Justification of the NLS approximation for a quasilinear water wave model

Guido Schneider[1] and C. Eugene Wayne[2]

April 8, 2011

Abstract

We show how for a quasilinear water wave model the NLS approximation can be justified. The model presents several new difficulties due to the quadratic terms which have to be eliminated by a normal-form transformation. Due to the quasilinearity of the problem there is some loss of regularity associated with the normal-form transformation and there is a nontrivial resonance present in the problem. The loss of regularity is dealt with by using a Cauchy–Kowalevskaya-like method to treat the initial value problem and the nontrivial resonance is dealt with via a rescaling argument.

Keywords: water waves; normal forms; Nonlinear Schrödinger equation

1 Introduction

The so-called 2D water wave problem in case of finite depth and no surface tension consists in finding the irrotational flow of an incompressible inviscid fluid in a canal of infinite length and fixed depth subject to gravitational force. Under these conditions the evolution of the system is completely determined by the elevation of the top surface $\eta = \eta(x,t)$ and the horizontal velocity $w = w(x,t)$ at the top surface, where $x \in \mathbb{R}$ denotes the spatial variable along the canal. By making the ansatz

$$\begin{pmatrix} \eta \\ w \end{pmatrix} = \varepsilon \Psi_{\mathrm{NLS}} + \mathcal{O}(\varepsilon^2)$$

with

$$\varepsilon \Psi_{\mathrm{NLS}} = \varepsilon A\left(\varepsilon(x + c_g t), \varepsilon^2 t\right) e^{i(k_0 x + \omega_0 t)} \varphi(k_0) + c.c., \tag{1}$$

[1]Institut für Analysis, Dynamik und Modellierung, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany, guido.schneider@mathematik.uni-stuttgart.de

[2]Dept. of Mathematics and Statistics and Center for Biodynamics, Boston University, 111 Cummington Street, Boston, MA 02215, USA, cew@math.bu.edu

in 1968 V.E. Zakharov [Zak68] derived the Nonlinear Schrödinger (NLS) equation

$$\partial_T A = i\nu_1 \partial_X^2 A + i\nu_2 A|A|^2, \tag{2}$$

with coefficients $\nu_j = \nu_j(k_0) \in \mathbb{R}$, in order to describe slow modulations in time and in space of the underlying temporally and spatially oscillating wave train $e^{i(k_0 x + \omega_0 t)}$. Here, $0 < \varepsilon \ll 1$ is a small perturbation parameter, $A(X,T) \in \mathbb{C}$ the complex-valued amplitude, and $\varphi(k_0) \in \mathbb{C}^2$ an eigenvector for the linearized equation, specified in more detail below. $T = \varepsilon^2 t \in \mathbb{R}$ is the slow time scale and $X = \varepsilon(x + c_g t) \in \mathbb{R}$ is the slow spatial scale. The basic spatial wave number $k = k_0$ and the basic temporal wave number $\omega = \omega_0$ are related via the linear dispersion relation of the water wave problem, namely

$$L(\omega, k) = \omega^2 - k \tanh k = 0. \tag{3}$$

The group velocity c_g of the wave packet is given by $c_g = \partial_k \omega|_{k=k_0, \omega=\omega_0}$.

It is the purpose of this paper to present a method which we expect will allow us in the future to prove the validity of the NLS approximation for the water wave problem in the case of finite depth and no surface tension.

Our model problem is given by the equation

$$\partial_t^2 u = -\omega^2 u - \rho^2 u^2, \tag{4}$$

where $\omega = \omega(-i\partial_x)$ and $\rho = \rho(-i\partial_x)$ are pseudo differential operators defined by their symbols in Fourier space. We choose

$$\omega(k)^2 = k \tanh k$$

such that (4) and the water wave problem have the same linear dispersion relation. We define ω uniquely through $\omega(k) > 0$ for $k > 0$ and $\omega(k) < 0$ for $k < 0$.

By the choice $\rho^2 = \omega^2$ the water wave problem in case of finite depth and no surface tension and (4) share the same principal difficulties which have to be overcome for a validity proof of the NLS approximation, namely:

- a quadratic nonlinearity,
- the quasilinearity,
- the trivial resonance at the wave number $k = 0$,
- and the nontrivial resonance at the wave number $k = k_0$ which is implied by the existence of the trivial resonance at $k = 0$.

However, the Lagrangian formulation of the water wave problem whose analysis is the future goal of the subsequent analysis is much more involved. Both the linear terms and nonlinear terms are more complicated than those of our model problem and in particular,

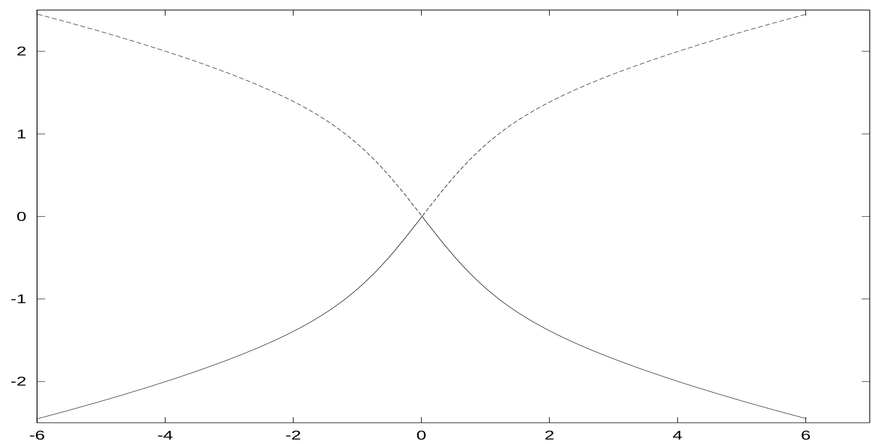

Figure 1: The curves $k \mapsto \pm\omega(k)$.

involve the Dirichlet–Neumann operator. Moreover, there is an additional, "artificial" eigenvalue curve in the Lagrangian formulation which produces a large number of further resonances. In order to illustrate the normal-form analysis which underlies our proof of the NLS approximation in a simple context we will first show that solutions of the model equation (4), which retains essential features of the water wave problem in case of finite depth and no surface tension , can be approximated by the Nonlinear Schrödinger equation.

Writing (4) as a first order system

$$\partial_t u = -\omega v, \qquad (5)$$
$$\partial_t v = \omega u + \omega u^2 \qquad (6)$$

we show that its solutions (u, v) can be approximated via the ansatz (1) by those of the NLS equation (2).

Notation. We denote Fourier transform by $(\mathcal{F}u)(k) = \widehat{u}(k) = \frac{1}{2\pi}\int u(x)e^{-ikx}\,\mathrm{d}x$. The Sobolev space H^r is equipped with the norm $\|u\|_{H^r} = (\int |\widehat{u}(k)|^2(1+|k|^2)^r\,\mathrm{d}k)^{1/2}$. Moreover, let $\|u\|_{C_b^n} = \sum_{j=0}^n \|\partial_x^j u\|_{C_b^0}$, where $\|u\|_{C_b^0} = \sup_{x\in\mathbb{R}} |u(x)|$.

Our result is

Theorem 1. *For all $k_0 \neq 0$ and for all $C_1, T_0 > 0$ there exist $T_1 > 0$, $C_2 > 0$, $\varepsilon_0 > 0$ such that for all solutions $A \in C([0, T_0], H^6(\mathbb{R}, \mathbb{C}))$ of the NLS equation (2) with*

$$\sup_{T\in[0,T_0]} \|A(\cdot, T)\|_{H^6(\mathbb{R},\mathbb{C})} \leq C_1$$

the following holds. For all $\varepsilon \in (0, \varepsilon_0)$ *there exists a solution of* (5)–(6) *which satisfies*

$$\sup_{t\in[0,T_1/\varepsilon^2]} \left\| \begin{pmatrix} u \\ v \end{pmatrix} (\cdot, t) - \varepsilon \Psi_{\mathrm{NLS}}(\cdot, t) \right\|_{(C_b^0(\mathbb{R},\mathbb{R}))^2} \leq C_2 \varepsilon^{3/2},$$

where $\varphi(k_0)$ *in the definition of* $\varepsilon\Psi_{\mathrm{NLS}}$ *in* (1) *can be chosen either as* $\begin{pmatrix} 1 \\ -i \end{pmatrix}$ *or* $\begin{pmatrix} 1 \\ i \end{pmatrix}$.

The error of order $\mathcal{O}(\varepsilon^{3/2})$ is small compared with the solution (u, v) and the approximation $\varepsilon\Psi_{\mathrm{NLS}}$ which are both of order $\mathcal{O}(\varepsilon)$ in L^∞ such that the dynamics of the NLS equation can be found also in (5)–(6). This fact should not be taken for granted. There are modulation equations (for example see [Sch95, GS01]) which, although derived by reasonable formal arguments, do not reflect the true dynamics of the original equations. One respect in which our theorem is not optimal is that we cannot show that $T_0 = T_1$. Nevertheless our estimates are on an $\mathcal{O}(1/\varepsilon^2)$ time scale and $T_1 \sim 1/C_1$ is of reasonable size so our theorem guarantees that we can observe typical NLS phenomena in our model equation. Moreover, this result is the first validity result for the NLS approximation in systems with quasilinear quadratic terms that we are aware of which holds for the qualitatively correct time scale. In [Kal87], for example, quadratic quasilinear terms are explicitly excluded.

The plan of the paper is as follows. In Section 2 we outline the underlying ideas. In Section 3 we introduce some notation and estimate the terms which remain after inserting the approximation into (5)–(6). In Section 4 we perform the normal-form transformation $W = U + G(U)$ with $G(U) = \mathcal{O}(U^2)$ and $U = (u, v)$. Special attention is given to the handling of the trivial resonance at the Fourier wavenumber $k = 0$ and of the nontrivial resonance at $k = k_0$. Due to the quasilinearity of the problem the normal-form transformation loses regularity, i.e., we have $G : H^{r+1/2} \to H^r$ so that the normal-form transformation cannot be inverted with the help of a Neumann series. Thus, Section 5 is devoted to the inversion of the normal-form transformation. In order to do so, we will use energy estimates. In Section 6 by using energy estimates in a scale of Banach spaces of analytic functions the error estimates are finally established for the transformed system. These estimates require that the nonlinear terms in the transformed system "lose" no more than one derivative, which coupled with the fact that our normal-form transformation loses half a derivative means that the nonlinearity in our original quasi-linear system is allowed to lose at most half a derivative. On the other hand, the Lagrangian formulation of the water wave problem in case of no surface tension falls into this class.

While this paper was under review we received a paper ([TW11]) in which the problem of approximating the motion of a wave packet on the surface of a fluid of infinite depth and no surface tension was solved by using special properties of the problem. For more details see the end of Section 2.

Notations: Throughout this paper many constants are denoted with the same symbol C and we always assume $0 < \varepsilon \ll 1$.

Acknowledgments: The work of Guido Schneider was partially supported by the Deutsche Forschungsgemeinschaft DFG under the grant Schn520/3-1/2. The work of Eugene Wayne was supported in part by the NSF through the grant DMS-0908093. Guido Schneider would like to thank the Mathematics Department of Boston University for its hospitality. Eugene Wayne thanks the Mathematics Institute at the University of Stuttgart for its hospitality and the DAAD for support for a visit of one month to Stuttgart during which much of this paper was completed.

2 The basic ideas

We consider an abstract evolutionary problem

$$\partial_t U = \Lambda U + N(U,U),$$

with Λ being a linear and N a symmetric bilinear operator. Suppose that U is formally approximated by $\varepsilon\Psi_{\mathrm{NLS}}$, i.e., that the residual

$$\mathrm{Res}(U) = -\partial_t U + \Lambda U + N(U,U)$$

is small for $U = \varepsilon\Psi_{\mathrm{NLS}}$. By modifying the formal approximation $\varepsilon\Psi_{\mathrm{NLS}}$ the residual can be made arbitrarily small, i.e., for all $\gamma > 0$ there exists a formal approximation $\varepsilon\Psi$ which to leading order is equal to the NLS approximation, $\varepsilon\Psi_{\mathrm{NLS}}$, such that

$$\mathrm{Res}(\varepsilon\Psi) = \mathcal{O}(\varepsilon^\gamma) \quad \text{and} \quad \varepsilon\Psi - \varepsilon\Psi_{\mathrm{NLS}} = \mathcal{O}(\varepsilon^2). \tag{7}$$

For the water wave problem the residual which contains complicated expansions of the Dirichlet–Neumann operator has been estimated in [CSS92].

In order to prove Theorem 1 we have to estimate the error

$$\varepsilon^\beta R = U - \varepsilon\Psi$$

for all $t \in [0, T_1/\varepsilon^2]$ to be of order $\mathcal{O}(\varepsilon^\beta)$ for some $\beta > 1$, i.e., we have to prove that R is of order $\mathcal{O}(1)$ for all $t \in [0, T_1/\varepsilon^2]$. The error R satisfies

$$\partial_t R = \Lambda R + 2\varepsilon N(\Psi, R) + \varepsilon^\beta N(R,R) + \varepsilon^{-\beta}\,\mathrm{Res}(\varepsilon\Psi). \tag{8}$$

For our equation, the linear operator Λ generates a uniformly bounded semigroup. The effects of the nonlinear term, $\varepsilon^\beta N(R,R)$, can be controlled over the relevant time interval if $\beta > 2$, which we assume henceforth. By choosing our approximation function $\varepsilon\Psi$ appropriately we can insure that $\varepsilon^{-\beta}\,\mathrm{Res}(\varepsilon\Psi) = \mathcal{O}(\varepsilon^2)$ and then the effects of this term on the evolution of R is also benign. Thus, the only remaining term is the linear term, $2\varepsilon N(\psi, R)$. Unfortunately, this term can perturb the linear evolution in such a way that the solutions begin to grow on time scales $\mathcal{O}(\varepsilon^{-1})$ and hence we would lose all control over the size of R on the desired time scale. Our approach to this problem is to eliminate this term via a normal-form transformation.

The idea of eliminating this term with a normal-form transformation

$$W = R + \varepsilon M(\Psi, R),$$

with M being a bilinear mapping, goes back to Kalyakin (cf. [Kal87] – see also [Sch98b, JMR00]). In order to eliminate $2\varepsilon N(\Psi, R)$ by this near identity change of variables, a non-resonance condition has to be satisfied. The eigenvalues $\lambda_j = \lambda_j(k)$ of the linearized problem (here $j = 1, 2$) as a function over the Fourier wave numbers k have to satisfy

$$|\lambda_p(k) - \lambda_1(k_0) - \lambda_q(k - k_0)| > 0 \tag{9}$$

for $p, q = 1, 2$ and all $k \in \mathbb{R}$ uniformly. It is easy to see that the eigenvalues $\lambda_j = i\omega_j$ of (5)–(6) with $\omega_j = \omega_j(k)$ given by the solutions of (3) do not satisfy (9) and in particular, there is a resonance at the wave number $k = 0$. This resonance is “trivial” for both (5)–(6) and the water wave problem, where we define a resonance to be trivial if the numerator of the normal-form transformation vanishes at the resonant wave number – otherwise it is called non-trivial. Trivial resonances ultimately cause no problems for the definition of the normal-form transformation. However, the presence of a resonance at the wave number $k = 0$ always implies another resonance for the wave number $k = k_0$ and this one turns out to be non-trivial. Therefore, the normal-form method of [Kal87] is no longer applicable and an improved method related to that used in [Sch98a] has to be applied. The method is based on a suitable wave number dependent rescaling of the error function R, followed by a number of special normal-form transformations.

This discussion and the construction of the normal-form transformation below emphasizes the difference between the meaning and effects of resonances in finite dimensional problems (or infinite dimensional problems with discrete spectrum) and those in infinite dimensional problems with continuous spectrum. This distinction was previously discussed in [McK97].

More recently the nature and effects of resonances in the water wave problem has also been examined for the 2D water wave problem in [Wu09] and for the 3D water wave problem in [GMS09] in establishing (almost) global existence results in case of infinite depth, i.e. $\omega^2 = |k|$. However, due to the different goal in [GMS09] the normal-from transformation does not have to be inverted and the loss of regularity occurs in such a way that the local existence method of the untransformed system still can be used.

In case of infinite depth and no surface tension the elimination of all quadratic terms is possible without loss of regularity as has been shown in [Wu09]. This has been used very recently in [TW11] to prove the NLS approximation property for the 2D water wave problem in the case of infinite depth and no surface tension. The differences between the water wave problem in the cases of infinite vs. finite depth are such that a transfer of the results from [TW11] to the case of finite depth does not seem obvious to us.

3 Notation and estimates for the residual

As noted in the introduction, the existence theory we use requires us to work in spaces of analytic functions. Therefore, we introduce $\widehat{Y}^p_{\sigma,r}$ equipped with the norm $\|\cdot\|_{\widehat{Y}^p_{\sigma,r}}$ given by

$$\|\widehat{u}\|_{\widehat{Y}^p_{\sigma,r}} = \|\widehat{u}w_{\sigma,r}\|_{L^p}$$

where the weight function

$$w_{\sigma,r}(k) = e^{\sigma|k|}\left(1+k^2\right)^{r/2}.$$

Moreover, let $\|u\|_{Y^p_{\sigma,r}} = \|\widehat{u}\|_{\widehat{Y}^p_{\sigma,r}}$. If $\widehat{u} \in \widehat{Y}^p_{\sigma,0}$, then u is analytic in the strip $\{z \in \mathbb{C} \mid |\mathrm{Im} z| < \sigma\}$. If $\widehat{u}$ has bounded support then the $\|\cdot\|_{\widehat{Y}^p_{\sigma,r}}$ norm is bounded by a constant times the $\|\cdot\|_{L^p}$ norm. We have Young's inequality

$$\|\widehat{u} * \widehat{v}\|_{\widehat{Y}^p_{\sigma,r}} \leq C(\|\widehat{v}\|_{L^1}\|\widehat{u}\|_{\widehat{Y}^p_{\sigma,r}} + \|\widehat{u}\|_{L^1}\|\widehat{v}\|_{\widehat{Y}^p_{\sigma,r}}) ,$$

where $*$ denotes convolution. Due to Sobolev's embedding theorem we have that $\widehat{Y}^p_{\sigma,r}$ can be embedded into L^1 with

$$\|\widehat{u}\|_{L^1} \leq C\|\widehat{u}\|_{\widehat{Y}^p_{\sigma,r}}$$

for every $\sigma > 0$ or if $rp/(p-1) > 1$ in case $\sigma = 0$.

Since our interest is the proof of an approximation result we only sketch the derivation of the NLS equation, the construction of an improved approximation and the proof of estimates for the residual which are quite standard. We refer to [Sch05, Sections 3.1 and 3.2] and the Appendix for more details.

Taking the Fourier transform of (5)–(6) we see that the linear part of the equation can be diagonalized as:

$$\partial_t U = \Lambda U + N(U,U), \tag{10}$$

with Λ being a linear, and N a bilinear mapping. In detail, in Fourier space we have

$$S = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} = S^*, \qquad \widehat{\Lambda}(k) = \begin{pmatrix} i\omega_1(k) & 0 \\ 0 & i\omega_2(k) \end{pmatrix},$$

$$\widehat{N}(\widehat{U},\widehat{V}) = S^{-1}\widetilde{N}(S\widehat{U},S\widehat{V}), \qquad \widetilde{N}(\widehat{U},\widehat{V})(k) = \omega_1(k)\begin{pmatrix} 0 \\ ((\widehat{U})_1 * (\widehat{V})_1)(k) \end{pmatrix},$$

where the eigenvalues $i\omega_{1/2}$ are given by

$$\omega_1(k) = \omega(k) = \sqrt{k\tanh(k)}\ , \quad \omega_2(k) = -\omega(k) = -\sqrt{k\tanh(k)} \tag{11}$$

for $k \geq 0$ and $\omega_j(k) = -\omega_j(-k)$. We introduce now the coefficients $\widehat{\alpha}^j_{mn}(k, k-\ell, \ell)$ of the bilinear mapping N by

$$(\widehat{N}(\widehat{u},\widehat{v})(k))_j = \int \sum_{m,n\in\{1,2\}} \widehat{\alpha}^j_{mn}(k,k-\ell,\ell)\widehat{u}_m(k-\ell)\widehat{v}_n(l)\,\mathrm{d}\ell, \tag{12}$$

where the coefficient functions $\widehat{\alpha}^j_{mn}(k,k-\ell,\ell)$ are proportional to the eigenvalues $\omega_1(k)$, for example, $\widehat{\alpha}^1_{11}(k,k-\ell,\ell)=\omega_1(k)/\sqrt{2}$. Since for our special model the coefficients $\widehat{\alpha}^j_{mn}$ only depend on the first variable, we write $\widehat{\alpha}^j_{mn}=\widehat{\alpha}^j_{mn}(k)$ in the following. With this notation (10) can be written as

$$\begin{aligned}\partial_t\widehat{U}_1(k,t) &= i\omega_1(k)\widehat{U}_1(k,t)+\int\sum_{m,n\in\{1,2\}}\widehat{\alpha}^1_{mn}(k)\widehat{U}_m(k-\ell,t)\widehat{U}_n(\ell,t)\,\mathrm{d}\ell,\\ \partial_t\widehat{U}_2(k,t) &= i\omega_2(k)\widehat{U}_2(k,t)+\int\sum_{m,n\in\{1,2\}}\widehat{\alpha}^2_{mn}(k)\widehat{U}_m(k-\ell,t)\widehat{U}_n(\ell,t)\,\mathrm{d}\ell.\end{aligned} \tag{13}$$

Note that from the form of the coefficients $\widehat{\alpha}^j_{mn}(k)$, we see that the worst growth in these coefficients is $\mathcal{O}(\sqrt{|k|})$ as $|k|\to\infty$. Thus, we expect the nonlinearity to "lose" half a derivative. Using Young's inequality for convolutions one can make this precise and we find that in either the Sobolev spaces or the spaces of analytic functions defined above one has estimates on the nonlinear term of the form

Lemma 2. *If $s\geq 2$ and $\sigma\geq 0$ there exists constants C_s and $C_{\sigma,s}$ such that the nonlinear terms in (13) satisfy the estimates*

$$\|N(U,V)\|_{H^s}\leq C_s(\|U\|_{H^{s+1/2}}\|V\|_{H^{s-1/2}}+\|U\|_{H^{s-1/2}}\|V\|_{H^{s+1/2}}), \tag{14}$$

$$\|N(U,V)\|_{Y^2_{\sigma,s}}\leq C_{\sigma,s}(\|U\|_{Y^2_{\sigma,s+1/2}}\|V\|_{Y^2_{\sigma,s-1/2}}+\|U\|_{Y^2_{\sigma,s-1/2}}\|V\|_{Y^2_{\sigma,s+1/2}}). \tag{15}$$

Remark 3. *The important point about this lemma is that we lose the maximum number of derivatives in only one of the two factors in the nonlinear term. This fact will play a role in our energy estimates in Section 6. The proof follows immediately from Leibniz's rule which also shows that the $s-1/2$ is not optimal and can be chosen smaller.*

We describe our approximation for the solution in more detail. We focus on the NLS approximation for the first component of (13) – the corresponding computation for the second component is almost identical. The basic idea is to write $U\approx\varepsilon(\tilde{\psi}_1+\tilde{\psi}_{-1})$, where $\tilde{\psi}_{\pm 1}$ are given by solutions of Nonlinear Schrödinger equation (NLS).

As remarked above, the residual

$$\mathrm{Res}(U)=-\partial_t U+\Lambda U+N(U,U) \tag{16}$$

is a measure how much U fails to be a solution of (10). If we were to use only the NLS approximation to the solution the residual would be $\mathcal{O}(\varepsilon^2)$. It turns out that the proof of the approximation theorem is greatly simplified if we choose the residual to be smaller than this. We make the residual smaller by approximating U not just with the NLS terms, but rather by a more complicated approximation:

$$\varepsilon\widetilde{\psi}_j=\sum_{j_2;|j_2|<5}\;\sum_{j_1;\beta_j(j_2,j_1)\leq 5}\varepsilon^{\beta_j(j_2,j_1)}\widetilde{\psi}^{j_1}_{j_2 j}\ ,\quad j=1,2\ , \tag{17}$$

where $\beta_1(j_2, j_1) = 1 + ||j_2| - 1| + j_1$, $\beta_1(j_2, j_1) = \beta_2(j_2, j_1)$ except for $\beta_2(1, j_1) = \beta_1(1, j_1) + 2$, and the terms

$$\widetilde{\psi}^{j_1}_{j_2 j} = A^{j_1}_{j_2 j}(\varepsilon(x + c_g t), \varepsilon^2 t) e^{i j_2 (k_0 x + \omega_0 t)} \ . \tag{18}$$

The terms with $j_1 = 0$, $j_2 = \pm 1$, $j = 1$ will again be given by NLS equation, while the higher order terms will be solutions of inhomogeneous linear PDE's or of algebraic equations. We describe in more detail the derivation of the equations for $\widetilde{\psi}^{j_1}_{j_2 j}$ in the Appendix, but at this point emphasize the following points:

- The lowest order terms in (17) are given by

  $$\begin{aligned}(\varepsilon\tilde{\psi}^0_1)_1 &= \overline{(\varepsilon\tilde{\psi}^0_{-1})_1} = \varepsilon A(\varepsilon(x + c_g)t, \varepsilon^2 t) e^{i(k_0 x + \omega_0 t)} \\ (\varepsilon\tilde{\psi}^0_1)_2 &= \overline{(\varepsilon\tilde{\psi}^0_{-1})_2} = 0 \ ,\end{aligned}$$

  where $A(X, T)$ is a solution of the NLS equation - we only modify the higher order terms.

- The higher order terms can be chosen in such a way that the residual $\mathrm{Res}(\varepsilon\tilde{\psi}) = \mathcal{O}(\varepsilon^6)$.

We mentioned in the introduction that our existence theory requires us to work in spaces of analytic functions. However, so far we have only required that the solutions of the NLS from which we construct our approximation lie in the Sobolev space H^6. We now prove that by a further small modification we can make our approximation an analytic function. We do this by replacing the $\widetilde{\psi}^{j_1}_{j_2 j}$ defined in (17) by functions that are "cut-off" in Fourier space, cf. [Alt02, BL03] for a similar prozedure. More precisely we define:

$$\psi^{j_1}_{j_2 j} : \widehat{\psi}^{j_1}_{j_2 j}(k) = \widehat{\widetilde{\psi}}^{j_1}_{j_2 j}(k) \text{ for } \{k \in \mathbb{R} \mid |k - j_2 k_0| \leq \delta\} \ ; \quad \widehat{\psi}^{j_1}_{j_2 j}(k) = 0 \text{ otherwise}, \tag{19}$$

for some $\delta > 0$ independent of $0 < \varepsilon \ll 1$. Our final approximation is then given by

$$\varepsilon\Psi_j = \sum_{j_2; |j_2| < 5} \ \sum_{j_1; \beta(j_2, j_1) \leq 5} \varepsilon^{\beta(j_2, j_1)} \psi^{j_1}_{j_2 j}. \tag{20}$$

We have the following estimates.

Lemma 4. *There exists $C_1 > 0$ and $C_2 > 0$ such that*

$$\|\varepsilon\Psi - \varepsilon\widetilde{\psi}\|_{C^0_b} \leq \|\varepsilon\widehat{\Psi} - \varepsilon\widehat{\widetilde{\psi}}\|_{L^1} \leq C_1 \varepsilon^6$$

and

$$\|\Psi\|_{Y^1_{\sigma, r}} \leq C_2 \|\Psi\|_{Y^1_{0,0}}.$$

Proof. The first estimate follows by noting that

$$\int \left|\chi_{|k-j_2k_0|<\delta}(k)-1\right| \left|\frac{1}{\varepsilon}\widehat{A}\left(\frac{k-j_2k_0}{\varepsilon}\right)\right| \mathrm{d}k$$
$$\leq \sup_{|k-j_2k_0|\geq\delta} \left|\frac{1}{(1+|\frac{k-j_2k_0}{\varepsilon}|^2)^{5/2}}\right| \cdot \|\varepsilon^{-1}\widehat{A}(\varepsilon^{-1}\cdot)\|_{\widehat{Y}^1_{0,5}} \leq C\varepsilon^5\|A\|_{H^6}.$$

The second follows since due to the compact support of $\widehat{\Psi}$ we have for all σ, $r \geq 0$ a $C = C(\sigma, r) > 0$ such that

$$\|\Psi\|_{Y^1_{\sigma,r}} \leq C\|\Psi\|_{Y^1_{0,0}}.$$

□

Because of the fact that we used the cut-off function for the approximation, we have the analyticity of the residual in a strip in the complex plane although the solutions of the NLS equation were only in the Sobolev space H^6. Note that since the Fourier transform of Ψ is non-zero only near $k = mk_0$ for $m = 0, \pm 1, \ldots \pm 5$ we can extend the definition into a strip of width $\sigma = \mathcal{O}(1)$ in the complex plane and still have an estimate on the residual of $\mathcal{O}(\varepsilon^6)$.

Thus, we finally find:

Lemma 5. *For all $C_A, T_0, \sigma, r > 0$ there exist $C_{\mathrm{Res}}, C_\Psi, \varepsilon_1 > 0$ such that the following holds for all $\varepsilon \in (0, \varepsilon_1)$.*

Let $A \in C([0,T_0], H^6(\mathbb{R},\mathbb{C}))$ be a solution of the NLS equation (2) *with*

$$\sup_{T\in[0,T_0]} \|A(T)\|_{H^6} \leq C_A.$$

Then the approximation Ψ defined in (20) *exists for all $T \in [0, T_0]$ and satisfies*

$$\begin{aligned}
\sup_{T\in[0,T_0]} \|\psi^{j_1}_{j_2}(T)\|_{Y^1_{\sigma,r}} &\leq C_\Psi, \\
\sup_{T\in[0,T_0]} \|S(\varepsilon\Psi(T)) - \varepsilon\Psi_{\mathrm{NLS}}(T)\|_{C^0_b} &\leq C_\Psi\varepsilon^2, \\
\sup_{T\in[0,T_0]} \|\operatorname{Res}(\varepsilon\Psi(T))\|_{Y^2_{\sigma,r}} &\leq C_{\mathrm{Res}}\varepsilon^{11/2}.
\end{aligned}$$

Proof. The first two estimates in the proof were explained above. The factor of S in the second estimate just accounts for the diagonalization of the linear part of the equation as explained after equation (10). Note that $\|\psi^{j_1}_{j_2}\|_{Y^2_{\sigma,r}} = \mathcal{O}(\varepsilon^{-1/2})$ due to the way the L^2 norm scales with ε. This is the reason for the difference of $-1/2$ between the formal order and the rigorous estimate. The estimate on the residual follows by a similar argument to the second – namely one extends the estimate on the Fourier transform into the complex plane (since the Fourier transform of Ψ is cut-off outside a neighborhood of mk_0). For complete details see the Appendix and [Sch05]. □

Remark 6. The first estimate in Lemma 5 is used for instance for the estimate

$$\|N(\Psi, R)\|_{Y^2_{\sigma,r}} \leq C\|\Psi\|_{Y^1_{\sigma,r+\frac{1}{2}}}\|R\|_{Y^2_{\sigma,r+\frac{1}{2}}}.$$

4 The normal-form transformation

In order to show that the solutions of the error equations (8) remain small over the very long time intervals ($t \sim \mathcal{O}(1/\varepsilon^2)$) needed for our approximation theorem we eliminate the term $2\varepsilon N(\Psi, R)$ from (8) via a normal-form transformation. There are several non-standard aspects of our normal-form transformations, but the one which causes the most technical difficulty is that the eigenvalues are continuous functions, rather than a discrete set of points, due to the continuous spectrum of the linearized problem. This makes it much harder to avoid resonances – though as pointed out in [McK97], the effects of these resonances may be less "deadly" than in the finite dimensional case.

Motivated by the form of the terms we want to eliminate from (25) we make a change of dependent variable of the form

$$\widetilde{R}_{j_1} = R_{j_1} + \varepsilon B_{j_1}(\Psi, R) \ , \quad j_1 = 1, 2 \ , \tag{21}$$

where

$$\widehat{B}_{j_1}(\Psi, R) = \sum_{j_2,j_3=1}^{2} \int \widehat{b}_{j_1;j_2,j_3}(k, k-\ell, \ell)\widehat{\Psi}_{j_2}(k-\ell)\widehat{R}_{j_3}(\ell)d\ell, \tag{22}$$

and where $\widehat{\Psi}_{j_2}$ and $\widehat{R}_{j_3}$ refer to the j_2 and j_3 components of $\widehat{\Psi}$ and $\widehat{R}$ respectively. A standard calculation which we explain in more detail below shows that the kernel function $\widehat{b}_{j_1;j_2,j_3}$ can be written as a quotient whose denominator is:

$$\omega_{j_1}(k) - \omega_{j_2}(k-\ell) - \omega_{j_3}(\ell) \ .$$

So long as the denominator remains away from zero our normal-form transformation is well defined. However, spots where the denominator vanish are known as resonances and given the formula (11) which defines $\omega_{1,2}$ it is obvious that there are at least two trouble spots:

(i) $k = 0$: Note that $k = 0$ is always a resonance if $j_2 = j_3$. However, note further that the nonlinear term also vanishes linearly at $k = 0$ (due to ω_1 that acts on the nonlinearity). Hence with the linearly vanishing denominator also the numerator of $\widehat{b}_{j_1;j_2,j_3}$ at $k = 0$ vanishes linearly and so $\widehat{B}_{j_1}$ in (22) can be well defined.

(ii) $k = k_0$: Because the Fourier transform approximate solution $\widehat{\Psi}(m)$ is concentrated around $m = \pm k_0$ we can approximate the denominator of the normal-form transformation by

$$\omega_{j_1}(k) \pm \omega_{j_2}(k_0) - \omega_{j_3}(k \mp k_0) \ . \tag{23}$$

(We will validate this approximation below.) Taking the "$-$" sign and assuming that $j_1 = j_2$ we see that $k = k_0$ is also a resonance. However, by scaling the dependent variable R near $k = 0$ one order less w.r.t. to ε than at the other wave numbers we can make the numerator of the kernel of the normal-form transformation to vanish for $k = k_0$, too. This allows the normal-form transformation to be well defined in spite of the resonance. A similar resonance and cancellation occurs at $k = -k_0$ for other choices of the signs in (23).

We now make the preceding observations more precise. We begin by rescaling the variable R to reflect the fact that the nonlinearity vanishes at $k = 0$. For a $\delta > 0$ sufficiently small, but independent of $0 < \varepsilon \ll 1$, define a weight function ϑ by its Fourier transform:

$$\widehat{\vartheta}(k) = \begin{cases} 1 & \text{for} \quad |k| > \delta, \\ \varepsilon + (1-\varepsilon)|k|/\delta & \text{for} \quad |k| \leq \delta. \end{cases}$$

We then rewrite a solution U of (10) as a sum of the approximation and an error, i.e.,

$$U = \varepsilon\Psi + \varepsilon^{\beta}\vartheta R, \tag{24}$$

with a $\beta \in (3, 7/2)$ and where, in a slight abuse of notation, ϑR is defined by $\widehat{\vartheta R} = \widehat{\vartheta}\widehat{R}$, i.e., we avoid writing the convolution $\vartheta * R$. Note that $\widehat{\vartheta}(k)\widehat{R}(k)$ is small at the wave numbers close to zero reflecting the fact that the nonlinearity vanishes at $k = 0$.

If we now insert U into (10) we find that R satisfies

$$\partial_t R = \Lambda R + 2\varepsilon\vartheta^{-1}N(\Psi, \vartheta R) + \varepsilon^{\beta}\vartheta^{-1}N(\vartheta R, \vartheta R) + \varepsilon^{-\beta}\vartheta^{-1}\,\mathrm{Res}(\varepsilon\Psi). \tag{25}$$

Since we need estimates for R on a time scale $\mathcal{O}(1/\varepsilon^2)$ and since ϑ^{-1} is at most of order $\mathcal{O}(1/\varepsilon)$ all terms on the right-hand side except for the linear ones are at least of order $\mathcal{O}(\varepsilon^2)$. In particular, if we can control the linear evolution all the remaining terms can be easily handled with the help of Gronwall's inequality. The evolution due to the term ΛR can be explicitly computed and causes no growth in R. Hence we need only control or eliminate the effects of the remaining linear terms.

We begin by examining the term $2\varepsilon\vartheta^{-1}N(\Psi, \vartheta R)$ in greater detail. Note that from (12), we can write the j_1-th component of this term as

$$\varepsilon\vartheta^{-1}\widehat{N}_{j_1}(\Psi, \vartheta R)(k) = \varepsilon\vartheta^{-1}(k) \sum_{j_2,j_3=1}^{2} \int \widehat{\alpha}^{j_1}_{j_2 j_3}(k)\widehat{\vartheta}(m)\widehat{\Psi}_{j_2}(k-m)\widehat{R}_{j_3}(m)dm \tag{26}$$

where the kernel function $\alpha^{j_1}_{j_2 j_3}(k)$ is proportional to $\omega_1(k)$. Recall further that according to (17), the approximating function Ψ can be written as

$$\Psi = \Psi_c + \varepsilon\Psi_s,$$

where both Ψ_c and Ψ_s have norm $\mathcal{O}(1)$ in any of the $Y^1_{\sigma,r}$ spaces (due to their compact support in Fourier space) but they have disjoint supports with $\operatorname{supp}(\widehat{\Psi}_c) \subset \{k \in \mathbb{R} \, ||k - k_0| \leq \delta, \ |k+k_0| \leq \delta\}$.

We find

Lemma 7. *There exists $C_L > 0$ such that*

$$\|\varepsilon\vartheta^{-1} N(\varepsilon\Psi_s, \vartheta R)\|_{Y^2_{\sigma,r-1/2}} \leq C_L \varepsilon^2 \|R\|_{Y^2_{\sigma,r}} \tag{27}$$

Proof. Recalling that

$$|\widehat{\alpha}^{j_1}_{j_2 j_3}(k)| \leq C \min(|k|^{1/2}, |k|) \ . \tag{28}$$

we see that there exists $C > 0$ such that

$$\left| \frac{\widehat{\alpha}^{j_1}_{j_2 j_3}(k)}{\vartheta(k)} \right| \leq C, \tag{29}$$

and hence, applying Young's inequality for convolutions as in Remark 6 gives the result. □

Thus, this term does not cause undue growth in the error R over the time scales of interest and we can ignore it. More precisely, if we rewrite (25) as

$$\begin{aligned} \partial_t R \ &= \ \Lambda R + 2\varepsilon\vartheta^{-1} N(\Psi_c, \vartheta R) + 2\varepsilon^2 \vartheta^{-1} N(\Psi_s, \vartheta R) \\ &\qquad + \varepsilon^{\beta} \vartheta^{-1} N(\vartheta R, \vartheta R) + \varepsilon^{-\beta}\vartheta^{-1} \operatorname{Res}(\varepsilon\Psi). \end{aligned} \tag{30}$$

then it is only the term $\varepsilon\vartheta^{-1} N(\Psi_c, \vartheta R)$ that needs to be removed by a normal-form transformation.

Remark 8. *In fact, we do not have to eliminate this term entirely, but rather eliminate it up to remainders that are of $\mathcal{O}(\varepsilon^2)$ which can as usual be controlled with the aid of Gronwall's inequality. This leads us, in the course of constructing the normal-form transformation below, to introduce a sequence of terms which we will denote $\varepsilon^2 \mathcal{E}^j$. We will show in the course of the argument that these terms can be bounded by $\mathcal{O}(\varepsilon^2)$ in the $Y^2_{\sigma,r-1}$ norm, if R is in a bounded neighborhood of the origin in $Y^2_{\sigma,r}$, and thus they can be ignored for the purpose of the normal-form transformation. The consequences of the loss of differentiability will be discussed in Section 6, but for the moment we note that it will turn out that so long as these remainder terms are of $\mathcal{O}(\varepsilon^2)$ and lose no more than one derivative, they can be ignored in the following discussion.*

Before constructing the first of the normal-form transformations we prove a lemma which will simplify the subsequent discussion and will allow us to extract the real 'dangerous' terms from $\varepsilon\vartheta^{-1} N(\Psi_c, \vartheta R)$. This lemma (and the simplifications it brings about) take advantage of the strong localization of Ψ_c near the wave numbers $\pm k_0$ in Fourier space.

Lemma 9. *Fix $p \in \mathbb{R}$. Assume that $\kappa = \kappa(k, k-m, m) \in C(\mathbb{R}^3, \mathbb{C})$. Assume further that ψ has a finitely supported Fourier transform and that $R \in Y^2_{\sigma,r}$.*

- *If κ is Lipshitz with respect to its second argument for $k-m$ in some neighborhood of $p \in \mathbb{R}$, then there exists $C_{\psi,\kappa,p} > 0$ such that*

$$\begin{aligned}\| \int \kappa(\cdot, \cdot - m, m)\widehat{\psi}(\frac{\cdot - m - p}{\varepsilon})\widehat{R}(m)dm - \int \kappa(\cdot, p, m)\widehat{\psi}(\frac{\cdot - m - p}{\varepsilon})\widehat{R}(m)dm\|_{Y^2_{\sigma,r}} \\ \leq C_{\psi,\kappa,p}\varepsilon\|R\|_{Y^2_{\sigma,r}}\end{aligned} \tag{31}$$

- *If κ is globally Lipshitz with respect to its third argument, then there exists $D_{\psi,\kappa} > 0$ such that*

$$\begin{aligned}\| \int \kappa(\cdot, \cdot - m, m)\widehat{\psi}(\frac{\cdot - m - p}{\varepsilon})\widehat{R}(m)dm \\ - \int \kappa(\cdot, \cdot - m, \cdot - p)\widehat{\psi}(\frac{\cdot - m - p}{\varepsilon})\widehat{R}(m)dm\|_{Y^2_{\sigma,r}} \\ \leq D_{\psi,\kappa}\varepsilon\|R\|_{Y^2_{\sigma,r}}\end{aligned} \tag{32}$$

Remark 10. *Note that there are two important aspects of this lemma – the first is that we fix the second argument of the kernel function κ to the value p (or the third to $k-p$) and the second is that the error which we make by this procedure is $\mathcal{O}(\varepsilon)$.*

Proof: We give the details of the proof for the first of the two cases in the Lemma. The very similar second case is left to the reader.

$$\begin{aligned}&\| \int \kappa(\cdot, \cdot - m, m)\widehat{\psi}(\frac{\cdot - m - p}{\varepsilon})\widehat{R}(m)dm - \int \kappa(\cdot, p, m)\widehat{\psi}(\frac{\cdot - m - p}{\varepsilon})\widehat{R}(m)dm\|^2_{Y^2_{\sigma,r}} \\ &= \int \left(\int (\kappa(k, k-m, m) - \kappa(k, p, m)\widehat{\psi}(\frac{k-m-p}{\varepsilon})\widehat{R}(m)dm\right)^2 e^{2\sigma|k|}(1+k^2)^r dk \\ &\leq \int \left(C_\kappa \int |(k-m) - p|\widehat{\psi}(\frac{k-m-p}{\varepsilon})\widehat{R}(m)dm\right)^2 e^{2\sigma|k|}(1+k^2)^r dk \\ &\leq C_\kappa^2(\int e^{\sigma\ell}(1+\ell^2)^{r/2}|\ell||\widehat{\psi}(\frac{\ell}{\varepsilon})|d\ell)^2\|R\|^2_{Y^2_{\sigma,r}} \leq C_{\psi,\kappa,p}\varepsilon^2\|R\|^2_{Y^2_{\sigma,r}},\end{aligned}$$

where to the next to last inequality we applied Young's inequality to bound the L^2 norm of the convolution and the last relied on the fact that $\widehat{\psi}$ has compact support. □

Remark 11. *The conclusions of Lemma 9 also hold if the integrals run only over a subset of $\mathbb{R}$.*

The first normal-form transformation: An important property of the nonlinear term in (4) is that due to the derivative acting on it the size of its Fourier transform depends on whether k is close to zero or not. (This property is shared with the water wave problem which is why we included it in our model equation.) In order to separate the behavior in these two regions more clearly we define projection operators P_0 and P_1 by the Fourier multipliers

$$\widehat{P}^0(k) = \chi_{|k|\leq\delta}(k) \quad\text{and}\quad \widehat{P}^1(k) = \mathbf{1} - \widehat{P}^0(k) \tag{33}$$

for a $\delta > 0$ sufficiently small, but independent of $0 < \varepsilon \ll 1$. (This is the same constant δ that appears in the definition of ϑ.) When necessary we will write

$$R = R^0 + R^1 \ ,$$

with $R^j = P^j R$, for $j = 0, 1$ and analogously with the other variables. Note that these *superscripts* should not be confused with the *subscripts* which denote the *components* of R.

Reconsider the part of (30) that we need to simplify, namely

$$\partial_t R = \Lambda R + 2\varepsilon\vartheta^{-1} N(\Psi_c, \vartheta R).$$

Applying the projection operators $P^{0,1}$ to this equation we see that it is equivalent to the system of equations

$$\begin{aligned} \partial_t R^0 &= \Lambda R^0 + 2\varepsilon\vartheta^{-1} P^0 N(\Psi_c, \vartheta R^0) + 2\varepsilon\vartheta^{-1} P^0 N(\Psi_c, \vartheta R^1), && (34)\\ \partial_t R^1 &= \Lambda R^1 + 2\varepsilon\vartheta^{-1} P^1 N(\Psi_c, \vartheta R^0) + 2\varepsilon\vartheta^{-1} P^1 N(\Psi_c, \vartheta R^1). && (35) \end{aligned}$$

Recall the form of the nonlinear term in (26). Since $\widehat{\Psi}_c(k-m) = 0$ unless $|(k-m) \pm k_0| < \delta$ and since $\widehat{R}^0(m) = 0$ for $|m| > \delta$ we see that $P^0 N(\Psi_c, \vartheta R^0) = 0$ if $\delta > 0$ is sufficiently small, but independent of $0 < \varepsilon \ll 1$. Thus, we can replace (34) by

$$\begin{aligned} \partial_t R^0 &= \Lambda R^0 + 2\varepsilon\vartheta^{-1} P^0 N(\Psi_c, \vartheta R^1), && (36)\\ \partial_t R^1 &= \Lambda R^1 + 2\varepsilon\vartheta^{-1} P^1 N(\Psi_c, \vartheta R^1) + 2\varepsilon\vartheta^{-1} P^1 N(\Psi_c, \vartheta R^0). && (37) \end{aligned}$$

and we changed the order of terms in (35). We will now attempt to construct normal-form transformations to eliminate the $\mathcal{O}(\varepsilon)$ terms from (36) and (37). After constructing the normal-form transformations we will then go back and examine their effect on the full equation (25).

Given the form of the terms we wish to eliminate in (36) and (37) we look for a transformation of the form

$$\widetilde{R}^0_j = R^0_j + \varepsilon B^{0,1}_j(\Psi_c, R^1) \ , \ \widetilde{R}^1_j = R^1_j + \varepsilon B^{1,1}_j(\Psi_c, R^1) + \varepsilon B^{1,0}_j(\Psi_c, R^0) \tag{38}$$

where

$$\begin{aligned} \widehat{B}^{0,1}_{j_1}(\Psi_c, R^1)(k) &= \sum_{j_2,j_3=1}^{2} \int \widehat{b}^{0,1}_{j_1;j_2,j_3}(k, k-m, m)\widehat{\Psi}^+_{c,j_2}(k-m)\widehat{R}^1_{j_3}(m)dm && (39)\\ &\quad + \sum_{j_2,j_3=1}^{2} \int \widehat{b}^{0,1}_{j_1;j_2,j_3}(k, k-m, m)\widehat{\Psi}^-_{c,j_2}(k-m)\widehat{R}^1_{j_3}(m)dm, \end{aligned}$$

with analogous formulas for $B^{1,0}$ and $B^{1,1}$. Note that we have used the fact that $\widehat{\Psi}_{c,j}(k) = \widehat{\Psi}^+_{c,j}(k) + \widehat{\Psi}^-_{c,j}(k)$, where $\widehat{\Psi}^{\pm}_{c,j}(k)$ represent the parts of $\widehat{\Psi}_{c,j}(k)$ localized near $\pm k_0$ respectively. In the following discussion we will focus on the terms containing $\widehat{\Psi}^+_{c,j}$ – those containing $\widehat{\Psi}^-_{c,j}$ are treated in an almost identical fashion.

Construction of $\mathbf{B^{0,1}}$: If we differentiate the expression for $\widetilde{R}^0$ in (38) w.r.t t we obtain

$$\partial_t \widetilde{R}^0_j = \partial_t R^0_j + \varepsilon B^{0,1}(\partial_t \Psi_c, R^1) + \varepsilon B^{0,1}(\Psi_c, \partial_t R^1) \ . \tag{40}$$

Recall that $\|\partial_t \Psi_c - \Lambda \Psi_c\|_{Y^2_{\sigma,r}} \leq C(\sigma,r)\varepsilon^2$ from the definition of Ψ_c, while (30) implies

$$\|\partial_t R^0_j - \Lambda R^0_j - 2\varepsilon\vartheta^{-1} P^0 \widehat{N}(\Psi_c, \vartheta R)\|_{Y^2_{\sigma,r-\frac{1}{2}}} \leq C\varepsilon^2(1 + \|R\|_{Y^2_{\sigma,r}}) \ . \tag{41}$$

Thus, provided the transformation $B^{0,1}$ is well-defined and bounded we have

$$\begin{aligned} \partial_t \widetilde{R}^0_j \ &= \ \Lambda\left(\widetilde{R}^0_j - \varepsilon B^{0,1}(\Psi_c, R^1)\right) + 2\varepsilon\vartheta^{-1} P^0 \widehat{N}(\Psi_c, \vartheta R^1) + \varepsilon B^{0,1}(\Lambda\Psi_c, R^1) \\ &\quad + \varepsilon B^{0,1}(\Psi_c, \Lambda R^1) + 2\varepsilon B^{0,1}(\Psi_c, \varepsilon\vartheta^{-1} P^1 \widehat{N}(\Psi_c, \vartheta R^1)) + \varepsilon^2 \mathcal{E}^1 \ , \end{aligned} \tag{42}$$

where we recall that the notation $\varepsilon^2\mathcal{E}^1$ means that the $Y^2_{\sigma,r-1}$ norm of this term can be bounded by $C\varepsilon^2$ if R is in some bounded neighborhood of the origin in $Y^2_{\sigma,r}$. Note that the term $\varepsilon B^{0,1}(\Psi_c, \varepsilon\vartheta^{-1}\widehat{N}(\Psi_c, \vartheta R^1))$ appears formally to be $\mathcal{O}(\varepsilon^2)$, but we will see below that the kernel of the transformation B will be $\mathcal{O}(\varepsilon^{-1})$ for certain wave numbers so that this term is only $\mathcal{O}(\varepsilon)$ for those wave numbers and must therefore be retained. It will be eliminated by a second normal-form transformation. The term is not $\mathcal{O}(1)$ since the terms B and ϑ^{-1} are $\mathcal{O}(\varepsilon^{-1})$ at different wave numbers in this term as will be shown below,

Since we want to eliminate all terms which are formally $\mathcal{O}(\varepsilon)$, this suggests that we choose $B^{0,1}$ so that

$$-\Lambda B^{0,1}(\Psi_c, R^1) + B^{0,1}(\Lambda\Psi_c, R^1) + B^{0,1}(\Psi_c, \Lambda R^1) = -2\vartheta^{-1} P^0 \widehat{N}(\Psi_c, \vartheta R^1). \tag{43}$$

We find that the kernel of $B^{0,1}$ should be of the form:

$$\widehat{b}^{0,1}_{j_1;j_2,j_3}(k, k-m, m) \ = \ \frac{-2\widehat{P}^0(k)\widehat{\alpha}^{j_1}_{j_2,j_3}(k)}{(i\omega_{j_1}(k) - i\omega_{j_2}(k-m) - i\omega_{j_3}(m))} \frac{\widehat{\vartheta}(m)}{\widehat{\vartheta}(k)}. \tag{44}$$

Due to the fact that the $\widehat{P}^0$ and $\widehat{\Psi}_c$ have supports localized near $k=0$ and $(k-m) = \pm k_0$ respectively this expression only has to be analyzed for $|(k-m) \pm k_0| < \delta$ and $|k| < \delta$. As a consequence for $\delta > 0$ sufficiently small, but independent of $0 < \varepsilon \ll 1$, we can also restrict to wave numbers m bounded away from 0. Hence from the two possible resonances discussed above only the resonance at $k=0$ will play a role for $B^{0,1}$. The kernel $\widehat{b}^{0,1}_{j_1;j_2,j_3}$

can then be estimated as follows. First note that if we consider the denominator of this expression near $k = 0$ then we have

$$\begin{aligned}&\mathrm{i}\omega_{j_1}(k) - \mathrm{i}\omega_{j_2}(k-m) - \mathrm{i}\omega_{j_3}(m)\\ &\qquad = \mathrm{i}\omega'_{j_1}(0)k - \mathrm{i}(\omega_{j_2}(-m) + \omega'_{j_2}(-m)k) - \mathrm{i}\omega_{j_3}(m) + \mathcal{O}(k^2)\end{aligned}$$

If $\omega_{j_3}(m) \neq \omega_{j_2}(m)$ this quantity is bounded below by some $\mathcal{O}(1)$ constant for all $|k| < \delta$. If, on the other hand, $\omega_{j_3}(m) = \omega_{j_2}(m)$ there exists a positive constant C such that

$$|\mathrm{i}\omega_{j_1}(k) - \mathrm{i}\omega_{j_2}(k-m) - \mathrm{i}\omega_{j_3}(m)| \geq C|k| \ . \tag{45}$$

Here, we have used the fact that $m \approx \pm k_0$ because of the support of $\widehat{\Psi}_c$ and the fact the fact that $\omega'_j(\pm k_0)$ is $\mathcal{O}(1)$ for all j (and is not equal to $\omega'_k(0)$). In either case, $|\widehat{\alpha}^{j_1}_{j_2,j_3}(k)| \leq C|k|$, and thus there exists $C \geq 0$ such that

$$|\widehat{\vartheta}(k)\widehat{b}^{0,1}_{j_1;j_2,j_3}(k, k-m, m)| \leq C \tag{46}$$

for all $|k| \leq \delta$.

Because of the factor of $\widehat{P}^0(k)$ which makes $\widehat{b}^{0,1}_{j_1;j_2,j_3}(k) = 0$ if $|k| > \delta$, $B^{0,1}$ is "smoothing" in the sense that if $R^1 \in Y^2_{\sigma,r}$ for some $r > 1$, then given any σ', r', there exists $C_{\sigma',r'}$ such that

$$\|\varepsilon B^{0,1}(\Psi_c, R^1)\|_{Y^2_{\sigma',r'}} \leq C_{\sigma',r'}\|R^1\|_{Y^2_{\sigma,r}} \ . \tag{47}$$

In particular, this estimate holds when $\sigma' = \sigma$ and $r' = r$. Note, however, that in spite of the factor of ε in front of $B^{0,1}$, we cannot assume that $C_{\sigma',r'} \sim \mathcal{O}(\varepsilon)$ because of the factor of $\vartheta^{-1}(k) \sim \varepsilon^{-1}$ for $k \approx 0$, in the formula for the kernel of $B^{0,1}$.

Construction of $\mathbf{B^{1,0}}$ and $\mathbf{B^{1,1}}$: If we differentiate the expression for $\widetilde{R}^1$ in (38) w.r.t t we obtain

$$\partial_t\widetilde{R}^1_j = \partial_t R^1_j + \varepsilon B^{1,1}(\partial_t\Psi_c, R^1) + \varepsilon B^{1,1}(\Psi_c, \partial_t R^1) + \varepsilon B^{1,0}(\partial_t\Psi_c, R^0) + \varepsilon B^{1,0}(\Psi_c, \partial_t R^0). \tag{48}$$

Replacing $\partial_t\Psi_c$ and $\partial_t R^j$ as above we find

$$\begin{aligned}\partial_t\widetilde{R}^1_j \ = \ & \Lambda\left(\widetilde{R}^1_j - \varepsilon B^{1,1}(\Psi_c, R^1) - \varepsilon B^{1,0}(\Psi_c, R^0)\right)\\ & + 2\varepsilon\vartheta^{-1}P^1\widehat{N}(\Psi_c, \vartheta R^1) + \varepsilon B^{1,1}(\Lambda\Psi_c, R^1) + \varepsilon B^{1,1}(\Psi_c, \Lambda R^1)\\ & + 2\varepsilon\vartheta^{-1}P^1\widehat{N}(\Psi_c, \vartheta R^0) + \varepsilon B^{1,0}(\Lambda\Psi_c, R^0) + \varepsilon B^{1,0}(\Psi_c, \Lambda R^0) + \varepsilon^2\mathcal{E}^2,\end{aligned} \tag{49}$$

We recall that the notation $\varepsilon^2\mathcal{E}^2$ means that the $Y^2_{\sigma,r-1}$ norm of this term can be bounded by $C\varepsilon^2$ if R is in some bounded neighborhood of the origin in $Y^2_{\sigma,r}$. The terms $\varepsilon B^{1,j}(\Psi_c, \varepsilon\vartheta^{-1}\widehat{N}(\Psi_c, \vartheta R^j))$ for $j = 0, 1$ are included in these error terms.

To avoid a resonance problem at $\pm k_0$ we will replace the terms $2\varepsilon\vartheta^{-1}P^1\widehat{N}(\Psi_c, \vartheta R^1)$ and $2\varepsilon\vartheta^{-1}P^1\widehat{N}(\Psi_c, \vartheta R^1)$ in (49) by $2\varepsilon\vartheta^{-1}P^1\widehat{N}(\Psi_c, \vartheta_0 R^1)$ and $2\varepsilon\vartheta^{-1}P^1\widehat{N}(\Psi_c, \vartheta_0 R^1)$, where

$\hat{\vartheta}_0(k) = \hat{\vartheta}(k) - \varepsilon$. The key fact that we will use below is that $\hat{\vartheta}_0(0) = 0$. Making this change introduces additional error terms $2\varepsilon^2\vartheta^{-1}P^1\widehat{N}(\Psi_c, R^1)$ and $2\varepsilon^2\vartheta^{-1}P^1\widehat{N}(\Psi_c, R^1)$ into (49). However, since $\hat{\vartheta}^{-1}(k)$ is $\mathcal{O}(1)$ on the support of $\hat{P}^1$, these terms can be included in the error term of the form $\varepsilon^2\mathcal{E}^2$.

Equation (49) can now be rewritten as

$$\begin{aligned}
\partial_t \widetilde{R}^1_j &= \Lambda\left(\widetilde{R}^1_j - \varepsilon B^{1,1}(\Psi_c, R^1) - \varepsilon B^{1,0}(\Psi_c, R^0)\right) \\
&\quad +2\varepsilon\vartheta^{-1}P^1\widehat{N}(\Psi_c, \vartheta_0 R^1) + \varepsilon B^{1,1}(\Lambda\Psi_c, R^1) + \varepsilon B^{1,1}(\Psi_c, \Lambda R^1) \\
&\quad +2\varepsilon\vartheta^{-1}P^1\widehat{N}(\Psi_c, \vartheta_0 R^0) + \varepsilon B^{1,0}(\Lambda\Psi_c, R^0) + \varepsilon B^{1,0}(\Psi_c, \Lambda R^0) + \varepsilon^2\mathcal{E}^2,
\end{aligned} \tag{50}$$

Since we want to eliminate all terms of $\mathcal{O}(\varepsilon)$, this suggests that we choose $B^{1,j}$ so that

$$-\Lambda B^{1,j}(\Psi_c, R^j) + B^{1,j}(\Lambda\Psi_c, R^j) + B^{1,j}(\Psi_c, \Lambda R^j) = -2\vartheta^{-1}P^1\widehat{N}(\Psi_c, \vartheta_0 R^j) \tag{51}$$

for $j = 0, 1$. However, we use Lemma 9 to replace this with an equation for $B^{1,j}$ which will result in a form for the normal-form transformation that is easier to bound, at the expense of introducing additional "error" terms all of which are $\mathcal{O}(\varepsilon^2)$ and will be included in $\varepsilon^2\mathcal{E}^2$. More specifically we apply Lemma 9 and make three changes in (51):

(A.1) We replace $B^{1,j}(\Lambda\Psi_c, R^j)$ by $B^{1,j}(\Lambda_0\Psi_c, R^j)$ where

$$\begin{aligned}
&\widehat{B}^{1,j}_{j_1}(\Lambda_0\Psi_c, R^j)(k) \\
&= \sum_{j_2,j_3=1}^{2} \int \widehat{b}^{1,j;+}_{j_1;j_2,j_3}(k) i\omega_{j_2}(k_0)\widehat{\Psi}^+_{c,j_2}(k-m)\widehat{R}^j_{j_3}(m)dm \\
&\quad + \sum_{j_2,j_3=1}^{2} \int \widehat{b}^{1,j;-}_{j_1;j_2,j_3}(k) i\omega_{j_2}(-k_0)\widehat{\Psi}^-_{c,j_2}(k-m)\widehat{R}^j_{j_3}(m)dm.
\end{aligned}$$

(A.2) We replace $B^{1,j}(\Psi_c, \Lambda R^j)$ by $B^{1,j}(\Psi_c, \Lambda_c R^j)$ where

$$\begin{aligned}
&\widehat{B}^{1,j}_{j_1}(\Psi_c, \Lambda_c R^j)(k) \\
&= \sum_{j_2,j_3=1}^{2} \int \widehat{b}^{1,j;+}_{j_1;j_2,j_3}(k)\widehat{\Psi}^+_{c,j_2}(k-m) i\omega_{j_3}(k-k_0)\widehat{R}^j_{j_3}(m)dm \\
&\quad + \sum_{j_2,j_3=1}^{2} \int \widehat{b}^{1,j;-}_{j_1;j_2,j_3}(k)\widehat{\Psi}^-_{c,j_2}(k-m) i\omega_{j_3}(k+k_0)\widehat{R}^j_{j_3}(m)dm.
\end{aligned}$$

(A.3) We replace $\vartheta^{-1}P^1\widehat{N}(\Psi_c, \vartheta_0 R^j)$ by $\vartheta^{-1}P^1\widehat{N}^+(\Psi^+_c, \vartheta_0 R^j) + \vartheta^{-1}P^1\widehat{N}^-(\Psi^-_c, \vartheta_0 R^j)$ where

$$\begin{aligned}
&\varepsilon\vartheta^{-1}P^1\widehat{N}^{j_1,\pm}(\Psi^\pm_c, \vartheta_0 R^j) \\
&\quad = -\varepsilon\vartheta^{-1}(k)\widehat{P}^1(k) \sum_{j_2,j_3=1}^{2} \int \widehat{\alpha}^{j_1}_{j_2,j_3}(k)\widehat{\vartheta}_0(k \mp k_0)\widehat{\Psi}^\pm_{c,j_2}(k-m)\widehat{R}^j_{j_3}(m)dm.
\end{aligned}$$

Inserting these changes into (51) we find that the kernel of $B^{1,j}$ should be of the form:

$$\hat{b}^{1,j;+}_{j_1;j_2,j_3} = \frac{2\widehat{P}^1(k)\widehat{\alpha}^{j_1}_{j_2,j_3}(k)}{(i\omega_{j_1}(k) - i\omega_{j_2}(k_0) - i\omega_{j_3}(k-k_0))} \frac{\widehat{\vartheta}_0(k-k_0)}{\widehat{\vartheta}(k)} \tag{52}$$

with a similar expression for $\widehat{b}^{1,j,-}_{j_1;j_2,j_3}$.

Remark 12. *The analysis of the kernel of $B^{0,1}$ would be simplified by the changes (A1)–(A3), too. However, we have not made these changes since $B^{0,1} = \mathcal{O}(\varepsilon^{-1})$ for certain wave numbers which would complicate the analysis of the subsequent second normal-form transformation.*

Due to the support of $\widehat{\Psi}_c$ and of $\widehat{P}^1$ the expression (52) only has to be analyzed for $|k-m-k_0| < \delta$ and $|k| \geq \delta$. We now consider the possible resonances in the denominator of (52).

- $k = 0$: Since $\widehat{P}^1(k) = 0$ for $|k| \leq \delta$, this resonance does not play a role in the analysis of either $B^{1,0}$ or $B^{1,1}$.
- $k = k_0$: There is a resonance at $k = k_0$ whenever $j_1 = j_2$. However, since the derivative of ω_j at k_0 is $\mathcal{O}(1)$, we have a bound on the denominator of the form

$$|i\omega_{j_1}(k) - i\omega_{j_2}(k_0) - i\omega_{j_3}(k-k_0)| \geq C|k-k_0| \tag{53}$$

  This singularity is offset, however, by the fact that the term $|\widehat{\vartheta}_0(k-k_0)| \leq C|k-k_0|$ and hence the kernel $\hat{b}^{1,j;+}_{j_1;j_2,j_3}$ can be extended continuously at $k = k_0$ with an $\mathcal{O}(1)$ bound on its size.

There are no other resonances for this term in the normal form and hence the kernel can be bounded for all values of k and m by an $\mathcal{O}(1)$ bound.

Applying Young's inequality one easily establishes that if $R^1 \in Y^2_{\sigma,r}$ then there exists $C > 0$ such that

$$\|\varepsilon B^{1,1}(\Psi_c, R^1)\|_{Y^2_{\sigma,r-1/2}} \leq C\varepsilon \|R^1\|_{Y^2_{\sigma,r}} \ . \tag{54}$$

Note that in this case there is a loss of "1/2 a derivative" – i.e. we get a bound of $B^{1,1}$ in the space $Y^2_{\sigma,r-1/2}$ rather than $Y^2_{\sigma,r}$. This is due to the growth of $\hat{\alpha}^{j_1}_{j_2,j_3} \sim |\omega_1(k)| \sim \sqrt{k}$ as $|k| \to \infty$. One the other hand, since we don't have to deal with the large values of $\vartheta^{-1}(k)$ near $k \approx 0$ we obtain a factor of ε on the right hand side of this estimate.

Due to the compact support of $\widehat{R}^0$ this loss of regularity is not present in the estimate for $B^{1,0}$. We find analogously

$$\|\varepsilon B^{1,0}(\Psi_c, R^0)\|_{Y^2_{\sigma,r}} \leq C\varepsilon \|R^0\|_{Y^2_{\sigma,r}} \ . \tag{55}$$

We can sum up the results of this first normal-form transformation as follows:

Proposition 13. *Define*

$$\widetilde{R}_j^0 = R_j^0 + \varepsilon B_j^{0,1}(\Psi_c, R^1)\ ,\ \widetilde{R}_j^1 = R_j^1 + \varepsilon B_j^{1,1}(\Psi_c, R^1) + \varepsilon B_j^{1,0}(\Psi_c, R^0) \tag{56}$$

for $j = 1, 2$. This transformation maps $(R_j^0, R_j^1) \in Y_{\sigma,r}^2 \times Y_{\sigma,r}^2$ into $(\widetilde{R}_j^0, \widetilde{R}_j^1) \in Y_{\sigma,r}^2 \times Y_{\sigma,r-1/2}^2$ for all $r > 1$ and $\sigma \geq 0$ and is invertible on its range. Furthermore, if we write the inverse transformations as

$$R_j^0 = \widetilde{R}_j^0 + \mathcal{B}_{0,j}^{-1}(\widetilde{R}^0, \widetilde{R}^1),\ \ R_j^1 = \widetilde{R}_j^1 + \mathcal{B}_{1,j}^{-1}(\widetilde{R}^0, \widetilde{R}^1)\ ,$$

then there exist constants C_0, C_1 such that the inverse transformations satisfy the estimates

$$\begin{aligned} \|\mathcal{B}_{0,j}^{-1}(\widetilde{R}^0, \widetilde{R}^1)\|_{Y_{\sigma,r}^2} &\leq C_0(\|\widetilde{R}^0\|_{Y_{\sigma,r}^2} + \|\widetilde{R}^1\|_{Y_{\sigma,r}^2}) \\ \|\mathcal{B}_{1,j}^{-1}(\widetilde{R}^0, \widetilde{R}^1)\|_{Y_{\sigma,r}^2} &\leq C_1\varepsilon(\|\widetilde{R}^0\|_{Y_{\sigma,r}^2} + \|\widetilde{R}^1\|_{Y_{\sigma,r}^2}) \end{aligned}$$

for $j = 1, 2$. Finally, if (R_j^0, R_j^1) satisfy the equations (36) *and* (37) *then $(\widetilde{R}_j^0, \widetilde{R}_j^1)$ satisfy*

$$\begin{aligned} \partial_t \widetilde{R}_j^0 &= \Lambda \widetilde{R}_j^0 + \varepsilon B^{0,1}(\Psi_c, \varepsilon\vartheta^{-1} P^1 N(\Psi_c, \vartheta\widetilde{R}^1)) + \varepsilon^2 \mathcal{E}^3 \\ \partial_t \widetilde{R}_j^1 &= \Lambda \widetilde{R}_j^1 + \varepsilon^2 \mathcal{E}^4, \end{aligned} \tag{57}$$

where we recall that the notation $\varepsilon^2 \mathcal{E}^j$ refers to terms whose $Y_{\sigma,r-1}^2$ norm is bounded by $C_j \varepsilon^2$ for $(\widetilde{R}^0, \widetilde{R}^1)$ in some fixed ball in $Y_{\sigma,r}^2$.

Proof. The proof of invertibility of the transformation is deferred until the next section. Assuming the invertibility for the moment the structure of the equations (57) follows immediately using $R_j^1 = \widetilde{R}_j^1 + \mathcal{O}(\varepsilon)$. □

The second normal-form transformation: We now construct a second normal-form transformation to remove the one remaining term of $\mathcal{O}(\varepsilon)$ from (57). Before doing so we analyze the offending term in more detail. Recall that $\Psi_c = \Psi_c^+ + \Psi_c^-$ where the Fourier transform of $\Psi_c^\pm$ is supported in a neighborhood of size δ of $\pm k_0$. Thus, we can write

$$\begin{aligned} &\varepsilon B^{0,1}(\Psi_c, \varepsilon\vartheta^{-1} P^1 N(\Psi_c, \vartheta\widetilde{R}^1)) = \varepsilon B^{0,1}(\Psi_c^+, \varepsilon\vartheta^{-1} P^1 N(\Psi_c^+, \vartheta\widetilde{R}^1)) \\ &\quad + \varepsilon B^{0,1}(\Psi_c^-, \varepsilon\vartheta^{-1} P^1 N(\Psi_c^-, \vartheta\widetilde{R}^1)) + \varepsilon B^{0,1}(\Psi_c^+, \varepsilon\vartheta^{-1} P^1 N(\Psi_c^-, \vartheta\widetilde{R}^1)) \\ &\qquad + \varepsilon B^{0,1}(\Psi_c^-, \varepsilon\vartheta^{-1} P^1 N(\Psi_c^+, \vartheta\widetilde{R}^1)). \end{aligned} \tag{58}$$

Each of the four terms can be rewritten as

$$\begin{aligned} &\varepsilon \widehat{B}_j^{0,1}(\Psi_c^l, \varepsilon\vartheta^{-1} P^1 N(\Psi_c^n, \vartheta\widetilde{R}^1))(k) \\ &\quad = \varepsilon^2 \sum_{k_1,k_2=1,2} \int \widehat{b}_{j;k_1,k_2}^{0,1}(k, k-\ell, \ell) \widehat{\Psi}_{c,k_1}^l(k-\ell) \\ &\qquad \times \vartheta^{-1}(\ell) \widehat{P}^1(\ell) \left(\sum_{k_3,k_4=1,2} \int \widehat{\alpha}_{k_3,k_4}^{k_2}(\ell-m) \widehat{\Psi}_{c,k_3}^n(\ell-m) \widehat{\vartheta}(m) \widehat{\widetilde{R}}_{k_4}^1(m) dm \right) d\ell, \end{aligned} \tag{59}$$

where the variable $l, n \in \{+,-\}$ and where we recall that

$$\widehat{b}^{0,1}_{j;k_1,k_2}(k, k-\ell, \ell) = \frac{2\widehat{P}^0(k)\widehat{\alpha}^j_{k_1,k_2}(k)}{i(\omega_j(k) - \omega_{k_1}(k-\ell) - \omega_{k_2}(\ell))} \frac{\widehat{\vartheta}(\ell)}{\widehat{\vartheta}(k)}.$$

We now apply Lemma 9 to simplify this expression as we did in $B^{1,j}$ $(j = 0,1)$. If we do so we obtain the expression

$$\begin{aligned}
&\varepsilon \widehat{B}^{0,1}_j(\Psi^l_c, \varepsilon\vartheta^{-1}(\cdot - lk_0)P^1 N(\Psi^n_c, \vartheta(\cdot - nk_0)\widetilde{R}^1))(k) \qquad (60)\\
&\quad = \varepsilon^2 \sum_{k_1,k_2=1,2} \int \widehat{b}^{0,1,l,n}_{j;k_1,k_2}(k)\widehat{\Psi}^l_{c,k_1}(k-\ell)\vartheta^{-1}(k-lk_0)\widehat{P}^1(k-lk_0) \\
&\qquad \times \left(\sum_{k_3,k_4=1,2} \int \widehat{\alpha}^{k_2}_{k_3,k_4}(k-lk_0)\widehat{\Psi}^n_{c,k_3}(\ell-m)\widehat{\vartheta}(k+(l+n)k_0)\widehat{\widetilde{R}}^1_{k_4}(m)dm \right) d\ell + \varepsilon^2\mathcal{E}^3,
\end{aligned}$$

with $l, n \in \{+,-\}$. (We interpret expressions like $l+n$ as if l and n were $+1$ and -1.)

We use the abbreviation:

$$\widehat{b}^{0,1,l,n}_{j;k_1,k_2}(k) = \frac{2\widehat{P}^0(k)\widehat{\alpha}^j_{k_1,k_2}(k)}{i(\omega_j(k) - \omega_{k_1}(lk_0) - \omega_{k_2}(k-lk_0))} \frac{\widehat{\vartheta}(k+(l+n)k_0)}{\widehat{\vartheta}(k)}.$$

With these modifications we can now prove that the last two of the terms in (58) are $\mathcal{O}(\varepsilon^2)$ and hence can be included in the $\varepsilon^2\mathcal{E}^3$ terms in (57):

Lemma 14. *There exists $C > 0$ such that*

$$\begin{aligned}
\|\varepsilon B^{0,1}(\Psi^+_c, \varepsilon\vartheta^{-1}P^1 N(\Psi^-_c, \vartheta\widetilde{R}^1))\|_{Y^2_{\sigma,r}} &\leq C\varepsilon^2\|\widetilde{R}^1\|_{Y^2_{\sigma,r}}, \\
\|\varepsilon B^{0,1}(\Psi^-_c, \varepsilon\vartheta^{-1}P^1 N(\Psi^+_c, \vartheta\widetilde{R}^1))\|_{Y^2_{\sigma,r}} &\leq C\varepsilon^2\|\widetilde{R}^1\|_{Y^2_{\sigma,r}}.
\end{aligned}$$

Proof. Since $B^{0,1}$ contains the factor $\widehat{P}^0(k)$ means that the integral over k which occurs in the $Y^2_{\sigma,r}$ norm runs only over the integral $|k| < \delta$. Thus, we can bound the $Y^2_{\sigma,r}$ norm by bounding the maximum of the kernel. The first term in Lemma 14 has the modified kernel

$$\varepsilon^2\widehat{b}^{0,1,+,-}_{j;k_1,k_2}(k)\widehat{\vartheta}^{-1}(k-k_0)\widehat{P}^1(k-k_0)\widehat{\alpha}^{k_2}_{k_3,k_4}(k-k_0)\widehat{\vartheta}(k). \qquad (61)$$

Since $\widehat{b}^{0,1,+,-}_{j;k_1,k_2}(k) = \widehat{\widetilde{b}}^{0,1,+,-}_{j;k_1,k_2}(k)/\widehat{\vartheta}(k)$ with $\widehat{\widetilde{b}}^{0,1,+,-}_{j;k_1,k_2}(k)$ being $\mathcal{O}(1)$ bounded, in the kernel (61) the factor $\widehat{\vartheta}(k)^{-1}$ cancels with the factor $\widehat{\vartheta}(k)$. Since all other terms in (61) are $\mathcal{O}(1)$ bounded for $|k| < \delta$ we have an $\mathcal{O}(\varepsilon^2)$ bound for the kernel (61). The second term in Lemma 14 can be estimated similarly. $\square$

Lemma 14 implies that the third and fourth terms in (58) need not be eliminated by the normal form transformation. Thus we now turn to the first two terms in this equation.

If we simplify the kernel of the first term in (58) with the aid of Lemma 9, we find its kernel has the form:

$$\varepsilon^2 \widehat{b}^{0,1,+,+}_{j;k_1,k_2}(k)\widehat{\vartheta}^{-1}(k-k_0)\widehat{P}^1(k-k_0)\widehat{\alpha}^{k_2}_{k_3,k_4}(k-k_0)\widehat{\vartheta}(k+2k_0) \tag{62}$$

plus errors that are of size $\mathcal{O}(\varepsilon^2)$. We obtain a very similar expression for the kernel of the second term in (58). Note that in contrast to the terms considered in Lemma 14 this expression does not contain a factor of $\hat{\vartheta}(k)$ to offset the $\hat{\vartheta}(k)$ in the denominator of $\widehat{b}^{0,1,+,+}_{j;k_1,k_2}(k)$ and thus they must be eliminated by a second normal form transformation.

We look for a transformation of the form

$$\begin{aligned} \mathcal{R}^0_j &= \widetilde{R}^0_j + \varepsilon D^{0,1,+}_j(\Psi^+_c, \Psi^+_c, \widetilde{R}^1) + \varepsilon D^{0,1,-}_j(\Psi^-_c, \Psi^-_c, \widetilde{R}^1) \\ \mathcal{R}^0_j &= \widetilde{R}^1_j \,. \end{aligned} \tag{63}$$

Differentiating the expression for $\mathcal{R}^0_j$ we find, just as in (40) and (51), that the terms of $\mathcal{O}(\varepsilon)$ in (57) will be eliminated if $D^{0,1,+}_j$ satisfies

$$\begin{aligned} \Big\{ & -\Lambda D^{0,1,+}(\Psi^+_c, \Psi^+_c, \widetilde{R}^1) + D^{0,1,+}(\partial_t\Psi^+_c, \Psi^+_c, \widetilde{R}^1) + D^{0,1,+}(\Psi^+_c, \partial_t\Psi^+_c, \widetilde{R}^1) \\ & + D^{0,1,+}(\Psi^+_c, \Psi^+_c, \partial_t\widetilde{R}^1) + \varepsilon\vartheta^{-1}B^{0,1}(\Psi^+_c, \varepsilon\vartheta^{-1}P^1N(\Psi^+_c, \vartheta\widetilde{R}^1))\Big\} = 0, \end{aligned} \tag{64}$$

or equivalently

$$\begin{aligned} \Big\{ & -\Lambda D^{0,1,+}(\Psi^+_c, \Psi^+_c, \widetilde{R}^1) + D^{0,1,+}(\Lambda\Psi^+_c, \Psi^+_c, \widetilde{R}^1) + D^{0,1,+}(\Psi^+_c, \Lambda\Psi^+_c, \widetilde{R}^1) \\ & + D^{0,1,+}(\Psi^+_c, \Psi^+_c, \Lambda\widetilde{R}^1) + \varepsilon\vartheta^{-1}B^{0,1}(\Psi^+_c, \varepsilon\vartheta^{-1}P^1N(\Psi^+_c, \vartheta\widetilde{R}^1))\Big\} = 0, \end{aligned} \tag{65}$$

with similar expressions for $D^{0,1,-}$. We find that we have to choose

$$\begin{aligned} & \varepsilon D^{0,1,+}_j(\Psi^+_c, \Psi^+_c, \widetilde{R}^1) \\ & \quad = \varepsilon^2 \sum_{k_1,k_2=1,2} \int \widehat{b}^{0,1,+,+}_{j;k_1,k_2}(k)\widehat{\Psi}^+_{c,k_1}(k-\ell)\vartheta^{-1}(k-k_0)\widehat{P}^1(k-k_0) \\ & \qquad \times \left(\sum_{k_3,k_4=1,2} \int \frac{\widehat{\alpha}^{k_2}_{k_3,k_4}(k-k_0)\widehat{\Psi}^+_{c,k_3}(\ell-m)\widehat{\vartheta}(k+2k_0)\widehat{\widetilde{R}}^1_{k_4}(m)}{-\omega_j(k)+\omega_{k_1}(k_0)+\omega_{k_3}(k_0)+\omega_{k_4}(k+2k_0)} dm \right) d\ell \end{aligned} \tag{66}$$

where we used as above in the kernel that $k-\ell \approx \ell - m \approx k_0$ due to the localization of $\widehat{\Psi}^+_c$ so we have $m \approx -2k_0$ which is made rigorous with Lemma 9. According to Young's inequality we have to estimate the kernel w.r.t. the sup norm. We already know that the numerator in this expression in $\mathcal{O}(\varepsilon)$. In order to estimate the denominator note that in this expression $k \approx 0$ due to the factor of $\widehat{P}^0$ in $\widehat{b}^{0,1,+,+}_{j;k_1,k_2}(k)$ and that $k_1 = k_3 = 1$ without summing over k_1 and k_3 in (66) due to (17). Hence

$$(-\omega_j(k) + \omega_{k_1}(k_0) + \omega_{k_3}(k_0) + \omega_{k_4}(k + 2k_0)) \approx 2\omega_{k_1}(k_0) + \omega_{k_4}(2k_0) \neq 0.$$

Regardless of the value of k_1 and k_4 this expression is bounded strictly away from zero. Hence the mapping $\varepsilon D^{0,1,+}$ is $\mathcal{O}(\varepsilon)$-bounded. We can construct and estimate the expression for $D^{0,1,-}$ in a very similar fashion. Therefore, the normal form is well defined and invertible. We find

Lemma 15. *If*

$$\mathcal{R}^0 = \widetilde{R}^0 + \varepsilon D^{0,1,+}(\Psi_c^+, \Psi_c^+, \widetilde{R}^1) + \varepsilon D^{0,1,-}(\Psi_c^-, \Psi_c^-, \widetilde{R}^1)$$

with $\varepsilon D^{0,1,\pm}$ defined as in (66)*, then for any $\sigma \geq 0$ and $r > 1$ there exists $C > 0$ such that*

$$\|\varepsilon D^{0,1,\pm}(\Psi_c^+, \Psi_c^+, \widetilde{R}^1)\|_{Y^2_{\sigma,r}} \leq C\varepsilon \|\widetilde{R}^1\|_{Y^2_{\sigma,r}} \ .$$

Remark 16. *Note that there is no loss of smoothness in this transformation due to the factor of $\widehat{P}^0$ in* (66) *via $\widehat{b}^{0,1,l,n}_{j;k_1,k_2}(k)$.*

Now, just as in Proposition 13 we have:

Proposition 17. *Fix $\sigma \geq 0$ and $r \geq 1$. Suppose $(\widetilde{R}^0, \widetilde{R}^1)$ satisfy the equations* (57)*. Define $(\mathcal{R}^0, \mathcal{R}^1)$ via the transformations* (63)*. Then for any $\rho > 0$, there exists ε_ρ such that for all $|\varepsilon| < \varepsilon_\rho$ the transformation* (63) *is invertible on the ball of radius ρ in $Y^2_{\sigma,r}$. Furthermore, $(\mathcal{R}^0, \mathcal{R}^1)$ satisfy the equations*

$$\begin{aligned} \partial_t \mathcal{R}^0 &= \Lambda \mathcal{R}^0 + \varepsilon^2 \mathcal{E}^5 \\ \partial_t \mathcal{R}^1 &= \Lambda \mathcal{R}^1 + \varepsilon^2 \mathcal{E}^6 \end{aligned} \tag{67}$$

Proof. The invertibility of the transformation in this case results from a simple application of the Neumann series since there is no loss of smoothness. The equation for $\mathcal{R}^0$ and $\mathcal{R}^1$ follow in the same way the equations for $\widetilde{R}^0$ and $\widetilde{R}^1$ were derived in the proof of Proposition 13. $\square$

Finally, we consider the composition of the two normal-form transformations, namely

$$\begin{aligned} \mathcal{R}^0 &= \widetilde{R}^0 + \varepsilon D^{0,1,+}(\Psi_c^+, \Psi_c^+, \widetilde{R}^1) + \varepsilon D^{0,1,-}(\Psi_c^-, \Psi_c^-, \widetilde{R}^1) \\ &= R^0 + \varepsilon B^{0,1}(\Psi, R^1) + \varepsilon D^{0,1,+}(\Psi_c^+, \Psi_c^+, R^1 + \varepsilon B^{1,0}(\Psi, R^0) + \varepsilon B^{1,1}(\Psi, R^1)) \\ &\qquad + \varepsilon D^{0,1,-}(\Psi_c^-, \Psi_c^-, R^1 + \varepsilon B^{1,0}(\Psi, R^0) + \varepsilon B^{1,1}(\Psi, R^1)) \\ &\equiv R^0 + \varepsilon F^0(R) \ , \end{aligned} \tag{68}$$

with a similar expression for $\mathcal{R}^1 \equiv R^1 + \varepsilon F^1(R)$. From Proposition 13 and Proposition 17 we see that

1. F^0 and F^1 are linear functions of R, and
2. The (composite) normal-form transformation loses at most half a derivative, i.e. there exists a constant C_F such that

$$\|\varepsilon F^1(R)\|_{Y^2_{\sigma,r-1/2}} \le C_F \varepsilon \|R\|_{Y^2_{\sigma,r-1/2}} \ .$$

   There is no loss of regularity in F^0 due to its compact support in Fourier space.

If we now insert use the information we have derived on the equations satisfied by the transformed variables we find the following proposition:

Proposition 18. *There exists a (linear) change of variables,*

$$\mathcal{R} = R + \varepsilon F(R)$$

defined for $R = (R_1, R_2) \in Y^2_{\sigma,r} \times Y^2_{\sigma,r}$ and invertible on its range such that in terms of the transformed variables the equation (25) *for the evolution of the error in our approximation takes the form*

$$\partial_t \mathcal{R} = \Lambda \mathcal{R} + \varepsilon^2 \ell(\mathcal{R}) + \varepsilon^\beta G(\mathcal{R}) + \varepsilon^{-\beta} \vartheta^{-1} \mathrm{Res}(\varepsilon \Psi) \ . \tag{69}$$

Furthermore the linear term $\varepsilon^2 \ell(\mathcal{R})$ and the bilinear term $\varepsilon^\beta G(\mathcal{R})$ satisfy the estimates

$$\|\varepsilon^2 \ell(\mathcal{R})\|_{Y^2_{\sigma,r-1}} \le C_L \varepsilon^2 \|\mathcal{R}\|_{Y^2_{\sigma,r}} \ ,$$

and

$$\|\varepsilon^\beta G(\mathcal{R})\|_{Y^2_{\sigma,r-1}} \le C_G \varepsilon^\beta \|\mathcal{R}\|_{Y^2_{\sigma,r}} \|\mathcal{R}\|_{Y^2_{\sigma,r-1}} \ .$$

Proof. The proof follows immediately from the estimates in Proposition 13, Proposition 17 and Lemma 2. □

5 Inverting the first normal-form transformation

To complete the derivation of the evolution equation for $(\mathcal{R}^0, \mathcal{R}^1)$ in Proposition 17 we now prove the invertibility of the first normal form transformation asserted in Proposition 13.

The difficulty in the inversion comes from the fact that $B_j^{1,1}$ loses half a derivative – i.e. in order to estimate $B_j^{1,1}(\Psi, R^1)$ in $Y^2_{\sigma,r}$ we must know that $R^1 \in Y^2_{\sigma,r+1/2}$. Therefore, inverting the normal-form transformation with the help of Neumann's series is not possible. Nonetheless the normal-form transformation is invertible due to a sort of energy estimate. We illustrate this inversion by looking first at a pair of model problems that exhibit this phenomenon in a somewhat simpler setting.

i) As our first example consider the transformation $u = F(v) = v + \varepsilon a \partial_x v$ with $0 < \varepsilon \ll 1$ a small parameter, and a some smooth function. As in our normal-form transformation,

F exhibits a loss of smoothness. Namely, in order to estimate $F(v)$ in the Sobolev space H^s we need to know that $v \in H^{s+1}$, in which case we have the estimate

$$\|u\|_{H^s} \leq \|v\|_{H^s} + \varepsilon \|a\|_{C_b^{s+1}} \|v\|_{H^{s+1}} \ .$$

Nevertheless, F is 1–1 and hence invertible on its range and we can estimate v in terms of u as follows:

$$\begin{aligned}
\int (\partial_x^s u)(\partial_x^s v) &= \int (\partial_x^s v)^2 + \varepsilon \int (\partial_x^s v)\partial_x^s(a\partial_x v) \\
&= \int (\partial_x^s v)^2 + \varepsilon \int a(\partial_x^s v)(\partial_x^{s+1} v) + \mathcal{O}(\varepsilon \|v\|_{H^s}^2) \\
&= \|\partial_x^s v\|_{L^2}^2 + \frac{\varepsilon}{2} \int a\partial_x((\partial_x^s v)^2) + \mathcal{O}(\varepsilon \|v\|_{H^s}^2) \\
&= \|\partial_x^s v\|_{L^2}^2 - \frac{\varepsilon}{2} \int (\partial_x a)(\partial_x^s v)^2 + \mathcal{O}(\varepsilon \|v\|_{H^s}^2) \ .
\end{aligned}$$

Summing up all estimates yields

$$\|v\|_{H^s}^2 \leq \|v\|_{H^s} \|u\|_{H^s} + C\varepsilon \|v\|_{H^s}^2$$

and hence

$$\|v\|_{H^s} \leq \|u\|_{H^s} + C\varepsilon \|v\|_{H^s},$$

which finally gives

$$\|v\|_{H^s} \leq \frac{1}{1 - C\varepsilon} \|u\|_{H^s} \ .$$

ii) The transformations we constructed in the previous section are expressed as convolutions of the Fourier transforms so our next example illustrates that the key property necessary in this context is that the kernel function in the convolution is purely imaginary and Lipschitz. Consider:

$$\widehat{u}(k) = \widehat{v}(k) + \int \widehat{b}(k)\widehat{a}(k-m)\widehat{v}(m)\,\mathrm{d}m$$

We assume that:

- $\widehat{b}(k)$ is pure imaginary.
- $\widehat{b}(k)$ is Lipshitz as a function of k.
- $\widehat{b}(k) \sim ik$ for $|k| \to \infty$

As before, a is assumed to be smooth and real valued. We then find

$$\begin{aligned}\int \overline{\widehat{v}(k)}\widehat{u}(k)+\widehat{v}(k)\overline{\widehat{u}(k)}dk &= 2\int \overline{\widehat{v}(k)}\widehat{v}(k)dk+\int \overline{\widehat{v}(k)}\widehat{b}(k)\widehat{a}(k-m)\widehat{v}(m)\,\mathrm{d}m\,\mathrm{d}k \\ &\quad +\int \widehat{v}(k)\overline{\widehat{b}(k)}\ \overline{\widehat{a}(k-m)}\ \overline{\widehat{v}(m)}\,\mathrm{d}m\,\mathrm{d}k \\ &= 2\int \overline{\widehat{v}(k)}\widehat{v}(k)dk+\int \overline{\widehat{v}(k)}\widehat{b}(k)\widehat{a}(k-m)\widehat{v}(m)\,\mathrm{d}m\,\mathrm{d}k \\ &\quad +\int \widehat{v}(m)\overline{\widehat{b}(m)}\ \overline{\widehat{a}(m-k)}\ \overline{\widehat{v}(k)}\,\mathrm{d}k\,\mathrm{d}m \\ &= 2\int \overline{\widehat{v}(k)}\widehat{v}(k)dk+\int \overline{\widehat{v}(k)}\widehat{a}(k-m)\widehat{v}(m)(\widehat{b}(k)+\overline{\widehat{b}(m)})\,\mathrm{d}k\,\mathrm{d}m\end{aligned}$$

where we used $\widehat{a}(\ell)=\overline{\widehat{a}(-\ell)}$ due to the fact that a is real-valued. Hence

$$2\|\widehat{v}\|_{L^2}^2 \leq 2\|\widehat{v}\|_{L^2}\|\widehat{u}\|_{L^2}+s_1$$

where with the Gagliardo–Nirenberg inequality

$$\begin{aligned}s_1 &= \left|\int \overline{\widehat{v}(k)}\widehat{v}(m)\widehat{a}(k-m)(\widehat{b}(k)+\overline{\widehat{b}(m)})\,\mathrm{d}k\,\mathrm{d}m\right| \\ &\leq \int |\overline{\widehat{v}(k)}\widehat{v}(m)||\widehat{a}(k-m)|C|k-m|\,\mathrm{d}m\,\mathrm{d}k \\ &\leq \|v\|_{L^2}^2\int |\widehat{a}(\ell)|C|\ell|\,\mathrm{d}\ell\end{aligned}$$

since $|\widehat{b}(k)+\overline{\widehat{b}(m)}| = |\widehat{b}(k)-\widehat{b}(m)| \leq C|k-m|$ if $\widehat{b}$ is Lipschitz-continuous and purely imaginary.

Note that if instead of estimating the normal-form transformation in the Sobolev spaces H^s we apply the above ideas in the spaces $Y^2_{\sigma,r}$, then using the estimate $e^{\sigma|k|}\leq e^{\sigma|k-m|}e^{\sigma|m|}$ for $\sigma\geq 0$ one can conclude in a very similar fashion the estimates

$$\|u\|_{Y^2_{\sigma,r}}\leq C\|v\|_{Y^2_{\sigma,r+1}} \quad \text{and} \quad \|v\|_{Y^2_{\sigma,r}}\leq C\|u\|_{Y^2_{\sigma,r}}$$

for b sufficiently small.

iii) Finally, we turn to the first normal-form transformation constructed in the previous section:

$$\begin{aligned}\widetilde{R}^0_j &= R^0_j+\varepsilon B^{0,1}_j(\Psi,R^1) \\ \widetilde{R}^1_j &= R^1_j+\varepsilon B^{1,1}_j(\Psi,R^1)+\varepsilon B^{1,0}_j(\Psi,R^0)\end{aligned} \tag{70}$$

for $j=1,2$. Recall that only the terms $B^{1,1}_j$ lose smoothness. Both $B^{0,1}_j$ and $B^{1,0}_j$ are bounded transformations from $Y^2_{\sigma,r}$ to $Y^2_{\sigma,r}$. Thus, we first consider just

$$\widetilde{R}^1_j = R^1_j+\varepsilon B^{1,1}_j(\Psi,R^1)+\varepsilon B^{1,0}_j(\Psi,R^0)\ . \tag{71}$$

From the previous section we know that

$$\widehat{B}_j^{1,1}(\Psi, R^1)(k) = \sum_{k_1,k_2=1,2} \sum_{\sigma=\pm} \int \widehat{b}_{j;k_1,k_2}^{1,1,\sigma}(k) \widehat{\Psi}_{c,k_1}^{\sigma}(k-m) \widehat{R}_{k_2}^1(m) dm$$

where from the explicit formula in (52) one can immediately verify that $\widehat{b}_{j;k_1,k_2}^{1,1,\sigma}(k)$ satisfies the conditions on the kernel that we required to establish the energy estimate in example (ii). Thus, we multiply both sides of (71) by R_j^1 and take the $Y_{\sigma,r}^2$ norm of both sides, and add together the estimates for $j=1$ and $j=2$ then just as in example (ii) we find

$$\|R^1\|_{Y_{\sigma,r}^2}^2 \le \|R^1\|_{Y_{\sigma,r}^2} \|\widetilde{R}^1\|_{Y_{\sigma,r}^2} + C_1 \varepsilon \|R^1\|_{Y_{\sigma,r}^2}^2 + C_2 \varepsilon (\|R^1\|_{Y_{\sigma,r}^2}^2 + \|R^0\|_{Y_{\sigma,r}^2}^2) \tag{72}$$

where $\|R^1\|_{Y_{\sigma,r}^2}^2 = \|R_1^1\|_{Y_{\sigma,r}^2}^2 + \|R_2^1\|_{Y_{\sigma,r}^2}^2$ and similarly for $\|\widetilde{R}^1\|_{Y_{\sigma,r}^2}$ and $\|R^0\|_{Y_{\sigma,r}^2}$.

This inequality implies that the transformation $R^1 \to \widetilde{R}^1$ is 1–1, hence invertible and satisfies the estimate

$$\|R^1\|_{Y_{\sigma,r}^2}^2 \ge \left(\frac{1}{1 - C_3 \varepsilon} \right) \left(\|\widetilde{R}^1\|_{Y_{\sigma,r}^2}^2 + \varepsilon C_4 \|R^0\|_{Y_{\sigma,r}^2}^2 \right) \tag{73}$$

so that we can write

$$R_j^1 = \widetilde{R}_j^1 + \varepsilon F(\widetilde{R}^1, R^0). \tag{74}$$

We now consider the transformation for R_j^0, which with the help of (73). We can write

$$\begin{aligned} \widetilde{R}_j^0 &= R_j^0 + \varepsilon B_j^{0,1}(\Psi, R^1) \\ &= R_j^0 + \varepsilon B_j^{0,1}(\Psi, \widetilde{R}^1) + \varepsilon^2 B_j^{0,1}(\Psi, F(\widetilde{R}^1, R^0)), \end{aligned} \tag{75}$$

or

$$R_j^0 = (\widetilde{R}_j^0 - \varepsilon B_j^{0,1}(\Psi, \widetilde{R}^1)) - \varepsilon^2 B^{0,1}(\Psi, F(\widetilde{R}^1, R^0)) \ . \tag{76}$$

Recall that $B^{0,1}$ is smoothing as we remarked in (47) and the extra power of ε insures that $\varepsilon^2 B^{0,1}(\Psi, F(\widetilde{R}^1, R^0))$ is also small. Thus (76) can be inverted by a Neumann series and we see that the normal-form transformation (70) is invertible and satisfies the estimates claimed in Proposition 13.

6 The error estimates

In order to solve (69), we use energy estimates in a scale of Banach spaces of analytic functions. By making the width σ of the strip of analyticity smaller as time evolves we can gain artificially some smoothing of the evolution. Since $\sigma = \mathcal{O}(1)$ in Lemma 5 and since we have to solve (69) on a time scale of order $\mathcal{O}(1/\varepsilon^2)$ the strip can be made smaller with a velocity of order $\mathcal{O}(\varepsilon^2)$. Hence define

$$\widehat{\mathcal{R}}(k,t) = \widehat{S}(k,t) \widehat{w}(k,t) = \widehat{w}(k,t) e^{-|k|(a - b\varepsilon^2 t)}$$

with constants $a, b > 0$ chosen below. If $w(t) \in L^2$, then $\mathcal{R}(t)$ is analytic in a strip of width $a - b\varepsilon^2 t$, i.e. $t \in [0, a/(b\varepsilon^2)]$. Computing the equation for w we find

$$\partial_t w = \Lambda w - |k| b\varepsilon^2 w + \varepsilon^2 \widetilde{\ell}(w) + \varepsilon^\beta \widetilde{G}(w) + \varepsilon^{-\beta}\vartheta^{-1}\widetilde{\mathrm{Res}}(\varepsilon\Psi), \tag{77}$$

where $\widetilde{\ell}(w) = S^{-1}(t)\ell(S(t)w)$, $\widetilde{G}(w) = S^{-1}(t)G(S(t)w)$, and $\widetilde{\mathrm{Res}}(\varepsilon\Psi) = S^{-1}(t)\mathrm{Res}(\varepsilon\Psi)$.

If we use the estimates on ℓ and G from Proposition 18, along with the fact that the support of $\mathrm{Res}(\varepsilon\Psi)$ is bounded in Fourier space, then we immediately obtain the following estimates for the terms in (77).

Corollary 19. *For any $r \geq 2$, there exist constants $\widetilde{C}_L$, $\widetilde{C}_G$ and $\widetilde{C}_R$ such that*

$$\begin{array}{rcl} \|\widetilde{\ell}(w)\|_{H^{r-1}} & \leq & \widetilde{C}_L \|w\|_{H^r}, \\ \|\widetilde{G}(w)\|_{H^{r-1}} & \leq & \widetilde{C}_G \|w\|_{H^r} \|w\|_{H^{r-1}}, \\ \|\varepsilon^{-\beta}\vartheta^{-1}\widetilde{\mathrm{Res}}(\varepsilon\Psi)\|_{H^r} & \leq & \widetilde{C}_R \varepsilon^2. \end{array}$$

We control the solutions of equation (77) using energy estimates and Gronwall's inequality. Fix some index $s \geq 6$ and define

$$\|w\|_{H^s}^2 = \|w\|_{L^2}^2 + \|w\|_{\mathring{H}^s}^2 \tag{78}$$

where

$$\|w\|_{\mathring{H}^s}^2 = \int |k|^{2s} |\widehat{w}(k)|^2 dk. \tag{79}$$

We have

$$\begin{aligned} \frac{1}{2}\partial_t \|w\|_{L^2}^2 = & -b\varepsilon^2 \int |k||\widehat{w}(k)|^2 dk + \varepsilon^2 \int |\widehat{w}(k)|\widehat{\widetilde{\ell}(w)}(k)| dk \\ & + \varepsilon^\beta \int |\widehat{w}(k)|\widehat{\widetilde{G}(w)}(k)| dk + \int |\widehat{w}(k)|\varepsilon^{-\beta}\widehat{\vartheta}^{-1}(k)\widehat{\widetilde{\mathrm{Res}}(w)}(k)| dk. \end{aligned} \tag{80}$$

Applying the Cauchy–Schwartz inequality and the estimates of Corollary 19, we find

$$\begin{array}{rcl} \frac{1}{2}\partial_t \|w\|_{L^2}^2 & \leq & -b\varepsilon^2 \|w\|_{\mathring{H}^{1/2}}^2 + \|w\|_{L^2}(\widetilde{C}_L \varepsilon^2 \|w\|_{H^3} + \widetilde{C}_G \varepsilon^\beta \|w\|_{H^2}\|w\|_{H^3} + \widetilde{C}_R \varepsilon^2) \\ & \leq & -b\varepsilon^2 \|w\|_{\mathring{H}^{1/2}}^2 + \varepsilon^2(\widetilde{C}_L + \widetilde{C}_R)\|w\|_{H^3}^2 + \widetilde{C}_G \varepsilon^\beta \|w\|_{H^3}^3 + \widetilde{C}_R \varepsilon^2. \end{array}$$

Now consider

$$\begin{aligned} \frac{1}{2}\partial_t \|w\|_{\mathring{H}^s}^2 = & -b\varepsilon^2 \int |k|^{2s+1}|\widehat{w}(k)|^2 dk + \varepsilon^2 \int |k|^{2s}|\widehat{w}(k)|\widehat{\widetilde{\ell}(w)}(k)| dk \\ & + \varepsilon^\beta \int |k|^{2s}|\widehat{w}(k)|\widehat{\widetilde{G}(w)}(k)| dk + \int |k|^{2s}|\widehat{w}(k)|\varepsilon^{-\beta}\widehat{\vartheta}^{-1}(k)\widehat{\widetilde{\mathrm{Res}}(w)}(k)| dk. \end{aligned} \tag{81}$$

If we once again apply the Cauchy–Schwartz inequality and the estimates in Corollary 19 we can bound the last three integrals in (81) by

$$\|w\|_{\mathring{H}^{s+1/2}}\left\{\widetilde{C}_L\varepsilon^2\|w\|_{H^{s+1/2}}+\widetilde{C}_G\varepsilon^{\beta}\|w\|_{H^{s-1/2}}\|w\|_{H^{s+1/2}}+\widetilde{C}_R\varepsilon^2\right\}. \tag{82}$$

Combining (81) and (82) gives

$$\frac{1}{2}\partial_t\|w\|^2_{\mathring{H}^s}\leq-\varepsilon^2(b-(\widetilde{C}_L+\widetilde{C}_R)-\widetilde{C}_G\varepsilon^{\beta-2}\|w\|_{H^{s-1/2}})\|w\|^2_{H^{s+1/2}}+\widetilde{C}_R\varepsilon^2. \tag{83}$$

Combining this with the estimate on the L^2 norm of w and using $\|w\|_{H^3}\leq 2\|w\|_{H^r}$ for all $r\geq 3$ we obtain the inequality

$$\frac{1}{2}\partial_t\|w\|^2_{H^s}\leq-\varepsilon^2(b-3(\widetilde{C}_L+\widetilde{C}_R)-3\widetilde{C}_G\varepsilon^{\beta-2}\|w\|_{H^{s-1/2}})\|w\|^2_{H^{s+1/2}}+2\widetilde{C}_R\varepsilon^2. \tag{84}$$

Applying Gronwall's inequality to (84) we obtain:

Proposition 20. *If* $b-3(\widetilde{C}_L+\widetilde{C}_R)-3\widetilde{C}_G\varepsilon^{\beta-2}\sup_{0\leq t\leq t_0}\|w(t)\|_{H^{s-1/2}}\geq 0$, *then*

$$\sup_{0\leq t\leq t_0}\|w(t)\|^2_{H^s}\leq(\|w(0)\|^2_{H^s}+2\widetilde{C}_R\varepsilon^2 t_0).$$

Take $t_0=\varepsilon^{-2}\widetilde{T}_0$ and $\|w(0)\|^2_{H^s}\leq 2\widetilde{C}_R\widetilde{T}_0$. Then choose b such that $b-3(\widetilde{C}_L+\widetilde{C}_R)-24\widetilde{C}_G\widetilde{C}_R\widetilde{T}_0\varepsilon^{\beta-2}\geq 0$. The Proposition 20 implies

Corollary 21. *For all* $0\leq\varepsilon^2 t\leq\widetilde{T}_0$,

$$\|w(t)\|^2_{H^s}\leq 4\widetilde{C}_R\widetilde{T}_0.$$

Finally we must check that the smoothing operator $S(t)$ is well defined. We require that the constants a and b in its definition be such that $\sigma>a$ and $a-b\varepsilon^2 t>a/2$ for all $0\leq\varepsilon^2 t\leq\widehat{T}_0$. In this case $S(t)$ is well defined. (Note that this means in particular that $\widehat{T}_0<\sigma/(2b)$.) Finally, we have

Corollary 22. *Choose* $T_1=\min(\widetilde{T}_0,\widehat{T}_0)$. *Then*

$$\begin{aligned}\sup_{0\leq\varepsilon^2 t\leq T_1}\|\mathcal{R}(t)\|^2_{Y^2_{a/2,s}}&\leq\sup_{0\leq\varepsilon^2 t\leq T_1}\|\mathcal{R}(t)\|^2_{Y^2_{a-b\varepsilon^2 t,s}}=\sup_{0\leq\varepsilon^2 t\leq\widetilde{T}_0}\|\mathcal{R}(t)\|^2_{Y^2_{a-b\varepsilon^2 t,s}}\\&=\sup_{0\leq\varepsilon^2 t\leq\widetilde{T}_0}\|S(t)w(t)\|^2_{Y^2_{a-b\varepsilon^2 t,s}}=\sup_{0\leq\varepsilon^2 t\leq\widetilde{T}_0}\|w(t)\|^2_{H^s}\leq 4\widetilde{C}_R\widetilde{T}_0.\end{aligned} \tag{85}$$

Since the $Y^2_{a/2,s}$ norm controls any Sobolev norm of $\mathcal{R}(t)$, we obtain

Corollary 23. *Choose* $T_1=\min(\widetilde{T}_0,\widehat{T}_0)$. *Then*

$$\sup_{0\leq\varepsilon^2 t\leq T_1}\|\mathcal{R}(t)\|^2_{H^s}\leq 4\widetilde{C}_R\widetilde{T}_0.$$

Combining this estimate with Proposition 13, Proposition 17, and Lemma 5 completes the proof of Theorem 1.

Remark 24. *The local existence and uniqueness of solutions combined with our error estimates gives the long time existence of solutions for free.*

7 Appendix

In this appendix we explain in more detail the higher order approximation introduced in Section 3. As noted there we use an approximation of the form:

$$\varepsilon\widetilde{\psi}_j = \sum_{j_2;|j_2|<5} \sum_{j_1;\beta_j(j_2,j_1)\leq 5} \varepsilon^{\beta_j(j_2,j_1)}\widetilde{\psi}^{j_1}_{j_2 j} \ , \tag{86}$$

where $\beta_1(j_2,j_1) = 1 + ||j_2|-1| + j_1$ and $\beta_1(j_2,j_1) = \beta_2(j_2,j_1)$ except for $\beta_2(1,j_1) = \beta_1(1,j_1)+2$. We show below that the terms proportional to ε (i.e. $\widetilde{\psi}^0_{\pm 1}$) are given by solutions of the NLS equation, while the higher order terms are defined by algebraic relations or inhomogeneous, linear PDE's.

Recall as well that the term $\widetilde{\psi}^{j_1}_{j_2}$ is assumed to have the form:

$$\widetilde{\psi}^{j_1}_{j_2 j} = A^{j_1}_{j_2 j}(\varepsilon(x+c_g t),\varepsilon^2 t)e^{ij_2(k_0 x+\omega_0 t)} \ . \tag{87}$$

From this we see that the index j_2 determines what multiple of the basic frequency and wave number a given term represents, j_1 represents what order we are studying in the approximation (i.e., $j_1 = 0$, corresponds to the lowest order approximation of a given wave number/frequency, $j_1 = 1$ to the next order approximation for that wave number, and so on) and j just labels the first or second component of these two component vectors.

The form assumed in (87) is important because it makes it easy to approximate the action of the Fourier multiplier operator with symbol $\hat{\omega}(k)$ on such functions with the aid of:

Lemma 25. *Let ω be a Fourier multiplier operator with symbol $\hat{\omega}(k)$. Assume that $\widetilde{\psi}^{j_1}_{j_2 j}$ has the form given in* (87)*. Then*

$$\begin{aligned}(\omega\widetilde{\psi}^{j_1}_{j_2 j})(x,t) = \Big\{&\hat{\omega}(j_2k_0)A^{j_1}_{j_2 j}(\varepsilon(x+c_g t),\varepsilon^2 t) - i\varepsilon(\partial_k\hat{\omega}(j_2k_0)\partial_X A^{j_1}_{j_2 j}(\varepsilon(x+c_g t),\varepsilon^2 t)) \\ &-\frac{1}{2}\varepsilon^2(\partial_k^2\hat{\omega}(j_2k_0)\partial_X^2 A^{j_1}_{j_2 j}(\varepsilon(x+c_g t),\varepsilon^2 t)) + \dots\Big\}e^{ij_2(k_0x+\omega_0 t)} \ .\end{aligned}$$

Proof. This results by computing of the inverse Fourier transform of $(\widehat{\hat{\omega}\widetilde{\psi}^{j_1}_{j_2 j}})(k)$ in which one expands the symbol $\omega(k)$ in a Taylor series around j_2k_0. Note that it is straightforward to estimate rigorously the error terms in this expansion but we won't need those estimates

- we estimate directly the difference between our approximation (which is based on this expansion) and the true solution of our equation in Section 6. $\square$

We now insert (87) into equation (10). We will focus on the case where the NLS equation occurs for the first component of $\widetilde{\psi}_{j_2}^{j_1}$ - namely $(\widetilde{\psi}_{j_2}^{j_1})_1$.

Substituting (87) into (10), the j-th component of the LHS of the equation yields:

$$\sum_{j_2;|j_2|<5}\ \sum_{j_1;\beta_j(j_2,j_1)\leq 5} \varepsilon^{\beta_j(j_2,j_1)}\Big\{ij_2\omega_0 A_{j_2 j}^{j_1} + \varepsilon c_g \partial_X A_{j_2 j}^{j_1} + \varepsilon^2 \partial_T A_{j_2 j}^{j_1}\Big\} e^{ij_2(k_0 x+\omega_0 t)}\ , \tag{88}$$

where we have suppressed the arguments of $A_{j_2 j}^{j_1}$ to save space.

Making the same insertion of the RHS, we will then equate terms proportional to the same power of ε and the same factor of $\exp(ij_2(k_0x+\omega_0 t)$ to determine the amplitude functions $A_{j_2 j}^{j_1}$.

The linear terms on the right hand side are easy to treat. Focussing on the first component and utilizing the expansion from just above we have

$$\begin{aligned} i(\omega_1\varepsilon\tilde{\psi}_1) \quad = \quad & \sum_{j_2;|j_2|<5}\ \sum_{j_1;\beta_1(j_2,j_1)\leq 5} \varepsilon^{\beta_1(j_2,j_1)}\Big\{i\hat{\omega}(j_2k_0)A_{j_2 1}^{j_1} \\ & +\varepsilon(\partial_k\hat{\omega})(j_2k_0)\partial_X A_{j_2 1}^{j_1} - i\frac{\varepsilon^2}{2}(\partial_k^2\hat{\omega})(j_2k_0)\partial_X^2 A_{j_2 1}^{j_1} + \dots\Big\} e^{ij_2(k_0x+\omega_0 t)}\ . \end{aligned} \tag{89}$$

(The expression for $i(\omega_2\varepsilon\tilde{\psi}_2)$ would just be the negative of this.)

Finally, we consider the nonlinear terms. Recall from Section 3 that the first component of the nonlinear term is of the form:

$$\sum_{m,n\in\{1,2\}} \hat{\alpha}_{m,n}^1(k)\hat{U}_m(k-\ell)\hat{U}_n(\ell)d\ell\ . \tag{90}$$

The exact form of $\hat{\alpha}_{m,n}^1$ is discussed in Section 3, but the important thing for the discussion here is that it is proportional to $\hat{\omega}_1(k)$. If we insert our approximation (87) into the nonlinear term we obtain a finite sum of terms of the form:

$$\varepsilon^{\beta_j(j_2,j_1)}\varepsilon^{\beta_{\tilde{j}}(\tilde{j}_2,\tilde{j}_1)}\int \hat{\alpha}_{j,\tilde{j}}^1(k)\tilde{\psi}_{j_2 j}^{j_1}(k-\ell)\tilde{\psi}_{\tilde{j}_2\tilde{j}}^{\tilde{j}_1}(\ell)d\ell \tag{91}$$

in Fourier space. Recall that in Fourier space

$$\tilde{\psi}_{j_2 j}^{j_1}(k) = \frac{1}{\varepsilon}\hat{A}(\frac{k-j_2k_0}{\varepsilon})e^{ij_2\omega_0 t}e^{ic_g(k-j_2k_0)t} \tag{92}$$

Inserting this expression into (91) and integrating we find (after inverting the Fourier transform) that this term has the form:

$$\begin{aligned} & \varepsilon^{\beta_j(j_2,j_1)+\beta_{\tilde{j}}(\tilde{j}_2,\tilde{j}_1)}\Big\{\hat{\alpha}_{j\tilde{j}}^1((j_2+\tilde{j}_2)k_0)A_{j_2 j}^{j_1}A_{\tilde{j}_2\tilde{j}}^{\tilde{j}_1} \\ & \quad -i\varepsilon(\partial_k\hat{\alpha}_{j\tilde{j}}^1((j_2+\tilde{j}_2)k_0))\partial_X\left(A_{j_2 j}^{j_1}A_{\tilde{j}_2\tilde{j}}^{\tilde{j}_1}\right) - \frac{\varepsilon^2}{2}(\partial_k^2\hat{\alpha}_{j\tilde{j}}^1((j_2+\tilde{j}_2)k_0))\partial_X^2\left(A_{j_2 j}^{j_1}A_{\tilde{j}_2\tilde{j}}^{\tilde{j}_1}\right) \\ & \qquad +\dots\Big\} e^{(j_2+\tilde{j}_2)(k_0x+\omega_0 t)}e^{ic_g(k-(j_2+\tilde{j}_2)k_0)t}\ . \end{aligned} \tag{93}$$

We now consider terms of various orders that appear on both side of the equation. The only terms of $\mathcal{O}(\varepsilon)$ arise from the case $j_1 = 0, j_2 = \pm 1$. They occur in the linear terms on both sides of the equation and their coefficients are identical so they cancel automatically and impose no constraints of the value of the amplitude functions $A^{j_1}_{j_2 j}$.

Next turn to the terms of $\mathcal{O}(\varepsilon^2)$. In addition to the contributions coming from the linear terms, the nonlinearity will generate contributions proportional to products of the leading order, $\mathcal{O}(\varepsilon)$ terms. Thus, we expect to have to consider terms proportional to $\exp(ij_2(k_0 x + \omega_0 t))$ for $j_2 = 0, \pm 2$, as well as the contributions coming from $j_2 = \pm 1$. The terms with j_2 negative are just complex conjugates of those with j_2 positive, so we will focus on the non-negative values of j_2.

Terms proportional to $\varepsilon^2 e^{ij_2(k_0 x + \omega_0 t)}$ with:

- $j_2 = 0$
$$\left(-\hat{\omega}_j(0) A^0_{0j} - 2\hat{\alpha}^j_{11}(0) A^0_{11} A^0_{-11}\right) \tag{94}$$

- $j_2 = 1$
$$\left(i\omega_0 A^1_{11} - i\hat{\omega}(k_0) A^1_{11} + c_g \partial_X A^0_{11} - (\partial_k \hat{\omega})(k_0) \partial_X A^0_{11}\right) \tag{95}$$

- $j_2 = 2$
$$\left(2i\omega_0 A^0_{2j} - i\hat{\omega}_j(2k_0) A^0_{2j} - \hat{\alpha}^j_{11}(2k_0) A^0_{11} A^0_{11}\right) \tag{96}$$

We want each of these expressions to equal zero. The expression in (94) vanishes automatically since both $\hat{\omega}_j(0) = 0$ and $\hat{\alpha}^j_{11}(0) = 0$. Thus, this imposes no conditions on A^0_{0j}. The expression in (95) also vanishes automatically since $\omega_0 = \hat{\omega}(k_0)$ $c_g = (\partial_k \hat{\omega})(k_0)$ by definition.

Finally, we can insure that (96) vanishes if we choose

$$A^0_{2j} = \frac{i\hat{\alpha}^j_{11}(2k_0)}{\hat{\omega}_j(2k_0) - 2\omega_0} A^0_{11} A^0_{11} \ . \tag{97}$$

where A^0_{11} will be fixed at the next order. Note that from the expression for $\hat{\omega}_j(k)$, the denominator $\hat{\omega}_j(2k_0) - 2\omega_0$ is non-zero such that this expression is well-defined.

Next we turn to the terms $\mathcal{O}(\varepsilon^3)$. We proceed exactly as in the case of the terms above, considering each of the values of j_2 generated by the nonlinearity in turn.

Terms proportional to $\varepsilon^3 e^{ij_2(k_0 x + \omega_0 t)}$ with:

- $j_2 = 0$
$$\left(c_g \partial_X A^0_{0j} - (\partial_k \hat{\omega}_j)(0) \partial_X A^0_{0j} + 2i(\partial_k \hat{\alpha}^j_{11})(0) \partial_X \left(A^0_{11} A^0_{-11}\right)\right) \tag{98}$$

- $j_2 = 1$

$$\left(\partial_T A^0_{11} + \frac{i}{2}(\partial_k^2 \hat{\omega})(k_0)\partial_X^2 A^0_{11} - 2\sum_j \hat{\alpha}^1_{1j}(k_0)(A^0_{11}A^0_{0j} + A^0_{-11}A^0_{2j}) \right) \tag{99}$$

and

$$\left(i(\omega_0 + \hat{\omega}(k_0))A^0_{12} - 2\sum_j \hat{\alpha}^2_{1j}(k_0)(A^0_{11}A^0_{0j} + A^0_{-11}A^0_{2j}) \right) \tag{100}$$

- $j_2 = 2$

$$\Big(i(2\omega_0 - \hat{\omega}_j(2k_0))A^1_{2j} + i(\partial_k \hat{\alpha}^j_{11})(2k_0)\partial_X \left(A^0_{11}A^0_{11} \right) - 2\hat{\alpha}^j_{11}(2k_0)A^0_{11}A^1_{11} \Big) \tag{101}$$

- $j_2 = 3$

$$\left(i(3\omega_0 - \hat{\omega}_j(3k_0))A^0_{3j} - 2\sum_{\tilde{j}} \hat{\alpha}^j_{1\tilde{j}}(3k_0)A^0_{11}A^0_{2\tilde{j}} \right) \tag{102}$$

Remark 26. *Note that we have omitted from these expressions terms like* $\omega_0 A^2_{1j} - \hat{\omega}(k_0)A^2_{1j}$ *which vanish regardless of the choice of* A^2_{1j}*.*

We can make each of these expressions vanish by the following choices: Equating (100) to zero determines A^0_{12} since $\omega_0 + \hat{\omega}(k_0) \neq 0$, equating (101) to zero determines A^1_{2j} since $2\omega_0 - \hat{\omega}_j(2k_0) \neq 0$ where A^1_{11} will be fixed at the next order, and finally equating (102) to zero determines A^0_{3j} since $3\omega_0 - \hat{\omega}_j(3k_0) \neq 0$, Next, set

$$A^0_{0j} = \left(-\frac{2i\partial_k \hat{\alpha}^j_{11}(0)}{c_g - \partial_k \hat{\omega}_j(0)} \right) A^0_{11}A^0_{-11} \ . \tag{103}$$

Inserting (103) and (97) into (99) for $j = 1$ gives

$$\partial_T A^0_{11} = -\frac{i}{2}(\partial_k^2 \hat{\omega})(k_0)\partial_X^2 A^0_{11} - i\nu |A^0_{11}|^2 A^0_{11} \ , \tag{104}$$

where

$$\nu = 2\sum_j \hat{\alpha}^1_{1j}(k_0) \left[\frac{\hat{\alpha}^j_{11}(2k_0)}{\hat{\omega}_j(2k_0) - 2\omega_0} + \frac{2\partial_k \hat{\alpha}^j_{11}(0)}{c_g - \partial_k \hat{\omega}_j(0)} \right] \ , \tag{105}$$

so we see that, as claimed, A^0_{11} is given by a solution of the Nonlinear Schrödinger equation.

Finally, we consider the terms of $\mathcal{O}(\varepsilon^4)$ – the terms of $\mathcal{O}(\varepsilon^5)$ are handled in an entirely analogous fashion. We proceed exactly as before by writing out the terms proportional to $\varepsilon^4 \exp(ij_2(k_0x + c_gt))$, though we must consider more choices of j_2 to account for the additional terms generated by the nonlinearity. We can make each of these new expressions vanish with choices much like those made above - only the cases $j_2 = 0$ and $j_2 = 1$ really need additional comment.

Terms proportional to $\varepsilon^4 e^{ij_2(k_0x+\omega_0t)}$ with:

- $j_2 = 0$
$$\Big(c_g\partial_X A^1_{0j} - \partial_k\hat{\omega}_j(0)\partial_X A^1_{0j} + \partial_T A^0_{0j} + 2i(\partial_k\hat{\alpha}^j_{11})\partial_X \left(A^1_{11}A^0_{-11} + A^0_{11}A^1_{-11}\right)\Big) \tag{106}$$
where the expression simplified due to $\partial_k^2\hat{\omega}_j(0) = \partial_k^2\hat{\alpha}^j_{11}(0) = 0$.

- $j_2 = 1$
$$\begin{aligned}\big\{\partial_T A^1_{11} + \frac{i}{2}(\partial_k^2\hat{\omega}(k_0))\partial_X^2 A^1_{11} + \frac{1}{3!}(\partial_k^3\omega(k_0))\partial_X^3 A^0_{11} \\ -2\sum_j(\hat{\alpha}^1_{1j}(k_0)(A^1_{11}A^0_{0j} + A^1_{-11}A^0_{2j} + A^0_{11}A^1_{0j} + A^0_{-11}A^1_{2j}) \\ +2i(\partial_k\hat{\alpha}^1_{1j})\partial_X(A^0_{11}A^0_{0j} + A^0_{-11}A^0_{2j}))\big\}\end{aligned} \tag{107}$$

Equating (107) to zero completes the definition of the terms above and defines A^1_{11} to be the solution of the linear, but inhomogeneous, Schrödinger equation, where we note that all the inhomogeneous terms of this equation have been defined at prior steps in this process.

Finally, we address A^1_{0j}. Equating (106) to zero determines A^1_{0j} since $c_g - \partial_k\hat{\omega}(k_0) \neq 0$. All other quantities in (106) have been defined at previous steps in the interative procedure. The resulting equation requires some integration w.r.t. X which would give an A^1_{0j} which in general no longer belongs to L^2. In order to avoid this integration we have to show that the only problematic term $\partial_T A^0_{0j}$ can be written as a derivative w.r.t X in order to insure that A^1_{01} has the appropriate decay properties as $X \to \pm\infty$.

Recall that from (103)
$$\partial_T A^0_{0j} = \left(-\frac{2i\partial_k\hat{\alpha}^j_{11}(0)}{c_g - \partial_k\hat{\omega}_j(0)}\right)((\partial_T A^0_{11})A^0_{-11} + A^0_{11}(\partial_T A^0_{-11}))\ . \tag{108}$$

We know that $A^0_{\pm 11}$ both satisfy the nonlinear Schrödinger equation,
$$\partial_T A^0_{\pm 11} = \mp\frac{i}{2}(\partial_k^2\hat{\omega}(k_0))\partial_X^2 A^0_{\pm 11} \mp i\nu|A^0_{\pm 11}|^2 A^0_{\pm 11}. \tag{109}$$

Inserting this expressions into (108) and recalling that $A^0_{-11} = \overline{A^0_{11}}$, we find that the nonlinear terms cancel and we are left with
$$\partial_T A^0_{0j} = \left(-\frac{\partial_k\hat{\alpha}^j_{11}(0)}{c_g - \partial_k\hat{\omega}_j(0)}\right)(\partial_k^2\hat{\omega}(k_0))\partial_X((\partial_X A^0_{11})A^0_{-11} - A^0_{11}(\partial_X A^0_{-11}))\ . \tag{110}$$

Inserting this expression into (106), we see that we can choose A_{01}^1 to be

$$\begin{aligned} A_{0j}^1 &= (c_g - \partial_k \hat{\omega}_j(0))^{-1} \Big\{ 2i(\partial_k \hat{\alpha}_{11}^j)(A_{11}^1 A_{-11}^0 + A_{11}^0 A_{-11}^1) \\ &\qquad - (\frac{\partial_k \hat{\alpha}_{11}^j(0)}{c_g - \partial_k \hat{\omega}_j(0)})(\partial_k^2 \hat{\omega}(k_0))((\partial_X A_{11}^0) A_{-11}^0 - A_{11}^0 (\partial_X A_{-11}^0)) \Big\} \,. \end{aligned} \tag{111}$$

As we remarked above, choosing the terms of $\mathcal{O}(\varepsilon^5)$ in the expansion is handled in a very similar fashion and we leave those calculations as an exercise. In order to close the system at $\mathcal{O}(\varepsilon^5)$ we have to compute the linear inhomogeneous Schrödinger equation for A_{11}^3 at $\mathcal{O}(\varepsilon^6)$, too.

References

[Alt02] Alterman, Deborah and Rauch, Jeffrey. Nonlinear geometric optics for short pulses. *J. Differ. Equations*, 178(2):437–465, 2002.

[BL03] Karen Barrailh and David Lannes. A general framework for diffractive optics and its applications to lasers with large spectrums and short pulses. *SIAM J. Math. Anal.*, 34(3):636–674, 2003.

[CSS92] W. Craig, C. Sulem, and P.-L. Sulem. Nonlinear modulation of gravity waves: a rigorous approach. *Nonlinearity*, 5(2):497–522, 1992.

[GMS09] P. Germain, N. Masmoudi, and J. Shatah. Preprint: arXiv:0906.5343v1, 2009.

[GS01] Thierry Gallay and Guido Schneider. KP description of unidirectional long waves. The model case. *Proc. Roy. Soc. Edinburgh Sect. A*, 131(4):885–898, 2001.

[JMR00] Jean-Luc Joly, Guy Metivier, and Jeffrey Rauch. Transparent nonlinear geometric optics and Maxwell-Bloch equations. *J. Differ. Equations*, 166:175–250, 2000.

[Kal87] L. A. Kalyakin. Asymptotic decay of a one-dimensional wave packet in a nonlinear dispersive medium. *Mat. Sb. (N.S.)*, 132(174)(4):470–495, 592, 1987.

[McK97] Henry P. McKean. How real is resonance? *Comm. Pure Appl. Math.*, 50(4):317–322, 1997.

[Sch95] Guido Schneider. Validity and limitation of the Newell-Whitehead equation. *Math. Nachr.*, 176:249–263, 1995.

[Sch98a] Guido Schneider. Approximation of the Korteweg-de Vries equation by the nonlinear Schrödinger equation. *J. Differential Equations*, 147(2):333–354, 1998.

[Sch98b] Guido Schneider. Justification of modulation equations for hyperbolic systems via normal forms. *NoDEA Nonlinear Differential Equations Appl.*, 5(1):69–82, 1998.

[Sch05] Guido Schneider. Justification and failure of the nonlinear Schrödinger equation in case of non-trivial quadratic resonances. *J. Differential Equations*, 216(2):354–386, 2005.

[TW11] Nathan Totz and Sijue Wu. Preprint: arXiv:1101.0545v1, 2011.

[Wu09] Sijue Wu. Almost global wellposedness of the 2-D full water wave problem. *Invent. Math.*, 177(1):45–135, 2009.

[Zak68] V.E. Zakharov. Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Sov. Phys. J. Appl. Mech. Tech. Phys.*, 4:190–194, 1968.