

# KAM for autonomous quasi-linear perturbations of mKdV

Pietro Baldi · Massimiliano Berti · Riccardo  
Montalto

Received: date / Accepted: date

**Abstract** We prove the existence of Cantor families of small amplitude, linearly stable, quasi-periodic solutions of *quasi-linear* (also called strongly nonlinear) autonomous Hamiltonian differentiable perturbations of the mKdV equation. The proof is based on a weak version of the Birkhoff normal form algorithm and a nonlinear Nash-Moser iteration. The analysis of the linearized operators at each step of the iteration is achieved by pseudo-differential operator techniques and a linear KAM reducibility scheme.

**Keywords** mKdV · KAM for PDEs · quasi-linear PDEs · Nash-Moser theory · quasi-periodic solutions

**Mathematics Subject Classification (2000)** 37K55 · 35Q53

## 1 Introduction and main result

In the paper [5] we proved the first existence result of quasi-periodic solutions for autonomous quasi-linear PDEs (also called “*strongly nonlinear*” in [25]), in particular of small amplitude quasi-periodic solutions of the KdV equation subject to a Hamiltonian quasi-linear perturbation. The approach developed in [5] (see also [4]) is of wide applicability for quasi-linear PDEs in 1 space dimension. In this paper we take the opportunity to explain the general strategy of [5] applied to a model which is slightly simpler than KdV.

---

Pietro Baldi  
Dipartimento di Matematica e Applicazioni “R. Caccioppoli”, Università di Napoli Federico II, Via Cintia, Monte S. Angelo, 80126 Napoli, Italy  
E-mail: [pietro.baldi@unina.it](mailto:pietro.baldi@unina.it)

Massimiliano Berti  
SISSA, Via Bonomea 265, 34136 Trieste, Italy  
E-mail: [berti@sissa.it](mailto:berti@sissa.it)

Riccardo Montalto  
Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich  
E-mail: [riccardo.montalto@math.uzh.ch](mailto:riccardo.montalto@math.uzh.ch)

We consider the cubic, focusing or defocusing, mKdV equation

$$u_t + u_{xxx} + \varsigma \partial_x(u^3) + \mathcal{N}_4(x, u, u_x, u_{xx}, u_{xxx}) = 0, \quad \varsigma = \pm 1, \quad (1.1)$$

under periodic boundary conditions  $x \in \mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$ , where

$$\mathcal{N}_4(x, u, u_x, u_{xx}, u_{xxx}) := -\partial_x [(\partial_u f)(x, u, u_x) - \partial_x((\partial_{u_x} f)(x, u, u_x))] \quad (1.2)$$

is the most general quasi-linear Hamiltonian (local) nonlinearity. Note that  $\mathcal{N}_4$  contains as many derivatives as the linear vector field  $\partial_{xxx}$ . It is a *quasi-linear* perturbation because  $\mathcal{N}_4$  depends linearly on the highest derivative  $u_{xxx}$  multiplied by a coefficient which is a nonlinear function of the lower order derivatives  $u, u_x, u_{xx}$ . The equation (1.1) is the Hamiltonian PDE

$$u_t = X_H(u), \quad X_H(u) := \partial_x \nabla H(u), \quad (1.3)$$

where  $\nabla H$  denotes the  $L^2(\mathbb{T}_x)$  gradient of the Hamiltonian

$$H(u) = \frac{1}{2} \int_{\mathbb{T}} u_x^2 dx - \frac{\varsigma}{4} \int_{\mathbb{T}} u^4 dx + \int_{\mathbb{T}} f(x, u, u_x) dx \quad (1.4)$$

on the real phase space

$$H_0^1(\mathbb{T}_x) := \left\{ u(x) \in H^1(\mathbb{T}, \mathbb{R}) : \int_{\mathbb{T}} u(x) dx = 0 \right\} \quad (1.5)$$

endowed with the non-degenerate symplectic form

$$\Omega(u, v) := \int_{\mathbb{T}} (\partial_x^{-1} u) v dx, \quad \forall u, v \in H_0^1(\mathbb{T}_x), \quad (1.6)$$

where  $\partial_x^{-1} u$  is the periodic primitive of  $u$  with zero average. The phase space  $H_0^1(\mathbb{T}_x)$  is invariant for the evolution of (1.1) because the integral  $\int_{\mathbb{T}} u(x) dx$  is a prime integral (the mass). For simplicity we fix its value to  $\int_{\mathbb{T}} u(x) dx = 0$ . We recall that the Poisson bracket between two functions  $F, G : H_0^1(\mathbb{T}_x) \rightarrow \mathbb{R}$  is defined as

$$\{F, G\}(u) := \Omega(X_F(u), X_G(u)) = \int_{\mathbb{T}} \nabla F(u) \partial_x \nabla G(u) dx. \quad (1.7)$$

We assume that the ‘‘Hamiltonian density’’  $f$  is of class  $C^q(\mathbb{T} \times \mathbb{R} \times \mathbb{R}; \mathbb{R})$  for some  $q$  large enough (otherwise, as it is well known, we cannot expect the existence of smooth invariant KAM tori). We also assume that  $f$  vanishes at  $u = u_x = 0$  and

$$|f(x, u, v)| \leq C(|u| + |v|)^5 \quad \forall (u, v) \in \mathbb{R}^2, \quad |u| + |v| \leq 1. \quad (1.8)$$

As a consequence the nonlinearity  $\mathcal{N}_4$  vanishes at order 4 at  $u = 0$  and (1.1) may be seen, close to the origin, as a ‘‘small’’ perturbation of the cubic mKdV equation

$$u_t + u_{xxx} + 3\varsigma u^2 u_x = 0. \quad (1.9)$$

Such equation is known to be completely integrable. Actually it is mapped into KdV by a Miura transform, and it may be described by global analytic action-angle variables, as it was proved by Kappeler-Topalov [21]. We also remark that, among the generalized KdV equations  $u_t + u_{xxx} \pm \partial_x(u^p) = 0$ ,  $p \in \mathbb{N}$ , the only known completely integrable ones are the KdV  $p = 2$  and the cubic mKdV  $p = 3$ .

It is a natural question to know whether the periodic, quasi-periodic or almost periodic solutions of (1.9) persist under small perturbations. This is the content of KAM theory. It is a difficult problem because of small divisors resonance phenomena, which are especially strong in presence of quasi-linear perturbations like  $\mathcal{N}_4$ .

In this paper (as well as in [5]) we restrict the analysis to the search of small amplitude quasi-periodic solutions. It is also a very interesting question to investigate possible extensions of this result to perturbations of finite gap solutions. A difficulty which arises in the search of small amplitude solutions is that the mKdV equation (1.1) is a *completely resonant* PDE at  $u = 0$ , namely the linearized equation at the origin is the linear Airy equation

$$u_t + u_{xxx} = 0$$

which possesses only the  $2\pi$ -periodic in time, real solutions

$$u(t, x) = \sum_{j \in \mathbb{Z} \setminus \{0\}} u_j e^{ij^3 t} e^{ijx}, \quad u_{-j} = \bar{u}_j. \quad (1.10)$$

Thus the existence of small amplitude quasi-periodic solutions of (1.1) is entirely due to the nonlinearity. Indeed, the nonlinear term  $\varsigma \partial_x(u^3)$  is the one that produces the main modulation of the frequency vector of the solution with respect to its amplitude (the well-known frequency-to-action map, or frequency-amplitude relation, or “twist”, see (4.10)) and that allows to “tune” the action parameters  $\xi$  so that the frequencies becomes rationally independent and diophantine. Note that the mKdV equation (1.1) does not depend on other external parameters which may influence the frequencies. This is a further difficulty in the study of autonomous PDEs with respect to the forced cases studied in [3]. Actually, in [3] we considered non-autonomous quasi-linear (and fully nonlinear) perturbations of the Airy equation and we used the forcing frequencies as independent parameters.

The core of the matter is to understand the perturbative effect of the quasi-linear term  $\mathcal{N}_4$  over *infinite times*. By (1.8), close to the origin, the quartic term  $\mathcal{N}_4$  is smaller than the pure cubic mKdV (1.9). Therefore, when we restrict the equation to finitely many space-Fourier indices  $|j| \leq C$ , we essentially enter in the range of applicability of finite dimensional KAM theory close to an elliptic equilibrium. The new problem is to understand what happens to the dynamics on the high frequencies  $|j| \rightarrow +\infty$ , since  $\mathcal{N}_4$  is a nonlinear differential operator of the *same* order (i.e. 3) as the constant coefficient linear (and integrable) vector field  $\partial_{xxx}$ .

Does such a strongly nonlinear perturbation give rise to the formation of singularities for a solution in finite time, as it happens for the quasi-linear wave equations considered by Lax [18] and Klainerman-Majda [22]? Or, on the contrary, does the KAM phenomenon persist nevertheless for the mKdV equation (1.1)? The answer to these questions has been controversial for several years. For example, Kappeler-Pöschel [20] (Remark 3, page 19) wrote: “It would be interesting to obtain perturbation results which also include terms of higher order, at least in the region where the KdV approximation is valid. However, results of this type are still out of reach, if true at all”.

We think that these are very important dynamical questions to be investigated, especially because many of the equations arising in Physics are quasi-linear or even fully nonlinear.

The main result of this paper proves that the KAM phenomenon actually persists, at least close to the origin, for quasi-linear Hamiltonian perturbations of mKdV (the same result is proved in [5] for KdV). More precisely, Theorem 1 proves the existence of Cantor families of small amplitude, linearly stable, quasi-periodic solutions of the mKdV equation (1.1) subject to quasi-linear Hamiltonian perturbations. It is not surprising that the same result applies for both the focusing and the defocusing mKdV because we are looking for small amplitude solutions. Thus the different sign  $\varsigma = \pm 1$  only affects the branch of the bifurcation.

From a dynamical point of view, note that the parameters  $\xi$  selected by the KAM Theorem 1 give rise to solutions of (1.1)-(1.2) which are global in time. This is interesting information because, as far as we know, there are no results of global or even local solutions of the Cauchy problem for (1.1)-(1.2), and such PDEs are in general believed to be ill-posed in Sobolev spaces (for a rough result of local well-posedness for (1.1)-(1.2) see [6]).

The iterative procedure we are going to present is able to select many parameters  $\xi$  which give rise to quasi-periodic solutions (hence defined for all times). This procedure works for parameters belonging to a finite dimensional Cantor like set which becomes asymptotically dense at the origin.

How can this kind of result be achieved? The proof of Theorem 1 – which we shall discuss in more detail later – is based on an iterative Nash-Moser scheme. As it is well known, the main step of this procedure is to invert the linearized operators obtained at each step of the iteration and to prove that the inverse operators, albeit they lose derivatives (because of small divisors), satisfy tame estimates in high Sobolev norms. The linearized equations are non-autonomous linear PDEs which depend quasi-periodically on time. The key point of this paper (and [5]) is that, using the symplectic decoupling of [10], some techniques of pseudo-differential operators adapted to the symplectic structure, and a linear Birkhoff normal form analysis, we are able to construct, for most diophantine frequencies, a time dependent (quasi-periodic) change of variables which conjugates each linearized equation into another one that is diagonal and has constant coefficients, that is, in “normal form”. This means that, in the new coordinates, we have *integrated the equations*. Then we easily invert the linearized operator (recall that the inverse loses derivatives because of small divisors) and we conjugate it back to solve the linear equation in the original set of variables. We remark that these quasi-periodic Floquet changes of variable map Sobolev spaces of arbitrarily high norms into itself and satisfy tame estimates. Hence the inverse operator also loses derivatives, but it satisfies tame estimates as well.

In the dynamical systems literature, this strategy is called “reducibility” of the equation and it is a quasi-periodic KAM perturbative extension of Floquet theory (Floquet theory deals with periodic solutions of finite dimensional systems). The difficulty to make it work in the present setting is due to the quasi-linear character of the nonlinearity in (1.1).

Before stating precisely our main result we shortly present some related literature. In the last years a big interest has been devoted to understand the effect of derivatives in the nonlinearity in KAM theory. For *unbounded* perturbations the

first KAM results have been proved by Kuksin [24] and Kappeler-Pöschel [20] for KdV (see also Bourgain [13]), and more recently by Liu-Yuan [19], Zhang-Gao-Yuan [30] for derivative NLS, and by Berti-Biasco-Procesi [7]-[8] for derivative NLW. For a recent survey of known results for KdV, we refer to [15]. Actually all these results still concern semi-linear perturbations.

The KAM theorems in [24], [20] prove the persistence of the finite-gap solutions of the integrable KdV under semilinear Hamiltonian perturbations  $\varepsilon \partial_x(\partial_u f)(x, u)$ , namely when the density  $f$  is independent of  $u_x$ , so that (1.2) is a differential operator of order 1. The key idea in [24] is to exploit the fact that the frequencies of KdV grow as  $\sim j^3$  and the difference  $|j^3 - i^3| \geq \frac{1}{2}(j^2 + i^2)$ ,  $i \neq j$ , so that KdV gains (outside the diagonal) two derivatives. This approach also works for Hamiltonian pseudo-differential perturbations of order 2 (in space), using the improved Kuksin's lemma proved by Liu-Yuan in [19]. However it does *not* work for the general quasi-linear perturbation in (1.2), which is a nonlinear differential operator of the *same* order as the constant coefficient linear operator  $\partial_{xxx}$ .

Now we state precisely the main result of the paper. The solutions we find are, at the first order of amplitude, localized in Fourier space on finitely many “tangential sites”

$$S^+ := \{\bar{j}_1, \dots, \bar{j}_\nu\}, \quad S := \{\pm j : j \in S^+\}, \quad \bar{j}_i \in \mathbb{N} \setminus \{0\} \quad \forall i = 1, \dots, \nu. \quad (1.11)$$

The set  $S$  is required to be even because the solutions  $u$  of (1.1) have to be real valued. Moreover, we also assume the following explicit hypothesis on  $S$ :

$$\frac{2}{2\nu - 1} \sum_{i=1}^{\nu} \bar{j}_i^2 \notin \left\{ j^2 + kj + k^2 : j, k \in \mathbb{Z} \setminus S, j \neq k \right\}. \quad (1.12)$$

Assumption (1.12) is a “non-degeneracy” condition. We assume it to prove that the Cantor-like set of amplitudes  $\xi \in \mathbb{R}_+^\nu$  for which the quasi-periodic solution (1.13) exists has positive measure, see Lemmata 28-29 and Remark 13.

**Theorem 1 (KAM for quasi-linear perturbations of mKdV)** *Given  $\nu \in \mathbb{N}$ , let  $f \in C^q$  (with  $q := q(\nu)$  large enough) satisfy (1.8). Then, for all the tangential sites  $S$  as in (1.11) satisfying (1.12), the mKdV equation (1.1) possesses small amplitude quasi-periodic solutions with diophantine frequency vector  $\omega := \omega(\xi) = (\omega_j)_{j \in S^+} \in \mathbb{R}^\nu$  of the form*

$$u(t, x) = \sum_{j \in S^+} 2\sqrt{\xi_j} \cos(\omega_j t + jx) + o(\sqrt{|\xi|}), \quad (1.13)$$

where

$$\omega_j := j^3 + 3\varsigma \left[ \xi_j - 2 \left( \sum_{j' \in S^+} \xi_{j'} \right) \right] j, \quad j \in S^+, \quad (1.14)$$

for a “Cantor-like” set of small amplitudes  $\xi \in \mathbb{R}_+^\nu$  with density 1 at  $\xi = 0$ . The term  $o(\sqrt{|\xi|})$  in (1.13) is a function  $u_1(t, x) = \tilde{u}_1(\omega t, x)$ , with  $\tilde{u}_1$  in the Sobolev space  $H^s(\mathbb{T}^{\nu+1}, \mathbb{R})$  of periodic functions, and Sobolev norm  $\|\tilde{u}_1\|_s = o(\sqrt{|\xi|})$  as  $\xi \rightarrow 0$ , for some  $s < q$ . These quasi-periodic solutions are linearly stable.

If the density  $f(u, u_x)$  is independent on  $x$ , a similar result holds for all the choices of the tangential sites, without assuming (1.12).

This result is deduced from Theorem 2. It was announced also in [4]-[5] under the stronger condition on the tangential sites

$$\frac{2}{2\nu-1} \sum_{i=1}^{\nu} \bar{j}_i^2 \notin \mathbb{Z}. \quad (1.15)$$

Let us make some comments.

1. In the case  $\nu = 1$  (time-periodic solutions), the condition (1.12) is always satisfied. Indeed, suppose, by contradiction, that there exist integers  $\bar{j}_1 \geq 1$ ,  $j, k \in \mathbb{Z}$  such that

$$2\bar{j}_1^2 = j^2 + jk + k^2. \quad (1.16)$$

Then  $j^2 + jk + k^2$  is even, and therefore both  $j$  and  $k$  are even, say  $j = 2n$ ,  $k = 2m$  with  $n, m \in \mathbb{Z}$ . Hence  $2\bar{j}_1^2 = 4(n^2 + nm + m^2)$ , and this implies that  $\bar{j}_1$  is even, say  $\bar{j}_1 = 2p$  for some positive integer  $p$ . It follows that  $2p^2 = n^2 + nm + m^2$ , namely  $p, n, m$  satisfy (1.16). Then, iterating the argument, we deduce that  $\bar{j}_1$  can be divided by 2 infinitely many times in  $\mathbb{N}$ , which is impossible.

2. When the density  $f(u, u_x)$  is independent of  $x$ , the  $L^2$ -norm

$$M(u) := \int_{\mathbb{T}} u^2 dx = \|u\|_{L^2(\mathbb{T})}^2 \quad (1.17)$$

is a prime integral of the Hamiltonian equation (1.1). Hence the solutions of (1.1) are in one-to-one correspondence with those of the Hamiltonian equation

$$v_t = \partial_x \nabla K(v) \quad \text{with} \quad K := H + \lambda M^2, \quad \lambda \in \mathbb{R}. \quad (1.18)$$

More precisely, if  $u(t, x)$  is a solution of (1.1), then  $v(t, x) := u(t, x - ct)$ , with  $c := -4\lambda M(u)$ , is a solution of (1.18). Vice versa, if  $v(t, x)$  solves (1.18), then the function  $u(t, x) := v(t, x + ct)$ , with  $c := -4\lambda M(v)$ , is a solution of (1.1) ( $M(v)$  is also a prime integral of the equation (1.18)).

The advantage of looking for quasi-periodic solutions of (1.18) is that, for  $\lambda = 3\varsigma/4$ , the fourth order Birkhoff normal form of  $K$  is diagonal (remark 1) and therefore no conditions on the tangential sites  $S$  are required (remark 13).

3. The diophantine frequency vector  $\omega(\xi) = (\omega_j)_{j \in S^+} \in \mathbb{R}^{\nu}$  of the quasi-periodic solutions of Theorem 1 is  $O(|\xi|)$ -close as  $\xi \rightarrow 0$  (see (1.14)) to the integer vector of the unperturbed linear frequencies

$$\bar{\omega} := (\bar{j}_1^3, \dots, \bar{j}_\nu^3) \in \mathbb{N}^{\nu}. \quad (1.19)$$

This makes perturbation theory more difficult. This is the difficulty due to the fact that the mKdV equation (1.1) is completely resonant at  $u = 0$ .

4. As shown by (1.13) the expected quasi-periodic solutions are mainly supported in Fourier space on the tangential sites  $S$ . The dynamics of the Hamiltonian PDE (1.1) restricted (and projected) to the symplectic subspaces

$$H_S := \left\{ v = \sum_{j \in S} u_j e^{ijx} \right\}, \quad H_S^\perp := \left\{ z = \sum_{j \in S^c} u_j e^{ijx} \in H_0^1(\mathbb{T}_x) \right\}, \quad (1.20)$$

where  $S^c := \{j \in \mathbb{Z} \setminus \{0\} : j \notin S\}$ , is quite different. We call  $v$  the *tangential* variable and  $z$  the *normal* one. On  $H_S$  the dynamics is mainly governed by a

finite dimensional integrable system (see Proposition 1), and we find it convenient to describe the dynamics in this subspace by introducing action-angle variable, see section 4. On the infinite dimensional subspace  $H_S^\perp$  the solution will stay forever close to the elliptic equilibrium  $z = 0$ .

In Theorem 1 it is stated that the quasi-periodic solutions are linearly stable. This information is not only an important complement of the result, but also an essential ingredient for the existence proof. Let us explain better what we mean. By the general procedure in [10] we prove that, around each invariant torus, there exist symplectic coordinates (see (6.13))

$$(\psi, \eta, w) \in \mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp$$

in which the mKdV Hamiltonian (1.4) assumes the normal form

$$\begin{aligned} K(\psi, \eta, w) = & \omega \cdot \eta + \frac{1}{2} K_{20}(\psi) \eta \cdot \eta + (K_{11}(\psi) \eta, w)_{L^2(\mathbb{T})} + \frac{1}{2} (K_{02}(\psi) w, w)_{L^2(\mathbb{T})} \\ & + K_{\geq 3}(\psi, \eta, w) \end{aligned} \quad (1.21)$$

where  $K_{\geq 3}$  collects the terms at least cubic in the variables  $(\eta, w)$ , see remark 4. In these coordinates the quasi-periodic solution reads  $t \mapsto (\omega t, 0, 0)$  and the corresponding linearized equations are

$$\begin{cases} \dot{\psi} = K_{20}(\omega t) \eta + K_{11}^T(\omega t) w \\ \dot{\eta} = 0 \\ \dot{w} - \partial_x K_{02}(\omega t) w = \partial_x K_{11}(\omega t) \eta. \end{cases} \quad (1.22)$$

Thus the actions  $\eta(t) = \eta(0)$  do not evolve in time and the third equation reduces to the forced PDE

$$\dot{w} = \partial_x K_{02}(\omega t)[w] + \partial_x K_{11}(\omega t)[\eta_0]. \quad (1.23)$$

Ignoring the forcing term  $\partial_x K_{11}(\omega t)[\eta_0]$  for a moment, we note that the equation  $\dot{w} = \partial_x K_{02}(\omega t)[w]$  is, up to a finite dimensional remainder (Proposition 3), the restriction to  $H_S^\perp$  of the ‘‘variational equation’’

$$h_t = \partial_x (\partial_u \nabla H)(u(\omega t, x))[h] = X_K(h),$$

where  $X_K$  is the KdV Hamiltonian vector field with quadratic Hamiltonian  $K = \frac{1}{2} ((\partial_u \nabla H)(u)[h], h)_{L^2(\mathbb{T}_x)} = \frac{1}{2} (\partial_{uu} H)(u)[h, h]$ . This is a linear PDE with quasi-periodically time-dependent coefficients of the form

$$h_t = \partial_{xx} (a_1(\omega t, x) \partial_x h) + \partial_x (a_0(\omega t, x) h). \quad (1.24)$$

In section 8 we prove the reducibility of the linear operator  $\dot{w} - \partial_x K_{02}(\omega t)w$ , which conjugates (1.23) to the diagonal system (see (8.64))

$$\partial_t v = -i \mathcal{D}_\infty v + f(\omega t) \quad (1.25)$$

where  $\mathcal{D}_\infty := \text{Op}\{\mu_j^\infty\}_{j \in S^c}$  is a Fourier multiplier operator acting in  $H_\perp^s$ ,

$$\mu_j^\infty := i(-m_3 j^3 + m_1 j) + r_j^\infty \in i\mathbb{R}, \quad j \in S^c,$$

with  $m_3 = 1 + O(\varepsilon^3)$ ,  $m_1 = O(\varepsilon^2)$ ,  $\sup_{j \in S^c} r_j^\infty = o(\varepsilon^2)$ , see (8.61), (8.62). The eigenvalues  $\mu_j^\infty$  are the *Floquet* exponents of the quasi-periodic solution. The solutions of the scalar non-homogeneous equations

$$\dot{v}_j + \mu_j^\infty v_j = f_j(\omega t), \quad j \in S^c, \quad \mu_j^\infty \in i\mathbb{R},$$

are

$$v_j(t) = c_j e^{\mu_j^\infty t} + \tilde{v}_j(t), \quad \text{where} \quad \tilde{v}_j(t) := \sum_{l \in \mathbb{Z}^\nu} \frac{f_{jl} e^{i\omega \cdot l t}}{i\omega \cdot l + \mu_j^\infty}$$

(recall that the first Melnikov conditions (8.66) hold at a solution). As a consequence, the Sobolev norm of the solution of (1.25) satisfies

$$\|v(t)\|_{H_x^s} \leq C \|v(0)\|_{H_x^s}, \quad \forall t \in \mathbb{R},$$

i.e. it does not increase in time.

We now describe in detail the strategy of proof of Theorem 1. Many of the arguments that we use are quite general and of wide applicability to other PDEs. Nevertheless, we think that a unique abstract theorem of existence and stability of quasi-periodic solutions applicable to *all* quasi-linear PDEs cannot be expected. Indeed the suitable pseudo-differential operators that are required to conjugate the highest order of the linearized operator to constant coefficients highly depend on the PDE at hand, see the discussion after (1.29).

There are two main issues in the proof:

1. **Bifurcation analysis.** Find approximate quasi-periodic solutions of (1.1) up to a sufficiently small remainder (which, in our case, should be  $O(u^4)$ ). In this step we also find the approximate “frequency-to-amplitude” modulation of the frequency with respect to the amplitude, see (4.10). This is the goal of sections 3 and 4.
2. **Nash-Moser implicit function theorem.** Prove that, close to the above approximate solutions, there exist exact quasi-periodic solutions of (1.1). By means of a Nash-Moser iteration, we construct a sequence of approximate solutions that converges to a quasi-periodic solution of (1.1) (sections 5-9).

The key step consists in proving the invertibility of the linearized operator and tame estimates for its inverse. This is achieved in two main steps.

- (a) **SYMPLECTIC DECOUPLING PROCEDURE.** The method in Berti-Bolle [10] allows to approximately decouple the “tangential” and the “normal” dynamics around an approximate invariant torus (section 6). It reduces the problem to the one of inverting a quasi-periodically forced PDE restricted to the normal subspace  $H_S^{\frac{1}{2}}$ . Its precise form is found in section 7.2.
- (b) **ANALYSIS OF THE LINEARIZED OPERATOR IN THE NORMAL DIRECTIONS.** In sections 7, 8 we reduce the linearized equations to constant coefficients. This involves three steps:
  - i. *Reduction in decreasing symbols*, sections 8.1-8.3 and 8.5,
  - ii. *Linear Birkhoff normal form*, section 8.4,
  - iii. *KAM reducibility*, section 8.6.

All the changes of variables used in the steps i)-iii) are  $\varphi$ -dependent families of symplectic maps  $\Phi(\varphi)$  which act on the phase space  $H_0^1(\mathbb{T}_x)$ . Therefore they preserve the Hamiltonian dynamical systems structure of the conjugated linear operators.



Let us discuss these issues in detail.

*Weak Birkhoff normal form.* According to the orthogonal splitting

$$H_0^1(\mathbb{T}_x) := H_S \oplus H_S^\perp$$

into the symplectic subspaces defined in (1.20), we decompose

$$u = v + z, \quad v = \Pi_S u := \sum_{j \in S} u_j e^{ijx}, \quad z = \Pi_S^\perp u := \sum_{j \in S^c} u_j e^{ijx}, \quad (1.26)$$

where  $\Pi_S, \Pi_S^\perp$  denote the orthogonal projectors on  $H_S, H_S^\perp$ .

We perform a “weak” Birkhoff normal form (weak BNF), whose goal is to find an invariant manifold of solutions of the third order approximate mKdV equation (1.1), on which the dynamics is completely integrable, see section 3. We construct in Proposition 1 a symplectic map  $\Phi_B$  such that the transformed Hamiltonian  $\mathcal{H} := H \circ \Phi_B$  possesses the invariant subspace  $H_S$  (see (1.20)). To this purpose we have to eliminate the term  $\int v^3 z dx$  (which is linear in  $z$ ). Then we check that its dynamics on  $H_S$  is integrable and non-isocronous. For that we perform the classical finite dimensional Birkhoff normalization of the Hamiltonian term  $\int v^4 dx$  which turns out to be integrable and non-isocronous.

Since the present weak Birkhoff map has to remove only finitely many monomials, it is the time 1-flow map of an Hamiltonian system whose Hamiltonian is supported on only finitely many Fourier indices. Therefore it is close to the identity up to *finite dimensional* operators, see Proposition 1. The key advantage is that it modifies  $\mathcal{N}_4$  very mildly, only up to finite dimensional operators (see for example Lemma 12), and thus the spectral analysis of the linearized equations (that we shall perform in section 8) is essentially the same as if we were in the *original* coordinates.

The weak normal form (3.7) does not remove (nor normalize) the monomials  $O(z^2)$ . We point out that a stronger normal form that removes/normalizes the monomials  $O(z^2)$  is also well-defined (it is called “partial Birkhoff normal form” in Kuksin-Pöschel [26] and Pöschel [27]). However, we do not use it because, for such a stronger normal form, the corresponding Birkhoff map is close to the identity only up to an operator of order  $O(\partial_x^{-1})$ , and so it would produce terms of order  $\partial_{xx}$  and  $\partial_x$ . For the same reason, we do not use the global nonlinear Fourier transform in [21] (Birkhoff coordinates), which is close to the Fourier transform up to smoothing operators of order  $O(\partial_x^{-1})$  (this is explicitly proved for KdV).

We remark that mKdV is simpler than KdV because the nonlinearity in (1.1) is cubic and not only quadratic, and, as a consequence, less steps of Birkhoff normal form are required to reach the sufficient smallness for the Nash-Moser scheme to converge (see Remark 11).

*Action-angle and rescaling.* At this point we introduce action-angle variables on the tangential sites (section 4) and, after the rescaling (4.5), we look for quasi-periodic solutions of the Hamiltonian (4.9). Note that the coefficients of the normal form  $\mathcal{N}$  in (4.13) depend on the angles  $\theta$ , unlike the usual KAM theorems [27], [23], where the whole normal form is reduced to constant coefficients. This is because the weak BNF of section 3 did not normalize the quadratic terms  $O(z^2)$ . These terms are dealt with the “linear Birkhoff normal form” (linear BNF) in section

8.4. In some sense the “partial” Birkhoff normal form of [27] is split into the weak BNF of section 3 and the linear BNF of sections 8.4.

The present functional formulation with the introduction of the action-angle variables allows to prove the stability of the solutions (unlike the Lyapunov-Schmidt reduction approach).

*Nonlinear functional setting and approximate inverse.* We look for a zero of the nonlinear operator (5.6), where the unknown is the torus embedding  $\varphi \mapsto i(\varphi)$ , and where the frequency  $\omega$  is seen as an “external” parameter. This formulation is convenient in order to verify the Melnikov non-resonance conditions required to invert the linearized operators at each step. The solution is obtained by a Nash-Moser iterative scheme in Sobolev scales. The key step is to construct (for  $\omega$  restricted to a suitable Cantor-like set) an approximate inverse (*à la* Zehnder [31]) of the linearized operator at any approximate solution. Roughly, this means to find a linear operator which is an inverse at an exact solution. A major difficulty is that the tangential and the normal dynamics near an invariant torus are strongly coupled.

*Symplectic approximate decoupling.* The above difficulty is overcome by implementing the abstract procedure in Berti-Bolle [10], which was developed in order to prove the existence of quasi-periodic solutions for autonomous NLW (and NLS) with a multiplicative potential. This approach reduces the search of an approximate inverse for (5.6) to the invertibility of a quasi-periodically forced PDE restricted to the normal directions. This method approximately decouples the tangential and the normal dynamics around an approximate invariant torus, introducing a suitable set of symplectic variables

$$(\psi, \eta, w) \in \mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp$$

near the torus, see (6.13). Note that, in the first line of (6.13),  $\psi$  is the “natural” angle variable which coordinates the torus, and, in the third line, the normal variable  $z$  is only translated by the component  $z_0(\psi)$  of the torus. The second line completes this transformation to a symplectic one. The canonicity of this map is proved in [10] using the isotropy of the approximate invariant torus  $i_\delta$ , see Lemma 8. In these new variables the torus  $\psi \mapsto i_\delta(\psi)$  reads  $\psi \mapsto (\psi, 0, 0)$ . The main advantage of these coordinates is that the second equation in (6.22) (which corresponds to the action variables of the torus) can be immediately solved, see (6.24). Then it remains to solve the third equation (6.25), i.e. to invert the linear operator  $\mathcal{L}_\omega$ . This is a quasi-periodic Hamiltonian perturbed linear Airy equation of the form

$$h \mapsto \mathcal{L}_\omega h := \Pi_S^\perp (\omega \cdot \partial_\varphi h + \partial_{xx}(a_1 \partial_x h) + \partial_x(a_0 h) + \partial_x \mathcal{R}h), \quad \forall h \in H_S^\perp, \quad (1.27)$$

where  $\mathcal{R}$  is a finite dimensional remainder. The exact form of  $\mathcal{L}_\omega$  is obtained in Proposition 3, see (7.23).

*Reduction to constant coefficients of the linearized operator in the normal directions.* In section 8 we conjugate the variable coefficients operator  $\mathcal{L}_\omega$  to a diagonal operator with constant coefficients which describes infinitely many harmonic oscillators

$$\dot{v}_j + \mu_j^\infty v_j = 0, \quad \mu_j^\infty := i(-m_3 j^3 + m_1 j) + r_j^\infty \in i\mathbb{R}, \quad j \notin S, \quad (1.28)$$

where the constants  $m_3 - 1, m_1 \in \mathbb{R}$  and  $\sup_j |r_j^\infty|$  are small, see Theorem 4. The main perturbative effect to the spectrum (and the eigenfunctions) of  $\mathcal{L}_\omega$  is due to

the term  $a_1(\omega t, x)\partial_{xxx}$  (see (1.27)), and it is too strong for the usual reducibility KAM techniques to work directly. The conjugacy of  $\mathcal{L}_\omega$  with (1.28) is obtained in several steps. The first task (obtained in sections 8.1-8.5) is to conjugate  $\mathcal{L}_\omega$  to another Hamiltonian operator of  $H_S^\perp$  with constant coefficients

$$\mathcal{L}_5 := \Pi_S^\perp (\omega \cdot \partial_\varphi + m_3 \partial_{xxx} + m_1 \partial_x + R_5) \Pi_S^\perp, \quad m_1, m_3 \in \mathbb{R}, \quad (1.29)$$

up to a small bounded remainder  $R_5 = O(\partial_x^0)$ , see (8.56). This expansion of  $\mathcal{L}_\omega$  in “decreasing symbols” with constant coefficients follows [3], and it is somehow in the spirit of the works of Iooss, Plotnikov and Toland [17]-[16] in water waves theory, and Baldi [2] for Benjamin-Ono. It is obtained by transformations which are very different from the usual KAM changes of variables. We underline that the specific form of these transformations depend on the structure of mKdV. For other quasi-linear PDEs the analogous reduction requires different transformations, see for example Alazard-Baldi [1], Berti-Montalto [12] for recent developments of these techniques for gravity-capillary water waves, and Feola-Procesi [14] for quasi-linear forced perturbations of Schrödinger equations.

The transformation of (1.27) into (1.29) is made in several steps.

1. *Reduction of the highest order.* The first step (section 8.1) is to eliminate the  $x$ -dependence from the coefficient  $a_1(\omega t, x)\partial_{xxx}$  of the Hamiltonian operator  $\mathcal{L}_\omega$ . For this purpose, we have to construct a symplectic diffeomorphism of  $H_S^\perp$  near  $\mathcal{A}_\perp := \Pi_S^\perp \mathcal{A} \Pi_S^\perp$ , where  $\mathcal{A}$  is a diffeomorphism of the form

$$u \mapsto (\mathcal{A}u)(\varphi, x) := (1 + \beta_x(\varphi, x))u(\varphi, x + \beta(\varphi, x)),$$

see (8.1). The starting point is to observe that  $\mathcal{A}$  is, for each  $\varphi \in \mathbb{T}^\nu$ , the time-one flow map of the time dependent Hamiltonian transport linear PDE

$$\partial_\tau u = \partial_x(b(\varphi, \tau, x)u), \quad b(\varphi, \tau, x) := \frac{\beta(\varphi, x)}{1 + \tau\beta_x(\varphi, x)}. \quad (1.30)$$

Actually the flow of (1.30) is the path of symplectic diffeomorphisms

$$u(\varphi, x) \mapsto (1 + \tau\beta_x(\varphi, x))u(\varphi, x + \tau\beta(\varphi, x)), \quad \tau \in [0, 1].$$

Thus, like in [5], we conjugate  $\mathcal{L}_\omega$  with the symplectic time-one flow map of the projected Hamiltonian equation

$$\partial_\tau u = \Pi_S^\perp \partial_x(b(\tau, x)u) = \partial_x(b(\tau, x)u) - \Pi_S \partial_x(b(\tau, x)u), \quad u \in H_S^\perp \quad (1.31)$$

generated by the the quadratic Hamiltonian  $\frac{1}{2} \int_{\mathbb{T}} b(\tau, x)u^2 dx$  restricted to  $H_S^\perp$ . By Lemma 15 (which was proved in [5]) such a symplectic map differs from  $\mathcal{A}_\perp$  only for finite dimensional operators.

This step may be seen as a quantitative application of the Egorov theorem, see [29], which describes how the principal symbol of a pseudo-differential operator (here  $a_1(\omega t, x)\partial_{xxx}$ ) transforms under the flow of a linear hyperbolic PDE (here (1.31)).

Because of the Hamiltonian structure, the previous step also eliminates the term  $O(\partial_{xx})$ , see (8.13). In section 8.2 we eliminate the time-dependence of the coefficient at the order  $\partial_{xxx}$ .

2. *Linear Birkhoff normal form.* In section 8.4 we eliminate the variable coefficient terms at the order  $O(\varepsilon^2)$ , which are present in the operator  $\mathcal{L}_\omega$ , see (7.23)-(7.24). This is a consequence of the fact that the weak BNF procedure of section 3 did not touch the quadratic terms  $O(z^2)$ . These terms cannot be reduced to constants by the perturbative scheme in section 8.6 (developed in [3]) which applies to terms  $R$  such that  $R\gamma^{-1} \ll 1$  where  $\gamma$  is the diophantine constant of the frequency vector  $\omega$  (the case in [3] is simpler because the diophantine constant is  $\gamma = O(1)$ ). Here, as well as in [5], since mKdV is completely resonant, such  $\gamma = o(\varepsilon^2)$ , see (5.3). The terms of size  $\varepsilon^2$  are reduced to constant coefficients in section 8.4 by means of purely algebraic arguments (linear BNF), which, ultimately, stem from the complete integrability of the fourth order BNF of the mKdV equation (1.9). More general nonlinearities should be dealt with the normal form arguments of Procesi-Procesi [28] for generic choices of the tangential sites.

*Complete diagonalization of (1.29).* In section 8.6 we apply the abstract KAM reducibility Theorem 4.2 of [3], which completely diagonalizes the linearized operator, obtaining (1.28). The required smallness condition (8.58) for  $R_5$  holds, after that the linear BNF of section 8.4 has put into constant coefficients the unbounded terms of nonperturbative size  $\varepsilon^2$ , and the conjugation procedure of sections 8.1-8.3 and 8.5 has arrived to a bounded and small remainder  $R_5$ .

*The Nash-Moser iteration to an invariant torus embedding.* In section 9 we perform the nonlinear Nash-Moser iteration which finally proves Theorem 2 and, therefore, Theorem 1. The smallness condition that is required for the convergence of the scheme is  $\varepsilon^2 \|\mathcal{F}(\varphi, 0, 0)\|_{s_0+\mu} \gamma^{-2}$  sufficiently small, see (9.5). It is verified because  $\|X_P(\varphi, 0, 0)\|_s \leq_s \varepsilon^{5-2b}$  (Lemma 5) and  $\gamma = \varepsilon^{2+a}$  with  $a > 0$  small. See also remark 11 for a comparison between the smallness condition required here with the one in [5].

*Notation.* We shall use the notation

$$a \leq_s b \iff a \leq C(s)b \text{ for some constant } C(s) > 0.$$

We denote by  $\pi_0$  the operator

$$u \mapsto \pi_0(u) := u - \frac{1}{2\pi} \int_{\mathbb{T}} u \, dx. \quad (1.32)$$

## 2 Functional setting

For a function  $u : \Omega_o \rightarrow E$ ,  $\omega \mapsto u(\omega)$ , where  $(E, \|\cdot\|_E)$  is a Banach space and  $\Omega_o$  is a subset of  $\mathbb{R}^\nu$ , we define the sup-norm and the Lipschitz semi-norm

$$\begin{aligned} \|u\|_E^{\text{sup}} &:= \|u\|_{E, \Omega_o}^{\text{sup}} := \sup_{\omega \in \Omega_o} \|u(\omega)\|_E, \\ \|u\|_E^{\text{lip}} &:= \|u\|_{E, \Omega_o}^{\text{lip}} := \sup_{\omega_1 \neq \omega_2} \frac{\|u(\omega_1) - u(\omega_2)\|_E}{|\omega_1 - \omega_2|}, \end{aligned} \quad (2.1)$$

and, for  $\gamma > 0$ , the Lipschitz norm

$$\|u\|_E^{\text{Lip}(\gamma)} := \|u\|_{E, \Omega_o}^{\text{Lip}(\gamma)} := \|u\|_E^{\text{sup}} + \gamma \|u\|_E^{\text{lip}}. \quad (2.2)$$

If  $E = H^s$  we simply denote  $\|u\|_{H^s}^{\text{Lip}(\gamma)} := \|u\|_s^{\text{Lip}(\gamma)}$ .

*Sobolev norms.* We denote by

$$\|u\|_s := \|u\|_{H^s(\mathbb{T}^{\nu+1})} := \|u\|_{H_{\varphi,x}^s} \quad (2.3)$$

the Sobolev norm of functions  $u = u(\varphi, x)$  in the Sobolev space  $H^s(\mathbb{T}^{\nu+1})$ . We denote by  $\|\cdot\|_{H_x^s}$  the Sobolev norm in the phase space of functions  $u := u(x) \in H^s(\mathbb{T})$ . Moreover  $\|\cdot\|_{H_\varphi^s}$  denotes the Sobolev norm of scalar functions, like the Fourier components  $u_j(\varphi)$ .

We fix  $s_0 := (\nu+2)/2$  so that  $H^{s_0}(\mathbb{T}^{\nu+1}) \hookrightarrow L^\infty(\mathbb{T}^{\nu+1})$  and any space  $H^s(\mathbb{T}^{\nu+1})$ ,  $s \geq s_0$ , is an algebra and satisfy the interpolation inequalities: for  $s \geq s_0$ ,

$$\|uv\|_s \leq C(s_0)\|u\|_s\|v\|_{s_0} + C(s)\|u\|_{s_0}\|v\|_s, \quad \forall u, v \in H^s(\mathbb{T}^d).$$

The above inequalities also hold for the norms  $\|\cdot\|_s^{\text{Lip}(\gamma)}$ .

We also denote

$$\begin{aligned} H_{S^\perp}^s(\mathbb{T}^{\nu+1}) &:= \{u \in H^s(\mathbb{T}^{\nu+1}) : u(\varphi, \cdot) \in H_S^\perp \forall \varphi \in \mathbb{T}^\nu\}, \\ H_S^s(\mathbb{T}^{\nu+1}) &:= \{u \in H^s(\mathbb{T}^{\nu+1}) : u(\varphi, \cdot) \in H_S \forall \varphi \in \mathbb{T}^\nu\}. \end{aligned}$$

*Matrices with off-diagonal decay.* A linear operator can be identified, as usual, with its matrix representation. We recall the definition of the  $s$ -decay norm (introduced in [9]) of an infinite dimensional matrix.

**Definition 1** Let  $A := (A_{i_1, i_2}^{i_2})_{i_1, i_2 \in \mathbb{Z}^b}$ ,  $b \geq 1$ , be an infinite dimensional matrix. Its  $s$ -decay norm  $|A|_s$  is defined by

$$|A|_s^2 := \sum_{i \in \mathbb{Z}^b} \langle i \rangle^{2s} \left( \sup_{i_1 - i_2 = i} |A_{i_1, i_2}^{i_2}| \right)^2. \quad (2.4)$$

For parameter dependent matrices  $A := A(\omega)$ ,  $\omega \in \Omega_o \subseteq \mathbb{R}^\nu$ , the definitions (2.1) and (2.2) become

$$|A|_s^{\text{sup}} := \sup_{\omega \in \Omega_o} |A(\omega)|_s, \quad |A|_s^{\text{lip}} := \sup_{\omega_1 \neq \omega_2} \frac{|A(\omega_1) - A(\omega_2)|_s}{|\omega_1 - \omega_2|}, \quad (2.5)$$

and  $|A|_s^{\text{Lip}(\gamma)} := |A|_s^{\text{sup}} + \gamma |A|_s^{\text{lip}}$ .

Such a norm is modeled on the behavior of matrices representing the multiplication operator by a function. Actually, given a function  $p \in H^s(\mathbb{T}^b)$ , the multiplication operator  $h \mapsto ph$  is represented by the Töplitz matrix  $T_i^{i'} = p_{i-i'}$  and  $|T|_s = \|p\|_s$ . If  $p = p(\omega)$  is a Lipschitz family of functions, then

$$|T|_s^{\text{Lip}(\gamma)} = \|p\|_s^{\text{Lip}(\gamma)}.$$

The  $s$ -norm satisfies classical algebra and interpolation inequalities proved in [3].

**Lemma 1** Let  $A = A(\omega), B = B(\omega)$  be matrices depending in a Lipschitz way on the parameter  $\omega \in \Omega_o \subset \mathbb{R}^\nu$ . Then for all  $s \geq s_0 > b/2$  there are  $C(s) \geq C(s_0) \geq 1$  such that

$$\begin{aligned} |AB|_s^{\text{Lip}(\gamma)} &\leq C(s) |A|_s^{\text{Lip}(\gamma)} |B|_s^{\text{Lip}(\gamma)}, \\ |AB|_s^{\text{Lip}(\gamma)} &\leq C(s) |A|_s^{\text{Lip}(\gamma)} |B|_{s_0}^{\text{Lip}(\gamma)} + C(s_0) |A|_{s_0}^{\text{Lip}(\gamma)} |B|_s^{\text{Lip}(\gamma)}. \end{aligned}$$

The  $s$ -decay norm controls the Sobolev norm, namely

$$\|Ah\|_s^{\text{Lip}(\gamma)} \leq C(s) \left( |A|_{s_0}^{\text{Lip}(\gamma)} \|h\|_s^{\text{Lip}(\gamma)} + |A|_s^{\text{Lip}(\gamma)} \|h\|_{s_0}^{\text{Lip}(\gamma)} \right).$$

Let now  $b := \nu + 1$ . An important sub-algebra is formed by the *Töplitz in time matrices* defined by

$$A_{(l_1, j_1)}^{(l_2, j_2)} := A_{j_1}^{j_2}(l_1 - l_2),$$

whose decay norm (2.4) is

$$|A|_s^2 = \sum_{j \in \mathbb{Z}, l \in \mathbb{Z}^\nu} \left( \sup_{j_1 - j_2 = j} |A_{j_1}^{j_2}(l)| \right)^2 \langle l, j \rangle^{2s}.$$

These matrices are identified with the  $\varphi$ -dependent family of operators

$$A(\varphi) := \left( A_{j_1}^{j_2}(\varphi) \right)_{j_1, j_2 \in \mathbb{Z}}, \quad A_{j_1}^{j_2}(\varphi) := \sum_{l \in \mathbb{Z}^\nu} A_{j_1}^{j_2}(l) e^{il \cdot \varphi}$$

which act on functions of the  $x$ -variable as

$$A(\varphi) : h(x) = \sum_{j \in \mathbb{Z}} h_j e^{ijx} \mapsto A(\varphi)h(x) = \sum_{j_1, j_2 \in \mathbb{Z}} A_{j_1}^{j_2}(\varphi) h_{j_2} e^{ij_1 x}.$$

Transformations of this kind were also used in [3] and in [9], [11]. All the transformations that we construct in this paper are of this type (with  $j, j_1, j_2 \neq 0$  because they act on the phase space  $H_0^1(\mathbb{T}_x)$ ).

**Definition 2** We say that

1. an operator  $(Ah)(\varphi, x) := A(\varphi)h(\varphi, x)$  is *symplectic* if each  $A(\varphi)$ ,  $\varphi \in \mathbb{T}^\nu$ , is a symplectic map of the phase space (or of a symplectic subspace like  $H_S^{\frac{1}{2}}$ );
2. an operator is *real* if it maps real-valued functions into real-valued functions;
3. the real operator  $\omega \cdot \partial_\varphi - \partial_x G(\varphi)$  is *Hamiltonian* if each  $G(\varphi)$ ,  $\varphi \in \mathbb{T}^\nu$ , is self-adjoint with respect to the  $L^2(\mathbb{T})$  complex scalar product.

A Hamiltonian operator is transformed, under a symplectic map, into another Hamiltonian operator, see [3]-section 2.3.

We conclude this preliminary section recalling the following well known lemmata about composition of functions (see, e.g., Appendix of [3]).

**Lemma 2 (Composition)** *Assume  $f \in C^s(\mathbb{T}^d \times B_1)$ ,  $B_1 := \{y \in \mathbb{R}^m : |y| \leq 1\}$ . Then  $\forall u \in H^s(\mathbb{T}^d, \mathbb{R}^m)$  such that  $\|u\|_{L^\infty} < 1$ , the composition operator  $\tilde{f}(u)(x) := f(x, u(x))$  satisfies  $\|\tilde{f}(u)\|_s \leq C\|f\|_{C^s}(\|u\|_s + 1)$  where the constant  $C$  depends on  $s, d$ . If  $f \in C^{s+2}$  and  $\|u + h\|_{L^\infty} < 1$ , then for  $k = 0, 1$*

$$\left\| \tilde{f}(u + h) - \sum_{i=0}^k \frac{\tilde{f}^{(i)}(u)}{i!} [h^i] \right\|_s \leq C\|f\|_{C^{s+2}} \|h\|_{L^\infty}^k (\|h\|_s + \|h\|_{L^\infty} \|u\|_s).$$

*The statement also holds replacing  $\|\cdot\|_s$  with the norms  $|\cdot|_{s, \infty}$  of  $W^{s, \infty}(\mathbb{T}^d)$ .*

**Lemma 3 (Change of variable)** *Let  $p \in W^{s,\infty}(\mathbb{T}^d, \mathbb{R}^d)$ ,  $s \geq 1$ , with  $\|p\|_{W^{1,\infty}} \leq 1/2$ . Then the function  $f(x) = x + p(x)$  is invertible, with inverse  $f^{-1}(y) = y + q(y)$  where  $q \in W^{s,\infty}(\mathbb{T}^d, \mathbb{R}^d)$ , and  $\|q\|_{W^{s,\infty}} \leq C\|p\|_{W^{s,\infty}}$ .*

*If, moreover,  $p$  depends in a Lipschitz way on a parameter  $\omega \in \Omega \subset \mathbb{R}^\nu$ , and  $\|D_x p\|_{L^\infty} \leq 1/2$  for all  $\omega$ , then  $\|q\|_{W^{s,\infty}}^{\text{Lip}(\gamma)} \leq C\|p\|_{W^{s+1,\infty}}^{\text{Lip}(\gamma)}$ . The constant  $C := C(d, s)$  is independent of  $\gamma$ .*

*If  $u \in H^s(\mathbb{T}^d, \mathbb{C})$ , then  $(u \circ f)(x) := u(x + p(x))$  satisfies*

$$\begin{aligned} \|u \circ f\|_s &\leq C(\|u\|_s + \|p\|_{W^{s,\infty}}\|u\|_1), \\ \|u \circ f - u\|_s &\leq C(\|p\|_{L^\infty}\|u\|_{s+1} + \|p\|_{W^{s,\infty}}\|u\|_2), \\ \|u \circ f\|_s^{\text{Lip}(\gamma)} &\leq C(\|u\|_{s+1}^{\text{Lip}(\gamma)} + \|p\|_{W^{s,\infty}}^{\text{Lip}(\gamma)}\|u\|_2^{\text{Lip}(\gamma)}). \end{aligned}$$

*The function  $u \circ f^{-1}$  satisfies the same bounds.*

### 3 Weak Birkhoff normal form

In this section it is convenient to analyze the mKdV equation in the Fourier representation

$$u(x) = \sum_{j \in \mathbb{Z} \setminus \{0\}} u_j e^{ijx}, \quad u(x) \longleftrightarrow u := (u_j)_{j \in \mathbb{Z} \setminus \{0\}}, \quad u_{-j} = \bar{u}_j, \quad (3.1)$$

where the Fourier indices are nonzero integers  $j$ , by the definition (1.5) of the phase space, and  $u_{-j} = \bar{u}_j$  because  $u(x)$  is real-valued. The symplectic structure (1.6) writes

$$\Omega = \frac{1}{2} \sum_{j \neq 0} \frac{1}{ij} du_j \wedge du_{-j}, \quad \Omega(u, v) = \sum_{j \neq 0} \frac{1}{ij} u_j v_{-j}, \quad (3.2)$$

the Hamiltonian vector field  $X_H$  in (1.3) and the Poisson bracket  $\{F, G\}$  in (1.7) are respectively

$$[X_H(u)]_j = ij \partial_{u_{-j}} H(u), \quad \{F, G\}(u) = - \sum_{j \neq 0} ij (\partial_{u_{-j}} F)(u) (\partial_{u_j} G)(u). \quad (3.3)$$

We shall sometimes identify  $v \equiv (v_j)_{j \in S}$  and  $z \equiv (z_j)_{j \in S^c}$ .

The Hamiltonian of the perturbed cubic mKdV equation (1.1) is  $H = H_2 + H_4 + H_{\geq 5}$  (see (1.4)) where

$$H_2(u) := \int_{\mathbb{T}} \frac{u_x^2}{2} dx, \quad H_4(u) := -\varsigma \int_{\mathbb{T}} \frac{u^4}{4} dx, \quad H_{\geq 5}(u) := \int_{\mathbb{T}} f(x, u, u_x) dx, \quad (3.4)$$

$\varsigma = \pm 1$  and  $f$  satisfies (1.8). According to the splitting (1.26)  $u = v + z$ , where  $v \in H_S$  and  $z \in H_S^\perp$ , we have  $H_2(u) = H_2(v) + H_2(z)$  and

$$H_4(u) = -\frac{\varsigma}{4} \int_{\mathbb{T}} v^4 dx - \varsigma \int_{\mathbb{T}} v^3 z dx - \frac{3\varsigma}{2} \int_{\mathbb{T}} v^2 z^2 dx - \varsigma \int_{\mathbb{T}} v z^3 dx - \frac{\varsigma}{4} \int_{\mathbb{T}} z^4 dx.$$

For a finite-dimensional space

$$E := E_C := \text{span}\{e^{ijx} : 0 < |j| \leq C\}, \quad C > 0, \quad (3.5)$$

let  $\Pi_E$  denote the corresponding  $L^2$ -projector on  $E$ .

In the next proposition we construct a symplectic map  $\Phi_B$  such that the transformed Hamiltonian  $\mathcal{H} := H \circ \Phi_B$  possesses the invariant subspace  $H_S$  defined in (1.20), and its dynamics on  $H_S$  is integrable and non-isochronous. To this purpose we have to eliminate the term  $\int v^3 z dx$  (which is linear in  $z$ ) and to normalize the term  $\int v^4 dx$  (which is independent of  $z$ ) in the quartic component of the Hamiltonian.

**Proposition 1 (Weak Birkhoff normal form)** *There exists an analytic invertible symplectic transformation of the phase space  $\Phi_B : H_0^1(\mathbb{T}_x) \rightarrow H_0^1(\mathbb{T}_x)$  of the form*

$$\Phi_B(u) = u + \Psi(u), \quad \Psi(u) = \Pi_E \Psi(\Pi_E u), \quad (3.6)$$

where  $E$  is a finite-dimensional space as in (3.5), such that the transformed Hamiltonian is

$$\mathcal{H} := H \circ \Phi_B = H_2 + \mathcal{H}_4 + \mathcal{H}_{\geq 5}, \quad (3.7)$$

where  $H_2$  is defined in (3.4),

$$\begin{aligned} \mathcal{H}_4 := & \frac{3\varsigma}{4} \left( \sum_{j \in S} |u_j|^4 - \sum_{j, j' \in S} |u_j|^2 |u_{j'}|^2 \right) - \frac{3\varsigma}{2} \int_{\mathbb{T}} v^2 z^2 dx \\ & - \varsigma \int_{\mathbb{T}} v z^3 dx - \frac{\varsigma}{4} \int_{\mathbb{T}} z^4 dx, \end{aligned} \quad (3.8)$$

and  $\mathcal{H}_{\geq 5}$  collects all the terms of order at least five in  $(v, z)$ .

*Proof* In Fourier coordinates (3.1) we have (see (3.4))

$$H_2(u) = \frac{1}{2} \sum_{j \neq 0} j^2 |u_j|^2, \quad H_4(u) = -\frac{\varsigma}{4} \sum_{j_1 + j_2 + j_3 + j_4 = 0} u_{j_1} u_{j_2} u_{j_3} u_{j_4}. \quad (3.9)$$

We look for a symplectic transformation  $\Phi$  of the phase space which eliminates or normalizes the monomials  $u_{j_1} u_{j_2} u_{j_3} u_{j_4}$  of  $H_4$  with at most one index outside  $S$ . By the relation  $j_1 + j_2 + j_3 + j_4 = 0$ , they are *finitely* many. Thus, we look for a map  $\Phi := (\Phi_F^t)|_{t=1}$  which is the time 1-flow map of an auxiliary quartic Hamiltonian

$$F(u) := \sum_{j_1 + j_2 + j_3 + j_4 = 0} F_{j_1 j_2 j_3 j_4} u_{j_1} u_{j_2} u_{j_3} u_{j_4}.$$

The transformed Hamiltonian is

$$\mathcal{H} := H \circ \Phi = H_2 + \mathcal{H}_4 + \mathcal{H}_{\geq 5}, \quad \mathcal{H}_4 = \{H_2, F\} + H_4, \quad (3.10)$$

where  $\mathcal{H}_{\geq 5}$  collects all the terms in  $\mathcal{H}$  of order at least five. By (3.9) and (3.3) we calculate

$$\mathcal{H}_4 = \sum_{j_1 + j_2 + j_3 + j_4 = 0} \left\{ -\frac{\varsigma}{4} - i(j_1^3 + j_2^3 + j_3^3 + j_4^3) F_{j_1 j_2 j_3 j_4} \right\} u_{j_1} u_{j_2} u_{j_3} u_{j_4}.$$

In order to eliminate or normalize only the monomials with at most one index outside  $S$ , we choose

$$F_{j_1 j_2 j_3 j_4} := \begin{cases} \frac{i\varsigma}{4(j_1^3 + j_2^3 + j_3^3 + j_4^3)} & \text{if } (j_1, j_2, j_3, j_4) \in \mathcal{A}, \\ 0 & \text{otherwise,} \end{cases} \quad (3.11)$$



where

$$\mathcal{A} := \{(j_1, j_2, j_3, j_4) \in (\mathbb{Z} \setminus \{0\})^4 : j_1 + j_2 + j_3 + j_4 = 0, \quad j_1^3 + j_2^3 + j_3^3 + j_4^3 \neq 0, \\ \text{and at least three among } j_1, j_2, j_3, j_4 \text{ belong to } S\}.$$

We recall the following elementary identity (Lemma 13.4 in [20]).

**Lemma 4** *Let  $j_1, j_2, j_3, j_4 \in \mathbb{Z}$  such that  $j_1 + j_2 + j_3 + j_4 = 0$ . Then*

$$j_1^3 + j_2^3 + j_3^3 + j_4^3 = -3(j_1 + j_2)(j_1 + j_3)(j_2 + j_3).$$

By definition (3.11),  $\mathcal{H}_4$  does not contain any monomial  $u_{j_1} u_{j_2} u_{j_3} u_{j_4}$  with three indices in  $S$  and one outside, because there exist no integers  $j_1, j_2, j_3 \in S, j_4 \in S^c$  satisfying  $j_1 + j_2 + j_3 + j_4 = 0$  and  $j_1^3 + j_2^3 + j_3^3 + j_4^3 = 0$ , by Lemma 4 and the fact that  $S$  is symmetric.

By construction, the quartic monomials with at least two indices outside  $S$  are not changed by  $\Phi$ . Also, by construction, the monomials  $u_{j_1} u_{j_2} u_{j_3} u_{j_4}$  in  $\mathcal{H}_4$  with all integers in  $S$  are those for which  $j_1 + j_2 + j_3 + j_4 = 0$  and  $j_1^3 + j_2^3 + j_3^3 + j_4^3 = 0$ . By Lemma 4, we split

$$\sum_{\substack{j_1, j_2, j_3, j_4 \in S \\ j_1 + j_2 + j_3 + j_4 = 0 \\ j_1^3 + j_2^3 + j_3^3 + j_4^3 = 0}} u_{j_1} u_{j_2} u_{j_3} u_{j_4} = A_1 + A_2 + A_3$$

where  $A_1$  is given by the sum over  $j_1, j_2, j_3, j_4 \in S, j_1 + j_2 + j_3 + j_4 = 0$  with the restriction  $j_1 + j_2 = 0$ ,  $A_2$  with the restriction  $j_1 + j_2 \neq 0$  and  $j_1 + j_3 = 0$ , and  $A_3$  with the restriction  $j_1 + j_2 \neq 0, j_1 + j_3 \neq 0$  and  $j_2 + j_3 = 0$ . We get

$$A_2 = \sum_{\substack{j, j' \in S \\ j' \neq -j}} |u_j|^2 |u_{j'}|^2 = \sum_{j, j' \in S} |u_j|^2 |u_{j'}|^2 - \sum_{j \in S} |u_j|^4, \quad A_1 = \sum_{j, j' \in S} |u_j|^2 |u_{j'}|^2, \\ A_3 = \sum_{\substack{j, j' \in S \\ j' \neq \pm j}} |u_j|^2 |u_{j'}|^2 = \sum_{j, j' \in S} |u_j|^2 |u_{j'}|^2 - 2 \sum_{j \in S} |u_j|^4,$$

whence (3.8) follows.  $\square$

*Remark 1* In the Birkhoff normal form for the Hamiltonian  $K = H + \lambda M^2$  defined in (1.18), three additional terms appear in (3.8), which are

$$\lambda \sum_{j, j' \in S} |u_j|^2 |u_{j'}|^2 + 2\lambda M(v)M(z) + \lambda M^2(z).$$

Then in (3.8) the sum  $(\lambda - \frac{3\zeta}{4}) \sum_{j, j' \in S} |u_j|^2 |u_{j'}|^2$  vanishes if we choose  $\lambda := 3\zeta/4$ .

#### 4 Action-angle variables

We introduce action-angle variables on the tangential directions by the change of coordinates

$$u_j := \sqrt{\tilde{\xi}_j + |j|\tilde{y}_j} e^{i\tilde{\theta}_j} \quad \text{for } j \in S; \quad u_j := \tilde{z}_j \quad \text{for } j \in S^c, \quad (4.1)$$

where (recall that  $u_{-j} = \bar{u}_j$ )

$$\tilde{\xi}_{-j} = \tilde{\xi}_j, \quad \tilde{\xi}_j > 0, \quad \tilde{y}_{-j} = \tilde{y}_j, \quad \tilde{\theta}_{-j} = -\tilde{\theta}_j, \quad \tilde{\theta}_j, \tilde{y}_j \in \mathbb{R}, \quad \forall j \in S. \quad (4.2)$$

To simplify notation, for the tangential sites  $S^+ := \{\bar{j}_1, \dots, \bar{j}_\nu\}$  we also denote  $\tilde{\theta}_{\bar{j}_i} := \tilde{\theta}_i$ ,  $\tilde{y}_{\bar{j}_i} := \tilde{y}_i$ ,  $\tilde{\xi}_{\bar{j}_i} := \tilde{\xi}_i$ ,  $i = 1, \dots, \nu$ .

The symplectic 2-form  $\Omega$  in (3.2) (i.e. (1.6)) becomes

$$\mathcal{W} := \sum_{i=1}^{\nu} d\tilde{\theta}_i \wedge d\tilde{y}_i + \frac{1}{2} \sum_{j \in S^c \setminus \{0\}} \frac{1}{i\bar{j}} d\tilde{z}_j \wedge d\tilde{z}_{-j} = \left( \sum_{i=1}^{\nu} d\tilde{\theta}_i \wedge d\tilde{y}_i \right) \oplus \Omega_{S^\perp} = d\Lambda \quad (4.3)$$

where  $\Omega_{S^\perp}$  denotes the restriction of  $\Omega$  to  $H_S^\perp$  (see (1.20)) and  $\Lambda$  is the Liouville 1-form on  $\mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp$  defined by  $\Lambda_{(\tilde{\theta}, \tilde{y}, \tilde{z})} : \mathbb{R}^\nu \times \mathbb{R}^\nu \times H_S^\perp \rightarrow \mathbb{R}$ ,

$$\Lambda_{(\tilde{\theta}, \tilde{y}, \tilde{z})}[\widehat{\theta}, \widehat{y}, \widehat{z}] := -\tilde{y} \cdot \widehat{\theta} + \frac{1}{2} (\partial_x^{-1} \widehat{z}, \widehat{z})_{L^2(\mathbb{T})}. \quad (4.4)$$

We rescale the ‘‘unperturbed actions’’  $\tilde{\xi}$  and the variables  $\tilde{\theta}, \tilde{y}, \tilde{z}$  as

$$\tilde{\xi} = \varepsilon^2 \xi, \quad \tilde{y} = \varepsilon^{2b} y, \quad \tilde{z} = \varepsilon^b z, \quad b > 1, \quad (4.5)$$

where  $b > 1$  will be fixed below (see (5.9) and Remark 3). The symplectic 2-form in (4.3) transforms into  $\varepsilon^{2b} \mathcal{W}$ . Hence the Hamiltonian system generated by  $\mathcal{H}$  in (3.7) transforms into the new Hamiltonian system

$$\begin{cases} \dot{\theta} = \partial_y H_\varepsilon(\theta, y, z), \\ \dot{y} = -\partial_\theta H_\varepsilon(\theta, y, z), \\ \dot{z} = \partial_x \nabla_z H_\varepsilon(\theta, y, z), \end{cases} \quad H_\varepsilon := \varepsilon^{-2b} \mathcal{H} \circ A_\varepsilon, \quad (4.6)$$

where

$$A_\varepsilon(\theta, y, z) := \varepsilon v_\varepsilon(\theta, y) + \varepsilon^b z, \quad v_\varepsilon(\theta, y) := \sum_{j \in S} \sqrt{\tilde{\xi}_j + \varepsilon^{2(b-1)} |j| y_j} e^{i\theta_j} e^{ijx}. \quad (4.7)$$

We still denote by

$$X_{H_\varepsilon} = (\partial_y H_\varepsilon, -\partial_\theta H_\varepsilon, \partial_x \nabla_z H_\varepsilon)$$

the Hamiltonian vector field in the variables  $(\theta, y, z) \in \mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp$ .

We now write explicitly the Hamiltonian  $H_\varepsilon(\theta, y, z)$  defined in (4.6). Recall the expression of  $\mathcal{H}$  given in (3.7). The quadratic Hamiltonian  $H_2$  in (3.4) transforms into

$$\varepsilon^{-2b} H_2 \circ A_\varepsilon = \text{const} + \sum_{j \in S^+} j^3 y_j + \frac{1}{2} \int_{\mathbb{T}} z_x^2 dx, \quad (4.8)$$

and, by (3.8), (3.7) we get (writing, in short,  $v_\varepsilon := v_\varepsilon(\theta, y)$ )

$$\begin{aligned} H_\varepsilon(\theta, y, z) &= e(\xi) + \alpha(\xi) \cdot y + \frac{1}{2} \int_{\mathbb{T}} z_x^2 dx - \frac{3\zeta}{2} \varepsilon^2 \int_{\mathbb{T}} v_\varepsilon^2 z^2 dx \\ &\quad + 3\zeta \varepsilon^{2b} \left( \frac{1}{2} \sum_{j \in S^+} j^2 y_j^2 - \sum_{j, j' \in S^+} j y_j j' y_{j'} \right) - \zeta \varepsilon^{1+b} \int_{\mathbb{T}} v_\varepsilon z^3 dx \\ &\quad - \frac{\zeta}{4} \varepsilon^{2b} \int_{\mathbb{T}} z^4 dx + \varepsilon^{-2b} \mathcal{H}_{\geq 5}(\varepsilon v_\varepsilon(\theta, y) + \varepsilon^b z) \end{aligned} \quad (4.9)$$

where  $e(\xi)$  is a constant, and  $\alpha(\xi) \in \mathbb{R}^\nu$  is the vector of components

$$\alpha_i(\xi) := \bar{j}_i^3 + 3\zeta \varepsilon^2 [\xi_i - 2(\xi_1 + \dots + \xi_\nu)] \bar{j}_i, \quad i = 1, \dots, \nu.$$

This is the “frequency-to-amplitude” map which describes, at the main order, how the tangential frequencies are shifted by the amplitudes  $\xi := (\xi_1, \dots, \xi_\nu)$ . It can be written in compact form as

$$\alpha(\xi) := \bar{\omega} + \varepsilon^2 \mathbb{A} \xi, \quad \mathbb{A} := 3\zeta D_S (I - 2U), \quad (4.10)$$

where  $\bar{\omega} := (\bar{j}_1^3, \dots, \bar{j}_\nu^3) \in \mathbb{N}^\nu$  (see (1.19)) is the vector of the unperturbed linear frequencies of oscillations on the tangential sites,  $D_S$  is the diagonal matrix

$$D_S := \text{diag}(\bar{j}_1, \dots, \bar{j}_\nu) \in \text{Mat}(\nu \times \nu),$$

$I$  is the  $\nu \times \nu$  identity matrix, and  $U$  is the  $\nu \times \nu$  matrix with all entries equal to 1. The matrix  $\mathbb{A}$  is often called the “twist” matrix. It turns out to be invertible. Indeed, since  $U^2 = \nu U$ , one has  $(I - 2U)(I - \frac{2}{2\nu-1}U) = I$ , and therefore

$$\mathbb{A}^{-1} = \frac{1}{3\zeta} \left( I - \frac{2}{2\nu-1} U \right) D_S^{-1}. \quad (4.11)$$

With this notation, one can also write

$$\frac{1}{2} \sum_{j \in S^+} j^2 y_j^2 - \sum_{j, j' \in S^+} j y_j j' y_{j'} = \frac{1}{2} (I - 2U)(D_S y) \cdot (D_S y). \quad (4.12)$$

*Remark 2* By remark 1, for the Hamiltonian  $K = H + \lambda M^2$ ,  $\lambda := 3\zeta/4$ , defined in (1.18) the twist matrix in the frequency-amplitude relation (4.10) becomes  $\mathbb{A} = 3\zeta D_S$ , which is diagonal.

We write the Hamiltonian in (4.9) (eliminating the constant  $e(\xi)$  which is irrelevant for the dynamics) as  $H_\varepsilon = \mathcal{N} + P$ , where

$$\begin{aligned} \mathcal{N}(\theta, y, z) &= \alpha(\xi) \cdot y + \frac{1}{2} (N(\theta)z, z)_{L^2(\mathbb{T})}, \\ (N(\theta)z, z)_{L^2(\mathbb{T})} &:= \int_{\mathbb{T}} z_x^2 dx - 3\zeta \varepsilon^2 \int_{\mathbb{T}} v_\varepsilon^2(\theta, 0) z^2 dx, \end{aligned} \quad (4.13)$$

describes the linear dynamics, and  $P := H_\varepsilon - \mathcal{N}$ , namely

$$\begin{aligned} P &:= \frac{3\zeta}{2} \varepsilon^{2b} (I - 2U)(D_S y) \cdot (D_S y) - \frac{3\zeta}{2} \varepsilon^2 \int_{\mathbb{T}} [v_\varepsilon^2(\theta, y) - v_\varepsilon^2(\theta, 0)] z^2 dx \\ &\quad - \zeta \varepsilon^{1+b} \int_{\mathbb{T}} v_\varepsilon(\theta, y) z^3 dx - \frac{\zeta}{4} \varepsilon^{2b} \int_{\mathbb{T}} z^4 dx + \varepsilon^{-2b} \mathcal{H}_{\geq 5}(\varepsilon v_\varepsilon(\theta, y) + \varepsilon^b z), \end{aligned} \quad (4.14)$$

collects the nonlinear perturbative effects.

## 5 The nonlinear functional setting

We look for an embedded invariant torus

$$i : \mathbb{T}^\nu \rightarrow \mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp, \quad \varphi \mapsto i(\varphi) := (\theta(\varphi), y(\varphi), z(\varphi)) \quad (5.1)$$

of the Hamiltonian vector field  $X_{H_\varepsilon}$  filled by quasi-periodic solutions with diophantine frequency  $\omega \in \mathbb{R}^\nu$ , that we regard as independent parameters. We require that  $\omega$  belongs to the set

$$\Omega_\varepsilon := \alpha([1, 2]^\nu) = \{\alpha(\xi) : \xi \in [1, 2]^\nu\} \quad (5.2)$$

where  $\alpha$  is the affine diffeomorphism (4.10). Since any  $\omega \in \Omega_\varepsilon$  is  $\varepsilon^2$ -close to the integer vector  $\bar{\omega} \in \mathbb{N}^\nu$  (see (4.10), (1.19)), we require that the constant  $\gamma$  in the diophantine inequality

$$|\omega \cdot l| \geq \gamma \langle l \rangle^{-\tau}, \quad \forall l \in \mathbb{Z}^\nu \setminus \{0\}, \quad \text{satisfies } \gamma = \varepsilon^{2+a} \quad \text{for some } a > 0. \quad (5.3)$$

In (5.9) we will fix  $a \in (0, 1/6)$  (see also the discussion in Remark 3). Note that the definition of  $\gamma$  in (5.3) is slightly stronger than the minimal condition, which is  $\gamma \leq c\varepsilon^2$  with  $c$  small enough. We assume  $a > 0$  just for simplicity. In addition to (5.3) we shall also require that  $\omega$  satisfies the first and second order Melnikov-non-resonance conditions (8.63).

We fix the amplitude  $\xi$  as a function of  $\omega$  and  $\varepsilon$ , as

$$\xi := \varepsilon^{-2} \mathbb{A}^{-1}[\omega - \bar{\omega}], \quad (5.4)$$

so that  $\alpha(\xi) = \omega$  (see (4.10)).

Now we look for an embedded invariant torus of the modified Hamiltonian vector field  $X_{H_{\varepsilon, \zeta}} = X_{H_\varepsilon} + (0, \zeta, 0)$ ,  $\zeta \in \mathbb{R}^\nu$ , which is generated by the Hamiltonian

$$H_{\varepsilon, \zeta}(\theta, y, z) := H_\varepsilon(\theta, y, z) + \zeta \cdot \theta, \quad \zeta \in \mathbb{R}^\nu. \quad (5.5)$$

Note that the vector field  $X_{H_{\varepsilon, \zeta}}$  is periodic in  $\theta$  (unlike the Hamiltonian  $H_{\varepsilon, \zeta}$ ). We introduce  $\zeta$  in order to adjust the average in the second equation of the linearized system (6.22), see (6.23). The vector  $\zeta$  has however no dynamical consequences. Indeed it turns out that an invariant torus for the Hamiltonian vector field  $X_{H_{\varepsilon, \zeta}}$  is actually invariant for  $X_{H_\varepsilon}$  itself, see Lemma 6. Hence we look for zeros of the nonlinear operator

$$\begin{aligned} \mathcal{F}(i, \zeta) &:= \mathcal{F}(i, \zeta, \omega, \varepsilon) := \mathcal{D}_\omega i(\varphi) - X_{H_\varepsilon}(i(\varphi)) + (0, \zeta, 0) \\ &= \begin{pmatrix} \mathcal{D}_\omega \theta(\varphi) - \partial_y H_\varepsilon(i(\varphi)) \\ \mathcal{D}_\omega y(\varphi) + \partial_\theta H_\varepsilon(i(\varphi)) + \zeta \\ \mathcal{D}_\omega z(\varphi) - \partial_x \nabla_z H_\varepsilon(i(\varphi)) \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{D}_\omega \Theta(\varphi) - \partial_y P(i(\varphi)) \\ \mathcal{D}_\omega y(\varphi) + \frac{1}{2} \partial_\theta (N(\theta(\varphi))z(\varphi), z(\varphi))_{L^2(\mathbb{T})} + \partial_\theta P(i(\varphi)) + \zeta \\ \mathcal{D}_\omega z(\varphi) - \partial_x N(\theta(\varphi))z(\varphi) - \partial_x \nabla_z P(i(\varphi)) \end{pmatrix} \end{aligned} \quad (5.6)$$

where  $\Theta(\varphi) := \theta(\varphi) - \varphi$  is  $(2\pi)^\nu$ -periodic and we use (here and everywhere in the paper) the short notation

$$\mathcal{D}_\omega := \omega \cdot \partial_\varphi. \quad (5.7)$$

The Sobolev norm of the periodic component of the embedded torus

$$\mathfrak{J}(\varphi) := i(\varphi) - (\varphi, 0, 0) := (\Theta(\varphi), y(\varphi), z(\varphi)), \quad \Theta(\varphi) := \theta(\varphi) - \varphi, \quad (5.8)$$

is  $\|\mathfrak{J}\|_s := \|\Theta\|_{H_\varphi^s} + \|y\|_{H_\varphi^s} + \|z\|_s$  where  $\|z\|_s := \|z\|_{H_{\varphi,x}^s}$  is defined in (2.3). We link the rescaling (4.5) with the diophantine constant  $\gamma = \varepsilon^{2+a}$  by choosing

$$\gamma = \varepsilon^{2+a} = \varepsilon^{2b}, \quad b = 1 + (a/2), \quad a \in (0, 1/6). \quad (5.9)$$

Other choices are possible, see Remark 3.

**Theorem 2** *Let the tangential sites  $S$  in (1.11) satisfy (1.12). For all  $\varepsilon \in (0, \varepsilon_0)$ , where  $\varepsilon_0$  is small enough, there exist a constant  $C > 0$  and a Cantor-like set  $\mathcal{C}_\varepsilon \subset \Omega_\varepsilon$ , with asymptotically full measure as  $\varepsilon \rightarrow 0$ , namely*

$$\lim_{\varepsilon \rightarrow 0} \frac{|\mathcal{C}_\varepsilon|}{|\Omega_\varepsilon|} = 1, \quad (5.10)$$

such that, for all  $\omega \in \mathcal{C}_\varepsilon$ , there exists a solution  $i_\infty(\varphi) := i_\infty(\omega, \varepsilon)(\varphi)$  of the equation  $\mathcal{F}(i_\infty, 0, \omega, \varepsilon) = 0$  (the nonlinear operator  $\mathcal{F}(i, \zeta, \omega, \varepsilon)$  is defined in (5.6)). Hence the embedded torus  $\varphi \mapsto i_\infty(\varphi)$  is invariant for the Hamiltonian vector field  $X_{H_\varepsilon}$ , and it is filled by quasi-periodic solutions with frequency  $\omega$ . The torus  $i_\infty$  satisfies

$$\|i_\infty(\varphi) - (\varphi, 0, 0)\|_{s_0+\mu}^{\text{Lip}(\gamma)} \leq C\varepsilon^{5-2b}\gamma^{-1} = C\varepsilon^{1-2a} \quad (5.11)$$

for some  $\mu := \mu(\nu) > 0$ . Moreover, the torus  $i_\infty$  is linearly stable.

Theorem 2 is proved in sections 6-9. It implies Theorem 1 where the  $\xi_j$  in (1.13) are the components of the vector  $\mathbb{A}^{-1}[\omega - \bar{\omega}]$ . By (5.11), going back to the variables before the rescaling (4.5), we get  $\tilde{\Theta}_\infty = O(\varepsilon^{5-4b})$ ,  $\tilde{y}_\infty = O(\varepsilon^{5-2b})$ ,  $\tilde{z}_\infty = O(\varepsilon^{5-3b})$ .

*Remark 3* The way to link the amplitude-rescaling (4.5) with the diophantine constant  $\gamma = \varepsilon^{2+a}$  in (5.3) is not unique.

The choice  $\varepsilon^{2b} < \gamma$  (i.e. “ $b > 1$  large”) reduces to study the Hamiltonian  $H_\varepsilon$  in (4.9) as a perturbation of an isochronous system (as in [23], [25], [27]). We can take  $b = 4/3$  in order to minimize the size of the perturbation  $P = O(\varepsilon^{7/3})$ , estimating uniformly all the terms in the last two lines of (4.9). As a counterpart we have to regard in (4.9) the constants  $\alpha := \alpha(\xi) \in \mathbb{R}^\nu$  (or  $\xi$  in (4.7)) as independent variables. This is the perspective described for example in [10]. Then the Nash-Moser scheme produces iteratively a sequence of  $\xi_n = \xi_n(\omega)$  and embeddings  $\varphi \mapsto i_n(\varphi) := (\theta_n(\varphi), y_n(\varphi), z_n(\varphi))$  at the same time.

The case  $\varepsilon^{2b} > \gamma$  (i.e. “ $b \geq 1$  small”), in particular if  $b = 1$ , reduces to study the Hamiltonian  $H_\varepsilon$  in (4.9) as a perturbation of a non-isochronous system à la Arnold-Kolmogorov (note that the quadratic Hamiltonian in (4.12) satisfies the usual Kolmogorov non-degeneracy condition). In this case, the constant  $\xi_j$  in (4.7) and the average of  $|j|y_j(\varphi)$  have the same size and therefore the same role. Then we may consider  $\xi_j$  as fixed, and tune the average of the action component  $y_j(\varphi)$  in order to solve the linear equation (6.28), which corresponds to the angle component. We use the invertible (averaged) “twist”-matrix (6.30) to impose that the right hand side in (6.28) has zero average.

The intermediate case  $\varepsilon^{2b} = \gamma$ , adopted in this paper (as well as in [5]), has the advantage to avoid the introduction of the  $\xi(\omega)$  as an independent variable, but it also enables to estimate uniformly the sizes of the components of  $(\Theta(\varphi), y(\varphi), z(\varphi))$  with no distinctions.

Now we prove tame estimates for the composition operator induced by the Hamiltonian vector fields  $X_{\mathcal{N}}$  and  $X_P$  in (5.6), which are used in the next sections. Since the functions  $y \mapsto \sqrt{\xi + \varepsilon^{2(b-1)}}|j|y$ ,  $\theta \mapsto e^{i\theta}$  are analytic for  $\varepsilon$  small enough,  $j \in S$  and  $|y| \leq C$ , the composition Lemma 2 implies that, for all  $\Theta, y \in H^s(\mathbb{T}^\nu, \mathbb{R}^\nu)$  with  $\|\Theta\|_{s_0}, \|y\|_{s_0} \leq 1$ , setting  $\theta(\varphi) := \varphi + \Theta(\varphi)$ , one has the tame estimate

$$\|v_\varepsilon(\theta(\varphi), y(\varphi))\|_s \leq_s 1 + \|\Theta\|_s + \|y\|_s.$$

Hence the map  $A_\varepsilon$  in (4.7) satisfies, for all  $\|\mathcal{J}\|_{s_0}^{\text{Lip}(\gamma)} \leq 1$  (see (5.8))

$$\|A_\varepsilon(\theta(\varphi), y(\varphi), z(\varphi))\|_s^{\text{Lip}(\gamma)} \leq_s \varepsilon(1 + \|\mathcal{J}\|_s^{\text{Lip}(\gamma)}). \quad (5.12)$$

In the following lemma we collect tame estimates for the Hamiltonian vector fields  $X_{\mathcal{N}}, X_P, X_{H_\varepsilon}$  (see (4.13), (4.14)) whose proof is a direct application of classical tame product and composition estimates.

**Lemma 5** *Let  $\mathcal{J}(\varphi)$  in (5.8) satisfy  $\|\mathcal{J}\|_{s_0+3}^{\text{Lip}(\gamma)} \leq C\varepsilon^{5-2b}\gamma^{-1} = C\varepsilon^{5-4b}$ . Then, writing in short  $\|\cdot\|_s$  to indicate  $\|\cdot\|_s^{\text{Lip}(\gamma)}$ , one has*

$$\begin{aligned} \|\partial_y P(i)\|_s &\leq_s \varepsilon^3 + \varepsilon^{2b}\|\mathcal{J}\|_{s+3} & \|\partial_\theta P(i)\|_s &\leq_s \varepsilon^{5-2b}(1 + \|\mathcal{J}\|_{s+3}) \\ \|\nabla_z P(i)\|_s &\leq_s \varepsilon^{4-b} + \varepsilon^{6-3b}\|\mathcal{J}\|_{s+3} & \|X_P(i)\|_s &\leq_s \varepsilon^{5-2b} + \varepsilon^{2b}\|\mathcal{J}\|_{s+3} \\ \|\partial_\theta \partial_y P(i)\|_s &\leq_s \varepsilon^3 + \varepsilon^{5-2b}\|\mathcal{J}\|_{s+3} & \|\partial_y \nabla_z P(i)\|_s &\leq_s \varepsilon^{2+b} + \varepsilon^{2b}\|\mathcal{J}\|_{s+3} \\ & & \|\partial_{yy} P(i) - \varepsilon^{2b} \mathbb{A}D_S\|_s &\leq_s \varepsilon^{1+2b} + \varepsilon^3\|\mathcal{J}\|_{s+3} \end{aligned}$$

( $\mathbb{A}, D_S$  are defined in (4.10)) and, for all  $\widehat{v} := (\widehat{\Theta}, \widehat{y}, \widehat{z})$ ,

$$\|\partial_y d_i X_P(i)[\widehat{v}]\|_s \leq_s \varepsilon^{2b}(\|\widehat{v}\|_{s+3} + \|\mathcal{J}\|_{s+3}\|\widehat{v}\|_{s_0+3}) \quad (5.13)$$

$$\|d_i X_{H_\varepsilon}(i)[\widehat{v}] + (0, 0, \partial_{xxx}\widehat{z})\|_s \leq_s \varepsilon^2(\|\widehat{v}\|_{s+3} + \|\mathcal{J}\|_{s+3}\|\widehat{v}\|_{s_0+3}) \quad (5.14)$$

$$\|d_i^2 X_{H_\varepsilon}(i)[\widehat{v}, \widehat{v}]\|_s \leq_s \varepsilon^2(\|\widehat{v}\|_{s+3}\|\widehat{v}\|_{s_0+3} + \|\mathcal{J}\|_{s+3}\|\widehat{v}\|_{s_0+3}^2). \quad (5.15)$$

In the sequel we also use that, by the diophantine condition (5.3), the operator  $\mathcal{D}_\omega^{-1}$  (see (5.7)) is defined for all functions  $u$  with zero  $\varphi$ -average, and satisfies

$$\|\mathcal{D}_\omega^{-1}u\|_s \leq C\gamma^{-1}\|u\|_{s+\tau}, \quad \|\mathcal{D}_\omega^{-1}u\|_s^{\text{Lip}(\gamma)} \leq C\gamma^{-1}\|u\|_{s+2\tau+1}^{\text{Lip}(\gamma)}. \quad (5.16)$$

## 6 Approximate inverse

In order to implement a convergent Nash-Moser scheme that leads to a solution of  $\mathcal{F}(i, \zeta) = 0$ , we now construct an *approximate right inverse* (which satisfies tame estimates) of the linearized operator

$$d_{i, \zeta} \mathcal{F}(i_0, \zeta_0)[\widehat{v}, \widehat{\zeta}] = \mathcal{D}_\omega \widehat{v} - d_i X_{H_\varepsilon}(i_0(\varphi))[\widehat{v}] + (0, \widehat{\zeta}, 0), \quad (6.1)$$

see Theorem 3. Note that  $d_{i, \zeta} \mathcal{F}(i_0, \zeta_0)$  is independent of  $\zeta_0$  (see (5.6)).

The notion of approximate right inverse is introduced in [31]. It denotes a linear operator which is an *exact* right inverse at a solution  $(i_0, \zeta_0)$  of  $\mathcal{F}(i_0, \zeta_0) = 0$ . We implement the general strategy in [10] which reduces the search of an approximate right inverse of (6.1) to the search of an approximate inverse on the normal directions only.

It is well known that an invariant torus  $i_0$  with diophantine flow is isotropic (see e.g. [10]), namely the pull-back 1-form  $i_0^* \Lambda$  is closed, where  $\Lambda$  is the Liouville 1-form in (4.4). This is tantamount to say that the 2-form  $\mathcal{W}$  (see (4.3)) vanishes on the torus  $i_0(\mathbb{T}^\nu)$ , because  $i_0^* \mathcal{W} = i_0^* d\Lambda = di_0^* \Lambda$ . For an ‘‘approximately invariant’’ torus  $i_0$  the 1-form  $i_0^* \Lambda$  is only ‘‘approximately closed’’. In order to make this statement quantitative we consider

$$\begin{aligned} i_0^* \Lambda &= \sum_{k=1}^{\nu} a_k(\varphi) d\varphi_k, \\ a_k(\varphi) &:= -([\partial_\varphi \theta_0(\varphi)]^T y_0(\varphi))_k + \frac{1}{2}(\partial_{\varphi_k} z_0(\varphi), \partial_x^{-1} z_0(\varphi))_{L^2(\mathbb{T})} \end{aligned} \quad (6.2)$$

and we quantify how small is

$$i_0^* \mathcal{W} = di_0^* \Lambda = \sum_{1 \leq k < j \leq \nu} A_{kj}(\varphi) d\varphi_k \wedge d\varphi_j, \quad A_{kj} := \partial_{\varphi_k} a_j - \partial_{\varphi_j} a_k. \quad (6.3)$$

Along this section we will always assume the following hypothesis (which will be verified at each step of the Nash-Moser iteration):

• **ASSUMPTION.** The map  $\omega \mapsto i_0(\omega)$  is a Lipschitz function defined on some subset  $\Omega_o \subset \Omega_\varepsilon$ , where  $\Omega_\varepsilon$  is defined in (5.2), and, for some  $\mu := \mu(\tau, \nu) > 0$ ,

$$\begin{aligned} \|\mathfrak{J}_0\|_{s_0+\mu}^{\text{Lip}(\gamma)} &\leq C\varepsilon^{5-2b}\gamma^{-1} = C\varepsilon^{5-4b}, & \|Z\|_{s_0+\mu}^{\text{Lip}(\gamma)} &\leq C\varepsilon^{5-2b}, \\ \gamma &= \varepsilon^{2+a} = \varepsilon^{2b}, & b &:= 1 + (a/2), \quad a \in (0, 1/6), \end{aligned} \quad (6.4)$$

where  $\mathfrak{J}_0(\varphi) := i_0(\varphi) - (\varphi, 0, 0)$ , and

$$Z(\varphi) := (Z_1, Z_2, Z_3)(\varphi) := \mathcal{F}(i_0, \zeta_0)(\varphi) = \omega \cdot \partial_\varphi i_0(\varphi) - X_{H_{\varepsilon, \zeta_0}}(i_0(\varphi)) \quad (6.5)$$

is the ‘‘error’’ function.

**Lemma 6 (Lemma 6.1 in [5])**  $|\zeta_0|^{\text{Lip}(\gamma)} \leq C\|Z\|_{s_0}^{\text{Lip}(\gamma)}$ . If  $\mathcal{F}(i_0, \zeta_0) = 0$ , then  $\zeta_0 = 0$ , and the torus  $i_0(\varphi)$  is invariant for  $X_{H_\varepsilon}$ .

Now we estimate the size of  $i_0^* \mathcal{W}$  in terms of  $Z$ . From (6.2), (6.3) one has  $\|A_{kj}\|_s^{\text{Lip}(\gamma)} \leq_s \|\mathfrak{J}_0\|_{s+2}^{\text{Lip}(\gamma)}$ . Moreover,  $A_{kj}$  also satisfies the following bound.

**Lemma 7 (Lemma 6.2 in [5])** The coefficients  $A_{kj}(\varphi)$  in (6.3) satisfy

$$\|A_{kj}\|_s^{\text{Lip}(\gamma)} \leq_s \gamma^{-1} (\|Z\|_{s+2\tau+2}^{\text{Lip}(\gamma)} + \|Z\|_{s_0+1}^{\text{Lip}(\gamma)} \|\mathfrak{J}_0\|_{s+2\tau+2}^{\text{Lip}(\gamma)}). \quad (6.6)$$

As in [10], we first modify the approximate torus  $i_0$  to obtain an isotropic torus  $i_\delta$  which is still approximately invariant. We denote the Laplacian  $\Delta_\varphi := \sum_{k=1}^{\nu} \partial_{\varphi_k}^2$ .

**Lemma 8 (Isotropic torus)** The torus  $i_\delta(\varphi) := (\theta_0(\varphi), y_\delta(\varphi), z_0(\varphi))$  defined by

$$y_\delta := y_0 + [\partial_\varphi \theta_0(\varphi)]^{-T} \rho(\varphi), \quad \rho_j(\varphi) := \Delta_\varphi^{-1} \sum_{k=1}^{\nu} \partial_{\varphi_j} A_{kj}(\varphi) \quad (6.7)$$

is isotropic. If (6.4) holds, then, for some  $\sigma := \sigma(\nu, \tau)$ ,

$$\|y_\delta - y_0\|_s^{\text{Lip}(\gamma)} \leq_s \|\mathfrak{J}_0\|_{s+\sigma}^{\text{Lip}(\gamma)}, \quad (6.8)$$

$$\|y_\delta - y_0\|_s^{\text{Lip}(\gamma)} \leq_s \gamma^{-1} \{ \|Z\|_{s+\sigma}^{\text{Lip}(\gamma)} + \|Z\|_{s_0+\sigma}^{\text{Lip}(\gamma)} \|\mathfrak{J}_0\|_{s+\sigma}^{\text{Lip}(\gamma)} \}, \quad (6.9)$$

$$\|\mathcal{F}(i_\delta, \zeta_0)\|_s^{\text{Lip}(\gamma)} \leq_s \|Z\|_{s+\sigma}^{\text{Lip}(\gamma)} + \|\mathfrak{J}_0\|_{s+\sigma}^{\text{Lip}(\gamma)} \|Z\|_{s_0+\sigma}^{\text{Lip}(\gamma)}, \quad (6.10)$$

$$\|\partial_i [i_\delta][\hat{v}]\|_s \leq_s \|\hat{v}\|_s + \|\mathfrak{J}_0\|_{s+\sigma} \|\hat{v}\|_s. \quad (6.11)$$

In the paper we denote equivalently the differential by  $\partial_i$  or  $d_i$ . Moreover we denote by  $\sigma := \sigma(\nu, \tau)$  possibly different (larger) “loss of derivatives” constants.

*Proof* It is sufficient to closely follow the proof of Lemma 6.3 of [5]. We mention the only difference: equation (6.11) of [5] is  $\|\mathcal{F}(i_\delta, \zeta_0)\|_s^{\text{Lip}(\gamma)} \leq_s \|Z\|_{s+\sigma}^{\text{Lip}(\gamma)} + \varepsilon^{2b-1}\gamma^{-1}\|\mathfrak{J}_0\|_{s+\sigma}^{\text{Lip}(\gamma)}\|Z\|_{s_0+\sigma}^{\text{Lip}(\gamma)}$ , with a big factor  $\varepsilon^{2b-1}\gamma^{-1} = \varepsilon^{-1}$  more with respect to the present bound (6.10). In (6.10) there is no such a factor, because, by the estimates for  $\partial_\theta\partial_y P, \partial_{yy} P, \partial_y \nabla_z P$  in Lemma 5, here we have  $\|\partial_y X_P(i)\|_s \leq_s \varepsilon^{2b}(1 + \|\mathfrak{J}\|_{s+3})$ . Hence (6.8), (6.9), (6.4) imply that

$$\|X_P(i_\delta) - X_P(i_0)\|_s \leq_s \|Z\|_{s+\sigma} + \|\mathfrak{J}_0\|_{s+\sigma}\|Z\|_{s_0+\sigma}. \quad (6.12)$$

Then the proof goes on as in [5], without the large factor  $\varepsilon^{2b-1}\gamma^{-1}$ .  $\square$

In order to find an approximate inverse of the linearized operator  $d_{i,\zeta}\mathcal{F}(i_\delta)$  we introduce a suitable set of symplectic coordinates nearby the isotropic torus  $i_\delta$ . We consider the map  $G_\delta : (\psi, \eta, w) \rightarrow (\theta, y, z)$  of the phase space  $\mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp$  defined by

$$\begin{pmatrix} \theta \\ y \\ z \end{pmatrix} := G_\delta \begin{pmatrix} \psi \\ \eta \\ w \end{pmatrix} := \begin{pmatrix} \theta_0(\psi) \\ y_\delta(\psi) + [\partial_\psi \theta_0(\psi)]^{-T} \eta + [(\partial_\theta \tilde{z}_0)(\theta_0(\psi))]^T \partial_x^{-1} w \\ z_0(\psi) + w \end{pmatrix} \quad (6.13)$$

where  $\tilde{z}_0(\theta) := z_0(\theta_0^{-1}(\theta))$ . It is proved in [10] that  $G_\delta$  is symplectic, using that the torus  $i_\delta$  is isotropic (Lemma 8). In the new coordinates,  $i_\delta$  is the trivial embedded torus  $(\psi, \eta, w) = (\psi, 0, 0)$ . The transformed Hamiltonian  $K := K(\psi, \eta, w, \zeta_0)$  is (recall (5.5))

$$\begin{aligned} K &:= H_{\varepsilon, \zeta_0} \circ G_\delta \\ &= \theta_0(\psi) \cdot \zeta_0 + K_{00}(\psi) + K_{10}(\psi) \cdot \eta + (K_{01}(\psi), w)_{L^2(\mathbb{T})} + \frac{1}{2} K_{20}(\psi) \eta \cdot \eta \\ &\quad + (K_{11}(\psi) \eta, w)_{L^2(\mathbb{T})} + \frac{1}{2} (K_{02}(\psi) w, w)_{L^2(\mathbb{T})} + K_{\geq 3}(\psi, \eta, w) \end{aligned} \quad (6.14)$$

where  $K_{\geq 3}$  collects the terms at least cubic in the variables  $(\eta, w)$ . At any fixed  $\psi$ , the Taylor coefficient  $K_{00}(\psi) \in \mathbb{R}$ ,  $K_{10}(\psi) \in \mathbb{R}^\nu$ ,  $K_{01}(\psi) \in H_S^\perp$  (it is a function of  $x \in \mathbb{T}$ ),  $K_{20}(\psi)$  is a  $\nu \times \nu$  real matrix,  $K_{02}(\psi)$  is a linear self-adjoint operator of  $H_S^\perp$  and  $K_{11}(\psi) : \mathbb{R}^\nu \rightarrow H_S^\perp$ . Note that the above Taylor coefficients do not depend on the parameter  $\zeta_0$ .

The Hamilton equations associated to (6.14) are

$$\begin{cases} \dot{\psi} = K_{10}(\psi) + K_{20}(\psi) \eta + K_{11}^T(\psi) w + \partial_\eta K_{\geq 3}(\psi, \eta, w) \\ \dot{\eta} = -[\partial_\psi \theta_0(\psi)]^T \zeta_0 - \partial_\psi K_{00}(\psi) - [\partial_\psi K_{10}(\psi)]^T \eta - [\partial_\psi K_{01}(\psi)]^T w \\ \quad - \partial_\psi \left\{ \frac{1}{2} K_{20}(\psi) \eta \cdot \eta + (K_{11}(\psi) \eta, w)_{L^2(\mathbb{T})} + \frac{1}{2} (K_{02}(\psi) w, w)_{L^2(\mathbb{T})} \right. \\ \quad \left. + K_{\geq 3}(\psi, \eta, w) \right\} \\ \dot{w} = \partial_x (K_{01}(\psi) + K_{11}(\psi) \eta + K_{02}(\psi) w + \nabla_w K_{\geq 3}(\psi, \eta, w)) \end{cases} \quad (6.15)$$

where  $[\partial_\psi K_{10}(\psi)]^T$  is the  $\nu \times \nu$  transposed matrix and the operators  $[\partial_\psi K_{01}(\psi)]^T$  and  $K_{11}^T(\psi) : H_S^\perp \rightarrow \mathbb{R}^\nu$  are defined by the duality relation  $(\partial_\psi K_{01}(\psi)[\hat{\psi}], w)_{L^2}$



$= \hat{\psi} \cdot [\partial_\psi K_{01}(\psi)]^T w$ , for all  $\hat{\psi} \in \mathbb{R}^\nu$ ,  $w \in H_S^\perp$ , and similarly for  $K_{11}$ . Explicitly, for all  $w \in H_S^\perp$ , and denoting  $\underline{e}_k$  the  $k$ -th versor of  $\mathbb{R}^\nu$ ,

$$K_{11}^T(\psi)w = \sum_{k=1}^\nu (K_{11}^T(\psi)w \cdot \underline{e}_k) \underline{e}_k = \sum_{k=1}^\nu (w, K_{11}(\psi) \underline{e}_k)_{L^2(\mathbb{T})} \underline{e}_k \in \mathbb{R}^\nu.$$

In the next lemma we estimate the coefficients  $K_{00}, K_{10}, K_{01}$  of the Taylor expansion (6.14). Note that on an exact solution we have  $Z = 0$  and therefore  $K_{00}(\psi) = \text{const}$ ,  $K_{10} = \omega$  and  $K_{01} = 0$ .

**Lemma 9** *Assume (6.4). Then there is  $\sigma := \sigma(\tau, \nu)$  such that*

$$\|\partial_\psi K_{00}\|_s^{\text{Lip}(\gamma)}, \|K_{10} - \omega\|_s^{\text{Lip}(\gamma)}, \|K_{01}\|_s^{\text{Lip}(\gamma)} \leq_s \|Z\|_{s+\sigma}^{\text{Lip}(\gamma)} + \|Z\|_{s_0+\sigma}^{\text{Lip}(\gamma)} \|\mathfrak{J}_0\|_{s+\sigma}^{\text{Lip}(\gamma)}.$$

*Proof* Follow the proof of Lemma 6.4 in [5]. The fact that here there is no factor  $\varepsilon^{2b-1}\gamma^{-1}$  is a consequence of the better estimate (6.10) for  $\mathcal{F}(i_\delta, \zeta_0)$  compared to the analogous estimate in [5].  $\square$

*Remark 4* If  $\mathcal{F}(i_0, \zeta_0) = 0$  then  $\zeta_0 = 0$  by Lemma 6, and Lemma 9 implies that (6.14) simplifies to the normal form

$$K = \text{const} + \omega \cdot \eta + \frac{1}{2} K_{20}(\psi) \eta \cdot \eta + (K_{11}(\psi) \eta, w)_{L^2(\mathbb{T})} + \frac{1}{2} (K_{02}(\psi) w, w)_{L^2(\mathbb{T})} + K_{\geq 3}.$$

We now estimate  $K_{20}, K_{11}$  in (6.14). The norm of  $K_{20}$  is the sum of the norms of its matrix entries.

**Lemma 10** *Assume (6.4). Then*

$$\|K_{20} - \varepsilon^{2b} \mathbb{A} D_S\|_s^{\text{Lip}(\gamma)} \leq_s \varepsilon^{2b+1} + \varepsilon^{2b} \|\mathfrak{J}_0\|_{s+\sigma}^{\text{Lip}(\gamma)}, \quad (6.16)$$

$$\|K_{11} \eta\|_s^{\text{Lip}(\gamma)} \leq_s \varepsilon^{5-2b} \|\eta\|_s^{\text{Lip}(\gamma)} + \varepsilon^{2b} \|\mathfrak{J}_0\|_{s+\sigma}^{\text{Lip}(\gamma)} \|\eta\|_{s_0}^{\text{Lip}(\gamma)}, \quad (6.17)$$

$$\|K_{11}^T w\|_s^{\text{Lip}(\gamma)} \leq_s \varepsilon^{5-2b} \|w\|_{s+2}^{\text{Lip}(\gamma)} + \varepsilon^{2b} \|\mathfrak{J}_0\|_{s+\sigma}^{\text{Lip}(\gamma)} \|w\|_{s_0+2}^{\text{Lip}(\gamma)}. \quad (6.18)$$

*In particular*  $\|K_{20} - \varepsilon^{2b} \mathbb{A} D_S\|_{s_0}^{\text{Lip}(\gamma)} \leq C \varepsilon^{5-2b}$ , and

$$\|K_{11} \eta\|_{s_0}^{\text{Lip}(\gamma)} \leq C \varepsilon^{5-2b} \|\eta\|_{s_0}^{\text{Lip}(\gamma)}, \quad \|K_{11}^T w\|_{s_0}^{\text{Lip}(\gamma)} \leq C \varepsilon^{5-2b} \|w\|_{s_0+2}^{\text{Lip}(\gamma)}.$$

*Proof* See the proof of Lemma 6.6 in [5].  $\square$

Consider the linear change of variables  $(\hat{\theta}, \hat{y}, \hat{z}) = DG_\delta(\varphi, 0, 0)[\hat{\psi}, \hat{\eta}, \hat{w}]$ , where  $DG_\delta(\varphi, 0, 0)$  is obtained by linearizing  $G_\delta$  in (6.13) at  $(\varphi, 0, 0)$ , and it is represented by the matrix

$$DG_\delta(\varphi, 0, 0) = \begin{pmatrix} \partial_\psi \theta_0(\varphi) & 0 & 0 \\ \partial_\psi y_\delta(\varphi) & [\partial_\psi \theta_0(\varphi)]^{-T} & -[(\partial_\theta z_0)(\theta_0(\varphi))]^T \partial_x^{-1} \\ \partial_\psi z_0(\varphi) & 0 & I \end{pmatrix}. \quad (6.19)$$

The linearized operator  $d_{i,\zeta} \mathcal{F}(i_\delta, \zeta_0)$  transforms (approximately, see (6.40)) into the operator obtained linearizing (6.15) at  $(\psi, \eta, w, \zeta) = (\varphi, 0, 0, \zeta_0)$  (with  $\partial_t \rightsquigarrow \mathcal{D}_\omega$ ), which is the linear operator

$$B[\hat{\psi}, \hat{\eta}, \hat{w}, \hat{\zeta}] = \begin{pmatrix} B_1[\hat{\psi}, \hat{\eta}, \hat{w}, \hat{\zeta}] \\ B_2[\hat{\psi}, \hat{\eta}, \hat{w}, \hat{\zeta}] \\ B_3[\hat{\psi}, \hat{\eta}, \hat{w}, \hat{\zeta}] \end{pmatrix},$$

where

$$\begin{aligned}
B_1 &:= \mathcal{D}_\omega \widehat{\psi} - \partial_\psi K_{10}(\varphi)[\widehat{\psi}] - K_{20}(\varphi)\widehat{\eta} - K_{11}^T(\varphi)\widehat{w}, \\
B_2 &:= \mathcal{D}_\omega \widehat{\eta} + [\partial_\psi \theta_0(\varphi)]^T \widehat{\zeta} + \partial_\psi [\partial_\psi \theta_0(\varphi)]^T [\widehat{\psi}, \zeta_0] + \partial_{\psi\psi} K_{00}(\varphi)[\widehat{\psi}] \\
&\quad + [\partial_\psi K_{10}(\varphi)]^T \widehat{\eta} + [\partial_\psi K_{01}(\varphi)]^T \widehat{w}, \\
B_3 &:= \mathcal{D}_\omega \widehat{w} - \partial_x \{ \partial_\psi K_{01}(\varphi)[\widehat{\psi}] + K_{11}(\varphi)\widehat{\eta} + K_{02}(\varphi)\widehat{w} \}.
\end{aligned} \tag{6.20}$$

**Lemma 11 (Lemma 6.7 in [5])** *Assume (6.4) and let  $\widehat{v} := (\widehat{\psi}, \widehat{\eta}, \widehat{w})$ . Then*

$$\begin{aligned}
\|DG_\delta(\varphi, 0, 0)[\widehat{v}]\|_s + \|DG_\delta(\varphi, 0, 0)^{-1}[\widehat{v}]\|_s &\leq s \|\widehat{v}\|_s + \|\mathcal{J}_0\|_{s+\sigma} \|\widehat{v}\|_{s_0}, \\
\|D^2G_\delta(\varphi, 0, 0)[\widehat{v}_1, \widehat{v}_2]\|_s &\leq s \|\widehat{v}_1\|_s \|\widehat{v}_2\|_{s_0} + \|\widehat{v}_1\|_{s_0} \|\widehat{v}_2\|_s + \|\mathcal{J}_0\|_{s+\sigma} \|\widehat{v}_1\|_{s_0} \|\widehat{v}_2\|_{s_0}
\end{aligned} \tag{6.21}$$

for some  $\sigma := \sigma(\nu, \tau)$ . The same estimates hold for the  $\|\cdot\|_s^{\text{Lip}(\gamma)}$  norm.

In order to construct an approximate inverse of (6.20) it is sufficient to solve the equation

$$\mathbb{D}[\widehat{\psi}, \widehat{\eta}, \widehat{w}, \widehat{\zeta}] := \begin{pmatrix} \mathcal{D}_\omega \widehat{\psi} - K_{20}(\varphi)\widehat{\eta} - K_{11}^T(\varphi)\widehat{w} \\ \mathcal{D}_\omega \widehat{\eta} + [\partial_\psi \theta_0(\varphi)]^T \widehat{\zeta} \\ \mathcal{D}_\omega \widehat{w} - \partial_x K_{11}(\varphi)\widehat{\eta} - \partial_x K_{02}(\varphi)\widehat{w} \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} \tag{6.22}$$

which is obtained by neglecting in  $B_1, B_2, B_3$  in (6.20) the terms  $\partial_\psi K_{10}, \partial_{\psi\psi} K_{00}, \partial_\psi K_{00}, \partial_\psi K_{01}$  and  $\partial_\psi [\partial_\psi \theta_0(\varphi)]^T [\cdot, \zeta_0]$  (these terms are naught at a solution by Lemmata 9 and 6).

First we solve the second equation in (6.22), namely  $\mathcal{D}_\omega \widehat{\eta} = g_2 - [\partial_\psi \theta_0(\varphi)]^T \widehat{\zeta}$ . We choose  $\widehat{\zeta}$  so that the  $\varphi$ -average of the right hand side is zero, namely

$$\widehat{\zeta} = \langle g_2 \rangle \tag{6.23}$$

(we denote  $\langle g \rangle := (2\pi)^{-\nu} \int_{\mathbb{T}^\nu} g(\varphi) d\varphi$ ). Note that the  $\varphi$ -averaged matrix  $\langle [\partial_\psi \theta_0]^T \rangle = \langle I + [\partial_\psi \theta_0]^T \rangle = I$  because  $\theta_0(\varphi) = \varphi + \Theta_0(\varphi)$  and  $\Theta_0(\varphi)$  is a periodic function. Therefore

$$\widehat{\eta} := \mathcal{D}_\omega^{-1} (g_2 - [\partial_\psi \theta_0(\varphi)]^T \langle g_2 \rangle) + \langle \widehat{\eta} \rangle, \quad \langle \widehat{\eta} \rangle \in \mathbb{R}^\nu, \tag{6.24}$$

where the average  $\langle \widehat{\eta} \rangle$  will be fixed below. Then we consider the third equation

$$\mathcal{L}_\omega \widehat{w} = g_3 + \partial_x K_{11}(\varphi)\widehat{\eta}, \quad \mathcal{L}_\omega := \omega \cdot \partial_\varphi - \partial_x K_{02}(\varphi). \tag{6.25}$$

• **INVERSION ASSUMPTION.** *There exists a set  $\Omega_\infty \subset \Omega_o$  such that for all  $\omega \in \Omega_\infty$ , for every function  $g \in H_{S^\perp}^{s+\mu}(\mathbb{T}^{\nu+1})$  there exists a solution  $h := \mathcal{L}_\omega^{-1}g \in H_{S^\perp}^s(\mathbb{T}^{\nu+1})$  of the linear equation  $\mathcal{L}_\omega h = g$ , which satisfies*

$$\|\mathcal{L}_\omega^{-1}g\|_s^{\text{Lip}(\gamma)} \leq C(s)\gamma^{-1} (\|g\|_{s+\mu}^{\text{Lip}(\gamma)} + \varepsilon^2 \gamma^{-1} \|\mathcal{J}_0\|_{s+\mu}^{\text{Lip}(\gamma)} \|g\|_{s_0}^{\text{Lip}(\gamma)}) \tag{6.26}$$

for some  $\mu := \mu(\tau, \nu) > 0$ .

By the above assumption there exists a solution

$$\widehat{w} := \mathcal{L}_\omega^{-1} [g_3 + \partial_x K_{11}(\varphi)\widehat{\eta}] \tag{6.27}$$

of (6.25). Finally, we solve the first equation in (6.22), which, substituting (6.24), (6.27), becomes

$$\mathcal{D}_\omega \widehat{\psi} = g_1 + M_1(\varphi) \langle \widehat{\eta} \rangle + M_2(\varphi) g_2 + M_3(\varphi) g_3 - M_2(\varphi) [\partial_\psi \theta_0]^T \langle g_2 \rangle, \quad (6.28)$$

where

$$\begin{aligned} M_1(\varphi) &:= K_{20}(\varphi) + K_{11}^T(\varphi) \mathcal{L}_\omega^{-1} \partial_x K_{11}(\varphi), & M_2(\varphi) &:= M_1(\varphi) \mathcal{D}_\omega^{-1}, \\ M_3(\varphi) &:= K_{11}^T(\varphi) \mathcal{L}_\omega^{-1}. \end{aligned} \quad (6.29)$$

To solve equation (6.28) we have to choose  $\langle \widehat{\eta} \rangle$  such that the right hand side in (6.28) has zero average. By Lemma 10 and (6.4), the  $\varphi$ -averaged matrix

$$\langle M_1 \rangle = \varepsilon^{2b} \mathbb{A} D_S + O(\varepsilon^{5-2b}). \quad (6.30)$$

Therefore, for  $\varepsilon$  small,  $\langle M_1 \rangle$  is invertible and  $\langle M_1 \rangle^{-1} = O(\varepsilon^{-2b}) = O(\gamma^{-1})$  (recall (5.9)). Thus we define

$$\langle \widehat{\eta} \rangle := -\langle M_1 \rangle^{-1} [\langle g_1 \rangle + \langle M_2 g_2 \rangle + \langle M_3 g_3 \rangle - \langle M_2 [\partial_\psi \theta_0]^T \rangle \langle g_2 \rangle]. \quad (6.31)$$

With this choice of  $\langle \widehat{\eta} \rangle$ , equation (6.28) has the solution

$$\widehat{\psi} := \mathcal{D}_\omega^{-1} [g_1 + M_1(\varphi) \langle \widehat{\eta} \rangle + M_2(\varphi) g_2 + M_3(\varphi) g_3 - M_2(\varphi) [\partial_\psi \theta_0]^T \langle g_2 \rangle]. \quad (6.32)$$

In conclusion, we have constructed a solution  $(\widehat{\psi}, \widehat{\eta}, \widehat{w}, \widehat{\zeta})$  of the linear system (6.22).

**Proposition 2** *Assume (6.4) and (6.26). Then,  $\forall \omega \in \Omega_\infty$ ,  $\forall g := (g_1, g_2, g_3)$ , the system (6.22) has a solution  $\mathbb{D}^{-1} g := (\widehat{\psi}, \widehat{\eta}, \widehat{w}, \widehat{\zeta})$  where  $(\widehat{\psi}, \widehat{\eta}, \widehat{w}, \widehat{\zeta})$  are defined in (6.32), (6.24), (6.31), (6.27), (6.23), and satisfy*

$$\|\mathbb{D}^{-1} g\|_s^{\text{Lip}(\gamma)} \leq_s \gamma^{-1} (\|g\|_{s+\mu}^{\text{Lip}(\gamma)} + \varepsilon^2 \gamma^{-1} \|\mathfrak{J}_0\|_{s+\mu}^{\text{Lip}(\gamma)} \|g\|_{s_0+\mu}^{\text{Lip}(\gamma)}). \quad (6.33)$$

*Proof* Recalling (6.29), by Lemma 10, (6.26), (6.4) we get  $\|M_2 h\|_{s_0} + \|M_3 h\|_{s_0} \leq C \|h\|_{s_0+\sigma}$ . Then, by (6.31) and  $\langle M_1 \rangle^{-1} = O(\varepsilon^{-2b}) = O(\gamma^{-1})$ , we deduce  $|\langle \widehat{\eta} \rangle|_{s_0}^{\text{Lip}(\gamma)} \leq C \gamma^{-1} \|g\|_{s_0+\sigma}^{\text{Lip}(\gamma)}$  and (6.24), (5.16) imply  $\|\widehat{\eta}\|_s^{\text{Lip}(\gamma)} \leq_s \gamma^{-1} (\|g\|_{s+\sigma}^{\text{Lip}(\gamma)} + \|\mathfrak{J}_0\|_{s+\sigma} \|g\|_{s_0}^{\text{Lip}(\gamma)})$ . The bound (6.33) is sharp for  $\widehat{w}$  because  $\mathcal{L}_\omega^{-1} g_3$  in (6.27) is estimated using (6.26). Finally  $\widehat{\psi}$  satisfies (6.33) using (6.32), (6.29), (6.26), (5.16) and Lemma 10.  $\square$

Let  $\widetilde{G}_\delta(\psi, \eta, w, \zeta) := (G_\delta(\psi, \eta, w), \zeta)$ . Let  $\|(\psi, \eta, w, \zeta)\|_s^{\text{Lip}(\gamma)}$  denote the maximum between  $\|(\psi, \eta, w)\|_s^{\text{Lip}(\gamma)}$  and  $|\zeta|_{s_0}^{\text{Lip}(\gamma)}$ . We prove that the operator

$$\mathbf{T}_0 := (D\widetilde{G}_\delta)(\varphi, 0, 0) \circ \mathbb{D}^{-1} \circ (DG_\delta)(\varphi, 0, 0)^{-1} \quad (6.34)$$

is an approximate right inverse for  $d_{i,\zeta} \mathcal{F}(i_0)$ .

**Theorem 3 (Approximate inverse)** *Assume (6.4) and the inversion assumption (6.26). Then there exists  $\mu := \mu(\tau, \nu) > 0$  such that, for all  $\omega \in \Omega_\infty$ , for all  $g := (g_1, g_2, g_3)$ , the operator  $\mathbf{T}_0$  defined in (6.34) satisfies*

$$\|\mathbf{T}_0 g\|_s^{\text{Lip}(\gamma)} \leq_s \gamma^{-1} (\|g\|_{s+\mu}^{\text{Lip}(\gamma)} + \varepsilon^2 \gamma^{-1} \|\mathfrak{J}_0\|_{s+\mu}^{\text{Lip}(\gamma)} \|g\|_{s_0+\mu}^{\text{Lip}(\gamma)}). \quad (6.35)$$

The operator  $\mathbf{T}_0$  is an approximate inverse of  $d_{i,\zeta}\mathcal{F}(i_0)$ , namely

$$\begin{aligned} & \| (d_{i,\zeta}\mathcal{F}(i_0) \circ \mathbf{T}_0 - I)g \|_s^{\text{Lip}(\gamma)} \\ & \leq_s \gamma^{-1} \| \mathcal{F}(i_0, \zeta_0) \|_{s_0+\mu}^{\text{Lip}(\gamma)} \| g \|_{s+\mu}^{\text{Lip}(\gamma)} \\ & \quad + \gamma^{-1} \{ \| \mathcal{F}(i_0, \zeta_0) \|_{s+\mu}^{\text{Lip}(\gamma)} + \varepsilon^2 \gamma^{-1} \| \mathcal{F}(i_0, \zeta_0) \|_{s_0+\mu}^{\text{Lip}(\gamma)} \| \mathfrak{J}_0 \|_{s+\mu}^{\text{Lip}(\gamma)} \} \| g \|_{s_0+\mu}^{\text{Lip}(\gamma)}. \end{aligned} \quad (6.36)$$

*Proof* In this proof we denote  $\| \cdot \|_s$  instead of  $\| \cdot \|_s^{\text{Lip}(\gamma)}$ . The bound (6.35) follows from (6.34), (6.33), (6.21). By (5.6), since  $X_{\mathcal{N}}$  does not depend on  $y$ , and  $i_\delta$  differs from  $i_0$  only for the  $y$  component, we have

$$\begin{aligned} & d_{i,\zeta}\mathcal{F}(i_0)[\widehat{v}, \widehat{\zeta}] - d_{i,\zeta}\mathcal{F}(i_\delta)[\widehat{v}, \widehat{\zeta}] = d_i X_P(i_\delta)[\widehat{v}] - d_i X_P(i_0)[\widehat{v}] \\ & = \int_0^1 \partial_y d_i X_P(\theta_0, y_0 + s(y_\delta - y_0), z_0)[y_\delta - y_0, \widehat{v}] ds =: \mathcal{E}_0[\widehat{v}, \widehat{\zeta}]. \end{aligned} \quad (6.37)$$

By (5.13), (6.8), (6.9), (6.4), we estimate

$$\| \mathcal{E}_0[\widehat{v}, \widehat{\zeta}] \|_s \leq_s \| Z \|_{s_0+\sigma} \| \widehat{v} \|_{s+\sigma} + (\| Z \|_{s+\sigma} + \| Z \|_{s_0+\sigma} \| \mathfrak{J}_0 \|_{s+\sigma}) \| \widehat{v} \|_{s_0+\sigma} \quad (6.38)$$

where  $Z := \mathcal{F}(i_0, \zeta_0)$  (recall (6.5)). Note that  $\mathcal{E}_0[\widehat{v}, \widehat{\zeta}]$  is, in fact, independent of  $\widehat{\zeta}$ . Denote the set of variables  $(\psi, \eta, w) =: \mathbf{u}$ . Under the transformation  $G_\delta$ , the nonlinear operator  $\mathcal{F}$  in (5.6) transforms into

$$\mathcal{F}(G_\delta(\mathbf{u}(\varphi)), \zeta) = DG_\delta(\mathbf{u}(\varphi))(\mathcal{D}_\omega \mathbf{u}(\varphi) - X_K(\mathbf{u}(\varphi), \zeta)), \quad (6.39)$$

where  $K = H_{\varepsilon,\zeta} \circ G_\delta$ , see (6.14)-(6.15). Differentiating (6.39) at the trivial torus  $\mathbf{u}_\delta(\varphi) = G_\delta^{-1}(i_\delta)(\varphi) = (\varphi, 0, 0)$ , at  $\zeta = \zeta_0$ , in the direction  $(\widehat{u}, \widehat{\zeta}) = (DG_\delta(\mathbf{u}_\delta)^{-1}[\widehat{v}], \widehat{\zeta}) = D\widetilde{G}_\delta(\mathbf{u}_\delta)^{-1}[\widehat{v}, \widehat{\zeta}]$ , we get

$$d_{i,\zeta}\mathcal{F}(i_\delta)[\widehat{v}, \widehat{\zeta}] = DG_\delta(\mathbf{u}_\delta)(\mathcal{D}_\omega \widehat{u} - d_{\mathbf{u},\zeta} X_K(\mathbf{u}_\delta, \zeta_0)[\widehat{u}, \widehat{\zeta}]) + \mathcal{E}_1[\widehat{v}, \widehat{\zeta}], \quad (6.40)$$

$$\mathcal{E}_1[\widehat{v}, \widehat{\zeta}] := D^2 G_\delta(\mathbf{u}_\delta) [DG_\delta(\mathbf{u}_\delta)^{-1} \mathcal{F}(i_\delta, \zeta_0), DG_\delta(\mathbf{u}_\delta)^{-1}[\widehat{v}]], \quad (6.41)$$

where  $d_{\mathbf{u},\zeta} X_K(\mathbf{u}_\delta, \zeta_0)$  is expanded in (6.20). In fact,  $\mathcal{E}_1$  is independent of  $\widehat{\zeta}$ . We split

$$\mathcal{D}_\omega \widehat{u} - d_{\mathbf{u},\zeta} X_K(\mathbf{u}_\delta, \zeta_0)[\widehat{u}, \widehat{\zeta}] = \mathbb{D}[\widehat{u}, \widehat{\zeta}] + R_Z[\widehat{u}, \widehat{\zeta}],$$

where  $\mathbb{D}[\widehat{u}, \widehat{\zeta}]$  is defined in (6.22) and  $R_Z[\widehat{\psi}, \widehat{\eta}, \widehat{w}, \widehat{\zeta}]$  is defined by difference, so that its first component is  $-\partial_\psi K_{10}(\varphi)[\widehat{\psi}]$ , its second component is

$$\partial_\psi [\partial_\psi \theta_0(\varphi)]^T [\widehat{\psi}, \zeta_0] + \partial_{\psi\psi} K_{00}(\varphi)[\widehat{\psi}] + [\partial_\psi K_{10}(\varphi)]^T \widehat{\eta} + [\partial_\psi K_{01}(\varphi)]^T \widehat{w},$$

and its third component is  $-\partial_x \{ \partial_\psi K_{01}(\varphi)[\widehat{\psi}] \}$  (in fact,  $R_Z$  is independent of  $\widehat{\zeta}$ ). By (6.37) and (6.40),

$$\begin{aligned} d_{i,\zeta}\mathcal{F}(i_0) & = DG_\delta(\mathbf{u}_\delta) \circ \mathbb{D} \circ D\widetilde{G}_\delta(\mathbf{u}_\delta)^{-1} + \mathcal{E}_0 + \mathcal{E}_1 + \mathcal{E}_2, \\ \mathcal{E}_2 & := DG_\delta(\mathbf{u}_\delta) \circ R_Z \circ D\widetilde{G}_\delta(\mathbf{u}_\delta)^{-1}. \end{aligned} \quad (6.42)$$

By Lemmata 9, 11, 6, and (6.10), (6.4), the terms  $\mathcal{E}_1, \mathcal{E}_2$  satisfy the same bound (6.38) as  $\mathcal{E}_0$ . Thus the sum  $\mathcal{E} := \mathcal{E}_0 + \mathcal{E}_1 + \mathcal{E}_2$  satisfies (6.38). Applying  $\mathbf{T}_0$  defined in (6.34) to the right in (6.42), since  $\mathbb{D} \circ \mathbb{D}^{-1} = I$  (see Proposition 2), we get  $d_{i,\zeta}\mathcal{F}(i_0) \circ \mathbf{T}_0 - I = \mathcal{E} \circ \mathbf{T}_0$ . Then (6.36) follows from (6.35) and the bound (6.38) for  $\mathcal{E}$ .  $\square$

## 7 The linearized operator in the normal directions

The goal of this section is to write an explicit expression of the linearized operator  $\mathcal{L}_\omega$  defined in (6.25), see Proposition 3. To this aim, we compute  $\frac{1}{2}(K_{02}(\psi)w, w)_{L^2(\mathbb{T})}$ ,  $w \in H_S^\perp$ , which collects all the terms of  $(H_\varepsilon \circ G_\delta)(\psi, 0, w)$  that are quadratic in  $w$ , see (6.14). We first recall some preliminary lemmata.

**Lemma 12 (Lemma 7.1-[5])** *Let  $H$  be a Hamiltonian function of class  $C^2(H_0^1(\mathbb{T}_x), \mathbb{R})$  and consider a map  $\Phi(u) := u + \Psi(u)$  satisfying  $\Psi(u) = \Pi_E \Psi(\Pi_E u)$ , for all  $u$ , where  $E$  is a finite dimensional subspace as in (3.5). Then*

$$\partial_u[\nabla(H \circ \Phi)](u)[h] = (\partial_u \nabla H)(\Phi(u))[h] + \mathcal{R}(u)[h], \quad (7.1)$$

where  $\mathcal{R}(u)$  has the “finite dimensional” form

$$\mathcal{R}(u)[h] = \sum_{|j| \leq C} (h, g_j(u))_{L^2(\mathbb{T})} \chi_j(u) \quad (7.2)$$

with  $\chi_j(u) = e^{ijx}$  or  $g_j(u) = e^{ijx}$ . The remainder in (7.2) is  $\mathcal{R}(u) = \mathcal{R}_0(u) + \mathcal{R}_1(u) + \mathcal{R}_2(u)$  with

$$\begin{aligned} \mathcal{R}_0(u) &:= (\partial_u \nabla H)(\Phi(u)) \partial_u \Psi(u), & \mathcal{R}_1(u) &:= [\partial_u \{\Psi'(u)^T\}][\cdot, \nabla H(\Phi(u))], \\ \mathcal{R}_2(u) &:= [\partial_u \Psi(u)]^T (\partial_u \nabla H)(\Phi(u)) \partial_u \Phi(u). \end{aligned} \quad (7.3)$$

**Lemma 13 (Lemma 7.3 in [5])** *Let  $\mathcal{R}$  be an operator of the form*

$$\mathcal{R}h = \sum_{|j| \leq C} \int_0^1 (h, g_j(\tau))_{L^2(\mathbb{T})} \chi_j(\tau) d\tau, \quad (7.4)$$

where the functions  $g_j(\tau), \chi_j(\tau) \in H^s$ ,  $\tau \in [0, 1]$  depend in a Lipschitz way on the parameter  $\omega$ . Then its matrix  $s$ -decay norm (see (2.4)-(2.5)) satisfies

$$\|\mathcal{R}\|_s^{\text{Lip}(\gamma)} \leq_s \sum_{|j| \leq C} \sup_{\tau \in [0, 1]} (\|\chi_j(\tau)\|_s^{\text{Lip}(\gamma)} \|g_j(\tau)\|_{s_0}^{\text{Lip}(\gamma)} + \|\chi_j(\tau)\|_{s_0}^{\text{Lip}(\gamma)} \|g_j(\tau)\|_s^{\text{Lip}(\gamma)}).$$

### 7.1 Composition with the map $G_\delta$

In the sequel we use the fact that  $\mathfrak{J}_\delta := \mathfrak{J}_\delta(\varphi; \omega) := i_\delta(\varphi; \omega) - (\varphi, 0, 0)$  satisfies, by (6.8) and (6.4),

$$\|\mathfrak{J}_\delta\|_{s_0 + \mu}^{\text{Lip}(\gamma)} \leq C \varepsilon^{5-2b} \gamma^{-1} = C \varepsilon^{5-4b}. \quad (7.5)$$

In this section we study the Hamiltonian  $K := H_\varepsilon \circ G_\delta = \varepsilon^{-2b} \mathcal{H} \circ A_\varepsilon \circ G_\delta$  defined in (6.14), (4.6). Recalling (4.7), (6.13),  $A_\varepsilon \circ G_\delta$  has the form

$$A_\varepsilon(G_\delta(\psi, \eta, w)) = \varepsilon v_\varepsilon(\theta_0(\psi), y_\delta(\psi) + L_1(\psi)\eta + L_2(\psi)w) + \varepsilon^b(z_0(\psi) + w) \quad (7.6)$$

where  $v_\varepsilon$  is defined in (4.7), and

$$L_1(\psi) := [\partial_\psi \theta_0(\psi)]^{-T}, \quad L_2(\psi) := [(\partial_\theta \tilde{z}_0)(\theta_0(\psi))]^T \partial_x^{-1}. \quad (7.7)$$

By Taylor's formula, we expand (7.6) in  $w$  at  $(\eta, w) = (0, 0)$ , and we get

$$(A_\varepsilon \circ G_\delta)(\psi, 0, w) = T_\delta(\psi) + T_1(\psi)w + T_2(\psi)[w, w] + T_{\geq 3}(\psi, w),$$

where

$$T_\delta(\psi) := A_\varepsilon(G_\delta(\psi, 0, 0)) = \varepsilon v_\delta(\psi) + \varepsilon^b z_0(\psi), \quad v_\delta(\psi) := v_\varepsilon(\theta_0(\psi), y_\delta(\psi)) \quad (7.8)$$

is the approximate isotropic torus in the phase space  $H_0^1(\mathbb{T})$  (it corresponds to  $i_\delta$  in Lemma 8),

$$T_1(\psi)w := \varepsilon^{2b-1}U_1(\psi)w + \varepsilon^b w, \quad T_2(\psi)[w, w] := \varepsilon^{4b-3}U_2(\psi)[w, w]$$

$$U_1(\psi)w = \sum_{j \in S} \frac{|j|[L_2(\psi)w]_j e^{i\theta_0(\psi)_j}}{2\sqrt{\xi_j + \varepsilon^{2(b-1)}|j|[y_\delta(\psi)]_j}} e^{ijx}, \quad (7.9)$$

$$U_2(\psi)[w, w] = - \sum_{j \in S} \frac{j^2[L_2(\psi)w]_j^2 e^{i\theta_0(\psi)_j}}{8\{\xi_j + \varepsilon^{2(b-1)}|j|[y_\delta(\psi)]_j\}^{3/2}} e^{ijx}, \quad (7.10)$$

and  $T_{\geq 3}(\psi, w)$  collects all the terms of order at least cubic in  $w$ . The terms  $U_1, U_2 = O(1)$  in  $\varepsilon$ . Moreover, using that  $L_2(\psi)$  in (7.7) vanishes as  $z_0 = 0$ , they satisfy

$$\begin{aligned} \|U_1 w\|_s &\leq_s \|\mathcal{J}_\delta\|_s \|w\|_{s_0} + \|\mathcal{J}_\delta\|_{s_0} \|w\|_s, \\ \|U_2[w, w]\|_s &\leq_s \|\mathcal{J}_\delta\|_s \|\mathcal{J}_\delta\|_{s_0} \|w\|_{s_0}^2 + \|\mathcal{J}_\delta\|_{s_0}^2 \|w\|_{s_0} \|w\|_s \end{aligned} \quad (7.11)$$

and also in the  $\|\cdot\|_s^{\text{Lip}(\gamma)}$ -norm. We expand  $\mathcal{H}$  by Taylor's formula

$$\mathcal{H}(u + h) = \mathcal{H}(u) + ((\nabla \mathcal{H})(u), h)_{L^2(\mathbb{T})} + \frac{1}{2}((\partial_u \nabla \mathcal{H})(u)[h], h)_{L^2(\mathbb{T})} + O(h^3).$$

Specifying at  $u = T_\delta(\psi)$  and  $h = T_1(\psi)w + T_2(\psi)[w, w] + T_{\geq 3}(\psi, w)$ , we obtain that the sum of all the components of  $K = \varepsilon^{-2b}(\mathcal{H} \circ A_\varepsilon \circ G_\delta)(\psi, 0, w)$  that are quadratic in  $w$  is

$$\begin{aligned} \frac{1}{2}(K_{02}w, w)_{L^2(\mathbb{T})} &= \varepsilon^{-2b}((\nabla \mathcal{H})(T_\delta), T_2[w, w])_{L^2(\mathbb{T})} \\ &\quad + \varepsilon^{-2b} \frac{1}{2}((\partial_u \nabla \mathcal{H})(T_\delta)[T_1 w], T_1 w)_{L^2(\mathbb{T})}. \end{aligned}$$

Inserting the expressions (7.9), (7.10) in the last equality we get

$$\begin{aligned} K_{02}(\psi)w &= (\partial_u \nabla \mathcal{H})(T_\delta)[w] + 2\varepsilon^{b-1}(\partial_u \nabla \mathcal{H})(T_\delta)[U_1 w] \\ &\quad + \varepsilon^{2(b-1)}U_1^T(\partial_u \nabla \mathcal{H})(T_\delta)[U_1 w] + 2\varepsilon^{2b-3}U_2[w, \cdot]^T(\nabla \mathcal{H})(T_\delta). \end{aligned} \quad (7.12)$$

**Lemma 14** *The operator  $K_{02}$  reads*

$$(K_{02}(\psi)w, w)_{L^2(\mathbb{T})} = ((\partial_u \nabla \mathcal{H})(T_\delta)[w], w)_{L^2(\mathbb{T})} + (R(\psi)w, w)_{L^2(\mathbb{T})} \quad (7.13)$$

where  $R(\psi)w$  has the "finite dimensional" form

$$R(\psi)w = \sum_{|j| \leq C} (w, g_j(\psi))_{L^2(\mathbb{T})} \chi_j(\psi). \quad (7.14)$$

The functions  $g_j, \chi_j$  satisfy, for some  $\sigma := \sigma(\nu, \tau) > 0$ ,

$$\|g_j\|_s^{\text{Lip}(\gamma)} \|\chi_j\|_{s_0}^{\text{Lip}(\gamma)} + \|g_j\|_{s_0}^{\text{Lip}(\gamma)} \|\chi_j\|_s^{\text{Lip}(\gamma)} \leq_s \varepsilon^{b+1} \|\mathcal{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}, \quad (7.15)$$

$$\begin{aligned} \|\partial_i g_j[\hat{v}]\|_s \|\chi_j\|_{s_0} + \|\partial_i g_j[\hat{v}]\|_{s_0} \|\chi_j\|_s + \|g_j\|_{s_0} \|\partial_i \chi_j[\hat{v}]\|_s + \|g_j\|_s \|\partial_i \chi_j[\hat{v}]\|_{s_0} \\ \leq_s \varepsilon^{b+1} (\|\hat{v}\|_{s+\sigma} + \|\mathcal{J}_\delta\|_{s+\sigma} \|\hat{v}\|_{s_0+\sigma}), \end{aligned} \quad (7.16)$$

where  $i = (\theta, y, z)$  (see (5.1)) and  $\hat{v} = (\hat{\theta}, \hat{y}, \hat{z})$ .

*Proof* Since  $U_1 = \Pi_S U_1$  and  $U_2 = \Pi_S U_2$ , the last three terms in (7.12) have all the form (7.14). We have to prove that they are also small in size.

By (4.8), (6.13), (7.7), the only term in  $\varepsilon^{-2b} H_2(A_\varepsilon(G_\delta(\psi, \eta, w)))$  that is quadratic in  $w$  is  $\frac{1}{2} \int_{\mathbb{T}} w_x^2 dx$ , so this is the only contribution to (7.12) coming from  $H_2$ .

It remains to consider all the terms coming from  $\mathcal{H}_{\geq 4} := \mathcal{H}_4 + \mathcal{H}_{\geq 5} = O(u^4)$ . The term  $\varepsilon^{b-1} \partial_u \nabla \mathcal{H}_{\geq 4}(T_\delta) U_1$ , the term  $\varepsilon^{2(b-1)} U_1^T (\partial_u \nabla \mathcal{H}_{\geq 4})(T_\delta) U_1$  and the term  $\varepsilon^{2b-3} U_2^T \nabla \mathcal{H}_{\geq 4}(T_\delta)$  have all the form (7.14) and, using the inequality  $\|T_\delta\|_s^{\text{Lip}(\gamma)} \leq \varepsilon(1 + \|\mathcal{J}_\delta\|_s^{\text{Lip}(\gamma)})$ , (7.11) and (6.4), the bound (7.15) holds. By (6.11) and using explicit formulae (7.7)-(7.10) we get (7.16).  $\square$

The conclusion of this section is that, after the composition with the action-angle variables, the rescaling (4.5), and the transformation  $G_\delta$ , the linearized operator to analyze is  $w \mapsto (\partial_u \nabla \mathcal{H})(T_\delta)[w]$ ,  $w \in H_S^\perp$ , up to finite dimensional operators which have the form (7.14) and size (7.15).

## 7.2 The linearized operator in the normal directions

In view of (7.13) we now compute  $((\partial_u \nabla \mathcal{H})(T_\delta)[w], w)_{L^2(\mathbb{T})}$ ,  $w \in H_S^\perp$ , where  $\mathcal{H} = H \circ \Phi_B$  and  $\Phi_B$  is the Birkhoff map of Proposition 1. We recall that  $\Phi_B(u) = u + \Psi(u)$  where  $\Psi$  satisfies (3.6) and  $\Psi(u) = O(u^3)$ . It is convenient to estimate separately the terms in

$$\mathcal{H} = H \circ \Phi_B = H_2 \circ \Phi_B + H_4 \circ \Phi_B + H_{\geq 5} \circ \Phi_B \quad (7.17)$$

where  $H_2, H_4, H_{\geq 5}$  are defined in (3.4).

We first consider  $H_{\geq 5} \circ \Phi_B$ . By (3.4) we get  $\nabla H_{\geq 5}(u) = \pi_0[(\partial_u f)(x, u, u_x)] - \partial_x \{(\partial_{u_x} f)(x, u, u_x)\}$  where  $\pi_0$  is the operator defined in (1.32). Since  $\Phi_B$  has the form (3.6), Lemma 12 (at  $u = T_\delta$ , see (7.8)) implies that

$$\begin{aligned} \partial_u \nabla (H_{\geq 5} \circ \Phi_B)(T_\delta)[h] &= (\partial_u \nabla H_{\geq 5})(\Phi_B(T_\delta))[h] + \mathcal{R}_{H_{\geq 5}}(T_\delta)[h] \\ &= \partial_x (r_1(T_\delta) \partial_x h) + r_0(T_\delta) h + \mathcal{R}_{H_{\geq 5}}(T_\delta)[h] \end{aligned} \quad (7.18)$$

where the multiplicative functions  $r_0(T_\delta)$ ,  $r_1(T_\delta)$  are

$$\begin{aligned} r_0(T_\delta) &:= \sigma_0(\Phi_B(T_\delta)), \quad r_1(T_\delta) := \sigma_1(\Phi_B(T_\delta)), \\ \sigma_0(u) &:= (\partial_{uu} f)(x, u, u_x) - \partial_x \{(\partial_{u_x u} f)(x, u, u_x)\}, \\ \sigma_1(u) &:= -(\partial_{u_x u_x} f)(x, u, u_x), \end{aligned} \quad (7.19)$$

the remainder  $\mathcal{R}_{H_{\geq 5}}(u)$  has the form (7.2) with  $\chi_j = e^{ijx}$  or  $g_j = e^{ijx}$  and, using (7.3), it satisfies, for some  $\sigma := \sigma(\nu, \tau) > 0$ ,

$$\begin{aligned} \|g_j\|_s^{\text{Lip}(\gamma)} \|\chi_j\|_{s_0}^{\text{Lip}(\gamma)} + \|g_j\|_{s_0}^{\text{Lip}(\gamma)} \|\chi_j\|_s^{\text{Lip}(\gamma)} &\leq_s \varepsilon^5 (1 + \|\mathcal{J}_\delta\|_{s+2}^{\text{Lip}(\gamma)}), \\ \|\partial_i g_j[\tilde{\eta}]\|_s \|\chi_j\|_{s_0} + \|\partial_i g_j[\tilde{\eta}]\|_{s_0} \|\chi_j\|_s + \|g_j\|_{s_0} \|\partial_i \chi_j[\tilde{\eta}]\|_s + \|g_j\|_s \|\partial_i \chi_j[\tilde{\eta}]\|_{s_0} \\ &\leq_s \varepsilon^5 (\|\tilde{\eta}\|_{s+\sigma} + \|\mathcal{J}_\delta\|_{s+2} \|\tilde{\eta}\|_{s_0+2}). \end{aligned} \quad (7.20)$$

Now we consider the contributions from  $H_2 \circ \Phi_B$  and  $H_4 \circ \Phi_B$ . By Lemma 12 and the expressions of  $H_2, H_4$  in (3.4) we deduce that

$$\partial_u \nabla (H_2 \circ \Phi_B)(T_\delta)[h] = -\partial_{xx} h + \mathcal{R}_{H_2}(T_\delta)[h], \quad (7.21)$$

$$\partial_u \nabla (H_4 \circ \Phi_B)(T_\delta)[h] = -3\zeta(\Phi_B(T_\delta))^2 h + \mathcal{R}_{H_4}(T_\delta)[h], \quad (7.22)$$

where  $\mathcal{R}_{H_2}(u)$ ,  $\mathcal{R}_{H_4}(u)$  have the form (7.2). By (7.3), they have size  $\mathcal{R}_{H_2}(T_\delta) = O(\varepsilon^2)$ ,  $\mathcal{R}_{H_4}(T_\delta) = O(\varepsilon^4)$ . More precisely, the functions  $g_j, \chi_j$  in  $\mathcal{R}_{H_4}(T_\delta)$  satisfy the bounds in (7.20) with  $\varepsilon^5$  replaced by  $\varepsilon^4$ . Regarding  $\mathcal{R}_{H_2}(T_\delta)$ , we need to find an exact formula for the terms of order  $\varepsilon^2$ .

The sum of (7.18), (7.21) and (7.22) gives a formula for  $\partial_u \nabla \mathcal{H}(T_\delta)[h]$ , where the terms of form (7.2) and order  $\varepsilon^2$  are confined in  $\mathcal{R}_{H_2}(T_\delta)$ . On the other hand, recalling (3.7),  $\mathcal{H} = H_2 + \mathcal{H}_4 + \mathcal{H}_{\geq 5}$ , and  $\partial_u \nabla H_2(T_\delta) = -\partial_{xx}$ , while  $\partial_u \nabla \mathcal{H}_{\geq 5}(T_\delta) = O(\varepsilon^3)$ . Therefore all the terms of order  $\varepsilon^2$  in  $\partial_u \nabla \mathcal{H}(T_\delta)$  can only come from  $\partial_u \nabla \mathcal{H}_4(T_\delta)$ . Using formula (3.8) for  $\mathcal{H}_4$ , we calculate

$$\Pi_S^\perp(\partial_u \nabla \mathcal{H}_4(T_\delta)[h]) = -3\zeta \Pi_S^\perp(T_\delta^2 h) \quad \forall h \in H_{S^\perp}^s.$$

Hence all the terms of order  $\varepsilon^2$  in  $\Pi_S^\perp(\partial_u \nabla \mathcal{H}(T_\delta)[h])$  are contained in the term  $-3\zeta \Pi_S^\perp(T_\delta^2 h)$  (and the term  $-3\zeta \Pi_S^\perp(T_\delta^2 h)$  is included in  $-3\zeta \Pi_S^\perp[(\Phi_B(T_\delta))^2 h]$  because  $\Phi_B(T_\delta) = T_\delta + \Psi(T_\delta)$ ). As a consequence,  $\Pi_S^\perp \mathcal{R}_{H_2}(T_\delta)$  is of size  $O(\varepsilon^3)$ , and its functions  $g_j, \chi_j$  (see (7.2)) satisfy (7.20) with  $\varepsilon^5$  replaced by  $\varepsilon^3$ .

By Lemma 14 and the results of this section we deduce:

**Proposition 3** *Assume (7.5). Then the Hamiltonian operator  $\mathcal{L}_\omega$  has the form,  $\forall h \in H_{S^\perp}^s(\mathbb{T}^{\nu+1})$ ,*

$$\mathcal{L}_\omega h := \mathcal{D}_\omega h - \partial_x K_{02} h = \Pi_S^\perp(\mathcal{D}_\omega h + \partial_{xx}(a_1 \partial_x h) + \partial_x(a_0 h) - \partial_x \mathcal{R}_* h) \quad (7.23)$$

where  $\mathcal{R}_* := \mathcal{R}_{H_2}(T_\delta) + \mathcal{R}_{H_4}(T_\delta) + \mathcal{R}_{H_{\geq 5}}(T_\delta) + R(\psi)$  (with  $R(\psi)$  defined in Lemma 14, and  $\mathcal{R}_{H_2}(T_\delta)$ ,  $\mathcal{R}_{H_4}(T_\delta)$ ,  $\mathcal{R}_{H_{\geq 5}}(T_\delta)$  defined in (7.18), (7.21), (7.22)), the functions

$$a_1 := 1 - r_1(T_\delta), \quad a_0 := 3\zeta(\Phi_B(T_\delta))^2 - r_0(T_\delta), \quad (7.24)$$

$r_0, r_1$  are defined in (7.19), and  $T_\delta$  in (7.8). They satisfy

$$\|a_1 - 1\|_s^{\text{Lip}(\gamma)} + \|a_0 - 3\zeta T_\delta^2\|_s^{\text{Lip}(\gamma)} \leq \varepsilon^3(1 + \|\mathcal{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}), \quad (7.25)$$

$$\|\partial_i a_1[\tilde{v}]\|_s + \|\partial_i(a_0 - 3\zeta T_\delta^2)[\tilde{v}]\|_s \leq \varepsilon^3(\|\tilde{v}\|_{s+\sigma} + \|\mathcal{J}_\delta\|_{s+\sigma} \|\tilde{v}\|_{s_0+\sigma}) \quad (7.26)$$

where  $\mathcal{J}_\delta(\varphi) := (\theta_0(\varphi) - \varphi, y_\delta(\varphi), z_0(\varphi))$  corresponds to  $T_\delta$ . The remainder  $\mathcal{R}_*$  has the form (7.2), and its coefficients  $g_j, \chi_j$  satisfy bounds (7.15)-(7.16).

*Remark 5* For  $K = H + \lambda M^2$ ,  $\lambda = 3\zeta/4$ , the coefficient  $a_0$  in (7.24) becomes

$$a_0 = 3\zeta \pi_0 [(\Phi_B(T_\delta))^2] - r_0(T_\delta),$$

where  $\pi_0$  is defined in (1.32). Thus the space average of  $a_0$  has size  $O(\varepsilon^3)$ .

Bound (7.15) imply, by Lemma 13, estimates for the  $s$ -decay norms of  $\mathcal{R}_*$ . The linearized operator  $\mathcal{L}_\omega := \mathcal{L}_\omega(\omega, i_\delta(\omega))$  depends on the parameter  $\omega$  both directly and also through the dependence on the torus  $i_\delta(\omega)$ . We have estimated also the partial derivative  $\partial_i$  with respect to the variables  $i$  (see (5.1)) in order to control, along the nonlinear Nash-Moser iteration, the Lipschitz variation of the eigenvalues of  $\mathcal{L}_\omega$  with respect to  $\omega$  and the approximate solution  $i_\delta$ .



## 8 Reduction of the linearized operator in the normal directions

The goal of this section is to conjugate the Hamiltonian linear operator  $\mathcal{L}_\omega$  in (7.23) to the constant coefficients linear operator  $\mathcal{L}_\infty$  defined in (8.64). The proof is obtained applying different kind of symplectic transformations. We shall always assume (7.5).

### 8.1 Space reduction at the order $\partial_{xxx}$

As a first step, we symplectically conjugate the operator  $\mathcal{L}_\omega$  in (7.23) to  $\mathcal{L}_1$  in (8.13), which has the coefficient of  $\partial_{xxx}$  independent on the space variable. Because of the Hamiltonian structure, this step also eliminates the terms  $O(\partial_{xx})$ .

We look for a  $\varphi$ -dependent family of symplectic diffeomorphisms  $\Phi(\varphi)$  of  $H_S^\perp$  which differ from

$$\mathcal{A}_\perp := \Pi_S^\perp \mathcal{A} \Pi_S^\perp, \quad (\mathcal{A}h)(\varphi, x) := (1 + \beta_x(\varphi, x))h(\varphi, x + \beta(\varphi, x)), \quad (8.1)$$

up to a small “finite dimensional” remainder, see (8.3). For each  $\varphi \in \mathbb{T}^\nu$ , the map  $\mathcal{A}(\varphi)$  is a symplectic map of the phase space, see Remark 3.3 in [3]. If  $\|\beta\|_{W^{1,\infty}} < 1/2$ , then  $\mathcal{A}$  is invertible (see Lemma 3), and its inverse and adjoint maps are

$$\begin{aligned} (\mathcal{A}^{-1}h)(\varphi, y) &= (1 + \tilde{\beta}_y(\varphi, y))h(\varphi, y + \tilde{\beta}(\varphi, y)), \\ (\mathcal{A}^T h)(\varphi, y) &= h(\varphi, y + \tilde{\beta}(\varphi, y)) \end{aligned} \quad (8.2)$$

where  $x = y + \tilde{\beta}(\varphi, y)$  is the inverse diffeomorphism (of  $\mathbb{T}$ ) of  $y = x + \beta(\varphi, x)$ .

The restricted map  $\mathcal{A}_\perp(\varphi) : H_S^\perp \rightarrow H_S^\perp$  is not symplectic. We have already observed in the introduction that  $\mathcal{A}(\varphi)$  is the time-1 flow map of the linear Hamiltonian PDE (1.30). The equation (1.30) is a linear transport equation, whose characteristic curves are the solutions of the ODE

$$\frac{d}{d\tau} x = -b(\varphi, \tau, x).$$

To obtain a symplectic transformation close to  $\mathcal{A}_\perp$ , we define a symplectic map  $\Phi$  of  $H_S^\perp$  as the time 1 flow of the Hamiltonian PDE (1.31). The linear operator  $\Pi_S^\perp \partial_x (b(\tau, x)u)$  is the Hamiltonian vector field generated by the quadratic Hamiltonian  $\frac{1}{2} \int_{\mathbb{T}} b(\tau, x)u^2 dx$  restricted to  $H_S^\perp$ . The flow of (1.31) is well defined in the Sobolev spaces  $H_{S^\perp}^s(\mathbb{T}_x)$  for  $b(\varphi, \tau, x)$  smooth enough, by standard theory of linear hyperbolic PDEs (see e.g. section 0.8 in [29]). The difference between the time 1 flow map  $\Phi$  and  $\mathcal{A}_\perp$  is a “finite-dimensional” remainder of size  $O(\beta)$ .

**Lemma 15 (Lemma 8.1 of [5])** *For  $\|\beta\|_{W^{s_0+1,\infty}}$  small, there exists an invertible symplectic transformation  $\Phi = \mathcal{A}_\perp + \mathcal{R}_\Phi$  of  $H_{S^\perp}^s$ , where  $\mathcal{A}_\perp$  is defined in (8.1) and  $\mathcal{R}_\Phi$  is a “finite-dimensional” remainder*

$$\mathcal{R}_\Phi h = \sum_{j \in S} \int_0^1 (h, g_j(\tau))_{L^2(\mathbb{T})} \chi_j(\tau) d\tau + \sum_{j \in S} (h, \psi_j)_{L^2(\mathbb{T})} e^{ijx} \quad (8.3)$$

for some functions  $\chi_j(\tau), g_j(\tau), \psi_j \in H^s$  satisfying for all  $\tau \in [0, 1]$

$$\|\psi_j\|_s + \|g_j(\tau)\|_s \leq_s \|\beta\|_{W^{s+2,\infty}}, \quad \|\chi_j(\tau)\|_s \leq_s 1 + \|\beta\|_{W^{s+1,\infty}}. \quad (8.4)$$

Moreover

$$\|\Phi h\|_s + \|\Phi^{-1}h\|_s \leq_s \|h\|_s + \|\beta\|_{W^{s+2,\infty}} \|h\|_{s_0} \quad \forall h \in H_{S^\perp}^s. \quad (8.5)$$

We conjugate  $\mathcal{L}_\omega$  in (7.23) via the symplectic map  $\Phi = \mathcal{A}_\perp + \mathcal{R}_\Phi$  of Lemma 15. Using the splitting  $\Pi_S^\perp = I - \Pi_S$ , we compute

$$\mathcal{L}_\omega \Phi = \Phi \mathcal{D}_\omega + \Pi_S^\perp \mathcal{A}(b_3 \partial_{yyy} + b_2 \partial_{yy} + b_1 \partial_y + b_0) \Pi_S^\perp + \mathcal{R}_I, \quad (8.6)$$

where the coefficients  $b_i(\varphi, y)$ ,  $i = 0, 1, 2, 3$ , are

$$b_3 := \mathcal{A}^T [a_1(1 + \beta_x)^3], \quad b_2 := \mathcal{A}^T [2(a_1)_x(1 + \beta_x)^2 + 6a_1\beta_{xx}(1 + \beta_x)], \quad (8.7)$$

$$b_1 := \mathcal{A}^T \left[ (\mathcal{D}_\omega \beta) + \frac{3a_1\beta_{xx}^2}{1 + \beta_x} + 4a_1\beta_{xxx} + 6(a_1)_x\beta_{xx} + ((a_1)_{xx} + a_0)(1 + \beta_x) \right],$$

$$b_0 := \mathcal{A}^T \left[ \frac{1}{1 + \beta_x} (\mathcal{D}_\omega \beta_x + a_1\beta_{xxxx} + 2(a_1)_x\beta_{xxx} + ((a_1)_{xx} + a_0)\beta_{xx}) + (a_0)_x \right],$$

and the remainder

$$\begin{aligned} \mathcal{R}_I := & -\Pi_S^\perp (a_1 \partial_{xxx} + 2(a_1)_x \partial_{xx} + ((a_1)_{xx} + a_0) \partial_x + (a_0)_x) \Pi_S \mathcal{A} \Pi_S^\perp \\ & - \Pi_S^\perp \partial_x \mathcal{R}_* \mathcal{A}_\perp + [\mathcal{D}_\omega, \mathcal{R}_\Phi] + (\mathcal{L}_\omega - \mathcal{D}_\omega) \mathcal{R}_\Phi. \end{aligned} \quad (8.8)$$

The commutator  $[\mathcal{D}_\omega, \mathcal{R}_\Phi]$  has the form (8.3) with  $\mathcal{D}_\omega g_j$  or  $\mathcal{D}_\omega \chi_j$ ,  $\mathcal{D}_\omega \psi_j$  instead of  $\chi_j$ ,  $g_j$ ,  $\psi_j$  respectively. Also the last term  $(\mathcal{L}_\omega - \mathcal{D}_\omega) \mathcal{R}_\Phi$  in (8.8) has the form (8.3) (note that  $\mathcal{L}_\omega - \mathcal{D}_\omega$  does not contain derivatives with respect to  $\varphi$ ). By (8.6), and decomposing  $I = \Pi_S + \Pi_S^\perp$ , we get

$$\mathcal{L}_\omega \Phi = \Phi (\mathcal{D}_\omega + b_3 \partial_{yyy} + b_2 \partial_{yy} + b_1 \partial_y + b_0) \Pi_S^\perp + \mathcal{R}_{II}, \quad (8.9)$$

$$\mathcal{R}_{II} := \{ \Pi_S^\perp (\mathcal{A} - I) \Pi_S - \mathcal{R}_\Phi \} (b_3 \partial_{yyy} + b_2 \partial_{yy} + b_1 \partial_y + b_0) \Pi_S^\perp + \mathcal{R}_I. \quad (8.10)$$

Now we choose the function  $\beta = \beta(\varphi, x)$  such that

$$a_1(\varphi, x)(1 + \beta_x(\varphi, x))^3 = b_3(\varphi) \quad (8.11)$$

so that the coefficient  $b_3$  in (8.7) depends only on  $\varphi$  (note that  $\mathcal{A}^T [b_3(\varphi)] = b_3(\varphi)$ ). The only solution of (8.11) with zero space average is (see e.g. [3]-section 3.1)  $\beta := \partial_x^{-1} \rho_0$ , where  $\rho_0 := b_3(\varphi)^{1/3} (a_1(\varphi, x))^{-1/3} - 1$ , and

$$b_3(\varphi) = \left( \frac{1}{2\pi} \int_{\mathbb{T}} (a_1(\varphi, x))^{-1/3} dx \right)^{-3}. \quad (8.12)$$

Applying the symplectic map  $\Phi^{-1}$  in (8.9) we obtain the Hamiltonian operator (see Definition 2)

$$\mathcal{L}_1 := \Phi^{-1} \mathcal{L}_\omega \Phi = \Pi_S^\perp (\omega \cdot \partial_\varphi + b_3(\varphi) \partial_{yyy} + b_1 \partial_y + b_0) \Pi_S^\perp + \mathfrak{R}_1 \quad (8.13)$$

where  $\mathfrak{R}_1 := \Phi^{-1} \mathcal{R}_{II}$ . Note that the term  $b_2 \partial_{yy}$  has disappeared from (8.13) because, by the Hamiltonian nature of  $\mathcal{L}_1$ , the coefficient  $b_2 = 2(b_3)_y$  (see [3]-Remark 3.5) and therefore, by (8.12),  $b_2 = 2(b_3)_y = 0$ .

**Lemma 16 (Lemma 8.2 of [5])** *The operator  $\mathfrak{R}_1$  in (8.13) has the form (7.4).*

Since  $a_1 = 1 + O(\varepsilon^3)$  and  $a_0 = 3\zeta T_\delta^2 + O(\varepsilon^3)$  (see (7.25)-(7.26) for the precise estimates), by the usual composition estimates we deduce the following lemma.

**Lemma 17** *There is  $\sigma = \sigma(\tau, \nu) > 0$  such that*

$$\begin{aligned} & \|\beta\|_s^{\text{Lip}(\gamma)} + \|b_3 - 1\|_s^{\text{Lip}(\gamma)} + \|b_1 - 3\zeta T_\delta^2\|_s^{\text{Lip}(\gamma)} + \|b_0 - 3\zeta(T_\delta^2)_x\|_s^{\text{Lip}(\gamma)} \\ & \leq_s \varepsilon^3 (1 + \|\mathfrak{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}), \end{aligned} \quad (8.14)$$

$$\begin{aligned} & \|\partial_i \beta[\tilde{v}]\|_s + \|\partial_i b_3[\tilde{v}]\|_s + \|\partial_i(b_1 - 3\zeta T_\delta^2)[\tilde{v}]\|_s + \|\partial_i(b_0 - 3\zeta(T_\delta^2)_x)[\tilde{v}]\|_s \\ & \leq_s \varepsilon^3 (\|\tilde{v}\|_{s+\sigma} + \|\mathfrak{J}_\delta\|_{s+\sigma} \|\tilde{v}\|_{s_0+\sigma}), \end{aligned} \quad (8.15)$$

where  $T_\delta$  is defined in (7.8). The transformations  $\Phi$ ,  $\Phi^{-1}$  satisfy

$$\|\Phi h\|_s^{\text{Lip}(\gamma)} + \|\Phi^{-1} h\|_s^{\text{Lip}(\gamma)} \leq_s \|h\|_{s+1}^{\text{Lip}(\gamma)} + \|\mathfrak{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)} \|h\|_{s_0+1}^{\text{Lip}(\gamma)} \quad (8.16)$$

$$\begin{aligned} \|\partial_i(\Phi h)[\tilde{v}]\|_s + \|\partial_i(\Phi^{-1} h)[\tilde{v}]\|_s & \leq_s \|h\|_{s+\sigma} \|\tilde{v}\|_{s_0+\sigma} + \|h\|_{s_0+\sigma} \|\tilde{v}\|_{s+\sigma} \\ & \quad + \|\mathfrak{J}_\delta\|_{s+\sigma} \|h\|_{s_0+\sigma} \|\tilde{v}\|_{s_0+\sigma}. \end{aligned} \quad (8.17)$$

Moreover the remainder  $\mathfrak{R}_1$  has the form (7.4), where the functions  $\chi_j(\tau)$ ,  $g_j(\tau)$  satisfy the estimates (7.15)-(7.16) uniformly in  $\tau \in [0, 1]$ .

## 8.2 Time reduction at the order $\partial_{xxx}$

The goal of this section is to get a constant coefficient in front of  $\partial_{yyy}$ , using a quasi-periodic reparametrization of time. We consider the change of variable

$$(Bw)(\varphi, y) := w(\varphi + \omega\alpha(\varphi), y), \quad (B^{-1}h)(\vartheta, y) := h(\vartheta + \omega\tilde{\alpha}(\vartheta), y), \quad (8.18)$$

where  $\mathbb{T}^\nu \rightarrow \mathbb{T}^\nu$ ,  $\vartheta \mapsto \varphi = \vartheta + \omega\tilde{\alpha}(\vartheta)$  is the inverse diffeomorphism of  $\vartheta = \varphi + \omega\alpha(\varphi)$  in  $\mathbb{T}^\nu$ . By conjugation, the differential operators become

$$B^{-1}\omega \cdot \partial_\varphi B = \rho(\vartheta)\omega \cdot \partial_\vartheta, \quad B^{-1}\partial_y B = \partial_y, \quad \rho := B^{-1}(1 + \omega \cdot \partial_\varphi \alpha). \quad (8.19)$$

By (8.13), using also that  $B$  and  $B^{-1}$  commute with  $\Pi_S^\perp$ , the conjugate operator  $B^{-1}\mathcal{L}_1 B$  is equal to

$$\Pi_S^\perp [\rho\omega \cdot \partial_\vartheta + (B^{-1}b_3)\partial_{yyy} + (B^{-1}b_1)\partial_y + (B^{-1}b_0)]\Pi_S^\perp + B^{-1}\mathfrak{R}_1 B. \quad (8.20)$$

We choose  $\alpha$  such that  $(B^{-1}b_3)(\vartheta) = m_3\rho(\vartheta)$  for some constant  $m_3 \in \mathbb{R}$ , namely

$$b_3(\varphi) = m_3(1 + \omega \cdot \partial_\varphi \alpha(\varphi)) \quad (8.21)$$

(recall (8.19)). The unique solution with zero average of (8.21) is

$$\alpha(\varphi) := \frac{1}{m_3}(\omega \cdot \partial_\varphi)^{-1}(b_3 - m_3)(\varphi), \quad m_3 := \frac{1}{(2\pi)^\nu} \int_{\mathbb{T}^\nu} b_3(\varphi) d\varphi. \quad (8.22)$$

Hence, by (8.20),

$$B^{-1}\mathcal{L}_1 B = \rho\mathcal{L}_2, \quad \mathcal{L}_2 := \Pi_S^\perp (\omega \cdot \partial_\vartheta + m_3\partial_{yyy} + c_1\partial_y + c_0)\Pi_S^\perp + \mathfrak{R}_2 \quad (8.23)$$

$$c_1 := \rho^{-1}(B^{-1}b_1), \quad c_0 := \rho^{-1}(B^{-1}b_0), \quad \mathfrak{R}_2 := \rho^{-1}B^{-1}\mathfrak{R}_1 B. \quad (8.24)$$

The transformed operator  $\mathcal{L}_2$  in (8.23) is still Hamiltonian, because the reparametrization of time preserves the Hamiltonian structure (see Section 2.2 and Remark 3.7 in [3]).

**Lemma 18** *There is  $\sigma = \sigma(\nu, \tau) > 0$  (possibly larger than  $\sigma$  in Lemma 17) such that*

$$|m_3 - 1|^{\text{Lip}(\gamma)} \leq C\varepsilon^3, \quad |\partial_i m_3[\tilde{\eta}]| \leq C\varepsilon^3 \|\tilde{\eta}\|_{s_0+\sigma} \quad (8.25)$$

$$\|\alpha\|_s^{\text{Lip}(\gamma)} \leq_s \varepsilon^3 \gamma^{-1} (1 + \|\mathfrak{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)})$$

$$\|\partial_i \alpha[\tilde{\eta}]\|_s \leq_s \varepsilon^3 \gamma^{-1} (\|\tilde{\eta}\|_{s+\sigma} + \|\mathfrak{J}_\delta\|_{s+\sigma} \|\tilde{\eta}\|_{s_0+\sigma})$$

$$\|\rho - 1\|_s^{\text{Lip}(\gamma)} \leq_s \varepsilon^3 (1 + \|\mathfrak{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)})$$

$$\|\partial_i \rho[\tilde{\eta}]\|_s \leq_s \varepsilon^3 (\|\tilde{\eta}\|_{s+\sigma} + \|\mathfrak{J}_\delta\|_{s+\sigma} \|\tilde{\eta}\|_{s_0+\sigma})$$

$$\|c_1 - 3\varsigma T_\delta^2\|_s^{\text{Lip}(\gamma)} + \|c_0 - 3\varsigma(T_\delta^2)_x\|_s^{\text{Lip}(\gamma)} \leq_s \varepsilon^5 \gamma^{-1} (1 + \|\mathfrak{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}), \quad (8.26)$$

$$\|\partial_i(c_1 - 3\varsigma T_\delta^2)[\tilde{\eta}]\|_s + \|\partial_i(c_0 - 3\varsigma(T_\delta^2)_x)[\tilde{\eta}]\|_s$$

$$\leq_s \varepsilon^5 \gamma^{-1} (\|\tilde{\eta}\|_{s+\sigma} + \|\mathfrak{J}_\delta\|_{s+\sigma} \|\tilde{\eta}\|_{s_0+\sigma}).$$

The transformations  $B$ ,  $B^{-1}$  satisfy the estimates (8.16), (8.17). The remainder  $\mathfrak{R}_2$  has the form (7.4), and the functions  $g_j(\tau)$ ,  $\chi_j(\tau)$  satisfy the estimates (7.15)-(7.16) for all  $\tau \in [0, 1]$ .

*Proof* To estimate  $\|\alpha\|_s^{\text{Lip}(\gamma)}$  we also differentiate (8.22) with respect to the parameter  $\omega$ . Note that  $c_1 - 3\varsigma B^{-1}(T_\delta^2) = O(\varepsilon^3)$ , and similarly  $c_0 - 3\varsigma B^{-1}((T_\delta^2)_x) = O(\varepsilon^3)$ . The factor  $\varepsilon^5 \gamma^{-1}$  in the last two inequalities comes from the estimate of the difference  $B^{-1}(T_\delta^2) - T_\delta^2 \simeq (T_\delta^2)_\varphi \alpha = O(\varepsilon^2 \varepsilon^3 \gamma^{-1})$ .  $\square$

### 8.3 Translation of the space variable

In this section we remove the space average from the coefficient in front of  $\partial_y$ . Consider the change of the space variable  $z = y + p(\vartheta)$  which induces on  $H_{S^\perp}^{\nu+1}(\mathbb{T}^{\nu+1})$  the operators

$$(Tw)(\vartheta, y) := w(\vartheta, y + p(\vartheta)), \quad (T^{-1}h)(\vartheta, z) = h(\vartheta, z - p(\vartheta)) \quad (8.27)$$

(which are a particular case of those used in section 8.1). The differential operators become  $T^{-1}\omega \cdot \partial_\vartheta T = \omega \cdot \partial_\vartheta + \{\omega \cdot \partial_\vartheta p(\vartheta)\} \partial_z$ ,  $T^{-1}\partial_y T = \partial_z$ . Since  $T, T^{-1}$  commute with  $\Pi_S^\perp$ , we get

$$\mathcal{L}_3 := T^{-1}\mathcal{L}_2 T = \Pi_S^\perp (\omega \cdot \partial_\vartheta + m_3 \partial_{zzz} + d_1 \partial_z + d_0) \Pi_S^\perp + \mathfrak{R}_3, \quad (8.28)$$

$$d_1 := (T^{-1}c_1) + \omega \cdot \partial_\vartheta p, \quad d_0 := T^{-1}c_0, \quad \mathfrak{R}_3 := T^{-1}\mathfrak{R}_2 T. \quad (8.29)$$

We choose

$$m_1 := \frac{1}{(2\pi)^{\nu+1}} \int_{\mathbb{T}^{\nu+1}} c_1 d\vartheta dy, \quad p := (\omega \cdot \partial_\vartheta)^{-1} \left( m_1 - \frac{1}{2\pi} \int_{\mathbb{T}} c_1 dy \right), \quad (8.30)$$

so that

$$\frac{1}{2\pi} \int_{\mathbb{T}} d_1(\vartheta, z) dz = m_1 \quad \forall \vartheta \in \mathbb{T}^\nu. \quad (8.31)$$

Recalling (8.26), we analyze the space average of  $c_1$  in more detail. To avoid ambiguity between the space variable  $y \in \mathbb{T}$  and the action  $y_\delta : \mathbb{T}^\nu \rightarrow \mathbb{R}^\nu$  of (7.8),

we rename  $x \in \mathbb{T}$  the space variable, and  $\varphi \in \mathbb{T}^\nu$  the variable on the torus (time variable). Let

$$\bar{v}(\varphi, x) := \sum_{j \in S} \sqrt{\xi_j} e^{i\ell(j) \cdot \varphi} e^{ijx}, \quad (8.32)$$

where  $\ell : S \rightarrow \mathbb{Z}^\nu$  is the odd injective map (see (1.11))

$$\ell(\bar{j}_i) := e_i, \quad \ell(-\bar{j}_i) := -e_i, \quad i = 1, \dots, \nu \quad (8.33)$$

and  $e_i = (0, \dots, 1, \dots, 0)$  denotes the  $i$ -th vector of the canonical basis of  $\mathbb{R}^\nu$ . In view of the next linear Birkhoff normal form step (whose goal is to normalize the term of size  $\varepsilon^2$ ), we observe that the component of order  $\varepsilon^2$  in  $T_\delta^2$  (see (7.8)) is  $\varepsilon^2 \bar{v}^2$ , with

$$\begin{aligned} \|T_\delta^2 - \varepsilon^2 \bar{v}^2\|_s^{\text{Lip}(\gamma)} &\leq \varepsilon^2 \|\mathfrak{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}, \\ \|\partial_i(T_\delta^2 - \varepsilon^2 \bar{v}^2)[\tilde{v}]\|_s &\leq \varepsilon^2 (\|\tilde{v}\|_{s+\sigma} + \|\mathfrak{J}_\delta\|_{s+\sigma} \|\tilde{v}\|_{s_0+\sigma}). \end{aligned} \quad (8.34)$$

Moreover, from (7.8), since  $(v_\delta, z_0)_{L^2(\mathbb{T})} = 0$ , and  $(\theta_0)_{-j} = -(\theta_0)_j$  for all  $j \in S$ , we have

$$\int_{\mathbb{T}} T_\delta^2 dx = \varepsilon^2 \int_{\mathbb{T}} v_\delta^2 dx + \varepsilon^{2b} \int_{\mathbb{T}} z_0^2 dx = \varepsilon^2 \sum_{j \in S} \xi_j + \varepsilon^{2b} \sum_{j \in S} |j| (y_\delta)_j + \varepsilon^{2b} \int_{\mathbb{T}} z_0^2 dx.$$

We define

$$\tilde{d}_1 := d_1 - 3\varsigma \varepsilon^2 \bar{v}^2, \quad \tilde{d}_0 := d_0 - 3\varsigma \varepsilon^2 (\bar{v}^2)_x, \quad (8.35)$$

and note that, by (8.31) and (8.32),

$$\frac{1}{2\pi} \int_{\mathbb{T}} \tilde{d}_1 dx = m_1 - \frac{3\varsigma \varepsilon^2}{2\pi} \int_{\mathbb{T}} \bar{v}^2 dx = m_1 - \varepsilon^2 c(\xi), \quad c(\xi) := 3\varsigma \sum_{j \in S} \xi_j. \quad (8.36)$$

Using the explicit formulae above, and Lemma 13 for the estimate of  $\mathfrak{R}_3$ , we get the following bounds.

**Lemma 19** *There is  $\sigma := \sigma(\nu, \tau) > 0$  (possibly larger than in Lemma 18) such that*

$$\begin{aligned} |m_1 - \varepsilon^2 c(\xi)|^{\text{Lip}(\gamma)} &\leq C \varepsilon^5 \gamma^{-1}, \quad |\partial_i m_1[\tilde{v}]| \leq C \varepsilon^{2b} \|\tilde{v}\|_{s_0+\sigma} \\ \|p\|_s^{\text{Lip}(\gamma)} &\leq \varepsilon^5 \gamma^{-2} + \|\mathfrak{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}, \\ \|\partial_i p[\tilde{v}]\|_s &\leq \varepsilon^5 \gamma^{-2} \|\mathfrak{J}_\delta\|_{s+\sigma} \|\tilde{v}\|_{s_0+\sigma}, \\ \|\tilde{d}_k\|_s^{\text{Lip}(\gamma)} &\leq \varepsilon^7 \gamma^{-2} + \varepsilon^2 \|\mathfrak{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}, \quad k = 0, 1, \\ \|\partial_i \tilde{d}_k[\tilde{v}]\|_s &\leq \varepsilon^5 \gamma^{-1} (\|\tilde{v}\|_{s+\sigma} + \|\mathfrak{J}_\delta\|_{s+\sigma} \|\tilde{v}\|_{s_0+\sigma}), \quad k = 0, 1. \end{aligned} \quad (8.37)$$

The matrix  $s$ -decay norm (see (2.4)) of the operator  $\mathfrak{R}_3$  satisfies

$$\begin{aligned} |\mathfrak{R}_3|_s^{\text{Lip}(\gamma)} &\leq \varepsilon^{1+b} \|\mathfrak{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}, \\ |\partial_i \mathfrak{R}_3[\tilde{v}]|_s &\leq \varepsilon^{1+b} (\|\tilde{v}\|_{s+\sigma} + \|\mathfrak{J}_\delta\|_{s+\sigma} \|\tilde{v}\|_{s_0+\sigma}). \end{aligned} \quad (8.39)$$

The transformations  $\mathcal{T}, \mathcal{T}^{-1}$  satisfy (8.16), (8.17).

*Remark 6* When  $K = H + \lambda M^2$ ,  $\lambda = 3/4$ , the constant coefficient  $m_1$  in (8.30) becomes of size

$$|m_1|^{\text{Lip}(\gamma)} \leq C\varepsilon^5\gamma^{-1}. \quad (8.40)$$

The inequality (8.40) is the key difference between the cases  $H + (3\zeta/4)M^2$  and  $H$  (compare (8.40) with (8.37), where  $m_1$  contains the non-perturbative term  $\varepsilon^2 c(\xi)$ ).

It is sufficient to estimate  $\mathfrak{R}_3$  (which has the form (7.4)) only in the  $s$ -decay norm (see (8.39)) because the next transformations will preserve it. Such norms will be used in the reducibility scheme of section 8.6.

#### 8.4 Linear Birkhoff normal form

Now we normalize the terms of order  $\varepsilon^2$  of  $\mathcal{L}_3$ . This step is different from the reducibility steps that we shall perform in section 8.6: the diophantine constant  $\gamma$  in (5.3) is  $\gamma = o(\varepsilon^2)$ , and therefore the terms of order  $\varepsilon^2$  are not perturbative, because  $\varepsilon^2\gamma^{-1}$  is not small (in fact, it is big). The reduction of this section is possible thanks to the special form of the term  $\varepsilon^2\mathcal{B}$  defined in (8.41): the harmonics of  $\varepsilon^2\mathcal{B}$  corresponding to a possible small divisor are naught, except  $\mathcal{B}_j^j(0)$ , see Lemma 20. Note that, since the previous linear transformations  $\Phi, B, T$  are  $O(\varepsilon^5\gamma^{-2})$ -close to the identity, the terms of order  $\varepsilon^2$  in  $\mathcal{L}_3$  are the same as in the original linearized operator.

First, we collect all the terms of order  $\varepsilon^2$  in the operator  $\mathcal{L}_3$  in (8.28). We have

$$\mathcal{L}_3 = \Pi_S^\perp (\omega \cdot \partial_\varphi + m_3 \partial_{xxx} + \varepsilon^2 \mathcal{B} + \tilde{d}_1 \partial_x + \tilde{d}_0) \Pi_S^\perp + \mathfrak{R}_3$$

where  $\tilde{d}_1, \tilde{d}_0, \mathfrak{R}_3$  are defined in (8.35), (8.29) and (recall (8.32))

$$\mathcal{B}h := 3\zeta \bar{v}^2 \partial_x h + 3\zeta (\bar{v}^2)_x h = \partial_x (3\zeta \bar{v}^2 h). \quad (8.41)$$

Note that  $\mathcal{B}$  is the linear Hamiltonian vector field of  $H_S^\perp$  generated by the Hamiltonian  $z \mapsto \frac{3\zeta}{2} \int_{\mathbb{T}} \bar{v}^2 z^2 dx$ .

We transform  $\mathcal{L}_3$  by a symplectic operator  $\Phi_2 : H_{S^\perp}^s(\mathbb{T}^{\nu+1}) \rightarrow H_{S^\perp}^s(\mathbb{T}^{\nu+1})$  of the form

$$\Phi_2 := \exp(\varepsilon^2 A) = I_{H_S^\perp} + \varepsilon^2 A + \varepsilon^4 \hat{A}, \quad \hat{A} := \sum_{k \geq 2} \frac{\varepsilon^{2(k-2)}}{k!} A^k, \quad (8.42)$$

where  $A(\varphi)h = \sum_{j, j' \in S^c} A_j^{j'}(\varphi) h_{j'} e^{ijx}$  is a Hamiltonian vector field. The map  $\Phi_2$  is symplectic, because it is the time 1 flow of a Hamiltonian vector field. We calculate

$$\begin{aligned} \mathcal{L}_3 \Phi_2 - \Phi_2 \Pi_S^\perp (\mathcal{D}_\omega + m_3 \partial_{xxx}) \Pi_S^\perp \\ = \varepsilon^2 \Pi_S^\perp \{ \mathcal{B} + (\mathcal{D}_\omega A) + m_3 [\partial_{xxx}, A] \} \Pi_S^\perp + \Pi_S^\perp \tilde{d}_1 \partial_x \Pi_S^\perp + R_3 \end{aligned} \quad (8.43)$$

where

$$\begin{aligned} R_3 := \varepsilon^4 \Pi_S^\perp \{ (\mathcal{D}_\omega \hat{A}) + m_3 [\partial_{xxx}, \hat{A}] + \mathcal{B}(A + \varepsilon^2 \hat{A}) \} \Pi_S^\perp \\ + \Pi_S^\perp \tilde{d}_1 \partial_x \Pi_S^\perp (\Phi_2 - I) + (\Pi_S^\perp \tilde{d}_0 \Pi_S^\perp + \mathfrak{R}_3) \Phi_2. \end{aligned} \quad (8.44)$$

*Remark 7*  $R_3$  has no longer the form (7.4). However  $R_3 = O(\partial_x^0)$  because  $A = O(\partial_x^{-1})$  (see Lemma 22), and therefore  $\Phi_2 - I_{H_x^{\frac{1}{2}}} = O(\partial_x^{-1})$ . Moreover the matrix decay norm of  $R_3$  is  $o(\varepsilon^2)$ .

In order to normalize the term of order  $\varepsilon^2$  of (8.43), we expand  $A_j^{j'}(\varphi) = \sum_{l \in \mathbb{Z}^\nu} A_j^{j'}(l) e^{il \cdot \varphi}$ , and for each  $j, j' \in S^c$ ,  $l \in \mathbb{Z}^\nu$ , we choose

$$A_j^{j'}(l) := \begin{cases} -\frac{\mathcal{B}_j^{j'}(l)}{i(\omega \cdot l + m_3(j'^3 - j^3))} & \text{if } \bar{\omega} \cdot l + j'^3 - j^3 \neq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (8.45)$$

This definition is well posed. Indeed, by (8.41) and (8.32),

$$\mathcal{B}_j^{j'}(l) := 3\varsigma_{ij} \sum_{\substack{j_1, j_2 \in S \\ j_1 + j_2 = j - j' \\ \ell(j_1) + \ell(j_2) = l}} \sqrt{\xi_{j_1} \xi_{j_2}}. \quad (8.46)$$

In particular  $\mathcal{B}_j^{j'}(l) = 0$  unless  $|l| \leq 2$ . For  $|l| \leq 2$  and  $\bar{\omega} \cdot l + j'^3 - j^3 \neq 0$ , the denominators in (8.45) satisfy

$$\begin{aligned} |\omega \cdot l + m_3(j'^3 - j^3)| &= |m_3(\bar{\omega} \cdot l + j'^3 - j^3) + (\omega - m_3\bar{\omega}) \cdot l| \\ &\geq |m_3||\bar{\omega} \cdot l + j'^3 - j^3| - |\omega - m_3\bar{\omega}||l| \geq 1/2 \end{aligned} \quad (8.47)$$

for  $\varepsilon$  small, because  $|\bar{\omega} \cdot l + j'^3 - j^3| \geq 1$  ( $\bar{\omega} \cdot l + j'^3 - j^3$  is a nonzero integer),  $\omega = \bar{\omega} + O(\varepsilon^2)$  and by (8.25).

*Remark 8* The operator  $A$  defined in (8.45) is Hamiltonian, because  $\mathcal{B}$  is Hamiltonian. The reason is a general fact: the denominators  $\delta_{l,j,k} := i(\omega \cdot l + m_3(k^3 - j^3))$  satisfy  $\bar{\delta}_{l,j,k} = \delta_{-l,k,j}$  and an operator  $G(\varphi)$  is self-adjoint with respect to the  $L^2(\mathbb{T})$  scalar product if and only if its matrix elements satisfy  $\overline{G_j^k(l)} = G_k^j(-l)$ , see [3]-Remark 4.5. Alternatively, we could solve the homological equation of this Birkhoff step directly for the Hamiltonian function whose flow generates  $\Phi_2$ .

By the definition (8.45), the term of order  $\varepsilon^2$  in (8.43) is zero on the Fourier indices  $(l, j, j')$  such that  $\bar{\omega} \cdot l + j'^3 - j^3 \neq 0$ , while it is equal to  $\varepsilon^2 \mathcal{B}_j^{j'}(l)$  for  $(l, j, j')$  such that  $\bar{\omega} \cdot l + j'^3 - j^3 = 0$ . Now we prove that the only nonzero components of  $\mathcal{B}$  that remain in (8.43) are  $\mathcal{B}_j^j(0)$ .

**Lemma 20** *If  $\bar{\omega} \cdot l + j'^3 - j^3 = 0$  and  $\mathcal{B}_j^{j'}(l) \neq 0$ , then  $l = 0$  and  $j = j'$ .*

*Proof* If  $\mathcal{B}_j^{j'}(l) \neq 0$ , then, by (8.46), there exist  $j_1, j_2 \in S$  such that  $j_1 + j_2 = j - j'$  and  $\ell(j_1) + \ell(j_2) = l$ . Hence, recalling (1.19) and (8.33),

$$0 = \bar{\omega} \cdot l + j'^3 - j^3 = \bar{\omega} \cdot \ell(j_1) + \bar{\omega} \cdot \ell(j_2) + j'^3 - j^3 = j_1^3 + j_2^3 + j'^3 - j^3.$$

This equality, together with  $j_1 + j_2 + j' - j = 0$ , implies that  $(j_1 + j_2)(j_1 + j')(j_2 + j') = 0$  by Lemma 4. Since  $j_1, j_2 \in S$ ,  $j' \in S^c$ , the set  $S$  is symmetric, and  $0 \notin S$ , we deduce that the factors  $j_1 + j'$  and  $j_2 + j'$  are nonzero. Hence  $j_1 + j_2 = 0$ , and therefore  $l = \ell(j_1) + \ell(-j_1) = 0$ .  $\square$

Thus, the only nonzero term of order  $\varepsilon^2$  in (8.43) is  $\mathcal{B}_j^j(0)$ . By (8.46), we calculate  $\mathcal{B}_j^j(0) = ij c(\xi)$ , where  $c(\xi)$  is defined in (8.36). Hence, by (8.45), Lemma 20 and (8.36), the term of order  $\varepsilon^2$  in (8.43) is

$$\varepsilon^2 \Pi_S^\perp \{ \mathcal{B} + (\mathcal{D}_\omega A) + m_3 [\partial_{xxx}, A] \} \Pi_S^\perp = \varepsilon^2 c(\xi) \partial_x \Pi_S^\perp. \quad (8.48)$$

*Remark 9* When  $K = H + \lambda M^2$ ,  $\lambda = 3\zeta/4$ , the operator in (8.41) becomes  $\mathcal{B}h = \partial_x (3\zeta \pi_0 (\bar{v}^2) h)$ . Hence  $\mathcal{B}_j^j(0) = 0$ , and the right-hand side term in (8.48) is zero, namely the first step of linear Birkhoff normal form completely eliminates all the terms of order  $\varepsilon^2$ .

We now estimate the transformation  $A$ .

**Lemma 21** (i) For all  $l \in \mathbb{Z}^\nu$ ,  $j, j' \in S^c$ ,

$$|A_j^{j'}(l)| \leq C(|j| + |j'|)^{-1}, \quad |A_j^{j'}(l)|^{\text{lip}} \leq \varepsilon^{-2}(|j| + |j'|)^{-1}. \quad (8.49)$$

(ii)  $(A_1)_j^{j'}(l) = 0$  for all  $l \in \mathbb{Z}^\nu$ ,  $j, j' \in S^c$  such that  $|j - j'| > 2C_S$ , where  $C_S := \max\{|j| : j \in S\}$ .

*Proof* (i) As already observed, for all  $|l| > 2$  one has  $\mathcal{B}_j^{j'}(l) = 0$ , and therefore  $A_j^{j'}(l) = 0$ . For  $|l| \leq 2$ ,  $j \neq j'$ , one has (since  $|\omega| \leq |\bar{\omega}| + 1$ )

$$|\omega \cdot l + m_3(j'^3 - j^3)| \geq |m_3||j'^3 - j^3| - |\omega \cdot l| \geq \frac{1}{4}(j'^2 + j^2) - 2|\omega| \geq \frac{1}{8}(j'^2 + j^2)$$

for  $(j'^2 + j^2) \geq C$ , for some constant  $C$ . Since also (8.47) holds, we deduce that, for all  $j \neq j'$ ,

$$A_j^{j'}(l) \neq 0 \quad \Rightarrow \quad |\omega \cdot l + m_3(j'^3 - j^3)| \geq c(|j| + |j'|)^2. \quad (8.50)$$

On the other hand, if  $j = j' \in S^c$ , and  $l \neq 0$ , then  $\mathcal{B}_j^{j'}(l) = 0$ , and therefore  $A_j^{j'}(l) = 0$ . For  $j = j'$  and  $l = 0$  we also have  $A_j^{j'}(l) = 0$  because  $\bar{\omega} \cdot l + j'^3 - j^3 = 0$ . Hence (8.50) holds for all  $j, j'$ . By (8.45), (8.50), (8.46) we deduce the first bound in (8.49). The Lipschitz bound follows similarly (use also  $|j - j'| \leq 2C_S$ ). (ii) follows by (8.45)-(8.46).  $\square$

The previous lemma means that  $A = O(|\partial_x|^{-1})$ . More precisely, we deduce the following bound.

**Lemma 22** (Lemma 8.19 of [5])  $|A \partial_x|_s^{\text{Lip}(\gamma)} + |\partial_x A|_s^{\text{Lip}(\gamma)} \leq C(s)$ .

It follows that the symplectic map  $\Phi_2$  in (8.42) is invertible for  $\varepsilon$  small, with inverse

$$\begin{aligned} \Phi_2^{-1} &= \exp(-\varepsilon^2 A) = I_{H_S^\perp} + \varepsilon^2 \check{A}, \quad \check{A} := \sum_{n \geq 1} \frac{\varepsilon^{2n-2}}{n!} (-A)^n, \\ |\check{A} \partial_x|_s^{\text{Lip}(\gamma)} + |\partial_x \check{A}|_s^{\text{Lip}(\gamma)} &\leq C(s). \end{aligned} \quad (8.51)$$

By (8.43) and (8.48) we get the Hamiltonian operator

$$\mathcal{L}_4 := \Phi_2^{-1} \mathcal{L}_3 \Phi_2 = \Pi_S^\perp (\mathcal{D}_\omega + m_3 \partial_{xxx} + (\varepsilon^2 c(\xi) + \tilde{d}_1) \partial_x) \Pi_S^\perp + R_4, \quad (8.52)$$

$$R_4 := (\Phi_2^{-1} - I) \Pi_S^\perp (\varepsilon^2 c(\xi) + \tilde{d}_1) \partial_x \Pi_S^\perp + \Phi_2^{-1} R_3. \quad (8.53)$$



**Lemma 23** *There is  $\sigma = \sigma(\nu, \tau) > 0$  (possibly larger than in Lemma 19) such that*

$$\begin{aligned} |R_4|_s^{\text{Lip}(\gamma)} &\leq_s \varepsilon^7 \gamma^{-2} + \varepsilon^2 \|\mathcal{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}, \\ |\partial_i R_4[\tilde{v}]|_s &\leq_s \varepsilon^{1+b} \|\tilde{v}\|_{s+\sigma} + \varepsilon^2 \|\mathcal{J}_\delta\|_{s+\sigma} \|\tilde{v}\|_{s_0+\sigma}. \end{aligned} \quad (8.54)$$

*Proof* Use (8.44), (8.42), (8.38), (8.39), (8.25) and Lemma 22.  $\square$

### 8.5 Space reduction at the order $\partial_x$

The goal of this section is to transform  $\mathcal{L}_5$  in (8.52) so that the coefficient of  $\partial_x$  becomes constant. We conjugate  $\mathcal{L}_4$  via a symplectic map of the form

$$\mathcal{S} := \exp(\Pi_S^\perp(w\partial_x^{-1}))\Pi_S^\perp = \Pi_S^\perp(I + w\partial_x^{-1})\Pi_S^\perp + \widehat{\mathcal{S}}, \quad (8.55)$$

where  $\widehat{\mathcal{S}} := \sum_{k \geq 2} \frac{1}{k!} [\Pi_S^\perp(w\partial_x^{-1})]^k \Pi_S^\perp$  and  $w : \mathbb{T}^{\nu+1} \rightarrow \mathbb{R}$  is a function. Note that the linear operator  $\Pi_S^\perp(w\partial_x^{-1})\Pi_S^\perp$  is the Hamiltonian vector field generated by the Hamiltonian  $-\frac{1}{2} \int_{\mathbb{T}} w(\partial_x^{-1}h)^2 dx$ ,  $h \in H_S^\perp$ . We calculate

$$\begin{aligned} \mathcal{L}_4 \mathcal{S} - \mathcal{S} \Pi_S^\perp(\mathcal{D}_\omega + m_3 \partial_{xxx} + m_1 \partial_x) \Pi_S^\perp \\ = \Pi_S^\perp(3m_3 w_x + \varepsilon^2 c(\xi) + \tilde{d}_1 - m_1) \partial_x \Pi_S^\perp + \tilde{R}_5, \end{aligned}$$

$$\begin{aligned} \tilde{R}_5 := \Pi_S^\perp \{ &(3m_3 w_{xx} + (\varepsilon^2 c(\xi) + \tilde{d}_1 - m_1) \Pi_S^\perp w) \pi_0 \\ &+ ((\mathcal{D}_\omega w) + m_3 w_{xxx} + (\varepsilon^2 c(\xi) + \tilde{d}_1) \Pi_S^\perp w_x) \partial_x^{-1} \\ &+ (\mathcal{D}_\omega \widehat{\mathcal{S}}) + m_3 [\partial_{xxx}, \widehat{\mathcal{S}}] + (\varepsilon^2 c(\xi) + \tilde{d}_1) \partial_x \widehat{\mathcal{S}} - m_1 \widehat{\mathcal{S}} \partial_x + R_4 \mathcal{S} \} \Pi_S^\perp, \end{aligned}$$

where  $\tilde{R}_5$  collects all the terms of order at most  $\partial_x^0$ . By (8.36), we solve  $3m_3 w_x + \varepsilon^2 c(\xi) + \tilde{d}_1 - m_1 = 0$  by choosing  $w := -(3m_3)^{-1} \partial_x^{-1} (\varepsilon^2 c(\xi) + \tilde{d}_1 - m_1)$ . For  $\varepsilon$  small the operator  $\mathcal{S}$  is invertible, and we get

$$\mathcal{L}_5 := \mathcal{S}^{-1} \mathcal{L}_4 \mathcal{S} = \Pi_S^\perp(\mathcal{D}_\omega + m_3 \partial_{xxx} + m_1 \partial_x) \Pi_S^\perp + R_5, \quad R_5 := \mathcal{S}^{-1} \tilde{R}_5. \quad (8.56)$$

Since  $\mathcal{S}$  is symplectic,  $\mathcal{L}_5$  is Hamiltonian (recall Definition 2). By (8.38), (8.37), (8.25), one has  $\|w\|_s^{\text{Lip}(\gamma)} \leq_s \varepsilon^7 \gamma^{-2} + \varepsilon^2 \|\mathcal{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}$ .

**Lemma 24** *There is  $\sigma = \sigma(\nu, \tau) > 0$  (possibly larger than in Lemma 23) such that*

$$\begin{aligned} |\mathcal{S}^{\pm 1} - I|_s^{\text{Lip}(\gamma)} &\leq_s \varepsilon^7 \gamma^{-2} + \varepsilon^2 \|\mathcal{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}, \\ |\partial_i \mathcal{S}^{\pm 1}[\tilde{v}]|_s &\leq_s \varepsilon^{2b} \|\tilde{v}\|_{s+\sigma} + \varepsilon^5 \gamma^{-1} \|\mathcal{J}_\delta\|_{s+\sigma} \|\tilde{v}\|_{s_0+\sigma}. \end{aligned}$$

*The remainder  $R_5$  satisfies the same estimates (8.54) as  $R_4$ .*

8.6 KAM reducibility and inversion of  $\mathcal{L}_\omega$ 

The coefficients  $m_3, m_1$  of the operator  $\mathcal{L}_5$  in (8.56) are constants, and the remainder  $R_5$  is a bounded operator of order  $\partial_x^0$  with small matrix decay norm, see (8.59). Then we can diagonalize  $\mathcal{L}_5$  by applying the iterative KAM reducibility Theorem 4.2 in [3] along the sequence of scales

$$N_n := N_0^{\chi^n}, \quad n = 0, 1, 2, \dots, \quad \chi := 3/2, \quad N_0 > 0. \quad (8.57)$$

In section 9, the initial  $N_0$  will (slightly) increase to infinity as  $\varepsilon \rightarrow 0$ , see (9.5). The required smallness condition (see (4.14) in [3]) is (written in the present notations)

$$N_0^{C_0} |R_5|_{s_0+\beta}^{\text{Lip}(\gamma)} \gamma^{-1} \leq 1 \quad (8.58)$$

where  $\beta := 7\tau + 6$  (see (4.1) in [3]),  $\tau$  is the diophantine exponent in (5.3) and (8.63), and the constant  $C_0 := C_0(\tau, \nu) > 0$  is fixed in Theorem 4.2 in [3]. By Lemma 24, the remainder  $R_5$  satisfies the bound (8.54), and using (7.5) we get (recall (5.9))

$$|R_5|_{s_0+\beta}^{\text{Lip}(\gamma)} \leq C\varepsilon^7 \gamma^{-2} = C\varepsilon^{3-2a}, \quad |R_5|_{s_0+\beta}^{\text{Lip}(\gamma)} \gamma^{-1} \leq C\varepsilon^7 \gamma^{-3} = C\varepsilon^{1-3a}. \quad (8.59)$$

We use that  $\mu$  in (7.5) is assumed to satisfy  $\mu \geq \sigma + \beta$  where  $\sigma := \sigma(\tau, \nu)$  is given in Lemma 24.

**Theorem 4 (Reducibility)** *Assume that  $\omega \mapsto i_\delta(\omega)$  is a Lipschitz function defined on some subset  $\Omega_o \subset \Omega_\varepsilon$  (recall (5.2)), satisfying (7.5) with  $\mu \geq \sigma + \beta$ , where  $\sigma := \sigma(\tau, \nu)$  is given in Lemma 24 and  $\beta := 7\tau + 6$ . Then there exists  $\delta_0 \in (0, 1)$  such that, if*

$$N_0^{C_0} \varepsilon^7 \gamma^{-3} = N_0^{C_0} \varepsilon^{1-3a} \leq \delta_0, \quad \gamma := \varepsilon^{2b} := \varepsilon^{2+a}, \quad a \in (0, 1/6), \quad (8.60)$$

then:

(i) **(Eigenvalues)**. *For all  $\omega \in \Omega_\varepsilon$  there exists a sequence*

$$\mu_j^\infty(\omega) := \mu_j^\infty(\omega, i_\delta(\omega)) := i(-\tilde{m}_3(\omega)j^3 + \tilde{m}_1(\omega)j) + r_j^\infty(\omega), \quad j \in S^c, \quad (8.61)$$

where  $\tilde{m}_3, \tilde{m}_1$  coincide with the coefficients  $m_3, m_1$  of  $\mathcal{L}_5$  in (8.56) for all  $\omega \in \Omega_o$ , and

$$\begin{aligned} |\tilde{m}_3 - 1|^{\text{Lip}(\gamma)} &\leq C\varepsilon^3, & |\tilde{m}_1 - \varepsilon^2 c(\xi)|^{\text{Lip}(\gamma)} &\leq C\varepsilon^5 \gamma^{-1}, \\ |r_j^\infty|^{\text{Lip}(\gamma)} &\leq C\varepsilon^{3-2a} \quad \forall j \in S^c \end{aligned} \quad (8.62)$$

for some  $C > 0$  (and  $c(\xi)$  is defined in (8.36)). All the eigenvalues  $\mu_j^\infty$  are purely imaginary. We define, for convenience,  $\mu_0^\infty(\omega) := 0$ .

(ii) **(Conjugacy)**. *For all  $\omega$  in the set*

$$\begin{aligned} \Omega_\infty^{2\gamma} := \Omega_\infty^{2\gamma}(i_\delta) := \left\{ \omega \in \Omega_o : |i\omega \cdot l + \mu_j^\infty(\omega) - \mu_k^\infty(\omega)| \geq \frac{2\gamma|j^3 - k^3|}{\langle l \rangle^\tau} \right. \\ \left. \forall l \in \mathbb{Z}^\nu, \forall j, k \in S^c \cup \{0\} \right\} \end{aligned} \quad (8.63)$$

there is a real, bounded, invertible linear operator  $\Phi_\infty(\omega) : H_{S^\perp}^s(\mathbb{T}^{\nu+1}) \rightarrow H_{S^\perp}^s(\mathbb{T}^{\nu+1})$ , with bounded inverse  $\Phi_\infty^{-1}(\omega)$ , that conjugates  $\mathcal{L}_6$  in (8.56) to constant coefficients, namely

$$\begin{aligned}\mathcal{L}_\infty(\omega) &:= \Phi_\infty^{-1}(\omega) \circ \mathcal{L}_5(\omega) \circ \Phi_\infty(\omega) = \omega \cdot \partial_\varphi + \mathcal{D}_\infty(\omega), \\ \mathcal{D}_\infty(\omega) &:= \text{diag}_{j \in S^c} \{\mu_j^\infty(\omega)\}.\end{aligned}\tag{8.64}$$

The transformations  $\Phi_\infty, \Phi_\infty^{-1}$  are close to the identity in matrix decay norm, with

$$|\Phi_\infty - I|_{s, \Omega_\infty^{2\gamma}}^{\text{Lip}(\gamma)} + |\Phi_\infty^{-1} - I|_{s, \Omega_\infty^{2\gamma}}^{\text{Lip}(\gamma)} \leq_s \varepsilon^7 \gamma^{-3} + \varepsilon^2 \gamma^{-1} \|\mathfrak{J}_\delta\|_{s+\sigma}^{\text{Lip}(\gamma)}.\tag{8.65}$$

Moreover  $\Phi_\infty, \Phi_\infty^{-1}$  are symplectic, and  $\mathcal{L}_\infty$  is a Hamiltonian operator.

*Proof* The proof closely follows the one of Theorem 4.1 in [3], which is based on Theorem 4.2, Corollaries 4.1, 4.2 and Lemmata 4.1, 4.2 of [3]. Here  $\omega \in \mathbb{R}^\nu$ , while in [3] the parameter  $\lambda \in \mathbb{R}$ , but Kirszbraun's Theorem on Lipschitz extension also holds in  $\mathbb{R}^\nu$ . The bound (8.65) follows by Corollary 4.1 of [3] and the estimate of  $R_5$  in Lemma 24 above.

To adapt the proof of [3] to the present case, the only changes in the statement of Theorem 4.2 of [3] are:  $\varepsilon^{3-2a}$  instead of  $\varepsilon$  in (4.18) of [3], and  $\varepsilon^{1+b}$  instead of  $\varepsilon$  in (4.23), (4.25) and (4.26) of [3]. The factor  $\varepsilon^{1+b}$  comes from the bound for  $\partial_i R_5$ , see Lemma 24 and (8.54).  $\square$

*Remark 10* Theorem 4.2 in [3] also provides the Lipschitz dependence of the (approximate) eigenvalues  $\mu_j^n$  with respect to the unknown  $i_0(\varphi)$ , which is used for the measure estimate (Lemma 25).

All the parameters  $\omega \in \Omega_\infty^{2\gamma}$  satisfy (specialize (8.63) for  $k = 0$ )

$$|\omega \cdot l + \mu_j^\infty(\omega)| \geq 2\gamma |j|^3 \langle l \rangle^{-\tau}, \quad \forall l \in \mathbb{Z}^\nu, j \in S^c,\tag{8.66}$$

and the diagonal operator  $\mathcal{L}_\infty$  is invertible.

In the following theorem we verify the inversion assumption (6.26) for  $\mathcal{L}_\omega$ .

**Theorem 5 (Inversion of  $\mathcal{L}_\omega$ )** *Assume the hypotheses of Theorem 4 and (8.60). Then there exists  $\sigma_1 := \sigma_1(\tau, \nu) > 0$  such that,  $\forall \omega \in \Omega_\infty^{2\gamma}(i_\delta)$  (see (8.63)), for any function  $g \in H_{S^\perp}^{s+\sigma_1}(\mathbb{T}^{\nu+1})$  the equation  $\mathcal{L}_\omega h = g$  has a solution  $h = \mathcal{L}_\omega^{-1} g \in H_{S^\perp}^s(\mathbb{T}^{\nu+1})$ , satisfying*

$$\|\mathcal{L}_\omega^{-1} g\|_s^{\text{Lip}(\gamma)} \leq_s \gamma^{-1} (\|g\|_{s+\sigma_1}^{\text{Lip}(\gamma)} + \varepsilon^2 \gamma^{-1} \|\mathfrak{J}_0\|_{s+\sigma_1}^{\text{Lip}(\gamma)} \|g\|_{s_0}^{\text{Lip}(\gamma)}).\tag{8.67}$$

*Proof* See the proof of Theorem 8.16 in [5].  $\square$

## 9 The Nash-Moser nonlinear iteration

In this section we prove Theorem 2. It will be a consequence of the Nash-Moser Theorem 6 below.

Consider the finite-dimensional subspaces

$$E_n := \{\mathfrak{J}(\varphi) = (\Theta, y, z)(\varphi) : \Theta = \Pi_n \Theta, y = \Pi_n y, z = \Pi_n z\}$$

where  $N_n := N_0^{\chi^n}$  are introduced in (8.57), and  $\Pi_n$  are the projectors (which, with a small abuse of notation, we denote with the same symbol)

$$\Pi_n \Theta(\varphi) := \sum_{|l| < N_n} \Theta_l e^{i l \cdot \varphi}, \quad \Pi_n z(\varphi, x) := \sum_{|(l,j)| < N_n} z_{lj} e^{i(l \cdot \varphi + jx)}, \quad (9.1)$$

where  $\Theta(\varphi) = \sum_{l \in \mathbb{Z}^\nu} \Theta_l e^{i l \cdot \varphi}$  and  $z(\varphi, x) = \sum_{l \in \mathbb{Z}^\nu, j \in S^c} z_{lj} e^{i(l \cdot \varphi + jx)}$  (for  $\Pi_n y(\varphi)$  similar definition as for  $\Pi_n \Theta(\varphi)$ ). We define  $\Pi_n^\perp := I - \Pi_n$ . The classical smoothing properties hold: for all  $\alpha, s \geq 0$ ,

$$\begin{aligned} \|\Pi_n \mathfrak{J}\|_{s+\alpha}^{\text{Lip}(\gamma)} &\leq N_n^\alpha \|\mathfrak{J}\|_s^{\text{Lip}(\gamma)} \quad \forall \mathfrak{J}(\omega) \in H^s, \\ \|\Pi_n^\perp \mathfrak{J}\|_s^{\text{Lip}(\gamma)} &\leq N_n^{-\alpha} \|\mathfrak{J}\|_{s+\alpha}^{\text{Lip}(\gamma)} \quad \forall \mathfrak{J}(\omega) \in H^{s+\alpha}. \end{aligned} \quad (9.2)$$

We define the constants

$$\mu_1 := 3\mu + 9, \quad \alpha := 3\mu_1 + 1, \quad \alpha_1 := (\alpha - 3\mu)/2, \quad (9.3)$$

$$\kappa := 3(\mu_1 + \rho^{-1}) + 1, \quad \beta_1 := 6\mu_1 + 3\rho^{-1} + 3, \quad 0 < \rho < \frac{1-3a}{C_1(2+3a)}, \quad (9.4)$$

where  $\mu := \mu(\tau, \nu)$  is the ‘‘loss of regularity’’ defined in Theorem 3 (see (6.35)) and  $C_1$  is fixed below.

**Theorem 6 (Nash-Moser)** *Assume that  $f \in C^q$  with  $q > s_0 + \beta_1 + \mu + 3$ . Let  $\tau \geq \nu + 2$ . Then there exist  $C_1 > \max\{\mu_1 + \alpha, C_0\}$  (where  $C_0 := C_0(\tau, \nu)$  is the one in Theorem 4),  $\delta_0 := \delta_0(\tau, \nu) > 0$  such that, if*

$$N_0^{C_1} \varepsilon^{b_*+2} \gamma^{-2} < \delta_0, \quad \gamma := \varepsilon^{2+a} = \varepsilon^{2b}, \quad N_0 := (\varepsilon^4 \gamma^{-3})^\rho, \quad b_* := 5 - 2b, \quad (9.5)$$

then, for all  $n \geq 0$ :

(P1)<sub>n</sub> *there exists a function  $(\mathfrak{J}_n, \zeta_n) : \mathcal{G}_n \subseteq \Omega_\varepsilon \rightarrow E_{n-1} \times \mathbb{R}^\nu$ ,  $\omega \mapsto (\mathfrak{J}_n(\omega), \zeta_n(\omega))$ ,  $(\mathfrak{J}_0, \zeta_0) := 0$ ,  $E_{-1} := \{0\}$ , satisfying  $|\zeta_n|_{s_0}^{\text{Lip}(\gamma)} \leq C \|\mathcal{F}(U_n)\|_{s_0}^{\text{Lip}(\gamma)}$ ,*

$$\|\mathfrak{J}_n\|_{s_0+\mu}^{\text{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1}, \quad \|\mathcal{F}(U_n)\|_{s_0+\mu+3}^{\text{Lip}(\gamma)} \leq C_* \varepsilon^{b_*}, \quad (9.6)$$

where  $U_n := (i_n, \zeta_n)$  with  $i_n(\varphi) = (\varphi, 0, 0) + \mathfrak{J}_n(\varphi)$ . The sets  $\mathcal{G}_n$  are defined inductively by:

$$\begin{aligned} \mathcal{G}_0 &:= \left\{ \omega \in \Omega_\varepsilon : |\omega \cdot l| \geq \frac{2\gamma}{\langle l \rangle^\tau} \quad \forall l \in \mathbb{Z}^\nu \setminus \{0\} \right\}, \\ \mathcal{G}_{n+1} &:= \left\{ \omega \in \mathcal{G}_n : |\omega \cdot l + \mu_j^\infty(i_n) - \mu_k^\infty(i_n)| \geq \frac{2\gamma_n |j^3 - k^3|}{\langle l \rangle^\tau} \right. \\ &\quad \left. \forall j, k \in S^c \cup \{0\}, l \in \mathbb{Z}^\nu \right\}, \end{aligned} \quad (9.7)$$

where  $\gamma_n := \gamma(1 + 2^{-n})$  and  $\mu_j^\infty(\omega) := \mu_j^\infty(\omega, i_n(\omega))$  are defined in (8.61) (and  $\mu_0^\infty(\omega) = 0$ ).

The difference  $\widehat{\mathfrak{J}}_n := \mathfrak{J}_n - \mathfrak{J}_{n-1}$  (where we set  $\widehat{\mathfrak{J}}_0 := 0$ ) is defined on  $\mathcal{G}_n$ , and it satisfies

$$\|\widehat{\mathfrak{J}}_1\|_{s_0+\mu}^{\text{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1}, \quad \|\widehat{\mathfrak{J}}_n\|_{s_0+\mu}^{\text{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1} N_{n-1}^{-\alpha_1} \quad \forall n > 1. \quad (9.8)$$

- (P2)<sub>n</sub>  $\|\mathcal{F}(U_n)\|_{s_0}^{\text{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} N_{n-1}^{-\alpha}$  where we set  $N_{-1} := 1$ .  
(P3)<sub>n</sub> (High norms).  $\|\mathfrak{J}_n\|_{s_0+\beta_1}^{\text{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1} N_{n-1}^\kappa$  and  $\|\mathcal{F}(U_n)\|_{s_0+\beta_1}^{\text{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} N_{n-1}^\kappa$ .  
(P4)<sub>n</sub> (Measure). The measure of the ‘‘Cantor-like’’ sets  $\mathcal{G}_n$  satisfies

$$|\Omega_\varepsilon \setminus \mathcal{G}_0| \leq C_* \varepsilon^{2(\nu-1)} \gamma, \quad |\mathcal{G}_n \setminus \mathcal{G}_{n+1}| \leq C_* \varepsilon^{2(\nu-1)} \gamma N_{n-1}^{-1}. \quad (9.9)$$

All the Lip norms are defined on  $\mathcal{G}_n$ , namely  $\|\cdot\|_s^{\text{Lip}(\gamma)} = \|\cdot\|_{s, \mathcal{G}_n}^{\text{Lip}(\gamma)}$ .

*Proof* To simplify notations, in this proof we denote  $\|\cdot\|_s^{\text{Lip}(\gamma)}$  by  $\|\cdot\|$ .

STEP 1: *Proof of (P1, 2, 3)<sub>0</sub>*. Recalling (5.6) we have  $\|\mathcal{F}(U_0)\|_s = \|\mathcal{F}(\varphi, 0, 0, 0)\|_s = \|X_P(\varphi, 0, 0)\|_s \leq_s \varepsilon^{5-2b}$  by Lemma 5. Hence (recall that  $b_* := 5 - 2b$ ) the smallness conditions in (P1)<sub>0</sub>-(P3)<sub>0</sub> hold taking  $C_* := C_*(s_0 + \beta_1)$  large enough.

STEP 2: *Assume that (P1, 2, 3)<sub>n</sub> hold for some  $n \geq 0$ , and prove (P1, 2, 3)<sub>n+1</sub>*. The proof of this step closely follows Step 2 in the proof of Theorem 9.1 of [5]. We just mention the main changes: here it is convenient to define

$$w_n := \varepsilon^2 \gamma^{-2} \|\mathcal{F}(U_n)\|_{s_0}, \quad B_n := \varepsilon^2 \gamma^{-1} \|\mathfrak{J}_n\|_{s_0+\beta_1} + \varepsilon^2 \gamma^{-2} \|\mathcal{F}(U_n)\|_{s_0+\beta_1}, \quad (9.10)$$

while the corresponding quantities defined in (9.18) of [5] have  $\varepsilon$  instead of  $\varepsilon^2$  (and then, with definition (9.10), the bounds (9.19) of [5] are also valid here without changes). In the present case, the estimates (9.20)-(9.21) of [5] for the quadratic Taylor remainder have to be adapted by replacing the factor  $\varepsilon$  with  $\varepsilon^2$ . The reason for this improvement is that the nonlinearity in the mKdV equation is cubic, whereas in the KdV equation considered in [5] the nonlinearity is just quadratic.

*Remark 11* Since the KdV, respectively mKdV, nonlinearity is quadratic, respectively cubic, the smallness condition required in [5] for the convergence of the Nash-Moser scheme is stronger than for Theorem 6: it is  $\varepsilon \|\mathcal{F}(\varphi, 0, 0)\|_{s_0+\mu} \gamma^{-2} \ll 1$  instead of  $\varepsilon^2 \|\mathcal{F}(\varphi, 0, 0)\|_{s_0+\mu} \gamma^{-2} \ll 1$ . As a consequence less steps of Birkhoff normal form are required (namely less monomials to work out in the original Hamiltonian) to reach the sufficient smallness  $\mathcal{F}(U_0) = O(\varepsilon^{5-2b})$  to make the Nash-Moser scheme to converge (in [5] it is needed  $\mathcal{F}(U_0) = O(\varepsilon^{6-2b})$ ).

STEP 3: *Prove (P4)<sub>n</sub> for all  $n \geq 0$* . For all  $n \geq 0$ , the difference  $\mathcal{G}_n \setminus \mathcal{G}_{n+1}$  is the union over  $l \in \mathbb{Z}^\nu$ ,  $j, k \in S^c \cup \{0\}$  of the sets  $R_{ljk}(i_n)$ , where

$$R_{ljk}(i_n) := \{\omega \in \mathcal{G}_n : |\omega \cdot l + \mu_j^\infty(i_n) - \mu_k^\infty(i_n)| < 2\gamma_n |j^3 - k^3| \langle l \rangle^{-\tau}\}. \quad (9.11)$$

Since  $R_{ljk}(i_n) = \emptyset$  for  $j = k$ , in the sequel we assume that  $j \neq k$ .

**Lemma 25** *For  $n \geq 1$ ,  $|l| \leq N_{n-1}$ , one has the inclusion  $R_{ljk}(i_n) \subseteq R_{ljk}(i_{n-1})$ .*

*Proof* The proof closely follows the one of Lemma 5.2 in [3]. The differences are that here the vector  $\omega$  is not confined along a fixed direction, here we have  $N_{n-1}$  instead of  $N_n$ , and the factor  $\varepsilon$  in (5.28) and (5.33) of [3] is replaced here by  $\varepsilon^7 \gamma^{-2} = \varepsilon^{3-2a}$ .

In the proof we use (9.8), (8.59), (8.25), (8.37), and the bounds (4.25), (4.26), (4.34) of [3] adapted to the present case (the bounds (4.25), (4.26) of [3] hold here with  $\varepsilon^{1+b}$  instead of  $\varepsilon$ , as already pointed out in the proof of Theorem 4; the bound (4.34) of [3] holds here with no change).  $\square$

By definition,  $R_{ljk}(i_n) \subseteq \mathcal{G}_n$  (see (9.11)). By Lemma 25, for  $n \geq 1$  and  $|l| \leq N_{n-1}$  we also have  $R_{ljk}(i_n) \subseteq R_{ljk}(i_{n-1})$ . On the other hand,  $R_{ljk}(i_{n-1}) \cap \mathcal{G}_n = \emptyset$  (see (9.7)). As a consequence,  $R_{ljk}(i_n) = \emptyset$  for all  $|l| \leq N_{n-1}$ , and

$$\mathcal{G}_n \setminus \mathcal{G}_{n+1} \subseteq \bigcup_{\substack{j,k \in S^c \cup \{0\} \\ |l| > N_{n-1}}} R_{ljk}(i_n) \quad \forall n \geq 1. \quad (9.12)$$

**Lemma 26** *Let  $n \geq 0$ . If  $R_{ljk}(i_n) \neq \emptyset$ , then  $|l| \geq C_1 |j^3 - k^3| \geq \frac{1}{2} C_1 (j^2 + k^2)$  for some constant  $C_1 > 0$  (independent of  $l, j, k, n, i_n, \omega$ ).*

*Proof* Follow the proof of Lemma 5.3 of [3], also using (8.62). Note that  $|\omega| \leq 2|\bar{\omega}|$  for all  $\omega \in \Omega_\varepsilon$ , for  $\varepsilon$  small enough, by (5.2) and (4.10).  $\square$

Now we study the measure of the resonant sets  $R_{ljk}(i_n)$  defined in (9.11). We have to analyze in more details the sublevels of the function

$$\omega \mapsto \phi(\omega) := i\omega \cdot l + \mu_j^\infty(\omega) - \mu_k^\infty(\omega), \quad (9.13)$$

appearing in (9.11) ( $\phi$  also depends on  $l, j, k, i_n$ ).

**Lemma 27** *There exists  $C_0 > 0$  such that for all  $j \neq k$ , with  $j^2 + k^2 > C_0$ , the set  $R_{ljk}(i_n)$  has Lebesgue measure  $|R_{ljk}(i_n)| \leq C\varepsilon^{2(\nu-1)} \gamma \langle l \rangle^{-\tau}$ .*

*Proof* For  $l \neq 0$ , decompose  $\omega = s\hat{l} + v$ , where  $\hat{l} := l/|l|$ ,  $s \in \mathbb{R}$ , and  $l \cdot v = 0$  (so that  $\omega \cdot l = s|l|$ ). Let  $\psi(s) := \phi(s\hat{l} + v)$ . The eigenvalues  $\mu_j^\infty$  are given in (8.61). By (8.36) and (5.4),  $\varepsilon^2 |c(\xi)|^{\text{lip}} \leq C_2$  for some constant  $C_2 > 0$  depending only on the set  $S$  of the tangential sites. Then, by (8.62) and (2.2),

$$\begin{aligned} |\tilde{m}_3(s_1) - \tilde{m}_3(s_2)| &\leq C\varepsilon^3 \gamma^{-1} |s_1 - s_2|, \\ |\tilde{m}_1(s_1) - \tilde{m}_1(s_2)| &\leq (C_2 + C\varepsilon^5 \gamma^{-2}) |s_1 - s_2| \leq 2C_2 |s_1 - s_2|, \\ |r_j^\infty(s_1) - r_j^\infty(s_2)| &\leq C\varepsilon^{3-2a} \gamma^{-1} |s_1 - s_2| \end{aligned}$$

for some  $C > 0$  and  $\varepsilon$  small enough, where, with a slight abuse of notations, we have written

$$\tilde{m}_i(s) = \tilde{m}_i(s\hat{l} + v), \quad i = 1, 3 \quad \text{and} \quad r_j^\infty(s) = r_j^\infty(s\hat{l} + v), \quad j \in S^c.$$

By (8.61) and Lemma 26,

$$\begin{aligned} |\psi(s_1) - \psi(s_2)| &\geq (|l| - C\varepsilon^3 \gamma^{-1} |j^3 - k^3| - 2C_2 |j - k| - 2C\varepsilon^{3-2a} \gamma^{-1}) |s_1 - s_2| \\ &\geq |j^3 - k^3| \left( C_1 - C\varepsilon^3 \gamma^{-1} - \frac{2C_2 |j - k|}{|j^3 - k^3|} - \frac{2C\varepsilon^{3-2a} \gamma^{-1}}{|j^3 - k^3|} \right) |s_1 - s_2| \\ &\geq \frac{C_1}{2} |j^3 - k^3| |s_1 - s_2| \end{aligned}$$

for  $\varepsilon$  small enough and  $j^2 + k^2 + |jk| > C_0 := 12C_2/C_1$ . As a consequence, the set  $\Delta_{ljk}(i_n) := \{s : s\hat{l} + v \in R_{ljk}(i_n)\}$  has Lebesgue measure

$$|\Delta_{ljk}(i_n)| \leq \frac{2}{C_1 |j^3 - k^3|} \frac{4\gamma_n |j^3 - k^3|}{\langle l \rangle^\tau} \leq \frac{C\gamma}{\langle l \rangle^\tau}$$

for some  $C > 0$ . The lemma follows by Fubini's Theorem.  $\square$

*Remark 12* When  $K = H + \lambda M^2$ ,  $\lambda = 3/4$ , using (8.40), the conclusion of Lemma 27 holds without restrictions on  $j, k$ .

It remains to estimate the measure of the finitely many resonant sets  $R_{ljk}(i_n)$  for  $j^2 + k^2 \leq C_0$ . Recalling (8.36) and the parity  $\xi_{-j} = \xi_j$ , we write  $c(\xi) = 6\zeta \mathbf{1} \cdot \xi$  where  $\mathbf{1}$  is the vector  $(1, \dots, 1) \in \mathbb{R}^\nu$  and  $\xi = (\xi_j)_{j \in S^+} \in \mathbb{R}^\nu$ . Hence, by (5.4),

$$\varepsilon^2 c(\xi) = 6\zeta \mathbf{1} \cdot \mathbb{A}^{-1}[\omega - \bar{\omega}] = 6\zeta \mathbb{A}^{-T} \mathbf{1} \cdot [\omega - \bar{\omega}] \quad (9.14)$$

where  $\mathbb{A}^{-T}$  is the transpose of  $\mathbb{A}^{-1}$ . We write the function  $\phi(\omega)$  in (9.13) as

$$\phi(\omega) = a_{jk} + b_{ljk} \cdot \omega + q_{jk}(\omega),$$

where

$$\begin{aligned} a_{jk} &:= -i(j^3 - k^3 + 6\zeta(j - k)\mathbf{1} \cdot \mathbb{A}^{-1}\bar{\omega}), \\ b_{ljk} &:= i(l + 6\zeta(j - k)\mathbb{A}^{-T}\mathbf{1}), \\ q_{jk}(\omega) &:= -i(\tilde{m}_3 - 1)(j^3 - k^3) + i(\tilde{m}_1 - \varepsilon^2 c(\xi))(j - k) + r_j^\infty - r_k^\infty \end{aligned}$$

(and  $\tilde{m}_3, \tilde{m}_1, \xi, r_j^\infty, r_k^\infty$  all depend on  $\omega$ ). By (8.62) and since  $j^2 + k^2 \leq C_0$  we deduce that  $|q_{jk}|^{\text{Lip}(\gamma)} \leq C\varepsilon^{3-2a}$ . Recalling (2.2) we get

$$|q_{jk}|^{\text{sup}} \leq C\varepsilon^{3-2a}, \quad |q_{jk}|^{\text{lip}} \leq \gamma^{-1}|q_{jk}|^{\text{Lip}(\gamma)} \leq C\varepsilon^{1-3a} \quad (9.15)$$

so that  $\phi(\omega)$  is a small perturbation of the affine function  $\omega \mapsto a_{jk} + b_{ljk} \cdot \omega$ . By the next lemma, the hypothesis (1.12) on the tangential sites  $S$  allows to verify that such function does not vanish identically.

**Lemma 28** *Assume (1.12). Then, for all  $j \neq k$ ,  $j^2 + k^2 \leq C_0$  it results  $a_{jk} \neq 0$ .*

*Proof* Using formulae (1.19) and (4.11), we calculate

$$\mathbf{1} \cdot \mathbb{A}^{-1}\bar{\omega} = -\frac{1}{3\zeta(2\nu - 1)} \sum_{i=1}^{\nu} \bar{j}_i^2.$$

Hence

$$a_{jk} = -i(j - k) \left( j^2 + jk + k^2 - \frac{2}{2\nu - 1} \sum_{i=1}^{\nu} \bar{j}_i^2 \right) \neq 0$$

by assumption (1.12) on the set  $S$ .  $\square$

Lemma 28 implies that  $\delta := \min\{|a_{jk}| : j^2 + k^2 \leq C_0, j \neq k\} > 0$ .

**Lemma 29** *Assume (1.12). If  $j^2 + k^2 \leq C_0$ , then  $|R_{ljk}(i_n)| \leq C\varepsilon^{2(\nu-1)}\gamma\langle l \rangle^{-\tau}$ .*

*Proof* Denote  $b := b_{ljk}$  for brevity. For  $j^2 + k^2 \leq C_0$ ,  $\omega \in R_{ljk}(i_n)$ , one has, by (9.11), (9.15),

$$|b \cdot \omega| \geq |a_{jk}| - |\phi(\omega)| - |q_{jk}(\omega)| \geq \delta - 2\gamma_n |j^3 - k^3| \langle l \rangle^{-\tau} - C\varepsilon^{3-2a} \geq \delta/2$$

for  $\varepsilon$  small enough. On the other hand,  $|b \cdot \omega| \leq 2|\bar{\omega}||b|$  because  $|\omega| \leq 2|\bar{\omega}|$  (see (5.2) and (4.10)). Hence  $|b| \geq \delta_1$  where  $\delta_1 := \delta/(4|\bar{\omega}|) > 0$ . Split  $\omega = s\hat{b} + v$  where  $\hat{b} := b/|b|$  and  $v \cdot \hat{b} = 0$ . Let  $\psi(s) := \phi(s\hat{b} + v)$ . By (9.15), for  $\varepsilon$  small enough, we get

$$|\psi(s_1) - \psi(s_2)| \geq (|b| - |q_{jk}|^{\text{lip}})|s_1 - s_2| \geq \frac{\delta_1}{2}|s_1 - s_2|.$$

Then we proceed similarly as in the proof of Lemma 27.  $\square$

The proof of (9.9) follows from the lemmata 25, . . . , 29, proceeding like in [3] (see the conclusion of the proof of Theorem 5.1 in [3]).  $\square$

**Proof of Theorem 2 concluded.** The conclusion of the proof of Theorem 2 follows exactly like in [5] (see “Proof of Theorem 5.1 concluded” in [5]).

*Remark 13* By remark 12, Lemma 28 (which is the only point in the paper where assumption (1.12) is used) is not needed any more. Thus Theorem 1 applies to  $K = H + (3\zeta/4)M^2$  without assuming hypothesis (1.12).

**Acknowledgements** This research was supported by the European Research Council under FP7 and PRIN 2012 “Variational and perturbative aspects of nonlinear differential problems”. This research was carried out in the frame of Programme STAR, financially supported by UniNA and Compagnia di San Paolo.

## References

1. Alazard T., Baldi P., *Gravity capillary standing water waves*, Arch. Ration. Mech. Anal. 217 (2015), no. 3, 741-830.
2. Baldi P., *Periodic solutions of fully nonlinear autonomous equations of Benjamin-Ono type*, Ann. Inst. H. Poincaré (C) Anal. Non Linéaire 30 (2013), 33-77.
3. Baldi P., Berti M., Montalto R., *KAM for quasi-linear and fully nonlinear forced perturbations of Airy equation*, Math. Annalen 359, 471-536 (2014).
4. Baldi P., Berti M., Montalto R., *KAM for quasi-linear KdV*, C. R. Acad. Sci. Paris, Ser. I 352 (2014) 603-607.
5. Baldi P., Berti M., Montalto R., *KAM for autonomous quasi-linear perturbations of KdV*, to appear on Ann. Inst. H. Poincaré (C) Anal. Non Linéaire.
6. Baldi P., Floridia G., Haus E., *Exact controllability for quasi-linear perturbations of KdV*, preprint.
7. Berti M., Biasco P., Procesi M., *KAM theory for the Hamiltonian DN LW*, Ann. Sci. Éc. Norm. Supér. (4), Vol. 46, fascicule 2 (2013), 301-373.
8. Berti M., Biasco P., Procesi M., *KAM theory for the reversible derivative wave equation*, Arch. Rational Mech. Anal., 212 (2014), 905-955.
9. Berti M., Bolle P., *Quasi-periodic solutions with Sobolev regularity of NLS on  $T^d$  with a multiplicative potential*, J. Eur. Math. Soc. 15 (2013), 229-286.
10. Berti M., Bolle P., *A Nash-Moser approach to KAM theory*, Fields Institute Communications, special volume “Hamiltonian PDEs and Applications”, to appear.
11. Berti M., Corsi L., Procesi M., *An abstract Nash-Moser theorem and quasi-periodic solutions for NLW and NLS on compact Lie groups and homogeneous manifolds*, Comm. Math. Phys. 334 (2015), no. 3, 1413-1454.
12. Berti M., Montalto R., *KAM for gravity capillary water waves*, preprint.
13. Bourgain J., *Gibbs measures and quasi-periodic solutions for nonlinear Hamiltonian partial differential equations*, 23-43, Gelfand Math. Sem., Birkhäuser Boston, Boston, MA, 1996.
14. Feola R., Procesi M. *Quasi-periodic solutions for fully nonlinear forced reversible Schrödinger equations*, J. Diff. Eq., 259, no. 7, 3389-3447, 2015.
15. Guan H., Kuksin S., *The KdV equation under periodic boundary conditions and its perturbations*, Nonlinearity 27 (2014), no. 9, R61-R88.
16. Iooss G., Plotnikov P.I., *Small divisor problem in the theory of three-dimensional water gravity waves*, Mem. Amer. Math. Soc. 200, no. 940 (2009).
17. Iooss G., Plotnikov P.I., Toland J.F., *Standing waves on an infinitely deep perfect fluid under gravity*, Arch. Rational Mech. Anal. 177 no. 3, (2005), 367-478.
18. Lax P., *Development of singularities of solutions of nonlinear hyperbolic partial differential equations*, J. Mathematical Phys. 5 (1964), 611-613.
19. Liu J., Yuan X., *A KAM Theorem for Hamiltonian Partial Differential Equations with Unbounded Perturbations*, Comm. Math. Phys. 307 (3) (2011), 629-673.
20. Kappeler T., Pöschel J., *KAM and KdV*, Springer, 2003.



21. Kappeler, T., Topalov, P. *Global well-posedness of mKdV in  $L^2(T, R)$* , Comm. Partial Differential Equations 30 (2005), no. 1-3, 435-449.
22. Klainerman S., Majda A., *Formation of singularities for wave equations including the nonlinear vibrating string*, Comm. Pure Appl. Math., 33, (1980), 241-263.
23. Kuksin S., *Hamiltonian perturbations of infinite-dimensional linear systems with imaginary spectrum*, Funktsional. Anal. i Prilozhen. 21, no. 3, 22-37, 95, 1987.
24. Kuksin S., *A KAM theorem for equations of the Korteweg-de Vries type*, Rev. Math. Phys., 10, 3, (1998), 1-64.
25. Kuksin S., *Analysis of Hamiltonian PDEs*, Oxford Lecture Series in Mathematics and its Applications, 19. Oxford University Press (2000).
26. Kuksin S., Pöschel J., *Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation*, Annals of Math. 2 143, (1996), 149-179.
27. Pöschel J., *Quasi-periodic solutions for a nonlinear wave equation*, Comment. Math. Helv., 71, no. 2, (1996) 269-296.
28. Procesi M., Procesi C., *A normal form for the Schrödinger equation with analytic nonlinearities*, Comm. Math. Phys. 312 (2012), 501-557.
29. Taylor M. E., *Pseudodifferential Operators and Nonlinear PDEs*, Progress in Mathematics, Birkhäuser, 1991.
30. Zhang J., Gao M., Yuan X. *KAM tori for reversible partial differential equations*, Nonlinearity 24 (2011), 1189-1228.
31. Zehnder E., *Generalized implicit function theorems with applications to some small divisors problems I-II*, Comm. Pure Appl. Math. 28 (1975), 91-140, and 29 (1976), 49-113.