KAM THEORY FOR EQUILIBRIUM STATES IN 1-D STATISTICAL MECHANICS MODELS

RAFAEL DE LA LLAVE

Abstract. We extend the Lagrangian proof of KAM for twist mappings [SZ89, LM01] to show persistence of quasi-periodic equilibrium solutions in statistical mechanics models. The interactions in the models considered here do not need to be of finite range but they have to decrease sufficiently with the distance.

When the interactions are range R, the models admit the dynamical interpretation of recurrences in $(\mathbb{R})^{2R}$. Note that the small perturbations in the Lagrangian are singular from the dynamical systems point of view since they may increase the dimension of phase space.

We show that in these models, given an approximate solution of the equilibrium equation with one Diophantine frequency, which is not too degenerate, there is a true solution nearby. As a consequence, we deduce that quasi-periodic solutions of the equilibrium equation with one Diophantine frequency persist under small modifications of the model.

The main result can also be used to validate numerical calculations or perturbative expansions.

We also show that Lindstedt series can be computed to all orders in these models.

1. Introduction

It is well known that equilibrium configurations in one-dimensional ferromagnetic models are equivalent to orbits of twist mappings. For example, the equilibrium configurations Frenkel-Kontorova model are equivalent to orbits of the standard map. See [MF94b] as well Section 1.1 and the references there.

In spite of the fact the two problems (orbits of twist maps and equilibrium states of ferromagnetic models), the intuition and the possible natural generalizations are very different for twist mappings and for equilibrium configurations. In this paper, we will consider some models that are very natural from the point of view of statistical mechanics but for which there is no easy equivalent in dynamical systems.

The main result of this paper is Theorem 1, which extends the Lagrangian proof of the twist mapping theorem in [SZ89, LM01] to some models in statistical mechanics in which the interactions are not nearest neighbor and indeed need not be finite range. From the point of view of statistical mechanics, this generalization is natural. In the statistical mechanics motivation of the Frenkel-Kontorova models they are used to describe the deposition of a material over a substratum of a different material. Frenkel-Kontorova models are also natural as description of spin systems over a one-dimensional crystal. In these motivations it is natural to consider extensions of the classical Frenkel-Kontorova models in which all the sites interact, even if the interaction decreases with the distance.

Adding a small interaction term for next nearest neighbor interactions, even if it seems like a small perturbation from the Lagrangian picture is a very singular perturbation from the point of view of mechanics. Adding a next-nearest neighbor to the Frenkel-Kontorova model makes the phase space 4 dimensional, rather than 2 dimensional and some of the terms are very large. We will give more details of this in Section 1.1.4.

If one is interested in ground states which are quasi-periodic orbits of the model, it is natural to expect that they persist under small (Lagrangian modifications of the model). This is accomplished by KAM theory. Nevertheless, as we have pointed out, from the dynamical systems point of view, the perturbations are not small. Adding a next nearest neighbor interaction, – no matter how small – even changes the dimension of the phase space.

In this paper, we will show that quasi-periodic orbits persist under perturbations of the models, provided that some non-degeneracy conditions are met. In the case that the interactions are nearest neighbor, this result is the standard twist theorem. When we consider infinite range interactions the equilibrium configurations cannot be considered as orbits of evolution problems, in particular, they cannot be formulated as orbits of Hamiltonian systems and the proofs of KAM theorems based in Hamiltonian formalism do not apply. Hence, we will use the Lagrangian formulation of the KAM theorem. Other KAM theorems which have a Lagrangian formulation but not a Hamiltonian one occur in [Mos88]. In contrast, there are KAM proofs in situations where there is a Hamiltonian formalism but not a Lagrangian one [JdlLZ99, DdlL00, GEHdlL05].

When the variables at each site in the statistical model are one dimensional, the interaction is of range R and there are some convexity properties in the model – which amount to ferromagnetism or twist – the equilibrium configurations (critical points of the action) correspond to orbits of "monotone recurrences" in $\mathbb{R}^{2R}$ and were considered in [Ang90]. These mappings satisfy very remarkable dynamical properties including shadowing theorems for many orbits. In this paper, we will show the persistence of quasi-periodic orbits with one independent frequency under some assumptions that include ferromagnetism as a particular case.

From the dynamical systems point of view, the fact that we get persistence of orbits with one independent frequency is somewhat surprising since one expects that the number of frequencies needed in the KAM theorem is the same as the number of degrees of freedom (see [Mos73]). Nevertheless, for a long time variational and topological methods [Ang90, KdlLR97, CdlL98, dlL00] had established existence of quasi-periodic orbits with only one frequency for models similar to the ones we consider. In particular [CdlL98] emphasized the case of interactions with infinite range but with a one-dimensional frequency.

The conditions of the present result are, as usual with KAM methods, more restrictive than the variational results in terms of smallness of certain terms, regularity requirements and the such, but less restrictive in terms of convexity properties and on the dimension of the variables. Of course, the KAM methods give more information than variational methods since they conclude that the orbits lie on smooth curves, whereas the variational methods only allow to conclude that they lie on perfect sets or just give a description as measures (we note that, even if it is customary to refer to the solutions produced by Aubry-Mather theory as quasi-periodic, in the more precise nomenclature of [SY98], they should be called *"almost automorphic"*').

We have formulated the KAM result following [SZ89, Zeh76, Mos66b, Mos66a] without reference to an integrable system. (For a comparison of different formulations of KAM theorem, see [dlL01].) Following the papers above, we show that, if we are given a function which satisfies the equilibrium equation with a good enough accuracy, and which satisfies some non-degeneracy condition, then, there is an exact solution of the equilibrium equation. The distance between the approximate solution and the exact one is bounded by the error with which the approximate solution solves the equation. Moreover, one can conclude some uniqueness of the solutions in a neighborhood of the approximate one.

This is what in numerical analysis is called *"a posteriori"* bounds. Since numerical or asymptotic procedures produce objects which solve the functional equation very approximately, an "a posteriori" KAM result serves as validation of the approximate (numerical or asymptotic) procedures, which produce objects which solve the equation up to an small error. Note that given an a-posteriori theorem, one does not need to validate the algorithm to produce the solution. One just needs to verify that the equilibrium equation is satisfied with an small error. In the case of numerical computations, reliable upper bounds on the error of the solution are computable (e.g. using interval arithmetic) these type of a-posteriori theorems are the basis of computer assisted proofs. See [Ran87, dlLR91, dlLR90, JdlLZ99, CC95, CC97] for computer assisted proofs in Hamiltonian systems and [CdlL05] for computer explorations in the systems considered in this paper. From the point of view of numerical explorations it is interesting to remark that the method presented here provides with an algorithm in which the Newton step has a cost of only $N \log(N)$ in the number of discretization points of the torus (this is in common with the algorithms in [SZ89, dlLGJV05]). In contrast, most of the implementations of KAM methods use algorithms of order N^3 operations per step. We will present some remarks on numerical algorithms for the quasi-periodic orbits in Section 7.1.

Relatedly, as already pointed out in [Mos66b, Mos66a, Zeh75], an *"a posteriori"* result for analytic functions with good enough quantitative bounds implies results for finite differentiable functions. We note that in the papers above, the finitely differentiable results include the assumption that the system is close to an integrable analytic one, but we have removed this assumption here. We hope to come back to this problem.

As a technical remark, we note that the proof of KAM theorem we present is not based on the transformation theory that appears in many proofs of KAM theorem, but rather is based on an iterative procedure in which corrections are added. The method is very similar to that of [Mos88, SZ89, LM01]. indeed, some of the identities we use will lead to efficient numerical implementations. We refer to [CdlL05] for a deeper discussion of numerical issues. See also Section 7.1.

Besides the KAM theory, we will also discuss some complementary results such as study the existence of Lindstedt series and their convergence.

1.1. **Models considered.** In Section 1.1.1, we discuss twist mappings of the annulus which is the best known and simplest model. For twist mappings, in the statistical mechanics language, the interactions are nearest neighbor and the variables at each site are one dimensional. In Section 1.1.3, we discuss statistical mechanics models of spin chains which include the possibility of long range interactions.

The models we consider are models in which the sites are arranged in a one-dimensional lattice $\mathbb{Z}$. That is, a configuration is a mapping from $\mathbb{Z} :\to \mathbb{R}$. From the point of view of statistical mechanics, it would have been natural to consider models defined in higher dimensional lattices or even on trees. These models on higher dimensional lattices do not admit interpretation as dynamical systems and they are not accessible to the methods in this paper. We hope to come back to this problem.

1.1.1. *Twist mappings of the annulus.* It is well known in Hamiltonian mechanics that orbits $\{x_n\}_{n\in\mathbb{Z}} \subseteq \mathbb{R}$ of a twist map of the annulus $\mathbb{A} \equiv \mathbb{T} \times \mathbb{R}$ can be identified with critical points of the functional given by the formal sum

$$\mathcal{S}(\{x_n\}) = \sum_{n\in\mathbb{Z}} S(x_n, x_{n+1}) \tag{1}$$

where S is the so-called generating function of the map.

The standard assumptions for the generating functions of the twist mappings of the annulus are:

$$\partial_x \partial_y S(x, y) \leq C < 0 \tag{2}$$

$$S(x + 1, y + 1) = S(x, y) \tag{3}$$

The condition (2) is called the twist condition in dynamical systems (it is called ferromagnetism in statistical mechanics) and (3) is implied by the assumption that the x_n can be interpreted as angles and that the variational principle has a physical meaning, The meaning of condition (3) is that, even if relative phases between two sites may be important for the variational principle, the variational principle of the model is unaffected by to addition of one unit to the states of all the sites. (i.e., if $\{x_n\}_{n\in\mathbb{Z}}$ is an orbit, so is $\{x_n + 1\}_{n\in\mathbb{Z}}$.) In some physical interpretations, when the x are angles, it would have been reasonable to assume $S(x+1, y) = S(x, y+1) = S(x, y)$ which clearly implies (3). For our purposes the assumption (3) is enough and we will not consider stronger assumptions even if they are natural for the physical models.

We recall that $\{x_n\}$ is an equilibrium configuration i.e. a critical point of (1) if and only if

$$\partial_{x_n} S(x_n, x_{n+1}) + \partial_{x_n} S(x_{n-1}, x_n) = 0 \qquad \forall n \in \mathbb{Z} \tag{4}$$

We will refer to the equations (4) as the Euler-Lagrange equations of (1). They are obtained by computing (formally since the sum (1) defining $\mathcal{S}$ is only a formal sum) $\partial_{x_n} \mathcal{S}(\{x_n\})$ by noting that x_n only appears in two terms in $\mathcal{S}(\{x_n\})$. Note that, even if (1) is just a formal sum, the equilibrium equation (4) is a well defined system of equations.

For our studies, the starting point will be the equations (4) and the relation with variational problems serves only motivation. Hence, we will not try to make more precise the well known connection between the variational principle and its critical point equations. Nevertheless, we recall that the Euler-Lagrange equations (4) are obtained by imposing that truncations of the sum (1) are stationary under modifications whose support is inside the domain of the truncation. On the other hand, we note that some of the identities that are crucial for our approach are directly related to the variational structure of the equilibrium equations. See Section 3.3 and the derivation of (58).

Setting

$$y_n = \partial_1 S(x_n, x_{n+1})$$

we have that the fact that x_n satisfies (4) is equivalent to saying $\{(x_n, y_n)\}_{n\in\mathbb{Z}}$ is an orbit of the mapping T defined by

$$(\tilde{x}, \tilde{y}) = T(x, y) \Leftrightarrow \begin{cases} y = \partial_x S(x, \tilde{x}) \\ \tilde{y} = -\partial_{\tilde{x}} S(x, \tilde{x}) \end{cases} \tag{5}$$

The reason why the equations in (5) define a map is that, because of assumption (2), given x, y we can use the first equation to determine a unique $\tilde{x}$, then, evaluate the second equation to compute $\tilde{y}$.

1.1.2. *Frenkel-Kontorova models.* The functionals (1) also admit motivations from solid state physics. For example, the case

$$S(x, y) = \frac{1}{2}(x - y - a)^2 + V(x) \tag{6}$$

with $V(x+1) = V(x)$ is commonly referred to as the Frenkel-Kontorova model which describes particles deposited in a periodic substratum. The interaction of the particles with the periodic substratum is modeled by the term $V(x)$ and the interaction with their nearest neighbors (modeled by the term $\frac{1}{2}(x - y - a)^2$).

Another possible interpretation of (6) – specifically for configurations for which $|x_n - x_{n+1}| \le 1$ — is interactions among spin chains. In this interpretation (2) is just that the interaction is ferromagnetic.

There is a wide variety of mathematical results for orbits of twist maps. Among recent surveys, we mention [LC00] for a survey emphasizing topological results and to [MF94a, Gol01] for surveys emphasizing variational results.

1.1.3. *Models from statistical mechanics with non-nearest neighbor interactions.* From the point of view of solid state physics it is natural to consider models analogue to (6) for which the interaction is not nearest neighbor. Nevertheless, it is natural to require invariance under translations. See [Rue69].

The invariance under translation implies that it suffices to consider models of the form:

$$\mathcal{S}(\{x_n\}_{n\in\mathbb{Z}}) = \sum_{L\in\mathbb{N}} \sum_{k\in\mathbb{Z}} H_L(x_k, \dots, x_{k+L}) \tag{7}$$

Remark 1. There is a small ambiguity in the definition of the models. For example, the model with $H_i = 0$, $i \neq 2$ and $H_2(x_0, x_1, x_2) = F(x_0, x_1) + F(x_1, x_2)$ is equivalent to the model with $H_i \neq 0$, $i \neq 1$ and $H_1(x_0, x_1) = F(x_0, x_1)$.

This ambiguity will not affect any of our reasonings. We will just state some assumptions on the interaction and draw conclusions from them. So that the results apply for one model provided that there is one of the equivalent interactions that satisfy the assumptions of the results here.

As a point of notation, we observe that it is natural to label the arguments of the interaction staring with zero. Hence, write $H_L(x_0, \dots, x_L)$. We denote $\partial_i H_L = \partial_{x_i} H_L$. Hence, the indices of derivatives run from 0 to L.

The analogue of (3) is

$$H_L(x_0, \dots, x_L) = H_L(x_0 + 1, \dots, x_L + 1) \tag{8}$$

This indicates that, even if the interactions may depend on the relative phases at different sites, they are independent of a global change of phase by an integer over all the sites.

The analogue of (2) that appears in the calculus of variations (e.g. in [MF94b, CdlL98]) is:

$$\begin{aligned} \partial_i \partial_j H_L &\leq 0 \qquad i \neq j \\ \partial_0 \partial_1 H_1 &\leq C < 0 \end{aligned} \tag{9}$$

which just requires that the two body interaction is strictly ferromagnetic while the many body interactions are not anti-ferromagnetic.

In this paper, we will require some weaker assumptions than the assumptions (9) in the calculus of variations. As usual in KAM theory, we will not need positivity assumptions but rather invertibility in a neighborhood of the approximate solution. We will assume that

$$|\det(\partial_0 \partial_1 H_1)| > C > 0 \tag{10}$$

and that the longer range interactions are small. See assumption H2 in Theorem 1 for a precise formulation of the assumption.

We will require that the interactions H_L decrease fast enough with L and that the nearest neighbor interactions dominate the rest of the interactions.

That is, we will require that

$$\sum_{L>2} \|H_L\|_q L^4 < \alpha \tag{11}$$

The precise definition of the norm used in (11) will be postponed till Section 2.3.2. We just anticipate that the norm is an analytic norm in a neighborhood, hence, it controls derivatives. The constant α will be related to the properties of the nearest neighbor interaction.

1.1.4. *An example.* A specific model that can serve as motivation is a Frenkel-Kontorova model with a range 2 interaction. The formal variational principle for the model is:

$$\begin{aligned}\mathcal{S}(\{x_n\}) = \frac{1}{2}\sum_n (x_{n+1} - x_n - a)^2 + \frac{A}{2}(x_{n+2} - x_n - b)^2 \\ + \frac{\varepsilon}{4\pi^2}(\cos(2\pi x_n) - 1)\end{aligned} \tag{12}$$

The equilibrium equations for the Lagrangian (12) are:

$$(x_{n+1} - 2x_n + x_{n-1}) + A\cdot(x_{n+2} - 2x_n + x_{n-2}) + \frac{\varepsilon}{2\pi}\sin(2\pi x_n) = 0 \tag{13}$$

The equation (13) defines a 4-dimensional map for $A \neq 0$. In that case, using (13), if we are given $x_{i-2}, x_{i-1}, x_i, x_{i+1}$, there is one and only one x_{i+2} so that (13) is satisfied. Hence, associated to (13), there is a mapping that given

$$X_i = (x_{i-2}, x_{i-1}, x_i, x_{i+1})$$

produces

$$X_{i+1} = (x_{i-1}, x_i, x_{i+1}, x_{i+2}).$$

Notice that, when $A = 0$, the equation (13) reduces to the – two dimensional – standard map. Hence, when A is small, (13) can be considered as a singular perturbation of the standard map. Note that the dimension of the phase space change. On the other hand, it seems quite natural that, for small A, the variational principle is very similar to the solutions of the unperturbed system.

As we will see, given the fact that our main result is an a-posteriori result, it will follow that the solutions produced are continuous, indeed, analytic in A.

Other generalizations of (12) which are interesting are:

$$\begin{aligned}\mathcal{S}(\{x_n\}) = \frac{1}{2}\sum_n (x_{n+1} - x_n - a)^2 + \sum_{j\geq 2} \frac{A_j}{2}(x_{n+j} - x_n - b_j)^2 \\ + \frac{\mu}{4\pi^2}(\cos(2\pi x_n) - 1)\end{aligned} \tag{14}$$

with A_j decaying sufficiently fast. The equilibrium equations for (14) are:

(15)
$$(x_{n+1} - 2x_n + x_{n-1}) + \sum_{j\geq 2} A_j \cdot (x_{n+j} - 2x_n + x_{n-j}) + \frac{\mu}{2\pi} \sin(2\pi x_n) = 0$$

Notice that if all the A_j are non-zero, the equilibrium equations are not a finite dimensional dynamical system.

In Section 5.1 we will study the existence of Lindstedt series for equations generalizing (15).

Remark 2. When A is small, it is natural to consider the model in (13) as a perturbation of the case $A = 0$. Nevertheless, we note that, from the point of view of dynamical systems, for $A > 0$ (even if it is small), the equilibrium equations define a dynamical systems in a 4 dimensional phase space. On the other hand, for $A = 0$, the equilibrium equations define a dynamical system in a two dimensional phase space.

We will present some tentative singular perturbation approach to this problem in Section 6. The calculations in this section will also lead to the conclusion that the systems do not admit a Hamiltonian formulation.

Remark 3. It is interesting to notice that there is an Aubry-Mather theory for quasi-periodic solutions of (13) provided that $A > 0$ (see [CdlL98, dlL00]).

1.2. **Equilibrium configurations.** The equilibrium configurations of models of the form (7) are defined as usual as the solution of Euler-Lagrange equations which we indicate formally as

$$\partial_{x_i} \mathcal{S}(\mathbf{x}) = 0 \qquad i \in \mathbb{Z} \tag{16}$$

For models of the form (7), the equations (16) are just

$$0 = \sum_{L} \sum_{\substack{k+j=i \\ j=0,\ldots,L \\ j\in\mathbb{Z}}} \partial_j H_L(x_k, \ldots, x_{k+j}, \ldots, x_{k+L}) \quad \forall i \in \mathbb{Z} \tag{17}$$

When the interactions are finite range, the sums in (17) become just finite sums. In general, our decay assumptions (see e.g., (11) and **H2** in Theorem 1) will be strong enough that the sums in (17) converge uniformly. Hence, in contrast with the sums in (2), (7) which are merely formal sums, the Euler-Lagrange equations (17) are well defined.

It is the equilibrium equations (17) that will be the basis of our study. Indeed, what we will do in this paper is to study procedures to construct solutions of equations (17). The fact that these equations have a variational motivation will not be used explicitly in our study of them, even if some of the identities we will use can be traced back to the variational structure.

In the case of finite range interactions, the assumptions (2) show that if we are given $x_{i-R}, x_{i-R+1}, \ldots, x_0, \ldots, x_{i+R-1}$ there is one and only one x_{i+R} which satisfies the equations (17) for $i = R$. In other words, given $x_{i-R}, x_{i-R+1}, \ldots, x_0, \ldots, x_{i+R-1}$, the equation (2) determines $x_{i-R+1}, x_{i-R+2}, \ldots, x_0, \ldots, x_{i+R}$, Hence, we can think of (17) for interactions of range R as defining a dynamical system in $(\mathbb{R})^{2R}$. Note that, because of the invariance under translation of the interaction, the dynamical system defined above is autonomous.

The systems thus defined, have been called "monotone recurrences" in [Ang90] when the second derivative satisfies (2). These monotone recurrences enjoy remarkable dynamical properties such as an Aubry-Mather theory on existence of quasi-periodic orbits with all frequencies. See [Ang90, CdlL98].

In this paper we will consider KAM results for models such as those considered in Section 1.1.3. We will prove that if there is a periodic function u such that $E_\omega[u]$ defined in (21) is small enough (with respect to other properties of the function and the rotation ω) then, there is another function u^* for which $E[u^*] = 0$. Moreover we can bound $u - u^*$ in terms of the original error of the approximation. Since the precise formulation of the result requires definitions of the norms used to measure the error and to formulate some non-degeneracy assumptions, we defer the formulation of Theorem 1 till we have introduced enough notation and definitions.

2. Statement of results

We will consider models of the form (7) but with the variables x_n being vectors in $\mathbb{R}$ as indicated in Section 1.1.

2.1. **Plane-like configurations, hull functions.** We will be interested in equilibrium configurations $\{x_n\}_{n\in\mathbb{Z}} \subset \mathbb{R}$ that can be written as:

$$x_n = h(n\omega) \tag{18}$$

where $\omega \in \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ satisfies

$$h(x + e) = h(x) + e \quad \forall e \in \mathbb{Z} \tag{19}$$

Notice that because of the periodicity assumption (19) h can be considered a map from the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ to itself. In our applications, we will assume that h is a diffeomorphism of the torus.

We will use the notation $h(\theta) = \theta + u(\theta)$ and often work with the u function which is periodic.

Remark 4. In solid state physics the function h is often referred as "hull" function of the configuration. In dynamical systems, the function h gives a semi-conjugacy between the dynamics and a rotation on a torus. KAM theory always looks for solutions of the form (18).

An obvious consequence of the form (19) is that $|x_i - \omega i| \le \|u\|_{C^0}$. Hence, the configurations that can be represented by hull functions are almost linear (usually called "plane-like" in homogenization theory). Our assumption that h is a diffeomorphism of the circle implies that $\|u\|_{C^0} \le 1$.

The property that a configuration is given by a hull function is equivalent to the fact that it satisfies the so-called "Birkhoff property" or "non-intersection property" or "self-conforming property", which was introduced in [Mat82, ALD83]. This property is very important in variational calculus. Under hypothesis such as the maximum principle (which is implied by (2)) all periodic minimizers satisfy the Birkhoff property.

2.2. **Equilibrium equations for hull functions.** We note that the Euler-Lagrange equations (17), that express that the configuration is in equilibrium, evaluated on a configuration described by a hull function $h = \mathrm{Id} + u$ are just

(20)

$$
\begin{aligned}
0 = \sum_L \sum_{k+j=i;j=0,\dots,L} \partial_j H_L\Big(& h(\theta + k\omega), \dots, h(\theta + (k+j)\omega), \dots \\
& , \dots, h(\theta + (k+L)\omega)\Big) \\
\sum_L \sum_{k+j=i} \partial_j H_L\Big(& \theta + k\omega + u(\theta + k\omega), \dots, \theta + (k+j)\omega + u(\theta + (k+j)\omega), \dots \\
& , \dots, \theta + (k+L)\omega + u(\theta + (k+L)\omega)\Big)
\end{aligned}
$$

If ω is irrational, (20) are satisfied if and only if $E[u](\theta)$ defined below vanishes identically

$$
\begin{aligned}
E_\omega[u](\theta) &\equiv \sum_L \sum_{j=0}^{L} \partial_j H_L\Big(h(\theta - j\omega), \dots, h(\theta), \dots, \\
&\qquad \dots h(\theta + (L-j)\omega)\Big) \\
&\equiv \sum_L \sum_{j=0}^{L} \partial_j H_L\Big(\theta - j\omega + u(\theta - j\omega), \dots, \theta + u(\theta), \dots, \\
&\qquad \dots \theta + (L-j)\omega + u(\theta + (L-j)\omega)\Big)
\end{aligned}
\tag{21}
$$

Note that under the periodicity assumptions for H_L and for u, $E_\omega[u](\theta)$ is a periodic function of θ.

When ω is understood – in most of this paper we will be considering ω a fixed Diophantine number – we will suppress the ω from the notation.

Remark 5. We also remark that $E[u](\theta)$ is the Euler-Lagrange equation associated to the functional $\mathcal{P}^\omega$ defined on periodic functions by

$$
\mathcal{P}[u] = \int_0^1 \sum_L \frac{1}{L} H_L\Big(\theta + u(\theta), \dots, \theta + L\omega + u(\theta + L\omega)\Big) \tag{22}
$$

This is an extension of the functional introduced in [Per74, Per79] for computational purposes in the case of nearest neighbor interactions. This functional was used in [Mat82] to establish existence of quasi-periodic solutions which were given by monotone – but possibly non-smooth – hull functions h.

2.2.1. *A symmetry of the equilibrium equations.* As motivation, we will start by examining some consequence of the invariance under translation.

Note that if $h(\theta) = \theta + u(\theta)$ is the hull function for a configuration x, then for $k, \ell \in \mathbb{Z}$

$$
\tilde{h}(\theta) = \theta + k\omega + u(\theta + k\omega) + \ell \tag{23}
$$

is the hull function for the configuration obtained shifting the argument in x by k and adding an integer to it. Since the interaction is invariant under translation (see (7)) and by addition of integers to a configuration (see (8)), it is clear that if the configuration $x_n = h(n\omega)$ is an equilibrium, the configuration whose hull function is $\tilde{h}$ given in (23) is also an equilibrium. If ω is irrational, then, $k\omega$ is dense in the torus. Therefore, if u is a continuous solution of the equilibrium equation (20), so is

$$u_\sigma = u(\theta + \sigma) + \sigma \tag{24}$$

for any choice of $\sigma \in \mathbb{T}$.

The fact that (24) is a solution of (20) for all σ can be also checked by substituting directly. It is valid for all ω, including rational ones.

In summary, in general, the quasi-periodic solution of the equilibrium equations are not unique and indeed appear in one parameter families. This corresponds to the choice of the origin of time in the of the torus parameterization, which is a symmetry of the problem. The diffeomorphism of the torus h_σ corresponding to u_σ is just the translation of the origin in h. That is $h(\theta + \sigma) = h_\sigma(\theta)$.

The fact that the solutions come in families will play an important role in the study of the equations of equilibrium. Using (24), we see that given one u, we can find a unique σ such that $\tilde{u}_\sigma$ has zero average.

An important consequence of the symmetry of the problem under translation and changes of phases is the identity:

$$E[u_\sigma](\theta) = E[u](\theta + \sigma). \tag{25}$$

This identity will play an important role in in Section 3.3.

Actually, the symmetry under changes of the origin of the phases is true not only for the equilibrium equations but also for the variational principle (22). We have

$$\mathcal{P}[u_\sigma] = \mathcal{P}[u]. \tag{26}$$

This identity will also play a role in Section 3.3.

Given the importance of the symmetry under shifts in the formulas, it is important for the analysis that the norms we use are also invariant under shifts. The norms we introduce in Section 2.3.1, are indeed invariant under shifts in the parameterization. The non-degeneracy conditions we will consider in Theorem 1 are also invariant under changes of the origin of the phase. Hence, we can assume without loss of generality that the approximate solutions we consider are such that the average of $E[u]$ is zero.

2.3. **Formulation of the main result.** The main result Theorem 1 below will be of the form that if there is a function which makes $E[u]$ small, and which satisfies non-degeneracy conditions, then close to it, we can find an exact solution close to the approximate one. This solution is unique – up to obvious symmetries — in a small neighborhood of the approximate solution we started with.

This formulation of KAM theorems, as a-posteriori results without reference to integrable systems have been common since [Mos66b, Mos66a]. See [dlL01] for a historical review.

In order to formulate that periodic functions are small, we will introduce appropriate norms in Section 2.3.1 and study some of the elementary properties. In Section 2.3.2, we will describe norms on interactions.

2.3.1. *Some families of norms in analytic periodic functions.* We will denote by

$$\mathbb{T}^\rho = \big\{ z \in \mathbb{C}/\mathbb{Z} \quad | \quad |\operatorname{Im} z| \le \rho \big\}$$

We denote by $\mathcal{A}_\rho$ the Banach space of functions from $\mathbb{T}^\rho$ taking values in $\mathbb{C}^L$ (when L is understood from the context we will omit it from the notation) such that

- They are real for a real argument.
- They are holomorphic in the interior of $\mathbb{T}^\rho$ and continuous on $\mathbb{T}^\rho$.

We consider $\mathcal{A}_\rho$ endowed with the norm

$$\|u\|_\rho = \sup_{z \in \mathbb{T}^{d,\rho}} |u(z)|. \tag{27}$$

This norm makes $\mathcal{A}_\rho$ a Banach space.

Since the set of L that we will consider will be unbounded, it is important to specify that the norm we will use in $\mathbb{C}^L$ is the supremum of the coordinates. Of course, for finite L all the norms in $\mathbb{C}^L$ are equivalent, but the constants given the equivalence could be unbounded as L grows, so that one needs to pay attention to the choice of norms. We have not optimized the choice of norms in $\mathbb{C}^L$. So that it is quite possible that other choices could lead to sharper results. This will also become important when we choose norms in the space of interactions in Section 2.3.2.

It is clear that if $\otimes$ is a bilinear operation of norm 1 (e.g an inner product, multiplication or matrices with their operator norms), we have:

$$\|u \otimes v\|_\rho \le \|u\|_\rho \cdot \|v\|_\rho \tag{28}$$

We also recall that we have Cauchy estimates for derivatives and for the Fourier coefficients in terms of the family of norms (27). We will write the Fourier series of a function $u \in \mathcal{A}_\rho$ as

$$u(\theta) = \sum_{k \in \mathbb{Z}} u_k \, e^{2\pi i k \theta}$$

Proposition 1. *Let $u \in \mathcal{A}_{\rho,L}$, with $\rho > 0$, $L \in \mathbb{N}$, then $D_\theta^j u \in \mathcal{A}_{\rho',L}$ for every $0 < \rho' < \rho$. Moreover,*

$$\|D_\theta^j u\|_{\rho'} \leq C\,(\rho - \rho')^{-j}\|u\|_\rho$$
$$|u_k| \leq e^{-2\pi\rho|k|}\|u\|_\rho$$

The proof of this proposition can be found in any book in complex analysis. In Section 2.3.2 we will present in detail similar standard proofs in other more complicated cases to make sure that we obtain the dependence in L.

The following lemma estimates the composition of functions and shows it is differentiable.

Lemma 1. *Let $u \in \mathcal{A}_{\rho,L}$. Let $U \subset \mathbb{C}^L$ be such that*

$$\operatorname{dist}(\mathbb{C}^L - U, u(\mathbb{T}^\rho)) \geq \delta > 0.$$

Let $f : U \to \mathbb{C}$ be an analytic function $\|f(z)\|_{L^\infty(U)} \leq M$. Let $\eta \in \mathcal{A}_{\rho,L}$.

Then, we have:

a) *If $||\eta - u||_\rho < \delta$, then $f \circ \eta \in \mathcal{A}_\rho$. Moreover,*

$$|f \circ \eta|_\rho \leq M.$$

b) *If $||u - \eta||_\rho \leq \delta/2$.*

Then, the mapping $\tilde{f} : \mathcal{A}_\rho \to \mathcal{A}_\rho$ defined by $\tilde{f}[\gamma] = f\circ(\eta+\gamma)$ is an analytic mapping from the set $\|\gamma\|_\rho < \delta/2$ to $\mathcal{A}_\rho$. Moreover, we have the following explicit formula for the derivative of the operator $D\tilde{f}$ and bounds for the reminder"

c)

$$(D\tilde{f})[\eta]\gamma = f' \circ \eta \cdot \gamma \tag{29}$$

If $\|\varphi\|_\rho < \delta/2$,

$$\begin{aligned}
&\|\tilde{f}[\eta + \varphi] - \tilde{f}[\eta] - D\tilde{f}[\eta]\varphi\|_\rho \\
&\qquad \leq 2M\delta^{-2}\|\varphi\|_\rho^2\|f \circ (\eta + \varphi) - f \circ \eta\|_\rho \\
&\qquad \leq 2M\|\varphi\|_\rho
\end{aligned} \tag{30}$$

The proof of Lemma 1 is an straightforward and standard application of the Taylor theorem with uniform estimates and the Cauchy estimates for the derivatives. We leave it for the reader.

Remark 6. We emphasize that the $D\tilde{f}$ in (30) refers to the derivative of the operator $\tilde{f}$ acting on a space of functions.

It is interesting to compare this with the derivative with respect to the variable θ of the function $\tilde{f}[\eta](\theta) = f \circ \eta(\theta)$. We have $\frac{d}{d\theta} f[\eta](\theta) = f' \circ \eta(\theta)$.

Even if $D\tilde{f}$ and $\frac{d}{d\theta} f[\eta](\theta)$ are conceptualy very different, they have the same formula. This will play an inportant role later leading to interesting cancellations.

2.3.2. *Regularity properties of the interactions.* We will assume that the interactions H_L are defined on a complex set and analytic there. We will need that the functions H_L are analytic in a complex domain large enough to allow their evaluation on the interactions we are considering. The goal of this section is to define norms in the interactions that measure their sizes so that we can state precisely the results. For the spaces that we will consider, it will be quite important to consider not only the sizes but also the domains. Even if we will not consider this in this paper, we hope that this will allow to extend the results to finite differentiable interactions using the techniques of [Mos66b, Mos66a, Zeh75] which characterize finite differentiable functions by their approximation properties. Thinking about this further extension, the result we will present will pay special attention to the dependence of the smallness conditions and the change required in the conclusions with respect to the domain of the function.

In the applications to physical problems, the assumption that the interactions are analytic in a large domain will be often satisfied.

Since our result will be formulated for an approximate solution, it is natural to consider domains for the interaction which are defined in the a neighborhood of the range of of a configuration.

Given $u \in \mathcal{A}_\rho$, we consider

$$
\begin{aligned}
\mathcal{D}_{L,u,\delta} = \Big\{ &(z_0, \dots, z_L) \subset (\mathbb{C})^{L+1} \;\Big|\; \\
&\exists\, \theta \in \mathbb{T}^{d,\rho}, |z_i - h(\theta + i\omega)| \le \delta, i = 0, \dots, L \Big\}
\end{aligned}
\tag{31}
$$

As usual, we suppress the dependence in ω from the notation unless it can cause confusion.

Again, we recall that the norms we use in $\mathbb{C}^L$ are the supremum norms.

Since our configurations will be such that they map real values into real values, in some applications it suffices to consider the simpler domains

$$
\tilde{\mathcal{D}}_{L,\delta} = \Big\{ (z_0, \dots, z_L) \subset (\mathbb{C})^{L+1} \quad \Big| \quad |\operatorname{Im}(z_i)| \le \delta \Big\}
\tag{32}
$$

Clearly,

$$\mathcal{D}_{L,u,\delta} \subset \tilde{\mathcal{D}}_{L,\|u\|_\rho+\delta}. \tag{33}$$

Since L will be unbounded, we will need to estimate the dependence in L of several standard results such as Cauchy estimates and the like. The proofs we present are not optimized very seriously.

With the choice of supremum norm in $\mathbb{C}^{L+1}$, we have

$$\begin{aligned} \sup_{\theta\in\mathbb{T}^\rho} |\big(u(\theta), u(\theta+\omega), \dots, u(\theta+L\omega)\big) \\ - (\tilde{u}(\theta), \tilde{u}(\theta+\omega), \dots, \tilde{u}(\theta+L\omega)| \\ \le ||u-\tilde{u}||_\rho \end{aligned} \tag{34}$$

Therefore, we have

$$\begin{aligned} H_L((u(\theta), u(\theta+\omega), \dots, u(\theta+L\omega)) \\ - H_L(\tilde{u}(\theta), \tilde{u}(\theta+\omega), \dots, \tilde{u}(\theta+L\omega)| \\ \le ||DH_L||_{L^\infty} ||u-\tilde{u}||_\rho \end{aligned} \tag{35}$$

On the other hand, we note that the estimate of the norm of a derivative in terms of the partial derivatives does have a dependence on L.

$$|DH_L| \le (L+1) \max_{j=0,\dots L} |\partial_{x_j} H_L|$$

The Cauchy bounds may also have a dependence in L. In Lemma 2, we state the version of Cauchy estimates which we will use (even if we do not know if it is optimal).

Lemma 2. *If* $\Omega \subset \tilde{\Omega}$ *and* $\operatorname{dist}(\Omega, \mathbb{C}^{L+1} - \tilde{\Omega}) \ge \delta$ *we have:*

$$||DH_L||_\Omega \le C(L+1)\delta^{-1} ||H_L||_{\tilde{\Omega}} \tag{36}$$

Proof. Given $z \in \Omega$ we can find circles γ_i centered in z_i with radius δ such that $\gamma = (\gamma_0, \dots, \gamma_L) \subset \Omega$.

Cauchy formula gives:

$$H_L(z) = \frac{1}{(2\pi i)^{L+1}} \int_{\gamma_0} dw_0 \cdots \int_{\gamma_L} dw_L \frac{H_L(w)}{(w_0 - z_0) \cdot \cdots (w_L - z_L)}$$

Therefore, given a direction η, $D_\eta H_L$, the directional derivative is:

(37)

$$\begin{aligned} D_\eta H_L(z) \\ = \frac{1}{(2\pi i)^{L+1}} \int_{\gamma_0} dw_0 \cdots \int_{\gamma_L} dw_L H_L(w) \Big[\frac{\eta_0}{(w_0 - z_0)^2 \cdot (w_1 - z_1) \cdot \cdots (w_L - z_L)} + \cdots \\ \frac{\eta_L}{(w_0 - z_0) \cdot (w_1 - z_1) \cdot \cdots (w_L - z_L)^2} \Big] \end{aligned}$$

Using that

$$\int_{\gamma_i} dw_i \frac{1}{|w_i - z_i|} \le 2\pi$$
$$\int_{\gamma_i} dw_i \frac{1}{|w_i - z_i|^2} \le 2\pi\delta^{-1}$$

we obtain the claimed result. □

Corresponding to the domains $\mathcal{D}_{L,u,\delta}$, $\tilde{\mathcal{D}}(L,\delta)$ we consider the spaces $\mathcal{H}_{L,u,\delta}$, $\mathcal{H}_{L,\delta}$ consisting of functions analytic in the interior and continuous in the whole domain. We endow these spaces with the supremum norm, which makes them Banach spaces.

$$\begin{aligned} \|H_L\|_{L,u,\delta} &= \sup_{z\in\mathcal{D}_{L,u,\delta}} |H_L(z)| \\ \|H_L\|_{L,\delta} &= \sup_{z\in\tilde{\mathcal{D}}_{L,\delta}} |H_L(z)| \end{aligned} \tag{38}$$

By (33), we have $\mathcal{H}_{L,u,\delta} \subset \mathcal{H}_{L,\|u\|_\rho+\delta}$ and $\|H_L\|_{L,u,\delta} \le \|H_L\|_{L,\|u\|_\rho+\delta}$.

2.3.3. *Statement of the main result of the paper.* Following standard practice in KAM theory, we will denote by C numbers that depend only on on combinatorial factors but are independent of the size of the domains considered, the Diophantine constants κ or the size of the error assumed. In our case, we will also require that they are independent of L, the range of the interactions. The meaning of this constants can change from one formula to the next.

The main result of this paper is

Theorem 1. *Let H be a translation invariant interaction as in* (7) *satisfying the periodicity condition* (8).

Let $\omega \in \mathbb{R}$. Let $h = \mathrm{Id} + u$, with $u \in \mathcal{A}_\rho$, $\int_{\mathbb{T}} u = 0$ be an analytic diffeomorphism of $\mathbb{T}$.

Assume:

H1) *ω is Diophantine, i.e., for some $\kappa > 0$, $\tau > 0$*

$$|p\omega - q| \ge \kappa |q|^{-\tau} \quad \forall p \in \mathbb{Z} - \{0\}, q \in \mathbb{Z}. \tag{39}$$

H2) *The interactions $H_L \in \mathcal{H}_{L,u,\delta}$ for some $\delta > 0$.*

Denote

$$M_L = \max(\|D^i H_L\|_{L,u,\delta}), \quad i = 0,1,2,3$$
$$\alpha = C \sum_{L\ge 2} M_L L^4$$

where C is a combinatorial constant that will be made explicit during the proof.

H3) *Assume that the inverses indicated below exist and have the indicated bounds.*

H3.1)
$$\|(\partial_0\partial_1 H_1)^{-1}(u(\theta), u(\theta+\omega))\|_\rho \leq T.$$

H3.2) *Define*
$$\mathcal{C}_{0,1,1}(\theta) = \partial_0\partial_1 H_1(u(\theta), u(\theta+\omega))h'(\theta)h'(\theta+\omega)$$
(See (56) *later for a justification of including the subindices in $\mathcal{C}$)*
Assume that:
$$\left(\int_{\mathbb{T}} \mathcal{C}_{0,1,1}^{-1}\right)^{-1} \leq U.$$

H4) *The following bounds measure the non-degeneracy and the accuracy with which the approximate solution solves the problem.*

H4.a) $\|\operatorname{Id} + u'\|_\rho \leq N_+$, $\|(\operatorname{Id} + u')^{-1}\|_\rho \leq N_-$.

H4.b) $\|E[u]\|_\rho \leq \varepsilon$.

Assume furthermore that the above upper bounds satisfy the following relations:

H4.i) $T\alpha < 1/2, \quad UT\alpha < 1/4$

H4.ii) $\|u\|_\rho + \rho \leq \frac{1}{2}\delta$

H4.iii) $\varepsilon \leq \varepsilon^*(N_+, N_-, \tau, \alpha, T, U, \delta)\kappa^4\rho^{4\tau+A}$

where ε^* is a strictly positive function which we will make explicit along the proof. The function ε^* makes quantitative the relation between the smallness conditions and the nondegeneracy conditions. $A \in \mathbb{R}^+$ is a number which will also be made explicit along the proof.

Then, there exists a periodic function $u^* \in \mathcal{A}_{\rho/2}$ such that

$$\tilde{E}_\omega[u^*] = 0 \tag{40}$$

and $\int u^* = 0$.

Moreover

$$\|u - u^*\|_{\rho/2} \leq C\kappa^{-2}\rho^{-2\tau-A}T\|E[u]\|_\rho \tag{41}$$

The function u^* is the only function in a ball of radius centered at u of radius $C\kappa^{-2}\rho^{-2\tau-A}T\|E[u]\|_\rho$ satisfying (40) *and the normalization. $\int u^* = 0$.*

The most important hypothesis of Theorem 1 is H4.iii) which requires that the test function satisfies the equilibrium equation with an error with is sufficiently small with respect to the other quantities that measure the degeneracy of the problem.

The hypothesis H3.1 is the twist condition for the nearest neighbor interaction. The condition H3.2 defines some quantity related to the twist. The quantity in H3.2 appears in H4.i. The importance of H4.i will come from the fact that it will allow us to show that the non-degeneracy condition in the nearest neighbor interaction allows us to control the non-degeneracy of all the other interactions.

Note also that the dependence of the smallness conditions on the domain of analyticity ρ is a power and also that the effect of the analyticity domain in the correction (41) is also a power of ρ. It is well known to experts that such a result can be used to prove a finitely differentiable result. We hope to come back to this problem.

We also note that such a result implies that there is Lipschitz dependence on the Diophantine frequencies. (Just take the solution for a frequency as an approximate solution for a nearby frequency). Indeed, by combining a result such as the one above with the Lindstedt series, one gets that the dependence on the frequencies is smooth in the sense of Whitney. See [Van02, dlLV00].

Theorem 1 implies the result of persistence of solutions for quasi-integrable systems. In our case, the integrable systems are the linear systems:

$$H_L^0(x_0, \ldots, x_L) = A_L \frac{1}{2} |x_0 - x_L|^2. \tag{42}$$

For systems of the form (42), given $\omega \in \mathbb{R}$, $x_n = \omega n$ is a solution, which corresponds to $u = 0$. If we consider a system $H_L = H_L^0 + \mu F_L$. If the perturbation satisfies $\|F\|_{L,\delta} L^2 < \infty$, for $|\mu|$ sufficiently small, we can consider $u = 0$ as an approximate solution of the system.

Note that, to verify the hypothesis of Theorem 1, ε – the error – is bounded by μ and the non-degeneracy conditions remain uniformly bounded as μ approaches zero. Hence, we can obtain the existence of solutions with corresponding frequencies. A more detailed discussion of results for quasi-integrable systems happens in Section 5.1.

3. Description of the proof of Theorem 1

The proof is based on an iterative procedure very similar to that of [Mos88, SZ89, LM01]. Given a function u so that $\|E[u]\|_\rho$ is sufficiently small compared with other properties of the function, the iterative procedure constructs another function $\tilde{u}$ defined in a smaller domain and such that

$$\|E[\tilde{u}]\|_{\rho'} \le C T A(\tau, N_+, N_-, T, U) \kappa^{-2} (\rho - \rho')^{-2\tau - 1} (\|E[u]\|_\rho)^2 \tag{43}$$

where A is a very easy algebraic function.

As it is well known in KAM theory, the estimates (43) imply that the iterative procedure can be repeated indefinitely and the resulting sequence converges to a function satisfying all the claims of Theorem 1. We will present details later.

The procedure to produce the improved solution we will use here is very similar to that of the papers [SZ89, Mos88, LM01]. Nevertheless, we have to deal with the fact that the interactions are not nearest neighbor. The method of the previous papers depended on some identities [Koz83, Mos88] obtained by combining derivatives of the approximate solution, and using the variational structure and the symmetry of the problem. These identities allow us to reduce the problem to a constant coefficients problem. As it turns out, similar identities hold in our case. and we can obtain a similar factorization. Once this factorization is obtained, we can deal treat the non-nearest neighbor interactions perturbatively.

3.1. **Heuristic description of the iterative step.** The iterative step is very similar to the iterative step in the previous papers [Mos88], [SZ89], [LM01]. The step is based on a modification of the Newton method which makes the linearized equation used in the Newton method readily solvable but does not change the quadratic convergence. See (53).

In this section we will describe the iterative procedure somewhat formally. We will specify the manipulations to be carried out with the functions but ignore questions of domains, definition and convergence of series involved. These questions will be addressed in Section 4 where we will develop estimates for the objects considered. These estimates will allow us to verify that the algorithm can indeed be well defined and that the steps indicated formally can be carried out (i.e. the compositions can be defined since the domains match). The estimates will also verify that the step improves the solution in the sense that they satisfy (43).

Undoubtedly some experts will be able to carry out the detailed estimates in Section 4 without reading it. In this presentation, we have followed the notation of [Mos88, SZ89, LM01] as much as possible.

In Sections 3.2, 3.3, 3.4 3.5 we will motivate and specify an iterative procedure that given a sufficiently approximate solution produces a better one. This procedure will be rigorously analyzed in the subsequent sections. In Section 4 we will present estimates that make precise the fact that the iterative procedure improves the approximation (albeit in a smaller domain). In Section 4.3 we will show that the procedure can be iterated indefinitely, and that converges to a solution. Given the estimates in Section 4, the estimates in Section 4.3 are very similar to those in [Zeh76]. In Section 4.4 we will show that the solution is unique in a ball.

3.2. **A Newton step.** The iterative procedure we will use will be a modification of the Newton procedure. The important thing for us that it leads to estimates as in (43). The step will be similar to the step in [LM01] – except in the perturbative treatment of the non-nearest neighbor interactions –. We have tried to follow the notation of [LM01] as much as possible.

As motivation, and to introduce the notation, we will start by discussing the standard Newton method. Given that we have an approximate solution u, a step of the Newton method consists in setting

$$\hat{u} = u + v$$

where v is obtained by solving

$$E'[u]v = -E[u] \tag{44}$$

where E' denotes the derivative of the functional E with respect to its argument.

Proceeding formally for the moment, (we will present precise estimates in Section 4) we compute (the computation can readily be justified using Lemma 1) that:

$$\begin{aligned}(E'[u]v)(\theta) &= \\ &= \sum_L \sum_{j=0}^{L} \sum_{i=0}^{L} \partial_j \partial_i H_L(\theta - j\omega + u(\theta - j\omega), \ldots, \theta + u(\theta), \\ &\qquad \ldots \theta + (L-j)\omega, u(\theta + (L-j)\omega)) v(\theta + (i-j)\omega)\end{aligned} \tag{45}$$

We introduce the notation

$$\begin{aligned} h(\theta) &= \theta + u(\theta) \\ h^{(i)}(\theta) &= \theta + i\omega + u(\theta + i\omega) \\ \gamma_L(\theta) &= (h(\theta), h^{(1)}(\theta), \ldots, h^{(L)}(\theta)) \\ \gamma_L^{(i)}(\theta) &= (h^{(i)}(\theta), h^{(1+i)}(\theta), \ldots, h^{(L+i)}(\theta)) \end{aligned} \tag{46}$$

In particular, $h^{(0)}(\theta) = h(\theta)$, $\gamma_L(\theta) = \gamma_L^{(0)}(\theta)$.

Using the notation above, (45) can be written more concisely as

$$(E'[u]v)(\theta) = \sum_L \sum_{j=0}^{L} \sum_{i=0}^{L} \partial_j \partial_i H_L(\gamma_L^{(j)}(\theta)) v^{(i-j)}(\theta) \tag{47}$$

3.3. **Two important geometric identities.** An identity for the operator $E'[u]$ introduced in (45) which will be extremely important later is

$$\frac{d}{d\theta} E[u](\theta) = (E'[u]h')(\theta) \tag{48}$$

The identity (48) can be verified by a direct calculation, taking derivatives with respect to θ in (21) and comparing with $E'[u]h'$ in (45).

More conceptually, we note that the identity (48) is just the derivative with respect to σ of the equation (25) $E[\tilde{u}_\sigma](\theta) = E[u](\theta + \sigma)$ evaluated at $\sigma = 0$. We recall that the equation (25) expressed the invariance of the problem under changes of the origin of the internal phase.

If we take derivatives with respect to σ in (25) and evaluate for $\sigma = 0$, taking into account that

$$\frac{d}{d\sigma} u_\sigma \big|_{\sigma=0} = \mathrm{Id} + u' = h'$$

we obtain (48).

Remark 7. It has been remarked in [BGK99] that the introduction of the correction term $E'[u]h'$ is similar to the use of Ward identities in quantum field theory.

Recall that the gist of Ward identities is that taking derivatives of solutions with respect to a symmetry of the theory, we obtain an identity that can be used to absolve terms in the perturbation theory or in the renormalization group. In our case, the symmetry of the theory is the covariance under the choice of the origin of the phase expressed by (25).

Another interpretation from the point of view of differential equations can be found in [Koz83, Mos88].

There other possible identities obtained through similar ideas. For example, [Zeh76] points out that other useful identities in the Hamiltonian formalism are related to the invariance of the equations under canonical changes of variables (The group structure). Of course, since our equations (or the partial differential equations of [Mos88]) do not admit a transformation theory, the two phenomena seem different. Nevertheless, the method of proof in [SZ89] uses symmetries of the equation – for flows – that allow one to reduce the Newton method to the case of the identity. I am very indebted to A. González-Enriquez who explained me many of these symmetries.

Another identity that will play an important role is:

Lemma 3. *With the notations above we have*

$$\int_{\mathbb{T}} h'(\theta)E[u](\theta)\,d\theta = 0. \tag{49}$$

Proof. By changing variables in each term of the sum below, we obtain:

$$\begin{aligned}
\sum_{j=0}^{L}\int_{\mathbb{T}} & d\theta\ h'(\theta)\partial_j H_L(h^{(-j)}(\theta)\dots h^{(L-j}(\theta)) \\
&= \sum_{j=0}^{L}\int_{\mathbb{T}} d\theta\ \partial_j H_L(h(\theta)\dots h^{L}(\theta))h'^{(j}(\theta) \\
&= \int_{\mathbb{T}} d\theta\ \frac{d}{d\theta}H_L(h(\theta)\dots h^{L}(\theta)) \\
&= 0
\end{aligned}$$

where we have used the periodicity properties of h (19) and the periodicity properties of H_L (3).

Recalling the expression of $E[u]$ given in (21), the desired result is obtained adding the previous calculation over L. ☐

More conceptually, we note that the equation (49) is related to the variational structure and the invariance under translation of the problem. We recall that, as mentioned before, the equilibrium configurations are critical points of the functional P introduced in (22). The variational principle (22) is invariant under changes of the origins of the phase and in (24) we had s $P[u_\sigma] = P[u]$. A simple calculation shows that:

$$\frac{d}{d\sigma}P[u_\sigma]\Big|_{\sigma=0} = \int h'(\theta)E[u](\theta)\,d\theta$$

and therefore, (49) is a consequence of the invariance under translation of the variational principle.

3.4. **The quasi-Newton method.** Unfortunately, the equation (44) is hard to solve in our case since it involves difference equations with non-constant coefficients.

The trick that works in our case is very similar to the one that was used in [Koz83, Mos88, SZ89] and specially in [LM01]. Namely, our step consists in solving the following equation, which is a modification of (44), the equation suggested by the Newton method.

$$h'(\theta)(E'[u]v)(\theta) - v(\theta)(E'[u]h')(\theta) = -h'(\theta)E(\theta) \tag{50}$$

The equation (50) is just (44) multiplied by the $d \times d$ matrix h' and added the extra term $v(E'[u]h')$.

We will show that indeed, the equation (50) can be solved by reducing it to constant coefficient equations plus an elementary perturbative argument. We will also show that the iterative step, leads to quadratic estimates of the form (43). The reason why the added extra term is small is that, because of (48) we can write

$$v(\theta)(E'[u]h')(\theta) = v(\theta)\frac{d}{d\theta}E[u](\theta) \tag{51}$$

So that, when we show in Section 4 that $\|v\|_{\rho'}$ is estimated by $\|E\|_\rho$ we will have that the extra term is controlled by $\|E\|_\rho^2$. The role of the added extra term will be to make the RHS of equation (50) be factorizable. This same phenomenon happened in [Mos88, SZ89].

3.5. **Solution of the equation of the quasi-Newton method.** The goal of this section is to specify the steps of an algorithm that can be used to solve the equation (50). Once we have specified how to break down (50) into auxiliary problems, we will present estimates for them in Section 4.

Following the previously mentioned references, we introduce a new variable w by

$$v(\theta) = h'(\theta)w(\theta) \tag{52}$$

Notice that the nondegeneracy assumption H4.a of Theorem 1 implies that h' is boundedly invertible, so that, the unknowns v and w are equivalent.

Substituting (52) by (47) into (50) and using the notation in (46), we obtain that the equation to be solved for the step of the modified Newton method is:

$$\begin{aligned}\sum_L \sum_{j=0}^{L}\sum_{i=0}^{L} &\partial_j\partial_i H_L \circ \gamma_L^{(-j)}(\theta)h'(\theta)h'^{(i-j)}(\theta)w^{(i-j)}(\theta)\\ &- \sum_L \sum_{j=0}^{L}\sum_{i=0}^{L} \partial_j\partial_i H_L \circ \gamma_L^{(-j)}(\theta)h'^{(i-j)}(\theta)w(\theta)h'(\theta)\\ = -h'(\theta)E[u](\theta)&\end{aligned} \tag{53}$$

We will analyze separately the terms that appear in the RHS of (53). We first fix L and then, we consider the terms that correspond to a certain i, j. Our goal is to show that the equations can be factored into simpler equations.

We first note that, when $i = j$, the terms in the RHS of (53) cancel. The two terms in the first sum cancel the two terms in the second sum.

When $i \neq j$ we observe that we have four terms involving the mixed derivatives $\partial_j \partial_i H_L \circ \gamma_L^{(j)}$, namely,

$$
\begin{aligned}
&\partial_j \partial_i H_L \circ \gamma_L^{(-j)}(\theta) h'(\theta) h'^{(i-j)}(\theta) w^{(i-j)}(\theta) \\
&+\partial_i \partial_j H_L \circ \gamma_L^{(-i)}(\theta) h'(\theta) h'^{(j-i)}(\theta) w^{(j-i)}(\theta) \\
&-\partial_j \partial_i H_L \circ \gamma_L^{(-j)}(\theta) h'^{(i-j)}(\theta) h'(\theta) w(\theta) \\
&-\partial_i \partial_j H_L \circ \gamma_L^{(-i)}(\theta) h'^{(j-i)}(\theta) h'(\theta) w(\theta)
\end{aligned} \tag{54}
$$

We rearrange (54) as

$$
\begin{aligned}
&\partial_j \partial_i H_L \circ \gamma_L^{(-j)}(\theta) h'(\theta) h'^{(i-j)}(\theta) [w^{(i-j)}(\theta) - w(\theta)] \\
&-\partial_i \partial_j H_L \circ \gamma_L^{(-i)} h'(\theta) h'^{(j-i)}(\theta) [w(\theta) - w^{(j-i)}(\theta)]
\end{aligned} \tag{55}
$$

An observation that is important for us is that the second term in (55) is just the first term shifted $(j-i)\omega$.

We introduce the notation

$$
\begin{aligned}
[\Delta_\ell w](\theta) &\equiv w(\theta + \ell\omega) - w(\theta) \\
&\equiv w^{(\ell)}(\theta) - w(\theta) \\
\mathcal{C}_{i,j,L} &\equiv \partial_j \partial_i H_L \circ \gamma_L^{-j}(\theta)\, h'(\theta) h'^{(i-j)}(\theta)
\end{aligned} \tag{56}
$$

With the notations (56) above, the four terms in (55) corresponding to a fixed i, j, $i \neq j$ can be written:

$$
\Delta_{j-i}[\mathcal{C}_{i,j,L}(\Delta_{i-j} w)] \tag{57}
$$

Therefore, the equation (53) can be written as:

$$
\sum_L \sum_{\substack{j,i=0 \\ j>i}}^{L} \Delta_{j-i} \mathcal{C}_{i-j,L} \Delta_{i-j} w = h' E[u] \tag{58}
$$

Remark 8. As a curiosity, we note that some of the identities between the terms in (54) can be considered physically as a consequence of the action-reaction principle. Note that the interpretation of each of the terms is that of the force on one body from another. They are clearly related to the fact that the equations come from a variational principle. The other identities used come from the invariance under shifting of the origin of the internal phase, which in turn is a consequence of the invariance under translations of the model and the invariance under global changes of the phase by an integer.

The basic idea we will use is that, under the hypothesis *i*) of Theorem 1, the equation (58) can be treated as a perturbation of the term corresponding to $L = 1$ corresponding to nearest neighbor interactions.

To accomplish this perturbative treatment, it will be important to study conditions for invertibility of the operators Δ_ℓ. A precise formulation of this study will be carried out in Section 4.1. For the moment, we will just perform formal manipulations to motivate and specify the procedure to be followed in Procedure 1. Later we will provide estimates that show that the procedure can be implemented and give bounds on the results.

The first observation is that the operator Δ_1 is diagonal on Fourier series. Hence, if we consider the equation for w given η

$$\Delta_\ell w = \eta \tag{59}$$

when $\eta = \sum \eta_k e^{2\pi i k\theta}$, provided that $\eta_0 = 0$, we can find (at least formal) solutions of (59) by setting, for $k \neq 0$:

$$w_k = \eta_k/(e^{2\pi i \ell k\cdot\omega} - 1) \tag{60}$$

(recall that we are assuming that ω is Diophantine, hence, in particular, $\ell k\omega \in \mathbb{Z} \implies k = 0$) These solutions η are unique up additive constants.

We will denote by Δ_1^{-1} the operator that given η produces the w with zero average. This makes into a linear operator.

These operators appear very often in KAM theory and have been extensively studied. In Section 4.1 we state the results of [Rüs75, Rüs76], which are optimal with respect to the loss of differentiability.

Hence, we can define the operators

$$\mathcal{L}_\ell^{\pm} = \Delta_{\pm 1}^{-1}\Delta_\ell$$

acting on all the functions in $\mathcal{A}_\rho$ and the operators

$$\mathcal{R}_\ell^{\pm} = \Delta_\ell \Delta_{\pm 1}^{-1}$$

defined for all functions with zero average.

Note that all the operators $\mathcal{L}_\ell^{\pm}$ and $\mathcal{R}_\ell^{\pm}$ have as range the set of functions with zero average.

Remark 9. The operators $\mathcal{R}_\ell^{\pm}$ $\mathcal{L}_\ell^{\pm}$ are not identical since $\mathcal{R}_\ell^{\pm}$ can only be defined on functions with integral zero. When applied to functions with integral zero, the operators $\mathcal{R}, \mathcal{L}$ agree. In other words, the only difference among them is that $\mathcal{R}_\ell^{\pm}$ have a domain which is a codimension 1 space while the domain of $\mathcal{L}_\ell^{\pm}$ is the whole space.

The key observation that allows to treat the equation (53) as a perturbation of the nearest neighbor case is the fact that, in spite of the fact that $\Delta_{\pm 1}^{-1}$ are unbounded operators, we have:

$$
\begin{aligned}
\|\Delta_{\pm 1}^{-1}\Delta_\ell\|_\rho &\le |\ell| \\
\|\Delta_\ell \Delta_{\pm 1}^{-1}\|_\rho &\le |\ell|.
\end{aligned} \tag{61}
$$

The (elementary) proof of this key result is postponed till Lemma 6.

Coming back to the solution of the equation (53), we observe that separating explicitly the nearest neighbor terms, we have that (58) can be written as

$$
\begin{aligned}
-h'E[u] &= \Delta_{-1}\mathcal{C}_{0,1,1}\Delta_1 w + \sum_{L\ge 2}\sum_{i>j}\Delta_{i-j}\mathcal{C}_{ij,L}\Delta_{j-i}w \\
&= \Delta_{-1}\left[\mathcal{C}_{0,1,1} + \sum_{L\ge 2}\sum_{i>j}\Delta_{-1}^{-1}\Delta_{i-j}\mathcal{C}_{i,j,L}\Delta_{j-i}\Delta_1^{-1}\right]\Delta_1 w
\end{aligned} \tag{62}
$$

By the twist assumption property H3, the operator $\mathcal{C}_{0,1,1}$ is boundedly invertible from the spaces $\mathcal{A}_{\rho'}$ (with zero integral) to itself.

Similarly, we observe that the operator $\mathcal{G}$ is defined on the functions $\mathcal{A}_{\rho''}$ with zero average. This is acceptable for our applications because the operator $\mathcal{G}$ in (62) is only applied to $\Delta_1 w$, which has zero average. We also note that the operator $\mathcal{G}$ always produces functions with zero average. All the terms in the sum defining $\mathcal{G}$ in (63) have in the left an operator $\mathcal{L}_{i-j}^-$ which produces functions with zero average.

The decay properties on the interaction which we assumed in Theorem 1 show that

$$
\mathcal{G} \equiv \sum_{L\ge 2}\sum_{i>j}\Delta_{-1}^{-1}\Delta_{i-j}\mathcal{C}_{i,j,L}\Delta_{j-i}\Delta_1^{-1} \tag{63}
$$

has a small norm from $\mathcal{A}_{\rho''}$ to $\mathcal{A}_{\rho''}$ for all $0 < \rho'' \le \rho$.

Indeed we have using the estimates assumed in Theorem 1 and the bounds for $||\Delta_{-1}^{-1}\Delta_{i-j}||_{\rho''}$ $||\Delta_{j-i}\Delta_1^{-1}||_{\rho''}$ obtained in Lemma 6

$$
\begin{aligned}
\|\mathcal{G}\|_{\rho''} &\le \sum_{L\ge 2} M_L \sum_{0\le i\le j\le L} |i-j|^2 \\
&\le C\sum_{L\ge 2} M_L L^4 \\
&= \alpha
\end{aligned} \tag{64}
$$

Hence, under the assumptions $i)$ of Theorem 1, the usual Neumann series shows that the operator $\mathcal{C}_{0,1,1} + \mathcal{G}$ is boundedly invertible from $\mathcal{A}_{\rho'}$ to $\mathcal{A}_{\rho'}$. Moreover, we have

$$
\begin{aligned}
\|(\mathcal{C}_{0,1,1} + \mathcal{G})^{-1} - (\mathcal{C}_{0,1,1})^{-1}\|_{\rho'} &\le 1/(1-T\alpha)\|(\mathcal{C}_{0,1,1})^{-1}\mathcal{G}\|_{\rho'} \\
&\le 2T\alpha
\end{aligned}
$$

The equation for $\mathcal{T} \in \mathbb{R}$ can be written as:

$$\left(\int_{\mathbb{T}} C_{0,1,1}^{-1}\right)\mathcal{T} + \int_{\mathbb{T}} \left[(\mathcal{C}_{0,1,1}+\mathcal{G})^{-1} - (\mathcal{C}_{0,1,1})^{-1}\right]\mathcal{T} = \int_{\mathbb{T}} (\mathcal{C}_{0,1,1}+\mathcal{G})^{-1}\varphi$$

Under the assumptions i), we see that the second term in the left hand side of the above equation can be treated as a perturbation of the first term. Therefore,

$$\begin{aligned} |\mathcal{T}| &\le U/(1-2UT\alpha)\|\varphi\|_{\rho'} \\ &\le 2U\|\varphi\|_{\rho'} \end{aligned}$$

Then, w can be obtained by solving the equation $\Delta_{-1}^{-1} w = \varphi + \mathcal{T}$. and, we obtain v by

In summary, the procedure to solve (62) is to follow the following steps

Procedure 1. (1) *Observe that, by Lemma 3, we have* $\int_{\mathbb{T}} h' E[u] = 0$.

(2) *Find a function* φ *(normalized so that* $\int \varphi = 0$*) solving the equation*

$$\Delta_{-1}\varphi = -h' E[u] \tag{65}$$

Therefore, for any constant $\mathcal{T}$

$$\Delta_1(\varphi + \mathcal{T}) = -h' E[u]$$

The equation (65) *is the standard small divisors equation, which has been studied in e.g.* [Rüs75].

(3) *Choose* $\mathcal{T}$ *in such a way that*

$$\int (\mathcal{C}_{0,1,1}+\mathcal{G})^{-1}(\varphi+\mathcal{T}) = 0. \tag{66}$$

To accomplish that, we will show that, under the non-degeneracy assumptions, the linear operator on $\mathcal{T}$ *defined by the LHS of* (66) *is is invertible. This follows from assumption* $H3$) *because the operator* $\mathcal{G}$ *is invertible.*

We emphasize that $\mathcal{T} \in \mathbb{R}$ *and that this step is dividing by a number, which is trivial if the number is non-zero.*

(4) *Obtain* w *by solving*

$$\Delta_1 w = (\mathcal{C}_{0,1,1}+\mathcal{G})^{-1}(\varphi+\mathcal{T})$$

(5) *We set*

$$\tilde{u}(\theta) = u(\theta) + h'(\theta)w(\theta)$$

as our improved solution

Remark 10. We note that the analysis above also allows to conclude uniqueness of the solution of the equation (50). Note that the factorization of the equation into steps is just an identity. Each of the steps provides with unique solutions. As we will see, uniqueness of the solution of this linearized problem will translate into uniqueness of the solution of the full non-linear problem. In turn, this will be useful for the study of problems with finite regularity.

4. Estimates for the iterative step

The goal of this section is to provide precise estimates for the iterative step described in Section 3.1.

The precise statement of the result of this section is the following Lemma 4. This lemma establishes that the procedure indicated in Section 3.1 can be carried out if there are some smallness assumptions (that guarantee that compositions can be defined). The most crucial results is the estimates (70) that establish that the error after the procedure is *quadratically small* with respect to the original error.

Finally, since the final estimates of the step depend on the constants that measure the non-degeneracy properties of the solution, we estimate how these non-degeneracy properties change. This will be important to show that the procedure can be iterated indefinitely using (69). This is very standard in KAM theory.

Remark 11. For further applications (e.g. Section 5.3) it is important to remark that the size of the change of u produced in the iterative step depends only on the size of the non-degeneracy conditions. In particular, if we have families of problems and families of approximate solutions so that they satisfy the hypothesis of the step uniformly, then, we can carry the step for all of them and then, the result has uniform estimates.

Lemma 4. *Under the assumptions of Theorem 1. Let* $0 < \rho' < \rho$. *Assume that*

$$\kappa^{-2}(\rho-\rho')^{-2\tau}(N_+)^2\|E[u]\|_\rho \le \delta/4. \tag{67}$$

Then, the procedure indicated above can be carried out to produce a solution $v \in \mathcal{A}_{\rho'}$ *of* (50) *satisfying* $\int v = 0$.

We have the following estimates:

$$\begin{aligned} \|w\|_{\rho''} &\le C\kappa^{-1}\sigma^{-\tau}\|\varphi\|_{\rho''} \\ &\le C\kappa^{-2}\sigma^{-2\tau}N_+\varepsilon \end{aligned} \tag{68}$$

Denote by $\tilde{u} = u + v, \tilde{h} = \mathrm{Id} + \tilde{h}$ to be the improved approximate solutions obtained applying the procedure, we have:

$$
\begin{aligned}
\|\tilde{u} - u\|_{\rho'} &\le C\kappa^{-2}(\rho - \rho')^{-2\tau}(N_+)^2 \|E[u]\|_\rho \\
\|\tilde{h}' - h'\|_{\rho'} &\le C\kappa^{-2}(\rho - \rho')^{-2\tau-1}(N_+)^2 \|E[u]\|_\rho
\end{aligned}
\tag{69}
$$

We also have have that $E[\tilde{u}]$ is well defined and, moreover, the solution solves the problem more accurately in the sense that:

$$
\begin{aligned}
\|E[\tilde{u}]\|_{\rho'} \le C \left[(N_+)^2 N_- + N_+^4 \sum_L M_L L \right] \cdot \\
\cdot \kappa^{-4}(\rho - \rho')^{-4\tau-1}\big(E[u]_\rho\big)^2
\end{aligned}
\tag{70}
$$

As a consequence of (69), we have the following estimates for the constants that measure the non-degeneracy (We use the same notation as in Theorem 1 but use the $\tilde{}$ to indicate that they are evaluated on the function $\tilde{u}$. We define

$$
\begin{aligned}
\Delta &= \varepsilon\kappa^{-2}(\rho - \rho')^{-2\tau}(N_+)^2 \\
\Delta' &= \varepsilon\kappa^{2}(\rho - \rho')^{-2\tau-1}(N_+)^2
\end{aligned}
$$

the RHS of the equations (69). We have:

$$
\begin{aligned}
&\tilde{T} \le T + C\delta^{-1}\Delta' \\
&\tilde{U} \le U + C\delta^{-1}\Delta K(U, T, \Delta) \\
&\tilde{N}_+ \le N_+ + C\Delta' \\
&\tilde{N}_+ \le N_- + C\Delta' K(N_+, N_+, \Delta')
\end{aligned}
\tag{71}
$$

where K are very simple algebraic functions that will be made explicitly in the proof.

Keep in mind that, as it is standard in KAM theory, when we repeat the procedure, we will show that ε will be much smaller than $(\rho - \rho')^{-2\tau-1}$, hence the estimate for the RHS in (70) is indeed much smaller. Also, the Δ, Δ' in (71) will be very small. Hence the intuition of the step is that it reduces the domain slightly, worsens slightly the non-degeneracy assumptions, but the error is drastically reduced. The losses of domain will will decrease rapidly enough that there is some domain left. Similarly, the non-degeneracy conditions will remain bounded. The quadratic convergence of the error to zero overcomes the other problems. This is perhaps the most important principle in the KAM method. See the subsequent sections for details.

In the rest of this section, we will prove Lemma 4. The most subtle part of the proof is the estimates in points (2) and (4) of the algorithm, but these estimates are classical estimates for small divisors. Rather optimal versions were obtained in [Rüs75, Rüs76].

The rest of the steps are more elementary in nature. As indicated before, the step (3) is just a perturbative argument based on Neumann series. From this, we get (69). Then, using Taylor's theorem for the operator, we get (70). Then, given (69), the estimates (71) are just based on the Neumann series (this is the reason why we get the functions K).

4.1. **Estimates for equations involving small divisors.** The following result is proved in [Rüs75].

Lemma 5. *Assume that* $\omega \in \mathbb{R}$ *satisfies* (39)*. Then, given any function* $\eta \in \mathcal{A}_{\rho,L}$ *satisfying* $\int \eta = 0$ *there is one and only one function* $\varphi \in L^2(\mathbb{T}, \mathbb{C}^L)$ *satisfying*

$$\Delta_{\pm 1}\varphi = \eta \qquad \int \varphi = 0 \tag{72}$$

Moreover, $\varphi \in \mathcal{A}_{\rho',L}$ *for all* $\rho' < \rho$ *and*

$$\|\varphi\|_{\rho'} \leq C\kappa^{-1}(\rho - \rho')^{-\tau}\|\eta\|_{\rho} \tag{73}$$

The constant C *is independent of* L*.*

Since we are using the supremum norm in $\mathbb{C}^L$, it is clear that to prove this result, we can just reduce to studying the equation for each of the components. Each component can be estimated by the norm of the function. Hence, it is clear that the final constant will be independent of the dimension of the image.

We refer to [Rüs75, Rüs76] for the proof of a more general result (e.g. it studies similar equations for functions defined over $\mathbb{T}^d$. Of course, the constants depend on the dimension of the domain. In this paper, we will fix functions defined over the torus).

To prove a weaker result with a worse exponent in $\rho - \rho'$ it suffices to notice that the estimates (39) provide an upper bound for the multipliers in (60). Using the Cauchy estimates that estimate $|\eta_k|$ in terms of $||\eta||_\rho$, one obtains the same result but with bounds $C\kappa^{-1}(\rho - \rho')^{-\tau-1}\|\eta\|_\rho$ which would also be enough for our case.

The more subtle estimates of [Rüs75] use also that the estimates (39) cannot be saturated for k that are very close. Another version of the estimates can be found in [dlL05] and the revised version of [dlL01]. $\square$

We denote by $\Delta_{\pm 1}^{-1}$ the mapping that to η associates φ solving (72).

Lemma 6. *For every $0 < \rho$ we have:*

$$
\begin{aligned}
&\|\Delta_\ell \Delta_{\pm 1}^{-1} \eta\|_\rho \le |\ell| \|\eta\|_\rho \\
&\|\Delta_{\pm 1}^{-1} \Delta_\ell \eta\|_\rho \le |\ell| \|\eta\|_\rho
\end{aligned}
\tag{74}
$$

Proof. Denoting by $\varphi = \Delta_1^{-1}\eta$, we have for $\ell > 0$,

$$
\begin{aligned}
\Delta_\ell \Delta_1^{-1}\eta(\theta) =& \Delta_\ell \varphi(\theta) = \varphi(\theta + \ell\omega) - \varphi(\theta) \\
=& \, [\varphi(\theta + \ell\omega) - \varphi(\theta + (\ell - 1)\omega)] \\
& + [\varphi(\theta + (\ell - 1)\omega) - \varphi(\theta + (\ell - 2)\omega)] \\
& + \cdots + \\
& + [\varphi(\theta + \omega) - \varphi(\theta)] \\
=& \eta(\theta + (\ell - 1)\omega) + \cdots + \eta(\theta)
\end{aligned}
\tag{75}
$$

If $\ell < 0$, we have

$$
\begin{aligned}
\Delta_\ell \Delta_1^{-1}\eta(\theta) =& \Delta_\ell \varphi(\theta) = \varphi(\theta + \ell\omega) - \varphi(\theta) \\
=& \, [\varphi(\theta + \ell\omega) - \varphi(\theta + (\ell + 1)\omega)] \\
& + [\varphi(\theta + (\ell + 1)\omega) - \varphi(\theta + (\ell + 2)\omega)] \\
& + \cdots + \\
& + [\varphi(\theta - \omega) - \varphi(\theta)] \\
=& - \eta(\theta - \ell\omega) - \cdots - \eta(\theta - \omega)
\end{aligned}
\tag{76}
$$

Hence, clearly in both cases, $\|\Delta_\ell \Delta_1^{-1}\|_\rho \le |\ell|$.

All the other cases are proved in the same way or can be deduced from the present ones. □

Remark 12. Note that the result here does not depend on the fact that ω is a Diophantine. The fact that ω is Diophantine is used to define $\Delta_{\pm 1}^{-1}$ as an operator on the whole space.

Note however that, for any irrational ω, the operator $\Delta_\ell \Delta_{\pm 1}^{-1}$ can be defined for trigonometric polynomials and we have the identities (75),(76). Hence, the operator $\Delta_\ell \Delta_{\pm 1}^{-1}$ is bounded in the space of trigonometric polynomials, which makes it uniquely defined and bounded in the whole space of analytic functions.

Remark 13. The diagonal elements of $\Delta_\ell \Delta_1^{-1}$ are

$$
\frac{e^{2\pi i \omega k \ell \omega} - 1}{e^{2\pi i k \omega} - 1} = \begin{cases} 1 + e^{2\pi i k \omega} + (e^{2\pi i k \omega})^2 + \cdots + (e^{2\pi i k \omega})^{\ell - 1} & \ell > 1 \\ -1 - (e^{2\pi i k \omega})^{\ell - 1} - (e^{2\pi i k \omega})^{\ell - 1} - \cdots - (e^{2\pi i k \omega}) & \ell < -1 \end{cases}
\tag{77}
$$

The identity (77) is just the elementary sum of the geometric series in the left. Of course, the identity for the diagonal factors can also be obtained from the identity (75).

Notice that, since the diagonal elements are bounded by $|\ell|$, we obtain that $\Delta_\ell \Delta_{\pm 1}^{-1}$ is also bounded in Sobolev spaces.

Remark 14. Note that the identity (77) shows that $\Delta_1 \Delta_2^{-1}$ is not bounded since the diagonal terms of this operator are

$$\frac{e^{2\pi i k \omega} - 1}{e^{2\pi i k 2\omega} - 1} = \frac{1}{e^{2\pi i k \omega} + 1} \tag{78}$$

Note that the denominator becomes arbitrarily close to zero as k ranges over the integers. Hence, it is not possible to use a simple perturbative argument to treat the general case as a perturbation of the interactions at length 2. We think that it would be quite interesting to see if there is a KAM method that can deal with interactions which are a small perturbation of a next-nearest neighbor interaction. Note that, even if the denominators in $\Delta_1 \Delta_2^{-1}$ can become small, the Diophantine properties of ω imply that to become very small, one has to take k large. Numerical experiments [CdlL05] seem to indicate that indeed one can find tori in this situation. We also note that formal Lindstedt type expansions can be written without too much problem. (See Section 5.1). Of course the variational methods in [CdlL98] do not have any problem dealing with this situation.

4.2. **Estimates for the step in Lemma 4.** In this section we prove estimates (68), (69) for the sizes of the changes produced by the iterative step. The proof is rather straightforward. We just follow the steps of Procedure 1 but we take care of ensuring that all the steps are well defined and give estimates for them.

Since we will have to loose domain repeatedly we will introduce auxiliary numbers $\rho'' = \rho - (\rho - \rho')/2$, We will denote $\rho - \rho' = \sigma$. Hence, $\rho - \rho'' = 2^{-1}\sigma$, $\rho' - \rho'' = 4^{-1}\sigma$, so that we can estimate from below the distance between two of these by $C\sigma$.

In step 1) we estimate $\|h' E[u]\|_\rho \le N_+ \varepsilon$ using the Banach algebra property. Then, by Lemma 5, we have $\|\varphi\|_{\rho''} \le C\kappa^{-1}\sigma^{-\tau} N_+ \varepsilon$.

The assumption H3 implies that $\|\mathcal{C}_{0,1,1}^{-1}\|'_\rho \le T\delta^{-2}$ as an operator from $\mathcal{A}_{\rho''}$ to itself. The assumption implies that $\|\mathcal{G}\| \le \alpha$ also as an operator from $\mathcal{A}_{\rho''}$ to itself.

Since, by assumption H3, $T\alpha < 1$, $\mathcal{C}_{0,1,1} + \mathcal{G}$ is invertible in $\mathcal{A}_{\rho''}$. Moreover,

$$\begin{aligned} \|(\mathcal{C}_{0,1,1} + \mathcal{G})^{-1} - \mathcal{C}_{0,1,1}^{-1}\| &\le \alpha(1 - \alpha T) \\ &\le 1/2T \end{aligned}$$

Using the fact that $\mathcal{C}_{0,1,1}$ is a multiplication operator, we write the equation for $\mathcal{T} \in \mathbb{R}$ as

$$
(79) \qquad \begin{aligned} \left(\int_{\mathbb{T}} \mathcal{C}_{0,1,1}^{-1}\right) \mathcal{T} + \int_{\mathbb{T}} \left[(\mathcal{C}_{0,1,1} + \mathcal{G})^{-1} - \mathcal{C}_{0,1,1}^{-1}\right] \mathcal{T} \\ = - \int_{\mathbb{T}} (\mathcal{C}_{0,1,1} + \mathcal{G})^{-1} \varphi \end{aligned}
$$

By assumption H3, $(\int \mathcal{C}_{0,1,1}^{-1})$ is an invertible matrix and its inverse has norm is less than U. We have shown that the operator $\mathcal{T} \to \int (\mathcal{C}_{0,1,1} + \mathcal{G})^{-1} - \mathcal{C}_{0,1,1}^{-1} \mathcal{T}$ has norm less than $1/2T$. By the assumptions $U \cdot T \leq 1/2$, we can treat (79) as a perturbation of the first term. Since the RHS of (79) has norm less than $1/2T \|\varphi\|_{\rho''}$, we have that

$$
\begin{aligned} |\mathcal{T}| &\leq 2U(1/2)T\|\varphi\|_{\rho''} \\ &\leq \frac{1}{2}\|\varphi\|_{\rho''} \end{aligned}
$$

Hence $\|\mathcal{T} + \varphi\|_{\rho''} \leq \frac{3}{2}\|\varphi\|_{\rho''}$.

Applying Lemma 5 we obtain (68).

Therefore,

$$
(80) \qquad \begin{aligned} \|\tilde{u} - u\|_{\rho'} \leq \|\tilde{u} - u\|_{\rho''} &\leq \|h'w\|_{\rho''} \\ &\leq C\kappa^{-2}\sigma^{-2\tau}(N_+)^2\varepsilon \end{aligned}
$$

Applying Cauchy inequalities, we obtain

$$
(81) \qquad \begin{aligned} \|\tilde{h}' - h'\|_{\rho'} &\leq (\rho'' - \rho')^{-1}\|\tilde{u} - u\|_{\rho''} \\ &\leq C\kappa^{-2}\sigma^{-2\tau-1}(N_+)^2\varepsilon \end{aligned}
$$

The equations (80), (81) give us (69) for the solution of (50).

We note that the solution consists on applying identities to factor the equation (50) into several different steps. We have shown that the solution is in $\mathcal{A}_{\rho'}$ for every $\rho' < \rho$. Nevertheless it is important to realize that the solution is unique among the solutions in $\mathcal{A}_{\hat{\rho}}$ for any $0 < \hat{\rho}$. Hence, the solution of (50) is unique among the solutions in $\mathcal{A}_{\hat{\rho}}$ for any $0 < \hat{\rho}$. This establishes the uniqueness claim in Lemma 4.

4.2.1. *Proof of* (70). The proof consists in showing that $\tilde{u}$ is still in the domain of the functional E. Then, we just add and subtract appropriate terms and estimate what remains using the Taylor estimates.

Because of (69) and the assumption (67) we obtain that

$$
(82) \qquad ||u - \tilde{u}||_{\rho'} \leq \delta/4
$$

where δ denotes – see the assumptions in Theorem 1 – the distance of the range of u to the boundary of the domain of the interaction.

Therefore, we have

$$
\begin{aligned}
E[u+v] &= h'^{-1}(h'E[u] + h'E'[u]v) + (E[u+v] - E[u] - E'[u]v) \\
&= h'^{-1}vE[u]h' + (E[u+v] - E[u] - E'[u]v) \\
&= h'^{-1}(v\frac{d}{d\theta}E[u]) + (E[u+v] - E[u] - E'[u]v)
\end{aligned}
\tag{83}
$$

The first identity is just adding and subtracting. The second equation uses that v solves equation (50) and the third identity is just (48).

Using the Cauchy inequality, the Banach algebra property and the estimates for v obtained in (80), we have:

$$
\|(h')^{-1}(v\frac{d}{d\theta}E[u])\|_{\rho'} \leq C\kappa^{-2}\sigma^{-2\tau-1}(N_+)^2 N_- \varepsilon^2 \tag{84}
$$

The equation (82) tells us that we can apply the estimates in Lemma 1 to each of the terms defining the error E and we obtain:

$$
\begin{aligned}
\|E[u+v] - E[u] - E'[u]v\|_{\rho'} &\leq \frac{1}{2}\sum_L M_L L \|v\|^2_{\rho'} \\
&\leq C\left(\sum_L M_L L\right)\kappa^{-4}\sigma^{-4\tau}(N_+)^4\varepsilon^2
\end{aligned}
\tag{85}
$$

Adding (84) and (85) and using the obvious estimates $\sigma^{-4\tau} \leq \sigma^{-4\tau-1}$ we obtain (70).

4.2.2. *Proof of the estimates for the change in the induction hypothesis in Lemma 4.* We use the notation introduced in (46) and denoting by $\tilde{\gamma}_L$ the one corresponding to $\tilde{u}$ instead of u. Given the estimates (69) and (67), we have (82) that tells us that we can apply Lemma 1.

We first observe that

$$
\begin{aligned}
\operatorname{dist}(\tilde{\gamma}_L(\mathbb{T}^{\rho'}), &\mathbb{C} - \operatorname{Domain}(H_L)) \\
&\geq \operatorname{dist}\big(\gamma_L(\mathbb{T}^{\rho}), \mathbb{C} - \operatorname{Domain}(H_L)\big) - ||\gamma_L - \tilde{\gamma}_L||_{\rho'}
\end{aligned}
$$

Since $||\gamma_L - \tilde{\gamma}_L||_{\rho'} = ||u - \tilde{u}||_{\rho'}$, applying (69), we see that the new function $\tilde{u}$ satisfies assumption H2 with

$$
\tilde{\delta} = \delta - \varepsilon\kappa^{-2}(\rho - \rho')^{-2\tau}(N_+)^2\|E[u]\|_\rho
$$

If we do that, we see that the M_L do not need to be changed because they are the supremum of functions over an smaller set.

We also note that, by Cauchy estimates, we have that, by the mean value theorem,

$$
\begin{aligned}
||\partial_0\partial_1 H_1(u(\cdot), u(\cdot+\omega)) &- \partial_0\partial_1 H_1(\tilde{u}(\cdot), \tilde{u}(\cdot+\omega))||_{\rho'} \\
&\leq 2||u-\tilde{u}||_{\rho'} \\
&= 2M_1\Delta
\end{aligned}
$$

Using that $\partial_0\partial_1 H_1(u(\theta), u(\theta+\omega))$ is invertible for all θ, we obtain, using the Neumann series that if Δ is small enough, so is $\partial_0\partial_1 H_1(\tilde{u}(\cdot), \tilde{u}(\cdot+\omega))$ and we get the bounds claimed in (71).

Adding and subtracting, we also get

$$
\begin{aligned}
||\mathcal{C}_{0,1,1} - &\tilde{\mathcal{C}}_{0,1,1}||_{\rho'} \\
\leq& ||\partial_0\partial_1 H_1(u(\cdot), u(\cdot+\omega)) - \partial_0\partial_1 H_1(\tilde{u}(\cdot), \tilde{u}(\cdot+\omega))h'(\cdot)h'(\cdot+\omega)||_{\rho'} \\
&+ ||\partial_0\partial_1 H_1(u(\cdot), u(\cdot+\omega))(h'(\cdot) - \tilde{h}'(\cdot)h'(\cdot+\omega)||_{\rho'} \\
&+ ||\partial_0\partial_1 H_1(u(\cdot), u(\cdot+\omega))h'(\cdot)(h'(\cdot+\omega) - h'(\cdot+\omega)||_{\rho'} \\
\leq& 2M_1\Delta N_+^2 + M_1\Delta' N_+ M_1\Delta' N_+ \\
=& \Delta'(2M_1(N_+^2 + N_+)
\end{aligned}
$$

From this, we deduce immediately the claim in (71) using the Neumann series.

4.3. **Iteration of the inductive step and convergence to a solution.** The rest is very standard and can be done by invoking some of the implicit function theorems in the literature. For example, [Zeh75, Zeh76, Van02, dlLV00]. See [dlL01] for an exposition of different methods. However, since we have used some non-standard spaces it is perhaps clearer to run over the argument giving the convergence. We hope that this will make this paper more self-contained.

We consider a system which satisfies the hypothesis of Theorem 1. We label with a subindex n all the elements corresponding to the n iterative step.

We start with a function defined in a domain parametrized by ρ_0. We choose a sequence of parameters $\rho_n = \rho_{n-1} - \rho_0 2^{-n-1}$. We try the iterative step so that the n iterative step starts with a function u_n defined in a domain ρ_{n-1} and ends up with a function u_{n+1} defined in a domain of radius ρ_{n+1}.

We note that assuming that we can take J steps to which we can apply Lemma 4 and that, in all the steps, the non-degeneracy conditions are bounded uniformly, we have:

$$
\begin{aligned}
\varepsilon_J \leq & \kappa^{-2}\rho_0^{-\eta}2^{-J(2\tau+A)}K(N_+,N_-,T,\tilde{T})\varepsilon_{J-1}^2 \\
\leq & \kappa^{-2-2\cdot 2}\rho_0^{-\eta)(1+2)}2^{-(J+2(J-1))\eta)}K^{1+2}\varepsilon_{J-2}^{2^2} \\
\leq & \kappa^{-2-2\cdot 2-\cdots-2\cdot 2^J}\rho_0^{-\eta)(1+2+\cdots+2^J)} \\
& \qquad 2^{-(J+2(J-1)+\cdots+2^{J-1}\cdot 1)\eta}K^{1+2+\cdot+2^J}\varepsilon_0^{2^J} \\
& \leq (\kappa^{-4}\rho_0^{-2\eta}2^BK^2)^{2^J}
\end{aligned}
\tag{86}
$$

where $B = \sum_{j=0}^{\infty} j2^{-j}$.

We see that under the assumptions in Theorem 1, the term in parenthesis in the RHS of (86) is smaller than 1. Indeed, by making ε_0 small enough, we can make it as small as desired.

4.4. **Uniqueness of the solution.** The proof of uniqueness is based on uniqueness of the solution of (50). In the language of [Zeh75], the uniqueness shows that there is an approximate inverse. We give the easy details.

We assume that, besides u, there was another solution $\hat{u}$. We note that, since they are solutions, applying Lemma 1

$$
0 = E[\hat{u}] - E[u] = E'[u](\hat{u}-u) + R(u,\hat{u}-u) \tag{87}
$$

with $||R||_\rho \leq M||\hat{u}-u||_\rho^2$.

Now, denoting as before $h = \mathrm{Id} + u$ and recalling that

$$
(E'[u]h')(\theta) = \frac{d}{d\theta}E[u] = 0
$$

we can write the equation (87) as:

$$
h'(\theta)(E'[u](\hat{u}-u))(\theta) - (\hat{u}-u)(\theta)(E'[u]h')(\theta) = -h'R \tag{88}
$$

The left hand side of (88) is the equation we studied in Section 3.4 and the subsequent sections.

Noting that the factorization (58) of the LHS of (88) achieved in Section 3.5 shows that $\int h'R = 0$. Using the uniqueness statements for the solution, we conclude that for any $0 < \rho' < \rho$,

$$
||u-\hat{u}||_{\rho'} \leq C\kappa^{-2}(\rho-\rho')^{-\eta}K||u-\hat{u}||_\rho^2 \tag{89}
$$

where K is an algebraic expression which depends on the non-degeneracy constants.

If we apply repeatedly the argument above choosing a sequence of domains

$$
\rho_n = \rho_{n-1} - 2^{-n}\rho_0,
$$

we obtain proceeding as in (86)

$$(90) \qquad ||u-\hat{u}||_{\rho'} \le \left(C^2\kappa^{-4}K^2 2^B \rho_0^{-2\eta}||u-\hat{u}||_{\rho}\right)^{2^n}$$

with B the same fixed number before

If $C^2\kappa^{-4}K^2 2^B \rho_0^{-2\eta}||u-\hat{u}||_{\rho} < 1$, we conclude, by taking the limit as $n \to \infty$ in (90) that

$$||u-\hat{u}||_{\rho_0/2} = 0$$

which is the desired conclusion about uniqueness of the solution.

5. Lindstedt series for quasiperiodic solutions in Extended systems.

If our model has a small parameter μ measuring the distance to integrable, it is natural to try to solve the equations (13) perturbatively. That is, as formal power series in μ. These formal power series have been considered for Hamiltonian systems and are called the Lindstedt series. In this section, we want to discuss the existence of these series to all orders (see Lemma 7) and the convergence in some cases (see Lemma 2).

5.1. **Existence of Lindstedt series to all orders.** In the case of Hamiltonian systems, the existence of Lindstedt series to all orders is proved in [Poi99]. The proof presented here is similar to the proof in [FdlL92, dlL01] for the twist mapping case.

Lemma 7. *Consider the model given by*

$$(91) \qquad H_L(x_0,\ldots,x_L) = \frac{1}{2}A_L(x_0-x_L)^2 + \mu H_L^1(x_0,\ldots,x_L;\mu)$$

where all the H_L^1 are analytic in all the variables.

Assume that

H1 *The frequency ω satisfies*

$$(92) \qquad \left|\sum_L 2A_L(\cos(2\pi k\omega)-1)\right| \ge \kappa|k|^{-\tau} \quad \forall k \in \mathbb{Z}^d - \{0\}$$

H2 *The interactions H_L satisfy for some $\rho > 0$ and for all μ of sufficiently small modulus*

$$(93) \qquad \begin{aligned} \sum_L ||H_L^1||_{\rho} &\le R \\ \sum_L |A_L| &\le R \end{aligned}$$

where A is uniform in μ.

Then, there exist $u_\mu(\theta) = \sum_0^\infty \mu^i u^i(\theta)$ *a formal power series in* μ *with coefficients analytic in* θ *solving the equation* (20) *in the sense of formal power series.*

The solution is unique if we impose the normalization condition

$$\int_{\mathbb{T}} d\theta \; u(\theta;\mu) = 0 \tag{94}$$

Moreover, in the case that the interactions are finite range and trigonometric polynomials, the result above can be improved in two ways:

- *We do not need the assumption* (92). *It suffices that*

$$\left| \sum_L 2A_L(\cos(2\pi k\omega) - 1) \right| \neq 0 \quad \forall k \in \mathbb{Z}$$

- *The* u_n *are trigonometric polynomials of degree less than* Kn.

It is interesting to compare the conditions of Theorem 1 and Lemma 7. Of course, one important difference is that Theorem 1 is not formulated with reference to an integrable model, whereas Lemma 7 is. Nevertheless, for the common ground of nearly integrable systems, they present differences. In particular, the Diophantine conditions are different.

Note that the non-degeneracy conditions of Theorem 1 are implicit in the choice of the unperturbed model. We also note that the condition on ω in Theorem 1 is the standard Diophantine condition whereas the Diophantine condition in Lemma 7 is the condition (92). The later condition depends on the coefficients of the integrable part of the interaction. We will explore the relation between the Diophantine coefficients of both results later. In Lemma 8 we will find some conditions on the coefficients A_L that ensure that the sets of ω satisfying the conditions of Lemma 7 are of full measure. Nevertheless the conditions (92) stronger than those in Theorem 1.

The most interesting difference among the conditions of Theorem 1 and Lemma 7 comes from the fact that Theorem 1 requires much weaker decay properties in the interaction than those in the main KAM result, Theorem 1.

Of course, one can wonder whether the Lindstedt series whose existence is asserted in Lemma 7 converge or not. In Lemma 2 we will present a result that asserts the convergence of the series provided that the model satisfies the hypothesis of Theorem 1 as well as the hypothesis of Lemma 7. The argument is an indirect one based on the KAM theorem following an argument of [Mos67].

Since some of the hypothesis of Lemma 7 are weaker than those of Theorem 1 we think it would be interesting to see if there is a direct proof of convergence of the Lindstedt series, specially if such a proof can deal with decay conditions such as those considered by the KAM approach Theorem 1. See [Eli96, CF94, Gal94, GG95] for proofs of some standard KAM theorems using the method of compensations.

Proof. In our case, the equation (20) reads

$$
\begin{aligned}
0 = &\sum_{L=1}^{\infty} +A_L(u(\theta+L\omega)+u(\theta-L\omega)-2u(\theta)) \\
&+\mu \sum_{L}\sum_{k+j=i} \partial_j H_L\Big(\theta + k\omega + u(\theta+k\omega;\mu), \ldots \\
&\quad , \theta(k+j)\omega + u(\theta+(k+j)\omega;\mu), \ldots \\
&\quad , \theta+(k+L)\omega + u(\theta+(k+L)\omega;\mu);\mu\Big)
\end{aligned}
\tag{95}
$$

We denote

$$[\Gamma u](\theta) = \sum_{L=1}^{\infty} A_L(u(\theta+L\omega)+u(\theta-L\omega)-2u(\theta))$$

We write equation (95) as

$$\Gamma u_\mu = \mathcal{N}(u_\mu)$$

where $\mathcal{N}(u)$ is defined as the other terms in (95).

Proposition 2. *In the conditions* (93)*, given a function*

$$u : \mathbb{T}^{\tilde{\rho}} \times \{\mu in \mathbb{C} | |\mu| < \beta\} \to \mathbb{C}$$

such that

- *For each fixed* θ*,* $u(\theta;\mu)$ *is polynomial in* μ*.*
- $u(\theta;0) = 0$*.*

Then, the sums

$$\sum_L H_L \gamma_\mu^L(\theta)$$

converge uniformly in $\theta \in \mathbb{T}^{\tilde{\rho}} \times \{\mu \in \mathbb{C} | |\mu| < a_1$ *for some* $a_1 > 0$*.*

The proof of Proposition2 is just the observation that for small enough μ then $u(\theta,\mu)$ is in the domain of definition of the all the H^L and, therefore the sup norm the H_L controls the composition. □

By the assumption (92), we see that equations for φ given η with $\eta(\theta)$ satisfying

$$\Gamma\varphi = \eta \tag{96}$$

have an analytic solution if and only if $\int \eta = 0$. Moreover, the solution is unique up to an additive constant. In particular, it is unique if we add the normalization $\int \varphi = 0$. The solution of (96) is given, in Fourier coefficients by

$$\varphi_k = \eta_k / \Big(\sum_L 2A_L(\cos(2\pi k\omega) - 1) \Big)$$

Hence, if η is an analytic function and ω satisfies the inequalities (92), we obtain that φ is analytic. If η is a trigonometric polynomial, provided that the RHS of (92) does not vanish, we obtain that there is trigonometric polynomial φ solving the equations (96).

Equating the terms independent of μ in (95), we obtain

$$\Gamma u^0 = 0$$

Which shows that u^0 is a constant. If we impose the normalization (94)

Equating terms of order μ^n, $n \geq 1$ on both sides of (95) we obtain

$$\Gamma u_n = \eta_n(u^0, u^1, \ldots, u^{n-1}) \tag{97}$$

where η_n is a polynomial whose coefficients depend on the derivatives of H_L up to order $n-1$. Note that, by the assumption H2, using Proposition 96, we conclude that if $u^0, \ldots, u^{n-1}$ are analytic functions on some strip so is η_n, therefore, we can formulate the equations for u^{n+1}. Our next task is that the equations can be solved and that the solution u^{n+1} is analytic in $\mathbb{T}^{\rho'}$ for any $\rho' < \tilde{\rho}$.

Note that if the interactions are finite range and all of the terms are trigonometric polynomials, it follows by induction that η_n is a trigonometric polynomial and that therefore, the u_n is a trigonometric polynomial.

Using the discussion of the solutions (96), we see that if we know $u^0, \ldots, u^{n-1}$ we can find u_n provided that $\int \eta_n = 0$ and that this u_n is unique under the normalization (94), which implies $\int u^n = 0$. Hence, Lemma 7 is proved once we show that we have

$$\int_{\mathbb{T}} d\theta \quad \eta_n(u^0(\theta), u^1(\theta), \ldots, u^{n-1}(\theta)) = 0 \tag{98}$$

To establish this, we now assume that we have $u^{\leq n}(\theta) = \sum_{i=0}^n \mu^i u^i(\theta)$ satisfying (95) up to errors of order μ^{n+1}.

We have, therefore that

$$\Gamma u^{[\leq n]} - \mathcal{N}(u^{[\leq n]}) = \mu^{n+1}\eta_n + \mu^{n+2} S_n \tag{99}$$

By (3), we have

$$
\begin{aligned}
0 &= \int (h^{\le n})' \big[\Gamma u^{[\le n]} - \mathcal{N}(u^{[\le n]})\big] \\
&= \mu^{n+1} \int \eta_n (h^{[\le n]})' + O(\mu^{n+2})
\end{aligned}
\tag{100}
$$

Taking into account that $(h^{[\le n]})' = \mathrm{Id} + O(\mu)$, we obtain that the RHS of (100) is $\mu^{n+1} \int \eta_n + O(\mu^{n+2})$ therefore, we obtain, equating the terms of μ^{n+1} in (100), we obtain $\int \eta_n = 0$ as desired. □

5.2. **On the Diophantine conditions** (92)**.** The conditions (92) are different from the standard Diophantine conditions. In this section, we study these conditions. Our first result is a result on the abundance of numbers satisfying (92).

Definition 1. *We say that a function F of period* 1 *is non degenerate when*

1. *It has a finite number of zeros in* $[0, 1)$*.*
2. *All the zeros are of finite order. That is, if $F(z_0) = 0$, there exists $N \in \mathbb{Z}$ such that $F^{(N)}(z_0) \neq 0$.*

Lemma 8. *With the notations of Lemma 7. Assume that the function*

$$
F(z) = \sum_{L=1} 2A_L(\cos(2\pi z) - 1)
$$

is a non-degenerate function in the sense of Definition 1.

Then, the set of ω satisfying (92) *is of full measure.*

Note that that the simplest case of the Lemma above is the case when $F(z)$ is a C^2 Morse function. This condition is generic among the set of coefficients $|A_j| \le Cj^{-(3+\delta)}$ endowed with the product topology. When the interaction is finite range, only a finite number of A_L are zero, so that the function $F(z)$ is a trigonometric polynomial which satisfies the assumption of Lemma 8.

As a corollary of Lemma 8, we obtain that, under the conditions of the Lemma – in particular, for finite range interactions – there is a full measure set of ω satisfying both (39) and (92). On the other hand, for any Diophantine ω, it is possible to choose A_L so that the denominators in (92) become zero (or grow very fast).

Note that the functions $F(z)$ always satisfy $F(0) = 0$. The conditions (92) are, roughly, that the values of $\omega k - n$ are not too close to the zeros of F. Hence, in particular, all the numbers that satisfy Definition 92 are Diophantine. On the other hand, it is not hard to produce numbers that are Diophantine but do not satisfy Definition 92.

Proof. By the assumption on F, since F has a finite number of critical points, it has a finite number of intervals of monotonicity. Therefore, given an interval I, $F^{-1}(I)$ is the union of a number of intervals smaller than the number of intervals of monotonicity of F. We have bounds for the length

$$|F^{-1}(I)| \leq A|I|^{\alpha}$$

Therefore, the sets

$$\mathcal{C}_{\kappa,\tau,k} = \{\omega||F(k\cdot\omega)| \leq \kappa|k|^{-\tau}\}$$

are the union of a finite number of regions bounded by planes.

If $B_r \subset \mathbb{R}$ is a ball of radius r we can estimate the measure

$$|\mathcal{C}_{\kappa,\tau,k} \cap B_r| \leq C\kappa^{\alpha}k^{-(\tau+1)\alpha}$$

We observe that if $(\tau+1)\alpha > 1$, we have

$$|\cup_{k\in\mathbb{Z}-\{0\}} \mathcal{C}_{\kappa,\tau,k} \cap B_r| \leq C\kappa^{\alpha}$$

This shows that, if we take κ small enough, we can make the measure of the sets where the inequality (92) fails as small as desired. Note that, as a consequence of the proof, if we take $\kappa \leq Cr^{1/\alpha}$, we can ensure that there is a point satisfying (92) in any ball of radius α. $\square$

5.3. **Convergence of the Lindstedt series.** The argument is based in an argument due to [Mos67], but it is simpler in this case. The basic idea is that we can apply a rapidly convergent quadratic procedure in a space of functions which are analytic in μ.

We note that an interaction of the form (91), if we take $u = 0$, the system satisfies the non-degeneracy assumptions of Theorem 1. Moreover, taking μ small enough, we obtain that the smallness assumption are satisfied. Of course, to apply Theorem 1 we need that the interactions decrease fast enough.

Theorem 2. *Consider a system of the form* (91) *satisfying the hypothesis of Lemma 7 Assume furthermore that for small μ, the system satisfies uniformly the hypothesis of Theorem 1 and that the frequency ω is Diophantine.*

Then, the Lindstedt series produced in Lemma 7 converges in a neighborhood of zero.

Note that this theorem requires both the Diophantine conditions of the Lemma 7 and the rapid decrease conditions for the interaction on Theorem 1. We think that it would be quite interesting to study whether there is a proof of convergence of the Lindstedt series without using fast decay properties of the interaction. This would lead to a proof of existence of smooth quasi-periodic solutions in situations not covered by Theorem 1.

Proof. Following [Mos67], we consider a space of functions which are jointly analytic in μ, θ and consider the original problem of finding solutions of (17) as a functional analysis problem in the space of analytic functions of the variables μ, θ.

We claim that there is Theorem analogous to Theorem 1 for functions depending analytically on extra parameters.

We just indicate the argument. Some more details on how to lift KAM theorems from functions to families of functions can be found in [dlLO00, Van02, dlLV00]. The paper [dlLO00] lifts the main theorem from [Zeh75] to spaces of families obtain KAM theorems for some very degenerate system.

We consider spaces of functions $\mathcal{A}_{r,\rho}$ whose domain is the set $|\mu| \leq r$, $\theta \in \mathbb{T}^\rho$.

We will assume that the functions are continuous in the whole domain and analytic in the interior. We endow these spaces with the norm of the supremum.

Our goal is to show an analogue of Theorem 1 for families depending on parameters. That is, we will assume that we have families of interactions depending analytically on parameters and satisfying the non-degeneracy assumption uniformly for the parameters. Then, if we have a analytic family that satisfies the equilibrium equations with enough accuracy, then there is another family which satisfies the equilibrium equation exactly. Moreover we also have the same estimates for the distance between the approximate solution and the true one.

We note that the Procedure 1 can be implemented in spaces of functions depending on the parameter μ. Provided that the estimates assumed in Theorem 1 hold uniformly, we observe that for each value of the parameter μ we obtain uniform estimates of the same form as those we had before.

We can also check that if the family $u(\theta; \mu$ depends analytically in the parameter μ, then the improved solution also depends analytically in μ. This is obvious from the fact that the procedure to find $\tilde{u}$ consists only in composing the function with the interaction terms, applying some elementary perturbation argument and then, solving some small divisors whose coefficients do not depend on μ.

Therefore, we conclude that the new solution belongs to $\mathcal{A}_{r,\tilde{\rho}}$ and that the new error satisfies the same estimates.

The rest of the argument for convergence does not need any change. As was emphasized in Theorem 1, we obtain that the size of ε which is allowed in the argument, only depends on the size of the non-degeneracy conditions.

Once we have that the result Theorem 1 can be adapted to families, we see that we can take as initial guess just $u(\theta, \mu) = 0$. Then, we obtain that there is a solution which is analytic.

By the uniqueness of the Lindstedt series, we obtain that the Lindstedt series coincides with the Taylor expansion of the function. $\square$

6. A normal hyperbolicity approach for quasi-integrable systems with finite range

In this section we study the normal hyperbolicity properties of the quasi-integrable systems whose interaction terms are in (91).

We note that the equilibrium equations are written explicitly in (15). In order to be able to use a a dynamical description, we will assume that the interaction has finite range. That is, we will assume that $A_j = 0$ for $j \geq J$.

For $\mu = 0$, the equilibrium equations are linear. They are:

(101)
$$(x_{n+1}+x_{n-1}-2x_n)+A_1(x_{n+2}+x_{n-2}-2x_n)+\cdots+A_{J-1}(x_{n+J}+x_{n-J}-2x_n) = 0$$

We can write (101) as a dynamical system in $\mathbb{R}^{2J}$. Denoting

$$\mathbf{x}_n = (x_{n+J}, x_{n+(J-1))}, \dots, x_{n-J+1}, x_{n-J})$$

The equilibrium equations can be written as a linear system

$$\mathbf{x}_{n+1} = \mathcal{M}\mathbf{x}_n$$

where

$$\begin{aligned}
\mathcal{M}\mathbf{x} =& \Big(\frac{1}{A_{J-1}}(x_{n+1} + x_{n-1} - 2x_n) \\
&+ \frac{A_1}{A_{J-1}}(x_{n+2} + x_{n-2} - 2x_n) \\
&+ \cdots + \\
&+ \frac{A_{J-2}}{A_{J-1}}(x_{n+(J-2)} + x_{n-(J-2)} - 2x_n) \\
&- x_{n-J} + 2x_n \\
&\quad , x_{n+J-1}, x_{N+J-2}, \dots, x_{n-J+1}\Big)
\end{aligned} \tag{102}$$

It is not hard to compute the characteristic equation for the operator $\mathcal{M}$. It suffices to try solutions of the form $x_n = \lambda^n$ in (101). The characteristic equation is:

(103)

$$
\begin{aligned}
0 &= \lambda^{J-1}(\lambda-1)^2 + A_1\lambda^{J-2}(\lambda^2-1)^2 + \cdots + A_{J-1}(\lambda^J-1)^2\\
&= (\lambda-1)^2\big[\lambda^{J-1} + A_1\lambda^{J-2}(\lambda+1) + \cdots + A_{J-1}(\lambda^{J-1}+\lambda^{J-2}+\cdots\lambda_1)\big]
\end{aligned}
$$

We note that the characteristic polynomial $P(\lambda)$ in (103) satisfies $P(1/lambda) = P(\lambda)\lambda^{2J}$. Even if we have not consider the Hamiltonian properties of the problem, we recall that the property above is true for all the Hamiltonian systems. One implication is that if λ_0 is a root of the characteristic polynomial, then $1/\lambda_0$ is also a root.

From the factorization in (103), we conclude has always a double root 1. Note also that the A coefficients are factors of the highest degree terms in the characteristic equation. We note that for the case $A = 0$ it has 0 as a root of multiplicity $J-1$. We conclude that if all the A's are small there will be $J-1$ roots (counted with multiplicity) near zero. Because of the symmetry of the polynomial, we conclude that their inverses will also be roots. Hence, for $A_J > 0$ and all the roots small, there will be two eigenvalues exactly 1, $J-1$ eigenvalues of small modulus and $J-1$ of large modulus.

The space corresponding to the eigenvalues 1 will be invariant and will be normally hyperbolic.

If we now add a non-integrable perturbation of the system, applying the theory of persistence of normally hyperbolic manifolds [Fen72, Fen74, HPS77], we conclude that there will be a two dimensional invariant manifold.

It is natural to conjecture that the system restricted to the invariant manifold is symplectic. This would explain the persistence of the one-frequency systems.

Of course, this approach does not give a clue of what can happen for systems far from integrable. Since center manifolds could fail to be even C^∞ it is not clear that this approach can produce straightforwardly the analytic results presented here. On the other hand, it seems that this normal hyperbolicity approach can produce results in the case that the x_i are multidimensional variables.

The reason for the conjecture that the system is symplectic restricted to the center manifold is that one can imagine that there is also a variational principle for the restriction. Indeed, in Physics one often finds arguments about *effective theory* [Ami84]. Perhaps the invariant manifold theorems could through some light on these problems. The effective lagrangian would be the lagrangin restricted to the invariant manifold.

We think that these question deserves further attention and it would be interesting to make precise these possibilities. This would require, in particular to make precise the theory of invariant manifolds in the Lagrangian formalism for systems of the type we have considered.

7. Some final remarks

7.1. **Numerical implementation.** Note that Procedure 1 is an algorithm that can be readily implemented as a numerical algorithm.

An efficient way of carrying out the computation is to keep the function u discretized as a Fourier series. Denote by N the number of Fourier coefficients kept.

The step 1 is diagonal in Fourier series so, it has a cost $O(N)$.

The operator $\mathcal{G}$ that is needed for step 2 is just a combination of translation and evaluation of derivatives. We note that, in Fourier series, translating or differentiating Fourier series is just an $O(N)$ operation.

If we use the Fast Fourier Transform algorithm (henceforth FFT), we can compute the products needed in $O(N \log(N))$. Again, the computation of $\mathcal{C}_{0,1,1} + \mathcal{G})^{-1}$ is diagonal on real space, so that if we use the FFT (cost $O(N \log(N))$), then the computation is just $O(N)$.

The computation of $\mathcal{T}$ is just division by a number. Actually, for the next step, the only thing that we need is to project $(\mathcal{C}_{0,1,1} + \mathcal{G})^{-1}(\varphi)$, which in Fourier series can be accomplished by setting to zero the zero'th order coefficient.

Then, the computation of w is $O(N)$ in Fourier coefficients.

The computation of $\tilde{u}$ is $O(N)$ in real space (some of the computations before give us a computation of h', which is diagonal in Fourier space anyway).

In summary, the steps of Procedure 1 are diagonal either in real space or in Fourier space. These diagonal operations cost $O(N)$ operations. The FFT needed to switch from real space to Fourier representation have a cost of $O(N \log(N))$.

Therefore, Procedure 1 can be implemented in $O(N \log(N))$ operations.

Preliminary implementations in [CdlL05] indicate that indeed it is possible to implement this algorithm quite efficiently.

We also note that the existence of a variational principle (22) can also be taken advantage using minimization methods. In the case that the system is ferromagnetic, it is not too hard to show that all the invariant circles are minimizers [MF94b]. Hence, in the case that the system is ferromagnetic, one can combine the good global properties of the variational method till one gets close enough to a solution so that one can use the Newton method. For the case of twist mappings this is the method that was used in [Per74, Per79].

7.2. **Some systems that are very close to integrable.** In the following we remark that there are some cases of the models (15) which can be made to be very close to integrable. These systems seem to have very remarkable statistical mechanics properties since for them, the waves move very easily. We think that this deserves further exploration.

It is amusing to note that (6) can be considered as a multistep discretization of the equation for the pendulum with step $\mu^{1/2}$. This is a method which is accurate up to errors $(\mu^{1/2})^4 = \mu^2$. Note that the pendulum equation, being a one degree of freedom Hamiltonian system has a conserved quantity and is integrable in open sets that do not intersect a level set of critical values of the energy – the separatrices. Hence, (6) can be considered as an integrable system up to errors μ^2.

If we choose $A = -2^{-4}$ in (13) we obtain a discretization of the pendulum which is accurate up to errors ε^3 and, therefore very close to integrable.

The calculation is standard in multistep methods. We set $x_n = x(\mu^{1/2} n\alpha)$ and we denote $x'_n = x'(\mu^{1/2}\alpha)$ and similarly for higher derivatives. We have:

$$
\begin{aligned}
x_{n+1} + x_{n-1} - 2x_n &= x''_n \mu\alpha^2 + \frac{2}{4!} x_n^{(4)} \mu^2 \alpha^4 + O(\mu^3) \\
x_{n+2} + x_{n-2} - 2x_n &= x''_n \mu\alpha^2 + \frac{2}{4!} 2^4 x_n^{(4)} \mu^2 \alpha^4 + O(\mu^3)
\end{aligned}
$$

Therefore

$$
(x_{n+1} + x_{n-1} - 2x_n) - 2^{-4}(x_{n+2} + x_{n-2} - 2x_n) = (1 + 2^{-4})\alpha^2 x''_n \mu + O(\mu^3)
$$

Choosing $\alpha = (1 + 2^{-4})^{-1/2}$ – this is just a change in time, we obtain that the equilibrium equation matches the pendulum equation to a higher the required order.

The calculations to higher order are also straightforward. If we carry out to higher order the expansions of $x_{n+j} + x_{n-j} - 2x_n$, multiply by A_j and add them, for $j = 2, \dots M$, the conditions that the coefficients for the even derivatives higher than the second vanish are just

$$
\begin{aligned}
&1 + 2^4 A_2 + 3^4 A_3 + \cdots M^4 A_M = 0 \\
&1 + 2^6 A_2 + 3^6 A_3 + \cdots M^6 A_M = 0 \\
&\cdots \\
&1 + 2^{M-1} A_2 + 3^{M-1} A_3 + \cdots M^{M-1} A_M = 0
\end{aligned}
$$

The above set of equations can be solved explicitly since they are the well known Vandermonde determinants. Hence, by choosing

$$A_j = -\prod_{k=2}^{j}(1-k^{-2})^{-1}(1-(j+1)^2k^{-2}),$$

in (15), we obtain systems that are integrable up to a order μ^M.

Acknowledgments

I thank R. Calleja, A. González-Enríquez and J. Vano for discussions on the subject. A. González-Enríquez and J. Vano supplied their preprint [GEV05] before publication, which suggested that a finite differentiable version of the result should be possible. A. González-Enríquez gave a very careful reading to a preliminary version of the manuscript, which improved the presentation and removed several mistakes.

The work of the author has been supported by NSF grants.

References

[ALD83] S. Aubry and P. Y. Le Daeron. The discrete Frenkel-Kontorova model and its extensions. I. Exact results for the ground-states. *Phys. D*, 8(3):381–422, 1983.

[Ami84] Daniel J. Amit. *Field theory, the renormalization group, and critical phenomena*. World Scientific Publishing Co., Singapore, second edition, 1984.

[Ang90] Sigurd B. Angenent. Monotone recurrence relations, their Birkhoff orbits and topological entropy. *Ergodic Theory Dynam. Systems*, 10(1):15–41, 1990.

[BGK99] J. Bricmont, K. Gawędzki, and A. Kupiainen. KAM theorem and quantum field theory. *Comm. Math. Phys.*, 201(3):699–727, 1999.

[CC95] A. Celletti and L. Chierchia. A constructive theory of Lagrangian tori and computer-assisted applications. In *Dynamics Reported*, pages 60–129. Springer, Berlin, 1995.

[CC97] Alessandra Celletti and Luigi Chierchia. On the stability of realistic three-body problems. *Comm. Math. Phys.*, 186(2):413–449, 1997.

[CdlL98] A. Candel and R. de la Llave. On the Aubry-Mather theory in statistical mechanics. *Comm. Math. Phys.*, 192(3):649–669, 1998.

[CdlL05] Renato Calleja and Rafael de la Llave. Computation of periodic, quasiperiodic and homoclinic orbits in systems with long range interactions. 2005. Manuscript.

[CE78] Pierre Collet and Jean-Pierre Eckmann. *A renormalization group analysis of the hierarchical model in statistical mechanics*. Springer-Verlag, Berlin, 1978. Lecture Notes in Physics, Vol. 74.

[CF94] L. Chierchia and C. Falcolini. A direct proof of a theorem by Kolmogorov in Hamiltonian systems. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 21(4):541–593, 1994.

[DdlL00] Amadeu Delshams and Rafael de la Llave. KAM theory and a partial justification of Greene's criterion for nontwist maps. *SIAM J. Math. Anal.*, 31(6):1235–1269 (electronic), 2000.

[dlL00] Rafael de la Llave. Variational methods for quasi-periodic solutions of partial differential equations. In *Hamiltonian systems and celestial mechanics (Pátzcuaro, 1998)*, volume 6 of *World Sci. Monogr. Ser. Math.*, pages 214–228. World Sci. Publishing, River Edge, NJ, 2000.

[dlL01] Rafael de la Llave. A tutorial on KAM theory. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, pages 175–292. Amer. Math. Soc., Providence, RI, 2001. Revised version available from `ftp.ma.utexas.edu/pub/papers/llave/tutorial.pdf`.

[dlL05] R. de la Llave. Improved estimates on measure occupied and the regularity of invariant tori in hamiltonian systems by zehnder's method. *Preprint*, 2005.

[dlLGJV05] R. de la Llave, A. González, À. Jorba, and J. Villanueva. KAM theory without action-angle variables. *Nonlinearity*, 18(2):855–895, 2005.

[dlLO00] R. de la Llave and R. Obaya. Decomposition theorems for groups of diffeomorphisms in the sphere. *Trans. Amer. Math. Soc.*, 352(3):1005–1020, 2000.

[dlLR90] R. de la Llave and David Rana. Accurate strategies for small divisor problems. *Bull. Amer. Math. Soc. (N.S.)*, 22(1):85–90, 1990.

[dlLR91] R. de la Llave and D. Rana. Accurate strategies for K.A.M. bounds and their implementation. In *Computer Aided Proofs in Analysis (Cincinnati, OH, 1989)*, pages 127–146. Springer, New York, 1991.

[dlLV00] R. de la Llave and J. Vano. A Whitney-Zehnder implicit function theorem. *Manuscript*, 2000.

[Dys71] Freeman J. Dyson. An Ising ferromagnet with discontinuous long-range order. *Comm. Math. Phys.*, 21:269–283, 1971.

[Eli96] L. H. Eliasson. Absolutely convergent series expansions for quasi periodic motions. *Math. Phys. Electron. J.*, 2:Paper 4, 33 pp. (electronic), 1996.

[FdlL92] Corrado Falcolini and Rafael de la Llave. A rigorous partial justification of Greene's criterion. *J. Statist. Phys.*, 67(3-4):609–643, 1992.

[Fen72] Neil Fenichel. Persistence and smoothness of invariant manifolds for flows. *Indiana Univ. Math. J.*, 21:193–226, 1971/1972.

[Fen74] N. Fenichel. Asymptotic stability with rate conditions. *Indiana Univ. Math. J.*, 23:1109–1137, 1973/74.

[Gal94] G. Gallavotti. Twistless KAM tori, quasi flat homoclinic intersections, and other cancellations in the perturbation series of certain completely integrable Hamiltonian systems. A review. *Rev. Math. Phys.*, 6(3):343–411, 1994.

[GEHdlL05] Alejandra González-Enríquez, Àlex Haro, and Rafael de la Llave. KAM theory with parameters and applications to degenerate systems. 2005. Manuscript.

[GEV05] A. Gonzalez-Enriquez and J. Vano. An estimate on smoothing and compostion with applications to KAM theory. Preprint, 2005.

[GG95] G. Gallavotti and G. Gentile. Majorant series convergence for twistless KAM tori. *Ergodic Theory Dynam. Systems*, 15(5):857–869, 1995.

[Gol01] Christophe Golé. *Symplectic twist maps*, volume 18 of *Advanced Series in Nonlinear Dynamics*. World Scientific Publishing Co. Inc., River Edge, NJ, 2001. Global variational techniques.

[HPS77] M.W. Hirsch, C.C. Pugh, and M. Shub. *Invariant manifolds*. Springer-Verlag, Berlin, 1977. Lecture Notes in Mathematics, Vol. 583.

[JdlLZ99] À. Jorba, R. de la Llave, and M. Zou. Lindstedt series for lower-dimensional tori. In *Hamiltonian Systems with Three or More Degrees of Freedom (S'Agaró, 1995)*, pages 151–167. Kluwer Acad. Publ., Dordrecht, 1999.

[KdlLR97] Hans Koch, Rafael de la Llave, and Charles Radin. Aubry-Mather theory for functions on lattices. *Discrete Contin. Dynam. Systems*, 3(1):135–151, 1997.

[Koz83] S. M. Kozlov. Reducibility of quasiperiodic differential operators and averaging. *Trudy Moskov. Mat. Obshch.*, 46:99–123, 1983. English translation: *Trans. Moscow Math. Soc.*, Issue 2:101–126, 1984.

[LC00] Patrice Le Calvez. *Dynamical properties of diffeomorphisms of the annulus and of the torus*, volume 4 of *SMF/AMS Texts and Monographs*. American Mathematical Society, Providence, RI, 2000. Translated from the 1991 French original by Philippe Mazaud.

[LM01] M. Levi and J. Moser. A Lagrangian proof of the invariant curve theorem for twist mappings. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, volume 69 of *Proc. Sympos. Pure Math.*, pages 733–746. Amer. Math. Soc., Providence, RI, 2001.

[Mat82] John N. Mather. Existence of quasiperiodic orbits for twist homeomorphisms of the annulus. *Topology*, 21(4):457–467, 1982.

[MF94a] J. N. Mather and G. Forni. Action minimizing orbits in Hamiltonian systems. In *Transition to Chaos in Classical and Quantum Mechanics (Montecatini Terme, 1991)*, pages 92–186. Springer, Berlin, 1994.

[MF94b] John N. Mather and Giovanni Forni. Action minimizing orbits in Hamiltonian systems. In *Transition to chaos in classical and quantum mechanics (Montecatini Terme, 1991)*, pages 92–186. Springer, Berlin, 1994.

[Mos66a] J. Moser. A rapidly convergent iteration method and non-linear differential equations. II. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:499–535, 1966.

[Mos66b] J. Moser. A rapidly convergent iteration method and non-linear partial differential equations. I. *Ann. Scuola Norm. Sup. Pisa (3)*, 20:265–315, 1966.

[Mos67] J. Moser. Convergent series expansions for quasi-periodic motions. *Math. Ann.*, 169:136–176, 1967.

[Mos73] J. Moser. *Stable and Random Motions in Dynamical Systems*. Princeton University Press, Princeton, N. J., 1973.

[Mos88] Jürgen Moser. A stability theorem for minimal foliations on a torus. *Ergodic Theory Dynamical Systems*, 8*(Charles Conley Memorial Issue):251–281, 1988.

[Per74] I. C. Percival. Variational principles for the invariant toroids of classical dynamics. *J. Phys. A*, 7:794–802, 1974.

[Per79] I. C. Percival. A variational principle for invariant tori of fixed frequency. *J. Phys. A*, 12(3):L57–L60, 1979.

[Poi99] H. Poincaré. *Les méthodes nouvelles de la mécanique céleste*, volume 1, 2, 3. Gauthier-Villars, Paris, 1892–1899.

[Ran87] D. Rana. *Proof of Accurate Upper and Lower Bounds to Stability Domains in Small Denominator Problems*. PhD thesis, Princeton University, 1987.

[Rue69] David Ruelle. *Statistical mechanics: Rigorous results*. W. A. Benjamin, Inc., New York-Amsterdam, 1969.

[Rüs75] H. Rüssmann. On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus. In *Dynamical Systems, Theory and Applications (Battelle Rencontres, Seattle, Wash., 1974)*, pages 598–624. Lecture Notes in Phys., Vol. 38, Berlin, 1975. Springer.

[Rüs76] H. Rüssmann. Note on sums containing small divisors. *Comm. Pure Appl. Math.*, 29(6):755–758, 1976.

[SY98] Wenxian Shen and Yingfei Yi. Almost automorphic and almost periodic dynamics in skew-product semiflows. *Mem. Amer. Math. Soc.*, 136(647):x+93, 1998.

[SZ89] D. Salamon and E. Zehnder. KAM theory in configuration space. *Comment. Math. Helv.*, 64(1):84–132, 1989.

[Van02] John A. Vano. *A Nash-Moser Implicit Function Theorem with Whitney Regularity and Applications*. PhD thesis, Univ. of Texas at Austin, 2002. MP_ARC #02-276.

[Zeh75] E. Zehnder. Generalized implicit function theorems with applications to some small divisor problems. I. *Comm. Pure Appl. Math.*, 28:91–140, 1975.

[Zeh76] E. Zehnder. Generalized implicit function theorems with applications to some small divisor problems. II. *Comm. Pure Appl. Math.*, 29(1):49–111, 1976.

Department of Mathematics, University of Texas, 1 University Station C1200, Austin, TX 78712-0257

E-mail address: llave@math.utexas.edu