(x^2 f, \partial_x f)_{H^s} - 6((1 + x^2)f, x\partial_x^2 f)_{H^s}$$

$$\leq C(\|f\|_{H^s(2)}\|f\|_{H^{s+1}}) + C(\|f\|_{H^s(2)}\|x\partial_x^2 f\|_{H^s})$$

So now consider

$$\|x\partial_x^2 f\|_{H^s}^2 = \|x\partial_x^2 f\|_{H^{s-2}}^2 + \|\partial_x^{s-1}(x\partial_x^2 f)\|_{L^2}^2 + \|\partial_x^s(x\partial_x^2 f)\|_{L^2}^2$$

We now treat the last term in the above, as the middle term can be handled in a similar fashion, and the first is easily dealt with.

$$
\begin{aligned}
\|\partial_x^s(x\partial_x^2 f)\|_{L^2}^2 &\leq C\|\partial_x^{s+1}f\|_{L^2}^2 + \|x\partial_x^{s+2}f\|_{L^2} \\
&\leq C\|f\|_{H^{s+4}}^2 + \int x^2\partial_x^{s+2}f\partial_x^{s+2}f dx \\
&\leq C\|f\|_{H^{s+4}}^2 + \int x^2\partial_x^{s+4}f\partial_x^s f dx + 4\int x\partial_x^{s+3}f\partial_x^s f dx \\
&\quad + 2\int \partial_x^{s+2}f\partial_x^s f dx \\
&\leq C(\|f\|_{H^{s+4}}^2 + \|f\|_{H^{s+4}}\|f\|_{H^s(2)})
\end{aligned}
$$

This estimate completes the proof. $\qquad\square$

Proof. **For Lemma 10**: Notice that the polynomials T_j are the first, third and fifth order polynomial expansions of $y/\sqrt{1-y^2}$ about $y=0$. Moreover, note that only odd powers appear in the expansion. So, by Taylor's theorem, there is a constant C such that $|y/\sqrt{1-y^2} - T_j(y)| \leq C|y|^{j+2}$.

We shall now use the Fourier transform version of the Sobolev norms in the following computation, which concludes the proof. Consider

$$
\begin{aligned}
&\|\lambda\Phi(\epsilon\cdot) - T_j(\epsilon\partial_X)\Phi(\epsilon\cdot)\|_{H^s}^2 \\
&= \int(1+k^2)^s|\widehat{\lambda\Phi(\epsilon x)} - \widehat{T_j(\epsilon\partial_X)\Phi(\epsilon x)}|^2 dk \\
&= \epsilon^{-2}\int(1+k^2)^s|(\frac{ik}{\sqrt{1+k^2}} - T_j(ik))\hat{\Phi}(k/\epsilon)|^2 dk \\
&\leq C\epsilon^{-2}\int(1+k^2)^s|k^{j+2}\hat{\Phi}(k/\epsilon)|^2 dk, \quad K=k/\epsilon \\
&= C\epsilon^{2j+3}\int(1+(\epsilon K)^2)^s|K^{j+2}\hat{\Phi}(K)|^2 dK \\
&\leq C\epsilon^{2j+3}\int(1+K^2)^s|K^{j+2}\hat{\Phi}(K)|^2 dK \\
&\leq C\epsilon^{2j+3}\|\partial_X^{j+2}\Phi\|_{H^s}^2 \\
&\leq C\epsilon^{2j+3}\|\Phi\|_{H^{s+j+2}}^2
\end{aligned}
$$

$\qquad\square$

Proof. **For Lemma 12**:

$$
\begin{aligned}
&\left(R^2 - R^1, \lambda[(\Psi^1 + \Psi^2)(R^1 + R^2)]\right)_{H^s} \\
&= \left(-\lambda(R^2 - R^1), (\Psi^1 + \Psi^2)(R^1 + R^2)\right)_{H^s} \\
&\leq -\left(\partial_t(R^1 + R^2), (\Psi^1 + \Psi^2)(R^1 + R^2)\right)_{H^s} \\
&\quad + \epsilon^{-11/2}\left(\text{Res}[\bar{\Psi}]^1 + \text{Res}[\bar{\Psi}]^2, (\Psi^1 + \Psi^2)(R^1 + R^2)\right)_{H^s} \\
&\leq -\left(\partial_t(R^1 + R^2), (\Psi^1 + \Psi^2)(R^1 + R^2)\right)_{H^s} + C\epsilon^3\|\bar{R}\|_{H^s\times H^s}
\end{aligned}
$$

$\qquad\square$

Proof. **For Lemma 13**: We shall be using the fact that, by the Sobolev embedding theorem, we have γ and its first s derivatives in L^∞.

$$|(f(x), \gamma(\epsilon x) g(x))_{H^s} - (g(x), \gamma(\epsilon x) f(x))_{H^s}|$$

$$= |\sum_{j=0}^{s} \left(\partial_x^j f(x), \partial_x^j (\gamma(\epsilon x) g(x)) \right)_{L^2} - \left(\partial_x^j g(x), \partial_x^j (\gamma(\epsilon x) f(x)) \right)_{L^2}|$$

$$= |\sum_{j=0}^{s} \sum_{l=0}^{j} c_{jl} \left\{ \left(\partial_x^j f(x), \partial_x^l (\gamma(\epsilon x)) \partial_x^{j-l} g(x) \right)_{L^2} - \left(\partial_x^j g(x), \partial_x^l (\gamma(\epsilon x)) \partial_x^{j-l} f(x) \right)_{L^2} \right\}|$$

$$= |\sum_{j=0}^{s} \sum_{l=0}^{j} c_{jl} \left(\partial_x^j f(x) \partial_x^{j-l} g(x) - \partial^j g(x) \partial_x^{j-l} f(x), \partial_x^l (\gamma(\epsilon x)) \right)_{L^2}|$$

$$= \epsilon |\sum_{j=0}^{s} \sum_{l=1}^{j} c_{jl} \left(\partial_x^j f(x) \partial_x^{j-l} g(x) - \partial^j g(x) \partial_x^{j-l} f(x), \epsilon^{l-1} \partial_X^l \gamma(\epsilon x) \right)_{L^2}|$$

$$= \epsilon |\sum_{j=0}^{s} \sum_{l=1}^{j} c_{jl} \int \left(\partial_x^j f(x) \partial_x^{j-l} g(x) - \partial^j g(x) \partial_x^{j-l} f(x) \right) \left(\epsilon^{l-1} \partial_X^l \gamma(\epsilon x) \right) dx|$$

$$\leq \epsilon \|\gamma\|_{W^{s,\infty}} \sum_{j=0}^{s} \sum_{l=1}^{j} c_{jl} \int | \left(\partial_x^j f(x) \partial_x^{j-l} g(x) - \partial^j g(x) \partial_x^{j-l} f(x) \right) | dx$$

$$\leq C \epsilon \|\gamma\|_{W^{s,\infty}} \|f\|_{H^s} \|g\|_{H^s}$$

$\square$

Proof. **For Lemma 14**: Functions which obey the inequality are bounded above by solutions to the following family of ordinary differential equations,

$$\dot{\eta}(T; \epsilon) = C(1 + \eta(T; \epsilon) + \epsilon^{5/2} \eta^{3/2}(T; \epsilon)), \quad \eta(0; \epsilon) = 0,$$

and so we prove the result for these equations.

By separation of variables, we have that $\eta(T; 0) = e^{CT} - 1$. We notice that for fixed T, $\eta(T; \epsilon)$ is a continuous and increasing function of ϵ. This follows since solutions of ODEs depend smoothly on their parameters and that the right hand side of the differential equations is increasing in ϵ.

Thus, by the intermediate value theorem, there exists ϵ_0 such that $\eta(T_0; \epsilon_0) = \epsilon_0^{-5}$. Moreover since ϵ^{-5} is a decreasing function for $\epsilon > 0$, we have $\eta(T_0; \epsilon) \leq \epsilon^{-5}$ for $\epsilon \in (0, \epsilon_0)$. We further note that for fixed ϵ, η is continuous and increasing in T. So we have $\eta(T; \epsilon) \leq \epsilon^{-5}$ for $T \in [0, T_0]$ and $\epsilon \in (0, \epsilon_0)$.

Thus we have,

$$\dot{\eta}(T; \epsilon) \leq C(1 + 2\eta(T; \epsilon)) \quad \eta(0; \epsilon) = 0,$$

for $T \in [0, T_0]$ and $\epsilon \in (0, \epsilon_0)$. We apply Gronwall's inequality to this to prove the result. $\square$

REFERENCES

[1] Walid Ben Youssef and Thierry Colin. Rigorous derivation of Korteweg-de Vries-type systems from a general class of nonlinear hyperbolic systems. *M2AN Math. Model. Numer. Anal.*, 34(4):873–911, 2000.

[2] J. L. Bona, W. G. Pritchard, and L. R. Scott. An evaluation of a model equation for water waves. *Philos. Trans. Roy. Soc. London Ser. A*, 302(1471):457–510, 1981.

[3] J. G. B. Byatt-Smith. An integral equation for unsteady surface waves and a comment on the Boussinesq equation. *J. Fluid Mech.*, 49:625–633, 1971.

[4] J. G. B. Byatt-Smith. The reflection of a solitary wave by a vertical wall. *J. Fluid Mech.*, 197:503–521, 1988.

[5] J. G. B. Byatt-Smith. The head-on interaction of two solitary waves of unequal amplitude. *J. Fluid Mech.*, 205:573–579, 1989.

[6] M. J. Cooker, P. D. Weidman, and D. S. Bale. Reflection of a high-amplitude solitary wave at a vertical wall. *J. Fluid Mech.*, 342:141–158, 1997.

[7] Walter Craig. An existence theory for water waves and the Boussinesq and Korteweg-de Vries scaling limits. *Comm. Partial Differential Equations*, 10(8):787–1003, 1985.

[8] P. G. Drazin and R. S. Johnson. *Solitons: an introduction*. Cambridge University Press, Cambridge, 1989.

[9] J. D. Fenton and M. M. Rienecker. A Fourier method for solving nonlinear water-wave problems: application to solitary-wave interactions. *J. Fluid Mech.*, 118:411–443, 1982.

[10] M. Hărăguş-Courcelle, D. P. Nicholls, and D. H. Sattinger. Solitary wave interactions of the Euler-Poisson equations. *Preprint*, 2002.

[11] M. Hărăguş-Courcelle and D. H. Sattinger. Inversion of the linearized Korteweg-de Vries equation at the multi-soliton solutions. *Z. Angew. Math. Phys.*, 49(3):436–469, 1998.

[12] L. A. Kalyakin. Long-wave asymptotics. Integrable equations as the asymptotic limit of nonlinear systems. *Uspekhi Mat. Nauk*, 44(1(265)):5–34, 247, 1989.

[13] Tadayoshi Kano and Takaaki Nishida. A mathematical justification for Korteweg-de Vries equation and Boussinesq equation of water surface waves. *Osaka J. Math.*, 23(2):389–413, 1986.

[14] Pius Kirrmann, Guido Schneider, and Alexander Mielke. The validity of modulation equations for extended systems with cubic nonlinearities. *Proc. Roy. Soc. Edinburgh Sect. A*, 122(1-2):85–91, 1992.

[15] David Lannes. Secular growth estimates for hyperbolic systems. Preprint, Université de Bordeaux I, 2002.

[16] T. Maxworthy. Experiments on collisions between solitary waves. *J. Fluid Mech.*, 76:177–185, 1976.

[17] J. W. Miles. Obliquely interacting solitary waves. *J. Fluid Mech.*, 79:157–169, 1977.

[18] Robert L. Sachs. Completeness of derivatives of squared Schrödinger eigenfunctions and explicit solutions of the linearized KdV equation. *SIAM J. Math. Anal.*, 14(4):674–683, 1983.

[19] Robert L. Sachs. A justification of the KdV approximation to first order in the case of N-soliton water wavers in a canal. *SIAM J. Math. Anal.*, 15(4):468–489, 1984.

[20] D. H. Sattinger and Yi Li. Matlab codes for nonlinear dispersive wave equations. 1998.

[21] Guido Schneider. The long wave limit for a Boussinesq equation. *SIAM J. Appl. Math.*, 58(4):1237–1245 (electronic), 1998.

[22] Guido Schneider and C. Eugene Wayne. The long-wave limit for the water wave problem. I. The case of zero surface tension. *Comm. Pure Appl. Math.*, 53(12):1475–1535, 2000.

[23] Guido Schneider and C. Eugene Wayne. Kawahara dynamics in dispersive media. *Phys. D*, 152/153:384–394, 2001. Advances in nonlinear mathematics and science.

[24] Guido Schneider and C. Eugene Wayne. The rigorous approximation of long-wavelength capillary-gravity waves. *Arch. Rational Mech. Anal.*, 162(3):247–285, 2002.

[25] C. H. Su and Rida M. Mirie. On head-on collisions between solitary waves. *J. Fluid Mech.*, 98(3):509–525, 1980.

[26] C. H. Su and Rida M. Mirie. Collisions between two solitary waves. part 2. a numerical study. *J. Fluid Mech.*, 115:475–492, 1982.

[27] Qisu Zho and Chau-Hsing Su. Overtaking collision between two solitary waves. *Phys. Fluids*, 29:2113–2123, 1986.