Quasi-periodic solutions for quasi-linear generalized KdV equations

Filippo Giuliani*
SISSA Preprint 38/2016/MATE

Abstract

We prove the existence of Cantor families of small amplitude, linearly stable, quasi-periodic solutions of *quasi-linear* autonomous Hamiltonian generalized KdV equations. We consider the most general quasi-linear *quadratic* nonlinearity. The proof is based on an iterative Nash-Moser algorithm. To initialize this scheme, we need to perform a bifurcation analysis taking into account the strongly perturbative effects of the nonlinearity near the origin. In particular, we implement a weak version of the Birkhoff normal form method. The inversion of the linearized operators at each step of the iteration is achieved by pseudo-differential techniques, linear Birkhoff normal form algorithms and a linear KAM reducibility scheme.

Keywords: KAM for PDE's; Quasi-linear PDE's; Quasi-periodic solutions; Nash-Moser theory; KdV

Contents

*SISSA, Via Bonomea 265, 34136, Trieste, Italy, fgiulian@sissa.it

1 Introduction

We prove the existence and the stability of Cantor families of quasi-periodic, small amplitude, solutions of the Hamiltonian quasi-linear generalized KdV equations

$$u_t + u_{xxx} + \mathcal{N}_2(x, u, u_x, u_{xx}, u_{xxx}) = 0, \tag{1.1}$$

under periodic boundary conditions $x \in \mathbb{T}$, where

$$\mathcal{N}_2(x, u, u_x, u_{xx}, u_{xxx}) := -\partial_x[(\partial_u f)(x, u, u_x) - \partial_x((\partial_{u_x} f)(x, u, u_x))] \tag{1.2}$$

and f is the most general *quasi-linear* Hamiltonian density

$$f(x, u, u_x) := c_1\, u_x^3 + c_2\, u_x^2\, u + c_3\, u^3 + c_4\, u_x^4 + c_5\, u_x^3\, u + c_6\, u_x^2\, u^2 + c_7\, u^4 + f_{\geq 5}(x, u, u_x), \tag{1.3}$$

where the coefficients $c_i, i = 1, 2, \ldots, 7$ are real numbers, and

$$f_{\geq 5}(x, u, u_x) := f_5(u, u_x) + f_{\geq 6}(x, u, u_x) \tag{1.4}$$

is the sum of the homogeneous component of f of degree five and all the higher order terms.
We assume that the Hamiltonian density f in (1.5) belongs to $C^q(\mathbb{T} \times \mathbb{R} \times \mathbb{R}; \mathbb{R})$ for some large q.

The equation (1.1) can be formulated as a Hamiltonian PDE $u_t = \partial_x \nabla_{L^2} H$, where $\nabla_{L^2} H$ is the $L^2(\mathbb{T})$ gradient of the Hamiltonian

$$H(u) = \int \frac{u_x^2}{2} + f(x, u, u_x)\, dx \tag{1.5}$$

on the real phase space

$$H_0^1(\mathbb{T}_x) := \left\{ u \in H^1(\mathbb{T}, \mathbb{R}) : \int_{\mathbb{T}} u(x)\, dx = 0 \right\} \tag{1.6}$$

endowed with the non-degenerate symplectic form

$$\Omega(u, v) := \int_{\mathbb{T}} (\partial_x^{-1} u)\, v\, dx, \quad \forall u, v \in H_0^1(\mathbb{T}_x), \tag{1.7}$$

where $\partial_x^{-1} u$ is the periodic primitive of u with zero average defined by

$$\partial_x^{-1} e^{\mathrm{i} jx} = \frac{1}{\mathrm{i} j}\, e^{\mathrm{i}\, j\, x} \quad \text{if } j \neq 0, \qquad \partial_x^{-1} 1 = 0.$$

The phase space $H_0^1(\mathbb{T}_x)$ is invariant under the flow of the equation (1.1).
The Poisson bracket induced by Ω between two functions $F, G \colon H_0^1(\mathbb{T}) \to \mathbb{R}$ is

$$\{F(u), G(u)\} := \Omega(X_F, X_G) = \int_{\mathbb{T}} \nabla F(u)\, \partial_x \nabla G(u)\, dx, \tag{1.8}$$

where X_F and X_G are the vector fields associated to the Hamiltonians F and G, respectively.

By (1.3) the nonlinearity $\mathcal{N}_2$ vanishes at order two at $u = 0$ and (1.1) may be seen, in a small neighbourhood of the origin, as a *small* perturbation of the Airy equation

$$u_t + u_{xxx} = 0. \tag{1.9}$$

The equation (1.1) is *completely resonant*, namely its linearized problem at the origin (1.9) possesses only the 2π-periodic in time solutions

$$u(t,x) = \sum_{j\in\mathbb{Z}\setminus\{0\}} u_j \, e^{\mathrm{i}\, j^3\, t}\, e^{\mathrm{i}\, j\, x}. \tag{1.10}$$

Then the existence of quasi-periodic solutions of (1.1) is due only to the presence of the nonlinearity. For this reason, we need to perform a bifurcation analysis which is mainly affected by the quasi-linear monomials of degree three and four in (1.3). Another difficulty is that, since the equation (1.1) is completely resonant, the diophantine frequency vector of the expected quasi-periodic solutions, if any, are $O(|u_j|^2)$-close to integer vectors.

We briefly present some literature related to this paper.
The KAM theory for PDE's has been developed in the eighties by Kuksin, with the pioneering work [22], and by Wayne [33], Craig-Wayne [11], Pöeschel [29] for the one dimensional nonlinear wave and Schrödinger equations, and, at a later time, in higher dimensional cases, by Bourgain [10], Eliasson-Kuksin [13], Berti-Bolle [7], Geng-Xu-You [16], Procesi-Procesi [30]-[31], Wang [32], Eliasson-Grebert-Kuksin [12].
The first results with unbounded perturbations have been proved by Kuksin in [23] and Kappeler-Pöeschel [21] for KdV (see [17] for a survey on known results for the KdV equation), by Liu-Yuan [26], Zhang-Gao-Yuan [34] for derivative NLS, and by Berti-Biasco-Procesi [5]-[6] for derivative NLW.
All the aforementioned papers treat semilinear problems, namely the case in which the nonlinearity depends on derivatives of order m, with $m \le n-1$, where n is the highest order of the derivatives appearing in the unperturbed system.
For quasi-linear and fully nonlinear PDE's, i.e. in the case $m = n$, the progress are more recent.
The first results in this direction are due to Iooss-Plotnikov [18]-[19], Iooss-Plotnikov-Toland [20], Plotnikov-Toland [28] for periodic solutions of water-waves equations. In the spirit of the method implemented in these papers, Baldi in [1] provides the existence of periodic solutions for the Benjamin-Ono equation.
Baldi, Berti, Montalto prove the first existence results of quasi-periodic solutions for quasi-linear and fully nonlinear PDE's, in the forced case for the Airy equation [2], and in the autonoumous case for the KdV and mKdV equation in [3] and [4]. In particular, they consider in [3] the Hamiltonian

$$H_{KdV} + \int_{\mathbb{T}} f_{\ge 5}(x,u,u_x)\, dx, \quad \text{where} \quad H_{KdV} := \int_{\mathbb{T}} \frac{u_x^2}{2} + u^3\, dx, \tag{1.11}$$

namely, the Hamiltonian (1.5) without the monomials of degree three and four in the variables (u, u_x), see (1.3). These works are based on Nash-Moser methods and a reducibility scheme that diagonalize completely the linearized system at any approximate solution. This procedure permits to prove also the linear stability of the solutions.
More recently, in [14] and [15] Feola-Procesi provide the existence and the stability of quasi-periodic solutions for quasi-linear and fully nonlinear perturbations of the Schrödinger equation in dimension one. We mention also the recent work by Montalto [27] on quasi-periodic solutions for the forced Kirchoff equation.

The aim of this paper is to generalize the results obtained in [3] and [4] considering the most general Hamiltonian density (1.3). We are interested in understanding the effect, over infinite times, of a *quadratic* and *quasi-linear* Hamiltonian perturbation in a small neighbourhood of the origin, where the polynomial perturbations of lowest degree are much *stronger*. This is significant in view of the study of small amplitude solutions for many fluid dynamics equations, like Degasperis-Procesi and water waves-type equations, which involve this kind of nonlinearities.

1.1 Main result

The solutions that we find are localized in Fourier space close to finitely many *tangential sites*

$$S^+ := \{\bar{\jmath}_1, \dots, \bar{\jmath}_\nu\}, \quad S := S^+ \cup (-S^+) = \{\pm j : j \in S^+\}, \quad \bar{\jmath}_i \in \mathbb{N} \setminus \{0\}, \quad \forall i = 1, \dots, \nu \tag{1.12}$$

and the linear frequencies of oscillation on the tangential sites are

$$\overline{\omega} := (\bar{\jmath}_1^3, \dots, \bar{\jmath}_\nu^3) \in \mathbb{N}^\nu. \tag{1.13}$$

The set S is required to be even because we look for real valued solutions of (1.1). Moreover, we also assume the following hypotesis on S:

(S) $\nexists\ j_1, j_2, j_3, j_4 \in S$ such that

$$j_1 + j_2 + j_3 + j_4 \neq 0, \ j_1^3 + j_2^3 + j_3^3 + j_4^3 - (j_1 + j_2 + j_3 + j_4)^3 = 0.$$

We decompose the phase space as

$$H_0^1(\mathbb{T}) := H_S \oplus H_S^\perp, \quad H_S := \mathrm{span}\{e^{\mathrm{i}\, j\, x} : j \in S\}, \quad H_S^\perp := \{u = \sum_{j \in S^c} u_j\, e^{\mathrm{i}\, j\, x} \in H_0^1(\mathbb{T})\},$$

and we denote by $\Pi_S, \Pi_S^\perp$ the corresponding orthogonal projectors. The subspaces H_S and $H_S^\perp$ are symplectic respect to the 2-form Ω (see (1.7)). We write

$$u = v + z, \quad v := \Pi_S u := \sum_{j \in S} u_j\, e^{\mathrm{i}\, j\, x}, \quad z = \Pi_S^\perp u := \sum_{j \in S^c} u_j\, e^{\mathrm{i}\, j\, x}, \tag{1.14}$$

where v is called the *tangent* variable and z the *normal* one. In the following, we will identify $v = (v_j)_{j \in S}$ and $z = (z_j)_{j \in S^c}$.

We shall also assume "non-resonant" and "non-degeneracy" conditions for the nonlinearity (1.3).

Definition 1.1. We say that the coefficients $c_1, \dots, c_7$ are *resonant* if the following holds

$$c_3 = c_7 = 2c_1^2 - c_4 = 7c_2^2 - 6c_6 = 0 \tag{1.15}$$

and we say that $c_1, \dots, c_7$ are *non-resonant* if (1.15) does not hold.

Moreover, we require the following "non-degeneracy" conditions on the coefficients $c_1, \dots, c_7$

(C1) fixed $\nu \in \mathbb{N}$, the coefficients $c_1, \dots, c_7$ satisfy

$$(7 - 16\nu)\, c_2^2 \neq 6\, (1 - 2\nu) c_6, \tag{1.16}$$

(C2) fixed $\nu \in \mathbb{N}$, the coefficients $c_1, \dots, c_7$ satisfy

$$\nu\, \frac{3c_6 - 4c_2^2}{9c_4 - 18c_1^2} \notin \{j^2 + k^2 + jk \,:\, j, k \in \mathbb{Z} \setminus \{0\},\, j \neq k\}. \tag{1.17}$$

Before stating the main result, we introduce a notion of "genericity" according to the one given by Biasco-Berti-Procesi [6], Procesi-Procesi [30] and Feola [15].

Definition 1.2. Fixed $\nu \in \mathbb{N}$ and given a non-trivial, i.e non identically zero, polynomial $P(z)$, with $z \in \mathbb{C}^\nu$, we say that a vector of integers $z_0 \in \mathbb{N}^\nu$ is *generic* if $P(z_0) \neq 0$.
We shall say that "*there is a generic choice of the tangential sites S for which some condition holds*" if this condition is satisfied by every vectors of integers $(\bar{\jmath}_1, \dots \bar{\jmath}_\nu)$ that are not zeros of some non trivial polynomial.

Theorem 1.3. *Given $\nu \in \mathbb{N}$, let $f \in C^q$ (with $q := q(\nu)$ large enough) satisfy* (1.3). *If $c_1, \dots, c_7$ in* (1.3) *are non-resonant (see Definition 1.1) and conditions* (C1)-(C2) *hold, then for a generic choice of tangential sites (see Definition 1.2 and* (1.12)*), in particular satisfying* (S)*, the 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*

$$u(t,x) = \sum_{j\in S^+} 2\sqrt{j\,\xi_j}\,\cos(\omega_j t + jx) + o(\sqrt{|\xi|}), \qquad \omega_j = j^3 + O(\xi_j) \tag{1.18}$$

for a Cantor-like set of small amplitudes $\xi \in \mathbb{R}^\nu_+$ with density 1 *at $\xi = 0$. The term $o(\sqrt{|\xi|})$ is small in some H^s-Sobolev norm, $s < q$. These quasi-periodic solutions are linearly stable.*

Let us make some comments on the assumptions of Theorem 1.3.

- The non-resonance condition stated in Definition 1.1 arises by asking that the frequency-amplitude map (4.18) is a diffeomorphism. The invertibility of this map is equivalent to require that $\det \mathbb{M} \neq 0$, where the determinant of $\mathbb{M}$ is a polynomial in the variables $(c_1, \dots, c_7, \bar{\jmath}_1, \dots, \bar{\jmath}_\nu)$. In Theorem 1.3 we fix non-resonant coefficients $c_1, \dots, c_7$ and we prove in Lemma 4.2 that the condition $\det \mathbb{M} \neq 0$ is satisfied for a generic choice of the tangential sites S. We remark that this explicit condition could be verified by fixing the integers $\bar{\jmath}_1, \dots, \bar{\jmath}_\nu$ and choosing the real parameters $c_1, \dots, c_7$ outside the zeros of some polynomial.

- For the measure estimates of Section 9.1, we shall avoid some lower order resonances by imposing the assumptions (H1) and $(\mathtt{H2})_{j,k}$ (see (9.33), (9.34)). These ones imply that some polynomials are non zero at $(c_1, \dots, c_7, \bar{\jmath}_1, \dots, \bar{\jmath}_\nu)$. If (C1)-(C2) hold and $c_1, \dots, c_7$ are non-resonant then these polynomials are not trivial in the variables $(\bar{\jmath}_1, \dots, \bar{\jmath}_\nu)$ (see Lemma 9.7 and Lemma 9.8) and, for a finite number of $j, k \in S^c$, (H1) and $(\mathtt{H2})_{j,k}$ are verified by fixing non-resonant parameters $c_1, \dots, c_7$ and by choosing a generic set of integers $\{\bar{\jmath}_1, \dots, \bar{\jmath}_\nu\}$.

- As in [3], we assume the Hypotesis (S), because we want to perform three steps of Birkhoff normal form. Indeed the smallness condition (9.4) required in Theorem 9.1 depends on the quadraticity of the nonlinearity in (1.1). We remark that he assumption (S) can be reformulate as a condition that is satisfied for a generic choice of the tangential sites.

The proof of Theorem 1.3 follows the scheme adopted in [3] and [4]. We now shortly present the strategy of the proof of Theorem 1.3 underlying the main differences with these works.

Bifurcation analysis. We cannot consider (1.1) as a perturbation problem for the linearized equation at the origin $u_t + u_{xxx} = 0$, because, as we said above, this equation is completely resonant, hence the frequency vector of its solutions does not satisfy any diophantine condition. Thus, the main modulation of the frequency vector of the solutions with respect to its amplitude is due to the nonlinearity $\mathcal{N}_2$, defined in (1.2). In order to control the shift of the linear frequencies under the effect of the nonlinearity near the origin and to find approximate quasi-periodic solutions for (1.1), in Section 3 we perform a weak version of the Birkhoff normal form algorithm. After two steps of this procedure, we are able to find a finite dimensional submanifold of the phase space foliated by approximately invariant tori, from which the expected quasi-periodic solutions of (1.1) bifurcate. On this subspace we introduce action-angle variables (see Section 4) and we use the "unperturbed" actions ξ of these tori as parameters for our problem. We require also that the frequency-amplitude map $\alpha(\xi)$ in (4.18), namely the function associating the actions to the frequencies, is a diffeomorphism (see Lemma 4.2), so that we could consider both as independent parameters.
The presence of the quasi-linear monomials of degree three and four in the Hamiltonian (1.5) makes significantly harder the computations of the new Hamiltonian after two steps of Birkhoff normal form with respect to the case examined in [3] for the Hamiltonian H_{KdV} (recall (1.11)). Because of the integrability of the KdV system, in [3] the *twist condition*, namely, the invertibility of the frequency-amplitude map, is obtained for every choice of the tangential set S (see (1.12)). On the contrary, for the general case (1.5) the twist condition depends on the choice of the parameters $c_1, \dots, c_7$ and the tangential sites $\bar{\jmath}_1, \dots, \bar{\jmath}_\nu$.

In Lemma 4.2 we provide the invertibility of the frequency-amplitude map for a large choice of the tangential sites and of the coefficients.

Nonlinear functional setting. After the rescaling (4.13), we look for quasi-periodic solutions with frequency vector ω for the (ω, ε)-parameter family of Hamiltonians (4.19). We assume that ω belongs to the image of the restriction of the frequency-amplitude map $\alpha(\xi)$ on a small compact subset of $\mathbb{R}^\nu$ that does not contain the origin (see (5.2)).
In Section 5 we formulate this problem as the search of the zeros of the nonlinear functional $\mathcal{F}(\omega, i(\omega))$ defined in (5.7), where ω is considered as an external parameter and $\varphi \mapsto i(\varphi)$ is a torus embedded in the phase space. We find a solution $i_\infty(\omega t)$ for $\mathcal{F} = 0$, which will correspond to a quasi-periodic solution with frequency vector ω of the original equation (1.1), by constructing, through a Nash-Moser iteration, a sequence $(i_n)_{n\geq 0}$ of approximate solutions that converges to it, see Theorem 5.2.

The inversion of the linearized operator at an approximate solution. The application of a Nash-Moser scheme involves, at any step, the inversion of the linearized operator at an approximate solution and this is, in fact, the main issue of the proof. Thanks to the abstract decoupling procedure developed by Berti-Bolle in [8], that exploits the Hamiltonian structure, the tangential and the normal linear dynamics around an approximately invariant torus can be studied separately, see Section 6. In particular, a suitable change of coordinates around this approximate quasi-periodic solution triangularizes the linearized problem and its inversion reduces to the study of a quasi-periodically forced PDE restricted to normal directions. The operator which has to be inverted, say $\mathcal{L}_\omega$, is pseudodifferential with variable coefficients and it is computed in Section 7.
In Section 8 we conjugate $\mathcal{L}_\omega$ to a diagonal operator, which describes infinitely many harmonic oscillators

$$\dot{v}_j + \mu_j^\infty v_j = 0, \qquad j \in S^c, \quad \mu_j^\infty \in \mathrm{i}\mathbb{R}. \tag{1.19}$$

The diagonalization of $\mathcal{L}_\omega$ is obtained with the same transformations defined in [3] and [4]. The main perturbative effect to the spectrum of $\mathcal{L}_\omega$ is due to the term $a_1(\omega t)\partial_{xxx}$ (see (7.33)) and the presence of u_x in the cubic part of the Hamiltonian density (1.3) affects this coefficient. In particular, $a_1 - 1 = O(\varepsilon)$, instead of $O(\varepsilon^3)$ as in [3]. In general, the corrections of the coefficients of $\mathcal{L}_\omega$ are bigger in size and this fact implies some difficulties in providing the smallness condition (8.133) required in Theorem 8.19. Moreover, the transformation used to conjugate $\mathcal{L}_\omega$ to a pseudodifferential operator with a coefficient in front of ∂_{xxx} independent of the x-variable (see Section 8.1) has form $\mathrm{I} + O(\varepsilon)$ and so it generates new terms of order ε^2. These terms are not perturbative for the reducibility scheme of Theorem 8.19 and we need to compute them in view of a linear Birkhoff normalization.
We also point out that we drop the assumption

$$j_1 + j_2 + j_3 \neq 0 \quad \text{for all } \ j_1, j_2, j_3 \in S \tag{1.20}$$

required in [3] to get "good" estimates on the transformations used to conjugate $\mathcal{L}_\omega$ to a diagonal operator. We better discuss this fact in Remark 8.8.

The Nash-Moser iteration, measure estimates and stability. In Section 9 we perform the nonlinear Nash-Moser iteration which proves Theorem 5.2 and, therefore, Theorem 1.3.
In the measure estimates for the sets of parameters $\mathcal{R}_{ljk}$, for which the second Melnikov conditions are violated (see (9.23)), some technical difficulties arise. Indeed, the corrections to the normal frequencies are big in size and the indices l, j, k are not tied by the conservation of the momentum, as, for instance, in [15], since the nonlinearity f in (1.3) depends on x. From these facts, some cases result to be *degenerate* and we shall impose some assumptions on the set S to avoid them (see Remark 9.6 and (9.33), (9.34)).
In Section 9.2 we prove the stability of the solution produced by the Nash-Moser algorithm exploiting the action-angle variables introduced in Section 4 and the diagonalization procedure performed in Section 8.

Acknowledgements

I am greatful to Massimiliano Berti for introducing me to the study of KAM theory and for the support offered for the elaboration of this paper. I also thank Michela Procesi and Roberto Feola for useful and stimulating discussions.

2 Preliminaries

2.1 Functional setting

Lipschitz norm. For a function $u\colon \Omega_0 \to E, \omega \to u(\omega)$, where $(E, \|\cdot\|_E)$ is a Banach space and Ω_0 is a subset of $\mathbb{R}^\nu$, we define the sup-norm and the lipschitz semi-norm

$$\begin{aligned} \|u\|_E^{\sup} &:= \|u\|_{E,\Omega_0}^{\sup} := \sup_{\omega\in\Omega_0} \|u(\omega)\|_E, \\ \|u\|_E^{lip} &:= \|u\|_{E,\Omega_0}^{lip} := \sup_{\omega_1\neq\omega_2} \frac{\|u(\omega_1)-u(\omega_2)\|_E}{|\omega_1-\omega_2|}, \end{aligned} \tag{2.1}$$

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

$$\|u\|_E^{Lip(\gamma)} := \|u\|_{E,\Omega_0}^{Lip(\gamma)} := \|u\|_E^{\sup} + \gamma \|u\|_E^{lip}. \tag{2.2}$$

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

Sobolev norms. We denote by

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

the Sobolev norms of functions $u = u(\varphi, x) \in H^s(\mathbb{T}^\nu \times \mathbb{T})$. We denote by $\|\cdot\|_{H^s_x}$, the Sobolev norm of functions $u(x)$ in the phase space of class H^s. We consider $s_0 := (\nu+2)/2$, hence we have that $H^{s_0}(\mathbb{T}^{\nu+1})$ is continuosly embedded in $L^\infty(\mathbb{T}^{\nu+1})$ and any space $H^s(\mathbb{T}^{\nu+1})$ with $s \geq s_0$ is an algebra and satisfies the interpolation inequalities: for $s \geq s_0$

$$\|u\, v\|_s \leq C(s_0)\, \|u\|_s \|v\|_{s_0} + C(s) \|u\|_{s_0} \|v\|_s, \quad \forall u, v \in H^s(\mathbb{T}^{\nu+1}). \tag{2.4}$$

The above inequalities also hold for the norm $\|\cdot\|^{Lip(\gamma)}$.
We also denote

$$\begin{aligned} H^s_{S^\perp}(\mathbb{T}^{\nu+1}) &:= \left\{ u \in H^s(\mathbb{T}^{\nu+1}) : u(\varphi, \cdot) \in H_S^\perp, \ \ \forall \varphi \in \mathbb{T}^\nu \right\}, \\ H^s_S(\mathbb{T}^{\nu+1}) &:= \left\{ u \in H^s(\mathbb{T}^{\nu+1}) : u(\varphi, \cdot) \in H_S \ \ \forall \varphi \in \mathbb{T}^\nu \right\}. \end{aligned} \tag{2.5}$$

We will use the notation $a \leq b$ to denote $a \leq C\, b$ for some constant $C > 0$. In particular, if the constant $C := C(s)$ depends on the index s, then we will use the notation $a \leq_s b$.

Matrices with off-diagonal decay. A linear operator can be identified with its matrix representation. We recall the definition of the s-decay norm (introduced in [7]) of an infinite dimensional matrix. This norm is used in [2] for the KAM reducibility scheme of the linearized operators.

Definition 2.1. The s-decay norm of an infinite dimensional matrix $A := (A_{i_1}^{i_2})_{i_1,i_2\in\mathbb{Z}^b}, b \geq 1$ is

$$|A|_s^2 := \sum_{i\in\mathbb{Z}^b} \langle i \rangle^{2\,s} \left(\sup_{i_1-i_2=i} |A_{i_1}^{i_2}| \right)^2. \tag{2.6}$$

For parameter dependent matrices $A := A(\omega), \omega \in \Omega_0 \subseteq \mathbb{R}^\nu$, the definitions (2.1) and (2.2) become

$$\begin{aligned} &|A|_s^{\sup} := \sup_{\omega\in\Omega_0} |A(\omega)|_s, \ \ |A|_s^{lip} := \sup_{\omega_1\neq\omega_2} \frac{|A(\omega_1)-A(\omega_2)|_s}{|\omega_1-\omega_2|}, \\ &|A|_s^{Lip(\gamma)} := |A|_s^{\sup} + \gamma |A|_s^{lip}. \end{aligned} \tag{2.7}$$

Such a norm is modelled 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 $h \to p\, h$ is represented by the Töplitz matrix $T_i^j = p_{i-j}$ and $|T|_s = \|p\|_s$. If $p = p(\omega)$ is a Lipschitz family of functions, then

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

The s-norm satisfies classical algebra and interpolation inequalities proved in [7].

Lemma 2.2. *Let $A = A(\omega), B = B(\omega)$ be matrices depending in a Lipschitz way on the parameter $\omega \in \Omega_0 \subseteq \mathbb{R}^\nu$. Then for all $s \geq s_0 > b/2$ there are $C(s) \geq C(s_0) \geq 1$ such that*

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

The s-decay norm controls the Sobolev norm, namely

$$\|Ah\|_s^{Lip(\gamma)} \leq C(s)\left(|A|_{s_0}^{Lip(\gamma)}\|h\|_s^{lip(\gamma)} + |A|_s^{Lip(\gamma)}\|h\|_{s_0}^{Lip(\gamma)}\right). \tag{2.8}$$

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 (2.6) is

$$|A|_s^2 = \sum_{j\in\mathbb{Z}, l\in\mathbb{Z}^\nu} \left(\sup_{j_1 - j_2 = j} |A_{j_1}^{j_2}(l)|\right)^2 \langle l, j\rangle^{2\,s}. \tag{2.9}$$

These matrices are identified with the φ-dependent family of operators

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

which act on functions of the x-variables as

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

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 2.3. We say that

(1) a map is symplectic if it preserves the 2-form Ω in (1.7);

(2) 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 $H_S^\perp$);

(3) the operator $\omega \cdot \partial_\varphi - \partial_x G(\varphi)$ is Hamiltonian if each $G(\varphi), \varphi \in \mathbb{T}^\nu$, is symmetric;

(4) an operator is real if it maps real-valued functions into real-valued functions.

A Hamiltonian operator is transformed, under a symplectic map, into another Hamiltonian operator, see [2]-Section 2.3.
We conclude this preliminary section recalling the following well known lemmata about composition of functions (see, e.g., Appendix in [2]).

Lemma 2.4. *(Change of variables) Let $p \in W^{s,\infty}(\mathbb{T}^d,\mathbb{R}), s \geq 1$, with $|p|_{1,\infty} \leq 1/2$. 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 $|q|_{s,\infty} \leq C|p|_{s,\infty}$. If, moreover, p depends in a Lipschitz way on a parameter $\omega \in \Omega \subseteq \mathbb{R}^\nu$, and $\|D_x p\|_{L^\infty} \leq 1/2$ for all ω, then $|q|^{Lip(\gamma)}_{s,\infty} \leq C|p|^{Lip(\gamma)}_{s+1,\infty}$. The constant $C := C(d,s)$ is independent of γ.*
If $u \in H^s(\mathbb{T}^d,\mathbb{C})$ then $(u \circ f)(x) := u(x+p(x))$ satisfies

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

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

Lemma 2.5. *(Tame product) Let $s \geq s_0 > d/2$. Then, for all $u, v \in H^s(\mathbb{T}^d)$, we have*

$$\|u\,v\|_s \leq C(s_0)\|u\|_{s_0}\|v\|_s + C(s)\|u\|_s\|v\|_{s_0}. \tag{2.10}$$

A function $f\colon \mathbb{T}^d \times B_1 \to \mathbb{C}$, where $B_1 := \{y \in \mathbb{R}^m : |y| < 1\}$, induces the composition operator

$$\tilde{f}(u)(x) := f(x, u(x), Du(x), \ldots, D^p u(x)) \tag{2.11}$$

where $D^k u(x)$ denotes the partial derivatives $\partial_x^\alpha u$ of order $|\alpha| = k$.

Lemma 2.6. *(Composition of functions) Assume $f \in C^r(\mathbb{T}^d \times B_1)$. Then for all $u \in H^{r+p}$ such that $|u|_{p,\infty} < 1$, the composition operator (2.11) is well defined and $\|\tilde{f}(u)\|_r \leq C\|f\|_{C^r}(\|u\|_{r+p} + 1)$, where the constant C depends on r, d, p. If $f \in C^{r+2}$ then for all $|u|_{p,\infty}, |h|_{,\infty} < 1/2$,*

$$\|\tilde{f}(u+h) - \sum_{i=0}^{k} \frac{\tilde{f}^{(i)}(u)}{i!}[h^i]\|_r \leq C\,\|f\|_{C^{r+2}}\|h\|^k_{L^\infty}(\|h\|_{r+p} + \|h\|_{L^\infty}\|u\|_{r+p}). \tag{2.12}$$

Lemma 2.7. *Let $d \in \mathbb{N}$, $d/2 < s_0 \leq s$, $p \geq 0$, $\gamma > 0$. Let F be a C^1-map satisfying the tame estimates: for all $\|u\|_{s_0+p} \leq 1, h \in H^{s+p}$,*

$$\|F(u)\|_s \leq C(s)(1 + \|u\|_{s+p}),$$
$$\|\partial_u F(u)[h]\|_s \leq C(s)(\|h\|_{s+p} + \|u\|_{s+p}\|h\|_{s_0+p}).$$

For $\Omega_0 \subset \mathbb{R}^\nu$, let $u(\omega)$ be a Lipschitz family of functions parametrized by $\omega \in \Omega_0$ with $\|u\|^{Lip(\gamma)}_{s_0+p} \leq 1$. Then

$$\|F(u)\|^{Lip(\gamma)}_s \leq C(s)(1 + \|u\|^{Lip(\gamma)}_{s+p}).$$

2.2 Fourier representation

In order to solve the homological equations along the Birkhoff normal form procedures performed in Sections 3 and 8, it is convenient to use the Fourier representation

$$u(x) = \sum_{j \in \mathbb{Z}\setminus\{0\}} u_j\, e^{\mathrm{i}\,j\,x}, \tag{2.13}$$

where the support of u excludes the zero because the elements of the phase space have zero average. Moreover, $\overline{u}_j = u_{-j}$, since the function u is real-valued. The symplectic structure (1.7) writes

$$\Omega = \frac{1}{2}\sum_{j\neq 0} \frac{1}{\mathrm{i}j}\, du_j \wedge du_{-j}, \qquad \Omega(u,v) = \sum_{j\neq 0} \frac{1}{\mathrm{i}j} u_j\, v_{-j}, \tag{2.14}$$

the Hamiltonian vector field X_H and the Poisson bracket (1.8) are respectively

$$[X_H(u)]_j = \mathrm{i}\,j\,\partial_{u_{-j}} H(u), \quad \{F,G\}(u) = -\sum_{j\neq 0} \mathrm{i}\,j\,(\partial_{u_{-j}}F)(u)(\partial_{u_j}G)(u). \tag{2.15}$$

We say that a homogeneous Hamiltonian of degree n

$$H(u) = \sum_{j_1,\dots,j_n \in \mathbb{Z}\setminus\{0\}} H_{j_1,\dots,j_n} u_{j_1} \dots u_{j_n} \tag{2.16}$$

preserves the *momentum* if it is supported on the set $\{(j_1,\dots,j_n) \in \mathbb{Z}^n \setminus \{\mathbf{0}\} : j_1 + \dots + j_n = 0\}$, where we denote with $\{\mathbf{0}\}$ the origin of any vector space $\mathbb{R}^n$, or, equivalently, if

$$\{H, M\} = 0, \quad M(u) = \int_{\mathbb{T}} u^2 \, dx.$$

We note that, by the presence of the x in the arguments of the function f in (1.5), the *momentum* is not preserved along the orbits of the equation (1.1).

3 Weak Birkhoff Normal form

The Hamiltonian (1.5) is $H = H_2 + H_3 + H_4 + H_{\geq 5}$, where

$$\begin{aligned} &H_2(u) := \frac{1}{2}\int_{\mathbb{T}} u_x^2 \, dx, \quad H_3(u) := \int_{\mathbb{T}} c_1 \, u_x^3 + c_2 \, u_x^2 \, u + c_3 \, u^3 \, dx, \\ &H_4(u) := \int_{\mathbb{T}} c_4 \, u_x^4 + c_5 \, u_x^3 \, u + c_6 \, u_x^2 \, u^2 + c_7 \, u^4 \, dx, \quad H_{\geq 5}(u) := \int_{\mathbb{T}} f_{\geq 5}(x, u, u_x) \, dx. \end{aligned} \tag{3.1}$$

For a finite dimensional space

$$E := E_C := \operatorname{span}\left\{ e^{\mathrm{i}\, j\, x} : 0 < |j| \leq C \right\}, \quad C > 0, \tag{3.2}$$

let Π_E denote the corresponding L^2-projector on E.
The notation $R(v^{k-q} z^q)$ indicates a homogeneous polynomial of degree k in (v, z) of the form

$$R(v^{k-q} z^q) = M[\underbrace{v,\dots,v}_{(k-q)\ times}, \underbrace{z,\dots,z}_{q\ times}], \quad M = k - \text{linear}.$$

We denote with $H_{n,\geq k}, H_{n,k}, H_{n,\leq k}$ the terms of type $R(v^{n-s} \, z^s)$, where, respectively, $s \geq k, s = k, s \leq k$, that appear in the homogeneous polynomial H_n of degree n in the variables (v, z).
In particular, we have

$$H_{3,\leq 1} = \int_{\mathbb{T}} \left\{ c_1 (v_x^3 + 3 \, v_x^2 \, z_x) + c_2 (v_x^2 \, v + 2 \, v_x \, v \, z_x + v_x^2 \, z) + c_3 (v^3 + 3 \, v^2 \, z) \right\} \, dx, \tag{3.3}$$

$$H_{3,\geq 2} = \int_{\mathbb{T}} \{ c_1 (z_x^3 + 3 \, z_x^2 \, v_x) + c_2 (z_x^2 \, z + z_x^2 \, v + 2 \, z_x \, z \, v_x) + c_3 (z^3 + 3 \, v^2 \, z) \} \, dx, \tag{3.4}$$

$$H_{4,0} = \int_{\mathbb{T}} \{ c_4 \, v_x^4 + c_5 \, v_x^3 \, v + c_6 \, v_x^2 \, v^2 + c_7 \, v^4 \} \, dx. \tag{3.5}$$

Proposition 3.1. *(**Weak Birkhoff Normal form**) Assume Hypothesis* (S)*. Then there exists an analytic invertible transformation of the phase space $\Phi_B \colon H_0^1(\mathbb{T}_x) \to H_0^1(\mathbb{T}_x)$ of the form*

$$\Phi_B(u) = u + \Psi(u), \quad \Psi(u) := \Pi_E \Psi(\Pi_E u), \tag{3.6}$$

where E is a finite dimensional space as in (3.2)*, such that the transformed Hamiltonian is*

$$\mathcal{H} = H \circ \Phi_B = H_2 + \mathcal{H}_3 + \mathcal{H}_4 + \mathcal{H}_5 + \mathcal{H}_{\geq 6}, \tag{3.7}$$

where H_2 is defined in (3.1),

$$
\begin{aligned}
&\mathcal{H}_3 = c_1 \int_{\mathbb{Z}} (z_x^3 + 3\, z_x^2\, v_x)\, dx + c_2 \int_{\mathbb{Z}} (z_x^2\, z + z_x^2\, v + 2\, v_x\, z_x\, z)\, dx + c_3 \int_{\mathbb{T}} (z^3 + 3v\, z^2)\, dx, \\
&\mathcal{H}_4 = H_{4,0}^{(4)} + \mathcal{H}_{4,2} + \mathcal{H}_{4,3} + \mathcal{H}_{4,4}, \quad \mathcal{H}_{4,2} = R(v^2\, z^2), \quad \mathcal{H}_{4,3} = R(v\, z^3), \\
&\mathcal{H}_{4,4} = \int_{\mathbb{T}} c_4\, z_x^4 + c_5\, z_x^3\, z + c_6\, z_x^2\, z^2 + c_7\, z^4\, dx, \quad \mathcal{H}_5 = \sum_{q=2}^{5} R(v^{5-q}\, z^q),
\end{aligned}
\tag{3.8}
$$

$H_{4,0}^{(4)}$ is defined in (3.25) *and $\mathcal{H}_{\geq 6}$ collects all the terms of order at least six in* (v, z).

The rest of this section is devoted to the proof of the Proposition 3.1.
We construct a symplectic map Φ_B as the composition of analytic and invertible transformations on the phase space that eliminates the terms linear in z and independent of it from the Hamiltonian (1.5). In this way, the Hamiltonian system (1.1) tranforms into one that is integrable and non-isocronous on the subspace $\{z = 0\}$.

Remark 3.2. We note that if $j_1, \ldots, j_N \in \mathbb{Z} \setminus \{0\}, j_1 + \cdots + j_N = 0$ and at most one of these integers does not belong to S, then $\max_{i=1,\ldots,N} |j_i| \leq (N-1) C_S$, where $C_S := \max_{j \in S} |j|$. Thus, the vector field $X_{F^{(N)}}$, generated by the finitely supported Hamiltonian

$$
F^{(N)} = \sum_{j_1 + \cdots + j_N = 0} F^{(N)}_{j_1 \ldots j_N} u_{j_1} \ldots u_{j_N},
$$

is finite rank, and, in particular, it vanishes outside the finite dimensional subspace $E := E_{(N-1)C_S}$ (see (3.2)) and it has the form

$$
X_{F^{(N)}}(u) = \Pi_E X_{F^{(N)}}(\Pi_E u).
$$

Hence its flow $\Phi^{(N)}$ is analytic and invertible on the phase space $H_0^1(\mathbb{T}_x)$.

Step one. First we remove the cubic terms independent of z and linear in z from the Hamiltonian H_3 defined in (3.1) . We look for a symplectic transformation $\Phi^{(3)}$ of the phase space which eliminates the monomials $u_{j_1}\, u_{j_2}\, u_{j_3}$ of H_3 with at most one index outside S.
We look for $\Phi^{(3)} := (\Phi^t_{F^{(3)}})_{|t=1}$ as the time-1 flow map generated by the Hamiltonian vector field $X_{F^{(3)}}$, with an auxiliary Hamiltonian of the form

$$
F^{(3)}(u) := \sum_{j_1 + j_2 + j_3 = 0} F^{(3)}_{j_1\, j_2\, j_3}\, u_{j_1}\, u_{j_2}\, u_{j_3}.
$$

The transformed Hamiltonian is

$$
\begin{aligned}
&H^{(3)} := H \circ \Phi^{(3)} = H_2 + H_3^{(3)} + H_4^{(3)} + H_{\geq 5}^{(3)}, \\
&H_3^{(3)} = H_3 + \{H_2, F^{(3)}\}, \quad H_4^{(3)} = \frac{1}{2}\{\{H_2, F^{(3)}\}, F^{(3)}\} + \{H_3, F^{(3)}\} + H_4,
\end{aligned}
\tag{3.9}
$$

where $H_{\geq 5}^{(3)}$ collects all the terms of order at least five in (v, z). In order to find the exact expression of $F^{(3)}$, we have to solve the homological equation

$$
H_3 + \{H_2, F^{(3)}\} = H_{3,\geq 2} \tag{3.10}
$$

or, equivalently, $\{H_2, F^{(3)}\} = -H_{3,\leq 1}$, see (3.3). In the Fourier representation, by (1.8) and (3.1), the equation (3.10) writes

$$
\sum_{j_1 + j_2 + j_3 = 0} \mathrm{i}\, (j_1^3 + j_2^3 + j_3^3)\, F^{(3)}_{j_1 j_2 j_3}\, u_{j_1}\, u_{j_2}\, u_{j_3} = \sum_{(j_1, j_2, j_3) \in \mathcal{A}_3} (-\mathrm{i}\, c_1\, j_1 j_2 j_3 - c_2\, j_1 j_2 + c_3)\, u_{j_1} u_{j_2} u_{j_3} \tag{3.11}
$$

where

$$\mathcal{A}_3 := \{(j_1, j_2, j_3) \in \mathbb{Z}^3 \setminus \{\mathbf{0}\} : j_1 + j_2 + j_3 = 0 \text{ and at least 2 indices among } j_1, j_2, j_3 \text{ belong to } S\}.$$

We note that if $(j_1, j_2, j_3) \in \mathcal{A}_3$ then $j_1^3 + j_2^3 + j_3^3 \neq 0$, because

$$j_1 + j_2 + j_3 = 0 \quad \Rightarrow \quad j_1^3 + j_2^3 + j_3^3 = 3\, j_1\, j_2\, j_3 \tag{3.12}$$

and $j_1, j_2, j_3 \in \mathbb{Z} \setminus \{0\}$.
Hence, to solve the equation (3.10) we choose

$$F^{(3)}_{j_1 j_2 j_3} := \begin{cases} \dfrac{-\mathrm{i}\, c_1\, j_1 j_2 j_3 - c_2\, j_1 j_2 + c_3}{\mathrm{i}(j_1^3 + j_2^3 + j_3^3)} & \text{if } (j_1, j_2, j_3) \in \mathcal{A}_3, \\ 0 & \text{otherwise.} \end{cases} \tag{3.13}$$

By construction, all the monomials of H_3 with at least two indices outside S are not modified by the transformation $\Phi^{(3)}$. Hence we have

$$H_3^{(3)} = c_1 \int_{\mathbb{Z}} (z_x^3 + 3\, z_x^2\, v_x)\, dx + c_2 \int_{\mathbb{Z}} (z_x^2\, z + z_x^2\, v + 2\, v_x\, z_x\, z)\, dx + c_3 \int_{\mathbb{T}} (z^3 + v\, z^2)\, dx. \tag{3.14}$$

Now we compute the fourth order term $H_4^{(3)}$ in (3.9). We have, by (3.10)

$$H_4^{(3)} = \frac{1}{2}\{\{H_2, F^{(3)}\}, F^{(3)}\} + \{H_3, F^{(3)}\} + H_4 = \frac{1}{2}\{H_{3,\leq 1}, F^{(3)}\} + \{H_3^{(3)}, F^{(3)}\} + H_4 \tag{3.15}$$

and by (3.11) and (3.13)

$$\begin{aligned} F^{(3)}(u) = &- \frac{c_1}{3} \int_{\mathbb{T}} v^3\, dx - c_1 \int_{\mathbb{T}} v^2\, z\, dx - \frac{c_2}{3} \int_{\mathbb{T}} (\partial_x^{-1} v)\, v^2\, dx - \frac{c_2}{3} \int_{\mathbb{T}} v^2\, (\partial_x^{-1} z)\, dx - \\ &- \frac{2\, c_2}{3} \int_{\mathbb{T}} v\, (\partial_x^{-1} v)\, z\, dx - \frac{c_3}{3} \int_{\mathbb{T}} (\partial_x^{-1} v)^3\, dx - c_3 \int_{\mathbb{T}} (\partial_x^{-1} v)^2\, (\partial_x^{-1} z)\, dx. \end{aligned} \tag{3.16}$$

Thus

$$\begin{aligned} \partial_x \nabla F^{(3)}(u) = &- c_1 \partial_x (v^2) - 2\, c_1 \partial_x \Pi_S[v\, z] + \frac{c_2}{3} \pi_0[v^2] - \frac{c_2}{3} \partial_{xx}[(\partial_x^{-1} v)^2] - \\ &- \frac{2\, c_2}{3} \partial_x \Pi_S[(\partial_x^{-1} v) z + (\partial_x^{-1} z) v] + \frac{2\, c_2}{3} \Pi_S[v\, z] + c_3 \pi_0[(\partial_x^{-1} v)^2] + \\ &+ 2\, c_3\, \Pi_S[(\partial_x^{-1} v)(\partial_x^{-1} z)] \end{aligned} \tag{3.17}$$

where π_0 denotes the projection on the space of functions with zero space average, namely

$$\pi_0[u] = u(x) - \frac{1}{2\pi} \int_{\mathbb{T}} u(x)\, dx.$$

By (3.3), we get

$$\begin{aligned} \nabla H_{3,\leq 1}(u) = &- 3\, c_1 \partial_x (v_x^2) - 6\, c_1 \partial_x \Pi_S[v_x\, z_x] - c_2 \partial_{xx} (v^2) - 2\, c_2 \partial_{xx} \Pi_S[v\, z] + \\ &+ c_2 \pi_0[v_x^2] + 2\, c_2 \Pi_S[v_x\, z_x] + 3\, c_3 \pi_0[v^2] + 6\, c_3 \Pi_S[v\, z]. \end{aligned} \tag{3.18}$$

Hence, by (1.8), (3.17), (3.18), we have

$$
\begin{aligned}
\frac{1}{2}\{H_{3,\le 1}, F^{(3)}\} = & \frac{3\,c_1^2}{2}\int_{\mathbb{T}} \partial_x(v_x^2)\,\partial_x(v^2)\,dx - \frac{c_1\,c_2}{2}\int_{\mathbb{T}} v^2\,\partial_x(v_x^2)\,dx + \frac{c_1\,c_2}{2}\int_{\mathbb{T}} \partial_x(v_x^2)\partial_{xx}[(\partial_x^{-1}v)^2]\,dx \\
& - \frac{3\,c_1\,c_3}{2}\int_{\mathbb{T}} \partial_x(v_x^2)\,(\partial_x^{-1}v)^2 + \frac{c_2^2}{6}\int_{\mathbb{T}} (\partial_x(v^2))^2\,dx + \frac{c_2^2}{6}\int_{\mathbb{T}} \partial_{xx}(v^2)\,\partial_{xx}[(\partial_x^{-1}v)^2]\,dx \\
& - c_2\,c_3\int_{\mathbb{T}} \partial_{xx}(v^2)\,(\partial_x^{-1}v)^2\,dx - \frac{c_1\,c_2}{2}\int_{\mathbb{T}} v_x^2\,\partial_x(v^2)\,dx + \frac{c_2^2}{6}\int_{\mathbb{T}} v_x^2\,\pi_0[v^2]\,dx \\
& - \frac{c_2^2}{6}\int_{\mathbb{T}} v_x^2\,\partial_{xx}[(\partial_x^{-1}v)^2]\,dx + \frac{c_2\,c_3}{2}\int_{\mathbb{T}} v_x^2\,\pi_0[(\partial_x^{-1}v)^2] + \frac{c_2\,c_3}{2}\int_{\mathbb{T}} (\pi_0[v^2])^2\,dx \\
& - \frac{3\,c_3^2}{2}\int_{\mathbb{T}} v^2\,\pi_0[(\partial_x^{-1}v)^2]\,dx + R(v^3\,z) + R(v^2\,z^2).
\end{aligned}
\tag{3.19}
$$

By (3.4), we get

$$
\begin{aligned}
\nabla H_3^{(3)}(u) = & -3\,c_1\,\partial_x(z_x^2) - 6\,c_1\,\partial_x\Pi_S^{\perp}[v_x\,z_x] - c_2\partial_{xx}(z^2) + c_2\pi_0[z_x^2] - 2\,c_2\partial_{xx}\Pi_S^{\perp}[v\,z] + \\
& + 2\,c_2\Pi_S^{\perp}[v_x\,z_x] + 3\,c_3\pi_0[z^2] + 2\,c_3\Pi_S^{\perp}[v\,z].
\end{aligned}
\tag{3.20}
$$

Thus by (1.8), (3.17), (3.20), we have

$$
\begin{aligned}
\{H_3^{(3)}, F^{(3)}\} = & 3\,c_1^2\int_{\mathbb{T}} \partial_x(z_x^2)\,\partial_x(v^2)\,dx - c_1\,c_2\int_{\mathbb{T}} v^2\,\partial_x(z_x^2)\,dx + \\
& + c_1\,c_2\int_{\mathbb{T}} \partial_x(z_x^2)\,\partial_{xx}[(\partial_x^{-1}v)^2]\,dx - 3\,c_1\,c_3\int_{\mathbb{T}} (\partial_x^{-1}v)^2\,\partial_x(z_x^2)\,dx + \\
& + c_1\,c_2\int_{\mathbb{T}} (\partial_x^{-1}v)^2\,\partial_x(z_x^2)\,dx - \frac{c_2^2}{3}\int_{\mathbb{T}} v^2\,\partial_{xx}(z^2)\,dx + \\
& + \frac{c_2^2}{3}\int_{\mathbb{T}} \partial_{xx}(z^2)\,\partial_{xx}[(\partial_x^{-1}v)^2]\,dx - c_2\,c_3\int_{\mathbb{T}} (\partial_x^{-1}v)^2\partial_{xx}(z^2)\,dx - \\
& - c_1\,c_2\int_{\mathbb{T}} z_x^2\,\partial_x(v^2)\,dx + \frac{c_2^2}{3}\int_{\mathbb{T}} z_x^2\,\pi_0[v^2]\,dx - \\
& - \frac{c_2^2}{3}\int_{\mathbb{T}} z_x^2\,\partial_{xx}[(\partial_x^{-1}v)^2]\,dx - c_2\,c_3\int_{\mathbb{T}} z_x^2\,\pi_0[(\partial_x^{-1}v)^2]\,dx - \\
& - 3\,c_1\,c_3\int_{\mathbb{T}} z^2\,\partial_x(v^2)\,dx + c_2\,c_3\int_{\mathbb{T}} v^2\,\pi_0[z^2]\,dx - \\
& - c_2\,c_3\int_{\mathbb{T}} z^2\,\partial_{xx}[(\partial_x^{-1}v)^2]\,dx + 3\,c_3^2\int_{\mathbb{T}} (\partial_x^{-1}v)^2\,\pi_0[z^2]\,dx + \\
& + R(v^3\,z) + R(v\,z^3).
\end{aligned}
\tag{3.21}
$$

Step two. We now construct a symplectic map $\Phi^{(4)}$ to eliminate the term $H_{4,1}^{(3)}$ (which is linear in z) and to normalize $H_{4,0}^{(3)}$ (which is independent of z). We need the following elementary lemma (Lemma 13.4 in [21]).

Lemma 3.3. *Let $j_1, j_2, j_3, j_4 \in \mathbb{Z}$ such that $j_1 + j_2 + j_3 + j_4 = 0$. Then*

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

We look for a map $\Phi^{(4)} := (\Phi^t_{F^{(4)}})_{|t=1}$ which is the time-1 flow map of an auxiliary Hamiltonian

$$
F^{(4)}(u) := \sum_{\substack{j_1+j_2+j_3+j_4=0,\\ \textit{at least 3 indices belong to } S}} F^{(4)}_{j_1 j_2 j_3 j_4}\,u_{j_1}u_{j_2}u_{j_3}u_{j_4},
$$

which has the same form of the Hamiltonian $H^{(3)}_{4,0}+H^{(3)}_{4,1}$. The transformed Hamiltonian is

$$H^{(4)}:=H^{(3)}\circ\Phi^{(4)}=H_2+H^{(3)}_3+H^{(4)}_4+H^{(4)}_{\geq 5},\quad H^{(4)}_4:=\{H_2,F^{(4)}\}+H^{(3)}_4 \tag{3.22}$$

and $H^{(4)}_{\geq 5}$ collects all the terms of order at least five in (v,z). We write

$$H^{(3)}_4(u)=\sum_{j_1+j_2+j_3+j_4=0}H^{(3)}_{4,\,j_1j_2j_3j_4}u_{j_1j_2j_3j_4}. \tag{3.23}$$

This makes sense since $H_{3,\leq 1},H^{(3)}_3$ and $F^{(3)}$ preserve the momentum, hence also $H^{(3)}_4$ does it. We choose the coefficients

$$F^{(4)}_{j_1j_2j_3j_4}:=\begin{cases}\dfrac{H^{(3)}_{4,\,j_1j_2j_3j_4}}{\mathrm{i}(j_1^3+j_2^3+j_3^3+j_4^3)} & \text{if } (j_1,j_2,j_3,j_4)\in\mathcal{A}_4,\\ 0 & \text{otherwise,}\end{cases} \tag{3.24}$$

where

$$\mathcal{A}_4:=\{(j_1,j_2,j_3,j_4)\in\mathbb{Z}^4\setminus\{\mathbf{0}\}\ :j_1+j_2+j_3+j_4=0,j_1^3+j_2^3+j_3^3+j_4^3\neq 0,\\ \text{and at most one among } j_1,j_2,j_3,j_4 \text{ outside } S\}.$$

By this definition, the symmetry of S and the Lemma 3.3, we have $H^{(4)}_{4,1}=0$, because there no exist $j_1,j_2,j_3\in S$ and $j_4\in S^c$ such that $j_1+j_2+j_3+j_4=0$, $j_1^3+j_2^3+j_3^3+j_4^3=0$. By construction, the terms $H^{(4)}_{4,i}=H^{(3)}_{4,i}, i=2,3,4$ are not changed by $\Phi^{(4)}$.
It remains to compute the resonant part of $H^{(3)}_{4,0}$, i.e. the terms of $H^{(3)}_4$ of type $R(v^4)$ supported on the modes (j_1,j_2,j_3,j_4) that do not belong to $\mathcal{A}_4$.
If we call

$$\mathcal{B}:=\{(j_1,j_2,j_3,j_4)\in S^4\ :\ j_1+j_2+j_3+j_4=0,\ j_1^3+j_2^3+j_3^3+j_4^3=0,\ j_1+j_2\neq 0\}$$

then by (3.5), (3.19) we have

$$\begin{aligned}H^{(4)}_{4,0}=&-\frac{3\,c_1^2}{2}\sum_{\mathcal{B}}(j_1+j_2)^2j_3j_4\,u_{j_1}u_{j_2}u_{j_3}u_{j_4}+\frac{c_2^2}{6}\sum_{\mathcal{B}}(j_3+j_4)^2\,u_{j_1}u_{j_2}u_{j_3}u_{j_4}-\frac{c_2^2}{6}\sum_{\mathcal{B}}j_3j_4\,u_{j_1}u_{j_2}u_{j_3}u_{j_4}\\&-\frac{c_2^2}{6}\sum_{\mathcal{B}}(j_1+j_2)^2(j_3+j_4)^2\frac{1}{j_1j_2}\,u_{j_1}u_{j_2}u_{j_3}u_{j_4}+\frac{c_2^2}{6}\sum_{\mathcal{B}}\frac{(j_1+j_2)^2\,j_3j_4}{j_1j_2}\,u_{j_1}u_{j_2}u_{j_3}u_{j_4}\\&+\frac{3}{2}c_3^2\sum_{\mathcal{B}}\frac{1}{\mathrm{i}j_1\,\mathrm{i}j_2}\,u_{j_1}u_{j_2}u_{j_3}u_{j_4}+\frac{c_2\,c_3}{2}\sum_{\mathcal{B}}u_{j_1}u_{j_2}u_{j_3}u_{j_4}-\frac{c_2\,c_3}{2}\sum_{\mathcal{B}}\frac{(j_1+j_2)^2}{j_1j_2}u_{j_1}u_{j_2}u_{j_3}u_{j_4}\\&-\frac{c_2c_3}{2}\sum_{\mathcal{B}}\frac{(j_3+j_4)^2}{j_1\,j_2}\,u_{j_1}u_{j_2}u_{j_3}u_{j_4}+\frac{c_2c_3}{2}\sum_{\mathcal{B}}\frac{j_3j_4}{j_1j_2}\,u_{j_1}u_{j_2}u_{j_3}u_{j_4}+c_4\sum_{\mathcal{B}\cup\{j_1+j_2=0\}}j_1\,j_2\,j_3\,j_4\,u_{j_1}u_{j_2}u_{j_3}u_{j_4}\\&-c_6\sum_{\mathcal{B}\cup\{j_1+j_2=0\}}j_1\,j_2\,u_{j_1}u_{j_2}u_{j_3}u_{j_4}+c_7\sum_{\mathcal{B}\cup\{j_1+j_2=0\}}u_{j_1}u_{j_2}u_{j_3}u_{j_4}.\end{aligned} \tag{3.25}$$

By Lemma 3.3, if $j_1+j_2+j_3+j_4=0,j_1^3+j_2^3+j_3^3+j_4^3=0$ then $(j_1+j_2)(j_1+j_3)(j_2+j_3)=0$. We develop all the sums in (3.25) with respect to the first index j_1. The possible cases are:

(i) $\{j_2\neq -j_1,j_3=-j_1,j_4=-j_2\}$ $\qquad$ (ii) $\{j_2\neq -j_1,j_3\neq -j_1,j_3=-j_2,j_4=-j_1\}$

(iii) $\{j_1+j_2=0\}$.

If $I := (I_{\bar{\jmath}_1}, \dots, I_{\bar{\jmath}_\nu}) \in \mathbb{R}^\nu_+$ with $I_j := |u_j|^2, j \in S$, we get

$$\begin{aligned} H^{(4)}_{4,0}(I) = &- 12\, c_1^2 \sum_{j \in S^+} j^4\, I_j^2 - 24 c_1^2 \sum_{\substack{j,j' \in S^+,\\ j \neq j'}} j^2\, j'^2\, I_j\, I_{j'} - \frac{7c_2^2}{3} \sum_{j \in S^+} j^2\, I_j^2 - \frac{8c_2^2}{3} \sum_{\substack{j,j' \in S^+,\\ j \neq j'}} (j^2 + j'^2)\, I_j\, I_{j'} \\ &- 3\, c_3^2 \sum_{j \in S^+} \frac{1}{j^2}\, I_j^2 - 2\, c_2\, c_3 \sum_{j \in S^+} I_j^2 - 8 c_2 c_3 \sum_{\substack{j,j' \in S^+,\\ j \neq j'}} I_j\, I_{j'} + 6\, c_4 \sum_{j \in S^+} j^4\, I_j^2 + 12 c_4 \sum_{\substack{j,j' \in S,\\ j \neq j'}} j^2\, j'^2\, I_j\, i_{j'} \\ &+ 2 c_6 \sum_{j \in S^+} j^2\, I_j^2 + 2 c_6 \sum_{\substack{j,j' \in S^+,\\ j \neq j'}} (j^2 + j'^2)\, I_j\, I_{j'} + 6\, c_7 \sum_{j \in S^+} I_j^2 + 12 c_7 \sum_{\substack{j,j' \in S^+,\\ j \neq j'}} I_j\, I_{j'} . \end{aligned} \tag{3.26}$$

The Hamiltonian system $H_2 + H_3^{(3)} + H_4^{(4)}$, obtained by truncation at order 4 of the transformed Hamiltonian $H \circ \Phi^{(3)} \circ \Phi^{(4)}$, possesses the invariant submanifold $\{z = 0\}$, and, restricted to this subspace, it is integrable. Indeed, if we introduce on H_S the action-angle variables $u \mapsto (\theta, I)$ by defining

$$u_j := v_j = \sqrt{I_j}\, e^{\mathrm{i}\theta_j}, \qquad I_j = I_{-j}, \quad \theta_{-j} = -\theta_j \quad j \in S, \tag{3.27}$$

the restriction of the Hamiltonian $H_2 + H_3^{(3)} + H_4^{(4)}$ to $\{z = 0\}$, namely $\frac{1}{2} \int v_x^2\, dx + H^{(4)}_{4,0}$, depends only on the actions $I_{\bar{\jmath}_1}, \dots, I_{\bar{\jmath}_\nu}$. We will prove later that, for a generic choice of the tangential sites, this system is also non-isochronous (actually it is formed by ν decoupled oscillators).
Due to the presence of a quadratic nonlinearity in the equation (1.1), we have to eliminate further monomials of $H^{(4)}$ in (3.22) in order to enter in a perturbative regime. Indeed, the minimal requirement for the convergence of the nonlinear Nash-Moser iteration is to eliminate the monomials $R(v^5)$ and $R(v^4\, z)$. Here we need the choice of the sites of Hypotesis ($\mathtt{S}$).

Step three. The homogeneous component of degree five of $H^{(4)}$ has the form

$$H_5^{(4)}(u) = \sum_{j_1 + \cdots + j_5 = 0} H^{(4)}_{5,\, j_1, \dots, j_5}\, u_{j_1} u_{j_2} u_{j_3} u_{j_4} u_{j_5},$$

indeed, the Hamiltonian $H_5^{(4)}$ preserves the momentum, because $f_5(u, u_x)$ does not depend on x (see (1.4)). We want to remove from $H_5^{(4)}$ the terms with at most one index among $j_1, \dots, j_5$ outside S. We consider the auxiliary Hamiltonian

$$F^{(5)} = \sum_{\substack{j_1 + \cdots + j_5 = 0,\\ \textit{at most one index outside } S}} F^{(5)}_{j_1, \dots, j_5}\, u_{j_1} \dots u_{j_5}, \quad F^{(5)}_{j_1, \dots, j_5} := \frac{H^{(4)}_{5,\, j_1, \dots, j_5}}{\mathrm{i}(j_1^3 + \cdots + j_5^3)}. \tag{3.28}$$

Hypotesis ($\mathtt{S}$) implies that

($\mathtt{S}_0$) there is no choice of 5 integers $j_1, \dots, j_5 \in S$ such that

$$j_1 + \cdots + j_5 = 0, \quad j_1^3 + \cdots + j_5^3 = 0, \tag{3.29}$$

($\mathtt{S}_1$) there is no choice of 4 integers $j_1, \dots, j_4 \in S$ and $j_5 \in S^c$ such that (3.29) holds.

Hence $F^{(5)}$ in (3.28) is well defined. Let $\Phi^{(5)}$ be the time-1 flow generated by $X_{F^{(5)}}$. The new Hamiltonian is

$$H^{(5)} := H^{(4)} \circ \Phi^{(5)} = H_2 + H_3^{(3)} + H_4^{(4)} + H_5^{(5)} + H^{(5)}_{\geq 6}, \quad H_5^{(5)} = \{H_2, F^{(5)}\} + H_5^{(4)}, \tag{3.30}$$

where $H^{(5)}_{\geq 6}$ collects all the terms of degree greater or equal than six, and, by the definition of $F^{(5)}$,

$$H_5^{(5)} = \sum_{q=2}^{5} R(v^{5-q} z^q). \tag{3.31}$$

Setting $\Phi_B := \Phi^{(3)} \circ \Phi^{(4)} \circ \Phi^{(5)}$ and renaming $\mathcal{H} := H^{(5)} = H \circ \Phi_B, \mathcal{H}_n = H_n^{(n)}$, by Remark (3.2), we conclude the proof of Proposition 3.1.

4 Action-angle variables

Consider the change of variable $v \mapsto (\theta, I)$ in (3.27), where the actions I are defined in the positive half space $\{v \in \mathbb{R}^\nu : v_i \geq 0, \forall i = 1, \dots, \nu\}$ and $\theta \in \mathbb{T}^\nu$. The symplectic form in (1.7) restricted to the subspace H_S transforms into the 2-form

$$\tilde{\Omega}_S = \sum_{j \in S^+} d\theta_j \wedge \frac{1}{j}\, dI_j. \tag{4.1}$$

Hence the Hamiltonian system $\mathcal{H}_{\leq 5} := H_2 + H_3^{(3)} + H_4^{(4)} + H_5^{(5)}$ restricted to $\{z = 0\}$ writes

$$\begin{cases} \dot{\theta}_j = j\, \dfrac{\partial}{\partial I_j} \mathcal{H}_{\leq 5}(\theta, I, 0), & j \in S^+, \\ \dot{I}_j = -\dfrac{\partial}{\partial \theta_j} \mathcal{H}_{\leq 5}(\theta, I, 0), & j \in S^+. \end{cases} \tag{4.2}$$

We have that

$$\tilde{h}(I) := \mathcal{H}_{\leq 5}(\theta, I, 0) := \sum_{j \in S^+} j^2\, I_j + H_{4,0}^{(4)}(I) \tag{4.3}$$

depends only by the actions I, and, if we call $\omega_j(I) := j\ \partial_{I_j} \tilde{h}(I)$, we have

$$\begin{cases} \dot{\theta}_j = \omega_j(I), & j \in S^+, \\ \dot{I}_j = 0, & j \in S^+. \end{cases} \tag{4.4}$$

By (3.26)

$$\begin{aligned} \omega_j(I) = & j^3 - 24\, c_1^2\, j^5\, I_j - 48 c_1^2\, j^3 \sum_{k \in S^+, k \neq j} k^2 I_k - c_2^2 \frac{14}{3}\, j^3\, I_j - \frac{16}{3} c_2^2\, j^3 \sum_{k \in S^+, k \neq j} I_k \\ & - \frac{16}{3} c_2^2\, j \sum_{k \in S^+, k \neq j} k^2 I_k - 6\, c_3^2\, \frac{1}{j}\, I_j - 2\, c_2\, c_3 j\, I_j - 8 c_2 c_3\, j \sum_{k \in S^+, k \neq j} I_k + 12\, c_4 j^5\, I_j \\ & + 24 c_4\, j^3 \sum_{k \in S^+, k \neq j} k^2\, I_k + 4 c_6 j^3\, I_j + 4 c_6 j^3 \sum_{k \in S^+, k \neq j} I_k + 4 c_6\, j \sum_{k \in S^+, k \neq j} k^2 I_k + 12\, c_7\, j\, I_j \\ & + 24 c_7\, j \sum_{k \in S^+, k \neq j} I_k. \end{aligned} \tag{4.5}$$

Hence, in a small neighbourhood of the origin of the phase space $H_0^1(\mathbb{T}_x)$, the submanifold $\{z = 0\}$ is foliated by invariant tori of amplitude ξ and frequency vector $\omega(\xi) := (\omega_j(\xi))_{j \in S^+}$ as in (4.5).
We shall select from this set of tori the approximately invariant quasi-periodic solutions to be continued and we will use their *unperturbed actions* ξ as parameters. Moreover, we shall require that the frequencies of these tori vary in a one-to-one way with the actions ξ. Thanks to this fact, we could control the conditions that we shall impose on the frequencies ω through the amplitudes, and viceversa.
If we call $\vec{1}$ the vector in $\mathbb{R}^\nu$ with all components equal to 1 and

$$D_S := \mathrm{diag}_{i=1,\dots,\nu}\{\bar{\jmath}_i\}, \qquad v_k := D_S^k\, \vec{1}, \qquad U := \vec{1}^T\, \vec{1} \tag{4.6}$$

then we can write, in a compact form, the vector with components $\omega_j(I)$, with $j \in S^+$, in (4.5), as

$$\omega(\xi) = \overline{\omega} + \mathbb{A}\, \xi, \tag{4.7}$$

where $\overline{\omega}$ is the vector of the linear frequencies (see (1.13)) and

$$\begin{aligned}\mathbb{A} :=\, & (24c_1^2 - 12c_4)D_S^5\{\mathrm{I} - 2D_{-2}UD_S^2\} + (\frac{14}{3}c_2^2 - 4c_6)D_S^3 + (4c_6 - \frac{16}{3}c_2^2)\{D_S^3U + D_SUD_S^2\} \\ & + 12(c_2c_3 - c_7)D_S + (24c_7 - 16c_2c_3)D_SU - 6c_3^2D_S^{-1}.\end{aligned} \tag{4.8}$$

The function of ξ in (4.7) is the *frequency-amplitude map*, which describes, at the main order, how the tangential frequencies are shifted by the amplitudes ξ.
In order to work in a neighbourhood of the unperturbed torus $\{I \equiv D_1\xi\}$ it is advantageous to introduce a set of coordinates $(\theta, y, z) \in \mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp$ adapted to it, defined by

$$\begin{cases} u_j := \sqrt{I_j}\, e^{\mathrm{i}\theta_j}\, e^{\mathrm{i}\,j\,x}, \qquad I_j := |j|(\xi_j + y_j), \qquad & j \in S, \\ u_j := z_j, & j \in S^c, \end{cases} \tag{4.9}$$

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

$$\xi_{-j} = \xi_j, \quad \xi_j > 0, \quad y_{-j} = y_j, \quad \theta_{-j} = -\theta_j, \quad \theta_j \in \mathbb{T},\ y_j \in \mathbb{R}, \quad \forall j \in S. \tag{4.10}$$

For the tangential sites $S^+ := \{\overline{\jmath}_1, \ldots, \overline{\jmath}_\nu\}$ we will also denote

$$\theta_{\overline{\jmath}_i} := \theta_i, \quad y_{\overline{\jmath}_i} := y_i, \quad \xi_{\overline{\jmath}_i} := \xi_i, \quad \omega_{\overline{\jmath}_i} = \omega_i, \quad i = 1, \ldots, \nu.$$

The symplectic 2-form Ω in (1.7) becomes

$$\mathcal{W} := \sum_{i=1}^{\nu} d\theta_i \wedge dy_i + \frac{1}{2}\sum_{j\in S^c} \frac{1}{\mathrm{i}j}\, dz_j \wedge dz_{-j} = \left(\sum_{i=1}^{\nu} d\theta_i \wedge dy_i\right) \oplus \Omega_{S^\perp} = d\Lambda, \tag{4.11}$$

where $\Omega_{S^\perp}$ denotes the restriction of Ω to $H_S^\perp$ and Λ is the contact 1-form on $\mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp$ defined by $\Lambda_{(\theta,y,z)} \colon \mathbb{R}^\nu \times \mathbb{R}^\nu \times H_S^\perp \to \mathbb{R}$,

$$\Lambda_{(\theta,y,z)}[\hat\theta, \hat y, \hat z] := -y \cdot \hat\theta + \frac{1}{2}(\partial_x^{-1} z, \hat z)_{L^2(\mathbb{T})}. \tag{4.12}$$

Working in a neighbourhood of the origin of the phase space, it is convenient to rescale the unperturbed actions ξ and the variables θ, y, z as

$$\xi \mapsto \varepsilon^2\xi, \quad y \mapsto \varepsilon^{2b}y, \quad z \mapsto \varepsilon^b\, z. \tag{4.13}$$

The symplectic form in (4.11) transforms into $\varepsilon^{2b}\,\mathcal{W}$. Hence the Hamiltonian system generated by $\mathcal{H}$ in (3.7) transforms into the new Hamiltonian system

$$\begin{cases} \dot\theta = \partial_y H_\varepsilon(\theta, y, z), \\ \dot y = -\partial_\theta H_\varepsilon(\theta, y, z), \\ \dot z = \partial_x \nabla_z H_\varepsilon(\theta, y, z), \end{cases} \qquad H_\varepsilon := \varepsilon^{-2b}\,\mathcal{H} \circ A_\varepsilon, \tag{4.14}$$

where

$$A_\varepsilon(\theta, y, z) := \varepsilon\, v_\varepsilon(\theta, y) + \varepsilon^b z, \quad v_\varepsilon(\theta, y) := \sum_{j\in S} \sqrt{|j|}\,\sqrt{\xi_j + \varepsilon^{2(b-1)} y_j}\, e^{i\theta_j} e^{ijx}. \tag{4.15}$$

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 defined in (4.17). The quadratic Hamiltonian H_2 in (3.1) becomes

$$\varepsilon^{-2b} H_2 \circ A_\varepsilon = const + \sum_{j\in S^+} j^3\, y_j + \frac{1}{2}\int_{\mathbb{T}} z_x^2\, dx, \tag{4.16}$$

and by (3.1), (3.21) and (3.25) we have (writing $v_\varepsilon := v_\varepsilon(\theta, y)$)

$$
\begin{aligned}
H_\varepsilon(\theta, y, z) &= e(\xi) + \alpha(\xi)\cdot y + \frac{1}{2}\int_{\mathbb{T}} z_x^2\, dx + \varepsilon \int_{\mathbb{T}} (3\, c_1 z_x^2\, (v_\varepsilon)_x + 3\, c_2 z_x^2\, v_\varepsilon + 2\, c_2 (v_\varepsilon)_x z_x z)\, dx \\
&+ \varepsilon^b \int_{\mathbb{T}} \big(c_1\, z_x^3 + c_2\, z_x^2\, z + c_3\, z^3\, dx\big)\, dx + \frac{\varepsilon^{2\,b}}{2}\,\mathbb{M}\, y\cdot y + \varepsilon^{2b}\int_{\mathbb{T}} (c_4\, z_x^4 + c_5\, z_x^3\, z \\
&+ c_6\, z_x^2\, z^2 + c_7\, z^4)\, dx + \varepsilon^2 R((v_\varepsilon(\theta,y))^2 z^2) + \varepsilon^{1+b} R(v_\varepsilon(\theta,y)\, z^3) + \varepsilon^3 R((v_\varepsilon(\theta,y))^3 z^2) \\
&+ \varepsilon^{2+b}\sum_{q=3}^{5} \varepsilon^{(q-3)(b-1)} R((v_\varepsilon(\theta,y))^{5-q} z^q) + \varepsilon^{-2b}\mathcal{H}_{\geq 6}(\varepsilon v_\varepsilon(\theta,y) + \varepsilon^b z)
\end{aligned}
\tag{4.17}
$$

where the function $e(\xi)$ is a constant and

$$
\alpha(\xi) = \overline{\omega} + \varepsilon^2\,\mathbb{M}\,\xi\,, \qquad \mathbb{M} := \mathbb{A}\, D_S \tag{4.18}
$$

is the frequency amplitude-map after the change of coordinates in (4.9) and the rescaling in (4.13). Usually $\mathbb{M}$ is called the *twist matrix* and we note that is symmetric.
We write the Hamiltonian in (4.17), eliminating the constant $e(\xi)$ which is irrelevant for the dynamics, as

$$
\begin{aligned}
& H_\varepsilon = \mathcal{N} + P, \qquad \mathcal{N}(\theta, y, z) = \alpha(\xi)\cdot y + \frac{1}{2}(N(\theta) z, z)_{L^2(\mathbb{T})}, \\
& \frac{1}{2}(N(\theta) z, z)_{L^2(\mathbb{T})} := \frac{1}{2}((\partial_z \nabla H_\varepsilon)(\theta, 0, 0)[z], z)_{L^2(\mathbb{T})} = \frac{1}{2}\int_{\mathbb{T}} z_x^2\, dx + \\
& + \varepsilon \int_{\mathbb{T}} c_1 z_x^2\, (v_\varepsilon)_x(\theta, 0)\, dx + \varepsilon \int_{\mathbb{T}} c_2\, z_x^2\, v_\varepsilon(\theta, 0)\, dx + 2\,\varepsilon\, c_2 \int_{\mathbb{T}} z\, z_x\, (v_\varepsilon)_x(\theta, 0)\, dx + \dots
\end{aligned}
\tag{4.19}
$$

where $\mathcal{N}$ describes the linear dynamics, and $P := H_\varepsilon - \mathcal{N}$ collects the nonlinear perturbative effects.

As we said before, we require that the map (4.18) is a diffeomorphism. This function is affine, thus its invertibility is equivalent to the nondegenerancy (or *twist*) condition

$$
\det \mathbb{M} := \det(D_S)\, \det\left(\frac{\partial^2}{\partial I_j\, I_k}\tilde{h}(I)\right)_{j,k\in\{1,\dots,\nu\}} \det(D_S) \neq 0. \tag{4.20}
$$

Remark 4.1. The inequality (4.20) is equivalent to the classical Kolmogorov condition that requires the invertibility of the Hessian of the Hamiltonian $\tilde{h}$ in (4.3). The presence of the diagonal matrix D_S in (4.20) is due to the symplectic form (1.7) and the choice of the action-angle variables (4.9).

In the following lemma we prove that the condition (4.20) is satisfied for non-resonant coefficients and a generic choice of the tangential sites (see Definition 1.2).

Lemma 4.2. *If the coefficients $c_1, \dots, c_7$ are non-resonant, for a generic choice of the tangential sites $\overline{\jmath}_1, \dots, \overline{\jmath}_\nu$ (see Definition 1.2) the condition (4.20) is satisfied.*

Proof. We write $\mathbb{M} = D_S^{-1}\,\mathbb{B}\, D_S$, with

$$
\begin{aligned}
\mathbb{B} :&= (24c_1^2 - 12c_4) D_S^6\{\mathrm{I} - 2D_{-2} U D_S^2\} + (\frac{14}{3}c_2^2 - 4c_6) D_S^4 \\
&+ (4c_6 - \frac{16}{3}c_2^2)\{D_S^4 U + D_S^2 U D_S^2\} - 6c_3^2 \mathrm{I} + 12(c_2 c_3 - c_7) D_S^2 \\
&+ (24c_7 - 16c_2c_3) D_S^2 U,
\end{aligned}
\tag{4.21}
$$

where I is the identity $\nu\times\nu$ matrix. The determinant of $\mathbb{B}$ is a polynomial in the variables $(\overline{\jmath}_1, \dots, \overline{\jmath}_\nu)$ and, if $c_3 \neq 0$, it is not trivial, namely it is not identically zero. Indeed, the monomial of minimal degree of this polynomial originates from the matrix $6\, c_3^2\, \mathrm{I}$, that is invertible, and so it cannot be naught.

Similarly, if $c_3 = 0$ and $2\,c_1^2 - c_4 \neq 0$ then the monomial of maximal degree, i.e. six, is not zero, beacuse $(24\,c_1^2 - 12\,c_4)\,D_S^6\{\mathrm{I} - 2\,D_S^{-2}\,U\,D_S^2\}$ is invertible.
If $c_3 = 2\,c_1^2 - c_4 = 0$ and $c_7 \neq 0$ then the monomial of minimal degree, i.e. two, is $12\,c_7\,D_S^2\,(2\,U - \mathrm{I})$, that is invertible, indeed

$$(2\,U - \mathrm{I})^{-1} = \mathrm{I} - \frac{2}{2\,\nu + 1}\,U,\}$$

where $2\,\nu + 1 \neq 0$, because $\nu \in \mathbb{N}$. If $c_3 = 2\,c_1^2 - c_4 = c_7 = 0$ then

$$\mathbb{B} = D_S^4\left\{(\frac{14}{3}c_2^2 - 4c_6)\mathrm{I} + (4c_6 - \frac{16}{3}c_2^2)\{U + D_{-2}UD_S^2\}\right\}$$

The matrix $U + D_S^{-2}UD_S^2$ has rank 2 and its image is spanned by the vectors $\vec{1} := (1, \dots, 1)$ and v_{-2} The eigenvalues of this matrix, different from zero, are

$$\lambda_1 := \nu + \sqrt{\left(\sum_{i=1}^{\nu} \bar{\jmath}_i^2\right)\left(\sum_{i=1}^{\nu} \bar{\jmath}_i^{-2}\right)}, \quad \lambda_2 := \nu - \sqrt{\left(\sum_{i=1}^{\nu} \bar{\jmath}_i^2\right)\left(\sum_{i=1}^{\nu} \bar{\jmath}_i^{-2}\right)}. \tag{4.22}$$

Then, if $7c_2^2 - 6\,c_6 \neq 0$ and $\alpha := (8\,c_2^2 - 6\,c_6)/(7\,c_2^2 - 6\,c_6)$, we require that

$$\begin{cases} 1 - \alpha\,\lambda_1 \neq 0, \\ 1 - \alpha\,\lambda_2 \neq 0. \end{cases} \tag{4.23}$$

The conditions (4.23) are satisfied for every choice of the tangential sites if $4\,c_2^2 = 3\,c_6$; otherwise, it is satisfied by generic integer vectors $(\bar{\jmath}_i)_{i=1}^{\nu}$. □

5 The nonliner functional setting

We look for an embedded invariant torus

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

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

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

where α is the function defined in (4.18) and, by Lemma 4.20, it is a diffeomorphism for a generic choice of the tangential sites.

Remark 5.1. We could consider any compact subset of $\{v \in \mathbb{R}^\nu : v_i > 0, \forall i = 1, \dots, \nu\}$ instead of the set $[1, 2]^\nu$ in the definition (5.2).

Since any $\omega \in \Omega_\varepsilon$ is ε^2-close to the integer vector $\overline{\omega} := (\bar{\jmath}_1^3, \dots, \bar{\jmath}_\nu^3) \in \mathbb{N}^\nu$, we require that the constant γ in the diophantine inequality

$$|\omega \cdot l| \geq \gamma\,\langle l\rangle^{-\tau}, \quad \forall l \in \mathbb{Z}^\nu \setminus \{0\} \tag{5.3}$$

satisfies

$$\gamma = \varepsilon^{2+a}, \quad \text{for some } a > 0. \tag{5.4}$$

Note that the definition of γ in (5.4) is slightly stronger than the minimal condition, namely $\gamma \leq c\,\varepsilon^2$, with $c > 0$ small enough. In addition to (5.3) we shall also require that ω satisfies the first and the second order Melnikov non-resonance conditions. We fix the amplitude ξ as a function of ω and ε, as

$$\xi := \varepsilon^{-2}\,\mathbb{M}^{-1}[\omega - \overline{\omega}], \tag{5.5}$$

so that $\alpha(\xi) = \omega$ (see (4.18)). Consequently, H_ε in (4.19) becomes a (ω, ε)-parameter family of Hamiltonians which possess an invariant torus at the origin with frequency vector close to ω.
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

$$H_{\varepsilon,\zeta}(\theta, y, z) := H_\varepsilon(\theta, y, z) + \zeta \cdot \theta, \quad \zeta \in \mathbb{R}^\nu. \tag{5.6}$$

We introduce ζ in order to control the average in the y-component of the linearized equations (6.23) (see (6.26)). However, the vector ζ has 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_ε} itself.
Thus, we look for zeros of the nonlinear operator

$$\mathcal{F}(i, \zeta) := \mathcal{F}(i, \zeta, \omega, \varepsilon) := \mathcal{D}_\omega i(\varphi) - X_{\mathcal{N}}(i(\varphi)) - X_P(i(\varphi)) + (0, \zeta, 0) \tag{5.7}$$

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

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

$$\mathcal{D}_\omega := \omega \cdot \partial_\varphi. \tag{5.8}$$

The Sobolev norm of the periodic component of the embedded torus

$$\mathfrak{I}(\varphi) := i(\varphi) - (\varphi, 0, 0) := (\Theta(\varphi), y(\varphi), z(\varphi)), \tag{5.9}$$

is

$$\|\mathfrak{I}\|_s := \|\Theta\|_{H^s_\varphi} + \|y\|_{H^s_\varphi} + \|z\|_s \tag{5.10}$$

where $\|z\|_s := \|z\|_{H^s_{\varphi,x}}$ is defined in (2.3).
We link the rescaling of the domain of the variables (4.13) with the diophantine constant $\gamma = \varepsilon^{2+a}$ by choosing

$$\gamma = \varepsilon^{2+a} = \varepsilon^{2\,b}, \quad b := 1 + (a/2). \tag{5.11}$$

Other choices are possible (see Remark 5.2 in [4]).

Theorem 5.2. *If $c_1, \ldots, c_7$ are non-resonant and conditions* (C1)-(C2) *hold, then for a generic choice of the tangential sites S, satisfying the assumption* (S)*, for all $\varepsilon \in (0, \varepsilon_0)$, where ε_0 is a positive constant small enough, there exist a constant $C > 0$ and a Cantor-like set $\mathcal{C}_\varepsilon \subseteq \Omega_\varepsilon$ (see (5.2)), with asymptotically full measure as $\varepsilon \to 0$, namely*

$$\lim_{\varepsilon \to 0} \frac{|\mathcal{C}_\varepsilon|}{|\Omega_\varepsilon|} = 1, \tag{5.12}$$

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

$$\|i_\infty(\varphi) - (\varphi, 0, 0)\|^{Lip(\gamma)}_{s_0+\mu} \leq C\, \varepsilon^{6-2b}\, \gamma^{-1} \tag{5.13}$$

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

Theorem 5.2 is proved in Sections 6 – 9. It implies Theorem 1.3 where the ξ_j in (1.18) are the components of the vector $\mathbb{M}^{-1}[\omega - \overline{\omega}]$.

Now we give tame estimates for the composition operator induced by the Hamiltonian vector fields $X_{\mathcal{N}}$ and X_P in (5.7).

Since the functions $y \to \sqrt{\xi + \varepsilon^{2(b-1)}y}, \theta \to e^{\mathrm{i}\,\theta}$ are analytic for ε small enough and $|y| \leq C$, the composition lemma 2.6 implies that, for all $\Theta, y \in H^s(\mathbb{T}^\nu, \mathbb{R}^\nu)$ with $\|\Theta\|_{s_0}, \|y\|_{s_0} \leq 1$, one has the tame estimate

$$\|v_\varepsilon(\theta(\varphi), y(\varphi))\|_s \leq_s 1 + \|\Theta\|_s + \|y\|_s. \tag{5.14}$$

Hence the map A_ε in (4.15) satisfies, for all $\|\mathfrak{I}\|_{s_0}^{Lip(\gamma)} \leq 1$

$$\|A_\varepsilon(\theta(\varphi), y(\varphi), z(\varphi))\|_s^{Lip(\gamma)} \leq_s \varepsilon(1 + \|\mathfrak{I}\|_s^{Lip(\gamma)}). \tag{5.15}$$

In the following lemma we collect tame estimates for the Hamiltonian vector fields $X_{\mathcal{N}}, X_P, X_{H_\varepsilon}$, see (4.19).

Lemma 5.3. *Let $\mathfrak{I}(\varphi)$ in (5.9) satisfy $\|\mathfrak{I}\|_{s_0+3}^{Lip(\gamma)} \leq C\,\varepsilon^{6-2b}\gamma^{-1}$. Then*

$$\|\partial_y P(i)\|_s^{Lip(\gamma)} \leq_s \varepsilon^4 + \varepsilon^{2b}\|\mathfrak{I}\|_{s+3}^{Lip(\gamma)}, \qquad \|\partial_\theta P(i)\|_s^{Lip(\gamma)} \leq_s \varepsilon^{6-2b}(1 + \|\mathfrak{I}\|_{s+3}^{Lip(\gamma)}), \tag{5.16}$$

$$\|\nabla_z P(i)\|_s^{Lip(\gamma)} \leq_s \varepsilon^{5-b} + \varepsilon^{6-b}\gamma^{-1}\|\mathfrak{I}\|_{s+3}^{Lip(\gamma)}, \qquad \|X_P(i)\|_s^{Lip(\gamma)} \leq_s \varepsilon^{6-2b} + \varepsilon^{2b}\|\mathfrak{I}\|_{s+3}^{Lip(\gamma)}, \tag{5.17}$$

$$\|\partial_\theta\partial_y P(i)\|_s^{Lip(\gamma)} \leq_s \varepsilon^4 + \varepsilon^5\gamma^{-1}\|\mathfrak{I}\|_{s+3}^{Lip(\gamma)}, \qquad \|\partial_y\nabla_z P(i)\|_s^{Lip(\gamma)} \leq_s \varepsilon^{b+3} + \varepsilon^{2b-1}\|\mathfrak{I}\|_{s+3}^{Lip(\gamma)}, \tag{5.18}$$

$$\|\partial_{yy}P(i) - \frac{\varepsilon^{2b}}{2}\mathbb{M}\|_s^{Lip(\gamma)} \leq_s \varepsilon^{2+2b} + \varepsilon^{2b+3}\gamma^{-1}\|\mathfrak{I}\|_{s+2}^{Lip(\gamma)} \tag{5.19}$$

and for all $\hat\imath := (\hat\Theta, \hat y, \hat z)$,

$$\|\partial_y d_i X_P(i)[\hat\imath]\|_s^{Lip(\gamma)} \leq_s \varepsilon^{2b-1}(\|\hat\imath\|_{s+3}^{Lip(\gamma)} + \|\mathfrak{I}\|_{s+3}^{Lip(\gamma)}\|\hat\imath\|_{s_0+3}), \tag{5.20}$$

$$\|d_i X_{H_\varepsilon}(i)[\hat\imath] + (0, 0, \partial_{xxx}\hat z)\|_s^{Lip(\gamma)} \leq_s \varepsilon(\|\hat\imath\|_{s+3}^{Lip(\gamma)} + \|\mathfrak{I}\|_{s+3}^{Lip(\gamma)}\|\hat\imath\|_{s_0+3}), \tag{5.21}$$

$$\|d_i^2 X_{H_\varepsilon}(i)[\hat\imath, \hat\imath]\|_s^{Lip(\gamma)} \leq_s \varepsilon(\|\hat\imath\|_{s+3}^{Lip(\gamma)}\|\hat\imath\|_{s_0+3}^{Lip(\gamma)} + \|\mathfrak{I}\|_{s+3}^{Lip(\gamma)}(\|\hat\imath\|_{s_0+3})^2). \tag{5.22}$$

In the sequel we will use that, by the diophantine condition (5.3), the operator $\mathcal{D}_\omega^{-1}$ (see (5.8)) is defined for all functions u with zero φ-average, and satisfies

$$\|\mathcal{D}_\omega^{-1}u\|_s \leq_s \gamma^{-1}\,\|u\|_{s+\tau}, \quad \|\mathcal{D}_\omega^{-1}u\|_s^{Lip(\gamma)} \leq_s \gamma^{-1}\|u\|_{s+2\tau+1}^{Lip(\gamma)}. \tag{5.23}$$

6 Approximate inverse

We will apply a Nash-Moser iterative scheme in order to find a zero of the functional $\mathcal{F}(i,\zeta)$ defined in (5.7). In particular, we shall construct a sequence of approximate solutions of

$$F(i,\zeta) = 0 \tag{6.1}$$

that converges to a solution in some Sobolev norm. In order to define this sequence we need to solve some linearized equations and this is the main difficulty for implementing the Nash-Moser algorithm.
Zehnder noted in [35] that it is sufficient to invert these equations only approximately to get a scheme with still quadratic speed of convergence. We refer to [35] for the precise notion of *approximate right inverse*, whose main feature is to be an *exact right inverse* when the equation is linearized at an exact solution. Hence, our aim is to construct an approximate right inverse of the linearized operator

$$d_{i,\zeta}\mathcal{F}(i_0,\zeta_0)[\hat\imath, \hat\zeta] = \mathcal{D}_\omega\hat\imath - d_i X_{H_\varepsilon}(i_0(\varphi))[\hat\imath] + (0, \hat\zeta, 0) \tag{6.2}$$

at any approximate solution i_0 of the equation (6.1), and to verify that satisfies some tame estimates.
Note that $d_{i,\zeta}\mathcal{F}(i_0,\zeta_0) = d_{i,\zeta}\mathcal{F}(i_0)$ is independent of ζ_0 (see (5.7)).
We will implement the general strategy in [8], [9] which reduces the search of an approximate right inverse of (6.2) 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.[8]), namely the pull-back 1-form $i_0^*\Lambda$ is closed, where Λ is the Liouville 1-form in (4.12). This is tantamount to say that the 2-form $\mathcal{W}$ in (4.11) vanishes on the torus $i_0(\mathbb{T}^\nu)$, because $i_0^*\mathcal{W} = i_0^* d\Lambda = d\, i_0^*\Lambda$. For an "approximately invariant" embedded torus i_0 the 1-form $i_0^*\Lambda$ is only "approximately closed". In order to make this statement quantitative we consider

$$i_0^*\Lambda = \sum_{k=1}^{\nu} a_k(\varphi)\, d\varphi_k, \quad a_k(\varphi) := -([\partial_\varphi \theta_0(\varphi)]^T y_0(\varphi))_k + \frac{1}{2}(\partial_{\varphi_k} z_0(\varphi), \partial_x^{-1} z_0(\varphi))_{L^2(\mathbb{T})} \tag{6.3}$$

and we quantify how small is

$$i_0^*\mathcal{W} = d\, i_0^*\Lambda = \sum_{1\le k<j\le \nu} A_{k\,j}(\varphi)\, d\varphi_k \wedge d\varphi_j, \quad A_{k\,j}(\varphi) := \partial_{\varphi_k} a_j(\varphi) - \partial_{\varphi_j} a_k(\varphi). \tag{6.4}$$

In order to get estimates for an approximate inverse we need to take in account the size of the "error" function

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

which gives a measure of how i_0 is near to be an exact solution.
Along this section we will always assume the following hypotesis (which will be proved at each step of the Nash-Moser iteration):

- **Assumption**. The map $\omega \mapsto i_0(\omega)$ is a Lipschitz function defined on some subset $\Omega_0 \subseteq \Omega_\varepsilon$, where Ω_ε is defined in (5.2), and, for some $\mu := \mu(\tau, \nu) > 0$,

$$\|\mathfrak{I}_0\|_{s_0+\mu}^{Lip(\gamma)} \le \varepsilon^{6-2b}\gamma^{-1}, \quad \|Z\|_{s_0+\mu}^{Lip(\gamma)} \le \varepsilon^{6-2b}, \quad \gamma = \varepsilon^{2+a}, \quad a \in (0, 1/6), \tag{6.6}$$

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

The next lemma proves that if i_0 is a solution of the equation (6.1), then the parameter ζ has to be naught, hence the embedded torus i_0 supports a quasi-periodic solution of the "original" system with Hamiltonian H_ε.

Lemma 6.1. *(Lemma 6.1 in [3]) We have*

$$|\zeta_0|^{Lip(\gamma)} \le C\|Z\|_{s_0}^{Lip(\gamma)}.$$

In particular, if $\mathcal{F}(i_0, \zeta_0) = 0$ then $\zeta_0 = 0$ and the torus $i_0(\varphi)$ is invariant for the vector field X_{H_ε}.

Now we estimate the size of $i_0^*\mathcal{W}$ in terms of the error function Z.
By (6.3), (6.4) we get

$$\|A_{k\,j}\|_s^{Lip(\gamma)} \le_s \|\mathfrak{I}_0\|_{s+2}^{Lip(\gamma)}.$$

Moreover, we have the following bound.

Lemma 6.2. *(Lemma 6.2 in [3]) The coefficients $A_{k\,j}(\varphi)$ in (6.4) satisfy*

$$\|A_{k\,j}\|_s^{Lip(\gamma)} \le_s \gamma^{-1}(\|Z\|_{s+2\tau+2}^{Lip(\gamma)} + \|Z\|_{s_0+1}^{Lip(\gamma)}\|\mathfrak{I}_0\|_{s+2\tau+2}^{Lip(\gamma)}). \tag{6.7}$$

As in [8], the idea is to analyze the operator linearized at an isotropic embedded torus i_δ, because the isotropy of the torus allows to construct a symplectic set of coordinates around it for which the linear tangential dynamic and the normal one are decoupled. Thus, the linear system becomes "triangular" and the hard part is to solve the equation in the normal directions (see Section 7).
Now we see that we can slightly modify i_0 (indeed, it is sufficient to move the y-component only) to obtain an isotropic torus i_δ, that is an approximate solution as well as i_0. At the end of this section, we will prove that we are able to construct an approximate right inverse of (6.2) starting from an approximate inverse of $d_{i,\zeta}\mathcal{F}(i_\delta, \zeta_0)[\hat{\imath}, \hat{\zeta}]$.

In the paper we denote equivalently the differential ∂_i or d_i. We use the notation $\Delta_\varphi := \sum_{k=1}^{\nu} \partial_{\varphi_k}^2$ and we denote by $\sigma := \sigma(\nu, \tau)$ possibly different (larger) "loss of derivatives" constants.

Lemma 6.3. *(**Isotropic torus**)(Lemma 6.3 in [3]) The torus $i_\delta = (\theta_0(\varphi), y_\delta(\varphi), z_0(\varphi))$ defined by*

$$y_\delta := y_0 + [\partial_\varphi \theta_0(\varphi)]^{-T} \rho(\varphi), \quad \rho_j(\varphi) := \Delta_\varphi^{-1} \sum_{k=1}^{\nu} \partial_{\varphi_j} A_{k\,j}(\varphi), \tag{6.8}$$

is isotropic. If (6.6) holds, then, for some $\sigma := \sigma(\nu, \tau)$,

$$\|y_\delta - y_0\|_s^{Lip(\gamma)} \le_s \gamma^{-1}(\|Z\|_{s+\sigma}^{Lip(\gamma)} \|\mathfrak{I}_0\|_{s_0+\sigma}^{Lip(\gamma)} + \|Z\|_{s_0+\sigma}^{Lip(\gamma)} \|\mathfrak{I}_0\|_{s+\sigma}^{Lip(\gamma)}), \tag{6.9}$$

$$\|\mathcal{F}(i_\delta, \zeta_0)\|_s^{Lip(\gamma)} \le_s \|Z\|_{s+\sigma}^{Lip(\gamma)} + \|Z\|_{s_0+\sigma}^{Lip(\gamma)} \|\mathfrak{I}_0\|_{s+\sigma}^{Lip(\gamma)}, \tag{6.10}$$

$$\|\partial_i i_\delta[\hat{\imath}]\|_s \le_s \|\hat{\imath}\|_s + \|\mathfrak{I}_0\|_{s+\sigma} \|\hat{\imath}\|_s. \tag{6.11}$$

We introduce a set of symplectic coordinates adapted to the isotropic torus i_δ. We consider the map $G_\delta \colon (\Psi, \eta, w) \to (\theta, y, z)$ of the phase space $\mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp$ defined by

$$\begin{pmatrix} \theta \\ y \\ z \end{pmatrix} := G_\delta \begin{pmatrix} \psi \\ \eta \\ w \end{pmatrix} := \begin{pmatrix} \theta_0(\psi) \\ y_\delta(\psi) + [\partial_\psi \theta_0(\psi)]^{-T} \eta + [(\partial_\theta \tilde{z}_0)(\theta_0(\psi))]^T \partial_x^{-1} w \\ z_0(\psi) + w \end{pmatrix} \tag{6.12}$$

where $\tilde{z}_0 := z_0(\theta_0^{-1}(\theta))$ (indeed $\theta_0 \colon \mathbb{T}^\nu \to \mathbb{T}^\nu$ is a diffeomorphism, because $\theta_0(\varphi) - \varphi$ is small). It is proved in [8] (Lemma 6.3) that G_δ in (6.12) is symplectic, using that the torus i_δ is isotropic. In the new coordinates, i_δ is at the origin, i.e. $(\psi, \eta, w) = (\psi, 0, 0)$. The transformed Hamiltonian $K := K(\psi, \eta, w, \zeta_0)$ is (recall (5.6))

$$\begin{aligned} K := H_{\varepsilon, \zeta_0} \circ G_\delta = \theta_0(\psi) \cdot \zeta_0 + K_{00}(\psi) + K_{10}(\psi) \cdot \eta + (K_{01}(\psi), w)_{L^2(\mathbb{T})} + \frac{1}{2} K_{20}(\psi) \eta \cdot \eta + \\ + (K_{11}(\psi)\eta, w)_{L^2(\mathbb{T})} + \frac{1}{2} (K_{02}(\psi) w, w)_{L^2(\mathbb{T})} + K_{\ge 3}(\psi, \eta, w) \end{aligned} \tag{6.13}$$

where $K_{\ge 3}$ collects the terms at least cubic in the variables (η, w). At any fixed ψ, the Taylor coefficient $K_{00}(\psi) \in \mathbb{R}, K_{10}(\psi) \in \mathbb{R}^\nu, K_{01}(\psi) \in H_S^\perp, K_{20}(\psi)$ is a $\nu \times \nu$ real matrix, $K_{02}(\psi)$ is a linear self-adjoint operator of $H_S^\perp$ and $K_{11}(\psi) \colon \mathbb{R}^\nu \to H_S^\perp$.
Note that the above Taylor coefficients do not depend on the parameter ζ_0.
The Hamilton equations associated to (6.13) are

$$\begin{cases} \dot{\psi} = K_{10}(\psi) + K_{20}(\psi)\eta + K_{11}^T(\psi) w + \partial_\eta K_{\ge 3}(\psi, \eta, w) \\ \dot{\eta} = -\,[\partial_\psi \theta_0(\psi)]^T \zeta_0 - \partial_\psi K_{00}(\psi) - [\partial_\psi K_{10}(\psi)]^T \eta - [\partial_\psi K_{01}(\psi)]^T w - \\ \qquad - \partial_\psi \left(\frac{1}{2} K_{20}(\psi)\eta \cdot \eta + (K_{11}(\psi)\eta, w)_{L^2(\mathbb{T})} + \frac{1}{2}(K_{02}(\psi) w, w)_{L^2(\mathbb{T})} + K_{\ge 3}(\psi, \eta, w) \right) \\ \dot{w} = \partial_x (K_{01}(\psi) + K_{11}(\psi)\eta + K_{02}(\psi) w + \nabla_w K_{\ge 3}(\psi, \eta, w)) \end{cases} \tag{6.14}$$

where $[\partial_\psi K_{10}(\psi)]^T$ is the $\nu \times \nu$ transposed matrix and $[\partial_\psi K_{01}(\psi)]^T, K_{11}^T(\psi) \colon H_S^\perp \to \mathbb{R}^\nu$ are defined by the duality relation

$$(\partial_\psi K_{01}(\psi)[\hat{\psi}], w)_{L^2(\mathbb{T})} = \hat{\psi} \cdot [\partial_\psi K_{01}(\psi)]^T w, \quad \forall \hat{\psi} \in \mathbb{R}^\nu, w \in H_S^\perp,$$

and similarly for K_{11}. Explicitly, for all $w \in H_S^\perp$, and denoting $\underline{e}_k$ the k-th versor of $\mathbb{R}^\nu$,

$$K_{11}^T(\psi) w = \sum_{k=1}^{\nu} (K_{11}^T(\psi) w \cdot \underline{e}_k)\, \underline{e}_k = \sum_{k=1}^{\nu} (w, K_{11}(\psi) \underline{e}_k)_{L^2(\mathbb{T})} \underline{e}_k \in \mathbb{R}^\nu. \tag{6.15}$$

In the next lemma we estimate the coefficients K_{00}, K_{10}, K_{01} in the Taylor expansion (6.13). The term K_{10} describes how the tangential frequencies vary with respect to ω. Note that on an exact solution (i_0, ζ_0) we have $K_{00}(\psi) = const, K_{10} = \omega$ and $K_{01} = 0$.

Lemma 6.4. *(Lemma 6.4 in [3]) Assume (6.6). Then there is $\sigma := \sigma(\tau,\nu)$ such that*

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

Remark 6.5. By Lemma 6.1 if $\mathcal{F}(i_0,\zeta_0)=0$ and, by Lemma 6.4, the Hamiltonian (6.13) simplifies to

$$K = const + \omega\cdot\eta + \frac{1}{2}K_{20}(\psi)\eta\cdot\eta + (K_{11}(\psi)\eta, w)_{L^2(\mathbb{T})} + \frac{1}{2}(K_{02}(\psi)w,w)_{L^2(\mathbb{T})} + K_{\geq 3}. \tag{6.16}$$

In general, the normal form (6.16) provides a control of the linearized equations in the normal bundle of the torus.

We now estimate K_{20}, K_{11} in (6.13). The norm of K_{20} is the sum of the norms of its matrix entries.

Lemma 6.6. *(Lemma 6.6 in [3]) Assume (6.6). Then for some $\sigma := \sigma(\nu,\tau)$ we have*

$$\|K_{20} - \frac{\varepsilon^{2b}}{2}\mathbb{M}\|_s^{Lip(\gamma)} \leq_s \varepsilon^{2b+2} + \varepsilon^{2b}\|\mathfrak{I}_0\|_{s+\sigma}^{Lip(\gamma)} + \varepsilon^3\gamma^{-1}\|\mathfrak{I}_0\|_{s_0+\sigma}^{Lip(\gamma)}\|Z\|_{s+\sigma}^{Lip(\gamma)}, \tag{6.17}$$

$$\|K_{11}\eta\|_s^{Lip(\gamma)} \leq_s \varepsilon^5\gamma^{-1}\|\eta\|_s^{Lip(\gamma)} + \varepsilon^{2b-1}(\|\mathfrak{I}_0\|_{s+\sigma}^{Lip(\gamma)} + \gamma^{-1}\|\mathfrak{I}_0\|_{s_0+\sigma}^{Lip(\gamma)}\|Z\|_{s+\sigma}^{Lip(\gamma)})\|\eta\|_{s_0}^{Lip(\gamma)}, \tag{6.18}$$

$$\|K_{11}^T w\| \leq_s \varepsilon^5\gamma^{-1}\|w\|_{s+2}^{Lip(\gamma)} + \varepsilon^{2b-1}(\|\mathfrak{I}_0\|_{s+\sigma}^{Lip(\gamma)} + \gamma^{-1}\|\mathfrak{I}_0\|_{s_0+\sigma}^{Lip(\gamma)}\|Z\|_{s+\sigma}^{Lip(\gamma)})\|w\|_{s_0+2}^{Lip(\gamma)}. \tag{6.19}$$

In particular

$$\|K_{20} - \frac{\varepsilon^{2b}}{2}\mathbb{M}\|_{s_0}^{Lip(\gamma)} \leq \varepsilon^6\gamma^{-1}, \quad \|K_{11}\eta\|_{s_0}^{Lip(\gamma)} \leq \varepsilon^5\gamma^{-1}\|\eta\|_{s_0}^{Lip(\gamma)}, \quad \|K_{11}^T w\|_{s_0}^{Lip(\gamma)} \leq \varepsilon^5\gamma^{-1}\|w\|_{s_0}^{Lip(\gamma)}.$$

We apply the linear change of variables

$$DG_\delta(\varphi,0,0)\begin{pmatrix}\hat\psi\\ \hat\eta\\ \hat w\end{pmatrix} := \begin{pmatrix}\partial_\psi\theta_0(\varphi) & 0 & 0\\ \partial_\psi y_\delta(\varphi) & [\partial_\psi\theta_0(\varphi)]^{-T} & -[(\partial_\theta\tilde z_0)(\theta_0(\varphi))]^T\partial_x^{-1}\\ \partial_\psi z_0(\varphi) & 0 & \mathrm{I}\end{pmatrix}\begin{pmatrix}\hat\psi\\ \hat\eta\\ \hat w.\end{pmatrix} \tag{6.20}$$

In these new coordinates the linearized operator $d_{i,\zeta}\mathcal{F}(i_\delta,\zeta_0)$ is "approximately" the operator obtained linearizing (6.14) at $(\psi,\eta,w,\zeta)=(\varphi,0,0,\zeta_0)$ with $\mathcal{D}_\omega$ instead of ∂_t, namely

$$\begin{pmatrix}\mathcal{D}_\omega\hat\psi - \partial_\psi K_{10}(\varphi)[\hat\psi] - K_{20}(\varphi)\hat\eta - K_{11}^T(\varphi)\hat w\\ \mathcal{D}_\omega\hat\eta + [\partial_\psi\theta_0(\varphi)]^T\hat\zeta + \partial_\psi[\partial_\psi\theta_0(\varphi)]^T[\hat\psi,\zeta_0] + \partial_{\psi\psi}K_{00}(\varphi)[\hat\psi] + [\partial_\psi K_{10}(\varphi)]^T\hat\eta + [\partial_\psi K_{01}(\varphi)]^T\hat w\\ \mathcal{D}_\omega\hat w - \partial_x\{\partial_\psi K_{01}(\varphi)[\hat\psi] + K_{11}(\varphi)\hat\eta + K_{02}(\varphi)\hat w\}.\end{pmatrix} \tag{6.21}$$

We give estimate on the composition operator induced by the transformation (6.20).

Lemma 6.7. *(Lemma 6.7 in [3]) Assume (6.6) and let $\hat\imath := (\hat\psi,\hat\eta,\hat w)$. Then, for some $\sigma := \sigma(\tau,\nu)$, we have*

$$\begin{aligned}&\|DG_\delta(\varphi,0,0)[\hat\imath]\|_s + \|DG_\delta(\varphi,0,0)^{-1}[\hat\imath]\|_s \leq_s \|\hat\imath\|_s + (\|\mathfrak{I}_0\|_{s+\sigma} + \gamma^{-1}\|\mathfrak{I}_0\|_{s+\sigma}^{Lip(\gamma)}\|Z\|_{s+\sigma})\|\hat\imath\|_{s_0}\\ &\|D^2G_\delta(\varphi,0,0)[\hat\imath_1,\hat\imath_2]\|_s \leq_s \|\hat\imath_1\|_s\|\hat\imath_2\|_{s_0} + \|\hat\imath_1\|_{s_0}\|\hat\imath\|_s + (\|\mathfrak{I}_0\|_{s+\sigma} + \gamma^{-1}\|\mathfrak{I}_0\|_{s_0+\sigma}\|Z\|_{s+\sigma})\|\hat\imath\|_{s_0}\|\hat\imath_2\|_{s_0}.\end{aligned} \tag{6.22}$$

Moreover the same estimates hold if we replace $\|\cdot\|_s$ with $\|\cdot\|_s^{Lip(\gamma)}$.

In order to construct an approximate inverse of (6.21) it is sufficient to solve the system of equations

$$\mathbb{D}[\hat\psi,\hat\eta,\hat w,\hat\zeta] := \begin{pmatrix}\mathcal{D}_\omega\hat\psi - K_{20}(\varphi)\hat\eta - K_{11}^T(\varphi)\hat w\\ \mathcal{D}_\omega\hat\eta + [\partial_\psi\theta_0(\varphi)]^T\hat\zeta\\ \mathcal{D}_\omega\hat w - \partial_x K_{11}(\varphi)\hat\eta - \partial_x K_{02}(\varphi)\hat w\end{pmatrix} = \begin{pmatrix}g_1\\ g_2\\ g_3\end{pmatrix} \tag{6.23}$$

which is obtained by (6.21) neglecting the terms that are naught at a solution, namely, by Lemmata (6.1) and (6.4), $\partial_\psi K_{10}, \partial_{\psi\psi}K_{00}, \partial_\psi K_{00}, \partial_\psi K_{01}$ and $\partial_\psi[\partial_\psi\theta_0(\varphi)]^T[\cdot,\zeta_0]$.

Remark 6.8. We will use the following notations for the averages of a function $v(\varphi, x)$

$$M_x[v] := \frac{1}{2\pi}\int_{\mathbb{T}} v(\varphi, x)\, dx, \quad M_\varphi[v] := \frac{1}{(2\pi)^\nu}\int_{\mathbb{T}^\nu} v(\varphi, x)\, d\varphi \tag{6.24}$$

and $M_{\varphi,x}[v] := M_x[M_\varphi[v]] = M_\varphi[M_x[v]]$.

First, we solve the second equation, namely

$$\mathcal{D}_\omega \hat{\eta} = g_2 - [\partial_\psi \theta_0(\varphi)]\hat{\zeta}. \tag{6.25}$$

We choose $\hat{\zeta}$ so that the φ-average of the right hand side of (6.25) is zero, namely

$$\hat{\zeta} = M_\varphi[g_2]. \tag{6.26}$$

Note that the φ-averaged matrix $M_\varphi[(\partial_\psi \theta_0)^T] = M_\varphi[\mathrm{I} + (\partial_\psi \Theta_0)^T] = \mathrm{I}$ because $\theta_0(\varphi) = \varphi + \Theta_0(\varphi)$ and $\Theta_0(\varphi)$ is periodic. Therefore

$$\hat{\eta} = \mathcal{D}_\omega^{-1}(g_2 - [\partial_\psi \theta_0(\varphi)]^T M_\varphi[g_2]) + M_\varphi[\hat{\eta}], \qquad M_\varphi[\hat{\eta}] \in \mathbb{R}^\nu, \tag{6.27}$$

where the average $M_\varphi[\hat{\eta}]$ will be fix when we deal with the first equation.
We now analyze the third equation, namely

$$\mathcal{L}_\omega \hat{w} = g_3 + \partial_x K_{11}(\varphi)\hat{\eta}, \quad \mathcal{L}_\omega := \omega\cdot\partial_\varphi - \partial_x K_{02}(\varphi). \tag{6.28}$$

If we fix $\hat{\eta}$, then solving the equation (6.28) is tantamount to invert the operator $\mathcal{L}_\omega$. For the moment we assume the following hypotesis (that will be proved in Section 8)

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

$$\|\mathcal{L}_\omega^{-1} g\|_s^{Lip(\gamma)} \le_s \gamma^{-1}(\|g\|_{s+\mu}^{Lip(\gamma)} + \varepsilon\gamma^{-1}\{\|\mathfrak{I}_0\|_{s+\mu}^{Lip(\gamma)} + \gamma^{-1}\|\mathfrak{I}_0\|_{s+\mu}^{Lip(\gamma)}\|Z\|_{s+\mu}^{Lip(\gamma)}\}\|g\|_{s_0}^{Lip(\gamma)}) \tag{6.29}$$

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

Remark 6.9. The term $\varepsilon\gamma^{-1}\|\mathfrak{I}_0\|_{s+\mu}^{Lip(\gamma)}$ arises because the remainder $\mathcal{R}_6$ in Section 8 contains the term $\varepsilon(\|\Theta_0\|_{s+\mu}^{Lip(\gamma)} + \|y_\delta\|_{s+\mu}^{Lip(\gamma)}) \le_s \varepsilon\|\mathfrak{I}_0\|_{s+\mu}^{Lip(\gamma)}$, see Lemma 8.18.
These big constants coming from the tame estimates for the inverse of the linearized operators at any approximate solution will be dominated by the quadraticity of the Nash-Moser scheme.

By the above assumption, there exists a solution of (6.28)

$$\hat{w} = \mathcal{L}_\omega^{-1}[g_3 + \partial_x K_{11}(\varphi)\hat{\eta}]. \tag{6.30}$$

Now consider the first equation

$$\mathcal{D}_\omega \hat{\psi} = g_1 + K_{20}\hat{\eta} - K_{11}^T(\varphi)\hat{w}. \tag{6.31}$$

Substituting (6.27), (6.30) in the equation (6.31), we get

$$\mathcal{D}_\omega \hat{\psi} = g_1 + M_1(\varphi)M_\varphi[\hat{\eta}] + M_2(\varphi)g_2 + M_3(\varphi)g_3 - M_2(\varphi)[\partial_\psi\theta_0]^T M_\varphi[g_2], \tag{6.32}$$

where

$$M_1(\varphi) := K_{20}(\varphi) + K_{11}^T(\varphi)\mathcal{L}_\omega^{-1}\partial_x K_{11}(\varphi), \quad M_2(\varphi) := M_1(\varphi)\mathcal{D}_\omega^{-1}, \quad M_3(\varphi) := K_{11}^T(\varphi)\mathcal{L}_\omega^{-1}. \tag{6.33}$$

In order to solve the equation (6.32) we have to choose $M_\varphi[\hat{\eta}]$ such that the right hand side in (6.32) has zero φ-average.

By Lemma 6.6 and (6.6), the φ-averaged matrix $M_\varphi[M_1] = \varepsilon^{2b} M + O(\varepsilon^{10}\gamma^{-3})$. Therefore, for ε small, $M_\varphi[M_1]$ is invertible and $M_\varphi[M_1]^{-1} = O(\varepsilon^{-2b}) = O(\gamma^{-1})$. Thus we define

$$M_\varphi[\hat\eta] := -(M_\varphi[M_1])^{-1}\{M_\varphi[g_1] + M_\varphi[M_2 g_2] + M_\varphi[M_3 g_3] - M_\varphi[M_2(\partial_\psi\theta_0)^T]\, M_\varphi[g_2]\}. \quad (6.34)$$

With this choice of $M_\varphi[\hat\eta]$ the equation (6.32) has the solution

$$\hat\psi := \mathcal{D}_\omega^{-1}\{g_1 + M_1(\varphi) M_\varphi[\hat\eta] + M_2(\varphi) g_2 + M_3(\varphi) g_3 - M_2(\varphi)[\partial_\psi\theta_0]^T M_\varphi[g_2]\}. \quad (6.35)$$

In conclusion, we have constructed a solution $(\hat\psi, \hat\eta, \hat w, \hat\zeta)$ of the linear system (6.23). We resume this in the following proposition, giving also estimates on the inverse of the operator $\mathbb{D}$ defined in (6.23).

Proposition 6.10. *(Proposition* 6.9 *in [3]) Assume* (6.6) *and* (6.29)*. Then, for all $\omega \in \Omega_\infty$, for all $g := (g_1, g_2, g_3)$, the system* (6.23) *has a solution $\mathbb{D}^{-1} g := (\hat\psi, \hat\eta, \hat w, \hat\zeta)$ where $(\hat\psi, \hat\eta, \hat w, \hat\zeta)$ are defined in* (6.35)*,* (6.27)*,* (6.30)*,* (6.26)*. Moreover, we have*

$$\|\mathbb{D}^{-1} g\|_s^{Lip(\gamma)} \leq_s \gamma^{-1}(\|g\|_{s+\mu}^{Lip(\gamma)} + \varepsilon\gamma^{-1}\{\mathfrak{I}_0\|_{s+\mu}^{Lip(\gamma)} + \gamma^{-1}\|\mathfrak{I}_0\|_{s_0+\mu}^{Lip(\gamma)}\|\mathcal{F}(i_0,\zeta_0)\|_{s+\mu}^{Lip(\gamma)}\}\|g\|_{s_0+\mu}^{Lip(\gamma)}). \quad (6.36)$$

Eventually we prove that the operator

$$\mathbf{T}_0 := (D\tilde G_\delta)(\varphi, 0, 0) \circ \mathbb{D}^{-1} \circ (DG_\delta(\varphi, 0, 0))^{-1} \quad (6.37)$$

is an approximate right inverse of $d_{i,\zeta}\mathcal{F}(i_0)$ where $\tilde G_\delta((\psi, \eta, w), \zeta)$ is the identity on the ζ-component. We denote the norm $\|(\psi, \eta, w, \zeta)\|_s^{Lip(\gamma)} := \max\{\|(\psi, \eta, w)\|, |\zeta|^{Lip(\gamma)}\}$.

Theorem 6.11. *(Theorem* 6.10 *in [3]) Assume* (6.6) *and the inversion assumption* (6.29)*. Then there exists $\mu := \mu(\tau, \nu)$ such that, for all $\omega \in \Omega_\infty$, for all $g := (g_1, g_2, g_3)$, the operator $\mathbf{T}_0$ defined in* (6.37) *satisfies*

$$\|\mathbf{T}_0 g\|_s^{Lip(\gamma)} \leq_s \gamma^{-1}(\|g\|_{s+\mu}^{Lip(\gamma)} + \varepsilon\gamma^{-1}\{\|\mathfrak{I}_0\|_{s+\mu}^{Lip(\gamma)} + \gamma^{-1}\|\mathfrak{I}_0\|_{s_0+\mu}^{Lip(\gamma)}\|\mathcal{F}(i_0,\zeta_0)\|_{s+\mu}^{Lip(\gamma)}\}\|g\|_{s_0+\mu}^{Lip(\gamma)}). \quad (6.38)$$

It is an approximate inverse of $d_{i,\zeta}\mathcal{F}(i_0)$, namely

$$\begin{aligned}&\|(d_{i,\zeta}\mathcal{F}(i_0) \circ \mathbf{T}_0 - \mathrm{I})g\|_s^{Lip(\gamma)} \leq_s \\ &\leq_s \gamma^{-1}\Big(\|\mathcal{F}(i_0,\zeta_0)\|_{s_0+\mu}^{Lip(\gamma)}\|g\|_{s+\mu}^{Lip(\gamma)} + \{\|\mathcal{F}(i_0,\zeta_0)\|_{s+\mu}^{Lip(\gamma)} + \varepsilon\gamma^{-1}\|\mathcal{F}(i_0,\zeta_0)\|_{s_0+\mu}^{Lip(\gamma)}\|\mathfrak{I}_0\|_{s+\mu}^{Lip(\gamma)}\}\|g\|_{s_0+\mu}^{Lip(\gamma)}\Big).\end{aligned} \quad (6.39)$$

7 The linearized operator in the normal directions

In this section we give an explicit expression of the linearized operator

$$\mathcal{L}_\omega := \omega \cdot \partial_\varphi - \partial_x K_{02}(\varphi). \quad (7.1)$$

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.
First we recall some preliminary lemmata.

Lemma 7.1. *(Lemma* 7.1 *in [3]) 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* (3.2)*. Then*

$$\partial_u[\nabla(H \circ \Phi)](u)[h] = (\partial_u \nabla H)(\Phi(u))[h] + \mathcal{R}(u)[h], \quad (7.2)$$

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

$$\mathcal{R}(u)[h] = \sum_{|j| \leq C}(h, g_j(u))_{L^2(\mathbb{T})}\chi_j(u) \quad (7.3)$$

with $\chi_j(u) = e^{\mathrm{i}jx}$ *or* $g_j(u) = e^{\mathrm{i}jx}$. *The remainder in* (7.3) *is*

$$\mathcal{R}(u) = \mathcal{R}_0(u) + \mathcal{R}_1(u) + \mathcal{R}_2(u)$$

with

$$\begin{aligned} &\mathcal{R}_0(u) := (\partial_u \nabla H)(\Phi(u))\partial_u \Psi(u), \quad \mathcal{R}_1(u) := [\partial_u \{\Psi'(u)^T\}][\cdot, \nabla H(\Phi(u))], \\ &\mathcal{R}_2(u) := [\partial_u \Psi(u)]^T (\partial_u \nabla H)(\Phi(u))\partial_u \Phi(u). \end{aligned} \tag{7.4}$$

Lemma 7.2. *(Lemma 7.3 in [3]) Let* $\mathcal{R}$ *be an operator of the form*

$$\mathcal{R}h = \sum_{|j|\le C} \int_0^1 (h, g_j(\tau))_{L^2(\mathbb{T})} \chi_j(\tau)\, d\tau, \tag{7.5}$$

where the functions $g_j(\tau), \chi_j(\tau) \in H^s, \tau \in [0,1]$ *depend in a Lipschitz way on the parameter* ω. *Then its matrix s-decay norm (see* (2.6)-(2.7)*) satisfies*

$$|\mathcal{R}|_s^{Lip(\gamma)} \le_s \sum_{|j|\le C} \sup_{\tau\in[0,1]} (\|\chi_j(\tau)\|_s^{Lip(\gamma)} \|g_j\|_{s_0}^{Lip(\gamma)} + \|\chi_j(\tau)\|_{s_0}^{Lip(\gamma)} \|g_j(\tau)\|_s^{Lip(\gamma)}). \tag{7.6}$$

7.1 Composition with the map G_δ

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

$$\|\mathfrak{I}_\delta\|_{s_0+\mu}^{Lip(\gamma)} \le C\,\varepsilon^{6-2b}\gamma^{-1}. \tag{7.7}$$

We now study the Hamiltonian $K := H_\varepsilon \circ G_\delta = \varepsilon^{-2b}\mathcal{H} \circ A_\varepsilon \circ G_\delta$ (see (4.19)). Recalling (4.15), $A_\varepsilon \circ G_\delta$ has the form

$$A_\varepsilon(G_\delta(\psi,\eta,w)) = \varepsilon v_\varepsilon(\theta_0(\psi), y_\delta(\psi) + L_1(\psi)\eta + L_2(\psi)w) + \varepsilon^b(z_0(\psi) + w) \tag{7.8}$$

where

$$L_1(\Psi) := [\partial_\psi \theta_0(\psi)]^{-T}, \quad L_2(\psi) := [(\partial_\theta \tilde{z}_0)(\theta_0(\psi))]^T \partial_x^{-1}. \tag{7.9}$$

By Taylor formula, we develop (7.8) in w at $(\eta, w) = (0,0)$, and we get

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

where

$$T_\delta(\psi) := A_\varepsilon(G_\delta(\psi,0,0)) = \varepsilon v_\delta(\psi) + \varepsilon^b z_0(\psi), \quad v_\delta(\psi) := v_\varepsilon(\theta_0(\psi), y_\delta(\psi)) \tag{7.10}$$

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

$$T_1(\psi)w := \varepsilon^{2b-1}U_1(\psi)w + \varepsilon^b w; \quad T_2(\psi)[w,w] := \varepsilon^{4b-3}U_2(\psi)[w,w] \tag{7.11}$$

$$U_1(\psi)w := \varepsilon \sum_{j\in S} \frac{|j|\,[L_2(\psi)w]_j\, e^{\mathrm{i}[\theta_0(\psi)]_j}}{2\sqrt{|j|}\sqrt{\xi_j + \varepsilon^{2(b-1)}[y_\delta(\psi)]_j}}, \tag{7.12}$$

$$U_2(\psi)[w,w] := -\varepsilon \sum_{j\in S} \frac{j^2\,[L_2(\psi)w]_j^2\, e^{\mathrm{i}[\theta_0(\psi)]_j}}{8|j|^{\frac{3}{2}}\{\xi_j + \varepsilon^{2(b-1)}[y_\delta(\psi)]_j\}^{\frac{3}{2}}}, \tag{7.13}$$

and $T_{\ge 3}(\psi, w)$ collects all the terms of order at least cubic in w. In the notation of (4.15), the function $v_\delta(\Psi)$ in (7.10) is $v_\delta(\psi) = v_\varepsilon(\theta_0(\psi), y_\delta(\psi))$. The terms U_1, U_2 in (7.12), (7.13) are $O(1)$ in ε. Moreover, using that $L_2(\psi)$ in (7.9) vanishes at $z_0 = 0$, they satisfy

$$\|U_1 w\|_s \le_s \|\mathfrak{I}_\delta\|_s \|w\|_{s_0} + \|\mathfrak{I}_\delta\|_{s_0}\|w\|_s, \quad \|U_2[w,w]\|_s \le_s \|\mathfrak{I}_\delta\|_s \|\mathfrak{I}_\delta\|_{s_0}\|w\|_{s_0}^2 + \|\mathfrak{I}_\delta\|_{s_0}^2 \|w\|_{s_0}\|w\|_s \tag{7.14}$$

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

$$\mathcal{H}(u+h) = \mathcal{H}(u) + ((\nabla\mathcal{H})(u), h)_{L^2(\mathbb{T})} + \frac{1}{2}((\partial_u\nabla\mathcal{H})(u)[h], h)_{L^2(\mathbb{T})} + O(h^3). \tag{7.15}$$

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 components of $K = \varepsilon^{-2b}(\mathcal{H} \circ A_\varepsilon \circ G_\delta)(\psi, 0, w)$ that are quadratic in w is

$$\frac{1}{2}(K_{02}w, w)_{L^2(\mathbb{T})} = \varepsilon^{-2b}((\nabla\mathcal{H})(T_\delta), T_2[w,w])_{L^2(\mathbb{T})} + \frac{\varepsilon^{-2b}}{2}((\partial_u\nabla\mathcal{H})(T_\delta)[T_1 w], T_1 w)_{L^2(\mathbb{T})}. \tag{7.16}$$

Inserting the expressions (7.12), (7.13) in the equality (7.16), we get

$$\begin{aligned} K_{02}(\psi)w =& (\partial_u\nabla\mathcal{H})(T_\delta)[w] + 2\varepsilon^{b-1}(\partial_u\nabla\mathcal{H})(T_\delta)[U_1 w] + \\ &+ \varepsilon^{2(b-1)}U_1^T(\partial_u\nabla\mathcal{H})(T_\delta)[U_1 w] + 2\,\varepsilon^{2b-3}U_2[w,\cdot]^T(\nabla\mathcal{H})(T_\delta). \end{aligned} \tag{7.17}$$

Lemma 7.3. *The operator K_{02} reads*

$$(K_{02}w, w)_{L^2(\mathbb{T})} = ((\partial_u\nabla\mathcal{H})(T_\delta)[w], w)_{L^2(\mathbb{T})} + (R(\psi)w, w)_{L^2(\mathbb{T})} \tag{7.18}$$

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

$$R(\psi)w = \sum_{|j|\leq C}(w, g_j(\psi))_{L^2(\mathbb{T})}\,\chi_j(\psi). \tag{7.19}$$

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

$$\|g_j\|_s^{Lip(\gamma)}\|\chi_j\|_{s_0}^{Lip(\gamma)} + \|g_j\|_{s_0}^{Lip(\gamma)} \leq_s \varepsilon^{1+b}\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \tag{7.20}$$

$$\begin{aligned} &\|\partial_i g_j[\hat\imath]\|_s\|\chi_j\|_{s_0} + \|\partial_i g_j[\hat\imath]\|_{s_0}\|\chi_j\|_s + \|g_j\|_s\|\partial_i\chi_j[\hat\imath]\|_{s_0} + \|g_j\|_{s_0}\|\partial_i\chi_j[\hat\imath]\|_s \\ &\leq_s \varepsilon^{1+b}\|\hat\imath\|_{s+\sigma} + \varepsilon^{2b-1}\|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s+\sigma} \end{aligned} \tag{7.21}$$

In conclusion, the linearized operator to analyze after the composition with the action-angle variables, the rescaling and the transformation G_δ is

$$w \mapsto (\partial_u\nabla\mathcal{H})(T_\delta)[w], \quad w \in H_S^\perp$$

up to finite dimensional operators which have form (7.19) and size (7.20).

7.2 The linearized operator in the normal directions

In this section we compute $((\partial_u\nabla\mathcal{H})(T_\delta)[w], w)_{L^2(\mathbb{T})}, w \in H_S^\perp$, recalling that $\mathcal{H} = H \circ \Phi_B$ and Φ_B is the Birkhoff map of Proposition 3.1. It is convenient to write separately the terms in

$$\mathcal{H} = H \circ \Phi_B = (H_2 + H_3) \circ \Phi_B + H_4 \circ \Phi_B + H_{\geq 5} \circ \Phi_B, \tag{7.22}$$

where $H_2, H_3, H_4, H_{\geq 5}$ are defined in (3.1). First we consider $H_{\geq 5} \circ \Phi_B$. By (3.1) we get

$$\nabla H_{\geq 5}(u) = \pi_0[(\partial_u f)(x, u, u_x)] - \partial_x\{(\partial_{u_x} f)(x, u, u_x)\}.$$

Since the Birkhoff transformation Φ_B has the form (3.6), Lemma 7.1 (at $u = T_\delta$) implies that

$$\begin{aligned} \partial_u\nabla(H_{\geq 5} \circ \Phi_B)(T_\delta)[h] &= (\partial_u\nabla H_{\geq 5})(\Phi_B(T_\delta))[h] + \mathcal{R}_{H_{\geq 5}}(T_\delta)[h] = \\ &= \partial_x(r_1(T_\delta)\,\partial_x h) + r_0(T_\delta)h + \mathcal{R}_{H_{\geq 5}}(T_\delta)[h] \end{aligned} \tag{7.23}$$

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

$$r_0(T_\delta) := \sigma_0(\Phi_B(T_\delta)), \quad \sigma_0(u) := (\partial_{uu} f)(x, u, u_x) - \partial_x\{(\partial_{uu_x} f)(x, u, u_x)\}, \tag{7.24}$$

$$r_1(T_\delta) := \sigma_1(\Phi_B(T_\delta)), \quad \sigma_1(u) := -(\partial_{u_x u_x} f)(x, u, u_x), \tag{7.25}$$

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

$$\|g_j\|_s^{Lip(\gamma)} \|\chi_j\|_{s_0}^{Lip(\gamma)} + \|g_j\|_{s_0}^{Lip(\gamma)} \leq_s \varepsilon^4 (1 + \|\mathfrak{I}_\delta\|_{s+2}^{Lip(\gamma)}),$$

$$\|\partial_i g_j[\hat\imath]\|_s \|\chi_j\|_{s_0} + \|\partial_i g_j[\hat\imath]\|_{s_0} \|\chi_j\|_s + \|g_j\|_s \|\partial_i \chi_j[\hat\imath]\|_{s_0} + \|g_j\|_{s_0} \|\partial_i \chi_j[\hat\imath]\|_s \leq_s \varepsilon^4 (\|\hat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+2} \|\hat\imath\|_{s_0+2}).$$

Now consider the contribution of $(H_2 + H_3 + H_4) \circ \Phi_B$. By Lemma 7.1 and (3.1) we have

$$\begin{aligned}
&\partial_u \nabla((H_2 + H_3 + H_4) \circ \Phi_B)(T_\delta)[h] = -h_{xx} - 6\, c_1\, \partial_x[\Phi_B(T_\delta)_x\, h_x] - 2\, c_2\, \partial_{xx}(\Phi_B(T_\delta) h) \\
&\quad + 2\, c_2\, \Phi_B(T_\delta)_x\, h_x + 6\, c_3\, \Phi_B(T_\delta)\, h - 12\, c_4\, \partial_x[(\Phi_B(T_\delta))_x^2\, h_x] - 3\, c_5\, \partial_x[(\Phi_B(T_\delta))_x^2\, h] \\
&\quad + 3\, c_5\, (\Phi_B(T_\delta))_x^2\, h_x - 2\, c_6\, \partial_x[\Phi_B(T_\delta)^2\, h_x] - 2\, c_6\, \partial_{xx}(\Phi_B(T_\delta)^2)\, h + 2\, c_6\, \Phi_B(T_\delta)_x^2\, h \\
&\quad + 12\, c_7\, \Phi_B(T_\delta)^2\, h + \mathcal{R}_{H_2}(T_\delta) + \mathcal{R}_{H_3}(T_\delta) + \mathcal{R}_{H_4}(T_\delta)[h],
\end{aligned} \tag{7.26}$$

where $\Phi_B(T_\delta)$ is a zero space average function, indeed Φ_B maps $H_0^1(\mathbb{T}_x)$ in itself by Proposition (3.1). The remainder $\mathcal{R}_{H_2}, \mathcal{R}_{H_3}, \mathcal{R}_{H_4}$ have the form (7.3) and, by (7.4), the size $(\mathcal{R}_{H_2} + \mathcal{R}_{H_3} + \mathcal{R}_{H_4})(T_\delta) = O(\varepsilon)$. We develop this sum as

$$(\mathcal{R}_{H_2} + \mathcal{R}_{H_3} + \mathcal{R}_{H_4})(T_\delta) = \varepsilon \mathcal{R}_1 + \varepsilon^2 \mathcal{R}_2 + \tilde{\mathcal{R}}_{>2}, \tag{7.27}$$

where $\tilde{\mathcal{R}}_{>2}$ has size $o(\varepsilon^2)$. Thus we get, for all $h \in H_S^\perp$,

$$\begin{aligned}
&\Pi_S^\perp \partial_u \nabla((H_2 + H_3 + H_4) \circ \Phi_B)(T_\delta)[h] = -h_{xx} + \Pi_S^\perp \{-6\, c_1\, \partial_x[\Phi_B(T_\delta)_x\, h_x] - 2\, c_2\, \partial_{xx}(\Phi_B(T_\delta) h) \\
&+ 2\, c_2\, \Phi_B(T_\delta)_x\, h_x + 6\, c_3\, \Phi_B(T_\delta)\, h - 12\, c_4\, \partial_x[(\Phi_B(T_\delta))_x^2\, h_x] - 3\, c_5\, \partial_x[(\Phi_B(T_\delta))_x^2\, h] + 3\, c_5\, (\Phi_B(T_\delta))_x^2\, h_x \\
&- 2\, c_6\, \partial_x[\Phi_B(T_\delta)^2\, h_x] - 2\, c_6\, \partial_{xx}(\Phi_B(T_\delta)^2)\, h + 2\, c_6\, \Phi_B(T_\delta)_x^2\, h + 12\, c_7\, \Phi_B(T_\delta)^2\, h\} \\
&+ \Pi_s^\perp(\varepsilon \mathcal{R}_1 + \varepsilon^2 \mathcal{R}_2 + \tilde{\mathcal{R}}_{>2})[h].
\end{aligned} \tag{7.28}$$

Now we expand $\Phi_B(u) = u + \Psi_2(u) + \Psi_{\geq 3}(u)$, where $\Psi_2(u)$ is a quadratic function of u, $\Psi_{\geq 3} = O(u^3)$ and both map $H_0^1(\mathbb{T}_x)$ in itself. At $u = T_\delta = \varepsilon v_\delta + \varepsilon^b z_0$ we get

$$\Phi_B(T_\delta) = T_\delta + \Psi_2(T_\delta) + \Psi_{\geq 3}(T_\delta) = \varepsilon v_\delta + \varepsilon^2 \Psi_2(v_\delta) + \tilde q, \tag{7.29}$$

where $\tilde q = \varepsilon^b z_0 + \Psi_2(T_\delta) - \Psi_2(v_\delta) + \Psi_{\geq 3}(T_\delta)$ and it satisfies

$$\|\tilde q\|_s^{Lip(\gamma)} \leq_s \varepsilon^3 + \varepsilon^b \|\mathfrak{I}_\delta\|_s^{Lip(\gamma)}, \quad \|\partial_i \tilde q[\hat\imath]\|_s \leq_s \varepsilon^b (\|\hat\imath\|_s + \|\mathfrak{I}_\delta\|_s \|\hat\imath\|_{s_0}). \tag{7.30}$$

Note that also $\tilde q$ has zero space average, indeed $\tilde q = \Phi_B(T_\delta) - \varepsilon v_\delta - \varepsilon^2 \Psi_2(v_\delta)$ and $\Phi_B(T_\delta), v_\delta, \Psi_2(v_\delta)$ belong to $H_0^1(\mathbb{T}_x)$.
We observe that the terms $O(\varepsilon)$ come from the monomials $R(v\, z^2)$ of $\mathcal{H}_3$ and the ones of size $O(\varepsilon^2)$ from $H_2 + \mathcal{H}_{4,2}$ (see (3.8)). Thus, we compare (7.28) with $\Pi_S^\perp(\partial_u \nabla(H_2 + \mathcal{H}_3 + \mathcal{H}_{4,2}))(T_\delta)[h]$, using (3.8), and, by (7.29), we obtain $\mathcal{R}_1 = 0$,

$$\Psi_2(v_\delta) = -c_1\, \partial_x(v_\delta^2) - \frac{c_2}{3} \partial_{xx}[(\partial_x^{-1} v_\delta)^2] + \frac{c_2}{3} \pi_0[v_\delta^2] + c_3 \pi_0[(\partial_x^{-1} v_\delta)^2] \tag{7.31}$$

and

$$
\begin{aligned}
\mathcal{R}_2[h] = & - 6c_1^2\{v_\delta\partial_{xx}(\Pi_S[(v_\delta)_x h_x]) - \partial_x((v_\delta)_x\partial_{xx}\Pi_S[v_\delta h])\}\\
& + 2c_1c_2\, v_\delta\, \partial_x(\Pi_S[(v_\delta)_x\, h_x]) + 2c_1c_2\, \partial_x((v_\delta)_x\, \partial_x\Pi_S[v_\delta\, h])\\
& - 2c_1c_2\, (\partial_x^{-1}v_\delta)\, \partial_{xx}\Pi_S[(v_\delta)_x\, h_x] + 2c_1c_2\, \partial_x\{(v_\delta)_x\, \partial_{xx}\Pi_S[(\partial_x^{-1}v_\delta)\, h]\}\\
& - \frac{2\,c_2^2}{3}\,(\partial_x^{-1}v_\delta)\,\partial_{xxx}\Pi_S[v_\delta\, h] + \frac{2\,c_2^2}{3}\,(\partial_x^{-1}v_\delta)\,\partial_x\Pi_S[(v_\delta)_x\, h_x]\\
& + \frac{2\,c_2^2}{3}\,\partial_x\{(v_\delta)_x\,\partial_x\Pi_S[(\partial_x^{-1}v_\delta)\, h] + 2c_2c_3\,(\partial_x^{-1}v_\delta)\,\partial_x\Pi_S[v_\delta\, h]\\
& - 2c_2c_3\, v_\delta\,\partial_x\Pi_S[(\partial_x^{-1}v_\delta)h] + 2c_1c_2\,\partial_x^{-1}\{v_\delta\,\partial_{xx}\Pi_S[(v_\delta)_x\, h_x]\}\\
& + 2c_1c_2\,\partial_x\{(v_\delta)_x\,\partial_{xx}\Pi_S[v_\delta\,(\partial_x^{-1}h)]\} + \frac{2\,c_2^2}{3}\,\partial_x^{-1}\{v_\delta\,\partial_{xxx}\Pi_S[v_\delta h]\}\\
& + \frac{2\,c_2^2}{3}\, v_\delta\,\partial_{xxx}\Pi_S[v_\delta\,(\partial_x^{-1}h)] - \frac{2\,c_2^2}{3}\,(\partial_x^{-1}\{v_\delta)\,\partial_x\Pi_S[(v_\delta)_x h_x]\}\\
& + \frac{2\,c_2^2}{3}\,\partial_x\{(v_\delta)_x\,\partial_x\Pi_S[v_\delta(\partial_x^{-1}h)] - 2c_2c_3\,\partial_x^{-1}\{v_\delta\,\partial_x\Pi_S[v_\delta\, h]\}\\
& - 2c_2c_3 v_\delta\,\partial_x\Pi_S[v_\delta\,(\partial_x^{-1}h)] - 2c_1c_2\, v_\delta\,\partial_x\Pi_S[(v_\delta)_x\, h_x]\\
& - 2c_1c_2\,\partial_x\{(v_\delta)_x\,\partial_x\Pi_S[v_\delta\, h]\} - \frac{4\,c_2^2}{3}\, v_\delta\,\partial_{xx}\Pi_S[v_\delta\, h]\\
& + \frac{2\,c_2^2}{3}\, v_\delta\,\Pi_S[(v_\delta)_x\, h_x] - \frac{2\,c_2^2}{3}\,\partial_x\{(v_\delta)_x\,\Pi_S[v_\delta\, h]\}\\
& + 4c_2c_3\, v_\delta\,\Pi_S[v_\delta\, h] + 6c_1c_3\,\partial_x^{-1}\{(\partial_x^{-1}v_\delta)\,\partial_x\Pi_S[(v_\delta)_x\, h_x]\}\\
& - 6c_1c_3\,\partial_x\{(v_\delta)_x\,\partial_x\Pi_S[(\partial_x^{-1}v_\delta)(\partial_x^{-1}h)]\} + 2c_2c_3\partial_x^{-1}\{(\partial_x^{-1}v_\delta)\,\partial_{xx}\Pi_S[v_\delta h]\}\\
& - 2c_2c_3\, v_\delta\,\partial_{xx}\Pi_S[(\partial_x^{-1}v_\delta)(\partial_x^{-1}h)] - 2c_2c_3\,\partial_x^{-1}\{(\partial_x^{-1}v_\delta)\,\Pi_S[(v_\delta)_x h_x]\}\\
& - 2c_2c_3\,\partial_x\{(v_\delta)_x\,\Pi_S[(\partial_x^{-1}v_\delta)(\partial_x^{-1}h)]\} - 6\,c_3^2\,\partial_x^{-1}\{(\partial_x^{-1}v_\delta)\,\Pi_S[v_\delta\, h]\}\\
& + 6c_3^2\, v_\delta\,\Pi_S[(\partial_x^{-1}v_\delta)(\partial_x^{-1}h)] + \frac{2}{3}c_2^2\, v_\delta\,\partial_{xxx}\Pi_S[(\partial_x^{-1}v_\delta)\, h].
\end{aligned}
\tag{7.32}
$$

In conclusion, we have the following proposition.

Proposition 7.4. *Assume* (7.7). *Then the Hamiltonian operator* $\mathcal{L}_\omega$, *for all* $h \in H^s_{S^\perp}(\mathbb{T}^{\nu+1})$, *has the form*

$$
\mathcal{L}_\omega h := \omega\cdot\partial_\varphi h - \partial_x K_{02} h = \Pi_S^\perp(\omega\cdot\partial_\varphi h + \partial_{xx}(a_1\, h_x) + \partial_x(a_0 h) - \varepsilon^2\partial_x\mathcal{R}_2 h - \partial_x\mathcal{R}_* h) \tag{7.33}
$$

where $\mathcal{R}_2$ *is defined in* (7.32),

$$
R_* := \tilde{\mathcal{R}}_{>2} + R_{H_{\geq 5}}(T_\delta) + R(\psi), \tag{7.34}
$$

with $R(\psi)$ *defined in Lemma 7.3, the functions*

$$
\begin{aligned}
a_1 :=& 1 + 6c_1\,(\Phi_B(T_\delta))_x + 2\,c_2\,\Phi_B(T_\delta) + 12c_4\,(\Phi_B(T_\delta))_x^2 + 3c_5\,\partial_x[\Phi_B(T_\delta)^2] + \\
& + 2c_6\,\Phi_B(T_\delta)^2 - r_1(T_\delta),
\end{aligned}
\tag{7.35}
$$

$$
\begin{aligned}
a_0 :=& 2c_2\,(\Phi_B(T_\delta))_{xx} - 6c_3\,\Phi_B(T_\delta) + 3c_5\,\partial_x[(\Phi_B(T_\delta))_x^2] + 2c_6\,\{\Phi_B(T_\delta)_x^2 + \\
& + 2\Phi_B(T_\delta)\,(\Phi_B(T_\delta))_{xx}\} - 12c_7\,\Phi_B(T_\delta)^2 - r_0(T_\delta)
\end{aligned}
\tag{7.36}
$$

the function r_1 *is defined in* (7.25), r_0 *in* (7.24), T_δ *and* v_δ *in* (7.10).
Furthermore, we have, for some $\sigma := \sigma(\nu,\tau) > 0$,

$$
\|a_1 - 1\|_s^{Lip(\gamma)} \leq_s \varepsilon\,(1 + \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}), \qquad \|\partial_i a_1[\hat{\imath}]\|_s \leq_s \varepsilon(\|i\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|i\|_{s_0+\sigma}), \tag{7.37}
$$

$$
\|a_0\|_s^{Lip(\gamma)} \leq_s \varepsilon\,(1 + \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}), \qquad \|\partial_i a_0[\hat{\imath}]\|_s \leq_s \varepsilon(\|i\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|i\|_{s_0+\sigma}), \tag{7.38}
$$

where $\mathfrak{I}_\delta(\varphi) := (\theta_0(\varphi) - \varphi, y_\delta(\varphi), z_0(\varphi))$ *corresponds to* T_δ*. The remainder* $\mathcal{R}_2$ *has the form* (7.3) *with*

$$\|g_j\|_s^{Lip(\gamma)} + \|\chi_j\|_s^{Lip(\gamma)} \le_s 1 + \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad \|\partial_i g_j[\hat\imath]\|_s + \|\partial_i \chi_j[\hat\imath]\|_s \le_s \|\hat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s_0+\sigma} \tag{7.39}$$

and also $\mathcal{R}_*$ *has the form* (7.3) *with*

$$\|g_j^*\|_s^{Lip(\gamma)}\|\chi_j^*\|_{s_0}^{Lip(\gamma)} + \|g_j^*\|_{s_0}^{Lip(\gamma)}\|\chi_j^*\|_s^{Lip(\gamma)} \le_s \varepsilon^3 + \varepsilon^{1+b}\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \tag{7.40}$$

$$\begin{aligned}&\|\partial_i g_j^*[\hat\imath]\|_s\|\chi_j^*\|_{s_0} + \|\partial_i g_j^*[\hat\imath]\|_{s_0}\|\chi_j^*\|_s + \|g_j^*\|_{s_0}\|\partial_i\chi_j^*\|_s + \|g_j^*\|_s\|\partial_i\chi_j^*\|_{s_0} \\ &\le_s \varepsilon^{1+b}\|\hat\imath\|_{s+\sigma} + \varepsilon^{2b-1}\|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s_0+\sigma}.\end{aligned} \tag{7.41}$$

The bounds (7.39) and (7.40) imply, by Lemma 7.2, estimates for the s-decay norms of $\mathcal{R}_2$ and $\mathcal{R}_*$. The linearized operator $\mathcal{L}_\omega := \mathcal{L}_\omega(\omega, i_\delta(\omega))$ depends on the parameter ω both directly and also through the dependence on the embedded torus $i_\delta(\omega)$. The estimates on the partial derivative respect to i (see (5.1)) allow us to control, along the Nash-Moser iteration, the Lipschitz variation of the eigenvalues of $\mathcal{L}_\omega$ with respect to ω and the approximate solution i_δ.

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 (7.33) to a constant coefficients linear operator $\mathcal{L}_\infty$. For this purpose, we shall apply the same kind of symplectic transformations used in [3], whose aim is to diagonalize the operator $\mathcal{L}_\omega$ up to a bounded remainder $\mathcal{R}_6$ (see (8.129)). This one has to satisfy the smallness condition (8.133) in order to initialize the KAM reducibility scheme of Theorem 8.19, that completes the diagonalization procedure.
The size of all these transformations will be greater than the ones used in [3] (see Section 8 in [3]) and, as a consequence, some non perturbative terms will be modified by them. Thus, in order to prove (8.133) we will have to overcome two main difficulties: (a) computing the terms of order ε and ε^2 after each transformation, since we need to normalize them through the Birkhoff steps of Section 8.5 and 8.6, (b) providing optimal estimates for the transformations and, consequently, for the remainder $\mathcal{R}_6$ (see (8.129)).

Consider

$$\overline{v}(\varphi, x) := \sum_{j\in S}\sqrt{|j|\xi_j}\, e^{\mathrm{i}\mathtt{l}(j)\cdot\varphi}\, e^{\mathrm{i}jx} \tag{8.1}$$

and $\mathtt{l}\colon S \to \mathbb{Z}^\nu$ is the odd injective map

$$\mathtt{l}\colon S \to \mathbb{Z}^\nu, \quad \mathtt{l}(\bar\jmath_i) := \mathtt{e}_i, \quad \mathtt{l}(-\bar\jmath_i) = -\mathtt{l}(\bar\jmath_i) = -\mathtt{e}_i, \quad i = 1, \dots, \nu, \tag{8.2}$$

denoting by $\mathtt{e}_i = (0, \dots, 1, \dots, 0)$ the i-th vector of the canonical basis of $\mathbb{R}^\nu$. We observe that

$$\|v_\delta - \overline{v}\|_s^{Lip(\gamma)} \le_s \|\mathfrak{I}_\delta\|_s^{Lip(\gamma)}, \qquad \|\partial_i(v_\delta - \overline{v})[\hat\imath]\|_s \le_s \|\hat\imath\|_s + \|\mathfrak{I}_\delta\|_s\|\hat\imath\|_{s_0}. \tag{8.3}$$

Remark 8.1. The function $\overline{v}(\varphi, x)$ in (8.1) corresponds to the torus $(\varphi, 0, 0)$ after the transformation A_ε defined in (4.15). In particular, this torus is invariant under the flow of the integrable Hamiltonian $\varepsilon^{-2b}\tilde h \circ A_\varepsilon$ (recalling (4.3)), which preserves the momentum. Hence, the square of the L^2 norm of $\overline{v}$ is independent of the time φ, as we can deduce by the properties of the map $\mathtt{l}$ defined in (8.2).
We shall expand the coefficients of the linearized operator at $y = z = 0$ to get the bounds on the transformations defined along this section, thus we will frequently use the inequalities (8.3) and the assumption (7.7). Moreover, we will use the fact that $\overline{v}$ satisfies the equation $L_{\overline{\omega}} = 0$, where $\overline{\omega}$ is the vector of the linear frequencies (see (1.13)) and $L_\omega := \omega \cdot \partial_\varphi + \partial_{xxx}$.

Remark 8.2. We recall that $\omega = \overline{\omega} + O(\varepsilon^2)$, see for instance (4.18). Moreover, note that $\mathcal{D}_\omega \overline{v} = \mathcal{D}_{\overline{\omega}} \overline{v} + \mathcal{D}_{\omega - \overline{\omega}} \overline{v}$ and

$$\mathcal{D}_{\omega-\overline{\omega}}\overline{v} = \sum_{j \in S} \mathrm{i}(\omega - \overline{\omega}) \cdot \mathtt{l}(j) \sqrt{|j|\xi_j}\, e^{\mathrm{i}\mathtt{l}(j)\cdot\varphi}\, e^{\mathrm{i}jx}.$$

Then $\|\mathcal{D}_{\omega-\overline{\omega}}\overline{v}\|_s^{Lip(\gamma)} \leq C\varepsilon^2$ and $\mathcal{D}_{\omega-\overline{\omega}}\overline{v}$ has zero spatial average.

We expand in powers of ε the coefficients a_0 and a_1 in (7.36) and (7.35) as

$$a_0 = \varepsilon a_{0,1} + \varepsilon^2 a_{0,2} + \mathtt{R}_{a_0}, \qquad a_1 - 1 = \varepsilon a_{1,1} + \varepsilon^2 a_{1,2} + \mathtt{R}_{a_1}, \tag{8.4}$$

where

$$\begin{aligned}
& a_{0,1} := 2c_2\, \overline{v}_{xx} - 6c_3\, \overline{v}, \qquad a_{1,1} := 6c_1 \overline{v}_x + 2c_2 \overline{v}_{xx}, \\
& a_{0,2} := 2c_2\, (\Psi_2(\overline{v}))_{xx} - 6c_3\, \Psi_2(\overline{v}) + 3c_5 \partial_x(\overline{v}_x^2) + 2c_6\{\overline{v}_x^2 + 2\overline{v}\overline{v}_{xx}\} - 12c_7\overline{v}^2, \\
& a_{1,2} := 6c_1(\Psi_2(\overline{v}))_x + 2c_2\Psi_2(\overline{v}) + 12c_4\overline{v}_x^2 + 3c_5\partial_x(\overline{v}^2) + 2c_6\overline{v}^2
\end{aligned}$$

and, by (8.3), $\|\mathtt{R}_{a_k}\|_s^{Lip(\gamma)} \leq \varepsilon^3 + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}$, for some $\sigma > 0$.

8.1 Space reduction at the order ∂_{xxx}

First we conjugate $\mathcal{L}_\omega$ in (7.33) to an operator $\mathcal{L}_1$ whose coefficient in front of ∂_{xxx} is independent on the space variable x. Because of the Hamiltonian structure, the terms $O(\partial_{xx})$ will be simultaneously eliminated.
We look for a φ-dependent family of symplectic diffeomorphisms $\Phi(\varphi)$ of $H_S^\perp$ which differ from

$$\mathcal{A}_\perp := \Pi_S^\perp \mathcal{A} \Pi_S^\perp, \quad (\mathcal{A}h)(\varphi, x) := (1 + \beta_x(\varphi, x))\, h(\varphi, x + \beta(\varphi, x)), \tag{8.5}$$

up to a small "finite dimensional" remainder, see (8.9).
If $\|\beta\|_{W^{1,\infty}} < \frac{1}{2}$ then $\mathcal{A}$ is invertible and its inverse and adjoin map are

$$(\mathcal{A}^{-1}h)(\varphi, y) := (1 + \tilde{\beta}_y(\varphi, y))\, h(\varphi, y + \tilde{\beta}(\varphi, y)), \quad (\mathcal{A}^T h)(\varphi, y) = h(\varphi, y + \tilde{\beta}(\varphi, y)) \tag{8.6}$$

For each $\varphi \in \mathbb{T}^\nu$, $\mathcal{A}(\varphi)$ is a symplectic transformation of the phase space, see Remark 3.3 in [2], but the restricted map $\mathcal{A}_\perp(\varphi)$ is not.
In order to find a symplectic diffeomorphism near $\mathcal{A}_\perp$ first we observe that $\mathcal{A}_\perp$ is the time-1 flow map of the linear Hamiltonian PDE

$$\partial_\tau u = \partial_x(b(\varphi, \tau, x)u), \quad b(\varphi, \tau, x) := \frac{\beta(\varphi, x)}{1 + \tau\beta_x(\varphi, x)}. \tag{8.7}$$

The equation (8.7) is a linear transport equation, whose characteristic curves are the solutions of the ODE

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

As in [3], we define a symplectic map Φ of $H_S^\perp$ as the time-1 flow of the Hamiltonian PDE

$$\partial_\tau u = \Pi_S^\perp \partial_x(b(\tau, x)u) = \partial_x(b(\tau, x)u) - \Pi_S \partial_x(b(\tau, x)u), \quad u \in H_S^\perp \tag{8.8}$$

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 (8.8) is well defined in the Sobolev spaces $H^s_{S^\perp}(\mathbb{T}_x)$ for $b(\tau, x)$ smooth enough, by standard theory of linear hyperbolic PDE's. We obtained a symplectic diffeomorphism Φ that differs from $\mathcal{A}_\perp$ by a "finite dimensional" remainder of small size, more precisely, of size $O(\beta)$.

Lemma 8.3. *(Lemma 8.2 in [3]) For $\|\beta\|_{W^{s_0+1,\infty}}$ small, there exists an invertible symplectic transformation $\Phi = \mathcal{A}_\perp + \mathcal{R}_\Phi$ of $H^s_{S^\perp}$, where $\mathcal{A}_\perp$ is defined in* (8.5) *and $\mathcal{R}_\Phi$ is a "finite dimensional" remainder*

$$\mathcal{R}_\Phi h = \sum_{j\in S}\int_0^1 (h, g_j(\tau))_{L^2(\mathbb{T})}\chi_j(\tau)\, d\tau + \sum_{j\in S}(h,\psi_j)_{L^2(\mathbb{T})}e^{\mathrm{i}jx} \tag{8.9}$$

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

$$\|\psi_j\|_s + \|g_j(\tau)\|_s \leq_s \|\beta\|_{W^{s+2,\infty}}, \quad \|\chi_j(\tau)\|_s \leq_s 1 + \|\beta\|_{W^{s+1,\infty}}. \tag{8.10}$$

Moreover

$$\|\Phi h\|_s + \|\Phi^{-1}h\|_s \leq_s \|h\|_s + \|\beta\|_{W^{s+2,\infty}}\|h\|_{s_0} \quad \forall h \in H^s_{S^\perp}. \tag{8.11}$$

We conjugate $\mathcal{L}_\omega$ in (7.33) via the symplectic map $\Phi = \mathcal{A}_\perp + \mathcal{R}_\Phi$ of Lemma (8.3). Using the splitting $\Pi_S^\perp = \mathrm{I} - \Pi_S$, we compute

$$\mathcal{L}_\omega\Phi = \Phi\mathcal{D}_\omega + \Pi_S^\perp\mathcal{A}(b_3\partial_{yyy} + b_2\partial_{yy} + b_1\partial_y + b_0)\Pi_S^\perp + \mathcal{R}_I, \tag{8.12}$$

where the coefficients are

$$b_3(\varphi,y) := \mathcal{A}^T[a_1\,(1+\beta_x)^3] \qquad b_2(\varphi,y) := \mathcal{A}^T[2(a_1)_x(1+\beta_x)^2 + 6\,a_1\,\beta_{xx}(1+\beta_x)] \tag{8.13}$$

$$b_1(\varphi,y) := \mathcal{A}^T\left[(\mathcal{D}_\omega\beta) + 3\,a_1\,\frac{\beta_{xx}^2}{1+\beta_x} + 4\,a_1\,\beta_{xxx} + 6\,(a_1)_x\beta_{xx} + (a_1)_{xx}(1+\beta_x) + a_0(1+\beta_x)\right] \tag{8.14}$$

$$b_0(\varphi,y) := \mathcal{A}^T\left[\frac{(\mathcal{D}_\omega\beta_x)}{1+\beta_x} + a_1\,\frac{\beta_{xxxx}}{1+\beta_x} + 2(a_1)_x\frac{\beta_{xxx}}{1+\beta_x} + (a_1)_{xx}\,\frac{\beta_{xx}}{1+\beta_x} + a_0\,\frac{\beta_{xx}}{1+\beta_x} + (a_0)_x\right] \tag{8.15}$$

and the remainder

$$\begin{aligned}\mathcal{R}_I := &-\Pi_S^\perp\partial_x(\varepsilon^2\mathcal{R}_2 + \mathcal{R}_*)\,\mathcal{A}_\perp - \Pi_S^\perp(a_1\partial_{xxx} + 2(a_1)_x\partial_{xx} + ((a_1)_{xx} + a_0)\partial_x + (a_0)_x)\Pi_S\mathcal{A}\Pi_S^\perp + \\ &+ [\mathcal{D}_\omega, \mathcal{R}_\Phi] + (\mathcal{L}_\omega - \mathcal{D}_\omega)\mathcal{R}_\Phi.\end{aligned} \tag{8.16}$$

The commutator $[\mathcal{D}_\omega, \mathcal{R}_\Phi]$ has the form (8.9) with $\mathcal{D}_\omega g_j$ or $\mathcal{D}_\omega\chi_j, \mathcal{D}_\omega\psi_j$ instead of χ_j, g_j, ψ_j respectively. Also the last term $(\mathcal{L}_\omega - \mathcal{D}_\omega)\mathcal{R}_\Phi$ in (8.16) has the form (8.9) (note that $\mathcal{L}_\omega - \mathcal{D}_\omega$ does not contain derivatives with respect to φ). By (8.12), and decomposing $\mathrm{I} = \Pi_S + \Pi_S^\perp$, we get

$$\mathcal{L}_\omega\Phi = \Phi(\mathcal{D}_\omega + b_3\partial_{yyy} + b_2\partial_{yy} + b_1\partial_y + b_0)\Pi_S^\perp + \mathcal{R}_{II}, \tag{8.17}$$

$$\mathcal{R}_{II} := \{\Pi_S^\perp(\mathcal{A} - \mathrm{I})\Pi_S - \mathcal{R}_\Phi\}(b_3\partial_{yyy} + b_2\partial_{yy} + b_1\partial_y + b_0)\Pi_S^\perp + \mathcal{R}_I. \tag{8.18}$$

In order to solve the equation

$$b_3(\varphi,y) = b_3(\varphi)$$

for some function $b_3(\varphi)$, so that the coefficient in front of ∂_{xxx} depends only on φ, we choose the function $\beta = \beta(\varphi,x)$ such that

$$a_1(\varphi,x)(1+\beta_x(\varphi,x))^3 = b_3(\varphi), \tag{8.19}$$

where we used that $\mathcal{A}^T[b_3(\varphi)] = b_3(\varphi)$. The only solution of (8.19) with zero space average is

$$\beta := \partial_x^{-1}\rho_0, \quad \rho_0 := b_3(\varphi)^{\frac{1}{3}}(a_1(\varphi,x))^{-\frac{1}{3}} - 1, \quad b_3(\varphi) := \left(\frac{1}{2\pi}\int_{\mathbb{T}}(a_1(\varphi,x))^{-\frac{1}{3}}\,dx\right)^{-3}. \tag{8.20}$$

Applying the symplectic map Φ^{-1} in (8.17) we obtain the Hamiltonian operator

$$\mathcal{L}_1 := \Phi^{-1}\mathcal{L}_\omega\Phi = \Pi_S^\perp(\omega\cdot\partial_\varphi + b_3(\varphi)\partial_{yyy} + b_1\partial_y + b_0)\Pi_S^\perp + \mathfrak{R}_1 \tag{8.21}$$

where $\mathfrak{R}_1 := \Phi^{-1}\mathcal{R}_{II}$. We used that, by the Hamiltonian nature of $\mathcal{L}_1$, the coefficient $b_2 = 2\,(b_3)_y$ and so, by the choice (8.20), we have $b_2 = 2\,(b_3)_y = 0$.

Lemma 8.4. *(Lemma 8.3 in [3]) The operator $\mathfrak{R}_1$ in (8.21) has the form (7.5).*

In the proofs of the estimates for the transformations and the coefficients, we will always use the index σ to denote a certain loss of derivatives, since we do not need to know exactly the total amount of this loss. This, in fact, involves only the regularity required for the Hamiltonian nonlinearity $f(x, u, u_x)$ in (1.5).

Lemma 8.5. *There is $\sigma := \sigma(\tau, \nu) > 0$ such that, for $k = 0, 1$,*

$$\|\beta\|_s^{Lip(\gamma)} \leq_s \varepsilon\,(1 + \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}) \qquad \|\partial_i \beta[\hat{\imath}]\|_s \leq_s \varepsilon(\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}) \tag{8.22}$$

$$\|b_3 - 1\|_s^{Lip(\gamma)} \leq_s \varepsilon^2\,(1 + \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}) \qquad \|\partial_i b_3[\hat{\imath}]\| \leq_s \varepsilon^2(\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}) \tag{8.23}$$

$$\|b_k\|_s^{Lip(\gamma)} \leq_s \varepsilon(1 + \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}) \qquad \|\partial_i b_k[\hat{\imath}]\|_s \leq_s \varepsilon(\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}). \tag{8.24}$$

The transformations Φ, Φ^{-1} satisfy

$$\|\Phi^{\pm 1} h\|_s^{Lip(\gamma)} \leq_s \|h\|_{s+1}^{Lip(\gamma)} + \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}\|h\|_{s_0+1}^{Lip(\gamma)} \tag{8.25}$$

$$\|\partial_i(\Phi^{\pm 1} h)[\hat{\imath}]\|_s \leq_s \|h\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma} + \|h\|_{s_0+\sigma}\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|h\|_{s_0+\sigma}\|\hat{\imath}\|_{s_0+\sigma}. \tag{8.26}$$

Moreover the remainder $\mathcal{R}_*$ has the form (7.5) where the functions $\chi_j(\tau), g_j(\tau)$ satisfy the estimates (7.40) uniformly in $\tau \in [0, 1]$.

Proof. Estimate (8.23): Consider the functions $g(t) = (1 + t)^{-\frac{1}{3}}$ and $\Upsilon(t) = (1 + t)^{-3}$, analytic in a small neighbourhood of the origin. Then we have

$$b_3 - 1 = \Upsilon(M_x[g(a_1 - 1) - g(0)]) - \Upsilon(0). \tag{8.27}$$

By the mean value theorem, $\|b_3 - 1\|_s \leq_s \|M_x[g(a_1 - 1) - g(0)]\|_s$. By Taylor expansion, we get

$$M_x[g(a_1 - 1) - g(0)] = g'(0)M_x[a_1 - 1] + \int_{\mathbb{T}}\int_0^1 (1 - s)\, g''(s(a_1 - 1))\,(a_1 - 1)^2\, ds\, dx \tag{8.28}$$

and we note that, by Remark 8.4,

$$M_x[a_1 - 1] = \varepsilon^2 M_x[a_{1,2}] + M_x[\mathtt{R}_{a_1}].$$

Moreover, $\|M_x[\mathtt{R}_{a_1}]\|_s \leq_s \varepsilon^3 + \varepsilon^{2b}\|\mathfrak{I}_\delta\|_{s+\sigma}$, because $M_x[v_\delta - \overline{v}] = M_x[\tilde{q}] = 0$ and $\mathtt{R}_{a_1}$ contains terms like $\varepsilon^2(v_\delta^2 - \overline{v}^2)$ and cubic in the x-derivatives of v_δ.
The second addend in the right hand side of (8.28) can be estimated by $\varepsilon^2(1 + \|\mathfrak{I}_\delta\|_{s+\sigma})$. Hence

$$\|b_3 - 1\|_s \leq_s \varepsilon^2(1 + \|\mathfrak{I}_\delta\|_{s+\sigma}). \tag{8.29}$$

Now we consider the partial derivative respect to the variable i (see (5.1)) of b_3, namely

$$\partial_i b_3[\hat{\imath}] = \Upsilon'(M_x[g(a_1 - 1) - g(0)])\, M_x[g'(a_1 - 1)\, \partial_i a_1[\hat{\imath}]].$$

The derivatives of the functions g and Υ, for ε small enough, are approximately 1. Therefore, the estimate

$$\|\partial_i b_3[\hat{\imath}]\|_s \leq_s \varepsilon^2(\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}) \tag{8.30}$$

derived from the estimate on $M_x[\partial_i a_1[\hat{\imath}]]$ and the fact that $M_x[\partial_i \overline{v}[\hat{\imath}]] = 0$. By (8.29) and (8.30) we conclude.

Estimate (8.22): Consider the functions $\phi(t) := (1 + t)^{-1}$ and $g(t) := (1 + t)^{-\frac{1}{3}}$. Recalling that $\beta_x = (b_3^{-1} a_1)^{\frac{1}{3}} - 1$, we have

$$\beta_x = g^{-1}(b_3^{-1}\, a_1 - 1) - g^{-1}(0) \quad \text{and} \quad b_3^{-1}\, a_1 - 1 = a_1\,(\phi(b_3 - 1) - \phi(0)) + (a_1 - 1).$$

Then, by (2.10),

$$
\begin{aligned}
\|\beta_x\|_s &\leq_s \|\phi(b_3-1)-\phi(0)\|_s\|a_1\|_{s_0} + \|\phi(b_3-1)-\phi(0)\|_{s_0}\|a_1\|_s + \|a_1-1\|_s \\
&\leq_s \|b_3-1\|_{s+\sigma} + \|b_3-1\|_{s_0+\sigma}\|a_1\|_s + \|a_1-1\|_s \leq_s \varepsilon(1+\|\mathfrak{I}_\delta\|_{s+\sigma}).
\end{aligned}
$$

Estimate (8.24): By (7.38), (7.37), (8.22) we get the estimates (8.24).
For the estimates (8.25), (8.26) on Φ, Φ^{-1} we apply Lemma 8.3 and the estimate (8.22) for β. We estimate the remainder $\mathcal{R}_*$ using (8.16), (8.18) and (7.40). □

8.2 Terms of order ε and ε^2

The diffeomorphism of the torus $\Phi = \mathcal{A}_\perp + \mathcal{R}_\Phi$ defined in Lemma 8.3 is, by (8.10) and (8.22), of the form $\mathrm{I} + O(\varepsilon)$, hence, the terms $O(\varepsilon^2)$ of $\mathcal{L}_\omega$ are modified by it.
From now on, the transformations we shall apply to reduce the linearized operator $\mathcal{L}_\omega$ to a constant coefficient operator will be $\mathrm{I} + O(\varepsilon^d)$ with $d > 1$, hence the terms of order $\varepsilon, \varepsilon^2$ will not be changed anymore. In this section, our goal is to identify them in view of the linear Birkhoff steps of Section 8.5 and 8.6.

We have to put in evidence the terms $O(\varepsilon), O(\varepsilon^2)$ of b_0, b_1, b_3 in (8.21) and the ones in the remainder $\mathfrak{R}_1$ defined in (8.51).

Coefficients b_k

First, we note that $b_k = \mathcal{A}^T\alpha_k = \alpha_k + (\mathcal{A}^T - \mathrm{I})\alpha_k$, $k = 0, 1$, where

$$
\alpha_1 := (\mathcal{D}_\omega\beta) + 3\,a_1\,\frac{\beta_{xx}^2}{1+\beta_x} + 4\,a_1\,\beta_{xxx} + 6\,(a_1)_x\beta_{xx} + (a_1)_{xx}(1+\beta_x) + a_0(1+\beta_x), \tag{8.31}
$$

$$
\alpha_0 := \frac{(\mathcal{D}_\omega\beta_x)}{1+\beta_x} + a_1\,\frac{\beta_{xxxx}}{1+\beta_x} + 2(a_1)_x\frac{\beta_{xxx}}{1+\beta_x} + (a_1)_{xx}\,\frac{\beta_{xx}}{1+\beta_x} + a_0\,\frac{\beta_{xx}}{1+\beta_x} + (a_0)_x. \tag{8.32}
$$

By (7.35), (8.20), we have

$$
\begin{aligned}
\beta = &-2\,c_1\Phi_B(T_\delta) - \frac{2}{3}\,c_2\partial_x^{-1}[\Phi_B(T_\delta)] - 4\,c_4\partial_x^{-1}[\Phi_B(T_\delta)_x^2] - c_5\,\pi_0[\Phi_B(T_\delta)^2] - \frac{2}{3}\,c_6\partial_x^{-1}[\Phi_B(T_\delta)^2] \\
&+ 8\,c_1^2\partial_x^{-1}[\Phi_B(T_\delta)_x^2] + \frac{8}{9}\,c_2^2\partial_x^{-1}[\Phi_B(T_\delta)^2] + \frac{8}{3}\,c_1c_2\pi_0[\Phi_B(T_\delta)^2] + \mathtt{R}
\end{aligned} \tag{8.33}
$$

where, by (7.30), $\|\mathtt{R}\|_s^{Lip(\gamma)} \leq_s \varepsilon^3 + \varepsilon^b\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}$. Then we write $\beta = \varepsilon\,\beta_1 + \varepsilon^2\,\beta_2 + \mathtt{R}_\beta$, where

$$
\begin{aligned}
\beta_1 :&= -2c_1\overline{v} - \frac{2}{3}c_2\partial_x^{-1}(\overline{v}), \\
\beta_2 :&= -2c_1\Psi_2(\overline{v}) - \frac{2}{3}c_2\partial_x^{-1}(\Psi_2(\overline{v})) - 4c_4\partial_x^{-1}(\overline{v}_x^2) - c_5\pi_0[\overline{v}^2] \\
&\quad - \frac{2}{3}c_6\partial_x^{-1}[\overline{v}^2] + 8c_1^2\partial_x^{-1}[\overline{v}_x^2] + \frac{8}{9}c_2^2\partial_x^{-1}[\overline{v}^2] + \frac{8}{3}c_1c_2\pi_0[\overline{v}^2]
\end{aligned} \tag{8.34}
$$

and $\mathtt{R}_\beta$ is defined by difference and satisfies, by (8.3),

$$
\|\mathtt{R}_\beta\|_s^{Lip(\gamma)} \leq_s \varepsilon^3 + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad \|\partial_i\mathtt{R}_\beta[\hat{\imath}]\|_s \leq_s \varepsilon(\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}).
$$

Now we can develop α_0 and α_1 in powers of ε. By (7.35), (7.36), (8.33) and by Remark 8.1 we obtain $\alpha_1 := \varepsilon\alpha_{1,1} + \varepsilon^2\alpha_{1,2} + \mathtt{R}_1$ and $\alpha_0 = \varepsilon\alpha_{0,1} + \varepsilon^2\alpha_{0,2} + \mathtt{R}_0$, where

$$
\begin{aligned}
\alpha_{1,1} &= 2\,c_2\overline{v}_{xx} - 6\,c_3\overline{v}, \\
\alpha_{1,2} &= L_{\overline{\omega}}[\beta_2] + \frac{8}{3}(\beta_2)_{xxx} - \frac{41}{3}\,\partial_x[(\beta_1)_x\,(\beta_1)_{xx}] + a_{0,2} + a_{0,1}\,(\beta_1)_x,
\end{aligned} \tag{8.35}
$$

and

$$\alpha_{0,1} = 2\,c_2\overline{v}_{xxx} - 6\,c_3\overline{v}_x,$$
$$\alpha_{0,2} = \partial_x L_{\overline{\omega}}[\beta_2] - 3\partial_x[(\beta_1)_x\,(\beta_1)_{xxx}] - 3\partial_x[(\beta_1)^2_{xx}] + a_{0,1}(\beta_1)_{xx} + (a_{0,2})_x. \tag{8.36}$$

The functions $\mathtt{R}_0$ and $\mathtt{R}_1$ are defined by difference and satisfy the following estimates

$$\|\mathtt{R}_k\|_s^{Lip(\gamma)} \leq_s \varepsilon^3 + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad \|\partial_i \mathtt{R}_k[\hat{\imath}]\|_s \leq_s \varepsilon(\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}), \quad k = 0, 1. \tag{8.37}$$

Remark 8.6. We note that the terms $O(\varepsilon)$ generated by the Hamiltonian $\int_{\mathbb{T}}(3c_1v_x + c_2v)\,z_x^2\,dx$ (see (3.8)) are cancelled by the diffeomorphism of the torus Φ.

Remark 8.7. The averages of $\alpha_{j,k}, j = 0, 1$ for $k = 1$ are zero and, for $k = 2$, we have

$$M_x[\alpha_{1,2}] = M_x[a_{0,2}] + M_x[a_{0,1}\,(\beta_1)_x] = -2c_6\,M_x[\overline{v}_x^2] - 12c_7M_x[\overline{v}^2] + \frac{4}{3}c_2^2M_x[\overline{v}_x^2] + 4c_2c_3M_x[\overline{v}^2],$$

$$M_x[\alpha_{0,2}] = M_x[a_{0,1}\,(\beta_1)_{xx}] = -4c_1c_2M_x[\overline{v}_{xx}^2] - 12c_1c_3M_x[\overline{v}_x^2].$$

We used the fact that $\partial_\varphi M_x[\overline{v}^2] = 0$, see Remark 8.1. Moreover, we note that, for a similar argument, $M_{\varphi,x}[\alpha_{k,2}] = M_x[\alpha_{k,2}]$, for $k = 0, 1$.

The transformation $\mathcal{A}^T - \mathrm{I}$ (see Section 8.1) is of order $O(\varepsilon)$, hence it generates new terms of order $O(\varepsilon^2)$ when it is applied to ones of order ε. In particular, by the regularity of the function $\overline{v}(\varphi, x)$, that is at least C^2, we have, for $k = 0, 1$, by Taylor expansion

$$\varepsilon(\mathcal{A}^T - \mathrm{I})\alpha_{k,1}(\varphi, y) = \varepsilon(\alpha_{k,1}(\varphi, y + \tilde{\beta}(\varphi, y)) - \alpha_{k,1}(\varphi, y)) = \varepsilon\partial_y(\alpha_{k,1})(\varphi, y)\,\tilde{\beta}(\varphi, y) + \mathtt{R}_{\tilde{\beta}},$$

where $\|\mathtt{R}_{\tilde{\beta}}\|_s \leq_s \varepsilon^3(1 + \|\mathfrak{I}_\delta\|_{s+\sigma})$ for some $\sigma > 0$.
We observe that $\tilde{\beta}(\varphi, y) = -(\mathcal{A}^T\beta)(\varphi, y)$ and by (8.34) we get, for $k = 0, 1$,

$$\varepsilon(\mathcal{A}^T - \mathrm{I})\alpha_{k,1}(\varphi, y) = -\varepsilon^2\,\partial_y(\alpha_{k,1})(\varphi, y)\,\beta_1(\varphi, y) + \mathtt{R}_{\tilde{\beta}}, \tag{8.38}$$

where we have renamed $\mathtt{R}_{\tilde{\beta}}$ the terms of order $o(\varepsilon^2)$.

Remainder $\mathfrak{R}_1$

The remaining terms of order ε^2 generated by the diffeomorphism of the torus Φ have the form (7.5) and originate from $\mathcal{R}_{II} = \Phi\mathfrak{R}_1$ (see (8.18)). Thus we analyze the expression

$$\begin{aligned}\mathcal{R}_{II} :=\, & \Pi_S^\perp(\mathcal{A} - \mathrm{I})\Pi_S[b_3\partial_{yyy} + b_1\partial_y + b_0] - R_\Phi(b_3\partial_{yyy} + b_1\partial_y + b_0) \\ & - \Pi_S^\perp\partial_x(\varepsilon^2\mathcal{R}_2 + \mathcal{R}_*)\mathcal{A}_\perp - \Pi_S^\perp[\partial_{xx}(a_1\partial_x) + \partial_x(a_0\cdot)]\Pi_S\mathcal{A}\Pi_S^\perp + [\mathcal{D}_\omega, \mathcal{R}_\Phi] \\ & + (\mathcal{L}_\omega - \mathcal{D}_\omega)\mathcal{R}_\Phi.\end{aligned} \tag{8.39}$$

We start from the first term in (8.39). As we said above, the transformation $\mathcal{A} - \mathrm{I}$ has size $O(\varepsilon)$. Hence, we look for the terms $O(\varepsilon)$ of $b_3\partial_{yyy} + b_1\partial_y + b_0$. We have, by (8.23), $b_3 = 1 + O(\varepsilon^2)$ and $b_k = \alpha_k + (\mathcal{A}^T - \mathrm{I})\alpha_k$ for $k = 0, 1$. Thus

$$b_3\partial_{yyy} + b_1\partial_y + b_0 = \partial_{yyy} + \varepsilon\partial_y(\alpha_{1,1}\,\cdot) + O(\varepsilon^2).$$

By Taylor expansion at the point $\beta = 0$, we get, for a function $u(\varphi, x)$

$$\begin{aligned}(\mathcal{A} - \mathrm{I})u(\varphi, x) &= (1 + \beta_x)u(\varphi, x + \beta) - u(\varphi, x) = u(\varphi, x + \beta) - u(\varphi, x) + \beta_x u(\varphi, x + \beta) = \\ &= u_x(\varphi, x)\beta(\varphi, x) + \beta_x(\varphi, x)u(\varphi, x) + O(\beta^2) = \\ &= \varepsilon\partial_x(\beta_1(\varphi, x)\,u(\varphi, x)) + O(\varepsilon^2).\end{aligned} \tag{8.40}$$

Therefore we have

$$\Pi_S^\perp(\mathcal{A}-\mathrm{I})\Pi_S[b_3\partial_{yyy}+b_2\partial_{yy}+b_1\partial_y+b_0]=\varepsilon^2\Pi_S^\perp[\partial_x(\beta_1\,\partial_x(\alpha_{1,1}\,\cdot))]+o(\varepsilon^2) \tag{8.41}$$

Now we extract the homogeneous terms of order ε from $\mathcal{R}_\Phi$ (see (8.9)). We recall the exact expressions of g_k and χ_k in (8.9) refering to the proof of Lemma 8.2 in [3]. We have

$$g_k(\tau,x):=-(\Phi^\tau)^T[b(\tau)\partial_x e^{\mathrm{i}kx}], \tag{8.42}$$

where $(\Phi^\tau)^T$ is the flow of the adjoint PDE

$$\partial_\tau z=\Pi_S^\perp\{b(\tau,x)\partial_x z\},\qquad b(\tau,x)=\frac{\beta(x)}{1+\tau\beta_x(x)}=\varepsilon\beta_1+O(\varepsilon^2). \tag{8.43}$$

This equation is well defined on $H^s_{S^\perp}(\mathbb{T}_x)$, because the function b is smooth enough. By (8.42) we have

$$g_k(\tau,x)=-b(\tau)\partial_x e^{\mathrm{i}kx}+(\mathrm{I}_{H_S^\perp}-(\Phi^\tau)^T)[b(\tau)\partial_x e^{\mathrm{i}kx}]$$

and, for $z\in H^s_{S^\perp}(\mathbb{T}_x)$, by (8.22) and (8.43), $\|(\Phi^\tau)^T z-z\|_s\leq_s\varepsilon C(\|z\|_{s+1}+\|\mathfrak{I}_\delta\|_{s+\sigma}\|z\|_{s+1})$, where C is the Lipschitz constant, in time, on the interval $[0,1]$ of the flow $(\Phi^\tau)^T$. Hence, by (8.42),

$$g_k=-\varepsilon\beta_1\,\partial_x e^{\mathrm{i}kx}+O(\varepsilon^2). \tag{8.44}$$

Now consider

$$\chi_k:=-\frac{1+\beta_x}{1+\tau\beta_x}\ \exp(\mathrm{i}k\gamma^\tau(x+\beta(x))),$$

where γ^τ is the flow of the characteristic ODE

$$\frac{d}{d\tau}x=-b(\tau,x). \tag{8.45}$$

By (8.43), the vector field of (8.45) has size $O(\varepsilon)$ and, by similar arguments used above for the flow of (8.43), we have $\gamma^\tau(x)-x=O(\varepsilon)$. By Taylor expansion of the function $\exp(\mathrm{i}k\gamma^\tau(x+\beta(x)))$ at $\beta=0$ we have

$$\chi_k=e^{\mathrm{i}kx}+O(\varepsilon). \tag{8.46}$$

Recalling (8.40) we have

$$\psi_k=(\mathcal{A}^T-\mathrm{I})e^{\mathrm{i}kx}=\varepsilon\partial_x(\beta_1 e^{\mathrm{i}kx})+O(\varepsilon^2)=\varepsilon(\beta_1)_x e^{\mathrm{i}kx}+\varepsilon\beta_1\,\partial_x e^{\mathrm{i}kx}+O(\varepsilon^2). \tag{8.47}$$

Eventually, by (8.44), (8.46) and (8.47), we have $\mathcal{R}_\Phi=\varepsilon\mathrm{R}_\Phi+O(\varepsilon^2)$, where

$$\begin{aligned}\mathrm{R}_\Phi(h):&=-\sum_{k\in S}(h,\beta_1\partial_x e^{\mathrm{i}kx})_{L^2(\mathbb{T})}e^{\mathrm{i}kx}+\sum_{k\in S}(h,(\beta_1)_x e^{\mathrm{i}kx})_{L^2(\mathbb{T})}e^{\mathrm{i}kx}+\sum_{k\in S}(h,\beta_1\partial_x e^{\mathrm{i}kx})_{L^2(\mathbb{T})}e^{\mathrm{i}kx}\\ &=\Pi_S[(\beta_1)_x\,h].\end{aligned} \tag{8.48}$$

By (8.48) the range of R_Φ is orthogonal to the subspace $H_S^\perp$, hence the term $\Phi^{-1}\,R_\Phi(b_3\partial_{yyy}+b_1\partial_y+b_0)$ will have size at least $O(\varepsilon^3)$, indeed $\Phi=\mathrm{I}_{H_S^\perp}+O(\varepsilon)$.
We ignore the terms $\varepsilon^2\mathcal{R}_2$ and $\mathcal{R}_*$ because are too small. Then, we can consider

$$(\mathcal{L}_\omega-\mathcal{D}_\omega)\mathrm{R}_\Phi=\Pi_S^\perp[\partial_{xx}(a_1\partial_x)+\partial_x(a_0\cdot)]\Pi_S^\perp\mathrm{R}_\Phi=0.$$

By (8.40) we have

$$\Pi_S^\perp[\partial_{xx}(a_1\partial_x)+\partial_x(a_0\cdot)]\Pi_S(\mathcal{A}-\mathrm{I})\Pi_S^\perp=\varepsilon^2\Pi_S^\perp[\partial_{xx}(a_{1,1}\partial_{xx}\Pi_S[\beta_1\,\cdot])+\partial_x(a_{0,1}\,\partial_x\Pi_S[\beta_1\,\cdot])]+o(\varepsilon^2). \tag{8.49}$$

It remains to study the commutator $[\mathcal{D}_\omega, \mathcal{R}_\Phi] = [D_{\overline{\omega}}, \mathcal{R}_\Phi^\varepsilon] + O(\varepsilon^3)$. We have

$$[D_{\overline{\omega}}, \mathcal{R}_\Phi^\varepsilon]h = \varepsilon D_{\overline{\omega}} \Pi_S[(\beta_1)_x h] - \varepsilon \Pi_S[(\beta_1)_x D_{\overline{\omega}} h] = \varepsilon \Pi_S[(D_{\overline{\omega}}(\beta_1)_x)h]$$

and so $\Phi^{-1}[\mathcal{D}_\omega, \mathcal{R}_\Phi] = o(\varepsilon^2)$.
Finally, by (8.41), (8.49), we obtained $\mathcal{R}_{II} = \varepsilon^2 \mathrm{R}_2 + o(\varepsilon^2)$, where, for $h \in H_S^\perp$,

$$\begin{aligned}
\mathrm{R}_2[h] &= \Pi_S^\perp\{\partial_x(\beta_1\, \Pi_S[\partial_x(\alpha_{1,1}h)]) - \partial_{xx}(a_{1,1}\partial_x\Pi_S[\partial_x(\beta_1 h)]) - \partial_x(\alpha_{1,1}\Pi_S[\partial_x(\beta_1 h)])\} \\
&= 4\, c_1 c_2\, \Pi_S^\perp\{-\partial_x(v_\delta\, \partial_x\Pi_S[(v_\delta)_{xx}\, h]) + \partial_{xx}((v_\delta)_x\partial_{xx}\Pi_S[(\partial_x^{-1}v_\delta)h]) \\
&+ \partial_{xx}(v_\delta\, \partial_{xx}\Pi_S[v_\delta h]) + \partial_x((v_\delta)_{xx}\, \partial_x\Pi_S[v_\delta h])\} \\
&+ \frac{4}{3}\, c_2^2\, \Pi_S^\perp\{-\partial_x((\partial_x^{-1}v_\delta)\, \partial_x\Pi_S[(v_\delta)_{xx}\, h]) + \partial_{xx}(v_\delta\partial_{xx}\, \Pi_S[(\partial_x^{-1}v_\delta)h]) \\
&+ \partial_x((v_\delta)_{xx}\, \partial_x\Pi_S[(\partial_x^{-1}v_\delta)h])\} \\
&+ 12\, c_1 c_3\, \Pi_S^\perp\{\partial_x(v_\delta\, \partial_x\Pi_S[v_\delta\, h]) - \partial_x(v_\delta\, \partial_x\Pi_S[v_\delta h])\} \\
&+ 4\, c_2 c_3\, \Pi_S^\perp\{\partial_x((\partial_x^{-1}v_\delta)\, \partial_x\Pi_S[v_\delta\, h]) - \partial_x(v_\delta\, \partial_x\Pi_S[(\partial_x^{-1}v_\delta)h])\} \\
&+ 12 c_1^2 \Pi_S^\perp\{\partial_{xx}((v_\delta)_x\, \partial_{xx}\Pi_S[v_\delta h])\}
\end{aligned} \tag{8.50}$$

Using (8.16), (8.18) we get

$$\mathfrak{R}_1 := \Phi^{-1}\mathcal{R}_{\mathrm{II}} = -\varepsilon^2\Pi_S^\perp\partial_x\mathcal{R}_2 + \mathcal{R}_* \tag{8.51}$$

where $\mathcal{R}_2$, defined in (7.32), has been renamed as

$$\mathcal{R}_2 := \mathcal{R}_2 - \partial_x^{-1}\mathrm{R}_2 \tag{8.52}$$

and we have renamed $\mathcal{R}_*$ the term $o(\varepsilon^2)$. Note that $\mathcal{R}_{II}^{\varepsilon^2}[h]$ has zero spatial average for every h belonging to $H^s_{S^\perp}(\mathbb{T}^{\nu+1})$ and the remainder $\mathcal{R}_*$ has the form (7.5).

8.3 Time reduction at the order ∂_{xxx}

The goal of this section is to make constant the coefficient of the highest order spatial derivative operator ∂_{yyy} by a quasi-periodic reparametrization of time. We consider the change of variable

$$(Bw)(\varphi, y) := w(\varphi + \omega\alpha(\varphi), y), \quad (B^{-1}h)(\vartheta, y) := h(\vartheta + \omega\tilde{\alpha}(\vartheta), y), \tag{8.53}$$

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

$$B^{-1}\omega\cdot\partial_\varphi B = \rho(\vartheta)\omega\cdot\partial_\vartheta, \quad B^{-1}\partial_y B = \partial_y, \quad \rho := B^{-1}(1 + \omega\cdot\partial_\varphi\alpha). \tag{8.54}$$

By (8.21), using also that B and B^{-1} commute with $\Pi_S^\perp$, we get

$$B^{-1}\mathcal{L}_1 B = \Pi_S^\perp[\rho\,\omega\cdot\partial_\vartheta + (B^{-1}b_3)\partial_{yyy} + (B^{-1}b_1)\partial_y + (B^{-1}b_0)]\Pi_S^\perp + B^{-1}\mathfrak{R}_1 B. \tag{8.55}$$

We choose α such that the new coefficient at order ∂_{yyy} is proportional to the function $\rho(\vartheta)$, namely

$$(B^{-1}b_3)(\vartheta) = m_3\,\rho(\vartheta), \quad m_3 \in \mathbb{R} \quad \Longrightarrow \quad b_3(\varphi) = m_3(1 + \omega\cdot\partial_\varphi\alpha(\varphi)). \tag{8.56}$$

The unique solution with zero average of (8.56) is

$$\alpha(\varphi) := \frac{1}{m_3}(\omega\cdot\partial_\varphi)^{-1}(b_3 - m_3)(\varphi), \quad m_3 := \frac{1}{(2\pi)^\nu}\int_{\mathbb{T}^\nu} b_3(\varphi)\,d\varphi. \tag{8.57}$$

Hence, by (8.55) we have

$$B^{-1}\mathcal{L}_1 B = \rho\,\mathcal{L}_2, \quad \mathcal{L}_2 := \Pi_S^\perp(\omega\cdot\partial_\vartheta + m_3\partial_{yyy} + c_1\partial_y + c_0)\Pi_S^\perp + \mathfrak{R}_2, \tag{8.58}$$

$$c_1 := \rho^{-1}(B^{-1}b_1), \quad c_0 := \rho^{-1}(B^{-1}b_0), \quad \mathfrak{R}_2 := \rho^{-1}B^{-1}\mathfrak{R}_1 B. \tag{8.59}$$

In order to control the corrections to the normal frequencies also at lower orders of size, we expand the constant coefficient m_3, defined in (8.57), in powers of ε. We have

$$m_3 = 1 + \varepsilon^2 d(\xi) + \mathtt{r}_{m_3} \tag{8.60}$$

where

$$\begin{aligned} d(\xi) : &= (12c_4 - 24\,c_1^2)\, M_{\varphi,x}[\overline{v}_x^2] + \varepsilon^2(2c_6 - \frac{8}{3}c_2^2) M_{\varphi,x}[\overline{v}^2] \\ &= (24c_4 - 48c_1^2)v_3 \cdot \xi + (4c_6 - \frac{16}{3}c_2^2)v_1 \cdot \xi \end{aligned} \tag{8.61}$$

and $|\mathtt{r}_{m_3}|^{Lip(\gamma)} \leq \varepsilon^3$. The transformed operator $\mathcal{L}_2$ in (8.58) is still Hamiltonian, since the reparametrization of time preserves the Hamiltonian structure (see Section 2.2 and Remark 3.7 in [2]).
We note that, by (8.59), for $k = 0, 1$, we have

$$c_k = b_k + (B^{-1} - \mathrm{I})b_k + (\rho^{-1} - 1)\, B^{-1}\, b_k$$

and $b_k = O(\varepsilon)$ is the biggest term in the expression above. We define, for $k = 0, 1$,

$$\tilde{c}_k := c_k - b_k = (B^{-1} - \mathrm{I})b_k + (\rho^{-1} - 1)B^{-1}b_k \tag{8.62}$$

and we estimate them in Lemma 8.9. The remainder $\mathfrak{R}_2$ in (8.59) has still the form (7.5) and, by (8.51),

$$\mathfrak{R}_2 := -\rho^{-1}B^{-1}\mathfrak{R}_1 B = -\varepsilon^2 \Pi_S^{\perp}\partial_x \mathcal{R}_2 + \mathcal{R}_* \tag{8.63}$$

where $\mathcal{R}_2$ is defined in (8.52) and we have renamed $\mathcal{R}_*$ the term of order $o(\varepsilon^2)$ in $\mathfrak{R}_2$.

Remark 8.8. In the proof of the estimates for the transformations B and $\mathcal{T}$, respectively defined in (8.53) and (8.75), we have to give a bound to the inverse of the operator $\mathcal{D}_\omega$ applied to the difference of a spatial and total (in space and time) average of some function in $H^s_{S^\perp}(\mathbb{T}^{\nu+1})$.
The main problem is that the estimate (5.23) is too rough to deal with functions $h(\varphi, x)$ of size greater or equal than ε^3, indeed, the terms $O(\varepsilon^3\gamma^{-1})$ are just not perturbative.
In the proofs of Lemma 8.9 and 8.10, we exploit the fact that if $h(\varphi, x)$ is a function supported on few harmonics, then we do not need to use the diophantine inequality (5.3) to give a bound to the divisors appearing in the Fourier coefficients of $\mathcal{D}_\omega^{-1}h$.
In this way, we overcome the problem discussed in Remark 8.11 in [3] and we can drop the hypotesis (1.20) on the tangential sites assumed in [3].

Lemma 8.9. *There is $\sigma = \sigma(\nu, \tau) > 0$ (possibly larger than the one in Lemma 8.5) such that*

$$|m_3 - 1|^{Lip(\gamma)} \leq C\,\varepsilon^2, \quad |\partial_i m_3[\hat{\imath}]| \leq \varepsilon^2 \|\hat{\imath}\|_{s_0+\sigma}, \tag{8.64}$$

$$\|\alpha\|_s^{Lip(\gamma)} \leq_s \varepsilon^4\gamma^{-1} + \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad \|\partial_i\alpha[\hat{\imath}]\|_s \leq_s \|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma} \tag{8.65}$$

$$\|\rho - 1\|_s^{Lip(\gamma)} \leq_s \varepsilon^3 + \varepsilon^{2b}\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad \|\partial_i\rho[\hat{\imath}]\|_s \leq_s \varepsilon^{2b}(\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}), \tag{8.66}$$

$$\|\tilde{c}_k\|_s^{Lip(\gamma)} \leq_s \varepsilon^{3-2a} + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad \|\partial_i\tilde{c}_k[\hat{\imath}]\|_s \leq_s \varepsilon(\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}). \tag{8.67}$$

Proof. Estimate (8.64): We have $m_3 - 1 = \fint_{\mathbb{T}^\nu}(b_3 - 1)\, d\varphi$, then, by (8.23),

$$|m_3 - 1| \leq \int_{\mathbb{T}^\nu} |b_3 - 1| d\varphi \leq \|b_3 - 1\|_{s_0} \leq C\varepsilon^2, \quad |\partial_i m_3[\hat{\imath}]| \leq \int_{\mathbb{T}^\nu} \partial_i b_3[\hat{\imath}] d\varphi \leq \|\partial_i b_3[\hat{\imath}]\|_{s_0} \leq \varepsilon^2\|\hat{\imath}\|_{s_0+2}.$$

Estimate (8.65): By (8.57) and the fact that m_3 is a constant near to 1, it is sufficient to give a bound to $b_3 - m_3$.
Consider the functions $g(t) = (1+t)^{-\frac{1}{3}}, \Upsilon(t) = (1+t)^{-3}$, defined in a small neighbourhood of the origin.
We have

$$b_3 - m_3 = (b_3 - 1) - M_\varphi[b_3 - 1] \overset{(8.27)}{=} \Upsilon[M_x[g(a_1 - 1) - g(0)]] - M_\varphi[\Upsilon[M_x[g(a_1 - 1) - g(0)]]]. \tag{8.68}$$

By the analiticity of Υ

$$\Upsilon(t) - \Upsilon(0) = \Upsilon'(0)\, t + \Upsilon_{\geq 2}[t], \qquad \Upsilon_{\geq 2}[t] := \sum_{k\geq 2} \frac{\Upsilon^{(k)}(0)}{k!}\, t^k,$$

for $|t|$ small enough. Hence, by (8.68),

$$\begin{aligned} b_3 - m_3 = \Upsilon'(0)\{M_x[g(a_1-1)-g(0)] - M_{\varphi,x}[g(a_1-1)-g(0)]\} \\ + \Upsilon_{\geq 2}[M_x[g(a_1-1)-g(0)]] - M_\varphi[\Upsilon_{\geq 2}[M_x[g(a_1-1)-g(0)]]]. \end{aligned} \tag{8.69}$$

The difference of the last two terms in the right hand side of (8.69) can be estimated by

$$\begin{aligned} &\|\Upsilon_{\geq 2}[M_x[g(a_1-1)-g(0)]] - M_\varphi[\Upsilon_{\geq 2}[M_x[g(a_1-1)-g(0)]]]\|_s \\ &\leq_s \|M_x[g(a_1-1)-g(0)]\|_{s_0} \|M_x[g(a_1-1)-g(0)]\|_s \overset{(8.23)}{\leq_s} \varepsilon^4(1+\|\mathfrak{I}_\delta\|_s). \end{aligned}$$

Now we prove a bound for the difference $M_x[g(a_1-1)-g(0)] - M_{\varphi,x}[g(a_1-1)-g(0)]$.
By Taylor expansion

$$g(a_1-1)-g(0) = g'(0)(a_1-1) + \frac{g''(0)}{2}(a_1-1)^2 + \frac{g'''(0)}{3!}(a_1-1)^3 + \frac{(a_1-1)^4}{6}\int_0^1 (1-s)^3\, g^{(4)}(s(a_1-1))\, ds$$

and the last term of the right hand side can be estimated by $\varepsilon^4(1+\|\mathfrak{I}_\delta\|_{s+\sigma})$.
The function a_1 in (7.37) is a linear combination of $\Phi_B(T_\delta), \Phi_B(T_\delta)^2$ (and their derivatives in the x-variable) and $r_1(T_\delta)$, whose coefficients depend on $c_1, \dots, c_7$ and other real constants. Without loss of generality, to simplify the notations, we can write $a_1 = 1 + \Phi_B(T_\delta) + \Phi_B(T_\delta)^2 + r_1(T_\delta)$ (recall (7.10), (7.29) and (7.25)). Thus, we have

$$M_x[a_1-1] = M_x[\Phi_B(T_\delta)^2] + M_x[r_1(T_\delta)], \quad M_x[(a_1-1)^2] = M_x[\Phi_B(T_\delta)^2] + 2M_x[\Phi_B(T_\delta)^3] + \mathsf{Q}_2(T_\delta),$$
$$M_x[(a_1-1)^3] = 4M_x[\Phi_B(T_\delta)^3] + \mathsf{Q}_3(T_\delta),$$

where $\|\mathsf{Q}_i(T_\delta)\|_s \leq_s \varepsilon^4 + \varepsilon^{2+b}\|\mathfrak{I}_\delta\|_{s+\sigma}$ for $i=2,3$. By (7.25) and the fact that $\Phi_B(T_\delta)$ has size $O(\varepsilon)$, $r_1(T_\delta)$ is a polynomial of degree three in the variables $(\Phi_B(T_\delta), \Phi_B(T_\delta)_x)$, up to a remainder that is bounded in H^s norm by $\varepsilon^4(1+\|\mathfrak{I}_\delta\|_{s+\sigma})$. Thus, we reduced to study the differences

$$M_x[\Phi_B(T_\delta)^2] - M_{\varphi,x}[\Phi_B(T_\delta)^2], \qquad M_x[\Phi_B(T_\delta)^3] - M_{\varphi,x}[\Phi_B(T_\delta)^3].$$

We have, up to constants,

$$\Phi_B(T_\delta)^2 = \varepsilon^2 v_\delta^2 + \varepsilon v_\delta \tilde{q} + \varepsilon^3 v_\delta \Psi_2(v_\delta) + \tilde{\mathsf{Q}}_2(T_\delta), \qquad \Phi_B(T_\delta)^3 = \varepsilon^3 v_\delta^3 + \tilde{\mathsf{Q}}_3(T_\delta),$$

where $\|\tilde{\mathsf{Q}}_i(T_\delta)\|_s \leq_s \varepsilon^4 + \varepsilon^{2+b}\|\mathfrak{I}_\delta\|_{s+\sigma}$ for $i=2,3$. By the definition of $\tilde{q}$ and the fact that v_δ and z_0 are orthogonal in $L^2(\mathbb{T})$, we have

$$\varepsilon M_x[v_\delta \tilde{q}] = \varepsilon^{2+b} M_x[\Psi_2'(v_\delta) v_\delta z_0] + \varepsilon M_x[v_\delta\, \Psi_3(T_\delta)], \tag{8.70}$$

thus $\|M_x[\varepsilon v_\delta \tilde{q}] - M_{\varphi,x}[\varepsilon v_\delta \tilde{q}]\|_s \leq_s \varepsilon^4 + \varepsilon^{2+b}\|\mathfrak{I}_\delta\|_{s+\sigma}$. It remains to estimate the differences of the averages of polynomial of degree two and three in the variables v_δ and its derivatives. These functions are of order ε^2 and ε^3, respectively, and supported on not many harmonics, because v_δ is not.
By (7.10) we get

$$M_x[v_\delta^2] - M_{\varphi,x}[v_\delta^2] = \varepsilon^{2(b-1)} \sum_{j\in S} |j|((y_\delta)_j - M_\varphi[(y_\delta)_j]).$$

We gain an extra smallness factor $\varepsilon^{2(b-1)}$ by the fact that $M_x[\overline{v}^2]$ is independent of φ (see Remark 8.1). Thus, we obtain $\varepsilon^2\|M_x[v_\delta^2] - M_{\varphi,x}[v_\delta^2]\|_s \leq_s \varepsilon^{2b}\|\mathfrak{I}_\delta\|_s$.
For the cubic terms in v_δ we use the following equality

$$M_x[v_\delta^3] - M_{\varphi,x}[v_\delta^3] = (M_x[\overline{v}^3] - M_{\varphi,x}[\overline{v}^3]) + M_x[v_\delta^3 - \overline{v}^3] - M_{\varphi,x}[v_\delta^3 - \overline{v}^3], \tag{8.71}$$

where $\|M_x[v_\delta^3 - \overline{v}^3] - M_{\varphi,x}[v_\delta^3 - \overline{v}^3]\|_s \leq_s \varepsilon^3\|\mathfrak{I}_\delta\|_s$.
We now analyze the first difference in the right hand side of (8.71). We cannot roughly bound it by ε^3 (see Remark 8.8). But we have

$$\mathcal{D}_\omega^{-1}\left(M_x[\overline{v}^3] - M_{\varphi,x}[\overline{v}^3]\right) = \sum_{\substack{j_1,j_2,j_3\in S,\\ j_1+j_2+j_3=0\\ \mathtt{l}(j_1)+\mathtt{l}(j_2)+\mathtt{l}(j_3)\neq 0}} \frac{\sqrt{\xi_{j_1}\xi_{j_2}\xi_{j_3}}}{\mathrm{i}\,\omega\cdot(\mathtt{l}(j_1)+\mathtt{l}(j_2)+\mathtt{l}(j_3))}\, e^{\mathrm{i}(\mathtt{l}(j_1)+\mathtt{l}(j_2)+\mathtt{l}(j_3))\cdot\varphi}. \tag{8.72}$$

We recall that $\omega = \overline{\omega} + O(\varepsilon^2)$, hence the denominator in (8.72) can be written as

$$\omega\cdot(\mathtt{l}(j_1)+\mathtt{l}(j_2)+\mathtt{l}(j_3)) = \overline{\omega}\cdot(\mathtt{l}(j_1)+\mathtt{l}(j_2)+\mathtt{l}(j_3)) + (\omega-\overline{\omega})\cdot(\mathtt{l}(j_1)+\mathtt{l}(j_2)+\mathtt{l}(j_3)) = j_1^3+j_2^3+j_3^3+O(\varepsilon^2)$$

and it is greater or equal than 1, indeed, if $j_1+j_2+j_3=0$, then $|j_1^3+j_2^3+j_3^3| = 3|j_1\,j_2\,j_3| \geq 3$. Thus, actually,

$$\|\mathcal{D}_\omega^{-1}\left(M_x[\overline{v}^3] - M_{\varphi,x}[\overline{v}^3]\right)\|_s \leq \varepsilon^3.$$

Finally, we get

$$\|b_3 - m_3\|_s \leq_s \varepsilon^3 + \varepsilon^{2b}\|\mathfrak{I}_\delta\|_{s+\sigma} \qquad \text{and} \qquad \|\mathcal{D}_\omega^{-1}(b_3 - m_3)\|_s \leq_s \varepsilon^4\gamma^{-1} + \|\mathfrak{I}_\delta\|_{s+\sigma}, \tag{8.73}$$

so $\|\alpha\|_s \leq_s \varepsilon^4\gamma^{-1} + \|\mathfrak{I}_\delta\|_{s+\sigma}$.
Now we look to the partial derivative

$$\partial_i\left(\frac{b_3 - m_3}{m_3}\right)[\hat\imath] = \frac{1}{m_3^2}\left[m_3\partial_i(b_3-m_3)[\hat\imath] - (b_3-m_3)\partial_i m_3[\hat\imath]\right]. \tag{8.74}$$

By (8.64) $m_3 - 1$ and $\partial_i m_3[\hat\imath]$ are of order ε^2, hence the estimate for $\partial_i\alpha[\hat\imath]$ comes from $\mathcal{D}_\omega^{-1}(\partial_i(b_3-m_3)[\hat\imath])$. By (8.69) we have

$$\begin{aligned}\partial_i(b_3-m_3)[\hat\imath] = \Upsilon'(0)\{&M_x[\partial_i(g(a_1-1)-g(0))[\hat\imath]] - M_{\varphi,x}[\partial_i(g(a_1-1)-g(0))[\hat\imath]]\}\\ &+ \partial_i\{\Upsilon_{\geq 2}[M_x[g(a_1-1)-g(0)]] - M_\varphi[\Upsilon_{\geq 2}[M_x[g(a_1-1)-g(0)]]]\}[\hat\imath]\end{aligned}$$

As before, the bigger terms are the partial derivatives of $M_x[g(a_1-1)-g(0)] - M_{\varphi,x}[g(a_1-1)-g(0)]$. We have

$$\partial_i(g(a_1-1)-g(0))[\hat\imath] = g'(0)\partial_i a_1[\hat\imath] + g''(0)(a_1-1)\,\partial_i a_1[\hat\imath] + \frac{g'''(0)}{2}(a_1-1)^2\,\partial_i a_1[\hat\imath] + \mathtt{T}(i_\delta,\hat\imath)$$

where $\|\mathtt{T}(i_\delta,\hat\imath)\|_s \leq_s \varepsilon^4(\|\hat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s_0+\sigma})$ and

$$\partial_i a_1[\hat\imath] = \partial_i\Phi_B(T_\delta)[\hat\imath] + 2\Phi_B(T_\delta)\partial_i\Phi_B(T_\delta)[\hat\imath] + \partial_i\tilde{q}[\hat\imath].$$

We note that $M_x[\partial_i\Phi_B(T_\delta)[\hat\imath]] = M_x[\partial_i\tilde{q}[\hat\imath]] = 0$. Thus, we focus on the terms

$$\Phi_B(T_\delta)\partial_i\Phi_B(T_\delta), \quad \Phi_B(T_\delta)^2\partial_i\Phi_B(T_\delta), \quad \Phi_B(T_\delta)\,\partial_i\tilde{q}[\hat\imath].$$

Further terms have Sobolev norm bounded by $\varepsilon^{2+b}(\|\hat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s_0+\sigma})$. We have

$$\begin{aligned}\Phi_B(T_\delta)\partial_i\Phi_B(T_\delta) &= \varepsilon^2 v_\delta\,\partial_i v_\delta[\hat\imath] + \varepsilon^3(\Psi_2(v_\delta) + \Psi_2'(v_\delta)v_\delta)\partial_i v_\delta[\hat\imath] + \varepsilon\partial_i(\tilde{q}\,v_\delta)[\hat\imath] + \tilde{\mathtt{T}}(i_\delta,\hat\imath),\\ \Phi_B(T_\delta)^2\partial_i\Phi_B(T_\delta)[\hat\imath] &= \varepsilon^3 v_\delta^2\,\partial_i v_\delta[\hat\imath] + \tilde{\mathtt{T}}(i_\delta,\hat\imath),\end{aligned}$$

where $\|\mathtt{T}(i_\delta, \hat\imath)\|_s \leq_s \varepsilon^{2+b}(\|\hat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s_0+\sigma})$. We start from the average of the partial derivative of $v_\delta \tilde q$. By (8.70) we get $\varepsilon\|\partial_i M_x[v_\delta \tilde q][\hat\imath]\|_s \leq_s \varepsilon^{2+b}(\|\hat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s_0+\sigma})$. Then, we reduce to study

$$M_x[v_\delta \partial_i v_\delta[\hat\imath]] - M_{\varphi,x}[v_\delta \partial_i v_\delta[\hat\imath]], \qquad M_x[v_\delta^2 \partial_i v_\delta[\hat\imath]] - M_{\varphi,x}[v_\delta^2 \partial_i v_\delta[\hat\imath]].$$

If we call $G(i_0(\varphi)) := y_\delta - y_0$, then we have

$$\partial_i v_\delta[\hat\imath] = \sum_{j\in S} \sqrt{|j|}\sqrt{\xi_j + \varepsilon^{2(b-1)}(y_\delta)_j} e^{\mathrm{i}(\theta_0)_j} \left(\mathrm{i}\hat\Theta_j + \varepsilon^{2(b-1)} \frac{\hat y_j + (\partial_i G(i_0(\varphi))[\hat\imath])_j}{2\,|j|\,(\xi_j + \varepsilon^{2(b-1)}(y_\delta)_j)} \right) e^{\mathrm{i}jx}$$

and

$$\begin{aligned} M_x[v_\delta \partial_i v_\delta[\hat\imath]] - M_{\varphi,x}[v_\delta \partial_i v_\delta[\hat\imath]] &= \varepsilon^{2(b-1)} \sum_{j\in S} \mathrm{i}\hat\Theta_j((y_\delta)_j - M_\varphi[(y_\delta)_j]) \\ &\quad + \frac{\varepsilon^{2(b-1)}}{2} \sum_{j\in S} \{(\partial_i G(i_0(\varphi))[\hat\imath])_j - M_\varphi[(\partial_i G(i_0(\varphi))[\hat\imath])_j]\}. \end{aligned}$$

Therefore, $\varepsilon^2\|M_x[v_\delta \partial_i v_\delta[\hat\imath]] - M_{\varphi,x}[v_\delta \partial_i v_\delta[\hat\imath]]\|_s \leq_s \varepsilon^{2b}(\|\hat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s_0+\sigma})$. Moreover, we have $\|\varepsilon^3 M_x[v_\delta^2 \partial_i v_\delta[\hat\imath]]\|_s \leq_s \varepsilon^3(\|\hat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s_0+\sigma})$. Hence, we get

$$\|\partial_i(b_3 - m_3)[\hat\imath]\|_s \leq_s \varepsilon^{2b}(\|\hat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s_0+\sigma})$$

and $\|\partial_i \alpha[\hat\imath]\| \leq_s \|\hat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s_0+\sigma}$.By Lemma 2.7 we deduce the inequality (8.65).

Estimate (8.66): Note that $\rho - 1 = B^{-1}\ ((b_3 - m_3)/m_3)$. Thus, by Lemma 2.4, (8.65), (8.73) we get

$$\begin{aligned} \|B^{-1}\ ((b_3 - m_3)/m_3)\|_s^{Lip(\gamma)} &\leq_s \|b_3 - m_3\|_{s+1}^{Lip(\gamma)} + \|\alpha\|_{s+s_0}^{Lip(\gamma)}\|b_3 - m_3\|_2^{Lip(\gamma)} \\ &\leq_s \varepsilon^3 + \varepsilon^{2b}\|\mathfrak{I}_\delta\|_{s+s_0+\sigma}^{Lip(\gamma)}. \end{aligned}$$

Estimate (8.67): Note that $\|\rho^{-1} - 1\|_s \leq_s \|\rho - 1\|_s$. By Lemma 2.4 and (2.10), (8.24), we get, for $k = 0, 1$,

$$\|(B^{-1} - \mathrm{I})\, b_k\|_s^{Lip(\gamma)} \leq_s \varepsilon^7\,\gamma^{-2} + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad \|(\rho^{-1} - 1)b_k\|_s^{Lip(\gamma)} \leq_s \varepsilon^4 + \varepsilon^{1+2b}\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}.$$

□

8.4 Translation of the space variable

The goal of this section is to remove the space average from the coefficient in front of ∂_y. This is a preliminary step for the descent method that we apply at Section 8.7.
Consider the change of variable

$$(\mathcal{T}w)(\vartheta, y) = w(\vartheta, y + p(\vartheta)), \quad (\mathcal{T}^{-1}h)(\vartheta, z) = h(\vartheta, z - p(\vartheta)). \tag{8.75}$$

The differential operators in $\mathcal{L}_2$ (see (8.58)) transform into

$$\mathcal{T}^{-1}\omega \cdot \partial_\vartheta \mathcal{T} = \omega \cdot \partial_\vartheta + \{\omega \cdot \partial_\vartheta p(\vartheta)\}\partial_z, \quad \mathcal{T}^{-1}\partial_y \mathcal{T} = \partial_z.$$

Since $\mathcal{T}, \mathcal{T}^{-1}$ commute with $\Pi_S^\perp$, we get

$$\mathcal{L}_3 := \mathcal{T}^{-1}\mathcal{L}_2\mathcal{T} = \Pi_S^\perp(\omega \cdot \partial_\vartheta + m_3\,\partial_{zzz} + D_S\,\partial_z + d_0)\Pi_S^\perp + \mathfrak{R}_3, \tag{8.76}$$

$$d_1 := (\mathcal{T}^{-1}c_1) + \omega \cdot \partial_\vartheta p, \quad d_0 := \mathcal{T}^{-1}c_0, \quad \mathfrak{R}_3 := \mathcal{T}^{-1}\mathfrak{R}_2\mathcal{T} \tag{8.77}$$

and we choose

$$m_1 := \frac{1}{(2\pi)^{\nu+1}} \int_{\mathbb{T}^{\nu+1}} c_1\, d\vartheta\, dy, \quad p := (\omega \cdot \partial_\vartheta)^{-1}\left(m_1 - \frac{1}{2\pi}\int_{\mathbb{T}} c_1\, dy\right) \tag{8.78}$$

so that

$$\frac{1}{2\pi}\int_{\mathbb{T}} d_1(\vartheta, z)\, dz = m_1 \quad \forall \vartheta \in \mathbb{T}^\nu. \tag{8.79}$$

We define

$$\tilde{d}_k := d_k - \varepsilon\, \alpha_{k,1} - \varepsilon^2(\alpha_{k,2} - \alpha_{k,1}\,(\beta_1)_x), \qquad k = 0, 1 \tag{8.80}$$

and we split $\mathfrak{R}_3 = -\varepsilon^2 \partial_x \overline{\mathcal{R}}_2 + \tilde{\mathcal{R}}_*$, where $\overline{\mathcal{R}}_2$ is obtained replacing v_δ with $\overline{v}$ in $\mathcal{R}_2$ and

$$\tilde{\mathcal{R}}_* := \mathcal{T}^{-1}\mathcal{R}_*\mathcal{T} + \varepsilon^2 \Pi_S^\perp \partial_x(\mathcal{R}_2 - \mathcal{T}^{-1}\mathcal{R}_2\mathcal{T}) + \varepsilon^2 \Pi_S^\perp \partial_x(\overline{\mathcal{R}}_2 - \mathcal{R}_2), \tag{8.81}$$

where $\mathcal{R}_*$ has been defined in (7.34) and modified along this section by adding terms $o(\varepsilon^2)$. We used that $\mathcal{T}^{-1}$ commutes with ∂_x and $\Pi_S^\perp$.
We define

$$c(\xi) := M_{\varphi,x}[\alpha_{1,2} + \alpha_{1,1}\,(\beta_1)_x]. \tag{8.82}$$

This quantity is a correction at order ε^2 to the eigenvalues of the linear operator $\mathcal{L}_\omega$, see (7.33). In particular, we have

$$m_1 = \varepsilon^2 c(\xi) + \mathtt{r}_{m_1}, \quad \text{with} \quad |\mathtt{r}_{m_1}|^{Lip(\gamma)} \leq \varepsilon^{3-2a}.$$

Lemma 8.10. *There is $\sigma := \sigma(\tau, \nu)$ (possibly larger than in Lemma 8.9) such that*

$$|m_1 - \varepsilon^2 c(\xi)|^{Lip(\gamma)} \leq \varepsilon^7\, \gamma^{-2}, \quad |\partial_i(m_1 - \varepsilon^2 c(\xi))[\hat{\imath}]| \leq \varepsilon^7\, \gamma^{-2} \|\hat{\imath}\|_{s_0+\sigma}, \tag{8.83}$$

$$\|p\|_s^{Lip(\gamma)} \leq_s \varepsilon^4 \gamma^{-1} + \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad \|\partial_i p[\hat{\imath}]\|_s \leq_s \|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}, \tag{8.84}$$

$$\|\tilde{d}_k\|_s^{Lip(\gamma)} \leq_s \varepsilon^{3-2a} + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad \|\partial_i \tilde{d}_k[\hat{\imath}]\|_s \leq_s \varepsilon(\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}) \tag{8.85}$$

for $k = 0, 1$. Moreover the matrix s-decay norm (see (2.6)*)*

$$|\tilde{\mathcal{R}}_*|_s^{Lip(\gamma)} \leq_s \varepsilon^3 + \varepsilon^2 \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad |\partial_i \tilde{\mathcal{R}}_*[\hat{\imath}]|_s \leq_s \varepsilon^2 \|\hat{\imath}\|_{s+\sigma} + \varepsilon^{2b-1}\|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}. \tag{8.86}$$

The transformations $\mathcal{T}, \mathcal{T}^{-1}$ satisfy (8.25), (8.26).

Proof. *Estimate* (8.84): By (8.59) and (8.78) we have

$$\begin{aligned} m_1 - M_x[c_1] = {} & (M_{\varphi,x}[b_1] - M_x[b_1]) + (M_{\varphi,x}[(\rho^{-1} - 1)b_1] - M_x[(\rho^{-1} - 1)b_1]) \\ & + (M_{\varphi,x}[(B^{-1} - \mathrm{I})b_1] - M_x[(B^{-1} - \mathrm{I})b_1]) \\ & + (M_{\varphi,x}[(\rho^{-1} - 1)(B^{-1} - \mathrm{I})b_1] - M_x[(\rho^{-1} - 1)(B^{-1} - \mathrm{I})b_1]). \end{aligned} \tag{8.87}$$

By (8.65), (8.66) and Lemma 2.10, we get $\|(\rho^{-1} - 1)(B^{-1} - \mathrm{I})b_1\|_s \leq_s \varepsilon^9 \gamma^{-2} + \varepsilon^6 \gamma^{-1} \|\mathfrak{I}_\delta\|_{s+\sigma}$. Thus, by (5.23)

$$\|\mathcal{D}_\omega^{-1}\{M_{\varphi,x}[(\rho^{-1} - 1)(B^{-1} - \mathrm{I})b_1] - M_x[(\rho^{-1} - 1)(B^{-1} - \mathrm{I})b_1]\}\|_s \leq_s \varepsilon^9 \gamma^{-3} + \varepsilon^6 \gamma^{-2} \|\mathfrak{I}_\delta\|_{s+\sigma}. \tag{8.88}$$

We note that $\rho^{-1} - 1$ is independent of x, hence $M_x[(\rho^{-1} - 1)b_1] = (\rho^{-1} - 1)M_x[b_1]$ and we can estimate the difference between the averages of $(\rho^{-1} - 1)b_1$ with

$$\|(\rho^{-1} - 1)M_x[b_1]\|_s \leq_s \varepsilon^5 + \varepsilon^{2(b+1)}\|\mathfrak{I}_\delta\|_{s+\sigma} \tag{8.89}$$

and use again (5.23) for $\|\mathcal{D}_\omega^{-1}(M_{\varphi,x}[(\rho^{-1} - 1)b_1] - M_x[(\rho^{-1} - 1)b_1])\|_s \leq_s \varepsilon^5 \gamma^{-1} + \varepsilon \|\mathfrak{I}_\delta\|_{s+\sigma}$.
By Taylor expansion and the fact that $\tilde{\alpha} = -\alpha + (B - \mathrm{I})\alpha$ (see (8.53)), we have

$$b_1(\vartheta + \omega\tilde{\alpha}(\vartheta), x) = b_1(\vartheta, x) - \omega \cdot \partial_\vartheta b_1(\vartheta, x)\, \alpha(\vartheta) + \mathtt{R}_{\tilde{\alpha}}(\vartheta, x)$$

where $\|\mathtt{R}_{\tilde{\alpha}}\|_s \leq_s \varepsilon^8 \gamma^{-2} + \varepsilon^4 \gamma^{-1} \|\mathfrak{I}_\delta\|_{s+\sigma}$. Moreover, by a change of variable

$$\int_{\mathbb{T}^{\nu+1}} (B^{-1} - \mathrm{I})b_1\, d\vartheta\, dx = \int_{\mathbb{T}^{\nu+1}} \omega \cdot \partial_\varphi \alpha(\varphi)\, b_1(\varphi, x)\, d\varphi\, dx. \tag{8.90}$$

From these facts and an integration by parts, we obtain

$$M_x[(B^{-1}-\mathrm{I})b_1] - M_{\varphi,x}[(B^{-1}-\mathrm{I})b_1] = \mathcal{D}_\omega\alpha\, M_x[b_1] - M_{\varphi,x}[(\mathcal{D}_\omega\alpha)\, b_1] + M_x[\mathtt{R}_{\tilde\alpha}] - M_{\varphi,x}[\mathtt{R}_{\tilde\alpha}]$$

and, by the estimate above for $\mathtt{R}_{\tilde\alpha}$ and the bound given by (8.73) for $\mathcal{D}_\omega\alpha$, we have

$$\|M_x[(B^{-1}-\mathrm{I})b_1] - M_{\varphi,x}[(B^{-1}-\mathrm{I})b_1]\|_s \leq_s \varepsilon^5 + \varepsilon^{2(b+1)}\|\mathfrak{I}_\delta\|_{s+\sigma}. \tag{8.91}$$

As before, we can use (5.23). We remark that

$$\int_{\mathbb{T}} b_1(\varphi,y)\,dy = \int_{\mathbb{T}} (\mathcal{A}^T\alpha_1)(\varphi,y)\,dy = \int_{\mathbb{T}} \alpha_1(\varphi, y+\tilde\beta(\varphi,y))\,dy = \int_{\mathbb{T}} \alpha_1(\varphi,x)(1+\beta_x(\varphi,x))\,dx,$$

hence, it remains to estimate

$$M_{\varphi,x}[b_1] - M_x[b_1] = (M_{\varphi,x}[\alpha_1] - M_x[\alpha_1]) + (M_{\varphi,x}[\alpha_1\beta_x] - M_x[\alpha_1\beta_x]). \tag{8.92}$$

The functions α_1 and $\alpha_1\beta_x$ are linear combinations of powers of $\Phi_B(T_\delta)$ (and its derivatives in the x-variable), $r_1(T_\delta)$, $r_0(T_\delta)$, whose coefficients depend on $c_1,\dots,c_7$ and other real constants. Hence, using the same reasoning adopted in the proof of the estimates (8.65), we get

$$\|\mathcal{D}_\omega^{-1}\{M_{\varphi,x}[\alpha_1] - M_x[\alpha_1]\}\|_s \leq_s \varepsilon^4\gamma^{-1} + \|\mathfrak{I}_\delta\|_{s+\sigma} \tag{8.93}$$

and the same estimate holds for $\mathcal{D}_\omega^{-1}\{M_{\varphi,x}[\beta_x\alpha_1] - M_x[\beta_x\alpha_1]\}$. By following analogous arguments used in the proof of the estimate (8.65) we conclude.

Estimate (8.83): By (8.59) and (8.78)

$$m_1 = \int_{\mathbb{T}^{\nu+1}} b_1\,dx\,d\varphi + \int_{\mathbb{T}^{\nu+1}} \tilde c_1\,dx\,d\varphi.$$

Moreover,

$$\int_{\mathbb{T}^{\nu+1}} b_1\,dx\,d\varphi = \int_{\mathbb{T}^{\nu+1}} (\varepsilon^2\alpha_{1,2} + \mathtt{R}_{a_1})\,dx\,d\varphi + \int_{\mathbb{T}^{\nu+1}} (\mathcal{A}^T-\mathrm{I})\alpha_1\,dx\,d\varphi.$$

Thus, the bound (8.83) comes from taking the maximum between

$$\left|\int_{\mathbb{T}^{\nu+1}} b_1\,d\varphi\,dy - \varepsilon^2\int_{\mathbb{T}^{\nu+1}} (\alpha_{1,2} + \alpha_{1,1}\,(\beta_1)_x)\,d\varphi\,dx\right| \leq \|\mathtt{R}_{a_1}\|_{s_0} + \|(\mathcal{A}^T-\mathrm{I})(\alpha_1 - \varepsilon\alpha_{1,1})\|_{s_0} \leq \varepsilon^3$$

and $\|\tilde c_1\|_{s_0} \leq \varepsilon^7\gamma^{-2} = \varepsilon^{3-2a}$.

Estimate (8.85): We observe that, by (8.38),

$$\tilde d_0 := \varepsilon^2(\mathcal{A}^T - I)\alpha_{0,2} + \mathcal{A}^T\mathtt{R}_0 + \mathcal{R}_{\tilde\beta} + (\mathcal{T}^{-1}-\mathrm{I})b_0 + \mathcal{T}^{-1}\tilde c_0.$$

By Lemma 2.6, 2.4 we have the following bounds

$$\begin{aligned}
\|\varepsilon^2(\mathcal{A}^T-I)\alpha_{0,2}\|_s &\leq_s \varepsilon^3(1+\|\mathfrak{I}_\delta\|_{s+\sigma}), \qquad & \|\mathcal{A}^T\mathtt{R}_0\|_s^{Lip(\gamma)} &\leq_s \varepsilon^3 + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)},\\
\|\mathcal{R}_{\tilde\beta}\|_s^{Lip(\gamma)} &\leq_s \varepsilon^3 + \varepsilon^{1+b}\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \qquad & \|\mathcal{T}^{-1}\tilde c_0\|_s^{Lip(\gamma)} &\leq_s \varepsilon^7\gamma^{-2} + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)},\\
\|(\mathcal{T}^{-1}-\mathrm{I})b_0\|_s &\leq_s \varepsilon^7\gamma^{-2} + \varepsilon\|\mathfrak{I}\|_{s+\sigma}.
\end{aligned}$$

From these estimates we get (8.85) for $k=0$. The estimate for $k=1$ can be obtained in the same way, considering that $\omega\cdot\partial_\vartheta p = O(\varepsilon^6\gamma^{-1})$ in low norm by (8.84). $\square$

8.5 Linear Birkhoff Normal Form (Step one)

Let us collect all the terms of order ε and ε^2 of $\mathcal{L}_3$ (see (8.76)) in the operators

$$\begin{aligned} \mathfrak{B}_1[h] &:= \alpha_{1,1}\,\partial_x h + \alpha_{0,1}\,h = \partial_x\{(2c_2 v_{xx} - 6c_3 v)\,h\}, \\ \mathfrak{B}_2[h] &:= \{\alpha_{1,2} - (\alpha_{1,1})_x\,\beta_1\}\,\partial_x h + \{\alpha_{0,2} - (\alpha_{0,1})_x\,\beta_1\}\,h - \partial_x \overline{\mathcal{R}}_2[h]. \end{aligned} \tag{8.94}$$

Note that $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are not the linear Hamiltonian vector fields of $H_S^\perp$ generated, respectively, by the Hamiltonians $R(v^2 z)$ and $R(v^2 z^2)$ in (3.1) at $v = \overline{v}$, as expected. Indeed, as we said in Remark 8.6, some Hamiltonians of type $R(v^2 z)$ have been eliminated by the diffeomorphism of the torus Φ defined in Section 8.1, and also the Hamiltonians $R(v^2 z^2)$ have been modified by that.
Renaming $\vartheta = \varphi, z = x$ we have

$$\mathcal{L}_3 = \Pi_S^\perp(\omega\cdot\partial_\varphi + m_3\partial_{xxx} + \varepsilon\mathfrak{B}_1 + \varepsilon^2\mathfrak{B}_2 + \tilde{d}_1\partial_x + \tilde{d}_0)\Pi_S^\perp + \tilde{\mathcal{R}}_* \tag{8.95}$$

where $\tilde{d}_1, \tilde{d}_0, \tilde{\mathcal{R}}_*$ are defined in (8.80) and (8.81).
The aim of this section is to eliminate $\mathfrak{B}_1$ from (8.95). In the next section we shall normalize the term $\mathfrak{B}_2$.
We conjugate $\mathcal{L}_3$ with a symplectic operator $\Phi_1\colon H^s_{S^\perp}(\mathbb{T}^{\nu+1}) \to H^s_{S^\perp}(\mathbb{T}^{\nu+1})$ of the form

$$\Phi_1 := \exp(\varepsilon A_1) = \mathrm{I}_{H_S^\perp} + \varepsilon A_1 + \varepsilon^2\frac{A_1^2}{2} + \varepsilon^3\hat{A}_1, \quad \hat{A}_1 := \sum_{k\ge 3}\frac{\varepsilon^{k-3}}{k!}A_1^k, \tag{8.96}$$

where $A_1(\varphi)h = \sum_{j,j'\in S^c}(A_1)_j^{j'}(\varphi)\,h_{j'}\,e^{\mathrm{i}jx}$ is a Hamiltonian vector field. The map Φ_1 is symplectic, because it is the time-1 flow of a Hamiltonian vector field. Therefore

$$\begin{aligned} &\mathcal{L}_3\Phi_1 - \Phi_1\Pi_S^\perp(\mathcal{D}_\omega + m_3\partial_{xxx})\Pi_S^\perp = \\ &= \Pi_S^\perp(\varepsilon\{\mathcal{D}_\omega A_1 + m_3[\partial_{xxx}, A_1] + \mathfrak{B}_1\} + \varepsilon^2\{\mathfrak{B}_1 A_1 + \mathfrak{B}_2 + \frac{1}{2}m_3[\partial_{xxx}, A_1^2] + \frac{1}{2}(\mathcal{D}_\omega A_1^2)\} + \tilde{d}_1\partial_x + R_3)\Pi_S^\perp \end{aligned} \tag{8.97}$$

where

$$R_3 := \tilde{d}_1\partial_x(\Phi_1 - \mathrm{I}) + \tilde{d}_0\Phi_1 + \tilde{\mathcal{R}}_*\Phi_1 + \varepsilon^2\mathfrak{B}_2(\Phi_1 - \mathrm{I}) + \varepsilon^3\{\mathcal{D}_\omega\hat{A}_1 + m_3[\partial_{xxx}, \hat{A}_1] + \frac{1}{2}\mathfrak{B}_1 A_1^2 + \varepsilon\mathfrak{B}_1\hat{A}_1\}. \tag{8.98}$$

Remark 8.11. R_3 has no longer the form (7.5). However $R_3 = O(\partial_x^0)$ because $A_1 = O(\partial_x^{-1})$ and therefore $\Phi_1 - \mathrm{I}_{H_S^\perp} = O(\partial_x^{-1})$. Moreover the matrix decay norm of R_3 is $o(\varepsilon^2)$.

In order to eliminate the order ε from (8.97), we choose

$$(A_1)_j^{j'}(l) = \begin{cases} -\dfrac{(\mathfrak{B}_1)_j^{j'}(l)}{i(\omega\cdot l + m_3(j'^3 - j^3))} & \text{if } \overline{\omega}\cdot l + j'^3 - j^3 \neq 0, \qquad j, j' \in S^c,\ l \in \mathbb{Z}^\nu \\ 0 & \text{otherwise} \end{cases} \tag{8.99}$$

This definition is well posed. Indeed, by (8.1) and (8.94)

$$(\mathfrak{B}_1)_j^{j'}(l) := \begin{cases} -2ij\,c_2\,(j-j')^2\sqrt{|j-j'|\xi_{j-j'}} - 6ij\,c_3\sqrt{|j-j'|\xi_{j-j'}} & \text{if } j - j' \in S, \quad l = \mathtt{1}(j-j') \\ 0 & \text{otherwise.} \end{cases} \tag{8.100}$$

In particular $(\mathfrak{B}_1)_j^{j'}(l) = 0$ unless $|l| \le 1$. Thus, for (l, j, j') such that $\overline{\omega}\cdot l + j'^3 - j^3 \neq 0$, the denominators in (8.99) satisfy

$$\begin{aligned} |\omega\cdot l + m_3(j'^3 - j^3)| &= |m_3(\overline{\omega}\cdot l + j'^3 - j^3) + (\omega - m_3\overline{\omega})\cdot l| \ge \\ &\ge |m_3|\,|\overline{\omega}\cdot l + j'^3 - j^3| - |\omega - m_3\overline{\omega}|\,|l| \ge 1/2, \ \ \forall |l| \le 1 \end{aligned} \tag{8.101}$$

for ε small enough, since $m_3 - 1$ and $\omega - \overline{\omega}$ are $O(\varepsilon^2)$. A_1 defined in (8.99) is a Hamiltonian vector field as $\mathfrak{B}_1$.

Lemma 8.12. *(Lemma* 8.16 *in [3]) If $j, j' \in S^c, j - j' \in S, l = \mathtt{1}(j - j')$, then*

$$\overline{\omega} \cdot l + j'^3 - j^3 = 3\, j\, j'\, (j' - j) \neq 0.$$

Corollary 8.13. *(Corollary* 8.17 *in [3]) Let $j, j' \in S^c$. If $\overline{\omega} \cdot l + j'^3 - j^3 = 0$ then $(\mathfrak{B}_1)_j^{j'} = 0$.*

By (8.99) and the previous corollary, the term of order ε in (8.97) is

$$\Pi_S^\perp(\mathcal{D}_\omega A_1 + m_3[\partial_{xxx}, A_1] + \mathfrak{B}_1)\Pi_S^\perp = 0. \tag{8.102}$$

We now prove that A_1 is a bounded transformation.

Lemma 8.14. *(Lemma* 8.18 *in [3])*

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

$$|(A_1)_j^{j'}(l)| \leq C(|j| + |j'|)^{-1}, \quad |(A_1)_j^{j'}(l)|^{lip} \leq \varepsilon^{-2}(|j| + |j'|)^{-1}. \tag{8.103}$$

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

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

Lemma 8.15. *(Lemma* 8.19 *in [3]) $|A_1\partial_x|_s^{Lip(\gamma)} + |\partial_x A_1|_s^{Lip(\gamma)} \leq C(s)$.*

It follows that the symplectic map Φ_1 in (8.96) is invertible for ε small, with inverse

$$\Phi_1^{-1} = \exp(-\varepsilon A_1) = \mathrm{I}_{H_S^\perp} + \varepsilon \check{A}_1, \ \check{A}_1 := \sum_{n \geq 1} \frac{\varepsilon^{n-1}}{n!}(-A_1)^n, \ |\check{A}_1\partial_x|_s^{Lip(\gamma)} + |\partial_x \check{A}_1|_s^{Lip(\gamma)} \leq C(s). \tag{8.104}$$

Since A_1 solves the homological equation (8.102), the ε-term in (8.95) is zero, and, with a straightforward calculation, the ε^2-term simplifies to $\mathfrak{B}_2 + \frac{1}{2}[\mathfrak{B}_1, A_1]$. We obtain the Hamiltonian operator

$$\mathcal{L}_4 := \Phi_1^{-1}\mathcal{L}_3\Phi_1 = \Pi_S^\perp(\mathcal{D}_\omega + m_3\partial_{xxx} + \tilde{d}_1\partial_x + \varepsilon^2\{\mathfrak{B}_2 + \frac{1}{2}[\mathfrak{B}_1, A_1]\} + \tilde{R}_4)\Pi_S^\perp, \tag{8.105}$$

$$\tilde{R}_4 := (\Phi_1^{-1} - \mathrm{I})\Pi_S^\perp[\varepsilon^2(\mathfrak{B}_2 + \frac{1}{2}[\mathfrak{B}_1, A_1]) + \tilde{d}_1\partial_x] + \Phi_1^{-1}\Pi_S^\perp R_3. \tag{8.106}$$

We split A_1 defined in (8.99), (8.100) into $A_1 = \overline{A}_1 + \tilde{A}_1$ where, for all $j, j' \in S^c, l \in \mathbb{Z}^\nu$,

$$(\overline{A}_1)_j^{j'}(l) := -\frac{2j\, c_2\, (j - j')^2 \sqrt{|j - j'|\xi_{j-j'}} + 6j\, c_3 \sqrt{|j - j'|\xi_{j-j'}}}{\overline{\omega} \cdot l + j'^3 - j^3} \tag{8.107}$$

if $\overline{\omega} \cdot l + j'^3 - j^3 \neq 0$, $j - j' \in S$, $l = \mathtt{1}(j - j')$, and $(\overline{A}_1)_j^{j'}(l) := 0$ otherwise.
By Lemma 8.12, for all $j, j' \in S^c, l \in \mathbb{Z}^\nu$,

$$(\overline{A}_1)_j^{j'}(l) = \begin{cases} -\dfrac{2}{3}\, c_2 \left(\dfrac{j - j'}{j'}\right) \sqrt{|j - j'|\xi_{j-j'}} - 2\, c_3\, \dfrac{1}{j'(j' - j)} \sqrt{|j - j'|\xi_{j-j'}} & \text{if } j - j' \in S, \\ 0 & \text{otherwise,} \end{cases} \tag{8.108}$$

namely

$$\overline{A}_1 h = -\frac{2}{3}c_2\Pi_S^\perp[\overline{v}_x\,(\partial_x^{-1}h)] + 2\, c_3\Pi_S^\perp[(\partial_x^{-1}\overline{v})(\partial_x^{-1}h)], \quad \forall h \in H_{S^\perp}^s(\mathbb{T}^{\nu+1}). \tag{8.109}$$

The difference is

$$(\tilde{A}_1)_j^{j'}(l) := -\frac{(2c_2\, j\, (j - j')^2 + 6\, c_3\, j)\sqrt{|j - j'|\xi_{j-j'}}\{(\omega - \overline{\omega}) \cdot l + (m_3 - 1)(j'^3 - j^3)\}}{(\omega \cdot l + m_3(j'^3 - j^3))(\overline{\omega} \cdot l + j'^3 - j^3)} \tag{8.110}$$

for $j, j' \in S^c, j - j' \in S, l = \mathtt{1}(j - j')$, and $(\tilde{A}_1)_j^{j'}(l) = 0$ otherwise. Then, by (8.105),

$$\mathcal{L}_4 = \Pi_S^\perp(\mathcal{D}_\omega + m_3\partial_{xxx} + \tilde{d}_1\,\partial_x + \varepsilon^2 T + R_4)\Pi_S^\perp, \tag{8.111}$$

where

$$T := \mathfrak{B}_2 + \frac{1}{2}[\mathfrak{B}_1, \overline{A}_1], \quad R_4 := \frac{\varepsilon^2}{2}[\mathfrak{B}_1, \tilde{A}_1] + \tilde{R}_4. \tag{8.112}$$

The operator T is Hamiltonian as $\mathfrak{B}_1, \mathfrak{B}_2, \overline{A}_1$, because the commutator of two Hamiltonian vector fields is Hamiltonian.

Lemma 8.16. *There is $\sigma = \sigma(\nu, \tau) > 0$ (possibly larger than in Lemma 8.10) such that*

$$|R_4|_s^{Lip(\gamma)} \leq_s \varepsilon^7\gamma^{-2} + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad |\partial_i R_4[\hat{\imath}]|_s \leq_s \varepsilon(\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}). \tag{8.113}$$

Proof. The proof follows the one of Lemma 8.20 in [3]. The only difference is the estimate on the coefficient $\tilde{d}_0$ (see (8.85)), that gives the term of size $\varepsilon^7\gamma^{-2}$ in (8.113), instead of $\varepsilon^5\gamma^{-1}$ in the inequality (8.95) in [3]. □

8.6 Linear Birkhoff Normal form (Step two)

The goal of this section is to normalize the term $\varepsilon^2 T$ from the operator $\mathcal{L}_4$ defined in (8.105). We cannot eliminate the terms $O(\varepsilon^2)$ at all, because some harmonics of $\varepsilon^2 T$, which correspond to null divisors, are not naught.
We conjugate the Hamiltonian operator $\mathcal{L}_4$ via a symplectic map

$$\Phi_2 := \exp(\varepsilon^2 A_2) = \mathrm{I}_{H_S^\perp} + \varepsilon^2 A_2 + \varepsilon^4\hat{A}_2, \quad \hat{A}_2 := \sum_{k\geq 2}\frac{\varepsilon^{2(k-2)}}{k!}A_2^k \tag{8.114}$$

where $A_2(\varphi) = \sum_{j,j'\in S^c}(A_2)_j^{j'}(\varphi)h_{j'}e^{\mathrm{i}jx}$ is a Hamiltonian vector field. We compute

$$\mathcal{L}_4\Phi_2 - \Phi_2\Pi_S^\perp(\mathcal{D}_\omega + m_3\partial_{xxx})\Pi_S^\perp = \Pi_S^\perp(\varepsilon^2\{\mathcal{D}_\omega A_2 + m_3[\partial_{xxx}, A_2] + T\} + \tilde{d}_1\partial_x + \tilde{R}_5)\Pi_S^\perp, \tag{8.115}$$

$$\tilde{R}_5 := \Pi_S^\perp\{\varepsilon^4(\mathcal{D}_\omega\hat{A}_2 + m_3[\partial_{xxx}, \hat{A}_2]) + (\tilde{d}_1\partial_x + \varepsilon^2 T)(\Phi_2 - \mathrm{I}) + R_4\Phi_2\}\Pi_S^\perp. \tag{8.116}$$

We define

$$(A_2)_j^{j'}(l) := \begin{cases} -\dfrac{T_j^{j'}(l)}{\mathrm{i}(\omega\cdot l + m_3(j'^3 - j^3))} & \text{if } \overline{\omega}\cdot l + j'^3 - j^3 \neq 0, \\ 0 & \text{otherwise.}\end{cases} \tag{8.117}$$

The definition is well posed. Indeed the matrix entries $T_j^{j'}(l) = 0$ for all $|j - j'| > 2C_S, l \in \mathbb{Z}^\nu$, where $C_S := \max\{|j| : j \in S\}$. Also $T_j^{j'}(l) = 0$ for all $j, j' \in S^c, |l| > 2$. Thus, arguing as in (8.101), if $\overline{\omega}\cdot l + j'^3 - j^3 \neq 0$, then $|\omega\cdot l + m_3(j'^3 - j^3)| \geq 1/2$. The operator A_2 is a Hamiltonian vector field because T is Hamiltonian.

Resonant terms

Now we compute the terms of $\varepsilon^2 T$ that cannot be removed by the Birkhoff map Φ_2.
By (8.109), (8.112) we get, for $h \in H^s_{S^\perp}$,

$$\begin{aligned}\mathfrak{B}_1\,\overline{A}_1[h] &= -\frac{4}{3}c_2^2\,\partial_x\Pi_S^\perp[\overline{v}_{xx}\,\Pi_S^\perp[\overline{v}_x\,(\partial_x^{-1}h)]] + 4c_2c_3\partial_x\Pi_S^\perp[\overline{v}_{xx}\Pi_S^\perp[(\partial_x^{-1}\overline{v})(\partial_x^{-1}h)]] \\ &\quad + 4c_2c_3\partial_x\Pi_S^\perp[\overline{v}\Pi_S^\perp[\overline{v}_x\,(\partial_x^{-1}h)]] - 12c_3^2\partial_x\Pi_S^\perp[\overline{v}\Pi_S^\perp[(\partial_x^{-1}\overline{v})(\partial_x^{-1}h)]] \\ \overline{A}_1\mathfrak{B}_1[h] &= -\frac{4}{3}c_2^2\,\Pi_S^\perp[\overline{v}_x\,\Pi_S^\perp[\overline{v}_{xx}h]] + 4c_2c_3\Pi_S^\perp[\overline{v}_x\,\Pi_S^\perp[\overline{v}\,h]] \\ &\quad + 4c_2c_3\Pi_S^\perp[(\partial_x^{-1}\overline{v})\Pi_S^\perp[\overline{v}_{xx}\,h]] - 12c_3^2\Pi_S^\perp[(\partial_x^{-1}\overline{v})\Pi_S^\perp[\overline{v}\,h]]\end{aligned}$$

whence, for all $j, j' \in S^c, l \in \mathbb{Z}^\nu$,

$$
\begin{aligned}
([\mathfrak{B}_1, \overline{A}_1])_j^{j'}(l) = \frac{4}{3} c_2^2\, \mathrm{i} & \sum_{\substack{j_1, j_2 \in S, j_1+j_2 = j-j', \\ j'+j_2 \in S^c, \mathtt{l}(j_1)+\mathtt{l}(j_2) = l}} \Big(\frac{j\, j_1^2\, j_2 - j_1\, j_2^2\, j'}{j'} \Big) \sqrt{|j_1\, j_2| \xi_{j_1} \xi_{j_2}} \\
& + 4 c_2 c_3\, \mathrm{i} \sum_{\substack{j_1, j_2 \in S, j_1+j_2 = j-j', \\ j'+j_2 \in S^c, \mathtt{l}(j_1)+\mathtt{l}(j_2) = l}} \Big(\frac{-j\, j_1^3 + j\, j_1\, j_2^2 - j_1^2\, j_2\, j' - j_2^3\, j'}{j' j_1 j_2} \Big) \sqrt{|j_1\, j_2| \xi_{j_1} \xi_{j_2}} \\
& + 12\, c_3^2\, \mathrm{i} \sum_{\substack{j_1, j_2 \in S, j_1+j_2 = j-j', \\ j'+j_2 \in S^c, \mathtt{l}(j_1)+\mathtt{l}(j_2) = l}} \Big(\frac{j j_1 - j' j_2}{j' j_1 j_2} \Big) \sqrt{|j_1\, j_2| \xi_{j_1} \xi_{j_2}}.
\end{aligned} \tag{8.118}
$$

If $([\mathfrak{B}_1, \overline{A}_1])_j^{j'}(l) \neq 0$ there are $j_1, j_2 \in S$ such that $j_1 + j_2 = j - j', j' + j_2 \in S^c, \mathtt{l}(j_1) + \mathtt{l}(j_2) = l$. Then

$$
\overline{\omega} \cdot l + j'^3 - j^3 = \overline{\omega} \cdot \mathtt{l}(j_1) + \overline{\omega} \cdot \mathtt{l}(j_2) + j'^3 - j^3 = j_1^3 + j_2^3 + j'^3 - j^3. \tag{8.119}
$$

Thus, if $\overline{\omega} \cdot l + j'^3 - j^3 = 0$, Lemma (3.3) implies that $(j_1 + j_2)(j_1 + j')(j_2 + j') = 0$. Now $j_1 + j', j_2 + j' \neq 0$ because $j_1, j_2 \in S, j' \in S^c$ and S is symmetric. Hence $j_1 + j_2 = 0$, which implies $j = j'$ and $l = 0$. In conclusion, if $\overline{\omega} \cdot l + j'^3 - j^3 = 0$, the only nonzero matrix entry $([\mathfrak{B}_1, \overline{A}_1])_j^{j'}(l)$ is

$$
\begin{aligned}
\frac{1}{2} ([\mathfrak{B}_1, \overline{A}_1])_j^j(0) = \frac{4}{3} c_2^2\, \mathrm{i} & \sum_{j_2 \in S, j_2 + j \in S^c} j_2^3\, |j_2| \xi_{j_2} + 8 c_2 c_3\, \mathrm{i} \sum_{j_2 \in S, j_2 + j \in S^c} j_2\, |j_2| \xi_{j_2} \\
& + 12\, c_3^2 \mathrm{i} \sum_{j_2 \in S, j_2 + j \in S^c} j_2^{-1}\, |j_2|\, \xi_{j_2}.
\end{aligned} \tag{8.120}
$$

Now consider $\mathfrak{B}_2$ defined in (8.94). We split $\mathfrak{B}_2 = B_1 + B_2 + B_3 + B_4 + B_5$, where

$$
\begin{aligned}
& B_1[h] := \alpha_{1,2}\, h_x, \quad B_2[h] := \alpha_{0,2}\, h, \quad B_3[h] := -(\alpha_{1,1})_x\, \beta_1\, h_x, \\
& B_4[h] := -(\alpha_{0,1})_x\, \beta_1, \quad B_5[h] := -\partial_x \overline{\mathcal{R}}_2[h].
\end{aligned} \tag{8.121}
$$

We denote by $(\alpha)_{j,l}$ the (j,l)-th Fourier coefficient of $\alpha(\varphi, x)$ as function of time and space. The Fourier representation of $B_i, i = 1, \dots, 4$ in (8.121) is

$$
\begin{aligned}
(B_1)_j^{j'}(l) &= \mathrm{i}\, j'\, (\alpha_{1,2})_{j-j', \mathtt{l}(j-j')}, \qquad (B_2)_j^{j'}(l) = (\alpha_{0,2})_{j-j', \mathtt{l}(j-j')} \\
(B_3)_j^{j'}(l) &= 4 c_1 c_2 \mathrm{i}\, j' (\overline{v}_{xxx} \overline{v})_{j-j', \mathtt{l}(j-j')} + \frac{4}{3} c_2^2 \mathrm{i} j' (\overline{v}_{xxx} (\partial_x^{-1} \overline{v}))_{j-j', \mathtt{l}(j-j')} \\
& \quad - 12 c_1 c_3 \mathrm{i}\, j'\, (\overline{v}\, \overline{v}_x)_{j-j', \mathtt{l}(j-j')} - 4 c_2 c_3 \mathrm{i}\, j' (\overline{v}_x (\partial_x^{-1} \overline{v}))_{j-j', \mathtt{l}(j-j')}, \\
(B_4)_j^{j'}(l) &= 4 c_1 c_2 (\overline{v}_{xxxx} \overline{v})_{j-j', \mathtt{l}(j-j')} + \frac{4}{3} c_2^2 (\overline{v}_{xxxx} (\partial_x^{-1} \overline{v}))_{j-j', \mathtt{l}(j-j')} \\
& \quad - 12 c_1 c_3 (\overline{v} \overline{v}_{xx})_{j-j', \mathtt{l}(j-j')} - 4 c_2 c_3 (\overline{v}_{xx} (\partial_x^{-1} \overline{v}))_{j-j', \mathtt{l}(j-j')}
\end{aligned}
$$

If $(B_k)_j^{j'}(l) \neq 0$, $k = 1, \dots, 4$ there are $j_1, j_2 \in S$ such that $j_1 + j_2 = j - j', l = \mathtt{l}(j_1) + \mathtt{l}(j_2)$ and (8.119) holds. Thus, if $\overline{\omega} \cdot l + j'^3 - j^3 = 0$, Lemma (3.3) implies that $(j_1 + j_2)(j_1 + j')(j_2 + j') = 0$, and, since $j' \in S^c$ and S is symmetric, the only possibility is $j_1 + j_2 = 0$. Hence $j = j'$, $l = 0$. In conclusion, if $\overline{\omega} \cdot l + j'^3 - j^3 = 0$, the only nonzero matrix element $(B_i)_j^{j'}(l), i = 1, \dots, 4$, by (8.7), is

$$
\begin{aligned}
& (B_1)_j^j(0) = \mathrm{i} j \sum_{k \in S} (-2 c_6 k^2 - 12 c_7 + \frac{4}{3} c_2^2 k^2 + 4 c_2 c_3)\, |k| \xi_k, \;\; (B_2)_j^j(0) = \sum_{k \in S} (-4 c_1 c_2\, k^4 - 12 c_1 c_3\, k^2)\, |k| \xi_k, \\
& (B_3)_j^j(0) = \mathrm{i} j \sum_{k \in S} (\frac{4}{3} c_2^2\, k^2 + 4 c_2 c_3) |k| \xi_k, \quad (B_4)_j^j(0) = \sum_{k \in S} (4 c_1 c_2\, k^4 + 12 c_1 c_3\, k^2)\, |k| \xi_k
\end{aligned} \tag{8.122}
$$

We note that $c(\xi)$ defined in (8.82) is equal to $-\mathrm{i}\sum_{i=1}^{4} j^{-1}\,(B_i)_j^j(0)$ (observe that the term $j^{-1}\,(B_i)_j^j(0)$ is independent of j) and we write

$$\begin{aligned} c(\xi) &= \sum_{k\in S^+} (-4c_6\,k^3 - 24c_7 k + \frac{16}{3}c_2^2\,k^3 + 16c_2c_3k)\,\xi_k \\ &= (\frac{16}{3}c_2^2 - 4c_6)v_3\cdot\xi + (16c_2c_3 - 24c_7)v_1\cdot\xi, \end{aligned} \tag{8.123}$$

where $v_3\cdot\xi = \sum_{j\in S^+} j^3\,\xi_j$ and $v_1\cdot\xi = \sum_{j\in S^+} j\,\xi_j$.
As before, the only possibility to get a zero at the denominator of (8.117) is $j_1+j_2=0$. Therefore

$$\begin{aligned} (B_5)_j^j(0) = \frac{4}{3}c_2^2\mathrm{i} &\sum_{j_2\in S, j_2+j\in S} j_2^3\,|j_2|\xi_{j_2} + 8c_2c_3\mathrm{i}\sum_{j_2\in S, j_2+j\in S} j_2\,|j_2|\xi_{j_2} \\ &+ 12c_3^2\mathrm{i}\sum_{j_2\in S, j_2+j\in S} j_2^{-1}\,|j_2|\xi_{j_2}. \end{aligned} \tag{8.124}$$

We note that for every odd function $f\colon S\to\mathbb{Z}$, by the simmetry of S, we have $\sum_{j_2\in S} f(j_2)\,\xi_{j_2}=0$. Thus, by (8.120) and (8.124), we get

$$(B_5)_j^j(0) + \frac{1}{2}([\mathfrak{B}_1,\overline{A}_1])_j^j(0) = \frac{4}{3}c_2^2\mathrm{i}\sum_{j_2\in S} j_2^3\,|j_2|\xi_{j_2} + 8c_2c_3\mathrm{i}\sum_{j_2\in S} j_2\,|j_2|\xi_{j_2} + 12c_3^2\mathrm{i}\sum_{j_2\in S} j_2^{-1}\,|j_2|\xi_{j_2} = 0.$$

Finally, we have

$$\mathcal{L}_5 := \Phi_2^{-1}\mathcal{L}_4\Phi_2 = \Pi_S^\perp(\mathcal{D}_\omega + m_3\partial_{xxx} + (\tilde d_1 + \varepsilon^2 c(\xi))\,\partial_x + R_5)\Pi_S^\perp, \tag{8.125}$$
$$R_5 := (\Phi_2^{-1} - \mathrm{I})\Pi_S^\perp(\tilde d_1 + \varepsilon^2 c(\xi))\partial_x + \Phi_2^{-1}\Pi_S^\perp\tilde R_5. \tag{8.126}$$

Lemma 8.17. *R_5 satisfies the same estimates* (8.113) *as R_4 (with a possibly larger σ).*

8.7 Descent method

The goal of this section is to transform $\mathcal{L}_5$ in (8.128) in order to make constant the coefficient in front of ∂_x. We conjugate $\mathcal{L}_5$ via a symplectic map of the form

$$\mathcal{S} := \exp(\Pi_S^\perp(w\partial_x^{-1}))\Pi_S^\perp = \Pi_S^\perp(\mathrm{I} + w\partial_x^{-1})\Pi_S^\perp + \hat{\mathcal{S}}, \quad \hat{\mathcal{S}} := \sum_{k\geq 2}\frac{1}{k!}[\Pi_S^\perp(w\partial_x^{-1})]^k\Pi_S^\perp, \tag{8.127}$$

where $w\colon\mathbb{T}^{\nu+1}\to\mathbb{R}$ is a function. Note that $\Pi_S^\perp(w\partial_x^{-1})\Pi_S^\perp$ is the Hamiltonian vector field generated by $-\frac{1}{2}\int_{\mathbb{T}} w(\partial_x^{-1}h)^2\,dx, h\in H_S^\perp$. We calculate

$$\mathcal{L}_5\mathcal{S} - \mathcal{S}\Pi_S^\perp(\mathcal{D}_\omega + m_3\partial_{xxx} + m_1\partial_x)\Pi_S^\perp = \Pi_S^\perp(3m_3w_x + \tilde d_1 + \varepsilon^2 c(\xi) - m_1)\partial_x\Pi_S^\perp + \tilde R_6, \tag{8.128}$$
$$\begin{aligned} \tilde R_6 :&= \Pi_S^\perp\{(3m_3w_{xx} + (\tilde d_1 + \varepsilon^2 c(\xi))\Pi_S^\perp w - m_1 w)\pi_0 + (\mathcal{D}_\omega w + m_3 w_{xxx} + (\tilde d_1 + \varepsilon^2 c(\xi))\Pi_S^\perp w_x)\partial_x^{-1} \\ &+ \mathcal{D}_\omega\hat{\mathcal{S}} + m_3[\partial_{xxx},\hat{\mathcal{S}}] + (\tilde d_1 + \varepsilon^2 c(\xi))\partial_x\hat{\mathcal{S}} - m_1\hat{\mathcal{S}}\partial_x + R_5\mathcal{S}\}\Pi_S^\perp \end{aligned}$$

where $\tilde R_6$ collects all the bounded terms. By (8.80), (8.82), we solve

$$3m_3w_x + \tilde d_1 + \varepsilon^2 c(\xi) - m_1 = 0$$

choosing $w := -(3m_3)^{-1}\partial_x^{-1}(\tilde d_1 + \varepsilon^2 c(\xi) - m_1)$. For ε sufficiently small, the operator $\mathcal{S}$ is invertible and, by (8.128),

$$\mathcal{L}_6 := \mathcal{S}^{-1}\mathcal{L}_4\mathcal{S} = \Pi_S^\perp(\mathcal{D}_\omega + m_3\partial_{xxx} + m_1\partial_x)\Pi_S^\perp + R_6, \quad R_6 := \mathcal{S}^{-1}\tilde R_6. \tag{8.129}$$

Since $\mathcal{S}$ is symplectic, $\mathcal{L}_6$ is Hamiltonian.

Lemma 8.18. *There is $\sigma := \sigma(\nu, \tau) > 0$ (possibly larger than in Lemma 8.16) such that*

$$|\mathcal{S}^{\pm 1} - \mathrm{I}|_s^{Lip(\gamma)} \leq_s \varepsilon^7 \gamma^{-2} + \varepsilon \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}, \quad |\partial_i \mathcal{S}^{\pm 1}[\hat{\imath}]|_s \leq_s \varepsilon (\|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma} \|\hat{\imath}\|_{s_0+\sigma}).$$

The remainder R_6 satisfies the same estimates of R_4 (with a possibly larger σ).

Proof. By (8.64), (8.83), (8.85), $\|w\|_s^{Lip(\gamma)} \leq_s \varepsilon^7 \gamma^{-2} + \varepsilon \|\mathfrak{I}_\delta\|_{s+\sigma}^{Lip(\gamma)}$, and the lemma follows by the definition of $\mathcal{S}$, see (8.127). Since $\hat{S} = O(\partial_x^{-2})$ the commutator $[\partial_{xxx}, \hat{S}] = O(\partial_x^0)$ and $|[\partial_{xxx}, \hat{S}]|_s^{Lip(\gamma)} \leq_s \|w\|_{s_0+3}^{Lip(\gamma)} \|w\|_{s+3}^{Lip(\gamma)}$. □

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

The coefficients m_3, m_1 of the operator $\mathcal{L}_6$ in (8.129) are constants, and the remainder R_6 is a bounded operator of order ∂_x^0 with small matrix decay norm. Then we can diagonalize $\mathcal{L}_6$ by applying the iterative KAM reducibility Theorem 4.2 in [2] along the sequence of scales

$$N_n := N_0^{\chi^n}, \quad n = 0, 1, 2, \ldots, \quad \chi := 3/2, \quad N_0 > 0. \tag{8.130}$$

In Section 9, the initial N_0 will (slightly) increase to infinity as $\varepsilon \to 0$, see (9.4). The required smallness condition (see (4.14) in [2]) is

$$N_0^{C_0} |R_6|_{s_0+\beta}^{Lip(\gamma)} \gamma^{-1} \leq 1, \tag{8.131}$$

where $\beta = 7\tau + 6$ (see (4.1) in [2]), τ is the diophantine exponent in (5.3) and (8.136), and the constant $C_0 := C_0(\tau, \nu) > 0$ is fixed in Theorem 4.2 in [2]. By Lemma 8.18, the remainder R_6 satisfies the bound (8.113), and using (7.7) we get

$$|R_6|_{s_0+\beta}^{Lip(\gamma)} \leq C \varepsilon^{7-2b} \gamma^{-1} = C \varepsilon^{3-2a}, \qquad |R_6|_{s_0+\beta}^{Lip(\gamma)} \gamma^{-1} \leq C \varepsilon^{1-3a}. \tag{8.132}$$

We use that μ in (7.7) is assumed to satisfy $\mu \geq \sigma + \beta$ where $\sigma := \sigma(\tau, \nu)$ is given in Lemma 8.18.

Theorem 8.19. ***(Reducibility)*** *Assume that $\omega \mapsto i_\delta(\omega)$ is a Lipschitz function defined on some subset $\Omega_0 \subseteq \Omega_\varepsilon$ (recall (5.2)), satisfying (7.7) with $\mu \geq \sigma + \beta$ where $\sigma := \sigma(\tau, \nu)$ is given in Lemma 8.18 and $\beta := 7\tau + 6$. Then there exists $\delta_0 \in (0, 1)$ such that, if*

$$N_0^{C_0} \varepsilon^{7-2b} \gamma^{-2} = N_0^{C_0} \varepsilon^{1-3a} \leq \delta_0, \quad \gamma := \varepsilon^{2+a}, \quad a \in (0, 1/6), \tag{8.133}$$

then

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

$$\mu_j^\infty(\omega) := \mu_j^\infty(\omega, i_\delta(\omega)) := -\mathrm{i}\tilde{m}_3(\omega)\, j^3 + \mathrm{i}\tilde{m}_1(\omega) j + r_j^\infty(\omega), \quad j \in S^c, \tag{8.134}$$

where $\tilde{m}_3, \tilde{m}_1$ coincide with the coefficients of $\mathcal{L}_6$ of (8.129) for all $\omega \in \Omega_0$. Furthermore, for all $j \in S^c$

$$|\tilde{m}_3 - 1|^{Lip(\gamma)} \leq C\varepsilon^2, \quad |\tilde{m}_1 - \varepsilon^2 c(\xi)|^{Lip(\gamma)} \leq C \varepsilon^{3-2a}, \tag{8.135}$$

for some $C > 0$. All the eigenvalues μ_j^∞ are purely imaginary. We define, for convenience, $\mu_0^\infty(\omega) := 0$.

(ii) ***(Conjugacy)***. *For all ω in the set*

$$\Omega_\infty^{2\gamma} := \Omega_\infty^{2\gamma}(i_\delta) := \left\{ \omega \in \Omega_0 : |\mathrm{i}\omega \cdot l + \mu_j^\infty(\omega) - \mu_k^\infty(\omega)| \geq \frac{2\gamma\, |j^3 - k^3|}{\langle l \rangle^\tau}, \ \forall l \in \mathbb{Z}^\nu, \ \forall j, k \in S^c \cup \{0\} \right\} \tag{8.136}$$

there is a real, bounded, invertible, linear operator $\Phi_\infty(\omega)\colon H^s_{S^\perp}(\mathbb{T}^{\nu+1}) \to H^s_{S^\perp}(\mathbb{T}^{\nu+1})$, with bounded inverse $\Phi_\infty^{-1}(\omega)$, that conjugates $\mathcal{L}_6$ in (8.129) *to constant coefficients, namely*

$$\begin{aligned} &\mathcal{L}_\infty(\omega) := \Phi_\infty^{-1}(\omega) \circ \mathcal{L}_5 \circ \Phi_\infty(\omega) = \omega \cdot \partial_\varphi + \mathcal{D}_\infty(\omega), \\ &\mathcal{D}_\infty(\omega) := diag_{j\in S^c}\{\mu_j^\infty(\omega)\}. \end{aligned} \tag{8.137}$$

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

$$|\Phi_\infty^{\pm 1} - \mathrm{I}|^{Lip(\gamma)}_{s,\Omega^{2\gamma}_\infty} \leq_s \varepsilon^7\gamma^{-3} + \varepsilon\gamma^{-1}\|\mathfrak{I}_\delta\|^{Lip(\gamma)}_{s+\sigma}. \tag{8.138}$$

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

Remark 8.20. Theorem 4.2 in [2] also provides the Lipschitz dependence of the (approximate) eigenvalues μ_j^n with respect to the unknown $i_0(\varphi)$, which is used for the measure estimate in Lemma 9.3.

Observe that all the parameters $\omega \in \Omega^{2\gamma}_\infty$ satisfy also the first Melnikov condition, namely

$$|\mathrm{i}\omega \cdot l + \mu_j^\infty(\omega)| \geq 2\gamma |j|^3 \langle l\rangle^{-\tau}, \quad \forall l \in \mathbb{Z}^\nu,\ j \in S^c, \tag{8.139}$$

because, by definition, $\mu_0^\infty = 0$, and the diagonal operator $\mathcal{L}_\infty$ is invertible.
In the following theorem we verify the inversion assumption (6.29) for $\mathcal{L}_\omega$.

Theorem 8.21. *Assume the hypotesis of Theorem 8.19 and* (8.133)*. Then there exists $\sigma_1 := \sigma_1(\tau,\nu) > 0$ such that, for all $\omega \in \Omega^{2\gamma}_\infty(i_\delta)$ (see* (8.136)*), for any function $g \in H^{s+\sigma_1}_{S^\perp}(\mathbb{T}^{\nu+1})$ the equation $\mathcal{L}_\omega h = g$ has a solution $h = \mathcal{L}_\omega^{-1} g \in H^s_{S^\perp}(\mathbb{T}^{\nu+1})$, satisfying*

$$\begin{aligned} \|\mathcal{L}_\omega^{-1} g\|^{Lip(\gamma)}_s &\leq_s \gamma^{-1}(\|g\|^{Lip(\gamma)}_{s+\sigma_1} + \varepsilon\gamma^{-1}\|\mathfrak{I}_\delta\|^{Lip(\gamma)}_{s+\sigma_1}\|g\|^{Lip(\gamma)}_{s_0}) \\ &\leq_s \gamma^{-1}(\|g\|^{Lip(\gamma)}_{s+\sigma_1} + \varepsilon\gamma^{-1}\{\|\mathfrak{I}_0\|^{Lip(\gamma)}_{s+\sigma_1+\sigma} + \gamma^{-1}\|\mathfrak{I}_0\|^{Lip(\gamma)}_{s_0+\sigma}\|Z\|^{Lip(\gamma)}_{s+\sigma_1+\sigma}\}\|g\|^{Lip(\gamma)}_{s_0}). \end{aligned} \tag{8.140}$$

9 The Nash-Moser nonlinear iteration

In this section we prove Theorem 5.2. It will be a consequence of the Nash-Moser theorem 9.1.
Consider the finite-dimensional subspaces

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

where $N_n := N_0^{\chi^n}$ are introduced in (8.130), and Π_n are the projectors (which, with a small abuse of notation, we denote with the same symbol)

$$\begin{aligned} &\Pi_n\Theta(\varphi) := \sum_{|l|<N_n} \Theta_l\, e^{\mathrm{i}l\cdot\varphi},\ \Pi_n y(\varphi) := \sum_{|l|<N_n} y_l\, e^{\mathrm{i}l\cdot\varphi},\ \text{where } \Theta(\varphi) = \sum_{l\in\mathbb{Z}^\nu}\Theta_l\, e^{\mathrm{i}l\cdot\varphi},\ y(\varphi) = \sum_{l\in\mathbb{Z}^\nu} y_l\, e^{\mathrm{i}l\cdot\varphi}, \\ &\Pi_n z(\varphi,x) := \sum_{|(l,j)|<N_n} z_{lj}\, e^{\mathrm{i}(l\cdot\varphi+jx)},\ \text{where} \quad z(\varphi,x) = \sum_{l\in\mathbb{Z}^\nu, j\in S^c} z_{lj}\, e^{\mathrm{i}(l\cdot\varphi+jx)}. \end{aligned} \tag{9.1}$$

We define $\Pi_n^\perp = \mathrm{I} - \Pi_n$. The classical smoothing properties hold, namely, for all $\alpha, s \geq 0$,

$$\|\Pi_n\mathfrak{I}\|^{Lip(\gamma)}_{s+\alpha} \leq N_n^\alpha\|\mathfrak{I}_\delta\|^{Lip(\gamma)}_s, \quad \forall\mathfrak{I}(\omega)\in H^s, \quad \|\Pi_n^\perp\mathfrak{I}\|^{Lip(\gamma)}_s \leq N_n^{-\alpha}\|\mathfrak{I}\|^{Lip(\gamma)}_{s+\alpha}, \quad \forall\mathfrak{I}(\omega)\in H^{s+\alpha}. \tag{9.2}$$

We define the following constants

$$\begin{aligned} &\mu_1 := 3\mu + 9, &&\alpha := 3\mu_1 + 1, &&\alpha_1 := (\alpha - 3\mu)/2, \\ &k := 3(\mu_1 + \rho^{-1}) + 1, &&\beta_1 := 6\mu_1 + 3\rho^{-1} + 3, &&0 < \rho < \frac{1-3a}{C_1(1+a)}. \end{aligned} \tag{9.3}$$

where $\mu := \mu(\tau,\nu) > 0$ is the "loss of regularity" given by the Theorem 6.37 and C_1 is fixed below. We note that the constants in (9.3) are the same of the ones defined in [3], but with a different (larger) μ.

Theorem 9.1. ***(Nash-Moser)*** *Assume that $f \in C^q$ with $q > S := s_0 + \beta_1 + \mu + 3$. 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 8.19), $\delta_0 := \delta_0(\tau, \nu) > 0$ such that, if*

$$N_0^{C_1} \varepsilon^{b_*+1} \gamma^{-2} < \delta_0, \quad \gamma := \varepsilon^{2+a} = \varepsilon^{2b}, \quad N_0 := (\varepsilon\gamma^{-1})^{\rho}, \quad b_* = 6 - 2b, \tag{9.4}$$

then, for all $n \geq 0$:

$(\mathcal{P}1)_n$ *there exists a function $(\mathfrak{I}_n, \zeta_n) \colon \mathcal{G}_n \subseteq \Omega_\varepsilon \to E_{n-1} \times \mathbb{R}^\nu, \omega \mapsto (\mathfrak{I}_n(\omega), \zeta_n(\omega)), (\mathfrak{I}_0, \zeta_0) := 0, E_{-1} := \{0\}$, satisfying $|\zeta_n|^{Lip(\gamma)} \leq C\|\mathcal{F}(U_n)\|_{s_0}^{Lip(\gamma)}$,*

$$\|\mathfrak{I}_n\|_{s_0+\mu}^{Lip(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1}, \quad \|\mathcal{F}(U_n)\|_{s_0+\mu+3}^{Lip(\gamma)} \leq C_* \varepsilon^{b_*}, \tag{9.5}$$

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

$$\begin{aligned}
&\mathcal{G}_0 := \{\omega \in \Omega_\varepsilon : |\omega \cdot l| \geq 2\gamma\langle l\rangle^{-\tau}, \ \forall l \in \mathbb{Z}^\nu \setminus \{0\}\}, \\
&\mathcal{G}_{n+1} := \left\{\omega \in \mathcal{G}_n : |\mathrm{i}\omega \cdot l + \mu_j^\infty(i_n) - \mu_k^\infty(i_n)| \geq \frac{2\,\gamma_n\,|j^3 - k^3|}{\langle l\rangle^\tau}, \ \forall j, k \in S^c \cup \{0\}, l \in \mathbb{Z}^\nu\right\},
\end{aligned} \tag{9.6}$$

where $\gamma_n := \gamma(1 + 2^{-n})$ and $\mu_j^\infty(\omega) := \mu_j^\infty(\omega, i_n(\omega))$ are defined in (8.134) *(and $\mu_0^\infty(\omega) = 0$). The differences $\hat{\mathfrak{I}}_n := \mathfrak{I}_n - \mathfrak{I}_{n-1}$ (where we set $\hat{\mathfrak{I}}_0 := 0$) is defined on $\mathcal{G}_n$, and satisfy*

$$\|\hat{\mathfrak{I}}_1\|_{s_0+\mu}^{Lip(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1}, \quad \|\hat{\mathfrak{I}}_n\|_{s_0+\mu}^{Lip(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1} N_{n-1}^{-\alpha}, \quad \forall n > 1. \tag{9.7}$$

$(\mathcal{P}2)_n$ $\|\mathcal{F}(U_n)\|_{s_0}^{Lip(\gamma)} \leq C_* \varepsilon^{b_*} N_{n-1}^{-\alpha}$ *where we set $N_{-1} := 1$.*

$(\mathcal{P}3)_n$ *(High Norms).* $\|\mathfrak{I}_n\|_{s_0+\beta_1}^{Lip(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1} N_{n-1}^k$ *and* $\|\mathcal{F}(U_n)\|_{s_0+\beta_1}^{Lip(\gamma)} \leq C_* \varepsilon^{b*} N_{n-1}^k$.

$(\mathcal{P}4)_n$ *(Measure). The measure of the "Cantor-like" sets $\mathcal{G}_n$ satisfies*

$$|\Omega_\varepsilon \setminus \mathcal{G}_0| \leq C_* \varepsilon^{2(\nu-1)} \gamma, \quad |\mathcal{G}_n \setminus \mathcal{G}_{n+1}| \leq C_* \varepsilon^{2(\nu-1)} \gamma N_{n-1}^{-1}. \tag{9.8}$$

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

Proof.

- *Proof of* $(\mathcal{P}_1)_0, (\mathcal{P}_2)_0, (\mathcal{P}_3)_0$. Recalling (5.7), we have, by the second estimate in (5.17),

$$\|\mathcal{F}(U_0)\|_s = \|\mathcal{F}((\varphi, 0, 0), 0)\|_s = \|X_P(i_0)\|_s \leq_s \varepsilon^{6-2b}.$$

Hence the smallness conditions in $(\mathcal{P}_1)_0, (\mathcal{P}_2)_0, (\mathcal{P}_3)_0$ hold taking $C_* := C_*(s_0 + \beta_1)$ large enough.

- *Assume that $(\mathcal{P}_1)_n, (\mathcal{P}_2)_n, (\mathcal{P}_3)_n$ hold for some $n \geq 0$, and prove $(\mathcal{P}_1)_{n+1}, (\mathcal{P}_2)_{n+1}, (\mathcal{P}_3)_{n+1}$.* By (9.3) and (9.4)

$$N_0^{C_1} \varepsilon^{b_*+1} \gamma^{-2} = N_0^{C_1} \varepsilon^{1-3a} = \varepsilon^{1-3a-\rho\, C_1(1+a)} < \delta_0$$

for ε small enough. If we take $C_1 \geq C_0$ then (8.133) holds. Moreover (9.5) imply (6.6), and so (7.7), and Theorem 8.21 applies. Hence the operator $\mathcal{L}_\omega := \mathcal{L}_\omega(\omega, i_n(\omega))$ defined in (7.33) is invertible for all $\omega \in \mathcal{G}_{n+1}$ and the last estimate in (8.140) holds. This means that the assumption (6.29) of Theorem 6.11 is verified with $\Omega_\infty = \mathcal{G}_{n+1}$. By Theorem 6.11 there exists an approximate inverse $\mathbf{T}_n(\omega) := \mathbf{T}_0(\omega, i_n(\omega))$ of the linearized operator $L_n(\omega) := d_{i,\zeta}\mathcal{F}(\omega, i_n(\omega))$, satisfying (6.38). By (9.4), (9.5)

$$\|\mathbf{T}_n g\|_s \leq_s \gamma^{-1}(\|g\|_{s+\mu} + \varepsilon\gamma^{-1}\{\|\mathfrak{I}_n\|_{s+\mu} + \gamma^{-1}\|\mathfrak{I}_n\|_{s_0+\mu}\|\mathcal{F}(U_n)\|_{s+\mu}\}\|g\|_{s_0+\mu}) \tag{9.9}$$

$$\|\mathbf{T}_n g\|_{s_0} \leq_{s_0} \gamma^{-1}\|g\|_{s_0+\mu} \tag{9.10}$$

and, by (6.39), using also (9.4), (9.5), (9.2),

$$\begin{aligned}
\|(L_n \circ \mathbf{T}_n - \mathrm{I})g\|_s \leq_s & \gamma^{-1}(\|\mathcal{F}(U_n)\|_{s_0+\mu}\|g\|_{s+\mu} + \|\mathcal{F}(U_n)\|_{s+\mu}\|g\|_{s_0+\mu} \\
& + \varepsilon\gamma^{-1}\|\mathfrak{I}_n\|_{s+\mu}\|\mathcal{F}(U_n)\|_{s_0+\mu}\|g\|_{s_0+\mu}) \qquad (9.11)\\
\|(L_n \circ \mathbf{T}_n - \mathrm{I})g\|_{s_0} \leq_{s_0} & \gamma^{-1}\|\mathcal{F}(U_n)\|_{s_0+\mu}\|g\|_{s_0+\mu} \\
\leq_{s_0} & \gamma^{-1}(\|\Pi_n\mathcal{F}(U_n)\|_{s_0+\mu} + \|\Pi_n^{\perp}\mathcal{F}(U_n)\|_{s_0+\mu})\|g\|_{s_0+\mu} \\
\leq_{s_0} & N_n^{\mu}\gamma^{-1}(\|\mathcal{F}(U_n)\|_{s_0} + N_n^{-\beta_1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1})\|g\|_{s_0+\mu}. \qquad (9.12)
\end{aligned}$$

The index β_1 in (9.3) is an ultraviolet cut, and it has to be define in order to obtain the convergence of the iteration scheme.
Now, for all $\omega \in \mathcal{G}_{n+1}$, we can define, for $n \geq 0$,

$$U_{n+1} := U_n + H_{n+1}, \quad H_{n+1} := (\hat{\mathfrak{I}}_{n+1}, \hat{\zeta}_{n+1}) := -\tilde{\Pi}_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n) \in E_n \times \mathbb{R}^{\nu}, \tag{9.13}$$

where $\tilde{\Pi}_n(\mathfrak{I},\zeta) := (\Pi_n\mathfrak{I},\zeta)$ with Π_n defined in (9.1). Since $L_n := d_{i,\zeta}\mathcal{F}(i_n)$, we write

$$\mathcal{F}(U_{n+1}) = \mathcal{F}(U_n) + L_nH_{n+1} + Q_n,$$

where

$$Q_n := Q(U_n, H_{n+1}), \quad Q(U_n, H) := \mathcal{F}(U_n + H) - \mathcal{F}(U_n) - L_nH, \quad H \in E_n \times \mathbb{R}^{\nu}. \tag{9.14}$$

Then, by the definition of H_{n+1} in (9.13), using $[L_n, \Pi_n]$ and writing $\tilde{\Pi}_n^{\perp}(\mathfrak{I},\zeta) := (\Pi_n^{\perp}\mathfrak{I}, 0)$ we have

$$\begin{aligned}
\mathcal{F}(U_{n+1}) &= \mathcal{F}(U_n) - L_n\tilde{\Pi}_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + Q_n = \mathcal{F}(U_n) - L_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + L_n\tilde{\Pi}_n^{\perp}\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + Q_n \\
&= \mathcal{F}(U_n) - \Pi_nL_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + (L_n\tilde{\Pi}_n^{\perp} - \Pi_n^{\perp}L_n)\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + Q_n \\
&= \Pi_n^{\perp}\mathcal{F}(U_n) + R_n + Q_n + Q_n'
\end{aligned} \tag{9.15}$$

where

$$R_n := (L_n\tilde{\Pi}_n^{\perp} - \Pi_n^{\perp}L_n)\mathbf{T}_n\Pi_n\mathcal{F}(U_n), \quad Q_n' := -\Pi_n(L_n\mathbf{T}_n - \mathrm{I})\Pi_n\mathcal{F}(U_n). \tag{9.16}$$

Lemma 9.2. *(Lemma 9.2 in [3]) Define*

$$w_n := \varepsilon\gamma^{-2}\|\mathcal{F}(U_n)\|_{s_0}, \qquad B_n := \varepsilon\gamma^{-1}\|\mathfrak{I}_n\|_{s_0+\beta_1} + \varepsilon\gamma^{-2}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}. \tag{9.17}$$

Then there exists $K := K(s_0, \beta_1) > 0$ such that, for all $n \geq 0$, setting $\mu_1 := 3\mu + 9$

$$w_{n+1} \leq KN_n^{\mu_1+\rho^{-1}-\beta_1}B_n + KN_n^{\mu_1}w_n^2, \qquad B_{n+1} \leq KN_n^{\mu_1+\rho^{-1}}B_n. \tag{9.18}$$

- *Proof of* $(\mathcal{P}_3)_{n+1}$. By (9.18) and $(\mathcal{P}_3)_n$

$$B_{n+1} \leq KN_n^{\mu_1+\rho^{-1}}B_n \leq 2C_*K\varepsilon^{b_*+1}\gamma^{-2}N_n^{\mu_1+\rho^{-1}}N_{n-1}^{k} \leq C_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{k}, \tag{9.19}$$

provided $2KN_n^{\mu_1+\rho^{-1}-k}N_{n-1}^{k} \leq 1, \forall n \geq 0$. Choosing k as in (9.3) and N_0 large enough, i.e. for ε small enough. By (9.17) and the bound (9.19) $(\mathcal{P}_3)_{n+1}$ holds.

- *Proof of* $(\mathcal{P}_2)_{n+1}$. Using (9.17), (9.18) and $(\mathcal{P}_2)_n, (\mathcal{P}_3)_n$, we get

$$w_{n+1} \leq KN_n^{\mu_1+\rho^{-1}-\beta_1}B_n + KN_n^{\mu_1}w_n^2 \leq KN_n^{\mu_1+\rho^{-1}-\beta_1}2C_*\varepsilon^{b_*+1}\gamma^{-2}N_{n-1}^{k} + KN_n^{\mu_1}(C_*\varepsilon^{b_*+1}\gamma^{-2}N_{n-1}^{-\alpha})^2$$

and $w_{n+1} \leq C_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{-\alpha}$ provided that

$$4KN_n^{\mu_1+\rho^{-1}-\beta_1+\alpha}N_{n-1}^{k} \leq 1, \quad 2KC_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{\mu_1+\alpha}N_{n-1}^{-2\alpha} \leq 1, \ \forall n \geq 0. \tag{9.20}$$

The inequalities in (9.20) hold by (9.4), taking α as in (9.3), $C_1 > \mu_1 + \alpha$ and δ_0 in (9.4) small enough. By (9.17), the inequality $w_{n+1} \leq C_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{-\alpha}$ implies $(\mathcal{P}_2)_{n+1}$.

- *Proof of* $(\mathcal{P}_1)_{n+1}$. The bound (9.7) for $\hat{\mathfrak{I}}_1$ follows by (9.13), (9.9) (for $s = s_0+\mu$) and $\|\mathcal{F}(U_0)\|_{s_0+2\mu} = \|\mathcal{F}((\varphi,0,0),0)\|_{s_0+2\mu} \leq_{s_0+2\mu} \varepsilon^{b_*}$. The bound (9.7) for $\hat{\mathfrak{I}}_{n+1}$ follows by (9.1), $(\mathcal{P}_2)_n$ and (9.3). It remains to prove that (9.5) holds at the step $n+1$. We have

$$\|\mathfrak{I}_{n+1}\|_{s_0+\mu} \leq \sum_{k=1}^{n+1} \|\hat{\mathfrak{I}}_k\|_{s_0+\mu} \leq C_*\varepsilon^{b_*}\gamma^{-1} \sum_{k\geq 1} N_{k-1}^{-\alpha_1} \leq C_*\varepsilon^{b_*}\gamma^{-1} \tag{9.21}$$

taking α_1 as in (9.3) and N_0 large enough, i.e. ε small enough. Moreover, using (9.1), $(\mathcal{P}_2)_{n+1}, (\mathcal{P}_3)_{n+1}$, (9.3) we get

$$\begin{aligned}\|\mathcal{F}(U_{n+1})\|_{s_0+\mu+3} &\leq N_n^{\mu+3}\|\mathcal{F}(U_{n+1})\|_{s_0} + N_n^{\mu+3-\beta_1}\|\mathcal{F}(U_{n+1})\|_{s_0+\beta_1} \\ &\leq C_*\varepsilon^{b_*}N_n^{\mu+3-\alpha} + C_*\varepsilon^{b_*}N_n^{\mu+3-\beta_1+k} \leq C_*\varepsilon^{b_*},\end{aligned}$$

which is the second inequality in (9.5) at the step $n+1$. The bound $|\zeta_{n+1}|^{Lip(\gamma)} \leq C\|\mathcal{F}(U_{n+1})\|_{s_0}^{Lip(\gamma)}$ is a consequence of Lemma (6.1).

9.1 Measure estimates

In this section we prove $(\mathcal{P}_4)_n$ for all $n \geq 0$. Fixed $n \in \mathbb{N}$, we have

$$\mathcal{G}_n \setminus \mathcal{G}_{n+1} = \bigcup_{l\in\mathbb{Z}^\nu, j,k\in S^c\cup\{0\}} R_{ljk}(i_n) \tag{9.22}$$

where

$$R_{ljk}(i_n) := \{\omega \in \mathcal{G}_n : |\mathrm{i}\omega\cdot l + \mu_j^\infty(i_n) - \mu_k^\infty(i_n)| < 2\,\gamma_n\,|j^3-k^3|\langle l\rangle^{-\tau}\}. \tag{9.23}$$

Since, by (5.3), $R_{ljk}(i_n) = \varnothing$ for $j = k$, in the sequel we assume that $j \neq k$.

Lemma 9.3. *(Lemma 9.3 in [3]) For $n \geq 1, |l| \leq N_{n-1}$, one has the inclusion $R_{ljk}(i_n) \subseteq R_{ljk}(i_{n-1})$.*

By definition, $R_{ljk}(i_n) \subseteq \mathcal{G}_n$ (see (9.23)). By Lemma 9.3, 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, $R_{ljk}(i_n) \cap \mathcal{G}_n = \varnothing$ (see (9.6)). As a consequence, $R_{ljk}(i_n) = \varnothing$ for all $|l| \leq N_{n-1}$, and

$$\mathcal{G}_n \setminus \mathcal{G}_{n+1} \subseteq \bigcup_{\substack{j,k\in S^c\cup\{0\}\\ |l|>N_{n-1}}} R_{ljk}(i_n) \quad \forall n \geq 1. \tag{9.24}$$

Lemma 9.4. *Let $n \geq 0$. If $R_{ljk}(i_n) \neq \varnothing$, then $|l| \geq C_1|j^3-k^3| \geq \frac{C_1}{2}(j^2+k^2)$ for some constant $C_1 > 0$ (independent of l, j, k, n, i_n, ω).*

By Lemma 9.4 it is sufficient to study the measure of the resonant sets $R_{ljk}(i_n)$ defined in (9.23) for $(l, j, k) \neq (0, j, j)$. In particular we will prove the following Lemma.

Lemma 9.5. *For all $n \geq 0$ and for a generic choice of the tangential sites, the measure $|R_{ljk}(i_n)| \leq C\varepsilon^{2(\nu-1)}\gamma\langle l\rangle^{-\tau}$.*

By (9.23), we have to bound the measure of the sublevels of the function $\omega \mapsto \phi(\omega)$ defined by

$$\phi(\omega) := \mathrm{i}\omega\cdot l + \mu_j^\infty(\omega) - \mu_k^\infty(\omega) = \mathrm{i}\omega\cdot l - \mathrm{i}m_3(\omega)(j^3-k^3) + \mathrm{i}m_1(j-k) + (r_j^\infty - r_k^\infty)(\omega) \tag{9.25}$$

Note that ϕ also depends on l, j, k, i_n. We recall that

$$m_3 = 1 + \varepsilon^2 d(\xi) + \mathtt{r}_{m_3}(\omega), \qquad m_1 = \varepsilon^2 c(\xi) + \mathtt{r}_{m_1}(\omega) \tag{9.26}$$

where

$$|\mathtt{r}_{m_3}|^{Lip(\gamma)} \leq C\varepsilon^3 \qquad |\mathtt{r}_{m_1}|^{Lip(\gamma)} \leq C\varepsilon^{3-2a} \tag{9.27}$$

and $d(\xi)$, $c(\xi)$ are defined in (8.61) and (8.123) respectively.

It will be useful to consider $\phi(\omega)$ in (9.25) as a small perturbation of an affine function in ω. We write it as

$$\phi(\omega) := a_{jk} + b_{ljk} \cdot \omega + q_{jk}(\omega), \qquad l \in \mathbb{Z}^\nu,\ j,k \in S^c, \tag{9.28}$$

where, by (4.18), (8.61), (8.123),

$$a_{jk} := -\,\mathrm{i}\{(j^3-k^3)[1-d(\mathbb{M}^{-1}\overline{\omega})] + (j-k)c(\mathbb{M}^{-1}\overline{\omega})\}, \tag{9.29}$$

$$\begin{aligned} b_{ljk} :=& \mathrm{i}\{l-(j^3-k^3)[(24c_4-48c_1^2)\mathbb{M}^{-T}v_3+(4c_6-\frac{16}{3}c_2^2)\mathbb{M}^{-T}v_1], \\ &+(j-k)[(-4c_6+\frac{16}{3}c_2^2)\mathbb{M}^{-1}v_3-(24c_7-16c_2c_3)\mathbb{M}^{-1}v_1]\} \end{aligned} \tag{9.30}$$

$$q_{jk}(\omega) := -\,\mathrm{i}\,\mathtt{r}_{m_3}(\omega)\,(j^3-k^3) + \mathrm{i}\,\mathtt{r}_{m_1}(\omega)\,(j-k) + r_j^\infty(\omega) - r_k^\infty(\omega) \tag{9.31}$$

and by (8.83), (9.27), (9.31),

$$\begin{aligned} |q_{jk}(\omega)|^{sup} &\leq \varepsilon^3|j^3-k^3| + \varepsilon^{3-2a}|j-k| + \varepsilon^{3-2a}, \\ |q_{jk}(\omega)|^{lip} &\leq |\mathtt{r}_{m_3}(\omega)|^{lip}|j^3-k^3| + |\mathtt{r}_{m_1}(\omega)|^{lip}|j-k| + |r_j^\infty - r_k^\infty|^{lip} \\ &\leq \varepsilon^3\gamma^{-1}|j^3-k^3| + \varepsilon^{3-2a}\gamma^{-1}|j-k| + \varepsilon^{1-3a}. \end{aligned} \tag{9.32}$$

Remark 9.6. The idea of the proof of Lemma 9.5 is that *generically* (see Definition 1.2) a_{jk} has to be sufficiently far from zero or the modulus of the "derivative" b_{ljk} has to be big enough.

We shall use the following *non-degeneracy* assumptions

$$(\mathtt{H1}) \qquad d(\xi)-1 \neq 0 \quad \text{at} \quad \xi = \mathbb{M}^{-1}\overline{\omega}, \tag{9.33}$$

$$(\mathtt{H2})_{j,k} \quad \text{Fixed } j,k \in S^c,\ j \neq k, \quad \det(\mathbb{M}+B(j,k)) \neq 0, \tag{9.34}$$

where

$$\begin{aligned} B(j,k) :=& -(24c_4-48c_1^2+\frac{12c_6-16c_2^2}{3(j^2+k^2+jk)})D_S^3UD_S^3 \\ &+(\frac{16c_2^2}{3}-4c_6+\frac{(16c_2c_3-24c_7)}{j^2+k^2+jk})D_SUD_S^3. \end{aligned} \tag{9.35}$$

In the next lemmata we prove that if the coefficients $c_1,\dots,c_7$ are non-resonant and conditions (C1)-(C2) hold, then there exist a generic choice of the tangential sites for which Lemma 9.10 and Lemma 9.13 hold true.

Lemma 9.7. *Fix $\nu \in \mathbb{N}$. If the coefficients $c_1,\dots,c_7$ are non-resonant and*

$$(7-16\nu)\,c_2^2 \neq 6\,(1-2\nu)c_6 \tag{9.36}$$

then the polynomial $P(\overline{\jmath}_1,\dots,\overline{\jmath}_\nu) := d(\mathbb{M}^{-1}\overline{\omega})-1$ is not identically zero. As a consequence, the assumption (H1) *is verified for a generic choice of the tangential sites.*

Proof. Suppose that $d(\mathbb{M}^{-1}\overline{\omega}) = 1$, namely

$$P(\overline{\jmath}_1,\dots,\overline{\jmath}_\nu) := \{(24c_4-48\,c_1^2)v_3+(4c_6-\frac{16}{3}c_2^2)v_1\}\cdot\mathbb{M}^{-1}\overline{\omega}-1 = 0. \tag{9.37}$$

We evaluate the polynomial P at the point $(\jmath_1, \ldots, \jmath_\nu) = \lambda(1, \ldots, 1) = \lambda\vec{1}$, for some λ to be determined, and we claim that this is not a zero. This implies that the polinomial P in (9.37) cannot be identically zero. We have

$$P(\lambda\vec{1}) = \{\lambda^5\,(24c_4 - 48\,c_1^2) + \lambda^3\,(4c_6 - \frac{16}{3}c_2^2)\}\,(\vec{1}\cdot\mathbb{M}(\lambda\vec{1})^{-1}\vec{1}) - 1$$

and $\mathbb{M}(\lambda\vec{1}) = a(\lambda)\mathrm{I} + b(\lambda)\,U$, where

$$a(\lambda) := (24c_1^2 - 12c_4)\lambda^6 + (\frac{14}{3}c_2^2 - 4c_6)\lambda^4 + (12c_2c_3 - 12c_7)\,\lambda^2 - 6\,c_3^2, \tag{9.38}$$

$$b(\lambda) := (-48c_1^2 + 24c_4)\lambda^6 + (-\frac{32}{3}c_2^2 + 8c_6)\lambda^4 + (-16c_2c_3 + 24c_7)\,\lambda^2. \tag{9.39}$$

We note that $a(\lambda) \neq 0$, because the coefficients are non-resonant. Moreover, by assumption (9.36) $a(\lambda) + \nu b(\lambda) \neq 0$ and we have

$$(\mathbb{M}(\lambda\vec{1}))^{-1} = \frac{\mathrm{I}}{a(\lambda)} - \frac{b(\lambda)}{a(\lambda)\,(a(\lambda) + b(\lambda)\,\nu)}\,U$$

and, by $\vec{1}\cdot\vec{1} = \nu,\ \vec{1}\cdot U\vec{1} = \nu^2$, we get

$$\vec{1}\cdot\mathbb{M}(\lambda\vec{1})^{-1}\vec{1} = \frac{\nu}{a(\lambda) + b(\lambda)\nu}. \tag{9.40}$$

Then $P(\lambda\vec{1}) = 0$ is equivalent to $p(\lambda) = 0$, where

$$\begin{aligned} p(\lambda) :&= \lambda^6\{24c_1^2 - 12c_4\} + \lambda^4\{(\frac{14}{3} - \frac{16\,\nu}{3})c_2^2 - 4(1-\nu)c_6\} \\ &+ \lambda^2\{(12 - 16\nu)c_2c_3 - 12(1 - 2\nu)c_7\} - 6c_3^2. \end{aligned}$$

Suppose that $c_3 \neq 0$, then $p(\lambda)$ is not trivial. If $c_3 = 0$ and $2c_1^2 \neq c_4$ then we conclude the same, because the monomial of degree six is not naught. If $c_3 = 0, 2c_1^2 = c_4$ then the monomial of minimum degree, namely three, it is not zero if $c_7 \neq 0$, indeed $\nu \in \mathbb{N}$. Suppose now that $c_3 = c_7 = 0, 2c_1^2 = c_4$. Eventually, by assumption (9.36) the monomial of maximum degree, namely four, is not naught and we conclude. $\square$

Lemma 9.8. *Fix $\nu \in \mathbb{N}$. If $c_1, \ldots, c_7$ are non-resonant and*

$$\nu\,\frac{3c_6 - 4c_2^2}{9c_4 - 18c_1^2} \notin \{j^2 + k^2 + jk\ :\ j, k \in \mathbb{Z}\setminus\{0\},\ j \neq k\}, \tag{9.41}$$

then the polynomials $P_{jk}(\jmath_1, \ldots, \jmath_\nu) := \det(\mathbb{M} + B(j,k))$ are not identically zero, for all $j, k \in S^c$, $j \neq k$.

Proof. By (9.35) we have

$$\begin{aligned} \mathbb{M} + B(j,k) = (24c_1^2 - 12c_4)D_S^6 &+ (\frac{14}{3}c_2^2 - 4c_6)D_S^4 + (4c_6 - \frac{16}{3}c_2^2)D_S^3 U D_S(\mathrm{I} - \frac{1}{j^2 + k^2 + jk}D_S^2) \\ &+ 12(c_2c_3 - c_7)D_S^2 - 6c_3^2\mathrm{I} + (16c_2c_3 - 24c_7)D_S U D_S(\mathrm{I} - \frac{1}{j^2 + k^2 + jk}D_S^2). \end{aligned}$$

If $c_3 \neq 0$ then the lowest order monomial of $\det(\mathbb{M} + B(j,k))$ is not zero and the same holds if $c_3 = 0$ and $c_7 \neq 0$. If $c_3 = c_7 = 0$ then the monomial of maximal degree is

$$D_S^3\left((24c_1^2 - 12c_4)\mathrm{I} + \frac{12c_6 - 16c_2^2}{3(j^2 + k^2 + jk)}U\right)D_S^3$$

and this is invertible if (9.41) holds. $\square$

Remark 9.9. By Lemma 9.8, if (C2) holds, then the assumptions $(\mathtt{H2})_{j,k}$ are satisfied by a generic choice of the tangential sites when j,k vary in a finite set of integers.

The rest of the section is devoted to the proof of Lemma 9.5.

Lemma 9.10. *Assume* (H1)*. Then, for a generic choice of the tangential sites, there exists $C_0 > 0$ such that for all $j \neq k$, $j,k \in S^c$, with $j^2 + k^2 > C_0$ and $l \in \mathbb{Z}^\nu$, we have $|R_{ljk}| \leq C\varepsilon^{2(\nu-1)}\gamma\langle l\rangle^{-\tau}$.*

Proof. If $j^2 + k^2 > C_0$ for some constant C_0, then $|j-k|/|j^3-k^3| \leq 2C_0^{-1}$ and

$$|a_{jk}| \geq |j^3 - k^3| \left\{ |1 - d(\mathbb{M}^{-1}\overline{\omega})| - \frac{2}{C_0}\, |c(\mathbb{M}^{-1}\overline{\omega})| \right\}.$$

If $d(\mathbb{M}^{-1}\overline{\omega}) \neq 1$ then, by taking C_0 large enough, we get $|a_{jk}| \geq \delta_0|j^3-k^3|$, for some $\delta_0 > 0$. This implies that for $\delta := \delta_0/2$ we have $|b_{ljk}\cdot\omega| \geq \delta|j^3-k^3|$. Indeed, by (9.23), (9.32)

$$|b_{ljk}\cdot\omega| \geq |a_{jk}| - |\phi(\omega)| - |q_{jk}(\omega)| \geq (\delta_0 - 2\gamma_n - |q_{jk}(\omega)|^{sup})|j^3-k^3| \geq \frac{\delta_0}{2}|j^3-k^3|,$$

for ε small enough (recall that $\gamma_n = o(\varepsilon^2)$).
If $b := b_{ljk}$ we have $|b\cdot\omega| \leq 2|b||\overline{\omega}|$, because $|\omega| \leq 2|\overline{\omega}|$. Hence $|b| \geq \delta_1\,|j^3-k^3|$ where $\delta_1 := \delta/(2|\overline{\omega}|)$. Split $\omega = s\hat{b} + v$ where $\hat{b} := b/|b|$ and $v\cdot b = 0$. Let $\Psi(s) := \phi(s\hat{b}+v)$. For ε small enough, by (9.32), we get

$$\begin{aligned} |\Psi(s_1) - \Psi(s_2)| &\geq (|b| - |q_{jk}|^{lip})|s_1 - s_2| \geq \left(\delta_1 - \frac{|q_{jk}|^{lip}}{|j^3-k^3|}\right)|j^3-k^3|\,|s_1-s_2| \\ &\geq \frac{\delta_1}{2}|j^3-k^3|\,|s_1-s_2|. \end{aligned}$$

As a consequence, the set $\Delta_{ljk}(i_n) := \{s : s\hat{l} + v \in R_{ljk}(i_n)\}$ has Lebesgue measure

$$|\Delta_{ljk}(i_n)| \leq \frac{2}{\delta_1\,|j^3-k^3|}\,\frac{4\,\gamma_n\,|j^3-k^3|}{\langle l\rangle^\tau} \leq \frac{C\,\gamma}{\langle l\rangle^\tau}$$

for some $C > 0$. The Lemma follows by Fubini's theorem. $\square$

Lemma 9.11. *There exists $M > 0$ such that for all $j \neq k$, $j,k \in S^c$, with $j^2 + k^2 \leq C_0$ (see Lemma 9.10) and $|l| \geq M$, we have $|R_{ljk}| \leq C\varepsilon^{2(\nu-1)}\gamma\langle l\rangle^{-\tau}$.*

Proof. For $l \neq 0$, we decompose $\omega = s\hat{l} + v$, where $\hat{l} := l/|l|$, $s \in \mathbb{R}$, and $l\cdot v = 0$. Let $\psi(s) := \phi(s\hat{l}+v)$. We remark that $c(\xi)$ and $d(\xi)$ are affine functions of the unperturbed actions ξ, hence

$$\varepsilon^2|c(\xi)|^{lip}, \varepsilon^2|d(\xi)|^{lip} \leq K$$

for some constant K depending only on the tangential sites and on the real coefficients $c_1,\ldots,c_7$. Then

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

Then, if we take M large enough and ε small, we have

$$\begin{aligned} |\psi(s_1) - \psi(s_2)| &\geq |j^3-k^3|\left(\frac{|l|}{|j^3-k^3|} - K - \frac{2\,K}{|j^2+k^2+jk|} - \frac{\varepsilon^{3-2a}\gamma^{-1}}{|j^3-k^3|}\right)|s_1-s_2| \\ &\geq \frac{\delta}{4}|j^3-k^3|\,|s_1-s_2|, \end{aligned}$$

where δ is a positive constant. Indeed, C_0 and K are fixed and it is sufficient to choose $|l|$ such that

$$\inf_{j\neq k, j^2+k^2\leq C_0} \frac{|l|}{|j^3-k^3|} - K - \frac{2K}{C_0} \geq \delta > 0.$$

As a consequence, the set $\Delta_{ljk}(i_n) := \{s : s\hat{l} + v \in R_{ljk}(i_n)\}$ has Lebesgue measure

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

for some $C > 0$. The Lemma follows by Fubini's theorem. $\square$

It remains to investigate R_{ljk} for a finite set of indeces (l, j, k). We need the following Lemma.

Lemma 9.12. *Suppose that $l \in \mathbb{Z}^\nu$ and $j, k \in S^c$ are such that $\overline{\omega} \cdot l \neq j^3 - k^3$ and*

$$|l| \leq M, \qquad |j^2 + k^2| \leq C_0 \tag{9.42}$$

for some positive constants M and C_0. Then R_{ljk} is empty.

Proof. We have

$$\begin{aligned}|\omega\cdot l + m_3(k^3-j^3)| = |m_3(\overline{\omega}\cdot l + k^3 - j^3) + (\omega - m_3\overline{\omega})\cdot l| \geq |m_3||\overline{\omega}\cdot l + k^3 - j^3|\\ - |\omega - m_3\overline{\omega}||l| \geq 1 - |\omega-\overline{\omega}|M - |m_3-1||\overline{\omega}|M \geq 1/2\end{aligned}$$

for ε small enough, because $|\omega - \overline{\omega}|, |m_3 - 1| \leq C\varepsilon^2$. Thus, by (9.25) we have

$$\begin{aligned}|\phi(\omega)| &\geq 1/2 - \varepsilon^2|c(\xi)||j-k| - |m_1 - \varepsilon^2 c(\xi)||j-k| - |r_j - r_k|\\ &\geq 1/2 - \varepsilon^2 \sup_{\xi\in[1,2]^\nu}(|c(\xi)|) - 2C\,\varepsilon^{3-2a} \geq 1/4.\end{aligned}$$

$\square$

Lemma 9.13. *If $j^2 + k^2 \leq C_0$, $j, k \in S^c$, $|l| \leq M$ (see Lemma 9.10 and 9.11) and $(\mathtt{H2})_{j,k}$ hold, then, for a generic choice of the tangential sites, $|R_{ljk}| \leq C\varepsilon^{2(\nu-1)}\gamma\langle l\rangle^{-\tau}$.*

Proof. We can write (9.25) as an affine function respect to the parameter ξ as

$$\begin{aligned}\phi(\xi) &= \mathrm{i}\,\overline{\omega}\cdot l - \mathrm{i}(j^3-k^3) + \mathrm{i}\varepsilon^2\{\mathbb{M}\xi\cdot l - d(\xi)(j^3-k^3) + c(\xi)(j-k)\} + q_{jk}(\alpha(\xi)),\\ q_{jk}(\alpha(\xi)) &= -\mathrm{i}r_{m_3}(\alpha(\xi))(j^3-k^3) + \mathrm{i}r_{m_1}(\alpha(\xi))(j-k) + r_j^\infty(\alpha(\xi)) - r_k^\infty(\alpha(\xi)).\end{aligned} \tag{9.43}$$

By the relation (4.18), we can estimate the Lipschitz constant of $\phi(\omega)$ with the derivative respect to ξ of the expression (9.43).
By Lemma 9.12, we consider the case $\overline{\omega}\cdot l = j^3 - k^3$. Thus

$$\begin{aligned}\phi(\xi) =& \mathrm{i}\,\varepsilon^2[\mathbb{M}\xi\cdot l - d(\xi)\,\overline{\omega}\cdot l + c(\xi)(j-k)] + q_{jk}(\alpha(\xi))\\ =& \mathrm{i}\,\varepsilon^2[\mathbb{M}\xi\cdot l - d(\xi)\overline{\omega}\cdot l + \frac{c(\xi)}{j^2+k^2+jk}\overline{\omega}\cdot l] + q_{jk}(\alpha(\xi))\\ =& \mathrm{i}\,\varepsilon^2[\mathbb{M} + B(j,k)]\,l\cdot\xi + q_{jk}(\alpha(\xi)).\end{aligned} \tag{9.44}$$

where $B(j, k)$ is defined in (9.35). By assumption ($\mathtt{H2}$), if $l \neq 0$, then

$$\delta_{ljk} := (\mathbb{M} + B(j,k))l \neq 0. \tag{9.45}$$

Hence, by (9.32), (9.44) and (9.45), for ε small enough, there exist a constant $C > 0$ such that

$$|\phi|^{lip} \geq \delta_{ljk} - |q_{jk}|^{lip} \geq C|j^3 - k^3|.$$

Then we conclude as in Lemma 9.11. $\square$

We have that Lemmata 9.10, 9.11, 9.13 implies Lemma 9.5. By (9.22) and Lemma 9.5 we get

$$|\mathcal{G}_0 \setminus \mathcal{G}_1| \le \sum_{l\in\mathbb{Z}^\nu, |j|,|k|\le C|l|^{1/2}} |R_{ljk}(i_0)| \le \sum_{l\in\mathbb{Z}^\nu} \frac{C\,\varepsilon^{2(\nu-1)}\gamma}{\langle l\rangle^{\tau-1}} \le C'\varepsilon^{2(\nu-1)}\gamma.$$

For $n \ge 1$, by (9.24),

$$|\mathcal{G}_n \setminus \mathcal{G}_{n+1}| \le \sum_{\substack{|l|>N_{n-1},\\ |j|,|k|\le C|l|^{1/2}}} |R_{ljk}(i_n)| \le \sum_{|l|>N_{n-1}} \frac{C\,\varepsilon^{2(\nu-1)}\gamma}{\langle l\rangle^{\tau-1}} \le C'\varepsilon^{2(\nu-1)}\gamma\, N_{n-1}^{-1}$$

because $\tau \ge \nu+2$. The estimate $|\Omega_\varepsilon \setminus \mathcal{G}_0| \le C\,\varepsilon^{2(\nu-1)}\gamma$ is elementary.

□

Conclusion of the Proof of Theorem 5.2. Theorem 9.1 implies that the sequence $(\mathfrak{I}_n, \zeta_n)$ is well defined for $\omega \in \mathcal{G}_\infty := \cap_{n\ge0}\mathcal{G}_n$, and $\mathfrak{I}_n$ is a Cauchy sequence in $\|\cdot\|^{Lip(\gamma)}_{s_0+\mu,\mathcal{G}_\infty}$, see (9.7), and $|\zeta_n|^{Lip(\gamma)} \to 0$. Therefore $\mathfrak{I}_n$ converges to a limit $\mathfrak{I}_\infty$ in norm $\|\cdot\|^{Lip(\gamma)}_{s_0+\mu,\mathcal{G}_\infty}$ and, by $(\mathcal{P}2)_n$, for all $\omega \in \mathcal{G}_\infty, i_\infty(\varphi) := (\varphi,0,0) + \mathfrak{I}_\infty(\varphi)$, is a solution of

$$\mathcal{F}(i_\infty, 0) = 0 \qquad \text{with} \qquad \|\mathfrak{I}_\infty\|^{Lip(\gamma)}_{s_0+\mu,\mathcal{G}_\infty} \le C\,\varepsilon^{6-2b}\gamma^{-1}$$

by (9.5). Therefore $\varphi \mapsto i_\infty(\varphi)$ is an invariant torus for the Hamiltonian vector field X_{H_ε} (recall (4.19)). By (9.8),

$$|\Omega_\varepsilon \setminus \mathcal{G}_\infty| \le |\Omega_\varepsilon \setminus \mathcal{G}_0| + \sum_{n\ge0} |\mathcal{G}_n \setminus \mathcal{G}_{n+1}| \le 2\,C_*\varepsilon^{2(\nu-1)}\gamma + C_*\varepsilon^{2(\nu-1)}\gamma \sum_{n\ge1} N_{n-1}^{-1} \le C\varepsilon^{2(\nu-1)}\gamma.$$

The set Ω_ε in (5.2) has measure $|\Omega_\varepsilon| = O(\varepsilon^{2\nu})$. Hence $|\Omega_\varepsilon \setminus \mathcal{G}_\infty|/|\Omega_\varepsilon| \to 0$ as $\varepsilon \to 0$ because $\gamma = o(\varepsilon^2)$, and therefore the measure of $\mathcal{C}_\varepsilon := \mathcal{G}_\infty$ satisfies (5.12).

9.2 Linear stability

We show that the solution $i_\infty(\omega t)$ is linearly stable, in the sense that the norm of the solutions of the Hamiltonian system associated to (4.19) linearized on the quasi-periodic solution i_∞ does not increase in time.
By Section 6, in particular by the Remark 6.5, the system related to (4.19) is conjugated to the linear system

$$\begin{cases} \dot\psi = K_{20}(\omega t)\eta + K_{11}^T(\omega t)w \\ \dot\eta = 0 \\ \dot w = \partial_x K_{02}(\omega t)w + \partial_x K_{11}(\omega t)\eta. \end{cases} \tag{9.46}$$

Thus the actions $\eta(t)$ do not evolve in time and the third equation of (9.46) reduces to the forced PDE

$$\dot w = \partial_x K_{02}(\omega t)w + \partial_x K_{11}(\omega t)\eta(0). \tag{9.47}$$

In Section 8 we proved the reducibility of the linear system (9.47), ignoring the quasi-periodic function $\partial_x K_{11}(\omega t)\eta(0)$. More precisely, we conjugated it to the diagonal system

$$\dot v_j + \mu_j^\infty v_j = 0, \qquad j \in S^c, \quad \mu_j^\infty \in \mathrm{i}\mathbb{R}, \tag{9.48}$$

where

$$\mu_j^\infty := \mathrm{i}(-m_3 j^3 + m_1 j) + r_j^\infty \tag{9.49}$$

with $m_3 = 1 + O(\varepsilon^2), m_1 = O(\varepsilon^2), r_j^\infty = O(\varepsilon^{3-2a})$. The eigenvalues μ_j^∞ are the *Floquet exponents* of the linear, quasi-periodically depending on time, system (9.47). Then equation (9.47) is reduced to

$$\dot{v}_j + \mu_j^\infty v_j = f_j(\omega t), \qquad j \in S^c \tag{9.50}$$

for some quasi-periodic function f_j. The solutions of the scalar non-homogeneous equation (9.50) are

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

We note that $\tilde{v}_j$ is well defined, indeed the first Melnikov conditions hold at a solution. As a consequence, if v is a solution of the system (9.47), then there exist a constant $C > 0$ such that

$$\|v(t)\|_{H^s_x} \le C \|v(0)\|_{H^s_x}, \quad \forall t \in \mathbb{R},$$

hence its Sobolev norm does not increase in time.

References

[1] Baldi P., Periodic solutions of fully nonlinear autonomous equations of benjamin-Ono type, Ann. Inst. H. Poincar (C) Anal. Non Linaire 30 (2013), 33-77.

[2] Baldi P., Berti M., Montalto R., KAM for quasi-linear and fully nonlinear forced perturbations of Airy equation, Math Annalen 359, 471-536 (2014).

[3] Baldi P., Berti M., Montalto R., KAM for autonomous quasi-linear perturbations of KdV, Ann. I. H. Poincar-AN (2015), http:// dx.doi.org/10.1016/j.anihpc. 2015.07.003.

[4] Baldi P., Berti M., Montalto R., KAM for autonomous quasi-linear perturbations of mKdV, Bollettino dell'Unione Matematica Italiana, Volume 9, Issue 2, 143-188.

[5] Berti M., Biasco L., Procesi M., KAM theory for the Hamiltonian DNLW, Ann. Sci. Ec. Norm. Sup. (4), VOL. 46, fascicule 2 (2013), 301-373.

[6] Berti M., Biasco L., Procesi M., KAM theory for the reversible derivative wave equation, Arch. Rational Mech. Anal., 212, (2014), 905-955.

[7] Berti M., Bolle P., Quasi-periodic solutions with Sobolev regularity of NLS on $\mathbb{T}^d$ with a multiplicative potential, Eur. Jour. Math. 15 (2013), 229-286.

[8] Berti M., Bolle P., A Nash-Moser approach to KAM theory, Fields Institute Communications, volume 75 "Hamiltonian PDEs and Applications", 255-284.

[9] Berti M., Bolle P., Quasi-periodic solutions for autonomous NLW on $\mathbb{T}^d$ with a multiplicative potential, in preparation.

[10] Bourgain J., Gibbs measures and quasi-periodic solutions for nonlinear Hamiltonian partial different equations, 23-43, Gelfand Math. Sem., Birkhäuser Boston, Boston, MA, 1996.

[11] Craig W., Wayne C., Newton's method and periodic solutions of nonlinear wave equation, Comm. Pure Appl. Math. 46, 1409-1498 (1993).

[12] Eliasson L.H., Grbert B., Kuksin S., KAM for the nonlinear Beam equation, preprint.

[13] Eliasson L.H., Kuksin S., KAM for non-linear Schrödinger equation, Annals of Math., 172 (2010), 371-435.

[14] Feola R., Procesi M., Quasi-periodic solutions for fully nonlinear forced reversible Schrödinger equations, Journal of Differential Equations, Volume 259, Issue 7, 3389-3447, (2015).

[15] Feola R., Quasi-periodic solutions for fully nonlinear NLS, PhD Thesis of Roberto Feola 2015.

[16] Geng J., Xu X., You J., An infinite dimensional KAM theorem and its application to the two dimensional cubic Schrödinger equation, Adv. Math. 226 (2011) 5361-5402.

[17] Huang G., Kuksin K., KdV equation under periodic boundary conditions and its perturbations, Nonlinearity, Volume 27 (9), (2014).

[18] Iooss G., Plotnikov P.I., Multimodal standing gravity waves: a completely resonant system, J. Math. Fluid Mech. 7 (2005), 110-126.

[19] Iooss G., Plotnikov P.I., Existence of multimodal standing gravity waves, J.Math.Fluid Mech.7 (2005), 349-364.

[20] 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.

[21] Kappeler T., Pöschel J., KAM and KdV, Springer 2003.

[22] Kuksin S., Hamiltonian perturbations of infinite-dimensional linear systems with imaginary spectrum, Funktsional. Anal. i Prilozhen. 21 no. 3, 22-37, 95 (1987).

[23] Kuksin S., A KAM theorem for equations of the Korteweg-de Vries type, Rev. Math. Phys., 10, 3, (1998), 1-64.

[24] Kuksin S., Analysis of Hamiltonian PDEs, Oxford Lecture Series in Mathematics and its Applications, 19. Oxford University Press (2000).

[25] Kuksin S,. Pöschel J., Invarian Cantor Manifolds of Quasi-Periodic Oscillations for a Nonlinear Schrödinger Equation, Ann. of Math. 143, (1996), 149-179.

[26] Liu J., Yuan X., A KAM theorem for Hamiltonian Partial Differential Equations with Unbounded Perturbations, Comm. Math. Phys., 307 (3) (2011), 629-673.

[27] Montalto R., Quasi-periodic solutions of forced Kirchoff equation, preprint.

[28] Plotnikov P.I., Toland, J.F., Nash-Moser theory for standing water waves, Arch. Ration. Mech. Anal. 159, (2001) 1-83.

[29] Pöschel J., A KAM-Theorem for some nonlinear partial differential equations, Ann. Sc. Norm. Sup. Pisa, Cl. Sci. (4) 23, 119-148 (1996).

[30] Procesi C., Procesi M., A normal form for the Schrödinger equation with analytic non-linearities, Communications in Mathematical Physics 312 (2012), no. 2, 501-557.

[31] Procesi C., Procesi M., A KAM algorithm for the non-linear Schrödinger equation, Advances in Math. 272 (2015), 399-470.

[32] Wang W. M., Supercritical nonlinear Schrödinger equations I: quasi-periodic solutions, Duke Math. J. 165, no. 6, (2016), 1129-1192.

[33] Wayne E., Periodic and quasi-periodic solutions of nonlineare wave equations via KAM theory, Comm. Math. Phys. 127, 479-528 (1990).

[34] Zhang J., Gao M., Yuan X., KAM tori for reversible partial differential equations, Nonlinearity 24 (2011), 1189-1228.

[35] Zehnder E., Generalized implicit function theorems with applications to some small divisors problems $I - II$, Comm. Pure Appl. Math., 28, 91-140, (1975); 29, 49-113, (1976).