A Nash-Moser approach to KAM theory

Massimiliano Berti, Philippe Bolle

Abstract Any finite dimensional embedded invariant torus of an Hamiltonian system, densely filled by quasi-periodic solutions, is isotropic. This property allows us to construct a set of symplectic coordinates in a neighborhood of the torus in which the Hamiltonian is in a generalized KAM normal form with angle-dependent coefficients. Based on this observation we develop an approach to KAM theory via a Nash-Moser implicit function iterative theorem. The key point is to construct an approximate right inverse of the differential operator associated to the linearized Hamiltonian system at each approximate quasi-periodic solution. In the above symplectic coordinates the linearized dynamics on the tangential and normal directions to the approximate torus are approximately decoupled. The construction of an approximate inverse is thus reduced to solving a quasi-periodically forced linear differential equation in the normal variables. Applications of this procedure allow to prove the existence of finite dimensional Diophantine invariant tori of autonomous PDEs.

1 Introduction

In the last years much work has been devoted to the existence theory of quasi-periodic solutions of PDEs. The main strategies developed to overcome the well known "*small divisors*" problem are:

1. KAM (Kolmogorov-Arnold-Moser) theory,
2. Newton-Nash-Moser implicit function theorems.

Massimiliano Berti
SISSA, Via Bonomea 265, Trieste, 34136, Italy e-mail: `berti@sissa.it`

Philippe Bolle
Avignon Université, Laboratoire de Mathématiques d'Avignon (EA2151), F-84018 Avignon, France, e-mail: `philippe.bolle@univ-avignon.fr`

The KAM approach consists in generating iteratively a sequence of canonical transformations of the phase space which bring, up to smaller and smaller remainders, the Hamiltonian system into a normal form with an invariant torus "at the origin": in the new coordinates, the invariant torus is the zero section of the trivial linear bundle $\mathbb{T}^\nu \times \mathbb{R}^\nu \times E \xrightarrow{\pi} \mathbb{T}^\nu$, see section 2. This iterative procedure requires, at each step, to invert the so called linear "homological equations". In the usual KAM scheme the normal form has constant coefficients (see (30)), the homological equations have constant coefficients and are solved imposing the "second order Melnikov" non-resonance conditions. The final KAM torus is linearly stable.

This scheme has been effectively implemented by Kuksin [20] and Wayne [28] for 1-d nonlinear wave (NLW) and Schrödinger (NLS) equations with Dirichlet boundary conditions. The required second order Melnikov non resonance conditions are violated in presence of multiple normal frequencies, for example, already for periodic boundary conditions.

Thus a more direct bifurcation approach was proposed by Craig and Wayne [13], see also [12], for 1-d NLW and NLS equations with periodic boundary conditions. After a Lyapunov-Schmidt decomposition, the search of the invariant torus is reduced to solve a functional equation by some Newton-Nash-Moser implicit function theorem in Banach scales of analytic functions of time and space. The main advantage of this approach is to require only the "first order Melnikov" non-resonance conditions for solving the linearized equations at each step of the iteration. These conditions are essentially the minimal assumptions. On the other hand, the main difficulty is that the linearized equations are PDEs with non-constant coefficients, represented by differential operators that are small perturbations of a diagonal operator having arbitrarily small eigenvalues. Hence it is hard to estimate their inverses in high norms. Craig-Wayne [13] solved this problem for periodic solutions and Bourgain [9] also for quasi-periodic solutions. This approach is particularly useful for PDEs in higher dimension due to the large (possibly unbounded) multiplicity of the normal frequencies. It has been effectively implemented by Bourgain [10], [11], for analytic NLS and NLW with Fourier multipliers on $\mathbb{T}^d$, $d \geq 2$, and by Berti-Bolle [4]-[5] for forced NLS and NLW equations with a multiplicative potential and finite differentiable nonlinearities. In Berti-Corsi-Procesi [8] this scheme has been then generalized into an abstract Nash-Moser implicit function theorem with applications to NLW and NLS on general compact Lie groups and homogeneous spaces.

We remark that in the above papers the transformations used to prove estimates in high norms for the inverse linearized operators at the approximate solutions are *not* maps of the phase space, as in the usual KAM approach, and therefore the dynamical system structure of the transformed system is lost.

The aim of this Note is to present an approach to normal form KAM theory based on a Nash-Moser implicit function theorem. Instead of performing directly a sequence of canonical maps of the phase space which (at the limit) conjugate the Hamiltonian to another one which possesses an invariant torus "at the origin", we construct an embedded invariant torus, with *equivalently* the normal form

(28) around it (Theorem 1), by a Nash-Moser iterative scheme in scales of Banach spaces. Note that the quadratic terms of the KAM normal form (28) are, in general, angle dependent.

The core of the present approach is to find a set of symplectic coordinates in which the tangential and the normal linearized equations at an (approximate) torus are (approximately) decoupled. This reduces the problem to the study of the quasi-periodically forced linearized equation in the normal directions. This symplectic construction preserves the Hamiltonian dynamical structure of the equations. Thus it is a decomposition of tangential/normal dynamics sharper than the usual Lyapunov-Schmidt reduction based on the splitting into bifurcation/range equations. The present KAM approach applies well also to PDEs whose flow is ill-posed.

As already mentioned, the main difficulty for implementing a Newton-Nash-Moser iterative scheme is to solve the (non homogeneous) linearized equations at each approximate quasi-periodic solution. This is a difficult matter for the presence of small divisors and because the tangential and normal components to the torus of the linearized equations are coupled by the nonlinearity.

It was noted by Zehnder [29] that, in order to get a "quasi Newton-Nash-Moser" scheme with still quadratic speed of convergence, it is sufficient to invert the linearized equation only approximately. In [29] Zehnder introduced the precise notion of approximate right inverse linear operator, see (43). Its main feature is to be an *exact* inverse of the linearized equation at an exact solution.

In this approach we construct an approximate right inverse for the functional equation satisfied by the embedding of an invariant torus of an Hamiltonian system, see (16), or (34)-(35). Let us explain more in detail the main ideas. The first observation is that an embedded invariant torus supporting a non-resonant rotation is isotropic. This is classical for finite dimensional Hamiltonian systems, see [19] or [17]-Lemma 33. Actually this property is also true for *infinite dimensional* Hamiltonian systems (Lemma 1) since it requires only that the Hamiltonian flow is well defined on the invariant torus and preserves the symplectic 2-form on it. Near an isotropic torus it is then possible to introduce the symplectic set of coordinates (23) in which the torus is "at the origin". It follows that the existence of an invariant torus and a nearby normal form like (28) are equivalent statements, see Theorem 1. Clearly, with second order Melnikov non-resonance conditions, it is also possible to obtain a constant coefficients normal form as (30), i.e. to prove the reducibility of the torus.

This observation is the bridge with the usual KAM proof based on normal forms transformations. The point is that the normal form (28) means more that the existence of the invariant torus, since it also provides a control of the linearized equations in the normal bundle of the torus. Actually, in the normal form coordinates, the linearized equations at the torus simplify, see (31). In particular the second component in (31) is decoupled from the others and the system (31) can be solved in a "triangular" way.

Of course, there is little interest in inverting the linearized equation at a torus that is already a solution. The point is that, at an approximate invariant torus, it is still possible to construct an approximate right inverse of the linearized equation, which is enough to get a "quasi-Nash-Moser" scheme à la Zehnder.

With this aim, in section 4 we extend the symplectic construction developed in section 2 for an invariant torus, to an approximate solution. Needless to say, an approximate invariant torus is only approximately isotropic (Lemma 5). Thus the first step is to deform it into a nearby exactly isotropic torus (Lemma 6). This enables to define the set of symplectic coordinates (66) in which the isotropic torus is "at the origin". In these new coordinates also the linearized equations (73) simplify and we may invert them approximately, namely solve only (74)-(75). Such system is obtained by neglecting in (73) the terms which are zero at an exact solution, see Lemmata 8 and 3. The linear system (74)-(75) may be solved in a triangular way, first inverting the action-component equation (76) (see (77)-(78)), which is decoupled from the other equations.

In the case of a Lagrangian (finite dimensional) torus, there is not the last normal component in the system (75), and one may immediately solve the equation (82) for the angle component, see (83)-(84). This completes the construction of an approximate inverse. This is another way to recover the classical results of Zehnder [29] and Salomon-Zehnder [27] for maximal dimensional tori, and it is closely related to the method in [14] by De la Llave, Gonzalez, Jorba, Villanueva.

On the contrary, in the general case of an isotropic torus, the present strategy has reduced the search of an approximate right inverse for the embedded torus equation, to the problem of solving the linear equation (79). This is a quasi-periodically forced linear PDE which is a small perturbation of the original linearized PDE, restricted to the normal directions. The existence of a solution for such an equation is very simple for partially hyperbolic (whiskered) tori, because there is no resonance in the normal directions. In the more difficult case of elliptic tori, where small divisors appear, this equation has the same feature as the quasi-periodically forced linear PDE restricted to the normal directions. It makes possible to exploit KAM results that have already been proved for the corresponding forced PDEs, as, for example, [4], [5], [2], [8].

For finite dimensional systems, this construction is deeply related to the Herman-Fejoz KAM normal form theorem used in [17] to prove the existence of elliptic invariant tori in the planetary solar system. Actually le "Théorème de conjugaison tordue" of Herman (Theorem 38 in [17]) is deduced by a Nash-Moser implicit functions theorem in Fréchet spaces.

This scheme may be effectively implemented for autonomous Hamiltonian PDEs, like, for example,

1. **(NLW)** Nonlinear wave equation

$$y_{tt} - \Delta y + V(x) y = f(x,y) \,, \quad x \in \mathbb{T}^d \,, \quad y \in \mathbb{R} \,, \tag{1}$$

   with a real valued multiplicative potential (we may clearly consider also a convolution potential $V * y$ as in [11]) and a real valued nonlinearity f.
2. **(NLS)** Hamiltonian nonlinear Schrödinger equation

$$\mathrm{i}u_t - \Delta u + V(x)u = f(x,u)\,, \quad x \in \mathbb{T}^d\,,\ u \in \mathbb{C}\,, \tag{2}$$

where $f(x,u) = \partial_{\bar u} F(x,u)$ and the potential $F(x,u) \in \mathbb{R}$, $\forall u \in \mathbb{C}$, is real valued. For $u = a + \mathrm{i}b$, $a,b \in \mathbb{R}$, we define the operator $\partial_{\bar u} := \frac{1}{2}(\partial_a + \mathrm{i}\partial_b)$.

3. **(KdV)** Quasi-linear Hamiltonian perturbed KdV equations

$$u_t + u_{xxx} + \partial_x u^2 + \mathcal{N}(x,u,u_x,u_{xx},u_{xxx}) = 0\,, \quad x \in \mathbb{T}\,, \tag{3}$$

where $\mathcal{N}(x,u,u_x,u_{xx},u_{xxx}) := -\partial_x\big[(\partial_u f)(x,u,u_x) - \partial_x((\partial_{u_x} f)(x,u,u_x))\big]$ is the most general Hamiltonian (local) nonlinearity, see (4).

The NLW and NLS equations are studied in [6] and the quasi-linear KdV in [3].

All the above PDEs are infinite dimensional Hamiltonian systems. Also in view of the abstract setting of section 2, we present their Hamiltonian formulation:

1. The NLW equation (1) can be written as the Hamiltonian system

$$\frac{d}{dt}\begin{pmatrix} y \\ p \end{pmatrix} = \begin{pmatrix} p \\ \Delta y - V(x)y + f(x,y) \end{pmatrix} = \begin{pmatrix} 0 & Id \\ -Id & 0 \end{pmatrix}\begin{pmatrix} \delta_y H(y,p) \\ \delta_p H(y,p) \end{pmatrix}$$

where $\delta_y H$, $\delta_p H$ denote the $L^2(\mathbb{T}^d_x)$-gradient of the Hamiltonian

$$H(y,p) := \int_{\mathbb{T}^d} \frac{p^2}{2} + \frac{1}{2}\big((\nabla y)^2 + V(x)y^2\big) + F(x,y)\,dx$$

and $\partial_y F(x,y) = -f(x,y)$. The variables (y,p) are "Darboux coordinates".

2. The NLS equation (2) can be written as the infinite dimensional complex system

$$u_t = \mathrm{i}\delta_{\bar u} H(u)\,, \quad H(u) := \int_{\mathbb{T}^d} |\nabla u|^2 + V(x)|u|^2 - F(x,u)\,dx\,.$$

Actually (2) is a real Hamiltonian PDE in the variables $(a,b) \in \mathbb{R}^2$, real and imaginary part of u. Denoting the real valued potential $W(a,b) := F(x,a+\mathrm{i}b)$, so that

$$\partial_{\bar u} F(x,a+\mathrm{i}b) := \frac{1}{2}(\partial_a + \mathrm{i}\partial_b)W(a,b) = f(x,a+\mathrm{i}b)\,,$$

the NLS equation (2) reads

$$\frac{d}{dt}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \Delta b - V(x)b + \frac{1}{2}\partial_b W(a,b) \\ -\Delta a + V(x)a - \frac{1}{2}\partial_a W(a,b) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 & -Id \\ Id & 0 \end{pmatrix}\begin{pmatrix} \delta_a H(a,b) \\ \delta_b H(a,b) \end{pmatrix}$$

with real valued Hamiltonian $H(a,b) := H(a+\mathrm{i}b)$ and δ_a, δ_b denote the L^2-real gradients.

3. The KdV equation (3) is the Hamiltonian PDE

$$u_t = \partial_x \delta H(u)\,, \qquad H(u) = \int_{\mathbb{T}} \frac{u_x^2}{2} - \frac{u^3}{3} + f(x,u,u_x)\,dx\,, \tag{4}$$

where δH denotes the $L^2(\mathbb{T}_x)$ gradient. A natural phase space for (4) is

$$H_0^1(\mathbb{T}_x) := \Big\{ u(x) \in H^1(\mathbb{T},\mathbb{R}) \ : \ \int_{\mathbb{T}} u(x)dx = 0 \Big\} .$$

In the present paper we shall focus on the geometric construction of the approximate right inverse for the equation satisfied by an embedded torus of an Hamiltonian system, stressing the algebraic aspects of the proof. In the papers [6], [3] the analytic estimates, and small technical variations, may disturb the geometric clarity of the approach.

2 Normal form close to an invariant torus

We consider the toroidal phase space

$$\mathscr{P} := \mathbb{T}^\nu \times \mathbb{R}^\nu \times E \qquad \text{where} \qquad \mathbb{T}^\nu := \mathbb{R}^\nu / (2\pi\mathbb{Z})^\nu$$

is the standard flat torus and E is a real Hilbert space with scalar product $\langle \ , \ \rangle$. We denote by $u := (\theta, I, z)$ the variables of $\mathscr{P}$. We call (θ, I) the "action-angle" variables and z the "normal" variables. We assume that E is endowed with a constant exact symplectic 2-form

$$\Omega_E(z,w) = \langle \bar{J} z, w \rangle , \quad \forall z, w \in E , \tag{5}$$

where $\bar{J} : E \to E$ is an antisymmetric bounded linear operator with trivial kernel. Thus $\mathscr{P}$ is endowed with the symplectic 2-form

$$\Omega := (dI \wedge d\theta) \oplus \Omega_E \tag{6}$$

which is exact, namely

$$\Omega = d\lambda \tag{7}$$

where d denotes the exterior derivative and λ is the 1-form on $\mathscr{P}$ defined by

$$\lambda_{(\theta,I,z)} : \mathbb{R}^\nu \times \mathbb{R}^\nu \times E \to \mathbb{R} ,$$

$$\lambda_{(\theta,I,z)}[\hat{\theta}, \hat{I}, \hat{z}] := I \cdot \hat{\theta} + \frac{1}{2} \langle \bar{J} z, \hat{z} \rangle , \quad \forall (\hat{\theta}, \hat{I}, \hat{z}) \in \mathbb{R}^\nu \times \mathbb{R}^\nu \times E . \tag{8}$$

The dot " $\cdot$ " denotes the usual scalar product of $\mathbb{R}^\nu$.

Remark 1. For the NLW equation $E = L^2 \times L^2$ with $L^2 := L^2(\mathbb{T}^d, \mathbb{R})$, the operator defining the symplectic structure is $\bar{J} = \begin{pmatrix} 0 & -Id \\ Id & 0 \end{pmatrix}$ and $\langle \ , \ \rangle$ is the L^2 real scalar product on E. The transposed operator $\bar{J}^T = -\bar{J}$ (with respect to $\langle \, , \, \rangle$) and the inverse $\bar{J}^{-1} = \bar{J}^T$. The same setting holds for the NLS equation with real valued Hamiltonian, writing it as a real Hamiltonian system in the real and imaginary part. For the

KdV equation $E = L^2_0(\mathbb{T},\mathbb{R}) := \{z \in L^2(\mathbb{T},\mathbb{R}) : \int_{\mathbb{T}} z(x)\,dx = 0\}$ the operator $\bar{J} = \partial_x^{-1}$ and $\langle\ ,\ \rangle$ is the L^2 real scalar product. Here the transposed operator $\bar{J}^T = -\bar{J}$ and the inverse $\bar{J}^{-1} = \partial_x$ is unbounded.

Given a Hamiltonian function $H : \mathscr{D} \subset \mathscr{P} \to \mathbb{R}$, we consider the Hamiltonian system

$$u_t = X_H(u)\,, \qquad \text{where} \qquad dH(u)[\cdot] = -\Omega(X_H(u),\cdot) \tag{9}$$

formally defines the Hamiltonian vector field X_H. For infinite dimensional systems (i.e. PDEs) the Hamiltonian H is, in general, well defined and smooth only on a dense subset $\mathscr{D} = \mathbb{T}^\nu \times \mathbb{R}^\nu \times E_1 \subset \mathscr{P}$ where $E_1 \subset E$ is a dense subspace of E. We require that, for all $(\theta, I) \in \mathbb{T}^\nu \times \mathbb{R}^\nu$, $\forall z \in E_1$, the Hamiltonian H admits a gradient $\nabla_z H$, defined by

$$d_z H(\theta,I,z)[h] = \langle \nabla_z H(\theta,I,z), h\rangle\,, \quad \forall h \in E_1\,, \tag{10}$$

and that $\nabla_z H(\theta,I,z) \in E$ is in the space of definition of the (possibly unbounded) operator $J := -\bar{J}^{-1}$. Then by (9), (5), (6), (10) the Hamiltonian vector field $X_H : \mathbb{T}^\nu \times \mathbb{R}^\nu \times E_1 \mapsto \mathbb{R}^\nu \times \mathbb{R}^\nu \times E$ writes

$$X_H = (\partial_I H, -\partial_\theta H, J\nabla_z H)\,, \qquad J := -\bar{J}^{-1}\,. \tag{11}$$

A continuous curve $[t_0,t_1] \ni t \mapsto u(t) \in \mathbb{T}^\nu \times \mathbb{R}^\nu \times E$ is a solution of the Hamiltonian system (9) if it is C^1 as a map from $[t_0,t_1]$ to $\mathbb{T}^\nu \times \mathbb{R}^\nu \times E_1$ and $u_t(t) = X_H(u(t))$, $\forall t \in [t_0,t_1]$. For PDEs, the flow map Φ^t_H may not be well-defined everywhere. The next arguments, however, will not require to solve the initial value problem, but only a functional equation in order to find solutions which are quasi-periodic, see (16).

We refer to [21] for a general functional setting of Hamiltonian PDEs on scales of Hilbert spaces.

Remark 2. In the example of remark 1 for NLW and NLS, we can take $E_1 := H^2 \times H^2$. Then the Hamiltonian vector field $J\nabla_z H : H^2 \times H^2 \to L^2 \times L^2$. For KdV we can take $E_1 = H^3_0(\mathbb{T})$ so that $J\nabla_z H = -\partial_x \nabla_z H : H^3_0(\mathbb{T}) \mapsto L^2_0(\mathbb{T})$.

We suppose that (9) possesses an embedded invariant torus

$$\varphi \mapsto i(\varphi) := (\theta_0(\varphi), I_0(\varphi), z_0(\varphi))\,, \tag{12}$$

$$i \in C^1(\mathbb{T}^\nu, \mathscr{P}) \cap C^0(\mathbb{T}^\nu, \mathscr{P} \cap \mathbb{T}^\nu \times \mathbb{R}^\nu \times E_1)\,, \tag{13}$$

which supports a quasi-periodic solution with non-resonant frequency vector $\omega \in \mathbb{R}^\nu$, more precisely

$$i \circ \Psi^t_\omega = \Phi^t_H \circ i\,, \quad \forall t \in \mathbb{R}\,, \tag{14}$$

where Φ^t_H denotes the flow generated by X_H and

$$\Psi^t_\omega : \mathbb{T}^\nu \to \mathbb{T}^\nu\,, \quad \Psi^t_\omega(\varphi) := \varphi + \omega t\,, \tag{15}$$

is the translation flow of vector ω on $\mathbb{T}^\nu$. Since $\omega \in \mathbb{R}^\nu$ is non-resonant, namely $\omega \cdot k \neq 0$, $\forall k \in \mathbb{Z}^\nu \setminus \{0\}$, each orbit of (Ψ^t_ω) is *dense* in $\mathbb{T}^\nu$. Note that (14) *only* requires that the flow Φ_H^t is well defined and smooth on the compact manifold $\mathscr{T} := i(\mathbb{T}^\nu) \subset \mathscr{P}$ and $(\Phi_H^t)_{|\mathscr{T}} = i \circ \Psi_\omega^t \circ i^{-1}$. This remark is important because, for PDEs, the flow could be ill-posed in a neighborhood of $\mathscr{T}$. From a functional point of view (14) is equivalent to the equation

$$\omega \cdot \partial_\varphi i(\varphi) - X_H(i(\varphi)) = 0 \,. \tag{16}$$

Remark 3. In the sequel we will formally differentiate several times the torus embedding i, so that we assume more regularity than (13). In the framework of a Nash-Moser scheme, the approximate torus embedding solutions i are indeed regularized at each step.

We require that $\theta_0 : \mathbb{T}^\nu \to \mathbb{T}^\nu$ is a diffeomorphism of $\mathbb{T}^\nu$ isotopic to the identity. Then the embedded torus $\mathscr{T} := i(\mathbb{T}^\nu)$ is a smooth graph over $\mathbb{T}^\nu$. Moreover the lift on $\mathbb{R}^\nu$ of θ_0 is a function

$$\theta_0 : \mathbb{R}^\nu \to \mathbb{R}^\nu, \quad \theta_0(\varphi) = \varphi + \Theta_0(\varphi) \,, \tag{17}$$

where $\Theta_0(\varphi)$ is 2π-periodic in each component φ_i, $i = 1, \dots, \nu$, with invertible Jacobian matrix $D\theta_0(\varphi) = Id + D\Theta_0(\varphi)$, $\forall \varphi \in \mathbb{T}^\nu$. In the usual applications $D\Theta_0$ is small and ω is a Diophantine vector, namely

$$|\omega \cdot k| \geq \frac{\gamma}{|k|^\tau} \,, \quad \forall k \in \mathbb{Z}^\nu \setminus \{0\} \,. \tag{18}$$

In such a case we say that the invariant torus embedding $\varphi \mapsto i(\varphi)$ is Diophantine. The torus $\mathscr{T}$ is the graph of the function (see (12) and (17))

$$j := i \circ \theta_0^{-1} \,, \quad j : \mathbb{T}^\nu \to \mathbb{T}^\nu \times \mathbb{R}^\nu \times E \,, \quad j(\theta) := (\theta, \tilde{I}_0(\theta), \tilde{z}_0(\theta)) \,, \tag{19}$$

namely

$$\mathscr{T} = \Big\{ (\theta, \tilde{I}_0(\theta), \tilde{z}_0(\theta)) \quad \text{where} \quad \tilde{I}_0 := I_0 \circ \theta_0^{-1} \,, \; \tilde{z}_0 := z_0 \circ \theta_0^{-1} \Big\}. \tag{20}$$

We want to introduce a symplectic set of coordinates (ψ, y, w) near the invariant torus $\mathscr{T} := i(\mathbb{T}^\nu)$ such that $\mathscr{T}$ is described by $\{y = 0, w = 0\}$ and the restricted flow is simply $\psi(t) = \varphi + \omega t$. We look for a diffeomorphism of the phase space of the form

$$\begin{pmatrix} \theta \\ I \\ z \end{pmatrix} = G \begin{pmatrix} \psi \\ y \\ w \end{pmatrix} := \begin{pmatrix} \theta_0(\psi) \\ I_0(\psi) + B_1(\psi) y + B_2(\psi) w \\ z_0(\psi) + w \end{pmatrix}$$

where $B_1(\psi) : \mathbb{R}^\nu \to \mathbb{R}^\nu$, $B_2(\psi) : E \to \mathbb{R}^\nu$ are linear operators. Note that in the first component G is just the diffeomorphism of $\mathbb{T}^\nu$ induced by the torus embedding and that G is linear in y, w (actually the third component of G is a translation in w).

Remark 4. The above change of variables G is a particular class of those used by Moser in [23], which also allow to "rotate" linearly the third component as $z_0(\psi)+C_1(\psi)y+C_2(\psi)w$.

In order to find a symplectic set of coordinates as above, namely to find $B_1(\psi)$, $B_2(\psi)$ such that G is symplectic, we exploit the isotropy of the invariant torus $i(\mathbb{T}^\nu)$, i.e. the fact that the 2-form Ω vanishes on the tangent space to $i(\mathbb{T}^\nu)\subset\mathscr{P}$,

$$0 = i^*\Omega = i^* d\lambda = d(i^*\lambda)\,. \tag{21}$$

In other words, the 1-form $i^*\lambda$ on $\mathbb{T}^\nu$ is closed. It is natural to use such property: also in the proof of the classical Arnold-Liouville theorem (see e.g. [24]), the first step for the construction of the symplectic action-angle variables is to show that a maximal torus supporting an non-resonant rotation is Lagrangian. We first prove the isotropy of an invariant torus as in [19], [17] (Lemma 5 will provide a more precise result).

Lemma 1. *The invariant torus $i(\mathbb{T}^\nu)$ is isotropic.*

Proof. By (14) the pullback

$$(i\circ\Psi_\omega^t)^*\Omega = (\Phi_H^t\circ i)^*\Omega = i^*\Omega\,. \tag{22}$$

For smooth Hamiltonian systems in finite dimension (22) is true because the 2-form Ω is invariant under the flow map Φ_H^t (i.e. $(\Phi_H^t)^*\Omega=\Omega$). In our setting, the flow (Φ_H^t) may not be defined everywhere, but Φ_H^t is well defined on $i(\mathbb{T}^\nu)$ by the assumption (14), and still preserves Ω on the manifold $i(\mathbb{T}^\nu)$, see the proof of Lemma 5 for details.

Next, denoting the 2-form $(i^*\Omega)(\varphi)=\sum_{i<j}A_{ij}(\varphi)d\varphi_i\wedge d\varphi_j$, we have

$$(i\circ\Psi_\omega^t)^*\Omega = (\Psi_\omega^t)^*\circ i^*\Omega = \sum\nolimits_{i<j}A_{ij}(\varphi+\omega t)d\varphi_i\wedge d\varphi_j\,,$$

and so (22) implies that $A_{ij}(\varphi+\omega t)=A_{ij}(\varphi)$, $\forall t\in\mathbb{R}$. Since the orbit $\{\varphi+\omega t\}$ is dense on $\mathbb{T}^\nu$ (ω is non-resonant) and each function A_{ij} is continuous, it implies that

$$A_{ij}(\varphi)=c_{ij}\,,\ \forall\varphi\in\mathbb{T}^\nu\,,\quad i.e.\ \ i^*\Omega=\sum\nolimits_{i<j}c_{ij}d\varphi_i\wedge d\varphi_j$$

is constant. But, by (7), the 2-form $i^*\Omega=i^*d\lambda=d(i^*\lambda)$ is also exact. Thus each $c_{ij}=0$ namely $i^*\Omega=0$. ☐

We now consider the diffeomorphism of the phase space

$$\begin{pmatrix}\theta\\ I\\ z\end{pmatrix} = G\begin{pmatrix}\psi\\ y\\ w\end{pmatrix} := \begin{pmatrix}\theta_0(\psi)\\ I_0(\psi)+[D\theta_0(\psi)]^{-T}y-\big[D\tilde z_0(\theta_0(\psi))\big]^T\bar J w\\ z_0(\psi)+w\end{pmatrix} \tag{23}$$

where $\tilde z_0(\theta):=(z_0\circ\theta_0^{-1})(\theta)$, see (20). The transposed operator $\big[D\tilde z_0(\theta)\big]^T : E\to\mathbb{R}^\nu$ is defined by the duality relation

$$\big[D\tilde z_0(\theta)\big]^T w \cdot \hat\theta = \langle w, D\tilde z_0(\theta)[\hat\theta]\rangle\,, \quad \forall w \in E\,,\ \hat\theta \in \mathbb{R}^\nu\,.$$

Lemma 2. *Let i be an isotropic torus embedding. Then G is symplectic.*

Proof. We may see G as the composition $G := G_2 \circ G_1$ of the diffeomorphisms

$$\begin{pmatrix}\theta\\ I\\ z\end{pmatrix} = G_1\begin{pmatrix}\psi\\ y\\ w\end{pmatrix} := \begin{pmatrix}\theta_0(\psi)\\ [D\theta_0(\psi)]^{-T} y\\ w\end{pmatrix}$$

and

$$\begin{pmatrix}\theta\\ I\\ z\end{pmatrix} \mapsto G_2\begin{pmatrix}\theta\\ I\\ z\end{pmatrix} := \begin{pmatrix}\theta\\ \tilde I_0(\theta) + I - \big[D\tilde z_0(\theta)\big]^T \bar J z\\ \tilde z_0(\theta) + z\end{pmatrix} \tag{24}$$

where $\tilde I_0 := I_0 \circ \theta_0^{-1}$, $\tilde z_0 := z_0 \circ \theta_0^{-1}$, see (20). We claim that both G_1, G_2 are symplectic, whence the lemma follows.

G_1 IS SYMPLECTIC. Since G_1 is the identity in the third component, it is sufficient to check that $(\psi, y) \mapsto (\theta_0(\psi), [D\theta_0(\psi)]^{-T} y)$ is a symplectic diffeomorphism on $\mathbb{T}^\nu \times \mathbb{R}^\nu$, which is a direct calculus.

G_2 IS SYMPLECTIC. We prove that $G_2^*\lambda - \lambda$ is closed and so (see (7)) $G_2^*\Omega = G_2^* d\lambda = dG_2^*\lambda = d\lambda = \Omega$. By (24) and the definition of pullback we have

$$\begin{aligned}(G_2^*\lambda)_{(\theta,I,z)}[\hat\theta, \hat I, \hat z] = {}& \big(\tilde I_0(\theta) + I - \big[D\tilde z_0(\theta)\big]^T \bar J z\big)\cdot\hat\theta\\ &+ \frac12\langle \bar J(\tilde z_0(\theta) + z), \hat z + D\tilde z_0(\theta)[\hat\theta]\rangle\,.\end{aligned}$$

Therefore (recall (8))

$$\begin{aligned}\big((G_2^*\lambda)_{(\theta,I,z)} - \lambda_{(\theta,I,z)}\big)[\hat\theta,\hat I,\hat z] = {}& \big(\tilde I_0(\theta) - \big[D\tilde z_0(\theta)\big]^T \bar J z\big)\cdot\hat\theta + \frac12\langle \bar J\tilde z_0(\theta), \hat z\rangle\\ &+ \frac12\langle \bar J\tilde z_0(\theta), D\tilde z_0(\theta)[\hat\theta]\rangle + \frac12\langle \bar J z, D\tilde z_0(\theta)[\hat\theta]\rangle\\ = {}& \tilde I_0(\theta)\cdot\hat\theta + \frac12\langle \bar J\tilde z_0(\theta), D\tilde z_0(\theta)[\hat\theta]\rangle\\ &+ \frac12\langle \bar J\tilde z_0(\theta), \hat z\rangle + \frac12\langle \bar J D\tilde z_0(\theta)[\hat\theta], z\rangle\,,\end{aligned} \tag{25}$$

having used that $\bar J^T = -\bar J$. We first note that the 1-form

$$(\hat\theta, \hat I, \hat z) \mapsto \langle \bar J\tilde z_0(\theta), \hat z\rangle + \langle \bar J D\tilde z_0(\theta)[\hat\theta], z\rangle = d(\langle \bar J\tilde z_0(\theta), z\rangle)[\hat\theta, \hat I, \hat z] \tag{26}$$

is exact. Moreover

$$\tilde I_0(\theta)\cdot\hat\theta + \frac12\langle \bar J\tilde z_0(\theta), D\tilde z_0(\theta)[\hat\theta]\rangle = (j^*\lambda)_\theta[\hat\theta] \tag{27}$$

(recall (8)) where $j := i \circ \theta_0^{-1}$ see (19). Hence (25), (26), (27) imply

$$(G_2^*\lambda)_{(\theta,I,z)} - \lambda_{(\theta,I,z)} = \pi^*(j^*\lambda)_{(\theta,I,z)} + d\Big(\frac{1}{2}\langle \bar{J}\tilde{z}_0(\theta), z\rangle\Big),$$

where $\pi : \mathbb{T}^\nu \times \mathbb{R}^\nu \times E \to \mathbb{T}^\nu$ is the canonical projection.

Since the torus $j(\mathbb{T}^\nu) = i(\mathbb{T}^\nu)$ is isotropic (Lemma 1) the 1-form $j^*\lambda$ on $\mathbb{T}^\nu$ is closed (as $i^*\lambda$, see (21)). This concludes the proof. □

Remark 5. A torus which is a graph over $\mathbb{T}^\nu$, i.e. $\theta \mapsto j(\theta) = (\theta, I_1(\theta), z_1(\theta))$ is isotropic if and only if $I_1(\theta) = \gamma + dU(\theta) - \frac{1}{2}[Dz_1(\theta)]^T \bar{J} z_1(\theta)$ for some constant $\gamma \in \mathbb{R}^\nu$ and $U : \mathbb{T}^\nu \to \mathbb{R}$. This follows from (27) and Corollary 1.

Since G is symplectic (note that Lemma 2 only requires i to be isotropic), the transformed Hamiltonian vector field

$$G^* X_H := (DG)^{-1} \circ X_H \circ G = X_K, \quad K := H \circ G,$$

is still Hamiltonian. By construction (see (23)) the torus $\{y = 0, w = 0\}$ is invariant and (16) implies $X_K(\psi, 0, 0) = (\omega, 0, 0)$ (see also Lemma 8). As a consequence, the Taylor expansion of the transformed Hamiltonian K in these new coordinates assumes the normal form

$$K = const + \omega \cdot y + \frac{1}{2} A(\psi) y \cdot y + \langle C(\psi) y, w\rangle + \frac{1}{2}\langle B(\psi) w, w\rangle + O_3(y, w) \qquad (28)$$

where $O_3(y,w)$ collects all the terms at least cubic in (y,w), and the operators A and B are symmetric. If, furthermore, ω is Diophantine we can perform, by standard perturbation theory, a symplectic change of coordinates which conjugates K to another Hamiltonian of the form

$$K_1 := \omega \cdot y + \frac{1}{2}\bar{A} y \cdot y + \langle C_1(\psi) y, w\rangle + \frac{1}{2}\langle B_1(\psi) w, w\rangle + O_3(y, w) \qquad (29)$$

where the constant matrix $\bar{A}$ is the average $\bar{A} := \int_{\mathbb{T}^\nu} A(\psi) d\psi$. This is the general normal form for a Hamiltonian near a Diophantine invariant torus.

Summarizing we have proved the following theorem:

Theorem 1. *Let $\mathscr{T} = i(\mathbb{T}^\nu)$ be an embedded torus, see* (12)-(13)*, which is a smooth graph over $\mathbb{T}^\nu$, see* (19)-(20)*, invariant for the Hamiltonian vector field X_H, and on which the flow is conjugate to the translation flow of vector ω, see* (14)-(15)*. Assume moreover that $\mathscr{T}$ is* ISOTROPIC*, a property which is automatically verified, in particular, if ω is non-resonant.*

Then there exist symplectic coordinates (ψ, y, w) in which $\mathscr{T}$ is described by $\mathbb{T}^\nu \times \{0\} \times \{0\}$ and the Hamiltonian assumes the normal form (28)*, i.e. the torus $\mathscr{T} = G(\mathbb{T}^\nu \times \{0\} \times \{0\})$ where G is the symplectic diffeomorphism defined in* (23)*, and the Hamiltonian $H \circ G$ has the Taylor expansion* (28) *in a neighborhood of the invariant torus. If, moreover, ω is Diophantine, see* (18)*, there is a symplectic change of coordinates in which the Hamiltonian assumes the normal form* (29).

Remark 6. If the torus $\mathscr{T}$ is isotropic, even if it is filled by periodic orbits (resonant torus), i.e. $\omega = 2\pi k/T$ for some $k \in \mathbb{Z}^\nu$, the previous theorem provides the normal form (28). For an application to Lagrangian tori see [1].

What is usually called a KAM normal form for isotropic Diophantine invariant tori is an Hamiltonian of the form

$$K_2 := \omega \cdot y + \frac{1}{2}\bar{A}y \cdot y + \langle \bar{C}y, w\rangle + \frac{1}{2}\langle \bar{B}w, w\rangle + O_3(y,w) \qquad (30)$$

where also the matrices $\bar{B}, \bar{C}$ are constant, see e.g. [15], [25], [20], [28]. The possibility to obtain such a normal form is related to the verification of the so called "second order Melnikov" non resonance conditions. This may be a difficult task for PDEs in higher space dimension because of the possible multiplicity of the normal frequencies, see e.g. [16], [26] for NLS.

The normal form (28) is relevant also in view of a Nash-Moser approach, because it provides a control of the linearized equations in the normal bundle of the torus. The linearized Hamiltonian system X_K at the trivial solution $(\psi, y, w)(t) = (\omega t, 0, 0)$ is

$$\begin{cases} \dot{\psi} - A(\omega t)y - [C(\omega t)]^T w = 0 \\ \dot{y} = 0 \\ \dot{w} - J\big(B(\omega t)w + C(\omega t)y\big) = 0. \end{cases} \qquad (31)$$

For applying the Nash-Moser scheme (section 4) we have to solve, at each step, the system of equations (31) with non-zero second members. Note that the second equation is decoupled from the others. Inserting its solution in the third equation we are reduced to solve a quasi-periodically forced linear equation in w. This may vary considerably for different PDEs. The difficulty is that $B(\omega t)$ is not constant. A way to solve it is to conjugate it to a constant coefficient equation (with second order Melnikov non resonance conditions), as for the normal form (30). For PDEs in higher space dimension this is not always possible and one proceeds with a multiscale analysis as in [10]-[11], [4]-[8] which requires only the first order Melnikov non-resonance conditions. Finally one solves the first equation in (31) for the angle component.

3 A Nash-Moser functional approach to KAM

We now describe the strategy for proving an abstract normal form KAM theorem by using a Nash-Moser implicit function theorem. We choose the setting of a perturbation of a parameter dependent family of isochronous linear Hamiltonian systems.

Let $\mathscr{O} \subset \mathbb{R}^\nu$ be an open set of parameters. We consider a family of Hamiltonians $H : [0, \varepsilon_0) \times \mathscr{O} \times \mathscr{P} \to \mathbb{R}$ like

$$H = H(\varepsilon, \alpha, u) = \mathscr{N}(\alpha, \theta, I, z) + \varepsilon P(\varepsilon, \alpha, \theta, I, z), \qquad (32)$$

which are perturbations of a parameter-dependent normal form

$$\mathcal{N}(\alpha,\theta,I,z) = \alpha \cdot I + \frac{1}{2}\langle N(\alpha,\theta)z,z\rangle \tag{33}$$

where $N(\alpha,\theta)$ is a symmetric operator. We suppose that, as in (10), (11), the Hamiltonian vector fields $z \mapsto JN(\alpha,\theta)z$, $J\nabla_z P(\alpha,\theta,I,z)$ are well defined and smooth maps from a dense subspace $E_1 \subset E$ into E. Note that $\mathcal{N}$ may depend on the angle variables θ (in the normal directions z).

Remark 7. In applications, the parameters α may vary with the "actions" of the unperturbed invariant tori (this approach was first used in [23]), or depend on the mass of a planet as in [17], or may be "external" parameters induced, for example, by the potential as in [21], [28], etc...

The normal form $\mathcal{N}$ possesses the invariant torus $\mathbb{T}^\nu \times \{0\} \times \{0\}$ on which the motion is endowed by the constant field α.

Remark 8. If the normal form $N(\alpha,\theta) = N(\alpha)$ is constant, i.e. it does not depend on the angles θ, the unperturbed torus $\mathbb{T}^\nu \times \{0\} \times \{0\}$ is said "reducible". In applications this is the common situation in order to start with perturbation theory.

The goal is then to prove that:

- for ε small enough, for "most" Diophantine vectors $\omega \in \mathcal{C} \subset \mathcal{O}$, there exists a value of the parameters $\alpha := \alpha_\infty(\omega,\varepsilon) = \omega + O(\varepsilon)$ and a ν-dimensional embedded torus $\mathcal{T} = i(\mathbb{T}^\nu)$ close to $\mathbb{T}^\nu \times \{0\} \times \{0\}$, invariant for the Hamiltonian vector field $X_{H(\varepsilon,\alpha_\infty(\omega,\varepsilon),\cdot)}$ and supporting quasi-periodic solutions with frequency ω. In view of (16), this is equivalent to looking for a solution $\varphi \mapsto i(\varphi) \in \mathbb{T}^\nu \times \mathbb{R}^\nu \times E$, close to $(\varphi,0,0)$, of the embedding equation

$$(\omega \cdot \partial_\varphi) i(\varphi) - X_{H(\varepsilon,\alpha_\infty(\omega,\varepsilon),\cdot)}(i(\varphi)) = 0\,, \tag{34}$$

  for some value $\alpha := \alpha_\infty(\omega,\varepsilon)$ of the parameters to be determined.

The set of frequencies $\omega \in \mathcal{C} \subset \mathcal{O}$ for which the invariant torus exists usually forms a Cantor like set. The measure of the set $\mathcal{C}$ (in particular that $\mathcal{C} \neq \emptyset$) clearly depends of the properties of the Hamiltonian H, in particular for infinite dimensional Hamiltonian system. The parameter $\alpha := \alpha_\infty(\omega,\varepsilon)$ is adjusted along the iterative Nash-Moser proof in order control the average of the first component of the Hamilton's equation (36), in particular for solving the linearized equation (82).

The function $\omega \mapsto \alpha_\infty(\omega,\varepsilon)$ is invertible and it may be proved to be smooth in ω (if the Hamiltonian H is smooth). Then, in applications, one may ask if, *given* $\beta \in \mathbb{R}^\nu$, there exists a value of $\omega = \alpha_\infty^{-1}(\beta)$ in the Cantor set of parameters $\mathcal{C} \subset \mathcal{O}$ for which (34) has a solution. In such a case one has proved the existence of a quasi-periodic solution of the given Hamiltonian $\beta \cdot I + \frac{1}{2}\langle N(\beta,\theta)z,z\rangle + \varepsilon P$. This perspective is the spirit of the Théorème de conjugaison hypothétique of Herman presented in [17].

Remark 9. Variants are possible. For example we could develop a KAM theorem for Hamiltonians which are perturbations of a non-isocronous (or Kolmogorov) normal form

$$H = \alpha \cdot I + \frac{1}{2} L(\alpha, \theta) I \cdot I + \langle M(\alpha, \theta) I, z \rangle + \frac{1}{2} \langle N(\alpha, \theta) z, z \rangle + \varepsilon P .$$

This is the setting, for example, considered in [29]. Actually this case may be reduced to (32) by a rescaling $\mathscr{R}_\varepsilon : (I, z) \mapsto (\varepsilon^{2a} I, \varepsilon^a z)$. Note that the transformed symplectic 2-form is $\mathscr{R}_\varepsilon^* \Omega = \varepsilon^{2a} \Omega$. A technical advantage of dealing with the parameter dependent family of isochronous normal forms (33) is that the linearized equations are simpler.

In order to find solutions of (34) we look for zeros of the nonlinear operator

$$\begin{aligned}
\mathscr{F}(\varepsilon, X) &:= (\omega \cdot \partial_\varphi) i(\varphi) - X_{H_\mu(\alpha,\cdot)}(i(\varphi)) \qquad (35) \\
&= \begin{pmatrix} \partial_\omega \theta_0(\varphi) - \partial_I H(\varepsilon, \alpha, i(\varphi)) \\ \partial_\omega I_0(\varphi) + \partial_\theta H(\varepsilon, \alpha, i(\varphi)) + \mu \\ \partial_\omega z_0(\varphi) - J \nabla_z H(\varepsilon, \alpha, i(\varphi)) \end{pmatrix} \\
&= \begin{pmatrix} \partial_\omega \theta_0(\varphi) - \alpha - \varepsilon \partial_I P(\alpha, i(\varphi)) \\ \partial_\omega I_0(\varphi) + \frac{1}{2} \partial_\theta \langle N(\alpha, \theta_0(\varphi)) z_0(\varphi), z_0(\varphi) \rangle + \varepsilon \partial_\theta P(\alpha, i(\varphi)) + \mu \\ \partial_\omega z_0(\varphi) - J N(\alpha, \theta_0(\varphi)) z_0(\varphi) - \varepsilon J \nabla_z P(\alpha, i(\varphi)) \end{pmatrix} \qquad (36)
\end{aligned}$$

in the unknowns

$$X := (\alpha, \mu, i(\varphi))$$

where the torus embedding

$$i(\varphi) := (\theta_0(\varphi), I_0(\varphi), z_0(\varphi)) := (\varphi, 0, 0) + (\Theta_0(\varphi), I_0(\varphi), z_0(\varphi)) \qquad (37)$$

and we use the shorter notation

$$\partial_\omega := \omega \cdot \partial_\varphi .$$

Note that $\mathscr{F}(\varepsilon, X) = 0$ is the equation $\partial_\omega i(\varphi) - X_{H_\mu(\alpha,\cdot)}(i(\varphi)) = 0$ for an embedded invariant torus of the non-exact Hamiltonian system generated by the Hamiltonian

$$H_\mu := H_\mu(\alpha, \cdot) : \mathbb{R}^\nu \times \mathbb{R}^\nu \times E \to \mathbb{R}, \quad H_\mu := H + \mu \cdot \theta . \qquad (38)$$

Remark that the Hamiltonian vector field X_{H_μ} is periodic in θ, unlike H_μ. Non-exact here means that $\Omega(X_{H_\mu}, \cdot) = -dH - \mu$ is a closed, non-exact 1-form on the phase space $\mathbb{T}^\nu \times \mathbb{R}^\nu \times E$. It is well-known that a non-exact Hamiltonian system does not possess invariant tori for $\mu \neq 0$. Actually, as proved in Lemma 3 below, if $\mathscr{F}(\varepsilon, X) = 0$ then $\mu = 0$ and so $\varphi \mapsto i(\varphi)$ is an invariant torus for X_H itself. The 'counter-term" $\mu \in \mathbb{R}^\nu$ is introduced as a technical trick to control the average of the second component of the equation (36), in particular for solving the linearized equation (76).

Lemma 3. *Let us define the "error function"*

$$Z(\varphi) = (Z_1, Z_2, Z_3)(\varphi) := \mathscr{F}(\varepsilon, i, \alpha, \mu) = (\omega \cdot \partial_\varphi) i(\varphi) - X_{H_\mu(\alpha,\cdot)}(i(\varphi)) \tag{39}$$

where $\varphi \in \mathbb{T}^\nu$. *Then*

$$\mu = \frac{1}{(2\pi)^\nu} \int_{\mathbb{T}^\nu} -[DI_0(\varphi)]^T Z_1(\varphi) + [D\theta_0(\varphi)]^T Z_2(\varphi) + [Dz_0(\varphi)]^T \bar{J} Z_3(\varphi)\, d\varphi\,. \tag{40}$$

In particular, if $\partial_\omega i(\varphi) - X_{H_\mu}(i(\varphi)) = 0$ *then* $\mu = 0$ *and so* $\varphi \mapsto i(\varphi)$ *is the embedding of an invariant torus of* X_H.

Proof. Let $i_{\psi_0}(\varphi) := i(\varphi + \psi_0)$ be the translated torus embedding, for all $\psi_0 \in \mathbb{T}^\nu$. Since H is autonomous the "restricted" Hamiltonian action functional (recall (8))

$$\Phi(\psi_0) := \int_{\mathbb{T}^\nu} \lambda_{i_{\psi_0}(\varphi)}[\partial_\omega i_{\psi_0}(\varphi)] - H(i_{\psi_0}(\varphi))\, d\varphi = \Phi(0)$$

is constant. Differentiating Φ at $\psi_0 = 0$ and integrating by parts ∂_ω we get, for all $\zeta \in \mathbb{R}^\nu$, (see (9))

$$\begin{aligned} 0 = D_{\psi_0} \Phi(\psi_0)[\zeta] &= -\int_{\mathbb{T}^\nu} \Omega\big(\partial_\omega i(\varphi) - X_H(i(\varphi)), Di(\varphi)[\zeta]\big) d\varphi \\ &= -\int_{\mathbb{T}^\nu} \Omega\Big(Z(\varphi) - \mu \cdot \frac{\partial}{\partial I}, Di(\varphi)[\zeta]\Big) d\varphi \end{aligned} \tag{41}$$

by the definition of Z in (39), (38), and denoting the vector field $(0, \mu, 0) = \mu \cdot \frac{\partial}{\partial I}$. Recalling (6)-(7) the integral

$$\int_{\mathbb{T}^\nu} \Omega\Big(\mu \cdot \frac{\partial}{\partial I}, Di(\varphi)[\zeta]\Big) d\varphi = \int_{\mathbb{T}^\nu} \mu \cdot D\theta_0(\varphi)[\zeta]\, d\varphi = (2\pi)^\nu \mu \cdot \zeta$$

because the periodic function $D(\theta_0 - id) = D\Theta_0$ (see (37)) has zero average. Hence, by (41) we deduce

$$\mu \cdot \zeta = \frac{1}{(2\pi)^\nu} \int_{\mathbb{T}^\nu} \Omega\big(Z(\varphi), Di(\varphi)[\zeta]\big) d\varphi\,, \quad \forall \zeta \in \mathbb{R}^\nu\,,$$

which recalling (5)-(6) gives (40). ☐

The optimal expected smallness condition for the KAM existence result, namely for finding solutions of the nonlinear equation $\mathscr{F}(X, \varepsilon) = 0$, is

$$\varepsilon \gamma^{-1} \ll 1 \tag{42}$$

where γ is the Diophantine constant in (18) of the frequency vector ω. This is certainly the case for finite dimensional Lagrangian tori (the optimality follows for example by a time rescaling argument). If ω has to satisfy other Diophantine conditions of first and second order Melnikov type the required smallness conditions may be stronger, see e.g. [3].

Remark 10. Other functional formulations are possible. We could look for zeros of

$$\mathscr{F}(\varepsilon, j, \alpha, c) = \begin{pmatrix} \partial_\omega \theta_0(\varphi) - \partial_I H(\varepsilon, \alpha, i(\varphi)) \\ H(j(\theta)) - c \\ \partial_\omega z_0(\varphi) - J\nabla_z H(\varepsilon, \alpha, i(\varphi)) \end{pmatrix}$$

where $j(\theta) = (\theta, I_1(\theta), z_1(\theta))$ defines an isotropic torus as described in remark 5 and $i = j \circ \theta_0$. The unknowns are the diffeomorphism θ_0 of $\mathbb{T}^\nu$, the component $z_0 = z_1 \circ \theta_0$ of the torus embedding, the constant $\gamma \in \mathbb{R}^n$, the potential $U : \mathbb{T}^\nu \to \mathbb{R}$, and the value of the Hamiltonian $c \in \mathbb{R}$ and α_0. Actually, because of the presence of the parameter α, we may impose $\gamma = 0$.

As already said, a solution of the nonlinear equation $\mathscr{F}(\varepsilon, X) = 0$ is obtained by a Nash-Moser iterative scheme. The first approximate solution is

$$X_0 = (\omega, 0, \varphi, 0, 0)$$

(namely $\alpha = \omega, \mu = 0$, $i(\varphi) = (\varphi, 0, 0)$) so that

$$\mathscr{F}(0, X_0) = O(\varepsilon)\,.$$

Then the strategy is to obtain iteratively better and better approximate solutions of the equation $\mathscr{F}(\varepsilon, X) = 0$ by a quasi-quadratic scheme. Given an approximate solution X, we look for a better approximate solution $X' = X + h$ by a Taylor expansion (for simplicity we omit to write the dependence on ε)

$$\mathscr{F}(X') = \mathscr{F}(X+h) = \mathscr{F}(X) + d_X\mathscr{F}(X)[h] + O(|h|^2)\,.$$

The idea of the classical Newton iterative scheme is to define h as the solution of $\mathscr{F}(X) + d_X\mathscr{F}(X)[h] = 0$. Since the invertibility of the linear operator $d_X\mathscr{F}(X)$ may be a quite difficult task, Zehnder [29] noted that it is sufficient to find only an approximate right inverse of $d_X\mathscr{F}(X)$, namely a linear operator $T(X)$ such that

$$d_X\mathscr{F}(X) \circ T(X) - Id = O(|\mathscr{F}(X)|)\,. \tag{43}$$

Remark that, at a solution $\mathscr{F}(X) = 0$, the operator $T(X)$ is an exact right inverse of $d_X\mathscr{F}(X)$. Thus, defining the new approximate solution

$$X' = X + h\,, \quad h := -T(X)\mathscr{F}(X)\,, \tag{44}$$

we get by (43) that

$$\mathscr{F}(X') = \mathscr{F}(X) - d_X\mathscr{F}(X)[T(X)\mathscr{F}(X)] + O(|h|^2) = O(|\mathscr{F}(X)|^2)\,. \tag{45}$$

This scheme can be called a "quasi-Newton" scheme. In typical PDEs applications, the approximate right inverse $T(X)$ "loses derivatives" due to the small divisors.

However, since the scheme (44) is quadratic by (45), it can nevertheless converge to a solution if $\mathscr{F}(X_0)$ is sufficiently small (depending also on the norm of T). The scheme (44) is usually implemented in Banach scales of analytic functions, as, for example,

$$A_\sigma := \Big\{ u(\varphi) = \sum_{k \in \mathbb{Z}^\nu} u_k e^{\mathrm{i}k\cdot\varphi} \, : \, \|u\|_\sigma^2 := \sum_{k \in \mathbb{Z}^\nu} |u_k|^2 e^{2|k|\sigma}(1+|k|^{2s_0}) < +\infty \Big\} \tag{46}$$

for some $\sigma > 0$, $s_0 > \nu/2$. The approximate inverse operator T is usually "unbounded", satisfies Cauchy type estimates like

$$\|Tg\|_{\sigma'} \le \frac{C}{\gamma(\sigma-\sigma')^\tau} \|g\|_\sigma \,, \quad \forall \sigma' < \sigma \,, \tag{47}$$

and there is $\beta > 0$ such that, $\forall \sigma' < \sigma$, $\forall g \in A_\sigma$,

$$\big\| \big(d_X \mathscr{F}(X) \circ T(X) - Id \big) g \big\|_{\sigma'} \le C \frac{\|\mathscr{F}(\varepsilon, X)\|_\sigma}{\gamma(\sigma-\sigma')^\beta} \|g\|_\sigma \,. \tag{48}$$

The constants $\tau, \beta > 0$ are the "loss of derivatives".

On the other hand, in Banach spaces of functions with finite differentiability, as the Sobolev scale

$$H^s := \Big\{ u(\varphi) = \sum_{k \in \mathbb{Z}^\nu} u_k e^{\mathrm{i}k\cdot\varphi} \, : \, \|u\|_s^2 := \sum_{k \in \mathbb{Z}^\nu} |u_k|^2 (1+|k|^{2s}) < +\infty \Big\} \,, \tag{49}$$

the quasi-Newton scheme (44) does not converge because after finitely many steps the approximate solutions are no longer regular. Following Moser [22], it is necessary to insert a smoothing procedure at each step (Nash-Moser scheme). The approximate inverse usually satisfies estimates like: there are constants $p, \rho > 0$ ("loss of derivatives") such that, for all $s \in [s_0, S]$, $\forall g \in H^{s+\rho}$,

$$\|T(X)g\|_s \le C(s, \|X\|_{s_0+p}) \big(\|g\|_{s+\rho} + \|g\|_{s_0} \|X\|_{s+\rho} \big) \,, \tag{50}$$

and

$$\begin{aligned} \big\| \big(d_X \mathscr{F}(X) \circ T(X) - Id \big) g \big\|_s \le C(s, \|X\|_{s_0+p}) \Big(& \|\mathscr{F}(X)\|_{s_0} \|g\|_{s+\rho} + \\ & + \|\mathscr{F}(X)\|_{s+\rho} \|g\|_{s_0} + \|X\|_{s+\rho} \|\mathscr{F}(X)\|_{s_0} \|g\|_{s_0} \Big) \,. \end{aligned} \tag{51}$$

In this note we will not insist in the analytical aspects of the convergence, for which we refer to [29], [7], or [3], [6].

The linearized operator of (35) is

$$d_X \mathscr{F}(\varepsilon, X)[\hat{X}] = (\omega \cdot \partial_\varphi)\hat{\imath} - D_i X_{H_\mu(\alpha,\cdot)}(i)[\hat{\imath}] - D_\alpha X_{H_\mu(\alpha,\cdot)}(i)[\hat{\alpha}] - (0, \hat{\mu}, 0) \,.$$

It is rather difficult to invert it because all the components of the Hamiltonian vector field are coupled by $O(\varepsilon)$-non-constant coefficient terms. In the next section, following the ideas of section 2, we present a symplectic change of variable which approximately decouples the tangential directions (*i.e.* $(\hat\theta, \hat I)$) and the normal ones (*i.e.* $\hat z$), and thus enables to find an approximate right inverse of $d_X \mathscr{F}(\varepsilon, X)$.

4 Approximate right inverse

We first report a basic fact about 1-forms on a torus. We regard a 1-form $a = \sum_{i=1}^{\nu} a_i(\varphi) d\varphi_i$ equivalently as the vector field $\mathbf{a}(\varphi) = (a_1(\varphi), \ldots, a_\nu(\varphi))$.

Given a function $g : \mathbb{T}^\nu \to \mathbb{R}$ with zero average, we denote by $u := \Delta^{-1} g$ the unique solution of $\Delta u = g$ with zero average.

Lemma 4. (Helmotz decomposition) *A smooth vector field* $\mathbf{a}$ *on* $\mathbb{T}^\nu$ *may be decomposed as the sum of a conservative and a divergence-free vector field:*

$$\mathbf{a} = \nabla U + \mathbf{c} + \rho\,, \quad U : \mathbb{T}^\nu \to \mathbb{R}\,,\ \mathbf{c} \in \mathbb{R}^\nu\,,\ \mathrm{div}\rho = 0\,,\ \int_{\mathbb{T}^\nu} \rho\, d\varphi = 0\,. \tag{52}$$

The above decomposition is unique if we impose that the mean value of U vanishes. We have $U = \Delta^{-1}(\mathrm{div}\,\mathbf{a})$, the components of ρ are

$$\rho_j(\varphi) = \Delta^{-1} \sum_{k=1}^{\nu} \partial_{\varphi_k} A_{kj}(\varphi)\,, \quad A_{kj} := \partial_{\varphi_k} a_j - \partial_{\varphi_j} a_k\,, \tag{53}$$

and $c_j = (2\pi)^{-\nu} \int_{\mathbb{T}^\nu} a_j(\varphi)\, d\varphi$, $j = 1, \ldots, \nu$.

Proof. $\mathrm{div}(\mathbf{a} - \nabla U) = 0$ if and only if $\mathrm{div}\,\mathbf{a} = \Delta U$. This equation has the solution $U := \Delta^{-1}(\mathrm{div}\,\mathbf{a})$ (note that $\mathrm{div}\,\mathbf{a}$ has zero average). Hence (52) is achieved with $\rho := \mathbf{a} - \nabla U - \mathbf{c}$. By taking the φ-average we get that each $c_j = (2\pi)^{-\nu} \int_{\mathbb{T}^\nu} a_j(\varphi)\, d\varphi$. Let us now prove the expression (53) of ρ_j. We have $\partial_{\varphi_k} \rho_j - \partial_{\varphi_j} \rho_k = \partial_{\varphi_k} a_j - \partial_{\varphi_j} a_k =: A_{kj}$ because $\partial_{\varphi_j} \partial_{\varphi_k} U - \partial_{\varphi_k} \partial_{\varphi_j} U = 0$. For each $j = 1, \ldots, \nu$ we differentiate $\partial_{\varphi_k} \rho_j - \partial_{\varphi_j} \rho_k = A_{kj}$ with respect to φ_k and we sum in k, obtaining

$$\Delta \rho_j - \sum_{k=1}^{\nu} \partial_{\varphi_k \varphi_j} \rho_k = \sum_{k=1}^{\nu} \partial_{\varphi_k} A_{kj}\,.$$

Since $\sum_{k=1}^{\nu} \partial_{\varphi_k \varphi_j} \rho_k = \partial_{\varphi_j} \mathrm{div}\rho = 0$ then $\Delta \rho_j = \sum_{k=1}^{\nu} \partial_{\varphi_k} A_{kj}$ and (53) follows. □

Corollary 1. *Any closed* 1*-form on* $\mathbb{T}^\nu$ *has the form* $a(\varphi) = \mathbf{c} + dU$ *for some* $\mathbf{c} \in \mathbb{R}^\nu$.

Corollary 2. *Let $a(\varphi)$ be a* 1*-form on* $\mathbb{T}^\nu$*, and let ρ be defined by* (53)*. Then $a - \sum_{j=1}^{\nu} \rho_j(\varphi) d\varphi_j$ is closed.*

We quantify how an embedded torus $i(\mathbb{T}^\nu)$ is approximately invariant for the Hamiltonian vector field X_{H_μ} in terms of the "error function" $Z(\varphi)$, defined in (39). A torus embedding $i(\varphi) = (\theta_0(\varphi), I_0(\varphi), z_0(\varphi))$ which is only approximately invariant may not be isotropic. Consider the pullback 1-form on $\mathbb{T}^\nu$ (see (8))

$$(i^*\lambda)(\varphi) = \sum_{k=1}^{\nu} a_k(\varphi) d\varphi_k\,, \tag{54}$$

where

$$a_k(\varphi) := \Big[\big[D\theta_0(\varphi)\big]^T I_0(\varphi) + \frac{1}{2}[Dz_0(\varphi)]^T \bar{J} z_0(\varphi) \Big]_k$$
$$= I_0(\varphi) \cdot \frac{\partial \theta_0}{\partial \varphi_k}(\varphi) + \frac{1}{2} \langle \bar{J} z_0(\varphi), \frac{\partial z_0}{\partial \varphi_k}(\varphi) \rangle . \qquad (55)$$

The 1-form $i^*\lambda$ is only approximately closed, namely the 2-form (recall (7))

$$i^*\Omega = d(i^*\lambda) = \textstyle\sum_{k<j} A_{kj}(\varphi) d\varphi_k \wedge d\varphi_j , \qquad (56)$$

$$A_{kj}(\varphi) = \partial_{\varphi_k} a_j(\varphi) - \partial_{\varphi_j} a_k(\varphi) ,$$

is small. We call the coefficients (A_{kj}) the "lack of isotropy" of the approximate torus embedding $\varphi \mapsto i(\varphi)$. In Lemma 5 below we quantify their size in terms of the error function Z defined in (39).

We first recall that the Lie derivative of a k-form β with respect to the vector field Y is $L_Y \beta := \frac{d}{dt}\big[(\Phi_Y^t)^* \beta\big]_{|t=0}$ where Φ_Y^t denotes the flow generated by Y.

Given a function $g(\varphi)$ with zero average, we denote by $u := \partial_\omega^{-1} g$ the unique solution of $\partial_\omega u = g$ with zero average.

Lemma 5. *The "lack of isotropy" coefficients A_{kj} satisfy, $\forall \varphi \in \mathbb{T}^\nu$,*

$$(\omega \cdot \partial_\varphi) A_{kj}(\varphi) = \Omega\big(DZ(\varphi)\underline{e}_k, Di(\varphi)\underline{e}_j\big) + \Omega\big(Di(\varphi)\underline{e}_k, DZ(\varphi)\underline{e}_j\big) \qquad (57)$$

where $(\underline{e}_1, \ldots, \underline{e}_\nu)$ denotes the canonical basis of $\mathbb{R}^\nu$. Thus, since each A_{kj} has zero mean value, if the frequency vector $\omega \in \mathbb{R}^\nu$ is non-resonant,

$$A_{kj}(\varphi) = \partial_\omega^{-1}\Big(\Omega\big(DZ(\varphi)\underline{e}_k, Di(\varphi)\underline{e}_j\big) + \Omega\big(Di(\varphi)\underline{e}_k, DZ(\varphi)\underline{e}_j\big)\Big) . \qquad (58)$$

Proof. We use Cartan's formula $L_\omega(i^*\Omega) = d\big((i^*\Omega)(\omega, \cdot)\big) + \big(d(i^*\Omega)\big)(\omega, \cdot)$. Since $d(i^*\Omega) = i^* d\Omega = 0$ by (7) we get

$$L_\omega(i^*\Omega) = d\big((i^*\Omega)(\omega, \cdot)\big) . \qquad (59)$$

Now we compute, for $\hat\psi \in \mathbb{R}^\nu$ (denoting the vector field $(0, \mu, 0) = \mu \cdot \frac{\partial}{\partial I}$)

$$(i^*\Omega)(\omega, \hat\psi) = \Omega(Di(\varphi)\omega, Di(\varphi)\hat\psi) = \Omega(X_H(i(\varphi)) + \mu \cdot \frac{\partial}{\partial I} + Z(\varphi), Di(\varphi)\hat\psi)$$
$$= -dH(i(\varphi))[Di(\varphi)\hat\psi] + \mu \cdot D\theta_0(\varphi)[\hat\psi] + \Omega(Z(\varphi), Di(\varphi)\hat\psi) .$$

We obtain

$$(i^*\Omega)(\omega, \cdot) = \textstyle\sum_{j=1}^{\nu} b_j(\varphi) d\varphi_j$$

$$b_j(\varphi) = (i^*\Omega)(\omega, \underline{e}_j) = -\frac{\partial (H \circ i)}{\partial \varphi_j}(\varphi) + \mu \cdot \frac{\partial \theta_0}{\partial \varphi_j}(\varphi) + \Omega(Z(\varphi), Di(\varphi)\underline{e}_j) .$$

Hence, by (59), the Lie derivative

$$L_\omega(i^*\Omega) = \sum_{k<j} B_{kj}(\varphi) d\varphi_k \wedge d\varphi_j \tag{60}$$

with

$$\begin{aligned} B_{kj}(\varphi) &= \frac{\partial b_j}{\partial \varphi_k}(\varphi) - \frac{\partial b_k}{\partial \varphi_j}(\varphi) \\ &= \frac{\partial}{\partial \varphi_k}(\Omega(Z(\varphi), Di(\varphi)\underline{e}_j)) - \frac{\partial}{\partial \varphi_j}(\Omega(Z(\varphi), Di(\varphi)\underline{e}_k)) \\ &= \Omega(DZ(\varphi)\underline{e}_k, Di(\varphi)\underline{e}_j) + \Omega(Di(\varphi)\underline{e}_k, DZ(\varphi)\underline{e}_j) \,. \end{aligned} \tag{61}$$

Recalling (15) and (56) we have, $\forall \varphi \in \mathbb{T}^\nu$,

$$(\psi^t_\omega)^*(i^*\Omega)(\varphi) = i^*\Omega(\varphi + \omega t) = \sum_{k<j} A_{kj}(\varphi + \omega t) d\varphi_k \wedge d\varphi_j \,.$$

Hence the Lie derivative

$$L_\omega(i^*\Omega)(\varphi) = \sum_{k<j} (\omega \cdot \partial_\varphi) A_{kj}(\varphi) d\varphi_k \wedge d\varphi_j \,. \tag{62}$$

Comparing (60)-(61) and (62) we deduce (57). □

The previous lemma provides another proof of Lemma 1. For an invariant torus embedding $i(\varphi)$ the "error function" $Z(\varphi) = 0$ (see (39)) and so each $A_{kj} = 0$. If ω is Diophantine (see (18)) then, by (58) the following size estimate holds

$$A_{kj} = O(Z\gamma^{-1}) \,. \tag{63}$$

This estimate can be made quantitative once the norms are specified. For example, in scales of analytic functions as (46), it gives $\|A_{kj}\|_{\sigma'} \leq \gamma^{-1}(\sigma - \sigma')^{-(\tau+1)}\|Z\|_\sigma$, for all $\sigma' < \sigma$. In the Sobolev spaces (49) it implies $\|A_{kj}\|_s \leq \gamma^{-1}\|Z\|_{s+\tau+1}$. Since in the sequel of this note we will only focus on the algebraic aspect of the proof, we shall write only formal estimates as (63). We refer to [6], [3] for the analytic quantitative estimates.

We now prove that near an approximate isotropic torus there is an isotropic torus.

Lemma 6. (Isotropic torus) *The torus embedding* $i_\delta(\varphi) = (\theta_0(\varphi), I_\delta(\varphi), z_0(\varphi))$ *defined by*

$$I_\delta(\varphi) = I_0(\varphi) - [D\theta_0(\varphi)]^{-T}\rho(\varphi) \,, \quad \rho_j := \Delta^{-1}\Big(\sum_{k=1}^{\nu} \partial_{\varphi_j} A_{kj}(\varphi)\Big) \tag{64}$$

is isotropic. Thus $I_\delta - I_0 = O(\gamma^{-1}Z)$.

Proof. By Corollary 2 the 1-form $i^*\lambda - \rho$ is closed with ρ_j defined in (64), see also (53), (54). Actually $i^*\lambda - \rho = i^*_\delta\lambda$ is the pullback of the 1-form λ under the modified torus embedding i_δ defined in (64), see (55). Thus the torus $i_\delta(\mathbb{T}^\nu)$ is isotropic. □

Let

$$Z_\delta(\varphi) := \mathscr{F}(\varepsilon, i_\delta, \alpha, \mu) = \partial_\omega i_\delta(\varphi) - X_{H_\mu(\alpha,)}(i_\delta(\varphi)) \tag{65}$$

be the error function of the isotropic torus embedding i_δ. We now show that the isotropic torus embedding i_δ is a good approximate solution as i. This is needed for proving the convergence of the iterative Nash-Moser scheme under the minimal smallness condition $\varepsilon\gamma^{-1} \ll 1$, see (42).

Lemma 7. $Z_\delta = O(Z)$.

Proof. Let $Z_\delta(\varphi) = (Z_{1,\delta}, Z_{2,\delta}, Z_{3,\delta})(\varphi)$. Since the difference between the torus embeddings i_δ and i affects only the I-component (Lemma 6), and the normal form Hamiltonian vector field $X_{\mathscr{N}}$ is independent of I (see (35)), the components $Z_{1,\delta}, Z_{3,\delta}$ differ from Z_1, Z_3 for $O(\varepsilon|I_\delta - I_0|) = O(Z\gamma^{-1}\varepsilon) = O(Z)$. Moreover

$$Z_{2,\delta} - Z_2 = \partial_\omega(I_\delta - I_0) + \varepsilon\big(\partial_\theta P(i_\delta) - \partial_\theta P(i)\big) = -\partial_\omega v + O(\varepsilon Z\gamma^{-1})$$

where $v(\varphi) := [D\theta_0(\varphi)]^{-T}\rho(\varphi)$. We claim that $\partial_\omega v = O(Z)$ whence the lemma follows. We have $\partial_\omega v = (\partial_\omega[D\theta_0(\varphi)]^{-T})\rho + [D\theta_0(\varphi)]^{-T}\partial_\omega\rho$. The second term $\big[D\theta_0(\varphi)\big]^{-T}\partial_\omega\rho = O(Z)$ because (see (64)) each $\partial_\omega\rho_j = \Delta^{-1}\sum_{k=1}^{\nu}\partial_{\varphi_j}\partial_\omega A_{kj} = O(Z)$ by (57). We now prove that also the first term $(\partial_\omega[D\theta_0(\varphi)]^{-T})\rho = O(Z)$. Since $\rho = O(Z\gamma^{-1})$ (see (63), (64)) it is sufficient to prove that

$$\partial_\omega[D\theta_0(\varphi)]^{-T} = -\big[D\theta_0(\varphi)\big]^{-T}\big(\partial_\omega[D\theta_0(\varphi)]^T\big)\big[D\theta_0(\varphi)\big]^{-T} = O(\varepsilon)\,.$$

Differentiating in φ the first component $\partial_\omega\theta_0(\varphi) = \alpha + \varepsilon(\partial_I P)(\alpha, i(\varphi)) + Z_1(\varphi)$ of (36), we deduce

$$\begin{aligned}\partial_\omega D\theta_0(\varphi) &= \varepsilon D_i(\partial_I P)(i(\varphi))Di(\varphi) + DZ_1(\varphi)\\ &= \varepsilon(D_\theta(\partial_I P)D\theta_0 + D_I(\partial_I P)DI_0 + D_z(\partial_I P)Dz_0)(\varphi) + DZ_1(\varphi)\end{aligned}$$

and so its transposed $\partial_\omega\big[D\theta_0(\varphi)\big]^T = O(\varepsilon + Z) = O(\varepsilon)$. □

In analogy with section 2 we now introduce a symplectic set of coordinates (ψ, y, w) near the isotropic torus $\mathscr{T}_\delta := i_\delta(\mathbb{T}^\nu)$ via the symplectic diffeomorphism

$$\begin{pmatrix}\theta\\ I\\ z\end{pmatrix} = G_\delta\begin{pmatrix}\psi\\ y\\ w\end{pmatrix} := \begin{pmatrix}\theta_0(\psi)\\ I_\delta(\psi) + [D\theta_0(\psi)]^{-T}y - \big[D\tilde{z}_0(\theta_0(\psi))\big]^T\bar{J}w\\ z_0(\psi) + w\end{pmatrix} \tag{66}$$

where $\tilde{z}_0 := z_0 \circ \theta_0^{-1}$. The map G_δ is symplectic by Lemma 2 because i_δ is isotropic (Lemma 6). In the new coordinates (ψ, y, w) the isotropic torus embedding i_δ is trivial, namely $i_\delta(\psi) = G_\delta(\psi, 0, 0)$.

Under the symplectic change of variable (66), the Hamiltonian vector field X_{H_μ} changes into

$$X_{K_\mu} = G_\delta^* X_{H_\mu} = (DG_\delta)^{-1}X_{H_\mu} \circ G_\delta \tag{67}$$

where (recall (38))

$$K_\mu := H_\mu \circ G_\delta = K + \theta_0(\psi)\cdot\mu\,, \qquad K := H \circ G_\delta\,. \tag{68}$$

In the above formula, θ_0 is the lift to $\mathbb{R}^\nu$ of the first componant of i_δ (see Lemma 6). The Taylor expansion of the new Hamiltonian $K_\mu : \mathbb{R}^\nu \times \mathbb{R}^\nu \times E \to \mathbb{R}$ at the trivial torus $(\psi, 0, 0)$ is

$$\begin{aligned} K_\mu = {} & \theta_0(\psi) \cdot \mu + K_{00}(\alpha, \psi) + K_{10}(\alpha, \psi) \cdot y + \langle K_{01}(\alpha, \psi), w \rangle \\ & + \frac{1}{2} K_{20}(\alpha, \psi) y \cdot y + \langle K_{11}(\alpha, \psi) y, w \rangle + \frac{1}{2} \langle K_{02}(\alpha, \psi) w, w \rangle + K_{\geq 3}(\alpha, \psi, y, w) \end{aligned} \tag{69}$$

where $K_{\geq 3}$ collects all the terms at least cubic in the variables (y, w). The Taylor coefficients of K (in the sequel we may omit to write their dependence on α) are $K_{00}(\psi) \in \mathbb{R}$, $K_{10}(\psi) \in \mathbb{R}^\nu$, $K_{01}(\psi) \in E$, $K_{20}(\psi) \in \mathrm{Mat}(\nu \times \nu)$ is a real symmetric matrix, $K_{02}(\psi)$ is a self-adjoint operator of E and $K_{11}(\psi) \in \mathscr{L}(\mathbb{R}^\nu, E)$.

The Hamiltonian system associated to K_μ then writes

$$\begin{cases} \dot\psi = K_{10}(\alpha, \psi) + K_{20}(\alpha, \psi) y + K_{11}^T(\alpha, \psi) w + \partial_y K_{\geq 3}(\psi, y, w) \\ \dot y = -[D\theta_0(\psi)]^T \mu - \partial_\psi K_{00}(\alpha, \psi) - [D_\psi K_{10}(\alpha, \psi)]^T y - [D_\psi K_{01}(\alpha, \psi)]^T w \\ \qquad - \partial_\psi \big(\tfrac{1}{2} K_{20}(\alpha, \psi) y \cdot y + \langle K_{11}(\alpha, \psi) y, w \rangle + \tfrac{1}{2} \langle K_{02}(\alpha, \psi) w, w \rangle + K_{\geq 3}(\psi, y, w) \big) \\ \dot w = J \big(K_{01}(\alpha, \psi) + K_{11}(\alpha, \psi) y + K_{02}(\alpha, \psi) w + \nabla_w K_{\geq 3}(\psi, y, w) \big). \end{cases} \tag{70}$$

As seen in section 2, if i_δ were an invariant torus embedding, the coefficient $K_{00}(\psi) = \text{const}$, $K_{10}(\psi) = \omega$ and $K_{01}(\psi) = 0$. Moreover also $\mu = 0$ by Lemma 3. We now express these coefficients in terms of the error function Z_δ of i_δ defined in (65) (equivalently Z, by Lemma 7).

Lemma 8. *The vector field*

$$\begin{aligned} X_{K_\mu}(\psi, 0, 0) & \stackrel{(70)}{=} \begin{pmatrix} K_{10}(\alpha, \psi) \\ -[D\theta_0(\psi)]^T \mu - \partial_\psi K_{00}(\alpha, \psi) \\ J K_{01}(\alpha, \psi) \end{pmatrix} \\ & = \begin{pmatrix} \omega \\ 0 \\ 0 \end{pmatrix} - \big(DG_\delta(\psi, 0, 0) \big)^{-1} Z_\delta(\psi). \end{aligned} \tag{71}$$

Proof. By (67) and $i_\delta(\psi) = G_\delta(\psi, 0, 0)$, we have

$$X_{K_\mu}(\psi, 0, 0) = DG_\delta(\psi, 0, 0)^{-1} X_{H_\mu}(i_\delta(\psi)) = DG_\delta(\psi, 0, 0)^{-1} \big(\partial_\omega i_\delta(\psi) - Z_\delta(\psi) \big)$$

and (71) follows because $DG_\delta(\psi, 0, 0)^{-1} D i_\delta(\psi)[\omega] = (\omega, 0, 0)$. □

We now write the coefficient $K_{10}(\alpha, \psi)$ which describes in (69) and (70) how the tangential frequencies vary with respect to α, and the coefficients $K_{11}(\alpha, \psi)$, $K_{20}(\alpha, \psi)$ which are $O(\varepsilon)$.

Lemma 9. *The coefficients*

$$K_{10}(\alpha,\psi) = [D\theta_0(\psi)]^{-1}\alpha + \varepsilon[D\theta_0(\psi)]^{-1}(\partial_I P)(\varepsilon,\alpha,i_\delta(\psi))$$
$$K_{11}(\alpha,\psi) = \varepsilon D_I \nabla_z P(i_\delta(\psi))[D\theta_0(\psi)]^{-T} + \varepsilon \bar{J}(D\tilde{z}_0)(\theta_0(\psi))(D_I^2 P)(i_\delta(\psi))[D\theta_0(\psi)]^{-T}$$
$$K_{20}(\alpha,\psi) = \varepsilon[D\theta_0(\psi)]^{-1}(D_I^2 P)(i_\delta(\psi)))[D\theta_0(\psi)]^{-T}$$

Proof. Differentiating $K = H \circ G_\delta$ we get $K_{10}(\psi) = [D\theta_0(\psi)]^{-1}(\partial_I H)(i_\delta(\psi))$ and the lemma follows by (32), (33). Similarly we deduce that

$$K_{11}(\psi) = D_I \nabla_z H(i_\delta(\psi))[D\theta_0(\psi)]^{-T} + \bar{J}(D_\theta \tilde{z}_0)(\theta_0(\psi))(D_I^2 H)(i_\delta(\psi))[D\theta_0(\psi)]^{-T}$$
$$K_{20}(\psi) = [D\theta_0(\psi)]^{-1}(D_I^2 H)(i_\delta(\psi)))[D\theta_0(\psi)]^{-T}$$

and by (32)-(33) the lemma follows. □

Under the linear change of variable (obtained linearizing (66) at $(\psi,y,w) = (\varphi,0,0)$)

$$\begin{pmatrix} \hat{\theta} \\ \hat{I} \\ \hat{z} \end{pmatrix} = DG_\delta(\varphi,0,0)\begin{pmatrix} \hat{\psi} \\ \hat{y} \\ \hat{w} \end{pmatrix} \tag{72}$$

the linearized operator $d_{i,\alpha,\mu}\mathscr{F}(\varepsilon,i_\delta,\alpha_0,\mu_0)$ is transformed approximately (see (88) for the precise expression of the error) into the one obtained when we linearize the Hamiltonian system (70) at $(\psi,y,w) = (\omega t,0,0)$ and differentiating also in α, μ at α_0,μ_0, namely

$$L(\hat{\psi},\hat{y},\hat{w},\hat{\alpha},\hat{\mu}) := \tag{73}$$
$$\begin{pmatrix} \partial_\omega \hat{\psi} - D_\alpha K_{10}(\alpha,\varphi)[\hat{\alpha}] - D_\psi K_{10}(\alpha,\varphi)[\hat{\psi}] - K_{20}(\alpha,\varphi)\hat{y} - K_{11}^T(\alpha,\varphi)\hat{w} \\ \partial_\omega \hat{y} + [D\theta_0(\varphi)]^T \hat{\mu} + [D^2\theta_0(\varphi)\hat{\psi}]^T[\mu_0] + \partial_\psi\big(D_\alpha K_{00}(\alpha,\varphi)[\hat{\alpha}]\big) \\ + D_{\psi\psi}K_{00}(\alpha,\varphi)\hat{\psi} + [D_\psi K_{10}(\alpha,\varphi)]^T\hat{y} + [D_\psi K_{01}(\alpha,\varphi)]^T\hat{w} \\ \partial_\omega \hat{w} - J\big(D_\alpha K_{01}(\alpha,\varphi)[\hat{\alpha}] + D_\psi K_{01}(\alpha,\varphi)[\hat{\psi}] + K_{11}(\alpha,\varphi)\hat{y} + K_{02}(\alpha,\varphi)\hat{w}\big) \end{pmatrix}.$$

For the convergence of the Nash Moser scheme, it is sufficient to invert the operator L defined in (73) only approximately, namely, in view of Lemmata 8 and 3, solve only the linear system

$$\mathbb{D}(\hat{\psi},\hat{y},\hat{w},\hat{\alpha},\hat{\mu}) := g(\varphi) = \begin{pmatrix} g_1(\varphi) \\ g_2(\varphi) \\ g_3(\varphi) \end{pmatrix} \tag{74}$$

with the "simpler" operator

$$\mathbb{D}(\hat{\psi},\hat{y},\hat{w},\hat{\alpha},\hat{\mu}) := \begin{pmatrix} \partial_\omega \hat{\psi} - D_\alpha K_{10}(\alpha,\varphi)[\hat{\alpha}] - K_{20}(\alpha,\varphi)\hat{y} - K_{11}^T(\alpha,\varphi)\hat{w} \\ \partial_\omega \hat{y} + [D\theta_0(\varphi)]^T\hat{\mu} + \partial_\psi\big(D_\alpha K_{00}(\alpha,\varphi)[\hat{\alpha}]\big) \\ \partial_\omega \hat{w} - J\big(D_\alpha K_{01}(\alpha,\varphi)[\hat{\alpha}] + K_{11}(\alpha,\varphi)\hat{y} + K_{02}(\alpha,\varphi)\hat{w}\big) \end{pmatrix}. \tag{75}$$

$\mathbb{D}$ is obtained from L in (73) neglecting the terms which are zero at an exact solution (α_0, μ_0, i_0) (with $\mu_0 = 0$) . System (74) may be solved in a triangular way. We first solve the second equation

$$\partial_\omega \hat{y} = -\partial_\psi \big(D_\alpha K_{00}(\alpha_0, \varphi)[\hat{\alpha}] \big) - [D\theta_0(\varphi)]^T \hat{\mu} + g_2 \,. \tag{76}$$

We choose $\hat{\mu}$ such that the φ-average of the right hand side

$$\langle\!\langle -\partial_\psi (D_\alpha K_{00}(\alpha_0, \varphi)[\hat{\alpha}]) - [D\theta_0(\varphi)]^T \hat{\mu} + g_2 \rangle\!\rangle = 0 \,.$$

Note that the average of the total derivative $\partial_\psi \big(\partial_\alpha K_{00}(\alpha_0, \varphi)[\hat{\alpha}] \big)$ is zero, and the averaged matrix $\langle\!\langle [D\theta_0(\varphi)]^T \rangle\!\rangle = Id + \langle\!\langle [D\Theta_0(\varphi)]^T \rangle\!\rangle = Id$ because $\Theta_0(\varphi)$ is periodic in φ. Hence we find

$$\hat{\mu} := \langle\!\langle g_2 \rangle\!\rangle \,, \tag{77}$$

and, by (76), we define

$$\hat{y} = -\partial_\omega^{-1} \big(\partial_\psi \big(D_\alpha K_{00}(\alpha_0, \varphi)[\hat{\alpha}] \big) + [D\theta_0(\varphi)]^T \langle\!\langle g_2 \rangle\!\rangle - g_2 \big) + c_1 \tag{78}$$

for some $c_1 \in \mathbb{R}^\nu$.

Next we consider the third equation

$$\partial_\omega \hat{w} - JK_{02}(\alpha_0, \varphi)\hat{w} = J(D_\alpha K_{01}(\alpha_0, \varphi)[\hat{\alpha}]) + JK_{11}(\alpha_0, \varphi)\hat{y} + g_3 \,. \tag{79}$$

Remark that (79) is a linear quasi-periodically forced PDE with a self adjoint operator K_{02} which is a perturbation of the normal form operator $N(\alpha, \theta)$ in (33). The solvability of (79) has to be checked case by case for a given PDE. We can say something when $N(\alpha, \theta) = N(\alpha)$ does not depend on θ, see remark 8. What is relevant is the nature of spectrum of the Hamiltonian vector field $JN(\alpha)$: if their eigenvalues are real or purely imaginary, simple or multiple, their asymptotic expansions, etc... If, for example, $JN(\alpha)$ has real spectrum, bounded away from zero, then also the linear operator

$$\partial_\omega - JK_{02}(\alpha_0, \varphi) \tag{80}$$

is invertible with good bounds for the inverse. This is the case for the continuation of isotropic tori of hyperbolic type, as considered in [29] and in [18]. If $JN(\alpha)$ has purely imaginary discrete spectrum (elliptic tori) the main work is to prove that for "most" frequencies ω the quasi-periodic linear operator (80) is invertible, and its inverse satisfies good estimates in high norms. This may be hard work, see the forced PDEs [4], [5], [2], [8]. However, if it is solved, it is possible to define the solution $\hat{w}$ of the linear equation (79) by

$$\hat{w} := \big(\partial_\omega - JK_{02}(\alpha_0, \varphi) \big)^{-1} \Big(J(D_\alpha K_{01}(\alpha_0, \varphi)[\hat{\alpha}]) + JK_{11}(\alpha_0, \varphi)\hat{y} + g_3 \Big) \,. \tag{81}$$

Finally we solve also the first equation of (75), namely

$$\partial_\omega \hat{\psi} = D_\alpha K_{10}(\alpha_0, \varphi)[\hat{\alpha}] + K_{20}(\alpha_0, \varphi)\hat{y} + K_{11}^T(\alpha_0, \varphi)\hat{w} + g_1 \,. \tag{82}$$

We look for $\hat{\alpha}$ such that the right hand side in (82) has zero average, namely

$$\langle\!\langle D_\alpha K_{10}(\alpha_0,\varphi)\rangle\!\rangle[\hat{\alpha}] + \langle\!\langle K_{20}(\alpha_0,\varphi)\hat{y}\rangle\!\rangle + \langle\!\langle K_{11}^T(\alpha_0,\varphi)\hat{w}\rangle\!\rangle + \langle\!\langle g_1\rangle\!\rangle = 0 . \tag{83}$$

By Lemma 9, $D_\alpha K_{10}(\alpha_0,\varphi) = D\theta_0(\varphi)^{-1} + O(\varepsilon)$, hence

$$\langle\!\langle D_\alpha K_{10}(\alpha_0,\varphi)\rangle\!\rangle = \langle\!\langle D\theta_0(\varphi)^{-1}\rangle\!\rangle + O(\varepsilon) = Id + O(\varepsilon\gamma^{-1})$$

because $D\theta_0 = Id + O(\varepsilon\gamma^{-1})$. Note that $\hat{y}$ and $\hat{w}$ depend on $\hat{\alpha}$ (see (78), (81)) but, since K_{20}, K_{11}^T are $O(\varepsilon)$ by Lemma 9, the equation (83) takes the form

$$(Id + R_\varepsilon)[\hat{\alpha}] = \Gamma \qquad \text{with} \quad R_\varepsilon \to 0 \;\text{ as }\; \varepsilon \to 0$$

For ε small enough, $Id + R_\varepsilon$ is invertible and (83) has a unique solution $\hat{\alpha}$.

Remark 11. Above we suppose that, for example, the size of the inverse $(\partial_\omega - JK_{02}(\alpha_0,\varphi))^{-1} = O(\gamma^{-1})$ and $\varepsilon\gamma^{-1}$ is small. Variants are possible.

Next, from (82) we find

$$\hat{\psi} = \partial_\omega^{-1}\big(D_\alpha K_{10}(\alpha_0,\varphi)[\hat{\alpha}] + K_{20}(\alpha_0,\varphi)\hat{y} + K_{11}^T(\alpha_0,\varphi)\hat{w} + g_1\big) + c_2 \tag{84}$$

for some constant $c_2 \in \mathbb{R}^\nu$.

Remark 12. The constants $c_1, c_2 \in \mathbb{R}^\nu$ in the definition of $\hat{y}$ in (78) and $\hat{\psi}$ in (84) are free (we can set for instance $c_1 = c_2 = 0$). Thus the operator $d_{i,\alpha,\mu}\mathscr{F}(\varepsilon,i,\alpha,\mu)$ has only a right inverse. About c_1, the presence of the parameter α gives the freedom to impose an additional condition for I_0 (for instance $I_0(0) = 0$, or the mean value of I_0 vanishes). The presence of the constant c_2 is connected to the fact that if $i(\varphi)$ is a solution then all the translates $i(\varphi + c)$ are solutions too. It is usual to impose that the mean value of $\theta(\varphi) - \varphi$ is 0.

In conclusion, the solution of the linear system (74) is

$$\mathbb{D}^{-1}g := \big(\hat{\psi}, \hat{y}, \hat{w}, \hat{\alpha}, \hat{\mu}\big)$$

defined in (77), (78), (81), (83), (84). Recalling (72) we finally define the linear operator

$$T_{i,\alpha,\mu} := D\widetilde{G}_\delta(\varphi,0,0) \circ \mathbb{D}^{-1} \circ DG_\delta(\varphi,0,0)^{-1} , \tag{85}$$

where we include in $\widetilde{G}_\delta$ also the parameters components, namely

$$\widetilde{G}_\delta(\psi,y,w,\alpha,\mu) := \big(G_\delta(\psi,y,w),\alpha,\mu\big) .$$

By construction, the operator $T_{i,\alpha,\mu}$ is an approximate right inverse of $d_{i,\alpha,\mu}\mathscr{F}$, because it has been obtained neglecting terms which vanish at an exact solution: we first substituted the approximate torus embedding i with the isotropic one i_δ (which coincide at a solution by Lemma 6) and then we neglected the terms K_{00}, K_{10}, K_{01}, μ_0 which are naught at a solution (Lemmata 8, 3). Let us give a more formal proof.

Lemma 10. *The operator $T_{i,\alpha,\mu}$ is an approximate right inverse of $d_{i,\alpha,\mu}\mathscr{F}(\varepsilon,i,\alpha,\mu)$.*

Proof. By (35), since $X_{\mathscr{N}}$ does not depend on I, and i_δ differs from i only for the I component, we have

$$\begin{aligned} d_{i,\alpha,\mu}\mathscr{F}(i,\alpha_0) - d_{i,\alpha,\mu}\mathscr{F}(i_\delta,\alpha_0) &= \varepsilon\big(d_{i,\alpha,\mu}X_P(i,\alpha_0) - d_{i,\alpha,\mu}X_P(i_\delta,\alpha_0)\big) \\ &= \varepsilon\int_0^1 \partial_I d_{i,\alpha,\mu}X_P(i_\delta + s(i-i_\delta),\alpha_0)[I_0 - I_\delta]ds \\ &=: \mathscr{E}_0 \end{aligned} \tag{86}$$

which is $O(Z)$ by Lemma 6 and (42).

We denote by $\mathtt{u} := (\psi,y,w)$ the symplectic coordinates induced by G_δ in (66). Under the symplectic map G_δ, the nonlinear operator $\mathscr{F}$ in (35) is transformed into

$$\mathscr{F}(G_\delta(\mathtt{u}(\varphi)),\alpha,\mu) = DG_\delta(\mathtt{u}(\varphi))\big(\partial_\omega \mathtt{u}(\varphi) - X_{K_\mu}(\mathtt{u}(\varphi),\alpha)\big) \tag{87}$$

where $K_\mu = H_\mu \circ G_\delta$, see (68). Differentiating (87) at the trivial torus embedding $\mathtt{u}_\delta(\varphi) := G_\delta^{-1}(i_\delta(\varphi)) = (\varphi,0,0)$ for the values of the parameters $(\alpha,\mu) = (\alpha_0,\mu_0)$, we get

$$\begin{aligned} d_{i,\alpha,\mu}\mathscr{F}(i_\delta,\alpha_0,\mu_0) &= DG_\delta(\mathtt{u}_\delta)\big(\partial_\omega - d_{\mathtt{u},\alpha,\mu}X_{K_\mu}(\mathtt{u}_\delta,\alpha_0,\mu_0)\big)D\widetilde{G}_\delta(\mathtt{u}_\delta)^{-1} + \mathscr{E}_1\,, \\ \mathscr{E}_1 &:= D^2G_\delta(\mathtt{u}_\delta)\big[DG_\delta(\mathtt{u}_\delta)^{-1}\mathscr{F}(i_\delta,\alpha_0,\mu_0), DG_\delta(\mathtt{u}_\delta)^{-1}\Pi[\cdot]\big]\,, \end{aligned} \tag{88}$$

where Π is the projection $(\hat\imath,\hat\alpha,\hat\mu) \mapsto \hat\imath$. In expanded form $d_{\mathtt{u},\alpha,\mu}X_{K_\mu}(\mathtt{u}_\delta,\alpha_0,\mu_0)$ is provided in (73). We split $\partial_\omega - d_{\mathtt{u},\alpha,\mu}X_{K_\mu}(\mathtt{u}_\delta,\alpha_0,\mu_0) = \mathbb{D} + R_Z$ where $\mathbb{D}$ is defined in (75) and R_Z is the part which vanishes in Z. By (86) and (88)

$$\begin{aligned} d_{i,\alpha,\mu}\mathscr{F}(i,\alpha) &= DG_\delta(\mathtt{u}_\delta)\circ\mathbb{D}\circ D\widetilde{G}_\delta(\mathtt{u}_\delta)^{-1} + \mathscr{E}_0 + \mathscr{E}_1 + \mathscr{E}_2 \\ \mathscr{E}_2 &:= DG_\delta(\mathtt{u}_\delta)\circ R_Z \circ D\widetilde{G}_\delta(\mathtt{u}_\delta)^{-1}\,. \end{aligned}$$

Applying T defined in (85) to the right, since $\mathbb{D}\circ\mathbb{D}^{-1} = Id$ we get

$$d_{i,\alpha,\mu}\mathscr{F}(i,\alpha_0,\mu_0)\circ T - Id = \mathscr{E}\circ T$$

where $\mathscr{E} := \mathscr{E}_0 + \mathscr{E}_1 + \mathscr{E}_2$ is $O(Z)$. □

Remark 13. In order to construct an approximate inverse for $d\mathscr{F}$, it is sufficient to have an approximate inverse of $\mathbb{D}$ in (74), i.e. we need in (81) only an approximate inverse for $\partial_\omega - JK_{02}(\alpha_0,\varphi)$.

The operator T usually satisfies estimates like (47)-(48) (in an analytic setting) or (50)-(51) (in a Sobolev scale) and the Nash-Moser iterative scheme with approximate right inverse converges.

References

1. Arcostanzo M., Arnaud M.C., Bolle P., Zavidovique M.: Tonelli Hamiltonians without conjugate points and C^0 integrability, preprint (2013).
2. Baldi P., Berti M., Montalto R.: KAM for quasi-linear and fully nonlinear forced perturbations of Airy equation, Math. Ann. **359**, no. 1-2, 471–536, (2014).
3. Baldi P., Berti M., Montalto R.: KAM for autonomous quasi-linear perturbations of KdV, preprint, (2014).
4. Berti P., Bolle P.: Sobolev quasi periodic solutions of multidimensional wave equations with a multiplicative potential, Nonlinearity, **25**, 2579–2613, (2012).
5. Berti M., Bolle P.: Quasi-periodic solutions with Sobolev regularity of NLS on $\mathbb{T}^d$ with a multiplicative potential, Eur. Jour. Math, **15**, 229–286, (2013).
6. Berti M., Bolle P.: Quasi-periodic solutions for autonomous NLW on $\mathbb{T}^d$ with a multiplicative potential, in preparation.
7. Berti M., Bolle P., Procesi M.: An abstract Nash-Moser theorem with parameters and applications to PDEs, Ann. I. H. Poincaré, **1** , 377–399, (2010).
8. Berti M., Corsi L., Procesi M.: An abstract Nash-Moser theorem and quasi-periodic solutions for NLW and NLS on compact Lie groups and homogeneous spaces, to appear on Comm. Math. Phys., 2014.
9. Bourgain J.: Construction of quasi-periodic solutions for Hamiltonian perturbations of linear equations and applications to nonlinear PDE, Internat. Math. Res. Notices, **11**, (1994).
10. Bourgain J.: Quasi-periodic solutions of Hamiltonian perturbations of $2D$ linear Schrödinger equations, Annals of Math. **148**, 363–439, (1998).
11. Bourgain J.: Green's function estimates for lattice Schrödinger operators and applications, Annals of Mathematics Studies **158**, Princeton University Press, Princeton, (2005).
12. Craig W.: Problèmes de petits diviseurs dans les équations aux dérivées partielles, Panoramas et Synthèses, **9**, Société Mathématique de France, Paris, (2000).
13. Craig W., Wayne C.E.: Newton's method and periodic solutions of nonlinear wave equation, Comm. Pure Appl. Math. **46**, 1409–1498, (1993).
14. De la Llave R., Gonzalez A., Jorba A., Villanueva J.: KAM theory without action-angle variables, Nonlinearity **18**, 855–895, (2005).
15. Eliasson, L. H.: Perturbations of stable invariant tori for Hamiltonian systems, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) 15 (1988), no. 1, 115147 (1989).
16. Eliasson L.H., Kuksin S.: KAM for non-linear Schrödinger equation, Annals of Math., 172, 371–435, (2010).
17. Fejoz J.: Démonstration du théorème d' Arnold sur la stabilité du système planétaire (d' après M. Herman), Ergod. Th. Dynam. Sys., **24**, 1–62, (2004).
18. Fontich E., De La Llave R., Sire Y.: A method for the study of whiskered quasi-periodic and almost-periodic solutions in finite and infinite dimensional Hamiltonian systems, Electronic Research Announcements in Math. Sciences, **16**, 9-22, S 1935-9179 AIMS, (2009).
19. Herman, M. R.: Inégalités "a priori" pour des tores lagrangiens invariants par des difféomorphismes symplectiques, Publ. Math., Inst. Hautes Étud. Sci. **70**, 47–101 (1989).
20. Kuksin S.: Hamiltonian perturbations of infinite-dimensional linear systems with imaginary spectrum, Funktsional. Anal. i Prilozhen. **21**, no. 3, 22–37, 95, (1987).
21. Kuksin S.: Analysis of Hamiltonian PDEs, Oxford University Press, Oxford, (2000).
22. Moser J.: A rapidly convergent iteration method and non-linear partial differential equations I-II, Ann. Scuola Norm. Sup. Pisa **3**, 20, 265–315, 499–535, (1966).
23. Moser J.: Convergent series expansions for quasi-periodic motions, Math. Ann. **169**, 136–176, (1967).
24. Moser J., Zehnder E.: Notes on Dynamical Systems, Courant Lecture Notes, **12**, (2005).
25. Pöschel, J.: A KAM-theorem for some nonlinear partial differential equations, Ann. Scuola Norm. Sup. Pisa Cl. Sci. **4** 23 , no. 1, 119–148, (1996).
26. Procesi M., Xu X.,: Quasi-Töplitz Functions in KAM Theorem, to appear in SIAM J. Math Anal., 45-4 2148–2181, (2013).

27. Salamon D., Zehnder E.: KAM theory in configuration space, Comm. Math. Helv. **64**, 84–132, (1989).
28. Wayne E.: Periodic and quasi-periodic solutions of nonlinear wave equations via KAM theory, Comm. Math. Phys. 127, 479–528, (1990).
29. Zehnder E.: Generalized implicit function theorems with applications to some small divisors problems I-II, Comm. Pure Appl. Math., 28, 91–140, (1975); 29, 49–113, (1976).