

# Reducible quasi-periodic solutions for the Non Linear Schrödinger equation

M. Procesi · C. Procesi

Received: date / Accepted: date

**Abstract** The present paper is devoted to the construction of small reducible quasi-periodic solutions for the completely resonant NLS equations on a  $d$ -dimensional torus  $\mathbb{T}^d$ . The main point is to prove that the normal form is reducible, block diagonal and satisfies the second Melnikov condition block wise. From this we deduce the result by a KAM algorithm.

**Keywords** KAM theory for PDEs · Quasi-periodic solutions.

**Mathematics Subject Classification (2000)** 37K55 · 35Q55

## 1 Introduction

The present paper is devoted to the construction of small reducible (see Definition 9) quasi-periodic solutions for the completely resonant NLS equations on a  $d$ -dimensional torus  $\mathbb{T}^d$ :

$$iu_t - \Delta u = \mathbf{k}|u|^{2q}u + \partial_{\bar{u}}G(|u|^2). \quad (1)$$

Here  $u := u(t, \varphi)$ ,  $\varphi \in \mathbb{T}^d$ ,  $\Delta$  is the Laplace operator and  $G(a)$  is a real analytic function whose Taylor series starts from degree  $q + 2$ ,  $q \geq 1$ . Finally  $\mathbf{k}$  is a coupling constant which can be normalised to  $\pm 1$ . Since in our results the sign is irrelevant we will set it equal to one. Note that we have restricted our attention to  $\varphi$ -independent non-linearities  $G$ . The corresponding symmetries imply the presence of  $d + 1$  constants of motion given by the  $L^2$  norm and the momentum (translation invariance).

---

The first author is supported by ERC project HamPDEs, under FP7

M. Procesi

Università di Roma, La Sapienza, Tel.: +390649913250, E-mail: mprocesi@mat.uniroma1.it

C. Procesi

Università di Roma, La Sapienza E-mail: procesi@mat.uniroma1.it

A relevant feature is that the NLS equation is completely resonant near  $u = 0$  (i.e. all the linear solutions are periodic), hence we look for our quasi-periodic solutions close to some *periodic* solutions

$$\sum_{i=1}^n \sqrt{\xi_i} e^{i|j_i|^2 t} e^{ij_i \cdot \varphi} \quad (2)$$

of the linear equation involving  $n$  frequencies  $S = \{j_1, \dots, j_n\} \subset \mathbb{Z}^d$ . It is well known, see for instance [8], that due to the presence of resonances, there exist choices of the frequencies  $S$  for which the solutions of the non-linear system differ drastically from the ones of the linear system. In order to avoid such phenomena we restrict to *generic* choices of  $S$ , where generic means that this list of vectors does not satisfy some explicit, although quite complicated, polynomial equations which express the *resonances* to be avoided.

Our results are obtained by exploiting the Hamiltonian structure of equation (1), and by studying a simplified Hamiltonian, denoted  $H_{Birk}$  see Formula (8), obtained from  $H$  by removing all terms of degree  $2q + 2$  which do not Poisson commute with the quadratic part. Its Hamiltonian vector field is tangent to infinitely many subspaces obtained by setting some of the coordinates equal to 0 (cf. [13], Prop. 1). On infinitely many of them the restricted system is completely integrable, thus the next step consists in choosing such a subset  $S$  which, for obvious reasons, is called of *tangential sites*.

In this step the linear solution of (2) deforms to a quasi-periodic solution

$$\sum_i \sqrt{\xi_i} e^{it(|j_i|^2 + \omega_i^{(1)}(\xi))} e^{ij_i \cdot \phi}, \quad \text{for } \omega_i^{(1)}(\xi) \text{ see Formula (53)}. \quad (3)$$

We then apply a KAM algorithm, starting from the small quasi-periodic orbits parametrised by  $\xi$  in a suitable domain. To be precise

**Theorem 1** *For any choice of  $n$  generic frequencies  $S = \{j_1, \dots, j_n\} \subset \mathbb{Z}^d$ , and for  $\varepsilon$  sufficiently small, there exists a compact set  $\mathcal{O}_\infty$  contained in  $\{(\xi_1, \dots, \xi_n)\}$ ,  $\varepsilon^2/2 \leq \xi_i \leq \varepsilon^2\}$  of measure of order  $\varepsilon^{2n}$ , parametrizing bijectively a set of quasi-periodic solutions of (1) which are a small perturbation of the solutions of type (3) of the equations associated to the hamiltonian  $H_{Birk}$ . Moreover the quasi-periodic solutions for all  $\xi \in \mathcal{O}_\infty$  are reducible KAM tori, see Definition 9.*

As is well known KAM algorithms require strong *non-degeneracy* conditions (namely the the eigenvalues of the operator linearized in a neighborhood of zero be bounded away from zero and distinct) not always valid, even for finite dimensional systems, this has for long time been an obstacle for applications to PDEs on tori. Indeed existence results for quasi-periodic solutions (with no control on the reducibility) for such equations were proved, starting from the late '90, by "multiscale" techniques (see [6],[3]). In the case of the resonant NLS we mention the paper [16] which covers our equation (1) and provides an existence result. The breakthrough result in KAM theory was in the paper [9], where the authors proved reducibility for a class of non-resonant NLS equations (see also [17] and [7], [10] for the beam equation). A main point in these papers was to impose a quantitative lower bound on the difference of the eigenvalues, this required the introduction of Töplitz Lipschitz functions.

In the case of resonant PDEs the first problem that arises in KAM schemes is to prove (when it holds) reducibility for the Birkhoff normal form, for the NLS this was done in [15]. Then, in order to proceed with the KAM scheme one needs further *non-resonance* assumptions on the normal form (the so called Melnikov conditions) which are in general much harder to prove than in the non-resonant cases.

In the case of the cubic NLS, i.e.  $q = 1$ , in [14] we have discussed in detail the KAM algorithm and proved the existence of families of stable and unstable quasi-periodic solutions. This required a very subtle combinatorial analysis (performed in [15]) which enabled us to prove the *second Melnikov conditions* (which amounts to proving that the NLS equation linearised at an approximate solution has distinct eigenvalues on the space of quasi-periodic functions). This combinatoric is not available for  $q > 1$  except in dimension  $d \leq 2$  (see. [11]).

In the present paper we discuss the general case  $d \geq 1, q \geq 1$  and prove that, for a generic choice of the excited frequencies  $S$ , the multiplicity of the eigenvalues of the corresponding linearised system is uniformly bounded, and moreover there is a normal form with a block diagonal non-degeneracy, see Proposition 1.

Using the properties of this normal form in section §4 we explain how a KAM algorithm leads to existence and reducibility of quasi-periodic solutions. This requires some minor variations in the KAM scheme of [14] in order to take into account the block diagonal structure (essentially one needs to control more derivatives in the  $\xi$  variables). Since the material is quite standard, but very heavy and lengthy, we only give a schematic proof, contained Propositions 3 and 4, see 4.3. For a detailed exposition we refer the reader to [14].

Finally we mention that very similar kind of problems appear in the preprint [10] where the authors study the non-linear beam equation.

## 2 The Hamiltonian formalism

Passing to the Fourier representation

$$u(t, \varphi) := \sum_{k \in \mathbb{Z}^d} u_k(t) e^{i(k, \varphi)} \quad (4)$$

we have, up to a rescaling of  $u$  and of time, in coordinates, dropping the part of  $G$  which will be placed in the perturbation, the Hamiltonian:

$$H := \sum_{k \in \mathbb{Z}^d} |k|^2 u_k \bar{u}_k + \sum_{k_i \in \mathbb{Z}^d: \sum_{i=1}^{2q+2} (-1)^i k_i = 0} u_{k_1} \bar{u}_{k_2} u_{k_3} \bar{u}_{k_4} \cdots u_{k_{2q+1}} \bar{u}_{k_{2q+2}}. \quad (5)$$

The complex symplectic form is  $i \sum_k du_k \wedge d\bar{u}_k$ , and we work on the scale of complex Hilbert spaces  $(u, \bar{u}) \in \bar{\ell}^{(a,p)} \times \bar{\ell}^{(a,p)}$ , where:

$$\bar{\ell}^{(a,p)} := \{u = \{u_k\}_{k \in \mathbb{Z}^d} \mid |u_0|^2 + \sum_{k \in \mathbb{Z}^d} |u_k|^2 e^{2a|k|} |k|^{2p} := \|u\|_{a,p}^2 \leq \infty\}, \quad (6)$$

Where  $a > 0, p > d/2$ . Note that both  $H$  and its Hamiltonian vector field  $X_H$  are analytic functions on these spaces<sup>1</sup> We denote as usual by  $\{A, B\}$  the associated

<sup>1</sup> it is well known that the NLS is locally well posed under much weaker regularity conditions. This is not the purpose of the present paper.

Poisson bracket and, if we want to stress the role of one of the two variables, we also write  $ad(A)$  for the linear operator  $B \mapsto \{A, B\}$ .<sup>2</sup>

We systematically apply the fact that we have  $d + 1$  conserved quantities: the  $d$ -vector *momentum*  $\mathbb{M} := \sum_{k \in \mathbb{Z}^d} |u_k|^2 k$  and the scalar *mass*  $\mathbb{L} := \sum_{k \in \mathbb{Z}^d} |u_k|^2$  with

$$\{\mathbb{M}, u_h\} = iu_h h, \quad \{\mathbb{M}, \bar{u}_h\} = -i\bar{u}_h h, \quad \{\mathbb{L}, u_h\} = iu_h, \quad \{\mathbb{L}, \bar{u}_h\} = -i\bar{u}_h. \quad (7)$$

The terms in equation (5) commute with  $\mathbb{L}$ . The conservation of momentum is expressed by the constraints  $\sum_{i=1}^{2q+2} (-1)^i k_i = 0$ .

We partition  $\mathbb{Z}^d = S \cup S^c$ ,  $S := (j_1, \dots, j_n)$  where the elements of  $S$  play the role of *tangential sites* and those of the complement  $S^c$  the *normal sites*. We divide  $u \in \bar{\ell}^{a,p}$  in two components  $u = (u_1, u_2)$ , where  $u_1$  has indexes in  $S$  and  $u_2$  in  $S^c$ . Here we always assume that  $S$  is subject to the constraints which make it *generic* and which are fully discussed in [13] and finally refined in [15]. Further constraints will appear later in this paper.

We apply a standard *semi-normal form* change of variables  $\Psi^{(1)} := e^{ad(F_{Birk})}$ , which is well defined and analytic:  $B_{\epsilon_0} \times B_{\epsilon_0} \rightarrow B_{2\epsilon_0} \times B_{2\epsilon_0}$ , for  $\epsilon_0$  small enough, see [13].

The map  $\Psi^{(1)}$  brings (5) to the form  $H = H_{Birk} + P^{2q+2}(u) + P^{4q+2}(u)$  where  $P^{2q+2}(u)$  is of degree  $2q+2$  in  $u$  and at least cubic in  $u_2$  while  $P^{4q+2}(u)$  is analytic of degree at least  $4q+2$  in  $u$ , finally

$$H_{Birk} := \sum_{k \in \mathbb{Z}^d} |k|^2 u_k \bar{u}_k + \sum_{\alpha, \beta \in \mathcal{C}'} \binom{2}{\alpha} \binom{2}{\beta} u^\alpha \bar{u}^\beta. \quad (8)$$

$$\mathcal{C}' : \alpha, \beta \in (\mathbb{Z}^d)^{\mathbb{N}} \quad \left| \begin{array}{l} |\alpha_2| + |\beta_2| \leq 2; \quad |\alpha| = |\beta| = q + 1, \\ \sum_k (\alpha_k - \beta_k) k = 0, \quad \sum_k (\alpha_k - \beta_k) |k|^2 = 0. \end{array} \right. \quad (9)$$

The constraint  $|\alpha_2| + |\beta_2| \leq 2$  comes from the definition of  $P^{2q+2}(u)$ , the other three constraints in this formula express the conservation of  $\mathbb{L}$ ,  $\mathbb{M}$  and of the *quadratic energy*:

$$\mathbb{K} := \sum_{k \in \mathbb{Z}^d} |k|^2 u_k \bar{u}_k. \quad (10)$$

Switching to polar coordinates we set

$$u_k := z_k \text{ for } k \in S^c, \quad u_{j_i} := \sqrt{\xi_i + y_i} e^{ix_i} = \sqrt{\xi_i} \left(1 + \frac{y_i}{2\xi_i} + \dots\right) e^{ix_i} \quad (11)$$

for  $i = 1, \dots, n$ . Here we consider the  $\xi_i > 0$  as parameters  $|y_i| < \xi_i$  while  $y, x, w := (z, \bar{z})$  are dynamical variables<sup>3</sup>. We denote by  $\ell^{(a,p)} := \ell_S^{(a,p)}$  the subspace of  $\bar{\ell}^{(a,p)} \times \bar{\ell}^{(a,p)}$  of the sequences  $u_i, \bar{u}_i$  with indices in  $S^c$  and denote the coordinates  $w = (z, \bar{z})$ . We define

$$\Lambda := \left[ \frac{1}{2}, \frac{3}{2} \right]^n.$$

<sup>2</sup>  $ad$  stands for *adjoint* in the language of Lie algebras.

<sup>3</sup> To be completely formal one should think of  $z, \bar{z}$  as independent dynamical variables, to this purpose they are often denoted by  $z^+, z^-$ , then one shows that the *real* subspace, where  $\bar{z}^+ = z^-$  is invariant w.r.t. the dynamics

We choose  $\varepsilon > 0$  small and note that for all  $\xi \in \varepsilon^2 \Lambda$  and for all  $r < \varepsilon/2$ , formula (11) is an analytic and symplectic change of variables  $\Phi_\xi$  in the domain

$$\begin{aligned} D_{a,p}(s,r) &= D(s,r) := \\ \{x,y,w : x \in \mathbb{T}_s^n, |y| \leq r^2, \|w\|_{a,p} \leq r\} &\subset \mathbb{T}_s^n \times \mathbb{C}^n \times \ell^{(a,p)}. \end{aligned} \quad (12)$$

Here  $\varepsilon > 0$ ,  $s > 0$  and  $0 < r < \varepsilon/2$  are auxiliary parameters while  $\mathbb{T}_s^n$  denotes the compact subset of the complex torus  $\mathbb{T}_\mathbb{C}^n := \mathbb{C}^n / 2\pi\mathbb{Z}^n$  where  $x \in \mathbb{C}^n$ ,  $|\operatorname{Im}(x)| \leq s$ . Moreover there exist universal constants  $c_1 < 1/2$ ,  $c_2$  such that if

$$r < c_1 \varepsilon \quad \text{and} \quad \sqrt{2nc_2} \kappa^p e^{(s+a\kappa)} \varepsilon < \varepsilon_0, \quad \kappa := \max(|j_i|) \quad (13)$$

the change of variables sends  $D(s,r) \rightarrow B_{\varepsilon_0}$  so we can apply it to our Hamiltonian.

We thus assume that the parameters  $\varepsilon$ ,  $r$ ,  $s$  satisfy (13). Formula (11) puts in action angle variables  $(y; x) = (y_1, \dots, y_n; x_1, \dots, x_n)$  the tangential sites, close to the action  $\xi$ , parameters for the system. The symplectic form is now  $dy \wedge dx + i \sum_{k \in S^c} dz_k \wedge d\bar{z}_k$ . By abuse of notations we still call  $H$  the composed Hamiltonian  $H \circ \Psi^{(1)} \circ \Phi_\xi$ .

Remark that, in polar coordinates the Hamiltonians  $\mathbb{L}$ ,  $\mathbb{M}$ ,  $\mathbb{K}$ , after dropping some constant terms which Poisson commute with everything become

$$\begin{aligned} \mathbb{M} &= \sum_i y_i j_i + \sum_{k \in S^c} k |z_k|^2, \quad \mathbb{L} = \sum_i y_i + \sum_{k \in S^c} |z_k|^2, \\ \mathbb{K} &= j^{(2)} \cdot y + \sum_{k \in S^c} |k|^2 |z_k|^2, \quad j^{(2)} := (|j_1|^2, \dots, |j_n|^2). \end{aligned} \quad (14)$$

*The standard form* By the rules of Poisson bracket, on the real space spanned by  $z, \bar{z}$ , we have  $\{\bar{a}, b\} = -\{a, b\} = \{a, \bar{b}\}$  so  $\{a, \bar{b}\} = -\{\bar{a}, b\} = \{b, \bar{a}\}$  is imaginary:

**Definition 1**  $(a, b) := i\{a, \bar{b}\}$  is a real symmetric form, called the *standard form*.

For the variables we have  $(z_h, z_k) = \delta_k^h$ ,  $(\bar{z}_h, \bar{z}_k) = -\delta_k^h$ , so the form is positive definite on the space spanned by the  $z$ , which give an orthonormal basis, negative on the space spanned by the  $\bar{z}$  and of course indefinite if we mix the two types of variables. Thus we may say that an element  $a$  in the real space spanned by  $z, \bar{z}$  is of *type  $z$*  (resp. *type  $\bar{z}$* ) if  $(a, a) = 1$  resp.  $(a, a) = -1$ . For a quadratic real Hamiltonian  $\mathcal{H} = \bar{\mathcal{H}}$  we have  $\{\mathcal{H}, \{a, \bar{b}\}\} = 0$  since  $\{a, \bar{b}\}$  is a scalar. The map  $x \mapsto i\{\mathcal{H}, x\}$  preserves the real subspace spanned by  $z, \bar{z}$  hence by the Jacobi identity  $(a, i\{\mathcal{H}, b\}) = (i\{\mathcal{H}, a\}, b)$  so the operator  $i\{\mathcal{H}, -\}$  is symmetric with respect to this form.

## 2.1 Functional setting

Following [12] we study *regular* functions  $F : \varepsilon^2 \Lambda \times D_{a,p}(s,r) \rightarrow \mathbb{C}$ , that is whose Hamiltonian vector field  $X_F(\cdot; \xi)$  is M-analytic from  $D(s,r) \rightarrow \mathbb{C}^n \times \mathbb{C}^n \times \ell_S^{a,p}$ . In the variables  $\xi$  we require  $C^{5d^2}$  regularity. Let us recall the definitions from [4]. Let us consider the Banach space  $V := \mathbb{C}^n \times \mathbb{C}^n \times \ell_S^{a,p}$  with  $(s,r)$ -weighted norm

$$\vec{v} = (x, y, z, \bar{z}) \in V, \quad \|\vec{v}\|_{V,s,r} := \frac{|x|_\infty}{s} + \frac{|y|_1}{r^2} + \frac{\|z\|_{a,p}}{r} + \frac{\|\bar{z}\|_{a,p}}{r} \quad (15)$$

where,  $|x|_\infty := \max_{h=1,\dots,n} |x_h|$ ,  $|y|_1 := \sum_{h=1}^n |y_h|$  and we restrict  $r, s$  with  $0 < s < 1$ ,  $0 < r < c_1 \varepsilon$ .

For a vector field  $X : D(s, r) \rightarrow V$ , described by the formal Taylor expansion:

$$X = \sum_{\nu, i, \alpha, \beta} X_{\nu, i, \alpha, \beta}^{(\nu)} e^{i(\nu, x)} y^i z^\alpha \bar{z}^\beta \partial_\nu, \quad \mathbf{v} = x, y, z, \bar{z}$$

we define the *majorant* and its *norm*:

$$\begin{aligned} MX &:= \sum_{\nu, i, \alpha, \beta} |X_{\nu, i, \alpha, \beta}^{(\nu)}| e^{s|\nu|} y^i z^\alpha \bar{z}^\beta \partial_\nu, \quad \mathbf{v} = x, y, z, \bar{z} \\ \|X\|_{s,r} &:= \sup_{(y,z,\bar{z}) \in D(s,r)} \|MX\|_{V,s,r}. \end{aligned} \quad (16)$$

The different weights ensure that, if  $\|X_F\|_{s,r}$  is sufficiently small, then  $F$  generates a close-to-identity symplectic change of variables from  $D(s/2, r/2) \rightarrow D(s, r)$ .

In our algorithm we deal with functions which depend in a  $C^\ell$  way on some parameters  $\xi$  in a compact set  $\mathcal{O} \subseteq \varepsilon^2 \Lambda$ , the integer  $\ell$  in fact will be chosen to be  $(2d+1)^2$ , the maximal size of diagonal blocks of the normal form of the NLS. To handle this dependence one introduces weighted norms for a map  $X : \mathcal{O} \times D(s, r) \rightarrow V$  setting:

$$\|X(\xi)\|_{s,r}^\lambda := \sum_{k \in \mathbb{N}^n: |k| \leq \ell} \lambda^{|k|} \|\partial_\xi^k X\|_{s,r}, \quad \|X\|_{s,r,\mathcal{O}}^\lambda := \sup_{\xi \in \mathcal{O}} \|X(\xi)\|_{s,r}^\lambda \quad (17)$$

where  $\lambda$  is a parameter of order  $\varepsilon^2$ . Sometimes when  $\mathcal{O}$  is understood we just write  $\|X\|_{s,r,\mathcal{O}}^\lambda = \|X\|_{s,r}^\lambda$ .

During the KAM algorithm we shall use a stronger norm, called *quasi-Töplitz*, depending on various parameters  $\vec{p}$  given by  $\lambda, s, r, \mathcal{O}$  and three parameters  $K, \vartheta, \mu$  with  $K$  a large positive integer and  $\frac{1}{2} < \vartheta, \mu < 4$ . We denote such norm by  $\|\cdot\|_{\vec{p}}^T$ . Since the definition is quite long we do not restate it in the present paper but refer the reader to [17] and [14], we only give in the appendix B an informal presentation of quasi-Töplitz Hamiltonians and their main properties. The main point is that this norm is closed w.r.t Poisson brackets and moreover it controls the behaviour of linear Hamiltonian vector fields with respect to linear stratifications (cf. Definition 5), this will be essential in the KAM algorithm for controlling the measure estimates.

**Definition 2** We define by  $\mathcal{V}_{\lambda,s,r,\mathcal{O}} = \mathcal{V}_{s,r}$ ,  $\mathcal{H}_{\lambda,s,r,\mathcal{O}} = \mathcal{H}_{s,r}$ , resp.  $\mathcal{H}_{\vec{p}}^T$  with  $\vec{p} = (\lambda, s, r, \vartheta, \mu, K, \mathcal{O})$ , the space of  $M$ -analytic vector fields, resp. regular analytic and finally quasi-Töplitz, Hamiltonians depending on a parameter  $\xi \in \mathcal{O}$  with the norms<sup>4</sup>

$$\|X\|_{s,r}^\lambda, \quad \|F\|_{s,r}^\lambda := \|X_F\|_{s,r}^\lambda < \infty, \quad \text{resp.} \quad \|X_F\|_{\vec{p}}^T < \infty, \quad (18)$$

where  $\vec{p} = (\lambda, s, r, \vartheta, \mu, K, \mathcal{O})$ .

<sup>4</sup> in fact Hamiltonians should be considered up to scalar summands and then this is actually a norm

The main properties of the majorant norm, contained in [2], Lemmata 2.10, 2.15, 2.17, express the compatibility of the norm with projections and Poisson brackets. Similarly the main properties of the quasi-Töplitz norm may be found in §8.17.1 and in Proposition 9.2 of [14].

Since  $\mathcal{H}_{\overline{p}}^T$  is a space of M-analytic functions it is naturally spanned by the monomials  $e^{i\nu \cdot x} y^i z^\alpha \bar{z}^\beta$ . It is natural to give degree 0 to the angles  $x$ , degree 2 to  $y$  and 1 to  $w$ . In this way each element  $F \in \mathcal{H}_{\overline{p}}^T$  is expanded as  $F = \sum_{j=0}^{\infty} F^{(j)}$ . We will need the projections onto various subspaces, in particular those defined by the degree, which are all continuous w.r.t the majorant norm.

**Definition 3** A linear operator  $L$  on  $\mathcal{H}_{\overline{p}}^T$  has degree  $j$  if it maps elements of degree  $k$  into elements of degree  $k + j$  for all  $k$ .

Notice that the degree of the composition of two linear operators of degree  $i, j$  is their sum  $i + j$ .

*Remark 1* If  $A \in \mathcal{H}_{s,r}$  has degree  $j$  the linear operator  $\text{ad}(A)$  has degree  $j - 2$ . By the Cauchy estimates (Lemma 2.17 of [2]) this is continuous as operator  $\mathcal{H}_{s,r} \rightarrow \mathcal{H}_{s',r'}$  with  $s' < s$  and  $r' < r$ .

**Definition 4** The *normal form*  $\mathcal{N}$  collects all the terms of  $H_{Birk}$  of degree  $\leq 2$  (dropping the constant terms). We then set  $P := H - \mathcal{N}$  or  $H = \mathcal{N} + P$ .

For a basic *finiteness property* of the normal form we need the following:

**Definition 5** A *linear stratification* of  $\mathbb{R}^n$  is a finite decomposition  $\mathbb{R}^n = \bigcup_j Y_j$  where the closure  $\bar{Y}_j$  of each  $Y_j$  is a linear affine subspace, and  $Y_j$  is obtained from  $\bar{Y}_j$  by removing a finite number of proper linear affine subspaces.

A given stratum  $Y$  lies in a minimal affine space  $\bar{Y}$  and the group  $T_Y$  of translations of  $\bar{Y}$  is the *group of translations of*  $Y$ .

*Remark 2* A linear stratification is obtained by choosing a finite list of linear affine subspaces  $A_i$ . Then to each  $A_i$  is associated, as stratum, the set of points in  $A_i$  which do not lie in any of the affine spaces  $A_j$  of the list which do not contain  $A_i$ .

Thus  $Y$  is obtained from any of its elements  $y$  as  $y + T_Y$  and then removing the lower dimensional strata.

*Remark 3* If each  $Y_i$  is defined by linear equations over  $\mathbb{Z}$  we speak of an *integral linear stratification*, this induces by intersection a linear stratification of  $\mathbb{Z}^n$ .

The integral points satisfying  $\ell$  independent linear equations with coefficients in  $\mathbb{Z}$  form a subgroup  $\Gamma$  of  $\mathbb{Z}^d$  isomorphic to  $\mathbb{Z}^{d-\ell}$  (with an integral basis which can be extended to a basis of  $\mathbb{Z}^d$ ). Correspondingly the intersection of an affine space  $\bar{Y}$  (defined over  $\mathbb{Z}$ ) of dimension  $k = d - \ell$  with  $\mathbb{Z}^d$  is of the form  $\Gamma + u$  where  $u \in \mathbb{Z}^d$  and  $\Gamma \subset \mathbb{Z}^d$  is a group of integral translations isomorphic to  $\mathbb{Z}^k$  (usually  $k$  is called *rank*). Each stratum has thus a rank and is obtained from such a translate by removing a finite number of translates of subgroups of strictly lower rank.

The linear stratifications appearing in this paper come from two sources, the theory of graphs and block diagonal structure of the Normal form and the Theory of cuts, see § 7.1 and Appendix B.

### 3 Properties of the normal form

Here we state the main properties used in the KAM algorithm. These properties are detailed and proved in §5.

We need to explain the following notation and fact. First given a sign  $\sigma = \pm 1$  and a variable  $z_k$  we write  $z_k^\sigma = z_k$  if  $\sigma = 1$  and  $\bar{z}_k$  if  $\sigma = -1$ .

Denote by  $\mathcal{F}$  the space of Hamiltonians of degree  $\leq 2$  (which form a Lie algebra under Poisson bracket). In our treatment we have a normal form Hamiltonian with *normal* variables  $z_k, \bar{z}_k$   $k \in S^c$ . The Hamiltonian is block diagonal w.r.t a block decomposition with finite *blocks*  $D_t \subset S^c$  indexed by a denumerable index set  $\mathfrak{T}$ . In turn this set is divided into an infinite set  $\mathfrak{T}_s$  and a finite set  $\mathfrak{T}_f$ . For each  $t \in \mathfrak{T}$  we have a finite set of indices  $k \in D_t \subset S^c$  and, for the finitely many elements in  $\mathfrak{T}_f$  also a sign function  $s(k) = \pm 1$  for  $k \in D_t$ . When we have such a sign we need to divide the corresponding variables  $\{z_k, \bar{z}_k\}_{k \in D_t}$  in the set  $z_k^{s(k)}$  and its conjugate. The variables span a symplectic space and the two sets each span a lagrangian subspace.

**Definition 6** An  $x$  independent quadratic Hamiltonian on the space  $\{z_k, \bar{z}_k\}_{k \in D_t}$  will be called *Lagrangian* if under Poisson bracket it preserves these two spaces, this means that we never have in its expression a term of type  $z_h z_k$  or  $\bar{z}_h \bar{z}_k$  (with  $h, k \in D_t$ ) if  $s(k) = s(h)$  or  $z_h \bar{z}_k$  if  $s(k) = -s(h)$ . In matrix terms this is the standard embedding of  $m \times m$  matrices into the Lie algebra of the symplectic group of dimension  $2m$ .

In fact, by Formula (21), this Lagrangian structure is equivalent to conservation of momentum (or of quadratic energy). We then have the main properties of the normal form summarized as follows:

**Proposition 1** For any generic choice of the tangential sites  $S = \{j_1, \dots, j_n\}$ , there exists a homogeneous algebraic hypersurface  $\mathfrak{A}$ , whose complement in  $\mathbb{R}_+^n$  is union of simply connected open regions  $\mathcal{R}_\alpha$  with the property:

For each  $\alpha$  there is an analytic family of symplectic changes of variables  $\mathcal{R}_\alpha \times D(s, r) \mapsto B_{\epsilon_0}$  which conjugates the NLS Hamiltonian to the following form  $H = \mathcal{N}^{(s)} + \mathcal{N}^{\text{nil}} + P$  where:

$$\begin{aligned} \mathcal{N}^{(s)} &:= \omega(\xi) \cdot y + \sum_{t \in \mathfrak{T}_s} \Omega_t(\xi) \sum_{k \in D_t} |z_k|^2 + \sum_{t \in \mathfrak{T}_f} \Omega_t(\xi) \sum_{k \in D_t} s(k) |z_k|^2, \\ \mathcal{N}^{\text{nil}} &= \sum_{t \in \mathfrak{T}_f} \mathcal{Q}_t^{\text{nil}}. \end{aligned} \quad (19)$$

The set  $\mathfrak{T}_s$  is a denumerable index set while  $\mathfrak{T}_f$  is finite. Each index  $t$  identifies a point  $\mathbf{r}_t \in S^c$  and an algebraic function  $\theta_t$ , homogeneous of degree  $q$ , chosen from a finite list  $\mathcal{Y}$  (see Formula (69)). The map  $t \mapsto (\mathbf{r}_t, \theta_t)$  is injective. To  $t$  is associated a finite set  $D_t \subset S^c$ .

The cardinality  $d_t$  of the set  $D_t$  is  $d_t \leq 2d + 1$  for  $t \in \mathfrak{T}_f$  and  $d_t \leq d + 1$  for  $t \in \mathfrak{T}_s$  (note that one may have that  $\mathbf{r}_t \notin D_t$ ).

Moreover the sets  $D_t$  give a disjoint decomposition of the normal sites  $S^c = \mathbb{Z}^d \setminus S$ .

*i) Non-degeneracy* We have  $\omega(\xi) = \mathbf{j}^{(2)} + \omega^{(1)}(\xi)$  where  $\omega^{(1)}(\xi)$  is homogeneous of degree  $q$  in the variables  $\xi$ . The map  $(\xi_1, \dots, \xi_m) \mapsto (\omega_1(\xi), \dots, \omega_m(\xi))$  is an algebraic local diffeomorphism for  $\xi$  outside some real algebraic hypersurface.

ii) **Asymptotic of the normal frequencies:** We have

$$\Omega_t(\xi) = |\mathbf{r}_t|^2 + \theta_t(\xi)$$

where all the functions  $\theta_t$  are chosen from the finite list  $\Upsilon$ .

Finally  $s(k) = \pm 1$  while  $\mathcal{Q}_t^{\text{nil}}$  is a Lagrangian (see Definition 6) nilpotent quadratic Hamiltonian in the variables  $\{z_k, \bar{z}_k\}_{k \in \mathbb{D}_t}$  independent of  $x$ .

iii) **Translation invariance:** The infinite set  $\mathfrak{T}_s$  decomposes into finitely many components  $\mathfrak{T}_s(i)$  with the following property.

For all  $t \in \mathfrak{T}_s(i)$  the dimension  $d_t$  is the same.

For each given  $\mathfrak{T}_s(i)$ , the elements  $\mathbf{r}_{t_2} - \mathbf{r}_{t_1}$ ,  $t_1, t_2 \in \mathfrak{T}_s(i)$  generate a subgroup of translations  $\Gamma_i \subset \mathbb{Z}^d$  of rank  $k_i = d - d_t + 1$ , (that is isomorphic to some  $\mathbb{Z}^{k_i}$ ,  $k_i \leq d$ ). Choosing an element  $t_0 \in \mathfrak{T}_s(i)$  gives a subset  $\Gamma_i^0 \subset \Gamma_i$  obtained from the group  $\Gamma_i$  by removing finitely many translates of subgroups of lower rank so that

$$\begin{aligned} \forall t_1, t_2 \in \mathfrak{T}_s(i), \quad \mathbb{D}_{t_2} = \mathbb{D}_{t_1} + \mathbf{r}_{t_2} - \mathbf{r}_{t_1}, \quad \mathbf{r}_{t_2} - \mathbf{r}_{t_1} \in \Gamma_i \\ \bigcup_{t \in \mathfrak{T}_s(i)} \mathbb{D}_t = \bigcup_{a \in \Gamma_i^0} \mathbb{D}_{t_0} + a. \end{aligned} \quad (20)$$

iv) **Affine structure:** Formula (20) determines an integral linear stratification  $\Sigma_0$  of  $\mathbb{Z}^d$  such that the set of roots  $\mathbf{r}_t$  is a union of strata.

We take all points in some  $\mathbb{D}_t$ ,  $t \in \mathfrak{T}_f$  to be zero dimensional strata.

Then for each  $i$ , the set  $\bigcup_{t \in \mathfrak{T}_s(i)} \mathbb{D}_t$  is the union of the parallel strata  $\Gamma_i^0 + k$ ,  $k \in \mathbb{D}_{t_0}$ .

v) **Piecewise Töplitz:** The function  $\theta_t(\xi)$  is the same for all  $t \in \mathfrak{T}_s(i)$ .

vi) **Constants of motion:** in the new coordinates, mass and momentum are:

$$\mathbb{L} = \sum_i y_i + \sum_k s(k) |z_k|^2, \quad \mathbb{M} = \sum_i y_i \mathbf{j}_i + \sum_{t \in \mathfrak{T}} \mathbf{r}_t \left( \sum_{k \in \mathbb{D}_t} s(k) |z_k|^2 \right), \quad (21)$$

moreover  $\mathcal{N}$  has as further constant of motion the quadratic energy:

$$\mathbb{K} = \sum_i |\mathbf{j}_i|^2 y_i + \sum_{t \in \mathfrak{T}} |\mathbf{r}_t|^2 \left( \sum_{k \in \mathbb{D}_t} s(k) |z_k|^2 \right). \quad (22)$$

vii) **Smallness:** Consider any compact domain  $\mathcal{O}$  in  $\mathcal{R}_\alpha \cap \varepsilon^2 \Lambda$ . If  $\varepsilon^3 < r < c_1 \varepsilon$ , the perturbation  $P$  in the new variables is small, more precisely we have the bounds:

$$\|X_P\|_{\frac{T}{p}} \leq C(\varepsilon^{2q-1} r + \varepsilon^{2q+3} r^{-1}), \quad (23)$$

where  $\vec{p} = (\lambda, s, r, \vartheta_0, \mu_0, N_0, \mathcal{O})$  with  $\vartheta_0 = 1$ ,  $\mu_0 = 2$  and  $N_0 = 8(d+1)! \kappa^{d+1}$  (see (13) for  $\kappa$ ) while  $C$  is independent of  $s, r$  and depends on  $\varepsilon, \lambda$  only through  $\lambda/\varepsilon^2$ .

*Proof* Most of these statements are part of Proposition 13 and Corollary 3. The smallness condition in the  $\lambda$  norm is the content of Theorem 1 of [13], the corresponding estimate for the quasi-Töplitz norm follows verbatim from part 4 of [14] extending Proposition 11.19.  $\square$

*Remark 4* Here *generic* means that the list of excited frequencies  $S$ , thought of as a vector in  $\mathbb{Z}^{dn}$ , is not a solution of a (complicated) polynomial  $P$  (in  $dn$  variables) with integer coefficients, the polynomial  $P$  is the product of explicit polynomials associated to a finite list of graphs which represent resonances to be avoided. Of course this list grows exponentially with the dimension and with  $q$  so, although the polynomial is described explicitly, it is not possible to write down the polynomial in a file. Nevertheless the *density* of generic choices clearly tends to 1 (cf. [14]).

*Remark 5* The index set  $\mathfrak{T}$  and the corresponding decomposition of the space as well as the affine structure depend only on the chosen connected component.

In the case of the cubic NLS the picture simplifies drastically since there is no nilpotent part and moreover, all the sets  $\mathbb{D}_t$  reduce to one element. This depends on the fact that the normal form is diagonalizable with distinct eigenvalues, while here we have to take into account the multiplicities.

**Proposition 2 (Melnikov conditions)** *For any  $\alpha$  and for all  $\xi \in \mathcal{R}_\alpha$ , the kernel of  $ad\mathcal{N}^{(s)}$  in the space  $\mathcal{F}$  (of hamiltonians with degree  $\leq 2$ ) is given by the subspace  $\mathcal{F}_{\ker}$ , of hamiltonians of the form*

$$\psi(\xi) \cdot y + \sum_{t \in \mathfrak{T}} Q_t(\xi, w)$$

where  $\psi(\xi) \in \mathbb{R}^n$  and  $Q_t(\xi, w)$  are  $x$  independent Lagrangian quadratic Hamiltonians depending only on the variables  $\{z_k, \bar{z}_k\}_{k \in \mathbb{D}_t}$ .

*Proof* The proof is in Proposition 14. □

This proposition induces a decomposition<sup>5</sup> of  $\mathcal{F}$  into:

**Definition 7** The subalgebra  $\mathcal{F}_{\ker}$  and its unique complement  $\mathcal{F}_{\text{rg}}$  stable under adjoint action of  $\mathcal{F}_{\ker}$ .

**Definition 8** We consider the free abelian group  $\mathbb{Z}^n$  with canonical basis  $e_i$  and define the following linear maps  $\eta : \mathbb{Z}^n \rightarrow \mathbb{Z}$ ,  $\pi : \mathbb{Z}^n \rightarrow \mathbb{Z}^d$ ,  $\pi^{(2)} : \mathbb{Z}^n \rightarrow \mathbb{Z}$

$$\eta : e_i \mapsto 1, \quad \pi : e_i \mapsto j_i, \quad \pi^{(2)} : e_i \mapsto |j_i|^2. \quad (24)$$

**Warning** In  $\mathbb{Z}^n$  we always use as norm  $|l|$  the  $L^1$  norm  $\sum_{i=1}^n |l^{(i)}|$ . On the other hand in  $\mathbb{Z}^d$ , and hence in  $S^c$ , we use the euclidean  $L^2$  norm.

The space  $\mathcal{F}$  decomposes as a direct sum of its parts of degree 0, 1, 2 respectively with basis

$$0) \quad e^{-i\nu \cdot x}, \quad 1) \quad e^{-i\nu \cdot x} z_k^\sigma, \quad 2) \quad e^{-i\nu \cdot x} z_h^{\sigma_1} z_k^{\sigma_2}, \quad e^{-i\nu \cdot x} y_i$$

satisfying mass and momentum conservation, deduced from Formula (21). We usually denote  $z^+ = z$ ,  $z^- = \bar{z}$  for convenience. We exclude the constants. We denote this degree decomposition as  $\mathcal{F} := F^0 \oplus F^{0,1} \oplus F^2$ ,  $F^2 = F^{0,2} \oplus F^{1,0}$ .

We now decompose  $F^{0,1}$  as orthogonal sum (with respect to the symplectic form), of subspaces (blocks)  $(\mathbb{D}_t, \nu)^{\pm 1}$  with  $t \in \mathfrak{T}$  and  $\nu \in \mathbb{Z}^n$  constrained by the conservation laws. Each  $(\mathbb{D}_t, \nu)^+$  has as basis the elements  $e^{i\nu \cdot x} z_k^{s(k)}$  where  $k \in \mathbb{D}_t$ .

<sup>5</sup> Roughly speaking this is just dividing a space of matrices into some block diagonal subalgebra and the stable off-diagonal part.

For each index  $(\mathfrak{t}, \nu)$  the constraints

$$\pi(\nu) + \mathbf{r}_{\mathfrak{t}} = \nu \cdot \mathbf{j} + \mathbf{r}_{\mathfrak{t}} = 0, \quad \eta(\nu) + 1 = 0 \quad (25)$$

express respectively the conservation of momentum and mass.

These blocks are obviously stable under  $ad(\mathcal{N})$  and on each block this action is invertible. On the blocks  $(\mathbf{D}_{\mathfrak{t}}, \nu)^\sigma$   $ad(\mathcal{N})$  is self adjoint for the standard form. If  $\mathfrak{t} \in \mathfrak{T}_s$  the standard form is positive definite if  $\sigma = 1$  negative if  $\sigma = -1$ .

The previous decomposition of  $F^{0,1}$  induces a block decomposition of  $F^{0,2}$ . The product of two blocks  $(\mathbf{D}_{\mathfrak{t}_i}, \nu_i)^{\pm 1}$ ,  $i = 1, 2$  produces a quadratic block in  $F^{0,2}$  stable under  $ad(\mathcal{N})$  with basis the products of basis elements. In order to avoid repetitions we index different quadratic blocks by  $(\nu, \mathfrak{t}_1, \mathfrak{t}_2)_{\sigma_1, \sigma_2}$  where the conservation laws are:

$$\eta(\nu) + 1 + \sigma_1 \sigma_2 = 0, \quad \pi(\nu) + \mathbf{r}_{\mathfrak{t}_1} + \sigma_1 \sigma_2 \mathbf{r}_{\mathfrak{t}_2} = 0 \quad (26)$$

while the basis elements are  $e^{i\sigma_1 \nu \cdot x} z_k^{\sigma_1 s(k)} z_h^{\sigma_2 s(h)}$  where  $k \in \mathbf{D}_{\mathfrak{t}_1}$  and  $h \in \mathbf{D}_{\mathfrak{t}_2}$ . For  $\sigma_1 \sigma_2$  fixed the blocks come in conjugate pairs and we usually exhibit the one with  $\sigma_1 = 1$ .

The action of  $ad(\mathcal{N})$  is invertible on all the blocks different from  $(0, \mathfrak{t}, \mathfrak{t}, \sigma, -\sigma)$  which add up to the part of  $\mathcal{F}_{\ker}$  in  $F^{(0,2)}$ .

#### 4 A KAM theorem

Having now prepared our normal form we need to explain how to perform a successful KAM algorithm.

We follow the same path as in [14] but with some differences due to the fact that our normal form is no more diagonal with different eigenvalues but rather block diagonal with uniformly bounded blocks corresponding to different eigenvalues.

The starting point for our KAM Theorem is a class of Hamiltonians

$$H := N + P, \quad N := \omega(\xi) \cdot \mathbf{y} + \sum_{\mathfrak{t} \in \mathfrak{T}} \mathcal{Q}_{\mathfrak{t}}(\xi, z, \bar{z}), \quad P = P(x, y, z, \bar{z}, \xi). \quad (27)$$

which is a small variation the class described in Proposition 1 from which we maintain all the notations. The Hamiltonians are defined in  $D(s, r) \times \mathcal{O}$ , where  $\mathcal{O} \subseteq \varepsilon^2 \Lambda$  is a compact domain of diameter of order  $\varepsilon^2$  contained in one of the components  $\mathcal{R}_\alpha \cap \varepsilon^2 \Lambda$  where the local diffeomorphism  $\omega(\xi)$  is injective. Here  $\omega(\xi) = \mathbf{j}^{(2)} + \omega^{(1)}(\xi) + \tilde{\omega}(\xi)$  where  $\mathbf{j}^{(2)}, \omega^{(1)}$  are the ones of the NLS normal form, defined in Proposition 1 item *ii*), while  $\tilde{\omega}(\xi)$  is small of the order of Formula (23). The finite (at most  $2d + 1$  degrees of freedom) dimensional quadratic Hamiltonians  $\mathcal{Q}_{\mathfrak{t}}$  depend on the variables  $z_k, \bar{z}_k$  with  $k \in \mathbf{D}_{\mathfrak{t}}$ . By the lagrangian structure of the blocks  $(\mathbf{D}_{\mathfrak{t}}, \nu)^\pm$  we represent these Hamiltonians by the matrices, denoted by  $\Omega_{\mathfrak{t}}(\xi)$ , acting on the blocks  $(\mathbf{D}_{\mathfrak{t}}, \nu)^+$ . Note that these are independent of  $\nu$ . We have that for all  $\mathfrak{t} \in \mathfrak{T}_g$  the matrix  $\Omega_{\mathfrak{t}}(\xi)$  is self-adjoint (the basis is orthonormal for the standard form). For all  $\mathfrak{t} \in \mathfrak{T}$  we have:

$$\Omega_{\mathfrak{t}}(\xi) = (|\mathbf{r}_{\mathfrak{t}}|^2 + \theta_{\mathfrak{t}}(\xi))I_{\mathfrak{t}} + \Omega_{\mathfrak{t}}^{\text{nil}} + \tilde{\Omega}_{\mathfrak{t}}(\xi). \quad (28)$$

where  $\mathbf{r}_t$  is the *root* of the stratum indexed by  $t \in \mathfrak{T}$ .  $\Omega_t^{\text{nil}}$  is a nilpotent matrix chosen from a finite list of matrices depending on  $\xi$  in an analytic way and homogeneous of degree  $q$ , similarly for  $\theta_t$ . Moreover the functions  $\theta_t$  satisfy the *piecewise Töplitz* property (iv) of Proposition 1. Finally the  $\tilde{\Omega}_t(\xi)$  define a quadratic Hamiltonian which is quasi Töplitz and small of the order of Formula (23). The same properties hold for the perturbation  $P$ .

It is well known that the Hamiltonian equations of motion for the normal form  $N$  admit the special solutions  $(x, 0, 0, 0) \rightarrow (x + \omega(\xi)t, 0, 0, 0)$ , that correspond to invariant tori in the phase space, for each  $\xi \in \mathcal{O}$ . These special solutions are solutions also of the perturbed system if and only if the Hamiltonian vector field associated to  $P$  vanishes on these invariant tori.

Our aim is to prove that, under suitable hypotheses, there is a set  $\mathcal{O}_\infty \subset \mathcal{O}$  of positive Lebesgue measure, so that, for all  $\xi \in \mathcal{O}_\infty$  the Hamiltonians  $H$  still admit invariant tori (close to the ones of the unperturbed system) given (in the new variables) by the equations  $y = z = 0$ . Moreover, the associated Hamiltonian vector field  $X_H$  restricted to these tori is  $\sum_{i=1}^n \omega_i^\infty(\xi) \partial_{x_i}$  while  $X_H$  linearised at a torus is block-diagonal in the normal variables with  $x$ -independent block matrices  $\Omega_t^\infty(\xi)$ .

**Definition 9 (reducible KAM tori)** A quasi-periodic solution of frequency  $\omega^\infty$  is a reducible KAM torus if (in appropriate symplectic variables) it is expressed by the equations  $y = z = 0$  and moreover the associated Hamiltonian vector field  $X_H$  restricted to the torus is  $\sum_{i=1}^n \omega_i^\infty(\xi) \partial_{x_i}$ , while  $X_H$  linearized at the torus is block-diagonal in the normal variables with  $x$ -independent block matrices  $\Omega_t^\infty(\xi)$  of uniformly bounded dimension.

#### 4.1 The algebraic algorithm

Let us explain the algebraic part of the algorithm. We have seen that the space  $\mathcal{F} := F^0 \oplus F^{0,1} \oplus F^2$ ,  $F^2 = F^{0,2} \oplus F^{1,0}$  can be canonically decomposed in an orthogonal sum of two parts,  $\mathcal{F} = \mathcal{F}_{\text{rg}} \oplus \mathcal{F}_{\text{ker}}$  where, for generic values of  $\xi$  the space  $\mathcal{F}_{\text{rg}}$  is the range of the normal form operator  $ad(\mathcal{N}^s)$  and  $\mathcal{F}_{\text{ker}}$  the Kernel. Denote by  $\Pi_{\text{rg}}$ ,  $\Pi_{\text{ker}}$  the corresponding projections.

Recall the degree projections  $\Pi^{\leq j}$ ,  $\Pi^j$ ,  $\Pi^{> j}$  onto polynomials of degree  $\leq j$ ,  $j, > j$ , we also denote by  $A^{\leq j} := \Pi^{\leq j} A$  for any Hamiltonian  $A$ . We shall also need to perform a *ultraviolet cut*, that is separate the subspace where the frequency  $\nu$  is bounded by some  $|\nu| \leq K$  and the rest, these projections will be denoted by putting  $K$  as pedex as in  $\Pi_{\text{rg}, \leq K}$ ,  $\Pi_{\text{ker}, \leq K}$ ,  $\Pi_{\leq K}$ ,  $\Pi_{> K}$  etc..

Up to now we have considered  $\mathcal{F}$  as an algebraic object by describing a basis but we will soon need to consider infinite linear combinations of these basis elements, defining regular quadratic Hamiltonians on some of the regions  $D(r, s)$ , moreover these will depend on parameters  $\xi$  on suitable compact domains, nevertheless the projections still make sense.

In fact it is convenient to define  $\Pi_{\text{rg}}$ ,  $\Pi_{\text{ker}}$  on the entire space of series of which  $\mathcal{F}$  is the part of degree  $\leq 2$ . We hence decompose

$$\mathcal{H}_{s,r} = \mathcal{F}_{\text{ker}} \oplus \mathcal{F}_{\text{rg}} \oplus \mathcal{H}_{s,r}^{>2}. \quad (29)$$

*Remark 6* [The goal] By definition, the normal form  $\mathcal{N}$  of Proposition 1 is in  $\mathcal{F}_{\ker}$ . In general, the condition for a hamiltonian  $H = N + P$ ,  $N = \Pi_{\ker}H$  to have KAM tori is  $\Pi_{\text{rg}}H = \Pi_{\text{rg}}P = 0$ . So our goal is to find a symplectic transformation  $\mathfrak{S}$  so that

$$\Pi_{\text{rg}}\mathfrak{S}(H) = 0.$$

The strategy is to construct this as a limit of a *quadratic Nash-Moser algorithm*.

**Warning** We deviate from the standard notation to put in evidence the intrinsic decomposition of  $\mathcal{H}_{r,s}$  given by Formula (29). In particular for a Hamiltonian  $H \in \mathcal{H}_{r,s}$  we will denote:

$$N := \Pi_{\ker}(H), \quad P_{\text{rg}} := \Pi_{\text{rg}}(H), \quad P^{>2} := \Pi^{>2}(H), \quad (30)$$

so that  $H = N + P_{\text{rg}} + P^{>2}$ . The first two terms correspond to the canonical decomposition of the Lie algebra (under Poisson bracket),  $\mathcal{F}$  into the subalgebra  $\mathcal{F}_{\ker}$  and its complement  $\mathcal{F}_{\text{rg}}$ , which is stable under the action of  $\mathcal{F}_{\ker}$ .

By Remark 1 if  $A$  has degree  $i$  we have that  $ad(A)$  has degree  $i - 2$  so as soon as  $i \geq 3$  these operators have positive degree and are nilpotent on  $\mathcal{F}$ , the cube is always 0.

We start with  $H_0 := N_0 + P_{\text{rg},0} + P_0^{>2}$ , where  $N_0$  is close to  $\mathcal{N}$  and  $P_{\text{rg},0}$  is appropriately small. We wish to find a convergent sequence of changes of variables, dependent on a sequence  $K_m$  of ultraviolet cuts,  $\mathfrak{S}_{m+1} := e^{adF_m}\mathfrak{S}_m$  so that at each step  $H_{m+1} = \mathfrak{S}_{m+1}(H_0)$  is such that  $N_m$  stays close to  $\mathcal{N}$ ,  $P_m^{>2}$  stays bounded while  $P_{\text{rg},m}$  converges to zero (super-exponentially).

At a purely formal level we would like that  $P_{\text{rg},m+1}$  is *quadratic* w.r.t.  $P_{\text{rg},m}$ . The generating function  $F_m \in \mathcal{F}_{\text{rg},\leq K_{m+1}}$  is fixed by solving the homological equation

$$\{N_m, F_m\} + \Pi_{\text{rg},m}\{P_m^{>2}, F_m\} = \Pi_{\leq K_{m+1}}P_{\text{rg},m}, \quad \Pi_{\text{rg},m} := \Pi_{\text{rg},\leq K_{m+1}} \quad (31)$$

which uniquely determines  $F_m$  as a *linear* function of  $P_{\text{rg},m}$  provided that the linear operator:

$$L_m := ad(N_m) + \Pi_{\text{rg},m}ad(P_m^{>2}) = ad(N_m) + \Pi_{\text{rg},m}ad(P_m^3) + \Pi_{\text{rg},m}ad(P_m^4),$$

is invertible on  $\mathcal{F}_{\text{rg},\leq K_{m+1}}$  (clearly we also need some quantitative control on the inverse).

*Remark 7* On  $\mathcal{F}$  the operators  $ad(N_m)$ ,  $\Pi_{\text{rg},m}ad(P_m^3)$ ,  $\Pi_{\text{rg},m}ad(P_m^4)$  have respectively degree 0,1,2 so it should be clear that  $L_m$  is invertible if and only if  $ad(N_m)$  is invertible and in this case one inverts

$$L_m = ad(N_m) \left( 1 + ad(N_m)^{-1} \Pi_{\text{rg},m}ad(P_m^{>2}) \right)$$

by inverting the second factor. This is of the form  $1 + A$  with  $A$  a sum of two linear operators of degree 1,2 respectively, so  $A^3 = 0$  and we invert  $1 + A$  with the 3 term Neumann series  $1 - A + A^2$ .

We now justify our choice by computing one *KAM step*, for notational convenience we drop the pedex  $m$  in  $H_m$  etc.. and substitute  $H_{m+1}$  with  $H_+$  etc... .

Let us compute  $H_+ := e^{adF}H$ . First split the operator  $e^{ad(F)} = 1 + ad(F) + E_F$ , by definition  $E_F$  is quadratic in  $F$  and hence quadratic in  $P_{\text{rg}}$ . Regarding the term

$$(1 + ad(F))(N + P_{\text{rg}} + P^{>2}) = N + P_{\text{rg}} + P^{>2} - \{N + P^{>2}, F\} + \{F, P_{\text{rg}}\}$$

we first notice that, since  $F$  is linear w.r.t.  $P_{\text{rg}}$  then the last summand is quadratic moreover since  $F$  solves the homological equation we have

$$P_{\text{rg}} - \{N + P^{>2}, F\} = (\Pi_{\text{ker}} + \Pi_{>2} + \Pi_{>K}\Pi_{\text{rg}})\{P^{>2}, F\} + \Pi_{>K}P_{\text{rg}}.$$

Then we deduce that

$$\begin{aligned} \Pi_{\text{ker}}e^{adF}H &:= N_+ = N + \Pi_{\text{ker}}\{P^{>2}, F\} + \Pi_{\text{ker}}Q(P_{\text{rg}}), \\ \Pi_{\text{rg}}e^{adF}H &:= P_+^{\leq 2} = \Pi_{>K}(P_{\text{rg}} + \Pi_{\text{rg}}\{P^{>2}, F\}) + \Pi_{\text{rg}}Q(P_{\text{rg}}), \\ \Pi_{>2}e^{adF}H &:= P_+^{>2} = P^{>2} + \Pi_{>2}\{P^{>2}, F\} + \Pi_{>2}Q(P_{\text{rg}}) \end{aligned}$$

where  $Q(P_{\text{rg}})$ , is quadratic in  $P_{\text{rg}}$  and collects the terms from  $E_F(H)$  and  $\{F, P_{\text{rg}}\}$ .

## 4.2 The quantitative estimates

*In the course of our computations we will often have a statement for functions  $f, g$  of the following type:*

there is a constant  $C$  depending only on some of the parameters  $d, n, q, t$  with  $|f| \leq C|g|$  we then will replace this statement by the formula

$$|f| \ll_{d,n,q,t} |g|, \quad \text{or just} \quad |f| \ll |g|.$$

*Notice that the relation  $|f| \ll |g|$  satisfies all the usual properties of the relation  $\leq$ , that is it is a partial order compatible with sums and products.*

At a more quantitative level, we start by choosing a domain  $\mathcal{O}_0 \subset \mathcal{R}_\alpha \cap \varepsilon^2\Lambda$  where some appropriate quantitative non-degeneracy conditions (see Lemma 11) hold.

By construction the measure of  $\mathcal{O}_0$  is of the order of  $\varepsilon^{2n}$ . Then we choose two positive parameters  $\tau, K_0$ , bounded from below by some estimate depending only on  $n, d, q$  and the set  $\mathcal{O}_0$ . Such a choice determines bounds on  $\varepsilon, r$  through formula 23 of the type  $r\varepsilon^{-1} + \varepsilon^3r^{-1} < CK_0^{-\tau}$ . When these bounds hold our algorithm converges for all  $\xi$  in some set whose complement, in the starting domain  $\mathcal{O}_0$ , has measure of order  $\varepsilon^{2n}K_0^{-f(\tau)}$  where  $f$  is some linear increasing function of  $\tau$ , see Proposition 4.

Each Hamiltonian  $H$  is constrained by several parameters,  $s, r, \Theta, L, M, a, \vartheta, \mu, K$ , which control the quantitative structure. The frequency cut-off parameter  $K$  grows exponentially while the other parameters shall be proved to be *telescopic*.

**Definition 10** We say that a positive parameter  $b$  is telescopic if for each step  $m$  we have  $b_0/2 < b_m < \frac{3}{2}b_0$  (usually  $b_m$  is either an increasing or a decreasing sequence).

Note that a sufficient condition is that

$$\sum_{m=0}^{\infty} |b_{m+1} - b_m| \leq \frac{b_0}{2}. \quad (32)$$

Finally the domain  $\mathcal{O}$  shrinks at each step but its measure remains bounded from below.

**Definition 11** The parameters  $s, r$  control the decreasing domain  $D(s, r)$  of definition in the dynamical variables.

The parameters  $\vartheta, \mu, K$  control the quasi-Töplitz norm of  $P$  and of the quadratic Hamiltonian associated to the  $\tilde{\Omega}_t$ , and we set  $\|P\|_{\vec{p}}^T \leq \Theta$ .

The parameters  $M, L$  control the derivatives up to order  $\ell = (2d+1)^2$  of  $\omega$  and  $\Omega_t$ , the fact that  $\omega$  is invertible and an estimate on the dilation factor.

$$\begin{aligned} |\partial_\omega \xi|_\infty &\leq L\varepsilon^{-2(q-1)}, \quad |\omega - j^{(2)}|_\infty, |\Omega_t - |\mathbf{r}_t|^2 I_t|_\infty \leq M\varepsilon^{2q} \\ |\partial_\xi^\alpha \omega|_\infty + |\partial_\xi^\alpha \theta_t| + |\partial_\xi^\alpha \Omega_t^{\text{nil}}|_\infty + |\partial_\xi^\alpha \tilde{\Omega}_t|_\infty &\leq \varepsilon^{2(q-|\alpha|)} M^{|\alpha|}. \end{aligned} \quad (33)$$

for all  $\alpha \neq 0$  such that  $|\alpha| \leq \ell$ .

Finally the parameter  $a$  is defined by

$$\begin{aligned} |(\omega \cdot \nu)^{-1}| &\leq \varepsilon^{-2q} a, \quad |(\omega \cdot \nu + \Omega_t)^{-1}| \leq \varepsilon^{-2q} a, \\ |(\omega \cdot \nu + L(\Omega_t) + \sigma\sigma' R(\Omega_\nu))^{-1}| &\leq \varepsilon^{-2q} a, \quad \forall \nu : |\nu| \leq S_0 \end{aligned} \quad (34)$$

where  $S_0$  is some fixed positive number, see Lemma 11. Here  $\Omega_t$  is defined by (28) and  $L(\Omega_t)$ , resp.  $R(\Omega_\nu)$ , are the operators of left respectively right multiplication by the matrices  $\Omega_t$ , resp.  $\Omega_\nu$ , on  $d_t \times d_\nu$  matrices. Finally  $|\cdot|$  is the  $L^2$  operator norm.

We choose as in [14]:

$$\begin{aligned} r_{m+1} &= (1 - 2^{-m-3})r_m, \quad s_{m+1} = (1 - 2^{-m-3})s_m, \\ \vartheta_{m+1} &= \vartheta_m + \vartheta_0 2^{-m-2}, \quad \mu_{m+1} = \mu_m - \mu_0 2^{-m-2}, \quad K_m = 4^m K_0. \end{aligned} \quad (35)$$

This choice allows us to perform the Cauchy and ultraviolet estimates.

We choose any sequence of nested domains  $\mathcal{O}_m$  as follows<sup>6</sup>. The domain  $\mathcal{O}_{m+1}$  is a subset of  $\mathcal{O}_m$  where not only (31) admits a unique solution but moreover

$$\|F_m\|_{\vec{p}'_m}^T \leq \varepsilon^{-2q} K_m^\tau \|P_{\text{rg}}\|_{\vec{p}_m}^T, \quad (36)$$

where  $\vec{p}_m = (\lambda_m, s_m, r_m, \vartheta_m, \mu_m, \mathcal{O}_m)$ ,  $\vec{p}'_m = (\lambda_m, s'_m, r'_m, \vartheta'_m, \mu'_m, \mathcal{O}_{m+1})$  with  $b'_m = (b_m + b_{m+1})/2$ , for  $b = s, r, \vartheta, \mu$ , and  $\lambda_m = \varepsilon^2 M_m^{-1}$ .

**Proposition 3** For  $K_0, \tau$  large and for all  $\varepsilon, r$  such that  $r\varepsilon^{-1} + \varepsilon^3 r^{-1} \ll K_0^{-\tau}$ , at each step  $m$ , in the set  $\mathcal{O}_m$ , one has the estimate

$$\varepsilon^{-2q} \|P_{\text{rg}, m}\|_{\vec{p}_m}^T \leq e^{-\frac{3}{2}m}. \quad (37)$$

Moreover all the constants  $s, r, \Theta, L, M, a, \vartheta, \mu$  are telescopic. Finally our algebraic algorithm converges on the set  $\cap_m \mathcal{O}_m$ , and we obtain a Hamiltonian  $H_\infty$  which has reducible KAM tori.

<sup>6</sup> of course we will need to show that we can make non-empty choices

*Proof* These estimates are all standard and follow the same lines as the corresponding ones in [14]. The procedure is recursive, suppose we have reached some step  $m$  and computed all the necessary estimates, then using for instance Lemma 10.20 of [14], one proves the estimate (37) for  $F_m$ .

Then  $F_m$  defines a symplectic change of variables and one proves, see Proposition 10.16 and section 10.17 of [14], that in the new Hamiltonian  $H_{m+1}$  both the perturbation  $P_{m+1}$  and the quadratic Hamiltonian associated to the  $\tilde{\Omega}_t^{(m+1)}$  are well defined in the domain  $D(s_{m+1}, r_{m+1})$  and quasi-Töplitz with parameters  $(K_{m+1}, \vartheta_{m+1}, \mu_{m+1})$ . Note that by definition of  $E_F$  and since  $F$  solves the Homological equation we have that  $Q(P_{\text{rg},m})$  is quasi-Töplitz. The variation in the parameters  $s, r, \vartheta, \mu$  is due to *Cauchy estimates*, namely it is used in order to control the norm of the Poisson bracket of two Hamiltonians, following Proposition 10.16 of [14].

The corrections of the parameters  $L, M, a, \Theta$  are obtained from the coordinate change, which is very close to the identity due to the super-exponential decay of the norm of  $F$ , this implies easily the telescopic nature of the parameters used, as  $|b_{m+1} - b_m| < \text{const } b_0 e^{-(3/2)^m}$  see Lemma 10.20 of [14].

Then one easily sees that the algorithm converges.

By the nature of the majorant norm, for all  $j$ , the part of homogeneous degree  $j$  of  $H_m$  converges to the part of homogeneous degree  $j$  of  $H_\infty$ .

In particular, since  $P_{\text{rg},m}$  converges to 0 this implies that  $N_m$  and  $P_m^{>2}$  converge to  $N_\infty$  and  $P_\infty^{>2}$  respectively.  $\square$

The main problem is thus to exhibit some domain of parameters  $\xi$  where the estimates (36) hold and prove that at the end of the algorithm we are left with a set  $\mathcal{O}_\infty := \bigcap_m \mathcal{O}_m$  (of Cantor type) of positive measure (in particular non-empty) of order  $\varepsilon^{2n}$ .

#### 4.3 Measure estimates

For the remaining discussion we drop the indices  $m$  and assume we are working at some undefined step, with  $F, H, N$  etc. the hamiltonians at this given step. By remark 7 the linear operator  $F \mapsto \Pi_{\text{rg}}\{F, P^{>2}\}$  is nilpotent of order three so in order to invert the homological equation we only need to invert the operator  $F \mapsto \text{ad}(N)(F)$  which is of degree 0 hence block diagonal on the decomposition  $F^0 \oplus F^{0,1} \oplus F^2$ .

Indeed, by the formulas (28) and (27),  $\text{ad } N$  acts on  $F^0$  and on  $F^{1,0}$  as  $\omega \cdot \partial_x$  and on each block  $(\nu, \mathfrak{t})_\sigma$  of  $F^{0,1}$  as

$$\sigma(\omega \cdot \nu + \Omega_{\mathfrak{t}}). \quad (38)$$

Finally on each block  $(\nu, \mathfrak{t}, \mathfrak{t}')_{\sigma, \sigma'}$  of  $F^{0,2}$ , thought of as  $d_{\mathfrak{t}} \times d_{\mathfrak{t}'}$  matrices,  $\text{ad } N$  acts as

$$\sigma(\omega \cdot \nu + L(\Omega_{\mathfrak{t}}) + \sigma \sigma' R(\Omega_{\mathfrak{t}'})) \quad (39)$$

where the notations are those of (34).

In each case we have to show that these linear operators (which we represent as matrices of dimension resp. one,  $d_{\mathfrak{t}}$  and  $d_{\mathfrak{t}'}$ ) are invertible and moreover estimate the norm (we choose the  $L^2$  operator norm) of the inverse. This will be done, when

the matrices are self-adjoint (i.e. both  $\mathfrak{t}$  and  $\mathfrak{t}'$  in  $\mathfrak{T}_g$ ), by lower bound estimates on the eigenvalues. By momentum conservation, the remaining set of matrices is finite and we estimate it by lower bounds on the determinants. We impose as constraint on the parameters  $\xi$  to have estimates of the form:

$$|(\omega \cdot \nu)^{-1}|, |(\omega \cdot \nu + \Omega_{\mathfrak{t}})^{-1}|, |(\omega \cdot \nu + L(\Omega_{\mathfrak{t}}) + \sigma\sigma'R(\Omega_{\mathfrak{t}'}) )^{-1}| \leq \varepsilon^{-2q} K^\varrho \quad (40)$$

for all  $\nu$  such that  $|\nu| \leq K$  and for some ( $K$  independent) value of  $\varrho$ .

**Lemma 1** *Fixing  $\varrho = \varrho(\tau) := (\tau - 1)/3((2d + 1)^2 + 4)$ , in the set of the  $\xi$  for which (40) holds one has the estimate (36).*

*Proof* The result follows directly from Lemma 14. We apply it with  $s, r = (s_m, r_m)$ ,  $\delta = 2^{-m-5}$ ,  $P \rightsquigarrow P_m$ ,  $F \rightsquigarrow F_m$  and  $K \rightsquigarrow K_{m+1} = 4^{m+1}K_0$ . Then by our choices  $\delta^{-2} \leq K$  (provided  $K_0 \geq 2^9$ ).  $\square$

**Definition 12** We choose  $\mathcal{O}_{m+1}$  to be the set of  $\xi \in \mathcal{O}_m$  for which (40) holds with  $\varrho = (\tau - 1)/3((2d + 1)^2 + 4)$ ,  $\omega \rightsquigarrow \omega_m$  and  $\Omega \rightsquigarrow \Omega_m$  for all  $\mathfrak{t}, \mathfrak{t}'$ .

**Proposition 4** *i) The measure of  $\mathcal{O}_m \setminus \mathcal{O}_{m+1}$  is bounded by  $\varepsilon^{2n} K_{m+1}^{-\frac{\tau}{b} + n + 1}$  with  $b = 3(4d)^{d+4}$ .*

*ii) Provided that  $\tau \geq b(n + 3)$ , we have  $|\mathcal{O}_\infty| \geq |\mathcal{O}_0| - \varepsilon^{2n} K_0^{-\frac{\tau}{b} + n + 2}$ , and thus is a parameter set of positive measure by taking  $K_0$  large.*

The proof of this Proposition occupies the rest of this section and concludes the proof of Theorem 1.

*Proof (Proof of Theorem 1)* By Proposition 3 we have reducible KAM tori for all  $\xi \in \mathcal{O}_\infty = \bigcap_m \mathcal{O}_m$  for all choices of the sets  $\mathcal{O}_m$  for which one has (36). By Lemma 1 the choice done in Definition 12 satisfies this requirement. Then Proposition 4 ensures that  $\mathcal{O}_\infty$  has positive measure provided  $K_0$  is large enough.  $\square$

In order to prove Proposition 4 we drop the index  $m$  and first give some notation.

**Definition 13** [Resonant sets] For all  $|\nu| \leq K$ ,  $\mathfrak{t}, \mathfrak{t}' \in \mathfrak{T}$  we define the *resonant* sets  $\mathfrak{R}_{\nu, \mathfrak{t}, \mathfrak{t}', \sigma, \sigma'}^\varrho$  with  $\sigma, \sigma' = 0, \pm 1$  as the sets of  $\xi \in \mathcal{O}$  where

$$|(\omega \cdot \nu + \sigma L(\Omega_{\mathfrak{t}}) + \sigma' R(\Omega_{\mathfrak{t}'}) )^{-1}| > \varepsilon^{-2q} K^\varrho. \quad (41)$$

Note that this formula includes all cases of the complement of the region given by Formula (40). When one (or both) of the  $\sigma$  equals zero then we drop it and the corresponding symbol  $\mathfrak{t}$ , for instance  $\mathfrak{R}_{\nu, \mathfrak{t}, \mathfrak{t}', \sigma, 0}^\varrho = \mathfrak{R}_{\nu, \mathfrak{t}, \mathfrak{t}'}^\varrho$ .

*Proof (Proof of Proposition 4)* Part ii) of Proposition 4, follows trivially from part i) since  $K_m = 4^m K_0$  and  $|\mathcal{O}_0| \sim \varepsilon^{2n}$ . Indeed by hypothesis  $-\frac{\tau}{b} + n + 1 \leq -2$  and hence  $|\mathcal{O}_\infty| \geq |\mathcal{O}_0| - \sum_m |\mathcal{O}_{m-1} \setminus \mathcal{O}_m| \geq$

$$\geq |\mathcal{O}_0| - \varepsilon^{2n} K_0^{-\frac{\tau}{b} + n + 1} \sum_m 4^{-m} \geq |\mathcal{O}_0| - \varepsilon^{2n} K_0^{-\frac{\tau}{b} + n + 2}.$$

In order to prove part i), the first thing (Lemma 2) is to show that for all choices of  $\varrho$  and for any  $|\nu| < K$ , and  $\mathfrak{t}, \mathfrak{t}', \sigma, \sigma'$ , we may impose the corresponding estimate

(40), by only removing, from the domain  $\mathcal{O}$ , a resonant set of parameters  $\mathfrak{R}_{\nu, \mathfrak{t}, \mathfrak{t}', \sigma, \sigma'}^{\varrho}$  of measure  $\leq C\varepsilon^{2n}K^{-c(\varrho-1)}$  for some  $C, c$  depending only on  $q, d, n$ .

Then we evaluate the measure of the union of resonant sets. Since  $\mathfrak{t}, \mathfrak{t}'$  vary in an infinite set it is not sufficient to simply count the resonant sets. In Lemma 3 we prove that for all  $\varrho$  and for all  $\nu, \mathfrak{t}, \mathfrak{t}', \sigma, \sigma'$  such that  $|\nu| \leq K$  and  $\sigma\sigma' \neq -1$  one has that the non-empty sets  $\mathfrak{R}_{\nu, \mathfrak{t}, \mathfrak{t}', \sigma, \sigma'}^{\varrho}$  belong to a finite list  $G_K$  with cardinality  $|G_K| < K^p$ . Hence we need to remove a set of measure  $\leq C\varepsilon^{2n}K^{-c(\varrho-1)}K^p$  where we choose  $c$  to be the minimum over the possible  $c$  and  $C$  the maximum.

In the case of the second Melnikov condition with  $\sigma\sigma' = -1$  we proceed in a slightly subtler way. We use the quasi-Töplitz property which is defined in terms of (among others) a free parameter  $\varrho_0$ , then we prove, in Lemma 4, that fixing  $\varrho_{d+2} = (4d)^{d+2}\varrho_0$  there exists a set  $R$  of measure at most  $\varepsilon^{2n}K^{-\varrho_0+n}$  such that for all  $\nu, \mathfrak{t}, \mathfrak{t}'$ :

$$\mathfrak{R}_{\nu, \mathfrak{t}, \mathfrak{t}', \sigma, -\sigma}^{\varrho_{d+2}} \subset R.$$

In conclusion fixing  $\varrho_0$  so that  $\varrho_{d+2} = \varrho$  we have proved that, provided  $\varrho$  is sufficiently large, by removing a small measure resonant set we may impose all the Melnikov conditions (40). We then verify (Corollary 1) that the measure is of order  $\varepsilon^{2n}K^{-\varrho/(4d)^{d+2}+n+1}$ . When we substitute  $\varrho = \varrho(\tau)$  we easily obtain item *i*) of Proposition 4.  $\square$

**Lemma 2** For  $|\nu| \leq K, \mathfrak{t}, \mathfrak{t}' \in \mathfrak{T}, \sigma, \sigma' = 0, \pm 1$ , we have:

$$|\mathfrak{R}_{\nu, \mathfrak{t}, \mathfrak{t}', \sigma, \sigma'}^{\varrho}| \leq C\varepsilon^{2n}K^{-c(\varrho-1)}, \quad (42)$$

where  $c = 1$  if either  $\sigma = \sigma' = 0$  or if  $\sigma' = 0$  and  $\mathfrak{t} \in \mathfrak{T}_g$  or if  $\mathfrak{t}, \mathfrak{t}' \in \mathfrak{T}_g$ . If  $\sigma' = 0$  and  $\mathfrak{t} \in \mathfrak{T}_f$  then  $c = d_{\mathfrak{t}}^{-1}$ , otherwise  $c = (d_{\mathfrak{t}}d_{\mathfrak{t}'})^{-1}$ .

*Proof* We give a brief overview of the proof, the details are in the appendix.

The first remark is that all the resonant sets with  $|\nu| \leq S_0$  are empty since formula (34) implies (40) provided that  $K_0^p$  is large enough.

Moreover if  $|u| := |\omega \cdot \nu + \sigma|\mathbf{r}_{\mathfrak{t}}|^2 + \sigma'|\mathbf{r}_{\mathfrak{t}'}|^2| > 2(2d+1)^2M\varepsilon^{2q}$  then the bounds (40) hold trivially and the resonant sets are empty.

Indeed, by (33), we have  $|\Omega_{\mathfrak{t}} - |\mathbf{r}_{\mathfrak{t}}|^2|_{\infty} \leq M\varepsilon^{2q}$  and we have to invert some matrix of the form  $u + Z$  with  $Z$  a matrix of size  $p \leq (2d+1)^2$  deduced from  $\Omega_{\mathfrak{t}}$  in such a way that also  $|Z|_{\infty} \leq M\varepsilon^{2q}$ . Since the operator norm of  $Z$  is bounded by  $pM\varepsilon^{2q}$  if  $2|Z|_2 \leq 2pM\varepsilon^{2q} \leq |u|$  we invert in Neumann series and obtain the estimate  $|(u + Z)^{-1}|_2 \leq 2|u|^{-1} \leq \text{const } \varepsilon^{-2q}$ . So we only need to consider  $S_0 \leq |\nu| \leq K$  and given  $\nu, \mathfrak{t}, \mathfrak{t}', \sigma, \sigma'$  we only need to work in the subset of  $\omega$  where

$$|\omega \cdot \nu + \sigma|\mathbf{r}_{\mathfrak{t}}|^2 + \sigma'|\mathbf{r}_{\mathfrak{t}'}|^2| < 2(2d+1)^2M\varepsilon^{2q}. \quad (43)$$

*Remark 8* The condition  $|\nu| > S_0$  is used so that one can perform the measure estimates not on the set  $\mathcal{O}$  but on its image  $\tilde{\mathcal{O}}$  under  $\omega$ , by multiplying by the dilation constant. For this, an upper bound for the absolute value of the determinant of the Jacobian of the inverse map is estimated, by Formula (33), as  $L^n \varepsilon^{-2n(q-1)}$ . This is justified by two remarks, first  $\tilde{\mathcal{O}}$  is contained in a hypercube of length  $2\varepsilon^{2q}$  and second the dilation constant is uniformly bounded in both ways.

Then the basis of our measure estimates is Lemma 7 which shows how to control the resonant set  $|f(\omega)| \leq \alpha$  of a  $C^k$  function  $f$  provided that one has a uniform lower bound on some derivative  $|\partial_\omega^h f|$  with  $|h| \leq k$ .

We now discuss each of the three inequalities of (40) separately, the technical parts are in the appendix.

On  $F^0$  (i.e.  $\sigma = \sigma' = 0$ ) the estimate is classical, see the appendix.

On  $F^{0,1}$  we treat separately  $\mathfrak{t} \in \mathfrak{F}_f$  and  $\mathfrak{t} \in \mathfrak{F}_g$ . In the second case, since by hypothesis the  $\Omega_{\mathfrak{t}}$  are self-adjoint one uses the typical KAM estimates on eigenvalues:

$$|\omega \cdot \nu + \lambda_{\mathfrak{t}}^{(i)}| \geq \varepsilon^{2q} K^{-\varrho}, \quad \forall \lambda_{\mathfrak{t}}^{(i)} \in \text{spec}(\Omega_{\mathfrak{t}}), \quad (44)$$

which are equivalent to the bounds (40) on the operator norm of the inverse matrix. To estimate the measure of the resonant set of this case, one just passes to the variables  $\omega$  and then shows that the derivative of the small divisor is bounded from below by  $S_0/n$ , see the appendix.

Regarding the  $\mathfrak{t} \in \mathfrak{F}_f$ , we cannot bound the  $L^2$  norm with the eigenvalues since they do not dominate the  $L^2$  norm and moreover may not be regular functions of  $\xi$ .

We can obtain the bound (40), by requiring

$$|\det(\omega \cdot \nu + \Omega_{\mathfrak{t}})| \geq \varepsilon^{2qd_{\mathfrak{t}}} K^{-\varrho+1} \quad (45)$$

here  $d_{\mathfrak{t}}$  is the dimension of the matrix  $\Omega_{\mathfrak{t}}$ .

The  $L_\infty$  norm of the matrix we are inverting is estimated as  $\ll M\varepsilon^{2q}$  by the two bounds (43) and (33). By Cramer's rule, we have that (45) implies the bounds (40) provided  $K_0$  is large enough (we use the exponent  $\varrho - 1$  in order to absorb the constants). Thus the measure of  $\mathfrak{R}_{\nu, \mathfrak{t}, \sigma}$  is dominated by the measure of the resonant set where (45) does not hold.

In this case again we obtain the measure estimates by a lower bound on some derivative. We need to perform  $d_{\mathfrak{t}}$  derivatives, thus we use the fact that all our functions are at least  $C^{d_{\mathfrak{t}}}$ .

We are left with the *second Melnikov conditions*  $\sigma, \sigma' = \pm 1$ . If at least one of the couple  $\mathfrak{t}, \mathfrak{t}'$  belongs to  $\mathfrak{F}_f$ , following the same reasoning as in the case of  $\sigma' = 0$ , we impose

$$|\det(\omega \cdot \nu + L(\Omega_{\mathfrak{t}}) + \sigma\sigma'R(\Omega_{\mathfrak{t}'}))| \geq \varepsilon^{2qd_{\mathfrak{t}}d_{\mathfrak{t}'}} K^{-\varrho+1} \quad (46)$$

By Cramer's rule, since  $|\omega \cdot \nu + \sigma|\mathbf{r}_{\mathfrak{t}}|^2 + \sigma'|\mathbf{r}_{\mathfrak{t}'}|^2|$  is small and by the second bound on (33), we have that (45) implies the bounds (40) provided  $K_0$  is large enough. This means that the measure of  $\mathfrak{R}_{\nu, \mathfrak{t}, \sigma}$  is dominated by the measure of the resonant set where (45) does not hold.

In this case in order to obtain the measure estimates, by a lower bound on some derivative, we need to perform  $d_{\mathfrak{t}}d_{\mathfrak{t}'}$  derivatives, thus we use the fact that all our functions are at least  $C^{d_{\mathfrak{t}}d_{\mathfrak{t}'}}$ .

When both  $\mathfrak{t}, \mathfrak{t}' \in \mathfrak{F}_g$  then the matrices we need to control are self-adjoint and we bound their inverse by controlling the eigenvalues, namely we need a condition

$$\begin{aligned} |\omega \cdot \nu + \lambda_{\mathfrak{t}}^{(i)} + \sigma\lambda_{\mathfrak{t}'}^{(j)}| &\geq \varepsilon^{2q} K^{-\varrho}, \\ \forall \sigma = \pm, \lambda_{\mathfrak{t}}^{(i)} \in \text{spec}(\Omega_{\mathfrak{t}}), \lambda_{\mathfrak{t}'}^{(j)} \in \text{spec}(\Omega_{\mathfrak{t}'}), \end{aligned} \quad (47)$$

This measure estimate is done just as in formula (44), indeed one just passes to the variables  $\omega$  and then shows that the derivative of the small divisor is bounded from below by  $S_0/n$ , see the appendix.  $\square$

**Lemma 3** *For all  $\varrho$  and for all  $\nu, \mathbf{t}, \mathbf{t}', \sigma, \sigma'$  such that  $|\nu| \leq K$  and  $\sigma\sigma' \neq -1$  one has that the non-empty sets  $\mathfrak{R}_{\nu, \mathbf{t}, \mathbf{t}', \sigma, \sigma'}^e$  belong to a finite list  $G_K$  with cardinality  $|G_K| < K^p$  with  $p = n + \frac{d}{2} + 1$ .*

*Proof* For the sets  $\mathfrak{R}_{\nu, \mathbf{t}, \sigma}^e$  we recall the conservation of momentum  $\mathbf{r}_{\mathbf{t}} = -\pi(\nu)$ , then a set is determined by  $|\nu| < K$  and  $\sigma = 0, \pm 1$ . We have at most  $3(2K)^n$  sets of this type.

In the sets  $\mathfrak{R}_{\nu, \mathbf{t}, \mathbf{t}', \sigma, \sigma'}^e$  we preliminarily notice that there are a finite number of cases where at least one of the pair  $\mathbf{t}, \mathbf{t}'$  belongs to  $\mathfrak{T}_f$ . Indeed let us suppose that  $\mathbf{t} \in \mathfrak{T}_f$ . Then  $|\mathbf{r}_{\mathbf{t}}| < C_0 \ll K_0$ ,  $\mathbf{r}_{\mathbf{t}'}$  is fixed by momentum conservation,  $\theta_{\mathbf{t}'}$  is chosen in the finite list  $\Upsilon$  (see formula (69)) and the pair identifies the index  $\mathbf{t}'$ . Consequently we have at most  $|\Upsilon|^2 (2C_0)^d (2K)^n$  sets of this type.

We are left with  $\mathfrak{R}_{\nu, \mathbf{t}, \mathbf{t}', \sigma, \sigma'}^e$  with both  $\mathbf{t}, \mathbf{t}' \in \mathfrak{T}_g$ . We notice that

$$\lambda_{\mathbf{t}}^{(i)} = |\mathbf{r}_{\mathbf{t}}|^2 + \theta_{\mathbf{t}} + \tilde{\lambda}_{\mathbf{t}}^{(i)}, \quad (48)$$

then if  $\sigma'\sigma = +$  the left hand side of (47) is trivially  $\geq 1/2$  for all  $\xi \in \mathcal{O}$ , if  $|\mathbf{r}_{\mathbf{t}}|^2 + |\mathbf{r}_{\mathbf{t}'}|^2 \geq 4K \geq 2|\omega||\nu|$  and we only need to consider a finite number of  $\mathbf{t}$  such that  $|\mathbf{r}_{\mathbf{t}}|^2 \leq 4K$ . Consequently we have at most  $|\Upsilon|^2 (4K)^{d/2} (2K)^n$  non-empty sets of this type.  $\square$

Treating the case  $\sigma'\sigma = -$  is a well-known problem in the study of NLS on  $\mathbb{T}^d$  which is overcome by exploiting the *quasi-Töplitz* property of the NLS Hamiltonian.

As explained in Appendix B, given  $K$  we can construct a stratification  $\Sigma^{(K)}$  of  $\mathbb{Z}^d$  with the following properties.

- Proposition 5**
- i) *To each stratum  $\Sigma$  is associated an order parameter  $\varrho_{\Sigma} \geq \varrho_0$  which can be taken from the finite list  $\varrho_i = (4d)^i \varrho_0$ ,  $i = 1, \dots, d+1$ .*
  - ii) *For each  $\mathbf{t} \in \mathfrak{T}_g$ ,  $\mathbf{D}_{\mathbf{t}}$  belongs to a set  $\Sigma_{\mathbf{t}}$  union of parallel strata each meeting  $\mathbf{D}_{\mathbf{t}}$  in a single point, all of codimension  $\ell \leq d_{\mathbf{t}} - 1$  whose union is a union of translates  $\mathbf{D}_{\mathbf{t}'} = \mathbf{D}_{\mathbf{t}} + u$  and with the same cut  $\varrho_{\Sigma}$ , moreover the eigenvalue  $\theta(k)$ ,  $k \in \Sigma_{\mathbf{t}}$  depends only on  $\mathbf{t}$ .*
  - iii) *If  $\ell = 0$  one has that the corresponding matrix (39) is invertible with the bounds (40).*
  - iv) *If  $\ell \neq 0$  but  $\pi(\nu) \cdot v$  is not constant on each stratum again the corresponding matrix (39) is invertible with the bounds (40).*

*Proof* i) The order parameter is that of the cut  $\varrho_j$  associated to the points of  $\Sigma$ .  
ii) Each  $\mathbf{D}_{\mathbf{t}}$  determines in the initial stratification a stratum for each of its points of codimension  $d_{\mathbf{t}} - 1$ . The stratification  $\Sigma^{(K)}$  is a refinement of the stratification  $\Sigma_0$  of Proposition 1 so the properties follow from that. iii) If  $\ell = 0$  this means that for every  $v \in B_K$  we have  $|v \cdot \mathbf{r}_{\mathbf{t}}| > 4K^{4d\varrho_0}$ , in particular we have  $\pi(\nu) \in B_K$  so  $|\pi(\nu) \cdot \mathbf{r}_{\mathbf{t}}| > 4K^{4d\varrho_0}$ . iv) In this case by definition we have  $|\pi(\nu) \cdot \mathbf{r}_{\mathbf{t}}| > 4K^{4d\varrho_{\Sigma}} \geq 4K^{4d\varrho_0}$ . So in both cases we need to observe that the matrix  $(\omega \cdot \nu + L(\Omega_{\mathbf{t}}) - R(\Omega_{\mathbf{t}'})$  has integral part  $\mathbf{j}^{(2)} \cdot \nu + |\mathbf{r}_{\mathbf{t}}|^2 - |\mathbf{r}_{\mathbf{t}'}|^2 = \mathbf{j}^{(2)} \cdot \nu - |\pi(\nu)|^2 - 2\pi(\nu) \cdot \mathbf{r}_{\mathbf{t}}$  since by conservation of momentum  $\pi(\nu) + \mathbf{r}_{\mathbf{t}} - \mathbf{r}_{\mathbf{t}'} = 0$ .

Now, if  $|\pi(\nu) \cdot \mathbf{r}_t| > 4K^{4d\varrho_0}$ , we have  $|\pi(\nu)|^2 + \pi(\nu) \cdot \mathbf{r}_t - \mathbf{j}^{(2)} \cdot \nu \geq |\pi(\nu) \cdot \mathbf{r}_t| - |\pi(\nu)|^2 - |\mathbf{j}^{(2)}| |\pi(\nu)| > 4K^{4d\varrho_0} - 2K^2$  is a large positive integer and since the remainder is small with  $\varepsilon^{2q}M$  by formula (33), we can invert the matrix for all  $\xi \in \mathcal{O}_m$  by Neumann series with a good estimate.  $\square$

For each stratum  $\Sigma$  of the stratification of  $\mathfrak{T}_s$ , given in Remark 17, we choose *one representative*  $\mathbf{t}_\Sigma$  (by fixing  $\mathbf{D}_{\mathbf{t}_\Sigma}$ ).

Then for each  $|\nu| \leq K$  such that  $\pi(\nu) \cdot x$  is constant for  $x$  on the stratum, we proceed as follows:

- For each  $\Sigma$  we determine (if it exists) the unique  $\mathbf{t}'$  so that  $(\mathbf{r}_{\mathbf{t}'}, \theta_{\mathbf{t}'}) = (\mathbf{r}_{\mathbf{t}_\Sigma} + \pi(\nu), \theta')$ ;
- if  $\mathbf{t}'$  exists we impose

$$|(\omega \cdot \nu + L(\Omega_{\mathbf{t}_\Sigma}) + \sigma\sigma' R(\Omega_{\mathbf{t}'})^{-1})| \leq \varepsilon^{-2q} K^{2d\varrho_\Sigma}.$$

The complementary of this set is by definition  $\mathfrak{R}_{\nu, \mathbf{t}_\Sigma, \mathbf{t}'}$  and has measure of order  $\varepsilon^{2n} K^{-2d\varrho_\Sigma}$ , by Lemma 2. Now by Lemma 15 we have less than  $K^{(2d-2)\varrho_\Sigma}$  strata of this type so setting

$$R := \bigcup_{|\nu| \leq K} \bigcup_{\Sigma, \theta'} \mathfrak{R}_{\nu, \mathbf{t}_\Sigma, \mathbf{t}'}^{2d\varrho_\Sigma} \quad (49)$$

we have

$$\varepsilon^{-2n} K^{-n} |R| \leq \sum_{\Sigma, \theta'} K^{-2d\varrho_\Sigma} \leq |\mathcal{I}| \sum_{j=0}^{d+1} K^{-2d\varrho_j} K^{(2d-2)\varrho_j} \leq K^{-\varrho_0}.$$

**Lemma 4** For all  $\nu, \mathbf{t}, \mathbf{t}'$  we have  $\mathfrak{R}_{\nu, \mathbf{t}, \mathbf{t}', \sigma, -\sigma}^{\varrho_{d+2}} \subset R$ .

*Proof* For each stratum  $\Sigma$  of  $\mathfrak{T}_s$  we have chosen a representative  $\mathbf{t}_\Sigma$ . Now for all  $\nu$  for which a small divisor may occur we have for the scalar products  $\pi(\nu) \cdot \mathbf{r}_{\mathbf{t}_\Sigma} = \pi(\nu) \cdot \mathbf{r}_t$  for all  $\mathbf{t}$  in this stratum. It then follows that

$$|\Omega_{\mathbf{t}} - \Omega_{\mathbf{t}_\Sigma}|_\infty = |\tilde{\Omega}_{\mathbf{t}} - \tilde{\Omega}_{\mathbf{t}_\Sigma}|_\infty \leq \varepsilon^{2q} M K^{-4d\varrho_\Sigma}, \quad \forall \mathbf{t} \in \Sigma$$

by applying Lemma 16. Consequently for each pair  $\mathbf{t}, \mathbf{t}'$  with  $\mathbf{t} \in \Sigma, \mathbf{t}' \in \Sigma'$  we have

$$|(\omega \cdot \nu + L(\Omega_{\mathbf{t}}) - R(\Omega_{\mathbf{t}'})^{-1})| = |(\omega \cdot \nu + L(\Omega_{\mathbf{t}_\Sigma}) - R(\Omega_{\mathbf{t}'_\Sigma}))^{-1}| |(1+A)^{-1}|$$

with

$$A = (\omega \cdot \nu + L(\Omega_{\mathbf{t}_\Sigma}) - R(\Omega_{\mathbf{t}'_\Sigma}))^{-1} (\Omega_{\mathbf{t}} - \Omega_{\mathbf{t}_\Sigma}), \quad |A| \leq M K^{-2d\varrho_\Sigma}$$

so that

$$|(\omega \cdot \nu + L(\Omega_{\mathbf{t}}) - R(\Omega_{\mathbf{t}'})^{-1})| \leq 2\varepsilon^{-2q} K^{2d\varrho_\Sigma} \leq \varepsilon^{-2q} K^{2d\varrho_\Sigma + 1}$$

In conclusion, since  $\varrho_\Sigma \leq \varrho_{d+1}$ , we have proved our claim.  $\square$

**Corollary 1** For  $\varrho$  large enough, the measure of the union of all resonant sets  $\mathfrak{R}_{\nu, \mathfrak{t}, \nu', \sigma, \sigma'}^{\varrho}$  with  $|\nu| \leq K$  is bounded by

$$\varepsilon^{2n} K^{-\frac{\varrho}{(4d)^{d+2}} + n + 1}$$

*Proof* This follows by Lemmas 2 3 and 4. Indeed we fix  $\varrho_{d+2} = \varrho$  so that  $\varrho_0 = \varrho/(4d)^{d+2}$ . Then the Lemmas imply that we need to remove a region of order

$$K^{-\frac{\varrho}{(2d+1)^2} + n + \frac{d}{2} + 1} + K^{-\frac{\varrho}{(4d)^{d+2}} + n} \leq K^{-\frac{\varrho}{(4d)^{d+2}} + n + 1},$$

provided  $\varrho$  and  $K$  are both large.

## 5 The normal form

We will work with many quadratic Hamiltonians in the variables  $w$  (thought of as a row vector). We represent a quadratic Hamiltonian in  $\mathcal{F}$  by a matrix  $F$  as

$$\mathcal{F}(w) = \frac{1}{2} w \cdot w J F^t = -\frac{1}{2} w F J w^t, \quad (50)$$

where  $J := -i\{w^t, w\}$  is the standard matrix of the symplectic form.

By explicit computation, and under simple genericity conditions, we have:

$$\mathcal{N} = \omega(\xi) \cdot y + \sum_{k \in S^c} |k|^2 |z_k|^2 + \mathcal{Q}(\xi; x, w), \quad \omega_i(\xi) := |j_i|^2 + \omega_i^{(1)}(\xi) \quad (51)$$

here  $\mathcal{Q}(\xi; x, w)$  is a quadratic Hamiltonian in the variables  $w$  with coefficients trigonometric polynomials in  $x$  given by Formula (30) of [13]. The frequency modulation  $\omega^{(1)}$  is homogeneous of degree  $q$  in  $\xi$  and given by the following explicit formula. We introduce

$$A_r(\xi_1, \dots, \xi_m) := \sum_{\sum_i k_i = r} \binom{r}{k_1, \dots, k_m}^2 \prod_i \xi_i^{k_i}. \quad (52)$$

and we have

$$\omega^{(1)}(\xi) = \nabla_{\xi} A_{q+1}(\xi) - (q+1)^2 A_q(\xi) \mathbf{1}. \quad (53)$$

By applying the results of [13] we decompose this very complicated infinite dimensional quadratic Hamiltonian into infinitely many decoupled finite dimensional systems reduced to constant coefficients by an explicit symplectic change of variables.

Since this construction is needed in the following we recall quickly Theorem 1 of [13].

**Theorem 2** [Theorem 1 of [13]] For all generic choices  $S = \{j_1, \dots, j_n\} \in \mathbb{Z}^{nd}$  of the tangential sites, there exists a phase shift map  $L : S^c \rightarrow \mathbb{Z}^n$ ,  $L : k \mapsto L(k)$ ,  $|L(k)| \leq 4qd$  such that the analytic symplectic change of variables:

$$\Psi : z_k = e^{-iL(k) \cdot x} z'_k, \quad y = y' + \sum_{k \in S^c} L(k) |z'_k|^2, \quad x = x', \quad (54)$$

from  $D(s, r/2) \rightarrow D(s, r)$  has the property that  $\mathcal{N}$  in the new variables:

$$\mathcal{N} \circ \Psi = \omega(\xi) \cdot y' + \sum_{k \in S^c} \tilde{\Omega}_k |z'_k|^2 + \tilde{Q}(w'), \quad (55)$$

has constant coefficients, where  $\omega(\xi)$  is defined in (51) and furthermore:

i) **Non-degeneracy** The map  $(\xi_1, \dots, \xi_m) \mapsto (\omega_1(\xi), \dots, \omega_m(\xi))$  is a local diffeomorphism for  $\xi$  outside a homogeneous real algebraic hypersurface.

ii) **Asymptotic of the normal frequencies:** We have  $\tilde{\Omega}_k = |k|^2 + \sum_i |j_i|^2 L^{(i)}(k)$ .

iii) **Reducibility:** The matrix  $\tilde{Q}(\xi)$  (see formula (50)) of the quadratic form  $\tilde{Q}(\xi, w')$  (see formula (64)) depends only on the variables  $\xi$ , its entries are homogeneous of degree  $q$  in these variables. It is block-diagonal and satisfies the following properties:

All of the blocks except a finite number are self adjoint of dimension  $\leq d+1$ . The remaining finitely many blocks have dimension  $\leq 2d+1$ .

All the (infinitely many) blocks are described, through Formula (50), by a finite list of matrices  $\mathcal{M}(\xi)$ .

iv) **Smallness:** If  $\varepsilon^3 < r < c_1\varepsilon$ , the perturbation  $\tilde{P} := P \circ \Psi$  is small, more precisely we have the bounds:

$$\|X_{\tilde{P}}\|_{s,r}^\lambda \leq C(\varepsilon^{2q-1}r + \varepsilon^{2q+3}r^{-1}), \quad (56)$$

where  $C$  is independent of  $r$  and depends on  $\varepsilon, \lambda$  only through  $\lambda/\varepsilon^2$ .

Regarding  $\omega^{(1)}$  given by Formula (53) we will need the following

**Lemma 5** For each  $i$  the polynomial  $\omega_i^{(1)}(\xi)$  equals  $-q(q+1)\xi_i^q$  plus terms which contain at least two variables.

*Proof* We do it for  $i=1$  by symmetry. From  $\nabla_\xi A_{q+1}(\xi)$  the terms  $\xi_j^q$  are obtained by

$$\partial_{\xi_1} (\xi_1^{q+1} + (q+1)^2 \sum_{j \neq 1} \xi_j^q \xi_1) = (q+1)\xi_1^q + (q+1)^2 \sum_{j \neq 1} \xi_j^q.$$

From the second summand we subtract  $(q+1)^2 \sum_j \xi_j^q$  getting the desired formula.  $\square$

## 5.1 The matrix blocks and the geometric graph $\Gamma_S$

The quadratic Hamiltonian  $\tilde{Q}$  is described, as we shall see in Formula (64), in terms of a 2-colored marked graph  $\Gamma_S$  with vertices in  $\mathbb{Z}^d$  and labels in  $\mathbb{Z}^n$  which encodes the possible interactions between the normal frequencies  $k \in S^c$ . In order to describe this in a combinatorial way we recall some notation from [13], see Definition 8 for the meaning of the maps  $\eta, \pi, \pi^{(2)}$ .

**Definition 14 (edges)** Consider a finite set  $X$  of elements in  $\mathbb{Z}^n$ .

$$X := \left\{ \ell = \sum_{i=1}^n \ell_i e_i, \ell_i \in \mathbb{Z} \quad \ell \neq 0, -2e_i \forall i, \quad \eta(\ell) \in \{0, -2\} \right\}. \quad (57)$$

Notice the *mass constraint*  $\sum_i \ell_i = \eta(\ell) \in \{0, -2\}$ . We call all these elements respectively the *black*,  $\eta(\ell) = 0$  and *red*  $\eta(\ell) = -2$  *edges* and denote them by  $X^{(0)}, X^{(-2)}$  respectively.

Each edge carries a *quadratic energy* (which is an integer):

$$K(\ell) := \frac{1 + \eta(\ell)}{2} (|\pi(\ell)|^2 + \pi^{(2)}(\ell)), \quad \frac{1 + \eta(\ell)}{2} = \pm \frac{1}{2}.$$

Of particular interest for the  $q$ -NLS is

$$X_q := \{\ell := \sum_{j=1}^{2q} \pm e_{i_j} = \sum_{i=1}^m \ell_i e_i, \quad \ell \neq 0, -2e_i \forall i, \quad \eta(\ell) \in \{0, -2\}\}. \quad (58)$$

Choose now a finite set of integral vectors  $S := \{j_1, \dots, j_n\}$  in  $\mathbb{Z}^d$ . For the NLS these are the tangential sites but the construction is purely geometric.

**Definition 15** Given  $\ell \in X^{(0)}$  denote by  $\mathcal{P}_\ell$  the set of pairs  $h, k \in \mathbb{Z}^d$  satisfying:

$$\sum_{j=1}^n \ell_j j_j + h - k = 0, \quad \sum_{j=1}^n \ell_j |j_j|^2 + |h|^2 - |k|^2 = 0 \quad \ell \in X^{(0)}. \quad (59)$$

If  $\ell \in X^{(-2)}$  we denote by  $\mathcal{P}_\ell$  the set of unordered pairs  $\{h, k\}$ ,  $h, k \in \mathbb{Z}^d$  satisfying:

$$\sum_{j=1}^n \ell_j j_j + h + k = 0, \quad \sum_{j=1}^n \ell_j |j_j|^2 + |h|^2 + |k|^2 = 0 \quad \ell \in X^{(-2)}. \quad (60)$$

It is convenient to formalise this definition in the language of graphs with two types of edges corresponding to the two conditions. When  $h, k$  satisfy formula (59) we join them by an oriented *black* edge, marked  $\ell$ , with source  $h$  and target  $k = h + \sum_{j=1}^m \ell_j j_j$ . Formula (60) is symmetric and we join  $h, k$  by an unoriented *red* edge marked  $\ell$ .

Note that the two conditions have a geometric meaning expressed through the quadratic energy. The first means that  $h$  lies on the hyperplane

$$H_\ell : \quad x \cdot \pi(\ell) = K(\ell) \quad (61)$$

while  $k$  lies on the parallel hyperplane  $H_{-\ell}$ .

The second condition means that  $h, k$  are opposite points on the sphere

$$S_\ell : \quad |x|^2 + x \cdot \pi(\ell) = K(\ell). \quad (62)$$

Our main object of interest are the connected components of this graph, called *geometric blocks*. In [13] we have shown that for a generic choice of  $S$  the set  $S$  is itself a connected component, called the *special component* all other components thus decompose  $S^c$ .

**Definition 16** For a given choice of edges  $X$  and vectors  $S$  we denote by  $\Gamma_{S,X}$  the resulting graph.

In the NLS, when  $q$  is fixed, we denote  $\Gamma_S := \Gamma_{S, X_q}$

*Remark 9* Note that if  $S \subset S'$  and  $X \subset X'$  we have  $\Gamma_{S,X} \subset \Gamma_{S',X'}$ .

**Proposition 6** *For generic choices of  $S$  the connected components of  $\Gamma_{S, X}$  have at most  $2d + 1$  vertices.*

*We have only finitely many components containing red edges, moreover the components which do not contain any red edge have vertices which are affinely independent (and hence at most  $d+1$ ) and are naturally decomposed in a finite family of subsets in each the connected components are obtained from one by translation.*

*Proof* Our graph is a subgraph of the graph defined in [13] relative to the edges of some  $X_p$  provided that we choose  $p$  large enough. Then by Theorem 3 of [13] for generic choices of  $\mathbf{j}_i$  we have the desired bound.  $\square$

In [13], Definition 17, we have associated to each connected component  $A$  a purely combinatorial graph  $\mathcal{A}$  which contains only the markings of the edges of  $A$  and hence encodes the information on the equations which the vertices of  $A$  must solve (equations (61) of [13]) of type as (61) and (62). We then have that there are only finitely many such combinatorial graphs, called *combinatorial blocks*, while there may be infinitely many  $A$  which have the same  $\mathcal{A}$ . In case the graph has no red edges the equations for the vertex  $x$  are all linear. This implies that the connected components of  $\Gamma_S$  which correspond to a given combinatorial graph with only black edges and with a chosen vertex are all obtained from a single one by translations by vectors which are orthogonal to the edges of the graph.

We also have chosen a preferred vertex, called the *root*, in each connected component, in such a way that the roots of the translates are the translates of this root. Formalizing:

- we have a map  $\mathbf{r} : S^c \rightarrow S^c$  with image the chosen set  $S^{c, \mathbf{r}}$  of roots.
- The fibers of this map are the connected components of the graph  $\Gamma_S$ .
- When we walk from the root  $\mathbf{r}(k)$  to  $k$  the parity  $\pm 1$  of the number of red edges on the path is independent of the path and we denote it by  $\sigma(k)$  (the *color* of  $k$ ).
- There are only finitely many elements  $k$  with  $\sigma(k) = -1$ , the finitely many corresponding roots are exactly the roots of the components with red edges.
- In any case the color of the root is always 1 (*black*).

The vector  $L(k) = \sum L_i(k)e_i$ , with  $k \in S^c$ , appearing in Theorem 2 is defined in Lemma 10 of [13] through the edges appearing in a path from the root to  $k$ . It depends only on the combinatorial graph. The vector  $L(k)$  tells us how to go from the root of the component  $A$  of the graph  $\Gamma_S$  to which  $k$  belongs, to  $k$ , namely:

$$\begin{aligned} k + \sum_i L_i(k)\mathbf{j}_i &= \sigma(k)\mathbf{r}(k), \quad |k|^2 + \sum_i L_i(k)|\mathbf{j}_i|^2 = \sigma(k)|\mathbf{r}(k)|^2, \\ 1 + \sum_i L_i(k) &= \sigma(k). \end{aligned} \tag{63}$$

This definition is well posed even if  $A$  is not a tree, so that one can *walk* from  $\mathbf{r}(k)$  to  $k$  in several ways, from our genericity conditions.

### 5.1.1 The NLS graph

We now apply the previous geometric construction to the NLS and the tangential sites  $S$  of Theorem 2. With the previous notations we have

$$\begin{aligned} \tilde{Q} &= \sum_{k \in S^c} \omega^{(1)}(\xi) \cdot L(k) |z'_k|^2 + \sum_{\ell \in X_q^0} c_q(\ell) \sum_{(h,k) \in \mathcal{P}_\ell} z'_h \bar{z}'_k \\ &+ \sum_{\ell \in X_q^{-2}} c_q(\ell) \sum_{\{h,k\} \in \mathcal{P}_\ell} [z'_h z'_k + \bar{z}'_h \bar{z}'_k]. \end{aligned} \quad (64)$$

where, given an edge  $\ell$ , we set  $\ell = \ell^+ - \ell^-$  and define:

$$c_q(\ell) := \begin{cases} (q+1)^2 \xi^{\frac{\ell^+ + \ell^-}{2}} \sum_{\substack{\alpha \in \mathbb{N}^m \\ |\alpha + \ell^+|_1 = q}} \binom{q}{\ell^+ + \alpha} \binom{q}{\ell^- + \alpha} \xi_i^\alpha & \ell \in X_q^0 \\ (q+1)q \xi^{\frac{\ell^+ + \ell^-}{2}} \sum_{\substack{\alpha \in \mathbb{N}^m \\ |\alpha + \ell^+|_1 = q-1}} \binom{q+1}{\ell^- + \alpha} \binom{q-1}{\ell^+ + \alpha} \xi_i^\alpha & \ell \in X_q^{-2} \end{cases} \quad (65)$$

We see now that the graph has been constructed in order to *decouple*  $\tilde{Q} = \sum_A \tilde{Q}_A$ . The sum runs over all geometric blocks  $A \in \Gamma_S$  and, if  $E_b(A), E_r(A)$  denotes the set of resp. black and red edges in  $A$ :

$$\tilde{Q}_A := \sum_{k \in A} \omega^{(1)}(\xi) \cdot L(k) |z'_k|^2 + \sum_{\ell \in E_b(A)} c_q(\ell) z'_h \bar{z}'_k + \sum_{\ell \in E_r(A)} c_q(\ell) [z'_h z'_k + \bar{z}'_h \bar{z}'_k] \quad (66)$$

is a quadratic Hamiltonian in the variables  $w'_A := z'_k, \bar{z}'_k$  with  $k$  running over the vertices of  $A$ , we have  $\{\tilde{Q}_A, \tilde{Q}_B\} = 0, \forall A \neq B$ . In [13] we have shown that

**Proposition 7** *For each geometric block  $A$  (with  $a$  vertices) we can divide the corresponding  $2a$  dynamical variables  $w'_A$  into two conjugate components  $w'_A = (u', \bar{u}')$  each spanning a  $\tilde{Q}_A$  stable Lagrangian subspace where  $u'_k = z_k$  if  $\sigma(k) = 1$ ,  $u'_k = \bar{z}_k$  if  $\sigma(k) = -1$ . In this basis  $-i\tilde{Q}_A$  has a block matrix  $C_A \oplus -C_A$ . By convention in the first block the root  $\mathbf{r}$  corresponds to  $z'_\mathbf{r}$ .*

Given two vertices  $h \neq k \in A$  we have that the matrix element  $c_{u'_h, u'_k}$  of  $C_A$  is non zero if and only if  $h, k$  are joined by an edge (marked say  $(i, j)$ ) and then

$$c_{u'_h, u'_k} = \sigma(k) c_q(\ell), \quad c_{u'_k, u'_h} = \sigma(k) (\omega^{(1)}(\xi), L(k)). \quad (67)$$

By definition  $L(k)$  depends only on the combinatorial graph  $\mathcal{A}$  of which  $A$  is a realization, therefore the matrix  $C_A = C_{\mathcal{A}}$  depends only on the combinatorial block  $\mathcal{A}$  (but the dynamical variables in  $-i\tilde{Q}_A$  depend on the geometric block).

We thus finally have a finite list  $\mathcal{G} := \{\mathcal{A}_1, \dots, \mathcal{A}_N\}$  of combinatorial graphs which may appear, together with a list of matrices  $C_{\mathcal{A}_i}$  which are explicitly described using Formula (67). The entries of these matrices are polynomials in the elements  $\sqrt{\xi_i}$  and homogeneous of degree  $q$  in  $\xi$ .

### 5.1.2 The space $F^{0,1}$

In the KAM algorithm we shall need to study in particular the action by Poisson bracket of  $\mathcal{N}$  on a special space of functions (linear Hamiltonians on  $\bar{\ell}^{(a,p)}$ ) called  $F^{0,1}$  defined in [13], for the readers convenience we recall the basic facts.

It is convenient to write a variable  $z$  or  $\bar{z}$  as  $z^\sigma$  where  $\sigma$  is 1 resp.  $-1$ .

**Definition 17** We set  $F^{0,1}$  to be the space of functions spanned by the basis elements  $e^{i\sigma\sigma(k)\nu \cdot x} z_k'^{\sigma} = e^{i\sigma([\sigma(k)\nu + L(k)] \cdot x)} z_k'^{\sigma}$ ,  $k \in S^c$  which preserve mass and momentum. <sup>7</sup>

One easily sees that  $F^{0,1}$  is a symplectic space under Poisson bracket. The formulas for mass and momentum in the new variables are for  $\mathbf{m} := e^{i\sigma\sigma(k)\nu \cdot x} z_k'^{\sigma}$ .

$$\{\mathbb{L}, \mathbf{m}\} = i\sigma\sigma(k) \left( \sum_i \nu_i + 1 \right) \mathbf{m}, \quad \{\mathbb{M}, \mathbf{m}\} = i\sigma\sigma(k) \left( \sum_i \nu_i \mathbf{j}_i + \mathbf{r}(k) \right) \mathbf{m}. \quad (68)$$

hence the conservation laws tell us that for an element  $e^{i\sigma\sigma(k)\nu \cdot x} z_k'^{\sigma} \in F^{0,1}$  the vector  $\nu \in \mathbb{Z}^d$  is constrained by the fact that  $-\sum_i \nu_i \mathbf{j}_i$  must be in the set of roots in  $S^c$  and moreover the mass constraint  $\sum_i \nu_i = -1$ .

For each connected component  $A$  of the graph  $\Gamma_S$  with root  $\mathbf{r}$ , any solution  $\nu$  of  $\sum_i \nu_i \mathbf{j}_i + \mathbf{r} = 0$ , where the mass of  $\nu$  is  $-1$ , determines in  $F^{0,1}$  a block denoted  $A, \nu$ . This is a symplectic space sum of two Lagrangian spaces  $(A, \nu)_+$  and  $(A, \nu)_- = \overline{(A, \nu)_+}$ .  $(A, \nu)_+$  has basis the elements  $e^{i\sigma\sigma(k)\nu \cdot x} z_k'^{\sigma(k)}$ ,  $k \in A$ . Thus  $F^{0,1}$  decomposes as orthogonal sum (with respect to the symplectic form), of these blocks  $A, \nu$ . Notice that, for a given geometric block  $A$  there are infinitely many blocks  $A, \nu$  as soon as  $n > d$ .

**Proposition 8** The matrix of  $-i\text{ad}(\mathcal{N})$  on the block  $(A, \nu)_+$  is the sum of the matrix  $C_A$  plus the scalar matrix  $[(|\mathbf{r}(m)|^2 + \sum_i \nu_i |\mathbf{j}_i|^2) + \omega^{(1)}(\xi) \cdot \nu] I_A$ .

*Proof* The action of  $-i\tilde{\mathcal{Q}}$  does not depend on  $\nu$ , it is only through  $-i\tilde{\mathcal{Q}}_A$  and gives the matrix  $C_A$ . As for the elements  $-i[\omega(\xi) \cdot y' + \sum_{k \in S^c} \tilde{\Omega}_k |z_k'|^2]$  by Formula (63), the term  $-i \sum_{k \in S^c} \tilde{\Omega}_k |z_k'|^2$  contributes on the first block the scalar  $|\mathbf{r}(m)|^2$ . As of  $-i\omega(\xi) \cdot y'$  it also contributes by a scalar, this time  $(\sum_i \nu_i |\mathbf{j}_i|^2 + \omega^{(1)}(\xi) \cdot \nu)$ .  $\square$

## 6 Normal form reduction

### 6.1 Fitting decomposition

Let us recall some basic definitions of linear algebra, given a linear operator  $A : V \rightarrow V$  where  $V$  is a finite dimensional vector space over a field  $F$  (we will work on function fields so of characteristic 0), we have the *Jordan decomposition*  $A = A_s + A_n$  where  $A_s$  is semisimple and  $A_n$  nilpotent with  $[A_s, A_n] = 0$ . Such a decomposition is unique so that  $A_s$  is called the semisimple part of  $A$  and  $A_n$  the nilpotent part. Semisimple may be defined in several ways but for us means that  $A_s$  is diagonalizable, not necessarily over  $F$  but in some finite extension field  $G \subset F$  which contains the eigenvalues of  $A$  (which are in fact the same as the

<sup>7</sup> we deviate from the notations of [13] and in  $F^{0,1}$  we also impose zero mass

eigenvalues of  $A_s$ ). The characteristic polynomials of  $A$  and  $A_s$  coincide, on the other hand  $A_s$  satisfies its minimal polynomial which has coefficients in  $F$  and as root all the eigenvalues with multiplicity 1.

If  $F$  contains the eigenvalues of  $A$ , call the distinct eigenvalues  $\alpha_1, \dots, \alpha_i$ , each may appear with some multiplicity, we have the *Fitting decomposition* of  $V = \oplus_i V_{\alpha_i}$  where each  $V_{\alpha_i}$  is a uniquely determined subspace which is stable under  $A$  and where  $A$  has the unique eigenvalue  $\alpha_i$ . It is easily seen that in fact  $V_{\alpha_i}$  is just the eigenspace of  $A_s$  for the eigenvalue  $\alpha_i$ .

In our case we are interested in matrices  $C(\xi)$  depending in a polynomial way from the parameters  $\sqrt{\xi_i}$ , we may consider the field  $F$  of rational functions in these parameters so that  $A = C(\xi)$  is a matrix of some size  $p$ , with coefficients in  $F$ , denote by  $M_p(F)$  the space of these matrices.

Take the Jordan decomposition and notice that, in this decomposition,  $A_s$  has no more polynomial entries but may acquire denominators. Then we have some finite field extension  $G \supset F$  and a matrix  $X \in M_p(G)$  such that  $XC(\xi)X^{-1}$  is a block diagonal form  $\oplus C_i$  with  $C_i \in M_{p_i}(G)$  has a unique eigenvalue  $\alpha_i \in G$  and the various eigenvalues are distinct. In fact  $X$  is defined up to a scalar multiplication so that we may further assume that all the entries of  $X$  are integral over the ring of polynomials in the  $\xi_i$ , that is satisfy each some monic polynomial (dependent of the entry)  $t^N + a_1 t^{N-1} + \dots + a_N$  where the coefficients  $a_i$  are polynomials in the  $\xi$ .

Now we have to interpret  $G$  as field of algebraic functions in the parameters  $\xi$ , this follows a standard path. The distinct eigenvalues are solutions of the minimal polynomial, then by removing the discriminant (which gives a algebraic hypersurface) on the complement these are distinct as functions, then one can further remove some real algebraic variety, which we may assume to be homogeneous, so that the open components of the complement are all simply connected (this follows for instance from the fact that one has an algebraic triangulation of the sphere such that it also triangulates the intersection of all these algebraic hypersurfaces with the sphere, then one takes the corresponding cones of the open triangles [5]). As a result we have that we may remove from the space of parameters  $\xi$  (which for us are real parameters) some algebraic hypersurface (of course real), such that outside this hypersurface the entries of  $X$  are true functions, which are as we have said algebraic so analytic, and integral in the sense explained, and finally assume distinct values for each value of  $\xi$  outside these algebraic hypersurfaces. Of course once we remove an algebraic hypersurface the complement is open and dense.

If  $A \in \text{End}(V)$  is a linear map and we decompose  $V = \oplus_{\alpha} V_{\alpha}$  into a Fitting decomposition, we have for the linear operator  $\text{ad}(A) : X \mapsto [A, X]$ , on the space  $\text{End}(V)$  of linear maps on  $V$ , the following facts.

$$\text{ad}(A)_s = \text{ad}(A_s), \quad \text{ad}(A)_n = \text{ad}(A_n).$$

The Fitting decomposition of  $\text{End}(V)$ , under  $\text{ad}(A)$ , is deduced from the decomposition in blocks induced by  $V = \oplus_{\alpha} V_{\alpha}$  that is first

$$\text{End}(V) = \oplus_{\alpha, \beta} \text{hom}(V_{\alpha}, V_{\beta}).$$

The subspace  $\text{hom}(V_{\beta}, V_{\alpha})$  is relative to the eigenvalue  $\alpha - \beta$  so that  $\text{ad}(A)$  on this space acts as  $(\alpha - \beta)\text{Id} + \text{ad}(A_n)$ . The map  $\text{ad}(A_n)$  is nilpotent, in fact if  $A_n^k = 0$

we have  $ad(A_n)^{2k} = 0$ . So for the off diagonal blocks,  $\alpha \neq \beta$  we have  $ad(A)$  on such a block is invertible with inverse

$$(\alpha - \beta)^{-1} \left( 1 + \sum_{j=1}^{2k-1} (-1)^j \left[ \frac{ad(A_n)}{(\alpha - \beta)} \right]^j \right).$$

It is possible that for different eigenvalues we may have  $\alpha_1 - \beta_1 = \alpha_2 - \beta_2$  so that this decomposition in general refines, but need not be equal, to the Fitting decomposition of  $ad(A)$ .

### 6.1.1 Blocks and matrices

We now apply this analysis to the finitely many matrices  $C_{\mathcal{A}_i}$  associated to the list  $\mathcal{G} := \{\mathcal{A}_1, \dots, \mathcal{A}_N\}$  of combinatorial graphs which may appear.

Thus we have finitely many distinct eigenvalues (with some multiplicity) of this finite list of matrices, which are algebraic functions of the parameters  $\xi$  and defined outside some homogeneous algebraic hypersurface. To make the discussion precise from now on one may remove from  $\mathbb{R}^n$  a homogeneous real algebraic hypersurface so that the connected components of the complement are all simply connected, and so that the distinct eigenvalues are functions in each component and with distinct values pointwise (see §6.1).

We fix one of these open simply connected regions, call it  $\mathcal{R}_\alpha$  in which these algebraic functions are well defined and then we denote this list of functions by

$$\mathcal{T} := \{\theta_1, \dots, \theta_O\}. \quad (69)$$

Contrary to what happens for the cubic NLS we may have various kinds of degeneracies in these eigenvalues, one eigenvalue may appear in two different matrices  $C_{\mathcal{A}_i}$  and with multiplicity  $> 1$ . This is discussed in §6.1 and 7.1.

Later we will need to restrict to some compact domain in one of the  $\mathcal{R}_\alpha \cap \Lambda$  in order to make estimates on the derivatives.

*Decomposing  $\mathcal{N}$ .* We now want to simplify  $\mathcal{N}$  by applying the Jordan and Fitting decomposition to the quadratic Hamiltonians  $\tilde{Q}_A$ . In order to do this we first write  $\tilde{Q}_A = \tilde{Q}_A^s + \tilde{Q}_A^n$  so that  $\tilde{Q}_A^s$  and  $\tilde{Q}_A^n$  commute,  $\tilde{Q}_A^s$  is represented by a semi-simple matrix and  $\tilde{Q}_A^n$  is represented by a nilpotent matrix. Note that, since  $\tilde{Q}_A$  is a polynomial in  $\sqrt{\xi}$ , then both its nilpotent and semi-simple part are rational functions in  $\sqrt{\xi}$ , so they are analytic outside an algebraic hypersurface (cf. §6.1).

Clearly  $\tilde{Q}^s = \sum_A \tilde{Q}_A^s$  and the matrix  $\frac{1}{2}Q_A^s = C_A^s \oplus (-C_A^s)$  where  $C_A^s$  is the semi-simple part of  $C_A$ .

Notice that all the  $\tilde{Q}_A$  such that  $A$  does not have red edges are represented by self-adjoint matrices and hence do not contribute to  $\tilde{Q}^n$ . Moreover since  $\mathbb{K}$  and  $\omega^{(1)}(\xi) \cdot y'$  commute with  $\mathcal{N}$  we can decompose  $\mathcal{N} = \mathcal{N}^s + \mathcal{N}^n$  (semi-simple and nilpotent parts) where  $\mathcal{N}^n = \tilde{Q}^n$  and  $\{\mathcal{N}^s, \mathcal{N}^n\} = 0$ .

We put  $\mathcal{N}^s$  in normal form by reducing each of the  $C_A^s$ . This means that we only need to work on a finite number of matrices.

**Proposition 9** *There exists a homogeneous algebraic hypersurface  $\mathfrak{A}$  such that the open region  $\mathbb{R}^n \setminus \mathfrak{A}$  decomposes into finitely many simply connected components  $\mathcal{R}_\alpha$  with the following property:*

*For each combinatorial graph  $\mathcal{A}$  the eigenvalues of the matrix  $C_{\mathcal{A}}$ , are algebraic, hence analytic, functions of  $\xi$  say  $\theta_1, \dots, \theta_k$  which are homogeneous of degree  $q$ .*

*For all  $\xi \in \mathbb{R}^n \setminus \mathfrak{A}$  there exists a linear symplectic change of coordinates  $u' \rightarrow U_{\mathcal{A}}(\xi)u' = u''$  such that:*

1.  $U_{\mathcal{A}}(\xi)$  is unitary with respect to  $\Sigma_{\mathcal{A}}$  and it can be chosen homogeneous of degree zero in  $\xi$ .

2.  $U_{\mathcal{A}}(\xi)$  is analytic in  $\xi$ .

3.  $U_{\mathcal{A}}(\xi)$  conjugates  $C_{\mathcal{A}}$  into a Fitting normal form: that is a block diagonal matrix such that each block has a unique eigenvalue.

*More precisely for each real eigenvalue  $\theta$ ,  $C_{\mathcal{A}}^s$  acts as  $\theta I$  on the eigenspace of  $\theta$ .*

*For each pair of conjugate complex eigenvalues  $\theta_{\pm} = a \pm ib$ , of multiplicity  $k$  we have a real  $2k$  dimensional space such that the two complex eigenvectors lie in its complexification. Then we have a basis of this subspace such that  $C_{\mathcal{A}}^s$  restricted to this subspace is a  $2 \times 2$  block matrix*

$$\begin{pmatrix} aI_k & -bI_k \\ bI_k & aI_k \end{pmatrix}.$$

*The matrix  $\sigma_{\mathcal{A}}$  of the standard form on this basis is  $\text{diag}(I_k, -I_k)$ .*

Note that two eigenvalues which are distinct as functions of  $\xi$  can be assumed to be distinct for all values of the parameters  $\xi_i \in \mathbb{R}^n \setminus \mathfrak{A}$ .

*Change of variables.* For each geometric block  $A$  we have the two Lagrangian blocks of variables where the Hamiltonian  $\tilde{Q}_A$  acts by Poisson bracket with the matrix  $C_{\mathcal{A}} \oplus (-C_{\mathcal{A}})$ , we thus perform on the two Lagrangian blocks the symplectic change of variables induced by the matrix  $U_{\mathcal{A}}$  which puts into Fitting decomposition  $C_{\mathcal{A}}$ . This is by construction a symplectic change of variables and we call with  $w''$  the new variables obtained. In these new variables we have a further decoupling of each  $\tilde{Q}_A$  as a sum  $\tilde{Q}_A = \sum_{\theta} \tilde{Q}_{A,\theta}$ , since the component  $A$  is determined by its root  $\mathfrak{r}$  we denote also  $\tilde{Q}_{A,\theta} = \tilde{Q}_{\mathfrak{r},\theta}$ .

For the  $\mathcal{A}$  with no red edges the variables  $u'_k = z'_k$ . So it is obvious that  $w'' = (Uz', \bar{U}\bar{z}')$ . In general we have always  $w'' = (Uu', \bar{U}\bar{u}')$  but we should specify which of the  $u'_k$  have  $\sigma(k) = 1$  or  $-1$ . We notice that by definition  $\{u'_k, \bar{u}'_h\} = i\delta_k^h \sigma(k)$  and since  $U$  is unitary with respect to  $\Sigma_{\mathcal{A}} = \text{diag}(\sigma(k))_{k \in A}$  then also  $\{u''_k, \bar{u}''_h\} = i\delta_k^h \sigma(k)$ , so we still may say that  $u''_k = (z'')^{\sigma(k)}$ . By abuse of notation we still call  $x, y, z_k, \bar{z}_k$  the new variables.

Since  $U_{\mathcal{A}}$  mixes only the variables  $u$  and since  $\mathbb{L}, \mathbb{M}, \mathbb{K}$  are scalar on the components  $u$  then in the new variables we have still

$$\mathbb{L} = \sum_i y_i + \sum_k \sigma(k) |z_k|^2, \quad \mathbb{M} = \sum_i j_i y_i + \sum_k \sigma(k) \mathfrak{r}(k) |z_k|^2, \quad (70)$$

$$\mathbb{K} = \sum_i |j_i|^2 y_i + \sum_k \sigma(k) |\mathfrak{r}(k)|^2 |z_k|^2. \quad (71)$$

**Definition 18** We denote by  $\mathfrak{S}^c$  the set of pairs  $\mathbf{r}, \theta$  where  $\mathbf{r}$  is a root of a block  $A$  in the geometric graph while  $\theta$  is one of the eigenvalues appearing in the corresponding Fitting decomposition of  $\mathcal{A}$ .

After the change of variables the new basis is still indexed by convention by the set  $S^c$ . Then  $S^c$  is decomposed as a union of finite blocks  $S_{\mathbf{r}, \theta}^c$  corresponding to the elements of  $\mathfrak{S}^c$ . Notice that only finitely many of the blocks indexed by  $\mathfrak{S}^c$  contain  $k$  with  $\sigma(k) = -1$ . Therefore all  $\tilde{Q}_{\mathbf{r}, \theta}$ , except finitely many, have the form,  $\tilde{Q}_{\mathbf{r}, \theta} = \theta(\sum_{k \in S_{\mathbf{r}, \theta}^c} |z_k|^2)$ .

The finite *bad blocks* have still a semisimple part  $\tilde{Q}_{\mathbf{r}, \theta}^s = \theta(\sum_k \sigma(k) |z_k|^2)$  and  $\sum_{k \in S_{\mathbf{r}, \theta}^c} \sigma(k) |z_k|^2$  acts as the scalar  $i, -i$  on the two corresponding Lagrangian spaces, since a basis of one of them is  $z_k^{\sigma(k)}$ ,  $k \in S_{\mathbf{r}, \theta}^c$  (and the other the conjugate). The normal form  $\mathcal{N}$  is

$$\mathcal{N} = \mathbb{K} + \mathbb{K}^1, \quad \mathbb{K}^1 = (\omega^{(1)}, y) + \sum_{(\mathbf{r}, \theta) \in \mathfrak{S}^c} \tilde{Q}_{\mathbf{r}, \theta} \quad (72)$$

where the decoupled quadratic Hamiltonian  $\tilde{Q}_{\mathbf{r}, \theta}$  corresponds to the block of the Fitting decomposition, of eigenvalue  $\theta$ , of the matrix  $C_{\mathcal{A}}$ .

*Remark 10* For a given  $k$  it is useful to denote by  $\theta(k)$  the corresponding eigenvalue. Note that the map  $k \mapsto \theta(k)$  depends only on the combinatorial graph.

In order to keep track of the information on the eigenvalue in the geometric and combinatoric graphs, we mark each vertex with its eigenvalue  $\theta(k)$ .

## 7 the kernel of $ad(\mathcal{N})$

We now study the subspace of regular analytic Hamiltonians of degree  $\leq 2$  which Poisson commute with  $\mathcal{N}$ ,  $\mathbb{M}$ ,  $\mathbb{L}$  for generic values of  $\xi$ . Note that they must commute separately with  $\mathbb{K}$  and  $\mathbb{K}^1$  since they are homogeneous of degree  $0, q$  respectively in  $\xi$ .

**Degree zero in  $w$ :** This space, which we denote by  $F^0$  can be divided into  $F^{0,0} \oplus F^{1,0}$  with basis  $e^{i\nu \cdot x}$  and  $e^{i\nu \cdot x} y_l$ . Since monomials are eigenvectors of all the operators we need to verify when all the eigenvalues are identically zero.

$$0 = \{\mathbb{L}, e^{i\nu \cdot x}\} \iff i \sum_i \nu_i = 0, \quad 0 = \{\mathbb{M}, e^{i\nu \cdot x}\} \iff i \sum_i \nu_i j_i = 0,$$

$$0 = \{\mathbb{K}, e^{i\nu \cdot x}\} \iff i \sum_i \nu_i |j_i|^2 = 0, \quad 0 = \{\mathbb{K}^1, e^{i\nu \cdot x}\} \iff i(\omega^{(1)}, \nu) = 0.$$

The last condition implies that  $\nu = 0$ , by 5. The same rules hold if the monomial has degree one in  $y$ . Hence the Kernel of  $ad(\mathcal{N})$  is  $x$  independent and hence of the form  $c + \vec{C} \cdot y$ .

**Degree one in  $w$ :** We analyze the action of  $ad(\mathcal{N})$  on a block of the space  $F^{0,1}$  given by an element of  $\mathfrak{S}^c$ . Given  $k \in S_{\mathbf{r}}^c$  in this block the monomial  $e^{i\sigma(k)\nu \cdot x} z_k^\sigma$  is an eigenvector for all our operators, by conservation of  $\mathbb{L}$ ,  $\mathbb{M}$  we have  $\sum_i \nu_i + 1 = 0$ ,  $\sum_i \nu_i j_i + \mathbf{r}(k) = 0$  and with eigenvalue for  $\mathbb{K}$ :

$$\{\mathbb{K}, e^{i\sigma(k)\nu \cdot x} z_k^\sigma\} \rightarrow i\sigma(k) \left( \sum_i \nu_i |j_i|^2 + |\mathbf{r}(k)|^2 \right),$$

as for  $\mathbb{K}^1$  we have the diagonal contribution

$$\{\omega^{(1)} \cdot y, e^{i\sigma\sigma(k)\nu \cdot x} z_k^\sigma\} \rightarrow i\sigma\sigma(k)\omega^{(1)} \cdot y$$

plus the contribution of the term  $\tilde{Q}_{\mathbf{r},\theta}$ . Now we claim that we may assume that the determinant of the matrix of  $ad(\mathcal{N})$  is non-zero as a function in  $\xi$ . This is clear if the scalar part  $\sum_i \nu_i |j_i|^2 + |\mathbf{r}(k)|^2 \neq 0$ . Since we are working in the space where we have the constraint  $\sum_i \nu_i j_i + \mathbf{r}(k) = 0$  this condition amounts to ask that

$$\sum_i \nu_i |j_i|^2 + \left| \sum_i \nu_i j_i \right|^2 \neq 0. \quad (73)$$

We then choose our tangential sites  $j_i$  satisfying this constraint for all  $\nu$  with coordinates  $|\nu_i| \leq a$  where  $a \leq 4dq$  is a common bound for the absolute value of the coordinates of each  $L(k)$  (which form a finite list). We then claim that for the remaining  $\nu$  the matrix  $C_{\mathcal{A}}$  has non-zero determinant, as polynomial in  $\xi$ . For such a  $\nu$  we have some coordinate  $\nu_i$  with  $|\nu_i| > a$ , we may assume that it is  $\nu_1$  then set  $\xi_j = 0, \forall j \neq 1$  and verify that still the determinant of the specialized matrix is  $\neq 0$ . All the off diagonal entries, which are given by formulas (65) vanish since they all contain at least 2 variables  $\xi$  due to Formula (57).

From Lemma 5, when we set  $\xi_j = 0, \forall j \neq 1$  we get  $\omega^{(1)} = -q(q+1)\xi_1^q e_1$ .

To prove our claim we need to analyze in more detail the matrix  $C_{\mathcal{A}}$  in particular its diagonal part which, from Formula (66) comes from  $\omega^{(1)}(\xi) \cdot L(k) |z'_k|^2$ . Finally after the specialization the diagonal matrix has entries a non zero constant times  $\xi_1^q (\nu_1 - L_1(k))$  which is non-zero by our constraint on  $\nu_1$  since the coordinates of each  $L(k)$  are in absolute value bounded by  $a$ .

We have proved that for all  $\xi$  outside a countable set of algebraic hypersurfaces the kernel of  $ad(\mathcal{N})$  is zero on functions of degree one commuting with  $\mathbb{L}, \mathbb{M}$ .

Since the determinants depend only on the semi-simple part we have the same statement also for  $\mathcal{N}^s$ ,

*Remark 11* Note that the genericity constraints (73) can be imposed since they are not zero as functions of  $j_i$ , indeed the only trivial case is when  $\nu = -e_i$  which contradicts  $\mathbf{r} \in S^c$ . Actually constraints like (73) appeared already in [13].

**Degree two in  $w$ :** We denote this space by  $F^{0,2}$ . We do not exhibit explicitly such a kernel (since we do not control the nilpotent part of  $\mathcal{N}$ ). We observe that  $\ker(ad(\mathcal{N}))$  is contained in  $\ker(ad(\mathcal{N}^s))$  which now depends only on the eigenvalues. The eigenvectors of degree two for  $ad(\mathcal{N}^s)$  in  $w$  are obtained by multiplying eigenvectors of degree one in all possible ways and the eigenvalues are sums of two eigenvalues. Since each eigenvalue appears together with its negative in the kernel we have the following two cases. This depends on whether we pair  $(\nu_1, A_1)_+$  with  $(\nu_2, A_2)_+$  (then we set  $\nu = \nu_1 + \nu_2$ ) or we pair  $(\nu_1, A_1)_+$  with  $(\nu_2, A_2)_-$  (and we set  $\nu = \nu_1 - \nu_2$ ).

Case 1:

$$\begin{aligned} \sum_i \nu_i &= -2, & \sum_i \nu_i j_i + \mathbf{r}(h) + \mathbf{r}(k) &= 0 \\ \sum_i \nu_i |j_i|^2 + |\mathbf{r}(h)|^2 + |\mathbf{r}(k)|^2, & \omega^{(1)}(\xi) \cdot \nu + \theta_h + \theta_k &= 0. \end{aligned} \quad (74)$$

Case 2:

$$\sum_i \nu_i = 0, \quad \sum_i \nu_i \mathbf{j}_i + \mathbf{r}(h) - \mathbf{r}(k) = 0.$$

$$\sum_i \nu_i |\mathbf{j}_i|^2 + |\mathbf{r}(h)|^2 - |\mathbf{r}(k)|^2 = 0, \quad \omega^{(1)}(\xi) \cdot \nu + \theta_h - \theta_k = 0. \quad (75)$$

*Remark 12* Of course in the kernel we always have the case  $\nu = 0$ , therefore we are in Case 2 with  $\mathbf{r}(h) = \mathbf{r}(k)$  and  $h, k$  in the same Fitting block.

On the other hand we may have a kernel arising from different Fitting blocks and also with  $\nu \neq 0$ . We want to encode this phenomenon in a new graph, whose edges we denote by  $Y = Y^0 \cup Y^{-2}$ , and defined as follows.

### 7.1 The graph $\Gamma_S^{(f)}$

**Definition 19** We say that  $\ell \in \mathbb{Z}^n \setminus \{0\}$  is an edge  $\ell \in Y^0$  if  $\eta(\ell) = \sum_i \ell_i = 0$  and we can obtain  $\omega^{(1)}(\xi) \cdot \ell$  as difference of two eigenvalues of the list (69) (of the matrices  $C_A$ ).

We say  $\ell \in \mathbb{Z}^n$  is an edge  $\ell \in Y^{-2}$  if  $\eta(\ell) = \sum_i \ell_i = -2$ ,  $\ell \neq -2e_i$  ( $i = 1, \dots, n$ ) and we can obtain  $\omega^{(1)}(\xi) \cdot \ell$  as sum of two eigenvalues of the list (69).

- Remark 13*
- i) Notice that  $\ell$  is uniquely determined by the polynomial  $\omega^{(1)}(\xi) \cdot \ell$  since  $-q(q+1)\ell_i$  is the coefficient of  $\xi_i^q$ , see Lemma 5.
  - ii) We notice that  $Y$  depends only on the structure of the matrices  $C_A$  which form a list of matrices associated to finitely many combinatorial graphs and do not depend on the choice of the  $\mathbf{j}_i$ .
  - iii) Then  $Y$  is contained in the set of edges  $X_p$  (see Definition 57) for some  $p$  sufficiently large.
  - iv) Notice that in the case of complex eigenvalues we must have that the imaginary parts cancel.

*The basic stratification.* The geometric graph  $\Gamma_{S, X_q \cup Y}$  is now used to define the basic stratification of  $\mathbb{Z}^d$  of which all the stratifications used in proposition 1 form a refinement.

The construction of the stratification is explained in [15], we take as single point strata all the points in the finite graphs with red edges. Then the graphs with only black edges come into a finite family such that the graphs of each family are all isomorphic and obtained each from the other by a translation. The set of translations, for a given graph  $\Gamma$  with  $a+1$  affinely independent points is defined as follows, one chooses a root  $\mathbf{r} \in \Gamma$  and takes the subgroup of elements  $u \in \mathbb{Z}^d$  orthogonal to the  $a$  linearly independent elements  $x - \mathbf{r}$ ,  $x \in \Gamma \setminus \{\mathbf{r}\}$ . Then generically  $\Gamma + u$  is still a connected component of the graph  $\Gamma_{S, X_q \cup Y}$  while for some special  $u$  this translate is contained in a larger connected component.

**Definition 20** The stratum associate to a point  $k$  lying in a graph  $\Gamma$  with only black edges is then defined by taking all the points  $h = k + u$  lying in an isomorphic graph  $\Gamma + u$  connected component of  $\Gamma_{S, X_q \cup Y}$ .

The graph  $\tilde{\Gamma}_{S, X_q \cup Y}$ . Starting from the geometric graph  $\Gamma_S$  we construct a new geometric graph, which we denote by  $\tilde{\Gamma}_{S, X_q \cup Y}$ , with vertices in  $S^c$  and edges  $X_q \cup Y$ . This is obtained from the geometric graph  $\Gamma_{S, X_q \cup Y}$ , which contains the graph  $\Gamma_S$  by keeping only the new edges which connect two roots (of  $\Gamma_S$ ) with the following constraints.

As explained in Remark 10 we have marked each vertex in  $S^c$  with its eigenvalue  $\theta(k) \in \mathcal{Y}$ . Consider any two connected components  $A, B$  of  $\Gamma_S$  with roots  $\mathbf{r}, \mathbf{r}'$ :

**Definition 21** If  $\mathbf{r}, \mathbf{r}'$  are connected in  $\Gamma_{S, X_q \cup Y}$  by a black oriented edge marked  $\ell \in Y^0$ , that is

$$\sum_i j_i \ell_i + \mathbf{r} - \mathbf{r}' = 0, \quad \sum_i |j_i|^2 \ell_i + |\mathbf{r}|^2 - |\mathbf{r}'|^2 = 0, \quad (76)$$

we keep this edge in  $\tilde{\Gamma}_{S, X_q \cup Y}$  if and only if there exists  $k \in A$  and  $h \in B$  such that

$$\omega^{(1)}(\xi) \cdot \ell + \theta(k) - \theta(h) = 0.$$

If  $\mathbf{r}, \mathbf{r}'$  are connected in  $\Gamma_{S, X_q \cup Y}$  by a red edge marked  $\ell \in Y^{-2}$  that is

$$\sum_i j_i \ell_i + \mathbf{r} + \mathbf{r}' = 0, \quad \sum_i |j_i|^2 \ell_i + |\mathbf{r}|^2 + |\mathbf{r}'|^2 = 0, \quad (77)$$

we keep this edge in  $\tilde{\Gamma}_{S, X_q \cup Y}$  if and only if there exists  $k \in A$  and  $h \in B$  such that

$$\omega^{(1)}(\xi) \cdot \ell + \theta(h) + \theta(k) = 0.$$

*Remark 14* If  $\mathbf{r}, \mathbf{r}'$  are connected by a black oriented edge marked  $\ell$  then  $\mathbf{r}', \mathbf{r}$  are connected by a black oriented edge marked  $-\ell$ , for red edges instead the relation is symmetric.

Notice moreover that for most points  $k \in S^c$ , in fact outside finitely many hyperplanes plus a finite set, one has  $\mathbf{r}(k) = k$  and  $\theta_k = 0$ .

**Proposition 10** *The connected components of  $\tilde{\Gamma}_{S, X_q \cup Y}$  for generic choices of  $S$  have at most  $2d + 1$  vertices moreover the components which do not contain any red edge have vertices which are affinely independent (and hence at most  $d + 1$ ).*

*Proof* Our graph is a subgraph of  $\Gamma_{S, X_q \cup Y}$  and we may apply Proposition 6.  $\square$

*Some geometry.*

*Remark 15* There is an equivalent way of looking at this construction, the conditions of Definition 21, define a graph, called the *root graph*, with vertices the roots  $\mathbf{r}$  and edges in  $Y$ , a connected component of the graph  $\tilde{\Gamma}_{S, X_q \cup Y}$  is obtained by a connected component  $C$  of this graph with vertices roots, by taking the union of the components of  $\Gamma_S$  with root  $\mathbf{r} \in C$ .

Then we see the following important fact. In  $\tilde{\Gamma}_{S, X_q \cup Y}$  the components containing only black edges are clearly obtained by the previous gluing construction from all, except finitely many, the connected components of  $\Gamma_S$  containing only black edges. In other words we have a set  $S_g^c$ , complement of a finite set, formed by all vertices of the connected components of  $\tilde{\Gamma}_{S, X_q \cup Y}$  containing only black edges.

We have that these components come into a finite family of usually infinite sets each obtained by translations from some subgroup of a given component.

We then have a linear stratification of  $\mathbb{Z}^n$ , which we still denote by  $Y_j$ , in which the roots of the components are a union of strata.

So a stratum  $Y$  of roots of dimension  $k$  determines a subgroup  $\Lambda$  of  $\mathbb{Z}^n$  of translations, of its closure, of rank  $k$  and for any  $\mathbf{r} \in Y$  we have that  $Y = \Lambda \setminus \Lambda' + \mathbf{r}$  where  $\Lambda'$  is a finite union of subgroups of  $\Lambda$  or rank strictly lower than  $k$ .

For a given stratum  $Y$  of roots, the graphs having roots in  $Y$  are all isomorphic as combinatorial graphs marked by the elements  $\theta$  under translation by the difference of the roots their union form a union of strata parallel to  $Y$  (i.e. inside cosets of  $\Lambda$ ).

**Definition 22** [Final graph] Notice that this construction induces a graph  $\Gamma_S^{(f)}$  also on the set  $\mathfrak{S}^c$ , see 18 of pairs  $(\mathbf{r}, \theta)$ . We define that  $(\mathbf{r}, \theta), (\mathbf{r}', \theta')$  are connected if  $\mathbf{r}, \mathbf{r}'$  are connected and  $\theta(k) = \theta, \theta(h) = \theta'$  for some  $k, h$  with  $\mathbf{r}(k) = \mathbf{r}, \mathbf{r}(h) = \mathbf{r}'$ .

In this graph then the eigenvalues  $\theta, \theta'$  are either both real or both complex. In the case of complex eigenvalues we must note that if  $(\mathbf{r}, \theta), (\mathbf{r}', \theta')$  are connected then so are  $(\mathbf{r}, \bar{\theta}), (\mathbf{r}', \bar{\theta}')$ .

**Lemma 6** *If  $(\mathbf{r}, \theta), (\mathbf{r}', \theta')$  are connected by a path of edges then they are connected by an edge. Such edge is black if the path contains an even number of red edges and is red otherwise.*

*Proof* We only need to prove our statement when the path is made of two edges. Suppose  $(\mathbf{r}, \theta), (\mathbf{r}_1, \theta_1)$  are connected by the edge  $\ell$  and  $(\mathbf{r}_1, \theta_1)$  is connected to  $(\mathbf{r}', \theta')$  by  $\ell'$ . If  $\ell$  and  $\ell'$  are both black we have  $\theta_1 - \theta = (\omega^{(1)}(\xi), \ell)$  and  $\theta' - \theta_1 = (\omega^{(1)}(\xi), \ell')$  which implies that  $\theta' - \theta = (\omega^{(1)}(\xi), \ell + \ell')$  and hence  $\ell + \ell' \in X$ . In the same way:

$$\sum_i j_i \ell_i + \mathbf{r} - \mathbf{r}_1 = \sum_i j_i \ell'_i + \mathbf{r}_1 - \mathbf{r}' = 0 \rightarrow \sum_i j_i (\ell_i + \ell'_i) + \mathbf{r} - \mathbf{r}' = 0,$$

same for the quadratic formulae. We have proved that  $(\mathbf{r}, \theta), (\mathbf{r}', \theta')$  are connected by a black oriented edge marked  $\ell + \ell'$ . Similarly if  $\ell, \ell'$  are red then  $(\mathbf{r}, \theta), (\mathbf{r}', \theta')$  are connected by a black oriented edge marked  $\ell - \ell'$ . Finally if  $\ell$  is red and  $\ell'$  is black  $(\mathbf{r}, \theta), (\mathbf{r}', \theta')$  are connected by a red edge marked  $\ell - \ell'$ .  $\square$

We need now some further constraints of the frequencies  $\mathbf{j}$ . Since we have seen that this final graph does not depend on  $S$  we can impose the condition that  $(\mathbf{r}, \theta), (\mathbf{r}, \theta'), \theta \neq \theta'$  is never connected by an edge, since we may assume that, if  $\theta \pm \theta' = (\omega^{(1)}(\xi), \ell)$  we have always  $\ell \cdot \mathbf{j} \neq 0$ .

Then we have

**Proposition 11** *The projection  $(\mathbf{r}, \theta) \mapsto \mathbf{r}$  is a map of the graph  $\Gamma_S^{(f)}$  to  $\Gamma_{S, X_a \cup Y}$ . It maps thus a connected component injectively into a connected component.*

In each connected component  $\mathfrak{D}$  of  $\Gamma_S^{(f)}$  we now want to choose a root, we do this by choosing as root a pair  $(\mathbf{r}, \theta)$  with  $\mathbf{r}$  a root of a components  $A$  of  $\Gamma_S$ , in other words we choose a root of the component of the *root graph*, in order to distinguish this from the other pairs appearing we shall use the symbol  $\mathbf{t} := (\mathbf{r}_t, \theta_t)$ . By the previous remarks we can make this choice *translation* invariant for the components

with black edges, we always make this choice. We shall denote this last set of roots by  $\mathfrak{T}$  which therefore is an indexing set for the connected components of  $\Gamma_S^{(f)}$ .

It is important to remark that we can divide  $\mathfrak{T}$  into a finite set  $\mathfrak{T}_f$  (which contains all the bad blocks) and the infinite *good* set  $\mathfrak{T}_g$  which by definition is the part which projects to the part of  $\tilde{\Gamma}_{S, X_q \cup Y}$  made of the components containing only black edges.

By Lemma 6 each vertex  $(\mathbf{r}, \theta)$  of the chosen component  $\mathfrak{D}$  is connected to the root  $\mathfrak{t}$  by an edge  $\ell(\mathbf{r}, \theta)$ .

Given a root  $\mathfrak{t} \in \mathfrak{T}$  let us denote by  $\mathfrak{D}_{\mathfrak{t}}$  the corresponding component of  $\Gamma_S^{(f)}$  and by  $\mathcal{D}_{\mathfrak{t}}$ , the *support* of  $\mathfrak{D}_{\mathfrak{t}}$  to be the set of the  $k \in S^c$  such that  $(\mathbf{r}(k), \theta_k) \in \mathfrak{D}_{\mathfrak{t}}$ . Notice that by construction this is a union of some of the subsets  $S_{\mathbf{r}, \theta}^c$  defined by the Fitting decomposition. If  $k \in \mathcal{D}_{\mathfrak{t}}$  we set  $\mathfrak{t}(k) = (\mathbf{r}_{\mathfrak{t}}(k), \theta_{\mathfrak{t}}(k)) := \mathfrak{t}$ . In general  $\mathbf{r}_{\mathfrak{t}}(k) \neq \mathbf{r}(k)$ ,  $\theta_{\mathfrak{t}}(k) \neq \theta(k)$ .

**Corollary 2** *The sets  $\mathcal{D}_{\mathfrak{t}}$  decompose the connected components of  $\tilde{\Gamma}_{S, X_q \cup Y}$ . This decomposition of the components with only black edges is invariant under translations.*

Set  $\ell_k$  through the Formulas

$$\mathbf{j} \cdot \ell_k + \mathbf{r}(k) = (\eta(\ell_k) + 1)\mathbf{r}_{\mathfrak{t}}(k), \quad \omega^{(1)}(\xi) \cdot \ell_k + \theta(k) = (\eta(\ell_k) + 1)\theta_{\mathfrak{t}}(k). \quad (78)$$

We also define a new color  $s(k) := \sigma(k)(\eta(\ell) + 1) = \pm 1$ . Note that we can reformulate the Formulas (63), in terms of  $L_k, \ell_k, s(k), \mathfrak{t}(k)$  by substituting in those Formulas the expressions given in Formulas (76), (77).

We see now that we have a parallel with 15. Take a degree 2 monomial  $\mathbf{m}$  product of two eigenvectors as in the previous discussion.

**Proposition 12** *A monomial  $\mathbf{m} = e^{i\nu \cdot x} z_h^\sigma z_k^{\sigma'}$  with  $\nu \neq 0$  is in the kernel of  $ad(\mathcal{N}^s)$  if and only if  $(\mathbf{r}(h), \theta_h), (\mathbf{r}(k), \theta_k)$  are connected by an edge  $\ell$  and  $\nu = \pm \ell$  (here the sign is determined by  $\sigma, \sigma' \sigma(k), \sigma(h)$ ).*

*In Formula (74) the edge is a red edge while in Formula (75) it is a black edge.*

*Proof* We apply the commutation rules; we only need to show that the possible monomials have  $\nu \neq -2e_i$  for all  $i$ . Indeed in this case we see that the equations for the conservation of  $\mathbb{M}$  and  $\mathbb{K}$  have the unique solution  $r(h) = \mathbf{j}_i \notin S^c$ , a contradiction.  $\square$

## 7.2 A new phase shift

At this point we can perform a new symplectic change of variables which is done with the same method of phase shift as in Formula (54).

We now define  $\Psi$ :

$$z_k = e^{-i\sigma(k)\ell_k \cdot x} z'_k, \quad \forall k \in S^c, \\ x = x', \quad y = y' + \sum_{k \in S^c} \sigma(k)\ell_k |z'_k|^2.$$

**Proposition 13** *i)  $\Psi$  is symplectic analytic and leaves  $\mathbb{M}, \mathbb{L}, \mathbb{K}, \mathbb{K}_1$  and  $\mathcal{N}$  in normal form. ii) For any function  $f \in \mathcal{R}_{s,r}$  of degree  $\leq 2$  and such that  $f$  Poisson commutes with  $\mathcal{N}^s$  we have that  $f \circ \Psi$  does not depend on  $x$ .*

*Proof* We first prove item ii). Since  $f$  is an absolutely convergent sum of monomials we only need to prove our claim on single monomials  $\mathbf{m}$ . If the degree in  $w$  is zero or one the statement follows trivially (by substituting the new variables). In case of degree two we substitute and get in one case

$$e^{i\sigma\sigma(h)\nu\cdot x} z_h^\sigma z_k^{\sigma'} = e^{i(\sigma\sigma(h)\nu - \sigma\sigma(h)\ell(h) - \sigma'\sigma(k)\ell(k))\cdot x} (z'_h)^\sigma (z'_k)^{\sigma'} = (z'_h)^\sigma (z'_k)^{\sigma'}$$

by applying Lemma 6 to the vertices  $(\mathfrak{A}, \mathfrak{t})$ ,  $(\mathfrak{r}(h), \theta_h)$ ,  $(\mathfrak{r}(k), \theta_k)$  and recalling that the last two vertices are joined by an edge marked  $\nu$ . The other case is identical.

It is possible that  $f$  contains monomials with  $\nu = 0$ . Then by Remark 12  $(\mathfrak{r}(h), \theta_h) = (\mathfrak{r}(k), \theta_k)$  and  $\sigma\sigma(h) = \sigma'\sigma(k)$  and such term remains  $x$  independent.

i) We can compute the new Hamiltonians explicitly by substitution:

$$\begin{aligned} \mathbb{L} &= \sum_i y_i + \sum_k \sigma(k) |z_k|^2 = \sum_i y'_i + \sum_{k \in S^c} \sigma(k) \eta(\ell_k) |z_k|^2 + \sum_k \sigma(k) |z_k|^2 \\ &= \sum_i y'_i + \sum_k s(k) |z'_k|^2 \\ \mathbb{M} &= \mathfrak{j} \cdot y + \sum_k \sigma(k) \mathfrak{r}(k) |z_k|^2 = \mathfrak{j} \cdot y' + \sum_{k \in S^c} \sigma(k) \mathfrak{j} \cdot \ell_k |z_k|^2 + \sum_k \sigma(k) \mathfrak{r}(k) |z_k|^2 \\ &= \mathfrak{j} \cdot y' + \sum_k s(k) \mathfrak{r}_t(k) |z'_k|^2 = \sum_{t \in \mathfrak{T}} \mathfrak{r}_t \left( \sum_{k \in \mathbb{D}_t} s(k) |z'_k|^2 \right). \end{aligned}$$

As for the nilpotent part  $\tilde{Q}^n$  in  $\mathcal{N}$  we notice that it is a sum of monomials as in Remark 12, so  $\tilde{Q}^n$  remains block diagonal and  $x$  independent by the same argument as in item ii).  $\square$

In order to simplify the notations we now drop the apex  $'$  and write  $y, z$  the new coordinates.

**Corollary 3** For  $\xi \in \mathcal{A}_\alpha$ , in the final coordinates the normal form of the Hamiltonian has the form

$$\mathcal{N} = \mathbb{K} + \mathbb{K}^1, \quad \mathbb{K}^1 = \omega^{(1)}(\xi) \cdot y + \sum_{t \in \mathfrak{T}_f} \mathcal{Q}_t + \sum_{t \in \mathfrak{T}_g} \theta_t \left( \sum_{k \in \mathbb{D}_t} |z_k|^2 \right) \quad (79)$$

$$\mathbb{K} = \sum_i |j_i|^2 y_i + \sum_{t \in \mathfrak{T}} |\mathfrak{r}_t|^2 \left( \sum_{k \in \mathbb{D}_t} s(k) |z_k|^2 \right). \quad (80)$$

As for the finitely many  $\mathcal{Q}_t$  each corresponds to a Hamiltonian which can be split in Jordan decomposition, with a scalar and a nilpotent part  $\mathcal{Q}_t^{\text{nil}}$ :

$$\mathcal{Q}_t = \mathcal{Q}_t^{\text{nil}} + \theta_t \left( \sum_{k \in \mathbb{D}_t} s(k) |z_k|^2 \right), \quad \forall t \in \mathfrak{T}_f.$$

The associated matrix corresponds to a map on two Lagrangian blocks each with a single non-zero eigenvalue  $\pm i\theta_t$  but contains possibly also some nilpotent part.

*Proof* When we substitute in Formula (72), setting  $\mathfrak{S}_f^c$  the support of the finitely many blocks in  $\mathfrak{T}_f$

$$\mathbb{K}^1 = \omega^{(1)}(\xi) \cdot y + \sum_{(\mathfrak{r}, \theta) \in \mathfrak{S}_f^c} \tilde{\mathcal{Q}}_{\mathfrak{r}, \theta} + \sum_{(\mathfrak{r}, \theta) \notin \mathfrak{S}_f^c} \theta \left( \sum_k |z_k|^2 \right)$$

$$= \omega^{(1)}(\xi) \cdot y' + \sum_{k \in S^c} \sigma(k) \omega^{(1)}(\xi) \cdot \ell_k |z'_k|^2 + \sum_{(\mathbf{r}, \theta) \in \mathfrak{S}_f^c} \tilde{Q}_{\mathbf{r}, \theta} + \sum_{(\mathbf{r}, \theta) \notin \mathfrak{S}_f^c} \theta \left( \sum_k |z'_k|^2 \right).$$

For the  $k$  in the support of  $\mathfrak{F}_g$  we have  $\sigma(k) = 1$ ,  $\eta(\ell_k) = 0$ , the coefficient of  $|z'_k|^2$  comes from two terms,  $\omega^{(1)}(\xi) \cdot \ell_k$  and  $\theta_k$  the sum is by Formula (78),  $\theta_{\mathbf{t}}(k) = \theta_{\mathbf{t}}$  which gives the last term of Formula (79), Formula (80) is similar.

The remaining terms are collected in the finite sum namely, for a given  $\mathbf{t} \in \mathfrak{F}_f$  the contribution to  $\mathcal{Q}_{\mathbf{t}}$  is the sum of the sum  $\sum_{(\mathbf{r}, \theta) \in D_{\mathbf{t}}} \tilde{Q}_{\mathbf{r}, \theta}$  plus the sum  $\sum_{k \in D_{\mathbf{t}}} \sigma(k) \omega^{(1)}(\xi) \cdot \ell_k |z'_k|^2$ . Formula (78) gives the statement on the eigenvalues.  $\square$

Remark also that, if  $k \in D_{\mathbf{t}}$ ,  $\mathbf{t} \in \mathfrak{F}_g$  then  $s(k) = 1$ . Thus  $\mathcal{N}^s$  is the sum of the terms

$$\sum_i (|j_i|^2 + \omega_i^{(1)}(\xi)) y_i + \sum_{\mathbf{t} \in \mathfrak{F}} (|\mathbf{r}(\mathbf{t})|^2 + \theta_{\mathbf{t}}) \left( \sum_{k \in D_{\mathbf{t}}} s(k) |z_k|^2 \right). \quad (81)$$

That is in  $y$ , given by  $\sum_i (|j_i|^2 + \omega_i^{(1)}(\xi)) y_i$ , the infinite real diagonal part given by  $\sum_{\mathbf{t} \in \mathfrak{F}_g} (|\mathbf{r}(\mathbf{t})|^2 + \theta_{\mathbf{t}}) \left( \sum_{k \in D_{\mathbf{t}}} |z_k|^2 \right)$  and finally the finite semisimple part involving the *bad* sites  $\sum_{\mathbf{t} \in \mathfrak{F}_f} (|\mathbf{r}(\mathbf{t})|^2 + \theta_{\mathbf{t}}) \left( \sum_{k \in D_{\mathbf{t}}} s(k) |z_k|^2 \right)$ . These finitely many blocks may also contribute to the nilpotent part of  $\mathcal{N}$ .

Recall that  $\sum_{k \in D_{\mathbf{t}}} s(k) |z_k|^2$  acts as the scalar  $i, -i$  on the two corresponding Lagrangian spaces, since a basis of one of them is  $z_k^{s(k)}$  (and the other the conjugate).

$F^{0,1}$  and the algebra  $\mathcal{F}_{\ker}$ . We now decompose  $F^{0,1}$  as orthogonal sum (with respect to the symplectic form), with blocks each decomposed as sum of two conjugate Lagrangian subspaces,  $(D_{\mathbf{t}}, \nu)^+ \oplus (D_{\mathbf{t}}, \nu)^-$  with  $\mathbf{t} \in \mathfrak{F}$  and  $\nu \in \mathbb{Z}^n$  constrained by the conservation laws.

For each index  $\mathbf{t}, \nu$  the constraints

$$\pi(\nu) + \mathbf{r}_{\mathbf{t}} = \nu \cdot \mathbf{j} + \mathbf{r}_{\mathbf{t}} = 0, \quad \eta(\nu) + 1 = 0$$

express respectively the conservation of momentum and mass. We have as basis for  $(D_{\mathbf{t}}, \nu)^+$  the elements  $e^{i\nu \cdot x} z_k^{s(k)}$  where  $k \in D_{\mathbf{t}}$  and for  $(D_{\mathbf{t}}, \nu)^-$  the conjugate elements.

These blocks are obviously stable under  $ad(\mathcal{N})$  and the action is deduced from corollary 3, note that on each block this action is invertible.

The product of two blocks  $(D_{\mathbf{t}_i}, \nu_i)^{\pm 1}$ ,  $i = 1, 2$  produces a quadratic block in  $F^{0,2}$  stable under  $ad(\mathcal{N})$  with basis the products of basis elements. In order to avoid repetitions we index different quadratic blocks by  $(\nu, \mathbf{t}_1, \mathbf{t}_2, \sigma_1, \sigma_2)$  where the conservation laws are:

$$\eta(\nu) + 1 + \sigma_1 \sigma_2 = 0, \quad \pi(\nu) + \mathbf{r}_{\mathbf{t}_1} + \sigma_1 \sigma_2 \mathbf{r}_{\mathbf{t}_2} = 0$$

while the basis elements are  $e^{i\sigma_1 \nu \cdot x} z_k^{\sigma_1 s(k)} z_h^{\sigma_2 s(h)}$  where  $k \in D_{\mathbf{t}_1}$  and  $h \in D_{\mathbf{t}_2}$ .

For  $\sigma_1 \sigma_2$  fixed the blocks come in conjugate pairs and we usually exhibit the one with  $\sigma_1 = 1$ .

**Proposition 14** *The action of  $ad(\mathcal{N})$  is invertible on all the blocks different from  $(0, \mathbf{t}, \mathbf{t}, \sigma, -\sigma)$ .*

*Proof* It is enough to prove that the action of the semisimple part  $ad(\mathcal{N}^s)$  is invertible on all these blocks.

Now from Formula (81) on a block  $(\nu, \mathfrak{t}, \mathfrak{t}', \sigma, \sigma')$  the action is via the scalar function

$$\sum_i (|j_i|^2 + \omega_i^{(1)}(\xi))\nu_i + \sigma(|\mathbf{r}(\mathfrak{t})|^2 + \theta_{\mathfrak{t}}) + \sigma'(|\mathbf{r}(\mathfrak{t}')|^2 + \theta_{\mathfrak{t}'})$$

so we only need to observe that if  $(\nu, \mathfrak{t}, \mathfrak{t}', \sigma, \sigma')$  is different from  $(0, \mathfrak{t}, \mathfrak{t}, \sigma, -\sigma)$  this function is non zero. If this expression equals 0 we must have separately

$$\sum_i |j_i|^2 \nu_i + \sigma|\mathbf{r}(\mathfrak{t})|^2 + \sigma'|\mathbf{r}(\mathfrak{t}')|^2 = 0, \quad \sum_i \omega_i^{(1)}(\xi)\nu_i + \sigma\theta_{\mathfrak{t}} + \sigma'\theta_{\mathfrak{t}'} = 0.$$

Now we also have conservation of momentum  $\pi(\nu) + \sigma\mathbf{r}(\mathfrak{t}) + \sigma'\mathbf{r}(\mathfrak{t}') = 0$ .

These properties then imply that the two roots  $\mathbf{r}(\mathfrak{t}), \mathbf{r}(\mathfrak{t}')$  are joined by the edge  $\nu$  but in the final graph this implies that  $\nu = 0$  and  $\mathbf{r}(\mathfrak{t}) = \mathbf{r}(\mathfrak{t}')$  the final graph was built so that these equalities imply that  $(\nu, \mathfrak{t}, \mathfrak{t}', \sigma, \sigma') = (0, \mathfrak{t}, \mathfrak{t}, \sigma, -\sigma)$ .  $\square$

This Proposition allows us to define the decomposition of  $F^{(0,2)}$  given by Definition 7 into the subalgebra  $\mathcal{F}_{\ker}$  and its unique complement  $\mathcal{F}_{\text{rg}}$  stable under adjoint action of  $\mathcal{F}_{\ker}$ .

## A Measure estimates

**Lemma 7** *Consider a  $C^q$  function  $f$  of  $n$  variables in a domain  $A$  contained in some hypercube of side of length  $\zeta$ . If for some  $k \in \mathbb{N}^n$  such that  $|k|_1 \leq q$  one has*

$$\inf_{x \in A} |\partial_x^k f| \geq c^{|k|} > 0$$

then for all  $\alpha \geq 0$  we have, setting:

$$A_0 := \{x \in A : |f(x)| \leq \alpha^{|k|}\}, \quad \text{meas}(A_0) \leq 2|k|\zeta^{n-1}\alpha c^{-1}. \quad (82)$$

*Proof* By induction. If  $|k| = 1$  this is a standard argument (which one can develop also for Lipschitz functions). Without loss of generality let us suppose that  $k = e_1$ . We consider the map  $F : x \mapsto (f(x), x_2, \dots, x_n)$  which, since  $|\partial_{x_1} f| > 0$ , is a diffeomorphism from  $A_0$  to some set  $B$  which is contained in the product of an  $n-1$ -dimensional hypercube with size of length  $\zeta$  times a segment of length  $2\alpha$ , hence the volume of  $F(A_0)$  is  $\leq \text{meas}(B) = 2\alpha\zeta^{n-1}$ .

The Jacobian of  $F^{-1}$  is  $\partial_x f^{-1}$  so in absolute value is  $\leq c^{-1}$ . Therefore the volume of the set  $A_0 = F^{-1}(F(A_0))$  can be estimated by  $2\alpha\zeta^{n-1}c^{-1}$ .

Let us now write (again without loss of generality)  $\partial_x^k = \partial_x^h \partial_{x_1}$  (where  $|h| = |k| - 1$ ) and set  $g(x) := c^{-1} \partial_{x_1} f(x)$ . So that we know

$$\inf_{x \in A} |\partial_x^h g| \geq c^{|h|}.$$

Then by the inductive hypothesis

$$A_1 := \{x \in A : |g(x)| < \alpha^{|h|}\}, \quad \text{meas}(A_1) \leq 2|h|\zeta^{n-1}\alpha c^{-1}.$$

On the region  $A \setminus A_1$  we have  $c|g(x)| = |\partial_{x_1} f(x)| \geq c\alpha^{|h|}$ . Then by case  $|k| = 1$  we have

$$A_2 := \{x \in A \setminus A_1 : |f(x)| \leq \alpha^{|k|}\}, \quad \text{meas}(A_2) \leq 2\zeta^{n-1}\alpha^{|k|}c^{-1}\alpha^{-|h|} = 2\zeta^{n-1}\alpha c^{-1}$$

and this concludes the proof since  $A_2 = A_0 \setminus A_1$  so  $A_0 \subset A_1 \cup A_2$  and:

$$\text{meas}(A_0) \leq \text{meas}(A_1) + \text{meas}(A_2) \leq 2|h|\zeta^{n-1}\alpha c^{-1} + 2\zeta^{n-1}\alpha c^{-1} = 2|k|\zeta^{n-1}\alpha c^{-1}.$$

$\square$

We shall often apply this Lemma as follows, we have our domain  $\mathcal{O}$  in the coordinates  $\xi$  and its image  $\tilde{\mathcal{O}}$  in the coordinates  $\omega$ . We want perform an estimate of a set  $A_0 \subset \mathcal{O}$  as before. We first perform the estimate for the corresponding  $B_0 \subset \tilde{\mathcal{O}}$  which satisfies the hypotheses of Lemma 82 with  $\zeta = \varepsilon^{2q}$ . The dilation factor of the inverse of the transformation  $\xi \mapsto \omega$  is bounded by  $L^n \varepsilon^{-2n(q-1)}$  (Remark 8). In all the estimates which we shall perform we have  $\alpha = \varepsilon^{2q}\beta$  where usually  $\beta = K^{-c\rho}$ . We deduce

**Corollary 4** *If the measure of  $B_0 = \omega(A_0)$  is bounded by  $C\zeta^{n-1}\alpha = C\varepsilon^{2q(n-1)}\varepsilon^{2q}\beta = C\varepsilon^{2qn}\beta$  that of  $A_0$  is bounded by  $C\beta L^n \varepsilon^{-2n(q-1)}\varepsilon^{2qn} \leq C'\varepsilon^{2n}\beta$ .*

Let  $f, g$  be inverse functions between two domains in  $\mathbb{R}^n$ , i.e.  $g \circ f = 1$  and both diffeomorphisms of some class  $C^p$ , we write in coordinates  $\omega_i = f_i(\xi_1, \dots, \xi_n)$  and  $\xi_i = g_i(\omega_1, \dots, \omega_n)$ . Then there are universal formulas due to Faà di Bruno, to express the derivatives  $b_\alpha := \partial_\xi^\alpha g$  as polynomials in terms of  $a_\beta := \partial_\xi^\beta f$  and the first derivatives  $b_i := \partial_{\omega_i} g$  (computed at the corresponding points,  $\xi, \omega = f(\xi)$ ).

We do not need the explicit formulas but only to know that the nature of such a polynomial is the following. We give a *weight*  $|\beta|$  to  $a_\beta$  and  $-1$  to  $b_i$ . Then the polynomial expressing  $b_\alpha$  is a linear combination with rational coefficients of monomials of weight  $-1$ .

So let  $h = |\alpha|$ , in each monomial we have some number  $k$  of factors  $b_i$  which contribute  $-k$  to the weight and other factors of type  $a_\beta$  of total weight  $k - 1$ , the total number  $k$  of factors  $b_i$  can be seen to be bounded by  $2h - 1$ .

We assume to have some estimates  $|\partial_\xi^\beta \omega| = |a_\beta| \leq \varepsilon^{2q-2|\beta|} M^{|\beta|}$ ,  $|\partial_{\omega_i} \xi| = |b_i| \leq \varepsilon^{2-2q} L$  and  $LM < c$ ,  $c > 1$ . In our setting these estimates come from Proposition 11 and are justified by the fact that we work on a domain of width  $\varepsilon^2$  and we start the algorithm from  $\omega$  homogeneous of degree  $q$ .

Under these estimates we obtain for  $b_\alpha = \partial_\omega^\alpha \xi$  an estimate, with  $h = |\alpha|$ , (and  $\varepsilon < 1$ )

$$|b_\alpha| \leq C\varepsilon^2 L \varepsilon^{-2hq} (LM)^{2(h-1)} \leq C\varepsilon^{2(1-hq)} L c^{2h-2} \quad (83)$$

where the order comes from estimating the worst monomial in the polynomial expression of  $b_\alpha$ . The constant  $C$  is some positive number, sum of the absolute values of the coefficients of this combinatorial expression and so depending only on  $n$  and  $\beta$ . Summarising

**Lemma 8** *Let  $P(\xi)$  be such that  $|\partial_\xi^h P(\xi)| \leq \varepsilon^{2(q-|h|)} M^{|h|}$  and consider the map  $\xi \rightarrow \omega(\xi)$  with derivatives  $|\partial_\xi^h \omega| \leq \varepsilon^{2(q-|h|)} M^{|h|}$  and its inverse  $\xi(\omega)$  with  $|\partial_\omega \xi| \leq \varepsilon^{2(1-q)} L$ . Then if  $LM \leq c$*

$$\begin{aligned} |\partial_\omega^j \xi| &\leq C\varepsilon^{2(1-|j|q)} L c^{2|j|-2} \\ |\partial_\omega^j P(\xi(\omega))| &\leq C'\varepsilon^{2q(1-|j|)} c^{2|j|-2} \end{aligned} \quad (84)$$

*Proof* The first is formula (83) and the second follows from the chain rule.  $\square$

**Lemma 9** *Consider a  $\mathfrak{p} \times \mathfrak{p}$  matrix  $A(\omega)$  and given  $\nu \in \mathbb{Z}^n$ , set*

$$f(\omega) = \det(\omega \cdot \nu I + A(\omega))$$

*and suppose that  $|\omega \cdot \nu| \leq \mathfrak{p}M\varepsilon^{2q}$  and*

$$|\partial_{\omega_1}^j A|_\infty \leq C_1 \varepsilon^{2q(1-|j|)} \quad (85)$$

*for some  $C_1 \geq \mathfrak{p}M$ . Then there exists  $C_2$  (depending on  $C_1, q, \mathfrak{p}$ ) such that if  $|\nu_1| > C_2$  we have*

$$|\partial_{\omega_1}^{\mathfrak{p}} f| > C_2^{\mathfrak{p}}$$

*Proof* Clearly

$$f(\omega) = P_A(\omega \cdot \nu) = (\omega \cdot \nu)^{\mathfrak{p}} + (\omega \cdot \nu)^{\mathfrak{p}-1} \text{tr}(A) + \dots + \det(A).$$

For the derivative we have that  $\partial_{\omega_1}^{\mathfrak{p}} [(\omega \cdot \nu)^{\mathfrak{p}-i} \sigma_i(A)]$  is a sum of terms  $\partial_{\omega_1}^h (\omega \cdot \nu)^{\mathfrak{p}-i} \partial_{\omega_1}^{\mathfrak{p}-h} \sigma_i(A)$ , with  $|h| \leq \mathfrak{p} - i$ . We estimate

$$|\partial_{\omega_1}^h (\omega \cdot \nu)^{\mathfrak{p}-i}| = |h! \nu_1^h (\omega \cdot \nu)^{\mathfrak{p}-i-h}| \leq h! \nu_1^h (\mathfrak{p}M\varepsilon^{2q})^{\mathfrak{p}-i-h}.$$

As for  $\partial_{\omega_1}^{\mathfrak{p}-h} \sigma_i(A)$  we have that  $\sigma_i(A)$  is a sum, with signs, of  $\binom{\mathfrak{p}}{i}$  products of  $i$  entries of  $A$ . For such a product  $\partial_{\omega_1}^{\mathfrak{p}-h}$  is again a sum of some  $C(h, i)$  products where each factor has been derived some  $u_i \geq 0$  times with  $\sum u_i = \mathfrak{p} - h$ . For these factors we apply the estimate (85) getting an estimate  $C_1^i \varepsilon^{2q(i-\mathfrak{p}+h)}$  for their product. Thus for each  $|\partial_{\omega_1}^h(\omega \cdot \nu)^{\mathfrak{p}-i} \partial_{\omega_1}^{\mathfrak{p}-h} \sigma_i(A)|$  we have an estimate  $h! \binom{\mathfrak{p}}{i} \nu_1^h (\mathfrak{p}M)^{\mathfrak{p}-i-h} C_1^i \leq h! \binom{\mathfrak{p}}{i} \nu_1^h C_1^{\mathfrak{p}-h}$ . Then:

$$|\partial_{\omega_1}^{\mathfrak{p}} f| \geq \mathfrak{p}! \nu_1^{\mathfrak{p}} - \sum_{|j| \leq \mathfrak{p}-1} \kappa_j |\nu_1|^j C_1^{\mathfrak{p}-|j|} \geq (C_2)^{\mathfrak{p}} \mathfrak{p}! - \sum_j \kappa_j C_2^{|j|} C_1^{\mathfrak{p}-|j|}.$$

The coefficients  $\kappa_j$  are purely combinatorial so given  $C_1$  we just need to choose  $C_1/C_2$  so that

$$\sum_j \kappa_j \left(\frac{C_1}{C_2}\right)^{\mathfrak{p}-|j|} < \mathfrak{p}!/2.$$

□

In our compact set  $\mathcal{O} \subset \varepsilon^2 \Lambda \subset \mathbb{R}^n$  we now consider the pointwise  $\lambda$  norm associated to the  $C^\ell$  maps from  $\mathcal{O} \rightarrow \mathbb{C}^{\mathfrak{p}}$  with the  $L_\infty$  norm,  $\|b(\xi)\|_\lambda := \sum_{|\alpha| \leq \ell} \lambda^{|\alpha|} |\partial_\xi^\alpha b(\xi)|_\infty$ .

**Lemma 10** *Take a  $\mathfrak{p} \times \mathfrak{p}$  matrix  $A(\xi)$  satisfying the bounds:*

$$|\partial_\xi^\alpha A|_\infty \leq \lambda^{-|\alpha|} \varepsilon^{2q}, \quad \forall |\alpha| \leq \ell$$

and a  $\mathfrak{p}$  vector  $b(\xi)$  with finite  $\lambda$  norm. Suppose now that  $A$  is invertible for all  $\xi \in \mathcal{O}$  and that  $|A^{-1}|_2 \leq \varepsilon^{-2q} K^e$  then  $y(\xi) = A^{-1}(\xi)b(\xi)$  satisfies the bounds

$$\|y\|_\lambda \leq \ell^n 2^{\ell^2} \mathfrak{p}^{3\ell+1} \varepsilon^{-2q} K^{(\ell+1)e} \|b\|_\lambda.$$

*Proof* We preliminarily note that

$$|\partial_\xi^\alpha b|_\infty \leq \lambda^{-|\alpha|} \|b\|_\lambda.$$

Then we derive  $\alpha$  times the equation  $Ay = b$  and obtain

$$A \partial_\xi^\alpha y = \partial_\xi^\alpha b - \sum_{\substack{j_1+j_2=\alpha, \\ j_1 \neq 0}} c_{j_1, j_2} (\partial_\xi^{j_1} A) \partial_\xi^{j_2} y, \quad 1 + \sum_{\substack{j_1+j_2=\alpha, \\ j_1 \neq 0}} c_{j_1, j_2} = 2^{|\alpha|}.$$

This gives the bound

$$|\partial_\xi^\alpha y|_\infty \leq |A^{-1}|_2 \left( \mathfrak{p} |\partial_\xi^\alpha b|_\infty + \mathfrak{p}^2 \sum_{j_1+j_2=\alpha, j_1 \neq 0} c_{j_1, j_2} |\partial_\xi^{j_1} A|_\infty |\partial_\xi^{j_2} y|_\infty \right)$$

We now prove by induction that

$$|\partial_\xi^\beta y| \leq \varepsilon^{-2q} \lambda^{-|\beta|} 2^{\ell|\beta|} \mathfrak{p}^{3|\beta|+1} K^{(|\beta|+1)e} \|b\|_\lambda$$

by simply substituting in the right hand side of the formula all our bounds

$$\begin{aligned} |\partial_\xi^\alpha y|_\infty &\leq \varepsilon^{-2q} K^e \left( \mathfrak{p} \lambda^{-|\alpha|} \|b\|_\lambda + \right. \\ &\mathfrak{p}^2 \sum_{\substack{j_1+j_2=\alpha \\ j_1 \neq 0}} c_{j_1, j_2} \lambda^{-|j_1|} \varepsilon^{2q} \varepsilon^{-2q} \lambda^{-|j_2|} 2^{\ell|j_2|} \mathfrak{p}^{3|j_2|+1} K^{(|j_2|+1)e} \|b\|_\lambda \left. \right)^{|j_1|+|j_2|=|\alpha|} \\ &\varepsilon^{-2q} \lambda^{-|\alpha|} \|b\|_\lambda K^e \left( \mathfrak{p} + \sum_{\substack{j_1+j_2=\alpha \\ j_1 \neq 0}} c_{j_1, j_2} 2^{\ell|j_2|} \mathfrak{p}^{3(|j_2|+1)} K^{(|j_2|+1)e} \right)^{|j_2| \leq |\alpha|-1} \\ &\leq \varepsilon^{-2q} \lambda^{-|\alpha|} \|b\|_\lambda K^{(|\alpha|+1)e} (\mathfrak{p} + \mathfrak{p}^{3|\alpha|} 2^{\ell(|\alpha|-1)} \sum_{\substack{j_1+j_2=\alpha \\ j_1 \neq 0}} c_{j_1, j_2}) \leq \\ &\leq \varepsilon^{-2q} \lambda^{-|\alpha|} \|b\|_\lambda K^{(|\alpha|+1)e} (\mathfrak{p} + \mathfrak{p}^{3|\alpha|} (2^{\ell|\alpha|} - 1)) \leq \varepsilon^{-2q} \lambda^{-|\alpha|} 2^{\ell|\alpha|} \mathfrak{p}^{3|\alpha|+1} K^{(|\alpha|+1)e} \|b\|_\lambda. \end{aligned}$$

multiplying by  $\lambda^{|\alpha|}$  and summing over  $\alpha$  (at most  $\ell^n$ ) we obtain the desired estimate. □

### A.0.1 NLS estimates

Consider the NLS equation restricted to some domain  $\mathcal{R}_\alpha \cap \varepsilon^2 \Lambda$ . This defines  $\omega$  and  $\Omega_t$  by setting  $\tilde{\Omega}_t = 0$ . By [14] we know that the perturbation  $P$  is quasi-Töplitz for some parameters  $K_0, \theta_0, \mu_0$ .

**Lemma 11** *i) There exist constants  $L_0, M_0$  so that formulæ (33) hold.*

*ii) Given  $S_0$ , we can choose the domain  $\mathcal{O}_0$  and a constant  $a_0$  so that Formulæ (34) hold for  $\omega, \Omega_t$  defined by  $\mathcal{N}$  with  $\mathcal{O} = \mathcal{O}_0$  and  $a = a_0$ .*

*Proof* i) We have that the functions  $\omega - j^{(2)}$  and its inverse are homogeneous and invertible, moreover the functions  $\theta_t$  and  $\Omega_t^{nil}$  are also homogeneous and from a finite list so we satisfy these formulas by taking the appropriate maximum on a compact domain of the unit sphere contained in some  $\mathcal{R}_\alpha$  and such that its cone from the origin contains the domain  $\mathcal{O}_0$ .

ii) By the Melnikov conditions, Proposition 14, the matrices under consideration are all invertible as functions outside the algebraic hypersurfaces where the determinant is 0. All these matrices evaluated at  $\xi = 0$  are an integer multiple of the identity, the remaining part is homogeneous and runs in a finite list. If this integer is non-zero then the estimates hold provided  $\varepsilon$  is small enough. Otherwise we have to restrict the domain so to avoid the finite number of hypersurfaces where one of these determinants vanish and estimate on the unit sphere.  $\square$

As for  $S_0$  we need to fix it as  $nC_2$  so to apply lemma 9 at all steps where  $A(\omega)$  are the matrices, appearing in Formulæ (40) in the coordinates  $\omega$ .

*Proof (Conclusion of the proof of Lemma 2)* By the non-degeneracy condition insured by Lemma 11, we may assume  $|\nu| > S_0$  so one of the coordinates  $\nu_i$  has  $|\nu_i| > S_0/n$  and we may assume this happens for  $\nu_1$ . We will fix  $S_0$  sufficiently large as seen in the course of the proof.

If  $\sigma = \sigma' = 0$  we claim that we remove from the set  $\mathcal{O}$  a region of order  $\varepsilon^{2n} K^{-e}$ . For this follow the strategy of Remark 8. Namely, apply Lemma 7 to  $f(\xi) = \omega \cdot \nu$  we then estimate the measure of the set  $A_0$  where  $|\omega \cdot \nu| < \varepsilon^{2q} K^{-e}$  as follows.

First by Formula (82) the measure of the image  $B_0$  of  $A_0$  under  $\omega$  is bounded by  $2\zeta^{n-1} \varepsilon^{2q} K^{-e} c^{-1}$  where  $\zeta$  is the size of the side of a hypercube containing  $B_0$  and  $c$  is a lower bound for one of the derivatives  $\nu_i = \partial_{\omega_i}(\omega \cdot \nu)$ . We may take for  $c$  the maximum of the absolute values of the  $\nu_i$  which is  $> S_0/n$ . As for  $\zeta$  we have  $\zeta \leq M\varepsilon^{2q}$  so we have a bound

$$\text{meas}(A_0) \leq L^n \varepsilon^{2n(q-1)} \text{meas}(B_0) \leq L^n \varepsilon^{-2n(q-1)} (M\varepsilon^{2q})^{n-1} K^{-e} \varepsilon^{2q} c^{-1} \ll \varepsilon^{2n} K^{-e}.$$

If  $t, t'$  are both in  $\mathfrak{I}_g$  we set for  $\sigma = 0, \pm 1$

$$f(\omega) = \omega \cdot \nu + \lambda_t^{(i)} + \sigma \lambda_{t'}^{(j)}.$$

By definition of  $M$  we have  $|\partial_\xi \lambda_t^{(j)}| \ll \varepsilon^{2q-1} M$  so by applying Lemma A.2. we know that  $|\partial_{\omega_1} f(\omega)| = |\nu_1 + \partial_{\omega_1}(\lambda_t^{(i)} + \sigma \lambda_{t'}^{(j)})| \geq |\nu_1| - |\partial_{\omega_1}(\lambda_t^{(i)} + \sigma \lambda_{t'}^{(j)})| \geq S_0/n - C\varepsilon^{2q-1}$  for some constant  $C$ , so taking  $S_0$  sufficiently large the proof is concluded by applying Lemma 7 with  $x$  replaced by  $\omega$ .

We get that the resonant set in the variables  $\omega$  has measure of order  $\varepsilon^{2q(n-1)+2q} K^{-e} = \varepsilon^{2qn} K^{-e}$ . In order to obtain the estimates in the variables  $\xi$  we multiply by the dilation factor  $L^n \varepsilon^{2n-2nq}$  getting  $\ll \varepsilon^{2n} K^{-e}$ .

When either  $t$  or  $t'$  are in  $\mathfrak{I}_f$  we need to study

$$f(\omega) = \det((\omega \cdot \nu)I + L(\Omega_t) + \sigma R(\Omega_{t'})).$$

This is a matrix of dimension  $\mathbf{p} = \delta_t \delta_{t'}$  (resp. of dimension  $\mathbf{p} = \delta_t$  if  $\sigma = 0$ ). By hypothesis setting  $F(\xi) = L(\Omega_t(\xi)) + \sigma R(\Omega_{t'}(\xi))$  we get  $|\partial_\xi^\alpha F|_\infty \leq 2\varepsilon^{2q-2|\alpha|} M$ .

We have that  $A(\omega) = F(\xi(\omega))$  satisfies estimates as in (85) and under the hypothesis  $|\omega \cdot \nu| \leq \mathbf{p} M \varepsilon^{2q}$  we can apply Lemma A.2. Since we are assuming  $|\nu_1| > S_0/n$  we only need to choose  $S_0$  large enough in order to prove  $|\partial_{\omega_1}^\mathbf{p} f| \geq C_2$  some positive constant.

By our definitions the resonance condition is  $|f(\omega)| \geq \varepsilon^{2q\mathbf{p}} K^{-e+1}$  where  $\mathbf{p} = d_t d_{t'}$ . Then we apply Lemma 7 with  $|k| = \mathbf{p}$  and  $\alpha = \varepsilon^{2q} K^{-e/\mathbf{p}}$ . We get that the resonant set in the variables  $\omega$  has measure of order  $2\mathbf{p} \zeta^{n-1} \varepsilon^{2q} K^{(-e+1)/\mathbf{p}}$ . So by Corollary 4 the resonant set in the variables  $\xi$  has measure of order  $\varepsilon^{2n-2q} \varepsilon^{2q} K^{-e/\mathbf{p}} = \varepsilon^{2n} K^{(-e+1)/\mathbf{p}}$ .  $\square$

Consider a normed space of sequences  $v := (v_i)$  we say that  $v \leq w$  if and only if  $v_i \leq w_i, \forall i$ .

Given  $k$  sequences  $v^{(j)}, j = 1, \dots, m$  we have by definition  $\sup_j v^{(j)} := (\sup_j v_i^{(j)})$ . Assume that the norm  $|\cdot|$  satisfies

$$0 \leq v \leq w \implies |v| \leq |w|.$$

**Lemma 12** *If  $v^{(j)} \geq 0, j = 1, \dots, m$  are  $m$  positive sequences we have then  $v^{(j)} \leq \sup_j v^{(j)} \leq \sum_j v^{(j)}$  hence*

$$\sup_j |v^{(j)}| \leq |\sup_j v^{(j)}| \leq \sum_j |v^{(j)}| \leq m \sup_j |v^{(j)}|. \quad (86)$$

*If  $a_i$  are positive numbers*

$$|\sum_j a_j v^{(j)}| \leq \sum_j a_j |v^{(j)}| \leq m |\sup_j a_j v^{(j)}| \leq m |\sum_j a_j v^{(j)}|. \quad (87)$$

**Lemma 13** *Consider a Hamiltonian*

$$Q(x, w) = \sum_{\substack{|\nu| \leq K, \alpha, \beta \\ |\alpha| + |\beta| \geq 1}} e^{\nu \cdot x} Q_{\nu, \alpha, \beta} z^\alpha \bar{z}^\beta,$$

*(i.e. independent of  $y$ , of degree at least one in  $w$  and in  $x$  a trigonometric polynomial of degree  $|\nu|_1 \leq K$ ). Denote its associated vector field by*

$$X_Q = X_Q^{(y)} + X_Q^{(w)}, \quad X_Q^{(y)} = \partial_x Q(x, w) \cdot \partial_y, \quad X_Q^{(w)} = -i \partial_z Q(x, w) \cdot \partial_{\bar{z}} + i \partial_{\bar{z}} Q(x, w) \cdot \partial_z,$$

*then one has that*

$$\|X_Q^{(y)}\|_{s,r} \leq K \|X_Q^{(w)}\|_{s,r} \quad (88)$$

*Proof* We can write (collecting in some arbitrary way the monomials)

$$Q = \sum_k A_k(x, w) z_k + B_k(x, w) \bar{z}_k.$$

Then, by definition

$$\|X_Q^{(y)}\|_{s,r} = r^{-2} \sup_{\|w\|_{a,p} \leq r} \sum_{j=1}^n |M \partial_{x_j} A_k(x, w) z_k + M \partial_{x_j} B_k(x, w) \bar{z}_k|$$

now we know that  $\sum_{j=1}^n |M \partial_{x_j} A_k(x, w)| \leq K |M A_k(x, w)|$  then by Cauchy-Schwartz we have

$$\|X_Q^{(y)}\|_{s,r} \leq r^{-2} K \sup_{\|w\|_{a,p} \leq r} \sqrt{\sum_k |z_k|^2 + |\bar{z}_k|^2} \sqrt{\sum_k |M A_k|^2 + |M B_k|^2}$$

and we have that  $|M A_k| \leq |M \partial_{z_k} Q|$  (same for  $B$ ) since  $\sum_k |z_k|^2 + |\bar{z}_k|^2 \leq r^2$  and the  $\ell^2$  norm is dominated by the  $\|\cdot\|_{a,p}$  norm the claim follows.  $\square$

Now in  $\ell_{a,p}$  we can define the norm

$$v = \{v_k\}_{k \in \mathbb{Z}^d} \quad |v|_{a,p}^2 = \sum_{\mathfrak{t}} e^{2a|\mathfrak{r}_{\mathfrak{t}}|} |\mathfrak{r}_{\mathfrak{t}}|^{2p} \sup_{k \in \mathcal{D}_{\mathfrak{t}}} |v_k|^2.$$

**Remark 16** This norm is built so that if for  $v, w \in \ell_{a,p}$  we have for all  $\mathfrak{t} \in \mathfrak{T}$  that  $\sup_{k \in \mathcal{D}_{\mathfrak{t}}} |v_k| \leq \sup_{k \in \mathcal{D}_{\mathfrak{t}}} |w_k|$  then  $|v|_{a,p} \leq |w|_{a,p}$ . One sees that the norms  $\|\cdot\|_{a,p}$  and  $|\cdot|_{a,p}$  are equivalent, namely  $c_{a,p} |v|_{a,p} \leq \|v\|_{a,p} \leq C_{a,p} |v|_{a,p}$ , with:

$$C_{a,p} := \sup_{\mathfrak{t}, k \in \mathcal{D}_{\mathfrak{t}}} \sqrt{d_{\mathfrak{t}}} e^{a(|k| - |\mathfrak{r}_{\mathfrak{t}}|)} \frac{|k|^p}{|\mathfrak{r}_{\mathfrak{t}}|^p}, \quad c_{a,p} = \inf_{\mathfrak{t}, k \in \mathcal{D}_{\mathfrak{t}}} e^{a(|k| - |\mathfrak{r}_{\mathfrak{t}}|)} \frac{|k|^p}{|\mathfrak{r}_{\mathfrak{t}}|^p}.$$

**Lemma 14** *Given a hamiltonian  $H = N + P$  with  $P$  a quasi-Töplitz function with parameters  $(K, \theta, \mu)$ , and  $N$  as in (30). Consider the linear operator*

$$L := \text{ad}(N) + \Pi_{\text{rg}, \leq K} \text{ad}(\Pi_{>2} P)$$

and the equation  $LF = \Pi_{\text{rg}, \leq K} P$ . In the set of  $\xi$  for which formulæ(40) hold, for all  $|\nu| < K$ ,  $\mathfrak{t}, \mathfrak{t}' \in \mathfrak{T}$ , such equation admits a unique solution  $F = \Pi_{\text{rg}} F$  which is a quasi-Töplitz function with parameters  $(K, \theta, \mu)$  and satisfies the bounds

$$\|X_F\|_{\vec{p}^+}^T \leq \delta^{-2} K^{3(\ell+4)\varrho} \|X_{F_{\text{rg}}}\|_{\vec{p}}^T,$$

where  $\vec{p}^+ = (s - 2\delta s, r - 2\delta r, K, \theta + 2\delta\theta, \mu - 2\delta\mu)$  and  $\vec{p} = (s, r, K, \theta, \mu)$ .

*Proof* We discuss in detail the estimates on the  $\lambda$  norm, the ones on the Töplitz norm follow verbatim from Proposition 10.16 of [14].

We recall that by remark 7, setting  $D = \text{ad}(\Pi_{\ker} P)$  and  $A = D^{-1} \Pi_{\text{rg}, \leq K} \text{ad}(\Pi_{>2} P)$  we have

$$L = D(1 + A) \quad \longrightarrow \quad L^{-1} = D^{-1}(1 - A + A^2).$$

Let us now work in the set of  $\xi$  where (40) hold. We first prove that for any quadratic Hamiltonian  $b \in \mathcal{F}_{\text{rg}, \leq K}$ , setting  $Y = D^{-1}b$ , we have

$$\|X_Y\|_{s,r}^\lambda \leq \varepsilon^{-2q} K^{(\ell+3)\varrho} \|X_b\|_{s,r}^\lambda. \quad (89)$$

As a first step we notice that if we divide  $Y = Y^{(0)} + Y^{(1)} + Y^{(2)}$  w.r.t the degree in  $w = z, \bar{z}$  then it is sufficient to bound each  $\|X_{Y^{(i)}}\|_{s,r}^\lambda$  in terms of the corresponding  $\|X_{b^{(i)}}\|_{s,r}^\lambda$ .

The term  $Y^{(0)}$  is trivial. We now study  $Y, b$  of degree one.

In studying linear real Hamiltonians of the form  $h = \sum_{\sigma, k, \nu} h_{\sigma, k}(\nu) e^{i\nu \cdot x} z_k^\sigma$ , by Lemma 13, we only need to bound the  $w$  component of the vector field, namely

$$\|X_h\|_{s,r} \leq 2(K+1)r^{-1} \left( \sum_k \left( \sum_\nu |h_{+,k}(\nu)| e^{s|\nu|} \right)^2 e^{2a|k|} |k|^{2p} \right)^{\frac{1}{2}}$$

Now we use this formulas with  $h = \partial_\xi^\alpha Y^{(1)}$ ,  $|\alpha| \leq \ell$ . Then, by using (87) with  $a_j = \lambda^{|\alpha|}$ ,  $m \leq (\ell+1)^n$ ,  $v_j = \{\sum_\nu e^{s|\nu|} |\partial_\xi^\alpha h_{s(k),k}(\nu)|\}_{k \in \mathbb{Z}^d}$  and  $|\cdot| = \|\cdot\|_{a,p}$ , we have that

$$\begin{aligned} \|X_Y\|_{s,r}^\lambda &\leq (2K+2)(\ell+1)^n r^{-1} \|M^\lambda Y\|_{a,p}, \\ M^\lambda h &:= \left\{ \sum_\nu e^{s|\nu|} \sum_{|\alpha| \leq \ell} \lambda^{|\alpha|} |\partial_\xi^\alpha h_{s(k),k}(\nu)| \right\}_{k \in \mathbb{Z}^d}. \end{aligned} \quad (90)$$

Now we have by Lemma 10 that for each  $\mathfrak{t}$  and for each  $k \in \mathbb{D}_\mathfrak{t}$

$$\sum_{|\alpha| \leq \ell} \lambda^{|\alpha|} \sup_{k \in \mathbb{D}_\mathfrak{t}} |\partial_\xi^\alpha Y_{s(k),k}(\nu)| \leq \varepsilon^{-2q} K^{(\ell+2)\varrho} \sum_{|\alpha| \leq \ell} \lambda^{|\alpha|} \sup_{k \in \mathbb{D}_\mathfrak{t}} |\partial_\xi^\alpha b_{s(k),k}(\nu)|$$

We exchange the sup with the sum over  $\nu$  using (86) with  $|\{v_\nu\}| = \sum_\nu e^{s|\nu|} |v_\nu|$ , then we have that

$$\sup_{k \in \mathbb{D}_\mathfrak{t}} |(M^\lambda Y)_k| \leq d_\mathfrak{t} \varepsilon^{-2q} K^{(\ell+2)\varrho} \sup_{k \in \mathbb{D}_\mathfrak{t}} |(M^\lambda b)_k|$$

By Remark 16 we conclude that  $|M^\lambda Y|_{a,p} \leq \ell \varepsilon^{-2q} K^{(\ell+2)\varrho} |M^\lambda b|_{a,p}$  and

$$\|M^\lambda Y\|_{a,p} \leq \ell C_{a,p} c_{a,p}^{-1} \varepsilon^{-2q} K^{(\ell+2)\varrho} \|M^\lambda b\|_{a,p}. \quad (91)$$

Then by (87)

$$\|M^\lambda b\|_{a,p} \leq \sum_{|\alpha| \leq \ell} \lambda^{|\alpha|} \left\| \left\{ \sum_\nu e^{s|\nu|} |\partial_\xi^\alpha b_{s(k),k}(\nu)| \right\}_{k \in \mathbb{Z}^d} \right\|_{a,p} \leq r \|X_b\|_{s,r}^\lambda.$$

Substituting (91) in (90) and then using this last bound we obtain

$$\|X_Y\|_{s,r}^\lambda \leq (2K+2)(\ell+1)^n \ell C_{a,p} c_{a,p}^{-1} \varepsilon^{-2q} K^{(\ell+2)\varrho} \|X_b\|_{s,r}^\lambda \leq \varepsilon^{-2q} K^{(\ell+3)\varrho} \|X_b\|_{s,r}^\lambda,$$

provided  $K$  is large (note that  $\varrho > 1$ ).

We need to treat now the case of (regular) quadratic Hamiltonians (with conservation of momentum) which belong to the range. To any quadratic hamiltonian  $Q(x, w)$  we associate an  $x$  dependent linear operator  $\tilde{Q}(x)$  on  $\ell_{a,p}$  by Poisson bracket. By Formula (50) this amounts to writing  $Q(x, w) = -\frac{1}{2} w \tilde{Q}(x) J w^t$ . Then in the basis  $z_k, \bar{z}_k$  we can define the majorant (of the operator  $\text{ad}(Q)$ ) in matrix terms as

$$\widehat{Q}_{k,h}^{\sigma,\sigma'} = \sum_{\nu} e^{s|\nu|} |\tilde{Q}_{k,h}^{\sigma,\sigma'}(\nu)|$$

Similarly to the linear case we only need to control

$$\|X_Q^{(w)}\|_{s,r} = \sup_{\|w\|_{a,p} \leq 1} \|\widehat{Q}w\|_{a,p}, \quad w = \{z_k, \bar{z}_k\}_{k \in \mathbb{Z}^d}.$$

Then passing to the equivalent norm  $|\cdot|_{a,p}$  of Remark 16, we have

$$\|X_Q\|_{s,r} \leq (K+1) c_{a,p}^{-1} C_{a,p} \sup_{|w|_{a,p} \leq 1} |Qw|_{a,p}.$$

Now as in the linear case we pass to the  $\lambda$  norm and get

$$\begin{aligned} \|X_Q\|_{s,r}^\lambda &\leq (\ell+1)^n (K+1) c_{a,p}^{-1} C_{a,p} \sup_{|w|_{a,p} \leq 1} |\widehat{Q}^\lambda w|_{a,p}, \\ (\widehat{Q}^\lambda)_{k,h}^{\sigma,\sigma'} &:= \sum_{\alpha} \lambda^{|\alpha|} \sum_{\nu} e^{s|\nu|} |\partial_{\xi}^{\alpha} \tilde{Q}_{k,h}^{\sigma,\sigma'}(\nu)| \end{aligned} \quad (92)$$

We consider  $Y = D^{-1}b$  where  $b$  is a quadratic Hamiltonian in the range. Since  $D$  preserves the range we have that also  $Y$  is in the range. Now we claim that for all  $w$  and for all  $t$  we have

$$\sup_{k \in \mathbb{D}_t} |(\widehat{Y}^\lambda w)_k| \leq \ell \varepsilon^{-2q} K^{(\ell+2)\varrho} \sup_{k \in \mathbb{D}_t} |(\widehat{b}^\lambda \tilde{w})_k|, \quad \tilde{w}_h = \sup_{k \in \mathbb{D}_t} |w_k| \quad \forall h \in \mathbb{D}_t. \quad (93)$$

If this holds true by remark 16 we have

$$|\widehat{Y}^\lambda w|_{a,p} \leq \ell \varepsilon^{-2q} K^{(\ell+2)\varrho} |\widehat{b}^\lambda \tilde{w}|_{a,p}$$

then since  $|\tilde{w}|_{a,p} = |w|_{a,p}$  (by definition) we have

$$\sup_{|w|_{a,p} \leq 1} |\widehat{Y}^\lambda w|_{a,p} \leq \ell \varepsilon^{-2q} K^{(\ell+2)\varrho} \sup_{|u|_{a,p} \leq 1} |\widehat{b}^\lambda u|_{a,p}.$$

Now we can apply (87) to take out the sum over  $\alpha$  from the norm  $|\cdot|_{a,p}$  and we get

$$|\widehat{b}^\lambda u|_{a,p} \leq (\ell+1)^n \sum_{\alpha} \lambda^{|\alpha|} |\widehat{\partial_{\xi}^{\alpha} b}|_{a,p} \leq (\ell+1)^n \|X_b\|_{s,r}^\lambda$$

$$\|X_Y\|_{s,r}^\lambda \leq (\ell+1)^n (K+1) c_{a,p}^{-1} C_{a,p} \ell \varepsilon^{-2q} K^{(\ell+2)\varrho} (\ell+1)^n \|X_b\|_{s,r}^\lambda \leq K^{(\ell+3)\varrho} \varepsilon^{-2q} \|X_b\|_{s,r}^\lambda.$$

For the proof of (93) we first use Lemma 10 to deduce

$$\sum_{\alpha} \lambda^{|\alpha|} \sup_{\substack{k \in \mathbb{D}_t \\ h \in \mathbb{D}_{t'}}} |\partial_{\xi}^{\alpha} \tilde{Y}_{k,h}^{s(k),s(h)}(\nu)| \leq \varepsilon^{-2q} K^{(\ell+2)\varrho} \sum_{\alpha} \lambda^{|\alpha|} \sup_{\substack{k \in \mathbb{D}_t \\ h \in \mathbb{D}_{t'}}} |\partial_{\xi}^{\alpha} \tilde{b}_{k,h}^{s(k),s(h)}(\nu)|.$$

By repeated use of (86) and (87) and using the fact that  $d_t d_{t'} \leq \ell$  the formula follows. This completes the proof of (89). Now we turn to  $F = L^{-1}P_{\text{rg}}$ . For any regular hamiltonian  $h$  and for all  $s' < s, r' < r$ , we can bound  $\|X_{\{P,h\}}\|_{s',r'}^\lambda$  by Cauchy estimates recalling that

$\|X_F\|_{s,r} \leq \Theta \leq \frac{3}{2}\Theta_0$  (here  $\Theta$  is one of the telescopic parameters defined in section 4.2). Then by using (89) at most three times we get

$$\|X_F\|_{s-2\delta s, r-2\delta r}^\lambda < \delta^{-2} K^{3(\ell+3)e} \|X_{F_{\text{reg}}}\|_{s,r}^\lambda. \quad (94)$$

Then passing to the Töplitz norm we follow word by word [14], Proposition 10.16.

By definition the quasi-Töplitz property for  $F$  depends only on the quadratic part  $F^{(2)}$ . Then proving that  $F$  is quasi-Töplitz requires that we produce a *piece-wise Töplitz* approximation and an error. As in Proposition 10.16 for the part of  $F$  with  $\sigma\sigma' = 1$  we take zero as piece-wise Töplitz approximation.

For given  $\nu$  we treat the quadratic part as sum of blocks  $F_{\nu,t,t'}$ . When  $\sigma\sigma' = -1$  by conservation of momentum  $\pi(\nu) + \mathbf{r}_t - \mathbf{r}_{t'} = 0$  so only  $\theta_{t'} = \theta_t$  must be fixed (and must belong to a finite list) then the couple  $(\pi(\nu) + \mathbf{r}_t, \theta')$  identifies (at most) one  $t'$  and hence the block  $D_{t'}$ . Then we identify the parallel strata  $Z$  to which  $D_t$  belong and the same for  $D_{t'}$ . We define the quasi-Töplitz approximation as follows. Recall that  $F_{\nu,t,t'}$  solves the equation discussed in remark (7) inverting the operator  $L_m$ . This solution is a 3 step algorithm in each step we either perform a Poisson bracket, which preserves the Quasi-Töplitz property or we solve an equation of type

$$\left( (\omega \cdot \nu) - 2\pi(\nu) \cdot \mathbf{r}_t - |\pi(\nu)|^2 + \theta_t + \tilde{\Omega}_t - \theta_{t'} - \tilde{\Omega}_{t'} \right) F_{\nu,t,t'} = P_{\nu,t,t'}, \quad (95)$$

by induction  $P_{\nu,t,t'}$  has a quasi-Töplitz approximation  $P_{\nu,t,t'}^{qt}$  same for  $\tilde{\Omega}_t$  and  $\tilde{\Omega}_{t'}$ . We distinguish two cases, if  $\pi(\nu) \cdot x$  is not constant on  $\Sigma_t$  we have seen that  $\pi(\nu) \cdot \mathbf{r}_t$  is large hence  $F_{\nu,t,t'}$  is quite small and we can again take zero as its piece-wise Töplitz approximation.

Otherwise in Formula (95), the only terms which are not constant on the stratum are  $\tilde{\Omega}_t - \tilde{\Omega}_{t'}$ . By induction they have a quasi-Töplitz approximation  $\tilde{\Omega}_t^{qt} - \tilde{\Omega}_{t'}^{qt}$  and the final error will produce the error of the quasi-Töplitz approximation.

That is we take the solution  $\bar{F}_{\nu,t,t'}^{qt}$  of

$$\left( (\omega \cdot \nu) - 2\pi(\nu) \cdot \mathbf{r}_t - |\pi(\nu)|^2 + \theta_t + \tilde{\Omega}_t^{qt} - \theta_{t'} - \tilde{\Omega}_{t'}^{qt} \right) \bar{F}_{\nu,t,t'}^{qt} = P_{\nu,t,t'}^{qt}. \quad (96)$$

as quasi-Töplitz approximation of  $F_{\nu,t,t'}$  on this stratum.

The final point is now to show that we can bound the inverse of the matrix on the left-hand side of (96) by  $K^\rho$ . on this stratum the constraint (41) on a single point is sufficient to constraint on all points of the stratum. For this one has to compute the value of  $\varrho$  to be used on the stratum and compare it with the error term in the quasi-Töplitz approximation.

$$\|X_F\|_{\vec{p}^+}^T < \delta^{-2} K^{3(\ell+4)e} \|X_{F_{\text{reg}}}\|_{\vec{p}}^T, \quad \vec{p}^+ = (s - 2\delta s, r - 2\delta r, K, \theta + 2\delta\theta, \mu - 2\delta\mu), \quad \vec{p} = (s, r, K, \theta, \mu) \quad (97)$$

□

## B Quasi Töplitz structure

*We group here an informal description of the main properties of quasi-Töplitz functions needed through the KAM algorithm. Recall that the notion of quasi-Töplitz is given through some parameters  $K, \vartheta, \mu$ .*

*First one has to identify, for each  $N \geq K$  a natural affine structure. This is done by introducing the notion of optimal presentation for a point  $m$  and the notion of cut.*

*The stratification by cuts.* The optimal presentation of a point  $m \in \mathbb{Z}^d$  is a combinatorial notion dependent on  $N$ . One presents the point  $m$  as the intersection of  $d$  hyperplanes  $(v_i, x) = p_i$  with  $v_i$  of norm  $\leq N$  and such that the list  $p_i, v_i$  is *minimal* in an explicit lexicographic order. This presentation, called *optimal* is unique and denoted  $m \xrightarrow{N} [v_i; p_i]$ .

Then, given an open interval  $I = (a, b)$ , one says that a point  $m \xrightarrow{N} [v_i; p_i]$  has a cut at  $I$  if no coordinate  $p_i$  lies in  $I$ . If it has such a cut then we have some  $p_\ell < a$ ,  $p_{\ell+1} > b$  and the first

$\ell$  equations of the optimal presentation of  $p$  define some affine space of codimension  $\ell$ , denoted  $[v_i; p_i]_\ell$ . This defines a linear stratification at least on the set of points which have such a cut.

If one gives  $d + 1$  disjoint increasing intervals  $I_j$ ,  $j = 1, \dots, d + 1$  clearly for any point  $m$  there is a minimum  $j$  such that  $m$  has a cut for  $I_j$ . To this cut is then associated an affine space and in this way one constructs the desired stratification as explained in remark 2. We introduce a parameter  $\varrho_0$  which will be fixed at the end. We start from intervals  $(\varrho_i, \varrho_{i+1})$  in the scale  $\log_N$  with

$$\varrho_1 = 4d\varrho_0, \quad \varrho_{i+1} = 4d\varrho_i$$

and impose for the condition that  $m \xrightarrow{N} [v_i; p_i]$  has a cut at  $\ell$  if there exists  $j$  such that,  $\log_N 2p_\ell < \varrho_j$ ,  $\log_N \frac{p_{\ell+1}}{4} \geq \varrho_{j+1}$ . This indeed is a *cut* associated to  $d + 1$  disjoint intervals and hence provides a stratification  $\Sigma^{N, \ell}$ , see section 3.1 of [17]. We shall usually drop the symbol  $\varrho$  and write just  $\Sigma^N$ .

Thus given any point  $m$  this construction determines the stratum  $\Sigma_m^N$  to which it belongs to and the linear space, of some codimension  $\ell$ , spanned by the stratum. If  $v$  is a vector with  $|v|_1 \leq N$  consider the scalar product  $(v, x)$ ,  $x \in \Sigma_m^N$ . If this is not constant on the stratum by the definitions it is *large with*  $4N^{\varrho_j+1}$ .

**Lemma 15** *The number of strata with a cut at  $\varrho_j$  is less than  $N^{(2d-1)\varrho_j}$ .*

*Proof* This is Remark 7.7 of [14] with  $p = 4N^{\varrho_j}$ , indeed summing over the codimension

$$\sum_{\ell=0}^d (2\kappa N)^{\ell d} (4N^{\varrho_j})^\ell \leq N^{d(\varrho_j+d+1)} \leq N^{(2d-2)\varrho_j}$$

provided that  $\varrho_0 > d(d+1)/(d-2)$ .  $\square$

The stratification  $\Sigma^N$ , provided that  $N$  is large enough, refines the stratification given by Definition 20 (cf. Theorem 5 of [14]). In fact given any element  $\mathfrak{t} \in \mathfrak{T}_s$ , this determines the set  $\mathfrak{D}_\mathfrak{t}$  and the root  $\mathfrak{r}_\mathfrak{t}$  (which in general is NOT in the set  $\mathfrak{D}_\mathfrak{t}$ ).

We then have that, the stratum through  $\mathfrak{r}_\mathfrak{t}$  and the ones passing through each individual point of  $\mathfrak{D}_\mathfrak{t}$  are all parallel to each other and the union of these strata is also a union of some family of sets  $\{\mathfrak{D}_\mathfrak{t}, \mathfrak{r}_\mathfrak{t}\}$ .<sup>8</sup>

*Remark 17* In other words we have a decomposition of the set  $\mathfrak{T}_s$  into sets  $\mathfrak{T}_s(i)$  (which are the ones introduced in Proposition 19), so that each stratum  $\mathfrak{T}_s(i)$  is the indexing set of the strata through the points in  $\mathfrak{D}_\mathfrak{t}$  and  $\mathfrak{r}_\mathfrak{t}$  for all  $\mathfrak{t} \in \mathfrak{T}_s(i)$ .

*$\tau$ -bilinear and quasi-Töplitz functions* The quasi-Töplitz property can not be given only on the quadratic terms of a Hamiltonian since we need a class of functions closed w.r.t. Poisson brackets, so we first relax the notion of quadratic Hamiltonian.

Given  $\frac{1}{2} \leq \mu, \vartheta \leq 4$  and  $\varrho_1 \leq 4d\tau \leq \varrho_{d+1}$  we define (cf. 8.6 of [14]) the  $(N, \theta, \mu, \tau)$ -bilinear functions to be the functions which are bilinear in the *high variables*  $z_m^\sigma, z_n^{\sigma'}$ . By this we mean that  $|m|, |n| > \vartheta N^{\varrho_{d+1}}$  and both  $m$  and  $n$  have a cut at  $(\mu N^\tau, \theta N^{4d\tau})$ . These functions may depend on  $x, y$  and on the *small variables*  $z_j^\sigma$  with  $|j| < \mu N^3$  in a possibly complicated way but with the constraint that the coefficients have *low momentum* (cf. 8.2 of [14]).

Finally we define the *piecewise Töplitz* functions as those  $(N, \theta, \mu, \tau)$ -bilinear functions which are constant when restricted to each stratum (cf. 8.10 of [14]).

We can now define the  $(K, \theta, \mu)$ -quasi-Töplitz functions. Informally speaking given a function  $f$ , for all  $N > K$ ,  $\varrho_1 \leq 4d\tau \leq \varrho_{d+1}$ , we project it on the  $(N, \theta, \mu, \tau)$ -bilinear functions and we say that  $f$  is quasi-Töplitz if all these projections are *well approximated* by a piecewise Töplitz function. To be more precise,  $\tau$  controls the size of the error function, namely the  $(N, \theta, \mu, \tau)$ -bilinear part of  $f$  is approximated by a piecewise Töplitz function with an error of the order  $N^{-4d\tau}$ , for all  $N \geq K$ .

The role of the parameters  $K, \theta, \mu$  is to ensure that if  $f, g$  are quasi-Töplitz with parameters  $K, \theta, \mu$  then  $\{f, g\}$  is quasi-Töplitz for all  $\vartheta' > \vartheta$  and  $\mu' < \mu$  provided  $K' > K$  is large enough.

<sup>8</sup> strictly speaking one needs a slight refinement of the definitions adding some more strata but with the same estimate on the number of strata.

The stratification  $\Sigma^N$  described above is connected to the quasi-Töplitz structure, indeed if  $m$  belongs to a stratum  $\Sigma_m^N$  of codimension  $\ell$  then, by construction,  $m$  has a cut at  $(\mu N^{\varrho_j}, \theta N^{4d\varrho_j})$  for all  $\frac{1}{2} \leq \mu, \vartheta \leq 4$  where  $\varrho_j$  is fixed by  $\Sigma^N$  and we shall denote it by  $\varrho_\Sigma$ . Then one can show that all points  $n$  close to  $m$  have a similar cut, see Lemmata 7.21 to 7.24 of [14]. In §8.12 of [14] there are several properties of diagonal quasi-Töplitz functions which are needed in the KAM algorithm. These properties are in fact also valid with the obvious modifications for the block diagonal quasi-Töplitz functions in the subalgebra  $\mathcal{F}_{\ker}$  of Definition 7.

This is due to the fact that, provided we start from  $N \geq K_0$  and  $K_0$  is large enough the corresponding stratifications refine the affine stratification of Proposition 1, part iii).

When we apply it to the hamiltonians in the KAM algorithm we exploit the fact that, by part iv) of the same proposition the corresponding function  $\theta_t$  is the same on all points of the stratum. Moreover we need the following

**Lemma 16** *Take a quasi-Töplitz function  $F$  quadratic in  $w$  and its projection  $\Pi_{\ker} F = \sum_t F_t$ . Then for each  $N \geq K$  consider the stratification  $\Sigma(N)$ . If  $\mathbf{r}_t, \mathbf{r}_{t'}$  belong to the same stratum  $Y$  (of codimension  $\ell$ ) and  $\mathbf{D}_t = \mathbf{D}_{t'} + T_Y$  (as in Proposition 1 iii.) then  $\theta_t = \theta_{t'}$  and*

$$|F_t - F_{t'}|_\infty^\lambda \leq N^{-4d\varrho_\Sigma} \|X_F\|_\frac{T}{p}$$

*Proof* This follows word by word as the proof of Lemma 8.20–Formula (92) of [14].  $\square$

## References

1. M. Berti and L. Biasco. Branching of Cantor manifolds of elliptic tori and applications to PDEs. *Comm. Math. Phys.*, 305(3):741–796, 2011.
2. M. Berti, L. Biasco, and M. Procesi. KAM theory for the Hamiltonian derivative wave equation. *Annales Scientifiques de l'ENS*, 46(2), 2013.
3. M. Berti, and P. Bolle. Quasiperiodic solutions with Sobolev regularity for NLS on  $\mathbb{T}^d$  and a multiplicative potential. *J. European Math. Society*, 15 (1): 229-286, 2013
4. M. Berti, P. Bolle, and M. Procesi. An abstract Nash-Moser theorem with parameters and applications to PDEs. *Annales de l'Institut Henri Poincaré (C) Non Linear Analysis*, 27(1):377 – 399, 2010.
5. J. Bochnak, M. Coste, and Marie-Francoise Roy. *Real algebraic geometry*, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1998. Translated from the 1987 French original, Revised by the authors.
6. J. Bourgain. *Green's function estimates for lattice Schrödinger operators and applications*, volume 158 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2005.
7. J. Geng and J. You. A KAM theorem for Hamiltonian partial differential equations in higher dimensional spaces. *Comm. Math. Phys.*, 262(2):343–372, 2006.
8. J. Colliander, M. Keel, G. Staffilani, H. Takaoka, and T. Tao. Transfer of energy to high frequencies in the cubic defocusing nonlinear Schrödinger equation. *Invent. Math.*, 181(1):39–113, 2010.
9. L.H. Eliasson and S.B. Kuksin. KAM for the nonlinear Schrödinger equation. *Ann. of Math. (2)*, 172(1):371–435, 2010.
10. L.H. Eliasson, and S.B. Kuksin KAM for the non-linear Beam equation 2: A normal form theorem arxiv:1502.02262
11. Van Nguyen, Bich. Characteristic polynomials, associated to the NLS. Ph.D. Thesis, La Sapienza 2013.
12. J. Pöschel. A KAM-theorem for some nonlinear partial differential equations. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 23(1):119–148, 1996.
13. C. Procesi and M. Procesi. A normal form for the Schrödinger equation with analytic non-linearities. *Communications in Mathematical Physics*, 312(2):501–557, 2012.
14. M. Procesi and C. Procesi. A KAM algorithm for the non-linear Schrödinger equation. *Advances in Math.*, 272:399–470, 2015.
15. M. Procesi, C. Procesi, and Nguyen Bich Van. The energy graph of the non-linear Schrödinger equation. *Rend. Lincei Mat. Appl.*, 24:1–73, 2013.

- 
16. W.M. Wang. Supercritical nonlinear Schrödinger equations i: Quasi-periodic solutions. Preprint, arXiv:1007.0156.
  17. X. Xu and M. Procesi. Quasi-Töplitz Functions in KAM Theorem. *SIAM J. of Math. Anal.*, 45(4):2148 – 2181, 2013.