A KAM theory for conformally symplectic systems: Efficient algorithms and their validation ☆

Renato C. Calleja [a,*], Alessandra Celletti [b], Rafael de la Llave [c]

[a] *Institute for Mathematics and Its Applications, University of Minnesota, 307 Church St SE, Minneapolis, MN 55455, United States*
[b] *Department of Mathematics, University of Roma Tor Vergata, Via della Ricerca Scientifica 1, 00133 Roma, Italy*
[c] *School of Mathematics, Georgia Institute of Technology, 686 Cherry St, Atlanta, GA 30332-1160, United States*

ARTICLE INFO

ABSTRACT

We present a KAM theory for some dissipative systems (geometrically, these are conformally symplectic systems, i.e. systems that transform a symplectic form into a multiple of itself). For systems with n degrees of freedom depending on n parameters we show that it is possible to find solutions with a fixed n-dimensional (Diophantine) frequency by adjusting the parameters.

We do not assume that the system is close to integrable, but we present the results in an *a-posteriori* format. Our unknowns are a parameterization of the quasi-periodic solution and some parameters in the system. We formulate an invariance equation that expresses that the system with the parameters leaves invariant the solution given by the embedding. We show that if there is a sufficiently approximate solution of the invariance equation, which also satisfies some non-degeneracy conditions, then there is a true solution nearby. The smallness assumptions above can be understood either in Sobolev or in analytic norms.

The a-posteriori format has several consequences: A) smooth dependence on the parameters, including the singular limit of zero dissipation; B) estimates on the measure of parameters covered by quasi-periodic solutions; C) convergence of perturbative expansions in dissipative analytic systems; D) bootstrap of regularity (i.e. that all tori which are smooth enough are analytic if the map is

---

* Corresponding author.
E-mail addresses: calleja@ima.umn.edu (R.C. Calleja), celletti@mat.uniroma2.it (A. Celletti), rafael.delallave@math.gatech.edu (R. de la Llave).

analytic); E) a numerically efficient criterion for the breakdown of the quasi-periodic solutions.

The proof is based on an iterative quadratically convergent method. The iterative step takes advantage of some geometric identities; these identities also lead to an efficient algorithm. If we discretize the parameterization with N terms, a modified Newton step requires $O(N)$ storage and $O(N \log(N))$ operations. The a-posteriori theorems allow one to be confident on the numerical results even very close to breakdown. The algorithm does not require that the system is close to integrable, so that a continuation algorithm is guaranteed to continue the tori till the breakdown (in practice, only limitation being the total memory and the precision). The algorithms have been implemented and run in other papers.

Contents

1. Introduction

Kolmogorov–Arnol'd–Moser (hereafter KAM) theory represented a breakthrough in the theory of the stability of nearly-integrable systems [55,5,67]. Under very general assumptions, KAM theory yields the persistence of quasi-periodic tori with Diophantine frequencies for the perturbed system, provided the perturbing parameter satisfies smallness conditions. The theorems of persistence are different when the systems considered have geometric properties (e.g., symplectic, volume preserving, reversible, which have been studied in classical works, see [70,71,11,10]). Depending on the geometric properties preserved, the continuation of the quasi-periodic solution requires to adjust parameters in the system.

In this paper we prove a KAM theorem in a geometric context different from those above. We consider "conformally symplectic" systems (maps and flows). A system is called conformally symplectic when it transports a symplectic form into a multiple of itself. KAM theory for some particular cases of conformally symplectic systems was obtained in [14].

Conformally symplectic systems appear in a variety of physical problems, notably mechanical systems with friction proportional to the velocity, in models of transport [34,89] and in optimization of discounted systems common in economics (see Section 2.3 and [9,51]).

Following [70], we will seek tori with a fixed frequency. Since conformally symplectic systems will be dissipative, finding a quasi-periodic orbit with a given frequency will require considering a family of systems depending on parameters and adjusting the parameters to ensure that the attractor is of the desired form. Indeed, for a system with fixed parameters, it could well happen that the only dynamics in the attractor is not quasi-periodic[1] (e.g., a strange non-chaotic attractor – [61]) or if it is quasi-periodic with a different function.

Therefore, the unknowns of our problem will be not just the quasi-periodic solutions, but the parameters to be adjusted. With a view to numerical applications, we will seek the quasi-periodic invariant torus by formulating a functional equation, namely (3.2) below, for the embedding of a torus whose dimension is the number of degrees of freedom. It is very important for applications that the number of variables in the embedding we are seeking is the number of degrees of freedom and not the dimension of the phase space. Having functions with few variables in our functional equations is important for numerics, because the difficulty of dealing with functions grows very rapidly with the number of variables (the *curse of dimensionality* in the words of [8]). We stress again that Eq. (3.2) is an equation not just for the embedding, but also for the parameters in the family.

The main result of this paper is provided by Theorem 20; its (somewhat technical) formulation is the standard form of "*a-posteriori* theorems" (see [33]).

If we are given a) an approximate solution of (3.2) (i.e. a function and a set of parameters so that, when substituted in (3.2), verify it up to a small error), b) that the approximate solution satisfies some non-degeneracy assumptions (the quantitative measures of these non-degeneracies are called *condition numbers* in numerical analysis), c) that the error is sufficiently small (depending on the condition numbers), then we conclude that there is a true solution of (3.2) and we can bound the distance between the approximate solution and the true solution by the error in the initial approximation (and the condition numbers). The method of proof will also allow us to conclude that the exact solution is unique in a neighborhood.

To make the above sketch precise, we just need to specify the sense in which the error is small, the sense in which the true solution is close to the approximate one (this, of course, entails defining some norms in spaces) and making explicit the non-degeneracy conditions.

---

[1] One of the results that we will prove are lower bounds on the measure of parameters for which the motion is quasi-periodic.

There are many choices of norms that we could use to make precise the notions of smallness alluded to above. In this paper we will present results both in analytic spaces and in Sobolev spaces. Having two types of norms will allow us to discuss bootstrap of regularity. Note that, since the algorithms give the Fourier series, the Sobolev norms are very accessible to the numerical computations and they are the basis of a very practical criterion for breakdown discussed in Section 9.2. See [15,24] for actual implementations.

The non-degeneracy conditions in Theorem 20 will be a generalization of the well-known twist conditions in KAM theory. Defining them properly in a far from integrable context will require some geometry. Note also that, since Eqs. (3.2) are equations for the embedding and the parameters, the non-degeneracy conditions will involve both the approximate solution and the dependence on parameters. Nevertheless, we will see that in the limit of zero dissipation the non-degeneracy assumptions reduce to twist conditions.

Note that given an a-posteriori theorem to conclude the existence of solutions one does not need to justify the way that the approximate solutions were produced, but only that they lead to small errors and that the non-degeneracy conditions are satisfied. This allows us also to validate some formal asymptotic expansions; in particular, the a-posteriori format is very well suited to obtain computer assisted proofs. One can take as the approximate solution assumed in Theorem 20 the output of a non-rigorous (but sensible and effective) program. One needs only to produce rigorous estimates on upper bounds for the error of the invariance equations and for the condition numbers. These rigorous estimates can be produced using extensions of interval arithmetic to functions' space (see, e.g., [64, 65,63,58]); this task is being pursued in [17].

Even for standard numerical computations, the a-posteriori format is particularly useful when we are working close to the breakdown of the tori, so that we can be confident that the calculations are correct and not corrupted by spurious solutions.

We also note that the simplicity of the proof presented here leads to very good estimates obtained in the traditional way using only pencil and paper. On the other hand, we note that we have not optimized the minimal regularity required or the regularity obtained. Of course, for many applications in Celestial Mechanics, the systems considered are analytic and regularity is not an issue, but the size of the perturbation is the dominant issue to obtain theorems that are applicable. For analytic systems, the computer assisted studies can give information not only on the numerical values of the perturbations allowed, but also on the domain of analyticity (see [15] for a non-rigorous study of the frontiers of domains of analyticity in some examples).

As we will see later, the method of proof also leads to very fast and efficient algorithms. These algorithms are non-rigorous, but very reliable in the traditional sense of numerical analysis. The strategy used in this paper to prove Theorem 20 is to develop a (quasi-)Newton method for the functional equation given by (3.2) below. This quasi-Newton method does not rely on transformation theory[2] (which requires dealing with functions with a number of variables equal to the dimension of the phase space), but rather it is just a modification of the standard Newton method for the invariance equation. Using some identities that come from the geometry, the quasi-Newton step factorizes into elementary steps, which can be done very efficiently either in Fourier or in real space (see Algorithm 33).

The conclusion is that, if we discretize the functions with N numbers (and remember that we are discretizing functions of a low number of variables), a Newton step requires $O(N)$ storage and $O(N \log(N))$ operations. Note that, even if we have a quadratically convergent method, we do not need to store (and much less to invert) a matrix of dimension N, which would have required $O(N^2)$ storage and $O(N^3)$ operations.

The combination of the fast algorithm and the ability to compute reliably close to the boundary of existence has been used in the exploration of the behavior near breakdown and in the study of the domains of analyticity ([15,24]; for previous investigations of the behavior close to breakdown using other methods, see [74–76]). This has led to a plethora of mathematical conjectures.

---

[2] A proof of persistence of tori in quasi-integrable dissipative systems based on transformation theory has appeared in [84].

The a-posteriori format with the local uniqueness leads almost immediately to other results (see Section 9). Among them: a) a result on bootstrap of regularity, b) a criterion for the breakdown of invariant tori. It was implemented numerically in [15] (a similar version of this criterion was presented for Hamiltonian systems in [21,20]), c) a proof of convergence of the asymptotic Lindstedt series around a dissipative system. We provide some arguments that indicate that the expansions starting near a conservative system could be divergent (see Section 10).

The method also leads to other results proved more or less in the same way, such as the persistence of whiskered tori in conformally symplectic systems [18] as well as a way to obtain fast algorithms for their computation along the lines of [46].

One respect in which the present method is inferior to transformation theory is in the study of finite regularity measured in C^m. The papers [68,69,90,91] developed the double smoothing method whose main idea is to approximate the original problem by a sequence of analytic ones. For Hamiltonian flows, this can be done smoothing the Hamiltonian. Smoothing maps while preserving the geometric structures is cumbersome (unless one is in a neighborhood of the identity). See [38] for a study of double smoothing while preserving a symplectic or volume structure.

For the sake of efficiency of exposition we present in great detail mainly results for maps, while in Appendix A we present the minor modifications required to obtain results for the case of flows.

Of course, one could have deduced the results for maps from results for time periodic flows. From our point of view, working with maps is more straightforward because the geometric reasons are easy to explain and in practical problems they appear naturally; indeed, in the review of [91] for Mathematical Reviews J. Moser suggested that *it seems desirable to have a direct proof for this basic result* [the preservation of tori for maps].

We finally remark that we provide estimates for the different algorithms, but we do not intend to give an explicit expression for the constants, though their dependence on parameters and norms is presented when necessary. For this reason, throughout this paper C denotes a generic positive constant. In practical applications – eventually carried out with the help of a computer – it is straightforward to write a sequence of functions that gives the constants entering in one step as functions of the previous ones, even if the final expression of the constants would be cumbersome to write. More quantitative results for concrete examples are taken up in [17]. With a view to these concrete estimates in applications, the paper presents somewhat more details than what is customary in other more theoretical papers.

1.1. Organization of the paper

The paper is organized as follows. In Section 2 we define conformally symplectic maps and flows. In Section 3 we formulate the invariance equation and we present several geometric identities which lead to the existence of an interesting system of coordinates in the neighborhood of an invariant torus. In particular, in Section 3.3 we use this system of coordinates to apply the theory of normally hyperbolic manifolds [36,37,49], as well as to obtain results on the regularity of the manifolds and on the behavior of perturbations near the quasi-periodic solutions. In particular, we show that all the quasi-periodic solutions are local attractors. This system of coordinates allows also to give a concise formulation of the non-degeneracy assumptions. We hope that presenting early some results on the geometry may serve as motivation for the form of the non-degeneracy conditions. In Section 4 we introduce spaces of analytic functions and Sobolev spaces and provide some standard technical results on them.

In Section 5, we state Theorem 20, which is the main result of this paper. Theorem 20 establishes the existence of solutions of the invariance equation, provided that we have approximate solutions which satisfy the non-degeneracy conditions. Of course, to read the statement of Theorem 20, only the definitions (and not the elementary results) in the preliminary material are needed.

In Section 5 we also state some related results, such us Theorem 29 on the local uniqueness of the solutions, Corollary 30 on the Lipschitz dependence on parameters of the solutions, and Corollary 32 on the measure in parameter space covered by quasi-periodic attractors.

The proof of Theorem 20 is based on a Newton-like method. The iterative step for the Newton's method is formulated in Section 6. The key idea of the iterative step is to use an adapted system of

coordinates near the solution so that the corrections can be done very efficiently. This is accomplished in Section 7.1. The estimates for the corrections applied in one iterative step are performed in Section 7. In Section 7.2 we make precise the statement that the error after one step is quadratic in the original error. After these quadratic estimates, there are standard abstract theorems that show that alternating the iteration with carefully chosen smoothings, the procedure converges. There are many variants of these ideas; a theorem well adapted to these methods appears in Appendix A of [21], while a slight improvement of it appears in Theorem 47 and we present a complete proof. For the sake of completeness, in Section 7.6 we also present a short proof of the convergence in the analytic case.

The proof of the uniqueness of the solution is presented in Section 8. In Section 9 we discuss some consequences of the a-posteriori formalism, such as the bootstrap of regularity and a criterion to compute the breakdown threshold. The perturbative expansions, namely the formal series solutions and their convergence, are discussed in Section 10. The algorithm to compute the parametric representation of quasi-periodic solutions for flows and the easy modifications required for the proof are presented in Appendix A.

2. Geometric preliminaries

We consider the phase space $\mathcal{M} = \mathbb{T}^n \times B$, $B \subseteq \mathbb{R}^n$ (B being an open, simply connected domain with a smooth boundary), endowed with the standard scalar product and a symplectic form Ω.

Note that this does not entail any loss of generality, since we can take $\mathcal{M}$ to be a subset of another manifold. Clearly, if we aim to look for an invariant torus, we can find a neighborhood of it of the form $\mathcal{M}$ and we will always work on $\mathcal{M}$.

We do not assume that Ω has the standard form; this generality is useful in several applications, for example when dealing with surfaces of section of Hamiltonian systems. We denote by $J = J(x)$ the matrix representing Ω at x, namely for any vectors u, v, one has

$$\Omega_x(u, v) = \big(u, J(x)v\big),$$

where $(\cdot, \cdot)$ denotes the Euclidean scalar product. We consider systems described by conformally symplectic mappings (see Section 2.1) or by conformally symplectic flows (see Section 2.2), which are defined as follows.

2.1. Conformally symplectic mappings

We introduce the notion of conformally symplectic maps (see [87,34,89,50,72,7,13] for studies of conformally symplectic geometry).

Definition 1. We say that a diffeomorphism f on $\mathcal{M}$ is *conformally symplectic*, if there exists a function $\lambda : \mathcal{M} \to \mathbb{R}$ such that[3]

$$f^*\Omega = \lambda\Omega. \tag{2.1}$$

When $n = 1$, any diffeomorphism is conformally symplectic with $\lambda(x) = \sigma |\det(Df(x))|$, $\sigma = +1, -1$ depending on whether the diffeomorphism is orientation preserving or reversing. When $n \geqslant 2$, the only possible λ is a constant function. In fact, taking the exterior derivatives of the left hand side of (2.1) one obtains

$$d\big(f^*\Omega\big) = f^* d\Omega = 0,$$

---

[3] By f^* we denote the pull-back via f (see [4,86] for an introduction to Cartan's calculus).

while from the right hand side of (2.1) one obtains:

$$d(\lambda \Omega) = d\lambda \wedge \Omega + \lambda \wedge d\Omega = d\lambda \wedge \Omega,$$

from which it follows that $d\lambda = 0$ for $n \geqslant 2$; since the manifold $\mathcal{M}$ is simply connected, one obtains that λ is constant.

Throughout this paper we will always consider the case λ equal to a constant, unless explicitly stated.

Note that if f is conformally symplectic, so it is the j-th iterate f^j. Indeed, when λ is constant one gets

$$\left(f^j\right)^* \Omega = \lambda^j \Omega.$$

In general, one has:

$$\left(f^j\right)^* \Omega = \lambda \circ f^{j-1}(x) \cdots \lambda(x) \Omega(x).$$

We remark that there exist more general definitions of conformally symplectic diffeomorphisms [7], but we prefer to use the formulation (2.1) with λ constant as it will be apt for several applications to physical problems.

An example of a conformally symplectic system that has appeared often in practice is the dissipative standard map, which is a 2-parameter family of maps, say $f_{\mu,\varepsilon}$, given by $f_{\mu,\varepsilon}(I, \varphi) = (\bar{I}, \bar{\varphi})$ with

$$\bar{I} = \left(I + \varepsilon V'(\varphi) + \mu\right)\lambda,$$
$$\bar{\varphi} = \varphi + \bar{I}, \tag{2.2}$$

where $V(\varphi)$ is a periodic, analytic function and $V'(\varphi)$ denotes its first derivative. Notice that for the mapping (2.2) one has that $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. The map (2.2) has been extensively investigated in the literature (see, e.g., [25,74–76]). The conservative case is obtained setting $\lambda = 1$ and $\mu = 0$.

For completeness we introduce also the following definition of exact conformally symplectic map, which applies also when studying the limit $\lambda = 1$.

Definition 2. If $\Omega = d\alpha$, we say that a diffeomorphism f is *exact conformally symplectic*, if there exists a single-valued function P such that

$$f^*\alpha = \lambda\alpha + dP.$$

The function P is called the primitive function of f. In the conservative case, it was extensively studied in [41]. Many of the properties of the primitive function for conservative systems have analogues in the conformally symplectic case.

Note also that, given a symplectic form, there can be several α's. The exact symplectomorphisms do no change, but their primitive functions depend on what is the α chosen.

As an example, if we take $\alpha = I d\varphi$, we see that in the standard map (2.2), we obtain:

$$f^*\alpha = \bar{I} d\bar{\varphi} = \lambda I d\varphi + dP(I, \varphi) + \lambda\mu d\varphi + \lambda^2 \mu dI + \lambda^2 \mu\varepsilon V''(\varphi) d\varphi$$

with $P(I, \varphi) = \lambda\varepsilon V(\varphi) + (\lambda^2/2)I^2 + \lambda^2\varepsilon V'(\varphi)I + \varepsilon^2(\lambda^2/2)V'^2(\varphi)$.

Therefore the standard map (2.2) is exact conformally symplectic if and only if $\mu = 0$. This can be seen more easily noting that the standard map can be written as $S = S_e \circ S_c$, where $S_e(I, \varphi) =$

$(I, I + \varphi)$, $S_c(I, \varphi) = (\lambda(I + \varepsilon V'(\varphi) + \mu), \varphi)$. It is easy to see that S_e is always exact symplectic, while S_c is exact conformally symplectic when and only when $\mu = 0$.

Remark 3. It should be clear that the results of this paper generalize to a somewhat more general context. In fact, if the phase space decomposes as

$$\mathcal{M} = \mathcal{M}_1 \times \cdots \times \mathcal{M}_j, \qquad \Omega = \Omega_1 \otimes \cdots \otimes \Omega_j, \quad j \geqslant 1,$$

it suffices to assume that

$$f^*\Omega = \lambda_1 \Omega_1 \otimes \cdots \otimes \lambda_j \Omega_j$$

with λ_j constants. This general set-up appears naturally in physical applications. It corresponds to j particles moving by a Hamiltonian interaction supplemented by a friction. Each particle experiences a frictional force proportional to its velocity, where the friction coefficient of each particle might be different. As we will see in Section 2.2, the friction coefficient of each particle is related to λ. The main ingredient is that the automatic reducibility discussed in Section 3.1 generalizes to the above context (more details are given in Remark 8).

2.2. Conformally symplectic flows

Definition 4. We say that a vector field X is a *conformally symplectic flow*, if there exists a function $\eta : \mathbb{R}^{2n} \to \mathbb{R}$ such that for the symplectic form Ω, we have:

$$L_X \Omega = \eta \Omega,$$

where L_X denotes the Lie derivative.

If η is constant, then the time t flow Φ_t satisfies

$$(\Phi_t)^*\Omega = \exp(\eta t)\Omega.$$

In the case that $\Omega = d\alpha$, there is a particularly interesting characterization of conformally symplectic flows. Denoting by i_X the contraction with the vector field X, one has

$$d(\eta\alpha) = \eta \, d\alpha = L_X \Omega = i_X \, d\Omega + d(i_X \Omega) = d(i_X \Omega),$$

so that $\eta\alpha$ and $i_X\Omega$ differ by a closed form. We say that the vector field X is *exact conformally symplectic*, when there exists a function H such that

$$i_X \Omega = \eta\alpha + dH. \tag{2.3}$$

When Ω is the standard form, say $\Omega = \sum_{j=1}^n d\varphi_j \wedge dI_j$, then $\alpha = I \, d\varphi$, so that Eq. (2.3) becomes

$$\dot{I} = -\frac{\partial H}{\partial \varphi} - \eta I,$$

$$\dot{\varphi} = \frac{\partial H}{\partial I}; \tag{2.4}$$

notice that Eqs. (2.4) are a generalization of the standard Hamilton's equations in symplectic geometry. In the 2-dimensional (1 degree of freedom) case, if $\text{div}(X) = -\eta$ and η is a constant, then the flow X_t changes the volume by a factor $\exp(-\eta t)$.

One important class of examples, extensively studied in the literature, is systems with a dissipation proportional to the velocity and subject to an external forcing through a potential force. In such a case, the time t map (obtained taking the solution of the flow at discrete times) will be conformally symplectic.

An example that has been studied several times in the literature is the spin-orbit problem [23], used to model the rotation of an oblate satellite around a planet. It is described by the equations

$$\dot{I} = -\frac{\partial V(\varphi, t)}{\partial \varphi} - \eta I + \mu,$$
$$\dot{\varphi} = I, \tag{2.5}$$

where $V(\varphi + 1, t) = V(\varphi, t)$, $V(\varphi, t) = V(\varphi, t + 1)$. All the vector fields (2.5) are conformally symplectic for the form $\Omega = d\varphi \wedge dI$. They are exact conformally symplectic, if and only if $\mu = 0$. They correspond to the (local) Hamiltonian $H_\mu = \frac{1}{2}I^2 + V(\varphi, t) - \mu\varphi$. Note that, even if H_μ is locally well defined, it is a globally defined function if and only if $\mu = 0$ (in old function language, H_μ is multi-valued when $\mu \neq 0$). Hence, the flow is exact conformally symplectic precisely when $\mu = 0$. The conservative case corresponds to $\eta = 0$ and $\mu = 0$. As it is well known, for conservative vector fields there are homotopically non-trivial invariant tori if and only if they are exact. Note that (2.2) is a discrete analogue of (2.5).

2.3. Conformally symplectic systems in variational problems with discounted systems

Conformally symplectic systems appear as the Euler–Lagrange equation of exponentially discounted variational principles with a Legendre condition. That is, for some real parameters $\eta > 0$ or $0 < \lambda < 1$ we are interested in critical points of

$$\mathcal{S}(q) = \int_0^\infty e^{-\eta t} L\big(q(t), \dot{q}(t)\big) \, dt,$$
$$\mathcal{S}(u) = \sum_{n=0}^\infty \lambda^n S(u_n, u_{n+1}) \tag{2.6}$$

respectively in the continuous-time case (with $L = L(q, \dot{q})$ being the Lagrangian function in phase space coordinates), and in the discrete-time case (with $S = S(u_n, u_{n+1})$ being the generating function). Models of this form are very common in finance, were future gains are exponentially discounted by the interest. They also appear in mechanics and other contexts [9]. The functionals (2.6) are very often bona-fide functionals, in contrast with the action principles of mechanics, which are often just formal integrals.

It is easy to see that the Euler–Lagrange equations for the above problems are, respectively,

$$\partial_1 L\big(q(t), \dot{q}(t)\big) - \frac{d}{dt}\partial_2 L\big(q(t), \dot{q}(t)\big) - \eta \partial_2 L\big(q(t), \dot{q}(t)\big) = 0,$$
$$\partial_1 S(u_n, u_{n+1}) + \lambda \partial_2 S(u_{n-1}, u_n) = 0, \tag{2.7}$$

where ∂_j denotes the derivative with respect to the j-th argument. Under the customary Legendre conditions, (2.7) can be transformed into conformally symplectic systems. Of course, when the Legendre conditions fail, the variational problem is very different from a conformally symplectic system.

Some issues such as the convergence of the critical points or mimimizers of (2.6) to the solutions of the Hamilton–Jacobi equation as $\eta \to 0$ or $\lambda \to 1$ are actively pursued in the literature (see [9,51]). Note that the results presented in Section 10 establish the convergence of the quasi-periodic solutions of the dissipative system to the corresponding solutions of the Hamiltonian one.

As customary in the comparison of KAM and variational methods, the KAM methods do not need convexity as the variational methods, but on the other hand the variational methods do not need assumptions on the size of the perturbations, and require less assumptions on smoothness than for KAM results.

3. Formulation of the functional equations satisfied by the embeddings of quasi-periodic tori

We denote by $\omega \in \mathbb{R}^n$ the *frequency* of motion, which we assume to satisfy the Diophantine condition

$$|\omega \cdot k - \ell| \geqslant \nu |k|^{-\tau}, \quad \ell \in \mathbb{Z}, \ k \in \mathbb{Z}^n \setminus \{0\}, \tag{3.1}$$

for suitable positive real constants $\nu \leqslant 1$, $\tau \geqslant 1$.

The set of Diophantine vectors is denoted by $\mathcal{D}_n(\nu, \tau)$. If the dimension of the space is obvious, we will omit the subindex n.

Given a family f_μ of conformally symplectic maps (that satisfies some non-degeneracy assumptions to be specified later) with $\mu \in \Lambda$, $\Lambda \subseteq \mathbb{R}^n$ open, we look for a value μ, say $\mu = \mu_e$, and an embedding $K : \mathbb{T}^n \to \mathcal{M}$, such that the following invariance equation is satisfied:

$$f_{\mu_e} \circ K(\theta) = K(\theta + \omega). \tag{3.2}$$

Eq. (3.2) will be the centerpiece of our treatment. For example, setting $\varepsilon = 0$ in (2.2) we can see that, for any λ, one gets

$$K(\theta) = (\theta, \omega), \qquad \mu_e = (\omega - \omega\lambda)/\lambda.$$

Notice that if (K, μ_e) satisfy (3.2), then $f_{\mu_e}(\text{Range}(K)) = \text{Range}(K)$ and, since K is an embedding, Range(K) is diffeomorphic to $\mathbb{T}^n$. Notice also that if (3.2) holds, then for any $\sigma \in \mathbb{T}^n$ the sequence $\{x_n\} = K(\sigma + n\omega)$ is an orbit of the map f_{μ_e}; therefore, the dynamics of $f_{\mu_e}|_{\text{Range}(K)}$ is diffeomorphic to a rotation.

Notice that our results are based in studying embeddings of a torus, so that we can study tori that are contractible (secondary tori) or meandering tori, which are not graphs in a system of coordinates. It is well known that secondary tori appear naturally near resonances [27,43] and that meandering tori appear in relation to oscillations [61]. The method presented here just requires that we find a parametrization solving approximately an invariance equation.

Remark 5. The solutions of (3.2) are never unique. Defining the shift T_σ such that $T_\sigma(\theta) = \theta + \sigma$, it is easy to see that if (K, μ_e) is a solution of (3.2), then $(K \circ T_\sigma, \mu_e)$ is also a solution for every $\sigma \in \mathbb{T}^n$.

Hence, Eq. (3.2) admits n-parameter families of solutions (obtained by choosing a different origin of phase of the solution). All these solutions correspond to the same geometric invariant object in phase space. In Sections 5.1 and 8 we will show that this is the only non-uniqueness of the problem in a neighborhood of the solution. In particular, the geometric tori are locally unique.

The geometric meaning of the lack of uniqueness pointed above is such that we shift the origin of coordinates in the reference torus we embed.

The problem of global uniqueness has been considered in [1], which contains global results for some particular systems – modifications of geodesic flows – with strong enough dissipation. The paper [1] shows that, under these circumstances, there is a unique Lagrangian manifold invariant under the flow. The paper [1] does not consider whether the motion in this manifold is given by

a rotation. Of course, there is no hope to obtain global uniqueness results unless one makes global assumptions.

Remark 6. For a family of conformal vector fields X_μ, fixing ω Diophantine in the sense defined in Appendix A, the invariance equation means to look for a value μ_e and an embedding K, such that

$$X_{\mu_e} \circ K(\theta) = (\omega \cdot \partial_\theta) K(\theta).$$

The following result, stating that invariant tori are Lagrangian, is already known for tori invariant by exact symplectic maps (or flows) (see, e.g., [30]), but we state it here for tori invariant by conformally symplectic mappings. Later, in Lemma 37, we will show that approximately invariant tori are approximately Lagrangian.

Proposition 7. *Let $n \geqslant 2$, let f be conformally symplectic with $|\lambda| \neq 1$ and let K satisfy* (3.2). *Then, one has*

$$K^* \Omega = 0. \tag{3.3}$$

If $\lambda = 1$, assuming furthermore that ω is irrational and that f is exact, then (3.3) *holds.*

The case $n = 1$ is trivial, since in a 1-dimensional manifold one can only define trivial 2-forms.

Proof. One easily obtains that

$$(f \circ K)^* \Omega = K^* f^* \Omega = \lambda K^* \Omega$$

and that

$$(K \circ T_\omega)^* \Omega = T_\omega^* K^* \Omega.$$

In coordinates we see that if $K^*(\Omega) \equiv \sum A_{ij}(\theta)\, d\theta_i \wedge d\theta_j$ (for suitable functions $A_{ij} = A_{ij}(\theta)$), then we have $A_{ij}(\theta + \omega) = \lambda A_{ij}(\theta)$. If $\lambda > 1$, we note that $A_{ij}(\theta) = \lambda^{-n}(A_{ij}(\theta + n\omega))$. Since A_{ij} is bounded, we obtain (3.3) taking the limit as $n \to \infty$, while if $\lambda < 1$ we take the limit as $n \to -\infty$. For the (slightly more complicated) symplectic case, compare with [90,30]. From $K^* \Omega = T_\omega^*(K^* \Omega)$ we deduce that, in coordinates, $K^* \Omega$ is constant, since the rotation is irrational. If $\Omega \equiv d\alpha$, we obtain that $K^* \Omega = d(K^* \alpha)$ and the only constant form which is an exact differential is identically zero. □

For further applications, it will be important to generalize the Lagrangian character of invariant tori to quasi-invariant tori, namely tori which satisfy the invariance equation (3.2) up to a small error term. The precise formulation of the results for quasi-invariant systems requires quantitative measures of the quasi-invariance as well as some results on the solutions of the difference equations as it will be done in Section 4.

3.1. Automatic reducibility

A key argument for our results is that in the neighborhood of an invariant torus, there is an explicit change of coordinates that makes the linearization of the invariance equation into a constant coefficient equation. We also note that this system of coordinates makes it also particularly simple to study the long term behavior of the variational equations, hence we can use this system of coordinates to obtain dynamical information such as Lyapunov exponents. The geometric interpretation of these identities is illustrated in Fig. 1. These geometric effects were already observed in [29,30] for the case of symplectic mappings.
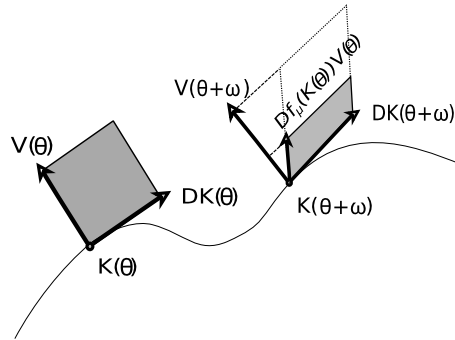
Fig. 1. Geometric explanation of automatic reducibility, where $V(\theta)$ is perpendicular to $DK(\theta)$ and such that the area of the shaded parallelogram is unitary.

In this section, we explain in detail the geometric reason for the so-called *automatic reducibility* of invariant tori (i.e. the fact that the preservation of a geometric structure leads to the existence of a system of coordinates, where the linear evolution is a constant coefficient matrix). Later, in Section 3.4, we will present a generalization to approximately invariant tori.

As we will see, this system of coordinates for approximately invariant solutions is crucial to obtain a Newton's step that has quadratic convergence, but "tame" estimates in the sense of Nash–Moser implicit function theorems. We think it is worth to start by covering first the exactly invariant case, since then all the arguments are geometrically natural. They can be adapted to the quasi-invariant case rather easily. Furthermore, we point out that we will use the coordinates in the exactly invariant case for other purposes, namely: A) to compute the Lyapunov exponents (Section 3.3), B) to establish local uniqueness (Section 8), C) to show that there exist perturbation theories to all orders, to establish their convergence and to develop fast algorithms for their computation (Section 10). Further applications of the adapted system of coordinates appear in [16].

3.2. An adapted system of coordinates for solutions of (3.2)

Taking the derivative of (3.2) (we write μ instead of μ_* to make the formulas less cluttered), we obtain:

$$(Df_\mu) \circ K \, DK - DK \circ T_\omega = 0. \tag{3.4}$$

Geometrically, the above equation (3.4) means that each of the vector fields $\partial_i K$ gets transported by Df_μ into itself. Note that the range of $DK(\theta)$ is the tangent space of Range(K) at $K(\theta)$. Since the range of a matrix does not change if we multiply it on the right, we find it convenient to introduce the normalization

$$N(\theta) \equiv \left(DK(\theta)^T DK(\theta)\right)^{-1}. \tag{3.5}$$

Let us define the function $V(\theta)$ as

$$V(\theta) \equiv J^{-1} \circ K(\theta) \, DK(\theta) N(\theta).$$

Expressing $K^*\Omega$ in coordinates,[4] the fact that K is Lagrangian is written as (see [30,3])

$$DK^T(\theta) J \circ K(\theta) DK(\theta) = 0. \tag{3.6}$$

Due to the Lagrangian character of the invariant torus, we have

$$\begin{aligned}
&\text{Range}\big(DK(\theta)\big) \cap \text{Range}\big(J^{-1} \circ K(\theta) DK(\theta)\big) \\
&= J^{-1} \circ K(\theta)\big[\text{Range}\big(J \circ K(\theta) DK(\theta)\big) \cap \text{Range}\big(DK(\theta)\big)\big] \\
&= \{0\}.
\end{aligned}$$

Hence, we have that $\text{Range}(DK(\theta)) \oplus \text{Range}(J^{-1} \circ K(\theta) DK(\theta))$ is a $2n$-dimensional vector space, which coincides with the whole tangent space. Therefore can write

$$Df_\mu \circ K(\theta) V(\theta) = V(\theta + \omega) A(\theta) + DK(\theta + \omega) S(\theta), \tag{3.7}$$

where, setting

$$P(\theta) \equiv DK(\theta) N(\theta), \tag{3.8}$$

we will see that $A(\theta)$ and $S(\theta)$ are given by

$$\begin{aligned}
A(\theta) &\equiv \lambda \, \text{Id}, \\
S(\theta) &\equiv P(\theta + \omega)^T Df_\mu \circ K(\theta) J^{-1} \circ K(\theta) P(\theta) - N(\theta + \omega)^T \gamma(\theta + \omega) N(\theta + \omega) A(\theta)
\end{aligned} \tag{3.9}$$

and

$$\gamma(\theta) \equiv DK(\theta)^T J^{-1} \circ K(\theta) DK(\theta). \tag{3.10}$$

When J is a complex structure, $J^{-1} = -J$, and the Lagrangian character of the torus is the same as $\gamma(\theta) = 0$. The same conclusion leading to $\gamma(\theta) \equiv 0$ happens also when J^2 is a multiple of the identity (which is what happens when one considers symplectic polar coordinates).

To prove (3.9) we just multiply (3.7) by appropriate factors and use geometric identities that come from the invariance and from the geometric properties of the torus (notably the Lagrangian character). For later purposes, it is important to remark that exactly the same procedure works for approximately invariant tori; in that case, of course, we will obtain that the identifications happen up to errors which can be estimated by the error in the invariance equation and in the Lagrangian character (which will, in turn, be controlled by the error in the invariance). Multiplying the left hand side of (3.7) by $DK(\theta + \omega)^T J \circ K(\theta + \omega)$ and using (3.4) we obtain

$$\begin{aligned}
&DK(\theta + \omega)^T J \circ K(\theta + \omega) Df_\mu \circ K(\theta) J^{-1} \circ K(\theta) DK(\theta) N(\theta) \\
&= \lambda \, DK(\theta + \omega)^T Df_\mu^{-T} \circ K(\theta) DK(\theta) N(\theta) \\
&= \lambda \, DK(\theta)^T DK(\theta) N(\theta) \\
&= \lambda \, \text{Id}, \tag{3.11}
\end{aligned}$$

---

[4] Recall that $K^*(\Omega)(u, v) = \Omega(DK u, DK v)$ for any vectors u, v, applying the general formulas of pull-back for forms [4].

where in the first line we have used the fact that f_μ is conformally symplectic,[5] namely

$$Df_\mu \circ K(\theta)^T J \circ K(\theta + \omega) Df_\mu \circ K(\theta) = \lambda J \circ K(\theta). \tag{3.12}$$

Moreover, from the right hand side of (3.7) one has

$$DK(\theta + \omega)^T J \circ K(\theta + \omega) \big[J^{-1} \circ K(\theta + \omega) DK(\theta + \omega) N(\theta + \omega) A(\theta) + DK(\theta + \omega) S(\theta) \big]$$
$$= DK(\theta + \omega)^T DK(\theta + \omega) N(\theta + \omega) A(\theta)$$
$$= A(\theta),$$

where we have used the definition of N and the Lagrangian character of the torus (see (3.6)). To compute $S(\theta)$ we multiply (3.7) by $P(\theta + \omega)^T = N(\theta + \omega)^T DK(\theta + \omega)^T$ and we obtain

$$P(\theta + \omega)^T Df_\mu \circ K(\theta) V(\theta) = P(\theta + \omega)^T V(\theta + \omega) A(\theta) + P(\theta + \omega)^T DK(\theta + \omega) S(\theta)$$
$$= N(\theta + \omega)^T \gamma(\theta + \omega) N(\theta + \omega) A(\theta) + S(\theta),$$

where, in the last line, we have just used the definition of $\gamma(\theta)$ (see (3.10)) and that

$$P(\theta + \omega)^T DK(\theta + \omega) = \text{Id}.$$

This completes the proof of (3.9).

Defining $M(\theta)$ as the $2n \times 2n$ matrix obtained juxtaposing the two $2n \times n$ matrices $DK(\theta)$, $V(\theta)$, namely

$$M(\theta) = \big[DK(\theta) \mid J^{-1} \circ K(\theta) DK(\theta) N(\theta) \big], \tag{3.13}$$

we obtain

$$Df_\mu \circ K(\theta) M(\theta) = M(\theta + \omega) \begin{pmatrix} \text{Id} & S(\theta) \\ 0 & \lambda \text{Id} \end{pmatrix}. \tag{3.14}$$

The geometric reason why (3.14) is true is illustrated in Fig. 1. We note that the vector field $DK(\theta)$ gets transported; geometrically $V(\theta)$ is a vector orthogonal to $DK(\theta)$, normalized so that the area of the parallelogram formed by them is equal to 1. Equivalently, the area of the parallelogram formed by $DK(\theta + \omega)$ and $V(\theta + \omega)$ is also equal to 1. The action of the derivative Df_μ on the parallelogram contracts the area by a factor λ; due to (3.7), the projection of $Df_\mu \circ K(\theta) V(\theta)$ onto $V(\theta + \omega)$ has to be λ times the length of $V(\theta + \omega)$.

Remark 8. The above construction generalizes to the case discussed in Remark 3 in which we consider different particles, each with its own friction coefficient. More precisely, we consider $\theta = (\theta_1, \ldots, \theta_j)$ and similarly $K(\theta) = (K_1(\theta), \ldots, K_j(\theta))$, where K_i takes values in the i-th copy of the manifold and it describes the motion of the i-th particle. We note that taking derivatives of the invariance equation we still obtain $Df_\mu \circ K(\theta) D_{\theta_i} K_i(\theta) = D_{\theta_i} K_i(\theta + \omega)$. Note also that, because the symplectic form is a product, we can define a symplectic conjugate $v(\theta) = (v_1(\theta_1), \ldots, v_j(\theta_j))$ with

---

[5] The conformally symplectic condition is equivalent to saying that $\Omega_{f(x)}(Df(x)u, Df(x)v) = \lambda \Omega_x(u, v)$ for any vectors u, v. Therefore, $(Df(x)u, J \circ f(x) Df(x)v) = \lambda(u, J(x)v)$; being valid for any vectors u, v, one gets $Df(x)^T J \circ f(x) Df(x) = \lambda J(x)$, which gives (3.12) taking $x = K(\theta)$.

$v_i(\theta_i) = J_i^{-1} \circ K_i(\theta) D_{\theta_i} K_i(\theta) N_i(\theta)$. The same geometric argument used in the text shows that we have

$$Df_\mu \circ K(\theta) v_i(\theta_i) = \lambda_i v_i(\theta + \omega) + S_i(\theta) D_{\theta_i} K_i(\theta + \omega)$$

$$Df_\mu \circ K(\theta) M(\theta) = M(\theta + \omega) \begin{pmatrix} \mathrm{Id} & S(\theta) \\ 0 & \Lambda \end{pmatrix},$$

where Λ is a diagonal matrix with entries $\lambda_1, \ldots, \lambda_j$.

Remark 9. One can make further changes of variables so that the matrix S in (3.14) takes a simpler form. For example we can consider

$$\tilde{M}(\theta) = M(\theta) \begin{pmatrix} \mathrm{Id} & B(\theta) \\ 0 & \mathrm{Id} \end{pmatrix}, \tag{3.15}$$

for a suitable matrix B to be determined as follows. Using (3.14) and (3.15) we see that

$$\begin{aligned} Df_\mu \circ K(\theta) \tilde{M}(\theta) &= M(\theta + \omega) \begin{pmatrix} \mathrm{Id} & S(\theta) \\ 0 & \lambda \, \mathrm{Id} \end{pmatrix} \begin{pmatrix} \mathrm{Id} & B(\theta) \\ 0 & \mathrm{Id} \end{pmatrix} \\ &= \tilde{M}(\theta + \omega) \begin{pmatrix} \mathrm{Id} & -B(\theta + \omega) \\ 0 & \mathrm{Id} \end{pmatrix} \begin{pmatrix} \mathrm{Id} & S(\theta) \\ 0 & \lambda \, \mathrm{Id} \end{pmatrix} \begin{pmatrix} \mathrm{Id} & B(\theta) \\ 0 & \mathrm{Id} \end{pmatrix} \\ &= \tilde{M}(\theta + \omega) \begin{pmatrix} \mathrm{Id} & -\lambda B(\theta + \omega) + S(\theta) + B(\theta) \\ 0 & \lambda \, \mathrm{Id} \end{pmatrix} \\ &= \tilde{M}(\theta + \omega) U(\theta + \omega), \end{aligned} \tag{3.16}$$

where the last equality defines $U(\theta + \omega)$. Hence, if we use the theory of solutions of cohomology equations when $|\lambda| \neq 1$ (see Section 4.2), we can choose $B(\theta)$ in such a way that

$$0 = -\lambda B(\theta + \omega) + S(\theta) + B(\theta). \tag{3.17}$$

Since (3.17) does not involve any small divisors, the solution B is as smooth as S. The geometric meaning of the construction above is that, when $|\lambda| \neq 1$, we choose a coordinate system in which there is a contracting invariant space transversal and complementary to the tangent of the tori.

In the case that $\lambda = 1$, we cannot solve (3.17), but we can have that the left hand side of (3.17) is a constant. Hence we can arrange that $U(\theta)$ in (3.16) becomes a constant upper-diagonal matrix with Id in the diagonal.

3.3. Relation with the regularity theory of normally hyperbolic manifolds

Remark 9 has important consequences for the dynamics, which we will now discuss. To simplify the exposition, we present only the case $|\lambda| < 1$; the case $|\lambda| > 1$ is reduced to the previous case by considering the inverse map.

The main observation is that Remark 9 has the dynamical interpretation that we can find a frame of reference in which the linearized dynamics is given by a constant diagonal matrix with eigenvalues $1, \lambda$, each of them with multiplicity n. This says, in particular, that the manifolds are normally hyperbolic. From Remarks 8 and 9 we have that for $j > 0$:

$$Df_\mu^j \circ K(\theta) = \tilde{M}(\theta + j\omega) \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & \lambda^j \, \mathrm{Id} \end{pmatrix} \tilde{M}^{-1}(\theta),$$

where $\tilde{M}$ is the matrix in Remark 9. We have, therefore, shown that there exists a decomposition

$$T_{K(\theta)}\mathcal{M} = \text{Range}\big(DK(\theta)\big) \oplus E^s_{K(\theta)} \tag{3.18}$$

(we use the standard notation from differential geometry, where $T_x\mathcal{A}$ denotes the tangent space of the manifold $\mathcal{A}$ at the point x), where E^s_x is the eigenspace corresponding to the eigenvalue λ in the constant system of coordinates. We have also shown that $\text{Range}(DK(\theta)) = T_{K(\theta)}K(\mathbb{T}^n)$ and that DK corresponds to the eigenvalue 1.

By construction, the splitting in (3.18) is invariant under Df_μ and we have that there is a constant C such that for all $j \in \mathbb{Z}$,

$$C^{-1}\lambda^j|v| \leqslant \big|Df^j_\mu \circ K(\theta)v\big| \leqslant C\lambda^j|v| \quad \Longleftrightarrow \quad v \in E^s_{K(\theta)},$$

$$C^{-1}|v| \leqslant \big|Df^j_\mu \circ K(\theta)v\big| \leqslant C|v| \quad \Longleftrightarrow \quad v \in T_{K(\theta)}K\big(\mathbb{T}^n\big). \tag{3.19}$$

The constants C are estimated by the norms of $\|\tilde{M}\|_{L^\infty}$, $\|\tilde{M}^{-1}\|_{L^\infty}$.

The properties (3.19) imply the assumptions of the theory of normal hyperbolicity with rate conditions (see, e.g., [36,37]). Indeed, the standard definition of normal hyperbolicity only requires the analogue of (3.19) with the exponential estimates of $Df^j_\mu|_{E^s}$ for positive j; the definition of normal hyperbolicity can accommodate also an unstable space and some (small) rate of growth of the derivative on the tangent space. Hence, applying the theory of normally hyperbolic manifolds, we obtain several consequences, among them, the following result.

Proposition 10. *If f_μ is conformally symplectic, $|\lambda| < 1$ and K satisfies (3.2), then the manifold $K(\mathbb{T}^n)$ is an attractor.*

We can prove Proposition 10 in several ways. For example, it suffices to appeal to the results in [36] on the dynamics on stable manifolds (in our case, the stable manifold is the whole space). The most direct proof is to show that if we consider the map expressed in the coordinates given by the frame constructed in Remark 9, we have that it is given by

$$(I, \varphi) \to (\lambda I, \varphi + \omega) + R(I, \varphi),$$

where $|R(I, \varphi)| \leqslant C|I|^2$, $|DR(I, \varphi)| \leqslant C|I|$. In such circumstances, it is clear that for I in a small neighborhood, I decreases exponentially under the forward iteration. Note that in Proposition 10 we do not need to assume any non-resonance property on ω. The only thing that we need is the possibility to construct the frame in Remark 9, which depends only on the fact that $K(\mathbb{T}^n)$ is a manifold and that it is Lagrangian, without assuming irrationality of the rotation when $|\lambda| < 1$.

Several further developments on the behavior near quasi-periodic solutions are obtained in [16].

We now develop some other consequences of the theory of normally hyperbolic manifolds for our case.

Proposition 11. *When $|\lambda| < 1$ the tori $\mathcal{K} = K(\mathbb{T}^n)$ are as differentiable as the map, when measured in the C^r regularity classes, $r \in \mathbb{N}$ (similarly for $|\lambda| > 1$, considering the inverse mapping).*

The reason why Proposition 11 is true is that in Remark 9 we have seen that there is a continuous splitting $T_x\mathcal{M} = T_x\mathcal{K} \oplus E^s_x$. When $v \in T_x\mathcal{K}$, we have $|Df^j(x)v| \leqslant C|v|$ for all $j \in \mathbb{Z}$. When $v \in E^s_x$ we have: $|Df^j(x)v| \leqslant C\lambda^j|v|$ for all $j \in \mathbb{N}$. In the language of [36,37], we have different growth rates in the tangent space and in the complementary distribution, which are given by $\rho_c = 1$, $\rho_s = \lambda$. In those papers, one can find that the manifold is C^ℓ, where $\ell = \min(-\log(\rho_s)/\log(\rho_c), r)$ where r is the regularity of the map. In our case, $\ell = r$. Note that the theory of regularity of [36,37] does not require any Diophantine property of the rotation.

Remark 12. Of course, the function K, conjugating the motion on the manifold to a rotation, may be less differentiable than the manifold $\mathcal{K}$, because the conjugation of a smooth map to a rotation may be less smooth than the map itself. The Diophantine properties of the rotation play an essential role in this loss of differentiability, but also the dimension n plays a role.

When the dimension $n = 1$, we have the powerful results of [47,83,53], which show that the conjugating map is $C^{r-\tau-\epsilon}$, $\epsilon > 0$ sufficiently small, independently from the map; when $n \geqslant 2$, the smooth conjugacy to rotations has other obstructions (see [48]).

Remark 13. Assume that $n = 1$, the rotation is Diophantine and that the map is analytic, while the conjugacy has to be C^r for any r; then, we can apply the bootstrap of regularity result provided in Theorem 54 to conclude that K is actually analytic and so are the decompositions into spectral bundles.

Remark 14. Very often one considers families depending on other parameters, so that the tori exist for some values of the parameter and not for others. The above considerations show that, when $n = 1$, the rotation is Diophantine and the mapping is analytic, then the conjugacy K has to remain analytic up to the breakdown. The only possibility left by the previous considerations is that the sufficiently smooth norms of the conjugacy K blow up (this was studied in [15]). Furthermore, since the conjugacy cannot break down if the manifolds remain smooth (and this is implied by the hyperbolicity), the only possibility is that the hyperbolicity also breaks down.

On the other hand, the automatic reducibility shows that the Lyapunov exponent is identically λ. Hence, the only way that hyperbolicity can break down is that the angle between the stable space and the tangent to the manifold goes to zero, so that the stable bundle merges with the tangent space. Breakdown of hyperbolicity by the merging of bundles was studied in [45,44]. In the problem at hand, the breakdown of hyperbolicity has been studied numerically in [24]. The numerical studies in [24] lead to remarkable conjectures.

Remark 15. When $n \geqslant 2$, to study the breakdown of the solutions of (3.2), one has also to consider, besides the breakdown of normal hyperbolicity, the phenomenon that happens at the breakdown of the smooth conjugacy of the maps of the circle to rotations. However, this boundary is very poorly understood, even at the level of numerical experiments (except for some particular classes such as linear skew products).

3.4. Automatic reducibility for approximately invariant tori

Of course, in the applications for iterative methods, we want to deal with approximately invariant tori, not with exactly invariant ones. The goal of this section is to show that, for approximately invariant tori, the automatic reducibility found before for invariant tori still holds up to an error that can be estimated by the error in the invariance equation. The precise estimates will be given once we introduce appropriate function spaces to measure the errors.

The procedure to establish these results is very similar to that of Section 3.1. Some of the identities used in Section 3.1 will hold only approximately, but it is important that we can estimate the error by that of the invariance equation. Because the errors in the reducibility are estimated by the error in the invariance, we will show that they do not affect the quadratic convergence of the algorithm, so that for the purposes of the Newton's step, the system can be considered as reducible. We emphasize that the calculations leading to (3.14) just rely on

a) taking derivatives of the invariance equation and then applying algebraic manipulations, which use

b) the conformally symplectic properties of the map,

c) the fact that the torus is Lagrangian.

In the case of an approximate invariant torus, we follow exactly the same algebraic manipulations (they are not geometrically natural and they require to use coordinates); however, a), c) are only

approximate identities and we show that (3.14) holds with an error which can be estimated by the error in the invariance equation. Whenever the rotation number is Diophantine or $\lambda \neq 1$, we show that the following expression holds:

$$Df_\mu \circ K(\theta) M(\theta) = M(\theta + \omega) \begin{pmatrix} \mathrm{Id} & S(\theta) \\ 0 & \lambda\,\mathrm{Id} \end{pmatrix} + R(\theta), \tag{3.20}$$

for a suitable error function $R = R(\theta)$, which will be bounded in Section 7.

The explicit expression for R can be found as follows. Let (3.2) be satisfied with an error $E = E(\theta)$, say

$$f_\mu \circ K(\theta) - K(\theta + \omega) = E(\theta); \tag{3.21}$$

by differentiating (3.21) we obtain

$$Df_\mu \big(K(\theta) \big) DK(\theta) - DK(\theta + \omega) = DE(\theta). \tag{3.22}$$

We will denote by

$$E_L(\theta) \equiv DK(\theta)^T J \circ K(\theta) DK(\theta), \tag{3.23}$$

the error in the Lagrangian character of the torus, which will be later (see Lemma 37b)) bounded by the error in the invariance equation.

If E_L is sufficiently small in the C^0 sense (which, as we will see in Section 7, follows from the smallness in the invariance error in a sufficiently smooth norm), then the spaces $DK(\theta)$ and $V(\theta) \equiv J^{-1} \circ K(\theta) DK(\theta) N(\theta)$ (where N is defined in (3.5)) are transversal and we can write as in (3.7) any vector in a unique way as a linear combination of the columns of $DK(\theta)$ and $V(\theta)$. Hence, there exist uniquely determined functions $\tilde{A}(\theta)$, $\tilde{S}(\theta)$, such that we can write

$$Df_\mu \circ K(\theta) V(\theta) = V(\theta + \omega) \tilde{A}(\theta) + DK(\theta + \omega) \tilde{S}(\theta). \tag{3.24}$$

Of course, in the approximately invariant case, $\tilde{A}$ and $\tilde{S}$ will not have the expressions given in (3.9). Indeed, we want to estimate $\tilde{A}(\theta) - A(\theta)$, $\tilde{S}(\theta) - S(\theta)$, where A, S are given in (3.9). Recall that, by the definition of R in (3.20), and by (3.22), (3.13), we have

$$R(\theta) = \big[DE(\theta) \,\big|\, V(\theta + \omega) \big(\tilde{A}(\theta) - A(\theta) \big) + DK(\theta + \omega) \big(\tilde{S}(\theta) - S(\theta) \big) \big]. \tag{3.25}$$

We now proceed to compute $\tilde{A}$, $\tilde{S}$, so that we can give expressions for the error in reducibility. Multiplying (3.24) on the left by $DK(\theta + \omega)^T J \circ K(\theta + \omega)$, we obtain as in (3.11), just $\lambda\,\mathrm{Id}$ since the calculation does not need any modification. Equating to this the multiplication of the right hand side of (3.24) by the same factor produces

$$\lambda\,\mathrm{Id} = \tilde{A}(\theta) + E_L(\theta + \omega) \tilde{S}(\theta), \tag{3.26}$$

where E_L is the error in the Lagrangian character defined in (3.23). Multiplying on the left both sides of (3.24) by $P(\theta + \omega)^T$ with P defined in (3.8), we obtain:

$$P(\theta + \omega)^T Df_\mu \circ K(\theta) J^{-1} \circ K(\theta) P(\theta) = N(\theta + \omega)^T \gamma(\theta + \omega) N(\theta + \omega) \tilde{A}(\theta) + \tilde{S}(\theta). \tag{3.27}$$

We see that (3.26), (3.27) can be considered as equations for $\tilde{A}, \tilde{S}$, because they determine uniquely such quantities (note that the diagonal terms are the identity and that the upper-diagonal terms in the system (3.26), (3.27) are small).

Solving the system of Eqs. (3.26), (3.27) by a contraction mapping principle, we obtain that the difference between $\tilde{A}, \tilde{S}$ – the approximate solutions of (3.26), (3.27) – and A, S – the exact solutions of (3.26), (3.27) – are bounded by a constant times the error of the approximate solution. That is, we can bound the size of $\tilde{A} - A$, $\tilde{S} - S$ by a constant times the size of E_L. Precise estimates will be given once we have defined appropriate function spaces, but we point out that, since they depend only on just using linear algebra and precise formulas, these estimates will be uniform provided that we take norms which are a Banach algebra under multiplication.

Now that the geometric procedure is specified, we proceed to develop estimates. This will require that we specify some spaces and that we develop estimates for some auxiliary equations, such as difference equations, which we will do in the next section.

4. Estimates on the solutions of the linearized equation

Since the KAM procedure is based on the application of a Newton's method, the estimates on the linearized equation are extremely important.

In our case, we will be concerned with an equation for $\varphi : \mathbb{T}^n \to \mathbb{C}$, given $\eta : \mathbb{T}^n \to \mathbb{C}$, of the form

$$\varphi(\theta + \omega) - \lambda \varphi(\theta) = \eta(\theta), \tag{4.1}$$

where $\lambda \in \mathbb{C}$, $\omega \in \mathbb{R}^n$ are given. Equations of the form (4.1) appeared in many contexts of dynamical systems; when $|\lambda| \neq 1$, they appear often in the study of hyperbolic dynamical systems, while when $|\lambda| = 1$, (4.1) is recognized as the standard small divisor equation. We remark that the cases $|\lambda| = 1$, $|\lambda| \neq 1$ are very different. When $|\lambda| \neq 1$, one can solve (4.1) by an elementary contraction mapping argument, which works for all real vectors ω. When $|\lambda| = 1$, it is well known that the argument is more subtle and, in particular, it depends on the arithmetic properties of ω and $\mathrm{Im}(\log(\lambda))$.

4.1. Several function spaces

In this section we present precise definitions of the norms and some elementary properties of the solutions of (4.1).

In this work we present two types of KAM theorems, one with the analytic estimates and another one with the estimates in Sobolev spaces. In each type of scale of function spaces, we present a theorem that assumes smallness in one space of the scale and we conclude existence of solutions on another space of the same scale. We also present estimates which are uniform in λ, as λ approaches 1 and estimates which assume that $|\lambda| \neq 1$ and are not uniform in λ.

The main results of this section are Lemma 18 and 19, which deal with the solution of (4.1) for the case $|\lambda| \neq 1$ and for the case that applies uniformly for all $\lambda \in \mathbb{R}$, including $\lambda = 1$.

The reason to present the results in two regularity scales is that the Sobolev norms are a rather straightforward byproduct of the algorithms we present here (which provide with the Fourier coefficients). Furthermore, we also present a bootstrap of regularity result, which states that the Sobolev solutions of high enough order are analytic. Another important reason to present estimates on both spaces is that, as shown in [15], we obtain that the breakdown of analytic circles happens when and only when the Sobolev norms of high enough order break up. This criterion is rather practical because it works without any fine-tuning, since it only relies on computing objects which are locally unique. For example, it avoids the computation of periodic orbits, which for many systems appear in a complicated way [35,59]. A comparison of this criterion based on blow up and other methods in the literature to compute breakdown can be found in [21, Appendix B], while implementations can be found in [19,20] for the conservative systems and in [15] for the conformally symplectic systems considered here.

The estimates uniform in λ are analytically more delicate since the difference equations involve small denominators. They also involve some more geometric obstructions. The reason why to include

both cases is that we want to pay particular attention to the case of small dissipation. This is a case that has received a great deal of attention in the applications, especially in Celestial Mechanics, [22, 14,23]. One of the good features of the method presented here is that it allows to continue seamlessly through the Hamiltonian case.

Definition 16. Given $\rho > 0$, we denote by $\mathbb{T}^n_\rho$ the set

$$\mathbb{T}^n_\rho = \left\{ z = x + iy \in \mathbb{C}^n / \mathbb{Z}^n \colon x \in \mathbb{T}^n, \ |y_j| \leqslant \rho, \ j = 1, \ldots, n \right\}.$$

Given $\rho > 0$, we denote by $\mathcal{A}_\rho$ the set of functions which are analytic in $\mathrm{Int}(\mathbb{T}^n_\rho)$ and extend continuously to the boundary. We endow $\mathcal{A}_\rho$ with the norm

$$\|f\|_{\mathcal{A}_\rho} = \sup_{z \in \mathbb{T}^n_\rho} |f(z)|. \tag{4.2}$$

More generally, if $\mathcal{C} \subset \mathbb{C}^n / \mathbb{Z}^n \times \mathbb{C}^n$ is a domain with a smooth boundary, we denote by $\mathcal{A}_\mathcal{C}$ the space of functions which are analytic in the interior of $\mathcal{C}$ and extend continuously to the boundary. We endow $\mathcal{A}_\mathcal{C}$ with the norm

$$\|f\|_{\mathcal{A}_\mathcal{C}} = \sup_{z \in \mathcal{C}} |f(z)|.$$

Given $m > 0$ and denoting the Fourier series of a function $f = f(z)$ as $f(z) = \sum_{k \in \mathbb{Z}^n} \widehat{f}_k \exp(2\pi i k \cdot z)$, we define the space H^m as

$$H^m = \left\{ f : \mathbb{T}^n \to \mathbb{C} \colon \|f\|_m \equiv \left(\sum_{k \in \mathbb{Z}^n} |\widehat{f}_k|^2 \left(1 + |k|^2 \right)^m \right)^{1/2} < \infty \right\}. \tag{4.3}$$

For a vector valued function $f = (f_1, f_2, \ldots, f_j)$, $j \geqslant 1$, we define the norm

$$\|f\|_{\mathcal{X}} = \sqrt{\|f_1\|^2_{\mathcal{X}} + \|f_2\|^2_{\mathcal{X}} + \cdots + \|f_j\|^2_{\mathcal{X}}},$$

where $\mathcal{X}$ is either $\mathcal{A}_\rho$ or H^m. For an $n_1 \times n_2$ matrix valued function F we define

$$\|F\|_{\mathcal{X}} = \sup_{v \in \mathbb{R}^{n_2}_+, \, |v| = 1} \sqrt{ \sum_{i=1}^{n_1} \left(\sum_{j=1}^{n_2} \|F_{ij}\|_{\mathcal{X}} v_j \right)^2 }.$$

Notice that if F is a matrix valued function and f is a vector valued function, then one has

$$\|Ff\|_{\mathcal{X}} \leqslant \|F\|_{\mathcal{X}} \|f\|_{\mathcal{X}}$$

for $\mathcal{X}$ being $\mathcal{A}_\rho$ or H^m with $m > \frac{n}{2}$.

To distinguish clearly between analytic and Sobolev norms we will use the notation $\|f\|_{H^m}$ instead of $\|f\|_m$.

It is well known that $\mathcal{A}_\rho$ and H^m, endowed with their corresponding norms (4.2), (4.3), are Banach spaces. It is also well known that $\mathcal{A}_\rho$ and H^m with $m > n/2$ are Banach algebras under pointwise multiplication [85]:

$$\|fg\|_{\mathcal{A}_\rho} \leqslant C\|f\|_{\mathcal{A}_\rho}\|g\|_{\mathcal{A}_\rho},$$

$$\|fg\|_{H^m} \leqslant C\|f\|_{H^m}\|g\|_{H^m}, \quad m > \frac{n}{2},$$

for a suitable real positive constant C.

It will be essential for the proof of the main result to have estimates on the composition of functions belonging to the Banach algebras introduced above. We include the composition estimates in the following lemma. As stated in the introduction, we use the same letter C for all constants appearing in the forthcoming estimates.

Lemma 17. *The following estimates for the composition of functions in Sobolev spaces and in spaces of analytic functions hold.*

A.1) *Let $f \in C^m$ be the space of functions with m continuous derivatives defined in the whole space. Then, for $g \in H^m \cap L^\infty(\mathbb{T}^n)$ one has:*

$$\|f \circ g\|_{H^m} \leqslant A_m(\|g\|_\infty)\|f\|_{C^m}(1 + \|g\|_{H^m}),$$

where A_m depends on $\|g\|_\infty$; $\|\cdot\|_\infty$ denotes the essential supremum norm.
A.2) *Let $f \in C^{m+2}$, $m > n/2$. Then, for $g, h \in H^m \cap L^\infty(\mathbb{T}^n)$ one has:*

$$\big\|f \circ g - f \circ h - Df \circ h(g - h)\big\|_{H^m} \leqslant \tilde{A}_m(\|g\|_\infty)\big\|D^2 f\big\|_{C^m}\|g - h\|_{H^m}^2.$$

B.1) *Let $f \in \mathcal{A}_{\mathcal{C}}$ be an analytic function on a domain $\mathcal{C} \subset \mathbb{C}^n/\mathbb{Z}^n \times \mathbb{C}^n$, where $\mathcal{C}$ is a compensated domain. Assume that $g = (g_1, \ldots, g_n)$ is such that $g(\mathbb{T}^n_\rho) \subset \mathcal{C}$ and $g_i \in \mathcal{A}_\rho$ with $\rho > 0$. Then $f \circ g \in \mathcal{A}_\rho$ and*

$$\|f \circ g\|_{\mathcal{A}_\rho} \leqslant \|f\|_{\mathcal{A}_{\mathcal{C}}},$$

where $\|f\|_{\mathcal{A}_{\mathcal{C}}} = \sup_{z \in \mathcal{C}} |f(z)|$.
B.2) *Similarly, if g, h are as above*

$$\big\|f \circ g - f \circ h - Df \circ h(g - h)\big\|_{\mathcal{A}_\rho} \leqslant C\big\|D^2 f\big\|_{\mathcal{A}_{\mathcal{C}}}\|g - h\|_{\mathcal{A}_\rho}^2.$$

Proof. A.1) is proven in [85, Section 13.3]. B.1) is obvious from the definition of the analytic norm as supremum.

For the other two cases, we just use the fundamental theorem of calculus to write

$$f \circ g(z) - f \circ h(z) - Df \circ h(z)(g - h)(z) = \int_0^1 dt \int_0^t ds\, D^2 f\big(g_s(z)\big)(g - h)^{\otimes 2}(z), \qquad (4.4)$$

where $g_s(z)$ is a path such that $g_0(z) = h(z)$, $g_1(z) = g(z)$.

In A.2), where we are assuming that the function f is defined everywhere, we can just take $g_s = sg + (1 - s)h$. This formula for g_s also works when $\mathcal{C}$ is a convex domain, but for more general domains we could need a more complicated path and the argument above only gives an estimate in the right hand side of B.2) by the square of the length of the path. The definition of compensated domains [31] is precisely that given any pair of points in the domain, we can find a path that joins them, whose length is not more than a constant times the distance between the points. Of course, all the convex domains are compensated. Note that, for the previous argument, since we are doing pointwise estimates, there is no need that the paths corresponding to different z are related.

Note that the same proof using (4.4) works both in the analytic and in the Sobolev case. Nevertheless, in the Sobolev case, since the Sobolev norms are not just pointwise estimates, one needs that the paths joining two points depend well on the point. Hence, we included the assumption that the functions are defined in the whole space. □

4.2. Estimates on cohomology equations

In this section, we collect estimates on the solutions of cohomology equations of the form (4.1), which are the main tool in KAM theory. The results collected here are very standard. We note that we provide estimates in two kinds of spaces: analytic and Sobolev spaces. We also present two types of estimates, one that corresponds to $|\lambda| \neq 1$ and the other that applies uniformly for all $\lambda \in \mathbb{R}$, including $\lambda = 1$. Roughly, the estimates for $|\lambda| \neq 1$ involve no loss of differentiability, but they include constants that depend on λ. When we present estimates that are valid for all values of λ in an interval including 1, we will use the name *the uniform case*. The estimates for $|\lambda| \neq 1$ will be referred to as *the non-uniform case*.

We also consider the dependence of the solutions on λ, but that is easy by observing that the derivatives with respect to λ also satisfy cohomology equations.

It is interesting to remark that when $|\lambda| \neq 1$, the cohomology equations have a unique solution for all the data. When $\lambda = 1$, they only have solutions for data in a space of codimension 1 (there is one obstruction), but when there is a solution, there is a one-dimensional space of solutions. This is reminiscent of what happens for finite dimensional matrices. Stating unified results for both the uniform and the non-uniform case will give the key for the formulation of the limit of zero dissipation.

Lemma 18. *Assume* $|\lambda| \neq 1$, $\omega \in \mathbb{R}^n$. *Then, given any Lebesgue measurable function* η, *there is one Lebesgue measurable function* φ *satisfying* (4.1). *Furthermore, for* $\rho > 0$, $m > 0$, *the following estimates hold*:

$$\|\varphi\|_{\mathcal{A}_\rho} \leqslant \left||\lambda| - 1\right|^{-1} \|\eta\|_{\mathcal{A}_\rho},$$

$$\|\varphi\|_{H^m} \leqslant \left||\lambda| - 1\right|^{-1} \|\eta\|_{H^m}. \tag{4.5}$$

Finally, one can bound the derivatives of φ *with respect to* λ *as*

$$\left\|D_\lambda^j \varphi\right\|_{\mathcal{A}_\rho} \leqslant \frac{j!}{\left||\lambda| - 1\right|^{j+1}} \|\eta\|_{\mathcal{A}_\rho}, \quad j \geqslant 1,$$

$$\left\|D_\lambda^j \varphi\right\|_{H^m} \leqslant \frac{j!}{\left||\lambda| - 1\right|^{j+1}} \|\eta\|_{H^m}, \quad j \geqslant 1. \tag{4.6}$$

Proof. Note that (4.1) is equivalent to

$$\varphi(\theta - \omega) - \frac{1}{\lambda}\varphi(\theta) = -\frac{1}{\lambda}\eta(\theta - \omega), \tag{4.7}$$

which is of the same form as (4.1), but it involves $1/\lambda$ in place of λ. Therefore, it suffices to consider the case $|\lambda| < 1$. Note that (4.7) implies

$$\varphi(\theta) = \eta(\theta - \omega) + \lambda\varphi(\theta - \omega)$$

$$= \eta(\theta - \omega) + \lambda\eta(\theta - 2\omega) + \lambda^2\varphi(\theta - 2\omega)$$

$$= \sum_{i=0}^{N} \lambda^i \eta\big(\theta - (i+1)\omega\big) + \lambda^{N+1}\varphi\big(\theta - (N+1)\omega\big).$$

Using Lusin's theorem and Poincaré recurrence theorem, it is easy to see that we have almost everywhere $\lim_{N \to \infty} \lambda^{N+1} \varphi(\theta - (N+1)\omega) = 0$. Therefore the only measurable solution of (4.1) is

$$\varphi(\theta) = \sum_{i=0}^{\infty} \lambda^i \eta(\theta - (i+1)\omega). \tag{4.8}$$

Since

$$\left\| \eta(\cdot - (i+1)\omega) \right\|_{\mathcal{A}_\rho} = \left\| \eta(\cdot) \right\|_{\mathcal{A}_\rho}$$

and

$$\left\| \eta(\cdot - (i+1)\omega) \right\|_{H^m} = \left\| \eta(\cdot) \right\|_{H^m},$$

we obtain that (4.8) converges uniformly in $\mathcal{A}_\rho$, H^m, whenever η belongs to these spaces. On spaces $\mathcal{X} = \mathcal{A}_\rho$ or $\mathcal{X} = H^m$, we also have that

$$\|\varphi\|_{\mathcal{X}} \leqslant \left(\sum_{i=0}^{\infty} |\lambda|^i \right) \|\eta\|_{\mathcal{X}},$$

which establishes (4.5).

To study the limit of conservative systems, we note that, taking derivatives with respect to λ of (4.8) and observing that the resulting series converges uniformly on $0 < |\lambda| \leqslant 1 - \varepsilon$ for any $\varepsilon > 0$, we obtain that

$$D_\lambda^j \varphi(\theta) = \sum_{i=j}^{\infty} \frac{i!}{(i-j)!} \lambda^{i-j} \eta(\theta - (i+1)\omega),$$

from which (4.6) follows straightforwardly; note that the estimates (4.5) and (4.6) become singular as $|\lambda| \to 1$. □

Next we consider the case λ in an interval containing 1 and we present the following result, which is standard in KAM theory (see [77]).

Lemma 19. *Consider* (4.1) *for* $\lambda \in [A_0, A_0^{-1}]$ *for some* $0 < A_0 < 1$ *and let* $\omega \in \mathcal{D}(\nu, \tau)$. *Assume that* $\eta \in \mathcal{A}_\rho$, $\rho > 0$ *(respectively* $\eta \in H^m$, $m > \tau$*) and that*

$$\int_{\mathbb{T}^n} \eta(\theta) \, d\theta = 0.$$

Then, there is one and only one solution of (4.1) *with zero average:* $\int_{\mathbb{T}^n} \varphi(\theta) \, d\theta = 0$. *Furthermore, if* $\varphi \in \mathcal{A}_{\rho-\delta}$ *for every* $\delta > 0$ *(resp.* $\varphi \in H^{m-\tau}$ *in the Sobolev estimate), then we have*

$$\|\varphi\|_{\mathcal{A}_{\rho-\delta}} \leqslant C \frac{1}{\nu} \delta^{-\tau} \|\eta\|_{\mathcal{A}_\rho},$$

$$\|\varphi\|_{H^{m-\tau}} \leqslant C \frac{1}{\nu} \|\eta\|_{H^m}, \tag{4.9}$$

where C is a constant that depends on A_0 and the dimension of the space, but it is uniform in λ and it is independent of the Diophantine constant ν.

Note that, when $\lambda = 1$, there are other solutions of (4.1), but all the solutions differ by a constant. The result of Lemma 19 is that the estimates of solutions of cohomology equations normalized to average zero are uniform when λ ranges over an interval that contains 1.

Proof. The analytic bound is established in [77]; in that reference, only the case $\lambda = 1$ is treated in detail, but all other cases can be treated by the same method as well. We note that, if we express η in Fourier coefficients, $\eta(\theta) = \sum_{j \in \mathbb{Z}^n} \widehat{\eta}_j \exp(2\pi i j \cdot \theta)$ and similarly for φ, we see that (4.1) is equivalent to having for all $j \in \mathbb{Z}^n$,

$$\big(\lambda - \exp(2\pi i j \cdot \omega)\big)\widehat{\varphi}_j = \widehat{\eta}_j. \tag{4.10}$$

Clearly, when $\lambda = 1$ and $j = 0$, it is impossible to satisfy (4.10) unless $\widehat{\eta}_0 = 0$. In such a case, we have that $\widehat{\varphi}_0$ is arbitrary. In all other cases, provided that $(\lambda - \exp(2\pi i j \cdot \omega)) \neq 0$, we can find $\widehat{\varphi}_j$ by setting

$$\widehat{\varphi}_j = \big(\lambda - \exp(2\pi i j \cdot \omega)\big)^{-1}\widehat{\eta}_j.$$

Hence, it suffices to estimate the multipliers using Cauchy bounds and (3.1), as it is done in [78] to get (4.9). □

Note that Lemma 19 involves only values of λ ranging over a real interval. A conjecture concerning complex values of λ will be given in Section 10. The result is false for sets that include circular segments of the circle $|\lambda| = 1$, because in such segments there are new small (or zero!) divisors that appear.

5. Statement of the main results

In this section we formulate the main results for maps which we state in Theorem 20. We note that this result is formulated in an *a-posteriori* format, namely we show that if there is a function which solves approximately the invariance equation (3.2) and satisfies some explicit non-degeneracy condition, then there is a true solution which is locally unique. Furthermore, we can bound the difference between the original approximate solution and the exact one by the original error in the invariance equation. It is quite important to note that we do not assume that the system is close to integrable, but only that we have an approximate solution. We note that Theorem 20 involves adjustment of parameters, as it is certainly needed in the dissipative case as indicated before (see [61]). In the conservative case, as it is well known, the adjustment of parameters is not needed and one can choose suitable initial conditions.

A general KAM theory with adjustment of parameters is developed in [70]. Nevertheless, the parameter count of Theorem 20 is somewhat different than the parameter count in [70] because, as shown in Section 3.1, the geometric structures present in our problems produce the automatic reducibility and fix some of the parameters. On the other hand, the requirement that the geometry is fixed imposes extra restrictions.

As noted before, we present statements in analytic and Sobolev norms as well as statements that are uniform in λ as λ approaches 1, and also statements that assume that $|\lambda|$ is away from 1.

In Section 8 we present local uniqueness results as well as some elementary consequences (Lipschitz dependence, measure estimates). In Section 10 we will also show that one can obtain perturbative expansions including the rather singular limit of zero dissipation. We also show that these expansions are convergent when there is some dissipation. In Section 9 we show the bootstrap of the regularity of solutions, from which we obtain a numerically accessible criterion for the study of the boundary of the analyticity domain of the solutions. The criterion roughly asserts that the boundary

of analyticity can be computed by a continuation method deforming the parameters while monitoring some Sobolev norms. This criterion has been already used for dissipative mappings in [15,24] and found to be quite practical and accurate. Of course, even if a method can theoretically reach arbitrarily close to breakdown, in practical implementations one has to deal with limitations in memory precision and time of machines. The results of [15,24] show that the method is quite practical. Similar justifications and implementations in the conservative cases can be found in [19,21,20].

The method of proof of Theorem 20 is to show that if we start a quadratically convergent method, it will converge; the quadratic convergence is used to overcome the small divisors that appear in the iterative step. A detailed formulation of the iterative step will be presented in Section 6. The step is based on some geometric identities *("automatic reducibility"*, already discussed in Section 3.1), which reduce the Newton's step to the solution of the standard difference equations with constant coefficients discussed in Section 4. We note that the automatic reducibility leads to a very efficient numerical algorithm (see Algorithm 33).

We present the estimates both in analytic spaces and in Sobolev spaces. Given the abstract formulation we use, this does not require much more work. We also present a self-contained proof in the analytic case (Section 7.6), which is much easier to read and which could be appropriate for a first reading.

For a function B we denote by $\overline{B}$ its average and by $(B)^0 = B - \overline{B}$.

Theorem 20. H1 *Let $\omega \in \mathcal{D}_n(\nu, \tau)$ according to* (3.1). *Let $\mathcal{M}$ be as in Section* 2.

H2 *Let f_μ with $\mu \in \Lambda$, $\Lambda \subseteq \mathbb{R}^n$ open, be a family of conformally symplectic mappings with respect to a symplectic form Ω, that is $f_\mu^* \Omega = \lambda \Omega$ (see Definition* 1) *with λ constant.*

Let $K_0 : \mathbb{T}^n \to \mathcal{M}$, $\mu_0 \in \mathbb{R}^n$ and define E, such that

$$f_{\mu_0} \circ K_0 - K_0 \circ T_\omega = E.$$

H3 *Assume that the following non-degeneracy condition holds*:

$$\det \begin{pmatrix} \overline{S} & \overline{S(B_b)^0} + \overline{\widetilde{A}_1} \\ (\lambda - 1)\,\mathrm{Id} & \overline{\widetilde{A}_2} \end{pmatrix} \neq 0, \tag{5.1}$$

where the $n \times n$ matrix S is an algebraic expression involving derivatives of K_0 written explicitly in (3.9). *The $n \times n$ matrices $\widetilde{A}_1, \widetilde{A}_2$ denote the first n and the last n columns of the $2n \times n$ matrix $\tilde{A} = M^{-1} \circ T_\omega D_{\mu_0} f_{\mu_0} \circ K_0$, where M is written explicitly in* (3.13). *The $n \times n$ matrix $(B_b)^0$ is the solution (with zero average) of $\lambda(B_b)^0 - (B_b)^0 \circ T_\omega = -(\widetilde{A}_2)^0$. We denote by*

$$\mathcal{T} \equiv \left\| \begin{pmatrix} \overline{S} & \overline{S(B_b)^0} + \overline{\widetilde{A}_1} \\ (\lambda - 1)\,\mathrm{Id} & \overline{\widetilde{A}_2} \end{pmatrix}^{-1} \right\|$$

and we refer to $\mathcal{T}$ as the twist *constant.*

A) *Analytic case*:

Assume **H1**–**H3** *and that $K_0 \in \mathcal{A}_\rho$ for some $\rho > 0$. Assume furthermore that for $\mu \in \Lambda$ we have that f_μ is a C^1-family of analytic functions on a domain – open connected set – $\mathcal{C} \subset \mathbb{C}^n / \mathbb{Z}^n \times \mathbb{C}^n$ with the following assumption on the domain.*

H4 *There exists a $\zeta > 0$, so that*

$$\mathrm{dist}(\mu_0, \partial \Lambda) \geqslant \zeta,$$

$$\mathrm{dist}\big(K_0(\mathbb{T}_\rho^n), \partial \mathcal{C}\big) \geqslant \zeta.$$

Furthermore, assume that the solution is sufficiently approximate in the following sense.

H5 *We assume that, for some $0 < \delta < \rho/2$, E satisfies the inequality*

$$\|E\|_{\mathcal{A}_\rho} \leqslant C\nu^{2\ell}\delta^{2\ell\tau};$$

*here and below C denotes a constant that can depend on τ, n, $\mathcal{T}$, $\|DK_0\|_\rho$, $\|N\|_\rho$, $\|M\|_\rho$ $\|M^{-1}\|_\rho$, where N, M are defined in (3.5), (3.13) as well as on ζ entering in **H4**. In such a case, ℓ takes the value 2. If we allow C to depend on λ with $|\lambda| \neq 1$, we can take $\ell = 1$.*

Then, there exists μ_e, K_e such that

$$f_{\mu_e} \circ K_e - K_e \circ T_\omega = 0. \tag{5.2}$$

The quantities K_e, μ_e satisfy

$$\|K_e - K_0\|_{\mathcal{A}_{\rho-\ell\delta}} \leqslant C\nu^{-\ell}\delta^{-\ell\tau}\|E\|_{\mathcal{A}_\rho},$$

$$|\mu_e - \mu_0| \leqslant C\|E\|_{\mathcal{A}_\rho}.$$

B) *Finitely differentiable, Sobolev case*:

*Assume **H1**–**H3** and that for $\mu \in \Lambda$, Λ being an open set in $\mathbb{R}^n$, we have that f_μ is a C^1-family of C^r functions (where $r \geqslant m + 13\ell\tau + 2$ with $m > \frac{n}{2} + \ell\tau$ and ℓ specified below) on a domain – open connected set – $\mathcal{C} \subset \mathbb{T}^n \times \mathbb{R}^n$.*

H4' *Assume that there exists a $\zeta > 0$, so that*

$$\operatorname{dist}(\mu_0, \partial\Lambda) \geqslant \zeta,$$

$$\operatorname{dist}(K_0(\mathbb{T}^n), \partial\mathcal{C}) \geqslant \zeta.$$

Furthermore, assume that the solution $K_0 \in H^{m+13\ell\tau}$ is sufficiently approximate in the following sense.
H5' *Assume that E and $\varepsilon^* > 0$ satisfy the inequality*

$$\|E\|_{H^{m-\ell\tau}} \leqslant \varepsilon^*,$$

where $\varepsilon^* = \varepsilon^*(\tau, \nu, n, \mathcal{T}, \|DK_0\|_{H^m}, \|N\|_{H^m}, \|M\|_{H^m}, \|M^{-1}\|_{H^m})$ is an explicit function.
Then, there exists K_e, μ_e satisfying (5.2) and such that they also satisfy the following distance bounds:

$$\|K_e - K_0\|_{H^m} \leqslant C\|E\|_{H^{m-\ell\tau}},$$

$$|\mu_e - \mu_0| \leqslant C\|E\|_{H^{m-\ell\tau}},$$

with $\ell = 2$ in the uniform case, while $\ell = 1$ if C depends on λ with $|\lambda| \neq 1$.

Remark 21. We note that the properties on the function f enter only very mildly, since it suffices to find bounds on some of its derivatives. Of course, in the analytic case, one can obtain the derivatives from estimates on the size in a slightly bigger domain.

Remark 22. Notice that the non-degeneracy condition **H3** has just the structure of the inverse of a matrix obtained averaging algebraic expressions of the derivatives of K_0. This makes it reasonably clear that if we modify K_0 slightly, the condition **H3** will also hold and that in a neighborhood of K_0 we can maintain the twist constant uniformly bounded. In particular, when we make small modifications by iterating the Newton procedure, the twist constant can be maintained. More details and quantitative estimates of the change will be developed in the proof.

Remark 23. Notice that the non-degeneracy condition in **H3** has a well-defined meaning when λ approaches 1, since for $\lambda = 1$, (5.1) just amounts to $\det(\overline{S}) \neq 0$, which is the standard Kolmogorov twist condition in KAM theory (see [30]), and $\det(\widetilde{\overline{A_2}}) \neq 0$, which is just the non-degeneracy of the family with respect to parameters.

Remark 24. When $\lambda = 1$, the existence of invariant tori requires that Ω is an exact form and that the mapping f is exact. For λ close to 1 one does not need that Ω is an exact form, nor that f is exact.

The reason is that the exactness comes into the proof because the automatic reducibility requires that the approximately invariant torus is approximately Lagrangian. This is proved showing that $K^*\Omega$ solves a cohomology equation with a small right hand side. In the case $|\lambda| \neq 1$ this is indeed enough to show that $K^*\Omega$ is small. In the $\lambda = 1$ case, this cohomology equation only allows to conclude that $K^*\Omega$ is almost constant and we need to use the exactness to conclude that the constant is zero.

In the case that Ω is exact, we can see that the existence of an approximately invariant torus implies that the map is approximately exact. The non-degeneracy condition **H3**, includes that we can change the cohomology of the map by changing μ. Hence, since f_μ is approximately exact, using **H3** we can make a small change of parameters so that the mapping becomes exact. This choice of parameters is implicit in the procedure. We see that, as λ approaches 1, the parameters μ approach zero, so that f_μ gets to be exact.

The solutions produced by Theorem 20 are essentially unique; as already remarked, the only possibility to get different but close solutions of the invariance equation is to change the origin in the parameterization. This will be made more explicit in Theorem 29.

The main idea in the proof Theorem 20 (in the notation of [90]) is that the linearized equation has a right inverse. To obtain the uniqueness result of Theorem 29 we remark that there is also a left inverse.

5.1. Uniqueness results

5.1.1. A preliminary normalization

In order to deal with the non-uniqueness pointed out in Remark 5, we note that it is possible to impose an extra normalization (see (5.3) below) for all possible candidates to a solution in a neighborhood of the solution. We note that the proof that the normalization can be achieved is elementary and it only uses the standard implicit function theorem. Hence, at the only price of complicating slightly the proximity assumptions in the statement of Theorem 29, one can formulate Theorem 29 without involving the normalization (see Remark 27).

The normalization (5.3) below also plays a role in the study of perturbative expansions, see Section 10. Of course, to discuss dependence on parameters, one needs to eliminate arbitrary choices, so that some normalization that makes the solutions unique is needed. We also note that some variants of this normalization are easy to impose in the algorithms that we are going to discuss.

Our chosen normalization is as follows. We want that the function $K_2^{(\sigma)} = K_2 \circ T_\sigma$ satisfies

$$\int_{\mathbb{T}^n} \left[M^{-1}(\theta) \left(K_2^{(\sigma)}(\theta) - K_1(\theta) \right) \right]_1 d\theta = 0, \tag{5.3}$$

where the subindex 1 in the braces means taking the first n components. In other words, if we write $K_2^{(\sigma)} - K_1 = MW_\sigma$, we are imposing that $[W_\sigma]_1$ has zero average.

Remark 25. The geometric meaning of the lack of uniqueness pointed out in Remark 5 is just that we can shift the origin of coordinates in the reference torus we are embedding. The geometric meaning of the normalization (5.3) is precisely that we make a definite choice of the phase. We note that the torus K_1 admits a natural system of coordinates. The integral in (5.3) can be understood geometrically as the average phase of $K_2^{(\sigma)}$ in the coordinates associated to K_1 in Section 3.2. Since the torus K_2 is

horizontal in the coordinates given by K_1 (their tangent spaces are very close), we see that a change in the phase of the embedding K_2 changes the average phase K_2 and we can obtain the normalization (5.3) applying the implicit function theorem. As we will see later, the fact that the shift in phase is the only non-uniqueness of the solution is related to the fact that we characterize the non-uniqueness of the solutions of the linearized invariance equation.

The normalization equation (5.3) can be considered as the average over the angle coordinates of the difference in the adapted coordinates (introduced in Section 3.2) in the neighborhood of K_1.

Proposition 26. *Let K_1, K_2 be solutions of (3.2), $\|K_1 - K_2\|_{C^1}$ be sufficiently small (with respect to quantities depending only on M – computed out of K_1 – and f). Then, there exists $\sigma \in \mathbb{R}^n$, such that $K_2^{(\sigma)} = K_2 \circ T_\sigma$ satisfies (5.3). Furthermore:*

$$|\sigma| \leqslant C \|K_1 - K_2\|_{C^0}, \tag{5.4}$$

where C can be chosen to be as close to 1 as desired by assuming that f_μ, K_1, K_2 are twice differentiable, $DK_1^T DK_1$ is invertible and $\|K_1 - K_2\|_{C^0}$ is sufficiently small.

The σ thus chosen is locally unique.

Proof. The proof is an easy application of the finite dimensional implicit function theorem.

We realize that, expressing W_σ in the coordinates given by M, we have

$$K_2^{(\sigma)} - K_1 = DK_1[W_\sigma]_1 + J^{-1} \circ K_1 DK_1 N_1[W_\sigma]_2.$$

If we take derivatives with respect to σ, we obtain:

$$D_\sigma K_2^{(\sigma)} = DK_1[D_\sigma W_\sigma]_1 + J^{-1} \circ K_1 DK_1 N_1[D_\sigma W_\sigma]_2.$$

Multiplying the left hand side by $DK_1^T J \circ K_1$ and by DK_1^T, using that the torus K_1 is Lagrangian and that $D_\sigma K_2^{(\sigma)} = DK_2 \circ T_\sigma$, we get

$$DK_1^T J \circ K_1 DK_2 \circ T_\sigma = [D_\sigma W_\sigma]_2,$$
$$DK_1^T DK_2 \circ T_\sigma = DK_1^T DK_1[D_\sigma W_\sigma]_1 + \gamma_1 N_1[D_\sigma W_\sigma]_2,$$

where γ_1 is as in (3.10) corresponding to the torus K_1. By (3.6) we obtain that $[D_\sigma W_\sigma]_2$ is small, if DK_1 is close to DK_2 and σ is small; under these conditions we obtain that $[D_\sigma W_\sigma]_1$ is close to the identity. An application of the implicit function theorem concludes the proof. In fact, if we define a function $F(x, \sigma) = \int_{\mathbb{T}^n}[M^{-1}(\theta)(K_2^{(\sigma)}(\theta) - K_1(\theta) - x)]_1 d\theta$, we recognize that $F(K_2 - K_1, 0) = 0$. Moreover,

$$D_\sigma F(x, \sigma) = \int_{\mathbb{T}^n} [D_\sigma W_\sigma]_1 d\theta$$

is close to the identity if K_2 is close to K_1. Applying the implicit function theorem, there exists a differentiable function $u = u(x)$, such that $F(x, u(x)) = 0$; from $u(x) = \sigma$ one obtains the estimate (5.4). $\square$

Remark 27. Notice that the normalization (5.3) does not use at all that K_1 is a solution of the invariance equation (3.2). We just use that M is invertible and that the inverse function theorem for the average can be used. Hence, it works just as well when K_1 is an approximate solution.

Remark 28. Note that from the inequality

$$\|K_2 - K_2 \circ T_\sigma\| \leqslant \|DK_2\| |\sigma| \leqslant C\|DK_2\| \|K_1 - K_2\|_{C^0},$$

we can derive bounds on $\|K_1 - K_2 \circ T_\sigma\|$ from bounds on $\|K_1 - K_2\|$.

The statement of Theorem 29 will be done under the assumption that the solution K_2 is normalized with respect to K_1. In numerical applications computing the shift is not very difficult and for concrete solutions one can get better estimates using direct numerical calculations, than those obtained using the triangle inequality as above.

5.1.2. Statement of the local uniqueness theorem

Theorem 29. *Let $\omega \in \mathcal{D}_n(\nu, \tau)$ according to (3.1). Let f_μ be a family of conformally symplectic mappings satisfying Definition 1 with λ constant. Let (K_1, μ_1), (K_2, μ_2) be solutions of (3.2). Assume also that K_2 satisfies (5.3).*

In the analytic case, let f_μ be a C^1-family of analytic functions, satisfying the non-degeneracy condition **H3** *at K_1, μ_1 and assume that assumption* **H4** *about the domain of f_{μ_1} and the range of K_1 holds. Let $M_+ \equiv \max(\|M\|_{\mathcal{A}_\rho}, 1)$, $M_- \equiv \max(\|M^{-1}\|_{\mathcal{A}_\rho}, 1)$. Assume that we have the following inequality:*

$$C\|D_\mu D_{(I,\varphi)} f_\mu\|_{\mathcal{A}_C} \nu^{-\ell} \delta^{-\ell\tau} M_+^3 M_- \max\big(\|W\|_{\mathcal{A}_{\rho+\ell\delta}}, |\mu_1 - \mu_2|\big) < 1. \tag{5.5}$$

Then, we have

$$K_1 = K_2, \qquad \mu_1 = \mu_2.$$

In the Sobolev case, let f_μ be a C^1-family of C^r functions, $r \geqslant m + 2$, $m > n/2 + \ell\tau$ satisfying the non-degeneracy condition **H3** *at K_1, μ_1. Let $M_+ \equiv \max(\|M\|_{H^m}, 1)$, $M_- \equiv \max(\|M^{-1}\|_{H^m}, 1)$, where M has been defined in (3.13). Assume that we have the following inequality:*

$$C\|D_\mu D_{(I,\varphi)} f_\mu\|_{C^m} \nu^{-\ell} M_+^3 M_- \max\big(\|W\|_{H^{m+\ell\tau}}, |\mu_1 - \mu_2|\big) < 1. \tag{5.6}$$

Then, we have

$$K_1 = K_2, \qquad \mu_1 = \mu_2.$$

In both analytic and Sobolev cases, we can take $\ell = 1$ if we allow the constants to depend on λ with $|\lambda| \neq 1$ and $\ell = 2$ if we allow the constants to be independent on λ.

The proof of Theorem 29 is given in Section 8. Of course, given Proposition 26 if we assume just that the solutions are sufficiently close (in a slightly stronger sense), we can assume that there is a normalized solution which is also normalized. Then Theorem 29 concludes that there exists $\sigma \in \mathbb{R}^n$ such that $K_1 = K_2 \circ T_\sigma$.

5.1.3. Some straightforward conclusions of uniqueness: Lipschitz dependence on parameters, measure estimates

An easy corollary of Theorems 20 and 29 is that if we consider a family of maps, which depends in a Lipschitz way on a parameter, then we obtain a Lipschitz dependence of the solution with respect to the parameter. Later, in Section 10.3, we will obtain sharper conclusions of differentiability on parameters, assuming, of course, that the problem is differentiable with respect to parameters. This is closely related to the existence and convergence of perturbative expansions.

Corollary 30. *Assume that the family $f_{\mu,\phi}$ depends also on a parameter ϕ, belonging to a metric space $(\mathcal{Y}, d)$, and assume that for each value of ϕ, the map $f_{\mu,\phi}$ satisfies the hypotheses of Theorem 20 with uniform constants. Assume also a Lipschitz dependence with respect to the parameter ϕ for $m > n/2 + \ell\tau$ for a fixed K. That is, assume that for a $\phi_0 \in \mathcal{Y}$ with ϕ, ϕ' sufficiently close to ϕ_0 and for all K in a sufficiently small neighborhood of K_{ϕ_0}, in the corresponding spaces we have:*

$$\| f_{\mu,\phi} \circ K - f_{\mu,\phi'} \circ K \|_{\mathcal{A}_\rho} \leqslant \tilde{A}_L \, d(\phi, \phi'),$$

$$\| f_{\mu,\phi} \circ K - f_{\mu,\phi'} \circ K \|_{H^m} \leqslant \tilde{A}_L \, d(\phi, \phi') \tag{5.7}$$

for a suitable constant $\tilde{A}_L$. Then, there exists a constant A_L such that the solution (ϕ, K_ϕ, μ_ϕ) of the invariance equation

$$f_{\mu_\phi,\phi} \circ K_\phi = K_\phi \circ T_\omega,$$

produced by applying Theorem 20 (and normalized so that we obtain uniqueness) is Lipschitz with respect to the parameter ϕ with a constant $C\tilde{A}_L$, i.e.

$$\| K_\phi - K_{\phi'} \|_{\mathcal{A}_{\rho-\ell\delta}} \leqslant C \tilde{A}_L \nu^{-\ell} \delta^{-\ell\tau} \, d(\phi', \phi)$$

in the analytic case with δ as in Theorem 20 and

$$\| K_\phi - K_{\phi'} \|_{H^m} \leqslant C \tilde{A}_L \, d(\phi', \phi)$$

in the Sobolev case with

$$|\mu_\phi - \mu_{\phi'}| \leqslant C \tilde{A}_L \, d(\phi', \phi),$$

where $\ell = 2$ in the uniform case and $\ell = 1$ if we allow C to depend on λ with $|\lambda| \neq 1$.

   The proof of Corollary 30 (see Section 8) relies on the remark that if (ϕ, K_ϕ, μ_ϕ) is a solution of (3.2) corresponding to the parameter ϕ, we have

$$\left\| f_{\mu_\phi,\phi'} \circ K_\phi - K_\phi \circ T_\omega \right\|_{\mathcal{A}_\rho} = \left\| f_{\mu_\phi,\phi'} \circ K_\phi - f_{\mu_\phi,\phi} \circ K_\phi \right\|_{\mathcal{A}_\rho} \leqslant \tilde{A}_L \, d(\phi, \phi').$$

Then, applying Theorem 20 and the local uniqueness for normalized solutions, Theorem 29, we obtain Corollary 30.

   A simple consequence of Corollary 30 is the following.

Corollary 31. *Assume that in the hypotheses of Theorem 20, we have an exact solution $f_{\mu_0} \circ K_0 - K_0 \circ T_{\omega_0} = 0$ with $K_0 \in \mathcal{A}_{\rho+\delta}$ ($K_0 \in H^{m+1}$, $m > n/2 + \ell\tau$), $\omega_0 \in \mathcal{D}_n(\nu, \tau)$. Fix $0 < \delta < \rho$; then, for $s > 0$ sufficiently small and for all $\omega \in \tilde{\mathcal{D}}_s \equiv \mathcal{D}_n(\nu, \tau) \cap B_s(\omega_0)$ (we denote by $B_s(\omega_0) \subset \mathbb{R}^d$, the ball of radius s around ω_0), there exist $K_\omega \in \mathcal{A}_{\rho-\ell\delta}$ ($K_\omega \in H^m$), $\mu_\omega \in \mathbb{R}^n$ such that*

$$f_{\mu_\omega} \circ K_\omega - K_\omega \circ T_\omega = 0,$$

and K_ω is normalized according to (5.3).

   Furthermore, the mapping $\omega \to (K_\omega, \mu_\omega)$ is Lipschitz, when considered as a mapping from the closed set $\tilde{\mathcal{D}}_s$ to $\mathcal{A}_{\rho-\ell\delta} \times \mathbb{R}^n$ ($H^m \times \mathbb{R}^n$), where $\ell = 2$ in the uniform case and $\ell = 1$ if we allow the Lipschitz constant to depend on λ with $|\lambda| \neq 1$.

We note that there are more sophisticated arguments that give that the dependence on the frequency is differentiable in the Whitney sense [88].

The proof of Corollary 31 is simply to observe that

$$\| f_{\mu_0} \circ K_0 - K_0 \circ T_\omega \|_{\mathcal{A}_\rho} = \| K_0 \circ T_{\omega_0} - K_0 \circ T_\omega \|_{\mathcal{A}_\rho} \leqslant C\delta^{-1} \| K_0 \|_{\mathcal{A}_{\rho+\delta}} |\omega - \omega_0|$$

in the analytic case and

$$\| f_{\mu_0} \circ K_0 - K_0 \circ T_\omega \|_{H^m} = \| K_0 \circ T_{\omega_0} - K_0 \circ T_\omega \|_{H^m} \leqslant C \| K_0 \|_{H^{m+1}} |\omega - \omega_0|$$

in the Sobolev case. Hence, when $|\omega - \omega_0|$ is small enough, we can take K_0, μ_0 as an approximate solution of the invariance equation for the frequency ω, so that the error in the invariance can be made arbitrarily small by making $|\omega - \omega_0|$ sufficiently small. We also note that the non-degeneracy conditions depend only on K_0, μ_0 and therefore they are uniform with the choices of ω.

It then follows from the application of Theorem 20 that we can have a solution K_ω, μ_ω. Furthermore, we have that $\| K_\omega - K_0 \|_{\mathcal{A}_{\rho-\ell\delta}}$ ($\| K_\omega - K_0 \|_{H^m}$) and $|\mu_\omega - \mu_0|$ are smaller than $C|\omega - \omega_0|$. It is important that we choose K normalized, for example satisfying the normalization (5.3).

Now, we observe that we can apply the argument again to all $\omega, \omega' \in \tilde{\mathcal{D}}_s$ and, eventually redefining the constants and making the smallness conditions stronger, we obtain

$$\| K_\omega - K_{\omega'} \|_{\mathcal{A}_{\rho-\ell\delta}}, \| K_\omega - K_{\omega'} \|_{H^m}, |\mu_\omega - \mu_0| \leqslant C|\omega - \omega'|.$$

An immediate consequence of Corollary 30 is the following.

Corollary 32. *In the conditions of Theorem 20, there is a positive measure set of μ, such that there is a K and an ω satisfying* (3.2).

The proof of Corollary 32 is just to observe that the set of Diophantine points is a set of positive measure. Since the mapping $\omega \to \mu_\omega$ is bi-Lipschitz, it sends a set of positive measure into a set of positive measure.

Notice that, due to Proposition 10, the solutions of (3.2) are attractors. Hence, we show that the set of parameters μ for which the attractor is quasi-periodic has positive measure. Furthermore, the set described by Corollary 32 is bi-Lipschitz equivalent to the set of Diophantine numbers $\mathcal{D}_n(\nu, \tau)$. This is a geometrically complicated set full of gaps. In the gaps of this set (which can be large in some topological sense) the attractors could have a dynamics more complicated than quasi-periodic (e.g., strange attractors, see [61]).

6. Formulation of the iterative step in the proof of Theorem 20

In this section we formulate the iterative step of the Newton's method and we argue that it leads to a fast and efficient algorithm (we present the algorithm in Section 6.2.2). Estimates showing that the error after a Newton's step is quadratic in the original errors (using the appropriate norms) will be developed in Section 7. Given these estimates, standard KAM theory ensures that, if the initial error is small enough, then the iteration procedure can be repeated indefinitely and it converges to a solution which is close to the initial approximation. In Section 7 we provide estimates on the dependence on parameters, characterizing the limit of small dissipation. The efficiency of the iterative step is due to the fact that we take advantage of some identities of geometric origin as described in Section 3.1.

6.1. The Newton's equation for the invariance equation

We start with an approximate solution of (3.2) up to an error term E, say

$$f_\mu \circ K - K \circ T_\omega = E, \tag{6.1}$$

where E is supposed to be small. Newton's method consists in finding corrections Δ, σ to K and μ respectively, such that the linear approximation of the transformation associated to $K + \Delta$, $\mu + \sigma$ quadratically reduces the error. Taking into account that

$$f_{\mu+\sigma} \circ (K + \Delta) = f_\mu \circ K + [Df_\mu \circ K]\Delta + [D_\mu f_\mu \circ K]\sigma + O\left(\|\Delta\|^2\right) + O\left(|\sigma|^2\right),$$

the resulting equation is

$$[Df_\mu \circ K]\Delta - \Delta \circ T_\omega + [D_\mu f_\mu \circ K]\sigma = -E. \tag{6.2}$$

6.2. A method to find approximate solutions of the linearized equation (6.2). A quasi-Newton method

Eq. (6.2) is not easy to solve, since it involves the unknown function Δ evaluated at different points and, moreover, the factor $Df_\mu \circ K$ appearing in the first term is not constant.

We will not solve (6.2) exactly, but we will find approximate solutions that still lead to a quadratically convergent procedure.

The main idea to find approximate solutions of (6.2) is to use the geometric identities developed in Section 3.1. Using the matrix valued function M introduced in (3.13), we change variables in (6.2) by setting

$$\Delta = MW \tag{6.3}$$

and we seek W instead of Δ. Note that in the iterative step M is known because it is an explicit expression (given in (3.13)) involving derivatives of K, which is known.

The geometric meaning of (6.3) is that M defines a frame of vector fields that transforms very simply under the map f_μ. The new unknown W is just the expression of Δ in the coordinates given by M. Using (6.3) we have that Eq. (6.2) is equivalent to

$$Df_\mu \circ KMW - (M \circ T_\omega)(W \circ T_\omega) + D_\mu f_\mu \circ K\sigma = -E;$$

using (3.20) one obtains that (6.2) is equivalent to:

$$M \circ T_\omega \left[\begin{pmatrix} \mathrm{Id} & S(\theta) \\ 0 & \lambda\,\mathrm{Id} \end{pmatrix} W - W \circ T_\omega \right] + D_\mu f_\mu \circ K\sigma = -E - RW, \tag{6.4}$$

where R is the error in (3.20). Since we expect that Δ, and therefore W, are estimated by E and, as we argued before, so is R, we obtain that the term RW in (6.4) is quadratic in E; therefore, we expect that we can omit this term without changing the quadratic nature of the method. Precise estimates for R and for the other quantities will be established in Section 7.

Our iterative step consists in solving the following equation (6.5), obtained dropping the term RW from (6.4):

$$\begin{pmatrix} \mathrm{Id} & S(\theta) \\ 0 & \lambda\,\mathrm{Id} \end{pmatrix} W - W \circ T_\omega = -M^{-1} \circ T_\omega E - M^{-1} \circ T_\omega D_\mu f_\mu \circ K\sigma. \tag{6.5}$$

As we will see, (6.5) reduces to difference equations with constant coefficients, so that it can be solved very efficiently by using Fourier methods. Eq. (6.5) can be expressed in components as

$$W_1 - W_1 \circ T_\omega = -SW_2 - \widetilde{E}_1 - (\widetilde{A}\sigma)_1,$$
$$\lambda W_2 - W_2 \circ T_\omega = -\widetilde{E}_2 - (\widetilde{A}\sigma)_2, \tag{6.6}$$

where, to simplify the notation, we have written

$$\widetilde{E} = M^{-1} \circ T_\omega E \equiv (\widetilde{E}_1, \widetilde{E}_2),$$
$$\widetilde{A} = M^{-1} \circ T_\omega D_\mu f_\mu \circ K. \tag{6.7}$$

Note that $\widetilde{A}$ is a $2n \times n$ matrix that we write as $\widetilde{A} = [\widetilde{A}_1 | \widetilde{A}_2]$ with $\widetilde{A}_1, \widetilde{A}_2$ being $n \times n$ matrices.

The system of equations (6.6) has an upper triangular structure. The second equation involves only W_2, but the first equation involves W_1 and W_2. So, it is natural to try to solve first the equation for W_2, substitute in the first equation of (6.6) and then find W_1.

Note that both equations in (6.6) for the unknowns W_1, W_2 are cohomology equations of the form studied in Section 4.2. The main subtlety of the procedure comes from the fact that these equations involve small divisors and obstructions which will be accommodated by choosing parameters. These obstructions and choices of parameters are different when $\lambda = 1$ and when $|\lambda| \neq 1$, but we will present a choice that works uniformly in both cases. We will first discuss these choices of parameters; once we do that, the discussion of estimates will become an application of the results of Section 4.2.

6.2.1. The choice of parameters

When $|\lambda| \neq 1$, the second equation can always be solved and has a unique solution, while for any λ the first equation involves small divisors and it requires that the right hand side has zero average. On the other hand, when $\lambda = 1$, both equations involve small divisors and require that the right hand side has zero average, but the solution of W_2 is not unique and it admits an arbitrary constant. In some way, the freedom of having an arbitrary average for W_2 compensates the extra obstruction required for $\lambda = 1$.

We now study the problem systematically. Given a function B we denote by $\overline{B}$ its average and by $(B)^0 = B - \overline{B}$, the no-average part.

We can divide (6.6) into two systems, one for the average and another one for the no-average part:

$$0 = -\overline{S}\,\overline{W_2} - \overline{S(W_2)^0} - \overline{\widetilde{E}_1} - \overline{\widetilde{A}_1}\sigma,$$
$$(\lambda - 1)\overline{W_2} = -\overline{\widetilde{E}_2} - \overline{\widetilde{A}_2}\sigma, \tag{6.8}$$

$$(W_1)^0 - (W_1)^0 \circ T_\omega = -(SW_2)^0 - (\widetilde{E}_1)^0 - (\widetilde{A}_1)^0\sigma,$$
$$\lambda(W_2)^0 - (W_2)^0 \circ T_\omega = -(\widetilde{E}_2)^0 - (\widetilde{A}_2)^0\sigma. \tag{6.9}$$

Unfortunately, the two systems (6.8) and (6.9) are not completely uncoupled due to the term $\overline{S(W_2)^0}$ appearing in the first equation of (6.8). Nevertheless, it is easy to uncouple the system because $(W_2)^0$ is an affine function of σ, since it satisfies (6.9). We define $(B_a)^0$, $(B_b)^0$ to be the zero average solutions of, respectively,

$$\lambda(B_a)^0 - (B_a)^0 \circ T_\omega = -(\widetilde{E}_2)^0,$$
$$\lambda(B_b)^0 - (B_b)^0 \circ T_\omega = -(\widetilde{A}_2)^0. \tag{6.10}$$

These equations (6.10) are readily solvable using the lemmas in Section 4.2. In particular, note that we have estimates which are uniform in λ as λ goes through 1.

We therefore see that we can transform (6.8) into

$$0 = -\overline{S}\,\overline{W_2} - \overline{S(B_a)^0} - \overline{S(B_b)^0}\sigma - \overline{\overline{E}_1} - \overline{\overline{A}_1}\sigma,$$
$$(\lambda - 1)\overline{W_2} = -\overline{\overline{\overline{E}}_2} - \overline{\overline{A}_2}\sigma. \tag{6.11}$$

The system (6.11), in spite of its typographically formidable appearance, is a finite dimensional system that we can write as

$$\begin{pmatrix} \overline{S} & \overline{S(B_b)^0} + \overline{\overline{A}_1} \\ (\lambda - 1)\,\mathrm{Id} & \overline{\overline{A}_2} \end{pmatrix} \begin{pmatrix} \overline{W_2} \\ \sigma \end{pmatrix} = \begin{pmatrix} -\overline{S(B_a)^0} - \overline{\overline{E}_1} \\ -\overline{\overline{\overline{E}}_2} \end{pmatrix}. \tag{6.12}$$

The non-degeneracy condition that we assumed in **H3** is precisely that the determinant of the matrix at the left hand side of (6.12) is not zero.

In summary, the algorithm to perform a quasi-Newton step is to: 1) form the auxiliary quantities entering into (6.11), 2) solve for $\overline{W_2}$, σ, 3) solve (6.9). In Section 6.2.2 we will present the algorithm and in Section 7 we will present estimates.

6.2.2. The algorithm for the improved approximation

The procedure described before in Section 6.2 leads to the algorithm described below for a given Diophantine frequency ω, where each step is denoted as follows: "$a \leftarrow b$" means that the quantity a is assigned by the quantity b.

Assume that we are given a family f_μ, and that we can compute Df_μ, $D_\mu f_\mu$. The family is quasi-conformal so that $f_\mu^* \Omega = \lambda \Omega$. The following algorithm computes an improved approximation for any $\lambda \in \mathbb{R}$.

Algorithm 33. *Given $K : \mathbb{T}^n \to \mathcal{M}$, $\mu \in \mathbb{R}^n$, we denote by $\lambda \in \mathbb{R}$ the conformal factor for f_μ. We perform the following computations*:

1) $E \leftarrow f_\mu \circ K - K \circ T_\omega$
2) $\alpha \leftarrow DK$
3) $N \leftarrow [\alpha^T \alpha]^{-1}$
4) $M \leftarrow [\alpha, J^{-1} \circ K\alpha N]$
5) $\beta \leftarrow M^{-1} \circ T_\omega$
6) $\widetilde{E} \leftarrow \beta E$
7) $P \leftarrow \alpha N$
   $A \leftarrow \lambda\,\mathrm{Id}$
   $\gamma \leftarrow \alpha^T J^{-1} \circ K\alpha$
   $S \leftarrow (P \circ T_\omega)^T Df_\mu \circ K J^{-1} \circ K P - (N \circ T_\omega)^T (\gamma \circ T_\omega)(N \circ T_\omega) A$
   $\tilde{A} \leftarrow M^{-1} \circ T_\omega D_\mu f_\mu \circ K$
8) $(B_a)^0$ solves $\lambda (B_a)^0 - (B_a)^0 \circ T_\omega = -(\widetilde{E}_2)^0$
   $(B_b)^0$ solves $\lambda (B_b)^0 - (B_b)^0 \circ T_\omega = -(\tilde{A}_2)^0$
9) *Find* $\overline{W}_2$, σ *solving*

$$0 = -\overline{S}\,\overline{W_2} - \overline{S(B_a)^0} - \overline{S(B_b)^0}\sigma - \overline{\overline{E}_1} - \overline{\overline{A}_1}\sigma,$$
$$(\lambda - 1)\overline{W_2} = -\overline{\overline{\overline{E}}_2} - \overline{\overline{A}_2}\sigma.$$

10) $(W_2)^0 = (B_a)^0 + \sigma (B_b)^0$

11) $W_2 = (W_2)^0 + \overline{W}_2$

12) $(W_1)^0$ *solves* $(W_1)^0 - (W_1)^0 \circ T_\omega = -(SW_2)^0 - (\widetilde{E}_1)^0 - (\widetilde{A}_1)^0 \sigma$

13) $K \leftarrow K + MW$

  $\mu \leftarrow \mu + \sigma.$

We formulate in Appendix A the equivalent of the Algorithm 33 for flows.

Remark 34. Algorithm 33 is constructed as follows. Step 1 follows from (6.1); step 2 defines α; steps 3, 4 follow respectively from (3.5), (3.13); step 5 defines β; steps 6, 7 follow respectively from (6.7), (3.9); step 8 follows from (6.10); step 9 gives the solution of (6.11); step 10 solves the second of (6.9); step 11 provides the solution W_2; step 12 solves the first of (6.9); step 13 determines the corrections Δ, σ with Δ as in (6.3).

Remark 35. There are two important points to underline in the above algorithm. One is that no large matrix (i.e. a matrix of dimension equal to the elements used in the discretization) is stored (and much less, inverted).

Another important point is that steps 2), 8), 10), 11), 12) are diagonal operations in the Fourier space, while all other steps are diagonal in the real space (a few of the steps, 10), 11) are diagonal in both spaces). Of course, once we obtain a representation of the function in discrete points or in Fourier space, we can obtain the other applying the Fast Fourier Transform (FFT). Therefore, if we decide to discretize the unknowns using N Fourier modes (as well as N discretization points), the storage required is $O(N)$ and the number of operations is $O(N \log N)$ – due to the FFT we use to switch from the representations in Fourier modes and in discrete space.

We will show that the algorithm has quadratic convergence.

Note that the algorithm does not require that the system is close to integrable. In particular, a continuation method can follow the tori all the way to breakdown.

Note that the algorithm is reasonably easy to implement and indeed has been implemented and run in [15,24].

Remark 36. Denoting by $J_0 = \begin{pmatrix} 0 & \mathrm{Id} \\ -\mathrm{Id} & 0 \end{pmatrix}$ the standard symplectic structure, we have that

$$M^T J \circ KM = J_0 + O(E). \tag{6.13}$$

Without changing the quadratic character of the algorithm, we can modify step 5) of the algorithm by using an approximate inverse M^{-1} obtained as $J_0^{-1} M^T J \circ K$.

Of course, from the theoretical point of view both methods for step 5 require $O(N)$ operations. If we use 5), the coefficient is the number of operations needed to invert $n \times n$ matrices, while in the case of (6.13) it is the number of operations needed to carry out the multiplication indicated. Note that J_0 is a very simple matrix, so that the multiplication is just a rearrangement of the coefficients, which in practice is easier to program. We do not need to link to linear algebra routines to compute $n \times n$ inverses and the only operations needed are very simple ones (such as those in the BLAS package [26]).

7. Estimates for the iterative step

7.1. Approximate reducibility

In this section, we present estimates on the approximate reducibility (see Lemma 37), which is a perturbation of the geometric arguments developed in Section 3.1. We present two versions of the estimates: one in the spaces of analytic functions and another in the space of Sobolev functions.

Lemma 37. *Let* $f_\mu : \mathcal{M} \to \mathcal{M}$ *be an analytic (respectively, C^r) conformal symplectic mapping. Let* $\omega \in \mathcal{D}(\nu, \tau)$.

Let $K : \mathbb{T}^n \to \mathcal{M}$ *be an embedding such that* $K \in \mathcal{A}_\rho$ *(respectively, $K \in H^m$, $m > n/2 + \tau + 1$ in the uniform case, $m > n/2 + 1$ in the non-uniform case, $r \geqslant m + 1$).*

Assume further that for some $\zeta > 0$:

1) $K(\mathbb{T}^n_\rho) \subset \text{domain}(f_\mu)$,

$$\text{dist}\big(K\big(\mathbb{T}^n_\rho\big), \partial \text{ domain}(f_\mu)\big) \geqslant \zeta > 0;$$

2) *the approximate invariance equation holds*:

$$f_\mu \circ K - K \circ T_\omega = E. \tag{7.1}$$

Then, we have

a) $\|Df_\mu \circ K\, DK - DK \circ T_\omega\|_{\mathcal{A}_{\rho-\delta}} \leqslant C\delta^{-1}\|E\|_{\mathcal{A}_\rho}$,

$\|Df_\mu \circ K\, DK - DK \circ T_\omega\|_{H^{m-1}} \leqslant C\|E\|_{H^m}$,

b) $\|K^*\Omega\|_{\mathcal{A}_{\rho-(\ell-1)\delta-\delta/2}} \leqslant C\nu^{-(\ell-1)}\delta^{-1-\tau(\ell-1)}\|E\|_{\mathcal{A}_\rho}$,

$\|K^*\Omega\|_{H^{m-(\ell-1)\tau-1}} \leqslant C\nu^{-(\ell-1)}\|E\|_{H^m}$,

where $\ell = 2$ *when C is independent of λ and $\ell = 1$ when C can depend on λ with $|\lambda| \neq 1$.*

Remark 38. The proof is based on repeating the calculations performed in Section 3.1, but keeping track of the estimates. In contrast with the calculation for exactly invariant systems done in Section 3.1, which could be performed with geometrically natural operations, in this section we need to perform geometrically unnatural operations such as comparing vectors at different points; they are possible because the phase space we are considering is a Euclidean manifold. Therefore in this section we will use a matrix formulation in preference to the more intrinsic geometric notation.

Proof. The proof of a) is just the chain rule applied to (3.2). The first step is to obtain estimates for the form E_Ω defined as:

$$T_\omega^* K^*\Omega - \lambda K^*\Omega = E_\Omega. \tag{7.2}$$

Note that, if h, g are diffeomorphisms (or more generally smooth maps with range $\mathcal{M}$), the matrix corresponding to $h^*\Omega - g^*\Omega$ is:

$$Dh^T J \circ h\, Dh - Dg^T J \circ g\, Dg = \big(Dh^T - Dg^T\big) J \circ h\, Dh + Dg^T (J \circ h - J \circ g) Dh$$
$$+ Dg^T J \circ g (Dh - Dg). \tag{7.3}$$

In our case, of course, we will apply the above formula (7.3) for $h = f_\mu \circ K$, $g = K \circ T_\omega$. Hence, we can estimate straightforwardly (7.2) using the above, Cauchy bounds (or the mean value theorem) and (7.1), as

$$\|E_\Omega\|_{\mathcal{A}_{\rho-\delta/2}} \leqslant C\delta^{-1}\|f_\mu \circ K - K \circ T_\omega\|_{\mathcal{A}_\rho} + \|\nabla J\|_{\mathcal{A}_\rho}\|f_\mu \circ K - K \circ T_\omega\|_{\mathcal{A}_\rho}$$
$$\leqslant C\delta^{-1}\|E\|_{\mathcal{A}_\rho},$$

$$\|E_\Omega\|_{H^{m-1}} \leqslant C\|E\|_{H^m}.$$

We now observe that, because $(f_\mu \circ K)^* \Omega = K^* f_\mu^* \Omega = \lambda K^* \Omega$, we have that $K^* \Omega$ satisfies (7.2), which in coordinates is a difference equation of the form we studied in Lemma 19, namely,

$$\left(DK^T \ J \circ K \ DK\right) \circ T_\omega - \lambda\left(DK^T \ J \circ K \ DK\right) = \tilde{E}_\Omega, \tag{7.4}$$

where $\tilde{E}_\Omega$ is the expression in coordinates of E_Ω. Furthermore, since for $\Omega = d\alpha$ we have $K^* \Omega = K^* d\alpha = d(K^* \alpha)$, we obtain that $DK^T \ J \circ K \ DK$ has zero average, and henceforth $\tilde{E}_\Omega$ has zero average. Applying the estimates obtained in Lemma 18 and Lemma 19 to (7.4), we obtain b). Notice that the limitation $m - (\tau + 1) > n/2$ and Sobolev's embedding theorem [85] ensure that we are dealing with continuous objects, which also enjoy the Banach algebra properties under multiplication and the composition estimates of Lemma 17. $\quad\square$

Lemma 39. *Let $f_\mu : \mathcal{M} \to \mathcal{M}$ be an analytic (respectively C^r) conformal symplectic mapping. Let $\omega \in \mathcal{D}(\nu, \tau)$. Let $K : \mathbb{T}^n \to \mathcal{M}$ be an embedding such that $K \in \mathcal{A}_\rho$ (respectively, $K \in H^m$, $m > n/2 + \tau + 1$ in the uniform case, $m > n/2 + 1$ in the non-uniform case, $r \geqslant m + 1$). Under the hypotheses of Lemma 37, assume that*

$$C\nu^{-1}\delta^{-(\tau+1)}\|E\|_{\mathcal{A}_\rho} \ll 1 \quad or \quad C\nu^{-1}\|E\|_{H^m} \ll 1. \tag{7.5}$$

Then, the matrix valued function

$$M(\theta) = \left[DK(\theta) \ \middle| \ J^{-1} \circ K(\theta) \ DK(\theta)N(\theta)\right]$$

satisfies

$$Df_\mu \circ K(\theta)M(\theta) = M(\theta + \omega)\begin{pmatrix} \mathrm{Id} & S(\theta) \\ 0 & \lambda\,\mathrm{Id} \end{pmatrix} + R(\theta),$$

where $S(\theta)$ is given by (3.9) and R satisfies:

$$\|R\|_{\mathcal{A}_{\rho-(\ell-1)\delta-\delta/2}} \leqslant C\nu^{-(\ell-1)}\delta^{-1-\tau(\ell-1)}\|E\|_{\mathcal{A}_\rho},$$
$$\|R\|_{H^{m-(\ell-1)\tau-1}} \leqslant C\nu^{-(\ell-1)}\|E\|_{H^m}, \tag{7.6}$$

where $\ell = 2$ when C is independent of λ or $\ell = 1$ when C can depend on λ with $|\lambda| \neq 1$.

Proof. Due to the assumption *(7.5)* and to b) of Lemma 37, then $\|K^* \Omega\|_{C^0}$ is sufficiently small, which implies that the torus $K(\mathbb{T}^n)$ is approximately Lagrangian and that

$$\mathrm{Range}\,DK(\theta) \cap \mathrm{Range}\left(J^{-1} \circ K(\theta)\,DK(\theta)\right) = \{0\} \quad \text{for all } \theta \in \mathbb{T}^n_{\rho-\delta}.$$

Due to the transversality of the spaces $\mathrm{Range}(J^{-1} \circ K(\theta)\,DK(\theta))$ and $\mathrm{Range}\,DK(\theta)$ for all θ, we obtain that $M(\theta)$ in (3.13) is a linear isomorphism. Due to a) of Lemma 37, the left column of M satisfies the bounds claimed in (7.6). The only thing that remains is to bound R_2 (compare with (3.25)), namely to study the right column of M provided by $J^{-1} \circ K\,DK\,N$.

We define the error on the Lagrangian character (see (3.23)) as

$$E_L \equiv DK^T \ J \circ K \ DK.$$

Due to b) of Lemma 37, the norm of E_L is bounded by the norm of E. Moreover, from (3.9), (3.26), (3.27), we obtain that

$$\tilde{A}(\theta) - A(\theta) = -E_L(\theta + \omega)\tilde{S}(\theta),$$

$$\tilde{S}(\theta) - S(\theta) = -N(\theta + \omega)^T \gamma(\theta + \omega)N(\theta + \omega)\big(\tilde{A}(\theta) - A(\theta)\big);$$

therefore the norms over $\mathcal{A}_\rho$ or H^m of $\|\tilde{A}(\theta) - A(\theta)\|$, $\|\tilde{S}(\theta) - S(\theta)\|$ are bounded by $\|E_L\|$. Due to (3.25) we obtain the estimates in (7.6). $\square$

7.2. Estimates for the increments in the step

In this section, we present estimates for the corrections Δ, σ obtained applying Algorithm 33 (we follow the notations introduced there). We present estimates using analytic and Sobolev norms. We note that, in the non-uniform case we can obtain better regularity estimates because we can use Lemma 18, rather than Lemma 19 to estimate the second equation of (6.6). Nevertheless, as pointed there, the constants depend on λ.

Lemma 40. *Let $f_\mu \in \mathcal{A}_\mathcal{C}$ be a family of conformally symplectic maps on a domain $\mathcal{C} \subset \mathbb{C}^n/\mathbb{Z}^n \times \mathbb{C}^n$, K an embedding, $\omega \in \mathcal{D}(\nu, \tau)$.*

A) *Analytic case: assume that the map f_μ is analytic and that for some $\zeta > 0$,*

$$K \in \mathcal{A}_{\rho+\delta}, \qquad K\big(\mathbb{T}_\rho^n\big) \subset \mathrm{domain}(f_\mu),$$

and

$$\mathrm{dist}\big(K\big(\mathbb{T}_\rho^n\big), \partial\big(\mathrm{domain}(f_\mu)\big)\big) \geqslant \zeta > 0.$$

Then, we have:

$$\|W\|_{\mathcal{A}_{\rho-\ell\delta}} \leqslant C\frac{1}{\nu^\ell}\delta^{-\ell\tau}\|E\|_{\mathcal{A}_\rho},$$

$$|\sigma| \leqslant C\|E\|_{\mathcal{A}_\rho},$$

where $\ell = 2$ in the uniform case, $\ell = 1$ in the non-uniform case and C is an explicit constant depending only on the dimension and on the non-degeneracy condition.

B) *Sobolev case: let f_μ be a C^r map, $r \geqslant m + 1$ and $m > \frac{n}{2} + \ell\tau$. Assume that*

$$K \in H^{m+1}$$

and that for some $\zeta > 0$:

$$\mathrm{dist}\big(K\big(\mathbb{T}_\rho^n\big), \partial\big(\mathrm{domain}(f_\mu)\big)\big) \geqslant \zeta > 0.$$

Then, we have:

$$\|W\|_{H^{m-\ell\tau}} \leqslant C\frac{1}{\nu^\ell}\|E\|_{H^m},$$

$$|\sigma| \leqslant C\|E\|_{H^m},$$

where $\ell = 2$ in the uniform case, $\ell = 1$ in the non-uniform case and C is an explicit constant depending only on the dimension and on the non-degeneracy condition.

Proof. We just follow the steps of Algorithm 33 estimating the output in terms of the input. We note that steps 8) and 12) involve solving cohomology equations; all the other steps are algebraic operations. The results of steps 8) and 12) are estimated using the results for difference equations in Section 4. The algebraic steps are estimated using the Banach algebra properties of the norms.

Note that we use the non-degeneracy conditions in **H3** of Theorem 20. As remarked before, note that these conditions hold provided that K stays in a neighborhood.

We remark that the estimates of the derivatives of f_μ with respect to the parameter μ immediately yield an estimate for $\tilde{A}$. Then, one has:

$$\left\| (D_\mu f_\mu) \circ K \right\|_{\mathcal{A}_\rho} \leqslant \sup_{\mathcal{C}} |D_\mu f_\mu|$$

and

$$\left\| (Df_\mu) \circ K \right\|_{\mathcal{A}_\rho} \leqslant \sup_{\mathcal{C}} |Df_\mu|.$$

In both cases, we define Q as an upper bound on the $\sup_{\mathcal{C}} |D_\mu f_\mu|$ and on the $\sup_{\mathcal{C}} |Df_\mu|$; then, for analytic norms we have that

$$\|\widetilde{A}\|_{\mathcal{A}_\rho} \leqslant Q \left\| M^{-1} \right\|_{\mathcal{A}_\rho}.$$

In the case of Sobolev norms we obtain the estimates from the composition

$$\left\| (D_\mu f_\mu) \circ K \right\|_{H^m} \leqslant A_m(\|K\|_\infty) \|D_\mu f_\mu\|_{C^m} \left(1 + \|K\|_{H^m}\right),$$

while in the analytic case we have

$$\left\| (D_\mu f_\mu) \circ K \right\|_{\mathcal{A}_\rho} \leqslant \|D_\mu f_\mu\|_{\mathcal{A}_\mathcal{C}}.$$

Similarly we have:

$$\left\| (Df_\mu) \circ K \right\|_{H^m} \leqslant A_m(\|K\|_\infty) \|Df_\mu\|_{C^m} \left(1 + \|K\|_{H^m}\right),$$

and

$$\left\| (Df_\mu) \circ K \right\|_{\mathcal{A}_\rho} \leqslant \|Df_\mu\|_{\mathcal{A}_\mathcal{C}}.$$

Another estimate that will be needed throughout the proof is an estimate for S. The main point is that S is an explicit algebraic expression involving only K and its derivatives, N, M^{-1}. Provided that these quantities remain in a small enough neighborhood of those corresponding to the initial guess, then S is uniformly bounded. For completeness we report below some explicit estimates in the analytic and Sobolev spaces that follow from the explicit formulas (see (3.9)):

$$\|S\|_{\mathcal{A}_\rho} \leqslant C \|Df_\mu\|_{\mathcal{A}_\mathcal{C}} \|K\|_{\mathcal{A}_{\rho+\delta}}^2 \|N\|_{\mathcal{A}_\rho}^2 |J^{-1}|_{\mathcal{B}_\zeta} + C \|N\|_{\mathcal{A}_\rho}^2 Q \left\| M^{-1} \right\|_{\mathcal{A}_\rho} \|\gamma\|_{\mathcal{A}_\rho},$$

$$\|S\|_{H^m} \leqslant C \|K\|_{H^{m+1}}^2 \|N\|_{H^m}^2 A_m(\|K\|_\infty) \|Df_\mu\|_{C^m} \left(1 + \|K\|_{H^m}\right) \left\| J^{-1} \right\|_{C^m}$$
$$+ C \|N\|_{H^m}^2 Q \left\| M^{-1} \right\|_{H^m} \|\gamma\|_{H^m},$$

where $\mathcal{B}_\zeta$ denotes a neighborhood of radius ζ around the image of $K(\theta)$.

The main point of the estimates above is that, provided that K, M, M^{-1}, N remain in a neighborhood, we have uniform bounds in S.

We now proceed to perform estimates of W, σ for $|\lambda| \neq 1$. Later, we will present estimates uniform for λ in an interval around 1.

The difference equations in steps 8), 9) of Algorithm 33 allow to conclude estimates for $\overline{W}_2$ and σ under the assumption **H3** of Theorem 20. Setting

$$\alpha_1 = \overline{S(B_b)^0} + \overline{\tilde{A}}_1, \qquad \beta_1 = -\overline{S(B_a)^0} - \overline{\tilde{E}}_1, \tag{7.7}$$

then we have

$$|\overline{W}_2| \leqslant \mathcal{T}\big(|\alpha_1| \|E\|_{\mathcal{A}_\rho} + |\beta_1| Q \left\|M^{-1}\right\|_{\mathcal{A}_\rho}\big),$$

$$|\sigma| \leqslant \mathcal{T}\big(|\lambda - 1| |\beta_1| + \|S\|_{\mathcal{A}_\rho} \|E\|_{\mathcal{A}_\rho}\big),$$

where

$$|\alpha_1| \leqslant C \frac{1}{||\lambda| - 1|} \left\|M^{-1}\right\|_{\mathcal{A}_\rho} Q \|S\|_{\mathcal{A}_\rho} + Q \left\|M^{-1}\right\|_{\mathcal{A}_\rho}$$

$$|\beta_1| \leqslant \left(\frac{1}{||\lambda| - 1|} \|S\|_{\mathcal{A}_\rho} + 1\right) \left\|M^{-1}\right\|_{\mathcal{A}_\rho} \|E\|_{\mathcal{A}_\rho}.$$

The estimates for W come from the fact that the components of W satisfy the difference equations in (6.6). Thus we obtain the following estimates, where the key point is that we can bound the norm of W by some other norm of E, times some quantities that are bounded provided that K (and therefore M, M^{-1}, S) remains in a sufficiently small neighborhood of the initial guess. Of course, the norms for W and the norms for E are in different spaces. In the analytic case, we lose some domain and the constants include a factor which is a power of the domain loss. From (6.6), Lemma 18 and Lemma 19, we finally obtain:

$$\|W_2\|_{\mathcal{A}_\rho} \leqslant \frac{C}{||\lambda| - 1|} \big(\|E\|_{\mathcal{A}_\rho} + \|D_\mu f_\mu\|_{\mathcal{A}_C} \left\|M^{-1}\right\|_{\mathcal{A}_\rho} |\sigma|\big)$$

$$\leqslant C \|E\|_{\mathcal{A}_\rho},$$

$$\|W_1\|_{\mathcal{A}_{\rho-\delta}} \leqslant C \delta^{-\tau} \frac{1}{\nu} \big[\|S\|_{\mathcal{A}_\rho} \|W_2\|_{\mathcal{A}_\rho} + \|E\|_{\mathcal{A}_\rho} + \|\tilde{A}_1\|_{\mathcal{A}_\rho} |\sigma|\big]$$

$$\leqslant C \delta^{-\tau} \|E\|_{\mathcal{A}_\rho}, \tag{7.8}$$

where C stands for a constant that depends on $\|DK\|_{\mathcal{A}_\rho}$ $\mathcal{T}$, $\|N\|_{\mathcal{A}_\rho}$ $\|M\|_{\mathcal{A}_\rho}$, $\|M^{-1}\|_{\mathcal{A}_\rho}$ and λ. Similar computations for the case of Sobolev spaces yield

$$|\sigma| \leqslant C \mathcal{T} \|E\|_{H^m}$$

and

$$\|W_2\|_{H^m} \leqslant \frac{C}{||\lambda| - 1|} \big(\|E\|_{H^m} + A_m(\|K\|_\infty) \nu^{-1} \|D_\mu f_\mu\|_{C^m} (1 + \|K\|_{H^m}) \|M^{-1}\|_{H^{m+1}} |\sigma|\big)$$

$$\leqslant C \|E\|_{H^m},$$

$$\|W_1\|_{H^{m-\tau}} \leqslant C \frac{1}{\nu} \big(\|S\|_{H^m} \|W_2\|_{H^m} + \|E\|_{H^m} + \|\tilde{A}_1\|_{H^m} |\sigma| \big)$$

$$\leqslant C \|E\|_{H^m}, \tag{7.9}$$

where C again depends on λ. The result for $|\lambda| \neq 1$ follows.

We conclude this section with estimates which are uniform in λ in an interval around 1. These estimates come from the fact that the solution of the first equation in (6.6) can be estimated by using Lemma 19, providing uniform estimates in λ for the difference equation (4.1).

These estimates are somewhat more subtle than in the previous case, because to find B, we have to solve a difference equation (6.10) that has small divisors. Note, in particular, that the equation for $(B_b)^0$ has a right hand side which is not small. We start by observing that we can still get easily

$$|\overline{W}_2|, |\sigma| \leqslant C \|E\|_{\mathcal{A}_\rho},$$

$$|\overline{W}_2|, |\sigma| \leqslant C \|E\|_{H^m}. \tag{7.10}$$

The reason for (7.10) is that the definition of α_1, β_1 in (7.7) only involves the function $S(B)^0$ in the real axis (or the C^0 norm). Hence, in the analytic case we can take as domain loss ρ, which remains uniformly bounded. In the Sobolev case, we see that the C^0 norm of $(B_b)^0$ is bounded by the H^m norm of A provided that $m > n/2 + \tau$. Furthermore, we see that we can bound $\|(B_a)^0\|_{C^0} \leqslant C \|E\|_{\mathcal{A}_\rho}$.

Like in the case for $\lambda \neq 1$, we can obtain estimates for W as follows (see (4.9)):

$$\|W_2\|_{\mathcal{A}_{\rho-\delta}} \leqslant C \delta^{-\tau} \frac{1}{\nu} \big(\|E\|_{\mathcal{A}_\rho} + \|D_\mu f_\mu\|_{\mathcal{A}_C} \|M^{-1}\|_{\mathcal{A}_\rho} |\sigma| \big)$$

$$\leqslant C \delta^{-\tau} \frac{1}{\nu} \|E\|_{\mathcal{A}_\rho},$$

$$\|W_1\|_{\mathcal{A}_{\rho-2\delta}} \leqslant C \delta^{-\tau} \frac{1}{\nu} \big(\|S\|_{\mathcal{A}_{\rho-\delta}} \|W_2\|_{\mathcal{A}_{\rho-\delta}} + \|E\|_{\mathcal{A}_\rho} + \|\tilde{A}_1\|_{\mathcal{A}_\rho} |\sigma| \big)$$

$$\leqslant C \delta^{-2\tau} \frac{1}{\nu^2} \|E\|_{\mathcal{A}_\rho}, \tag{7.11}$$

where C denotes a constant that is independent of λ.

Similarly, for the case of the Sobolev spaces we obtain:

$$\|W_2\|_{H^{m-\tau}} \leqslant C \frac{1}{\nu} \big(\|E\|_{H^m} + A_m \big(\|K\|_\infty \big) \|D_\mu f_\mu\|_{C^m} \big(1 + \|K\|_{H^m} \big) \|M^{-1}\|_{H^m} |\sigma| \big)$$

$$\leqslant C \|E\|_{H^m},$$

$$\|W_1\|_{H^{m-2\tau}} \leqslant C \frac{1}{\nu} \big(\|S\|_{H^{m-\tau}} \|W_2\|_{H^{m-\tau}} + \|E\|_{H^m} + \|\tilde{A}_1\|_{H^{m+1}} |\sigma| \big)$$

$$\leqslant C \|E\|_{H^m}. \quad \square \tag{7.12}$$

7.3. Estimates for the convergence of the iterative step

In the present section we state and prove Lemma 42 which provides estimates for Algorithm 33. The estimates in Lemma 42 allow to apply an abstract implicit function theorem (see Theorem 47 later), which establishes the convergence of the iterative scheme (or some modification involving smoothing) and the bounds for the solutions claimed in Theorem 20. Note that in the analytic case, we will also present a self-contained proof that does not require to use an abstract implicit function theorem. In the language of [90], Lemma 42 shows that the process we have presented in Algorithm 33 describes an *approximate right inverse* of the derivative of the functional given by the invariance equation.

We have already shown (Lemma 40) that an application of Algorithm 33 provides a step that makes the correction bounded (in a less smooth norm) by the original error. Now, let us restate the result in Lemma 40 using a convenient operator notation. Following [90,62,21], we describe Algorithm 33 by introducing a linear operator $\eta[K, \mu]$, depending on the approximate solution (K, μ). In this context, the operator η produces the correction (Δ, σ) from the error functional $\mathcal{E} = \mathcal{E}[K, \mu]$, where

$$\mathcal{E}[K, \mu] = f_\mu \circ K - K \circ T_\omega. \tag{7.13}$$

The procedure is completely specified in Algorithm 33. We will denote this process by

$$(\Delta, \sigma) = -\eta[K, \mu]E,$$

where $\Delta = -(\eta[K, \mu]E)_1$, $\sigma = -(\eta[K, \mu]E)_2$; note that Δ, σ depend linearly on E.

Remark 41. We note that Lemma 40 provides linear estimates in the error $\mathcal{E}[K, \mu]$ for the operator η, both in analytic and Sobolev spaces. One can also verify that Algorithm 33 is defined for all (K, μ) in an open ball in the corresponding spaces. Therefore, we have the following result.

A) Analytic case: under the hypotheses of Lemma 40A), the linear operator $\eta : \mathcal{A}_\rho \to \mathcal{A}_{\rho - \ell\delta}$ is defined for every (K, μ) in the unit ball in $\mathcal{A}_\rho$ centered at (K_0, μ_0), and satisfies

$$\big\| \eta[K, \mu]E \big\|_{\mathcal{A}_{\rho - \ell\delta}} \leqslant C \frac{1}{\nu^\ell} \delta^{-\ell\tau} \|E\|_{\mathcal{A}_\rho}.$$

B) Sobolev case: under the hypotheses of Lemma 40B), the linear operator $\eta : H^m \to H^{m - \ell\tau}$ is defined for every (K, μ) in the unit ball in H^m centered at (K_0, μ_0), and satisfies

$$\big\| \eta[K, \mu]E \big\|_{H^{m - \ell\tau}} \leqslant C \frac{1}{\nu^\ell} \|E\|_{H^m}.$$

The constants ℓ and C behave as in Lemma 40, i.e. $\ell = 2$ in the uniform case, while $\ell = 1$ if $\lambda \neq 1$ and C depends also on λ.

It is well known in Nash–Moser theory that a quadratically convergence step converges even if each step involves a loss of derivatives [70]. In the context of our problem, the quadratic estimates on the step amount to proving that, in a space of less regular functions, the norm of the quantity

$$D\mathcal{E}\Delta + D_\mu\mathcal{E}\sigma + E \tag{7.14}$$

is proportional to the square of the norm of the error E. The quadratic estimates will come from a simple computation on the norm of (7.14) and from applying the estimates in Lemma 37, Lemma 40.

Lemma 42. *Let $\eta[K, \mu]$ be the linear operator produced by an application of Algorithm* 33.
Assume that

$$C\nu^{-1}\delta^{-(\tau+1)}\|E\|_{\mathcal{A}_\rho} \ll 1 \quad \text{or} \quad C\nu^{-1}\|E\|_{H^m} \ll 1,$$

so that the conclusions of Lemma 37 *hold. Then, under the assumptions of Lemma* 40A) *and* B)*, we have the following quadratic estimates.*

A) *Analytic case*:

$$\left\| D\mathcal{E}[K,\mu]\Delta + D_\mu \mathcal{E}[K,\mu]\sigma + E \right\|_{\mathcal{A}_{\rho-\ell\delta}} \leqslant C \frac{1}{\nu^{2\ell-1}} \delta^{-1-\tau(2\ell-1)} \|E\|^2_{\mathcal{A}_\rho}.$$

B) *Sobolev case (for $r \geqslant m+1$, $m > n/2 + \ell\tau$)*:

$$\left\| D\mathcal{E}[K,\mu]\Delta + D_\mu \mathcal{E}[K,\mu]\sigma + E \right\|_{H^{m-\ell\tau}} \leqslant C \frac{1}{\nu^{2\ell-1}} \|E\|^2_{H^m}.$$

The constants ℓ and C behave as in Lemma 40.

Proof. Using the definition of $\mathcal{E}[K,\mu]$ in (7.13), M in *(3.13)*, R in *(3.20)* and adding and subtracting, one can verify the following identity.

$$
\begin{aligned}
D\mathcal{E}\Delta + D_\mu\mathcal{E}\sigma + E &= D\mathcal{E}\Delta + D_\mu\mathcal{E}\sigma - RM^{-1}\Delta + RM^{-1}\Delta + E \\
&= \left(Df_\mu \circ K - RM^{-1}\right)\Delta - \Delta \circ T_\omega + D_\mu f_\mu \circ K\sigma + E + RM^{-1}\Delta \\
&= \left(M \circ T_\omega \begin{pmatrix} \mathrm{Id} & S(\theta) \\ 0 & \lambda\,\mathrm{Id} \end{pmatrix} M^{-1}\right)\Delta - \Delta \circ T_\omega + D_\mu f_\mu \circ K\sigma + E + RM^{-1}\Delta \\
&= RM^{-1}\Delta, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (7.15)
\end{aligned}
$$

where we have used (3.20) and the fact that the operator η is obtained following Algorithm 33 to solve Eq. (6.5).

To obtain the result, we estimate the right hand side of the last equality in (7.15) by using Lemma 39 and the estimates in Remark 41. $\quad\square$

7.4. Estimates for the error of the improved solution

A natural consequence of the quadratic estimates in Lemma 42 above and the linear estimates in Lemma 40 is the following estimate of the error of the improved solution, which follows from Taylor's theorem applied to the functional given by the invariance equations.

We present the proof which is just a direct application of Lemma 17. This will allow us to give a self-contained proof in the analytic case. The main subtlety to keep in mind is that, since the estimates for the increment blow up if δ – the loss of domain – goes to zero, we cannot ensure that the range of $K + \Delta$ is in the domain of f_μ, so that the new error makes sense. Hence, in the direct proof, if one fixes the rate of domain losses, one has to ensure that the error decreases fast enough, so that the composition $f_\mu \circ (K + \Delta)$ makes sense (in the abstract theorem, this corresponds to the choices of smoothing steps to ensure that we never leave a neighborhood).

Lemma 43. *Let $\eta[K,\mu]$ be as in Lemma 42 and denote*

$$\Delta = -\left(\eta[K,\mu]E\right)_1 \quad and \quad \sigma = -\left(\eta[K,\mu]E\right)_2.$$

*Assume that $\mathcal{C}$, ζ and Λ are as in **H4** – for the analytic estimates – (or as in **H4'** for the Sobolev estimates) of Theorem 20. Furthermore, assume that*

$$C\nu^{-1}\delta^{-(\tau+1)}\|E\|_{\mathcal{A}_\rho} < \zeta \quad or \quad C\nu^{-1}\|E\|_{H^m} < \zeta, \qquad (7.16)$$

and that ζ is small enough, so that the estimates of Lemma 42 hold.

Then, we obtain the following composition estimates for $r \geqslant m + 2$, $m > n/2 + \ell\tau$:

A) *Analytic case:*

$$\left\|\mathcal{E}[K + \Delta, \mu + \sigma]\right\|_{\mathcal{A}_{\rho - \ell\delta}} \leqslant C \frac{1}{\nu^{2\ell}} \delta^{-2\ell\tau} \|E\|_{\mathcal{A}_\rho}^2.$$

B) *Sobolev case:*

$$\left\|\mathcal{E}[K + \Delta, \mu + \sigma]\right\|_{H^{m - \ell\tau}} \leqslant C \frac{1}{\nu^{2\ell}} \|E\|_{H^m}^2.$$

The constants ℓ and C behave as in Lemma 40.

Proof. The proof of the composition estimates follows from Taylor's theorem and from the estimates in Lemma 42. Define the remainder of the Taylor expansion as

$$\mathcal{R}\left[(K, \mu), (K', \mu')\right] = \mathcal{E}[K', \mu'] - \mathcal{E}[K, \mu] - D\mathcal{E}[K, \mu](K' - K) \\ - D_\mu \mathcal{E}[K, \mu](\mu' - \mu).$$

We notice that (7.16) guarantees that $|\sigma| < \zeta$ and

$$\|\Delta\|_{\mathcal{A}_{\rho - \ell\delta}} < \zeta \quad \text{or} \quad \|\Delta\|_{H^{m - \ell\tau}} < \zeta.$$

In particular, $K + \Delta \in \mathcal{C}$ and $\mu + \sigma \in \Lambda$. So one has:

$$\mathcal{E}[K + \Delta, \mu + \sigma] = E + D\mathcal{E}[K, \mu]\Delta + D_\mu \mathcal{E}[K, \mu]\sigma + \mathcal{R}\left[(K, \mu), (K + \Delta, \mu + \sigma)\right].$$

The first three terms of the above expansion are estimated in Lemma 42. The remainder is estimated as follows. In the analytic setting we have:

$$\|\mathcal{R}\|_{\mathcal{A}_{\rho - \ell\delta}} \leqslant C\left(\|\Delta\|_{\mathcal{A}_{\rho - \ell\delta}}^2 + |\sigma|^2\right) \leqslant \left[\left(C \frac{1}{\nu^\ell} \delta^{-\ell\tau} \|E\|_{\mathcal{A}_\rho}\right)^2 + \left(C\|E\|_{\mathcal{A}_\rho}\right)^2\right],$$

so that

$$\left\|\mathcal{E}[K + \Delta, \mu + \sigma]\right\|_{\mathcal{A}_{\rho - \ell\delta}} \leqslant C \frac{1}{\nu^{2\ell}} \delta^{-2\ell\tau} \|E\|_{\mathcal{A}_\rho}^2.$$

In the Sobolev setting we have:

$$\|\mathcal{R}\|_{H^{m - \ell\tau}} \leqslant C\left(\|\Delta\|_{H^{m - \ell\tau}}^2 + |\sigma|^2\right) \leqslant \left[\left(C \frac{1}{\nu^\ell} \|E\|_{H^m}\right)^2 + \left(C\|E\|_{H^m}\right)^2\right],$$

so that

$$\left\|\mathcal{E}[K + \Delta, \mu + \sigma]\right\|_{H^{m - \ell\tau}} \leqslant C \frac{1}{\nu^{2\ell}} \|E\|_{H^m}^2. \quad \square$$

7.5. An abstract implicit function theorem. Conclusion of the proof of Theorem 20

To conclude the proof of Theorem 20, we can just apply the abstract implicit function theorem that we state below for completeness (see Theorem 47). We follow the formulation of Theorem A.6 in the appendix of [21] with some modifications presented later. The theorem holds for an arbitrary scale of Banach spaces for which smoothing operators are available. The proof is done by combining the Newton's step, which looses derivatives, with smoothing that restores them. The procedure converges if the order of the Banach spaces is bounded. The main hypothesis is that the initial guess satisfies the equation very approximately as well as some other explicit non-degeneracy conditions.

For the sake of making the paper more self-contained, we also present an explicit proof of the analytic case of Theorem 20 in Section 7.6.

We consider a one-parameter family of Banach spaces $\mathcal{X}^r$ in the interval $0 \leqslant r' \leqslant r \leqslant \infty$,

$$\mathcal{X}^0 \supseteq \mathcal{X}^{r'} \supseteq \mathcal{X}^r \supseteq \mathcal{X}^\infty$$

and with norms satisfying

$$\|g\|_{\mathcal{X}^{r'}} \leqslant \|g\|_{\mathcal{X}^r}$$

for all $g \in \mathcal{X}^r$ and $0 \leqslant r' \leqslant r$.

In a scale of Banach spaces we define smoothing operators as follows.

Definition 44. Given a scale of Banach spaces $\{\mathcal{X}^r\}$, we say that $\{S_t\}_{t \in \mathbb{R}^+}$ is a family of smoothing operators when

i) $\lim_{t \to \infty} \|(S_t - \mathrm{Id})u\|_{\mathcal{X}^0} = 0$,
ii) $\|S_t u\|_{\mathcal{X}^m} \leqslant C t^{m-\ell} \|u\|_{\mathcal{X}^\ell}$, for all $0 \leqslant \ell \leqslant m$ and for all $u \in \mathcal{X}^\ell$,
iii) $\|(\mathrm{Id} - S_t)u\|_{\mathcal{X}^\ell} \leqslant C t^{-(m-\ell)} \|u\|_{\mathcal{X}^m}$, for all $0 \leqslant \ell \leqslant m$ and for all $u \in \mathcal{X}^m$.

In our case the scales of Banach Spaces will be either spaces of analytic functions or Sobolev spaces. Smoothing operators in Sobolev spaces are standard in functional analysis (see for example [85,21,73]). In the case of analytic functions the smoothing is obtained by rescaling the size of the strip on which the analytic functions are defined.

As pointed out in [90], one important consequence of the existence of smoothing operators is the validity of interpolation inequalities.

Proposition 45. *Let $f \in \mathcal{X}^s$; for $r \leqslant s$, $0 \leqslant \theta \leqslant 1$, we have*

$$\|f\|_{\mathcal{X}^{\theta r + (1-\theta)s}} \leqslant C \|f\|_{\mathcal{X}^r}^\theta \|f\|_{\mathcal{X}^s}^{1-\theta}. \tag{7.17}$$

The proof in [90] is very elementary. It suffices to observe that, for all $t > 0$ we have $f = S_t f + (\mathrm{Id} - S_t)f$, so that we obtain the bound:

$$\|f\|_{\mathcal{X}^{\theta r + (1-\theta)s}} \leqslant \|S_t f\|_{\mathcal{X}^{\theta r + (1-\theta)s}} + \left\|(\mathrm{Id} - S_t)f\right\|_{\mathcal{X}^{\theta r + (1-\theta)s}}$$
$$\leqslant C t^{\theta r + (1-\theta)s - s} \|f\|_{\mathcal{X}_s} + C t^{-(r - (\theta r + (1-\theta)s))} \|f\|_{\mathcal{X}_r}.$$

Computing the minimum of the function in t at the right hand side, provides the interpolation inequality (7.17).

Remark 46. Of course, for concrete examples of spaces, the interpolation inequalities were done much earlier and by different methods. For analytic functions, (7.17) is given by Hadamard's *three circles theorem*, while for finite differentiable functions or Sobolev functions the result was obtained by other methods [40,57,54,31].

To make the notation of the theorem more compatible with this paper, the elements of the Banach spaces will have two components, which we denote as (K, μ). Later, we will also separate the components of the approximate inverse. In an arbitrary scale of Banach spaces with smoothing we have the following result.

Theorem 47. *Let $\alpha > 0$ and $p > \alpha$ and let $\mathcal{X}^q$ for $p - \alpha \leqslant q \leqslant p + 13\alpha$ be a scale of Banach spaces with smoothing operators. Let $\mathcal{B}_q$ be the unit ball in $\mathcal{X}^q$, $\tilde{\mathcal{B}}_q = (K_0, \mu_0) + \mathcal{B}_q$ the unit ball translated by $(K_0, \mu_0) \in \mathcal{X}^q$ and let $\mathcal{B}(\mathcal{X}^q, \mathcal{X}^{q-\alpha})$ be the space of bounded linear operators from $\mathcal{X}^q$ to $\mathcal{X}^{q-\alpha}$. Consider the functional $\mathcal{E} = \mathcal{E}[K, \mu]$ with $\mathcal{E} : \tilde{\mathcal{B}}_q \to \mathcal{X}^q$ and let $\eta = \eta[K, \mu]$, with $\eta : \tilde{\mathcal{B}}_q \to \mathcal{B}(\mathcal{X}^q, \mathcal{X}^{q-\alpha})$. Consider a pair $(K, \mu) \in \tilde{\mathcal{B}}_q$ and denote by E the function obtained by evaluating the functional $\mathcal{E}$ at (K, μ), i.e. $E = \mathcal{E}[(K, \mu)]$, and by $\Delta = -(\eta[(K, \mu)]E)_1$, $\sigma = -(\eta[K, \mu]E)_2$. Furthermore, we assume:*

i) *$\mathcal{E}(\tilde{\mathcal{B}}_q \cap \mathcal{X}^q) \subset \mathcal{X}^q$ for $p - \alpha \leqslant q \leqslant p + 13\alpha$;*
ii) *$\mathcal{E}|_{\tilde{\mathcal{B}}_q \cap \mathcal{X}^q} : \tilde{\mathcal{B}}_q \cap \mathcal{X}^q \to \mathcal{X}^q$ has two continuous Fréchet derivatives for $p - \alpha \leqslant q \leqslant p + 13\alpha$;*
iii) *$\|(\Delta, \sigma)\|_{\mathcal{X}^{q-\alpha}} \leqslant C\|E\|_{\mathcal{X}^q}$, $(K, \mu) \in \tilde{\mathcal{B}}_q$, for $q = p - \alpha, p + 13\alpha$;*
iv) (*quadratic estimates*)

$$\left\| D\mathcal{E}[K, \mu]\Delta + D_\mu \mathcal{E}[K, \mu]\sigma + E \right\|_{\mathcal{X}^{p-\alpha}} \leqslant C\|E\|_{\mathcal{X}^p}^2,$$

where $(K, \mu) \in \tilde{\mathcal{B}}_p$;
v) *$\|E\|_{\mathcal{X}^{p+13\alpha}} \leqslant C(1 + \|(K, \mu)\|_{\mathcal{X}^{p+13\alpha}})$, $(K, \mu) \in \tilde{\mathcal{B}}_{p+13\alpha}$.*

Then, if we can find $(K_0, \mu_0) \in \mathcal{B}_{p+13\alpha}$ such that $\|E_0\|_{\mathcal{X}^{p-\alpha}}$ is sufficiently small (where $E_0 = \mathcal{E}[K_0, \mu_0]$), there exists $(K_e, \mu_e) \in \mathcal{X}^p$, such that $\mathcal{E}[K_e, \mu_e] = 0$. Moreover,

$$\left\|(K_e - K_0, \mu_e - \mu_0)\right\|_{\mathcal{X}^p} \leqslant C\|E_0\|_{\mathcal{X}^{p-\alpha}}. \tag{7.18}$$

Remark 48. Note that, even if we are referring everything to the unit ball of the spaces for convenience, we could use any ball. It suffices to redefine the norms multiplying them by a constant. Of course, in such a case, the smallness conditions in the end could depend on the size of the balls.

Remark 49. Note that (7.18) provides bounds on a smoother space from bounds in a rougher space. This is not paradoxical, because we are assuming that (K_0, μ_0) is in $\mathcal{B}_{p+13\alpha}$. Given the assumption v), this implies that $\|E_0\|_{\mathcal{X}^{p+13\alpha}}$ is bounded. Hence, by the interpolation inequalities (7.17) we obtain that the smallness assumption on $\|E_0\|_{\mathcal{X}^{p-\alpha}}$ implies the smallness assumption on $\|E_0\|_{\mathcal{X}_s}$, $p - \alpha \leqslant s \leqslant p + 13\alpha$.

Remark 50. The proof of Theorem 47 is very similar to that in Appendix A in [21], but since we performed some modifications we prefer to insert the whole proof for completeness. In particular, the parameters defining the scales of Banach spaces have been improved ($p + 13\alpha$ in this paper and $p + 17\alpha$ in [21]); moreover, in assumption v) the norms are computed in the same space, while in [21] they were computed in different spaces. These different composition estimates in v) are due to the fact that the operator $\mathcal{E}$ does not lose regularity under composition in our case, while this loss of regularity was allowed in [21]. Since the operator does not lose derivatives and since we are always working on bounded sets, the hypothesis v) becomes just that the operator is bounded.

Proof. The proof relies on an iterative procedure combining the ideas of [81,90]. Given $(K, \mu) \in \mathcal{X}^{p+13\alpha}$ and $E = \mathcal{E}[K, \mu]$ with the property that $\|E\|_{\mathcal{X}^{p-\alpha}}$ is sufficiently small compared with the other properties of the function, the iterative procedure constructs (K_e, μ_e) such that $\mathcal{E}[K_e, \mu_e] = 0$.

Let $\kappa > 1$, $\beta, \gamma, \delta > 0$, $0 < \psi < 1$ be real numbers which will be specified later. We construct a sequence $\{(K_n, \mu_n)\}_{n \geqslant 0}$ by defining

$$(K_{n+1}, \mu_{n+1}) = (K_n, \mu_n) - S_{t_n} \eta[K_n, \mu_n] E_n,$$

where $t_n = e^{\beta \kappa^n}$. By induction we prove that the following properties and inequalities are satisfied:

$(p1; n)$ $((K_n, \mu_n) - (K_0, \mu_0)) \in \mathcal{B}_p$;
$(p2; n)$ $\|E_n\|_{\mathcal{X}^{p-\alpha}} \leqslant \psi e^{-\gamma \alpha \beta \kappa^n}$;
$(p3; n)$ $1 + \|(K_n, \mu_n)\|_{\mathcal{X}^{p+13\alpha}} \leqslant \psi e^{\delta \alpha \beta \kappa^n}$.

Suppose that conditions $(p1; j)$, $(p2; j)$, and $(p3; j)$ are true for $j < n$. We start by proving $(p1; n)$. First, we notice that $(p2; n-1)$, assumption iii) and assumption ii) of Definition 44 imply that

$$\begin{aligned}
\big\|(K_n, \mu_n) - (K_{n-1}, \mu_{n-1})\big\|_{\mathcal{X}^p} &= \big\| S_{t_{n-1}} \eta[K_{n-1}, \mu_{n-1}] E_{n-1} \big\|_{\mathcal{X}^p} \\
&\leqslant C e^{2\alpha \beta \kappa^{n-1}} \big\| \eta[K_{n-1}, \mu_{n-1}] E_{n-1} \big\|_{\mathcal{X}^{p-2\alpha}} \\
&\leqslant C \psi e^{\alpha \beta \kappa^{n-1}(2-\gamma)}.
\end{aligned}$$

Then if $\gamma > 2$, $\{(K_n, \mu_n)\} \subset \mathcal{X}^p$ converges to some $(K, \mu) \in \mathcal{X}^p$. In order to prove $(p1; n)$, we remark that, using $\kappa^j \leqslant j(\kappa - 1)$, we obtain

$$\begin{aligned}
\big\|(K_n, \mu_n) - (K_0, \mu_0)\big\|_{\mathcal{X}^p} &\leqslant C \psi \sum_{j=1}^{\infty} e^{\alpha \beta \kappa^j (2-\gamma)} \\
&\leqslant C \psi \sum_{j=1}^{\infty} e^{\alpha \beta j (\kappa-1)(2-\gamma)} \\
&\leqslant C \psi \frac{e^{\alpha \beta (\kappa-1)(2-\gamma)}}{1 - e^{\alpha \beta (\kappa-1)(2-\gamma)}},
\end{aligned} \tag{7.19}$$

which shows that $\|(K_n, \mu_n) - (K_0, \mu_0)\|_{\mathcal{X}^p} \leqslant C\psi$ for $\gamma > 2$ and β large enough.

In order to prove $(p2; n)$, we add and subtract the terms E_{n-1},

$$D\mathcal{E}[K_{n-1}, \mu_{n-1}] \eta[K_{n-1}, \mu_{n-1}] E_{n-1},$$

and

$$D\mathcal{E}[K_{n-1}, \mu_{n-1}] \big[S_{t_{n-1}} \eta[K_{n-1}, \mu_{n-1}] E_{n-1} \big]_1 + D_\mu \mathcal{E}[K_{n-1}, \mu_{n-1}] \big[S_{t_{n-1}} \eta[K_{n-1}, \mu_{n-1}] E_{n-1} \big]_2$$

to E_n. Here $\Delta_j = -(\eta[K, \mu] E_j)_1$, $\sigma_j = -(\eta[K, \mu] E_j)_2$. Next we collect the terms in three groups to obtain

$$\|E_n\|_{\mathcal{X}^{p-\alpha}} \leqslant \big\|E_n - E_{n-1} + D\mathcal{E}[K_{n-1},\mu_{n-1}]\big[S_{t_{n-1}}\eta[K_{n-1},\mu_{n-1}]E_{n-1}\big]_1$$
$$+ D_\mu\mathcal{E}[K_{n-1},\mu_{n-1}]\big[S_{t_{n-1}}\eta[K_{n-1},\mu_{n-1}]E_{n-1}\big]_2\big\|_{\mathcal{X}^{p-\alpha}}$$
$$+ \big\|D\mathcal{E}[K_{n-1},\mu_{n-1}]\Delta_{n-1} + D_\mu\mathcal{E}[K_{n-1},\mu_{n-1}]\sigma_{n-1} + E_{n-1}\big\|_{\mathcal{X}^{p-\alpha}}$$
$$+ \big\|D\mathcal{E}[K_{n-1},\mu_{n-1}]\big[(\mathrm{Id} - S_{t_{n-1}})\eta[K_{n-1},\mu_{n-1}]E_{n-1}\big]_1$$
$$+ D_\mu\mathcal{E}[K_{n-1},\mu_{n-1}]\big[(\mathrm{Id} - S_{t_{n-1}})\eta[K_{n-1},\mu_{n-1}]E_{n-1}\big]_2\big\|_{\mathcal{X}^{p-\alpha}}. \tag{7.20}$$

We estimate the first term in (7.20) using assumption iii) and the formula for the quadratic remainder of Taylor's expansion:

$$\big\|E_n - E_{n-1} + D\mathcal{E}[K_{n-1},\mu_{n-1}]\big[S_{t_{n-1}}\eta[K_{n-1},\mu_{n-1}]E_{n-1}\big]_1$$
$$+ D_\mu\mathcal{E}[K_{n-1},\mu_{n-1}]\big[S_{t_{n-1}}\eta[K_{n-1},\mu_{n-1}]E_{n-1}\big]_2\big\|_{\mathcal{X}^{p-\alpha}}$$
$$\leqslant C\psi^2 e^{2\alpha\beta\kappa^{n-1}(2-\gamma)}.$$

Concerning the second term of (7.20), using assumption iv) we obtain

$$\big\|D\mathcal{E}[K_{n-1},\mu_{n-1}]\Delta_{n-1} + D_\mu\mathcal{E}[K_{n-1},\mu_{n-1}]\sigma_{n-1} + E_{n-1}\big\|_{\mathcal{X}^{p-\alpha}} \leqslant C\|E_{n-1}\|_{\mathcal{X}^p}^2.$$

We estimate $\|E_{n-1}\|_p^2$ by using the interpolation inequality, assumption v) and the induction hypotheses $(p2; n-1)$ and $(p3; n-1)$:

$$\|E_{n-1}\|_{\mathcal{X}^p}^2 \leqslant C\|E_{n-1}\|_{\mathcal{X}^{p-\alpha}}^{26/14}\|E_{n-1}\|_{\mathcal{X}^{p+13\alpha}}^{2/14}$$
$$\leqslant C\|E_{n-1}\|_{\mathcal{X}^{p-\alpha}}^{26/14}\big(1 + \big\|(K_{n-1},\mu_{n-1})\big\|_{\mathcal{X}^{p+13\alpha}}\big)^{2/14}$$
$$\leqslant C\psi^2 e^{\alpha\beta\kappa^{n-1}(-\frac{26\gamma}{14}+\frac{2\delta}{14})}.$$

Concerning the third term of (7.20), we use the properties of the smoothing operators and the fact that the Fréchet derivative, $(D\mathcal{E}, D_\mu\mathcal{E})$, is bounded:

$$\big\|D\mathcal{E}[K_{n-1},\mu_{n-1}]\big[(\mathrm{Id} - S_{t_{n-1}})\eta[K_{n-1},\mu_{n-1}]E_{n-1}\big]_1$$
$$+ D_\mu\mathcal{E}[K_{n-1},\mu_{n-1}]\big[(\mathrm{Id} - S_{t_{n-1}})\eta[K_{n-1},\mu_{n-1}]E_{n-1}\big]_2\big\|_{\mathcal{X}^{p-\alpha}}$$
$$\leqslant C\big\|(\mathrm{Id} - S_{t_{n-1}})\eta[K_{n-1},\mu_{n-1}]E_{n-1}\big\|_{\mathcal{X}^p}$$
$$\leqslant Ce^{-12\alpha\beta\kappa^{n-1}}\big\|\eta[K_{n-1},\mu_{n-1}]E_{n-1}\big\|_{p+12\alpha}$$
$$\leqslant Ce^{-12\alpha\beta\kappa^{n-1}}\|E_{n-1}\|_{\mathcal{X}^{p+13\alpha}}$$
$$\leqslant Ce^{-12\alpha\beta\kappa^{n-1}}\big(1 + \big\|(K_{n-1},\mu_{n-1})\big\|_{\mathcal{X}^{p+13\alpha}}\big)$$
$$\leqslant C\psi e^{\alpha\beta\kappa^{n-1}(\delta-12)}.$$

Finally, the desired inequality $(p2; n)$ is satisfied whenever

$$C\big(\psi^2 e^{2\alpha\beta\kappa^{n-1}(2-\gamma)} + \psi^2 e^{\alpha\beta\kappa^{n-1}(\frac{2\delta}{14}-\frac{26\gamma}{14})} + \psi e^{\alpha\beta\kappa^{n-1}(\delta-12)}\big) \leqslant \psi e^{-\gamma\alpha\beta\kappa^n}$$

or equivalently

$$C\left(\psi e^{-\alpha\beta\kappa^{n-1}(2(\gamma-2)-\gamma\kappa)} + \psi e^{-\alpha\beta\kappa^{n-1}(-\frac{2\delta}{14}+\frac{26\gamma}{14}-\gamma\kappa)} + e^{-\alpha\beta\kappa^{n-1}(12-\delta-\gamma\kappa)}\right) \leqslant 1. \qquad (7.21)$$

Condition (7.21) is satisfied for β sufficiently large, ψ sufficiently small and provided that

$$\gamma(2-\kappa) > 4,$$

$$\gamma(26 - 14\kappa) > 2\delta,$$

$$12 - \gamma\kappa > \delta. \qquad (7.22)$$

This concludes the proof of $(p2;n)$. To prove $(p3;n)$, we start by remarking that

$$1 + \left\|(K_n, \mu_n)\right\|_{\mathcal{X}^{p+13\alpha}} \leqslant 1 + \sum_{j=0}^{n-1} \left\|S_{t_j}\eta[K_j, \mu_j]E_j\right\|_{\mathcal{X}^{p+13\alpha}}$$

$$\leqslant 1 + C\sum_{j=0}^{n-1} e^{\alpha\beta\kappa^j} \left\|\eta[K_j, \mu_j]E_j\right\|_{\mathcal{X}^{p+12\alpha}}$$

$$\leqslant 1 + C\sum_{j=0}^{n-1} e^{\alpha\beta\kappa^j} \left\|E_j\right\|_{\mathcal{X}^{p+13\alpha}}$$

$$\leqslant 1 + C\sum_{j=0}^{n-1} e^{\alpha\beta\kappa^j} \left(1 + \left\|(K_j, \mu_j)\right\|_{\mathcal{X}^{p+13\alpha}}\right)$$

$$\leqslant 1 + C\psi\sum_{j=0}^{n-1} e^{\alpha\beta(1+\delta)\kappa^j},$$

so that one obtains

$$\left(1 + \left\|(K_n, \mu_n)\right\|_{\mathcal{X}^{p+13\alpha}}\right)e^{-\delta\alpha\beta\kappa^n} \leqslant e^{-\delta\alpha\beta\kappa^n} + C\psi\sum_{j=0}^{n-1} e^{\alpha\beta\kappa^j(1+\delta-\kappa\delta)}. \qquad (7.23)$$

To have $(p3;n)$ it suffices that the right hand side of (7.23) is less than 1. Therefore, if $\delta > \frac{1}{\kappa-1}$, the right hand side of (7.23) will be less than 1 for sufficiently large β. If we consider $\kappa = 4/3$, $\gamma = 61/10$ and $\delta = 7/2$, then (7.22) and $\delta > \frac{1}{\kappa-1}$ are satisfied simultaneously. To complete the induction, we fix β large enough so that (7.23) and (7.21) are satisfied.

Finally, with our choices of β and γ, we fix ψ to be $\psi = \|E\|_{\mathcal{X}^{p-\alpha}}e^{\alpha\beta\gamma}$, which, together with (7.19), leads to the estimate

$$\left\|(K_e - K_0, \mu_e - \mu_0)\right\|_{\mathcal{X}^p} \leqslant C\psi\frac{e^{\alpha\beta(\kappa-1)(2-\gamma)}}{1 - e^{\alpha\beta(\kappa-1)(2-\gamma)}} \leqslant C_{\gamma,\alpha,\beta,\kappa}\|E\|_{\mathcal{X}^{p-\alpha}},$$

which completes the proof. $\quad\square$

7.5.1. Choices of spaces and parameters

We apply Theorem 47 to prove the finite differentiable version of Theorem 20. To do so, we need to make appropriate choices of the parameter p, of the loss of differentiability α, and of the scales of Banach spaces $\mathcal{X}^q$ for $p - \alpha \leqslant q \leqslant p + 13\alpha$. In our case we choose the scale of Banach spaces to be the Sobolev space H^q for the function K times the space $\Lambda \subset \mathbb{R}^n$ for the parameter μ, i.e. $\mathcal{X}^q = H^q \times \Lambda$ with the product norm. We remark that smoothing operators for the spaces H^q are readily available, see [85,21].

Our choice of p and α depends on the estimates on the increment of the step and the estimates on the approximate reducibility in Section 7. According to Lemma 37 and 40 we choose the Sobolev exponent $p = m$ with $m > \frac{n}{2} + \ell\tau$ and the loss of differentiability $\alpha = \ell\tau$. This allows us to use the estimates in Remark 41 to justify iii) in Theorem 47. Moreover, item ii) of the abstract theorem requires that the functional

$$\mathcal{E}[K, \mu] = f_\mu \circ K - K \circ T_\omega,$$

defined in H^q for $m - \ell\tau \leqslant q \leqslant m + 13\ell\tau$, has two continuous Fréchet derivatives. The estimates in [60,6] on composition operators guarantee that if the function f_μ has two continuous derivatives more than the highest Sobolev exponent, then the composition operator $f_\mu \circ K \in H^{m+13\ell\tau}$ must have at least two continuous Fréchet derivatives. Therefore, f_μ should be in the space C^r for $r \geqslant m + 13\ell\tau + 2$.

7.6. End of the proof of the results of Theorem 20 for analytic regularity

In this section we present a proof of the convergence of the iterative step in analytic spaces. We start by remarking that the convergence statement follows from the abstract Theorem 47. Nevertheless, since the proof is fairly simple, we think it is worthwhile to present a self-contained proof, which is very similar to that in [90] (see for instance the qualitative estimates in [92]); a pedagogical exposition can be found in [29]. We point out that this proof is based on the technique of "analytic smoothing", which is the oldest technique in KAM theory, going back to Kolmogorov [55]. One constructs increasingly approximate solutions in smaller analyticity domains, but the loss of domain slows down so that we end up with a positive radius of analyticity.

Remark 51. For the experts in KAM theorem, we recall that the papers [68,69] introduced the technique of double smoothing perfected in [92] which allows to obtain finite differentiability results from the analytic ones.

The main observation of the method is that $C^{k+\alpha}$ spaces can be characterized by the speed of approximation of analytic functions defined on decreasing domains.

Then, starting in a C^r problem, one can smooth it and obtain a sequence of analytic problems, which yield a sequence of analytic solutions. The regularity of the problem translates into a fast convergence of the sequence of problems, which in turn yields the speed of convergence of the solutions using the *a-posteriori* format of the theorem. Therefore we obtain the smoothness of the solutions of the final problem.

As it is shown in [68,69,92], the double smoothing method yields better differentiability results, than the one-step smoothing which is the basis of results such as Theorem 47. Nevertheless, we remark that for our purposes a double smoothing is not so useful for the following reasons.

(1) Since we rely on geometric identities, we would need that also the smoothed problems preserve the geometric properties. It is not so straightforward to show that a smooth diffeomorphism can be approximated by analytic diffeomorphisms, which also preserve the same geometric structure and satisfy quantitative bounds (in the symplectic, volume preserving contact case this is done in [38]. It seems that one can adapt the methods of [38] to the present case, but the quantitative statements are not so straightforward).

(2) We are interested in presenting results in Sobolev norms rather than $C^{k+\alpha}$ norms, because Sobolev norms are much easier to compute numerically. Then, Sobolev spaces are not easy to characterize using approximations by analytic functions.

We are now able to present the estimates for the iterative step in the analytic space $\mathcal{A}_\rho$. Precisely, we compute estimates for the increment and estimates for the new reminder (provided the composition of the correction can be well defined). To simplify the typography, we use the index $h \in \mathbb{Z}$ to denote the steps of the iteration and we will use subindices to indicate the quantities after h steps. We start by making the choice of domain losses. A convenient choice is

$$\delta_h \equiv \frac{\rho_0}{2^{h+2}}$$

with ρ_0 equal to ρ of Theorem 20 and

$$\rho_{h+1} = \rho_h - \delta_h$$

for $h \geqslant 0$. Note that $\rho_h \geqslant \frac{\rho_0}{2}$ for $h \geqslant 1$. Define

$$\varepsilon_h \equiv \big\| \mathcal{E}(K_h, \mu_h) \big\|_{\mathcal{A}_{\rho_h}}$$

and let

$$d_h \equiv \|\Delta_h\|_{\mathcal{A}_{\rho_h}}, \qquad v_h \equiv \|D\Delta_h\|_{\mathcal{A}_{\rho_h}}, \qquad s_h \equiv |\sigma_h|.$$

Remark 52. By Lemma 40 we have the following inequalities:

$$d_h \leqslant \widehat{C}_h v^a \delta_h^{-\alpha} \varepsilon_h,$$
$$v_h \leqslant \widehat{C}_h v^a \delta_h^{-\alpha-1} \varepsilon_h,$$
$$s_h \leqslant \widehat{C}_h \varepsilon_h,$$

where $\widehat{C}_h$ are explicit constants depending in a polynomial manner on $\|M_h\|_{\mathcal{A}_{\rho_h}}$, $\|M_h^{-1}\|_{\mathcal{A}_{\rho_h}}$, $\|N_h\|_{\mathcal{A}_{\rho_h}}$ and $\mathcal{T}_h$ with $a = -2$, $\alpha = 2\tau$. If we allow C_h to depend also on λ, we can take $a = -1$, $\alpha = \tau$.

Remark 53. In the following we will denote by C a constant depending on ν, τ, n, ζ, ρ_0, $|J^{-1}|_{\mathcal{B}_\zeta}$, and that is a polynomial in $\|M_0\|_{\mathcal{A}_{\rho_0}}$, $\|M_0^{-1}\|_{\mathcal{A}_{\rho_0}}$, $\|N_0\|_{\mathcal{A}_{\rho_0}}$, $\mathcal{T}_0$. We will denote as C_h the maximum of the constants $\widehat{C}_h$ and of the constants $\widetilde{C}_h$ introduced in the Taylor estimate of Lemma 43 as

$$\varepsilon_{h+1} \leqslant \widetilde{C}_h v^{2a} \delta_h^{-2\alpha} \varepsilon_h^2.$$

We assume that C is large enough, for instance $C > 2C_0$. In the proof, we will give smallness conditions, so that $C_h \leqslant C$ for every $h \geqslant 0$. Since we look for a solution (K_e, μ_e) which is near to (K_0, μ_0), it is natural to expect (as shown later) that the quantities $\|M_h\|_{\mathcal{A}_{\rho_h}}$, $\|M_h^{-1}\|_{\mathcal{A}_{\rho_h}}$, $\|N_h\|_{\mathcal{A}_{\rho_h}}$ and $\mathcal{T}_h$ will be close to $\|M_0\|_{\mathcal{A}_{\rho_0}}$, $\|M_0^{-1}\|_{\mathcal{A}_{\rho_0}}$, $\|N_0\|_{\mathcal{A}_{\rho_0}}$ and $\mathcal{T}_0$, respectively.

We prove by induction that for all integers $h \geqslant 0$ the following properties hold:

$(p1; h)$

$$\|K_h - K_0\|_{\mathcal{A}_{\rho_h}} \leqslant \kappa_K \varepsilon_0 < \zeta,$$
$$|\mu_h - \mu_0| \leqslant \kappa_\mu \varepsilon_0 < \zeta$$

with

$$\kappa_K \equiv C \nu^a \rho_0^{-\alpha} 2^{2\alpha+1}, \qquad \kappa_\mu \equiv 2C; \tag{7.24}$$

$(p2; h)$

$$\varepsilon_h \leqslant (\kappa_0 \varepsilon_0)^{2^h - 1} \varepsilon_0$$

with

$$\kappa_0 \equiv C \nu^{2a} \rho_0^{-2\alpha} 2^{6\alpha};$$

$(p3; h)$ $C_h \leqslant C$.

Note that $(p1; 0)$, $(p2; 0)$ and $(p3; 0)$ are trivial. Assume that $(p1; h)$, $(p2; h)$, $(p3; h)$ are true for $h = 1, \ldots, H$. Then, by Lemma 43 we obtain the Taylor estimate

$$\varepsilon_h = \left\| \mathcal{E}(K_{h-1} + \Delta_{h-1}, \mu_{h-1} + \sigma_{h-1}) \right\|_{\mathcal{A}_{\rho_h}} \leqslant C \nu^{2a} \delta_{h-1}^{-2\alpha} \varepsilon_{h-1}^2, \tag{7.25}$$

where C, a, α are as in Remarks 52 and 53. Without loss of generality we can assume $C \geqslant 1$, $\rho_0 < 1$. We have:

$$\begin{aligned}
\varepsilon_h &\leqslant C \nu^{2a} \rho_0^{-2\alpha} 2^{2\alpha h} 2^{2\alpha} \varepsilon_{h-1}^2 \\
&\leqslant \left(C \nu^{2a} \rho_0^{-2\alpha} 2^{4\alpha} \right) 2^{2\alpha(h-1)} \left(C \nu^{2a} \rho_0^{-2\alpha} 2^{4\alpha} 2^{2\alpha(h-2)} \varepsilon_{h-2}^2 \right)^2 \\
&\leqslant \left(C \nu^{2a} \rho_0^{-2\alpha} 2^{4\alpha} \right)^{1+2+\cdots+2^{h-1}} 2^{2\alpha((h-1)+2(h-2)+\cdots+2^{h-2})} \varepsilon_0^{2^h}.
\end{aligned}$$

Using that $1 + 2 + \cdots + 2^{h-1} = 2^h - 1$, $(h-1) + 2(h-2) + \cdots + 2^{h-2} = 2^{h-1} \sum_{j=1}^{h-1} j 2^{-j} = 2^h - (h+1)$, one obtains:

$$\begin{aligned}
\varepsilon_h &\leqslant \left(C \nu^{2a} \rho_0^{-2\alpha} 2^{4\alpha} \right)^{2^h - 1} 2^{2\alpha(2^h - (h+1))} \varepsilon_0^{2^h} \\
&\leqslant \left(C \nu^{2a} \rho_0^{-2\alpha} 2^{4\alpha} 2^{2\alpha} \varepsilon_0 \right)^{2^h - 1} \varepsilon_0
\end{aligned}$$

for $h = 1, \ldots, H$.

Let us now prove $(p1; H+1)$, $(p2; H+1)$ and $(p3; H+1)$. We assume the induction assumption $(p1; h)$, $(p2; h)$, $(p3; h)$ and that

$$\|N_h\|_{\mathcal{A}_{\rho_h}} \leqslant 2\|N_0\|_{\mathcal{A}_{\rho_0}},$$

$$\|M_h\|_{\mathcal{A}_{\rho_h}} \leqslant 2\|M_0\|_{\mathcal{A}_{\rho_0}},$$

$$\|M_h^{-1}\|_{\mathcal{A}_{\rho_h}} \leqslant 2\|M_0^{-1}\|_{\mathcal{A}_{\rho_0}},$$

$$\mathcal{T}_h \leqslant 2\mathcal{T}_0$$

for $h = 0, \ldots, H$.

First, we prove $(p1; H+1)$ as follows. Using $j + 1 \leqslant 2^j$, we have:

$$\|K_{H+1} - K_0\|_{\mathcal{A}_{\rho_{H+1}}} \leqslant \sum_{j=0}^{H} d_j \leqslant \sum_{j=0}^{H}\left(\widehat{C}_j \nu^a \delta_j^{-\alpha}\varepsilon_j\right)$$

$$\leqslant C\nu^a \rho_0^{-\alpha} 2^{2\alpha}\varepsilon_0 \sum_{j=0}^{H}\left(2^{j\alpha}\kappa_0^{2^j-1}\varepsilon_0^{2^j-1}\right)$$

$$\leqslant C\nu^a \rho_0^{-\alpha} 2^{2\alpha}\varepsilon_0 \sum_{j=0}^{H}\left(2^{\alpha}\kappa_0\varepsilon_0\right)^j$$

$$\leqslant C\nu^a \rho_0^{-\alpha} 2^{2\alpha+1}\varepsilon_0,$$

assuming that ε_0 is sufficiently small, e.g., $2^{\alpha}\kappa_0\varepsilon_0 \leqslant \frac{1}{2}$. In conclusion,

$$\|K_{H+1} - K_0\|_{\mathcal{A}_{\rho_{H+1}}} \leqslant \kappa_K \varepsilon_0$$

with κ_K as in (7.24). Moreover, we have:

$$|\mu_{H+1} - \mu_0| \leqslant \sum_{j=0}^{H} s_j \leqslant \sum_{j=0}^{H}\widehat{C}_j\varepsilon_j$$

$$\leqslant C\sum_{j=0}^{H}(\kappa_0\varepsilon_0)^{2^j-1};$$

assuming that ε_0 is sufficiently small, e.g., $\kappa_0\varepsilon_0 \leqslant \frac{1}{2}$, we conclude:

$$|\mu_{H+1} - \mu_0| \leqslant 2C\varepsilon_0.$$

We take ε_0 small enough so that $\kappa_K\varepsilon_0 < \zeta$ and $2C\varepsilon_0 < \zeta$. Since $(p1; H+1)$ is true, we use Taylor estimate (7.25) with $H+1$ in place of h to obtain $(p2; H+1)$.

In order to prove $(p3; H+1)$ we need other hypotheses that are easier, but more tedious to verify:

$$\|N_h - N_0\|_{\mathcal{A}_{\rho_h}} \leqslant \widetilde{C}\|DK_h - DK_0\|_{\mathcal{A}_{\rho_h}},$$

$$\|M_h - M_0\|_{\mathcal{A}_{\rho_h}} \leqslant \widetilde{C}\|DK_h - DK_0\|_{\mathcal{A}_{\rho_h}},$$

$$\|M_h^{-1} - M_0^{-1}\|_{\mathcal{A}_{\rho_h}} \leqslant \widetilde{C}\|DK_h - DK_0\|_{\mathcal{A}_{\rho_h}},$$

$$|\mathcal{T}_h - \mathcal{T}_0| \leqslant \widetilde{C}\|DK_h - DK_0\|_{\mathcal{A}_{\rho_h}},$$

where $\widetilde{C}$ is a uniform constant. The above inequalities come from the fact that N, M, M^{-1} and $\mathcal{T}$ are algebraic expressions of DK, Df_μ and $D_\mu f_\mu$. The inverses of the previous matrix valued functions can be computed using Neumann series. To estimate DK_{H+1} note that

$$
\begin{aligned}
\| DK_{H+1} - DK_0 \|_{\mathcal{A}_{\rho_{H+1}}} &\leqslant \sum_{j=0}^{H} \nu_j \\
&\leqslant \sum_{j=0}^{H} \widehat{C}_j \nu^a \delta_j^{-\alpha-1} \varepsilon_j \\
&\leqslant C \nu^a \rho_0^{-\alpha-1} 2^{2(\alpha+1)} \sum_{j=0}^{H} 2^{(\alpha+1)j} (\kappa_0 \varepsilon_0)^{2^j - 1} \varepsilon_0 \\
&\leqslant C \nu^a \rho_0^{-\alpha-1} 2^{2\alpha+3} \varepsilon_0.
\end{aligned} \tag{7.26}
$$

Indeed, we have bounded the sum in the second inequality as well as the last expression by taking ε_0 sufficiently small, e.g., $2^{\alpha+1} \kappa_0 \varepsilon_0 \leqslant \frac{1}{2}$. The other estimates follow from (7.26). Again, if we take $\widetilde{C} C \nu^a \rho_0^{-\alpha-1} 2^{2\alpha+3} \varepsilon_0$ small enough, we are able to verify $(p3; H+1)$, since C_{H+1} is an algebraic expression of $\|M_H\|_{\mathcal{A}_{\rho_H}}$, $\|M_H^{-1}\|_{\mathcal{A}_{\rho_H}}$, $\|N_H\|_{\mathcal{A}_{\rho_H}}$ and $\mathcal{T}_H$.

8. Proof of Theorem 29 and Corollary 30

In this section we prove the uniqueness result of Theorem 29 and Corollary 30. We present in detail the Sobolev case, which, of course, implies the analytic case. We also present the analytic case directly.

Remember that we have:

$$
\begin{aligned}
f_{\mu_1} \circ K_1 &= K_1 \circ T_\omega, \\
f_{\mu_2} \circ K_2 &= K_2 \circ T_\omega.
\end{aligned} \tag{8.1}
$$

We denote by

$$
\tilde{R} \equiv f_{\mu_2} \circ K_2 - f_{\mu_1} \circ K_1 - Df_{\mu_1} \circ K_1 (K_2 - K_1) - D_\mu f_{\mu_1} \circ K_1 (\mu_2 - \mu_1). \tag{8.2}
$$

We anticipate that, because of Taylor's theorem, $\tilde{R}$ is quadratic in $K_2 - K_1$, $\mu_2 - \mu_1$. We also note that, using the reducibility as in Section 3.1, for some specific M obtained using K_1 in (3.13), we have:

$$
Df_{\mu_1} \circ K_1(\theta) M(\theta) = M(\theta + \omega) \mathcal{B}(\theta), \tag{8.3}
$$

where $\mathcal{B}(\theta) = \begin{pmatrix} \text{Id} & S(\theta) \\ 0 & \lambda \, \text{Id} \end{pmatrix}$. As before, we introduce the notation

$$
K_2(\theta) - K_1(\theta) = M(\theta) W(\theta), \tag{8.4}
$$

where M is the matrix constructed in (3.13) corresponding to the solution K_1, namely $M(\theta) = [DK_1(\theta) \mid J^{-1} \circ K_1(\theta) DK_1(\theta) N_1(\theta)]$, $N_1(\theta) = (DK_1(\theta)^T DK_1(\theta))^{-1}$. Subtracting the identities (8.1), and using (8.2), (8.3), (8.4), we obtain

$$0 = f_{\mu_2} \circ K_2 - K_2 \circ T_\omega - f_{\mu_1} \circ K_1 + K_1 \circ T_\omega$$

$$= Df_{\mu_1} \circ K_1 (K_2 - K_1) + D_\mu f_{\mu_1} \circ K_1 (\mu_2 - \mu_1) + \tilde{R} - (K_2 - K_1) \circ T_\omega$$

$$= M(\theta + \omega)\big[\mathcal{B}(\theta) W(\theta) + M^{-1}(\theta + \omega) D_\mu f_{\mu_1} \circ K_1 (\mu_2 - \mu_1) - W(\theta + \omega)$$

$$+ M^{-1}(\theta + \omega) \tilde{R} \big]. \tag{8.5}$$

We emphasize that (8.5) is an identity satisfied by the difference between the two solutions and by the related quantities we have introduced. Since $M(\theta + \omega)$ is invertible we have the following identity

$$\mathcal{B}(\theta) W(\theta) - W(\theta + \omega) + M^{-1}(\theta + \omega) D_\mu f_{\mu_1} \circ K_1 (\mu_2 - \mu_1)$$

$$+ M^{-1}(\theta + \omega) \tilde{R}(\theta) = 0. \tag{8.6}$$

The proof of Theorem 29 will be obtained by observing that (8.6) implies estimates for W and for $|\mu_1 - \mu_2|$ in terms of $\tilde{R}$ (note that (8.6) is very similar to the equations we had studied in the Newton's procedure) and that Taylor's theorem implies estimates for $\tilde{R}$ in terms of W and $\mu_1 - \mu_2$. Then, putting together the two estimates, we will obtain that $\max(\|W\|, |\mu_1 - \mu_2|) \leqslant C \max(\|W\|, |\mu_1 - \mu_2|)^2$. One small wrinkle is that the norms in both sides are different, but, as we will show, one can use interpolation inequalities to take care of that.

8.1. Estimates of the function W in (8.6)

The importance of (5.3) is that applying Lemma 19 to (8.6) we obtain the following estimates on the function W and on $|\mu_2 - \mu_1|$ (recall that in order to obtain estimates on W we use that W_1 has zero average.)

Note that we are not using the full strength of Lemma 19, which asserts the existence of solutions. In our case, we know that W exists and that it satisfies some equation. Lemma 19 gives us estimates for the solution.

To simplify the typography, we introduce $M_+ = \max(\|M\|_{H^m}, 1)$, $M_- = \max(\|M^{-1}\|_{H^m}, 1)$ and we obtain for $m > n/2 + \ell\tau$:

$$\|W\|_{H^{m-\ell\tau}} \leqslant C M_+ M_- \frac{1}{\nu^\ell} \|\tilde{R}\|_{H^m},$$

$$|\mu_2 - \mu_1| \leqslant C M_+ M_- \|\tilde{R}\|_{H^m}, \tag{8.7}$$

where $\ell = 1$ if we allow that the constants depend on λ with $|\lambda| \neq 1$ and $\ell = 2$ if we allow the constants to be independent on λ.

Using Taylor's theorem for the composition with Sobolev spaces, recalling (8.4) and the definition of M_+, we obtain an estimate for the function $\tilde{R}$:

$$\|\tilde{R}\|_{H^{m-\ell\tau}} \leqslant \frac{1}{2} \|D_\mu D_{(I,\varphi)} f_\mu\|_{C^m} M_+^2 \big(\|W\|_{H^m}^2 + |\mu_2 - \mu_1|^2 \big). \tag{8.8}$$

8.2. Interpolation and end of the proof of Theorem 29

Putting together (8.7) and (8.8), we obtain:

$$\max\big(\|W\|_{H^{m-\ell\tau}}, |\mu_1 - \mu_2| \big)$$

$$\leqslant C \|D_\mu D_{(I,\varphi)} f_\mu\|_{H^m} \nu^{-\ell} M_+^3 M_- \max\big(\|W\|_{H^m}, |\mu_1 - \mu_2| \big)^2. \tag{8.9}$$

Using the well-known interpolation inequalities (see, e.g., [85,21] and Proposition 45), we obtain

$$\|W\|_{H^m} \leqslant C \|W\|^{1/2}_{H^{m-\ell\tau}} \|W\|^{1/2}_{H^{m+\ell\tau}},$$

where C is a constant depending on m and n; we transform (8.9) into

$$\max\big(\|W\|_{H^{m-\ell\tau}}, |\mu_1 - \mu_2|\big)$$
$$\leqslant C \|D_\mu D_{(I,\varphi)} f_\mu\|_{C^m} \nu^{-\ell} M_+^3 M_- \max\big(\|W\|_{H^{m+\ell\tau}}, |\mu_1 - \mu_2|\big) \max\big(\|W\|_{H^{m-\ell\tau}}, |\mu_1 - \mu_2|\big)$$

(C depending on m, n), so that $W = 0$, namely $K_1 = K_2$ and $\mu_1 = \mu_2$ if (5.6) is satisfied.

8.3. The analytic case

The analytic case of Theorem 29 is implied by the previous one. Nevertheless, it is instructive to give just the minor changes needed for a direct proof. Note that the analytic case has a free parameter δ. Proceeding as before, and using the notation $M_\pm = \max(\|M^{\pm 1}\|_{\mathcal{A}_\rho}, 1)$, applying the analytic version of Lemma 19 we obtain

$$\|W\|_{\mathcal{A}_{\rho-\ell\delta}} \leqslant C M_+ M_- \frac{1}{\nu^\ell} \delta^{-\ell\tau} \|\tilde{R}\|_{\mathcal{A}_\rho}$$
$$|\mu_2 - \mu_1| \leqslant C M_+ M_- \|\tilde{R}\|_{\mathcal{A}_\rho},$$

with $\ell = 1$ if we allow that constants to depend on λ with $|\lambda| \neq 1$ and $\ell = 2$ if we allow the constants to be independent on λ. The only difference in the argument is that rather than using the interpolation of Sobolev norms, we use Hadamard's three circle theorem [2], which gives

$$\|W\|_{\mathcal{A}_\rho} \leqslant C \|W\|^{1/2}_{\mathcal{A}_{\rho-\ell\delta}} \|W\|^{1/2}_{\mathcal{A}_{\rho+\ell\delta}}.$$

Using (5.5) one obtains $K_1 = K_2$ and $\mu_1 = \mu_2$.

8.4. Proof of Corollary 30

Proof. To prove Corollary 30 we first note that thanks to Theorem 29, the solutions are unique in the neighborhood we are considering. We know that

$$f_{\mu_\phi, \phi} \circ K_\phi = K_\phi \circ T_\omega$$

and for $\phi' \neq \phi$ we consider the function e defined as

$$e = f_{\mu_\phi, \phi'} \circ K_\phi - K_\phi \circ T_\omega.$$

Then, we have that for ϕ in a metric space $(\mathcal{Y}, d)$ and for $\mathcal{X}$ being $\mathcal{A}_\rho$ or $H^{m-\ell\tau}$, due to (5.7):

$$\|e\|_{\mathcal{X}} = \|f_{\mu_\phi, \phi'} \circ K_\phi - K_\phi \circ T_\omega - f_{\mu_\phi, \phi} \circ K_\phi + K_\phi \circ T_\omega\|_{\mathcal{X}} \leqslant \tilde{A}_L \, d(\phi, \phi').$$

Therefore if $d(\phi, \phi')$ is small enough, there exist $(K_{\phi'}, \mu_{\phi'})$ such that

$$f_{\mu_{\phi'}, \phi'} \circ K_{\phi'} = K_{\phi'} \circ T_\omega$$

and

$$\|K_\phi - K_{\phi'}\|_{\mathcal{A}_{\rho - \ell\delta}} \leqslant C\tilde{A}_L \nu^{-\ell} \delta^{-\ell\tau} d(\phi, \phi'),$$

$$|\mu_\phi - \mu_{\phi'}| \leqslant C\tilde{A}_L d(\phi, \phi'),$$

with $\ell = 2$ in the uniform case and $\ell = 1$ if C depends on λ with $|\lambda| \neq 1$,

$$\|K_\phi - K_{\phi'}\|_{H^m} \leqslant C\tilde{A}_L d(\phi, \phi'),$$

$$|\mu_\phi - \mu_{\phi'}| \leqslant C\tilde{A}_L d(\phi, \phi'),$$

which is the desired Lipschitz property with Lipschitz constant $A_L = C\tilde{A}_L$. This concludes the Proof of Corollary 30. □

9. Further consequences of the *a-posteriori* formalism

As pointed out in [21], the fact that we have an a-posteriori theorem leads to the two following consequences.

(1) The solutions of the invariance equation, associated to analytic maps, which are Sobolev solutions of high enough order, are also analytic.
(2) A parameter value ϕ_0 of a map is on the boundary of the parameters with quasi-periodic attractors if and only if a Sobolev norm of high enough order of the conjugacy of quasi-periodic attractors for nearby parameters ϕ blows up as ϕ approaches ϕ_0.

The key to both results is the a-posteriori format of Theorem 20 that shows that given an approximate solution (either in an analytic or in a Sobolev sense) that satisfies appropriate non-degeneracy conditions, then there is a locally unique solution in the same spaces.

9.1. Bootstrap of regularity

The result on the bootstrap of regularity is obtained by observing that if the Sobolev regularity of a solution is high enough, then a truncation will be an approximate solution in the analytic sense. Indeed, the analytic a-posteriori theorem shows that there is an analytic solution close to the truncated solution. Finally, the local uniqueness result in Sobolev spaces (see Theorem 29), together with the fact that the Sobolev regularity is high enough, shows that the original and analytic solutions agree. The bootstrap of regularity proof for Sobolev spaces in the context of twist maps is given in [21] (for a full proof for twist maps in C^m spaces see [38,39]). In the case of conformally symplectic maps we obtain the following result.

Theorem 54. *Assume that the hypotheses of Theorem* 20 *hold and that the map* f_μ *is a real analytic map. Let* $K_s \in H^p$ *for* $p > \frac{n}{2} + 2\ell\tau$, *and* $\mu_s \in \Lambda$ *solve*

$$f_{\mu_s} \circ K_s - K_s \circ T_\omega = 0.$$

Then, there exists a $\rho > 0$ *such that* $K_s \in \mathcal{A}_\rho$. *The constant* ℓ *is* 1 *if we assume that* $|\lambda| \neq 1$ *and* $\ell = 2$ *if we assume that* λ *belongs to an interval containing* 1.

Proof. We start with $K_s \in H^p$, $p > \frac{n}{2} + 2\ell\tau$ and $\mu_s \in \Lambda$ a solution of the invariance equation, i.e. $\mathcal{E}(K_s, \mu_s) = 0$, obtained by applying Theorem 20.

We consider an approximation to the solution K obtained by truncating the Fourier series at the L-th Fourier mode, $K^{\leqslant L}(\theta) = \sum_{|k| \leqslant L} K_k e^{2\pi i k \cdot \theta}$. From the definition of $K^{\leqslant L}$, we obtain the following estimates in the C^0 and C^1 norms for every L sufficiently large:

$$\| K_s - K^{\leqslant L} \|_{C^0} \leqslant C L^{-p+\frac{n}{2}}, \qquad \| K_s - K^{\leqslant L} \|_{C^1} \leqslant C L^{-p+\frac{n}{2}+1}. \tag{9.1}$$

Later, we will use these estimates to guarantee that the function $K^{\leqslant L}$ satisfies the non-degeneracy conditions, the twist condition **H3**, and the assumption **H4'** on the domain of f_μ whenever L is large enough. This is due to the fact that the pair (K_s, μ_s) satisfies these conditions in the Sobolev space H^p.

To obtain an estimate in the $\mathcal{A}_{\rho_L}$ norm, for some $\rho_L > 0$, we will consider an $\alpha > 1$ so that if $\rho_L = \frac{1}{L} \frac{\log \alpha}{2\pi}$, we have that

$$\| K^{\leqslant L} \|_{\mathcal{A}_{\rho_L}} \leqslant \left(e^{2\pi \rho_L} \right)^L \sum_{|k| \leqslant L} |\widehat{K}_k| \leqslant \alpha \| K_s \|_{H^p}. \tag{9.2}$$

Indeed, using the estimates in (9.1) and (9.2), we obtain estimates for the invariance equation in the space of analytic functions $\mathcal{A}_{\frac{\rho_L}{2}}$. From the C^0 estimate in (9.1), we know that

$$\| \mathcal{E}(K^{\leqslant L}, \mu_s) \|_{C^0} = \| \mathcal{E}(K^{\leqslant L}, \mu_s) - \mathcal{E}(K_s, \mu_s) \|_{C^0} \leqslant \left(\| f_{\mu_s} \|_{C^1} + 1 \right) \| K^{\leqslant L} - K_s \|_{C^0} \leqslant C L^{-p+\frac{n}{2}},$$

and we can use composition estimates to verify that, if $\rho_{s,p} \equiv \| K_s \|_{H^p}$, then

$$\| \mathcal{E}(K^{\leqslant L}, \mu_s) \|_{\mathcal{A}_{\rho_{s,p}}} \leqslant \| f_\mu \|_{\mathcal{A}_{(\| K_s \|_{H^p})}} + \| K^{\leqslant L} \|_{\mathcal{A}_{\rho_L}}.$$

Using the log-convexity of the supremum of the analytic functions (Hadamard three circle theorem), we can interpolate the previous inequalities and obtain estimates in $\mathcal{A}_{\frac{\rho_L}{2}}$:

$$\| \mathcal{E}(K^{\leqslant L}, \mu_s) \|_{\mathcal{A}_{\frac{\rho_L}{2}}} \leqslant C L^{-\frac{p}{2}+\frac{n}{4}}. \tag{9.3}$$

The estimate in (9.3) is the last ingredient that we need to apply the analytic case of Theorem 20 using $K^{\leqslant L}$ and μ_s as an approximate solution and noticing that if L is large enough then we have that $C \nu^{-2\ell} (\frac{\rho_L}{2})^{-2\ell\tau} \| \mathcal{E}(K^{\leqslant L}, \mu_s) \|_{\mathcal{A}_{\frac{\rho_L}{2}}} < 1$.

The conclusion of Theorem 20 is that there exists a $K_e \in \mathcal{A}_{\frac{\rho_L}{4}}$ and $\mu_e \in \Lambda$ such that $\mathcal{E}(K_e, \mu_e) = 0$. Moreover, for $0 < \delta < \frac{\rho_L}{4}$ we obtain the following estimates for K_e and μ_e:

$$\| K_e - K^{\leqslant L} \|_{\mathcal{A}_{\frac{\rho_L}{2}-\ell\delta}} \leqslant C \nu^{-\ell} \delta^{-\ell\tau} L^{-\frac{p}{2}+\frac{n}{4}},$$

$$| \mu_e - \mu_s | \leqslant C L^{-\frac{p}{2}+\frac{n}{4}}. \tag{9.4}$$

We obtain estimates on the Sobolev norm of $K^{\leqslant L} - K_s$ by using the smoothing operators in Sobolev spaces defined in 44 and choosing the smoothing parameters $\beta, t \in \mathbb{R}^+$ to be $\beta = \frac{p}{2} - \frac{n}{4} - \ell\tau$ and $t = (\nu^{-\ell} \| K \|_{H^p}^{-1})^{\frac{1}{\beta}}$:

$$\| K^{\leqslant L} - K_s \|_{H^p} \leqslant \| S_t (K^{\leqslant L} - K_s) \|_{H^p} + \| (Id - S_t)(K^{\leqslant L} - K_s) \|_{H^p}$$

$$\leqslant C t^\beta \| K^{\leqslant L} - K_s \|_{H^{p-\beta}} \leqslant C \nu^{-\ell} L^{-\frac{p}{2}+\frac{n}{4}+\ell\tau}. \tag{9.5}$$

Notice that the last inequality is a consequence of the definition of the Sobolev norm,

$$\left\| K^{\leqslant L} - K_s \right\|_{H^{p-\beta}}^2 = \sum_{|k|>L} \left(1 + |k|^2\right)^{p-\beta} |\widehat{K}_k|^2 \leqslant CL^{-2\beta} \sum_{|k|>L} \left(1 + |k|^2\right)^{p} |\widehat{K}_k|^2 \leqslant CL^{-2\beta} \| K_s \|_{H^p}^2.$$

We remark that when $L > \frac{1}{\delta}$, the first estimate in (9.4) implies that

$$\left\| K_e - K^{\leqslant L} \right\|_{H^p} \leqslant C\nu^{-\ell} L^{-\frac{p}{2}+\frac{n}{4}+\ell\tau}. \tag{9.6}$$

Therefore, we combine (9.6) and (9.5) to obtain the inequality

$$\| K_e - K_s \|_{H^p} \leqslant C\nu^{-\ell} L^{-\frac{p}{2}+\frac{n}{4}+\ell\tau}.$$

This last estimate, together with the second estimate in (9.4), implies that for L large enough the inequality (see Theorem 29)

$$C \| D_\mu D_{(I,\varphi)} f_\mu \|_{H^{p-\ell\tau}} \nu^{-\ell} M_+^3 M_- \max\left(\left\| M^{-1}(K_e - K_s) \right\|_{H^p}, |\mu_e - \mu_s| \right) < 1 \tag{9.7}$$

should hold. By the uniqueness in H^p, we have that $K_e = K_s$. In particular, $K_s \in \mathcal{A}_{\frac{\rho_L}{4}}$.   □

9.2. Criterion for the breakdown of analyticity

The second consequence of the a-posteriori formalism states a criterion for the breakdown of analyticity of quasi-periodic attractors. The justification of the criterion is based on Theorem 54. Note that using Theorem 20 and Theorem 29 we can obtain the bootstrap of regularity uniformly in a region.

Theorem 55. *Let $f_{\mu,\phi}$ be a family of analytic mappings, satisfying the hypotheses of Theorem 20. We assume that for some $\phi_0 > 0$ the mapping $f_{\mu_{\phi_0},\phi_0}$ has a Sobolev regular quasi-periodic solution $(K_{\phi_0}, \mu_{\phi_0})$ that satisfies the non-degeneracy assumption (5.1). Then, if $|\phi - \phi_0|$ is small enough depending on the size of the Sobolev norm of K_{ϕ_0} and on $|\mu_{\phi_0}|$, then $f_{\mu_\phi,\phi}$ has an analytic solution, which is locally unique.*

We use Theorem 55 to construct and to justify the correctness of numerically accessible algorithms to estimate the breakdown of analyticity. A prototype algorithm to estimate the breakdown is given below.

Algorithm 56.
Choose a path in the parameter space starting with the integrable case
Initialize
The parameters and the solution at the integrable case
Repeat
    Increase the parameters along the path
    Run the iterative step
    If (*Iterations of the Newton's step do not converge*)
        Decrease the increment in parameters
    Else (*Iteration success*)
        Record the values of the parameters
        and the Sobolev norm of the solution
    If *Non-degeneracy conditions fail*
        Return "inconclusive"
Until *Sobolev norm exceeds a threshold*

The ending point of the algorithm is the estimated critical value, namely when the Newton's steps converge, but the Sobolev norm exceeds a threshold.

Algorithm 56 has been implemented to estimate the breakdown of analyticity of quasi-periodic attractors of conformally symplectic maps in [15]; for similar calculations in the symplectic case see [19].

A recurrent remark concerning schemes like Algorithm 56 is that one can make the results more convincing by observing that the norms blow up according to a power law. Renormalization group predicts that there is a power law blow up for each Sobolev norm and that there is a simple relation between the scaling exponents corresponding to Sobolev norms (see [28,15]). These empirically found scaling relations are consistent with a renormalization group description of the breakdown. A version of this renormalization group was proposed in [76].

10. Perturbative expansions

In several applications, the conformally symplectic mappings (or flows) include parameters, some of which may be small. For example, in celestial mechanics applications, one can often consider the masses of the planets as small (compared to the masses of the Sun) or even the dissipation to be small.

In these circumstances, it is natural (and standard in theoretical physics) to obtain perturbative expansions of the objects of interest in terms of the small parameter. From the mathematical point of view, there are several natural results one can consider. One result shows that indeed one can compute the expansion in power series and that the formal expansion can be carried to all orders (when the family of maps is analytic). A second type of results is to prove estimates on the reminder and the general term, thus showing that the formal power series indeed defines a convergent series. We will present results of all these types. Theorem 59 shows that there are perturbation results to all orders, Theorem 60 shows that, when the unperturbed case is dissipative, the perturbative series indeed defines an analytic function. Note that Theorem 59 applies also to the case that the unperturbed system is conservative. Indeed, when the unperturbed system is conservative and the perturbations are not, we present arguments that suggest that for typical systems, the perturbative expansions exist to all orders, but do not converge. The proof of Theorems 59 and 60 is based on the use of the *automatic reducibility* coming from the geometry. We note that the proofs of the above results also lead to an efficient algorithm to compute the perturbative expansions. Indeed, we obtain a quadratic algorithm, whose step doubles the number of terms computed so far in the perturbative expansion. The cost in operations is only polynomial in the degree.

10.1. Basic set up

Since we will deal with analytic functions, it is convenient to consider maps defined in complex extensions, namely

$$f : \Lambda \times V \times A_n \times \mathbb{T}_\rho^n \to A_n \times \mathbb{T}_\rho^n,$$

where $\Lambda \subset \mathbb{C}^n$, $A_n \subset \mathbb{C}^n$, $0 \in V \subset \mathbb{C}$ are open sets. We think of the variables in $A_n, \mathbb{T}_\rho^n$ to be the dynamical action-angle variables. The variables in Λ are the variables μ that we have considered so far. Note that we have already developed a theory for families of maps in this form. The variable in V, which we will denote by ε and which we will refer to as *external parameter*, has the meaning of the perturbation parameter. Since often the perturbation parameter is small, it is convenient to assume that $\varepsilon = 0$ is in the range of parameters Λ.

The goal of this section is to study how the results obtained for ε fixed depend on the value of ε. That is, we will investigate the function $\mu(\varepsilon)$ and we study whether one can obtain perturbative expansions in ε of the result. We write the families as

$$f_{\mu,\varepsilon}(I, \varphi) = (\bar{I}, \bar{\varphi}),$$

so that we consider μ, ε as parameters. We assume that each of the mappings $f_{\mu,\varepsilon}$ is conformally symplectic, namely

$$(f_{\mu,\varepsilon})^* \Omega = \lambda_\varepsilon \Omega,$$

for some function λ_ε. We assume that for $\varepsilon = 0$ the family $f_{\mu,0}$ admits a solution μ_0, K_0 of the invariance equation

$$f_{\mu_0,0} \circ K_0 = K_0 \circ T_\omega, \tag{10.1}$$

satisfying the non-degeneracy condition of Theorem 20. An important particular case for the assumption (10.1) is that $f_{\mu_0,0}$ is an integrable mapping, but this is not needed. Finally, we assume that ω is Diophantine (see (3.1)). Note that Theorem 20 implies that, changing ε in a sufficiently small neighborhood, there exists a solution $(\mu_\varepsilon, K_\varepsilon)$ of the equation

$$f_{\mu_\varepsilon,\varepsilon} \circ K_\varepsilon = K_\varepsilon \circ T_\omega. \tag{10.2}$$

Theorem 20 gives that the solution is locally unique and that it depends in a Lipschitz way on ε (in appropriate topologies for the mapping K).

In this section we aim to study the functions $\mu_\varepsilon, K_\varepsilon$. First, we will show that there are solutions of the parameterized equations (10.2) in the sense of power series in ε and later we will show that these power series are actually analytic (notice that there are some differences among the hypotheses of the two results). As it is well known, the existence of solutions to all orders are required to satisfy some Diophantine condition, such as (3.1) (or the Brjuno–Rüssmann conditions in [79]). The second difference is that we can establish analyticity on parameters only when all the maps are contractive. Of course, when all maps are symplectic, the results are well known [70].

Remark 57. We have considered only the case in which the perturbing parameter is one-dimensional. However, the results easily generalize to deduce the analyticity with respect to more parameters. Indeed, if $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$, $\varepsilon_i \in \mathbb{C}$, we can consider $\varepsilon_1, \ldots, \varepsilon_{i-1}, \varepsilon_{i+1}, \ldots, \varepsilon_n$ fixed and we can apply the same technique to show the convergence of expansions in ε_i. Then, it suffices to use Hartogs theorem [56] to conclude that μ, K are jointly analytic in all variables.

10.2. Formal series solutions

We are seeking solutions of (10.2) of the form

$$\mu_\varepsilon = \sum_{i=0}^\infty \mu^{(i)} \varepsilon^i, \qquad K_\varepsilon = \sum_{i=0}^\infty K^{(i)} \varepsilon^i,$$

for some unknown coefficients $\mu^{(i)}$, $K^{(i)}$. More precisely, we seek $\mu^{(i)}$, $K^{(i)}$ such that, using a truncation of μ_ε and K_ε to the order N, one has

$$f_{\sum_{i=0}^N \mu^{(i)}\varepsilon^i, \varepsilon} \circ \sum_{i=0}^N K^{(i)} \varepsilon^i = \sum_{i=0}^N K^{(i)} \circ T_\omega \varepsilon^i + o(\varepsilon^{N+1}). \tag{10.3}$$

The $o(\varepsilon^{N+1})$ in (10.3) can be understood in two meanings, which are indeed equivalent:

A) Formal sense: if we substitute the series, expand and group all the terms in ε, then all the coefficients of ε^i, $i \leqslant N$, vanish.

B) Asymptotic sense: if we substitute

$$\mu_\varepsilon^{\leqslant N} = \sum_{i=0}^{N} \mu^{(i)} \varepsilon^i, \qquad K_\varepsilon^{\leqslant N} = \sum_{i=0}^{N} K^{(i)} \varepsilon^i,$$

into (10.2), we obtain for some $\rho' > 0$:

$$\left\| f_{\mu_\varepsilon^{\leqslant N}, \varepsilon} \circ K_\varepsilon^{\leqslant N} - K_\varepsilon^{\leqslant N} \circ T_\omega \right\|_{\mathcal{A}_{\rho'}} \leqslant C \varepsilon^{N+1}. \tag{10.4}$$

Remark 58. For the series that we are considering, both notions are equivalent as it can be easily seen. Note that the function at the left hand side of (10.4) is analytic in ε. Since there are only non-negative powers of ε involved, the coefficients of the expansion in ε up to the order N depend only on the coefficients of the expansion of μ_ε, K_ε up to the order N. Hence, if we substitute a polynomial truncation of the power series, we obtain that the coefficients of order $i \leqslant N$ of the power series are given by the formulas obtained in the formal power series expansions. If they vanish, then we obtain (10.4).

Theorem 59. *Assume that the families $f_{\mu,\varepsilon}$ are as above, that ω satisfies $f_{\mu,0} \circ K_0 = K_0 \circ T_\omega$ and that Assumption* **H3** *in Theorem 20 is satisfied. Assume furthermore that ω satisfies (3.1). Then, there is a formal power series solution of (10.2). Moreover, there is one and only one such a series that, besides (10.2), satisfies also the normalization (5.3).*

The reason why we impose the normalization (5.3) is that the solutions of the invariance equation are not unique, since we can always change the origin of the phases in the parameterization, and still obtain a solution. Of course, if these changes of phases are performed arbitrarily for each value of the parameter, there is no possibility of obtaining any smooth dependence. Of course, other normalizations could work just as well.

Proof. Since $f_{\mu^{(0)},0} \circ K^{(0)} = K^{(0)} \circ T_\omega$ and $(f_{\mu^{(0)},0})^* \Omega = \lambda_0 \Omega$ for some value λ_0, proceeding as in Section 3.1, there exist M, S as in (3.14) such that

$$Df_{\mu^{(0)},0} \circ K^{(0)}(\theta) M(\theta) = M(\theta + \omega) \mathcal{B}(\theta), \qquad \mathcal{B}(\theta) \equiv \begin{pmatrix} \text{Id} & S(\theta) \\ 0 & \lambda_0 \, \text{Id} \end{pmatrix}. \tag{10.5}$$

Note that since $K^{(0)}, \mu^{(0)}$ are exact solutions of the invariance equation, the formulas in (10.5) are exact as indicated in Section 3.1. As standard in perturbation theory, we expand (10.2) in ε and match the coefficients of equal powers of ε on both sides of the equation. This will give us recursive equations that, as we will see, determine $K^{(i)}, \mu^{(i)}$, provided that we have computed all the previous ones. Equating terms of order ε in (10.2), we find

$$Df_{\mu^{(0)},0} \circ K^{(0)} K^{(1)} - K^{(1)} \circ T_\omega + (D_\mu f)_{\mu^{(0)},0} \circ K^{(0)} \mu^{(1)} + (D_\varepsilon f)_{\mu^{(0)},0} \circ K^{(0)} = 0. \tag{10.6}$$

More generally, matching the terms of order ε^i we obtain

$$Df_{\mu^{(0)},0} \circ K^{(0)} K^{(i)} - K^{(i)} \circ T_\omega + (D_\mu f)_{\mu^{(0)},0} \circ K^{(0)} \mu^{(i)}$$
$$= P_i\big(K^{(1)}, \ldots, K^{(i-1)}, \mu^{(1)}, \ldots \mu^{(i-1)}\big), \tag{10.7}$$

where P_i is a polynomial expression in $K^{(1)}, \ldots, K^{(i-1)}, \mu^{(1)}, \ldots, \mu^{(i-1)}$, whose coefficients are given by the polynomial derivatives of $f_{\mu,\varepsilon}$ evaluated at $\mu = \mu^{(0)}$, $\varepsilon = 0$, and composed with $K^{(0)}$. Of course

(10.6) is an equation for $(\mu^{(1)}, K^{(1)})$, while (10.7) is an equation for $(\mu^{(i)}, K^{(i)})$. Eqs. (10.6) and (10.7) have to be supplemented with the equations obtained expanding (5.3), namely

$$\int_{\mathbb{T}^n} \left[M^{-1}(\theta) \left(K^{(i)}(\theta) - K^{(0)}(\theta) \right) \right]_1 d\theta = 0, \tag{10.8}$$

so that we adjust the average of the first component of W, being $W = \sum_{i=0}^{\infty} W^{(i)} \varepsilon^i$, as it was done in (5.3) which implies a shifting of the origin of the angle coordinate. Eq. (10.7) can be conveniently studied using (10.5). As in the Newton's step, we substitute (10.5) into (10.7) and, introducing the notation of the Newton's step, namely $K^{(i)}(\theta) = M(\theta) W^{(i)}(\theta)$, we obtain

$$\mathcal{B}(\theta) W^{(i)}(\theta) - W^{(i)}(\theta + \omega) + M^{-1}(\theta + \omega)(D_\mu f)_{\mu^{(0)},0} \circ K^{(0)}(\theta) \mu^{(i)}$$
$$= M^{-1}(\theta + \omega) P_i \left(K^{(1)}, \dots, K^{(i-1)}, \mu^{(1)}, \dots, \mu^{(i-1)} \right). \tag{10.9}$$

We recall that expressing equation (10.9) in components (denoted with subscripts) we obtain the following equations:

$$\lambda_0 \left(W^{(i)} \right)_2(\theta) - \left(W^{(i)} \right)_2(\theta + \omega) + \left[M^{-1}(\theta + \omega)(D_\mu f)_{\mu^{(0)},0} \circ K^{(0)} \right]_2 \mu^{(i)}$$
$$= \left[M^{-1}(\theta + \omega) P_i(\theta) \right]_2,$$
$$\left(W^{(i)} \right)_1(\theta) - \left(W^{(i)} \right)_1(\theta + \omega) + S(\theta) \left(W^{(i)} \right)_2(\theta) + \left[M^{-1}(\theta + \omega)(D_\mu f)_{\mu^{(0)},0} \circ K^{(0)}(\theta) \right]_1 \mu^{(i)}$$
$$= \left[M^{-1}(\theta + \omega) P_i(\theta) \right]_1.$$

These equations can be solved as in Section 6 using the non-degeneracy assumption and the Diophantine property of ω. We note that for $\lambda_0 \neq 1$, it suffices to solve one equation involving small divisors and another equation without small divisors. When $\lambda_0 = 1$, we have to solve two equations involving small divisors. Recall that the existence of solutions of the small divisor equations requires that some average vanishes. Hence, depending on whether $|\lambda_0| \neq 1$ or $\lambda_0 = 1$ one needs to deal with one or two obstructions. Of course, we also obtain different estimates on the solutions depending on whether we solve one or two equations, but this is not our concern here. In the case when $|\lambda_0| \neq 1$, we start by solving for $(W^{(i)})_2$. Because the equation does not have small divisors, once we fix $\mu^{(i)}$, the $(W^{(i)})_2$ is determined uniquely. The $\mu^{(i)}$ is determined in such a way that the equation for $(W^{(i)})_1$ – which involves small divisors – has a solution. Under a non-degeneracy condition of the form (5.1), we can determine $(W^{(i)})_1$, $(W^{(i)})_2$, $\mu^{(i)}$. The solution is unique up to a constant, which is chosen so that (10.8) is satisfied. In the case when $\lambda_0 = 1$, we obtain that the equation for $(W^{(i)})_2$ can be solved by choosing properly $\mu^{(i)}$. Nevertheless, the solution is determined only up to an additive constant, which can be chosen in such a way that the equation for $(W^{(i)})_1$ is solvable. Again, this determines $(W^{(i)})_1$ up to an additive constant, which is chosen so that (10.8) is satisfied. $\quad\square$

10.3. Convergence of the perturbative series expansions

In this section we prove that the formal series expansions of the previous sections are convergent. The proof is a rather simple consequence of the "a-posteriori" format of Theorem 20.

Theorem 60. *In the conditions of Theorem 59, assume furthermore that $|\lambda| \neq 1$. Then, the normalized power series obtained in Theorem 59 converges to an analytic function.*

Proof. First of all, we observe that Theorem 20 produces a family of solutions $(\mu_\varepsilon, K_\varepsilon)$ for all $|\varepsilon|$ sufficiently small. The proof of the theorem shows that the solutions are Lipschitz functions with respect

to ε. We now turn to prove that they are actually differentiable (in the complex sense) and therefore analytic in ε. We note that the first order expansion gives us a formal derivative that satisfies the inequality for $\eta > 0$:

$$\left\| f_{\mu_{\varepsilon+\eta}, \varepsilon+\eta} \circ \left(K_\varepsilon^{(0)} + \eta K_\varepsilon^{(1)} \right) - \left(K_\varepsilon^{(0)} + \eta K_\varepsilon^{(1)} \right) \circ T_\omega \right\|_{\mathcal{A}_{\rho-\delta}} \leqslant C \delta^{-\alpha} \nu^{-2} \eta^2$$

for suitable constants $C > 0$, $\alpha > 0$. We finally remark that $(\mu_\varepsilon^{(0)} + \eta \mu_\varepsilon^{(1)}, K_\varepsilon^{(0)} + \eta K_\varepsilon^{(1)})$ is an approximate solution of the invariance equation for $\varepsilon + \eta$ up to an error bounded by $O(\varepsilon^2)$; on the other hand, the non-degeneracy conditions remain the same.

For η small enough, we can apply Theorem 20 to obtain the existence of a solution $(\tilde{\mu}_{\varepsilon+\eta}, \tilde{K}_{\varepsilon+\eta})$ of (3.2) for the parameters $\varepsilon + \eta$, satisfying

$$\left| \tilde{\mu}_{\varepsilon+\eta} - \mu_\varepsilon^{(0)} - \eta \mu_\varepsilon^{(1)} \right| \leqslant C \eta^2,$$

$$\left\| \tilde{K}_{\varepsilon+\eta} - K_\varepsilon^{(0)} - \eta K_\varepsilon^{(1)} \right\|_{\mathcal{A}_{\rho-2\delta}} \leqslant C \eta^2.$$

Using Theorem 29 we obtain that $(\tilde{\mu}, \tilde{K})$ is precisely the solution produced by a direct application of Theorem 20. Therefore, the resulting solution satisfies the inequalities:

$$\left| \mu_{\varepsilon+\eta} - \mu_\varepsilon^{(0)} - \eta \mu_\varepsilon^{(1)} \right| \leqslant C \eta^2,$$

$$\left\| K_{\varepsilon+\eta} - K_\varepsilon^{(0)} - \eta K_\varepsilon^{(1)} \right\|_{\mathcal{A}_{\rho-2\delta}} \leqslant C \eta^2.$$

As a consequence, the quantity $\mu_\varepsilon^{(1)}$ is the derivative of μ_ε and $K_\varepsilon^{(1)}$ is the derivative of K_ε, whenever we consider K_ε in the domain $\mathbb{T}_{\rho-2\delta}^n$. Once we know that $(\mu_\varepsilon, K_\varepsilon)$ is differentiable for ε in a complex neighborhood of zero, we know that its Taylor expansion in ε converges in the domain $\mathbb{C}^n \times \mathbb{T}_{\rho-2\delta}^n$. Since we know that the functions $K_\varepsilon, \mu_\varepsilon$ are analytic, it is easy to argue that the (convergent) Taylor expansions of these functions are the functions that we have computed in the perturbative expansions. Using Remark 58, we note that the coefficients of the Taylor expansion solve (10.7) and (10.8). Since the solutions of these equations are unique, we conclude that, indeed, the formal solutions are the Taylor series of the analytic functions $K_\varepsilon, \mu_\varepsilon$ and that, therefore, these Taylor series converge in a sufficiently small domain. $\quad\square$

Remark 61. Perturbative expansions provide a tool to determine an explicit relation between the frequency ω and the drift μ. For example, in the case of the dissipative standard map (2.2), one can proceed as follows. Let Δ_λ be defined as the finite difference operator acting on a function $u = u(\theta)$, $\theta \in \mathbb{T}$, as $\Delta_\lambda u(\theta) = u(\theta + \frac{\omega}{2}) - \lambda u(\theta - \frac{\omega}{2})$; then, parametrizing the invariant torus as $\varphi = \theta + u(\theta)$ and using the invariance equation, one gets

$$\Delta_1 \Delta_\lambda u(\theta) - \lambda V'\big(\theta + u(\theta)\big) + \gamma = 0, \tag{10.10}$$

where $\gamma = \omega(1 - \lambda) - \mu \lambda$. Multiplying (10.10) by $1 + u_\theta$ and taking the average, one gets

$$\gamma + \overline{u_\theta \Delta_1 \Delta_\lambda u} = 0,$$

namely

$$\mu \lambda = \omega(1 - \lambda) + \langle u_\theta \Delta_1 \Delta_\lambda u \rangle,$$

which relates the frequency ω and the drift μ (see [14]).

10.3.1. Some conjectures on the convergence of the perturbative expansions

Note that the hypotheses of Theorem 59 on the existence of power series are less restrictive than those of Theorem 60 on the convergence of the power series expansions thus obtained. The existence of formal power series is valid both for $\lambda_0 = 1$ and for $|\lambda_0| \neq 1$, whereas the convergence is established only when $|\lambda_0| \neq 1$. There is a good reason for these restrictions. In fact, we note that the equations $\lambda_0(W^{(i)})_2 - (W^{(i)})_2 \circ T_\omega = \eta$ could fail to have solutions for all η whenever

$$\lambda_0 = \exp(2\pi i k \cdot \omega), \quad k \in \mathbb{Z}^n \setminus \{0\}. \tag{10.11}$$

In such a case, we can only obtain solutions for η such that $\hat{\eta}_{ak} = 0$, $a \in \mathbb{Z}$. Since these numbers are dense on the unit circle, we cannot develop a theory for open sets of λ_0 which contain 1. The following heuristic argument supports our first conjecture.

Conjecture 62. *If $\lambda_0 = 1$ and $\lambda_{\mu_\varepsilon} \not\equiv 1$, then the perturbative expansion diverges for a generic family $f_{\mu,\varepsilon}$.*

The heuristic argument for Conjecture 62 follows from the open mapping theorem: if $\lambda_{\mu_\varepsilon}$ is nontrivial (which is a generic condition), in any complex neighborhood of zero in the ε-plane, we see that there would be some values ε^* for which $\lambda_{\mu_{\varepsilon^*}}$ is of the form (10.11). For these values, we cannot guarantee that there exists a derivative with respect to ε by the previous argument. Indeed, it seems plausible that these derivatives do not exist because the right hand side of the equation we have to solve does not satisfy the constraints. This would be clear if f_{μ_ε}, μ_ε, K_ε were arbitrary, but of course, they are closely related. The same argument applies when $|\lambda_0| = 1$, but nevertheless the series is defined to all orders. These numbers are of full measure in the unit circle.

Conjecture 63. *When $\lambda_0 = 1$ and ω is Diophantine, the perturbative series are Gevrey. For a family which satisfies some non-degeneracy assumption, the functions K_ε, μ_ε are analytic in a domain which is a ball in the complex plane minus a sequence of balls with centers in a curve and whose radii are bounded by a function that decreases exponentially fast with the distance to the origin.*

In the paper [52], there is an argument that shows that, using Theorem 20, the second part of Conjecture 63 is a consequence of the first part. The first part of Conjecture 63 seems plausible because of analogies with other cases.

Conjecture 64. *Let f_0 be a non-degenerate system with an analytic invariant torus, K_0. Let ω be a frequency that does not satisfy a non-resonance condition of the form*

$$\lim_{N \to \infty} \max_{q \in \mathbb{Z}^n, \, |q| \leqslant N, \, p \in \mathbb{Z}} |\omega \cdot q - p|^{-1} \exp(-\alpha N) = 0$$

for all $\alpha \in \mathbb{R}^+$. Then, there exists an analytic family f_ε – indeed polynomial in ε – for which it is impossible to obtain an asymptotic expansion. Indeed, such functions are generic.

Similar phenomena in other contexts have been discussed in [82,66,80]. Note that a consequence of Conjecture 64 is that there is a residual set of frequencies for which the perturbative expansions can be defined to all orders, but nevertheless diverge for all $\varepsilon \neq 0$.

10.4. A fast algorithm for the computation of perturbative expansions. An alternative proof of Theorem 60

The main observation is that Algorithm 33 can be lifted to analytic families. The quadratic convergence of the algorithm implies that the number of terms in the expansion doubles at every step.

Also, using the usual KAM estimates, we can establish that the series converges. We just describe the iterative step starting from the equation

$$f_{\mu_\varepsilon^{(N)}} \circ K_\varepsilon^{(N)} - K_\varepsilon^{(N)} \circ T_\omega = E_\varepsilon^{(N)};$$

proceeding as in Section 3.1, in analogy to (10.5), we obtain that

$$Df_{\mu_\varepsilon^{(N)}} \circ K_\varepsilon^{(N)}(\theta) M_\varepsilon^{(N)}(\theta) = M_\varepsilon^{(N)}(\theta + \omega) \mathcal{B}_\varepsilon^{(N)}(\theta) + O(\varepsilon^N),$$

$$\mathcal{B}_\varepsilon^{(N)}(\theta) = \begin{pmatrix} \mathrm{Id} & S_\varepsilon^{(N)}(\theta) \\ 0 & \lambda_\varepsilon^N \, \mathrm{Id} \end{pmatrix}.$$

Note that in this case, of course, all the objects involved in the solution of the cohomology equations are functions depending on ε. The fact that the estimates are uniform in ε leads to the conclusion that exactly the same estimates used in the iterative step hold for the convergence in the sup-norm of families (see [32]). Since the uniform limit of analytic functions is also analytic, we obtain that the solution depends analytically on parameters. We also note that this is a practical algorithm to compute the perturbative series expansions. Using the methods of automatic differentiation for functions of two variables ([12], see [42] for a modern review), one can implement the operators involved in the evaluation of the Newton's method.

Compared with a direct numerical solution of (10.7) this method is not only faster, but it also has the advantage that, being a Newton's method, it is numerically stable; on the contrary, in the order by order method, the errors in the low-order affect the high-order terms, but they are never corrected, so that the errors accumulate on higher order terms. On the other hand, the Newton's method keeps on correcting even the low-order terms.

Acknowledgments

Appendix A. A dissipative KAM scheme for flows

In this section, we discuss a practical scheme to compute parameterizations of invariant tori for flows. We present only the formal calculations; the convergence of the scheme can be proved in a very similar way as it was done for mappings. We will not discuss the method in the greatest generality as possible, but we will assume that the symplectic form is given by $\Omega = d\varphi \wedge dI$. In this context, we consider a conformally symplectic family of vector fields of the form (2.4), (2.5), that we write as

$$\dot{I} = -\frac{\partial H(I, \varphi)}{\partial \varphi} + \lambda I + \mu,$$

$$\dot{\varphi} = \frac{\partial H(I, \varphi)}{\partial I}, \tag{A.1}$$

defined over a manifold $\mathcal{M} \subset \mathbb{T}^n \times \mathbb{R}^n$ and being μ a parameter in $\mathbb{R}^n$.

Let $\omega \in \mathbb{R}^n$ be a Diophantine frequency for flows, namely $\omega \in \mathcal{D}_n(\nu, \tau)$ with

$$\mathcal{D}_n(\nu, \tau) = \left\{ \omega \in \mathbb{R}^n \colon |\omega \cdot k| \geqslant \nu |k|^{-\tau} \ \forall k \in \mathbb{Z}^n \backslash \{0\} \right\}.$$

We parameterize an invariant attractor with frequency ω as $(I, \varphi) = K(\theta)$ for $\theta \in \mathbb{T}^n$, such that $\dot{\theta} = \omega$ for a suitable embedding $K : \mathbb{T}^n \to \mathcal{M}$. Let ∂_ω be the partial derivative operator $\partial_\omega \equiv \omega \cdot \partial_\theta$; denoting by F_μ the vector field associated to (A.1), we obtain that K must satisfy the invariance equation

$$\partial_\omega K(\theta) - F_\mu \circ K(\theta) = 0. \tag{A.2}$$

Taking the gradient of (A.2) we obtain

$$D\big(\partial_\omega K(\theta)\big) - \nabla\big(F_\mu \circ K(\theta)\big) DK(\theta) = 0.$$

The Lagrangian character of the torus implies that K satisfies

$$DK(\theta)^T J \circ K(\theta) DK(\theta) = 0.$$

The equivalent of the automatic reducibility of Section 3.1 is obtained through the following proposition.

Proposition 65. *Let $N(\theta) \equiv (DK(\theta)^T DK(\theta))^{-1}$; let M be the matrix obtained by juxtaposing the matrices $DK(\theta)$, $J DK(\theta) N(\theta)$, i.e.*

$$M(\theta) \equiv \big[DK(\theta) \mid J DK(\theta) N(\theta) \big]. \tag{A.3}$$

Then, setting $A \equiv \nabla F_\mu \circ K$, we have:

$$\partial_\omega M(\theta) - A(\theta) M(\theta) = M(\theta) \begin{pmatrix} 0 & S(\theta) \\ 0 & \lambda \, \mathrm{Id} \end{pmatrix}. \tag{A.4}$$

with

$$S(\theta) \equiv N(\theta) DK(\theta)^T J \big(A(\theta) + A(\theta)^T \big) DK(\theta) N(\theta).$$

Proof. Let us compute separately the first and second half of the columns of $W(\theta) \equiv \partial_\omega M(\theta) - A(\theta) M(\theta)$. The first half columns are zero, since (A.2) implies (for simplicity of notation we omit the argument θ):

$$\partial_\omega (DK) = (\nabla F_\mu \circ K) DK$$

or

$$\partial_\omega (DK) = A\, DK. \tag{A.5}$$

The second half columns M_2 are given by

$$M_2(\theta) \equiv \partial_\omega (J DK N) - A J DK N;$$

due to (A.5) we obtain

$$M_2 = J\partial_\omega(DK)N + JDK\,\partial_\omega(N) - AJDKN$$

$$= JADKN + JDK\,\partial_\omega(N) - AJDKN$$

$$= (JA - AJ)DKN + JDK\,\partial_\omega(N).$$

The conformally symplectic condition for flows can be written as

$$(\nabla F_\mu \circ K)J + J(\nabla F_\mu \circ K)^T = -\lambda J. \tag{A.6}$$

From (A.6) one obtains that $AJ = -JA^T - \lambda J$, so that

$$M_2 = J(A + A^T)DKN + \lambda JDKN + JDK\,\partial_\omega(N).$$

Moreover, using $NDK^T DK = \mathrm{Id}$, it follows that

$$\partial_\omega(N) = -NDK^T(A + A^T)DKN; \tag{A.7}$$

in fact, from

$$\partial_\omega(N)\,DK^T DK + N\partial_\omega(DK^T)DK + NDK^T\,\partial_\omega(DK) = 0$$

and being $\partial_\omega(DK) = ADK$, $\partial_\omega(DK^T) = DK^T A^T$, one has

$$\partial_\omega(N)N^{-1} + NDK^T A^T DK + NDK^T ADK = 0,$$

which gives (A.7). Since the vectors $\{\frac{\partial K(\theta)}{\partial \theta_i}, J\frac{\partial K(\theta)}{\partial \theta_i}\}_{i=1,\dots,n}$ form a basis of $\mathbb{R}^{2n}$, we can find $n \times n$ matrices $S(\theta)$, $T(\theta)$, such that

$$M_2(\theta) = DK(\theta)S(\theta) + JDK(\theta)N(\theta)T(\theta). \tag{A.8}$$

Multiplying by $-DK^T J$, one obtains

$$T = -DK^T JM_2 = \lambda\,\mathrm{Id}.$$

Multiplying (A.8) by $N(\theta)DK(\theta)^T$ one obtains

$$S = -NDK^T[AJDKN - \partial_\omega(JDKN)]$$

$$= -NDK^T(AJ - JA)DKN,$$

being $DK^T JDK = 0$ and

$$\partial_\omega(JDKN) = JADKN + JDK\,\partial_\omega(N).$$

Since (A.6) implies that $AJ = -JA^T - \lambda J$, namely $AJ - JA = -J(A + A^T) - \lambda J$, one has

$$S = NDK^T J(A + A^T)DKN,$$

being $NDK^T JDKN = 0$, so that we finally obtain (A.4). $\quad\square$

Assume that (A.2) is satisfied up to an error term E, say

$$\partial_\omega K - F_\mu \circ K = E.$$

As in Section 6.1, we denote by Δ, σ the corrections to K, μ and we write the linearized equation as

$$\partial_\omega \Delta - A\Delta - (\nabla_\mu F_\mu \circ K)\sigma = -E.$$

To solve this equation we make the change of variables

$$\Delta(\theta) \equiv M(\theta)W(\theta),$$

with M as in (A.3) and W to be determined as follows. We aim to find the corrections Δ, σ, such that the error of the approximate linearized equation is quadratically reduced. Let us assume that

$$\partial_\omega M(\theta) - A(\theta)M(\theta) = M(\theta)\begin{pmatrix} 0 & S(\theta) \\ 0 & \lambda\,\mathrm{Id} \end{pmatrix} + R,$$

for some error function $R = R(\theta)$. The iterative step is obtained by solving the following equation (A.9), where the term RW is neglected:

$$\partial_\omega W + \begin{pmatrix} 0 & S(\theta) \\ 0 & \lambda\,\mathrm{Id} \end{pmatrix} W = -\tilde{E} + M^{-1}(\nabla_\mu F_\mu \circ K)\sigma, \tag{A.9}$$

with $\tilde{E} \equiv M^{-1}E$. We have thus obtained differential equations with constant coefficients, which can be solved using Fourier methods. Let

$$\tilde{A} \equiv -M^{-1}(\nabla_\mu F_\mu \circ K);$$

denoting the components of W as (W_1, W_2), we have

$$\partial_\omega W_1(\theta) + S(\theta)W_2(\theta) = -\tilde{E}_1(\theta) - \tilde{A}_1(\theta)\sigma,$$
$$\partial_\omega W_2(\theta) + \lambda W_2(\theta) = -\tilde{E}_2(\theta) - \tilde{A}_2(\theta)\sigma, \tag{A.10}$$

where $\tilde{E} \equiv (\tilde{E}_1, \tilde{E}_2)$, $\tilde{A} \equiv [\tilde{A}_1|\tilde{A}_2]$ with $\tilde{A}_1$, $\tilde{A}_2$ being $n \times n$ matrices. For any $|\lambda| \neq 0$ the second equation can always be solved, while for any λ the first equation involves small divisors and the right hand side of (A.10) must have zero average. Using the same notation as in Section 6.2.1, namely setting $W_1 = \overline{W}_1 + (W_1)^0$, $W_2 = \overline{W}_2 + (W_2)^0$, $(W_2)^0 = (B_a)^0 + \sigma(B_b)^0$, one is led to choose $\overline{W}_2$, σ as the solution of

$$\begin{pmatrix} \overline{S} & \overline{S(B_b)^0} + \overline{\tilde{A}_1} \\ \lambda\,\mathrm{Id} & \overline{\tilde{A}_2} \end{pmatrix} \begin{pmatrix} \overline{W}_2 \\ \sigma \end{pmatrix} = \begin{pmatrix} -\overline{S(B_a)^0} - \overline{\tilde{E}_1} \\ -\overline{\tilde{E}_2} \end{pmatrix}, \tag{A.11}$$

provided the following non-degeneracy condition is satisfied

$$\det \begin{pmatrix} \overline{S} & \overline{S(B_b)^0} + \overline{\tilde{A}_1} \\ \lambda\,\mathrm{Id} & \overline{\tilde{A}_2} \end{pmatrix} \neq 0.$$

To conclude, we summarize the algorithm for computing the improved approximation for flows as follows.

Algorithm 66. *Given $K : \mathbb{T}^n \to \mathcal{M}$, $\mu \in \mathbb{R}^n$ perform the following computations*:

  1) $a \leftarrow \partial_\omega K$

  2) $b \leftarrow F_\mu \circ K$

  3) $E \leftarrow a - b$

 4.1) $\alpha \leftarrow DK(\theta)$

 4.2) $N \leftarrow (\alpha^T \alpha)^{-1}$

 4.3) $\beta \leftarrow \alpha N$

 4.4) $\gamma \leftarrow J\beta$

 4.5) $M \leftarrow [\alpha \mid \gamma]$

 4.6) $\tilde{E} \leftarrow M^{-1} E$

 4.7) $A \leftarrow \nabla F_\mu \circ K$

 4.8) $G \leftarrow \nabla_\mu F_\mu \circ K$

 4.9) $S \leftarrow N\alpha^T J (A + A^T) \beta$

4.10) $\tilde{A} \leftarrow -M^{-1} G$

  5) $(B_a)^0$ solves $\partial_\omega (B_a)^0 + \lambda(B_a)^0 = -(\tilde{E}_2)^0$, $(B_b)^0$ solves $\partial_\omega(B_b)^0 + \lambda(B_b)^0 = -(\tilde{A}_2)^0$

  6) *Find* $\overline{W}_2, \sigma$ *solving* (A.11)

  7) $(W_2)^0 = (B_a)^0 + \sigma (B_b)^0$

  8) $W_2 = (W_2)^0 + \overline{W}_2$

  9) W_1 solves $\partial_\omega W_1(\theta) + S(\theta) W_2(\theta) = -\tilde{E}_1(\theta) - \tilde{A}_1(\theta)\sigma$

10) $K \leftarrow K + MW, \mu \leftarrow \mu + \sigma.$

References

[1] A.A. Agrachëv, Invariant Lagrangian submanifolds of dissipative systems, Uspekhi Mat. Nauk 65 (5(395)) (2010) 185–186.

[2] L. Ahlfors, Complex Analysis, McGraw–Hill, 1976.

[3] V.I. Arnol'd, V.V. Kozlov, A.I. Neishtadt, Mathematical aspects of classical and celestial mechanics, in: Dynamical Systems, III, Springer-Verlag, Berlin, 1993, pp. 1–291, vii–xiv.

[4] R. Abraham, J.E. Marsden, Foundations of Mechanics, Benjamin/Cummings, Reading, MA, 1978.

[5] V.I. Arnol'd, Proof of a theorem of A.N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations, Russian Math. Surveys 18 (5) (1963) 9–36.

[6] J. Appell, P.P. Zabrejko, Nonlinear Superposition Operators, Cambridge Tracts in Math., vol. 95, Cambridge University Press, Cambridge, 1990.

[7] A. Banyaga, Some properties of locally conformal symplectic structures, Comment. Math. Helv. 77 (2) (2002) 383–398.

[8] Richard Bellman, Dynamic Programming. Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 2010, reprint of the 1957 edition, with a new introduction by Stuart Dreyfus.

[9] Alain Bensoussan, Perturbation Methods in Optimal Control, Wiley/Gauthier-Villars Ser. Modern Appl. Math., John Wiley & Sons Ltd., Chichester, 1988, translated from the French by C. Tomson.

[10] H.W. Broer, G.B. Huitema, M.B. Sevryuk, Quasi-Periodic Motions in Families of Dynamical Systems. Order Amidst Chaos, Springer-Verlag, Berlin, 1996.

[11] H.W. Broer, G.B. Huitema, F. Takens, B.L.J. Braaksma, Unfoldings and bifurcations of quasi-periodic tori, Mem. Amer. Math. Soc. 83 (421) (1990), viii+175.

[12] R.P. Brent, H.T. Kung, Fast algorithms for manipulating formal power series, J. Assoc. Comput. Mach. 25 (4) (1978) 581–595.

[13] G. Bande, D. Kotschick, Moser stability for locally conformally symplectic structures, Proc. Amer. Math. Soc. 137 (7) (2009) 2419–2424.

[14] A. Celletti, L. Chierchia, Quasi-periodic attractors in celestial mechanics, Arch. Ration. Mech. Anal. 191 (2) (2009) 311–345.

[15] R. Calleja, A. Celletti, Breakdown of invariant attractors for the dissipative standard map, Chaos 20 (1) (2010) 013121.

[16] R. Calleja, A. Celletti, R. de la Llave, Local behavior near quasi-periodic solutions of conformally symplectic systems, J. Dynam. Differential Equations (2013), in press.

[17] Renato Calleja, Alessandra Celletti, Rafael de la Llave, KAM estimates for the dissipative standard map, 2013, in preparation.

[18] Renato Calleja, Alessandra Celletti, Rafael de la Llave, Whiskered tori for conformally symplectic systems, 2013, in preparation.

[19] R. Calleja, R. de la Llave, Fast numerical computation of quasi-periodic equilibrium states in 1D statistical mechanics, including twist maps, Nonlinearity 22 (6) (2009) 1311–1336.

[20] R. Calleja, R. de la Llave, Computation of the breakdown of analyticity in statistical mechanics models: numerical results and a renormalization group explanation, J. Stat. Phys. 141 (6) (2010) 940–951.

[21] R. Calleja, R. de la Llave, A numerically accessible criterion for the breakdown of quasi-periodic solutions and its rigorous justification, Nonlinearity 23 (9) (2010) 2029–2058.

[22] A. Celletti, Periodic and quasi-periodic attractors of weakly-dissipative nearly-integrable systems, Regul. Chaotic Dyn. 14 (1) (2009) 49–63.

[23] A. Celletti, Stability and Chaos in Celestial Mechanics, Springer-Verlag, Berlin, 2010 (published in association with Praxis Publ. Ltd., Chichester).

[24] R. Calleja, J. -Ll Figueras, Collision of invariant bundles of quasi-periodic attractors in the dissipative standard map, Chaos 22 (3) (2012) 033114.

[25] B.V. Chirikov, A universal instability of many-dimensional oscillator systems, Phys. Rep. 52 (5) (1979) 264–379.

[26] J.J. Dongarra, J. Du Croz, S. Hammarling, R.J. Hanson, An extended set of FORTRAN basic linear algebra subprograms, ACM Trans. Math. Software 14 (1988) 1–17.

[27] A. Delshams, R. de la Llave, T.M. Seara, A geometric mechanism for diffusion in hamiltonian systems overcoming the large gap problem: Heuristics and rigorous verification on a model, Mem. Amer. Math. Soc. 179 (844) (2006) 1–141.

[28] R. de la Llave, Smooth conjugacy and S–R–B measures for uniformly and non-uniformly hyperbolic systems, Comm. Math. Phys. 150 (2) (1992) 289–320.

[29] R. de la Llave, A tutorial on KAM theory, in: Smooth Ergodic Theory and Its Applications, Seattle, WA, 1999, in: Proc. Sympos. Pure Math., vol. 69, Amer. Math. Soc., Providence, RI, 2001, pp. 175–292.

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

[31] R. de la Llave, R. Obaya, Regularity of the composition operator in spaces of Hölder functions, Discrete Contin. Dyn. Syst. 5 (1) (1999) 157–184.

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

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

[34] C.P. Dettmann, G.P. Morris, Proof of Lyapunov exponent pairing for systems at constant kinetic energy, Phys. Rev. E 53 (6) (2006) R5545–R5548.

[35] Corrado Falcolini, Rafael de la Llave, A rigorous partial justification of Greene's criterion, J. Stat. Phys. 67 (3–4) (1992) 609–643.

[36] N. Fenichel, Asymptotic stability with rate conditions, Indiana Univ. Math. J. 23 (1973–1974) 1109–1137.

[37] N. Fenichel, Asymptotic stability with rate conditions. II, Indiana Univ. Math. J. 26 (1) (1977) 81–93.

[38] A. González Enriquez, R. de la Llave, Analytic smoothing of geometric maps with applications to KAM theory, J. Differential Equations 245 (2008) 1243–1298.

[39] A. González-Enríquez, J. Vano, An estimate of smoothing and composition with applications to conjugation problems, J. Dynam. Differential Equations 20 (1) (2008) 239–270.

[40] J. Hadamard, Sur le module maximum d'une fonction et de ses derives, Bull. Soc. Math. France 42 (1898) 68–72.

[41] À. Haro, The primitive function of an exact symplectomorphism, Nonlinearity 13 (5) (2000) 1483–1500.

[42] A. Haro, Automatic differentiation tools in computational dynamical systems, 2013, in preparation.

[43] A. Haro, R. de la Llave, New mechanisms for lack of equiparation of energy, Phys. Rev. Lett. 89 (7) (2000) 1859–1862.

[44] À. Haro, R. de la Llave, Manifolds on the verge of a hyperbolicity breakdown, Chaos 16 (1) (2006) 013120, 8.

[45] À. Haro, R. de la Llave, A parameterization method for the computation of invariant tori and their whiskers in quasi-periodic maps: numerical algorithms, Discrete Contin. Dyn. Syst. Ser. B 6 (6) (2006) 1261–1300 (electronic).

[46] Gemma Huguet, Rafael de la Llave, Yannick Sire, Fast numerical algorithms for the computation of invariant tori in hamiltonian systems, Discrete Contin. Dyn. Syst. Ser. A 17 (4) (2011) 1309–1353.

[47] M.-R. Herman, Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations, Publ. Math. Inst. Hautes Études Sci. (49) (1979) 5–233.

[48] M.-R. Herman, Une méthode pour minorer les exposants de Lyapounov et quelques exemples montrant le caractère local d'un théorème d'Arnolḋ et de Moser sur le tore de dimension 2, Comment. Math. Helv. 58 (3) (1983) 453–502.

[49] M.W. Hirsch, C.C. Pugh, M. Shub, Invariant Manifolds, Lecture Notes in Math., vol. 583, Springer-Verlag, Berlin, 1977.

[50] Stefan Haller, Tomasz Rybicki, On the group of diffeomorphisms preserving a locally conformal symplectic structure, Ann. Global Anal. Geom. 17 (5) (1999) 475–502.

[51] Renato Iturriaga, Héctor Sánchez-Morgado, Limit of the infinite horizon discounted Hamilton–Jacobi equation, Discrete Contin. Dyn. Syst. Ser. B 15 (3) (2011) 623–635.

[52] À. Jorba, R. de la Llave, M. Zou, Lindstedt series for lower-dimensional tori, in: Hamiltonian Systems with Three or More Degrees of Freedom, S'Agaró, 1995, Kluwer Acad. Publ., Dordrecht, 1999, pp. 151–167.

[53] Y. Katznelson, D. Ornstein, The differentiability of the conjugation of certain diffeomorphisms of the circle, Ergodic Theory Dynam. Systems 9 (4) (1989) 643–680.

[54] A. Kolmogoroff, On inequalities between the upper bounds of the successive derivatives of an arbitrary function on an infinite interval, Amer. Math. Soc. Transl. (4) (1949) 19.

[55] A.N. Kolmogorov, On conservation of conditionally periodic motions for a small change in Hamilton's function, Dokl. Akad. Nauk SSSR (N.S.) 98 (1954) 527–530; English translation in: Stochastic Behavior in Classical and Quantum Hamiltonian Systems, Volta Memorial Conf., Como, 1977, in: Lecture Notes in Phys., vol. 93, Springer-Verlag, Berlin, 1979, pp. 51–56.

[56] S.G. Krantz, Function Theory of Several Complex Variables, AMS Chelsea Publishing, Providence, RI, 2001, reprint of the 1992 edition.

[57] E. Landau, Einige ungleichungen für zweimal differentiarbare funktionen, Proc. Lond. Math. Soc. 13 (1896) 43–58.

[58] Oscar E. Lanford III, Computer-assisted proofs in analysis, in: Proceedings of the International Congress of Mathematicians, vols. 1, 2, Berkeley, Calif., 1986, Amer. Math. Soc., Providence, RI, 1987, pp. 1385–1394.

[59] Héctor E. Lomelí, Renato Calleja, Heteroclinic bifurcations and chaotic transport in the two-harmonic standard map, Chaos 16 (2) (2006) 023117, 8.

[60] M. Lanza de Cristoforis, Higher order differentiability properties of the composition and of the inversion operator, Indag. Math. 5 (4) (1994) 457–482.

[61] M. Levi, Qualitative analysis of the periodically forced relaxation oscillations, Mem. Amer. Math. Soc. 32 (244) (1981), vi+147.

[62] S. Łojasiewicz Jr., E. Zehnder, An inverse function theorem in Fréchet-spaces, J. Funct. Anal. 33 (2) (1979) 165–174.

[63] Ramon E. Moore, R. Baker Kearfott, Michael J. Cloud, Introduction to Interval Analysis, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009.

[64] Ramon E. Moore, Interval Analysis, Prentice Hall Inc., Englewood Cliffs, NJ, 1966.

[65] Ramon E. Moore, Computational Functional Analysis, Ellis Horwood Ser. Math. Appl., Ellis Horwood Ltd., Chichester, 1985.

[66] J. Moser, On the integrability of area preserving Cremona mappings near an elliptic fixed point, Bol. Soc. Mat. Mexicana (2) 5 (1960) 176–180.

[67] J. Moser, On invariant curves of area-preserving mappings of an annulus, Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II (1962) 1–20.

[68] J. Moser, A rapidly convergent iteration method and non-linear partial differential equations. I, Ann. Sc. Norm. Super. Pisa 3 (20) (1966) 265–315.

[69] J. Moser, A rapidly convergent iteration method and non-linear differential equations. II, Ann. Sc. Norm. Super. Pisa 3 (20) (1966) 499–535.

[70] J. Moser, Convergent series expansions for quasi-periodic motions, Math. Ann. 169 (1967) 136–176.

[71] J. Moser, Stable and Random Motions in Dynamical Systems, Princeton University Press, Princeton, NJ, 1973.

[72] Robert McLachlan, Matthew Perlmutter, Conformal Hamiltonian systems, J. Geom. Phys. 39 (4) (2001) 276–300.

[73] J. Nash, The imbedding problem for Riemannian manifolds, Ann. of Math. (2) 63 (1956) 20–63.

[74] D.A. Rand, Existence, nonexistence and universal breakdown of dissipative golden invariant tori. I. Golden critical circle maps, Nonlinearity 5 (3) (1992) 639–662.

[75] D.A. Rand, Existence, nonexistence and universal breakdown of dissipative golden invariant tori. II. Convergence of renormalization for mappings of the annulus, Nonlinearity 5 (3) (1992) 663–680.

[76] D.A. Rand, Existence, nonexistence and universal breakdown of dissipative golden invariant tori. III. Invariant circles for mappings of the annulus, Nonlinearity 5 (3) (1992) 681–706.

[77] 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, WA, 1974, in: Lecture Notes in Phys., vol. 38, Springer-Verlag, Berlin, 1975, pp. 598–624.

[78] H. Rüssmann, On optimal estimates for the solutions of linear difference equations on the circle, Celestial Mech. 14 (1) (1976) 33–37.

[79] H. Rüssmann, On the one-dimensional Schrödinger equation with a quasiperiodic potential, in: Nonlinear Dynamics, Internat. Conf., New York, 1979, in: Ann. N.Y. Acad. Sci., vol. 357, New York Acad. Sci., New York, 1980, pp. 90–107.

[80] M. Saprykina, Domain of analyticity of normalizing transformations, Nonlinearity 19 (7) (2006) 1581–1599.

[81] J.T. Schwartz, Nonlinear Functional Analysis, Gordon and Breach, New York, 1969.

[82] C.L. Siegel, Über die Existenz einer Normalform analytischer Hamiltonscher Differentialgleichungen in der Nähe einer Gleichgewichtslösung, Math. Ann. 128 (1954) 144–170.

[83] Ya.G. Sinai, K.M. Khanin, Smoothness of conjugacies of diffeomorphisms of the circle with rotations, Uspekhi Mat. Nauk 44 (1(265)) (1989) 57–82, 247; English translation in: Russian Math. Surveys 44 (1) (1989) 69–99.

[84] Letizia Stefanelli, Ugo Locatelli, Kolmogorov's normal form for equations of motion with dissipative effects, Discrete Contin. Dyn. Syst. Ser. B 17 (7) (2012) 2561–2593.

[85] M.E. Taylor, Partial Differential Equations. III. Nonlinear Equations, Springer-Verlag, New York, 1997, Corrected reprint of the 1996 original.

[86] W. Thirring, Classical Mathematical Physics, third edition, Springer-Verlag, New York, 1997.

[87] Izu Vaisman, Locally conformal symplectic manifolds, Int. J. Math. Math. Sci. 8 (3) (1985) 521–536.

[88] J. Vano, A Nash–Moser implicit function theorem with Whitney regularity and applications, PhD thesis, University of Texas, 2002.

[89] M.P. Wojtkowski, C. Liverani, Conformally symplectic dynamics and symmetry of the Lyapunov spectrum, Comm. Math. Phys. 194 (1) (1998) 47–60.

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

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

[92] E. Zehnder, Moser's implicit function theorem in the framework of analytic smoothing, Math. Ann. 219 (2) (1976) 105–121.